95

Advances in Biochemical Engineering/Biotechnology

Series Editor: T. Scheper

Advances in Biochemical Engineering/Biotechnology

Series Editor: T. Scheper

Recently Published and Forthcoming Volumes

Microscopy Techniques
Volume Editor: Rietdorf, J.
Vol. 95, 2005

Regenerative Medicine II
Clinical and Preclinical Applications
Volume Editor: Yannas, I.V.
Vol. 94, 2005

Regenerative Medicine I
Theories, Models and Methods
Volume Editor: Yannas, I.V.
Vol. 93, 2005

Technology Transfer in Biotechnology
Volume Editor: Kragl, U.
Vol. 92, 2005

Recent Progress of Biochemical and Biomedical Engineering in Japan II
Volume Editor: Kobayashi, T.
Vol. 91, 2004

Recent Progress of Biochemical and Biomedical Engineering in Japan I
Volume Editor: Kobayashi, T.
Vol. 90, 2004

Physiological Stress Responses in Bioprocesses
Volume Editor: Enfors, S.-O.
Vol. 89, 2004

Molecular Biotechnology of Fungal β-Lactam Antibiotics and Related Peptide Synthetases
Volume Editor: Brakhage, A.
Vol. 88, 2004

Biomanufacturing
Volume Editor: Zhong, J.-J.
Vol. 87, 2004

New Trends and Developments in Biochemical Engineering
Vol. 86, 2004

Biotechnology in India II
Volume Editors: Ghose, T.K., Ghosh, P.
Vol. 85, 2003

Biotechnology in India I
Volume Editors: Ghose, T.K., Ghosh, P.
Vol. 84, 2003

Proteomics of Microorganisms
Volume Editors: Hecker, M., Müllner, S.
Vol. 83, 2003

Biomethanation II
Volume Editor: Ahring, B.K.
Vol. 82, 2003

Biomethanation I
Volume Editor: Ahring, B.K.
Vol. 81, 2003

Process Integration in Biochemical Engineering
Volume Editors: von Stockar, U., van der Wielen, L.A.M.
Vol. 80, 2003

Microbial Production of L-Amino Acids
Volume Editors: Faurie, R., Thommel J.
Vol. 79, 2003

Phytoremediation
Volume Editor: Tsao, D.T.
Vol. 78, 2003

Chip Technology
Volume Editor: Hoheisel, J.
Vol. 77, 2002

Modern Advances in Chromatography
Volume Editor: Freitag, R.
Vol. 76, 2002

History and Trends in Bioprocessing and Biotransformation
Vol. 75, 2002

Microscopy Techniques

Volume Editor: Jens Rietdorf

With contributions by
T. W. J. Gadella · R. Gräf · A. B. Houtsmuller · T. Kohl · K. Miura
A. Miyawaki · H. Mizuno · E. B. van Munster · T. Nagai · J. Rietdorf
P. Schwille · J.-B. Sibarita · C.-K. Sun · M. Ueda · T. Wazawa
T. Zimmermann

Springer

Advances in Biochemical Engineering/Biotechnology reviews actual trends in modern biotechnology. Its aim is to cover all aspects of this interdisciplinary technology where knowledge, methods and expertise are required for chemistry, biochemistry, micro-biology, genetics, chemical engineering and computer science. Special volumes are dedicated to selected topics which focus on new biotechnological products and new processes for their synthesis and purification. They give the state-of-the-art of a topic in a comprehensive way thus being a valuable source for the next 3–5 years. It also discusses new discoveries and applications.

Special volumes are edited by well known guest editors who invite reputed authors for the review articles in their volumes.

In references *Advances in Biochemical Engineering/Biotechnology* is abbreviated as *Adv Biochem Engin/Biotechnol* as a journal.

Visit the Adv Biochem Engin/Biotechnol home page at springeronline.com

Library of Congress Control Card Number 2004113697

ISSN 0724-6145
ISBN-10 3-540-23698-8 **Springer Berlin Heidelberg New York**
ISBN-13 978-3-540-23698-6 **Springer Berlin Heidelberg New York**
DOI 10.1007/b14097

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in Germany

Typesetting: Fotosatz-Service Köhler GmbH, Würzburg
Cover: KünkelLopka GmbH, Heidelberg; design & production GmbH, Heidelberg

Printed on acid-free paper 02/3141xv – 5 4 3 2 1 0

Series Editor

Prof. Dr. T. Scheper

Institute of Technical Chemistry
University of Hannover
Callinstraße 3
30167 Hannover, Germany
scheper@iftc.uni-hannover.de

Volume Editor

Dr. J. Rietdorf

European Molecular Biology Laboratory
Advanced Light Microscopy Facility
Meyerhofstraße 1
69117 Heidelberg, Germany
jens.rietdorf@embl-heidelberg.de

Editorial Board

Advances in Biochemical Engineering/Biotechnology Also Available Electronically

For all customers who have a standing order to Advances in Biochemical Engineering/Biotechnology, we offer the electronic version via SpringerLink free of charge. Please contact your librarian who can receive a password for free access to the full articles by registering at:

springerlink.com

If you do not have a subscription, you can still view the tables of contents of the volumes and the abstract of each article by going to the SpringerLink Homepage, clicking on "Browse by Online Libraries", then "Chemical Sciences", and finally choose Advances in Biochemical Engineering/Biotechnology.

You will find information about the

– Editorial Board
– Aims and Scope
– Instructions for Authors
– Sample Contribution

at springeronline.com using the search function.

Attention all Users of the "Springer Handbook of Enzymes"

Information on this handbook can be found on the internet at **springeronline.com**

A complete list of all enzyme entries either as an alphabetical Name Index or as the EC-Number Index is available at the above mentioned URL. You can download and print them free of charge.

A complete list of all synonyms (more than 25,000 entries) used for the enyzmes is available in print form (ISBN 3-540-41830-X).

Preface

Advances in microscopy techniques have recently had a tremendous impact on research in biochemistry and molecular biology. While the study of biochemical reactions was formerly confined to cuvette-based measurements of purified biomolecules, it is now possible for investigators to make such analyses inside the complex environment of living cells and organisms.

While some of these advanced techniques were in fact developed some time ago, it is only now that they are becoming available to a wider range of researchers as components of commercial equipment. As a consequence, there is a strong demand for the ability to help individuals from a biomedical, rather than from an instrumentation background, in order for them to make the best use of these new tools in their research. To meet these requirements, the contributions in this book mainly stem from instructors present at a series of very successful training courses held in the Advanced Light Microscopy Facility at the European Molecular Biology Laboratory over the past 5 years.

Inevitably, in a project of this size, it has been impossible to cover all the ongoing developments of techniques. This special issue emphasises the analysis of dynamic events where recent advances in the respective microscopy field have been of particular relevance to the biomedical community. Microscopy topics covered by this book are Higher Harmonic Generation (HHG, Chapter 2), Spinning Disk Confocal (Chapter 3), Total Internal Reflection Fluorescence (TIRF, Chapter 4), Fluorescence Correlation Spectroscopy (FCS, Chapter 5), Fluorescence Lifetime Imaging (FLIM, Chapter 6), Fluorescence Recovery after Photobleaching (FRAP, Chapter 7), Deconvolution (Chapter 8) and Spectral Imaging and Linear Unmixing (Chapter 9). Resonance Energy Transfer (FRET) is not addressed in a single chapter but is covered by parts of other contributions, e.g. chapters 1, 4, 6 and 9. It is important to realise that not only the microscopic techniques themselves that are evolving significantly, but likewise the stains and probes (Engineering of Fluorescent Proteins, chapter 1) and the computational technologies to improve and analyse multi-dimensional image data (Tracking Movement in Cell Biology, chapter 10).

I thank all the authors who took on the time consuming task of preparing their contributions and agreeing to share their expertise in this volume, and, in addition, the staff at Springer Verlag for their help at all stages of preparation and production.

Heidelberg, January 2005 Jens Rietdorf

Contents

Adv Biochem Engin/Biotechnol (2005) 95: 1–15
DOI 10.1007/b102208

Engineering Fluorescent Proteins

Atsushi Miyawaki (✉) · Takeharu Nagai · Hideaki Mizuno

Laboratory for Cell Function Dynamics, Advanced Technology Development Group, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako-city, Saitama, 351-0198 Japan
matsushi@brain.riken.jp

Abstract Green fluorescent protein from the jellyfish *Aequorea victora* (GFP) and GFP-like proteins from Anthozoa species encode light-absorbing chromophores intrinsically within their respective protein sequences. Recent studies have made progress in obtaining bright variants of these proteins which develop chromophores quickly and efficiently, as well as novel fluorescent proteins that photoactivate or photoconvert, i.e., become fluorescent or change colors upon illumination at specific wavelengths. Also, monomeric versions of these proteins have been engineered for fusion protein applications. Simple GFP variants and circularly permuted GFP variants have been used to develop fluorescent probes that sense physiological signals such as membrane potential and concentrations of free calcium. Further molecular characterization of the structure and maturation of these proteins is in progress, aimed at providing information for rational design of variants with desired fluorescence properties.

Keywords Green fluorescent protein · GFP-like proteins · Chromophore · Photoactivation · Photoconversion · Calcium imaging

List of Abbreviations

FP	Fluorescent protein
GFP	Green fluorescent protein
PA-GFP	Photoactivatable GFP
RFP	Red fluorescent protein
YFP	Yellow fluorescent protein

1 Engineering for Brighter Fluorescence

1.1 Fluorescence Development

Green fluorescent protein (GFP), originally isolated from the jellyfish *Aequorea victoria* (*Aequorea* GFP), has been the subject of continued interest since it was cloned in 1992 [1, 2]. Moreover, the discovery of novel GFP-like proteins from Anthozoans (coral animals) has significantly expanded the range of colors available for cell biological applications [3, 4]. Here, the term 'fluorescent protein' (FP) will be used to describe a protein that can become spontaneously fluorescent. Due to their autocatalytic chromophore generation, and the possibility of spectral manipulation of these proteins via mutagenesis, FPs of this sort make attractive tools for a myriad of biological applications. After protein synthesis, many FPs mature slowly through a multi-step process that consists of folding, chromophore formation, and chromophore modification. *Aequorea* GFP, for example, requires folding and chromophore formation processes, which for this protein may be interdependent [5]. Sometime during or after the protein adopts its native conformation, an internal tripeptide, Ser^{65}-Tyr^{66}-Gly^{67}, forms a chromophore, 4-(*p*-hydroxybenzylidene)-5-imidazolinone, by nucleophilic attack of Gly^{67}-Nα on the carbonyl group of Ser^{65}, followed by dehydration and oxidation of the α-β bond of Tyr^{66} (Fig. 1A). The wavelengths of light absorbed and emitted by the chromophore depend on the local chemistry, and as a result substitutions in or near the chromophore may produce large changes in fluorescence. Yellow fluorescent protein (YFP), a yellow variant of GFP, has a Tyr residue substituted at position Thr^{203} (T203Y). The π-π interaction between Tyr^{203} and the chromophore phenol ring most likely reduces the excited state energy, thereby increasing the excitation and emission wavelengths (Fig. 1B) [2]. In addition to folding and chromophore formation, some GFP-like proteins undergo a process of chromophore modification in order to reach a mature state with the desired fluorescence properties. This event takes place in the final folded molecule. For example, DsRed is a GFP-like protein that fluoresces first green and then red, implying that its chromophore structure is modified during maturation. Recent structural studies have shown that a tripeptide in DsRed (Gln^{66}-Tyr^{67}-Gly^{68}) analogous to the chromophore-forming sequence of *Aequorea* GFP forms a 4-(*p*-hydroxybenzylidene)-5-imidazolinone chro-

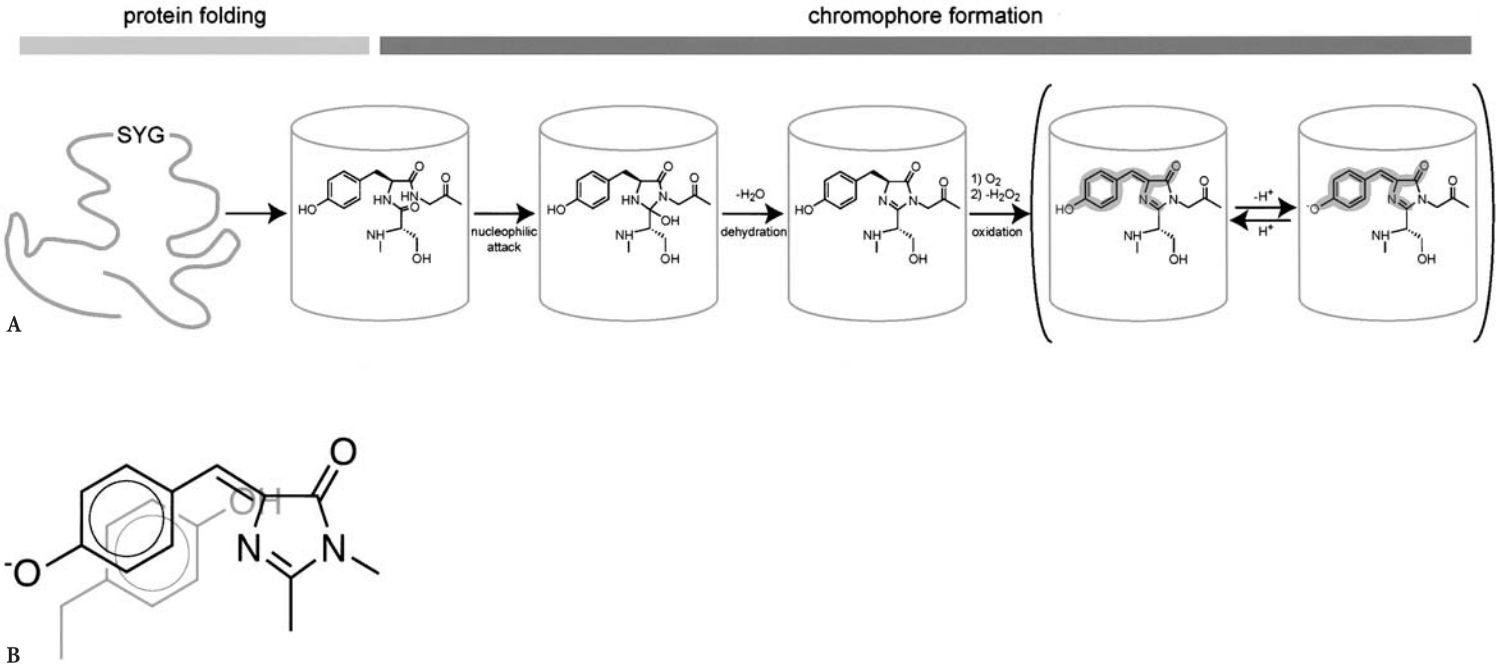


Fig. 1 **A** Diagram representing *Aequorea* GFP folding and chromophore formation. The β-can structure represents the native conformation of the protein, while the denatured form is depicted as an irregular chain. π-conjugation for visible-light absorption is indicated in *gray*. **B** Chromophore structure of YFP. Reprinted with permission from Miyawaki et al., Curr. Opin. Chem. Biol. 7, 558 (2003). Copyright © 2003 Elsevier

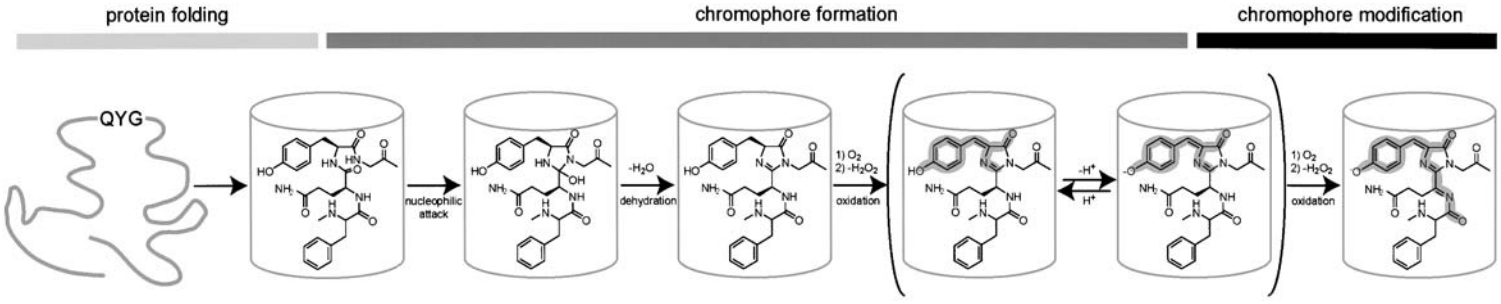


Fig. 2 Schematic of folding, chromophore formation, and chromophore modification of DsRed. The β-can structure represents the native conformation of the protein, while the denatured form is shown as an irregular chain. It is not known at which step oligomerization occurs. π-conjugation for visible-light absorption is indicated in *gray*. Reprinted with permission from Miyawaki et al., Curr. Opin. Chem. Biol. 7, 560 (2003).

mophore identical to GFP to generate its green intermediate [6]. Subsequently, the Cα-Nα bond of Gln^{66} oxidizes as the protein matures to become red (Fig. 2). The following section describes recent improvements in both YFP and DsRed maturation speed and/or fluorescence efficiency.

1.2 Improvements in Maturation of YFP

Two bright versions of YFP, Venus [7] and citrine [8], have recently been developed that mature quickly and efficiently, allowing for immediate detection of fluorescence signals after gene-introduction. Venus has a new mutation, F46L, and four common folding mutations, F64L, M153T, V163A, and S175G. The superiority of Venus to EYFP in terms of maturation is best observed following the re-naturation/re-oxidation profile of each protein. The chromophores of urea-denatured Venus and EYFP were reduced with 5 mmol/l dithionite [9], and re-naturation and re-oxidation were initiated by dilution at 37 °C (Fig. 3). Since oxidation is the rate-limiting step, the observed overall rate of fluorescence recovery under these circumstances should represent the rate of oxidation of the cyclized chromophore. The speed and yield of the re-naturation from denatured/reduced protein was significantly improved in Venus, which contains the F46L mutation (Fig. 3). To examine whether this mutation facilitates the maturation of any other *Aequorea* GFP variants, we introduced F46L into enhanced versions of blue, cyan, and green fluorescent proteins (Clontech). None of these mutants, however, displayed enhanced maturation

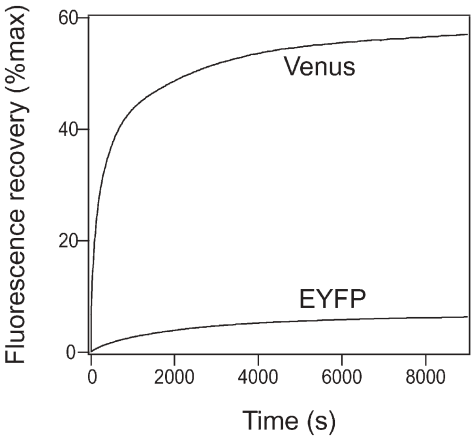


Fig. 3 Time course of fluorescence recovery of Venus and EYFP from their denatured/reduced state. For preparation of denatured/reduced protein samples, 5 mmol/l dithionite was added to denaturation buffer (8 mol/l urea and 1 mmol/l DTT). Recovery of fluorescence was initiated by a 100-fold dilution into re-naturation buffer (35 mmol/l KCl, 2 mmol/l $MgCl_2$, 50 mmol/l tris pH 7.5, 1 mmol/l DTT) at 37 °C. The emission at 530 nm was monitored by excitation at 515 nm. Reprinted with permission from Miyawaki et al., Curr. Opin. Chem. Biol. 7, 559 (2003). Copyright © 2003 Elsevier

[7]. On the other hand, the citrine version of YFP possesses a single mutation, Q69M [8].

Citrine and Venus have two main features in common. First, they both are yellow variants of GFP which have mutations, Q69M and F46L, respectively, which are specific to YFP. Second, they have similar backgrounds, arising from concerted efforts to improve the genetically encoded Ca^{2+} indicators, camgaroo [10] and pericam [11], respectively. As will be discussed below, camgaroo is a YFP variant into which calmodulin (CaM) has been inserted in place of Tyr145 of YFP; pericam was created by inserting a circularly-permuted YFP (cpYFP) between CaM and M13 (a CaM-target peptide). Since the original versions of camgaroo and pericam showed poor expression at 37 °C, their cDNAs were randomly mutated and screened in *E. coli* for maximal fluorescence. The Q69M and F46L mutations were found to be the most beneficial for improving the brightness of camgaroo and pericam, respectively, and these mutations were subsequently transferred into YFP to generate citrine and Venus. To engineer brighter GFP variants, therefore, it appears that the most efficient method may be to choose dimmer variants as parental constructs for mutagenesis. There is no guarantee, however, that the mutations discovered through this process will be effective for improving current bright variants. Though the crystal structures of citrine [8] and Venus [12] have been determined at 2.2 Å resolution, it is still unclear how the Q69M and F46L mutations act to accelerate the rate of oxidation of the Cα-Cβ bond of Tyr66.

The improved brightness should also be discussed in terms of the extent of quenching of the fluorophore by environmental factors. Among *Aequorea* GFP variants, YFPs are relatively acid sensitive, and uniquely quenched by halide ions, including chloride (Cl^-). Proton (H^+) and Cl^- synergistically affect the charge state of the chromophore of YFP, thereby suppressing its fluorescence. Citrine and Venus show increased resistance to both H^+ and Cl^- [7, 8].

1.3 Improvements in Maturation of DsRed

Compared with *Aequorea* GFP, DsRed has three unique features. First, the maturation process of DsRed includes a chromophore modification step which is responsible for changing its emission color from green to red (Fig. 2). Second, DsRed matures more efficiently at 37 °C than at room temperature, in contrast to *Aequorea* GFP, which folds preferentially at lower temperatures [13]. This may be directly traced to the differing water temperatures of the source organisms' habitats; *Aequorea victoria* is found in the Pacific Northwest, while *Discosoma* is native to the Indo-Pacific Ocean. Third, like many other Anthozoan GFP-like proteins, DsRed forms a tight tetrameric complex [14]. This tetrameric complex formation has been shown to be required for the development of red fluorescence [15], but it has not yet been determined which maturation step requires this process. Such tight oligomerization has probably evolved to maximize thermotolerance under conditions of intense tropical

sunlight. Very recently, Campbell et al. reported the successful engineering of monomeric RFP (mRFP1) from DsRed [16]. Their approach employed site-directed mutagenesis to break the tetrameric structure followed by random mutagenesis to rescue the original red fluorescence. The success of this work offers great hope regarding the conversion of other oligomeric fluorescent proteins into monomers. Another solution to the problem of RFP oligomerization reported by Campbell et al. [16] was to concatenate two dimer forming subunits of a mutant of DsRed with a spacer, which facilitates their dimer interactions through intramolecular contacts.

Incomplete chromophore modification of DsRed gives rise to residual green fluorescence, prohibiting the combined use of this molecule with green-emitting FPs in dual-color labeling experiments. The commercially available DsRed2 (Clontech) demonstrates only modest improvements in chromophore modification, but more recently, the release of engineered variants of DsRed known as T1 [17] and E57 [18] have enabled researchers to overcome the problems associated with slow and incomplete modification. Comparative analysis of these mutants in the context of the crystal structure of DsRed suggests available space around the chromophore is crucial for fast and complete modification [18]. Conversely, a long-lived green state can be advantageous for analysis of the history of gene expression in a cell. A new mutant of DsRed, E5, is particularly useful as it changes color from green to red in a predictable time course [19]. This feature makes it possible to use the ratio of green to red emission as an estimate of the time elapsed after initiation of reporter gene expression. Therefore, E5 functions as a fluorescent timer that yields temporal and spatial information regarding target promoter activity.

2
Engineering for Photoactivation and Photoconversion

2.1
Photoactivatable GFP

In the past two years, three new fluorescent protein variants have been generated that allow the selective activation or color conversion of fluorescent signal after specific illumination. The first variant, PA-GFP (photoactivatable GFP) [20] is based on wild-type *Aequorea* GFP, which has a bimodal absorption or excitation spectrum with two peak maxima, at 395 and 475 nm, corresponding to the protonated and the deprotonated states of the chromophore, respectively. When excited at 475 nm, wild-type *Aequorea* GFP emits maximal fluorescence at 503 nm, while excitation at 395 nm yields a maximum at 508 nm [2]. The latter large Stokes shift results from excited state deprotonation of the chromophore, as phenols become greatly more acidic in their excited states. Thus, excitation of the protonated chromophores gives emission at greater than 500 nm, similar to the direct excitation of the deprotonated chromophore.

Although the proton transfer is eventually reversible, the protonated chromophore is irreversibly isomerized to the deprotonated form upon intense illumination at 395 nm. The light transforms the species with the protonated chromophore, which absorbs at 395 nm, into the deprotonated species, which absorbs at 475 nm [2]. A recent cryospectroscopy, mass spectroscopy, and crystallography study on the photoconverted product showed that this photoconversion is a one-photon process paralleled by decarboxylation of a glutamic acid residue at position 222 [21]. Patterson and Lippincott-Schwartz found that substitution of histidine for threonine at position 203 was effective in decreasing the initial absorbance at 475 nm [20]. The resulting mutant, PA-GFP, exhibits up to a 100-fold increase in green fluorescence excitation at 488 nm when illuminated with 413-nm light.

2.2 Kaede

Around the same time, a green-emitting FP cloned from the stony coral *Trachyphyllia geoffroyi* was serendipitously discovered by Ando et al. to be useful as an optical cell marker [22]. The discovery occurred when the researchers accidentally left a test tube of the protein on a lab bench near a window, and found after a while that it had turned red. The naturally-engineered FP was named Kaede, the Japanese word for "maple leaf". Kaede contains a tripeptide, His^{62}-Tyr^{63}-Gly^{64}, which acts as a green chromophore that can be photoconverted to red. In its green state, it has two absorption peaks at 380 and 508 nm, corresponding to the neutral and ionized form, respectively. The neutral form of this molecule is highly sensitive to irradiation with UV or violet light (350–400 nm), which produce excitation and photoconversion. Interestingly, it has been observed that the photoconverted Kaede dissociates into 19- and 10-kDa fragments on SDS/PAGE. The structural basis for the green-to-red photoconversion has recently been presented [23]. As in *Aequorea* GFP, a chromophore, 4-(*p*-hydroxybenzylidene)-5-imidazolinone, derived from the

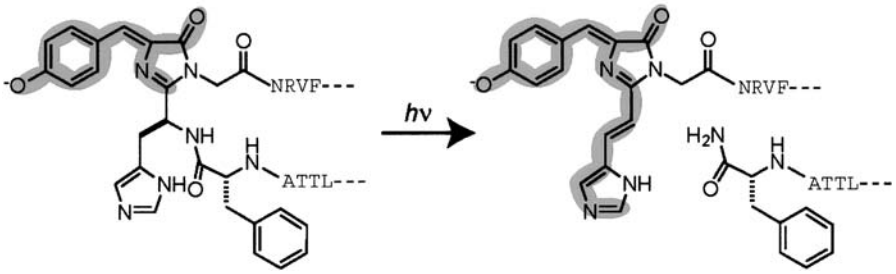


Fig. 4 Scheme for the formation and photo-induced extension of the chromophore of Kaede. Structures derived from Phe^{61}, His^{62}, Tyr^{63}, and Gly^{64} are drawn, and the neighboring amino acids (single-letter code) are added. π-conjugation for visible-light absorption is indicated in *gray*

tripeptide mediates green fluorescence in Kaede. UV irradiation causes an unconventional cleavage within Kaede protein between the amide nitrogen and the α carbon (Cα) at His^{62} via a formal β-elimination reaction, which requires the whole, intact protein for its catalysis. The subsequent formation of a double bond between His^{62}-Cα and -Cβ extends the π-conjugation to the imidazole ring of His^{62}, creating a new red-emitting chromophore, 2-[(1*E*)-2-(5-imidazolyl)ethenyl]-4-(*p*-hydroxybenzylidene)-5-imidazolinone (Fig. 4).

2.3 Kindling Fluorescent Protein

Lastly, Chudakov et al. showed that asCP, a unique GFP-like non-fluorescent chromoprotein from the sea anemome *Anemonia sulcata*, becomes fluorescent ("kindles") upon green light irradiation, with maximal emission at 595 nm [24]. Interestingly, asCP also has a histidine residue at position 203 (numbering in accordance with GFP alignment). More recently, they have generated a practically-useful mutant of asCP (designated KFP1) by replacing the alanine at position 148 with glycine [25]. KFP1 displays a roughly 30-fold increase in red fluorescence following excitation at 532 nm. This process is reversible, and can be controlled by both the light intensity level and the total light dose.

2.4 Histidine for Photochemical Reactions

It should be noted that all three of these variants carry histidine residues within or near the chromophore, suggesting the active involvement of histidine in photochemical reactions. Histidine, however, appears to play different roles depending on whether it is situated outside the chromophore at position 203 (PA-GFP and KFP1) or within the chromophore (Kaede). Studies by Ando et al. [22] and Chudakov et al. [24] have addressed the molecular mechanisms behind color variation occurring within individual coral animals. When exposed to sunlight, the tentacles and disk of the coral animals turn a shade of red in proportion to the degree of photoconversion. Then they revert to their previous colors as newly-synthesized proteins are added. Mechanisms such as this may be responsible for the great variety of color observed in coral reefs.

3 Engineering for Disruption of Oligomerization

Whether a fluorescent protein tends to form oligomers is an important consideration, as such interactions can interfere with the function of the host protein to which it is fused. Unfortunately, all of the *Anthozoan* GFP-like proteins

characterized so far form obligate oligomers [26]. While oligomerization does not prevent their use for reporting gene expression or marking cells, it does preclude their use in fusion protein applications. Similarly, fusion of DsRed, which normally forms a tetramer, to host proteins often disrupts their normal behavior, although there are some exceptions. As the monomeric RFP (mRFP1) matures ten times faster than its parental protein, it exhibits similar brightness to DsRed in living cells despite its lower molar extinction coefficient, fluorescence quantum yield and photostability. Because it is monomeric, mRFP1 has enabled red-fluorescence labelings that were not possible before with DsRed. Also, the excitation and emission maxima of mRFP1 are 584 and 607 nm, respectively, which gives good spectral separation from other FP signals. This work provides hope that other oligomeric FPs might also be converted into monomers. Indeed, the far-red variant HcRed1, made from a parent chromoprotein that seems to form obligate tetramers, has also been engineered to form dimers [27]. Similarly, a green-emitting FP recently cloned from *Galaxeidae* coral, Azami Green (AG) has been engineered to monomers (mAG) [28]. The anemone fluorescent protein epFP611 can function as a monomer, but only at low concentrations in the presence of detergent [29].

A further problem is the potential aggregation of fluorescent proteins, which impedes any cellular application and leads to cellular toxicity. Although the molecular mechanisms of FP aggregation remain unclear, there are two possible explanations. First, aggregation may be due to electrostatic or hydrophobic interactions between FPs. The possible contribution made by electrostatic interactions has been supported by recent work in which non-aggregating mutants were successfully generated by removing basic residues located near the amino termini of several fluorescent proteins [30], including DsRed. So, it may also be possible to make non-aggregating mutants by removing hydrophobic side chains on the surface of oligomeric complexes. It should be noted that *Renilla* GFP becomes soluble as a result of its dimerization; a hydrophobic patch becomes hidden at the dimerization interface and allows the surface of the dimer to become hydrophilic. The second possibility is that aggregation may follow FP oligomerization. Thus, the problem might be made worse still if host proteins are also oligomeric, as fusion to FPs might result in crosslinking into massive aggregates. Indeed, DsRed tends to produce more serious aggregation when fused to a host protein although, in an exception to this trend, fusion of DsRed to protein kinase C-γ retains the dynamic redistribution of the enzyme after stimulation. Overall, this aggregation problem would most easily be solved by using monomeric FPs.

Aequorea GFP has a propensity to weakly dimerize (Kd=0.1–0.3 mmol/l) [2]. Non-dimerizing CFP and YFP (mCFP and mYFP) have been constructed and utilized successfully in FRET experiments to determine how lipid-modified proteins assemble in the microdomains of the plasma membrane [31].

4
Engineering for Visualization of Cellular Functions

4.1
Genetically-Encodable Probes

In the past several years, various probes for cellular functions have been generated using FPs. Those probes employ simple GFP variants, circularly permuted GFP variants [10], or pairs of GFP variants that permit fluorescence resonance energy transfer (FRET) [32, 33]. Because the probes can be introduced by gene transfer techniques, they have significant advantages over conventional organic dyes. For instance, whereas conventional optical imaging of brain tissue stained with voltage-sensitive dyes is a noninvasive technique for recording the activities of a number of neurons simultaneously, it collects signals from all cell types including glial cells, which represent a large fraction of the total membrane surface in brain. By contrast, the selective introduction of genetically encoded probes into certain neurons has enabled the elimination of glial signals. Moreover, it is possible to place the probes within specific subcellular compartments where the desired signals predominate. The following section describes development of the probes that employ simple GFP variants or circularly permuted GFP variants.

4.2
Voltage Sensors

The prototype of the voltage-sensitive fluorescent protein is 'FlaSh' [34]. Insertion of wild-type GFP after the sixth transmembrane domain of a nonconducting mutant of the Shaker potassium channel has led to the generation of the sensor. In the construct, the voltage-driven rearrangement of the channel is converted into the change in fluorescence intensity of the GFP. It shows relatively complex and slow kinetics (τ_{off}>85 ms), suggesting that the structural rearrangement is related to the C-type inactivation of the channel. Recently, FlaSh has been significantly improved and diversified [35]. The substitution of various GFP variants has expanded its readout from single wavelength intensity modulating to dual wavelength ratiometric, and has produced probes with fast kinetics. In addition, mutations in the channel domain have tuned the probes. For example, the deletion of the N-terminal ball domain has improved the response rate and the Shaker S4 mutation L366A has negatively shifted the narrow dynamic range by approximately 30 mV, thereby enabling efficient detection of voltage changes around –60~70 mV occurring in retinal neurons.

Another genetically encoded voltage probe has been generated using a similar approach but a different channel. The probe termed 'SPARC' contains a mammalianized wild-type GFP inserted into an intracellular loop of a reversibly nonconducting form of the rat μI skeletal muscle voltage-gated sodium channel [36]. It can detect depolarizing pulses as short as 2 ms without inactivation

during extended depolarizations, which indicates that the fluorescence change is due to gating charge movement of a single domain of the channel.

4.3 Calcium Sensors

Our understanding of the structure-photochemistry relationships of GFP has enabled the development of genetic calcium probes based on a single GFP variant. Camgaroo-1 was constructed by inserting calmodulin (CaM) between positions 145 and 146 of YFP [10]. The Ca^{2+}-dependent conformational change of CaM induces the ionization of the chromophore, resulting in fluorescence increase of up to sevenfold. The chromophore development of this probe has been improved (camgaroo-2) by an amino acid substitution, Q69M [8].

Two groups fused circularly permuted green fluorescent proteins (cpGFPs) in which the amino and carboxyl portions had been interchanged around position 145 and reconnected by short spacers between the original termini, to CaM and its target peptide, M13. Chimeric proteins G-CaMP [37] and pericam [11] are fluorescent and their spectral properties change reversibly with Ca^{2+} concentration, probably due to the interaction between CaM and M13, leading to an alteration of the environment surrounding the chromophore.

G-CaMP is a single wavelength intensity modulating probe for Ca^{2+}. Fast mobilization of Ca^{2+} upon depolarization was observed in myotubes expressing the G-CaMP probe [37]. Recently, G-CaMP was applied to a functional analysis of odor-evoked patterns of activity in the neural assemblies in *Drosophila* antennal lobe [38]. The probe was expressed in olfactory receptor neurons or projection neurons, and patterns of glomerular activity were imaged in the antennal lobe presynaptically or postsynaptically, respectively. Odors elicited specific patterns of glomerular activity that were conserved in multiple flies, and the specific responsivity of a given glomerulus was found to be a consequence of the specificity of a single odorant receptor expressed by the incoming sensory neurons. Those observations are consistent with the "one neuron-one receptor" principle [39] that holds true for both vertebrate and insect olfactory systems.

Three types of pericam have been obtained by mutating several amino acids adjacent to the chromophore [11]. Of these, "flash pericam" becomes bright with Ca^{2+} like G-CaMP, whereas "inverse pericam" dims. On the other hand, "ratiometric pericam" has an excitation wavelength changing in a Ca^{2+}-dependent manner, thereby enabling dual excitation ratiometric Ca^{2+} imaging (Fig. 5A). Ratiometric pericam permits quantitative Ca^{2+} measurement by minimizing the effects of several artifacts that are unrelated to changes in free Ca^{2+} concentration ($[Ca^{2+}]$) (Fig. 5B). It has been successfully used to monitor changes in $[Ca^{2+}]$ in cardiomyocyte mitochondria. That study has demonstrated that mitochondrial $[Ca^{2+}]$ oscillates synchronously with cytosolic $[Ca^{2+}]$ during beating [40]. A laser-scanning confocal microscopy (LSCM) system has been modified for obtaining confocal images of Ca^{2+} using ratiometric pericam [41]. Rapid

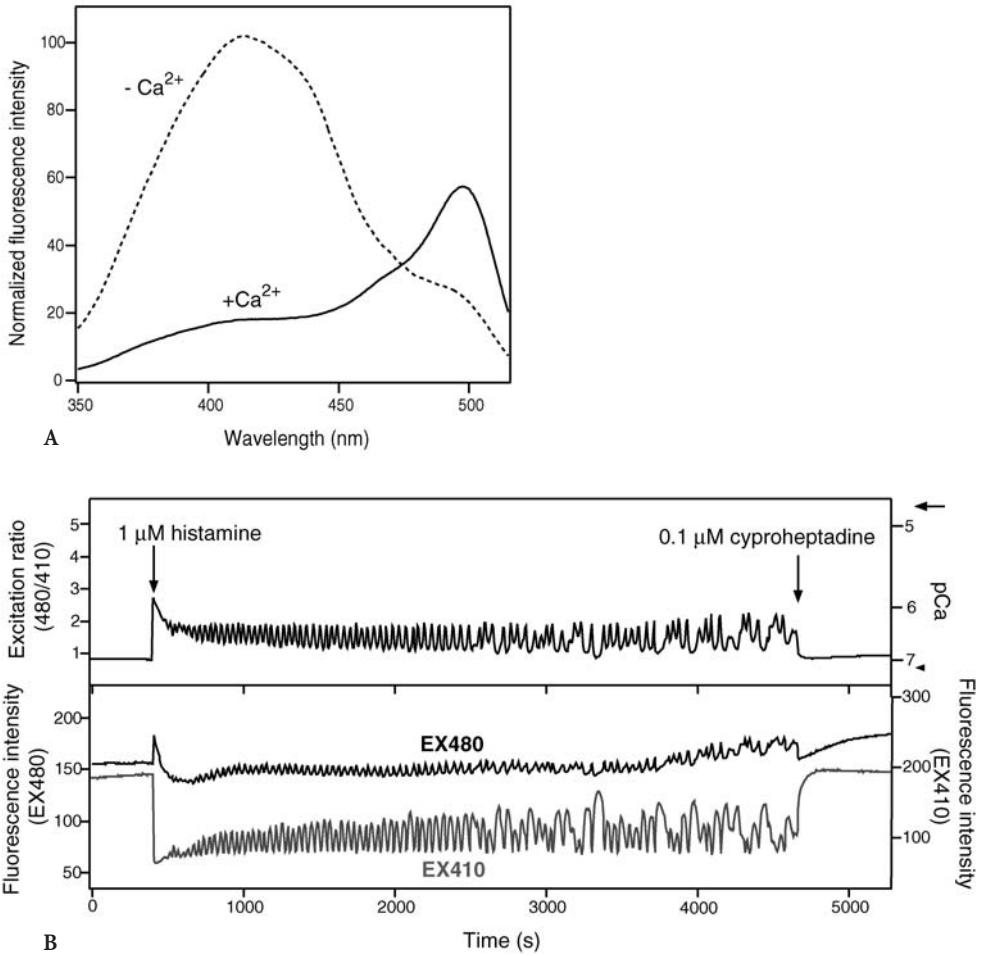


Fig. 5 **A** Excitation and emission spectra of ratiometric pericam in the presence and absence of calcium. **B** Typical $[Ca^{2+}]_i$ transients and oscillations induced by receptor-stimulations in HeLa cells expressing ratiometric-pericam. The sampling interval was 3~5 s. The right-hand ordinate calibrates $[Ca^{2+}]_i$ in μM with R_{max} and R_{min} indicated by an *arrow* and *arrowhead*, respectively. Reprinted with permission from Nagai et al., Proc. Natl. Acad. Sci. USA 98, 3197 (2001). Copyright © 2001 National Academy Sciences, USA

exchange between two laser beams has been achieved using acousto-optic tunable filters. Samples are scanned on each line sequentially by a violet laser diode (408 nm) and a diode-pumped solid state laser (488 nm). In this way, the ratios of the excitation peaks can be obtained at frequencies of up to 200 Hz, which enables visualization of Ca^{2+} dynamics within a motile mitochondrion.

Acknowledgements This work was supported by grants from CREST of JST (Japan Science and Technology), the Japanese Ministry of Education, Science and Technology, and HFSP (Human Frontier Science Program).

References

1. Prasher DC, Eckenrode VK, Ward WW, Prendergast FG, Cormier MJ (1992) Gene 111:229
2. Tsien RY (1998) Annu Rev Biochem 67:509
3. Matz MV, Fradkov AF, Labas YA, Savitsky AP, Zaraisky AG, Markelov ML, Lukyanov SA (1999) Nat Biotechnol 18:969
4. Labas YA, Gurskaya NG, Yanushevich YG, Fradkov AF, Lukyanov KA, Lukyanov SA, Matz MV (2002) Proc Natl Acad Sci USA 99:4256
5. Remington SJ (2002) Nat Biotechnol 20:28
6. Gross LA, Baird GS, Hoffman RC, Baldridge KK, Tsien RY (2000) Proc Natl Acad Sci USA 97:11990
7. Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A (2002) Nat Biotechnol 20:87
8. Griesbeck O, Baird GS, Campbell RE, Zacharias DA, Tsien RY (2001) J Biol Chem 276:29188
9. Reid BG, Flynn GC (1997) Biochemistry 36:6786
10. Baird GS, Zacharias DA, Tsien RY (1999) Proc Natl Acad Sci USA 96:11241
11. Nagai T, Sawano A, Park ES, Miyawaki A (2001) Proc Natl Acad Sci USA 98:3197
12. Rekas A, Alattia JR, Nagai T, Miyawaki A, Ikura M (2002) J Biol Chem 277:50573
13. Mizuno H, Sawano A, Eli P, Hama H, Miyawaki A (2001) Biochemistry 40:2502
14. Baird GS, Zacharias DA, Tsien RY (2000) Proc Natl Acad Sci USA 97:11984
15. Sacchetti A, Subramaniam V, Jovin TM, Alberti S (2002) FEBS Lett 525:13
16. Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, Tsien RY (2002) Proc Natl Acad Sci USA 99:7877
17. Bevis BJ, Glick BS (2002) Nat Biotechnol 20:1159
18. Terskikh AV, Fradkov AF, Zaraisky AG, Kajava AV, Angres B (2002) J Biol Chem 277:7633
19. Terskikh A, Fradkov A, Ermakova G, Zaraisky A, Tan P, Kajava AV, Zhao X, Lukyanov S, Matz M, Kim S, Weissman I, Siebert P (2000) 290:1585
20. Patterson GH, Lippincott-Schwartz J (2002) Science 297:1873
21. van Thor JJ, Gensch T, Hellingwerf KJ, Johnson LN (2002) Nat Struct Biol 9:37
22. Ando R, Hama H, Yamamoto-Hino M, Mizuno H, Miyawaki A (2002) Proc Natl Acad Sci USA 99:12651
23. Mizuno H, Mal TK, Tong KI, Ando R, Furuta T, Ikura M, Miyawaki A (2003) Mol Cell 12:1051
24. Chudakov DM, Feofanov AV, Mudrik NN, Lukyanov S, Lukyanov KA (2003) J Biol Chem 278:7215
25. Chudakov DM, Belousov VV, Zaraisky AG, Novoselov VV, Staroverov DB, Zorov DB, Lukyanov S, Lukyanov KA (2003) Nat Biotechnol 21:191
26. Zhang J, Campbell RE, Ting AY, Tsien RY (2002) Nat Rev Mol Cell Biol 3:906
27. Gurskaya NG, Fradkov AF, Terskikh A, Matz MV, Labas YA, Martynov VI, Yanushevich YG, Lukyanov KA, Lukyanov SA (2001) FEBS Lett 507:16
28. Karasawa S, Araki T, Yamamoto-Hino M, Miyawaki A (2003) J Biol Chem 278:34167
29. Wiedenmann J, Elke C, Spindler KD, Funke W (2000) Proc Natl Acad Sci USA 97:14091
30. Yanushevich YG, Staroverov DB, Savitsky AP, Fradkov AF, Gurskaya NG, Bulina ME, Lukyanov KA, Lukyanov SA (2002) FEBS Lett 511:11

31. Zacharias DA, Violin JD, Newton AC, Tsien RY (2002) Science 296:913
32. Miyawaki A (2003) Curr Opin Neurobiol 13:591
33. Miyawaki A (2003) Dev Cell 4:295
34. Siegel MS, Isacoff EY (1997) Neuron 19:735
35. Guerrero G, Siegel MS, Roska B, Loots E, Isacoff EY (2002) Biophys J 83:3607
36. Ataka K, Pieribone VA (2002) Biophys J 82:509
37. Nakai J, Ohkura M, Imoto K (2001) Nat Biotechnol 19:137
38. Wang JW, Wong AM, Flores J, Vosshall LB, Axel R (2003) Cell 112:271
39. Wang F, Nemes A, Mendelsohn M, Axel R (1998) Cell 93:47
40. Robert V, Gurlini P, Tosello V, Nagai T, Miyawaki A, Di Lisa F, Pozzan T (2001) EMBO J 20:4998
41. Shimozono S, Fukano T, Nagai T, Kirino Y, Mizuno H, Miyawaki A (2002) Sci STKE 125:PL4

Adv Biochem Engin/Biotechnol (2005) 95: 17–56
DOI 10.1007/b102209

Higher Harmonic Generation Microscopy

Chi-Kuang Sun (✉)

Graduate Institute of Electro-Optical Engineering and Department of Electrical Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC
sun@cc.ee.ntu.edu.tw

Abstract Higher harmonic-generation, including second harmonic generation and third harmonic generation, leaves no energy deposition to the interacted matters due to its virtual-level transition characteristic, providing a truly non-invasive modality and is ideal for in vivo imaging of live specimens without any preparation. Second harmonic generation microscopy

provides images on stacked membranes and arranged proteins with organized nano-structures due to the bio-photonic crystalline effect. Third harmonic generation microscopy provides general cellular or subcellular interface imaging due to optical inhomogeneity. Due to their virtual-transition nature, no saturation or bleaching in the generated signal is expected. With no energy release, continuous viewing without compromising sample viability can thus be achieved. Combined with its nonlinearity, higher harmonic generation microscopy provides sub-micron three-dimensional sectioning capability and millimeter penetration in live samples without using fluorescence and exogenous markers, offering morphological, structural, functional, and cellular information of biomedical specimens without modifying their natural biological and optical environments.

Keywords Second harmonic generation · Third harmonic generation · Scanning microscopy · Higher harmonic generation · Multi harmonic generation

List of Abbreviations and Symbols

2D	Two dimensions
3D	Three dimensions
4D	Four dimensions
CCD	Charge coupled device
E-field	Electric field
EFISHG	Electric field induced second harmonic generation
EO	Electro-optic
fps	Frames per second
fs	Femtosecond
GaN	Gallium nitride
HHG	Higher harmonic generation
HHGM	Higher harmonic generation microscopy
hpf	Hours past fertilization
Hz	Hertz
InGaN	Indium gallium nitride
IR	Infrared
J	Joule
KHz	Kilohertz
MHz	Megahertz
GHz	Gigahertz
NA	Numerical aperture
NLC	Nematic liquid crystal
OCT	Optical coherent tomography
OPA	Optical parametric amplifier
OPO	Optical parametric oscillator
PMT	Photo multiplier tube
P^{NL}	Nonlinear polarization
ps	Picosecond
SHG	Second harmonic generation
THG	Third harmonic generation
TPEF	Two-photon excited fluorescence
UV	Ultraviolet
χ	Optical susceptibility
$\chi^{(2)}$	Second order nonlinear susceptibility
$\chi^{(3)}$	Third order nonlinear susceptibility

1 Introduction

Higher harmonic generation microscopy (HHGM), including second harmonic generation (SHG) microscopy and third harmonic generation (THG) microscopy, is an emerging microscopic technique for a wide range of biological, medical, and material investigations. Advances in the developments of HHGM have provided researchers with novel means by which 3D non-invasive visualization of biological and non-biological specimens can be achieved with high penetration and high spatial resolution. Higher harmonic generations (HHG), including SHG and THG, are known to leave no energy deposition to the interacted matters due to their virtual energy conservation characteristic, that is, the emitted HHG photon energy is the same as the total absorbed excitation photon energy. This virtual energy-conservation characteristic provides the optical non-invasive nature desirable for microscopy applications, especially for live cells and clinical imaging. Different from multi-photon excited fluorescence microscopy, no energy release is required during the HHG processes, thus no in-focus cell damage and no photobleaching are expected from the HHG process itself. Due to its nonlinear nature, similar to the multi-photon excited fluorescence processes, the generated SHG intensity depends on the square of the incident light intensity, while the generated THG intensity depends on the third power of the incident light intensity. This nonlinear dependency allows localized excitation and is ideal for intrinsic optical sectioning. Due to their coherent nature during the generation processes, HHG, especially SHG, is sensitive to the local biological structures and is particular useful for structural studies. In contrast, THG microscopy can be utilized as a general-purpose microscopy technique for morphological studies.

In the past four decades, SHG has been widely applied to the study of photonic materials and interfacial regions without a center of symmetry, and was later combined with a microscope for SHG scanning microscopy [1, 2] in the 1970s. Today, SHG scanning microscopes have been widely used in material studies for surface monolayer detection [3], ferroelectric domain structures [4], nonlinear crystal characterization [5], and electric field distribution imaging [6, 7]. SHG in biological specimen was first observed in 1971 by Fine and Hansen in collageneous tissues [8]. In 1986, Freund, Deutsch, and Sprecher demonstrated a laser scanning microscopy in collagen fibrils using SHG as the contrast to study rat-tail tendon [9]. In 1989, another interesting finding was made by Huang and his coworkers who used SHG to probe the nonlinear optical properties of purple membrane-poly(vinyl alcohol) films [10]. The SHG was attributed to the naturally oriented dipole layers. Similar to the fact that SHG can be used to visualize the electric field, SHG was also shown to have intrinsic sensitivity to the voltage (or low frequency electric field) across a biological membrane [11] and was combined with microscopy for membrane potential imaging [12, 13]. Recently, it is concluded that besides electric field, SHG contrast in biological specimens can be provided by nanocrystalline structures, including stacked membranes, aligned protein structures, and microtubule arrays [14, 15].

In 1995, Tsang [16] reported that THG is a fundamental physical process occurring at all interfaces free from the constraint of a phase-matching condition and wavelength restriction. Taking advantage of this characteristic, in 1997 Silberberg and his coworkers demonstrated THG microscopy for interface imaging in an optical fiber [17]. THG microscopy was then applied to image laser-induced breakdown in glass [18] and liquid crystal structures [19]. We further demonstrated function-THG microscopy for defect distribution imaging in GaN semiconductors utilizing the resonant enhancement effect [20]. For THG applications in biology, in 1996 Alfano and his coworkers found significant THG in chicken tissues [21]. After the first demonstration of THG microscopy, it was then quickly applied to image biological specimen due to its interface sensitivity [22–25]. Detailed cellular and subcellular organelles can be clearly resolved with THG even with dynamic imaging. With cubic dependence on the illumination intensity, THG provides even better optical sectioning resolution than SHG or two-photon excited fluorescence (TPEF) using the same illumination wavelength, but is more sensitive to attenuation of the illumination light. With combined THG and SHG imaging on live zebrafish embryos, the complex developmental processes throughout the 1.5-mm-thick zebrafish embryos can be observed in vivo without any treatment. No optical damage was observed even with high illumination after long-term observations and all the examined embryos developed normally at least to the larval stage, indicating prolonged viability under the viewing condition of HHG [26]. With a 1.5 mm penetration capability, <500 nm lateral resolution can be achieved by using THG microscopy in live specimens [27]. Besides its capability for general morphological and interfacial imaging, THG microscopy can also provide functional imaging. For example, THG microscopy allows one to temporally visualize the release of Ca^{2+} [28]. Functional THG microscopy is also demonstrated by labeling the samples with gold [29] or silver [30] nanoparticles taking advantage of the surface-plasmon-resonance effects in noble-metal nanoparticles.

2 Principles of Higher Harmonic Generation

SHG and THG processes are both nonlinear optical processes, related to the interaction of intense light with matters. The SHG process describes the generation of a light wave with twice the frequency (with half of the wavelength) of the original one while the THG process describes the generation of a light wave that triples the frequency (with one-third of the wavelength) of the original one. The higher harmonic light wave generation is coupled from the induced nonlinear polarization under intense laser excitation. The interaction of nonlinear polarization and the excitation light is related through a nonlinear susceptibility χ. For example, SHG can be generated through a second-order nonlinear polarization $P^{(2)}$ created according to [31]

$$P_i^{(2)}(2\omega) = \frac{1}{2}\,\varepsilon_0\,\chi_{ijk}^{(2)}(2\omega:\omega,\omega)\,E_j(\omega)\text{ and }E_k(\omega)\,. \tag{1}$$

The subindices *i*, *j*, *k* indicate base directions (x, y, z) in Cartesian coordinates. $\chi_{ijk}^{(2)}(2\omega{:}\omega,\omega)$ represents a second order nonlinear susceptibility tensor, a tensor of third rank, responsible for SHG process while $E_j(\omega)$ and $E_k(\omega)$ are the laser field amplitudes in *j* and *k* directions. $P_i^{(2)}(2\omega)$ is the induced SHG nonlinear polarization in *i* direction at twice the fundamental frequency, which couples to and generates light field radiation $E_i(2\omega)$ in twice the original angular frequency 2ω. Summation over *j* and *k* is implied in Eq. (1) by the Einstein summation convention. Please notice that $\chi_{ijk}^{(2)}(2\omega{:}\omega,\omega)$ vanishes for a medium with centro-symmetry. For a non-centro-symmetry medium, not all the tensor elements exist.

With low frequency electric fields breaking the centro-symmetry, SHG can also be generated due to low frequency electric field even in centro-symmetric media, providing a tool to image electric fields. This process is usually described as a third-order nonlinear process with a third-order nonlinear polarization tensor, a tensor of fourth rank, described by

$$P_i^{(3)}(2\omega) = \frac{3}{4}\,\varepsilon_0\,\chi_{ijkl}^{(3)}(2\omega:0,\omega,\omega)\,E_j(0)\,E_k(\omega)\,E_l(\omega)\,, \tag{2}$$

where $E_j(0)$ is the low frequency electric field. This process is usually referred to as electric-field-induced-second-harmonic-generation (EFISHG).

Similarly, the THG process is related to the creation of a third order nonlinear polarization $P^{(3)}$ (at 3ω angular frequency) according to

$$P_i^{(3)}(3\omega) = \frac{1}{4}\,\varepsilon_0\,\chi_{ijkl}^{(3)}(3\omega:\omega,\omega,\omega)\,E_j(\omega)\,E_k(\omega)\,E_l(\omega) \tag{3}$$

where $\chi_{ijkl}^{(3)}(3\omega{:}\omega,\omega,\omega)$ represents a third order nonlinear susceptibility tensor, a tensor of fourth rank, responsible for the THG process.

SHG and THG can be visualized by considering the interaction in terms of the exchange of photons between various frequencies of the fields. According to this picture, which is illustrated in Fig. 1, two or three photons of angular

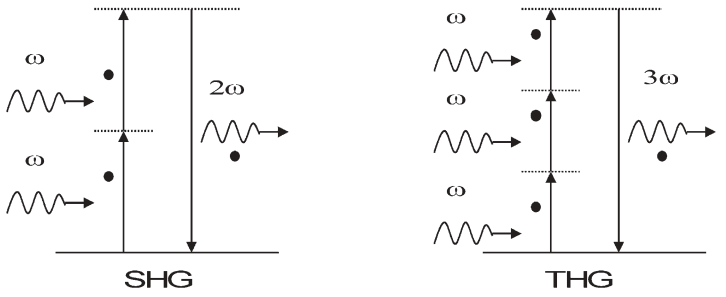


Fig. 1 Energy level diagrams describing SHG and THG

frequency ω are destroyed and a photon of angular frequency 2ω (for SHG) or 3ω (for THG) is created in a single quantum-mechanical process. The solid lines in the figure represent the atomic or molecular ground states, and the dashed lines represent what is known as virtual states. These virtual levels are not energy eigenlevels of the atoms or molecules, but rather represent the combined energy of one of the energy eigenstates of the atom or molecules and one or more photons of the radiation field. Due to its virtual level transition characteristics, HHG are known to leave no energy deposition to the interacted matters, since no real electron transition is involved and the emitted photon energy will be exactly the same as the total absorbed photon energy. This energy conservation characteristic provides the optical "non-invasiveness" desirable for microscopy applications, especially for live specimen imaging since there is no energy deposition into the imaged sample during the HHG processes.

According to Eqs. (1)–(3), the generated SHG intensity depends on the square of the incident light intensity, while the generated THG intensity will depend on the third power of the incident light intensity. Similar to multi-photon induced fluorescence process, this nonlinear dependence allows localized excitation and is ideal for intrinsic optical sectioning in scanning laser microscopy. Usually the third-order nonlinear susceptibility $\chi^{(3)}(3\omega{:}\omega,\omega,\omega)$ is much weaker than the second-order nonlinear susceptibility $\chi^{(2)}(2\omega{:}\omega,\omega)$ under a normal illumination condition; thus THG should be much harder to observe. However, not all materials have second-order nonlinear susceptibility. For centro-symmetric media, the lowest order nonlinear susceptibility will be $\chi^{(3)}$ instead of $\chi^{(2)}$. This is why people have rarely heard of SHG from a bulk glass. The random distribution of oxides in glass creates optical centro-symmetry in the optical wavelength scale, thus inhibiting the SHG process. Only non-centro-symmetry media are allowed to generate SHG.

On the other hand, all materials have non-zero third order susceptibility, the magnitudes of which vary according to material properties, symmetry, and wavelength. When the medium around the focal point is homogeneous, the third-harmonic waves generated before and after the focal point interfere destructively, resulting in zero net THG [32]. However, when there is inhomogeneity near the focal point, such as an interface between two media, the symmetry along the optical axis is broken and a measurable amount of the third harmonic wave is generated. Due to its nonlinear nature, the third-harmonic light is generated only in close proximity to the focal point. Therefore, high depth resolution can be obtained, allowing THG microscopy to perform optical sectioning and to construct 3D images of transparent samples. Since all materials have non-vanishing third-order susceptibilities, THG microscopy can be utilized as a general-purpose microscopy technique, with no need of fluorescence labeling.

According to photon momentum conservation, the generated higher harmonic photons will be emitted in the same direction as the annihilated photons, unless non-collinear phase matching processes due to a periodic local environment are involved. This forward direction emission property restricts the light

collection geometry in most microscopy applications. While some SHG microscopic studies adopt reflection geometry for their easy combination with current TPFM setups or due to difficulties in transmission collection, most HHGMs utilize transmission detection, which is different from the reflection detection in most laser scanning fluorescence microscope nowadays.

3 Technology Developments

3.1 Laser Sources

When SHG in biological specimen was first observed in 1971 from collageneous tissues [8], a Q-switched ruby laser at 694 nm was used while 347 nm SHG generated in the UV region faced strong absorption in bio-specimens. In 1986, Freund, Deutsch, and Sprecher demonstrated the first biological study of SHG microscopy in connective tissues [9] based on a 1064 nm Q-switched Nd:YAG laser with a nanosecond-scale pulsewidth. With the recent rapid advance in femtosecond laser technology, femtosecond lasers are now applied to SHG and THG microscopy due to their high peak intensities for efficient HHG compared with Q-switched lasers. For example, in 1997 Alfano and coworkers demonstrated SHG tomography of animal tissues by use of 100-fs laser pulses at 625 nm from a modelocked dye laser [33]. Nowadays, most researchers adopt fs Ti:sapphire lasers with a broad tuning range of 690–1070 nm as the excitation sources for efficient SHG and high penetration in biological tissues [12, 21, 34–39] while TPEF can be simultaneous excited in one simple microscope setup [15, 37, 38]. Besides Ti:sapphire lasers, we also utilize a fs Cr:forsterite laser [14, 25–27] for its high penetration, high cell viability, and capability for simultaneous THG microscopy.

For THG applications in biology, in 1996 Alfano and his coworkers found THG signals in chicken tissues using a 10 Hz 1064 nm Nd:YAG laser with a 30 ps pulsewidth [21]. It is also intuitive to use a fs Ti:sapphire laser as the excitation source due to it capability for simultaneous acquisition of TPEF and better spatial resolution due to shorter excitation wavelength [40]. However the strong absorption of the generated THG in the UV region limits its application to thin biological specimens. To avoid strong UV absorption of the generated THG in thick biological specimens, most THG microscopy utilizes fs excitation sources with a wavelength longer than 1200 nm. The lasers that are often used for THG microscopy include an optical parametric oscillator (OPO) synchronously pumped by a fs Ti:sapphire laser at a wavelength of 1500 nm [24, 28] with a repetition rate of ~80 MHz, an optical parametric amplifier (OPA) at 1200 nm pumped by a Ti:sapphire amplifier with a 250 kHz repetition rate [22, 23], a fiber laser at 1560 nm with a repetition rate of 50 MHz [41], and a Cr:forsterite laser [14, 25–27, 30, 42] at 1230 nm with a repetition rate of 110 MHz.

To realize in vivo non-invasive HHGM with high penetration and high cell viability for long-term observations, it is important to choose the desired laser source to avoid any unnecessary laser-tissue interactions while high spatial resolution, high sensitivity, and high penetration can all be achieved. Since the generation of optical higher harmonics has a weak dependency on the excitation wavelength, it is thus allowed to choose the desirable light source that can provide high penetration through turbid specimens with minimized unwanted light-tissue interactions including scattering, absorption, and photodamage. According to previous studies [14, 43, 44], light attenuation (including both absorption and scattering) in most live biological specimens reaches a minimum around 1200–1350 nm wavelength due to the combination of diminishing scattering cross-section with wavelength and avoiding the resonant molecular absorption of common tissue constituents such as water, melanin, and hemoglobin. Moving laser wavelength to longer than 1350 nm faces strong water absorption that will cause strong attenuation of excitation light with much reduced penetration and water heating. Previous studies comparing contrast and penetration depth between 800 nm and 1300 nm light sources for optical coherent tomography (OCT) in live tissues suggested superior performance at 1200–1300 nm [45, 46]. Moving the operating wavelength of HHGM to 1200–1350 nm spectral region can not only increase the penetration depth in live specimens but also reduce the multiphoton absorption cross-section section [47] and thus reduce the potential photodamage and phototoxicity.

Previous reports [47–50] on the on-focus multi-photon-absorption-induced photo-damages induced by using fs Ti:sapphire lasers in TPEF microscopy indicate the importance to reduce possible tissue damage while performing HHGM also with the high intensity fs light source. These photodamage phenomena limit the maximum optical intensity applicable in a HHGM, which also limit the nonlinear signal intensity and thus the penetration depth. Previous studies suggested that with an 80-MHz Ti:sapphire laser and a NA~1 objective, less than 7 mW average power could be applied to live subject to prevent possible optical damages [48–50]. A previous study [51] also indicated that increasing the excitation laser wavelength to 1047 nm will allow long-term TPEF imaging of mammalian embryos without compromising viability, while the maximum average illumination power can be increased to 13 mW while a total exposure of 2 J could be applied to one embryo within 24 h imaging period.

We have previously investigated the wavelength dependent cell damage induced by high intensity fs near infrared lasers [47]. The study was performed with a fs Ti:sapphire laser at ~800 nm and a fs Cr:forsterite laser at 1230 nm. With a longer output wavelength from a Cr:forsterite laser, multi-photon absorption and auto-fluorescence were all found to be significantly suppressed and the destructive plasma formation was found to be greatly reduced. Sustained multi-photon spectra, with suppressed intensity and narrowed linewidth, can be observed in most plant specimens with a tightly focused Cr:forsterite laser beam under long term irradiation with more than 100 mW laser average

power. In contrast, multi-photon absorption induced destructive plasma formation was frequently observed with a tightly focused Ti:sapphire laser beam within seconds if the applied laser average power was higher than 10 mW. Our recent in vivo embryonic study using the 110-MHz fs Cr:forsterite laser at 1230 nm indicated that the zebrafish embryos can be continuously imaged over the course of hours to days without observable damage and the zebrafish embryos can all develop into normal healthy larva even with 100 mW average illumination power, corresponding to a total exposure of >1000 J per embryo [26].

Another important issue is the optical pulse repetition rate selection. To maintain cell vitality, the optical peak intensity and the pulse energy should be lowered down to the safety zone, which might limit the signal strength of the stimulated HHG, thus limit the scanning speed and the image frame rate. Increasing the pulse repetition rate can solve this problem. Different from absorption stimulated fluorescence, the virtual-transition processes of HHG do not create any real-transition-resulted upper-state electrons. The selection of the pulse repetition rate is thus not restricted by the relaxation time of those photoexcited electrons. By increasing the pulse repetition rate higher than 100-MHz or even GHz while maintaining the same low pulse energy, we can increase the average intensity of the detected nonlinear signals and thus increase the acquisition speed, penetration depth, and image frame rate [39]. For most of our current applications, an all-solid-state Cr:forsterite fs laser [6, 7, 14, 20, 25–27, 30, 42] centered at 1230-nm with a repetition rate of 110 MHz is implemented as the light source while a 2GHz-repetition-rate fs Ti:sapphire laser centered around 800 nm is also selected for real-time SHG microscopy [39]. The recent rapid advance in fs laser technology allows the repetition rate of a fs Cr:forsterite laser to be increased up to 420 MHz with a 30 fs pulsewidth and a 500 mW average power [52]. However this laser has not been applied to scanning microscopy yet.

3.2 PMT-Based Higher Harmonic Generation Microscope

Most laser scanning HHGM is adopted from a commercial confocal scanning system combined with a microscope while all optics are modified to allow the passage of the excitation IR light. Figure 2 shows an example schematic setup of a photo-multiplier-tube- (PMT-) based HHGM built in the author's laboratory. A home-built Cr:forsterite laser, which operates at 1230 nm with a repetition rate of 110 MHz and a pulse width of 140 fs (350 mW average output), was used as the light source to allow both SHG/THG to fall within the visible spectrum and provide high tissue penetration at the same time. We also combine a tunable fs Ti:sapphire laser emitting around 700–980 nm for efficient TPEF or for higher SHG spatial resolution. The laser output was first shaped and collimated by a telescope to avoid power loss at the XY galvanometer mirror scanner and to fit the back aperture of the focusing objective. The collimated beam was then

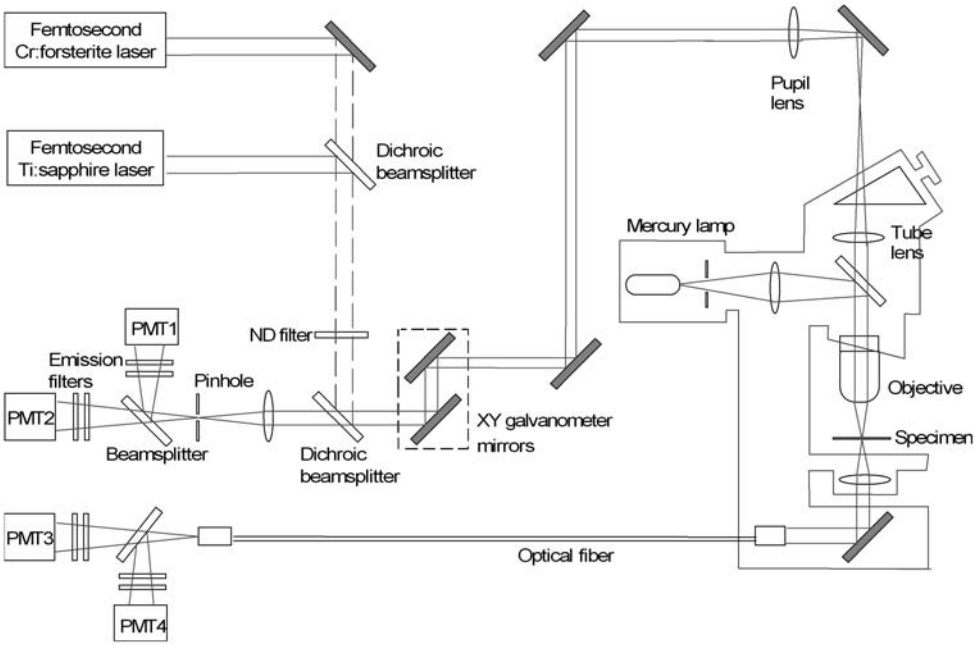


Fig. 2 Schematic diagram of a PMT-based HHGM

coupled into the scanning system connecting to a microscope with an aperture fitting tube lens. Real-time scanning was accomplished through the high-speed galvanometer mirrors inside the scanner. In order to visualize all cell activities inside the live zebrafish embryo, the working distance of the chosen objective has to be longer than the embryo thickness (>1 mm), including the chorion. The excitation laser pulse was focused into the desired location inside the specimen and scanning with a spot size close to its diffraction limit with a 2-mm-working-distance high NA IR objective. The forward-propagated HHG is collected using a high-NA condenser. For example, by using a numerical aperture (NA) 0.9 long-working-distance objective, the diffraction-limited spatial resolution of our system is ~1 μm in axial and 400 nm (500 nm) in radial directions for THG (SHG) microscopy using a 1230-nm excitation wavelength. For HHG intensity mapping, two PMTs (PMT3 and PMT4), which are synchronized with the galvanometer mirrors, are used to record the filtered SHG and THG intensities point by point to form a two-dimensional sectioned images. Stereographic 3D images of the live specimen can be acquired through accurate height control of the sample cell or objective along the light propagation axis. The microscope objective employed to focus the laser beam can also act as the collection lens to collect the backward propagation SHG, THG, and multiphoton excited fluorescence, which are separated from the laser beam with a dichroic beamsplitter. Different luminescence wavelengths and HHG are further separated using another dichroic beamsplitter and directed into two different

PMTs (PMT1 and PMT2) with appropriate filters. To observe moving live specimens, the recording speed could be chosen to be 0.25 to 2.5 s per frame with 512×512-scanned points. Some commercial scanning systems allow even higher scanning speed.

3.3 CCD-Based Higher Harmonic Generation Microscope

For thin and transparent specimens, the sectioned 2D HHG images, excited by the projected dithering laser beam on the objective focal plane, can be directly visualized with a CCD camera. Figure 3 shows an example setup of a CCD-based scanning SHG microscope. We utilized this setup for direct visualization of electric field distribution in an IC circuit [7]. The light source was the fs Cr:forsterite laser centered at 1230 nm. The laser was focused onto the sample with a high NA objective. Incorporating an 8-kHz XY scanner (VSH-8, GSI Lumonics) enables the focused laser spot to scan across the focal plane at more than 40 frames per second. The induced SHG image at a wavelength of 615 nm was then projected onto a high frame-rate CCD camera (Retiga-1350, Qimaging) for direct visualization while a color filter was used to remove the fundamental light at 1230 nm.

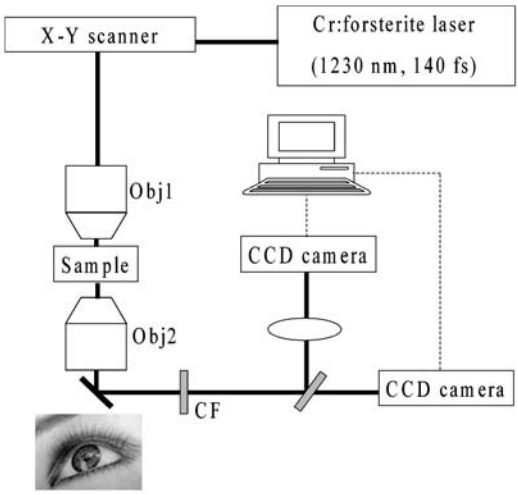


Fig. 3 Schematic diagram of a CCD-based HHGM. CF: color filter

3.4 Spectrum-Based Higher Harmonic Generation Microscope

We have previously demonstrated a spectrum-based multi-modal nonlinear spectral microscope [25]. A 3D mechanical stage was combined with a spec-

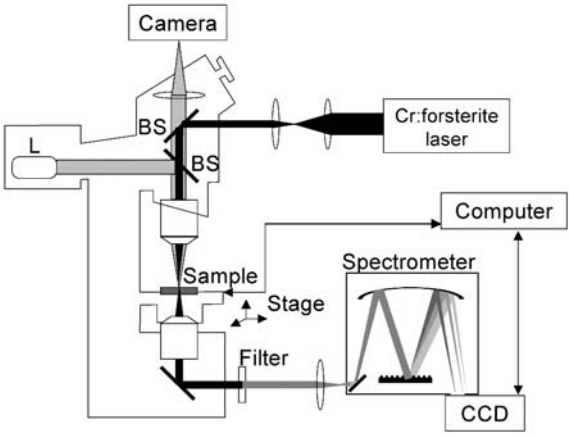


Fig. 4 Schematic diagram of a multi-modal nonlinear spectral microscope

trometer to achieve an x-y-z-λ 4D image. Figure 4 shows an example setup for this spectrum-based multi-modality nonlinear spectral microscope. The femtosecond Cr:forsterite laser output was first attenuated by neutral-density filters and shaped by a telescope, then focused into samples with a high NA objective. The photo-excited nonlinear spectrum was collected using an opposing objective and then directed into a spectrometer coupled to a TE-cooled CCD detector. We filtered out the fundamental IR beam with an optical glass filter. A mechanical stage was used to achieve a 3D scan in order to obtain x-y-z-λ images with a detailed nonlinear spectrum corresponding to each pixel recorded, including SHG and THG. The disadvantage of this system is its slow scanning rate. The pixel acquisition time usually ranges from 0.02 s to 0.5 s. On the other hand, this system provides full spectral information so that different modal signals, including SHG, THG, TPEF, or even three and four photon excited fluorescences [53] can all be clearly resolved.

4
Contrasts in Higher Harmonic Generation Microscopy

4.1
Contrast in SHG Microscopy

Since the first observation of SHG from collageneous tissues in 1971 by Fine and Hansen [8], collagen fibrils provide excellent contrast in biological SHG microscopy, including the first SHG microscopy study of connective tissues in 1986 by Freund, Deutsch, and Sprecher [9]. Freund and his coworkers attributed the SHG generation to the polarity (that is one type of non-centro-symmetry) of rat-tail tendon, which was then correlated to polar collagen fibrils. In

1989, J. Y. Huang and his coworkers used SHG to probe the nonlinear optical properties of purple membrane-poly(vinyl alcohol) films [10]. The SHG is attributed to the naturally oriented dipole layers. Recently Alfano and coworkers have also reported SHG from animal tissues [21]. Even though they attributed the SHG only from the surface term that is due to the broken symmetry at the boundary, they also found some SHG intensity dependence on the tissue constitutes with asymmetric structures such as collagen. Combining this effect, they demonstrated SHG tomography for mapping the structure of animal tissues by use of 100-fs laser pulses at 625 nm [33].

We have also studied the SHG contrast mechanism in biological tissues. Our recent study indicated that strong SHG signals in biological specimens mainly originate from highly organized nano-structures. Numerous biological structures including stacked membranes and arranged proteins are highly organized in the scale much smaller than one optical wavelength and are found to exhibit strong optical activities through SHG interactions, behaving similar to man-made nonlinear optical crystals or nano-photonic crystals. Previous observations of SHG on collagen fibrils [8, 9, 21], purple membranes [10], and muscle fibrils [21, 33] are just a few examples. A similar conclusion is also made by Campagnola and his coworkers [15]. Recently the angular dependence of the hyper-Rayleigh scattered light intensity from a suspension of bacteriorhodopsin has been analyzed by Clays, Van Elshocht, and Persoons [54]. The observation calls for combination of the second-order nonlinear hyperpolarizability of the retinal chromophore with the linear refractive index of the *apo*-protein matrix. They concluded that the purple membrane, containing orderly patched bacteriorhodopsin protein, is a natural nonlinear photonic bandgap material due to the structuring of the small nonlinear chromophore in the large linear matrix [54, 55], thus generating forward emission SHG.

Recent studies on man-made nano-periodic structures, e.g. super-lattices or semiconductor Bragg mirrors, indicate strong enhancement in SHG [55, 57] due to the nonlinear photonic crystal effect [58]. It is interesting to notice that the periodicity of the observed nonlinear photonic crystal does not necessarily have to be on the order of the optical wavelength or a fraction of the optical wavelength inside the materials [59], but could be much smaller than the affected wavelength with a nanometer scale [54–57]. This strong SHG enhancement could be simply understood as the break down of optical isotropy within the materials. With the same principle, highly organized nano-structures based on nonlinear biomolecules would also break optical isotropy and behave as SHG-active nano-optical crystals. Examples of highly organized biological nano-structures include grana in the chloroplast, mitotic spindles, cellulose microfibrils, and collagen bundles. These biological structures act as SHG nano-optical crystals and provide the sources of contrast in SHG microscopy.

4.1.1 Plant Cell Wall

Plant cell walls consist of cellulose microfibrils, which are orderly arranged to form macrofibrils with a dimension of about 10 nm. Within the microfibrils, micelles represent another degree of highly ordered crystalline structure. These crystalline structures produce optical anisotropy and provide active SHG. Figure 5A shows the nonlinear emission spectrum that was measured in the transmission direction from the cell wall of a parenchyma in maize (*Zea mays*) stem using the 1230 nm fs Cr:forsterite laser as the excitation source [16]. Symmetric SHG and THG spectra centered at 615 and 410 nm can be observed, with intensities similar to/or stronger than that of the residual TPEF centered at 680 nm. In order to confirm the nonlinear nature of the measured signals, Fig. 6 shows the measured SHG and THG powers (which are proportional to the measured counts on a CCD with a fixed integration time) of 615 nm and 410 nm light vs input IR power from the cell wall of maize stem by using a

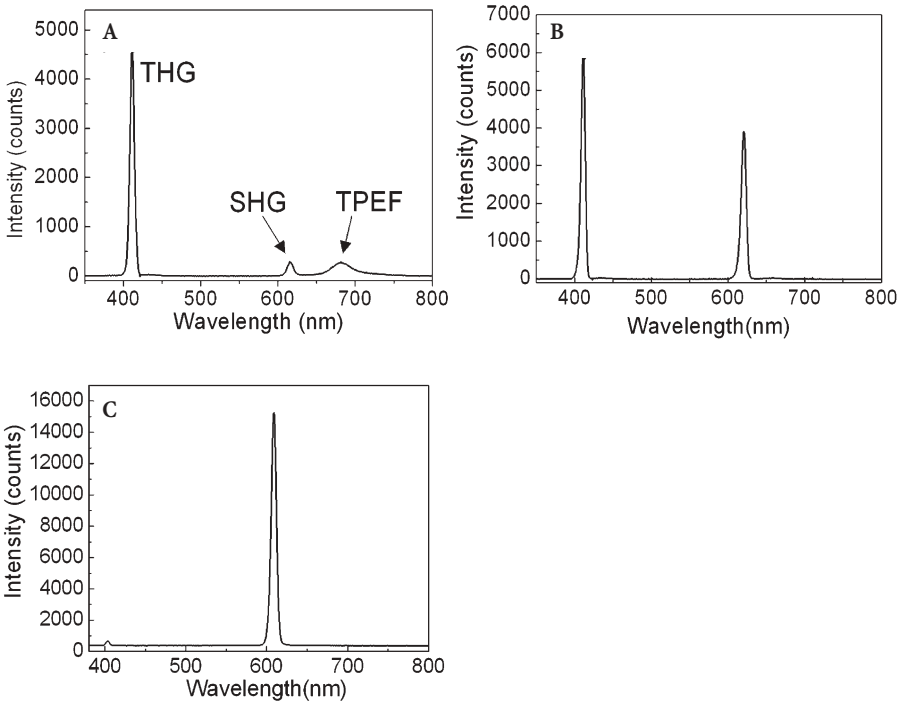


Fig. 5A–C Nonlinear emission spectra measured from: **A** the cell wall of a parenchyma in maize (*Zea mays*) stem; **B** the stone cell of a pear (*Pyrus serotina* R.) fruit; **C** potato (*Solanum tuberosum* L.) starch granule. These spectra were recorded under similar illumination intensity, with normalized 0.1 s integration time. Note the extraordinarily strong SHG from the starch granule

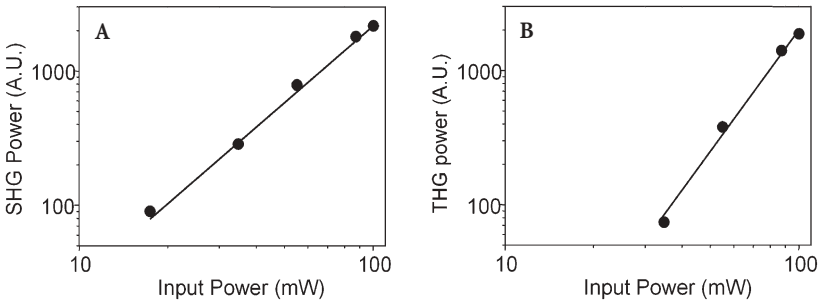


Fig. 6A, B **A** Average SHG power (615 nm) vs average input IR power (1230 nm) from the cell wall of maize stem. **B** average THG power (410 nm) vs average input IR power (1230 nm) from the cell wall of maize stem. The *solid circles* are experimental data and the *solid lines* are lines with slopes of 2 and 3 for A and B, respectively

0.75NA objective. Solid circles are the experimental data and solid lines correspond to fitting of the output-input relations. Slopes of 2 and 3 can be found in the log-log plots, corresponding to square and cubic dependency of the generated signals on the excitation intensity, confirming the SHG and THG nature of the measured signals.

In order to confirm the origins of the SHG generation, scanning images were used to identify the location of the nonlinear signal sources. Figure 7A shows a sectioned (x-y) SHG image taken from the ground tissue of maize stem at

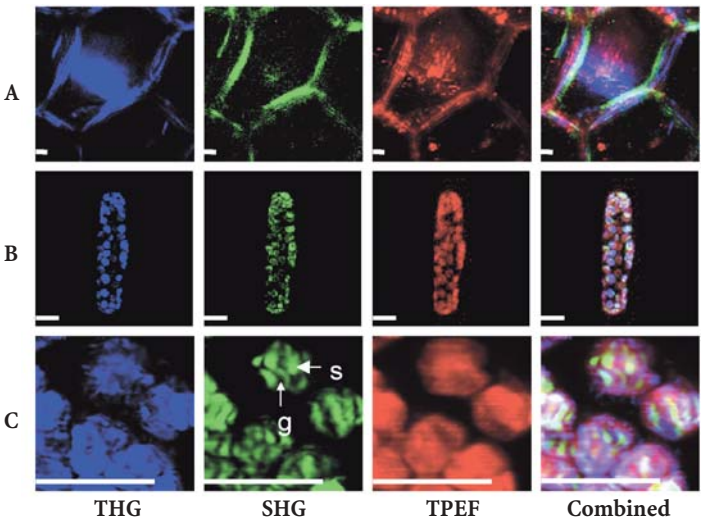


Fig. 7A–C Scanning THG (*shown in blue*), SHG (*shown in green*), TPEF (*shown in red*), and the combined multimodal images. Images were taken from: **A** ground tissue cells in a maize stem 420 µm from the sample surface; **B** a live mesophyll cell from *Commelina communis* L.; **C** enlarged view of B. *Scale bar*: 15 µm. g: grana. s: starch granule

a depth of 420 μm from the sample surface using a multi-modal nonlinear spectral microscope. TPEF of residue fluorescence and THG images are also shown for comparison. SHG shows mainly the longitudinal cell walls while TPEF indicates the distribution of the auto-fluorescent molecules. SHG reflects the orderly arranged cellulose microfibrils that break the 3D optical centrosymmetry in the cell wall. The combined multi-modal image is also shown with blue, green, and red colors representing THG, SHG, and TPEF respectively. The source of SHG was further confirmed by the strong signal obtained from the stone cell of pear (*Pyrus serotina* R.) fruit (Fig. 5B). The extensive secondary wall development of the sclerenchyma generates significant SHG signals.

4.1.2 Starch Granule

Starch granule, a strong birefringent structure, consists of crystalline amylopectin lamellae organized into spherical blocklets and large concentric growth rings [60]. These structural features are presumably responsible for the strong SHG we observed (Fig. 5C). For example, the SHG signal from potato (*Solanum tuberosum* L.) starch granule is so strong that is visible to the naked eyes [14]. Its alternating crystalline and semi-crystalline rings [61] with spatially modulated nonlinear properties could behave as 3D nonlinear photonic bandgap crystals [58]. The unexpected strong SHG activity might not only be the result of its super-helical amylopectin nano-structures, but also suggest possible SHG non-collinear phase matching provided with its reciprocal lattice basis vectors of the 3D photonic bandgap crystals. Spatial frequency of high order structures between the order of 100 nm up to the order of 10 μm, depending on illumination wavelength and composition materials, could all provide the non-collinear phase matching base-vector for SHG process [58], providing even stronger SHG activity due to coherence. This significant enhancement of SHG interactions in a 3D structure could also be treated as a simultaneous availability of a high density of states and improvement of effective coherent length (similar to the phase matching effect) due to these wavelength-scale photonic "bandgap" structures [57].

4.1.3 Mineral Deposition

Mineral deposition in plant cells forms crystalline structures, and thus also generate significant SHG signals. Figure 8A shows scanning HHG (including SHG and THG) images taken from the adaxial surface of a maize leaf. The distribution of mineral deposition in silica cells (profiled by THG) can be clearly seen through SHG.

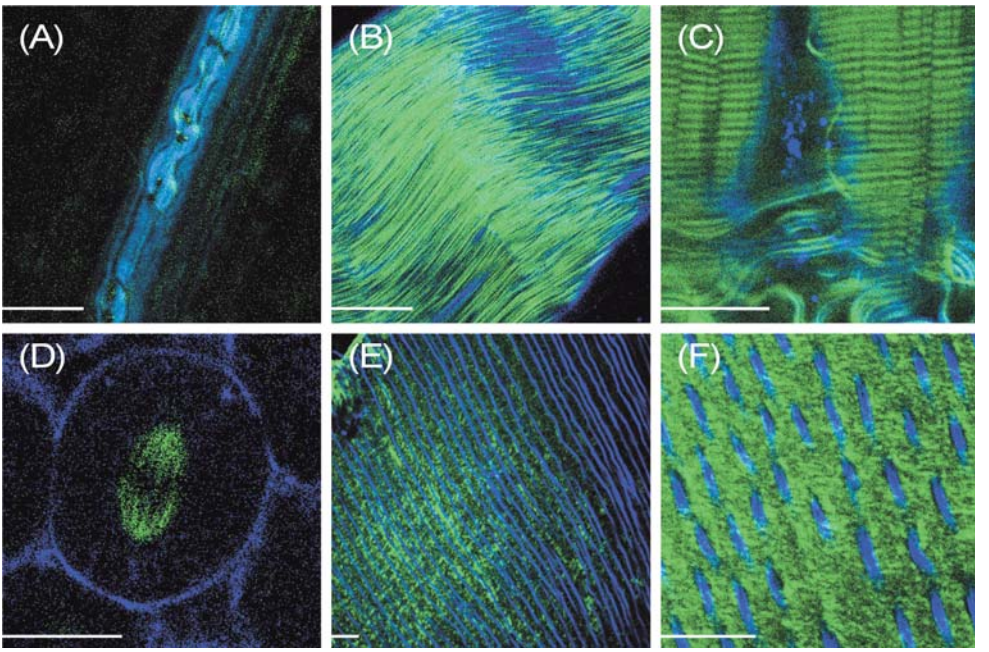


Fig. 8A–F Higher-harmonic generation image of: **A** silica cells on the adaxial surface of a maize leaf; **B** mouse tendon; **C** mouse skeletal muscle; **D** the mitosis process in the embryonic blastoderm at 1-k cell stage in a zebrafish embryo; **E,F** dentine of a human tooth. THG is *shown in blue* and SHG is *shown in green*. *Scale bar*: 20 µm

4.1.4 Grana

Laminated membrane structures with a nano-periodicity are also potential candidates for producing strong SHG activity. In chloroplasts, in addition to the TPEF signals generated from the highly auto-fluorescent photosynthetic pigments, SHG appears in different sub-organelle compartments. Figure 7B,C shows scanning (x-y) SHG images of chloroplasts inside a live mesophyll cell of a day flower (*Commelina communis* L) leaf. The distribution of chlorophyll inside the chloroplasts is also shown by scanning the corresponding TPEF signals around 680 nm. Matching with TEM images of similar specimens, we concluded [14] that the signals of SHG are the result of the orderly stacked thylakoid membranes in grana (crescent shaped) and the highly structured starch granules (oval or round shaped) inside the chloroplasts. The stacked thylakoid membranes of grana with a 20-nm period and the orderly deposited amylopectin in the starch granules provide the structural requirement for efficient SHG, resembling the behavior of photonic crystals.

4.1.5 Connective Tissues

Animal connective tissues were well-known to generate efficient SHG [8, 9]. Collagen, or collagenous fibers, are the principle and most abundant fibers of the connective tissues. The collagen fiber is consisted of a bundle of fine subunits called collagen fibrils. The diameter of the collagen fibril can be as small as 15 nm, while those in dense connective tissues of tendon may measure up to 200 nm. Inside the collagen fibril exhibits a sequence of closely spaced transverse bands which repeat every 68 nm along the length of the fiber [61], providing necessary condition for bio-photonic crystal effect of SHG activity. The scanning HHGM image taken from the mouse tendon is shown in Fig. 8B. Tendon is known to consist of parallel bundles of collagenous fibers, producing strong SHG.

4.1.6 Skeletal Muscle

Another interesting SHG source in animal tissues is skeletal muscle. A skeletal muscle consists of bundles of muscle fibers called fasciculi. In turn, each fasciculus consists of a bundle of elongate muscle fiber cells. Within the muscle cell are longitudinal units, the myofibrils, which are made up of myosin and actin myofilaments [60]. The myofilaments are organized in a specific manner that gives the muscle fiber a cross-striated appearance. The functionally significant repeating (with ~2 μm period) unit of the myofibril is the sarcomere, which is the segment of the myofibril between two Z lines, containing one section of A band and one section of I band. Inside a sarcomere, the orderly arranged myofilaments provide the necessary nano-sized periodicity for nonlinear photonic crystal effects. Figure 8C shows a HHGM section of the leg skeletal muscle from a one-year-old mouse, illustrating bright SHG in a pattern clearly related to the period of sarcomeric repeats in the myofilament lattice with a ~2-μm period. The sarcomere band pattern of SHG extends continuously across the full width of muscle fibers. On the bottom of the image, intense SHG is also observed from wavy collagen fibrils in the epimysium layer surrounding muscle fibers. The strong SHG emission inside one sarcomere suggests that the SHG from muscle are not just contributed from the birefringence and polarity induced by the orientation of fiber bundle nor just from the periodicity between different sarcomeres, but from the orderly arranged myofilaments inside a sarcomere with a period on the order of nanometers. To explore the molecular source of SHG within the myofilament lattice, Campagnola and his coworkers [15] combined SHG/TPEF to examine the responsible source and the bright SHG stripes were found to be mainly contributed from the A-bands with densely packed myosin/actin filaments inside.

4.1.7 Spindle Array

Mitotic spindles are made up of spatially organized dynamic microtubules, which are 25 nm diameter proteinaceous nanotubes [62] formed by the self-assembly of α and β tubulin heterodimers. Therefore, SHG can be generated from the spindle arrays as the cells undergo mitosis. Campagnola and his coworkers [15] imaged tubulin structures in *C. elegans* embryos and found that SHG could also be generated within spindle array that are not composed of coiled-coil structures. With our higher-harmonic generation microscope, mitosis processes inside a live zebrafish embryo were monitored in vivo. The formation and movement of spindles during the anaphase can be revealed by the SHG modality. Figure 8D shows the mitosis process in the embryonic blastoderm at the 1-k cell stage (2.5-hpf) during the anaphase. The cell membranes can be visualized through THG.

4.1.8 Dentine

Strong SHG was previously observed by Kao et al. in human tooth from enamel that encapsulates the dentine [63], where the SHG was attributed to the highly ordered structures in enamel that encapsulates the dentine. In dentine, strong SHG emission is still observable [64]. Figure 8E shows a HHGM section inside the dentine of a human tooth. While THG provides contrast for the observation of the dentinal tubules, SHG can be observed throughout the dentine, due to highly crystallized non-mineral organic matrix based on collagen fibrils. Figure 8F shows an enlarged HHGM section taken from a different region in the same tooth.

4.1.9 Electric Field and Membrane Potential

Besides studies based on second order optical nonlinearity, SHG microscopy can also be applied to electric field or voltage imaging based on the EFISHG effect, which is a third order optical nonlinearity. Utilizing EFISHG in gallium nitride, we have successfully demonstrated an electric field distribution imaging inside a semiconductor sample using scanning SHG microscopy [6]. Compared with traditional electro-optical sensor probe technique, EFISHG microscopy of electric fields has all the advantage of laser scanning microscope, with high optical resolution and great 3D sectioning power. It is also a background free experiment if the EFISHG material possesses the centro-symmetric property. Recently we have successfully achieved 3D electric field distribution mapping in an IC circuit utilizing nematic liquid crystals as the EFISHG probe material with a sub-micron resolution [7]. Taking advantage of the EFISHG effect, SHG was also shown to have intrinsic sensitivity to the voltage (low frequency electric field) across a biological membrane [11] and was combined with microscopy for

membrane potential imaging [12, 13]. A membrane-staining dye was also used to enhance membrane imaging with SHG microscopy [12, 13, 34]. 1064-nm Nd:YAG lasers or 880-nm Ti:sapphire lasers were used in these studies.

4.2 Polarization Anisotropy in SHG Microscopy

Since the SHG signal is contributed from highly organized crystalline structures, the polarization dependency of the generated SHG on the excitation light thus reflects the local arrangements and structures of the underlying bio-molecules, resembling the symmetry group of a photonic crystal that is characterized by $\chi^{(2)}_{ijk}(2\omega{:}\omega,\omega)$ tensors. By studying the polarization relationship between the coherently generated SHG and the fundamental excitation within the sub-femtoliter foci, different contributing matrix elements of the $\chi^{(2)}_{ijk}(2\omega{:}\omega,\omega)$ tensors can thus be determined explicitly. Recently we have demonstrated that, by studying these polarization dependencies of skeletal muscle fibers using a HHGM, tensor elements of $\chi^{(2)}_{ijk}(2\omega{:}\omega,\omega)$ can be analyzed and obtained [65], which reflect the detailed arrangements and structures of the constructing myosin/actin filaments inside myofibrils.

Inside a myofibril, the orderly arranged myofilaments provide the nanocrystalline-effect which determines the crystal symmetry and the polarization dependency of the measured SHG. We assigned the parallel long axes of myosin/actin filaments as the z axis (Fig. 9B), which is the same as the long axes of parallel aligned muscle fibers and myofibrils. The spatial arrangement of the helical myosin/actin filaments constituting the myofibrils has a hexagonal symmetry in the plane perpendicular to the z axis [61]. Since our spatial selectivity is smaller than the size of a myofibril (~1 μm), it should be the spatial symmetry of the local arrangement of the underlayed myosin/actin filaments (8~20 nm thick) determining the susceptibility tensor of the studied SHG. By taking a scanning image we were measuring 512×512 spatial points. Since the analysis was based on the scanned images, the nonlinear susceptibility we measured was therefore the ensemble average of the local arrangements of the myosin/actin filaments inside different spatial points. With parallel-aligned filaments but loosely arranged nano-hexagonal cells without orientation uniformity in the plane perpendicular to the z-axis, the spatial symmetry of the ensemble-averaged myofilaments becomes cylindrical rather than hexagonal, when examined in an optical scale. It is also interesting to notice that under Kleiman symmetry, which is close to our case, hexagonal and cylindrical symmetries share the same tensor elements.

The non-vanishing effective SHG tensor coefficients $d_{il} = \frac{1}{2}\varepsilon_0\chi^{(2)}_{il} = \frac{1}{2}\varepsilon_0\chi^{(2)}_{ijk}$ in a hexagonal symmetry medium are d_{31}, d_{33}, d_{15} and d_{14} [66]. Here we simplify $l = \begin{bmatrix} 1 & 6 & 5 \\ 6 & 2 & 4 \\ 5 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 11 & 12 & 13 \\ 21 & 22 & 23 \\ 31 & 32 & 33 \end{bmatrix} = jk$ [32]. We naturally assume the symmetry axis as

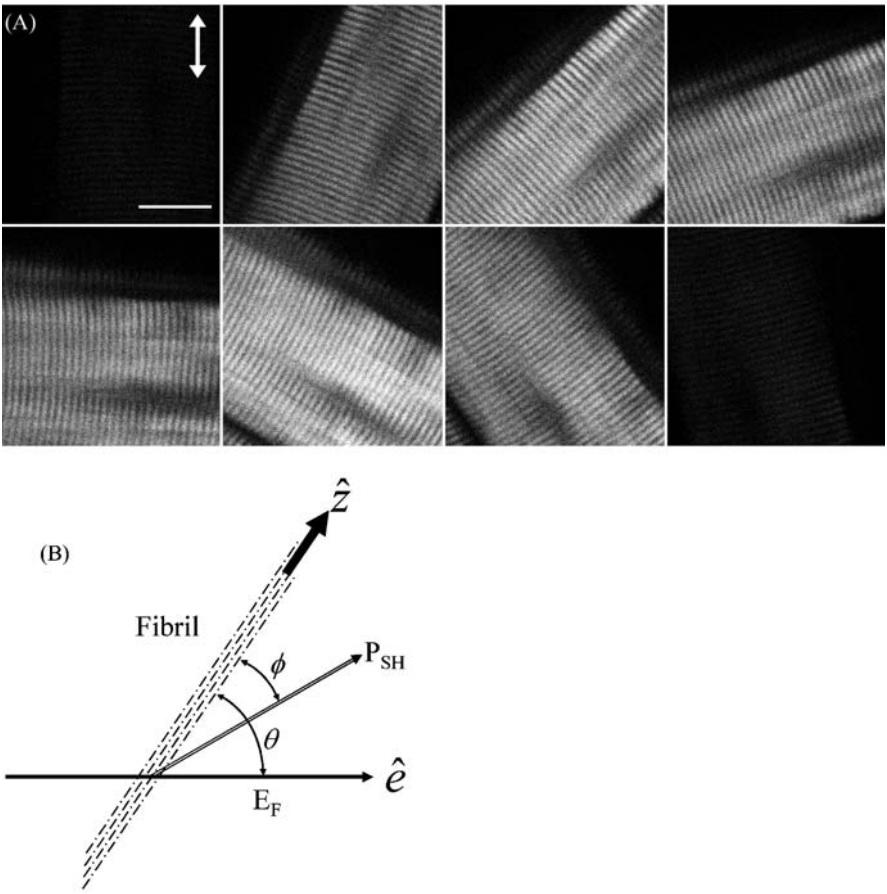


Fig. 9A, B **A** Optically-sectioned SHG image series showing the alternation of signal intensity as the angle between the myosin/actin filaments and fundamental polarization (*shown in arrow*) varies. **B** Schematic drawing showing the relative directions of the assigned parameter. *Scale bar*: 20 μm

the parallel long axes of myosin/actin filaments ($\hat{\mathbf{z}}$ axis). Assuming Kleinman's symmetry holds, then $d_{15}=d_{31}$ and $d_{14}=0$, which is the same as the SHG tensor coefficients under cylindrical symmetry and we follow this parameter setting in the subsequent analysis. The vector expression for the second-order nonlinear polarization $\mathbf{P}^{(2)}$ under cylindrical symmetry assumption is [65–67]

$$\mathbf{P}^{(2)} = a\,\hat{\mathbf{z}}\,(\hat{\mathbf{z}}\cdot\mathbf{E})^2 + \hat{\mathbf{z}}\,|\mathbf{E}|^2 + c\,\mathbf{E}\,(\hat{\mathbf{z}}\cdot\mathbf{E})\,, \tag{4}$$

where $\hat{\mathbf{z}}$ represents the unit vector along the symmetry z axis, $\mathbf{E}=E\,\hat{\mathbf{e}}$ is the incident electric field with fundamental light frequency, and $\hat{\mathbf{e}}$ is the polarization unit vector. In the SHG measurement the incident fundamental field is linearly polarized. *a*, *b*, and *c* are parameters related to the effective SHG coefficient

tensors. Their relations to the d_{il} coefficients are: $d_{31}=b$, $d_{15}=c/2$, and $d_{33}=(a+b+c)$ [67]. Without helicity in myosin/actin filaments, b and c should vanish since $+\hat{\mathbf{z}}$ and $-\hat{\mathbf{z}}$ would be symmetric in this case.

Due to symmetry in the plane perpendicular to the z-axis, we experimentally choose the optical propagation direction $\hat{\mathbf{k}}$ perpendicular to the myosin/actin filament axis $\hat{\mathbf{z}}$ (and incident field polarization $\hat{\mathbf{e}}$). In this case, we can control the incident field polarization **E** in both ordinary and extraordinary axis, which are perpendicular or parallel to the filament axis $\hat{\mathbf{z}}$ respectively. By decomposing the induced SHG polarization into components parallel and perpendicular to the filament axis as

$$\mathbf{P}_{\parallel}^{(2)} = \{d_{31}\sin^2\theta + d_{33}\cos^2\theta\}\, E^2\,\hat{\mathbf{z}}\,, \tag{5}$$

$$\mathbf{P}_{\perp}^{(2)} = 2\, d_{15}\sin\theta\cos\theta\, E^2\,\hat{\mathbf{z}}\times\hat{\mathbf{k}}\,, \tag{6}$$

one can compare the measured polarization behavior with the theory. θ is the angle between $\hat{\mathbf{z}}$ and $\hat{\mathbf{e}}$. The measured SHG intensity I_{SHG} is proportional to $|\mathbf{P}_{\perp}^{(2)}|^2 + |\mathbf{P}_{\parallel}^{(2)}|^2$, and from Eqs. (5) and (6) we have

$$I_{\mathrm{SHG}} \propto E^4\{(d_{31}\sin^2\theta + d_{33}\cos^2\theta)^2 + {d_{15}}^2\sin^2(2\theta)\}. \tag{7}$$

The polarization of the SHG field, which is a combination of SHG electric fields $E_{\perp}^{\mathrm{SHG}}$ and $E_{\parallel}^{\mathrm{SHG}}$ induced by $\mathbf{P}_{\perp}^{(2)}$ and $\mathbf{P}_{\parallel}^{(2)}$ respectively, has an inclination angle φ relative to the filament axis, which can also be determined by Eqs. (5) and (6).

Figure 9A show the acquired SHG intensity dependency on the angle θ between myosin/actin filaments orientation $\hat{\mathbf{z}}$ and the incident fundamental polarization $\hat{\mathbf{e}}$. No polarizer was placed before the detection PMT. The arrow indicates the fundamental polarization direction $\hat{\mathbf{e}}$. It is interesting and straightforward to find that the maximum of SHG emission intensity does not occur as the myosin/actin filaments being parallel or perpendicular to the funda-

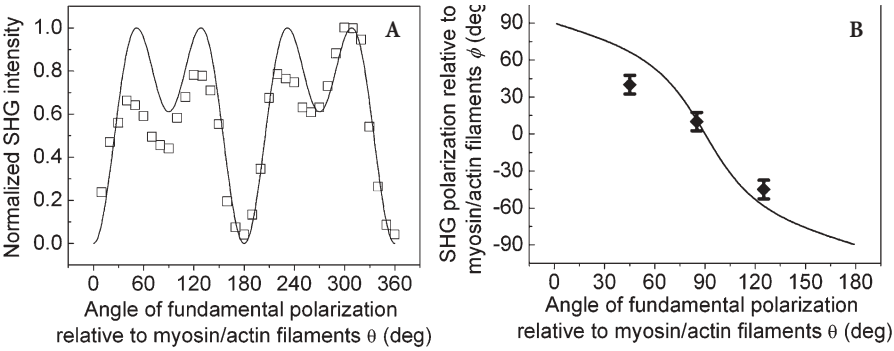


Fig. 10A, B The experimental data showing the SHG: **A** intensity (*shown in open rectangles*); **B** polarization (*shown in diamonds with error bar*) as a function of the angle between fundamental polarization and myosin/actin filaments (θ) along with the calculated fitting (*solid lines*)

mental polarization, while at these situations a global and a local intensity minima can be found instead. Moreover, there are two maxima and two minima within a 180-degree rotation of the sample. The polarization of the emitted SHG can be determined by inserting a polarizer into the collection path.

The complete SHG signal intensity and polarization dependencies on myosin/actin filaments orientation relative to the incident fundamental polarization throughout the 360° rotation are given in Fig. 10. Excellent fitting (solid line) can be obtained based on a cylindrical symmetry assumption with $d_{33}/d_{31} \cong 0.09$ and $d_{15}/d_{31}=1.15$. The SHG intensity and polarization in myofibrils thus follow the cylindrical symmetry (C_∞), reflecting the detailed arrangements of the constructing actin/myosin filaments as expected. The effective SHG susceptibility tensor can be written down as

$$I_{C_\infty}^{(\mathrm{SHG})}\,(1230\ nm \rightarrow 615\ nm) = d_{\mathrm{eff}} \begin{bmatrix} 0 & 0 & 0 & 0 & 1.15\pm0.01 & 0 \\ 0 & 0 & 0 & 1.15\pm0.01 & 0 & 0 \\ 1 & 1 & 0.09\pm0.03 & 0 & 0 & \end{bmatrix} \tag{8}$$

The $d_{33} \cong 0$ result implies that it is difficult to produce an axial second-order polarization with an optical field parallel to the myosin/actin filaments and that the radiated SHG are contributed mainly from $d_{31}=d_{zxx}$ and $d_{15}=d_{xzx}$. Note that the non-zero d_{31} and d_{15} suggest the chirality in myosin/actin filaments inside a myofibril. This suggestion complies with a previous conjecture, which states that the nonlinear optical properties in several biological materials (e.g. collagen matrix and muscle) are induced or enhanced by the chirality in their structure [68, 69].

4.3 Contrast in THG Microscopy

Since SHG occurs from the nano-crystalline structures under study, it is highly desirable to use other non-invasive HHG effect to image general cellular features so that proper correlation of structures can be obtained. Due to the optical dispersion property in biological tissues, THG was proven to be generated from regions with optical inhomogeneity [16, 17, 32] and was applied to image general cellular structures [14, 22–27].

THG is generally a weak process but is dipole allowed; therefore it occurs in all materials, including materials with or without inversion symmetry. In 1995, Tsang [16] reported that when using focused high intensity ultrashort laser pulses, this normally weak THG process becomes highly operative at a simple air-dielectric interface and is much stronger than the bulk of most dielectric materials. It was also reported [16] that the interface THG is a fundamental physical process occurring at all interfaces free from the constraint of a phase-matching condition and wavelength restriction. This could be explained [32] by treating the THG process as a tightly focused Gaussian beam. According to Boyd [32], the

efficiency of THG vanishes when the beam is focused inside a uniform sample for the case of positive phase mismatch ($\Delta k=k_3-3k_1>0$, true for normally dispersive materials) or phase match ($\Delta k=0$). Efficient THG will only occur when the laser beam was focused in some inhomogeneous regions like interfaces, unless the material possesses anomalous dispersion. Taking advantage of this characteristic, in 1997 Silberberg and his coworkers demonstrated THG microscopy for interface imaging in an optical glass fiber [17]. THG microscopy was then applied to image laser-induced breakdown in glass [18], liquid crystal structures [19], and defect distribution in semiconductors [20].

For THG applications in biology, in 1996 Alfano and his coworkers found THG in chicken tissues using a 10-Hz 1064-nm Nd:YAG laser with a 30-ps pulsewidth [21]. After the first demonstration of THG microscopy, it was then quickly applied to image biological specimen due to its interface sensitivity, by using 1200 nm fs light from an OPA [22, 23], 1500 nm fs light from an OPO [24], 800 nm fs light from a Ti:sapphire laser [40], or with 1230 nm fs light directly from a Cr:forsterite laser [14, 25–27, 42]. Detailed cellular and subcellular organelles can be clearly resolved with THG. With THG imaging on zebrafish embryos, continuous viewing can be achieved without compromising viability under the viewing condition with 1230-nm fs light with a 110-MHz repetition rate [26, 27]. With a cubic dependence on the illumination intensity as shown in Fig. 6B, THG provides even better optical sectioning resolution than SHG or TPEF under the condition of using the same illumination wavelength, but is more sensitive to attenuation of the illumination light.

It is also interesting to compare THG with its corresponding SHG images. In general, SHG reflects the orderly arranged nano-structures while THG provides general interfacial profiles due to inhomogeneity. For example, strong THG is induced due to optical inhomogeneity within and surrounding the plant cell wall (Fig. 7A). In Fig. 8, THG picks up morphological interfaces, such as the boundary of silica cells and the walls of dentinal tubules. THG also reveals the cell membrane, collagen fibers, interface between muscle fibers, and the muscular I-bands. Not just cell profiles and structure interfaces, THG also provides information on the various sub-organellar interfaces (Fig. 7B,C, Fig. 8D) like nuclear membranes. It is also interesting to notice that, since THG is allowed for isotropic materials and can be applied to image general morphology, thus has weaker dependence on the incident light polarization. No polarization dependency was found in the THG images of Fig. 8A,D,E,F while strong polarization dependency was observed for THG in muscular I bands [65].

5 Real-Time SHG Imaging

The observation of SHG with a sensitivity comparable to that of linear optical effects could only be possible with high input laser power, which is unsuitable for the study of fragile and sensitive tissues in biological specimens. Hence,

SHG imaging found limited applications in studies of live specimens until a way of enhancing the nonlinear signal could be found. The advent of mode-locked lasers in the eighties finally provided an adequate light source to exploit the characteristics of nonlinear optical phenomena. The high peak power of fs pulses maximizes the SHG intensity for a given average input power while, however, acute photodamage is also induced due to high optical flux.

König et al. have demonstrated that the loss of cell viability occurred at an average power of ~7 mW for 150-fs pulses (with a 80-MHz repetition rate), corresponding to a peak power of 0.56 kW, a peak intensity of 0.85 TW/cm^2, or a pulse energy of ~0.1 nJ [48–50]. In the same study they also verified the nonlinear nature of the optical damage in the fs pulse regime. From an experimental point of view, the breakdown threshold can be defined by the observation of bubble formation at the laser focus [70]. The results of A. Vogel et al. showed that multi-photon ionization and free-electron-induced chemical bond breaking, not relating to heating or thermoelastic stresses, probably mediate the highly localized ablation of intracellular structures and intranuclear chromosome dissection [70]. To avoid the nonlinear process dominated optical damage, the maximal obtainable SHG power is limited by the applicable peak intensity and pulse energy instead of its average applicable power.

To realize real-time SHG microscopy of live specimens without sacrificing cell viability, the peak intensity and the pulse energy of the applied fs illumination can not exceed a tolerable threshold value. The average SHG signal can thus be increased in two ways – by increasing the number of foci or by increasing the repetition rate of a laser.

Kawata et al. [71] have recently demonstrated a CCD-based multifocus SHG microscope in which the SHG detection efficiency and the image acquisition rate are improved by several tens of times compared with typical single-focus scanning techniques. By use of a microlens array scanner, the laser beam of an 82-MHz mode-locked Ti:sapphire laser was split into ~100 beamlets, and the same number of foci were formed in the specimen. By projecting the SHG image excited by the parallel scanned foci enables real-time imaging of the live specimen. Since the frame rate of scanning laser microscopes depends on how much signal light is available from a specimen, the frame rate of the demonstrated system is thus improved ~100 times compared with that obtained with a single-focus scanning method. The imaging properties and the usefulness of the microscope are demonstrated by SHG imaging of living HeLa cells and cultured rat cardiac myocytes [71]. In their system, ~100 foci were produced from a single beam of Ti:sapphire laser of 170 mW power at the focal plane, corresponding to 1.7 mW per focus. Cell viability was thus maintained by using multifocus scanning, while the frame rate can get faster without increasing the illumination pulse energy at the focal spots. The drawback of this real-time system is that it cannot be applied to optically thick samples, the projected image of which on to the imaging CCD will be strongly distorted.

For optically thick samples, we have demonstrated a real-time PMT-based SHG microscope using the combination of a 2-GHz repetition rate fs Ti:sap-

phire laser and a fast XY galvanometer mirror scanner [39]. With identical peak intensity and pulse energy, the mean power of the emitted SHG will linearly increase with the laser repetition rate while no nonlinear photodamage will take place if we keep the peak intensity and pulse energy well below the photodamage threshold. While acquiring dynamic images in nonlinear microscopy, the pixel dwelling time must be long enough for the detector to collect sufficient photons. As a result, increasing the repetition rate and thus increasing the nonlinear signal intensity can efficiently reduce the pixel dwelling time as well as the acquisition time per frame. By increasing the repetition rate of the fs laser pulse train ~25 times, SHG average intensity could be significantly improved while cell viability is maintained due to the under-threshold peak intensity and pulse energy.

Figure 11A shows the real-time in vivo SHG sectioned image series of moving zebrafish muscle fibers taken with the 2-GHz repetition rate Ti:sapphire laser. Individual sarcomeres can be visualized through the in vivo SHG images with a bright A-band and a dark I-band as a period. The average power incident on fish muscle fibers was 150 mW, corresponding to a 75-pJ pulse energy and a ~0.1-TW/cm^2 peak intensity with a 60×/NA1.2 water immersion objective. This peak intensity is much smaller than the peak intensity threshold for the loss in cell viability demonstrated by König et al. [48] while the average incident power, however, is one order of magnitude larger than that in [48]. In consequence, no photodamage can be observed while the fast movement of muscle fibers inside a free-moving zebrafish can be in vivo traced in real time. Furthermore, the contraction of muscle fibers inside a vertebrate can also be monitored in real-time through SHG microscopy, which is the first time to our knowledge. According to our result, the maximum contraction of each sarcom-

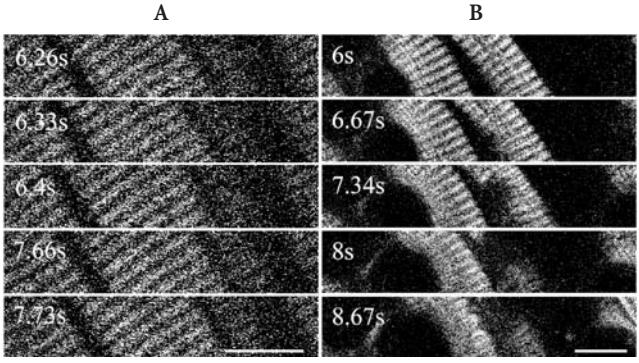


Fig. 11 **A** Time series of SHG in vivo microscopic images in live zebrafish muscle tissue taken with a 2-GHz, 150-mW Ti:sapphire laser under a frame rate of 15 fps. No damage can be observed with such intensity (0.1 TW/cm^2) and pulse energy (75-pJ). **B** Similar image series in live fish muscle tissue taken with an 82-MHz, 115-mW Ti:sapphire laser. Optical damage such as bubble formation due to the excessively strong peak power (1.6 TW/cm^2) and pulse energy (1.4 nJ) is obvious. *Scale bar*: 10 μm

ere is ~100 nm, which is approximately 5% of its total length. Due to the speed limitation of the utilized galvanometer scanner, the maximum frame rate was ~15 frame per second (fps) for images with 100-×512-pixel resolution, which was used in Fig. 11.

On the other hand, if we apply the conventional 82-MHz Ti:sapphire laser with a similar average incident power (115 mW), the corresponding pulse energy and peak intensity will be as high as 1.4 nJ and 1.6 TW/cm^2, respectively. Under this extremely strong incident optical field, most biological materials will suffer from violent nonlinear photodamage in company with bubble formations. This destructive process in live zebrafish muscles is shown in Fig. 11B. It can be observed that the optical damage occurs in less than one second after the onset of laser irradiation. A bubble can be seen clearly after 2 s irradiation and one of the muscle fibers has been totally destroyed. If the pulse energy of the 82-MHz laser was attenuated to the safety region (100 pJ) approximately the same as that of the 2-GHz laser as shown in Fig. 11A (with a higher peak intensity 0.11 TW/cm^2), the average power will decay to only ~8 mW, which is a commonly accepted power level when implementing nonlinear microscopy with fs Ti:sapphire lasers. However, with identical measurement parameter settings as in Fig. 11A (including PMT voltage, black and white levels, and frame rate), no SHG from muscle fibers can be observed at all in the real-time scanning image.

6
Applications

6.1
Electric Field Imaging

As a result of the wide use of electronic and optoelectronics devices, the need for an efficient tool to characterize integrated circuits is growing. Electric field sensing is a direct and non-invasive way to perform this kind of characterization. Electron-beam-based systems are widely used but require heavy technical efforts. Thus it is beneficial to find out easier alternatives for observing electric field. Electro-optic (EO) sampling has previously been widely used to probe electric field in circuits [72]. However, to obtain the distribution of electric field (E-field) with EO techniques, the EO crystal needs to scan through the region of interest point by point and their transverse resolution is limited by the size of the probe head (~10 μm) while the axial resolution is even worse. Utilizing EFISHG to probe E field intensity has been studied for a long time. The intensity of the EFISHG is proportional to the square of the optical excitation intensity as well as the square of the probed E-field magnitude (voltage) [73]. E-fields in different interfaces have been studied with EFISHG effects, including Au/Si [74], Si/SiO_2 heterojunctions [75], and Au/GaAs Schottky barriers [76]. EFISHG has also been applied to monitor bio-membrane potentials [13] and to map the

piezo-electric fields in bulk GaN [20] and InGaN quantum wells [6] with 3D resolution. Similar to scanning SHG microscopy, EFISHG techniques can provide excellent 3D sectioning capability because of the quadratic power dependence. The visualization of 3D E-field distribution with sub-μm resolution could therefore be achieved. With the help of a fast scanner, the focused laser spot can scan across the focal plane faster than video rate. Thus one can directly visualize real-time images of the E-field through a reflection type SHG microscope similar to the one in Fig. 3.

We have recently demonstrated direct visualization of E-field through a scanning SHG microscope with the aid of the strong EFISHG effect in nematic liquid crystal (NLC) [7]. We choose NLC as the SHG-active media for several reasons. First, strong EFISHG have been observed in NLC [77, 78]. Since the EFISHG from NLC is generated inside the bulk, 3D E-field visualization is thus achievable, different from previous interface techniques [74–76]. Second, NLC is non-conducting and transparent, which allows the fundamental and the SHG signal to transmit with minimum attenuation. Third, with non-prealigned NLC, there is no SHG background without E-field applied, resulting in a background-free technique. Increasing the exciting light intensity can thus enhance the magnitude of EFISHG signal without increasing the background noise, improving the signal/noise ratio. Moreover, for real-time visualization, a strong signal is required. The efficiency of EFISHG effect of NLC is much higher than from the surfaces and semiconductors, such as GaN. In addition, the polarization dependency of the EFISHG effect in the non-prealigned NLC enables one to determine the direction of the E-field vector.

The setup of the scanning SHG microscope was similar to the one in Fig. 3. The light source was the fs Cr:forsterite laser centered at 1230-nm. The laser was focused onto the sample with a 60×/NA1.2 water-immersion objective (Olympus UPLAPO60×W/IR), resulting in a sub-μm beam waist. Incorporating an 8-kHz XY scanner (VSH-8, GSI Lumonics) enabled the focused laser spot to scan across the focal plane at >40 fps. The induced EFISHG image at a wavelength of 615 nm was then projected onto a high frame-rate CCD camera (Retiga-1350, Qimaging) for direct visualization while a color filter was used to remove the fundamental light at 1230 nm.

Figure 12A shows the SHG response in NLC vs applied voltage in a log-log scale. The slope of the fitting line is 2.06, showing the square dependence on applied voltage as expected for EFISHG. Since there was no pre-alignment of the applied NLC, no directional preference will be selected by the bulk NLC before applying the electric field due to optical centro-symmetry. It is important to notice that no SHG can be detected as the applied bias goes to zero, indicating a background-free measurement. Figure 12B shows the dependence of EFISHG intensity on the angle between the E-field and the laser linear polarization. A sinusoidal response can be observed, which is attributed to the projection of E-field in the direction of pump polarization. As shown in Fig. 12B, the EFISHG maximum is reached when the polarization is parallel to the static E-field. On the other hand, a minimum that is below the measurement

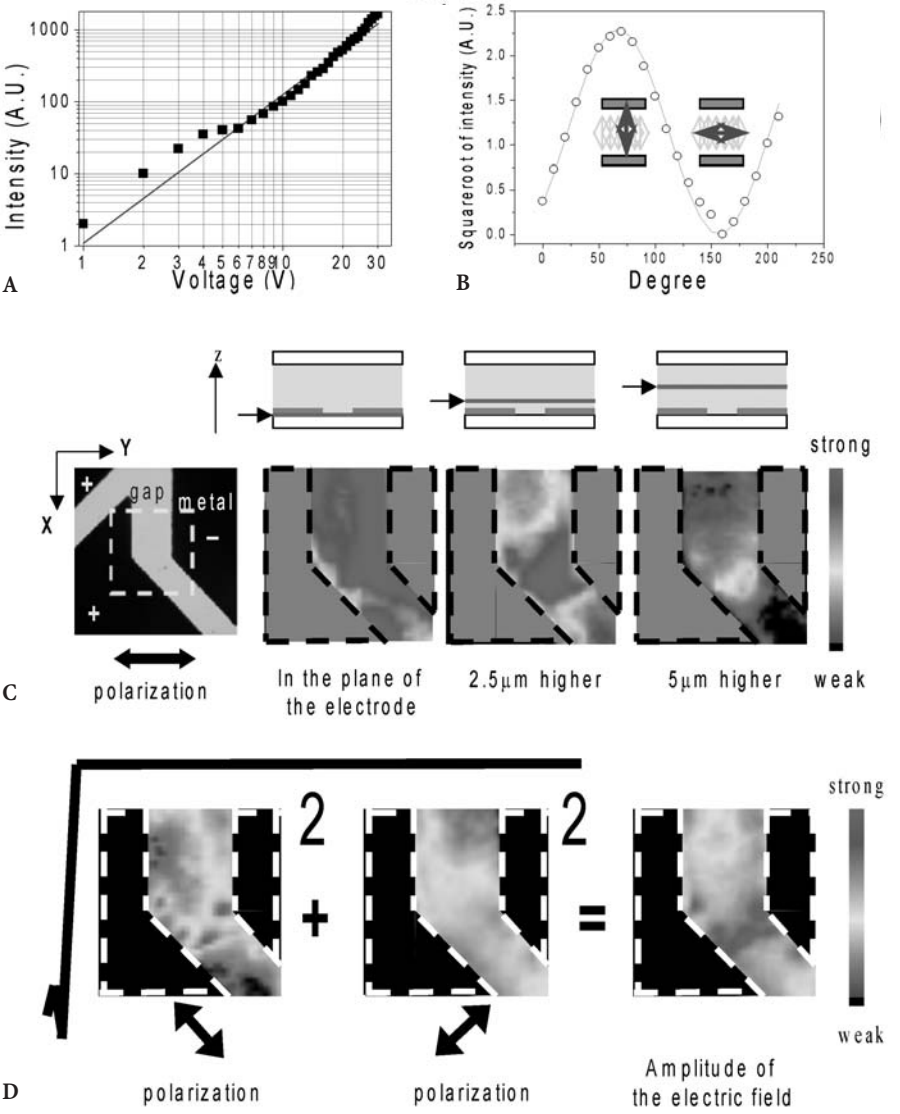


Fig. 12A–D Visualization of E-field using EFISHG in NLC: **A** SHG intensity in NLC vs applied voltage. Square dependence on applied voltage is observed as expected for EFISHG. Notice the low SHG background when the applied voltage is zero; **B** SHG response vs excitation light polarization. The maximum occurs when the pump field is parallel to the E-field; **C** sectioned images measuring the Y component of the E-field with different sectioning heights. The picture of the test sample and the corresponding imaging area is also shown for comparison; **D** determination of the amplitude of the E-field projected in the XY plane. We visualize the E-field components in two perpendicular directions ($\hat{Y}+\hat{X}$ and $\hat{Y}-\hat{X}$) with linear polarized lights. To obtain E-field amplitude distribution in the XY plane, the value of each pixel in these two images was squared, added together, and then taken its square root. Image size: 20×20 μm^2

sensitivity can be observed when the pump polarization is perpendicular to the E-field. The dominant component of the EFISHG tensor leading to the observation is $\chi^{(3)}_{xxxx}$ (2ω:0,ω,ω). With linear polarized light, it is then possible to determine the E-field vector projected in the XY plane since EFISHG selects the E-field component parallel to the excitation polarization.

Figure 12C shows the sectioned image measuring the Y component of the E-field with different sectioning heights. The picture of the test sample and the corresponding imaging area is also shown for comparison. The images were obtained by taking the square root of the EFISHG signal intensity. With different colors representing different amplitudes, it is obvious to see the magnitude of the Y component of the electric field decreases sharply with height, as expected. Excellent Z resolution on the order of ~1 μm can be obtained.

By applying two perpendicular excitation polarizations, the amplitude as well as the direction of the E-field vector in the XY plane can be obtained. On the other hand, applying circular polarized light can obtain the E-field intensity in the XY plane with one measurement. After taking the square root on the pixel values of the EFISHG image excited by a circular polarized light, the amplitude of the E-field can be obtained. However, the E-field direction has to be obtained with two independent measurements. Figure 12D shows the determination of the amplitude of the electric field projected in the XY plane. We visualize the E-field components in two perpendicular directions ($\hat{Y}+\hat{X}$ and $\hat{Y}-\hat{X}$) with linear polarized lights. To obtain the E-field amplitude distribution in the XY plane, the value of each pixel in these two processed images was squared, added together, and then taken its square root. On the other hand, to obtain the directions of the E-field vectors, the amplitude of one E-field component was divided by the value of the corresponding pixel in the other sectioned image. The angle (θ) between the electric field and the polarization was then calculated from the quotient (q) with $\tan\theta=q$. However, there is an ambiguity for determining θ because we could not tell θ from ($\pi-\theta$) due to that only the absolute values of E-field components are measured. In order to overcome this ambiguity, an extra single point measurement can be applied with a third laser polarization.

6.2 Developmental Biology

Highly penetrative and non-invasive in vivo microscopy with <500 nm spatial resolution has the potential of offering new insight into the embryonic morphological changes and the complex developmental processes. HHG, including SHG and THG, leaves no energy deposition to the interacted matters due to their virtual-level-transition characteristics, providing the optical non-invasive nature desirable for life microscopy. By using endogenous HHG as the contrast agents, no fluorescence is required and the common issues of photodamage, phototoxicity, photobleaching, or dye availability can all be eliminated. Based on a Cr:forsterite fs laser centered at 1230 nm, which is the transparency window of most biological specimens, we have recently demonstrated the

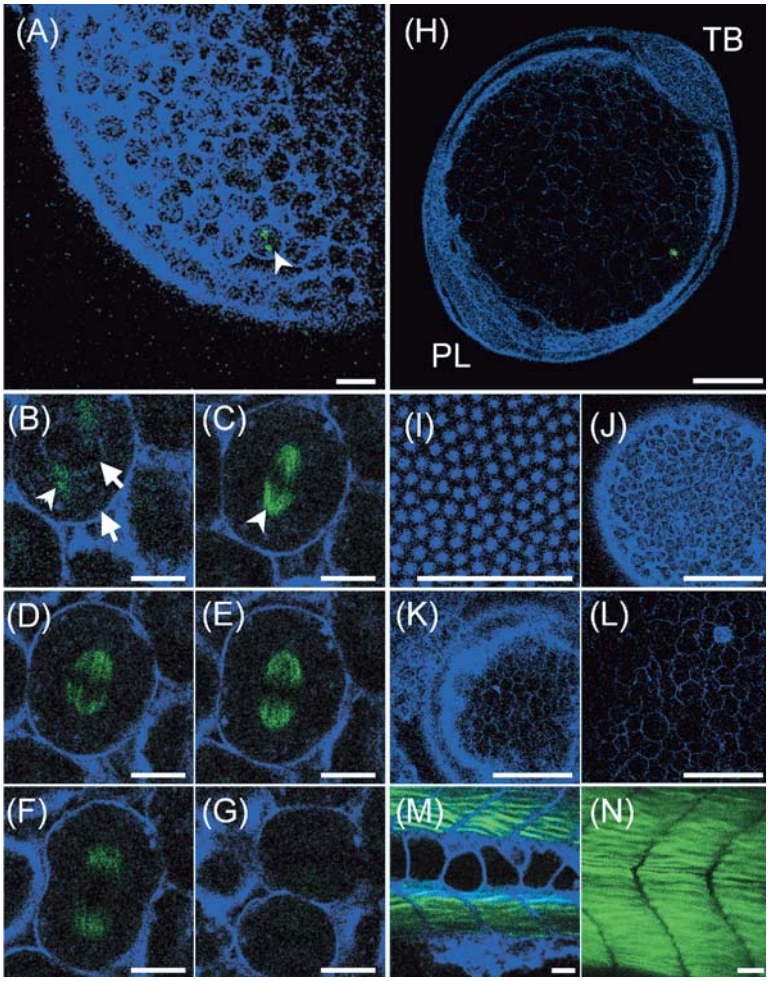
non-invasive and highly penetrative HHGM that is highly suitable for in vivo embryonic developmental observations. The complicated development within a >1 mm thick zebrafish *(Danio rerio)* embryo from initial cell proliferation, gastrulation, to the tissue formation were all observed clearly in vivo without any treatment on the live specimens. The excellent sub-micron resolution and the non-invasive characteristic of HHGM allows us to capture the subtle developmental information on the cellular or sub-cellular levels occurred deep inside the live embryos and larvae [26, 27].

Our home-built laser scanning HHGM for the developmental biology study was adapted from an Olympus FV300 scanning system combined with an Olympus B×51 upright microscope while all optics were modified to allow the passage of 1200–1350-nm infrared light. The collimated laser beam of the Cr:forsterite laser was coupled into the scanning system connecting to the microscope with an aperture fitting tube lens. Real-time scanning was accomplished through the high-speed galvanometer mirrors inside FV300. In order to visualize all cell activities inside the live zebrafish embryo, the working distance of the chosen objective has to be longer than the thickness (>1.5 mm) of the embryo, including the chorion. We choose a 2-mm working-distance high NA IR objective (LUMPlanFl/IR 60×/water/NA 0.90, Olympus) for this study. The forward-propagated HHG is collected using a high-NA (1.2) oil-immersion condenser. For HHG intensity mapping, two PMTs (R928P, Hamamatsu) were used to record the filtered SHG and THG intensities respectively point by point to form 2D sectioned images. To observe live moving specimens, the recording speed was chosen to be 0.4–4 fps with 512×512-scanned points. The live and free-moving wild-type zebrafish embryos under observation were kept inside an aquarium formed by slide glasses. No treatment was applied to the live specimens.

The optical sectioning power and non-invasive nature of SHG and THG allow us to observe 3D cell proliferation processes inside embryos in vivo. For instance, the spatial variations of linear or the third order nonlinear susceptibilities between nuclear membrane and cytoplasm or between cell membrane and the surrounding fluid can provide detectable THG signals. As a result, the cell structure in the embryo and the distribution of the organelles can be revealed by in vivo THG microscopy (Fig. 13). On the other hand, strong SHG can be observed in the spindles as the blastoderm cells undergo mitosis (arrowhead in Fig. 13A,B) due to an optical crystalline effect. The dynamic changes of spindle and membrane between two daughter cells can be imaged in vivo by the developed HHGM (Fig. 13B–G). Endogenous SHG signal vanishes as the spindle microtubules disperse in the cells and exhibit no more crystalline characteristic. No optical damage can be observed during the whole proliferation process even with 100 mW incident average power on the embryo after continuous long-term observations. Compared with common TPEF techniques using a 730~800-nm fs Ti:sapphire laser at a similar experimental condition, Chinese hamster ovary cells were found to be unable to form clones with more than 6 mW average power illumination and complete cell destruction occurs at an average power more than 10 mW [48]. Our successful long-term observation

with 100 mW average power indicates not only the reduction of linear absorption with longer IR light of 1230 nm, but also the reduction of multi-photon damages. We can successively monitor the development of the embryos throughout a 12-h period of time continuously and the examined embryos can all develop normally to larval stage after they were kept in the container for one day.

The excellent depth-resolution and high penetration capability of the HHGM are demonstrated in Fig. 13H–L. Figure 13H shows an in vivo optical sectioning in a live zebrafish embryo at a depth of 700 μm from the chorion, showing the developing polster and tail bud (chorion are outside the viewing area). The yolk granule membranes and the semicrystalline membrane proteins in the yolk cells can be picked up by THG and SHG signals, respectively. Moving the focus to the sample surface, THG signals reveal the 0.5~0.7 μm width



granular canals distributed on the chorion surface (Fig. 13I), demonstrating its sub-μm resolution. At a depth of 420 μm from the chorion, the blastomere at the embryonic shield can be observed. Individual cell nucleus, cellular membrane, and the boundary of the blastomere can all be observed through high-resolution THG microscopy due to their optical susceptibility differences with the surrounding media (Fig. 13J). Moving 50 μm deeper into the blastoderm (Fig. 13K), the deep cell layers (at left), the yolk cells (at right), and the internal yolk syncytial layers between them all showed up in the scanning sectioned images, confirming the capability of THG microscopy to be a general non-invasive imaging tool for cell and tissue morphology. At a 570-μm depth in the embryo (Fig. 13L), the densely packed yolk granules were clearly observed with excellent spatial resolution. Currently we are able to demonstrate >1.5 mm high penetration ability into the embryo [27].

Besides observing the embryo developments, HHGM was applied to the examination of general tissue structures and the crystalline nano-structures in live free-swimming larvae. The general morphological structures can be revealed by THG signals including the larva skin, eyes, brains, boundary of somite, and notochord. On the other hand, the detailed distribution of early developing muscle fibers in a somite can be observed through SHG (Fig. 13M,N). Again this high penetration microscopy allows us to image through a live free-swimming zebrafish larva while sub-micron-scaled morphological developments can be revealed.

To our knowledge, no other technique can provide such high-resolution, high-contrast in vivo images of developmental processes deep inside live

Fig. 13A–N Mitosis processes inside a live zebrafish embryo in vivo monitored with higher harmonic generation microscopy: **A** an optical section of the embryo at the dome stage. The imaging depth is about 400-μm from the chorion surface. THG (*shown in blue*) can pick up all interfaces including external yolk syncytial layers, cell membranes, and nuclear membranes while SHG (*shown in green*) shows the microtubule-formed spindle biconical array; **B–G** time series of the mitosis with a 70-s interval between each image; **B** the cell at center is in its anaphase. The intact and dissolved nuclear membranes are revealed by THG (*arrow*) while the centrosome (*arrowhead*) is revealed by SHG; **C** the separation of chromosome and spindle fibers is revealed by SHG signals (*arrowhead*); **D–F** the center cell goes into telophase. Decreased SHG intensity indicates the dispersed spindles. THG shows the beginning of cytokinesis; **G** newly-formed cell membrane indicates the end of a mitosis cycle. **H–L** In vivo higher harmonic generation sectioning inside a live zebrafish embryo at the 2-somite stage; **H** a sectioning showing the whole embryo at a depth of 700 μm from the chorion surface (ventral view). PL: polster; TB: tail bud; **I** THG image of the chorion surface, showing the <1 μm diameter granular canals; **J–L** the depth-resolved optical sections at depths of 420 μm, 470 μm, and 570 μm inside the embryo. **M, N** In vivo higher harmonic generation sectioning inside a live zebrafish larva at the 20-somite stage; **M** an optical section at the center of the larva showing the segments inside the vacuolated notochord and the distribution of somites alongside the notochord. Interfaces between somites are clearly revealed with THG; **N** SHG imaging inside somites showing individual muscle fiber with SHG. *Scale bar*: 20 μm for A–G, I, M, and N; 100 μm for H and J–L

vertebrate embryos and larvae without extra labeling or handling. The excellent 3D (better than 400 nm) resolution of the demonstrated technology, a truly non-invasive and highly penetrative microscopy, allows one to capture the subtle developmental information on the cellular or sub-cellular levels occurring deep inside the live embryos and larvae without alternating their natural biological and optical environments. This technique can not only provide the 3D in vivo observation of cytoarchitecture dynamics during embryogenesis with sub-micrometer resolution and millimeter penetration depth, but could also make strong impact in developmental biology studies, for example, studying cell-fate determination at early developmental stage.

6.3 Non-invasive Biopsy of Skin and Tumor Diagnosis

Traditional biopsy requires the removal, fixation, and staining, and examination of tissues, cells, or fluids from the living body. The histological procedure is invasive and painful. Non-invasive in-vivo optical biopsy is thus required, which should provide non-invasive highly-penetrative 3D imaging with sub-micron spatial resolution. OCT [45, 46] is one of the potential methods for optical biopsy of skin, which can provide high penetration using 1.3 μm wavelength light [45, 46] or sub-micron axial resolution with a 725-nm broadband light source [79]. Optical biopsy based on scanning TPEF microscopy is another choice [80, 81]. Compared with OCT, laser-scanning nonlinear microscopy provides much higher lateral resolution while the sectioned 2D images are in the plane parallel to the surface, in contrast to the cross-sectional (tangential) images provided by OCT. Previous TPEF biopsy of skin based on 780-nm femtosecond light [80, 81] provides high resolution imaging from the skin surface through the epidermal-dermal junction. Without using exogenous fluorophores, autofluorescence in the epidermis layer is predominately from the reduced pyridine nucleotides. In the dermis the contrast is provided by the autofluorescence of the collagen and elastin. However for future clinical applications without surgery, current 700–850 nm based laser scanning technology still presents several limitations including low penetration depth, in-focus cell damages, multi-photon phototoxicity due to high optical intensity in the 800 nm wavelength region, and toxicity due to required exogenous fluorescence markers. Previous research has also indicated that using shorter wavelength for more efficient TPEF produces greater photodamage [48]. The reports on the on-focus multi-photon-absorption-induced photo-damage induced by Ti:sapphire lasers in TPEF microscopy indicate the importance to reduce possible tissue damage while performing optical biopsy of skin with an high intensity light source.

We have recently demonstrated a higher-harmonic generation biopsy of skin using the fs Cr:forsterite laser centered at 1230 nm [42]. According to previous studies, light attenuation (including both absorption and scattering) in human skin reaches a minimum around 1300 nm wavelength [43] due to the combi-

nation of diminishing scattering cross-section with wavelength and avoiding the resonant molecular absorption of common tissue constituents such as water, melanin, and hemoglobin. Moving the operating wavelength of an optical biopsy to 1200–1300 nm spectral region can not only increase the penetration depth in skin but also reduce the multiphoton absorption cross-sections and thus reduce the potential photodamage and phototoxicity.

For the preliminary study, we adapted a system similar to the one used in developmental biology study so that the scattered and reflected backward SHG signal could also be collected in the reflection geometry using the focusing objective.

Figure 14A,B shows the backward-SHG (in green), forward SHG (in red), forward-THG (in blue), and the combined cross-sectional (x-z) as well as the paradermal (x-y) scanning sectioned images of a fixed integument sample removed from the caudal dorsal plane of a B6 mouse. Due to momentum conservation, the generated HHG photons prefer forward emission and can thus

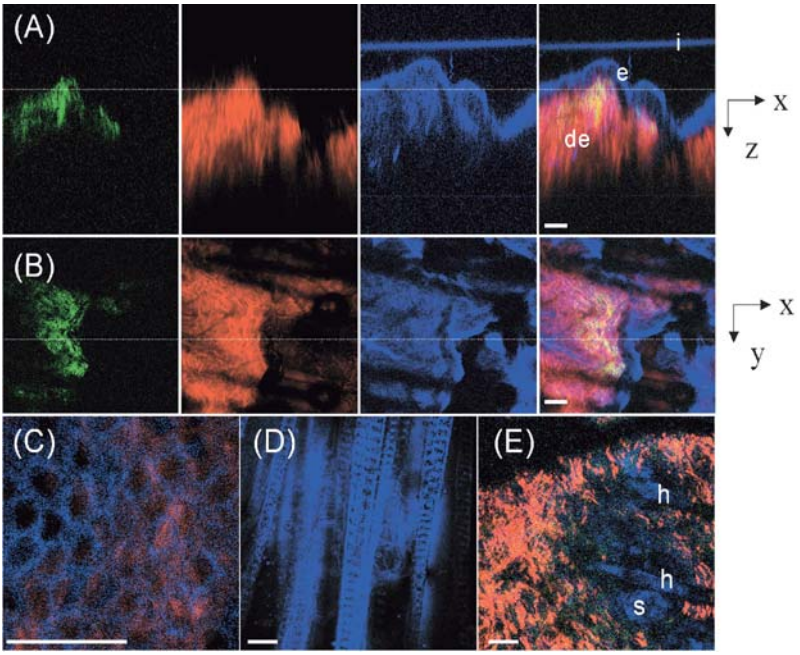


Fig. 14A–E Backward-SHG (*shown in green*), forward-SHG (*shown in red*), forward-THG (*shown in blue*), and the combined cross-sectional (x-z, into the skin (A) as well as the paradermal (x-y, parallel to the surface (B) scanning images of mouse skin. *Dotted lines* in (B) mark the corresponding z(y)-position for B(A). **C** Sectioned paradermal image in the Stratum basal layer. **D** Sectioned image of hairs above skin. **E** Sectioned paradermal image deep in the dermis. Image size: 235×310 μm^2 for A; 235×235 μm^2 for B,D,E; 59×59 μm^2 for C. i: coverglass-formalin interface. e: epidermis layer. de: dermis layer. h: hair. s: sebaceous gland. *Scale bar*: 30 μm

be collected using the transmission geometry. However for clinical non-invasive optical biopsy, reflection geometry is preferred due to optically thick human bodies. From the cross-sectional images, the general histological structures in both the epidermis and dermis layer can be identified using THG due to its sensitivity to local optical inhomogeneity. The coverglass-formalin interface was also picked up by THG in the top part of the cross-sectional image. Forward-SHG, on the other hand, reflects the distribution of connected tissues in the dermis layer. It is interesting to compare the backward-SHG with the forward-SHG images. It can be found that the backward-SHG not only displays the distribution of the generation sources (like connective tissues), but also the scattering and reflection properties of the tissues. In this specific set of data, the detected backward-SHG intensity was on the order of 1/10 of the forward-SHG intensity, indicating the importance of scattering in the collection of backward-SHG. Taking a paradermal sectioning in the surface region (Fig. 14B), the observed THG images were dominated by the top contour profile of the Stratum corneum due to the large index difference between the Stratum corneum-formalin interface while weak THG signals pick up fine structures in both epidermal and dermal layers including wavy collagen fibers in connected tissues.

Figure 14C also shows a sectioned image in the Stratum basal layer on top of the epidermis-dermis junction. The cell membranes of densely packed basal cells can be clearly resolved by THG signals. Figure 14D shows a clear hair image taken from the region above epidermis. Moving the focal plane deep into the dermal layer, Fig. 14E shows a typical sectioned image where the SHG picks up the connective tissues consisting of collagen fibers and the THG reflects the hairs and the sebaceous glands around.

Our study suggests that HHGM could provide sub-micron resolution deep-tissue non-invasive biopsy images without using fluorescence and exogenous markers, thus being useful for non-invasive biopsy of skin diseases and for characterization of mucosal tissues of the cervix, colon, ear, nose, throat, oral cavity, and esophagus, where diseases start at the epithelial layer a few hundred microns beneath the surface [35].

Alfano and his coworkers have demonstrated in cheek pouch mucosa tissue that the spatial mapping of the backward SHG signals [35] generated from the fibrillar collagen can reveal the depth differentiation between normal, dysplasia, and a more advanced cancerous state. The content and structure of collagen is essential in governing the delivery of therapeutic molecules in tumors. Jain and his coworkers also recently demonstrated [36] that by using SHG to image the fibrillar collagen in tumor growing in mice, one could estimate the relative diffusion hindrance, quantify the dynamics of collagen modification after pharmacologic intervention, and provide mechanistic insight into improved diffusive transport induced by the hormone relaxin. Their study indicated that SHG microscopy could offer researchers and clinicians an enhanced ability to estimate the relative penetrabilities of molecular therapeutics.

Our previous study [82] also indicates that HHGM can distinguish normal and abnormal tissues in the human oral cavity at the cellular level, thus pro-

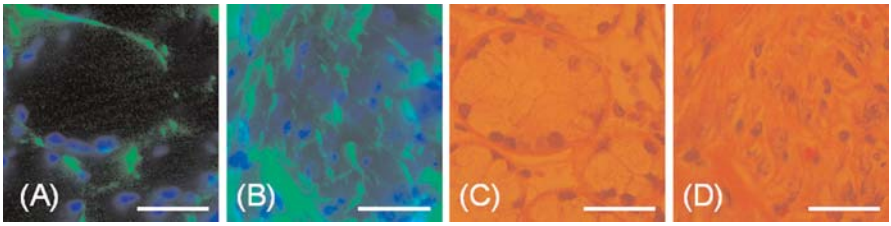


Fig. 15A–D **A** THG (*shown in blue*) and SHG (*shown in green*) scanning images of normal submucosa tissue. **B** THG and SHG scanning images of abnormal submucosa tissue, showing groups of tumor cells and overgrown connective tissues. **C,D** Conventional image of H & E stained sections showing the same area as A and B, respectively. *Scale bar*: 30 µm

viding a potential tool for non-invasive diagnosis of early oral cancers. Under normal clinical examination, early oral cancers are difficult to differentiate from innocuous tissues. Therefore, the early stages of oral neoplasia are often neglected. For the detection of oral neoplasia, incisional biopsy remains to be the most reliable diagnostic method. However, even an experienced clinician cannot easily choose a representative location of an oral lesion for biopsy. Therefore, it is worthwhile to develop a convenient and non-invasive diagnostic tool that can detect oral cancer in the early stage.

Figure 15A shows a typical HHGM image of normal submucosa tissue taken from the oral cavity of a patient with squamous cell carcinoma. THG reflects the nucleus of the glands while SHG reflects the distribution the connective tissue. Figure 15B is a comparison HHGM images illustrating groups of tumor cells showing fusiform differentiation, also in submucosa of the surgically removed oral carcinoma tissue of the same patient. Figure 15C,D shows the conventional microscopic images of the H&E stained histological sectioning of the same areas corresponding to Fig. 15A,B. Our preliminary study also indicated the capability of higher harmonic generation for future non-invasive biopsy of skin and mucosal cancers.

7 Conclusions

Light microscopy, fluorescence in particular, has been used extensively in the correlation between structures, molecules, and functions in modern biological sciences. Either intrinsic or extrinsic fluorescent probes are needed in fluorescence microscopy using single- or multi-photon excitation schemes. Hence, dye penetration, probe toxicity, and photo-bleaching/damage are the limiting factors frequently encountered. Different from fluorescence processes, higher harmonic generation processes are based on virtual transitions. The marked advantage of this virtual transition during wavelength conversion is the lack of energy deposition, thus no photo-damage or bleaching from the process is

expected. Higher harmonic generation microscopy can hence be considered as a truly "non-invasive" modality and is ideal for in vivo imaging of live specimens without any preparation. Due to its nonlinear nature, the generated SHG intensity depends on square of the incident light intensity, while the generated THG intensity depends on cubic of the incident light intensity. This nonlinear dependency allows localized excitation and is ideal for intrinsic optical sectioning in scanning laser microscopy. SHG microscopy can provide images on stacked membranes and arranged proteins with organized nano-structures due to the bio-photonic crystal effect. On the other hand, THG microscopy can provide general cellular or subcellular interface imaging due to optical inhomogeneity.

Due to the virtual nature of higher harmonic generations, no saturation or bleaching in the generated signal is expected. With no energy release, continuous viewing without compromising sample viability can thus be achieved. Through proper selection of laser wavelength, sub-micron resolution and 1.5 mm penetration was simultaneously accomplished by using higher harmonic generation microscopy for the 3D in vivo imaging of embryonic dynamics in live zebrafish embryos [27]. No optical damage was observed even with high illumination after long-term observations and all the examined embryos developed normally at least to the larval stage, indicating prolonged viability under the viewing condition of higher harmonic generations [26].

Higher harmonic generations provide a truly non-invasive microscopy and biopsy tool with sub-micron 3D spatial resolution and millimeter penetration without using fluorescence and exogenous markers. HHG can be applied to live specimen without compromising viability, while high resolution morphological, structural, functional, and cellular information of biological and medical specimens can be obtained. With the flexibility to be combined with fluorescence based technology, this novel optical tool is likely to become a major imaging modality for biomedical studies.

References

1. Hellwarth R, Christensen P (1974) Opt Commun 12:318
2. Sheppard CJR, Kompfner R, Gannaway J, Walsh D (1977) IEEE J Quantum Electron 13:100
3. Boyd GT, Shen YR, Hänsch TW (1986) Opt Lett 11:97
4. Uesu Y, Kurimura S, Yamamoto Y (1995) Appl Phys Lett 66:2165
5. Gauderon R, Lukins PB, Sheppard CJR (1998) Opt Lett 23:1209
6. Sun CK, Chu SW, Tai SP, Keller S, Abare A, Mishra UK, DenBaars SP (2001) Scanning 23:182
7. Chen IH, Chu SW, Bresson F, Tien MC, Shi JW, Sun CK (2003) Opt Lett 28:1338
8. Fine F, Hansen WP (1971) Appl Opt 10:2350
9. Freund I, Deutsch M, Sprecher A (1986) Biophys J 50:693
10. Huang JY, Chen Z, Lewis A (1989) J Phys Chem 93:3314
11. Bouevitch O, Lewis A, Pinevsky I, Wuskell JP, Loew LM (1994) Biophys J 65:672

12. Campagnola PJ, Wei MD, Lewis A, Loew LM (1999) Biophys J 77:3341
13. Peleg G, Lewis A, Linial M, Loew LM (1999) Proc Natl Acad Sci USA 96:6700
14. Chu SW, Chen IS, Liu TM, Sun CK, Lee SP, Lin BL, Cheng PC, Kuo MX, Lin DJ, Liu HL (2002) J Microsc 208:190
15. Campagnola PJ, Millard AC, Terasaki M, Hoppe PE, Malone CJ, Mohler WA (2002) Biophy J 81:493
16. Tsang TYF (1995) Phys Rev A 52:4116
17. Barad Y, Eisenberg H, Horowitz M, Silberberg Y (1997) Appl Phys Lett 70:922
18. Squier JA, Müller M (1999) Appl Opt 38:5789
19. Yelin D, Silberberg Y, Barad Y, Patel JS (1999) Appl Phys Lett 74:3107
20. Sun CK, Chu SW, Tai SP, Keller S, Mishra UK, DenBaars SP (2000) Appl Phys Lett 77:2331
21. Guo Y, Ho PP, Tirksliunas A, Liu F, Alfano RR (1996) Appl Opt 35:6810
22. Müller M, Squier J, Wilson KR, Brakenhoff GJ (1998) J Microsc 191:266
23. Squier JA, Müller M, Brakenhoff GJ, Wilson KR (1998) Opt Express 3:315
24. Yelin D, Silberberg Y (1999) Opt Express 5:169
25. Chu SW, Chen IH, Liu TM, Cheng PC, Sun CK, Lin BL (2001) Opt Lett 26:1909
26. Sun CK, Chu SW, Chen SY, Tsai TH, Liu TM, Lin CY, Tsai HJ (2004) J Struct Biol147:19
27. Chu SW, Chen SY, Tsai TH, Liu TM, Lin CY, Tsai HJ, Sun CK (2003) Opt Express 11:3093
28. Canioni L, Rivet S, Sarger L, Barille R, Vacher P, Viosin P (2001) Opt Lett 26:515
29. Yelin D, Oron D, Thiberge S, Moses E, Silberberg Y (2003) Opt Express 11:1385
30. Chu SW, Chen SY, Tsai TH, Cheng CC, Liu TM, Sun CK, Wu SB, Biring S, Wang JK, Wang YL (2003) Conference on Lasers and Electro-Optics, Baltimore, MD, USA
31. Haus HA (1984) Waves and fields in optoelectronics, 1st edn. Prentice-Hall, Englewood Cliffs, New Jersey
32. Boyd R (1992) Nonlinear optics, 1st edn. Academic Press, New York
33. Guo Y, Ho PP, Savage H, Harris D, Sacks P, Schantz S, Liu F, Zhadin N, Alfano RR (1997) Opt Lett 22:1323
34. Moreaux L, Sandre O, Mertz J (2000) J Opt Soc Am B 17:1685
35. Guo Y, Savage HE, Liu F, Schantz SP, Ho PP, Alfano RR (1999) Proc Natl Acad Sci USA 96:10854
36. Brown E, McKee T, di Tomaso E, Pluen A, Seed B, Boucher Y, Jain RK (2003) Nat Med 9:796
37. Moreaux L, Sandre O, Blanchard-Desce M, Mertz J (2000) Opt Lett 25:320
38. Zoumi A, Yeh A, Tromberg BJ (2002) Proc Natl Acad Sci USA 99:11014
39. Chu SW, Liu M, Sun CK, Lin CY, Tsai HJ (2003) Opt Express 11:933
40. Yelin D, Oron D, Korkotian E, Segal M, Silberberg Y (2002) Appl Phys B 74:S97
41. Millard AC, Wiseman PW, Fittinghoff DN, Wilson KR, Squire JA, Müller M (1999) Appl Opt 38:7393
42. Sun CK, Chen CC, Chu SW, Tsai TH, Chen YC, Lin BL (2003) Opt Lett 28:2488
43. Anderson RR, Parish JA (1981) J Invest Dermat 77:13
44. Cheng PC, Pan SJ, Shih A, Kim KS, Liou WS, Park MS (1998) J Microsc 189:199
45. Schmitt JM, Knuttel A, Yadlowsky M, Eckhaus MA (1994) Phys Med Biol 39:1705
46. Bouma BE, Tearney GJ, Bilinsky IP, Golubovic B, Fujimoto JG (1996) Opt Lett 21:1839
47. Chen IH, Chu SW, Sun CK, Cheng PC, Lin BL (2002) Opt Quant Electron 34:1251
48. König K, So PTC, Mantulin WW, Gratton E (1997) Opt Lett 22:135
49. Tirlapur UK, König K, Peuckert C, Krieg R, Halbhuber JJ (2001) Exp Cell Res 263:88
50. Cheng PC, Lin BL, Kao FJ, Gu M, Xu MG, Gan X, Huang MK, Wang YS (2001) Micron 32:661
51. Squirell JM, Wokosin DL, White JG, Bavister BD (1999) Nat Biotechnol 17:763
52. Thomann I, Bartels A, Corwin KL, Newbury NR, Hollberg L, Diddams SA, Nicholsom JW, Yan MF (2003) Opt Lett 28:1368

53. Sun CK, Chu SW, Liu TM (2002) OSA Annual Meeting, Orlando, FL
54. Clays K, Van Elshocht S, Persoons A (2000) Opt Lett 25:1391
55. Clays K, Van Elshocht S, Chi M, Lepoudre E, Persoons A (2001) J Opt Soc Am B 18:1474
56. Zhao T, Chen ZH, Chen F, Shi WS, Lu HB, Yang GZ (1999) Phys Rev B 60:1697
57. Dumeige Y, Vidakovic P, Sauvage S, Sagnes I, Levenson JA, Sibilia C, Centini M, D'Aguanno G, Scalora M (2001) Appl Phys Lett 78:3021
58. Berger V (1998) Phys Rev Lett 81:4136
59. Broderick NGR, Ross GW, Offerhaus HL, Richardson DJ, Hanna DC (2000) Phys Rev Lett 84:4345
60. Gallant DJ, Bouchet B, Baldwin PM (1997) Carbohydr Polym 32:177
61. Ross MH, Romrell LJ (1989) Histology: a text and atlas. Williams and Wilkins, Baltimore, USA
62. Cassimeris L, Inoue S, Salmon ED (1988) Cell Motil Cytoskel 10:185
63. Kao FJ, Wang YS, Huang MK, Huang SL, Cheng PC (2000) SPIE Proc 4082:119
64. Lin CJ, Kao FJ (2003) Focus on microscopy. Genoa, Italy
65. Chu SW, Chen SY, Chern GW, Tsai TH, Szu-Yu Chen, Chen YC, Lin BL, Sun CK (2004) Biophys J 86:3914
66. Butcher PN, Cotter D (1990) The elements of nonlinear optics. Cambridge University Press, London
67. Stoller PK, Reiser M, Celliers PM, Rubenchik AM (2002) Biophys J 82:3330
68. Verbiest T, Elshocht SV, Kauranen M, Hellemans L, Snauwaert J, Nuckolls C, Katz TJ, Persoons A (1998) Science 282:913
69. Kim BM, Eichler J, Reiser KM, Rubenchik AM, Da Silva LB (2000) Laser Surg Med 27:329
70. Vogel A, Noack J, Hüttmann G, Paltauf G (2002) Proc SPIE 4633A:1
71. Kobayashi M, Fujita K, Kaneko T, Takamatsu T, Nakamura O, Kawata S (2002) Opt Lett 27:1324
72. Weingarten KJ, Rodwell MJW, Bloom DM (1988) IEEE J Quantum Electron 24:198
73. Lee CH, Chang RK, Bloembergen N (1967) Phys Rev Lett 18:167
74. Ohlhoff C, Meyer C, Lupke G, Loffler T, Pfeifer T, Roskos HG, Kurz H (1996) Appl Phys Lett 68:1699
75. Dadap JI, Hu XF, Anderson MH, Downer MC, Lowell JK, Aktsipetrov OA (1996) Phys Rev B 53:7607
76. de Jong W, Etteger AF, van't Hof CA, van Hall PJ, Rasing T (1995) Surf Sci 331/333:1372
77. Saha SK, Wong GK (1979) Appl Phys Lett 34:423
78. Barnik MI, Blinov LM, Dorozhkin AM, Shtykov NM (1983) Mol Cryst Liq Cryst 98:1
79. Povazay B, Bizheva K, Unterhuber A, Hermann B, Sattmann H, Fercher AF, Drexler W, Apolonski A, Wadsworth WJ, Knight JC, Russell P, Vetterlein M, Scherzer E (2002) Opt Lett 27:1800
80. So PTC, Kim H, Kochevar IE (1998) Opt Express 3:339
81. Masters BR, So PTC (2001) Opt Express 8:2
82. Chen IH (2002) Master Thesis, National Taiwan University

Adv Biochem Engin/Biotechnol (2005) 95: 57–75
DOI 10.1007/b102210

Live Cell Spinning Disk Microscopy

Ralph Gräf[1] (✉) · Jens Rietdorf[2] · Timo Zimmermann[2]

[1] A.-Butenandt-Institut/Zellbiologie, Ludwig-Maximilians-Universität München, Schillerstrasse 42, 80336 München, Germany
rgraef@lrz.uni-muenchen.de
[2] Advanced Light Microscopy Facility, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany

Abstract In vivo microscopy of dynamic processes in cells and organisms requires very fast and sensitive acquisition methods. Confocal laser scanning microscopy is inherently speed-limited by the requirement of beam scanning movements. In contrast to single beam scanning systems, the parallelized approach of multi-beam scanning is much faster. Spinning disk confocal microscopes are therefore very suited for fast in vivo imaging. The principles of spinning disk microscopy will be explained in this chapter and a thorough comparison of the performance of single beam and multi-beam scanning systems is made and illustrated with an example of in vivo imaging in *Dictyostelium discoideum*.

Keywords Spinning disk microscopy · In-vivo imaging · Confocal microscopy · Real-time imaging

List of Abbreviations

AOTF	Acousto-optical tunable filter
AU	Airy disc unit
CCD	Charge Coupled Device
CLSM	Confocal laser scanning microscope
CSU	Confocal scanning unit
FRAP	Fluorescence Recovery after Photobleaching
GFP	Green Fluorescent Protein
MBCM	Multi-Beam Confocal Microscopy
NA	Numerical aperture
PMT	Photomultiplier Tube
SBCM	Single Beam Confocal Microscopy

1 Introduction

In biological imaging, confocal laser scanning microscopy (CLSM) has in the last decade significantly extended our ability to visualize highly complex samples as multidimensional datasets (space, time, colors). In parallel, the introduction of fluorescent protein variants as in vivo tags of structures of interest has opened up new ways to observe cellular processes inside the living cell or tissue (for review see Miyawaki et al., this issue).

In this chapter we will present and discuss a confocal microscopy variant that is very well suited to in vivo imaging.

The most common type of confocal microscope uses a single focused laser beam to sequentially point-scan a region (single beam confocal microscope, referred to as SBCM in the following text). The fluorescence created by the passage of this focused beam through the sample is sent through a narrow aperture in the intermediate image plane (the confocal pinhole) onto a detector and is thus reduced to the photons coming from the plane of focus of the objective, but not from regions above or below it (Fig. 1A). By this rejection of out-of-focus contributions an optical section is created containing only the information from the focal plane. This basic operational principle as it was already realized for Marvin Minsky's prototype in 1955 is used in most commercially available confocal microscopes today.

Long before confocal microscopes became a standard imaging tool in biology, however, another more parallelized approach to confocal imaging was developed using a technique significantly predating most electronic imaging inventions: In 1884 Paul Nipkow created a device that transmitted images electrically. It was the first television camera and made use of a rotating disk with a spiral pattern of holes that broke down two-dimensional information into a sequential series of signals that could be reconstituted into an image using a complementary disk with the same pattern.

In 1968, M. Petrán and his collaborators applied the Nipkow disk principle to develop a tandem scanning reflected light microscope in which the single

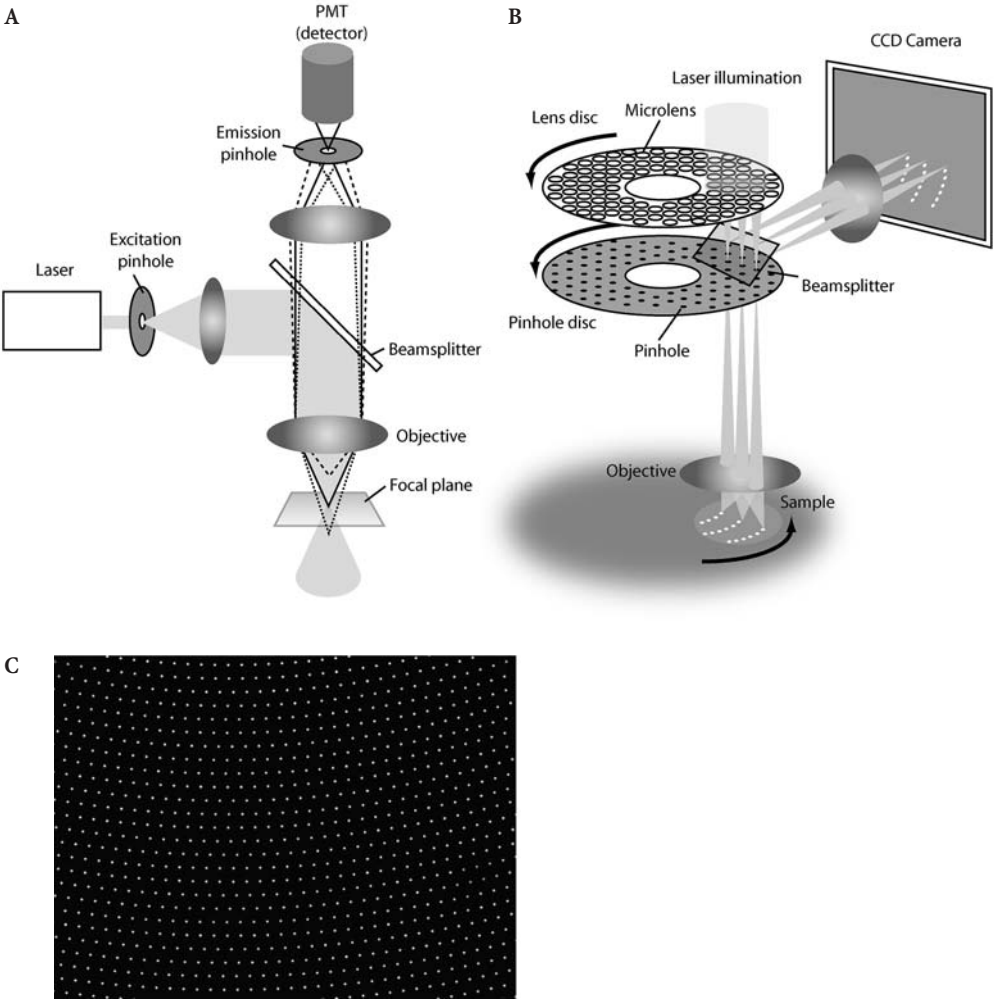


Fig. 1A–C Operating principles of single and multi-beam scanning confocal microscopes: **A** schematic drawing of a single beam scanning confocal microscope; **B** schematic drawing of a multi-beam scanning confocal microscope (Yokogawa CSU 10); **C** the constant pitch helical pinhole pattern of the Yokogawa spinning disk in the image field. During rotation of the disk, the pinholes evenly sweep the whole field of view

beam scanning confocal approach was parallelized to utilize multiple beams and corresponding pinholes [1]. Although this approach overcame the severe speed disadvantage of the single beam scanning method, it had significant problems of its own. For fluorescence imaging, the technology suffered from little excitation light reaching the sample due to the limited pinhole area (approximately 1%). Additional drawbacks were the requirement of high precision in the pinhole placement for designs with opposing excitation and emission

pinholes or problems with the scattered excitation light inside the detection system for setups using the same pinhole for excitation and emission [2].

When confocal microscopes became more widely available due to improvements in computer and imaging technology, the favored approach was single beam scanning. Recently, a significantly improved disk design by Yokogawa Inc. as well as progressive improvements in camera design have re-established the alternative multi-beam scanning technique (multi-beam confocal microscope, referred to as MBCM in the following text), especially for the requirements of in-vivo imaging. In combination with the Green Fluorescent Protein (GFP) technology applied in live-cell imaging, fast, multi-beam scanning microscopy is now a powerful tool for cell biology.

It is our aim in this chapter to explain the operational principle of multi-beam (tandem) scanning microscopy, to distinguish it from single beam scanning approaches, to compare their possibilities, to present commercially available designs and to demonstrate suitable applications.

2 Principle of Operation

The principle of operation of spinning disk microscopes will mainly be explained referring to a confocal scanning unit (CSU) designed by Yokogawa Corp [2]. It represents a very modern variant of the basic concept and in its design some of the initial limitations of the tandem scanning technology (little excitation light, uneven illumination, backscattering) were addressed and overcome. It is incorporated into several commercially available spinning disk setups (Perkin Elmer, VisiTeC).

A spinning disk confocal microscope consists of a rotating disk with multiple pinholes and a CCD camera (Fig. 1B). The pinholes on the disk (Fig. 1B) are arranged in a pattern that allows every location of an image to be covered when the disk is rotated.

In spinning disk microscopes an even field of illumination is created (e.g. by widening laser illumination into a circular field) that irradiates a section of the spinning disk (Fig. 1B). While most light does not pass the disk, the light going through the pinholes forms a set of minibeams corresponding to the pinhole pattern and sweeps the image field because of the disk rotation. Every minibeam in itself is confocal, with the same aperture serving as the excitation as well as the emission pinhole for a single mini-beam (Fig. 1C, [2, 3]). Designs using opposing pinholes on the disk (Petran et al. 1968) are not realized in the current instruments.

In the Yokogawa design a constant pitch helical pinhole pattern is rotated in order to provide even and stripe-free illumination of the entire field of view. Of the 20,000 pinholes on the disk, approximately 1000 cover the field of view of the microscope at any time (Fig. 1C). Due to the pinhole arrangement, every single image position is covered with 1/12th rotation of the disk. As the disk

rotates with 1800 rpm (i.e. 30 rps) this amounts to 30×12=360 full frames that are acquired per second [2].

To avoid crosstalk between the spots of individual minibeams, the pinholes are spaced significantly apart [3]. Accordingly, only a small area of the disk is covered by pinholes (1–4%) and most of the excitation light does not reach the sample because it is blocked by the disk. In the Yokogawa design this problem is overcome by a second disk in front of the pinhole disk (Fig. 1C). It contains microlenses arranged in the same pattern as the pinholes. These collect the excitation light and focus it into the pinholes thereby significantly increasing the excitation light throughput from approx. 1% to 40–60% [2].

A confocal image is formed almost instantaneously and can be directly viewed through the eyepiece of the scanhead.

In SBCMs, the emitted light is detected by photomultiplier tubes (PMTs) that read out intensities over time. Spatial information is not perceived. The image is reconstituted by plotting the intensity value of a certain time-point to the corresponding x-y-position of the scanning beam. MBCMs use a two-dimensional detector (a CCD camera) to record the intensity and spatial position of all minibeams simultaneously (Fig. 1B). The frame rate is defined by the camera exposure time and the frame readout speed of the camera. Maximally it can go up to 1/12 of the rotation frequency of the spinning disk (i.e. 360 fps in case of the CSU-10).

The characteristic differences between SBCMs and MBCMs consist in (1) serial against a parallelized scanning approach and (2) the mode of detection (PMTs vs CCD camera). All further differences between SBCMs and MBCMs result from these two initial factors.

3
Comparison of Single Beam and Multi-beam Scanning Confocal Imaging

The following section will discuss how the differences in the images taken on both systems arise.

3.1
Image Acquisition Rate

In a SBCM the image acquisition rate is limited by the speed of the scan mirrors. The fastest scanning units currently available are operating at resonance frequency, thereby achieving 512-line-frame rates close to video rate or more than 100 Hz for reduced 32-line frame formats.

In a MBCM the image acquisition rate is limited by the speed of the camera readout. In the CSU-10 version of the Yokogawa scanhead the rotation frequency of the disc was limiting the acquisition rate to 360 Hz, which is no longer the case for the CSU-21 that can rotate with higher speeds.

As soon as light intensity becomes the limiting factor which is clearly the case for live specimen fluorescence imaging applications, images with similar

quality can be acquired much faster on a MBCM, due to its two- to fivefold overall higher efficiency (see below) and about 1000-fold higher integration time per pixel (pixel dwell time) allowing for higher light throughput per time unit without saturation of the fluorophore (see below).

To increase the image acquisition rate while keeping the light flux through the sample constant there are the principal possibilities to either sacrifice spatial resolving power or reduce the observed volume. Using the zoom function and changing the pinhole diameter the size of voxels and thereby resolving power of the system is easily and flexibly adjustable in the SBCM while in the MBCM the resolving power can only be changed stepwise by changing the objective, using an optovar or reading out the camera in binning mode (see also "Optical sectioning and confocality" below). In addition the SBCM can scan rather complexly shaped regions. Taken together, optimizing the observation volume is easier on a SBCM. This advantage however does not make the SBCM the superior tool for fast image acquisition except for applications where line regions are to be observed as for example in scanning a single line along an axon to measure the propagation of action potentials.

3.2 Efficiency, Photobleaching and Phototoxicity

According to theoretical estimates the overall light loss on optical components along the optical path of a filter based SBCM is expected to be similar to the loss inside the CSU10 based MBCMs (calculations according to [4], based on specifications in [5]), thus, differences in efficiency mainly rely on the different sensitivities of the detectors, which is in good agreement with our own findings and those of others [2].

The quantum efficiency of current CCD cameras (up to 75%) is significantly higher than that of PMTs (maximum 25% in most commercially available systems). Moreover, with the latest development of 'on chip amplifying' CCDs that achieve a strong amplification of the incoming signal while preserving a good spatial resolution, appropriate confocal acquisition of very fast biological processes inside living tissue has become feasible. While the higher efficiency in the detection of photons will reduce photobleaching and phototoxicity by making better use of the emitted photons, another advantage of MBCMs over SBCMs for live specimen imaging results from the higher efficiency in excitation of the fluorophores by the MBCM.

In single beam scanning, a high intensity laser beam passes over the sample and illuminates every region intensely, but only for a very short period of time (typically 2–3 µs). Only the light put into the sample and read out from the sample during this time is available to transport the information for the generation of the image. As a result, the excitation light has to be intense in order to excite enough fluorophores during this short time. In multi-beam scanning, the excitation light is split into many mini-beams of correspondingly lower intensity. However, several of these beams pass over the same region sequentially and the emis-

sion of all their passes is collected during the exposure time of the camera exposure. The illumination time per pixel is therefore significantly (1000-fold) longer.

We have analyzed the effects of these very different types of illumination by extending model calculations made by [6] for single beam scanning to a multiple beam situation(see Appendix at the end of the chapter).

We find that in the MBCMs significantly less excitation light is needed for the generation of an equivalent image due to two reasons: (1) the detection system is more efficient (see above); (2) the fluorophore saturation level is low. In SBCMs excitation light levels are close to the saturation of the fluorophore, a situation where the number of photons emitted by a fluorophore no longer increases proportionally with increasing excitation light intensity is easily reached, while in MBCMs the excitation light level is typically far from saturation of the fluorophore. The reduction of excitation light will at the same time also avoid potential photodamage by photochemical reactions independent of the fluorophore label (e.g. flavins) and in our opinion constitutes an additional important advantage for MBCM over SBCMs in the observation of living samples.

3.3 Multichannel Imaging

For many biological applications, the imaging of more than one fluorescence channel is required. This can be done on spinning disk systems by sequentially acquiring the different image channels onto the CCD and changing the excitation light and/or the emission filter between the channels. While switching between different excitation light lasers employing acousto-optical tunable filters (AOTFs) is fast (2 ms), changing emission filters involves filter wheels and is therefore ten times slower. Sequential multichannel acquisition involves a reduction in the image acquisition rate which can become a severe limitation for colocalisation studies of moving structures. Fast moving structures in the image might move between the acquisition of the channels and thus lead to mismatches between image channels.

Parallel acquisition is possible on MBCMs but requires modifications to the detection system. In such modified setups, the emitted fluorescence is split into longer and shorter wavelength components by a beamsplitter and either projected onto separate CCD cameras or onto different regions of the same CCD chip (DualView Beamsplitter). The recent introduction of sensitive color CCD cameras (like the Hamamatsu 3CCD camera) also has further potential for fast multichannel imaging (T. Nagai, personal communication) without significant alterations to the detection system.

Due to problems of significant crosstalk and cross-excitation between fluorophores (see review by Zimmermann in this issue), multichannel imaging is very often performed sequentially even on potentially parallel acquisition systems like SBCMs. By switching excitations line by line, SBCMs possess a sequential acquisition mode that avoids mismatches between channels, which is not available for MBCMs. For most applications, however, the subsecond

frame rates for multichannel imaging achievable on spinning disk confocals are adequate.

3.4 Regional Bleaching

Fluorescence Recovery after Photobleaching (FRAP) (for a review, see Houtsmuller in this issue) and photoactivation [7] or photoconversion of fluorescent proteins (see Miyawaki et al. in this issue) are powerful techniques to investigate protein dynamics inside living cells. These techniques require a bleaching or activation step, i.e. a short irradiation of a defined image sub-region with intense laser light and is easily performed on SBCMs. Although some spinning disk systems also use laser light for excitation, the laser cannot be used for spot or region bleaching in the existing setups. It is widened for the illumination of the whole field of view and cannot be specifically positioned within the image. Regional bleaching on spinning disk systems could however be performed with an additional positionable laser. This would require customization of the system, but it would be a formidable application considering the excellent fast time-lapse capacity of spinning disk systems.

3.5 Optical Sectioning and Confocality

The thickness of an optical section in a SBCM microscope can be varied by adjusting the diameter of the detection pinhole, while the pinhole diameter and therefore sectioning depth is fixed in MBCMs. The depth discrimination of the system is determined by the numerical aperture (NA) and magnification of the objective lens. For the CSU-10 from Yokogawa Inc. pinhole diameter and spacing are calculated to match a 1.4 NA 100× lens for green excitation light and yield a 500 nm diameter projection of the pinhole in the object plane which is about 1 Airy Unit (AU) for an excitation wavelength of 550 nm and an optical section of about 0.8 μm will be generated.

To achieve other section depths using the CSU-10, the investigator is limited by the available objective lenses. Using a 1.4 NA 60× lens the projection of the invariant pinhole into the object plane will be larger than 1 AU resulting in a larger z-depth of 1.25 μm. Using 1.3 NA 40× the section is about 1.8 μm. (For the Olympus DSU described below, discs with different optical properties are available to match the respective objective lenses and thereby higher flexibility is generated). Contributions of out-of-focus fluorescence into neighboring pinholes is to be expected in multi-beam systems depending on the fluorescence intensity and distribution inside the sample, thereby compromising their confocality [2]. Theoretical depth discrimination capability is therefore better in single beam setups. It should however be kept in mind that in order to achieve a high resolution excellent contrast is required which on a single beam confocal is usually difficult to achieve during live specimen imaging [8].

Under low signal to noise conditions the multi-beam confocal may achieve even better resolution due to its higher detection efficiency (see above).

3.6 Conclusion

Spinning disk microscopes are well suited for in vivo imaging and under certain conditions perform better than SBCMs for several reasons: (1) high frame rates can be achieved; (2) the use of CCD detectors provides high detection efficiencies; and (3) low intensity excitation light is used. Some of these advantages over SBCMs vanish if multichannel measurements are to be performed. Sequential acquisition compromises the advantage of fast acquisition and generates problems with respect to co-localisation studies while attempts to simultaneous acquisition using several cameras or beamsplitters are either tricky to align and expensive or severely compromise the light budget if emission beamsplitters or color cameras are employed. None of the commercially available variants of spinning disk microscopes are equipped for regional bleaching as required for photobleaching and photoactivation measurements. The overall resolution achieved of MBCMs is expected to be similar or better to that of SBCMs under low S/N condition as in live cell imaging due to higher detection efficiency. The main differences of the two system types are summarized in Table 1.

Table 1 Overview of single and multi beam confocal microscope features for in-vivo imaging applications

Microscope type	Single beam confocal microscope	Multi beam confocal microscope
Acquisition speed	Limited frame rate +	High frame rate +++
Detection efficiency	Photomultiplier tube +	CCD camera ++
Bleaching rates for equivalent images	Higher, more emission photons required +	Lower, less emission photons required ++
Multichannel imaging	Simultaneous detection, sequential detection ++	Sequential detection +
Region specific bleaching	Possible ++	Not possible (w/o additional hardware) –
Optical sectioning, axial resolution	Adjustable pinhole diameter ++	Fixed pinhole diameter, possible crosstalk between neighboring pinholes +

4
Available Designs for Video-Rate Confocal Microscopy

In this section a short overview over commercially available multi-beam scanning systems is given. In addition, other existing confocal designs that also provide high frame rates are briefly mentioned.

4.1
Multi-beam Designs

4.1.1
Laser-Based Spinning Disk Confocals

Since it is a self-contained solution, Yokogawa confocal scanning units (CSU) with the double disk design are employed in several commercial solutions (e.g. Yokogawa, Perkin Elmer, VisiTec, Solamere). In addition to the CSU-10, the more recent CSU-21 with higher disk rotation speeds and switchable filters is also available. The systems are operated with a variety of lasers and a CCD camera. For switching between fluorescence channels, solutions with filterwheels as well as with AOTFs exist. For simultaneous two-channel imaging, setups with two cameras are also available.

4.1.2
White Light Based Spinning Disk Confocals

Two commercially available designs use Xenon or Mercury arc lamps for excitation instead of laser light. Because of the broad emission spectrum of the light source, the excitation light is controlled and easily varied by the use of excitation filters. This flexibility is not available with lasers as these are limited to specific excitation lines. They can be easily added onto an existing widefield system at relatively low cost and quick switches between widefield and confocal imaging are possible with them. They do provide images with improved z-resolution, but the optical sectioning capability is inferior to single beam scanning confocals due to restrictions in the light budget that are compensated by a decrease in confocality (see above).

The CARV module (ATTO Bioscience) contains a Nipkow disk with pinholes and can be added to a variety of widefield microscopes. The Olympus Disk Scanning Unit (DSU) can be fitted to Olympus microscopes and uses a rotating disk with a pattern of slits instead of pinholes. Its optical sectioning performance can be altered by the insertion of disks with different slit widths.

4.1.3
Two-Photon Multi-Beam Scanning

A very different concept of multi-beam scanning is realized in Multifocal Multiphoton laser scanning Microscopy (TriM Scope, LaVision BioTec). In this

design a beam multiplexer is used to create an array of scanning foci (array can be single or rectangular/quadratic shape) that is used for parallel scanning at high rates. Functionally it is very different in that it uses multiphoton excitation to create its scanning foci and does not involve pinholes on the detection side. A similar multiphoton approach is provided by [9]. Time-multiplexed multifocal multiphoton microscopy overcomes the general problem of multifocal microscopy, crosstalk between single foci [10].

4.2 Video-Rate Confocal Microscopes of Other Designs

In addition to multi beam scanning, other solutions for fast confocal imaging have been introduced over the years. With the increasing flexibility and speed of single beam scanning confocals many of them however have become obsolete and have gone out of manufacture. Slit scanning confocals [11] use a stripe of excitation light instead of a single point. As this stripe has to be moved across the sample in only one direction, scanning time per frame is significantly reduced so that images can be acquired at high frame rates with a CCD camera. The x-y resolution is however asymmetrical and the confocality is also compromised. Some existing commercial solutions (Meridian InSight, BioRad ViewScan DVC 250) are not manufactured anymore. Recently, Carl Zeiss AG has introduced the LSM 5 LIVE based on the slit scanning concept. Its distinguishing feature is a CCD line detector that allows very fast readout. Because of this, much higher frame rates at high scanning resolutions can be achieved than would be possible with a conventional CCD camera.

Fast scanning confocal microscopes increase the frame speed to video-rate by accelerated scanning of a single point. For this, either a resonant scanner or an acousto-optical deflector (AOD) is required [12]. The widely used Noran Odyssey XL employs an AOD for the fast scanning. Due to the AOD properties, the emission light cannot be completely descanned and has to pass through a slit instead of a pinhole aperture, thereby affecting confocality. The model is not produced anymore, but VisiTech International has recently introduced the VTeye based on the same concept. As a third way, Leica incorporated the resonant scanner approach into its existing range of confocal microscopes to offer a fast, fully confocal spectral setup (TCS SP2 AOBS RS). The limitation here is only in the zoom-in properties, as zooming can only be done centred due to the resonant scanner properties. As the imaging properties are like for any single beam scanning confocal, the comparison to spinning disk designs is as described above, but with higher frame rates gained at the cost of less detected photons.

5
Single and Multi-Beam Confocal Imaging in the Analysis of Centrosome Dynamics in *Dictyostelium*

The analysis of centrosome dynamics in *Dictyostelium discoideum* by four-dimensional imaging is a good example for an imaging task where three-dimensional movements of motile, small, fluorescent particles such as centrosomes have to be recorded over time.

The centrosome is a nucleus-associated organelle that serves as the main microtubule-organizing center in animal, plant and fungal cells. Its functions include mitotic spindle organization, cytokinesis, cell cycle progression, organelle positioning and the directionality of cell migration [13]. The centrosome duplicates once per cell cycle and the two centrosomal entities are separated during mitotic spindle formation where they are forming the spindle poles. The control of centrosome number, i.e. one per nucleus, seems to be essential for dividing cells to maintain their euploid state, since centrosome amplification leading to supernumerary centrosomes is a hallmark of tumor cells [14–16]. *Dictyostelium* amoebae are a good model system for centrosome research [17]. Thus, we have created a *Dictyostelium* mutant characterized by green fluorescent supernumerary centrosomes and a weak cytokinesis defect [18, 19]. This was achieved by overexpression of DdCP224, a member of the XMAP215-family of microtubule-associated proteins. The chromosomes in these mutants were visualized by co-expression of GFP-tagged histone2B. These mutants were used to investigate the dynamics of supernumerary and nucleus-associated centrosomes during mitosis [19]. The microscopic setup had to meet several requirements. The small size of *Dictyostelium* centrosomes (diameter of approximately 0.5 μm) requires the good spatial resolution of a confocal microscope. Their motility demands good temporal resolution and rapid acquisition of z-stacks over time, since the green-fluorescent centrosomes tend to move out of the confocal plane. Furthermore, *Dictyostelium* cells are very sensitive to phototoxic effects, especially during mitosis and if they are exposed to blue light required for GFP-excitation. Therefore, this was a suitable imaging problem to analyze differences in performance of spinning disk confocal microscopy (Perkin-Elmer-Wallac Ultraview) and single beam scanning confocal microscopy (Zeiss LSM510 META).

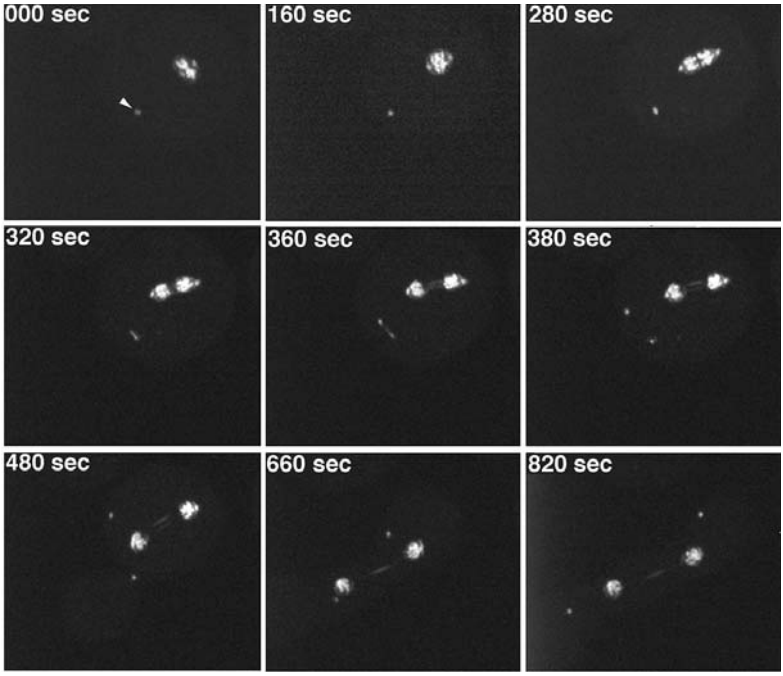
Since mitotic progression was inhibited by illumination with too high light intensities, long before any bleaching effect of GFP was detectable, the prevention of phototoxic effects turned out to be the most important task, independent of the imaging system used. Therefore, low laser powers had to be used which implied different consequences. In the case of spinning disk microscopy, relatively long exposure times of 500 ms and 2×2 binning had to be used resulting in a relatively low frame rate of approximately 1.3 frames/s and reduced pixel resolution. In the case of single beam laser scanning microscopy, low fluorescence signal intensities at a given laser setting require either high PMT voltages, which increases image noise, or a larger pinhole diameter, which de-

creases resolution. The frame rate depends on scanning options, such as the window size (in pixels), scan speed, bi- or unidirectional scanning, averaging of accumulated scans or the use of a line step (e.g. scanning of every second line with interpolation of the missing ones). To reach a rate of 1.3 frames/s a relatively small window size of 512×256 was chosen and the scan speed was set to the maximum. A line-step factor of 2 compensated the time needed to average two scans of each line, which was necessary to reduce noise.

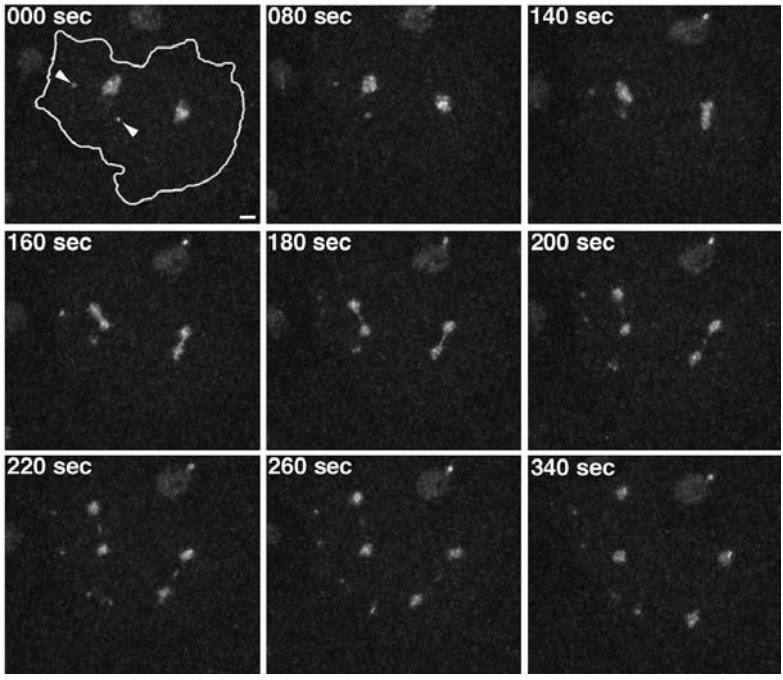
It was not possible to compare the two confocal systems with the same laser power settings since both systems employ different lasers and filter sets. Therefore, settings were chosen which allowed viewing of the cells for at least 30 min without indications for phototoxic effects.

Both types of scan-heads were mounted on an inverted microscope and, therefore, glass-bottom petri dishes were used to view the cells. After the cells had settled on the glass surface (~5 min), the medium was replaced by phosphate buffer and the cells were flattened by agar overlay [20]. This reduced the high mobility of centrosomes in the z-direction and the number of z-levels required to follow the centrosome. The time delay between each stack was 10 s. Under these conditions, the *Dictyostelium* mutants were viewed at 21 °C for up to 30 min.

The image series shown in Fig. 2 represent maximum intensity projections of image stacks acquired with each of the two confocal systems. Both sequences show mitotic progression of a cell with GFP-labeled spindle poles and supernumerary centrosomes (arrowheads at time-point "0 s" in Fig. 2) as well as GFP-labeled chromosomes. Figure 2A shows a mononucleated cell with one supernumerary centrosome and Fig. 2B a bi-nucleated cell with two supernumerary centrosomes. Imaging started in early (Fig. 2A) and late (Fig. 2B) metaphase, respectively, with a bipolar mitotic spindle and chromosomes arranged in a metaphase plate. The supernumerary centrosomes are not involved in spindle dynamics but duplicate in anaphase (time-point "320 s" in Fig. 2A and "180 s" in Fig. 2B). Both sequences illustrate essentially the same process, but the image sequence obtained by spinning disk microscopy reveals more details regarding the discernibility of the chromosomes and the duplication event of the supernumerary centrosomes, where a tiny spindle is visible between the two daughter centrosomes (Fig. 2A, time-point "360 s"). By contrast, the images obtained by single beam laser scanning microscopy are noisier and less detailed (Fig. 3). The chromosomes, for instance, mostly appear as one mass instead of individual entities. This is due to the chosen compromise between speed, sensitivity and resolution. Improvement of any of the factors would have to be done at the cost of another. For example, taking out the line step to improve the y-resolution would have to be compensated by deactivating the line-averaging to maintain the same frame rate, thus leading to even more noise in the image. However, the scientific question, if and when supernumerary centrosomes duplicate, is solved equally well by both time-lapse sequences, although the spinning disk movie is of higher quality.


000 sec
160 sec
280 sec
320 sec
360 sec
380 sec
480 sec
660 sec
820 sec

A


000 sec
080 sec
140 sec
160 sec
180 sec
200 sec
220 sec
260 sec
340 sec

B

Fig. 2A, B Comparison of the duplication of supernumerary centrosomes as acquired on a spinning disk confocal microscope (Perkin-Elmer-Wallac Ultraview) and a single beam laser scanning confocal microscope (Zeiss LSM510 META): **A** spinning disk confocal (MBCM). Live observation of a mitotic DdCP224-GFP/GFP-histone2B cell with initially one supernumerary centrosome (*arrowhead in the first image*). Image acquisition started in late metaphase after duplication of the nucleus-associated centrosome whose daughter centrosomes are located at the spindle poles. The time is indicated in seconds. Each image represents a maximum intensity z-projection of five confocal slices. Settings: z-distance, 0.5 μm each; exposure time, 500 ms; frame rate, 1.3 fr/s; binning, 2×2; **B** single beam confocal (SBCM). Live observation of a mitotic DdCP224-GFP/GFP-histone2B cell (cell edges are outlined in the first image) with initially two supernumerary centrosomes (*arrowheads in the first image*). Image acquisition started in late metaphase. The time is indicated in seconds. Each image represents a maximum intensity z-projection of five confocal slices. Settings: z-distance, 1 μm each; 512×256 pixels; line-step, 2; averaging, twofold; pinhole size, 2.8 airy disk units; scan direction, unidirectional; scan speed, maximum; frame rate, 1.3 fr/s; *scale bars* 2 μm

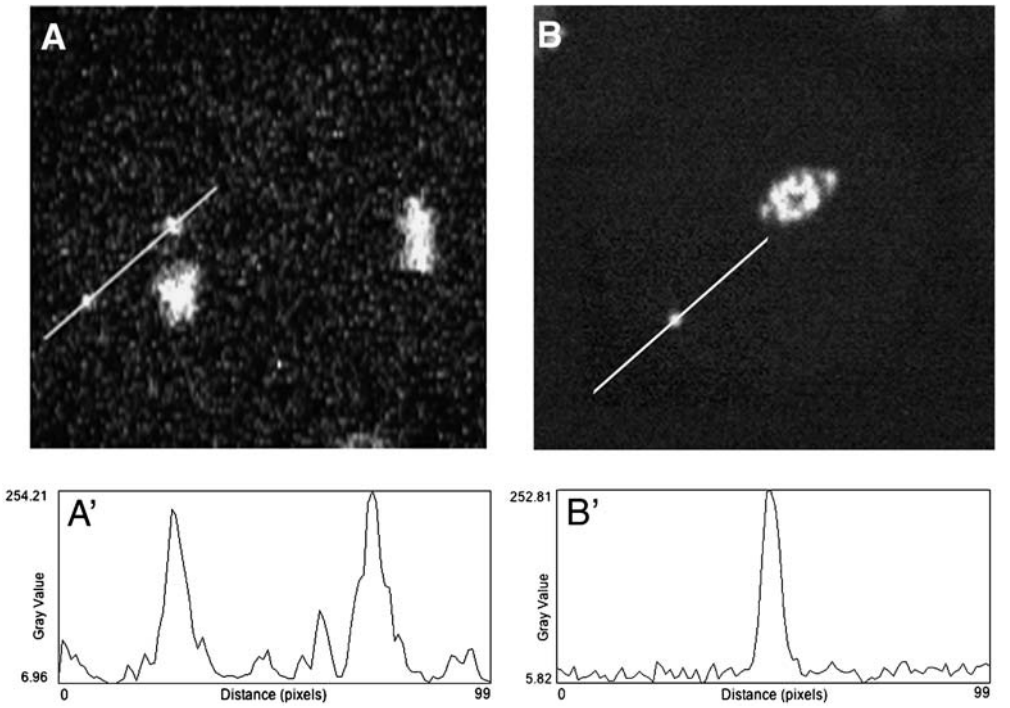


Fig. 3A, B Comparison of signal to noise ratio between laser scanning confocal microscopy and spinning disk confocal microscopy. Single slices derived from the image sequences of Fig. 2 are shown in **A** and **B**. The respective plots of fluorescence intensity (*displayed as gray levels of an 8-bit gray scale*) along a straight line through the center of supernumerary centrosomes (A′, B′) reveal superior signal to noise ratio of spinning disc microscopy vs single beam scanning microscopy

In cases of stronger signals, the frame rate (by shorter exposures) or the resolution (by reduction of binning) can be improved accordingly for spinning disk microscopes. At the single beam scanning microscope the frame rate used above is already close to the limit for such imaging tasks since the scan speed was set already close to the maximum. It can only be further improved by bidirectional scanning or omission of frame or line averaging. However, all these steps will reduce image quality. All improvements of the image quality, such as slower scanning, averaging of more scans or relinquishment of the line-step factor will significantly reduce the frame rate.

References

1. Petran M, Hadravsky M, Egger MD, Galambos R (1968) J Opt Soc Am 58:661
2. Inoue S, Inoue T (2002) Methods Cell Biol 70:87
3. Kino GS (1995) In: Pawley JB (ed) Handbook of biological confocal microscopy. Plenum Press, New York, p 155
4. Stelzer EHK (1995) In: Pawley JB (ed) Handbook of biological confocal microscopy. Plenum Press, New York, p 139
5. Kawamura S, Negishi H, Ostuki S, Tomosada N (2002) Yokogawa Tech Rep Engl Ed 33:17
6. Tsien RY, Waggoner A (1995) In: Pawley JB (ed) Handbook of biological confocal microscopy. Plenum Press, New York, p 267
7. Patterson GH, Lippincott-Schwartz J (2002) Science 297:1873
8. Stelzer EHK (1998) J Microsc 189:15
9. Bewersdorf J, Pick R, Hell SW (1998) Opt Lett 23:655
10. Egner A, Andresen V, Hell SW (2002) J Microsc 206:24
11. Amos WB, White JG (1995) In: Pawley JB (ed) Handbook of biological confocal microscopy. Plenum Press, New York, p 403
12. Tsien RY, Bacskai BJ (1995) In: Pawley JB (ed) Handbook of biological confocal microscopy. Plenum Press, New York, p 459
13. Rieder CL, Faruki S, Khodjakov A (2001) Trends Cell Biol 11:413
14. Fisk HA, Mattison CP, Winey M (2002) Curr Opin Cell Biol 14:700
15. Nigg EA (2002) Nat Rev Cancer 2:815
16. Doxsey S (2002) Mol Cell 10:439
17. Gräf R, Brusis N, Daunderer C, Euteneuer U, Hestermann A, Schliwa M, Ueda M (2000) Curr Topics Dev Biol 49:161
18. Gräf R, Daunderer C, Schliwa M (2000) J Cell Sci 113:1747
19. Gräf R, Euteneuer U, Ho TH, Rehberg M (2003) Mol Biol Cell 14:4067
20. Fukui Y, Yumura S, Yumura TK (1987) In: Spudich JA (ed) Methods in cell biology, vol 28. Academic Press, Orlando, FL, p 347
21. Dittrich PS, Schwille P (2001) Appl Phys B 73:829

Appendix

For comparing single and multi-beam scanning we follow a model calculation made by [6] for single beam scanning of fluorescein at 1 mW intensity at 488 nm and extend it for multi-beam scanning.

The following values are being used:

- Laser power at 488 nm: 1 mW
- Objective NA: 1.25
- Peak excitation intensity in a Gaussian spot I=1.25×10^{24} photons/cm^{2}s
- Fluorescein optical cross section cs=3.06×10^{-16} cm^2/molecule
- Fluorescein lifetime rate constant k_f=2.2×10^8 s^{-1}
- Fluorescein quantum efficiency Q_e=0.9
- Fluorescein bleaching efficiency Q_b=3×10^{-5}

The photon emission rate of a fluorophore is determined by the intensity dependent excitation rate, the lifetime of the fluorophore and its quantum efficiency. The excitation rate and the fluorophore lifetime allow the calculation of the saturation of the fluorophore. The photon emission rate as well as bleaching rate is directly proportional to it.

$$\text{Excitation rate constant } k_a = cs \times 1 \tag{1}$$

$$\text{Saturation } S = \frac{k_a}{k_a + k_f} \tag{2}$$

$$\text{Emission rate } k_{em} = Q_e k_f S \tag{3}$$

$$\textit{Bleaching rate } k_{em} = Q_b k_f S \tag{4}$$

Total Sample Irradiation Per Frame

How much light goes into a sample for a single exposure in the two setups? For this, an identical laser source and identical frame rates are assumed for both systems. In single beam scanning, the beam moves with finite speed over all regions of the sample. The total irradiation R of the sample is thus the product of the intensity of the laser beam I and the total acquisition time t for a frame.

$$R = I \times t \tag{5}$$

The comparison to a multi-beam system is actually quite simple: With an identical laser, the same intensity I is emitted for the same acquisition time t. The difference in photons actually hitting the sample is just determined by the transmission efficiency E of the spinning disk.

$$R = I \times t \times E \tag{6}$$

For identical frame rates and laser powers, the difference of total irradiation between a single beam confocal and a Yokogawa scan head would thus be a factor of approx. 2 (efficiency E 0.4–0.6).

Fluorescence emission

With this difference of total irradiation how much fluorescence is actually collected? This involves two aspects, the amount of emission photons and the efficiency of detection. In single beam scanning at 1 mW excitation intensity, fluorescein molecules in the region of peak intensity are 63% saturated, i.e. only 37% are not excited at any given time according to 1 and 2. The fluorescence emission rate k_{em} amounts to 1.26×10^8 photons/s according to 3.

In multi-beam scanning with a Yokogawa disk, the intensity of a single mini-beam is significantly reduced and can be calculated as

$$I_{multi\_beam} = \frac{I}{n_{pinholes}} \times E \tag{7}$$

For n=1000 pinholes and a transmission efficiency E=0.5, the intensity of a minibeam would amount to 1/2000th of the single beam intensity.

The saturation level of fluorescein in such a mini-beam is 0.09% according to 1, 2 and 7. The fluorescence emission rate is 1.72×10^5 photons/s according to 3.

Whereas the difference in excitation intensity between single and multi-beam scanning is 1:2000, the difference in emission is only 1:731. This means significantly higher excitation efficiency at low intensities. This is an effect of the decreasing numbers of fluorophores available in the unexcited state at high saturation levels.

Fluorescence Emission Per Frame

With the fluorescence emission rates k_{em} of single and multi-beam scanning setups known, how many photons are actually emitted by a fluorophore in a single frame?

For single beam scanning, the amount of emitted photons per fluorophore per frame N_{em} is

$$N_{em} = k_{em} \times t \tag{8}$$

where t is the pixel dwell time of the beam. Using the calculation example and a pixel dwell time of 2.5 µsec this amounts to 300 photons emitted by a single fluorophore per frame.

For a spinning disk, the calculation can not be performed with pixel dwell times and the number of beam passes, since these parameters (radial velocity, number of passing pinholes) vary over the image field according to the radial position along the disk. The product of beam dwell times and beam passes is however a constant over the whole field (even illumination).

To derive a value equivalent to the single beam pixel dwell time for a spinning disk microscope, the total irradiation of the image field per frame has to be divided by the number of image pixels to get a representative illumination

value for each pixel under even illumination. The division of this value by the intensity of a mini-beam provides the total time of illumination centered on a pixel and thus an equivalent to the single beam pixel dwell time.

Accordingly, for multi-beam scanning the amount of emitted photons per fluorophore per frame N_{em} is

$$N_{em} = k_{em} \times \frac{R}{n \times I_{multi\_beam}} \tag{9}$$

where R is the total illumination of the sample per frame according to 6 and n is the number of image pixels.

Using this calculation, approximately 420 photons are emitted by a fluorophore under multi-beam conditions per frame.

This is an idealized calculation assuming identical areas of illumination for the single and multi-beam setups. In reality, only 63% of the circularly widened excitation beam goes into the rectangular (1.3:1.0) detection area of the CCD, so that less excitation light is actually used in the multi-beam system than for the calculation. Taking this into account, approximately equal amounts of fluorescence photons can be expected for single and multi-beam systems.

Bleaching Rates

The photobleaching of fluorophores by illumination is still a poorly understood process [6]. Assuming linear bleaching kinetics even at the high intensities present in single beam scanning, as indicated by fluorescence correlation spectroscopy data [21], equal amounts of photons emitted per frame mean that the bleaching rate is more or less equal in multi-beam scanning and in single beam scanning microscopy.

Adv Biochem Engin/Biotechnol (2005) 95: 77–106
DOI 10.1007/b102211

Total Internal Reflection Fluorescence Microscopy in Single Molecule Nanobioscience

Tetsuichi Wazawa[1] · Masahiro Ueda[2] (✉)

[1] Laboratory of Nanodevices (endowed by OMRON Co.), Graduate School of Frontier Biosciences, Osaka University, 1–3 Yamadaoka, Suita, Osaka 565-0871, Japan
wazawa@fbs.osaka-u.ac.jp
[2] PRESTO, JST and Soft Biosystem Group, Laboratories for Nanobiology, Graduate School of Frontier Biosciences, Osaka University, 1–3 Yamadaoka, Suita, Osaka 565-0871, Japan
ueda@phys1.med.osaka-u.ac.jp

Abstract Recent development in total internal reflection fluorescence microscopy (TIRFM) has made it possible to directly monitor the behaviors of biomolecules at the single-molecule level both in vitro and in living cells. This technique has opened a new era in bioscience, so-called single-molecule nanobioscience. In this review, we will summarize the theoretical basis of TIRFM, explain the instruments of TIRFM for single-molecules imaging, and discuss some novel applications of TIRFM to single-molecule nanobioscience.

Keywords TIRF · Evanescent · Single molecule · Nanobioscience · *Dictyostelium*

List of Abbreviations

cAMP	Cyclic adenosine 3′5′-monophosphate
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
FRET	Fluorescence resonance energy transfer
GFP	Green fluorescent protein
GPCR	G-protein-coupled receptors
NA	Numerical aperture
S/N	Signal to noise
TIRF	Total internal reflection fluorescence
TIRFM	Total internal reflection fluorescence microscopy
TMR	Tetramethylrhodamine
YFP	Yellow fluorescent protein

1 Introduction

This review is intended as a guide outlining the applications of total internal reflection fluorescence microscopy (TIRFM) to single-molecule nanobioscience. TIRFM was originally used for biological imaging of multi-molecules and organelles in living cells (reviewed by Axelrod [1–3]). Recently, TIRFM has been further refined for single-molecule imaging of biomolecules in vitro and also in living cells, and currently it is one of the most popular forms of microscopy used for this purpose. TIRFM has made it possible to follow biochemical reactions, structural changes and movements of biomolecules at the single molecule level in real time, and the results give new insights on molecular mechanisms of biomolecules [4–10].

TIRFM can be used to observe fluorophores attached to biomolecules and living cells in aqueous solution near a glass surface. The illumination method utilized for the excitation of fluorophores in TIRFM is conceptually simple. When the excitation light for fluorophores is incident above some "critical angle" upon the glass/water interface, the light is totally internally reflected and generates a thin electromagnetic field in the water with the same wavelength as the incident light. This field is called the "evanescent field" and it can excite fluorophores very near the glass surface. Because the intensity of the evanescent field decays exponentially with the distance from the glass surface, fluorophores further from the surface are not excited, thereby reducing the background noise dramatically. As noted in Funatsu et al. (1995), background noise is the major problem to be overcome during single-molecule imaging in an aqueous environment [11]. Noise is mainly derived from Raman scattering of water molecules and out of focus fluorophores, and noise is estimated to be more than 10 times higher than the fluorescence from a single fluorophore under conventional microscopes. The critical reduction of background noise was achieved by limiting the illumination of excitation light to very near the glass surface. TIRF was ideal for this purpose because it provides a means of selectively exciting

fluorophores near the glass surface. The theoretical and technical aspects of TIRFM will be dealt with in the following sections.

Single-molecule imaging technique has already been successfully applied to a wide range of biological systems. For example, this technique has been used to measure enzymatic activity of myosin, a molecular motor responsible for force generation in muscle contraction [11, 12]. Furthermore, a single fluorescently-labeled kinesin molecule, another molecular motor, has been observed moving along a microtubule using TIRFM [13]. These represent the first time a single, working molecule has been visualized. Single-molecule fluorescence spectroscopy has also been achieved using TIRFM, providing a powerful tool for elucidating conformational dynamics of proteins [14, 15]. In addition to soluble proteins, membrane proteins such as ion channels incorporated into artificial lipid membrane have been imaged with TIRFM [16, 17]. More recently, TIRFM has been applied in cell biology to visualize signaling molecules in living cells at the single-molecule level [18, 19]. Thus, the development of single-molecule imaging techniques with TIRFM has opened a new era in biology, so-called single-molecule nanobioscience [9]. The novel applications of TIRFM will be discussed later and some of the limitations on single molecule fluorescence measurements will be described.

2 Theory: Evanescent Waves by Total Internal Reflection

In this section, the theoretical aspects of the evanescent waves will be outlined, especially the amplitude and intensity of the waves. Because the evanescent wave is discussed considerable detail by Axelrod [1] and Fornel [20], our treatment of this topic will be brief. Let us consider a planar light wave on *x*-*z* plane penetrating into medium 1 (Fig. 1). When the light beam of infinite width strikes

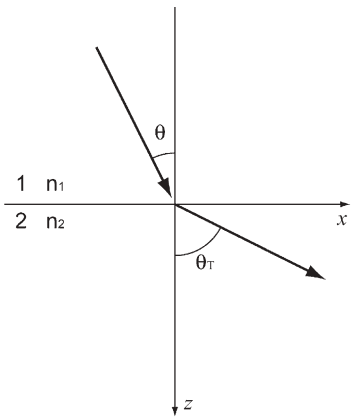


Fig. 1 The geometry of incidence and transmission

the interface between the two media (1 and 2), the beam is partially transmitted into medium 2 in an angle in a refractive index dependent manner. We denote refractive indices as n_1 and n_2 for media 1 and 2, respectively, where $n_1 > n_2$. Also we denote the incidence and refraction angles measured from the norm (z axis) as θ and θ_T, respectively. The phase speeds (v_1 and v_2) for the light beams in the media are $v_1 = c/n_1$ and $v_2 = c/n_2$, respectively, where c is the speed of light in a vacuum. To propagate into medium 2, the transmitted light is required to have the same phase at the interface anywhere and anytime (Fig. 2a). In other words, when the incidence light attains a length of $\sin\theta$ at the speed of v_1, the transmitted light should reach a length of $\sin\theta_T$ at a speed of v_2. This can be described by Snell's law,

$$\frac{\sin\theta}{v_1} = \frac{\sin\theta_T}{v_2} \tag{1}$$

and

$$n_1 \sin\theta = n_2 \sin\theta_T . \tag{2}$$

If the incidence angle is equal to the critical angle, $\theta_c = \sin^{-1}(n_2/n_1)$, the transmitted wave goes parallel to the interface because the transmitted wave has the same phase as the incidence beam only when the refraction angle is 90° (Fig. 2b). When the incidence angle θ is greater than $\theta_c = \sin^{-1}(n_2/n_1)$, the refraction angle θ_T cannot be found, thus, the wave only penetrates transiently into a limited space, and hence, total internal reflection occurs (Fig. 2c). In this condition, the electromagnetic wave is permitted to penetrate into only a limited depth as explained by Maxwell's equations. By solving Maxwell's equations, the electric field of the transmitted wave, $\mathbf{E}_T$, can be written as

$$\mathbf{E}_T = \mathbf{A}_T \exp\left[i\left\{ \left(\frac{2\pi}{\lambda_2} \right) (x \sin\theta_T + z \cos\theta_T) - \omega_T t \right\} \right], \tag{3}$$

where $\mathbf{A}_T$, λ_2, ω_T, z, and t are the amplitude of the electric field, the wavelength in medium 2, the angular frequency in medium 2, the perpendicular distance from the interface and time, respectively. By Snell's law (Eq. 2), we have

$$\cos\theta_T = \sqrt{1 - \sin^2\theta_T} = \pm i \sqrt{\frac{\sin^2\theta}{n^2} - 1} , \tag{4}$$

where $i = (-1)^{1/2}$ and $n = n_2/n_1$ ($= \sin\theta_c$). Using Eqs. (3) and (4), it follows that

$$\mathbf{E}_T = \mathbf{A}_T \exp\left[i\left\{ \frac{2\pi \sin\theta_T}{\lambda_2} x - \omega_T t \right\} \right] \exp\left[\mp z \frac{2\pi}{\lambda_2} \sqrt{\frac{\sin^2\theta}{n^2} - 1} \right]. \tag{5}$$

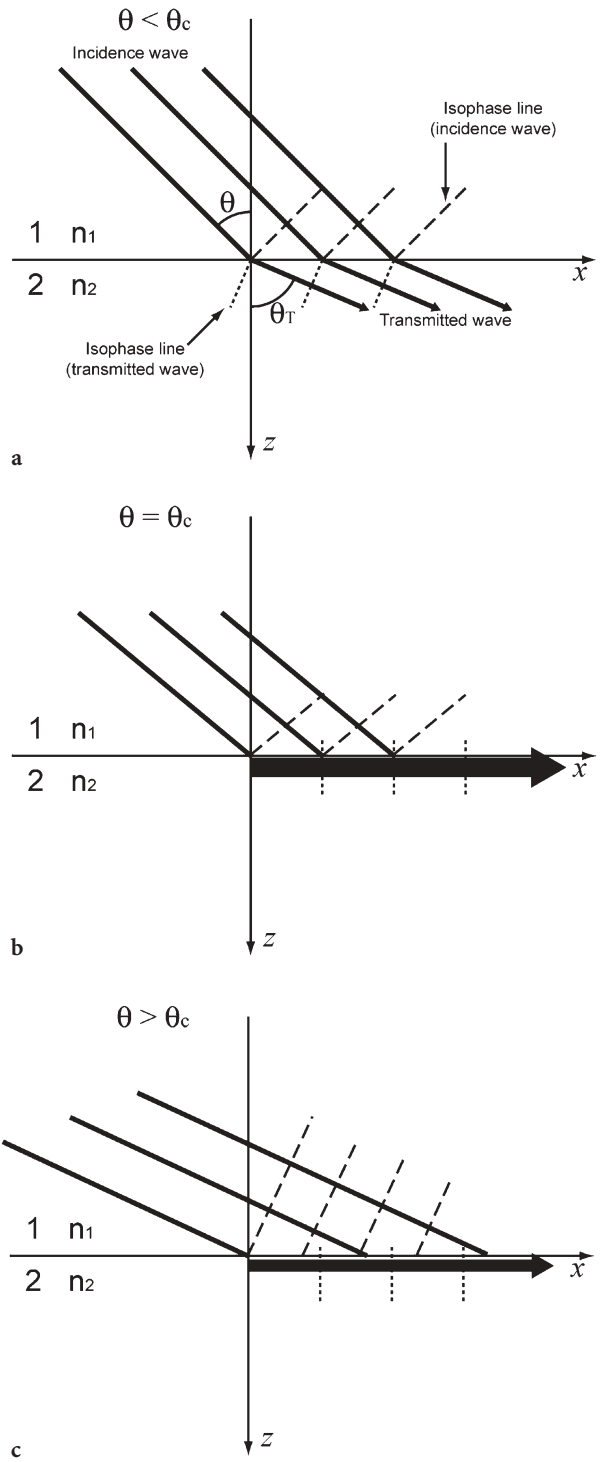


Fig. 2a–c Incidence-angle dependency of phases of incidence and transmitted waves

The sign of the argument of the second exponential must be negative, because the amplitude of the electric field would diverge as z approached ∞ if it were positive. Similarly, the sign in Eq. (4) must be positive. Equation (5) indicates that the transmitted wave goes towards the positive x direction. For the intensity of the wave I_T, we have

$$I_T = |\mathbf{E}_T|^2 = |\mathbf{A}_T|^2 \exp\left[-z\frac{4\pi}{\lambda_2}\sqrt{\frac{\sin^2\theta}{n^2} - 1}\right]. \tag{6}$$

By Eqs. (5) and (6), it is shown that the amplitude and intensity diminish exponentially with the perpendicular distance z from the interface between media 1 and 2. This type of wave is called as the evanescent wave. The penetration depth d, 1/e value in Eq. (6) is

$$d = \frac{\lambda_2}{4\pi\sqrt{\frac{\sin^2\theta}{n^2} - 1}}$$

or (7)

$$d = \frac{\lambda}{4\pi\sqrt{n_1^2\sin^2\theta - n_2^2}}$$

where $\lambda = n_2\lambda_2$, wavelength of the incident light in vacuum, The penetration depth d is typically in an order of wavelength except for $\theta \to \theta_c$. For example, as a typical TIRFM configuration, if we take medium 1 as fused silica (n_1=1.46), medium 2 as water (n_2=1.33), wavelength in vacuum, 532 nm, and the incidence angle, 70°, then d=126 nm. Thus, by using the evanescent wave, the fluorophore can be selectively illuminated very near to the silica/water interface in the water medium.

Next the components of $\mathbf{E}_T$ are examined. For a linearly-polarized incidence light beam, according to Fresnel's formula, the ratio of the amplitudes of electric field is given by

$$\frac{A_{Tp}}{A_{Ip}} = \frac{2\cos\theta}{\cos\theta_T + n\cos\theta}, \tag{8}$$

$$\frac{A_{Ts}}{A_{Is}} = \frac{2\cos\theta}{n\cos\theta_T + \cos\theta} \tag{9}$$

where A_{Ip} and A_{Tp} are the amplitudes in the electric field for p-polarized incidence and transmitted waves, and A_{Is} and A_{Ts} are those for s-polarized waves, respectively. Using Eqs. (4), (8) and (9), the components of electric field for the transmitted wave at z=0, denoted by E_{0x}, E_{0y} and E_{0z}, are given by

$$
\begin{aligned}
E_{0x} &= A_{\mathrm{Tp}}(-\cos\theta_T) \\
&= A_{\mathrm{Ip}}\frac{2\cos\theta}{\cos\theta_T + n\cos\theta}(-\cos\theta_T) \\
&= A_{\mathrm{Ip}}\frac{2\cos\theta\sqrt{\sin^2\theta - n^2}}{\sqrt{n^4\cos^2\theta + \sin^2\theta - n^2}}e^{-i(\varphi_{\mathrm{P}} + \pi/2)}, \\
E_{0y} &= A_{\mathrm{Ts}} \\
&= A_{\mathrm{Is}}\frac{2\cos\theta}{\sqrt{1 - n^2}}e^{-i\varphi_{\mathrm{s}}}, \\
E_{0z} &= A_{\mathrm{Tp}}(\sin\theta_{\mathrm{T}}) \\
&= A_{\mathrm{Ip}}\frac{2\cos\theta\sin\theta}{\sqrt{n^4\cos^2\theta + \sin^2\theta - n^2}}e^{-i\varphi_{\mathrm{P}}},
\end{aligned}
\tag{10}
$$

where

$$
\begin{aligned}
e^{-i\varphi_{\mathrm{P}}} &= \frac{n^2\cos\theta - i\sqrt{\sin^2\theta - n^2}}{\sqrt{n^4\cos^2\theta + \sin^2\theta - n^2}} \\
e^{-i\varphi_{\mathrm{s}}} &= \frac{\cos\theta - i\sqrt{\sin^2\theta - n^2}}{\sqrt{1 - n^2}}
\end{aligned}
\tag{11}
$$

and we used the relation such that $-i = e^{-i\pi/2}$. The electric fields decay exponentially with the distance from the interface (Eq. 5), thus the electric fields of p- and s-polarizations are given by

$$
\begin{aligned}
\mathbf{E}_{\mathrm{P}} &= (\mathbf{e}_x E_{0x} + \mathbf{e}_z E_{0z})\exp\left[-z\frac{2\pi}{\lambda_2}\sqrt{\frac{\sin^2\theta}{n^2} - 1}\right], \\
\mathbf{E}_{\mathrm{s}} &= (\mathbf{e}_y E_{0y}\exp\left[-z\frac{2\pi}{\lambda_2}\sqrt{\frac{\sin^2\theta}{n^2} - 1}\right].
\end{aligned}
\tag{12}
$$

where $\mathbf{e}_x$, $\mathbf{e}_y$, and $\mathbf{e}_z$ are unit vectors for x, y, z components, respectively. From Eq. (12) it is clear that the evanescent wave by the incidence of p-polarization light has both the x component parallel to the direction of propagation and the z component. However, the evanescent wave by the incidence of s-polarized light has only the y component, which is purely transverse to the direction of propagation.

From Eq. (10), each component of the intensity of the evanescent wave by the p- or s-polarized incidence light at z=0 can be written as

$$I_{0x} = |A_{lp}|^2 \frac{4\cos^2\theta(\sin^2\theta - n^2)}{n^4\cos^2\theta + \sin^2\theta - n^2},$$

$$I_{0y} = |A_{ls}|^2 \frac{4\cos^2\theta}{1 - n^2}, \tag{13}$$

$$I_{0z} = |A_{lp}|^2 \frac{4\cos^2\theta\sin^2\theta}{n^4\cos^2\theta + \sin^2\theta - n^2}.$$

Since the intensities of the evanescent field for the p- and s-polarized incidence light at z=0 are $I_{0p}=I_{0x}+I_{0z}$ and $I_{0s}=I_{0y}$, respectively, it follows that

$$I_{0p} = |A_{lp}|^2 \frac{4\cos^2\theta(\sin^2\theta - n^2)}{n^4\cos^2\theta + \sin^2\theta - n^2},$$

$$I_{0s} = |A_{ls}|^2 \frac{4\cos^2\theta}{1 - n^2}. \tag{14}$$

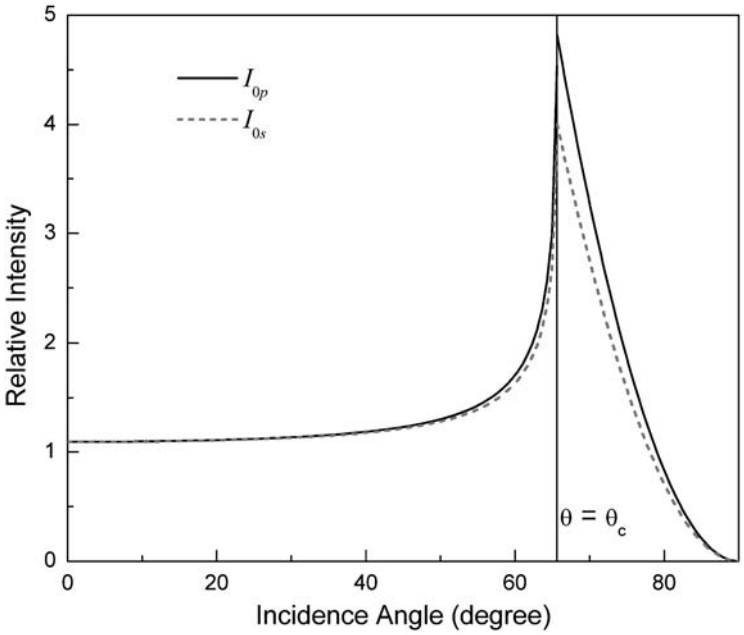


Fig. 3 Plot of the relative intensities of the evanescent waves as a function of the incidence angle for p- and s-polarization. In the calculation of the intensities, the media 1 and 2 are fused silica (n_1=1.46) and water (n_2=1.33), respectively, and the wavelength was 532 nm

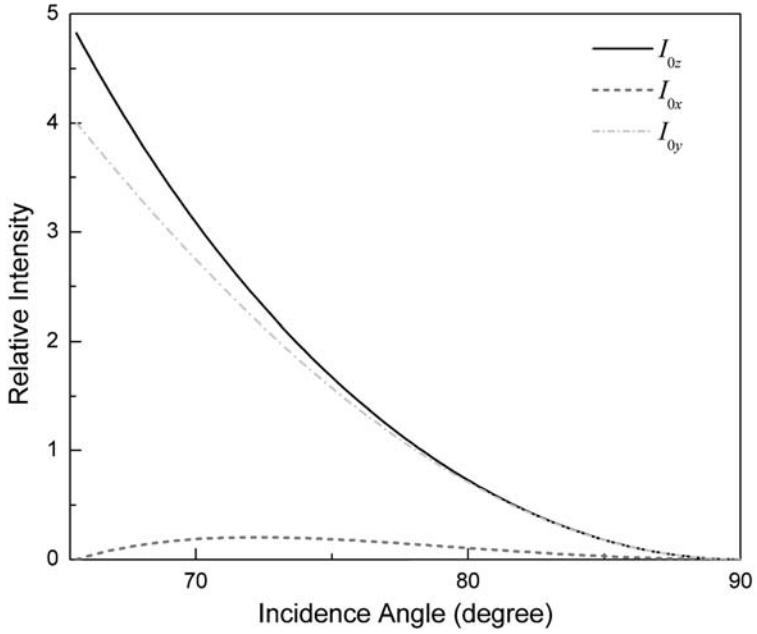


Fig. 4 Plot of the intensities of the evanescent waves for x and z components at p-polarization and y component at s-polarization

Figure 3 shows the intensities of the electric field as a function of incidence angle where media 1 and 2 are silica (n_1=1.46) and water (n_2=1.33), respectively. The intensities of both p- and s-polarizations increase with incidence angle for $\theta<\theta_c$, and they decrease and reach 0 at θ=90°. At $\theta=\theta_c$, they have maximal values, which are typically several times greater than that of the incidence light. When incidence angle is close to θ_c, the fluorescence of fluorophores excited by the transmitted light may become several times brighter owing to enhancement of the electric field. The x and z components of the intensities of the evanescent wave for the p-polarized incidence are shown in Fig. 4. The z components monotonously decreases and reaches 0 at 90° as a result of the cosine factor. The x component is 0 at $\theta=\theta_c$ and θ=90° as described by the terms, $\sin^2\theta-n^2$ and $\cos^2\theta$, respectively, and hence is bell-shaped. The intensity of x component is always lower than that of the z component because of the "$-n^2$" factor in the fraction.

3
Instruments of TIRFM

3.1
Prism-Type TIRFM

Here we will describe the optical configuration of prism-type TIRFM on an inverted microscope, designed especially for single molecule studies. In prism-type TIRFM, an illumination laser beam is introduced onto the glass/liquid interface using a prism from the topside of the microscope, whereas fluorescence from fluorophores is collected with a microscope objective located on the other side (Fig. 5). That is, the illumination optics to generate the evanescent wave is completely independent from the focusing optics of the microscope. Because there is considerable freedom in designing the spectral transmittance or reflectance of filters and dichroic mirrors of the microscope, so that autofluorescence arising from optical materials such as glasses and refractive index matching oils can be easily lowered and/or eliminated. This advantage allows us to set up single molecule microscopes with extremely-low background.

Figure 5 shows a schematic diagram of a TIR on an inverted microscope. The laser beam is allowed to pass through a neutral density filter, depolarizer, achromatic lens, prism and glass slide, and then the beam is totally internally reflected at the glass/water interface. Fluorescence from fluorophores excited by the evanescent wave is collected by an objective lens and allowed to pass through filters and focusing optics. The fluorescence is finally introduced into the measurement system such as video camera, spectrograph or photodiodes.

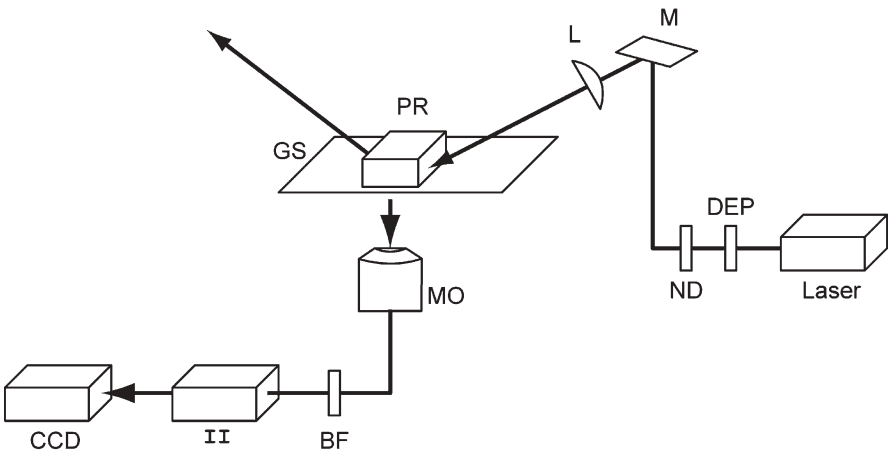


Fig. 5 Schematic diagram of a prism-type TIR setup on an inverted microscope: DEP, depolarizer; ND, neutral density filter; M, mirror; L, focusing lens; PR, prism; GS, glass slide; MO, microscope objective; BF, barrier filter; II, imager intensifier; CCD, charge-coupled device

The laser as a light source should be selected depending on fluorophores to be measured and the design of experiments. To perform single-molecule imaging of commonly used dyes such as Cy3, Cy5, and tetramethylrhodamine, a laser beam at a power of about 10 mW is incident on a specimen plane at an illuminated area of 250×110 μm^2 ($1/e^2$ value) [15]. Considering the loss of laser power due to the reflection at the surface of mirrors and lenses, it is desirable to choose a laser with several times higher power than needed for the incidence at the specimen plane. Water-cooled Ar and Ar-Kr gas lasers are often used as a multi-purpose light source because of their high power and the variety of different lasing wavelengths (e.g. Coherent, Innova 70C-Spectrum). In fact, they can be used to image fluorescent proteins (CFP, GFP, and YFP), Cy3, tetramethylrhodamine, Texas Red and some of Alexa and BODIPY dyes in single-molecule experiments. However, such water-cooled lasers are not suitable to use in nanometry experiments, because the mechanical vibration produced from the circulating water interferes with the measurement of pico-Newton forces and nanometer displacements [21]. Diode lasers and solid-state lasers are most commonly utilized as a light source as they are vibration-free, small in size and less expensive. Some of the most common lasing wavelengths are 405, 473, 488, 532, 635, and around 680 nm. In addition, He-Ne lasers at 632.8 nm are also used for the dyes Cy5, Alexa 633 and Alexa 647.

The light beam emitted from the laser is passed through a neutral density filter to modulate the laser power, and then a focusing lens (Fig. 6). The purpose of the focusing lens is to concentrate the laser beam with an appropriate area and power on the specimen (without the lens, a very high power output of laser is required to obtain sufficient power density at the specimen plane). For single-molecule imaging, the microscope magnification would be more than ×100, and the diameter of the laser beam itself would be much larger than the

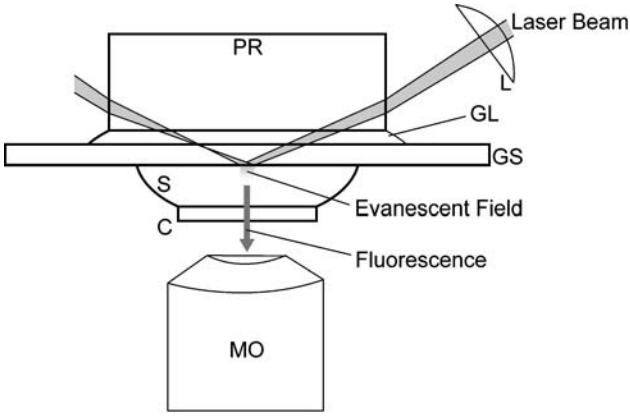


Fig. 6 Detailed view around a specimen of a prism-type TIR on an inverted microscope: L, focusing lends; PR, prism; GL, glycerol; S, aqueous solution; C, coverslip; MO, microscope objective

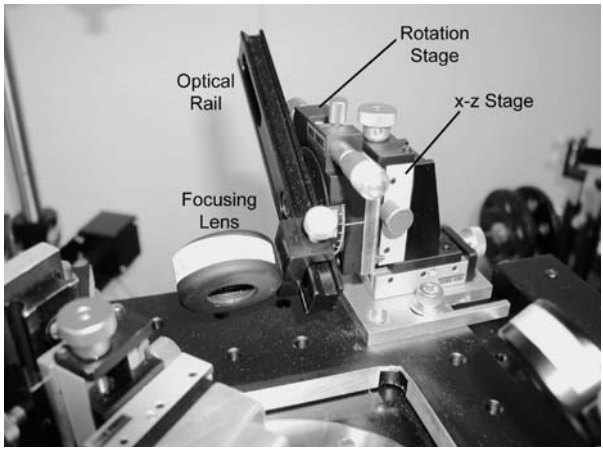


Fig. 7 Photograph of the mount mechanism of a focusing lens

size of the microscope field. The laser beam is not necessary to be strictly focused at the specimen plane, but it may be slightly defocused (Fig. 7). An achromatic convex lens is commonly used as the focusing lens. The lens can be mounted on a carrier on an optical rail that is assembled with a rotation and *x-z* stages (Fig. 7). The size of the illuminated area can be adjusted by the translocation of the focusing lens along the optical rail that is parallel to the path of the laser beam. A depolarizer is often inserted into the system to reduce fluorescence polarization effects, though isotropy of the electric field cannot be obtained with a single-beam illumination. The mirror (M in Fig. 5) is also an essential component of the system to position accurately the illuminated area. It should be noted that the holder of the mirror must be rigidly immobilized with vibration-free rods because vibration of the mirror holder will change directly the position of the illuminated area.

After passing through the focusing lens, the laser beam is incident on a prism and glass slide through a matching liquid between them, striking a glass/water interface, causing the beam to be totally internally reflected (Fig. 6). The prism is crucial because the laser beam is brought to an incidence angle at the glass/water interface greater than the critical angle. Although prisms are available in a variety of shapes, we recommend a cube prism rather than 45- or 60-degree prisms. The cube prism makes it easier to adjust the incidence angle and prevents the matching liquid from seeping up the slope of the prisms, which will disturb the refraction of the beam. For the purpose of single molecule fluorescence detection, one of preferable materials for the prism would be a synthetic fused silica with low-autofluorescence. Other materials such as BK7 may cause much higher autofluorescence arising from the transmission of very intense laser beam which would greatly increase the background light in the microscope fields. The prism can be mounted on a removable holder attached to a two-axis translator that moves vertically and laterally. The vertical move-

ment is important as it allows the matching liquid between the prism and the glass slide to spread when pressure is applied to the liquid layer. The gap between the prism and the glass slide should be completely filled with the matching liquid. As a matching liquid, we often use fluorescence-microscope-grade glycerol, although refractive index matching oil available commercially may also be applicable. The refractive index of glycerol is 1.47 which is very near that of fused silica so there is very little distortion of the laser beam and hardly any reflection at the interface. In addition, glycerol is less irritating and less harmful, thus it is suitable for its use in small closed spaces such as darkrooms. As for the glass slide, we usually use low-fluorescence fused silica as well as the prism. The glass slide directly affects the background of the microscope field. For example, significant autofluorescence was observed in some commercially borosilicate glass slides for emissions between 650 and 850 nm when an excitation wavelength of 457.9 nm was used (Fig. 8). When we tried to simultaneously image single Cy3 and Cy5 molecules immobilized on a borosilicate glass slide, single Cy3 molecules could be imaged but single Cy5 molecules could not, owing to its high background light. How clean the glass surface is also critically affects the background noise. In order to clean the glasses in moder-

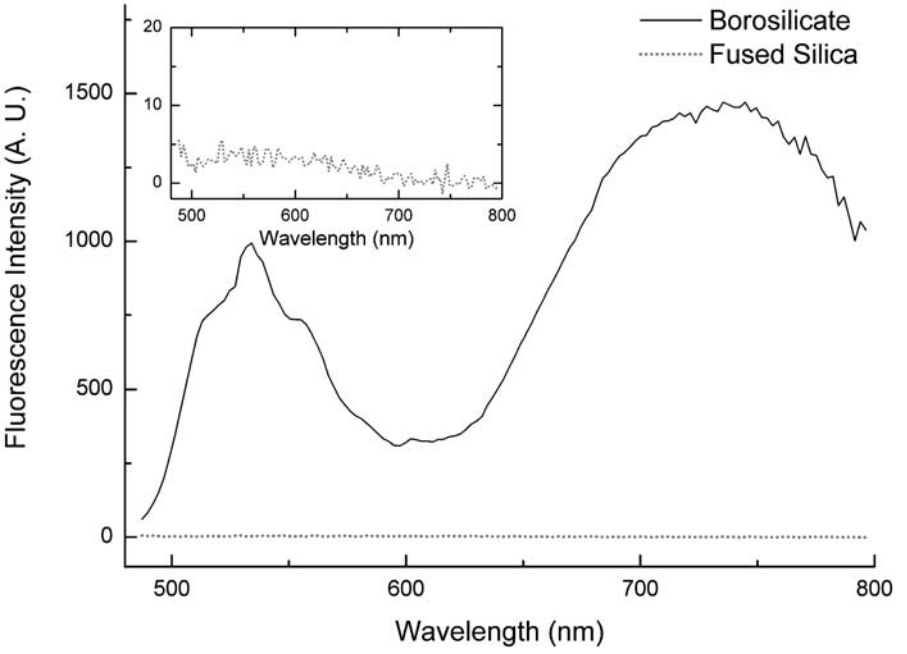


Fig. 8 Fluorescence emission spectra of commercially available borosilicate and fused silica glass slides. The spectra were measured with a TIRFM-based fluorescence spectroscope [15]. The borosilicate glass slide was obtained from Matsunami Glass (Osaka, Japan) and low-fluorescence grade fused silica from Hikari Kobo (Tokyo, Japan). Excitation wavelength was 457.9 nm

ate conditions, they can be ultrasonicated in laboratory-grade detergents, pure methanol, pure ethanol, pure acetone or 0.1 mol/l NaOH. In addition, the glasses may be heated to 400 °C for several hours. After the washing processes, the glasses should be exhaustively rinsed with and stored in ultra pure water.

The gap between the glass slide and coverslip is filled with aqueous solution, in which the sample may be dissolved or immersed. Since single molecule fluorescence imaging is a kind of extremely low-level light detection, contamination should be as minimized as possible. Hence, the aqueous solutions must be prepared with freshly-purified ultra pure water. The aqueous stock solutions such as the pH buffers and salines can be frozen until use in order to eliminate any increase in bacteria. The additives to the aqueous solutions can also influence the background of microscope field. As a pH buffer, we often use phosphate, HEPES, PIPES, citrate, and acetate.

Fluorescence from dye molecules excited by the evanescent wave propagates into the water medium, passes through the coverslip, and then goes into the microscope objective (Fig. 6). The numerical aperture of the microscope objective is typically high for single molecule imaging as it increases the ability of the objective to collect light. However, the high numerical aperture significantly limits the thickness of water medium between glass slide and coverslip. When we use an oil-immersion objective with a NA of 1.4, a permitted thickness for the medium is less than 50 μm. Moreover, the illuminated region in this optical configuration would not be optically optimal, because most of objective lens is designed to take an image at the plane just above the coverslip with the lowest aberration. Therefore, this results in aberration and the loss of the total number of detected photons [22]. In practice we use a high NA objective such as Olympus PlanApo 60× (NA1.45).

The light collected by the microscope objective is allowed to pass through a filter and focusing lens, and then strikes on the faceplate of a photo-detector. Band pass filters are commonly used as a barrier filter in the optical system. It is necessary to exclusively transmit the fluorescence of dyes to be observed and to eliminate any unnecessary background light such as Rayleigh and Raman scattering, randomly reflected and refracted excitation light and fluorescence from unknown contaminations. For single molecule fluorescence imaging, wide-transmission-range filters such as notch and long pass filters alone may cause unexpected errors. We should take the transmission range of a band pass filter between the excitation wavelength and the Raman scattering. The shorter boundary of the transmission is determined so that the blocking of the excitation light is guaranteed. The longer boundary would be limited by Raman scattering for water and the peak wavelength is calculated by

$$\frac{1}{\lambda_R} = \frac{1}{\lambda_E} - \delta_W \tag{15}$$

where δ_W=0.00033 nm^{-1} and λ_R and λ_E are the wavelengths of Raman scattering and excitation, respectively [23]. In practice, the transmission wavelength must

be tuned further, depending on fluorophores to be used and experiments to be performed. The excitation light should be rejected by reflection at the surface of the filter, since the transmission of a very intense laser beam through glass material could result in autofluorescence. Thus, the filter should be designed so that the excitation beam is rejected by the thin film coat on the filter, and therefore colored glass filters themselves alone may not be adequate for single molecule fluorescence imaging. Moreover, the substrate of the filters can be low-fluorescence grade fused silica, rather than BK7, to reduce autofluorescence more completely. Thus, the filtered light is passed through a series of focusing lens, and then strikes on the faceplate of a sensitive photodetector.

A prism-type TIRFM can be constructed on an upright microscope by replacing a condenser with a prism (Fig. 9). The laser beam is concentrated by a focusing lens, incident on the prism, and totally internally reflected at the glass/water interface. An advantage of this configuration is that petri dishes can be used to hold the specimens so that a water-immersion microscope objective is operational. Thus, the microscope objective can be positioned to produce optimal optics which minimizes aberration and the loss of photons detected, as long as the specimen is sufficiently thin. This type of configuration is advantageous for observing cells, although under these conditions, it is more difficult to align the laser beam.

A surface plasmon resonance at the interface between the solution and a thin metal coat on the glass can be used to generate an evanescent field, and the field generated is stronger than that produced by TIR [24].

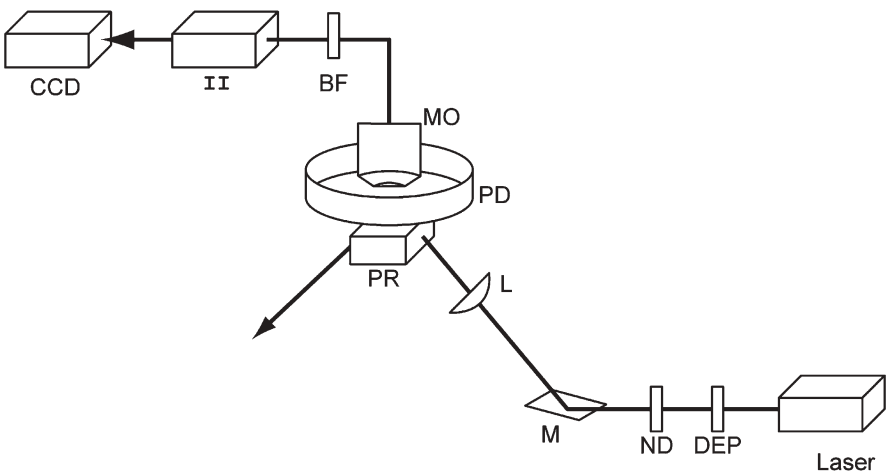


Fig. 9 Schematic diagram of a prism-type TIR setup based on an upright microscope: DEP, depolarizer; ND, neutral density filter; M, mirror; L, focusing lens; PR, prism; PD, petri dish; MO, microscope objective; BF, barrier filter; II, imager intensifier; CCD, charge-coupled device

3.2
Objective-Type TIRFM

Figure 10 illustrates schematic drawing of objective-type TIRFM. A high numerical aperture objective lens is mounted on an inverted microscope. A laser beam is passed through a neutral density filter (ND) and a beam expander (BE) to adjust its power and diameter. When the laser polarized linearly is used, the polarization of the laser beam is converted from linear to circular by a quarter-wave plate ($\lambda/4$). The laser beam is focused by a lens (L) on the back focal plane of the objective, so that specimens are illuminated uniformly with Koehler illumination. By shifting the position of the mirror (M) located between the lens (L) and dichroic mirror (DM), the path of the incident laser light is shifted from the center to the edge of the objective. At the center position, the microscope can be used as a standard epi-fluorescence microscope (Fig. 10b).

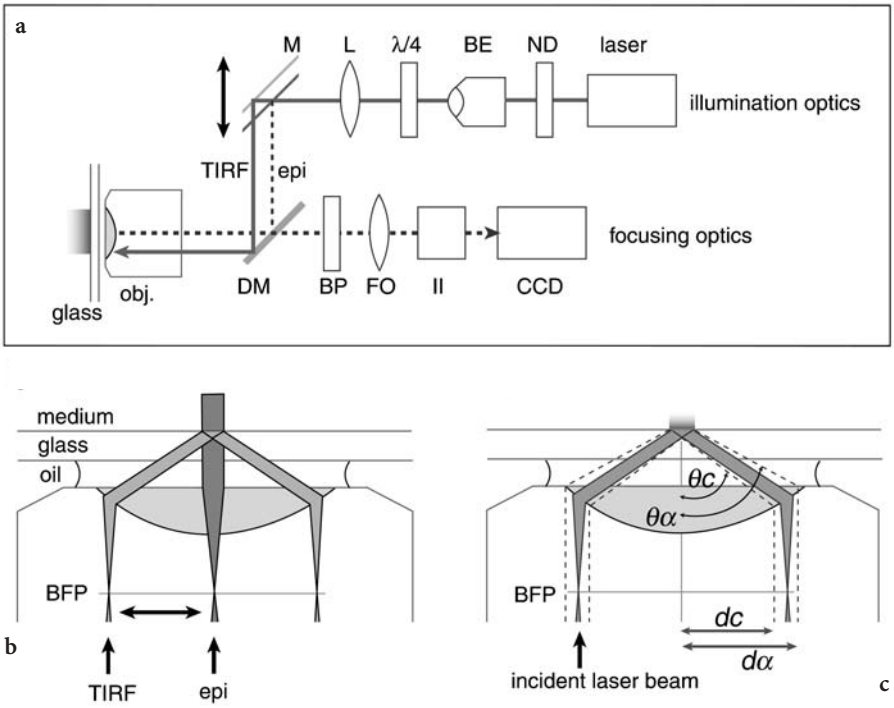


Fig. 10a–c Schematic diagram of objective-type TIRFM: **a** configuration: ND, neutral density filter; BE, beam expander; λ/4, quarter-wave plate; L, focusing lens; M, mirror; DM, dichroic mirror; BP, band pass filter; FO, focusing optics of microscope; II, imager intensifier; CCD, charge-coupled device; **b** illumination can be switched between epi-fluorescence microscopy and objective-type TIRFM by shifting the position of the mirror (M). BFP, back focal plane; **c** schematic drawing of objective-type TIRFM. The incident light is focused on the back focal plane at a position between $d\alpha$ and dc corresponding to $\theta\alpha$ and θc, respectively

At the edge between $\theta\alpha$ and θc, the laser beam is incident above "critical angle" at the glass/water interface at where the beam is totally internally reflected, generating evanescent field in the water. Thus, by shifting the position of the one mirror, the type of illumination can be switched between epi-fluorescence microscopy and objective-type TIRFM.

Fluorescence from fluorophores is collected by the same objective. The background of scattered light is rejected with a band pass filter (BP). The fluorescence images are intensified by an image intensifier (e.g. Video Scope, VS4-1845 or Hamamatsu, GaAsP C8600-05) and acquired by a highly sensitive camera such as EB-CCD camera (e.g. Hamamatsu, EB-CCD, C7190-23). Alternatively, a cooled CCD camera with a high sensitivity can be used for single molecule imaging. For simultaneous observation of dual color and fluorescent resonance energy transfer (FRET) imaging, a dual-view optics (e.g. Hamamatsu, W-view optics, A8509) can be placed before the image intensifier.

Objective-type TIRFM requires an objective lens with a very high NA practically larger than 1.4. To make TIR at the glass/water interface, the laser light should be focused on the back focal plane between $d\alpha$ and dc which corresponds to $\theta\alpha$ and θc, respectively (Fig. 10c). The angle $\theta\alpha$ is determined by $n_g \sin\theta\alpha = \mathrm{NA}$, where n_g is reflective index of the glass (1.52) and NA is the numerical aperture of the objective. θc is the critical angle of the glass/medium interface, which is determined by $n_g \sin\theta c = n_m \sin 90°$, where n_m is the reflective index of the aqueous medium. In the case of water, $n_m = 1.33$ and then $n_g \sin\theta c = 1.33$. $\theta\alpha$ must be greater than θc to generate an evanescent wave, thus an objective with a NA>1.33 is required for objective-type TIRFM. To observe living cells (1.37 refractive index) adhered on the glass, $n_g \sin\theta c = 1.37$, and thus an objective with a NA>1.37 is required. One of the most popular objectives for objective-type TIRFM has a NA of 1.45. In this case, $\theta\alpha$ and θc are 72.5° and 61.0° for water (or 64.3 ° for the cells), respectively.

Objective lenses with a NA>1.4 are commercially available. For the observation of purified biomolecules in an aqueous solution (1.33 refractive index), NA1.4 PlanApo 60× oil and 100× oil can be used to obtain images of single-molecules. The objective lenses with the higher NA available are NA1.65 Olympus Apo 100× oil High Reso, NA1.45 Olympus PlanApo 60× oil and so on. These lenses work well for observing living cells having the higher refractive index (about 1.37). The NA1.65 objective requires special glass and oil with a refractive index of 1.78. The NA1.45 objectives uses standard glass and oil with a refractive index of 1.52. Except for observing cells having with particularly dense organelles, the NA1.45 objectives are suitable for single-molecule observation in vitro and in living cells by objective-type TIRFM.

As described in the instruments of prism-type TIRFM, the optical materials such as filters, mirrors, coverslips, immersion oil should be chosen carefully to minimize autofluorescence and scattering.

4
Applications to Single-Molecule Nanobiology

Prism-type TIRFM provides a wide field of view with extremely low background noise over a wide range of wavelength. Also, these systems have less restrictions for designing optical systems in regards to the illumination and focusing requirements. The first observation of fluorescent single molecules in aqueous solutions was performed with prism-type TIRFM, regarded as one of the securest methods for single molecule imaging at the full video speed at room temperature [11]. After this breakthrough, prism-type TIRFM has been extensively developed to perform the experiments requiring extremely-low background and complex optical systems such as fluorescence spectroscopy [15, 25, 26], fluorescence resonance energy transfer [14], fluorescence polarization [27] and simultaneous measurement of fluorometry and nanometry [21].

Prism-type TIRFM is useful to obtain high quality images with high S/N ratios but there is it a restriction on the thickness of the sample. Because the space above the coverslip is occupied with a prism, it is sometimes difficult to apply prism-type TIRFM to thicker samples such as living cells and tissues. Tokunaga et al. [12] have developed a new type of TIRFM, the objective-type TIRFM (prismless TIRFM), which has a free space above the coverslip. This advantage in objective-type TIRFM made it possible to combine single-molecule imaging techniques with scanning probe microscopy such as atomic force and intermolecular force microscopy [28, 29]. Also, these types of microscopes are useful to combining with other techniques utilized in cell biology, such as microinjection, micromanipulation, and changing the media. Additionally, objective-type TIRFM is now commercially available.

One can measure the time courses, trajectories and distribution of molecular properties for individual biomolecules in vitro and in living cells using TIRFM. Thereby, this microscope has been employed to investigate motility of molecular motors, interactions of macromolecules, enzyme reactions, conformational dynamics of macromolecules, and the events associated with cell signaling. There have already been many reviews dealing with the new insights on biomolecules and biosystems that have emerged from single molecule studies [4–10]. There is a number of single molecule imaging techniques other than TIRFM applied to biomolecules and biosystems.

4.1
Motility of Motor Proteins

The ability of TIRFM to detect the positions and its changes of single molecules with high precision has been used to investigate motility of molecular motors. In fact, single molecule imaging techniques were first applied to molecular motors [11]. Myosin, kinesin, dynein and nucleotide polymerases are motor proteins that move along filamentous rail molecules such as F-actin, microtubule, and RNA or DNA. The first direct observation of the motility of mole-

cular motors was performed on kinesin [13]. Kinesin has very high duty ratio (ratio of interaction time over the total time of ATPase cycle), more than 99%, and thus hardly dissociates from microtubule [30]. Hence, a single kinesin molecule on a microtubule can move distances longer than the diffraction limit without dissociating from microtubule. The high duty ratio was the key for imaging of molecular motor motility at the single molecule level. In the case of skeletal muscle myosin II, which has a very low duty ratio, processive movements has not been observed at the single molecule level under TIRFM. Several types of molecular motors such as unconventional myosins, kinesin superfamily proteins, RNA polymerase have been visualized to move processively [31–35]. One of the insights emerging from the single molecule studies on molecular motors is that motor proteins move by biased Brownian motion [6, 9].

GFP (or YFP)-tagging techniques have been combined with single molecule imaging techniques to investigate molecular motors. Fluorescence from GFP (or YFP) is strong enough to be observed at the single molecule level. Single molecule imaging techniques have very high sensitivity and specificity to detect fluorophores, thus, it is not necessary to purify the motor proteins in order of milligrams to a high degree of purity to measure its motile properties. In fact, GFP-tagged myosin was prepared in vitro using T7 RNA polymerase and rabbit reticulocyte lysate (in vitro translation system), and then its ATPase activity was assay under TIRFM at the single molecule level [36]. The observation of GFP-tagged single kinesin molecule motility has been performed [37] and similar studies have been conducted using Myosin V, an unconventional myosin [35]. The fluorescent proteins can be used to mark the position of the molecules. However, it should be used carefully for more intricate purposes, e.g., a donor or acceptor of fluorescence resonance energy transfer (FRET), because GFP and its variants frequently exhibit blinking [38]. At the present stage of the techniques, the photochemical and photophysical properties of the fluorescent proteins are so obstinate that results obtained from the proteins are often difficult to be interpreted.

4.2 Enzymatic Reactions

The example discussed here is single molecule measurements of ATPase cycle by molecular motors. Many kinds of biomolecules have ATPase or GTPase activity. The ATPase reaction of myosin has been visualized at the single molecule level using fluorescent ATP analogues [11, 12, 39, 40]. Under the microscope, the free single fluorescent ATP molecule in aqueous solutions cannot be imaged clearly as a fluorescent spot at the video rate, because the fluorescent ATP molecules undergo rapid Brownian motion. When the fluorescent ATP molecule binds to myosin immobilized on a glass surface, the fluorescence arising from the ATP molecule becomes visible as a single spot. After the fluorescent ATP molecule is hydrolyzed into ADP and inorganic phosphate, and then released from the immobilized myosin, the fluorescent spot suddenly disappears. Thus,

the myosin ATPase reactions can be visualized by observing this on/off kinetics of a fluorescent ATP molecule [11]. The distribution of the ATP binding duration measured from the single molecule experiments could be well fitted to a single exponential curve, revealing the dissociation rates of ATP from myosin molecules. This calculated dissociation rate agreed well with the ATP turnover rate that was measured in an ensemble of myosin molecules. This may be the first confirmation of the compatibility between an ensemble and a single molecule experiments. Such single molecule imaging techniques of ATPase reaction have been combined with optical trapping nanometry to simultaneously measure individual ATPase cycles and the mechanical events of a single myosin molecules, revealing the relationship between the mechanical and chemical events of the myosin ATPase cycle [21]. This work suggests that myosin molecules have a hysteresis or memory state, which stores chemical energy from ATP hydrolysis. Xie and colleagues have reported a single molecule approach to determine enzymatic kinetics in vitro [41, 42]. They also suggested the existence of a memory state in a flavoenzyme. Similar methods can be used to determine the dissociation rates of ligands to receptor molecules in living cells by using a fluorescently-labeled ligand analog [19].

4.3 Interactions of Protein Molecules

The interaction between molecules can be directly detected by observing co-localization of fluorescently-modified molecules or, more strictly, fluorescence resonance energy transfer (FRET) between the molecules of interest.

Molecular chaperons assist in folding of protein molecules by using chemical energy released from ATP hydrolysis. The dynamics of the interaction between a molecular chaperon, GroEL, and co-chaperon GroES or its substrate, lactoglobin, was analyzed at the single molecule level [43, 44]. GroEL and its associated proteins were fluorescently-modified separately with dyes of different excitation and emission wavelengths, and their binding was detected by imaging their co-localization. These single molecule observations revealed a new kinetic intermediate in GroEL-assisted protein folding processes. The intermediate is not observed in ensemble measurements involving a large number of molecules, because the dynamic properties of biomolecules are averaged in ensemble measurements.

Single pair FRET has also been used to image macromolecular interactions at the single molecule level. FRET takes place only if the donor and acceptor fluorophores are very near to each other (distance, <10 nm). FRET is the most suitable method for detecting the complex formation between two molecules, whereas observation of co-localization of two molecules is sometimes difficult to interpret because of the diffraction limit. Ishii et al. (1999) measured FRET between single Cy3 and Cy5 fluorophores conjugated separately to α-tropomyosin subunits [14]. When the two subunits of α-tropomyosin attached to the different fluorophores formed a homodimer, the two fluorophores were close enough

for FRET to occur at the energy transfer efficiency near 100%. Similar methods have been used to detect conformational changes of single ribozyme molecules [45]. TIRFM for FRET experiments can be configured with the equipment for dual color imaging or fluorescence spectroscopy.

4.4 Conformational Dynamics of Macromolecules

To examine the photophysical properties of fluorescent molecules and the conformational dynamics of single macromolecules etc., prism-TIRFM-based fluorescence spectroscopes have been employed. For spectroscopic measurements, the baseline of single molecule spectra must be as low as possible, and the prism TIRFM matches this requirement. We constructed a spectroscope combined with a prism TIRFM, in which a grating spectrograph was coupled to a cooled CCD camera [15]. The conformational dynamics of myosin was investigated using the spectroscopic measurement of an environment-sensitive fluorophore conjugated to myosin (Fig. 11). The fluorescence spectra of fluorophores attached to myosin showed a fluctuation with time over a time scale of seconds, suggesting slow spontaneous conformational changes in the myosin molecule. Suzuki et al. (2002) developed another configuration for single molecule fluorescence spectroscope based upon prism TIRFM [25]. They introduced a specially-designed wedge prism into the focusing optical system in order to obtain spectral images of individual fluorophores with an imaging camera device. The fluorescence images of single fluorescent molecules under observation were dispersed by wavelengths and the resulting dispersed images were focused on the faceplate of an image-intensifier-coupled cooled CCD camera. Thus, fluorescence spectra of many single fluorescent molecules in the microscope field could be acquired during a single measurement. Ma et al. (2000) also developed a similar single molecule spectroscope to that of Suzuki and colleagues [25, 26]. They utilized a transmission grating as a dispersing optic in place of the dispersion prism.

The sensitivity of detecting the time courses and rare events has been applied to the observation of molecular dynamic processes in macromolecules. The on/off blinking processes of GFP molecules was investigated by single molecule imaging using a prism TIRFM [38, 46]. These investigators speculated that the blinking was likely responsible for a photo-induced conformational change of GFP. Single molecule fluorescence spectroscopy was applied using a synthetic photo-functional rigid rod macromolecule immersed in pure methanol [47]. They observed jumps in the fluorescence spectrum in a time scale of seconds, which suggested that the spectral change would be attributed to slow conformational changes due to thermal fluctuation. Single molecule measurement of fluorescence polarization were also performed to probe conformation changes of motor proteins [26, 48, 49]. A divalently-reactive fluorophore (bisiodoacetamidorhodamine) was rigidly conjugated to calmodulin bound to myosin V at a known orientation. The fluorescence polarization of the

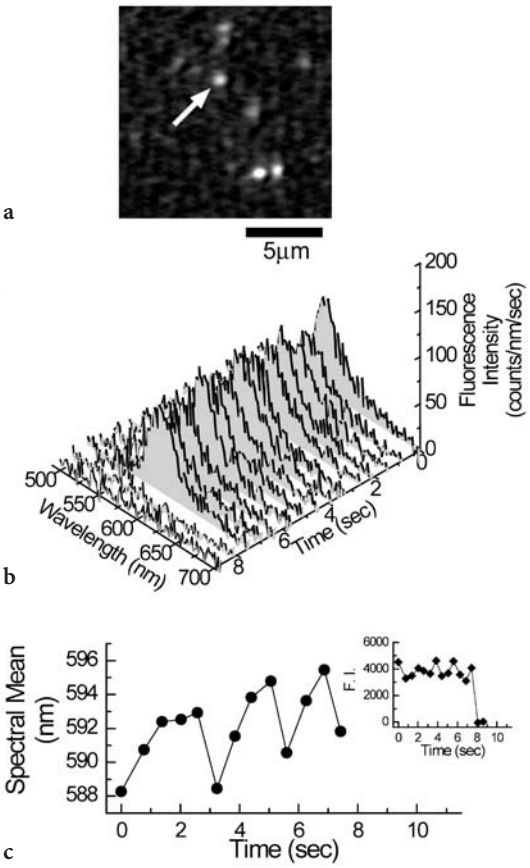


Fig. 11a–c Spectral fluctuations of single tetramethylrhodamine (TMR)-labeled myosin molecules: **a** fluorescence images of single TMR-myosin molecules. *Arrow* represents the fluorescence spot analyzed; **b** a series of fluorescence spectra from a single TMR-myosin (*arrow* in a) with time. Spectra were taken sequentially every 0.61 s with the collection time of 0.5 s for each spectrum; **c** fluctuations in the fluorescence spectrum of single TMR-myosin. The spectral mean was plotted as a function of time. (*inset*) Total fluorescence intensity obtained by integrating fluorescence intensity for wavelengths from 540 to 600 nm for each spectrum

fluorophore was measured at the single molecule level during the myosin V motility, and the tilting of a working myosin molecule was successfully detected.

4.5 Ion Channels

Ion channels regulate ion currents and electrochemical potentials across plasma membranes, which are responsible for excitability of nerve, muscle and other type of cells. Single channel current recording using the patch clamp method has revealed electric properties of many kinds of channels at a single

molecule level [50]. However, the detailed mechanism of how the conformational states of channels are related to their functions remains to be unsolved. Ide and his colleagues have been developing a new experimental system, in which electrical currents and fluorescence signals from single ion channels can be measured simultaneously [16, 17, 51]. Fluorescently labeled ion channels were incorporated into a planar lipid bilayer formed horizontally on the agar-coated glass and observed using an objective type TIRFM. Further development of this technique combined with FRET or dual-view optics will make it possible to observe the interactions between ion channels and their ligands or regulators, and also to follow simultaneously the conformational changes and ion currents in vitro.

4.6
Single Molecule Imaging in Living Cells

Living cells are complex but well-organized systems comprising various kinds of molecular machines. Cells can respond flexibly and adaptively to environmental changes via dynamic processes in reaction networks of molecular machines. The question of how extracellular and intracellular signals are transmitted and processed in the bio-molecular networks is a central theme in life science. Single molecule imaging techniques have made it possible to detect quantitatively the locations, movements, turnovers, and complex formations of signaling molecules at the single molecule level in living cells, providing powerful tools to elucidate molecular mechanisms of cell signaling [9, 10].

As illustrated in Fig. 12a, the objective-type TIRFM can be used for this purpose, because this type of microscopes provides a free space above the glass surface and thus it has no limitation on the thickness of samples. The surface of the cell close to the glass surface is locally illuminated by the evanescent wave, lighting up biological events on the cell surface at the single molecule level. In some type of cells, it is possible to observe the other (apical) surface by adjusting the incident light carefully [18].

Sako et al. (2000) first reported single molecule imaging in living cells. They demonstrated that single fluorophore attached to epidermal growth factor (EGF) molecules bound to their receptors (EGFR) could be observed on the surface of a living mammalian cell [18]. Fluorescently-labeled EGF (Cy3-EGF) in solution will not produce clear images of single molecules because of the rapid Brownian motion. When Cy3-EGF binds to its receptor on the surface of cells, Cy3 can be seen as a clear fluorescent spot. By following Cy3-EGF bound to the receptor, they could directly visualize the dimerization of the EGF-EGFR complex. By using a fluorescently-labeled antibody that recognized specifically the phosphorylated receptor, autophosphorylation of the EGFR was also observed at the single molecule level. The dimerization and autophosphorylation of EGFR are known to be essential processes for early signal transduction of EGF.

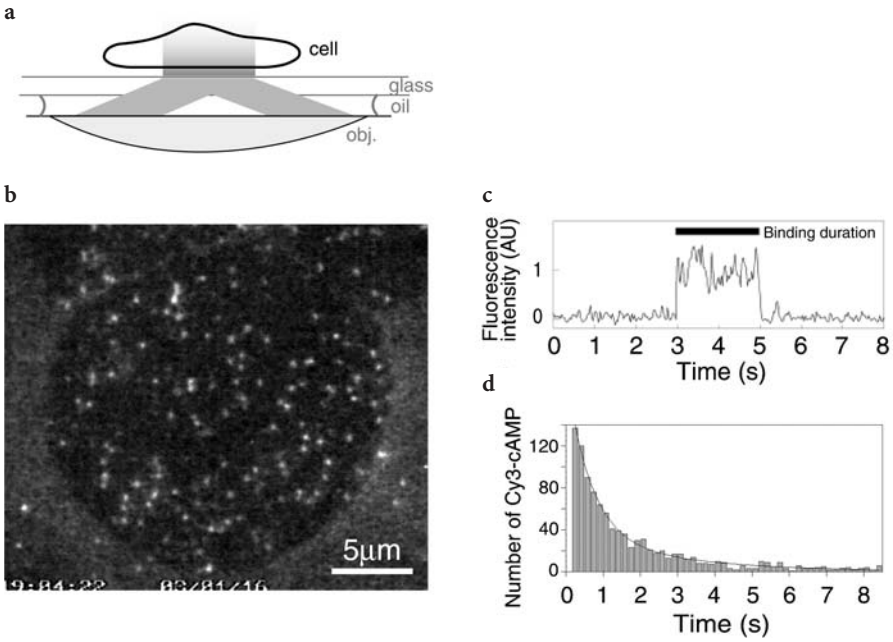


Fig. 12 **a** Single molecule imaging in living cells using objective-type TIRFM. **b** Single molecule imaging of Cy3-cAMP bound to the receptors on the surface of living *Dictyostelium* cells. **c** The time course of fluorescence intensity of Cy3-cAMP spots bound to cells. **d** Histogram of the binding duration of Cy3-cAMP. The *line* represents the fitting of data to a sum of the exponential functions $\exp(-k_{-1}t)$, where k_{-1} is the dissociation rates of Cy3-cAMP from the receptors

Single-molecule imaging techniques can also be used to monitor GFP (or YFP)-tagged proteins in living cells [52–55]. Compared with cyanine dyes such as Cy3 and Cy5, fluorescence emitted from GFP variants is weak and quickly photo-bleached under the continuous illumination of excitation light (generally within about 5–10 s). Furthermore, GFP variants have higher molecular weights than the fluorescence dyes, so that GFP-tagged proteins sometimes lose their functions by steric inhibition. Nevertheless, GFP variants have some advantages for single molecule imaging in living cells. For example, it is easy to prepare cells expressing GFP-tagged proteins using genetic engineering. The labeling ratio of GFP to proteins can be 100%. However at present, CFP and BFP, which are kinds of GFP variants, are not suitable for single molecule imaging because of the rapid photo-bleaching.

4.7 Single Molecule Measurements of Kinetic Rates in Living Cells

Taking the ligand binding experiments in *Dictyostelium* cells as an example [19], a single molecule approach to receptor kinetics in living cells will be

explained here (Fig. 12). *Dictyostelium* cells exhibit a chemotactic response to cyclic adenosine 3′5′-monophosphate (cAMP), which is mediated by a cell surface receptor and G-protein-linked signaling pathway. To elucidate how the cells sense concentration gradients of cAMP, a fluorescent-labeled cAMP (Cy3-cAMP) was prepared, and its binding to the receptors on the living cells was observed by objective-type TIRFM (Fig. 12b) [19].

As shown in Fig. 12c, the binding duration of individual Cy3-cAMP molecules can be determined by measuring the time difference between the appearance and the disappearance of the Cy3-cAMP spots on the cells. The ligand binding is stochastic events in nature, thus an individual event of a single ligand molecule is not particularly informative. However, by means of making a histogram of the binding duration, the distribution of the binding duration can be determined statistically (Fig. 12d). Fitting the distribution with a sum of exponential functions, $\exp(-k_{-1}t)$, the k_{-1} values can be obtained. The k_{-1} values are the dissociation rates of the ligands bound to its receptors, and reflect the kinetic state of the receptors. In the case of G-protein-coupled receptors (GPCR), the dissociation rates k_{-1} depend on the interaction of GPCR with its coupled G-proteins.

This analysis was performed separately for Cy3-cAMP bound to the anterior pseudopods and the posterior tails of the *Dictyostelium* cells undergoing chemotaxis, revealing that Cy3-cAMP receptor complexes on anterior pseudopods dissociated about three times faster than those on posterior tails (Fig. 13). The dissociation curves were fitted to a sum of two exponential functions, indicating the presence of at least two receptors that adopt different kinetic states. Further characterization of this difference in the receptor states between anterior and posterior region suggests that the difference in the receptor states

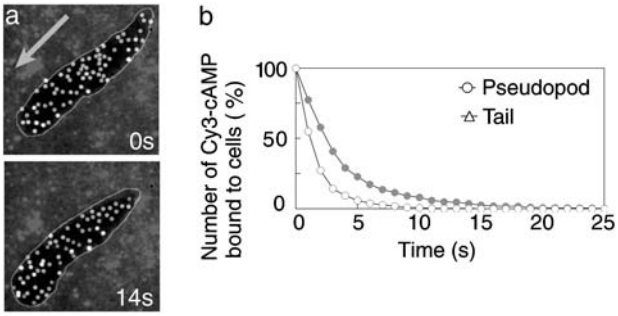


Fig. 13a, b Single-molecule analysis of the binding kinetics of Cy3-cAMP on Dictyostelium cells undergoing chemotaxis: **a** receptor occupancy of the cell under a gradient of Cy3-cAMP. The *arrow* represents the direction of the source of Cy3-cAMP. Time is given in seconds. Individual Cy3-cAMP spots and the cell contour are traced; **b** the release curves of Cy3-cAMP spots bound to the anterior pseudopods or the posterior tails. The release curves of bound Cy3-cAMP were fitted to two exponentials. At anterior pseudopod, k_{-1}=1.1 s^{-1} (71%) and 0.39 s^{-1} (29%). At the posterior tail, k_{-1}=0.39 s^{-1} (76%) and 0.16 s^{-1} (24%). Chemical gradients in cAMP concentration may be converted into the differences in the receptor states

reflect the difference in the coupling of cAMP receptor with G-proteins. Chemical gradients in cAMP concentration may be converted into the differences in the receptor states. Thus, single molecule imaging in living cells allows us to directly monitor the behavior of signaling molecules in relation to intracellular environments, cell polarity, and the cell's response.

Histograms and statistic analysis have been extensively used in the single channel recording experiments and single molecule enzyme kinetics [41, 42, 50].

5
Some Advantages and Problems of Fluorescent Dyes in Single Molecule Fluorescence Measurements

One of the advantages of using fluorescent dyes for single molecule study is their small mass and volume. Other than fluorophores, particle probes such as colloidal golds and latex beads have been used for single molecule studies. These probes, however, sometimes affect the native properties of biomolecules by interfering with the freedom of motion of the attached site because they have a much higher mass and volume. Fluorescent dyes also may interfere with motion, but the degree of the effect may be smaller owing to their smaller mass and size. In addition, fluorescent dyes can be conjugated specifically to many kinds of chemical groups of biomolecules in order to prepare a wide variety of functional probes. Moreover, these functional fluorescent probes can be prepared at a high purity by conventional organic chemistry and chromatography. The particle probes cannot be synthesized in such high precision. Below some of the restrictions in the usage of fluorescent dyes for single molecule fluorescence measurements will be discussed.

Photobleaching of fluorescent dyes is the most limiting factor for single molecule fluorescence experiments. Photobleaching restricts the total number of photons to be detected at the sensor of the experimental apparatus. For example, an order of 10^4~10^5 photons are typically detected from a single Cy3 molecule [56]. The total number of photons determines not only the data length obtained from a single dye molecule but also the S/N ratio of measurement. Because the fluorescence intensity of a fluorescent dye generally depends on the power of excitation light beam, the power must be adjusted so that a desired data length and S/N ratio is obtained. Higher excitation beam power results in higher fluorescence intensity but simultaneously shorter duration of the emitted fluorescence.

In general, the fluorescence intensity of single fluorescent molecules is almost in the range of the single photon counting level. In our TIRFM system, when we observe single Cy3 molecules for tens of seconds, its intensity detected on the faceplate of a detector (cooled CCD, image intensifier or avalanche photodiode) would be approximately 10^3 (photons/s). Therefore, at the full video rate (30 Hz), the number of photons per video frame is approximately 30. Assuming that there are no changes in the environmental conditions of the

fluorophore or the fluorophores do not change between the singlet and triplet states, the deviation of fluorescence intensity at best can be formulated by a Poisson process. Let X be the number of photons detected per sampling interval. If we assume the Poisson distribution of X, then the variance can be written as

$$\mathrm{Var}(X) = \lambda , \tag{16}$$

where

$$\lambda = \mathrm{E}(X) . \tag{17}$$

Therefore, the standard deviation σ_M of the number of photons measured by a detector can be computed as

$$\sigma_M = \sqrt{Q\lambda + \varepsilon_R^2} , \tag{18}$$

where Q and ε_R are the quantum efficiency and readout noise of the detector, respectively. For example, when λ=30, Q=30% (typical value for GenIII image intensifiers) and ε_R=0 (as the best case), we have σ_M=3. This value is not negligible because the average number of photons sensed by the detector is $Q\lambda$=9 and hence the S/N ratio is only $Q\lambda/\sigma_M$=3 at best. We often find large fluctuations in the time course of single molecule fluorescence (e.g. Fig. 14). Even if we develop better detectors with a higher quantum efficiency and lower readout noise in the future, the fluctuations of the fluorescence intensity will exist,

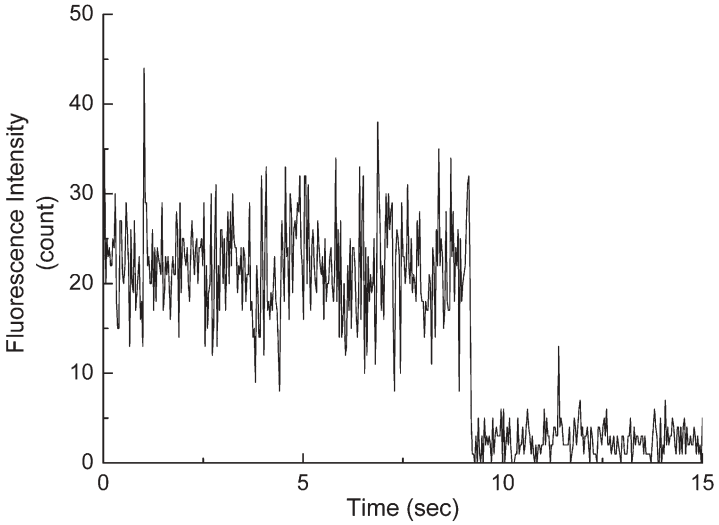


Fig. 14 Typical time course of the fluorescence intensity of single Cy3 molecules. The fluorescence intensity was measured with an avalanche photodiode single photon counting module (SPCM-AQR-16, Perkin Elmer, Canada). Excitation wavelength, 514.5 nm. Sampling interval, 30 ms

because the fluorescence intensity is extremely low. Thus, we have to design an experimental setup and laser powers so that we can obtain both the desired data length and S/N ratio, making the best use of the limited number of photons. Moreover, it should be noted that we cannot acquire the data length and S/N ratio beyond such limitation.

A microscope with low background light and a detector with low dark counts are also essential for single molecule fluorescence imaging. Let λ_F be the fluorescence intensity of a single fluorescent molecule, λ_B the intensity of the background light and λ_D the dark count (unit, photons per sampling interval). The S/N ratio r_M of the measurement in account of the background light and dark count can be computed as

$$r_M = \frac{Q\lambda_F}{\sqrt{Q(\lambda_F + \lambda_B) + \lambda_D + \varepsilon_R^2}} \,. \tag{19}$$

This equation indicates that if the background and dark count increase, then r_M decreases and approaches zero. One may think that even if the background is very high, we may be able to measure the fluorescence intensity of a single fluorescent molecule by simply subtracting the mean intensity of background from a measured total intensity. However, this notion is incorrect because the fluorescence intensity of interest ($Q\lambda_F$) can be overcome by high background

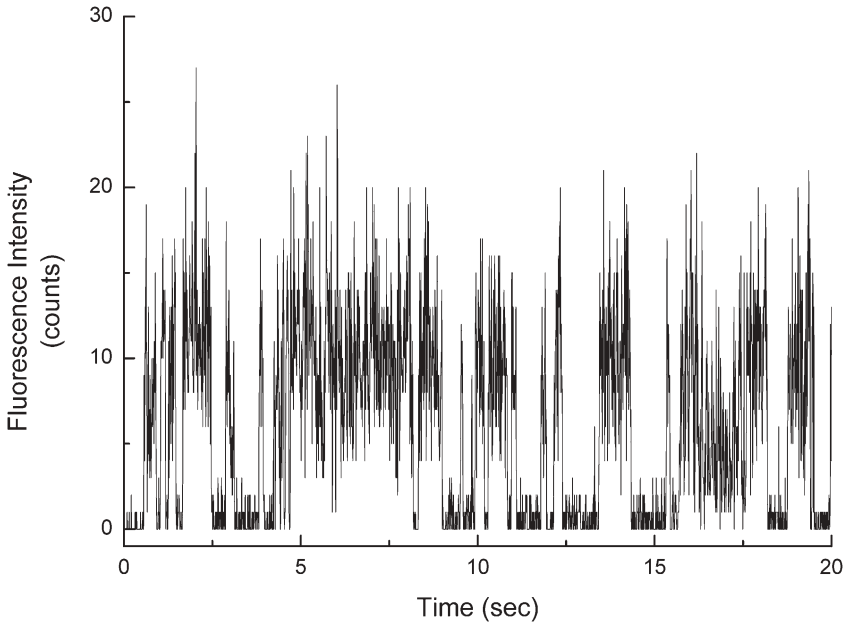


Fig. 15 An example of the blinking measured from a single tetramethylrhodamine molecules. The fluorescence intensity was measured with an avalanche photodiode single photon counting module. Excitation wavelength, 514.5 nm. Sampling interval, 10 ms

($Q\lambda_B$) and dark count (λ_D). It is not practical to measure the fluorescence intensity of single fluorescent molecule buried in high background light.

In addition to photobleaching, there is a phenomenon referred to as blinking. This is a process whereby the fluorescence intensity of single fluorescent molecule changes suddenly and reversibly, and thus it interferes with the measurement of fluorescence intensity (Fig. 15). Blinking has been observed to occur in the time scales from microseconds to seconds. GFP and its variants frequently exhibit blinking. The mechanism of blinking is not completely understood, but may be responsible for the isomerization of fluorophore and intersystem crossing between singlet and triplet states [57].

Acknowledgements The authors thank Prof. T. Yanagida, members and ex-members of his laboratories in Osaka University, ERATO, ICORP and CREST for helpful discussion and valuable comments. The authors also thank Dr. J. West and F. Brozovich for critical reading of the manuscript.

References

1. Axelrod D, Hellen EH, Fulbright RM (1992) Total internal reflection fluorescence. In: Lakowicz JR (ed) Topics in fluorescence spectroscopy. Vol 3. Biochemical applications. Plenum Press, New York, p 289
2. Axelrod D (1989) Methods Cell Biol 30:245
3. Axelrod D (2001) Traffic 2:764
4. Harada Y, Funatsu T, Tokunaga M, Saito K, Higuchi H, Ishii Y, Yanagida T (1997) Methods Cell Biol 55:117
5. Weiss S (1999) Science 283:1676
6. Yanagida T, Kitamura K, Tanaka H, Iwane AH, Esaki S (2000) Curr Opin Cell Biol 12:20
7. Forkey JN, Quinlan ME, Goldman YE (2000) Prog Biophys Mol Biol 74:1
8. Ishii Y, Ishijima A, Yanagida T (2001) Trends Biotech 19:211
9. Ishijima A, Yanagida T (2001) Trends Biochem Sci 26:438
10. Sako Y, Yanagida T (2003) Nat Rev Mol Cell Biol Suppl SS1–5
11. Funatsu T, Harada Y, Tokunaga M, Saito K, Yanagida T (1995) Nature 374:555
12. Tokunaga M, Kitamura K, Saito K, Iwane AH, Yanagida T (1997) Biochem Biophys Res Commun 235:47
13. Vale RD, Funatsu T, Pierce DW, Romberg L, Harada Y, Yanagida T (1996) Nature 380:451
14. Ishii Y, Yoshida T, Funatsu T, Wazawa T, Yanagida T (1999) Chem Phys 247:163
15. Wazawa T, Ishii Y, Funatsu T, Yanagida T (2000) Biophys J 78:1561
16. Ide T, Yanagida T (1999) Biochem Biophys Res Commun 265:595
17. Ide T, Takeuchi Y, Aoki T, Yanagida T (2002) Jpn J Physiol 52:429
18. Sako Y, Minoguchi S, Yanagida T (2000) Nature Cell Biol 2:168
19. Ueda M, Sako Y, Tanaka T, Devreotes P, Yanagida T (2001) Science 294:864
20. Fornel F (2001) Evanescent waves. Springer, Berlin Heidelberg New York
21. Ishijima A, Kojima H, Funatsu T, Tokunaga M, Higuchi H, Tanaka H, Yanagida T (1998) Cell 92:161
22. Ambrose WP, Goodwin PM, Nolan JP (1999) Cytometry 36:224
23. Tsien RY, Waggoner AS (1995) Fluorophores for confocal microscopy: photophysics and photochemistry. In: Pauley J (ed) Handbook of confocal microscopy. Plenum Press, New York, p 267

24. Yokota H, Saito K, Yanagida T (1998) Phys Rev Lett 80:4606
25. Suzuki Y, Tani T, Sutoh K, Kamimura S (2002) FEBS Lett 512:235
26. Ma Y, Shortreed MR, Yeung ES (2000) Anal Chem 72:4640
27. Forkey JN, Quinlan ME, Shaw MA, Corrie JE, Goldman YE (2003) Nature 422:399
28. Kitamura K, Tokunaga M, Iwane AH, Yanagida T (1999) Nature 397:129
29. Aoki T, Sowa Y, Yokota H, Hiroshima M, Tokunaga M, Ishii Y, Yanagida T (2001) Single Mol 3:183
30. Romberg L, Pierce DW, Vale RD (1998) J Cell Biol 140:1407
31. Harada Y, Funatsu T, Murakami K, Nonoyama Y, Ishihama A, Yanagida, Y (1999) Biophys J 76:709
32. Okada Y, Hirokawa N (1999) Science 283:1152
33. Sakamoto T, Amitani I, Yokota E, Ando T (2000) Biochem Biophys Res Commun 272:586
34. Inoue Y, Iwane AH, Miyai T, Muto E, Yanagida T (2001) Biophys J 81:2838
35. Nishikawa S, Homma K, Komori Y, Iwaki M, Wazawa T, Hikikoshi Iwane A, Saito J, Ikebe R, Katayama E, Yanagida T, Ikebe M (2002) Biochem Biophys Res Commun 290:311
36. Iwane AH, Funatsu T, Harada Y, Tokunaga M, Ohara O, Morimoto S, Yanagida T (1997) FEBS Lett 407:235
37. Pierce DW, Vale RD (1998) Methods Enzymol 298:154
38. Dickson RM, Cubitt AB, Tsien RY, Moerner WE (1997) Nature 388:355
39. Oiwa K, Eccleston JF, Anson M, Kikumoto M, Davis CT, Reid GP, Ferenczi MA, Corrie JE, Yamada A, Nakayama H, Trentham DR (2000) Biophys J 78:3048
40. Oiwa K, Jameson DM, Croney JC, Davis CT, Eccleston JF, Anson M (2003) Biophys J 84:634
41. Lu HP, Xun L, Xie S (1998) Science 282:1877
42. Xie S (2001) Single Mol 4:229
43. Yamasaki R, Hoshino M, Wazawa T, Ishii Y, Yanagida T, Kawata Y, Higurashi T, Sakai K, Nagai J, Goto Y (1999) J Mol Biol 292:965
44. Taguchi H, Ueno T, Tadakuma H, Yoshida M, Funatsu T (2001) Nat Biotechnol 19:861
45. Zhuang X, Kim H, Pereira MJ, Babcock HP, Walter NG, Chu S (2002) Science 296:1473
46. Moerner WE, Peterman EJ, Brasselet S, Kummer S, Dickson RM (1999) Cytometry 36:232
47. Shinohara K, Yamaguchi S, Wazawa T (2001) Polymer 42:7915
48. Peterman EJ, Sosa H, Goldstein LS, Moerner WE (2001) Biophys J 81:2851
49. Sosa H, Peterman EJ, Moerner WE, Goldstein LS (2001) Nat Struct Biol 8:540
50. Sakmann B, Neher E (eds) (1995) Single channel recording. Plenum, New York
51. Ide T, Takeuchi Y, Yanagida T (2002) Single Mol 3:33
52. Sako Y, Hibino K, Miyauchi T, Miyamoto Y, Ueda M, Yanagida T (2000) Single Mol 1:151
53. Iino R, Koyama I, Kusumi A (2001) Biophys J 80:2667
54. Watanabe N, Mitchison TJ (2001) Science 295:1083
55. Hibino K, Watanabe TM, Kozuka J, Iwane AH, Okada T, Kataoka T, Yanagida T, Sako Y (2003) CHEMPHYSCHEM 4:748–753
56. Schmidt T, Kubitscheck U, Rohler D, Nienhaus U (2002) Single Mol 3:327
57. Widengren J, Schwille P (2000) J Phys Chem A 104:6416

Adv Biochem Engin/Biotechnol (2005) 95: 107–142
DOI 10.1007/b102212

Fluorescence Correlation Spectroscopy with Autofluorescent Proteins

Tobias Kohl[1] · Petra Schwille[2] (✉)

[1] Pastor-Sander-Bogen 92, 37083 Göttingen, Germany
tkohl1@web.de
[2] Institute of Biophysics, BioTec, TU Dresden; c/o Max-Planck-Institute for Molecular Cell Biology and Genetics, Pfotenhauerstrasse 108, 01307 Dresden, Germany
schwille@mpi-cbg.de

Abstract Fluorescence correlation spectroscopy (FCS) is a versatile technique operating at the single-molecule level, that successfully meets many challenges of modern biological research. Based on the detection of mobile fluorescent molecules diffusing in and out of a diffraction-limited laser focus, the method allows to resolve particle dynamics within cells and their compartments. Previous FCS studies have described various parameters of protein function, namely mobility, transport and localization phenomena, enzymatic turnovers of biochemical substrates and molecular association and dissociation reactions. Recent progress in the application of FCS to intracellular systems has particularly taken advantage of detecting autofluorescent proteins and their genetically encoded fusions to cellular proteins. This review discusses recent applications of FCS analysis with and on fluorescent proteins, particularly highlighting chemical and physical properties. Inherent limitations of the presented approaches are discussed in detail and strategies for optimisation of experimental systems outlined.

Keywords FCS · GFP · DsRed · Cross-correlation analysis · Single molecule · In vivo · Two-photon excitation · FRET · Blinking · Fluorescence resonance energy transfer

List of Abbreviations

FCS Fluorescence correlation spectroscopy
FRET Fluorescence resonance energy transfer
RFP Red fluorescent protein
TPE Two-photon excitation

1 Introduction

Biological research in the post-genomic era has so far been particularly concerned with predicting and studying protein function. To this end, experiments are increasingly performed in the live cell environment, taking advantage of fluorescence-based microscopic techniques that allow for real-time data acquisition, instead of employing cell extracts or purified recombinant proteins in biochemical assays. These in vivo methods rely on fluorescent reporter systems consisting of the protein of interest labelled with a fluorescent dye molecule. To minimize any interference with the cellular system, this is ideally realized by a genetic fusion with fluorescent proteins, so-called genetically encoded probes.

The present article focuses on a particular optical technique with single-molecule sensitivity, namely Fluorescence Correlation Spectroscopy (FCS) [1, 2], with a specific emphasis on its suitability for intracellular analysis of fluorescent proteins. FCS analyses minute spontaneous fluctuations in the fluorescence emission of small molecular ensembles to study underlying inter- and intramolecular dynamics. Combining single molecule sensitivity with large statistical averaging, fluctuation autocorrelation analysis allows for a highly quantitative simultaneous analysis of multiple molecular parameters, namely particle concentration and mobility, intramolecular dynamics and binding reactions. In addition to single-colour applications by autocorrelation analysis, dual-colour cross-correlation analysis has been established and proved particularly beneficial for the study of molecular association and dissociation events [3]. Since FCS observes fluorescent molecules at nanomolar or even lower concentrations, it is well suited to collect quantitative data under physiological conditions and monitor dynamic equilibria in vivo [4]. While bulk studies are typically performed at fairly high particle concentrations, subtle regulatory processes or cellular background activities might only be resolved at the single molecule level. Moreover, fluorescence correlation spectroscopy can be combined with two-photon excitation (TPE) conferring further advantages, in particular excitation of two or more spectrally different dyes with one laser line, and reducing cell damage and sample consumption by photobleaching [5–7]. Due to its sensitivity, FCS resolves intramolecular photodynamic processes of fluorescent proteins, like pH-dependent blinking [8–11].

Nevertheless, just like any other fluorescence-based technique, FCS benefits from the continued efforts for discovery and redesign of optimised fluorescent probes [12, 13]. Along with reporting recent applications of intracellular FCS [14–17] this article thus aims to provide a summary on fluorescent proteins (for thorough discussion see Miyawaki in this issue) from the specific viewpoint of their applicability in FCS experiments, with respect to their beneficial and limiting properties.

2
Theoretical and Practical Concepts of FCS

At first a brief introduction to the technique of Fluorescence Correlation Spectroscopy (FCS) shall be given, to lay the grounds for the discussion of applications following later in this review. Its large potential for simultaneous examination of molecular parameters like particle concentration, mobility, rate constants of intramolecular dynamics and molecular interactions has contributed to the present popularity and wide-spread application of FCS. Thus, the interested reader can be referred to a number of recent reviews and articles covering FCS in much more detail, addressing theoretical framework, technical setups, reported applications and practical aspects [18–21].

2.1 Autocorrelation Analysis

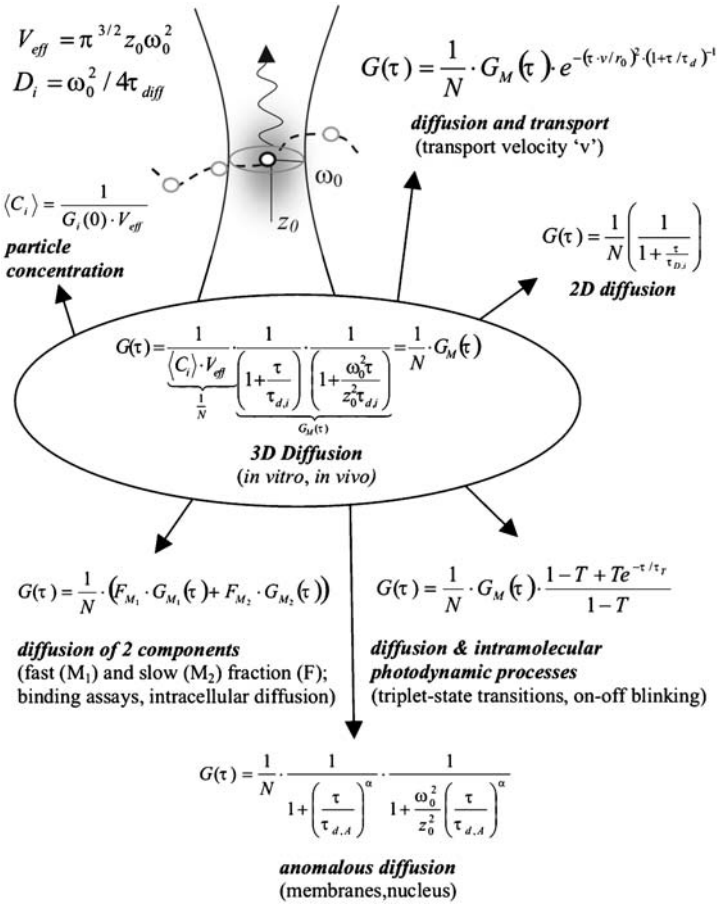
FCS is based on the universal formalism of correlation analysis, comparing time signals for series of lag times τ. In the common case, the signal results from fluorescence excitation and detection within a confocal volume [22]. The normalized fluorescence autocorrelation function is generally defined by

$$G(\tau) = \langle \delta F(t)\ \delta F(t+\tau)\rangle / \langle F(t)\rangle^2 \quad (1)$$

where $\delta F(t) \equiv F(t) - \langle F(t)\rangle$ describes the time-dependent fluctuation of fluorescence signals from a volume element $V_{eff} = \pi^{3/2}\omega_0^2 z_0$ (2). V_{eff} constitutes the detection volume (with ω_0 and z_0 describing the lateral and axial radius) that largely overlaps with the laser focus (Scheme 1). Fluorescence fluctuations principally result from fluorescing particles diffusing into and out of V_{eff} with a typical diffusion coefficient D_i (Scheme 1). The resulting fluctuation decay time τ describes the average particle residence time τ_{diff} within V_{eff}. Fluctuation amplitudes and their duration and temporal frequency of occurrence, are characteristic for the particle number within V_{eff}, the mode of particle movement and the presence of fast photodynamic processes. These contributions are represented in G(τ) by the particle number 'N', the characteristic diffusion behaviour '$G_M(\tau)$' and an additional factor induced by fast intra-molecular photodynamic processes. Scheme 1 presents the basic fitting models for autocorrelation curves G(τ). Starting with free two- or three-dimensional diffusion of a single species, the autocorrelation function can be modified adding intra-molecular photodynamic processes, introducing additional molecular species with different mobility, or considering obstructed 'anomalous' diffusion. Typical sources of fast intra-molecular fluctuations are excitation intensity dependent transitions of the dye to the long-living first excited triplet state and other less well-defined dark states, but also fluorophore isomerization, and reversible protonation [8, 9, 11, 23]. An exponential decay factor ($1 - T + Te^{-\tau/\tau_T}$) with the relaxation time τ_T corresponding to the triplet relaxation time, and the triplet fraction T of dark molecules, has to be introduced to account for these additional fluorescence fluctuations (Scheme 1).

Fluorescent dyes can also show distinct intra-molecular transitions between two or more differently bright states, e.g. a fast triplet transition and a slow on-off blinking, which have to be accounted for by an individual decay term for each transition. Figure 1 illustrates how autocorrelation curves change their shape due to intra-molecular transitions.

The local concentration of fluorescent molecules $\langle C_i\rangle$ can be determined for single species from correlation-curve amplitudes G(0) (extrapolated for $\tau \to 0$) (Scheme 1). Particle numbers N are proportional to inverse autocorrelation amplitudes G(0). If particles show a non-Brownian diffusion due to corralling or particle interaction with fixed structures, as has been described within membranes (2D) and cells an anomalous diffusion model can be applied to fit the data (Scheme 1) [24–26]. The autocorrelation function for anomalous 3D diffusion is



Scheme 1 Fitting functions for FCS analysis. Basic fitting functions and scheme for particle diffusion through the focal volume. Many experimental situations require combinations of the presented models. The complexity of a model to be applied is limited by the signal-to-noise ratio of recorded fluorescence fluctuations and the number of well-known parameters that can be fixed prior to analysis. The *grey shading* of the focus indicates the extension of V_{eff} and the Gaussian intensity distribution of the laser light within the focus

adapted by introducing an anomaly coefficient α indicating the degree of obstruction.

2.2 Cross-Correlation Analysis

Instead of autocorrelating a single signal trace, different signal traces originating from at least two spectrally different dyes within the same confocal volume element can be compared at high temporal resolution in a cross-correlation

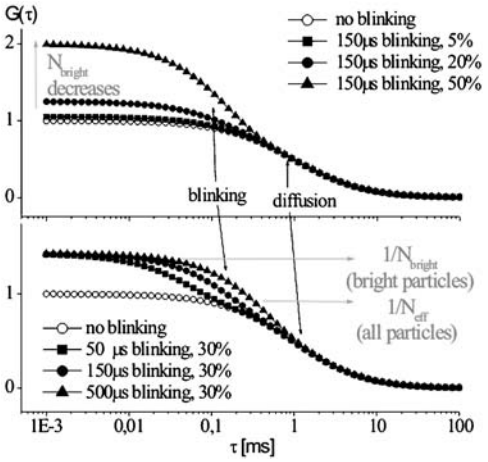


Fig. 1A, B Blinking in diffusion curves. The effect of blinking is visualized: **A** for an increasing dark fraction [%] and a fixed blinking decay time [µs]; **B** for a constant dark fraction and increasing blinking decay times. The number of bright particles decreases upon blinking, while the number of all detected particles stays constant (N=1). The correlation function acquires an additional shoulder and changes its shape. Diffusion takes place with τ_d=1 ms for all curves

mode (Fig. 2). Different dyes are either excited with two overlapping laser beams or TPE while fluorescence is collected in overlapping detection volumes of two separate detection pathways [27]. This way of analysis quantifies the degree of similarity between fluorescence fluctuations obtained from different dyes, and relies on fluorescence events that take place simultaneously. Only concomitant movement of spectrally different dyes accounts for cross-correlating fluctuation signals. Phenomena constricted to each individual fluorophore, like triplet-blinking, occur independently in each detection channel and do not contribute to the cross-correlation curve. The cross-correlation function for two diffusing species "i" and "j" of different colours is defined as

$$G_{ij}(\tau) = \langle \delta F_i(t)\, \delta F_j(t+\tau)\rangle / \{\langle F_i(t)\rangle \langle F_j(t)\rangle\} \quad (3)$$

The concentration of double-labelled substrate molecules can in ideally calibrated setups be easily determined from the auto- and cross-correlation amplitudes as shown in Fig. 2, since the cross-correlation amplitude is directly proportional to the number of double-labelled particles in V_{eff}. The percentage of double-labelled particles is thus given by the relative cross-correlation amplitude.

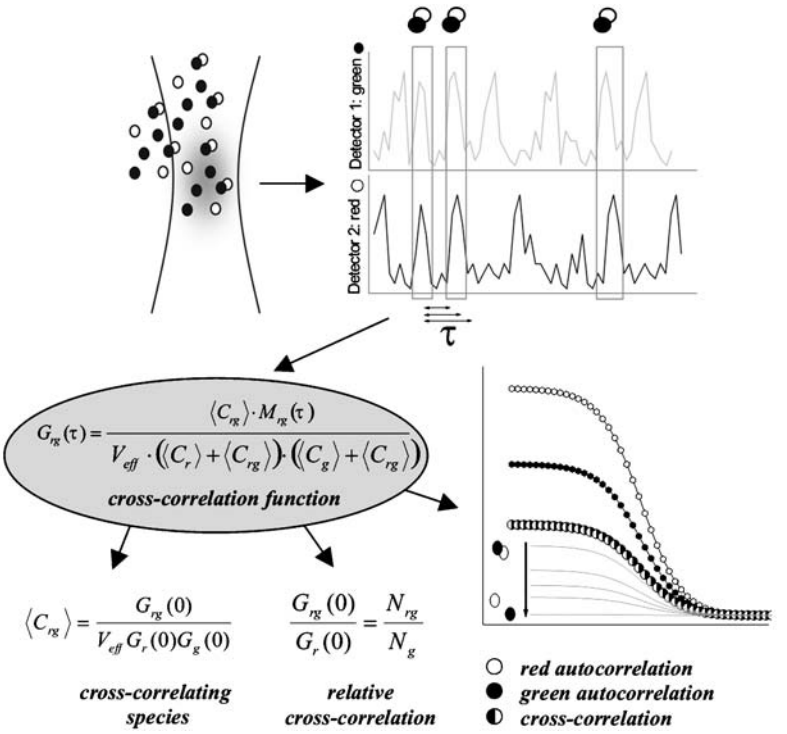


Fig. 2 Auto and cross-correlation analysis. Two traces of fluorescence fluctuations result from diffusion of single and double-labelled particles through V_{eff}. Simultaneous events in both channels have been highlighted in boxes and give rise to a cross-correlation amplitude $G_{rg}(0)$. Separation of connected labels results in a decrease of the cross-correlation amplitude while autocorrelation amplitudes stay constant. Changes in the number of double-labelled particles can be quantified both in absolute values $\langle C_{rg}\rangle$ and the relative cross-correlation

3 From Synthetic Dyes to Fluorescent Proteins

Single-molecule optical analysis relying on fluorescent dye molecules imposes stringent requirements such as high signal-to-noise ratios and steady photon fluxes for each observed molecule. Apart from high quantum yields and large absorption cross sections, fluorophores thus have to display a high photostability at up to several 100 kW/cm^2 laser intensities. It has been assumed that a conventional dye molecule emits about 10^5 to 10^6 photons before being irreversibly bleached [28] which corresponds to a maximum of several seconds of continuous excitation/emission for intensities applied with most single-molecule based techniques. Due to the high temporal resolution of modern optical methods, ideal dyes should also show a weak tendency for fast intra-molecular photophysical transitions to dark states, specified by singlet-to-triplet state

quantum yields and on-off blinking phenomena induced by pH or excitation light. Both, protein fluorophores and synthetic dyes were found to meet these requirements for single-molecule applications.

3.1 Protein Fluorophores Instead of Synthetic Dyes?

The selection of a suitable dye system crucially depends on the particle to be labelled (e.g. lipids, DNA, RNA, proteins) and whether in vitro or in vivo systems are to be examined. Both, chemical labelling by covalently linking small and activated organic dyes to biological molecules, and genetic labelling with fluorescent proteins in protein fusion constructs have been established as standard methods. Among the dyes that can be covalently linked to biological molecules and have been successfully applied in single-molecule based applications are rhodamine, cyanine, Bodipy and Alexa dyes. The expected chemical environment in a given application further specifies a dye's suitability, since either hydrophobic or hydrophilic dye properties have to be taken into account [29]. Moreover, not all dye molecules are equally suited for two-photon excitation. On the other hand, interactions of labelled molecules with their conjugated dyes can affect a dye's photophysical behaviour and have to be taken into account for assay design [30, 31]. Remarkably many synthetic dyes exhibit ~10 times slower bleaching kinetics as reported for protein fluorophores [32]. In contrast, a major disadvantage of applying synthetic dyes on biological molecules, particularly proteins, is the difficulty of labelling at a certain site and at a fixed molar ratio. Moreover, if to be applied in vivo, the labelled molecule has to be delivered into cells by microinjection, loading reagents, or electroporation, all of which introduce additional experimental handicaps and pitfalls.

Genetically encoded reporter constructs based on fluorescent proteins circumvent many of these problems. Proteins that may be hard to purify can be labelled and expressed intracellularly, and fusions readily be targeted to sub-cellular compartments. In addition, fluorescent proteins like EGFP are more or less inert in the cytosol and very rarely cause photodynamic toxicity [13, 14, 29]. Taken together, these advantages render protein fluorophores superior to synthetic dyes for most intracellular applications. The number of successfully applied fluorescent reporter constructs has significantly grown over the last years [12] (see Miyawaki within this issue). New fluorophores have been developed applying mutagensis strategies to known species, and have also been newly discovered in reef organisms. Naturally, not all fluorophores known to date are equally suited for single-molecule based approaches, and display common and individual advantages and limitations that will be discussed in more detail in this review. Recently, Zhang et al. and Miyawaki reviewed current trends and developments in finding fluorescent probes for cell biological research, addressing a large body of protein fluorophores, and practical details for their application [12, 13].

3.2 Fluorescent Proteins

With regard to future FCS applications based on fluorescent proteins, three questions have to be raised. First, which fluorophores have been successfully used for FCS so far? Second, which were their features? And third, which are interesting new candidate proteins for FCS analysis with appropriate photophysical and biochemical characteristics? In order to address these questions, a short overview of fluorescent proteins based on Table 1 will first be given.

3.2.1 GFP and its Mutants

Until recently, the green fluorescent protein (avGFP) from the jellyfish *Aequorea victoria* was the only fluorescent protein to be widely mutated and altered for biological applications. Meanwhile, a family of GFP-like proteins from different organisms has emerged that share the common β-barrel fold structure and intrinsic chromophores but represent a vast spectral range, with (E)GFP and DsRed being the most prominent representatives [33].

After wild-type avGFP had been optimised for proper folding at 37 °C and efficient expression within cells, efforts were directed to the design of spectral variants to allow for multi-colour applications of fluorescent proteins [34, 35]. When the protein has acquired its β-can fold, chromophore formation occurs autocatalytically by cyclization and dehydration of Ser-65, Tyr-66 and Gly-67 to a colourless and non-fluorescent imidazolone ring followed by oxidation to an extended conjugated system giving rise to fluorescence. Mutations have been inserted in and around the chromophore to alter spectral properties, being grouped into 7 classes [34]. Depending on its composition and neighbouring amino acids, the chromophore predominantly exists in either a protonated or deprotonated form, e.g. with a Tyr-66 phenol group or a phenolate, respectively. Although effectively shielded by the β-can, external pH shifts can lead to reversible chromophore protonation/deprotonation in most GFP variants, and shift their spectral properties.

As shown in Table 1, spectral variants with blue, cyan and yellow-green emission characteristics have been developed out of wtGFP and are now commercially available. Continuing efforts are made to improve fluorophore brightness and decrease their sensitivity against acidosis and Cl^- concentration, as with the third generation of YFP mutants, 'Citrine' and 'Venus', or the newly presented cyan-green CGFP [13, 36, 37].

Since GFP mutants emitting blue light have to be excited with near-UV lasers, they are less favourable dyes for sensitive intracellular applications due to the significant interference of UV radiation with cellular components. Nevertheless, advanced spectroscopic concepts such as three-photon excitation might shed new interest on blue fluorescent proteins, e.g. for combined two- and three-pho-

Table 1 A selection of fluorescent proteins and their properties

Name	Excitation/ emission maxima (nm)	Extinction coefficient ($M^{-1}cm^{-1}$)	Quantum yield (%)	Biophysical applications	Comments	Color	References
EBFP[a]	383/445	31,000	25	– No FCS reports – Low photostability	– Dim dye – Candidate protein for three-photon excitation	Blue	[107]
ECFP[a]	434/477	26,000	40	– No FCS reports – Tested less suitable for single-molecule microscopy	– Low photostability	Blue	[32, 107]
T-Sapphire	399/511	44,000	60	– No FCS reports – Largest Stokes shift reported for protein fluorophore – No pH-sensitivity	– New variant of the 'Sapphire' GFP mutant – Candidate protein for three-photon excitation – Fluorescence process involves an internal proton transfer	Green	[52]
EGFP[a]	489/508	55,000	60	– FCS – Single-molecule microspopy – (pH-dependent) blinking	– Most widely applied label due to its brightness and photostability	Green	[107]

Table 1 (continued)

Name	Excitation/ emission maxima (nm)	Extinction coefficient ($M^{-1}cm^{-1}$)	Quantum yield (%)	Biophysical applications	Comments	Color	References
MAG	492/505	41,800	80		– Reported monomeric mutant of AG – Native AG is a tetramer	Green	[53]
EYFP[a]	514/527	84,000	61	– FCS – Single-molecule microspopy – (pH- and intensity-dependent) blinking	– 'Citrine' and 'Venus' are improved versions with regard to photostability, reduced pH- and chloride-sensitivity	Green-yellow	[107]
PAGFP	505/525	–	–		– Photo-activable (413 nm) mutant of EYFP	Green-yellow	[50]
DsRed= drFP583	558/583	72,500[b]	68[b]	– FCS – Intensity-dependent blinking – No pH sensitivity – remarkably photostable in FCS	– Commercially available as DsRed1[a] – >20–30 h maturation time[b] – Obligate tetramer and aggregation – DsRed2[a]:threefold maturation speed, 68% brightness relative to DsRed1 – Aggregation tendency can be reduced	Orange-red	[11, 32, 39]

Table 1 (continued)

Name	Excitation/ emission maxima (nm)	Extinction coefficient ($M^{-1}cm^{-1}$)	Quantum yield (%)	Biophysical applications	Comments	Color	References
T1	554/586	30,100	42	– 36% brightness relative to DsRed1	– DsRed- Express[a] – Maturation half-time 0.7 h – Tetramer like DsRed1	Orange-red	[46]
tdimer2(12) (tandem)[c]	552/579	120,000	68	– FCS (see this review, intensity-dependent blinking)	– DsRed-mutant – Maturation half-time 2 h – Improved brightness	Orange-red	[42]
mRFP1 (monomer)	584/607	44,000	25	– FCS (see this review, intensity-dependent blinking) – Decreased photostability	– Monomeric, dim DsRed-mutant – Maturation half-time <1 h – Red-shifted compared to DsRed	Red	[42]
HcRed1[a] (dimer)	592/645	–	–	– FCS (see this review, intensity-dependent blinking)	– t-HcRed1 similarly bright as mRFP1 – Source for optimized far-red proteins	Red	[43, 108]
t-HcRed1 (tandem)	590/637	160,000	4				

Table 1 (continued)

Name	Excitation/ emission maxima (nm)	Extinction coefficient ($M^{-1}cm^{-1}$)	Quantum yield (%)	Biophysical applications	Comments	Color	References
EqFP611	559/611	78,000	45	– FCS – Intensity-dependent blinking – Large Stokes shift	– Tetramer, dissociates at high dilution – Intracellular maturation: ~6 h at 28 °C; none at 37 °C	Red	[78]
KAEDE	508/518 572/582	98,800 60,4000	80 33	– Emission switch from green to red after UV illumination	– Tetramer – Quantum yield is pH resistant	Green→ red	[51]

EBFP: Enhanced Blue Fluorescent Protein; ECFP: Enhanced Cyan Fluorescent Protein; EGFP: Enhanced Green Fluorescent Protein; EYFP: Enhanced Yellow Fluorescent Protein.

[a] Available from Clontech, Palo Alto, USA.

[b] Reported values show large discrepancies probably due to different protein purity and maturation (Baird 2000; Matz 1999; Patterson 2002; Campbell 2002).

[c] 'Tandem' indicates that two monomers have been fused to produce a one peptide chain dimer, while 'dimer' or 'tetramer' indicates that the protein is an obligate oligomer.

ton excitation in multicolour applications. So far, no mutant of Aequorea GFP exhibits emission maxima longer than 529 nm although there is a tremendous need for red-shifted fluorescent indicators also in single-molecule based approaches: Multi-colour applications require efficient spectral separation of fluorescence from different maximally bright fluorophores, and longer wavelength excitation/emission properties of the red-shifted dyes move them further away from the spectral regime where autofluorescence is present.

3.2.2 Red Fluorescent Proteins (RFPs)

The discovery of a naturally red fluorescent protein in *Discosoma sp*. named DsRed (drFP583) suddenly made red-shifted protein tags available, and started analogous intense efforts to optimise DsRed [38]. While only sparse dimerization of wtGFP could be observed and easily be eliminated DsRed is an obligate tetramer [39, 40]. Therefore, DsRed is likely to tetramerize any of its fusion partners and also shows the tendency to form aggregates possibly conferring cellular toxicity [13]. To eliminate this problem, both non-aggregating tetrameric mutants and also a dim monomeric version of DsRed, named mRFP1 (Monomeric Red Fluorescent protein 1) have been generated, reducing basic-acidic surface interactions between monomers [41, 42]. Also, dimerization of monomers in tandem fusions has been presented as a strategy to reduce and suppress oligomerization and aggregation [42, 43].

As a further drawback, native DsRed also exhibits slow and complex fluorophore maturation due to an additional autocatalytic modification, extending the chromophore's conjugation system and allowing for red fluorescence [44, 45]. Non-mature protein with 475-nm excitation/500-nm emission maxima transforms into mature protein with 558-nm excitation/585-nm emission maxima and requires >48 h to reach 90% of maximal fluorescence [39]. Fluorophore maturation has been significantly accelerated in a number of mutants, e.g. T1, mRFP1 and E57 and was proposed to depend upon the space around the fluorophore [46, 47].

Unfortunately, no monomeric bright red fluorescent protein with photophysical properties similar to GFP mutants or the tetrameric red proteins has been reported to date. Applying RFPs thus still requests one of three possible compromises: 1) using bright obligate tetramers with accelerated maturation and easily-reduced tendency for higher aggregation, 2) applying comparatively dark monomers, 3) applying RFP tandem fusions.

Apparently, strategies developed to abolish tetramerization and enhance brightness will have to be combined to yield the ultimate red protein tag, since all Anthozoan GFP-like proteins known to date seem to form obligate tetramers [33]. The presented drawbacks of red fluorescent proteins do not categorically preclude their use in intracellular applications and single-molecule based technologies, as shall be discussed in this review. Moreover two indirect strategies have been proposed to improve DsRed applications: pulsed expression of DsRed

enriched mature DsRed within cells and reduced spectral ambiguities, while co-expression of dark DsRed mutants titrated bright DsRed tags and reduced oligomerisation of tagged proteins [48, 49].

Two additional fluorescent proteins have to be mentioned in this section, because they provide exciting features for cell biology and also insight into the complexity of chromophore maturation. Patterson and Lippinicott-Schwartz [50] recently presented a photo-activatable YFP mutant (PA-GFP) that was found to increase fluorescence 100-fold after intense irradiation with 413 nm. Even more impressing, a green fluorescing protein from the stony coral *Trachyphyllia geoffroyi* had been cloned that emits bright green fluorescence after synthesis but switches to bright and stable red fluorescence upon UV illumination [51]. The protein was termed "Kaede" and unfortunately also forms a tetrameric complex.

3.2.3 Alternative Fluorescent Probes

Recent developments of alternative fluorescent probes have been reviewed by Zhang et al. and will be mentioned briefly in this review, pointing towards major promises and drawbacks for biological applications [13]. Griffin et al. reported specific covalent labelling of recombinant protein molecules inside live cells based on the formation of covalent complexes between a biarsenical dye (e.g. FlAsH, ReAsH) and a tetra-cysteine motif fused to proteins [54]. The method combines the advantages of site-specificity, by genetic introduction of the tetra-cystein motif, and the usage small synthetic dyes of different fluorescent emission [55]. Nevertheless, significant background staining and the necessity of delivery protocols for biarsenic dyes impose significant disadvantages. A different approach to design new fluorophores was presented by Bae et al. [56]: the fluorescence properties of a GFP mutant were changed by 'selective pressure incorporation' (SPI) of an amino acid analogue within the GFP chromophore in auxotrophic *E. coli* host strains. So far, the method is extremely difficult to apply within mammalian cells and only produced very wide emission spectra.

3.3 Applying Fluorescent Proteins as Tags

The application of fluorescent proteins as tags raises various questions about their general suitability as fusion partners. Its hard to predict whether tagging with a globular ~30 kDa domain will affect the function and native intracellular distribution of the protein of interest, and by which means, e.g. optimised intervening protein linker sequences, artificial effects can be minimized. Classical biochemical approaches like two-hybrid systems for mapping molecular interactions of random fusion proteins succeeded in spite of the apparent sterical limitations. Considering the enormous number of tagged proteins in

sensitive FRET studies, starting with calmodulin and more recently applying protein kinase B (Miyawaki et al. 1997; Calleja et al. 2003) proves that essential protein characteristics might not even be affected by tagging with two protein fluorophores [57, 58]. On the other hand, obligate fluorophore oligomerization like DsRed tetramerization results in untypical molecular constellations. This can affect the function of a tagged protein or be disadvantageous for the optical technique to be applied. Consequently, researchers have to provide evidences for proper protein function in both a molecular and cellular context, at best by gene replacement and reconstitution or alternative less elaborate approaches [59]. The question of how to fuse fluorescent proteins to proteins of biochemical interest has been addressed by several publications and proved to be worth considering [13, 60–63]. Apart from standard N- or C-terminal tagging, protein fluorophores were successfully inserted into other proteins of interest, and vice versa. Both fusion partners retained their native function or purposely showed altered fluorescence properties. Only weak fluorescence quenching of protein fluorophores attached to other proteins has been reported, e.g. 80% average brightness values of in vivo EGFP for an adenylate kinase-EGFP fusion protein [15]. This may result from steric and chemical protection of the chromophore within a rigid β-barrel. Moreover, intracellular expression systems can be tuned to produce fixed and reproducible ratios of different protein fusions which is highly advisable for dual colour applications. While the fluorophore ratio is fixed when expressing bi-coloured fusion-proteins, simultaneous expression of two separate fusions can be facilitated using internal ribosome entry sites in bicistronic gene arrangements [64]. Intracellular expression of fusion proteins also allows for fine tuning expression levels within the concentration range of 10 nmol/l to 1 μmol/l typically used for intracellular FCS analysis. Particularly low expression levels that mimic physiological conditions have been attempted applying linearized plasmids in transient expression [15] and inducible ecdyson or tetA based expression systems [65, 66].

4
Properties of Fluorescent Proteins Analysed by FCS

4.1
Photodynamic Processes Affect the Applied Method

Fluorescent proteins are nowadays ubiquitous in microscopic experiments, usually being imaged in fusion to cellular proteins at different levels of spatial and temporal resolution. Quantitative standard biological assays have been performed at the bulk level [67] depending mainly on the spectroscopic properties of autofluorescent proteins. These approaches have been addressing protein localization, gene expression, trafficking, sensitivity to local environment such as pH, and issues amendable to FRET (Fluorescence Resonance Energy Transfer), like Ca^{2+} sensing or protease reactions [57, 68]. Fluorescence Lifetime

Imaging Microscopy (FLIM) based approaches have presented intracellular FRET studies for conformational changes and protein-protein interactions and can be performed with high temporal resolution at the bulk level [58, 69]. In parallel, fluorescence-based imaging has increasingly been applied to the direct observation of single molecules to address their intra-molecular dynamics along with translational mobility both in vitro and in vivo [70–72]. Cellular autofluorescence decreasing S/N ratios and the choice of appropriate dyes showed to be critical to these studies [32]. Nearly all protein fluorophores analysed at the single molecule level unveiled unexpected properties such as reversible photobleaching and on/off blinking in microseconds to seconds time ranges [8, 73, 74]. These characteristics have important implications for studying single-molecule trajectories or intramolecular dynamics such as conformational changes.

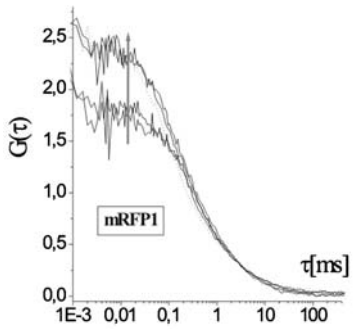
In contrast to monitoring one molecule at a time in single-molecule microscopy, fluorescence correlation spectroscopy (FCS) combines single molecule sensitivity with averaging over a larger number of dyes, continuously diffusing in and out of the detection volume. This approach to single molecule analysis proved capable of efficiently discriminating intramolecular photodynamics from translational dynamics in a highly quantitative manner and is less sensitive to the long-living dark states reported for GFP [8, 10, 11, 73, 74]. Recent publications on FCS have documented its capability of tolerating autofluorescence and blinking phenomena, and their quantitative inclusion in data evaluation [6, 8, 9, 11].

4.2 EGFP, YFPs and DsRed analysed by FCS

Thorough photophysical FCS analyses have been carried out on various fluorescent proteins so far: EGFP, YFPs, DsRed. These studies produced valuable information for fluorophore usage as fusion partners and successive data evaluation. They also document which photophysical parameters of reporter molecules ought to be determined in vitro to reduce the number of unknown parameters in vivo. We briefly outline these fluorophore studies and continue to hint to implied options and limitations.

4.2.1 EGFP

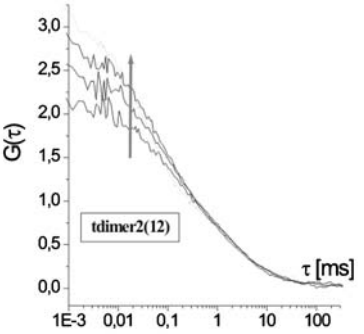
EGFP differs from wtGFP in 4 amino acids, of which F64L improves folding at 37 °C and S65T was introduced to favour the deprotonated (phenolate) bright form of the chromophore, formed autocatalytically from Ser-65, Tyr-66 and Gly-67 [34]. FCS analysis after excitation with 488 nm unveiled that the EGFP chromophore shows a complex protonation equilibrium between 3 chromophore species, 2 protonated dark and one deprotonated bright state [8]. Due to characteristic (de-) protonation time constants, the chromophore exhibits



light intensity η *[kHz]* τ_d

5kW/cm²	2.9 cpm	1.4ms
4kW/cm²	2.1 cpm	1.3ms
2kW/cm²	1.9 cpm	1.2ms
1kW/cm²	1.1 cpm	1.5ms
7kW/cm²	3.8 cpm	1ms (dotted line)

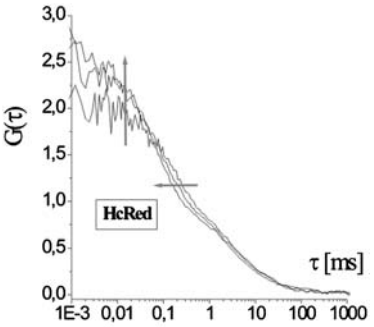
A



light intensity η *[kHz]* τ_d

13kW/cm²	8 cpm	2.0ms
8kW/cm²	3 cpm	2.7ms
5kW/cm²	3 cpm	2.0 ms
26kW/cm²	6 cpm	2.3ms (dotted line)

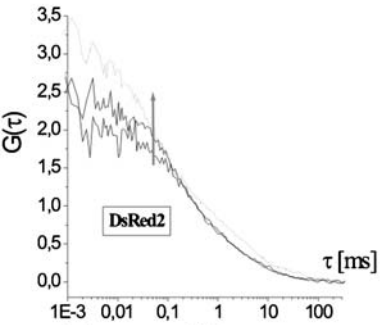
B



light intensity η *[kHz]* τ_d

10kW/cm²	3.5cpm	3ms
7kW/cm²	2.1cpm	3ms
5kW/cm²	1.9cpm	4ms

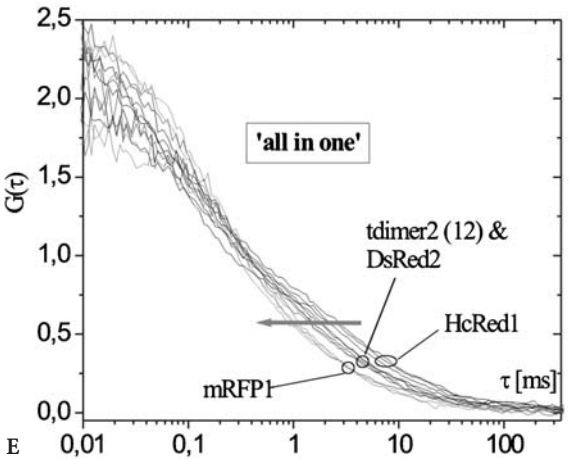
C



light intensity η *[kHz]* τ_d

8 kW/cm²	1.9cpm	3.4ms
5 kW/cm²	1.6cpm	2.3ms
16 kW/cm²	4.5cpm	4.0ms (dotted line)

D



E

blinking that leads to specific contributions to EGFPs autocorrelation curves as shown in simulated curves in Fig. 1. The blinking relaxation times have been called chemical decay factor τ_C and increase from 50 µs to 450 µs upon a shift from pH 5 to 7 due to proton exchange with the solvent. The dark fraction F_C increases for low pH values (≈80% at pH 5). From pH 8 to 11 a ~13% fraction of the fluorophore shows pH independent blinking with τ_B≈450 µs that is due to internal protonation and slightly sensitive to excitation power. Apparently, a fitting function including only one exponential decay term was sufficient to describe these blinking phenomena within the tested pH range. With higher laser powers EGFP correlation curves showed an additional contribution (≈30 µs) that was assigned to triplet state population.

Interestingly, a Y66W GFP mutant lacking a protonable hydroxyl on the chromophore was examined by Haupts et al. (1999) and no pH-dependent fast fluorescence blinking detected [35]. In reverse, pH-sensitive GFPs can be applied as highly accurate pH sensors for in vivo applications read out by imaging methods or FCS [8, 75]. FCS applications in the cytosol take place at pH~7 and thus face little pH dependent blinking [76]. It is advisable to determine blinking parameters for the expected pH range before the evaluation of intracellular measurements. Also, application of GFPs in cases such as cellular acidosis, e.g. within hippocampal neurons, requires consideration of blinking phenomena and might be simplified by usage of pH-resistant mutants like "Sapphire" unless they have to be applied as pH sensors [34, 52].

4.2.2 YFPs

In analogous studies, the yellow-shifted GFP mutants S65G/S72A/T203Y, S65G/S72A/T203F and Citrine showed both, pH-dependent and excitation intensity-dependent blinking [9, 10]. The flickering rate increased with excitation

Fig. 3A–E Comparison of RFPs in the cytosol of Hek293 cells. Autocorrelation curves were recorded at different excitation intensities, normalized and plotted for: **A** mRFP1; **B** tdimer2(12); **C** HcRed1; **D** DsRed2; **E** superimposed to allow for visual comparison. Diffusion properties (τ_d in 'ms') and particle numbers (used for normalization to N=1) could be determined and revealed molecular brightness values (η given in 'counts per molecule' [kHz]). *Dotted lines* (A, B, D) represent atypical diffusion curves attributable to bleaching. According to our interpretation, dark fractions increased for all tested RFPs from ~40% to ~60% (indicated by *vertical arrows*) with growing intensities (543 nm) that were determined before entering the microscope. Diffusion curves (E) suggested significant differences in particle mobility (indicated by *horizontal arrows*, E). While mRFP1 seemed to show the highest intracellular mobility (E, B) tdimer2(12) performed as the brightest dye (A) in our measurements. In contrast, mRFP1 (B), DsRed2 (C) and HcRed1 (D) showed photobleaching even at fairly low intensities. DsRed2 (C) and HcRed1 displayed a slow second diffusing component that made curve fitting rather difficult and favoured bleaching. This comparison of dynamic RFP-behaviour within cells fits well with their biochemical characterization and predicted behaviour

intensity, accordingly the decay time decreased with τ_B<100 µs, showing a constant dark fraction of F≈60%. A second dark state (10–100 µs) resulting from chromophore protonation could be observed below pH 8. Thus YFPs required two exponential decay terms with τ_C and τ_B to fit data of the low to moderate pH measurements.

4.2.3 DsRed

DsRed displays both chemically and photophysically remarkable features. In FCS, its strong light-dependent blinking seems to involve three different fluorophore states, at least two of which are fluorescent [10, 11]. Two characteristic fluctuation time scales $\tau_{D(1)}$ and $\tau_{D(2)}$ reaching from several tens to several hundreds of microseconds with intensity dependent dark fractions D_1 and D_2 were assigned to conformational rearrangements of the chromophore. Moreover, the fluorescence spectrum of DsRed red-shifts ≈8 nm after increasing excitation intensity. In contrast to most GFP mutants, the DsRed fluorophore doesn't show any pH-dependent blinking for pH 6 to 11. Upon two-photon excitation, the light-dependent flickering was not detected and DsRed seemed to be remarkably photostable. Obligate tetramerization of DsRed and formation of higher order aggregates could be detected by FCS analysis even at low nanomolar concentrations since ~1.5- and ~2.4-fold lower diffusion coefficients (~4- and 13-fold larger MW) compared to monomeric GFP have been determined for purified DsRed. The nature of fluorescence emission of DsRed has also been addressed in single-molecule analysis with important implications for FCS [77]. Monomers within a tetramer were proposed to transfer energy between each other. In consequence, emission results with equal probability from each of the four chromophores. Thus, an excitonic mode of excitation transfer seems to be active, while independent emission from individual monomers can be excluded.

86% of DsRed tetramers were found to contain at least one immature green monomer. Consequently, DsRed tetramers present an inhomogeneous population of dye molecules that show high sensitivity to excitation intensities due to complex energy transfer processes among the monomers. The resulting complexity and heterogeneity of emission behaviour implies that FCS analysis offers a rather phenomenological description of DsRed fluorescence and is therefore well suited to analyse particle numbers and mobilities. No similar blinking phenomena as reported for DsRed has been described for monomeric GFP mutants.

A new red protein eqFP611 from *Entacmaea quadricolor* displayed both monomeric and tetrameric features in initial studies [78]. FCS analysis at low nmol/l concentrations predicted a mass ratio of 2.7 between DsRed and eqFP611 and produced a diffusion coefficient similar to the monomeric GFP mutant Citrine. In contrast, size-exclusion chromatography and the recently resolved crystal structure of eqFP611 indicated that eqFP611 too can form tetramers

based on extensive monomer surface interactions [79]. Interestingly, six mutated residues in the monomeric DsRed variant mRFP1 can also be observed at the corresponding positions of eqFP611. Thus eqFP611 seems to be a promising substrate for mutational optimisation of its features, e.g. shifting its maturation optima from 24.5 to 37 °C and decreasing its oligomerization tendency.

4.3 On-Off Blinking of Protein Fluorophores in FCS

Because dark fractions of blinking protein fluorophores appear in steady state fluorescence measurements as a reduced quantum yield and efficiency, they can reduce a fluorophore's potential as a fluorescent marker. In part, this limitation can be counterbalanced by time-resolved FCS. If the blinking decay time τ_B gets however too close to the diffusional decay times τ_d or even larger, the number of fluorophores present in V_{eff} could be significantly underestimated in the case of large dark fractions. This effect would also decrease the maximum cross-correlation amplitude to be expected in dual colour cross-correlation analysis. Thus, blinking can set a lower limit for V_{eff} and resulting diffusion times, unless underestimation of particle numbers can be tolerated. To separate blinking from diffusional fluctuations, different sizes of V_{eff} can be readily compared. The application of TPE seemingly eliminates intensity-dependent blinking in most fluorescent proteins, due to different selection rules. Here also, intensity dependent photoconversion to long-living non-emitting states can also reduce apparent diffusion times analogously to photobleaching, and can easily be detected applying different intensities.

In summary, FCS can quantitatively account for blinking phenomena that proved to be common to frequently used fluorophores. Interestingly, the fraction of intensity dependent blinking seems to increase with red-shifted emission of protein fluorophores. It is tempting to speculate whether this might be due to the extended π-electron systems present in red fluorescent dyes.

4.4 Fluorophore Oligomerization, Aggregation and Dim Fluorophores

Fluorophore aggregation can occur by self-aggregation of fluorophores (see above), but also in fusions to oligomer-forming partners, and affects FCS data evaluation. Oligomerization of fluorescently tagged proteins leads to a distribution of molecular brightness that has to be considered for proper determination of particle numbers [16, 21]. There exist several methods of choice to examine the accurate distribution of monomeric vs oligomeric molecules that can be applied on the same data sets used for FCS analysis: Fluorescence Intensity Distribution Analysis (FIDA) and Photon Counting Histogram (PCH) analysis both consider intensity distributions in fluctuation traces [80, 81]. Fluorescence Intensity Multiple Distributions Analysis (FIMDA) combines fea-

tures of FCS and FIDA, allowing for simultaneous determination of molecular brightness and diffusion times [82]. Besides that, an FCS based approach to analyse polydispersity has been presented and higher-order autocorrelation introduced [83, 84]. Singular bright particles that present a significant or dominating contribution to the correlation curve usually have to be excluded from FCS analysis by running short repetitive measurement intervals to select and average only traces devoid of aggregates [22].

Mutant fluorescent proteins optimised for shifted fluorescence excitation/emission properties or fast maturation and weak oligomerization (DsRed mutants) often display very low molecular brightness values due to photobleaching or small quantum yields. Figure 3 presents an initial attempt to measure and compare autocorrelation curves of DsRed2 and its mutants tdimer2(12) and mRFP1 as well as HcRed1. The application of comparatively dark red fluorophores in vivo takes advantage of the decreased cellular autofluorescence in the red compared to the green emission range (see below). It requires an adaptation of the experimental setup, like applying TPE or varying the size of the laser focal spot, both affecting local molecular brightness and 'signal to noise' ratios (see discussion in this review).

4.5 Photobleaching and Photostability

Sufficient photostability of dyes is an important prerequisite for FCS studies in general and particularly inside live cells. Photobleaching affects both auto- and crosscorrelation amplitudes and introduces a significant artefact risk in FCS analysis [16, 21]. Moreover, cells of several picoliters volume will contain only 10^5 to 10^6 dye molecules at dye concentrations of 10–100 nmol/l and thus are readily subjected to dye depletion. As reviewed here, EGFP, various YFPs and RFPs have been found to work well with FCS applications displaying sufficient brightness and photostability. Photophysical properties of autofluorescent proteins have been compared for ECFP, EGFP, EYFP and DsRed [32]. ECFP seems less suited for FCS due to its low brightness and high susceptibility to photobleaching, while EYFP was shown to be equally suited as EGFP. Despite its strong excitation-dependent blinking, DsRed was found to exhibit very little photobleaching in FCS analysis within an excitation intensity range of 0.5–100 kW/cm^2 and long measurement intervals [11]. In contrast, low photostability of DsRed and DsRed-mutants has been reported, although Harms et al. indicated that dark states complicate the interpretation of photobleaching data [32, 42]. In addition, ECFP, EGFP, EYFP and DsRed were shown to be excitable by two-photon excitation [85].

4.6
Two-Photon Excitation of Fluorescent Proteins

Two photon excitation (TPE) and its application to FCS analysis has been described and discussed in detail and shall be briefly reviewed with regard to fluorescent proteins and intracellular usage [6, 7, 86, 87].

Two-photon excitation requires quasisimultaneous (10^{-15} s) absorption of two photons of approximately double the wavelength as usual and thus is limited to a small self-confined volume element displaying sufficiently high photon flux densities. High photon flux rates are usually provided by employing pulsed laser systems. TPE confers particular advantages to intracellular applications, since both cell damage and sample depletion by photobleaching are significantly reduced and limited to a small focal spot compared to one-photon excitation taking place throughout the created light cone. Consequently, TPE was applied for long-term data acquisition of TMR in live cells without significant loss of fluorescence signal [6]. TPE of the protein fluorophores ECFP, EYFP, EGFP and DsRed has been examined and excitation cross-spectra have been recorded, while TPE FCS has been performed on rsGFP, EGFP and DsRed [6, 14, 85, 88]. Intersystem crossing to intensity-dependent dark state is typically much less likely since different photophysical transitions are involved in TPE.

Due to these different selection rules for excited fluorophores, TPE spectra of different dyes can show significant overlap in contrast to their one-photon excitation spectra. Consequently, both synthetic and protein fluorophores with largely different emission have been efficiently excited with a single IR laser line in TPE based cross-correlation studies [7, 17, 88]. Analogous three-photon excitation with IR lasers requiring ~10-fold higher intensity densities allows for exciting UV-absorbing dyes and might be used in combination with TPE [89].

5
Single Colour Applications of Fluorescent Proteins In Vivo

The number of intracellular FCS applications based on fluorescent proteins is constantly increasing. Single colour applications can provide data about the localization and concentrations of fluorophores and their fusions, intramolecular photo-dynamics and particle mobility.

5.1
Concentration Measurements

FCS allows for highly precise and locally confined determinations of particle concentrations, that help to monitor and correlate genetic and physiological processes within cellular compartments, as was demonstrated within *E. coli*, the only reported FCS application within bacteria to our knowledge [90].

5.2 A Sensor for Environmental Conditions

With regard to its pH sensitivity, GFP in combination with FCS is considered an intracellular pH sensor [8–10]. Internal calibration even within cells can be performed by examining both particle numbers and photophysical dynamics in parallel, and pH-sensitive protein fluorophores can efficiently be targeted to different cellular compartments [75]. Another interesting approach, though not necessarily depending upon FCS, consists in the insertion of allosteric sites into protein fluorophores that influence fluorescence properties upon ligand binding [13, 63].

5.3 Mobility as a Precise Measure of Particle Action

Determination and comparison of particle mobility can shed light on all events affecting particle motion. This can account for particle interactions such as protein-protein interactions and ligand binding, particle partitioning into different cellular compartments or active transport phenomena.

5.3.1 Active Transport

Depending on the characteristic shape of autocorrelation curves, active transport of GFP in vivo could be distinguished from diffusion within plastid tubules of plant cells and its properties described in more detail [91]. Bright GFP batches and not single molecules were actively transported in a plug flow mode, while laminar flow could be excluded. The approach of this study can be applied to any other elongated narrow structure, like dendrites. Even more powerful, FCS strategies presented for the analysis of fluorescent particles in microstructures, like applying two laser foci and examining spatial cross-correlation, can as well be transferred to organic structures [92].

5.3.2 Diffusion: A Comparison off In Vitro and In Vivo Applications

Assessing changes in particle size on the basis of diffusion times derived from FCS measurements has been successfully accomplished in a number of in vitro binding experiments and was initially applied to study quantitative hybridization kinetics of DNA probes to RNA [93, 94]. Moreover, a detailed and systematic examination of autocorrelation analysis of ligand-protein binding equilibria has been presented, and the handling of brightness distributions arising from oligomerization of labelled monomers has been discussed [81, 83, 95].

However, autocorrelation analysis offers a limited resolution for two-component solutions (e.g. free ligand, bound and slower ligand) since the diffusion

time scales only with the third root of the molecular weight of spherical particles. Thus, large mass changes result in comparatively small changes in τ_{Diff} and D_i. Meseth et al. determined a resolution limit at a minimum ratio of 1.6 in diffusion times τ_{Diff} for equally bright particles, which corresponds to a factor of 4 in molecular mass [96]. For less ideal conditions like different dye-brightness, a much bigger ratio for τ_{Diff} is required for proper separation of two components. With regard to the inhomogeneity of the intracellular environment (see below for discussion) and the stringent requirements described above, this kind of analysis, though very powerful for binding studies in vitro, won't be applicable to most in vivo situations.

Nevertheless, differences in mobilities of EGFP tagged proteins can be readily exploited to analyse partitioning in different cellular compartments like the cytosol and the plasma membrane, and to document changes in cellular localization [15, 97]. Molecules in the cytosol and the plasma membrane differ in their mode of diffusion (3D- and 2D-diffusion, respectively), and the degree of anomalous diffusion that has to be assumed to fit the obtained data. This experimental approach nicely tracks changes in localization and distribution of proteins involved in signalling from the plasma membrane to the cytosol and nucleus, and thus presents an efficient tool for analogous investigations.

In order to perform rigorous mobility analysis for in vivo situations, cell-type specific intracellular viscosities (see below) and a wide distribution of mobilities within the same compartment of a given cell have to be taken into account. Mobility analysis of EGFP in the cytosol of Hek293 and Hela cells resulted in quite similar values for $D_{EGFP} \approx 2.3–2.5 \times 10^{-7}$ cm^2/s but a lower value $D_{EGFP} \approx 1.3 \times 10^{-7}$ cm^2/s within AT-1 and Cos-7 cells [14, 15, 25, 97]; for comparison $D_{EPFP} \approx 8.7 \times 10^{-7}$ cm^2/s was determined in water. Thus, a comparison of fusion protein mobilities derived from different cell lines requires normalization to a reference control such as EGFP. Moreover, values for D_{EGFP} in the cytosol showed standard deviations of at least 10% for different measurement positions that have to be assigned to interactions with organelles and unknown particles, and make proper statistical approaches to mobility based in vivo studies obligatory. The exact position of the measurement volume V_{eff} within cells, in particular the proximity of membranes, can affect obtained mobility values. Nevertheless, extensive PCH and FCS analysis by Chen et al. excluded disturbing and detectable interactions of EGFP with intracellular compounds [14]. Comparing mobilities of EGFP and EGFP-fusions in the cytosol yielded intriguing implications about fusion protein behaviour: while some constructs showed the theoretically predicted difference in D_is according to molecular weight ratios another fusion protein moved much slower than predicted and showed a significantly larger mobility standard deviation [15, 25, 97]. This may account for a non-spherical conformation, or more interesting, for unexpected and specific interactions that are of particular interest for proteins involved in signalling. On the other hand, finding predicted D_is based on molecular weight ratios to a reference fluorophore confirms the absence of specific interactions and thus presents a nice internal mobility control.

Interestingly, diffusion behaviour in the nucleus was found to depend on the cell cycle: in metaphase nuclei it was comparable to the cytosol, while showing significantly different mobility properties in interphase nuclei [25].

Experimental data from in vivo measurements have been fitted both with multi-component models and assuming anomalous diffusion [6, 15, 25]. Certain applications like measuring in fragmented cell membranes clearly require anomalous diffusion models [24]. Applying either model automatically implies a slightly different explanation for the observed phenomena: for multi-component fits, the dye is supposed to partly interact and move together with slow particles. For anomalous diffusion, the dyes movement appears restricted by inert obstacles obstructing molecular motion by an excluded volume interaction which can even result in a confinement of the detection volume as within small cavities [25, 26]. Accordingly, Verkman recently reviewed solute and macromolecule diffusion in cellular compartments and described three factors accounting for slowed intracellular diffusion: 1) slowed diffusion in fluid-phase cytoplasm, 2) probe binding to intracellular components, and 3) probe collisions with intracellular components, called molecular crowding [98].

5.4 Practical and Theoretical Considerations for FCS on Protein Fluorophores In Vivo

5.4.1 Dealing with Background Fluorescence in Buffers and Cells

Several reports documented that FCS can accommodate fluorescent background quite well. This ranges from easy estimates based on negative controls to a formal mathematical treatment and inclusion into analytical formulas. Fluorescent background can be found within cells or buffers including fluorescent components such as Triton-X 100. Cellular autofluorescent molecules constitute a heterogeneous group including NADH, flavins and flavoproteins, collagen and elastin, lipofuscin and can be found throughout the cell. Its properties seem to depend on normal and transformed cellular conditions and metabolic activity [99].

Immobile or slow sources of autofluorescence will contribute an uncorrelated fluorescence signal and can easily be photobleached or corrected for mathematically [6]. The concentration of diffusible components contributing correlated fluorescent signals has been estimated to range between 0.3–15 nmol/l in the cytosol of mammalian cells and 23 nmol/l in the nucleus and also average molecular brightness values were determined in studies using different cell lines and modes of excitation [6, 14, 100]. Obviously, the collected cellular background is strongly dependent on chosen excitation/emission parameters, e.g. is a factor of 2 higher at 488 nm compared to 543 nm, and significantly lower for 2PE [6, 32]. Flavins in particular exhibit strong fluorescence emission within the detection range of EGFP and EYFP [32]. Blab et al. examined TPE cross spectra of flavin mononucleotide and concluded significantly smaller excita-

tion rates of flavin for TPE compared to one photon excitation [85]. As a rough estimate, Schwille et al. proposed to introduce background fluorescence into the autocorrelation function for fluorescence $F_{background}>0.1\ F_{total}$ and high molecular brightness $\eta_{lBackground}$ [6]. Strong background fluorescence otherwise could lead to an overestimation of the actual particle number and underestimation of molecular brightness values and also affect cross-correlation amplitudes. The relation of apparent and real particle numbers and brightness values can be described formally and help to correct data interpretation [14]. Particularly low nanomolar concentrations of reporter molecules in cells (<10 nmol/l) can require more elaborate data analysis, therefore intracellular concentrations for fluorescent proteins like EGFP should range within 10–100 nmol/l for convenient and reliable analysis [6]. Comparison of the molecular brightness of reporter molecules in vitro and in vivo which is linked to the apparent particle number can serve as a direct and easy measure for the impact of autofluorescence [6, 14]. When applying one photon excitation, the size of V_{eff} (primarily controlled by the pinhole diameter in the detection unit) has an important impact on background phenomena. With intensities well below the saturation threshold, the fluorescence collected from each dye molecule increases in larger detection volumes V_{eff}. Yet for one-photon excitation, the relative increase of the photon yield per detected molecule η in vivo is much smaller than in vitro, since the scattering background contribution is significantly favoured in larger volumes. Interestingly, nearly equal values of η in vitro and in vivo could be obtained after TPE for both synthetic dyes and EGFP which can be explained by the small excitation/detection volume collecting less background from autofluorescence [6, 14]. For one-photon excitation, a compromise between background contributions and molecular brightness needs to be found that allows for reliable data interpretation.

These initial 'signal to noise' considerations are particularly important with regard to the application of non-ideal dyes. They help to design appropriate negative controls and optimise the measurement system with regard to fluorophore excitation/emission maxima and dye concentrations. Theoretical brightness comparisons of different dyes can be readily accomplished for known absorption coefficients and quantum yields. Nevertheless, experimentally obtained molecular brightness values η depend dramatically on excitation, used emission filters, and V_{eff}. They can be low even for bright dyes if crosstalk has to be minimized in multicolour applications.

5.4.2 Choice of Cell Line

In the first place, the choice of a cell line depends on biological requirements of the system to be studied. However, it should be highlighted that with regard to intracellular particle mobility and the degree of autofluorescence, different cell lines can't necessarily be considered equal for optical methods with single molecule sensitivity. Slightly different compositions of the intracellular matrix

might severely affect autofluorescence and particle mobility, and thus the dye bleaching, due to their influence on average residence times. Comparing previous studies, the viscosity in the cytosol was found to be 2.6- to 10-fold higher than in aqueous solution [25]. The growing number of confocal applications allows for a comparison of highly accurate mobilities in so far used cell lines. Schwille et al. provide a very systematic study of particle brightness and mobility, including a comparison of cytosolic measurements with the dye CMTMR in Hela, EMT6 and RBL cells [6]. The investigated dye molecules display both different mobilities and partitioning to differently mobile fractions. Differences in EGFP mobility among different cell types have been addressed above. It can't be excluded that the standard deviation of mobilities within a single compartment also varies with different cell lines [6]. In summary, there are qualitative differences in viscosity and background fluorescence between different cell lines though these can only vaguely be quantified.

6
Dual-Colour Applications: Cross-Correlation Analysis

6.1
Different Applications of Cross-Correlation Analysis

The application of two spectrally distinct fluorophore tags within a given system allows for a much more detailed study of molecular interactions and molecule behaviour in general, than in one-colour experiments. Cross-correlation analysis presents a direct and precise measure for particle interactions. Moreover, simultaneously obtaining two autocorrelation functions provides data about particle numbers and mobility that present important internal reference and calibration parameters. Starting with the analysis of DNA hybridisation, the technique has been further evaluated in model systems addressing real-time enzyme kinetics and the exclusive application of protein fluorophores and has been successfully applied in vivo [3, 16, 17, 88, 101]. Studying events at the single protein level holds the biggest promise for in vivo applications. Degradative processes such as protease reactions will be typically described by decreasing cross-correlation amplitudes [88]. Specific and reversible protein-protein interactions can be illustrated by both decreasing and increasing cross-correlation amplitudes [17]. Cross-correlation analysis also can be applied to address phenomena of indirect particle interactions within complex aggregates and colocalization within small diffusing compartments, such as active transport or co-localization in endocytic vesicles smaller than V_{Eff} [16]. It seems that the case of decreasing cross-correlation amplitudes for degradation events, starting at a typical maximum amplitude, is somewhat easier to address, since maximum amplitudes of interactions at equilibrium are primarily unknown and require proper functioning of the tagged binding partners.

6.2 Comparing Cross-Correlation Analysis to FRET

Fluorescence Resonance Energy Transfer (FRET) presents a popular approach to study molecular interactions and conformational changes both in vitro and in vivo, and emerged as a promising spectroscopic tool to study single-molecule dynamics [57, 70, 102–104]. In FRET studies, molecular processes can be evidenced by spectral changes of the emission signal because of distance-dependent energy transfer from an excited donor dye to a long-wavelength acceptor, and additionally by determining decreased donor lifetimes in Fluorescence Lifetime Microscopy (FLIM) [105]. Hence, FRET has been used for the development of efficient intracellular ligand-binding and protease assays and also probed for protein-protein interactions and conformational changes in live cells [57, 58, 68, 69, 106]. However, FRET requires spatial distances between donor and acceptor dyes of typically 20–60 Å and thus faces a fundamental sterical limit reducing its versatility. Dual-colour fluorescence cross-correlation spectroscopy evades this limitation by only depending on concomitant signal fluctuations from a confocal detection volume in two spectrally distinct emission channels. FRET analysis itself provides no intrinsic calibration parameters like particle concentrations and dynamic particle properties.

6.3 Cross-Correlation Analysis Based Solely on Fluorescent Proteins

Obviously, finding an appropriate combination of two proteins for in vivo cross-correlation analysis would significantly simplify these studies. In addition to identifying protein fluorophores with high brightness and photostability, a small spectral overlap of fluorescence emission is required to minimize detector crosstalk, e.g. representation of a green dye's emission in the red detection channel. Depending on particle number and brightness ratios, strong detector crosstalk (>10%) results in a significant overestimation of double-labelled particle numbers, which can be corrected for by including 'crosstalk'-particles and their brightness into the correlation functions, analogously to polydispersity. Molecular blinking phenomena help to distinguish 'real' from artifactual cross-correlation, since the blinking shoulder is not present in real cross-correlation curves [21].

DsRed and its mutants display emission spectra sufficiently red-shifted for two-colour applications with GFP mutants. The only combination of GFP mutants to display a similar small spectral overlap is presented by EBFP and EYFP, but no FCS application of EBFP has been reported so far and its excitation maximum (388 nm) is unfavourable for intracellular applications.

Accordingly, Kohl et al. reported an in vitro protease assay for two-photon cross-correlation analysis based solely on the fluorescent proteins rsGFP and DsRed [88]. A GFP-peptide-DsRed construct named STEV-ST was purified and subjected to proteolysis separating the fluorophores. With regard to future

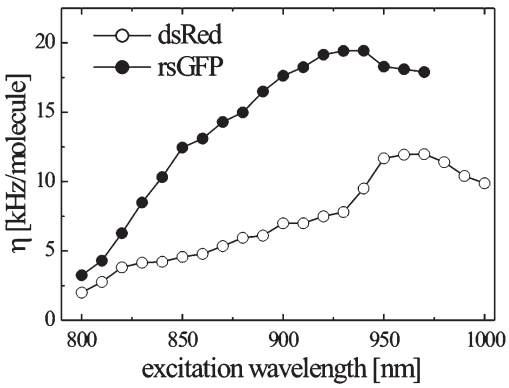


Fig. 4 TPE of GFP and DsRed. Wavelength dependence of the fluorescence emission yield η was determined for rsGFP and DsRed with two-photon-excitation. Photon yields per molecule were measured under conditions well below photobleaching. The optimal wavelength for joint excitation of both dyes was 940 nm with a maximum intensity of 20 mW

applications in live cells, simultaneous two-photon excitation of both fluorophores was attempted and efficiently realized at 940 nm with intensities safely below photobleaching (Fig. 4).

Equal dye brightness under the chosen conditions and minimized detector crosstalk of less than 10% presented convincing starting conditions for cross-correlation analysis based on rsGFP and DsRed. The proteolytic cleavage of STEV-ST was monitored by real-time FCS analysis as illustrated in Fig. 5a. Online kinetics of cleavage reactions at different enzyme concentrations were presented and allowed for easy discrimination of reaction rates depending on enzyme concentration. Moreover, the combinability of FRET and FCS analysis was demonstrated based on internal calibration by particle brightness values (Fig. 5b). The influence of FRET on cross-correlation analysis due to changing ratios of molecular brightness upon binding/unbinding of labelled particles could be quantified. In summary, sufficient proof of principle for protein-based dual-colour cross correlation analysis has been presented, along with an interesting extension of FCS to FRET phenomena.

6.4 Revisiting RFPs

Whether transfer of the dye pair GFP/DsRed or any analogous combinations to in vivo applications will work successfully depends on intracellular photon yields per molecule and the extent of fluorophore oligomerization. Nearly equal molecular brightness values for in vivo and in vitro situations were reported for EGFP implying that molecular brightness issues are unlikely to impede intracellular applications of the presented DsRed/GFP pair. Tetramerization of a reporter construct like STEV-ST devoid of any physiological role probably won't

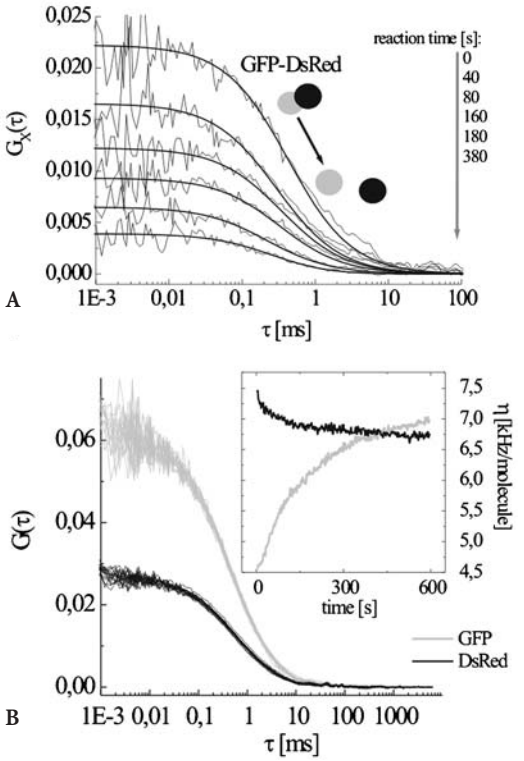


Fig. 5A, B Proteolytic cleavage analysed by: **A** cross-correlation analysis; **B** FRET on single-molecule-scale. In A, during the course of a specific proteolytic reaction GFP and DsRed get separated leading to gradually decreasing cross-correlation amplitudes $G_X(0)$ determined in 40-s intervals. In B, alternatively, autocorrelation functions and photon counts per molecule in kHz (*inset*) for rsGFP and DsRed were determined in parallel during a proteolytic digest. Changes in fluorescence intensity of FRET donor and acceptor were detected immediately after enzyme addition, whereas autocorrelation G(0) values and fluorescent particle numbers remained constant. The monitored increase of rsGFP fluorescence corresponds to approximately 35% FRET in the intact substrate

affect its function while presenting a delicate issue for protein-protein interactions in signalling. Remarkably, DsRed has been applied as the acceptor dye in several intracellular FRET studies addressing both conformational and protein interaction processes [58, 69]. Tetramerization only partly interfered with protein function, mainly by affecting cellular distribution. Nevertheless, DsRed fusions and DsRed alone were found to show variable behaviour in different cells, ranging from no detectable aggregation to punctate localization [48]. A wide distribution of oligomeric states up to high-order aggregates is rather obvious for visually punctate fluorescence and will impair protein function and FCS studies.

In summary, DsRed and its mutants represent interesting fusion partners for in vivo cross-correlation analysis although drawbacks cannot be excluded. Nevertheless, researches should consider the variety of DsRed mutants existing to date and evaluate compromises such as monomeric but rather dark (mRFP1) or bright but bulky dimeric tags such as tdimer2(12).

6.5 Cross-Correlation Analysis In Vivo

Performing cross-correlation analysis in vivo on genetically tagged proteins is most desirable since it avoids the need for any dye delivery and promises a highly quantitative characterization of all involved molecular species. The first cross-correlation studies in vivo based on external uptake of fluorescent molecules have been reported recently, simultaneously with the application of GFP and DsRed for in vitro cross-correlation analysis [16, 17, 88]. Thus, the next step of setting up a cross-correlation application solely based on fluorescent proteins in vivo is not far-fetched and shall be discussed on the basis of these studies.

So far, attempts to measure dual-colour cross-correlation in vivo rely on externally labelling at least one of the target molecules and introducing it into the cell. Studying endocytosis of fluorescent particles by FCS facilitates the issue of dye delivery and was accomplished to analyse the fate of A and B subunits of cholera toxin after cellular uptake [16]. Cross-correlation of the holotoxin consisting of the Cy2-labeled A subunit and the Cy5-labeled B subunit could be detected after endosomal uptake and was diminished significantly after the toxin had reached the Golgi. In this study, small endosomal vesicles constituted the diffusing entity as could be told from diffusional properties and particle brightness. A reduction in cross-correlation thus indicated that A and B subunits adopted individual pathways in vesicle trafficking. The study was performed using one photon excitation with two laser lines, applying the commercial Zeiss ConfoCor2 setup. Very high signal-to-noise levels could be obtained, which has been attributed to very bright diffusing particles, as the endocytic vesicles contained several fluorophores. Additionally, a thorough discussion of bleaching phenomena that can lead to artifactual cross-correlation is provided.

Another recent intracellular cross-correlation study succeeded in monitoring reversible protein binding. Using two-photon excitation, binding and unbinding of Alexa 633-labeled calmodulin and EGFP-labelled protein kinase II (CaMKII) according to changes in intracellular calcium levels could be monitored [17, further publication in preparation]. While the calcium-calmodulin-dependent kinase was expressed intracellularly as a fusion protein to EGFP, synthetically labelled calmodulin was delivered via electroporation.

7 Summary and Outlook

Fluorescence correlation spectroscopy clearly approved of its future potential for biochemical research on both, in solution and in the living cell. Two initial intracellular dual-colour cross-correlation assays clearly demonstrated its capability to address reversible protein-protein interactions and ligand-binding in cells. Detailed intracellular mobility analysis of fusion-proteins were presented. Most of the reported studies were performed on home-built setups, including TPE with a pulsed laser system, but also with a commercial FCS setup (Zeiss ConfoCor II) featuring only one-photon excitation. Both modes of excitation succeeded in performing dual-colour FCS within cells, stimulating enhanced propagation of (dual-colour) FCS in general. Compared to FRET analysis in binding assays, FCS is not limited by steric requirements to its reporter molecules, whereas FRET appears to be the method of choice to study conformational changes. In contrast to optical methods relying on a pixel-by-pixel analysis of fluorescence images, FCS directly observes single diffusing reporter molecules and consequently allows for a simultaneous acquisition of multiple molecular parameters. Fluorescent proteins showed to be largely suitable for FCS: a) to be applied both in vitro and in vivo, b) for single colour as well as cross-correlation applications and c) for one-photon and two-photon excitation. During the next decade, fluorescent proteins will probably present the ultimate fluorescent probe for labelling proteins in vivo: they allow for highly precise and intracellular tagging, can be manipulated and optimised by genetic engineering and partly proved to act as biologically inert particles. Although smaller fluorescent tags would confer steric advantages no smaller candidate protein has been reported or indicated to our knowledge.

Acknowledgements We thank Elke Haustein for helpful discussion and proofreading the manuscript. Plasmids encoding the fluorescent proteins mRFP1 and tdimer2(12) were kindly provided by Roger Y. Tsien (University of California at San Diego).

References

1. Magde D, Elson EL, Webb WW (1972) Phys Rev Lett 29:705
2. Eigen M, Rigler R (1994) Proc Natl Acad Sci USA 91:5740
3. Schwille P, Meyeralmes FJ, Rigler R (1997) Biophys J 72:1878
4. Schwille P (2001) Cell Biochem Biophys 34:383
5. Denk W, Strickler JH, Webb WW (1990) Science 248:73
6. Schwille P, Haupts U, Maiti S, Webb WW (1999) Biophys J 77:2251
7. Heinze KG, Koltermann A, Schwille P (2000) Proc Natl Acad Sci USA 97:10377
8. Haupts U, Maiti S, Schwille P, Webb WW (1998) Proc Natl Acad Sci USA 95:13573
9. Schwille P, Kummer S, Heikal AA, Moerner WE, Webb WW (2000) Proc Natl Acad Sci USA 97:151
10. Heikal AA, Hess ST, Baird GS, Tsien RY, Webb WW (2000) Proc Natl Acad Sci USA 97:11996

11. Malvezzi-Campeggi F, Jahnz M, Heinze KG, Dittrich P, Schwille P (2001) Biophys J 81:1776
12. Miyawaki A (2002) Cell Struct Function 27:343
13. Zhang J, Campbell RE, Ting AY, Tsien RY (2002) Nature Rev Mol Cell Biol 3:906
14. Chen Y, Muller JD, Ruan QQ, Gratton E (2002) Biophys J 82:133
15. Ruan QQ, Chen Y, Gratton E, Glaser M, Mantulin WW (2002) Biophys J 83:3177
16. Bacia K, Majoul IV, Schwille P (2002) Biophys J 83:1184
17. Kim SA, Heinze KG, Waxham MN, Schwille P (2002) Biophys J (Annual Meeting Abstracts) 82:44a
18. Schwille P (2000) Cross-correlation analysis in FCS. In: Rigler R, Elson E (eds) Fluorescence correlation spectroscopy – theory and applications. Springer, Berlin Heidelberg New York, p 360
19. Krichevsky O, Bonnet G (2002) Rep Prog Phys 65:251
20. Haustein E, Schwille P (2002) Fluorescence correlation spectroscopy. Biophysics Online Textbook
21. Bacia K, Schwille P (2003) Methods (Duluth) 29:74
22. Rigler R, Mets U, Widengren J, Kask P (1993) Eur Biophys J 22:169
23. Widengren J, Mets U, Rigler R (1995) J Phys Chem 99:13368
24. Schwille P, Korlach J, Webb WW (1999) Cytometry 36:176
25. Wachsmuth M, Waldeck W, Langowski J (2000) J Mol Biol 298:677
26. Gennerich A, Schild D (2000) Biophys J 79:3294
27. Schwille P, Meyeralmes FJ, Rigler R (1997) Biophys J 72:1878
28. Ambrose WP, Goodwin PM, Jett JH, Van Orden A, Werner JH, Keller RA (1999) Chem Rev 99:2929
29. Dittrich P, Malvezzi-Campeggi F, Jahnz M, Schwille P (2001) Biol Chem 382:491
30. Eggeling C, Fries JR, Brand L, Günther R, Seidel CAM (1998) Proc Natl Acad Sci USA 95:1556
31. Nazarenko I, Pires R, Lowe B, Obaidy M, Rashtchian A (2002) Nucleic Acids Res 30: 2089
32. Harms GS, Cognet L, Lommerse PHM, Blab GA, Schmidt T (2001) Biophys J 80:2396
33. Labas YA, Gurskaya NG, Yanushevich YG, Fradkov AF, Lukyanov KA, Lukyanov SA, Matz MV (2002) PNAS 99:4256
34. Tsien RY (1998) Annu Rev Biochem 67:509
35. Palm GJ, Wlodawer A (GREEN FLUORESCENT PROTEIN 302 PG. 378–394. 1999 [Figures], [Plates], [Colour plates]
36. Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Atsushi M (2002) Nat Biotechnol 20:87
37. Griesbeck O, Baird GS, Campbell RE, Zacharias DA, Tsien RY (2001) J Biol Chem 276:29188
38. Matz MV, Fradkov AF, Labas YA, Savitsky AP, Zaraisky AG, Markelov ML, Lukyanov SA (1999) Nat Biotechnol 17:969
39. Baird GS, Zacharias DA, Tsien RY (2000) Proc Natl Acad Sci USA 97:11984
40. Sacchetti A, Subramaniam V, Jovin TM, Alberti S (2002) FEBS Lett 525:13
41. Yanushevich YG, Staroverov DB, Savitsky AP, Fradkov AF, Gurskaya NG, Bulina ME, Lukyanov KA, Lukyanov SA (2002) FEBS Lett 511:11
42. Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, Tsien RY (2002) Proc Natl Acad Sci USA 99:7877
43. Fradkov AF, Verkhusha VV, Staroverov DB, Bulina ME, Yanushevich YG, Martynov VI, Lukyanov S, Lukyanov KA (2002) Biochem J 368:17
44. Gross LA, Baird GS, Hoffman RC, Baldridge KK, Tsien RY (2000) Proc Natl Acad Sci USA 97:11990

45. Yarbrough D, Wachter RM, Kallio K, Matz MV, Remington SJ (2001) Proc Natl Acad Sci USA 98:462
46. Bevis BJ, Glick BS (2002) Nat Biotechnol 20:83
47. Terskikh A, Fradkov A, Ermakova G, Zaraisky A, Tan P, Kajava AV, Zhao X, Lukyanov S, Matz MV, Kim S, Weissman I, Siebert P (2000) Sci 290:1585
48. Erickson MG, Moon DL, Yue DT (2003) Biophys J 85:599
49. Gavin P, Devenish RJ, Prescott M (2002) Biochem Biophys Res Commun 298:707
50. Patterson GH, Lippincott-Schwartz J (2002) Science 297:1873
51. Ando R, Hama H, Yamamoto-Hino M, Mizuno H, Miyawaki A (2002) PNAS 99:12651
52. Zapata-Hommer O, Griesbeck O (2003) BMC Biotechnol 3:5
53. Karasawa S, Araki T, Yamamoto-Hino M, Miyawaki A (2003) J Biol Chem 278:34167
54. Griffin BA, Adams SR, Tsien RY (1998) Science 281:269
55. Adams SR, Campbell RE, Gross LA, Martin BR, Walkup GK, Yao Y, Llopis J, Tsien RY (2002) J Am Chem Soc 124:6063
56. Bae JH, Rubini M, Jung G, Wiegand G, Seifert MHJ, Azim MK, Kim JS, Zumbusch A, Holak TA, Moroder L, Huber R, Budisa N (2003) J Mol Biol 328:1071
57. Miyawaki A, Llopis J, Heim R, Mccaffery JM, Adams JA, Ikura M, Tsien RY (1997) Nature 388:882
58. Calleja V, Ameer-Beg SM, Vojnovic B, Woscholski R, Downward J, Larijani B (2003) Biochem J 372:33
59. Insall R (2003) Biochem J 372
60. Leaney JL, Benians A, Graves FM, Tinker A (2002) J Biol Chem 277:28803
61. Baird GS, Zacharias DA, Tsien RY (1999) Proc Natl Acad Sci USA 11241
62. Biondi RM, Baehler PJ, Reymond CD, Veron M (1998) Nucleic Acids Res 26:4946
63. Doi N, Yanagawa H (1999) FEBS Lett 453:305
64. Wyborski DL, Bauer JC, Vaillancourt P (2001) Biotechniques 31:618
65. Gossen M, Bujard H (1992) PNAS 89:5547
66. No D, Yao TP, Evans RM (1996) Proc Natl Acad Sci USA 93:3346
67. Piston DW, Patterson GH, Knobel SM (Methods in cell biology, vol 58: green fluorescent proteins 58 PG. 31-+. 1999 [Figures], [Plates], [Colour plates]
68. Tawa P, Tam J, Cassady R, Nicholson DW, Xanthoudakis S (2001) Cell Death Differentiation 8:30
69. Tramier M, Gautier I, Piolot T, Ravalet S, Kemnitz K, Coppey J, Durieux C, Mignotte V, Coppey-Moisan M (2002) Biophys J 83:3570
70. Weiss S (1999) Science 283:1676
71. Sako Y, Minoghchi S, Yanagida T (2000) Nature Cell Biol 2:168
72. Schutz GJ, Kada G, Pastushenko VP, Schindler H (2000) EMBO J 19:892
73. Dickson RM, Cubitt AB, Tsien RY, Moerner WE (1997) Nature 388:355
74. Garcia-Parajo MF, Segers-Nolten GMJ, Veerman JA, Greve J, van Hulst NF (2000) Proc Natl Acad Sci USA 97:7237
75. Llopis J, Mccaffery JM, Miyawaki A, Farquhar MG, Tsien RY (1998) Proc Natl Acad Sci USA 95:6803
76. Lodish H, Baltimore D, Berk A, Zipursky SL, Matsudaira P, Darnell J (1995) Molecular cell biology, 3rd edn. Scientific American Books, New York, p 640
77. Garcia-Parajo MF, Koopman M, van Dijk EMHP, Subramaniam V, van Hulst NF (2001) Proc Natl Acad Sci USA 98:14392
78. Wiedenmann J, Schenk A, Rocker C, Girod A, Spindler KD, Nienhaus GU (2002) PNAS 99:11646
79. Petersen J, Wilmann PG, Beddoe T, Oakley AJ, Devenish RJ, Prescott M, Rossjohn J (2003) J Biol Chem M307896200

80. Kask P, Palo K, Fay N, Brand L, Mets U, Ullmann D, Jungmann J, Pschorr J, Gall K (2000) Biophys J 78:1703
81. Chen Y, Muller JD, So PTC, Gratton E (1999) Biophys J 77:553
82. Palo K, Metz U, Jager S, Kask P, Gall K (2001) Biophys J 79:2858
83. Starchev K, Buffle J, Perez E (1999) J Colloid Interface Sci 213:479
84. Palmer AG, Thompson NL III (1987) Biophys J 52:257
85. Blab GA, Lommerse PHM, Cognet L, Harms GS, Schmidt T (2001) Chem Phys Lett 350:71
86. Denk W, Piston DW, Webb WW (1995) Two-photon molecular excitation in laser-scanning microscopy. In: Pawley JB (ed) Handbook of biological confocal microscopy. Plenum Press, New York, chap 28, p 445
87. Berland KM, So PTC, Gratton E (1995) Biophys J 68:694
88. Kohl T, Heinze KG, Kuhlemann R, Koltermann A, Schwille P (2002) PNAS 99:12161
89. Piston DW (1999) Trends Cell Biol 9:66
90. Cluzel P, Surette M, Leibler S (2000) Science 287:1652
91. Kohler RH, Schwille P, Webb WW, Hanson MR (2000) J Cell Sci 113:3921
92. Dittrich PS, Schwille P (2002) Anal Chem 74:4472
93. Schwille P, Oehlenschläger F, Walter NG (1996) Biochemistry 35:10182
94. Boukari H, Nossal R, Sackett DL (2003) Biochemistry 42:1292
95. Chen Y, Muller JD, Tetin SY, Tyner JD, Gratton E (2000) Biophys J 79:1074
96. Meseth U, Wohland T, Rigler R, Vogel H (1999) Biophys J 76:1619
97. Saito K, Ito E, Takakuwa Y, Tamura M, Kinjo M (2003) FEBS Lett 541:126
98. Verkman AS (2002) Trends Biochem Sci 27:27
99. Croce AC, Spano A, Locatelli D, Barni S, Sciola L, Bottiroli G (1999) Photochem Photobiol 69:364
100. Brock R, Hink MA, Jovin TM (1998) Biophys J 75:2547
101. Kettling U, Koltermann A, Schwille P, Eigen M (1998) Proc Natl Acad Sci USA 95:1416
102. Ha T, Enderle T, Ogletree DF, Chemla DS, Selvin PR, Weiss S (1996) Proc Natl Acad Sci USA 93:6264
103. Deniz AA, Laurence TA, Beligere GS, Dahan M, Martin AB, Chemla DS, Dawson PE, Schultz PG, Weiss S (2000) Proc Natl Acad Sci USA 97:5179
104. Selvin PR (2000) Nat Struct Biol 7:730
105. Oida T, Sako Y, Kusumi A (1993) Biophys J 64:676
106. Immink RGH, Gadella TWJ, Ferrario S, Busscher M, Angenent GC (2002) Proc Natl Acad Sci USA 99:2416
107. Patterson G, Day RN, Piston D (2001) J Cell Sci 114:837
108. Gurskaya NG, Fradkov AF, Terskikh A, Matz MV, Labas YA, Martynov VI, Yanushevich YG, Lukyanov KA, Lukyanov SA (2001) FEBS Lett 507:16

Adv Biochem Engin/Biotechnol (2005) 95: 143–175
DOI 10.1007/b102213

Fluorescence Lifetime Imaging Microscopy (FLIM)

Erik B. van Munster · Theodorus W. J. Gadella (✉)

Swammerdam Institute for Life Sciences & Centre for Advanced Microscopy,
Section Molecular Cytology, Kruislaan 316, 1098 SM Amsterdam, The Netherlands
munster@science.uva.nl, gadella@science.uva.nl

Abstract Fluorescence lifetime imaging microscopy (FLIM) is a technique to map the spatial distribution of nanosecond excited state lifetimes within microscopic images. FLIM systems have been implemented both in the frequency domain, using sinusoidally intensity-modulated excitation light and modulated detectors, and in the time domain, using pulsed excitation sources and time-correlated or time-gated detection. In this review we describe the different modes in which both frequency-domain and time-domain FLIM instruments have been constructed in wide-field and in point-scanning (confocal) microscopes. Also, novel additional strategies for constructing FLIM-instruments are discussed. In addition to technical implementation, this chapter gives an overview of the application of FLIM in cell biological en biomedical studies. Especially for in situ protein-protein interaction studies using fluorescence resonance energy transfer (FRET), FLIM has proven to be a robust and established technique in modern cell biology. Other application areas, including usage of lifetime contrast for ion-imaging, quantitative imaging, tissue characterization and medical applications, are discussed.

Keywords Fluorescence lifetime imaging microscopy · Frequency domain · Time-correlated single photon counting · Fluorescence resonance energy transfer · Protein-protein interactions

List of Abbreviations

7-AAD	7-Aminoactinomycin D
AC	Alternating current
AOM	Acousto-optic modulator
APP	Amyloid precursor protein
BCECF	2′,7′-Bis-(2-carboxyethyl)-5(and-6)-carboxyfluorescein
BTC	Benzothiazole coumarin
cAMP	Cyclic-adenosine monophosphate
CCD	Charge-coupled device
CFP	Cyan fluorescent protein (derived from GFP)
cGMP	Cyclic-guanosine monophosphate
CHO	Chinese hamster ovary
DC	Direct current
DsRed	Red fluorescent protein from *Discosoma*
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EOM	Electro-optic modulator
FLIE	Fluorescence lifetime imaging endoscopy
FLIM	Fluorescence lifetime imaging microscopy
GFP	Green fluorescent protein from *Aequoria victoria*
MCP	Micro channel plate
MQAE	*N*-Ethoxycarbonylmethyl-6-methoxyquinolinium bromide
NADH	Reduced nicotinamide dinucleotide
PBFI	Potassium-binding fluorescence indicator
PI-TP	Phosphatidylinositol-transfer protein
PKC	Protein kinase C

PLD Phospholipase D
PMT Photo multiplier tube
PS-1 Presenilin-1
REP Rab escort protein
SNAFL Seminaphthofluorescein
SPQ 6-Methoxy-*N*-(3-sulfopropyl)quinolinium
TCSPC Time-correlated single-photon counting
TeNT *Tetanus* toxin
TIR Total internal reflection
YFP Yellow fluorescent protein (derived from GFP)

1 Introduction

Fluorescence lifetime imaging microscopy (FLIM) is a technique to determine the spatial distribution of excited state lifetimes in microscopic samples. This can mean everything from a single decay time, to an entire decay profile, in two or three dimensions. Typically, FLIM instruments are designed to measure lifetimes in the nanosecond range, since the lifetimes of most fluorochromes used in modern fluorescence microscopy fall within this range. In this chapter, an overview is presented of the various techniques used in FLIM instruments today and of application areas in biology and biomedicine.

2 FLIM Instrumentation

2.1 Historic Origins

FLIM has its roots in two fields of research: 1) microscopy and 2) fluorescence spectroscopy. In the latter field of research, non-spatially-resolved fluorescence lifetime measurements were performed since 1926 [1], i.e. long before FLIM was developed. Typically, bulk measurements were carried out using cuvettes. Not surprisingly, most of the methodology and nomenclature used in FLIM today, e.g. 'frequency-domain', and 'time-domain', have their origins in instruments that were used for cuvette-based lifetime measurements.

The first instrument combining time resolved fluorescence spectroscopy with microscopy, dates back to 1959 [2]. In this instrument, only single point measurements could be done, so strictly speaking no actual imaging was done. The first instrument measuring spatially resolved lifetimes was described in 1989 [3]. Various instruments were developed in the early 1990's independently by different groups. Nowadays, there are many FLIM instruments, mostly used for biomedical applications.

2.2 General Properties of FLIM Instruments

What all FLIM instruments have in common is that 1) the excitation light is intensity-modulated or pulsed, and 2) that the emitted fluorescence light is measured time-resolved. Since the lifetimes that have to be resolved are typically in the nanosecond range, this means that both the modulation of the excitation light and the detection need to be performed at extremely high speeds. Consequently, most of the conventional instrumentation used for steady-state fluorescence microscopy cannot be used. In the next section several modes of implementation of FLIM are discussed.

2.3 Classification of FLIM Instruments

Virtually all FLIM instruments in use today have been custom-build, in whole or in part, often with a specific application in mind. As a result, a great variety of instruments exists. To treat them all here would be impossible. Instead, we classified instruments in Table 1, based on the technique used to obtain the actual lifetime information (A to C) and based on the microscopy technique used (1 to 5). In this table, references to technical descriptions are presented. In the following section the properties of each class are discussed as well as the practical consequences for FLIM instruments in this class.

2.3.1 Frequency-Domain FLIM

The idea to use frequency-domain detection for the measurement of fluorescent lifetimes dates back to 1921 [4], although the idea to measure small phase changes to determine short time intervals is much older [5]. The first instruments to measure non-spatially-resolved lifetimes in the 1920s were all based on frequency-domain detection.

Table 1 Overview of different existing FLIM instruments

	1 Wide-field	2 Confocal	3 Multi-photon	4 Struct. illum.	5 TIR
A) Frequency-domain	[14, 46, 50, 160, 161]	[9]	[10, 162]	[59]	
B) Time-domain	[18, 94, 115, 152, 153, 163–165]	[166]	[162, 167–170]	[171, 172]	[173, 174]
C) Pump-probe		[20, 175, 176]			

In frequency-domain FLIM, the intensity of the excitation light is continuously modulated. Due to the (non-instant) fluorescence decay, the fluorescence emission will display a phase shift and a decrease in modulation. This can be understood if we look at the fluorescence intensity of a mono-exponentially decaying fluorochrome after an infinitesimal short pulse of excitation light at t=0 as described by Eq. (1):

$$I_{pulse}(t) = \frac{1}{\tau} \cdot \exp\left(\frac{-t}{\tau}\right) \tag{1}$$

with $I_{pulse}(t)$ the normalized intensity, t time, and τ, the fluorescence lifetime. If the excitation is not an infinitesimal short pulse, the resulting fluorescence intensity is described by the convolution of the excitation signal (E(t)) with Eq. (1):

$$I(t) = E(t) \otimes I_{pulse}(t) \tag{2}$$

A convolution corresponds to a multiplication in the Fourier domain. Fourier transforming $I_{pulse}(t)$, we obtain

$$\int_0^\infty \exp(i\omega t) \cdot \frac{1}{\tau} \cdot \exp\left(\frac{-t}{\tau}\right) dt = (1 + i\omega\tau)^{-1} \tag{3}$$

Rewriting this in the form $A \cdot \exp(i\varphi)$ enables us to see what happens to the phase and amplitude of each frequency component of the excitation signal after multiplication:

$$(1 + i\omega\tau)^{-1} = (1 + \omega^2\tau^2)^{-1/2} \cdot \exp(i \cdot \tan^{-1}(\omega\tau)) \tag{4}$$

Thus, each frequency component in the excitation signal E(t) will undergo a phase shift according to

$$\tan^{-1}(\omega\tau) \tag{5}$$

and an attenuation of the amplitude described by

$$(1 + \omega^2\tau^2)^{-1/2} \tag{6}$$

Since a DC component in the excitation signal will not be affected (ω=0), a change in amplitude is proportional to the modulation depth of the component.

Thus, it is possible to determine the lifetime τ, from the observed phase shift or from the decrease in modulation depth of the emitted fluorescent relative to the excitation signal, using Eqs. (5) and (6) (see also Fig. 1). Two lifetimes result from this: τ_φ, the lifetime based on the phase shift, and τ_M, the lifetime based on the decrease in modulation depth. In the derivation above, i.e. using a mono-exponential decay, $\tau_\varphi = \tau_M$. In order to analyze more complex

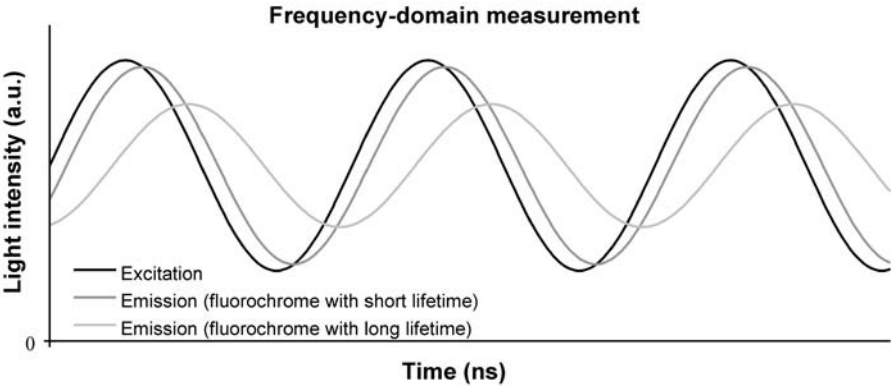


Fig. 1 Principle of frequency-domain lifetime measurement (see text). By using sinusoidally modulated excitation light and measuring the phase-shift and demodulation of the emitted fluorescence the lifetime of the fluorochrome is determined

decays, measurements can be repeated using multiple modulation frequencies. In non-spatially-resolved fluorescence spectroscopy measurement of phase shift and demodulation over a range of modulation frequencies is commonly used to resolve complex decays [6, 7]. In FLIM, measurements are usually limited to a single frequency, although multiple frequency FLIM has been developed and applied [8]. The optimal modulation frequency, i.e. the frequency that results in the most accurate lifetime measurements, is dictated by the lifetimes to be measured. Typically, an instrument is designed for a limited range of fluorochromes. Since most fluorochromes commonly used in (biomedical) fluorescence microscopy have lifetimes ranging from 1 to 10 ns, most frequency-domain FLIM instruments operate at frequencies between 10 and 100 MHz.

To apply frequency-domain lifetime measurement to FLIM, some special measures have to be taken: 1) the light source has to be modulated, and 2) phase and modulation have to be extracted from the recorded signal. In scanning microscopes this is relatively simple: the modulated fluorescence signal is usually measured directly using a photo multiplier tube (PMT), enabling the detection of phase and modulation depth using conventional electronics [9] or through cross-correlation [10, 11]. In wide-field frequency-domain FLIM, cross-correlation is used to obtain phase and modulation depth. For this, typically, a modulated image intensifier is used, modulated at the same frequency as the excitation light. For every pixel, the intensity as a function of the phase difference between excitation light and image intensifier gain is determined by recording a sequence of images at varying phase differences. The conventional way of determining phase and modulation depth from these recorded images is by Fourier analysis [12, 13]. Assuming the phase difference to be spread equidistantly over the 360-degree range, F_{sin}, F_{cos}, and F_{DC}, are determined for every pixel using

$$F_{sin,\,n} = \sum_{k=0}^{K} \sin(2\pi nk/K) \cdot I_k$$

$$F_{cos,\,n} = \sum_{k=0}^{K} \sin(2\pi nk/K) \cdot I_k \tag{7}$$

$$F_{DC} = \sum_{k=0}^{K} I_k$$

where K represents the number of recorded images, I_K the intensity in the k-th image, and n the harmonic of interest. Phase and modulation depth are then determined by

$$\varphi_n = \tan^{-1}(F_{sin,\,n}/F_{cos,\,n}) \tag{8}$$

$$M_n = 2\frac{\sqrt{F_{sin,\,n}^2 + F_{cos,\,n}^2}}{F_{DC}} \tag{9}$$

where φ represents phase, and M represents modulation depth, defined here as the amplitude of the modulation divided by the average intensity (M=1 corresponds to a modulation depth of 100%). Since it is not the absolute phase and modulation depth that is of interest, but rather the phase shift and the decrease in modulation depth relative to the excitation light, phase and modulation depth of the excitation light need to be determined as well. For this, a reflecting or scattering object is used to measure the excitation light directly. Lifetimes are determined using Eqs. (10) and (11):

$$\tau_\varphi = \frac{\tan(\varphi_{em} - \varphi_{ex})}{\omega} \tag{10}$$

$$\tau_M = \frac{1}{\omega}\sqrt{\frac{1}{\left(\frac{M_{em}}{M_{ex}}\right)^2 - 1}} \tag{11}$$

where φ_{em} and φ_{ex} represent phase, and M_{em} and M_{ex} represent modulation depth of the emission and excitation, respectively. ω is the angular frequency used for modulation. In the case of monoexponential decay, τ_φ and τ_M are equal. If $\tau_\varphi < \tau_M$, this is an indication that the decay is multi-exponential and has different lifetime components [14]. In this way it is very easy to discriminate simple from complex decay in frequency-domain FLIM. Usually, lifetimes are determined based on the fundamental frequency only (n=1), but in theory any harmonic present in the signal can be analyzed, yielding more information about additional lifetime components in the case of multi-exponential decays [8]. For references to frequency-domain instrumentation see Table 1, A1–A4.

2.3.2 Time-Domain FLIM

Conceptually, time-domain lifetime measurement is easier to understand than frequency-domain lifetime measurement. In time-domain lifetime measurement, a short (relative to the fluorescence lifetime) pulse of excitation light is given, after which the emitted fluorescence is measured time resolved [15], resulting directly in decay curves like that described in Eq. (1) (see also Fig. 2). Due to the requirement of short light pulses and fast detection, time-domain measurements became possible only about 40 years later than frequency-domain measurements using a flashlamp as excitation source [16].

In practice, time-domain measurements can be done using time correlated single photon counting (TCSPC) whereby the arrival time of the first photon after each pulse is monitored at very high time resolution [17]. By recording the arrival times of a large number of photons, a representation of the decay curve is obtained. In order for this approach to work, the chance of detection of a photon after a pulse should be low. If this is not the case, the distribution will be biased toward shorter lifetimes. It has been estimated that, for TCSPC to work for lifetime measurements, the detection efficiency should be 1% or lower [15]. This means that TCSPC always is relatively slow. Advantage is that the actual decay curve is measured directly.

Another time-domain technique is the collection of photons in a fixed number (typically 2 to 8) of discrete time intervals using gated detection. For the simple case of a mono-exponential decay, provided the pulse duration $\ll\tau$, the emitted fluorescent intensity can be written as

$$I(t) = I_0 \cdot \exp\left(\frac{-t}{\tau}\right) \tag{12}$$

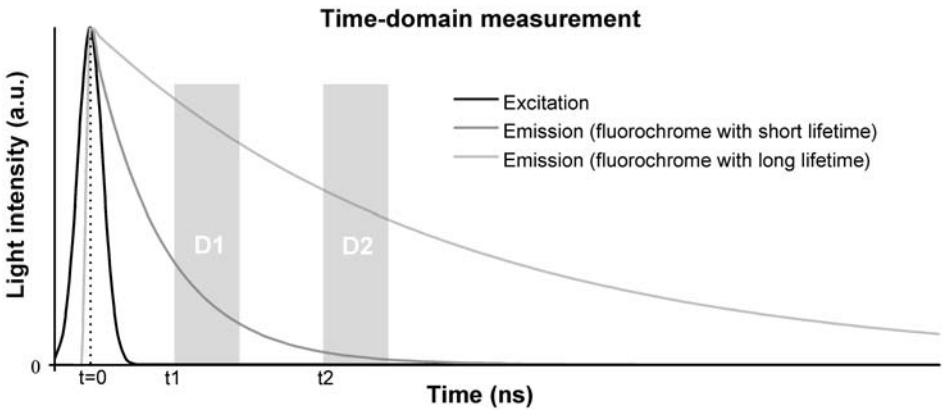


Fig. 2 Principle of time-domain lifetime measurement (see text). Fluorochromes are excited using a short pulse of light, after which the emitted fluorescence is measured time-resolved. Usually, fluorescence is recorded in two or more discreet time intervals

with I (t) the fluorescent intensity as function of time, I_0, the fluorescent intensity immediately after the pulse, t time after the light pulse, and τ the excited state lifetime. In this case, it can be shown that measurement of integrated intensity in two time intervals suffice to determine τ, using

$$\tau = \frac{t_1 - t_2}{\ln(D_1 - D_2)} \tag{13}$$

with t_1 and t_2 the time delay between the excitation pulse and the start of detection interval 1 and 2, respectively, and D_1 and D_2, the integrated intensities in intervals 1 and 2 (see Fig. 2). When the light pulse is not small relative to the fluorescence lifetime, corrections have to be made for the shape of the light pulse, since the measured fluorescence decay is the convolution of the decay and the shape of the light pulse. This may also be necessary for the temporal response characteristics of the detector, unless the rise and fall time are much smaller than the fluorescence lifetime. In cases where the decay is multi-exponential, more measurements will be required, and more stringent application of all the necessary corrections for pulse shape and detector profile should be used.

To apply time-domain detection to FLIM, the microscope should be equipped with a pulsed light source, e.g. a pulsed laser. For the detection in scanning microscopy modalities, a PMT and either dedicated TCSPC electronics or a system for gated detection are used. In wide-field microscopy, TCSPC is extremely difficult due to the fact that not only the arrival time but also the spatial origin within the sample of each photon has to be monitored. This can be done using a quadrant detector [18] but, in order to build up an image, the instrument has to wait for enough photons to get a reliable representation of the decay curve for each location, leading to extremely long acquisition times. Therefore, the much faster gated image intensifier approach is usually used instead of TCSPC for time-domain wide-field microscopy. For reference to time-domain FLIM instrumentation see Table 1, B1–B5.

2.3.3 Pump-Probe FLIM

A method less common for lifetime measurements is the so-called pump-probe or double-pulse approach. Like time- and frequency-domain detection, the technique originates in non-spatially-resolved fluorescence spectroscopy [19]. In this technique, two very short excitation pulses follow each other. The first pulse excites fluorochromes inside the detection volume to full or partial saturation. The second pulse, or probe pulse, arrives at a variable (ns) time delay. If the time delay between the pulses is short compared to the fluorescence lifetime, most of the fluorochromes will still be in the excited state when the second pulse arrives so that the second pulse cannot excite additional fluorochromes and thus does not lead to additional fluorescence. If the time delay is long, most fluorochromes will have relaxed back to their ground state, so that the second pulse leads to

additional fluorescence. By measuring the fluorescence intensity as function of the delay between the two pulses, the lifetime can be determined. A variation of this technique is to use a second pulse at the emission wavelength, instead of the excitation wavelength. This way, the state of the molecules is read using stimulated emission [20]. Advantage of the pump-probe approach is that no gated or gain-modulation detection is required.

To apply the pump-probe technique for FLIM, either two pulses with a controllable delay between them are required as excitation source or two sinusoidally varying signals, whereby the signal is obtained by cross-correlation [11, 21]. In practice, two pulses are obtained by splitting the light from a single pulsed laser using a beam splitter, and introducing an extra pathlength in one of the beams, after which the two beam are recombined. The pathlength-difference between the two beams thus controls the delay between the two pulses. For the detection, no special measures are necessary: measuring the total fluorescence intensity as function of the delay between the pulses does not require time-resolved measurement. One of the disadvantages of pump-probe implementations of FLIM is the requirement of high power pulsed laser sources and problems related to photobleaching and phototoxicity. For reference to pump-probe FLIM instrumentation see Table 1, C2.

Besides the distinction between the way the lifetime is measured, a distinction can be made based on the optics used. Since the development of the first FLIM instruments, many different types of fluorescence microscopy have been used. FLIM has been integrated in one form or another with most types of microscopes available today. Each type of microscopy has its own requirements and difficulties in using it for FLIM.

2.3.4 Wide-Field Microscopy

What is generally meant by wide-field microscopy is any microscope whereby image formation takes place by the optics without scanning: the lens directly forms an image, which can be projected on a camera or observed through the eye piece. Before the development of confocal microscopy and other scanning microscopy modes, this was the only way to perform microscopy. In the biomedical sciences, wide-field fluorescence microscopy is still widely used and offers a number of advantages over confocal and other advanced microscopy modes.

To use this type of microscope for FLIM, the camera has to be able to measure time resolved. Typically, a gain-modulated image intensifier is used for this, although recently the use of a modulated CCD camera for this has been reported [22, 23]. The light source has to be intensity-modulated. This can be pulsed, for time-domain FLIM, or continuously, for frequency-domain FLIM. If a laser is used for excitation, the beam will have to be expanded to illuminate not a single point in the specimen but the entire field of view in order to obtain wide-field images. For references to wide-field FLIM instrumentation see Table 1, A1 and B1.

2.3.5 Confocal Microscopy

In confocal microscopy [24], image formation takes place by scanning the object thereby rejecting the light originating from out-of-focus locations by using a pinhole in the image plane. The main advantage over wide-field microscopy is that the images have a much higher resolution in the z-direction, as well as a little bit more resolution in the lateral direction. This makes it possible to record 3D stacks of objects by recording slices of an image. To use a confocal microscope for FLIM, the detected fluorescence must be measured time-resolved (in the case of time-domain FLIM), or demodulated (in the case of frequency-domain FLIM). Since the fluorescence is usually a single signal measured by a PMT, this is simpler than in wide-field FLIM where an entire image has to be recorded time-resolved, or demodulated. The light source, which is usually a laser, must be pulsed or modulated. For references to confocal FLIM instrumentation see Table 1, A2, B2, C2.

2.3.6 Multiphoton Microscopy

What is generally meant by multi-photon microscopy is a technique where by the wavelength of the light used for excitation is a constant factor (2 for two-photon microscopy) higher than the excitation wavelength of the fluorochrome [25, 26]. This means that excitation of the fluorochrome is achieved by simultaneous absorption of two (infrared) photons rather than by a single photon. Consequently, the excitation probability is not linear with the excitation intensity (as with conventional fluorescence), but depends on the second power of the excitation intensity in the case of two-photon microscopy. Due to this quadratic intensity-dependence, excitation takes place only at the exact focal area of the focused laser beam. Areas outside this focus are not excited. This means that there is an inherit sectioning effect, enabling the recording of 3D images by scanning a specimen in three dimensions. Unlike confocal microscopy, no pinholes have to be used achieve this sectioning. To achieve the high powers needed for multiphoton microscopy, usually a pulsed laser is used. The power during these very short pulses is extremely high compared to the average power of the laser. To use multiphoton microscopy for FLIM, the same measures have to be taken as for confocal microscopy, i.e. measure the fluorescence time-resolved or demodulated. Since a pulsed laser is already used in multiphoton microscopy, no special measured have to be taken to modulate the excitation light. For references to multi-photon FLIM instrumentation, see Table 1, A3 and B3.

2.3.7
Structured Illumination

Although technically not a fundamentally different microscopy technique, a number of microscope techniques have appeared that can be regarded as systems to obtain a sectioned image, i.e. with a high z-resolution, while maintaining wide-field aspect. Typically, excitation light is not spread uniformly over the field of view, but is structured. This can be anything from a pattern of stripes [27], an array of points [28] a line [29], or a programmable pattern [30]. By scanning this structure over the object a sectioned image is obtained, either directly or after some processing of the images, enabling real-time, or near real-time, observation of the sectioned image. To use these systems for FLIM measurements the same type of adaptations are required as for wide-field microscope. The added advantage over wide-field systems is the sectioning capabilities of these systems. For references to structured illumination FLIM instrumentation see Table 1, A4 & A5.

2.3.8
Total Internal Reflection (TIR)

A more recent development is the use of total internal reflection (TIR) techniques for fluorescence microscopy [31, 32]. In this technique, excitation light is reflected back inside a medium of a relatively high index of refraction (typically glass), from an interface with a medium with a lower index of refraction (typically, water, medium, or a cell membrane). The angle at which this reflection takes place is greater than or equal to the angle for total internal reflection. Importantly, this total internal reflection will generate a so-called evanescent wave inside the medium of the lower refractive index that extents 10–100 nm inside the medium and is able to excite fluorochromes. Thus is it possible to selectively excite fluorochromes in an extremely thin layer just above the interface. Apart from the extreme sectioning capabilities, TIR shares aspects with wide-field microscopy: a wide field-of-view is excited, and a large area is imaged by image-forming optics. To integrate TIR with FLIM, the excitation light has to be modulated or pulsed, and the recorded image should be gain-modulated or gated, for instance using an image intensifier. For references to total internal reflection FLIM instrumentation see Table 1, B5.

2.4
Comparison between FLIM Systems

Comparing the frequency-domain and the time-domain approach, it was found that one is not fundamentally better than the other in terms of signal to noise ratio, when measuring the fluorescence lifetime of a mono-exponentially decaying fluorochrome [33]. In both systems, it is crucial, however, to use the optimal settings for each particular measurement. In frequency-domain FLIM, the

modulation frequency should be chosen with care. In time-domain, the number of sampling points to reconstruct the decay curve is important for achieving the best signal [34]. Time-domain offers the possibility to measure the decay curve itself, provided enough time-intervals are measured. From this, a lifetime can be obtained by exponential fitting. The number of time intervals determine how many different lifetime components can be resolved. If only a few time intervals are measured, as is often the case, usually only a single lifetime value is obtained. Frequency-domain FLIM measurements are more indirect. Actual decay curves are not measured. Lifetimes are determined from recorded images by additional calculations. To resolve more complex multi-exponential decays, measurements using different modulation frequencies or the use of higher harmonics have to be used [8]. The difficulty in time-domain is the pulsed light sources. Typically, a relatively expensive pulsed laser is used for this. These lasers are not available in as many wavelength as continuous lasers often used as excitation source (HeNe, HeCd and Ar-Ion lasers).

Ideally, the choice for a specific FLIM system should be based on the specific application and its requirements in terms of speed, spatial resolution and lifetimes to be resolved. In practice, however, this choice is often dictated by financial considerations, equipment already available, and the expertise available. Typically, FLIM instruments are custom-build by the users, although a number of companies nowadays make products that are specifically made to help transform an existing microscope into a FLIM system [35–37].

2.5 Light Sources for FLIM

The main difference between light sources for conventional, steady state, fluorescence microscopy, and light sources for FLIM is that the latter should be modulated at radio frequencies or pulsed with very short pulse duration. This has a number of consequences, the main one being that light sources that work well for steady state fluorescence microscopy are not very suitable for FLIM.

2.5.1 Mercury Arc Lamps

Traditionally, mercury arc lamps are the excitation source of choice for steady state wide-field fluorescence microscopy. They combine a high irradiance with a UV- and blue-rich spectrum, making it ideal to excite most fluorochromes used in fluorescence microscopy. Modern FLIM instruments do not use arc lamps, simply because both direct modulation of arc lamps is difficult at frequencies needed to resolve lifetimes of most fluorochromes. Externally modulating the light originating from an arc-lamp is also difficult due to the fact that the light does not form a narrow collimated beam like laser light, but originates from an extended source. This makes it difficult to use external modulators like acousto-optic- or electro-optic modulators. It is not impossible to

use an arc lamp. The instrument build by Venneta in 1959 is based on externally modulated light from an arc lamp, using a primitive form of acousto-optic modulation [2].

2.5.2 Continuous Wave Lasers

Continuous wave lasers are relatively inexpensive and easy to operate compared to pulsed lasers. The most popular lasers, argon-ion (351, 458, 488, and 514 nm), helium-neon (543, 594, 612, and 632 nm), and helium cadmium (325 and 442 nm) are very suited for fluorescence application, and have been used as standard light sources for confocal microscopy [38]. In order to use these lasers for FLIM, their intensities have to be modulated at the high frequencies. Modulation of lasers can be done externally using acousto-optic or electro-optic modulators (AOMs and EOMs, respectively). An AOM [39] consists of a piece of transparent material with a piezo crystal attached. If a high frequency electrical signal of the correct frequency is applied to the crystal, a standing wave is produced in the material. As a result, the AOM behaves as a phase grating that switches on and off at twice the applied frequency. If a laser beam is sent through the AOM, light will be refracted to higher orders when the grating is on, and left unobstructed when the grating is off. This way, a modulated laser beam is created in the zeroth order that is isolated using an iris stop. Typically, AOMs have one single resonance frequency for which the modulation depth will be maximal. This complicates changing of modulation frequencies. AOMs have been used in a number of wide-field frequency-domain FLIM instruments.

An EOM [40] can be seen as a variable waveplate whose retardation depends on the voltage applied across. When light entering the EOM is polarized, and an analyzer is placed at the output window, perpendicular or parallel to the input polarization, the EOM acts as an intensity modulator. A problem with EOMs is that the power consumption increases with increasing modulation frequency, making things impractical at high frequencies often used in FLIM (up to 100 MHz). Advantage is that there is no resonance frequency, making it very easy to tune to different frequencies. Although not used as widely as AOMs, EOMs have been used for frequency-domain FLIM, both wide-field and scanning [41, 42].

When using lasers as light source for wide-field application, the beam has to be expanded to fill the entire field of view. Due to the inherit coherence of laser light, speckles are formed if no measures are taken. This speckle problem is not unique to FLIM, but emerges whenever coherent laser light is used for illumination of an extended area [43]. Speckle formation is avoided in existing FLIM instruments by passing the light to a constantly moving diffuser, or through a mechanically vibrating multi-mode fiber. Actually, speckles do not disappear, but are averaged out within the integration time typically used to record an image.

2.5.3 Pulsed Lasers

Nowadays, the standard source for short light pulses is the pulsed laser. For FLIM, this can be a modified version of a continuous wave laser such as a mode-locked Argon-Ion laser, which typically produces pulses 70 ps width, but more often is a titanium:sapphire laser, producing pulses in the order of 100 fs. Disadvantage of this laser is that its output lies typically between 720 and 1000 nm, making it useless as light source for conventional fluorescence. As a source for two-photon fluorescence, however, they are excellent. In two-photon fluorescence its effective excitation range is 360 to 500 nm. Furthermore, the short pulse width increases the peak intensity of the pulses to the high intensities needed for two-photon excitation.

Typically, pulsed light sources are used for time-domain measurements, but frequency-domain measurements are possible as well, since the pulse rate can be made very constant and usually is in the range used for frequency-domain FLIM measurements (around 80 MHz). An added advantage is that a pulsed signal is very rich in harmonic-content, when compared to a modulated signal coming out of an AOM, or EOM, enabling multi-frequency measurements.

2.5.4 Light-Emitting-Diodes

A relatively new development is the use of Light-Emitting-Diodes (LEDs) as light source for wide-field FLIM [44–46]. Advantages are obvious: LEDs are extremely cheap, can be directly modulated at MHz frequencies and are easy to handle (no cooling required). An extra advantage is that the modulation frequency can be varied very easily, enabling multi-frequency FLIM to resolved more complex decays [44]. Nowadays, LEDs are available in many wavelengths for excitation of most fluorochromes. Emitted power is less than what can be achieved with laser sources, but for many applications this is not a problem.

At the moment, LEDs are constantly being improved in terms of efficiency, output power, and available wavelength. If developments continue, it could very well be that, in time, LEDs are the light source of choice, not only for wide-field FLIM, but also for conventional steady-state wide-field fluorescence microscopy.

2.6 Detectors for FLIM

Another point where FLIM differs from steady state fluorescence microscopy is the detection. In order to resolve nanosecond lifetimes, measurements have to be time-resolved, to a degree far beyond what is possible with fast cameras.

2.6.1 Photo Multiplier Tubes

Photo multiplier tubes (PMTs) are the detector of choice for scanning microscopy modes, whereby a single intensity is measured as function of the position. Basically, PMTs work by electronic multi-stage amplification of a free electron originating from a photocathode by accelerating it within the tube, and letting it hit multiple subsequent dynodes. For FLIM, the response time of the detector should be fast enough for both time- and frequency-domain measurements. Nowadays, PMTs are very fast, and finding a PMT with a sub ns rise time is usually no problem.

2.6.2 Image Intensifiers

For wide-field FLIM, the image of the object is not scanned but is formed by the optics. In order to measure this image-time resolved, a gated or modulated image intensifier is used. Basically, an image intensifier consists of a photocathode on which the image formed by the microscope is projected. The electrons freed from the cathode are accelerated through a so-called microchannel-plate (MCP), which is a structure consisting of tiny pores (microchannels), that will keep the electrons in a straight path. After the MCP, the electron will impinge on a phosphor screen, thereby giving rise to emitted light, similar to the way light is produced in a conventional TV screen. Due to the MCP, the information on the location of the generation of the electron is conserved when the electron arrives on the phosphor screen. In other words, the image formed on the phosphor screen is a replica of the image that is projected on the photo-cathode. By varying the voltage on the MCP, the gain of the intensifier can be controlled. This way the intensifier can be modulated or turned on and off at very fast rates. The image on the phosphor screen is usually imaged onto a conventional, slow-scan, CCD camera to record the image.

This way it is possible to use gated detection (for time-domain FLIM), or intensity-modulation (for frequency-domain FLIM). Disadvantage of the use of image intensifiers is that they affect the image. Due to various causes, the output image on the phosphor screen is less sharp than the input image. Modulated or gated image intensifiers are relatively expensive. Often, they are the most expensive item of the entire FLIM set-up and usually also the most fragile: overexposure of the photo cathode can irreversibly damage the device.

Nevertheless, image intensifiers are still very popular for wide-field FLIM, the main reason being lack of serious alternatives.

2.7 Recent Developments

2.7.1 3D Reconstruction of Wide-Field FLIM

In terms of the FLIM instrumentation itself, a recent development in the analysis and processing of lifetime data is the extraction of 3D information, using techniques developed and applied for conventional steady-state microscopy [47–49], from wide-field FLIM recordings [50]. Basically, images are recorded in z-stacks, after which reconstruction is used. Apart from acquiring 3D information, the main advantage is that the lifetime data themselves become more reliable. This is because out-of-focus light that can affect measurements in wide-field microscopy, is digitally filtered out. Unlike intensity in steady-state fluorescence microscopy, lifetimes do not simple 'add up'. Hence, removing out-of-focus light yields more reliable measured lifetimes, especially in cases where there is a lot of out of focus blur originating from fluorochromes with different lifetimes.

2.7.2 Global Analysis

Another development is the use of so-called 'global analysis' [51, 52] for frequency-domain FLIM. Like many things in FLIM, the original idea originates in the area of fluorescence spectroscopy [53]. When using frequency-domain FLIM, a single measurement will result in two lifetimes, one resulting from the phase shift, and one resulting from the decrease in modulation depth: τ_φ and τ_M in Eqs. (10) and (11). In the case of a mono-exponential decay, these lifetimes are identical to each other and to the lifetime of the sample. If the lifetime consists of two components, like (e.g. due to FRET [54]), the phase lifetime is lower than the modulation lifetime. Unfortunately, two measured lifetimes do not suffice to resolve three independent parameters (the lifetimes of the two components plus the relative strengths of the two). Global analysis makes use of the measured lifetimes of multiple pixels in one image to resolve these three parameters under the assumption that the lifetimes of the two components are invariable over the whole image, whereas the relative strength varies from pixel to pixel.

2.7.3 Combination of FLIM with Other Modalities

Another development is the combination of FLIM with other modalities, such as polarization or anisotropy measurements. This has resulted in the measurement of spatially resolved anisotropy decay measurements in microscopy [41, 55–57]. Basically, conventional FLIM measurements are combined with methods to determine the polarization state of the fluorescence after using polarized excitation light. This is done by measuring fluorescent intensity and

lifetime through a polarizer parallel and perpendicular to the excitation polarization direction. Time resolved anisotropy measurements can be used to determine rotational diffusion constants and thus to size and degrees of freedom of bound and unbound fluorochromes.

Analogous to this development is the combination of spectrally measurements with FLIM (sFLIM, [58, 59]), resulting in spectrally resolved lifetimes. In practice, this can be done relatively easy, by inserting a spectrograph or dispersing element between the object and the detector.

3 Applications of FLIM

Any phenomenon influencing the excited state fluorescence lifetime can be studied with FLIM. By acquiring FLIM-micrographs, the spatial distribution of these phenomena can be imaged. An excellent review of phenomena affecting excited state lifetimes of fluorophores and general application areas for FLIM has been published recently [60].

3.1 FRET-FLIM

Without any doubt, the most powerful FLIM-application in biology is fluorescence resonance energy transfer (FRET)-microscopy. In FRET, a quantum of energy is transferred in a nonradiative process from a donor fluorophore to an acceptor chromophore. This process only occurs if i) the donor fluorescence emission spectrum overlaps with the acceptor absorbance, ii) the donor and acceptor chromophores are in very close proximity (generally less than 9 nm), and iii) if the transition dipole moments of the donor and acceptor chromophores are not perpendicular. As a consequence of FRET, the donor fluorescence quantum yield is reduced. If the acceptor molecule is also a fluorophore, the acceptor fluorescence is increased (the so-called sensitized-emission). Hence, FRET leads to donor quenching, the energy of which can be 'converted' into acceptor fluorescence. The donor quenching is caused by an additional deactivation pathway from the donor-excited state (the energy transfer pathway) which does not yield a donor emission photon. Besides quenching, the extra deactivation pathway from the donor excited state also increases the rate by which the donor can relax from the excited state into the ground state. In other words the donor fluorescence (or excited-state) lifetime in a FRET situation (τ_{FRET}) is reduced as compared to the control donor fluorescence lifetime τ (see Eq. 14 in which

$$\tau_{FRET} = \tau(1 - E) \tag{14}$$

where *E* is the energy transfer efficiency). This reduction in lifetime can be imaged using FLIM. In this way the spatial distribution of FRET can be visualized in microscopic samples. For reviews of FRET see [60, 61].

One could argue that it would be much less technically demanding just to image the decreased donor fluorescence intensity rather than measuring its lifetime. The problem with quantitative imaging of intensities is that they do not depend only on the local quantum yield of the fluorophores but also on i) the local concentration of the fluorophore, ii) the optical path of the microscope, iii) the local excitation light intensity and iv) the local fluorescence detection efficiency. Especially the local concentration of the fluorophore is not easily determined separately from the local quantum yield. In contrast, FLIM does provide an independent estimate of the local donor lifetime (proportional to the donor quantum yield), independent of i–iv mentioned above. Hereby FLIM is a robust technique for measuring FRET.

The enormous potential of FRET in biology is caused by the fact that FRET only occurs at donor-acceptor proximity between 0–9 nm. The actual distance dependency of the FRET-efficiency E is described by

$$\mathrm{E} = \frac{R_0^6}{R_0^6 + r^6} \tag{15}$$

in which R_0 is the Förster radius and r is the distance between the donor and acceptor chromophores. Depending on the choice of donor-acceptor pair, R_0 usually is in the range of 3–6 nm [62]. The FRET-efficiency E is equal to 50% at $r{=}R_0$ and at larger distances E quickly drops to zero because of the inverse sixth-power distance-dependency. Hence, FRET is only observed at distances $r{<}1.5R_0$ (i.e. 0–9 nm). This distance scale is in the range of protein size and hence only occurs upon physical interactions between donor and acceptor molecules. Hence, by equipping proteins, DNA or lipids with acceptor and donor chromophores one can study molecular interactions in single cells. In this way FRET-FLIM combines both nm resolution (revealing molecular interactions) and the general sub-micrometer optical spatial resolution (mapping these interactions within the cellular context). As such FRET-FLIM has been reviewed both highlighting application areas and technical implementation [63–66].

3.1.1 FRET-FLIM Studies of Protein-Protein Interactions

FRET-FLIM has been extensively applied to measure protein-protein interactions, especially in cell signaling-studies. Using FRET-FLIM for the study of epidermal-growth factor receptor (EGFR or ErB1), has elucidated various details in relation to receptor-oligomerization and -activation. FRET-FLIM revealed the existence of predimerized receptors using low temperature incubations and donor/acceptor labeled EGF [54]. The phosphorylation status of the EGFR has been visualized by FRET-FLIM using donor labeled EGFR and acceptor labeled phosphotyosin-specific antibodies [67]. Using this approach also lateral propagation of the ErB1-receptor activation has been visualized in single living cells following local stimulation [68]. Also new FRET-(FLIM) techniques including

rFLIM and emFRET have been applied recently for the study of erB1-signalling [41]. Similarly, FRET-FLIM has been applied to study the interaction between receptor protein-tyrosine phosphatase alfa in living cells using CFP/YFP tagged receptors and truncations thereof indicating that the transmembrane domain might mediate dimerization [69]. Not only signaling at the membrane but also downstream targets in the cytosol have been visualized using FRET-FLIM. As such, several molecular events in protein kinase C (PKC)-signaling have been visualized in (living) cells, including proteolytic processing of dual labeled PKC-beta1 [70], activation through phosphorylation of PKCalfa using fluorescently tagged PKC and labeled anti-phosphotyrosine antibodies [71], interaction between PKCalfa and beta1 integrin [72], association between PKCalfa and fascin [73] and complex formation involving PKCalfa, integrin beta and ezrin [74, 75]. Also at the latest stages of signal transduction, FRET-FLIM proved useful in elucidating interactions between transcription factors in the intact nucleus. FRET-FLIM enabled the specific visualization of homo- and heterodimers of MADS box transcription factors in nuclei of living plant protoplasts [76]. Also in mammalian cells FRET-FLIM has been applied to study homodimerization of transcription factor CAATT/enhancer binding protein alfa [77, 78].

FRET-FLIM has been proven to be useful in monitoring interactions between key proteins involved in diseases such as interactions between Rab-GTPases and the Rab escort proteins (REP1 and REP2) which are important in Rab-prenylation diseases [79]. Furthermore, interactions between amyloid precursor protein (APP) and presenilin (PS-1) have been studied both of which have been implicated to be involved in the neurodegenerative Alzheimer disease [80, 81]. More recently, insight has been obtained in the molecular mechanism of activation of upstream components in the signaling cascade to PS-1. It was found that PS-1 is recruited to the Notch1-receptor upon ligand (delta 1)-stimulation of Notch1 [82]. Dysferlin, mutations of which have been implicated in limb girdle muscular dystrophy type 2B and Miyoshi myopathy, has been shown to interact with both Annexin A1 and A2 in a Ca^{2+} and sarcolemmal disruption-dependent manner from which a model is proposed for the function of dysferlin in the aforementioned diseases [83]. FRET-FLIM also revealed homomultimerization of the membrane-active 2B protein from enteroviruses. These interactions play a role in permeabilizing endomembranes for calcium [84]. Finally, insight was generated in the molecular mechanism of the zinc protease tetanus toxin (TeNT). FRET-FLIM showed that TeNT and Thy-1 (a 15-kDa protein) are closely associated at the plasma membrane and that TeNT is enriched in detergent-insoluble lipid rafts [85].

3.1.2 FRET-FLIM Studies of Lipid-Protein Interactions

By incorporating fluorescent lipids in biological membranes, association of (fluorescently-tagged) peripheral membrane proteins to these membranes can

be visualized using FRET-FLIM. Using this strategy, [86] have studied the association of phosphatidylinositol transfer proteins (PITPs) to membranes. Similarly FRET-FLIM was employed to study the activation of phospholipase D (PLD) by EGF by monitoring its translocation to membranes using fluorescently tagged-PLD and fluorescent lipids incorporated in membranes of living cells [87]. In principle FRET-FLIM could also be used to probe the local lipid environment of (trans)membrane proteins [88].

3.1.3 FRET-FLIM Studies of DNA Structure and DNA-Protein Interactions

By using DNA-binding fluorescent dyes, chromatin structures can be visualized in single cells. By employing Hoechst 33258 as donor which specifically binds to AT-rich regions in the DNA, and 7-aminoactinomycin D (7-AAD) which specifically binds to GC-rich regions in the DNA as acceptor, FRET-FLIM revealed non-homogeneous FRET patterns in individual nuclei. Furthermore, it was found that the FRET-efficiency was increased during the cell cycle when moving from G(0/1) into G(2)/M. Hence by using this approach highly condensed heterochromatic regions in the nucleus can be visualized [89–91]. Interestingly, as can be expected from the nature of the distribution of donors and acceptors within the 'linear' DNA-structure, non-exponential decays have been detected in these FRET-FLIM studies. In addition to FRET-FLIM (monitoring proximity of donor/acceptor dyes), lifetime-contrast of single dyes can be used as well for studying chromatin structure (see below).

In a recent study, SYTOX-dyes and fluorescently-tagged proteins were used in a FRET-FLIM study to monitor DNA-protein interactions in single cells. As a control, strong FRET was monitored between GFP-labeled histon2B and SYTOX-orange. Also FRET could be detected between SYTOX (labeling chromatin) and transcriptional activator proteins [92].

3.1.4 Fluorescent Tags for FRET-FLIM

In principle, any couple of fluorophores can be used for FRET, provided that the emission spectrum of the donor overlaps with the absorption of the acceptor. For a review of FRET-couples (and R0 values) of chemical dyes see [62]. Furthermore, donors with a high fluorescence quantum-yield and acceptors with a high molar absorbance will display increased FRET. For FLIM it will be important to tune the instrument-performance to ensure maximal sensitivity to small changes in lifetimes at the control donor lifetime. Usually this is easily achieved. Many FRET-pairs have been used for FRET-FLIM including chemical probes as Fluorescein-Rhodamine [54, 93], calcein-sulforhodamine B [94], and Cy3-Cy5, [70]. Since 1996, the availability of genetic-encoded fluorophores such as CFP, GFP, YFP has boosted application of FRET-FLIM enormously [95]. Nowadays fluorescent-tagging of proteins no longer depends on laborious protein pu-

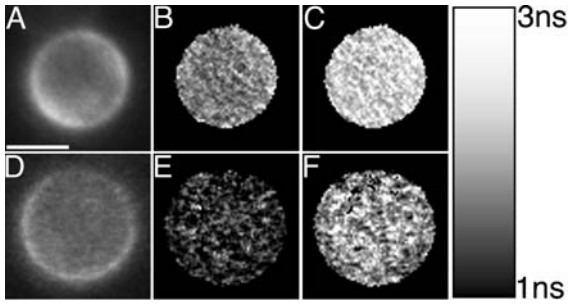


Fig. 3A–F FLIM-FRET analysis of Cowpea protoplasts co-expressing: **A–C** a myristoylated CFP and unfused YFP (control); **D–F** a myristoylated YFP and a prenylated CFP (FRET situation). Images A and D are CFP fluorescence intensity images, B and E are micrographs of τ_φ of CFP, whereas C and F are micrographs of τ_M of CFP. All lipidated constructs are located at the plasma membrane, whereas unfused YFP (not shown) is located in the cytosol and does not interact with the CFP-construct. The mean value of τ_φ and τ_M are 2.04 ns, 2.62 ns, 1.35 ns and 2.14 ns, respectively for images B, C, E and F. Hence, from τ_φ and τ_M a FRET efficiency of 34% and 18% is calculated using Eq. (14), respectively. The lifetime values are indicated in a *greyscale* ranging from *black* (1 ns) to *white* (3 ns) as indicated next to the figure. Images courtesy of Joop Vermeer. For more details see Vermeer et al. [108]. *Bar* indicates 10 μm

rification, chemical labeling, and microinjection back into the cells, but just can be accomplished using transfection or stable transformation of cells. In addition to the ease of producing fluorescently-tagged proteins in living cells, usage of genetically-encoded signal-sequences and cell-type specific promotors enables control over subcellular localization (organalle targeting) and cell-type specific expression in tissues. GFP can be used as donor with Cy3 as acceptor [67]. However, very popular is the completely genetic encoded FRET-pair CFP/YFP (see Fig. 3 [61] for a review, [66, 93] for methodologic examples, and many of the above-mentioned protein-protein FRET-FLIM studies as other examples). The availability of monomeric versions of the orange-red fluorescing proteins from corals [96] will further expand the palette of available genetic encoded fluorophores for FRET, and potentially should enable triple-FRET or even dual donor-acceptor FRET involving four colors.

3.1.5 Quantitative FRET-FLIM Analysis

Because FRET-FLIM is so widely applied in biology, it has also received special attention with respect to quantitative data analysis. Usually, the biologist is not per se interested in the actual lifetime values, but he/she is interested in the degree of complex formation and the proximity/orientation of the donor/acceptor pair in the complex. FLIM enables not just a contrast in these parameters but can provide quantitative estimation of the degree of complex formation and intermolecular distances in situ. To this end, special analysis routines have been developed based on (global) analysis of multiple lifetimes in images using a

priori knowledge on the biology: e.g. assuming two lifetime states: a non-interacting unquenched donor and an interacting quenched donor [51, 97, 98]. Using such algorithms images of the fraction of activated (phosphorylated) receptors have been produced [52, 97]. Also new FLIM-FRET techniques based on in-growth of the sensitized emission [99], monitoring time-resolved anisotropy [41, 100], or combining spectral information with lifetime information [58, 59, 101] adding new parameters for monitoring FRET, will undoubtedly contribute to further improve accuracy in quantitative FRET-estimation.

3.1.6 Comparison of FLIM with Other FRET Microscopy Techniques

FLIM is not the only technique for observing FRET. If both donor and acceptor chromophores are fluorophores, FRET can also be estimated with steady state techniques, including acceptor photobleaching [102, 103], spectral imaging [61, 104] and ratio-imaging (or filter-FRET) [105, 106] techniques. Recently all possible modes of FRET-microscopy (also including yet unexplored FRET-imaging modes) have been reviewed by Erijman and Jovin [107].

With each method of FRET estimation there are advantages and disadvantages. Here, specifically the three most widely used other FRET-imaging techniques besides FRET-FLIM: i) acceptor-bleaching, ii) filter-FRET and iii) spectral imaging are briefly discussed in relation to FLIM. Also elsewhere, reviews and experimental studies have appeared comparing FLIM with other FRET-microscopy techniques [61, 63, 108].

In acceptor-photobleaching the donor intensity is measured before and after selective photodestruction of the acceptor [103]. The increase in donor fluorescence upon acceptor bleaching is proportional to E. Advantages of acceptor-bleaching are the relative ease of performing the experiment and the availability of internal control areas in one image (by selective bleaching of areas in the image). The problems of using this technique are i) its inherent irreversibility and the induced photodamage, ii) the necessity of a (second) selective laser source that is absorbed by the acceptor but not by the donor (e.g. 514 nm-Ar-laser for the YFP/CFP couple), iii) sensitivity of the FRET-estimation to donor-bleaching in taking the donor images before and after bleaching, iv) movement of the donor molecule between the acquisition of the pre- and post-bleach donor images will introduce artifacts, and v) the method is sensitive to photo chromicity of the chromophores that sometimes can be switched to dark (low quantum yield) but absorbing states or to states displaying blue shifted fluorescence spectra. Especially for GFPs the latter phenomena have been described [109]. It is of note that acceptor bleaching can be combined with FLIM. In this case not the increase in donor intensity but the increase in donor-lifetime upon acceptor bleaching is monitored [67]. In this way the method becomes more insensitive towards slight donor-bleaching.

Filter-FRET (or ratio-imaging) methods have also become quite popular for measuring FRET. Usually three images with different filter sets are obtained:

a donor-image (donor excitation-donor emission wavelengths), an acceptor image (acceptor excitation-acceptor emission wavelengths) and a 'FRET-image' (donor excitation-acceptor emission wavelength). The true FRET contribution is calculated from these three images by correcting the 'FRET-image' for contributions of residual donor-emission and direct acceptor-excitation yielding a 'pure' sensitized acceptor-emission image [105, 106]. This procedure also requires control measurements yielding correction factors. Advantage of the method is its ease of implementation and speed, but disadvantages are the high sensitivity to errors caused by background fluorescence, excitation intensity fluctuations, (partial) photobleaching, pixel shifts (due to filter changes) and possible inner filtering by objects in the specimen. This severely complicates (and in most cases prohibits) quantitative estimation of E as is possible in FRET-FLIM.

Fluorescence spectral imaging microscopy (FSPIM) combines spatial resolution with spectral resolution. At every position (or across a line) within the specimen a complete spectrum is obtained [110]. Advantages of FSPIM are that by obtaining complete emission spectra, it gives a fast qualitative estimate for the presence of FRET: one can quickly distinguish FRET from non-FRET spectra just by scoring sensitized acceptor emission vs donor emission contributions in the emission spectra. Also errors due to wrong filter selection, autofluorescence contributions or presence of other dyes are quickly recognized from the shape of the spectra but may be more difficult to see using filter-FRET methods. Disadvantages are that at high acceptor concentration the direct acceptor excitation at the donor excitation wavelength can yield false-positive FRET signals, making the technique less suitable for quantitative FRET-estimation. However, by performing dual excitation and spectral decomposition procedures [104] one can calculate FRET-efficiencies.

The advantages of FLIM with respect to the other methods of FRET estimation are its robustness (less sensitivity to artifacts) and the possibilities for quantitative estimation of E. Especially in situations i) with unknown relative concentrations of donor and acceptor fluorophores (as for interaction studies), ii) with a non-specific acceptor interacting with more targets than the donor [59] and in situations where a quantitative estimate of E is required, FRET-FLIM and acceptor-bleaching are the methods of choice. If FRET-estimation as a function of time is required for one living cell, acceptor-bleaching is not suitable. In this case FRET-FLIM, but also filter-FRET or spectral measurements are the methods of choice (the latter two especially if qualitative comparison is sufficient). If the relative donor -acceptor abundancy is known (e.g. by usage of a FRET-indicator with a 1:1 donor: acceptor stoichiometry such as yellow-cameleons [111]), spectral methods, combining donor quenching and acceptor sensitization (FSPIM, filter-FRET), are preferred over the donor methods that only register donor quenching (FLIM, acceptor bleaching). Undoubtedly, multimode FRET-imaging, combining lifetime and spectral modalities will further improve the accuracy of FRET-estimation. Recently, a number of such systems have been described [58, 59, 101, 108].

3.2 Ion Imaging

By using fluorescent dyes that change their quantum-yield upon binding (or chelating) of an ion or small soluble molecules, local concentrations of such ions or solutes can be imaged using FLIM. For instance calcium can be quantified by measuring increased fluorescence lifetimes of calcium-green or Quin-2 (blue excitation) [10, 112–114] or Calcium-Crimson (orange excitation) [115] upon calcium-binding. This avoids UV-excitation as is necessary for exciting the conventional ratiometric calcium dyes as INDO and FURA.

Also pH-imaging has been described with FLIM using chemical dyes that display a pH-dependent shift in lifetimes such as carboxy-SNAFL-1 and -SNAFL-2 [114, 116], [117, 118], BCECF [119, 120], carboxyfluorescein [121] and the novel low-PH LysoSensor probes [122], for a comparison see also [123].

Recently a lifetime-probe (Newport Green DCF) for imaging Zn^{2+} has been described [124], and also reports have appeared of non-ratiometric lifetime probes for quantifying Na^+ [125] and for Mg^{2+} (magnesium-green, Mag-quin-2, Mag-quin-1), Cl^- (MQAE, SPQ), heavy metals (bis-BTC) and K^+ (PBFI) [126].

For many of these ions there are also ratiometric dyes and genetic-encoded sensors. If available, the use of ratiometric dyes that change their excitation or emission spectrum is often preferable for imaging of ions or small solutes since it allows a faster and more easy way of detection. In many studies both ratiometric approaches and FLIM approaches for ion imaging have been compared (see references above). Well known chemical probes for ratiometric ion-imaging are INDO and Fura for calcium-imaging and BCECF and SNAFL for pH-imaging. Recently, also a variety of genetically-encoded ratiometric sensors for measuring ion-concentration or small molecules using mutated GFPs, or FRET-based sensors using fusion proteins containing CFP/YFP have become available, e.g. for calcium-imaging using cameleons [111], for Zn^{2+}-imaging [127], for pH-imaging using pHluorin [128], and Cl^-/pH sensitive versions of YFP [129, 130] and FRET-sensors for cAMP [131], cGMP [132] and nitric oxide [133]. The enormous advantage of genetic encoded sensors is the ease of loading and specific targeting to the desired cell type or subcellular location. It is expected that tandem GFP-fusions of newly engineered specific binding proteins or peptides will further expand the possibility of ratiometric imaging; for a recent review see [134].

3.3 Oxygen Imaging

Oxygen is a well-known collisional quencher of fluorescence. Oxygen-quenching also decreases the fluorescence lifetime and in this way oxygen concentrations can be measured inside cells by FLIM. The sensitivity for oxygen becomes larger for fluorophores (or luminophores) with large fluorescence lifetimes. Especially long-lived ruthenium dyes (τ>800 µs) have been used as sensitive

reporters of oxygen using FLIM [135, 136]. Using such long-lived luminophores requires special FLIM-equipment, e.g. operating at much lower pulse repetition rates with long detection intervals or at very low modulation-frequencies (kHz range). Recently, a long-lived palladium-porphyrin compound was also applied for FLIM-measurements of the oxygen tension in mouse retina [137].

3.4
Probing the Microenvironment, Contrast Generation and Quantitative Imaging

Especially small chemical dyes can exhibit marked differences in fluorescence quantum yield or lifetime in different environments. As discussed above, differences in quantum yield are very difficult to measure with normal fluorescence microscopy since it cannot be imaged independently from dye-concentration. With FLIM, the sensitivity of chemical dyes can be exploited to selectively probe the microenvironment around the dye. A well known example is the hydrophobic DNA-binding dye ethidium blue which has a marked higher lifetime when bound to DNA as compared to membranes. This property has been used for discriminating between membrane-bound and DNA-bound ethidium blue in FLIM-micrographs [20]. Also lifetime-contrast has been employed to study the chromatin-conformation in living cells using DNA-binding fluorescent dyes. Examples are SYTO13 [138], YOYO-1 [90]. FRET-FLIM studies of chromatin structure have also been reported using differentially-colored DNA-binding dyes [89, 91].

Binding of fluorescent substrates to pockets inside enzymes can cause a change in fluorescence lifetime. For instance, the fluorescence lifetime of NADH increases from 0.5 to 1 ns upon binding to proteins, enabling imaging of free and protein-bound NADH by FLIM [139]. Also binding of nonfluorescent substrates to autofluorescent enzymes can be imaged by exploiting autofluorescence lifetime-changes upon substrate binding. An example of this is FLIM-imaging of binding of a substrate to a flavin-containing autofluorescent redox-enzyme [140].

For quantitative imaging of distribution of fluorescent components in images it is of importance to correct for deviations in local fluorescent quantum yields. Especially for fluorescent lipids that are known to display marked reduction of fluorescent quantum yield upon microaggregation in membranes FLIM can be used to assist quantitative imaging. As such FLIM has been used to confirm that intensity-variations in fluorescence micrographs were due to concentration variations and not to lifetime-differences of fluorescently-tagged lipids in spermatozoa [141, 142]. If aggregation decreases the fluorescent lifetime, FLIM can also be used to study the aggregation state of molecules as shown by Jakobs et al. [143] for DsRed aggregation in *E. coli*.

Since GFPs display a relatively constant lifetime in cells (unless they are involved in FRET), FLIM has been applied for colocalization studies of spectrally similar GFPs by exploiting the lifetime-difference between these GFP-variants [144]. By assuming constant lifetimes for each GFP-variant, global analysis

methods can give quantitative distribution maps of coexpressed GFPs with similar spectral properties [52]. However, it is of note, that biophysical parameters such as refractive index can affect the lifetime, even of the very much shielded chromophore of GFP [145].

Interestingly, the lifetime-contrast can be used to obtain superresolution in co-localization studies for spectrally distinct fluorophores. Heilemann et al. [146] showed that by assigning lifetimes to Cy5 (2 ns) and to JF9 (4 ns) a pattern-recognition technique can yield a spatial resolution below 30 nm.

The lifetime contrast is also useful for studying autofluorescence distribution in tissues or skin. Many biomolecules contribute to autofluorescence such as NAD(P)H, porphyrin, flavin, collagen, elastin and melanin. Since the lifetimes differ for these compounds, contrast can be generated in complex biological samples without (fluorescence) staining. This lifetime-contrast has been used for 3D-imaging of skin [147, 148] and in studies of oxidative stress in CHO-cells [149]. Since in most cases the exact molecular nature of the autofluorescence is unknown and given the typical multicomponent nature of the autofluorescence emission, stretched-exponential fitting routines have been described for analyzing autofluorescence lifetime-images [150], enabling faster analysis, less noise in the lifetime images and a more easy way of representing lifetime-heterogeneity.

3.5 Medical Diagnosis

It has been found that tumors can be detected in a very early stage of development by pretreatment with porphyrin photosensitizers. The tumor-cells display a marked increased fluorescence lifetime as compared to surrounding control cells [151–154]. Also lifetime-differences in normal autofluorescence have been detected between normal and tumor-cells. Recent studies revealed that FLIM could assist in histopathological assessment of fixed (unstained) breast cancer tissues. The autofluorescence lifetime contrast revealed statistically relevant differences ($p<0.05$) between benign and malignant stroma samples [155].

The notion that lifetime-contrast can be used to detect tumor cells, inspired the development of new endoscopes using lifetime-contrast for screening patients. Many endoscopes (fluorescence lifetime imaging endoscopy, FLIE) have been implemented in the frequency domain [12, 156, 157] allowing the use of phase-suppression techniques enabling real-time (8 Hz frame rate) display of lifetime-contrast. Since it is easy to simultaneously equip the endoscopes with a high-power IR-laser to immediately burn away the tumor upon its identification, FLIE enables a fast and relatively non-invasive method for diagnosis and treatment of small tumors. Recently, a FLIE-instrument has been described based on the time-domain approach [158]. Instrumental was the availability of a compact picosecond diode laser as excitation source. Using this device Siegel and colleagues studied animal tissue sections, knee joints and human teeth and they could obtain functional contrast using fluorescence lifetime imaging.

Analysis of the induced autofluorescence using stretched exponentials [150] proved useful for this time-domain FLIE instrument [158]. Also in an earlier study, FLIM was used for the detection of early stages of dental caries or plaque formation [159].

4 Conclusions and Outlook

As clearly demonstrated from the diversity in application areas, it can be stated that FLIM is a robust and mature technique. For the quantitative analysis of molecular interactions in situ, FLIM-FRET has become an established technique in cell biology. Nowadays, in many cases, cytological proof of molecular interactions is required for publication of manuscripts in high quality life sciences-journals, making FLIM a pivotal technique for complementing traditional techniques for measuring protein-protein interactions such as immunoprecipitation, affinity chromatography or yeast-two-hybrid screens (not providing such in situ information). Consequently, we expect that the largest impact of FLIM will come from the further proliferation of FLIM into biological and biomedical sciences rather than from development of novel and improved FLIM-instruments. For this reason, increasing effort is being put in making FLIM available for biomedical users, both by the research groups that develop FLIM instrumentation, as well as by commercial parties, like microscope manufacturers. To make this a success, easy (automated) operation of commercial instrumentation and straightforward unambiguous data analysis capable of fast and reproducible extraction of quantitative information from the specimen will be of vital importance. In these areas there is still a lot to gain using combinations of existing approaches. Together with the increasing availability of genetically-encoded protein-modules and biosensors for all sorts of signaling cascades and other interesting physiological processes, we expect a bright future for FLIM.

Acknowledgements The authors would like to thank Joop Vermeer for providing the FLIM images used in Fig. 3.

References

1. Gaviola Z (1926) Z Phys 42:853
2. Venetta BD (1959) Rev Sci Instrum 30:450
3. Wang XF, Uchida T, Minami S (1989) Appl Spectrosc 43:840
4. Wood RW (1921) Proc R Soc London A 99:362
5. Abraham, Lemoine (1899) CR Hebd Seance Acad Sci 129:206
6. Lakowicz JR (1999) In: Lakowicz JR (ed) Principles of fluorescence spectroscopy. Kluwer Academic, New York, p 141
7. Spencer RD, Weber G (1969) Ann NY Acad Sci 158:361

8. Squire A, Bastiaens PIH (1999) J Microsc 193:36
9. Carlsson K, Liljeborg A (1997) J Microsc 185:37
10. So PTC, French PMW, Yu W, Berland KM, Dong CY, Gratton E (1995) Bioimaging 3:49
11. Dong CY, Buehler C, So PTC, French T, Gratton E (2001) Appl Opt 40:1109
12. Schneider PC, Clegg RM (1997) Rev Sci Instrum 68:4107
13. Gadella TWJ (1999) In: Mason WT (ed) Fluorescent and luminescent probes. Academic Press, New York, p 467
14. Gadella TWJ, Jovin TM, Clegg RM (1993) Biophys Chem 48:221
15. Lakowicz JR (1999) In: Lakowicz JR (ed) Principles of fluorescence spectroscopy. Kluwer Academic, New York, p 95
16. Bennet RG (1960) Rev Sci Instrum 31:1275
17. O'Connor DV, Phillips D (1984) Time-correlated single photon counting. Academic Press, New York
18. Emiliani V, Sanvitto D, Tramier M, Piolot T, Petrasek Z, Kemnitz K, Durieux C, Coppey-Moisan M (2003) Appl Phys Lett 83:2471
19. Lytle FE, Parrish RM, Barnes WT (1985) Appl Spectrosc 39:444
20. Dong CY, So PT, French T, Gratton E (1995) Biophys J 69:2234
21. Dong CY, Buehler C, So PTC, Gratton E (1997) Biophys J 72:TU422
22. Mitchell AC, Wall JE, Murray JG, Morgan CG (2002) J Microsc 206:225
23. Mitchell AC, Wall JE, Murray JG, Morgan CG (2002) J Microsc 206:233
24. Pawley JB (1995) Handbook of biological confocal microscopy. Plenum Press, New York
25. Denk W, Piston DW, Webb WW (1995) In: Pawley JB (ed) Handbook of biological confocal microscopy. Plenum Press, New York
26. Denk W, Strickler JH, Webb WW (1990) Science 248:73
27. Neil MAA, Squire A, Juskaitis R, Bastiaens PIH, Wilson T (2000) J Microsc 197:1
28. Petran N, Hadravsky M, Egger MD, Galambos R (1968) J Opt Soc Am 58:661
29. Brakenhoff GJ, Visscher K (1992) J Microsc 165:139
30. Hanley Q, Verveer PJ, Gemkow MJ, Arndt-Jovin D, Jovin TM (1999) J Microsc 196:317
31. Beaumont V (2003) Biochem Soc Trans 31:819
32. Axelrod D (2003) Biophotonics B (Methods Enzymol)361:1
33. Carlsson K, Philip J (2002) Proc SPIE-Int Soc Opt Eng 4622:70
34. Gerritsen HC, Asselbergs MAH, Agronskaia AV, Van Sark W (2002) J Microsc 206:218
35. Lambert Instruments: www.lambert-instruments.com
36. LaVision: http://www.lavision.de/
37. Becker & Hickl: http://www.becker-hickl.de
38. Gratton E, van de Ven MJ (1995) In: Pawley JB (ed) Handbook of biological confocal microscopy. Plenum Press, New York, p 69
39. Chang IC (1995) In: Bass M (ed) Handbook of optics, vol 2. McGraw-Hill, New York, p 12.1
40. Maldonado TA (1995) In: Bass M (ed) Handbook of optics, vol 2. McGraw-Hill, New York, p 13.1
41. Lidke DS, Nagy P, Barisas BG, Heintzmann R, Post JN, Lidke KA, Clayton AH, Arndt-Jovin DJ, Jovin TM (2003) Biochem Soc Trans 31:1020
42. Carlsson K, Liljeborg A (1998) J Microsc 191:119
43. Iwai T, Asakura T (1996) Proceedings of the IEEE 84:765
44. Herman P, Maliwal BP, Lin HJ, Lakowicz JR (2001) J Microsc 203:176
45. van Geest LK (2003) J Fluoresc 13:365
46. van der Oord CJR, Stoop KWJ, van Geest LK (2001) Proc SPIE-Int Soc Opt Eng 4252: 115

47. Shaw PJ (1995) In: Pawley JB (ed) Handbook of biological confocal microscopy. Plenum Press, New York, p 373
48. Agard DA, Sedat JW (1983) Nature 302:676
49. Castleman KR (1979) Digital image processing. Prentice-Hall, New Jersey
50. Squire A, Bastiaens PIH (1999) J Microsc 193:36
51. Verveer PJ, Bastiaens PIH (2002) J Microsc 209:1
52. Verveer PJ, Squire A, Bastiaens PIH (2000) Biophys J 78:2127
53. Lakowicz JR (1999) Principles of fluorescence spectroscopy. Kluwer Academic, New York
54. Gadella TWJ, Jovin TM (1995) J Cell Biol 129:1543
55. Clayton AH, Hanley QS, Arndt-Jovin DJ, Jovin TM (2001) Biophys J 80:656.35
56. Clayton AHA, Hanley QS, Arndt-Jovin DJ, Subramaniam V, Jovin TM (2002) Biophys J 83:1631
57. Siegel J, Suhling K, Leveque-Fort S, Webb SED, Davis DM, Phillips D, Sabharwal Y, French PMW (2003) Rev Sci Instrum 74:182
58. Tinnefeld P, Herten DP, Sauer M (2001) J Phys Chem 105:7989
59. Hanley QS, Arndt-Jovin DJ, Jovin TM (2002) Appl Spectrosc 56:155
60. Clegg RM, Holub O, Gohlke C (2003) Methods Enzymol 360:509
61. Gadella TWJ, van der Krogt GNM, Bisseling T (1999) Trends Plant Sci 4:287
62. Wu PG, Brand L (1994) Anal Biochem 218:1
63. Chen Y, Mills JD, Periasamy A (2003) Differentiation 71:528
64. Bastiaens PI, Squire A (1999) Trends Cell Biol 9:48
65. Periasamy A (2001) J Biomed Opt 6:287
66. Krishnan RV, Masuda A, Centonze VE, Herman B (2003) J Biomed Opt 8:362
67. Wouters FS, Bastiaens PIH (1999) Curr Biol 9:1127
68. Verveer PJ, Wouters FS, Reynolds AR, Bastiaens PIH (2000) Science 290:1567
69. Tertoolen LG, Blanchetot C, Jiang G, Overvoorde J, Gadella TW Jr, Hunter T, den Hertog J (2001) BMC Cell Biol 2:8
70. Bastiaens PIH, Jovin TM (1996) Proc Natl Acad Sci USA 93:8407
71. Ng T, Squire A, Hansra G, Bornancin F, Prevostel C, Hanby A, Harris W, Barnes D, Schmidt S, Mellor H, Bastiaens PI, Parker PJ (1999) Science 283:2085
72. Ng T, Shima D, Squire A, Bastiaens PI, Gschmeissner S, Humphries MJ, Parker PJ (1999) EMBO J 18:3909
73. Anilkumar N, Parsons M, Monk R, Ng T, Adams JC (2003) EMBO J 22:5390
74. Ng T, Parsons M, Hughes WE, Monypenny J, Zicha D, Gautreau A, Arpin M, Gschmeissner S, Verveer PJ, Bastiaens PI, Parker PJ (2001) EMBO J 20:2723
75. Legg JW, Lewis CA, Parsons M, Ng T, Isacke CM (2002) Nat Cell Biol 4:399
76. Immink RG, Gadella TW Jr, Ferrario S, Busscher M, Angenent GC (2002) Proc Natl Acad Sci USA 99:2416
77. Elangovan M, Day RN, Periasamy A (2002) J Microsc 205:3
78. Chen Y, Periasamy A (2004) Microsc Res Tech 63:72
79. Larijani B, Hume AN, Tarafder AK, Seabra MC (2003) J Biol Chem 278:46798
80. Berezovska O, Bacskai BJ, Hyman BT (2003) Sci Aging Knowledge Environ 2003:PE14
81. Berezovska O, Ramdya P, Skoch J, Wolfe MS, Bacskai BJ, Hyman BT (2003) J Neurosci 23:4560
82. Ramdya P, Skoch J, Bacskai BJ, Hyman BT, Berezovska O (2003) J Neurochem 87:843
83. Lennon NJ, Kho A, Bacskai BJ, Perlmutter SL, Hyman BT, Brown RH Jr (2003) J Biol Chem 278:50466
84. van Kuppeveld FJ, Melchers WJ, Willems PH, Gadella TW Jr (2002) J Virol 76:9446
85. Herreros J, Ng T, Schiavo G (2001) Mol Biol Cell 12:2947
86. Larijani B, Allen-Baume V, Morgan CP, Li M, Cockcroft S (2003) Curr Biol 13:78

87. Hughes WE, Larijani B, Parker PJ (2002) J Biol Chem 277:22974
88. Gadella TWJ, Arndt-Jovin D, Jovin TM (1994) J Fluoresc 4:295
89. Murata S, Herman P, Lakowicz JR (2001) Cytometry 43:94
90. Murata S, Herman P, Lakowicz JR (2001) J Histochem Cytochem 49:1443
91. Murata S, Herman P, Lin H, Lakowicz JR (2000) Cytometry 41:178
92. Cremazy F, Manders E, Bastiaens P, Hager G, Van Munster E, Gadella T, van Driel R (2004) (submitted for publication)
93. Bacskai BJ, Skoch J, Hickey GA, Allen R, Hyman BT (2003) J Biomed Opt 8:368
94. Oida T, Sako Y, Kusumi A (1993) Biophys J 64:676
95. Tsien RY (1998) Annu Rev Biochem 67:509
96. Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, Tsien RY (2002) Proc Natl Acad Sci USA 99:7877
97. Verveer PJ, Squire A, Bastiaens PIH (2001) J Microsc 202:451
98. Clayton AH, Hanley QS, Verveer PJ (2004) J Microsc 213:1
99. Harpur AG, Wouters FS, Bastiaens PIH (2000) Nat Biotechnol 19:167
100. Clayton AH, Hanley QS, Arndt-Jovin DJ, Subramaniam V, Jovin TM (2002) Biophys J 83:1631
101. Neher R, Neher E (2004) J Microsc 213:46
102. Bastiaens P, Majoul I, Verveer PJ, Soling H, Jovin T (1996) EMBO J 15:4246
103. Bastiaens PIH, Jovin TM (1998) In: Celis JE (ed) Cell biology: a laboratory handbook, vol 3. Academic Press, London, p 136
104. Zimmermann T, Rietdorf J, Girod A, Georget V, Pepperkok R (2002) FEBS Lett 531:245
105. Gordon GW, Berry XH, Liang B, Levine B, Herman B (1998) Biophys J 74:2702
106. Hoppe A, Christensen K, Swanson JA (2002) Biophys J 83:3652
107. Jares-Erijman EA, Jovin TM (2003) Nat Biotechnol 21:1387
108. Vermeer JEM, van Munster EB, Vischer NO Jr, Gadella TWJ (2004) J Microsc (in press)
109. Creemers TMH, Lock AJ, Subramaniam V, Jovin TM, Völker S (2000) Proc Natl Acad Sci USA 97:2974
110. Balaban RS, Kurtz I, Cascio HE, Smith PD (1986) J Microsc 141:31
111. Miyawaki A, Griesbeck O, Heim R, Tsien RY (1999) Proc Natl Acad Sci USA 96:2136
112. Lakowicz JR, Szmacinski H, Nowaczyk K, Johnson ML (1992) Cell Calcium 13:131
113. Lakowicz JR, Szmacinski H, Nowaczyk K, Lederer WJ, Kirby MS, Johnson ML (1994) Cell Calcium 15:7
114. Sanders R, Gerritsen HC, Draaijer A, Houpt PM, Levine YK (1994) SPIE Proc 2197:56
115. Periasamy A, Wodnicki P, Wang XF, Kwon S, Gordon GW, Herman B (1996) Rev Sci Instrum 67:3277
116. Sanders R, Draaijer A, Gerritsen HC, Houpt PM, Levine YK (1995) Anal Biochem 227:302
117. Andersson RM, Carlsson K, Liljeborg A, Brismar H (2000) Anal Biochem 283:104
118. Carlsson K, Liljeborg A, Andersson RM, Brismar H (2000) J Microsc 199:106
119. Hanson KM, Behne MJ, Barry NP, Mauro TM, Gratton E, Clegg RM (2002) Biophys J 83:1682
120. Behne MJ, Meyer JW, Hanson KM, Barry NP, Murata S, Crumrine D, Clegg RW, Gratton E, Holleran WM, Elias PM, Mauro TM (2002) J Biol Chem 277:47399
121. Vroom JM, De Grauw KJ, Gerritsen HC, Bradshaw DJ, Marsh PD, Watson GK, Birmingham JJ, Allison C (1999) Appl Environ Microbiol 65:3502
122. Lin HJ, Herman P, Kang JS, Lakowicz JR (2001) Anal Biochem 294:118
123. Lin HJ, Herman P, Lakowicz JR (2003) Cytometry 52A:77
124. Thompson RB, Peterson D, Mahoney W, Cramer M, Maliwal BP, Suh SW, Frederickson C, Fierke C, Herman P (2002) J Neurosci Methods 118:63

125. Szmacinski H, Lakowicz JR (1997) Anal Biochem 250:131
126. Szmacinski H, Lakowicz JR, Johnson ML (1994) Methods Enzymol 240:723
127. Miyawaki A, Tsien RY (2000) Methods Enzymol 327:472
128. Miesenbock G, De Angelis D, Rothman J (1998) Nature 394:192
129. Metzger F, Repunte-Canonigo V, Matsushita S, Akemann W, Diez-Garcia J, Ho C, Iwasato T, Grandes P, Itohara S, Joho R, Knopfel T (2002) Eur J Neurosci 15:40
130. Nagai T, Ibata K, Park E, Kubota M, Mikoshiba K, Miyawaki A (2002) Nat Biotechnol 20:87
131. Zaccolo M, De Giorgi F, Cho C, Feng L, Knapp T, Negulescu P, Taylor S, Tsien R, Pozzan T (2000) Nature Cell Biol 2:25
132. Honda A, Adams S, Sawyer C, Lev-Ram V, Tsien R, Dostmann W (2001) Proc Natl Acad Sci USA 98:2437
133. Pearce L, Gandley R, Han W, Wasserloos K, Stitt M, Kanai A, McLaughlin M, Pitt B, Levitan E (2000) Proc Natl Acad Sci USA 97:477
134. Miyawaki A (2003) Curr Opin Neurobiol 13:591
135. Hartmann P, Ziegler W, Lubbers DW (1997) Adv Exp Med Biol 428:605
136. Gerritsen HC, Sanders R, Draaijer A, Ince C, Levine YK (1997) J Fluoresc 7:11
137. Shonat RD, Kight AC (2003) Ann Biomed Eng 31:1084
138. van Zandvoort MA, de Grauw CJ, Gerritsen HC, Broers JL, Egbrink MG, Ramaekers FC, Slaaf DW (2002) Cytometry 47:226
139. Lakowicz JR, Szmacinski H, Nowaczyk K, Johnson ML (1992) Proc Natl Acad Sci USA 89:1271
140. Gadella TWJ (1997) Microsc Anal 13
141. Gadella BM, Lopes-Cardozo M, van Golde LM, Colenbrander B, Gadella TW Jr (1995) J Cell Sci 108(3):935
142. Gadella BM, Gadella TW Jr, Colenbrander B, van Golde LM, Lopes-Cardozo M (1994) J Cell Sci 107(8):2151
143. Jakobs S, Subramaniam V, Schonle A, Jovin TM, Hell SW (2000) FEBS Lett 479:131
144. Pepperkok R, Squire A, Geley S, Bastiaens PIH (1999) Curr Biol 9
145. Suhling K, Siegel J, Phillips D, French PM, Leveque-Fort S, Webb SE, Davis DM (2002) Biophys J 83:3589
146. Heilemann M, Herten DP, Heintzmann R, Cremer C, Muller C, Tinnefeld P, Weston KD, Wolfrum J, Sauer M (2002) Anal Chem 74:3511
147. Konig K, Riemann I (2003) J Biomed Opt 8:432
148. Masters BR, So PT, Gratton E (1998) Ann NY Acad Sci 838:58
149. Konig K, So PT, Mantulin WW, Tromberg BJ, Gratton E (1996) J Microsc 183(3):197
150. Lee KC, Siegel J, Webb SE, Leveque-Fort S, Cole MJ, Jones R, Dowling K, Lever MJ, French PM (2001) Biophys J 81:1265
151. Cubeddu R, Canti G, Taroni P, Valentini G (1993) Photochem Photobiol 57:480
152. Kohl M (1993) Appl Phys B 56:131
153. Schneckenburger H, König K, Dienerberger T, Hahn R (1994) Opt Eng 33:2600
154. Andersson-Engels S, Canti G, Cubeddu R, Eker C, Klinteberg C, Pifferi A, Svanberg K, Svanberg S, Taroni P, Valentini G, Wang I (2000) Lasers Surg Med 26:76
155. Tadrous PJ, Siegel J, French PM, Shousha S, Lalani N, Stamp GW (2003) J Pathol 199:309
156. Wagnieres G, Mizeret J, Studzinsky A, van den Bergh H (1997) J Fluoresc 7:75
157. Mizeret J, Wagnieres G, Stepinac T, van den Berg H (1997) Lasers Med Sci 12:209
158. Siegel J, Elson DS, Webb SE, Lee KC, Vlandas A, Gambaruto GL, Leveque-Fort S, Lever MJ, Tadrous PJ, Stamp GW, Wallace AL, Sandison A, Watson TF, Alvarez F, French PM (2003) Appl Opt 42:2995
159. Konig K, Schneckenburger H, Hibst R (1999) Cell Mol Biol (Noisy-le-grand) 45:233

160. Lakowicz JR, Berndt K (1991) Rev Sci Instrum 62:1727
161. van Munster EB, Gadella TWJ (2004) J Microsc 213:29
162. Gratton E, Breusegem S, Sutin J, Ruan Q, Barry N (2003) J Biomed Opt 8:381
163. Minami S (1990) J Photochem Photobiol A 53:11
164. Cubeddu R (1991) Proc SPIE-Int Soc Opt Eng 1525:17
165. Ni T, Melton L (1991) Appl Spectrosc 45:938
166. Buurman EP, Sanders R, Draaijer A, Gerritsen HC, van Veen JJF, Houpt PM, Levine YK (1992) Scanning 14:155
167. Piston DW, Sandison DR, Webb WW (1992) Proc SPIE-Int Soc Opt Eng 1640:379
168. Straub M, Hell SW (1998) Appl Phys Lett 73:1769
169. Straub M, Hell SW (1998) Bioimaging 6:177
170. Systsma J, Vroom JM, De Grauw CJ, Gerritsen HC (1998) J Microsc 191:39
171. Cole MJ, Siegel J, Webb SED, Jones R, Dowling K, French PMW (2000) Opt Lett 25:1361
172. Webb SED, Gu Y, Leveque-Fort S, Siegel J, Cole MJ, Dowling K, Jones R, French PMW (2002) Rev Sci Instrum 73:1898
173. Fujiwara N, Tsukahara S, Watarai H (2001) Langmuir 17:5337
174. Schneckenburger H, Stock K, Strauss WSL, Eickholz J, Sailer R (2003) J Microsc 211:30
175. Buist AH, Müller M, Gijsbers EJ, Brakenhoff GJ, Sosnowski TS, Norris TB, Squier J (1997) J Microsc 186:212
176. Müller M, Ghauharali R, Visscher K, Brakenhoff GJ (1995) J Microsc 177:171

Adv Biochem Engin/Biotechnol (2005) 95: 177–199
DOI 10.1007/b102214

Fluorescence Recovery after Photobleaching: Application to Nuclear Proteins

Adriaan B. Houtsmuller (✉)

Department of Pathology, Josephine Nefkens Institute, Erasmus MC,
University Medical Centre Rotterdam, PO-Box 1738, 3000 DR Rotterdam, The Netherlands
a.houtsmuller@erasmusmc.nl

Abstract Fluorescence redistribution after photobleaching (FRAP) has received increasing attention ever since it was first introduced into cell biological research. The method was developed in the 1970s, when its biological application mainly focused on the mobility of fluorescently labelled constituents of the cell membrane. The development of confocal scanning microscopy in the 1980s facilitated accurate investigation of the behaviour of molecules in the inside of cells without specialised equipment. However, FRAP did not yet become as popular as it is today, probably because of the dedicated and time-consuming methodology required to purify and label proteins or other compounds and, moreover, to inject them into cells. The revolution created by the development of GFP-technology finally lead to a tremendous boost of FRAP applications in studying the behaviour of proteins in the living cells. Finally, the ongoing increase of speed and memory of personal computers allows computer modelling of FRAP experiments for analysis of complex 3-D FRAP data, and for the development of new FRAP assays. Here we discuss several variants of FRAP on the basis of its application to the investigation of the behaviour of proteins in the living cell nucleus.

Keywords FRAP · Green fluorescent protein · Protein mobility · Cell nucleus · Protein-DNA interaction · DNA repair · Transcription · Replication

List of Abbreviations

AR Androgen receptor
ER Estrogen receptor
FLIP Fluorescence loss in photobleaching
FRAP Fluorescence recovery after photobleaching
FRET Fluorescence resonance energy transfer
GR Glucocorticoid receptor
NER Nucleotide excision repair
UV Ultra-violet

1 Introduction

In the past decade fluorescence recovery after photobleaching (FRAP) has received tremendously increased attention, especially in the field of live cell investigation of protein mobility. FRAP makes use of the property of fluorescent molecules that they not only can be visualised in fluorescence microscopes, but also can be photobleached, i.e. they can be made non-fluorescent by illuminating them at relatively high intensity of the excitation light. After selectively photobleaching a small volume inside a larger volume, the subsequent recovery of fluorescence in the illuminated region can be followed in time. The extent to which, and speed at which this recovery occurs are measures for the fraction of mobile molecules and the speed at which they move, respectively. Various novel findings in many fields of cell biology that came from the application of FRAP have been elaborately discussed in a large number of reviews (e.g. [1–7]). In this review we will focus on the utilisation of FRAP to study the mobility and behaviour of proteins in the inside of the cell nucleus. Specifically, we will focus on proteins and protein complexes that can interact with DNA, such as transcription and DNA repair factors. Live cell investigation of these DNA-transacting proteins provides an ideal model system to demonstrate the basic principles of FRAP, since they basically display two types of behaviour: they are either mobile, moving through the nucleus or they are (transiently) immobile when they interact with (usually immobile) DNA. As will be shown here, FRAP is a powerful method to study precisely this behaviour. In this chapter, we will first introduce the principles of FRAP in a stepwise manner, from determining overall protein mobility to measuring the immobile fraction. In addition, we will show how FRAP can be applied to determine duration of immobilisation when proteins are transiently immobilised. In each subsection, examples from biological research will be given to show how these qualitative and quantitative FRAP approaches contributed to understanding nuclear protein function. After this we will present a number of variations of FRAP, discuss their interpretation, and summarise some potential pitfalls in application and analysis of FRAP experiments. Throughout the review, examples of typical FRAP curves will be provided that were generated by computer

simulations. We will present a detailed description of the computer models that underlie these simulations.

2
Fluorescence Recovery After Photobleaching (FRAP)

2.1
Fluorescence

In biological research, fluorescent dyes are generally used to label molecules or cellular compartments. The labelled entities can be visualised by fluorescence microscopes, among which in current days the confocal microscope is one of the most frequently used. A large number of fluorescent dyes are available, in many cases as labels of specific biomolecules such as antibodies to detect and localise a specific constituent in a cell. In addition, a considerable number of proteins and other biologically relevant molecules have fluorescent properties themselves. The best known today is without doubt the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria*. The unique property of this protein, and of its many colour variants, is that they require no specialised cellular machinery to reach their final fluorescent form. This allows researchers to express the protein in many different cell types, and to genetically label proteins with important cellular functions. The fact that labelling is encoded by the gene of interest avoids the elaborate purification, chemical labelling and injection into cells. This has lead to an impressive revolution in cell and molecular cell biology and it clearly is this revolution that has boosted the application of a relatively old fluorescence technique: fluorescence recovery after photobleaching [8].

2.2
Mobility

One potential problem in visualisation of fluorescent compounds is that after a certain period of excitation dependent on the intensity of the excitation light they loose their fluorescent capacity, a process known as photobleaching. Although in general photobleaching may hamper or even prevent extended time-lapse monitoring or imaging low dye concentrations, it can also be put to use to study the dynamic behaviour of labelled molecules. The technique to do so is named after the phenomenon that it makes use of: fluorescence recovery after photobleaching. The basic FRAP experiment is straightforward (Figs. 1 and 2): a region of limited dimensions within a larger volume is illuminated shortly at the excitation wavelength of the dye to be bleached, for instance with an intense laser beam. Immediately after illumination molecules in the exposed region will no longer be fluorescent. If the labelled molecules have a fixed position, the region will remain bleached and this situation will remain unchanged (Fig. 1A). However, if the fluorescent molecules are mobile, for instance because

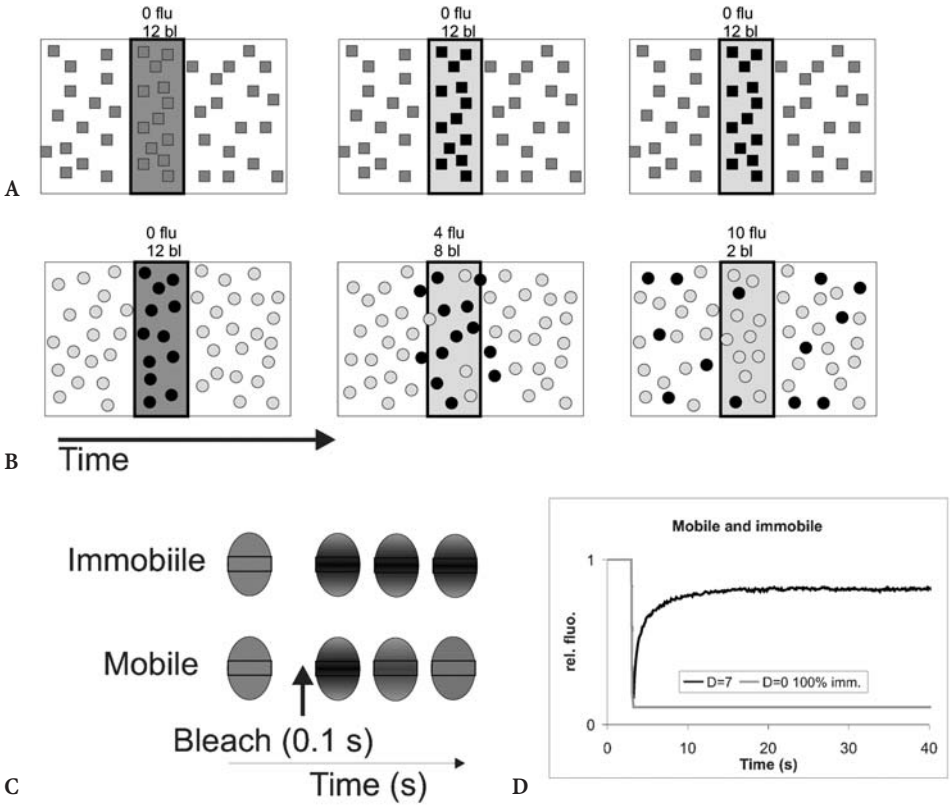


Fig. 1A–D The FRAP principle. When a part of a volume (*shaded rectangle*) containing fluorescent molecules is illuminated at high intensity excitation light, the molecules inside that volume lose their fluorescent property, a process also termed photobleaching: **A** when the molecules are immobile (*grey squares*), the bleached molecules in the strip and fluorescent molecules outside the strip will stay in place, so the bleached area will remain non-fluorescent after photobleaching; **B** when molecules are mobile (*grey spheres*), for instance because they diffuse, bleached molecules in the strip and fluorescent ones outside will redistribute throughout the nucleus after photobleaching, resulting in a gradual recovery of fluorescence after photobleaching. When the bleached area is not very small compared to the total volume (like in many biological cases) the final recovery will not reach the initial value; **C** schematic of strip-FRAP experiments on cell nuclei (*grey ellipses*) in the two cases in A and B. A narrow strip region in the centre of the nucleus is bleached with a short intense bleach pulse and the recovery of fluorescence in the strip is monitored at regular time intervals; **D** graph of the fluorescence recovery process in the two cases depicted in A and B. Relative fluorescence ($F(t)/F(0)$) is plotted against time. The curves are based on computer simulated FRAP experiments (see text). The molecules in case B, where they were all mobile, had a diffusion coefficient of 7 $\mu m^2/s$ in the simulated experiment

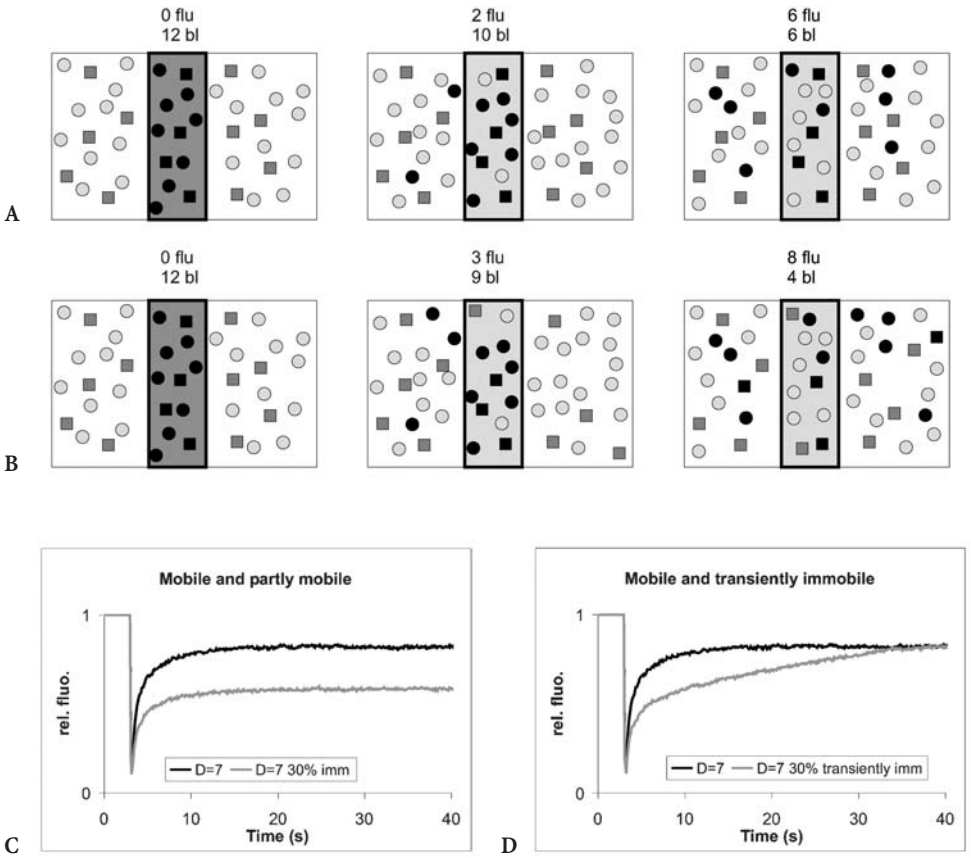


Fig. 2A–D the FRAP principle in two, more complicated cases than the ones in Fig. 1: **A** a situation where only a part of the molecules is immobilised (*grey squares*). In this case only the mobile fraction of molecules (*grey spheres*) causes recovery of fluorescence. Since the immobile do no longer contribute to the signal after photobleaching, fluorescence after complete redistribution will be considerable lower than before bleaching. After correction for the fraction of bleached mobile molecules, the size of the immobile fraction can be calculated from the FRAP curves; **B** a situation where a fraction of the molecules are transiently immobile, i.e. they bind for a while and release. In this case the recovery of fluorescence will initially be due to redistribution of mobile molecules. A secondary recovery of fluorescence will occur due to fluorescent immobile molecules outside the strip becoming mobile and redistribute. The graphs at C and D represent the recovery of fluorescence in the cases in A and B plotted against time; **C** *black curve* represents free molecules in Fig. 1B. The *grey curve* represents the situation in A where a 30% fraction of the molecules is permanently immobilised, resulting incomplete recovery of fluorescence; **D** *black curve* is the same as in C. The *grey curve* now represents a situation where a steady state exists in which always a 30% fraction of the molecules is immobilised, but individual ones are immobile for 45 s, resulting in an initial recovery due to diffusion of free molecules and a secondary (linear) recovery of fluorescence due to release of fluorescent immobile molecules. The time point where the recovery reaches a plateau is a good estimate of the average binding time of individual molecules. The curves are based on computer simulated FRAP experiments (see text)

they diffuse, after the illumination the non-fluorescent and fluorescent molecules inside and outside the area will remix (Fig. 1B), leading to an increase of fluorescence in the bleached region which continues until bleached and fluorescent molecules have completely redistributed over the entire volume. If the bleached area is relatively large compared to the volume to which the investigated molecules are restricted, the final recovery of fluorescence will not be to the pre-bleach level (Fig. 1B). The process can be followed by monitoring fluorescence either in the bleached area or in the total volume in which the investigated molecules reside (Fig. 1C).

Almost all current application of FRAP in the field of nuclear research are performed using modern confocal microscopes. An advantage of the confocal microscope is that any region can be selected for bleaching. Usually a small area is selected somewhere inside the nucleus and the area is scanned prior to bleaching at a low laser intensity to determine a prebleach fluorescence value. After that a short bleach pulse is given at a high laser intensity. Subsequently the recovery of fluorescence inside the area is monitored at low intensity, preferably until complete recovery of the fluorescent signal is reached. Usually the selected area is kept small, but if cells are studied that contain fluorescently tagged molecules at a physiologically relevant concentration, the obtained signal may be quite low. In such a situation it is advisable to select a larger area to increase the signal to noise ratio, for instance a narrow strip spanning the entire nucleus. With the newest versions of the confocal microscope it is possible to scan an area of 20×1 µm (200×10 pixels at optimal resolution) in only 20 ms. After a statistically relevant number of cells has been sampled, the average mobility of the fluorescent molecules can be determined by averaging the normalised fluorescent data of individual cells. There are several ways of normalising FRAP data. The most straight forward way is to express the data relative to the prebleach value: $I_{norm,t}=(I_t-I_{background})/(I_{prebleach}-I_{background})$, where $I_{prebleach}$ is the measurement before the bleach (or the average of a number of recordings before bleaching), and $I_{background}$ is the signal level when no fluorescence is present. A second method of normalisation is used to fit data to analytically derived descriptions of diffusion and is expressed as: $I_{norm,t}=(I_t-I_0)/(I_{final}-I_0)$, where I_{final} is the final value when no more fluorescence recovery occurs. This way of normalising the data yields a curve that starts at zero right after bleaching and reaches one at recovery, allowing to fit the data (e.g. using least squares methods) to any equation that represents the diffusion process. Since many different approaches have yielded equations that were described in abundant literature we will not address the topic in this review but refer to the following citations. Axelrod and coworkers were the first to develop a detailed analytical solution for the basic FRAP principle, which was later refined by Soumpasis [9]. They were able to do so by using a single laser beam to both bleach a small region, i.e. the spot created by a focused laser beam (at high laser intensity), and detect the recovery of fluorescence (at low laser intensity) inside that same region. This approach enhanced the analysis of the data, since the bleached region exactly overlapped the region monitored afterwards. More-

over, the authors provided ways to determine whether observed mobility was due to flow or to diffusion. However, these analytical solutions were limited to membrane studies. Therefore, other mathematical approximations have been used to fit data obtained with confocal microscopy. A one-dimensional model of diffusion into a strip was originally used to study diffusion in nuclear and ER membranes [10] and later extended to investigate nuclear protein diffusion [11]. Various other equations representing 2-D diffusion models can be found in Carrero et al. and the literature cited therein [12]. A very thorough 3-D approach was presented by Blonk et al. [13]. In addition Monte Carlo simulation may be used to generate FRAP curves to fit experimental data [11]. The curves presented in the figures of this review are generated using this approach. This will be discussed in more detail later.

After the advent of GFP technology, the first FRAP experiments in research of protein behaviour in the cell nucleus were applied to determine the mobility of proteins or other biomolecules in the living cell nucleus. It had long been a matter of great concern whether DNA-transacting processes like transcription, replication and repair and also RNA-splicing were governed by highly organised and compartmentalised holo-enzymes, (sometimes termed factories), comprised of many different factors. This hypothesised high degree of organisation and formation of stable protein structures suggested a relatively low mobility of nuclear constituents. The development of GFP tagging made it possible to thoroughly investigate the issue in its most relevant context: the living cell. One of the first investigations in which FRAP provided information on the mobility of nuclear proteins were focused on constituents of the nucleotide excision repair (NER) machinery, a repair system that removes single strand DNA damage caused by ultra violet (UV) light. The authors investigated cell lines that stably expressed functional GFP-tagged NER-factors at physiological levels. The latter was considered important since over-expression of the factors under investigation may result in a large fraction of the molecules being mobile simply because they are redundant. In another bad-case-scenario, over-expression may lead to precipitation of the excess of protein, for instance caused by misfolding due to insufficient availability of chaperones. In this type of situations mobility studies may not be very informative.

By application of the confocal FRAP experiments described above to two repair factors (ERCC1, XPA) and a repair/transcription factor (TFIIH) it was shown that nuclear proteins have a much higher mobility than previously anticipated [11]. In the absence of DNA damage the NER factors ERCC1/XPF and XPA appeared to be freely mobile and diffuse through the nucleus as separate entities. After this discovery, the surprisingly high mobility of DNA repair factors was also found for other nuclear proteins using a similar FRAP approach. In a study of three proteins involved in different nuclear processes, Misteli and coworkers showed a surprisingly high mobility of HMG-17, SF2/ASF and fibrillarin, which are involved in transcription, pre-mRNA splicing and rRNA processing respectively [14]. Others also found, using FRAP that molecules can be highly mobile in the nucleus. Various studies revealed that dextrans up to a

molecular weight of 500 kD and also fluorescently labelled oligonucleotides diffuse freely through the nucleus [15–17] and have access to most regions of the nucleus [18].

Complementary to the investigation of DNA metabolising factors, the mobility of their template, chromatin, was also studied using FRAP. In a non-confocal microscope set-up, using an aperture diaphragm to minimise out-of-focus-light detection, FRAP was applied to cells of which the DNA was stained with an ethidium bromide derivate [19]. After bleaching small spots in the stained chromatin the spots remained non-fluorescent for periods of more than one hour, indicating that interphase chromatin is largely immobile, at least at the spatio-temporal resolution of the described experiment.

These examples show that already the very straightforward, largely qualitative application of FRAP to determine whether biomolecules are mobile or not has provided important conceptual implications for the way nuclear processes are organised in the living cell.

2.3 Immobilisation

As already pointed out above, FRAP is not only able to determine whether molecules are mobile or not, but can also be applied to quantitatively determine a second mobility parameter, i.e. immobile fraction. If only a fraction of the investigated fluorescently tagged molecules is immobile, those outside the bleached area will never enter it, similar to the situation in Fig. 1A, whereas those inside the area will never leave (Fig. 2A). Immobile molecules that were inside the area contributed to the fluorescent signal before bleaching, but will no longer do so after bleaching. Therefore, recovery of fluorescence in the bleached area will reach a lower final level compared to when no immobile fraction would be present (Fig. 2C). To estimate the immobile fraction the data is usually normalised in a different way than described above: $I_{\mathrm{norm,t}}=(I_{\mathrm{t}}-I_0)/(I_{\mathrm{prebleach}}-I_0)$, where I_0 is the intensity immediately after bleaching. Using this way of normalising the data yields a curve of which the prebleach value is one and the fluorescence level immediately after bleaching is zero. Provided that the fraction of mobile bleached molecules is negligible compared to the total amount of fluorescence molecules, i.e. if the FRAP curve would return to prebleach levels when no immobile fraction was present, the immobile fraction can be estimated as $1-\mathrm{I}_{\mathrm{norm,final}}$ (Fig. 2C). However, in most biological applications the volume in which the fluorescent molecules are present is very small, for instance in the cell nucleus. As stated above, a substantial portion of the fluorescent molecules may then be bleached, giving rise to an incomplete recovery for which the signal should be corrected if one wants to determine immobile fraction. There are several ways to deal with the problem. First, and probably best is to study the same proteins in a situation where they are freely mobile, in case of gene transcription or DNA repair factors, when they are inactive. For DNA repair this is not a great problem, since one can study cells without DNA damage. However,

in the case of essential transcription factors, it is more problematic to do so. In that case the signal can be divided by the total nuclear fluorescence outside the bleached region. However, this method may have a potential drawback when cells are studied under condition where photobleaching due to monitoring occurs, a problem that will be discussed in detail later. An alternative way to estimate the fraction of fluorescent mobile molecules removed by the bleach could be obtained from experiments on cells expressing free GFP. In fact, it would be of great value if such control experiments were always presented allowing a better comparison of data obtained in different labs using different microscopes and FRAP settings.

Although the observation alone that proteins are highly mobile in densely compacted compartments such as the cell nucleus has attributed to understanding of how biomolecules function, it is the fact that FRAP can also be used to determine the percentage of immobilised molecules that makes it a strong tool for investigating protein behaviour in living cells. This becomes strikingly apparent in the case of the investigation of nuclear proteins that have a function in regulating DNA metabolism (transcription, replication and repair). It is obvious that in order to perform their function, these factors have to bind to DNA, either directly or indirectly through other chromatin constituents. Since DNA is essentially immobile, at least during a typical FRAP experiment, the activity of DNA-transacting enzymes and supportive factors leads to their immobilisation. The first reports on the application of FRAP to study DNA binding in vivo were from the investigation of repair of UV-lesions of DNA. By application of a FRAP procedure (see also below) specifically developed to study immobilisation in small compartments (like the nucleus) it was shown that NER factors ERCC1/XPF, XPA, and the transcription/repair factor TFIIH bind to DNA damage (and are thus immobilised) in UV dose dependent amounts [11, 20, 21], reaching a plateau at an immobile fraction of approximately 30–40%, at a UV-dose of 8–16 J/m^2 (UV-C irradiation was used). In addition it was shown that the apparent diffusion coefficient of the freely mobile fraction was not affected by the induction of DNA damage. Note that this can be judged already by visual inspection of the FRAP curves from simple strip-FRAP experiments (Fig. 2C): the (incomplete) recovery of fluorescence in the situation where there is an immobile fraction (grey line) reaches a plateau at the same time after photobleaching as in the case where there is no immobile fraction, the only difference being that the final level is lower.

2.4 Transient Immobilisation

In the FRAP experiments described above to determine immobile fraction of repair proteins, the immobile fraction is constant at the time scale of the FRAP experiment. However, it is also possible that individual molecules immobilise for a shorter period, leading to a turn-over of immobile and mobile molecules during the FRAP experiment. If the system studied is in steady state, i.e. if an

equal amount of molecules bind and release per unit time, this scenario will result in a secondary recovery of fluorescence. FRAP analysis can be applied to determine the third 'mobility parameter': average duration of transient immobilisation of individual molecules. In Fig. 2B such a situation is depicted. It can be seen that the release of molecules that were immobile at the time of bleaching will give rise to a secondary recovery of fluorescence in the bleached area (Fig. 2D). Such a scenario may well be expected in the case of transcription, replication or DNA repair since binding to DNA by these factors in many cases may be limited to the time needed for their action. For instance a DNA repair factor like the endonuclease heterodimer ERCC1/XPF can be expected to bind to repair complexes at the DNA for at least the time required to cut the 5′ end of the damaged DNA strand. After the DNA lesion has been repaired successfully it is well possible that the repair complex releases, since the DNA is now repaired. Indeed, it was shown that individual NER repair factors are immobilised to damage for approximately three to five minutes [11, 20, 21]. In addition, it was recently reported for the androgen receptor (AR) that the wild type AR shows a reduced mobility compared to a non-DNA-binding mutant, and that this drop is caused by the transient binding of ARs to DNA in the order of 45–90 s [22]. For this, the authors used two complementary FRAP assays, a strip-FRAP and a combined FLIP and FRAP assay, which are discussed in detail below; see Fig. 4.

Although it is possible to determine the average duration of immobilisation in cases where the immobilisation is relatively long compared to the time the mobile fraction redistributes completely within the nuclear volume, often transient immobilisation will be too short to distinguish from (slow) diffusion (see also below). Nevertheless, in many cases conclusions can be drawn from the qualitative observation that a reduced mobility compared to a situation where the protein is inactive is probably due to interaction with immobile elements in the nucleus. For estrogen receptors (ER), it was reported that binding to ligand slowed down the previously freely mobile ER. In addition, the authors report a distinct slow down induced by antagonists as well as by proteasome inhibitors, which they contribute to binding to the (operationally defined) nuclear matrix [23]. Moreover, their data suggest that in the presence of ligand, the tagged cofactor SRC-1 shows dynamics similar to the ER, suggesting interaction of these molecules prior to DNA binding. Similarly, the RNA splicing factor ASF was reported to move considerably slower through the nucleus than expected for free diffusion. The authors argue that frequent but transient interactions with relatively immobile nuclear binding sites causes the observed slow nucleoplasmic mobility [24].

2.5 Residence Time

The previous sections discussed the application of FRAP to determine the overall nucleoplasmic mobility of nuclear proteins. FRAP has another, more read-

ily appreciated application in the investigation of residence time of nuclear factors that accumulate in small nuclear substructures such as nucleoli, nuclear bodies, telomeres or DNA double strand break related foci. To determine residence time of proteins at foci, it is possible to conduct two complementary photobleaching assays, FRAP and FLIP (fluorescence loss in photobleaching). Again, the FRAP experiment is straight forward: the accumulation is bleached and fluorescence recovery monitored at regular intervals after bleaching. The time it takes to fully recover is a measure for the residence time of proteins in the accumulation. In the complementary FLIP experiment, a region at a distance from the accumulation is bleached and the loss of fluorescence in the distant accumulation is monitored (see below).

Similar to the results of measurement of overall nuclear mobility, FRAP and FLIP revealed in many cases a surprisingly high exchange of many factors with nuclear bodies. One of the most elaborate studies was conducted by Misteli and coworkers who combined what was termed kinetic modelling with FRAP of factors that regulate transcription of rRNA genes in the nucleolus, providing a kinetic framework for RNA polymerase I transcription [25]. The authors show a high turnover of transcription initiation factors at rRNA genes. In addition, it was shown by others that also transcription factor TFIIH is associated with nucleoli in a highly dynamic fashion [20]. Similarly, nuclear bodies such as Cajal bodies, and also DNA double strand break foci and telomeres were investigated by FRAP and FLIP. In these studies the authors show that different factors with different functions associate to foci in a differential manner [26–30]. Replication factors also interact with replication foci in a differential way. Cardoso and coworkers have reported that PCNA, but not RPA34, is stably associated with replication foci [31].

Apart from studying naturally occurring nuclear substructures, researchers have also used various methods to introduce artificial accumulations into cell nuclei. One of the first reports comes from steroid receptor research. Hager and co-workers investigated the activity of GFP-tagged glucocorticoid receptors in cells containing long tandem arrays of the MMTV promoter. It was shown that liganded GR-GFP accumulated to these arrays in such amounts that they could be readily visualised by fluorescence microscopy as bright fluorescent spots [32]. In addition, FRAP and FLIP measurements suggested a transient interactions between MMTV-promoters and GRs in the order of seconds to a minute, suggesting a 'hit-and-run' mechanism for GR transcription initiation. In a similar way, nuclear factors involved in various types of DNA repair were investigated, by inducing DNA damage in a subregion of the nucleus [20, 33]. Interestingly, these studies show that DNA repair factors of UV-damage tend to bind considerable longer to damage than transcription factors like steroid receptors and general transcription factors [11, 20, 34]. An obvious exception to this is RNA polymerase II which was reported from FRAP studies to be immobilised for more than 10 min [35].

3
FRAP Variants

3.1
Applications

In this section we will discuss various FRAP variants that have been used to study the behaviour of nuclear proteins. The development of confocal microscopy in the 1980s made it possible to extend the straightforward methods provided by early researchers. The scanning capacity of this type of microscopes made it possible to bleach specific areas and measure the redistribution of fluorescence in other regions, or even scan the entire cell. This enabled various more dedicated approaches that will be discussed below: FRAP-FIM (FRAP for immobilisation measurement) (Fig. 3A) [11], combined strip-FRAP (Fig. 3B) and FLIP-FRAP (Fig. 3C,E) [22], FLIP (Fig. 3D) [28–30, 32–34] and iFRAP (Fig. 3F) [25].

A FRAP variation termed FRAP-FIM was developed to study the immobilisation of nuclear repair factors. The method differs from the 'conventional' FRAP applications, in that the spot is illuminated at a relatively low intensity for a relatively long time [11], typically 4–12 s. If all molecules are immobile, only those in the beam will get bleached, resulting in a dark hole in the nucleus when a new image is taken at the confocal plane, even if this image is taken after a long period of time. However, if all molecules are mobile, a significant part (depending on diffusion rate) will pass through the beam in this period, get bleached and diffuse through the nucleus. If a period of redistribution of the bleached and unbleached mobile molecules is allowed after the bleaching, this will result in a homogeneously distributed decrease of fluorescence in the entire nucleus. In an intermediate situation, when a part of the molecules is mobile, the resulting image will be an intermediate between the two extreme situations. To obtain images of these extremes, cells expressing GFP can be used that are either paraformaldehyde-fixed or alive, representing the 100% immobile and 100% mobile scenarios respectively. Next, the fluorescence intensity ratio in images taken after and before bleaching are calculated as a function of the distance to the bleach spot. To do so the image is divided into concentric rings around the bleach spot. Finally, the weighted sum of the fluorescence ratio profiles (FRPs) of the extremes, $a(\mathrm{FRP}_{\mathrm{immobile}})+b(\mathrm{FRP}_{\mathrm{mobile}})$, is fitted to the curves obtained in cells under investigation and the weight factors a and b represent the immobile and mobile fraction respectively.

Of special interest is the combination of the strip-FRAP (Figs. 3B and 4A) and FLIP-FRAP method (Figs. 3C and 4B) that can be used to study transient immobilisation [22]. In the combined FLIP-FRAP experiment a region distant from the bleach region is monitored simultaneously with the FRAP region. In such an experiment (Figs. 3C and 4B), a strip at one pole of the nucleus is bleached for a relatively long period (typically 4–8 s) at a moderate excitation intensity. Subsequently the fluorescence is monitored in that region (FRAP), but

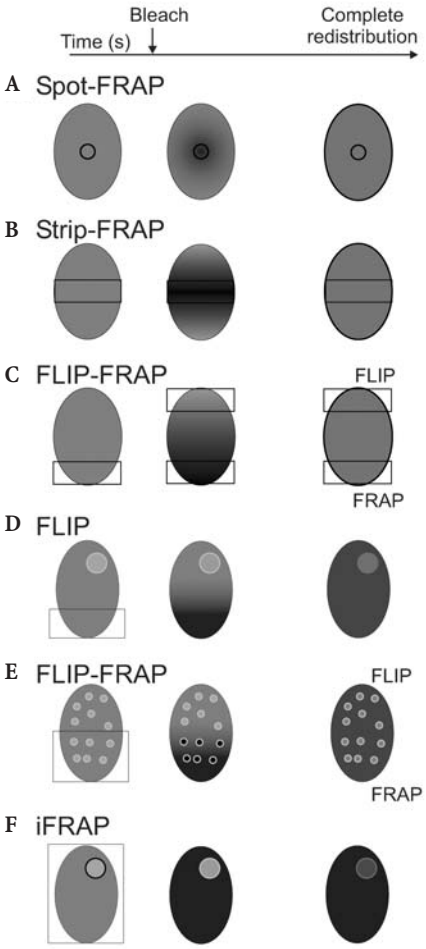
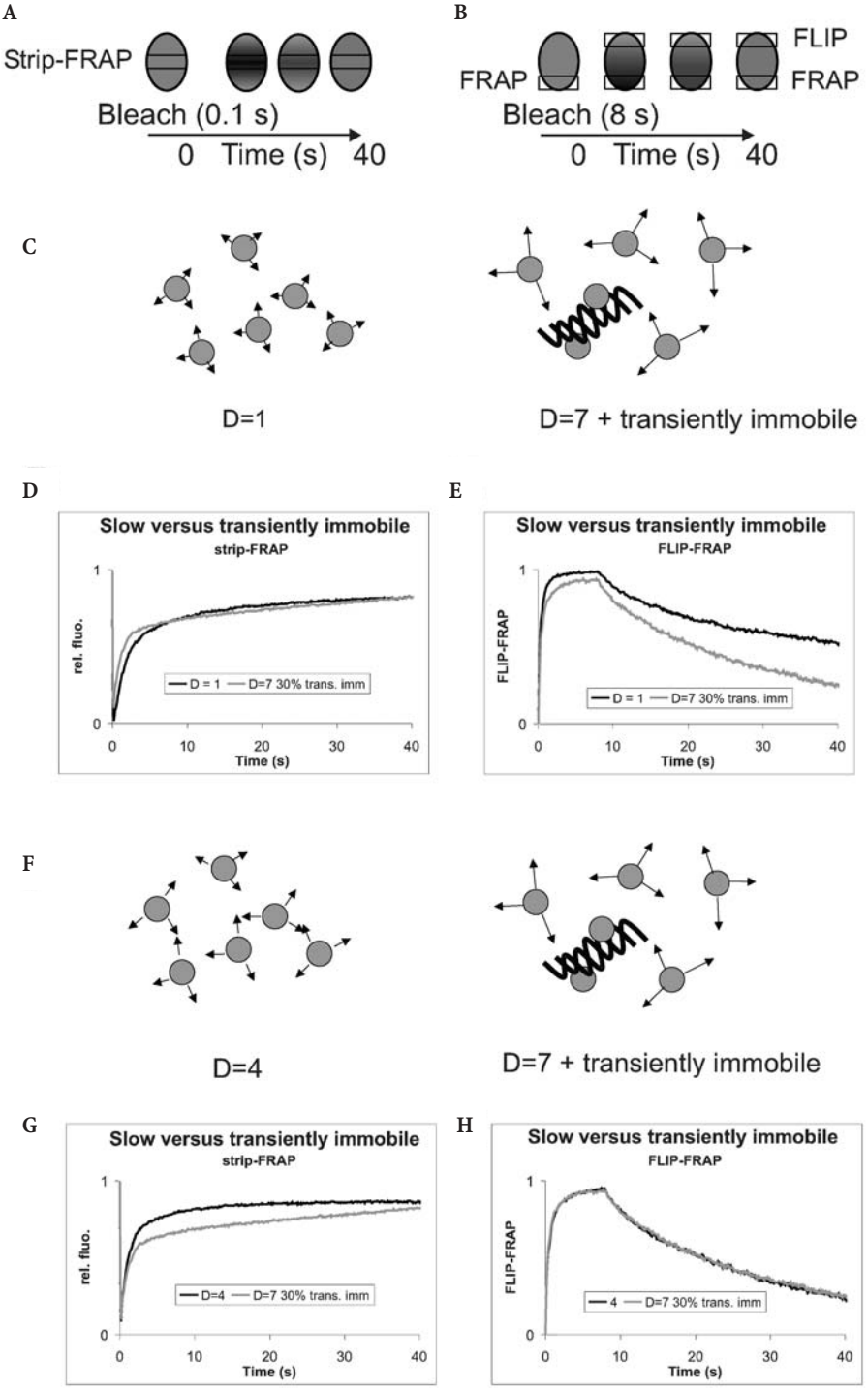


Fig. 3A–F Examples of different FRAP applications that have been developed to study the dynamics of nuclear proteins. The curves are based on computer simulated FRAP experiments (see text): **A** spot bleaching – the most straightforward FRAP experiment based on bleaching only a diffraction limited spot. A specific application of the method was developed to study the immobilisation of nuclear repair factors (see text and [9]); **B** strip bleaching – when signals are relatively low, e.g. due to low expression of the tagged protein, a larger area, for instance a strip spanning the entire nucleus can be bleached, and the fluorescence monitored at regular intervals; **C** combined FLIP and FRAP – the FLIP-FRAP method differs from the strip-FRAP in that two areas are monitored after bleaching. D–F represent methods to determine the residence times of proteins associated with nuclear accumulation such as nucleoli, nuclear bodies, telomeres locally damaged nuclei or DNA double-strand break associated foci; **D** fluorescence loss in photo bleaching (FLIP). An area at a distance of nuclear accumulation (*bright circle*) is bleached, either repetitively, or with one prolonged bleach pulse at low intensity. The velocity at which fluorescence is lost in the distant accumulation is a measure for the residence time at the accumulation, the point in time where a new steady state is reached being a good measure for residence time; **E** combined FLIP and FRAP. One half of the nucleus is bleached, including foci (*bright circles*) in that half; **F** iFRAP (inverse FRAP). iFRAP was developed to estimate the rate of dissociation of molecules from the nucleolus (see text)


A
Strip-FRAP
Bleach (0.1 s)
0 Time (s) 40
B
FRAP
FLIP
FRAP
Bleach (8 s)
0 Time (s) 40
C
D=1
D=7 + transiently immobile
D
Slow versus transiently immobile
strip-FRAP
rel. fluo.
D = 1 D=7 30% trans. imm
Time (s)
E
Slow versus transiently immobile
FLIP-FRAP
FLIP-FRAP
D = 1 D=7 30% trans. imm
Time (s)
F
D=4
D=7 + transiently immobile
G
Slow versus transiently immobile
strip-FRAP
rel. fluo.
D=4 D=7 30% trans. imm
Time (s)
H
Slow versus transiently immobile
FLIP-FRAP
FLIP-FRAP
4 D=7 30% trans. imm
Time (s)

also in the area at the other side of the nucleus (FLIP). The power of combining two different FRAP protocols (strip-FRAP and FLIP-FRAP) is exemplified in Fig. 4. In a scenario where molecules transiently bind to immobile elements in the nucleus (Fig. 4C, left panel), it will often, if not always, be possible to define another scenario where the entire population of molecules is mobile (but moves slower) (Fig. 4C, right panel) which results in similar FRAP curves (Fig. 4D). For instance, using simple strip-FRAP methods in a situation where molecules diffuse at a D of 7 $\mu m^2/s$ of which 30% is immobile, individual molecules being on average immobilised for 45 s, it will be difficult to distinguish this from a situation where the entire population diffuses at a D close to 1 $\mu m^2/s$ (Fig. 4D). However, a way to cope with this problem is to perform a complementary FLIP-FRAP experiment. The two situations described above yield similar curves in the strip-FRAP experiments but are clearly separated in the accompanying FLIP-FRAP experiment (Fig. 4E). From this, one may erroneously conclude that the FLIP-FRAP experiment is apparently more suited to identify the presence of secondary fluorescence recovery. However, scenarios also exist that give similar curves in the FLIP-FRAP experiment, but not in the strip-FRAP (Fig. 4F): if D=4 $\mu m^2/s$ the FLIP-FRAP curve are similar to the faster but transiently immobile molecules whereas the strip-FRAP curves are now different (Fig. 4G,H). In addition to the clear advantage of using different FRAP approaches, it also allows one to apply global analysis of the data, a statistical method to enhance measurement accuracy often used in fluorescence methodology.

When binding times are very short, typically less than approximately 5 s, the combined strip-FRAP/FLIP-FRAP method, or any other FRAP-method is not capable to distinguish this scenario from a situation where all molecules are freely mobile (but moving slower). However, several approaches to cope with this problem have been developed. In the investigation of nuclear factors, it can

Fig. 4A–H Comparison of two different variants of FRAP: strip-FRAP and FLIP-FRAP (see also Fig. 3B, C). The combined use of the two protocols may allow to discriminate between transient binding and slow diffusion. The curves are based on computer simulated FRAP experiments (see the text): **A**, **B** schematic drawings of the strip-FRAP (see also Fig. 1C) and FLIP-FRAP methods. The FLIP-FRAP method differs from the strip-FRAP in that two areas are monitored after bleaching. Briefly, a strip at one pole of the nucleus is bleached for a relatively long period at a moderate excitation intensity. Subsequently the fluorescence is monitored in that region (FRAP), but also in the area at the other side of the nucleus (FLIP). Subsequently the difference between the two (normalised) fluorescence levels is plotted against time; **C** schematic drawing of two scenarios where molecules are either free, but relatively slow (D=4 $\mu m^2/s$, top panel), or relatively fast (D=7 $\mu m^2/s$), but transiently immobilised such that 30% is immobile in steady state and individual molecules are immobilised for 45 s (*bottom panel*); **D**, **E** strip-FRAP and FLIP-FRAP curves of the scenarios depicted in C. In this case strip-FRAP can discriminate between the two cases, whereas the FLIP-FRAP curves are nearly identical; **F** schematic drawing of a situation where freely mobile molecules are slower (D=1 $\mu m^2/s$, *top panel*) than in C; **G, H** strip-FRAP curves are identical whereas the FLIP-FRAP method can now discriminate between the two scenarios

be expected that if immobilisation is observed, this is due to binding to DNA or other immobile structures in the nucleus. If the DNA binding domain of the molecule is known, and mutations are known that specifically disable the DNA-binding capacity, one may compare the mobility of mutants that lack the ability to bind to DNA with wild type mobility [22, 36]. If functional mutants are not available, it is possible to inhibit the process in which the protein under surveillance is involved. For instance the mobility of general transcription factors can be compared in untreated and transcription-inhibited cell exposed to agents like DRB or actinomycin D [20, 35]. Finally, a powerful set of experiments may be to study protein mobility at different temperatures [14, 17, 20]. Since diffusion changes linearly with temperature a limited effect is expected when temperature is dropped from 310 K to 300 K (assuming that the effect on viscosity on the cellular interior is also limited, since diffusion is also linearly dependent on the reciprocal of viscosity). However, if the molecules are transiently immobilised in a temperature dependent fashion, e.g. in an enzymatic reactions, their binding time is expected to increase considerably resulting in a notable drop in mobility.

In addition to the above described method to determine overall nucleoplasmic mobility, methods have also been developed to determine the residence times of proteins associated with nuclear accumulations such as nucleoli, nuclear bodies, telomeres, locally damaged nuclei or DNA double-strand break associated foci. These accumulation are frequently studied using the most popular variant of the basic FRAP experiment discussed already above: FLIP (Fig. 3D). The FLIP approach to study accumulations or small compartments, is to bleach an area at a distance of a nuclear accumulation, either repetitively, or with one prolonged bleach pulse at low intensity. If molecules in the accumulation are permanently bound, fluorescence will not go down in the accumulation, since bleached molecules do not exchange with the unbleached in the accumulation. In contrast, fluorescence in the accumulation is expected to decrease if molecules in the accumulation have limited residence times. The velocity at which fluorescence is lost is a measure for the turn-over at the accumulation, the point in time where a new steady state is reached being a good measure for the average residence time in the accumulation. A variant of FLIP for measuring residence times in multiple foci was introduced in the investigation of foci associated with DNA double strand break [26] and of telomeric proteins [27] (Fig. 3E). Briefly, one half of the nucleus is bleached, including the foci present in that half. Subsequently the entire image is monitored at regular time intervals. Image analysis is then used to determine the changes in fluorescence level in each of the bleached and unbleached foci. Another sophisticated method, termed iFRAP (inverse FRAP) was introduced by Misteli and coworkers (Fig. 3F) [25]. It was developed to estimate the rate at which molecules associated with the nucleolus exchange with the surrounding nucleoplasm. The method optimally makes use of the advanced possibilities provided by the operating software of modern confocal microscopes to indicate bleach regions with complicated shape. In iFRAP the entire nuclear volume is bleached with

exception of the investigated accumulation (nucleolus). Immediately after bleaching, the loss of fluorescence in the accumulation fully represents the releasing molecules, whereas in normal FLIP the loss is the result of unbalance of dissociation and association of fluorescent molecules. Therefore determination of the rate of exchange does not require further complicated analytical methods.

4 Pitfalls

In the interpretation and analysis of FRAP experiments there are some unexpected pitfalls that we will discuss below. A first potential difficulty may arise when fluorophore levels are low requiring monitoring at relatively high excitation intensity. Correction for the resulting monitor bleaching may lead to erroneous results when an immobile fraction is present. The second potential pitfall concerns the influence of exchange of bleached and fluorescent molecules with compartments distant from the bleach region. This potential problem specifically may lead to erroneous interpretation of FRAP results when the protein under surveillance accumulate in nuclear subcompartments like the nucleolus, in replication or other foci, or in locally applied DNA damage. Third, the phenomenon of reversible bleaching or blinking may affect FRAP analysis. In the next three paragraphs we will discuss these items.

When the concentration of the tagged protein is low, which is frequently observed when factors are stably expressed at physiological levels, for instance in transfected cells or knock-in mice, fluorescence can only be detected using a laser intensity at which a certain degree of acquisition bleaching will occur. Specifically when the investigator is interested to quantify an observed or expected secondary recovery, this may cause a substantial problem: the secondary recovery may be cancelled out or even exceeded by monitor bleaching during monitoring fluorescence after bleaching (Fig. 5A,B). The obvious method to deal with this problem is to monitor a region similar to the one subjected to FRAP, without applying the bleach pulse, derive the bleach rate and use that to subsequently correct the FRAP curves. Although this works in the case of a freely mobile population of molecules, the correction method is not applicable when a (transiently) immobile fraction is present (Fig. 5C). This is due to the fact that in the control experiment performed to derive bleach rate, the immobile fraction will contribute more to the loss of fluorescence than the mobile one since immobile molecules are constantly in the monitored area, whereas mobile bleached are replaced by mobile fluorescent. This immobile fraction, however, does not contribute to monitor bleaching observed after photobleaching, since it is then largely bleached by the bleach pulse. The result is an overcompensation of the FRAP curve (Fig. 5C). A frequently used alternative correction for acquisition bleaching is to monitor the entire nucleus and normalise the fluorescence intensity in the bleached area to the entire fluorescence level in the nucleus (minus the bleached region). However, when a confocal

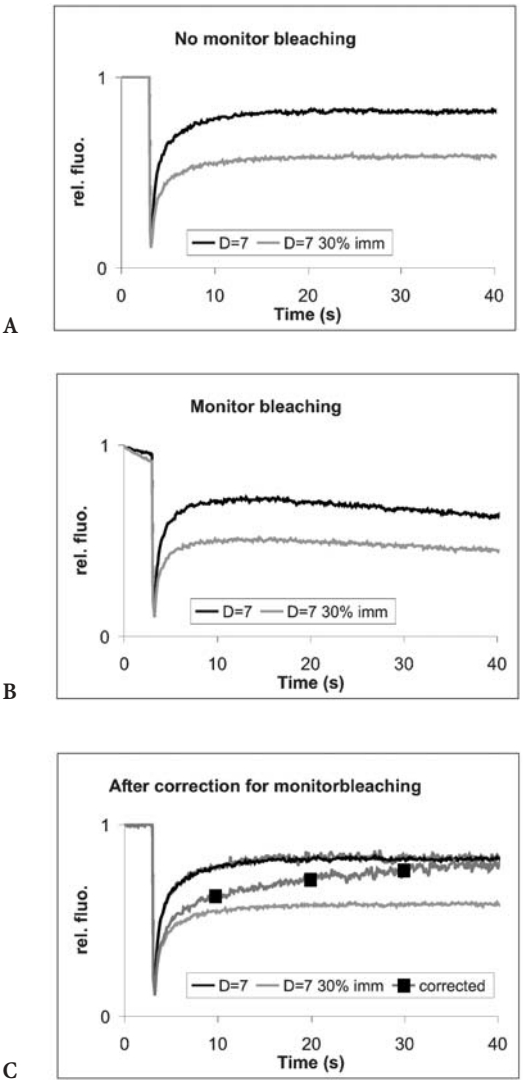


Fig. 5A–C The effect of correction for monitor bleaching in FRAP experiments. The curves are based on computer simulated FRAP experiments (see text): **A** strip-FRAP curves of freely mobile (D=7 $\mu m^2/s$) molecules and the same of which 30% is permanently immobilised; **B** strip-FRAP of the same situations as in A with the difference that now a considerable monitor bleaching is simulated. This may in practice often occur when molecules are investigated at low concentration; **C** to correct for monitor bleaching the experiments were repeated without the bleach pulse. The corrected curve fits well in a situation where the molecules are all freely mobile. However, when an immobile fraction is present, the correction overcompensates the monitor bleaching that actually occurred. This is due to the fact that during the control experiment the immobile fraction bleaches with different kinetics than the free fraction. In the FRAP experiment, after the bleach pulse the immobile fraction is bleached and the signal will have different bleaching characteristics

microscope is used, this method is also hampered by the same phenomenon as in the above example. If an immobile fraction is present, it will contribute in a different way to acquisition bleaching (outside the bleach area) than the mobile fraction which is constantly diffusing to and from the confocal plane. In cases where considerable acquisition bleaching occurs this way of normalising the data will lead to considerable underestimation of the immobile fraction. The longer the experiment is continued, the smaller the fraction will appear, and eventually, he immobile fraction will be fully obscured, when it is completely bleached by the monitoring beam.

A second problem may occur when proteins under surveillance accumulate in one or multiple foci (see above) or when they exchange with neighbouring compartments (shuttling between nucleus and cytoplasm. In many cases, residence times can very well be determined by FLIP, FRAP or iFRAP (Fig. 3). However, a potential problem may arise if the overall nucleoplasmic mobility of the protein is to be determined. Even if the area that is bleached contains no foci, nucleoli, or local damage, a secondary recovery may be observed due to the exchange of bleached and fluorescent molecules with the distant accumulation(s). In the case the focal residence time is in the range of the redistribution time of the protein, this secondary recovery will partly overlap the initial recovery, leading to an apparent slow down of the proteins mobility. A similar effect will take place if the protein shuttles between nucleus and cytoplasm.

A third potential threat to proper FRAP analysis is the 'blinking' behaviour of GFP. It has been shown that many fluorescent proteins rapidly switch between a dark non-fluorescent state and a fluorescent state. The period GFPs are in the dark state (off-time) is not dependent on laser whereas the on-times are [37, 38]. Since the bleach pulse is at a much higher intensity as the monitoring after bleaching, it can be expected that a part of the recovery of fluorescence is due to decrease of the pool of molecules in the off-state. This is not a great problem if all conditions are kept constant, since it will only introduce a constant bias in the results However, comparison between different variants of GFP that have different blinking properties is hampered.

5
Monte Carlo Simulation of FRAP Experiments

From the above it may have become clear that the application of FRAP to constituents of the cell nucleus has dramatically changed the view of the way nuclear processes are organised. This is in spite of the fact that much of the early work, with some exceptions, has been based on semi-quantitative and qualitative interpretation of the FRAP data. It is expected that to push further the limits of FRAP application to processes in the cell nucleus, more sophisticated quantification will be required for future work.

Most of the analytical techniques that have been developed for FRAP quantification are based on mathematical analytical models of diffusion of the

labelled molecules, requiring the solution of known diffusion equations of transport. In these types of analyses, photobleaching is considered as an irreversible process and often simplified into an initial condition for the concentration of chromophores, since the duration of the applied bleach pulse is supposed to be very small compared to the characteristic diffusion time of the studied molecules. Further simplifications are applied to these models by using a reduced set of spatial dimensions or neglecting boundary effects, or by replacing the full point spread function of the focused laser beam with some geometrical approximation. Clearly, these methods capture the essence of FRAP experiments [8, 9, 12, 13], a very thorough 3-D approach being the one developed by Blonk and coworkers [13]. However, when considering experiments in cellular compartments such as the nucleus, some of the approximations may sometimes be oversimplifications, as the system is three-dimensional, finite and bounded, and molecules may be immobile for longer or shorter periods of time. Therefore, it may also be useful to apply another approach to the analysis of FRAP data, i.e. computer modelling of the FRAP procedures and the behaviour of molecules inside small volumes [39]. We have applied in most of our studies a straightforward Monte Carlo method [11, 20, 22], and the examples in this review are also based on this approach. Simulations were performed using experimentally obtained parameters, describing the properties of the microscope lens (beam shape and 3-D intensity distribution, during monitoring and during bleach pulse), the fluorescent label (quantum yield, susceptibility to bleaching, 'blinking' behaviour) and properties of the cell nucleus (size and shape). All these parameters were kept constant in the simulations. Simulations were then performed varying the three protein mobility parameters described above, diffusion coefficient, immobile fraction and duration of binding of individual molecules. Diffusion was simulated by randomly picking a direction to step into with a step size derived from the equation: D=stepsize^2/(6·cycletime), where cycletime in our simulations ranges from 20 to 100 ms. An immobile fraction was established by giving molecules a chance to bind at each cycle, derived from the equation: $P_{\text{bind,1 cycle}}$=(immobile fraction)·(mobile fraction)$^{-1}$·cycletime^{-1}. This software was used to generate the curves presented in this chapter.

6 Discussion

FRAP is a powerful tool for investigating protein behaviour in nuclei of living cells. FRAP experiments are most informative when an inducible system can be used, allowing to compare protein behaviour in their active and inactive state. Alternatively, comparison of the mobility of functional mutants can be compared with wild type mobility. The possibility to estimate immobile fraction and, moreover, the average duration of immobilisation of individual proteins is very useful in studying the dynamics of nuclear proteins, and specifically those that are involved in DNA transacting processes, transcription, repair

and replication. In addition, accumulation in subnuclear regions allows direct visualisation of the exchange (or the absence of it) of proteins with the subnuclear structure. Such accumulations may be either naturally occurring foci or artificial accumulations on for instance stably integrated large promotergene arrays [32] or locally inflicted DNA damage [20, 33]. FRAP probably is less effective when applied to proteins of which the function is not known.

Due to the power of FRAP in the study of nuclear processes, an impressive amount of data has accumulated indicating that many nuclear factors are highly mobile. However, assuming free Brownian diffusion may very likely be an oversimplification of the actual situation. Since many nuclear factors have DNA binding properties and since the nucleus is packed with proteins and DNA it may be expected that the overall mobility of such factors reflects not only (relatively) free movement but also very transient interactions with chromatin or other nuclear structures. In addition, diffusion of factors that do not show interaction may very well be anomalous rather than completely free, due to the crowded nature of the nucleus. The presence of relatively dense chromatin territories may cause diffusing proteins to temporarily get trapped inside such areas giving rise to anomalities in their mobility. However, in many cases FRAP data fit quite well to diffusion models [22] suggesting that at the spatial and temporal resolution of confocal FRAP experiments the effects of anomalous or restricted diffusion may be too weak to be detected by FRAP. Determining the effective diffusion coefficient at which a free fraction moves may then be of less functional relevance than determining their immobilisation characteristics. At least in the case of DNA-metabolising proteins, these immobilisation parameters are likely to represent the timing of the activity of the proteins when they are associated with immobile DNA-protein complexes. Nevertheless, effective diffusion coefficients may, in well controlled experiments, provide an indication of the molecular size of complexes in which a tagged factor resides, although future work is required to provide decisive evidence.

Concluding, FRAP has clearly shown its applicability in nuclear research and is expected to contribute largely to further unravelling and quantifying nuclear processes such as transcription, replication, RNA splicing and DNA repair. In addition, the combined application of FRAP with other quantitative techniques, like FRET (fluorescence resonance energy transfer) and FCS (fluorescence correlation spectroscopy) will be instrumental for future research of the functional organisation of the cell nucleus.

References

1. Janicki SM, Spector DL (2003) Nuclear choreography: interpretations from living cells. Curr Opin Cell Biol 15:149–157
2. Belmont A (2003) Dynamics of chromatin, proteins, and bodies within the cell nucleus. Curr Opin Cell Biol 15:304–310
3. Chubb JR, Bickmore WA (2003) Considering nuclear compartmentalization in the light of nuclear dynamics. Cell 112:403–406

4. Lippincott-Schwartz J, Altan-Bonnet N, Patterson GH (2003) Photobleaching and photoactivation: following protein dynamics in living cells. Nat Cell Biol S7–14
5. Misteli T (2001) Nuclear structure – protein dynamics: implications for nuclear architecture and gene expression. Science 291:843–847
6. Houtsmuller AB, Vermeulen W (2001) Macromolecular dynamics in living cell nuclei revealed by fluorescence redistribution after photobleaching. Histochem Cell Biol 115:13–21
7. Vermeulen W, Houtsmuller AB (2002) The transcription cycle in vivo: a blind watchmaker at work. Mol Cell 10:1264–1266
8. Axelrod D, Koppel DE, Schlessinger J, Elson E, Webb WW (1976) Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. Biophys J 16:1055
9. Soumpasis DM (1983) Theoretical analysis of fluorescence photobleaching recovery experiments. Biophys J 41:95–97
10. Ellenberg J, Siggia ED, Moreira JE, Smith CL, Presley JF, Worman HJ, Lippincott-Schwartz J (1997) Nuclear membrane dynamics and reassembly in living cells: targeting of an inner nuclear membrane protein in interphase and mitosis. J Cell Biol 138:1193–1206
11. Houtsmuller AB, Rademakers S, Nigg AL, Hoogstraten D, Hoeijmakers JH, Vermeulen W (1999) Action of DNA repair endonuclease ERCC1/XPF in living cells. Science 284:958–961
12. Carrero G, McDonald D, Crawford E, de Vries G, Hendzel MJ (2003) Using FRAP and mathematical modeling to determine the in vivo kinetics of nuclear proteins. Methods 29:14–28
13. Blonk JCG, Don A, Van Aalst H, Birmingham JJ (1993) Fluorescence photobleaching recovery in the confocal scanning light microscope. J Microsc 169:363–374
14. Phair RD, Misteli T (2000) High mobility of proteins in the mammalian cell nucleus. Nature 404:604–609
15. Seksek O, Biwersi J, Verkman AS (1997) Translational diffusion of macromolecule-sized solutes in cytoplasm and nucleus. J Cell Biol 138:131–142
16. Lukacs GL, Haggie P, Seksek O, Lechardeur D, Freedman N, Verkman AS (2000) Size-dependent DNA mobility in cytoplasm and nucleus. J Biol Chem 275:1625–1629
17. Politz JC, Tuft RA, Pederson T, Singer RH (1999) Movement of nuclear poly(A) RNA throughout the interchromatin space in living cells. Curr Biol 9:285–291
18. Verschure PJ, van der Kraan I, Manders EM, Hoogstraten D, Houtsmuller AB, van Driel R (2003) Condensed chromatin domains in the mammalian nucleus are accessible to large macromolecules. EMBO Rep 4:861–866
19. Abney JR, Cutler B, Fillbach ML, Axelrod D, Scalettar BA (1997) Chromatin dynamics in interphase nuclei and its implications for nuclear structure. J Cell Biol 137:1459–1468
20. Hoogstraten D, Nigg AL, Heath H, Mullenders LHF, van Driel R, Hoeijmakers JHJ, Vermeulen W, Houtsmuller AB (2002) Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair in vivo. Mol Cell 10:1163–1174
21. Rademakers S, Volker M, Hoogstraten D, Nigg AL, Mone MJ, Van Zeeland AA, Hoeijmakers JH, Houtsmuller AB, Vermeulen W (2003) Xeroderma pigmentosum group A protein loads as a separate factor onto DNA lesions. Mol Cell Biol 23:5755–5767
22. Farla P, Hersmus R, Nigg AL, Mari PO, Geverts B, Dubbink EJ, Trapman J, Houtsmuller AB (2004) The androgen receptor ligand domain stabilises DNA binding in vivo. J Struct Biol 147:50–61
23. Stenoien DL, Patel K, Mancini MG, Dutertre M, Smith CL, O'Malley BW, Mancini MA (2001) FRAP reveals that mobility of oestrogen receptor-alpha is ligand and proteasome-dependent Nat Cell Biol 3:15–23

24. Kruhlak MJ, Lever MA, Fischle W, Verdin E, Bazett-Jones DP, Hendzel MJ (2000) Reduced mobility of the alternate splicing factor (ASF) through the nucleoplasm and steady state speckle compartments. J Cell Biol 150:41–51
25. Dundr M, Hoffmann-Rohrer U, Hu Q, Grummt I, Rothblum LI, Phair RD, Misteli T (2002) A kinetic framework for a mammalian RNA polymerase in vivo. Science 298:1623–1626
26. Essers J, Houtsmuller AB, van Veelen L, Paulusma C, Nigg AL, Pastink A, Vermeulen W, Hoeijmakers JHJ, Kanaar R (2002) Nuclear dynamics of RAD52 group homologous recombination proteins in response to DNA damage. EMBO J 21:2030–2037
27. Mattern KA, Swiggers SJJ, Nigg AL, Lowenberg B, Houtsmuller AB, Zijlmans MJM (2004) Dynamics of protein binding to telomeres in living cells: implications for telomere structure and function. Mol Cell Biol 24:5587–5594
28. Sleeman JE, Trinkle-Mulcahy L, Prescott AR, Ogg SC, Lamond AI (2003) Cajal body proteins SMN and Coilin show differential dynamic behaviour in vivo. J Cell Sci 116: 2039–2050
29. Handwerger KE, Murphy C, Gall JG (2003) Steady-state dynamics of Cajal body components in the Xenopus germinal vesicle. J Cell Biol 160:495–504
30. Herrmann A, Sommer U, Pranada AL, Giese B, Kuster A, Haan S, Becker W, Heinrich PC, Muller-Newen G (2004) STAT3 is enriched in nuclear bodies. J Cell Sci 117:339–349
31. Sporbert A, Gahl A, Ankerhold R, Leonhardt H, Cardoso MC (2002) DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters. Mol Cell 10:1355–1365
32. McNally JG, Muller WG, Walker D, Wolford R, Hager GL (2000) The glucocorticoid receptor: rapid exchange with regulatory sites in living cells. Science 287:1262–1265
33. Lukas C, Falck J, Bartkova J, Bartek J, Lukas J (2003) Distinct spatiotemporal dynamics of mammalian checkpoint regulators induced by DNA damage. Nat Cell Biol 5:255–260
34. Becker M, Baumann C, John S, Walker DA, Vigneron M, McNally JG, Hager GL (2002) Dynamic behavior of transcription factors on a natural promoter in living cells. EMBO Rep 3:1188–1194
35. Kimura H, Sugaya K, Cook PR (2002) The transcription cycle of RNA polymerase II in living cells. J Cell Biol 159:777–782
36. Schaaf MJ, Cidlowski JA (2003) Molecular determinants of glucocorticoid receptor mobility in living cells: the importance of ligand affinity. Mol Cell Biol 23:1922–1934
37. Moerner WE, Peterman EJ, Brasselet S, Kummer S, Dickson RM (1999) Optical methods for exploring dynamics of single copies of green fluorescent protein. Cytometry 36:232–238
38. Garcia-Parajo MF, Segers-Nolten GM, Veerman JA, Greve J, van Hulst NF (2000) Real-time light-driven dynamics of the fluorescence emission in single green fluorescent protein molecules. Proc Natl Acad Sci USA 97:7237–7242
39. Siggia ED, Lippincott-Schwartz J, Bekiranov S (2000) Diffusion in inhomogeneous media: theory and simulations applied to whole cell photobleach recovery. Biophys J 79:1761–1770

Adv Biochem Engin/Biotechnol (2005) 95: 201–243
DOI 10.1007/b102215

Deconvolution Microscopy

Jean-Baptiste Sibarita (✉)

Curie Institute, Research Division, UMR144-CNRS, 75005 Paris, France
sibarita@curie.fr

Abstract Since its introduction in 1983, deconvolution microscopy has become a key image-processing tool for visualizing the cellular structures of fixed and living specimens in three dimensions and at subresolution scale. The last 20 years have seen the development of many different applications based on deconvolution microscopy, including a wide variety of optical setup and deconvolution algorithms. This chapter aims to summarize and to describe

in detail the major features of this technology, from theoretical aspects to practical solutions. It will begin by explaining the principle of image formation in three-dimensional optical sectioning microscopy. As deconvolution microscopy provides, in essence, a means of overcoming the limits of optical microscopy, the second part of this chapter is dedicated to the theoretical and experimental description of image generation through a microscope. Methods will be detailed for the determination of point spread function, as a crucial step for the characterization of any optical system and a key preliminary step for image deconvolution. The challenges faced and the various possibilities for determining this function precisely will be discussed. All possible sources of aberrations and image degradation processes will be discussed. In the third part of this chapter, we will introduce the acquisition setup and requirements for compliance between acquisition and deconvolution processes. Typical setups for fixed and living cell observation will be detailed, with key features for optimizing speed and reducing artifacts. In the fourth and last part of this chapter, we will describe, in theoretical terms, the various restoration algorithms commonly used in the field of optical microscopy and will provide results obtained with some of the commercially available packages. We shall conclude by considering the prospects for future solutions (currently under development) aiming to handle more easily the huge amounts of data generated by rapid multi-dimensional living cell microscopy. Designed for use by standard cell biologists and hardware and software engineers and developers, this chapter has been written to provide a clear explanation of the wide-reaching and powerful domain of deconvolution microscopy.

Keywords Deconvolution · Wide-field microscopy · Confocal microscopy · Image restoration algorithm · Point spread function

1 Introduction

Light microscopy provides cell biologists with the unique possibility of examining living samples under conditions similar to those found in the native state. Recent progress in cellular, molecular biology and microscopy has made possible the acquisition of multidimensional data concerning rapid cellular activities. It is now feasible to observe the cellular activities of various organelles within the cell, in three dimensions, at high spatial and temporal resolutions. These observations are achieved by means of optical sectioning microscopy, in which images are recorded while the focal plane is rapidly and precisely raised through the sample. The three-dimensional (3D) acquisition process is repeated at several time points during a given period, leading to the collection of image stacks. This results in four-dimensional (4D) acquisition: three spatial dimensions plus the time dimension. The acquisition of multiple labels by means of such a technique is called five-dimensional microscopy (5D), with wavelength (λ) acting is the fifth dimension.

Living cell microscopy has undergone taken a major leap forward in the last ten years, thanks to the introduction of natural fluorescent probes such as green fluorescent protein (GFP) [2–4] and its derivatives (cyan FP, yellow FP, etc.). These advances, combined with increases in the sensitivity of high-resolution detectors, have made it possible to observe living cells for a long

period of time at low light excitation levels, minimizing the disturbance of cell activity.

The greatest limitation of optical microscopy is spatial resolution, which is in the range of the wavelength of light used. The ultimate goal of cell microscopy is to capture the activity of cell components. In this respect, the resolution of limited numerical aperture wide-field microscopy may not be sufficiently high to optically resolve small organelles. This results in blurring, each focal plane being contaminated with out-of-focus information from the adjacent planes above and below that examined, making the data difficult, if not impossible to analyze. Confocal scanning light microscopy (CSLM) [5, 6] and confocal spinning disk microscopy [7] both provide higher spatial resolution than conventional wide-field microscopy. Unfortunately, these techniques are slower and require higher levels of light excitation than wide-field microscopy. They are therefore not always compatible with the high-speed multi-dimensional living cell microscopy required for the observation of subcellular phenomena. Deconvolution is a computational restoration method that provides an alternative means of producing high-resolution images. Unlike confocal techniques, which physically remove the out-of-focus emission information by means of a pinhole, deconvolution is a mathematical processing method in which computations for the acquired stacks reassign the diffracted light to its original location. As the emitted signal is collected in its entirety by means of a highly sensitive CCD camera, post-deconvolution images may in some cases provide higher resolution than confocal microscopy. Moreover, the sensitivity and dynamic range of wide-field deconvolution microscopy are much higher than those of confocal microscopy, and this may prove critical for the observation of small, dark objects (e.g. vesicles or tubules) in the vicinity of larger, brighter structures (e.g. the Golgi apparatus).

Deconvolution is a heavy computational method now accessible to any laboratory, thanks to the fantastic progress of computer science. A stack of images can now be processed by inexpensive computers, in a time period as short as a few milliseconds to a few hours, depending on acquisition size and the deconvolution algorithm used. Moreover, spatial deconvolution is not strictly limited to wide-field microscopy. Instead, it can also be applied to confocal data, increasing resolution and facilitating more profound structural analysis.

The choice between the various techniques available depends on many parameters, including spatial and temporal resolution, spectral requirements, space and time observation windows, bleaching, and the available equipment. Regardless of the technique chosen, data collection must be optimized and an appropriate deconvolution algorithm selected [8, 9]. The deconvolution process is directly linked to the image formation process. The quality of deconvolution thus depends on the quality of the microscopy. This chapter aims to guide users through the fantastic and wide-ranging world of deconvolution microscopy.

2
Image Formation and Optical Distortions in Three-Dimensional Fluorescence Microscopy

Recent advances in fluorescence microscopy, detectors, motorization and digital image processing have made possible the routine collection of two- and three-dimensional data for biological samples. In three dimensions, the object is optically sectioned plane by plane in a process involving the continuous displacement of the focal plane along the optical axis (Fig. 1). The objective or the sample may be moved to change the focal plane and acquire the corresponding 2D section image. For living cell experiments, particularly if high speed sectioning is required, it is recommended to move the objective rather than the whole sample to prevent morphological changes in the object due to its displacement. Both confocal microscopy and wide-field microscopy involve optical sectioning. Confocal microscopy is based on the illumination of a small, highly focused and diffraction-limited volume with a beam of light, and detection of the fluorescence emitted from the excited volume with a single-element detector. A pinhole is placed in the plane of the image, just in front of the detector, to eliminate a portion of the out-of-focus light emitted, thereby increasing axial resolution. For each focal position, a two-dimensional image is captured point by point, by scanning this single spot in the observation plane. Although fundamentally different, both scanning confocal microscopy and wide-field microscopy are based on the same process of image formation produced by a point source object (Fig. 1).

Even the most perfect of objectives produces a distorted image of the object. The degradation observed during image formation is mostly due to optical blur and photon noise. Optical blur is a consequence of light diffraction through an optical system, resulting in the limitation of optical resolution. This limitation is physical and reproducible and is inherent to any optical system. It can be characterized both theoretically and experimentally, and a precise understanding of this limitation forms the basis of most deconvolution algorithms.

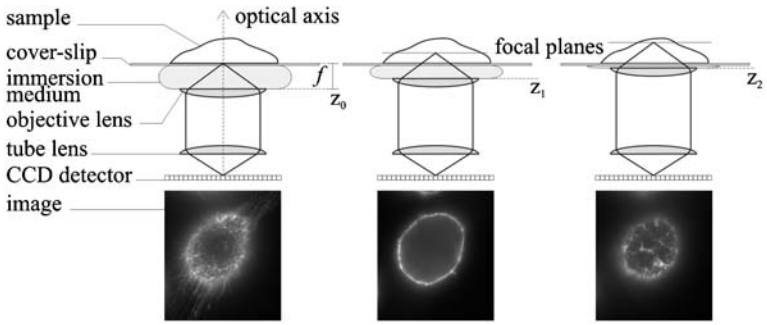


Fig. 1 Optical sectioning microscopy is performed by moving the focal plane of the objective along the optical axis. With infinite optics, the focal plane remains constant and located at a distance *f* from the objective, where *f* is the focal length of the objective

Noise is a random process that is also responsible for image degradation. In digital microscopy, noise arises from the statistical distribution of the photons and electronic devices. Both sources are known and have been described statistically [10]. Photons are the predominant source of noise. The quantum nature of light results in the detected photons following a Poisson distribution. The electronic noise is Gaussian in nature and is negligible in modern detectors in comparison to photon noise. This often results in image quality being purely a function of the number of detected photons, which depends on integration time and the quantum efficiency of the detector. The noise present in the image considerably limits the efficiency of deconvolution algorithms. It must be taken into account, either directly in the algorithms themselves, or by adapted filtering. The level of noise depends on the observation conditions. Most living cell acquisitions aiming to capture rapid 3D events have low signal-to-noise ratios, because cell viability and fluorescence must be preserved.

Figure 2 shows a simulation of an object subject to blur and noise degradation, as observed through a microscope. One very noticeable effect of the blur is the difference in intensity distribution between small and large objects.

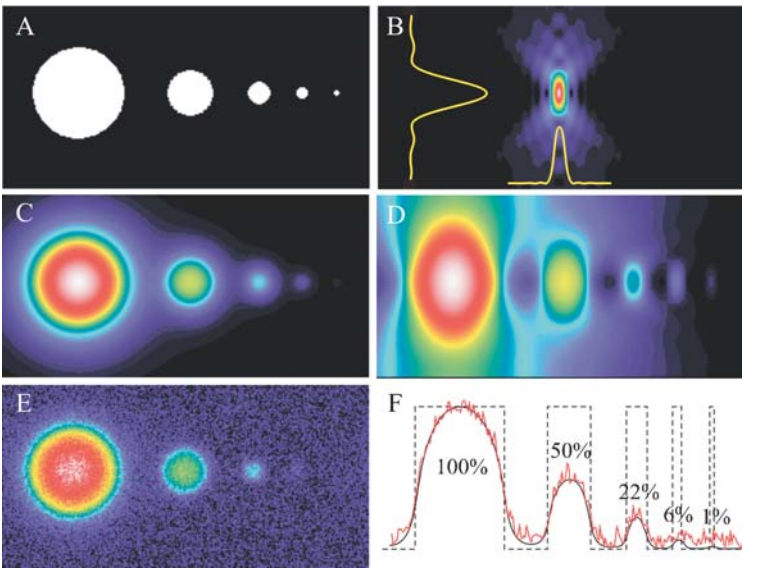


Fig. 2 Three-dimensional simulation of an object observed through a microscope, in the presence of optical blur and noise. The object consists of five spheres with different diameters but the same fluorescence density: **A** initial object; **B** point spread function (PSF) of the microscope; **C**, **D** lateral and axial cross sections of the object after convolution with the microscope's PSF; **E** lateral cross section of the object after blurring and the addition of noise; **F** intensity profiles and percentage of the object's maximum intensity of original (*dashed*), blurred (*black*) and blurred+noisy (*red*) data. Following blurring, the smaller the object, the weaker its maximum intensity is likely to be. Noise reduces the likelihood of detecting small and highly attenuated objects

2.1 Image Formation in an Ideal, Aberration-Free Microscope

Theoretical models [11] of the 3D imaging formation systems in optical sectioning microscopy assume that the optical systems are free of aberrations. The microscope is characterized by its 3D point spread function (PSF), which is the image of a single point source object through the objective. In these ideal models, the PSF is shift-invariant, which means that the image of a single point source is the same, regardless of its position in the object space.

An ideal, aberration-free objective can be described in terms of the wave-like nature of light. A light wave, of wavelength λ, emitted from a point source in the object plane, is a spherical wave of wavelength λ centered on this point source. A perfect lens transforms a spherical wave into another spherical wave. The image of the point source is a point image. According to Huygens' principle, a spherical wave front consists of a multitude of single point sources, each emitting a spherical wave of wavelength λ [11, 12]. All waves from the same wave front have the same phase origin, and interfere together in the image plane, resulting in a diffraction pattern known as the Airy pattern, consisting of a series of concentric spheres. From this pattern, the lateral resolution limit of a perfect objective with finite numerical aperture can be described by the Rayleigh criterion:

$$r_{latteral} = \frac{0.61\lambda}{n\sin\alpha} = \frac{0.61\lambda}{NA} \tag{1}$$

where n is the refractive index of the medium between the specimen and the objective, α the half cone angle captured by the objective, and λ the emission wavelength. NA is the numerical aperture of the objective. This formula gives the minimum distance that can be separated laterally by an ideal objective of numerical aperture NA, for a given wavelength λ, in two dimensions (Fig. 3).

The two-dimensional analytical formulation of the PSF of an aberration-free microscope [8, 11] can be as follows:

$$h(v) = 2\int_{2}^{1} P(\rho)\, J_0(v\rho)\rho d\rho$$
$$\text{with } \rho = \frac{\sqrt{x^2+y^2}}{a} \text{ and } v = \frac{2\pi NA}{\lambda}\sqrt{x^2+y^2} \tag{2}$$

where P is the circular pupil function of radius a, J_0 the first order Bessel function, λ the emission wavelength and NA the numerical aperture.

Spherical waves issued from the wave front interfere not only in the image plane, but also throughout the 3D space. Consequently, the image of the point source located in the object plane is a three-dimensional diffraction pattern, centered on the conjugate image of the point source located in the image plane.

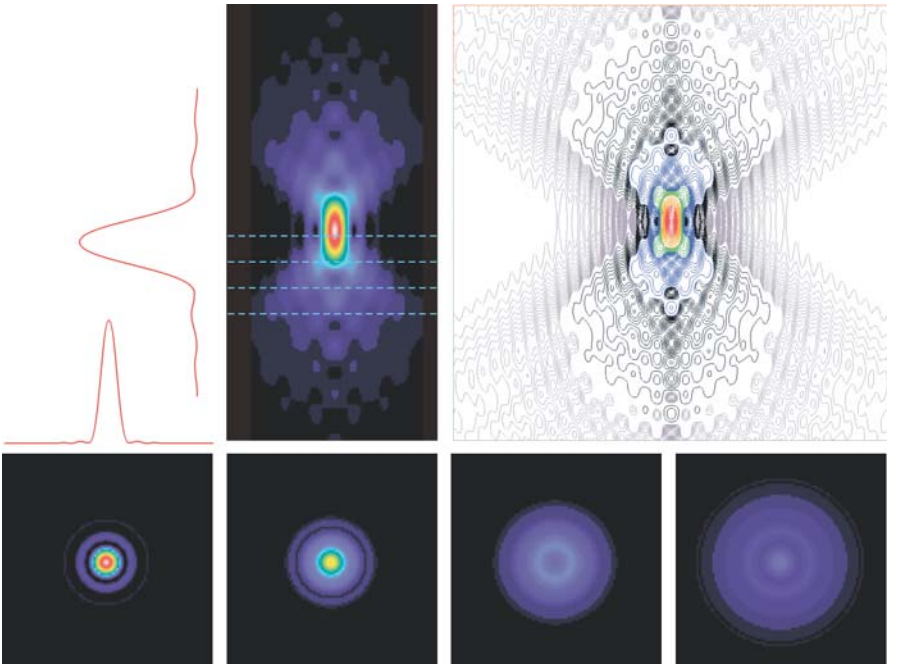


Fig. 3 Diffraction pattern of an ideal, aberration-free objective, in one, two and three dimensions

With a perfect lens, this function is symmetric along the optical axis [11]. The three-dimensional analytical formulation of the PSF for a perfect, ideally corrected and aberration-free objective is given by [6, 13]

$$h(v,u) = 2\int_0^1 P(\rho)\, J_0(v\rho)\, \exp\left(\frac{iu\rho^2}{2}\right)\rho d\rho$$
$$\text{with } u = \frac{8\pi}{\lambda} zn \sin^2(\alpha/2) \tag{3}$$

In the axial direction, intensity distribution is similar in shape to the Airy disk in the lateral direction. As for lateral resolution, axial resolution defines the minimum distance that can be separated in the z-axis. It is described by the distance between the maximum intensity of the central bright region and the first point of minimum intensity along the z-axis, and is given by

$$r_{axial} = \frac{2\lambda n}{NA^2} \tag{4}$$

Unfortunately, this 3D image formation model has been shown to be experimentally inaccurate [13, 14], especially for oil immersion objectives of high

numerical aperture. The next section describes the spherical aberrations that may arise in this case.

2.2 Image Formation in an Oil Immersion Objective with Aberrations

Classical image formation theory, as described above, assumes that the 3D image formation process of a fluorescence microscope is linear and shift-invariant. This implies that a single 3D PSF – the image of a single point source – is sufficient to describe completely image formation throughout the 3D object space. Unfortunately, shift invariance is not always demonstrated, and PSF is all too often a function of the location of the point source in the object space. Most frequently, shift invariance does not apply in the axial direction. Other sources of aberration, including stigmatism, coma and curvature of the field, are negligible and well corrected in high-quality objectives.

Axial shift variance results from the use of objectives under non-optimal conditions. This is unfortunately the case for the observation of biological specimens, particularly in living cell microscopy, in which the emitted light passes through media with different refraction indices. Optical paths differ from the optimal paths for which the objectives were designed, generating spherical aberrations that can severely affect the ideal image of a point source. In general, only the plane position just above a coverslip of a given thickness and refractive index and separated from the objective by immersion oil of a specific thickness and refraction index will produce an aberration-free image, as predicted by classical image formation theory. Any object placed elsewhere within this plane is subject to spherical aberrations.

In practice, it is not easy to carry out microscopy in ideal conditions. For example, the refractive index of a living cell culture is closer to that of water (n=1.33) than to that of immersion oil (n=1.516), and is far from the value for which optics are designed. Consequently, when light is focused deep into a cell well below the coverslip (the point for which the objective is designed), it passes through various layers with different refractive indices, resulting in the generation of spherical aberrations. Spherical aberrations are characterized by axial asymmetry in the shape of the PSF, an increase in the flare of the PSF, and a decrease in its maximal intensity. They decrease both resolution and brightness in wide-field and confocal microscopy. They gradually increase as the focal plane moves deeper into the sample (Fig. 16). They are therefore not seen when recording beads mounted on the coverslip.

Gibson and Lanni [15] proposed a theoretical model taking into account the consequences of the optical path modifications that occur when light rays follow trajectories other than that for which the optics were designed. They calculated the optical path difference (ODP), which is the difference between ideal trajectories and real trajectories, taking into account parameters determining path fluctuation: thickness (*toil*) and refractive index (*noil*) of the immersion oil, thickness (*tc*) and refractive index (*nc*) of the coverslip and the refractive

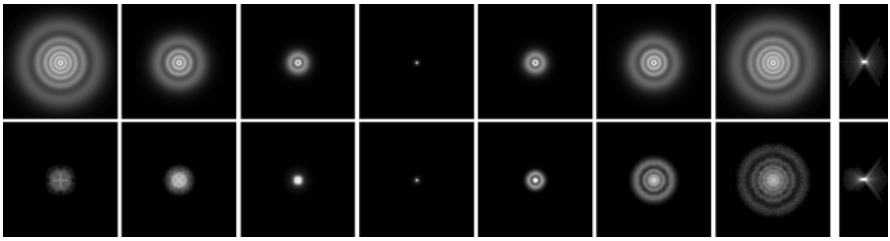


Fig. 4 Comparison of theoretical aberration-free PSF (*top*) and empirical PSF (*bottom*) in the green emission range, for a ×100 1.4 *NA* objective, with an oil immersion refractive index of $n=1.518$

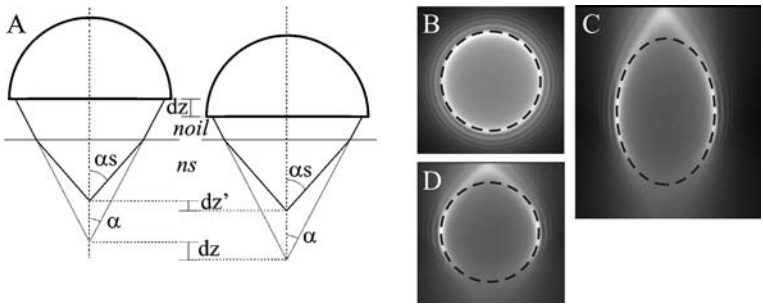


Fig. 5 **A** Axial distance mismatch due to the presence of layers with different refractive indices along the optical path. **B** Lateral section of the bead. **C, D** Observation of a 6 µm fluorescent calibration bead without (C) and with (D) distance calibration

index (*ns*) of the sample. Any deviation of any of these parameters from the optimal values for which the optics were designed will cause spherical aberrations. This model partially describes the spherical aberrations, which can be observed in practice when recording point spread functions from microbeads various distances from the coverslip or when light passes through successive layers with different refractive indices (Fig. 5A).

2.3 Distance Calibration

Another source of distortion that must be taken into account in optical sectioning microscopy is the geometric distortion generated by focusing light through layers with different refractive indices (Fig. 5).

If we ignore the difference between the refractive indices of the immersion oil and the coverslip, the displacement error can be expressed as

$$dz' = \frac{\tan\alpha}{\tan\alpha s}\, dz = \frac{\tan(\arcsin(NA/n))}{\tan(\arcsin(NA/ns))}\, dz \tag{5}$$

This equation quantifies the axial distortions arising from the use of media with different refractive indices between the immersion oil and the sample. This implies that, for the entire light cone to be focused, *NA* must be less than *ns*. For objectives with high numerical apertures, $NA>ns$, and when $ns<n$ (typically when observing living cells with an 1.4 *NA* oil immersion objective) the numerical aperture falls below *ns* and some of the light is lost due to reflection to the medium interface.

This mismatch error may be as high as 3, in the case of living cell (ns=1.33) observed with an oil immersion objective ($noil$=1.516) with a high numerical aperture. This error can be corrected experimentally by attempting to match the immersion medium of the objective with that of the sample (e.g. using a water immersion objective for living samples or by matching the mounting medium and the immersion oil in terms of refractive index for fixed samples). Practically, it is necessary to calibrate the entire acquisition system by imaging fluorescent calibration beads. Figure 5 illustrates the importance of calibrating the microscope. Calibration can be achieved by inputting the measured axial correction factor directly into the acquisition software or, if this is not possible, by interpolating the data once acquired. This step is crucial for acquisitions of three-dimensional data with optical sectioning. Sampling errors can cause major problems in deconvolution processes using a theoretical PSF, in which the details of sampling must be precisely known. More generally, without this correction step, acquisitions can be achieved either by oversampling (when $ns<n$) or by downsampling (when $ns>n$). However, oversampling may lead to photobleaching and damage to the cell whereas downsampling may result in a loss of information.

2.4 Determination of the Point Spread Function

The PSF describes the response of the microscope, for each point of the sample imaged. This information is required for most deconvolution algorithms because it describes the way in which the objective distorts the image during the acquisition process. Assuming that the image formation process is linear and shift-invariant (this point will be discussed later), a single function completely describes the functioning of the instrument at any point in the specimen. The accuracy and quality of the PSF are essential to ensure the correct performance of any deconvolution algorithm. Noise, incorrect estimates of aberrations and incorrect scaling of the PSF may cause major artifacts in the restored image.

The PSF may be calculated theoretically or measured empirically. There are advantages and disadvantages to both methods. In addition to simplifying the process by avoiding the need for time-consuming PSF extraction, the main advantage of theoretical PSF calculation is the absence of noise. It is therefore possible to determine, with similar levels of accuracy, the 3D contributions of the PSF within and outside the focal plane, at a distance at which detectors are not sensitive to capture any signal above the noise. Nevertheless, there are many

reasons for preferring an empirical model. First, although good theoretical PSF models exist (see above), they apply only to perfect lenses and well defined and calibrated optical paths, which are, of course, never found in practice. Small variations in the regularity of the lens may change the shape of the PSF in ways that cannot be predicted by theoretical models. Second, some spherical aberrations are very difficult to predict theoretically. Any discrepancy between the theoretically estimated PSF and the actual PSF of a microscope decrease the efficiency of the deconvolution and may result in artifacts in the results.

Although theoretical estimates cannot provide an exact measure of the real PSF of objectives, they may nonetheless be of value for estimating expected values in ideal experimental conditions. This requires precise knowledge of the microscope setup and the use of a theoretical PSF generator, taking into account the spherical aberration primarily due to the mismatch of refractive indices along the optical path, as described above. This model requires precise knowledge of the numerical aperture of the microscope, the objective to object path length, the emission wavelength, the refractive index of both immersion and sample media and the thickness of the coverslip. In addition, (x, y, z) sampling is also required to represent the final result in a discrete manner. Sampling identical to that for acquisition is used, in general, so that the extracted PSF can be used directly in deconvolution or compared with the measured PSF. If a theoretical PSF-based deconvolution is performed on acquired data sets, then absolute calibration of the microscope optics must be carried out first (see previous section).

For accurate determination of the microscope's PSF, it is advisable to measure the PSF under the same conditions as used for the sample itself. Each objective is unique, and any variation in the regularity of the lens may affect PSF shape, particularly for immersion objectives with a high *NA*. Moreover, PSF measurement can detect problems in the microscope setup that may not be evident when observing complex biological samples. There are many possible sources of problems, including oil mismatch, objective defect, vibrations from the heating system or CCD cooling, and unexpected changes in the properties of the oil when working at different temperatures (e.g. room temperature and 37 °C). In the face of these potential sources of signal degradation, it is strongly recommended that the microscope's optical properties, stability and reproducibility should be checked with microbead samples (Fig. 6).

PSF can be measured experimentally in a simple manner involving the acquisition of images of subresolution fluorescent beads. These microbeads, commercially available as a concentrated solution, are used to represent a point source for evaluation of the system. The smallest beads possible should be used in theory, but small beads are weaker in intensity and bleach more rapidly. In practice, beads with a diameter less than half the resolution are used: beads of 0.15 μm in diameter, from Molecular Probes Inc., for a 1.4 *NA* oil immersion objective with a lateral resolution of 0.4 μm at an emission wavelength of 500 nm. Beads are diluted to isolate each PSF without overlap, and are dried onto the coverslip. The coverslip is then reversed, mounted in the same medium as used

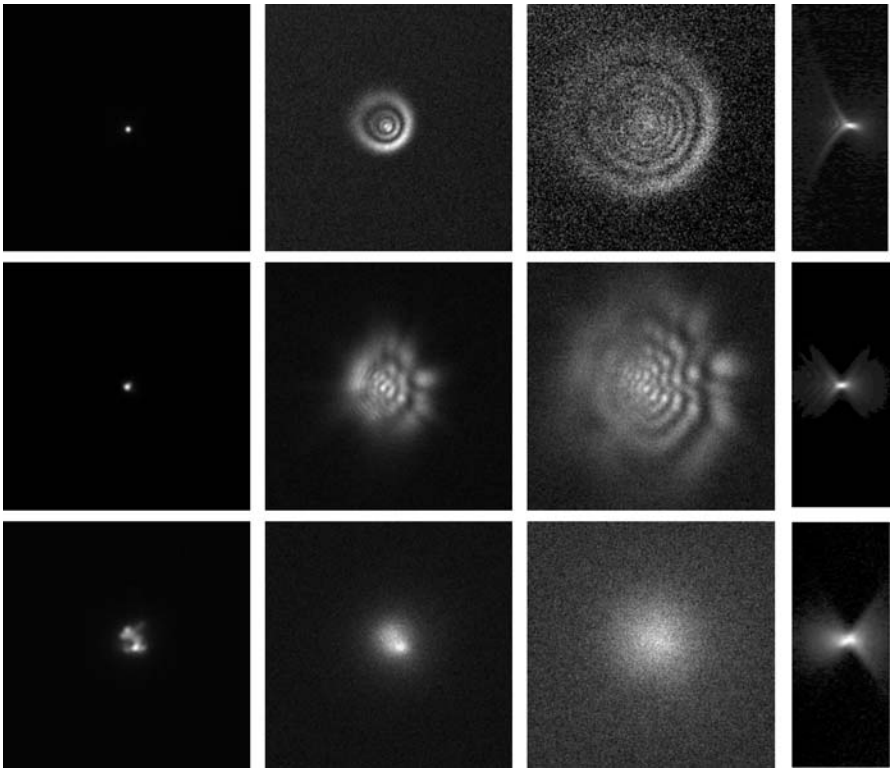


Fig. 6 Examples of artifacts: optical problem (*top*), oil mixing (*middle*) and vibrations (*bottom*)

to embed the sample, and attached firmly to the slide with nail polish to ensure that the specimen remains stable during the acquisition. Optical sections of bead samples must be achieved, using an optimal sampling method, compatible with optic axial resolution and deconvolution, as for the sample itself (see below). An image stack can then be collected, adjusting excitation and exposure time parameters to maximize the dynamics in the center of the bead without saturating the signal, and to prevent bleaching. To increase the signal-to-noise ratio of the image, especially in the out-of-focus regions in which signal intensity is weak, it is recommended to average acquisitions of multiple beads. The acquisition protocol can be facilitated by using the same bead for the averaging, keeping bleaching minimal. The signal-to-noise ratio can be increased significantly by applying circular averaging with respect to the optical axis, if, of course, the original bead acquisitions show the symmetry predicted by the theory (Fig. 7).

If possible, all conditions resulting in a change in PSF shape should be characterized. Identified sources of changes in optical properties include (1) emission wavelength, (2) axial distance to the coverslip and (3) refractive indices.

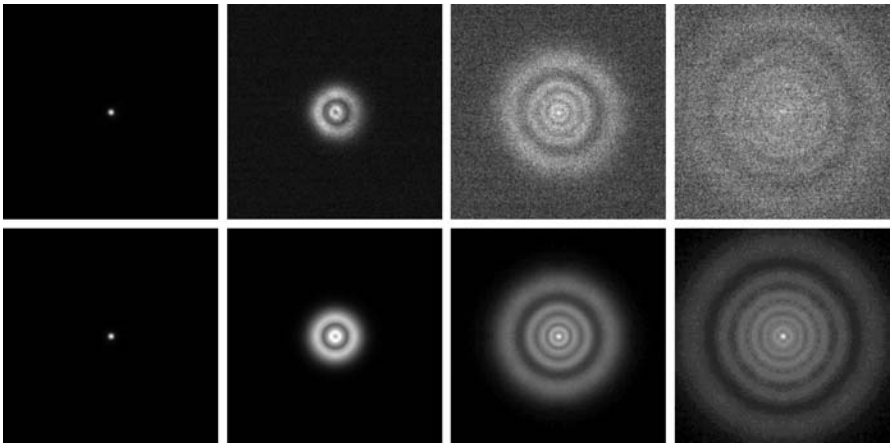


Fig. 7 Experimental PSF extraction: raw data (*top*), and data after bead averaging and circular symmetry processing (*bottom*)

1. As PSF is a function of emission wavelength [16], the objective must be characterized with different color beads. Multicolor beads (Tetraspec, Molecular Probes Inc.), conjugated with four different fluorescent probes, provide a broad range of types of emitted light (UV, green, red and IR), facilitating spectral characterization of the objective. Moreover, these measurements can be used to check, and even to correct after acquisition, the non-apochromatism of the lens (Fig. 8).
2. As described above, spherical aberrations break shift invariance in the axial direction. Consequently, if thick objects need to be observed in their entirety (more than about 20 μm), or if objects located far from the coverslip are to be observed, it is advisable to characterize the microscope's PSF at various z levels. This can be done theoretically, or experimentally by recording flu-

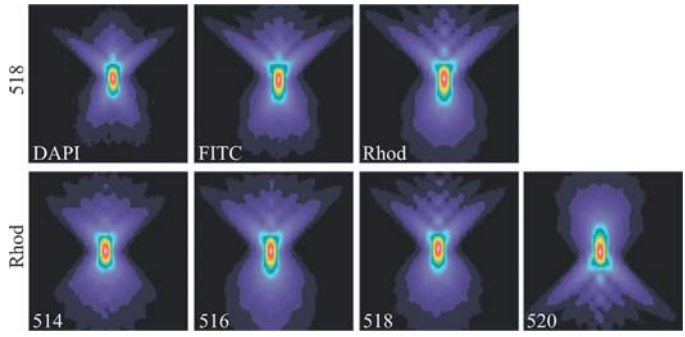


Fig. 8 *Top*: PSF for Dapi, GFP and rhodamine wavelengths, using an oil with a refractive index of 1.518. *Bottom*: PSF for the rhodamine wavelength, using oils with refractive indices of 1.514, 1.516, 1.518 and 1.520

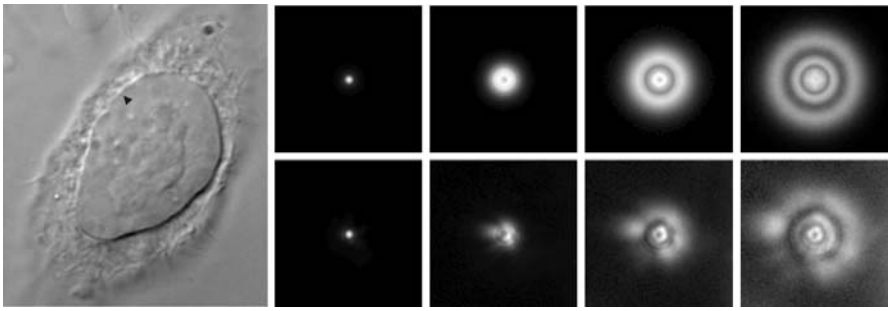


Fig. 9 Distortion of a PSF due to the natural fluctuation of local refractive indices. Fluorescent beads were injected into a cell before fixation. *Top*: Experimental PSF determined in the absence of a cell. *Bottom*: Distorted PSF determined from a bead within a cell. *Left*: DIC image of the cell, showing local refractive index variations; the *black arrow* indicates the location of the bead

orescent microbeads lying on the surfaces of gels of a given thickness and refractive index.

3. Ideally, PSF should be measured under the same conditions as the specimen itself. In practical terms, this implies that beads should be observed after injection into the specimen, which is generally not possible. Neither theory nor experimental methods can predict the change in PSF caused by the passage of light through a heterogeneous medium of unknown refractive index. Estimation of the local refractive index can be achieved by acquiring pre-DIC images on the sample. Its knowledge allows to model the PSF distortions for each point of the whole volume, using ray-tracing techniques [17], without the need to inject microbeads into the cell. This local information is of value because it makes possible to perform deconvolution with precise knowledge of the PSF. However, under such circumstances, the PSF cannot be assumed to be shift-invariant, limiting deconvolution due to the extremely large amount of computation time required (Fig. 9).

This illustrates the difficulties and limitations of the PSF extraction process for the observation of biological specimens. Neither theory nor experimental determinations can precisely predict the PSF that should be used for the deconvolution of a complex biological sample. Empirical measurements are clearly the most appropriate method, but comparison of the values obtained with those predicted by the theoretical model can help to solve problems in the optical setup. One good solution would be to fit an analytical model to the experimental acquisition, to make the best use of both approaches. One recent study adopted such a method [18] recovering iteratively the theoretical shape of the pupil function, including phase information, from bead acquisitions. This made it possible to model aberrations causing axial asymmetry. However, this method cannot be used to overcome problems with shift invariance, an essential requirement of deconvolution algorithms that often does not apply completely in practice.

Thus, the image formation process in optical sectioning microscopy differs from classical image formation theory. Variations in refractive index along the optical path and lens imperfections are the main causes of spherical aberrations. These aberrations are difficult to describe perfectly in theoretically terms and cause a breakdown of axial symmetry and shift invariance properties. Shift invariance property is in some cases very useful for highly simplified mathematical descriptions of the image formation process, mostly for deconvolution (see below for more details). Shift invariance can be assumed to apply in a certain range of thickness, particularly if the optical path is close to the ideal case for which the optics were designed. It is therefore important to characterize experimentally and to adjust the optical setup before routine acquisition and deconvolution, to stick as closely as possible to the ideal case. This can be achieved by selecting the most appropriate objective with the highest possible quality and by adjusting the refractive index layers along the optical path. As each lens is unique, different objectives must be tested and the objective providing the PSF closest to the theoretical value should be used. In practice, it is a good idea to characterize the microscope using the coverslip and to assume axial shift invariance in a thickness range of 10 to 15 μm. This makes it possible to observe a cell monolayer by optical sectioning deconvolution microscopy. If structures deeper within the cells are to be investigated, then the microscope should be characterized at a z distance close to the volume of interest. In all cases, the z-calibration procedure described in the previous section must always be carried out first.

3 Image Acquisition

In optical sectioning microscopy, a three-dimensional stack of two-dimensional images is recorded. Each 2D image corresponds to object acquisition at one focal plane position.

3.1 Sampling

As acquisitions are digital, sampling constraints must be checked to maximize the resolution of the objective. The Nyquist theorem states that for a frequency-limited system, which is the case for any numerical aperture-limited objective with a given resolution limit r, a sampling Δ of half of the resolution is sufficient for the digital recording of the entire signal with no loss of information. For 3D microscopy, the theorem can be expressed for the lateral and axial directions as follows:

$$\Delta_{xy} \leq \frac{r_{xy}}{2} \text{ and } \Delta_z \leq \frac{r_z}{2} \tag{6}$$

where r_{xy} and r_z denote the lateral and the axial resolutions, and Δ_{xy} and Δ_z the lateral and axial samplings. In other words, images must be sampled using at least two points per resolvable element.

In wide-field microscopy, lateral sampling is achieved by the CCD chip and is a function of the CCD pixel size P_S, camera binning b, and objective magnification M, as follows:

$$\Delta_{xy} = \frac{P_S b}{M} \tag{7}$$

By incorporating the lateral resolution equation (Eq. 1) into Eq. 7, the lateral sampling criterion can be rewritten as

$$\frac{NA}{M} \leq \frac{0.61\lambda}{2P_S b} \tag{8}$$

This relationship provides the ideal objective range, for a microscope setup with a given CCD chip, for checking the Nyquist theorem and ensuring compatibility with deconvolution.

For scanning confocal microscopy, lateral sampling is subject to no such constraints and may be ideally adjusted with the zoom command to satisfy the Nyquist theorem. For both wide-field and confocal microscopes, axial sampling must also be adjusted by selecting an appropriate plane spacing distance.

However, this criterion produces the minimum necessary sampling step for obtaining a discrete signal from a real continuous signal. In practice, it is recommended to slightly oversample the signal to optimize separation of the high-frequency noise from the signal itself. Ideally, it is good practice to sample three points per resolvable element. However, practical aspects must be taken into account, particularly when observing living samples. Oversampling increases the number of images required for a given acquisition volume. This results in a significant increase in phototoxicity, acquisition time and photobleaching, each of which may cause artifacts in optical sectioning microscopy combined with deconvolution, and must be minimized:

- Phototoxicity results from prolonged illumination when observing living biological samples. It results in irreversible light-induced damage, mainly due to emission of free radicals, which may lead to changes in the behavior of the living sample, and even to the death of the sample.
- Increasing the acquisition time per time point for living samples may generate artifacts in the deconvolution process. Given that the goal of deconvolution is to "reassign" the light to its original place, the objects must not move within the resolution limit of the microscope during the optical sectioning acquisition process.
- Fluorescence is produced by a cycle in which an electron of a chromophore is excited from its ground state to an excited state, and then relaxes back to

the ground state, resulting in the emission of a photon. Instead of displaying the expected fluorescence relaxation, the excited state can undergo a variety of photochemical reactions that destroy fluorescence. This fluorophore destruction is called photobleaching, regardless of the underlying mechanism responsible. The repeated and prolonged exposures needed to acquire multidimensional data by optical sectioning microscopy result in the continuous loss of cell fluorescence. This loss of fluorescence, if arising during a single stack acquisition, may result in the incorrect estimation of the real light intensity distribution. Consequently, the restoration step may lead to the artifactual reassignment of light. This can be prevented by carrying out a pre-processing step to correct z-bleaching [19]. Undesirable fluctuations in illumination, such as lamp jittering, must also be corrected. If bleaching occurs gradually during time-lapse acquisition, it results in a progressive decrease in signal-to-noise ratio from stack to stack. As noise limits the efficiency of the deconvolution process, this may result in unstable data that are non-reproducible from one stack to another.

Another important criterion is sampling volume: ideally, all the emitted light should be collected for reassignment to its original location. In wide-field microscopy, the light emission cone extends to a few microns above and below the emission plane. This implies that a few microns above and below the peripheral planes containing the object should be acquired so that all the out-of-focus information can be taken into account in the deconvolution process. This is not feasible practically, resulting in a deconvolution process that is less efficient in the peripheral planes than in the middle planes. This "missing information" problem is even greater if only the subvolumes of interest are acquired. In such cases, corresponding mostly to living-cell microscopy in which rapid processes must be acquired over a short period of time, the number of planes must decrease with the observation area. This results in some of the signal not being captured and therefore not being taken into account for deconvolution. This limitation can be partially overcome by means of an intermediate approach involving the use of optical sections with a varying interplane distance [20]: small enough in the volume of interest for checking of the sampling criterion, but larger further away. Unfortunately, this acquisition protocol is not compatible with the high-speed acquisition of the streaming mode described below (Fig. 10).

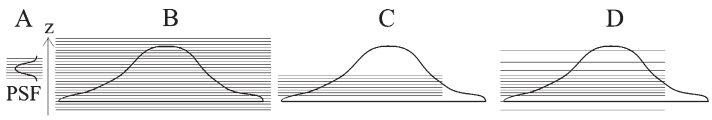


Fig. 10 **A** PSF and optimal sampling (sampling criterion). **B** Ideal full-object sampling. **C** Practical subobject sampling, in which only the volume of interest is acquired, to reduce exposure time. **D** Intermediate object sampling using varying steps, in which only some of the out-of-focus information is collected

3.2 Optical Setup

In this section, wide-field deconvolution microscopy will be treated as an alternative to confocal microscopy because it is more flexible, and under some circumstances, more appropriate. Despite the use of common equipment, fixed and living sample acquisition protocols will be treated separately because the different constraints associated with these protocols ultimately end up leading to individual setups.

A deconvolution microscope may be assembled from individual components or purchased as an integrated solution. Although time-consuming, assembly offers a much higher level of flexibility and is generally less expensive than an integrated solution. This option makes it possible to build up a system to fulfill precisely the requirements of the research group, whereas integrated systems provide a more general solution. Moreover, there is continual progress in imaging techniques, optical devices, detectors, motorization and other elements, and a flexible "custom-assembled" setup makes it much easier to update the system. The essential components of a deconvolution microscope include: a standard microscope mounted on an anti-vibration table, a microstepping motor to control the focus, a fluorescence illumination device – including excitation filter, dichroic mirror, emission filter and shutter elements – a cooled high-resolution CCD camera and software for optimizing the control and synchronization of all these elements. It may be useful to maintain control of the bright-field transmission mode, to facilitate localization of the tagged fluorescent proteins within the cell. The main differences between fixed and living sample experiments, regardless of the temperature and atmospheric controls for living cells, are the fluorescence setup and the way in which multicolor images are collected.

3.2.1 Fluorescence

In fixed-sample experiments, standard single-color filter blocks are used in combination with white-light illumination and a shutter. Multilabels are then acquired, color-stack by color-stack, by simply changing the dichroic mirror block device from one position to another. This process makes use of the high intensity of the standard equipment, provided by single dichroic filter blocks. This equipment is more efficient and specific than multicolor dichroic filter blocks. With the standard system, only the dichroic block, the shutter and the z-motor have to be motorized. If a standard mechanical shutter is used, it can be left open during stack collection for faster acquisition.

For living-cell microscopy, multilabeling acquisition is constrained by the fact that objects are likely to move during the collection of stacks. Multicolor images must therefore be acquired as rapidly as possible, and the sequential stack-color mode described above is inappropriate. There is a single solution to this problem: using a multiband pass dichroic mirror and dissociating

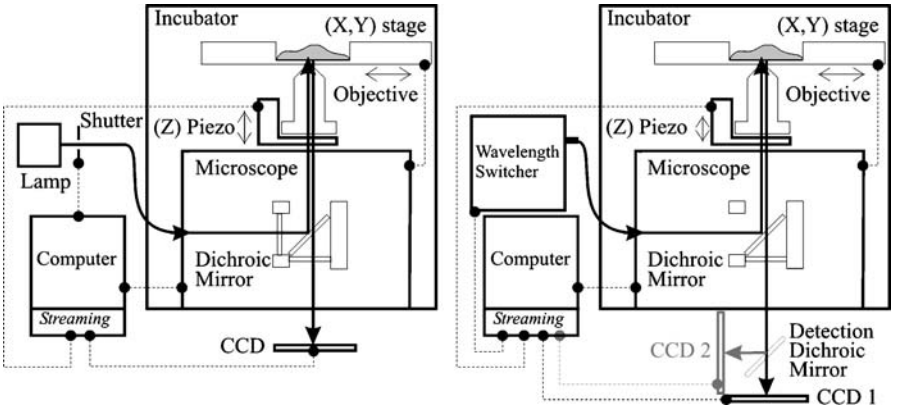


Fig. 11 Typical optical setup of a fast multidimensional microscope for fixed (*left*) and living (*right*) cell observation. For fixed-sample acquisition, image stacks are acquired sequentially, one color at a time. Single dichroic mirrors are used and exchanged to switch between colors. For living-cell microscopy, the dichroic mirror is fixed and contains a multiband filter. In sequential mode, for each focal position, excitation wavelengths are changed with a fast wavelength switcher, and only one CCD camera is needed. In simultaneous mode, several CCD cameras are used in combination with an adapted detection dichroic mirror

excitation and emission. For each plane, different wavelengths are sequentially or simultaneously acquired:

- In the sequential mode, excitation wavelengths for a given observation succeed each other. Emission wavelength can be fixed with a multi-band emission filter or synchronized with excitation part, using single-band emission filters.
- In the simultaneous mode, multi-excitation band-pass filters are used to excite all the fluorophores at the same time. Dichroic emission beam splitters are used at the detection level to separate emission wavelengths and to record multicolor images simultaneously. Images are recorded either by splitting the image fields on a single camera or by using several cameras (Fig. 11).

3.2.2 Detection

The detector and the means of controlling it are of key importance for high-speed acquisitions. With the ever-increasing sensitivity and speed of detectors, digital cameras make it possible to continually increase the speed of recording of weak signals. Up to 100 frames per second can be recorded, depending on the type of camera and the size of the region observed. An understanding of the acquisition process depends on knowledge of the workings of CCD cameras, the details of which can be found elsewhere [12].

A CCD chip consists of an array of small isolated photodiodes. The acquisition process consists of three steps: (i) charge integration, (ii) charge transfer

and (iii) read-out. During charge integration, photons are captured by each individual diode during a given period of integration, also known as the exposure time. Photon absorption results in the formation of electron-hole pairs, which move toward their respective electrodes, and are accumulated in a potential well. During charge transfer, the charges stored in each pixel are shifted to a serial readout register, by means of a cascade of synchronized parallel and serial shifts. Once in the serial register, the readout process starts. Packets of electrons are sequentially read, amplified and converted into an analog signal proportional to the number of stored charges in each pixel. Finally, this raster scan of the CCD chip is converted into a digital signal and loaded into the memory of the computer via a frame grabber.

In a standard full-frame CCD chip, three steps must be performed sequentially before the next image acquisition process begins. For frame transfer and interline transfer CCD technologies, the presence of an intermediate buffer makes it possible to discharge the potential well rapidly by rapid parallel shift. Once shifted to the readout buffer, the next acquisition sequence can start, continuing in parallel with the readout process.

Neighboring pixels can be grouped together to form a larger "super-pixel", and this process can be used to increase the acquisition rate. This technique, known as binning, has two positive consequences: first, the signal-to-noise ratio is increased by increasing the absorption surface, and second, the number of pixels transferred is reduced. Unfortunately, binning increases the pixel size of the image, thereby reducing sampling frequency. Consequently, the constraints imposed by the sampling criterion (see above) may not be fulfilled, possibly decreasing image quality and resulting in images incompatible with the deconvolution process. In practical terms, the same tradeoff is always made between speed, sensitivity and resolution (Fig. 12).

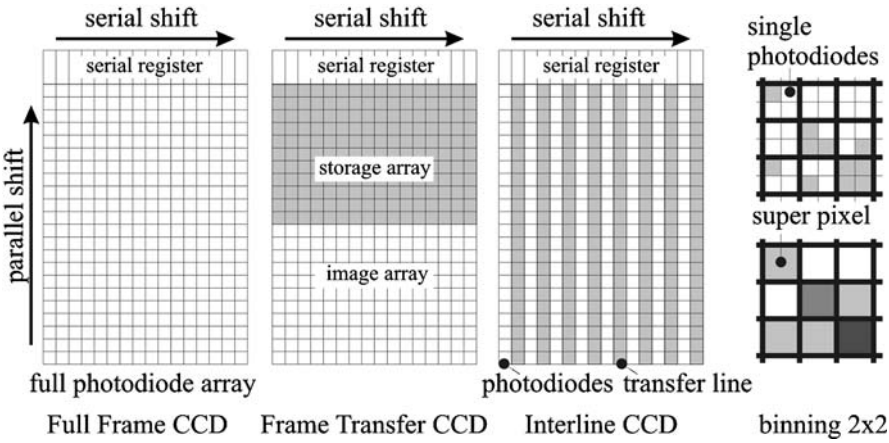


Fig. 12 CCD architecture of full frame, frame transfer and interline CCD and details of the binning mode

3.2.3 Streaming

Image acquisition with frame and interline transfer CCD cameras can clearly be optimized by making the integration and read-out processes function in parallel. This maximizes acquisition speed, with image stacks collected in RAM at maximum speed. This operating mode, called streaming, is optimal if the readout time (TR) is shorter than the exposure time (TE). In this case, acquisition time depends only on exposure time. In practice, once the exposure time is determined from single images produced with a sufficiently high signal-to-noise ratio, the acquisition region should be adjusted with respect to the camera readout speed. For rapid living-cell microscopy, binning and streaming are often combined (Fig. 13).

In multidimensional microscopy, focus and wavelength changer devices must be controlled between image acquisitions. To maintain streaming efficiency, z-streaming and lambda-streaming have been developed for high-speed multicolor stack collections. During this process, the z-position and/or excitation wavelength is changed during the shift register step in streaming mode. This solution requires high-speed devices responding within a few milliseconds, such as piezo-electric focus devices, DG4 or monochromator illumination devices.

Nevertheless, during streaming, acquisition parameters such as exposure time and plane distance cannot be changed. This constrains exposure time, which must be identical for each color during a fast multicolor acquisition, and precludes the use of a variable z-step as described earlier.

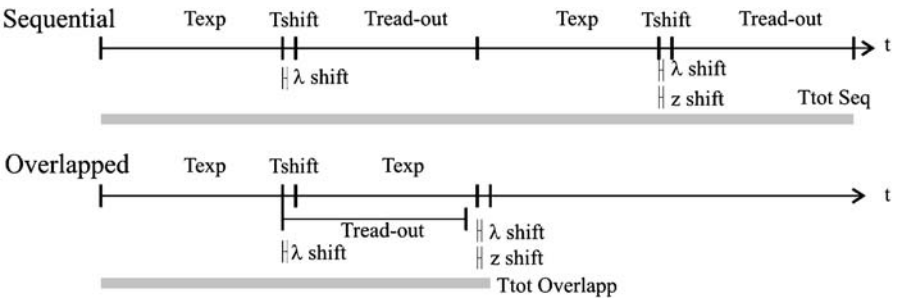


Fig. 13 Time-course diagrams of sequential and overlapping (streaming) acquisition modes

4 Deconvolution

Deconvolution is a computerized inversion method for restoring an image distorted during the image formation process, based on prior knowledge of the degradation phenomenon. In microscopy, the goal is to reassign the optical blur to its original position and to reduce statistical noise. Image restoration belongs

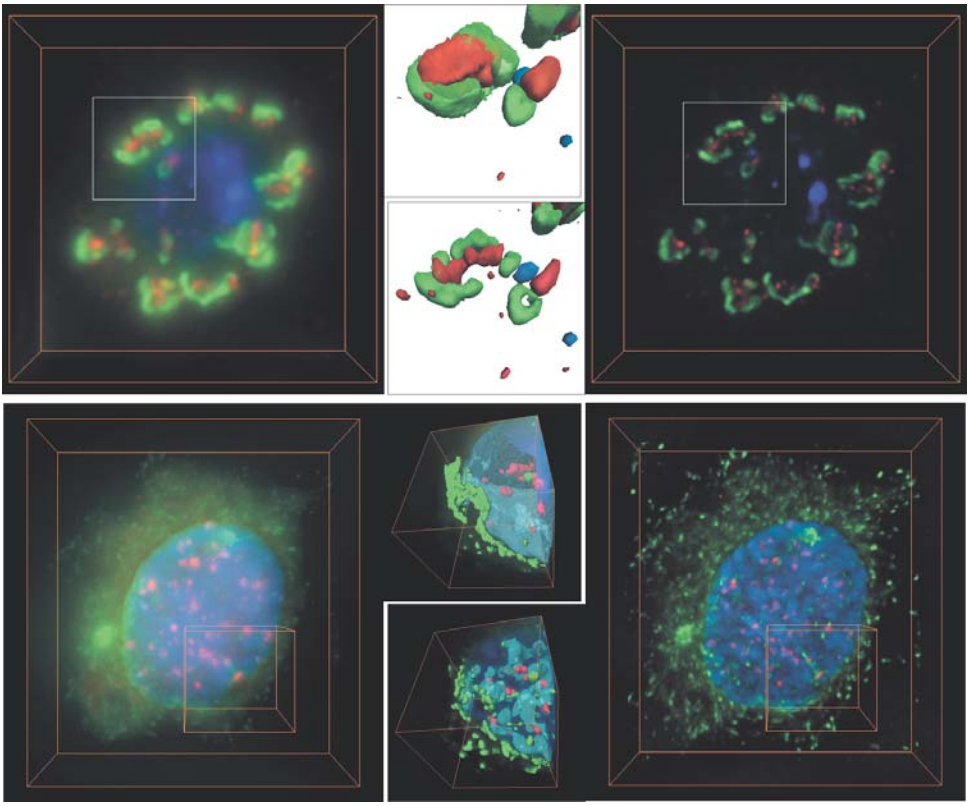


Fig. 14 *Top*: Volume rendering by maximum intensity projection (*left and right*) and surface rendering (*middle*) of drosophila S2 cells stained for two different antigens localized to the Golgi apparatus. Golgi in Drosophila cells consists of several independent units dispersed into the cytoplasm. The deconvolution clearly reveals that these antigens are localized to different but closely apposed regions of Golgi units. Courtesy of I. Kouranti and A. Echard, Institut Curie. *Bottom*: Volume rendering by maximum intensity projection (*left and right*) and surface rendering (*middle*) of HeLa cells labeled for $p150^{Glued}$ (*green*), centromeres (CREST, *red*) and DNA (DAPI, *blue*). Initial data (*left*), restored data (*right*) and zoom on the surface rendering (*middle*) to illustrate the gain of resolution provided by the deconvolution and its contribution to the segmentation process. This prophase cell displays the classic pattern of labeling of microtubule plus ends by p150. After deconvolution, centromeric regions from one pair are resolved into two contiguous spherical regions and the DNA condensing events occurring in the prophase nucleus are visible in the segmented data. Nevertheless, the pattern corresponding to bright green dashes, easily visible in the volume rendering image, displays segmentation into ball-like structures. This illustrates the limitation of biological object segmentation, which cannot always be resolved, even with the best deconvolution methods available. Courtesy of F. Cordelières, Institut Curie

to the ill-posed problem category and does not generate a single solution. The result is an estimate of the object $\hat{o}$ that is closer to the observed object o than the acquired image i. In addition to improving visualization of the details of the image, deconvolution also facilitates segmentation and analysis, as illustrated in Fig. 14.

In the next section we will provide some mathematical definitions of the problem and will then go on to detail the most commonly used algorithms.

4.1 Degradation Model and Mathematical Formalism

The degradation process can be modeled as an operator H, which, together with the noise N, operates on an input image o to produce a distorted image i. Digital deconvolution is a mathematical process that computes an object's approximation $\hat{o}$, given the acquired image i, knowledge of the degradation model H, and the noise N. In general, knowledge of N is limited to knowledge of its statistical nature. In optical microscopy, degradation is mostly due to the optical blur generated by out-of-focus contributions of the object in the focal plane, and statistical noise. Assuming that noise follows a Poisson distribution (see above), it can be approximated by a Gaussian additive distribution for the purposes of simplification.

4.1.1 Continuous Formulation

In the case of continuous functions, the general relationship between input and output can be expressed as follows:

$$i(x, y, z) = N(\iiint o(x', y', z')\, h(x, y, z, x', y', z')\, dx'\, dy'\, dz') \tag{9}$$

where i is the acquired image, o the object acquired, h the PSF and N the noise. Assuming linearity and shift invariance in the image formation process (under certain conditions detailed above), the triple integral can be simplified to

$$i(x, y, z) = N(\iiint o(x', y', z')\, h(x - x', y - y', z - z')\, dx'\, dy'\, dz') \tag{10}$$

which is a standard 3D convolution integral that can be written as

$$i(x, y, z) = N(h(x, y, z) \otimes o(x, y, z)) \tag{11}$$

This equation describes, by means of a simple mathematical convolution operator, the way in which the acquired object o is degraded during the image formation process: first by being blurred with the instrument function h and secondly by noise degradation N. The convolution can be made in the Fourier domain, making it possible to use the mathematical property that a convolu-

tion in the spatial domain becomes a point-by-point multiplication in the Fourier domain

$$FFF\{h(x, y, z) \otimes o(x, y, z)\} = H(u, v, w)\, O(u, v, w) \tag{12}$$

where italic capital letters represent the Fourier transformations of the corresponding lower-case variables, and *FFF*{} the 3D Fourier transform operator. The Fourier transformation of the PSF, *H*, is called the optical transfer function (OTF).

4.1.2 Discrete Formulation

The discrete mathematical formulation of image formation is as follows:

$$i_m = N(\sum_l h_{l-m}\, o_l) = N(a_m) \tag{13}$$

This equation is referred to as the classical imaging equation, from which many iterative algorithms are derived, with a_m is introduced as the noise-free image. At any point *m* of coordinates $m=(x, y, z)$ in three dimensions, the acquired image for the point i_m is equal to the sum of contributions of the entire observed object *o*, weighted by the blurring function *h,* which is the microscope's PSF. The noise *N* associated with each point is independent of the image formation system.

4.1.3 Matrix Formulation

The classical imaging equation can also be expressed in matrix notation as follows:

$$\mathbf{i} = N(\mathbf{Ho}) \tag{14}$$

where **i, o** and **n** are one-dimensional matrices of size M, obtained by concatenation of their respective rows. **H** is a block circulant blurring matrix of size M×M. This notation can be used for the formulation of a class of well-known direct restoration algorithms.

In fluorescence microscopy, the convolution operator implies a linear and shift-invariant point spread function *h*. Although we showed earlier that this assumption does not hold true in all cases, this approximation will be assumed in the future because it simplifies computation of the solution.

An obvious way to reverse the convolution equation in the Fourier domain and to recover *O* given *I* and *H* would be to divide *I* by *H*. Unfortunately, in the presence of noise, the Fourier transformation of the PSF has a finite support and falls rapidly to zero because of the lack of resolution. This results in divisions by very small or zero values, making it impossible to divide *I* by *H*. In addition, noise is amplified during division, leading to the generation of ma-

jor artifacts. Such problems are called ill-posed or ill-conditioned problems, for which a unique solution of the object cannot be directly recovered without additional knowledge. All restoration methods, whether direct or iterative, aim to generate an estimate of the object o, denoted $\hat{o}$, generally by minimizing a function or criterion.

There are various families of deconvolution algorithms, each with its own advantages and limitations. These algorithms are powerful for the processing of images with known properties (e.g. for a given signal-to-noise range), but produce poor results in other cases. Unfortunately, the algorithms providing the best results with simulated data do not necessarily provide the best results with real acquired data. No best method has yet been identified in practical terms. We will now describe the algorithms most frequently used in published studies, providing results obtained with commercially available packages.

4.2 Wiener Filtering

This restoration technique involves estimating $\hat{\mathbf{o}}$, such that $\mathbf{H}\hat{\mathbf{o}}$ approximates $\mathbf{i}$ in the least squares sense [21]. In the absence of constraint, this is achieved by finding the value of $\hat{\mathbf{o}}$ that minimizes $\|\mathbf{i} - \mathbf{H}\hat{\mathbf{o}}\|^2 - \|\mathbf{n}\|^2$. Assuming that the norm of the noise is small, this is equivalent to minimizing the function

$$\mathbf{\Phi}(\hat{\mathbf{o}}) = \|\mathbf{i} - \mathbf{H}\hat{\mathbf{o}}\|^2 \tag{15}$$

with respect to $\hat{\mathbf{o}}$. The minimization of $\mathbf{\Phi}$ is straightforward and is achieved by setting to the zero vector the derivative of $\mathbf{\Phi}$ with respect to $\hat{\mathbf{o}}$:

$$\frac{\partial \mathbf{\Phi}(\hat{\mathbf{o}})}{\partial(\hat{\mathbf{o}})} = \mathbf{0} \tag{16}$$

The solution is the classical unconstrained inverted equation

$$\hat{\mathbf{o}} = \frac{\mathbf{H}^{\mathrm{T}}\mathbf{i}}{\mathbf{H}^{\mathrm{T}}\mathbf{H}} \tag{17}$$

which has been reported to be unsolvable in practice, mainly because of the non-existence of $\mathbf{H}$'s inverse.

A more flexible approach involves adding constraints by minimizing functions of the form $\|\mathbf{Q}\hat{\mathbf{o}}\|^2$, where $\mathbf{Q}$ is a linear operator on $\mathbf{o}$, subject to the constraint

$$\|\mathbf{i} - \mathbf{H}\hat{\mathbf{o}}\|^2 = \|\mathbf{n}\|^2 \tag{18}$$

with $\|\mathbf{i} - \mathbf{H}\hat{\mathbf{o}}\|^2 = \|\mathbf{i} - \mathbf{H}\hat{\mathbf{o}}\|^{\mathrm{T}}(\mathbf{i} - \mathbf{H}\hat{\mathbf{o}})$ and $\|\mathbf{n}\|^2 = \mathbf{n}^{\mathrm{T}}\mathbf{n}$. This minimization problem can be resolved with Lagrange multipliers, by minimizing the criterion function

$$\mathbf{\Phi}(\hat{\mathbf{o}}) = \|\mathbf{i} - \mathbf{H}\hat{\mathbf{o}}\|^2 + \alpha\|\mathbf{Q}\hat{\mathbf{o}}\|^2 \tag{19}$$

where α, the Lagrange multiplier, is a regularization parameter. Again, $\mathbf{\Phi}$ is minimized by its derivation with respect to $\hat{\mathbf{o}}$, and setting the result to the zero vector, which gives

$$\hat{\mathbf{o}} = \frac{\mathbf{H}^{\mathrm{T}}\mathbf{i}}{\mathbf{H}^{\mathrm{T}}\mathbf{H} + \eta\mathbf{Q}^{\mathrm{T}}\mathbf{Q}} \tag{20}$$

with η=1/α. This equation, also known as the Tikhonov-Miller solution, can also be derived from Bayesian theory. Approaches to the determination of an optimal value of η will be discussed in the statistical approach to restoration methods (see below).

From Eq. (20), the well known Wiener invert filter [21] can be obtained and expressed in the Fourier space as follows:

$$\hat{O}(u, v, w) = \frac{H^{*}(u, v, w)}{|H(u, v, w)|^2 + K} I(u, v, w) \tag{21}$$

where H^* is the complex conjugate of H and $|H(u,v,w)|2=H^*(u,v,w)\ H(u,v,w)$. In the Wiener invert filter, η is set to 1 and K is defined as P_n/P_o, where P_n and P_o represent the power spectra of the noise and the object, respectively. In practice, P_n and P_o are unknown and K is a constant, in the range 0.001 to 0.1 [23].

This direct restoration method is straightforward and requires only modest computation power. However, it is sensitive to PSF, and small errors in the estimation of PSF may result in major artifacts. This algorithm is often used in practice as a first approximation in iterative procedures. The iterative approaches presented below, although more demanding in terms of computation power, have been shown to be less sensitive to errors and to generate more stable estimates of o.

4.3 Nearest Neighbors

This technique does not use all the out-of-focus information, but does use that contained in the two adjacent planes above and below the plane of interest: the nearest neighbors [22, 23]. This method is based on the principle that these two planes make the most substantial contribution to the central plane. Thus, only the contributions of these two planes are taken into account for correction of the central plane, with other out-of-focus contributions ignored. The general convolution relationship between the acquired image i and the object o can be simplified and rewritten for the k-th plane as a function of the information constrained in the adjacent (k-1)-th and (k+1)-th planes:

$$i_{\mathrm{k}} = o_{\mathrm{k}} \otimes h_0 + o_{\mathrm{k-1}} \otimes h_{-1} + o_{\mathrm{k+1}} \otimes h_{+1} \tag{22}$$

where indices denote the plane number for i and o, and the plane distance relative to the focal plane h_0 for h. Assuming that $o_{\mathrm{k}\pm1} \approx i_{\mathrm{k}\pm1}$ and ignoring the

difference between h_{-1} and h_{+1}, such that $h_{+1}=h_{-1}=h_1$, the object in the k-th plane can be approximated by

$$\hat{o}_k = [i_k - c(i_{k-1} + i_{k+1}) \otimes h_{+1}] \otimes h_0^{-1} \tag{23}$$

where c is an adjustable constant, between 0.45 and 0.5.

This technique has the advantage of being very rapid, because it requires only two two-dimensional convolutions per plane. However, the efficiency of this method is greatly limited by the approximation in which light contributions from planes other than the two nearest neighbors are not taken into account. Recent advances in computer hardware have made it possible to use more sophisticated algorithms applied to the 3D light distribution in its entirety, in a reasonable time range.

4.4 Constrained Iterative Algorithms

The goal of this algorithm family is to develop a solution through an iterative process under positive constrains, which, when convolved by the known blurring function, will provide the acquired image. This type of algorithm is derived from the classical imaging equation (Eq. 13), assuming that the noise n_m is negligible [23, 24]. The classical imaging equation is then reduced to

$$i_m = \sum_l h_{l-m} o_l = a_m \tag{24}$$

4.4.1 The Jansson Van-Cittert Algorithm

By subtracting a_m from both sides of the equation, raising both sides to the power of p and by adding o_m to both sides, we get

$$o_m = o_m + (i_m - \sum_l h_{l-m} o_l)^p \tag{25}$$

Putting this equation into its iterative form results in

$$\hat{o}_m^{k+1} = K[\hat{o}_m^k + w_m(i_m - \sum_l h_{l-m} \hat{o}_l^k)^p] \tag{26}$$

where K is a normalizing energy constant and w_m a weighting function. w_m and p can be adjusted to enforce the non-negativity of the solution for all iterations.

If $w_m=1-(i_k-A)2/A2$, with A equal to the maximum value of $i/2$, and $p=1$, then this equation corresponds to the Jansson Van-Citter formula. In this case, the non-negativity constraint is forced by setting $\hat{o}_m^{k+1} = 0$ if $\hat{o}_m^{k+1} < 0$. The initialization of $\hat{o}_m^{k+1}$ is given by $\hat{o}_m^0 = i_m$. This iterative algorithm calculates a new estimate of the object as a function of the previous estimate, the acquired image and the known PSF, using an additive correction factor.

4.4.2 The Gold Algorithm

Based on exactly the same principle, the multiplicative formulation of constrained iterative algorithms can be obtained first by dividing both sides of the classical noise-free imaging equation (Eq. 24) by a_m and by raising both sides to the power of p, then by multiplying both sides by o_m, and finally by rewriting the result in its iterative form:

$$\hat{o}_m^{k+1} = K\hat{o}_m^k \frac{i_m}{\sum_l h_{l-m}\hat{o}_l^k}^p \tag{27}$$

For p=1, this equation corresponds to the Gold formula. Non-negativity condition is always ensured for optical objects ($\hat{o}^0$ positive), and, as for the additive case, the initialization of $\hat{o}_m^{k+1}$ is given by $\hat{o}_m^0 = i_m$.

First, both additive and multiplicative formulations of the iterative constraint algorithms suffer from slow convergence of the solution (although the multiplicative formulation gives a more rapid convergence). Second, they give poor results if the original data are noisy, particularly because noise not taken into account may create resonance and generate constructive artifactual signals. These two problems can be partially solved by adding a step to control noise throughout the iterative process. Practically, the first guess is filtered with a Gaussian or Wiener filter and each of four to seven iterations, depending on the signal-to-noise ratio, is smoothed using a Gaussian filter of variable width. The wider the filtering, the better the noise control is likely to be, but the less sharp the result. The user must take the decision concerning the acceptability of such a compromise, depending on the subsequent analyses to be performed.

4.5 Statistical Algorithms

Statistical information about the noise present in the images is incorporated into this family of algorithms. Like constrained iterative algorithms, they are recursive but they are built to determine an acquired object, blurred by the optical system and corrupted by random noise. The quantum nature of light results in the distortion of the image by noise. Noise is caused by photon counting (Poisson), detector readout (Gaussian) and analog-to-digital conversion (uniform). Statistical algorithms are based on finding an estimate $\hat{o}$ of the object o according to the nature of the noise. They can be classified into two subfamilies, according to the use or non-use of a regularization function. Both subfamilies are derived from the maximum likelihood estimation (MLE) algorithm, written to minimize the Poisson noise in the image. In the case of algorithms using a regularization function, noise is sometimes assumed to follow a Gaussian distribution for reasons of mathematical simplification [26]. The origins of these algorithms are described by the Bayesian theory of the

signal [27], in which the object o is statistically represented by its prior density function $P(o)$ and introducing the conditional density function $P(i|o)$ to model the blurred object i given the object o. Bayesian rules give the posterior density $P(o|i)$ of object o for a given blurred object i as

$$P(o|i) = \frac{P(i|o)\,P(o)}{P(i)} \tag{28}$$

This method is based on maximizing the probability function and estimating the maximum value of the object o, a posteriori, given the known image i. As $P(i)$ is independent of o, this is equivalent to minimizing the negative logarithm of the numerator of $P(o|i)$, providing a function Φ of the form

$$\Phi(o) = -\ln P(i|o) - \gamma \ln P(o) = L(i, o) + \gamma\Omega(o) \tag{29}$$

$L(i,o)$ is the likelihood estimation function of the object, $\Omega(o)$ is a smoothing function and γ is a regularization parameter. The likelihood function depends on the noise model: Poisson or Gaussian. It measures how well the estimate fits the acquired data. The smoothing function, also called the penalty function, attenuates the artifacts generated by noise amplification. γ is a balance between the need to fit the data and the need to stabilize the estimate. Setting Ω to 0 yields the maximum likelihood algorithm.

4.5.1 Maximum Likelihood Estimation Algorithms

Maximum likelihood estimation is a mathematical strategy for computing an object corrupted by noise. In the case of microscopy, the noise is linked to the statistical nature of quantum photon emission and follows a Poisson distribution, described by the following equation:

$$P_{Pois}(x|\mu) = \frac{\mu^x e^{-\mu}}{x!} \tag{30}$$

This equation gives the probability of obtaining a noisy datum x, given its noise-free averaged value μ. As each pixel constituting an acquired image is statistically independent from the others, it follows a Poisson distribution, and the whole image probability is then equal to the product of individual probabilities, giving

$$P_{Pois}(i|o) = P_{Pois}(i|a) = \prod_m P_{Pois}(i_m|a_m) = \prod_m \frac{a_m^{i_m} e^{-a_m}}{i_m!} \tag{31}$$

The log-likelihood function of the noise is obtained by taking the logarithm of the image probability:

$$L_{Pois}(i, o) = -\ln P_{Pois}(i|o) = \sum_m (a_m - i_m \ln(a_m) + \ln(i_m!)) \tag{32}$$

In the case of a Poisson noise process, the smoothing function Ω is set to zero and Φ is minimized by minimizing the likelihood function. The solution is straightforward and is computed by setting the derivative of the likelihood to zero, with respect to o_m, for fixed i, resulting in

$$\frac{\partial L(i,o)}{\partial o_m} = \sum_l \left(h_{m-l} \frac{i_l}{a_l} \right) - 1 = 0 \tag{33}$$

By adding 1 to both sides of the equation, raising both sides to the power p, adding o_m to both sides and writing the result in the iterative form, we obtain

$$\hat{o}_m^{k+1} = K \hat{o}_m^k \sum_l \left(h_{m-l} \frac{i_l}{\sum_n h_{l-n} \hat{o}_n^k} \right)^p \tag{34}$$

where $\hat{o}_m^0 = i_m$ and K is a constant ensuring energy conservation. The MLE ensures the non-negativity of the solution for optical objects ($\hat{o}^0$ positive), and no additional constraints are required. The Richardson and Lucy algorithm is obtained by assuming p=1.

These algorithms are the most frequently used to restore noisy data, mainly because they are easy and straightforward to use but also because the constraints they impose on noise statistics make them robust to noise. Nevertheless, they display very slow convergence (between 50 and 1000 iterations) and require more computing time per iteration than classical constrained iterative algorithms (two convolutions per iteration). Although, the use of a value greater than 1 for p makes the convergence faster, it also results in less stable convergence. With the absence of a regularization function Ω, the noise present during acquisition can lead to artifacts. In this case, the first estimate strongly influences the result and the first estimate should be smoothed with a Gaussian or Wiener filter, as for the iterative constrains algorithms, to prevent noise amplification during iteration. Another way of preventing such problems is to terminate the algorithm before convergence.

In the case of microscopy, Poisson noise is signal-dependent. It is therefore important to take into account the additive background b, mainly produced by the dark-field current of the detector, giving

$$\hat{o}_m^{k+1} = K \hat{o}_m^k \sum_l \left(h_{m-l} \frac{i_l}{\sum_n h_{l-n} \hat{o}_n^k + b_l} \right)^p \tag{35}$$

where b is the additive background image, which can be measured by taking dark images.

4.5.2 Algorithms Using a Regularization Function

One way to avoid convergence problems and to increase the speed of MLE algorithms is to add a regularization function. The goal is to minimize Eq. (29) in order to find a stable estimation of $\hat{o}$, and various strategies have been developed to achieve this. The noise is assumed to follow a Gaussian distribution rather than a Poisson distribution, to ensure that the algorithm can be solved mathematically and is compatible with Bayesian theory. This approximation has been validated for data with a high signal-to-noise ratio, and should in theory be restricted to such data. However, in practice, this family of algorithms work pretty well on noisy data, even though they are not optimized for such data. The corresponding Gaussian distribution of standard deviation σ is described by the following equation:

$$P_{Gaus}(x|\mu) = e^{-\frac{(i-\mu)^2}{2\sigma^2}} \tag{36}$$

Using an analog method for Poisson noise, the Gaussian log-likelihood function of the noise is given by

$$L_{Gaus}(i,o) = -\ln P_{Gaus}(i|o) = \frac{1}{2\sigma^2}\sum_m (i_m - a_m)^2 = \frac{1}{2\sigma^2}\|i - h \otimes o\|^2 \tag{37}$$

As for Poisson noise, the penalty function could be omitted and a Gaussian MLE algorithm could be written by simply minimizing the Gaussian likelihood. However, this is not really useful in practice because the MLE algorithm is better adapted. For a Gaussian noise distribution, the regularization function used may be the Gaussian density function $P(o)$, giving the quadratic penalty function

$$P(o) = \Omega(o) = \|o\|^2 \tag{38}$$

The Φ function to be minimized then becomes a quadratic function:

$$\Phi(\hat{o}) = \frac{1}{2\sigma^2}\|i - h \otimes \hat{o}\|^2 + \gamma\|\hat{o}\|^2 \tag{39}$$

equivalent to that obtained with matrix notation (Eq. 19). Its direct minimization yields the well known Wiener filter or unconstrained Tikhonov-Miller solution developed above, which has been shown to have practical limitations. Equation (19) can be minimized iteratively, adding the physical non-negativity constraint to the solution $\hat{o}_m^i \geq 0$.

The conjugate gradient method has been adapted for iterative minimization of the quadratic function Φ [28–30]. It is one of the most efficient iterative methods for identifying the extremes of a function of n variables, assuming that the gradient of the function can be calculated. This non-stationary method

moves forward by generating vector sequences of iterates (i.e., successive approximations to the solution), residuals corresponding to the iterates, and search directions used in updating the iterates and residuals. The iterates $\hat{o}^k$ are updated in each iteration by a multiple α^k of the search direction vector $\mathbf{p}^k$

$$\hat{\mathbf{o}}^k = \hat{\mathbf{o}}^{k-1} + \alpha^k \mathbf{p}^k \tag{40}$$

Correspondingly, the residuals $\mathbf{r}^k = \mathbf{b} - \mathbf{A}\hat{\mathbf{o}}^k$ are updated as

$$\mathbf{r}^k = \mathbf{r}^{k-1} - \alpha^k \mathbf{A}\mathbf{p}^k \tag{41}$$

with, in our case, $\mathbf{A} = (\mathbf{H}^T\mathbf{H} + \eta)$ and $\mathbf{b} = \mathbf{H}^T\mathbf{i}$.

In the absence of nonlinear constraints, an optimal value of α^k can be calculated analytically and is given by

$$\alpha^k = -\frac{\mathbf{p}^{k^T}\mathbf{r}^{k-1}}{\mathbf{p}^{k^T}\mathbf{A}\mathbf{p}^k} \tag{42}$$

However, if there is a nonlinear non-negativity constraint on the estimate, this value is no longer valid, resulting in very slow convergence of the solution [26], rendering this algorithm much less useful than the standard MLE algorithm.

The iterative constrained Tikhonov Miller (ICTM) algorithm has been used to take into account the non-negativity constraint of $\hat{\mathbf{o}}$ and to generate a better estimate of α^k [28, 30]. It involves a one-dimensional iterative minimization algorithm, like the Brent or Gold algorithm [32], and aims to minimize $\Phi(C(\hat{\mathbf{o}}^{k-1} + \alpha^k \mathbf{p}^k))$, with $C(.)$ the non-negativity constraint operator, which sets the negative values of its argument to zero. The final algorithm consists of a main iterative loop for calculating the conjugate direction, and a subiterative loop that finds the best estimate of α^k and calculates the new estimate of $\hat{\mathbf{o}}$. This increases the computation time because, for each estimate of α^k in the subiteration loop, a complete evaluation of $\Phi(C(\hat{\mathbf{o}}^{k-1} + \alpha^k \mathbf{p}^k))$ must be carried out, and this is the most cumbersome part of the algorithm. Moreover, this algorithm lacks mathematical rigor because the gradient of $\Phi(\hat{\mathbf{o}}^k)$ is used to minimize the gradient of $\Phi(C(\hat{\mathbf{o}}^k))$.

An alternative approach to increasing computation speed by increasing the speed of determination of α^k was investigated by Veerver et al. [26]. This method involves determining the value of α^k that sets the gradient to zero in the search direction, and then solving

$$\mathbf{p}^{k^T} \nabla \Phi (C(\mathbf{o}^{k-1} + \alpha^k \mathbf{p}^k)) = 0 \tag{43}$$

which gives

$$\alpha^k = -\frac{\mathbf{p}^{k^T}\mathbf{T}(\mathbf{o})\mathbf{r}^{k-1}}{\mathbf{p}^{k^T}\mathbf{T}(\mathbf{o})\mathbf{A}\mathbf{T}(\mathbf{o})\mathbf{p}^k} \tag{44}$$

where $\mathbf{T}(\mathbf{o})$ is a diagonal matrix defined by

$$\mathbf{T}(\mathbf{o})_{ij} = \begin{cases} 1 & \text{if i = j and } \mathbf{o}_i > 0 \\ 0 & \text{otherwise} \end{cases} \tag{45}$$

This direct solution is more rigorous mathematically than that for ICTM algorithms. It is also much faster because no subiteration loop is required.

In contrast to the situation for Poisson noise, the variance of Gaussian noise is, by definition, independent from the signal. Therefore, the background image can be omitted in each of the equations without changing the noise properties, simply by subtracting it from the data in the preprocessing step.

4.6 Blind Deconvolution Algorithms

In the blind deconvolution approach, the object and the PSF are assumed to be unknown and are estimated throughout the iterative process [34–36]. The algorithm uses the standard MLE algorithm described above, together with a PSF estimation for each iteration. The object is computed, using the MLE estimation, as follows:

$$\hat{o}_m^{k+1} = K\hat{o}_m^k \sum_l \left(\hat{h}_{m-l}^k \frac{i_l}{\sum_n \hat{h}_{l-n}^k \hat{o}_n^k} \right) \tag{46}$$

Using exactly the same mathematical reasoning, PSF is estimated by maximizing the log likelihood function with respect to h_m, which gives

$$\hat{h}_m^{k+1} = \frac{\hat{h}_m^k}{N} \sum_l \left(\hat{o}_{m-l}^k \frac{i_l}{\sum_n \hat{o}_{l-n}^k \hat{h}_n^k} \right) \tag{47}$$

where N is a normalization constant relating to the unit volume. Object initialization is performed using $\hat{o}_m^0 = i_m$, and the first estimate of PSF $\hat{h}_m^0$ is calculated theoretically from the microscope parameters. Like the standard MLE algorithm, the non-negativity of both $\hat{o}_m^{k+1}$ and $\hat{h}_m^{k+1}$ are ensured. Additive constraints on the PSF can be imposed, such as both axial circular symmetry and frequency band-limited properties, which can be used to check the PSF. These constraints are added at each iteration, after the computation of the new estimate $\hat{h}_m^{k+1}$. Axial circular symmetry is achieved plan-by-plan, by averaging the values equidistant from the middle of the image, where the optical axis should exist. The frequency band-limited constraint is imposed by setting all the values of the Fourier transformation of $\hat{h}_m^{k+1}$ such that they lie above the known frequency cutoff. In three-dimensional microscopy, lateral and axial resolutions are different, and the PSF can be filtered separately in each of the two directions.

Like the standard MLE algorithm, this family of algorithms suffers from slow convergence and the estimates obtained are not stable. First guess smoothing and regularization functions can be applied to increase the convergence speed, as described above.

This family of algorithms may be used for non-specialist purposes, or when the point spread function is unknown. It takes twice as long as the MLE per iteration but can produce better results than the use of a theoretical PSF, especially if unpredicted aberrations are present. Nevertheless, we always recommend that the acquisition system be characterized and calibrated (see above on PSF determination), to minimize aberrations before acquisition, because the better the acquisition, the better the result of the deconvolution is likely to be, regardless of the algorithm used.

4.7 Results

In this section, we have tested various algorithms, using commercial packages. It is very difficult to compare algorithms because the results obtained depend heavily on image quality (signal-to-noise ratio, sampling), and on algorithm parameters and optimizations. Which criteria should be used to compare images? Image quality, which is very difficult to assess for biological objects, is not the only parameter in practice. Speed is also of great importance because in some cases it is preferable to use a given algorithm simply because it is ten times faster than another "more rigorous" algorithm. It is also difficult to compare the speeds of the different algorithms, because this speed depends on the way the algorithms are implemented, optimizations and stopping criteria, and a given algorithm may give different results and execution speeds when implemented via different software packages. We will therefore avoid detailing speeds here. We will instead show results obtained with different families of algorithms, and will try to assess the advantages and limitations of each, and their robustness to noise. However, we will also present some typical artifacts that may arise.

Figure 15 displays deconvolution results obtained with five algorithms, from the various families described above. We display three optical sections and one axial cross section for each. We have also measured intensity profiles and added arrows to illustrate in more detail the efficiency of these algorithms and to draw attention to specific parts of the images.

Artifacts may arise for various reasons. For example, the algorithm chosen may not be efficient enough to process the data in the presence of too much noise. Alternatively, the choice of algorithm parameters may be poor, PSF estimation may be inaccurate or the data may have been subsampled. There is little point in trying to illustrate all these artifacts on real samples because their appearance depends on object morphology and noise level. We have discussed fully the key role played by noise and incorrect PSF estimation in the generation of artifacts. To illustrate these points, Fig. 16 shows the result of deconvolution on images of fluorescent microbeads internalized by a cell before fixa-

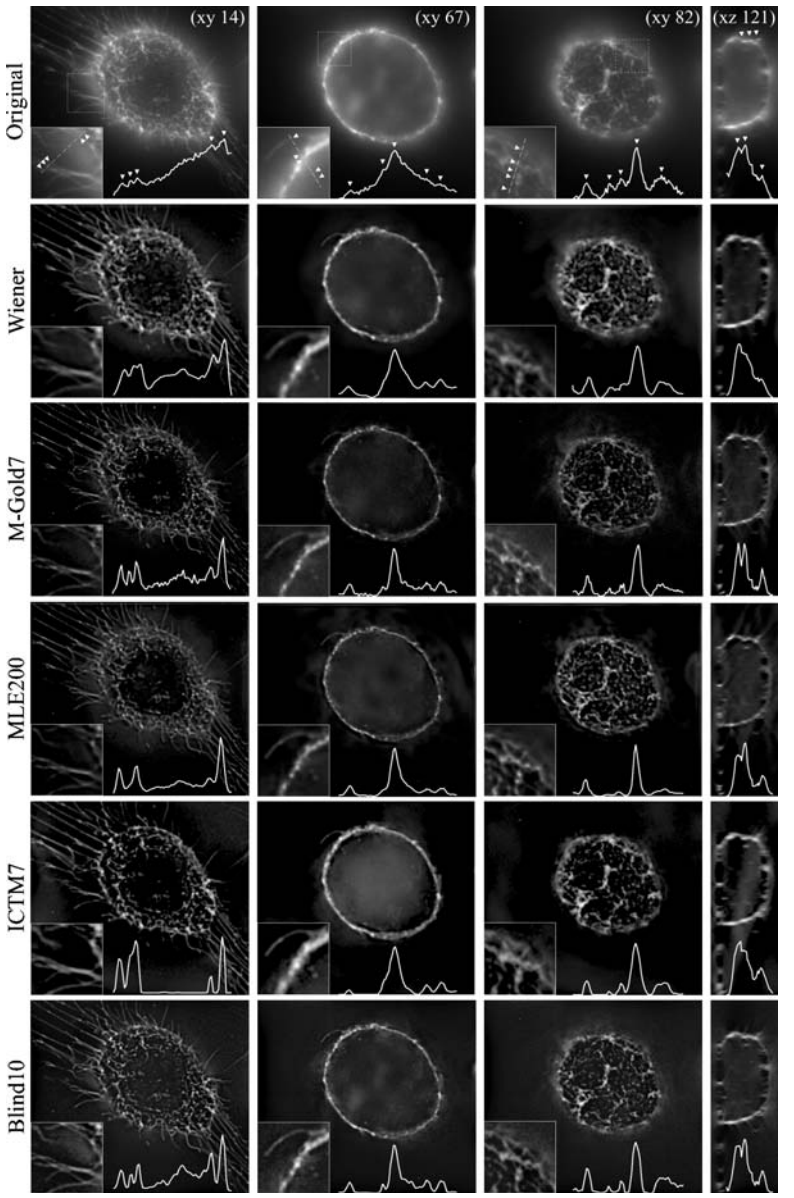


Fig. 15 Image deconvolution of a fixed MDCK cell labeled with concanavalin A – FITC. Images were acquired with a 3D microscope, as described in the text. From *left to right*: optical sections of the bottom (on the coverslip), middle and top of the cell, and axial cross-section of the same cell. From *top to bottom*: Original data, and data restored using: the Wiener invert filter, the modified Gold iterative constrained algorithm with 7 iterations (Metamorph, Universal Imaging), the MLE algorithm with 200 iterations, the quick ICTM algorithm with 7 iterations (Huygens, SVI) and finally the blind MLE deconvolution algorithm with 10 iterations (Autodeblur, Autoquant)

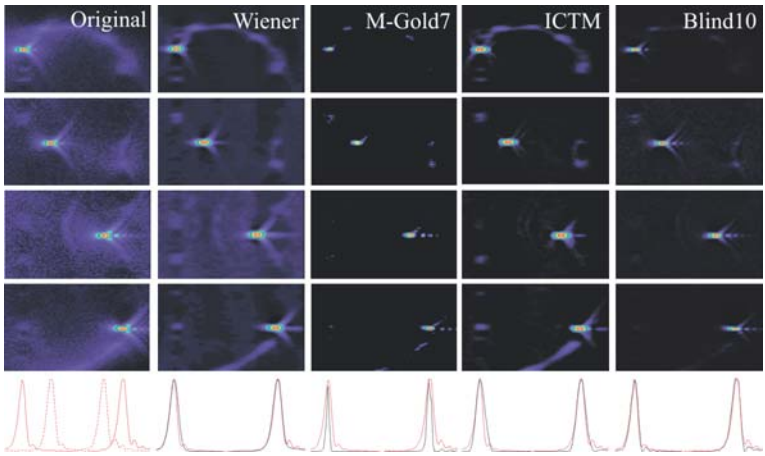


Fig. 16 Image deconvolution of fluorescent microbeads located at various distances from the coverslip. Rhodamine microbeads, 100 nm in diameter, were internalized within an MDCK cell before fixation. From *left to right*, axial cross-sections on original data, and data restored using: the Wiener invert filter, the modified Gold iterative constrained algorithm with 7 iterations (Metamorph, Universal Imaging), the quick ICTM statistical algorithm with 7 iterations (Huygens, SVI) and finally the blind MLE deconvolution algorithm with 10 iterations (Autodeblur, Autoquant). For all the algorithms, except for the blind approach, the PSF was modeled using a microbead located on the coverslip. The *bottom line* displays intensity profiles measured along bead cross-sections

tion and located at various distances from the coverslip. This indicates the efficiency of the deconvolution algorithms in terms of axial resolution, and illustrates the artifacts that arise when a single PSF model is used, given that axial shift invariance is assumed in all the algorithms. As expected, efficiencies decreased and artifacts increased with increasing distance of the bead from the coverslip (Fig. 16).

Finally, to illustrate robustness of the algorithms to noise, deconvolution results obtained with four different algorithms applied to image stacks acquired with different exposure times are shown in Figs. 17 and 18. As expected, the signal-to-noise ratio of the images considerably affects the quality of the results for most of the algorithms. Moreover, some typical artifacts become visible on images with low signal-to-noise ratios, with the emergence of a background texture with an amplitude in the order of magnitude of the signal. As predicted by theory, statistical algorithms, which require heavier computation, are more robust to noise. Nevertheless, the addition of a noise control parameter to constrained iterative algorithms increases their robustness to noise, and good results can be obtained in a very short time.

It is not easy to compare algorithms. However, although it is impossible to draw definitive conclusions concerning the efficiencies of algorithms from only a few data sets, a few general comments can be made. First, for all the algo-

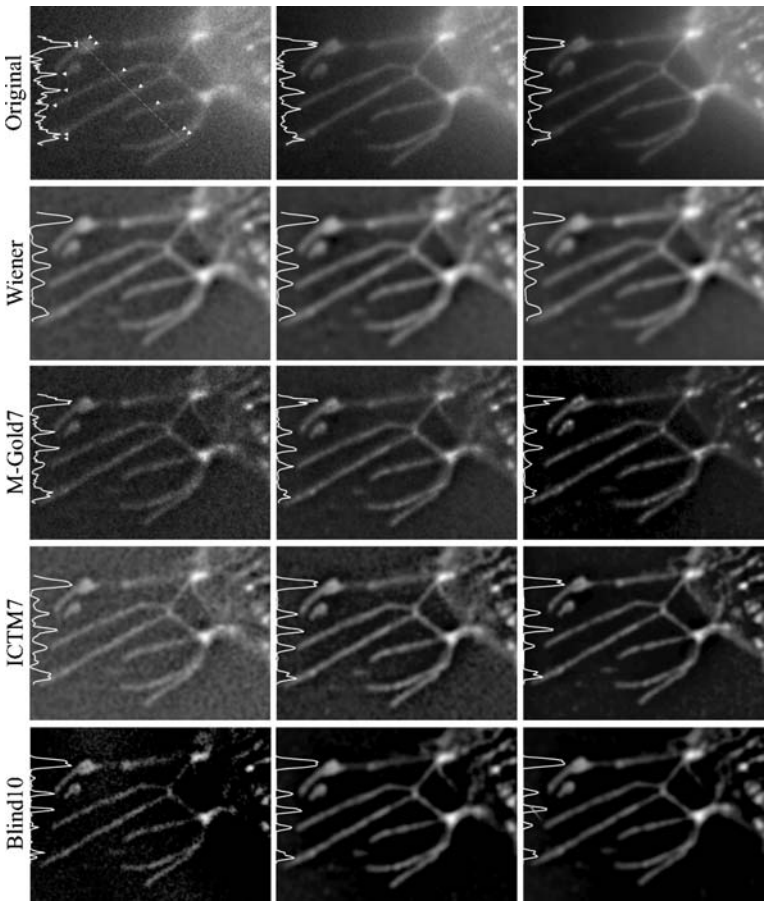


Fig. 17 Deconvolution of image stacks acquired with different exposure times, using four different algorithms. Original images of a single plane acquired for, from *left to right*, 50 ms, 100 ms and 200 ms per plane. From *top to bottom*, images restored using: the Wiener invert filter, the modified Gold iterative constrained algorithm with 7 iterations (Metamorph, Universal Imaging), the ICTM algorithm with 7 iterations (Huygens, SVI) and finally the blind MLE deconvolution algorithm with 20 iterations (Autodeblur, Autoquant)

rithms and processed images presented, deconvolution clearly improved image quality. It enhanced contrast and decreased blurring, mainly in the axial direction. Such an improvement should always occur, provided that the PSF estimate used is correct. Even planes distant from the coverslip, for which we know the PSF is not perfectly valid, do not present strong artifacts. Second, the Wiener invert filter is clearly much less efficient than other iterative methods. However, the ability of this algorithm to reduce noise nonetheless renders it useful for first guesses for iterative algorithms. All noisy data were improved, even with classical iterative algorithms such as MGold, which includes a noise control

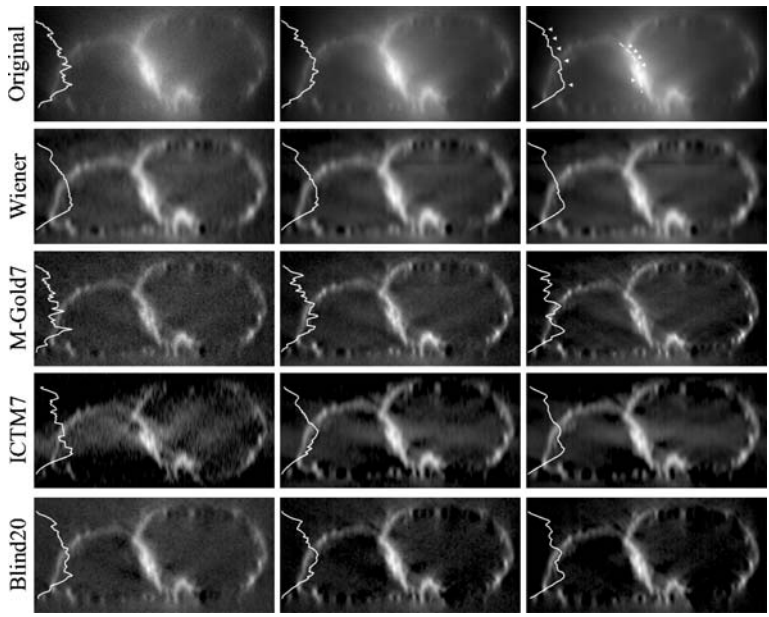


Fig. 18 Deconvolution of image stacks acquired with different exposure times, using four different algorithms. Axial Cross sections of original image stacks acquired for, from *left to right*, 50 ms, 100 ms and 200 ms per plane. From *top to bottom*, cross sections of restored stacks using: the Wiener invert filter, the modified Gold iterative constrained algorithm with 7 iterations (Metamorph, Universal Imaging), the ICTM algorithm with 7 iterations (Huygens, SVI) and finally the blind MLE deconvolution algorithm with 20 iterations (Autodeblur, Autoquant)

parameter. Moreover, as if slightly less robust to noise than statistical algorithms, MGold greatly improves the image quality in the axial direction. MLE algorithms give good results, but are not very useful in the absence of a regularization parameter, due to very slow convergence. Even after 200 iterations the MLE gave unsatisfactory results, whereas only seven iterations of the QICTM gave better results. All the results obtained with the blind deconvolution algorithm were good, although some artifacts began to appear with noisy data. Nevertheless, microscope characterization must still be carried out correctly as this step is crucial for collecting "perfect" data and hence obtaining good final results. Finally, even when the estimate of PSF used is accurate and acquisition conditions are perfect, artifacts may arise in the restored data. As a general rule, all the structures present in the restored data should be visible in the original data, even if they are unclear. Thus, if a new structure appears, it is certainly an artifact.

4.8 Deconvolution of Large Data Sets by Distributed Computing

The deconvolution technique has become so common that it has now essentially been integrated into microscopy and has become a part of the acquisition process itself. This so-called "deconvolution microscopy" is based on recent increases in the computational power available. In parallel, fluorescence microscopy has made a major leap forward in terms of multidimensional data collection, greatly increasing the amount of data to be processed. It is now not unusual to obtain gigabytes of data in thousands of files in the space of an hour, and the difficulty lies less in processing several "huge" stacks than in processing the growing flux of stacks generated by this ultra-fast microscopy. Living-cell time-lapse microscopy is now faced with the same problems as nuclear physics studies involving huge particle accelerators: how to almost instantaneously capture and process huge amounts of collected data. Of course, the order of magnitude of the problem differs between instruments: whereas light microscopy deals with terabytes (10^6 megabytes) of data, particle accelerators deal with petabytes (10^9 megabytes). However, in the near future, most cell biology research laboratories will be equipped with a wide variety of rapid microscopes, whereas only a few particle accelerators are available worldwide, so the challenge is probably equally great in both disciplines.

The deconvolution of such large amounts of data is critical and may become a brake on progress. The computer responsible for image processing is generally connected to a local area network (LAN). Frequently, other workstations in this LAN are not in full-time use: at night, in particular, most computers are idle. This "down time" could be used more effectively. Moreover, with the developments of low-cost PC farms, it is becoming easier and cheaper to obtain high computational capacities. We have developed a method for distributed computations to take advantage of these possibilities [37].

In our approach, we do not parallelize the deconvolution algorithm itself to increase the deconvolution speed of each stack. Instead, we distribute independent tasks, at the image stack level. Thus, our parallel calculations are "coarse-grained", the computational element being a single stack. Basically, processing jobs are queued into a pipeline. At its endpoint, job consumer threads (A-N) concurrently invoke the deconvolution of queued stacks on remote servers (remote computers A-N) (Fig. 19).

Our computational model is derived from some early attempts to distribute, over a network of computers, numerical calculations requiring the transfer of non-negligible amounts of data between client and servers [38, 39]. However, recent advances in network technologies have increased reliability and performances. In particular, web-based approaches to distributed computing are even beginning to emerge [40].

However, network performance still presents an obstacle because each stack must be sent from the client application to a server before remote processing; transport time is therefore non-negligible. The extent to which speed is in-

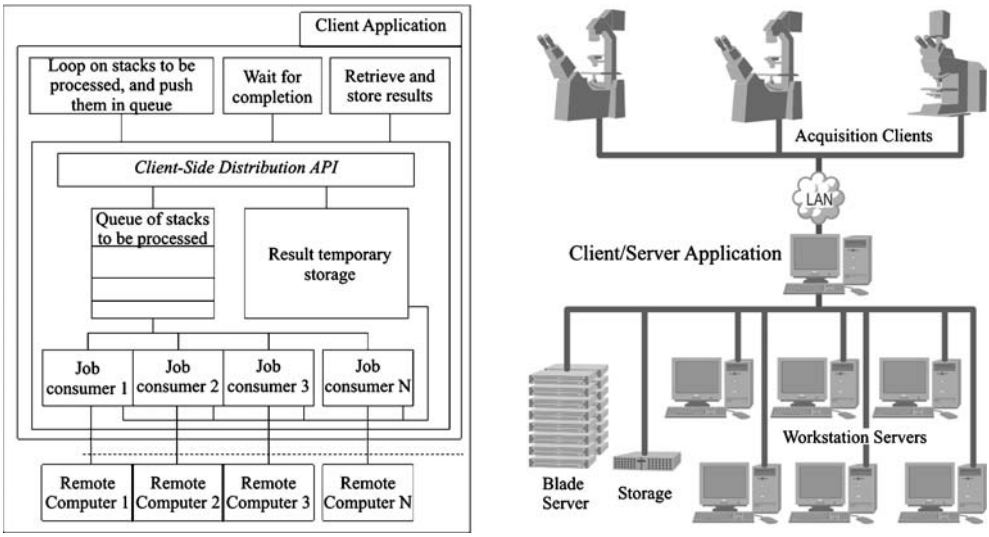


Fig. 19 Computational model of the distribution process (*left*) and distributed deconvolution microscopy architecture (*right*)

creased clearly depends on the time T_P to process a single stack and the time T_N to send this stack back and forth between the client and a server. Figure 20 shows a simulation of the "speedup factor" as a function of the ratio T_N/T_P.

The major issue when assessing the potential value of distributed computation is this T_N/T_P ratio. To visualize this mentally, just consider T_P as the "parallelizable" part of the computation, whereas T_N is "uncompressible" because however many computers there are, there is only one physical network link available. When T_N/T_P is low, this means that the algorithm is complex with respect to the amount of data to be transferred and that the process may be accelerated effectively by using multiple computers. When T_N/T_P is high, the problem is less a matter of algorithmic complexity than of data transfer performance; in such cases, shared memory or massively parallel computers may be more appropriate.

Although it is sometimes argued that network technology (and speed) will improve in the future, it should be noted that processors are likely to undergo similar improvements. Thus, if T_N/T_P is low for a given computational problem in 2004, it will probably not be significantly better for the same problem in 2014 – just imagine the "race towards performance" in which both processor and network manufacturers are engaged… The outcome clearly depends on the "winner of the race".

To illustrate the method, we processed 120 stacks of 1.4 Mb each (160 Mb of raw data) off-line using the modified Gold algorithm with 10 iterations and the MLE algorithm with 50 iterations. One to eight similar remote computers (PIV-2 GHz) linked on a 1 Gb/s LAN were used to assess the increase in speed.

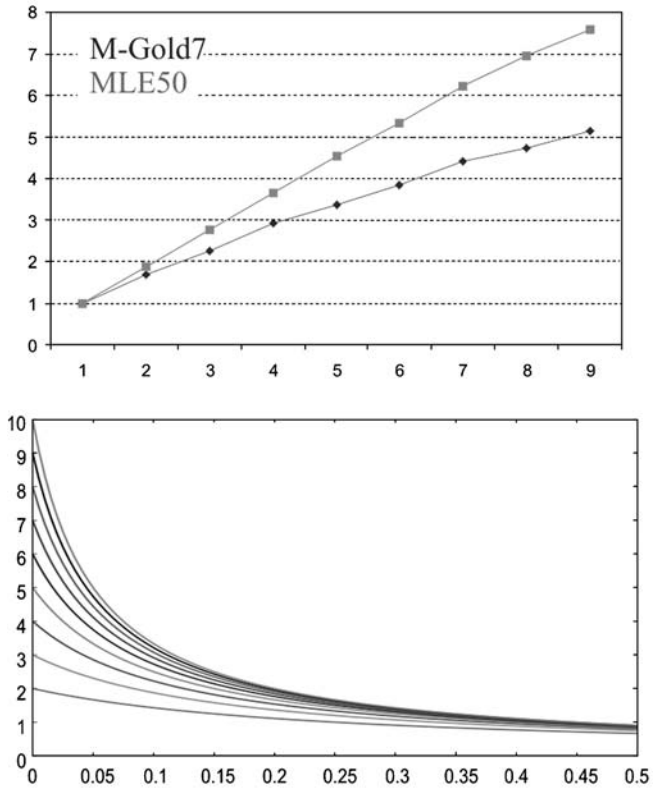


Fig. 20 *Top*: Measured speedup factor using one to eight remote similar computers in parallel, for 2 different algorithms: MGold with 7 iterations and MLE with 50 iterations. Computation was achieved using 120 stacks of 1.4 Mb each, and computers connected via a 1 Gb/s LAN. *Bottom*: Simulation of the speedup factor as a function of the ratio T_N/T_P for two to ten remote servers

A stack of 500×400 pixels × 9 planes was acquired every second over a 2-min period with only 50 ms exposure time per plane. The acquisition of each stack in z-streaming mode took less than 500 ms, allowing more than 500 ms for fluorescence relaxation. Figure 20 shows the measured speedup factor as a function of the number of remote servers. The speedup factor is defined as the time required to restore all the stacks in sequential mode, divided by the time measured to restore all the stacks in parallel mode. The only difference between the algorithms is the computing time per stack.

In the fastest case, it took only 1 min 45 s to process all the data using eight remote servers plus the client computer, rather than about 9 min in sequential mode, giving a speedup factor greater than 5.

In conclusion, for this typical high-speed time-lapse microscopy experiment, the total acquisition time was greater than the total processing time. This technology could therefore be seen as leading the way towards online process-

ing. We imagine that in the very near future, it will be possible to visualize the deconvolution during the acquisition. Achieving on-screen real-time navigation within the 3D model of a live cell, the dream may finally become a reality.

Acknowledgments Vincent Fraisier, Fabrice Cordelières, Henri Magnin, Jean Salamero, Zvi Kam, Nate O'Connor, Jeff Stuckey, Tim Holmes, David Biggs, Jan De Mey, Pascal and Joy Jaillon.

References

1. Agard DA (1984) Ann Rev Biophys Bioeng 13:191
2. Chalfie M (1994) Science 263:802
3. Cubitt AB (1995) Trends Biochem Sci 20:448
4. Ormo M (1996) Science 273:1392
5. Pawley JB (1995) Handbook of biological confocal microscopy. Plenum, New York
6. Wilson T (1990) Confocal microscopy. Academic Press, London
7. Kino GS (1995) Intermediate optics in Nipkoff disk microscopes In: Pawley JB (1995) Handbook of biological confocal microscopy. Plenum, New York, p 155
8. Wallace W, Schaefer LH, Swedlow JR (2001) BioTechniques 31:1075
9. McNally JG, Karpova T, Cooper J, Conchello JA (1999) Methods 19:373
10. Sheppard C, Gan X, Gu M, Roy M (1995) In: Pawley JB (1995) Handbook of biological confocal microscopy. Plenum, New York, p 363
11. Born M, Wolf E (1964) Principle of optics. Pergamon, New York
12. Inoué S, Spring KR (1997) Video-microscopy. Plenum Press, New York
13. Erhardt A, Zinser D, Komitowski D, Bille J (1995) Appl Opt 24:194
14. Hiraoka Y, Sedat JW, Agard D (1990) Biophys J 57:325
15. Gibson SF, Lanni F (1991) J Op Soc Am 8:1601
16. Scalettar BA, Swedlow JR, Sedat J, Agard DA (1996) J Microsc 182:50
17. Kam Z, Hanser B, Gustafsson MGL, Agard DA, Sedat J (2001) Proc Natl Acad Sci USA 98:3790
18. Hanser B, Gustafsson M, Agard D, Sedat J (2003) Opt Lett 28:801
19. Visser TD, Groen FCA, Brakenhoff GJ (1991) J Microsc 163:189
20. Carrington W (1990) Proceedings of SPIE The International Society for Optical Engineering 1205:72
21. Gonzalez R, Woods R. (1993) Digital image processing. Addison Wesley
22. Castelman KR (1979) Digital image processing. Prentice Hall, Englewood Cliffs
23. Agard DA, Hiraoka Y, Shaw P, Sedat JW (1989) Methods Cell Biol 30:353
24. Jannson PA, Hunt RM, Plyler EK (1970) J Opt Soc Am 60:596
25. Meinel ES (1986) J Opt Soc Am A. 3:787
26. Verveer P, Jovin T (1997) J Microsc 188:191
27. Richardson W (1972) J Opt Soc Am 62:55
28. Lagendikk RL (1990) PhD Thesis
29. Carrington WA, Lynch RM, Moore EDW, Isenberg G, Fogarty KE, Fay ES (1995) Science 268:1483
30. Van der Vort H, Strasters K (1995) J Microsc 178:165
31. Verveer P, Gemkow M, Jovin T (1999) J Microsc 193:50
32. Press WH, Flannery BP, Teukosky SA, Vetterling WT (1999) Numerical recipes in C. University Press, Cambridge UK

33. Verveer P, Gemkow M, Jovin T (1999) J Microsc 193:50
34. Holmes TJ (1992) J Op Soc Am A 9:1052
35. Krishnamurthi V, Liu YH, Bhattacharyya S, Turner JN, Holmes TJ (1995) Appl Opt 34:6633
36. Holmes TJ, Bhattacharyya S, Cooper JA, Hanzel D, Krishnamurthi V, Lin WC, Roysam B, Szarowski DH, Turner JN (1995) In: Handbook of biological confocal microscopy. Plenum, New York, p 389
37. Sibarita JB, Magnin H, De Mey J (2002) IEEE Proc
38. Magnin H (1991) PhD Thesis INPG, University of Grenoble, France
39. Magnin H, Coulomb JL (1990) Proc. STRUCOME conference, Paris, France, p 563
40. Natrajan A, Crowley M, Wilkins-Diehr N, Humphrey MA, Fox AD, Grimshaw AS, Brooks CL (2001) Proc 10th IEEE International Symposium on High Performance Distributed Computing, San Francisco

Adv Biochem Engin/Biotechnol (2005) 95: 245–265
DOI 10.1007/b102216

Spectral Imaging and Linear Unmixing in Light Microscopy

Timo Zimmermann (✉)

Advanced Light Microscopy Facility, European Molecular Biology Laboratory,
Meyerhofstrasse 1, 69117 Heidelberg, Germany
tzimmerm@embl.de

Abstract Fluorescence microscopy is an essential tool for modern biological research. The wide range of available fluorophores and labeling techniques allows the creation of increasingly complex multicolored samples. A reliable separation of the different fluorescence labels is required for analysis and quantitation, but it is complicated by the significant overlap of the emission spectra. This problem can be addressed on the acquisition and the processing side by the use of spectral imaging in conjunction with linear unmixing of the image data. This method allows the reliable separation of even strongly overlapping fluorescence signals and has become an important tool in colocalization and in FRET studies. In this chapter, the microscope techniques available for spectral imaging are presented and the theory of linear unmixing is explained. Possible limitations as well as approaches for image optimization are discussed to help to realize the full potential of this novel method. Biological applications that can be improved by spectral imaging and linear unmixing are presented.

Keywords Spectral imaging · Linear unmixing · Fluorescent proteins · FRET

List of Abbreviations

CFP	Cyan fluorescent protein
EGFP	Enhanced GFP
EYFP	Enhanced YFP
FRET	Fluorescence resonance energy transfer
FTS	Fourier transform spectroscopy
GFP	Green fluorescent protein
SNR	Signal to noise ratio
YFP	Yellow fluorescent protein

1 Introduction

Fluorescence microscopy is an essential tool for modern biological research, especially in cellular and molecular biology. In its methods it is a continuously expanding field. New fluorophores are introduced [1, 2] and more and more spectral variants of fluorescent proteins are made available as markers [3–5].

Multichannel fluorescence imaging makes use of this diversity of available markers to visualize different aspects of the same specimen with specific fluorescent labels. In this way the localization of several cellular proteins can be compared during processes like cell division [6] or secretion [7]. More and more channels contain complex information and quantitative analysis is required on these data. A quantitative approach depends on exact data. On closer observation, there are however inherent problems with multichannel fluorescence images for this. Fluorescence signals may not be completely separated in the different channels and an unambiguous identification (and quantitation) may therefore not be possible. This is an important consideration especially when the amounts of the different labels are very different or when the choice of available labels is limited as it is the case for fluorescent proteins.

What is required are techniques to reliably analyze complex multichannel data. Recently, processing methods developed for multiband satellite images have successfully been applied also to the analysis of microscope images [8, 9]. By this linear unmixing of multichannel data, "clean" representations of the fluorophores in the sample can be created and be quantitatively analyzed.

In this chapter, the problems of multichannel fluorescence imaging are explained and methods to correct the image data are presented. Important parameters of the methods as well as possible limitations are discussed and applications for these techniques are shown.

2 The Problem: Crosstalk

Upon excitation with light of appropriate wavelengths, fluorophores will reemit some of the acquired energy as fluorescence. The emitted fluorescence

light is distributed over a wide spectral range according to the chemical properties of the fluorophore. This characteristic emission wavelength distribution is the emission spectrum of the fluorophore (Fig. 1A). The emission spectra of most fluorophores share two characteristics. Their onset in the shorter wavelength range is clearly defined by a steep increase towards the emission peak. The decrease of emission contributions towards the longer wavelengths is much less steep. A fluorophore emission spectrum therefore has a long emission tail towards the red that covers up to 100 nm of spectral range behind the peak emission. It contains a significant amount of the total emission of a fluorophore.

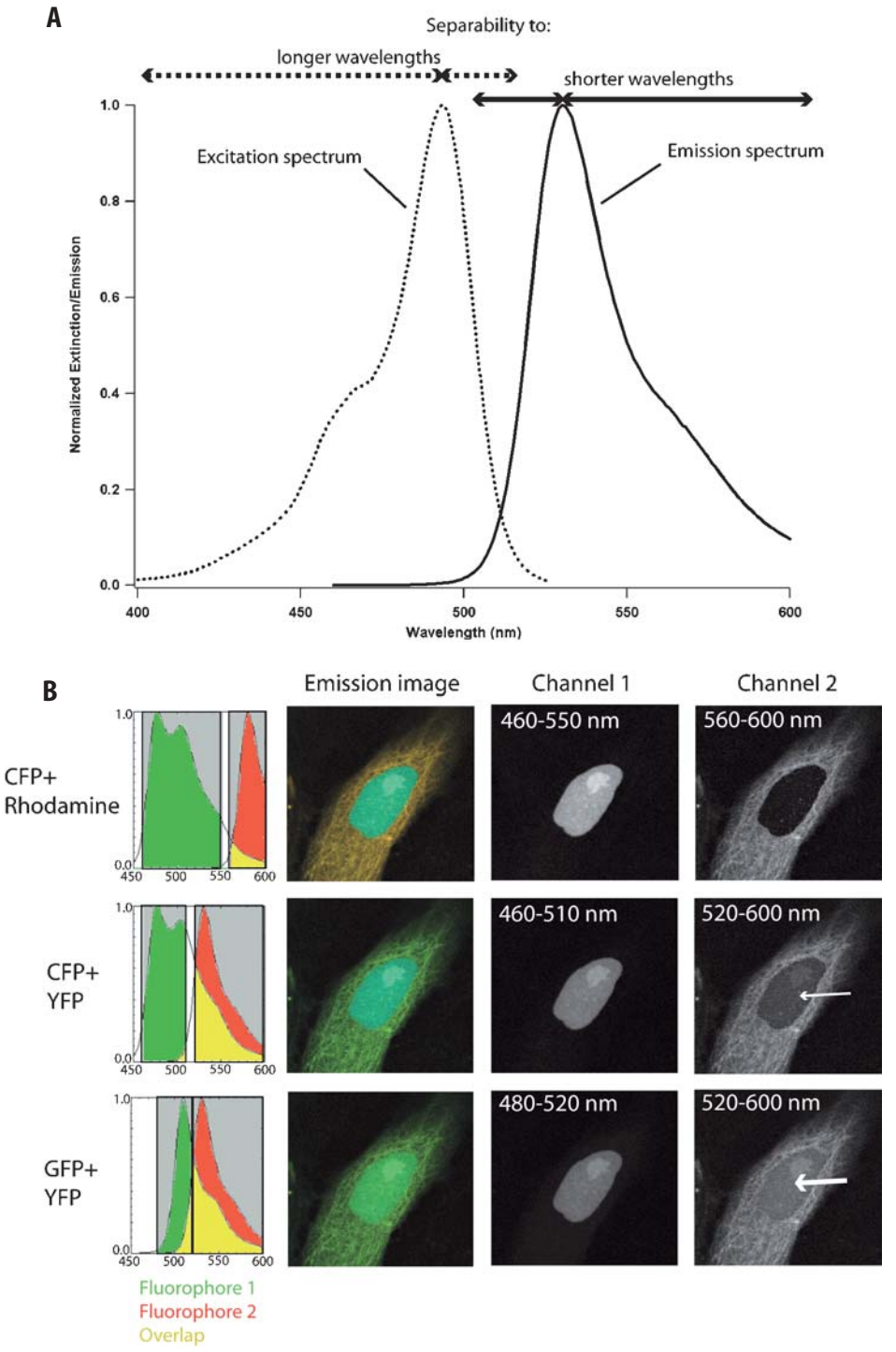
The overlap of different fluorophore spectra due to their wide emission range is an inherent problem of multichannel fluorescence imaging (Fig. 1B). In the absence of additional signals, a fluorophore can be detected through a longpass filter that covers its whole emission spectrum. If additional fluorophores are present, a bandpass filter around the emission peak is required to constrict the spectral detection range and to thus reduce the crosstalk.

While looking into the microscope, the problem may not be so severe. The fluorophores have distinct color hues that are discernible by the eye. We can thus distinguish between a green GFP signal and the more yellowish autofluorescence that may also be there. This is not the case for most of the detection devices used for fluorescence imaging. The spectral information (color) is lost and only intensity information is preserved. Like in a black and white photograph, it is then not possible anymore to draw conclusions about the real color of an object. Accordingly it is hard to assign a structure to a specific fluorescence label (Fig. 1B). This is especially important in cases where labels may colocalize.

For many standard fluorescence imaging applications the problem of crosstalk between fluorophores can be easily overcome. The perceived 'brightness' (amount of emitted photons) of a fluorophore signal in an image channel is determined by two factors: the spectral range detected by the emission filter *and* the efficiency of fluorophore excitation. This efficiency is determined by the absorption spectrum of the fluorophore (Fig. 1A) and the wavelength of the light used. Therefore even a fluorophore having a significant overlap in the emission with another fluorescent label may not produce significant crosstalk if the excitation wavelength is chosen properly. As the excitation efficiency of a fluorophore steeply decreases behind its excitation peak, fluorophores of longer wavelength properties can usually be specifically excited at their excitation maximum without co-exciting a fluorescent label of shorter wavelength (Fig. 1A).

In standard fluorescence imaging, two complementary mechanisms are thus used to separate multiple fluorescent labels reliably.

– A bandpass filter around the emission maximum separates the generated emission of a fluorophore from contributions of co-excited labels of longer wavelength.


A
Separability to:
longer wavelengths
shorter wavelengths
Excitation spectrum
Emission spectrum
Normalized Extinction/Emission
Wavelength (nm)
B
Emission image
Channel 1
Channel 2
CFP+
Rhodamine
460-550 nm
560-600 nm
CFP+
YFP
460-510 nm
520-600 nm
GFP+
YFP
480-520 nm
520-600 nm
Fluorophore 1
Fluorophore 2
Overlap

– Light at the excitation maximum of one fluorophore does not usually further excite shorter wavelength fluorescent labels. Thus, no crosstalk is generated by them.

This approach works well with the standard labels used in fluorescence imaging (e.g. DAPI, FITC, TRITC, Cy5 or equivalent labels). There are however drawbacks and limitations:

– The acquisition of the channels is sequential. Only one excitation can be used at the same time to avoid crosstalk.
– A bandpass emission filter specific for a certain fluorophore detects only approximately 50% of the available photons of this fluorophore. The rest is rejected because they are outside the spectral range of the filter.
– The approach does not work for fluorophores with highly overlapping spectra or with unusual Stokes shifts (spectral distance between excitation and emission).

This translates into limitations in speed (sequential imaging), sensitivity (detected photons) and labeling (available fluorophores).

In advanced imaging applications, and especially in in-vivo imaging, these limitations are very significant.

– Sequential imaging means more time is required for the acquisition of a single timepoint and temporal resolution is impaired. In addition, the fluorescence channels themselves may be mismatched if a fast movement occurs between the channel acquisitions.
– The amount of available fluorophore in the sample may be very small. In such cases sensitivity in the detection is required.
– The choice of spectral variants of fluorescent proteins for in-vivo imaging is still very limited. Those that work best have considerable spectral overlap that cannot be separated by using specific filtersets. Many of the available in-vivo dyes have spectral characteristics that are not easily matched to standard filtersets.

Methods that accurately correct channel crosstalk instead of just avoiding it by sequential imaging are a possibility to overcome these limitations.

Fig. 1 **A** Excitation and emission spectra of fluorophores. On the excitation side, fluorophores are separable from longer wavelength fluorophores because of the steep decrease of their absorption properties towards the red. Inversely, the emission spectrum is clearly limited towards the blue, allowing separation from shorter wavelength dyes. **B** Examples of increasing crosstalk of fluorophore emissions. CFP and rhodamine are almost completely separated in their emission and no crosstalk is perceived when they are imaged through suitable filters. CFP+YFP and GFP+YFP show increasing amounts of crosstalk between their emissions (indicated in *arrows of increasing size*). Different emission spectra were assigned to the same two-channel data set by computer simulation

3 Solutions

Multichannel fluorescence imaging of fluorophores with overlapping spectra is confronted with similar problems as the very different field of remote sensing. Multiband images taken by satellites contain different aspects of the same scene. These image bands do not contain clearly separated information about objects in the image; this information is instead distributed over several or all of the image channels. The information for a certain type of object does however have a certain characteristic distribution over the image bands, a 'signature'. For years, methods have been applied in the remote sensing field to extract specific object information by using the known signatures of objects and resolving their amount of contribution to the total signal of the multiband dataset [10].

The object signature of multiband satellite images is quite similar to the representation of a fluorescence emission spectrum in multiple filtersets. Recently the same approach used in satellite imaging has also been applied to the imaging of multiple fluorophores [8, 9, 11].

Three analysis methods are commonly used in satellite imaging and have been tested for multichannel microscopy [8]:

- Supervised classification analysis
- Primary component analysis
- Linear unmixing

The first two methods are classification based. Classification analysis of spectral data has for some time now been used in spectral karyotyping [12] where objects (in this case chromosomes) have only one characteristic signature. Those approaches are however not suitable for the quantitative analysis of samples with possibly colocalized labels as is the case for immunostainings or GFP preparations of tissues or cells. Linear unmixing is the method best suited to analyze mixed contributions to a pixel, as is the case for colocalized labels [8, 9].

3.1 Linear Unmixing

3.1.1 Principle

Linear unmixing is based on the assumption that the total detected signal S for every channel λ can be expressed as a linear combination of the contributing fluorophores FluoX:

$$S(\lambda) = A_1 \times \mathrm{Fluo1}(\lambda) + A_2 \times \mathrm{Fluo2}(\lambda) + A_3 \times \mathrm{Fluo3}(\lambda) \ldots \quad (1)$$

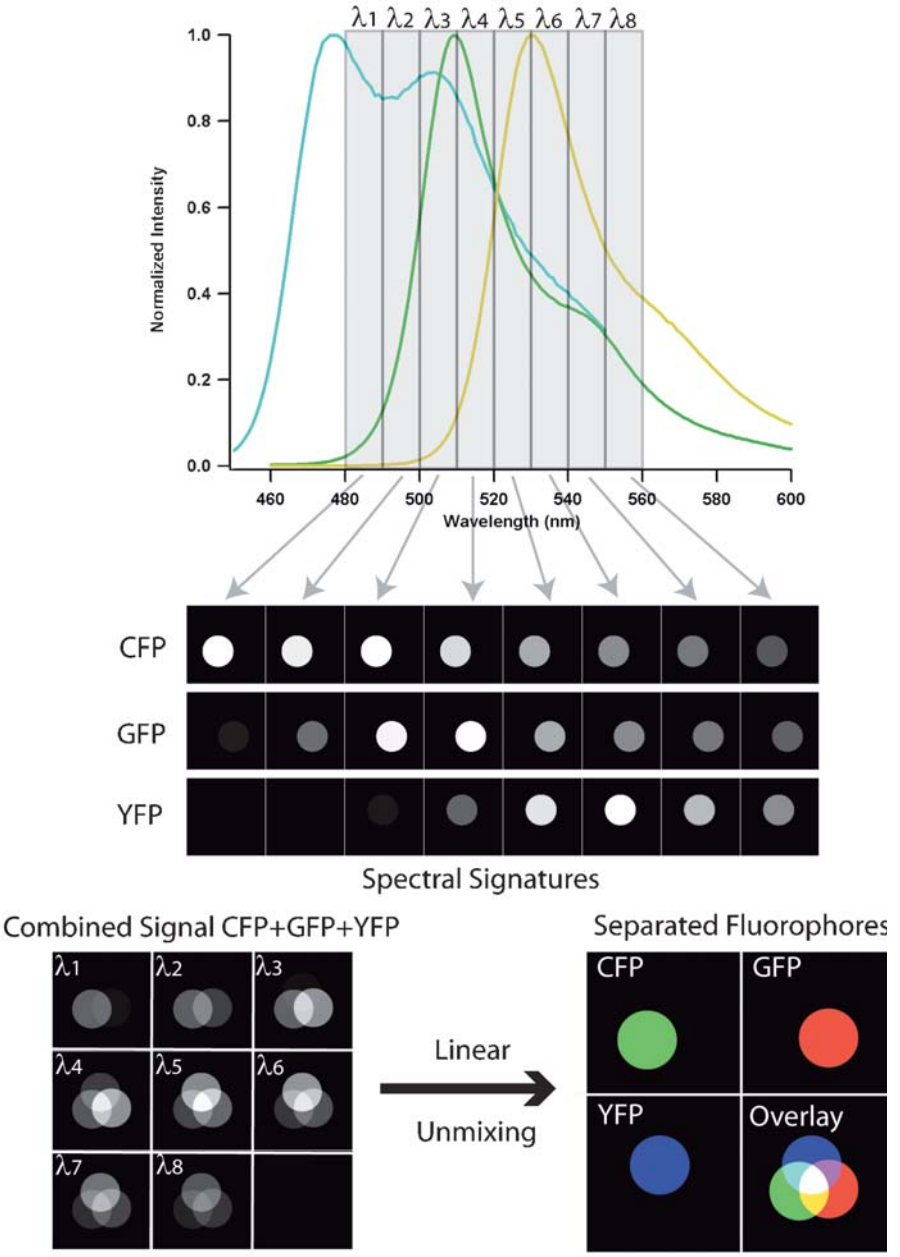


Fig. 2 Spectral imaging of fluorescence signals. Contributions of CFP, GFP and YFP to eight successive spectral channels are shown. The distribution of emission signal to the channels is a direct representation of the fluorophore emission spectrum and constitutes a spectral signature. With linear unmixing using these spectral signatures as reference, even combined and mixed signals can be clearly separated into the fluorophores that contribute to the total signal

A_x represents the amount of contribution by a specific fluorophore.
More generally this can be expressed as

$$S(\lambda) = \sum A_1 \times R_i(\lambda) \tag{2}$$

or

$$S = A \times R \tag{3}$$

where R represents the reference emission spectra of the fluorophores [11].

With the signal S detected and the reference emission spectra R known, the contributions A of the fluorophores in the sample are determined by calculating contribution values that most closely match the detected signals in the channels. This can be done by a least-square fitting approach that minimizes the square difference between calculated and measured values with the following set of differential equations:

$$\frac{\partial \sum_j \{S(\lambda_j) - \sum_i A_i R_i(\lambda_j)\}^2}{\partial A_i} = 0 \tag{4}$$

where j represents the number of detection channels and i the number of fluorophores.

The linear equations are usually solved with the singular value decomposition method [9, 13]. After the calculation of the weighing matrix A, clear representations of the separated fluorophores can be created (Figs. 2 and 3).

For the linear unmixing of spectral data, several criteria have to be met (see Table 1):

- The number of spectral detection channels must be at least equal to the number of fluorophores in the sample. If this is not the case, multiple solutions are possible and no unique result can be attained for spectral separation.
- All fluorophores present in the sample have to be considered for the unmixing calculation. If this is not done, the results will inevitably be false! The unmixing calculation is however not affected by taking into account fluorophore spectra in addition to the ones present in the sample. Zero contri-

Table 1 Requirements for correct results in linear unmixing

Unmixing parameter	$<N_{fluorophores}$	$=N_{fluorophores}$	$>N_{fluorophores}$
Detection channels	No unique solution (underdetermined)	Unique solution (determined)	Unique solution (overdetermined)
Factors in the unmixing equation	False result	Correct result	Correct result

$N_{fluorophores}$ stands for the number of fluorescent labels present in the sample.

bution values will be assigned to them. In the case of background contributions (e.g. autofluorescence), these also have to be defined spectrally and be treated as additional spectra.

It has to be considered that linear unmixing is performed on measured microscope data. Accordingly the values used for the unmixing calculations unavoidably include measurement errors that depend on the measurement conditions. By using a fitting approach (least squares, see above) valid results can be obtained even with such deviations and an estimate of the quality of the fit is possible (for overdetermined but not for determined equation systems). Factors influencing the quality of linear unmixing will be discussed in the section 'relevant parameters'.

Not all cases that are possible for equation systems in linear algebra have to be considered for measured data. As a fluorophore is either present or absent in the sample, negative contribution values are not possible and can be excluded from the available solutions. By using constrained unmixing algorithms, the calculated fluorophore contributions are made to sum to unity to avoid leftover values. This facilitates comparisons between separated fluorophores and keeps the unmixed data in close relation to the measured values.

In addition to the linear unmixing method described above, alternative approaches for the spectral unmixing of microscopic data have also been implemented, especially for datasets consisting of only few spectral channels [14]. These approaches are based on the correlation of the intensity values of a pixel in different image channels (as can be visualized in scatterplots similar to the ones used in cytofluorimetry). The unmixing is achieved by finding the distribution angles of the desired fluorophores in the scatterplot and by orthogonalizing them into separate channels ("stretching" them onto different axes of the plot). The method in principle does not require a priori information about the spectra because the main distributions can be found by line fitting. This only works reliably however only if significant amounts of the labels are present without colocalization.

3.1.2 Methods

3.1.2.1 Unmixing on the Emission Side

Spectral imaging has been implemented in several ways on widefield or confocal microscopes. Some of the solutions require specific instrumentation, but the method is also very generally applicable since any multi-channel fluorescence image can be considered as a series of spectral images. On a standard widefield fluorescence microscope, spectral separation of overlapping fluorophores was shown to be improved by determining and correcting for the crossover of individual fluorophores into different filter sets [15, 16]. Fourier

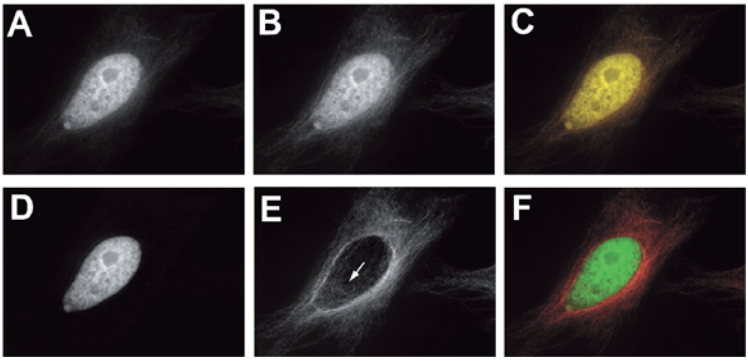


Fig. 3A–F Application example for spectral imaging and linear unmixing. Excitation based unmixing of HeLa cells containing Histone-EGFP and Alexa 488 stained microtubules. The overlap of the two emission spectra is significant as the emission peaks (EGFP: 508 nm, Alexa 488:515 nm) are only 7 nm apart. The sample was imaged on a widefield microscope with just one emission filter (530/50) and beamsplitter (500) under two different monochromator excitations using autoexposures for the channels: **A** image taken with 470±7 nm excitation; **B** image taken with 490±7 nm excitation; **C** overlay image of the two channels without linear unmixing. The crosstalk of the SFP-labeled nucleus is clearly visible as yellow; **D–F** images of the cells after linear excitation unmixing: D – Histone-GFP; E – Alexa 488 labeled microtubules. Even the microtubules on top of the nucleus are detected in this widefield image; (*arrow*) F – overlay image of the unmixed fluorophores

transform spectroscopy (FTS) has been implemented by coupling an interferometer to a widefield microscope to obtain detailed spectral information at every position in a sample [9, 17]. Initially just used with classification algorithms to do spectral karyotyping [12], the applications were subsequently extended to microscope data with colocalized labels. Analyzing FTS image data with linear un-mixing using singular value decomposition, up to seven fluorophores were simultaneously imaged and distinguished [9]. The same linear unmixing approach was then also applied to spectral data obtained with two-photon confocal microscopy and a liquid crystal tunable filter to separate up to four FP variants [8]. These approaches acquire the spectral data sequentially as a series of images (λ-stack). This is time-consuming and usually requires several minutes to acquire one λ-stack. The methods further suffer from fluorophore specific photobleaching rates which become most significant in sequential image acquisition. This poses serious problems for imaging living samples where the localization of the FPs can rapidly change during data acquisition and photobleaching may cause unpredictable effects on cell physiology. These problems with sequential data acquisition can be overcome by acquiring the spectral information into parallel detection channels. This has now been implemented on commercial confocal microscopes such as the LSM510 Meta from Zeiss which uses a grating for the spectral dispersion of the signal onto a multidetector array [11, 13]. Flexible parallel acquisition of spectral image data has also been implemented on prism based spectral confocal micro-

Table 2 Overview of the hardware capabilities ofdifferent microscope setups for spectral imaging

Acquisition mode	Overdetermined ($N_c \gg N_f$)	Determined ($N_c \approx N_f$)
Parallel	Zeiss LSM 510 Meta[a]	Leica SP/SP2/SP2-AOBS[b] Bio-Rad Rainbow[b] Beamsplitter setups Confocal microscopes
Sequential	Zeiss LSM 510 Meta[a] Leica SP/SP2/SP2-AOBS[b] Bio-Rad Rainbow[b] LCTF detection SpectraCube (FTS)	Widefield microscopes w. Filterwheels/filtercubes

[a] The Zeiss Meta detector can read out 8 of its 32 channels at the same time. To read out all 32 channels, 4 sequential acquisitions have to be made.
[b] Sequential: λ-series into a single detector. Parallel: Multiple detectors (3–4).
N_c: number of detection channels.
N_f: number of fluorophores.

scopes, the Leica AOBS and SP systems [14]; see also [18]. More recently, a filter-based spectral confocal approach was also introduced (Bio-Rad Rainbow)

It has become apparent that linear unmixing of microscope data can be realized on a wide variety of microscopic setups. This is because in many cases only a few channels ($N_{channels} \geq N_{fluorophores}$) are required for spectral imaging and unmixing. The number of fluorophores in a typical biological sample is very often limited to two or three. These may however be strongly overlapping as is the case for FPs and therefore require unmixing. An overview of systems and their capabilities is given in Table 2.

3.1.2.2 Unmixing on the Excitation Side

The range of possible approaches for linear unmixing can be significantly extended by looking at the second characteristic property of fluorophores, their excitation spectra. As for emission spectra, linear unmixing can also be applied based on the fluorophores' excitation spectra. Instead of exciting at one wavelength and collecting the emitted fluorescence into spectrally different detection channels, it is possible to excite sequentially at different wavelengths and detect the respective fluorescence with one detector only (Fig. 3). The total emission can be collected into just one detector so that the SNR for the acquired data is very high. The data quality is an important parameter for linear unmixing (see section 'relevant parameters'). Data acquired in this way can then be analyzed with the same algorithm used for emission based unmixing (see section 'linear unmixing'). The excitation-based implementation of spectral un-

mixing should be ideal on set-ups equipped with excitation sources that allow fast switching of the excitation light.

Excitation-based unmixing is being used in commercially available setups for widefield as well as for two-photon imaging. In contrast to the parallel detection possibility for emission unmixing, excitation unmixing is unavoidably based on sequentially acquired data.

As the number of fluorophores separable by linear unmixing is limited by the number of channels available for analysis (see Table 1), the combination of excitation and emission unmixing even increases the number of fluorophores that can be distinguished in a sample.

The quality of the unmixed data depends directly on the acquisition settings and the quality of the spectral data. It is therefore important to understand thoroughly the parameters affecting spectral imaging and linear unmixing to be able to judge the strengths and weaknesses of the diverse acquisition possibilities.

3.1.3 Relevant Parameters

Since spectral imaging is an established method in the field of satellite imaging and remote sensing, factors affecting efficiency have already been thoroughly considered [10]. The method is predominantly limited by factors such as image background or detector noise, but the appropriate selection of the number and bandwidth of the detection channels with respect to the overlap of the fluorophores to be distinguished also plays an important role [19, 20]. These limitations become most critical when imaging living specimens, where the signals of interest are usually weak and photodamage by intense irradiation has to be avoided.

3.1.3.1 Reference Spectra

For correct unmixing it is essential to use proper reference spectra. This is best done by measuring reference samples that contain only one of the fluorophores. In cases where it is certain that the signals are not colocalized in the regions used as references, samples with multiple labels could also be taken. This approach does however bear the risk of taking spectra with slight contributions of other fluorophores as references. In the unmixed image, this would result in assignment errors that would make any analysis of colocalization impossible.

For the processing of spectral data, it is advisable to take only the reference spectra of the fluorophores contained in the sample for the unmixing calculation. Inclusion of unnecessary reference spectra may degrade the resulting image because noise and other aberrations contained in the spectral data would result in a higher unmixing error under these conditions.

Reducing the number of reference spectra is one approach to optimize unmixing. In methods like iterative endmember ejection, spectra that are clearly not present in a pixel (negative contribution) are not used in the unmixing calculation.

3.1.3.2 Background

Removing any signal not originating from the fluorophores to be analyzed by background subtraction is an essential prerequisite for the linear unmixing analysis [9]. As with ratiometric measurements [21], failure to correct the background properly leads to significant intensity-dependent artifacts in the processed images. However, in cases of spectrally homogeneous background, there is the option to treat the background as a further fluorophore and thus separate it from the specimen specific fluorescent signals by linear unmixing.

3.1.3.3 Noise

Since linear unmixing is a pixel-based method, it is susceptible to errors introduced into the original images or the reference spectra by the Poisson noise of the fluorescence light itself and the detector readout noise. These sources of noise become important at low light levels in live specimen imaging where usually the illumination light and exposure time have to be kept at a minimum to preserve the physiological integrity of the sample. Computer simulations show that in the absence of noise, unmixing efficiencies are independent of the number of detection channels used (Fig. 3B). However, in the presence of detector readout noise, the errors in the processed image increase relative with the number of detection channels (Fig. 4A,B). For a constant detector readout noise, the detector signal to noise ratio (dSNR, does not include other sources of noise) will decrease according to 1/n (n=number of detection channels) for each detector. This decrease of the dSNR is only partially compensated by the noise averaging effect that occurs due to the increased number of detection channels (improvement according to $n^{-1/2}$; see [22]). Consequently, the dSNR decreases according to $n^{-1/2}$ and thus sampling the spectral information into few detection channels with broad bandwidth should result in superior quality of the unmixed data compared to sampling into a large number of detection channels with narrow bandwidths. This observation is relevant for measurements with high detector readout noise or with low signals. The total SNR (all noise types) of a measurement is affected by the number of detection channels proportional to the contribution of detection noise to the total noise.

Detection noise is signal independent and therefore additional detectors will increase the total noise of a measurement. This is not the case for the Poisson distributed noise of the emission light (photon shot noise), a significant factor for weak fluorescence signals. It is not significantly affected by the number of

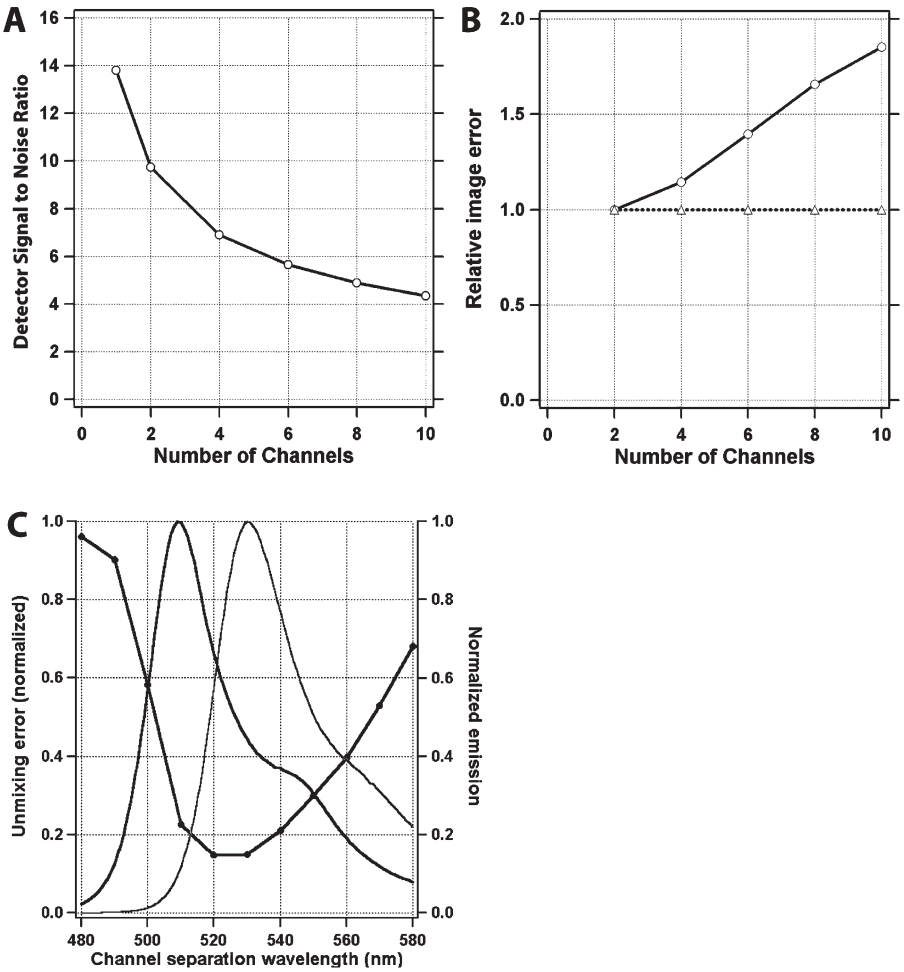


Fig. 4A–C Examples of factors influencing the efficiency of spectral unmixing. The simulation data were created and processed with routines written in Interactive Data Language (Research Systems, Inc.): **A** decrease of the detector signal to noise ratio in dependence of the number of channels used for sampling the spectral information; **B** dependence of unmixing efficiency on the number of detection channels in the absence and presence of noise. The relative increase of the unmixing error is independent of the actual noise level. Image error values in the graph are normalized to the image error obtained for two channel unmixing. *Solid line with circles*: relative image error in the presence of detector noise. *Broken line with triangles*: relative image error without noise; **C** influence of detection channel characteristics on the unmixing exemplified for two detection channels. A GFP signal was simulated and unmixed against YFP with detector noise added. Moving the border between the two detection channels to different wavelengths demonstrates the existence of an optimal position corresponding to a minimal unmixing error. The spectra for GFP (*dark gray*) and YFP (*light gray*) are superimposed on the plot (*right axis*)

detection channels acquiring the total signal. This is due to the fact that the standard deviation of the Poisson noise depends on the intensity of the signal. With this signal-dependent noise, less signal per channel also means less noise in that channel. Although the photon shot noise SNR ($N/\sqrt{N}$ for N photons) is lower for weaker signals per channel, the total photon noise of all channels remains constant ($SNR = \frac{\sum N_j}{\sqrt{\sum (\sqrt{N_j})^2}}$ for j channels). Thus results obtained with few or many channels are in this aspect similar.

3.1.3.4 Detector Channel Arrangements and Spectral Overlap of the Fluorophores

In contrast to significantly overdetermined systems, the filter settings for setups with approximately as many channels as fluorophores should be chosen carefully for optimal results [19, 20]. An optimized arrangement of the detector channels to obtain the least unmixing error can be determined according to the spectra of the fluorophores (Fig. 4C). This is most easily demonstrated with two partially overlapping fluorophore spectra (a,b) detected with two channels. The optimized solution corresponds to a maximal separability, found by calculating the endmember separability with the SVD method (see above) for the diverse filter setups. Alternatively, it can also be found using a 'figure of merit' as a measure of the efficiency in the use of emitted photons [19]. For most fluorophore combinations, the optimal cutoff between two channels is located at the intersection of the falling slope of the first with the rising slope of the second fluorophore spectrum.

However, even with optimized channel settings, the unmixing efficiency is directly affected by the amount of spectral overlap between the fluorophores. In the absence of noise, the degree of overlap does not affect unmixing efficiency. In the presence of noise the errors of the unmixed data increase with the spectral overlap of the fluorophores. One way to overcome this is by over-sampling, which has been nicely exploited on a widefield microscope by applying sequential imaging with detection filter arrangements partially overlapping in sequential acquisitions [23].

In the case of significantly overdetermined setups (non-overlapping), the separability (and therefore the quality of the unmixed image) is solely determined by the amount of overlap between the used fluorophore spectra.

3.1.3.5 Number of Channels

Both determined (n channels=n fluorophores) as well as overdetermined systems (n channels>n fluorophores) can be used for spectral imaging and linear unmixing [19, 20]. Are there differences in performance?

As shown above, readout noise increases with the number of detection channels used. In cases of significant readout noise, this affects the quality of the unmixing. On the other hand, an estimate of the error of the fit is possible for overdetermined, but not for determined systems by comparing the fit with the raw data of each channel.

Whereas the positioning of the detection channels is very relevant for determined setups (see above), significantly overdetermined systems are less susceptible to this influence.

The following conclusions can thus be drawn:

The best unmixing results are obtained with few and wide channels as this gives higher signals per channel and minimizes possible readout noise problems. If the gains of the channels can be set independently, the separability of the signals can be enhanced in this way. Such an approach does however require fine tuning of the settings, as only optimized settings will give an improved result. Not properly chosen settings will give inferior results.

Overdetermined systems may not inherently give better results, but can be used without the fine tuning of filter settings.

3.2 Subtraction

In some cases of overlapping fluorophores, a straightforward subtraction approach, as used in many formulas for sensitized emission detection (see below), can be used instead of linear unmixing. Here, the contribution of the first fluorophore A into the detection channel for the second fluorophore B is determined and expressed as a normalized value R_A:

$$R_A = \frac{A_2}{A_1} \tag{5}$$

To get the second fluorophore without contribution of the first, a simple subtraction is performed using the information of the first channel:

$$B = Ch_2 - (R_A \times Ch_1) \tag{6}$$

This approach works also for fluorescent proteins [24]. It also works for more than two fluorophores (see below).

There are however limitations to this technique. It requires at least one spectral channel to contain one fluorophore without any contribution of the other:

$$R_x(\lambda) - \sum_i R_i(\lambda) = 0, \quad x \in \{1 \ldots i\} \tag{7}$$

That channel is used for the subtraction. In cases of mixed contributions in all channels it is not applicable! In cases of multiple fluorophores this condition

has to be met in the beginning and again after each subtraction of a fluorophore. The subtraction technique is therefore limited to only a subset of the cases that can be solved by linear unmixing. In practice this means that not all fluorophore combinations can be used and that more restrictive filter settings have to be used that do not sample all of the available signal.

A second shortcoming of the subtraction technique is that while bleed-through is corrected, the additional signal contained in other channels is not further utilized. In linear unmixing, the total signal of all channels is distributed onto the fluorophores according to their contributions. The unmixed signals can therefore be significantly brighter than their representation in the single spectral channels. With the subtraction method, this is not possible.

The susceptibility of the subtraction technique to noise artifacts is equal to the linear unmixing technique.

In comparison to the subtraction method, linear unmixing is more flexible, uses more of the available information and is thus the approach best suited to process data with overlapping fluorophores.

4 Applications

Almost any fluorescence microscope can be used for spectral imaging and the data can be stored as image files. But how to perform the linear unmixing? By now, unmixing tools are provided in the software of several confocal microscopes (Zeiss, Leica, BioRad). They are now also available for some widefield microscopes (Olympus BioSystems). In addition it is possible to implement linear unmixing easily in any image processing package or as an ImageJ plugin. Appendix A contains a numerical recipe for simple linear unmixing that can be used in any software.

With the processing tools available, spectral imaging and linear unmixing can be applied to many forms of biological imaging.

4.1 Timelapse Imaging

Because of the highly dynamic nature of processes in living samples, it is essential for in vivo co-localization experiments or ratiometric methods to gather information of two or more fluorescent molecules at the same time. With GFP-tagged proteins this is usually difficult to achieve due to the lack of spectral variants that can be excited simultaneously but do not bleed through significantly in the emission channels. Further complications arise when the need to resolve fast processes, such as the movement of membrane transport carriers [25], precludes the use of time consuming sequential image acquisition which can overcome bleed through problems with GFP variants to some extent [26]. For these problems, spectral imaging offers the possibility to excite two spec-

trally similar fluorophores, for example EGFP and EYFP, with just one wavelength (e.g. 488 nm) and detect them simultaneously without significant losses of the emitted fluorescence. In contrast to common methods involving glass filters to separate the emitted light from distinct fluorophores, spectral imaging collects almost all of the fluorescence emitted, which is critical for work with living samples. Thus spectral imaging, even with only two fluorophores being involved, should always be the method of choice when working with living samples.

4.2 Fluorescence Resonance Energy Transfer

A further application of spectral imaging is in FRET microscopy, which is an important tool for imaging the dynamics of cellular processes like protein-protein interactions or post-translational modifications [27]. However, an efficient FRET signal requires significant overlap between the emission spectrum of the donor and the excitation spectrum of acceptor fluorophore. This requirement is inevitably accompanied by a significant overlap of the emission spectra of the donor and acceptor and complicates the determination of FRET efficiencies [27, 28]. As spectral imaging is well suited to separate even highly overlapping donor and acceptor emissions, FRET imaging with already established donor acceptor pairs can be facilitated [13] and applied to previously unused FP-based donor acceptor pairs with increased FRET efficiencies due to the increased spectral overlap of the donor emission and acceptor excitation [18].

FRET in living samples is very often observed with methods that detect and quantify the sensitized emission of the acceptor. In the most common methods, the channel crosstalk of the measurement is corrected by subtraction of image channels [27] (see section 'Subtraction'). It is easily possible to replace this step with linear unmixing. As linear unmixing covers also cases of strongly overlapping fluorophores, the method can thus be extended and less restrictive filter settings that collect more of the signal can be used (contributions of both fluorophores to both channels are allowed).

References

1. Zhang J, Campbell RE, Ting AY, Tsien RY (2002) Nat Rev Mol Cell Biol 3:906
2. Chan WC, Maxwell DJ, Gao X, Bailey RE, Han M, Nie S (2002) Curr Opin Biotechnol 13:40
3. Lippincott-Schwartz J, Patterson GH (2003) Science 300:87
4. Miyawaki A, Sawano A, Kogure T (2003) Nat Cell Biol Suppl:S1
5. Hu CD, Kerppola TK (2003) Nat Biotechnol 21:539
6. Gerlich D, Beaudouin J, Gebhard M, Ellenberg J, Eils R (2001) Nat Cell Biol 3:852
7. Stephens DJ, Lin-Marq N, Pagano A, Pepperkok R, Paccaud JP (2000) J Cell Sci 113(12):2177
8. Lansford R, Bearman G, Fraser SE (2001) J Biomed Opt 6:311
9. Tsurui H, Nishimura H, Hattori S, Hirose S, Okumura K, Shirai T (2000) J Histochem Cytochem 48:653

10. Landgrebe D (2000) IEEE Sig Proc Mag
11. Dickinson ME, Bearman G, Tilie S, Lansford R, Fraser SE (2001) Biotechniques 31:1272
12. Schrock E, du Manoir S, Veldman T, Schoell B, Wienberg J, Ferguson-Smith MA, Ning Y, Ledbetter DH, Bar-Am I, Soenksen D, Garini Y, Ried T (1996) Science 273:494
13. Hiraoka Y, Shimi T, Haraguchi T (2002) Cell Struct Funct 27:367
14. Olschewski F (2002) Imaging Microsc 4:22
15. Castleman KR (1994) Bioimaging 2:160
16. Kato N, Pontier D, Lam E (2002) Plant Physiol 129:931
17. Malik Z, Cabib D, Buckwald RA, Talmi A, Garini Y, Lipson SG (1996) J Microsc 182:133
18. Zimmermann T, Rietdorf J, Girod A, Georget V, Pepperkok R (2002) FEBS Lett 531:245
19. Neher R, Neher E (2004) J Microsc 213:46
20. Zimmermann T, Rietdorf J, Pepperkok R (2003) FEBS Lett 546:87
21. Bolsover SR, Silver RA, Whitaker M (1993) In: Shotton D (ed) Electronic light microscopy. Wiley-Liss, New York, p 181
22. Sheppard CJR, Gu M, Roy M (1992) J Microsc 168:209
23. Garini Y, Gil A, Bar-Am I, Cabib D, Katzir N (1999) Cytometry 35:214
24. Zimmermann T, Siegert F (1998) Biotechniques 24:458
25. Shima DT, Scales SJ, Kreis TE, Pepperkok R (1999) Curr Biol 9:821
26. Ellenberg J, Lippincott-Schwartz J, Presley JF (1999) Trends Cell Biol 9:52
27. Wouters FS, Verveer PJ, Bastiaens PI (2001) Trends Cell Biol 11:203
28. Gordon GW, Berry G, Liang XH, Levine B, Herman B (1998) Biophys J 74:2702

Appendix

Numerical Recipe for Linear Unmixing

To make the principle of linear unmixing more easily understandable, the solution for a two channel/two fluorophore situation is shown.

$Ch_{x,y}$ represent the signals in detection channels x and y, and A_x, B_x and A_y, B_y the normalized contributions of *FluoA* or *FluoB* to channels x and y as they are known from the spectral signatures of the fluorescent proteins.

Deduction:

$$Ch_x = A_x\,FluoA + B_x\,FluoB$$

$$Ch_y = A_y\,FluoA + B_y\,FluoB$$

$$Q = \frac{Ch_x}{Ch_y} \Rightarrow Ch_x = QCh_y \Rightarrow$$

$$A_x\,FluoA + B_x\,FluoB = Q(A_y\,FluoA + B_y\,FluoB)$$

$$A_x\,FluoA + B_x\,FluoB = QA_y\,FluoA + QB_y\,FluoB$$

$$A_x\,FluoA - QA_y\,FluoA = QB_x\,FluoB - B_x\,FluoB$$

$$FluoA(A_x - QA_y) = FluoB(QB_y - B_x)$$

$$\frac{FluoA}{FluoB} = \frac{QB_y - B_x}{A_x - QA_y}$$

Application:
In order to determine the fluorescence emitted by each of two individual fluorophores (*FluoA*, *FluoB*) in co-localization or FRET experiments, only four equations have to be applied for every image pixel i:

$$Q(i) = \frac{Ch_x(i)}{Ch_y(i)} \tag{A1}$$

$$R(i) = \frac{FluoA(i)}{FluoB(i)} = \frac{B_yQ(i) - B_x}{A_x - A_yQ(i)} \tag{A2}$$

The total signal $S(i) = \sum_{k=1}^{n} Ch_x(i)$ as the sum of all channels can then be divided into the contributions of FluoA and FluoB by

$$\mathrm{FluoA}\,(i) = \frac{S\,(i)}{1 = \dfrac{1}{R\,(i)}} \tag{A3}$$

$$\mathrm{FluoB}\,(i) = \frac{S\,(i)}{1 + R\,(i)} \tag{A4}$$

or computationally more simply as

$$\mathrm{FluoB}\,(i) = \frac{S\,(i)}{1 + R\,(i)} \tag{A3'}$$

$$\mathrm{FluoA}\,(i) = S\,(i) - \mathrm{FluoB}\,(i) \tag{A4'}$$

Adv Biochem Engin/Biotechnol (2005) 95: 267–295
DOI 10.1007/b102218

Tracking Movement in Cell Biology

Kota Miura (✉)

Cell Biology and Biophysics Programme, EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany
miura@embl.de

Abstract This article is an overview of techniques for measuring movement of proteins, vesicles and cells using digital image processing. Diverse techniques have been developed during the last decade. Both strong and weak points of each technique are discussed, together with short description on the actual application. Among these techniques, the optical flow estimation technique is explained in more detail and an example of protein movement analysis is described.

Keywords Tracking · Movement · Cell biology · Digital image processing

1 Introduction

Movement is an essential function in cellular system. Chemical interactions of proteins are microscopic movements of proteins, namely, diffusion, transport and the collision among the proteins. Vesicles containing proteins move to specific sites within cell for their destined functions. Chromosomes move to separate from each other during cell division. Cytoskeleton dynamically forms bundles and networks to afford routes for the intracellular trafficking and to control the physical architecture of the cell. Cells themselves move by migration. Within multicellular structure, cell movement drives the morphogenesis and preserves the homeostasis. Measurement of these various movements provides quantitative information that is inevitable for understanding the cellular system.

Digital image processing for analyzing these movements became a general technique in cell biology. Many techniques for tracking and measuring the movement have developed during the last decade. Since there has been no concise review that has focused on these techniques, I tried to collect diverse ways for measuring movement in cell biology to give a broad overview of these techniques. Each technique will be explained briefly and a short description on example applications will be added. Among all these techniques I will make a longer explanation on the gradient method based optical flow estimation since this technique is a powerful but has been mostly overlooked. I will show an example analysis of intracellular protein movement using this technique. The general procedure for image processing in biological applications was recently summarized [1]. Some of the techniques explained here were examined with their precision and compared quantitatively [2]. Readers may refer to those articles as well.

2 Techniques

2.1 Digital Video Processing

The term "digital image processing" encompasses a broad range of techniques for enhancing and modifying digital images using the computer. These techniques can be roughly categorized into two types depending on the nature of the image to deal with. One is the static image processing and the other is the dynamic image processing. The static image does not have a temporal dimension. For the processing of static images, software such as Adobe Photoshop, NIHimage[1] or its Java application ImageJ[2] equip many filtering operations and

[1] NIHimage home page (http://rsb.info.nih.gov/nih-image/).
[2] ImageJ home page (http://rsb.info.nih.gov/ij/).

the processing can readily be done. Dynamic image is what generally called "video sequences" or simply "movies". Analysis of the movement in biology inevitably deals with dynamic processes that change with time. In most cases, researchers are required to compose scripts or macros in image processing software for efficient and adequate analyses of these dynamic events.

2.2 Manual Tracking

Tracking of the moving objects, such as proteins, vesicles, organelle and cells enables measurement of their movement speed and directions. The simplest way to measure these movements in the digital image sequence is manual tracking. In earlier days when the computer-assisted method was not generally available, researchers overlaid tracing paper or transparent plastic foil on the video-monitor screen, traced the edge of the object at different time points and measured their position to calculate their speed and movement direction. Availability of digital image processing enabled biological researchers to facilitate these processes by using personal computer, the simplest of which is analogous to the classic "tracing paper" method. That is to trace the edge of the object using imaging software such as NIHimage. Then thresholding of the image leaves only the trace. Centroid is the average of all pixel coordinates inside the trace and is the most commonly used feature of object for representing its position. Centroid coordinate $p_c(x,y)$ is calculated as follows:

$$p_c(x,y) = \left(\frac{\sum x_i}{n_x}, \frac{\sum y_j}{n_y} \right) \Bigg|_{x,y \in \Re} \tag{1}$$

where $\Re$ is region surrounded by the contour. The position of the object can be measured for every frame of the sequence manually. Then the movement speed and direction can be calculated from the measured coordinates.

The basic principle of program-based particle tracking is similar to the manual tracking. It consists of two major steps. The first step is what I have described above as "tracing the edge of the object", and is termed "segmentation" in the digital image processing. The segmentation enables us to calculate the coordinate of the object position, such as its centroid. In the second step successive position of the object must be linked to calculate the speed and direction.

2.3 Thresholding

"Thresholding" is a simple way to do the segmentation quantitatively. By setting a threshold value, the image can be converted to a binary image (black and white image). Take an example of an eight-bit image of a cell labeled with fluorescence in Fig. 1a. The average pixel intensity (brightness) of this image is 68, that of the cell is ca. 105 and the average background intensity is ca. 30. By setting a thresh-

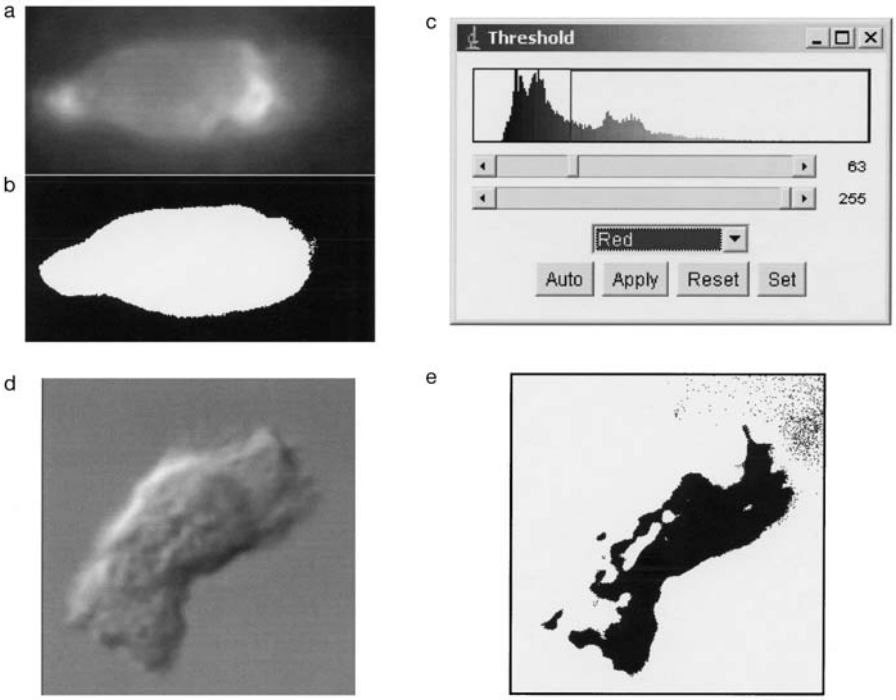


Fig. 1a–e Image thresholding: **a** A *Physarum* cell stained by Rhodamine-Phalloidin; **b** the same image after image thresholding; **c** threshold window of the image processing free ware ImageJ. Original image (a) was binarized by a threshold value 63. Pixels with intensity higher than 63 were converted to white and the remaining pixels were set to black in b; **d** a DIC image of a *Physarum* cell; **e** its threshold image (0–128: *black*. 129–255: *white*). Edge of the upper half of the cell is brighter than the background. Thresholding cannot detect this part of the cell edge

old value of 63 (Fig. 1c), pixels with brightness above 63 become white and the remaining pixels become black (Fig. 1b). After thresholding the image, the centroid of the object thus its position can be easily calculated using Eq. (1). In conventional image processing software such as NIHimage or ImageJ, centroid calculation is already included as a function.

Thresholding is a simple task, but this operation does not work in many cases. When the object is not optically flat, or the contrast of object edge is low, thresholding does not detect the edge of the object properly. For example in fluorescence images, a constant threshold value fails to detect a constant edge if the total fluorescence intensity decays by photobleaching. When a cell is observed through a DIC microscope, there are both brighter and darker portions against the background (Fig. 1d) that only a half of the cell edge is detected by the thresholding (Fig. 1e). Low signal-to-noise ratio also hinders a correct segmentation at the object edge since noise in the background tends to be included to the selected region. Uneven illumination or fluorescence signal may cause

the similar edge detection problem. To solve the photobleaching problem, the threshold level can be dynamically adjusted for each frame depending on the average brightness of the image frame (e.g. lower threshold value for a lower average intensity). However, the low S/N ratio or uneven illumination problem will still remain. For these reasons, various other techniques have developed.

2.4 Gaussian Fitting

For spherical signals such as fluorescence beads or fluorescently labeled single molecule, the signal intensity distribution can be fitted to two-dimensional Gaussian curves [3–5]:

$$I(x, y) = z_0 + z_n \exp\left\{-\frac{(x - X_n)^2 + (y - Y_n)^2}{W_n^2}\right\} \tag{2}$$

where $I(x, y)$ is the intensity distribution of an image, z_0 is the background intensity, z_n is the height of the peak. W_n is the width of the curve that peaks at (X_n, Y_n). The peak position is the signal position: $P_n(x,y)=(X_n, Y_n)$. Although this fitting method is restricted to spherical object, it yields the most precise measurements even with a very low signal-to-noise ratio [2].

2.5 Pattern Matching – Local Matching Methods

A popular technique for searching the position of target objects is "pattern matching". In this technique, a kernel containing a template pattern of the object is compared with different positions within the image to find a position with the best match (Fig. 2a). A popular way to evaluate the resemblance of the pattern is the cross-correlation function $C(x,y)$ [6]:

$$C(x, y) = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} I(x + i, y + j)\{K(i, j) - \overline{K}\} \tag{3}$$

where $I(x,y)$ is the intensity distribution of a image frame and $K(i,j)$ is a $n \times m$ pixels kernel which contains the template image pattern. $\overline{K}$ is the mean intensity of the kernel. The $C(x,y)$ becomes largest at the position where the pattern matching is at its best. In the actual application, the template pattern is sampled from the image frame $I_k(x,y)$. Then the $C(x, y)$ between the kernel (Fig. 2b) and the consecutive image frame $I_{k+1}(x,y)$ (Fig. 2c) is calculated (Fig. 2d). The template kernel K matches the best where $C(x,y)$ is maximized. We call this position (x_c, y_c). Gelles et al. (1988) introduced the following formula for obtaining the peak position (x_c, y_c):

$$x_c = \frac{\sum x\{C(x, y) - T\}}{\sum \{C(x, y) - T\}} \tag{4}$$

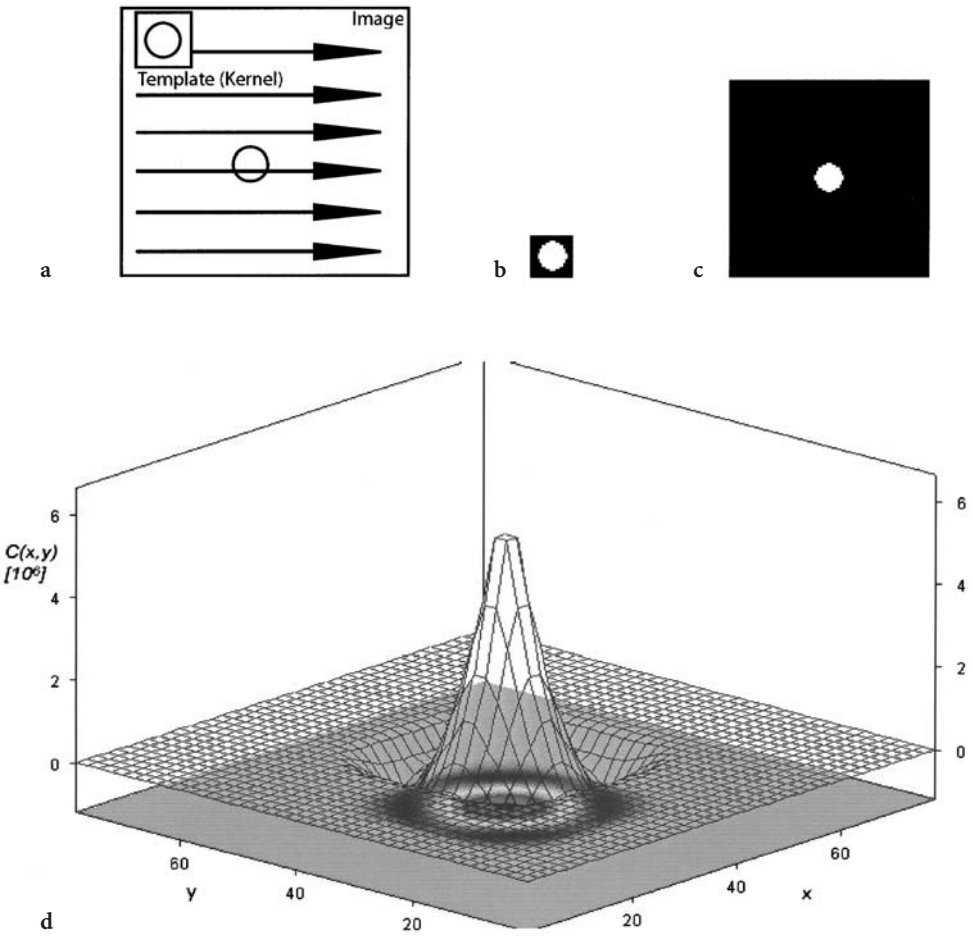


Fig. 2a–d Pattern matching method: **a** a kernel containing a template image is compared with every position within an image frame to evaluate their similarity; **b–d** an example of calculating the cross-correlation of the template image kernel (b) and an image frame (c). $C(x,y)$ was calculated according to Eq. (2) and the result is plotted in a three-dimensional graph (d). The peak position of $C(x,y)$ corresponds to the position of the object

$$y_c = \frac{\sum y\{C(x,y) - T\}}{\sum \{C(x,y) - T\}} \tag{5}$$

where T is a threshold value and negative values were discarded. (x_c, y_c) corresponds to the centroid of the magnitude of correlation that is thresholded by T.

Since the cross-correlation function (Eq. 3) tends to give higher values at the brighter regions rather than feature similarity, a normalized form of cross-correlation can also be used:

$$C_N(x,y) = \frac{\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}[I(x+i,y+j) - \bar{I}]\{K(i,j) - \bar{K}\}}{M_{Ix,y} \cdot M_k} \qquad (6)$$

$$M_{Ix,y} = \sqrt{\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}[I_{x+i,y+j}]^2} \qquad M_{Ik} = \sqrt{\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}[K_{i,j}]^2}$$

where $\bar{I}$ is the mean intensity of a portion of the image overlapping the kernel, M_k and $M_{Ix,y}$ are the root mean square value of the kernel and the corresponding portion of the image, respectively. The precision of the measurement is high [2] but has a drawback that it fails tracking the object when its shape changes.

The cross-correlation based matching method has been used extensively in single particle tracking (SPT). SPT is a powerful technique for measuring the mobility of membrane bound and the movement of motor protein with a nanometer precision [6–9]. In these studies, a single protein was attached to a very small gold particle or labeled with a fluorophore and its movement was captured by video microscopy. Theoretical examination showed that different modes of protein movement can be discriminated with nanometer resolution by measuring the mean square displacement of the labeled proteins [10]. Various types of membrane protein motions, such as immobile, directed, confined, tethered, normal diffusion and anomalous diffusion, were resolved by these studies, revealing the possible effects of the kinetics of the membrane proteins [11]. An automatic tracking program for multiple proteins was developed by Ghosh and Webb [12]. This program was used recently for measuring the movement of actin patches in yeast cells [13]. Details on the SPT microscopy technique was recently reviewed [14].

2.6 Trajectory Recovering – Nearest Neighbor Method

To recover the trajectory, the particle position in each time frame must be linked to the position in the next time frame. This operation can also be called the "registration". The position of the object $P_n(x,y)$ at the n-th frame must be linked to the coordinate of the centroid of the object $P_{n+1}(x',y')$ at a time point n+1-*th* frame to measure the displacement. If there is only a single object within the sequence, the object at the position $P_n(x,y)$ would obviously move to the position $P_{n+1}(x',y')$. However, what could be done if there were multiple objects moving in a single image sequence? If the time resolution is high enough compared to the movement speed of the objects, the best guess would be that the object that is the nearest to (x,y) is the position of the object after its displacement. Then the solution is to search for the surrounding of the position (x, y) in the n+1-*th* frame and determine the nearest object as $P_{n+1}(x', y')$ [7]. Nearest point detection can be done automatically by programming this process in the macro language of NIH image (or imageJ), by using the function

"analyze particles". For example, using NIHimage, multiple beads embedded in an actin filament gel were tracked and their mean square displacement was calculated to characterize the organization of the actin filament gel [15, 16].

A modified version of the nearest-neighbor method is a probability assignment method. A probability of finding $P_{n+1}(x,y)$ can be assigned for each pixels in the surrounding area of $P_n(x,y)$ in the n+1-th frame of [3, 4]:

$$C = \exp\left\{-\left[\frac{\Delta R}{R_d}\right]^2\right\} \cdot \exp\left\{\frac{[-\Delta Z]^2}{Z_n}\right\} \tag{7}$$

where ΔR is the change in the position, R_d is a measure of the expected change in the position, ΔZ is the change in the peak intensity and Z_n is the mean peak intensity of the spot. Single $P_{n+1}(x',y')$ can be chosen from multiple candidates in accordance with a higher probability for that position.

2.7 Trajectory Recovering – Graph Theory Based Method

Nearest-neighbor based methods fail to recover the true trajectory in some cases. The example in Fig. 3a, showing three particles flowing in the same direction, is such a case. There are several ways to link the particle at the time point k to a particle at the time point k+1. Although the true trajectory is Fig. 3c, the nearest-neighbor method would choose the shortest but incorrect link Fig. 3d. Similar miss-linking becomes highly probable when more particles are present. Danuser and his co-workers recently developed a new way to overcome these problems (Fig. 3b) [17]. In this method, all possible links are compared between the position of the object in the k-th frame and the candidate object positions in the k+1-th frame (Fig. 3b, circles). Then links are compared with their "costs" and "flow" in terms of the graph theory. Costs correspond to the distance between the points. Flow corresponds to the number of successful displacement of objects. The combination of links with a higher flow is selected first, and then the combination with smaller costs is selected as a more probable combination. In case of an example shown in the figure, the link between the successive positions is determined as in Fig. 3c, since the flow is maximized in this situation. While only two time points are considered in the example shown in the figure, an increased precision can be achieved by considering three consecutive time points, which is called the "three layered graph". In addition, the "costs" calculation can be more robust by considering the persistency of the movement and brightness of the objects as the following formula:

$$\begin{aligned} c = {} & \gamma_1 \left[1 - \frac{(\mathbf{r}_k - \mathbf{r}_{k-1}) \cdot (\mathbf{r}_{k+1} - \mathbf{r}_k)}{\|\mathbf{r}_k - \mathbf{r}_{k-1}\| \times \|\mathbf{r}_{k+1} - \mathbf{r}_k\|}\right] \\ & + \gamma_2 \left[1 - 2\frac{\sqrt{\|\mathbf{r}_k - \mathbf{r}_{k-1}\| \times \|\mathbf{r}_{k+1} - \mathbf{r}_k\|}}{\|\mathbf{r}_k - \mathbf{r}_{k-1}\| + \|\mathbf{r}_{k+1} - \mathbf{r}_k\|}\right] + \gamma_3 \left[\frac{\sigma_X(I)}{\bar{I}}\right] \end{aligned} \tag{8}$$

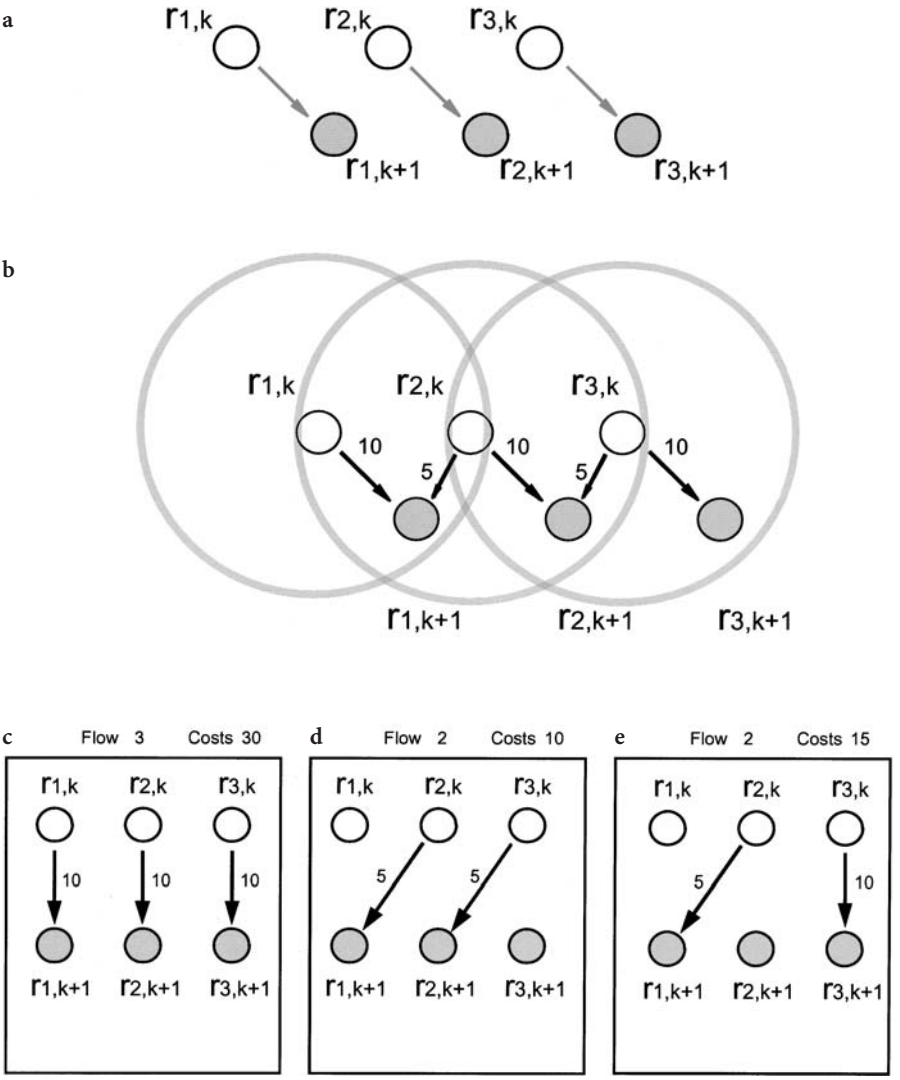


Fig. 3a–e Graph theory based approach for the recovering the trajectory: **a, b** the movement of three objects r1, r2 and r3 at the *k*-th image frame to the consecutive *k*+1-th frame are shown. *White circles* are the object positions in the *k*-th frame and *gray-filled circles* are the object positions in the *k*+1-th frame, *e.g.* $r_{1,k+1}$ is the position of object r1 in the *k*+1-th frame. The movement is towards the right-bottom and the displacement is identical for all of the objects. For finding a consecutive position in the *k*+1-th frame, a limited area (*gray circles*) surrounding each of the object in the *k*-th frame is searched. In case of $r_{1,k}$, there is only one object $r_{1,k+1}$ that is available. In case of $r_{2,k}$, two candidates, $r_{1,k+1}$ and $r_{2,k+1}$ exists. The object $r_{3,k}$ also has two candidates. Because there are multiple candidates, three combinations for linking the objects are possible, which are shown in **c–e**. In d and e, $r_{1,k}$ does not move. "Flow" and "costs" can be calculated for each condition and are also shown in the figure. The drawing is in accordance with Vallotton et al. (2003)

where $\mathbf{r}_n$ is the position vector at the n-th frame. $\sigma_X(I)$ is the standard deviation of the object brightness and $\bar{I}$ is the mean brightness over three frames. The first term evaluates the persistency of the direction and equals to zero if the direction of trajectory is same between k–1 to k+1. The second term evaluates the consistency of the speed and equals to zero if the overall speed is constant. The third term evaluates the consistency of the brightness of the spot and is zero if the brightness remains constant. γ_1, γ_2 and γ_3 are the weighting values for the each term. These values are tuned according to the characteristics of the object movement. Vallotton et al. (2003) used empirical values (γ_1, γ_2, γ_3)= (0.4, 0.6, 0.2). One disadvantage of this graph theory approach is that this method cannot cope with the disappearance of the object, which often occurs in the biological samples.

2.8 Active Contour

The active contour method, or Snakes, is an approach for simultaneous edge detection and object tracking [18]. In this technique, the user first places an elastic closed chain in the surrounding of the object to be tracked. This chain is the "active contour", which dynamically and automatically changes its shape until it detects the edge of the object. It works is as follows: Let the closed chain be represented as the function of a pair of parameterized coordinate $v(x(s),y(s))$. The contour will be stabilized by minimizing the equation

$$E_{snake} = \int E_{internal}(v(s))\,ds + \int E_{external}(v(s))\,ds \tag{9}$$

where $E_{internal}$ is the internal energy, which determines the tension and rigidity of the chain. $E_{external}$ is the external force that balances against the internal force. The main component of $E_{external}$ is the image force E_{image}, which is derived from the pattern of the image intensity distribution where each segment of the chain resides. Appropriate operators can be applied to the image so that E_{image} becomes lowest at the edge of the object. A typical definition of E_{image} uses the image gradient in such a way that the energy is lowest at the object edge:

$$E_{image} = -\left|\frac{\partial I(x,y)}{\partial x} + \frac{\partial I(x,y)}{\partial y}\right|^2 = -|\nabla I|^2 \tag{10}$$

where ∇I is the gradient of image intensity (Fig. 4). The active contour dynamically changes its shape and stabilizes at the equilibrium between internal and external forces. To lock the active contour at the edge of the object more efficiently and precisely, various other types of E_{image} have been developed (cf. [19–23]). Ray et al. introduced additional terms such as shape or size constraints to tune the behavior of the active contour for tracking the leukocyte movement [24]. A demonstration of the active contour technique can be found on the internet[3].

[3] http://www.markschulze.net/snakes/.

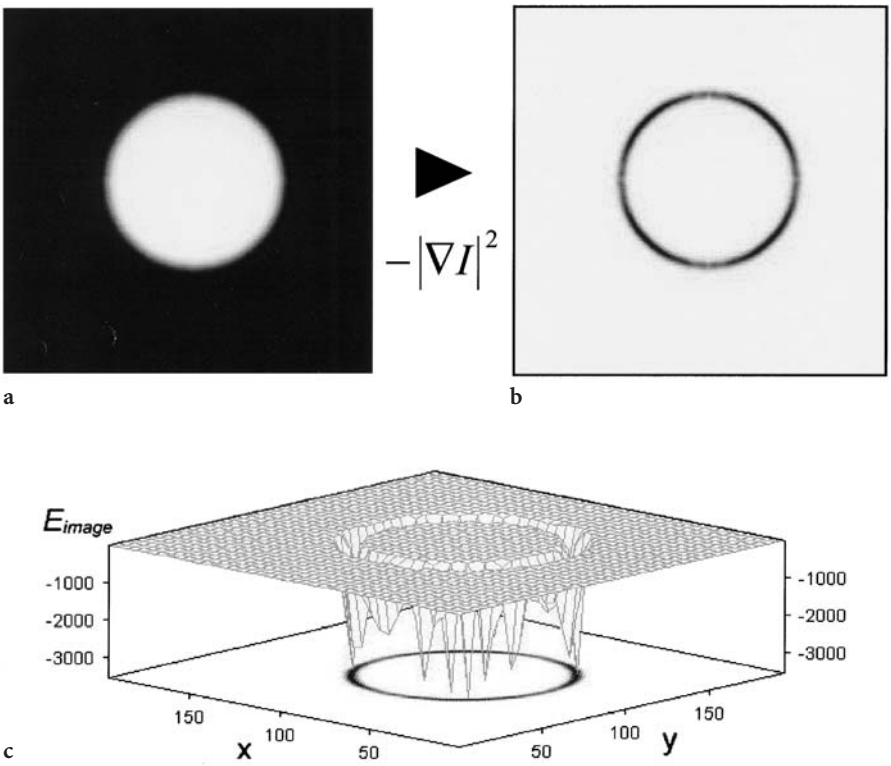


Fig. 4a–c An example of Image Force: **a** original image; **b** its image force (squared- absolute gradient: Eq. 10). The image force becomes lowest at the edge of the circular object shown in (a); **c** as an active contour dynamically searches for a position with a lowest energy (Eq. 9), the active contour locks on the object edge where E_{image} forms a valley

The stabilized contour in the n-th frame can be inherited to the n+1-th image frame, thus enabling the processive detection of the object edge during displacement. Since the active contour is highly flexible, this method is advantageous for tracking a combined movement of both displacement and deformation. Single cells migration is a typical example of such a combined movement. Active contour based edge detection technique is tolerant to shape changes of the target object. The user needs, however, to assign the initial position and size of the "chain", so that the method cannot be fully automatic for multiple objects measurement. Besides, active contour will not follow the movement of the object if the object moves more than half the diameter of the object itself between two frames.

As a further application to correlate the intracellular signaling and pseudopod protrusion dynamics, the active contour-based measurement of the cell surface activity was conjugated with the simultaneous measurement of GFP-labeled protein dynamics near the cell edge [25].

A major problem appears when trajectories of objects are crossing to each other since determination of single trajectory becomes difficult. Especially when portion of the object edge overlaps, ambiguity of the boundary prevents segmentation. The probability of the crossing events can be lowered by preparing low-density samples. Increasing the temporal resolution is another possibility. Another solution could be modifications to the algorithm itself. Zimmer et al. (2002) developed an active contour algorithm with repulsion force. When two active contours come close together, the contours repulse each other so in a way that the individual contour is maintained for the each object [26].

3 Optical Flow Estimation

Another way of measuring movement is what is generally called the optical flow estimation. Optical flow is "the distribution of apparent velocities of movement of brightness patterns in an image" [27]. In video sequences the projection of temporal axis to the *x*-*y* plane results in an optical flow image. Since the optical flow is a result of the movement, it contains information on movement speed and direction. The optical flow estimation recovers these quantitative measures of the movement, which enables the statistical treatments of all movement that occurs in the sequence. A velocity vector field is the calculation result in which every movement, namely speed and direction, within the sequence is mapped (see Fig. 7a,b). The largest difference compared to the other tracking technique is that this operation does not require the segmentation step.

Optical flow detection can be categorized into two types in terms of basic algorithm: the "matching method" and the "gradient method". In the matching method, displacement is measured by searching for a particular region in the consecutive frame by matching the pattern of the previous frame. In the gradient method, optical flow is detected by assuming no changes in the signal intensity pattern at different time points and by using equations that correlate the spatial and temporal intensity gradient.

3.1 Pattern Matching – Global Matching Methods

The principle of the matching method is same as the pattern matching algorithm explained already (see above), but the pattern matching is tested more globally within the image frame. The template kernels are sampled from every region in the frame. Then the next frame is compared with each of those template kernels to search for a best matching position. The area for the evaluation of the pattern similarity in the next frame can be restricted to a limited region surrounding the original position of the kernel to save time and to increase the precision.

Let's take as an example a 5×5 pixels template kernel. The first template kernel is sampled from the top-left corner of *n*-th image frame. Then the

matching of this template kernel to the shifted positions within the n+1-th image frame is evaluated. After the evaluation is finished, the second template kernel can be sampled from n-th image frame at the position next to the first frame, five pixels shifted in either the x or y direction. The size of template kernel size can be varied. A smaller kernel detects matching of the detailed features and intensity patterns but larger features and patterns would be omitted. Various correlation evaluation algorithms have been designed and comparison of their accuracy can be found elsewhere [28].

When the matching region is found, displacement of that pattern between the frames is obtained. The velocity and the direction of the movement are calculated from the displacement. If the movement is purely translational and does not change shape, a specific position that matches with the template kernel could easily be detected.

Breen and Williams used the matching method to measure the cell movement velocity within the multicellular structure of *Dictyostelium discoideum* [29]. They used the sum of the absolute difference (SAD) for the evaluation of the pattern matching. The sum of absolute difference is

$$SAD(x,y) = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} \left| I(x+i, y+j) - K(i,j) \right| \tag{11}$$

$SAD(x,y)$ is minimized when the image pattern matches with the overlapping portion of the image. Zoccolan et al. measured the extent of deformation of leech skin by the optical flow estimation based on SAD [30–32]. McNally and colleagues have extended the matching method for analyzing three-dimensional cell movement within multicellular structure of *Dictyostelium* [33] and revealed different modes of cell movement in wild type cells and myosin-II null mutant cells during the multicellular phase [34, 35]. An application of the pattern matching technique for a 3D sequence is reported [36].

3.2 Gradient Methods

Whereas the pattern matching method has been the major algorithm used in cell biology, optical flow estimation has another type of algorithm more generally used in broader areas. It is called the "gradient method" and relies on image differentials. The gradient method algorithm has advantages over the pattern matching method. The measurement is robust, and unlike the pattern matching algorithm, the gradient method can be more tolerant to the deformation of the moving object. Preconditioning such as the determination of threshold value or setting appropriate combination of filters to leave out the target objects is not prerequisite. Detailed descriptions of different techniques for the gradient method can be found elsewhere [37-40].

3.2.1 The Optical Flow Equation

The general assumption in the gradient method is that the total intensity of the image sequence is constant:

$$\frac{\partial I(x,y,t)}{\partial x}u + \frac{\partial I(x,y,t)}{\partial y}v + \frac{\partial I(x,y,t)}{\partial t} = 0 \tag{12}$$

$$\mathbf{v}\,(u,v) = \mathbf{v}\left(\frac{dx}{dt}, \frac{dy}{dt}\right) \tag{13}$$

where $I(x,y,t)$ is the intensity distribution of the image and $\mathbf{v}$ is the optical flow vector. Equation (12) links the partial derivatives of the brightness pattern of the image sequence and the optical flow velocity. Since there are two unknowns, another constraint is required. Various constraints have been proposed for obtaining the optical flow vectors.

3.2.2 Spatial Global Optimization Method (SGO)

If speed and direction of object movement vary smoothly along the space, the smoothness of the vector field can be set as the second constraint. This method is called the spatial global optimization (SGO). The spatial smoothness of the optical flow can be assumed as follows [27]:

$$\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial v}{\partial x}\right)^2 + \left(\frac{\partial v}{\partial y}\right)^2 \rightarrow \min \tag{14}$$

Then an error function $E(x, y)$ can be formulated as the difference between Eqs. (12) and (14). $\mathbf{v}(u, v)$ can be estimated that minimizes the error function $E(x,y)$. This constraint is only valid when the movement of objects is roughly in a similar direction.

3.2.3 Spatial Local Optimization Method (SLO)

Instead of taking the whole vector field into account, a constant optical flow within a local region can be assumed [41, 42]. Let us consider a local region S and assume that all the vectors are identical within this area (Fig. 5):

$$\frac{\partial \mathbf{v}}{\partial x} = \frac{\partial \mathbf{v}}{\partial y} = 0\big|_{x,y \in S} \tag{15}$$

This method is called the spatial local optimization (SLO). The SLO constraint assumes no shape change of the object between consecutive frames. The opti-

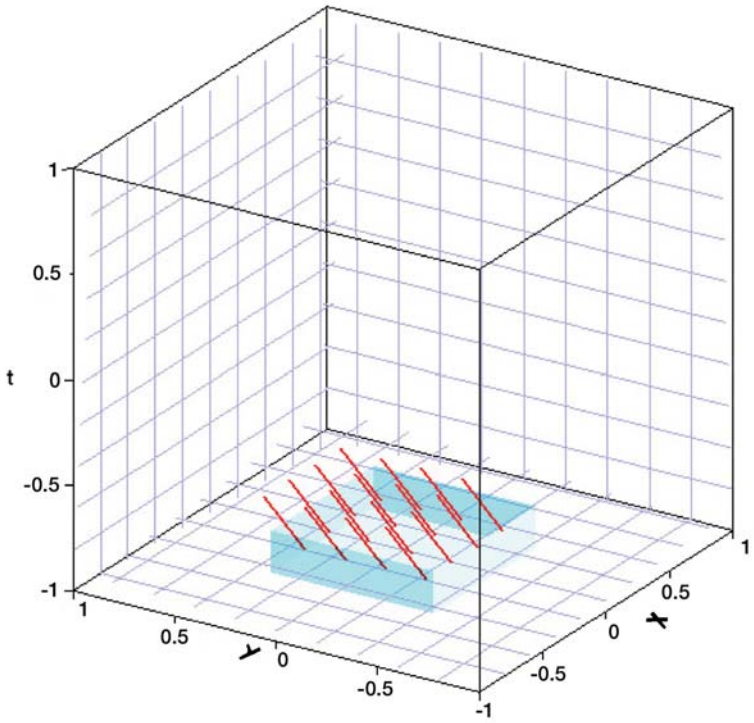


Fig. 5 Constraints of the optical flow estimation – the spatial local optimization. An image sequence is shown as a three-dimensional matrix $I(x,y,t)$. *Red bars* are the displacement vectors. In SLO approach, we assume that displacement vectors in a small region in X-Y plane (*light-blue block*) are all same. This then becomes the second constraint for the optical flow estimation (see Eq. 13)

cal flow can be estimated by the least squares method using Eqs. (12) and (15). Since we assume that the solution for the gradient equations within the local region S is the same, the error of the vector field within S at the time point k is

$$E = \sum_i \sum_j (I_x(i,j,k)\,u + I_y(i,j,k)\,v + I_t(i,j,k))^2 \tag{16}$$

Using the least squares method, $\boldsymbol{v}(u,v)$ can be determined by two equations $\partial E/\partial u=0$ and $\partial E/\partial v=0$. Instantaneous optical flow can be estimated from a small number of image frames.

Although the SLO method is a standard approach, it assumes a rigid object motion that is a rare case in biology. To relax this strong restriction on the shape change, Cai et al. invoked the flexible and flat nature of biological skin and formulated the in-plane incompressibility constraint within the local region [43]:

$$\frac{\partial \mathbf{v}}{\partial x} + \frac{\partial \mathbf{v}}{\partial y} = 0 \tag{17}$$

which allows the deformation of the object assuming the total brightness of the object is constant.

3.2.4 Constraints Involving the Temporal Domain

The optical flow estimation technique was initially developed for the artificial vision of robots. For this reason, the estimation of the optical flow was designed to be fast, from only two frames of a sequence. The calculation speed was important in these cases. After the optical flow estimation was applied for high-precision measurements of the movement, the consideration of the temporal domain, namely the utilization of multiple frames in the sequence, became more important.

3.2.4.1 Temporal local optimization method (TLO)

In contrast to the SGO and SLO methods, the temporal local optimization method (TLO) assumes that the optical flow field is constant temporally (Fig. 6) [44]:

$$\frac{\partial \mathbf{v}}{\partial t} = 0 \tag{18}$$

The constant vector can be assumed for N frames. Then for a stack of frames $0 \leq k < N$, following error function can be generated:

$$E = \sum_k \left(I_x(i,j,k)\,u + I_y(i,j,k)\,v + I_t(i,j,k)\right)^2 \tag{19}$$

By the least squares method, $\boldsymbol{v}(u,v)$ can be determined by two equations $\partial E/\partial u = 0$ and $\partial E/\partial v = 0$. The constraint Eq. (18) does not assume the spatial smoothness of the vectors. For this reason, TLO method does not depend on the constant object shape unlike SGO and SLO methods. In addition, velocity vectors at each pixel can be independently determined that the resulting optical flow estimation has a higher spatial resolution. TLO method was applied to the analysis of cell movement of *Dictyostelium discoideum* to study the cell movement pattern within multicellular structure and revealed the geometry of cell-to-cell signaling waves that coordinate the movement of tens of thousands of cells [45, 46].

In addition to the above constraints, the second-order techniques [47, 48], the energy-based method [49] and the phase-based method [50] has been proposed for the optical flow estimation. These techniques are yet to be applied in cell biology. Quantitative comparison of the algorithm among Horn and Schunck, Lucas and Kanade, Uras et al., Nagel, Anadan, Singh, Heeger and Fleet methods showed that the Lucas and Kanade's method, an SLO method, has the highest precision [38]. All these constraints examined in Barron's paper are included

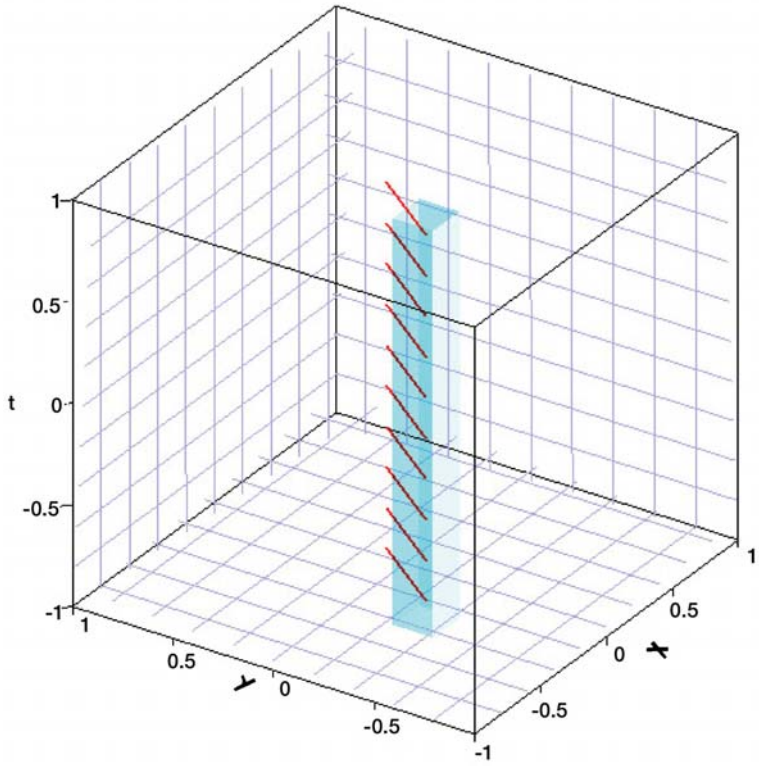


Fig. 6 Constraints of the optical flow estimation – the temporal local optimization. An image sequence is shown as a three-dimensional matrix $I(x,y,t)$. *Red bars* are the displacement vectors. In TLO approach, we assume that displacement vectors in a small region along the time-axis (*light-blue block*) are all same; the second constraint for the optical flow estimation (see Eq. 18)

as a plugin in ImageJ and are readily available for the public use (see ImageJ website [2]). Although the TLO method has been a minor approach, the comparison between SLO and TLO methods showed that TLO has a higher precision for measuring speed and direction of moving objects [44].

3.2.4.2 Structure Tensor Method

The SLO approach has been extended to the structure tensor model approach. This approach considers the temporal domain in addition to the spatial smoothness by treating the tensor field of the matrix $I(x,y,t)$ [40]. The trajectory of moving objects appears as a structure that penetrates the $I(x,y,t)$ matrix that is inclined against the t axis. The structural tensor of this trajectory is subjected to the eigenvalue analysis for obtaining the vector of the brightest part of the trajectory. This approach updated the minimization procedure from

the least minimum squared method in SLO to a more robust total least squares method [40]. Sliding velocity of fluorescently labeled actin filament was estimated by the structure tensor approach [51]. In this study, the motion velocity was estimated by the peak position of the velocity histogram of the vectors. Leave growth kinetics was also measured using this algorithm [52–54].

3.2.5 Combination of Different Techniques

Several recent reports combine different approaches to merit different advantages. The active contour technique was combined with the gradient method based optical flow estimation to add directional bias to the external force $E_{external}$ to follow the successive cell positions that is apart from the previous position [24]. In another case, the gradient method was combined with the pattern matching method also to add directional bias to the estimate the optical flow to measure the plant root growth [55]. Furthermore, the cross-correlation based pattern matching method was used to subtract the background movement of the tissue from the apparent cell movement to leave the real cell movement [56]. These reports indicate that the measurement becomes more precise with the custom combination of multiple methods depending on the nature of the target movement.

4 Choosing a Tracking Method: Which is the Best?

Each researcher must select one of the tracking methods for their interest. How should one choose a method out of all these different ways? The answer is not simple. "The best" depends on the hardware setup and the nature of the movement of the object. Table 1 summarizes advantages and disadvantages of each method as a guide line.

The manual tracking and the thresholding could be the first step for tracking any object. The properties of the image of object can be roughly grasped by the researcher through these initial trials. Then the researcher can refine the choice of tracking method more easily. Most of tracking methods are not included as a standard function in conventional image analysis software. To use these methods, the researcher must code a tracking program. Characteristic of the object movement and image properties understood during the initial trial by the manual tracking and the thresholding ease the coding of the tracking program as well.

Examination of following characters of the object is helpful for choosing a tracking method. The first point is whether the object alters its shape while it is changing its position. In methods such as pattern matching algorithm, the shape change causes pattern mismatching that the object tracking becomes unsuccessful. The Gaussian fitting algorithm also fails tracking the shape-

Table 1 Advantages and disadvantages of different methods for tracking movement

Method	Advantages	Disadvantages	Measurement suggestion
Manual tracking	– Flexible application – No limit in the object displacement distance per frame if the user is confident for their correlation	– Requires intensive work due to the manual segmentation (manual drawings) – Requires registration	Any
Thresholding	– Flexible application – The function is equipped in most of the imaging applications	– Correctness of the segmentation depends on the image type – Very sensitive to the image brightness change – Requires registration	Optically flat objects
Gaussian fitting	– Very low S/N (4<) – The nano-meter precision – Requires registration	– Only symmetric objects (spherical)	Particle tracking of protein attached to a bead
Active contour	– Shape changes of the object can be precisely measured – Suited for objects with a combination of displacement and large shape changes	– User must initially define the position of the "active contour" – Requires registration when successive positions are apart (typically, more than a diameter of the initial active contour) – Parameter settings require a skilled experience	Dynamics of shape changes Cell migration dynamics
Pattern matching (local)	– Low S/N – Displacement of the object in successive point is only limited by search area for the pattern matching	– Requires registration – Shape-changing object is unfavourable	Movement of objects with constant shape and features (movement of constant-shape organelle)

Table 1 (continued)

Method	Advantages	Disadvantages	Measurement suggestion
Optical flow estimation (pattern matching – global)	– No registration required – All movement in the sequence is measured – Requires only two frames from the sequence	– Change in the feature is assumed to be smooth over the sequence – Shape-changing object is unfavourable – Intensive calculation – Sensitive to the image brightness change	Dynamics of changes in pattern of rigid structures, such as microtubule orientation dynamics; tissue deformation.
Optical flow estimation Gradient method – SGO/SLO	– No registration required – All movement in the sequence is measured – Requires only two frames from the sequence – Less calculation than TLO	– Shape-changing object is unfavourable – Measurable displacement per frame is limited by algorithm for gradient calculation – Sensitive to the image brightness change	Movement of objects with constant shape and features (movement of constant-shape organelle)
Optical flow estimation Gradient method – TLO	– No registration required – All movement in the sequence is measured – Low S/N – High resolution vector field – Precision higher than SGO/SLO	– Requires multiple frames (3<). – Constant speed of the object is assumed – Measurable displacement per frame is limited by algorithm for gradient calculation – Sensitive to the image brightness change	Movement of structurally unstable intracellular structure The cell movement within tissue

changing object. The second point is the movement distance of the object per frame. If this distance is too large, the user must consider either increasing the capture rate or giving up the automatic tracking of object. The third point is the density of the objects. If there are many objects with overlapping tracks, most of the algorithm fails in recovering the true trajectory of objects. In this case, user must decrease the density of the object in experiments or increase the capture rate to avoid the crossing of the objects. Use of the optical flow estimation by the gradient approach is another solution.

In general, the automatic tracking works better when the time resolution of the movement is higher. The object displacement per frame then becomes smaller that the trajectory recovery becomes easier. Increasing the time resolution, however, has a drawback. It decreases the signal-to-noise ratio (S/N ratio) especially in the case of fluorescence imaging. Decreased S/N ratio lowers the precision of segmentation. Even with algorithms that do not depend on segmentation such as the optical flow estimation methods, lower S/N causes more erroneous measurement due to either incorrect pattern matching or the deviation of the derivatives from the real value.

An optimum time resolution of the image sequence can be found by balancing the frame rate and the S/N ratio. The optimum time resolution is thus determined by the following parameters:

1. Hardware setup: Magnification of the object lens, maximum speed of the capture rate and the quality of the CCD camera. High quality setup enables higher capture rates with higher S/N ratio.
2. The brightness and optical characteristics of the object determines the possible capture rate.
3. Objects that moves at a higher speed require a higher capture rate. Higher capture rate could limit the exposure time in case of fluorescence imaging, which then causes less contrast images and a low S/N ratio.

For these reasons the optimum time resolution must be determined for each of different combinations of the hardware setups and the optical and the movement characteristics of the object.

5 Analysis of Protein Movement within Cultured Cells

An example of the optical flow estimation of protein movement within cultured cells using the gradient method is described in the following. The TLO method (Eq. 18) was used [44, 46]. The program was written in IgorPro 4.08 (Wavemetrics Inc. OR, USA) programming language. The pure vector field estimated from video sequences acquired through fluorescence microscopy contains many vectors due to noise in the background. To remove these unwanted vectors, several filtering was applied to the resulting vector field to eliminate noise-derived vectors. The evaluation of the measurement precision will be presented elsewhere

[57]. I analyzed the movement of a microtubule binding protein eb1 labeled with green fluorescent protein in Vero cell. eb1 is thought to be involved in controlling the stability of microtubule. The sequence was taken using a spinning disk confocal microscope (UltraVIEW, Perkin-Elmer, UK) at two frames per minute.

A series of tiff-format image frames was converted to a stack, which was then treated as a three-dimensional matrix $I(x,y,t)$. Partial derivatives of the image sequence must be first calculated as seen in Eqs. (12) and (18). There are three popular ways for obtaining the first-order partial derivatives[4]: Sobel kernel, Roberts kernel and Prewitt kernel (also is called two-point central difference kernel). The Prewitt kernel was used as follows:

$$\frac{\partial I}{\partial t} = \left[\sum_{i=-1}^{1} \sum_{j=-1}^{1} \{I(x+i, y+j, t+1) - I(x+i, y+j, t-1)\}/2 \right] /9 \tag{20a}$$

$$\frac{\partial I}{\partial x} = \left[\sum_{j=-1}^{1} \sum_{k=-1}^{1} \{I(x+1, y+j, t+k) - I(x-1, y+j, t+k)\}/2 \right] /9 \tag{20b}$$

$$\frac{\partial I}{\partial y} = \left[\sum_{i=-1}^{1} \sum_{k=-1}^{1} \{I(x+i, y+1, t+k) - I(x+i, y-1, t+k)\}/2 \right] /9 \tag{20c}$$

I used 40 frames to obtain a velocity vector field for 20 min (Fig. 7a). As many as 40 frames were used because the required number of frames for a high precision measurement is determined to be between 30 and 50 frames [57]. The vector field consists of one vector for each single pixel in the TLO method. The plotting of all these vectors results in a plane field covered only with vectors. Thus for the visualization purpose, 25 vectors within each of 5×5 pixel kernel were averaged to represent a single vector for that kernel. The statistics of speed and direction (see below), however, were derived from the original vector field. Grid points (blue circles) located at the center of each kernel are the origin of each vector (see magnified view of the vector field at Fig. 7b.)

The length of the vector is the speed of the movement. The orientation of the vector is the movement direction. Figure 7c,d shows histograms of velocity and direction obtained from the all the vectors within the vector field (Fig. 7a). The direction of the movement is measured as an angle of the vectors against a line auxiliary drawn between the origin of that vector and the reference point (blue cross). The histogram shows that the movement is mostly outward, as the peak position is at ±180° (Fig. 7d).

The vector field contains apparent vectors due to the intensity changes in the background coming from noise. Apparent vectors also appear at the boundary of a high intensity region and low intensity region. This is because small fluctuation in the intensity at the high-contrast boundary causes large changes in the gradient and result a large errors. To avoid errors of this type, I further

[4] http://supportweb.cs.bham.ac.uk/documentation/visilog5/html/refguide/chap8.html.

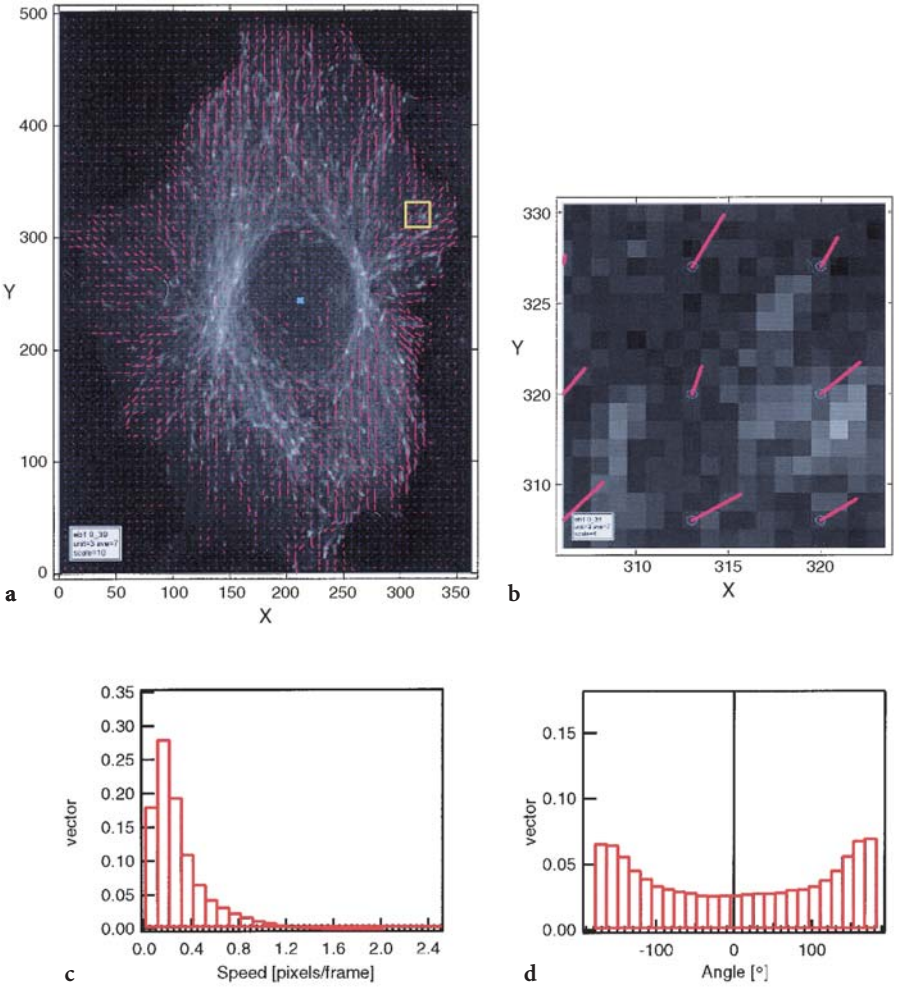


Fig. 7a–d The optical flow estimation of the eb1 movement and its analysis: **a** a vector field image of the eb1 movement. An image sequence of the fluorescently labeled protein eb1 in a Vero cell was processed and the vector field was obtained. *Small red bars* are the vectors. The first frame of the image sequence is shown in the background. The *light-blue cross* at the center of the cell is a reference point for the measurement of vector orientation (see d); **b** a magnified view of the region indicated by yellow square in (a). *Red bars* are vectors each at the position indicated by *blue circle*; **c** histogram of the magnitude of the vectors in (a). This corresponds to the vector distribution of velocity of the eb1 movement in the Vero cell; **d** histogram of the vector directionality. Direction was defined as the angle against the reference point (light-blue cross) shown in (a). 0° is the movement straight towards the reference point, the cell center. eb1 was mostly moving away from the reference points (±180°)

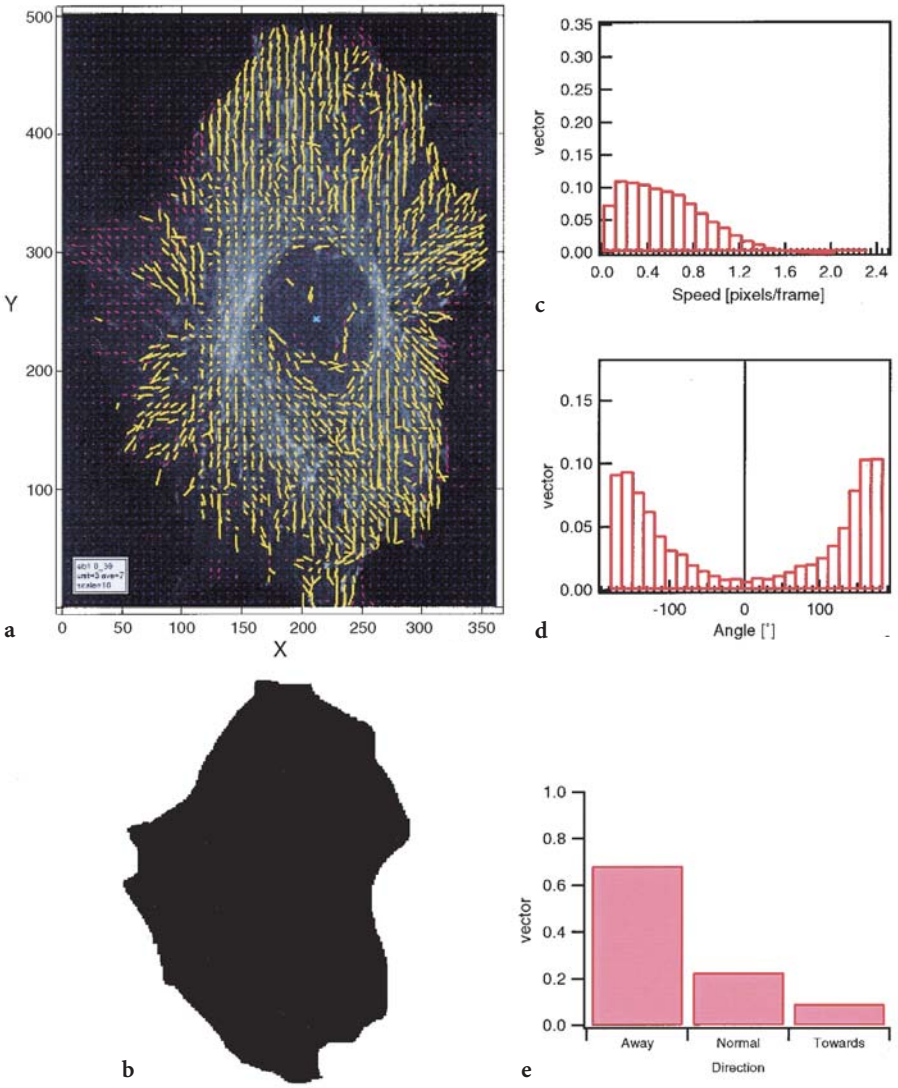


Fig. 8a–e Removal of noise derived vectors: image masking and intensity filtering: **a** the vector field after the image masking and intensity filtering of the vector field shown in Fig. 7. *Red bars* are vectors before filtering and *yellow bars* are the vectors after the filtering; **b** a mask used for the image masking. *White regions* were masked; **c** histogram of speed. The peak broadened to higher speed due to the removal of noise- derived apparent vectors; **d** histogram of direction. Removal of the noise-derived vectors clarified the outward movement of eb1 protein; **e** the directionality was categorized into three: Away (±120–180°), Normal (±60–120°) and Towards (±0–60°) the cell center. About 70% of the movement of signal detected in the image sequence was directed away from the cell center

filtered the vector field to isolate only the true vectors originated from the movement of the protein.

The vector field can be filtered by an image masking filter and an intensity filter (Fig. 8). The *image masking filter* masks the vectors outside the cell in the background (Fig. 8b). At the same time, regions with a high-contrast edge can also be masked to eliminate the erroneous vectors generated at these positions (not shown in Fig. 8b, for this sequence did not contain erroneous high-contrast edges). *Intensity filtering* sets a range of pixel intensity that corresponds to the signal of our interest and isolates only the vectors from those pixels. Since I was only interested in the movement of bright intensity signals but not the dark background where the protein was absent, only the vectors from pixels with a high intensity was isolated by this filtering (Fig. 8a). The intensity filtering largely eliminates the vectors coming from noise, which is evident in the histograms of speed and direction. As a result, the distribution of the speed shifted to a higher value (Fig. 8c). The outward eb1 protein movement became clearer, as the peak at ±180° became sharper in Fig. 8d than in Fig. 7d. To simplify these results, the histogram in Fig. 8d was categorized into three different movement directions against the cell center: Away (±120~180°), Normal (±60~120°) and Towards (±0~60°). In this case, about 70% of the movement

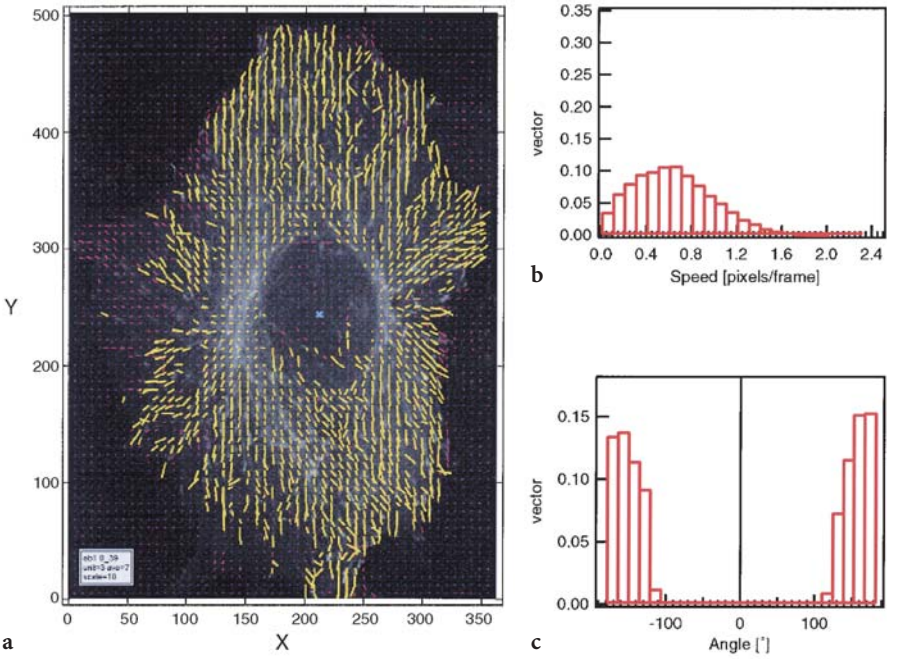


Fig. 9a–c Direction filtering: **a** the vector field after the direction filtering (*in yellow*), showing only the vectors directed away (±120~180°) from the cell center; **b** histogram of speed; **c** histogram of direction. Note that the vector is distributed only within ±120~180° due to the filtering

of the signal detected in the image sequence was directed away from the cell center (Fig. 8e).

To analyze the speed of the movement in a specific direction, direction filtering can be applied. The eb1 movement was outward from the cell center (Fig. 8e). To analyze the speed of this outward movement, vectors directed Away (±120~180°) from the cell center were isolated (Fig. 9a). A clear peak in the histogram of speed appeared after the direction filtering (Fig. 9b).

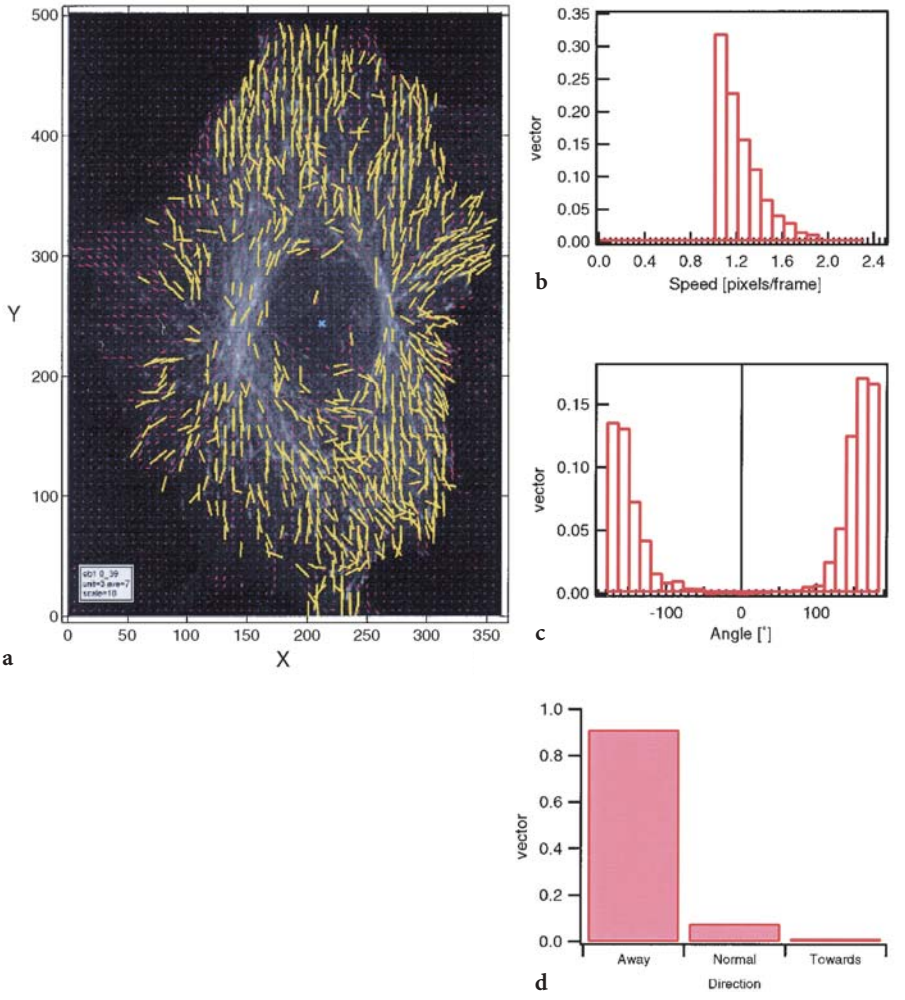


Fig. 10a–d Speed filtering: **a** the vector field after the speed filtering (*in yellow*), showing only the vectors within a range of speed (1.0~2.0 pixels/frame); **b** histogram of speed. Note that the vector is distributed within 1.0~2.0 pixels/frame due to the filtering; **c** histogram of direction; **d** three-direction category histogram shows that the higher speed population was moving away from the cell center, the directionality much clearer than Fig. 8d

Table 2 A summary of eb1 movement analysis

	Speed [pixels/frame] ±s.d.	Mean of the direction [°] ±confidence interval	Pearson's coefficient *r* of directionality
Original vector field	0.30±0.26	−179±1	0.23
Image masking and intensity filtering	0.55±0.36	−177±1	0.53
Direction filtering	0.64±0.36	–	–
Speed filtering	–	−176±0	0.81

Conversely, it is also possible to analyze the directionality of a population moving at a high speed. In Fig. 10a, only the vectors with a speed between 1 and 2 pixels per frame are shown (see also Fig. 10b). The movement direction of this high speed population was clearly outward, as it is evident in the histogram shown in Fig. 10c,d. Table 2 is the summary of these filtering and analyses. Mean value of the movement direction was calculated by von Mises circular statistics [58]. Pearson's coefficient *r* is a concentration parameter for the directionality. *r* is 0 when the direction is randomly oriented and 1 when the movement is confined in a single direction. Increase in the precision of the measurement is reflected in an increase in the Pearson's coefficient *r*.

When the outward movement of eb1 protein is measured by manual tracking, average velocity of the eb1 structure moving along the microtubule was 1.57±1.07 pixels/frame (n=16). The speed measured by the optical flow estimation is smaller, 0.64±0.36 pixels/frame, about a half of that by the manual tracking. This is most likely because in case of the optical flow estimation, all movements are measured including those which we usually do not regard significance, such as signals with ambiguous shapes or those that moves only a short distance. The lower value of the movement speed is thus the result of a more objective approach for measuring the movement.

Acknowledgement I thank Emmanuel Reynaud (EMBL Heidelberg) for providing the sample sequence of eb1. I am grateful to Atsushi Nomura (Yamaguchi University) for an introduction to the gradient approach for the optical flow estimation.

References

1. Eils R, Athale C (2003) J Cell Biol 161:477
2. Cheezum MK, Walker WF, Guilford WH (2001) Biophys J 81:2378
3. Schütz GJ, Schindler H, Schmidt T (1997) Biophys J 73:1073
4. Anderson CM, Georgiou GN, Morrison IE, Stevenson GV, Cherry RJ (1992) J Cell Sci 101(2):415
5. Tardin C, Cognet L, Bats C, Lounis B, Choquet D (2003) Embo J 22:4656

6. Gelles J, Schnapp BJ, Sheetz MP (1988) Nature 331:450
7. Geerts H, De Brabander M, Nuydens R, Geuens S, Moeremans M, De Mey J, Hollenbeck P (1987) Biophys J 52:775
8. Schnapp BJ, Gelles J, Sheetz MP (1988) Cell Motil Cytoskeleton 10:47
9. Sheetz MP, Turney S, Qian H, Elson EL (1989) Nature 340:284
10. Qian H, Sheetz MP, Elson EL (1991) Biophys J 60:910
11. Saxton MJ, Jacobson K (1997) Annu Rev Biophys Biomol Struct 26:373
12. Ghosh RN, Webb WW (1994) Biophys J 66:1301
13. Carlsson AE, Shah AD, Elking D, Karpova TS, Cooper JA (2002) Biophys J 82:2333
14. Ritchie K, Kusumi A (2003) Methods Enzymol 360:618
15. Tseng Y, Kole TP, Wirtz D (2002) Biophys J 83:3162
16. Apgar J, Tseng Y, Fedorov E, Herwig MB, Almo SC, Wirtz D (2000) Biophys J 79:1095
17. Vallotton P, Ponti A, Waterman-Storer CM, Salmon ED, Danuser G (2003) Biophys J 85:1289
18. Kass M, Witkin A, Terzopoulos D (1987) Int J Computer Vision 1:321
19. Xu C, Prince JL (1998) IEEE Trans Image Process 7:359
20. Xu C, Prince JL (1998) Signal Process 71:131
21. Geiger D, Gupta A, Costa LA, Vlontzos J (1995) IEEE Trans Pattern Anal Machine Intell 17:294
22. Leymarie F, Levine MD (1993) IEEE Trans Pattern Anal Machine Intell 15:617
23. Cohen LD, Cohen I (1993) IEEE Trans Pattern Anal Machine Intell 15:1131
24. Ray N, Acton ST, Ley K (2002) IEEE Trans Med Imaging 21:1222
25. Dormann D, Libotte T, Weijer CJ, Bretschneider T (2002) Cell Motil Cytoskeleton 52:221
26. Zimmer C, Labruyere E, Meas-Yedid V, Guillen N, Olivo-Marin JC (2002) IEEE Trans Med Imaging 21:1212
27. Horn BKP, Schunck BG (1981) Artif Intell 17:185
28. Giachetti A (2000) IVC 18:247
29. Breen EJ, Williams KL (1994) Microbiol UK 140:1241
30. Zoccolan D, Pinato G, Torre V (2002) J Neurosci 22:10790
31. Zoccolan D, Giachetti A, Torre V (2001) J Neurosci Methods 110:65
32. Zoccolan D, Torre V (2002) J Neurosci 22:2283
33. Awasthi V, Doolittle K, Parulkar G, McNally J (1994) Bioimaging 2:98
34. Clow PA, McNally JG (1999) Mol Biol Cell 10:1309
35. Doolittle KW, Reddy I, McNally JG (1995) Dev Biol 167:118
36. Bornfleth H, Edelmann P, Zink D, Cremer T, Cremer C (1999) Biophys J 77:2871
37. Teklap M (1995) Digital video processing. Prentice Hall, Englewood Cliffs, N.J.
38. Barron JL, Fleet DJ, Beauchemin SS (1994) Int J Comput Vis 12:43
39. Beauchemin SS, Barron JL (1995) ACM Comp Surv 27:433
40. Haussecker H, Spies H (1999) In: Jähne B, Haussecker H, Geissler P (eds) Handbook of computer vision and applications, vol 2. Academic Press, San Diego, p 310
41. Kearney JK, Thompson WB, Boley DL (1987) IEEE Trans Pattern Anal Mach Intell 9:229
42. Lucas BD, Kanade T (1981) DARPA Proceedings of Imaging Understanding Workshop, p 121
43. Cai H, Richter CP, Chadwick RS (2003) Biophys J 85:1929
44. Nomura A, Miike H, Koga K (1991) Pattern Recog Lett 12:183
45. Abe T, Early A, Siegert F, Weijer C, Williams J (1994) Cell 77:687
46. Siegert F, Weijer CJ, Nomura A, Miike H (1994) J Cell Sci 107:97
47. Nagel HH (1983) Comput Vision Graphics Image Process 21:85
48. Uras S, Girosi F, Verri A, Torre V (1988) Biol Cybern 60:79
49. Heeger D (1988) Int J Comput Vis 1:279
50. Fleet D, Jepson A (1990) Int J Comput Vis 5:77

51. Uttenweiler D, Veigel C, Steubing R, Gotz C, Mann S, Haussecker H, Jahne B, Fink RH (2000) Biophys J 78:2709
52. Barron JL, Liptay L (1997) Bioimaging 5:82
53. Walter A, Spies H, Terjung S, Küsters R, Kirchgeßner N, Schurr U (2002) J Exp Bot 53:689
54. Schmundt D, Stitt M, Jähne B, Schurr U (1998) Plant J 16:505
55. van der Weele CM, Jiang HS, Palaniappan KK, Ivanov VB, Palaniappan K, Baskin TI (2003) Plant Physiol 132:1138
56. Acton ST, Wethmar K, Ley K (2002) Microvasc Res 63:139
57. Miura K, Pepperkok R (in preparation)
58. Zar JH (1999) Biostatistical analysis. Prentice Hall, Upper Saddle River, NJ

Author Index Volumes 51 – 95

Author Index Volumes 1 – 50 see Volume 50

Subject Index