Anal. Chem. **2000,** *72,* 443−447

Single-Molecule Identification by Spectrally and Time-Resolved Fluorescence Detection

Michael Prummer, Christian G. Hübner,* Beate Sick, Bert Hecht, Alois Renn, and Urs P. Wild

Eidgenössiche Technische Hochschule Zürich, CH-8092 Zürich, Switzerland

A method to identify single molecules rapidly and with high efficiency based on simple probability considerations is proposed. In principle, any property of a detected photon in a single-molecule fluorescence experiment, e.g., emission wavelength, arrival time after pulsed excitation, and polarization, can be analyzed within the framework of the outlined methodology. Monte Carlo simulations show that less than 500 photons are needed to assign an observed single molecule to one out of four species with a confidence level higher than 99.9%. We show that single dye molecules of four different dyes embedded in a polymer film can be identified with time-correlated single-photon counting spectrally resolved in two channels.

In recent years, the possibility to detect fluorescence from single molecules at ambient temperatures has initiated several attempts to identify and distinguish individual molecules in solution.[1−8] One of the most prominent future applications of such a technique is DNA sequencing by labeling individual bases with different fluorescent dyes.[9] Single-molecule sequencing offers the advantage that error sources inherent to PCR amplification can be eliminated. It offers unique precision and speed for the enormous task of, for example, the Human Genome Project as well as for clinical applications.

Different properties of fluorescence photons, namely, spectrum resolved in two channels,[1] fluorescence lifetime,[6] and fluorescence lifetime in conjunction with fluorescence intensity,[8] have been analyzed to identify individual molecules. In general, for the identification, parameters computed from bunches of photons were considered, such as fluorescence lifetime[6] and photon burst size.[8] In the case of fluorescence lifetime, these parameters are obtained by fits of a model function to a photon arrival time histogram obtained from a single molecule. The criterion of the identification is based on the agreement of the single-molecule fluorescence lifetime with the bulk lifetime within a given confidence interval.

Recent developments in time-correlated counting electronics allow time-resolved photon-by-photon analysis and multichannel detection.[10] Taking advantage of these improvements, we propose a method based on simple probability considerations which can be extended to any observable in a single-molecule fluorescence experiment.

In our method, each photon emerging from the molecule is analyzed with regard to a probability distribution obtained from reference experiments. The probability distributions that serve as a fingerprint are established from measurements of the observables for pure bulk samples of the different species. The observables may be arrival time after pulsed excitation and after the last emitted photon, emission wavelength, and polarization. For each photon, a certain probability for having either of the molecules as the emission source can be computed. By combining these probabilities photon by photon, eventually the molecule can be identified.

An important benefit of the method is that the calculations are simple and can be performed on-line. Therefore, first, the molecules can simultaneously be sorted and separately collected for further use or verification of the identification. Second, the analysis can be stopped when the required accuracy is reached, which allows for a further optimization of the identification procedure. Third, no model function is required for the analysis.

We present for the first time single-molecule identification of four different dyes with spectral- and time-resolved fluorescence. The combination of time-correlated single-photon counting with spectral analysis dramatically decreases the number of photons needed for single-molecule identification as compared to identification based on lifetime exclusively. That means, as much information as possible from each photon is utilized in order to identify molecules with as few photons as possible. Monte Carlo simulations were performed with the aim to get an estimate how many photons are needed for identification.

THEORETICAL SECTION

The normalized photon histogram from molecules of sort m_i ($1 \leq i \leq N$, where i denotes the different species, with respect

* Corresponding author: (fax) +41-1-6321021; (e-mail) huebner@phys.chem.ethz.ch.

(1) Soper, S. A.; Davis, L. M.; Shera, E. B. *J. Opt. Soc. Am. B* **1992**, *9*, 1761−9.
(2) Nie, S.; Zare, R. *Anal. Chem.* **1995**, *70*, 431−437.
(3) Brand, L.; Eggeling, C.; Zander, C.; Drexhage, K. H.; Seidel, C. A. M. *J. Phys. Chem. A* **1997**, *101*, 4313−4321.
(4) Fries, J. R.; Brand, L.; Eggeling, C.; Kollner, M.; Seidel, C. A. M. *J. Phys. Chem. A* **1998**, *102*, 6601−6613.
(5) Kollner, M.; Fischer, A.; Arden-Jacob, J.; Drexhage, K. H.; Muller, R.; Seeger, S.; Wolfrum, J. *Chem. Phys. Lett.* **1996**, *250*, 355−360.
(6) Zander, C.; Sauer, M.; Drexhage, K. H.; Ko, D. S.; Schulz, A.; Wolfrum, J.; Brand, L.; Eggeling, C.; Seidel, C. A. M. *Appl. Phys. B* **1996**, *63*, 517−523.
(7) Zander, C.; Drexhage, K. H.; Han, K. T.; Wolfrum, J.; Sauer, M. *Chem. Phys. Lett.* **1998**, *286*, 457−465.
(8) VanOrden, A.; Machara, N. P.; Goodwin, P. M.; Keller, R. A. *Anal. Chem.* **1998**, *70*, 1444−1451.
(9) Dorre, K.; et al. *Bioimaging* **1997**, *5*, 139−52.
(10) Becker, W.; Hickl, H.; Zander, C.; Drexhage, K. H.; Sauer, M.; Siebert, S.; Wolfrum, J. *Rev. Sci. Instrum.* **1999**, *70*, 1835−1841.

10.1021/ac991116k CCC: $19.00 © 2000 American Chemical Society
Published on Web 12/22/1999

to one of the above-mentioned observables x, provides the relative occurrence of photons in the respective bin. This relative occurrence is equal to the probability of detecting a photon in bin b ($0 \le b \le K$) of the observable x

$$P(b|m_i) = \frac{c_i(b)}{\sum_{\tilde{b}=0}^{K} c_i(\tilde{b})} \quad (1)$$

where $c_i(b)$ is the number of counts in bin b. Assuming that the number of counts not originating from the molecule is negligible compared to the number of photons emitted by the molecule, the probability of molecule m_i being the source of a count in bin b is

$$P(m_i|b) = \frac{P(m_i)P(b|m_i)}{\sum_{q=1}^{N} P(m_q)P(b|m_q)} \quad (2)$$

provided that the molecule is one out of the N possible kinds of molecules and no additional species of light-emitting molecules is present in the sample. If the occurrences of the different molecules in the sample are comparable, as would be the case in DNA sequencing, eq 2 can be approximated by

$$P(m_i|b) = \frac{P(b|m_i)}{\sum_{q=1}^{N} P(b|m_q)} \quad (3)$$

The probabilities $P(m_i|b)$ from eq 3 establish a look-up table for the different types of molecules with an entry for each bin b and each species. These entries are the probabilities that can be addressed for each detected photon. If more properties of a fluorescence photon are taken into account, their respective probabilities according to eq 2 or eq 3 have to be multiplied provided that the observables are statistically independent. In this case, the probabilities $P(m_i|b)$ are functions of more than one observable. If L photons are detected from one molecule in the respective bins b_j, where the index j denotes the photon number, the probability that these photons are emitted by the molecule m_i can be calculated according to

$$P_L(m_i) = \frac{\sum_{j=1}^{L} P(m_i|b_j)}{\sum_{q=1}^{N}\sum_{j=1}^{L} P(m_q|b_j)} \quad (4)$$

The course of the analysis is as follows. First, the observables are measured in reference samples containing only one species of molecule. The resulting photon histograms with respect to the observables are normalized to obtain the probability distribution $P(b|m_i)$. If only photons within a range of values of the observables are to be considered, e.g., within a photon arrival time window, the normalization has to be performed in the limits of this window. A look-up table is generated with the aid of eq 3 which assigns to each bin of the observable x a certain probability for each type of molecule. This look-up table is multidimensional if more than one observable is measured.

To identify the molecules, for each photon the probabilities $P(m_i|b)$ can be read from the look-up table and the probabilities $P_L(m_i)$ according to eq 4 are calculated. The analysis can be stopped as soon as the probability $P_L(m_i)$ has reached a preset accuracy level. The fraction of correctly identified molecules corresponds to this accuracy level.

EXPERIMENTAL SECTION

In our experiment, the arrival time after pulsed excitation and the emission wavelength of the fluorescence photons were observed. An active mode-locked Nd:YAG laser (Coherent Antares, Palo Alto, CA) frequency-doubled to 532 nm was used to excite the molecules. The width of the pulses was ∼150 ps with a repetition rate of 76 MHz. The excitation light was fed through an optical fiber for spatial filtering and focused to a diffraction-limited spot into the sample by an oil immersion objective (numerical aperture 1.3, Leica, Bensheim, Germany). The fluorescence photons were collected by the same objective and separated from the excitation light by a dichroic mirror (50% transmission at 550 nm). An additional holographic notch filter at 532 nm suppressed scattered excitation light.

For fast and efficient detection of fluorescence photons, two avalanche photodiodes (APD; SPQ 141, EG&G, Vandreuil, Canada) with dark count rates of 100 counts/s were used. The aperture of the optical fiber and the small active area of the APD defined a confocal volume of ∼100 aL, thus providing a low background light level. The fluorescence light was spectrally split by a second dichroic mirror (50% transmission at 575 nm) before being detected by the APDs. Both APD signals were fed through a routing device into a PC card-based single-photon counting device (SPC 402, Becker & Hickl, Berlin, Germany). A trigger from the mode-locker driver served as the synchronization signal. Having dead times of 100 and 150 ns, respectively, the APDs and the SPC card were capable of detecting photons with a rate of 8 MHz. Considering an overall detection efficiency of the setup of $\eta \simeq 0.065$, this number perfectly matched the excitation repetition rate. For each photon, the arrival time τ with respect to the excitation pulse (256 channels, 100 ps width), the time t with respect to the start of the experiment (50 ns resolution), and the information on which detector counted the photon were stored. The timing resolution was ∼400 ps fwhm. A home-build piezo-driven $x-y-z$ scanning stage was used for raster scanning the sample.

The four dyes rhodamine 6G (R6G), sulforhodamine B (SRB), dibenzanthranthene (DBATT), and 1,1′-dioctadecyl-3,3,3′,3′-tetramethylindocarbocyanine perchlorate (DiI) were chosen for the experiments because of their distinct emission spectra and fluorescence lifetimes (cf. Table 1). Thin films with a thickness of ∼50 nm were spin cast on microscope cover slips from a solution of poly(methyl methacrylate) (PMMA) in toluene with dye concentrations in the order of 10^{-9} M. Raster scanning of these samples led to a fluorescence burst each time a molecule traversed the confocal volume. This way, the transit time of the molecules could be conveniently adjusted to simulate identification in a flowing sample stream which would be investigated in DNA

Table 1. Right Column: Results of the MC Simulations of Single-Molecule Identification with Sequential Statistics

dye	emission maximum[b] (nm)	fluorescence lifetime[b] (ns)	no. of photons[a]	
			1 channel	2 channels
DBATT	586	10.15	285	102
SRB	584	4.15	7562	647
R6G	555	3.75	8070	267
DiI	572	3.05	1638	724

[a] Average number of photons required for identification with 99.9% accuracy. [b] Bulk emission maximum and fluorescence lifetime.

sequencing. Bulk samples were prepared in the same way as the single-molecule samples but with micromolar dye concentrations. In this case, several hundred molecules were in the confocal volume, which allowed for recording of the arrival time histograms without a burst analysis.

For the simulation of the experiment, the look-up tables were generated by assuming single-exponential fluorescence decay behavior. A single look-up table suffices for the simulation of the experiment without spectral resolution, whereas two look-up tables, one for each spectral channel, were computed for the simulation of the spectrally resolved experiment. The look-up tables were generated within a finite time window since in the experiment a time window has to be set to account for the laser repetition rate and to reject photons not emitted by the molecule. To compare the simulation with the experiment, the bulk values obtained from the reference experiments were used for the fluorescence lifetimes as well as for the count ratio between the two spectral channels (cf. Table 1).

Random photon arrival times were generated with the aid of a Monte Carlo (MC) method obeying exponential probability distributions according to one of the four molecules at a time. Gaussian time jitter was added to account for a finite instrument response function. To simulate the experiment with two spectrally resolved channels, these were randomized obeying the count ratio between both channels. The simulated data were then analyzed following the procedure outlined in the previous section.

RESULTS AND DISCUSSION

MC Simulations. Table 1 summarizes the results of the MC simulations. It compares the number of photons needed to achieve a confidence level of 99.9% for four different dyes for detection with and without spectral resolution. It shows that the dyes SRB and R6G can be identified with a small number of photons only if spectral resolution is exploited. The dye DBATT, which has a significantly longer fluorescence lifetime than the other dyes, is identified by a small number of photons even without the need of spectral resolution. The most dramatic decrease of the number of photons required for identification when spectral resolution is exploited is observed for R6G since its spectral mean is significantly lower than that of the other dyes, whereas its lifetime is close to that of SRB. Each average is based on the simulation of 10 000 single-molecule photon bursts. About 10 of the simulated molecules were misidentified in accordance with theory.

Recently, the number of photons required to identify one molecule out of two by fluorescence lifetime without spectral information with 99.9% accuracy was theoretically predicted to be well above 1000 if their lifetimes are 3 and 4 ns, respectively.[11] The prediction was based on the precision of the determination of fluorescence lifetime, i.e., based on a model function. Here, e.g., DiI and SRB having fluorescence lifetimes of 3.05 and 4.15, respectively, are both distinguished within a set of four dyes with ~700 photons. Compared to the theoretical prediction in ref 11, our approach is more realistic since a finite instrument response function is taken into account. In addition, to match the experimental conditions, only photons detected within a 12 ns time window are used for the analysis yet photons outside this window still contribute to the total number of photons needed.

Single-Molecule Identification Experiments. Graphical representations of the look-up tables constructed according to eq 4 from the bulk measurements are shown in Figure 1 together with the measured photon arrival time histograms (a). The stacked probabilities (eq 3) to have a given molecule as the source of a photon is plotted versus the photon arrival time for both spectral channels (b). Due to the pulse repetition rate of 13 ns, almost two complete fluorescence decays occur within the range of 25 ns of the time-to-amplitude converter (TAC). Therefore, for each of them, a separate time window is set. The borders of the time window were adjusted to exclude laser light scattered by the matrix surrounding the molecule. This way, the background light that still might pass the notch filter can be reduced to a negligible amount. On the other hand, to optimize the number of analyzed photons, the time windows were kept as large as possible. The photons close to the boundaries of the TAC range are discarded since the TAC has nonlinearities at both limits of its range.

The shape and the size of the areas representing the probabilities determine the efficiency of the identification procedure. Two remarkable features can be observed in the photon arrival time histograms. A peak at time 13 ns, the time position of the laser pulse, in the histogram of the dye DBATT in the short wavelength channel shows that scattered laser light still is detected. The peak is visible only for DBATT since its fluorescence intensity in the short-wavelength channel is much lower as compared to the other dyes. Second, there is a significant difference in the shape of the histograms at short times between the two channels. This difference is due to rotation of the molecules within the polymer matrix, which leads to a polarization relaxation of the emission.[12] The dichroic mirror has a polarization-dependent reflectance, which makes this relaxation visible in the arrival time histogram.

In the single-molecule identification experiment, the respective look-up table is consulted for each count depending on which detector registered the photon. The obtained probabilities lead together with eq 4 to the joint probabilities of observing either one of the molecules.

A typical image of a sample of SRB in a PMMA film obtained by confocal scanning microscopy is shown in Figure 2. The count rate as a function of time along the white line is plotted in the inset. The transit of the molecule through the confocal spot is seen as a burst of photons having a duration of ~15 ms. Depending on the type of dye, during the transit time ap-

(11) Sauer, M.; Arden-Jacob, J.; Drexhage, K. H.; Gobel, F.; Lieberwirth, U.; Muhlegger, K.; Muller, R.; Wolfrum, J.; Zander, C. *Bioimaging* **1998**, *6*, 14−24.

(12) Rice, S.; Kenney-Wallace, G. *Chem. Phys.* **1980**, *47*, 161.

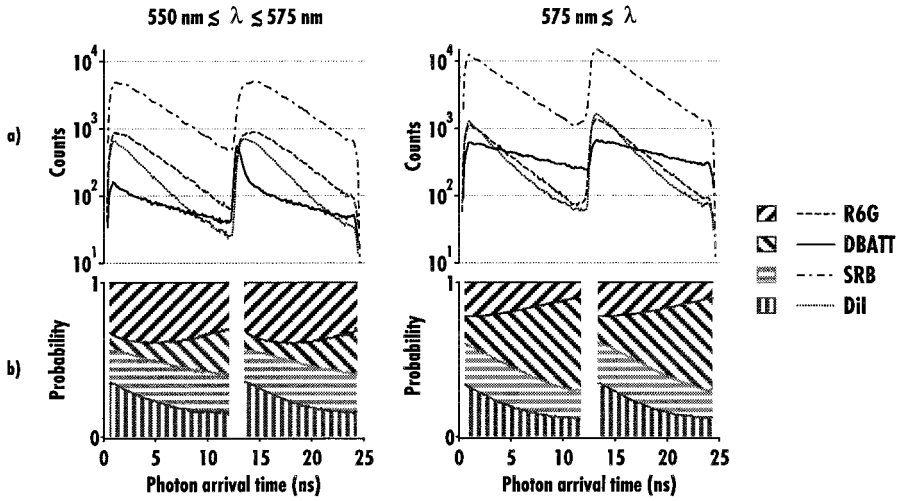

Figure 1. (a) Photon arrival time histograms in the two spectrally distinguished channels obtained from bulk measurements of the four dyes R6G, SrB, DBATT, and DiI. (b) Graphical representation of the look-up tables computed from the histograms a according to eq 3 in two time windows within the TAC range.

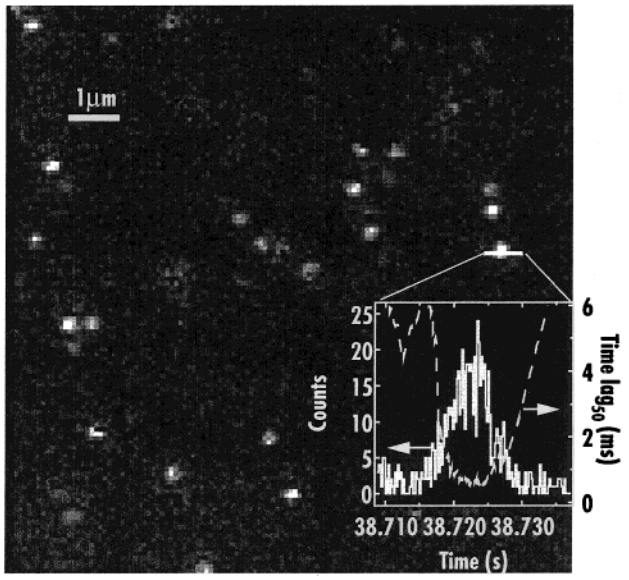

Figure 2. Scanning confocal optical microscopy image of individual SRB molecules in a PMMA film. Inset: histogram (bin width 0.2 ms) of the photon burst emitted by the highlighted molecule during a single transit through the confocal volume (solid line, left axis) and time lag between 50 consecutive photons (dashed line, right axis).

proximately 500−3000 photons were detected in a burst. The number of dark counts during the dwell time is at least 2 orders of magnitude lower and therefore neglected in the analysis.

The dashed line in Figure 2 is a plot of the time lag between 50 consecutive counts. If the time lag is below 2 ms, the identification procedure is started. It is stopped either when the molecule is identified or when the time lag between 50 consecutive counts exceeds the limit of 2 ms. In the latter case, the molecule is regarded as not identified. The decision when to start and to stop the analysis is the only part of the experiment depending on the count rate. Thus, the above-mentioned number of photons and the time lag have to be adjusted in such a way that the most weakly luminescing species is clearly distinguished from the background without the loss of too many photons.

Figure 3 visualizes the identification experiment for four different single molecules, one of each dye, embedded in a PMMA

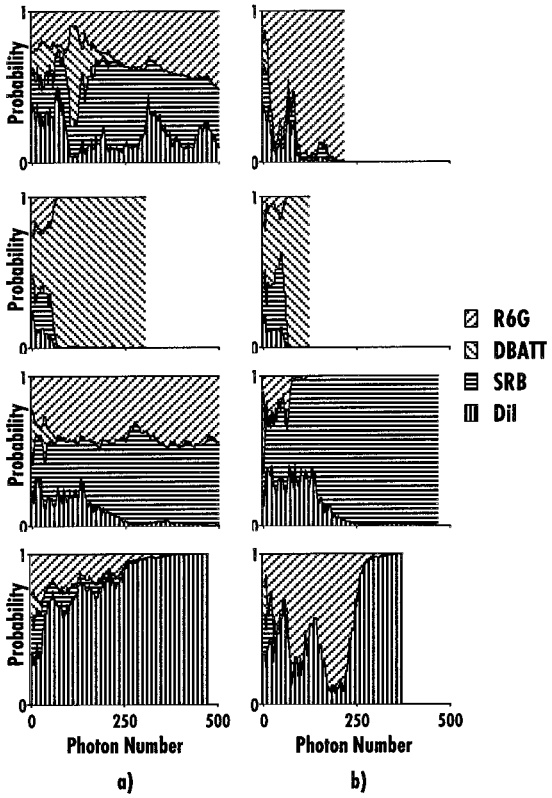

Figure 3. Analysis of single-molecule identification experiments for four molecules of the respective dyes: stacked probabilities to have either dye versus photon number analyzed without (a) and with (b) spectral resolution.

film. From the top to the bottom rows, respectively, pure dilute samples of R6G, DBATT, SRB, and DiI were investigated in this experiment. The stacked probabilities according to eq 4 are plotted versus the photon number for four selected molecules. Each panel shows how the probability of either one of the molecules being observed develops with the photon number. As soon as the confidence level of 99.9% is reached, the analysis is stopped. For the sake of comparability with the MC simulation, the fluorescence photons of each burst are analyzed without (a) and with (b) taking

spectral resolution into account. In the case of spectrally resolved time-correlated single-photon counting (b), the four molecules are identified with a small number of photons. In contrast, if spectral resolution is discarded (a), no decision can be made regarding the dyes R6G and SRB within the transit time of the molecule in this example. Figure 3 demonstrates the stochastic nature of the identification experiment. For small numbers of photons, the probabilities show large fluctuations which decrease as the number of detected photons increases.

Due to the polarization dependence of the dichroic mirror reflectance, the count ratio between both spectral channels depends on the orientation of the single molecule. Within the PMMA film, the molecules obey a fixed orientation. For this reason, linearly polarized light was used in order to excite mainly molecules with the proper orientation. A molecule in a liquid that rotates quickly compared to the transit time through the confocal volume takes an isotropic effective orientation removing the polarization dependence of the count ratio.

In the DiI sample, which showed the best photostability, 54 photon bursts were analyzed yielding no misidentification. On average, 191 photons were needed for the identification. This number is even smaller than that predicted by the MC simulation. A possible explanation for the gain in efficiency might be the polarization dependence of the dichroic mirror, which leads to a further discrimination with respect to polarization anisotropy.

Compared to other single-molecule identification methods, the approach presented here has three advantages. The photodestruction quantum yield and out-of-focus molecules, which may both obscure the results obtained if the fluorescence burst size is evaluated,[8] have no influence on fluorescence lifetime and spectrum. It is very unlikely that a molecule undergoes photodestruction prior to identification because of the small number of photons needed. In fact, while scanning the sample only in one direction, thus exciting the same molecules several times, we never observed bleaching during the first scan since the photodestruction quantum yield of the dyes is at most in the order of 10^{-5}, which corresponds to a number of detectable photons of at least 5000. Also, no theoretical models are required since our analysis is based on probability distributions. Additionally, since all instrument response functions are already inherent in the probability distribution, there is no need for a deconvolution.

CONCLUSIONS

We have shown that spectrally resolved time-correlated single-photon counting is suitable to identify single molecules that are similar in fluorescence lifetime, but different in emission spectra, and vice versa. With our method of analysis, the information associated with each photon can be transparently exploited for the identification of its single molecular emitter. The method can be extended to more observables and higher resolutions. Increasing the spectral resolution to 256 channels is well conceivable, and MC simulations show that this will lead to a further decrease by almost 1 order of magnitude of the number of photons needed for identification. An improved set of dyes with more distinct properties can gain efficiency of identification to an even higher level. If the molecules are distinct with respect to another observable, resolution of this observable has to be provided and one more histogram needs to be recorded which is then used for the analysis in the same way. For example, to utilize rotational mobility, a polarizing beam splitter has to be inserted in one of the two detection paths with an additional APD. Thus, if only a small number of photons emerging from one single molecule can be detected, the single-molecule misclassification probability can be decreased to the level necessary for reliable DNA sequencing. The method offers the chance to simplify and miniaturize the components to produce an inexpensive, easy to use, yet extremely sensitive screening device for clinical applications.

ACKNOWLEDGMENT

This work was supported by the Swiss National Science Foundation and ETH Zurich.

Received for review September 29, 1999. Accepted November 21, 1999.

AC991116K