

Multiparameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 reverse transcriptase:primer/template complexes

P. J. Rothwell*[†], S. Berger^{†‡§}, O. Kensch*^{†¶}, S. Felekyan[‡], M. Antonik[‡], B. M. Wöhrl*, T. Restle*, R. S. Goody*^{¶||}, and C. A. M. Seidel^{¶||}

[†]Abteilung Spektroskopie und Photochemische Kinetik, Max-Planck-Institut für Biophysikalische Chemie, Am Fassberg 11, D-37077 Göttingen, Germany; and [‡]Abteilung Physikalische Biochemie, Max-Planck-Institut für Molekulare Physiologie, Otto-Hahn-Strasse 11, D-44227 Dortmund, Germany

Edited by James A. Spudich, Stanford University School of Medicine, Stanford, CA, and approved December 23, 2002 (received for review July 5, 2002)

By using single-molecule multiparameter fluorescence detection, fluorescence resonance energy transfer experiments, and newly developed data analysis methods, this study demonstrates directly the existence of three structurally distinct forms of reverse transcriptase (RT):nucleic acid complexes in solution. Single-molecule multiparameter fluorescence detection also provides first information on the structure of a complex not observed by x-ray crystallography. This species did not incorporate nucleotides and is structurally distinct from the other two observed species. We determined that the nucleic acid substrate is bound at a site far removed from the nucleic acid-binding tract observed by crystallography. In contrast, the other two states are identified as being similar to the x-ray crystal structure and represent distinct enzymatically productive stages in DNA polymerization. These species differ by only a 5-Å shift in the position of the nucleic acid. Addition of nucleoside triphosphate or of inorganic pyrophosphate allowed us to assign them as the educt and product state in the polymerization reaction cycle; i.e., the educt state is a complex in which the nucleic acid is positioned to allow nucleotide incorporation. The second RT:nucleic acid complex is the product state, which is formed immediately after nucleotide incorporation, but before RT translates to the next nucleotide.

HIV-1 reverse transcriptase (RT) is a key enzyme in the life cycle of HIV. This multifunctional enzyme is responsible for the complex process of transcribing viral RNA into double-stranded DNA for integration into the host cell genome. The enzyme is a heterodimer composed of subunits which share a common N terminus and have subdomains referred to as fingers, palm, thumb, and connection (Fig. 1). Although the subdomains of each subunit are structurally similar, the overall folding within the two subunits is quite different (1). The large subunit, p66, contains RNA- and DNA-dependent DNA polymerase as well as RNase H activities. The p51 subunit, which is inactive in the heterodimer but active in homodimers (2, 3), is thought to play a role in stabilizing the conformation of p66.

Crystallographic studies of RT:nucleic acid primer/template (p/t) complexes have provided valuable insights into the structural features and conformational changes induced by p/t binding (1, 4–7). To date, these crystallographic models have indicated a single p/t-binding mode. In contrast, recent solution-based kinetic studies on p/t binding and nucleotide incorporation by RT suggest a heterogeneous mixture of several different binding modes (8). To confirm directly the existence of several species, and to obtain structural and functional information about each, we have used single-molecule spectroscopy to investigate RT:p/t complexes in solution.

Single-molecule techniques have proven to be valuable tools for resolving static and dynamic heterogeneity of an ensemble (9–13). For this investigation, we use a newly developed multiparameter fluorescence detection (MFD) technique. MFD is characterized by the simultaneous and independent acquisition of fluorescence intensity (F), lifetime (τ) and anisotropy (r) as a

function of time (t) in two spectral windows (λ) on a molecule-by-molecule basis (9). The spectroscopic information can be used to classify molecules into the different subpopulations that compose an ensemble and to characterize their size, identity, diffusion times, or local environment. The use of multiple, simultaneously acquired parameters improves the resolution of distinct species significantly (14, 15).

To obtain structural information at the single-molecule level, spectroscopic techniques can be combined with fluorescence resonance energy transfer (FRET). In FRET, the energy from an excited donor dye, D , is nonradiatively transferred to an acceptor dye, A . The transfer depends strongly on the separation distance and provides a tool for measuring distances in the range of 20–100 Å (16). Single molecule FRET has been successfully applied to monitor conformational changes of immobilized ribozymes and proteins (13, 17–22). However, quantitative analyses are difficult to perform because FRET depends not only on distance, but on several time-varying parameters such as molecular orientations, dye quenching, or dye bleaching.

To provide the high resolution necessary to distinguish between structurally similar species, FRET experiments were performed by using the MFD technique, which accounts for inhomogeneous broadening due to local effects such as quenching, orientation, and bleaching.

It is important to note that the two reporter dyes represent, in effect, two degrees of freedom. Thus, only two-dimensional analysis with independent parameters can separate globally correlated FRET effects from individual dye effects. One possibility provided by MFD is to use the donor fluorescence lifetime and the measured intensity ratio of donor and acceptor signal, F_D/F_A . In the study presented here, these techniques are used to determine the number of RT:p/t complexes which occur and to characterize their structural properties.

Materials and Methods

We used an RT mutant containing a single accessible cysteine at position 287 (see Fig. 1) of the p66 subunit (23). This cysteine, situated within the thumb subdomain of p66, was labeled with the green donor fluorophore Alexa⁴⁸⁸-C₅ maleimide (24). The red acceptor dye Cy5-*N*-hydroxysuccinimide ester was attached

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: RT, reverse transcriptase; p/t, primer/template; MFD, multiparameter fluorescence detection; F , fluorescence intensity; τ , lifetime; r , anisotropy; t , time; λ , spectral windows; FRET, fluorescence resonance energy transfer; DE , dead-end complex; A , acceptor fluorophore; D , donor fluorophore; dNTP, deoxy-nucleoside triphosphate; dp/dt, DNA primer/DNA template; R_{DA} , donor-acceptor distance; sMFD, single-molecule MFD.

[†]P.J.R., S.B., and O.K. contributed equally to this work.

[§]Present address: Solvias AG Klybeckstrasse 191, Postfach, CH-4002 Basel, Switzerland.

[¶]Present address: DIREVO Biotech AG, Biochemical Assay Development, Nattermannallee 1, D-50829 Köln, Germany.

^{||}To whom correspondence should be addressed. E-mail: roger.goody@mpi-dortmund.mpg.de or cseidel@gwdg.de.

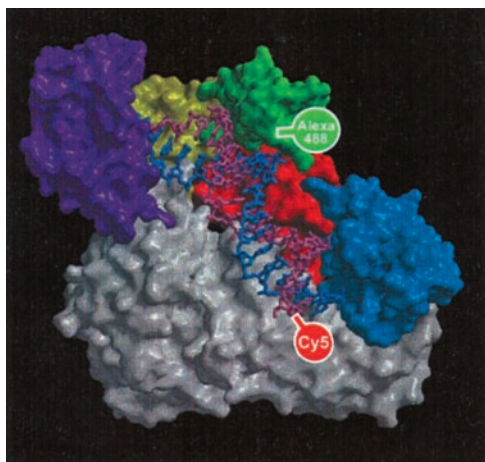


Fig. 1. Crystal structure of the RT:dp/dt complex. The protein is displayed as a molecular surface. The different regions of p66 are colored as follows: fingers, purple; palm, yellow; thumb, green; connection, red; and RNase H domain, blue; whereas p51 is colored gray. The nucleic acid substrate is shown as a stick representation with the primer colored purple and the template colored blue. The positions of the labels are shown: the donor dye (Alexa⁴⁸⁸) is attached to the protein and the acceptor dye (Cy5) is attached at the 5' end of the primer.

via a C₆-aminolink to the 5' end of the primer strand of a 19/35 DNA/DNA primer/template, a construct referred to as Cy5-[5'-DNA primer/DNA template (dp/dt)]. The sequence of the primer/template is based on the HIV-1 viral primer binding site. The primer sequence is 5'-d(TTG TCC CTG TTC GGG CGC C)-3' and the template sequence is 5'-d(GGG TTT GCT AAC CAT AGG CGC CCG AAC AGG GAC AA)-3'. In the case of the terminated primer, the 3' cytosine was replaced by a dideoxy-cytosine, and the corresponding p/t is referred to as d_{dp/dt}.

The Cy5-(5'-dp/dt) construct was designed such that the distance between the residue at position 287 and the 5' end of the primer was expected to be ≈40 Å, based on previously published crystallographic structures (6). This distance is close to the Förster radius, R_0 , of 49 Å for Alexa⁴⁸⁸-Cy5, which is a prerequisite for accurate distance calculations.

The aqueous measurement buffer contained 10 mM KCl, 50 mM Tris·HCl, pH 7.8, 6 mM MgCl₂, and 400 μM ascorbic acid. Because of the extremely low concentrations used in single-molecule studies (50 pM RT and 200 pM dp/dt), high binding affinities are required to ensure stable complex formation. By reducing the KCl concentration from the 50 mM used in previous studies on this system (8) to 10 mM, we were able to reduce the dissociation constant, K_d , dramatically. Data from classical ensemble titrations indicated an upper limit for the K_d value at this salt concentration of <100 pM. The actual single molecule measurements confirmed that the affinity was high enough to ensure complex formation, because most RT molecules sampled contained bound (labeled) substrate. Ensemble measurements on single nucleotide incorporation under these salt conditions showed similar kinetic behavior to that seen at higher salt, with three kinetic phases being observed (8).

Before measurements, a “stock” complex solution was made at a concentration of 250 nM protein and 1 μM dp/dt. This solution was equilibrated on ice for at least 30 min. Immediately before a measurement, the stock solution was diluted rapidly into the measurement buffer 1:10, followed by a further 1:10 dilution before a final 1:50 dilution to a final concentration of 50 pM protein and 200 pM dp/dt. A droplet of the sample solution (50 μl) was applied to a coverslip which formed the bottom of

a closed chamber with a water-saturated atmosphere. Coating the coverslip with κ-casein led to complete suppression of adsorption of the sample molecules during the measurement time of up to an hour.

Multiparameter Fluorescence Detection. sMFD is performed by using a confocal (pinhole diameter = 100 μm) epi-illuminated microscope with excitation by a pulsed linearly polarized argon-ion laser at 496 nm. The laser is focused into the solution with a 60 × 1.2 water immersion objective. Fluorescence detection is performed with the same objective, but with an additional pinhole, resulting in a detection volume of 2 fl, as estimated from fluorescence correlation spectroscopy (FCS) measurements. Further set-up parameters obtained via FCS are the ratio ω_0/z_0 of 4, and the characteristic diffusion times t_D for Rhodamine 110 and Alexa⁴⁸⁸-K287C of 0.3 and 1.3 ms, respectively. Here, ω_0 and z_0 are the distances from the center of the laser beam focus in the radial and axial directions, respectively, at which the collected fluorescence intensity has dropped by a factor of e^2 , compared with its peak value. Sample molecules diffusing freely through the solution occasionally pass through the detection volume, resulting in a brief (≈1 ms) burst of fluorescence. Dilute solutions of molecules (≈50 pM) ensure that only single molecules are detected, each molecule producing a single burst. The fluorescence signal is divided into parallel and perpendicular components in wavelength ranges below and above 620 nm (green and red, respectively; filters: HQ535/50, HQ730/140). The photons are detected by four avalanche photodiodes coupled to a counting board (SPC 431, Becker & Hickel, Berlin) and a personal computer. For each photon, the arrival time after the laser pulse, the time since the last photon, the polarization, and the wavelength are recorded. Fluorescence bursts are distinguished from the background signal of 1–2 kHz by applying certain threshold intensity criteria (9).

FRET-Analysis Techniques. FRET occurs between two dyes when the emission spectrum of an excited donor fluorophore overlaps the absorption spectrum of a nearby acceptor fluorophore. The efficiency, E , of FRET depends strongly on the dye separation [donor-acceptor distance (R_{DA})]:

$$E = \left[1 + \left(\frac{R_{DA}}{R_0} \right)^6 \right]^{-1} \quad [1]$$

R_0 is a DA-pair-specific constant known as the Förster radius and depends on the spectral overlap integral between D and A , $J(\lambda)$, the donor fluorescence quantum yield, Φ_{FD} , the refractive index, n , and the geometric factor, κ^2 , which accounts for the steady state relative dipole orientation between D and A : $R_0 = 9780(\lambda)^{-4}n^{-4}\Phi_{FD}$.

E can be calculated by two methods: (i) either by measuring the corrected donor and acceptor fluorescence intensities, F_D and F_A , or (ii) by the donor lifetimes in the presence and absence of an acceptor, $\tau_{D(A)}$ and $\tau_{D(0)}$.

$$E = \left[1 + \frac{F_D a_{trans} \Phi_{FA}}{F_A \Phi_{FD}} \right]^{-1} = 1 - \frac{\tau_{D(A)}}{\tau_{D(0)}} \quad [2]$$

For intensity measurements, the fluorescence quantum yields of the acceptor, Φ_{FA} , and donor, Φ_{FD} , must be known (0.40 and 0.63 respectively for these experiments), and the factor a_{trans} (=0.80 for Cy5) has been introduced to account for the fraction of the acceptor dye in a fluorescent trans conformation (25).

Detailed analysis of the fluorescence decay of the subensemble of FRET-active molecules indicated fluorescence quenching of the donor dye resulting in a double exponential fluorescence decay with the fluorescence lifetimes, τ_{DN} , and species amplitudes, x_{DN} ($\tau_{D1} = 4.27$ ns, $x_{D1} = 0.56$, and $\tau_{D2} =$

1.11 ns, $x_{D2} = 0.44$). Quenching does not perturb the FRET process as expressed by the “reduced Förster radius,” R_{0r} defined below in 4; however, this competing process must be taken into account for the calculation of the appropriate FRET-parameters (26). Knowing that the maximum-likelihood estimator used for the calculation of single-molecule fluorescence decay data gives results that are close to a fluorescence-weighted average fluorescence lifetime, we developed a correction procedure to compute the fluorescence lifetime of the donor not perturbed by reversible quenching. Because correlation analysis and sMFD indicate that the fluctuations occur in the submillisecond time range, it is justified to use the species-weighted average fluorescence lifetimes, $\bar{\tau}_{D(0)}$ and $\bar{\tau}_{D(A)}$, as the appropriate variables in the calculations of the FRET-efficiency. Therefore, a polynomial correction function was generated by simulating FRET to obtain the quadratic correction factors of $a = -0.5330$, $b = 0.9583$, and $c = 0.0139$ (see *Supporting Text*, which is published as supporting information on the PNAS web site, www.pnas.org). These values were used to rescale all experimentally determined lifetimes values to their photophysically meaningful ones.

The signal intensities, S_G and S_R , are calculated by dividing the number of photons by the burst duration. F_D and F_A are determined from S_G and S_R by correcting for the background counts, B_G and B_R , detection efficiencies, g_G and g_R , and spectral crosstalk, α .

$$F_D = \frac{S_G - B_G}{g_G} = \frac{F_G}{g_G} \quad [3]$$

$$F_A = \frac{(S_R - B_R) - \alpha(S_G - B_G)}{g_R} = \frac{F_R}{g_R}$$

Combining 1 and 2, we obtain the separation of two dyes as a function of their intensities:

$$R_{DA} = R_{0r} \left(a_{trans} \Phi_{FA} \frac{F_D}{F_A} \right)^{1/6} \quad [4]$$

We define R_{0r} ($= 53 \text{ \AA}$) as the “reduced Förster radius,” which differs from R_0 by being independent of Φ_{FD} : $R_{0r} = 9780 \cdot [J(\lambda) \cdot \kappa^2 \cdot n^{-4}]^{(1/6)}$. The detection efficiencies always appear together as the ratio g_G/g_R , which is measured for each experiment and typically has a value of 0.52. The distribution in calculated ranges is dominated by the shot noise in the photon counting, and therefore follows Poisson statistics. By following error propagation rules, R_{DA} will have a Gaussian distribution, with an SD, σ , explicitly given by:

$$\sigma(R_{DA}) = \frac{1}{6} R_{0r} \left[\frac{F_D}{F_A} a_{trans} \Phi_{FA} \right]^{1/6} \cdot \left[\left(\frac{1}{F_G} + \frac{\alpha}{F_R} \right)^2 (S_G + B_G) + \left(\frac{1}{F_R} \right)^2 (S_R + B_R) \right]^{1/2} \quad [5]$$

Quantitative formulae are provided by 4 and 5 for determining the separation between two dyes based on intensity measurements. For 4 and 5, we assumed κ^2 to be 2/3. The validity of this assumption is explicitly checked by measuring the species-selective anisotropies. The value for σ must be calculated from photon counts, not count rates, as suggested in 5. In our case, the photon flux varies in the bursts, being averaged on the millisecond time scale seen by FCS. Therefore, we use the values measured for the count rates, in kHz, as average photon counts in 5. Empirical evidence available from a variety of experiments

suggests that this use of the measured count rates to calculate σ is accurate.

The fluorescence lifetime is determined for each burst in two steps: (i) by generating a histogram of the photon arrival times, and (ii) by fitting the histograms to a single exponential by using a maximum-likelihood estimator and iterative convolutions to account for the scatter contribution (9, 27). From 1 and 2, the relationship between the fitted $\bar{\tau}_{D(A)}$ values and R_{DA} can be determined.

$$R_{DA} = \sqrt[6]{\frac{\bar{\tau}_{D(A)} \Phi_{FD}}{\bar{\tau}_{D(0)} - \bar{\tau}_{D(A)}}} R_{0r} \quad [6]$$

We use the average lifetime value $\bar{\tau}_{D(0)} = 3.1$ ns. The stepwise correction of the data for background and lifetime is shown in Fig. 5, which is published as supporting information on the PNAS web site.

Assuming mobile donor and acceptor dyes, the combination of the accuracy of lifetime measurements and the precision of intensity measurements via sMFD advances FRET as a tool to measure not only relative but also absolute distances. In addition, plotting R_{DA} vs. $\bar{\tau}_{D(A)}$ reveals complementary information about the existence of multiple dye states.

For R_0 , a value for the orientation factor κ^2 is required. If the dyes are known to be rotationally mobile over the lifetime of the excited state, it is possible to use the value 2/3 obtained by integrating over all of the possible orientations. The mobility of the donor dye is described by the rotational correlation time, ρ_D , which can be determined indirectly from the donor dye’s anisotropy, r_D . Anisotropy is the degree of polarization in the fluorescent signal and depends upon ρ_D , $\bar{\tau}_{D(A)}$, and the initial anisotropy, $r_0 = 0.375$, as shown in the Perrin equation:

$$r_D = \frac{r_0}{\frac{\bar{\tau}_{D(A)}}{\rho_D} + 1} \quad [7]$$

MFD records the polarization of each photon, enabling us to determine r for each molecule. This method has been described in detail (27). Briefly, the steady-state anisotropy is calculated for each burst, correcting for scattered light, changes in polarization due to the optical components, and the different detection efficiencies of parallel and perpendicular photons (G factor).

Results and Discussion

Investigating the complexes formed between donor-labeled RT and acceptor-labeled Cy5-(5'-dp/dt) by sMFD, Fig. 2 shows the use of multidimensional information in diagrams of the donor lifetime in the presence of acceptor, $\bar{\tau}_{D(A)}$, vs. donor-acceptor distance, R_{DA} , (Fig. 2 Upper) and donor anisotropy, r_D , (Fig. 2 Lower). The number of molecules (fluorescence bursts) in each bin are gray-scale coded from white (lowest) to black (highest). The corresponding 1D-histograms are given as projections. The data for RT: Cy5-(5'-dp/dt) complexes (Fig. 2A) from several measurements readily show that we are able to resolve three distinct RT:dp/dt complexes, designated as *Ia*, *Ib*, and *II*.

For further analysis, it is very useful to plot the function $R_{DA}(\bar{\tau}_{D(A)})$ (6) on the same graph (magenta sigmoidal line) to distinguish local quenching from FRET-effects. If additional local donor quenching/dequenching exists, then the fitted donor lifetimes will be shorter/longer than the expected $\bar{\tau}_{D(A)}$, horizontally shifting those bursts by $\Delta\bar{\tau}_{D(A)}$ (green arrow in Fig. 2B). For local acceptor quenching/dequenching, F_A will be changed and, according to 4, the calculated distances will vertically shift up/down by ΔR_{DA} (red arrow). Therefore, any peak that falls on the line given by 6 is not quenched by any other mechanism

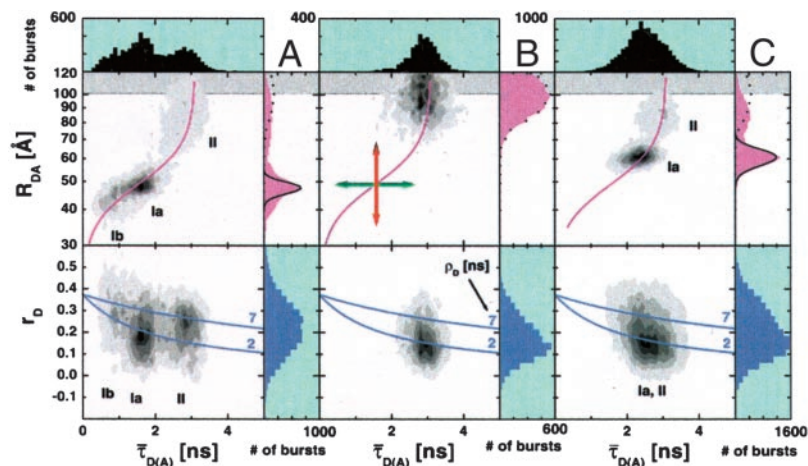


Fig. 2. Quantitative single-molecule multiparameter FRET analysis. (*Upper*) The “species weighted” average lifetimes $\bar{\tau}_{D(A)}$ are plotted vs. R_{DA} together with an overlaid curve computed by **6** ($\bar{\tau}_{D(0)} = 3.1$ ns, $\Phi_{FD} = 0.63$, $R_{0r} = 53$ Å). The gray scale indicates an increasing number of bursts. The one-dimensional R_{DA} histograms (magenta color) also display theoretical predictions of single distance distribution profiles (solid lines) and the maximum reasonable R_{DA} value of ≈ 101 Å for the DA-pair used (dotted lines for donor only species). (*Lower*) $\bar{\tau}_{D(A)}$ is plotted vs. r_D together with the overlaid Perrin equation (7) computed for two rotational correlation times ρ_D (2 and 7 ns). The value 0.375 is used for r_0 . (A) Donor-labeled RT and Cy5-(5'-dp/dt). Species *I* is broadly distributed along the line of **6**, indicating that it consists of two closely related species (*Ia* and *Ib*), which are both mobile and differ slightly in distance, but not in local environment. (B) Donor-labeled RT and unlabeled dp/dt. Only one species is observed, limiting the maximum reasonable value of R_{DA} to 101 Å. (C) Donor-labeled RT and Cy5-(5'-dp/dt) after limited incorporation of four nucleotides. The shifted-species *I* peak is distributed around a mean distance of 61 Å.

besides FRET. Correspondingly, we can conclude that movement of peaks along this line is caused by changes in R_{DA} and not by changes in Φ_{FD} or Φ_{FA} that are due to differences in the local environment. It should be noted here that traditional experiments measuring only intensity cannot detect local donor quenching and cannot distinguish between acceptor quenching and distance changes.

The assignment of species *Ia* and *Ib* is based on the close proximity of these two species in transfer efficiency. The distribution has a maximum DA distance of 48 Å (*Ia*) and a pronounced shoulder at smaller distances (*Ib*). This distance is in close agreement with the expected distances between the labeled bases calculated from the crystal structure (≈ 40 Å) plus the length of the dye linkages (≈ 5 Å) (6). The deviation between the actual broad distribution of the R_{DA} and the theoretical distribution predicted by **5** for Poisson (shot noise) distributed intensities (solid line in the 1D R_{DA} histograms) supports the existence of at least two distinct subpopulations that differ only slightly in distance. Furthermore, we can rule out broadening caused by dye quenching, because the distribution of the overlapping peaks follows the theoretical $R_{DA}(\bar{\tau}_{D(A)})$ curve. Species *II* is characterized by an R_{DA} of ~ 87 Å. This distance indicates that the substrate is far from the nucleic acid binding cleft and does not correspond to any previously known structure. The observed distance is slightly smaller than the maximum reasonable value, $R_{DA}(\text{Max}) \sim 101$ Å (dotted line in the 1D R_{DA} histograms). This value is determined by measuring a sample of donor-labeled RT in complex with unlabeled dp/dt (Fig. 2B). Because there is no acceptor in the donor-only labeled RT:dp/dt complexes, there is no FRET, and subsequently, there should be no red signal. However, background, dark counts, and crosstalk of donor signal still contribute to a mean red signal of 1.4 kHz of “apparent acceptor.” From this signal, we can calculate a noise-limited maximum discernible separation distance for the DA pair used here (therefore, larger distances are shaded in gray). In agreement with the theoretical prediction (dotted line in the 1D R_{DA} histogram), the broadness of the peak is due to the small numbers of red photons in each burst, causing the intensity ratio F_D/F_A to vary greatly with a change of only a few red photons.

To check whether the differences in R_{DA} observed for the three complexes might be caused by species-specific orientational effects of the dyes, the corresponding histograms of $\bar{\tau}_{D(A)}$ vs. r_D were analyzed (Fig. 2 Lower). Overlaid are lines calculated from the Perrin equation (7) computed for two fixed rotational correlation times of 7 and 2 ns (blue lines). In Fig. 2A, species *Ia* and *Ib* have a broad ρ_D -distribution of rotational correlation times ranging from 1 to 7 ns and having a center at ≈ 2 ns, which indicates predominantly high mobility and validates the assumed value of $2/3$ for κ^2 . We suggest that species *I* contains two or more subspecies with similar positions, representing different steps in the polymerization reaction. The different rotational correlation times indicate different donor dye mobilities, perhaps because of clamping of the finger and thumb domains during polymerase activity. In contrast, the dye mobility in species *II* is significantly reduced ($\rho_D \sim 7$ ns). Selective time-resolved anisotropy analysis permits estimation of worst-case scenarios of possible DA-distance errors caused by orientational effects. The obtained maximum errors of ± 3 Å for species *Ia* and *Ib* and ± 16 Å for species *II* are too small to explain the observed distance difference of 39 Å (see Fig. 6, which is published as supporting information on the PNAS web site).

Further evidence for the existence of these two different principle classes of RT:dp/dt complexes was obtained by measuring the specific dissociation rates, k_d , in a displacement experiment. A 1,000-fold excess of unlabeled dp/dt was added to the solution, and the amount of each species exhibiting specific FRET was followed over time. The fluorescence parameter used to distinguish the species is the ratio of donor to acceptor fluorescence intensity, F_D/F_A . A time-dependent decrease in both RT:cy5-(5'-dp/dt) complexes is observed together with a concomitant increase in RT:dp/dt complex containing only donor (Fig. 3). Kinetic analysis provides dissociation rates for species *I* and *II*, which differ by a factor of 2: $k_d(I) = 6.2 \pm 0.3 \times 10^{-3} \text{s}^{-1}$ and $k_d(II) = 2.6 \pm 0.5 \times 10^{-3} \text{s}^{-1}$ (8).

The activity of RT under single-molecule conditions was examined by performing selective incorporation of four nucleotides, and terminating the chain with a dideoxynucleotide (Fig. 2C). The extension of the primer increases the DA distance and shifts species *I* to a larger R_{DA} , indicating that this complex is the

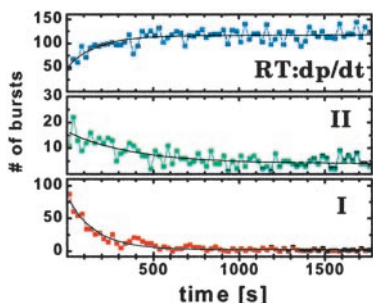


Fig. 3. Dissociation of the acceptor-labeled DNA from the donor-labeled RT. Monitoring three characteristic F_D/F_A ranges for the species in Fig. 2 [species *I*: 0.7–3.8; species *II*: 5.7–14.1; species RT:dp/dt (unlabeled): 19.5–291.7], the number of molecules detected in 25-sec intervals is plotted vs. time after a 1,000-fold excess of unlabeled dp/dt was added to the solution. The number of bursts of species *I* and *II* decrease with two different characteristic dissociation rate constants k_d ($6.2 \times 10^{-3} \text{ s}^{-1}$ and $2.6 \times 10^{-3} \text{ s}^{-1}$) accompanied by a corresponding increase in the number of molecules of the trapped species (donor-labeled RT with unlabeled DNA).

enzymatically productive one. The measured distance change of 13 Å corresponds with the expected extension of four base pairs of B-DNA. Additionally, ρ_D is low (2 ns), indicating a mobile dye. The distance for species *II* remains unchanged in the presence of nucleotides. An interesting feature of experiments in which the primer molecule was extended by four nucleotides (Fig. 2C) is that in contrast to the starting situation (Fig. 2A), the agreement for the end product between the theoretical R_{DA} -Gaussian distribution and the experimental 1D-histogram of species *I* is excellent. We hypothesize that the broadly distributed species *I* consists of two different conformations. After termination of the chain with a dideoxy-nucleoside triphosphate, the presence of the next cognate nucleotide in solution stalls the RT:dp/dt complex in a single conformation.

To test the hypothesis that species *I* consists of two subpopulations, we observed the activity of RT when bound to a preterminated Cy5-(d.dp/dt). In the presence of a 200 μM mixture of all four nucleoside triphosphates [deoxy-nucleoside triphosphates (dNTPs); Fig. 4A], the entire population is shifted into a single, Gaussian-distributed R_{DA} -peak at 51 Å with a high rotational mobility (solid horizontal line). This behavior is analogous to that seen in the extension experiment (Fig. 2C) and probably corresponds to a state in which RT is prepared to accept the dNTP for incorporation, but incorporation is prevented by the terminal dideoxynucleotide. In the presence of 200 μM sodium pyrophosphate, NaPP_i, the population is again shifted, but this time toward shorter distances (Fig. 4B, dashed horizontal line). In addition, the R_{DA} -peak is broader than a single Gaussian, and ρ_D remains high and broadly distributed. As sketched in Fig. 4, we suggest that the first and major complex with a shot-noise limited R_{DA} -distribution, which we refer to as the productive complex in its educt state (*Ia*: P-E), is capable of facile incorporation of nucleotides at a rate constant of $\approx 50 \text{ s}^{-1}$, as measured previously (8), and probably corresponds to the structure seen in the x-ray crystallographic analysis of Huang *et al.* (6). The second complex, species *Ib*, is characterized by a broad R_{DA} distribution in which the mean distance between the dyes has shortened by $\sim 5 \text{ Å}$ (Fig. 4, black arrows and the dashed and solid lines).

The corresponding $\bar{\tau}_{D(A)}-r_D$ diagrams indicate that the mean rotational correlation times of the donor dyes in the two species, $\rho_D(Ia) = 2 \text{ ns}$ and $\rho_D(Ib) = 7 \text{ ns}$, differ significantly, which poses the question whether, assuming the same κ^2 , the calculation of R_{DA} values is justified. Thus, donor and acceptor anisotropy decays were analyzed by selective subensemble spectroscopy of

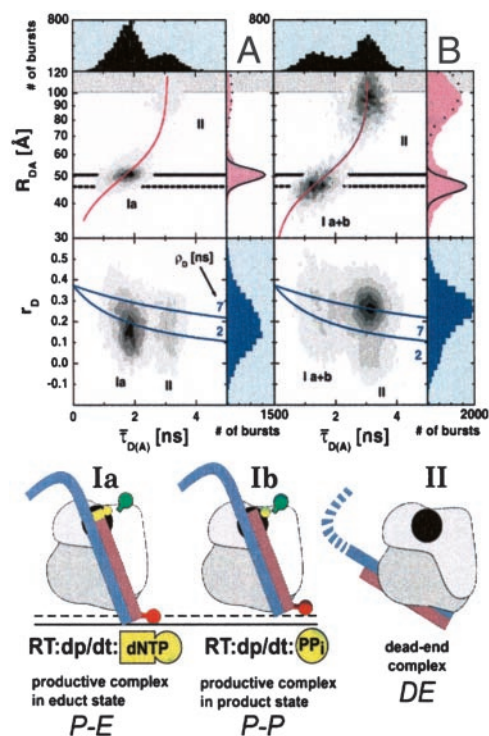


Fig. 4. (A) Addition of 200 μM dNTPs to a complex of RT:d.dp/dt. The $\bar{\tau}_{D(A)}-R_{DA}$ histogram shows only species *Ia* and *II*. The productive complex (*Ia*) interacts with the dp/dt in a state closely resembling the known RT:dp/dt structures. Here, the dNTP is thought to occupy a binding site in the polymerization active site and is, therefore, prebound for incorporation into the primer strand (purple). This state is the productive complex in the educt state (P-E). The solid black lines in the histogram at $R_{DA} = 51 \text{ Å}$, and below in the cartoon, indicate the position of the dp/dt bound in the P-E state. The p66 subunit is colored light gray, and the p51 subunit is colored dark gray. The polymerase active site of p66 is colored black. The fluorescence dyes Alexa⁴⁸⁸ and Cy5 are indicated by balloons colored green and red, respectively. In the P-E state, a dNTP occupies the binding pocket. In the one-dimensional R_{DA} histogram, the distribution of species *Ia* is observed to be shot-noise limited. (B) Addition of 200 μM sodium pyrophosphate (NaPP_i) to a complex of RT:d.dp/dt. The presence of PP_i moves the peak toward shorter distances by 5 Å, indicated by species *Ib*. The presence of PP_i shifts the primer terminus into the binding pocket, forming the productive complex in the product state, P-P. The dashed black lines in the histogram at 46 Å, and below in the cartoon, indicate the position of the dp/dt bound in the product state, in the presence of the PP_i shifting the dp/dt into the binding cleft. The peak is not Gaussian-distributed, and the rotational correlation time, ρ_D , remains high. The position of species *II* remains unchanged, indicating a dead end (DE) complex. Preliminary results indicate that the nucleic acid substrate in the DE complex is bound at a site on the p51 subunit, far removed from the nucleic acid-binding tract observed by crystallography. The cartoon indicates a suggested binding orientation for the DE complex.

species *Ia* and *Ib* (Supporting Text, Section II). For both species, the low rotational correlation time of the acceptor allows us to exclude the possibility that the observed changes are due solely to different κ^2 values; instead, they are caused primarily by a difference in distance between the dyes.

The results shown in Fig. 4 are consistent with a model in which the second complex, *Ib*, must first undergo an isomerization step before nucleotide incorporation can occur and would represent a “slipped” structure in which the dp/dt is bound in the correct orientation but is not correctly positioned in the binding cleft for nucleotide binding (8). An attractive possibility is that the 3' end of the primer has moved further into the polymerase active site, so that the terminal nucleotide is now occupying the binding site for incoming nucleoside triphosphate. This state is

actually the expected state occurring in the polymerization mechanism immediately after nucleotidyl transfer and release of PP_i . Therefore, the complex Ib is referred to as a productive complex in its product state, $P-P$. Before the next nucleotide binding can occur, the primer-template $P-P$ complex has to move to free the active site, i.e., by generating the productive complex in its educt state, Ia , with a larger R_{DA} , as indicated in the sketch of Fig. 4. According to this interpretation, the slower of the two fast phases seen in the nucleotide incorporation kinetics (8) corresponds to this movement, which is an essential step that must occur after nucleotide incorporation and represents the fundamental translocation step in the polymerization reaction. This step seems to be correlated with the previously reported phase occurring at $\approx 2 \text{ s}^{-1}$ (8), after which binding of the next nucleotide and incorporation occurs at $\approx 50 \text{ s}^{-1}$, the rate constant of the fastest phase seen in the kinetic experiments (8).

The third complex, species II , seen in Figs. 2 and 4, appears to have a quite different structure, as indicated by a broad R_{DA} distribution peaking at a long DA distance of $\approx 87 \text{ \AA}$. We shall refer to species II as the dead-end complex, DE , because such a complex was originally postulated to explain the slowest kinetic phase observed for single nucleotide incorporation (8). In view of the very different R_{DA} , the failure to incorporate nucleotides rapidly, and specific slower dissociation rates (Fig. 3), we suggest that the DE complex has a completely different configuration to the productive species and, as shown in the earlier kinetic work, that the dp/dt substrate must first dissociate from the enzyme and rebind before incorporation can occur, a process which seems to be very slow under the conditions used here. One suggestion pertaining to its structure which can be made at present is that the simplest conceivable possibility, i.e., that the dp/dt is bound with the wrong polarity, with the blunt end of the double-stranded region instead of the 3' end of the primer at the polymerase active site, can be excluded, because such a complex would have a high rather than low FRET efficiency.

In future work, a large number of protein mutants with single reactive cysteine residues can be prepared together with dp/dt hybrids labeled in different positions to obtain a sufficiently large number of distance constraints from experiments on RT: dp/dt complexes to allow construction of a structural model. From this

work, expect results supporting a binding mode of the primer/template not located in the DNA-binding cleft like the productive complex seen in the x-ray structure. The observed binding should indicate a conformation for the dead-end complex with the substrate bound to the p51 subunit. One possible orientation for DE is shown in Fig. 4. Furthermore, there is an indication for a conformational change of RT compared with the conformation adopted in the productive complex.

It is not yet clear that all three complexes identified are physiologically relevant. For the two productive complexes, probable structures have been assigned, and it is very likely that they represent two complexes occurring in the polymerase mechanism. It is possible that the DE complex identified here and in kinetic studies is not of biological significance, but it is equally possible that it reflects a mode of interaction of RT with its nucleic acid substrate which might be of importance in the complex process of reverse transcription, which involves interaction with a large number of different types of nucleic acid substrates and includes the process of strand transfer (28).

The results reported here demonstrate the power of measuring several fluorescence parameters simultaneously at the single-molecule level. The new two-dimensional FRET analysis removes the problem of broadening caused by local dye quenching and allows for direct DA distance determination with subnanometer spatial resolution. Although the existence of several RT: dp/dt complexes had been inferred indirectly from kinetic experiments, it would not have been possible to obtain direct physical evidence on the nature of the complexes by using ensemble measurements. Future studies of complex assemblies of biological macromolecules, which is one of the most important remaining challenges in biology, will be greatly facilitated by methods of the type described here.

We thank E. Schweinberger for assistance with the experiments, E. Hausteiner for analytical software, Michael Hess for help with Fig. 1, and Axel Scheidig for helpful discussions. R.S.G. and C.A.M.S. acknowledge the Volkswagenstiftung. R.S.G. acknowledges the Deutsche Forschungsgemeinschaft (SFB 1524) and the Max Planck Society for financial support. C.A.M.S. acknowledges the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BioFuture Grant 0311865).

- Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1992) *Science* **256**, 1783–1790.
- Restle, T., Müller, B. & Goody, R. S. (1990) *J. Biol. Chem.* **265**, 8986–8988.
- Restle, T., Müller, B. & Goody, R. S. (1992) *FEBS Lett.* **300**, 97–100.
- Sarafianos, S. G., Das, K., Tantillo, C., Clark, A. D., Ding, J., Whithcomb, J. M., Boyer, P. L., Hughes, S. H. & Arnold, E. (2001) *EMBO J.* **20**, 1449–1461.
- Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., et al. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6320–6324.
- Huang, H., Chopra, R., Verdine, G. L. & Harrison, S. C. (1998) *Science* **282**, 1669–1675.
- Boyer, P. L., Sarafianos, S. G., Arnold, E. & Hughes, S. H. (2001) *Proc. Natl. Acad. Sci. USA* **97**, 3056–3061.
- Wöhrl, B. M., Krebs, R., Goody, R. S. & Restle, T. (1999) *J. Mol. Biol.* **292**, 333–344.
- Eggeling, C., Berger, S., Brand, L., Fries, J. R., Schaffer, J., Volkmer, A. & Seidel, C. A. M. (2001) *J. Biotechnol.* **86**, 163–180.
- Lu, H. P., Xun, L. & Xie, X. S. (1998) *Science* **282**, 1877–1882.
- Moerner, W. E. & Orrit, M. (1999) *Science* **283**, 1670–1676.
- Xie, X. S. & Trautman, J. K. (1998) *Annu. Rev. Phys. Chem.* **49**, 441–480.
- Weiss, S. (2000) *Nat. Struct. Biol.* **7**, 724–729.
- Herten, D. P., Tinnefeld, P. & Sauer, M. (2000) *Appl. Phys. B* **71**, 765–771.
- Kask, P., Palo, K., Fay, N., Brand, L., Mets, Ü., Ullmann, D., Jungmann, J., Pschorr, J. & Gall, K. (2000) *Biophys. J.* **78**, 1703–1713.
- van der Meer, B. W., Cooker, G. & Chen, S. Y. (1994) *Resonance Energy Transfer: Theory and Data* (VCH, New York).
- Zhuang, X., Kim, H., Pereira, M. J. B., Babcock, H. P., Walter, N. G. & Chu, S. (2002) *Science* **296**, 1473–1476.
- Russell, R., Zhuang, X., Babcock, H. P., Millett, I. S., Doniach, S. & Chu, S. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 155–160.
- Zhuang, X., Bartley, L. E., Babcock, H. P., Russell, R., Ha, T., Herschlag, D. & Chu, S. (2000) *Science* **288**, 2048–2051.
- Dahan, M., Deniz, A. A., Ha, T. J., Chemla, D. S., Schultz, P. G. & Weiss, S. (1999) *Chem. Phys.* **247**, 85–106.
- Deniz, A. A., Laurence, T. A., Beligere, G. S., Dahan, M., Martin, A. B., Chemla, D. S., Dawson, P. E., Schultz, P. G. & Weiss, S. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 5179–5184.
- Ha, T., Zhuang, X. W., Kim, H. D., Orr, J. W., Williamson, J. R. & Chu, S. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 9077–9082.
- Kensch, O., Restle, T., Wöhrl, B. M., Goody, R. S. & Steinhoff, H. J. (2000) *J. Mol. Biol.* **301**, 1029–1039.
- Kensch, O., Connolly, B. A., Steinhoff, H. J., McGregor, A., Goody, R. S. & Restle, T. (2000) *J. Biol. Chem.* **275**, 18271–18278.
- Widengren, J., Schweinberger, E., Berger, S. & Seidel, C. A. M. (2001) *J. Phys. Chem. A* **105**, 6851–6866.
- Lakowicz, J. R. (1999) *Principles of Fluorescence Spectroscopy* (Kluwer Academic/Plenum, New York).
- Schaffer, J., Volkmer, A., Eggeling, C., Subramaniam, V., Striker, G. & Seidel, C. A. M. (1999) *J. Phys. Chem. A* **103**, 331–336.
- Negroni, M. & Buc, H. (2001) *Nat. Cell Biol.* **2**, 151–155.