A toolkit and benchmark study for FRET-restrained high-precision structural modeling

Stanislav Kalinin[1], Thomas Peulen[1,5], Simon Sindbert[1,5], Paul J Rothwell[1,2], Sylvia Berger[1], Tobias Restle[2,3], Roger S Goody[2], Holger Gohlke[4] & Claus A M Seidel[1]

We present a comprehensive toolkit for Förster resonance energy transfer (FRET)-restrained modeling of biomolecules and their complexes for quantitative applications in structural biology. A dramatic improvement in the precision of FRET-derived structures is achieved by explicitly considering spatial distributions of dye positions, which greatly reduces uncertainties due to flexible dye linkers. The precision and confidence levels of the models are calculated by rigorous error estimation. The accuracy of this approach is demonstrated by docking a DNA primer-template to HIV-1 reverse transcriptase. The derived model agrees with the known X-ray structure with an r.m.s. deviation of 0.5 Å. Furthermore, we introduce FRET-guided 'screening' of a large structural ensemble created by molecular dynamics simulations. We used this hybrid approach to determine the formerly unknown configuration of the flexible single-strand template overhang.

In recent years, single-molecule fluorescence spectroscopy has come of age. Single-molecule FRET (smFRET) serves as a 'spectroscopic ruler'[1] by reporting distance information between donor and acceptor fluorophores placed within a certain proximity. The method is providing important insights into the structural heterogeneity and function of biomolecules under *in vitro*[2] and *in vivo*[3] conditions. Yet a common misconception about FRET is that it is too inaccurate for structural modeling. The uncertain fluorophore positions with respect to their attachment points and the orientation dependence of FRET efficiency (the 'κ^2 problem') are considered fundamental limitations. We argue that explicitly modeling dye behavior[4–7] is the key to increasing the spatial resolution of FRET with flexibly linked dyes, thus making it a truly quantitative tool. This requires calculating the distribution of dye positions while considering the structure of the biomolecule. In contrast, the conversion of FRET data into distances between the labeling sites results in an unnecessary loss of accuracy.

Several approaches to derive FRET-restrained structures have been published[8–17]. In particular, several groups[12,13,15] have successfully developed and applied FRET-guided rigid-body docking with restraints obtained from experiments on immobilized molecules. A complementary probabilistic data analysis approach termed the 'nano-positioning system'[10,11,18] aims to find fixed fluorophore positions consistent with experimental data. However, many questions remain. Although mean dye position displacement due to flexible linkers is well understood[4,6,13,19], other issues have received less attention. The effects of averaging FRET efficiency over donor-acceptor distance distributions, the potential impact of the spatial arrangement of structural units on dye distributions, and missing information on the quality of structural models as influenced by the uncertainties of 'input' FRET data are all important considerations. There is also little evidence as to the accuracy of FRET-restrained three-dimensional (3D) modeling, which could be gained through comparison to known structures. (To our knowledge, FRET-derived models were quantitatively tested against independent structural data only once[20], using low-resolution FRET data.) Finally, a productive combination of FRET and state-of-the-art *in silico* modeling to generate candidate model structures is needed.

We have established a comprehensive toolkit and pipeline for FRET-restrained high-precision structural modeling that considers the mobility and structural heterogeneity of biomolecules. To validate our approach, we performed a benchmark study with the DNA primer template in complex with a human immunodeficiency virus type 1 (HIV-1) reverse transcriptase (RT), showing that our model is within a 0.5-Å r.m.s. deviation from the crystal structure. FRET-guided screening of a large structural ensemble created by molecular dynamics (MD) simulations allowed us to characterize the previously unresolved flexible single-strand template overhang.

Software for FRET-restrained positioning and screening (FPS) is freely available in the **Supplementary Software** and at http://www.mpc.hhu.de/software.

RESULTS

Using a fast approximation of dye position distributions

There is ample theoretical[5,7,21,22] and experimental[23] evidence for the existence of donor-acceptor distance distributions due to dye linkers (**Supplementary Fig. 1**). **Figure 1** illustrates

---

[1]Lehrstuhl für Molekulare Physikalische Chemie, Heinrich-Heine-Universität (HHU), Düsseldorf, Germany. [2]Max-Planck-Institut für Molekulare Physiologie, Dortmund, Germany. [3]Institut für Molekulare Medizin, Universität zu Lübeck, Lübeck, Germany. [4]Institut für Pharmazeutische und Medizinische Chemie, HHU, Düsseldorf, Germany. [5]These authors contributed equally to this paper. Correspondence should be addressed to S.K. (stanislav.kalinin@uni-duesseldorf.de) or C.A.M.S. (cseidel@hhu.de).

RECEIVED 22 APRIL; ACCEPTED 5 OCTOBER; PUBLISHED ONLINE 11 NOVEMBER 2012 DOI:10.1038/NMETH.2222

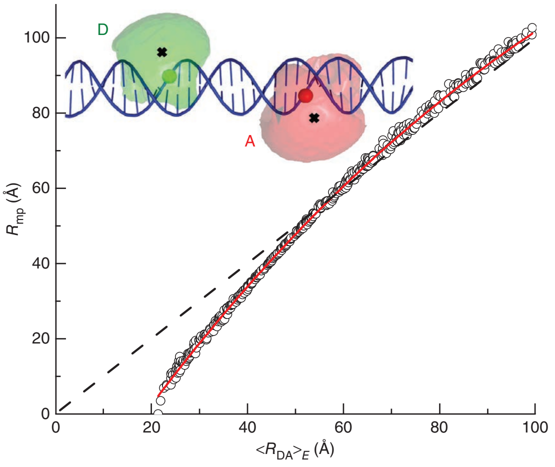

Figure 1 | Effect of dye position distributions on FRET efficiency. The distance between mean dye positions (R_{mp}) is plotted against the distance formally calculated from the FRET efficiency, $\langle R_{\mathrm{DA}}\rangle_E$. An $\langle R_{\mathrm{DA}}\rangle_E$-to-$R_{\mathrm{mp}}$ conversion function (red line) was generated by fitting a third-order polynomial to $\langle R_{\mathrm{DA}}\rangle_E/R_{\mathrm{mp}}$ value pairs (open circles) calculated for a set of randomly oriented accessible volumes (AVs) of Alexa 488 and Cy5 (R_0 = 52 Å) for dsDNA. The r.m.s. deviation between the data and the polynomial approximation is 0.9 Å over the whole range of $\langle R_{\mathrm{DA}}\rangle_E$ and 0.6 Å for 30 Å < $\langle R_{\mathrm{DA}}\rangle_E$ < 70 Å. The black dashed line serves as the reference for direct correlation with a slope of 1. The difference between R_{mp} and $\langle R_{\mathrm{DA}}\rangle_E$ can reach 10 Å in the range accessible to FRET (~30–70 Å). The DNA schematic shows the space accessible to Alexa 488 (green) and Cy5 (red) attached to a dsDNA via a C6 linker, calculated using the AV algorithm. The mean positions of the dyes are depicted as black crosses; the attachment atoms of the linkers, as spheres. The distance between the dyes' attachment points (the C5 atoms) is 43.6 Å, whereas the distance between the dyes' mean positions (R_{mp}) is 52.6 Å. D, donor and A, acceptor dyes.

the space accessible to the donor (D, Alexa 488) and acceptor (A, Cy5) fluorophores attached to DNA using standard C6 linkers. The mean dye positions are clearly far from attachment points and the accessible space is large, requiring that FRET observables be averaged over a distribution of donor-acceptor distances. The distance between mean dye positions, R_{mp}, and that calculated from the FRET efficiency, $\langle R_{\mathrm{DA}}\rangle_E$ (Online Methods), exhibit a large difference of up to 10 Å, or ~30% (**Fig. 1**). This confirms that considering distance distributions is essential for quantitative FRET, particularly when R_{mp} is below the Förster radius R_0.

However, the expected FRET efficiency can only be calculated with high precision if the spatial distributions of donor and acceptor are accurately predicted. Dye behavior can be realistically modeled using MD simulations[5,7,13], but these are time consuming. As a fast alternative with no sampling problems, in many cases a geometric accessible volume (AV) algorithm[11,24] can give a reasonable approximation[23] in under 1 s of computational time, making the screening of large structural ensembles feasible. The assumptions of the AV approach can be verified by analyzing complementary fluorescence parameters from a multiparameter fluorescence detection (MFD) experiment, such as fluorescence lifetime distribution and anisotropy, to rule out immobile dyes. Moreover, the FRET pair can be calibrated using molecules of known structure, such as dsDNA. We argue that systematic errors related to the AV approach are likely to average out if many donor-acceptor distances are measured, and we propose using the AV algorithm (**Supplementary Note 1**) for FRET-restrained positioning and screening. That way, ångström precision can be achieved when positioning labeled macromolecules, even if a single FRET measurement does not provide the distance between two labeling sites with high accuracy.

Workflow for FRET-restrained positioning and screening

Six steps are needed to generate a FRET-restrained structural model (**Fig. 2a**). (i) Generate a starting model using known structures, homology modeling or *ab initio* modeling. (ii) Design a network of dye positions from the starting model, taking the distributions of the coupled dyes into account. (iii) Measure FRET by single-molecule MFD (smMFD), thereby avoiding ensemble averaging; analyze photon distributions and time-resolved anisotropies of the dyes for rigorous data analysis and error estimation of donor-acceptor distances. (iv) Search for and evaluate possible structural models on the basis of their agreement with the FRET data by FPS. (We used two complementary approaches: discovering possible arrangements by rigid-body docking of known substructures and screening models in a large structure library.) (v) Rank models according to their violation of FRET restraints and cluster by similarity to judge their uniqueness and determine the corresponding confidence levels. (vi) Determine model precision by bootstrapping.

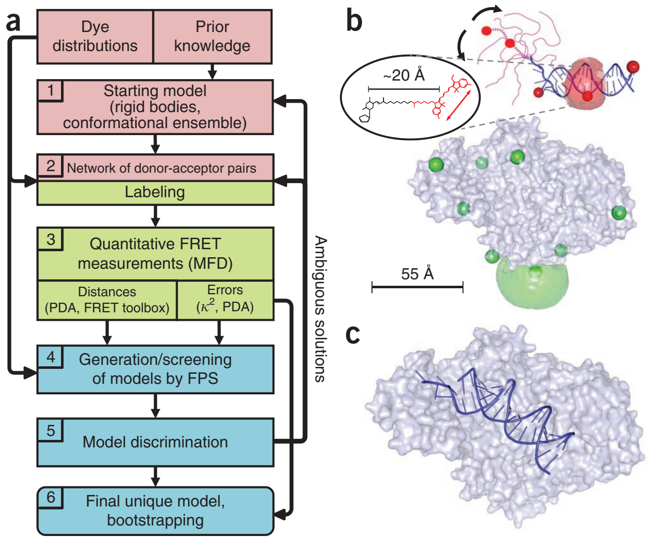

Figure 2 | FRET-restrained positioning and screening (FPS). (**a**) FRET-restrained high-precision structural modeling comprises three main parts: experimental design (pink shading), sample measurement and analysis (green), and generation and validation of structural models (blue). (**b**) Separate protein (gray) and dsDNA (blue) structures from the 1R0A crystal structure, and cartoon of the unresolved ssDNA overhang (magenta). Naming and details of the eight donor labeling positions (small green spheres) and five acceptor positions (small red spheres) are in the Online Methods, **Supplementary Figures 1 and 2** and **Supplementary Table 1**. The acceptor dyes (Cy5) are attached via C6 linkers (shown in ellipse). AV clouds are generated as in **Supplementary Note 1** and are shown for labeling positions p51E194C (large green volume) and dp(10) (large red volume). Parameters used for generation of AV clouds are given in **Supplementary Note 1**. (**c**) Structural model of the RT:dp/dt complex obtained by rigid-body docking.

We applied this workflow to HIV-1 RT in complex with a 19/35 DNA primer/DNA template (dp/dt) (**Fig. 2b**,**c**). RT is a heterodimer composed of a 66-kD chain (p66) and a 51-kD chain (p51) that transcribes viral RNA into dsDNA[25]. We characterized both the rigid double-stranded and flexible single-stranded DNA parts of the complex (**Fig. 2b**). Several crystal structures exist for the productive complex in the open educt state (P-E)[26,27], the state immediately before incorporation of the next nucleotide[28], thus allowing us to determine the accuracy of FPS for dsDNA (**Fig. 2c**). The ssDNA template overhang was previously unresolved by X-ray crystallography[25,29,30]. However, by combining FPS and MD simulations, we found a well-defined configurational space of the overhang preferentially interacting with one region of the protein.

Detailed workflow using HIV-1 RT case study

Here we demonstrate FRET-restrained high-precision structural modeling by applying our workflow to model dsDNA and ssDNA in the RT:dp/dt complex.

Step 1: starting models. The complex partners and the labeling positions are illustrated in **Figure 2b**. As prior knowledge, we chose the crystal structure with a 2.8-Å resolution from ref. 26 (PDB ID: 1R0A; justification given in **Supplementary Note 2**), where the RT:dp/pt complex is in the open educt state (P-E)[28].

To test the accuracy of FRET-restrained modeling, we separated the dp/dt from the protein and applied FPS. Alternatively, we docked the protein from 1R0A with a canonical B-DNA and relaxed the resulting complex structure by MD simulations. To determine the template overhang conformation missing in the crystal structure in the final step, we generated a starting model for the MD simulations by attaching the single strand to the crystallized DNA such that it projects straight out from the protein.

Step 2: network of donor-acceptor pairs. We chose eight donor and five acceptor label positions on the enzyme and the primer/template DNA (**Supplementary Table 1**), respectively (**Fig. 2b**). Overall, 36 independent smFRET measurements were performed for the RT:dp/dt complex.

Step 3.1: quantitative smFRET measurements by smMFD. smMFD experiments avoid ensemble averaging by analyzing single-molecule events. Distance information is usually deduced from the FRET efficiency (E)[31,32], which can be calculated either from the donor and acceptor fluorescence, F_{D} and F_{A}, or from donor fluorescence lifetimes in the presence ($\tau_{\mathrm{D(A)}}$) and absence ($\tau_{\mathrm{D(0)}}$) of the acceptor[33].

$$E=\frac{F_{\mathrm{A}}}{(\gamma' F_{\mathrm{D}}+F_{\mathrm{A}})}=1-\frac{\tau_{\mathrm{D(A)}}}{\tau_{\mathrm{D(0)}}},\quad \text{with}\quad \gamma'=\Phi_{\mathrm{FA}}/\Phi_{\mathrm{FD(0)}} \tag{1}$$

In equation (1), the correction factor γ' accounts for fluorescence quantum yields Φ_{F} of the donor and acceptor. In MFD, all fluorescence parameters are acquired simultaneously[33], which enables a multidimensional analysis. The correlated FRET analysis by equation (1) helps to avoid most pitfalls of ensemble FRET measurements, such as incomplete labeling, fluorophore quenching and the inability to resolve multiple FRET states[9,28,33].

In **Figure 3a**, 2D burst-frequency histograms of the $F_{\mathrm{D}}/F_{\mathrm{A}}$ signal ratio and donor anisotropy (r_{D}) versus $\tau_{\mathrm{D(A)}}$ are presented for the complex RT(p51<u>K173C</u>):dp(10)/dt. In agreement with ref. 28, three complex types were found: dead-end (D-E), productive complex in the product state (P-P) and productive complex in the educt state (P-E; see **Supplementary Note 3**). The observed populations followed the theoretically expected dependencies between $\tau_{\mathrm{D(A)}}$ and $F_{\mathrm{D}}/F_{\mathrm{A}}$ (**Supplementary Note 4**) or r_{D}, respectively. This indicates that no substantial dye quenching took place, which could have resulted in errors in the recovered R_{DA}, and that no long-lived immobile dye population was present (as characterized by a long mean rotation correlation time ρ). Our 2D FRET analysis of all other data sets can be found in **Supplementary Data 1**.

Here we calculate E from intensities (equation (1)) because F_{D} and F_{A} obey well-defined statistics[34,35], whereas the lifetime information ensures that the observed effects are due to FRET. Using both 2D analysis and photon distribution analysis (PDA; see step 3 below; analysis in the 1D $F_{\mathrm{D}}/F_{\mathrm{A}}$ histogram in **Fig. 3a**)[34,36] gives unsurpassed sensitivity for characterization of FRET populations derived from smFRET experiments. For the template overhang, we used distances obtained from sub-ensemble time-correlated single-photon counting (seTCSPC) data (**Supplementary Note 5** and **Supplementary Data 2**).

Step 3.2: input data for FPS—distances and uncertainties. PDA simultaneously provides mean distances $\langle R_{\mathrm{DA}}\rangle_E$ and uncertainties (ΔR_{DA}) by explicitly taking photon statistics into account. PDA needs fewer free parameters than the traditional approach of fitting multiple Gaussian peaks. It provides meaningful fit-quality parameters that allow one to justify the chosen fit model. To estimate the errors of fitted parameters, $\Delta R_{\mathrm{DA}}(E)$, we explored the parameter space for sets of variables providing acceptable fits (**Supplementary Note 4**). If E distributions are fitted by Gaussians using general-purpose fitting software, error estimation is usually difficult because the proper model function and the s.d. of data points are unavailable.

Uncertainties in the mutual orientation of donor and acceptor (κ^2 errors, $\Delta R_{\mathrm{DA}}(\kappa^2)$) can be estimated by analyzing anisotropy decays[37] accessible in MFD (**Supplementary Note 6** and **Supplementary Data 3**). A typical distribution of possible κ^2 values compatible with experimental anisotropy decays is shown in **Figure 3b**. Because of the weak dependence of R_0 on κ^2 (ref. 32), this broad distribution results in only 10% uncertainty in the distance. As a consequence, correcting $\langle R_{\mathrm{DA}}\rangle_E$ values for the mean κ^2 for all allowed orientations of D and A (**Fig. 3b**) yields almost the same final structures as the assumption of $\langle\kappa^2\rangle = 2/3$ (**Supplementary Note 6**). The overall uncertainties are determined following well-known error-propagation rules (Online Methods). This procedure was applied to all distances measured for the dsDNA and for the template overhang (**Supplementary Tables 2** and **3**).

For generating or judging FRET-based structure models, the knowledge of $\langle E\rangle$ and, thus, $\langle R_{\mathrm{DA}}\rangle_E$ is sufficient. These parameters can be calculated for any putative structure using AV or MD models of dye position distributions and are directly comparable with experimental data. In practice, it is useful to convert $\langle R_{\mathrm{DA}}\rangle_E$ into R_{mp} (**Fig. 1**) to avoid repeated calculations or transformations of AVs during iterative structure optimization (**Supplementary Note 7**).

To calculate $\langle R_{\mathrm{DA}}\rangle_E$ from the AV model, we assumed static averaging of distances and dynamic reorientation (not necessarily

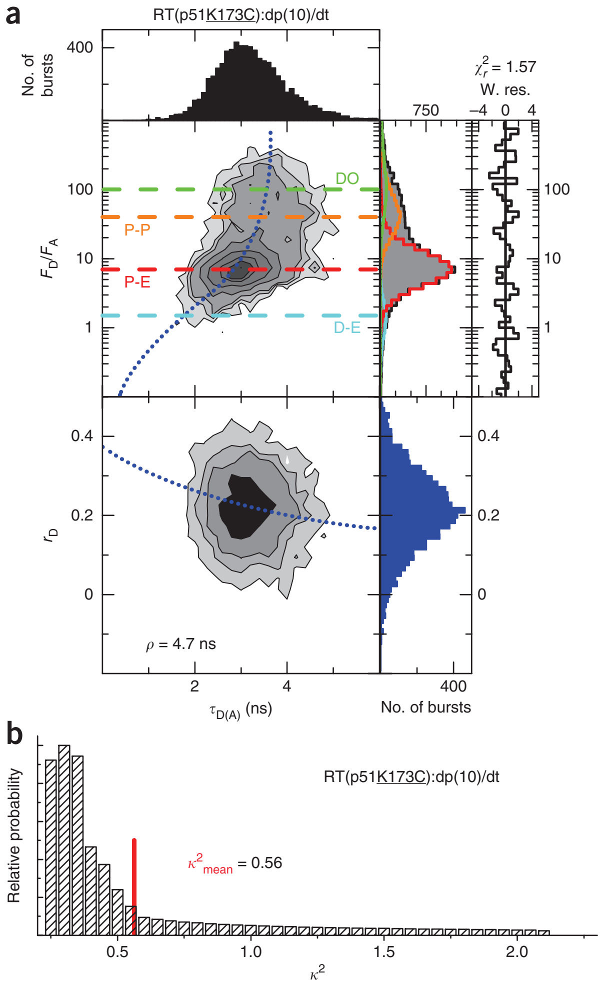

Figure 3 | ** Distance measurements and error estimation. (a**) 2D burst-frequency histograms of F_D/F_A versus donor fluorescence lifetime $\tau_{D(A)}$ (center left subpanel) and donor fluorescence anisotropy r_D versus $\tau_{D(A)}$ (bottom left subpanel) for RT(p51K173C):dp(10)/dt. The number of molecules (bursts) is scaled from white (lowest) to black (highest). 1D histograms are shown as projections. The dotted blue line (center) represents the theoretical relationship between F_D/F_A and $\tau_{D(A)}$ (static FRET line; **Supplementary Note 4**). The dotted blue line (bottom) represents the expected relationship between r_D and $\tau_{D(A)}$ and is given by the Perrin equation $r_D = r_0/(1+\tau_{D(A)}/\rho)$, with mean rotational correlation time ρ = 4.7 ns, r_0 = 0.374. The 1D F_D/F_A histogram is fitted (black line) by PDA using three FRET states with Gaussian distributed distances ($\sigma_{app} = 0.09 \times \langle R_{DA}\rangle_E$; **Supplementary Note 4**), 55% P-E $\langle R_{DA}\rangle_{E\ (P\text{-}E)}$ = 62 Å, 23% P-P $\langle R_{DA}\rangle_{E\ (P\text{-}P)}$ = 88 Å and 4% D-E $\langle R_{DA}\rangle_{E\ (D\text{-}E)}$ = 47 Å, and with 18% donor only (DO). The fit yields χ_r^2 = 1.57 (weighted residuals (w. res.) are plotted to the right of the 1D F_D/F_A histogram). FRET states: P-E, protein in educt state (red); P-P, protein in product state (orange); D-E, dead-end complex (cyan) (**Supplementary Note 3**). Dashed horizontal lines in the F_D/F_A versus $\tau_{D(A)}$ plot correspond to burst distributions of individual FRET states at right. (**b**) Distribution of possible κ^2 values for RT(p51K173C):dp(10)/dt derived from time-resolved anisotropy analyses. The uncertainty in κ^2 results in an error of 10.0% in $\langle R_{DA}\rangle_E$ (**Supplementary Note 6**).

resulting in $\langle\kappa^2\rangle = 2/3$) on the timescale of FRET. We have previously shown[6,23] that this approximation works well for dyes attached to DNA and RNA via long methylene linkers. Although the reorientation timescale is not as fast for donor dyes attached to RT (**Fig. 3a**), fast rotations of acceptor bound to DNA justify the assumption of dynamic κ^2 averaging.

Step 4a: rigid-body docking of dsDNA with FRET restraints. To find the position and orientation of dsDNA (from crystal structure or B-DNA) with respect to RT in best agreement with FRET data, the weighted data-model deviation (χ_E^2) has to be minimized for the set of n distances

$$\chi_E^2 = \sum_{i=1}^{n} \frac{\left(R_{DA(i)} - R_{model(i)}\right)^2}{\Delta R_{DA(i)}^2} \qquad (2)$$

This optimization problem can be defined for $\langle R_{DA}\rangle_E$ values or, more conveniently, for converted mean position distances R_{mp}, which are computed using a polynomial description (**Fig. 1** and **Supplementary Note 7**). To solve this problem, we assumed partners in the complex to be rigid bodies and applied a rigid-body dynamics approach to dock them using FRET restraints. Although a large number of more sophisticated approaches exist (for example, refs. 38,39), to our knowledge none allow for the explicit modeling of fluorophores and averaging of measured quantities over distributions of donor-acceptor distances.

Here we estimate the coordinates of mean dye positions by AV simulations and then fix them with respect to the labeled substructure. If a distance is measured between certain donor and acceptor dyes, this is accounted for by adding a 'spring' connecting the dyes' mean positions that has an equilibrium length of R_{mp} and a strength derived from the corresponding ΔR_{DA} (see **Supplementary Note 8** for implementation). Relaxing this system is equivalent to minimizing the χ_E^2 parameter given by equation (2). Two steps are distinguished in the rigid-body dynamics approach.

'Search': in the first step, we generate a large number of complexes starting from random configurations of the binding partners, excluding those with steric clashes. To prevent clashes between RT and DNA, we introduce strong repulsive forces between atoms approaching each other by a distance smaller than the sum of their van der Waals radii. In this way, the positioning is guided by an overall quality parameter (reduced χ_r^2; see Online Methods) that accounts for violations of FRET restraints and of van der Waals radii.

'Refinement': in the second step, AVs are recalculated, accounting for possible interactions (steric clashes) between the dyes and parts of the biomolecule structure that the dyes are not attached to. The resulting mean dye positions are used to reoptimize the structure.

Step 4b: screening of structural ensembles. An alternative strategy for finding a structure compatible with FRET data is to generate a large ensemble of putative structures (for example, by extracting conformations from an MD trajectory) and to filter the results with respect to agreement with FRET data by calculating χ_r^2 for each structure. Structures with a low χ_r^2 and good configurational quality in terms of MD are then selected. An obvious advantage of this approach is that state-of-the-art algorithms for conformational sampling can be used instead of rigid-body docking. Below we screened in two ways: (i) by refining a docked structure (B-DNA) that has been relaxed by MD simulations or (ii) by searching for a template overhang structure within a large ensemble, in the absence of prior knowledge.

Step 5: analysis of docking/screening results. We clustered structures obtained by docking into groups of solutions with similar

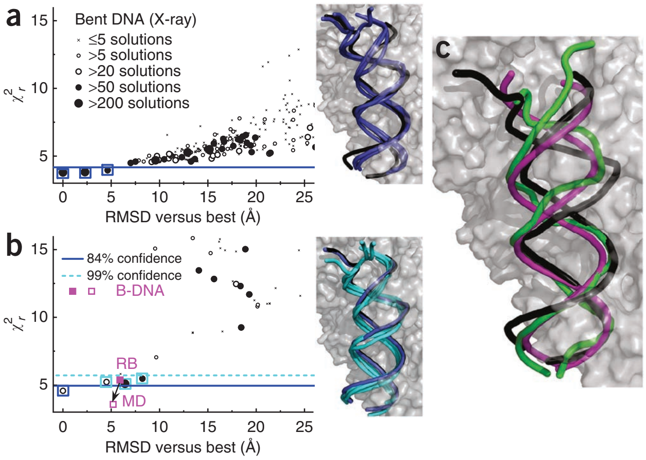

Figure 4 | Cluster analysis of docking results. (**a**,**b**) Rigid-body docking of RT and bent dp/dt (both from crystal structure 1R0A). Left, cluster analysis of obtained solutions for initial search phase with 6-Å clash tolerance (**a**) and second refinement iteration with 1-Å clash tolerance (**b**). Horizontal lines represent confidence thresholds as defined by $\chi_r^2 < \chi_{r,\min}^2 + 0.378$ (blue solid line, ~1.4 σ) and $\chi_r^2 < \chi_{r,\min}^2 + 1.09$ (cyan dashed line, ~2.6 σ). The thresholds are derived from the χ^2 distribution with 14 d.f. ($\chi^2 < 19.3$ and $\chi^2 < 29.2$, respectively). Right, DNA from X-ray structure (black; obscured in **b** owing to agreement with docked DNA) overlaid with FPS solutions (blue, cyan) indicated in cluster plots by colored squares. The cluster plot in **b** also shows the solutions for rigid-body docking B-DNA (RB, filled magenta square) and refinement of B-DNA via MD simulations (MD, open magenta square). The improvement with respect to agreement with FRET data through MD refinement is indicated by the black arrow. (**c**) Overlay of DNA from crystal structure (1R0A, black) with docked B-DNA (magenta) and MD relaxed B-DNA (green).

χ_r^2 values and low r.m.s. deviation within the group (**Fig. 4** and **Supplementary Note 9**). After a coarse search step (**Fig. 4a**; 6-Å clash tolerance), three groups of solutions were found below an 84% confidence threshold with respect to the best possible fit with $\chi_{r,\min}^2$ (**Fig. 4a**,**b** and Online Methods). In relation to the X-ray structure, the r.m.s. deviation values for these groups were 6.3 Å, 5.0 Å and 9.7 Å, respectively. After one refinement iteration consisting of recalculating AVs and reducing the clash tolerance to 2 Å, only one solution remained below the threshold (1.3-Å r.m.s. deviation; data not shown). A second refinement with 1-Å clash tolerance further separated the best structure from the other solutions, with an r.m.s. deviation value of only 0.5 Å with respect to the X-ray structure (**Fig. 4b**). The agreement was remarkably good considering the dye linker lengths of ~20 Å. Even if we required a significance level of >99% (**Fig. 4b**) and accept the solutions with r.m.s. deviations of ~4–8 Å, the deviation from the X-ray structure was still much smaller than the sum of dye linker lengths, which justifies the term 'high-precision FRET'. Compared to typical κ^2 uncertainties (~9%, **Supplementary Tables 2** and **3**), the resulting errors were also small, most likely because of averaging toward $\langle\kappa^2\rangle = 2/3$ when multiple donor-acceptor distances were considered; the same applies to possible systematic errors of the AV approach. Accounting for clashes had a clear effect on the selectivity, but even with unrealistically mild restrictions (as in the search step, **Fig. 4a**) FRET models already agreed well with the known structure (**Supplementary Note 10**).

As an independent validation, we applied rigid-body docking to straight B-DNA and the protein structure from 1R0A (**Fig. 4c** and **Supplementary Note 11**). Clearly, rigid-body docking could not account for bending of DNA; however, the resulting model agreed reasonably well with the X-ray structure (r.m.s. deviation = 5.9 Å). At the same time, because of its high χ_r^2 value of 5.36, this model could be rejected with ~95% confidence (**Fig. 4b**). This confirmed the high resolution power of FPS: FRET 'senses' that the shape of straight B-DNA is not optimal.

To find out whether the DNA becomes bent on binding, the docked RT:B-DNA complex was used as a starting model for relaxation by MD (**Supplementary Note 12**). Ten MD trajectories were screened for agreement with 20 FRET distances measured for the crystallized part of the complex. Screening of MD data revealed a clear correlation between χ_r^2 for FRET data and the MD agreement with the X-ray structure (**Supplementary Note 11**). In **Figure 4c**, a representative of the 50 best MD structures is shown. Relaxation by MD simultaneously improved the agreement with the X-ray structure (r.m.s. deviation = 5.2 Å, not accounting for the three bases of the single-stranded template overhang) and with FRET data (**Fig. 4b**), demonstrating the usefulness of combining MD simulations with high-precision FRET.

Step 6: estimation of precision and accuracy. The uncertainties of FRET-restrained modeling include possible alternative solutions (local χ_r^2 minima) with data-model deviations and uncertainties

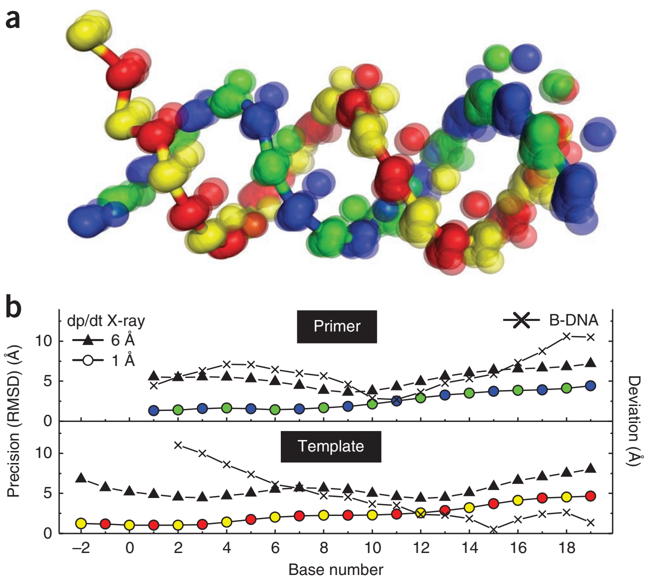

Figure 5 | Bootstrapping of docked dsDNA dp/dt. (**a**) Ensemble of structures generated with perturbed distances and a 1-Å clash tolerance. For better comparison, the phosphorus (P) atoms of the DNA backbone are alternately colored yellow/red or blue/green for the primer and template strands, respectively. The DNA is oriented as in **b**. (**b**) Uncertainty of P atom positions calculated for each P atom using equation (4) after the initial search with a 6-Å clash tolerance (triangles) and after two refinement steps with 2-Å (not shown) and 1-Å (circles) clash tolerances, respectively. The average r.m.s. deviation values for all P are 5.9 Å and 2.9 Å, respectively. The deviations between the docked B-DNA and the DNA of the crystal structure are also shown (crosses).

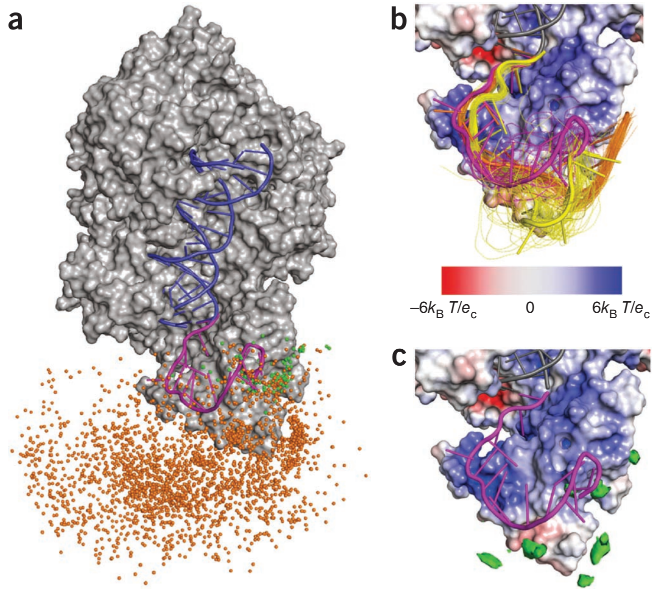

Figure 6 | Structure of ssDNA dp/dt overhang obtained by screening of MD trajectories. (**a**) Ensemble of ssDNA overhang structures generated by all-atom MD simulations in explicit solvent (orange; 2,855 conformations) and conformations filtered by sub-ensemble time-correlated single-photon counting (seTCSPC) using a confidence interval of 1 σ (green; 150 conformations). Dots represent the N1 atom of the nucleobase of the nucleotide dt(−15). The structure that best fits the FRET data is shown as a magenta cartoon ($\chi^2_r = \chi^2_E/16 = 0.88$). (**b**) Conformations within a confidence interval of 1 σ of the seTCSPC data. The PDB file with all conformers is given as **Supplementary Data 4**. The electrostatic potential of the protein as determined by adaptive Poisson-Boltzmann solver (APBS) has been mapped onto the protein surface using a color scale. (**c**) Green isopleths show regions of preferred residence of the N1 atom of the nucleobase of nucleotide dt(−15) as determined from MD simulations. The isopleths encompass regions with at least 2% of the maximal residence likelihood. Note the overall good agreement between these regions and the location of the 1-σ confidence interval determined by seTCSPC (**b**).

similar to those of the best solution itself. Assuming that a unique solution had been found at this point, we applied a procedure similar to bootstrapping[40] to estimate the precision for the best structure.

The model distance set calculated for the optimized structure (**Fig. 4b**) was perturbed by simultaneously adding normally distributed random numbers with a mean of 0 and an s.d. given by the experimental errors $\{\Delta R_{DA}\}$ (step 3.2) to all model distances. The structure was then reoptimized with the perturbed distance restraints. The procedure was repeated 100 times, yielding 100 perturbed structures for the double-stranded dp/dt part (**Fig. 5a**). This set of structures represents a distribution of possible positions of all backbone atoms consistent with experimental data. For each phosphorus (P) atom, the uncertainty was calculated (**Fig. 5b**) from the resulting structural ensemble (Online Methods, equation (4)). For the best FRET model, we estimated an average precision of 2.9-Å r.m.s. deviation for all P atoms of the dsDNA. The X-ray structure (0.5-Å r.m.s. deviation from the FRET structure, all dsDNA atoms) was well within these uncertainty limits. On the other hand, the model with docked B-DNA showed large deviations from the best solution and could be excluded (**Fig. 5b**). The solutions obtained at earlier stages of rigid-body docking were clearly less precise and less accurate, showing that the refinement steps are essential.

For MD-relaxed B-DNA, the uncertainty is given by the largest r.m.s. deviation value for any of the 50 structures below the χ_r^2 threshold with respect to the best solution (**Supplementary Note 11**). This estimation yielded 4.1-Å r.m.s. deviation for P atoms. Thus, given the differences between the structures (**Fig. 4b**) and the precision of individual models, the FRET (2.9 Å for rigid-body docked dsDNA, **Fig. 5**; 4.1 Å for MD-relaxed B-DNA) and the X-ray (2.8 Å; **Supplementary Note 2**) structures could not be distinguished within the limits of precision. However, it is clear that the B-DNA was bent by binding to a significant extent (**Fig. 4b**).

Alternatively, cross-validation or similar tests[41] can be performed. However, in contrast with X-ray or NMR data, the redundancy of FRET data is usually low. New labeling positions are often chosen to distinguish between ambiguous solutions (**Fig. 2a**), and discarding a few FRET restraints might make the position of a unit completely undefined. For this reason, bootstrapping is preferred over procedures in which some data points are completely discarded. For RT, both error estimation methods work well (**Supplementary Note 13**).

Extension to flexible parts of the complex

We generated a conformational ensemble of the template overhang (**Fig. 2b**), applying all-atom MD simulations in explicit solvent[42]. Ten trajectories (**Fig. 6a**) were filtered using 16 distances determined for the acceptor positions dp/dt(−6) and dp/dt(−15) on the template overhang (**Supplementary Table 4**). In **Figure 6a** the structure of the overhang that best fits the FRET data is depicted together with the approximate 1-σ confidence interval (150 conformations) representing the N1 atom of the nucleobase of nucleotide dt(15). The conformational ensemble satisfying FRET data is represented by three major configurations in **Figure 6b** and **Supplementary Data 4** preferentially located in positively charged regions. The isopleths for the N1 atom determined by MD simulations (**Fig. 6c**) illustrate good agreement between these regions and the structures satisfying FRET restraints. The structures wind around the fingers domain, with the 5′ end binding to RT in a loop structure in positively charged protein regions. Although we cannot exclude the existence of other minor conformer populations for template overhang, seTCSPC data showed no excessive broadening of donor-acceptor distance distributions measured for the template overhang as compared to the dsDNA part (**Supplementary Note 14** and **Supplementary Tables 4** and **5**). Thus, a substantial population of free unbound overhang can be excluded.

Whereas interactions of the single-strand template region in close proximity to the primer terminus directly affect active site geometry and, thus, fidelity of the enzyme, the effects of template-RT interactions beyond position −6 are not immediately obvious. However, during reverse transcription, the enzyme has to pass regions with extended secondary and even short double-stranded structures. It is conceivable that extensive interactions between RT and the template relatively far upstream of the site of nucleotide incorporation help to resolve such obstacles.

DISCUSSION

FPS improves the precision of FRET-derived structures dramatically. In combination with advanced computer simulations, it allows for a detailed molecular description of the proposed structure models. For this hybrid approach, we established a procedure to characterize the uniqueness and precision of FRET-restrained models that is based on a precise spectroscopic estimation of 'input' uncertainties. With the single-molecule advantage of FRET, heterogeneous systems can be investigated easily. Our approach takes advantage of MFD data of freely diffusing molecules. This has two distinct advantages: (i) data are free of problems due to protein immobilization and surface artifacts, and (ii) statistics are better and single-molecule events are perfectly averaged.

Moreover, the FPS toolkit includes an interface for predicting donor-acceptor distances for a given structural model and labeling positions, which is indispensable for planning FRET experiments and comparing them with the experimental results.

Currently, we derive our knowledge of biomolecular structure largely from traditional methods such as X-ray crystallography that determine highly resolved but static models. However, biomolecules are dynamic and undergo intrinsic motions[43,44].

With MFD, FRET has a key advantage: it allows for the observation of multiple biomolecular conformations in solution[9] with high time resolution determined by the fluorescence lifetime of the dyes on the order of a few nanoseconds. Moreover, together with filtered fluorescence correlation spectroscopy[45], FRET harbors the potential to study conformational control of biomolecular function in complex systems and associate it with detailed dynamic structures without averaging.

Our application of FPS to the flexible single-strand template overhang of RT:dp/dt revealed a preferential structure with the 5′ end of the overhang bound to the fingers domain of RT, corroborating the finding that overall dp/dt binding affinity to RT increases with the overhang length by a factor of 7 (ref. 29). There is biochemical evidence that the properly bound template overhang plays an important role in translocation of nucleic acids during processive DNA synthesis[25], for example, by helping to resolve secondary structures within the substrate and by proper alignment of the primer terminus within the active site, thus affecting fidelity of DNA synthesis.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We dedicate this paper to the memory of our brilliant colleague, Robert M. Clegg, a pioneer in the application of FRET in the life sciences and a remarkable human being. We would like to thank E. Schweinberger, O. Kensch and B.M. Wöhrl for assistance with the experiments, E. Haustein for analytical software, H. Sanabria for help with data visualization and A. Scheidig for helpful discussions. We acknowledge financial support from the Volkswagen Foundation (to R.S.G. and C.A.M.S.; grant no. I/74470), Max-Planck Society (to R.S.G.), German Federal Ministry of Education and Research (BMBF) (BioFuture grant no. 0311865) and German Science Foundation within SPP 1258 (grant no. SE 1195/12-2) (to C.A.M.S.) and 'Fit for Excellence' initiative of HHU (to H.G.). T.P. thanks the International Helmholtz Research School of Biophysics and Soft Matter (IHRS BioSoft), and S.S. thanks the NRW Research School of Biological Structures in Molecular Medicine and Biotechnology (BioStruct) for scholarships.

AUTHOR CONTRIBUTIONS

S.K. and C.A.M.S. designed studies and assisted in data analysis and interpretation; P.J.R. and S.B. prepared samples and performed experiments; S.K. developed the software; T.P. and S.S. analyzed the data; H.G. performed MD simulations and assisted with data interpretation; T.R. and R.S.G. supervised the biochemical experiments and assisted with data collection and interpretation; S.K., T.P., S.S., T.R., H.G. and C.A.M.S. wrote the manuscript; C.A.M.S. supervised the project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Stryer, L. Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem.* **47**, 819–846 (1978).
2. Ha, T. *et al.* Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci. USA* **93**, 6264–6268 (1996).
3. Sakon, J.J. & Weninger, K.R. Detecting the conformation of individual proteins in live cells. *Nat. Methods* **7**, 203–205 (2010).
4. Merchant, K.A., Best, R.B., Louis, J.M., Gopich, I.V. & Eaton, W.A. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proc. Natl. Acad. Sci. USA* **104**, 1528–1533 (2007).
5. Best, R.B. *et al.* Effect of flexibility and *cis* residues in single-molecule FRET studies of polyproline. *Proc. Natl. Acad. Sci. USA* **104**, 18964–18969 (2007).
6. Woźniak, A.K., Schröder, G., Grubmüller, H., Seidel, C.A.M. & Oesterhelt, F. Single-molecule FRET measures bends and kinks in DNA. *Proc. Natl. Acad. Sci. USA* **105**, 18337–18342 (2008).
7. Hoefling, M. *et al.* Structural heterogeneity and quantitative FRET efficiency distributions of polyprolines through a hybrid atomistic simulation and Monte Carlo approach. *PLoS ONE* **6**, e19791 (2011).
8. Mekler, V. *et al.* Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108**, 599–614 (2002).
9. Margittai, M. *et al.* Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *Proc. Natl. Acad. Sci. USA* **100**, 15516–15521 (2003).
10. Andrecka, J. *et al.* Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **105**, 135–140 (2008).
11. Muschielok, A. *et al.* A nano-positioning system for macromolecular structural analysis. *Nat. Methods* **5**, 965–971 (2008).
12. Choi, U.B. *et al.* Single-molecule FRET-derived model of the synaptotagmin 1-SNARE fusion complex. *Nat. Struct. Mol. Biol.* **17**, 318–324 (2010).
13. Brunger, A.T., Strop, P., Vrljic, M., Chu, S. & Weninger, K.R. Three-dimensional molecular modeling with single molecule FRET. *J. Struct. Biol.* **173**, 497–505 (2011).
14. Sabir, T., Schroder, G.F., Toulmin, A., McGlynn, P. & Magennis, S.W. Global structure of forked DNA in solution revealed by high-resolution single-molecule FRET. *J. Am. Chem. Soc.* **133**, 1188–1191 (2011).
15. McCann, J.J., Zheng, L.Q., Chiantia, S. & Bowen, M.E. Domain orientation in the N-terminal PDZ tandem from PSD-95 is maintained in the full-length protein. *Structure* **19**, 810–820 (2011).
16. Balci, H., Arslan, S., Myong, S., Lohman, T.M. & Ha, T. Single-molecule nanopositioning: structural transitions of a helicase-DNA complex during ATP hydrolysis. *Biophys. J.* **101**, 976–984 (2011).
17. Boura, E. *et al.* Solution structure of the ESCRT-I complex by small-angle X-ray scattering, EPR, and FRET spectroscopy. *Proc. Natl. Acad. Sci. USA* **108**, 9437–9442 (2011).
18. Muschielok, A. & Michaelis, J. Application of the nano-positioning system to the analysis of fluorescence resonance energy transfer networks. *J. Phys. Chem. B* **115**, 11927–11937 (2011).
19. Olofsson, M., Kalinin, S., Zdunek, J., Oliveberg, M. & Johansson, L.B.Å. Tryptophan-BODIPY: A versatile donor-acceptor pair for probing generic changes of intraprotein distances. *Phys. Chem. Chem. Phys.* **8**, 3130–3140 (2006).
20. Knight, J.L., Mekler, V., Mukhopadhyay, J., Ebright, R.H. & Levy, R.M. Distance-restrained docking of rifampicin and rifamycin SV to RNA polymerase using systematic FRET measurements: developing benchmarks of model quality and reliability. *Biophys. J.* **88**, 925–938 (2005).
21. Dolghih, E., Roitberg, A.E. & Krause, J.L. Fluorescence resonance energy transfer in dye-labeled DNA. *J. Photochem. Photobiol. A Chem.* **190**, 321–327 (2007).
22. VanBeek, D.B., Zwier, M.C., Shorb, J.M. & Krueger, B.P. Fretting about FRET: Correlation between kappa and R. *Biophys. J.* **92**, 4168–4178 (2007).

23. Sindbert, S. *et al.* Accurate distance determination of nucleic acids via Förster resonance energy transfer: implications of dye linker length and rigidity. *J. Am. Chem. Soc.* **133**, 2463–2480 (2011).
24. Cai, Q. *et al.* Nanometer distance measurements in RNA using site-directed spin labeling. *Biophys. J.* **93**, 2110–2117 (2007).
25. Götte, M., Rausch, J.W., Marchand, B., Sarafianos, S. & Le Grice, S.F.J. Reverse transcriptase in motion: Conformational dynamics of enzyme-substrate interactions. *Biochim. Biophys. Acta* **1804**, 1202–1212 (2010).
26. Peletskaya, E.N., Kogon, A.A., Tuske, S., Arnold, E. & Hughes, S.H. Nonnucleoside inhibitor binding affects the interactions of the fingers subdomain of human immunodeficiency virus type 1 reverse transcriptase with DNA. *J. Virol.* **78**, 3387–3397 (2004).
27. Huang, H., Chopra, R., Verdine, G.L. & Harrison, S.C. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**, 1669–1675 (1998).
28. Rothwell, P.J. *et al.* Multi-parameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 reverse transcriptase:primer/template complexes. *Proc. Natl. Acad. Sci. USA* **100**, 1655–1660 (2003).
29. Patel, P.H. *et al.* Insights into DNA polymerization mechanisms from structure and function analysis of HIV-1 reverse transcriptase. *Biochemistry* **34**, 5351–5363 (1995).
30. Upadhyay, A.K., Talele, T.T. & Pandey, V.N. Impact of template overhang-binding region of HIV-1 RT on the binding and orientation of the duplex region of the template-primer. *Mol. Cell. Biochem.* **338**, 19–33 (2010).
31. Lakowicz, J.R *Principles of Fluorescence Spectroscopy* 2nd edn. (Springer, 2006).
32. Van Der Meer, B.W., Coker, G. III. & Chen, S.-Y.S. *Resonance Energy Transfer: Theory and Data*. (Wiley, 1994).
33. Sisamakis, E., Valeri, A., Kalinin, S., Rothwell, P.J. & Seidel, C.A.M. Accurate single-molecule FRET studies using multiparameter fluorescence detection. *Methods Enzymol.* **475**, 455–514 (2010).
34. Antonik, M., Felekyan, S., Gaiduk, A. & Seidel, C.A.M. Separating structural heterogeneities from stochastic variations in fluorescence resonance energy transfer distributions via photon distribution analysis. *J. Phys. Chem. B* **110**, 6970–6978 (2006).
35. Gopich, I.V. & Szabo, A. Single-macromolecule fluorescence resonance energy transfer and free-energy profiles. *J. Phys. Chem. B* **107**, 5058–5063 (2003).
36. Kalinin, S., Felekyan, S., Valeri, A. & Seidel, C.A.M. Characterizing multiple molecular states in single-molecule multiparameter fluorescence detection by probability distribution analysis. *J. Phys. Chem. B* **112**, 8361–8374 (2008).
37. Dale, R.E., Eisinger, J. & Blumberg, W.E. The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. *Biophys. J.* **26**, 161–193 (1979).
38. Lyskov, S. & Gray, J.J. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res.* **36**, W233–W238 (2008).
39. van Dijk, M. & Bonvin, A. Pushing the limits of what is achievable in protein-DNA docking: benchmarking HADDOCK's performance. *Nucleic Acids Res.* **38**, 5634–5647 (2010).
40. Efron, B. Discussion: jackknife, bootstrap and other resampling methods in regression-analysis. *Ann. Stat.* **14**, 1301–1304 (1986).
41. Brünger, A.T. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
42. Case, D.A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
43. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
44. Tokuriki, N. & Tawfik, D.S. Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).
45. Felekyan, S., Kalinin, S., Sanabria, H., Valeri, A. & Seidel, C.A.M. Filtered FCS: species auto- and cross-correlation functions highlight binding and dynamics in biomolecules. *ChemPhysChem* **13**, 1036–1053 (2012).

ONLINE METHODS

All samples, applied methods and supporting data are described in detail in the supplementary information. All data used for FPS are compiled in the **Supplementary Tables 2–5**.

Materials. Sample preparation and labeling were done according to ref. 46. We used RT mutants containing single accessible cysteine positions that were labeled with Alexa 488 as donor dye. Cy5 was used as acceptor dye and was attached to different positions on a 19/35 DNA primer/DNA template. Further details are given in the **Supplementary Methods** and **Supplementary Figure 1**. The Alexa 488–Cy5 dye pair has a Förster radius $R_0 = 52$ Å for $\langle \kappa^2 \rangle = 2/3$.

The donor labeling positions (green spheres) are named according to the RT subunit (p51 or p66) and the position of the introduced cysteine mutation to which they were coupled: p51Q6C, p51K173C, p51E194C, p51K281C p66Q6C, p66T27C, p66E194C and p66K287C. Five acceptor labeling positions on the primer/template DNA (red spheres) are named according to strand (dp or dt) and the position of the labeled nucleotide. The last paired nucleotides are referred to as position 1. For labeled positions on the template overhang, the position is negative with regard to position 1, and for the primer-coupled fluorophores, the number is positive (**Supplementary Fig. 2**). Thus, the labeled dp/dt complexes are named dp(1)/dt, dp(10)/dt, dp(19)/dt, dp/dt(−6) and dp/dt(−15) (**Supplementary Table 1**). As an example, FRET-pairs are named RT(p51E194C):dp/dt(−15).

Experimental conditions. The aqueous measurement buffer contained 10 mM KCl, 50 mM Tris·HCl, pH 7.8, 6 mM $MgCl_2$ and 400 μM ascorbic acid. See **Supplementary Methods** for further details.

Experimental methods. The multiparameter fluorescence detection (MFD) is performed using an inverted confocal microscope with excitation by a pulsed laser at 496 nm. Fluorescence detection is performed with an additional pinhole defining a detection volume of 2 femtoliters (fl). Sample molecules diffusing through the detection volume cause a brief (~1-ms) burst of fluorescence. Dilute solutions of molecules (~50 pM) ensure that only single molecules are detected. The fluorescence signal is divided into parallel and perpendicular components and in wavelength ranges below and above 620 nm (green and red, respectively), and single photons are detected by four avalanche photodiodes. For each photon, the arrival time after the laser pulse, the time since the last photon, the polarization, and the wavelength range are recorded. Fluorescence bursts are distinguished from the background signal of 1–2 kHz by applying certain threshold intensity criteria[47]. See **Supplementary Methods** for further details.

Analysis methods. Distances and their uncertainties (**Supplementary Tables 2** and **3**) were determined by PDA and time-resolved sub-ensemble anisotropies. The positional distribution of certain species was analyzed by sub-ensemble time-correlated single-photon counting (**Supplementary Note 5** and **Supplementary Tables 4** and **5**).

Toolkit for FPS. Software is available for download at http://www.mpc.hhu.de/software. Details on FPS, especially on the generation and screening of models, are given in **Supplementary Note 8**. The software can use all types of distance restraints, including, for example, FRET distances derived from surface smFRET or ensemble TCSPC experiments.

Methods for rigid-body docking of dsDNA with FRET restraints in step 4. (For more details see **Supplementary Note 8**.)

Input distances for FPS. Typically the mean FRET efficiency $\langle E \rangle$ is measured in an experiment. However, because of different averaging of E and the D-A distance R_{DA}, $\langle E \rangle$ is not directly related to the distance between mean dye positions R_{mp} (refs. 6,23). However, a formal distance, referred to as donor-acceptor FRET-averaged distance $\langle R_{DA} \rangle_E$, is calculated by $\langle E \rangle = 1/(1+\langle R_{DA} \rangle_E^6/R_0^6)$. The Förster radius R_0 is calculated for $\langle \kappa^2 \rangle = 2/3$ unless stated otherwise. The optimization problem can be defined for R_{mp} or $\langle R_{DA} \rangle_E$ values. The first option is easier to implement in combination with rigid-body dynamics, whereas directly calculating deviations between experimental and model $\langle R_{DA} \rangle_E$ is more appropriate for structure screening. PDA and similar techniques[48,49] directly provide $\langle R_{DA} \rangle_E$. To obtain R_{mp} values, an $\langle R_{DA} \rangle_E$-to-R_{mp} conversion function can be generated (**Fig. 1** and **Supplementary Note 7**).

Search. In the first step, we generate a large number of rigid-body models that correspond to local minima of χ_r^2 (see equation (8.3) in **Supplementary Note 8**). For this, the optimization is started from a random configuration of the binding partners, excluding those with clashes between them. The Verlet algorithm[50] (**Supplementary Note 8**) is applied to model translational and rotational motions of units until the system is relaxed and certain convergence criteria are fulfilled. Although probably not the most efficient optimization algorithm, rigid-body dynamics is straightforward to use and allows exploration of the local minima. In addition, clashes between molecules are prevented by introducing strong repulsive forces between atoms approaching each other by a distance smaller than the sum of their van der Waals radii. In this way, rigid-body dynamics effectively minimizes the reduced chi-squared parameter that accounts for violations of FRET restraints and of van der Waals radii, $\chi_r^2 = (\chi_E^2 + \chi_{clash}^2)/(n - p)$, where n is the number of FRET restraints and p is the number of degrees of freedom—here, 6. Initially, clashes are allowed to some extent to ensure penetration of DNA into the nucleic acid binding cleft, and structures showing considerable clashes are filtered out later by a χ_r^2 threshold (see below). To ensure exhaustive sampling of the configurational space of the binding partners, docking with random initial conditions[15] is repeated many (here, 10^4) times.

Refinement. The second step is to remodel the AVs accounting for possible interactions (steric clashes) between the dyes and parts of the biomolecule structure the dyes are not attached to. These modified AVs are then used to calculate new mean dye positions, which is followed by a reoptimization of the structure. At the same time, we reduce the clash tolerance to make clashes between complex partners practically forbidden. This procedure can be repeated several times for each structure until new iterations do not further improve the agreement with experimental data or change the solution significantly.

Analysis of docking/screening results. Solutions are considered ambiguous if the respective χ_r^2 values do not differ significantly. We typically apply a threshold $\chi_r^2 < \chi_{r,\min}^2 + [2/(n-p)]^{1/2}$, which roughly corresponds to the variance of the chi-squared distribution of 2× (degrees of freedom)[51] (blue lines in **Fig. 4a**,**b**). The fact that $\chi_{r,\min}^2$ is often larger than 1 is attributed to systematic experimental errors and to possible limitations of the AV and/or rigid-body models. In this work, this problem is compensated for by oversampling. Other criteria defining different levels of significance can be applied here in a straightforward way.

Estimation of precision. In most of cases, a clear peak assignment to the P-P and P-E state is possible (**Supplementary Tables 2** and **3**). The overall 'input' distance uncertainties in equation (2) are then estimated following the propagation rules

$$\Delta R_{\mathrm{DA}(\pm)}^2(E,\kappa^2) = \Delta R_{\mathrm{DA}(\pm)}^2(E) + \Delta R_{\mathrm{DA}}^2(\kappa^2) \tag{3a}$$

In equation (3a), terms indicated with a ± subscript can be asymmetric with respect to R_{DA}. More advanced κ^2 estimation procedures[18,52] can be also conveniently incorporated into our set of tools. In four cases, the peak assignment (see **Supplementary Tables 2** and **3**) was unclear owing to comparable amplitudes. In these cases, the errors in distances also include this uncertainty. If two peaks with $R_{\mathrm{DA}(1)}$ and $R_{\mathrm{DA}(2)}$ (with $R_{\mathrm{DA}(1)} < R_{\mathrm{DA}(2)}$) have comparable amplitudes x_1 and x_2, the amplitude-weighted average distance $\langle \mathrm{R}\rangle_{E,x} = (a_1 R_{\mathrm{DA}(1)} + a_2 R_{\mathrm{DA}(2)})/(a_1 + a_2)$ is taken for modeling as R_{DA}, whereas the uncertainties are given by equation (3b)

$$\begin{aligned}\Delta R_{\mathrm{DA}(+)}(E,\kappa^2,\mathrm{assign}) &= R_{\mathrm{DA}(2)} + \Delta R_{\mathrm{DA}(2,+)}(E,\kappa^2) - \langle R\rangle_{E,x} \\ \Delta R_{\mathrm{DA}(-)}(E,\kappa^2,\mathrm{assign}) &= \langle R\rangle_{E,x} - (R_{\mathrm{DA}(1)} - \Delta R_{\mathrm{DA}(1,-)}(E,\kappa^2))\end{aligned} \tag{3b}$$

Procedures to estimate various contributions to the input uncertainties are described in detail in **Supplementary Note 15**.

For each phosphorus atom, the position uncertainty (described by the r.m.s. deviation) is estimated from an ensemble of perturbed structures using equation (4)

$$\mathrm{r.m.s.d.} = \left\langle \left| \vec{R}_{\mathrm{best\ model}} - \vec{R}_{\mathrm{perturbed\ model}} \right|^2 \right\rangle^{1/2} \tag{4}$$

MD simulations are described in **Supplementary Note 12**. The discrimination between models and the determination of quality parameters is described in **Supplementary Note 9**.

46. Rothwell, P.J. *Structural Investigations on HIV-1 RT Using Single Pair Fluorescence Energy Transfer* PhD thesis, Univ. Dortmund (2002).
47. Eggeling, C. *et al.* Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *J. Biotechnol.* **86**, 163–180 (2001).
48. Nir, E. *et al.* Shot-noise limited single-molecule FRET histograms: Comparison between theory and experiments. *J. Phys. Chem. B* **110**, 22103–22124 (2006).
49. Santoso, Y., Torella, J.P. & Kapanidis, A.N. Characterizing single-molecule FRET dynamics with probability distribution analysis. *ChemPhysChem* **11**, 2209–2219 (2010).
50. Martyna, G.J., Tobias, D.J. & Klein, M.L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).
51. Soong, T.T. *Fundamentals of Probability and Statistics for Engineers* (Wiley, 2004).
52. Isaksson, M., Norlin, N., Westlund, P.O. & Johansson, L.B.Å. On the quantitative molecular analysis of electronic energy transfer within donor-acceptor pairs. *Phys. Chem. Chem. Phys.* **9**, 1941–1951 (2007).