A nano-positioning system for macromolecular structural analysis

Adam Muschielok[1], Joanna Andrecka[1], Anass Jawhari[1,2], Florian Brückner[1,2], Patrick Cramer[1,2] & Jens Michaelis[1]

Very often, the positions of flexible domains within macromolecules as well as within macromolecular complexes cannot be determined by standard structural biology methods. To overcome this problem, we developed a method that uses probabilistic data analysis to combine single-molecule measurements with X-ray crystallography data. The method determines not only the most likely position of a fluorescent dye molecule attached to the domain but also the complete three-dimensional probability distribution depicting the experimental uncertainty. With this approach, single-pair fluorescence resonance energy transfer measurements can now be used as a quantitative tool for investigating the position and dynamics of flexible domains within macromolecular complexes. We applied this method to find the position of the 5′ end of the nascent RNA exiting transcription elongation complexes of yeast (*Saccharomyces cerevisiae*) RNA polymerase II and studied the influence of transcription factor IIB on the position of the RNA.

In recent years, high-resolution structural models of large macromolecular complexes such as the ribosome[1], the RecBCD helicase[2] or RNA polymerases[3,4] have been obtained using X-ray crystallography. Although these structures provide detailed insight into the molecular architecture of complex biological systems, the position of flexible domains can usually not be determined because of averaging effects.

Single-molecule methods, on the other hand, provide the possibility of directly obtaining structural information because they allow the study of real-time conformational changes of macromolecular complexes[5]. In combination with fluorescence resonance energy transfer (FRET)[6], a technique that has been termed a molecular ruler[7], one can in principle measure distances within a macromolecule in real-time. However, because of experimental problems such as variations in quantum yield[8] or dependence of FRET on the orientations of the two dye molecules[9], there are few examples in the literature of quantitative distance measurements[8,10–12] or position determination[13–16] using single-pair FRET (sp-FRET). Instead, these data are more often interpreted in a qualitative fashion monitoring conformational changes and length increases or decreases[17–20].

Using triangulation of several FRET distance measurements, it is possible to determine a previously unknown position[13–16,21–23]. Although these experiments are able to estimate the most likely position, they do not show how existing experimental uncertainties might affect the position determined. Therefore, these positions must be interpreted with great caution because one has no information about the experimental accuracy. In principle, one can conduct control measurements that provide validity tests of the position determined[14], but to arrive at a quantitative technique, experimental uncertainties must be taken into account.

Here we used bayesian parameter estimation[21], a probability-based analysis method, to compute the three-dimensional probability density function for the position of a so-called 'antenna dye molecule' (ADM) attached to a flexible domain within a macromolecular complex. By determining its position, we inferred the position of the flexible domain. We measured the FRET efficiencies between the ADM and several 'satellite dye molecules' (SDMs) attached to known positions. Within the bayesian framework we were able to account for errors (due to orientational effects) in the Förster radii determined, errors in the measured FRET efficiencies and uncertainties in the SDM positions (due to the attachment of the SDMs by means of flexible linkers). Moreover, we were able to account for geometric constraints such as occluded volumes and maximum distances. From the calculated probability density function, we directly obtained the most likely ADM position and the associated three-dimensional credibility regions reflecting the experimental uncertainties. This method has similarities with the Global Positioning System (GPS), and we therefore termed this technique 'Nano Positioning System' (NPS).

As an example, we used NPS to study the position of nascent RNA in yeast RNA polymerase II (Pol II) elongation complexes. We had previously used sp-FRET–based triangulation to follow the path of the nascent RNA and shown that RNA leaves the active-site cleft of the polymerase through the so-called exit tunnel and then follows a path across the dock domain[14].
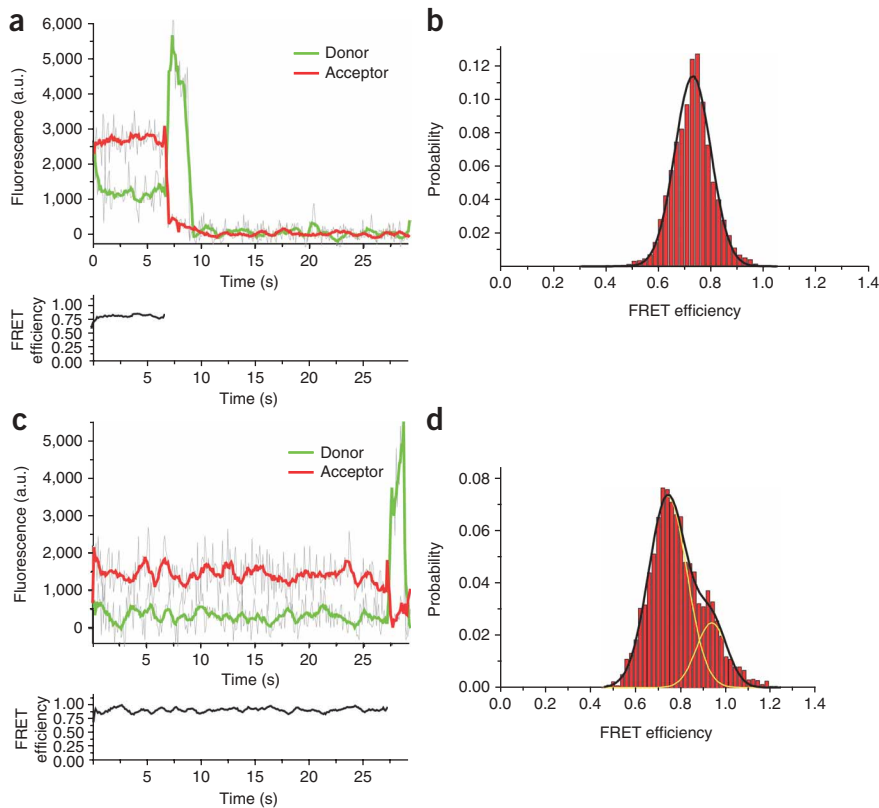
Figure 1 | Single-pair FRET time traces and histograms for RNA 29–Rpb7-S16C complexes. (**a–d**) Data recorded in the absence (**a,b**) and presence (**c,d**) of TFIIB. Displayed in **a,c** are the averaged fluorescence intensities of the donor (green) and the acceptor molecules (red) together with the unfiltered data (gray) and computed FRET efficiencies. The histogram of all complexes recorded in the absence of TFIIB (**b**) can be fitted by a single gaussian. In contrast, in presence of TFIIB (**d**), a double gaussian (yellow curves) is necessary to describe the data. The single-molecule fluorescence time trace in **c** belongs to a subpopulation of about 20% of all complexes that yielded a FRET efficiency substantially higher than that of complexes measured in the absence of TFIIB. Histograms for all other complexes are shown in **Supplementary Figures 2** and **3**. a.u., arbitrary units.

Here we applied NPS to determine the position of an ADM attached to the 3′ or 5′ end of a 29-nucleotide (nt)-long RNA in a Pol II elongation complex. Moreover, we have investigated how the latter position is influenced by the presence of transcription factor IIB (TFIIB). The results allow us to explain discrepancies between recent RNA chemical cross-linking experiments[22] and single-molecule fluorescence data on the position of the RNA[14], and they have implications for our understanding of the transition from transcription initiation to elongation.

RESULTS
Single-pair FRET measurements
We studied fluorescently labeled Pol II transcription elongation complexes as described recently[14]. We attached a donor dye that acted as the ADM to the 5′ end of a 29-nt-long RNA molecule (RNA 29). The acceptor dyes used as SDMs were either attached to the template DNA or at one of various positions on the hetero-dimer Rpb4/7 (**Supplementary Fig. 1** online). Elongation complexes were assembled, purified, immobilized and studied in a custom-built total internal reflection fluorescence (TIRF) microscope[14,20]. A typical time trace of donor fluorescence intensity, acceptor fluorescence intensity and the computed FRET efficiency

is shown in **Figure 1a**. We collected several hundred of these time traces and constructed histograms (**Fig. 1b**) from which we extracted the mean FRET efficiency. Histograms of all SDM-ADM pairs investigated are shown in **Supplementary Figures 2** and **3** online. All data are summarized in **Table 1**.

We had previously measured the FRET efficiencies between a donor dye attached to the 5′ end of the RNA and an acceptor dye attached to the template DNA at positions −10 (DNA 1) and +3 (DNA 2), and to Rpb7 at positions Rpb7-Cys150 and Rpb7-Cys94 (ref. 14). Here, we also present measurements for two more acceptor positions, attached to Rpb7-S16C and Rpb4-S73C, giving us a total of six different FRET measurements.

To study the effect of TFIIB on the position of the RNA, we added excess amounts of TFIIB to the immobilized elongation complexes in the sample chamber. In the presence of 3-, 10- and 100-fold (12 nM) excesses of TFIIB, we observed altered FRET efficiencies[14] for a fraction of the molecules (∼20% for 100-fold excess) (**Fig. 1c**). The computed FRET histogram of complexes in the presence of TFIIB can be fitted with a double gaussian (**Fig. 1d**); that is, there are two possible conformations of the RNA in the Pol II elongation complexes. We performed control experiments wherein we preincubated Pol II with a sixfold excess of TFIIB before the addition of nucleic acids and Rpb4/7 and obtained similar results (**Supplementary Fig. 4** online). We also verified that TFIIB was capable of binding to the prepared elongation complexes (**Supplementary Fig. 5** online). Furthermore, we performed experiments with a TFIIB construct in which the B-finger (residues 62–83) had been deleted. Again, we observed a fraction of the complexes with altered FRET values leading to similar double-peaked histograms (**Supplementary Fig. 4**). Notably, we did not observe dynamic switching between the two FRET levels in the single-molecule traces; the single-molecule FRET efficiency could always be associated with one of the two peaks in the histogram. Therefore, TFIIB leads to a different but stable conformation of the RNA in a fraction of the complexes.

To test the accuracy of NPS in determining a known position, we attached the ADM to the 3′ end of the RNA (RNA 1) and performed measurements with the acceptor dye attached to either DNA 1, DNA 2, template DNA at position + 9 (DNA 3), Rpb7-Cys150 or Rpb7-S16C.

Owing to experimental uncertainties, only qualitative distance information can be obtained from sp-FRET measurements. A quantitative analysis requires consideration of the uncertainties associated with the sp-FRET measurements, so we developed a probability-based analysis method to determine the distances

Table 1 | Förster radii and measured FRET efficiencies

SDM (Alexa 647) labeling site	ADM labeling site	Isotropic Förster radius (Å)	FRET efficiency (%)	Number of molecules
DNA 1	RNA 1[a]	47	50.7 ± 0.35	221
DNA 2	RNA 1[a]	47	80.0 ± 0.14	322
DNA 3	RNA 1[a]	47	61.0 ± 0.40	356
Rpb7-Cys150	RNA 1[b]	57	41.5 ± 0.39	68
Rpb7-S16C	RNA 1[a]	48	39.4 ± 0.38	105
DNA 1	RNA 29[b] without TFIIB	62	67.3 ± 0.6	305
DNA 2	RNA 29[b] without TFIIB	62	42.2 ± 0.7	374
Rpb7-Cys150	RNA 29[b] without TFIIB	65	60.4 ± 0.6	211
Rpb7-Cys94	RNA 29[b] without TFIIB	66	52.2 ± 1.1	202
Rpb7-S16C	RNA 29[b] without TFIIB	62	73.6 ± 0.7	144
Rpb4-S73C	RNA 29[b] without TFIIB	63	52.9 ± 2.1 / 75.2 ± 2.9[c]	78
DNA 1	RNA 29[b] with TFIIB	62	38.2 ± 0.8	392
DNA 2	RNA 29[b] with TFIIB	62	63.0 ± 2.7	243
Rpb7-Cys150	RNA 29[b] with TFIIB	65	77.0 ± 1.0	191
Rpb7-Cys94	RNA 29[b] with TFIIB	66	71.2 ± 4.3	222
Rpb7-S16C	RNA 29[b] with TFIIB	62	93.6 ± 4.1	129

[a]ADM: Alexa 555. [b]ADM: tetramethylrhodamine. [c]Two maxima were observed, which is consistent with two possible SDM positions (**Supplementary Fig. 1**).

between the SDMs and the ADM, as well as the position of the ADM, using the sp-FRET experiments above.

Bayesian parameter estimation applied to NPS

If there were no measurement errors, the measurement of one dye pair would limit the position of the ADM to the surface of a sphere, with the SDM position being the center and the measured distance being the radius. Hence, the position of the ADM should be the intersection of as many spheres as dye pairs were measured. For three distance measurements, it is easy to calculate the intersection points of the spheres that correspond to the constraints imposed by each distance measurement. However, because of errors in the Förster radii determined and the measured FRET efficiencies as well as uncertainties in the SDM positions caused by flexible linkers, the true position of the ADM can deviate substantially from the intersection point. For any given experimental error, the uncertainty in actual position depends strongly on the angle between the sphere surfaces at the intersection point (as illustrated in two dimensions, **Fig. 2a,b**). In the case of four or more distance measurements, the spheres will generally not have a common
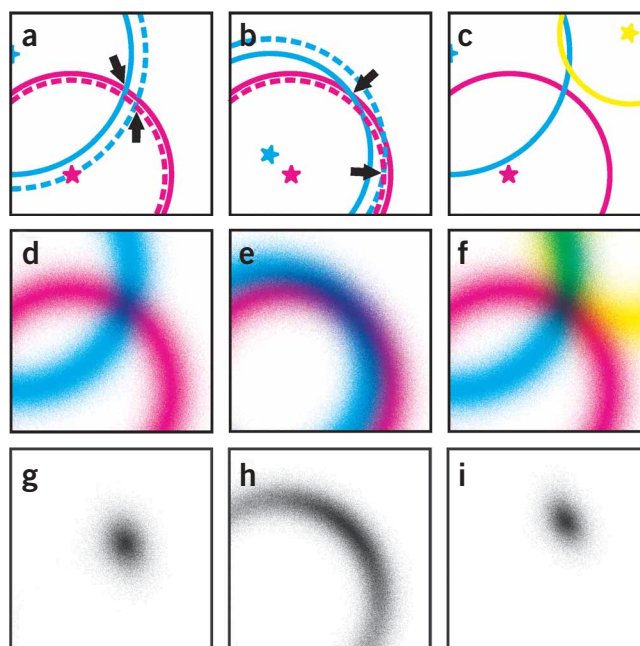
intersection point (**Fig. 2c**). Thus, two questions arise: how to analyze an arbitrary number of distance measurements consistently, and how to assign an error to the calculated position of the ADM. It is obvious that the hard constraints (spheres) must be softened to take the error sources into account. If we assume that the uncertainties of the SDM positions are isotropic, then the constraints will be spherically symmetric and we end up with fuzzy spheres (**Fig. 2d–f**). Mathematically, these constraints can be described by probability densities and one can calculate their overlap (**Fig. 2g–i**), which is defined as the product of the individual densities and has a shape that reflects our knowledge of the ADM position.

The idea described is realized mathematically by bayesian statistics[21], a general theory that applies probability theory to data analysis. In the bayesian conception, information about variables is represented by probability distributions. In this sense, sharply peaked distributions correspond to precise knowledge, diffuse and broad distributions to vague knowledge.

To apply bayesian parameter estimation to NPS, we must define a physical model to be used to explain the experiments. In this model, we assume that the positions of the ADM and several SDMs are static and describe them by the parameters x and s_i, respectively (**Fig. 3**), where $i = 1 \ldots N$, N being the number of different SDM positions. Further, we assume Förster-type energy transfer between each ADM-SDM pair, which introduces the Förster radius R_i, a parameter that defines the length scale of the energy transfer[6].

In the next step of the analysis, we calculate the probability density of the model parameters given (i) the measured data, in our case the measured FRET efficiencies E_i, and (ii) the background information I. In general, the background information represents all knowledge about the experiment performed, such as the model used to describe the data or the constraints for the parameters—for example, in the case of NPS, the X-ray structure. Formally, this probability density is called the 'posterior', and we denote it as $p(x,\{s_i\},\{R_i\}|\{E_i\},I)$. In this notation, the vertical bar is read to mean 'given', the variables to the right of it are the conditions, and a
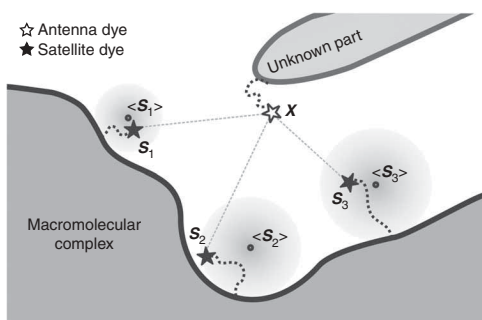

Figure 2 | The concept of fuzzy spheres, presented using the two-dimensional analog, fuzzy circles. (**a–c**) Schematic diagrams of the conventional approach for obtaining the intersection (indicated by arrows) of circles (cyan, magenta and yellow) with known centers (stars). In general, the circles with experimentally determined radii (solid lines) differ from the real situation (dashed lines). Depending on the angle of intersection, the uncertainty in position varies. If the circles intersect at ∼90°, the error is on the order of the accuracy of the distance measurement (**a**) but can be much larger if the angle is small (**b**). More than two circles may not have a common intersection at all (**c**); in three dimensions, this is the case for more than three spheres. (**d–f**) The overlay (purple in **d,e**, black in **f**) of fuzzy circles naturally reflects the uncertainty of the position determination for the situations shown in **a–c**, respectively. (**g–i**) The overlap of fuzzy circles shown in **d–f**, respectively. In **d–i**, the density of dots corresponds to the probability of the overlap.

☆ Antenna dye
★ Satellite dye

Figure 3 | Schematic representation of 'antenna' and 'satellite' dye positions. The antenna dye molecule (ADM; open star) at position x is attached to a part of the construct with unknown position. The prior of the ADM position is constant outside of the macromolecular complex (above the thick line) and zero in the interior of the macromolecular complex. Several satellite dye molecules (SDMs; filled stars) at positions s_i are attached to the macromolecular complex by flexible linkers. The shaded areas around the most likely SDM positions $<s_i>$ (small circles) represent the isotropic gaussian priors of the SDM positions s_i. The actual distances between the SDMs and the ADM are represented by dotted lines. From these distances, one can calculate the expected FRET efficiencies.

comma is read as the conjunction 'and'. The curly brackets denote a set of variables; for example, $\{s_i\} = s_1, \ldots, s_N$. Using Bayes' theorem[21], we can express the posterior by

$$p(x,\{s_i\},\{R_i\}|\{E_i\},I) \propto p(x,\{s_i\},\{R_i\}|I)\, p(\{E_i\}|x,\{s_i\},\{R_i\},I)$$
(1)

The expression $p(x,\{s_i\},\{R_i\}|I)$ is called the 'prior' and represents the knowledge of the model parameters given background information only (that is, before the experiment has been performed). In our case, it encodes the constraint of the ADM and SDM positions to the accessible volume, as well as the known Förster radii. The expression $p(\{E_i\}|x,\{s_i\},\{R_i\},I)$, the so-called 'likelihood', mathematically connects the measured data with the model parameters and represents our expectation of how the data E_i would look if we had to simulate it with the given set of parameters. Because the model parameters are independent, the prior may be factored into separate priors for each parameter. Similarly, the likelihood may be factored because the FRET measurements are independent (**Supplementary Methods** online).

According to equation (1), the posterior is a function of all SDM positions, all Förster radii and the ADM position. However, to obtain information about the ADM position only, we must consider all possible values of the Förster radii and SDM positions. Consequently, we integrate the posterior in all parameters but the ADM position to obtain $p(x|\{E_i\},I)$. By performing this calculation in the case of spherically symmetric SDM position priors (**Supplementary Methods**), we showed that the ADM position posterior decomposes into a product of spherically symmetric functions and the ADM position prior:

$$p(x|\{E_i\},I) \propto p(x|I)\Pi_i K_i(x)$$
(2)

The functions $K_i(x)$ represent the fuzzy spheres mentioned earlier, and their radial profiles display the uncertainty of the distance measurement. We generalized equation (2) to apply NPS to situations with arbitrarily shaped SDM position priors and use multiple gaussians to encode the SDM position (**Supplementary Methods**). In that case the functions $K_i(x)$ are no longer spherically symmetric but consist of a superposition of fuzzy spheres that themselves still possess radial profiles (**Supplementary Figs. 6** and **7** online).

To visualize the result, we calculated credibility regions for the ADM position (**Supplementary Methods**). These are the smallest volumes that enclose a probability equal to the specified credibility level. The surface of the credibility region may be displayed together with the macromolecular complex, in our case the Pol II elongation complex. In contrast to displaying the point of maximum probability only, the credibility region gives a direct impression not only of the estimated position of the ADM but also of the experimental uncertainties.
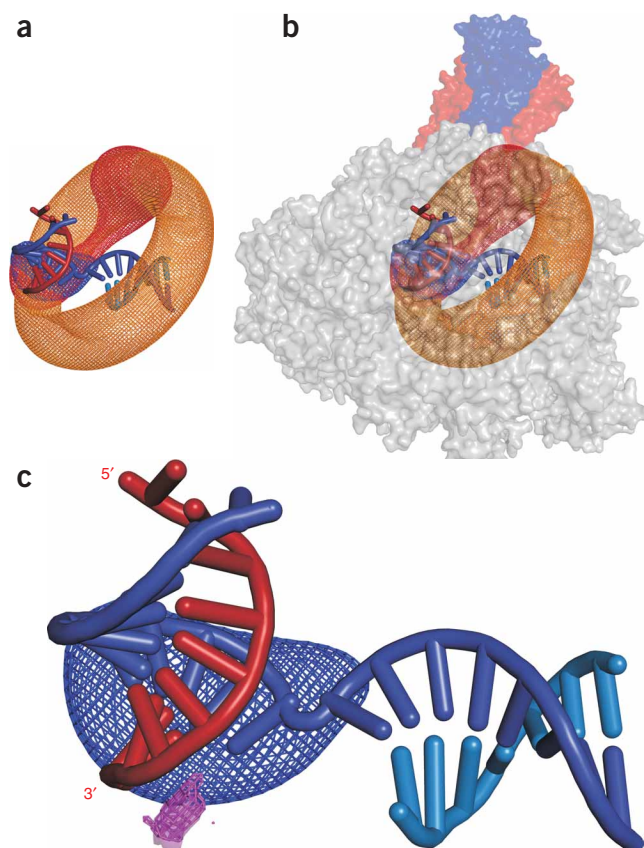

Figure 4 | Experimental test of NPS: the position of an ADM attached to the 3′ end of the RNA in a Pol II elongation complex. (**a**) Regions of 68% credibility for the ADM attached to the RNA 3′ end, determined using either three satellites (DNA 1, DNA 2, DNA 3; orange), four satellites (in addition, Rpb7-S16C; red) or five satellites (in addition, Rpb7-Cys150; blue), together with a cartoon representation of RNA (red), template DNA (dark blue) and non-template DNA (light blue). (**b**) Result of NPS inside the crystallographically determined[34] Pol II elongation complex (transparent surface representation; Protein Data Bank accession code 1Y1W). (**c**) Comparison of NPS result without (blue mesh) and with (magenta mesh) restriction of ADM position to free space within the Pol II complex when computing the accessible region of dye molecule (magenta surface) attached to 3′ end of RNA.
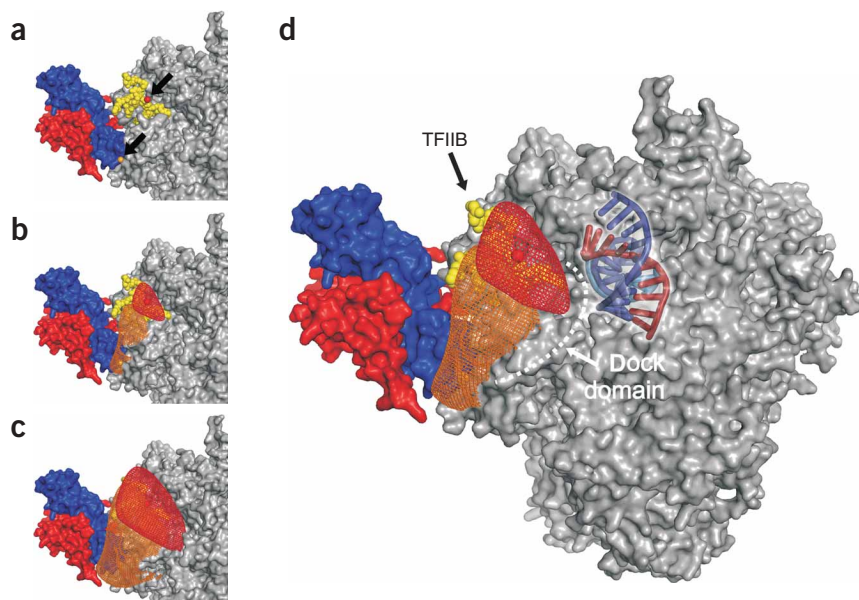
Figure 5 | Computed position of the ADM attached to 5′ end of the 29-nt RNA. (**a**–**c**) Diagrams showing the back view of a Pol II elongation complex[34] (Protein Data Bank accession code 1Y1W), together with our estimation of the ADM position displayed as a mesh surface. The ADM is attached to the 5′ end of RNA 29. A portion including the region around the dock domain is shown in **a**–**c**. In addition, we modeled the TFIIB position (yellow spheres) according to Pol II core–TFIIB structure[30] (Protein Data Bank 1R5U). Meshes illustrating the experimental uncertainty are shown for the ADM in the absence (red) and presence (orange) of TFIIB. Displayed are the most likely positions (**a**, highlighted by arrows), as well as credibility regions enclosing 38% (**b**), 95% (**c**) and 68% (**d**) probabilities.

the 5′ end of RNA 29 within the Pol II elongation complex structure and determine the location of the ADM (**Fig. 5**). A traditional localization analysis would only be able to determine the position of maximum probability (if all uncertainties had been included) (**Fig. 5a**). In contrast, NPS gives much more information and allows us to calculate 38%, 68% and 95% credibility regions (**Fig. 5b–d**).

Finally we analyzed the FRET data obtained in the absence and presence of TFIIB. Whereas in the absence of TFIIB the ADM attached to the 5′ end of RNA 29 is located just to the side of the dock domain of Pol II, the presence of TFIIB shifts the position much closer to Rpb4/7. The experimental uncertainties of the position without TFIIB, along the three principal directions, were computed to be ±3.0 Å, ±5.8 Å and ±8.7 Å and those in the presence of TFIIB were ±4.2 Å, ±5.6 Å and ±15.5 Å.

DISCUSSION

NPS uses single-molecule FRET measurements to determine the position of flexible domains within macromolecular complexes and accurately estimates the uncertainties in position, whereas previous methods determine the most likely position[13–16,22–26] only. As a result, one obtains the three-dimensional probability density of the ADM position. The accuracy of the position determination depends not only on the experimental uncertainties, but also on the number of satellites used and their relative orientation and separation. The largest experimental uncertainty arises from the variation of different relative orientations of the fluorophores. In the current model, we assumed no reorientations of the dye molecules on timescales longer than the fluorescence lifetime. Thus, dependent on the measured fluorescence anisotropy, a relatively large uncertainty can arise from different possible (mean) relative orientations of the two dye molecules. The respective experimental uncertainty could be reduced if (i) the orientations could be measured directly or (ii) the distribution of mean orientations were known. In principle, such information could be obtained by rotation of the excitation polarization[27] or by defocused imaging[28], and a combination of NPS with such polarization techniques will be an important future development.

The shape of the probability density does not imply any mobility of the ADM, as our model is a static one. A large or elongated density is only a consequence of the incomplete information gained in the experiment and is caused by the experimental uncertainties. Fast dynamic changes—for example, the frequent switching between two possible conformations—could potentially cause an improper position assignment. In principle, one could use methods with higher time resolution, such as time-correlated single photon counting or burst integrated fluorescence lifetime of freely diffusing complexes, to overcome this problem. In contrast, if dynamical changes occur that are slow compared to the time resolution of the

Experimental test of the NPS system

To test our method, we first analyzed the FRET data for the ADM attached to the known position at the 3′ end of the RNA (RNA 1) and determined how well NPS was able to locate this position. By comparing equal credibility regions in the computed ADM position probability density, one can see that the information about the ADM position increases with increasing number of satellites (**Fig. 4a,b**). Note that the relative positions of the satellites are important for the overall accuracy of the method (compare to **Fig. 2**).

Using an ADM at the 3′ end of the RNA illustrates how prior information can help in the localization. The surrounding protein does not leave much free space for the dye molecule, and therefore, using prior information about protein position constrains the possible locations of the ADM and substantially improves the localization accuracy (**Fig. 4c**). The probability densities determined with and without this constraint overlap well with the accessible volume of the dye molecule, showing that NPS is indeed well suited for determining positions within macromolecular structures. To summarize the respective experimental uncertainties, we diagonalized the covariance matrix of the ADM position density to obtain uncertainties in the three principal directions (**Supplementary Methods**). The uncertainties of these position determinations were ±1.5 Å, ±1.9 Å and ±4.7 Å when the ADM was excluded from the Pol II interior and ±4.6 Å, ±5.4 Å and ±9.5 Å when no constraints were used.

Position of the RNA in the presence and absence of TFIIB

Next we applied NPS to determine a previously unknown position. To this end, we used NPS to analyze the FRET data for the ADM at

experiment, as was observed in Pol II elongation complexes with shorter RNA molecules[14], one can directly include these dynamics in the analysis and compute multiple ADM positions by performing the computation for the different maxima in the FRET histograms.

It is important to note how the shape and size of the credibility regions determined are influenced by the experimental errors introduced into the calculation. Although one might argue that this again induces some uncertainty because some of the experimental errors can only be estimated, it is easy to repeat the calculation with different error estimates and check for inconsistencies. Thus, the method has the potential to help determine unknown experimental errors arising from, for example, orientational effects of the donor-acceptor pairs.

The method described can be expanded by analyzing the data globally. This is advantageous if one wants to extract information from correlations between different measurements. In our example, we used the same SDMs for determining the position of the ADM both with and without TFIIB. The position of the particular SDMs is unlikely to change when TFIIB is added. Therefore, one could extract more information if both measurements were analyzed together. However, doing so requires a fully numerical solution of the problem.

One can also extend this concept to a situation in which no X-ray structure is available. By measuring the FRET distances between dye molecules attached to several positions within a complex, it is possible to extract the relative positions and plot the respective credibility regions, which incorporate the experimental uncertainties. This can be done using the bayesian framework described in this paper.

Our recently published results[14] and the data presented here show that the 5′ end of RNA transcripts of lengths of 26 and 29 nt are located on or close to the dock domain of Pol II. We further show here that addition of TFIIB to preformed elongation complexes results in a subpopulation of complexes that contain the 5′ end of the nascent RNA in proximity of the Rpb4/7 heterodimer. This is consistent with UV cross-linking of emerging RNA transcripts with a length between 26 and 32 nt (ref. 29) to the Rpb7 subunit, indicating that TFIIB is bound to the elongation complexes during these studies. This raises the question how long TFIIB remains associated with the transcription complex. It was suggested that the initial growth of the RNA results in a clash with the so-called B-finger domain of TFIIB and triggers displacement of the B-finger from the active-site cleft during the transition from initiation to elongation and promoter escape[30]. Recent experiments have indicated that this displacement is triggered by the collapse of the DNA bubble that is initially formed upstream[31]. In light of the data presented here, TFIIB could remain bound to Pol II even after this transition. TFIIB might then help to direct RNA toward the Rpb4/7 subcomplex. This hypothesis is supported by recent data showing that Rpb4/7 remains bound to Pol II during elongation[32] and is involved in recruiting 3′ processing factors[33].

In conclusion, we have developed a new type of method that allows one to obtain quantitative structural information about the position and dynamics of flexible domains within macromolecular complexes, together with an accurate uncertainty estimate. The method is based on single-molecule data and can therefore be applied to situations wherein only a small fraction of the complex is in an altered conformation.

METHODS

Software. All numerical calculations were performed in a custom-written MATLAB (The MathWorks) program. A compiled version is available online (**Supplementary Software**), as well as at http://www.cup.uni-muenchen.de/pc/michaelis/, where updated versions may also be available. The complete source code is available upon request. Analytical calculations were done in Maple (Waterloo Maple Inc.) and visualization was done in PyMOL (DeLano Scientific).

Preparation of Pol II elongation complexes. We prepared Pol II elongation complexes according to a procedure previously published[14]. For the determination of the position of the ADM attached to the 3′ end of RNA (*RNA 1*), we used a 17-nt RNA oligomer (5′-AUGCAUAAAGACCAGGC-3′, IBA GmbH) with either Alexa 555 (DNA 1, DNA 2, DNA 3, Rpb7-S16C) or tetramethylrhodamine (Rpb7-Cys150) attached to the 3′ end.

ADM and SDM position prior. To calculate the ADM and SDM position priors, we used geometric constraints based on the crystallographic data for the Pol II elongation complexes and the chemical nature of the linkers, which were modeled as flexible chains (**Supplementary Methods**). Using these constraints, we calculated the respective accessible volumes. The SDM and ADM priors were assumed constant within the accessible volume for each SDM and the ADM.

Förster radius prior. To calculate the prior for each Förster radius, we determined the isotropic Förster radius (**Table 1**) from the spectral overlap of the dye pairs and the donor quantum yield[14], and performed Monte Carlo simulations of the orientation factor that enters into R_i. In the simulations, we assumed that the transition dipoles of each donor and acceptor dye pair perform a random walk around a preferred orientation within their fluorescence lifetimes, which results in an axially symmetric distribution of dye orientaions[9]. To characterize the extent of the random walk, we determined the depolarization on the timescale of the fluorescence lifetime by measuring fluorescence anisotropies of the dye molecules (**Supplementary Table 1** online). Finally, we calculated priors for R_i using the simulated distribution of orientations (**Supplementary Methods** and **Supplementary Fig. 8** online).

FRET efficiencies. We collected the data using the single-molecule TIRF microscope described previously[14]. In short, we recorded time traces of the donor and acceptor dye intensities with a frame rate of 10 Hz. We averaged the traces with a ten-frame sliding-average filter and calculated the FRET efficiencies taking into account the leakage of the donor emission into the acceptor channel, as well as differences in donor and acceptor quantum yields and detection efficiencies calculated molecule by molecule[14].

We estimated the average FRET efficiencies, E_i, from gaussian least-square fits to the histograms of the measured FRET efficiencies. The fits for measurements with (double gaussian) and without (single gaussian) TFIIB were done independently. In the case of the Rpb7-Cys94 measurement with TFIIB, the fit yielded a large uncertainty for the position of the side peak. We therefore used the value obtained from the measurement in absence of TFIIB to fit the main peak. We determined our systematic error by comparing the position of the peak in absence of TFIIB to that of the main

peak in the presence of TFIIB for all measurements and found a systematic error of 0.6%. Only in the case of Rpb7-S16C in the presence of TFIIB did we find a larger systematic error of 4% because the observed FRET efficiencies were very high and thus small errors in background subtraction led to substantial deviations. We then used error propagation to calculate the errors of the measured FRET efficiencies, ΔE_i, from the systematic and statistical errors (**Supplementary Methods**).

Because we computed E_i and ΔE_i, we used a gaussian likelihood, $p(E_i|\boldsymbol{x},\boldsymbol{s}_i,R_i,I)$, which is a probability density with a peak at the expected FRET efficiency, which is equal to $(1 + |\boldsymbol{x} - \boldsymbol{s}_i|^6 / R_i^6)^{-1}$, and a s.d. ΔE_i. Note that the expected FRET efficiency is a function of model parameters only.

ADM position posterior. The posterior for the ADM position, $p(\boldsymbol{x}|\{E_i\},I)$, was evaluated on a cubic lattice (spacing 1 Å). We calculated the maximum posterior position, the measurement errors and the probability density levels corresponding to the credibility regions displayed (**Supplementary Methods**). The ADM position posteriors were transformed and saved as density maps (**Supplementary Data**) such that the credibility regions may be displayed as isosurfaces in PyMOL.

Other methods. The bayesian parameter estimation applied to NPS, especially the choice of the priors and the calculation of the ADM position density, is described in the **Supplementary Methods**.

Note: Supplementary information is available on the Nature Methods website.

AUTHOR CONTRIBUTIONS
A.M. performed calculations, designed analysis, wrote the analysis program and wrote the paper; J.A. performed experiments and wrote the paper; F.B. provided reagents; A.J. performed experiments and provided reagents; P.C. and J.M. designed experiments and wrote the paper.

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Ramakrishnan, V. Ribosome structure and the mechanism of translation. *Cell* **108**, 557–572 (2002).
2. Singleton, M.R. *et al.* Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks. *Nature* **432**, 187–193 (2004).
3. Cramer, P., Bushnell, D.A. & Kornberg, R.D. Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* **292**, 1863–1876 (2001).
4. Vassylyev, D.G. *et al.* Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* **448**, 157–162 (2007).
5. Weiss, S. Fluorescence spectroscopy of single biomolecules. *Science* **283**, 1676–1683 (1999).
6. Forster, T. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys-Berlin* **437**, 55–75 (1948).
7. Bregeon, D. *et al.* Transcriptional mutagenesis induced by uracil and 8-oxoguanine in *Escherichia coli*. *Mol. Cell* **12**, 959–970 (2003).
8. Kapanidis, A.N. *et al.* Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314**, 1144–1147 (2006).
9. 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).
10. Myong, S. *et al.* Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase. *Science* **317**, 513–516 (2007).
11. Kapanidis, A.N. *et al.* Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci. USA* **101**, 8936–8941 (2004).
12. Rothwell, P.J. *et al.* Multiparameter single-molecule fluorescence spectroscopy reveals heterogeneity of HIV-1 reverse transcriptase:primer/template complexes. *Proc. Natl. Acad. Sci. USA* **100**, 1655–1660 (2003).
13. Rasnik, I. *et al.* DNA-binding orientation and domain conformation of the *E. coli* Rep helicase monomer bound to a partial duplex junction: single-molecule studies of fluorescently labeled enzymes. *J. Mol. Biol.* **336**, 395–408 (2004).
14. Andrecka, J. *et al.* Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **105**, 135–140 (2008).
15. Schröder, G.F. & Grubmüller, H. FRETsg: biomolecular structure model building from multiple FRET experiments. *Comput. Phys. Commun.* **158**, 150–157 (2004).
16. 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).
17. Zhuang, X. *et al.* Correlating structural dynamics and function in single ribozyme molecules. *Science* **296**, 1473–1476 (2002).
18. McKinney, S.A. *et al.* Structural dynamics of individual Holliday junctions. *Nat. Struct. Biol.* **10**, 93–97 (2003).
19. Diez, M. *et al.* Proton-powered subunit rotation in single membrane-bound F0F1-ATP synthase. *Nat. Struct. Mol. Biol.* **11**, 135–141 (2004).
20. Lewis, R. *et al.* Conformational changes of a Swi2/Snf2 ATPase during its mechano-chemical cycle. *Nucleic Acids Res.* **36**, 1881–1890 (2008).
21. Sivia, D.S. *Data Analysis: A Bayesian Tutorial* Ch. 1–3, 3–77 (Clarendon Press, Oxford, UK, 1996).
22. Alber, F. *et al.* Determining the architectures of macromolecular assemblies. *Nature* **450**, 683–694 (2007).
23. Medintz, I.L. *et al.* A fluorescence resonance energy transfer-derived structure of a quantum dot-protein bioconjugate nanoassembly. *Proc. Natl. Acad. Sci. USA* **101**, 9612–9617 (2004).
24. Radman-Livaja, M. *et al.* Architecture of recombination intermediates visualized by in-gel FRET of lambda integrase-Holliday junction-arm DNA complexes. *Proc. Natl. Acad. Sci. USA* **102**, 3913–3920 (2005).
25. Sun, X. *et al.* Architecture of the 99 bp DNA-six-protein regulatory complex of the lambda *att* site. *Mol. Cell* **24**, 569–580 (2006).
26. Mekler, V. *et al.* Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* **108**, 599–614 (2002).
27. Jung, C. *et al.* Simultaneous measurement of orientational and spectral dynamics of single molecules in nanostructured host-guest materials. *J. Am. Chem. Soc.* **129**, 5570–5579 (2007).
28. Toprak, E. *et al.* Defocused orientation and position imaging (DOPI) of myosin V. *Proc. Natl. Acad. Sci. USA* **103**, 6495–6499 (2006).
29. Ujvari, A. & Luse, D.S. RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat. Struct. Mol. Biol.* **13**, 49–54 (2006).
30. Bushnell, D.A. *et al.* Structural basis of transcription: an RNA polymerase II-TFIIB cocrystal at 4.5 angstroms. *Science* **303**, 983–988 (2004).
31. Pal, M., Ponticelli, A.S. & Luse, D.S. The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol. Cell* **19**, 101–110 (2005).
32. Jasiak, A.J. *et al.* Genome-associated RNA polymerase II includes the dissociable RPB4/7 subcomplex. *J. Biol. Chem.* **283**, 26423–26427 (2008).
33. Runner, V.M., Podolny, V. & Buratowski, S. The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3' processing factors. *Mol. Cell. Biol.* **28**, 1883–1891 (2008).
34. Kettenberger, H., Armache, K.J. & Cramer, P. Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* **16**, 955–965 (2004).