

Dissertation zur Erlangung des Doktorgrades der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Single molecule fluorescence studies of the RNA polymerase II elongation complex



Joanna Andrecka
aus Myslowice, Polen

2009

Erklärung

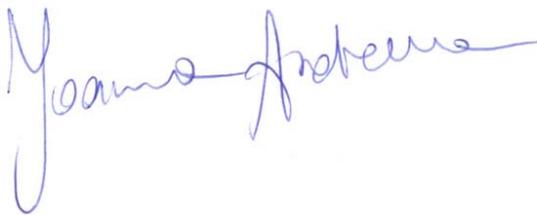
Diese Dissertation wurde im Sinne von §13 Abs. 3 der Promotionsordnung vom 29. Januar 1998 von Herrn Prof. Dr. Jens Michaelis betreut.

Ehrenwörtliche Versicherung

Diese Dissertation wurde selbständig und ohne unerlaubte Hilfe erarbeitet.

München, am 22. Juni 2009

Joanna Andrecka

A handwritten signature in blue ink that reads "Joanna Andrecka". The signature is written in a cursive style with a long horizontal stroke at the end.

Dissertation eingereicht am 22. Juni 2009

1. Gutachter: Prof. Dr. Jens Michaelis
2. Gutachter: Prof. Dr. Patrick Cramer

Mündliche Prüfung am 17. Juli 2009

Acknowledgements

I would like to thank my supervisor, Prof. Jens Michaelis for creating a highly motivating scientific environment, lots of freedom but also the pressure he put on me. I am deeply grateful for his support and understanding of my never-ending trips, which he sometimes called "urlaub - projects".

Moreover, I would like to thank my collaborator Prof. Patrick Cramer. I have got the unique opportunity to work in his lab at the time the group was still small - it was an unforgettable experience. I am deeply thankful to Claudia, Karim, Eli, Flo and Anass for their help.

I would also like to thank all, present and former, members of the AK Bräuchle for creating a very special atmosphere in our lab. It is difficult to express how great time I had here.

Special thanks to:

Moritz for his "live-advices";

Robert for making everyday-lab life much easier and for being the most amazing Bavarian-guy I have ever met;

Adam for all your help, NPS and answering all my stupid questions... Dzięki!!!

Peter for vmd script, climbing and night-printing session;

Bärbel for all our „salads" and „coffee breaks" ... and of course not only for that;

Iko for encouraging me ☺

All my students, who had to suffer from me - Markus, Julia, Andreas, Thilo... Thanks!

Flo, beside our "collaboration", thank you for having a real home at Saalburgstrasse 9.

Ania, thanks for bringing me here and that I could always rely on you.

All my friends in Kraków: Agacie- żabie, Kliczowi, Ani Kubasiak, Ani Nikiel, Klemensowi, Markowi, Wojtkowi, Michałowi... Dziękuję Wam Kochani, bez Was nie dałabym rady!!!

oraz Marysi, za drugi dom...

Moim rodzicom, których zaniedbałam...

*I wish to dedicate this work to Prof. Zygmunt Wasylewski,
my first supervisor, for whom nothing was impossible*

Table of contents

SUMMARY

PUBLICATIONS

INTRODUCTION

1	FLUORESCENCE	1
1.1	<i>Phenomenon of fluorescence</i>	1
1.2	<i>Jabłoński diagram</i>	2
1.3	<i>Fluorophores</i>	4
1.3.1	<i>Natural fluorophores</i>	4
1.3.2	<i>Fluorescent labeling of biomolecules</i>	5
1.4	<i>Fluorescence Resonance Energy Transfer</i>	7
1.5	<i>FRET measurements and distance determination</i>	8
1.5.1	<i>Incomplete labeling and averaging effect</i>	8
1.5.2	<i>Förster radius</i>	9
1.5.3	<i>Orientation factor κ^2</i>	10
1.6	<i>Single molecule fluorescence</i>	11
1.7	<i>TIRF</i>	12
1.8	<i>FRET data used as quantitative distance information</i>	15
2	BAYESIAN DATA ANALYSIS	17
2.1	<i>Bayes' theorem and marginalization</i>	17
2.2	<i>Bayesian approach in data analysis</i>	20
2.3	<i>Bayesian analysis in biology and medicine</i>	21
2.4	<i>NPS - bayesian analysis for single molecule FRET data</i>	23
2.4.1	<i>Intuitive description of the analysis method</i>	23
3	RNA POLYMERASE II	26
3.1	<i>Transcription</i>	26
3.2	<i>Transcription machinery</i>	27
3.3	<i>RNA Pol II transcription cycle</i>	30
3.4	<i>Structure of RNA Pol II elongation complex</i>	33
3.5	<i>Single molecule studies of transcription</i>	37
3.6	<i>Open questions and scope of this work</i>	39

METHODS

4	MOLECULAR BIOLOGY METHODS	40
4.1	<i>Single cysteine Rpb4/7 mutants</i>	40
4.2	<i>Transformation</i>	42
4.3	<i>Expression of recombinant protein in E.coli</i>	43
4.4	<i>Purification of the Rpb4/7 subcomplex</i>	44
4.5	<i>Protein labeling</i>	46
4.6	<i>Nucleic acid scaffolds</i>	47
4.7	<i>Preparation of Pol II - artificial bubbles elongation complex</i>	50
4.8	<i>Competition measurements with TFIIB</i>	50
5	FLUORESCENCE METHODS	52
5.1	<i>Anisotropy measurements</i>	52
5.2	<i>Förster radii measurements</i>	52
5.3	<i>Preparation of sample chamber for sp-FRET measurements</i>	53
5.4	<i>Experimental setup for sp-FRET</i>	55
5.5	<i>Data collection and analysis</i>	58

5.5.1	<i>Camera settings used in single molecule FRET measurements</i>	58
5.5.2	<i>Analysis program</i>	58
5.5.3	<i>Calculation of FRET efficiencies</i>	60
6	DATA ANALYSIS METHODS	62
6.1.1	<i>Conventional Method - Triangulation</i>	62
6.2	<i>NPS</i>	63
6.2.1	<i>Bayesian parameter estimation applied to NPS</i>	63
6.2.2	<i>Likelihood</i>	64
6.2.3	<i>Prior</i>	65
6.2.4	<i>Posterior</i>	68
RESULTS AND DISCUSSION		
7	SINGLE-MOLECULE TRACKING OF RNA EXITING FROM RNA POLYMERASE II	69
7.1	<i>Engineering fluorescent labeled active Pol II elongation complexes</i>	70
7.2	<i>Single-pair FRET results</i>	72
7.2.1	<i>FRET traces</i>	72
7.2.2	<i>Dynamic switching</i>	73
7.2.3	<i>Direct acceptor excitation</i>	73
7.2.4	<i>From FRET efficiency to distance</i>	75
7.3	<i>Accurate localization of the RNA 5'-end</i>	76
7.3.1	<i>Triangulation</i>	76
7.3.2	<i>Accuracy test and error estimation of the triangulated positions</i>	77
7.4	<i>Biological implications</i>	79
7.4.1	<i>RNA traverses the exit tunnel and disengages from the surface</i>	79
7.4.2	<i>Extending RNA associates with Pol II before it becomes flexible</i>	81
7.4.3	<i>Conclusion and outlook</i>	82
8	POSITION OF 5'END OF THE 29NT RNA DETERMINED BY NPS AND INFLUENCE OF TRANSCRIPTION FACTOR IIB ON THE POSITION OF RNA	84
8.1	<i>Additional data required to use NPS</i>	84
8.2	<i>Experimental test of the NPS system</i>	86
8.3	<i>Positions of RNA17, RNA20, RNA23 and RNA26 revealed by NPS</i>	87
8.4	<i>Position of the 5'end of 29nt RNA determined by NPS</i>	91
8.5	<i>Influence of transcription factor IIB on the position of the RNA</i>	93
8.6	<i>Conclusion and outlook</i>	96
9	NANO POSITIONING SYSTEM REVEALS THE COURSE OF UPSTREAM AND NONTEMPLATE DNA WITHIN POL II ELONGATION COMPLEX	97
9.1	<i>Single molecule data used to determine the nontemplate DNA position</i>	97
9.2	<i>Positions of the ADMs revealed using NPS</i>	101
9.3	<i>Model of the nontemplate and upstream DNA</i>	102
9.4	<i>Detailed picture and discussion</i>	104
9.4.1	<i>Strand separation</i>	104
9.4.2	<i>Nontemplate DNA within the bubble</i>	105
9.4.3	<i>Loops network and its interaction with DNA</i>	106
9.4.4	<i>Upstream DNA in the elongation complex</i>	108
9.5	<i>Conclusion</i>	110
10	OUTLOOK AND FUTURE PERSPECTIVES	111

APPENDIX

REFERENCES

CURRICULUM VITAE

Summary

Transcription is the first step in gene expression, where the information carried by DNA is transcribed into messenger RNA (mRNA). The mRNA synthesis in the eukaryote nucleus is carried out by one enzyme – RNA polymerase II (Pol II). The structure of the free enzyme is known, as well as the structure of the transcribing enzyme together with DNA template and RNA product. However, some flexible domains within the elongation complex could not be solved. In this work single molecule fluorescence resonance energy transfer (smFRET) was used to study structure of the yeast RNA polymerase II elongation complex, in particular the position of nucleic acids which could not be visualized in the x-ray structure due to their flexibility.

First, RNA exiting from the Pol II elongation complex was tracked. Measuring FRET efficiencies allows for calculating the distance between two dye molecules. Determined distances between the RNA 5'-end and three known locations within the elongation complex were used for determination of its position by means of triangulation.

Next, a new data analysis method was developed. This method combines smFRET data, x-ray crystallographic data and a rigorous statistical analysis, which accounts for the experimental uncertainties. It was named “Nano Positioning System” (NPS) due to its analogy to the Global Positioning System. With this approach the position of the 5'-end of the nascent RNA was determined. Additionally, the influence of the transcription factor IIB (TFIIB) on the position of the RNA was studied.

Finally, NPS was used to find several positions within nontemplate and upstream DNA. These results were used to build a complete model of the DNA across the elongation complex. Together with earlier studies determining the position of the nascent RNA, it completes the picture of the Pol II elongation complex.

Publications

Parts of this work have been published:

Andrecka J., Lewis R., Bruckner F., Lehmann E., Cramer P., and Michaelis J. (2008)

Single-molecule tracking of mRNA exiting from RNA polymerase II

Proc Natl Acad Sci USA 105, 135-140.

Muschielok A., Andrecka J., Jawhari A., Bruckner F., Cramer P., and Michaelis J. (2008)

A nano-positioning system for macromolecular structural analysis

Nature Methods 5, 965- 971.

Andrecka J., Treutlein B., Izquierdo Arcusa M. A., Muschielok A., Lewis R., Cheung A., Cramer P., and Michaelis J. (2009)

Nano Positioning System reveals the course of upstream and nontemplate DNA within the RNA polymerase II elongation complex

Manuscript in preparation.

Introduction

1 Fluorescence

1.1 Phenomenon of fluorescence

The term *fluorescence* was introduced by George Gabriel Stokes in 1852 (Stokes, 1852) and was given as the description of the essence of the mineral fluorite, composed of calcium fluoride, which showed emission of the visible light when illuminated with "invisible radiation" (UV radiation).

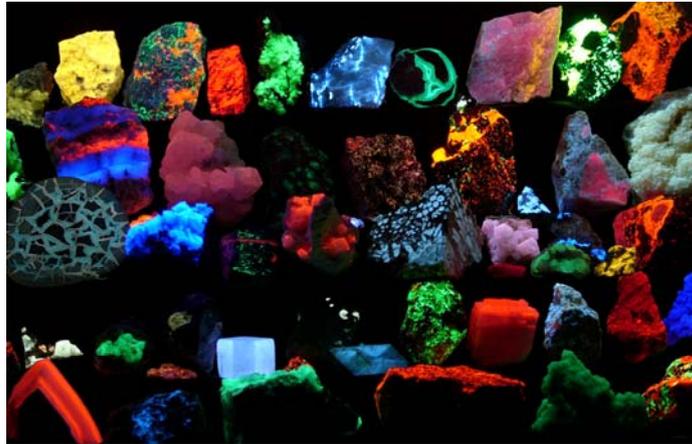


Figure 1: Fluorescent minerals

Fluorescent minerals are minerals that emit visible light when activated with ultraviolet light (called "invisible radiation" by G.G. Stokes in 1852), x-rays or an electron beam. This intriguing phenomenon was called fluorescence after the mineral fluorite, which sometimes glows blue when exposed to sunlight. There are around 500 minerals having fluorescent properties. This picture was taken from: <http://en.wikipedia.org/wiki/Fluorescence>.

Fluorescence as a type of luminescence is an optical phenomenon, in which the molecular absorption of a photon triggers the emission of another photon with longer wavelength. Absorption of energy can bring an electron into its excited state. Fluorescence is the emission of a photon as a consequence of the electron returning from the excited singlet state to the ground state. This return occurs rapidly with typical emission rates of 10^8 s^{-1} .

1.2 Jabłoński diagram

The fluorescence phenomenon can be described by the Jabłoński diagram, which exists in a variety of forms and illustrates the possible processes after excitation.

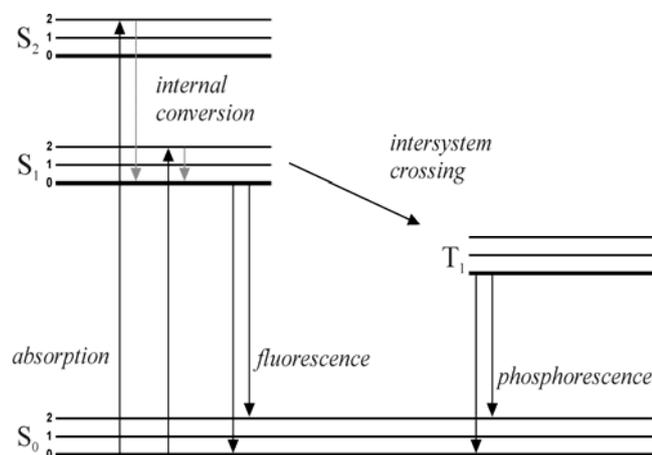


Figure 2: Jabłoński diagram

One possible form of the Jabłoński diagram. S_0 , S_1 , S_2 depict ground, first and second excited electronic singlet states, respectively. 0, 1, 2 denote different vibrational energy levels. Possible processes after excitation are shown.

A typical diagram is shown in Figure 2. Absorption usually occurs by the molecules in the electronic ground state S_0 and the fluorophores are excited to the higher vibrational levels of S_1 or S_2 . Beside a few exceptions, molecules relax rapidly to the lowest vibrational level of S_1 in a process called internal conversion. From there, molecules usually return to the higher excited vibrational levels of the electronic ground state, which quickly reaches thermal equilibrium. Molecules in the S_1 state can also undergo a spin conversion to the first triplet state, T_1 , in a process called intersystem crossing. Emission from this state, termed *phosphorescence*, is a second type of luminescence and will be not important for this thesis. Other processes like quenching, energy transfer or solvent interactions can be presented by Jabłoński diagram. It is also used to explain common features of fluorescence, like the fact that the energy of the emission is less than that of the absorption. This general characteristic of fluorescence called Stoke's shift was named after Sir George Gabriel Stokes, who described it for the first time in 1852 (Stokes, 1852).

Aleksander Jabłoński was a Polish physicist, born in 1898. He entered the University of Kharkov to study physics in 1916. His study was interrupted by the First World War. At the end of 1918, when an independent Poland was recreated after more than 120 years of occupation by neighboring powers, Jabłoński arrived in Warsaw to continue his studies of physics at the University of Warsaw.

As an enthusiastic musician, Aleksander Jabłoński played the first violin at the Warsaw Opera from 1921 to 1926 parallel to his studies at the university. His teacher, Stanisław Barcewicz, was a famous Polish violin player at this time. Although Jabłoński left the Opera in 1926 and devoted himself entirely to the scientific work, music remained his great passion.

He received his PhD in 1930 with a paper "On the influence of the change of the wavelength of excitation light on the fluorescence spectra". After that, Jabłoński spent two years (1930-1931) as a fellow of the Rockefeller Foundation in Germany working with Peter Pringsheim at the Friedrich Wilhelm University in Berlin and with Otto Stern in Hamburg. In 1934, he acquired his habilitation from the University of Warsaw with the thesis "On the influence of intermolecular interactions on the absorption and emission of light".

The problem that intrigued Jabłoński for many years was the polarization of photoluminescence. To explain the experimental facts, he distinguished the transition moments in absorption and in emission and analyzed various factors responsible for the depolarization of luminescence. In 1931 Jabłoński started to work in his second main field of research, namely the pressure broadening of spectral lines.

In April 1938, Jabłoński accepted a faculty appointment at the Stefan Batory University at Vilno, where he developed experimental studies of pressure broadening of atomic spectral lines. In particular, he initiated the pioneering investigations of the temperature dependence of the widths of pressure broadened spectral lines. These studies, whose first results were published in two communications in "Nature", were interrupted by the outbreak of World War II.

After the war in November 1945, Jabłoński started to work again at the Department of Physics of the University of Warsaw under Stefan Pieńkowski. He moved to Toruń, where in fall 1945 a new University was founded named after Nicholas Copernicus, who was born in Toruń. On the 1st of January 1946, Jabłoński was nominated as full professor of N. Copernicus University and his first historic lecture for students of science at Toruń took place on February 17, 1946. It must be remembered, however, that this beginning happened in the very difficult post-war years in a country totally destroyed by the World War II. Despite all difficulties Jabłoński with great energy started to organize an authentic scientific center at Toruń. As the chairman of the Physics Department from its very beginning in 1946 to his retirement in 1968 Jabłoński created a modern laboratory, in which he developed his own field of research in atomic and molecular optics and he helped to initiate researches in other fields, such as those in solid state physics, magnetic resonance studies and others.

In the 1950's Jabłoński developed the theory of quenching and depolarization of photoluminescence. This theory was used as a basis for the interpretation of many experiments performed at Toruń by his co-workers in the late 1950's and early 1960's. At that time Jabłoński introduced another quantity, describing the degree of polarization called "emission anisotropy", which is now generally preferred and recommended.

The stimulus which Aleksander Jabłoński has provided and is still providing to all his co-workers and pupils can hardly be overestimated. For all of them Professor Aleksander Jabłoński will always hold a special place. Many of his former students in Toruń and in other universities continue and extend his work in the field of luminescence, photophysics, biophysics, dye lasers and pressure broadening of spectral lines.

Professor Aleksander Jabłoński died in 1980.

(Szudy, 1998)

1.3 Fluorophores

1.3.1 Natural fluorophores

Fluorescent substances are called *fluorophores*. The first known fluorophore was quinine, which is present in tonic water. Exposed to sunlight, quinine gives a faint blue glow, usually visible at the water surface. The first observation of fluorescence of quinine, not yet called fluorescence, was described by Sir John Frederick William Herschel in 1845 (Herschel, 1845). To this day, the fluorescence of quinine remains one of the most beautiful examples of this phenomenon.

Fluorescence typically occurs from aromatic molecules. Fluorophores can be divided into two general classes – *intrinsic fluorophores*, in which fluorescence occurs naturally, and *extrinsic fluorophores*, which are added to a sample that does not display the desired spectral properties. A good example of an intrinsic fluorophore is the indol group of tryptophan. It absorbs near 280 nm and emits near 340 nm. The fluorescence emission spectrum of the indol group shows a high sensitivity to solvent polarity. It is blue shifted, when the tryptophan is buried within the protein or red shifted, when it is exposed to polar solvent e.g. if protein is unfolded. There are other components of living cells, which display significant fluorescence: reduced nicotinamide adenine dinucleotide (NADH), oxidized flavins such as adenine dinucleotide (FAD) and adenine mononucleotide (FMN), pyridoxal phosphate, as well as chlorophyll (plants). Membranes are not fluorescent. However, it is possible to label them with e.g. diphenylhexatriene (DPH), which spontaneously aggregates with the non-polar region of the membranes formed by lipid tails. DPH is almost not soluble in water and its fluorescence is quenched in such condition. Fluorescence emission occurs only from those DPH molecules which are bound to the membranes.

Another example of intrinsic fluorescence is the green fluorescent protein (GFP). It can absorb and emit light due to a unique covalently attached chromophore, which is

formed post-translationally within the protein upon cyclisation and oxidation of residues 65-67, Ser-Tyr-Gly. This protein was found in *Aequorea Victoria*. However, not GFP is responsible for the glow often seen on pictures of this jellyfish - is due to the reflection of the flash used to photograph the jellies. In case of *Aequorea Victoria*, a protein called *aequorin* releases blue light upon binding with calcium. This blue light is then absorbed by GFP, which in turn emits green light.

1.3.2 Fluorescent labeling of biomolecules

For most biological applications of fluorescence, specific labeling of the biological compounds is required. Therefore fluorescent dye derivatives, which react with specific groups, are used. For example, in order to obtain covalently labeled nucleic acids, oligonucleotides are synthesized with a primary amino group linked via flexible carboxyhydrate chain. Subsequently, succinimidyl ester derivatives of fluorescent dyes are used for labeling, by forming a stable amide bond upon reaction with amino groups present in oligonucleotides (Figure 3(A)). The same reaction can be used for nonspecific labeling of proteins (reaction with amino groups of the N- terminus and lysine residues).

To attach a fluorescent dye specifically to a protein, the thiol group of cysteine can be used. The application of maleimide derivatives of fluorescent dyes leads to a covalent linkage of the dye to the protein via a thioether moiety (Figure 3(B)). This method is well suited for proteins containing only a single exposed cysteine. However, many proteins have more than one cysteine residue and it is necessary to remove them prior to the labeling procedure. Frequently, even one point mutation can negatively influence function and/or stability of a protein. Therefore, the development of a new protein labeling method, like introducing an unnatural amino acids upon the expression of the protein of interest, seems to be crucial for a progress the fluorescence field (Prescher and Bertozzi, 2005; Xie and Schultz, 2006).

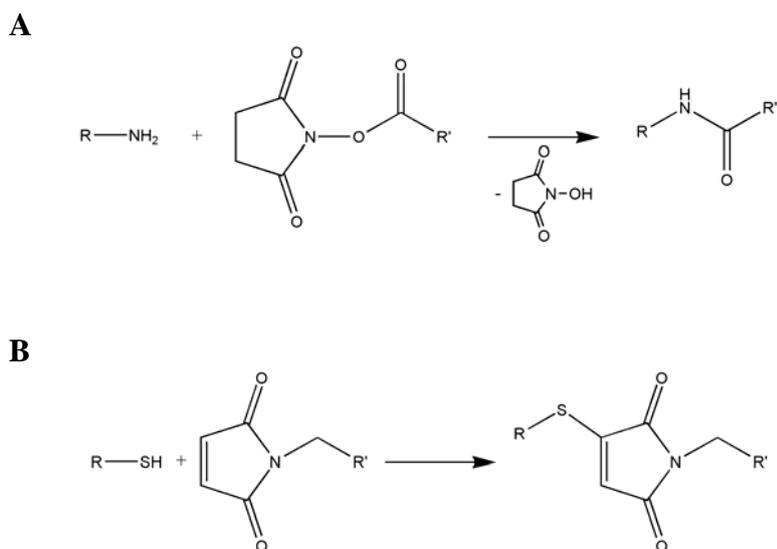


Figure 3: Fluorescent labeling

(A) Reaction of the succinimidyl ester derivative of the fluorescent dye (-R') with the primary amine of nucleic acids or proteins (R - NH₂). **(B)** Reaction of the maleimide derivative of the fluorescent dye (-R') with the thiol group of cysteine (R - SH).

It is also possible to express proteins with GFP fused. GFP was purified and its properties were studied by Osamu Shimomura in the 1960s (Shimomura et al., 1962). Later, many different mutants of GFP have been engineered. The first major improvement was a single point mutation (S65T) reported in 1995 by Roger Tsien (Heim et al., 1995). This mutation dramatically improved the spectral characteristics of GFP resulting in increased fluorescence, photostability and a shift of the major excitation peak to 488nm with the peak emission kept at 509 nm. GFP and its derivatives have thoroughly redefined fluorescence microscopy and its application in cell biology and other biological disciplines. Fluorescent proteins expressed with another protein are usually much less harmful than conventional chemical markers. The importance of GFP was recognized in 2008, when the Nobel Committee awarded Osamu Shimomura, Marty Chalfie and Roger Tsien.

1.4 Fluorescence Resonance Energy Transfer

The excited state energy of an excited dye molecule (called donor, D) can be transferred to another dye molecule (called acceptor, A) in a process called Fluorescence Resonance Energy Transfer (FRET). FRET is a result of long range dipole - dipole interaction and occurs without the appearance of a photon. One can observe this phenomenon if the donor emission spectrum overlaps with the acceptor absorption spectrum and the distance between the two dyes is shorter than a certain value (usually < 10 nm). The transfer efficiency E is strongly dependent on the distance, r , between donor (D) and acceptor (A) (equation [1]). R_0 is the Förster distance at which FRET is 50% efficient:

$$E = \frac{R_0^6}{R_0^6 + r^6} \quad [1]$$

When the D - A distance is near R_0 , the FRET efficiency is very sensitive for distance changes. Therefore, one of the applications of FRET is to get information about the distance between two sites on a macromolecule (Figure 4).

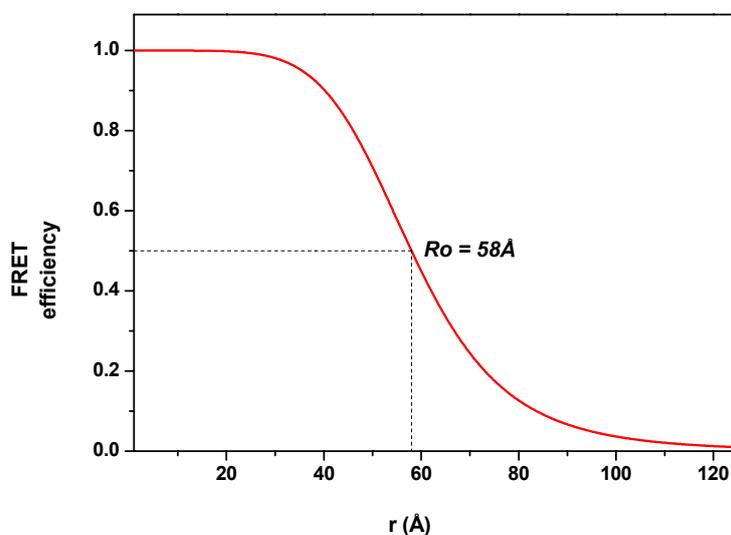


Figure 4: Dependency of the FRET efficiency on distance, r , between Donor and Acceptor.

1.5 FRET measurements and distance determination

Förster distances, R_0 , ranging from 20-90 Å are convenient to study biological macromolecules. Therefore, FRET was termed a “molecular ruler” according to the paper from the late 1960s (Stryer and Haugland, 1967). However, FRET data analysis brings several problems, which are shortly described below.

1.5.1 Incomplete labeling and averaging effect

The first step in distance determination is obtaining absolute FRET efficiency values. Here, complication can arise due to non-stoichiometric labeling and/or static quenching, which can lead to dark states. In bulk FRET measurements, it is the largest source of errors. Different species of molecules give different input into the absolute FRET efficiency value. For example, if the sample is incompletely labeled with acceptor, the measured FRET efficiency is smaller than the true value.

Another problem appears if the biological system studied is intrinsically heterogeneous. There could be two different sources of heterogeneity - distribution of structures or dynamics (structure changes between two or more states). Crucial information about these heterogeneities is lost in conventional FRET measurements, in which the signal from many molecules is averaged.

In contrast, single-molecule methods permit detection and characterization of individual molecules at a time. It solves the problem of the incompletely labeled sample and can be used to recover the information lost due to ensemble averaging (see *Introduction* 1.6; 3.5).

1.5.2 Förster radius

Assuming a proper determination of the FRET efficiency value the simple formula [1] can be used to calculate the distance, r . However, it is necessary to know the Förster radius, R_0 , in order to calculate the D - A distance.

$$\langle R_0 \rangle = \left[\frac{9(\ln 10)\Phi^D \kappa^2 J(\nu)}{128\pi^5 N_A n^4} \right]^{\frac{1}{6}} \quad [2]$$

Φ^D is the quantum yield of the donor, $J(\nu)$ is the spectral overlap between the donor's emission and the acceptor's absorption, N_A is Avogadro's number, n is the index of refraction of the medium and κ^2 is the orientation factor determined by the relative orientation of the two dipoles.

$J(\nu)$, Φ^D , n and κ^2 must be known in order to calculate R_0 . The refractive index is usually assumed to be near that of water ($n = 1.33$) or small organic molecules ($n = 1.39$).

Spectroscopic properties of chromophores depend strongly on their molecular environment. In the case of fluorescently labeled biological probes, quantities such as extinction coefficient, radiative rate, quantum yield, rotational freedom and spectra can differ from those of the un-conjugated fluorophores. Failure to account for these factors contributes uncertainties in quantitative interpretation of FRET and can cause systematic errors. Therefore, it is important to determine $J(\nu)$, Φ^D for each donor-acceptor pair.

1.5.3 Orientation factor κ^2

Last but not least factor in the analysis of the FRET efficiency is the κ^2 factor, which depends on the relative orientation of donor and acceptor and is given by:

$$\begin{aligned}\kappa^2 &= (\cos\theta_T - 3 \cos\theta_D \cos\theta_A)^2 \\ \kappa^2 &= (\sin\theta_D \sin\theta_A \cos\varphi - 2 \cos\theta_D \cos\theta_A)^2\end{aligned}\quad [3]$$

θ_T is the angle between the emission transition dipole of the donor and the absorption transition dipole of the acceptor, θ_D and θ_A are the angles between these dipoles and the vector joining the donor and the acceptor, φ is the angle between the planes (Figure 5).

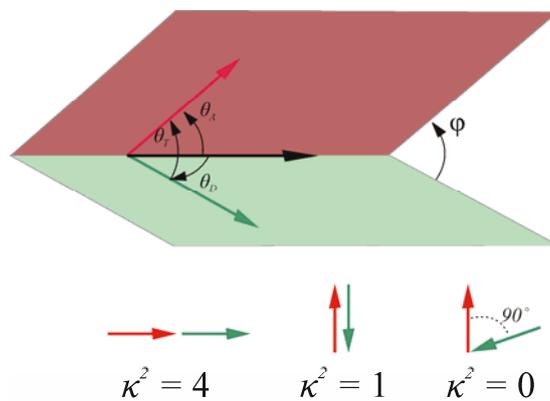


Figure 5: Dependency of the orientation factor κ^2 on the direction of the emission dipole of the donor (green) and the absorption dipole of the acceptor (red)

The κ^2 factor can range from 0 to 4 (the dependency on the relative orientation of donor and acceptor is shown in Figure 5). Since the sixth root of κ^2 is taken to calculate the distance, variation of κ^2 from 1 to 4 will result in 26% change in distance, r . However, if the dipoles are orientated perpendicular to one other, $\kappa^2 = 0$. This results in serious errors in distance calculation. This problem has been discussed in detail (Dale et al., 1979).

At present, it is difficult to measure κ^2 . Therefore, there is always concern about the relative orientation of donor and acceptor. However, it is possible to set limits on the range of κ^2 that, in turn, set limits on the range of D - A distances (Dale et al., 1979). These limits can be determined from the anisotropies of the donor and acceptor. κ^2 is generally assumed to be equal to 2/3, which is the value for orientation of donor and acceptor randomized by rotational diffusion prior to energy transfer.

1.6 Single molecule fluorescence

The first single molecule optical detection experiments were performed exactly 20 years ago by frequency-modulation (Moerner and Kador, 1989) and later by fluorescence excitation (Orrit and Bernard, 1990). Since then, single molecule methods have become a very important tool to study biological system.

Compared with bulk techniques, there are several advantages of single molecule approaches. The most important property is that the individual behavior of each molecule can be distinguished. The time resolution of single-molecule fluorescence experiments allows for an investigation of structural changes in real time, resulting in a dynamic view of complex biological processes.

Single molecule FRET (smFRET) was introduced in 1996 (Ha et al., 1996). FRET signal can be detected from freely diffusing molecules in solution or from confined molecules using a variety of immobilization methods. Collecting FRET data from one molecule at a time solves the averaging problem and difficulties due to incomplete labeling. It is also a tool to measure with better precision providing additional information about dynamics of individual donor - acceptor pairs. Therefore, Stryer's idea of FRET being used a "molecular ruler", has become more realistic.

FRET is being used to study a variety of biological systems and has become a common method. Fast progress in the single molecule field stimulated the

development of new techniques and tools (e.g. fluorescent dyes). Since one can “observe” one molecule at a time the best dye for single molecule experiments is a dye which can be characterized by a high number of fluorescence cycles (before photobleaching), high count rates and, what is very important for FRET approaches, does not show any blinking behavior.

1.7 TIRF

Total Internal Reflection Fluorescence (TIRF) is a common technique to observe FRET at the single molecule level. It was developed by Daniel Axelrod and his co-workers in the early 1980's (Axelrod, 1984) and is based on well known physical phenomena.

Snell's law (also known as law of refraction) describes the behavior of light passing through a boundary between two different isotropic media, such as water and glass. The law says that the ratio of the sine of the angles of incidence and of refraction is a constant and depends on the refractive indexes of the media n_1, n_2 :

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{n_2}{n_1} \quad [4]$$

When $\theta_1 > \theta_{critical}$, refraction does not occur and the incident ray undergoes *total internal reflection* from the interface. In order to calculate the critical angle of incident, θ_2 is assumed to be 90° and equ.[4] is solved for $\theta_{critical}$:

$$\theta_{critical} = \arcsin\left(\frac{n_2}{n_1}\right) \quad [5]$$

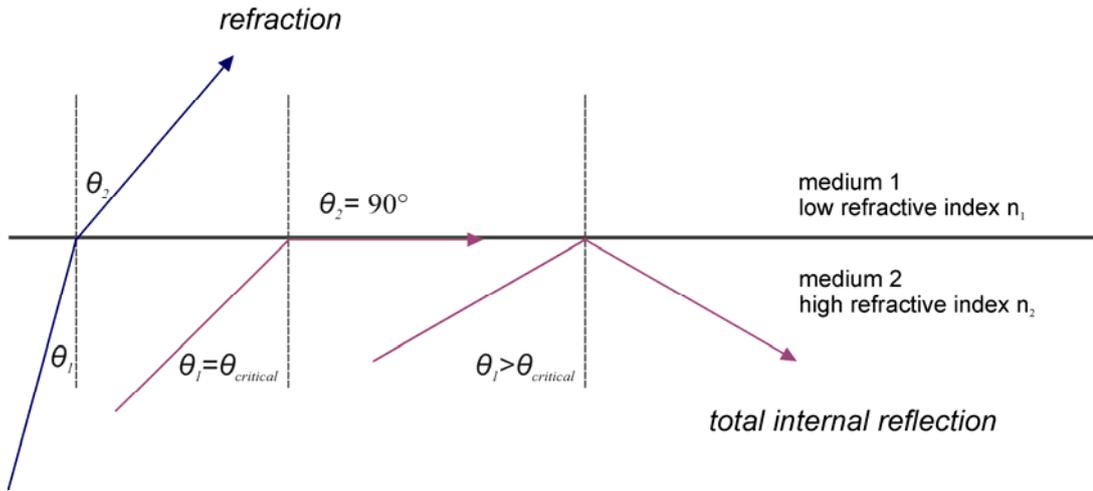


Figure 6: Total internal reflection

When the light passes through a boundary between two different isotropic media, it refracts in the incident medium at an angle given by Snell's law. When $\theta_1 > \theta_{critical}$, no refracted ray appears, and the incident ray undergoes total internal reflection from the interface. Before the ray totally internally reflects, the light refracts at the angle of 90° ; and travels directly along the surface between the two refractive media.

During total internal reflection at the interface, a so called *evanescent wave* is generated. The evanescent wave has its maximum of intensity at the surface and exponentially decays with increasing distance from the interface, but can still be used to excite fluorophores:

$$E(z) = E(0) \exp(-z/d) \quad [6]$$

$E(z)$ is the energy at a perpendicular distance, z , from the interface, and $E(0)$ is the energy at the interface. The penetration depth, d , is dependent on the wavelength of the incident illumination, λ , the angle of incidence, θ_1 , and the refractive indices of the media, n_1 , and at the interface, n_2 , according to the equation:

$$d = \frac{\lambda}{4\pi} (n_1^2 \sin^2 \theta_1 - n_2^2)^{-1/2} \quad [7]$$

It is important to note that, for total internal reflection to occur and produce an evanescent field, the refractive index of the medium, n_1 , must be greater than that of the specimen medium, n_2 , and the angle of incidence, θ_I , must be greater than the critical angle $\theta_{critical}$.

Only molecules that are at the surface are excited and fluoresce, while molecules in the bulk of solution, at distances larger than 100-200 nm are not excited and, respectively, do not fluoresce. Therefore, TIRF efficiently rejects the background signal of the bulk of solution and allows for very sensitive detection - down to single molecules.

TIRF originally was a prism based technique. The sample attached to the glass slide (usually quartz glass) is observed through a cover slide by high aperture objective lens. The backside of the slide is connected to the prism (Figure 7) so that the laser beam strikes the slide/sample interface at the critical angle.

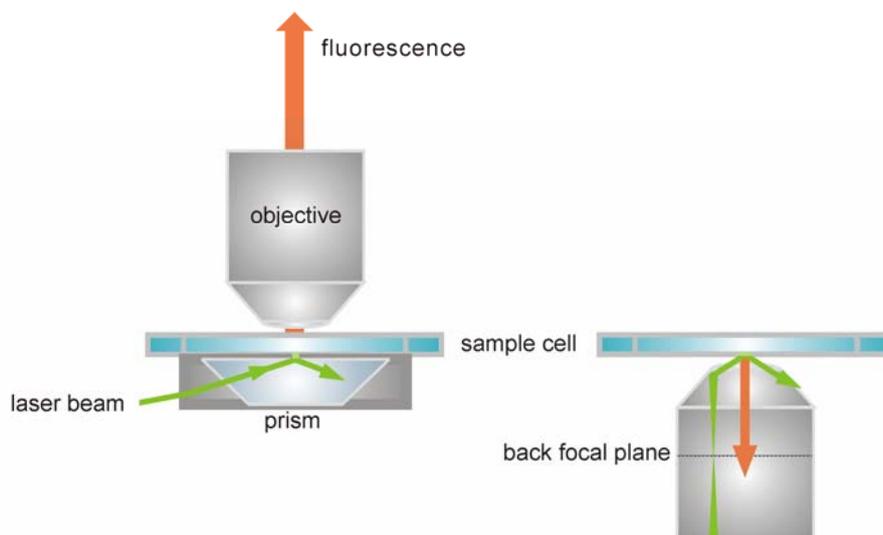


Figure 7: Prism and objective type TIRF

Two types of illumination which allow to obtain the TIR effect. (Left) Prism type TIRF; a focused laser beam is introduced into the microscope coverslip by means of a prism attached to its surface, and the beam incidence angle is adjusted to the critical angle. (Right) Objective type TIRF; light is focused into the back focal plane of the objective. The focus is moved away from the center towards the edge of the back focal plane in order to obtain the critical angle.

The TIR effect can also be achieved by illuminating a sample through an objective lens (Figure 7). The numerical aperture directly determines the range of angles, at which light can exit the objective. The relationship between the numerical aperture and the achievable illumination incidence angles is described by the following equation: *Numerical Aperture (NA)* = $n \cdot \sin\alpha$, n represents the refractive index and α is one-half the objective angular aperture.

Typically, the refractive index of the sample is about 1.33 and it requires illumination with an objective having a numerical aperture of greater than 1.33 in order to achieve total internal reflection (ideally 1.45 or higher). Laser light must be focused into the back focal plane of the objective. To obtain the critical angle, one has to move the focus away from the center towards the edge of the back focal plane (Figure 7).

1.8 FRET data used as quantitative distance information

FRET was termed a “molecular ruler”, as mentioned before, but for a long time it was not often used for precise distance determination. This apparent paradox has two reasons: the problem in obtaining absolute FRET efficiency values and the difficulty of calculating distances, r , from FRET efficiencies, E (see *Introduction 1.5*).

To derive a relationship between fluorescence measurements and the FRET efficiency values one can use the formalism of Clegg (Clegg, 1992). It accounts for all species of labeled molecules: the fraction with only donor and only acceptor molecules, the fraction with both donor and acceptor and the fraction of unlabeled molecules. In addition FRET is corrected for direct excitation of acceptor at the donor wavelength, λ_{donor} . Besides, all spectroscopic properties, which can be different for a fluorophore attached to the biological sample compared to the free fluorophore, must be determined (e.g. extinction coefficient, quantum yield). Moreover, errors of all measured variables must be taken into account, when associated error in the determined distance is calculated.

This method was used previously to study specific conformational changes in *Escherichia coli* 30S ribosomal subunit (Hickerson et al., 2005; Majumdar et al., 2005). The measured distance changes provided restraints for modeling the movement that occurs within the 30S ribosome subunit upon formation of the 70S ribosome.

Over the last decade, single-molecule fluorescence methods have been developed into extremely versatile tools. This fast progress made possible to use FRET in order to obtain quantitative distance information about proteins and protein complex structures (Diez et al., 2004). Using single molecule techniques solves the averaging problem of different populations present in the sample. It allows for the observation of individual molecules at a time and therefore gives additional information about dynamics.

It was demonstrated that the information about several distances within a labeled complex can be used for a distance-constrained triangulation procedure (Mekler et al., 2002; Rasnik et al., 2004). In addition, the time resolution of single-molecule fluorescence experiments allows for the real time-investigation of conformational processes and thus dynamic pictures of biological systems emerge, which are oftentimes used to refine the understanding of molecular processes (Margittai et al., 2003).

It is important to mention that the problem of calculating distances was always overcome by assuming κ^2 to be 2/3 (see *Introduction* 1.5.3). However, even if this assumption was true, only the most likely position could be estimated and it was not calculated how existing experimental uncertainties, in particular error in κ^2 , might effect the determined position. Moreover, information about dyes molecules position was often neglected. However, if one parameter can not be determined experimentally this lack of knowledge should be included into the data analysis. Different way of thinking about experimental uncertainties will be a subject of the next chapter.

2 Bayesian Data Analysis

Probability theory is nothing else than common sense reduced to calculation

Pierre Simon, Marquis de Laplace

The Bayesian understanding of the concept of probability is a *measure of a state of knowledge*, in contrast to interpreting it as a frequency or a physical property of a system. Its name is derived from the 18th century statistician Thomas Bayes, who pioneered some of the concepts. This different way of thinking will be briefly introduced in this chapter.

2.1 Bayes' theorem and marginalization

Bayes' theorem relates the probabilities p , of events X and Y :

$$p(X | Y, I) = \frac{p(Y | X, I)p(X | I)}{p(Y | I)} \quad [8]$$

Here, vertical bar “|” means ‘given’ so that all items to the right of that symbol are taken as being true, comma is read as the conjunction ‘and’. Bayes' theorem is based on two rules. First is the *product rule*:

$$p(X, Y | I) = p(X | Y, I)p(Y | I) \quad [9]$$

The *product rule* states that the probability that both, X and Y are true is equal to the probability that X is true given Y is true times the probability that Y is true (irrespective of X). Bayes' theorem [8] follows directly from the *product rule* [9].

Since the statement: “ Y and X are both true” means same as “ X and Y are both true”,
 $p(X, Y | I) = p(Y, X | I)$. Therefore the product rule [9] can be applied:

$$p(Y, X | I) = p(Y | X, I)p(X | I). \quad [9a]$$

Eq. [8] and [9a] result in eq. [9b] which is exactly Bayes' theorem [8]:

$$p(X | Y, I)p(Y | I) = p(Y | X, I)p(X | I) \quad [9b]$$

The second rule called *sum rule* states that the sum of the probabilities that X is true and that \bar{X} is false is equal to one:

$$p(X | I) + p(\bar{X} | I) = 1 \quad [10]$$

In all equations the probabilities are made conditional on I , so that we keep in mind that there is always relevant background information.

The marginalization equation for our standard X and Y takes the following form (Sivia, 1996):

$$p(X | I) = p(X, Y | I) + p(X, \bar{Y} | I) \quad [11]$$

It comes directly from the *product rule* and the *sum rule*. First one can expand $p(X, Y | I)$ by using the *product rule* [9]:

$$p(X, Y | I) = p(Y, X | I) = p(Y | X, I)p(X | I)$$

Then, by adding a similar expression for $p(X, \bar{Y} | I)$ to both sides of the equation, one gets:

$$p(X, Y | I) + p(X, \bar{Y} | I) = [p(Y | X, I) + p(\bar{Y} | X, I)] \times p(X | I) \quad [11a]$$

The *sum rule* brings the quantity in square brackets on the right side to unity and then directly to marginalization equation [11]. It states, verbally, that the probability that X is true, irrespective of whether or not Y is true, is equal to the sum of the probability that both X and Y are true and the probability that X is true and Y is false.

However, if we have a whole set of alternative possibilities: $Y_1, Y_2, Y_3 \dots Y_N = \{Y_k\}$ instead of having a proposition Y and its negative counterpart, then the probability that X is true is:

$$p(X | I) = \sum_{k=1}^M p(X, Y_k | I) \quad [11b]$$

This is the general form of the marginalization equation [11], which in case of *continuum limit* is:

$$p(X | I) = \int_{-\infty}^{\infty} p(X, Y | I) dY \quad [11c]$$

which can be derived using again the *product rule*:

$$p(X, Y_k | I) = p(Y_k | X, I) \times p(X | I) \quad \text{as long as:} \quad \sum_{k=1}^M p(Y_k | X, I) = 1$$

The normalization requirement is important if $\{Y_{kj}\}$ form a mutually exclusive and exhaustive set of probabilities (which means that, if one of Y_k 's is true then all others must be false, but one of them must be true). Marginalization gives us a tool to deal with nuisance parameters, such as unwanted background, which necessarily enter the analysis but are of no intrinsic interest.

2.2 Bayesian approach in data analysis

The rules of probability can be used in data analysis. It becomes clear if one replaces X and Y in Bayes' theorem with model *parameters* P and *data* D :

$$p(P | D, I) \propto p(D | P, I)p(P | I). \quad [12]$$

Bayes' theorem gives the ability to calculate the probability that a hypothesis (certain values of the parameters) is true given the data. One can use the knowledge about the probability that would have observed the measured data if the hypothesis is true. The various terms in Bayes' theorem have formal names in data analysis:

$p(P | D, I)$ is called the **posterior**, it represents our knowledge about the truth of the hypothesis in light of the data.

$p(D | P, I)$ is called the **likelihood**. Likelihood represents our expectance of how the data would look like if it had been simulated with the given set of parameters.

$p(P | I)$ is the probability density of the model parameters P given the background information I only. It is called **prior** and represents the knowledge (or ignorance) about the model parameters before the experiment had been performed.

The proportionality in equation [12] replaces the equality from equation [8]. It is, because the term $p(Y | I)$, here: $p(D | I)$, has been omitted.

The missing denominator is a constant (does not depend on the hypothesis, and the data does not vary). It is called the *evidence* or total probability and can be used to compare different models.

2.3 Bayesian analysis in biology and medicine

Twenty five years ago, Edwin T. Jaynes (Jaynes, 1984) predicted the following scenario:

“Bayesian methods ... apply also to a mass of new problems that cannot be formulated at all in orthodox terms; and computers are now ... performing very nontrivial data analysis in such diverse fields as spectrum estimation, medical instrumentation, ... and what will probably become the largest area of application, biological macromolecular structure determination.”

Bayesian approach is still only rarely used in biology. However, novel techniques have recently appeared where an enormous amount of data is collected. For example, the proteomic methods which simultaneously detect and measure the expression of hundreds or thousands of proteins present in a biological sample. This area of research is gaining increased attention in biomedical research. Therefore developing of new analysis methods is required and one can be sure that Bayesian methods will find an application in this fast developing field and Jaynes' words will become to be truth.

One of such an example is Bayesian analysis of microarray data (<http://www.bamarray.com/>, 2009). Microarray is a multiplex technology used in molecular biology and in medicine, which evolved from Southern blotting, where fragmented DNA is attached to a substrate and then probed with a known gene or its fragment. Nowadays microarray systems can consist of arrayed series of thousands of microscopic spots of DNA oligonucleotides, each containing picomoles of a specific DNA sequence. This can be a short section of a gene or other DNA elements that are used as probes to hybridize cDNA or cRNA samples (called targets). Probe-target hybridization is usually detected and quantified by fluorescence-based detection of fluorophore-labeled targets to determine relative abundance of nucleic acid sequences in the target. DNA microarrays can be used to measure changes in expression levels. The analysis of DNA microarrays poses a large number of statistical problems, including normalization of the data. A basic difference between

microarray data analysis and more traditional biomedical research is the dimensionality of the data. A large clinical study might collect 100 data items per patient for thousands of patients. A medium-size microarray study will obtain many thousands of numbers per sample for hundred samples. Many analysis techniques treat each sample as a single point in a space with thousands of dimensions, then attempt by various techniques to reduce the dimensionality of the data to something humans can visualize. BAM (Bayesian Analysis of Variance for Microarrays) is a new statistical technique for detecting differentially expressing genes from microarray data (Ishwaran and Rao, 2005).

A similar problem of an enormous amount of data appears in the proteomics field. One popular proteomic method is matrix assisted laser desorption and ionization, time-of-flight mass spectrometry (MALDI-TOF). A number of preprocessing steps have to be performed before MALDI-TOF data can be modeled, regardless of the ultimate approach used for inference. These steps include baseline correction, normalization, denoising, and transformation. It was demonstrated (Morris et al., 2006) how Bayesian analysis can be used to model MALDI - TOF proteomics data.

Bayesian probability theory is also well-suited to formalize and solve macromolecular structure determination problems. It was demonstrated that a full Bayesian analysis of NMR data is feasible (Rieping et al., 2005). A major advantage of the Bayesian approach over optimization-based techniques is its ability to cope with nuisance parameters. Moreover, the precision of all hypothesis parameters was obtained. That way, it was possible to provide an objective measure of precision for NMR structures.

The Bayesian approach was also used for single molecule experiments, such as single molecule tracking (Yoon et al., 2008). The mean of diffusion coefficients, number of dimensions and fluorescence intensity of the spots of interest was used as prior information. This method could detect weak signals and make estimates closer to real trajectories than the deterministic method. Also, it could restore lost positions caused by single blinks since it considers spatiotemporal information in the sequence of images.

2.4 NPS - bayesian analysis for single molecule FRET data

Previously, simple triangulation of several FRET distance measurements was used to determine an unknown position of RNA within the RNA polymerase II structure (Andrecka et al., 2008). However, this method could only estimate the most likely position and did not take into account existing experimental uncertainties that might affect the determined position. Moreover, it was impossible to analyze more than 3 distance measurements consistently.

Here, for the first time, Bayesian analysis is used for FRET data in order to get structural information about macromolecular complex (Muschiello et al., 2008). Bayesian parameter estimation is used to compute the three dimensional probability density for the position of the 'antenna dye molecule' (ADM) using measured FRET efficiencies from several 'satellite dye molecules' (SDMs). The ADM is a dye molecule attached to a macromolecular complex which position is unknown and is our object of interest. The SDMs are dye molecules attached within a macromolecular complex as a network of "satellites". This method due to its similarities to the Global Positioning System (GPS) was termed 'Nano Positioning System' (NPS).

2.4.1 Intuitive description of the analysis method

To understand the influence of experimental uncertainties (in the case of triangulation) it is good to imagine the situation presented below (Figure 8).

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, due to errors in the determined Förster radii or measured FRET efficiencies as well as uncertainties about the actual SDM positions (the dye molecules are attached via flexible linkers) the true position of the ADM can deviate significantly from the intersection point.

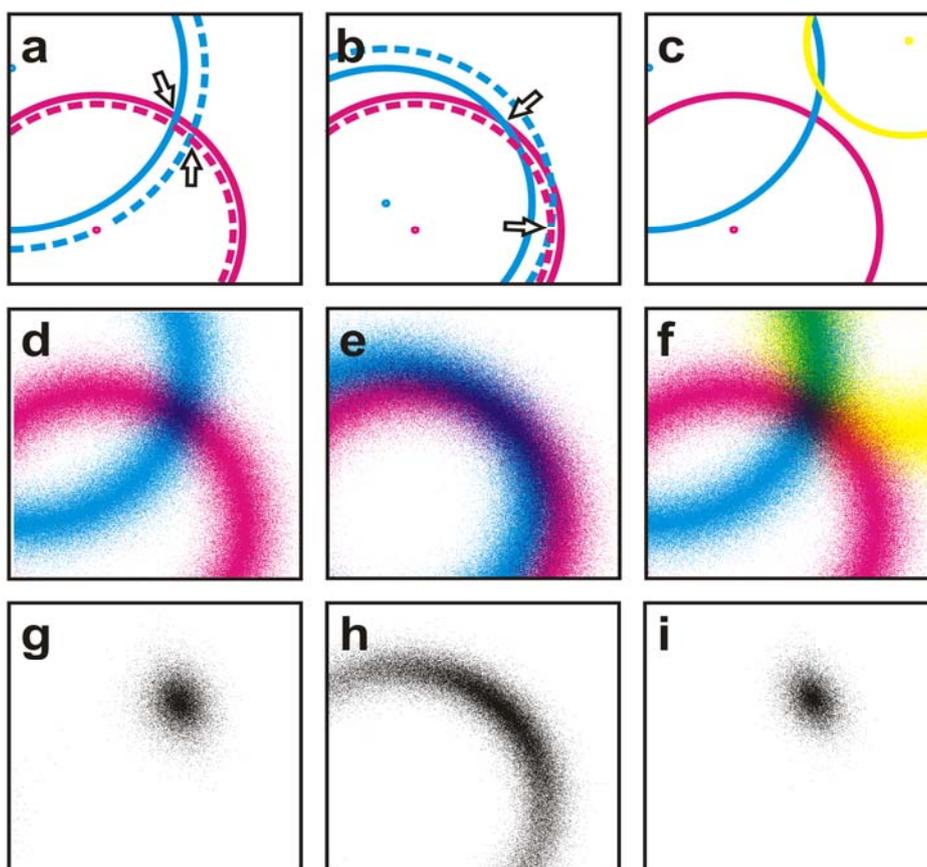


Figure 8: The concept of fuzzy spheres

To present the concept of fuzzy spheres, one can use the two-dimensional analogue, fuzzy circles. Schematic diagrams show the conventional approach for calculating the intersection (indicated by arrows) of circles with known centres. In general, the circles with an experimentally determined radius (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 $\sim 90^\circ$ 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). Due to experimental errors more than two circles may not have a common intersection at all (c). The overlaps of fuzzy circles are shown in grey scale (g-i) and they naturally reflect the uncertainty of the position determination. The density of black points corresponds to the probability of the overlap. For any given experimental error, the uncertainty in actual position depends strongly on the angle between the sphere surfaces at the intersection point. Taken from (Muschiolok et al., 2008).

In the case of four or more distance measurements, the spheres will in general not have a common intersection point (Figure 8(c)). It is obvious that the hard constraints (spheres) have to be softened in order to take the error sources into account. If we assume the uncertainties of the SDM positions to be isotropic, then it is evident that the constraints will be spherically symmetric and we end up with fuzzy spheres (Figure 8(d)-(f)). Mathematically these constraints can be described by probability densities and one can calculate their overlap (Figure 8(g)-(i)), which is defined as the product of the individual densities and has a shape that reflects our knowledge of the ADM position.

Within the Bayesian framework, it is possible to account for experimental errors as well as for errors which simply come from the lack of knowledge (Muschielok et al., 2008). From the calculated probability density distribution, one can directly obtain the most likely ADM position and the associated three dimensional credibility intervals that reflect the experimental uncertainties.

3 RNA Polymerase II

3.1 Transcription

Protein expression is a complex, multi-step process essential for every living cell based on the information carried by DNA. However, as DNA is not the direct template for protein synthesis, an intermediate in the form of another nucleic acid, messenger RNA (mRNA), is required. Transcription of the genetic information from DNA into mRNA is the first step of gene expression.

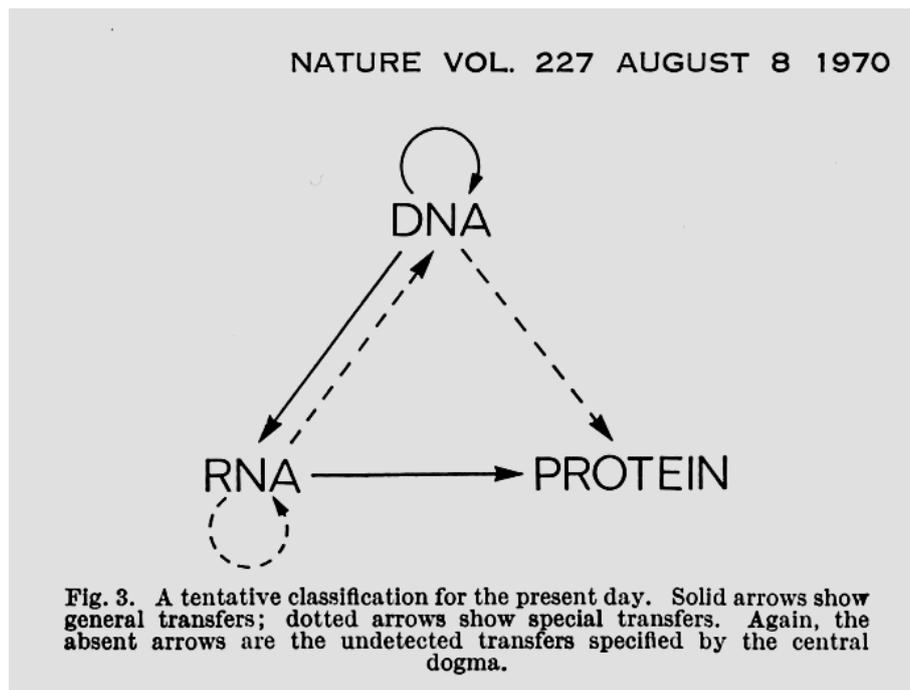


Figure 9: The flow of genetic information, proposed by F. Crick in 1970

Fragment of the original paper where the concept of flow of genetic information was proposed. The Central Dogma of Molecular Biology was enunciated before, in 1958, also by F. Crick, stated that the information can not be transferred from protein to either protein or nucleic acid.

The concept of mRNA as the information-carrying intermediate in protein synthesis was proposed by F. Jacob and J. Monod (Jacob, 1961) and experimentally confirmed by S. Brenner, F. Jacob and M. Meselson (Brenner et al., 1961). This novel concept accelerated the search for an enzyme that produces mRNA molecule in a DNA-templated manner.

The enzyme transcribing DNA into RNA was discovered independently by J. Hurwitz (Hurwitz et al., 1961) and S. Weiss (Weiss and Nakamoto, 1961) and termed "RNA polymerase". The earliest evidence for the DNA dependence of RNA polymerase was the finding that the base composition of mRNA is complementary to that of the DNA template (Hall and Spiegelman, 1961).

Ten years later an inverted flow of genetic information was found in retroviruses. The RNA-dependent DNA polymerase transfers the information from a viral RNA-based genome to DNA in the process of reverse transcription (Baltimore, 1970; Temin and Mizutani, 1970).

3.2 Transcription machinery

Transcription is carried out by the protein complexes named DNA dependent RNA polymerases. In prokaryotic cells, thousands of different genes are transcribed by a single enzyme. The overall architecture of this machine is characterized by a core enzyme including the five subunits: β , β' , two α and ω and as well as additional subunit σ , which is essential for binding of the polymerase to its promoter. Structures of bacterial RNA polymerases elongation complexes from *Thermus aquaticus* and *Thermus thermophilus* (Vassylyev et al., 2002; Vassylyev et al., 2007b; Zhang et al., 1999) are known. The two largest subunits β and β' form a crab claw - like structure harbouring the DNA binding channel with the catalytic center. One α subunit is associated with β and the other with the β' -subunit. The small ω subunit localizes around the C-terminus of β' .

Bacterial and eukaryotic RNA polymerases exhibit striking structural similarities in their overall shape and relative position of the subunits. A comparison of the subunit composition of different multi-subunit RNA polymerases is shown in Table 1. In eukaryotes, the task of gene transcription is distributed among three enzymes, RNA polymerase I, II and III. A fourth RNA polymerase was recently discovered in plants (Herr et al., 2005; Kanno et al., 2005), where it is claimed to play a role in the silencing of repetitive DNA sequences by RNA-directed DNA methylation (Till and Ladurner, 2007). Each of the eukaryotic RNA polymerases is dedicated to transcription of a specific set of genes and is subject to a specific regulation.

RNA Polymerase I (Pol I) is specialized in high-level transcription of rDNA. Thereby, it creates a single precursor RNA transcript, which is processed into a mature 28S, 5.8S and 18S rRNA.

All mRNAs are produced by a single enzyme - RNA polymerase II. Additionally to different protein-coding genes, Pol II transcriptome includes also the genes of non-coding RNAs such as small nuclear (sn), small nucleolar (Esnouf) or micro (mi) RNAs.

RNA Polymerase III (Pol III) is dedicated to the synthesis of small non-coding RNAs, including transfer RNAs, 5S ribosomal RNA, U6 small nuclear RNA and many other RNAs of known and unknown functions. It is the most diverse group of all RNA polymerases transcripts. A wide collection of genes transcribed by this enzyme encodes structural or catalytic RNAs, which are shorter than 400 base pairs.

Pol I, II, and III comprise 14, 12, and 17 subunits, and have a total molecular weight of 589, 514, and 693 kDa, respectively. Ten subunits form a structurally conserved *core*, and additional subunits are located on the periphery. The two, largest subunits are the homologues of the β and β' subunits of the bacterial enzyme. Five subunits are identical in all three eukaryotic polymerases and two other are shared between Pol I and Pol III systems (Table 1).

Table 1: Subunit composition of multisubunit RNA polymerases (Cramer et al., 2008)

RNA polymerase	Pol I	Pol II	Pol III	Archaea	Bacteria
Ten-subunit core	A190	Rpb1	C160	A' + A''	β'
	A135	Rpb2	C128	B (B' + B'')	β
	AC40	Rpb3	AC40	D	α
	AC19	Rpb11	AC19	L	α
	Rpb6 (ABC23)	Rpb6	Rpb6	K	ω
	Rpb5 (ABC27)	Rpb5	Rpb5	H	-
	Rpb8 (ABC14.5)	Rpb8	Rpb8	-	-
	Rpb10 (ABC10 β)	Rpb10	Rpb10	N	-
	Rpb12 (ABC10 α)	Rpb12	Rpb12	P	-
	A12.2	Rpb9	C11	-	-
Rpb4/7 complexes	A14	Rpb4	C17	F	-
	A43	Rpb7	C25	E'	-
TFIIF-like subcomplex ^a	A49	(Tfg1/Rap74)	C37	-	-
	A34.5	(Tfg2/Rap30)	C53	-	-
Pol III-specific subcomplex	-	-	C82	-	-
	-	-	C34	-	-
	-	-	C31	-	-
Number of subunits	14	12	17	11 (12)	5
Molecular weight in kDa (species)	589 (<i>S. cerevisiae</i>)	514 (<i>S. cerevisiae</i>)	693 (<i>S. cerevisiae</i>)	380 (<i>P. furiosus</i>)	375 (<i>T. aquaticus</i>)

^a The two subunits in Pol I and Pol III are predicted to form heterodimers that resemble parts of the Pol II initiation/elongation factor TFIIF, which is composed of subunits Tfg1, Tfg2, and Tfg3 in *Saccharomyces cerevisiae*, and of subunits Rap74 and Rap30 in human.

3.3 RNA Pol II transcription cycle

Gene transcription by RNA Polymerase II involves multiple steps (Cramer, 2007; Hahn, 2004). The mRNA transcription cycle can be roughly divided into three stages: initiation, elongation and termination.

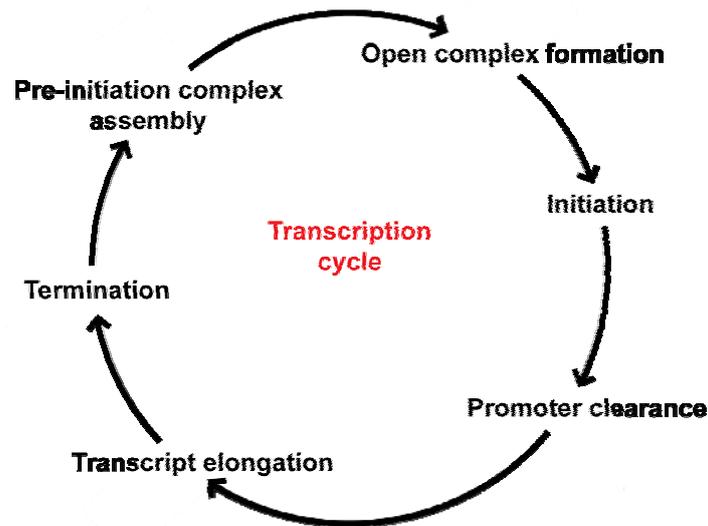


Figure 10: The transcription cycle. Adapted from (Svejstrup, 2004)

During initiation, RNA polymerase recognizes a DNA promoter sequence and melts or separates the two strands to form the single stranded transcriptional bubble. The preinitiation complex (PIC) is formed when RNA Pol II is positioned on the promoter (Bushnell et al., 2004; Chen et al., 2007; Miller and Hahn, 2006). Then, DNA is melted and the template strand slips inside the cleft of the enzyme. After this transition into the open complex, the enzyme can start synthesis of the first phosphodiester bond of RNA. Release of the promoter-assembled transcription factors and transition to the elongation phase occurs after synthesis of about 30 nucleotides (Figure 10).

Elongation is an active phase of the cycle, which involves movement of the polymerase and production of a growing transcript. The polymerase moves with single nucleotide steps along the DNA template, but is also capable of reverse translocation, termed backtracking, which can lead to pausing or arrest of the enzyme.

The termination step completes the transcription cycle. Guided by the DNA sequence at the end of the gene, Pol II terminates transcription and the RNA is cleaved and polyadenylated upon a signal on the transcript. Pol II is then removed from the DNA and recycled to enter another round of transcription (Gilmour and Fan, 2008). Although the cleavage and polyadenylation reactions have been studied extensively (Wahle, 1992; Wahle and Keller, 1992), little is known about how Pol II transcription terminates and how the enzyme is released from the DNA.

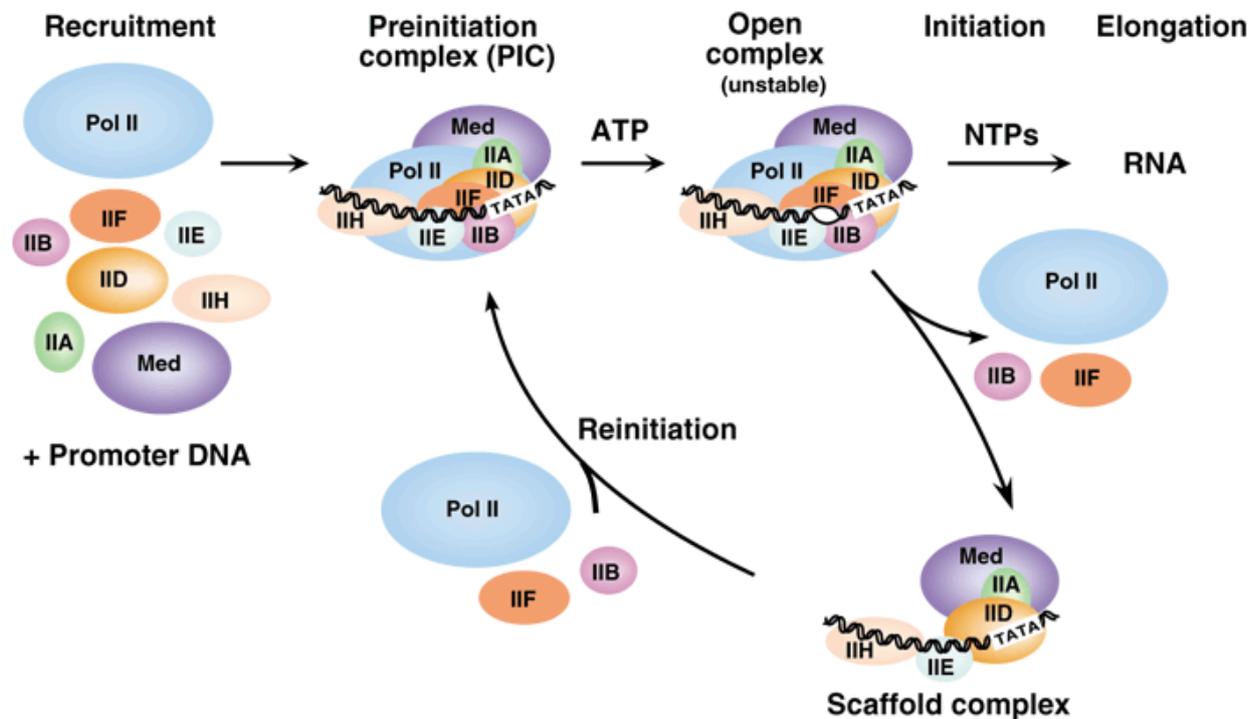


Figure 11: The pathway of transcription initiation and reinitiation for Pol II (Hahn, 2004)

The mRNA transcription cycle is the most complex, and it involves a plethora of factors in addition to Pol II, many of which are important for gene regulation. Most of the regulation takes place during the initiation phase. Recognition of the promoter sequences, as well as formation of the preinitiation complex and, subsequently, of the open complex is facilitated by several, transcription factors and coactivators, specific for Pol II (Table 2). Other factors are responsible for chromatin modification in order to allow polymerase to move along the transcribed gene (Armstrong, 2007; Kulaeva et al., 2007; Li et al., 2007).

The mRNA production by the elongating Pol II is controlled by several protein factors and coupled to the subsequent events of mRNA biogenesis, like cleavage, capping, polyadenylation and nuclear export. The C-terminal domain of Rpb1 (CTD), with its specific phosphorylation pattern, plays a key role in recruiting factors coordinating the transcription cycle (Buratowski, 2003; Hirose and Manley, 2000; Meinhart et al., 2005).

Table 2: *Saccharomyces cerevisiae* Pol II general transcription factors and coactivators

Factor	number of subunits	Function (Hahn, 2004)	
TFIIA	2	Stabilizes TBP and TFIID-DNA binding. Blocks transcription inhibitors. Positive and negative gene regulation.	
TFIIB	1	Binds TBP, Pol II and promoter DNA. Helps fix transcription start site.	
TFIID	TBP	1	Binds TATA element and deforms promoter DNA. Platform for assembly of TFIIB, TFIIA and TAFs.
	TAFs	14	Binds INR and DPE promoter elements. Target of regulatory factors.
Mediator	24	Binds cooperatively with Pol II. Kinase and acetyltransferase activity. Stimulates basal and activated transcription. Target of regulatory factors.	
TFIIF	3	Binds Pol II and is involved in Pol II recruitment to PIC and in open complex formation.	
TFIIE	2	Binds promoter near transcription start. May help opening or stabilizing the transcription bubble in the open complex.	
TFIIH	10	Functions in transcription and DNA repair. Kinase and two helicase activities. Essential for open complex formation. Mutations in IIH can cause human disease.	

3.4 Structure of RNA Pol II elongation complex

...life is a structure...

John Cage

from J. Cage's performance:

I have nothing to say and I am saying it...

Structural studies of eukaryotic transcription have been focused on RNA Polymerase II (Pol II), an enzyme that produces all mRNA in eukaryotic cells. Regulation of Pol II transcription underlies cell differentiation and development therefore it was an object of special interest for structural studies over the last decades.

In early 1990s, initial electron microscopic studies revealed the overall shape of the enzyme at 16 Å resolution (Darst et al., 1991). This structure, as the 3D reconstruction of 2D crystals, showed a number of striking features of Pol II: an arm surrounding a 25Å diameter channel, which extends into a groove of about the same diameter; a narrower channel, about half this diameter that branches off from the 25Å diameter channel; and a tunnel through the enzyme that opens into the floor of the narrow channel. This very first structural information, gave insight into the interaction of the enzyme with DNA and RNA in the course of transcription (Figure 12). Later, two different conformations of the enzyme were observed by EM (Asturias et al., 1997), corresponding to the open and the closed form of polymerase.

A 6 Å crystallographic map of core Pol II was obtained with phases derived from a Ta₆Br₁₂²⁺ heavy atom cluster. For the first time, more robust information of the cleft and the mobile clamp could be observed suggesting functional roles for two mobile protein domains: the tip of a flexible arm forms a downstream DNA clamp; and a hinged domain which may serve as an RNA clamp, enclosing the transcript from about 8 - 18 residues upstream of the 3'-end in a tunnel (Fu et al., 1999).

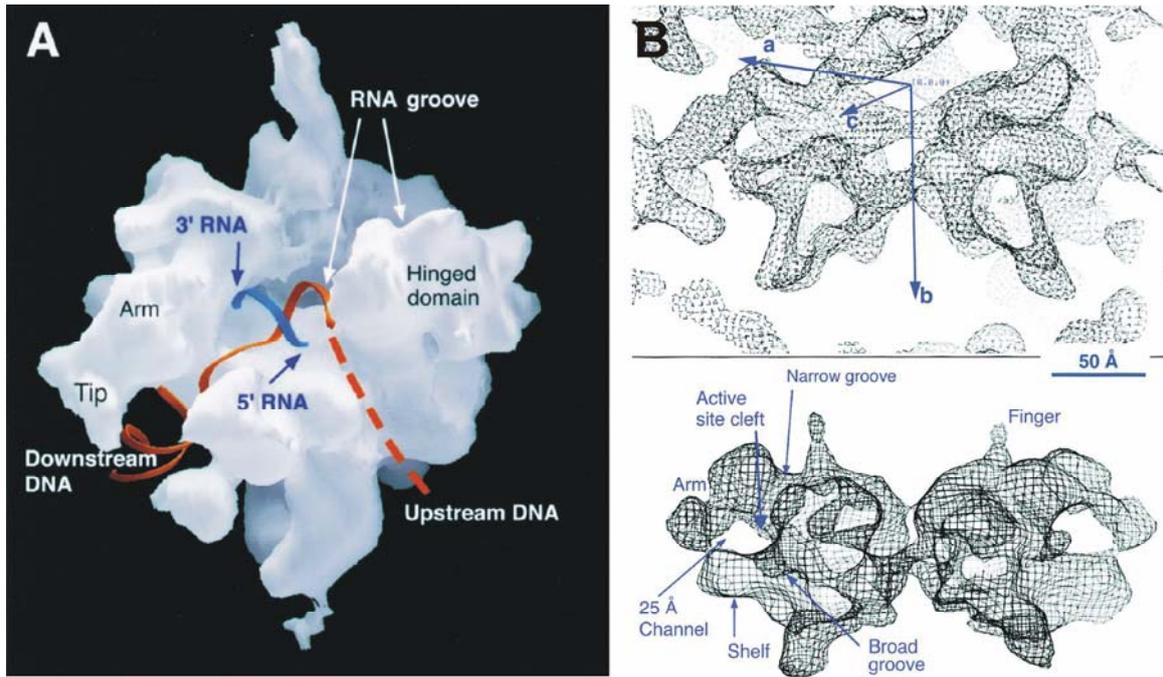


Figure 12: History of Pol II structure determination: 6 Å electron density map being the first density obtained from 3D crystals in comparison with molecular envelope derived from 16 Å EM structure. Adapted from (Fu et al., 1999)

(A) Stereo view of the surface (white) of RNA polymerase II derived from the X-ray map, with models for the RNA transcript upstream and downstream DNA. The likely trajectory of upstream DNA (dashed line). (B) (top) Molecular Envelopes of RNA Polymerase II in Crystals. The envelope displayed as a single isocontour representing about 90% of the molecular volume. a, b, and c denote the unit cell axes of the 3D crystal. (bottom) The molecular envelope from electron crystallography with a contour level representing about 70% of the molecular volume, previously identified features are indicated.

The low resolution structural data was excelled by a backbone model of core Pol II at 3.1 Å resolution (Cramer et al., 2000). This model revealed the subunit architecture of Pol II for the first time. Arrangement and secondary structure elements as well as functional domains like jaws, clamp, wall and the active site could be identified.

Finally, nearly atomic models of core Pol II and a Pol II-DNA/RNA complex could be obtained at 2.8 Å and 3.3 Å resolution, respectively (Cramer et al., 2001b; Gnatt et al., 2001a) and a complete picture of the 12 subunit RNA polymerase II complex which included the Rpb4/7 subcomplex at 4.2 Å resolution (Armache et al., 2003).

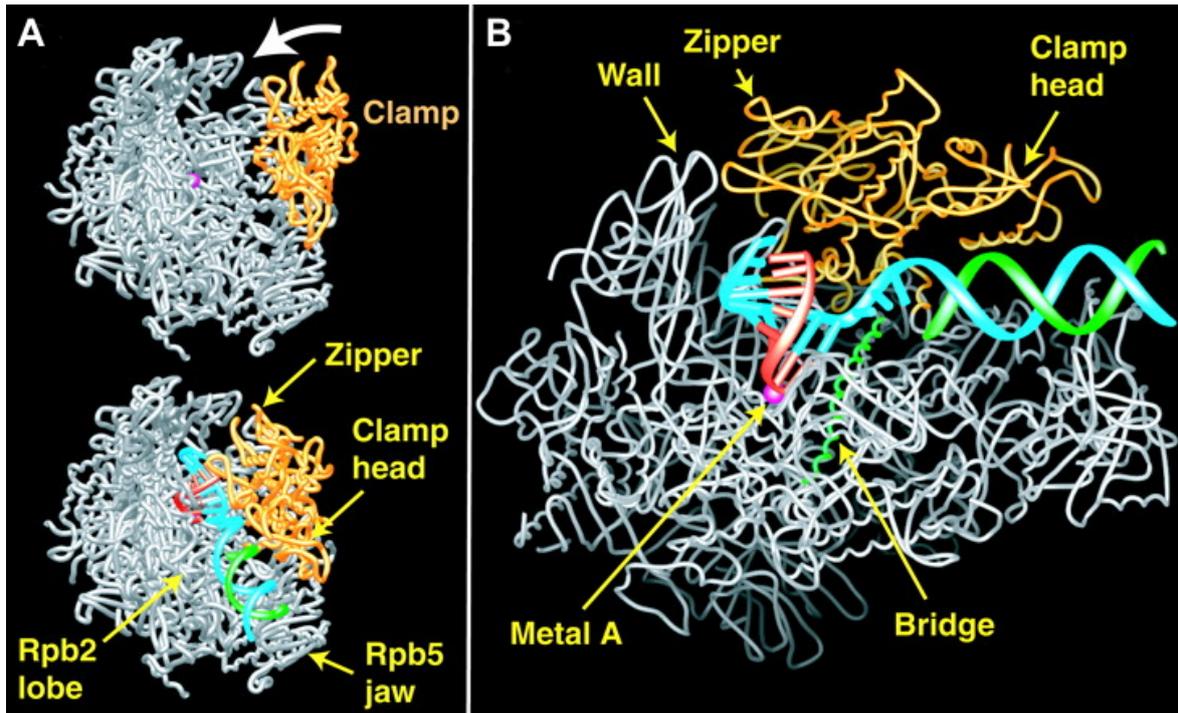


Figure 13: Crystal structure of polymerase II transcribing complex

(A) Comparison of structures of free (top) Pol II and transcribing complex (bottom). Clamp (yellow) closes on DNA and RNA. (B) RNA polymerase II transcribing complex. Rpb2 was removed to reveal nucleic acids. Rpb1 bridge helix is highlighted in green and active site metal is shown as pink sphere. This figure is adapted from rom (Gnatt et al., 2001a)

One year later the elongation complex structure, its interactions with NTP and TFIIIS with 4.0 Å was solved (Kettenberger et al., 2004). It detected template DNA (positions +9 to -10) and nascent RNA (within the DNA-RNA hybrid at positions +1 to -8 and two nucleotides upstream of the hybrid, positions -9 and -10). Nontemplate DNA (positions +2 to +9) in the downstream duplex region was also visible in this structure. In these studies artificial bubbles were used (see *Methods*, Figure 16). However, nontemplate ssDNA strand within the transcription bubble and the upstream dsDNA have still not been observed as well as the nascent RNA product beyond the hybrid, although both were present in the scaffold used for crystallization.

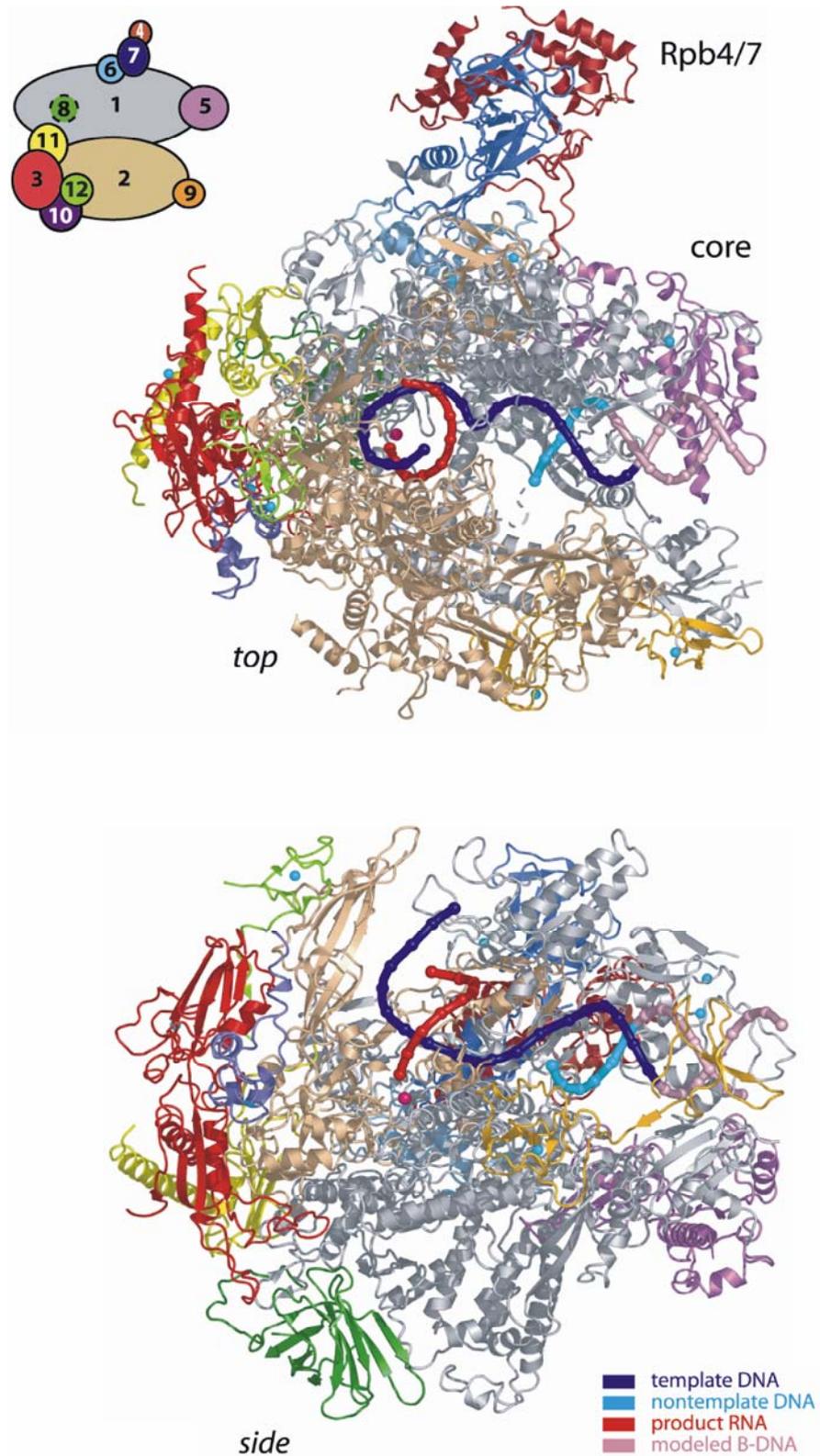


Figure 14: Structure of complete 12-subunit RNA polymerase II elongation complex

Two views of a ribbon model of the protein subunits and nucleic acids, top view (top) and side view (bottom). The Pol II subunits Rpb1–Rpb12 are colored according to the key on top. Template DNA, nontemplate DNA, and product RNA are shown in blue, cyan, and red, respectively. This figure is adapted from (Kettenberger et al., 2004a).

3.5 Single molecule studies of transcription

The power of single-molecule fluorescence techniques for transcription studies had so far been demonstrated only for viral and bacterial RNA polymerases. Fluorescence microscopy was used for visualizing the nonspecific binding of RNAP to DNA providing evidence for sliding as a mechanism for promoter search (Kabata et al., 1993). Later, TIRF in combination with an optical tweezer was used to observe RNAP binding to the DNA in a sequence dependent manner (Harada et al., 1999).

Binding of individual polymerase molecules to the promoter DNA was also studied with atomic force microscopy (AFM) (Guthold et al., 1994). This technique was additionally used to study open complex formation, it was shown that the DNA contour length of the polymerase open complex was reduced by approximately 30 nm (~90 bp) relative to the free DNA. The DNA bend angle measured varied from 55 to 88 degrees. These results strongly supported the notion that, during transcription initiation, the promoter DNA wraps nearly 300 degrees around the polymerase (Rivetti et al., 1999). Magnetic tweezer, another single molecule force spectroscopy technique was employed to study changes in DNA super-coiling associated with the formation of the open complex (Revyakin et al., 2004).

Fluorescence resonance energy transfer was used to monitor distances within single molecules of the abortively initiating transcription initiation complexes (Kapanidis et al., 2006; Revyakin et al., 2006). Abortive cycles of synthesis and release of short RNA products occur before the polymerase escapes the promoter sequence. The mechanism of initial transcription was explained using single molecule FRET technique. It was shown that, in the very early stage of transcription the polymerase remains bound to the promoter DNA and pulls downstream DNA into itself producing short RNA products. When the nascent RNA reaches ~13 nucleotides the polymerase can enter the elongation phase. This step of the transcription cycle was studied as well (Margeat et al., 2006; Rivetti et al., 2003).

An NTP incorporation cycle is composed of multiple reaction steps, namely translocation, NTP binding, hydrolysis and pyrophosphate release. The most fundamental question concerns about how catalysis is coupled to mechanical translocation. The first model, the *Brownian ratchet model*, assumes that the enzyme can slide back and forth on the DNA driven by thermal energy. In the second model, the *power-stroke model*, chemical energy derived from the NTP incorporation drives the forward translocation of the polymerase. Single molecule force spectroscopy has been used to study the dynamics of elongation by the bacterial enzyme (Adelman et al., 2002). These experiments could elucidate kinetic aspects of transcriptional pausing and backtracking (Forde et al., 2002; Shaevitz et al., 2003). By improving the resolution of the tweezer experiments, details of the translocation mechanism were obtained (Abbondanzieri et al., 2005; Harada et al., 2001; Skinner et al., 2003; Thomen et al., 2005).

The last phase of transcription, termination, has been examined the least so far and only two examples exist in the literature where termination was studied by single molecule technique, namely by the optical tweezer (Larson et al., 2008; Yin et al., 1999).

3.6 Open questions and scope of this work

Fluorescence Resonance Energy Transfer (FRET) in combination with single molecule techniques can be used for structural studies. It provides the possibility to obtain very precise information about the distance between two dyes attached to a macromolecule. Information about at least three distances within the structure can be used to determine a previously unknown position by means of triangulation.

In this work, single molecule FRET was used to map the path of the nascent RNA within yeast RNA polymerase II elongation complex. Information about three distances was used for this task. Fourth distance measured was used as consistency control. These experiments could estimate the most likely position of the RNA but they were not able to calculate how existing experimental uncertainties might effect the determined position.

In order to use FRET as a quantitative technique, experimental uncertainties have to be taken into account. Therefore, a new analysis method has been developed. The Nano Positioning System (NPS) combines x-ray crystallographic information, single-molecule FRET data and bayesian parameter estimation, a probability based analysis method. This work shows for the first time the application of the NPS for the determination of the position of the 29 nucleotide long RNA and the influence of transcription factor IIB (TFIIB) on its position.

In the last part of this work one more piece of the transcription puzzle is put to its place by the presentation of a complete Pol II elongation model. FRET efficiencies were measured between dye molecules attached to the non-template or up-stream DNA and several dye molecules attached to known positions within the Pol II elongation complex. Recently developed Nano Positioning System was then used to determine nontemplate and upstream DNA positions. Obtained results provided a framework to define the path of the nucleic acids across the elongation complex of RNA polymerase II.

Methods

4 Molecular Biology Methods

Standard molecular biology procedures such as isolation of DNA, restriction enzyme analysis, polymerase chain reaction (PCR), cloning, agarose gel electrophoresis, etc., were essentially carried out as described (Sambrook and Russell, 2001).

4.1 Single cysteine Rpb4/7 mutants

The eucaryotic RNA polymerase II enzyme consists of a 10 subunit core and Rpb4/7 subcomplex. The 10 subunit core of RNA polymerase II is purified from yeast cells (Edwards et al., 1990), whereas the Rpb4/7 heterodimer can be over-expressed in bacterial expression system (Sakurai et al., 1999). There are only five native cysteines in the Rpb4/7 complex, four within the Rpb7 subunit at positions C38, C47, C94, C150 and one at position C185 of Rpb4. Therefore Rpb4/7 lends itself for the attachment of a single dye molecule to a defined and known location within the elongation complex.

In order to produce two mutants with a single cysteine residue at locations C94 and C150, respectively, the site directed mutagenesis was performed replacing native cysteines with serines. Moreover, the Rpb4/7 (-)Cys mutant with all native cysteines replaced by serines was used to obtain two additional mutants with the cysteines at the new position S16C and S73C of Rpb7 and Rpb4, respectively (Figure 15).

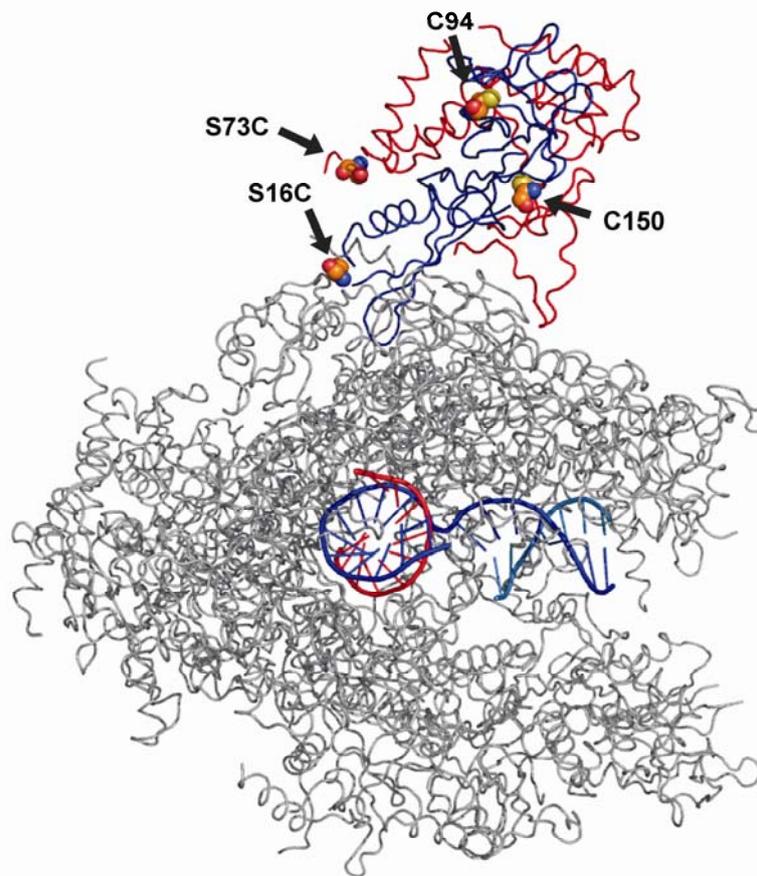


Figure 15: Single cysteine Rpb4/7 mutans

Positions of 2 native cysteines (C94 and C150 of Rpb7, spheres representation) and 2 serines replaced by cysteines (S16C and S73C of Rpb7 and Rpb4, respectively). These positions were used as attachment points for fluorescent dyes. For each of the mutants only one of the cysteins shown was present. RNA polymerase II core (10 subunits) is presented in gray, Rpb4 in red and Rpb7 in blue, all as ribbons. Nucleic acids are presented as cartoons, template DNA in dark blue, nontemplate DNA in light blue and RNA in red.

Site-directed mutagenesis by an overlap extension method was performed in order to obtain the Rpb4/7 mutants. Mutations were carried out by complementary “overlapping” primers, which in combination with “flanking” primers were used to generate two DNA fragments having overlapping ends. These fragments were combined in a subsequent “fusion” reaction, in which the overlapping ends anneal, allowing the 3' overlap of each strand to serve as a primer for the 3' extension of the complementary strands.

Rpb4/7 constructs were amplified from full length Rpb4/7 (Table 3). For the ligation of each construct into the vector pET-21b (Novagen), all primers carried either NdeI (forward) or NotI (reverse) restriction sites and were inserted into the vector using these restriction sites.

Table 3: Primers used for mutagenesis of Rpb4/7 subcomplex

Rpb4_NdeI_for	5' - ggaattccatatgaatgtttctacatcaacctttc
Rpb7_NotI_rev	5' - ctgctagctgcgccgcaatagcaccCAAATAATcttctttg
Rpb7-C38/47S_for	5' - ggttcgagtacgggtaaattcggatataattcttagtgctcctag
Rpb7-C38/47S_rev	5' - ctaggacactaagaatataatccgaatttaccctactcgaccc
Rpb7-C94S_for	5' - cagtcgtttcaagttctcagcacggg
Rpb7-C94S_rev	5' - cccgtgctgagaacttgaaacgactg
Rpb7-C150S_for	5' - gttAAAATTgaaggtagtatcagtcaag
Rpb7-C150S_rev	5' - cttgactgatactagcttcaattttaac
Rpb4-C185S_for	5' - ggttctttggccagtgacacagctgatg
Rpb4-C185S_rev	5' - catcagctgtgtcactggcCAAAGAACC
Rpb7-S16C_for	5' - cccttcatccgtgttttttcggctcctcg
Rpb7-S16C_rev	5' - cgaggaccgAAAAAACACGGATGAAGGG
Rpb4-C73S_for	5' - gcatttAAAAGATGTCAAAAAAACAC
Rpb4-C73S_rev	5' - gtgttttttttgacatcttttAAATGC

4.2 Transformation

A 50 µl aliquot of competent cells and plasmid DNA were thawed on ice. 1 µl of plasmid DNA (or ligation mixture) was added to XL-1 blue (used for cloning) or BL21(DE3) RIL cells (used for protein expression), which were incubated on ice for 15 minutes, followed by a 'heat shock' performed at 42 °C for 60 seconds. After short cooling on ice and adding 1ml of LB medium, the cells were incubated at 37 °C for 45 minutes. The cells were finally plated on LB agar plates containing antibiotics. The plates were incubated overnight at 37 °C.

4.3 Expression of recombinant protein in E.coli

A fresh colony was picked within 24 h after having started transformation, and a starter culture was grown over-night in 50 ml LB medium containing antibiotics. LB medium (two liter per five liter flask) supplemented with antibiotics was inoculated and shaken (180 rpm) at 37 °C until the culture reached log phase ($OD_{600} \approx 0.7 - 0.8$).

After cooling the *E. coli* suspension on ice, IPTG was added in a 1:2000 ratio and the cells were grown over night at 18 °C. Cells were collected by centrifugation (4000 rpm, SLS6000 rotor, 4°C) and suspended in Lysis Buffer, flash-frozen in liquid nitrogen and stored at -80 °C.

In case when the expression in LB Medium failed, so called Auto-inducing Medium (Studier, 2005) was used. A starter culture was grown over-night in standard LB medium. For preparing 1 liter of Auto-inducing Medium about 925 ml of ZY medium was complemented with 50 ml of 20 x NPS buffer, 20 ml 50 x 5052 solution and 1 ml 1M $MgSO_4$. Auto-inducing Medium (two liter per five liter flask) supplemented with antibiotics was inoculated and shaken (180 rpm) at 30 °C for about 3 - 4 h until the culture reached log phase (OD 0.7 - 0.8). Then, the temperature was shifted to 18 °C and the cells were grown over night.

Luria Bertani (LB) medium:

1% tryptone

0.5% yeast extract

0.5% NaCl

pH 7.0

1.5–2 % (w/v) of Bacto-Agar was added to the medium to prepare LB-agar plates

Auto-inducing Medium:

ZY medium	20 x NPS buffer	50 x 5052 solution	MgSO₄ solution
10 g tryptone	66 g (NH ₄) ₂ SO ₄	250 g glycerol	1 M MgSO ₄
5 g yeast extract	136 g KH ₂ PO ₄	25 g glucose	
925 ml H ₂ O	142 g Na ₂ HPO ₄	100 g α-lactose	
	900 ml H ₂ O	730 ml H ₂ O	

Supplements/Antibiotics:

100 mg/ml ampicillin (in H₂O)

50 mg/ml chloramphenicol (in ethanol)

1 M IPTG (in H₂O)

4.4 Purification of the Rpb4/7 subcomplex

Recombinant yeast Rpb4/7 was expressed in E. coli BL21(DE3) (Stratagene). Cells were grown in 2 x 2 L of culture in Auto-induction medium. After approximately 4 h, when OD₆₀₀ ≈ 0.6 was reached, the temperature was shifted from 30°C to 20 °C. After 11 h, the cells were harvested by centrifugation (15 min at 5000 rpm in a SLC6000 rotor), suspended in 50 ml buffer 1 and lysed by means of sonification. The lysate was cleared by centrifugation (30 min at 15000 rpm in a SS34 rotor) and applied onto a NiNTA column (Quiagen; 1 ml column volume) previously equilibrated with buffer 1. Subsequently the column was washed with 3 ml of buffer 1, 3 ml buffer 1 containing additionally 10 mM imidazole and 3 ml of buffer 1 containing additionally 20 mM imidazole. Elution was performed subsequently with 3 ml of buffer 1 additionally containing 50 mM imidazole and 3 ml of buffer 1 additionally containing

200 mM imidazole. Peak fractions were pooled, diluted 1:3 with buffer 2 and applied on a ResourceQ column (GE Healthcare, 6 ml column volume), equilibrated in buffer 2. Rpb4/7 was eluted with a linear gradient from 0-1000 mM NaCl in buffer 2. The purified Rpb4/7 heterodimer was concentrated to 6 mg/ml and aliquots were stored at $-80\text{ }^{\circ}\text{C}$ in assembly buffer.

Buffer 1

150 mM NaCl
5 % (v/v) glycerol
50 mM Tris pH 7.5
10 mM β -mercaptoethanol
protease inhibitors

Buffer 2

50 mM Tris pH 7.5
5 mM DTT
1 mM EDTA
(1M NaCl for gradient)

Assembly buffer

50 mM Hepes
40 mM $(\text{NH}_4)_2\text{SO}_4$
5 μM ZnCl_2
5% glycerol
5 mM DTT

100 x protease inhibitor mix (P.I.)

1.42 mg Leupeptin
6.85 mg Pepstatin A
850 mg PMSF
1650 mg benzamidine
dry ethanol to 50 ml

/ stored at $-20\text{ }^{\circ}\text{C}$; added immediately before usage

4.5 Protein labeling

To allow for quantitative FRET experiments, site-specific labeling was desired. For this purpose, the reactivity of thiol groups being present in cysteines was used. For labeling of thiol groups, maleimide derivatives of fluorescent dyes were used leading to a covalent linkage of the dye to the protein via a thioether moiety.

Dye labeling of the single-cysteine mutants was conducted using 8 - 10 times molar excess of Alexa647-C2-Maleimide (Molecular Probes) in assembly buffer (without DTT) for one hour at 25° C. Labeled protein was further purified using G-50 spin-columns (Amersham Biosciences, Freiburg, Germany) with assembly buffer containing 10 mM DTT.

The degree of labeling, *DOL*, was determined using a UV-VIS spectrophotometer (Varian Inc. Cary 50). Protein concentration, *c*, was calculated using the extinction coefficient $\epsilon = 10240 \text{ M}^{-1} \text{ cm}^{-1}$, which was calculated with ProtParam tool from ExPASy (<http://au.expasy.org/tools/protparam.html>, 2009). Dye absorption at 280 nm was taken into account and in case of Alexa647, 3% of the absorption maximum (A_{dye}) was subtracted:

$$c = (A_{280} - A_{\text{dye}} \times 0.03) / \epsilon$$

$$DOL = (A_{\text{dye}} / \epsilon_{\text{dye}}) / c$$

A labelling efficiency of ~20% was obtained. In order to verify the specific labeling of the Rpb7 subunit SDS-PAGE was performed and Alexa647 labeled protein was viewed using Amersham Typhoon 9410 Imager. The presence of a single fluorescent band corresponding to Rpb7 or Rpb4 (for Rpb4-S73C mutant) confirmed the high specificity of the labeling procedure (Figure 16).

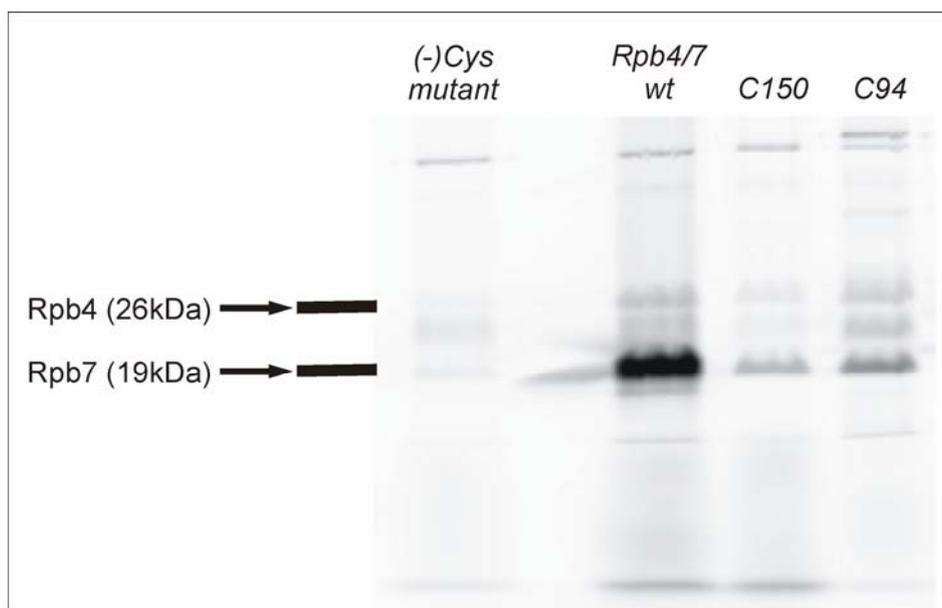


Figure 16: Specific detection of the Rpb4/7 complex labeled with Alexa647

SDS-PAGE gel image was taken using Amersham Typhoon 9410 Imager. Rpb4 and Rpb7 subunit positions are marked. In order to check for unspecific labeling the mutant without any cysteines, (-)Cys, was used (first lane). Rpb4/7 wilde type with all 5 native cysteines was used as positive control (second lane). Last two lanes show labeled Rpb4/7 single cysteine mutants C150 and C94 of the Rpb7 subunit. All mutants were labeled using an identical procedure as described.

4.6 Nucleic acid scaffolds

Artificial elongation bubbles were constructed from RNA and DNA oligomers using partially mismatched DNA template and non-template strands (Figure 17). The DNA and RNA strands were purchased from IBA (Göttingen, Germany) and Biomers (Ulm, Germany). The same sequences have been used previously for solving the complete Pol II elongation complex structure (Kettenberger et al., 2004a).

DNA-RNA bubbles were annealed using equimolar amounts of template DNA, non-template DNA and RNA in TE buffer (10 mM Tris-HCl, 1mM EDTA) at a final concentration of 2 μ M. The mixture was heated to 95° C for 5 minutes followed by fast cooling to 55° C and then slow cooling to 4° C for one hour. Ready DNA-RNA bubbles were stored at -20 °C.

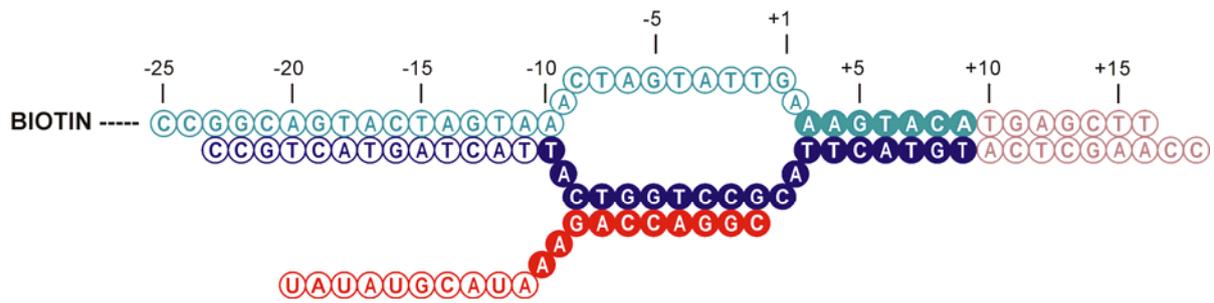


Figure 17: Nucleic acid scaffold

Sequence of the bubbles, which have been used previously for solving the complete Pol II elongation complex structure (Kettenberger et al., 2004a). Filled circles represent bases which are visible in the structure. This construct contains 20nt long RNA, for determining the nascent RNA position, longer RNA was used (Table 4).

Table 4: Oligomers used for bubbles preparation

<i>name</i>	<i>sequence (5' → 3')</i>	<i>modification</i>
T-DNA	ccaagctcaagtacttacgcctggtcattactagtactgcc	-
T-DNAforRNA4	ccaagctcaagtacttacgccagggtcattactagtactgcc	-
T-DNA (-10)	ccaagctcaagtacttacgcctggtc <u>a</u> tactagtactgcc	alexa647 at <u>T28</u>
T-DNA (+3)	ccaagctcaagtact <u>t</u> acgcctggtcattactagtactgcc	alexa647 at <u>T16</u>
T-DNA (+9)	ccaagctca <u>t</u> gtacttacgcctggtcattactagtactgcc	alexa647 at <u>T10</u>
NT-DNA	ccggcagtactagtaaactagtattgaaagtacatgagctt	5'end biotin
NT-DNA (+1)	ccggcagtactagtaaactagtatt <u>t</u> aaagtacatgagctt	5'end biotin alexa555/TMR - <u>T26</u>
NT-DNA (-2)	ccggcagtactagtaaactagta <u>t</u> tgaaagtacatgagctt	5'end biotin alexa555/TMR - <u>T24</u>

Methods

NT-DNA (-4)	ccggcagtactagtaaactag <u>t</u> attgaaagtacatgagctt	5'end biotin alexa555/TMR - <u>T22</u>
NT-DNA (-7)	ccggcagtactagtaaact <u>t</u> agtattgaaagtacatgagctt	5'end biotin alexa555/TMR - <u>T19</u>
NT-DNA (-12)	ccggcagtactag <u>t</u> aaaactagtattgaaagtacatgagctt	5'end biotin alexa555/TMR - <u>T14</u>
NT-DNA (-15)	ccggcagtact <u>t</u> agtaaactagtattgaaagtacatgagctt	5'end biotin alexa555/TMR - <u>T11</u>
NT-DNA (-18)	ccggcag <u>t</u> actagtaaactagtattgaaagtacatgagctt	5'end biotin alexa555/TMR - <u>T8</u>
RNA1	augcauaaaagaccagg <u>c</u>	3'end alexa647/555/TMR
RNA4	augcauaaaagacc <u>u</u> ggc	alexa647 - <u>U14</u>
RNA10	<u>a</u> agaccaggc	5'end alexa647
RNA17	<u>a</u> ugcauaaaagaccaggc	5'end TMR
RNA20	<u>u</u> auaugcauaaaagaccaggc	5'end TMR
RNA23	<u>a</u> uauauaugcauaaaagaccaggc	5'end TMR
RNA26	<u>c</u> gaauauauaugcauaaaagaccaggc	5'end TMR
RNA29	<u>a</u> cacgaauauauaugcauaaaagaccaggc	5'end TMR

4.7 Preparation of Pol II - artificial bubbles elongation complex

The Pol II-DNA-RNA-Rpb4/7 complex was assembled as described (Armache et al., 2003). 3 pmol of nucleic acid scaffold were mixed with 2 pmol of 10 subunit core Pol II in 15 μ l of assembly buffer. After 10 minutes, 5 fold excess of the Rpb4/7 subcomplex (10 pmol) was added. The complete elongation complex was purified using Microcon–YM100 centrifugal filter units (Millipore, Molsheim, France). The membrane was first blocked using 100 μ l of 1 mg/ml BSA in assembly buffer (30sec, 14 000 rcf), then Pol II-DNA-RNA-Rpb4/7 complex was loaded onto the membrane and purified against 2 x 450 μ l assembly buffer. Assembled complexes were stored at +4 °C for a maximum of 2 days.

4.8 Competition measurements with TFIIB

The TFIIB protein used in these experiments was expressed for 15 hours at 18° C in *E.coli* BL21(DE3)RIL cells (Stratagene). Cells were harvested by centrifugation, resuspended in buffer A and lysed by French Press.

After centrifugation, the supernatant was loaded onto a Ni-NTA column (Qiagen) equilibrated with buffer A. The column was washed with 10 ml of buffer B and 10 ml of buffer A. Proteins were eluted stepwise with buffer A containing 10, 20, 50 and 200 mM imidazole. Eluted fractions were diluted 3-fold with buffer C and further purified by cation exchange chromatography (MonoS, Amersham). The column was equilibrated with buffer D and proteins were eluted with a linear gradient of 15 column volumes from 100 mM to 1 M NaCl.

After concentrating the sample, it was applied to a Superose-6 HR gel filtration column (Amersham) equilibrated with assembly buffer. Pooled peak fractions were concentrated, flash-frozen in liquid nitrogen and stored in -80° C.

Buffer A

150 mM NaCl

50 mM Tris pH 7.5

5% glycerol

10 mM β -mercaptoethanol

Buffer B

2M NaCl

50 mM Tris pH 7.5

5% glycerol

10 mM β -mercaptoethanol

Buffer C

50 mM Tris pH 7.5

5 mM DTT

Buffer D

100 mM NaCl

50 mM Tris pH 7.5

5 mM DTT

For the TFIIB competition experiments, elongation complexes containing RNA29 were used. TFIIB was added in 3 fold, 10 fold and 100 fold excess of TFIIB directly to the chamber with elongation complexes attached. Alternatively excess of TFIIB was incubated with nucleic acid scaffolds, followed by addition of RNA polymerase II and Rpb4/7 and purification by using Microcon-YM100 filter as described above.

5 Fluorescence Methods

5.1 Anisotropy measurements

Fluorescence anisotropy measurements were conducted using a steady state fluorescence spectrometer (Edinburgh Instruments F900). The instrumental G-Factor was determined individually for each measurement. The dye molecules were excited at 530 nm (TAMRA, Alexa555) and 630 nm (Alexa647). Emission spectra were collected from 540 - 650nm (TAMRA, Alexa555) and 640 - 750nm (Alexa647), Anisotropy was calculated as an average of 20 data points around the maximum of emission.

Each measurement was performed on elongation complexes at room temperature using freshly prepared complexes with only one dye-labeled component (DNA, RNA or Rpb7, respectively) at a concentration of 100 nM in a volume of 60 μ l of assembly buffer.

5.2 Förster radii measurements

The standard procedure (Vamosi et al., 1996) was used to determine the Förster radii (R_0). To determine donor quantum yield, Rhodamine 101 in ethanol was used as a standard.

First, absorptions of both donor sample and standard were measured (concentration \approx 2 μ M). Donor sample was diluted to obtain an absorption \sim 0.05. Reference sample was diluted in EtOH to absorption \sim 0.016 (if the expected QY was \sim 30%, e.g. TMR).

Emission spectra from 555 to 700 nm were taken with excitation at 550 nm. Taking into account the dilution factor, the quantum yield was calculated by using the value of absorption at 550 nm and fluorescence (integrated emission spectrum). Donor emission spectra from 528-700 nm with an excitation wavelength of 525 nm and acceptor absorption spectra from 400-700 nm were taken in order to compute the overlap integral.

The Förster radius was calculated (R_0^{iso}), with an assumption of free rotational diffusion of the dye molecule ($\kappa^2=2/3$) and the refractive index $n = 1.3492$, using *PhotochemCAD* (<http://www.photochemcad.com/>, 2009).

All fluorescence measurements were conducted in bulk using a steady state fluorescence spectrometer (Edinburgh Instruments F900) and for all UV-VIS measurements a spectrophotometer (Varian Cary 50) was used. All spectra were corrected for nonlinearity of the detector.

5.3 Preparation of sample chamber for sp-FRET measurements

Flow cells were constructed similar to a procedure published by Bennink (Bennink M. L., 2001) from a sandwiched piece of sealing film (Nescofilm), a quartz glass slide (with two holes for tubings) and a coverslip.

After thorough cleaning with detergent (Hellmanex II, Hellma) and ultra-pure water (Millipore Simplicity 185) using the manufacturer's instructions, slides and coverslips were dried in nitrogen flow. Remaining particles were oxidized using a butane-gas torch. The cleaned glass slides were then incubated in a two-percent (v/v) solution of (3-aminopropyl)-triethoxysilane (Sigma-Aldrich, Schnelldorf, Germany) in spectroscopically clean acetone for 30 minutes for silanisation. The silanised glass slides were then rinsed with ultrapure water and dried with nitrogen. Hereafter, a

solution of mPEG-succinimidyl propionate (400 mg/ml, MW 5000 Da, Nektar Therapeutics, Huntsville, Alabama, U.S.A.) and biotinylated PEG-NHS (15 mg/ml, MW 3400 Da, Nektar Therapeutics) in carbonate-bicarbonate buffer at pH 9.4 was applied (incubated for 1 hour), followed by cleaning with ultra-pure water and drying with nitrogen. The PEGylated quartz glass slides were then assembled together with a pre-cut sealing film and the coverslip and heated to 130° C for one minute, in order to allow for the thermoplastic film to seal the channel for the later sample chamber.

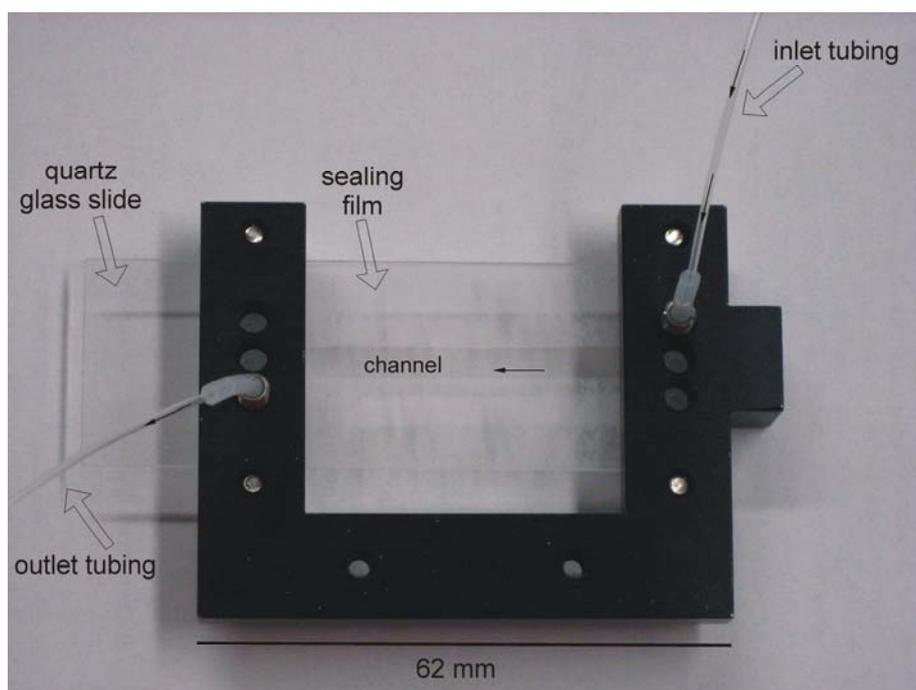


Figure 18: Sample chamber for sp-FRET measurements

Poll elongation complexes were attached to the surface via a PEG-Biotin/Neutravidin/Biotin attachment. To this end the surface of the sample chamber was incubated with 0.5 mg/ml Neutravidin (Molecular Probes) in phosphate buffered saline (PBS) for 15 minutes at room temperature. After exchanging the buffer with assembly buffer, the chamber was pre-bleached by illuminating the entire chamber with the DPSS laser (Cobolit Calypso, 491 nm) at full power (50 mW) (Figure 18). Pre-assembled elongation complexes (~20 pM) were loaded and incubated for 15 minutes. Afterwards, unbound complexes were removed by washing extensively with assembly buffer.

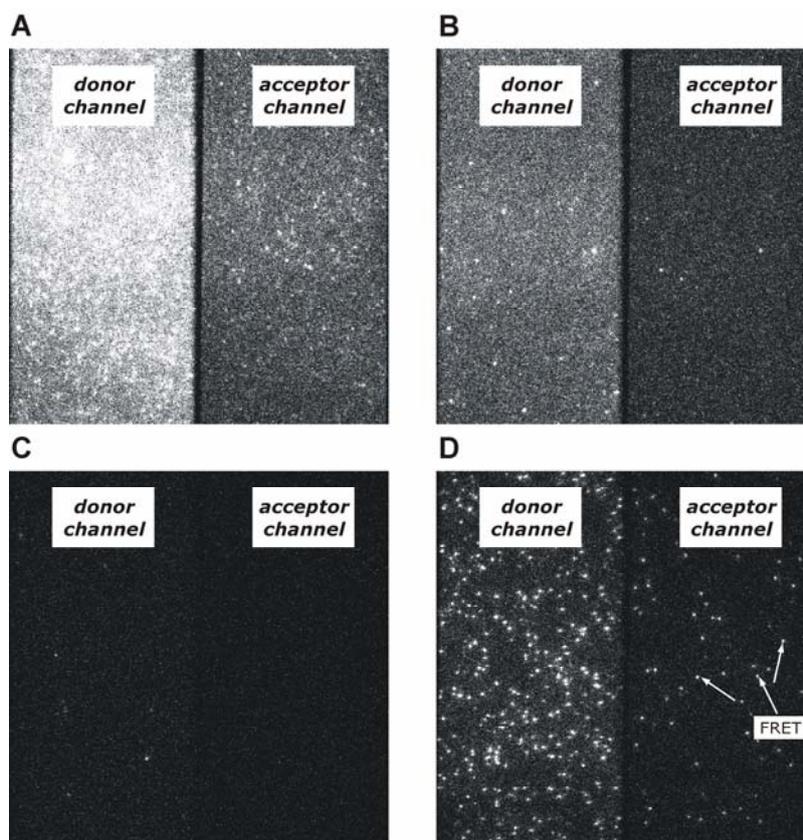


Figure 19: Chamber bleaching and loading the sample

(A) Picture of the chamber illuminated using blue laser (491 nm, full power - 50 mW) before bleaching (B) and after bleaching. (C) Chamber before loading sample, illuminated using green laser (532 nm, 25 mW). (D) picture of the chamber with loaded sample (here: 50 pM of donor-acceptor labeled 40 bp DNA).

5.4 Experimental setup for sp-FRET

Single-pair FRET experiments were performed on a homebuilt prism-based total internal reflection fluorescence microscope (TRIFM) (Figure 21). In this apparatus, the sample was excited using both a frequency-doubled Nd-YAG laser (Crystalaser) at 532 nm for the excitation of donor molecules and FRET pairs and a He-Ne laser (Uniphase) at 633 nm for the direct excitation of the acceptor. The light from the two lasers was combined spatially using a dichroic mirror (Chroma Z532RDC). During the measurement the excitation could be alternated between the two laser sources using two software controlled shutters (Uniblitz LS6). Alternating between FRET and direct

acceptor excitation allows for the discrimination of dynamical changes in FRET efficiencies due to conformational changes against changes in FRET efficiencies caused by fluctuations in acceptor brightness (*Results and Discussion*, Figure 31). The excitation light was focused to a prism mounted on top of the sample chamber. For total internal reflection the incident angle was around 74° . The area of illumination was roughly $70 \times 35 \mu\text{m}^2$ with a power density of $1.5 \mu\text{W}/\mu\text{m}^2$. The observed area could be scanned by translating the flow-chamber with a micrometer stage (Nanomax TS, P7MAX302, Thorlabs). Fluorescence intensity was collected via a water-immersion objective (Plan Apo 60x, N.A. 1.2, Nikon) and directed to an intensified, back-illuminated EMCCD-camera (iXon DV897DCS-BV, Andor). Two channels of detection were introduced by physically blocking one half of the EMCCD-chip using a slit in the image plane of the detection path and spatially offsetting the donor and acceptor fluorescence via a dichroic beam splitter (Chroma 645DCXR). Emission wavelengths from 550 – 610 nm (Omega Optical 3RD550-610) and 660 – 760 nm (Chroma HQ710/100M) were isolated using filters in each detection channel.

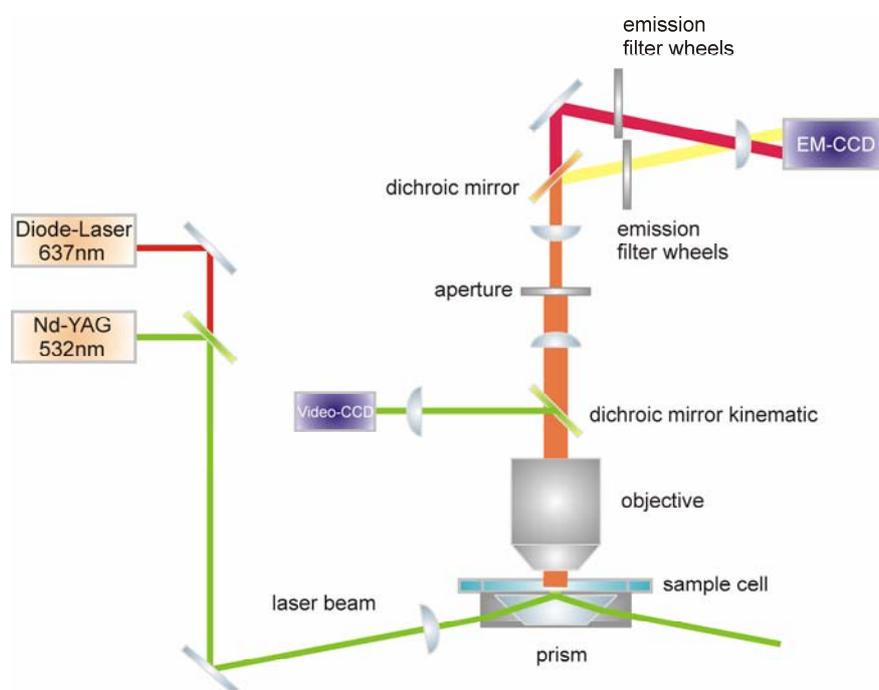


Figure 20: Schematic picture of the setup used for sp-FRET measurements

The excitation beams were focused to a prism mounted on top of the sample cell. For total internal reflection the incident angle was around 74° . Fluorescence intensities were collected via objective and directed to an intensified, back-illuminated EM-CCD camera. Two channels of detection were introduced. Donor and acceptor fluorescence intensities were separated via a dichroic beam splitter and isolated using filters in each detection channel.

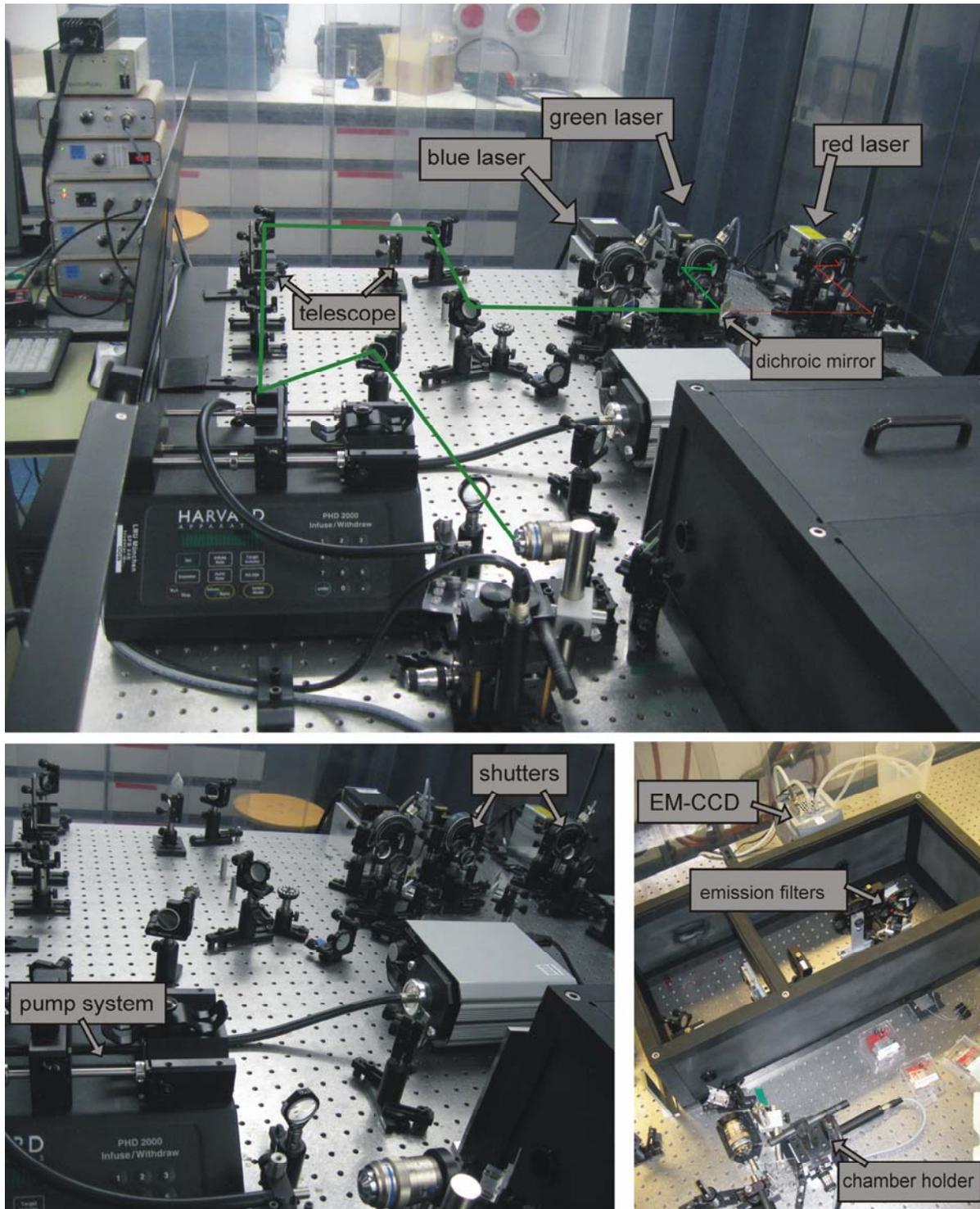


Figure 21: Setup used for sp-FRET measurements

Upper panel: General overview of the setup. Laser beam path marked in green and red. Lower left panel: Excitation path. Lower right panel: Emission path. In order to reduce noise this part of the setup is placed in a black box so that no light from outside can get into the CCD camera.

5.5 Data collection and analysis

5.5.1 Camera settings used in single molecule FRET measurements

SETUP ACQUISITION

Setup CCD:

Acquisition Mode: *Kinetic*, Triggering: *Internal*, Readout Mode: *Image*,
Exposure Time: *0.1015 secs.*, Number of Accumulations: *1*,
Accumulation Cycle Time: *0.1015 secs*, Kinetic Series Length: *250 – 400*,
Kinetic Cycle Time: *0.10324 secs.*, Number of prescans: *0*, Frame Transfer: *Yes*,
Baseline Clamp: *No*, Vertical Pixel Shift Speed: *3.3 μ s*,
Vertical Clock Voltage, Amplitude: *Normal*, Readout Rate: *3 MHz at 14-bit*,
Pre-Amplifier Gain: *5x*, Output Amplifier: *Electron multiplying*,
Electron Multiplier, Gain Enabled: *Yes*, Electron Multiplier DAC Setting: *225*

Binning:

Sub Image: *512 x 512 (Full)*, Binning: *1 x 1*

5.5.2 Analysis program

The acquired data was analyzed using a software written in MATLAB. Fully automated routine to find FRET pairs, calculate and subtract the local background and to compute fluorescence trajectories was used. Detailed description of the program used can be found in the dissertation of Robert Lewis (LMU, 2009).

The very first step was the creation of a so called “bead map”. For this purpose, fluorescent particles with a broad emission spectrum which could be detected in both channels were immobilized in the chamber. Two out of at least five of these particles were then manually selected and their coordinates were used to determine the relative read-out (specifically the pixel deviation in x and y direction between the two channels was determined). In order to find single fluorescent molecules first 25 frames from the acceptor channel and frames 16-38 of the donor channel were averaged and the absolute minimum of number of counts per pixel of the entire video

was subtracted from each pixel. The local background was found in every box of the 16x16 pixel grid, by means of the local minima. Next, in this “model of local minima” a 20x20 pixels grid was homogenized (*imfilter* function). After scaling of intensities from 0 to 255, a threshold above which molecule should be found was set. In this step, the video was scanned with a 7x7 pixel mask-box, searching for local maxima. Signals from actual molecules were checked for shape and presence of other molecule nearby. Important to note, the program searched for the molecules in the acceptor channel. The next step was to find the corresponding molecule in the donor channel. True (original) acceptor and donor intensities were extracted (as a sum of the intensities in the 7x7 pixels box) and corrected by subtracting the local background.

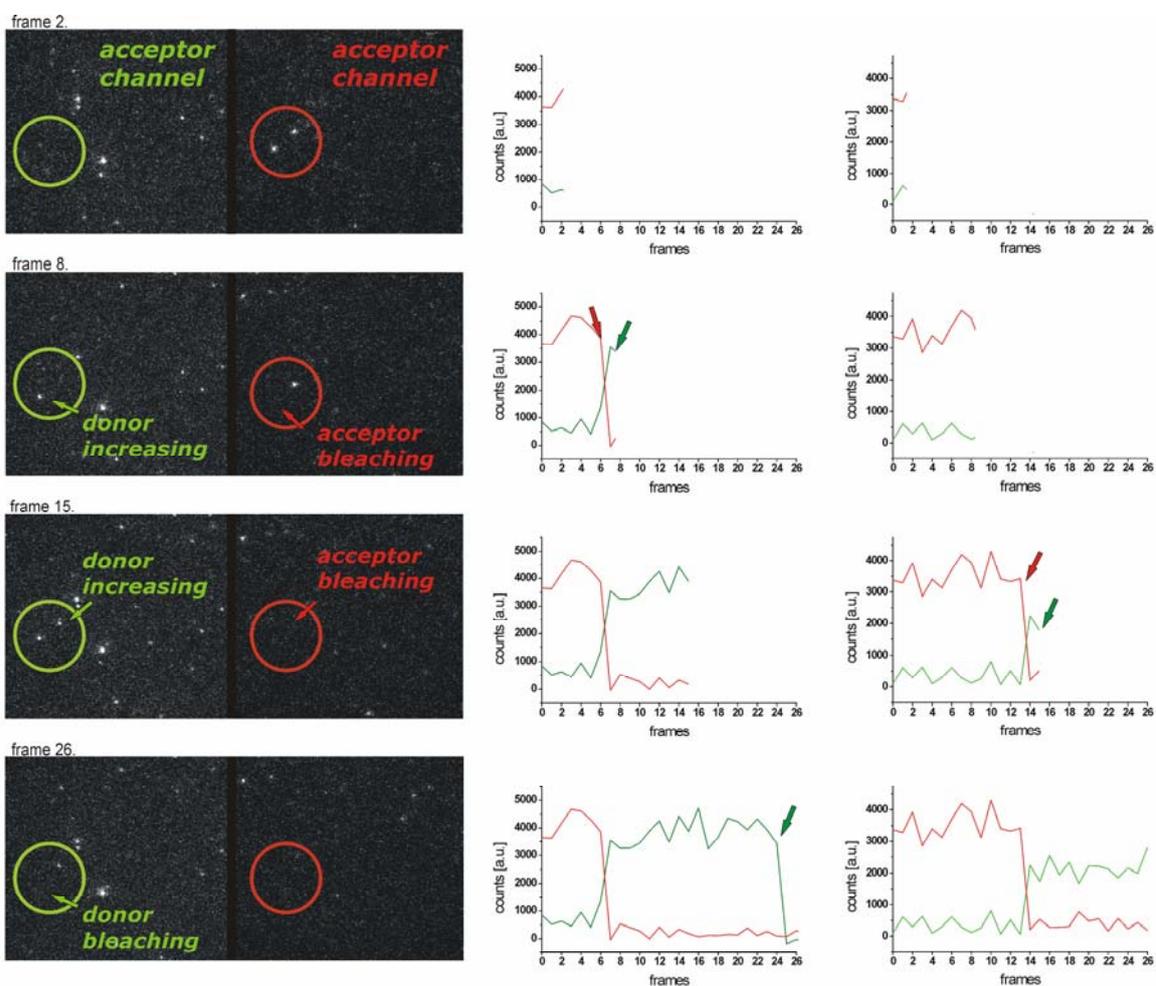


Figure 22: Original data from the camera and extracted donor/acceptor intensities

Left panel: Selected frames from the original data from the camera showing typical behavior of donor and acceptor molecules. Right panel: Time traces of donor (green) and acceptor (red) (frame 2., 8., 15., and frame 26.). Acquisition time of one frame was 100 ms. All traces were corrected for background and used to extract FRET values.

5.5.3 Calculation of FRET efficiencies

For each molecule, the donor and acceptor intensities were collected as described above (Figure 22). Each time trace was corrected for background, channel cross-talk, variations in the fluorophores' quantum yields, and the corresponding detection efficiencies of the two channels.

For calculation of the FRET efficiency of the individual molecules FRET pairs, we used the following formula:

$$E = \frac{I_A - \beta I_D}{I_A + \gamma I_D}$$

where γ and β are correction factors defined as:

$$\gamma = \frac{I_A - I'_A}{I'_D - I_D}$$

$$\beta = \frac{I'_A}{I'_D}$$

I_A and I_D are background corrected intensities from the acceptor and donor channels and I and I' are the intensities before and after acceptor photobleaching, respectively.

β and γ are correction factors; β accounts for the leakage of the donor emission into the acceptor channel, while γ is a factor that includes the quantum yields of the fluorophores and the detection efficiencies of the two channels.

Direct excitation of the acceptor molecules by the green laser was neglected since it was less than 2% of the donor intensity.

Correction factors were calculated for every FRET pair individually. FRET pairs, which did not show acceptor bleaching were discarded from the analysis.

FRET efficiency values were calculated for every time point after filtering the original data sets using a 10 point sliding average filter. For every molecule, also a mean FRET value was calculated. So called “frame wise histograms” of every time point and “molecule wise histograms” of every molecule were prepared. Histograms were then fitted with one or more gaussians and the mean FRET efficiency as well as the standard error were extracted from the fit.

For conventional triangulation, the donor - acceptor separation, r , was calculated. The following formula was used:

$$r = R_0 \left(\frac{1}{E} - 1 \right)^{1/6}$$

Distances, r , calculated from the FRET efficiencies of three donor – acceptor pairs were then used in order to gain information about the unknown position (see *Methods* 6.1.1).

6 Data Analysis Methods

Two methods were used in order to determine the unknown positions within Pol II elongation complex. First, simple triangulation was used to map the RNA exit path (see *Results and Discussion*, chapter 7). Later, Bayesian parameter estimation was applied by means of NPS (see *Results and Discussion*, chapter 8 and chapter 9). Both methods are described below.

6.1.1 Conventional Method - Triangulation

Triangulation of at least three FRET distance measurements allows for determination of the unknown position of the ‘antenna dye molecule’ (ADM). Mathematically, one has to construct spheres around each of the three known positions of the ‘satellite dye molecules’ (SDMs) with the radii determined by means of FRET. The intersection point of the three spheres determines the unknown position. It should be noted that, mathematically, there might be as many as two intersection points. However, one of these points can be ruled out due to distance constraints. Alternatively the correct intersection point can be determined by measuring a fourth distance of a SDM to the ADM. The following procedure was used to calculate the intersection point of 3 spheres (see *Appendix*, A1). The centers of the spheres (attachment points of the acceptor dye molecules, SDMs) were defined by atom selections in the pdb file (1Y1W) and are listed in Table 5.

Table 5: Sphere centers (acceptor attachment points) used for triangulation

<i>attachment point</i>	<i>atom</i>	<i>chain</i>	<i>resi</i>	<i>name</i>	<i>x/Å</i>	<i>y/Å</i>	<i>z/Å</i>
T-DNA(-10)	31450	T	28	C6	77.878	35.805	-16.395
T-DNA(+3)	31206	T	16	C6	95.713	59.792	-15.336
Rpb7-C150	27049	G	150	SG	119.061	27.641	-65.684
Rpb7-C94	26613	G	94	SG	117.336	9.872	-75.956

Sphere's radii were used as input variables. The vmd script (see *Appendix, A1*) was load in the TCL - console of VMD (<http://www.ks.uiuc.edu/Research/vmd>, 2008), the procedure was called by "*rad1 rad2 rad3 cid1 spr* ", where "*rad1*", "*rad2*", "*rad3*" were scalars and represented the distances measured by FRET; "*c1*" was the color-specification number. The result of the routine was the point of intersection which was presented as sphere of radius defined by "*spr*". The result was saved as a *.dat* file at the path specified in the last section of the script (see *Appendix, A1*).

6.2 NPS

6.2.1 Bayesian parameter estimation applied to NPS

In order to apply the Bayesian parameter estimation to the NPS (see *Introduction 2.2*), a physical model describing the experiments must be defined. Additionally, model parameters, priors for the parameters and a likelihood, which mathematically connects the measured data with the parameters, have to be determined.

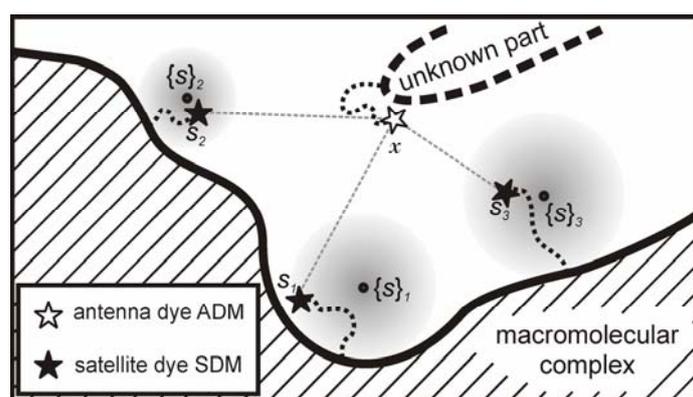


Figure 23: Schematic representation of “antenna” and “satellite” dye positions

The antenna dye molecule (ADM; open star) was attached at position x to a part of the complex of which the position was not known. Several satellite dye molecules (SDM; filled stars) were attached to the complex by flexible linkers. Therefore the position of satellites can not be precisely predicted. The shaded spheres around the most likely positions $\{s\}_i$ represent priors of the SDM positions s_i . The actual distances between the SDMs and the ADM are presented by dotted lines. From these distances, one can calculate the expected FRET efficiencies.

Here, a static model for the positions of the dye molecules was assumed and according to the Förster theory, measurement i of FRET between a particular SDM at position $s_i = (s_{ix}, s_{iy}, s_{iz})$ and the ADM at position $x_i = (x_{ix}, x_{iy}, x_{iz})$, resulted in data E_i . It is important to remember that x_i (position of ADM) and s_i (position of SDMs), are model parameters. The overall situation was illustrated in Figure 23.

6.2.2 Likelihood

Given the positions x and s_i it is expected to observe the FRET efficiency \mathcal{E}_i which depends on the distance between the dye molecules and the Förster radius R_i :

$$\mathcal{E}_i(x, s_i, R_i) = \left(1 + |x - s_i|^6 / R_i^6\right)^{-1} \quad [13]$$

Note that R_i is another model parameter and that \mathcal{E}_i is a function of the parameters but no parameter itself.

In order to compute the likelihood, the fact that the distance measurements between the ADM and different SDMs were assumed to be statistically independent and therefore, $p(\{E\} | x, \{s\}, \{R\}, I)$ factorizes into N factors. Each measurement gives the corresponding FRET efficiency E_i , with error ΔE_i and contributes to the likelihood with a Gaussian factor:

$$p(\{E\} | x, \{s\}, \{R\}, I) = \prod_i p(E_i | x, s_i, R_i, I) = \prod_i \frac{\exp\left\{-[\mathcal{E}_i(x, s_i, R_i) - E_i]^2 / 2(\Delta E_i)^2\right\}}{\Delta E_i \sqrt{2\pi}} \quad [14]$$

6.2.3 Prior

Structure based **geometric constraints** were encoded in the prior. It was assumed that x (position of the ADM) could be anywhere outside of the structure of the elongation complex but stays within a certain distance. Hence, $p(x | I)$ is constant in this region and zero everywhere else. NPS excludes antenna and satellite dye molecule positions that cause a sterical clash with the complex. The clash is defined by Van der Waals contact of either dye or linker with the protein - DNA complex (Figure 24).

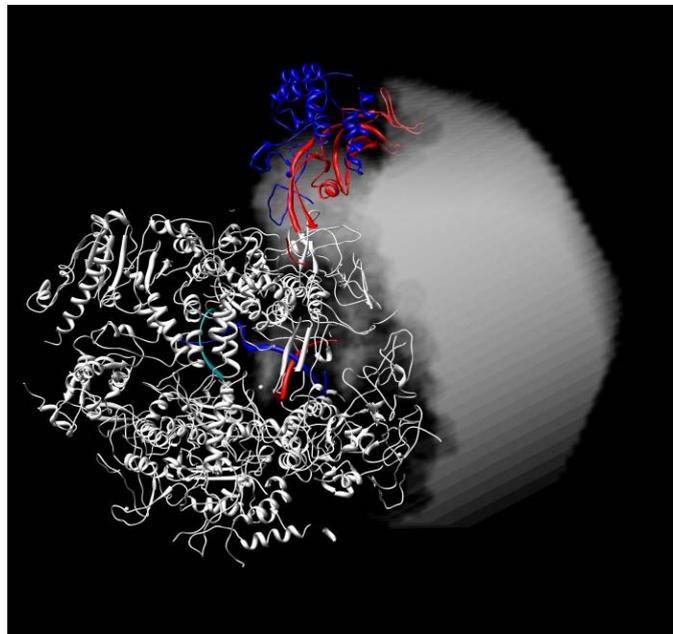


Figure 24: Antenna dye molecule position prior

It was assumed that the antenna dye molecule position prior $p(x | I)$ is constant inside V_A , the volume that is accessible to the antenna dye, and zero outside of V_A . Accessible volume for 29nt long RNA is shown in grey. A flexible linker ($L_{\text{linker}} = 92 \text{ \AA}$ and $D_{\text{linker}} = 9.5 \text{ \AA}$) was attached to the 5' end of the RNA product seen in the crystal structure.

Positions of satellite dye molecules (acceptor) were modeled and used as prior information. SDMs were linked to the elongation complexes using flexible linkers, of which the length was estimated to be 13 \AA (12 C-atoms in the case of DNA and RNA) or 7 \AA (6 C-atoms for Rpb4/7). The diameter of the linker was set to 4.5 \AA and the

diameter of the dye molecule to 7 Å. With this geometrical constraints the total accessible volume was calculated for all SDMs positions.

This model takes into account only the accessible volume of the nucleic acid – protein complex, but does not account for electrostatic interaction. Accessible volume was then approximated by a sum of 15 gaussians with individual widths.

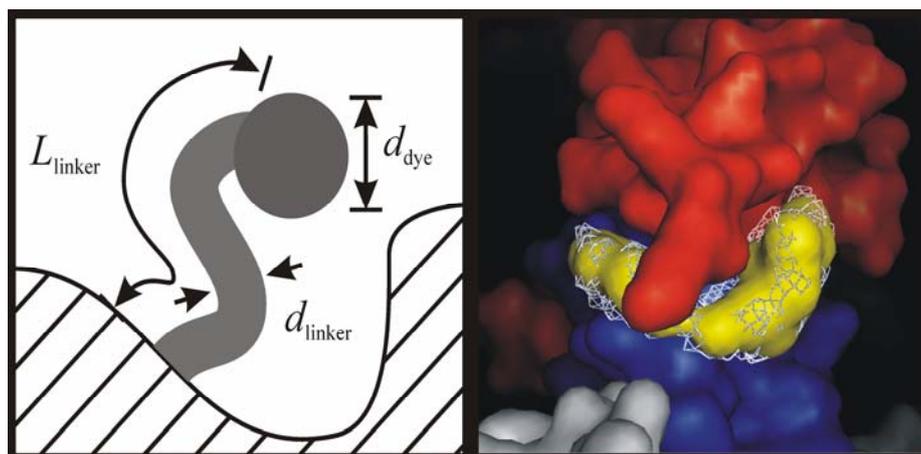


Figure 25: Satellite position prior

Modeled position of satellite dye molecule. SDM was approximated as sphere of diameter d_{dye} and linked to the elongation complex by flexible linker of dimensions L_{linker} and d_{linker} . This model takes into account only the accessible volume of the nucleic acids – protein complex. Right panel: example of satellite position modeling of C73-Rpb4.

Since κ^2 was assumed to be 2/3 in the Förster radii calculations (see *Introduction*, 5.1) obtained values, due to orientational effects, have rather large uncertainties (Dale et al., 1979).

In order to account for these uncertainties, the anisotropies of all ADMs and SDMs positions were measured (see *Methods*, 5.1). Assuming that there is no additional rotational movement beyond the time scale of the fluorescence lifetime, Monte Carlo Simulation was performed in order to calculate the probability densities of the Förster radii assuming an isotropic distribution of the average dye molecule orientation (Dale et al., 1979; Muschielok et al., 2008).

Table 6: Satallite dye molecules attachment points used for modeling prior

<i>attachment point</i>	<i>x/Å</i>	<i>y/Å</i>	<i>z/Å</i>	<i>L_{linker}</i>	<i>d_{linker}</i>	<i>d_{dye}</i>
T-DNA(-10)	77.878	35.805	-16.395	13 Å	4.5 Å	7 Å
T-DNA(+3)	95.713	59.792	-15.336	13 Å	4.5 Å	7 Å
T-DNA(+9)	93.245	79.840	-23.401	13 Å	4.5 Å	7 Å
RNA1	104.989	45.651	8.867	7 Å	4.5 Å	7 Å
RNA4	92.247	41.749	0.537	13 Å	4.5 Å	7 Å
RNA10	85.384	29.111	-15.487	7 Å	4.5 Å	7 Å
Rpb7-C150	119.061	27.641	-65.684	7 Å	4.5 Å	7 Å
Rpb7-C94	117.336	9.872	-75.956	7 Å	4.5 Å	7 Å
Rpb4-C73	147.396	18.485	-40.487	7 Å	4.5 Å	7 Å
Rpb7-C16	133.790	19.266	-27.015	7 Å	4.5 Å	7 Å

For all donor - acceptor pairs, the calculated probabilities were fitted using a sum of 10 pseudo-gaussians. The Gaussians parametrized the simulated distribution and further were used to calculate the **Förster radii priors**, $p(R_i | I)$.

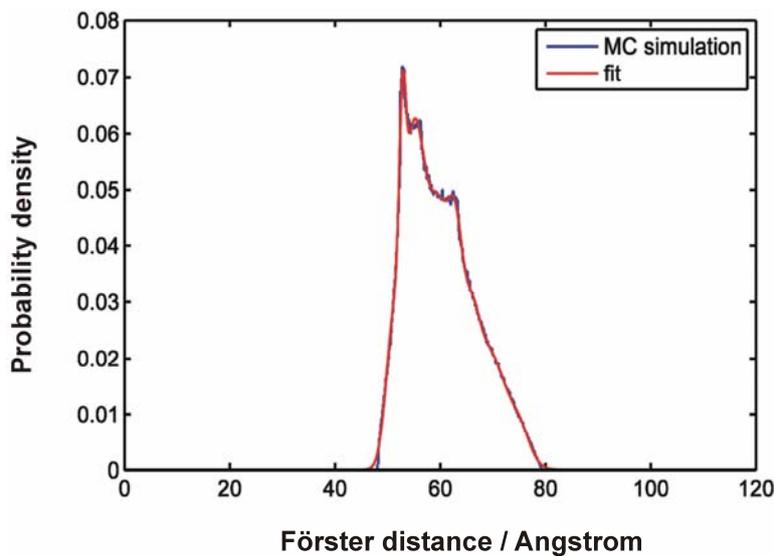


Figure 26: Förster radius prior

Given measured Förster (R_0^{iso}) radii and anisotropies of donor and acceptor the Monte Carlo simulations were performed. Calculated probabilities were fitted using a sum of 10 gaussians.

6.2.4 Posterior

Within the approximations used for likelihood and prior, the posterior probability can be computed. The result is a probability distribution which depends on x , s and R . However, we are only interested in the position of the ADM and consequently, calculate $p(x | E, I)$ by integrating over all possible combinations of SDM positions and Förster radii that result in an intersection in x :

$$p(x | \{E\}, I) = \iint p(x, \{s\}, \{R\} | \{E\}, I) d\{s\} d\{R\} \quad [15]$$

After carrying out the integration, formally called marginalization, the posterior can be written as the product of the prior in x and the product of functions $K_{ij}(x)$.

$$p(x | E, I) = p(x | I) \prod_i K_i(x) \quad [16]$$

Eq. 16 gives the probability density function for the ADM given a set of FRET measurements. For a visualization of the result, one can calculate credibility regions of the ADM position $p(x | E, I)$. These are the smallest volumes that enclose a probability equal to the specified credibility level. The surface of the credibility region can be displayed together with the macromolecular complex (in this case the RNA Polymerase II elongation complex). In contrast to displaying the point with maximum probability, the credibility region gives a direct impression of the estimated position of the ADM and of the experimental uncertainties.

Results and Discussion

7 Single-molecule tracking of RNA exiting from RNA polymerase II

Crystallographic analysis of the Pol II elongation complex detected the position of the nascent RNA within the DNA-RNA hybrid and two nucleotides after the point of DNA-RNA strand separation (Kettenberger et al., 2004b; Westover et al., 2004b). The last ordered RNA nucleotide is located at the entrance to a tunnel, which is formed between the polymerase wall, clamp and lid. Beyond this putative exit tunnel, two prominent surface grooves on either side of the dock domain could in principle further accommodate exiting RNA (Cramer et al., 2000a; Cramer et al., 2001a).

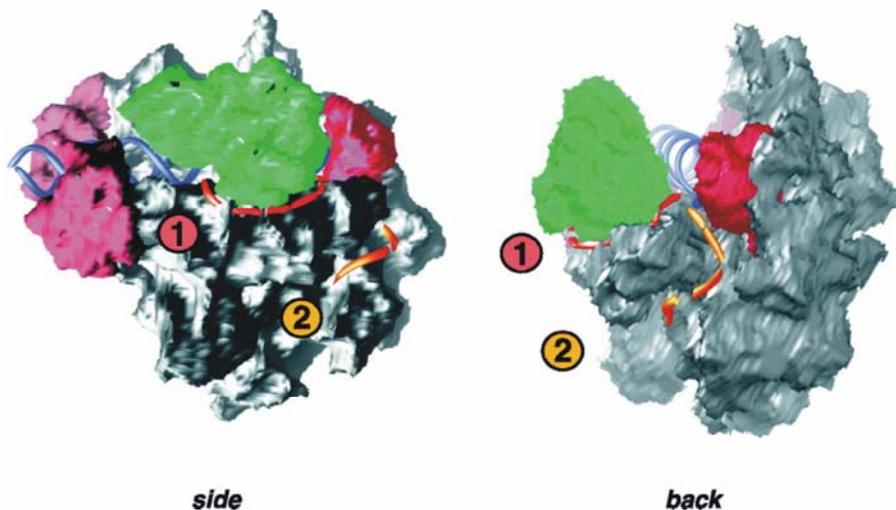


Figure 27: Two proposed RNA exit grooves and funnel beneath the active site

The model is shown in two perpendicular directions of view (side, back). To the side and back views have been added dashed lines corresponding to about 10 nucleotides of RNA, lying in well-defined grooves leading away from the hybrid-binding region (groove 1, red; groove 2, orange). The nontemplate strand of the DNA within the transcription bubble and the upstream DNA duplex are not shown. Adapted from (Cramer et al., 2000b).

Groove 1 winds along the base of the clamp towards the Rpb4/7 sub-complex, whereas groove 2 leads along Rpb11 towards Rpb8. Recently, nascent RNA could be cross-linked to Rpb7, providing apparent support for groove 1 (Ujvari and Luse, 2006). RNA beyond position -10 was present in one of the crystallographic studies of the elongation complex (Kettenberger et al., 2004b), but could not be observed in the tunnel or in the subsequent grooves, suggesting that its interactions are transient and cannot be detected in medium-resolution electron density maps.

To answer the question whether nascent RNA indeed exits through the proposed tunnel, and if it follows a defined surface path beyond the tunnel, a single-molecule fluorescence technique was employed. By measuring several distances between RNA 5' end for varying length of RNA and known positions within the polymerase structure the RNA exit path could be revealed.

7.1 Engineering fluorescent labeled active Pol II elongation complexes

The possibility to assemble complete Pol II elongation complexes from endogenous yeast 10-subunit core enzyme, recombinant Rpb4/7 subcomplex and synthetic nucleic acid scaffolds enabled the preparation of defined elongation complexes with fluorescent labels at discrete positions. In order to reliably predict label locations, nucleic acid scaffolds were used that closely resemble those in the complete Pol II elongation complex structure (Kettenberger et al., 2004b).

Different scaffolds were prepared with RNA lengths ranging from 17 to 29 nucleotides, and comprising a donor fluorophore (TMR) at their 5'-end (RNA17, RNA20, RNA23, RNA26, RNA29) (see *Methods*, Table 4 and Figure 17). An acceptor fluorophore (Alexa 647) was attached to the DNA template strand either at the position -10 or position +3. Two other acceptor positions that were used as labels were coupled to cysteine residues in the Rpb4/7 complex (Rpb7-C150 and Rpb7-C94) after other cysteines had been replaced by serines (see *Methods*, 4.1). Labeling of these two mutants resulted in ~20% labeling efficiency and is shown in figure 16. The positions of all labels within the Pol II structure are presented in Figure 28.

The complexes were active in RNA chain elongation in ensemble (Brueckner et al., 2007) and single-molecule experiments (Andrecka et al., 2008). They may therefore be viewed as stable intermediates in the early elongation process.

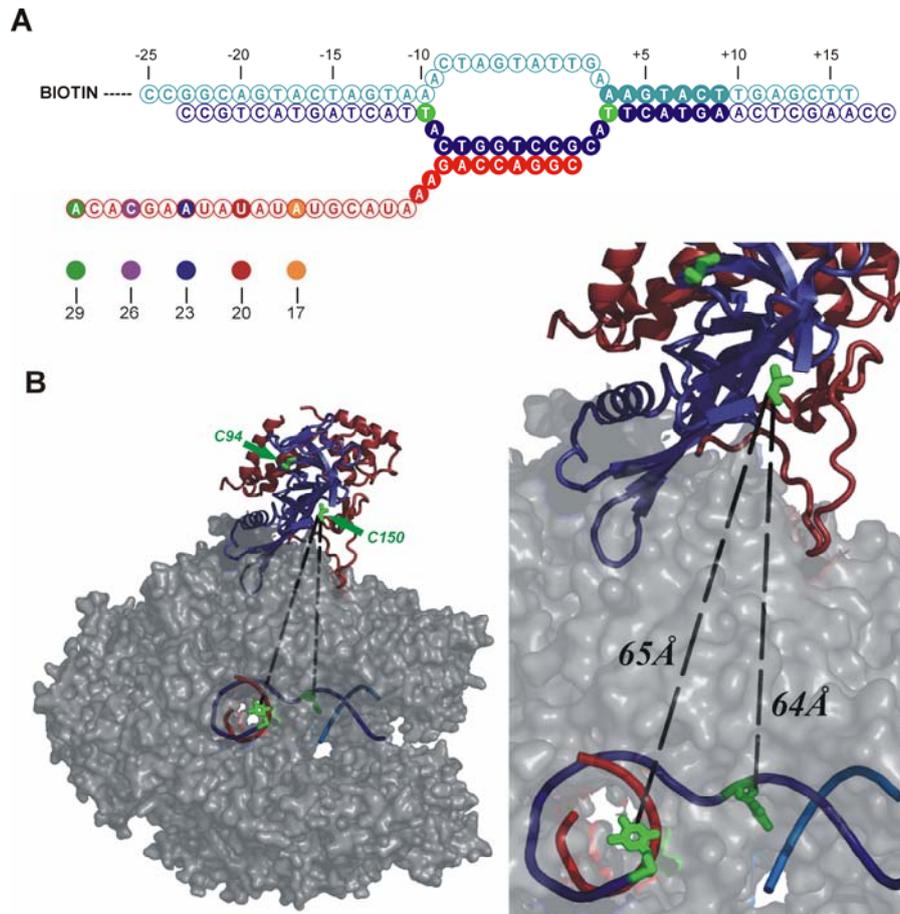


Figure 28: Pol II structure and labelling

(A) Schematic showing the oligonucleotides used in the single-molecule experiments for the formation of elongation complexes. Filled and open circles denote nucleotides whose positions are known and unknown, respectively, from crystallographic studies (Kettenberger et al., 2004b). Highlighted are the labelling positions on the template DNA and RNA. (B) Top-view of the complete Pol II elongation complex structure (Kettenberger et al., 2004b), indicating labelling sites. The core polymerase is shown in surface representation (in grey), and the backbone of the template DNA (blue), non-template DNA (cyan) and RNA (red) and the heterodimer Rpb4/7 (Rpb4 in red and Rpb7 in blue) are displayed using cartoon diagrams. The positions that were used to attach dye molecules, that is Rpb7, residues Rpb7-C94 and Rpb7-C150, as well as the template DNA at position -10 and +3, are shown in green. The right picture shows a close-up view of the labelling region highlighting the distances between Rpb7-C150 and tDNA(-10) as well as Rpb7-C150 and tDNA(+3).

7.2 Single-pair FRET results

7.2.1 FRET traces

Using a TIRF microscope (see *Methods*, 5.4 and 5.5), single-pair FRET (sp-FRET) signal from individual elongation complexes (see *Methods*) was recorded as fluorescence intensities of donor and acceptor molecules as a function of time. These donor-acceptor trajectories were used to compute the respective sp-FRET efficiencies. Using the data from hundreds molecules histograms of FRET efficiencies could be constructed. Figure 29 shows an example of sp-FRET histogram and characteristic time trajectory of donor fluorescence, acceptor fluorescence and computed sp-FRET efficiency. All histograms are shown in *Appendix A2.1*, Figure A1 and Figure A2.

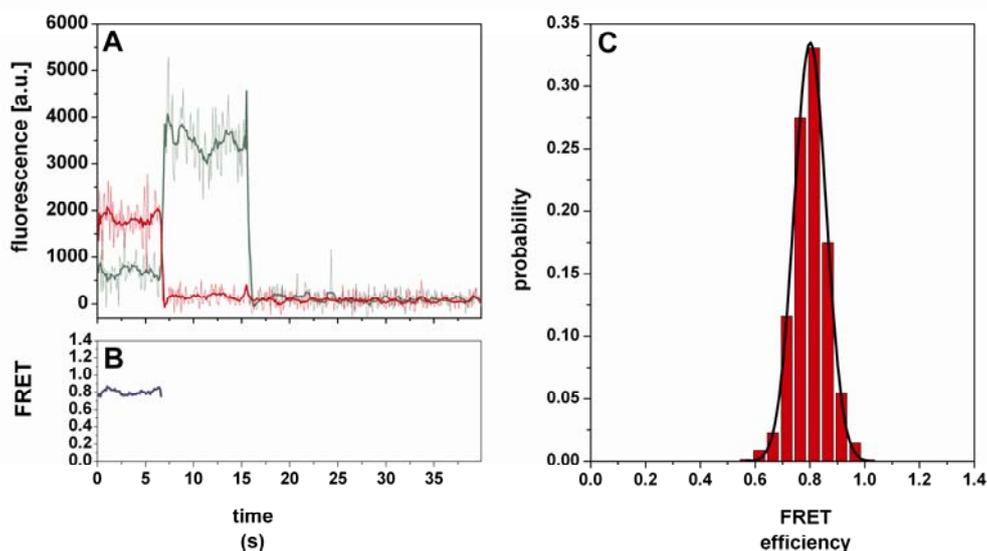


Figure 29: Single-pair FRET time trace and histogram

Exemplary time trace and complete histogram for tDNA(-10) and RNA23. (A) The time trace shows the fluorescence intensities of the donor (green) and the acceptor (red) (thin lines correspond to the actual signal and thick lines to a ten point sliding average) as well as the computed FRET signal (blue). The trajectory shows constant fluorescence intensities for the donor and acceptor molecule until bleaching of the acceptor after ~ 7 s. At this point, the donor intensity increases and remains constant, until the donor itself photo-bleaches after ~ 17 s. The computed FRET efficiency is constant (B). (C) The histogram of 356 such sp-FRET trajectories shows a single peak, which can be fitted with a gaussian distribution that is centred at a FRET efficiency of $E = 0.81$.

The trajectory of FRET efficiency E shown in Figure 29 is rather constant, which was the typical behavior observed for all elongation complexes with RNA17, RNA20 or RNA23. The histograms of hundreds of such sp-FRET trajectories show a single peak which can be fitted with a gaussian distribution that is centred at a FRET efficiency of E (see Table 7).

7.2.2 Dynamic switching

In contrast, when the RNA is extended from 23 to 26 nucleotides the FRET signal no longer remains constant. The FRET efficiency of the RNA26 elongation complexes switches repeatedly between two levels and the corresponding histograms show two peaks (Figure 30).

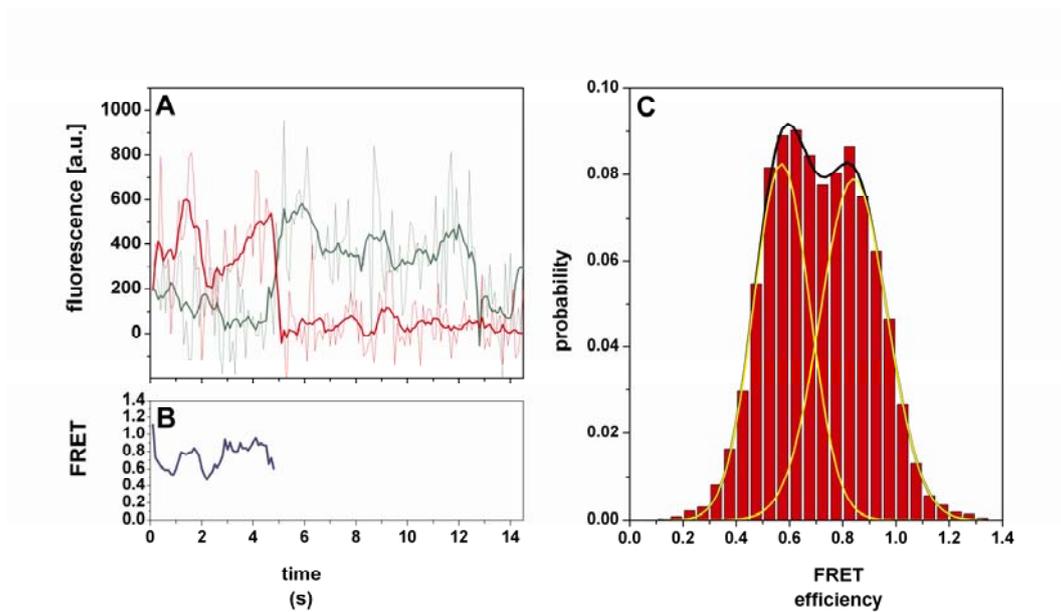


Figure 30: Single-pair FRET time trace and histogram

Exemplary time trace and complete histogram for tDNA(-10) and RNA26. (A) The time trace shows the fluorescence intensities of the donor (green) and the acceptor (red) (thin lines correspond to the actual signal and thick lines to a ten point sliding average) as well as the computed FRET signal (blue). Here, the FRET efficiency (C) is not constant but varies between $E \sim 0.85$ and $E \sim 0.6$. Accordingly, the histogram (D) of 224 such trajectories can be fitted with two gaussians centred at $E = 0.58$ and $E = 0.85$.

7.2.3 Direct acceptor excitation

Fluctuations in FRET efficiency could be caused by either distance variations or photo-physical changes in the donor and/or acceptor molecule. One possibility to check for changes in the acceptor brightness is exciting the acceptor directly in an alternating excitation scheme (Kapanidis et al., 2004; Muller et al., 2005). A typical example of a trajectory showing modulation of the FRET efficiency when being excited with green light and constant fluorescence intensity when excited with red light is shown in Figure 30. Furthermore, the total intensity of emitted fluorescence was analyzed. It was found, that the observed fluctuations are not caused by changes of donor brightness. Thus, the observed fluctuations of E cannot be attributed to photo-physical changes of the donor or acceptor molecule, instead the single-molecule probe reports on a real-time re-ordering of the growing RNA chain.

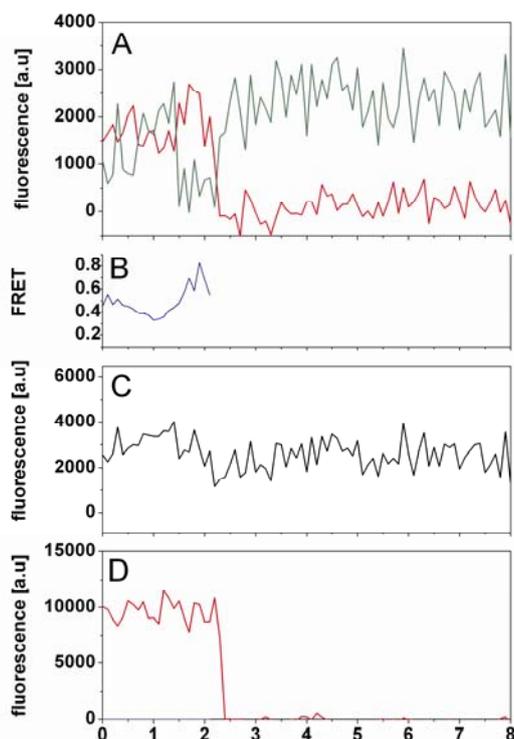


Figure 31: Alternating laser excitation

Exemplary trace showing an elongation complex with the RNA26-tDNA(-10) FRET pair. The excitation laser was alternated frame by frame between 633nm (direct acceptor excitation) and 532nm (FRET excitation, lower panel). (A) Signal from directly excited acceptor. (A) Donor and acceptor intensities while excited with the green laser. (B) FRET efficiency is not constant but varies between $E \sim 0.5-0.8$, while the total intensity (C) remains roughly constant. (D) Signal from acceptor while directly excited with the red laser.

7.2.4 From FRET efficiency to distance

For each length of RNA four independent sets of experiments were performed measuring the FRET efficiency of the complexes with the donor molecule attached to the RNA 5'-end and the acceptor attached to either tDNA(-10), or to tDNA(+3), or one of the Rpb4/7 single cysteine mutants. In order to calculate distances Förster radii were measured (with the assumption of κ^2 to be 2/3) so that the known formula [1] could be used (see *Intoduction*, 1.4). All results are shown in Table 7.

Table 7: Overview of the sp-FRET results used to determine nascent RNA position

donor (TMR)	acceptor (Alexa647)	FRET (%)	width	R_0 (Å) ¹	distance (Å)	# of molecules	comments
RNA17	tDNA(-10)	92.1	0.24	57	38	337	
RNA17	tDNA(+3)	55.8	0.19	57	55	336	
RNA17	Rpb7-C150	69.1	0.28	59	52	323	
RNA17	Rpb7-C94	68.7	0.38	60	52	91	control, 4 th distance
RNA20	tDNA(-10)	87.0	0.21	60	43	227	
RNA20	tDNA(+3)	49.6	0.19	60	60	412	
RNA20	Rpb7-C150	67.9	0.23	62	55	226	
RNA20	Rpb7-C94	63.8	0.2	63	57	76	control, 4 th distance
RNA23	tDNA(-10)	81.4	0.19	59	46	356	
RNA23	tDNA(+3)	43.3	0.14	59	62	361	
RNA23	Rpb7-C150	69.1	0.3	62	54	230	
RNA23	Rpb7-C94	67.4	0.23	62	55	105	control, 4 th distance
RNA26	tDNA(-10)	57.8/84.6	0.21/0.23	53	50/40	224	two peaks
RNA26	tDNA(+3)	42.3/58.8	0.19/0.35	53	55/50	272	two peaks
RNA26	Rpb7-C150	55.9/85.5	0.25/0.29	55	53/41	96	two peaks
RNA26	Rpb7-C94	50.0/79.3	0.28/0.2	55	55/44	92	control, 4 th distance
RNA29	tDNA(-10)	67.2	0.18	62	55	305	
RNA29	tDNA(+3)	42.7	0.18	62	65	374	
RNA29	Rpb7-C150	59.4	0.19	65	61	148	
RNA29	Rpb7-C94	51.2	0.17	66	65	96	control, 4 th distance
tDNA(-10)	Rpb7-C150	56.2	0.24	63	61	281	control, known d
tDNA(+3)	Rpb7-C150	50.6	0.36	63	63	50	control, known d

¹Accuracy +/- 2 Å

²Mean of the histogram

³Standard error of the mean

7.3 Accurate localization of the RNA 5'-end

7.3.1 Triangulation

The unknown relative position of the dye molecule was triangulated as described in the Methods part (see *Methods*, 6.1). For each length of RNA, the distances of its 5'-end to the three (tDNA(-10), tDNA(+3) and Rpb7-C150) known points in space were used. Results for each RNA length are presented below (Figure 32).

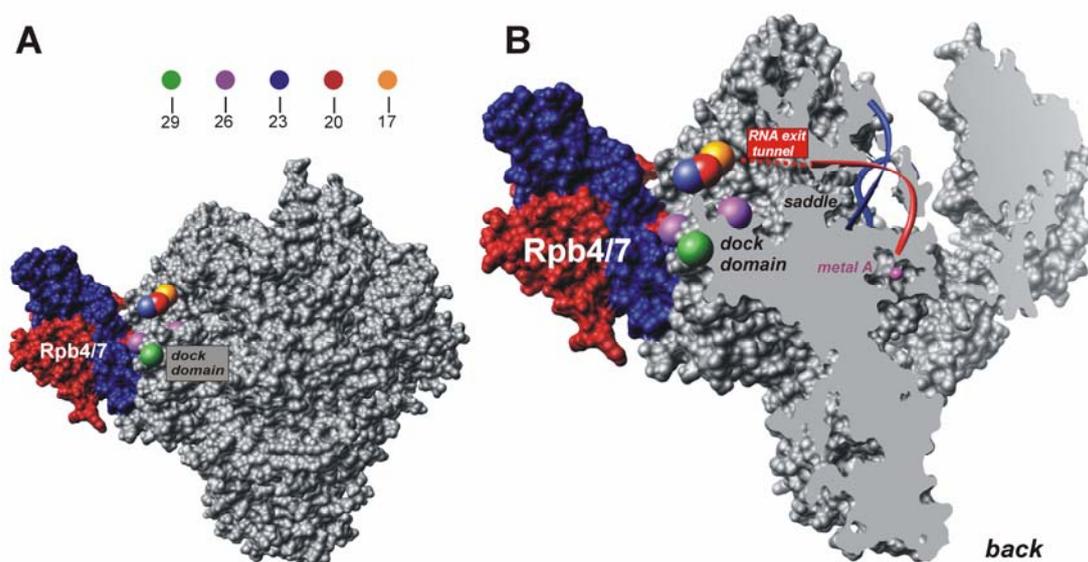


Figure 32: The position of the nascent RNA within the elongation complex

(A) Position of the dye molecule attached to the 5'-end is illustrated by a single sphere with a radius of 5 Å for RNA17 (in orange), RNA20 (in red), RNA23 (in blue), RNA29 (green) and by a pair of spheres for RNA26 (purple), indicating the presence of two states. (B) Cut-away view revealing the paths of the RNA on the interior and exterior.

The fourth measurement (between Rpb7-C94 and 5'-end of the RNA29) was crucial to solve the position of RNA26. FRET measurements of this position resulted in double gaussian FRET efficiency distribution (see *Results and Discussion*, 7.2.2) and therefore all four possible triangulations were performed. This additional measurement was used to decide which of the performed triangulation was correct.

7.3.2 Accuracy test and error estimation of the triangulated positions

The determined positions have to be taken with great caution since one has no information about the experimental accuracy. Control measurements were conducted to provide validity tests of the determined position. Determined position of each 5' end of the nascent RNA was checked by measuring the distance to yet another known point within the structure. To this end another single cysteine mutant of Rpb7, namely Rpb7-C94 (Figure 28) was used and the distance between the acceptor attached to this position and the donor on the RNA was measured. The measured distance could then be compared to the distance between the intersection-point determined by the triangulation performed (using tDNA(-10), tDNA(+3) and Rpb7-C150) and the position of Rpb7-C94 from the x-ray structure. The comparison yielded deviations of on average 2.5 Å (never more than 5 Å, see Table 8), thus providing the estimate of the experimental error. Moreover, distances between two known positions within the structure were measured, namely, distance from Rpb7-C150 to tDNA(-10) and from Rpb7-C150 to tDNA(+3). Measured distances, 61Å and 63Å respectively, compared to their separation in the structure, 65 Å and 64Å were in good agreement (see Picture 28 and Table 8).

Table 8: Accuracy of position determination

The table shows the distances between the determined position of the RNA 5'-end and the crystallographically determined position of Rpb7-C94 and the results of the FRET measurements with the acceptor attached to Rpb7-C94. The difference between these values (shown in the third column) is a good estimate of the experimental accuracy.

	Distance in structure (Å)	Measured distance (Å)	Difference (Å)
RNA 17	48	52	4
RNA 20	54	55	1
RNA 23	53	54	1
RNA 26_1	48	44	4
RNA 26_2	60	55	5
RNA 29	65	65	0
tDNA(-10)→ Rpb7-C150	65	61	4
tDNA(+3)→ Rpb7-C150	64	63	1

Furthermore, triangulation allows for a self-consistent test for the accuracy of the sp-FRET distance determination. If fourth distance was measured, as described above, one could perform four independent three-distances-trilaterations and compare the resulting four intersection points. An example of this procedure for the case of RNA29 is shown in Figure 33. The four positions that are determined by the four trilaterations are all within ~ 3 Å from each other. Thus, the dye position was determined very accurately and the Foerster distances were correct.

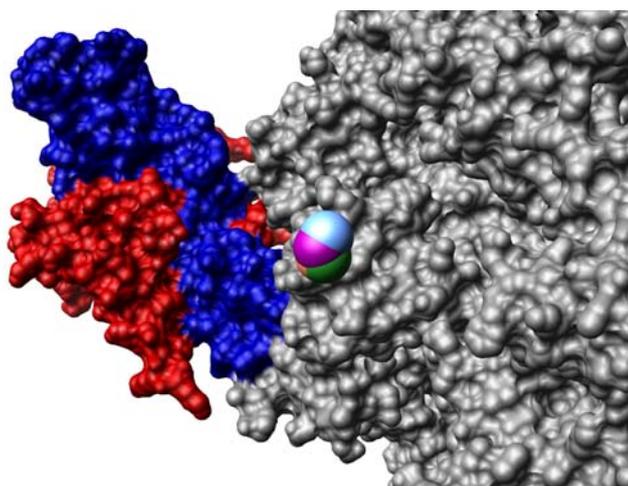


Figure 33: Multiple triangulations of the position of RNA 29

The spheres show the position of RNA29 (back view of the polymerase) determined using the distances between RNA 29 and: (i) DNA1, DNA2 and Rpb7/C150 (green sphere), (ii) DNA1, DNA2 and Rpb7/C94 (orange sphere) (iii) DNA1, Rpb7/C150 and Rpb7/C94 (magenta sphere) and (iv) DNA2, Rpb7/C150 and Rpb7/C94 (blue sphere). The center of the four spheres all lie within 3 Å, indicating that the position of the dye label attached to the RNA can be determined very accurately.

7.4 Biological implications

7.4.1 RNA traverses the exit tunnel and disengages from the surface

As it was mentioned at the beginning of this chapter in the crystal structure the last ordered RNA nucleotide is located at the entrance to a tunnel. This tunnel was called “RNA exit channel” for the bacterial RNA polymerase and was proposed to accommodate exiting RNA (Korzheva et al., 2000). In yeast similar funnel is formed between the polymerase wall, clamp and lid and leads from the active center cleft to the exterior (Cramer et al., 2001a; Gnatt et al., 2001b; Kettenberger et al., 2004b; Westover et al., 2004b).

Triangulation analysis for RNA17, RNA20 and RNA23 yields the corresponding positions of the 5'-end (Figure 32). The 5'-end of the RNA17 was located and its position can only be explained if the RNA extends from the crystallographically observed position -10 through the previously proposed exit tunnel beneath the polymerase lid. Thus, these data confirm that the RNA leaves the Pol II core through the exit tunnel as had been shown previously for the bacterial RNA polymerase (Korzheva et al., 2000; Vassylyev et al., 2007a). Moreover, the 5'-end of the RNA17 location is in good agreement with the position of the RNA17 in bacterial structure, which is the last nucleotide solved (Figure 34).

The distance between the last nucleotide visible in the crystal structure at position -10 and the observed position at -17 is ~28 Å, indicating that the RNA is in the stacked conformation within the exit tunnel. This result agrees with the bacterial elongation complex structure (Vassylyev et al., 2007a) where the RNA in the exit channel was showed in stacked conformation. Beyond the tunnel, for the lengths of 17-23 nucleotides, the RNA is compact and the distance between RNA17 -RNA23 is only ~11 Å. In this region the RNA end is too far from the polymerase surface for direct interactions.

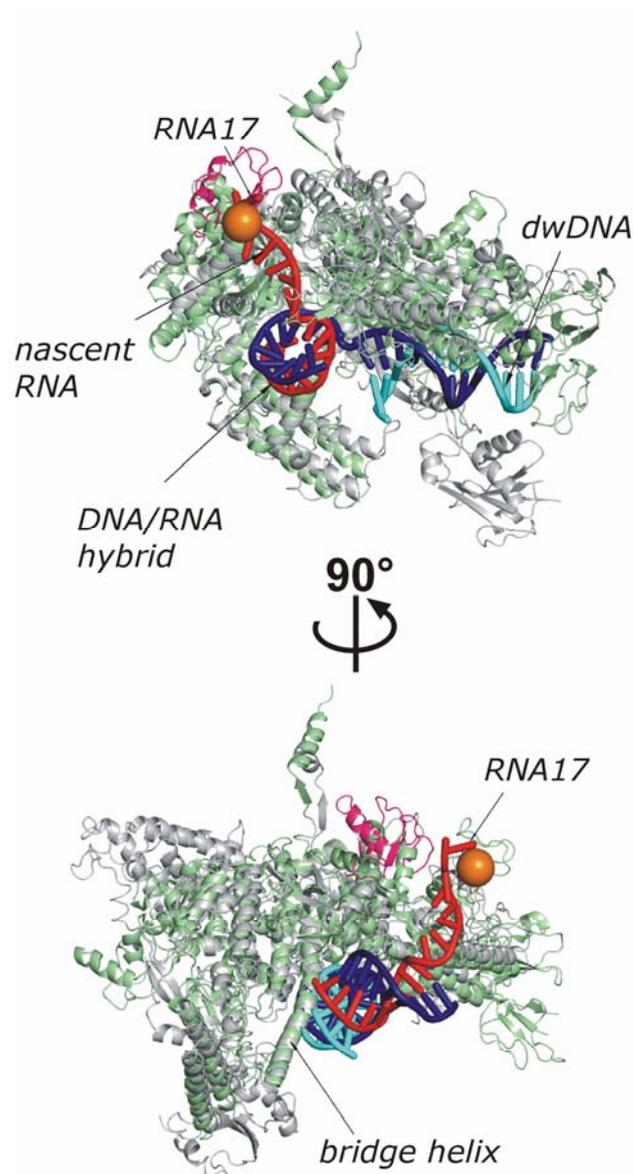


Figure 34: Comparison of the position of RNA17 in bacterial and yeast RNA polymerase

Alignment of the bacterial RNA polymerase elongation complex of *Thermus Thermophilus* (pdb: 2O5J) and yeast RNA polymerase II elongation complex (pdb: 1Y1W). Rpb1 and beta prime subunits presented as cartoons in grey and green, respectively. The template DNA (blue), non-template DNA (cyan) and RNA (red) of bacterial structure are shown using cartoons representation. Position of RNA17 in yeast polymerase determined by triangulation is shown as orange sphere of radius 5Å. Last nucleotide solved in bacterial structure corresponds to RNA17.

7.4.2 Extending RNA associates with Pol II before it becomes flexible

For elongation complexes containing RNA26 the smFRET measurements resulted in histograms that could be fitted with a double Gaussian, thus yielding two values for the distance. Therefore, trilateration does not yield a single position but two positions between which the RNA end is fluctuating (Figure 32). While for short lengths the RNA lacks contact with the surface of the polymerase, at a length of 26 nucleotides the RNA associates with the surface of the polymerase in the region of the dock domain. Furthermore, the conformation of the RNA between positions -23 and -29 is more extended, since the 5'-end of RNA26 is located 15 Å or 20 Å from the end of RNA23 and it advances another 17 Å or 21 Å between the positions -26 and -29. This observation has further implications for understanding the transition from transcription initiation to elongation. In an initially transcribing complex growing RNA sterically clashes with the initiation factor TFIIB finger domain (Bushnell et al., 2004). Beyond the tunnel RNA also clashes with the TFIIB ribbon domain (Chen and Hahn, 2003), which is located on the Pol II dock forming contacts with Rpb1 between residues 409 and 419 (Bushnell et al., 2004). Both of the observed positions of RNA26 are exactly in that area, in close proximity to the side chains of serine S409 and arginine R416, respectively. Moreover, the surface charge distribution of the polymerase in the region of the dock domain has both positive and negative patches providing several possible electrostatic interaction sites (Figure 35).

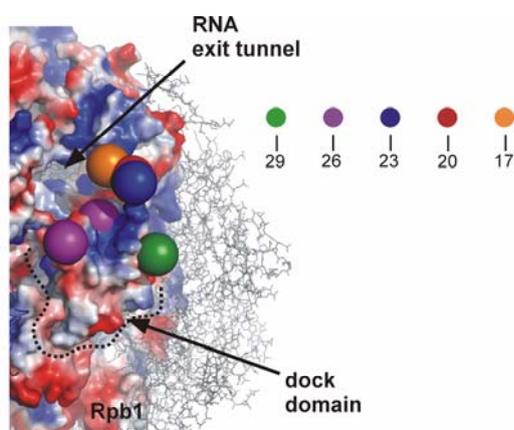


Figure 35: Surface charge distribution and RNA interaction sites

The surface of Pol II Rpb1 subunit is coloured according to the electrostatic potential. Positive, neutral and negative charges are shown in blue, white and red, respectively.

Thus growing RNA may contribute to TFIIB release during the transition from initiation to elongation, although strain that accumulates during initial DNA scrunching and subsequent bubble collapse is apparently more important (Kapanidis et al., 2006; Pal et al., 2005). Since RNA extends over the dock domain, and not via the flanking grooves, it might prevent re-association of the TFIIB ribbon with Pol II during mRNA elongation. TFIIB could also remain bound to Pol II even after this transition and then help to re-direct RNA. The question which model describes the position of the RNA29 can be answer by determining its position without and with TFIIB present (see *Results and Discussion*, 8.5).

7.4.3 Conclusion and outlook

A very general implication of this work is that the applied single-molecule triangulation technique based on multiple FRET measurements provides an accurate tool for determining the positions of flexible domains in large multi-protein complexes. In particular, single-molecule fluorescence analysis revealed structural features of a eukaryotic multi-subunit RNA polymerase that had thus far escaped detection by x-ray crystallographic means due to inherent flexibility and/or due to limited diffraction data resolution.

However, this simple triangulation method did not account for experimental errors and obtained results are, in fact, most likely positions. Errors in the determined Förster radii (due to variations in the quantum yield or orientational effects), systematic errors of the measured FRET efficiencies and uncertainties due to the attachment of dye molecule via flexible linker can strongly influence the determined position. Due to all these uncertainties it was not possible to use more than three distances since there would be no intersection of four or more spheres. To overcome these problems a novel analysis method was needed.

Moreover, there was discrepancy between recent chemical cross-linking experiments (Ujvari and Luse, 2006) and data presented here. The 5' end of RNA transcripts of a length of 26 nucleotides (RNA26) and 29 nucleotides (RNA29) were located on or in close to the dock domain of Pol II. Cross-linking studies have shown that RNA can be cross-linked to Rpb7 subunit in complexes containing the 26- to 32- nucleotide transcript. One explanation of this discrepancy would be different complexes used in both experiments. For single molecule FRET measurements purified RNA polymerase II was used. Cross-linking was performed on yeast extract which could still contain some of the transcription factors which could influence position of the growing RNA. Since RNA sterically clashes with the initiation factor TFIIB finger domain (Bushnell et al., 2004) this factor would be the first candidate to check its influence on the RNA position.

Influence of TFIIB on the position of RNA and a new data analysis method will be two issues discussed in the next chapter.

8 Position of 5'end of the 29nt RNA determined by NPS and influence of transcription factor IIB on the position of RNA

As it was shown in chapter 7 triangulation of three FRET distance measurements allows to determine a previously unknown position. However, this simple triangulation could estimate only most likely position and was not able to estimate how experimental uncertainties might influence the results. It was also not possible to use more than three measurements at once.

Here, Bayesian parameter estimation was used to compute the three dimensional probability distribution for the position of the 'antenna dye molecule' (ADM) using measured FRET efficiencies from several 'satellite dye molecules' (SDMs). The ADM is a dye molecule attached to a macromolecular complex which position is unknown. The SDMs are dye molecules attached within a macromolecular complex as a network of "satellites". Within the Bayesian framework it was possible to account for errors in the determined Förster radii, errors in the measured FRET efficiencies, uncertainties due to the attachment of the dyes via flexible linkers, and for geometric constraints. Using this new method, called Nano Positioning System (NPS), position of the 5'end of the RNA was determined and influence of IIB transcription factor on its position was studied.

8.1 Additional data required to use NPS

Since the real position of the 'satellite dye molecules' is not known (only the attachment point of SDMs to the polymerase) uncertainties due to the flexible linkers should be taken into account. To this end SDMs positions were modeled as described in *Methods*, 6.2.3 and this prior was used to compute position of ADM. Locations of SDMs are presented in Figure 36.

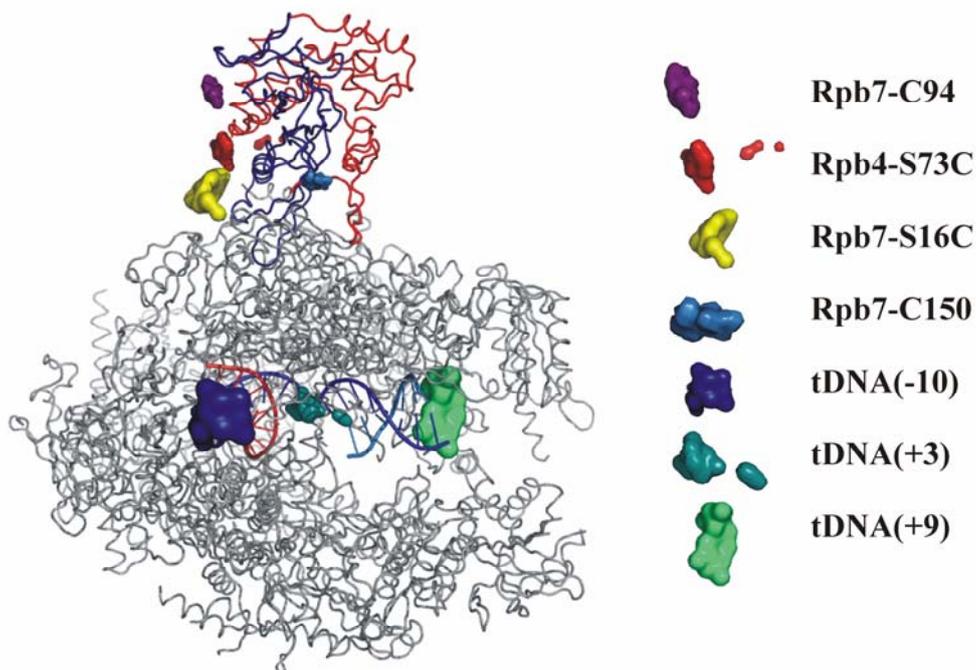


Figure 36: Locations of the “satellite dye molecules”

Top view of the elongation complex with Pol II core (grey), Rpb4 (blue), Rpb7 (red), template DNA strand (blue), non-template DNA strand (cyan) and RNA product (red). The location of the satellite dye molecule positions is indicated by small spheres. For some labelling sites two (or more) likely positions were modelled (see Methods for details). The size of the spheres corresponds to 68% credibility volumes computed using Gaussian distributions around the modelled positions with a standard deviation of 2 Å.

Moreover, within the Bayesian framework, it was possible to account for errors in the determined Förster radii (due to variations in the quantum yield or orientational effects, see *Introduction*, 6.2). To calculate errors related to determined Förster radii one needs to measure anisotropy of ADM and all SDMs in order to perform Monte Carlo simulation (all anisotropy data are presented in Table 10).

8.2 Experimental test of the NPS system

In order to test the new analysis method, the ADM (donor) was attached to a known position, the 3' end of the RNA (RNA1) and tested how well NPS can locate this position. SDMs attached to the template DNA at positions -10, +3 and +9, and to Rpb7 at positions Rpb7-C150 and Rpb7-C16 were used. All data are summarized in Table 9 and histograms are shown in *Appendix*, A2.2 Figure A3.

Table 9: Measured FRET efficiencies and Förster radii for RNA1 position calculation

SDM	isotropic Förster radius [Å]	Mean and standard deviation [Å] of the Förster radius prior	FRET efficiency [%]	# of molecules
tDNA(-10) ¹	47	45.0 ± 6.0	50.4 ± 2.0	221
tDNA(+3) ¹	47	44.7 ± 6.5	80.0 ± 2.0	322
tDNA(+9) ¹	47	44.8 ± 6.4	60.9 ± 2.0	356
Rpb7-C150 ²	57	55.6 ± 5.6	41.1 ± 2.0	68
Rpb7-C16 ²	48	46.5 ± 5.3	39.3 ± 2.0	105

1) SDM: Alexa 647, ADM: Alexa 555

2) SDM: Alexa 647, ADM: TMR

Figure 37 shows the computed ADM position probability density. It is very informative to compare the accuracy of the measurement as a function of the number of SDM molecules. By comparing equal credibility regions, one can see that the information about the ADM position increases with increasing number of satellites. Moreover, it is important to note that the relative positions of the satellites play an important role for the overall accuracy of the method (compare to Figure 38).

The RNA1 position is also a good example for illustrating how prior information can help in the localization. The surrounding protein does not leave a lot of free space for the dye molecule and therefore using prior information about protein position substantially improves the localization accuracy (Figure 37(B)). The determined probability density with and without this constrain overlaps well with the accessible volume of the dye molecule showing that NPS is indeed well suited for determining positions within macromolecular structures.

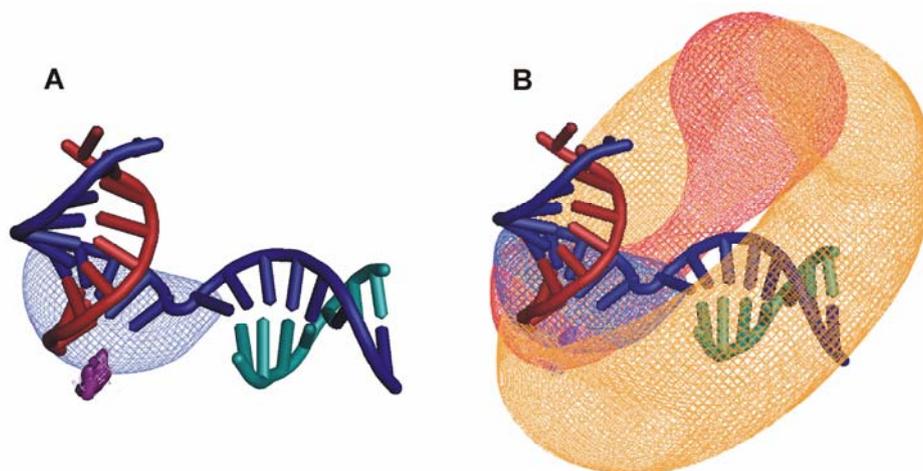


Figure 37: Experimental test of NPS

The position of an ADM attached to the 3' end of the RNA in a Polymerase II elongation complex. (A) Comparison of NPS result without (blue) and with restriction of ADM position to free space within the Pol II complex (magenta) to the computed accessibility region of dye molecule attached to 3' end of RNA (purple). (B) 68% credibility region for the ADM attached to the RNA 3' end determined using either three satellites (tDNA(-10), tDNA(+3), tDNA(+9), orange), four satellites (in addition Rpb7-S16C, red) or five satellites (in addition Rpb7-C150, blue) together with a cartoon representation of RNA (red), template DNA (dark blue) and non-template DNA (light blue).

In order to summarize the respective experimental uncertainties, the covariance matrix of the ADM position density was diagonalised to get uncertainties in the three principal directions. The uncertainty of the position determination was $\pm 2.7 \text{ \AA}$, $\pm 3.9 \text{ \AA}$ and $\pm 7.9 \text{ \AA}$ when the ADM was excluded from the Pol II interior, and $\pm 4.6 \text{ \AA}$, $\pm 5.0 \text{ \AA}$ and $\pm 9.1 \text{ \AA}$ when no constraints were used.

8.3 Positions of RNA17, RNA20, RNA23 and RNA26 revealed by NPS

Here, NPS was used to calculate positions of RNA17, RNA20, RNA23 and RNA26 which were previously determined by means of conventional triangulation (see Results and Discussion, 7.3.1).

Previously measured distances between a donor dye attached (ADM) to the 5' end of the RNA and an acceptor dye attached to the template DNA at positions -10 and +3,

and to Rpb7 at positions Rpb7-C150 and Rpb7-C94 (SDMs) were used (Results and Discussion, Table 7). This calculation resulted in an elongated probability density of the RNA positions (Figure 38).

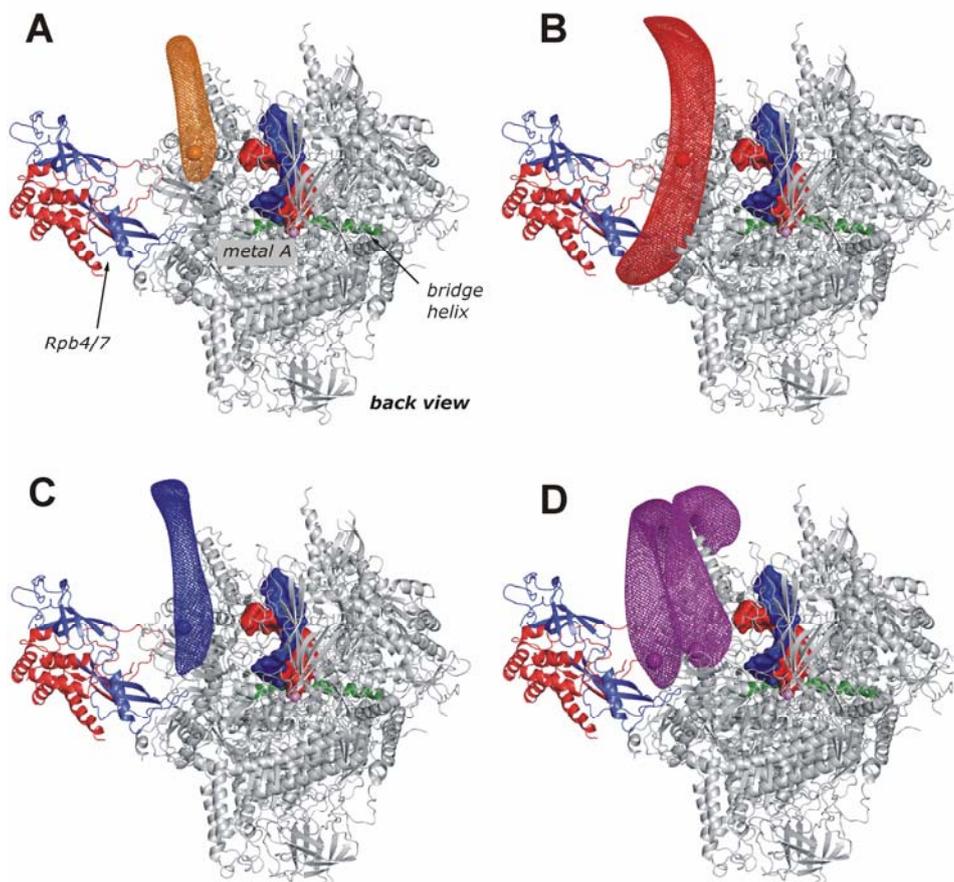


Figure 38. NPS results for RNA17, RNA20, RNA23 and RNA26

Back view of the Pol II presented as cartoons, core enzyme in grey, Rpb4/7 in red and blue. Position of the RNA determined with conventional triangulation shown as spheres – orange (RNA17), red (RNA20), blue (RNA23), purple (RNA26, two positions). 38% credibility region of the position of (A) RNA17, (B) RNA20 and (C) RNA23, (D) RNA26, respectively, determined by NPS. For these calculations information about distances to 4 satellites were used, namely tDNA(-10), tDNA(+3), Rpb7-C150, Rpb7-C94.

Those large and elongated densities are the consequences of the incomplete information gained in the experiment performed, caused by the relative orientation of SDMs and lack of information from every direction. In order to solve this problem additional data were taken (Table 10).

FRET efficiency was measured between 5'end of the RNA and Rpb4-C73 giving in total 5 satellite-antenna distances which could be used for NPS. Results are presented in Figure 38.

Table 10: Data used for NPS calculation

donor (TMR)	acceptor (Alexa647)	FRET (%)	R_o^{iso} (Å)	Anisotropy Donor	Anisotropy Acceptor
RNA17	tDNA(-10)	92.1	57	0.27	0.30
RNA17	tDNA(+3)	55.8	57	0.27	0.32
RNA17	Rpb7-C150	69.1	59	0.27	0.23
RNA17	Rpb7-C94	68.7	60	0.27	0.22
RNA17	Rpb4-C73	60.2	60	0.27	0.27
RNA20	tDNA(-10)	87.0	60	0.25	0.30
RNA20	tDNA(+3)	49.6	60	0.25	0.32
RNA20	Rpb7-C150	67.9	62	0.25	0.23
RNA20	Rpb7-C94	63.8	63	0.25	0.22
RNA20	Rpb4-C73	63.7	61	0.25	0.27
RNA23	tDNA(-10)	81.4	59	0.24	0.30
RNA23	tDNA(+3)	43.3	59	0.24	0.32
RNA23	Rpb7-C150	69.1	62	0.24	0.23
RNA23	Rpb7-C94	67.4	62	0.24	0.22
RNA23	Rpb4-C73	63.3	61	0.24	0.27
RNA26	tDNA(-10)	57.8/84.6	53	0.25	0.30
RNA26	tDNA(+3)	42.3/58.8	53	0.25	0.32
RNA26	Rpb7-C150	55.9/85.5	55	0.25	0.23
RNA26	Rpb7-C94	50.0/79.3	55	0.25	0.22
RNA26	Rpb4-C73	47.1/59.4	53	0.25	0.27

In case of RNA26 all single molecule FRET histograms resulted in a bimodal distribution (*Appendix*, A2.2 Figure A2), also the additional data taken for Rpb4-C73 in order to use NPS showed similar distribution (Table 10). Those 5 measurements give 16 possible pairs which can be used for NPS calculation. To find the best solution one can then use evidence to decide which combination was the best.

16 possible NPS calculation mentioned above have been performed. However, the evidences did not show the significant differences between the pairs (see *Appendix*, Table A3). Moreover, their positions were very similar. Two exemplary densities are shown in Figure 39.

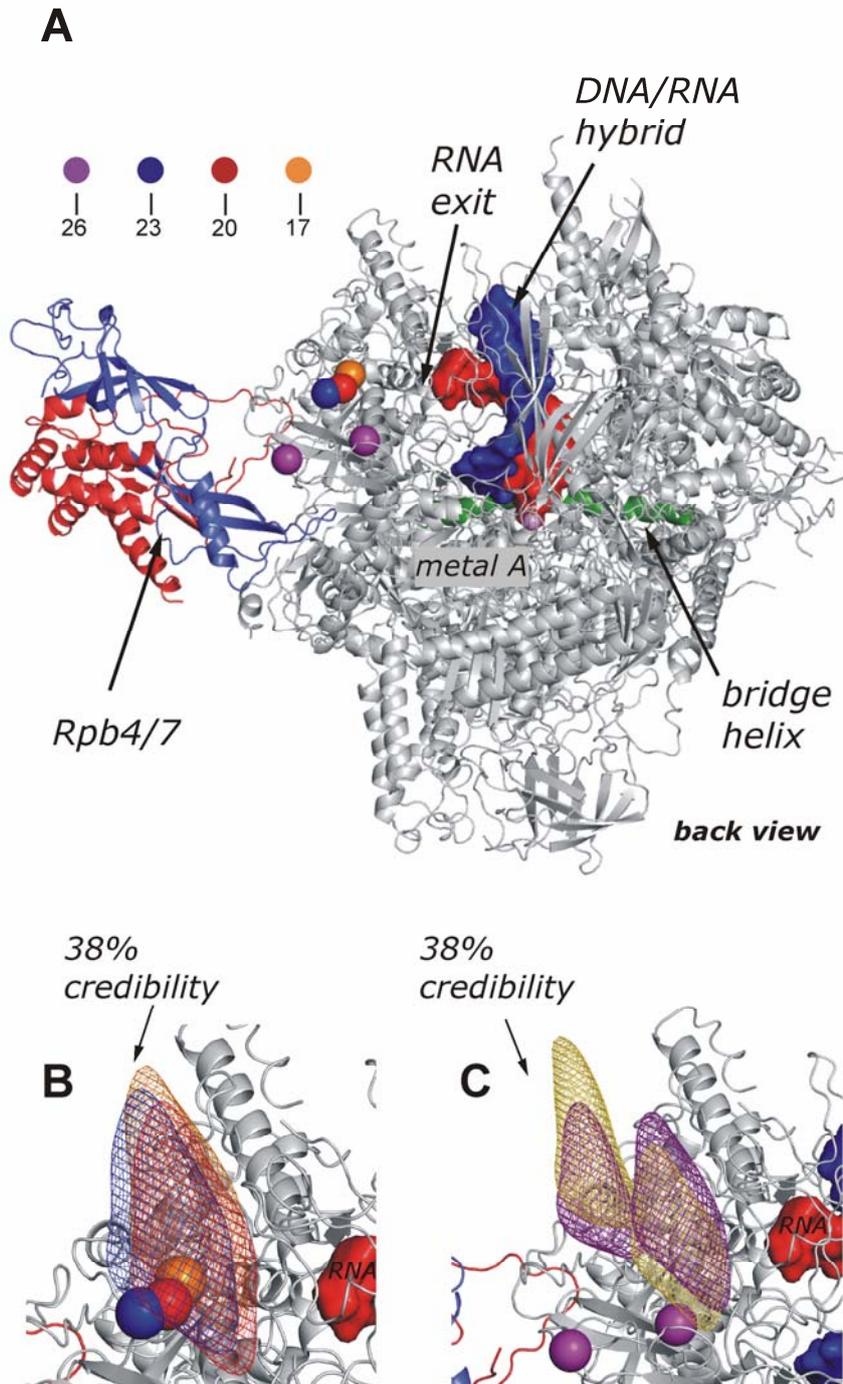


Figure 39. NPS results for RNA17, RNA20, RNA23 and RNA26 with additional information about distance to fifth satellite, Rpb4-C73

(A) back view of the Pol II presented as cartoons, core enzyme in grey, Rpb4/7 in red and blue. Position of the RNA determined with conventional triangulation shown as spheres with a radii 3Å – orange (RNA17), red (RNA20), blue (RNA23), purple (RNA26). (B, C) comparison of both, RNA positions determined with triangulation (spheres) and NPS (mesh). 38% credibility regions of the position of RNA17, RNA20, RNA23 (B) and RNA26 (C), determined by NPS are shown. In case of RNA26 two (of possible 16) combinations are shown in purple and yellow.

8.4 Position of the 5'end of 29nt RNA determined by NPS

Previous data (chapter 7) located 5'end of the RNA29 on the dock domain of the polymerase what was in contradiction to the cross-linking data (Ujvari and Luse, 2006). Here, the position of RNA29 was determined using NPS in presence and absence of TFIIB. Donor dye that acted as the antenna dye molecule (ADM) was attached to the 5'-end of a 29 nucleotide long RNA molecule. The acceptor dyes were used as satellite dye molecules (SDMs). Compared to earlier studies two additional acceptor positions were used namely, Rpb7-S16C and Rpb4-S73C giving a total of six different FRET measurements. All data are summarized in Table 11.

Table 11: Measured FRET efficiencies and Förster radii for RNA29(TMR)

SDM (Alexa647)	Isotropic Förster Radius (Å)	FRET efficiency (%) without TFIIB	# of molecules without TFIIB
tDNA(-10)	62	67 ± 2	305
tDNA(+3)	62	42 ± 2	374
Rpb7-C150	65	60 ± 2	211
Rpb7-C94	66	52 ± 2	202
Rpb7-C16	62	74 ± 2	144
Rpb4-C73	63	53/75± 2	78

Table 12: Anisotropies measured and used for Förster radii simulations

Labeling site	dye	Fluorescence anisotropy	SD
RNA1	TMR	0.21	0.00693
RNA1	Alexa555	0.31	0.01103
RNA29	TMR	0.21	0.00757
tDNA(-10)	Alexa647	0.30	0.00988
tDNA(+3)	Alexa647	0.32	0.01600
tDNA(+9)	Alexa647	0.31	0.00441
Rpb7-C150	Alexa647	0.23	0.01555
Rpb7-C94	Alexa647	0.22	0.00361
Rpb7-S73C	Alexa647	0.27	0.00751
Rpb7-S16C	Alexa647	0.20	0.00606

In order to apply NPS information about the anisotropy of the ADMs and SDMs is needed. It was used to simulate the Försted radius of every donor-acceptor pair and calculation of its error (Table 12).

NPS was applied to the measured FRET data of the position of the 5' end of a 29 nucleotide long RNA in a Pol II elongation complex. Figure 40 shows the computed credibility regions. ADM attached to the 5' end of RNA 29 was located just to the side of the dock domain.

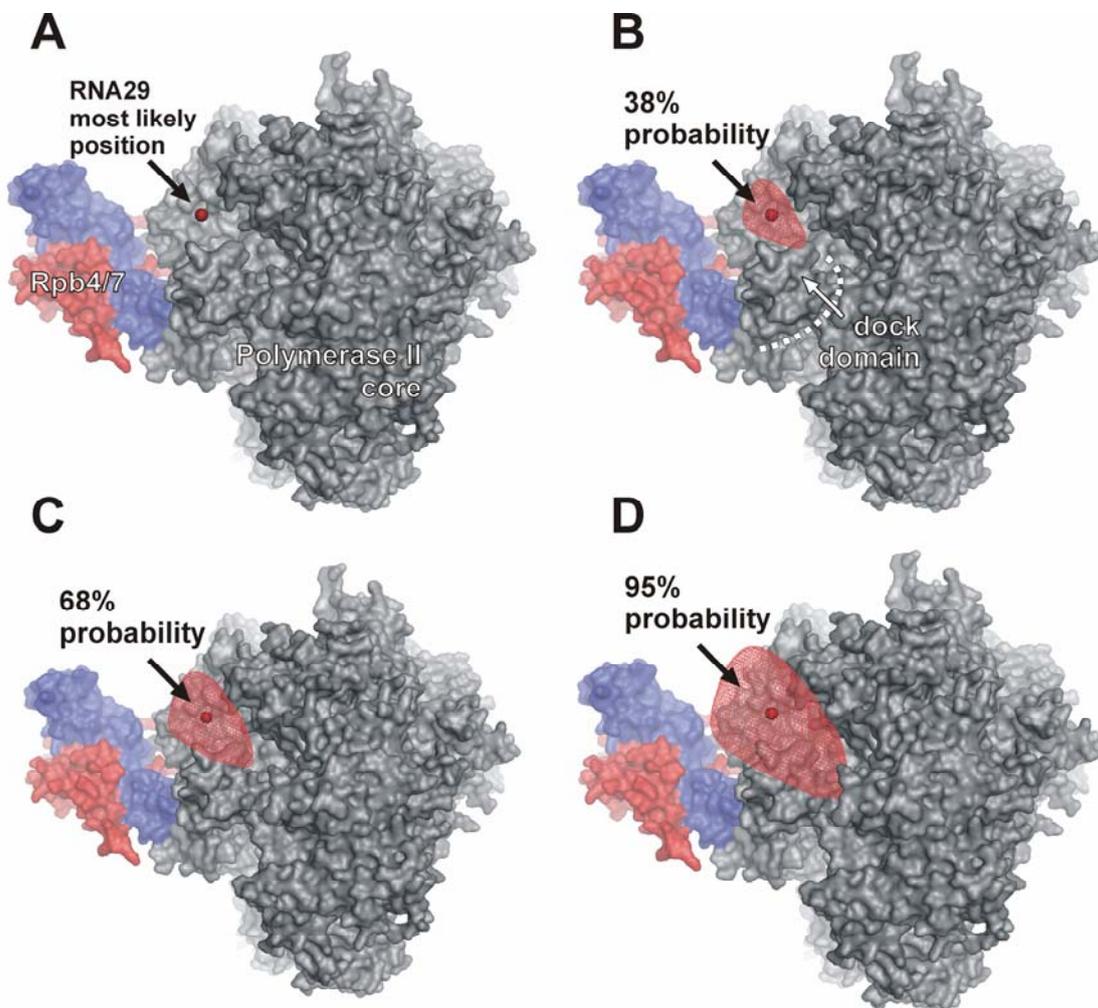


Figure 40: Computed position of the RNA29 using NPS

Back view of a Pol II elongation complex (Kettenberger et al., 2004b) (1Y1W) together with estimation of the RNA29 position shown as a meshed red surface. The surfaces enclose (B) 38%, (C) 68% and (D) 95% probability. The most likely position is represented by small red sphere.

8.5 Influence of transcription factor IIB on the position of the RNA

In order to study the effect of TFIIIB on the position of the RNA, TFIIIB was added to the sample chamber. In the presence of a 3 fold, 10 fold and 100 fold excess of TFIIIB a second peak in the computed FRET histograms was observed (Figure 41).

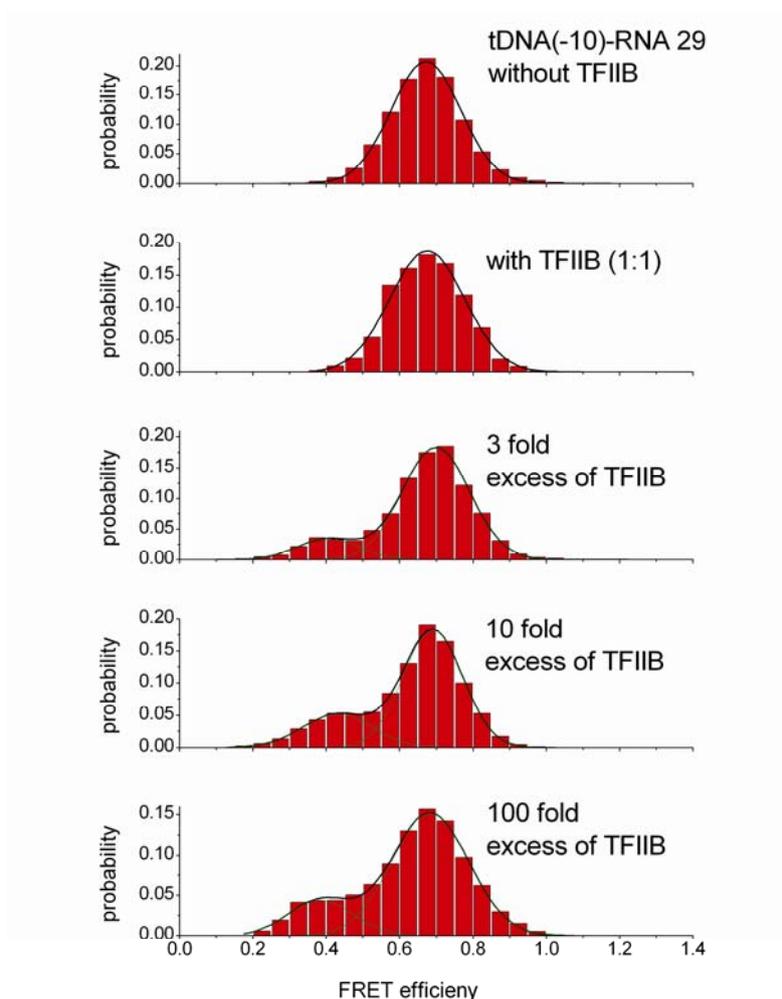


Figure 41: Effect of TFIIIB on RNA position

Histograms of tDNA(-10)-RNA 29 without TFIIIB and in presence of TFIIIB as well as in a 3 fold, 10fold or 100 fold excess of TFIIIB. The addition of TFIIIB leads to the appearance of a side peak at lower FRET efficiencies which increases with increasing concentration of TFIIIB. However, even at a 100 fold excess of TFIIIB only ~20% of the observed complexes have a lower FRET value; the initial conformation is still dominant.

Control experiments were performed with pre-incubated Pol II with varying amount of TFIIIB, prior to the addition of nucleic acids and Rpb4/7. Since growing RNA was proposed to clash with the initiation factor TFIIIB finger domain (Bushnell et al., 2004) additional experiments were performed with a TFIIIB construct in which the B-finger had been deleted. In both cases, a fraction of the complexes with altered FRET values leading to similar double-peaked histograms was observed.

It is important to note that dynamic switching between the two FRET levels in the single-molecule traces was not observed; the single-molecule FRET efficiency could always be associated with one of the two peaks in the histogram. Therefore, TFIIIB leads to a different but stable conformation of the RNA in a fraction of the complexes.

Histograms of measured distances between the 5' donor labeled end of the nascent RNA and an acceptor dye attached to the template DNA and to Rpb7 all showed second peak after TFIIIB added. All data regarding altered FRET values are presented in Table 13.

Table 13: FRET efficiencies and Förster radii used for RNA position determination after TFIIIB added

SDM (Alexa647)	Isotropic Förster Radius (Å)	FRET efficiency (%) with TFIIIB	# of molecules with TFIIIB
tDNA(-10)	62	38 ± 2	392
tDNA(+3)	62	63 ± 3	243
Rpb7-C150	65	77 ± 2	191
Rpb7-C94	66	71 ± 2	222
Rpb7-C16	62	94 ± 2	129

Information from those FRET sub-populations was used to find the new position of the nascent RNA. This new position differs from the originally obtained 5'end of RNA29 location (Figure 42). It was found in proximity of the Rpb4/7 heterodimer.

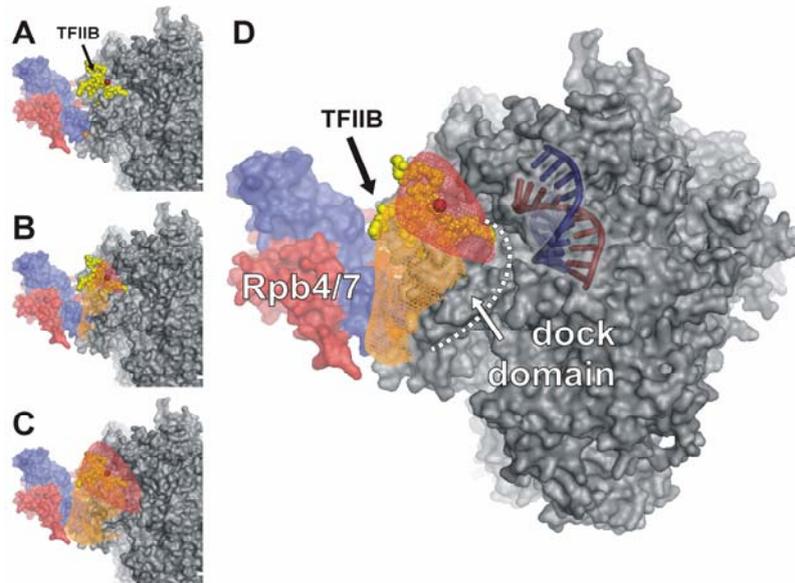


Figure 42: Position of the ADM attached to 5' end of 29 nucleotide long RNA

Back view of a Pol II elongation complex (PDB: 1Y1W) together with estimation of the ADM position displayed as a meshed surface. The ADM was attached to the 5' end of a 29nt long RNA. (B) A zoom into the region around the dock domain is shown. In addition, TFIIB position according to Pol II core structure is shown (Bushnell et al., 2004) (PDB: 1R5U) and displayed as yellow spheres. Meshes illustrating the experimental uncertainty are shown for the ADM in the absence (in red) and presence of TFIIB (in orange). Displayed are the most likely positions (A), as well as credibility regions contoured at 38% (B), 95% (C) and 68% (D) probability.

The new position of the RNA29 found in presence of TFIIB is consistent with UV cross-linking of emerging RNA (Ujvari and Luse, 2006). RNA transcript with a length between 26 and 32nt was cross-linked to the Rpb7 subunit. It indicates that TFIIB could have been 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 triggering displacement of the B-finger from the active site cleft during the transition from initiation to elongation and promoter escape (Bushnell et al., 2004). Recent experiments indicated that this displacement was triggered by the collapse of the DNA bubble that is initially formed upstream (Pal et al., 2005). In the light of the data presented here TFIIB could remain bound to Pol II even after this transition. TFIIB might then help to direct RNA towards the Rpb4/7 sub-complex. This hypothesis is supported by recent data showing that Rpb4/7 remains bound to Pol II during elongation and is involved in recruiting 3' processing factors (Runner et al., 2008).

8.6 Conclusion and outlook

In this studies the new technique, namely, Nano Positioning System was used for the first time. The power of this method is its ability to account for experimental uncertainties and therefore giving more reliably results. It also allows for including in the calculations all existing information about the system (prior knowledge).

It was shown that the position of the flexible domain (here RNA29) within the macromolecular complex (here RNA polymerase II) can be determined using NPS. Moreover, difference in the position could be detected when TFIIB was present. In the future, influence of other transcription factors on the polymerase structure could be investigated using this approach.

Single molecule fluorescence techniques can not compete with X-ray crystallography. The resolution gives rather rough position of the domain studied. However, by increasing the number of satellites it should be possible to obtain more detailed picture of the studied system. In the next chapter NPS was used to study course of the non-template and up-stream DNA with much better resolution which finally allowed for modeling of the DNA position.

9 Nano Positioning System reveals the course of upstream and nontemplate DNA within Pol II elongation complex

Despite of current progress in structural studies of the Pol II elongation complex a complete picture did not arrive, since the course of the upstream DNA and the nontemplate strand in the transcription bubble remained unknown. The upstream DNA duplex contributes to EC stability, as shown for a bacterial EC (Sidorenkov et al., 1998). The nontemplate DNA strand is also required for EC stability, and for maintaining the upstream edge of the bubble (Kireeva et al., 2000a; Kireeva et al., 2000b). It maintains the lateral stability of the polymerase by reducing the probability of backtracking (Wang and Landick, 1997) and influences polymerase pausing and arrest (Nudler et al., 1994). Furthermore, several transcription factors apparently interact with the nontemplate strand in the bubble, including the elongation factor RfaH (Sevostyanova et al., 2008), and initiation factor TFIIE (Chen et al., 2007). Thus, understanding of nontemplate and upstream DNA position is important.

9.1 Single molecule data used to determine the nontemplate DNA position

Here, NPS was used to determine the position of upstream DNA and the nontemplate strand in the Pol II EC. FRET efficiencies were measured between 'antenna dye molecules' (ADMs) attached to the nontemplate DNA strand and several 'satellite dye molecules' (SDMs) attached to positions within the EC that are known from crystallographic studies.

To determine those positions, a fluorescence donor (Alexa 555 or TMR) that acted as the ADM to the nontemplate DNA strand at positions +1, -2, -4, -7, -12, -15, or -18 were attached (see *Methods*, Table 4). The acceptor dye molecule (Alexa 647) was used like a 'satellite' at a known position and was either attached to the template DNA (at positions -10, +3 or +9) or RNA (at -1, -4 or -10) or at one of two positions on the heterodimer Rpb4/7 (Rpb7-C150 and Rpb4-C73) (Table 14).

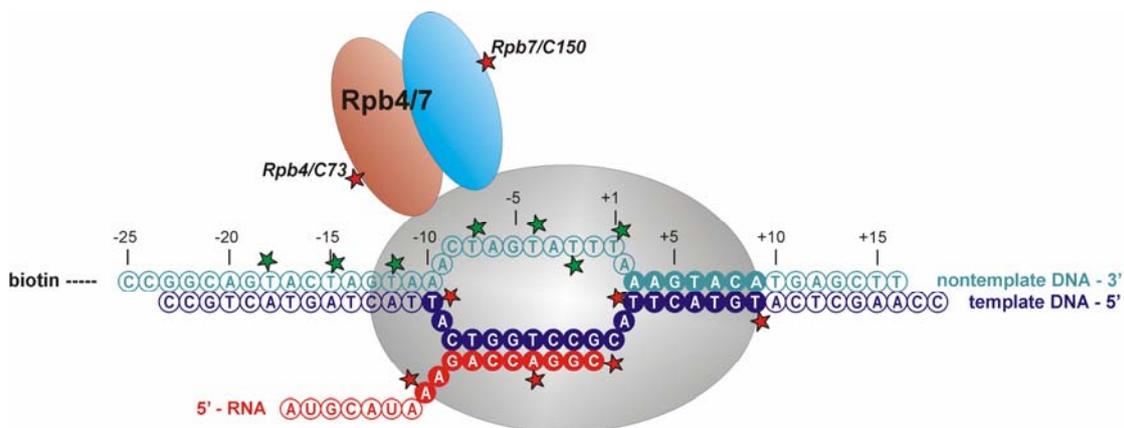


Figure 43: Experimental Design

Labelling positions. Elongation complexes were formed using mismatched nucleic acid scaffolds. The template DNA strand, nontemplate DNA strand, and product RNA are colored blue, cyan, and red, respectively. Positions of attached dye molecules are indicated by green stars for donor and red stars for acceptor dye molecules. Pol II core (grey) and Rpb4/7 (red/blue) are shown schematically. A fluorescent dye molecule (Alexa 555 or TMR) that acted as the fluorescence donor in the smFRET experiments was attached to position +1, -2, -4, -7 -12, -15 or -18 of the nontemplate DNA. Since these are the positions that are to be determined the molecule is called the ‘antenna’. The acceptor dye molecule (Alexa 647) was used like a ‘satellite’ at a known position and was either attached to the template DNA (at positions -10, +3 or +9) or RNA (at -1, -4 or -10) or at one of two positions on the heterodimer Rpb4/7 (Rpb7-C150 and Rpb4-C73).

Table 14: Antenna-Satellite system

Overview of satellite positions. For each antenna position only those satellites whose distance to the antenna was expected to fall in the sensitive range for FRET measurements were used. Moreover, depending on the expected distance either Alexa 555 or TMR were used as fluorescence donor.

Antenna (DONOR)	Satellites (ACCEPTOR = alexa 647)
NT-DNA(+1)	T-DNA(-10)*, T-DNA(+9)*, RNA1*, RNA4*, RNA10*, Rpb7/C150**
NT-DNA(-2)	T-DNA(-10)*, T-DNA(+3)*, T-DNA(+9)*, RNA1*, RNA4*, RNA10*, Rpb7/C150**
NT-DNA(-4)	T-DNA(-10)*, T-DNA(+3)*, T-DNA(+9)*, RNA1*, RNA4*, RNA10*, Rpb7/C150**
NT-DNA(-7)	T-DNA(-10)*, T-DNA(+3)*, T-DNA(+9)*, RNA1*, RNA4*, RNA10*, Rpb7/C150**
NT-DNA(-12)	T-DNA(+3)***, T-DNA(+9)*, RNA1*, RNA4*, Rpb7/C150**, Rpb4/C73**
NT-DNA(-15)	T-DNA(-10)*, T-DNA(+3)***, T-DNA(+9)*, RNA1*, RNA4*, Rpb7/C150**, Rpb4/C73**
NT-DNA(-18)	T-DNA(-10)*, T-DNA(+3)**, T-DNA(+9)**, RNA1*, RNA4*, RNA10*, Rpb7/C150**, Rpb4/C73**

* alexa 555 used as donor

** TMR used as donor

For each SDM-ADM pair, hundreds of smFRET time-traces were recorded. The smFRET data was then entered into histograms which were fitted using Gaussians to extract the mean FRET efficiency between ADM and SDM. For each pair the isotropic Förster distance R_0^{iso} was determined experimentally (Table 14). Histograms and exemplary time traces for one of the positions are presented in Figure 44.

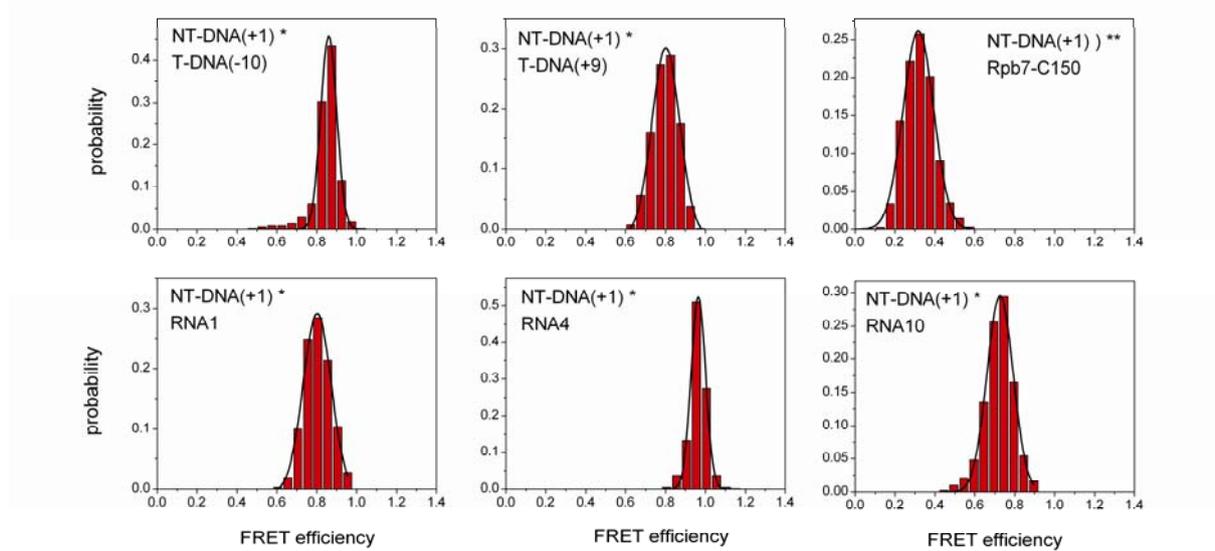


Figure 44: Complete histograms of the FRET efficiency between NT-DNA(+1) and all SDMs used to determine this position

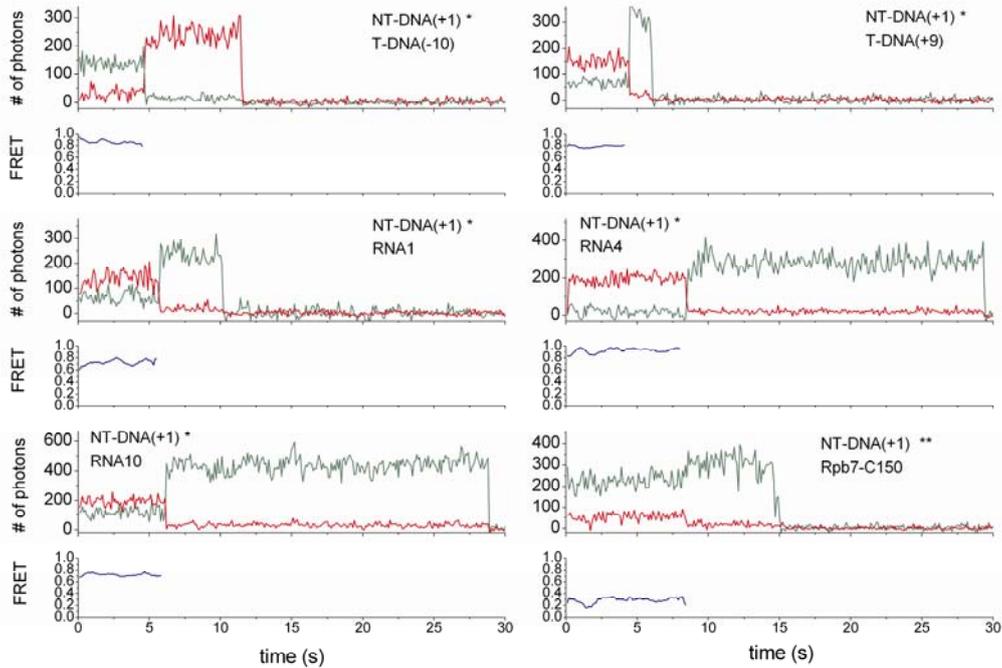


Figure 45: Examples of the time traces of the donor and acceptor fluorescence intensities and FRET efficiency between NT-DNA(+1) and each SDMs used to determine this position

Results and Discussion

Table 15: Overview of the results of the sp-FRET measurements used to determine position of nontemplate and up-stream DNA

Antenna (Donor)	Satellite (Acceptor)	Anisotropy Donor	Anisotropy Acceptor	R_0^{iso} (Å)	#	FRET efficiency	SE
NT-DNA(+1) *	T-DNA(-10)	0.29	0.30	54	255	86	0.00136
NT-DNA(+1) *	T-DNA(+9)	0.29	0.31	54	194	80	0.00211
NT-DNA(+1) *	RNA1	0.29	0.15	54	108	80	0.00476
NT-DNA(+1) *	RNA4	0.29	0.15	54	204	96	0.00141
NT-DNA(+1) *	RNA10	0.29	0.20	53	247	73	0.00188
NT-DNA(+1) **	Rpb7/C150	0.19	0.23	64	110	32	0.00223
NT-DNA(-2) *	T-DNA(-10)	0.31	0.30	50	465	84	0.00208
NT-DNA(-2) *	T-DNA(+3)	0.31	0.32	50	442	91	0.00202
NT-DNA(-2) *	T-DNA(+9)	0.31	0.31	50	244	75	0.00421
NT-DNA(-2) *	RNA1	0.31	0.15	50	183	64	0.00166
NT-DNA(-2) *	RNA4	0.31	0.15	50	393	90	0.0014
NT-DNA(-2) *	RNA10	0.31	0.20	49	330	76	0.00163
NT-DNA(-2) **	Rpb7/C150	0.19	0.23	62	94	35	0.0034
NT-DNA(-4) *	T-DNA(-10)	0.30	0.30	49	330	78	0.00135
NT-DNA(-4) *	T-DNA(+3)	0.30	0.32	49	281	88	0.00224
NT-DNA(-4) *	T-DNA(+9)	0.30	0.31	49	818	61	0.00179
NT-DNA(-4) *	RNA1	0.30	0.15	49	237	46	0.00166
NT-DNA(-4) *	RNA4	0.30	0.15	49	342	81	0.00117
NT-DNA(-4) *	RNA10	0.30	0.20	48	626	57	0.00034
NT-DNA(-4) **	Rpb7/C150	0.21	0.23	64	104	36	0.00508
NT-DNA(-7) *	T-DNA(-10)	0.29	0.30	46	356	80	0.00116
NT-DNA(-7) *	T-DNA(+3)	0.29	0.32	46	522	83	0.00439
NT-DNA(-7) *	T-DNA(+9)	0.29	0.31	46	463	57	0.00305
NT-DNA(-7) *	RNA1	0.29	0.15	46	199	40	0.00214
NT-DNA(-7) *	RNA4	0.29	0.15	46	477	78	0.00159
NT-DNA(-7) *	RNA10	0.29	0.20	45	461	60	0.00275
NT-DNA(-7) **	Rpb7/C150	0.18	0.23	64	173	54	0.00277
NT-DNA(-12) *	T-DNA(+3)	0.27	0.32	49	89	53	0.00207
NT-DNA(-12) **	T-DNA(+3)	0.18	0.32	59	592	90	0.00456
NT-DNA(-12) *	T-DNA(+9)	0.27	0.31	49	257	46	0.00432
NT-DNA(-12) *	RNA1	0.27	0.15	49	142	45	0.00419
NT-DNA(-12) *	RNA4	0.27	0.15	49	295	86	0.0089
NT-DNA(-12) **	Rpb4/C73	0.18	0.27	62	132	32	0.00442
NT-DNA(-12) **	Rpb7/C150	0.18	0.23	64	139	57	0.00247
NT-DNA(-15) *	T-DNA(-10)	0.31	0.30	49	175	89	0.00094
NT-DNA(-15) *	T-DNA(+3)	0.31	0.32	49	108	38	0.00612
NT-DNA(-15) **	T-DNA(+3)	0.16	0.32	60	107	68	0.00487
NT-DNA(-15) *	T-DNA(+9)	0.31	0.31	49	234	30	0.0045
NT-DNA(-15) *	RNA1	0.31	0.15	49	290	33	0.00499
NT-DNA(-15) *	RNA4	0.31	0.15	49	412	63	0.00458
NT-DNA(-15) **	Rpb4/C73	0.16	0.27	63	73	29	0.00613
NT-DNA(-15) **	Rpb7/C150	0.16	0.23	64	129	48	0.00202
NT-DNA(-18) *	T-DNA(-10)	0.27	0.30	48	1315	63	0.0056
NT-DNA(-18) **	T-DNA(+3)	0.17	0.32	58	378	38/57 [§]	0.012/0.023 [§]
NT-DNA(-18) **	T-DNA(+9)	0.17	0.31	58	471	38/71 [§]	0.0035/0.0137 [§]
NT-DNA(-18) *	RNA1	0.27	0.15	48	325	27	0.00377
NT-DNA(-18) *	RNA4	0.27	0.15	48	99	49	0.00632
NT-DNA(-18) *	RNA10	0.27	0.20	48	177	68	0.0042
NT-DNA(-18) **	Rpb4/C73	0.17	0.27	63	50	22	0.01133
NT-DNA(-18) **	Rpb7/C150	0.17	0.23	62	94	35	0.00514

* alexa 555 used as donor

** TMR used as donor

§ second peak

9.2 Positions of the ADMs revealed using NPS

The three-dimensional probability densities for the position of the ADM were computed using NPS. Förster radii, SDM and ADM priors, as well as likelihood were used as described before (see *Methods*, 6.2). The densities were displayed as credibility volumes (Figure 46). The close spacing of the ADM positions defined the overall course of the DNA.

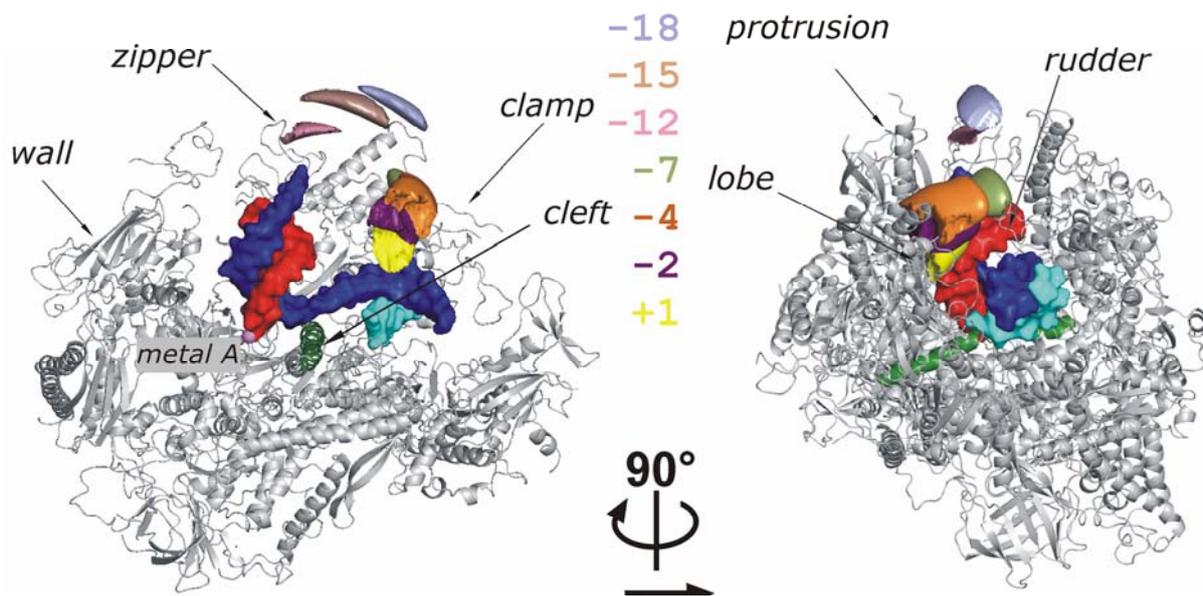


Figure 46: Position of nontemplate and upstream DNA in Pol II elongation complexes

NPS results calculated from measured FRET values. 38% credibility volumes obtained for nontemplate DNA at register +1 (in yellow), -2 (in purple), -4 (in orange), -7 (in green), -12 (in pink), -15 (in dark salmon) and -18 (light blue) are displayed using surface representation. Left panel: Side view of Pol II core enzyme shown in cartoon representation (grey), Rpb2 was omitted for clarity to reveal the nucleic acids. Right panel: Alternate view of the polymerase (rotation by 90° as indicated). Here, all 10 polypeptides of the core enzyme are shown. The previously determined parts of the template DNA (in dark blue), nontemplate DNA (in light blue) and RNA (in red) (Kettenberger et al., 2004a) are shown using surface representation.

9.3 Model of the nontemplate and upstream DNA

The calculated ADM probability density functions were used to construct a model of the nontemplate strand from register +2 to -23 and of the template strand from -11 to -23. This represents the single stranded region of the nontemplate strand and the upstream duplex DNA, neither of which are visible in the crystal structure of the Pol II elongation complex 1Y1W (Kettenberger et al., 2004a). The modeled DNA was positioned such that an ADM labeled base passed adjacent to the corresponding ADM position (but not entirely within it), since the ADM position describes the position of the dye molecule and not that of the actual base. Base stacking was maintained throughout the single-stranded part of the model. It was assumed that the point of rewinding occurs at register -11 and a straight B-form DNA helix was inserted at this point to represent the upstream duplex. The model was built using *coot* (Emsley and Cowtan, 2004) and energy minimized with *phenix.refine* (Afonine et al., 2005).

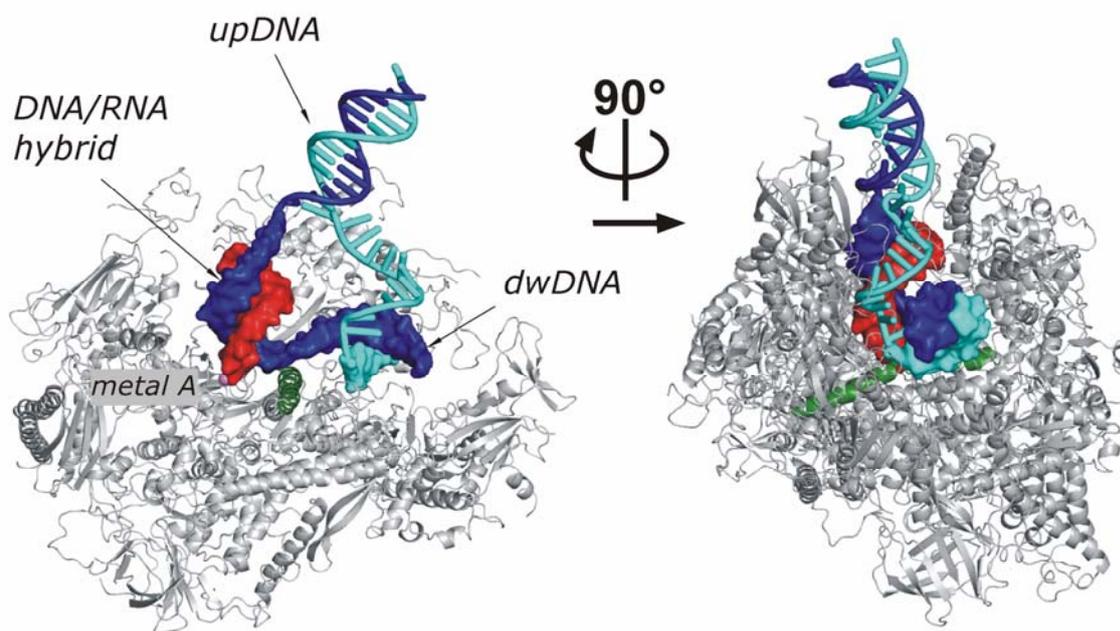


Figure 47: Model of nontemplate and upstream DNA in Pol II elongation complexes

Obtained probability densities were used to build a model of nontemplate and upstream DNA. Modeled DNA is shown in cartoon representation. Left panel: Side view of Pol II core enzyme shown in cartoon representation (grey), Rpb2 was omitted for clarity to reveal the nucleic acids. Right panel: Alternate view of the polymerase (rotation by 90° as indicated). Here, all 10 polypeptides of the core enzyme are shown. The previously determined parts of the template DNA (dark blue), nontemplate DNA (light blue) and RNA (red) (Kettenberger et al., 2004a) are shown using surface representation.

Since the densities used for modeling correspond to the positions of the dye attached to the DNA (SDMs), not to DNA itself, the most natural way to test the model was building the antenna dye position, given the model. The accessible volume for the ADMs calculated (see *Methods*, 6.2.3) from the modeled nontemplate DNA overlapped well with the experimentally determined credibility volumes for all ADM (Figure 48 shows two examples).

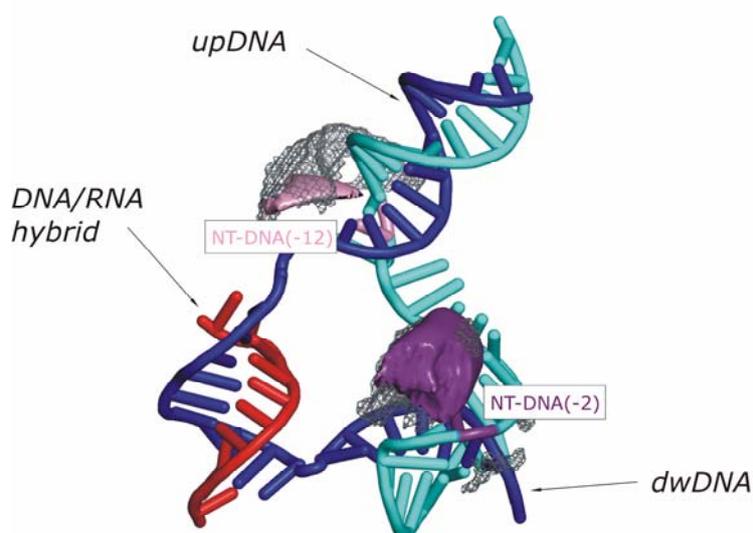


Figure 48: Consistency check of built model

38% credibility volume for the dye molecule attached to nontemplate DNA at register -2 (purple) and -12 (pink) determined using NPS (surface representation) and computed accessible volume for a dye molecule attached to the respective position of the modeled nontemplate DNA (grey meshes).

9.4 Detailed picture and discussion

9.4.1 Strand separation

Downstream DNA strand separation is essential for the polymerase translocation and nucleotide insertion. However, this process is not well understood, since to date a separated nontemplate DNA has not been observed in any of the elongation complex structures.

Kornberg and co-workers observed fraying between nontemplate and templated DNA in elongation complex structures already at position +5 (Westover et al., 2004a). An early melting of the DNA duplex would be required for elongation models, where several nucleotides line up with the template strand prior to incorporation at the active site of the enzyme (Zhang and Burton, 2004). However, in these crystallographic studies the nontemplate DNA only extended up to the +5 position and thus the observed fraying could be an artifact caused by the length of the nontemplate DNA used in these studies. In contrast, Kettenberger et al. used mismatched nucleic scaffolds, where the nontemplate DNA extended from the +15 to the -26 position (Kettenberger et al., 2004a). In this structure nontemplate DNA was observed from positions +3 to +9 and was base paired with template strand till position +3.

In the model presented here, the strands of the downstream DNA duplex are separated at register +2 at fork loop 2 (Figure 50, passing near the conserved residue R504, which is important for transcription *in vitro* (Naji et al., 2008). From there, the nontemplate strand is directed towards the lobe of Rpb2. At register -4, the nontemplate strand approaches a loop on top of the lobe domain (loop β 10- β 11, Rpb2 residues 272-278) on the Rpb2 side of the cleft. This loop is invariant between yeast and human Pol II and includes an exposed lysine residue (K277) that likely interacts with the nontemplate strand.

9.4.2 Nontemplate DNA within the bubble

After strand separation the nontemplate strand turns towards the Rpb1 side of the cleft and continues to the rudder, an extension from the clamp domain (Fig 50). The backbone of the nontemplate DNA at register -7 to -9 is close to the Rpb1 residues 309-315 in the rudder as well as to conserved Lys 471 of fork loop 1 at position -10. Nontemplate and template strands re-anneal at register -11 to form the upstream DNA duplex. Upstream DNA exits from the Pol II - cleft between the Rpb1 clamp and the Rpb2 protrusion domains. The axis of the upstream DNA duplex is parallel to the axes of helices $\alpha 8$ in the clamp and $\alpha 11$ in the protrusion.

The bases in the upstream half of the nontemplate DNA are directly exposed to solvent (positions -5 to -10). This observation explains the results of micrococcal nuclease cleavage experiments on the bacterial elongation complex which showed that DNA was degraded rapidly from the upstream end up to position -5 (Wang and Landick, 1997), while positions downstream of this site were stable. Cleavage of the exposed nontemplate DNA shifted the polymerase backward suggesting that the exposed segment of the nontemplate DNA directly or indirectly stabilizes the lateral position of the transcription complex.

The nontemplate strand in the bubble was previously modeled to penetrate the groove between the lobe and protrusion (Korzheva et al., 2000), this results show that this region passes loop $\beta 10$ - $\beta 11$ close to the four-stranded β -sheet of the lobe. Deletions in this domain cause defects in the formation of the bacterial open promoter complex (Nechaev et al., 2000; Severinov and Darst, 1997) indicating that interactions between loop $\beta 10$ - $\beta 11$ and the nontemplate strand may contribute to DNA melting or stabilization of the bubble (Figure 48).

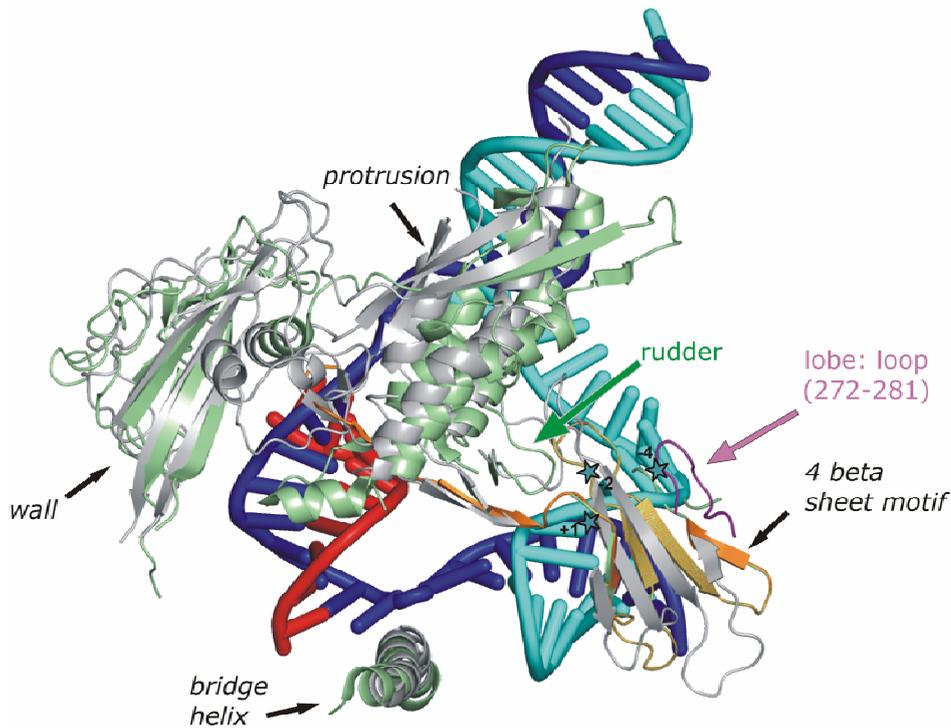


Figure 49: Comparison to crosslinking in the bacterial polymerase

Overlay of the elongation complex structure of Pol II from yeast (grey, 1Y1W) and the *Thermus thermophilus* enzyme (green, 1IW7) using side view. Chemical crosslinking between bases at positions +1, -2 and -4 (indicated by stars) were observed to residues 130- 183 of the β subunit. The crosslinking regions agree perfectly with the model for the course of the nontemplate DNA.

9.4.3 Loops network and its interaction with DNA

The model suggests that the rudder, a loop protruding from the clamp into the cleft, plays crucial roles during transcription. The DNA nontemplate strand passes above the Rpb1 residues 309-315 in the rudder, and re-anneals with the template strand to form the exiting upstream DNA duplex, suggesting that the rudder is important for proper DNA rewinding and maintenance of the upstream end of the bubble. The interaction between rudder, fork loop1 and DNA may play an important role in polymerase movement along DNA as was suggested previously by MD simulations which observed fork loop 1 movement accompanied by a movement of the nontemplate DNA in response to an artificial force (Suenaga et al., 2006).

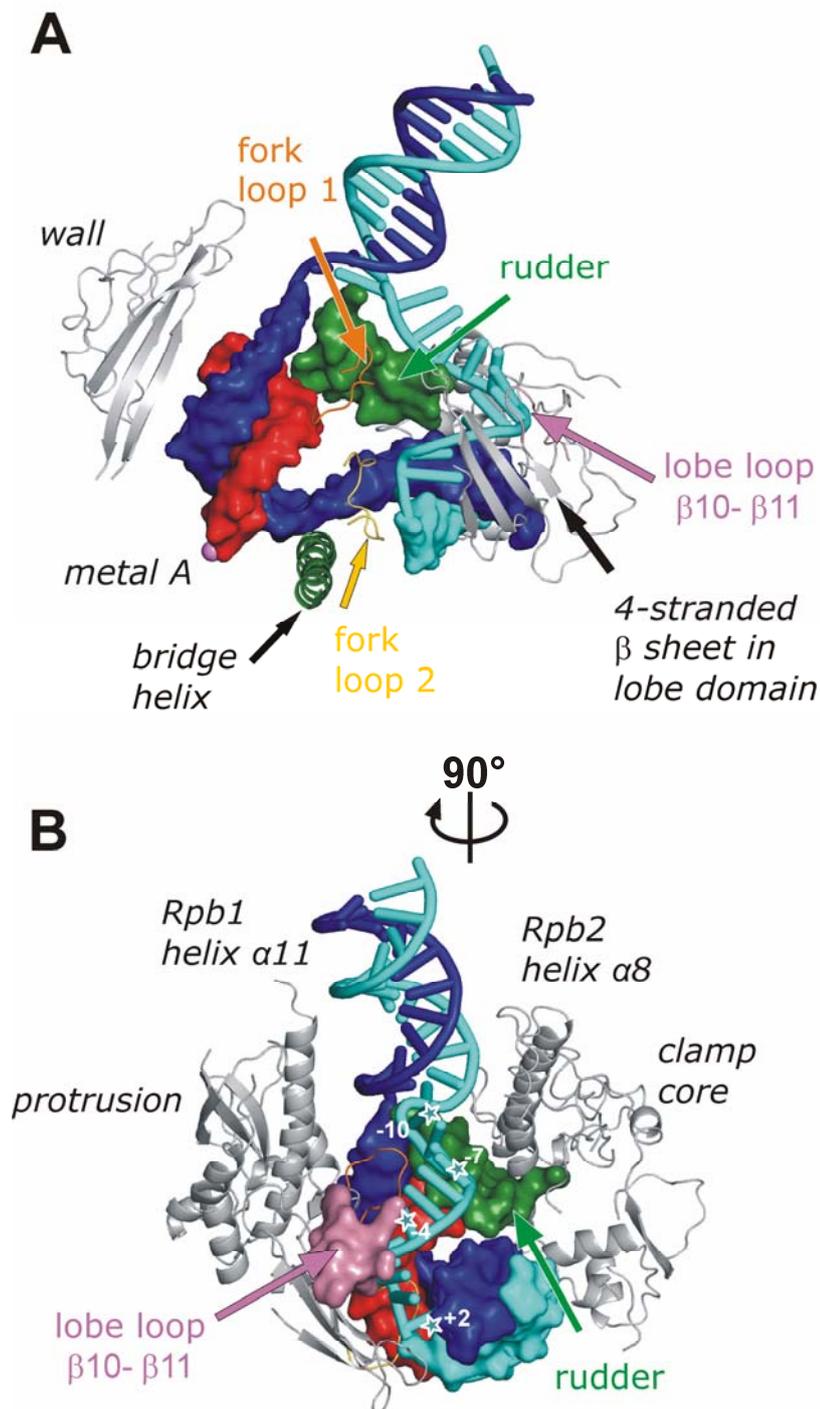


Figure 50: Interaction network

Zoom into the elongation complex showing important structural features that define the path of the nontemplate DNA. Previously known DNA and RNA positions are shown in surface representation and modeled position of nontemplate and upstream DNA using cartoons. (A) Side view, polymerase elements important for the pathway of the nontemplate DNA are highlighted: rudder (in green), fork loop1 (in orange), fork loop 2 (yellow) and lobe loop (residues 272-278, pink). (B) 90° rotated view.

The model also explains why the rudder is critical for promoter opening and EC stability *in vitro* (Kuznedelov et al., 2002; Naji et al., 2008), and for transcription elongation *in vivo* (Malagon et al., 2006). These results argue against the original proposal that two other loops extending from the clamp, the zipper and lid, separate DNA strands at the upstream bubble (Cramer et al., 2001b; Gnatt et al., 2001a). The lid instead is involved in RNA separation from DNA at the end of the hybrid (Kettenberger et al., 2004a; Westover et al., 2004b). However, a bacterial polymerase mutant that lacks the lid maintains the correct hybrid length, but only when nontemplate DNA is present (Naryshkina et al., 2006; Touloukhonov and Landick, 2006), suggesting that maintenance of the upstream end of the bubble and maintenance of the hybrid length are coupled. In addition, other transcription factors which bind to the nontemplate strand during initiation (Chen et al., 2007) and elongation (Sevostyanova et al., 2008) will contribute to the interaction network and further stabilize the upstream end of the bubble.

9.4.4 Upstream DNA in the elongation complex

As the upstream DNA exits from polymerase, its course becomes less well defined, indicating increasing mobility. At the most upstream ADM (register -18), smFRET time traces were not constant but rather showed dynamic switching between two values (see *Appendix*, A2.3 Figure A9). The resulting histograms revealed two peaks, which were fitted by two Gaussians (Figure A9).

Using NPS two different positions of the ADM were determined. The dominant position, which accounts for ~70% of the data, is shown in Figure 46 and was used for building the DNA model. The alternative position is shifted towards the protrusion by about 10 Å (Figure 51).

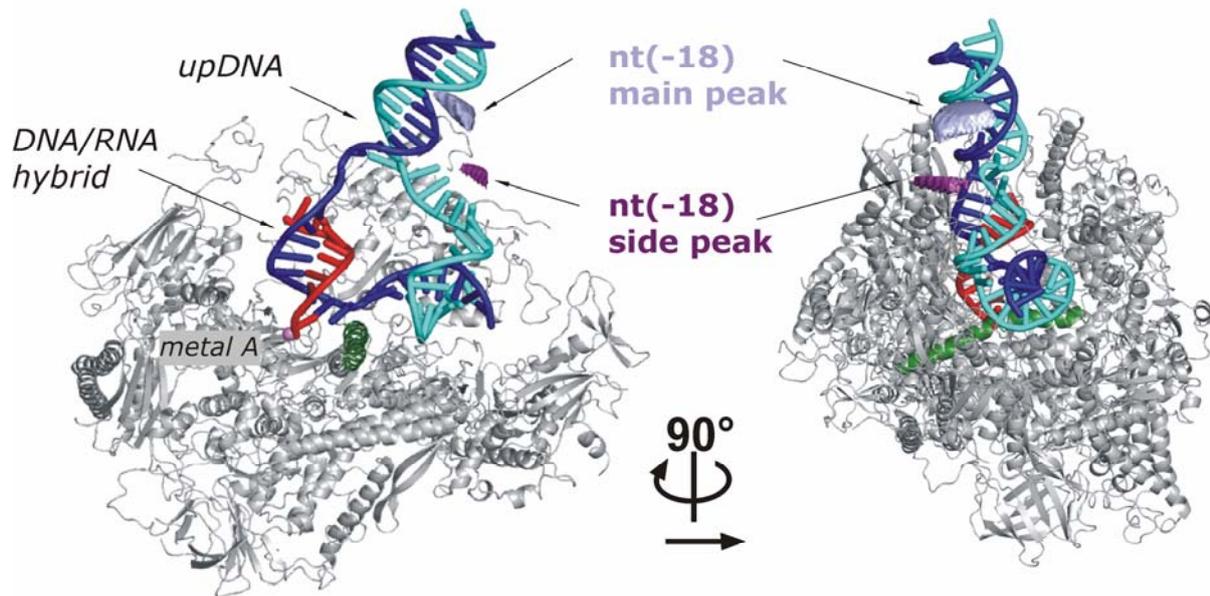


Figure 51: Two observed positions of upstream DNA

For the most upstream ADM attached to nontemplated DNA at position -18 dynamic switching between two states was observed, resulting in double peaked histograms (Appendix). Thus, NPS was used to model both positions. The most likely position which accounts for about 70% of the data was used for building the model of the upstream DNA. The position of the sidepeak in the histograms (~30% probability) was shifted by about 10Å towards the protrusion.

The angle between the downstream and upstream DNA duplexes in the model build is approximately 80°, in agreement with previous studies using atomic force microscopy (Rees et al., 1993). Moreover, the presented results suggest that the location of the upstream DNA changes dramatically during the transition from initiation to elongation. In the closed complex, upstream DNA is located to the polymerase upstream face outside the cleft (Chen and Hahn, 2004), whereas in the EC it is located above the cleft between the polymerase protrusion and clamp. This relocation of upstream DNA may occur during open complex formation, when the DNA is melted and the template strand slips into the active center, or during promoter escape, when the initial transcription bubble collapses and initiation factors are released (Ujvari and Luse, 2006). These and other dynamic structural transitions during the transcription cycle can be addressed using NPS in the future.

9.5 Conclusion

Crystallographic studies of the RNA polymerase II (Pol II) elongation complex (EC) revealed the locations of downstream DNA and the DNA-RNA hybrid, the course of the nontemplate DNA strand in the transcription bubble and the upstream DNA duplex remained the open question.

NPS was employed to solve this problem. By increasing the number of satellite dye molecules used in single molecule FRET experiments it was possible to build a complete model of the Pol II EC containing all nucleic acid regions. This studies gave insight into the separation of the nontemplate from the template strand, interaction of the bubble region with several loops within elongation complex and showed dynamic behaviour of the upstream DNA while leaving the enzyme. These and other dynamic structural transitions during the transcription cycle can be addressed using NPS in the future.

10 Outlook and future perspectives

In this work one more piece of the transcription puzzle was tried to be fit into place. The structure of the elongation complex was studied using single molecule fluorescence approach. FRET efficiencies were measured between ‘antenna dye molecules’ (ADMs) attached to unknown positions on RNA and DNA and several ‘satellite dye molecules’ (SDMs) attached to known positions within the Pol II elongation complex. The recently developed Nano Positioning System (NPS) was then used to determine antenna’s positions. Obtained results provided a framework for more detailed modeling to define the path of the nucleic acids across the elongation complex of RNA polymerase II.

General implication of this work is that the applied single-molecule trilateration technique based on multiple FRET measurements provides an accurate tool for determining the positions of flexible domains in large multi-protein complexes. This probability based analysis method combines single molecule FRET data with X-ray experiments. The great advantage of this approach is that it accounts for various experimental errors and uncertainties due to lack of information about the system. Here, Nano Positioning System (NPS) was used for the first time. It was shown that it can be applied to the large macromolecular complexes (like eukaryotic RNA polymerase II).

The future development of the NPS technique will diverse in few directions. First, it is necessary to include information about orientation of the SDM and ADM. Here, more detailed modeling will be required. Second, by measuring FRET signals also between SDMs and ADM itself one can obtain “FRET data network”. By doing that it should be possible to get information about unknown position even if the x-ray structure is not known. Finally, NPS can be combined with electron microscopy data. In case of multi-subunit complexes it is often difficult to distinguish between densities and corresponding subunits, in this sense NPS gives a great tool to gain additional information about the structure studied. NPS can be used, for example, to investigate

transcription initiation. This very first step of transcription is at the same time most complicated. NPS is well suited to answer some very basic question, e.g. position and orientation of transcription factors like, TBP or TFIIB.

The time resolution of single-molecule fluorescence experiments allows for an investigation of structural changes in real time. The new direction would be then investigation the transcribing enzyme, one molecule at the time. It could result in a dynamic view of such a complex biological process like transcription, providing valuable insights into its unclear mechanistic aspects and eventually also its regulation. Moreover, fluorescence data can be completed with the single molecule force spectroscopy.

Appendix

Appendix

A1 Script used for calculation the intersection of 3 spheres:

```
proc overl {rad1 rad2 rad3 cid1 spra} {  
    # set variabels for Radii  
    set r1 $rad1  
    set r2 $rad2  
    set r3 $rad3  
  
    # set variabels for atomselections of Sphere centers  
    set sel1 [atomselect top "chain name and resid number and name 1"]  
    set sel2 [atomselect top " chain name and resid number and name 2"]  
    set sel3 [atomselect top " chain name and resid number and name 3"]  
  
    # get the coordinates of the atom  
    lassign [$sel1 get {x y z}] coord1  
    lassign [$sel2 get {x y z}] coord2  
    lassign [$sel3 get {x y z}] coord3  
  
    set centr [vecsacle 0.3333333333 [vecadd $coord1 $coord2 $coord3]]  
  
    # make new coordinate system  
    # Distance vectors  
    set dis12 [vecsub $coord1 $coord2]  
    set dis13 [vecsub $coord1 $coord3]  
    set dis23 [vecsub $coord2 $coord3]  
  
    # New basis vectors  
    set xt [vecnorm $dis23]  
    set yt [vecnorm [veccross $dis13 $dis23]]  
    set zt [veccross $xt $yt]  
  
    # centers in new basis  
    set pos3 {0 0 0}  
    set pos2 "[veclength $dis23] 0 0"  
    set pos1 "[vecdot $dis13 $xt] 0 [vecdot $dis13 $zt]"  
  
    # calculate intesection points  
    set is1t [eval expr pow($r3, 2) - pow($r2, 2) + pow([veclength $dis23],2)]  
    set is1 [eval expr $is1t / (2*[veclength $dis23])]   
  
    set is22 [eval expr pow($r3, 2) - pow($is1,2)]  
    set is3t [eval expr (pow($r3,2) -pow($r1,2)-2*[lindex $pos1 0]*$is1 + pow([lindex  
$pos1 0],2) + pow([lindex $pos1 2],2))]   
  
    set is3 [eval expr $is3t / (2 * [lindex $pos1 2])]   
  
    set is2 [eval expr sqrt($is22 - pow($is3, 2))]
```

Appendix

```
set ISP1 "$is1 $is2 $is3"
set ISP2 "$is1 -$is2 $is3"

# transform points into original basis

# make elements of transformation matrix

    set RTM11 [lindex $xt 0]
    set RTM12 [lindex $xt 1]
    set RTM13 [lindex $xt 2]
    set RTM21 [lindex $yt 0]
    set RTM22 [lindex $yt 1]
    set RTM23 [lindex $yt 2]
    set RTM31 [lindex $zt 0]
    set RTM32 [lindex $zt 1]
    set RTM33 [lindex $zt 2]
    set RTM41 [lindex $coord3 0]
    set RTM42 [lindex $coord3 1]
    set RTM43 [lindex $coord3 2]

# calculate coordinate retransformation

set IP11 [eval expr $is1 * $RTM11 +$is2 * $RTM21 +$is3 * $RTM31 + $RTM41]
set IP12 [eval expr $is1 * $RTM12 +$is2 * $RTM22 +$is3 * $RTM32 + $RTM42]
set IP13 [eval expr $is1 * $RTM13 +$is2 * $RTM23 +$is3 * $RTM33 + $RTM43]

set IP21 [eval expr $is1 * $RTM11 -$is2 * $RTM21 +$is3 * $RTM31 + $RTM41]
set IP22 [eval expr $is1 * $RTM12 -$is2 * $RTM22 +$is3 * $RTM32 + $RTM42]
set IP23 [eval expr $is1 * $RTM13 -$is2 * $RTM23 +$is3 * $RTM33 + $RTM43]

# and draw the spheres at intersection points and store them in a File

draw color $cid1
draw sphere "$IP11 $IP12 $IP13" radius 2 resolution 50
draw color $cid2
draw sphere "$IP21 $IP22 $IP23" radius 2 resolution 50
set Erg1 [open d:/folder/Inters1.dat a+]
set Erg2 [open d:/folder/Inters2.dat a+]
puts $Erg1 "$IP11 $IP12 $IP13"
puts $Erg2 "$IP21 $IP22 $IP23"
close $Erg1
close $Erg2

}
```

A2 Histograms

A2.1 Single-molecule tracking of RNA exiting from RNA polymerase II

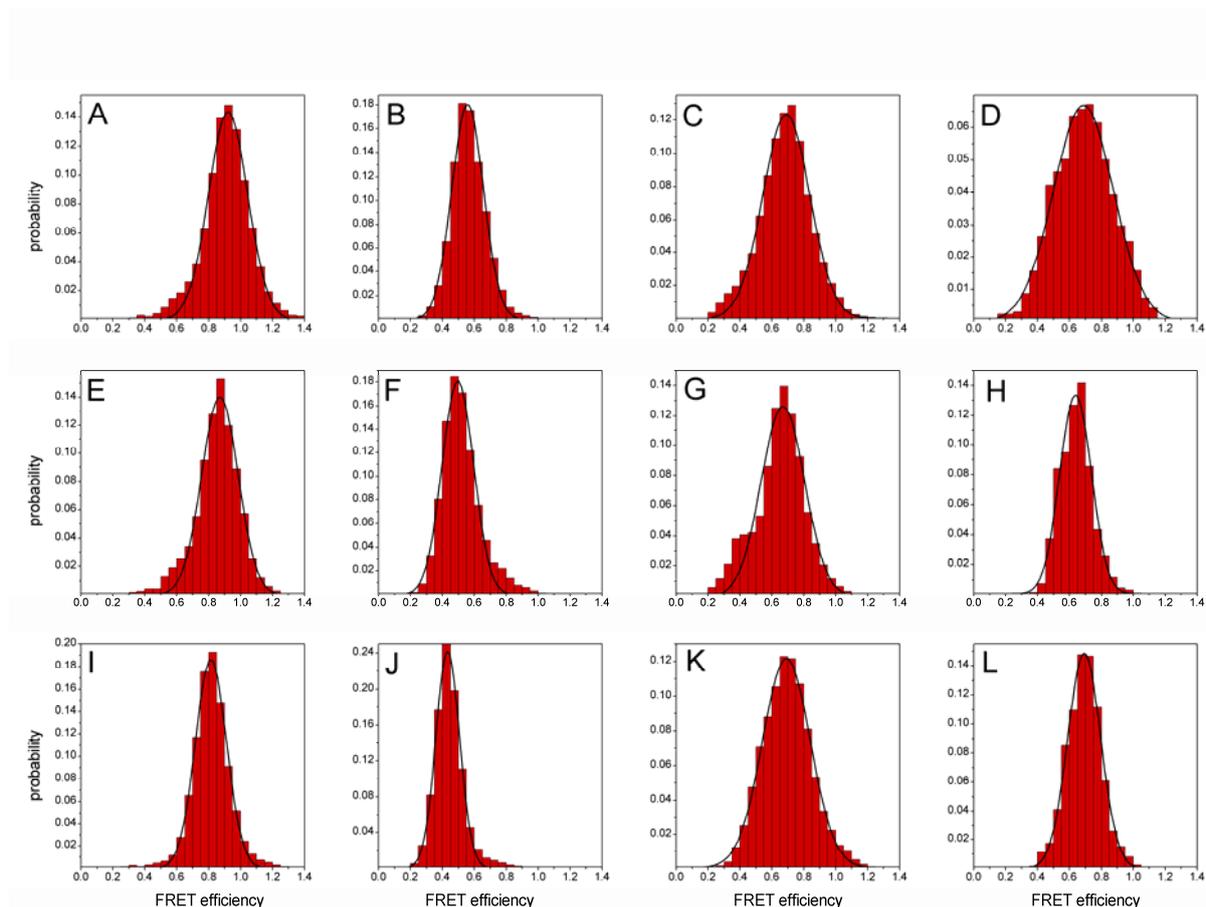


Figure A1: Complete histograms of the FRET pairs for RNA17, RNA20 and RNA23

(A) DNA1-RNA17, (B) DNA2-RNA17, (C) Rpb7/C150-RNA17, (D) Rpb7/C94-RNA17, (E) DNA1-RNA20, (F) DNA2-RNA20, (G) Rpb7/C150-RNA20, (H) Rpb7/C94-RNA20 (I) DNA1-RNA23, (J) DNA2-RNA23, (K) Rpb7/C150-RNA23 and (L) Rpb7/C94-RNA23. The histograms show a single peak distribution and were each fitted with single gaussian yielding mean FRET efficiencies (see Table xxx) that were used for triangulation and for determining the positions of RNA17, RNA20 and RNA23.

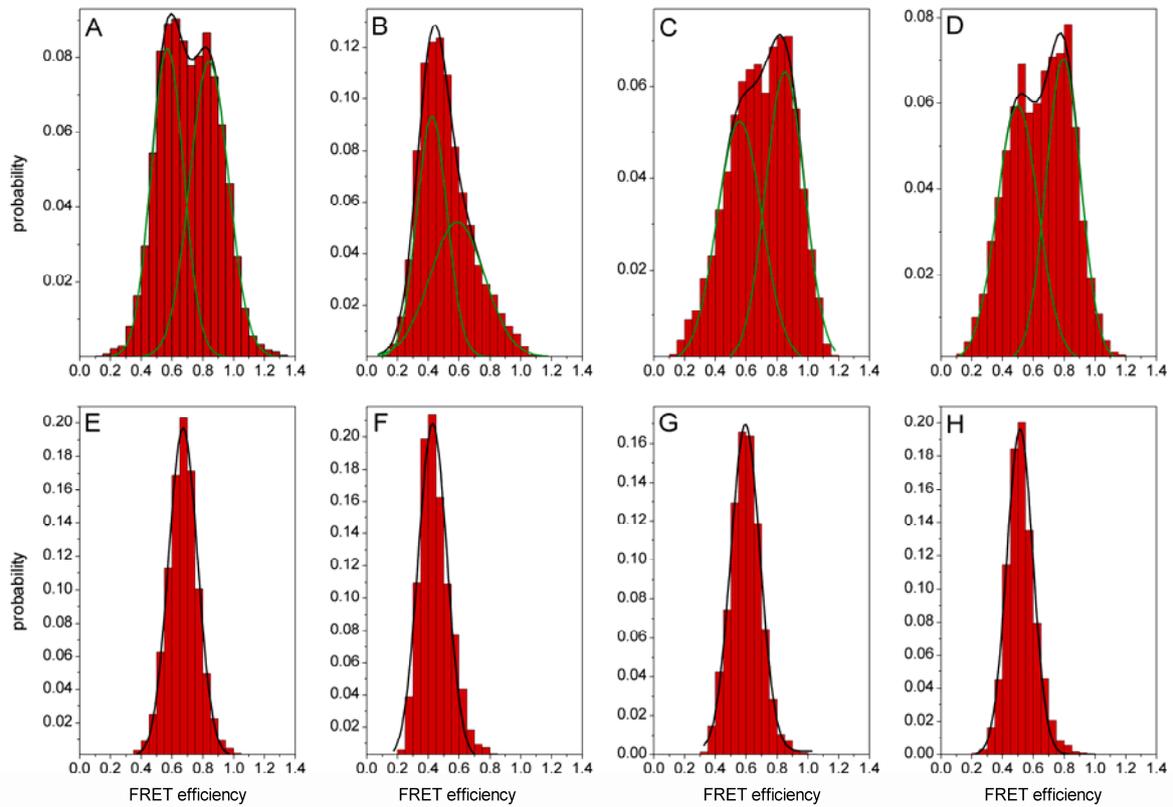


Figure A2: Complete histograms of the FRET pairs for RNA26 and RNA29

(A) DNA1-RNA26, (B) DNA2-RNA26, (C) Rpb7/C150-RNA26, (D) Rpb7/C94-RNA26, (E) DNA1-RNA29, (F) DNA2-RNA29, (G) Rpb7/C150-RNA29, (H) Rpb7/C94-RNA29. The histograms were fitted with a single (E-H) or double (A-D) Gaussian indicated by the black and green lines (all results of the fits are summarized in Table xxx).

A2.2 Position of 5' end of the 29nt RNA determined by NPS and influence of transcription factor IIB on the position of RNA

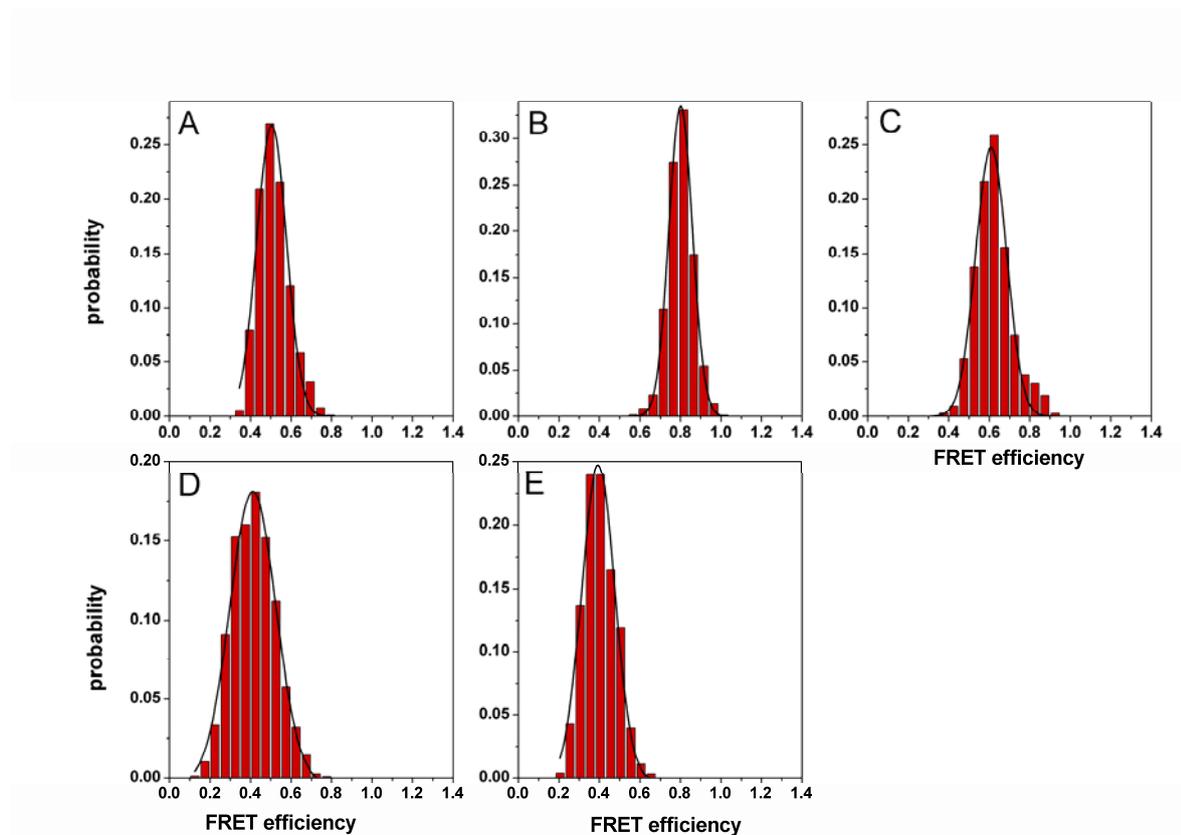


Figure A3: FRET histograms of RNA1 measurements

Shown are the histograms of the measurements with donor attached to the RNA3' end and the acceptor attached to tDNA(-10) (A), tDNA(+3) (B), tDNA(+9) (C), Rpb7-C150(D) and Rpb7-C16(E). As a donor dye Alexa 555 (A-C, E) or TMR (D) and as an acceptor dye Alexa 647 was used. Gaussian fits to the histograms are shown by black line.

A2.3 Nano Positioning System reveals the course of upstream and nontemplate DNA within Pol II elongation complex

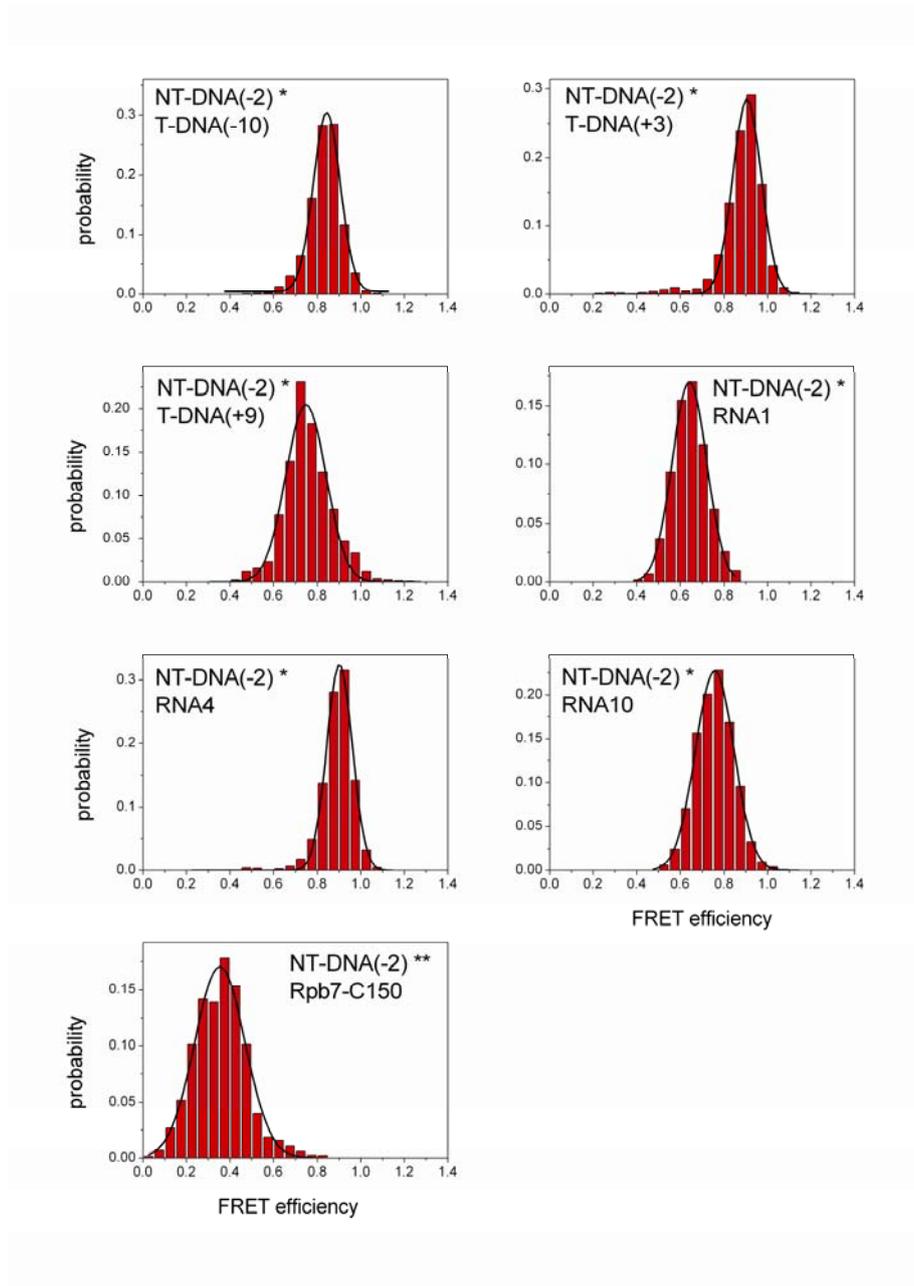


Figure A4: Complete histograms of the FRET efficiency between NT-DNA(-2) and all SDMs used to determine this position

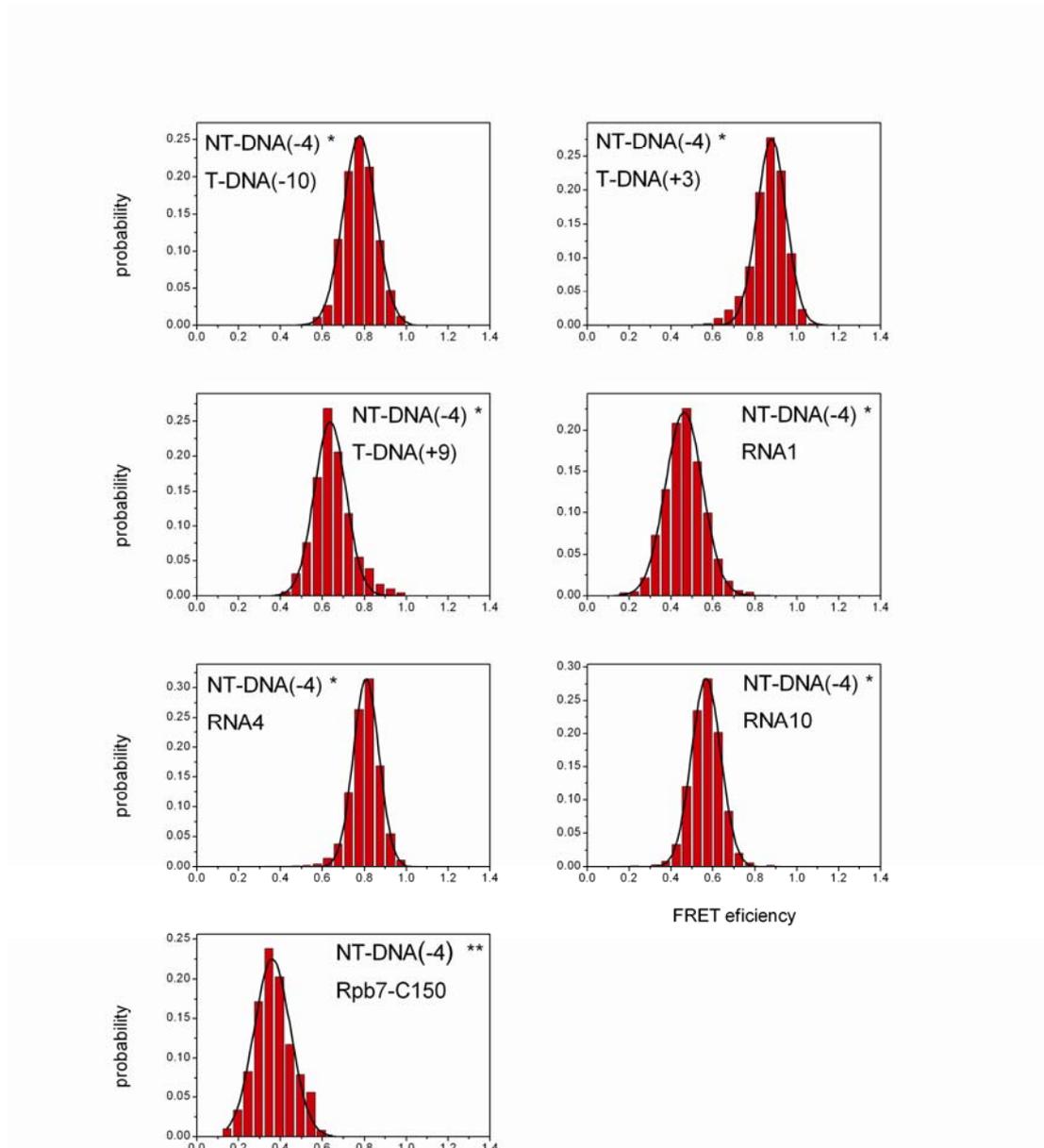


Figure A5: Complete histograms of the FRET efficiency between NT-DNA(-4) and all SDMs used to determine this position

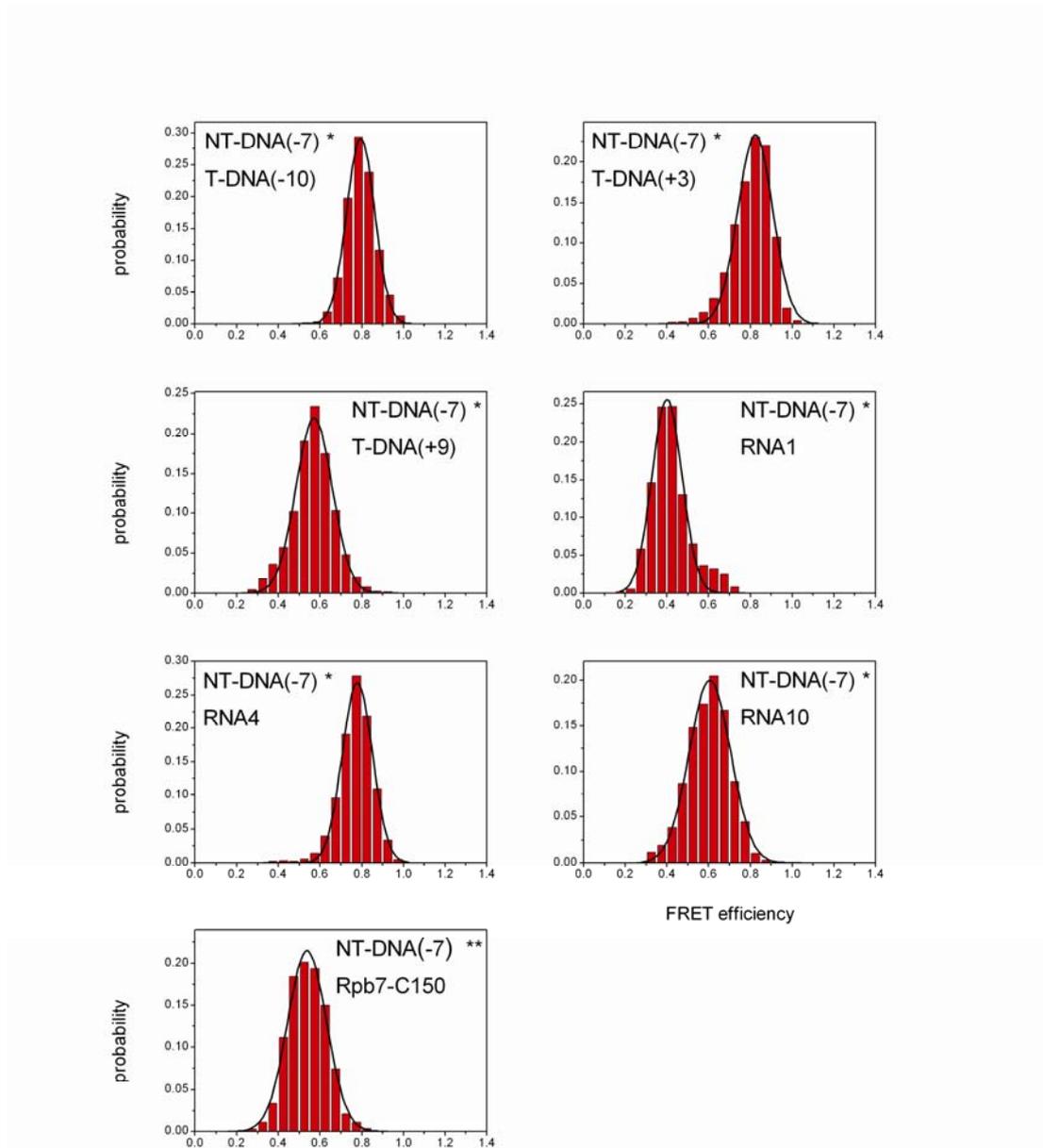


Figure A6: Complete histograms of the FRET efficiency between NT-DNA(-7) and all SDMs used to determine this position

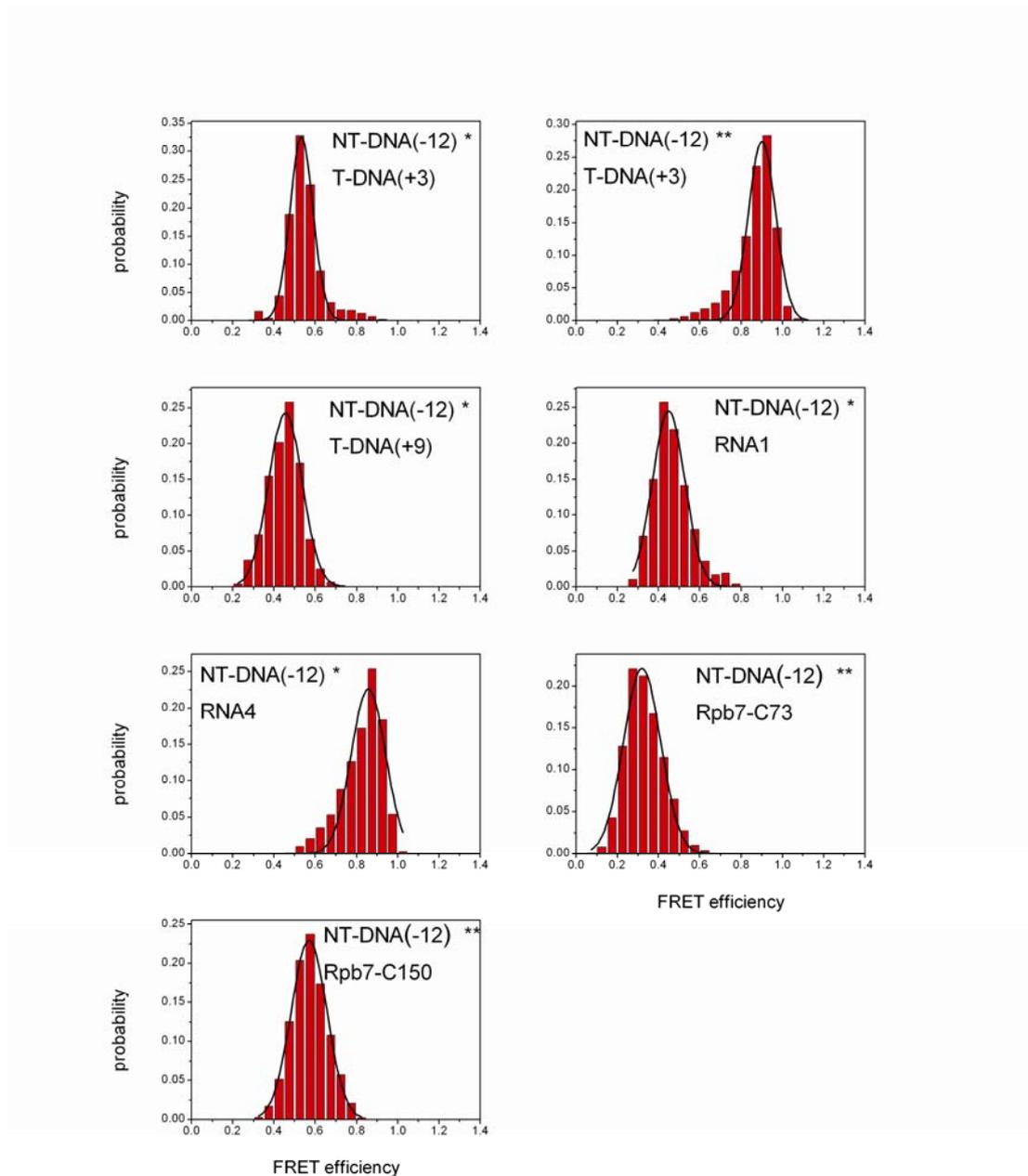


Figure A7: Complete histograms of the FRET efficiency between NT-DNA(-12) and all SDMs used to determine this position

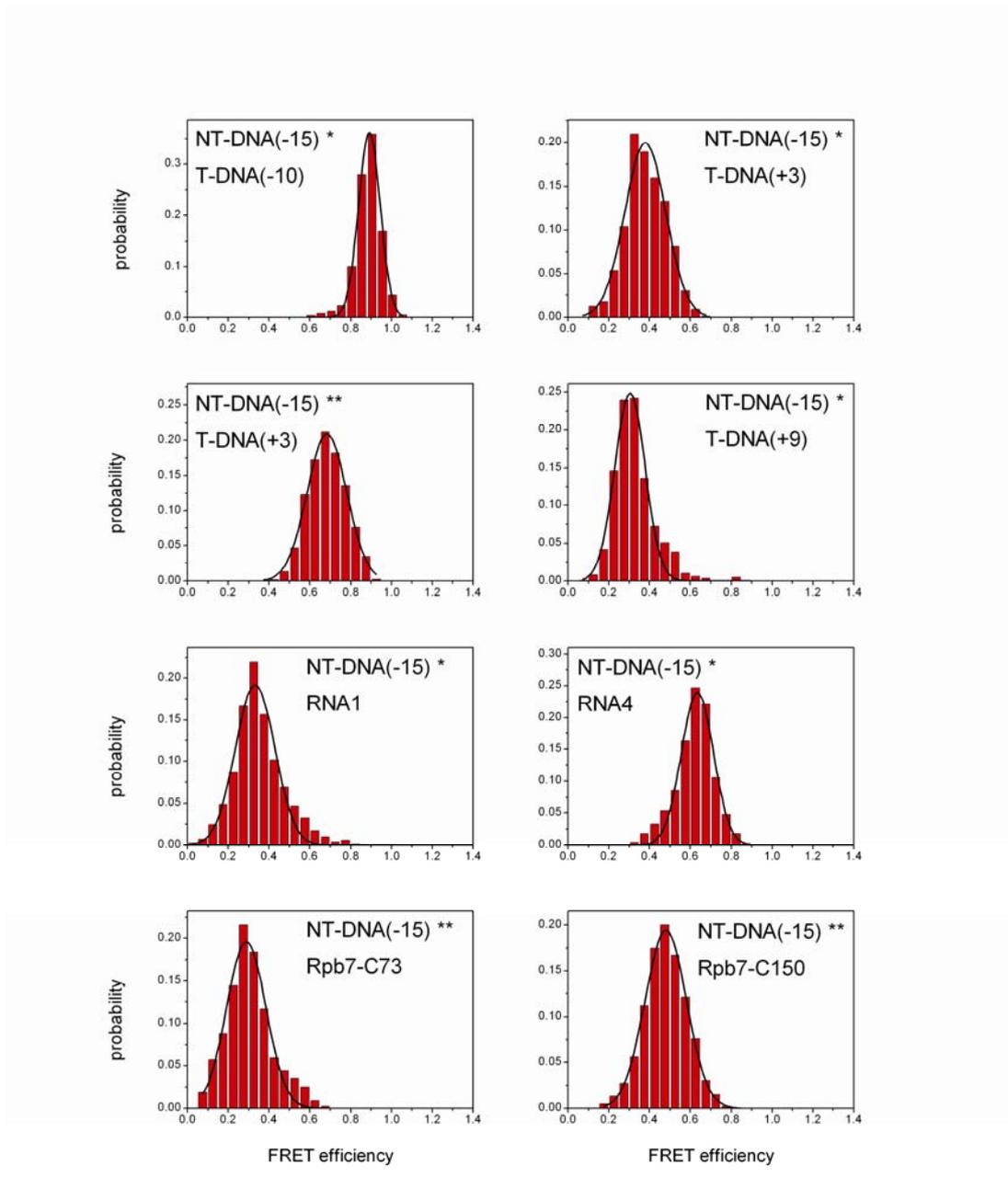


Figure A8: Complete histograms of the FRET efficiency between NT-DNA(-15) and all SDMs used to determine this position

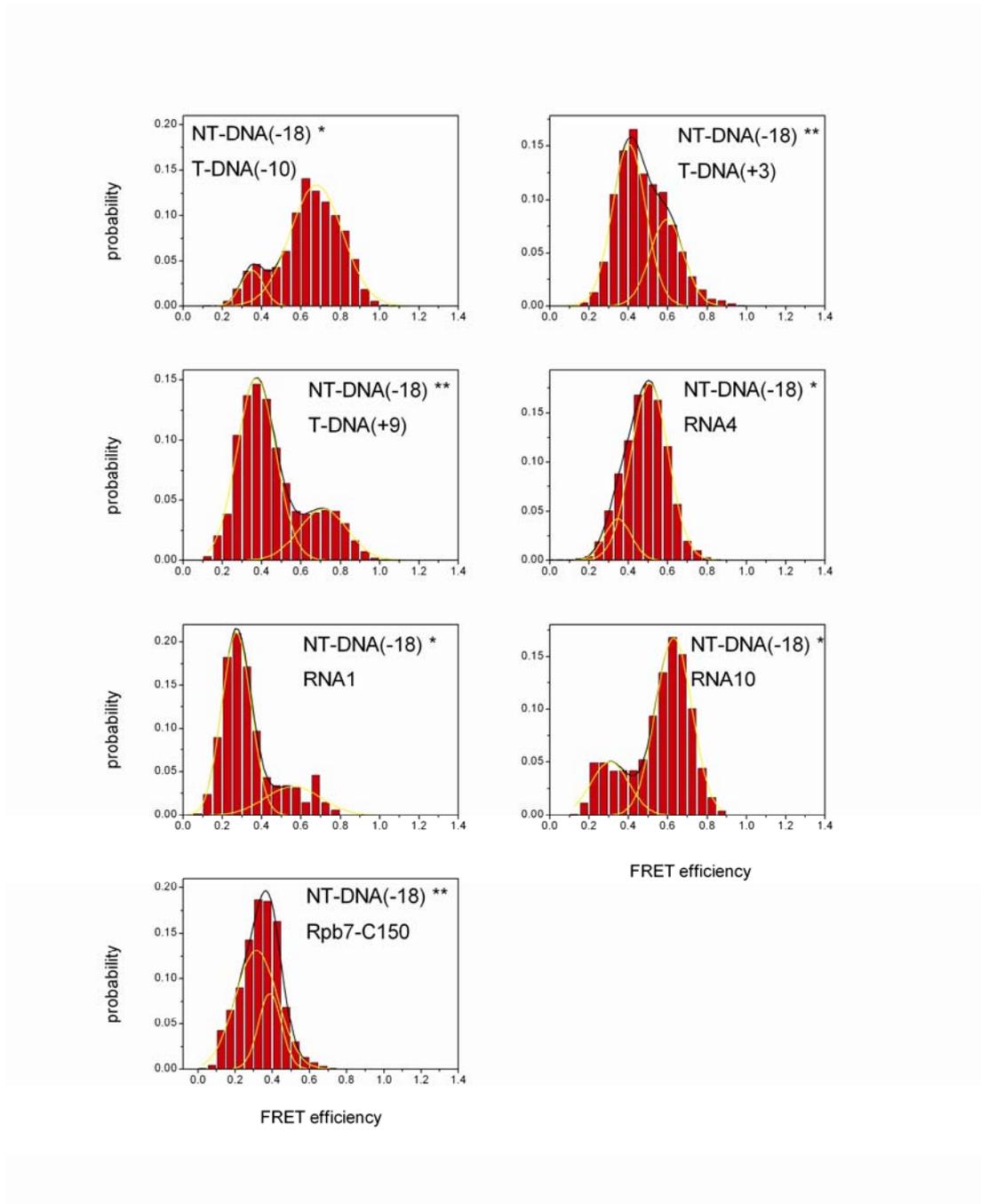


Figure A9: Complete histograms of the FRET efficiency between NT-DNA(-18) and all SDMs used to determine this position

* alexa 555 used as donor
 ** TMR used as donor

A3 Evidence calculated for the 32 possible RNA26 positions

Table A3. Evidence calculated for the 32 possible RNA26 positions

The FRET efficiency of the RNA26 elongation complexes switches repeatedly between two levels and the corresponding histograms show two peaks. Therefore one can perform 32 different triangulations resulting in 16 pair of RNA26 positions. The table presents the calculated evidences for all positions of RNA26. Original FRET data used for the calculations are shown for every triangulation.

pair	tDNA(-10)	tDNA(+3)	C150	C94	C73	evidence	tDNA(-10)	tDNA(+3)	C150	C94	C73	evidence
1.	58	42	56	50	47	11.2399	85	59	86	80	59	12.8217
2.	85	42	56	50	47	12.0715	58	59	86	80	59	11.6436
3.	85	59	56	50	47	12.0602	58	42	86	80	59	11.8332
4.	85	59	86	50	47	11.7230	58	42	56	80	59	11.5600
5.	85	59	86	80	47	12.9000	58	42	56	50	59	11.4064
6.	58	59	56	50	47	11.0193	85	42	86	80	59	12.7533
7.	58	59	86	50	47	10.1255	85	42	56	80	59	11.8442
8.	58	42	86	80	47	11.8526	85	59	56	50	59	11.9862
9.	58	42	86	50	47	9.8360	85	59	56	80	59	11.6019
10.	58	59	86	80	47	11.5235	85	42	56	50	59	11.8955
11.	58	42	56	80	47	11.5701	85	59	86	50	59	11.6892
12.	58	59	56	80	47	10.8443	85	42	86	50	59	11.1769
13.	58	59	86	50	59	10.4748	85	42	56	80	47	12.0368
14.	58	59	56	80	59	10.9544	85	42	86	50	47	11.3618
15.	85	42	86	80	47	12.8353	58	59	56	80	59	11.242
16.	85	59	56	80	47	11.7093	58	42	86	50	59	12.0368

References

- Abbondanzieri, E.A., Greenleaf, W.J., Shaevitz, J.W., Landick, R., and Block, S.M. (2005). Direct observation of base-pair stepping by RNA polymerase. *Nature* *438*, 460-465.
- Adelman, K., La Porta, A., Santangelo, T.J., Lis, J.T., Roberts, J.W., and Wang, M.D. (2002). Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behavior. *Proc Natl Acad Sci U S A* *99*, 13538-13543.
- Afonine, P.V., Grosse-Kunstleve, R.W., and Adams, P.D. (2005). A robust bulk-solvent correction and anisotropic scaling procedure. *Acta Crystallogr D Biol Crystallogr* *61*, 850-855.
- Andrecka, J., Lewis, R., Bruckner, F., Lehmann, E., Cramer, P., and Michaelis, J. (2008). Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proc Natl Acad Sci U S A* *105*, 135-140.
- Armache, K.-J., Kettenberger, H., and Cramer, P. (2003). Architecture of the initiation-competent 12-subunit RNA polymerase II. *Proc. Natl. Ac. Sc. USA* *100*, 6964-6968.
- Armstrong, J.A. (2007). Negotiating the nucleosome: factors that allow RNA polymerase II to elongate through chromatin. *Biochem Cell Biol* *85*, 426-434.
- Asturias, F., Meredith, G., Poglitsch, C., and Kornberg, R. (1997). Two conformations of RNA polymerase II revealed by electron crystallography. *J. Mol. Biol.* *272*, 536-540.
- Axelrod, D. (1984). Total internal reflection fluorescence in biological systems. *Journal of Luminescence* *31-32*, 881.
- Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* *226*, 1209-1211.
- Bennink M. L., L.S.H., Leno G. H., Zlatanova J., de Grooth B. G., Greve J. (2001). Unfolding individual nucleosomes by stretching single chromatin fibers with optical tweezers. *Nat. Struct. Biol.*, 606 – 610.
- Brueckner, F., Hennecke, U., Carell, T., and Cramer, P. (2007). CPD damage recognition by transcribing RNA polymerase II. *Science* *315*, 859-862.
- Buratowski, S. (2003). The CTD code. *Nat Struct Biol* *10*, 679-680.

References

- Bushnell, D.A., Westover, K.D., Davis, R.E., and Kornberg, R.D. (2004). Structural basis of transcription: an RNA polymerase II-TFIIB cocystal at 4.5 Angstroms. *Science* *303*, 983-988.
- Chen, H.-T., Warfield, L., and Hahn, S. (2007). The positions of TFIIF and TFIIE in the RNA polymerase II transcription initiation complex. *Nat Struct Mol Biol* *8*, 696-703.
- Chen, H.T., and Hahn, S. (2003). Binding of TFIIB to RNA polymerase II: Mapping the binding site for the TFIIB zinc ribbon domain within the preinitiation complex. *Mol Cell* *12*, 437-447.
- Chen, H.T., and Hahn, S. (2004). Mapping the location of TFIIB within the RNA polymerase II transcription preinitiation complex: a model for the structure of the PIC. *Cell* *119*, 169-180.
- Clegg, R.M. (1992). Fluorescence resonance energy transfer and nucleic acids. *Methods Enzymol* *211*, 353-388.
- Cramer, P. (2007). Finding the right spot to start transcription. *Nat. Struct. Mol. Biol.* *14*, 686-687.
- Cramer, P., Armache, K.J., Baumli, S., Benkert, S., Brueckner, F., Buchen, C., Damsma, G.E., Dengl, S., Geiger, S.R., Jasiak, A.J., *et al.* (2008). Structure of eukaryotic RNA polymerases. *Annu Rev Biophys* *37*, 337-352.
- Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000a). Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* *288*, 640-649.
- Cramer, P., Bushnell, D.A., Fu, J., Gnatt, A.L., Maier-Davis, B., Thompson, N.E., Burgess, R.R., Edwards, A.M., David, P.R., and Kornberg, R.D. (2000b). Architecture of RNA polymerase II and implications for the transcription mechanism [see comments]. *Science* *288*, 640-649.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001a). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* *292*, 1863-1876.
- Cramer, P., Bushnell, D.A., and Kornberg, R.D. (2001b). Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* *292*, 1863-1876.
- Dale, R.E., Eisinger, J., and Blumberg, W.E. (1979). The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. *Biophys J* *26*, 161-193.
- Darst, S.A., Edwards, A.M., Kubalek, E.W., and Kornberg, R.D. (1991). Three-dimensional structure of yeast RNA polymerase II at 16Å resolution. *Cell* *66*, 121-128.

References

- Diez, M., Zimmermann, B., Borsch, M., König, M., Schweinberger, E., Steigmiller, S., Reuter, R., Felekyan, S., Kudryavtsev, V., Seidel, C.A., and Graber, P. (2004). Proton-powered subunit rotation in single membrane-bound F₀F₁-ATP synthase. *Nat Struct Mol Biol* *11*, 135-141.
- Edwards, A.M., Darst, S.A., Feaver, W.J., Thompson, N.E., Burgess, R.R., and Kornberg, R.D. (1990). Purification and lipid layer crystallization of yeast RNA polymerase II. *Proc. Natl. Acad. Sci. USA* *87*, 2122-2126.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* *60*, 2126-2132.
- Esnouf, R.M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* *15*, 132-134.
- Forde, N.R., Izhaky, D., Woodcock, G.R., Wuite, G.J., and Bustamante, C. (2002). Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by *Escherichia coli* RNA polymerase. *Proc Natl Acad Sci U S A* *99*, 11682-11687.
- Fu, J., Gnatt, A.L., Bushnell, D.A., Jensen, G., J., Thompson, N.E., Burgess, R.R., David, P.R., and Kornberg, R.D. (1999). Yeast RNA polymerase II at 5 Å resolution. *Cell* *98*, 799-810.
- Gilmour, D.S., and Fan, R. (2008). Derailing the locomotive: transcription termination. *J Biol Chem* *283*, 661-664.
- Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001a). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* *292*, 1876-1882.
- Gnatt, A.L., Cramer, P., Fu, J., Bushnell, D.A., and Kornberg, R.D. (2001b). Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 Å resolution. *Science* *292*, 1876-1882.
- Guthold, M., Bezanilla, M., Erie, D.A., Jenkins, B., Hansma, H.G., and Bustamante, C. (1994). Following the assembly of RNA polymerase-DNA complexes in aqueous solutions with the scanning force microscope. *Proc Natl Acad Sci U S A* *91*, 12927-12931.
- Ha, T., Enderle, T., Ogletree, D.F., Chemla, D.S., Selvin, P.R., and Weiss, S. (1996). Probing the interaction between two single molecules: fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc Natl Acad Sci U S A* *93*, 6264-6268.
- Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol* *11*, 394-403.

References

- Harada, Y., Funatsu, T., Murakami, K., Nonoyama, Y., Ishihama, A., and Yanagida, T. (1999). Single-molecule imaging of RNA polymerase-DNA interactions in real time. *Biophys J* 76, 709-715.
- Harada, Y., Ohara, O., Takatsuki, A., Itoh, H., Shimamoto, N., and Kinoshita, K., Jr. (2001). Direct observation of DNA rotation during transcription by *Escherichia coli* RNA polymerase. *Nature* 409, 113-115.
- Heim, R., Cubitt, A.B., and Tsien, R.Y. (1995). Improved green fluorescence. *Nature* 373, 663-664.
- Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. *Science* 308, 118-120.
- Herschel, J.F.W. (1845). On a Case of Superficial Colour Presented by a Homogeneous Liquid Internally Colourless. *Phil. Trans. R. Soc.* 135, 323-328.
- Hickerson, R., Majumdar, Z.K., Baucom, A., Clegg, R.M., and Noller, H.F. (2005). Measurement of internal movements within the 30 S ribosomal subunit using Forster resonance energy transfer. *J Mol Biol* 354, 459-472.
- Hirose, Y., and Manley, J.L. (2000). RNA polymerase II and the integration of nuclear events. *Genes Dev* 14, 1415-1429.
- <http://au.expasy.org/tools/protparam.html> (2009). ExpPASy.
- <http://www.bamarray.com/> (2009). Bayesian analysis of microarray data.
- <http://www.ks.uiuc.edu/Research/vmd> (2008). VMD.
- <http://www.photochemcad.com/> (2009). PhotochemCAD
- Ishwaran, H., and Rao, J.S. (2005). Spike and slab variable selection: frequentist and Bayesian strategies. *The Annals of Statistics* 33, 730-773.
- Jaynes, E.T. (1984). What's wrong with Bayesian methods? (unpublished manuscript). <http://baves.wustl.edu/etj/articles/bayesian.methods.pdf>.
- Kabata, H., Kurosawa, O., Arai, I., Washizu, M., Margaron, S.A., Glass, R.E., and Shimamoto, N. (1993). Visualization of single molecules of RNA polymerase sliding along DNA. *Science* 262, 1561-1563.
- Kanno, T., Huettel, B., Mette, M.F., Aufsatz, W., Jaligot, E., Daxinger, L., Kreil, D.P., Matzke, M., and Matzke, A.J. (2005). Atypical RNA polymerase subunits required for RNA-directed DNA methylation. *Nat Genet* 37, 761-765.

References

- Kapanidis, A.N., Lee, N.K., Laurence, T.A., Doose, S., Margeat, E., and Weiss, S. (2004). Fluorescence-aided molecule sorting: analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc Natl Acad Sci U S A* *101*, 8936-8941.
- Kapanidis, A.N., Margeat, E., Ho, S.O., Kortkhonjia, E., Weiss, S., and Ebright, R.H. (2006). Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* *314*, 1144-1147.
- Kettenberger, H., Armache, K.-J., and Cramer, P. (2004a). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol. Cell* *16*, 955-965.
- Kettenberger, H., Armache, K.J., and Cramer, P. (2004b). Complete RNA polymerase II elongation complex structure and its interactions with NTP and TFIIS. *Mol Cell* *16*, 955-965.
- Kireeva, M.L., Komissarova, N., and Kashlev, M. (2000a). Overextended RNA:DNA hybrid as a negative regulator of RNA polymerase II processivity. *J Mol Biol* *299*, 325-335.
- Kireeva, M.L., Komissarova, N., Waugh, D.S., and Kashlev, M. (2000b). The 8-nucleotide-long RNA:DNA hybrid is a primary stability determinant of the RNA polymerase II elongation complex. *J Biol Chem* *275*, 6530-6536.
- Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., and Darst, S.A. (2000). A structural model of transcription elongation. *Science* *289*, 619-625.
- Kulaeva, O.I., Gaykalova, D.A., and Studitsky, V.M. (2007). Transcription through chromatin by RNA polymerase II: histone displacement and exchange. *Mutat Res* *618*, 116-129.
- Kuznedelov, K., Korzheva, N., Mustaev, A., and Severinov, K. (2002). Structure-based analysis of RNA polymerase function: the largest subunit's rudder contributes critically to elongation complex stability and is not involved in the maintenance of RNA-DNA hybrid length. *Embo J* *21*, 1369-1378.
- Larson, M.H., Greenleaf, W.J., Landick, R., and Block, S.M. (2008). Applied force reveals mechanistic and energetic details of transcription termination. *Cell* *132*, 971-982.
- Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. *Cell* *128*, 707-719.
- Majumdar, Z.K., Hickerson, R., Noller, H.F., and Clegg, R.M. (2005). Measurements of Internal Distance Changes of the 30 S Ribosome Using FRET with Multiple Donor-Acceptor Pairs: Quantitative Spectroscopic Methods. *Journal of Molecular Biology* *351*, 1123.
- Malagon, F., Kireeva, M.L., Shafer, B.K., Lubkowska, L., Kashlev, M., and Strathern, J.N. (2006). Mutations in the *Saccharomyces cerevisiae* RPB1 gene conferring hypersensitivity to 6-azauracil. *Genetics* *172*, 2201-2209.

References

- Margeat, E., Kapanidis, A.N., Tinnefeld, P., Wang, Y., Mukhopadhyay, J., Ebright, R.H., and Weiss, S. (2006). Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. *Biophys J* *90*, 1419-1431.
- Margittai, M., Widengren, J., Schweinberger, E., Schroder, G.F., Felekyan, S., Haustein, E., Konig, M., Fasshauer, D., Grubmuller, H., Jahn, R., and Seidel, C.A. (2003). Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. *Proc Natl Acad Sci U S A* *100*, 15516-15521.
- Meinhart, A., Kamenski, T., Hoepfner, S., Baumli, S., and Cramer, P. (2005). A structural perspective of CTD function. *Genes Dev* *19*, 1401-1415.
- Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A.N., Niu, W., Ebright, Y.W., Levy, R., and Ebright, R.H. (2002). Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell* *108*, 599-614.
- Miller, G., and Hahn, S. (2006). A DNA-tethered cleavage probe reveals the path for promoter DNA in the yeast preinitiation complex. *Nat Struct Mol Biol* *13*, 603-610.
- Moerner, W.E., and Kador, L. (1989). Optical detection and spectroscopy of single molecules in a solid. *Physical Review Letters* *62*, 2535.
- Morris, J.S., Brown, P.J., Herrick, R.C., Baggerly, K.A., and Coombes, K.R. (2006). Bayesian Analysis of Mass Spectrometry Proteomics Data using Wavelet Based Functional Mixed Models. UT MD Anderson Cancer Center Department of Biostatistics Working Paper Series.
- Muller, B.K., Zaychikov, E., Brauchle, C., and Lamb, D.C. (2005). Pulsed interleaved excitation. *Biophys J* *89*, 3508-3522.
- Muschielok, A., Andrecka, J., Jawhari, A., Bruckner, F., Cramer, P., and Michaelis, J. (2008). A nano-positioning system for macromolecular structural analysis. *Nat Meth* *5*, 965.
- Naji, S., Bertero, M.G., Spitalny, P., Cramer, P., and Thomm, M. (2008). Structure-function analysis of the RNA polymerase cleft loops elucidates initial transcription, DNA unwinding and RNA displacement. *Nucleic Acids Res* *36*, 676-687.
- Naryshkina, T., Kuznedelov, K., and Severinov, K. (2006). The role of the largest RNA polymerase subunit lid element in preventing the formation of extended RNA-DNA hybrid. *J Mol Biol* *361*, 634-643.
- Nechaev, S., Chlenov, M., and Severinov, K. (2000). Dissection of two hallmarks of the open promoter complex by mutation in an RNA polymerase core subunit. *J Biol Chem* *275*, 25516-25522.
- Nudler, E., Goldfarb, A., and Kashlev, M. (1994). Discontinuous mechanism of transcription elongation. *Science* *265*, 793-796.

References

- Orrit, M., and Bernard, J. (1990). Single pentacene molecules detected by fluorescence excitation in a p-terphenyl crystal. *Phys Rev Lett* *65*, 2716-2719.
- Pal, M., Ponticelli, A.S., and Luse, D.S. (2005). The role of the transcription bubble and TFIIB in promoter clearance by RNA polymerase II. *Mol Cell* *19*, 101-110.
- Prescher, J.A., and Bertozzi, C.R. (2005). Chemistry in living systems. *Nat Chem Biol* *1*, 13-21.
- Rasnik, I., Myong, S., Cheng, W., Lohman, T.M., and Ha, T. (2004). 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.
- Rees, W., Keller, R.W., Vesenka, J.P., Yang, G., and Bustamante, C. (1993). Evidence of DNA bending in transcription complexes imaged by scanning force microscopy. *Science* *260*, 1646-1649.
- Revyakin, A., Ebright, R.H., and Strick, T.R. (2004). Promoter unwinding and promoter clearance by RNA polymerase: detection by single-molecule DNA nanomanipulation. *Proc Natl Acad Sci U S A* *101*, 4776-4780.
- Revyakin, A., Liu, C., Ebright, R.H., and Strick, T.R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* *314*, 1139-1143.
- Rieping, W., Habeck, M., and Nilges, M. (2005). Inferential Structure Determination. *Science* *309*, 303-306.
- Rivetti, C., Codeluppi, S., Dieci, G., and Bustamante, C. (2003). Visualizing RNA extrusion and DNA wrapping in transcription elongation complexes of bacterial and eukaryotic RNA polymerases. *J Mol Biol* *326*, 1413-1426.
- Rivetti, C., Guthold, M., and Bustamante, C. (1999). Wrapping of DNA around the E.coli RNA polymerase open promoter complex. *EMBO J* *18*, 4464-4475.
- Runner, V.M., Podolny, V., and Buratowski, S. (2008). The Rpb4 subunit of RNA polymerase II contributes to cotranscriptional recruitment of 3' processing factors. *Mol Cell Biol* *28*, 1883-1891.
- Sakurai, H., Mitsuzawa, H., Kimura, M., and Ishihama, A. (1999). The Rpb4 subunit of fission yeast *Schizosaccharomyces pombe* RNA polymerase II is essential for cell viability and similar in structure to the corresponding subunits of higher eukaryotes. *Mol Cell Biol* *19*, 7511-7518.
- Sambrook, J., and Russell, D.W. (2001). *Molecular cloning – A laboratory Manual*, 3rd edition edn (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

References

Severinov, K., and Darst, S.A. (1997). A mutant RNA polymerase that forms unusual open promoter complexes. *Proc Natl Acad Sci U S A* *94*, 13481-13486.

Sevostyanova, A., Svetlov, V., Vassylyev, D.G., and Artsimovitch, I. (2008). The elongation factor RfaH and the initiation factor σ bind to the same site on the transcription elongation complex. *Proceedings of the National Academy of Sciences* *105*, 865-870.

Shaevitz, J.W., Abbondanzieri, E.A., Landick, R., and Block, S.M. (2003). Backtracking by single RNA polymerase molecules observed at near-base-pair resolution. *Nature* *426*, 684-687.

Shimomura, O., Johnson, F.H., and Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, *Aequorea*. *J Cell Comp Physiol* *59*, 223-239.

Sidorenkov, I., Komissarova, N., and Kashlev, M. (1998). Crucial role of the RNA:DNA hybrid in the processivity of transcription. *Mol. Cell* *2*, 55-64.

Sivia, D.S. (1996). Data analysis
a Bayesian tutorial (Oxford: Clarendon Press).

Skinner, G.M., Baumann, C.G., Quinn, D.M., Molloy, J.E., and Hoggett, J.G. (2003). Promoter binding, initiation and elongation by bacteriophage T7 RNA polymerase: A single-molecule view of the transcription cycle. *J Biol Chem*.

Stokes, G.G. (1852). On the Change of Refrangibility of Light. *Phil. Trans. R. Soc.* *142*, 463-562.

Stryer, L., and Haugland, R.P. (1967). Energy transfer: a spectroscopic ruler. *Proc Natl Acad Sci U S A* *58*, 719-726.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* *41*, 207-234.

Suenaga, A., Okimoto, N., Futatsugi, N., Hirano, Y., Narumi, T., Ohno, Y., Yanai, R., Hirokawa, T., Ebisuzaki, T., Konagaya, A., and Taiji, M. (2006). Structure and dynamics of RNA polymerase II elongation complex. *Biochem Biophys Res Commun* *343*, 90-98.

Svejstrup, J.Q. (2004). The RNA polymerase II transcription cycle: cycling through chromatin. *Biochim Biophys Acta* *1677*, 64-73.

Szudy, J. (1998). Born 100 Years Ago: Aleksander Jabłoński (1898-1980). Uniwersytet Mikołaja Kopernika, Toruń, Poland.

Temin, H.M., and Mizutani, S. (1970). RNA-dependent DNA polymerase in virions of Rous sarcoma virus. *Nature* *226*, 1211-1213.

References

- Thomen, P., Lopez, P.J., and Heslot, F. (2005). Unravelling the mechanism of RNA-polymerase forward motion by using mechanical force. *Phys Rev Lett* *94*, 128102.
- Till, S., and Ladurner, A.G. (2007). RNA Pol IV plays catch with Argonaute 4. *Cell* *131*, 643-645.
- Toulokhonov, I., and Landick, R. (2006). The role of the lid element in transcription by *E. coli* RNA polymerase. *J Mol Biol* *361*, 644-658.
- Ujvari, A., and Luse, D.S. (2006). RNA emerging from the active site of RNA polymerase II interacts with the Rpb7 subunit. *Nat Struct Mol Biol* *13*, 49-54.
- Vamosi, G., Gohlke, C., and Clegg, R.M. (1996). Fluorescence characteristics of 5-carboxytetramethylrhodamine linked covalently to the 5' end of oligonucleotides: multiple conformers of single-stranded and double-stranded dye-DNA complexes. *Biophys J* *71*, 972-994.
- Vassilyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassilyeva, M.N., Borukhov, S., and Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* *417*, 712-719.
- Vassilyev, D.G., Vassilyeva, M.N., Perederina, A., Tahirov, T.H., and Artsimovitch, I. (2007a). Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* *448*, 157-162.
- Vassilyev, D.G., Vassilyeva, M.N., Perederina, A., Tahirov, T.H., and Artsimovitch, I. (2007b). Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* *448*, 157.
- Wahle, E. (1992). The end of the message: 3'-end processing leading to polyadenylated messenger RNA. *Bioessays* *14*, 113-118.
- Wahle, E., and Keller, W. (1992). The biochemistry of 3'-end cleavage and polyadenylation of messenger RNA precursors. *Annu Rev Biochem* *61*, 419-440.
- Wang, D., and Landick, R. (1997). Nuclease Cleavage of the Upstream Half of the Nontemplate Strand DNA in an *Escherichia coli* Transcription Elongation Complex Causes Upstream Translocation and Transcriptional Arrest. *J. Biol. Chem.* *272*, 5989-5994.
- Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004a). Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* *119*, 481-489.
- Westover, K.D., Bushnell, D.A., and Kornberg, R.D. (2004b). Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science* *303*, 1014-1016.

References

Xie, J., and Schultz, P.G. (2006). A chemical toolkit for proteins--an expanded genetic code. *Nat Rev Mol Cell Biol* 7, 775-782.

Yin, H., Artsimovitch, I., Landick, R., and Gelles, J. (1999). Nonequilibrium mechanism of transcription termination from observations of single RNA polymerase molecules. *Proc Natl Acad Sci U S A* 96, 13124-13129.

Yoon, J.W., Bruckbauer, A., Fitzgerald, W.J., and Klenerman, D. (2008). Bayesian inference for improved single molecule fluorescence tracking. *Biophys J* 94, 4932-4947.

Zhang, C., and Burton, Z.F. (2004). Transcription factors IIF and IIS and nucleoside triphosphate substrates as dynamic probes of the human RNA polymerase II mechanism. *J Mol Biol* 342, 1085-1099.

Zhang, G., Campbell, E.A., Minakhin, L., Richter, C., Severinov, K., and Darst, S.A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution [see comments]. *Cell* 98, 811-824.

CURRICULUM VITAE

Name Joanna Andrecka
Nationality Polish
Date of birth 12.05.1980
Place of birth Mysłowice, Poland

Education

XI 2005 – VI 2009 PhD Studies in the group of Prof. Jens Michaelis,
Department of Chemistry and Biochemistry,
Ludwig-Maximilians-University, Munich, Germany

X 2004 - X 2005 PhD Studies of Biochemistry and Biophysics,
Faculty of Biotechnology, Department of Physical Biochemistry,
Jagiellonian University of Cracow, Poland

X 2001 – VI 2007 Master Studies of Trombone,
Instrumental Faculty, Department of Winds, Percussion and Accordion,
Academy of Music, Cracow, Poland

X 1999 - VI 2004 Interdisciplinary, Individual MSc Studies of Mathematical and
Natural Sciences, Jagiellonian University of Cracow, Poland
MSc thesis: *Expression of the human dopamine receptor D1 in E.coli*
Department of Physical Biochemistry, Faculty of Biotechnology,
Jagiellonian University of Cracow, Poland
Supervisor: Prof. Zygmunt Wasylewski

IX 1995 - VI 1999 High School, Katowice, Poland

