

**Structural Investigations on HIV-1 RT using
single pair Fluorescence Resonance Energy
Transfer**

Dissertation

**zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)
des Fachbereichs Chemie
der Universität Dortmund**

**Paul James Rothwell
Max-Planck-Institut für
Molekulare Physiologie
Dortmund Juni 2002**

The work was performed in the period from October 1998 to June 2002 in the “Max-Planck-Institut für Molekulare Physiologie” in Dortmund, Abteilung Physikalische Biochemie under the supervision of Prof. Dr. Roger S. Goody

1. Gutachter: Prof. Dr. R. S. Goody
2. Gutachter: Prof. Dr. T. N. Mitchell

Contents

1. INTRODUCTION.....	1
1.1 The structure of HIV-1.....	2
1.2 The life cycle of HIV-1.....	3
1.3 Reverse transcription.....	5
1.4 Structure of HIV-1 RT	9
1.5 Substrate binding and nucleotide incorporation by HIV-1 RT	13
1.6 Single molecule spectroscopy.....	15
1.7 Aims	16
2. MATERIALS	18
2.1 Chemicals.....	18
2.2 Enzymes / Molecular weight markers.....	19
2.3 Other materials	19
2.4 Bacterial strains	20
2.5 Plasmids.....	20
2.6 Growth Medium and buffers	20
2.6.1 Growth medium	20
2.6.2 DNA buffers.....	21
2.6.3 Protein gel buffers.....	21
2.6.4 Column buffers.....	22
2.6.5 RT measurement buffers	23
2.6.6 RT storage buffers.....	23
3 METHODS	24
3.1 Molecular biology	24
3.1.1 Determination of DNA concentration	24
3.1.2 Phenol extraction to purify DNA	24

Contents

3.1.3 Purification of DNA	24
3.1.4 Restriction digestion of DNA	25
3.1.5 Ligation	25
3.1.6 Production of competent cells for electroporation	25
3.1.7 Transformation of <i>E. coli</i> using electroporation.....	25
3.1.8 Preparation of plasmid DNA	26
3.1.9 Production of glycerol stocks	26
3.1.10 TBE-agarose gel electrophoresis	26
3.1.11 In vitro mutagenesis by use of the overlap extension polymerase chain reaction	26
3.1.12 Construction of the p6HRT166[SII] Strep-tag II containing vector	29
3.1.13 Colony screening for plasmid selection	30
3.1.14 DNA sequencing	30
3.1.14 Protein expression	31
3.2 Protein biochemical methods	32
3.2.1 Denaturing SDS polyacrylamide gel electrophoresis	32
3.2.2 Purification of recombinant RT	33
3.2.3 Purification of the p51[6H] subunit	34
3.2.4 Purification of the p66 [6H-SII] subunit	34
3.2.5 Protein labelling	35
3.2.6 Formation of RT from individually labelled subunits	36
3.3 Nucleic acids	36
3.3.1 Oligonucleotides	36
3.3.2 Dye labelled oligonucleotides	36
3.3.3 Radioactive 5'-end labelling of oligonucleotides with ³² P(ATP).....	37
3.3.4 Thin layer chromatography	37
3.3.5 Denaturing polyacrylamide gel electrophoresis	37
3.3.6 Hybridisation of DNA/DNA primer / templates	38
3.4 Enzymatic methods.....	38
3.4.1 RNA dependent DNA polymerase activity	38
3.4.2 Determination of the rate of single nucleotide incorporation by RT	39
3.5 Biophysical methods	40
3.5.1 Multidimensional single molecule fluorescence spectroscopy using multiparameter fluorescence spectroscopy (MFD)	40
3.5.1.1 Setup / Confocal microscope.....	41
3.5.1.2 Data acquisition	43
3.5.1.3 Data selection	44
3.5.1.4 Data analysis.....	46

3.5.1.4.1 Time correlated single photon counting to determine fluorescence anisotropy and fluorescent lifetime	46
3.5.1.4.4 Multi-Channel-Scalar (MCS) and fluorescence intensity determination	49
3.5.1.5 Fluorescence resonance energy transfer (FRET)	50
3.5.1.6 Measurement chamber preparation	54
3.5.1.7 Sample preparation	54
4. RESULTS AND DISCUSSION	55
4.1 Protein:nucleic acid interactions	55
4.1.1 Cloning, expression, and purification of HIV-1 RT	55
4.1.2 Mutagenesis using overlap extension PCR	56
4.1.3 Expression and purification of HIV-1 RT	57
4.1.4 Protein labelling with Alexa 488	59
4.1.5 Characterisation of labelled proteins	61
4.1.5.1 RNA-dependent DNA polymerase activity	61
4.1.5.2 Single nucleotide incorporation kinetics	62
4.1.5.3 Quantum yield determination of Alexa 488 by BIFL	64
4.1.6 Single molecule experiments	65
4.1.6.1 Single molecule studies using the model system p66 ^{K287C[A488]} /p51[6H]	65
4.1.6.2 Structural studies using p66 ^{K287C[A488]} /p51[6H]: p_5'_Cy5/t	70
4.1.6.3 Dissociation rate constant- K_{off} measurements	73
4.1.6.4 Basic model of RT:p/t interactions	74
4.1.6.5 Addition of pyrophosphate or dNTPs to a preterminated p/t	76
4.1.6.6 Role of the template in p/t positioning	81
4.1.6.7 Construction of a low resolution model of the productive and dead end complexes	83
4.1.6.7.1 Determination of the position of the donor fluorophore on the protein	84
4.1.6.7.2 determination of the position of the acceptor fluorophore on the p/t	86
4.1.6.7.3 Donor only samples	88
4.1.6.7.4 Structural studies using the substrate p_5'_Cy5/t	90
4.1.6.7.5 Structural studies using the substrate p_10bp_Cy5/t	92
4.1.6.7.6 Structural studies using the substrate p_3'_Cy5-t	94
4.1.6.7.7 Structural studies using the substrate p/t_7o_Cy5	95
4.1.6.7.8 Structural studies using the substrate p/t_5'_Cy5	97
4.1.6.7.9 The structure of the P and DE complexes	99
4.2 Subunit interactions and conformational changes	103
4.2.1 Cloning of p66[6H-SII], and purification of the individual subunits	103
4.2.1.2 Purification of the p51[6H] subunit	103
4.2.1.2 Purification of the p66[6H-SII] subunit	104
4.2.2 Protein labelling	105

Contents

4.2.3. Single molecule experiments	109
4.2.3.1. The structure of the heterodimeric enzyme	109
4.2.3.2. The structure of monomeric p66	111
5. SUMMARY	114
6. FUTURE STUDIES	116
7. REFERENCE LIST	118
8. APPENDICES	129
Appendix 1. Nucleic acids.....	129
8.1.1 Oligonucleotides used for MFD experiments (5'-3').....	129
8.1.2 Mutagenesis Oligonucleotides (5'-3').....	129
8.1.3. Fluorescent modified bases used	130
Appendix 2. programs used during RT:p/t interaction modelling	131
8.2.1 Display_sphere.....	131
8.2.2. Generate_overlap	135
8.2.3 Input cards.....	141
8.2.3.1 Input card for the p_5_Cy5/t (P-complex).....	141
8.2.3.2 Input card for the p_10bp_Cy5/t (P-complex).....	142
8.2.3.3 Input card for the p_3_Cy5/t (P-complex).....	142
8.2.3.4 Input card for the p/t_7o_Cy5 (P-complex).....	142
8.2.3.5 Input card for the p/t_5_Cy5 (P-complex).....	143
8.2.3.6 Input card for the p_5_Cy5/t (DE-complex)	143
8.2.3.7 Input card for the p_10bp_Cy5/t (DE-complex).....	143
8.2.3.8 Input card for the p_3_Cy5/t (DE-complex)	144
8.2.3.9 Input card for the p/t_7o_Cy5 (DE-complex)	144
8.2.3.10 Input card for the p/t_5_Cy5 (DE-complex)	144

Abbreviations

A	acceptor
AIDS	acquired immunodeficiency syndrome
ALV	avian leukosis virus
APD	avalanche photodiode
bp	base pair
BIFL	burst integrated fluorescence lifetime
Ci	curie
cpm	counts per minute
CTS	central termination sequence
D	donor
Da	dalton
dATP	deoxyadenosine triphosphate
dCTP	deoxycytosine triphosphate
DE	dead end complex
dGTP	deoxyguanine triphosphate
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleic triphosphate
dsDNA	double stranded DNA
dTTP	deoxythymidine triphosphate
<i>E. coli</i>	<i>Escherichia coli</i>
EIAV	equine infectious anaemia virus
Fab	fragment antigen binding
FCS	fluorescence correlation spectroscopy
FRET	fluorescence resonance energy transfer
HIV	human immunodeficiency virus
HTLV	human T cell leukaemia viruses
LB	Luria broth
LTR	long terminal repeat
MCS	multichannel scalar
MFD	multiparameter fluorescence detection
MLE	maximum likelihood estimator
MLV	mouse leukaemia virus

Abbreviations

MMTV	mouse mammary tumour virus
MPMV	mason-Pfizer monkey virus
mRNA	messenger RNA
NNRTI	non nucleoside reverse transcriptase inhibitor
NRTI	nucleoside reverse transcriptase inhibitor
P	productive complex
PAGE	polyacrylamide gel electrophoresis
PBS	primer binding site
PP	preproductive complex
PCR	polymerase chain reaction
PPT	polypurine tract
p/t	primer/template
rpm	revolutions per minute
RNA	ribonucleic acid
RNase H	ribonuclease H
RSV	rous sarcoma virus
RT	reverse transcriptase
SIV	simian immunodeficiency virus
SMS	single molecule spectroscopy
spFRET	single pair FRET
TAC	time to amplitude converter
TCSPC	time correlated single photon counting
tRNA	transfer ribonucleic acid
U	unit
v/v	volume/volume
w/v	weight/volume

Abbreviations

Amino acids

One letter code	Three letter code	Name
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine

1. Introduction

The first retrovirus was discovered in 1911 by Peyton Rous (1911). Rous successfully transmitted solid tumours of chickens by transplanting tissue and was able to isolate the causative agent, Rous sarcoma virus (RSV). This was followed by the discovery of many other examples of acutely transforming retroviruses, such as mouse leukaemia virus (MLV) and mouse mammary tumour virus (MMTV) (Bittner.J.J. 1936).

Retroviruses have received much attention in recent years due to the discovery of the first human retrovirus in the early '80s (Poiesz *et al.* 1980) and the identification of an etiological role for retroviruses in human disease (Barre-Sinoussi *et al.* 1983; Gallo *et al.* 1982; Gallo *et al.* 1984; Popovic *et al.* 1984; Robert-Guroff *et al.* 1982; Sarngadharan *et al.* 1984). In humans, retroviruses have been shown to cause leukaemia through the human T cell leukaemia viruses, HTLV-1 and HTLV-2, and acquired immunodeficiency syndrome (AIDS) through the human immunodeficiency viruses, HIV-1 and HIV-2 (Barre-Sinoussi *et al.* 1983; Gallo *et al.* 1984; Levy *et al.* 1984; Poiesz *et al.* 1980).

Morphologically, although retroviruses resemble other viruses in several respects, they differ from all others in containing an RNA genome that replicates through a DNA intermediate. Retroviruses can be divided into groups based on their morphology in negatively stained electron micrograph pictures:

- A-type: Non-enveloped immature particles only seen inside cells, believed to result from endogenous retrovirus like genetic elements.
- B-Type: Enveloped, extracellular particles with a condensed, acentric core and prominent envelope spikes, e.g. MMTV.
- C-Type: Similar to B-type, but with a central core and barely visible spikes, e.g. most mammalian and avian retroviruses: MLV, HTLV, and HIV.
- D-Type: Slightly larger virus (120 nm) with less prominent spikes, e.g. Mason-Pfizer monkey virus (MPMV).

1.1 The structure of HIV-1

HIV-1 is a member of the C-type retroviruses and belongs to the lentivirus, or “slow” virus, genus. Other examples of lentiviruses include the simian immunodeficiency virus (SIV), visna virus and equine infectious anaemia virus (EIAV).

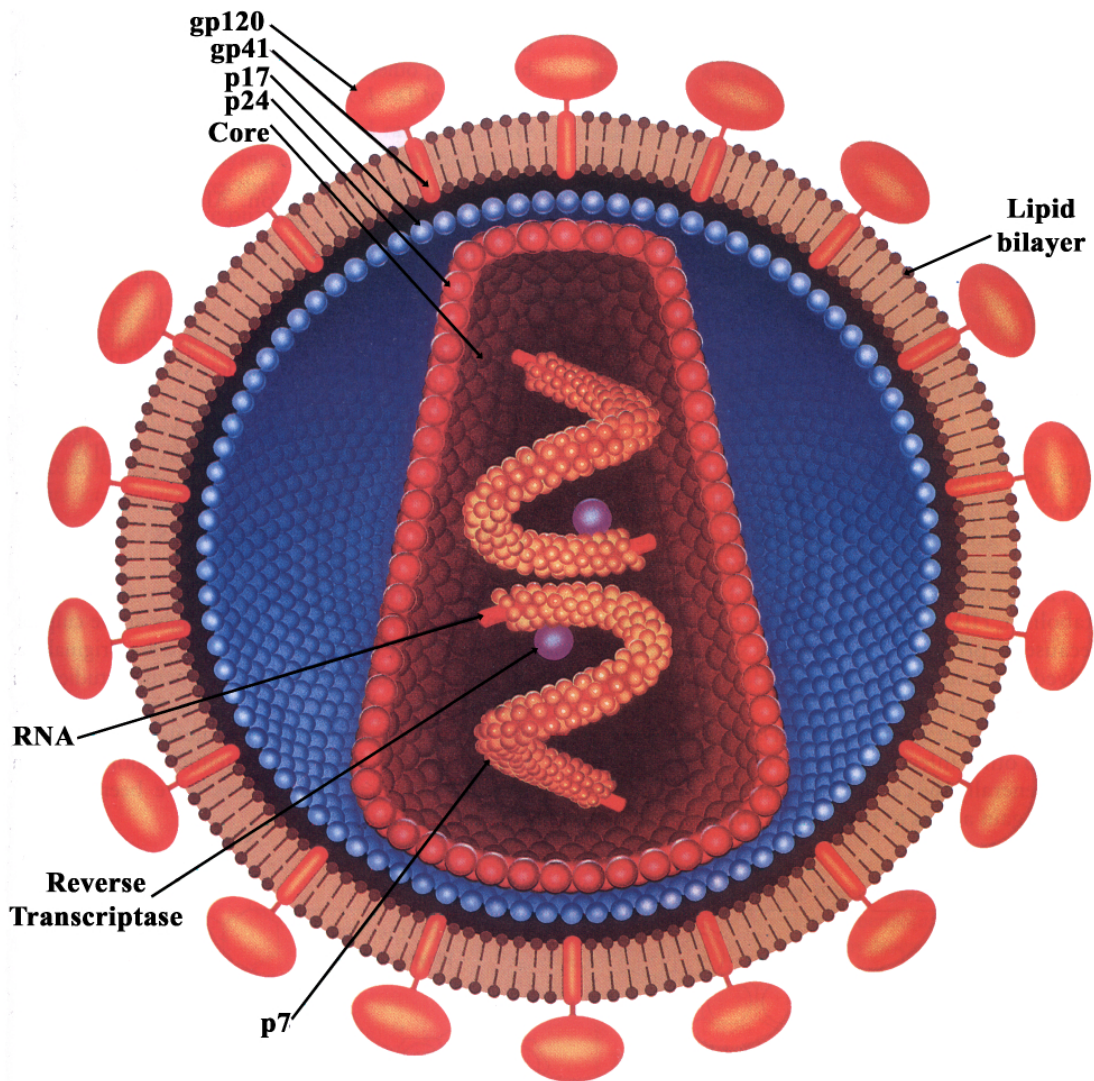


Figure 1.1. The structure of HIV-1 (modified from (Darnell et al. 1990)).

The structure of the HIV-1 particle is shown in Figure 1.1. The HIV-1 particle is roughly spherical and approximately 100 nm in diameter. The outer coat of HIV-1 consists of a lipid bilayer derived from the human host cell. Exposed surface

glycoproteins composed of gp120 are anchored to the lipid bilayer via the transmembrane protein, gp41. The lipid bilayer also contains proteins derived from the host cell, including the class I and II major histocompatibility complexes which, in their normal location, are important in controlling the host cell immune response (Arthur *et al.* 1992). A matrix shell comprising approximately 2000 copies of the matrix protein (p17) lines the inner surface of the viral membrane. Located in the centre of the virus is a conical capsid core which is made up of approximately 2000 copies of capsid protein, p24. The capsid particle encapsulates two copies of viral RNA. The viral RNA is stabilised in a ribonucleotide complex with approximately 2000 copies of the nucleocapsid protein, p7. The core also contains three virally encoded enzymes: protease, reverse transcriptase (RT), and integrase. The accessory proteins Nef, Vif, and Vpr, as well as a cellular derived tRNA, also appear to be packaged. Three additional accessory proteins that function in the host cell, Rev, Tat, and Vpu, do not appear to be present in the core (Turner and Summers 1999).

1.2 The life cycle of HIV-1

AIDS is caused by infection with HIV. The virus is transmitted by sexual contact, by the transfer of whole blood or blood products, or from mother to child through the placenta. The virus targets CD4 bearing cells, which include dendritic cells, and macrophages and monocytes of the immune system (Roitt *et al.* 1989).

The life cycle of HIV-1 is shown in Figure 1.2. The initial event of the life cycle is the interaction of gp120 viral protein with the CD4 receptor on a host cell. An additional interaction with a chemokine receptor, including CXCR4 and CCR5 (Clapham and Weiss 1997; Moore 2002), is needed to promote membrane fusion. Membrane fusion is followed by the entry of the viral core into the cytoplasm. The two strands of viral RNA are then reverse transcribed into DNA by the viral RT.

Once synthesised, the viral DNA is transported to the nucleus as part of a preintegration complex composed of integrase, matrix, RT and Vpr, as well as the cellular host protein HMG-I. After active transport to the nucleus, the viral RNA is covalently integrated into the host cell genome by the catalytic activity of the integrase (Turner and Summers 1999). It is believed that a message from within the infected cell, not of viral origin, is necessary for integration (Greene 1993).

After integration, the virus then produces messenger RNAs (mRNAs) that encode the information for the viral proteins using host cell enzymes, including RNA polymerase II. The RNA polymerase has to be activated by two molecular switches located in two regions at each end of the proviral DNA called long terminal repeats, LTRs (see Figure 1.3). The LTRs are prepared for transcription by binding of NF- κ B/rel protein, which function to increase the transcriptional activity of the gene (Greene 1993).

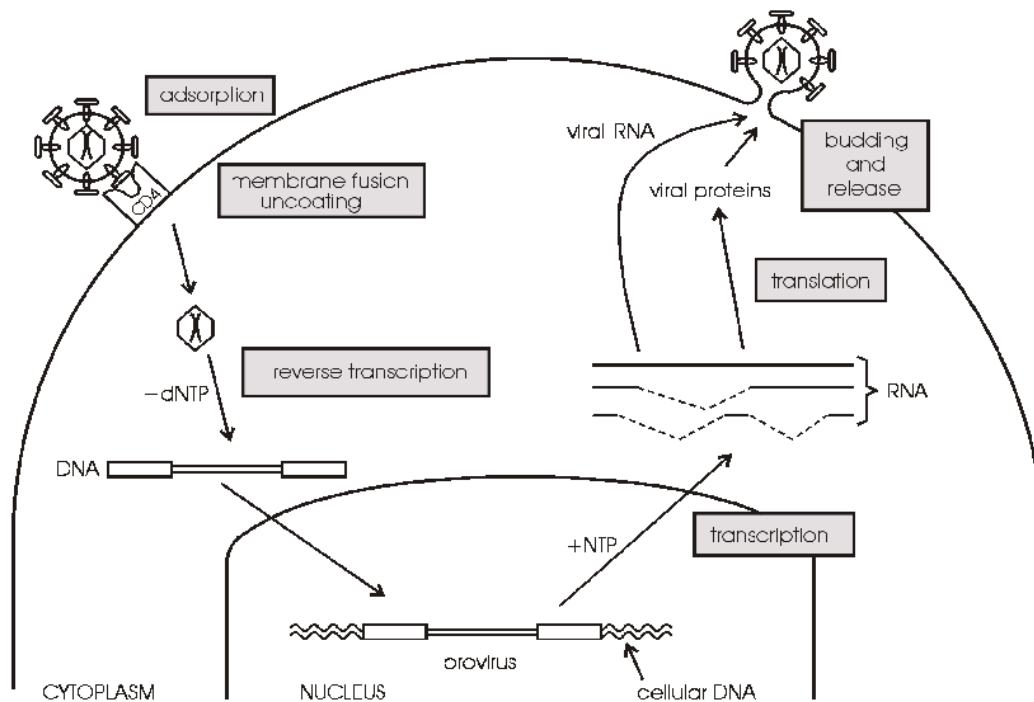


Figure 1.2. Life cycle of HIV-1

Initially, short spliced RNA transcripts of about 2000 bases in length are produced which code for the regulatory proteins of the virus, e.g. Tat, Rev, and Nef. The Tat protein binds to the trans activation region (*tar*) within the viral RNA leading to a phosphorylation of the RNA polymerase by a human kinase (Herrman and Rice 1995). The phosphorylation leads to a 1000 fold acceleration of transcription of the proviral DNA by RNA polymerase (Greene 1993). The Nef protein is believed to function in modifying the host cell to make it more suitable for manufacturing the HIV virion (Greene 1993). The third regulatory protein, Rev, is important for

switching early phase infection into late stage infection by promoting transport of longer RNA transcripts to the cytoplasm (Turner and Summers 1999).

In the late stage two new classes of RNA transcripts appear: long (unspliced) transcripts of 9,200 bases and medium length (singly spliced) of some 4,500 bases. Once the long and medium length transcripts reach the cytoplasm, the cellular machinery begins constructing the components of the new virus. *gag* encodes the core proteins, *pol* encodes reverse transcriptase (RT), protease, integrase, and *env* encodes the two envelope proteins. In addition three other proteins are encoded for by the *vpr*, *vif*, and *vpu* genes (see Figure 1.3). These proteins are important in construction of the new virus. The viral particles then migrate to the outer surface of the cell and obtain their lipid bilayer when budding from the host cell. Subsequent to budding, the polyproteins are cleaved by the viral protease to produce the independent enzymes as well as the matrix, capsid, and nucleocapsid protein leading to the formation of the mature virus.

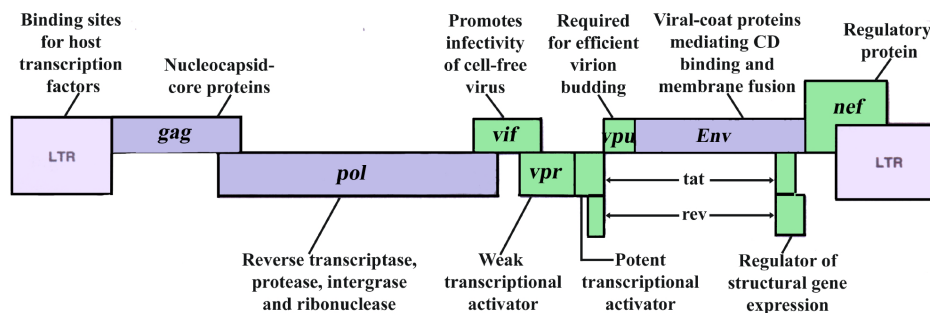


Figure 1.3. The genome of HIV-1.

1.3 Reverse transcription

Reverse transcription is the transcription of an RNA genome into a double stranded DNA molecule. The process was first discovered by Baltimore and Temin in RNA tumour virus and RSV, respectively (Baltimore 1970; Mizutani *et al.* 1970). All previously known transfers of genetic information were from DNA to RNA and hence the synthesis of DNA from RNA was termed reverse transcription.

During the course of reverse transcription, RT interacts with a variety of different nucleic acid substrates:

RNA/RNA - initiation of minus strand DNA synthesis

DNA/RNA - elongation of minus strand synthesis

RNA/DNA - initiation of plus strand DNA synthesis

DNA/DNA - elongation of plus strand synthesis

The process of reverse transcription of the HIV-1 genome occurs mainly in the cytosol of the cell shortly after viral entry. Reverse transcription by RT occurs in a multiprotein complex containing the matrix protein, nucleocapsid protein, and possibly Nef and Vif (Turner and Summers 1999). To perform the process of reverse transcription, RT has a RNA- and DNA-dependent polymerase activity as well as a Ribonuclease H (RNase H) activity to cleave the template RNA during reverse transcription.

Figure 1.4 illustrates the process of reverse transcription. Like other DNA polymerases RT needs a nucleic acid primer carrying a 3'-OH group to initiate DNA synthesis. All retroviruses utilise a host encoded tRNA species as a primer for reverse transcription. HIV-1 RT uses the human tRNA^{Lys3} which is annealed approximately 200 nucleotides downstream of the 5' end of the viral RNA genome at the primer binding site (PBS). The PBS is complementary to the 18 3'-terminal nucleotides of human tRNA^{Lys3}. RT specifically recognises the tRNA/RNA hybrid and reverse transcription is initiated. After initiation, reverse transcription proceeds to the 5'-end of the RNA genome (step 1) (Isel *et al.* 1996). The first product of reverse transcription is the minus strand strong stop DNA [(-)ssDNA] which is produced when RT reaches the 5'-end of the genomic RNA. The (-)ssDNA is transferred to the 3'-end of the genomic RNA to the so called R-region which is complementary to the 3'-end of the newly produced (-)ssDNA. The (-)ssDNA is used to prime the synthesis of the full length (-) strand DNA (step 2). During (-) strand DNA synthesis, the viral RNA is degraded except for two purine rich regions, referred to as the polypurine tract (ppt) and the central polypurine tract (cPPT). Both PPTs are highly resistant to RNase H degradation and are used as primers for (+) strand DNA synthesis (step 3).

The two distinct (+) strand segments formed are called the downstream (D+) and upstream (U+) regions and are initiated from the cPPT and PPT respectively. During synthesis of the U+ DNA the PBS region is regenerated using the tRNA^{Lys3} as a template, which is subsequently cleaved by the RNase H activity of RT. Both PPTs are also removed at this stage by a mechanism which remains unclear. The PBS region generated during (-) strand synthesis is hybridised to that generated from (+) strand synthesis, by a circularisation mechanism (step 4). The (+) strand PBS region is used to elongate the (+) strand DNA to the central termination sequence (CTS), which is located downstream of the cPPT (step 5). Termination of DNA synthesis occurs about 100 nucleotides (nt) downstream of the cPPT. During this process RT must displace the previously synthesised D+ segment in order to terminate DNA synthesis. The full length DNA therefore contains a (+) strand DNA overlap in its centre (step 6). The DNA product also contains a duplicated U3-R-U5 region [termed long terminal repeats (LTRs)] due to the two strand transfer reactions.

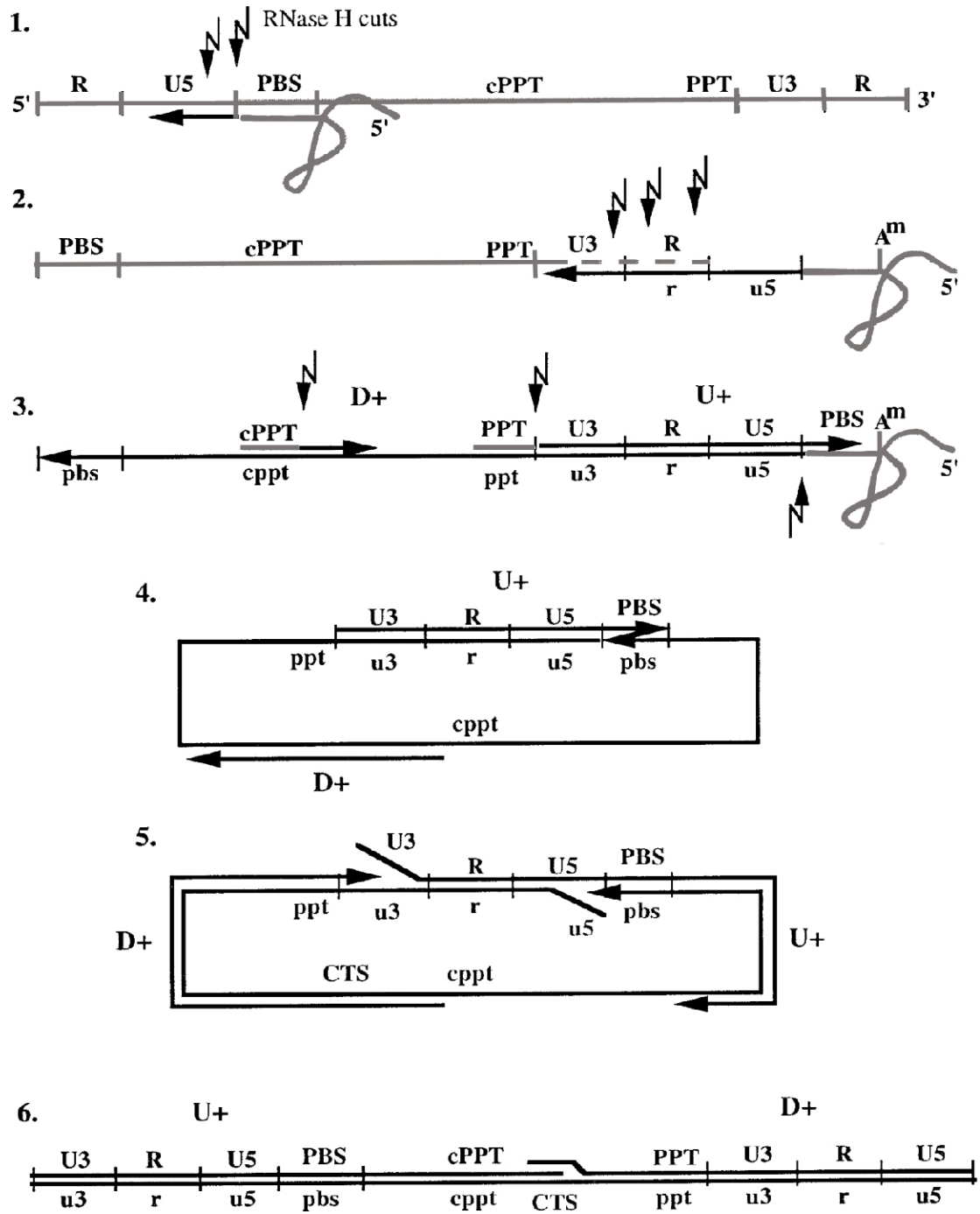


Figure 1.4. The process of reverse transcription (Gotte et al. 1999).

1.4 Structure of HIV-1 RT

HIV-1 RT is a heterodimer composed of two related subunits, a 66 kDa subunit (p66) and a 51 kDa subunit (p51). p51 is generated by the proteolytic cleavage of p66 between residues F440 and Y441 and lacks the RNase H domain. Structural data for HIV-1 RT is available from X-ray crystallography.

The solved structures can be divided into three main classes:

- 1) RT complexed with non-nucleoside RT inhibitors (NNRTI's), e.g. Nevirapine, TIBO, α -APA and HEPT (Das *et al.* 1996; Kohlstaedt *et al.* 1992; Ren *et al.* 1995; Smerdon *et al.* 1994)
- 2) RT:nucleic acid complexes e.g. [RT:DNA/DNA], [RT:RNA:DNA], [RT:RNA (pseudoknot)] and [RT:DNA/DNA:dNTP] (Huang *et al.* 1998; Jacobo-Molina *et al.* 1993; Jaeger *et al.* 1998; Sarafianos *et al.* 2001)
- 3) unliganded RT (Esnouf *et al.* 1995; Hsiou *et al.* 1996; Rodgers *et al.* 1995)

Shown in Figure 1.5 is the ternary complex [RT:DNA/DNA:dNTP] (Huang *et al.* 1998). Based on the resemblance of the p66 subunit to a right hand, the subdomains of RT (p66 and p51) have been designated as fingers, palm and thumb, which together form a “cleft” to grip the nucleic acid substrate. The connection subdomain, which exists in both subunits, links the polymerase domain to the RNase H domain.

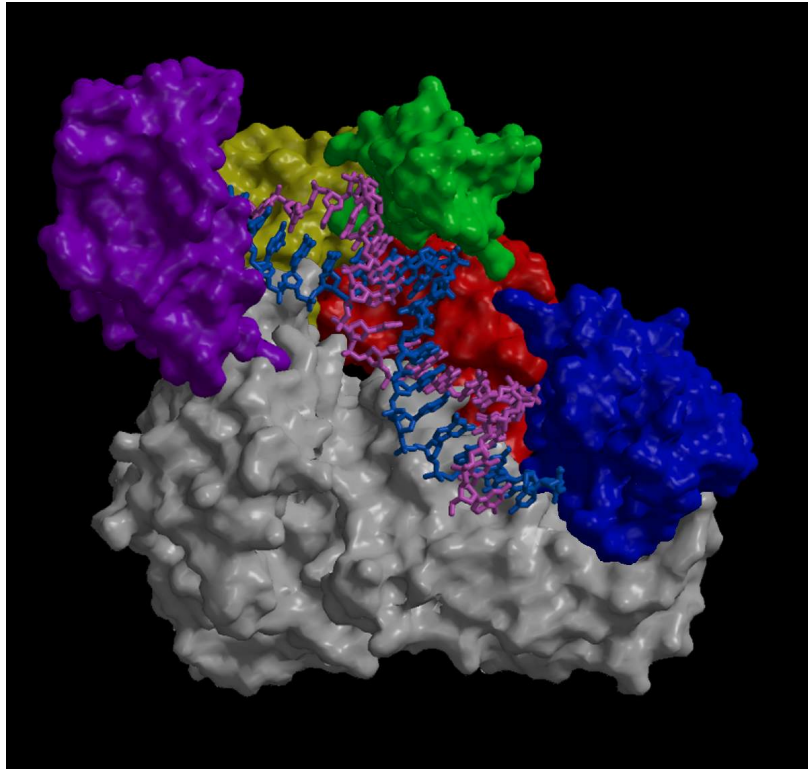


Figure 1.5. The structure of the HIV-1 RT:p/t:dNTP complex. The protein is displayed as a molecular surface with the different regions of p66 being coloured as follows: fingers (purple), palm (yellow), thumb (green), connection (red), and RNase H domain (blue), while p51 is coloured grey. The nucleic acid substrate is shown as a stick representation with the primer coloured purple and the template coloured blue.

HIV-1 RT is only enzymatically active as a dimer (Restle *et al.* 1990; Restle *et al.* 1992). The formation of active RT occurs via a two step mechanism, involving rapid association of the two subunits into an inactive dimer, followed by a slow conformational change yielding the fully active form of the enzyme (Divita *et al.* 1995a). The first interaction between the two subunits occurs via a tryptophan (Trp)-rich hydrophobic cluster located in the connection subdomains. This is followed by a conformational change that stacks the thumb subdomain of p51 onto the RNase H domain of p66 and arranges the fingers subdomain of p51 in the palm subdomain of p66 (Divita *et al.* 1995a; Divita *et al.* 1995b).

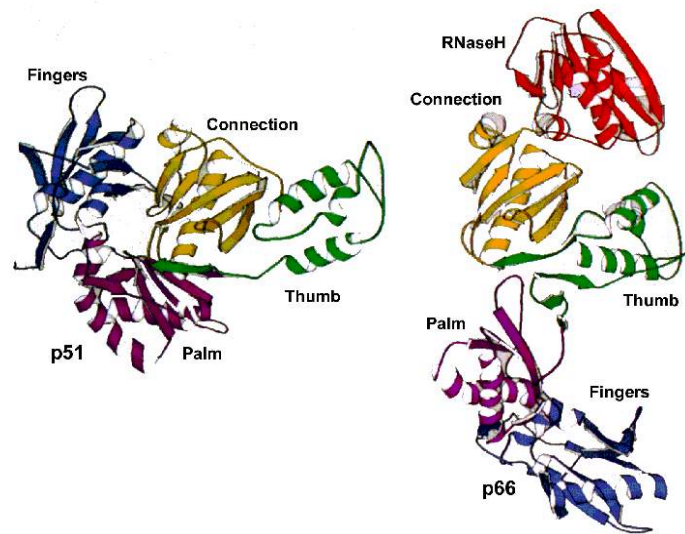


Figure 1.6. Comparison of the domain organisation of the two subunits of the heterodimeric RT. The orientation of the connection domains has been chosen to be identical in the two subunits.

One striking feature of this dimeric enzyme is the highly asymmetric organisation of the subunits (Figure 1.6). Although the subdomains in both p66 and p51 are folded similarly, the subdomains have a totally different spatial arrangement. For p51 the thumb is rotated away from the palm, permitting significant contact with the RNaseH domain. The p51 subunit has no catalytic activity and is thought to stabilise the heterodimer.

The unliganded enzyme adopts a predominantly “closed” conformation, in which the fingers and thumb subdomain of p66 come into close contact (Hsiou *et al.* 1996; Rodgers *et al.* 1995). Recent studies using EPR spectroscopy showed that the unliganded enzyme can exist in two distinct conformations (Kensch *et al.* 2000). One conformation would correspond to that observed by crystallography for the unliganded enzyme. The second conformation seen indicates an open conformation. However, due to the limitations of the methodology, it could only be concluded that the tips of fingers and thumb subdomains are more than 30 Å apart in the open conformation.

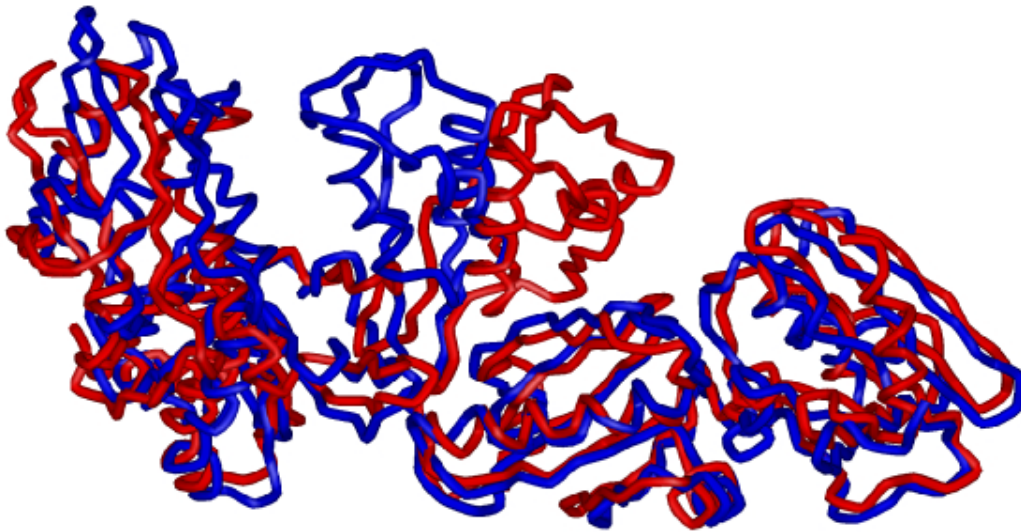


Figure 1.7. The open and closed conformations of p66 from HIV-1 RT. The open conformation is shown in red and the closed conformation in blue.

Comparison of the unliganded RT structures with the DNA-bound RT and the NNRTI-bound RT structures reveals that the p66 thumb subdomain can exhibit two different upright conformations (Figure 1.7). In the RT:p/t complex, the p66 thumb subdomain adopts an upright position that can be described as resulting from a rigid-body rotation of the p66 thumb compared with the thumb position in the unliganded RT. This movement results in an opening of the structure, thus generating a large DNA binding cleft from the N-terminus of the polymerase domain to the C-terminal RNase H active site. The position of the other subdomains remains largely unchanged. The nucleic acid substrate adopts a mixed A'-like and B-form structure when bound to RT. The first 6 base pairs in the polymerase active site adopts an A'-like form, and the remainder of the bound substrate assumes a B-DNA conformation. NNRTI binding induces an additional hinge movement of the p66 thumb near the thumb's knuckle, causing the p66 thumb to adopt a configuration that is even more extended than in the DNA-bound RT structure.

On binding of the incoming nucleotide substrate in the presence of bound p/t the enzyme undergoes a further structural rearrangement in the fingers subdomain (Huang *et al.* 1998). The closure of the fingers forms a pocket that accommodates the incoming dNTP. The physiological significance of the rather large domain

movements is currently unknown. However, it is believed that these conformational changes play an important role during the polymerisation process.

1.5 Substrate binding and nucleotide incorporation by HIV-1 RT

RT catalyses the synthesis of DNA via an ordered mechanism in which the p/t binds prior to the dNTP (Majumdar *et al.* 1988) (Figure 1.8). Polymerisation begins with the binding of a p/t to the unliganded enzyme (RT) to form the RT-p/t complex (step 1), followed by a structural change in either the protein or the p/t (step 2). Nucleotide incorporation into the RT^{*}-p/t complex is initiated by the binding of a dNTP to RT^{*}-p/t to form the RT^{*}-p/t-dNTP complex (step 3). The rate limiting step of polymerisation is the conversion of RT^{*}-p/t-dNTP to an activated complex, RT^{**}-p/t-dNTP (step 4). The cycle is completed by the nucleophilic attack by the 3'-OH primer terminus on the α -phosphate of the dNTP that results in the formation of a phosphodiester bond and removal of the pyrophosphate (PP_i) product (steps 4 and 5). After incorporation of the nucleotide, the polymerase can either dissociate from the p/t and restart the cycle at another 3'-OH primer terminus (distributive mode of polymerisation) or translocate along the elongated DNA product, p₊₁/t, towards the new 3'-terminus (processive mode of polymerisation).

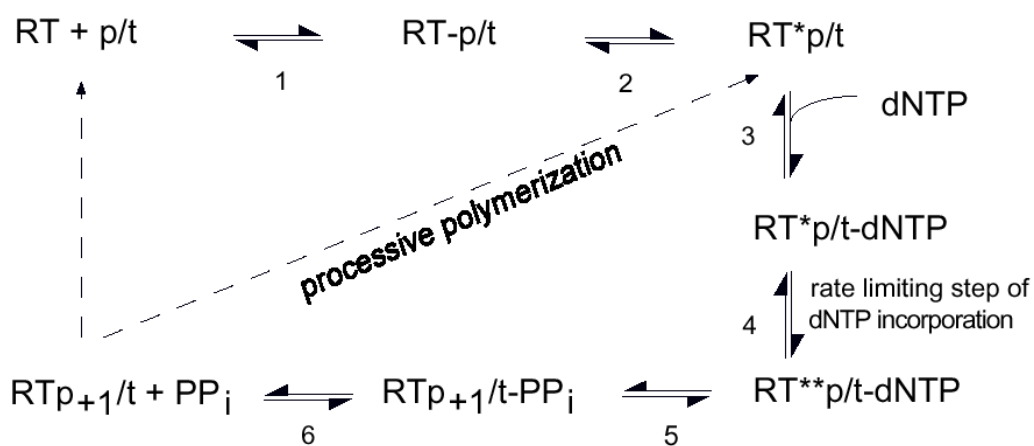


Figure 1.8. Proposed mechanism of nucleotide incorporation by RT.

Previous transient kinetic experiments on p/t binding and nucleotide incorporation by RT showed the existence of three distinct RT:nucleic acid complexes (Figure 1.9). The major evidence for the existence of three complexes was provided by single turnover, single nucleotide incorporation kinetics. By using this method it is possible to indirectly monitor different complexes due to different rates of nucleotide incorporation. The results showed that dNTP incorporation occurs with a biphasic exponential burst followed by a linear phase. The exponential burst consists of a fast phase with a rate of 20-60 s⁻¹ and a slow phase with rates of 0.5-2 s⁻¹. The relative distribution of these two burst amplitudes is defined by the nature of the nucleic acid substrate (DNA/DNA, DNA/RNA) (Wöhrl *et al.* 1999). The [DNA/RNA:RT] complex shows primarily fast incorporation (>80 %) whereas less than 45 % of the [DNA/DNA:RT] complex incorporates dNTPs rapidly.

The fast phase, corresponding to a productive complex (P), is capable of facile nucleotide incorporation and probably corresponds to that observed by x-ray crystallography. For the complex corresponding to the second exponential phase, termed the preproductive complex (PP), the enzyme or substrate must undergo a small conformational change before nucleotide incorporation can occur. In the case of the slow linear phase, it is thought that the p/t substrate is bound to RT in an orientation that does not allow nucleotide incorporation. The third complex was therefore called the dead end complex (DE). Before nucleotide incorporation can occur, the p/t must first dissociate and reassociate in either the P or PP manner before nucleotide incorporation can occur.

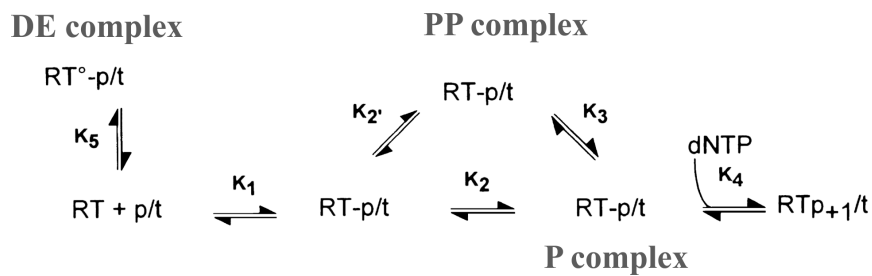


Figure 1.9. Proposed model of RT:nucleic acid interactions. * indicates a structural change occurring in the enzyme or the substrate

1.6 Single molecule spectroscopy

During the last decade, significant advances have been made in the field of single molecule spectroscopy (SMS). Using highly diluted aqueous samples it is possible to observe and characterise individual fluorescently labelled molecules due to different fluorescent properties, such as intensity, lifetime, and anisotropy. Physical, chemical, and biological properties can be measured, which would otherwise be lost in a conventional ensemble measurement. The technique has been successfully applied to the study of motor proteins (Noji *et al.* 1997; Wang *et al.* 1998), enzymes (Lu *et al.* 1998), and ribozymes (Zhuang *et al.* 2000).

Fluorescence resonance energy transfer (FRET) is a spectroscopic process by which energy is passed nonradiatively between molecules over long distances (10-100 Å). The long-range resonance dipole-dipole interaction occurs through space and therefore depends on the distance between the donor and acceptor molecules. The technique has been applied to determine the structure of the Klenow fragment:DNA complexes (Furey *et al.* 1998), and nucleic acid structures, such as the four way DNA junction (Duckett *et al.* 1993) and duplex DNA molecules (Clegg *et al.* 1993). In recent years, the sensitivity of FRET has been extended to the single molecule level (for reviews see (Ha 2001; Weiss 1999; Weiss 2000)). By looking at individual changes of a single molecule it has been possible to monitor conformational dynamics of biomolecules during catalysis (Ha *et al.* 1999) and protein folding (Deniz *et al.*

2000), as well as to distinguish between populations of DNA molecules (Deniz *et al.* 1999) due to differential FRET.

1.7 Aims

Previous crystallographic strategies to determine the structure of RT:nucleic acids have relied on the presence of additional agents to stabilise the complex. In the first strategy, the presence of a bound antibody fragment (Fab) seems to stabilise the nucleic acid in a single conformation. In the second strategy the nucleic acid substrate was crosslinked to the protein in the productive conformation. From previous kinetic studies three modes of interaction were seen between RT and its nucleic acid substrate (see section 1.5). The nature of the binding modes seen could be assigned in terms of different rate constants, but no structural information was obtained. The first aim of this project is to use single pair (sp) FRET to determine structural information on additional RT:p/t complexes not solved by X-ray crystallography.

It was previously demonstrated that formation of active RT is a two step mechanism, involving rapid association of the two subunits into an inactive dimer, followed by a slow conformational change yielding the fully active form of the enzyme (Divita *et al.* 1995a). The first interaction between the two subunits occurs via a trp-rich hydrophobic cluster located in the connection subdomains. This is followed by a conformational change that stacks the thumb subdomain of p51 onto the RNaseH domain of p66 and arranges the finger subdomain of p51 in the palm subdomain of p66 (Divita *et al.* 1995b). It is not presently known which structures the subunits adopt when they are in the monomeric state. Such questions might be important in understanding the driving force and dynamics of this protein/protein interaction. Conventional structural methods like NMR spectroscopy or X-ray crystallography have failed to address the issue since the experimental conditions used, in particular the high protein concentrations employed (μM - mM range as compared subnanomolar range for single molecule measurements), lead to the formation of homodimers (p66/p66 or p51/p51). Choice of the appropriate positions for fluorescent probes should allow these structural changes to be monitored on dimerisation as well as answering questions concerning the conformations of the monomers.

Studies using EPR spectroscopy showed that the enzyme in the substrate free state exists in two distinct conformations (Kensch *et al.* 2000) (see section 1.4). In the presence of a p/t substrate the enzyme adopts an open conformation. On binding of the incoming nucleotide substrate in the presence of p/t bound to the enzyme a structural rearrangement in the fingers domain occurs. A further aim of this work was to develop strategies for the dual labelling of RT in order to determine the rate of interconversion between the open and closed enzyme in the substrate free state.

2. Materials

2.1 Chemicals

chemical	supplier
Acetic acid	Roth
Acrylamide/Bis-acrylamide (29:1) 40%	Roth
Agarose	Sigma
Alexa Fluor [®] 488C5 Maleimide, Na ⁺ salt (A488)	Molecular Probes
Alexa Fluor [®] 594C5 Maleimide, Na ⁺ salt (A594)	Molecular Probes
Alexa Fluor [®] 633C5 Maleimide, Na ⁺ salt (A633)	Molecular Probes
Ammoniumpersulphate (APS)	Serva
Ampicillin (Amp)	Serva
Bromophenol blue	Pharmacia
Chloroform	Baker
Coomassie brilliant blue G250+R-250	Serva
Deoxynucleotide trisphosphates (dNTP)	Roche
Dideoxynucleotide trisphosphates (ddNTP)	Roche
Dimethylsulphoxide (DMSO)	Serva
Dithiothreitol (DTT)	Gerbu
Ethyl diamintetraacetate (EDTA)	Gerbu
Ethidium bromide (EtBr)	Serva
Ethanol	Sigma
Glycerol	Gerbu
Isopropyl- β -D-thiogalactopyranoside (IPTG)	Gerbu
Imidazole	Difco
Kanamycin (Kan)	Pharmacia
β -Mercaptoethanol	Sigma
Sodium chloride (NaCl)	Baker
Sodium dodecylsulphate (SDS)	ICN
Phenol	Roth
Phenyl methyl sulphonyl fluoride (PMSF)	Roth
Poly(rA)/oligo(dT) ₁₂₋₁₈	Pharmacia
Potassium chloride (KCl)	Fluka
Radioactively labelled nucleotides: [³ H]-dTTP	ICN

$[\gamma^{32}]$ -ATP	NEN
N, N, N', N', -tetramethylethylenediamin (TEMED)	Merk
Tetrasodium pyrophosphate (Na ₄ PP _i)	Sigma
Trihydroxymethylaminomethane (TRIS)	Roth
Urea	Baker
Yeast extract	Baker

2.2 Enzymes / Molecular weight markers

Pfu turbo polymerase	Stratagene
Restriction enzymes	Roche, New England Biolabs, Pharmacia
T4 DNA-ligase	Roche
T4 polynucleotide kinase	New England Biolabs
Taq polymerase	New England Biolabs
1kb standard	Gibco-BRL

2.3 Other materials

Diethylaminoethyl (DEAE) paper	Whatmann
FPLC-Sephadex G25 (gel filtration)	Pharmacia
Ni-NTA Sephadex-superflow (affinity)	Qiagen
Streptactin macroprep (affinity)	IBA
DEAE-Sepharose (weak anion exchanger)	Pharmacia
5ml SP-Hitrap (cation exchanger)	Pharmacia
Nap-10 column	Pharmacia
Viva spin filters	Viva Life Science
Millipore spin filters	Millipore
G25 gel filtration spin column	Pharmacia
PCR purification kit	Qiagen
Gel extraction kit	Qiagen
Miniprep spin kit	Qiagen
Midiprep kit	Qiagen

2.4 Bacterial strains

Escherichia coli M15

Nal^S, Str^S, Rif^S, lac, ara, gal, mtl, recA⁻

MPI f. med. Forschung,
Abt. Biophysik,
Heidelberg

2.5 Plasmids

pRT166

Amp^R, HIV-1 RT p66 (isolate BH10)

Vector: pkk 233-2

MPI f. med. Forschung,
Abt. Biophysik,
Heidelberg
Pharmacia (Weiss 1988)

p6HRT51

*Amp^R, HIV1 RT p 6 histidine (his) (6H) 51
(isolate HXB2)*

Vector: pdS56RBSII

Stuart Le Grice
National Cancer Institute
NCI-Frederick,

pDMI.1 (Certa *et al.* 1986)

Kan^R, lac I^Q-Repressor

2.6 Growth Medium and buffers

2.6.1 Growth medium

2xTY* Medium (pH 7.3)

1.6 % (w/v) tryptone
1.0 % (w/v) yeast
0.5 % (w/v) NaCl
1.0 % (v/v) glycerol
1 mM MgCl₂
50 mM sodium hydrogen
phosphate

LB Medium (pH 7.5)

0.5 % (w/v) yeast extract
1.0 % (w/v) tryptone
1.0 % (w/v) NaCl

2.6.2 DNA buffers

Native DNA loading buffer	25 % (v/v) ficoll 0.125 % (w/v) bromophenol blue 0.125 % (w/v) xylene cyanol 10 mM EDTA
Formamide DNA loading buffer	80 % (v/v) Formamide 1 x TBE buffer 0.1 % (w/v) xylene cyanol 0.1 % (w/v) bromophenol blue
TBE buffer (pH 8.3)	90 mM Tris-HCl 90 mM boric acid 2 mM EDTA

2.6.3 Protein gel buffers

Stacking gel buffer (pH 6.8)	0.25 M Tris-HCl 0.2 % (w/v) SDS
Resolving gel buffer (pH 8.8)	0.75 M Tris-HCl 0.2 % (w/v) SDS
Gel running buffer (pH 8.5)	0.25 M Tris-HCl 2 M glycine 1 % (w/v) SDS
protein loading buffer (pH 8.0)	120 mM Tris-HCl 8.5 % (w/v) SDS 35 % (v/v) glycerol 400 mM Dithiothreitol (DTT) 0.05 % (w/v) bromophenol blue

Coomassie stain solution	10 % (v/v) acetic acid 50 % (v/v) methanol 0.1 % (w/v) coomassie brilliant blue G250+R-250
Destain solution	50 % (v/v) methanol 10 % (v/v) acetic acid

2.6.4 Column buffers

Buffer W[Ni] (pH 8.0)	100 mM Tris-HCl
Buffer W[SII] (pH 8.0)	100 mM Tris-HCl 1 mM EDTA
Buffer E [SII] (pH 8.0)	100 mM Tris-HCl 1 mM EDTA 2.5 mM desthiobiotin
Buffer RT-D-10 (A) (pH 7.5)	50 mM Tris-HCl 25 mM NaCl 10 % (v/v) glycerol 1 mM EDTA
Buffer RT-D-10 (B) (pH 7.5)	50 mM Tris-HCl 200 mM NaCl 10 % (v/v) glycerol 1 mM EDTA

2.6.5 RT measurement buffers

RT-Assay buffer (pH 8.0)

- 80 mM KCl
- 50 mM Tris-HCl
- 10 mM MgCl₂
- 1 mM DTT
- 0.05 % (v/v) Triton-X-100

SMS buffer (pH 7.6)

- 10 mM KCl
- 50 mM Tris-HCl
- 10 mM MgCl₂
- 140 μM vitamin C

SMS protein buffer (pH 7.6)

- 500 mM KCl
- 50 mM Tris-HCl
- 10 mM MgCl₂
- 140 μM vitamin C

2.6.6 RT storage buffers

SMS storage buffer (pH 7.6)

- 500 mM Tris-HCl
- 60 mM MgCl₂
- 500 mM KCl

RT-D-15 buffer (pH 7.5)

- 50 mM Tris-HCl
- 25 mM NaCl
- 15 % (v/v) glycerol
- 1 mM EDTA

3 Methods

3.1 Molecular biology

3.1.1 Determination of DNA concentration

To determine the concentration of DNA in solution the absorbance was measured at 260 nm in a quartz cuvette ($d = 1$ cm). An absorbance (Abs) of one corresponds to 50 $\mu\text{g/ml}$ for double stranded DNA (dsDNA), 40 $\mu\text{g/ml}$ for single stranded DNA and RNA, and 30 $\mu\text{g/ml}$ for oligonucleotides.

3.1.2 Phenol extraction to purify DNA

One volume of phenol and chloroform was added to the nucleic acid sample. The solution was mixed and centrifuged at 13,000 rpm for 2 minutes using a benchtop centrifuge. The upper layer was collected and an additional volume of chloroform added to the lower phase. The sample was mixed, centrifuged and the upper phase pooled with that previously collected. The solvent was exchanged to water using a G25 gel filtration spin column (Pharmacia).

3.1.3 Purification of DNA

DNA produced by PCR amplification was purified using the “PCR purification kit” (Qiagen). DNA fragments produced from restriction enzyme digestion were separated by agarose gel electrophoresis and fragment migration monitored using UV light. The band of interest was excised from the gel and purified using the “Gel extraction kit” (Qiagen). Both kits employed are based on the adsorption of DNA to a silica membrane at high salt concentrations, while contaminants such as proteins, and low molecular weight DNA fragments, such as PCR primers, are not retained on the membrane. The bound DNA fragments can then be further washed, in the case of a gel extraction to remove agarose, and eluted using water. The purification of DNA using both kits was performed as per manufacturer instructions.

3.1.4 Restriction digestion of DNA

Restriction enzyme digests of DNA fragments were performed as per manufacturer's instructions. The reaction was stopped by addition of native DNA-loading buffer. Fragments produced by restriction enzymes digestion were separated using agarose gel electrophoresis (section 3.1.3).

3.1.5 Ligation

For ligation 1-10 fmol of linear plasmid DNA was mixed with a 10 fold molar excess of fragment DNA. Ligation was performed in T4 DNA ligase buffer in a volume of 12.5 μ l, using 0.2 units of T4 DNA ligase for 12-16 h at 16 °C.

3.1.6 Production of competent cells for electroporation

1 litre of LB medium supplemented with 25 μ g/ml kanamycin (LB^{Kan}) was inoculated with 10 ml of an overnight culture of M15 cells (containing pDMI.1). Cells were grown to an OD₅₉₅ of 0.5 before harvesting by centrifugation at 4500 g for 15 min at 4 °C. The cells were resuspended in 200 ml of ice cold 5 % glycerol and recentrifuged as above. The same procedure was performed 3 times in order to minimise the salt content. The pellet was then resuspended in 2 ml of 10 % glycerol and portioned into 80 μ l fractions, which were shock frozen using liquid nitrogen. Competent cells were stored at -80 °C.

The transformation rate was determined by transforming the cells with 0.1 ng of pBlue Script (plasmid) and plating on LB agar plates supplemented with 100 μ g/ml Ampicillin. The number of colonies should be larger than 10⁸ /1 μ g plasmid DNA transformed.

3.1.7 Transformation of *E. coli* using electroporation

80 μ l of competent cells of strain M15/pDMI.1 were mixed with 3 μ l of ligation mix and cooled for 1 minute on ice. The cells were then transferred to a precooled electroporation cuvette. The cuvette was pulsed (1500 V, 5 ms) and the cells immediately mixed with 1 ml of LB medium. Cells were grown at 37 °C for 1 h before plating on LB-agar plates supplemented with 25 μ g/ml kanamycin and 100 μ g/ml ampicillin. Cells were grown for 12-16 h at 37 °C before colony selection.

3.1.8 Preparation of plasmid DNA

2ml of LB medium supplemented with a 100 µg/ml ampicillin and 25 µg/ml kanamycin (LB^{Amp/Kan}) were inoculated from glycerol stocks and grown overnight at 37 °C. The overnight culture was used to inoculate (1:1000 dilution) 2 ml or 100 ml LB medium, for a small scale or large scale DNA preparation, respectively. Cells were grown for 12-16 h at 37 °C in an orbital shaker, followed by harvesting by centrifugation. Plasmid DNA was purified using the “Miniprep spin kit” (Qiagen) for a small scale purification, or the “Midiprep kit” (Qiagen) for a large scale purification as per manufacturer’s instructions. Mini-preparation of DNA produced between 10-20 µg plasmid DNA, while midi-preparation yielded between 75-100 µg plasmid DNA.

3.1.9 Production of glycerol stocks

300 µl of autoclaved 50 % glycerol solution was added to 700 µl of cells at an OD₅₉₅ of 0.7. The solution was mixed and cells were stored at –80 °C.

3.1.10 TBE-agarose gel electrophoresis

DNA samples were analysed for length and purity using agarose gel electrophoresis. Depending on the size of the DNA fragment, the agarose concentration was between 0.5 and 1.5 % (w/v). The required amount of agarose was solubilised by heating in TBE buffer. Ethidium bromide was added to a final concentration of 0.006 % (w/v). The gel was poured into the gel casting equipment and allowed to polymerise. Samples were prepared in DNA loading buffer and gels were run horizontally at 90 V using a “Minisub DNA Cell” (Biorad) in TBE buffer until fragment separation was complete. A 1 kb DNA ladder (GibcoBRL) was used as a molecular weight standard.

3.1.11 In vitro mutagenesis by use of the overlap extension polymerase chain reaction

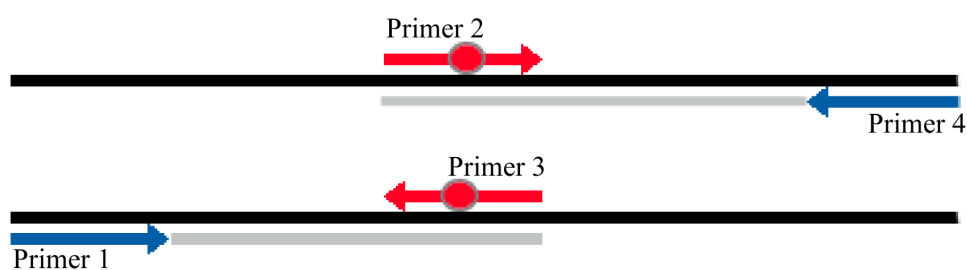
This technique was developed by Ho *et al.* (1989) and further optimised by Herlitz and Koenen (1990). Use of this method allows production of a mutation at any point in the DNA sequence without the need for introduction of new restriction sites, or the close proximity of a restriction site. In this form of PCR, 4 primers are

used and a total of three PCR reactions. In Figure 3.1 is displayed a simple scheme of this method. For the successful use of this method it is important that between primers 1 and 4 (blue) restriction sites are located for cloning and that primers 2 and 3 (red) contain the desired codon change for the introduction of a mutation.

Two PCR reactions were carried out under the conditions described below (see A).

1 fmol	DNA-template (final concentration 2 pM)
3 μ l	10 μ M Primer 1 (final concentration 0.6 μ M)
3 μ l	10 μ M Primer 2 (final concentration 0.6 μ M)
1 μ l	10 mM dNTP mix (final concentration 0.2 mM)
5 μ l	10 x cloned Pfu buffer
1 U	cloned Pfu Polymerase
made up to 50 μ l with sterile water.	

A)



B)

I) Filling the 3' overlap of the mutated DNA strands produced (see A)



II) Amplification of the double stranded DNA

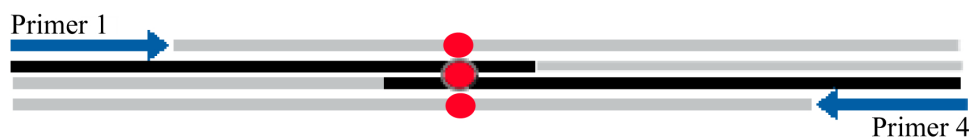


Figure 3.1. Schematic representation of site directed mutagenesis. In the first stage 2 PCRs were carried out in parallel (A) to produce two fragments with an overlap region defined by primers 2 and 3 (red). The fragments were then hybridised and the single stranded region filled up (BI) followed by amplification (BII).

The two resulting PCR-products contain the desired mutation and a short region of overlap defined by the length and position of primers 2 and 3. The PCR fragments were purified by the use of the “PCR purification kit” (Qiagen) and the quantities of the fragments estimated from an agarose gel. 100 fmol of each fragment were added to TE buffer to a final volume of 15 μ l. The double stranded DNA was denatured by heating to 95 °C for 3 min followed by hybridisation of the complementary regions from PCR 1 and PCR 2 at 37 °C for 5 min. 2 μ l of the hybridisation mix was then added to the mixture below.

2 μ l	hybridisation mix (13 fmol) from the first round of PCR
5 μ l	10 x cloned Pfu buffer
1 μ l	10 mM dNTP mix
1 U	Pfu Polymerase
made up to 40 μ l with sterile water	

The mixture was then heated to 72 °C for 10 min in order to allow the Pfu polymerase to fill up the 3'-end of the fragment. The outer flanking oligonucleotides (primers 1 and 4) were then added to a final concentration of 0.6 μ M for the amplification of the desired full length fragment. The fragment was cleaned using a “PCR clean up kit” (Qiagen) before digestion with the appropriate restriction enzymes and ligation into the desired vector. The introduction of the desired mutation was confirmed by DNA sequencing.

For a typical PCR reaction conditions are shown below. The hybridisation temperature differs depending on the melting temperature of the primers used:

denature:	3 min 95 °C
hybridisation:	1 min 46-60 °C
extension:	2 min 72 °C

The cycle was repeated 25 times before a final extension step (72 °C, 10 min) to ensure that the single stranded region was filled up.

3.1.12 Construction of the p6HRT166[SII] Strep-tag II containing vector

In order to purify the p66 subunit of RT, a Strep-tag II affinity tag (Maier *et al.* 1998; Zwicker *et al.* 1999) was added onto the C-terminus of p66. In Figure 3.2 is depicted the strategy for the introduction of the Strep-tag II. A PCR reaction was performed using a primer complementary to a region upstream of the EcoRV site and an antisense primer complementary to the last 30 bases, before the HindIII site of pRT166^{C38S/C280S}, followed by the sequence for the Strep-tag II, and a new HindIII site. The fragment produced contained the C-terminal coding region of p66, the Strep-tag II and restriction sites for cloning.

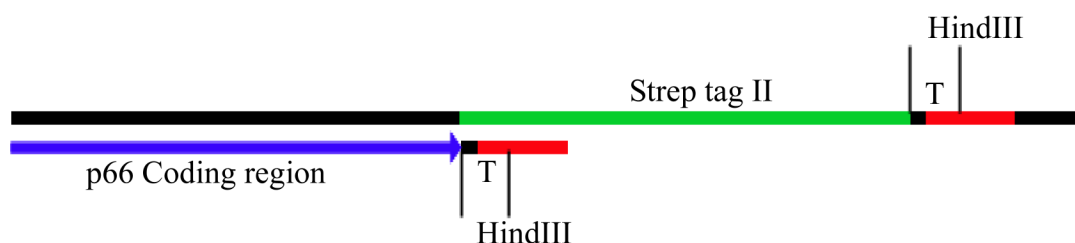


Figure 3.2. Strategy for introduction of Strep-tag II (SII)- The lower strand shows the template, pRT166. The coding sequence is shown in blue, while T indicates the termination site, and the HindIII site is shown in red. The upper strand shows the Strep tag II anti primer, which was used in combination with the EcoRV sense primer to generate the Strep-tag II labelled protein. 30 bases of the Strep-tag II anti primer (black) are complementary to the template. This is followed by the Strep-tag II sequence (green), a new termination site (T), and a new HindIII site.

The fragment was amplified by PCR with Taq polymerase using the following conditions.

2 fmol	DNA-template (final concentration 2 pM)
3 µl	10 µM EcoRV sense (final concentration 0.6 µM)
3 µl	10 µM C-strep-tag II anti (final concentration 0.6 µM)
0.5 µl	10 mM dNTP mix (final concentration 0.2 mM)
1 µl	25 mM MgCl ₂
2.5 µl	10 x Taq polymerase buffer
Made upto 24.5 µl using sterile water.	

The plasmid template was denatured by heating to 95°C for 5 min before the addition of 1 U of *Taq* polymerase. The fragment was amplified using the following PCR cycle conditions:

95°C	30 s
52°C	30 s
72°C	2 min 30 s

The cycle was repeated 25 times before a final extension step (72 °C, 10 min) to ensure the single stranded region was filled up.

The fragment produced and the p6HRT151^{C280S} plasmid were digested using EcoRV and HindIII. The fragment was then ligated into p6HRT151. The new plasmid produced, p6HRT166[SII], Contained the p66 coding sequence with an N-terminal hexa-histidine (His)-tag and a C-terminal Strep-tag II.

3.1.13 Colony screening for plasmid selection

8-16 colonies were picked randomly from LB^{Amp/Kan} agar plates spread with transformants. The colonies were inoculated into LB^{Amp/Kan} medium and grown overnight at 37 °C. Plasmids were then purified using either a “Miniprep Spin kit” (Qiagen) or the cells diluted 1:1000 into 100 ml of fresh LB^{Amp/Kan} medium for overnight growth and purification using the “Midiprep kit” (Qiagen) as per manufacturers instructions. Colonies were screened using restriction enzyme digestion (section 3.1.4) followed by sequencing (see section 3.1.14)

3.1.14 DNA sequencing

For DNA sequencing of cloned fragments the “PrismTM Ready Reaction Big DyeDeoxyTM Terminator cycle sequencing kit” (Applied Biosystems) was used. This method of DNA sequencing is modified from the technique of (Sanger *et al.* 1977). The method is based on the selective termination of DNA synthesis through the incorporation of fluorescently labelled dideoxynucleotides. Each of the four dideoxynucleotides is coupled to a different fluorophore and can therefore be detected due to different fluorescent properties. With the addition of a thermostable polymerase it is possible to sequence DNA using PCR cycling.

For one sequencing reaction (20 μ l), 1 μ g plasmid DNA, 3 pmol primer, and 8 μ l termination mix were mixed and the following PCR conditions applied:

95°C	1 min followed by 25 cycles using the following parameters:
95°C	30 s
46°C	1 min
60°C	4 min

The DNA was precipitated by the addition of ethanol to a final concentration of 70 % and pelleted by centrifugation (13,000 rpm, 4 °C, 45 min) using a benchtop centrifuge. The pellet was washed twice with 70 % ethanol followed by centrifugation (13,000 rpm, 4 °C, 10 min). The pellet was dried and resuspended in 4 μ l of deionised formamide in 8.3 mM EDTA, pH 8.0. The DNA fragments produced by PCR were analysed using an ABI model 373 DNA sequencer (Perkin Elmer, Applied Biosystems) and separated using polyacrylamide gel electrophoresis. The different fluorescence maxima of the fragments were used to distinguish between the four labelled dideoxynucleotides. The gel electrophoresis was performed in house by the “Sequencing and Synthesis” group.

3.1.14 Protein expression

2ml of 2xTY^{*Amp/Kan} were inoculated with glycerol stocks of *E. coli* M15 cells transformed with either the plasmid pRT166, p6HRT51, or p6HRT66[SII] and grown for 8 h at 37 °C. For small scale expression studies, the culture was diluted 1:1000 in 2 ml of fresh medium, and for large scale expression 1:1000 in 500 ml of fresh medium, and grown overnight at 37 °C. The appropriate amount of the preculture was used to inoculate fresh 2xTY^{*Amp/Kan} medium to a starting OD₅₉₅ of 0.09. Amounts and volumes varied depending on the amount of medium to be inoculated. Cells were grown to an OD₅₉₅ of 0.7 before gene induction, p66 or p51, by the addition of isopropyl- β -D-thiogalactopyranoside (IPTG) to a final concentration of 100 μ g/ml. After 3 h the cells were cooled to 4 °C and all subsequent steps were performed at this temperature. Cells were harvested by centrifuging at 10,000 g for 10 min at 4 °C. The bacterial pellets were then weighed and frozen. Levels of protein expression after induction were determined by sodium dodecylsulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

3.2 Protein biochemical methods

3.2.1 Denaturing SDS polyacrylamide gel electrophoresis

SDS-PAGE is a technique used to analyse the quantity, purity, and molecular weights of proteins. During this work 10 % SDS-PAGE gels were used for the analysis of protein purity. For two 10 % gels, the resolving gel was made as follows:

Resolving gel (10 %):	deionised water	2.3 ml
	resolving gel buffer	5.0 ml
	acrylamide stock solution (40%)	2.5 ml

To this was added 50 μ l of 10 % ammonium persulphate (APS) and 8 μ l of N, N, N', N'-Tetramethyl-ethylenediamine (TEMED). The resolving gel was cast using the "Miniprotein II" (Biorad) casting system, leaving 15 mm for the stacking gel.

Stacking gel (4 %):	deionised water	3.2 ml
	stacking gel buffer	2.5 ml
	acrylamide stock solution (40%)	0.5 ml

After the resolving gel was set, 25 μ l of 10 % APS and 8 μ l of TEMED were added to the stacking gel mix, which was then cast above the resolving gel. Protein samples were prepared by adding an equal amount of 2x protein loading buffer and heating for 5 min at 95 °C. Pure HIV-1 RT was used as a marker. Gels were run at 200 V until the bromophenol blue front had entered the buffer solution. SDS-PAGE gels were stained using Coomassie brilliant blue G250+R-250.

Fluorescently labelled proteins (2 μ g) were also analysed for purity using SDS-PAGE. The gel was briefly washed using distilled water before being loaded on the "FLA5000" (Fuji) docking plate. For the visualisation of Alexa Fluor[®] 488 (A488) an argon-ion laser was used with excitation being performed at 473 nm. A488 fluorescence was separated from the excitation source using a 510 nm cut-off filter. For Alexa Fluor[®] 633 (A633) and Cy5, a 635 nm krypton laser was used in combination with a 665 nm cut off filter. For 2 μ g of fluorescently labelled RT, a sensitivity of 400 V on the photomultiplier was used.

3.2.2 Purification of recombinant RT

For the production of heterodimeric RT a 1:15 (w/w) ratio of p51-containing cells and p66-containing cells were mixed. The cell mixture was resuspended in an equal amount (w/v) of buffer W[Ni]+1 M NaCl. Cells were lysed by microfluidisation (Microfluidics corporation) at 650 psi. In addition, sonication was used to break down DNA within the lysate.

The lysate was centrifuged at 46,000 g for 45 min at 4 °C. The supernatant was collected and MgCl₂ was added to a final concentration of 10 mM to facilitate the dimerisation of the two RT subunits. The solution was stirred slowly for 10 h. Recentrifugation of the solution at 46,000 g for 45 min at 4 °C removed any precipitated protein as well as some DNA.

The p51 subunit contains a His-tag which, will under the appropriate conditions, bind to a Ni-NTA column. A Ni-NTA column was equilibrated in buffer W[Ni]+1 M NaCl. The protein was loaded onto the column at a flow rate of 1 ml/min. The column was washed with buffer W[Ni]+ 1 M NaCl until the Abs₂₈₀ level reached within 10 % of the baseline Abs₂₈₀ level for buffer W[Ni]+ 1 M NaCl. Loosely bound proteins were removed by washing with buffer W[Ni]+ 5 mM imidazole. When the Abs₂₈₀ of the flowthrough had dropped to a stable baseline, the protein was eluted using a gradient of 0-100 % buffer W[Ni]+500 mM imidazole over 100 min and fractions of 1 ml were collected (flow rate 1 ml/min). Fractions were analysed on 10 % SDS-PAGE gels and those containing the heterodimer were pooled.

To the pooled fractions an equal volume of RT-D-10 (A) buffer was added and the protein was dialysed for 16 h against RT-D-10 (A). The solution was then loaded onto a pre-equilibrated [in RT-D-10 (A)] 25 ml DEAE-Sepharose column connected in series to a 5 ml SP-Hitrap at a flow rate of 1 ml/min. The column was then washed with 150 ml RT-D-10 (A). The DEAE-Sepharose column was then disconnected and the protein eluted using an NaCl gradient, with the heterodimer eluting at 25 % RT-D-10 (B) (50 mM NaCl). 1.5 ml fractions were collected and analysed using SDS-PAGE. Fractions containing RT were pooled and concentrated using a 15 ml 30 kDa viva spin filter (Viva Life Science), followed by buffer exchange to RT-D-15. The protein concentration was determined by Abs₂₈₀ measurements and calculated using a molar extinction coefficient of

260,000 M⁻¹ cm⁻¹. The protein was aliquoted into 1 mg, 500 µg, and 10 µg portions and shock frozen using liquid nitrogen. The protein was stored at -80 °C.

3.2.3 Purification of the p51[6H] subunit

The p51 subunit was purified in essentially the same manner as the heterodimer, however in the last step the protein was eluted from the SP-Hitrap column using 70 mM NaCl (c.f. 50 mM NaCl for heterodimeric RT). The concentration of the protein was determined from the Abs₂₈₀ and calculated using a molar extinction coefficient of 124,180 M⁻¹ cm⁻¹. The protein was then aliquoted into 1 mg portions and shock frozen using liquid nitrogen. The protein was stored at -80 °C.

3.2.4 Purification of the p66 [6H-SII] subunit

Cells were lysed as described in 3.2.1. However, the cells were resuspended in buffer W[Ni]+1 M NaCl. Phenyl methyl sulphonyl fluoride (PMSF) was added to a final concentration of 0.4 mM in order to protect the p66 subunit from proteolytic degradation. The lysate was centrifuged at 46,000 g for 60 min at 4 °C and the supernatant collected. The supernatant was filtered through a 0.2 µm filtration cap (Nalgene) before loading onto a Ni-NTA column (equilibrated in buffer W[Ni]+ 1 M NaCl+0.4 mM PMSF) at a flow rate of 1 ml/min. The column was washed with buffer W[Ni]+1 M NaCl+0.4 mM PMSF at a flow rate of 1 ml/min until the Abs₂₈₀ level was close to the baseline. The column was further washed with buffer W[Ni]+5 mM imidazole before elution with buffer W[Ni]+200 mM imidazole. The eluate was concentrated to less than 1ml in a 15 ml 10 kDa Viva spin filter (Viva Life Science) before buffer exchange to buffer W[SII] + 0.4 mM PMSF using a Nap-10 gel filtration column (Pharmacia). The protein solution was then loaded onto a 10 ml Streptactin macroprep column at a flow rate of 2 ml/min. The column was washed with buffer W[SII] + 1 M NaCl + 0.4 mM PMSF until a stable OD₂₈₀ value was reached before elution with buffer E. The protein was concentrated on a 15 ml 10 kDa Viva spin filter (Viva Life Science) followed by buffer exchange to RT-D-15. The protein concentration was determined by measuring the Abs₂₈₀ and calculated using a molar extinction coefficient of 136,270 M⁻¹ cm⁻¹. After aliquoting the protein into 1 mg portions it was shock frozen in liquid nitrogen. In order to minimise the time

required to purify p66[6H-SII] no SDS-PAGE analysis is performed during the purification. The protein was stored at -80°C .

3.2.5 Protein labelling

All proteins were labelled (p66/p51[6H], p66[6H-SII] or p51[6H]) with the Alexa dyes Alexa Fluor[®] 488, Alexa Fluor[®] 594 or Alexa Fluor[®] 633 (Molecular Probes), abbreviated as A488, A594 and A633, respectively. Typically, 1-2 mg of protein were labelled via an introduced cysteine with an Alexa maleimide derivative. The reaction was carried out in a volume of 300 μl (150 mM Tris-HCl (pH 7.0)) containing a 5-10 fold molar excess of the dye over the protein to be labelled. The reaction was shaken gently for 2 h at 4°C . Excess dye was removed by gel filtration over a 30 ml sephadex G-25 column in buffer W[Ni]+1 M NaCl. Labelled RT was concentrated initially to 150 μl with a 15 ml 30 kDa Viva spin filter (Viva Life Science) before further concentration, and buffer exchange to RT-D-15 using a 0.5 ml 30 kDa Ultrafree spin filter (Millipore) to a final volume of 30 μl . The individual labelled subunits were concentrated with spin filters with a cut off range of 10 kDa. Abs_{280} and Abs_x were employed to determine the protein concentration and labelling efficiency, where x is the absorbance maxima of the dye used. The protein was then aliquoted into 2 μl portions and the purity and grade of labelling assessed by SDS-PAGE.

For double labelling of the p66 monomer, 1mg of p66[6H-SII]^{W24C/K287C/C38S/C280S} was labelled with A488 and A633 using a 10 fold molar excess of A488 and a 15 fold molar excess of A633. The labelling, purification, and concentration determination were performed as described above.

	A488	A594	A633
Abs [nm]	493	588	622
ϵ [$\text{M}^{-1}\text{cm}^{-1}$]	72,000	96,000	143,000
Emission [nm]	516	612	640

Table 3.1. Spectral properties of dyes used for protein labelling.

3.2.6 Formation of RT from individually labelled subunits

2 mg of p51[6H] labelled with A633 was mixed with 1 mg of p66[6H-SII] labelled with A488 in RT-D-10 to an end volume of 1 ml. DTT, MgCl₂, and PMSF were added to end concentrations of 1 mM, 10 mM, and 0.4 mM respectively. The dimerisation reaction was allowed to proceed overnight at 4 °C with gentle shaking, followed by centrifugation (13,000 rpm, 30 min, 4 °C) in an Eppendorf benchtop centrifuge to remove precipitated protein. The protein solution was taken and the buffer exchanged to buffer W[SII]+0.4 mM PMSF using a Nap-10 gel filtration column (Pharmacia). The protein was then loaded onto a 10 ml Streptactin macroprep column (flow rate 1 ml/min). The column was washed with buffer W[SII]+1 M NaCl ((flow rate 1 ml/min) to remove any remaining free dye, before elution with buffer E. The eluate was concentrated to 150 µl with a 30 kDa Viva-spin filter (Viva Life science) before concentration to 30 µl using a 10 kDa ultrafree spin filter (Millipore), followed by buffer exchange to RT-D-15. The Abs₂₈₀, and Abs's of the two dyes were recorded before aliquoting and shock freezing the protein in liquid nitrogen. The protein was stored at -80 °C.

3.3 Nucleic acids

3.3.1 Oligonucleotides

All Oligonucleotides for experiments using multidimensional fluorescence detection (MFD) spectroscopy were ordered from IBA and were purified to fluorescence correlation spectroscopy (FCS) grade. Oligonucleotides used for mutanegenesis were ordered from MWG Biotech and purified to HPLC grade (see appendix 1)

3.3.2 Dye labelled oligonucleotides

5'-labelled and internally labelled oligonucleotides were coupled to Cy5-NHS-Ester by a (CH₂)₆-NH₂-bridge to the C₅ of the pyrimidine group of thymidine (see appendix 2). The 3'-Cy5 label was introduced via a phosphoamidite labelled base, being the first base incorporated during the synthesis. The 3'-primer terminated oligonucleotide was produced by the dideoxynucleotide (ddNTP) being introduced as the first base of the synthesis.

3.3.3 Radioactive 5'-end labelling of oligonucleotides with $^{32}\text{P}(\text{ATP})$

Radioactive labelling at the 5'-end of a DNA strand is used to determine the purity of nucleic acid substrates and for enzymatic analysis of nucleotide incorporation. Typically 100 pMol of DNA was labelled with 10 μCi of $[\gamma^{32}]\text{-ATP}$ using 10 units of polynucleotide kinase in a volume of 10 μl . The mixture was incubated for 1 h at 37 °C. Oligonucleotides were purified by phenol: chloroform extraction. The upper phase was taken, the volume recorded and a 1 μl (diluted 1:10) sample taken for thin layer chromatography (TLC) (section 3.3.4). The upper phase was applied to a G25 gel filtration spin column (Pharmacia) and centrifuged at 3000 rpm for 2 min. The flow-through was collected, the volume noted and a 1 μl sample taken (diluted 1:5). A further 25 μl of deionised water was added to the G25 gel filtration spin column. A further centrifugation step was carried out (3000 rpm, 1 min) and a 1 μl sample of the flow through taken (diluted 1:5). The diluted samples were taken for analysis of oligonucleotide recovery using TLC (section 3.3.4).

3.3.4 Thin layer chromatography

To determine the concentration of radiolabelled oligonucleotides 2 x 1 μl of the collected samples (see section 3.3.3) were loaded onto a TLC plate (Polygram, cel 300 PEI/UV₂₅₄), and thin layer chromatography was performed in 0.6 M KH_2PO_4 , pH 3.5. After the leading edge of the solvent front had reached the end of the plate, the plate was dried, and then exposed to a photosensitive plate for 5 min. The plate was scanned using a "Phosphoimager" (Biorad) and radioactivity quantified using the program "Molecular Analyst" (Biorad). By comparing the counts per minute (cpm) of the labelled oligonucleotide before and after purification from free $[\gamma^{32}]\text{-ATP}$ it is possible to determine its concentration. Oligonucleotide recovery was typically in the range of 70-90 %.

3.3.5 Denaturing polyacrylamide gel electrophoresis

This method allows the separation of DNA fragments at high resolution (i.e. \pm 1 nucleotide). Nucleotide incorporation rates by RT were analysed with this method using radioactivity as a reporter group. A gel was made up to a thickness of 0.4 mm, and a height of at least 40 cm using the Biorad gel casting system. During this work,

gels were made using a 10 % acrylamide (29:1 acrylamide:bis-acrylamide) containing 7M urea in TBE buffer. Before sample loading, the gel was preheated to a temperature of 56 °C. The nucleic acid samples were mixed 1:1 with formamide DNA loading buffer and the p/t denatured by heating to 95 °C for 3 min. Electrophoresis was performed in TBE buffer with a voltage between 2000 V and 2500 V at a constant temperature of 56 °C for approximately 1 h. The samples were fixed in the gel by incubating in a solution of 10 % acetic acid followed by drying on Whatman GB002 blotting paper in a gel dryer (80 °C, 2 h). Analysis of nucleotide incorporation was either performed by autoradiography or by fluorescence.

3.3.6 Hybridisation of DNA/DNA primer / templates

For the hybridisation of primer to template, equal molar quantities of each were mixed in a buffer containing 50 mM Tris-HCl (pH 7.0) and 20 mM NaCl. The mixture was heated to 95 °C for 30 seconds and then allowed to cool to room temperature over 2 h in a heating block.

3.4 Enzymatic methods

3.4.1 RNA dependent DNA polymerase activity

The RNA-dependent DNA polymerase activity of wild type RT and labelled RTs was determined by monitoring the incorporation of [³H] dTTP into a poly(rA)/oligo(dT)₁₂₋₁₈ p/t. The reaction was initiated by the addition of 5 µl of a 21 nM RT solution to the following mix: 0.2 U/ml poly(rA)/Oligo(dT)₁₂₋₁₈, 50 µM dTTP, and 8.3 pM (=2.1 µCi/ml), [Methyl-³H] dTTP. After 10 min, 2 x 6 µl of the reaction mix were loaded onto 2 cm² DEAE filters and allowed to dry. The filters were washed twice with 2 x SSC buffer (0.3 M NaCl, 30 mM sodium citrate) to remove free dNTPs and enzyme. After a final wash with 100 % ethanol, the papers were allowed to dry.

In order to determine the specific activity of RT, a maximal and minimum count rate was determined (see eq 3.1). 4 x 6 µl of the reaction were incubated without RT and loaded onto DEAE filters. 2 samples were washed (minimal count value [cpm_{min}]) using the above described procedure and 2 samples were not (maximum count value [cpm_{max}]). After drying, the DEAE papers were added to 3 ml scintillation fluid (“*OptiPhase HiSafe*”, Whatman) in a scintillation vial. The counts generated were measured over 60 s using a scintillation counter (Wallac 1409).

$$\text{Specific activity} = \frac{\text{cpm}_{\text{Probe}} - \text{cpm}_{\text{min}}}{\text{cpm}_{\text{max}} - \text{cpm}_{\text{min}}} \cdot \frac{n_{\text{dTTP}}}{m_{\text{RT}}} \quad \text{Eq. 3.1}$$

$\text{cpm}_{\text{probe}}$, cpm_{min} , cpm_{max} , are the count rates for the RT mutant, and the minimum and maximal count rates determined by scintillation counting. n_{dTTP} is the molar amount of dTTP added, while m_{RT} is the relative molecular mass of RT. The units of specific activity are [nmol/(mg.10 min)].

3.4.2 Determination of the rate of single nucleotide incorporation by RT

The principle of the quench flow technique is the rapid mixing of two reaction components followed by quenching of the reaction in a quench flow apparatus after a defined time. For times less than 100 μs in the apparatus used, the reaction time is defined by the length of tubing through which the reactants travel before they are mixed with quenching solution. For longer time points, the reaction is allowed to proceed for a defined time before addition of quenching solution. Experiments are designed so that a series of samples are collected which have reacted for a defined length of time.

Quench flow experiments were performed in a Kintech rapid quench flow apparatus to determine the rate of single nucleotide incorporation into a p/t by wild type RT, as well as labelled RTs. In figure 3.3 is shown the basic components of the quench flow apparatus. "Loading loop A" (yellow) contains a complex of RT and p/t (400 nM RT:200 nM p/t). The enzyme was in two fold excess over the p/t in order to exclude multiple turnovers which would complicate the analysis. "Loading loop B" (red) contains the next dNTP to be incorporated into the p/t at concentration of 200 μM . Syringe A and C (blue) contain the reaction buffer (SMS buffer), which is used to force the reaction components into the reaction loop. Syringe B (green) contains 0.6 % trifluoroacetic acid (TFA), which is used to quench the reaction. A series of samples were collected of different time points from 0.005 s to 20 s and p and p+1 were separated and analysed using denaturing polyacrylamide electrophoresis (section 3.3.5).

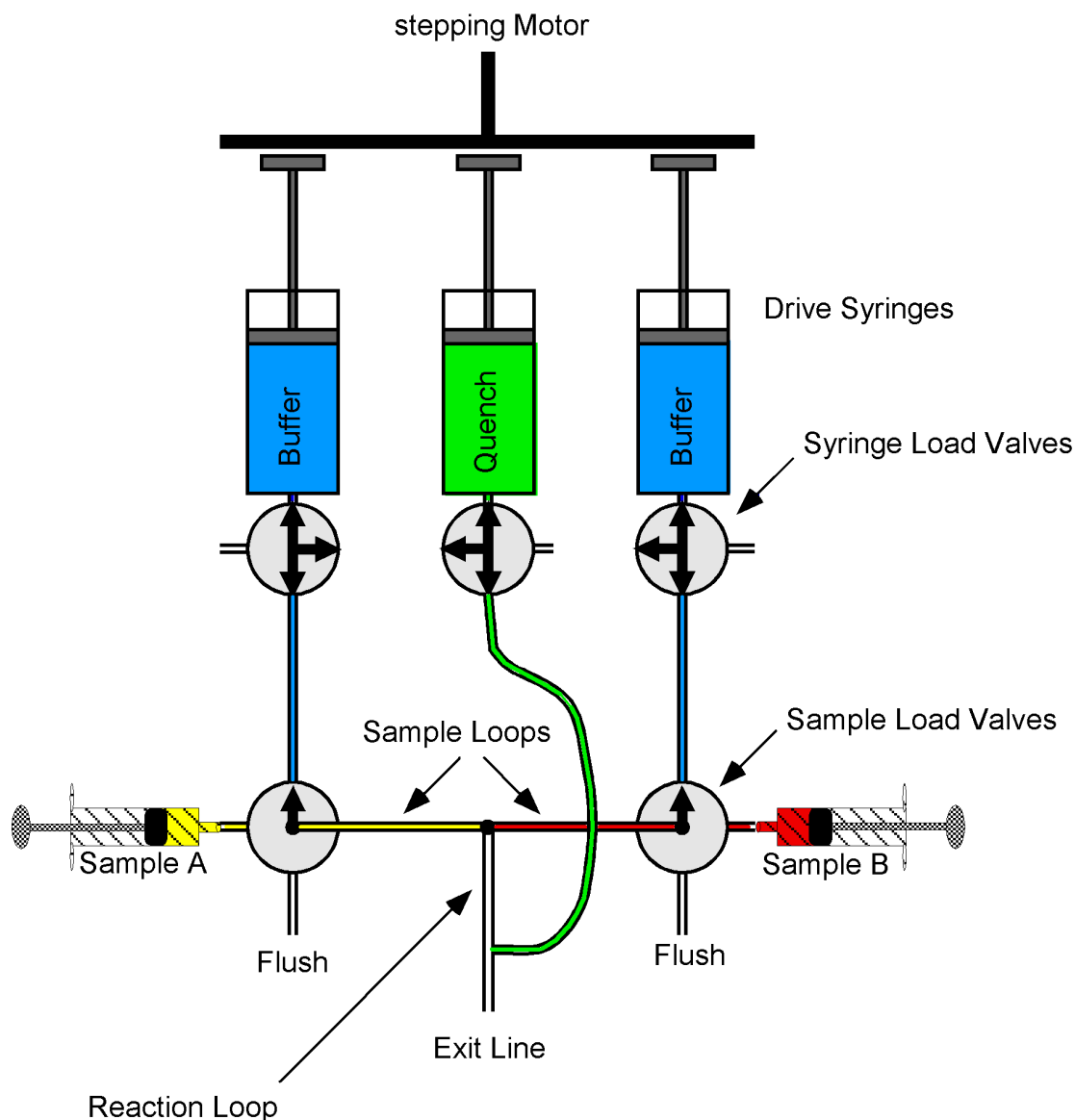


Figure 3.3. Quench flow setup.

3.5 Biophysical methods

3.5.1 Multidimensional single molecule fluorescence spectroscopy using multiparameter fluorescence spectroscopy (MFD)

The spectroscopic technique used during this work for the data acquisition and analysis of the fluorescence properties of single molecules is known as multiparameter fluorescence detection, (MFD) (Eggeling *et al.* 2001) (formerly known as burst-integrated fluorescence-lifetime, BIFL (Eggeling *et al.* 1998)). This

technique was developed by the group of Dr. Claus Seidel at the Max-Planck-Institute for Biophysical Chemistry in Göttingen. (Eggeling 2000). The initial idea of the MFD-technology was proposed by Keller *et al.* (1996). The first application of single molecule-fluorescence spectroscopy was for the identification of fluorescent molecules and the pursuit of observing molecular conformational changes through changes of fluorescence parameters at the single molecule level (Brand *et al.* 1997; Brand 1998; Fries *et al.* 1998; Schaffer *et al.* 1999; Zander *et al.* 1996). Through the simultaneous acquisition of all full fluorescence information (intensity, anisotropy, lifetime and spectral range of the detected photon), it is possible to determine the processes that influence fluorescence. Of major importance to this technique is the method of data analysis. For MFD, in contrast to fluorescence correlation spectroscopy (FCS), no statistical averaging is used over several events or on a heterogeneous sample, as a result of which no raw data is lost. An overview of the possible applications of this technique was described by Eggeling (2000), while the fundamental electronic construction was described by Brand (1998). In the following section, the MFD technique and the steps involved in the data analysis, will be described.

3.5.1.1 Setup / Confocal microscope

The setup used during the single molecule experiments is shown in Figure 3.4. Fluorescent molecules are excited using an actively mode-locked argon-ion laser operated at 496 nm (Sabre, Coherent, Palo Alto, CA, USA; APE, Berlin, Germany). The excitation pulses have a width of 190 ps and a repetition rate of 73 MHz. A dual-band dichroic mirror separates the incoming excitation wavelength from the outgoing fluorescence signal (488/636PC, AHF Analysentechnik, Tübingen, Germany). The objective is a 60 x water immersion lens (Olympus, UPlanApo, 60x, 1.2 N.A.). The mean excitation intensity at 496 nm within the detection volume is 28 kW/cm². The detection volume of approx. 2 f_l is calculated from FCS measurements. It is restricted by a pinhole (diameter 100 μm) and the beam waist ω_0 of 0.56 μm. Further set-up parameters obtained via FCS are the ratio ω_0/z_0 of 4, and the characteristic diffusion time t_D of the fluorescent dye (for Rhodamine 110 and RT^[A488] of 0.3 ms and 1.5 ms, respectively). Here, ω_0 and z_0 are the distances from the centre of the laser beam focus, in the radial and axial direction respectively. The fluorescence signal is divided

into its parallel and perpendicular components with respect to the linear polarised excitation beam by a polarising beam splitter cube (VISHT11 Gsänger, Planegg, Germany), which is then subsequently divided into green and red fluorescence components by further dichroic mirrors (DB: 595 DCRX, AHF) resulting in four signal paths (green parallel and green perpendicular, red parallel and red perpendicular). Band pass filters (IF: HQ535/50 and 730DF140, AHF) are used in the final step to separate the signal in spectral regions from background. Four avalanche photodiodes (APD) (AQR 14, EG & G Vandreuil, Quebec, Canada) are used to monitor the photons. A PC-BIFL-card (SPC 431, Becker & Hickl GmbH Berlin, Germany) registers for each event the arrival time after the laser pulse, the inter-photon time and the detector channel in which the photon arrives corresponding to spectral range and polarisation.

The inset of Figure 3.3 shows a typical multichannel scalar (MCS) trace (bin width 1 ms). A single molecule diffusing through the volume element generates a burst of fluorescence photons. These bursts are identified via small inter-photon times relative to the background of 1-2 kHz. Only bursts which meet certain selection criteria (signal threshold of 42.1 μs for 30 subsequent events, minimum burst size 100 photons) are analysed. Arrival times of the photons in the selected bursts are used to generate fluorescence decay histograms for each signal path. With further analysis, the parameters fluorescence intensity (F), lifetime (τ), anisotropy (r) and their spectral range (λ) are determined.

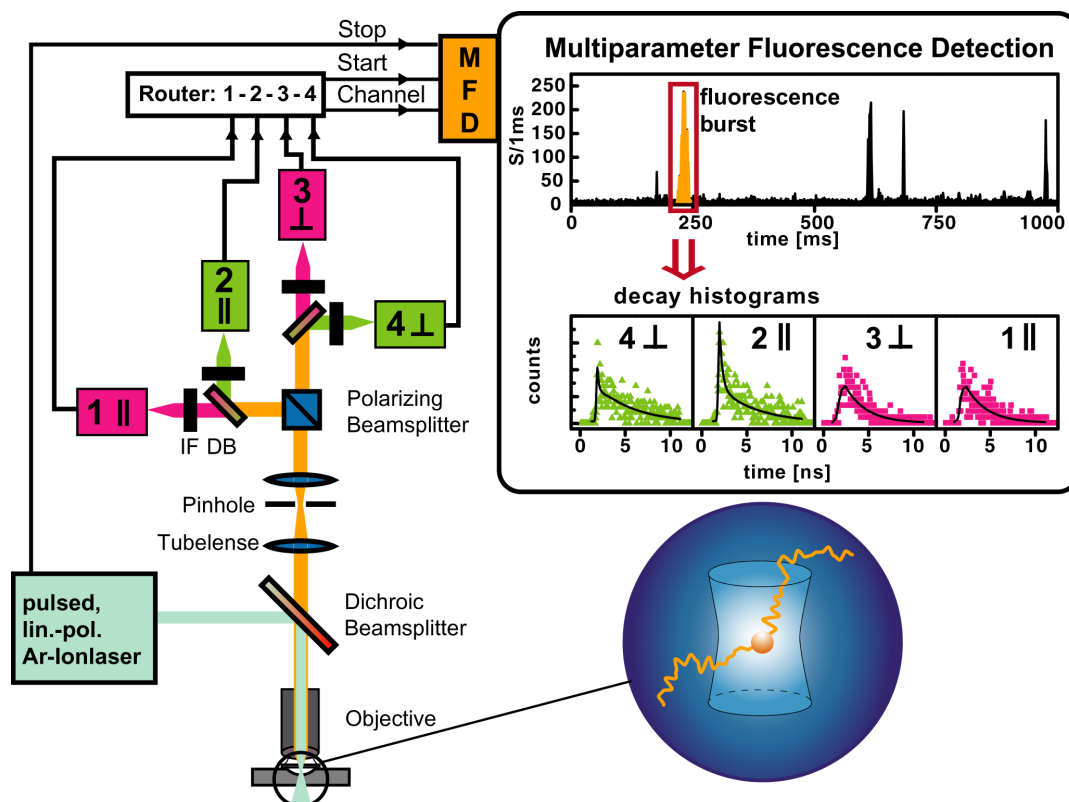


Figure 3.4. Setup for multiparameter fluorescence detection spectroscopy.

3.5.1.2 Data acquisition

The principle of data acquisition using the MFD-method is displayed in Figure 3.5. A regular pulse of 73 MHz of width 190 ps (orange vertical line) is used to excite fluorescent molecules and based on the excitation efficiency of ca. 2 %, only some of these pulse result in the detection of signal photons (small vertical lines). For each detected signal photon, four different parameters are stored: (1) The temporal detection sequence of the photons in form of the event number, $ev \#$ spectral range, λ , of the detected fluorescence (green or red); (2) The channel number, $ch \#$ (1 to 4), in which each signal photon is recorded, to determine the relative polarisation of the signal photon (parallel or perpendicular) with respect to the linear polarisation of the exciting laser to determine the anisotropy, r ; and the wavelength range (above or below 620 nm). (3) The arrival time of the signal photon relative to the incident laser pulse, $\Delta\tau$ (grey horizontal rectangles), measured by time-correlated single-photon counting (TCSPC) to determine the fluorescence lifetime over the picosecond to nanosecond time scale, τ , in an arrival time histogram ; and (4) the time lag, Δt (blue arrows), to the preceding signal photon in the μs - to ms - time range to calculate the

fluorescence intensity, F , on the macroscopic time-scale and to enable a specific and photon-exact fluorescence burst selection.

As illustrated in Figure 3.4, two time scales with different resolution are relevant in the experiment, where a single molecule is excited by a periodic train of laser pulses: the macroscopic time of the experiment measured by the time lag, Δt , to the preceding signal photon with an increment of 50 ns/step; and the arrival time of the signal photon to characterise the fluorescence decay, which is measured relative to the subsequent laser pulse using the inverted mode of TCSPC with a time resolution of 50 ps/channel. For subsequent selective analysis with high time resolution this information is stored together with spectral and polarisation properties for each registered event in a row of a four-dimensional data matrix for the green and red signals.

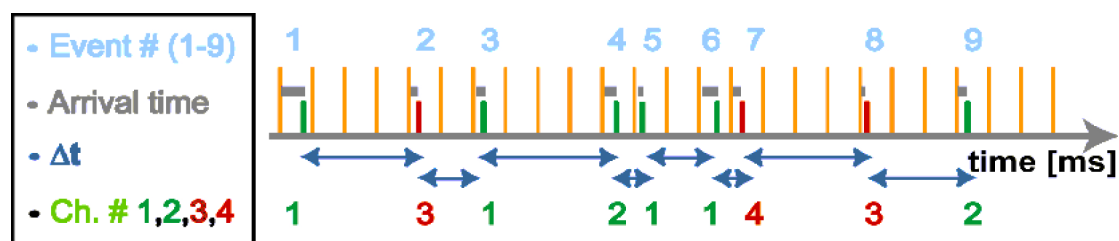


Figure 3.5. Principle of the data acquisition of the MFD-method: for each detected photon (green and red vertical line) four parameters are stored: event #; fluorescence decay time, Dt (grey beams); time interval, Δt , between two detected Photons (blue arrows); and channel number, $ch\#$.

3.5.1.3 Data selection

For data selection, the information from all 4 channels (green parallel and perpendicular, red parallel and perpendicular) is summed to produce a multichannel scalar (MCS). The first step in data selection is to distinguish between fluorescence and background. Therefore, the burst selection is confined to the actual fluorescence photons in the burst, avoiding the inclusion of extraneous background photons. Burst selection is performed by using the interphoton time, Δt , as a selection parameter. The interphoton time, Δt , between two consecutive detected background photons is very large (Figure 3.6, ev# 1-1000), and therefore results in a large value for Δt , whereas fluorescence photons emitted during a dye transit cause a photon burst and thus have a small interphoton time, Δt (Figure 3.6, ev# 1000-2000). A signal with a lower Δt is

classified as a fluorescence burst if a certain number of consecutive photons are below a chosen threshold value. In addition, the Δt -data is smoothed by means of a Lee-filter (Enderlein *et al.* 1997).

Due to Poissonian fluctuations, there is a possibility that the number of photocounts may be below the chosen threshold value even after the beginning or before the end of a burst. This leads to a removal of the edges of fluorescent bursts, which are not recognised as belonging to the same burst. In the burst size distribution, this leads to an underestimation of burst sizes at high burst size numbers, and to a large increase in burst misclassification in the number of bursts with low burst size numbers.

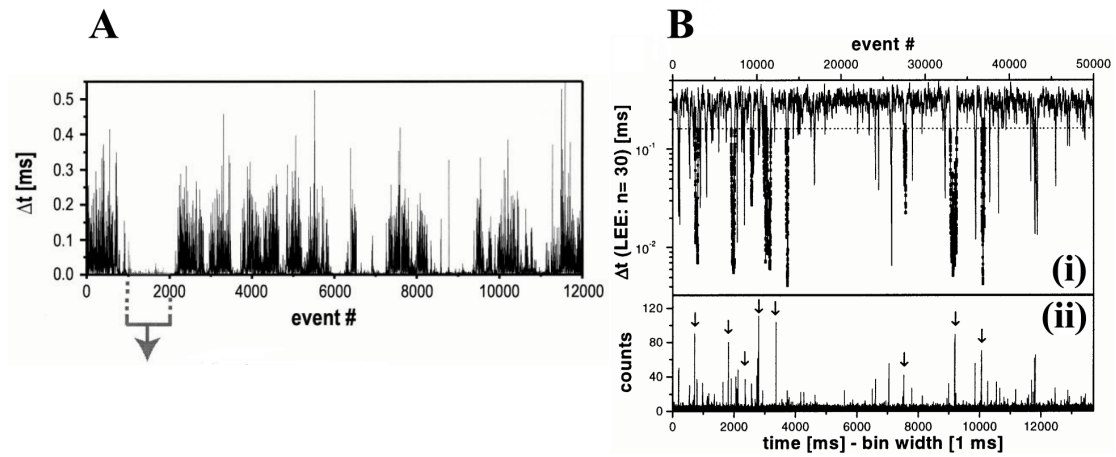


Figure 3.6. Principle of data selection. A) Δt [ms] against event number. For background signal a large value of Δt is seen (ev # 1-1000). A fluorescent burst leads to a reduction of Δt (ev # 1000-2000), and are therefore selected for data analysis (indicated by arrow). Bi) Lee filtered data. The trace shows a mixture of Rh123/EYFP in buffer. Excitation was performed at 502 nm (pulse width of 190 ps, repetition rate 73 MHz, focal excitation irradiance 60 kW cm^{-2}). Interphoton time, Δt , of the smoothed data (lee-filtered: width, $2m+1 = 61$, constant filter parameter, $S_0=10$, and a threshold value, $t_{th}=0.16 \text{ ms}$) against event number. Bursts meeting this criteria are selected (dashed line) for further analysis. Bii) MCS trace with a bin width of 1 ms, calculated from the Δt trace of (Bi), versus the macroscopic measurement time of the experiment. Selected bursts are indicated by arrows (reproduced from (Eggeling *et al.* 2001)).

The Lee-filter of width, m , and a standard deviation, S_0 , used on the discrete data window, k , of time interval, Δt (j), on M photons with the event numbers ($ev\#$) i , is defined by:

$$\overline{\Delta t}(k) = \frac{1}{2m+1} \sum_{j=-m}^m \Delta t(k+j) \quad m < k < M-m \quad \text{eq 3.2}$$

With the variance being defined by:

$$\sigma_k^2 = \frac{1}{2m+1} \sum_{j=-m}^m [\Delta t(k+j) - \overline{\Delta t}(k+j)]^2 \quad 2m < k < M-2m \quad \text{eq 3.3}$$

The filtered data, Δt^* , is given by:

$$\Delta t^*(i) = \overline{\Delta t}(i) + (\Delta t(i) - \overline{\Delta t}(i)) \frac{\sigma_k^2}{\sigma_k^2 + S_0^2} \quad \text{eq 3.4}$$

Where S_0 is a constant filter parameter. The main purpose of the Lee filter is preferential smoothing of the photon shot noise, and therefore resulting in correct classification of fluorescent bursts. A burst is defined by any consecutive number of channels with $\overline{\Delta t}(i) > \Delta t_{th}$, where t_{th} is the predefined threshold value. During this work M was set as 100 photons, while t_{th} was set at 0.0421 ms. Therefore, bursts with a minimal amount of successive photons with a Lee filtered Δt -value under a certain threshold criterion are selected for further analysis.

3.5.1.4 Data analysis

3.5.1.4.1 Time correlated single photon counting to determine fluorescence anisotropy and fluorescent lifetime

By the use of time correlated single photon counting (TCSPC) in combination with the MFD-method, time resolved fluorescence spectroscopy is possible. A detailed description of this method is given by Brand (Brand 1998). The TCSPC technique is essentially a "stopwatch" technique. The excitation pulse is split such that a photodiode is triggered at the same time that the sample is excited. It can be

imagined that a stopwatch is started at this point. When a fluorescence photon is detected by the stopwatch, it is stopped and the "time" measured is recorded.

Of major importance to this method is the time to amplitude converter (TAC), in which the temporal interval between the detection of a signal photon and photons released from the laser pulse are recorded. The laser pulse releases a linearly rising voltage ramp that is stopped by the detection of photons. The voltage value measured is initially converted into a time and then further converted into the decay time, $\Delta\tau$. In the case of a high pulse repetition rate, the TAC can be operated in inverted mode whereby the detection event of a photon is used as a start-signal of the tension ramp, and the next laser pulse serves as a stopping event.

To determine the fluorescence parameters of interest based on this small number of detected fluorescence photons, it is important to note that the registered signal contains fluorescence, j , and variable fractions, γ , of scattered light, s . This scattered light fraction skews the fluorescence decay and anisotropy. Furthermore, refraction by high numerical aperture microscope objective lenses changes the linear polarisation orientation of the excitation light and fluorescence emission. Therefore, an analytical method developed by Koshioka *et al.* (1995) is applied to describe the imaged fluorescence, j_{\parallel} and j_{\perp} , by the use of two correction factors, l_1 and l_2 . The x-axis is defined as parallel and the y-axis as perpendicular to the polarisation of the exciting laser, the x-y plane being coincident with the sample plane. Thus, the fluorescence signal, j_x , collected parallel to the x-axis, which is registered by detector 1, is mixed with y- and z-axial polarised components. Similar considerations hold for detector 2, which monitors the fluorescence, j_y , parallel to the y-axis. A factor, $G = E_{\perp} / E_{\parallel}$, must also be used to correct for the slightly different detection efficiencies, E_{\parallel} and E_{\perp} , of the two channels. The data in the fluorescence arrival time histograms are accumulated in k channels with a total arrival time window, $T = 12.5$ ns. Therefore, the continuous time, t , in the signal decay patterns is replaced by the time, $t_i = iT/k$, of the channel, i , and counted events, $c_{x,i}$ and $c_{y,i}$, of channel, i . Two signal decay histograms are measured for the parallel and perpendicular channels.

Defining an effective scatter-polarisation ratio, α , between the perpendicular and parallel polarised scattered light components, s_y and s_x , the absolute value of the anisotropy, r_s , corrected for the scatter contributions is given by Eq. 3.5.

$$r_S = \frac{G \left(\sum_{v=1}^k c_{x,v} \right) - \left(\sum_{v=1}^k c_{y,v} \right) - \gamma \left(1 - \frac{3}{2} \alpha \right) c_{x,y,z}^{\Sigma}}{(1-3l_2)G \left(\sum_{v=1}^k c_{x,v} \right) + (2-3l_1) \left(\sum_{v=1}^k c_{y,v} \right) - \gamma \left(1 - 3l_2 - \frac{3}{2} \alpha (l_1 - 2l_2) \right) c_{x,y,z}^{\Sigma}} \quad \text{eq 3.5}$$

with

$$c_{x,y,z}^{\Sigma} = G \left(\sum_{v=1}^k c_{x,v} \right) + 2 \left(\sum_{v=1}^k c_{y,v} \right)$$

$$\gamma = (G s_x + 2 s_y) / (G(s_x + j_x) + 2(s_y + j_y))$$

$$\alpha = 2 s_y / (G s_x + 2 s_y)$$

There are two ways to obtain the rotational correlation time, ρ : (1) Direct global maximum likelihood estimator (MLE) analysis of the arrival time histograms of the two detection channels, and (2) computation by the Perrin-equation. Within the spherical rotator model, the Perrin-equation (Eq. 3.6) allows a molecular description of the steady-state anisotropy, r , by linking r to characteristic dynamic fluorescence properties described by lifetime, τ , and rotational correlation time, ρ , with the anisotropy, $r(0)$, at time zero.

$$r = \frac{r(0)}{1 + \tau/\rho} \quad \text{Eq 3.6}$$

The decay time, $\Delta\tau$, of the detected photons is used to determine the fluorescent lifetime, τ . Arrival time histograms with $K=256$ channels and a total time window T (12.5 ns) are generated from $\Delta\tau$ and are used to determine the characteristic decay time, i.e. the fluorescence lifetime, τ . Due to the limited number of photons recorded, the data are fitted using a pattern recognition technique based on a MLE which has the lowest misclassification probability for the low number of events used here (Brand *et al.* 1997; Brand 1998; Enderlein *et al.* 1997; Maus *et al.* 2001; Sauer *et al.* 1998; Zander *et al.* 1996). Depending on the macroscopic time, the MLE determines two variables for each signal decay pattern $Mi(\tau, T, k, \gamma_B, B)$ in the channel i : the appropriate fluorescence lifetime τ of the fluorescence, F , and the fraction γ_B of background signal B due to Raman scattering (Eq. 3.7). The background signal, B , caused by Raman scattering is identical with the instrument response function, PB , which can be experimentally determined using pure solvent.

$$M_i(\tau, T, k, \gamma_B, B) = \gamma_B \frac{B_i(T, k)}{\sum_{j=1}^k B_j(T, k)} + (1 - \gamma_B) \frac{F_i(\tau, T, k)}{\sum_{j=1}^k F_j(\tau, T, k)} \quad \text{eq 3.7}$$

The quality of the lifetimes determined is assessed on the basis of a minimum reduced $2I_r^*$, using the different fluorescence lifetime probabilities $M_i(\tau)$ (Kullback-Leibler minimum reduced $2I_r^*$) (Kullback 1959). The reduced $2I_r^*$ serves as an MLE quality fit parameter, and serves to reduce the variation of the τ and γ_B (Brand *et al.* 1997; Brand 1998; Eggeling 2000; Zander *et al.* 1996). $2I_r^*$ is defined by the following equation, where C_i is the count in the i^{th} channel and C are the total counts. PB is the instrument response function, and is equivalent to B.

$$2I_r^* = \frac{2}{k-3} \sum_{i=1}^k c_i \ln \left(\frac{c_i}{C M_i(\tau, T, k, \gamma_B, PB)} \right) \quad \text{eq 3.8}$$

the Standard deviation, $\sigma(\tau)$, of the determined fluorescence lifetime, τ , is defined by the following dependency, where N is the number of photons.

$$\sigma(\tau) \propto \frac{\tau}{\sqrt{N}} \quad \text{eq 3.9}$$

3.5.1.4.4 Multi-Channel-Scalar (MCS) and fluorescence intensity determination

A multichannel scalar (MCS) records the count rate of events as a function of time, yielding the fluorescence count rate, S . The fluorescence events selected using the Δt and the Lee-filter are averaged for a certain time, in our case 1 ms. The number of counts/ms can then be plotted against time. The fluorescence signal intensity, S , of the selected bursts is calculated simply by dividing the number of photons by the total duration of the burst.

3.5.1.5 Fluorescence resonance energy transfer (FRET)

Fluorescence resonance energy transfer (FRET) is a spectroscopic process by which energy is passed nonradiatively between molecules over long distances (10-100 Å). The “donor” molecule (D), which must be fluorescent, absorbs a photon and transfers the energy nonradiatively to the “acceptor” (A), which need not be fluorescent. The transfer of energy leads to a reduction in the fluorescence intensity, F , and the lifetime, τ , of the donor due to the additional decay pathway in the presence of an acceptor. The long range resonance dipole-dipole interaction occurs through space and therefore depends on the distance between the donor and acceptor molecules. The relationship between energy transfer and the distance separating the donor and acceptor is given by the Förster relationship (eq 3.10, and Figure 3.6):

$$E = \frac{1}{1 + (R/R_0)^6} \quad \text{Eq 3.10}$$

The Förster radius, R_0 is the distance at which energy transfer is 50% and is dependent on the donor-acceptor pair used. The following relationship defines the Förster radius

$$R_0 [\text{Å}] = 9.78 \times 10^3 \cdot (\Phi_D \cdot \kappa^2 \cdot n^{-4} \cdot J)^{1/6} \quad \text{Eq 3.11}$$

The Förster radius is dependent on the donor quantum yield, Φ_D , the refractive index of the medium, n , the overlap integral between the donor emission and the acceptor absorbance, J , and the orientation factor, κ^2 . All of these factors effect the energy transfer and thus the measurement of distance between the donor and acceptor. The orientation factor, κ^2 , is the source of most errors in distance measurements. κ^2 varies between 0 and 4, depending on the orientation of the fluorophores. It is generally assumed that the orientation of the donor-acceptor pair is random corresponding to an average value of κ^2 of 2/3

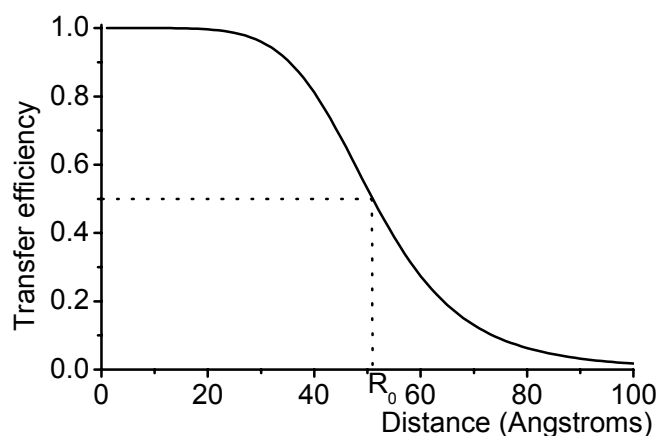


Figure 3.7 Relationship between energy transfer due to FRET and distance. The curve is calculated from equation 3.10, with a Förster radius of 51 Å, the calculated value for the dye pair A488-Cy5 (Berger 2002).

The relationship between transfer efficiency and donor-acceptor separation is shown graphically in Figure 3.7. From this relationship, it is clear that the achievable distance resolution by this technique is $(1 \pm 0.5) R_0$. The standard Förster radius for the donor-acceptor pair A488-Cy5 is 51 Å (Berger 2002). The rate of energy transfer, E , and hence the distance, can be calculated by a variety of methods (Clegg 1992; Selvin 1995):

- 1) From the decrease in fluorescence intensities of the donor, or increases acceptor fluorescence.
- 2) From the decrease in the donor lifetime.
- 3) From changes in fluorescence anisotropy of the donor and acceptor.

In this work the energy transfer, and subsequently the distance, was determined from the fluorescence intensity ratio of donor/acceptor fluorescence, F_D/F_A ; and the donor lifetime, τ_D .

$$E = \left[1 + \frac{F_D}{F_A} \frac{a_{trans} \Phi_A}{\Phi_D} \right]^{-1} \quad \text{eq 3.12}$$

$$E = 1 - \frac{\tau_{D(A)}}{\tau_{D(0)}} \quad \text{eq 3.13}$$

The fluorescence intensity signal of a burst, S , is calculated by dividing the number of photons by the total duration of the burst. The signal must be corrected for background counts, B , detector efficiencies, g , and spectral crosstalk, α , in order to obtain the donor and acceptor fluorescence, F_D and F_A respectively.

$$F_D = \frac{S_G - B_G}{g_G} \quad \text{eq 3.14}$$

$$F_A = \frac{(S_R - B_R) - \alpha(S_G - B_G)}{g_R} \quad \text{eq 3.15}$$

It follows that:

$$\frac{F_D}{F_A} = \frac{g_R F_G}{g_G F_R} = \frac{g_R}{g_G} \frac{S_G - B_G}{(S_R - B_R) - \alpha(S_G - B_G)} \quad \text{eq 3.16}$$

The subscripts G and R denote the green and red signals. By combining equations eq 3.10 and eq 3.12, the separation of two dyes as a function of their intensities can be determined. For intensity measurements, the fluorescence quantum yields of the acceptor, Φ_A , and donor, Φ_D , must be known. Additionally, the factor α_{trans} must be introduced to account for the fraction of time the acceptor dye, Cy5, is in a non-fluorescent conformation.

$$R_{DA} = R_{0r} \left(a_{\text{trans}} \Phi_{FA} \frac{F_D}{F_A} \right)^{1/6} \quad \text{eq 3.17}$$

$$\text{where } R_{0r} = 9780 \left(\frac{J(\lambda) \kappa^2}{n^4} \right)^{1/6} \quad \text{eq 3.18}$$

By determining the distance using the ratio of F_D/F_A , the quantum yield of the donor can be disregarded since changes in F_D/F_A are independent of the donor quantum yield, Φ_D . Therefore a ‘reduced Förster radius’, R_{0r} can be defined (eq 3.18) which differs from R_0 by being independent of Φ_D . For these experiments, $\Phi_A = 0.4$ and $R_{0r} = 53 \text{ \AA}$. The distance distribution in the calculated ranges is dominated by the shot noise in the photon counting, and therefore follows Poisson statistics. By following error propagation rules, R_{DA} will have a Gaussian distribution, with a standard deviation explicitly given by:

$$\sigma(R_{DA}) = \frac{1}{6} R_{0r} \left[\frac{g_R F_G}{g_G F_R} a_{trans} \Phi_{FA} \right]^{1/6} \left[\left(\frac{1}{F_G} + \frac{\alpha}{F_R} \right)^2 (S_G + B_G) + \left(\frac{1}{F_R} \right)^2 (S_R + B_R) \right]^{1/2} \quad \text{eq 3.19}$$

Equations 3.17 provides the empirical formula for determining the separation between two dyes based on intensity measurement. However, the quantum yields, spectral crosstalk, and detector efficiencies must be determined reliably. From equations 3.17 and 3.13 the relationship between the fitted $\tau_{D(A)}$ values and R_{DA} can be determined.

$$R_{DA} = R_{0r} \left(\frac{\tau_{D(0)} \tau_{D(A)} k_{FD}}{\tau_{D(0)} - \tau_{D(A)}} \right)^{1/6} \quad \text{eq 3.20}$$

k_{FD} is the fluorescence decay rate of the donor, and has the value $1.6 \times 10^8 \text{ s}^{-1}$. For $\tau_{D(0)}$, we use the value 4.1 ns.

The fitting procedure’s ability to converge on a single lifetime value is highly dependent on the number of photons being analysed (section 3.5.1.4.1). For this reason, it is more reliable to use equation 3.17 to determine separations. However, by plotting R_{DA} versus $\tau_{D(A)}$ on the same graph, information about the dye states, which can effect the analysis of distances, can be determined.

If the dyes are known to be rotationally mobile over the lifetime of the excited state, it is possible to use the value 2/3 for κ^2 obtained by integrating over all the possible orientations. The mobility of the dye is described by the rotational correlation time, ρ , which can be determined indirectly from the dye’s anisotropy, r . Anisotropy

is the degree of polarisation in the fluorescence signal, and is dependent upon both ρ and τ as shown in the Perrin equation (eq 3.6)

The initial anisotropy, r_0 , accounts for inherent depolarisation due to misalignments between the laser, the excitation dipole, and the emission dipole. MFD records the polarisation of each photon for the determination of r for each molecule (see section 3.5.1.4.1). By plotting r versus τ histograms, and overlaying eq 3.6 for various values of ρ , it is possible to determine the rotational mobility of the dye. For values of ρ less than τ , the value of $2/3$ for κ^2 can be justified.

3.5.1.6 Measurement chamber preparation

In order to suppress protein absorption, cover slides were coated with κ -casein (Sigma). A standing droplet of 200 μ l of a 1 mg/ml κ -casein solution was added to the coverslide and left for 10 min before removal by aspiration. The contact area was washed by pipetting 200 μ l of bi-distilled water onto this area followed by removal. The process was repeated 10 times before the coverslide was stored in the water saturated atmosphere of the measurement chamber.

3.5.1.7 Sample preparation

10 μ M protein solutions were made up in SMS storage buffer and stored at 4 °C. Before the measurements, a “stock” complex solution was made at a concentration of 250 nM protein and 1 μ M p/t. This was equilibrated on ice for at least 30 min. Immediately before a measurement, the stock solution was diluted rapidly in the measurement buffer 1:10 followed by a further 1:10 dilution before a final 1:50 dilution to a final concentration of 50 pM protein and 200 pM p/t. A droplet of the sample solution (50 μ l) was applied to a κ -casein coated cover slip which formed the bottom of a closed chamber with a water-saturated atmosphere. Measurements were performed at 25 °C for up to 1 h.

For measurements using the double labelled RT dimer and monomeric p66, 10 μ M protein solutions were made as above. However the protein was diluted and measured in SMS protein buffer. In the last dilution step the solutions were diluted 1:12.5 to give a final protein concentration of 200 pM.

4. Results and Discussion

4.1 Protein:nucleic acid interactions

4.1.1 Cloning, expression, and purification of HIV-1 RT

In HIV-1, p66 and p51 are encoded for by the same gene, *pol*. For the viral production of RT, the p51 subunit is formed by the cleavage of a protease sensitive region within p66 by the viral protease. For the labelling of RT with thiol reactive reporter groups, cysteine residues had to be introduced in a site specific manner. Previous work (Kensch 2000) showed that mutating the 3 solvent accessible cysteine of RT (p66-C38, C280; p51-C280) to serine residues had no effect on activity. In addition, the cysteine at position 38 in p51 is inaccessible in the heterodimeric enzyme. Cysteine codons were introduced site specifically into either the p66 or p51[6H] coding sequence of plasmids pRT166 and p6HRT51 using overlap extension PCR. Cysteine can therefore be introduced into each subunit independently and the new mutant proteins expressed in *E. coli*. Co-homogenisation of the p66 and p51[6H] expressing *E. coli* cells produces the heterodimeric enzyme, which can be purified by metal chelate chromatography via the N-terminal His-tag of p51.

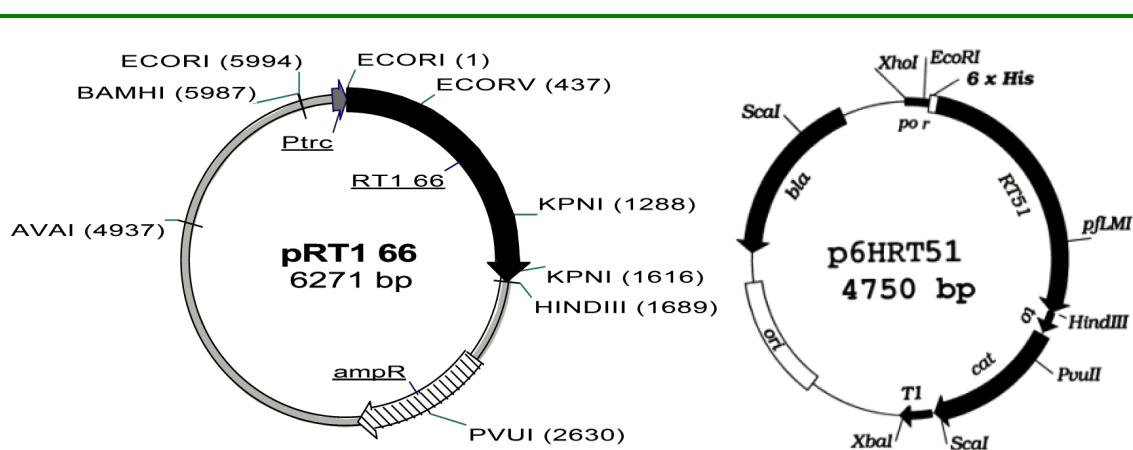


Figure 4.1. Plasmids used for mutagenesis during this work.

4.1.2 Mutagenesis using overlap extension PCR

Cysteine codons were introduced into the RT coding regions of the expression vectors pRT166 and p6HRT51 (Figure 4.1) by overlap extension PCR (section 3.1.11). The following example of this technique is shown, p6HRT51^{C280S/} → p6HRT51^{C280S/K281C}.

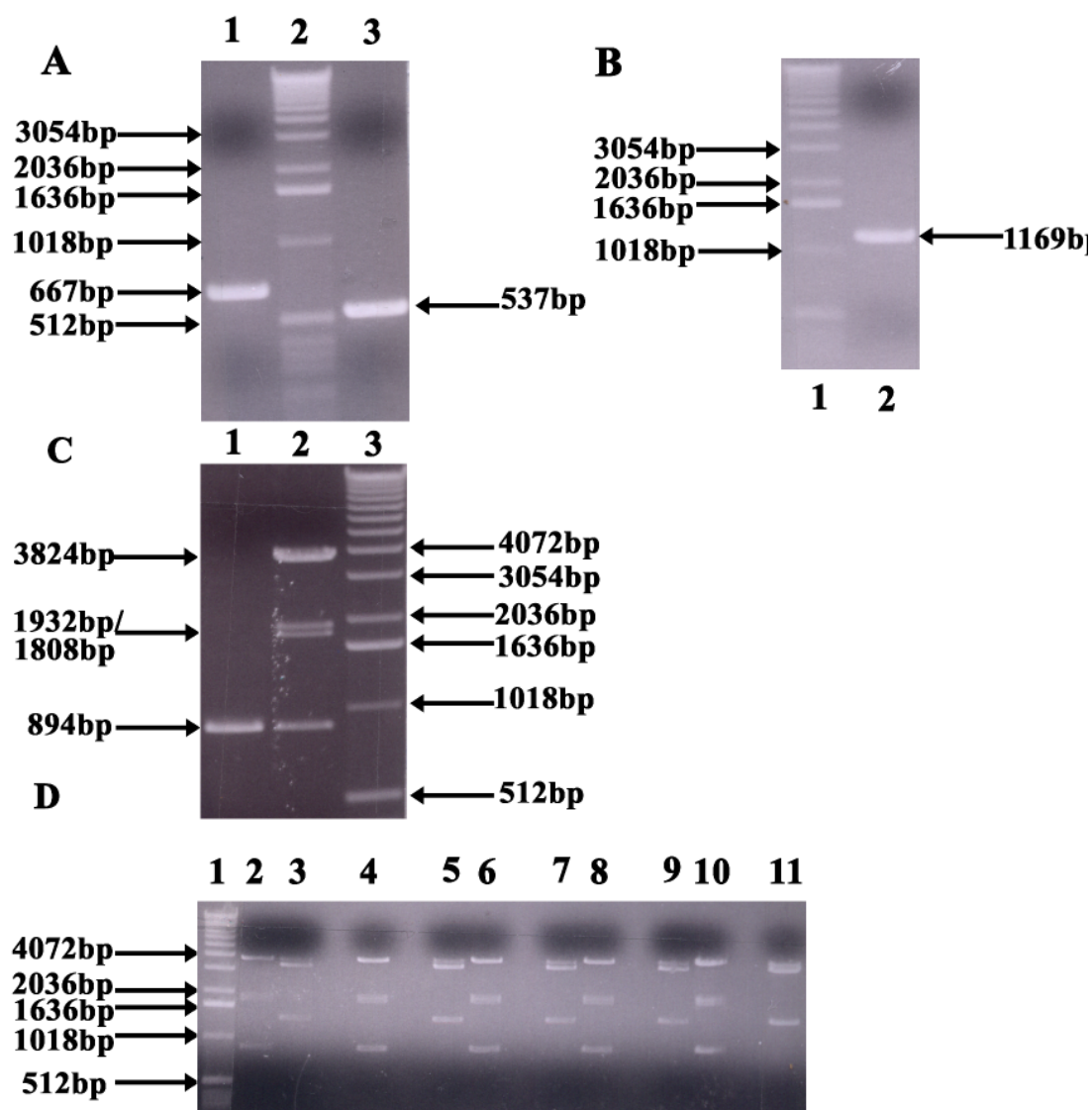


Figure 4.2 Products from the three PCR reactions. A) reaction products from PCR 1 (lane 3) and PCR 2 (lane 1), and the 1 kb ladder (lane 2). B) PCR product from the second round of PCR (lane 1), and the 1 kb ladder. C) fragments produced from restriction digestion of PCR 3 (lane 1), and p6HRT51^{C280S} (lane 3), and the 1 kb ladder (lane 2). D) Digest of plasmids purified from transformed bacteria digested with EcoRV/HindIII, lanes 2, 4, 6, 8, 10 (fragment length: 855/3859 [p6HRT51^{C280S}] and 1807/1933 [pDMI.1]) and EcoRI/HindIII, 3, 5, 7, 9, 11 (fragment length: 1351/3361 [p6HRT51^{C280S}] and 3740 [pDMI.1]).

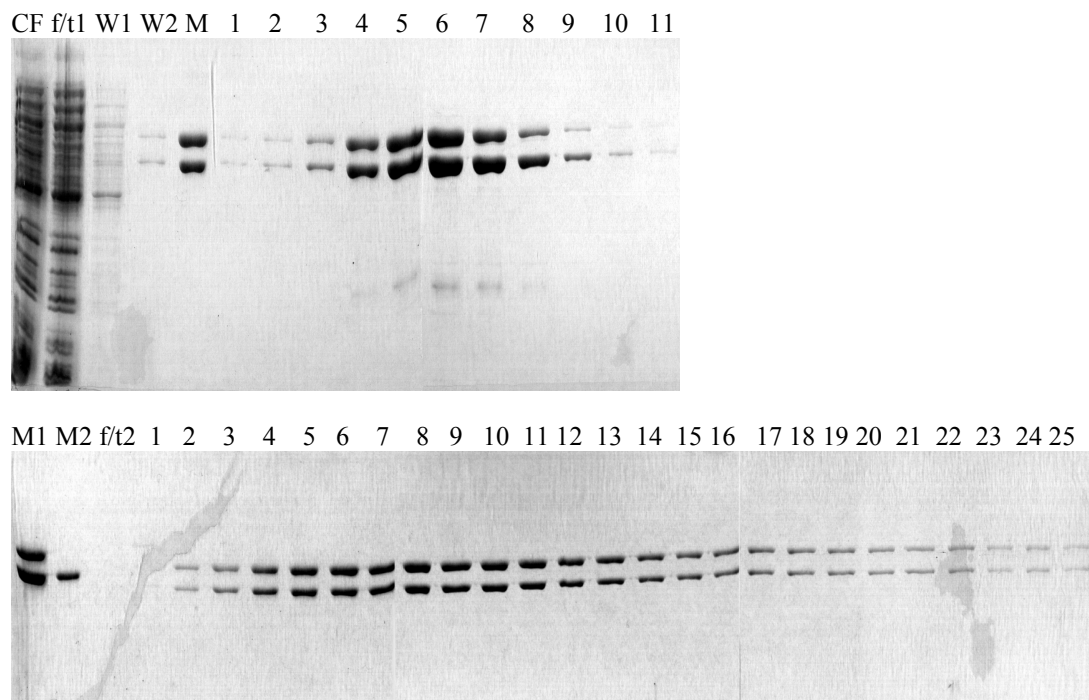
Two PCR reactions were performed in parallel using the plasmid p6HRT51^{C280S/} as a template with the following primers 1) K281C sense and Qx51 anti, and 2) K281C anti and S68C sense. The reaction products of length 537 bp (lane 3) and 667 bp (lane 1) are shown in Figure 4.2A. After purification, the two products from PCR 1 and PCR 2 were hybridised and the single stranded region filled using Pfu polymerase.

A third PCR reaction was performed with the outer primers, Qx51 anti and S68C sense, producing a fragment containing the mutation K281C and restriction sites for EcoRV and HindIII. The reaction product of 1169 bp (Figure 4.2B-lane 2) was gel purified. The PCR product and the parent plasmid p6HRT51^{C280S} were digested with EcoRV and HindIII (Figure 4.2C). The fragments of 3824 bp (lane 2) and 894 bp (lane 1) resulting from the digestion p6HRT51^{C280S} and PCR3 respectively were gel purified, and ligation was performed as described in section 3.1.5.

The ligation mixture was transformed into *E. coli* strain M15+pDMI.1 and transformants were selected on LB^{Amp/Kan}. Plasmids were purified using a plasmid miniprep kit (section 3.1.8) before selective digestion with EcoRV/HindIII and EcoRI/HindIII (Figure 4.2D). The size of the fragments produced determines if the plasmid is a positive or an undesired ligation product. The sequence of the newly inserted fragment was determined by DNA sequencing (section 3.1.14).

4.1.3 Expression and purification of HIV-1 RT

The p66 and p51 subunits of HIV-1 RT were expressed in *E. coli* M15/pDMI.1. The plasmid pDMI.1 encodes the Laq repressor which represses the *trc* and *tac* promoters of pRT166 and p6HRT51, respectively. p66 and p51[6H] (gene) expression was induced by the addition of IPTG and cells were grown for 3 h before harvesting. For cells expressing p66, the final OD₅₉₅ was between 1.5-2.0 and for p51[6H] a final OD₅₉₅ of between 10-12 was obtained. p66/p51[6H] can be purified by Ni-NTA affinity chromatography (section 3.2.2).



CF= Cytosolic Fraction , f/t1=Flow through from Ni-NTA column, W1= Wash 1 (buffer W + 1 M NaCl), W2=Wash 2 (buffer W + 5 mM imidazole, M1=2 µg RT Marker, M2= 1 µg p51 marker, f/t2=Flow through from Sp-Hitrap column, n, fraction number

Figure 4.4. Purification of p66 /p51[6H]^{K173C} using Ni-NTA, (upper panel) followed by purification using a combination of DEAE-Sepharose and an SP-hitrap column (lower panel).

The following example is shown for protein purification p66^{C38S/C280S}/p51[6H]^{C280S/K173C}. For the heterodimer purification a 1:15 ratio of p66 to p51[6H] containing cells was cohomogenised due to differential expression levels of the two subunits (Kensch 2000). After centrifugation, the supernatant was collected and the enzyme dimerised for 10 h at 4 °C in the presence of 10 mM MgCl₂. In a first step the RT enzyme was purified over a 6 ml Ni-NTA affinity chromatography column. The bound RT was washed with buffer W + 1 M NaCl to disrupt ionic interactions, e.g. protein:nucleic acid interactions, followed by buffer W + 5 mM imidazole to remove proteins loosely associated with the Ni-NTA matrix. The protein was eluted using an imidazole gradient from 0 - 500 mM imidazole over 100 min. Fraction 1-7 were pooled, while the later fractions which contain a higher percentage of p51[6H] were discarded (fractions 8-11). The pooled fractions were dialysed against RT-D-10(A) before further purification over a DEAE-Sepharose anion exchange column in series with a SP-Hitrap cation exchange column. Bacterial proteins and DNA are retained by DEAE-Sepharose while the heterodimeric RT and

p51[6H] are bound by the SP-Hitrap. After disconnection of the DEAE-Sepharose column, the heterodimeric enzyme is separated from p51[6H] by use of a NaCl gradient (0 – 200mM). During this purification all excess p51[6H] was removed on the Ni-NTA column. The RT produced is greater than 99 % pure and typically a yield of 1-2 mg/10 g of p66 expressing cells is obtained.

The protein mutants produced during this work are shown in table 4.1. All proteins were expressed to a similar level apart from the double p66 mutant (p66^{C38S/C280S/K287C/T470C}), which was expressed to a low level. All mutant proteins contained 2 cysteine to serine mutations in p66 (C38S, C280S) and a single cysteine to serine mutation in p51[6H] (C280S) and are referred to by the introduced cysteine.

<i>Heterodimer (p66/p51[6H])</i>	<i>Name used</i>
p66 ^{C38S/C280S/Q6C} /p51[6H] ^{C280S}	p66 ^{Q6C} /p51[6H]
p66 ^{C38S/C280S/W24C} /p51[6H] ^{C280S}	p66 ^{W24C} /p51[6H]
p66 ^{C38S/C280S/T27C} /p51[6H] ^{C280S}	p66 ^{T27C} /p51[6H]
p66 ^{C38S/C280S/E194C} /p51[6H] ^{C280S}	p66 ^{E194C} /p51[6H]
p66 ^{C38S/C280S/K287C} /p51[6H] ^{C280S}	p66 ^{K287C} /p51[6H]
p66 ^{C38S/C280S/K287C/T470C} /p51[6H] ^{C280S}	p66 ^{K287C/T470C} /p51[6H]
p66/p51[6H] ^{C280S/Q6C}	p66/p51[6H] ^{Q6C}
p66/p51[6H] ^{C280S/K173C}	p66/p51[6H] ^{K173C}
p66/p51[6H] ^{C280S/E194C}	p66/p51[6H] ^{E194C}
p66/p51[6H] ^{C280S/K281C}	p66/p51[6H] ^{K281C}

Table 4.1. Protein mutants produced during this work

4.1.4 Protein labelling with Alexa 488

In order to study the interaction between acceptor labelled nucleic acid substrate and donor labelled proteins by FRET, the following proteins were labelled with the green fluorophore Alexa Fluor[®] 488 C5-maleimide [A488]: p66^{Q6C}/p51[6H], p66^{W24C}/p51[6H], p66^{T27C}/p51[6H], p66^{E194C}/p51[6H], p66^{K287C}/p51[6H], p66/p51[6H]^{Q6C}; p66/p51[6H]^{E194C}; p66/p51[6H]^{K173C}; p66/p51[6H]^{K281C} as described in section 3.2.5.

After labelling, the excess dye was initially removed from the labelled RT by immobilising the protein on a Ni-NTA column via the N-terminal His-tag of p51[6H] and washing with different buffers, e.g high salt (1 M NaCl), or detergents (Tween-20

and Triton X 100). However, even after extensive washing on the column it was extremely difficult to remove the detergents from the protein. Also for Tween-20 washed RT, after elution with imidazole from the Ni-NTA column, crystals formed in the labelled protein solutions, leading to damage of the Millipore spin filters. In addition, removal of free dye using the Ni²⁺ column resulted in a low protein recovery of approximately 50 %. The optimal method for labelled protein purification was by gel filtration using a 30 ml bed volume sephadex G25 gel filtration column equilibrated in a high salt buffer (buffer W+1 M NaCl), which resulted in a 90 % recovery of protein.

Labelled protein

p66^{Q6C[A488]}/p51[6H]

p66^{W24C[A488]}/p51[6H]

p66^{T27C[A488]}/p51[6H]

p66^{E194C[A488]}/p51[6H]

p66^{K287C[A488]}/p51[6H]

p66/p51[6H]^{Q6C[A488]}

p66/p51[6H]^{K173C[A488]}

p66/p51[6H]^{E194C[A488]}

p66/p51[6H]^{K281C[A488]}

Table 4.2. Heterodimeric proteins labelled during this work.

After purification, two absorption measurements were carried out at 280 nm and 493 nm and the labelling efficiency was determined as follows:

$$\% \text{ labelling} = \frac{c_{A488}}{c_{RT}} = \frac{\text{Abs}_{493\text{nm}} \cdot \epsilon_{RT}}{\text{Abs}_{280\text{nm}} \cdot \epsilon_{A488}} = \frac{\text{Abs}_{493\text{nm}} \cdot 260000 \text{M}^{-1} \text{cm}^{-1}}{\text{Abs}_{280\text{nm}} \cdot 72000 \text{M}^{-1} \text{cm}^{-1}} = 1.02 - 1.10$$

eq. 4.1

Where $c_{A488} = [A488]$,

$c_{RT} = [\text{reverse transcriptase}]$,

$\text{Abs}_{280} = \text{Absorbance at 280 nm}$

$\text{Abs}_{493} = \text{Absorbance at 493 nm}$

$\epsilon_{RT} = \text{Molar extinction coefficient of reverse transcriptase (260,000 M}^{-1} \text{cm}^{-1})$

$\epsilon_{A488} = \text{Molar extinction coefficient of A488 (72,000 M}^{-1} \text{cm}^{-1})$

Typically measured labelling efficiencies were in the range of 102-110 %, due to a small amount of free dye remaining associated with the protein. Figure 4.6 shows a 10 % polyacrylamide gel of the labelled proteins, p66/p51[6H]^{E194C[A488]} (lane 1 and 3) and p66^{E194C[A488]}/p51[6H] (lane 2 and 4). The gel was first scanned for fluorescence using the “FLA5000 flouorimager” (Fuji) (lane 1 and 2), followed by staining with coomassie. The A488 fluorophore was excited at a wavelength of 473 nm, and fluorescence emission measured above 510 nm, using a 510 nm cut off filter (see section 3.2.1), followed by staining with Coomassie blue. As can be seen a certain percentage of free dye remains associated with the subunit that contains no free cysteine even after purification. Lanes 2 and 4 shows a Coomassie stain of the labelled proteins, showing a high degree of purity of the labelled protein.

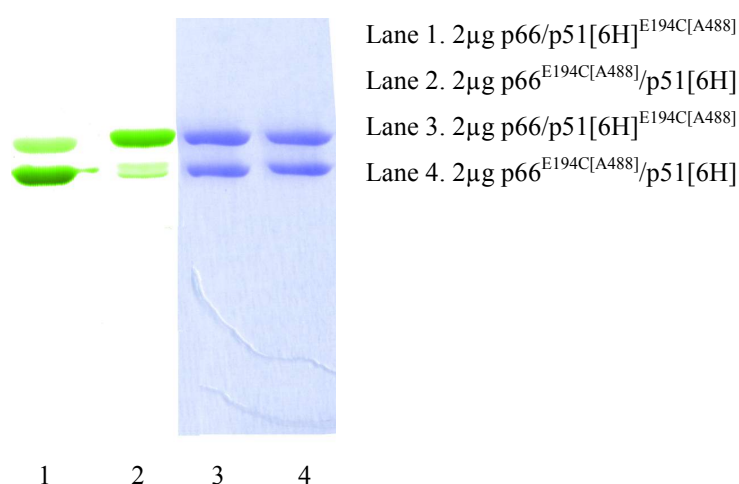


Figure 4.6. Protein labelling of p66/p51[6H]^{E194C[A488]} and p66^{E194C[A488]}/p51[6H]. 1 mg of each protein was labelled and proteins were purified using a 30 ml Sephadex G25 column. After concentration and quantification. 2 µg of each protein were loaded onto a 10 % SDS-PAGE gel. The gel was first scanned for fluorescence (lane 1 and 2), before staining with Coomassie (lanes 3 and 4).

4.1.5 Characterisation of labelled proteins

4.1.5.1 RNA-dependent DNA polymerase activity

All labelled proteins were tested using an RNA-dependent DNA polymerase activity assay (section 3.4.1) to determine whether the labelling procedure impaired enzyme activity (Figure 4.7). All proteins maintained close to wild type activity after labelling, with the exception of p66^{W24C[A488]}/p51[6H]. The unlabelled p66^{W24C}/p51[6H] (data not shown) displayed activity close to the wild type enzyme.

However, after coupling of the A488 dye the activity was reduced to less than 50 %. The crystal structure of the ternary complex (RT:p/t:dNTP) (Huang *et al.* 1998), in which the first three bases of the template overhang are resolved, reveals that W24 interacts with the 3rd nucleotide in the template overhang. The large A488 coupled at this position may cause steric hindrance for correct positioning of the p/t. Interestingly, for p66^{T27C[A488]}/p51[6H], in which A488 is coupled only three amino acid away from W24, there is very little reduction in activity, possibly indicating an ordered interaction of the template overhang with the finger domain of p66. Furthermore, both RT mutants labelled at the N-terminus (Q6C) of either p66 or p51 showed a greatly increased activity compared to wild type. Although this assay, gives some indication of the activity of RT, it is not extremely accurate. The values obtained are governed by dissociation events rather than catalysis.

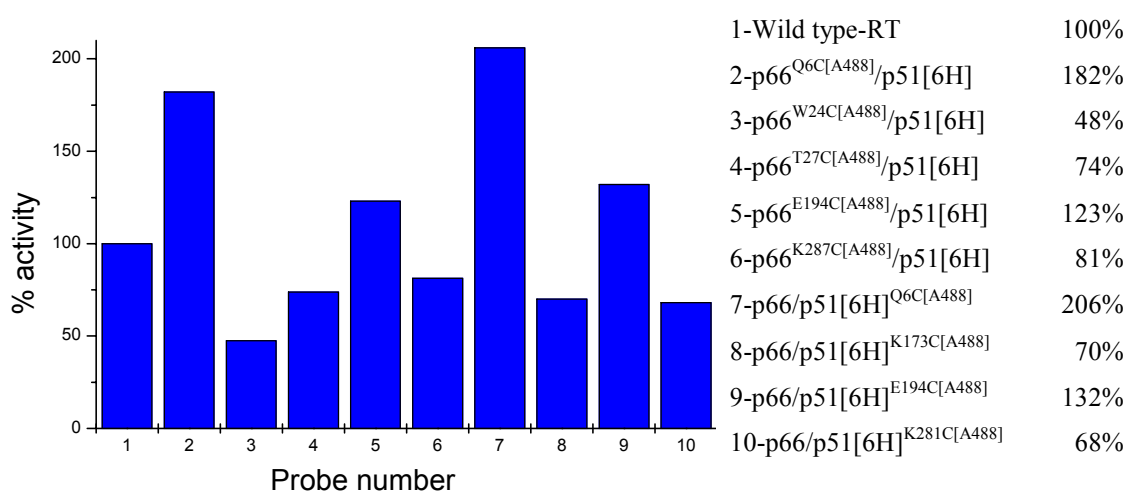


Figure 4.7. RNA-dependent DNA polymerase activity of labelled mutants. All labelled proteins were tested for their ability to incorporate [³H]dTTP into a poly(rA)/oligo(dT) primer/template. The wild type enzyme was set at 100% activity.

4.1.5.2 Single nucleotide incorporation kinetics

Quench flow experiments were performed as described in section 3.4.2 to determine if the low salt buffer used during the single molecule experiments had an effect on single nucleotide incorporation catalysis by RT. The autoradiograph of the 10% denaturing polyacrylamide gel for the analysis of nucleotide incorporation by the labelled protein p66/p51[6H]^{K281C[A488]} is shown in Figure 4.8. For the analysis of

nucleotide incorporation, the ratio of unextended primer (0) to extended product (+1, +2, +3) was considered. Therefore, it was possible to determine the amount of extended primer with respect to time.

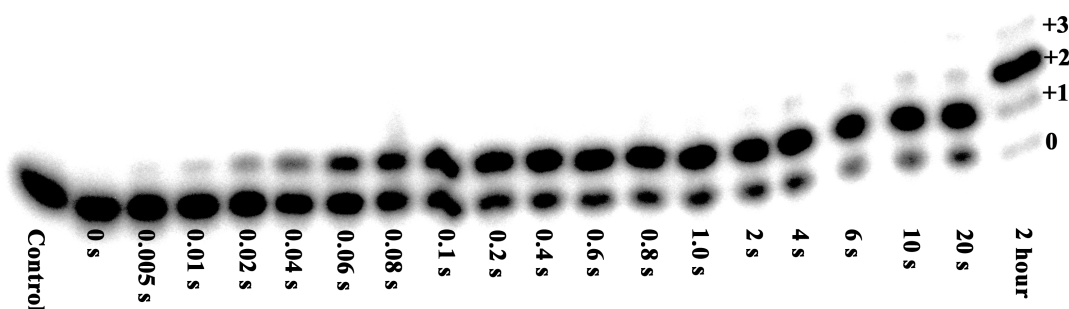


Figure 4.8. Single nucleotide incorporation by p66/p51[6H]^{K281C}[A488]. Autoradiograph showing the primer (0) and the amount of extended primer (+1, +2, +3) against time. Also shown are the control (no RT added), and a sample quenched after 2 hour. Experiments were performed in SMS buffer at 25 °C

The results of the analysis are shown in Figure 4.9. Fitting of the data with a double exponential gave a lower chi-squared value than with a single exponential. Determination of the rates (k_{pol1} and k_{pol2}) for the two burst phases yielded values of 11.9 s^{-1} and 0.53 s^{-1} , and the slope gave a value of 0.18 s^{-1} . A summary of the results is shown in table 4.3. The relative ratio between the amplitudes of k_{pol1} , k_{pol2} and the slope is 66%, 10%, and 24% respectively.

	Values obtained	(Wöhrl <i>et al.</i> 1999)
k_{pol1}	11.9 s^{-1}	$20\text{-}60 \text{ s}^{-1}$
k_{pol2}	0.53 s^{-1}	$0.5\text{-}2. \text{ s}^{-1}$
slope	0.18 s^{-1}	Not determined

Table 4.3. Comparison of data obtained in the current studies with those obtained previously (Wöhrl *et al.* 1999).

When comparing these results to those obtained previously (Wöhrl *et al.* 1999) a reduction in the rate of k_{pol1} by a factor of 2-5 is seen. The ratio between the phases is slightly altered with the fast phase now comprising 66% of the burst compared to 40%. Nevertheless the relative ratio between the two burst phases (k_{pol1} and k_{pol2})

compared to the slow steady state phase is 76% to 24% and is comparable to the previous kinetic experiments (Wöhrl *et al.* 1999). The results indicate that under the buffer conditions used for single molecule experiments a heterogeneous binding mode for RT:p/t complexes is observed.

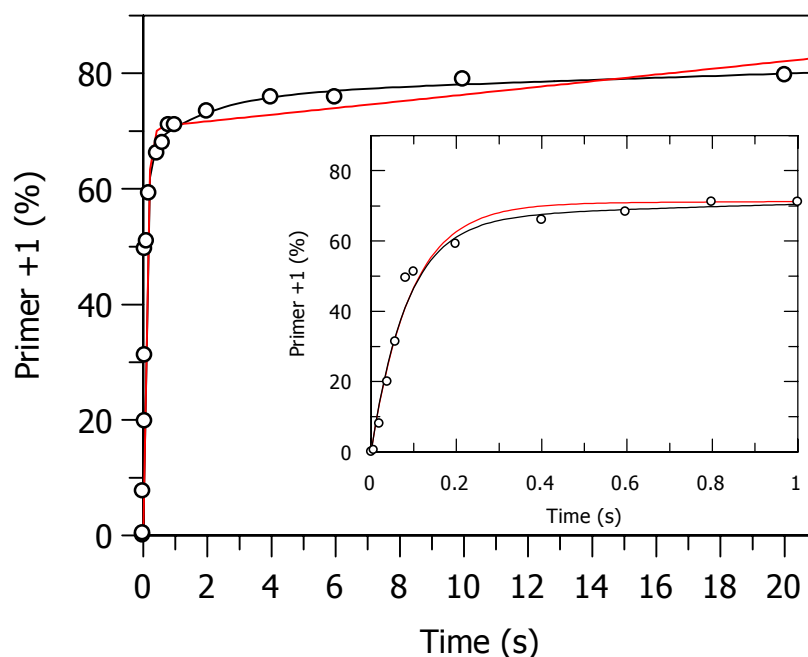


Figure 4.9. Single nucleotide incorporation by $p66/p51[6H]^{K281C[A488]}$. Data were fitted using a double exponential equation plus slope (black line) or to a single exponential plus slope (red line). A preformed complex of 400 nM RT and 200 nM p/t was rapidly mixed with 200 μ M dATP. The analysis of the data with a double exponential equation plus slope yielded values for k_{pol1} and k_{pol2} of $11.9 \pm 1.7 \text{ s}^{-1}$ and $0.53 \pm 0.97 \text{ s}^{-1}$ respectively.

4.1.5.3 Quantum yield determination of Alexa 488 by BIFL

For the determination of the transfer efficiency, E , and the donor-acceptor-separation, R_{DA} , the fluorescence quantum yield of the donor-dye coupled to the protein is required. The fluorescence quantum yield was calculated using the mean lifetime for each mutant relative to the fluorescence lifetime of the fluorescence dye Rh110. The mean lifetime was determined from a measurement of the labelled protein in complex with an unlabelled DNA/DNA p/t [referred to as p/t(U)] at a protein and substrate concentration in the range of 10^{-9} M in SMS buffer. Generally, a bi-exponential fit of the fluorescence decay was performed where the contribution of light scattering was considered. The fit also provides the amplitude of each obtained lifetime value, which was taken into account when calculating the mean lifetime. Due

to statistical reasons, the lifetime is not taken from a single molecule experiment because a minimum number of 100 photons in a burst is required. Therefore, only a mono-exponential fit is carried out in a single molecule experiment and provides less accurate lifetime determination compared to the bi-exponential fit of a more concentrated solution.

All quantum yields obtained are high (see table 4.3), indicating for all mutants, no significant quenching occurs in the presence of p/t(U). The lowest quantum yield was obtained for the mutant p66^{K287C[A488]}/p51[6H]. This may be due to interactions between A488 coupled to C287 with the p/t substrate. However, p66^{Q6C[A488]}/P51[6H] also has a reduced quantum yield, and from the crystal structure, this should have very little interaction with the rest of the protein or the p/t substrate. The highest quantum yields obtained are for p66^{E194C[A488]}/p51[6H] and p66/p51[6H]^{E194C[A488]}. The labels at these positions are coupled to rear of the protein and are thus far away from the nucleic acid binding cleft.

Labelled protein	τ_{av}^1 [ns]	τ_{av}^2 [ns]	Φ_{FD} from τ_{av}
A) p66 ^{Q6C[A488]} /p51[6H]	3,58	2,85	0,68
B) p66 ^{T27C[A488]} /p51[6H]	3,67	3,11	0,74
C) p66 ^{E194C[A488]} /p51[6H]	4,47	3,61	0,84
D) p66 ^{K287C[A488]} /p51[6H]	3,51	2,71	0,64
E) p66/p51[6H] ^{Q6C[A488]}	3,67	3,08	0,74
F) p66/p51[6H] ^{K173C[A488]}	3,70	3,11	0,75
G) p66/p51[6H] ^{E194C[A488]}	3,80	3,72	0,87
H) p66/p51[6H] ^{K281C[A488]}	3,78	3,21	0,76

Table 4.3. Lifetimes and calculated quantum yields for the labelled proteins used for the single molecule experiments.

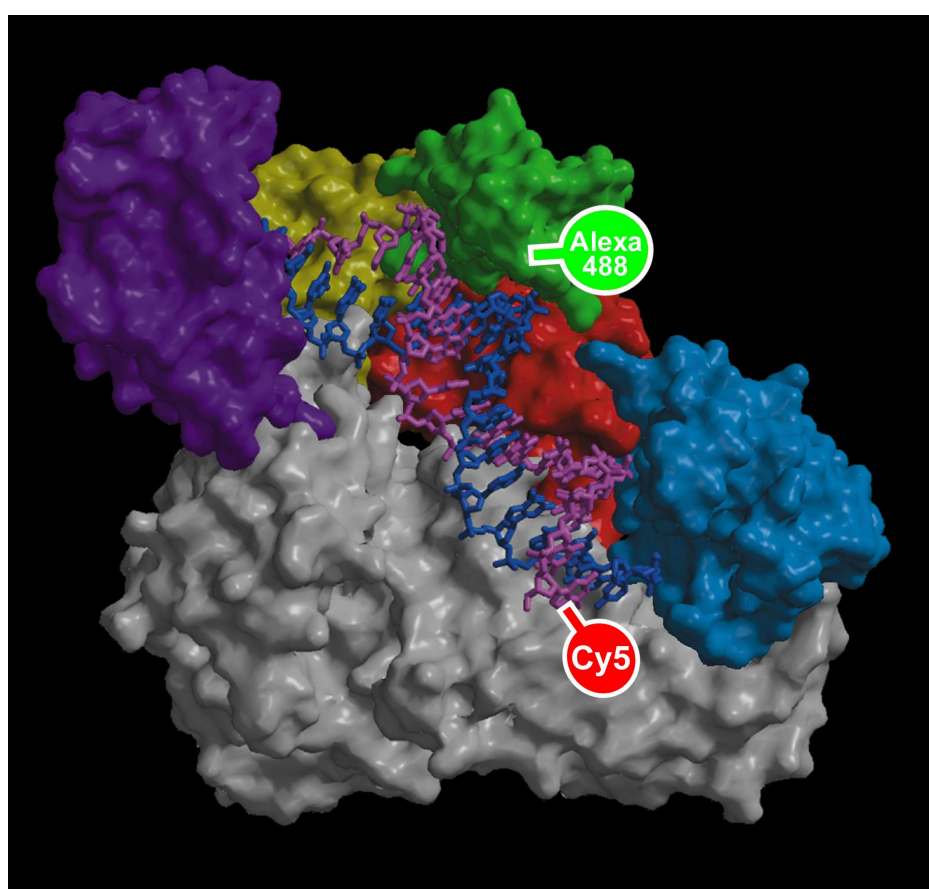
4.1.6 Single molecule experiments

4.1.6.1 Single molecule studies using the model system

p66^{K287C[A488]}/p51[6H]

Results obtained from previous kinetic experiments suggested the existence of three different RT:nucleic acid complexes coexisting in solution (Wöhrl *et al.* 1999): A so called productive complex (P) capable of facile nucleotide incorporation; a preproductive complex (PP), in which the nucleic acid substrate or protein must first

undergo an isomerisation event before nucleotide incorporation can occur; and a dead end complex (DE) from which the nucleic acid substrate must first dissociate from the enzyme, and rebind in a P or PP manner before incorporation can occur (see section 1.5). The ratio of the different complexes varies depending on the substrate used (DNA/DNA p/t vs DNA/RNA p/t). For a DNA/DNA p/t the P-, PP- and DE-complex consist approximately 25 %, 45 % and 30 % of the total population respectively. Whereas for DNA/RNA the ratio was 72 %, 16%, and 12% for the P-, PP-, and DE-complexes, respectively (Wöhrl *et al.* 1999).



Cy5_TTGTC CCTGTTCGGGCGCC
AACAGGGACAAGCCCGCGGTACGTCTAATTGG

Figure 4.10. The structure of HIV-1 RT. The protein is displayed as a molecular surface with the different regions of p66 being coloured as follows: fingers (purple), palm, (yellow), thumb (green), connection (red), and RNase H domain (blue), while p51 is coloured grey. The nucleic acid substrate is shown as a stick representation with the primer coloured purple and the template coloured blue. The positions of the donor (Alexa⁴⁸⁸-C₅ maleimide) and acceptor (Cy5 attached via C6-aminolink) labels are also displayed. The sequence of the 19/35 primer/template is based on the HIV-1 viral primer binding site (PBS) (Ratner *et al.* 1985).

One of the first considerations for the single molecule experiments was the affinity of the binary complex between RT and its p/t substrate, due to the low concentrations needed for single molecule experiments (50-100 pM protein, 200 pM-400 pM p/t). It was shown previously that the affinity of a DNA/DNA p/t complex for RT was in the low nanomolar range (Divita *et al.* 1993; Krebs *et al.* 1997; Thrall *et al.* 1998) under the buffer conditions used in these studies. Experiments by Kensch (Kensch 2000) showed that the K_D is dramatically reduced by lowering the KCl concentration to 10 mM. An accurate value for the K_D using equilibrium ensemble measurements could not be obtained, due to the very low concentrations required and thus poor signal quality for such a tight interaction. Data were obtained which suggested an upper limit of K_D of ca. 100 pM (Kensch 2000).

In order to investigate the three nucleic acid binding modes seen for HIV-1 RT, and to gain structural information on the different complexes, spFRET was used. By using a donor labelled protein in combination with an acceptor labelled DNA/DNA p/t substrate different complexes can be sorted by spFRET due to different rates of energy transfer. For structural analysis of the three complexes a DNA/DNA p/t substrate was chosen. Due to the more equal distribution between the three complexes, compared to DNA/RNA (see above), sorting of complexes due to differential rates of FRET should prove easier.

To study the distance-dependence of FRET between a donor dye, D, and an acceptor, A, p66^{K287C[A488]}/p51[6H] was used in combination with a 19/35 p/t in which the acceptor dye, Cy5 is coupled to the 5'-end of the primer, a construct referred to as p_5_Cy5/t (Figure 4.10). The p_5_Cy5/t construct was designed such that the distance between the residue at position 287 and the 5'-end of the primer was approximately 40 Å in the structure characterised by X-ray crystallography (Huang *et al.* 1998), fitting nicely into the accessible range of FRET (20 to 100 Å).

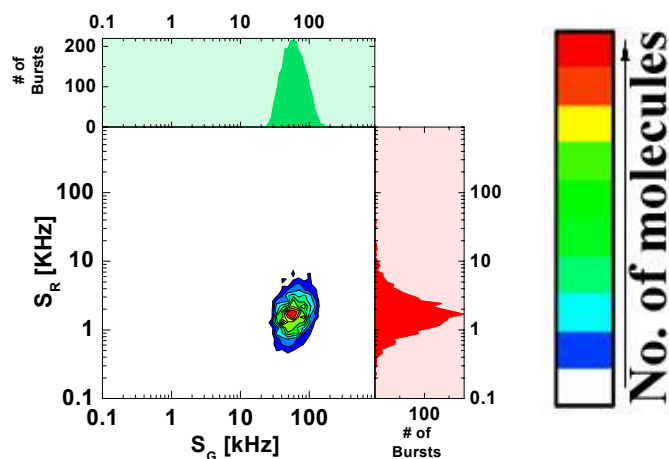


Figure 4.11. Signal intensity plot of $p66^{K287C[A488]}/p51[6H]$ in complex with $p/t(U)$. Reaction conditions are as follows $50 \text{ pM } p66^{K287C[A488]}/p51[6H] : 200 \text{ pM } 19/35 \text{ p/t } [p/t(U)]$ measured in SMS buffer at $25 \text{ }^\circ\text{C}$.

The basic information obtained from an MFD experiments are the green and red signals (S_G and S_R). Figure 4.11 shows the signal intensity plot for $p66^{K287C[A488]}/p51[6H]$ in complex with $p/t(U)$. In the top panel is a 1D histogram of the green signal (S_G) against the number of molecules. Each bar in the histogram is generated by binning molecules which have the same characteristics, in this case S_G of a defined count rate. S_G is correlated against the red signal (S_R) (right hand side panel). The 2D plot is the correlation between S_G and S_R and is coded from blue (lowest) to red (highest). For the donor only, $p66^{K287C[A488]}/p51[6H]:p/t(U)$ complex there is one “island” of intensity in the 2D histogram which has a high green signal (59 kHz), corresponding to the unquenched donor, and a low red signal (1.7 kHz) which is equal to background signal and crosstalk.

When looking at the $[p66^{K287C[A488]}/p51[6H]:p_5'_Cy5/t]$ sample two islands are apparent indicating the presence of at least two species (Figure 4.12). The first species has a greatly reduced S_G (21 kHz) and a corresponding increase in S_R (26 kHz) due to efficient FRET. The second species has a slightly reduced S_G (53.5 kHz) and an increased S_R (2.4 kHz) which is due to low FRET.

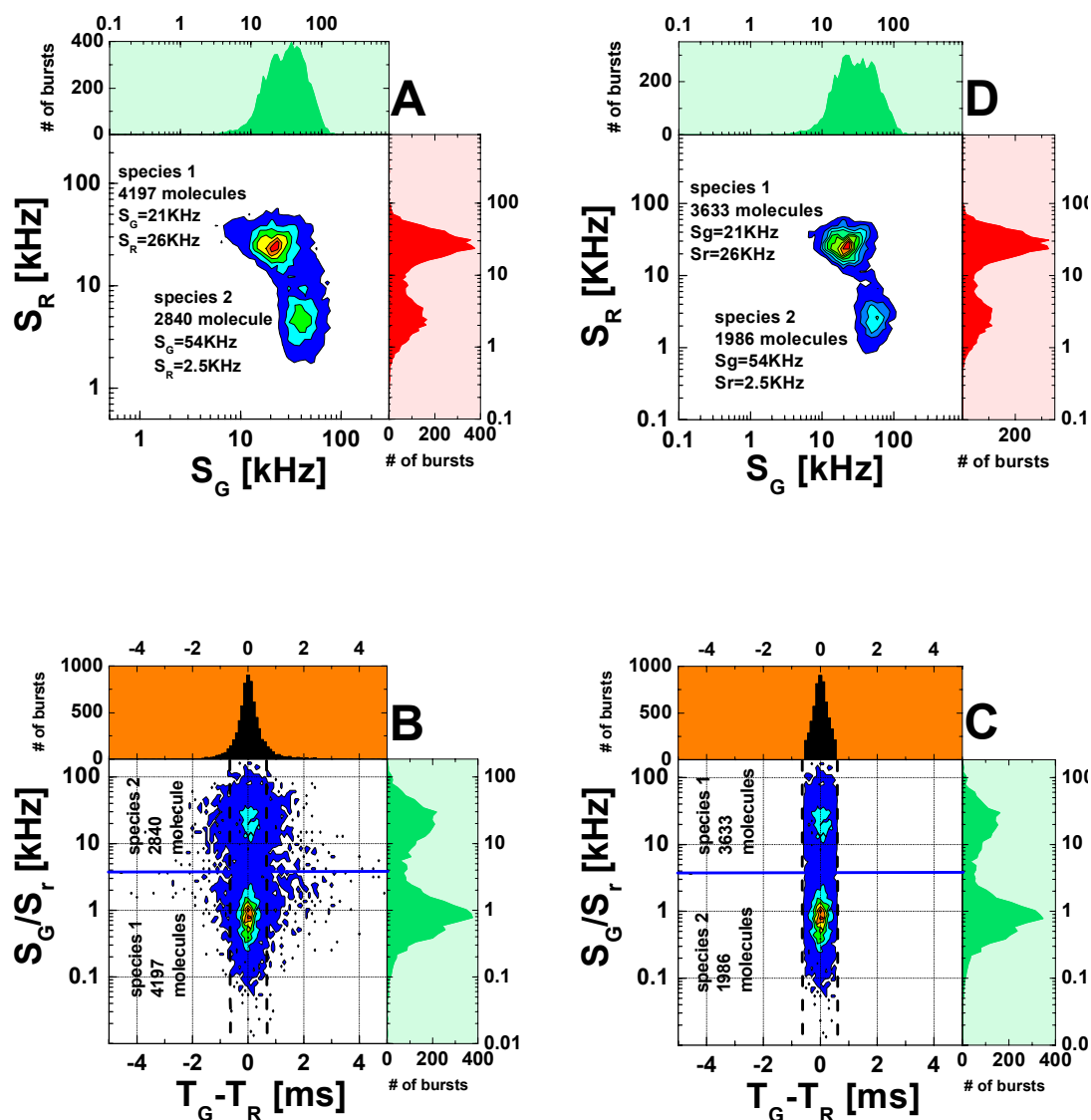


Figure 4.12 Cleaning procedure to remove bursts containing photobleached molecules. A) S_G vs S_R for K287C:p_5_Cy5/t sample. The No. of molecules in each burst is indicated, as well as the green and red signal intensities. B) S_G/S_r vs $T_G - T_R$ showing the limits used for removing photobleached molecule (dashed line ± 0.55). The S_G/S_r level used to separate species I and species II for analysis is also indicated (Blue solid line). C) S_G/S_r vs $T_G - T_R$ after cleaning, and D) cleaned S_G vs S_R plot. Reaction conditions: 50 pM p66^{K287C[A488]}/p51[6H], 200 pM p/t measured in SMS buffer at 25 °C

In FRET measurements at the single molecule level it is important that donor and acceptor dyes are fluorescent during the time of each detected fluorescence burst. To discriminate bursts where at least one of the dyes is photo-bleached, due to excitation, the difference of the photon weighted centre of gravity of “green” and “red” photons is used, $T_G - T_R$. For this analysis the “green” and the “red” photons of a single molecule event are considered separately. The photonic centre of gravity for

the “green” and “red” events can then be determined, in respect to the time at which it occurs. If no photo-bleaching occurs, the photonic centre for the “green”, T_G , and the “red”, T_R , occur at the same time and therefore $T_G - T_R$ is zero. However, if during a transition of a molecule photo-bleaching of the acceptor dye occurs, the number of photons detected in the “red” channels is reduced. This means that “green” photons are detected for a longer time than “red” photons and the two centres of gravity are shifted resulting in a value of $T_G - T_R$ higher than zero. If the green dye is photodestroyed the value of $T_G - T_R$ will be less than zero. For the data analysis of the single molecule experiments in this work, a value of $-0.55 \text{ ms} < T_G - T_R < +0.55 \text{ ms}$ was chosen.

This selection removes mostly molecules from the centre region between species I and species II. The cleaned intensity plots shows approximately 65% of the molecules are present in the species I peak and 35 % of the molecules in the species II peak. From the intensity plots it is difficult to determine whether the two species observed are due to different conformations of the donor dye, possibly leading to quenching or due to FRET. Although distances can be calculated from the donor or acceptor intensities independently, intensity measurements alone cannot detect donor quenching, and cannot distinguish between acceptor quenching and distance changes.

4.1.6.2 Structural studies using p66^{K287C[A488]}/p51[6H]: p_5'_Cy5/t

In order to gain distance information on the different complexes detected the data were converted into donor-acceptor distance, R_{DA} ; as described in section 3.5.1.5

Figure 4.13. plots donor-acceptor distance, R_{DA} , against donor lifetime, $\tau_{D(A)}$, in the top panel, and $\tau_{D(A)}$ against the donor anisotropy r_D in the lower panel. The iso- τ curve (black sigmoidal line-upper panel (eq 3.20)) displayed in the upper panel is the relationship between $\tau_{D(A)}$ and R_{DA} calculated from the fluorescence intensity ratio (F_D/F_A) (see section 3.5.1.5). If donor quenching exists, then the fitted donor lifetimes will be shorter than the expected τ_D , shifting those bursts to the left. For acceptor quenching, F_A will be reduced and according to equation 3.16, the calculated distances will shift up. Consequently, any peaks which fall on the line given by eq 3.20 are due to FRET events, and not changes in the donor quantum yield, Φ_D , or acceptor quantum yield, Φ_A , due to differences in the local environment. On the right upper panel is a 1D histogram of R_{DA} . Superimposed on this histogram is the expected

Gaussian distribution given by eq. 3.19 (dotted black line). On the lower panel in Figure 4.13 are plotted rotational correlation times of 7 and 2 ns (black lines) calculated from the Perrin equation (eq. 3.6). Acquisition of anisotropy data allows us to determine the degree of mobility of the donor fluorophore, and hence ascertain whether using $2/3$ as an approximation for κ^2 is valid.

Figure 4.13C plots the data from , p66^{K287C[A488]}/p51[6H] in complex with p/t(U) (analogous to Figure 4.11). Since there is no acceptor, there is no FRET, and subsequently there should be no red signal. However, background, dark counts, and crosstalk still contribute to a red signal of 1.7 kHz. From this signal a noise limited maximum discernible distance resolution (MDR) can be calculated (grey shaded area). The donor only sample has a single broad peak centred at 92 Å. The broadness of the peak is due to the small numbers of red photons in each burst, causing the ratio F_D/F_A to vary greatly with a change of only one or two red photons. Nevertheless, the fit between the theoretical distribution and the measured background is excellent for the donor only species.

Looking at anisotropy of the donor only species [p66^{K287C[A488]}/p51[6H]:p/t(U)], it can be seen that the data lies along the line for $\rho_D = 2$ ns. As the rotational correlation time is small compared to $\tau_{D(A)}=4.1$ ns, the donor at this position has a high degree of freedom, and therefore using $2/3$ as an approximation for κ^2 is valid.

For p66^{K287C[A488]}/p51[6H] in complex with p_5_Cy5/t (Figure 4.13A), two peaks are apparent and indicate the presence of at least two different species, corresponding to species I and II described in section 4.1.6.1. Species I is the majority species, broadly distributed along the iso- τ line with a maximum at $R_{DA} = 45$ Å. There is close agreement between the R_{DA} value and the expected distance calculated from the crystal structure (~ 40 Å) plus dye linkages. In the 1D distance histogram, species I cannot be fitted using equation 3.19. As the distribution of the peak follows the iso- τ line, we can rule out broadening due to dye quenching. The anisotropy plot (Figure 4.13A (lower panel)) shows a broad peak for species I with rotational correlation times ranging from 2 ns to 7 ns. This taken together with non Gaussian distributed R_{DA} peak indicates the existence of two subspecies in the species I distribution.

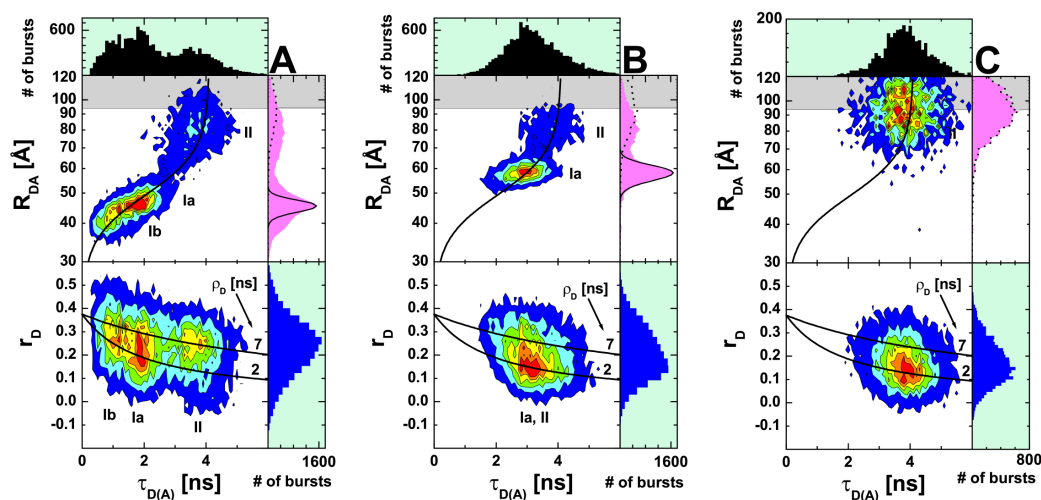


Figure 4.13. $\tau_{D(A)}$, plotted versus R_{DA} (upper panel) and against r_D , (lower panel) for A) $p66^{K287C[A488]}/p51[6H]:p_5_Cy5/t$, B) after addition of 50 μM dATP, dTTP, dGTP and ddCTP and C) $p66^{K287C[A488]}/p51[6H]:p/t(U)$. Also plotted is the Iso- τ line (black line-upper panel) and rotational correlation times calculated from the Perrin equation (black line-lower panel). The grey area is the MDR of 92 Å. The 1D R_{DA} -histograms also display theoretical predictions of single distance distribution profile (solid line) (see section 3.5.1.5) Reaction conditions: 50 pM $p66^{K287C[A488]}/p51[6H]$, 200 pM p/t measured in SMS buffer at 25 °C.

Species II (Figure 4.13A and 4.13B) has a single, well defined peak located at $R_{DA} = 83$ Å. In the 1D histogram, it is clearly seen to be separate from the donor only species (see Figure 4.13C). This distance would position the substrate far from the nucleic acid binding cleft, which does not correspond to any previously determined structure of an RT:p/t complex. The anisotropy plot [Figure 4.13A (lower panel)] indicates that the dye has a greatly reduced mobility. However, the error margin associated with low mobility is only 10 Å, so the characteristics of this binding mode must be due a large distance, and not orientational effects.

From the distances obtained, it would be expected that species I would correspond to the P-complex observed by crystallography (Huang *et al.* 1998; Jacobo-Molina *et al.* 1993). In order to confirm this the activity of RT was probed by performing selective incorporation of four nucleotides, terminating the chain with a dideoxy-nucleotide (Figure 4.13B). The extension of the primer increases the DA distance, and shifts species I to larger R_{DA} , indicating that this is the P-complex. The measured distance change of 13 Å corresponds to the expected extension of 4 base pairs. In contrast to Figure 4.13A, the theoretical distribution and the distance

histogram are in close agreement. Additionally, ρ_D is low (2 ns), indicating a mobile dye. The distance for species II remains unchanged in the presence of nucleotides.

4.1.6.3 Dissociation rate constant- K_{off} measurements

Further evidence for the existence of two different classes of RT:p/t complexes is obtained by measuring the specific dissociation rates, k_{off} , in a displacement experiment. A one thousand fold excess of p/t(U) was added to the solution and the amount of each $p66^{K287C[A488]}/p51[6H]:p_5_Cy5/t$ species was followed over time (Figure 4.14). The fluorescence parameter used to distinguish the species is ratio of donor to acceptor fluorescence signals, S_G/S_R . A distinct decrease in both $p66^{K287C[A488]}/p51[6H]:p_5_Cy5/t$ complexes is observed (Figure 4.14B and C) together with a concomitant increase in RT:p/t(U) complex containing only donor (Figure. 4.14C). Species I and II were fitted using a mono exponential decay function and dissociation rates for species I and II which differ by a factor of 2: $k_{off}(I) = 0.0062 \pm 0.0003 \text{ s}^{-1}$ and $k_{off}(II) = 0.0026 \pm 0.0005 \text{ s}^{-1}$ were obtained. Previous ensemble experiments by (Kensch 2000) using a 16/36 DNA/DNA p/t yielded k_{off} values of 0.017 s^{-1} and 0.003 s^{-1} , in reasonable agreement to the values obtained in this study.

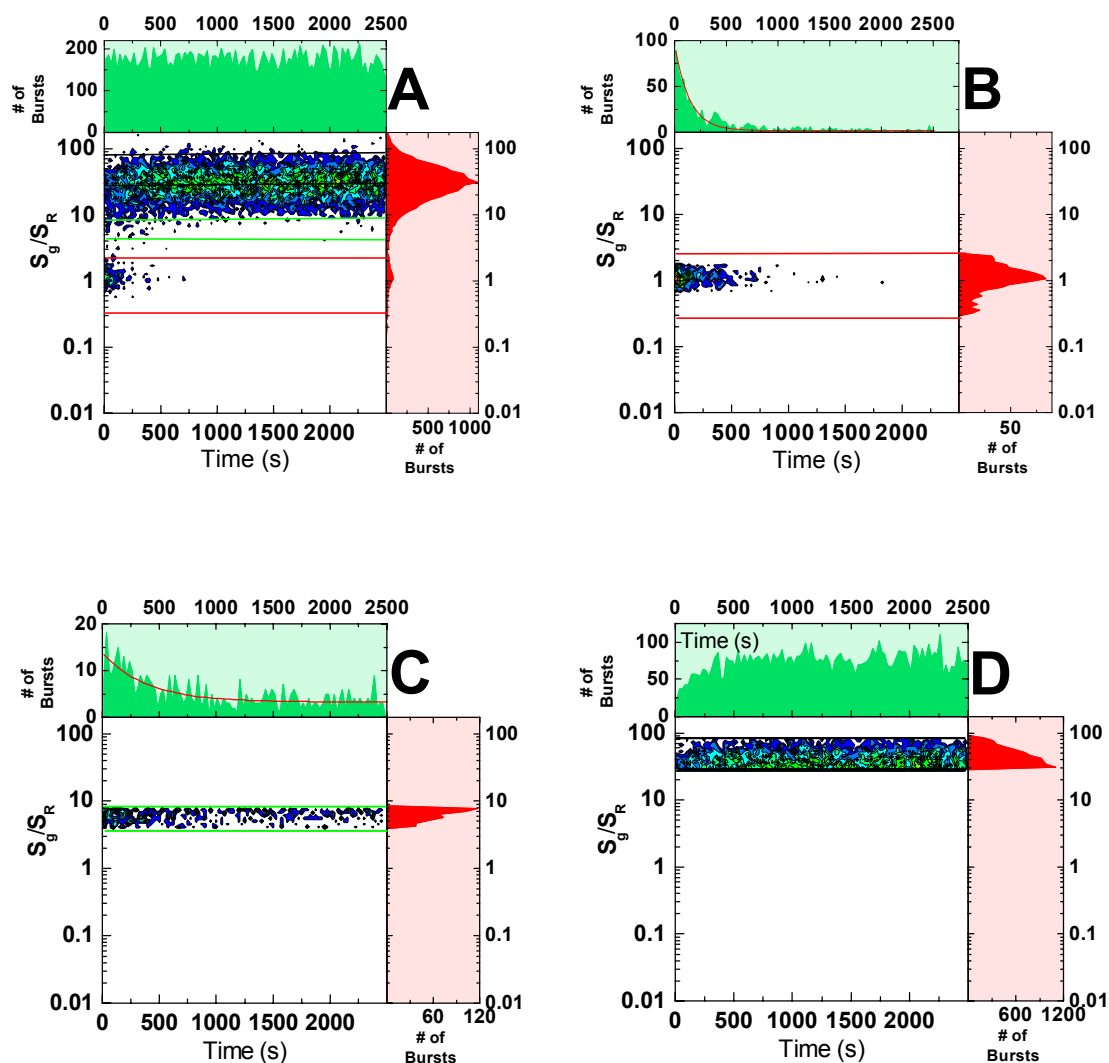


Figure 4.14. S_G/S_R vs mean macrotime plots for k_{off} measurements. A 1000 fold excess of DNA/DNA p/t was added to a 50 pM $p66^{K287C[A488]}/p51[6H]:200pMol p_5 Cy5/t$ (A), and the decay of species I (B) and species II (C), as well as (D) the increase of donor only species ($p66^{K287C[A488]}/p51[6H]:DNA/DNA p/t$) was followed. Acquisition of data was started 13 seconds after the addition of the unlabelled p/t, and the plots represent an average of three measurements. Measurements performed in SMS buffer at 25 °C.

4.1.6.4 Basic model of RT:p/t interactions

Figure 4.15 shows a simple model of how the p/t may interact with the enzyme based on the R_{DA} information obtained for the three RT:p/t complexes. When comparing the results with those of previous kinetic experiments (Wöhrl *et al.* 1999) the first and major complex with a shot-noise limited R_{DA} distribution is most likely to be the P-complex (Ia), and probably corresponds to the structure seen by (Huang *et al.* 1998). Taking the additional length of dye linkage into account, there is close

agreement between the expected distances calculated from the crystal structure (~ 40 Å) and the donor acceptor distance R_{DA} of 45 Å calculated for species Ia.

Species Ib is characterised by a broader R_{DA} distribution, which might be indicative of a less defined species. The mean distance between the dyes in this complex has shortened by ~ 6 Å, compared to the P-complex. One possibility is that the 3'-end of the primer has moved further into the polymerase active site, so that the terminal nucleotide is now occupying the binding site for incoming nucleoside triphosphate. This is the expected state occurring in the polymerisation reaction immediately after nucleotidyl transfer. Therefore, before nucleotide binding can occur, the p/t complex has to move back to free the active site (i.e. by generating complex Ia). Accordingly, the slower of the two fast phases seen in nucleotide incorporation kinetics (Wöhrl *et al.* 1999) may correspond to this movement, which is an essential step that must occur after nucleotide incorporation and represents the fundamental translocation step in the polymerisation reaction. Complex Ib was therefore assigned as the preproductive (PP) complex.

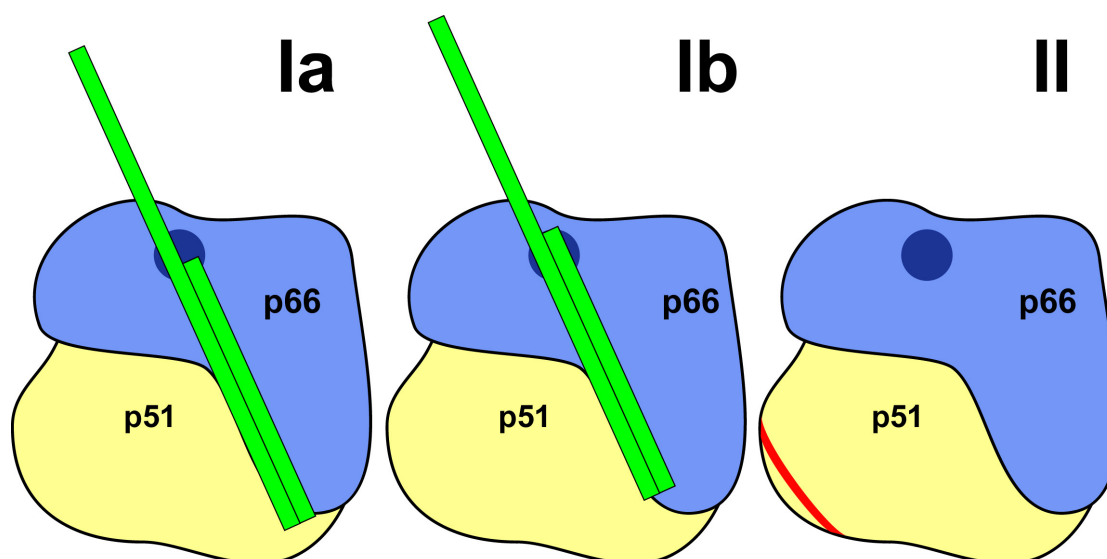


Figure 4.15 Cartoon illustrating the RT: p/t interactions. The p66 subunit is coloured in light blue, and the p51 subunit is coloured yellow. The polymerase active site of p66 is coloured dark blue. The productive complex (Ia) interacts with the p/t (green) in a state closely resembling the known RT: p/t structures (Huang *et al.* 1998; Jacobo-Molina *et al.* 1993). In the pre-productive complex (Ib) the primer terminus is shifted such that it now occupies the dNTP-binding pocket. For the third complex (II) the red line shows the area of contact of the 5'-terminus of the primer with the protein, generated using a distance of 87 Å between donor and acceptor dyes.

The third complex (species II) seen appears to have a very different structure, as indicated by a broad R_{DA} distribution peaking at a long DA-distance of approximately 83 Å. In Figure 4.15 the red line delineates the area of the protein in which the acceptor dye interacts with the protein. The DE-complex was originally postulated to explain the slowest kinetic phase observed for single nucleotide incorporation, in which the nucleic acid substrate must first dissociate from the enzyme and rebind in the P- or PP-manner before nucleotide incorporation can occur. In view of the very different R_{DA} , the failure to incorporate nucleotides rapidly, and specific dissociation rates (Figure 4.14), species II may correspond to the DE complex. The DE complex appears to bind in a completely different manner than the P-species. As shown in the earlier kinetic work, the p/t substrate must first dissociate from the enzyme and rebind before incorporation can occur, a process which appears to be very slow under single molecule conditions. From the studies undertaken it can be discounted that the p/t is bound with the wrong polarity, with the blunt end of the double stranded region instead of 3'-end of the primer at the polymerase active site. Such a complex would have a high rather than low FRET efficiency.

4.1.6.5 Addition of pyrophosphate or dNTPs to a preterminated p/t

From the structural data it appears that in the PP-complex, the 3'-terminus of the primer is shifted such that it occupies the dNTP-binding pocket (Figure 4.15). In this state addition of a dNTP should displace the 3'-terminus from the dNTP pocket, leading to the formation of the P-complex. Conversely, pyrophosphate should drive the reaction backwards, allowing the removal of the last nucleotide in the primer, and presumably bases thereafter.

To test the hypothesis that species I consists of two subpopulations, the activity of RT was tested when bound to a terminated p_5_Cy5/t (referred to as p_3_Term_5_Cy5/t). The sequence of the primer is the same as p_5_Cy5, however the first base incorporated during oligonucleotide synthesis, was a ddCTP. Figure 4.16A is analogous to Figure 4.13A, indicating that the presence of ddCTP on the 3' primer terminus does not have a significant effect on the distribution of the three complexes observed.

In order to ensure the same experimental conditions, The same mixture of dNTP was added as the previous incorporation experiments (section 4.1.6.2). After the addition of dNTPs to an end concentration of 200 μ M species I is shifted into a single

Gaussian distributed peak with a high rotational mobility. The single species is characterised by a distance of 48 Å, and analogous to species (Ia, P) (see Figure 4.15). This behaviour is identical to that seen in the extension experiment (Figure 4.13B), and probably corresponds to a state where RT is prepared to accept the dNTP for incorporation, but is prevented from doing so by the ddNTP already incorporated into the DNA. This result provides additional evidence that the nucleic acid binding mode seen for the PP-complex (Ib) is a result of a shifting of the 3'-terminus of the primer into the dNTP-binding pocket.

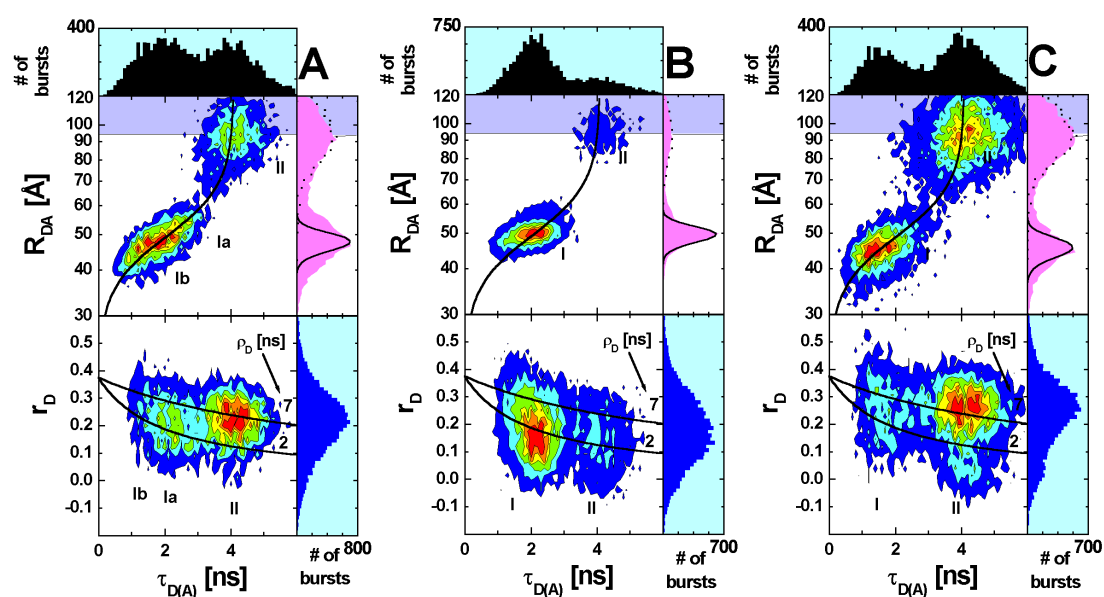


Figure 4.16. Donor lifetime, $\tau_{D(A)}$, plotted versus DA-distance, R_{DA} (upper panel) and against donor anisotropy, r_D , (lower panel) for A) 50 pM p66^{K287C[A488]}/p51[6H]:200 pM p_3_term_5_Cy5/t, B) after addition of 50 μ M dATP, dTTP, dGTP and ddCTP, or C) after addition of 200 μ M Na₄PPi. Also plotted is the Iso- τ line (black line-upper panel) and rotational correlation times calculated from the Perrin equation. The purple rectangle is the maximal discernible resolution we can achieve for this system.

Measurements performed in SMS buffer at 25 °C

One interesting aspect of these results is the reequilibration between species I and species II. When the primer was extended by 4 bases (section 4.1.6.2), and terminated by a ddNTP no significant reequilibration was seen between species I and species II. This may be due to the fact that significant structural changes in the nucleic acid conformation may have to occur during translocation (e.g. A'-like to B form DNA transition) along the p/t before the tight ternary complex (RT:p/t:dNTP) is formed, whereas in the case of the [p66^{K287C[A488]}/p51[6H]:p_3_term_5_Cy5/t] sample

the presence of the nucleotide in the dNTP binding pocket would only require a small change in the nucleic acid structure to attain the ternary complex. This re-equilibration is fast compared to the measurement time. Any changes that must occur are happening in a time range faster than 30 s, the dead time for these experiments.

Significant changes are also seen in the mobility of the donor dye for the [p66^{K287C[A488]}/p51[6H]:p_3_term_5_Cy5/t] sample. Both species are now situated on/or below a rotational correlation time of 2 ns. For the P-complex this may be explained by a relative movement of the fingers subdomain compared to the thumb domain on dNTP binding which could result in an increased mobility for the dye coupled at position 287. It may also be due to a repositioning of the template overhang on transition from the binary complex [RT:p/t] (Jacobo-Molina *et al.* 1993) to the ternary complex [RT:p/t:dNTP] (Huang *et al.* 1998) resulting in increased dye mobility for this labelled mutant. However, due to the lack of structural information on the template overhang in the binary complex it is hard to determine whether this is the case. For the DE complex the significance of the increased mobility is hard to determine. Species II seems to play no role in the polymerase activity of RT, and therefore binding of a dNTP should have no effect on the donor dye mobility. However, it is possible that a dNTP can occupy the dNTP-binding pocket for this binding mode increasing the rotational mobility of the donor dye coupled to the thumb domain of p66.

In the second experiment, Na₄PP_i was added to an end concentration of 200 μM, a level slightly higher than physiological concentrations (Barshop *et al.* 1991). In the presence of pyrophosphate, the peak for species I is shifted towards shorter distances. In addition, the peak is not Gaussian distributed, and ρ remains high. For species II, there is a slight increase in distance, and an anisotropic splitting in two distinct peaks. One peak is centred at a value higher than 7 ns, indicating a highly restricted dye. The second peak is shifted to a value below 2 ns. It may reflect a more open conformation of the enzyme in the presence of pyrophosphate. We also however, cannot rule out that this increased mobility is due to a fluorescent contamination from the added pyrophosphate.

In contrast to the addition of dNTPs there is an increase in the level of species II and a shifting of the peak, above the MDR. This suggests that formation of the donor only species occurs, probably due to the presence of Pyrophosphate in the

dNTP binding pocket leading to a reduction in affinity for the p/t. The process of pyrophorolysis appears to be slow under the conditions used. After an initial reequilibration between species I and II, the level of species I remains constant.

One of the major classes of inhibitors against HIV-1 are the so called nucleoside RT inhibitors (NRTIs). NRTIs compete against natural dNTPs both for recognition by RT as a substrate, and for incorporation into the nascent viral DNA chain. NRTIs inhibit RT catalysed proviral DNA synthesis by two mechanisms (Goody *et al.* 1991). Firstly they act as competitive inhibitors for binding and/or catalytic incorporation with respect to the analogous dNTP substrate. Secondly, and of much more significance, after incorporation DNA synthesis is terminated due to the lack of a 3'-OH group. Since HIV-1 RT lacks a formal proof reading activity, a single NRTI incorporation event should be sufficient to suppress viral DNA synthesis. However, the emergence of HIV-1 strains resistant to this class of inhibitors is a major problem for antiretroviral therapies (Goody 1995). Three mechanisms have been proposed for resistance to NRTIs: Discrimination due to impaired binding of nucleoside analogues, p/t repositioning and, phosphorolytic removal of incorporated chain terminating NRTIs either using Pyrophosphate (Reardon 1993) or ATP (Meyer *et al.* 1998) as a "pyrophosphate" source (Figure 4.17) (for a review see (Sluis-Cremer *et al.* 2000)). In order to utilise a phosphorolytic mechanism to remove chain terminators, a binding mode analogous to the PP-complex would have to exist as an intermediate.

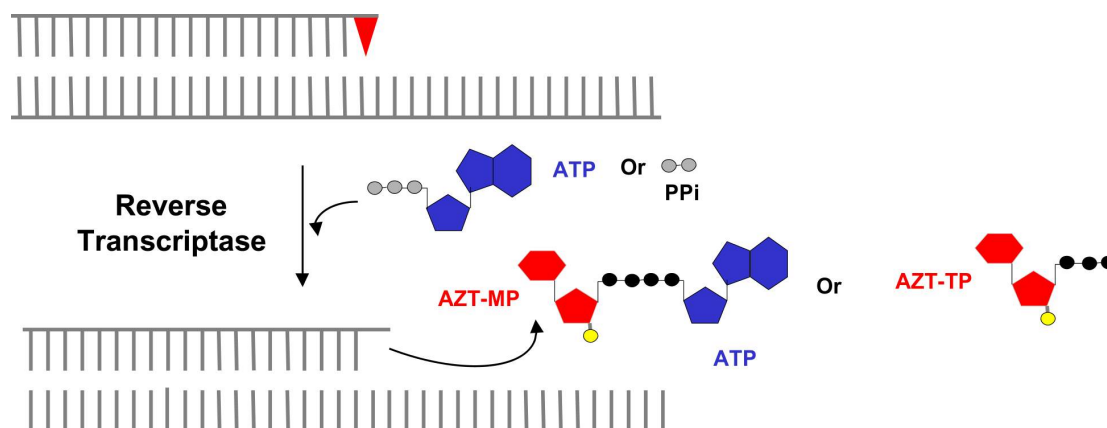


Figure 4.17. The role of pyrophosphate (PPi) and ATP in removing 3'-terminal 3'-azido-3'-deoxythymidine (AZT) from a primer. The initial event is a nucleophilic attack by PPi or the γ -phosphate of ATP on the 3'-5' phosphodiester bond between primer 3'-terminal AZT and the penultimate primer nucleotide. The reaction leads to the removal of the terminal AZT either as AZTTP or as an adenosine-AZT 5', 5' tetraphosphate, leaving a nucleotide with a free 3'-OH at the primer terminus.

The ability of RT to remove a chain terminator, in this case ddCTP, by the process of pyrophorolysis using pyrophosphate was investigated. In our studies we see no significant removal of ddCTP in the presence of pyrophosphate. Removal of ddCTP, and presumably bases thereafter, would lead to a time dependent increase in FRET efficiency as the 5' Cy5 on the primer is translocated towards the donor situated in the thumb domain of p66. However, Pyrophosphate does effect the modes of binding seen. In the presence of Pyrophosphate the P-complex is lost and replaced by the PP-complex in which the 3'-terminus is now situated in the dNTP binding pocket. The affinity for the p/t is also reduced, which can be seen from the increase of the donor only species.

It was previously suggested by Reardon (Reardon 1993) that the rate constant for pyrophorolysis by wild type RT is slow, and was similar to the rate constant for dissociation of a terminated p/t. In agreement with these results no significant pyrophorolytic activity by RT under the conditions used is seen. Therefore, additional factors are probably needed to increase the efficiency of this process. Studies using AZT terminated p/t showed that two mutations in RT associated with AZT resistance (R67N, K70R) increase the rate of pyrophorolysis. This is thought to be mediated by an increased affinity of RT for the p/t, combined with an increased rate of Pyrophosphate associated pyrophorolysis (Arion *et al.* 1998).

4.1.6.6 Role of the template overhang in p/t positioning.

It was previously shown that wild type HIV-1 RT is resistant to inhibition by ddNTPs when the template extension is short (Boyer *et al.* 1994). However, two RT mutants resistant to the NRTIs, ddI and ddGTP (L74V and E89G, respectively) showed the same resistance to inhibition independent of the template extension length. This was thought to be due to a greater discrimination between the NRTIs and the natural dNTPs due to p/t repositioning, although the mechanism of this discrimination was unclear.

In order to determine if there is a structural basis for this phenomenon, the effect of the template extension length on the heterogeneous binding modes seen between RT and its p/t substrate were examined. Shown in Figure 4.18 are the $\tau_{D(A)}$ vs R_{DA} and r_D plots for p66^{K287C[A488]}/p51[6H] in complex with: A) p_5_Cy5/t⁺³, B) p_5_Cy5/t⁺⁶, C) p_5_Cy5/t⁺⁹, and D) p_5_Cy5/t⁺¹⁶.

When the template extension length was only three we see a reduction in affinity for the P-complex and the PP-complex, as indicated by the increase in level of the low FRET species. The low FRET species is now situated close to the MDR line, and therefore probably consists of not only the DE complex but also free enzyme. The PP-complex is also very broadly distributed, indicating a disordered binding mode. The broad nature of the PP-complex for the p/t⁺³ substrate, may indicate a high degree of mobility of the p/t-substrate bound in this manner. There is also a relative increase of the PP-complex compared to the P-complex, as well as a greater distance separation between the two species compared to measurements with the 19/35 p/t (Figure 4.18D).

The rotational mobility for the P- and PP-complex is centred slightly below 2 ns, indicating a highly mobile dye. For the low FRET species two peaks are apparent, one situated at 7 ns and one situated at 0.3 ns. The species situated at 7 ns probably corresponds to the DE-complex (see Figure 4.18D and section 4.1.6.2). The species with the greatly increased mobility may be the free enzyme, as a similar behaviour is observed after the addition of Pyrophosphate

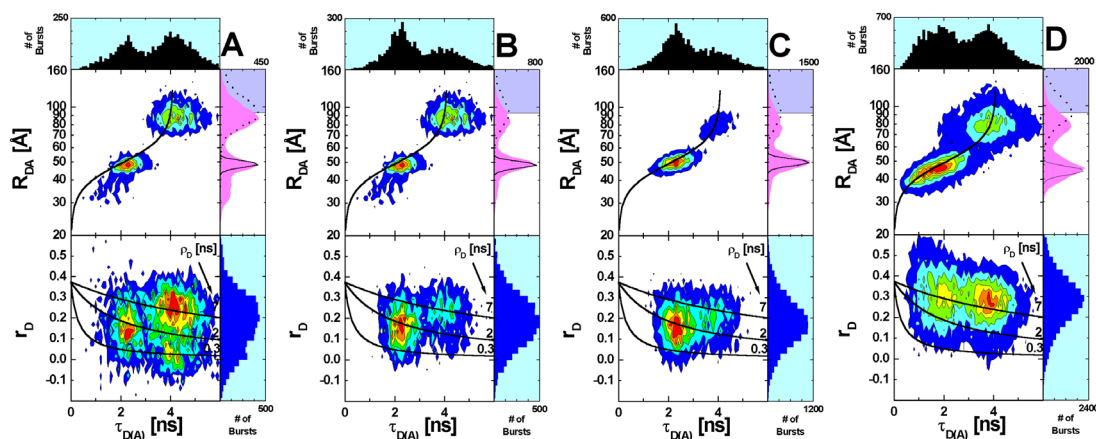


Figure 4.18. Donor lifetime, $\tau_{D(A)}$, plotted versus DA-distance, R_{DA} (upper panel) and against donor anisotropy, r_D , (lower panel) for A) 50 pM $p66^{K287C[A488]}/p51[6H]:200$ pM p_5_Cy5/t^{+3} , B) 50 pM $p66^{K287C[A488]}/p51[6H]:200$ pM p_5_Cy5/t^{+6} , C) 50 pM $p66^{K287C[A488]}/p51[6H]:200$ pM p_5_Cy5/t^{+9} , D) 50 pM $p66^{K287C[A488]}/p51[6H]:200$ pM p_5_Cy5/t^{+16} . Also plotted is the Isot line (black line-upper panel) and rotational correlation times calculated from the Perrin equation. The purple rectangle is the maximal discernible resolution we can achieve for this system. Measurements performed in SMS buffer at 25 °C

The results using a +6 overhang (Figure 4.18B) show a more ordered mode of binding. In this case, the P-complex predominates over the PP-complex and has a better defined peak at 48 Å. Although the PP-complex is still present, the amount is reduced greatly as compared to experiments with the +3 template extension p/t. The affinity for the two productive complexes has also increased, since the level of low FRET species has decreased significantly compared to the +3 overhang. It is also shifted to a value lower than the MDR, indicating that the template overhang may be important for the affinity of both P-complexes. The mobility of the P-and the PP-complex is broadly distributed, and is centred at a value of 2 ns, indicating a mobile dye. For the low FRET species one peak is apparent centred at 7 ns, characteristic of the DE complex.

For the +9 overhang (Figure 4.18C) the pattern observed between the three species is comparable to that seen with the 19/35 p/t, although there are slightly lower levels of the DE-complex. The P-species seems to be better defined with this substrate. It may be that additional interactions with a longer template lead to a broader range of binding modes being seen. r_D remains largely unchanged going from a +6 to a +9 overhang.

For the AZT resistant RTs (see section 4.1.6.5), there has been recent speculation about the role that mutations that confer resistance to AZT play in the process of pyrophorolysis (Boyer *et al.* 2001). One theory is that the mutations around the dNTP-binding site produce steric hindrance for AZT binding. However, previous kinetic evidence renders this unlikely due to the fact that there are no significant differences in affinity between dTTP and AZT (Krebs *et al.* 1997). The second theory suggests that the mutations “hold” the 3'-terminus of the p/t in the dNTP-binding pocket when AZT is incorporated by altered interaction with the template overhang, and hence allow its removal by pyrophorolysis. This is thought to be achieved either by pyrophosphate or by ATP (see Figure 4.17).

In the previous studies using shorter template overhangs, the author suggested that discrimination was due to impaired binding of NRTI's due to template repositioning (Boyer *et al.* 1994). However, the previous studies using shorter template overhangs failed to take into account the presence of pyrophosphate contaminations of nucleotide preparations. The experiments above indicate that shorter template overhangs lead to a greater separation between the P- and PP-complex. It was shown previously that the 3'-terminus of the p/t occupies the dNTP binding pocket in the case of the PP-complex. A greater occupation of this state would allow the enzyme to remove NRTI's via pyrophorolysis. Mutations implicated in drug resistance to NRTIs, which lead to the disruption of the p/t interaction, favouring the PP-complex, could lead to a configuration suitable for nucleotide removal by pyrophorolysis.

4.1.6.7 Construction of low resolution models of the productive and dead end complexes.

In order to determine a more detailed map of the two major complexes seen, the P- and DE-complex, a large number of mutants containing single cysteine were generated, and labelled with the green donor fluorophore A488 (see section 4.1.4). The labelled cysteine mutants (see Figure 4.19A) were tested with a number of different p/t-substrates of the same sequence in which the position of the acceptor fluorophore was varied (see Figure 4.19B). By combining the distance information achieved it should be possible to generate enough constraints to determine a low resolution structure for both major binding modes.

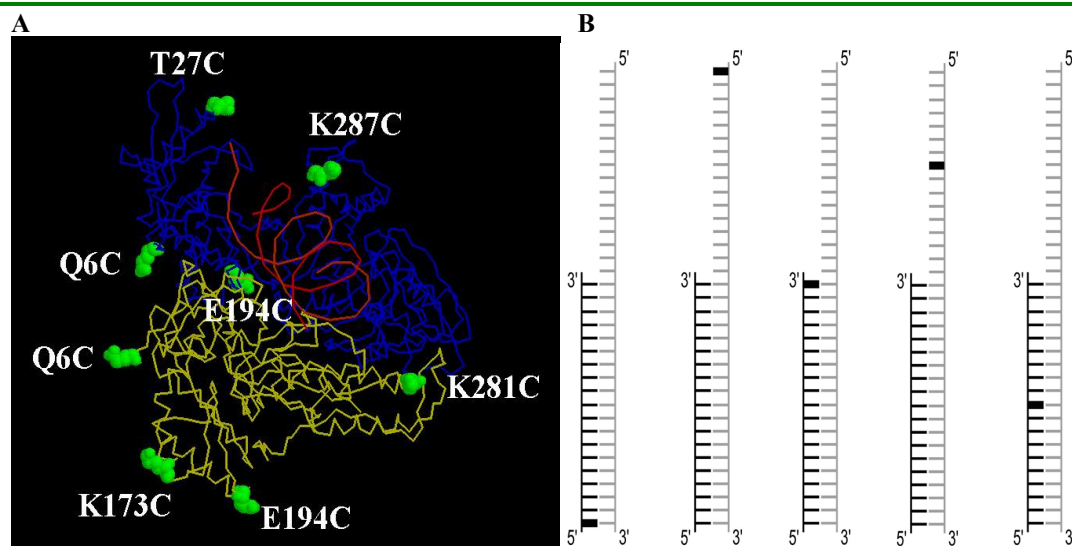


Figure 4.19. A) Position of cysteine mutants used for introduction of a site specific label. p66 is shown in blue, while p51 is shown in yellow. The residues mutated to cysteine are shown in green while the p/t is shown in red. B) DNA/DNA p/t substrates used in this study. The primer strand is shown in black, and the template in grey. The base to which the acceptor dye is coupled is shown in bold. The substrates are named from left to right: p_5_Cy5/t, p/t_5_Cy5, p_3_Cy5/t, p/t_7o_Cy5, and p_10bp_Cy5/t.

The crystal structure chosen for the basis of modelling was the ternary complex, RT:p:t:dNTP (Huang *et al.* 1998) (PDB code=1RTD). This structure was chosen due to a more detailed structural determination of the binding mode of the p/t in the P complex compared to the binary complex between RT:p/t (Jacobo-Molina *et al.* 1993) (PDB code=2HMI). That is to say the template overhang is resolved to +3 bases (cf. +1 for 2HMI), and the length of the double stranded region is longer, 21 bp (CF 18 bp for 2HMI).

4.1.6.7.1 Determination of the position of the donor fluorophore on the protein

The A488 dye is covalently linked to the protein via a C₆-bridge to an introduced cysteine residue. From the C β of the residue to the dye centre a distance of 15 Å is expected. In order to position the active centre of the dye the program “display sphere”, developed by Axel Scheidig, (see appendix II) was used in combination with the molecular modelling program “ono7” using the structure RTD1 as a “template”. The program generates a 15 Å distance shell around the C β atom of the selected residue. The C β atom was then moved to the edge of the 15 Å sphere in the direction of the amino acid side chain using the grab command in “ono7”. The A488 dye centre was then used as a fixed point to determine the position of the

acceptor on each substrate. Although to some extent this method reduces uncertainties in the donor position, it is unclear whether the enzyme adopts the same conformation for the DE-complex and the P-complex. However, results obtained using EPR spectroscopy suggest that the RT:p/t complex adopts an open conformation in the presence of a p/t (Kensch *et al.* 2000).

Species were assigned as P or DE on three criteria: The first criteria for assigning the P-complex is the similarity of the R_{DA} obtained by MFD experiments to that calculated from the crystal structure RTD1 between the C β of the A488 labelled residue to the Cy5 labelled base. The second selection criterion concerns the characteristic substrate binding properties of each labelled protein. For p66^{T27C[A488]}/p51[6H] and p66^{K287C[A488]}/p51[6H] we see approximately equal amounts of the P- and DE-complex in combination with most labelled substrates. For p66/p51[6H]^{K173C[A488]}, p66^{Q6C[A488]}/p51[6H] and p66^{E194C[A488]}/p51[6H] we see higher amounts of the P-complex compared to DE-complex. For the labelled proteins p66/p51[6H]^{Q6C[A488]}, p66/p51[6H]^{E194C[A488]}, and p66/p51[6H]^{K281C[A488]} a higher proportion of the P-complex, or a single binding mode, which based on expected distance (R_{EXP}) can be assigned as the P-complex. The third criterion is whether the distances used generate an overlap area (see section 4.1.6.7.8).

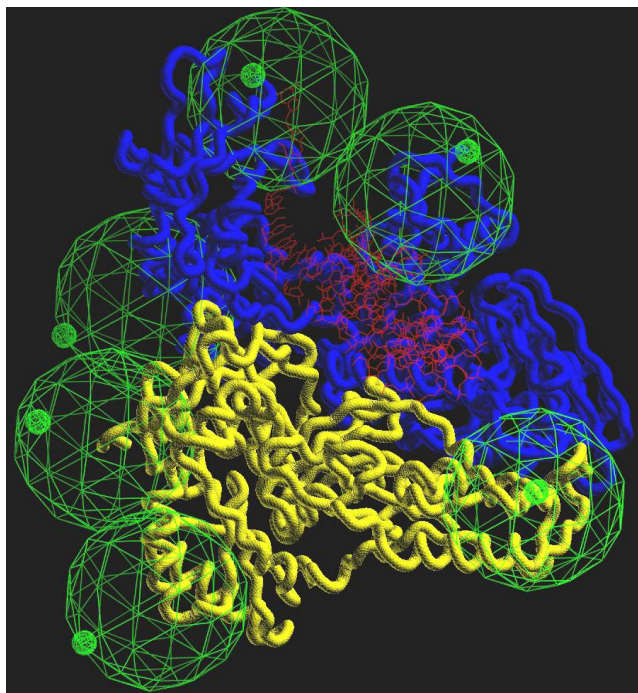


Figure 4.20. Positioning of the A488 dye. p66 is shown in blue, p51 in yellow and the p/t substrate shown in red. The 15 Å distance spheres are shown as a green wire mesh. The chosen position of the A488 dye for 6 of the labelled protein is shown as a small green sphere.

4.1.6.7.2 determination of the position of the acceptor fluorophore on the p/t

In order to obtain the overlap area and hence the position of the acceptor, the experimental distances were loaded into the program “generate_overlap”, developed by Axel Scheidig (see appendix 2). Shown in Table 4.4 is the basic input format of the program. The first three columns are defined by the xyz co-ordinates of the donor dye coupled to the protein. The 4th column is the R_{DA} obtained from the MFD experiments generated from one binding mode (i.e. P or DE). The 5th column is the width of the fitted Gaussian distribution. The program generates distance shells using the half width of the histogram where the centre is defined by the XYZ co-ordinates of the donor fluorophore. If an overlap is generated, the area of overlap is output as a pdb-file.

-200.0	200.0	-200.0	200.0	-200.0	200.0	Maximal grid size
2.0	2.0	2.0				Mesh size
90.110	177.650	65.630	69.0	12.1		p66 ^{Q6C[A488]} /p51[6H]
48.790	135.030	85.690	69.0	14.4		p66 ^{E194C[A488]} /p51[6H]
115.020	133.490	90.420	86.2	18.6		p66 ^{K287C[A488]} /p51[6H]
35.520	123.980	55.250	45.4	10.4		p66/p51[6H] ^{Q6C[A488]}
83.600	187.660	34.210	62.3	12.5		p66/p51[6H] ^{K173C[A488]} *
110.490	176.320	12.020	200.0	400.0		p66/p51[6H] ^{E194C[A488]}
123.940	138.600	12.410	62.1	7.4		p66/p51[6H] ^{K281C[A488]}
54.780	115.850	-7.040	35.2	4.4		Atom output
O						
WAT						Chain
Z						
END						

Table 4.4. Input card for the program “generate_overlap”. The example shown is for the p_5/t substrate bound in the productive manner in combination with all RT mutants. On the right hand side are shown the mutants to which the distances correspond. * K173C was not used in this analysis due to problems in overlap generation using this labelled mutant. The coloured rows correspond to the distance spheres shown in Figure 4.21.

In Figure 4.21 a simplified example of the procedure used to determine the acceptor position is shown. The coloured wireframe spheres represent the mean distance from each of the A488 labelled mutants to the Cy5 of the p_5_Cy5/t substrate bound in the P-manner. The overlap area generated, i.e. the acceptor position, is shown as a red spacefill area. For all combinations of protein mutants and p/t a similar procedure was performed in order to determine the acceptor position, and hence the nucleic acid binding mode for both the P-and DE-complex.

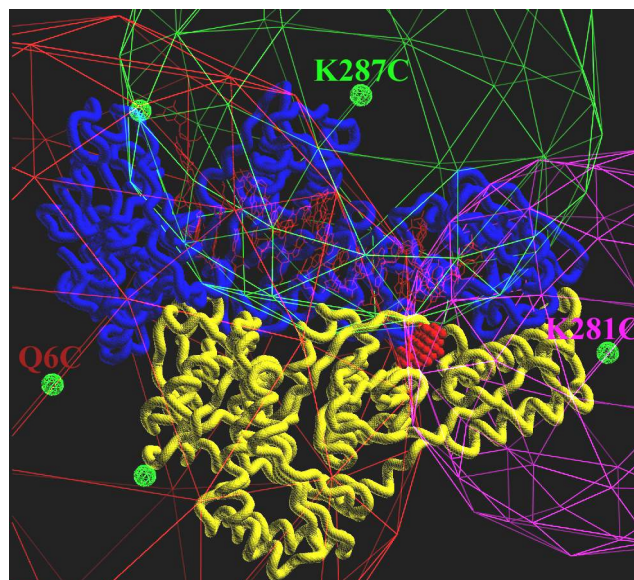


Figure 4.21. Generation of overlap area for the *p_5_Cy5/t*. RT is shown as a tube representation with the p66 subunit coloured blue and the p51 subunit coloured yellow. The p/t substrate is shown as a red wire frame, and the experimentally determined acceptor position is shown as a red spacefill. The distance shells used to generate this overlap area are shown for p66^{Q6C}/p51[6H] (red), p66^{K287C}/p51[6H] (green), and p66/p51[6H]^{K281C} (magenta).

4.1.6.7.3 Donor only samples

In order to determine the maximal discernible distance resolution (MDR), measurements were carried out using the protein alone complexed with an unlabelled DNA/DNA substrate, p/t(U). The maximal distance is not defined by the reduced Förster radius of 53 Å, in which case the value would be the same for all mutants measured. Rather it is defined to some extent by the quantum yield of the donor, spectral crosstalk, and background signal in the red channels. Values greater than 79.5 Å ($R_{0r} \pm 0.5$) are not accurate. However, if distances obtained are below the MDR, we can rule out that they are generated from the labelled protein alone.

Figure 4.22 shows the $\tau_{D(A)}$ vs R_{DA} (top panel) and r_D (lower panel) for the donor only sample (RT:p/t(U)). The table below shows the values obtained for the MDR and the donor lifetime for each labelled protein. The distances range from 90 Å to 105 Å, suggesting a slight difference in fluorescence properties for each labelled protein. That is to say, in this “distance” range a change of only a few photons recorded in the red channels has a large effect on calculated distance. The τ_D of the A488 coupled at different positions on the protein ranges from 3.9 to 4.1 ns. This indicates that in the presence of p/t(U) no significant quenching of the donor

fluorescence occurs. The expected lifetime of A488 C₅ maleimide is 4.1 ns (Berger 2002).

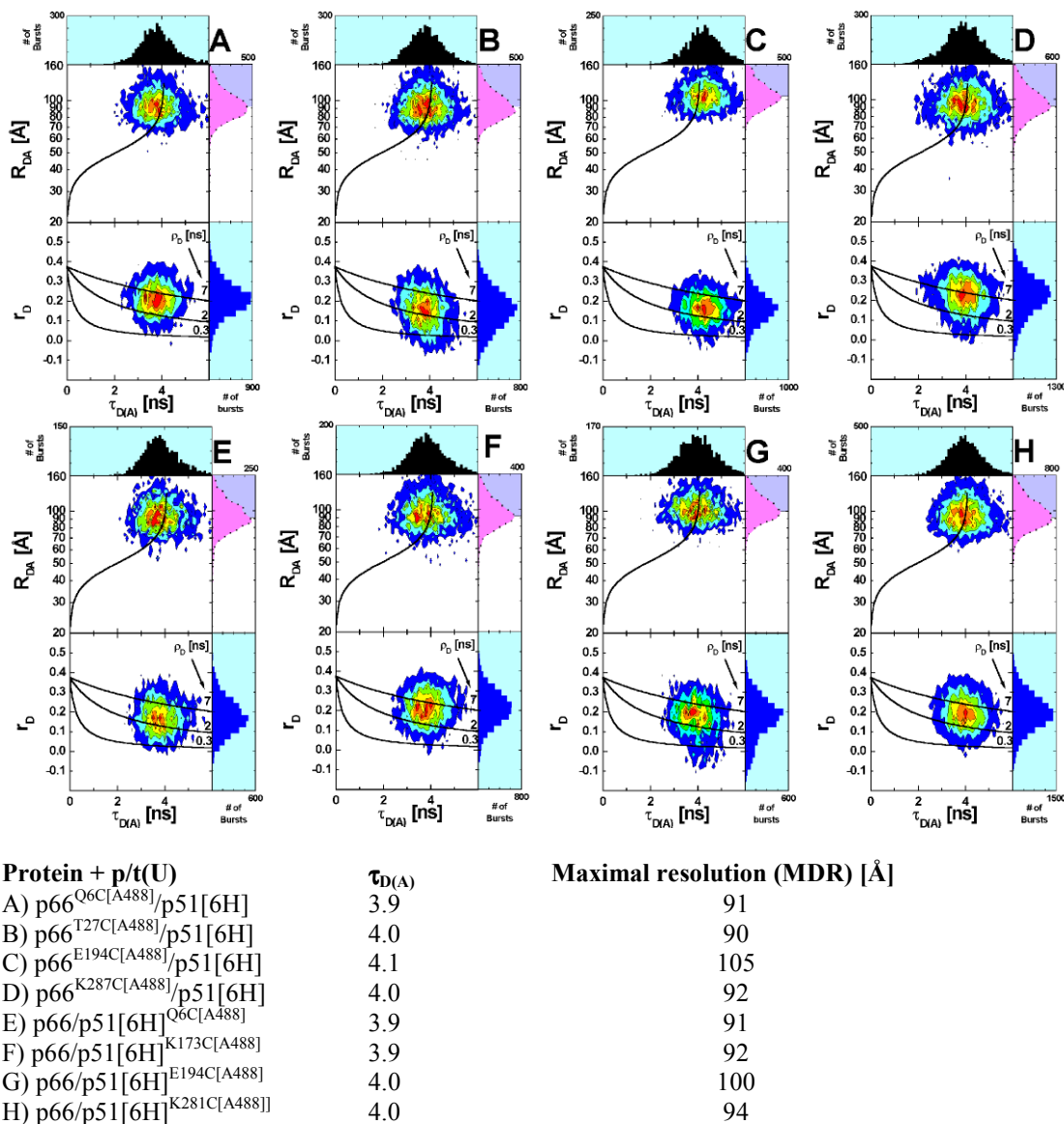


Figure 4.22. $\tau_{D(A)}$, plotted versus R_{DA} (upper panel) and against donor anisotropy, r_D , (lower panel) for different labelled proteins (indicated above) in complex with p/t(u). Also plotted is the Iso- τ line (black line-upper panel) and rotational correlation times calculated from the Perrin equation (black-line-lower panel). The purple area (1D distance plot) is the maximal discernible resolution we can achieve for each mutant. Reaction conditions 50 pM RT:200 pM p/t measured in SMS buffer at 25 °C.

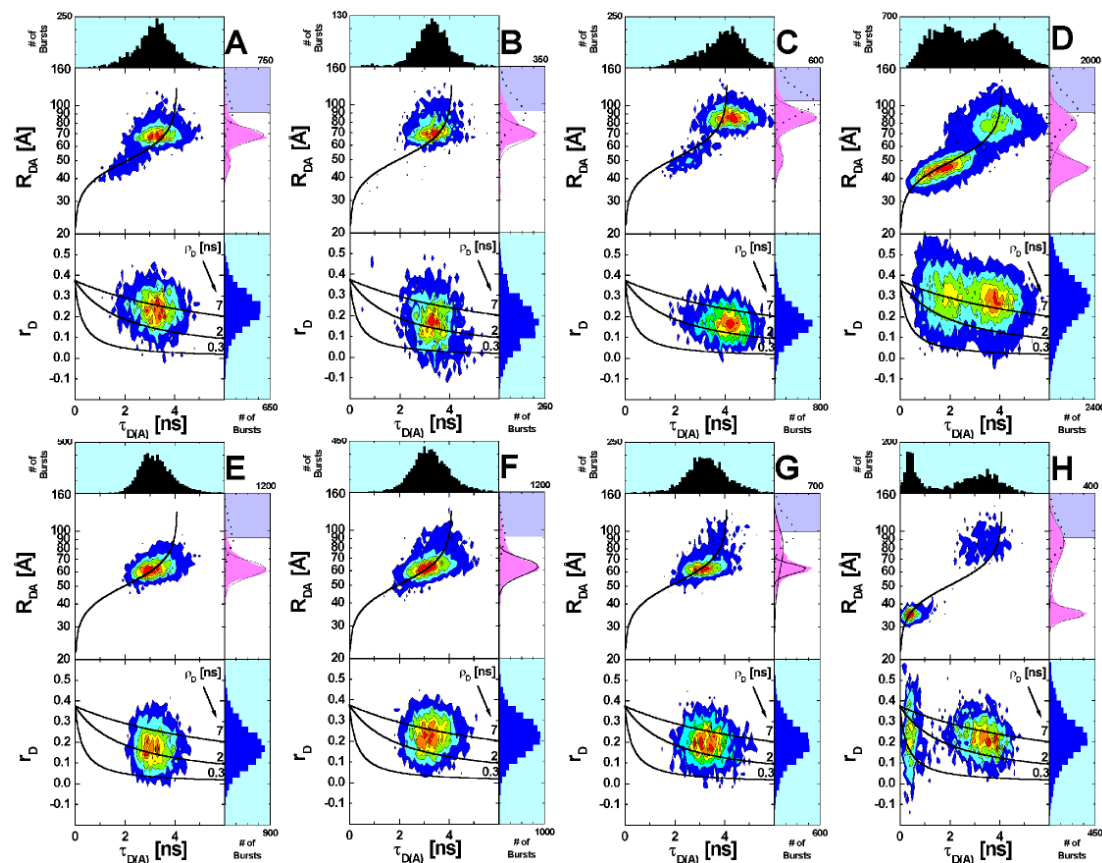
When looking at the anisotropy it can be seen that values range from rotational mobilities of 2 to 7 ns, which is within the range of the donor lifetime of 4.1 ns, hence the value of 2/3 as κ^2 is valid in all cases assuming the acceptor has free rotational mobility. For the labelled proteins, p66^{T27C[A488]}/p51[6H] (B), p66^{K287C[A488]}/p51[6H]

(D), and p66/p51[6H]^{E194C[A488]} (G) there is an additional “tail” on the anisotropy at lower values. The high rotational mobility indicates the presence of free A1488 dye in the protein preparations. The free dye is small (MW, 900Da) compared to the protein (mw 117,000 Da), and therefore does not tend to depolarise light.

4.1.6.7.4 Structural studies using the substrate p_5_Cy5/t

In Figure 4.23 are the $\tau_{D(A)}$ vs R_{DA} (upper panel) and r_D (lower panel) graphs obtained using the substrate p_5_Cy5/t, in combination with all the donor labelled mutants. The table below shows the distances obtained using MFD and their classification as P (R_{DA-P}), or DE (R_{DA-DE}) based on the expected distance from the crystal structure (R_{EXP}). Also shown are the $\frac{1}{2}$ widths obtained from fitting with a Gaussian ($\frac{1}{2} W$). On the upper right hand panel are shown the Gaussian fits for each peak, and the MDR area (purple box-upper right side panel).

For p66^{Q6C[A488]}/p51[6H] (A) two species are seen, the species centred at 69 Å was assigned as P due to R_{EXP} of 63 Å. The second complex centred at 48 Å was assigned as the DE-complex. For p66^{T27C[A488]}/p51[6H] (B) a single species is seen at 69 Å which has a broad range of mobilities, which was assigned as the P-complex. A “tail” is seen at longer distances which could be generated by the DE-complex, donor only, or free dye. Due to the fact that the peak for this species cannot be resolved, no distance was assigned for the DE complex in this case. For p66^{K287C[A488]}/p51[6H] (D) the high FRET species centred at 45 Å was assigned as the P-complex based on the R_{EXP} , and the low FRET species centred at 79 Å as the DE-complex (see section 4.1.6.2). The low rotational mobility for the DE complex centred at 7 ns is characteristic for p66^{K287C[A488]}/p51[6H], and therefore used as an additional selection criteria.



Protein + p_5_Cy5/t	R_{exp} [Å]	R_{DA-P} [Å]	$\frac{1}{2} W$ [Å]	R_{DA-DE} [Å]	$\frac{1}{2} W$ [Å]
A) p66 ^{O6C[A488]} /p51[6H]	63	69.0	12.1	47.8	8
B) p66 ^{T27C[A488]} /p51[6H]	65	69	14.4		
C) p66 ^{E194C[A488]} /p51[6H]	82	86.2*	18.6	51.7	14.5
D) p66 ^{K287C[A488]} /p51[6H]	42	45.4	10.4	79.9*	27.7
E) p66/p51[6H] ^{Q6C[A488]}	56	62.3	12.5		
F) p66/p51[6H] ^{K173C[A488]}	59	63.1**	13.6**		
G) p66/p51[6H] ^{E194C[A488]}	59	62.1	7.4	81.1**	45
H) p66/p51[6H] ^{K281C[A488]}	34	35.2	4.4	87.8*	39.4

Figure 4.23. $\tau_{D(A)}$, plotted versus R_{DA} (upper panel) and against r_D , (lower panel) for different labelled proteins (indicated above) in complex with p_5_Cy5/t. Also plotted is the Iso- τ line (black line-upper panel) and rotational correlation times calculated from the Perrin equation (black line-lower panel). The purple box is the MDR. R_{exp} is the expected distance between the donor and acceptor calculated from the crystal structure 1RTD. R_{DA-P} and R_{DA-DE} are the distances assigned for modelling of the productive and dead end complex, respectively. Also shown is the $\frac{1}{2}$ width ($\frac{1}{2} W$) of the fitted Gaussian distributions. * indicates distances outside of the range $(1 \pm 0.5) R_{0r}$. ** Distances not used for distance analysis. For the P-complex W was reduced by 50% for the acceptor position determination (section 4.1.6.7.8).

Reaction conditions 50pM RT:200pM p/t measured in SMS buffer at 25 °C.

For p66^{E194C[A488]}/p51[6H] (C) the major species centred at 84 Å was assigned as the P-species, due to the R_{EXP} of 82 Å, and the high FRET species centred at 49 Å was assigned as the DE-complex.

For p66/p51[6H]^{Q6C[A488]} (E) a single peak centred at 62.3 Å is seen, which based on the R_{EXP} of 56 Å, was assigned as the P-complex. It may be in the case of the [p66/p51[6H]^{Q6C[A488]}:p_5_Cy5/t] both the P- and the DE-complex are equidistant. However, testing of the protein p66/p51[6H]^{Q6C[A488]} using the other labelled p/t always results a single species, suggesting this mutant is only capable of binding the p/t in the P-manner. This may also explain the high activity displayed by this mutant (section 4.1.5.1). Results obtained using this labelled protein indicates that under the SMS conditions used, the affinity is high enough to ensure complete complex formation. p66/p51[6H]^{Q6C[A488]} (E) is far removed from the nucleic acid binding tract observed by crystallography and is therefore unlikely to mediate a tighter interaction through direct contact with the p/t substrate bound in the P-manner. It is more likely that it would disrupt interactions of additional protein:nucleic acid complexes. It can also be ruled out that acceptor bleaching, or incomplete labelling of the p/t-substrate are observed.

For p66/p51[6H]^{K173C[A488]} (F) one resolvable species is seen at an R_{DA} of 63.1 Å, which was assigned as the P complex due to the R_{EXP} of 59 Å. A “tail” is seen at higher distances, which was not used in the distance analysis, although it could conceivably be generated from the DE-complex. For p66/p51[6H]^{E194C[A488]} (G) the species at 62.1 Å was assigned as P based on the R_{EXP} of 59 Å. The broadly distributed species at 81.1 Å was assigned as the DE-complex. In the case of p66/p51[6H]^{K281C[A488]} (H) the high FRET species centred at 35 Å ($R_{EXP} = 34$ Å) was assigned as P. The second species at 87.8 Å was assigned as the DE-complex.

4.1.6.7.5 Structural studies using the substrate p_10bp_Cy5/t

For this p/t the Cy5 is coupled 10 bases away from the 5' end of the primer strand. This position was chosen due to the presence of a thymidine in the sequence at this position, as well as the solvent accessibility seen for this position in the crystal structure.

For p66^{Q6C[A488]}/p51[6H] (A) two species are seen. The high FRET species at 42.8 Å was assigned as P based on the R_{EXP} of 39 Å. The minor species at 80 Å was assigned as the DE complex. For p66^{T27C[A488]}/p51[6H] (B) and p66^{K287C[A488]}/p51[6H] (D) the high fret species was defined as the P complex, due to the R_{EXP} . The low FRET species for both mutants was therefore assigned as the DE complex. In the case

of p66^{E194C[A488]}/p51[6H] (C) the major species centred at 49.7 Å was classified as P due to the R_{EXP} and the low FRET species at 92 Å as the DE.

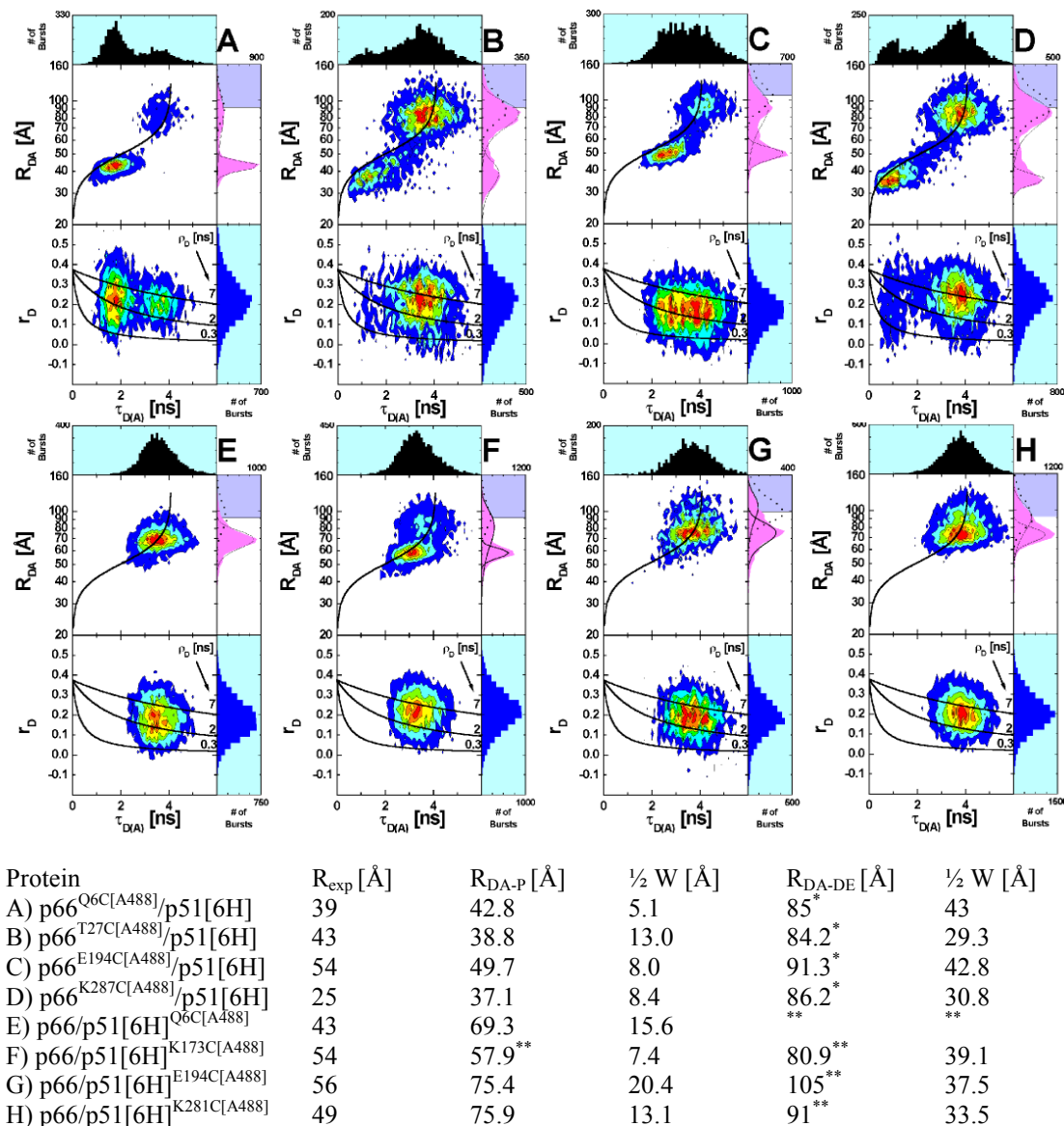


Figure 4.24. $\tau_{D(A)}$, plotted versus R_{DA} (upper panel) and against r_D , (lower panel) for different labelled proteins (indicated above) in complex with p_{10bp}-Cy5/t. (for additional plot details see Figure 4.24). Reaction conditions 50 pM RT:200 pM p/t measured in SMS buffer at 25 °C.

Binding of the p/t substrate by the p66/p51[6H]^{Q6C[A488]} (E) is again characterised by a single species located at 69.3 Å. There is a large variation between this value and the one expected from the crystal structure. However, when taking the combined length of both dye linkers into account, 30 Å, it is possible that this species is generated from the P-complex. The R_{DA} value was therefore taken for the P-

complex. For p66/p51[6H]^{K173C[A488]} (F) the species centred at 57.9 Å was assigned as P, and the broadly distributed species at 80.9 Å was assigned as DE. For both p66/p51[6H]^{E194C[A488]} (G) and p66/p51[6H]^{K281C[A488]} (H), a single species is seen with a “tail” at longer distances, which is centred on/or above the MDR line. Although there is a large difference between the measured values for each mutant and the R_{EXP} the species centred below the MDR line were assigned as P. The second species in both cases is at a distance characterised by the donor only sample. This indicates that this species may be either free protein or a non-FRET showing complex, due to large distances.

4.1.6.7.6 Structural studies using the substrate p_3'_Cy5-t

By using the substrate with the label at the 3'-terminus the double stranded region of the p/t can be fully characterised. For p66^{Q6C[A488]}/p51[6H] (A) the major species centred at 43 Å was assigned as the P-complex based on the R_{EXP} of 24 Å, and the minor species centred at 74.8 Å assigned as DE. For p66^{T27C[A488]}/p51[6H] (B) and p66^{K287C[A488]}/p51[6H] (D) a disrupted binding mode is seen. The high FRET species should correspond to the P-complex even though in the case of both labelled mutants the level of this species is low. The reason for this is unclear. It would be expected that the bulky Cy5-dye on the 3'-terminus may disrupt positioning of the p/t, however this would be the case for all labelled proteins. It could be the case that the mobility of the dye is restricted due to close proximity of both dyes to one another. However, the rotational mobilities are within an acceptable range. Nevertheless, the minor species (high FRET) was assigned as the P complex, as even when taking into account the additional length of the both dye linkers (30 Å) it is impossible to reach a value of 84 Å or 87 Å for p66^{T27C[A488]}/p51[6H] (B) and p66^{K287C[A488]}/p51[6H] (D), respectively. Therefore the species with the high R_{DA} for these mutants was assigned as DE-complex. For p66^{E194C[A488]}/p51[6H] (C) the species at 49.7 Å was assigned as P due to the R_{EXP} of 54 Å, and the low FRET species was assigned as DE. For p66/p51[6H]^{Q6C[A488]} (E), a single peak at 69.3 Å is seen, corresponding to the P-complex. The major species at 56.1 Å for p66/p51[6H]^{K173C[A488]} was set as the P-complex, while the minor species at 90 Å was assigned as the DE complex. For p66/p51[6H]^{E194C[A488]} (G) the species at 71.4 Å was assigned as the P-complex based on the R_{EXP} and the species at 87.8 Å was assigned as the DE-complex. For

p66/p51[6H]^{K281C[A488]} the major peak at 75.4 Å due to the R_{EXP} was set as the P-complex.

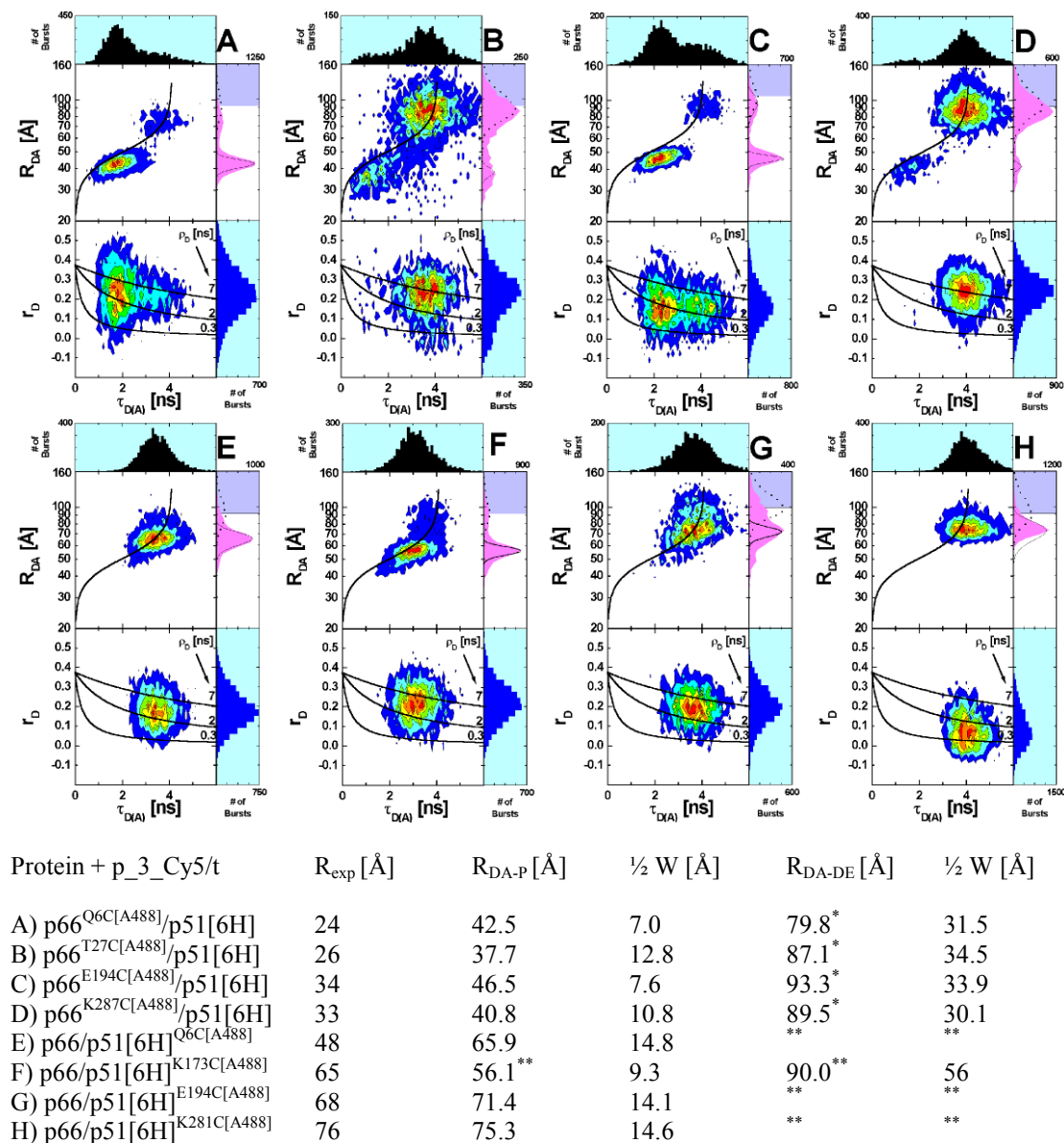
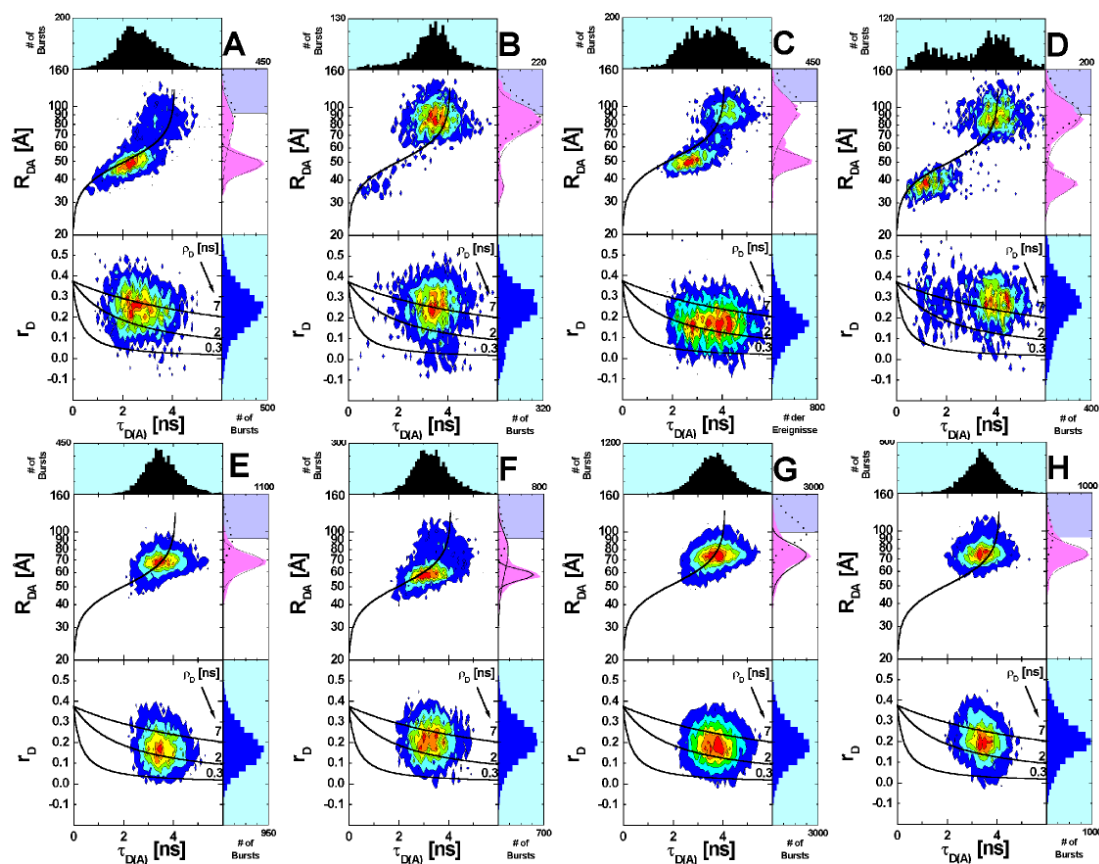


Figure 4.25. $\tau_{D(A)}$, plotted versus R_{DA} (upper panel) and against r_D , (lower panel) for different labelled proteins (indicated above) in complex with p_3_Cy5/t. . (for additional plot details see Figure 4.24). For the P-complex W was reduced by 50% for the acceptor position determination. Reaction conditions 50 pM RT:200 pM p/t measured in SMS buffer at 25 °C.

4.1.6.7.7 Structural studies using the substrate p/t_7o_Cy5

For the substrate p/t_7o_Cy5 it is the first case in which distances cannot be directly assigned due to R_{EXP}. Therefore the R_{EXP} shown in Figure 4.26 are calculated from each mutant to the 3rd base in the template overhang.



Protein +p/t τ_{70_Cy5}	R_{exp} [Å]	R_{DA-P} [Å]	$\frac{1}{2} W$ [Å]	R_{DA-DE} [Å]	$\frac{1}{2} W$ [Å]
A) p66 ^{Q6C[A488]} /p51[6H]	39	49.2	9.9	81.6*	40.5
B) p66 ^{T27C[A488]} /p51[6H]	12	37.5	9.2	87.2*	30
C) p66 ^{E194C[A488]} /p51[6H]	59	50.2	9.0	92.7*	39
D) p66 ^{K287C[A488]} /p51[6H]	27	38.3	9.0	90.4*	34.5
E) p66/p51[6H] ^{Q6C[A488]}	43	68.6	15.7	**	**
F) p66/p51[6H] ^{K173C[A488]}	80	58.1**	8.0**	77.6**	40.4
G) p66/p51[6H] ^{E194C[A488]}	89	74.8	21.4	**	**
H) p66/p51[6H] ^{K281C[A488]}	80	75.2	17.7	**	**

Figure 4.26. $\tau_{D(A)}$, plotted versus R_{DA} (upper panel) and against r_D , (lower panel) for different labelled proteins (indicated above) in complex with p/t τ_{70_Cy5} (for additional plot details see Figure 4.24). For the P complex W was reduced by 50% for the acceptor position determination. Reaction conditions 50 pM RT:200 pM p/t measured in SMS buffer at 25 °C.

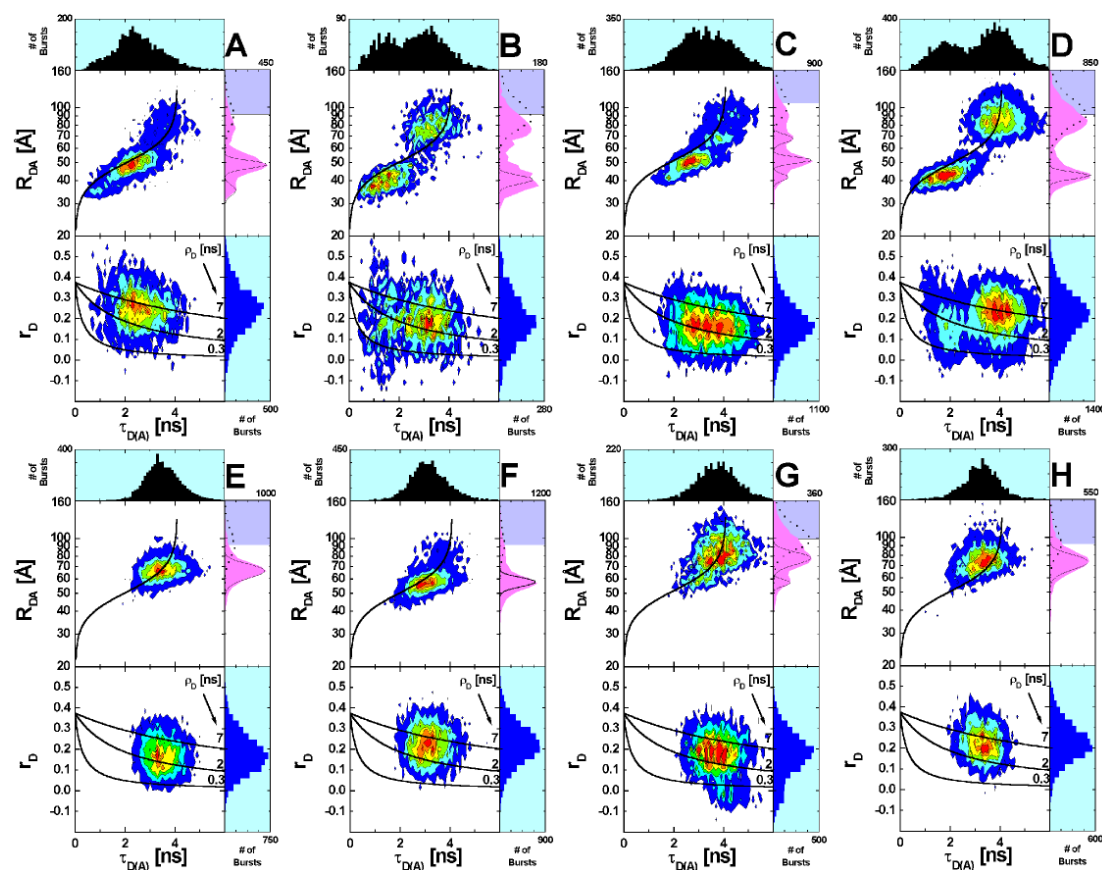
For p66^{Q6C[A488]}/p51[6H] (A) the major complex was located at an R_{DA} of 48.6 Å and was assigned as the P-complex. For p66^{T27C[A488]}/p51[6H] (B) there is a low amount of a high FRET species. This was assigned as the P-complex due to the R_{EXP} . It would be virtually impossible to reach a distance of 80 Å away from this mutant even if the template overhang is orientated away from this position. If the template orientates along the fingers of p66, there may be a reduction in affinity for the

productive complex due to interactions between the two donor and acceptor fluorophores. For p66^{K287C[A488]}/p51[6H] (D) there is an approximately equal amount of a high FRET and a low FRET species. The high FRET species centred at 38 Å was assigned as P, and the low FRET species with a low rotational mobility was assigned as DE. The single binding mode seen for p66/p51[6H]^{Q6C[A488]} (E) was assigned as the P-complex. For p66/p51[6H]^{K173C[A488]} (F) the major complex at a distance of 58 Å was assigned as the P-complex, while the low FRET species was assigned as the DE-complex. For both p66/p51[6H]^{E194C[A488]} (G) and p66/p51[6H]^{K281C[A488]} (H) the single resolvable species was assigned as P.

4.1.6.7.8 Structural studies using the substrate p/t_5'_Cy5

Although no distance information is available for the position of the acceptor for this substrate bound in the P-manner, due to the characteristics of each mutant it was possible to assign species as P and DE. For p66^{Q6C[A488]}/p51[6H] (A) the major complex centred at an R_{DA} of 49 Å was assigned as the P-complex, due to the high level of this species. The DE complex assignment was difficult due to the existence of two additional species centred at 35 Å, and 77 Å. Initially, the high FRET species centred at 35 Å was assigned as the DE complex, however no overlap could be generated for the acceptor position (see section 4.1.6.7.8). Therefore the low FRET-species was assigned as the DE-complex. The high FRET-species maybe generated from the PP-complex, or may represent a different conformation of the template. For both p66^{T27C[A488]}/p51[6H] (B) and p66^{K287C[A488]}/p51[6H] (D), there is an approximately equal amount of both species. The low FRET species in both cases was assigned as the DE complex due to the low mobility of the A488 dye, a characteristic for this binding mode. The high FRET- species was therefore assigned as P. For p66^{E194C[A488]}/p51[6H] (C), a complex binding pattern is seen. The distance distribution reveals at least three complexes. The major species located at a distance of 52 Å, was assigned as the P-complex. Due to the complex binding behaviour seen for this mutant, a distance was not assigned for the DE-complex. For p66/p51[6H]^{Q6C[A488]} (E) a single binding mode was seen which was assigned as the P complex. For p66/p51[6H]^{K173C[A488]} (F) the major species centred at 56.9 Å was assigned as P, and the minor low FRET species as the DE complex. For p66/p51[6H]^{E194C[A488]} (G), two species are seen. Neither of the distances generated from this mutant resulted in an overlap area so was not used in modelling of the P-

and DE-complex. Also, the anisotropy for this measurement indicates the presence of either free dye, or more likely a “dirt” contamination in the measurement solution. This is indicated by the peak seen at an r_D of 0. For p66/p51[6H]^{K281C[A488]}, a single peak is seen at 74 Å, and due to the binding characteristics of this mutant was assigned as the P-complex.



Protein + p/t_5_Cy5	R_{exp} [Å]	R_{DA-P} [Å]	$\frac{1}{2} W$ [Å]	R_{DA-DE} [Å]	$\frac{1}{2} W$ [Å]
A) p66 ^{Q6C[A488]} /p51[6H]	39	49.0	10.6	79.7*	31.9
B) p66 ^{T27C[A488]} /p51[6H]	12	40.2	10.3	79.6*	29.4
C) p66 ^{E194C[A488]} /p51[6H]	59	51.0	7.7	**	**
D) p66 ^{K287C[A488]} /p51[6H]	27	42.3	11	80.3*	24,5
E) p66/p51[6H] ^{Q6C[A488]}	43	61.7	15	**	**
F) p66/p51[6H] ^{K173C[A488]}	80	56.9**	9.5	73.9**	41.2
G) p66/p51[6H] ^{E194C[A488]}	89	76.5**	11.2**	59**	4.8**
H) p66/p51[6H] ^{K281C[A488]}	80	75.1	21.2	**	**

Figure 4.27. $\tau_{D(A)}$, plotted versus R_{DA} (upper panel) and against r_D , (lower panel) for different labelled proteins (indicated above) in complex with p/t_5_Cy5. (for additional plot details see Figure 4.24). For the P-complex W was reduced by 50% for the acceptor position determination. Reaction conditions 50 pM RT:200 pM p/t measured in SMS buffer at 25 °C.

4.1.6.7.9 The structure of the P and DE complexes

The results of the overlap analysis are shown in Figure 4.28 for the P-complex and Figure 4.29 for the DE-complex. The protein is displayed in spacefill representation with p66 coloured blue and p51 coloured yellow. The base to which the dye is coupled is shown in a red stick representation. For the two template labelled substrates, p/t_7o_cy5 and p/t_5'_Cy5, the last base (+3) resolved in the ternary complex (RT:p/t:dNTP) is shown (Huang *et al.* 1998). The labelled proteins used to generate the overlap areas are indicated in Figures 4.23-4.27. p66/p51[6H]^{K173C[A488]} (F) was not used in the analysis due to difficulties in generation of overlap areas, suggesting the positioning of the donor fluorophore was incorrect.

The overlap area generated for the p_5_Cy5/t-substrate bound in the productive manner is shown in Figure 4.28A. The overlap area is situated approximately 11 Å from the base to which the acceptor is coupled. Taking into account the dye linker, there is close agreement between the overlap area generated and the crystal structure. For p_10bp_Cy5/t (Figure 4.28B) a large overlap area is generated, even though all distances obtained for the P-complex are within $(1 \pm 0.5) R_{0r}$. The distance between the base and the overlap area at the closest distance is 11 Å, and at the furthest distance is 29 Å. However, the centre of the overlap area is situated close to the base to which the dye is coupled. In the case of the p_3_Cy5/t (4.28C) all distances obtained were also in the range of $(1 \pm 0.5) R_{0r}$. The overlap area generated is small and is situated approximately 11 Å from the 3'-terminus, which when taking the additional length of the linker into account is in close agreement with the crystal structure.

For the substrates in which the Cy5 is positioned in the template overhang, p/t_7o_Cy5 (Figure 4.28D) and p/t_5'_Cy5 (Figure 4.29E) a similar behaviour is seen. The overlap area generated is situated in the DNA binding cleft of RT, suggesting a specific interaction of the acceptor fluorophore with the protein. The data for the p/t_5'_Cy5 however revealed a complex binding behaviour (Figure 4.27). For the labelled proteins p66^{Q6C[A488]}/p51[6H] (A), p66^{E194C[A488]}/p51[6H] (C), and p66/p51[6H]^{E194C[A488]} (D), an additional complex is seen at shorter distances. For the above mentioned proteins, all donor fluorophores are situated within a similar area and may be therefore detecting a different binding mode. If the other labelled proteins are too far from the acceptor position the binding mode would not be seen, but however would lead to a species at a higher R_{DA} due to a non-FRET species. The

centre of a non-FRET species would be centred at the MDR border. When comparing the measurements using T27C(B) and K287C(D) in combination with the p_5_Cy5/t (Figure 4.23) and p/t_5_Cy5 (Figure 4.27) it can be seen that the low FRET (high R_{DA}) distribution is more asymmetric in the case of the p/t_5_Cy5 substrate which may be indicative of an additional complex.

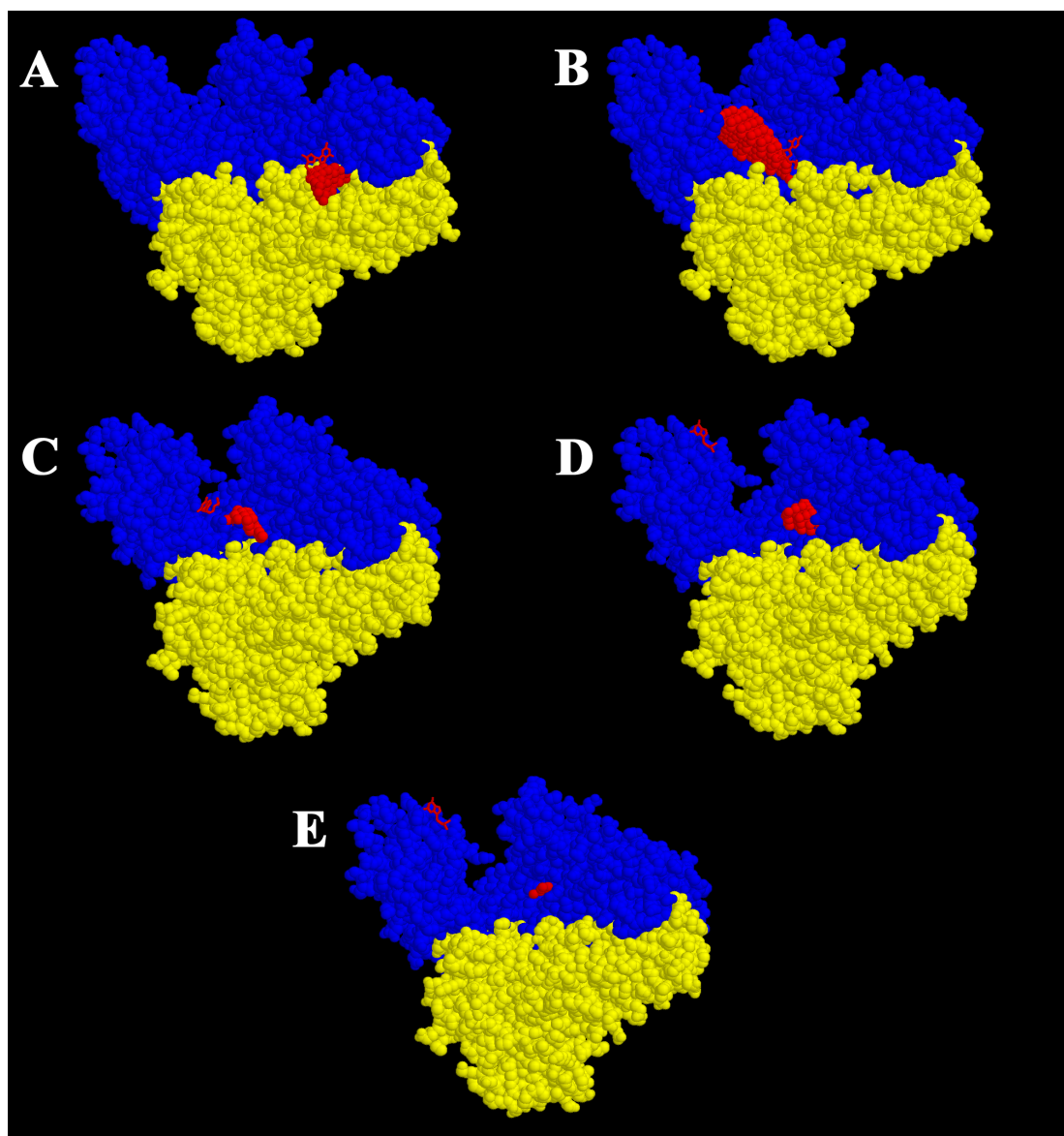


Figure 4.28. Modelling of the productive complex. p66 is coloured in blue, and p51 in yellow. The overlap area for the acceptor dye is shown as a red spacefill area and the base to which the acceptor dye is coupled is shown as a red stick representation. A) p_5_Cy5/t, B) p_10bp_Cy5/t, C) p_3_Cy5/t, D) p/t_7o_Cy5, and E) p/t_5_Cy5. For the two template labelled substrates the last base resolved (+3) from the crystal structure is shown.

From the structure of the ternary complex it is known that the template interacts with RT for at least 3 bases into the overhang (Huang *et al.* 1998). Biochemical studies on wild type RT and drug resistant mutants indicate that the template is important for positioning of the p/t to at least 6 bases into the overhang (Boyer *et al.* 1994). This is also supported by the studies undertaken here (section 4.1.6.6). From the results it is unclear how many bases of the template overhang interact with the protein, due to an interaction of the Cy5 label with the DNA binding tract. However, in the case of the p/t_5_Cy5 measurements additional complexes are seen. If the low R_{DA} species seen are due to the P-mode of binding, it would indicate that the template would curve around the fingers of p66, and would be located close to the N-terminus of p66 near the interface with p51.

After modelling of the P-complex, the remaining distances were used to generate overlap areas for the acceptor bound in the DE-mode. From the overlap areas generated it was impossible to develop a model for this binding mode. For only one substrate, p_5_Cy5/t (Figure 4.29) the distance information obtained is within the range $(1 \pm 0.5) R_{0r}$ to give a high degree of certainty about the position of the substrate. For the other substrates most of the distance information obtained is characterised by a large R_{DA} values. The pdb files generated were of large size and the overlap areas seen covered a large area of the protein. It is unclear why, in the case of the second complex, the R_{DA} values are large. There could be several reasons for this effect:

- 1) Non fluorescent acceptor molecule. Experiments using p66/p51[6H]^{Q6C[A488]} indicate a single binding mode for this labelled mutant. With all the substrates tested the single species seen is well separated from the MDR. This indicates that the Cy5-molecule is not bleached during the measurement time, and rules out the presence of a non-fluorescent acceptor. A non fluorescent acceptor molecule would reduce the green intensity, and would lead to a relative decrease in F_D/F_A and hence a reduction in distance.
- 2) Donor or acceptor quenching. Due to the type of analysis used (see section 4.1.6.2) it can be ruled out that donor or acceptor quenching occurs. A decrease in the recorded acceptor signal would lead to an increase in F_D/F_A , and therefore lead to species above the MDR line. A decrease in donor fluorescence due to quenching would lead to a reduction in τ and shift species seen to the left of the Iso- τ line.

- 3) Rotational effects. By recording the r_D it is possible to determine if rotational effects of the donor fluorophore affect the R_{DAS} obtained. In the case of two labelled proteins, $p66^{T27C[A488]}/p51[6H]$ and $p66^{K287C[A488]}/p51[6H]$ there is a reduction in mobility of the donor fluorophore for the DE-complex, which could affect the distances calculated. For the other labelled proteins there seem to be no rotational differences between the DE- and P- complexes. However, it cannot be ruled out that the acceptor dye is rotationally restricted.

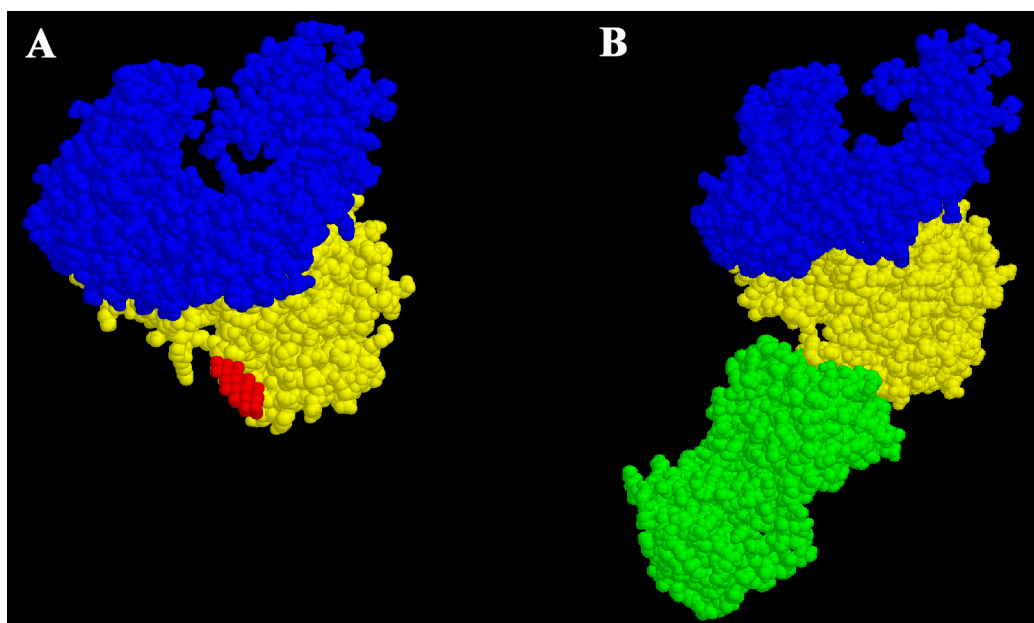


Figure 4.29. A) Modelling of the dead end complex for the substrate p_5_Cy-t . $p66$ is coloured in blue, and $p51$ yellow. The overlap area for the acceptor dye is shown as a red spacefill area. B) Structure of $[Fab:RT-p/t]$ (2HMI) (Jacobo-Molina et al. 1993). The Fab fragment is shown in green.

From the structural data obtained for p_5_Cy5/t it appears as if the 5'-end of the template is bound by the palm of the $p51$ subunit. The initial studies using $p66^{K287C[A488]}/p51[6H]$ (section 4.1.6.4) indicated that the substrate is interacting with the protein in this area. Using the other labelled proteins it was possible to reduce the uncertainty in the position of the substrate in the case of p_5_Cy5/t . It appears as if the label in this case is positioned at a site close to the Fab fragment used to crystallise the binary complex (RT:p/t). This may explain why the strategy of stabilising the RT:p/t complex using a Fab fragment was successful. Most of the conclusions which can be drawn from this binding mode were discussed in section 4.1.6.4.

4.2 Subunit interactions and conformational changes

4.2.1 Cloning of p66[6H-SII], and purification of the individual subunits

It was previously observed that RT adopts two distinct conformations in solution (Kensch *et al.* 2000) In order to study different conformational states of the RT using fluorescence methods, it was necessary to label a single protein with a donor fluorophore and an acceptor fluorophore. To label both subunits separately with different fluorophores a system for purification of the individual subunits was required.

For the p51 subunit this is made possible by the use of the N-terminal His-tag. However, the conventional p66 purification procedure requires 4 days and the amounts of recovered protein are low (Restle 2002). Therefore an expression vector was created, p6HRT166[SII] (see Figure 4.30 and section 3.1.12), from which the protein produced (p66[6H-SII]) contains an N-terminal His-tag and a C-terminal Strep-tag II. The protein can be purified in less than one day using affinity chromatography.

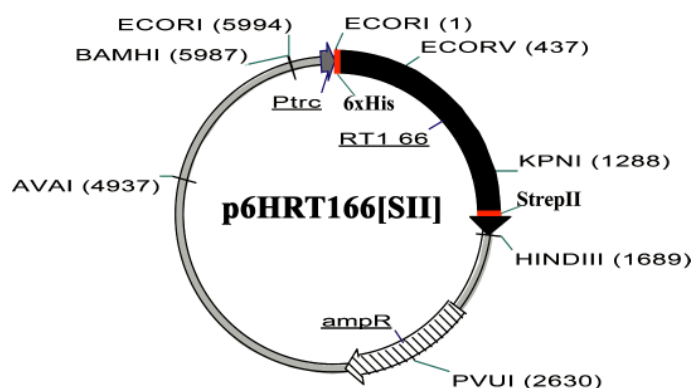
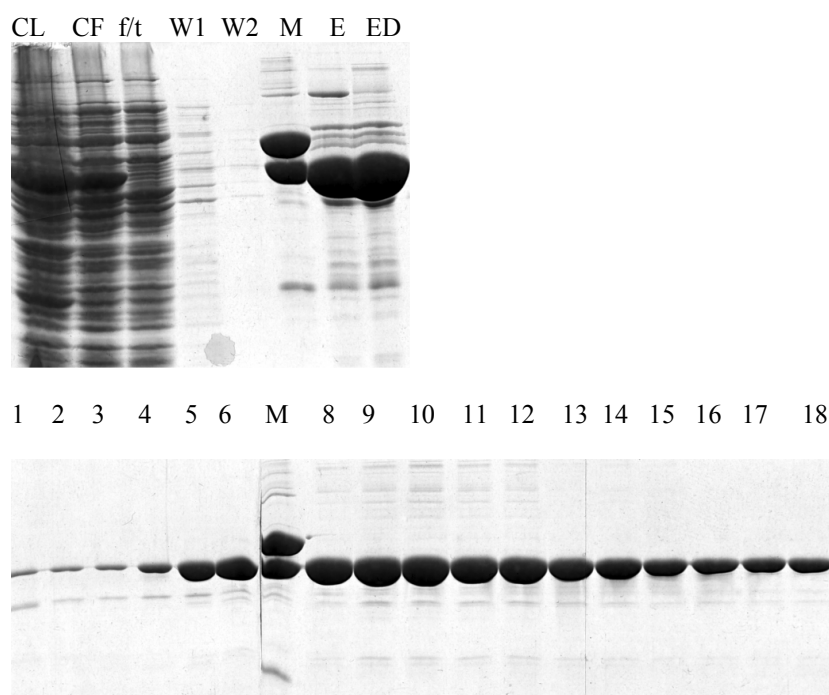


Figure 4.30. Expression vector for production of p66[6H-SII].

4.2.1.2 Purification of the p51[6H] subunit

The p51 purification (section 3.2.3) is performed in essentially the same manner as for the heterodimer (see section 3.2.2). Shown in Figure 4.31 is a 10 % SDS-polyacrylamide gel for the purification of p51[6H]^{Q6C}. After loading of the cellular lysate (CL) and washing with different buffers (W1 and W2). p51 is eluted from the Ni-NTA column using 200 mM imidazole and a single fraction collected

(lane E). After overnight dialysis against RT-D-10, a precipitate was formed, but as can be seen from lane ED, this mostly consists of high molecular weight proteins and not p51[6H]. After buffer exchange to RT-D-10, p51 is purified as for section 3.2.2, however a higher NaCl concentration is needed to elute the p51 subunit from the SP-Hitrap column (70mM NaCl cf. 50mM for heterodimer). Fractions 6-17 were pooled, and concentrated. The purity of p51[6H] was greater than 99% and approximately 30 mg of p51 were obtained from 10 g of cells.



CL=cellular lysate, CF=cytosolic fraction, f/t=flow through, W1= wash 1 (buffer W + 1 M NaCl), W2=wash 2 (buffer W + 5 mM Imidazole), M=5 μ g RT marker, E,=elute from Ni-NTA column, ED=elute after overnight dialysis into RT-D-10, n=fraction number

Figure 4.31. Purification p51[6H]^{Q6C} using Ni-NTA (upper panel), followed by purification using a combination of DEAE-Sepharose and SP-hitrap columns (lower panel).

4.2.1.2 Purification of the p66[6H-SII] subunit

Purification of p66[6H-SII] using two affinity columns ensures that only full-length p66 is obtained. Proteolytic degradation is to some extent prevented by the use of PMSF but is not completely blocked. The p66 subunit is initially purified using a Ni-NTA column, which also has the advantage that biotin is removed before using the streptactin column. Biotin binds to streptactin irreversibly resulting in inactivation of the column. In the second step the C-terminal Strep-tag II of p66 is selected for using

the streptactin column. Purification by this method ensures that only full length p66 is obtained.

No SDS-PAGE is performed during the purification to ensure a faster purification time. p66 purified by this method is over 99 % pure (see Figure 4.34) and protein yields are typically in the range of 1-2 mg/10 g cells. However due to the higher OD₅₉₅ value obtained from p66[6H-SII] expressing cells compared to p66 expressing cells, a 10 litre expression produces over 100 g cells (cf. p66 10-15 g). Therefore, the net yield of p66 expressing cells is higher. The p66[6H-SII] and p51[6H] mutants produced during this work are shown in table 4.5

<i>p66[6H-SII]</i>	<i>p51[6H]</i>
p66 ^{C38S/C280S/K287C}	p51[6H] ^{/C280S/K281C}
p66 ^{C38S/C280S/W24C/K287C}	p51[6H] ^{/C280S/Q6C}
p66 ^{C38S/C280S/Q6C}	

Table 4.5. Monomeric proteins produced during this work.

4.2.2 Protein labelling

To study different conformational states of RT, two dual labelling strategies were chosen using the mutants p66[6H-SII]^{K287C}/p51[6H]^{Q6C} and p66[6H-SII]^{K287C}/p51[6H]^{K281C} (Figure 4.32). The positions were chosen due to distance changes seen between the mutants upon movement of the thumb subdomain of p66 (see Figure 4.32).

When the enzyme is in the closed conformation, a distance of 53 Å is expected between Q6C(p51)-K287C(p66) and for K281C(p51)-K287C(p66) a distance of 75 Å. On opening of the binding cleft the distance increases for Q6C(p51)-K287C(p66) to 63 Å leading to a change in transfer efficiency of approximately 20 %, and reduces to 60 Å for K281C(p51)-K287C(p66) leading to a change in transfer efficiency of 28 % see Figure 4.33.

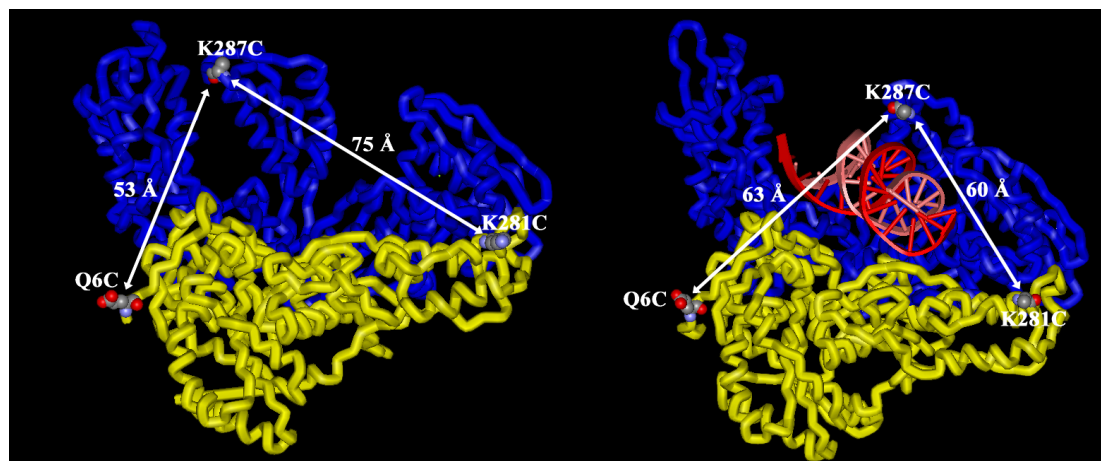


Figure 4.32. Systems chosen for dual labelling of RT. The p66 subunit (blue) and the p51 subunit (yellow) are shown as a tube representation. The positions labelled with a fluorophore are shown as spacefill. The changes in distances on going from closed unliganded enzyme (left) to open DNA bound enzyme (right) are also displayed.

p66[6H-SII]^{K287C} was labelled with A488 and A594, and the monomeric p51[6H]^{Q6C} and p51[6H]^{K281C} were labelled with A633 and A488 respectively, as described in section 3.2.5.

In a first step, excess dye was removed from the A488 or A594 labelled p66[6H-SII]^{K287C} by gel filtration. Then the protein was immobilised on the streptactin affinity matrix via the C-terminal Strep-tag II of p66, followed by washing with a high salt buffer to disrupt ionic interactions. This method resulted in a labelling efficiency of 97 % suggesting less free dye was present after purification, as compared to labelling of the heterodimer (section 4.1.4).

For p51[6H]^{K281C}, a low labelling efficiency of 50 % was observed. The reason for the low labelling efficiency seen for p51[6H]^{K281C} is unclear since even if the p51 subunit had formed a homodimeric enzyme, K281C should be accessible in the p51 conformation, and also in the “pseudo” p66 like conformation. It cannot be ruled out that monomeric p51 has a different structure to that found in the heterodimeric enzyme, reducing the accessibility of the introduced cysteine.

The labelling of p51^{Q6C} with A633 was also problematic. After addition of the A633 dye, a precipitate formed during the labelling reaction. This was removed using centrifugation and free dye was removed by gel filtration over a 30 ml Sephadex G25 column (buffer W+1M NaCl). The labelling efficiency was 100 % for p51[6H]^{Q6C[A633]}. An interesting aspect of the p51 labelling experiments is the fact that

only the introduced cysteine is labelled suggesting C38 is buried even in the monomeric structure of p51.

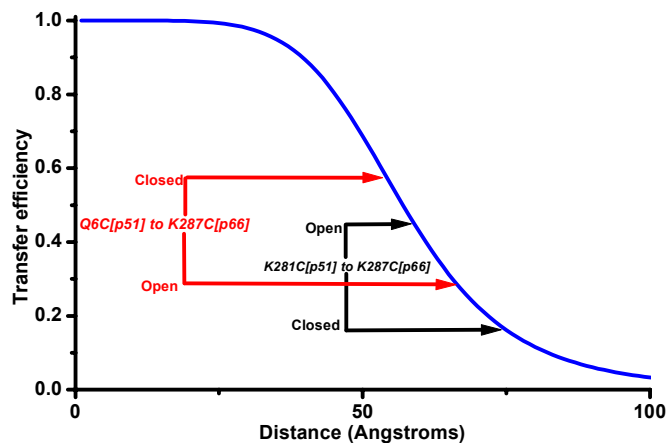


Figure 4.33 Transfer efficiency plot (blue) for the donor-acceptor combination A488-A633 (Förster radius = 57 Å). Shown in red and black are the distance changes expected on opening and closing of p66 for the systems K281C (p51)-K287C(p66) and Q6C(p51)-K87C(p66), respectively.

The heterodimer was reconstituted in the presence of MgCl_2 , with a 2.5 fold molar excess of $\text{p51}[6\text{H}]^{\text{Q6C}[\text{A633}]}$ over $\text{p66}[6\text{H-SII}]^{\text{K287C}[\text{A488}]}$. After addition of $\text{p51}[6\text{H}]^{\text{Q6C}[\text{A633}]}$ a blue precipitate formed. After an overnight dimerisation step this was removed by centrifugation and the protein purified using the streptactin macroprep column as described above. Since the amount of recovered protein was near the maximum that could be expected from the starting amount of p66, the precipitate formed during dimerisation consisted mostly of the labelled p51 subunit. The monomeric subunits and the heterodimer were analysed by SDS-PAGE (see Figure 4.35). The gels were first scanned for A488 and A633 (see section 3.2.1), followed by staining with Coomassie blue.

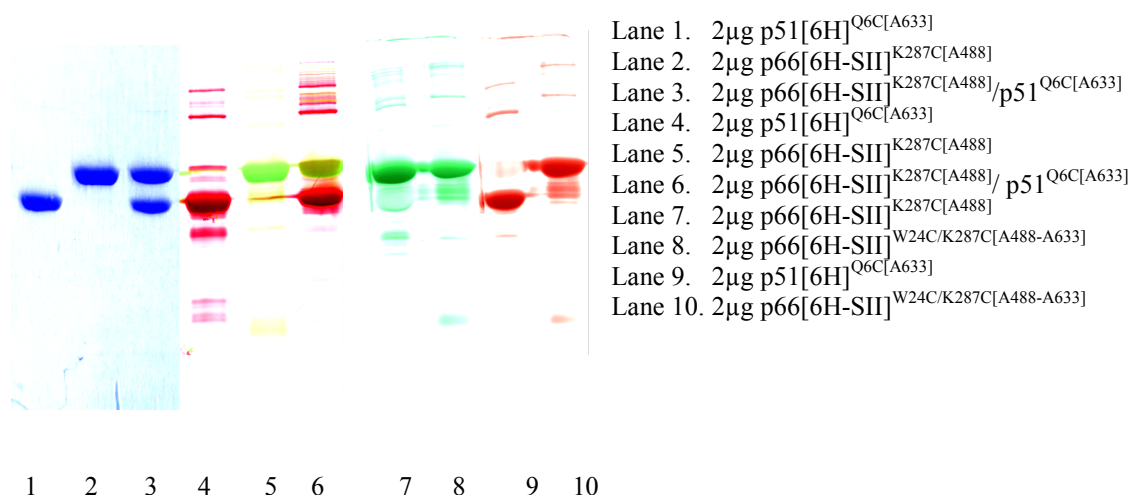


Figure 4.34. SDS-PAGE of double labelled proteins. Lanes 1-3 coomassie stain, and overlaid scans (4-6) of A488 (green) and A633 (red) of individual subunits and the heterodimeric enzyme. Lanes 7 and 9 show the non overlaid scans. Lanes 8 and 10 show the double labelled monomeric p66

The Coomassie stain of the individual subunits, and the heterodimer (lanes 1-3) shows that the subunits were not degraded during dimerisation. Lanes 4-6 show the overlaid scans for A488 and A633. The overlay of these two scans demonstrates that both fluorophores are present and coupled to the correct subunit. Although other proteins appear to be present in the case of the reconstituted heterodimer they constitute less than 2 % of the total fluorescence signal. Lanes 7 and 9 show the individual scans of the two subunits. The reconstituted double labelled protein p66[6H-SII]^{K287C[A488]}/p51[6H]^{Q6C[A633]} displayed a polymerase activity of 114 % when compared to wild type RT (see section 4.3.2), showing that formation of the heterodimer by this method resulted in fully active protein, and that the Strep-tag II had no adverse effects on activity.

It is presently not known which structures the subunits adopt when they are in the monomeric state. In order to gain structural information on the monomeric conformation of p66, the protein p66[6H-SII]^{W24C/K287C} was labelled randomly with the dyes A488 and A633 (see section 3.2.5). The two positions chosen for labelling are in the fingers (W24C) and thumb (K287C) subdomains of p66 (see Figure 4.34). Depending on whether p66 assumes the conformation observed for the heterodimer, or a structure closer to p51, large differences in energy transfer would be expected between the two labelled cysteine.

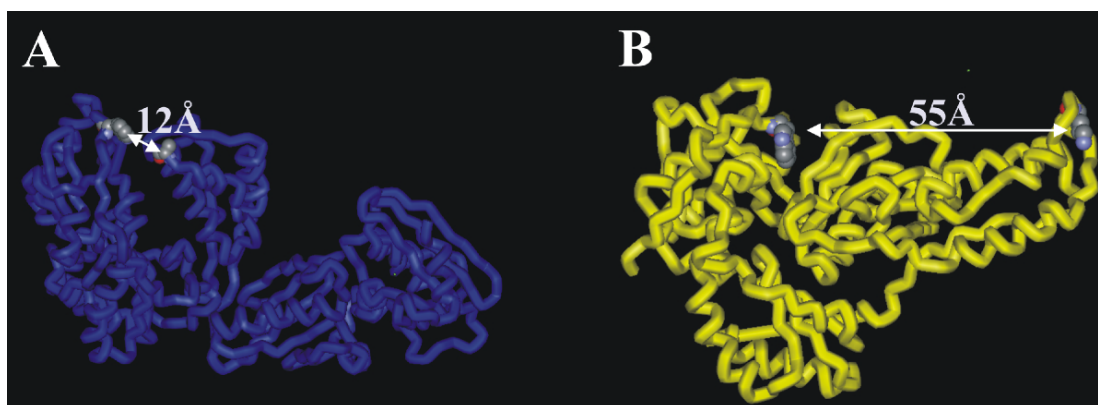


Figure 4.35. expected distances between W24C and K287C based on the structure of p66 (blue) and p51 (yellow) in the heterodimeric RT.

After labelling, the excess dye was removed by gel filtration followed by immobilisation and washing on the streptactin macroprep column. The labelling efficiency was 47% for A488 and 53% for A633. In Figure 4.35 lanes 8 and 10 show the A488 and A633 fluorescence respectively. There seems to be a slight degradation of the p66 subunit indicated by the presence of multiple bands below the p66 subunit. However, the fluorescence seen is less than 5% of the total signal.

<i>p66</i> [6H-SII]	<i>p51</i> [6H]
<i>p66</i> [6H-SII] ^{K287C[A488]}	<i>p51</i> [6H] ^{Q6C[A633]}
<i>p66</i> [6H-SII] ^{K287C[A594]}	<i>p51</i> [6H] ^{K281C[A488]}
<i>p66</i> [6H-SII] ^{K287C/W24C[A488-A633]}	

Table 4.6. Monomeric proteins labelled during this work.

4.2.3. Single molecule experiments

4.2.3.1. The structure of the heterodimeric enzyme

The protein *p66*^{K287C[A488]}/*p51*[6H]^{Q6C[A633]} was tested to determine if we could resolve the different structures seen in the apo-enzyme (Kensch *et al.* 2000) and measure the rate of interconversion between the open and closed state. In Figure 4.36 are the $\tau_{D(A)}$ vs R_{DA} and r_D graphs obtained using the double labelled protein. In the case of the free enzyme (Figure 4.36A), two species are present centred at distances of 82 Å and 116 Å. The ratio between the species is 57% and 43 % for the 82 Å and the 116 Å species. For the 82 Å species there appears to be significant quenching of the

donor or acceptor fluorophore. This is indicated by the position of the island, situated above and left of the Iso- τ line. For the donor fluorophore, the rotational mobility of the dye is between 2-7 ns, suggesting the dye has a mobility in an acceptable range, and therefore $2/3$ can be used for κ^2 . For the 116 Å centred species, the value obtained is above the MDR of 92 Å for p66^{K287C[A488]}/p51[6H] (see section 4.1.6.7.1). From the 82 Å species it can be seen that quenching of the acceptor fluorophore occurs. Comparing Q6 in the open and closed conformations of RT, there are no major changes in the position. Therefore if acceptor quenching exists for the 82 Å species it would also occur for the 116 Å species. Quenching of the acceptor fluorophore would increase the value of F_D/F_A (see section 3.5.1.5), and hence increase the R_{DA} . Although this leads to inaccuracies in the distances determined, it can be ruled out that the species is generated from a donor only labelled protein.

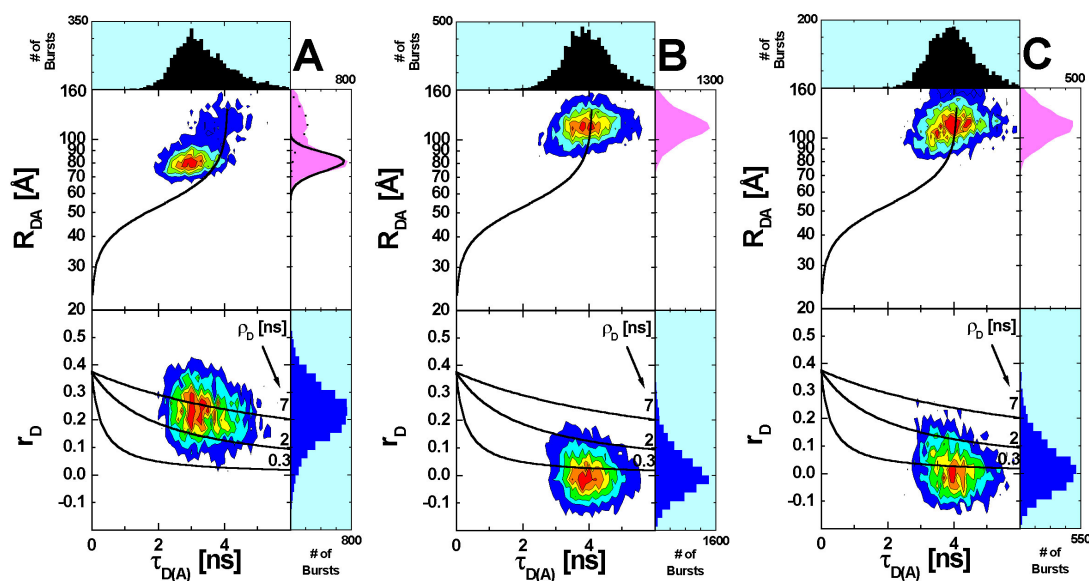


Figure 4.36. $\tau_{D(A)}$, plotted versus R_{DA} (upper panel) and against r_D (lower panel) for p66^{K287C[A488]}/p51[6H]^{Q6C[A633]}. Also plotted are the Iso- τ line (black line-upper panel) and rotational correlation times calculated from the Perrin equation. A) 200 pM p66^{K287C[A488]}/p51[6H]^{Q6C[A633]}, B) after addition of p/t(U) (final concentration 100 nM), and C) after addition of dNTPs to a final concentration of 200 μ M. Measurements were performed in SMS-protein buffer at 25°C.

In the presence of p/t(U) (Figure 4.36B) the species at 82 Å is lost and replaced by a single species at 116 Å. This suggests that the 116 Å species corresponds to the open conformation solved by crystallography (Huang *et al.* 1998; Jacobo-Molina *et al.* 1993) and that the 82 Å species corresponds to the closed conformation of RT (Hsiou *et al.* 1996; Rodgers *et al.* 1995). The distances expected

between the K287C (p66) and Q6C (p51) is 53 Å for the closed conformation and 66 Å for the open conformation. There is significant deviation between the expected values and the measured values. This is probably due to the quenching of the acceptor fluorophore.

On addition of nucleotides, no distance change is seen. However, this was not expected as both fluorophores are not sensitive to the structural changes in the fingers subdomain of p66 occurring upon nucleotide binding.

In the previous set of experiments, the rate of interconversion between the open and closed conformation was investigated. However, two species are seen which are stable over time. The average diffusion time for the free protein is 1.5ms, and changes in structure occurring over this time range would result in a single broadly distributed species, which is not seen. Due to the time information stored from MFD experiments, it was also possible to select bursts characterised by a longer diffusion time (over 10 ms). No difference in the pattern between the two species was observed. This suggests that the rate of interconversion between the open and closed conformations of RT is occurring in time ranges higher than 10 ms

4.2.3.2. The structure of monomeric p66

Due to the labelling efficiency of both dyes it would be expected that approximately 25% of the protein consists of donor only (D-D), 50% of donor acceptor (D-A, A-D), and 25% consist of acceptor only (A-A). By the use of single molecule spectroscopy it is possible to sort and disregard fluorescence events occurring due to donor only or acceptor only labelled proteins. A further advantage of single molecule spectroscopy is that due to the low concentrations used (pM range) it is unlikely that homodimeric RT (p66/p66) is formed. The dissociation constant for p66/p66 is in the nM range

In Figure 4.37 is plotted the $\tau_{D(A)}$ vs R_{DA} and r_D graph obtained using the double labelled RT monomer p66[6H-SII]^{W24C/K287C[A488-A633]}. This plot shows that at least two distributions are present. The first species characterised by an extremely high distance centred at 123 Å is probably due to donor labelled p66 (D-D). The second species centred at 77 Å is a low FRET-species, which tails at lower values of R_{DA} . This species is characterised by a broad distribution, broader than the fitted error distribution (eq. 3.19). This may be due to different properties of the donor or

acceptor dye depending on whether coupled to the protein at W24C or K287C, or *vice versa*. The r_D is in an acceptable range with a rotational mobility slightly above 2 ns.

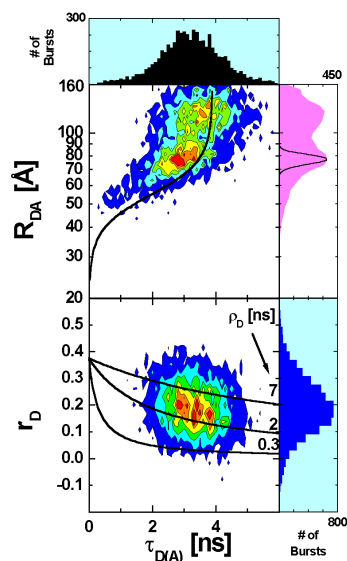


Figure 4.36. $\tau_{D(A)}$, plotted versus R_{DA} (upper panel) and against r_D , (lower panel) for the $p66[6H-SII]^{W24C/K287C}$ [A488-A633]. Also plotted is the Iso- τ line (black line-upper panel) and rotational correlation times calculated from the Perrin equation (black lines-lower panel). 200 pM p66 measured in SMS-protein buffer at 25 °C.

If the monomeric p66 adopts the same structure as for the heterodimer a distance of 12 Å would be expected between the two labelled positions. When p66 adopts a p51 like conformation a distance of 55 Å would be predicted. The data obtained suggest that the monomeric p66 adopts a p51 like conformation in solution. Experiments were also performed in which unlabelled p51[6H] was mixed with p66-W24C-K287C[A488-A633], in order to study the dimerisation process. Although the results showed a time dependent decrease in the species centred at 77 Å, it was not possible to analyse the increase in the high FRET (low distance) species formed.

It was previously suggested that based on energy minimisation considerations that p66 free in solution would adopt a “closed” p51 like structure, rather than the “open” p66 structure (Wang *et al.* 1994). It was also shown that the dimerisation of HIV-1 RT is a two step process (Restle *et al.* 1990; Restle *et al.* 1992). After the initial loose association of the two subunits a slow conformational isomerisation must occur before the enzyme displays catalytic activity.

Due to the structure of p66 observed in the experiments performed, it seems highly likely that the slow conformational change occurring is due to rearrangements

of the p66 subunit. The role of p51, which displays no catalytic activity in the heterodimer, would be to stabilise the energetically less favourable open p66 structure. The results also help to explain the observation that the monomeric subunits show no activity. If the monomeric p66 adopted similar structure to that seen in the heterodimer, the monomeric enzyme should be active. However, if it exists in the p51 like conformation, it would be inactive.

5. Summary

Using spFRET it was shown that interaction of RT with a p/t leads to the production of several different complexes. This is in contrast to what would be expected from crystallographic studies, but in accord with kinetic evidence on heterogeneity of RT:p/t complexes. The complexes were designated as productive (P), preproductive (PP), and dead end (DE). The p/t bound in the P manner corresponds to the binding mode characterised by crystallography. For the PP complex, the 3'-terminus of the primer is shifted such that it occupies the dNTP-binding pocket. Further evidence was provided by experiments using a pre-terminated p/t. In the presence of nucleotides, the PP complex is lost and replaced by the P complex. Addition of pyrophosphate led to the formation of the PP complex. In the case of the DE complex, the binding of p/t by RT occurs at a site not previously implicated in substrate recognition. The three complexes seen are in good agreement with the previous kinetic model proposed to explain p/t binding by RT (Wöhrl *et al.* 1999).

By studying complexes in which the template overhang was shortened, it was also possible to provide evidence for the role of the template overhang in nucleic acid positioning. With shorter templates, a greater distance separation is seen between the P and PP complexes, as well as a higher amount of the PP complex. It is possible that mutations implicated in NRTI resistance that disrupt p/t through altered interactions with the template may favour the PP complex, in which removal of a chain terminator should be possible via pyrophorolysis by pyrophosphate or nucleotide triphosphates.

Software was developed in order to map the binding mode in the P and DE complexes. The modelling of both binding modes was partially successful. For the P complex, the double stranded region of the p/t is in close agreement to the binding mode observed by crystallography and verifies the use of this technique for distance determination. For the template extension, it appears as if the dye sticks to the protein in the DNA-binding tract of p66. The structural data obtained for the DE complex, although of lower quality, indicate a binding site for the p/t far removed from the DNA-binding tract of p66.

During this work purification/labelling strategies have also been developed for monomeric p66 and p51. Formation of the heterodimer leads to a fully active enzyme containing two different fluorophores introduced in a site specific manner. Testing of double labelled enzyme confirms the previously obtained EPR results that the free

enzyme exists in two distinct conformations in solution. Although we could not determine the rate of interconversion between the “open” and “closed” forms of RT, the process must occur over time ranges greater than 10 ms under the conditions used.

Using p66 labelled with a donor and acceptor fluorophore it was also possible to obtain structural information on the monomeric form of p66. The results obtained suggest that in the monomeric form p66 does not exist in a structure similar to that seen in the active heterodimeric form. It appears likely that monomeric p66 exists in a p51-like conformation (i.e. like that seen for p51 in the heterodimer) and that a major structural rearrangement takes place after interaction with p51 to form the active heterodimer.

6. Future Studies

Structural data obtained on simplified RT:primer/template (p/t) complexes indicate heterogeneous binding modes, designated as productive (P), preproductive (PP) and dead-end complexes (DE). So far, the roles and the structures of two of these complexes (preproductive and dead-end complex) are unknown. Nevertheless, the structural data obtained so far on the two previously uncharacterised complexes leads to several interesting questions about their function.

Studies on the PP complex. From the data obtained it appears that in the PP complex, the 3'-terminus of the primer is shifted such that it occupies the dNTP-binding pocket. In this state, interaction with pyrophosphate could drive the polymerisation reaction backwards, allowing the removal of the last nucleotide in the growing primer, which could be a chain terminator derived from nucleoside prodrugs such as AZT and 3TC. It is believed that this process contributes to resistance in mutant viruses (Boyer *et al.* 2001). Using 3'-dideoxy-terminated p/ts it may be possible to investigate the role and mechanism of this process. An extension of this work would be to determine the effect of drug resistant RT mutants on the binding modes identified in the present work.

Studies on the DE complex. Based on the structural information gained, it should be possible to develop strategies to induce a single binding mode to favour the DE complex. Fabs have recently been described in the literature (Chiba *et al.* 1996; Chiba *et al.* 1997; Ohba *et al.* 2001) which bind to a discontinuous epitope in the thumb region of p66, which is not accessible in p51, and have been shown to block p/t binding. By the use of antibody fragments (Fabs) directed against the p66 domain of RT it may be possible to block formation of the P and PP complex. This may be useful in the crystallisation of the DE complex. It must be stressed that without the structural information on the DE complex obtained from the single-molecule measurements planning strategies of this kind would not be possible.

Approaching the function of the DE complex from the opposite direction, by favouring the productive complex, could also be useful. The binding site for the p/t to RT in the DE complex appears to be at a location close to the binding site of an Fab used to obtain first crystals of an RT:p/t complex (Jacob-Molina *et al.* 1993; Sarafianos *et al.* 2001). If the binding mode of this complex is disrupted in the

presence of the Fab, it should be possible to demonstrate this using the quenched flow method.

RT is only enzymatically active as a dimer (Restle *et al.* 1990; Restle *et al.* 1992). One striking feature of the dimeric enzyme is the highly asymmetric organisation of the subunits (Figure 3). Substances destabilising or blocking this protein/protein interaction inactivate the enzyme and can therefore block the retroviral life cycle (Morris *et al.* 1999). It has been demonstrated that formation of active RT is a two step mechanism, involving rapid association of the two subunits into an inactive dimer, followed by a slow conformational change yielding the fully active form of the enzyme (Divita *et al.* 1995a). The first interaction between the two subunits occurs via a Trp-rich hydrophobic cluster located in the connection subdomains. This is followed by a conformational change that stacks the thumb subdomain of p51 onto the RNaseH domain of p66 and arranges the fingers subdomain of p51 in the palm subdomain of p66 (Divita *et al.* 1995a).

So far there has not been a satisfactory method of monitoring the dimerisation process of RT (Divita *et al.* 1995a). Previous studies have been able to monitor the association of the p66 and the p51 subunit. However no direct method to monitor the isomerisation of the enzyme has been developed. The ability to label both subunits individually should allow the development of FRET based systems to monitor this step of heterodimer formation. These studies could provide the basis for assay to be used in screening substances as potential inhibitors for one or both steps involved in dimer formation

7. Reference List

- Arion, D., Kaushik, N., McCormick, S., Borkow, G., and Parniak, M. A. (1998). Phenotypic mechanism of HIV-1 resistance to 3'-azido-3'-deoxythymidine (AZT): increased polymerization processivity and enhanced sensitivity to pyrophosphate of the mutant viral reverse transcriptase. *Biochemistry* **37**, 15908-15917.
- Arthur, L. O., Bess Jr, J. W., Sowder II, R. C., Benveniste, R. E., Mann, D. L., Chermann, J., and Henderson, L. E (1992). Cellular Proteins Bound to Immunodeficiency Viruses: Implications for Pathogenesis and Vaccines. *Science* **258**, 1935-1938.
- Baltimore, D. (1970). RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* **226**, 1211-1213.
- Barre-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS). *Science* **220**, 868-871.
- Barshop, B. A., Adamson, D. T., Vellom, D. C., Rosen, F., Epstein, B. L., and Seegmiller, J. E. (1991). Luminescent immobilized enzyme test systems for inorganic pyrophosphate: assays using firefly luciferase and nicotinamide- mononucleotide adenylyl transferase or adenosine-5'-triphosphate sulfurylase. *Anal.Biochem.* **197**, 266-272.
- Berger, S. (2002). *Personal Communication*.
- Bittner.J.J. (1936). Some Possible Effects of Nursing on the Mammary Gland Tumor Incidence in Mice. *Science* **84**, 162.
- Boyer, P. L., Sarafianos, S. G., Arnold, E., and Hughes, S. H. (2001). Selective excision of AZTMP by drug-resistant human immunodeficiency virus reverse transcriptase. *J.Virol.* **75**, 4832-4842.

Boyer, P. L., Tantillo, C., Jacobo-Molina, A., Nanni, R. G., Ding, J., Arnold, E., and Hughes, S. H. (1994). Sensitivity of wild-type human immunodeficiency virus type 1 reverse transcriptase to dideoxynucleotides depends on template length; the sensitivity of drug-resistant mutants does not. *Proc.Natl.Acad.Sci.U.S.A* **91**, 4882-4886.

Brand, L. Zeitaufgelöster Nachweis einzelner Moleküle in Lösung. (1998). *Georg-August-Universität Göttingen*. (Thesis)

Brand, L., Eggeling, C., Zander, C., Drexhage, K. H., and Seidel, C. A. M (1997). Single-molecule identification of coumarin-120 by time-resolved fluorescence detection. *J.Physical Chemistry A* **101**, 4313-4321.

Certa, U., Bannwarth, W., Stuber, D., Gentz, R., Lanzer, M., Le Grice, S., Guillot, F., Wendler, I., Hunsmann, G., Bujard, H., and . (1986). Subregions of a conserved part of the HIV gp41 transmembrane protein are differentially recognized by antibodies of infected individuals. *EMBO J.* **5**, 3051-3056.

Chiba, J., Nakano, M., Suzuki, Y., Aoyama, K., Ohba, H., Kobayashi, T., Yasuda, A., Kojima, A., and Kurata, T. (1997). Generation of neutralizing antibody to the reverse transcriptase of human immunodeficiency virus type 1 by immunizing of mice with an infectious vaccinia virus recombinant. *J.Immunol.Methods* **207**, 53-60.

Chiba, J., Yamaguchi, A., Suzuki, Y., Nakano, M., Zhu, W., Ohba, H., Saito, A., Shinagawa, H., Yamakawa, Y., Kobayashi, T., and Kurata, T. (1996). A novel neutralization epitope on the 'thumb' subdomain of human immunodeficiency virus type 1 reverse transcriptase revealed by a monoclonal antibody. *J.Gen.Virol.* **77 (Pt 12)**, 2921-2929.

Clapham, P. R. and Weiss, R. A. (1997). Spoit for choice of co-receptors. *Nature* **388**, 230-231.

Clegg, R. M. (1992). Fluorescence resonance energy transfer and nucleic acids. *Methods Enzymol.* **211**, 353-388.

Clegg, R. M., Murchie, A. I., Zechel, A., and Lilley, D. M. (1993). Observing the helical geometry of double-stranded DNA in solution by fluorescence resonance energy transfer. *Proc.Natl.Acad.Sci.U.S.A* **90**, 2994-2998.

Darnell, J., Lodish, H., and Baltimore, D. (1990). Cancer. In 'Molecular Cell Biology'. Scientific American Books

Das, K., Ding, J., Hsiou, Y., Clark, A. D., Jr., Moereels, H., Koymans, L., Andries, K., Pauwels, R., Janssen, P. A., Boyer, P. L., Clark, P., Smith, R. H., Jr., Kroeger Smith, M. B., Michejda, C. J., Hughes, S. H., and Arnold, E. (1996). Crystal structures of 8-Cl and 9-Cl TIBO complexed with wild-type HIV-1 RT and 8-Cl TIBO complexed with the Tyr181Cys HIV-1 RT drug-resistant mutant. *J.Mol.Biol.* **264**, 1085-1100.

Deniz, A. A., Dahan, M., Grunwell, J. R., Ha, T., Faulhaber, A. E., Chemla, D. S., Weiss, S., and Schultz, P. G. (1999). Single-pair fluorescence resonance energy transfer on freely diffusing molecules: observation of Forster distance dependence and subpopulations. *Proc.Natl.Acad.Sci.U.S.A* **96**, 3670-3675.

Deniz, A. A., Laurence, T. A., Beligere, G. S., Dahan, M., Martin, A. B., Chemla, D. S., Dawson, P. E., Schultz, P. G., and Weiss, S. (2000). Single-molecule protein folding: diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proc.Natl.Acad.Sci.U.S.A* **97**, 5179-5184.

Divita, G., Muller, B., Immendorfer, U., Gautel, M., Rittinger, K., Restle, T., and Goody, R. S. (1993). Kinetics of interaction of HIV reverse transcriptase with primer/template. *Biochemistry* **32**, 7966-7971.

Divita, G., Rittinger, K., Geourjon, C., Deleage, G., and Goody, R. S. (1995a). Dimerization kinetics of HIV-1 and HIV-2 reverse transcriptase: a two step process. *J.Mol.Biol.* **245**, 508-521.

Divita, G., Rittinger, K., Restle, T., Immendorfer, U., and Goody, R. S. (1995b). Conformational stability of dimeric HIV-1 and HIV-2 reverse transcriptases. *Biochemistry* **34**, 16337-16346.

Duckett, D. R., Murchie, A. I., Bhattacharyya, A., Clegg, R. M., Diekmann, S., von Kitzing, E., and Lilley, D. M. (1993). The structure of DNA junctions and their interaction with enzymes. *Eur.J.Biochem.* **211**, 285-295.

- Eggeling, C. (2000) Analyse von photochemischer Kinetik und Moleküldynamik durch mehrdimensionale Einzelmolekül-Fluoreszenzspektroskopie.. *Georg-August-Universität Göttingen*. (Thesis)
- Eggeling, C., Berger, S., Brand, L., Fries, J. R., Schaffer, J., Volkmer, A., and Seidel, C. A. (2001). Data registration and selective single-molecule analysis using multi-parameter fluorescence detection. *J.Biotechnol.* **86**, 163-180.
- Eggeling, C., Fries, J. R., Brand, L., Gunther, R., and Seidel, C. A. (1998). Monitoring conformational dynamics of a single molecule by selective fluorescence spectroscopy. *Proc.Natl.Acad.Sci.U.S.A* **95**, 1556-1561.
- Enderlein, J., Robbins, D. L., Ambrose, W. P., Goodwin, P. M., and Keller, R. A (1997). The statistics of single molecule detection: an overview. *Bioimaging* **5**, 88-98.
- Esnouf, R., Ren, J., Ross, C., Jones, Y., Stammers, D., and Stuart, D. (1995). Mechanism of inhibition of HIV-1 reverse transcriptase by non- nucleoside inhibitors. *Nature Structural Biology* **2**, 303-308.
- Fries, J. R., Brand, L., Eggeling, C., Köllner, M., and Seidel, C. A. M (1998). Quantitative identification of different single-molecules by selective time-resolved confocal fluorescence spectroscopy. *J.Physical Chemistry A* **102**, 6601-6613.
- Furey, W. S., Joyce, C. M., Osborne, M. A., Klenerman, D., Peliska, J. A., and Balasubramanian, S. (1998). Use of fluorescence resonance energy transfer to investigate the conformation of DNA substrates bound to the Klenow fragment. *Biochemistry* **37**, 2979-2990.
- Gallo, R. C., Mann, D., Broder, S., Ruscetti, F. W., Maeda, M., Kalyanaraman, V. S., Robert-Guroff, M., and Reitz, M. S. (1982). Human T-Cell Leukemia-Lymphoma Virus (HTLV) is in T but not B Lymphocytes from a Patient with Cutaneous T-Cell Lymphoma. *Proc.Natl.Acad.Sci.U.S.A* **79**, 5680-5683.
- Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., Safai, B., White, G., Foster, P., and Markham, P. D. (1984). Frequent Detection and Isolation of Cytopathic Retroviruses (HTLV-III) from Patients with AIDS and at Risk for AIDS. *Science* **224**, 500-503.

-
- Goody, R. S. (1995). Rational drug design and HIV: hopes and limitations. *Nat.Med.* **1**, 519-520.
- Goody, R. S., Muller, B., and Restle, T. (1991). Factors contributing to the inhibition of HIV reverse transcriptase by chain-terminating nucleotides in vitro and in vivo. *FEBS Lett.* **291**, 1-5.
- Gotte, M., Li, X., and Wainberg, M. A. (1999). HIV-1 reverse transcription: a brief overview focused on structure- function relationships among molecules involved in initiation of the reaction. *Arch.Biochem.Biophys.* **365**, 199-210.
- Greene, W. C. (1993). AIDS and the immune system. *Sci.Am.* **269**, 98-105.
- Ha, T. (2001). Single-molecule fluorescence resonance energy transfer. *Methods* **25**, 78-86.
- Ha, T., Ting, A. Y., Liang, J., Caldwell, W. B., Deniz, A. A., Chemla, D. S., Schultz, P. G., and Weiss, S. (1999). Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. *Proc.Natl.Acad.Sci.U.S.A* **96**, 893-898.
- Herlitze, S. and Koenen, M. (1990). A general and rapid mutagenesis method using polymerase chain reaction. *Gene* **91**, 143-147.
- Herrman, C. H. and Rice, A. P. (1995). Lentivirus Tat proteins specifically associate with a cellular protein kinase, TAK, that hyperphosphorylates the carboxyl-terminal domain of the large subunit of RNA polymerase II: candidate for a Tat co.-factor. *J.Virol.* **69**, 1612-1620.
- Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., and Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**, 51-59.
- Hsiou, Y., Ding, J., Das, K., Clark, A. D., Jr., Hughes, S. H., and Arnold, E. (1996). Structure of unliganded HIV-1 reverse transcriptase at 2.7 Å resolution: implications of conformational changes for polymerization and inhibition mechanisms. *Structure.* **4**, 853-860.

-
- Huang, H., Chopra, R., Verdine, G. L., and Harrison, S. C. (1998). Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. *Science* **282**, 1669-1675.
- Isel, C., Lanchy, J. M., Le Grice, S. F., Ehresmann, C., Ehresmann, B., and Marquet, R. (1996). Specific initiation and switch to elongation of human immunodeficiency virus type 1 reverse transcription require the post-transcriptional modifications of primer tRNA³Lys. *EMBO J.* **15**, 917-924.
- Jacobo-Molina, A., Ding, J., Nanni, R. G., Clark, A. D., Jr., Lu, X., Tantillo, C., Williams, R. L., Kamer, G., Ferris, A. L., Clark, P., and . (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. *Proc.Natl.Acad.Sci.U.S.A* **90**, 6320-6324.
- Jaeger, J., Restle, T., and Steitz, T. A. (1998). The structure of HIV-1 reverse transcriptase complexed with an RNA pseudoknot inhibitor. *EMBO J.* **17**, 4535-4542.
- Keller, R. A., Ambrose, W. P., Goodwin, P. M., Jett, J. H., Martin, J. C., and Wu, M. (1996). Single-molecule fluorescence analysis in solution. *Applied Spectroscopy* **50**, 12A-32A.
- Kensch, O. (2000) Untersuchungen zur Konformation und Dynamik der Reversen Transkriptase von HIV-1 durch ESR- und Einzelmolekülfluoreszenzspektroskopie. *Universität Dortmund* (Thesis)
- Kensch, O., Restle, T., Wöhrle, B. M., Goody, R. S., and Steinhoff, H. J. (2000). Temperature-dependent equilibrium between the open and closed conformation of the p66 subunit of HIV-1 reverse transcriptase revealed by site-directed spin labelling. *J.Mol.Biol.* **301**, 1029-1039.
- Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., and Steitz, T. A. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science* **256**, 1783-1790.

- Koshioka, M., Sasaki, K., and Masuhara, H. (1995). Time-dependent fluorescence depolarization analysis in three dimensional spectroscopy. *Applied Spectroscopy* **49**, 224-228.
- Krebs, R., Immendorfer, U., Thrall, S. H., Wohrl, B. M., and Goody, R. S. (1997). Single-step kinetics of HIV-1 reverse transcriptase mutants responsible for virus resistance to nucleoside inhibitors zidovudine and 3-TC. *Biochemistry* **36**, 10292-10300.
- Kullback, S. (1959). information, theory and statistics. New York.
- Levy, J. A., Hoffman, A. D., Kramer, S. M., Landis, J. A., Shimabukuro, J. M., and Oshiro, L. S. (1984). Isolation of Lymphocytopathic Retroviruses from San Francisco Patients with AIDS. *Science* **225**, 840-842.
- Lu, H. P., Xun, L., and Xie, X. S. (1998). Single-molecule enzymatic dynamics. *Science* **282**, 1877-1882.
- Maier, T., Drapal, N., Thanbichler, M., and Bock, A. (1998). Strep-tag II affinity purification: an approach to study intermediates of metalloenzyme biosynthesis. *Anal.Biochem.* **259**, 68-73.
- Majumdar, C., Abbotts, J., Broder, S., and Wilson, S. H. (1988). Studies on the mechanism of human immunodeficiency virus reverse transcriptase. Steady-state kinetics, processivity, and polynucleotide inhibition. *J.Biol.Chem.* **263**, 15657-15665.
- Maus, M., Cotlet, M., Hofkens, J., Gensch, T., and De Schryver, F. G. (2001). An experimental comparison of the maximum likelihood estimation and nonlinear least-squares fluorescence lifetime analysis of single molecules. *Anal.chem.* **73**, 2078-2086.
- Meyer, P. R., Matsuura, S. E., So, A. G., and Scott, W. A. (1998). Unblocking of chain-terminated primer by HIV-1 reverse transcriptase through a nucleotide-dependent mechanism. *Proc.Natl.Acad.Sci.U.S.A* **95**, 13471-13476.
- Mizutani, S., Boettiger, D., and Temin, H. M. (1970). A DNA-depenent DNA polymerase and a DNA endonuclease in virions of Rous sarcoma virus. *Nature* **226**, 1211-1213.

- Moore, J. P. (2002). Coreceptors: Implications for HIV Pathogenesis and therapy. *Science* **276**, 51-52.
- Morris, M. C., Berducou, C., Mery, J., Heitz, F., and Divita, G. (1999). The thumb domain of the P51-subunit is essential for activation of HIV reverse transcriptase. *Biochemistry* **38**, 15097-15103.
- Noji, H., Yasuda, R., Yoshida, M., and Kinoshita, K. (1997). Direct observation of the rotation of F1-ATPase. *Nature* **386**, 299-302.
- Ohba, H., Soga, T., Tomozawa, T., Nishikawa, Y., Yasuda, A., Kojima, A., Kurata, T., and Chiba, J. (2001). An immunodominant neutralization epitope on the 'thumb' subdomain of human immunodeficiency virus type 1 reverse transcriptase revealed by phage display antibodies. *J.Gen.Virol.* **82**, 813-820.
- Poiesz, B. J., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna, J. D., and Gallo, R. C. (1980). Detection and Isolation of Type C Retrovirus Particles from Fresh and Cultured Lymphocytes of a Patient with Cutaneous T-Cell Lymphoma. *Proc.Natl.Acad.Sci.U.S.A* **77**, 7415-7419.
- Popovic, M., Sarngadharan, M. G., Read, E., and Gallo, R. C. (1984). Detection, Isolation, and Continuous Production of Cytopathic Retroviruses (HTLV-III) from Patients with AIDS and Pre-AIDS. *Science* **224**, 497-500.
- Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., and . (1985). Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* **313**, 277-284.
- Reardon, J. E. (1993). Human immunodeficiency virus reverse transcriptase. A kinetic analysis of RNA-dependent and DNA-dependent DNA polymerization. *J.Biol.Chem.* **268**, 8743-8751.
- Ren, J., Esnouf, R., Garman, E., Somers, D., Ross, C., Kirby, I., Keeling, J., Darby, G., Jones, Y., Stuart, D., and . (1995). High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nature Structural Biology* **2**, 293-302.
- Restle, T. 2002. (Personal Communication)

- Restle, T., Muller, B., and Goody, R. S. (1990). Dimerization of human immunodeficiency virus type 1 reverse transcriptase. A target for chemotherapeutic intervention. *J.Biol.Chem.* **265**, 8986-8988.
- Restle, T., Muller, B., and Goody, R. S. (1992). RNase H activity of HIV reverse transcriptases is confined exclusively to the dimeric forms. *FEBS Lett.* **300**, 97-100.
- Robert-Guroff, M., Nakao, Y., Notake, K., Ito, Y., Sliski, A., and Gallo, R. C. (1982). Natural Antibodies to Human Retrovirus HTLV in a Cluster of Japanese Patients with Adult T Cell Leukemia. *Science* **215**, 975-978.
- Rodgers, D. W., Gamblin, S. J., Harris, B. A., Ray, S., Culp, J. S., Hellmig, B., Woolf, D. J., Debouck, C., and Harrison, S. C. (1995). The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. *Proc.Natl.Acad.Sci.U.S.A* **92**, 1222-1226.
- Roitt, I. M., Brostoff, J., and Male, D. K. (1989). In 'Immunology'. (Gower Medical, London:
- Rous, P (1911). Transmission of a malignant new growth by means of a cell-free filtrate. *J.Exp.Med.* **13** , 397.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc.Natl.Acad.Sci.U.S.A* **74**, 5463-5467.
- Sarafianos, S. G., Das, K., Tantillo, C., Clark, A. D., Jr., Ding, J., Whitcomb, J. M., Boyer, P. L., Hughes, S. H., and Arnold, E. (2001). Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. *EMBO J.* **20**, 1449-1461.
- Sarnadharan, M. G., Popovic, M., Bruch, L., Schupbach, J., and Gallo, R. C. (1984). Antibodies Reactive with Human T-Lymphotropic Retroviruses (HTLV-III) in the Serum of Patients with AIDS. *Science* **224**, 506-508.
- Sauer, M., Drexhage, K. H., Lieberwirth, U., Müller, R., Nord, S., and Zander, C. (1998). Dynamics of the electron transfer reaction between an oxazine dye and DNA oligonucleotides monitored on the single-molecule level. *chemical physics letters* **284**, 153-163.

- Schaffer, J., Volkmer, A., Eggeling, C., Subramaniam, V., Striker, G., and und Seidel, C. A. M (1999). Identification of single molecules in aqueous solution by time-resolved fluorescence anisotropy. *J.Physical Chemistry A* **103**, 331-336.
- Selvin, P. R. (1995). Fluorescence resonance energy transfer. *Methods Enzymol.* **246**, 300-334.
- Sluis-Cremer, N., Arion, D., and Parniak, M. A. (2000). Molecular mechanisms of HIV-1 resistance to nucleoside reverse transcriptase inhibitors (NRTIs). *Cell Mol.Life Sci.* **57**, 1408-1422.
- Smerdon, S. J., Jager, J., Wang, J., Kohlstaedt, L. A., Chirino, A. J., Friedman, J. M., Rice, P. A., and Steitz, T. A. (1994). Structure of the binding site for nonnucleoside inhibitors of the reverse transcriptase of human immunodeficiency virus type 1. *Proc.Natl.Acad.Sci.U.S.A* **91**, 3911-3915.
- Thrall, S. H., Krebs, R., Wohrl, B. M., Cellai, L., Goody, R. S., and Restle, T. (1998). Pre-steady-state kinetic characterization of RNA-primed initiation of transcription by HIV-1 reverse transcriptase and analysis of the transition to a processive DNA-primed polymerization mode. *Biochemistry* **37**, 13349-13358.
- Turner, B. G. and Summers, M. F. (1999). Structural biology of HIV. *J.Mol.Biol.* **285**, 1-32.
- Wang, J., Smerdon, S. J., Jager, J., Kohlstaedt, L. A., Rice, P. A., Friedman, J. M., and Steitz, T. A. (1994). Structural basis of asymmetry in the human immunodeficiency virus type 1 reverse transcriptase heterodimer. *Proc.Natl.Acad.Sci.U.S.A* **91**, 7242-7246.
- Wang, M. D., Schnitzer, M. J., Yin, H., Landick, R., Gelles, J., and Block, S. M. (1998). Force and velocity measured for single molecules of RNA polymerase. *Science* **282**, 902-907.
- Weiss, S. Klonierung, Expression und biochemische Charakterisierung der HIV-1 Reversen Transkriptase. 1988 *Ruprecht-Karls-Universität Heidelberg*. (Thesis)

Weiss, S. (1999). Fluorescence spectroscopy of single biomolecules. *Science* **283**, 1676-1683.

Weiss, S. (2000). Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nature Structural Biology* **7**, 724-729.

Wöhrl, B. M., Krebs, R., Goody, R. S., and Restle, T. (1999). Refined model for primer/template binding by HIV-1 reverse transcriptase: pre-steady-state kinetic analyses of primer/template binding and nucleotide incorporation events distinguish between different binding modes depending on the nature of the nucleic acid substrate. *J.Mol.Biol.* **292**, 333-344.

Zander, C., Sauer, M., Ko, D. S., Schulz, A., Wolfrum, J., Brand, L., Eggeling, C., and Seidel, C. A. (1996). Detection and characterization of single molecules in aqueous solution. *Applied Physics.B* **63**, 517-523.

Zhuang, X., Bartley, L. E., Babcock, H. P., Russell, R., Ha, T., Herschlag, D., and Chu, S. (2000). A single-molecule study of RNA catalysis and folding. *Science* **288**, 2048-2051.

Zwicker, N., Adelhelm, K., Thiericke, R., Grabley, S., and Hanel, F. (1999). Strep-tag II for one-step affinity purification of active bHLHzip domain of human c-Myc. *Biotechniques* **27**, 368-375.

8. Appendices

Appendix 1. Nucleic acids

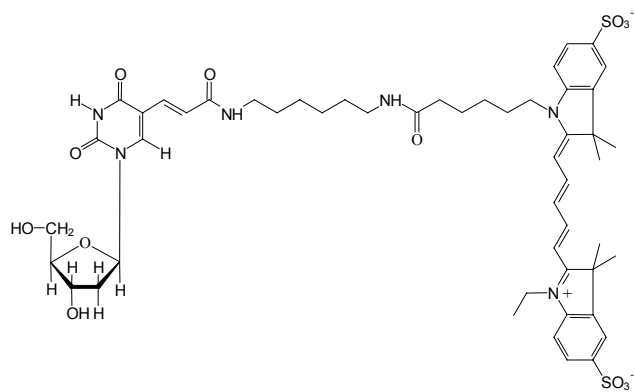
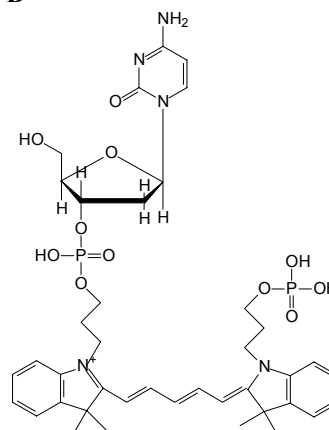
8.1.1 Oligonucleotides used for MFD experiments (5'-3')

P	TTGTCCCTGTTCTGGGCGCC
T	TGGTTAATCTCTGCATGGCGCCCGAACAGGGACAA
P_5_Cy5	TTGTCCCTGTTCTGGGCGCC
P_10bp_Cy5	TTGTCCCTGTTCTGGGCGCC
P_3'_Cy5	TTGTCCCTGTTCTGGGCGCC
T_7o_Cy5	GGGTTAATCTCTGCATGGCGCCCGAACAGGGACAA
T_5'_Cy5	TGGTTAATCTCTGCATGGCGCCCGAACAGGGACAA
T ⁺³	CATGGCGCCCGAACAGGGACAA
T ⁺⁶	CTGCATGGCGCCCGAACAGGGACAA
T ⁺⁹	TCTCTGCATGGCGCCCGAACAGGGACAA

8.1.2 Mutagenesis Oligonucleotides (5'-3')

EcoRV-sense	GGA TTA GAT ATC AGT ACA ATG TGC TTC C
S68C-sense	TTG CCA TAA AGA AAA AAG ACT GTA CTA AAT GGA GA
S68C-anti	TCT CCA TTT AGT ACA GTC TTT TTT CTT TAT GGC AA
Qx66-anti	ATC AGG CTG AAA ATC TTC TCT CAT CCG C
C280S-sense	TAA AGT AAG GCA ATT ATC TAA ACT CCT TAG AGG AA
C280S-anti	TTC CTC TAA GGA GTT TAG ATA ATT GCC TTA CTT TA
C38S-sense	AGC ATT AGT AGA AAT TTC TAC AGA AAT GGA AAA GG
C38S-anti	CCT TTT CCA TTT CTG TAG AAA TTT CTA CTA ATG CT
Fx51-sense	GCT TTG TGA GCG GAT AAC AAT TAT AAT AGA
K287C-sense	AAT TCC TTA GAG GAA CCT GTG CAC TAA CAG AAG TAA T
K287C-anti	ATT ACT TCT GTT AGT GCA CAG GTT CCT CTA AGG AGT T
T27C-sense	TAA ACA ATG GCC ATT GTG TGA AGA AAA AAT AAA AGC
T27C-anti	5'GCT TTT ATT TTT TCT TCA CAC AAT GGC CAT TGT TTA3'
W24C-sense	CCA AAA GTT AAA CAA TGT CCA TTG ACA GAA GAA AA
W24C-anti	TTT TCT TCT GTC AAT GGA CAT TGT TTA ACT TTT GG
Qx51-anti	GGA TCT ATC AAC AGG AGT CCA AGC TC
K281C-sense	AGT AAG GCA ATT ATC TTG TCT CCT TAG AGG AAC CA
K281C-anti	TGG TTC CTC TAA GGA GAC AAG ATA ATT GCC TTA CT
T470C-sense	AAA GGT TGT CCC CCT ATG TAA CAC AAC AAA TCA G
T470C-anti	CTG ATT TGT TGT GTT ACA TAG GGG GAC AAC CTT T
S68C-sense	TTG CCA TAA AGA AAA AAG ACT GTA CTA AAT GGA GA
S68C-anti	TCT CCA TTT AGT ACA GTC TTT TTT CTT TAT GGC AA
K173C-sense	AGA GCC TTT TAA ATG TCA AAA TXX AGA CA
K173C-anti	TGT CTG GAT TTT GAC ATT TAA AAG GCT CT
Q6C-sense	CTT CCC ATT AGC CCT ATT TGT ACT GT
Q6C-anti	TTA CTG GTA CAG TAC AAA TAG GGC TA
C-Strep-tag II	AGG CTA AGC TTA TTT TTC GAA TTG AGG ATG TGA CCA TAG TAT TTT CCT GAT TCC AGC ACT GAC
E194C-sense	T GTA GGA TCT GAC TTA TGT ATA GGG CAG CAT AGA A
E194C-anti	T TCT ATG CTG CCC TAT ACA TAA GTC AGA TCC TAC A

8.1.3. Fluorescent modified bases used

A**B**

(A) The structure of AminoC6_Cy5 dThymine and, (B) The structure of 3'_Cy5 labelled dCytosine

Appendix 2. programs used during RT:p/t interaction modelling

8.2.1 Display_sphere

```

/*
Display_sphere.c

Input card
-10.0 15.0 -12.0 23.0 -10.0 30.0 => Limits for grid points
 1.0  1.0  1.0      => Interval for grid points
 1.0  2.0  3.0  5.0  1.0  => xyz coordinates of donor A, lower and upper limit
 O      => atom type for output pdb file
 WAT    => residue type for output pdb file
 X      => chain ID for output pdb file
 END
*/

#include<stdio.h>
#include<ctype.h>
#include<string.h>
#include<math.h>
#include<stdlib.h>

FILE *input_data_file,*output_log_file,*output_pdb_file;

main (int argc , char *argv [])
{
  char  xname[80],wname[80],pname[80];
  char  type[10];
  char  atom_type[3],resid_type[4],chain_id[2];
  double      x,y,z;
  int  atom_num=0,resid_num=0;
  double distance_a ;
  double x_a , y_a , z_a , aver_dist_a , error_dist_a ;
  double x_start , x_stop , delta_x ;
  double y_start , y_stop , delta_y ;
  double z_start , z_stop , delta_z ;

  switch (argc) {
    case 3 : strcpy(xname,argv[1]);
             strcpy(wname,argv[2]);
             strcpy(pname,argv[3]);
             break;
    case 2 : strcpy(xname,argv[1]);
             strcpy(wname,argv[1]);
             strcpy(pname,argv[1]);
             strcat(xname,".dat");
             strcat(wname,".log");
             strcat(pname,".pdb");
             break;
    default :
             printf("FRET data file: ");
             scanf("%s",xname);

```

```

        printf("Log file name: ");
        scanf("%s",wname);
        printf("PDB file name: ");
        scanf("%s",pname);
        break;
    }
    if(!(input_data_file = fopen(xname,"r"))) {
        printf("no such file %s\n",xname);
        exit(0);
    }
    output_log_file = fopen(wname,"w");
    output_pdb_file = fopen(pname,"w");

    do {
        if (fscanf (input_data_file, "%lf%lf%lf%lf%lf%lf",
            &x_start,&x_stop,&y_start,&y_stop,&z_start,&z_stop) != 6) goto Fehler1 ;
        printf("Record1: %lf%lf%lf%lf%lf%lf\n",
            x_start,x_stop,y_start,y_stop,z_start,z_stop);

        if (fscanf (input_data_file, "%lf%lf%lf",
            &delta_x,&delta_y,&delta_z) != 3) goto Fehler2 ;
        printf("Record2: %lf%lf%lf\n",
            delta_x,delta_y,delta_z);

        if (fscanf (input_data_file, "%lf%lf%lf%lf%lf",
            &x_a,&y_a,&z_a,&aver_dist_a,&error_dist_a) != 5) goto Fehler3 ;
        printf("Record3: %lf%lf%lf%lf%lf\n",
            x_a,y_a,z_a,aver_dist_a,error_dist_a);

        if (fscanf (input_data_file, "%s ",atom_type) != 1) goto Fehler9 ;
        printf("Record9: %s \n",atom_type);
        if (fscanf (input_data_file, "%s ",resid_type) != 1) goto Fehler10 ;
        printf("Record10: %s \n",resid_type);
        if (fscanf (input_data_file, "%s ",chain_id) != 1) goto Fehler11 ;
        printf("Record11: %s \n",chain_id);

        printf("Start of calculation \n");

        x = x_start ;
        y = y_start ;
        z = z_start ;

        for (x = x_start ; x <= x_stop ; x = x + delta_x)
            { for (y = y_start ; y <= y_stop ; y = y + delta_y)
                { for (z = z_start ; z <= z_stop ; z = z + delta_z)
                    {
                        distance_a = sqrt((x-x_a)*(x-x_a) + (y-y_a)*(y-y_a) + (z-z_a)*(z-z_a)) ;
                        if ( distance_a >= (aver_dist_a - (error_dist_a)/(2)))
                            { if ( distance_a <= (aver_dist_a + (error_dist_a)/(2)))
                                {
                                    atom_num++;
                                    resid_num++;
                                }
                            }
                    }
                }
            }
        fprintf (output_pdb_file,"ATOM %4d%3s %4s %1s%4d %8.3f%8.3f%8.3f 1.00 20.00\n",
            atom_num,atom_type,resid_type,chain_id,resid_num,x,y,z) ;
    }
}

```

```
fprintf(output_log_file,"Protein XYZ Min-Max-Bereich...: %8.3f %8.3f %8.3f %8.3f %8.3f %8.3f
\n",
    x_start,x_stop,y_start,y_stop,z_start,z_stop);
fprintf(output_log_file,"Gridstufung für Berechnung...: %8.3f %8.3f %8.3f\n",
    delta_x,delta_y,delta_z);
fprintf(output_log_file,"XYZ of donor A and lower and upper limit: %8.3f %8.3f %8.3f %8.3f
%8.3f\n",
    x_a,y_a,z_a,aver_dist_a,error_dist_a);

    if(!strcmp(type,"END ")) break;
} while (0);

fprintf(output_pdb_file,"END");
printf("End of calculation \n");

fclose(output_log_file);
fclose(output_pdb_file);
fclose(input_data_file);

return 0 ;

exit(0);

Fehler1:
printf ("Error reading min_max range of protein \n") ;
fclose (input_data_file) ;
return 1 ;

Fehler2:
printf ("Error reading delta_xyz line \n") ;
fclose (input_data_file) ;
return 1 ;

Fehler3:
printf ("Error reading donor A line \n") ;
fclose (input_data_file) ;
return 1 ;

Fehler9:
printf ("Error reading atom name \n") ;
fclose (input_data_file) ;
return 1 ;

Fehler10:
printf ("Error reading residue name \n") ;
fclose (input_data_file) ;
return 1 ;

Fehler11:
printf ("Error reading chain id \n") ;
fclose (input_data_file) ;
return 1 ;
}

/* Der "File-Pointer" darf vom Programmierer als "Pointer auf ein 'Objekt
vom Typ FILE" angesehen werden, der von fopen geliefert wird und fuer
```

alle nachfolgenden File-Operation (Lesen, Schreiben, Schliessen, ...) als Identifier verwendet wird (dass sich hinter dem Type FILE eine in stdio.h beschriebene Struktur verbirgt, ist fuer den Programmierer von geringem Interesse). Der File-Pointer muss in der Form

```
FILE *infile_p
```

vereinbart werden, wobei fuer den hier gewaehlten Namen infile_p ein beliebiger Name stehen darf. */

/* Ein File wird z. B. mit

```
infile_p = fopen (argv [1] , "r" );
```

geoeffnet, wobei zwei Argumente angegeben werden muessen:

* Das erste Argument ist der Filename, der als relativer oder absoluter Filename angegeben werden darf, also gegebenenfalls auch den gesamten Pfad (und unter DOS eine Laufwerksbezeichnung) enthalten kann.

* Das zweite Argument kennzeichnet die beabsichtigte Verwendung des geoeffneten Files, besonders wichtig sind folgende Moeglichkeiten:

"r" oeffnet zum Lesen,

"w" oeffnet zum Schreiben, der Inhalt eines mit dem angegebenen Namen bereits existierenden Files geht dabei verloren,

"a" ('append') oeffnet zum Schreiben ab File-Ende eines bereits existierenden Files.

Als Return-Wert liefert fopen den fuer nachfolgende Aktionen benoetigten File-Pointer oder NULL (bei Misserfolg, passiert z. B. beim Oeffnen eines nicht existierenden Files mit "r" oder eines existierenden schreibgeschuetzten Files mit "w"), sollte beim Oeffnen unbedingt abgefragt werden. Dies kann (kuerzer als oben programmiert) z. B. mit

```
if ((infile_p = fopen (argv [1] , "r")) != NULL)
{
  ** File erfolgreich geoeffnet **
}
```

erfolgen, weil in C eine Wertzuweisung (infile_p = ...) eingeklammert werden darf und die Klammer noch einmal den zugewiesenen Wert repraesentiert, der dann fuer eine Vergleichsoperation verwendet werden kann. */

/* Mit der 'stdio'-Funktion

```
i = fgetc (infile_p);
```

wird genau ein Zeichen gelesen und als 'int'-Wert abgeliefert (in file1.c wird dieser Return-Wert ignoriert, weil nur die Anzahl der zu lesenden Zeichen ermittelt werden soll).

Die Frage, warum eine Funktion, die ein Zeichen liest, nicht einen Return-Wert vom Typ 'char' abliefern, wird durch den Sonderfall beantwortet: Wenn das File-Ende erreicht ist, liefert 'fgetc' das "End-of-File"-Zeichen EOF, fuer das (zumindest in einem Zeichensatz mit 256 Zeichen) kein Platz waere, weil mit einem Byte (Typ 'char') nur genau 256 Zeichen darstellbar sind. EOF ist in stdio.h definiert, ueblicherweise als -1, was allerdings den Programmierer nicht

```

interessieren muss. In file1.c wird jedes gelesene Zeichen mit EOF
verglichen.
*/

```

8.2.2. Generate_overlap

```

/*
generate_overlap.c

Input card
-200.0 200.0 -200.0 200.0 -200.0 200.0 => Limits for grid points; extends over protein
 1.0 1.0 1.0 => Interval for grid points
 1.0 2.0 3.0 5.0 1.0 => xyz coordinates of donor A, mean value and error
 1.0 2.0 3.0 5.0 1.0 => xyz coordinates of donor B, mean value and error
 1.0 2.0 3.0 5.0 1.0 => xyz coordinates of donor C, mean value and error
 1.0 2.0 3.0 5.0 1.0 => xyz coordinates of donor D, mean value and error
 1.0 2.0 3.0 5.0 1.0 => xyz coordinates of donor E, mean value and error
 1.0 2.0 3.0 5.0 1.0 => xyz coordinates of donor F, mean value and error
 1.0 2.0 3.0 5.0 1.0 => xyz coordinates of donor G, mean value and error
 1.0 2.0 3.0 5.0 1.0 => xyz coordinates of donor H, mean value and error
 O => atom type for output pdb file
 WAT => residue type for output pdb file
 X => chain ID for output pdb file
 END
*/

#include<stdio.h>
#include<ctype.h>
#include<string.h>
#include<math.h>
#include<stdlib.h>

FILE *input_data_file,*output_log_file,*output_pdb_file;

main (int argc , char *argv [])
{
  char xname[80],wname[80],pname[80];
  char type[10];
  char atom_type[3],resid_type[4],chain_id[2];
  double x,y,z;
  int atom_num=0,resid_num=0;
  double distance_a , distance_b, distance_c ;
  double distance_d , distance_e, distance_f ;
  double distance_g , distance_h ;
  double x_a , y_a , z_a , aver_dist_a , error_dist_a ;
  double x_b , y_b , z_b , aver_dist_b , error_dist_b ;
  double x_c , y_c , z_c , aver_dist_c , error_dist_c ;
  double x_d , y_d , z_d , aver_dist_d , error_dist_d ;
  double x_e , y_e , z_e , aver_dist_e , error_dist_e ;
  double x_f , y_f , z_f , aver_dist_f , error_dist_f ;
  double x_g , y_g , z_g , aver_dist_g , error_dist_g ;
  double x_h , y_h , z_h , aver_dist_h , error_dist_h ;
  double x_start , x_stop , delta_x ;
  double y_start , y_stop , delta_y ;

```



```

double z_start , z_stop , delta_z ;

switch (argc) {
  case 3 : strcpy(xname,argv[1]);
           strcpy(wname,argv[2]);
           strcpy(pname,argv[3]);
           break;
  case 2 : strcpy(xname,argv[1]);
           strcpy(wname,argv[1]);
           strcpy(pname,argv[1]);
           strcat(xname,".dat");
           strcat(wname,".log");
           strcat(pname,".pdb");
           break;
  default :
           printf("FRET data file: ");
           scanf("%s",xname);
           printf("Log file name: ");
           scanf("%s",wname);
           printf("PDB file name: ");
           scanf("%s",pname);
           break;
}
if(!(input_data_file = fopen(xname,"r"))) {
  printf("no such file %s\n",xname);
  exit(0);
}
output_log_file = fopen(wname,"w");
output_pdb_file = fopen(pname,"w");

do {
  if (fscanf (input_data_file, "%lf%lf%lf%lf%lf%lf",
    &x_start,&x_stop,&y_start,&y_stop,&z_start,&z_stop) != 6) goto Fehler1 ;
  printf("Record1: %lf%lf%lf%lf%lf%lf\n",
    x_start,x_stop,y_start,y_stop,z_start,z_stop);

  if (fscanf (input_data_file, "%lf%lf%lf",
    &delta_x,&delta_y,&delta_z) != 3) goto Fehler2 ;
  printf("Record2: %lf%lf%lf\n",
    delta_x,delta_y,delta_z);

  if (fscanf (input_data_file, "%lf%lf%lf%lf%lf",
    &x_a,&y_a,&z_a,&aver_dist_a,&error_dist_a) != 5) goto Fehler3 ;
  printf("Record3: %lf%lf%lf%lf%lf\n",
    x_a,y_a,z_a,aver_dist_a,error_dist_a);

  if (fscanf (input_data_file, "%lf%lf%lf%lf%lf",
    &x_b,&y_b,&z_b,&aver_dist_b,&error_dist_b) != 5) goto Fehler4 ;
  printf("Record4: %lf%lf%lf%lf%lf\n",
    x_b,y_b,z_b,aver_dist_b,error_dist_b);

  if (fscanf (input_data_file, "%lf%lf%lf%lf%lf",
    &x_c,&y_c,&z_c,&aver_dist_c,&error_dist_c) != 5) goto Fehler5 ;
  printf("Record5: %lf%lf%lf%lf%lf\n",
    x_c,y_c,z_c,aver_dist_c,error_dist_c);

  if (fscanf (input_data_file, "%lf%lf%lf%lf%lf",
    &x_d,&y_d,&z_d,&aver_dist_d,&error_dist_d) != 5) goto Fehler6 ;
  printf("Record6: %lf%lf%lf%lf%lf\n",

```

```

    x_d,y_d,z_d,aver_dist_d,error_dist_d);

if (fscanf(input_data_file, "%lf%lf%lf%lf%lf",
    &x_e,&y_e,&z_e,&aver_dist_e,&error_dist_e) != 5) goto Fehler7 ;
printf("Record7: %lf%lf%lf%lf%lf\n",
    x_e,y_e,z_e,aver_dist_e,error_dist_e);

if (fscanf(input_data_file, "%lf%lf%lf%lf%lf",
    &x_f,&y_f,&z_f,&aver_dist_f,&error_dist_f) != 5) goto Fehler8 ;
printf("Record8: %lf%lf%lf%lf%lf\n",
    x_f,y_f,z_f,aver_dist_f,error_dist_f);

if (fscanf(input_data_file, "%lf%lf%lf%lf%lf",
    &x_g,&y_g,&z_g,&aver_dist_g,&error_dist_g) != 5) goto Fehler12 ;
printf("Record9: %lf%lf%lf%lf%lf\n",
    x_g,y_g,z_g,aver_dist_g,error_dist_g);

if (fscanf(input_data_file, "%lf%lf%lf%lf%lf",
    &x_h,&y_h,&z_h,&aver_dist_h,&error_dist_h) != 5) goto Fehler13 ;
printf("Record10: %lf%lf%lf%lf%lf\n",
    x_h,y_h,z_h,aver_dist_h,error_dist_h);

if (fscanf(input_data_file, "%s ",atom_type) != 1) goto Fehler9 ;
printf("Record9: %s\n",atom_type);
if (fscanf(input_data_file, "%s ",resid_type) != 1) goto Fehler10 ;
printf("Record10: %s\n",resid_type);
if (fscanf(input_data_file, "%s ",chain_id) != 1) goto Fehler11 ;
printf("Record11: %s\n",chain_id);

printf("Start of calculation\n");

x = x_start ;
y = y_start ;
z = z_start ;

for (x = x_start ; x <= x_stop ; x = x + delta_x)
  { for (y = y_start ; y <= y_stop ; y = y + delta_y)
      { for (z = z_start ; z <= z_stop ; z = z + delta_z)
          {
              distance_a = sqrt((x-x_a)*(x-x_a) + (y-y_a)*(y-y_a) + (z-z_a)*(z-z_a)) ;
              distance_b = sqrt((x-x_b)*(x-x_b) + (y-y_b)*(y-y_b) + (z-z_b)*(z-z_b)) ;
              distance_c = sqrt((x-x_c)*(x-x_c) + (y-y_c)*(y-y_c) + (z-z_c)*(z-z_c)) ;
              distance_d = sqrt((x-x_d)*(x-x_d) + (y-y_d)*(y-y_d) + (z-z_d)*(z-z_d)) ;
              distance_e = sqrt((x-x_e)*(x-x_e) + (y-y_e)*(y-y_e) + (z-z_e)*(z-z_e)) ;
              distance_f = sqrt((x-x_f)*(x-x_f) + (y-y_f)*(y-y_f) + (z-z_f)*(z-z_f)) ;
              distance_g = sqrt((x-x_g)*(x-x_g) + (y-y_g)*(y-y_g) + (z-z_g)*(z-z_g)) ;
              distance_h = sqrt((x-x_h)*(x-x_h) + (y-y_h)*(y-y_h) + (z-z_h)*(z-z_h)) ;

if ( distance_a >= (aver_dist_a - (error_dist_a)/(2)))
  { if ( distance_a <= (aver_dist_a + (error_dist_a)/(2)))
    { if ( distance_b >= (aver_dist_b - (error_dist_b)/(2)))
      { if ( distance_b <= (aver_dist_b + (error_dist_b)/(2)))
        { if ( distance_c >= (aver_dist_c - (error_dist_c)/(2)))
          { if ( distance_c <= (aver_dist_c + (error_dist_c)/(2)))
            { if ( distance_d >= (aver_dist_d - (error_dist_d)/(2)))
              { if ( distance_d <= (aver_dist_d + (error_dist_d)/(2)))
                { if ( distance_e >= (aver_dist_e - (error_dist_e)/(2)))
                  { if ( distance_e <= (aver_dist_e + (error_dist_e)/(2)))
                    { if ( distance_f >= (aver_dist_f - (error_dist_f)/(2)))
                      { if ( distance_f <= (aver_dist_f + (error_dist_f)/(2)))

```

```

        { if ( distance_g >= (aver_dist_g - (error_dist_g)/(2)))
          { if ( distance_g <= (aver_dist_g + (error_dist_g)/(2)))
            { if ( distance_h >= (aver_dist_h - (error_dist_h)/(2)))
              { if ( distance_h <= (aver_dist_h + (error_dist_h)/(2)))
                {
                  atom_num++;
                  resid_num++;
                  fprintf(output_pdb_file,"ATOM %4d%3s %4s %1s%4d %8.3f%8.3f%8.3f 1.00 20.00\n",
                    atom_num,atom_type,resid_type,chain_id,resid_num,x,y,z);
                }
              }
            }
          }
        }
      }
    }
  }
}

fprintf(output_log_file,"Protein XYZ Min-Max-Bereich...: %8.3f %8.3f %8.3f %8.3f %8.3f %8.3f
\n",
  x_start,x_stop,y_start,y_stop,z_start,z_stop);
fprintf(output_log_file,"Gridstufung für Berechnung...: %8.3f %8.3f %8.3f\n",
  delta_x,delta_y,delta_z);
fprintf(output_log_file,"XYZ of donor A and lower and upper limit: %8.3f %8.3f %8.3f %8.3f
%8.3f\n",
  x_a,y_a,z_a,aver_dist_a,error_dist_a);
fprintf(output_log_file,"XYZ of donor B and lower and upper limit: %8.3f %8.3f %8.3f %8.3f
%8.3f\n",
  x_b,y_b,z_b,aver_dist_b,error_dist_b);
fprintf(output_log_file,"XYZ of donor C and lower and upper limit: %8.3f %8.3f %8.3f %8.3f
%8.3f\n",
  x_c,y_c,z_c,aver_dist_c,error_dist_c);
fprintf(output_log_file,"XYZ of donor D and lower and upper limit: %8.3f %8.3f %8.3f %8.3f
%8.3f\n",
  x_d,y_d,z_d,aver_dist_d,error_dist_d);
fprintf(output_log_file,"XYZ of donor E and lower and upper limit: %8.3f %8.3f %8.3f %8.3f
%8.3f\n",
  x_e,y_e,z_e,aver_dist_e,error_dist_e);
fprintf(output_log_file,"XYZ of donor F and lower and upper limit: %8.3f %8.3f %8.3f %8.3f
%8.3f\n",
  x_f,y_f,z_f,aver_dist_f,error_dist_f);
fprintf(output_log_file,"XYZ of donor G and lower and upper limit: %8.3f %8.3f %8.3f %8.3f
%8.3f\n",
  x_g,y_g,z_g,aver_dist_g,error_dist_g);
fprintf(output_log_file,"XYZ of donor H and lower and upper limit: %8.3f %8.3f %8.3f %8.3f
%8.3f\n",
  x_h,y_h,z_h,aver_dist_h,error_dist_h);

if(!strcmp(type,"END ")) break;

```

```
} while (0);

fprintf(output_pdb_file,"END");
printf("End of calculation \n");

fclose(output_log_file);
fclose(output_pdb_file);
fclose(input_data_file);

return 0 ;

exit(0);

Fehler1:
printf ("Error reading min_max range of protein \n") ;
fclose (input_data_file) ;
return 1 ;

Fehler2:
printf ("Error reading delta_xyz line \n") ;
fclose (input_data_file) ;
return 1 ;

Fehler3:
printf ("Error reading donor A line \n") ;
fclose (input_data_file) ;
return 1 ;

Fehler4:
printf ("Error reading donor B line \n") ;
fclose (input_data_file) ;
return 1 ;

Fehler5:
printf ("Error reading donor C line \n") ;
fclose (input_data_file) ;
return 1 ;

Fehler6:
printf ("Error reading donor D line \n") ;
fclose (input_data_file) ;
return 1 ;

Fehler7:
printf ("Error reading donor E line \n") ;
fclose (input_data_file) ;
return 1 ;

Fehler8:
printf ("Error reading donor F line \n") ;
fclose (input_data_file) ;
return 1 ;

Fehler9:
printf ("Error reading atom name \n") ;
fclose (input_data_file) ;
return 1 ;

Fehler10:
```

```

printf ("Error reading residue name \n");
fclose (input_data_file);
return 1;

Fehler11:
printf ("Error reading chain id \n");
fclose (input_data_file);
return 1;

Fehler12:
printf ("Error reading donor G line \n");
fclose (input_data_file);
return 1;

Fehler13:
printf ("Error reading donor H line \n");
fclose (input_data_file);
return 1;
}

/* Der "File-Pointer" darf vom Programmierer als "Pointer auf ein 'Objekt
vom Typ FILE" angesehen werden, der von fopen geliefert wird und fuer
alle nachfolgenden File-Operation (Lesen, Schreiben, Schliessen, ...)
als Identifikator verwendet wird (dass sich hinter dem Type FILE eine
in stdio.h beschriebene Struktur verbirgt, ist fuer den Programmierer
von geringem Interesse). Der File-Pointer muss in der Form

        FILE *infile_p

vereinbart werden, wobei fuer den hier gewaehlten Namen infile_p ein
beliebiger Name stehen darf. */

/* Ein File wird z. B. mit

        infile_p = fopen (argv [1], "r");

geoeffnet, wobei zwei Argumente angegeben werden muessen:

* Das erste Argument ist der Filename, der als relativer oder absoluter
  Filename angegeben werden darf, also gegebenenfalls auch den gesamten
  Pfad (und unter DOS eine Laufwerksbezeichnung) enthalten kann.

* Das zweite Argument kennzeichnet die beabsichtigte Verwendung des
  geoeffneten Files, besonders wichtig sind folgende Moeglichkeiten:

  "r" oeffnet zum Lesen,
  "w" oeffnet zum Schreiben, der Inhalt eines mit dem angegebenen Namen
  bereits existierenden Files geht dabei verloren,
  "a" ('append') oeffnet zum Schreiben ab File-Ende eines bereits
  existierenden Files.

Als Return-Wert liefert fopen den fuer nachfolgende Aktionen benoetigten
File-Pointer oder NULL (bei Misserfolg, passiert z. B. beim Oeffnen eines
nicht existierenden Files mit "r" oder eines existierenden schreib-
geschuetzten Files mit "w"), sollte beim Oeffnen unbedingt abgefragt
werden. Dies kann (kuerzer als oben programmiert) z. B. mit

        if ((infile_p = fopen (argv [1], "r")) != NULL)

```

```

{
  ** File erfolgreich geoeffnet **
}

```

erfolgen, weil in C eine Wertzuweisung (`infile_p = ...`) eingeklammert werden darf und die Klammer noch einmal den zugewiesenen Wert repraesentiert, der dann fuer eine Vergleichsoperation verwendet werden kann. */

/* Mit der 'stdio'-Funktion

```
i = fgetc (infile_p);
```

wird genau ein Zeichen gelesen und als 'int'-Wert abgeliefert (in `file1.c` wird dieser Return-Wert ignoriert, weil nur die Anzahl der zu lesenden Zeichen ermittelt werden soll).

Die Frage, warum eine Funktion, die ein Zeichen liest, nicht einen Return-Wert vom Typ 'char' abliefert, wird durch den Sonderfall beantwortet: Wenn das File-Ende erreicht ist, liefert 'fgetc' das "End-of-File"-Zeichen EOF, fuer das (zumindest in einem Zeichensatz mit 256 Zeichen) kein Platz waere, weil mit einem Byte (Typ 'char') nur genau 256 Zeichen darstellbar sind. EOF ist in `stdio.h` definiert, ueblicherweise als -1, was allerdings den Programmierer nicht interessieren muss. In `file1.c` wird jedes gelesene Zeichen mit EOF verglichen. */

8.2.3 Input cards

8.2.3.1 Input card for the p_5_Cy5/t (P-complex)

-200.0	200.0	-200.0	200.0	-200.0	200.0
2.0	2.0	2.0			
90.110	177.650	65.630	69.0	12.1	
48.790	135.030	85.690	69.0	14.4	
115.020	133.490	90.420	86.2	18.6	
35.520	123.980	55.250	45.4	10.4	
83.600	187.660	34.210	62.3	12.5	
110.490	176.320	12.020	200.0	400.0	
123.940	138.600	12.410	62.1	7.4	
54.780	115.850	-7.040	35.2	4.4	
O					
WAT					
Z					
END					

8.2.3.2 Input card for the p_10bp_Cy5/t (P-complex)

```

-200.0  200.0  -200.0  200.0  -200.0  200.0
2.0      2.0      2.0
  90.110 177.650  65.630  42.8   10.2
  48.790 135.030  85.690  38.8   26.0
 106.000 139.820  97.840  49.7   16.0
  35.520 123.980  55.250  37.1   16.4
  83.600 187.660  34.210  69.3   31.4
 110.490 176.320  12.020 200.0  400.0
 123.940 138.600  12.410  75.4   40.8
  54.780 115.850  -7.040  75.9   26.2
O
WAT
Y
END

```

8.2.3.3 Input card for the p_3_Cy5/t (P-complex)

```

-200.0  200.0  -200.0  200.0  -200.0  200.0
2.0      2.0      2.0
  90.110 177.650  65.630  42.5   7.0
  48.790 135.030  85.690  37.7  12.8
 115.020 133.490  90.420  46.5   7.6
  35.520 123.980  55.250  40.8  10.8
  83.600 187.660  34.210  65.9  14.8
 110.490 176.320  12.020 200.0 400.0
 123.940 138.600  12.410  71.4  14.1
  54.780 115.850  -7.040  75.3  14.6
N
WAT
X
END

```

8.2.3.4 Input card for the p/t_7o_Cy5 (P-complex)

```

-200.0  200.0  -200.0  200.0  -200.0  200.0
2.0      2.0      2.0
  90.110 177.650  65.630  48.6   9.8
  48.790 135.030  85.690  36.9   9.2
 115.020 133.490  90.420  50.2   9.0
  35.520 123.980  55.250  38.3   9.0
  83.600 187.660  34.210  68.6  15.7
 110.490 176.320  12.020 200.0 400.0
 123.940 138.600  12.410  74.8  21.4
  54.780 115.850  -7.040  75.2  17.7
O
WAT
W
END

```

8.2.3.5 Input card for the p/t_5_Cy5 (P-complex)

```

-200.0  200.0  -200.0  200.0  -200.0  200.0
2.0      2.0      2.0
  90.110  177.650  65.630  49.0  10.6
  48.790  135.030  85.690  40.2  10.3
 115.020  133.490  90.420  51.0   7.7
  35.520  123.980  55.250  42.3  11.0
  83.600  187.660  34.210  61.7  15.0
 110.490  176.320  12.020 200.0 400.0
 123.940  138.600  12.410  76.5  11.2
  54.780  115.850  -7.040  75.1  21.2
O
WAT
V
END

```

8.2.3.6 Input card for the p_5_Cy5/t (DE-complex)

```

-200.0  200.0  -200.0  200.0  -200.0  200.0
2.1      2.1      2.1
  90.110  177.650  65.630  47.8  16.0
  48.790  135.030  85.690 200.0 400.0
 106.000  139.820  97.840  51.7  29.0
  35.520  123.980  55.250  79.9  13.9
  83.600  187.660  34.210 200.0 400.0
 110.490  176.320  12.020 200.0 400.0
 123.940  138.600  12.410 200.0 400.0
  54.780  115.850  -7.040  87.8  39.4
N
WAT
U
END

```

8.2.3.7 Input card for the p_10bp_Cy5/t (DE-complex)

```

-200.0  200.0  -200.0  200.0  -200.0  200.0
2.0      2.0      2.0
  90.110  177.650  65.630  85.0  86.0
  48.790  135.030  85.690  84.2  58.6
 106.000  139.820  97.840  91.3  85.6
  35.520  123.980  55.250  86.2  61.6
  83.600  187.660  34.210 200.0 400.0
 110.490  176.320  12.020 200.0 400.0
 123.940  138.600  12.410 200.0 400.0
  54.780  115.850  -7.040 200.0 400.0
C
WAT
T
END

```


8.2.3.8 Input card for the p_3_Cy5/t (DE-complex)

```

-200.0  200.0  -200.0  200.0  -200.0  200.0
2.0      2.0      2.0
  90.110 177.650  65.630  79.8  63.0
  48.790 135.030  85.690  87.1  69.0
 115.020 133.490  90.420  93.3  67.8
  35.520 123.980  55.250  89.5  60.2
  83.600 187.660  34.210 200.0 400.0
 110.490 176.320  12.020 200.0 400.0
 123.940 138.600  12.410 200.0 400.0
  54.780 115.850  -7.040 200.0 400.0
O
WAT
S
END

```

8.2.3.9 Input card for the p/t_7o_Cy5 (DE-complex)

```

-200.0  200.0  -200.0  200.0  -200.0  200.0
2.0      2.0      2.0
  90.110 177.650  65.630  81.6  91.0
  48.790 135.030  85.690  87.2  60.0
 115.020 133.490  90.420  92.7  78.0
  35.520 123.980  55.250  90.4  69.0
  83.600 187.660  34.210 200.0 400.0
 110.490 176.320  12.020 200.0 400.0
 123.940 138.600  12.410 200.0 400.0
  54.780 115.850  -7.040 200.0 400.0
O
WAT
R
END

```

8.2.3.10 Input card for the p/t_5_Cy5 (DE-complex)

```

-200.0  200.0  -200.0  200.0  -200.0  200.0
2.0      2.0      2.0
  90.110 177.650  65.630  79.7  63.8
  48.790 135.030  85.690  79.6  58.8
 115.020 133.490  90.420 200.0 400.0
  35.520 123.980  55.250  80.3  49.3
  83.600 187.660  34.210 200.0 400.0
 110.490 176.320  12.020 200.0 400.0
 123.940 138.600  12.410 200.0 400.0
  54.780 115.850  -7.040 200.0 400.0
O
WAT
Q
END

```

Curriculum Vitae

Name: Paul Rothwell
Date of birth: 13/12/75
Place of birth: Shotley Bridge,
County Durham,
England
U.K.

Education

Oct 1997-1998: University of York, York, U.K.
Master of Research (MRes) in Biomolecular Science
Projects: *Expression studies on an inactive precursor of penicillin acylase (Department of Chemistry, University of York)*
Towards determining the role of SecB in transport of heat-labile Enterotoxin (Department of Biology, University of York)
The use of a micro-volume pressure chamber to characterise states in the acto-myosin and acto-myosin.ADP complexes (Max-Planck-Institut fuer Molekulare Physiologie Dortmund)

Sep 1994-1997: University of York, York, U.K.
Bachelor of Science (BSc) in Biochemistry
Grade: 2:2
Projects: *Sequence independent ligand binding in an oligopeptide transporter protein (Department of Chemistry, University of York)*

June 1992-1994: St Benedicts RC High School, Whitehaven, Cumbria, U.K.
A levels
Grades: Biology (B), Chemistry (C), Mathematics (C), General studies (C)

June 1993-1994: St Benedicts RC High School, Whitehaven, Cumbria, U.K.
G.C.S.E
Grades: Science (B,B), Mathematics (C), English Literature (C), English Language (C), History (D), Craft Design and communications (D), Business Studies (E)

Publications

Multi-parameter single-molecule fluorescence spectroscopy reveals three different HIV-1 reverse transcriptase :primer/template complexes

P.J. Rothwell, S. Berger, O. Kensch, S. Felekyan, M. Antonik, B.M. Wöhrl, T. Restle, R.S. Goody and C.A.M. Seidel

(In preparation)

Conferences

4th International Weber symposium on “Innovative Fluorescence Methodologies in Biochemistry and Medicine”, Kaanapali, Maui, Hawaii (June 23- June 27, 1999 (Poster))

Volkswagen stiftung symposium “New Trends in Physics, Chemistry, and Biology with Single Molecules”, Wiesbaden, Germany (July 14th-16th, 1999 (**Poster**))

4th Annual meeting on “Structure-Function Analysis of Drug Resistant HIV-1 Reverse Transcriptase Having New Improved Selection Properties”, Consiglio Nazionale delle Ricerche, Italy November 11th-14th, 1999 ((**Talk**))

Volkswagen stiftung “Physics, Chemistry, and Biology with Single Molecules”, Banz, Germany (March 5th-7th, 2001 (**Poster-1st Prize for the best poster presentation**))

Acknowledgements

I would first and foremost like to thank my PhD supervisor Prof. Dr. Roger S. Goody for the opportunity to work in his department, and for his advice and support during the course of my PhD. I would also like to thank my second supervisor Prof. Dr. T. N. Mitchell for excellent back up supervision.

I am very grateful to all members of the HIV-1 group for help and advice during my studies in Dortmund. I would like to thank: Oliver Kensch who gave me a crash course in molecular biology and was an integral part of the establishment of my project; Tobias Restle and Birgitta Wöhrl for their continued support, and enthusiasm for the work, as well as for reading my dissertation; and last but not least our excellent technicians- Karin Vogel-Bachmayr and Martina Wischnewski.

Special thanks are due to Axel Scheidig, who developed software for the analysis of the MFD data.

I would also like to thank all members of the Single-Molecule Fluorescence Spectroscopy at the Max-Planck-Institut für biophysikalische Chemie (Karl-Friedrich-Bonhoeffer-Institut) in Göttingen for making my trips to there enjoyable and fulfilling. Special thanks go to Dr Claus Seidel for his support and suggestions for single molecule experiments. Also, Sylvia Berger and Suren Felekyan who were responsible for analysing the vast amount of single molecule data collected (40 Gb) and for converting it into a usable format, as well as for introducing me to the basics of single molecule spectroscopy.

Last, but not least, I would like to thank Tresfore Dambe for a careful reading of the manuscript.