AFRICAN SWINE FEVER VIRUS DNA POLYMERASE X: BIOPHYSICAL INTERACTION STUDIES AND NMR ASSIGNMENTS OF THE POLYMERASE-DEOXYGUANOSINE TRIPHOSPHATE COMPLEX

By

Markus Wolfgang Voehler

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry

December, 2007

Nashville, Tennessee

Approved:

Professor Brian O. Bachmann

Professor Frederick P. Guengerich

Professor Andrzej M. Krezel

Professor Sandra J. Rosenthal

Professor Michael P. Stone

To my children, Nicole, Dominic and Melanie

and

To my wife, Barbara, never ceasing to support

ACKNOWLEDGMENTS

Not in my wildest dreams did I imagine to work on such a thesis when I arrived in Nashville to work as NMR spectroscopist at Vanderbilt University 15 years ago. The opportunity arose and thanks to the encouragement of Dr. Andrzej M. Krezel, Drs. M. Thomas and Constance M. Harris, Dr. Michael P. Stone and my wife Barbara Voehler, I stepped out and “dove” into this admittedly sometimes wild endeavor. Juggling my full attention between an expanding NMR facility and the research project with graduate school was not always easy, but I consider myself very fortunate having been given the opportunity to participate in this program. I also knew that I always could count on those persons and had their support throughout the years.

I would like to thank my advisor, Dr. Michael P. Stone for giving me the opportunity to do this research in his laboratory, for everything he has taught me, all the resources he provided, and the patience he had, when the research did not move forward as quickly as desired due to my other commitments in the NMR facility.

I would like to thank my Ph.D. committee: Dr. Brian O. Bachmann, Dr. Frederick P. Guengerich, Dr. Andrzej M. Krezel, and Dr. Sandra J. Rosenthal. Special thanks goes to Drs. Guengerich and Krezel who served on the Academic Research Monitoring

committee and provided much insight to the project, stimulated new ideas through their comments and discussions, engaged in hands on support through their laboratory, and taught me invaluable lessons in the classroom.

I would like to thank all the members of the Stone laboratory. Special thank goes to Kyle Brown for his friendship and encouragement and some former members of the laboratory, Drs. Tandace Scholdberg, Keither Merritt, Irene Zegar, and Nathalie Schnetz-Boutaud. Two undergraduate students, Amy Gordon and Jerry Yen, spent many hours helping with the protein production for which I am very grateful.

I am also thankful for some great colleagues in the NMR facility, Dr. Donald Stec and Margaret Allamann, as well as Dr. Jaison Jacob, who taught me invaluable lesions in biomolecular NMR and remained a source of knowledge and a friend even after leaving Vanderbilt.

I would like to thank all the members of the Chemistry Department, Center for Structural Biology, Center in Molecular Toxicology, and the Vanderbilt-Ingram Cancer Center for their support, contributions, and resources they make available. Special thanks goes to Mrs. Paulette Lynch, Jackie Brown, and Mr. Robert Padgett for ordering all the materials, Mrs. Ellen Rochelle for her Macintosh support, Dr. James Riggins and Robert Eoff for running the activity assays, and Dr. HakJun Kim for his help with

NMRView. I am very appreciative of Drs. Geoffrey Mueller and Eugene DeRose at the National Institute of Environmental Health Sciences for discussing many software issues. Thanks to Dr. Frank Delaglio who kindly provided the titration analysis module of the NMRPipe software package. I am thankful to Vanderbilt University for allowing me to remain fully employed while participating in this wonderful educational institution as a student. This work would not have been possible without the financial support of the National Institute of Health, the Center in Molecular Toxicology, and Vanderbilt University.

Most importantly, a special thank goes out to my family. My wife, Barbara Voehler, lovingly supported my work on a daily bases and was there to share successes and disappointments. Although I missed out on many school or sports events I do not remember a complaint from my children. I thank you, Nicole, Dominic, and Melanie for understanding and being the joy and inspiration you are!

Finally, I want to give God the glory for providing the strength to complete this task.

TABLE OF CONTENTS

Page

ACKNOWLEDGMENT iii

LIST OF TABLES xii

LIST OF FIGURES xiv

LIST OF ABBREVIATIONS xviii

Chapter

I. GENERAL INTRODUCTION 1

POLYMERASES 1

Importance of Genome Integrity *2*

Additional Functionalities of Polymerases *3*

DNA Polymerase Families *5*

Common Subdomains of Polymerases *6*

The Palm Subdomain 8

The Thumb Subdomain 8

The Finger Subdomain 8

Other Subdomains 10

Mechanism of Nucleotidyl Transfer *11*

THE AFRICAN SWINE FEVER VIRUS POLYMERASE X 13

Brief Background *14*

ASF Virus and ASFV Pol X *15*

Activity and Fidelity of Pol X *17*

Structure Elements of ASFV Polymerase *21*

Secondary Structure Comparisons of Pol X with other Family X Polymerases 22

Residue Comparison of Pol X with Pol β 25

II. EXPRESSION, PURIFICATION AND SOLUBILIZATION OF THREE ASFV POL X CONSTRUCTS 27

INTRODUCTION 27
MATERIAL AND METHODS 28
ASFV Pol X Constructs *28*
Sequencing *29*
PAGE *29*
Stock Cultures *30*
ASFV POL X_{orig} CONSTRUCT 30
Protein Expression in LB Medium *30*
Protein Expression in Minimal Medium M9 ..... *32*
Purification of Pol X_{orig} Construct *34*
ASFV POL X_{iz} CONSTRUCT 38
Expression of Pol X_{iz} Construct *38*
Purification of Pol X_{iz} Construct *39*
Microdrop Screening to Optimize the Solvent Conditions for ASFV Pol X_{iz} *44*
Experimental Setup 46
Screen 1: Individual Buffer Systems 49
Screen 2 and 3: pH Variations and Salt Additives 50
Screen 4: NaCl Concentration 51
Screen 5: Include Stabilizers 51
Screen 6 and 7: Conditions for Sodium Acetate and HEPES-Na Based Buffers 52
Conclusions of Solubility Screen Trials 53
ASFV POL X_{gpm} CONSTRUCT 54
ACTIVITY ASSAYS OF POL X 60
Activity of Pol X_{gpm} Pre- and Post HPLC *61*
Activity Pol X_{orig} in the Presence of a Primer-Template or Gapped DNA *63*
CONCLUSIONS 65

III. FLOURESCENCE AND ITC STUDIES OF ASFV POL X – NUCLEOTIDE AND ASFV POL X - DNA – COMPLEXES 68

INTRODUCTION 68
Binding Studies by Fluorescence *74*
Fluorescence Measurements in Protein 75

Binding Studies by Isothermal Titration Calorimetry (ITC) *78*
MATERIAL AND METHODS 80
Chemicals and Buffers *80*
Nucleotides and Oligonucleotides *81*
44-mer 84
36-mer 85
34-mer 85
16-mer 86
Template-Primer 17-9-mer, single-stranded DNA (ss-DNA) 87
Experimental Set-Up and Data Analysis for Fluorescence Measurements *88*
Experimental Set-Up and Data Analysis for ITC Measurements *89*
General Measurement Procedure 91
RESULTS 93
Fluorescence Binding Studies *93*
Binding Study with a 44-mer and 36-mer Duplex DNA 93
Binding Study with a 34-mer DNA Duplex Containing Two Hairpins 100
Binding Study with a 16-mer DNA Duplex 102
Binding Study with a Primer-Template DNA Duplex 106
Binding Studies by Isothermal Titration Calorimetry *107*
CONCLUSIONS 111
Fluorescence Measurements *111*
ITC Measurements *112*

IV. BINDING STUDIES USING NMR 115

INTRODUCTION 115
MATERIALS AND METHODS 120
Experimental Set-Up and Data Analysis of NMR Measurements *120*
Gpm-Buffer 120
NMR Samples 120
Data Analysis 122
Relaxation Measurements By NMR *124*
RESULTS 125
NMR Studies of a Binary Complex Between the ASFV Pol X and Nucleotides *125*
Impact of Magnesium 125
Dissociation Constants for Deoxynucleotide Triphosphates 126

Purines: dGTP and ATP 128
Pyrimidines: dCTP and dTTP 133
NMR Studies of ASFV Pol X - DNA Complexes *134*
Complex Formation with the 36-mer Double Hairpin DNA 135
Complex Formation with the 44-mer Double Hairpin DNA 138
Ternary Complex Formation with 16-mer DNA in the Presence of ddGTP 141
NMR Relaxation Studies for the 36- and 44-mer DNA – Pol X Binary Complex 142
NMR Binding Studies Involving Single Strand And Template-Primer DNA 147
$dC(pdC)_6$ and $dG(pdG)_6$ 149
$dC(pdC)_8$ 149
$dA(pdA)_{11}$ and $dT(pdT)_{11}$ 151
$dA(pdA)_{11}dC(pdC)_5$ 152
$dA(pdA)_{11}dC(pdC)_6$: $dA(pdA)_{11}$ Template-Primer ds-DNA 155
Additional Attempts to Improve Spectral Quality for the Pol X-DNA Complexes *155*
DISCUSSION AND CONCLUSIONS 157
Complex with Nucleotide *157*
Complex with Single- and Double-Stranded-DNA *158*

V. NMR ASSIGNMENT OF ASFV POL X AS BINARY COMPLEX WITH DEOXYGUANOSINE TRIPHOSPHATE 165

INTRODUCTION 165
Brief History *167*
Fundamental Aspects of NMR *168*
Scalar Couplings *171*
Dipolar Couplings *173*
Multi-Pulse Experiments and Gradients *175*
Indirect Detection *176*
Three-Dimensional (3D) Experiments *178*
MATERIALS 181
NMR Sample Preparation *181*
RESULTS 181

Oxidized and Reduced Form of ASFV Pol X *181*
Backbone Assignment *191*
Sidechain Assignment *199*
DISCUSSION AND CONCLUSION 205
Two Conformers of Pol X: Oxidized and Reduced *205*
Chemical Shift Comparison Between Free and dGTP-Complexed Pol X *209*

VI. PERFORMANCE OF CRYOGENIC PROBES AS A FUNCTION OF IONIC STRENGTH AND SAMPLE TUBE GEOMETRY 213

INTRODUCTION 213
MATERIAL AND METHODS 217
Sucrose Test Samples *217*
Protein Samples *217*
RESULTS AND DISCUSSION 220
Factors Affecting the Signal-To-Noise Ratio *220*
NMR Spectroscopy in High Salt Solution Using Cryogenic Probes *221*
Probe Tuning 221
90^0 Degree Pulse 222
Signal-to-Noise Comparisons for a Sucrose Test Sample 224
Systematic Comparison of Signal-To-Noise Data on a Protein Sample *227*
Constant Protein Concentration 228
Constant Amount of Protein 229
CONCLUSION 232

VII. CONCLUDING REMARKS 234

SOLUBILITY OF POL X AT HIGH CONCENTRATIONS 234
OXIDIZED AND REDUCED FORM OF POL X 235
COMPLEX FORMATION 237
CONCLUSION 240

APPENDIX

A. CLEAVAGE OF THE LEADER PEPTIDE 242
Activity of Cleaved Pol X_{orig} *244*

B. SOLUBILITY SCREEN FOR POL X_{iz} 246

Table 1: Main Buffers and pH *247*
Table 2: Main Buffers and pH, 1 mM Protein Solution *248*
Table 3: Main Buffers and pH, 2.5 mM Protein Solution *249*
Table 4: NaCl Addition, 2.5 mM Protein Solution *250*
Table 5: Stabilizers, 2.5 mM Protein Solution *251*
Table 6: Stabilizers, 1mM Protein Solution *252*
Table 7: Stabilizers, 1mM Protein Solution *253*

C. EXPRESSION AND PURIFICATION OF ASFV POL X_{gpm} 254

Solutions for protein expression and purification *254*
Expression and purification protocol *255*

D. ANALYSIS OF THE NMR-TITRATION EXPERIENTS 259

NMRPipe *259*

E. TABLE E-1: TITRATION EXPERIMENT OF ASFV POL X WITH 36-MER DNA 264

TABLE E-2: TITRATION EXPERIMENT OF ASFV POL X WITH 44-MER DNA 266
TABLE E-3: TITRATION EXPERIMENT OF ASFV POL X WITH SS-DNA 268

F. NMR DATA PROCESSING SCRIPTS FOR NMRPIPE 273

HSQC *273*
HNCO *273*
HNCA *274*
CBCA(CO)NH *275*
HNCACB *276*
HBHA(CO)NH *277*
(H)CC(CO)NH *278*
H(CCCO)NH *279*

G. NMR ASSIGNMENT OF ^{1}H, ^{13}C AND ^{15}N RESONANCES OF ASFV POL X_{gpm} IN A COMPLEX WITH DEOXYGUANOSINE-TRIPHOSPHATE 281

REFERENCES 292

LIST OF TABLES

Table		Page
1-1	Main structural elements observed for some Family X polymerases	24
3-1	Physical properties of nucleotide triphosphates	82
3-2	Physical properties and sequences of used single stranded DNA	87
3-3	Binding constants for two different pol X constructs with 44- and 36-mer at various NaCl concentrations.	99
3-4	Binding constants between pol X_{gpm} and a self-annealing 34-mer	102
3-5	Binding constants of pol X_{gpm} with DNA and nucleotides	104
3-6	Binding constants of a template-primer DNA	107
3-7	Summary of ITC measurements	109
4-1	Interactions study of nucleotides and pol X by NMR	128
4-2	Average T_2 times for pol X at four different conditions	144
4-3	Average T_2 times for two different pol X-DNA complexes	144
4-4	Estimated T_2 relaxation values at different conditions	145
5-1	Physical properties of NMR active nuclei used in biomolecular systems	169
5-2	List of backbone experiments used in this work, including HBHA(CO)NH for proton Hα and Hβ assignment.	195
5-3	Select acquisition (A) and processing (B) parameters for backbone experiments	196
5-4	A. Select acquisition parameters for sidechain experiments	201
	B. Select processing parameters for sidechain experiments	202

6-1 Summary of experimental conditions and results comparing ^{15}N-HSQC spectra measured on a 21 kDa pol X polymerase in different NMR tubes 231

LIST OF FIGURES

Figure		Page
1-1	Crystal structures of two DNA polymerase complexes with duplex DNA based on the PDB coordinates pol β (PDB code 1BPY) and pol T7 (PDB code 1T7P)	7
1-2	Three crystal structures indicating the domain movement of pol β	9
1-3	Active site of ternary complex in the pre-catalytic state of nucleotidyl transfer	12
1-4	Secondary structure elements on solution structure of free ASFV pol X	22
2-1	Cell expression curves in minimal medium cultures under various conditions	34
2-2	Imidazole based purification of pol X_{iz} using a Ni-NTA resin	41
2-3	SDS-PAGE of pol X_{iz} purified on a cobalt or nickel based affinity column	43
2-4	Chromatogram of gel filtration with Sephacryl S-100 HR column	43
2-5	Schematic drawing of sample preparation for a single solubility screen	47
2-6	Purification of pol X_{gpm}	57
2-7	HPLC purification of pol X_{gpm}	58
2-8	MALDI TOF mass spectrum of ^{15}N labeled pol X_{gpm}	58
2-9	Superposition of ^{15}N-HSQC spectra of pol X_{gpm} after CM-20 ion exchange and HPLC purification	60
2-10	Activity assay of pol X_{gpm} samples taken before and after HPLC purification	62
2-11	Activity assay of pol X_{orig}	64
3-1	Schematic fluorescence emission spectra	76

3-2 44-mer duplex DNA 85

3-3 36-mer duplex DNA 85

3-4 34-mer duplex DNA 86

3-5 16-mer duplex DNA 86

3-6 Position of tryptophan and tyrosine in solution structure of pol X 95

3-7 Representative fluorescence emission spectra from a titration of pol X_{gpm} 97

3-8 Proton NMR spectrum at 500 MHz of 16-mer CpC gapped DNA 103

3-9 ITC measurements of pol X, 40 μM, with all four deoxynucleotides 108

3-10 Comparison of the binding between pol X and adenosine-triphosphate or deoxyadenosine-triphosphate 109

4-1 Schematic representation HSQC spectrum changes due to different exchange rates in the presence of a small ligand 118

4-2 Spectrum matrix output from the NMRPipe module titrView 124

4-3 Pol X alone, with dCTP and dCTP plus Mg^{2+} 126

4-4 Comparison of effects seen by purines and pyrimidines in HSQC spectra 127

4-5 Titration spectra of pol X_{gpm} with dGTP 129

4-6 Mapping of dGTP and ATP interactions with pol X 131

4-7 Superposition of pol X ^{15}N-HSQC spectra at different 36-mer DNA concentrations 137

4-8 Pol X-36-mer interaction surface map 138

4-9 ^{15}N-HSQC spectra of pol X - 44-mer DNA complex 139

4-10 44-mer - pol X interaction surface mapping 140

4-11 Ternary ^{15}N-HSQC spectra for pol X with ddGTP and various amounts of 16-mer 142

4-12 Binding interface with dC(pdC)$_8$ 150

4-13 Binding interface with dA(pdA)$_{11}$ 152

4-14 Superposition of ^{15}N-HSQC spectra when titrating pol X with dA(pdA)$_{11}$dC(pdC)$_6$ 153

4-15 Binding interface with dA(pdA)$_{11}$dC(pdC)$_6$ 154

4-16 Conformational changes in pol β structures 161

4-17 Gapped DNA modeled on to free pol X structure visualizing potential complex formation modes in accordance with the NMR mapping results 163

5-1 Important scalar coupling values between nuclei used in fully carbon and nitrogen labeled protein experiments 172

5-2 Schematic representation of a 3D experiment 180

5-3 Amide chemical shift change on native pol X over time 183

5-4 Chemical shift change for residues Val 65 and Lys 85 over time 186

5-5 Monitoring the resonance shifts after addition of 100 mM DTT-d10 to an NMR sample containing the pol X – dGTP complex 188

5-6 Activity tests on ASFV pol X in the oxidized and reduced form 190

5-7 Correlations in backbone experiments used 192

5-8 NvAssign CBCA layout to assign sequential clusters 198

5-9 Example of a spectral layout for the proton and carbon sidechain assignment 204

5-10 Alignment of the two ASFV pol X structures in white with the rat pol β structure 207

5-11 Amide chemical shift differences between native and complex pol X 210

5-12 Chemical shift changes between amide resonance of native pol X and pol X_{gpm} – dGTP complex mapped to the structure of native pol-X 211

6-1 Dual NMR sample tube arrangement 219

6-2 A. $\pi/2$ pulse length of a 10 mM sucrose test samples containing NaCl concentrations varying between 0 and 4M 223

6-3 Signal to noise data of 10 mM sucrose, 3 mm NMR tube as a function of the NaCl concentration 225

A-1 SDS-PAGE of a pilot cleavage reactions with increasing incubation time 242

A-2 SDS-PAGE of an NMR sample digestion 243

A-3 Primer extension assay on pol X_{orig} after cleavage with rEK 245

LIST OF ABBREVIATIONS

$\hbar$	Planck constant (6.626068 10^{-34} Js)
1D	One dimensional
2D	Two dimensional
3D	Three dimensional
Ac	Acetate
AP	Apurinc, abasic site
ASFV	African Swine Fever Virus
ATP	Adenosine 5'-triphosphate
BER	Base excision repair
BME	β-mercaptoethanol
BSA	Bovine serum albumin
CGE	Capillary Gel Electrophoresis
CHAPS	3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate
CT	Constant time
dATP	2'- deoxyadenosine 5'-triphosphate
dCTP	2'- deoxycytidine 5'-triphosphate
ddATP	2',3'-dideoxyadenosine 5'-triphosphate
ddCTP	2',3'-dideoxycytidine 5'-triphosphate
ddGTP	2',3'-dideoxyguanosine 5'-triphosphate
ddTTP	2',3'-dideoxythymidine 5'-triphosphate
dGTP	2'- deoxguanosine 5'-triphosphate
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic Acids
dNTP	deoxynucleotide triphosphate

ds-DNA	Double stranded Deoxyribonucleic Acids
DTT	Dithiothreitol or (*2S,3S*)-1,4-Bis-sulfanylbutane-2,3-diol (IUPAC)
dTTP	2'-deoxythymidine 5'-triphosphate
dUMPNPP	2'-deoxyuridine 5'-(alpha,beta-imido)triphosphate
EDTA	Ethylene diamine tetracetic acid
ER	Endoplasmic reticulum
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High performance liquid chromatography
HSQC	Heteronuclear single Quantum Coherence
IPTG	Isopropyl β-D-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
k_B	Boltzmann constant ($1.3805\ 10^{-23} JK^{-1}$)
LMW	Low Molecular Weight marker (Pharmacia, 14.4, 20.1, 30, 43, 67 and 94 kDa)
MALDI	Matrix-assisted laser desorption/ionization
MES	2-(N-morpholino)ethanesulfonic acid
Mf	mol fraction
MWCO	Molecular Weight Cut Off
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear Magnetic Resonance
ORF	Open reading frame
PDB	Protein Data Bank
PIPES	1,4-Piperazinediethanesulfonic acid (IUPAC)
Pol X	African Swine Fever Virus Polymerase X
Pol	Polymerase
Ppm	Parts per million
rEK	Recombinant Enterokinase
RNA	Ribonucleic Acids

SCT	Semi constant time
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
ss-DNA	Single stranded Deoxyribonucleic Acids
T_1	Longitudinal relaxation
T_2	Transverse relaxation
TdT	Terminal Deoxynucleotidyl Transferase
TFA	Trifluoroacetic acid
Tris	Trishydroxymethylaminomethane or 2-amino-2-hydroxymethyl-1,3-propanediol (IUPAC)
TSP-d4	Trimethylsilyl-2,2,3,3-tetradeuteropropionic acid
UV	Ultraviolet
VIS	Visible
τ_c	Overall correlation time
τ_m	Mixing time

CHAPTER I

GENERAL INTRODUCTION

Polymerases

Today, two DNA polymerases have been identified in the African Swine Fever Virus (ASFV). One of them is involved in the replication of the viral genome, while the other polymerase (ASFV pol X, or also referred to as pol X) has a high sequence and structural homology to polymerases of the Family X, but its function is not identified yet. This work will provide an enhanced understanding on the interaction of pol X with nucleotides and single- or double-stranded oligodeoxynucleotides. But what are polymerases in general, what is their importance or characteristics, and how do they work? These questions will be addressed in the following paragraphs.

The genome of an organism contains the entire hereditary information for an individual species in the form of chromosomes within the cell. Each chromosome is a single macromolecule consisting of deoxyribonucleic acids (DNA), or ribonucleic acids (RNA) in some viruses, where the sequence of the individual nucleic acids represents a specific code, the genes that encode for individual proteins, RNA genes, or regulatory

regions. Today, about 20,000 to 25,000 protein-coding genes have been identified in the human genome (*1*), but this represents barely 2% of the roughly three billion bases present in the genome. The remaining 98% are deemed non-coding DNA and have been referred to as “junk DNA”. Yet over the past few years, this picture has experienced rapid shift with studies such as the one from Pennisi et al (*2*), who observed that 80% of a base subset (although including a large portion of non-coding DNA) still shows signs of expression and therefore might have more biological importance than attributed to it previously.

Importance of Genome Integrity

Maintaining the integrity of the genome is crucial to any form of life (*3*). Before the cell divides, all three billion bases of each mammalian cell must be faithfully duplicated with a minimum of mistakes. Nonetheless, it is estimated that a mammalian genome undergoes about 100,000 modifications per day (*4, 5*). Besides errors occurring during the duplication process, various other sources can lead to alterations of the gene long after the replication has taken place. The endogenous mutagens originate within the cell, cause amongst others DNA hydrolysis, oxidation and methylation. They play a major role in mutagenesis, carcinogenesis, and aging (*6*). Other alterations arise from

exogenous factors, such as environmental toxins, ionizing radiation or chemotherapeutic agents (*7*) also produce lesions on the DNA.

The slightest modification in the DNA genome can lead to a wide variety of events. A base substitution in the gene leads to an alteration in the protein this gene is encoding, which might alter its activity, disable certain functions, or modulate its regulatory behavior. Mutations, sometimes a single alteration, often lead to catastrophic consequences and are the basis for a wide variety of diseases including cystic fibrosis, xeroderma pigmentosum, sickle cell anemia, or a variety of cancerous diseases. (*8*). The organism must therefore provide a powerful mechanism to faithfully replicate the genome in the first place, but equally important restore accurately any changes or modifications that were imposed on the DNA.

Additional Functionalities of Polymerases

DNA polymerases are key enzymes involved in both, replicating or repairing the genome and catalyze the polymerization reaction of deoxynucleotides in the 5' → 3' direction to extend an existing 3' hydroxyl group on the primer single strand. The proper Watson-Crick base is added to the primer according to the template strand in order to extend the primer, forming a complementary double stranded DNA. Some polymerases,

though, are template independent and extend a primer without a template. Depending on the number of nucleotides added by a polymerase within one association/dissociation cycle the polymerase is classified to be either processive or distributive. Processive polymerases are capable of adding several nucleotides to a growing DNA strand before dissociating from the template. Those involved in replication tend to be highly processive, often adding thousands of nucleotides in one cycle, while polymerases involved in repair typically have lower processivity. If the enzyme dissociates after each round of extension it is known to be distributive. A few DNA polymerases, including pol β, pol λ, poļ μ, terminal deoxynucleotidyl transferase (TdT), yeast pol IV and pol X are known to display distributive synthesis at least in the absence of the 5'-phosphate group (*9-11, 12, 22*).

In addition to their main functionality of extending a primer or inserting nucleotides into a gapped DNA, many polymerases are multi-enzymatic proteins. This means that they exhibit other activities, which are directed by various other structural modules within the molecule besides the polymerization domain. One commonly seen functionality in polymerases is the 3'-5' exonuclease activity (*13*) that allows the enzyme to excise an incorrect base and immediately replace it by the correct one. Because of its corrective nature, this activity is also known as proofreading. Other polymerases like pol

I contain an additional 5'-3' exonuclease to cleave a RNA primer that might be present immediately upstream from the DNA synthesis site and replace it with deoxynucleotides (*14*). A third example of such an added functionality is the 8 kDa 5′-deoxyribose 5′-phosphate-lyase (dRP-lyase) domain as seen in pol β (*15*). It enables the excision of a sugar-flap moiety prior to the insertion of a new nucleotide during the repair process.

DNA Polymerase Families

The number of known polymerases is continuing to grow and currently at least five polymerases are recognized in *Escherichia coli,* nine in *Saccharomyces cerevisiae*, and 16 in humans, as reported by Hübscher et al (*16*) or more recently by Bebenek et al (*17*). Based on sequence homology and structural similarities, polymerases have been grouped into seven families: A, B, C, D, X, Y, and reverse transcriptase (RT). All families but RT have members who are involved in the DNA repair process. So far no eukaryotic polymerases have been found in the families C and D, which are also the least characterized families. Examples for Family A members are the extensively studied bactriophage T7 and the eukaryotic mitochondrial pol γ (both replicative DNA polymerases) and the repair polymerase of *E. coli* pol I (Klenow fragment). Major replicative members belonging to Family B are pol α, δ, ε, and ζ (*18, 19*). They are

distinct by their remarkable accuracy during replication that is supported by a strong 3'-5' exonuclease activity. Probably the best known representative from the Family X is pol β (*20, 21*). Other members are pol σ, λ, μ, TdT, pol X (*5, 22*) or the yeast representative pol 4. Polymerases of this family are mainly involved in variety of repair processes. The Family Y polymerases differ from others by their low fidelity and ability for translesion DNA synthesis (*23*). Pol η, ι, κ, Rev1, and the translesion synthesis *E. coli* polymerases Pol IV and V (*24, 25*) are all part of this family.

Common Subdomains of Polymerases

Although a large variety of individual functions can be found in the known DNA polymerases and their sequence homology even within a family is often small, all determined structures so far indicate a common architecture. It has been compared to a hand, consisting of a "thumb", "palm" and "finger" domain (*26*). Figure 1-1 shows this concept on two examples of polymerase complexes, pol β (*27*), the smallest mammalian polymerase (39 kDa) known so far belonging to the Family X, and the bacteriophage T7, a Family A (*28*) polymerase. The subdomains have originally been defined by a structural alignment of the palm subdomain. A functional alignment of the catalytic participants has been used as well (*29*) and resulted in some confusing nomenclature for

the DNA polymerase pol β, where the functional equivalent of the thumb subdomain in pol β is structurally equivalent to the finger subdomain of other polymerases and vice versa (*15*). During this work, the original, more common nomenclature will be used, placing the fingers of pol β between the palm and the 8 kDa subdomain and the thumb at the C-terminal end.

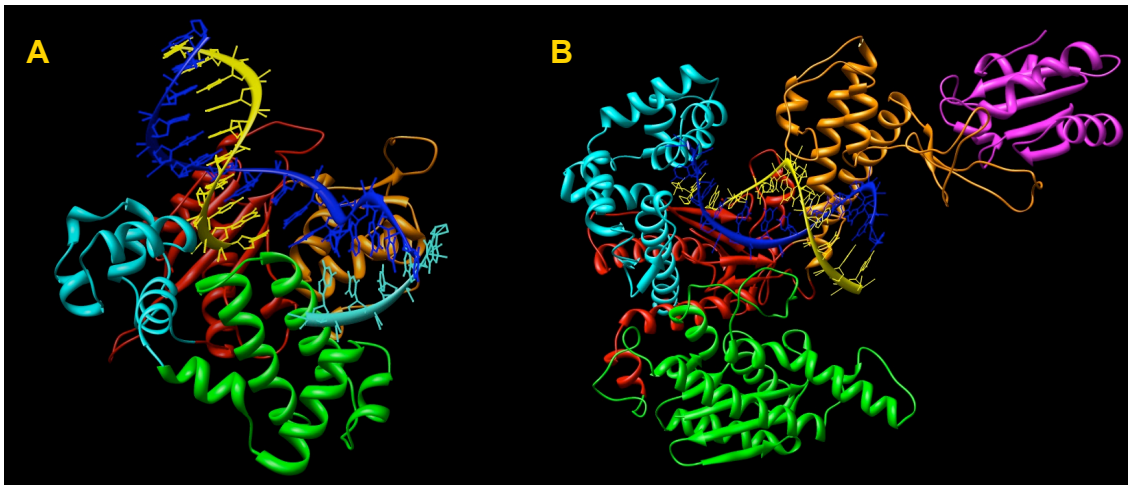


Figure 1-1: Crystal structures of two DNA polymerase complexes with duplex DNA based on the PDB coordinates pol β (PDB code 1BPY) Figure **A**, and pol T7 (PDB code 1T7P) Figure **B**. Common domains are colored as followed: palm domain in red, thumb in orange, finger in cyan, DNA template in blue, primer in yellow. The downstream DNA single strand in the pol β structure is light blue and the 8 kDa domain in green. For the pol T7, green indicates the 3'→5' exonuclease domain and magenta the thioredoxin protein complexed with the pol T7.

The Palm Subdomain

The palm domain represents the core of polymerases, where the nucleotidyl transfer takes place hence it is not surprising that it is highly conserved for all polymerases. The structural conservation is most pronounced in a distinct β-sheet where three highly conserved acidic residues are located, forming a carboxylate triad. This triad is essential for the catalysis of the nucleotidyl transfer reaction (*30-33*). The finger and thumb domains have much less conserved structural analogies. Still, certain elements stand out in each domain.

The Thumb Subdomain

The thumb subdomain (finger in pol β) exhibits some structural analogy by the presence of parallel or anti-parallel α-helices. They are generally involved in a number of important interactions across the minor groove of the DNA and show conserved interactions with the DNA backbone (*28, 34*).

The Finger Subdomain

Crystal structures of several polymerases (*35-37*) have shown an increased flexibility of the finger subdomain. Intense studies on this important feature on pol β

showed a distinct movement of this subdomain (thumb in pol β) between an open and closed state, providing crucial orienting interactions with the incoming nucleotide triphosphate (*38-40*). It remains to be seen whether this is a common mechanistic feature. Binary structures of pol β with DNA show an open thumb domain, similar to the one described for the apo form (*41*), but in the presence of a ddCTP, the crystal structure shows the thumb in the closed form (*39*), temporarily trapping the nucleotide in the complex (Figure 1-2, B and C). This movement and the observed similarities of the apo- and binary complex structures lead to the proposal that it might represent the

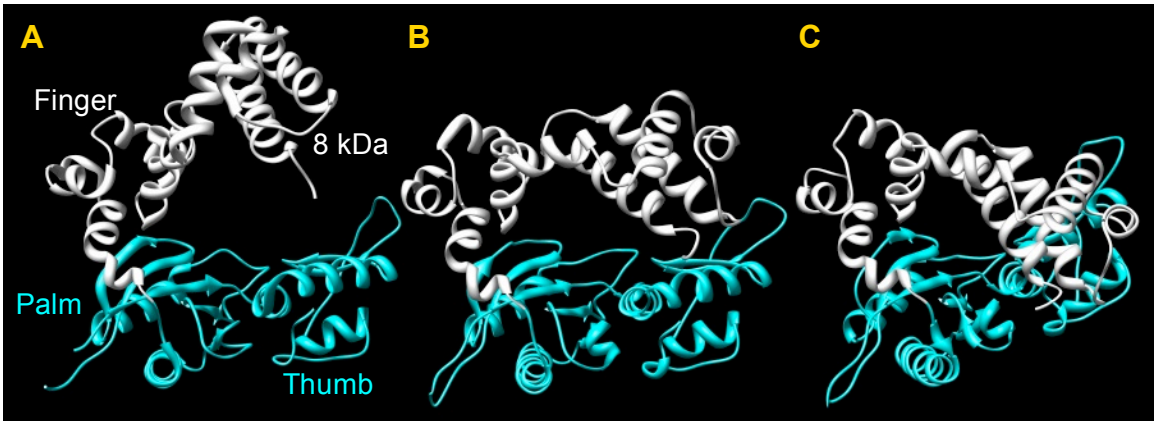


Figure 1-2: Three crystal structures indicating the domain movement of pol β. **A**: Free pol β (PDB code 1BPD). **B**: Complex of pol β with 7 bp complementary DNA, open thumb, but finger and 8 kDa domain swung down (PDB code 9ICJ). **C**: Ternary complex with 16-mer template, 10-mer primer and 5-mer downstream primer in the presence of ddCTP, Mg^{2+} and Na^{+} (PDB code 1BPY) indicating an additional thumb closure. Cyan indicates the palm and thumb domain.

anticipated conformational change (*42*) and possibly manifest the rate-limiting step that has been observed in pre-steady state kinetic experiments (*43*).

Several crystal structures for the Family Y polymerases have shown unusually small finger and thumb domains, leading to the hypothesis that this might enable them to more easily accommodate lesions and explain their low fidelity.

Other Subdomains

The structures of pol β and TdT as shown in Figure 1-1 point out another important property of some polymerases, the presence of additional functional domains colored in green. The pol T7 structure shows the 3'→5' exonuclease domain used for the proofreading of the DNA and an extended loop on the thumb domain providing the thioredoxin binding interface. Thioredoxin binds very tightly to the polymerase and therefore is co-crystallized in the structure. Its presence increases the processivity of DNA synthesis of this complex. The 8 kDa fragment on the pol β structure also provides additional functionality and is present in pol λ pol μ and TdT. It has been shown to bind strongly to the 5'-phosphate group of gapped DNA and possess the dRP-lyase activity (*15*).

Mechanism of Nucleotidyl Transfer

Studies of crystal structures from pol β have led to an explanation of the important catalytic mechanism, how the polymerization reaction takes place (*27, 29, 39, 44*). Structures of a number of other polymerases (*44*) have confirmed this mechanism to be general. The challenge was to trap a relevant intermediate without interfering with the reaction mechanism. Attempts were made to trap structures with dideoxy-terminated primers, resulting in a missing O3' which had to be modeled into the structure. Alternatively, Mg^{2+} (known to be essential for the polymerization reaction) was substituted with Ca^{2+} to inhibit the catalysis of the nucleotide insertion. Both attempts distort the geometry of the active site. An ingenious idea by Batra et al (*44, 45*) allowed the structure determination of a pre-catalytic complex of pol β in the presence of Mg^{2+} and the 3'-OH on the primer by using a nonhydorlyzable 2'-deoxyuridine 5'-(alpha,beta-imido)triphosphate (dUMPNPP), which served as a nucleotide analog. He showed that the substitution of the binding oxygen by a bridging nitrogen did not distort the geometry of the complex while maintaining all the coordination sites in place.

The catalytic site involves three aspartic acids that are conserved for all known polymerases and a two-metal ion mechanism was proposed by Steitz (*30-33*).

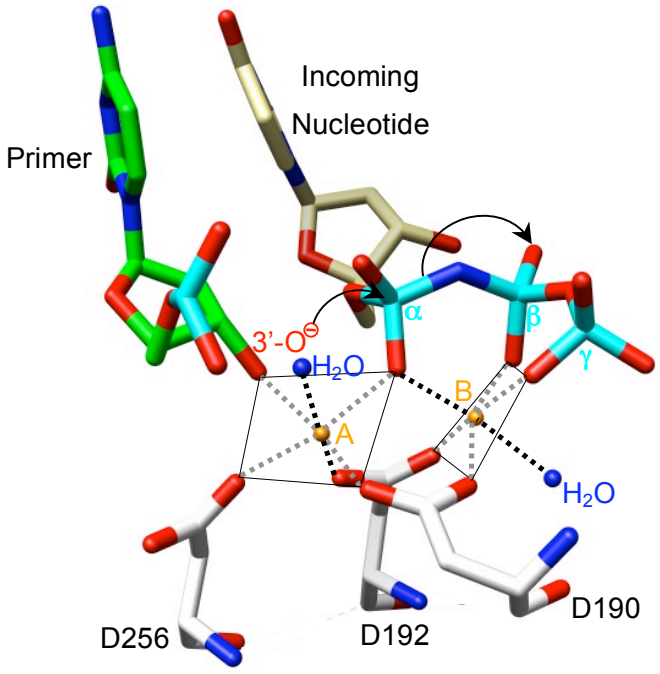


Figure 1-3: Active site of ternary complex in the pre-catalytic state of nucleotidyl transfer. This model is based on a recent crystal structure (*44*) (PDB code 2FMS), where the pre-catalytic state was trapped with dUMPNPP. Coordination of the catalytic and nucleotide metal binding sites (orange) to the active site aspartates D190, D192 and D256, two water molecules (blue), the 3'-hydroxyl group of the primer and the triphosphate of the incoming nucleotide (dUMPNPP) are indicated with dashed lines.

The proper geometry of the binding site is facilitated by the presence of the two metal ions. The catalytic metal (metal A) coordinates with all three aspartates (D190, D192 and D256), a water molecule, interacts with an α-phosphate-oxygen and the primer 3'-OH. The nucleotide-binding metal (metal B) is coordinated to two aspartates (D190 and D192), a water molecule and the oxygen atoms of all three phosphates of the

triphosphate group. A detailed description of the proposed mechanism was described by Lin at al (*45*) and entails the following steps. The pKa of the 3'-OH was proposed to be lowered in the presence of metal A, promoting the start of the catalysis reaction with a proton transfer from the 3'-OH to D256. The resulting nucleophile attacks the α-phosphate followed by an α-P-oxygen bond breakage. This leads to the release of the pyrophosphate with its associated metal B and a successive return of the active site to the pre-catalytic configuration, ready for another round of catalysis. The metal ions have not been observed in the binary complex with DNA and appear only to be present with the incoming nucleotide. Binding of the correct nucleotide with metal B was found to induce the thumb closure for pol β, but was not observed for pol λ indicating, that it might not be required for the catalytic step and could serve other functions.

The African Swine Fever Virus Polymerase X

The African Swine Fever Virus contains the smallest naturally occurring DNA polymerase found so far. With 174 amino acid residues and a molecular weight of slightly over 20 kDa the ASFV pol X is about half the size of the next smallest DNA polymerase, pol β. Therefore, it is not surprising that it became the first polymerase with a solution structure determined by NMR instead of crystallographic methods (*46, 47*).

Brief Background

The ASFV originated in the Sub-Sahara areas of Africa and is a highly contagious disease of domestic pigs and wild boars. The virus causes a lethal hemorrhagic disease in the pigs with a mortality rate of 100%. Less virulent forms have been described outside Africa (*48*). ASFV also infects African wild pigs, including warthogs or bushpigs, as well as soft ticks that parasite warthogs maintaining an endemic cycle, although no obvious disease has been detected in these hosts or other species. Active virus can be recovered years after infection (*49*) and might indicate that the ASFV is able to produce a persistent infection. The virus was first reported outside of Africa 1957 in Lisbon, Portugal and spread throughout the Iberian peninsula, into France, Belgium, and other European and Mediterranean countries (*50*). Through a strictly enforced slaughter policy, both Spain and Portugal managed to eradicate the disease in the late 1990s. The ASFV crossed the Atlantic in the 1970s and outbreaks were reported on some Caribbean Islands. Any outbreak of the ASFV disease has to be reported to the World Organization for Animal Health (OIE), which maintains a list of major occurrences. The Republic of Georgia reported one of the most recent outbreaks in June 2007, the first ever seen in this part of Europe. Approximately 14,000 hogs died

or had to be destroyed during this incident showing the permanent threat for hog farming. Unfortunately there is no vaccine available so far.

ASF Virus and ASFV Pol X

ASFV is an icosahedral, cytoplasmic deoxyvirus and has recently been classified as the sole member of the Family *Asfarviridae (51)*. It has been speculated to be the missing link between the large cytoplasmic DNA iridovirus and poxvirus groups (*51*). Virus particles are about 200 nm in diameter and contain a central nucleoid of about 80 nm surrounded by an inner lipid membrane (*52*). The protein capsid consist of a double membrane complex with its origin at the endoplasmic reticulum (*53*) and an outer envelope derived from the plasma membrane (*54*). Assembly of the virus takes place in discrete perinuclear areas, also termed virus factories (*55*). The protein p73 (*56*) represents 35% of the viral protein mass and another 25% is produced by proteolytic cleavage of the p220 polyprotein (*52*). A total of at least 34 structural proteins have been identified (*57*).

The ASFV genome consists of a 170-190 kbp double stranded DNA, dependent on the virus strain, and a conserved central region of 125 kbp, encoding for 151 polypeptides. Yáñez et al (*56*) determined the genome sequence of the avirulent BA71V

strain and identified a number of enzymes involved in the gene transcription, protein modification and DNA replication pathways. Of special interest to this work were two sequences that showed a high similarity to two types of DNA polymerases: one was attributed to a viral replication DNA polymerase with analogy to the Family B polymerase (*58*) and the other one showed similarity to the Family X DNA polymerase, designated as ASFV pol X.

Further sequence alignments of pol X (*56*) and biochemical analysis (*9*) pointed to a close resemblance of pol X with the pol β enzyme. It was determined that the molecule was an analog to the C-terminal half of pol β, including the catalytic nucleotide binding site in the palm and the thumb subdomain. Therefore similar functionality with regard to these subdomains was anticipated. The finger and 8 kDa subdomains which are missing in pol X, have been described (*40*) to contain an important helix-hairpin-helix motif in pol β, including a metal ion, and interact with the backbone phosphates of the DNA strand. In addition, the fingers are mainly interacting with the primer while the 8 kDa unit does so with the downstream oligonucleotide, positioning a gapped DNA on the enzyme. Therefore it can be anticipated that these missing subdomains have an impact on the interaction between pol X and DNA.

Regardless, ASFV pol X was shown (*9*) to have DNA-directed polymerase activity with considerable fidelity, most efficiently catalyzing single-nucleotide gaps. The action appears to be highly distributive during DNA primer extension, but some processivity was observed when a gapped DNA was used, especially in the presence of a 5'-phosphate group. Lamarche et al (*59*) also described an AP lyase activity, similar to the one described for pol β (*60*). Conversely no 3'→5' exouclease activity could be found. The analogies of pol X to pol β led to the hypothesis that pol X participates in a base excision repair (BER) like pathway as well (*9*). The ASFV genome encodes for additional enzyme homologues required for the BER repair mechanism, supporting this hypothesis. They include a DNA glycosylase, AP endonuclease, phosphodiesterase, and DNA ligase (*4*). With its small size and presumably simplified functions, pol X was expected to serve as a good model to study and understand the structure function relationship for a template directed nucleotide insertion.

Activity and Fidelity of Pol X

According to the biochemical characterization reported by Oliveros et al (*9*), pol X performs template directed nucleotide insertions and is able to select the correct one among the four nucleotides, which is then faithfully inserted. The experiment was carried

out using four template-primer oligonucleotides that differed only in the first template base, containing either one of the four bases. Nucleotide insertion for the 16 possible combinations was evaluated by adding all four nucleotides to each template-primer. A clean insertion of only the correct nucleotide was observed in all cases, even in the presence of a 20-fold higher concentration for the non-complementary nucleotides, showing high fidelity of the polymerase. The single nucleotide gap was tested with three nucleotides and showed insertion of the complementary dATP and ddATP, without discrimination towards the ddATP and a very inefficient insertion of dGTP. It should be noted that the incubation time for these experiments was only about 15 minutes.

Tsai and coworkers (*61*) reported a different result. Quantitative measurements of several kinetic parameters showed that the catalytic efficiency of pol X was only about $1/5000^{th}$ that of pol β for the correct base-pair incorporation and un-proportionally high for the formation of the G:G mismatch formation. Furthermore the fidelity of pol X was determined to be extremely low for all nucleotide combinations, making it questionable whether this enzyme actually could be classified as a repair enzyme. By far the lowest fidelity was seen for the incorporation of dGTP opposite G, forming the incorrect base-pair G:G at a ratio of 1 in 1.9. This is the lowest nucleotide incorporation specificity observed so far for a template directed nucleotide polymerase and surpasses the error

prone pol η by a factor of at least ten (*62*). The insertion of dGTP shows generally the lowest fidelity, defined as $[(k_{pol}/K_{d,app})_{cor}+(k_{pol}/K_{d,app})_{inc}]/(k_{pol}/K_{d,app})_{inc}$, where the subscripts "cor" and "inc" refer to the correct and incorrect incorporation, respectively. Ratios for the so determined fidelity of G opposite G, T and A were reported as 1.9, 25, and 30, respectively, a low fidelity that would be incompatible with a repair function of this polymerase but rather point to a mutase activity of pol X.

Several attempts to explain these observations are reported in the literature (*59, 61, 63, 64*) but a conclusive answer is still lacking. One suggestion was made that pol X might have evolved into an enzyme specifically targeting the incorporation of dGTP. The reasoning was that the ASFV host cell, the swine macrophage, produces and releases reactive oxygen species introducing a range of damage, in particular DNA lesions resulting from the oxidation of dGMP to 8-oxo-7,8-dihydro-2'-dGMP (8-oxo-G) (*4*). The only glycosylase homologue found in the ASFV genome is known to specifically excise 8-oxo-G lesions, making this specific functionality appealing. After the excision, the repair would involve the insertion of dGTP opposite the template C, which happens to have the highest efficiency for insertion. While this might explain the preference for dGTP binding, it does not provide an explanation for the overall extremely low fidelity. A possible explanation for that includes the notion that pol X is involved in mutagenesis

aimed at providing variability to the virus. Other members of the Family X polymerases, like TdT or pol μ, have been mentioned to exhibit some mutase like functions (*46*). Similarly, reports about some variation in four proteins encoded by the ASFV genome and the observation of some point mutations at different positions along the genome (*65, 66*), plus the fact that a less virulent strain has been observed outside of Africa, might support this hypothesis. On the other hand, the number of reported mutations is small which would more likely point to naturally occurring events, unless it is the result of pol X that specifically and predominantly acts on single-nucleotide gaps containing undamaged template nucleotides. Another interesting possibility is differential behavior of pol X in vivo compared to the in-vitro situation (*61*). In vivo, pol X might associate with other molecules that in turn could enhance the fidelity and/or catalytic efficiency. The small size and therefore limited functionalities of pol X makes such a model appealing, where some of the functionality would be provided by other molecules, but so far no precedence for this is known.

A quantitative examination of the fidelity and catalytic efficiency by the group of Blanco and Salas (*63*) lead to a drastically different result, confirming the initial report mentioned above. Their conclusion was that the misinsertion frequency for a one base gap oligonucleotide with the 5'-phosphate on the downstream primer was in the range of

those observed for pol β, hence ASFV pol X would have similar functionality as pol β and participate in the viral BER process to maintain viral genomic integrity during infection. The fidelity values reported by Salas et al (*63*) were between 40 and 700 times higher then the ones reported by Tsai and coworkers (*61*). The general trend for the individual base-pairs was maintained, meaning that the fidelity for a G:G formation was the lowest at 1 in 1,400 (Tsai (*61*): 1.9) and C:C had the highest fidelity at 1 in 300,000 (Tsai (*61*): 7,700). Salas (*63*) did also not see a distinct enhancement of insertion efficiency for dGTP opposite G but reported a 1,000-fold incorporation advantage for the correct dCTP base. Differences in experimental conditions such as pH, buffer composition, and DNA sequences were re-examined by both groups (*59, 63*) but ruled out as a reason for the discrepancy. This leaves a very basic question about the functionality of the polymerase unanswered and further work is needed to find the proper explanation for this discrepancy.

Structure Elements of ASFV Polymerase

Two NMR structures of the free pol X were determined by Mullen and coworkers (*46)* and Tsai and coworkers (*46, 47*) Both structures resemble very closely the palm

and thumb domain of pol β (*27*). Figure 1-4 shows the assignment of the secondary structure elements for pol X as they will be referred to in this work.

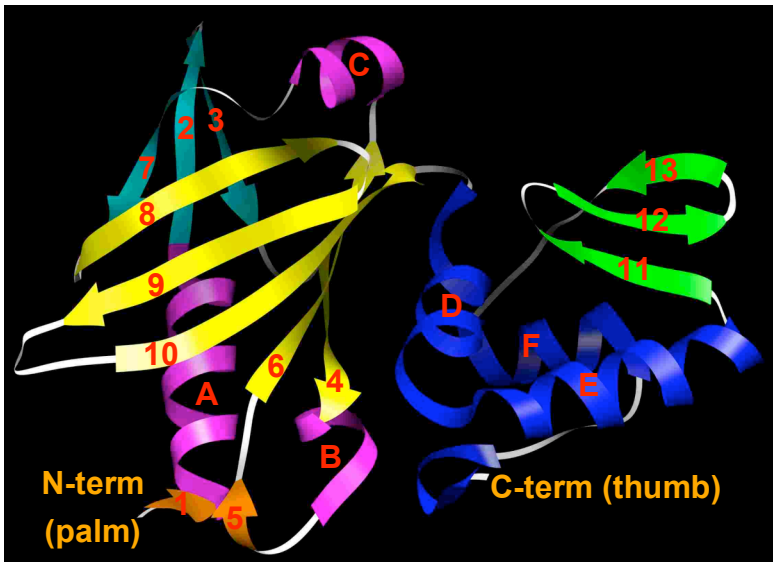


Figure 1-4: Secondary structure elements on solution structure of free ASFV pol X (PDB code 1JAJ). Palm subdomain with β strands 1-10 and α-helices A-C, C-terminal or thumb domain consisting of β-strands 11-13 and α-helices D-F.

Secondary Structure Comparisons of Pol X with other Family X Polymerases

While the majority of the pol X structure agrees very well with the one of pol β some notable differences exist. The biggest difference is the presence of a β-sheet, consisting of strands β2, β3 and β7 in pol X, that is absent in pol β, where the same region contains an α–helix L instead. Moreover, αC is a new helix in pol X, replacing an unstructured loop in the pol β structure. Three other loop regions in pol β are not present

in the pol X structure. One of them connects β5 with β6 (β8, β9 analog in pol X). The strand β5 is shorter in pol β followed by a long loop connecting to strand β6 after a turn. β9 and β10 are connected by a short four residue turn in pol X but contain a 17 residue long, unstructured loop in pol β. The third loop is located in the thumb subdomain, where the three stranded β sheet (β11, β12, β13) seen in pol X is reduced to the first two strands in pol β and a long loop makes the connection to the following α-helix αO. The remainder of the two structures overlay well. Structure comparisons with other Family X DNA polymerases revealed a similar picture for the palm and thumb subdomains, as was described for pol β. The results are listed in Table 1-1.

Three features can be identified, where pol X is consistently different. Pol X is the only enzyme of this group that forms a β-sheet (β2, β3, β7) instead of a long α-helix (αL or α12). Unique for pol X is also the presence of α-helix αC. The other structures show an unstructured loop in this location, with the exception of TdT, where αC superimposes with the N-terminal end of a long helix α12. The β-sheet (β4, β6, β8-10) is very characteristic for all five structures. While the locations for the three aspartates are well preserved, strand β5 is substantially shorter in pol β and completely missing, e.g. replaced by an unstructured connector in pol μ.

Table 1-1: Main structural elements observed for some Family X polymerases. The structures listed were obtained from the Protein Data Bank. Pol X served as the reference structure and corresponding regions for each of the polymerases were compared to the features in that region of pol X.

	Pol X	Pol β	Pol λ	Pol μ	TdT
PDB code	1JAJ	1BPY	1RZT	2IHM	1JMS
catalytic aspartate	superimpose	superimpose	superimpose	superimpose loosely	superimpose
analog of pol X β2, β3, β7-sheet	β2, β3, β7	αL	αL	αL	α12
analog of pol X α-helix C	αC	loop	loop	loop	superimposes with N-terminal of α12
analog of pol X β8	β8	short β5	similar to pol X	β8 strand missing	similar to pol X
analog of pol X β8 loop to β9	short loop connecting β8 and β9	long loop from β5 to β6	similar to pol X	long loop from αL to β8	18 aa loop (facing C-terminal)
analog of pol X β9 loop to β10	4 aa turn	12 aa loop	9 aa loop	15 aa loop	15 aa loop
β7-sheet in C-terminal	β11, β12, β13-sheet	short β8, β9, long loop	short β6, β7, long loop	Short β6, β7, long loop	short β6, β7, long loop

Two long loops, one replacing the turn between β9 and β10, the other one replacing β13, are present in all structures besides pol X. A third long loop is added in the TdT structure and reaches across the domain interface from the β sheet over to the thumb domain. These flexible loops do not seem to alter the overall structure of the palm and thumb domain, but could have a functional impact.

Residue Comparison of Pol X with Pol β

Some residues that have been found to be functionally important in pol β are conserved in pol X and are pointed out here. The first similarity is the conservation of a cis-peptide bond causing a sharp kink between the α-helices αD and αE. A similar feature is found in pol β between αM and αN. Both structures also agree in the position of the highly conserved active site aspartate residues, D49, D51 and D100, which are almost super-imposable and indicate a general structural conservation of the active site. Besides the three aspartate residues, others i.e. R283, Y271, D276, and R258 have been shown to play a crucial role in the polymerase fidelity, and catalytic nucleotide binding efficiency (*67, 22*). The four residues are positioned on the pol X equivalent helix αD and αE and β10. Only R283 is conserved in pol X by R127, while the other residues differ, H115 (Y271), V120 (D276) and F102 (R258), likely to affect the functionality. Vande Berg et al (*68*) report a D276V mutation on pol β and see increased binding of the mutated enzyme with the nucleotide. Y271 and F272 (pol X: H115 and F116) form the YF motif and interact with the minor groove of the DNA. Other interactions between DNA and pol β have been described for K230, T233, K234 for the palm and T292, Y296, R283 with the thumb (*39*). They correspond to the pol X residues K72, S75, F76 and K136, Y140, R127, respectively, and are mostly conserved. Charged residues, not

present in pol β, were proposed (*46*) to interact with the DNA in pol X. They are K131-133 on αE and K59, K60 and K63 on αC. An interesting difference results from cysteines C81 and C86, where they are reported to form a disulfide bond in one structure of pol X, while remaining reduced in the other. A more detailed discussion on this topic and experimental data will be provided in Chapter V.

Overall, pol X shows the most compact structure of all Family X polymerases described here. It can be envisioned that the basic nucleotidyl transfer reaction is the same because it takes place in these highly conserved subdomains. The missing flexible loops might alter the functionality of pol X, but this has not been reported yet. The absence of the entire finger and the 8 kDa domain in pol X has eliminated the proofreading and dRP lyase capability (*9*) and an important location that is responsible for the DNA binding, as well as proper positioning of gapped DNA or nicked substrates in pol β (*22, 27, 39, 69*). However, pol X appears to bind DNA with a similar affinity then pol β (*46, 47*) and therefore must provide an alternative way of accomplishing this. The enhanced positive charges on αC and αE in pol X might provide a hint towards how pol X accomplishes the interaction with DNA. Several studies, including this one, aim to provide an explanation and more specific insight into the DNA - pol X binding site (*70-72*).

CHAPTER II

EXPRESSION, PURIFICATION AND SOLUBILIZATION OF THREE ASFV POL X CONSTRUCTS

Introduction

The open reading frame (ORF) pO174L of ASFV pol X was first described by Yáñez et al (*56*), as he was completing the nucleic acid sequence analysis of the African Swine Fever Virus. Besides the gene for pol X, it is noteworthy that several other enzymes of the DNA repair system were found to be encoded in the viral genome, specifically those participating in the base excision repair. The role of these proteins in the life cycle of the ASFV remains uncertain but ASFV is known to mainly replicate in macrophages and monocytes, where an oxidative environment prevails. The reactive oxygen species generated in that environment are required for the microbicidal functions but also cause oxidative DNA damage. Therefore, it has been hypothesized that a repair system for these small DNA lesions is desirable and the BER process would be well suited. The AP endonuclease sequence homolog is encoded by the pE296R gene and pNP419L represents a putative DNA ligase. For a complete BER system, an initiating glycosylase would be needed to excise the base containing the lesion and produce the

abasic site. So far no gene that would encode for a general glycosylase has been identified in the ASFV genome. Alternatively, spontaneous depurinations of DNA generate lesions with an abasic site as well and might be a source for this step here. The protein sequence of p0174L was shown to be most similar to the C-terminal regions of DNA poly β, TdT (*73*) and the yeast ORF designated YCR14C (*11*) according to the amino acid sequence.

Material and Methods

ASFV Pol X Constructs

During this work, three different constructs of the ASFV pol X were used. The original construct was kindly provided by Luis Blanco (Universidad Autónoma, Madrid, Spain) and contained an N-terminal leader peptide of 35 amino acids, including a 6x His tag and an enterokinase cleavage site. The second construct contained a C-terminal His tag and was kindly prepared by Irene Zegar, who was a member of this laboratory at that time. The third construct was kindly given by the late Gregory Mullen (University of Connecticut Health Center, Farmington, CT). These constructs will be referred to as pol X_{orig}, pol X_{iz}, and pol X_{gpm}, respectively.

Sequencing

For all three constructs, the plasmid was extracted at least once following an overnight culture growth of 250 mL and was stored at -80 ºC for future use. A Qiagen Plasmid Maxi Kit (Qiagen, Valencia, CA) was used for the extraction and the procedure was performed according to the manufacturer instructions following the "Midi and Maxi Protocol". Upon plasmid extraction, the DNA sequence was determined to ensure a valid plasmid. In-house sequencing was done by the Vanderbilt University Sequencing Core and the data was analyzed with MacVector™ version 5.0.2 (Oxford Molecular Group, Campbell, CA). To assure the proper protein expression on the pol X_{orig} construct, purified proteins before and after cleavage were sequenced by the Edman degradation method, using the in-house protein sequencing facility. The first six amino acids were determined on a Procise Sequencer (Applied Biosystems, Foster City, CA).

PAGE

For each batch of protein produced, an SDS-PAGE was run to ensure the quality and purity of the protein. Initially gels were run on Laemmli System SE260 (Hoefer, San Francisco, CA) and gels were prepared according to the manufacturer specifications,

16% acrylamide for separation and a 4% acrylamide stacking gel. Later, pre-cast NuPage 4-12% Bis-Tris gels (Invitrogen, Carlsbad, CA) were used.

Stock Cultures

Stock cultures of a freshly grown, non-induced pol X cultures, harvested at OD_{600} of ~0.8 were prepared for all constructs in multiple copies and stored at -80 ºC. These samples were prepared as followed: 0.5 mL of the freshly grown culture were combined with 0.95 mL glycerol and 0.15 mL of a mixture containing 0.5 M $MgSO_4$ and 0.2 M Tris•HCl at pH 8.0.

ASFV Pol X_{orig} Construct

Protein Expression in LB Medium

The open reading frame containing the DNA polymerase X gene O174L from ASFV was inserted by the group of Luis Blanco into a pRSET-A bacterial expression vector (*74*) at the bamHI/PstI site, which encoded the recombinant protein as fusion with a multifunctional leader peptide of 35 amino acids. The N-terminal leader peptide also contained a hexahistidyl sequence for purification with an affinity resin, a tyrosine residue for radioiodination, an enterokinase proteolytic cleavage site to remove the

leader peptide and an ampicillin resistance gene. The bacterial host was *E. coli* strain BL21(DE3) pLysS, which contained a chromosomal copy of the T7 DNA polymerase gene. Recombinant protein expression was under the control of the lacUV5 promoter, which was inducible by isopropyl β-D-thiogalactopyranoside (IPTG), a lactose analog. Initially, the lacUV promoter that was responsible for the expression of the T7-polymerase was repressed; hence no T7 polymerase was produced. Upon addition of IPTG to the culture medium, the lac repressor was displaced causing an expression of T7 polymerase, which in turn recognizes the T7 promoter on the gene containing the ORF of the protein of interest and therefore started the production of the desired protein as long as T7 polymerase is available. The expressed protein was 199 amino acids long unless the leader peptide was cleaved by enterokinase, which would leave only five additional residues.

An initial overnight expression was set up in 50 mL Luria-Bertani (LB) broth culture medium (Sigma), containing 50 µg/mL ampicillin at 37 °C, and an aliquot of 10 mL of this culture were added to 800 mL sterile LB medium, containing the antibiotic in the same concentration. The cultures were grown at 37 ºC under constant shaking until they reached an OD_{600} of 0.8. At this point expression was induced with 1 M IPTG to a final concentration of 0.4 mM. After 4 h of inoculation, the cells were collected by

centrifugation at 7,300 g for 20 min at 4 ºC (Beckman, JLA-9.100 rotor) and frozen at -80 ºC until purification.

Protein Expression in Minimal Medium M9

Protein expression in LB-medium yields large amounts of proteins. For NMR studies, proteins > 12 kDa require the enrichment of NMR active, isotopic labels such as ^{15}N and/or ^{13}C, since the natural abundance of ^{13}C and ^{15}N at 1% and 0.1%, respectively, is too low for those measurements. To obtain uniformly labeled protein with these isotopes, expression in a minimal medium is commonly used (*75*). Preparation of 1 liter minimal medium (M9) was as follows: 22.6 g Na_2HPO_4 $7H_2O$, 6.0 g KH_2 PO_4, 1.0 g NaCl, and 8 mg $CaCl_2$ $2H_2O$ were dissolved in deionized water, adjusted to pH 8.0 and sterilized. Sterile filtered solutions for the following supplements were added: 2 mL of 1 M $MgSO_4$, 1 mL ampicillin (50 mg/mL), 5 mL $^{15}NH_4Cl$ 20% w/v (>98% ^{15}N, Cambridge Isotope Laboratories) and 4 mL $^{13}C_6$-D-glucose 20% w/v (>98% ^{13}C, Cambridge Isotope Laboratories). For the expression of only ^{15}N labeled protein, $^{12}C_6$-D-glucose was used instead of the $^{13}C_6$-D-glucose.

Overnight cultures were initially grown in 50 mL of the same M9 medium as used for the large batch. The procedure was analogous to the one described for LB medium

but the cells grew much more slowly in this nutrition deprived medium. The cells were induced at an OD_{600} between 0.7 and 1.0 and typically did not surpass a final OD of 1.2 after 4 h. Several modifications to this initial scheme improved the cell growth significantly. First, overnight cultures were grown in the LB medium as described above to a high density (*76*). The cultures were spun at 500 g for 15 min, the supernatant removed and the tube washed with M9 medium before the pellets were dissolved in additional M9 medium and added to the large medium. The second modification involved the addition of either yeast extracts or a vitamin cocktail that were described to reduce the lag phase in the growth curve (*77*) (see also Figure 2-1, curve C-E). A trace metal mixture was also added as described in Weber et al (*78*). Thirdly, equilibrating the main culture medium to 37 ºC before adding the overnight culture also increased the rate of the initial growth phase. Figure 2-1 shows the effect of these individual modifications.

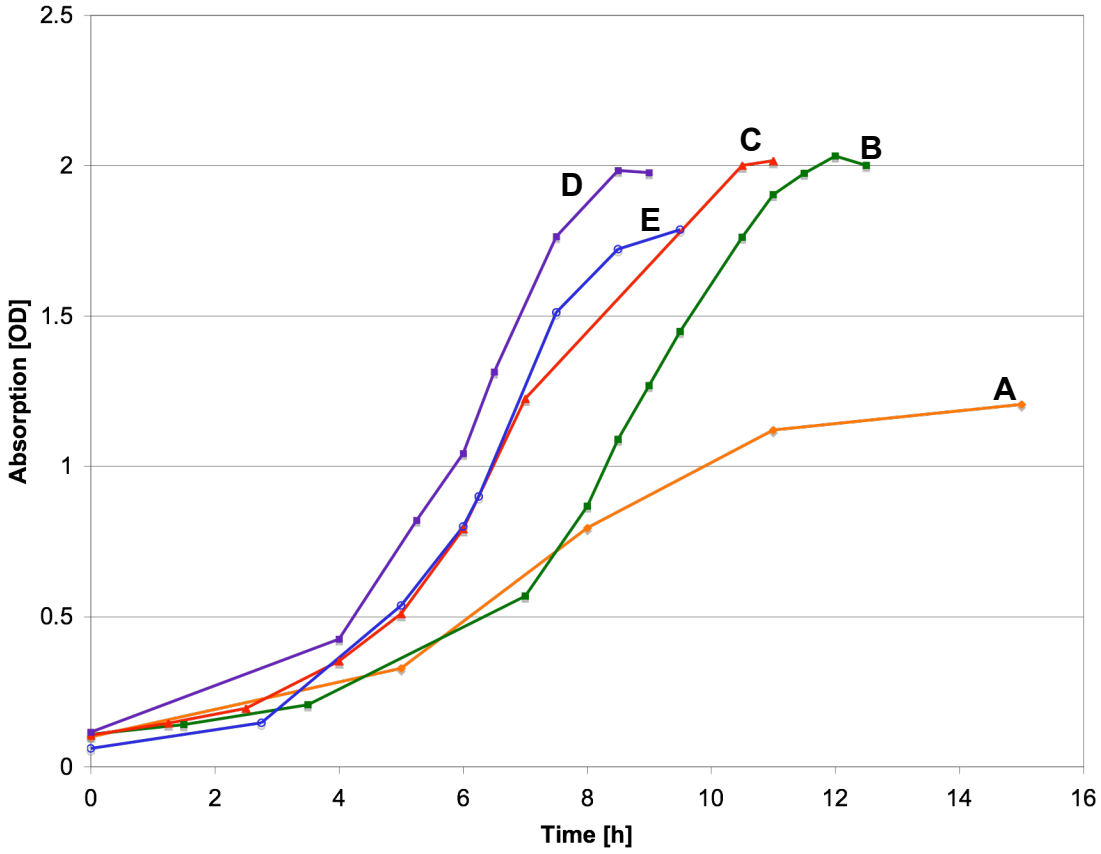


Figure 2-1: Cell expression curves in minimal medium cultures under various conditions; **a.** plain minimal medium, **b.** addition of trace metals to medium, **c.** trace minerals and yeast added, **d.** trace minerals and yeast added, main culture temperature equilibrated to 37 ºC prior to addition of overnight culture, **e.** trace minerals and vitamin solution added, main culture temperature pre-equilibrated to at 37 ^{0}C.

Purification of Pol X_{orig} Construct

The initial purification of this protein was done under denaturing conditions. The temperature was maintained at 4 ºC for all processes and buffers were kept at the same temperature. Pellets were collected and combined in a Falcon tube and 35 mL lysis

buffer was added (100 mM NaH_2PO_4, 10 mM Tris•HCl and 6 M guanidine HCl, adjusted to pH 8.0). Lysis of the cells was accomplished by 3 to 5 30 s sonification bursts (Branson Sonifier 450, Danbury, CT) until no solid cell material was visible anymore, keeping the sample on ice. Cell lysate was spun at 18,000 g for 20 min (Beckman, JA-12 rotor) and the supernatant was collected. 10 mL Ni-NTA resin, equilibrated with the lysis buffer, was added to the soluble fraction and nutated for 45 min. The protein-resin solution was then packed into a column. Washing and elution steps were performed using a buffer of 100 mM NaH_2PO_4, 10 mM Tris•HCl, 10 mM imidazole, and 6 M guanidine HCl and 10 to 20 column volumes were used in each step. A pH-dependent elution scheme was employed. The initial wash at pH 8.0 removed all the non-binding proteins. Additional wash cycles at pH 6.3 and 5.9 were performed to remove endogenous proteins containing histidine residues. Elution of the target protein occured in the same buffer adjusted to pH 4.5, where the histidine residues of the 6x His tag were protonated and therefore no longer bound to the Ni-NTA. The eluent was monitored with an A_{280nm} ISCO UA-5 absorbance detector and a type 6 flow cell unit (Teledyne ISCO, INC, Lincoln, NE). Fractions of 1 mL were collected.

Re-folding of the protein occurred during dialysis. The fractions obtained in the previous separation were combined in a 15 mL Slide-A-Lyzer, MWCO 10 kDa (Pierce,

Rockford, IL). Several dialysis conditions were utilized to minimize the protein precipitation. A buffer containing 50 mM sodium phosphate, 150 mM NaCl at pH 7.2 was tried first since it was a commonly used NMR buffer. This buffer condition caused precipitation during dialysis. The sample was spun and the supernatant was diluted and washed 3 to 5 times with 5 mM Tris•HCl buffer, pH 7.5 and concentrated to a final volume of 0.2 mL. The 1D ^{1}H as well as the ^{15}N-HSQC NMR spectra indicated a well-folded protein in a single conformation.

Trying to minimize precipitation, several other dialysis conditions were tested in successive batches. Dialysis in 5 mM Tris•HCl and 5 mM $MgCl_2$ at pH 7.0 also resulted in precipitation, which could be partially dissolved at pH 5.5 and was completely solubilized at pH 4.0. The pH was raised to 6.0 after adding 50 mM NaCl and no further precipitation was seen. About 19 mg of protein was obtained from a 3-liter minimal growth medium.

Since the goal was to cleave the 35-residue leader peptide from this construct, the next buffer system was equivalent to the cleavage buffer and consisted of 20 mM Tris•HCl, 50 mM NaCl, and 2 mM $CaCl_2$ at pH 7.4. These conditions were similar to the previous ones used, also caused a small amount of precipitation and yielded 18.3 mg of soluble protein. Dialyses on subsequent samples used for cleavage of the leader peptide

were carried out in this buffer, but the pH during the first dialysis step was lowered to 5.5 before it was raised to its final value of 6.5. Cleaving the leader peptide with recombinant enterokinase was possible for microgram samples, but milligram amounts, as needed for NMR measurements, did not yield sufficient product due to the solubility of the cleaved protein. More details can be found in Appendix A.

At a later point pol X_{orig} was purified under native conditions, using standard phosphate-based buffers as described for the pol X_{iz} purification for the Ni-NTA column and 500 mM imidazole to elute the protein. An additional S-100 size exclusion column run with 20 mM Tris•HCl, 50 mM NaCl, and 2 mM $CaCl_2$ at pH 6.5, yielded pure protein in high yield.

All these different conditions indicated that it was more important to maintain a low pH to keep the protein in solution and that the composition of the buffer salts had only a secondary impact on the solubility. Low salt concentrations were tolerated well by this protein, although increasing the salt concentration was shown to promote the solubility.

ASFV Pol X_{iz} Construct

Expression of Pol X_{iz} Construct

The open reading frame gene O174L representing the ASFV polymerase X was cloned into a bacterial expression vector pET-29b(+) from Novagen (Novagen, Madison, WI). This prokaryotic expression vector includes a *lac* operator and pol T7 promoter, meaning that over-expression was induced by IPTG as described above for the p-RSET-A vector. Differences to the previous vector included a kanamycin resistance gene and the hexahistidine tag on the C-terminal. Insertion of the O174L gene occurred between the Nde I and Xho I cloning site, leaving an additional glutamic acid between the pol X sequence and the six histidines at the C-terminal end. The plasmid was transformed into an *E. coli* strain BL21(DE3) pLysS. Stock solutions and plasmid extracts were obtained as described before. Expressions in both LB and minimal medium were performed in the same way as described for the pol X_{orig} protein, with the exception of using 20 µg/mL kanamycin and 25 µg/mL chloramphenicol as antibiotics, utilizing the optimized expression that included trace minerals and yeast or vitamin mixtures in the M9 media. Over-expression of the pol X_{iz} was indicated in the SDS gel, Figure 2-2 (B). Pre- and post-induction samples were loaded on lines 1 and 2. It appeared that line 2 showed a stronger band at 21 kDa but since this band coincided with a strong band in the pre-

induced gel the comparison based only on these two samples was inconclusive. The same band was also present in the sample from the first wash cycle of the Ni-NTA column as shown in lane 3, indicating that another protein with similar molecular weight was expressed in *E. coli*, but did not have a high affinity to the Ni-column. The purified protein in line 4 showed the presence of overlapping proteins in the pre- and post-induced samples.

Purification of Pol X_{iz} Construct

The purification of this protein was done under native conditions instead of the denaturing condition used for pol X_{orig}. Pellets were taken up in 40 mL lysis buffer containing 50 mM NaH_2PO_4, 300 mM NaCl, 5 mM imidazole, 0.5% of the buffer 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES-Na) and 0.4 mM serine protease inhibitor phenylmethylsulfonylfluoride (PMSF) at pH 8.0. The cells were lysed with 3 to 5 x 30 s sonication bursts, spun for 20 min in a fixed angle JA-12 rotor (Beckman) at 18,000 g. An aqueous Ni-NTA resin solution was washed, equilibrated and 10 mL of the final 50% Ni-NTA solution were added to the supernatant, which then was nutated for 45 min. The protein-resin complex was packed into a column and non-bound proteins were removed with at least 10 column volumes of wash buffer containing 50 mM NaH_2PO_4,

300 mM NaCl, and 5 mM imidazole (pH 7.5). The elution was monitored at 280 nm and a typical elution profile is shown in Figure 2-2 (A). Additional wash cycles and elution of the target protein were achieved by either modulating the pH or increasing the concentration of imidazole.

Buffers used for the pH dependent elution scheme were based on 50 mM sodium phosphate, 300 mM NaCl, and 20 mM imidazole. Analogous to the denaturing pH dependent purification, the pH of the buffer was lowered to pH 6.5 after the initial wash cycle and elution occurred at pH 4.5. The same buffer was used for the imidazole based method, but instead of modifying the pH, an extended wash in the presence of 20 mM imidazole was performed prior to elution of the target protein with 500 mM imidazole that was removed by successive dialysis or gel filtration.

Purified protein was obtained by both methods within 1-2 column volumes. Imidazole has an UV absorbance at 280 nm adding a substantial baseline shift at the point of peak elution as indicated in Figure 2-2 (A), and the high concentration of imidazole had to be removed from the sample after the column run. Nonetheless, imidazole elution is a milder and often preferred method for native protein elution from a metal affinity resin and was the main method used to purify pol X_{iz}. The pH-based elution

made the peak detection easier, but required a low pH of 4.5, which is not always compatible with a target protein.

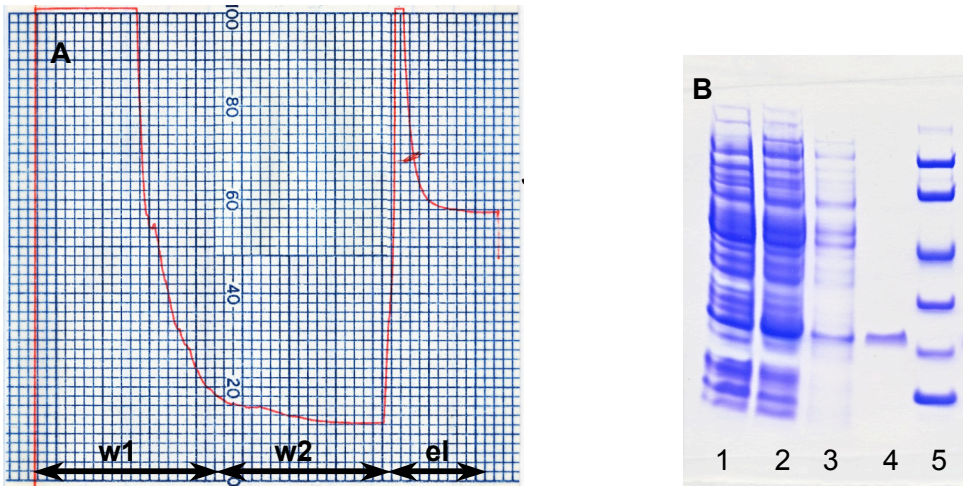


Figure 2-2: Imidazole based purification of pol X_{iz} using a Ni-NTA resin. **A.** A_{280} absorbance recorded during the purification using a Ni-NTA and wash buffer w1 (50 mM NaH_2PO_4, 300 mM NaCl, 5 mM imidazole, pH 7.5), w2 as w1, but 20 mM imidazole and elution (el) as w1 and w2, but 500 mM imidazole. **B.** Chomassie stained SDS-PAGE showing lanes with pre-induction (1), post-induction (2), sample from column wash cycle w1 (3), elution of purified protein (4) and LMW (14.4, 20.1, 30, 43, 67 and 94 kDa)

As an alternative to the Ni-NTA resin, a cobalt-based metal affinity resin referred to as Talon (Clontech, Palo Alto, Ca) was used. Advantages of this resin, as claimed by the manufacturer, included a tetradentate Co chelator that would bind more strongly to the metal in an electronegative pocket and reduce the possibility of metal leakage even in the presence of strong denaturating agents. Talon was also described to bind more

selectively to the polyhistidine-tagged proteins due to a more stringent spatial requirement compared to the Ni based resins, while reducing the affinity, so that the host protein could be eluted under milder conditions. The buffers for the purification of pol $X_{iz/orig}$ with the Talon resin contained 50 mM sodium phosphate and 300 mM NaCl at pH 7.0 for the initial wash cycle. For the second wash cycle, 5 mM imidazole was added, and 200 mM imidazole was used for the elution. For the pH-dependent elution, the buffer was changed to 50 mM sodium acetate and 300 mM NaCl at pH 5.0. Figure 2-3 shows a comparison of the three different purification methods. Best protein purity was achieved using the imidazole based elution scheme on a cobalt based affinity column as indicated in lane 3 of Figure 2-3 (A), followed by a pH 5.0 elution with sodium acetate on the same column. Imidazole-based purification on the Ni-NTA column was more prone to some higher molecular weight impurity as is indicated in Figure 2-3 (B).

In cases where some additional protein bands were visible on the SDS-PAGE gels as indicated in Figure 2-3 (B), further purification of the protein was performed. Larger samples were concentrated using an Amicon Ultra-4 10 kDa spin filter followed by a gel filtration Sephacryl S-100 high-resolution chromatography column (18”) (Amersham Pharmacia Biotech, Piscataway, NJ). Separation was done with 20 mM

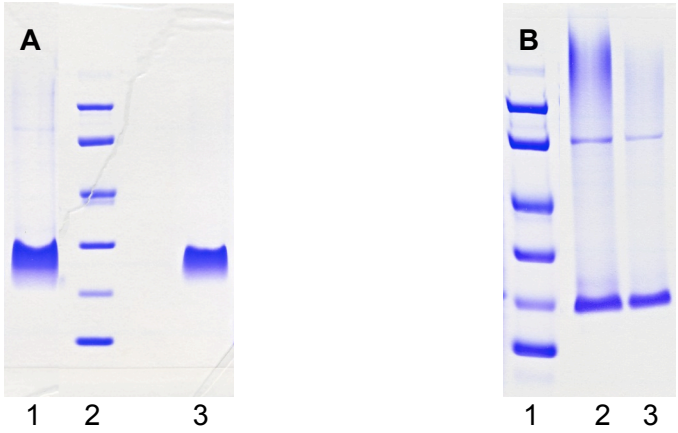


Figure 2-3: SDS-PAGE of pol X_{iz} purified on a cobalt- or nickel-based affinity column. **A.** Purification on Talon resin, pH and imidazole based elution; lane 1 elution of target protein with sodium acetate, pH 5.0, lane 2, LMW marker (14.4, 20.1, 30, 43, 67 and 94 kDa), lane 3 from elution with 200 mM imidazole. **B.** Purification on Ni-NTA resin, lane 1, LMW marker, lanes 2 and 3, 10 µL and 5 µL samples after elution with 500 mM imidazole.

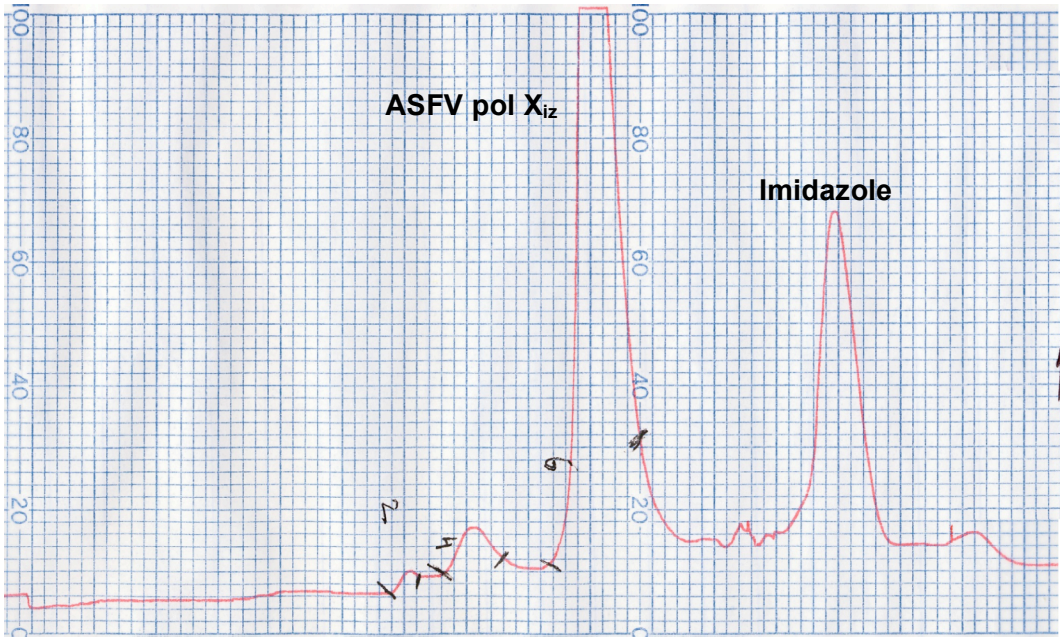


Figure 2-4: Chromatogram of gel filtration with Sephacryl S-100 HR column. Elution buffer was 20 mM sodium acetate, 1 mM $MgCl_2$ and 0.02% NaN_3 at pH 6.0. Main fraction consists of the target protein, a later strong peak was identified as imidazole.

sodium acetate, 1 mM $MgCl_2$, and 0.02% NaN_3 at pH 6.0, monitoring A_{280} with an ISCO UA-5 flow cell and recorder. The graph in Figure 2-4 shows a typical elution profile from the size exclusion column. It was possible to separate several impurities from the main protein fraction, and imidazole was removed during this separation step as well. The size exclusion step took the place of the dialysis. After the determination of the protein concentration by UV, a final concentration of 5 mM dithiotreitol (DTT) was added, the sample concentrated, and for NMR samples, 5% D_2O was added.

Alternatively, a buffer exchange was achieved by dialysis using MWCO 10 kDa Slide-A-Lyzer cartridges from Pierce into the same sodium acetate buffer used for the S-100 column. It was not uncommon to see some protein precipitation collecting in the cartridge. Protein yields for all the purification methods on this construct varied between 5 and 15 mg per 1.6 L minimal medium expression.

Microdrop Screening to Optimize the Solvent Conditions for ASFV Pol X_{iz}

During the early work with this enzyme it became apparent that the buffer conditions reported for the biochemical assays (*9*) were not suitable for the micromolar to low millimolar concentrations required in NMR samples. While several of the more common buffer conditions were tried, none appeared to retain the polymerase in solution

for an extended time period. There are many conditions influencing the solubility of a protein, most importantly the pH, which impacts the choice of the buffer salt or the salt concentration. Other factors such as reducing agents or detergents may also have a significant impact on the long-term stability of a protein, but to find a proper combination between the many options is a tedious, time-consuming task requiring a large amount of protein.

Bagby et al (*79*) have reported a method using microdialysis to test the solubility of proteins at NMR concentrations in various buffers. Another report by Lepre and Moore (*80*) addressed the same issue by using microdrops in a sealed reservoir allowing for simultaneous screening of multiple conditions. This method was based on the well-established 'hanging drop' technique used extensively by the X-ray crystallographer community, but its aim was inverted for this application. Instead of screening with precipitants to form crystals, the new purpose of the screen was to find stabilizers that promoted protein solubility.

The general microdrop strategy was to place drops of protein solutions that were close to their solubility limit on a glass cover slip, mixed each drop with a different buffer, inverted the slip and placed it on a reservoir containing the corresponding buffer. The cover slip and reservoir were sealed allowing the system to equilibrate. Water would

slowly diffuse from the drop to the reservoir solution, increasing the protein concentration in the drop. If the buffer had a stabilizing effect, the protein would remain soluble, otherwise it would exceed its solubility limit and precipitate.

While the buffer conditions equilibrated over the first 24 to 48 h, the protein stability in this new buffer environment could be followed over days and weeks. It was often seen that a protein could be fully dissolved at NMR concentrations but gradually precipitated over the following hours or days.

Experimental Setup

A batch of 4 x 800 mL pol X_{iz} was grown in LB medium and purified over a Co-affinity column, eluted with imidazole. The protein containing fractions were dialyzed twice against 20 mM sodium acetate and 50 µM ethylene diamine tetracetic acid (EDTA) at pH 6.0. To minimize the impact of sodium acetate in the solubility test, the third dialysis was performed against 2 mM sodium acetate in order to maintain the pH. Under these conditions the samples were stable for only a short time, long enough to prepare the screen but precipitation formed after a few hours. The concentration was determined to be 1 mM, representing a protein concentration used for NMR measurements. Another sample was concentrated to 2.5 mM to test the solubility limit during the screen.

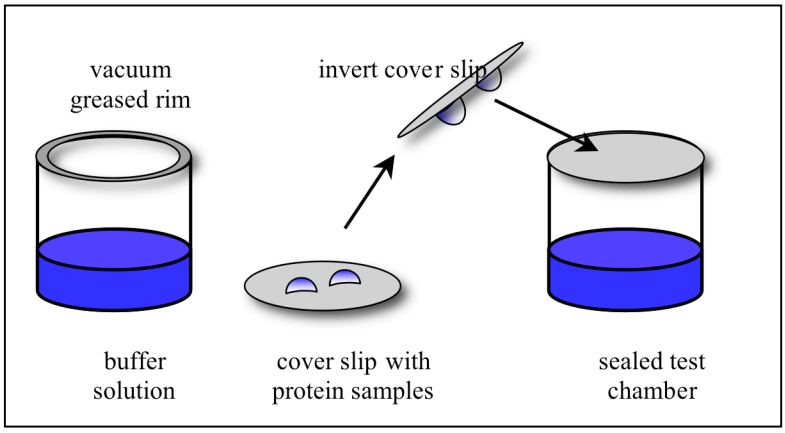


Figure 2-5: Schematic drawing of sample preparation for a single solubility screen. Each well of a 24 well VDX linbro plate is filled with 1 mL of test buffer and the rim covered with vacuum grease to provide a tight seal. Two drops of a 2 µL protein solution are placed on the siliconized cover slip and mixed with 1 µL of buffer solution from the well. The cover slip is inverted and placed on the buffer well, making sure it is sealed well.

Multiple screens were set up to test a number of buffer and stabilizer combinations. The screens were performed in standard 24 well VDX plates with pre-cast sealant (Hampton Research, Aliso Viejo, CA). Where not noted otherwise, 200 mM buffer stock solutions at various pH and stabilizer solutions were used. The procedure followed the steps depicted in Figure 2-5, where first 0.5 mL of the main buffer were pipetted into the well. Additional solutions, e.g. salts or stabilizers, were added depending on the individual test, and the difference to 1 mL was made up with water. One or two 2 µL aliquots of the above described protein solution were pipetted onto the siliconized glass cover slips, immediately followed by 2 µL buffer solution from the

reservoir buffer. Gentle mixing of the drops was achieved through slow pumping on the pipette. The cover slip was inverted and placed atop of the corresponding well. The pre-cast silicon vacuum grease assured a tight seal for each individual well. For the duration of the screen, the wells were kept inside the incubator at room temperature to minimize environmental temperature fluctuations. Changes were monitored by visual inspection of the drop through a 10x magnifying glass against a black background. The scoring system proposed by Lepre and Moore was used to describe the drops; a score of 0 indicated a clear solution, 0.5 very light cloudiness throughout the drop, 1 indicated a stronger precipitation in the center of the drop, 2 and 3 growing precipitation at ~50% and ~75% coverage, respectively, and a score of 4 indicated precipitation completely covering the drop. Additional observations were indicated with letters; c indicating crystals, r for a ring-like precipitation, d for spot precipitation on the glass cover and “dry” indicated that the seal was not tight enough and the drop eventually dried out.

The objective of an initial screen was to identify a suitable buffer system and pH condition. Commonly used NMR buffers were tested and included sodium- and potassium phosphate, sodium acetate, 1,4-piperazinediethanesulfonic acid (PIPES), trishydroxymethylaminomethane (Tris), 2-(N-morpholino)ethanesulfonic acid (MES), and sodium cacodylate. The focus was on buffer systems that allowed for a slightly acidic pH

with the exception of Tris, since initial data and a calculated pI of 10.0 would suggest a better solubility at slightly reduced pH. Observations were recorded over a time course of 70 days and the individual data are listed in Appendix B, Table 1.

Screen 1: Individual Buffer Systems

The phosphate based buffers, which are much desired for NMR work because of the absence of protons, showed a unified negative result in this screen. Heavy precipitation was observed regardless of pH (4.0-7.0) and cation type (sodium or potassium). The same was true for sodium citrate. Interestingly, dissolving the protein in Tris at pH 7.0 and 8.0 or PIPES pH 6.5 also caused slight to severe precipitation over the first few days. Those were the buffers used for the biochemical assays by Oliveros et al (*9*) and the NMR studies by Mullen and coworkers (*46, 81*), and shows, that the buffer salt itself is not always sufficient to dissolve a protein and keep in solution for an extended time period, but requires additional components. The microdrop containing the protein dissolved in water only confirmed the observation that the solubility under these conditions is of a short-term nature. By far the best performance was observed for sodium acetate in a pH range of 4.5-6.5, followed by sodium cacodylate at pH 6.5 and HEPES-Na at pH 6.6.

Screen 2 and 3: pH Variations and Salt Additives

The following two screens involved the same base buffers but varying pH in the presence of 20 mM $MgCl_2$, $CaCl_2$, $(NH_4)_2SO_4$, and NH_4Ac. Sodium citrate was no longer included in these series. Some of the salt additions provided a stabilizing effect towards the protein, which can be seen in Appendix B, Tables 2 and 3. The protein sample used for the third screen was 2.5 mM compared to 1 mM for the first two screens and several buffers were used in both screens to investigate the impact of the higher protein concentration. Adding only the four salts to the aqueous protein solution did not improve the stability and precipitation appeared within days. The phosphate-based samples continued to show precipitation within the first week for all samples and the solubility of the MES based buffer was not convincing either. A decreased stability was found for the HEPES-Na based buffer and the protein precipitated quickly from these solutions, which could also be attributed to the higher protein concentration. Sodium cacodylate alone provided good solubility and with the exception of the additional $(NH_4)_2SO_4$, further improvement on the long-term stability was seen in this buffer, regardless of the protein concentration. No detrimental effect by any of the additional salts and higher sample concentration was found for sodium acetate and all solutions with this buffer remained clear throughout the screen. A delayed precipitation pattern was seen for the Tris-based

system. The combination of $CaCl_2$ and phosphate based buffers was not tried, since the two salts form insoluble products.

Screen 4: NaCl Concentration

Another screen focused on the effect of sodium chloride in the buffer solution, results are listed in Appendix B, Table 4. Stock solutions of 1 M NaCl were added to the previously screened reservoir buffers to achieve final concentrations of 25 mM, 50 mM, and 100 mM NaCl. The 2.5 mM protein sample was used for this test. Similar to previous findings, protein solutions containing sodium acetate or cacodylate performed best. Solubility in the Tris buffer was only maintained at lower salt concentrations. Interestingly, bad solubility was found for PIPES up to 100 mM NaCl, which was consistent with later findings that 500 mM NaCl were required for long-term solubility of the protein in PIPES.

Screen 5: Include Stabilizers

The solubility of the protein in the three buffers, Tris, sodium acetate and sodium phosphate was examined in the presence of some common stabilizers like DMSO, glycerol, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) and a

ligand, ATP. Appendix B, Table 5 summarizes the observations. Protein solubility was good in the sodium acetate buffer in the presence of all those compounds, except 30% DMSO and ATP. The only other positive result in this screen was the combination of Tris and glycerol.

Screen 6 and 7: Conditions for Sodium Acetate and HEPES-Na Based Buffers

The last two screens focused on combinations of the sodium acetate and HEPES-Na based buffers in the presence of various stabilizers and the results are listed in Appendix B, Tables 6 and 7. Previous screens showed excellent solubility of pol X_{iz} in sodium acetate based buffers. A final screen included reducing agents like DTT and β-mercapto-ethanol (BME), two different pH's and additional organic solvents and detergents. Only the presence of 3 mM SDS, DMSO at higher concentrations, and ATP caused some precipitation. The HEPES-Na screen was included since data in screen 1 indicated some potential benefit with this buffer system, but the results in Appendix B, Table 7 show, that the protein did not remain in solution for a longer time period with most HEPES-Na based combinations.

Conclusions of Solubility Screen Trials

The challenge to dissolve the pol X_{iz} protein at millimolar concentrations in a suitable buffer system was addressed by a microdrop screening method, that provided a tool to systematically compare the solubility and long-term stability in dozens of buffer conditions at NMR concentrations, but required only small amounts of protein.

The screens showed, that the pol X_{iz} dissolved well in sodium acetate based buffers at a pH between 5.0 and 6.5 and remained in solution for weeks at millimolar concentrations in these buffers. Only the additions of 3 mM SDS, DMSO, and ATP interfered with that solubility. Sodium cacodylate buffers exhibited almost identical characteristics as the sodium acetate buffers. Given the choice between the two, the sodium acetate buffer was preferred over the sodium cacodylate one due to the toxicity of the arsenic present in the cacodylate buffer.

Pol X_{iz} remained well dissolved in only a few of the Tris-based buffers. Tris alone was not able to keep the protein soluble but adding 25 mM NaCl or glycerol sustained solubility. Interestingly, higher NaCl concentrations and the addition of $MgCl_2$ reduced the solubility, which was also observed in some of the purification schemes. Tris buffer solutions did well in a pH range of 7.5 and 9.0, which was higher than what the protein appeared to like and might have contributed to the reduced solubility in this buffer

system. Unfortunately none of the preferred phosphate based buffers provided ample solubility for the protein and could not be considered for use. Neither did MES, HEPES-Na or sodium citrate based buffers.

The most suitable three buffers, sodium acetate, sodium cacodylate and Tris, contain protons giving rise to strong NMR signals, which could interfere with some of the NMR measurements. All three buffer salts were found to be available in a fully deuterated form and the sample preparation could include a final buffer exchange step to introduce the deuterated buffer.

As a result of this microscreen, the purification of pol X_{iz} was done in a buffer system containing 20 mM sodium acetate, 1 mM $MgCl_2$ and NaN_3 at pH 6.0.

ASFV Pol X_{gpm} Construct

The construct, kindly provided by the late Gregory Mullen, contained the open reading frame of the ASFV pol X gene, inserted into a pET23a plasmid (Novagen). Expression was done in *E. coli* strain BL21(DE3) containing the pLysS vector, as was the case for the previous two constructs. A major difference was the absence of any additional amino acid residues. The protein over-expressed by this vector contained the 174 residues of ASFV pol X as described by Yáñez et al (*56*). Therefore concerns of any

potential interference due to a C- or N-terminal peptide were diminished and the molecular weight was minimal. A disadvantage of this construct was the need for native purification since there was no purification aid (e.g 6x His tag) available. The expression and purification procedure was followed as reported in the literature (*81, 82*). A detailed protocol including a description of all the solutions and conditions can be found in Appendix C.

After the purification on the CM-20 ion exchange column, four low molecular weight bands were seen as shown in Figure 2-6 (B). Although these bands were not very intense, they indicated some impurity of the protein. To improve the purity, an additional separation on a S-100 seize exclusion column was included but the additional bands in the SDS-PAGE remained present. Additional purification was successfully done by HPLC using a Phenomex Luna C-8, 25 cm X 4.6 mm column. The protein was first concentrated with an Amicon Ultra-15, 10 kDa MWCO spin filter to about 1.5 mL and injected in 100 to 150 µL increments. Solvents were A = 0.1% TFA in H_2O and B = 0.1% trifluoroacetic acid (TFA) in acetonitrile. Reversed phase chromatography was run with a gradient and a flow rate of 2 mL/min, using a Beckmann HPLC System Gold (Beckmann Instruments, Fullerton, CA). Starting conditions were 10% B, increasing the organic phase during 30 min to 70% B, isocratic flow for the following 10 min, reverse back to

10% B during the next 10 min and remain at this composition for another 10 min. The retention time for pol X_{gpm} under these conditions was 22.3 min corresponding to an elution at 55% B. Fractions were subsequently dried in a CentriVap console (Labconco, Kansas City, MO) and stored in the refrigerator until used.

Several small impurities were removed during this HPLC purification step as is evident from the trace in Figure 2-7. MALDI-TOF mass spectra were collected to verify the proper molecular mass and purity of the sample. In the example shown in Figure 2-8, 1 µL of ^{15}N labeled, HPLC purified pol X_{gpm} was mixed with 1 µL matrix containing 10 mg/mL sinapinic acid dissolved in 400 µL H_2O, 100 µL TFA 5% and 500 µL acetonitrile. The measurement was done on a Voyager Elite mass spectrometer (Applied Biosystems, Foster City, CA). Calibration was done using the myoglobin standard. A mass of 20,555.7 Da was measured compared to the expected 20,562.4 Da.

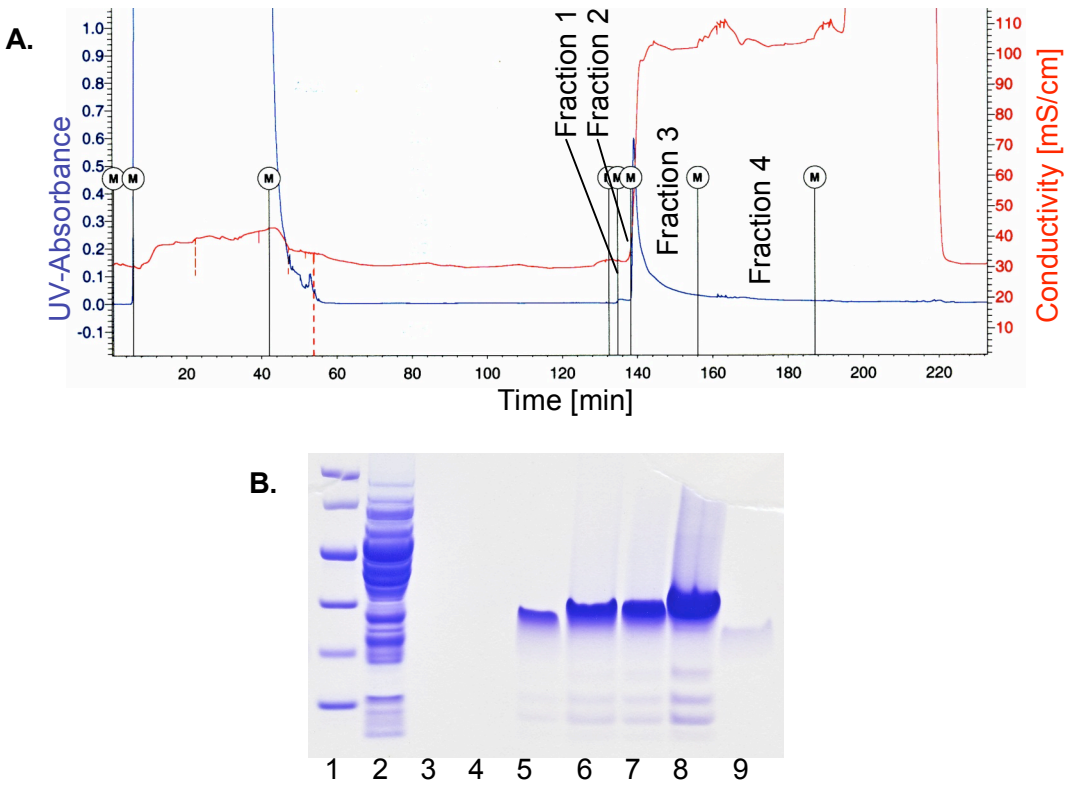


Figure 2-6: Purification of pol X_{gpm}. **A.** Typical CM-20 ion exchange profile, wash cycle with 20 mM PIPES, 2 mM BME, 5 mM EDTA, 0.5 mM PMSF, 0.02% NaN_3 and 250 mM NaCl, pH 6.5 during the first 130 min. Elution of pol X_{gpm} with 1 M NaCl in the previous buffer including fractions 1 to 4 and a column wash with 2 M NaCl between 190 and 220 min. Blue curve represents A_{280} absorbance, red curve the conductivity in mS/cm. **B.** 4-12% Bis-Tris SDS PAGE, lanes 1: LMW marker (14.4, 20.1, 30, 43, 67 and 94 kDa), 2: 20 µL sample from the wash cycle, 3: 20 µL fraction 1, 4: 20 µL fraction 2, 5: 5 µL fraction 3, 6: 20 µL fraction 3, 7: 5 µL of small sample taken close to maximum peak absorbance during elution, lane 8: 20 µL of same sample as used for lane 7, lane 9: 20 µL fraction 4.

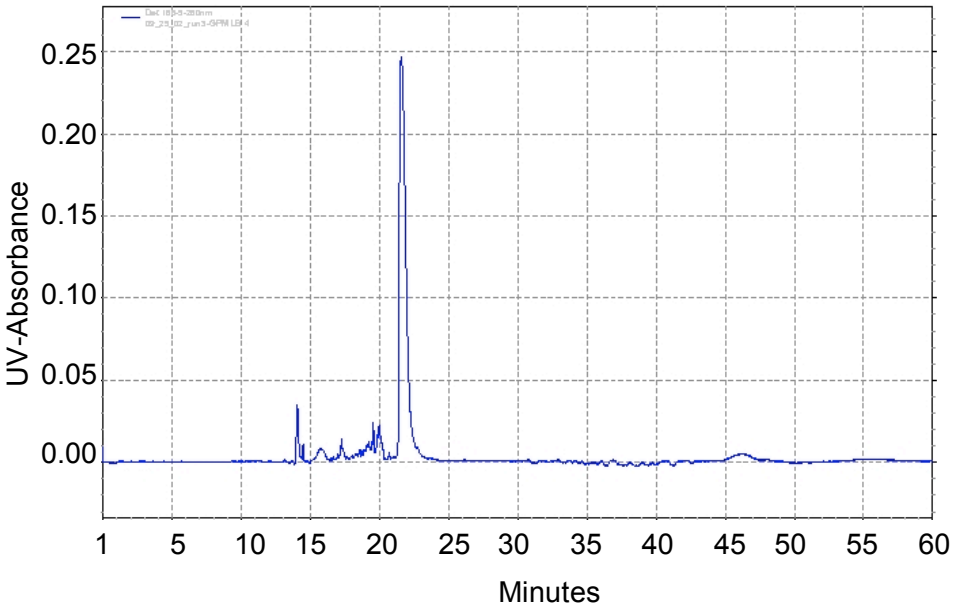


Figure 2-7: HPLC purification of pol X_{gpm}. Solvents were A = 0.1% TFA in H_2O and B = 0.1% TFA in acetonitrile. A gradient from 10% B to 70% B during the first 30 min was used, isocratic conditions for the following 10 min, reversing back to 10% B during the next 10 min and remain at this composition for another 10 min. The flow rate was 2 mL/min.

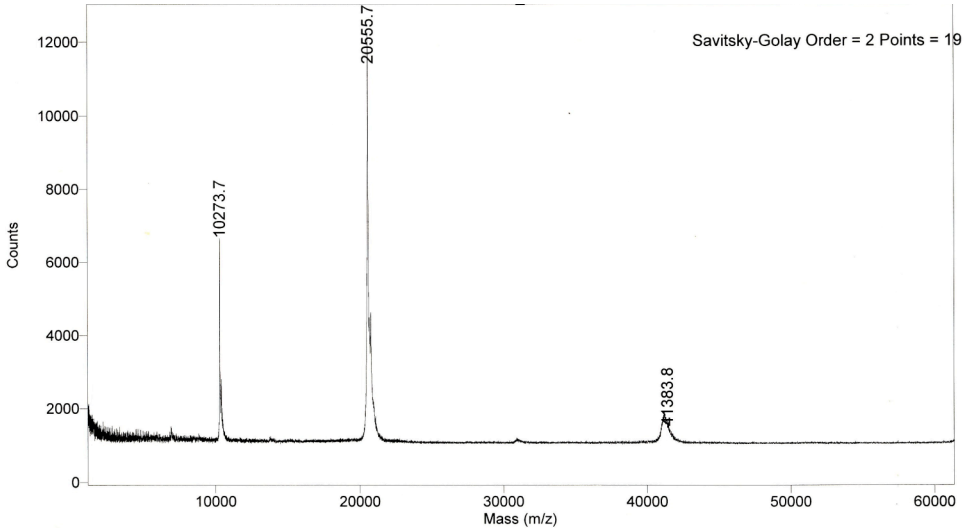


Figure 2-8: MALDI TOF mass spectrum of ^{15}N labeled pol X_{gpm}. The following acquisition parameters were used: 82 scans, positive ion, linear mode, accelerating voltage of 25,000 V, grid voltage at 93% and guide wire voltage at 0.2%, laser power 2,700, low mass gate at 1,000 Da, mirror ratio of 1.060, a pulse delay of 150 ns and pressure at 2.63×10^{-7} bar. A Savitzky–Golay smoothing filter was used for data processing.

A comparison of ^{15}N-HSQC spectra before and after HPLC purification confirmed the added purity and is shown in Figure 2-9. Many of the small correlation peaks with a proton chemical shift between 8 to 8.5 ppm were attributed to impurities and were no longer present in the spectrum taken after HPLC purification. Another concern was whether the protein would be properly folded after the HPLC purification. The NMR spectra indicated that the protein refolded into the same confirmation where all resonances appeared at the same position. Additional activity assays were performed to show retained enzyme activity and are discussed below. Later in the work it was seen, that the bulk of the impurities originated from the front of the elution peak at 1 M NaCl during the CM-20 separation. Sacrificing a small portion of the initial fraction yielded a much cleaner product (>95% estimated from NMR spectra) with acceptable yields making the HPLC purification step unnecessary, while still providing pure protein.

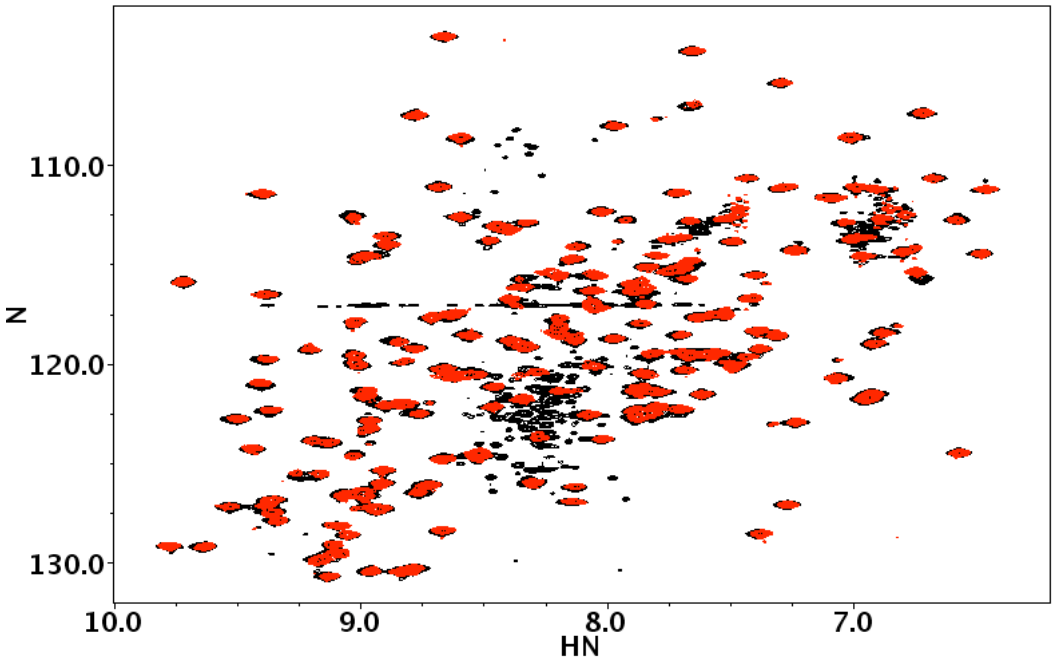


Figure 2-9: Superposition of ^{15}N-HSQC spectra of pol X_{gpm} after CM-20 ion exchange column purification (black) and additional HPLC purification (red). The same NMR buffer was used for both samples, containing 20 mM PIPES, 20 mM $MgCl_2$, 0.02% NaN_3 and 500 mM NaCl at pH 6.5.

Activity Assays of Pol X

It has been shown that ASFV pol X has activity to incorporate deoxynucleotide triphosphates (dNTP) and extend a DNA primer according to the template (*9*). It was important to ensure that the purified enzyme retained activity for all three individual constructs. Two assays were used to test for enzyme activity: the insertion and extension assays. The insertion assay not only monitors the activity of the enzyme, but it also provides information on the fidelity of a polymerase. Only one type of dNTP was

added to the enzyme in this assay. A high fidelity polymerase would extend the primer by the correct dNTP that was complementary to the template at that position. All other dNTP's would not be inserted. For the extension assay, also called polymerization assay, the enzyme was incubated in the presence of all four dNTP's allowing for multiple primer extension steps, often to the full length of the template.

The 5' end of the primer DNA was labeled with ^{32}P-γ ATP and purified on a 20% PAGE. Samples were incubated in 50 mM Tris•HCl, pH 7.5, 10 mM $MgCl_2$, 5 mM DTT, 0.1 mg/mL BSA, and 4% glycerol. Enzymes and dNTP were added in various concentrations as indicated for the individual experiment. Incubations were carried out at 37 ºC for 30 min, and the reactions were stopped by heating the sample to 70 ºC for 15 min in the presence of 10 mM EDTA. The samples were resolved on a 20% denaturing PAGE and monitored by autoradiography.

Activity of Pol X_{gpm} Pre- and Post HPLC

A comparison of the activity before and after the HPLC purification indicated, that the pol X_{gpm} protein refolded properly and retained activity as shown in Figure 2-10. In this case full extension of a gapped DNA, a 28/15-mer template-primer with a 12-mer downstream phosphorylated primer, was seen. Incorporation for all three non-

complemantary dNTP's was observed as well, with the exception of dATP. These observations reflect the literature, where in some cases an incorporation with high fidelity (*9, 63*), in other cases the opposite, very low fidelity, has been reported (*61, 64*).

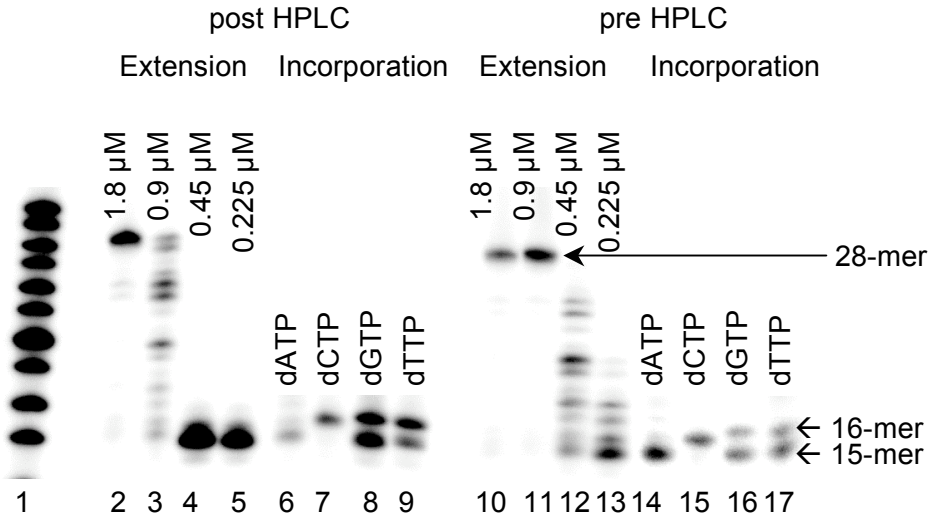


Figure 2-10: Activity assays of pol X_{gpm} samples taken before and after HPLC purification. The assays were performed under conditions described in the text. Polymerase concentrations for the extension were as indicated in the figure, 1.8 µM enzyme was used for the incorporation assay. A 28/15/12-mer double-stranded DNA with a single G-gap was used. Template-Primer sequences were as followed: 28-mer template 5'-AGTGGCCGTCGTGGTACTCACTGTGATC-3', the 15-mer primer was ^{32}P-5'-GATCACAGTGAGTAC-3' and the complementary 12-mer sequence was 5'-pACGACGGCCAGT-3'. Lane 1: DNA marker 14-32 oligomer in 2 base increments, lanes 2 to 9, sample of HPLC purified pol X_{gpm} were used, samples for 10 to 17 were taken after CM-20 ion exchange column. Lanes 2-5 and 10-13 represent the extension in the presence of all 4 dNTP's, 6-9 and 14-17 the incorporation with the specified dNTP.

Activity Pol X_{orig} in the Presence of a Primer-Template or Gapped DNA

Further comparisons were done on uncleaved pol X_{orig} protein, using a template-primer with 17-mer and 9-mer ss-DNA, as well as a gapped ds-DNA construct with the same template-primer, but including a phosphorylated 7-mer downstream primer creating a single base gap. Both, extension and insertion assays were performed and are shown in Figure 2-11 (A-D). At 2.5 and 5 μM pol X_{orig}, minimal extension of 2 bases was observed for the template primer construct. Only one base was inserted for the gaped DNA, Figure 2-11 (C).

The insertion assay showed a faithful insertion of two dCTP opposite G of the template, but an additional and unexpected third insertion band was observed as well. The extension on the gapped DNA, Figure 2-11 (D), showed a main product of dCTP properly inserted opposite G, but an additional dCTP opposite G, possibly displacing the terminal cytosine of the 7-mer. A further band indicates some degree of misincorporation of dGTP that would lead to G → C transversion. As indicated above, Tsai et al have reported the unfaithful incorporation of bases for ASFV pol X, which appears to be seen in these assays as well.

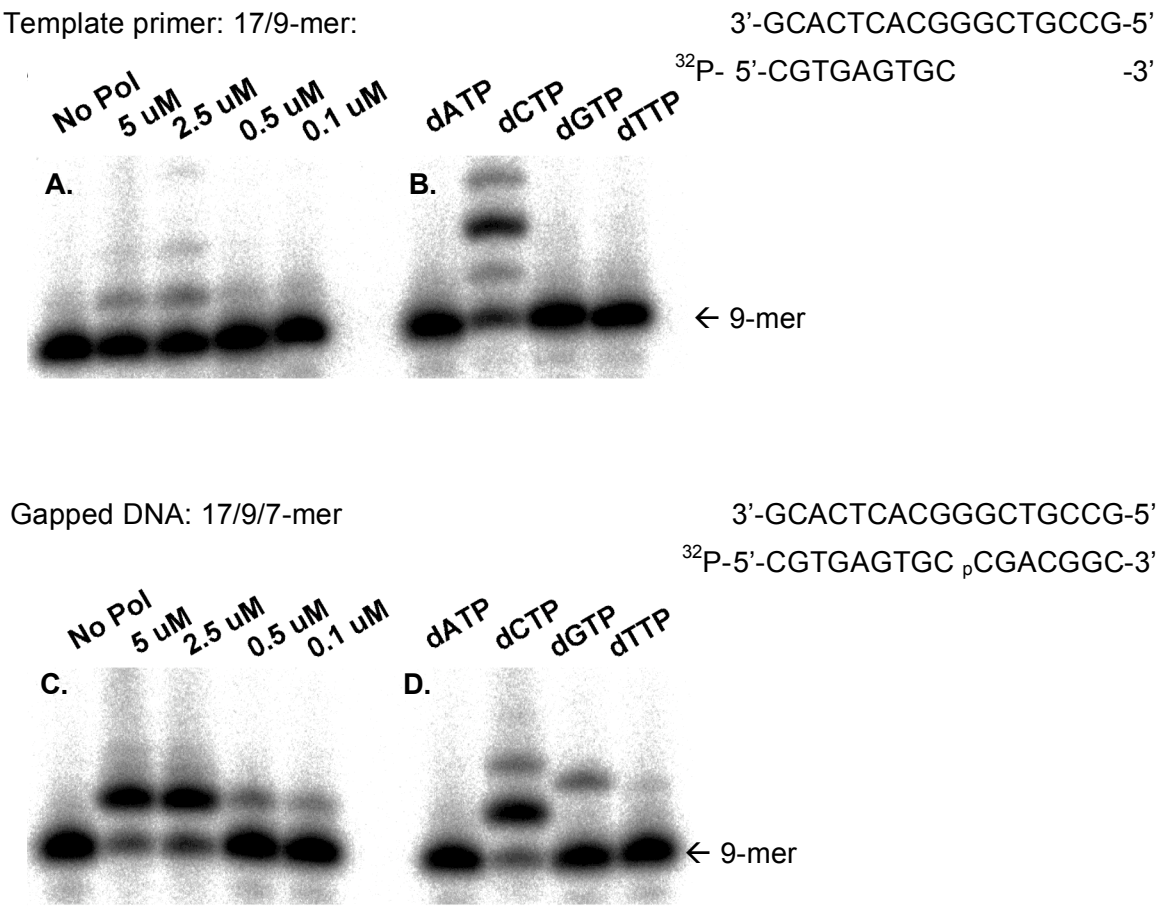


Figure 2-11 **A-D**: Activity assay of pol X_{orig} were performed under conditions described in the text. Polymerase concentrations for the extension assays **A.** and **C.** were as indicated in the figure, 5 μM enzyme concentration was used for the incorporation assays **B.** and **D.** 400 nM DNA and 100 μM dNTP's were used in all assays. The 17-mer template sequence was 5'-GCCGTCGGGCACTCACG-3', the 9-mer primer was ^{32}P-5'-CGTGAGTGC-3' and the complementary 7-mer sequence was 5'-pCGACGGC-3'. Panel **A.** and **B.** show the result for the 17/9 template-primer oligonucleotides, **C.** and **D.** for the single gap double stranded DNA.

Conclusions

In this chapter the expression and purification of three different ASFV pol X constructs were described. The proteins differed in that pol X_{orig} had a 35 amino acid leader peptide containing a 6x His tag for easier purification, likewise pol X_{iz} contained a 6x His tag but on the C-terminal end with a single linker residue, while pol X_{gpm} represented the unmodified, full length 174 residue ASFV polymerase. It was shown that all of these proteins could be expressed and purified to a level required for NMR samples. The purification under denaturing conditions initially performed for the pol X_{orig} was later modified to an imidazole based Ni-NTA elution followed by a S-100 seize exclusion column. To retain solubility of the protein, a low pH of 5.0 to 6.5 seemed to be crucial and higher salt concentrations were advantageous. Reducing the overall protein size by cleaving off the N-terminal leader peptide with recombinant enterokinase proved challenging. While the cleavage worked fine for small-scale pilot tests, larger sample with NMR quantities produced small yields. Based on today's knowledge it is assumed that the cleaved protein was largely insoluble in the buffer provided and was removed together with the enterokinase capture agarose.

Alternatively the pol X_{iz} was expressed. Although a large amount of protein was eluted with either the Ni-NTA or Talon column using both, a pH gradient or imidazole

based elution, substantial precipitation occurred during the subsequent dialysis. This precipitation indicated a solubility problem and the commonly used buffer combinations showed no improvement. An extensive solubility screen was performed resulting in sodium acetate and sodium cacodylate based buffers at pH 5.0 to 6.5 to be the most suitable buffer systems in which pol X_{iz} remained soluble over several weeks.

The preferred protocol to produce NMR suitable samples on pol X_{iz} involved a purification on the Co-affinity column, elution with imidazole and a subsequent dialysis or size exclusion column against 20 mM sodium acetate, 1 mM $MgCl_2$, 0.02% NaN_3, and optionally 50 mM NaCl at pH 6.5.

The full length, 174 residue containing pol X_{gpm} was expressed and purified according to literature that became available just as the conditions for the previous constructs were worked out. An additional HPLC purification step was introduced which resulted in better protein purity while maintaining enzyme activity. Alternatively, discarding the initial 5 to 10% of the elution during the ion exchange column yielded protein with comparable purity. The report of the NMR structure on free ASFV pol X protein caused a major shift in the aim of this project towards studies of the binary/ternary complexes and the decision was made to pursue further work with this construct, which included the use of the same buffer conditions described in the

literature (*46*) This made a comparison of NMR data for a pol X-complex with the ones published for the free pol X possible.

CHAPTER III

FLOURESCENCE AND ITC STUDIES OF ASFV POL X – NUCLEOTIDE AND ASFV POL X - DNA – COMPLEXES

Introduction

The focus of this chapter is on the interaction between pol X and its interacting partners, the nucleotides and oligonucleotides. These biochemical interactions are often of non-covalent and reversible nature. As a matter of fact, this is vital for life itself. Many biological processes rely on the fact, that a substrate is bound to a specific protein or protein complex for a certain time in order to accomplish a task. After this task is accomplished, the substrate needs to be released again, in order for the protein to start the process over on a new substrate. This is also the case for the pol X enzyme, where one function is to insert a proper nucleotide in a gap on the DNA strand. Other examples include a multitude of biological processes, where small molecules are directly involved in manipulation regulatory pathways, energy production, signaling or many more examples. The whole immune system is based on some reversible binding and regulation of small molecules or ions, causing some specific response in the biological signaling pathways. By no means are these biochemical reactions limited to protein-

small molecule interactions. They are equally important between protein-protein or protein-DNA interactions, where in some cases entire macromolecule machineries assembles to fulfill one or multiple task. All these reversible interactions are essential for the proper function of cells and life in general.

In a general biological context, interactions of two non-identical molecules are typically referred to as ligand receptor interaction. Although loosely defined, the receptor is typically the larger protein and the ligand the smaller molecule.

The affinity between a receptor and a ligand describes the strength of the binding-forces between the two. It is an indication of how well a certain ligand fits into a receptor, or how strong for example some electrostatic interactions, hydrogen bonding, hydrophobic or Van der Waals forces are on the interface. The quantitative measure of the affinity is the association – or dissociation constant K_a, K_d, respectively and the free energy ΔG^o. A low K_d or negative ΔG^0 indicates a strong affinity. The rates at which these associations and dissociations occur are defined by the rate constants k_{on} and k_{off}. An entire field evolves around the study of these equilibrium and rate constants in order to accurately describe how a receptor interacts with its ligand. These systems are commonly quite complex, involving multiple cascade processes where many events influence each other.

The above-mentioned rate and equilibrium constants can be described as followed for a simple receptor-ligand interaction:

$$R + L \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} RL \qquad [3.1]$$

where R is the receptor, L the ligand, k_{on} and k_{off} the rate constants for the complex R L formation or dissociation, respectively.

$$K_d = \frac{[R]*[L]}{[RL]} = \frac{1}{K_a} = \frac{k_{off}}{k_{on}} \qquad [3.2]$$

$$\Delta G^o = R*T*\ln K_d = -R*T*\ln K_a \qquad [3.3]$$

The average amount of time a ligand will be part of the complex is described by the time $t_{1/2}$, which is equal to $\ln 2/k_{off}$. For a small molecule ligand, k_{on} is often determined by the time it takes for a ligand to make contact with the receptor. The on-rate then is diffusion limited and falls in the range of about 10^6 to 10^8 M^{-1} s^{-1}. For a tightly bound complex with a K_d in the nM range, this would result in an off-rate of 10^{-3} to 0.1 s^{-1} and a $t_{1/2}$ of seconds to hours, in which the ligand can be bound to the receptor. This time is much shorter for weakly bound systems, where the complex might persist only for µs to ms when the K_d is millimolar. The importance of this becomes evident, when designing an experiment to measure either rate or equilibrium constants. Some measurements are more suitable for certain time regimes than others. While these

considerations are true for diffusion controlled k_{on} rates, it is important to realize, that many biological interactions exhibit much slower binding behavior, meaning lower k_{on} rates, which means that the time regimes for $t_{1/2}$ are shifted.

The determination of equilibrium constants requires an accurate measurement of all individual molecular concentrations at any given time. Typically one parameter can be recorded and the others have to be derived from the overall system description. For a first order kinetic system one can assume a linear response between the portion of the receptor in the bound state compared to the total protein concentration and describe the response (I) at a given titration point as followed:

$$I(t) = I^{o} + (I_{\max} - I^{o})\frac{[RL]}{[R_{tot}]} \qquad [3.4]$$

In a common experimental setup, the initial concentration of the receptor is determined and known. This total receptor concentration does not change over the course of the experiment and can be called $[R_{tot}]$. What does change is its distribution between fraction that is bound [RL] to the ligand and the free state $[R_{free}]$. Any one of the available species can be monitored, depending on the method utilized. Another known parameter is the ligand concentration. Often, a well-characterized, concentrated stock solution of ligand is incrementally added to the receptor so that the total ligand concentration $[L_{tot}]$ at any given time is known as well. The difference between the total

ligand concentration and the one bound to the receptor is the free ligand concentration [L_{free}]. What always has to be accounted for is the dilution effect on ligand and receptor due to the addition of volume with the titrated ligand. In cases where the ligand can be added as a highly concentrated solution in a very small volume, this dilution effect is negligible. The following relations are true at any time during a titration experiment:

$$[R_{tot}] = [R_{free}] + [RL] \quad [3.5]$$

$$[L_{tot}] = [L_{free}] + [RL] \quad [3.6]$$

Substituting these general concentrations into Eq. 3.2 the following description of K_d results:

$$K_d = \frac{([R_{tot}] - [RL]) * ([L_{tot}] - [RL])}{[RL]} \quad [3.7]$$

Since the total receptor and ligand concentration is known, this equation has to be solved for the complex concentration and can be with the following rearrangements:

$$K_d\,[RL] = ([R_{tot}] - [RL]) * ([L_{tot}] - [RL]) \quad [3.8]$$

$$K_d\,[RL] = [R_{tot}]\,[L_{tot}] - [R_{tot}]\,[RL] - [L_{tot}]\,[RL] + [RL]^2 \quad [3.9]$$

$$0 = [RL]^2 - [R_{tot}]\,[RL] - [L_{tot}]\,[RL] - K_d\,[RL] + [R_{tot}]\,[L_{tot}] \quad [3.10]$$

$$0 = [RL]^2 - ([R_{tot}] + [L_{tot}] + K_d)\,[RL] + [R_{tot}]\,[L_{tot}] \quad [3.11]$$

The solution of the quadratic equation yields the following relationship:

$$[RL] = \frac{([R_{tot}] + [L_{tot}] + K_d) \pm \sqrt{([R_{tot}] + [L_{tot}] + K_d)^2 - 4([R_{tot}][L_{tot}])}}{2} \quad [3.12]$$

Combining Eq. 3.4 with 3.12 produces the final quadratic equation that will be used for experimental data fitting:

$$I(t) = I^{\circ} + (I_{\max} - I^{\circ}) \frac{([R_{tot}] + [L_{tot}] + K_d) \pm \sqrt{([R_{tot}] + [L_{tot}] + K_d)^2 - 4([R_{tot}][L_{tot}])}}{2[R_{tot}]} \qquad [3.13]$$

Although this equation contains only one unknown parameter (K_d) for each individual titration point, it can only be solved with an iterative fitting procedure. Several mathematical packages are available to accomplish this task. Either Prism Version 4.0c (GraphPad Software, San Diego, CA) or xcrvfit Version 3.0.6 were used for the analysis (xcrvfit: A graphical X-windows program for binding curve studies and NMR spectroscopic analysis, developed by Boyko, R. and Sykes, B.D., University of Alberta, website: http://www.bionmr.ualberta.ca/bds/software/xcrvfit).

A wide range of different methods is available to measure the dissociation constants. As mentioned above, depending on the type of binding, some methods are more suitable for a specific system than others. Factors that play a role in choosing a suitable method for the determination of binding are not limited only to the degree of binding, but solubility, concentration, amount available or physical properties of the receptor and ligand need to be considered. Commonly used methods are: chromatographic methods for semi-quantitative determinations, such as size exclusion columns, affinity columns or even filter based binding assays; gel mobility shift assays,

also called retardation gels; calorimetric methods or spectroscopic methods, including: ultraviolet spectroscopy (UV), circular dichroism (CD), fluorescence spectroscopy or NMR spectroscopy.

Binding Studies by Fluorescence

There is an enormously large and fast expanding range of fluorescence spectroscopy techniques available. Methods like chip based microarray sensors for detecting of multiple ligands simultaneously, flow cytometry, fluorescence microscopy and many more methods have evolved as essential and powerful tools for investigating biochemical systems. This study focuses on utilizing the fluorescent signal provided by the intrinsic tyrosine and tryptophan residue of pol X. Generally these residues will influence the fluorescence signal in a manner critically dependent on the folding of the protein. They act as sensitive monitors of conformational changes in response to denaturation, pH or temperature changes or complex formation with a ligand. Here, this technique was used to probe the interaction of a binary or ternary complex between pol X and nucleotide or oligodeoxynucleotides. There are a few important fundamentals that bear explanation at this point in order to interpret the data collected.

Fluorescence Measurements in Protein

When interpreting the emission spectrum of proteins the following factors need to be considered: 1) Tyrosine fluorescence is often not seen in the presence of tryptophan. This is due to a smaller intensity of tyrosine. The quantum yields after excitation at 280 nm are 0.2, 0.1 and 0.04 for tryptophan, tyrosine and phenylalanine, respectively (*83*). The absorption coefficient for tryptophan and tyrosine at 280 nm are 5540 and 1480 $M^{-1}cm^{-1}$, respectively (*84*), while phenylalanine absorbs only at 240-265 nm. Taking these two factors into account it becomes clear, why most protein fluorescence spectra are dominated by the tryptophan emission, where the signal of free tyrosine is about five times less intense than the one of tryptophan, but in a protein sample, it is often reduced even further. Reasons for this difference are a combination of interactions such as hydrogen bonding to peptide carbonyl groups, the presence of charged neighboring groups and non-radiative energy transfer from tyrosine to tryptophan or disulfide bonds. While the degree of quenching depends on the protein studied, all of these interactions result in a diminished or quenched tyrosine fluorescence signal. 2) The λ_{max} of tyrosine does not vary much at physiological pH and occurs at 303 nm. It is very little dependent on the polarity of its environment. This is in stark contrast to the behavior of tryptophan. Its λ_{max} depends heavily on the polarity of the immediate environment in the protein.

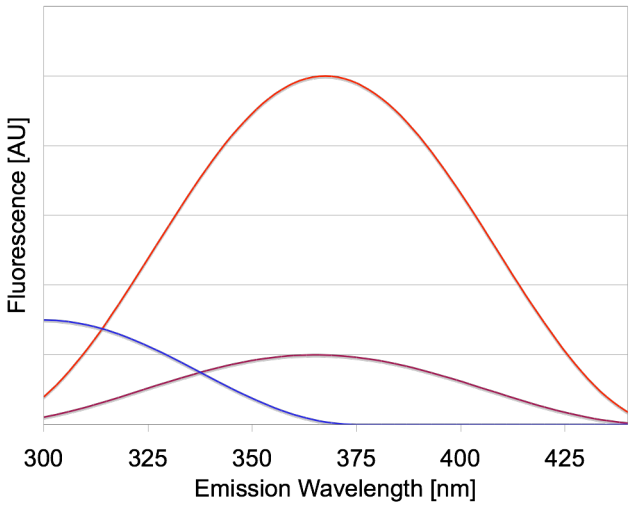


Figure 3-1: Schematic fluorescence emission spectra of equimolar solutions of tryptophan, red λ_{ex} 280 nm, purple λ_{ex} 295 nm and tyrosine, blue λ_{ex} 280 nm. Fluorescence intensity [AU] is in arbitrary units.

Values range from 308 to 352nm, 303nm for a highly non-polar environment of the tryptophan buried in the core of the protein and 352 nm for fully solvent exposed tryptophan as is the case in denatured proteins. A λ_{max} between the two extremes could either indicate that the tryptophan is buried towards the core of the protein, but within a more polar microenvironment, causing the red shift. Alternatively the residue could be positioned more towards the surface, where it is more exposed to the solvent or be involved in hydrogen bonds. Both factors would lead to a shift of the emission to longer wavelength and might not be distinguishable without further information. 3) As a general rule for protein measurements, an excitation of 275 nm will be used to observe both,

tyrosine and tryptophan residues. To observe the tryptophan alone, an excitation of 295 nm is common. The emission spectrum from tyrosine and tryptophan overlap substantially if tryptophan is in a non-polar environment and emits at lower wavelength. Depending on the residue ratio in the protein, the contribution of the tyrosine to the total signal intensities can be substantial, even though the tyrosine signal is much weaker. When exciting tryptophan selectively at 295 nm, no emission spectrum for tyrosine is observed anymore. The downside of this excitation is, that 295 nm falls into the tail of the tryptophan excitation spectrum, therefore fewer signal is excited resulting in a weaker emission spectrum. Even at an excitation of 280 nm, the tyrosine signal contribution becomes negligible though, if the tryptophan is red-shifted far enough.

Buffers should be chosen so that they provide a stable environment for the protein studied and do not interfere with the measurement. Two buffer additives commonly used in protein work may interfere and should be avoided: the presence of imidazole and oxidized DTT are not compatible with fluorescence spectroscopy, since they have an absorbance in the same wavelength range as tryptophan. Both can be removed by dialyses prior to the measurement or avoided all together. Reduced DTT can be added immediately before the experiment if no replacement is suitable.

Binding Studies by Isothermal Titration Calorimetry (ITC)

Isothermal titration calorimetry is a powerful, yet rather simple method to study and characterize protein-ligand interactions. Ligands may be other proteins or peptides, nucleic acids, carbohydrates, lipids or small organic molecules, metals, ions or protons that interact with a given protein (*85-87*). Furthermore ITC has also been utilized for protein folding studies (*88*). ITC measurements rely on the detection of any intrinsic heat change that occurs during the titration with a ligand. Heat is absorbed or generated in almost all binding events, which makes ITC an universal detector. The measured binding enthalpy can be the bases for the analysis of structural changes, activity or functions of molecules and has been successfully used to optimize drug targets (*89*). When executed with the proper control experiments, ITC can also provide insight into thermodynamic information about the molecular binding mechanism, reflected in the reaction enthalpy and entropy, which in turn determines the free energy of the event studied.

Some of the main advantages of the ITC measurement are: 1) A single measurement can provide the binding constant (K_b) of a given system, the stoichiometry of binding, or its enthalpy (ΔH) and entropy (ΔS). 2) No chemical tagging or immobilization of binding components is required, therefore true in solution reactions are monitored without the potential for undesired side effects interfering with the actual

binding site. 3) Colored solutions do not interfere with ITC measurements, as is the case in other spectroscopic methods. It is even possible to introduce turbid solutions or particulate suspension, if the reactant is not involved in this. If the reactant is aggregating or precipitating during the course of the measurement, it will falsify the results. 4) The measurements are fast and informative compared with some other methods. 5) A wide range of binding constants can be accurately determined with this method, K_b from 10^2 to 10^9 or larger, if competitive binders are utilized (*90*). 6) There is no limitation in the molecular size as seen with other methods.

With all these advantages, ITC seems to be the ideal method to measure any kind of binding events. Nonetheless, there are limitations that need to be considered as well. 1) In many systems, the resulting reaction enthalpy change is small and may cause inaccurate measurements, at least a very stable environment is required in this case. Additionally large proton ionization heats, as they are present for example in Tris•HCl and other buffers (*86, 91*) or pH mismatches between the reactant and titrant solution will create comparatively large amounts of heat during the mixing, interfering with the measurement. Therefore it is essential that both solutions be equilibrated to the exact same buffer conditions, preferably through dialysis. In the presence of low molecular weight compounds, where dialysis is not possible, one has to carefully adjust the pH,

salt, temperature and other buffer conditions to minimize the difference between the reactant and titrant solution. 2) The volume of the reactant cell is about 1.8 mL, leading to a substantial requirement of reactant for these measurements that is even larger for the titrant, which is added in excess. 3) The presence of DTT has shown to produce erratic baselines and should be avoided. Other reducing agents, such as BME may be used instead.

Keeping these limitations in mind, the following questions can be readily answered in simple experiments: Is there an interaction between the protein and ligand, what kind of affinity or stoichiometry is present, is the interaction well behaved, is there full activity of the protein. The determination of more specific parameters will require a significantly greater commitment in taking care of small interferences and possible errors (*92*).

Material and Methods

Chemicals and Buffers

Gpm-buffer: This buffer was used to dissolve the free pol X_{gpm} and contains 20 mM PIPES, 20 mM $MgCl_2$, 10 mM DTT, 0.02% NaN_3, and 500 mM NaCl at pH 6.5. Variations to these standard conditions were used in some cases and are pointed out

then. Those variations included a different salt concentration (NaCl), in some cases DTT was substituted by BME or deuterated PIPES and DTT were used for some NMR experiments.

Nucleotides and Oligonucleotides

Deoxynucleotide triphosphates were purchased from Sigma Aldrich (St. Louis, MO) as sodium salts. Stock solutions were prepared by weighing out an appropriate amount of the salt, dissolving and possibly diluting the deoxynucleotide triphosphates in the corresponding buffer. Dideoxynucleotide triphosphates were purchased from Fermentas (Glen Burnie, MD) as sets of four containing 20 mM, aqueous solutions titrated to pH 7.0 with NaOH. Final concentrations for all stock solutions were determined by UV spectrometric addition method.

Oligonucleotides were synthesized and purified by MWG Biotech (High Point, NC), using their unique high purified salt free (HPSF®) purification technique. To anneal the DNA, the samples were heated to 85 ^{0}C for 10 min using a water bath and then allowed to cool down over several hours. In cases, where two single-stranded DNA were annealed, further purification was required. The mixture of single-stranded and double-stranded DNA resulting from the annealing process was separated by a DNA grade bio-

gel hydroxylapatite column using a BioLogic-LP chromatography system (both: Bio-Rad Laboratories, Hercules, CA). After initial equilibration in 10 mM NaH_2PO_4, 0.1 M NaCl, 50 µM Na_2EDTA at pH 7.0 on the column, elution of the individual components was achieved by a gradient from 10 to 200 mM NaH_2PO_4 in the otherwise same buffer at a flow rate of 1.2 mL/min. Fractions were collected and lyophilized, re-suspended in 1 mL of H_2O and desalted into gpm-buffer (100 mM NaCl) on a Sephadex G-25 column (Sigma-Aldrich, St. Louis, MO) using the BioLogic-LP system.

Table 3-1: Physical properties of nucleotide triphosphates (as listed by Fermentas)

Nucleotide	Name	MW [g/mol]	Extinction coefficient * [$Mol^{-1}cm^{-1}$]
dGTP	3'-deoxyguanosine 5'-triphosphate	507.2	ε_{253} = 13700
dATP	3'-deoxyadenosine 5'-triphosphate	491.2	ε_{259} = 15200
dCTP	3'-deoxycytidine 5'-triphosphate	467.1	ε_{271} = 9000
dTTP	3'-deoxythymidine 5'-triphosphate	482.1	ε_{267} = 9600
ddGTP	2',3'-dideoxyguanosine 5'-triphosphate	491.0	ε_{253} = 13600
ddATP	2',3'-dideoxyadenosine 5'-triphosphate	475.0	ε_{261} = 15200
ddCTP	2',3'-dideoxycytidine 5'-triphosphate	451.0	ε_{271} = 9300
ddTTP	2',3'-dideoxythymidine 5'-triphosphate	466.0	ε_{260} = 9600

* Values at pH 7.0, 20mM K_3PO_4

Verification of the product was done by MALDI-TOF mass spectrum measurements on a Voyager-Elite (PerSeptive Biosystems,) instrument in the negative reflector mode. The matrix for these DNA measurements contained 0.5 M 3-hydroxypicolinic acid and 0.1 M ammonium citrate. Purity was further verified by capillary gel electrophoresis (CGE), performed on a PACE 5500 instrument (Beckman Instruments, Fullerton, CA), using an eCAP ssDNA 100-R kit. Electrophoresis ran at 12 kV for 30 min and the UV absorbance of the DNA was monitored at 254 nm. The lyophilized material was dissolved in small amounts with the buffer used for the experiment.

Concentrations were determined by UV-measurements at A_{260nm} on a Cary 400 (Varian Palo Alto, Ca). Extinction coefficients were calculated by the nearest-neighbor method reported by Cantor at al (*93*).

Proton 1D and 2D magnitude COSY NMR spectra helped verify the duplex formation and presence of a single conformation. With a few exceptions the oligonucleotides were designed to be self-annealing by the formation of two hairpin loops allowing for a monomolecular system. The initial hypothesis for using two hairpin loops was, that the loop would bulge out sufficiently and upon selection of a proper duplex core length restrict the back and forward sliding of the DNA on the polymerase

interface. Special attention was given to design a sequence so that the single strands had little potential to form a self-complementary duplex but rather would anneal with the two hairpin loops. The online DNA analysis tool from IDT (Integrated DNA Technologies, Coralville, IA) was used to design the DNA sequence. The duplex cores were designed to contain a double gap that upon addition of ddNTP would be reduced to the preferred single gap (*9, 63*).

44-mer

The 44-mer sequence was based on the one used for the processivity assay reported by Oliveros et al (*9*). The duplex core of the sequence was shortened and the 6 base gap reduced to a GpG gap. A T_5 sequence was introduced twice to allow the formation of the hairpin loop. As shown in the same report, the 5'-phosphate allowed for more efficient gap filling and therefore was included in the design.

5’- $_P$ACG AC**T TTT T**GT CGT GGT ACT CAC TGT G**TT TTT** CAC AGT GAG TA – 3’

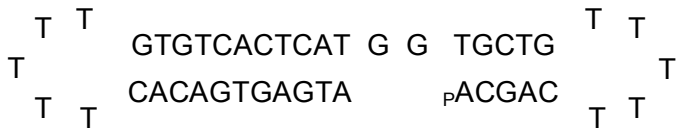


Figure 3-2: 44-mer duplex DNA, MW = 13,604.8 g/mol, ε_{260} = 414,800 $Mol^{-1}cm^{-1}$

36-mer

In order to reduce the duplex core size, a 36-mer with a high G●C content in the core was designed.

5’- $_P$GTC GG**T TTT T**CC GAC GGC GCA TCG **TTT TT**C GAT GCG -3’

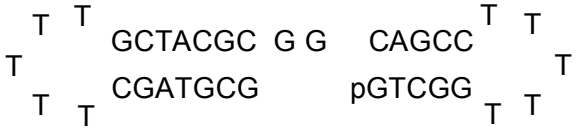


Figure 3-3: 36-mer duplex DNA, MW = 11,136.1 g/mol, ε_{260} = 323,800 $Mol^{-1}cm^{-1}$

34-mer

This 34-mer was designed to closer resemble the 44-mer, but had a shorter core and loop. Some modifications were required to fulfill all the mentioned requirements and

create a stable duplex core. Varani et al (*94*) and Gutell et al (*95*) described the formation of the shorter, but stable GAAA loop used. The gap sequence was changed to accommodate insertion of a dGTP.

5' –$_P$ACG ACA AAG GTC GTC CTC TCA CGA AAG CGT GAG A –3'

```
  A A  GCACTCT C C  TGCTG G A
  A G  CGTGAGA      pACGAC A A
```

Figure 3-4: 34-mer duplex DNA, MW = 10,549.8 g/mol, ε_{260} = 339,200 $Mol^{-1}cm^{-1}$

16-mer

The shortest double hairpin duplex used was a 16-mer. Searle et al (*96*) reported a high melting temperature of certain short intramolecular double hairpin structures containing a 5'-GAA loop. In analogy to this, the 16-mer with a CpC gap was designed. The short loop was forming a tight turn, which no longer was bulging out.

5' $_P$GCG AAG CCC CCG AAG G – 3'

```
    GCC C C  CGA
  A           A
    AGG      pGCG
```

Figure 3-5: 16-mer duplex DNA, MW = 4,981.2 g/mol, ε_{260} = 154,400 $Mol^{-1}cm^{-1}$

Template-Primer 17-9-mer, single-stranded DNA (ss-DNA)

To assess the impact of the hairpin loop, a regular template-primer DNA was used based on the 44-mer sequence with some minor modifications to ensure stable duplex formation. Sequences and physical properties of the 17-mer template and the 9-mer primer are listed in Table 3-2, along with some the ss-DNA used.

Table 3-2: Physical properties and sequences of used single stranded DNA

Name	Sequence	MW [g/mol]	ε_{260} [$Mol^{-1}cm^{-1}$]
17-mer template	5'-GCC GTC GGG CAC TCA CG-3'	5,172.4	153,400
9-mer primer	5'-CGT GAG TGC-3'	2,754.9	85,000
C-7-mer	$dC(pdC)_6$	1,962.3	50,600
G-7-mer	$dG(pdG)_6$	2,242.5	72,100
G-9-mer	$dG(pdG)_8$	2,900.9	92,300
A-12-mer	$dA(pdA)_{11}$	3,696.5	14,7400
T-12-mer	$dT(pdT)_{11}$	3,588.4	9,7800
C-6-A-12-mer	$dC(pdC)_5dA(pdA)_{11}$	5,431.6	18,9200

An additional template-primer $dC(pdC)_5dA(pdA)_{11}$: $dA(pdA)_{11}$ was prepared following the procedure described.

Experimental Set-Up and Data Analysis for Fluorescence Measurements

All fluorescence measurements were conducted in a Fluoromax-3 spectrometer (Horiba Jobin Yvon, Edison, NJ) at room temperature. After initialization, the spectrometer was calibrated using the internal calibration routine and the Raman spectrum of water was measured to confirm the calibration. Protein was added to a 1 mL cuvette at concentrations in the low micromolar range. Excitation spectra were run for pol X with λ_{ex} at 300, 307, 335 and 350 nm. According to these excitation spectra, optimal excitation of tryptophan was observed at 295 nm, while tyrosine showed maximum peak intensity at 275 nm. Using these excitation frequencies, typically two to three runs were performed and the data averaged. A delay of 30 s between each scan was used. To observe the maximal signal, the slit width for the emission was chosen at its maximum of 5 nm and the scanned wavelength range was from 305 to 400 nm for both excitation frequencies. The samples were scanned at a speed of 0.5 nm/s. Millimolar stock solutions of the oligonucleotide were prepared, so that only small amounts had to be added to the protein solution. Mixing was ensured by pumping the solution several times with a 200 µL pipette and continued stirring of the solution in the fluorometer.

Experimental Set-Up and Data Analysis for ITC Measurements

ITC measurements were performed on a VP-ITC system (MicroCal LLC, Northampton, MA). The integrated analysis software was Origin, version 7.0 (OriginLab, Northampton, MA). A cell feedback network measures the temperature difference of the exo- or endothermic reaction between the sample and reference cell (ΔT_1). The two coin shaped cells are mounted in an adiabatic environment and only connect to the outside through some narrow access tubes to minimize any heat exchange. A second sensor measures the temperature difference between the cells and the jacket (ΔT_2). As the measurement proceeds, the interactive feedback loop maintains ΔT_1 at zero. The power required to compensate for the thermal reaction is a direct measure of the total heat change during the reaction.

The syringe represents a high precision multi-functional device. Titrant is added to the fill port by means of another syringe and tube connections until the entire syringe is filled. The injector consists of a high precision step motor that pushes the plunger down in well-defined increments utilizing a screw mechanism. A paddle like broadening at the bottom of the syringe provides the necessary surface to mix the cell reactant with the titrant in the sample cell. Proper de-gassing of all used solutions is important to reduce the baseline noise during the detection.

Choosing the proper concentrations is important for a successful measurement. The formula in Eq. 3.14 provides some useful guideline for the selection of the reactant concentration.

$$C = K_a * M_{tot} * n \quad [3.14]$$

K_a is the binding constant, M_{tot} the total concentration of macromolecules in the cell at the start of the experiment and n the stoichiometry parameter. C is a unit-less constant and the value of interest. A value in the range of 5 to 500 for constant C is required to get good results for an ITC run. Outside that range, the binding is either too weak and the response is below the detection limit of the calorimeter. The sensitivity of the instrument used is about 0.1 µCal. Typical injections should produce an average heat of at least 3 to 5 µCal. On the other side, if the binding is too strong, all the ligand is bound immediately, producing the same degree of heat for each titration step. Upon saturation, there is an immediate change and no heat is produced any longer. The transition in this case is too sudden and the data cannot be analyzed any longer. Dilution or a competitive binding scheme may extend the measurable range in this case.

The concentration for the ligand added by the syringe is chosen by the following estimation (*85*):

$$[L_{tot}] = 7 * n * C \quad [3.15]$$

Estimations in Eq. 3.14 and 3.15 are only useful if K_a and n are known. Otherwise best guesses are required to run some scouting experiments before the parameters can be optimized.

General Measurement Procedure

Buffer solutions containing 20 mM PIPES, 20 mM $MgCl_2$, 1 mM BME, 0.02% NaN_3, and either 300 mM or 500 mM NaCl at pH 6.5 were used. Pol X samples were equilibrated multiple times using Amicon 10k spin filters. The nucleotides and oligonucleotides were dissolved in the same buffers and the pH between the two samples was carefully monitored. Freshly degassed buffer was used to dilute the stock solutions, samples were prepared immediately prior to the measurement and the final solutions were degassed again. Reactant cell and syringe were washed thoroughly with water and degassed buffer. A buffer-buffer titration was performed before each measurement series to validate proper system performance.

A 3 mL syringe containing a long, blunt end needle was used to wash and fill the reactant cell. The needle was initially fully inserted to the bottom of the cell and pulled out slowly as solution was injected, ensuring that no bubbles were trapped in the cell.

To fill the titration syringe, the ligand solution was placed into a small, tall vial. The paddle of the syringe was lowered into the vial to barely touch the bottom. With the aid of a temporary 1 mL syringe connected to the fill port, the ligand was pulled into the titration syringe. Once full, the plunger of the titration syringe was activated and lowered just enough to pass the fill port to close off the system. Now the plunger was pushed completely down and up twice, emptying the ligand into the small vial. This was done to ensure proper functionality of the assembly and relief any pressure buildup or bubble formation.

The temporary syringe was disconnected, the needle inserted into the reactant cell and the entire syringe rotated at 300 rpm for proper mixing during the titration. The following parameters were set for these titrations: 40 injections, 25 ºC run temperature, reference power of 5 µCal/s, initial equilibration delay of 1200 s with the fast equilibration mode and auto titration activated, 10 s injection time, resulting in a 0.5 µL/s ligand addition, a delay of 180 s in-between each injection. The titration was run under computer control. More detailed instructions for operating the calorimeter and its computer controls can be taken from the instruction manual (*97*).

Data analysis was performed with the integrated software, Origin 7.0. The titration points were fit to quadratic equation for single binding sites, as described for the

fluorescence measurements in the introduction. Results included the binding constant, reaction enthalpy and entropy values with their respective errors. A detailed description of the software and data analysis is provided by the manufacturer (*98*).

Results

Fluorescence Binding Studies

Binding Study with a 44-mer and 36-mer Duplex DNA

One of the first fluorescence measurements was designed to get a better understanding of the interaction between pol X and DNA under different sample conditions. Since multiple pol X constructs were available, it was of interest to see, whether there was any detectable difference between them. The pol X_{orig} was not considered in this study, because of its long N-terminal leader peptide, making it less useful for NMR studies. Only the pol X_{iz} and pol X_{gpm} constructs were compared with each other. The main difference between the two constructs consisted in the presence of a C-terminal His-tag on pol X_{iz} while pol X_{gpm} did not contain any additional residues.

Additionally, this first test series also addressed the question of complex stability at various salt concentrations. Previous NMR work (*46, 81*) on free pol X_{gpm} was conducted at 500 mM NaCl in order to retain protein stability. On the other hand, protein-

DNA complexes become less stable at higher salt concentration, which was a concern. Therefore the salt dependence of the dissociation constant pol X had to be investigated. One of the advantages of fluorescence measurements is the increased sensitivity compared to the NMR experiments. Protein solutions with 10 to 100 times lower concentrations can be used for these measurements. At these concentrations, pol X remained stable for a short time in the gpm buffer containing as little as 50 mM NaCl, allowing the measurement of the dissociation constant between pol X and oligonucleotides. Addition of the oligonucleotide to the protein sample at low salt concentration had to be done slowly. When added too fast, the concentrated DNA stock solution caused some local precipitation, which dissolved after some time.

The positions of the tryptophan and tyrosine residues were known once the structure of free pol X was published (*46, 47*). Figure 3-6 shows the position of these two residues, purple for tryptophan and orange for tyrosine. The structure shows, that the tryptophan is positioned in the palm of the protein, in the short loop connecting strand β9 and β10, close to α-helix αA on the surface of the molecule. Although at first glance located in a remote place of the protein, it is directly connected to the β-sheet, that contains the catalytic active site and interactions with the nucleotide or oligonucleotide might still alter the fluorescence of this residue sufficiently.

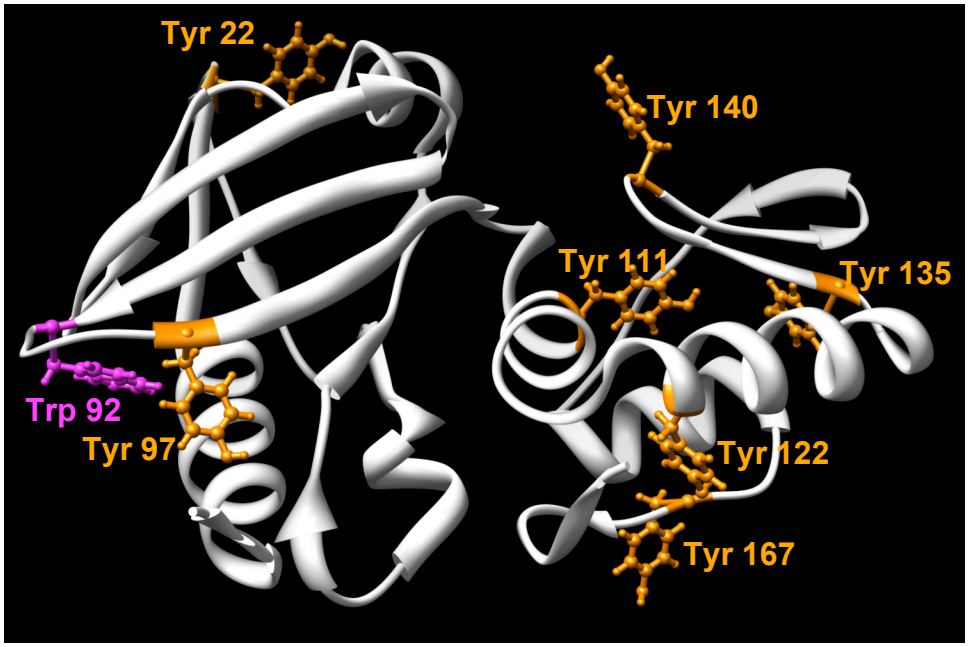


Figure 3-6: Position of tryptophan and tyrosine in solution structure of free pol X, as published by Maciejewski et al (*46*), (PDB code 1JAJ). Highlighted in purple is the only tryptophan residue, the seven tyrosine residues are colored in orange.

Five of the seven tyrosine residues are located in the C-terminal subdomain of the protein where interactions between the protein and the nucleotide or oligonucleotide were expected, hence could induce a strong change in the fluorescence signal. Tyrosine 97 is facing towards the nucleotide binding site and its signal was expected to be impacted strongly as well. Only tyrosine 22 appears to be removed from the sites where most of the interactions were expected. Although individual tyrosine signals cannot be distinguished, their distribution throughout the molecule in the putative DNA and nucleotide-binding site raised the anticipation of a strong impact on these signals in the

fluorescence spectrum. It is noteworthy, that with the exception of tyrosine 111, all these residues are to some degree exposed to the surface of the molecule where a high degree of solvent accessibility was expected.

Figure 3-7 shows two representative emission titration curves. The consecutive measurements were done in a sample containing 8.2 µM pol X_{gpm} in the presence of 500 mM NaCl. The 44-mer DNA duplex was titrated to the polymerase. The fluorescence emission spectra were measured after excitation at either 275 nm or 295 nm. A blank measurement, which included only the buffer was subtracted from all spectra, but is also shown in the graph.

Comparing the two emission spectra series, one can conclude: 1) The spectra excited at 295 nm showed a slight shift of the main tryptophan emission wavelength from 347 nm to 342 nm. This new maximum showed up immediately after the first addition of DNA as a second peak, while the original peak remained as a shoulder throughout the measurement. This indicated, that a portion of the tryptophan moved towards a slightly less polar environment. Even at an excess of 5 mol fractions though, both forms appeared to remain present, without any major change in their ratio. One explanation

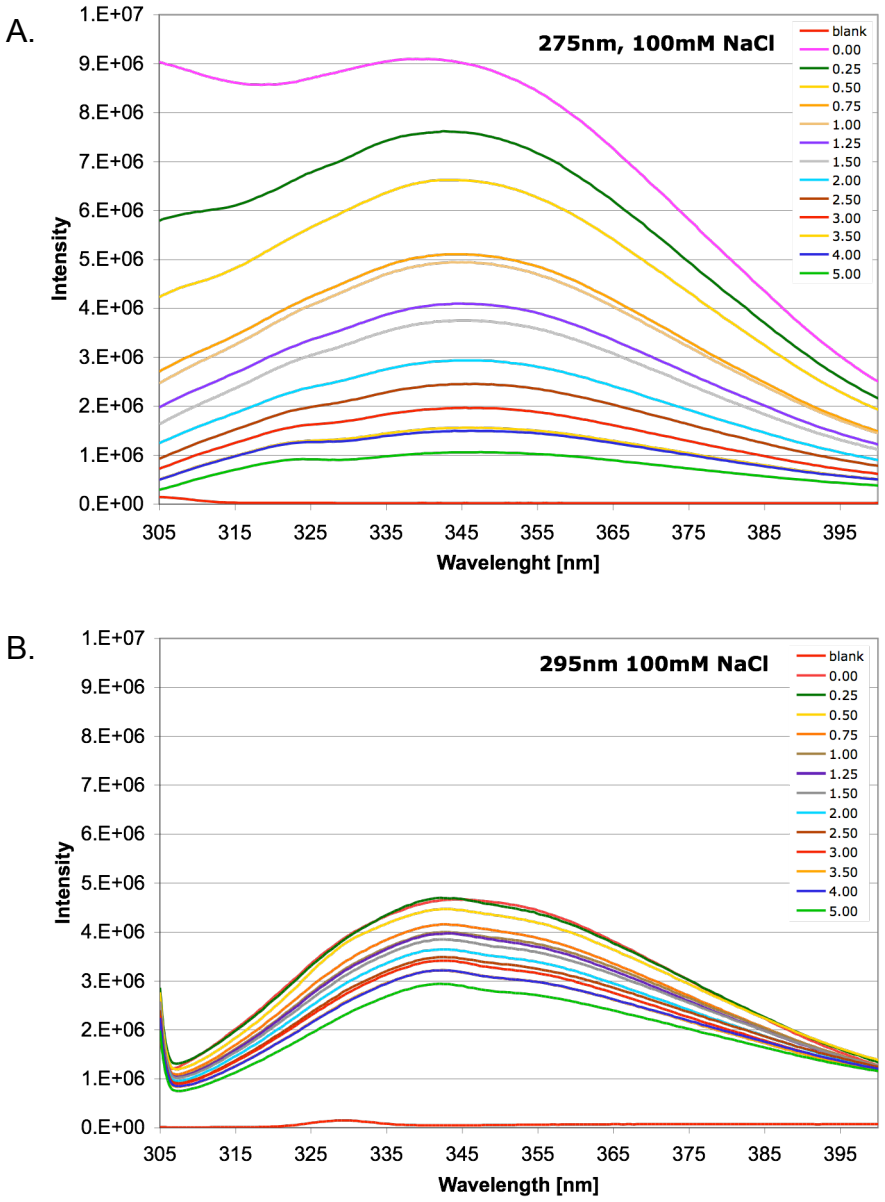


Figure 3-7: Representative fluorescence emission spectra from a titration of pol X_{gpm} with 44-mer DNA. **A.** Excitation at 275nm and **B.** 295 nm. Each figure shows an overlay of 20 titration spectra from pol X_{gpm} at 8.2 μM, and 44-mer DNA at 3.1 mM. The buffer contains 20 mM PIPES, 20 mM $MgCl_2$, 2 mM BME, 0.02% NaN_3, and 500 mM NaCl. The most intense spectrum represents pol X_{gpm} without DNA. The signal was quenched as DNA was added to the polymerase. The legend shows the amount of DNA added in mol fractions: blank, which was subtracted from each spectrum, 0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 2.00, 2.50, 3.00, 3.50, 4.00 and 5.00.

for this would be that upon addition of DNA, the protein changed its conformation but remained in constant exchange throughout, which was reflected in the two wavelengths present. 2) The overall signal intensity and the relative signal change of the emission spectrum excited at 295 nm was weaker. This could be expected, since 295 nm was in the tail of the tryptophan excitation spectrum. The advantage though was the selective excitation of only this residue. For the determination of the binding constant, the spectrum intensity was taken at the maximum emission at 342 nm. Because the signal intensity of the intrinsic tryptophan signal in these spectra was weak and the change between spectra was smaller, it was more challenging to obtain a reliable fit using these emission spectra. 3) There was a very strong change in the signal intensity for tyrosine at 306 nm, when excited at 275 nm. This was an anticipated change, since there were many tyrosine residues in the anticipated binding pocket.

The conclusions from these comparisons were, that the environment of the tryptophan was changing as DNA was added. It was possible that the protein was in multiple conformations and that change was slow relative to the fluorescence time scale. The tryptophan was thought to be reversibly moving between a rather solvent exposed, polar environment, and a slightly less polar position, most likely pointing towards the core of the palm.

Binding constants of two pol X constructs, pol X_{iz} and pol X_{gpm} with the 44- and 36-mer duplex DNA, containing a two base GpG gap, are summarized in Table 3-3. The spectrum intensity at the maximum emission was analyzed by a quadratic fit. As pointed out above, the tryptophan signal in the emission spectrum after excitation at 275 nm gave the strongest response and was therefore generally used for analysis of the K_d. The fluorescence signal was almost completely quenched following the addition of one mol fraction of DNA.

Table 3-3: Binding constants for two different pol X constructs with 44- and 36-mer at various NaCl concentrations.

Construct	NaCl concentration	DNA	K_d λ_{ex} = 275 nm λ_{em} = 342 nm	Error
Pol X_{iz}	500 mM	44-mer	8.5 µM	± 0.7 µM
Pol X_{gpm}	500 mM	44-mer	8.7 µM	± 0.4 µM
Pol X_{gpm}	100 mM	44-mer	5.0 µM	± 0.7 µM
Pol X_{gpm}	50 mM	44-mer	2.1 µM	± 0.5 µM
Pol X_{gpm}	100 mM	36-mer	8.3 µM	± 1.0 µM

Over all these values showed, that the DNA binding to pol X was in the low micromolar range. No substantial difference in binding was detected for the two constructs, indicating that the additional C-terminal His-tag did not interfere with the DNA

binding at this level. The varying salt concentration did have a small affect on the binding and showed a slightly tighter bound DNA to the polymerase at lower salt concentrations. This would be expected, but the small difference should not pose a significant problem even if working at 500 mM NaCl buffers. The dissociation constant for the 36-mer had about the same value as the high salt value for the 44-mer. Since this measurement was done in only 100 mM NaCl, it can be expected that the binding of the 36-mer at 0.5 M NaCl is even smaller then the one seen for the 44-mer. The shorter core size of the 36-mer provided a smaller binding interface and most likely accounted for the decrease in binding. It could be conclude, that neither the salt concentration in the range of 50 to 500 mM NaCl, nor the change from a 44-mer to a 36-mer double hairpin DNA had a major impart on the binding to pol X.

Binding Study with a 34-mer DNA Duplex Containing Two Hairpins

Another series of experiments tested for the binding in the presence of the 34-mer double hairpin with a two base pair CpC gap and a reduced hairpin loop. According to the report from Varani et al (*94*) and Woese et al (*95*), the nucleotide sequence GAAA or GCAA formed a stable tetra-loop. Although the 44-mer with the penta T-loop showed sufficient binding as determined by the fluorescence, it did not provide a suitable NMR

spectrum, as will be shown in this Chapter IV. One rationale for the bulky loop was, that it might hinder the DNA from moving back and forth on the polymerase by "locking" it down with these loops. Conversely, the NMR data suggested that the loops might have interfered with the binding, which led to the exploration of DNA with a smaller loop, as was present in the 34-mer.

Another difference was the presence of a CpC gap in the template instead of the GpG gap used previously. Literature reports (*61*) became available describing that pol X would bind preferably and tighter to the deoxyguanosine triphosphate, which prompted the change to the CpC gap and also made considerable modification of the DNA sequence necessary. While the previous sequence was based on the one reported by Pelletier for his structural work on the ternary complex on pol β (*40*), this new sequence was chosen to resemble the sequence originally used by Oliveros et al (*9*) for his activity studies on pol X.

The polymerase concentration was reduced from 8 μM to 2 μM in order to minimize the inner filter effect. A maximum excess of two mol fractions of DNA was added to a freshly prepared pol X_{gpm} solution. Looking at the results in Table 3-4 immediately suggest, that the 34-mer binds about ten times stronger to the polymerase then the above described 44- or 36-mer. Because modifications in the sequence were

the only major change, the tighter binding was attributed to them. Whether the tighter loop or the change in the core sequence was the cause, could not be determined. An attempt to titrate additional dideoxynucleotide triphosphate did not yield sufficient additional quenching of the signal.

Table 3-4: Binding constants between pol X_{gpm} and a self-annealing 34-mer, formed by two GAAA tetra-loops containing a CpC gap at the center of the double stranded core. Two μM of pol X_{gpm} was titrated with 34-mer, 100 μM, up to a final concentration of 4 μM.

Construct	NaCl concentration	DNA	K_d λ_{ex} = 275 nm λ_{em} = 342 nm	Error
Pol X_{gpm}	500 mM	34-mer	0.619 μM	± 0.176 μM
Pol X_{gpm}	500 mM	34-mer	0.735 μM	± 0.116 μM

Binding Study with a 16-mer DNA Duplex

Encouraged by the positive effect seen for the 34-mer, the DNA sequence was further shortened. A shorter DNA was also advantageous to the NMR work especially for a potential structural elucidation. Besides shortening the core, the hairpin loops were further reduced by an additional base as suggested by (*95, 96*), yielding a 16-mer sequence.

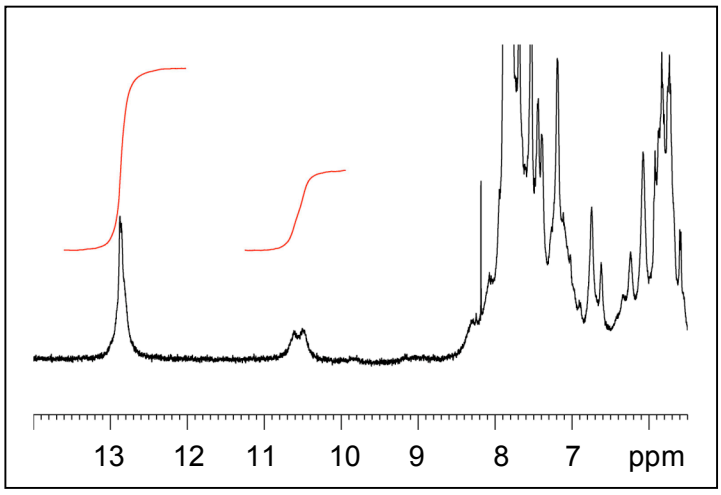


Figure 3-8: Proton NMR spectrum at 500 MHz of 16-mer CpC gapped DNA, 288 K, 32 scans, binominal water suppression with node at a distance of 10 ppm from water.

According to that report, such a loop would still be stable, but would no longer form a bulky hairpin. Instead, a rod-like structure would be formed. It was advantageous to keep hairpins in the sequence since they provided additional stability and allowed for an extended Watson-Crick double strand formation between the short primer and template sequences. A 1D ^{1}H NMR spectrum, displayed in Figure 3-8, showed the expected imino signals at 12.86 ppm and two broader, upfield shifted signals at 10.63 and 10.52 ppm, which could be attributed to the hydrogen bonded guanine imino protons from the G-A base pair in each respective hairpin (*96, 99*). The intensity ratio was 2:4 and reflected correctly the two imino hydrogen bonds for the G-C pairs on either side of the gap. The presence of all imino protons at the proper chemical shift in the NMR

spectrum provided strong evidence that a stable duplex containing two hairpins was formed.

Table 3-5: Binding constants of pol X_{gpm} with DNA and nucleotides. A self-annealing 16-mer, containing two GAA hairpin loops and a CpC gap as well as nucleotides were used. Pol X_{gpm}, (2μM), was titrated with 16mer and ddGTP, both 100 μM, up to a final concentration of either 4 μM or 10 μM before adding ddGTP. Pol X_{gpm} (8.5 μM) was used for the dGTP titration and dGTP at 125 μM was added up to 2 mol fractions.

Construct	NaCl concentration	1st addition	2nd addition	K_d λ_{ex} = 275 nm λ_{em} = 306 nm	Error
Pol X_{gpm}	500 mM	16-mer		2.3 μM	± 2.5 μM
Pol X_{gpm}	500 mM	16-mer	ddGTP	0.48 μM	± 0.44 μM
Pol X_{gpm}	500 mM	ddGTP		0.53 μM	± 0.81 μM
Pol X_{gpm}	500 mM	ddGTP	16-mer	0.47 μM	± 0.48 μM
Pol X_{gpm}	500 mM	dGTP		1.2 μM	±0.2 μM

Table 3-5 shows a summary of the fluorescence measurements using this 16-mer DNA fragment. As was the case for the other two oligonucleotide sequences, a dissociation constant in the low micromolar range was found. The values for the binding constants laid in-between the 44- or 36-mer oligonucleotides and the 34-mer. This could be an indication that the core of the 16-mer had become too short and no longer bound as tight as the longer 34-mer. The longer 44- and 36-mer contained the bulky loop,

which was proposed to weaken the binding interaction. Addition of 5-fold excess on 16-mer confirmed, that saturation was achieved.

In addition to the 16-mer, the behavior of the ternary complex was tested. After 16-mer was added at 5-fold excess to pol X, a second titration followed into the same solution. This time the complementary dideoxyguanosine triphosphate was titrated. Saturation was seen at two mol fractions of ddGTP. As mentioned previously, the emission spectrum after the first titration with the 16-mer was already weak, but in this case additional changes could be observed after ddGTP addition, when the emission spectrum of tyrosine at 306 nm was monitored instead of the tryptophan emission. Fitting the data to the quadratic equation resulted in a K_d of 0.2 to 0.9 μM, with an average of 0.48 μM. Given the small changes of the fluorescence signal and the potentially competitive nature of the oligonucleotide/nucleotide binding events, this value could only be considered as a rough estimate.

The same was true for the reverse experiment where ddGTP was added first followed by the 16-mer. The overall binding constants for the ternary complex at the end of both titration schemes were comparable but smaller then the one for the oligonucleotide alone which would indicate that the nucleotide provided for a tighter binding of the ternary complex.

Binding Study with a Primer-Template DNA Duplex

A final study described here involved a primer-template DNA fragment instead of a double hairpin DNA. The sequence, as described in methods, was chosen so that the individual oligonucleotides would refrain from self-annealing and the resulting 3' extension site of the template would contain a GpG base. The construct was a 17-9-mer template-primer. The experimental design was slightly modified and three independent titrations were performed. The first titration probed the binding behavior of the primer-template DNA with pol X. In two subsequent titrations, either dideoxycytosine triphosphate or dideoxyguanosine triphosphate were added up at a two-fold excess of the polymerase concentration and the template-primer was then titrated to each of these solutions. Table 3-6 summarizes the result of these experiments. This primer-template oligomer was binding tight to the polymerase. Its dissociation constant was comparable to the one measured for the 34-mer, but stronger than those seen for the 44-mer, 36-mer and 16-mer. Interestingly the presence of the dideoxynucleotides did not result in a change of the dissociation constant, as was the case for the 16-mer. It could be argued that the template-primer already bound with similar strength to the polymerase so that the presence of the nucleotide did not contribute to a measurable enhancement anymore.

Table 3-6: Binding constants of a template-primer DNA. The individual oligomers are 17 bp and 9 bp long. The first two complementary bases on the template following the 3' end of the primer were two guanine bases. Pol X_{gpm} at 1.5 µM was titrated with the oligomer, 100 µM, to a final concentration of 2 µM, in the absence and presence of dideoxynucleotides (3 µM). Averages of 3 determinations are listed.

Construct	NaCl concentration	1st addition	2nd addition	K_d λ_{ex} = 275 nm λ_{em} = 306 nm	Error
Pol X_{gpm}	300 mM	none	17-9-mer	0.28 µM	± 0.03 µM
Pol X_{gpm}	300 mM	5 mf ddCTP	17-9-mer	0.30 µM	± 0.12µM
Pol X_{gpm}	300 mM	5 mf ddGTP	17-9-mer	0.41 µM	± 0.02µM

Binding Studies by Isothermal Titration Calorimetry

The NMR titration studies described in the following chapter revealed an unexpected behavioral difference between individual nucleotides. To independently confirm these findings, ITC measurements on the binary pol X complex were performed, including the purines deoxguanosine triphosphate (dGTP), adenosine triphosphate (ATP) and deoxyadenosine triphosphate (dATP), as well as the pyrimidines deoxycytidine triphosphate (dCTP) and thymidine-triphosphate (TTP). Multiple titrations were performed with all nucleotides mentioned above. The molar excess of nucleotide compared to the polymerase was either 2.5 or 3-fold.

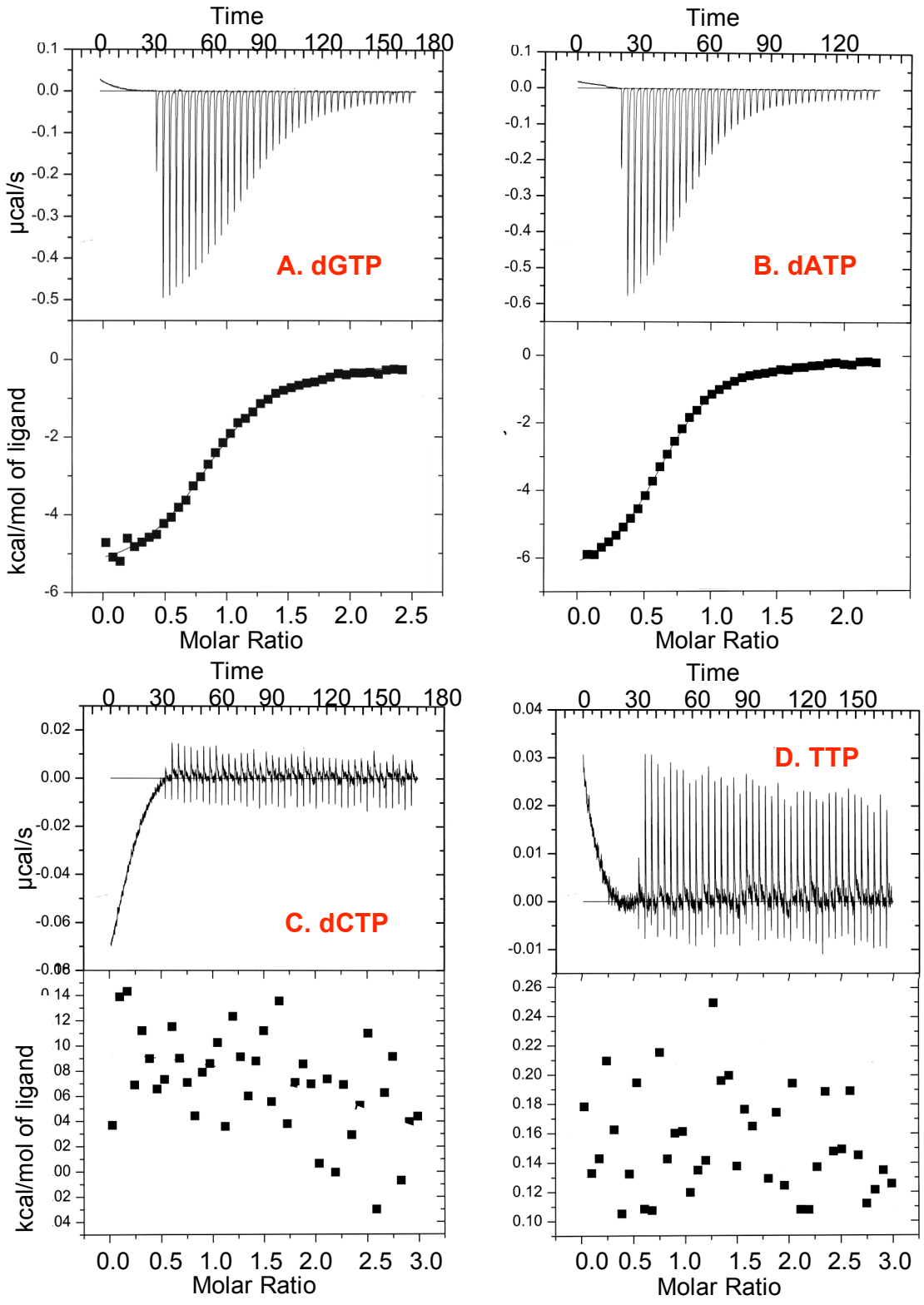


Figure 3-9: ITC measurements of pol X_{gpm}, 40 µM, with all four deoxynucleotides at 600 µM in the presence of 300 mM NaCl, as described in the text.

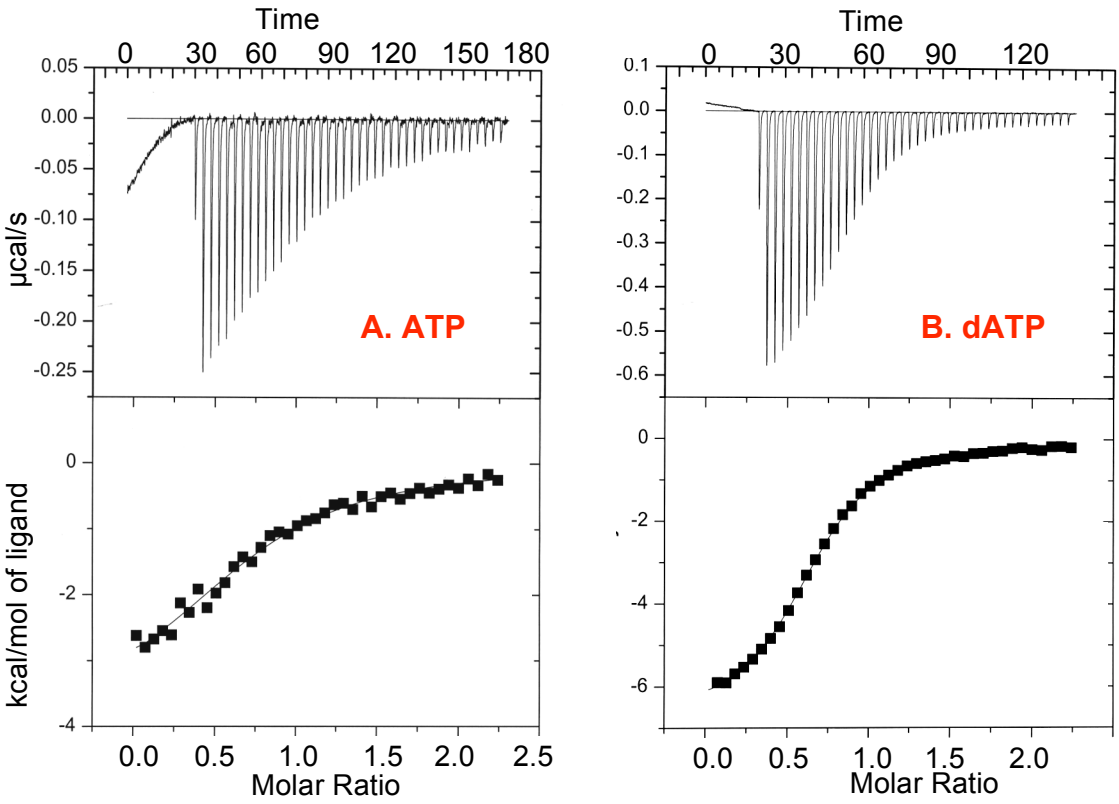


Figure 3-10: Comparison of the binding between pol X_{gpm} and **A.** adenosine triphosphate and **B.** deoxyadenosine triphosphate

Table 3-7: Summary of ITC measurements. Dissociation constants and thermodynamics for binding of individual nucleotides with pol X_{gpm} at 25 ºC. ΔS is calculated from ΔH and K_a. No analysis was possible for dCTP and TTP.

Nucleotide	NaCl concentration	Pol X [µM]	ligand [µM]	K_d [µM]	Error [µM]	ΔH [cal/mol]	Error [cal/mol]	ΔS [cal/mol/deg]
dGTP	300mM	40	600	2.7	± 0.5	-4610	± 66	10.0
dATP	300mM	40	600	3.2	± 0.1	-6924	± 62	1.9
dCTP	300mM	40	800	n/a	n/a	n/a	n/a	n/a
TTP	300mM	40	800	n/a	n/a	n/a	n/a	n/a
ATP	300mM	40	600	12.3	± 1.1	-3981	± 185	9.1
dGTP	500mM	40	600	3.7	± 0.2	-5622	± 58	6.0

Figure 3-9 displays a representative ITC titration curve for each nucleotide. It was immediately apparent, that under comparable conditions, the ITC study showed distinct differences between the purines and pyrimidines. Even though the nucleotide concentration was increased for the pyrimidines, no analyzable signal was detected during these measurements. The deoxypurines compared very well with each other and confirmed findings from the NMR titration as well (Chapter IV). The binding constant of 2.7 ±0.5 μM for dGTP at 300 mM NaCl was within the error of the one found by NMR, where an average binding constant of 2.4 ± 0.7 μM was reported. A similar comparison could be made for the ATP, where the NMR measurements yielded an average binding constant of 10.2 ±1.0 μM compared to the ITC determination of 12.3 ±1.1 μM. The comparison between the ATP and dATP showed a substantial difference in the binding. A similar trend was observed for dGTP and ddGTP during the fluorescence titrations and was reported in Table 3-5. There the dissociation constant for dGTP was determined to be 1.2 μM versus 0.5 μM for ddGTP, which was about a factor of three lower. This might indicate that pol X exhibits a differential binding behavior not only between the purines and pyrimidines, but also for nucleotides that are alike but have a different degree of hydroxylation of the sugar. Oliveros (*9*) has reported that ddNTP's more readily inhibited

DNA synthesis than was previously observed for pol β. The enhanced binding of the dideoxyguanosine triphosphate to pol X might explain this behavior.

Additionally, the interaction between pol X and dGTP at two salt concentrations was examined by ITC. As shown in Table 3-7, there was a small change in the binding when modifying the salt concentration. At 300 mM NaCl, a K_d of 2.7 ±0.5 μM was measured, while it was 3.7 ±0.2 μM at 500 mM salt. Although the change was very small, it was in the expected direction.

Conclusions

Fluorescence Measurements

With the help of these fluorescence measurements, it was possible to gain some important insight into the DNA-protein binding. 1) The tryptophan emission spectrum displayed two peaks upon addition of oligonucleotides. Therefore a conformational change affecting the tryptophan must take place moving it towards a more non-polar environment. While the effect of this change was clearly visible, the change appeared to be modest. The tryptophan was expected to remain exposed to the surface, but experienced some additional hydrophobic shielding. 2) Although the tryptophan residue was somewhat remote from the putative binding sites, clearly detectable responses were

measurable upon complex formation. 3) The DNA dumbbell design was helpful to retain a stable DNA duplex and minimize the number of macromolecules in the system, but care must be taken in the selection of the proper design of the hairpin loop. While the hope was to “lock” the DNA on to the polymerase with the aid of bulkier hairpins, they appeared in some cases to be in the way and reduce the protein-DNA interaction. The shorter tetra- and tri-loops appeared to be favorable. 4) Monitoring the intrinsic emission proofed in times challenging and generally small changes in the fluorescence intensity were encountered, leading to larger errors for individual data fits. Although briefly considered, the use fluorophores was dismissed by the concern of introducing non-specific interactions into the pol X-DNA complex system, which was barely known in the initial stages of this work. With today’s knowledge, the use of fluorophores would be strongly considered to detect more distinct signal changes in a lower concentration regime.

ITC Measurements

The dissociation constants for dNTP with pol X were also determined by Mullen and coworkers (*46*). In their work, they report K_d values of 1.6, 6.4, 11.9 and 4.0 μM for the displacement of dGTP, dATP, dTTP and dCTP, respectively. A similar buffer was

used: 5 mM PIPES, 5 mM $MgCl_2$, 125 mM NaCl at pH 6.5. No reducing agent was used and a lower salt concentration of 125 mM NaCl compared to the 300 to 500 mM NaCl used for these ITC measurements. The measurements were done by a competition assay based on the fluorescence measurement of a fluorescent probe 2',3'-trinitrophenyl-ATP-Mg^{2+} (TNP-ATP- Mg^{2+}). This probe bound to pol X with high affinity in a 1:1 stoichiometry and was competitively replaced by the four deoxynucleotides. Regardless of these experimental differences, the measurements for the purines yielded similar results. The K_d values determined by ITC were 2.7 and 3.2 μM for dGTP and dATP, respectively, at 300 mM NaCl. It was interesting to see, that ITC measurements worked well for the purines, but not for the pyrimidines, even though the reported (*46*) dissociation constants between the two classes of nucleotides do not differ much. Indeed, for all three methods used, the purines yielded a consistent result, while the pyrimidines did not. The conclusion that methodical differences lead to this divergence represents one possibility. Such differences were given by the reaction competition during the fluorescence measurement. For the ITC measurement, a potential cancellation of exo- and endothermic reactions might occur during the titration with pyrimidines, while a distinct exothermic binding reaction was present for the purines. During NMR titrations, individual amides were directly monitored. A distinct difference

between the purines and the pyrimidines was also reported for some biochemical assays (*59, 63*). The K_m and k_{cat} for example was significantly different for pyrimidines, when measuring these kinetic parameters for template-primer and single gap substrates with the correct base insertion by pol X (*63*). Tsai and coworker (*59*, *61*) also reported such a difference and saw preferred incorporation of dGTP opposite G, disfavoring the correct dCTP incorporation and instead produced the G:G mispair with exceptionally low fidelity.

While the underlying cause remains elusive, the data suggests that the pyrimidines might miss out on some important electrostatic interactions, possibly some stacking interactions of the smaller aromatic ring system or hydrogen bonds that are geometrically more tailored towards the purines, supporting the notion of a polymerase with a tailored specificity towards oxygen damage repair (*63*).

CHAPTER IV

BINDING STUDIES USING NMR

Introduction

NMR provides a powerful tool to study interactions between molecules, often at an atomic resolution in solution. A short introduction to NMR will be given in Chapter V and attention here will be focused on specific issues relating to the binding studies.

Interaction studies can be of various nature, for example a ligand might induce some conformational changes to a protein, form intermolecular complexes of various stabilities or initiate chemical reactions. Depending on the interaction, different NMR parameters are observable, such as chemical shifts (δ), coupling constants (J) or relaxation rates (T_1 and T_2), and the proper parameters have to be individually chosen. Often, changes in the chemical shifts of the amide nuclei are observed, since they are sensitive to their chemical environment and small changes during a titration can cause a substantial change in δ.

Many exchange processes can be observed as well. Three exchange rates can be distinguished and are differentiated in the case of chemical shift into slow exchange,

with an exchange rate k << | δ_A - δ_B |, intermediate exchange k ≈ | δ_A - δ_B | and fast exchange k >> | δ_A - δ_B |. Typical chemical shift differences are on the order of 10-500 Hz. These chemical shift differences are field dependent so processes that appear to be in fast exchange at low fields can sometimes be found to be at intermediate or fast exchange at higher field spectrometers. This is particularly useful when encountering intermediate exchange, which then can be shifted depending on the field strength.

The spectral appearance also changes depending on the exchange rate and will be discussed on the example of a small ligand interacting with a protein. The principal is also true for larger ligands such as DNA fragments or peptides, but in that case the molecular weight of the overall complex might be an additional factor. If the molecular weight of the complex is substantially greater, the rotational correlation time (τ_c) will be increased causing additional line broadening; hence reduced peak intensity.

Figure 4-1 shows a schematic representation of the most commonly observed interactions. Figure 4-1 (A) represents the fast exchange case. A single resonance line for both populations (A and B) is observed (δ_{obs}) at the chemical shift representing the weighted average of the individual species:

$$\delta_{obs} = \alpha\, \delta_B + (1 - \alpha)\, \delta_B \qquad [4.1]$$

where α is the fraction of the population of species A at equilibrium. The equilibrium

changes during the course of the titration, therefore the chemical shift of the peak shifts according to the above equation. The line broadening is negligible and remains comparable to the natural linewidth.

The other extreme is the slow exchange represented in Figure 4-1 (C) Here the lifetime of each species is sufficiently long to observe an NMR signal at either position. In this case, the exchange rate is much smaller than the frequency difference between the sites, and the two lines are only minimally broadened by the exchange. A special situation of this case presents itself when the two species interact strongly forming a tight complex, or in the extreme situation undergo a chemical reaction, where no equilibrium between the two species is present anymore, but the chemical shift is a result of the newly formed product, as might be seen in a cross-link formation. The newly formed product or tight complex creates a different environment for some of the nuclei, causing their chemical shifts to move sometimes far away from the original positions. As more ligand is added to the protein, the intensity of the initial peak is diminishing, while the product peak grows at the same rate. No spectral changes will be observable passed the stoichiometric addition of ligand.

Figure 4-1 (B) shows the situation where signals reflect an intermediate exchange rate. This situation is typically not desirable, since the signals can be

substantially broadened, sometimes to the point where they are no longer visible. Depending on whether the exchange rate is less than or greater than the frequency difference between the to populations, either one or two signals may be observed.

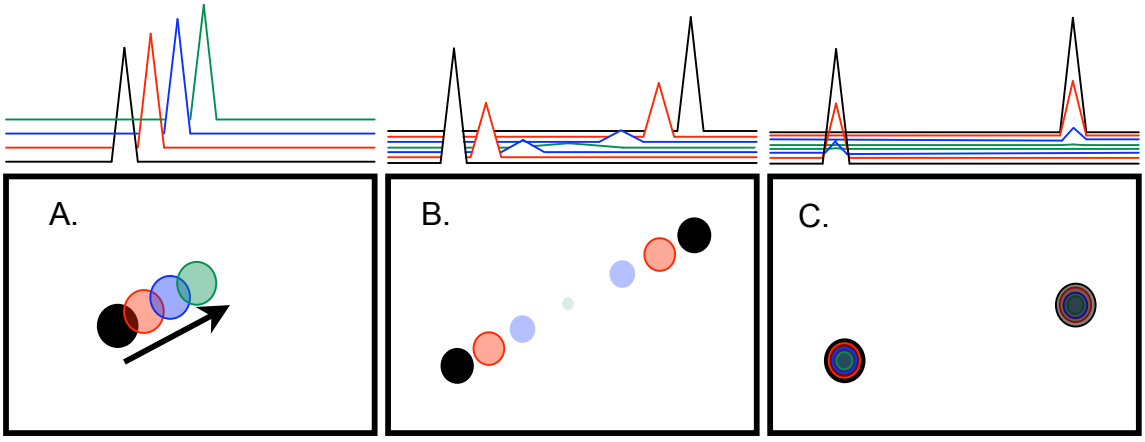


Figure 4-1: Schematic representation HSQC spectrum changes due to different exchange rates in the presence of a small ligand. **A.** Fast exchange, **B.** intermediate exchange and **C.** slow exchange. Chemical reactions or tight binding systems have a similar appearance.

The determination of the equilibrium dissociation constant K_d, using NMR as a method of detection, requires a series of two dimensional (2D) ^{15}N-HSQC spectra at various ligand concentrations. Each individual correlation-peak in the HSQC spectrum represents the nitrogen and proton frequencies of the amide nuclei of individual amino acid residues in the enzyme. Therefore these studies provide detailed insight into

structural and kinetic features at atomic resolution, provided the correlation peaks are assigned, which will be the aim of Chapter V.

In these 2D spectra, the chemical shift change has to be quantified to analyze the spectrum and extract the binding constant. Since peaks do not shift exclusively in either the X- or Y-axes of the 2D spectrum, it is common practice to use chemical shift changes that are calculated as the resultant of both vectors as follows:

$$\delta_i = \sqrt{\left(\delta_{HN}\right)^2 + c * \left(\delta_N\right)^2} \qquad [4.2]$$

where (δ_{HN}) and (δ_N) are the chemical shift changes in ppm in either the proton or nitrogen axes, respectively. The scaling factor *c* of 0.2 accounts for the different chemical shift ranges in the proton (5 ppm) and nitrogen (25 ppm). The least square fit analysis described Chapter IV is used to calculate the K_d, based on the protein- and ligand concentrations and δ_i.

The situation is different when the system is in slow exchange or tight binding (Figure 4-1, C). Here, the intensity of the original peak will be monitored as it decreases continuously, while a new peak grows into the spectrum at an unknown position. Therefore, it is often difficult or impossible to identify the new correlation unambiguously. These substantial up- or downfield shifts are a result of the ligand interacting with the protein, significantly changing the environment for some amides and their correlation

peaks. If the assignment for both sets of peaks is known, than the equilibrium constant can be determined from either peak, giving some redundancy. Overlapping signals allow often the analysis of only one or the other peak.

Materials and Methods

Experimental Set-Up and Data Analysis of NMR Measurements

Gpm-Buffer

The NMR buffer contained 20 mM PIPES, 20 mM $MgCl_2$, 10 mM DTT, 0.02% NaN_3, and 500 mM NaCl. Variations to these standard conditions were sometimes used and will be pointed out at the appropriate place. Typical variations include the salt concentration (NaCl) or the substitution of DTT with BME or deuterated PIPES and DTT for some NMR measurements.

NMR Samples

In order to perform these NMR studies, ^{15}N isotopically labeled enzyme had to be prepared in milligram quantities, as described in Chapter II. Purified pol X_{gpm} was available lyophilized. The enzyme was taken up into an NMR buffer that initially did not contain DTT. This modification was necessary due to the absorbance of DTT at A_{283} Pol

X contains one tryptophan residue and its molar extinction coefficient is 15,440 $M^{-1}cm^{-1}$ at 280 nm (*100*).

The following experimental scheme was employed: Before each 2D NMR measurement, a 1D ^{1}H NMR experiment was acquired to ensure consistency of the spectrometer setup between the individual titration points. The water resonance was suppressed utilizing the DANTE pulse train (*101, 102*) or the excitation sculpting technique (*103*). ^{15}N-HSQC spectra were recorded in the sensitivity-enhanced mode using pulse field gradients for coherence selection (*104*). A selective, optimized water flip back pulse was incorporated to minimize the perturbation of the water signal, and a WATERGATE sequence for water suppression (*105*) was used with standard acquisition parameters. All ^{15}N-HSQC acquisition parameters were maintained for each titration experiment with the exception of the proton pulse length and transmitter offset that were optimized for each experiment. These titration experiments were done at enzyme concentrations of 100 μM to 500 μM. This made the use of a cryogenic probe at 500 MHz or 600 MHz essential (see Chapter VI). Typical experiment durations for each titration points were 2-3 hours, and 8-15 different ligand/enzyme concentration ratios were measured.

Data Analysis

Data analysis was performed with the help of four individual software packages. Detailed information and parameter settings can be found in Appendix D. Fourier transformation of the time domain data and visualization of the titration points was done by NMRPipe/NMRDraw (*106*) (Version V 2.5 NMR Science, North Potomac, MD). NMRView (*107*) (C-Version 5, NMRView J, One Moon Scientific) was used to extract the peak intensity information, which then served as the input data to either of the programs Prizm or xcrvfit (see Chapter III) to calculate dissociation constants.

Individual titrations were analyzed according to the following general procedure. Upon conversion of the Bruker time-domain data to NMRPipe format, the spectral series of 2D HSQC experiments were processed into a 3D data matrix. Next, a suitable peaklist was created. If no assignments were available a generic list was produced with the standard peak-picking tool within NMRPipe. From the 3D data matrix, peaks were picked for each individual plane and the peak file syntax was adapted. For known peak assignments, the peaklist for the first spectrum was replaced with the peaklist containing the assignment. Using a separate script (assD.tcl) each peak was visually inspected, corrected where necessary and added or deleted. The final, correct peaklist served as a reference for the entire series.

For pol X, as many as 2,000 resonances were available per titration. The NMRPipe interface “titrView” provided an invaluable tool to screen these data. An example is shown in Figure 4-2.

The changes in intensity and shift of the peaks were classified into four groups (G, M, B, S): G meant that the changing peaks were well separated from any other peaks and suitable for further analysis. Group M contained peaks that changed, but overlapped with other peaks. Group B had severely overlapping peaks. Group S contained peaks with a partial overlap, which was manifested as a peak shoulder. In this case, two possible cases could be differentiated: 1) One pair of partially overlapping peaks, where only one peak shifted 2) Peaks with a very small shift change, e.g. the new peak overlaps with the initial peak position. Neither of these peaks could be used for further analysis.

The peaks assigned to groups G and M were extracted from the NMRView peaklist into a separate file, and the relaxation tool in NMRview was utilized to determine the peak intensities for the titration series of each assigned residue. These intensities were then used together with the concentration information of ligand and protein at each titration point, to perform the nonlinear least square fit using the programs Prizm or xcrvfit.

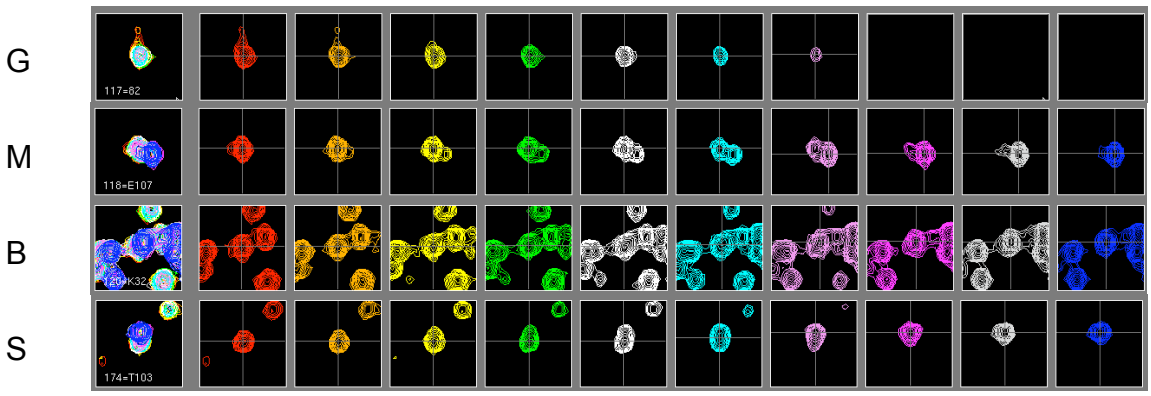


Figure 4-2: Spectrum matrix output from the NMRPipe module titrView. Individual strips are labeled as: G indicating well suitable peaks for K_d determination, M for peaks that might not be assigned properly during the titration analysis due to a small chemical shift difference resulting in peak overlap, B for heavily overlapping peaks and S for peaks with a shoulder.

Relaxation Measurements By NMR

An inverse detection scheme as proposed by Kay et al (*108*) was utilized for the ^{15}N-T_2 relaxation measurements. Acquisition parameters included a 1024 x 256 complex data matrix, 16 or 32 scans dependent on the sample concentration, 2.5 s relaxation delay and a T_2 CPMG sequence length of 12, 24, 36, 60 and 72 ms. Quadrature detection in F_1 was done according to States et al (*109*). Experiments were processed using NMRPipe (*106*) and NMRView (*107*).

Results

NMR Studies of a Binary Complex Between the ASFV Pol X and Nucleotides

Impact of Magnesium

A first measurement confirmed, that nucleotides bound to pol X only in the presence of Mg^{2+}, as shown in Chapter I and more importantly, caused a change in the HSQC spectrum indicating ligand-protein interaction. The binary complex was formed between pol X_{gpm} and dCTP. Lyophylized polymerase was dissolved in a buffer that contained 20 mM cacodylic acid, 10 mM DTT and 0.02% sodium azide. The dissolved protein had a concentration of 400 µmol/L. A ^{15}N-HSQC spectrum of the free pol X was measured. Upon addition of dCTP, another HSQC spectrum was measured and the two spectra were compared (Figure 4-3, A). No significant chemical shifts, as seen in Figure 4-3 A, were observed between the free pol X_{gpm} spectrum (black) and the spectrum with pol X_{gpm} in the presence of dCTP but in the absence of Mg^{2+} (blue). Changes as indicated in Figure 4-3 (B) were seen after adding 20 mM $MgCl_2$ to the sample containing pol X_{gpm} and a stoichiometric amount of dCTP (Figure 4-3 B, green spectrum), or a 4-fold excess of dCTP (Figure 4-3 B, purple spectrum). Some peaks shifted so far, that they could not be seen in the vicinity of the original peaks, and are indicated by the arrows in Figure 4-3, B. At a 4-fold excess of dCTP, one also observes

more than one resonance for some of the amines, indicating the possibility for multiple conformations.

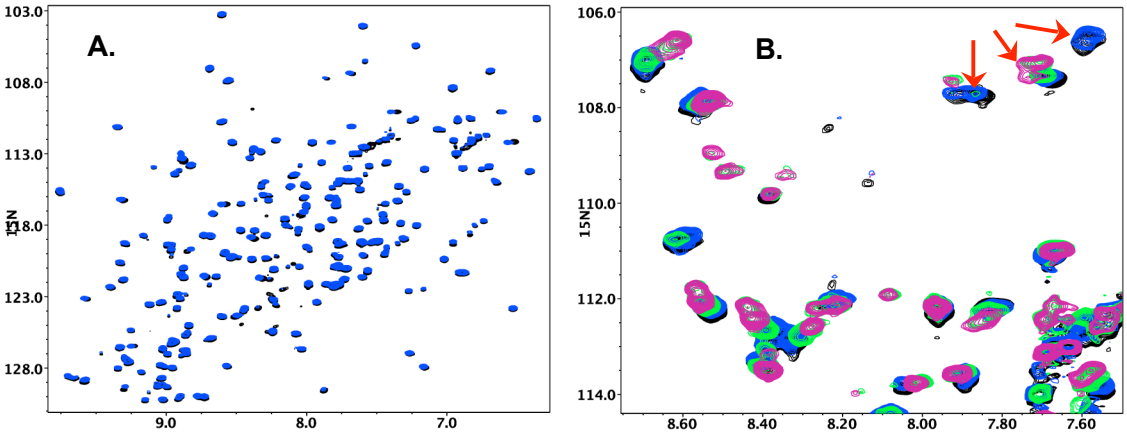


Figure 4-3: Pol X_{gpm} alone, with dCTP and dCTP plus Mg^{2+}. **A.** Full HSQC spectra of pol X_{gpm} only (black) and in the presence of equimolar amount of dCTP (blue). **B.** Expansion including all four spectra: pol X only (black), plus dCTP only (blue), plus 20mM Mg^{2+} (green) and a 4-fold excess of dCTP (purple).

Dissociation Constants for Deoxynucleotide Triphosphates

As a comparison to the ITC experiments and to gain more specific information on the binding interface, several nucleotides were titrated to pol X_{gpm}. Figure 4-4 shows a superposition of HSQC spectra, comparing pol X_{gpm} in the presence of the purines dGTP and ATP in Figure 4-4 (A) and the two pyrimidines dCTP and dTTP in Figure 4-4 (B). The spectra of the two purine complexes deoxyguanosine triphosphate and adenosine triphosphate were compared and chemical shift differences between the two complexes

are indicated in Figure 4-4 (A). The two deoxypyrimidine triphosphate complexes were in good agreement with each other. The red dCTP spectrum shows several distinct, small peaks, indicating the existence of at least two conformations at 4-fold excess. They are also present in the dTTP spectrum. Several resonances shift differently in the purine and the pyrimidine spectra and examples were pointed out with a circle. The experimental conditions and results for all four titrations were summarized in Table 4-1. Residues with chemical shift changes larger then the linewidth of the peak after addition of nucleotides were mapped on the existing structure of free pol X (PDB code 1JAJ).

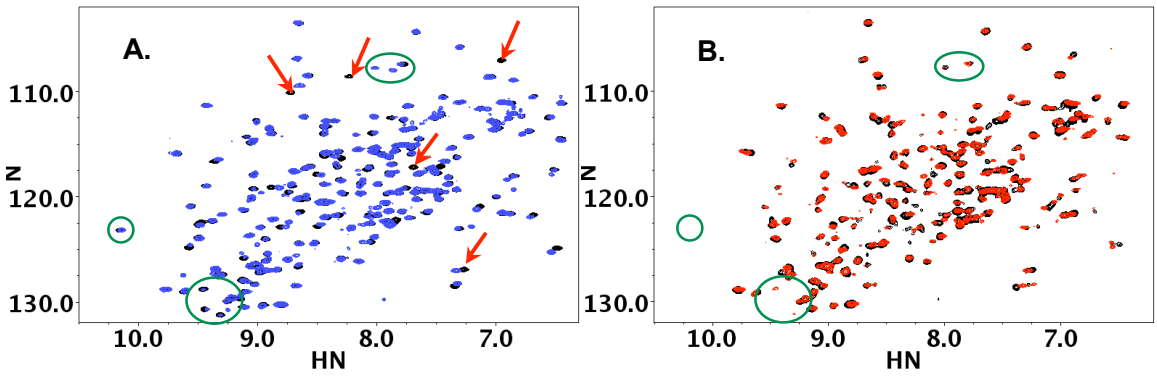


Figure 4-4: Comparison of effects seen by purines and pyrimidines in HSQC spectra: Superposition of HSQ spectra of pol X_{gpm} with **A.** two mol fractions of dGTP (black) or ATP (blue) added, the arrows indicate correlations that differ for these two nucleotides. **B.** five mol fractions of dTTP (black) and dCTP (red) added. The circles highlight differences between the purine- and pyrimidine-complex spectra.

Table 4-1: Interactions study of nucleotides and pol X_{gpm} by NMR.

	dGTP	ATP	dCTP	dTTP
pol X_{gpm} conc.	137 µM	120 µM	151 µM	231 µM
Nucleotide conc.	2.32 mM	2.83 mM 28.3 mM[1)]	2.5 mM, 25 mM[1)]	5 mM, 10 mM[1)]
Mol fractions added	0, 0.05, 0.1, 0.15, 0.2, 0.4, 0.6, 0.8, 1, 2	0, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 2, 5, 15	0, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 4	0, 0.25, 0.5, 0,75, 1, 2, 4
NaCl conc.	300 mM	300 mM	300 mM	300 mM
^{1}H frequency	600 MHz	600 MHz	500 MHz	500 MHz
Acquisition matrix	1024 x 256	1024 x 256	1024 x 256	1024 x 128
Scans	32	32	32	32
Shifts used for K_d analysis [2)]	35 - 38	31 - 37	13	20
K_d [3)]	2.4 ± 0.7 µM	10.2 ± 1.0 µM	49.7 µM	40.2 µM

[1)] Higher concentration of stock solution used for above stoichiometric additions

[2)] Number of usable peaks varies depending on type of analysis.

[3)] K_d value as average of all residues analyzed.

Purines: dGTP and ATP

The addition of dGTP or ATP caused chemical shift changes as described for the slow exchange. The intensity of the affected peaks was reduced upon addition of the purines and new resonances of the complexed amides appeared in the HSQC spectrum as indicated in Figure 4-5.

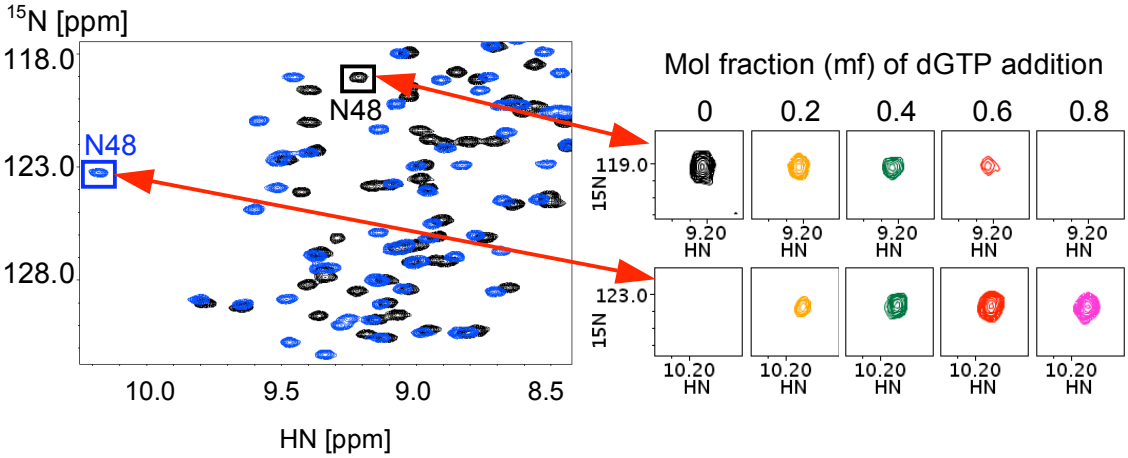


Figure 4.5: Titration spectra of pol X_{gpm} with dGTP. Left side: Expansion of ^{15}N-HSQC of free pol X (black) and 2 mf dGTP complex (blue). Right side: Tile plots of N48 peaks at different mol fraction of dGTP addition (0, 0.2, 0.4, 0.6, and 0.8 mf). Upper row indicated N48 peak of free pol X, lower row of the dGTP complex.

The analysis of the dGTP complex was possible once the assignment for the shifted residues was known. The change in resonance frequency caused some of the complex peaks to overlap with peaks from the free pol X, or overlapping regions from the free pol X spectrum were simplified by peaks shifting away for the complexed spectrum. Therefore the set of residues analyzed was not identical for both analyses. The peak intensities were used to determine the K_d and were fit to the nonlinear, quadratic equation with the program Prizm. The dissociation constants reported in Table 4-1 represent an average of all determined values for each residue and nucleotide in each

analysis series; e.g. it represents the overall average of all values obtained. The error is a reflection of the differences between the individual analysis series.

The following residues were used for the analysis of the dGTP-complex. The peaks classification followed the definition outlined in the introduction: Group G: L2, G7, V37, G38, L40, R42, L47, N48, D51, L52, L53, C86, L88, I90, Q98, L99, D100, A104, F116, T117, G118, Y122, I124, R125, E156, K157, F165, I169, R173, group M: K9, N12, S39, E43, K45, V56, E58, K63, H64, L66, V79, V80, K85, E91, F102, A106, E107, I113, H115.

Figure 4-6 (A) indicates that residues D49, V50, V87, F89 and L101 were not included in the analysis, although the chemical shift of the residues around them had changed. All of them were located on the same β-sheet. Closer examination of the data showed, that V87 was the only residue so far not assigned, and the other four residues were not included due heavy peak overlap. These residues were most likely affected as well and colored red in the surface map. Since the proline residues were not detected in the HSQC spectrum, they also remained unassigned and in this area of interest include P110, P119 and P170. The catalytic triad with residues D49, D51 and D100 was indicated in green.

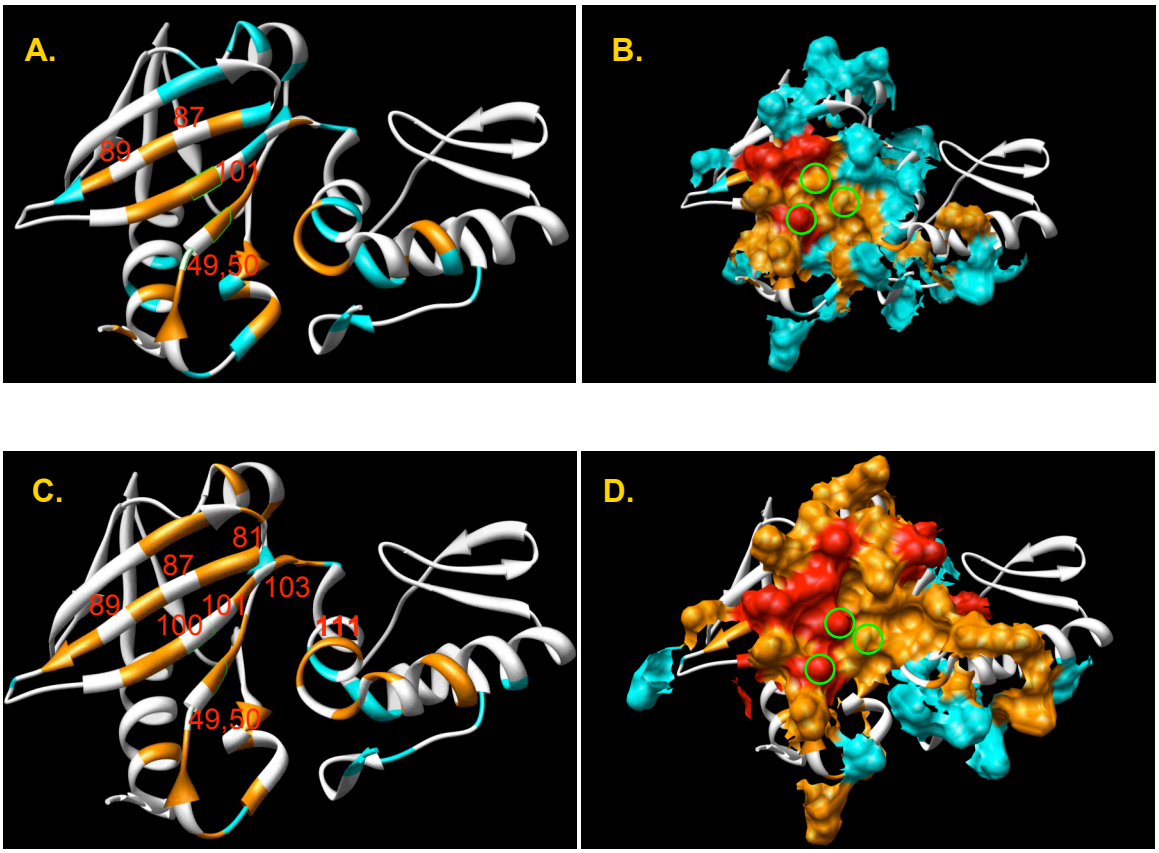


Figure 4-6: Mapping of dGTP (**A.** and **B.**) and ATP (**C.** and **D.**) interactions with pol X. Individual residues that show a chemical shift change are colored on a ribbon and in a surface representation allowing for differentiation of position and surface interface. Orange indicates well-resolved HSQC peaks (group G) while peaks in cyan belong to group M, where some degree of peak overlap exists. The red labels on the surfaces indicate overlapping peaks and V87, which was not assigned. The aspartic acid residues of the catalytic triad D49, D51 and D100 are outlined in green.

The analysis for the ATP titration followed the same scheme as was used for the dGTP. It was possible to identify the peaks and adapt the peaklists from the dGTP assignment since the peak shifts between the dGTP- and ATP-pol X complexes were

comparable. Only six peaks shifted substantially into different position, G38, F116, G118, V120 and R124, but could still be assigned.

The following peaks were in group G for the ATP-complex: L2, G7, G38, L40, E43, K45, L47, N48, D51, L52, L53, E58, K63, K79, V80, G82, L85, C86, L88, I90, E91, Q98, L99, F102, A104, L105, H115, F116, G118, V120, Y122, I124, R125 and group M contained: E44, V56, E93, E107, I113, A129, E156, K157, F165, R168, R173 and L174. Both groups of peaks were used for the K_d determination. Residues D49, V50, C81, E83, R84, F89, Y97, D100, L101, T103 and Y111 were overlapped and therefore not classified, although some chemical shift was indicated. Their location in the molecule would indicate, that they are involved in the binding interface though. Mapping the residues affected by the purine addition to the pol X structure (Figure 4-6) revealed that the interaction surfaces (A and C) were very similar to each other. The three residues that had been indicated as the catalytic triad were located within the outlined nucleotide binding site. It was also interesting to see that the affected residues did cross over to the C-terminal domain, which was an indication that the nucleotide was positioned in the interface site. Some conformational change between the two subdomains upon binding of the nucleotide was also envisioned since the affected surface was involving residues to the top of the β-sheet 8-10 and αC. The K_d was 3-fold stronger for the dGTP

compared to the ATP. The values corroborate those measured by ITC at 2.4 μM and 10.2 μM for the dGTP and ATP, respectively. These K_d values suggest that dGTP and ATP nucleotides bind to the polymerase even in the absence of DNA.

Pyrimidines: dCTP and dTTP

A similar study was done for the pyrimidines dCTP and dTTP. The experimental conditions were the same as for the purines except that a final excess of five times the nucleotide was added to pol X_{gpm}. The red spectrum in Figure 4-4 (B) shows several smaller peaks, indicating that saturation was not reached even at a 5-fold excess of dCTP. This was more pronounced for dCTP compared to dTTP and confirmed the weaker binding indicated by the dissociation constant. Data was analyzed as described above for those peaks that shifted in analogy to the dGTP spectrum and could be assigned. Residues belonging to group G were: G7, V11, V37, V38, S39, R42, E43, L47, N48, D51, L52, K63, F116, G118, L174. Group M: K85, C86, L88, T117, I124, Q139. Inclusion of a number of peaks was not possible due to heavy overlap. Mapping these peaks to the pol X structure did not indicate any major changes. The surface coverage for the assignable peaks was comparable to the one shown for ATP and dGTP. Nonetheless, a significantly reduced dissociation constant was observed. The nonlinear,

quadratic fit for the residues listed above resulted in a K_d of 49.7 μM for dCTP and 40.2 μM for dTTP. These values are roughly 20 times lower then the ones seen for dGTP and could be responsible for some of the discussed differences in the polymerase activity. These measurements also confirmed that there is a distinct difference between the purines and the pyrimidines.

NMR Studies of ASFV Pol X - DNA Complexes

A structural study of the ternary complex between pol X, a DNA fragment and the incoming base would reveal the most detailed information and provide a mechanistic understanding of the DNA repair process. Encouraged by the fluorescence data, where strong affinity between several DNA fragments and pol X was observed, a series of NMR studies was undertaken to work out the conditions for these systems. Since NMR has a low sensitivity, sample concentrations had to be increased compared to the fluorescence measurements. Such sample concentrations can alter a biological system, cause secondary effects like aggregation, loss of solubility or impact the affinity of a complex.

For all the experiments described in this section a similar protocol was used: The pol X_{gpm} samples were prepared in gpm-buffer, containing 500 mM NaCl and 20 mM

Mg^{2+}, if not stated differently. The concentration was determined by an A_{280nm} standard addition method using 4-5 data points. DNA samples were dissolved in the same buffer as used for the polymerase. Annealing was achieved by heating up the sample to 80 ºC for 10 min followed by a cool down over the course of several hours. The DNA sample was tested for purity using capillary gel electrophoresis (CGE). A 1D ^{1}H NMR spectrum of DNA alone showed the presence of the signature imino protons as described before. 1D ^{1}H NMR experiments of the protein and complex were measured before each 2D HSQC experiment to assure proper sample conditions.

Complex Formation with the 36-mer Double Hairpin DNA

A 36-mer oligonucleotide as described in the Chapter III were used. Distance measurements on the existing structure indicated that the duplex core of the 36-mer should be sufficiently long to span the hypothesized DNA interaction site of pol X. Figure 4-7 shows the measured ^{15}N-HSQC spectra at 600 MHz superimposed with several titration spectra at 0, 0.02, 0.05, 0.10, 0.20, 0.35, 0.50 and 0.65 mol fractions (mf) of the 36-mer. A different behavior compared to the nucleotide titration spectra was immediately apparent. Here, none of the correlation peaks was noticeably shifting but the intensity of the peaks was varying to a large extent. While all but four peaks were still

present after the addition of 0.2 mol fractions of the 36-mer, 26 peaks were no longer visible after the addition of 0.35 mf DNA and 72 peaks had vanished after the addition of 0.65 mf, which represents about half of all the peaks. For about 32 residues no major change was detected for the amide signal intensity. Table E-1, Appendix E summarizes this data and lists the peaks involved in each category. When increasing the DNA concentration to a stoichiometric ratio only few additional changes occurred. The number of scans for HSQC spectra with 0.65 and 0.8 mol fractions of DNA was increased from 4 to 32 to better observe the low intensity signals.

The surface map of the residues shown in Figure 4-8 provides an easier look than the data listed in Appendix E, Table E-1 and shows where the changes took place. The orange surface represents residues, that were affected after DNA additions of less then 0.35 mf, cyan those below 0.65 mf. Residues colored in white showed various degrees of peak intensity reductions, while those in magenta were unaffected at 0.65 mf. The area of strongest change was along the subdomain interface and spanned from the C-terminal side of α-helix E over to the β–sheet β11 and β12, including the loop region between the two strands. These two distinct regions (subdomain interface and connection from αE to αC) within the protein were identified to be strongest affected in the presence of DNA, creating an L-shape interface. The residues not affected by the

addition of the 36-mer were clustered on the other side of the molecule, mainly on αA, β2 and β3.

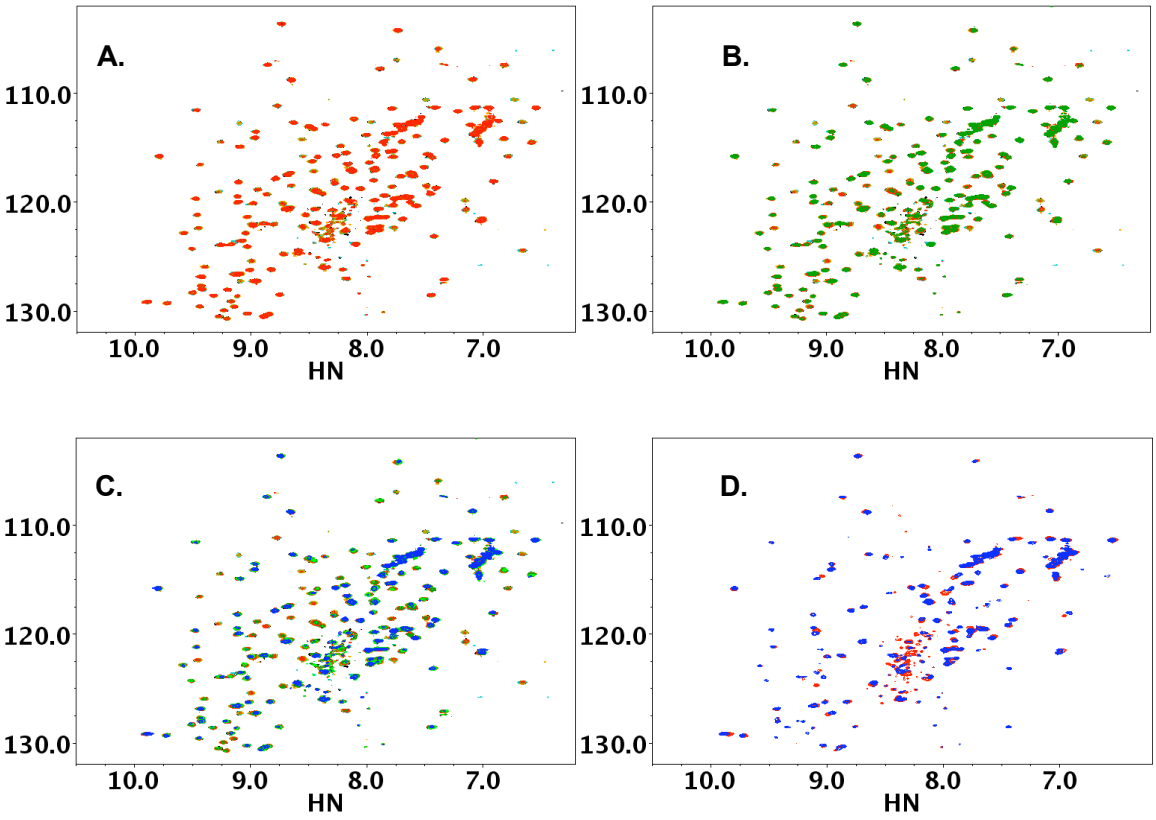


Figure 4-7: Superposition of pol X ^{15}N-HSQC spectra at different 36-mer DNA concentrations, expressed in mol fractions (mf) of DNA, as listed in the color scheme on the right. **A.** Superposition of spectra from 0 to 0.2 mf, **B.** 0 to 0.35 mf, **C.** 0 to 0.65 mf. All spectra were recorded with four scans. **D.** 32 scan ^{15}N-HSQC spectra, 600 MHz, of 0.65 mf in blue and at stoichiometric concentration in red.

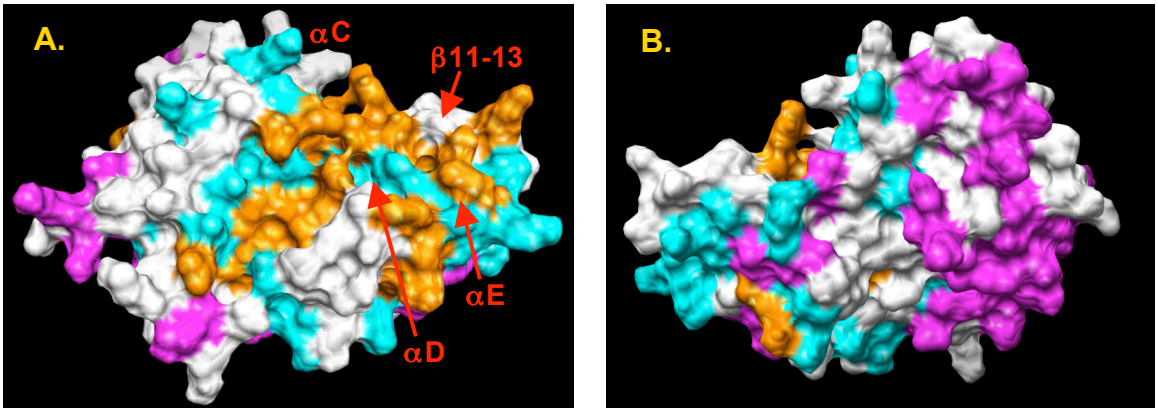


Figure 4-8: Pol X-36-mer interaction surface map. **A.** Pol X with the palm on the left and C-terminal on the right. **B.** Rotation of A. by 180 degree. Residues colored in orange have completely disappeared after addition of 0.35 mf of 36-mer, those in cyan are no longer visible at 0.65 mf. Magenta indicates residues with no appreciable intensity change at 0.65 mf and residues in white show a reduced intensity at 0.65 mf, but their peaks are still visible.

Complex Formation with the 44-mer Double Hairpin DNA

The result from the study with the 36-mer suggested that the pol X-DNA complex was in intermediate exchange at the NMR timescale. Next the 44-mer DNA sequence was used, to see whether the core length would make a difference on the exchange rate. Fluorescence studies had indicated a tight complex with the 44-mer with a K_d of 8 μM even in the presence of 500 mM NaCl.

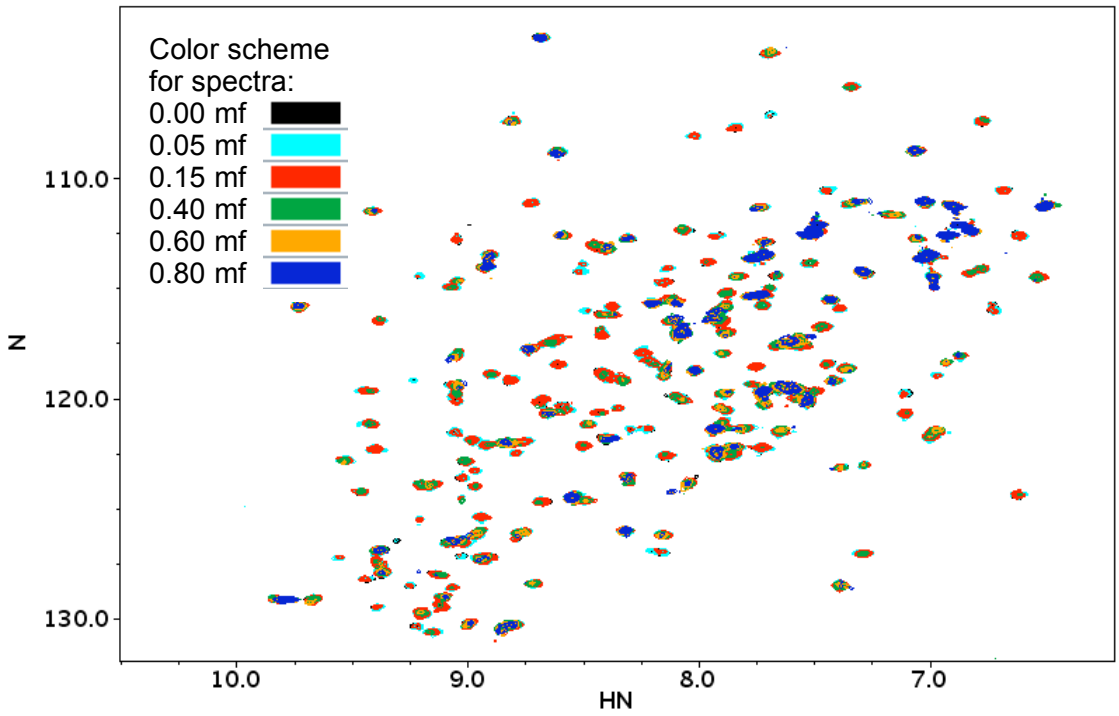


Figure 4-9: ^{15}N-HSQC spectra of pol X - 44-mer DNA complex, 600 MHz, 32 scans, at differing mol fractions (mf) as listed in the figure.

NMR measurements were performed in the same way as described for the 36-mer DNA. The pol X_{gpm} concentration was 276 μM and 44-mer was added in mol fractions of 0.05, 0.15, 0.25, 0.4, 0.6 and 0.8.

While the complex formation between the 36-mer and 44-mer DNA with pol X_{gpm} was found to be very comparable, there were some distinct changes (Figures 4-7 and 4-9). First, upon addition of 0.6 mf of 44-mer only 20 residues remained unchanged and this number was even smaller at 0.8 mf. This compared to 45 unchanged residues for the 36-mer, where no additional change was observed upon further DNA addition. For

the 44-mer on the other side, a continued trend was seen towards decreased peak intensity of additional peaks. Second, a number of peaks were found to be split into two or more peaks at 0.6 and 0.8 mf of 44-mer DNA. This was not the case for the 36-mer. Table E-2 in Appendix E summarizes the observed changes for the individual residues after adding 44-mer oligonucleotide. Consistent with the overall decreased peak intensity for the 44-mer, more peaks were no longer visible at 0.6 mf compared to the 36-mer. Otherwise similar residues were affected during both DNA titrations as indicated in Figure 4-10.

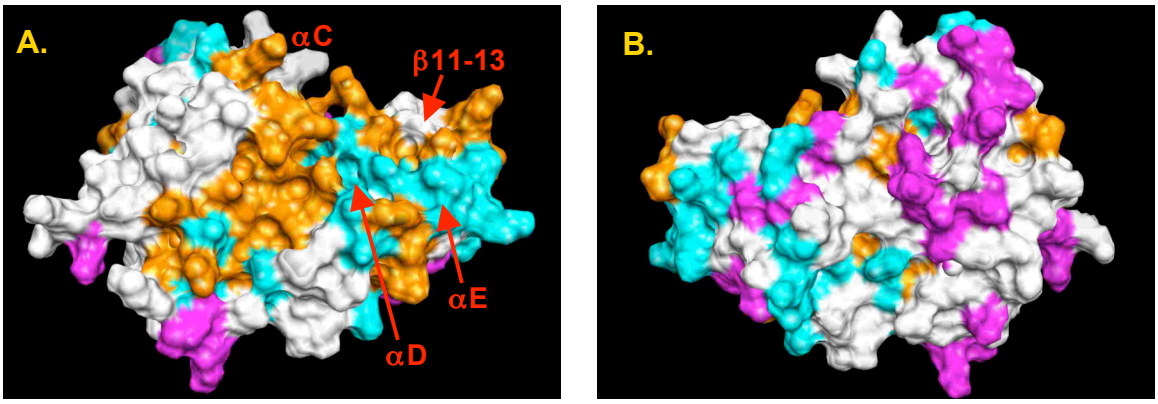


Figure 4-10: 44-mer - pol X_{gpm} interaction surface mapping. **A.** Pol X with the palm on the left and C-terminal on the right. **B.** Rotation of A. by 180 degree. Residues colored in orange have completely disappeared at 0.40 mf of 44-mer addition and those in cyan are no longer visible at 0.60 mf. Magenta indicates residues with little or no intensity change at 0.60 mf compared to the original spectrum and the residues in white show much reduced intensities at 0.60 mf, but still visible peaks.

Ternary Complex Formation with 16-mer DNA in the Presence of ddGTP

The 36-mer and 44-mer DNA sequence, both with a T_5 hairpin loop, exhibited similar behavior as monitored by the HSQC spectra and showed exchange-broadened resonances for the majority of amides at sub-stoichiometric DNA concentrations. Such spectra are unsuitable for structural NMR work and therefore the 16-mer was studied next. As described in the previous chapter, this sequence had a CpC-gap and a non-bulky loop. The K_d of 2.3 μM was determined for the 16-mer by fluorescence (Chapter II), which was between the one for the 34-mer and the 44-mer. The presence of ddGTP increased the binding to the 16-mer 5-fold. For the NMR titration, the dideoxyguanosine triphosphate was added to pol X_{gpm} before the titration with the 16-mer. Standard ^{15}N-HSQC were measured at 500 MHz, 32 scans in a 4 mm NMR tube containing 300 μL of pol X_{gpm}, 250 μM, in the gpm-buffer at 500 mM NaCl. Similar to the dGTP spectrum, the 10-fold excess of ddGTP produced a number of chemical shift perturbations for the pol X_{gpm} ^{15}N-HSQC spectrum as shown in Figure 4-11. Several correlation peaks showed multiple resonances, similar to the ones reported for ddCTP. Although the differences between the dGTP and ddGTP spectra were larger than expected the ddGTP complex spectrum could be used as a baseline for the titration with the 16-mer. The main focus was to maintain the HSQC resonances of pol X while forming a complex with DNA.

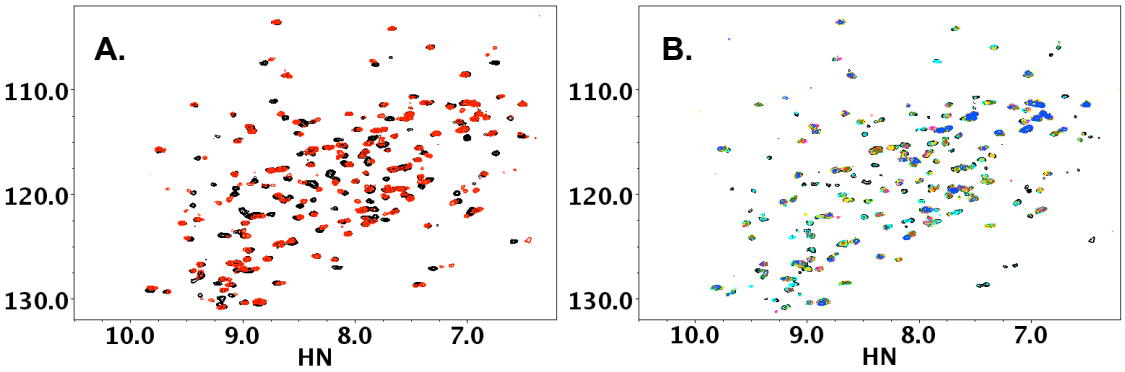


Figure 4-11: Ternary ^{15}N-HSQC spectra for pol X_{gpm} with ddGTP and various amounts of 16-mer, 500 MHz, 32 scans. **A.** Before (black) and after addition of 10 mf ddGTP (red), and **B.** after addition of varying mol fractions (mf) of 16-mer DNA as indicated in the color scheme.

Color scheme for spectra:
0.0 mf
0.1 mf
0.2 mf
0.4 mf
0.6 mf
1.0 mf
2.0 mf

The anticipation was that a ternary complex was more stable hence lead to an HSQC spectrum that would contain all correlations. This was not the case, as is evident from Figure 4-11. The resulting spectra resembled the by now very familiar scenario where upon addition of 16-mer successively less amide residue signals were seen in the spectrum.

NMR Relaxation Studies for the 36- and 44-mer DNA – Pol X Binary Complex

An extensive series of transverse relaxation (T_2) measurements were conducted to determine T_2 values for pol X alone or in the presence of DNA. These relaxation

values would give some indication of the overall molecular motion, and therefore show, whether a monomeric or dimeric complex would be present under the various conditions.

Two different pol X constructs were used in the absence and presence of the 36-mer and 44-mer DNA. One measurement also involved the pol X-dGTP complex. Table 4-2 to 4-3 summarizes the average T_2 times determined for all these complexes, while Table 4-4 provides calculated T_2 values for the various possible pol X-complex combinations.

The T_2 relaxation times listed in Table 4-2 were measured at 500 MHz at either 293 or 298 K. The values reported confirme, that pol X was in the monomeric state under the conditions used and could not be a dimer according to these measurements. The calculated T_2 values assumed a spherical molecule, therefore the measured relaxation times were lower, indicating a slower tumbling than would be the case for an ideal spherical shape. Changing the temperature by only five degrees showed the expected trend of a 5-10 ms decrease in T_2 relaxation time. The two pol X constructs showed no difference in the overall tumbling behavior. This was expected since the difference between the two proteins was only in a C-terminal His-tag. The lower salt concentration used for Pol X_{iz} was not expected to have any impact on the tumbling unless aggregation would occur under these conditions. This appeared not to be the

case. The addition of dGTP also did not change the relaxation time, due to the small molecular weight addition.

Table 4-2: Average T_2 times for pol X at four different conditions

500 MHz	T_2 [ms] @ 298 K	T_2 [ms] @ 293 K
pol X_{iz}	65.1 ±3.7 (144)[1)]	
pol X_{gpm}	61.1 ±3.8 (157)[2)]	
pol X_{gpm}		55.8 ±3.8 (152)[3)]
pol X_{gpm} + 5 mol fractions dGTP		59.5 ±2.1 (165)[4)]

[1)] 0.4 mM pol X_{iz} 20 mM NaAc, 50mM NaCl, 0.02% NaN_3, 5 mM $MgCl_2$, pH 6.0

[2)] 0.5 mM pol X_{gpm}, gpm-buffer, 500 mM NaCl

[3)] 1.0 mM pol X_{gpm}, gpm-buffer, 500 mM NaCl

[4)] 1.0 mM pol X_{gpm}, gpm-buffer, 500 mM NaCl plus 5 mol fractions of dGTP.

The number of residues available for the T_2 analysis is given in the parenthesis.

Table 4-3: Average T_2 times for two different pol X-DNA complexes.

600 MHz	T_2 [ms] 36-mer [1)]	T_2 [ms] 44-mer [2)]
pol X_{gpm} + 0 mol fraction DNA	65.0 ± 2.7 (158)	70.4 ± 5.3 (150)
pol X_{gpm} + 0.02 mol fraction DNA	63.8 ± 3.4 (157)	
pol X_{gpm} + 0.10 mol fraction DNA	59.8 ± 4.0 (149)	
pol X_{gpm} + 0.15 mol fraction DNA		57.2 ± 5.8 (100)
pol X_{gpm} + 0.20 mol fraction DNA	57.2 ± 4.3 (114)	
pol X_{gpm} + 0.35 mol fraction DNA	52.9 ± 3.5 (72)	
pol X_{gpm} + 0.40 mol fraction DNA		50.3 ± 7.1 (33)
pol X_{gpm} + 0.50 mol fraction DNA	52.6 ± 4.3 (39)	
pol X_{gpm} + 0.60 mol fraction DNA		47.6 ± 4.1 (15)
pol X_{gpm} + 0.65 mol fraction DNA	46.7 ± 5.0 (38)	
pol X_{gpm} + 0.80 mol fraction DNA	42.9 ± 3.7 (35)	

[1)] 460 µM pol X_{gpm} titrated with the indicated amount of 36-mer double hairpin DNA

[2)] 276 µM pol X_{gpm}, titrated with the indicated amount of 44-mer double hairpin DNA

All samples were dissolved in gpm-buffer, 500 mM NaCl at 25^0C, 600 MHz. The number of residues available for the T_2 analysis is given in the parenthesis.

Table 4-4: Estimated T_2 relaxation values at different conditions. The calculation assumed an internuclear H-N distance of 0.102 nm, dipolar and CSA terms were used as described in (*108*), S^2 was assumed to be 0.85 and t_e $105x10^{-12}$. Molecular sizes were chosen according to the monomer size of pol X, its potential dimer and either one in a complex with the 36-mer (11 kDa) or 44-mer (13.6 kDa).

Molecular Size [kDa]	B_o [MHz]	T [K]	T_2 [ms] (average)	T_2 [ms] (range)
21	500	293	66	34 – 84
21	500	298	74	39 - 94
21	600	298	70	36 - 89
42	500	298	39	20 - 50
42	600	298	36	18 - 47
32 (+ 36mer)	600	298	47	24 - 61
35 (+ 44mer)	600	298	43	22 - 56
53 (dimer + 36mer)	600	298	29	14 - 37
56 (dimer + 44mer)	600	298	27	14 - 35

Table 4-3 lists relaxation times measured for the 36-mer and 44-mer complexes. All these T_2 relaxation times were obtained at 600 MHz and 298 K. Before adding any DNA good agreement existed between the calculated relaxation time of 70 ms and the measured ones of 65 ms and 70 ms. Various amounts of each DNA were added to the pol X_{gpm} sample and the mol fractions at which the transverse relaxation times were determined are listed in Table 4-3. Considerable changes in T_2 times were observed for both oligonucleotides as their respective concentration was increased. Values for these relaxation times decreased as more DNA was added and reached a low value of 42.9 ms and 41.1 ms at a mol fraction of 0.8 for the 36-mer and 44-mer, respectively. As

indicated in parenthesis, the number of amide correlation peaks that could be used for the analysis was reduced dramatically at that point due to the exchange broadening. The remaining peaks that were observable were located in a part of the molecule that was less affected by exchange broadening. Since these measurements were used to determine the overall tumbling of the molecule, these amides could still serve as molecular probes. The decreasing relaxation times reflected the average growth in complex. The more the polymerase was saturated on average, the lower the apparent relaxation time was.

When comparing the T_2 relaxation times from the complex with the calculated values, it could be conclude, that these pol X – DNA complexes existed as a monomers. Attempts to map the transverse relaxation times for individual residues had to be dismissed, since the extent of the exchange contribution was not known. In order to define intra-molecular motion, a full characterization of the molecular dynamics would be required, including longitudinal and heteronuclear relaxation measurements and a fitting procedure to a motional model.

NMR Binding Studies Involving Single Strand And Template-Primer DNA

A recent study from Jezewska et al (*70*) described interactions between pol X and single stranded DNA (ss-DNA) using quantitative fluorescence and analytical ultracentrifugation measurements. They used poly-etheno-derivatized oligonucleotides as probes for their fluorescence measurements. This probe provided a strong signal, which allowed the use of low sample concentrations. Measurements were done in 10 mM sodium cacodylate, 1mM $MgCl_2$, and 10% (v/v) glycerol, pH 7.0. They found, that these derivatized, ss-DNA exhibited different associations with pol X, depending on the length. No binding was detected with shorter than five oligonucleotides. A first binding event, involving the so-called proper binding site, required 7 (±1) residues and was the only binding site involved up to a 12-mer oligonucleotide. A transition phase was reported between a 12- and a 16-mer. The total binding site involved a second binding event and required a minimal oligonucleotide length of 16 (±1) residues. Past 20-24 residues, the binding of a second pol X molecule to the single stranded DNA was reported. The reported intrinsic binding constant K_p ranged from $5.4x10^4$ for the 7-mer to $6.8\ x10^5$ for the 20-mer (K_d: 19 to 1.5 μM, respectively), which was comparable to the K_d values observed for in this study. The nature of these fluorescence and

ultracentrifugation experiments allowed only the observation of the global effects while speculations could be made about the actual protein-DNA binding interface.

NMR provided an ideal tool to identify the specific residues on pol X that were involved in these interactions. Using ss-DNA was believed to be an alternative tool to investigate the binding interaction of pol X with ds-DNA.

In analogy to the report by Jezewska et al (*70*), a series of oligonucleotides at various length and base content was prepared. The NMR detection allowed the use of all naturally occurring bases, eliminating any potentially adverse affects of the derivatized oligonucleotide used by Jezewska et al. Oligonucleotides were dissolved in gpm-buffer at 500 mM NaCl. The following stock solutions were used: $dC(pdC)_6$ at 1.21 mM, $dG(pdG)_6$ at 2.37 mM, $dC(pdC)_8$ at 1.22 mM, $dA(pdA)_{11}$ at 4.11 mM, $dT(pdT)_{11}$ at 10.0 mM, $dA(pdA)_{11}$ $dCp(dC)_5$ at 6.13 mM, dideoxynucloetide triphosphates: ddTTP at 1.56 mM, ddATP at 1.98 mM and ddGTP at 1.89 mM. A freshly prepared ^{15}N labeled pol X_{gpm} stock solution of 742 µM was prepared in the same buffer as the oligonucleotides. Standard HSQC experiments as described before were performed at 500 MHz, 16 or 32 scans, in 3 mm NMR tubes with a sample volume of 160 µL. The oligonucleotides were uniformly added at mol fractions of 0.25, 0.5, 0.75, 1.0, 3.0 and 5.0.

$dC(pdC)_6$ and $dG(pdG)_6$

The titration for both $dC(pdC)_6$ and $dG(pdG)_6$, resulted in a chemical shift perturbation for Q146 only, even at a 5-fold excess of the oligonucleotide. According to Jezewska et al (*70*), the 7-mer should be sufficiently long to interact with the proper ss-DNA binding site, but no changes were observed by NMR. The most likely explanation was the 5-fold higher NaCl concentration present in the NMR samples, which was requirement for sample stability for the concentrated NMR samples. Other less likely reasons could include the difference between the poly-dC/poly-dG and the poly-dεA used in those experiments.

$dC(pdC)_8$

Upon addition of $dC(pdC)_8$ several changes were observed in the ^{15}N-HSQC spectrum. These changes were much more gradual and required a 5-fold oligonucleotide excess to become visible. The intensity of the correlation peaks for only four residues (Figure 4-12, orange colored), R84, K85, D100 and Q146, weakened after a stoichiometric addition of $dC(pdC)_8$ and were no longer detectable at the 5-fold excess, or had shifted too far from the original position that they were not identified anymore. A reduced intensity for 21 amide resonances was seen (cyan) and slight

chemical shift perturbations were observed for 24 amide peaks (green). The classification for all residues is listed in Appendix E, Table E-3. The main perturbations were visible at the C-terminal side of α-helix E, the loop between β11, β12 and β9, β10. It was now possible to locate the proper binding site there. Several other perturbations were visible on αB and αF. They might indicate some movement between the two subdomains.

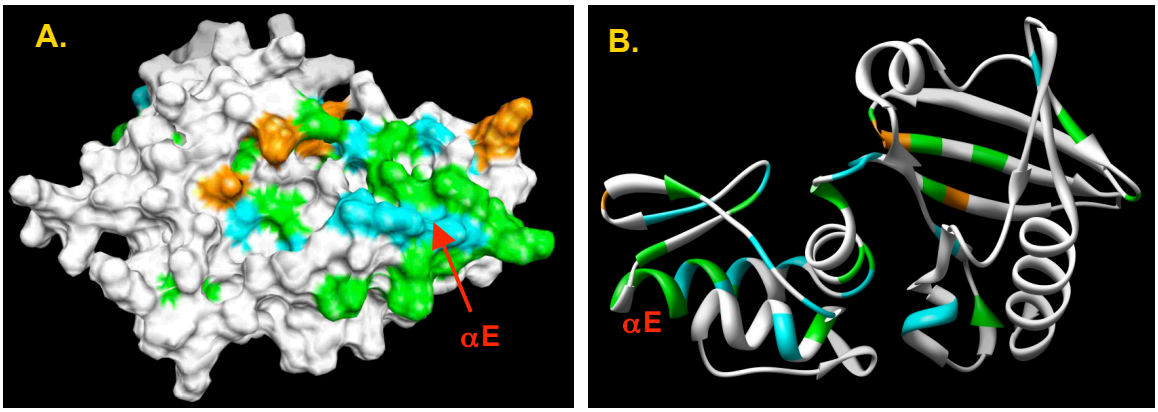


Figure 4-12: Binding interface with $dC(pdC)_8$. **A.** Pol X with the palm on the left and C-terminal on the right side. **B.** Rotation of A. by 180 degree. Color code for the classification of amide peak changes in the ^{15}N-HSQC spectrum after addition of 5 mf $dC(pdC)_8$ single strand DNA: orange indicates residues whose amide peaks are no longer detected at their original position at 5 mf; cyan, peaks with strong chemical shift perturbations; green, peaks with minor chemical shift perturbations and white indicates no change was seen.

$dA(pdA)_{11}$ and $dT(pdT)_{11}$

According to the report from Jezewska et al (*70*), a 12-mer ss-DNA should completely cover the proper binding site of pol X. The NMR experiments was performed with $dA(pdA)_{11}$ and $dT(pdT)_{11}$. Additions of 5-fold excess caused a comparable response on the protein for both ss-DNA sequences. Compared to the 9-mer, an extended perturbation at the C-terminal domain and an additional affect on α-helix C could be observed. The subdomain interface also experiences additional changes. A list of all affected residues is contained in Appendix E, Table E-3. An additional 2-fold excess of the complementary dideoxynucleotide (ddTTP or ddATP, respectively) was added to both samples, to find out, whether a stable ternary complex was formed. This caused a dramatic change in the spectra with only few residues left unaffected. These unchanged residues are colored in magenta in Figure 4-13 (B), and the data is listed in Table E-3, Appendix E.

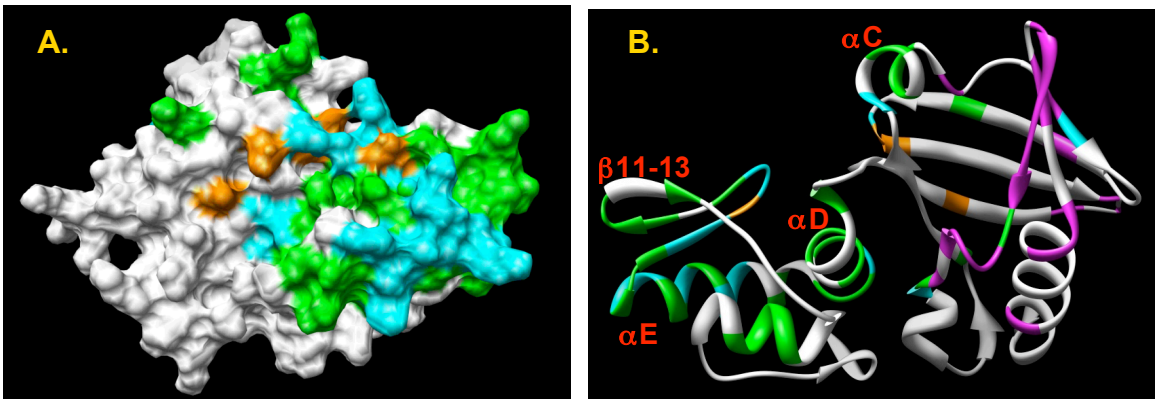


Figure 4-13: Binding interface with $dA(pdA)_{11}$. **A.** Pol X with the palm on the left and C-terminal on the right side. **B.** Rotation of A. by 180 degree. Color code for the classification of amide peak changes in the ^{15}N-HSQC spectrum after addition of 5 mf $dA(pdA)_{11}$ single strand DNA: orange indicates residues whose amide peaks are no longer detected at their original position at 5 mf; cyan, peaks with strong chemical shift perturbations; green, peaks with minor chemical shift perturbations and white indicates no change was seen. Magenta colored residues are the ones that did not show changes after addition of ddTTP.

$dA(pdA)_{11}dC(pdC)_5$

This 18-mer was expected to span the total binding site (*70*) of pol X and was similar in size to some of the duplex DNA fragments described earlier. The individual ^{15}N-HSQC spectra (Figure 4-14) revealed a distinct difference to the double stranded DNA complex measurements. Although the intensity of the amide peaks was reduces in both cases when more DNA was added, more ss-DNA was required to show this effect. For the 18-mer ss-DNA, the majority of peaks was still present in the spectrum at a stoichiometric ratio, while a majority of the peaks was no longer detectable past 0.6 mf

for the duplex DNA. This would indicate that the exchange broadening with the ss-DNA was not as sever, as was the case for the ds-DNA. Since saturation occurred later, the overall binding constant must be weaker as well.

Another difference was the shifting of the peaks. No peak shifting was observed for the ds-DNA. At a 5-fold excess of $dA(pdA)_{11}dC(pdC)_6$, the majority of affected peaks were either in complete exchange broadening, showed a large chemical shift perturbation or both.

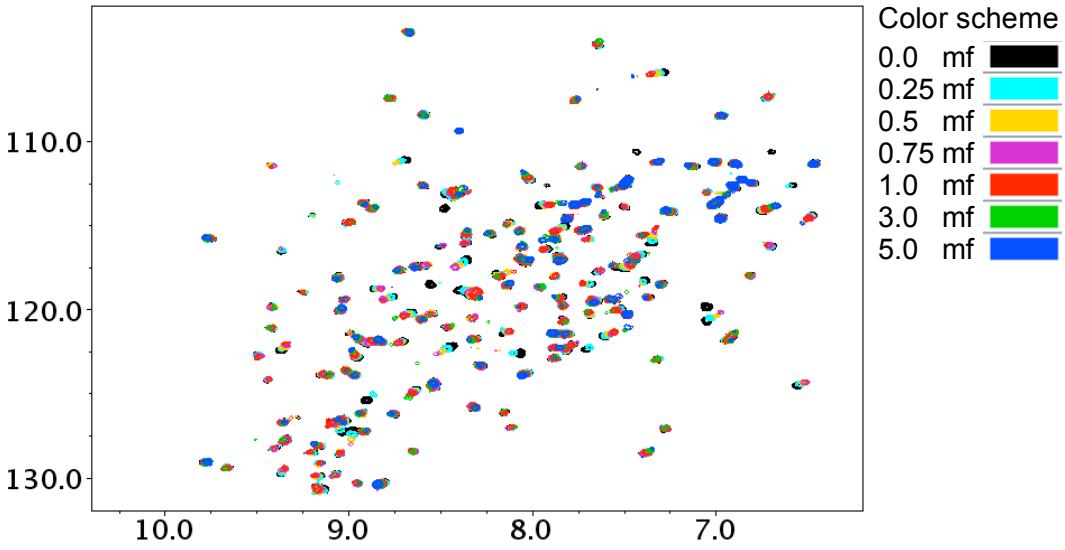


Figure 4-14: Superposition of ^{15}N-HSQC spectra when titrating pol $X_{gpm,}$ with $dA(pdA)_{11}dC(pdC)_6$. Mol fractions of $dA(pdA)_{11}dC(pdC)_6$ are indicated in the color scheme.

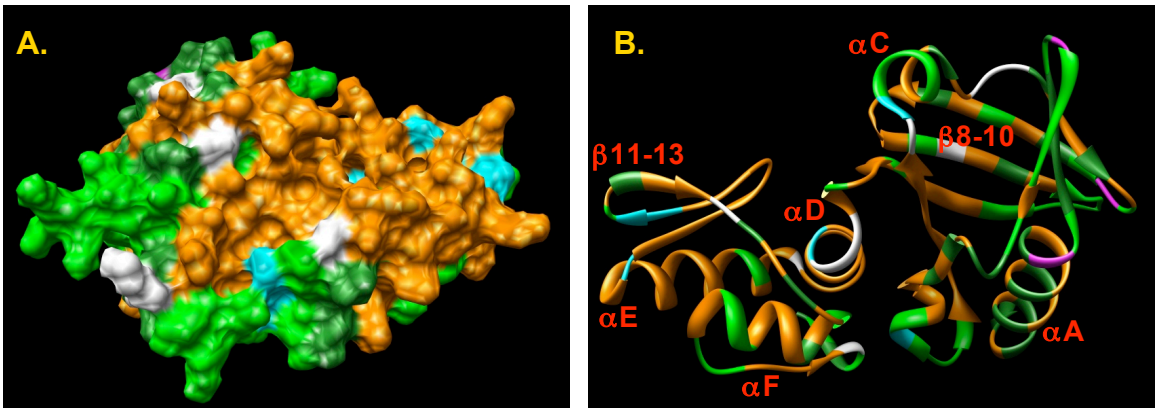


Figure 4-15: Binding interface with $dA(pdA)_{11}dC(pdC)_6$. **A.** Pol X with the palm on the left and C-terminal on the right side. **B.** Rotation of A. by 180 degree. Color code for the classification of amide peak changes in the ^{15}N-HSQC spectrum after addition of 5 mf $dA(pdA)_{11}dC(pdC)_6$ single strand DNA: orange indicates residues whose amide peaks are no longer detected at their original position at 5 mf; cyan, peaks that experience a large shift, green: reduced intensity (dark green) and shifted peaks(light green), magenta: no appreciable change and white: not assigned residue or proline**.**

Mapping the changes to the pol X structure showed an affect throughout the protein (Figure 4-15), with the residues, whose resonances had disappeared completely at 5 mf producing an L-shaped surface in the subdomain interface and C-terminal, similar to the one reported for the 36-mer and 44-mer ds-DNA (Figure 4-15, orange). The data for this titration is listed in Table E-3, Appendix E.

$dA(pdA)_{11}dC(pdC)_6$: $dA(pdA)_{11}$ Template-Primer ds-DNA

A final experiment in this series was performed using the $dA(pdA)_{11}dC(pdC)_6$: $dA(pdA)_{11}$ template-primer ds-DNA sequence. The 18-mer was annealed in the presence of the 12-mer by heating the sample to 80 ^{0}C and slowly cooling it. Upon addition of 0.25 mf of this duplex, large numbers of amide peaks were already exchange broadened and more resonances were no longer detectable as the duplex concentration was raised to 0.5 mf and 1.0 mf. At 1.0 mf, only 18 amide resonances were left identifiable. This observation was in analogy to the previously reported one for the double hairpin, gapped DNA at various lengths. A 5- and 13-fold excess of ddGTP was added to see, whether shifting the equilibrium would recover some correlation peaks, which was not the case.

Additional Attempts to Improve Spectral Quality for the Pol X-DNA Complexes

A number of additional attempts were made to improve the spectral quality of the pol X-DNA complex. Since none of them yielded the desired improvement, the attempts are listed here:

1) Pol X_{gpm} - 36-mer DNA: reduce NaCl concentration to 100 mM through dialysis; loss of DNA signals.

2) Pol X_{gpm} - 16-mer – ddGTP ternary complex: reduce NaCl concentration to 100 mM in spin filter; spectrum not different.

3) Pol X_{gpm} - 44-mer DNA: TROSY measurement; no improvement or additional peaks

4) Pol X_{gpm} - 44-mer DNA: reduce temperature to 278 K; all but some sidechain signals disappeared completely

5) Pol X_{gpm} - 44-mer DNA: increase temperature to 303 and 308 K; heavy degradation off pol X

6) Dissolve pol X_{gpm} in 20 mM cacodylic acid, add to lyophilized 44-mer, 12-fold excess; no precipitation, but also no additional peaks visible in HSQC spectrum

7) Dissolve pol X_{gpm} in 20 mM cacodylic acid, add 44-mer in same buffer; massive precipitation

8) Dissolve pol X_{gpm} in gpm-buffer, add to lyophilized 44-mer in same buffer; massive precipitation

9) Pol X_{iz} in 20 mM sodium acetate, 20 mM $MgCl_2$ and 0.02% NaN_3, add 36-mer; immediate precipitation

10) Pol X_{gpm} - 44-mer DNA: use 500 mM KCl instead of NaCl; no difference

Discussion and Conclusions

Complex with Nucleotide

As discussed in Chapter III, Maciejewski et al (*46*) had reported similar values for the purines, but not the pyrimidines. In their work they reported K_d values of 1.6, 11.9 and 4.0 μM for the displacement of dGTP, dTTP and dCTP, respectively, as determined by a fluorometric competition assays. This compares to 2.4, 40.2 and 49.7 μM for dGTP, dTTP and dCTP, respectively, determined by NMR titration. The K_d values for dGTP compared well, as was the case for all three methods, NMR, ITC and fluorescence. This was not the case for the pyrimidines. The K_d values differed between the NMR and fluorescence based determinations by a factor of 4 to 12 and no values were obtained by ITC. While an indirect competition method was used for the fluorescence based determination, the other two methods were direct measurements. ITC monitors the overall reaction heat and shows exothermic reactions for the purines, while no signal was observable for the pyrimidines. With the NMR measurement individual amide resonances were used to directly monitor the complex formation. With the residue specific information provided by the NMR titrations it was possible to describe a specific area of the molecule that is affected upon the addition of nucleotide. The data showed, that the addition of any of the four nucleotides induced major chemical shift changes in

the subdomain interface of the molecule, likely including a conformational change, because about a third of all residues in pol X were affected upon addition of the nucleotide. Such an extensive effect would seem unlikely for the nucleotide binding interaction alone. It was shown that the majority of the affected residues were concentrated in an area close to the aspartic acids D49, D51 and D100 and span the two subdomains, in particular α-helices B, D and E and the β-strands β 4-6, and 8-10. The forthcoming structure determination of the complex should answer the question on how much conformational change is involved, what the new confirmation looks like and what its implications towards catalysis and fidelity are.

Complex with Single- and Double-Stranded-DNA

The NMR titration of pol X with DNA fragments indicated that the exchange was in an intermediate time regime on the NMR time scale compared to the slow exchange observed for the nucleotides. During the titration with the nucleotides peak intensities in the NMR spectra remained constant, and the resonances experienced some chemical shift perturbation upon complex formation. When adding oligonucleotides though, the peak intensity was drastically reduced and a majority of resonances were no longer observable at stoichiometric ratios. The bigger ligand molecule formed an increased

molecular weight of the complex, which would reduce the NMR signal, but the observed reduction exceeded this expected range. All attempts to move the system out of the intermediate exchange regime were unsuccessful.

The residues affected by the 36-mer and 44-mer additions were mapped to the pol X structure using the sub-stoichiometric titration results, where intensity changes of the amide resonances could be monitored. This surface map revealed comparable interactions between the oligonucleotide and pol X. An L-shaped area of chemical shift perturbations was identified and involved some of the previously pointed out (*46, 70*), positively charged residues on αE, L131, L132, R125, and R127. Further interactions were seen on the β-sheet β11-12, which contains the positively charged residues, N134, L136, N138, Q139, L144, N145, and Q146. More residues were affected in the subdomain interface including αD, β4, β6, and, β10. This area forms a shallow cleft in the free pol X structure. No positively charged residues are located there. This would suggest, that this area might not be directly involved in DNA binding, but some conformational change occurring between the two subdomains causes the perturbation. The 36-mer and even more the 44-mer do show additional changes on K63, located on αC, K79 and K85 on β9, which are located close to αC. The conclusion was drawn, that the ds-DNA spans from the C-terminal side of αE through the vicinity of β11-12 over to

αC. The data suggested, that the DNA was positioned between the C-terminal side of αC and β8, which forms a cleft. These explanations would be consistent with preliminary observations made by Maciejewski et al (*46*) and support the hypothesis made by Jezewska et al (*70*) that αE and αC would be involved by longer oligonucleotides. The same group (*71, 72*) recently proposed models for ds-DNA and ds-gapped-DNA as well. Their fluorescence work led to the proposal that the oligonucleotide was located either in the subdomain interface or along αE, while the gapped DNA would move along αE, loop around in the vicinity of the C- and N-terminal of the enzyme and align along the subdomain interface along αD to leave the binding site near αC.

The data presented here suggests three possibilities: 1) The DNA is positioned on the binding interface involving the C-terminal end of αE, β-strands 11 and 12, and parts of the palm domain involving the main β-sheet and αC, but not the N-terminal end of αE as suggested by Jezewska (*71, 72*). Sharp bends in DNA have been reported for the ternary structures of pol β in the presence of gapped and nicked DNA (*27*). Based on these structures, the binary complex with gapped DNA on pol X would resemble striking similarities with pol β. The DNA from this ternary pol β structure (*27*) (PDB code 1BPY) was modeled onto the free pol X structure (PDB code 1JAJ) (*46*) to visualize the proposed complex and is shown in Figure 4-17 (A). This structure does not explain the

perturbations seen in the subdomain interface. 2) To explain the perturbations in the subdomain interface, two processes can be envisioned: either a conformational change or a repositioning of the DNA. Structures of pol β indicate (*27, 40, 41*) that pol β undergoes substantial conformational change depending on whether it is present in the free, DNA binary complex or a DNA ddCTP ternary complex (Figure 4-16, A-C).

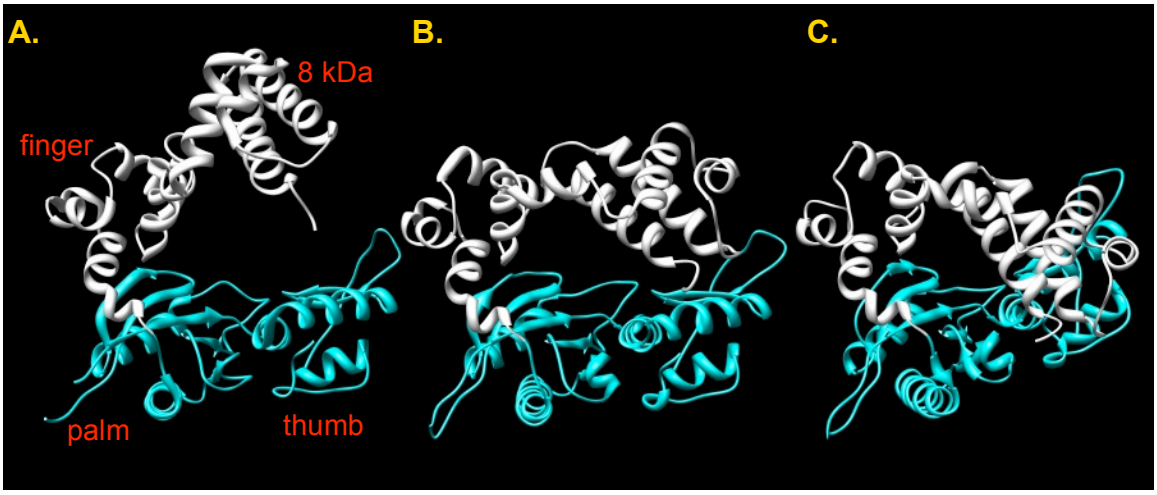


Figure 4-16: Conformational changes in pol β structures. Only the protein structures are shown, magenta indicates the palm and thumb domain, which are the two domains present in pol X, finger and 8 kDa domains are colored in white. **A**. Free pol β (*41*), (PDB code 1BPD). **B.** Binary complex of pol β with 7 bp complimentary DNA (*40*), (PDB code 9ICJ). **C.** Ternary complex of pol β (*27*), (PDB code 1BPY).

In analogy to these conformational changes observed for pol β, the complex formation of pol X with DNA might also undergo some conformational change, and the

perturbations observed in the subdomain interface would indicate a motion between the tow domains (Figure 4-17 C). The missing finger and 8 kDa domain in pol X also opens the possibility of an additional change. It is conceivable, that the DNA swings down from the position described above, interacting with the subdomain interface as indicated in Figure 4-17 (B). This subdomain is not as rich on charged residues as the one shown in Figure 4-17 (A) and contains the nucleotide-binding site. A positioning of the DNA over the subdomain-binding site would not cause the perturbations observed in the αC helix and the upper part of the β-sheet 8-10. Therefore a move between the two positions is proposed, in which the DNA would interact with all these sites in pol X, which also would be consistent with the extensive exchange broadening of the resonances observed in the HSQC spectra. 3) A combination of the DNA movement and conformational change together cannot be ruled out and would also be consistent with the exchange broadening seen in the NMR spectra.

Using short ss-DNA oligonucleotides, the HSQC spectra retained a majority of pol X resonance peaks at stoichiometric concentrations or higher. Chemical shift perturbations were observed with the ss-DNA compared to the exclusive observation of exchange broadening as was the case for the ds-DNA. This indicated a different exchange behavior for the ss-DNA.

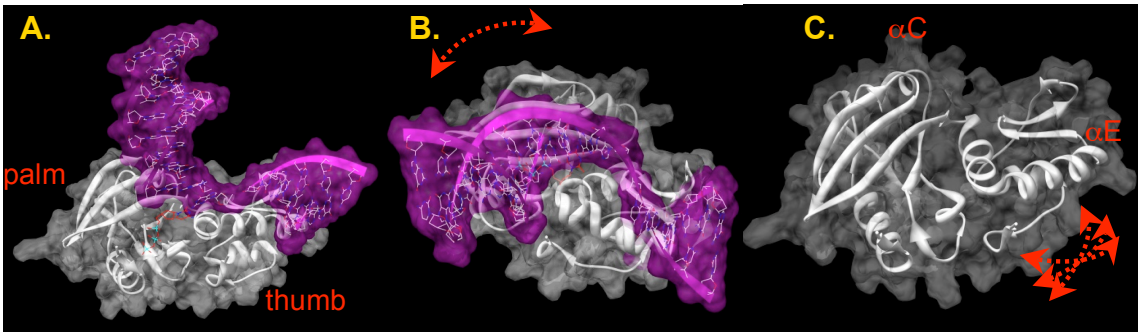


Figure 4-17 (A-C): Gapped DNA modeled on to free pol X structure visualizing potential complex formation modes in accordance with the NMR mapping results. A and B. Gapped DNA structure from ternary pol β complex (*27*) (PDB code 1BPY) modeled on pol X structure (*46*) (PDB code 1JAJ) in two positions covering the binding interface (A) αC, αD, αE and β strands 11,12 and (B), αE, β strands 11,12 and the subdomain interface. C. Indicates the conformational movement between the two domains.

Through the study of small ss-DNA fragments it was possible to identify the C-terminal domain as the initial contact site for DNA. As was the case for ds-DNA, the C-terminal side of αE and the loop connecting β11 and β12 showed first a chemical shift perturbation and with longer DNA fragment, the effect spread over to αC. The subdomain interface was always affected, more extensively for larger ss-DNA fragments.

One noteworthy observation was made for tryptophan 92. This residue did not experience any appreciable chemical shift perturbation or exchange broadening during the titration experiments. Its remote location did explain this behavior. Furthermore this

observation was consistent with the small response that was detected during the fluorescence measurements.

CHAPTER V

NMR ASSIGNMENT OF ASFV POL X AS BINARY COMPLEX WITH DEOXYGUANOSINE TRIPHOSPHATE

Introduction

NMR spectroscopy has proven to be a very powerful tool for the analysis of organic molecules and their structure or the structure determination of biomolecules. Its capability to provide atomic level information required in deriving high resolution structures for biomolecules is only matched by the x-ray diffraction method. Either method has its unique strengths and shortcomings and therefore the most suitable technique must be chosen for any given task. Some of the strengths of the NMR based biomolecular structure determinations lie in the fact that molecules are examined in solution, which is typically their natural environment. This also implies that molecular motions, inter- or intra-molecular interactions such as protein-drug, protein-protein or protein-DNA interactions, changes due to temperature, pH or other environmental parameters can be readily observed. Similar observations might be more challenging through diffraction methods using crystallized molecules. Producing suitable crystals of biomolecules represents frequently the bottleneck for that method and there is

sometimes additional concern that the crystal packing could distort the free structure. As any method, NMR also suffers from limitations that are mainly related to the size of the molecule. Currently, the maximum limit approaches 40-60 kDa with the most sophisticated NMR methods. NMR measurements are of inherent low sensitivity, although constant progress is made on extending these limits.

Other spectroscopic methods, such as fluorescence or UV spectroscopy also provide structural information for biomolecules. Although based on similar phenomena, NMR spectroscopy differs in a number of aspects. Here a photon is absorbed which promotes a nuclear spin from its ground state to an excited state. To do so, the spins have to be aligned in an external magnetic field, which is uniquely different to other methods. The field strength directly impacts the energy difference between the two states and the resulting signal intensity. A second distinction lies in the lifetime of the excited state. It is in the order of 10^9 times longer for NMR than for the methods using electronic excited states, resulting in much more narrow spectral lines. Dealing with signal linewidths of less then a Hertz (Hz) in organic molecules and a few Hz for biomolecular molecules requires near perfect magnetic field homogeneity. As an example, measurements at 800 x 10^6 Hz (800 MHz) proton frequency or 18.8 Tesla

require homogeneity with a sub-hertz resolution. Consequently the complexity and cost of NMR instrumentation is very high.

Brief History

The basic principle of NMR measurements lies in the intrinsic nuclear magnetic moment of each spin and was first measured by Rabi et al 1938 (*110*). Felix Bloch and Edward Mills Purcell independently observed 8 years later (*111, 112*) transitions caused by resonance during a radiofrequency sweep in a static field and received the Nobel prize in physics for this discovery in 1952. Richard Ernst and Weston A. Anderson 1965 (*113*) introduced pulsed excitation and signal averaging followed by a Fourier transformation of the time domain data which was much faster compared to the radio-frequency sweep and allowed for signal averaging by multiple scans, giving rise to stronger signals. This opened the way to more complex experiments and the first two dimensional (2D) NMR spectrum was introduced by Jean Jeener (*114*) and implemented by the Aue, Bartholdi and Ernst 1975 (*115*). Richard Ernst received a Nobel Prize in Chemistry 1991 for his contributions. Although higher dimensionalities were expected to be possible then, it took another 12 years, before Griesinger, Sørenson and Ernst (*116*) proposed 3D NMR experiments in 1987. In the meantime, Kurt Wüthrich (*117*) published

the first NMR derived assignments on a protein leading to structural determination of proteins by NMR. His book “NMR of Proteins and Nucleic Acids” (*118*) remains a classic in the field. This work was also recognized by the Nobel Prize in Chemistry 2002, emphasizing the importance of NMR based methods in modern science

Fundamental Aspects of NMR

The following few paragraphs discuss briefly the fundamental aspects of NMR and provide some understanding of considerations required during experimental set-up, as well as the information content gained from an NMR spectrum. Many excellent books and articles are available with a wealth of detailed information and only a few selected ones are referenced here (*118-124*)

Nuclei with an odd mass number, such as ^{1}H, ^{13}C, ^{15}N or ^{31}P exhibit a nuclear dipole moment arising from the spin angular momentum due to their unpaired proton. The same can be said for nuclei with an even mass, but odd charge, like ^{2}H or ^{14}N. All other nuclei, most importantly ^{12}C as the most abundant naturally occurring carbon isotope, do not have an angular momentum and therefore cannot be observed by NMR. The spin angular momentum is quantized and the quantum states depend on the spin quantum number I of the observed nuclei. For spin one-half nuclei (I = ½), which

includes all biologically interesting nuclei (^{2}H is an exception), the two quantum states (m) are + ½ $\hbar$ and - ½ $\hbar$ and are related to the magnetic moment of a nuclear spin $\vec{\mu}$ by the gyromagnetic ratio γ whose values are listed in Table 5-1. The product of γ and the magnetic field strength result in the Larmor frequency, which is the observed resonance frequency of a nucleus measured in the NMR spectrum.

Table 5-1: Physical properties of NMR active nuclei used in biomolecular systems. Listed values are adapted from the periodic table of the NMR Guide (Bruker Biospin, Germany). The resonance frequency for all nuclei is listed at 11.744 T and is proportional to the gyromagnetic ratio, the absolute sensitivity is a product of the natural abundance times relative sensitivity, which is the sensitivity at constant magnetic field and equal number of nuclei.

Nucleus	gyromagnetic ratio γ [10^7*rad/s*T]	Resonance frequency at 11.744 T (500 MHz for ^{1}H)	spin quantum number *I*	Absolute sensitivity (^{1}H = 1.00)	Natural Abundance
^{1}H	26.753	500.000 MHz	½	1.00	99.98%
^{2}H	4.106	76.753 MHz	1	1.45 10^{-6}	1.5 10^{-2}%
^{13}C	6.728	125.721 MHz	½	1.76 10^{-4}	1.108%
^{15}N	-2.712	50.664 MHz	½	3.85 10^{-6}	0.365%
^{31}P	10.841	202.404 MHz	½	6.63 10^{-2}	100.0%

The energy of interaction between the dipole field and the external field is called Nuclear Zeeman Hamiltonian and expressed as:

$$\hat{H} = - \mu_z \bullet B_{0z} = - \gamma \hbar B_{0z} m_z \qquad [5.1]$$

where the magnetic field along the z-axis is B_{0z}. The energy difference between the ground and excited state is calculated from Eq. 5.1,

$$\Delta E = \hat{H}_{\beta} - \hat{H}_{\alpha} = -\gamma \hbar B_0 \qquad [5.2]$$

Since the energy of a photon is $E = \hbar \nu$, the resulting Larmor frequency can be extracted from Eq. 5.2:

$$\nu = \frac{\Delta E}{\hbar} = - \frac{\gamma B_0}{2\pi} \qquad [5.3]$$

Averaging these observations over multiple spins present in a NMR sample results in a macroscopic magnetization M_0 that represents the population difference between the two states and ultimately determines the observable NMR signal intensity. As can be seen from the Boltzmann distribution in Eq. 5.4, the population difference depends on the magnetic field, gyromagnetic constant and the temperature T.

$$\frac{N_{\beta}}{N_{\alpha}} = e^{\Delta E/k_B T} = e^{-\gamma \hbar B_o / 2\pi k_B T} \qquad [5.4]$$

where k_B is the Boltzmann constant. This relationship explains why NMR is a relatively insensitive measuring technique, higher fields are advantageous and the use of a high γ nucleus is important.

Equation 5.3 indicates that the frequency is directly dependent on the main magnetic field and γ. This relationship would provide only a single line for the NMR spectrum of a given nucleus. The resolving power of the technique stems from the fact

that the effective field for each nucleus B_{eff} differs from the main field depending on the presence of electrons that surround the nucleus and either expose or shield it to various extent. Therefore each nucleus has a characteristic shielding factor σ giving it a unique resonance frequency in the spectrum.

$$B_{eff} = (1 - \sigma) B_o \qquad [5.5]$$

where σ is the chemical shift tensor. Shielding and the resulting frequencies are compared to a standard frequency given by the tetramethylsilane proton signal. The difference between two frequencies is field dependent and for universal comparison converted to the *chemical shift* scale in ppm, which is defined as:

$$\delta = \frac{\nu - \nu_o}{\nu_o} \times 10^6 \qquad [5.6]$$

with units of parts per million [ppm].

Scalar Couplings

Another important concept arises from the interactions of observable nuclei via their bonding electrons. This scalar coupling, also referred to as J-coupling, involves NMR active nuclei that are connected by chemical bonds. The concept might be easiest explained on an example, where the observation of one nucleus, *I* adjacent to another, *S*, is considered. Both spins are present in two spin states as shown above. The

observed spin *I* is affected by the two states of *S* and as a result the bonding electrons become polarized from the magnetic dipole *S* which either increases or decreases the magnetic field on the observed *I* nucleus. Since the population difference between the two spin states on *S* is approximately equal (Eq. 5.4), half of the observed signals *I* experience an increase of the local magnetic field, while it is decreased for the other half. The signal *I* is therefore split into two lines equally shifted to higher and lower frequency around the original frequency with the separation between the two lines representing the value for the J-coupling. The same arguments are true for the observation of spin *S*, resulting in equal splitting of the resonances for both nuclei. The magnitude of the coupling is dependent on the gyromagentic ratio of the nuclei involved and hence independent of the field.

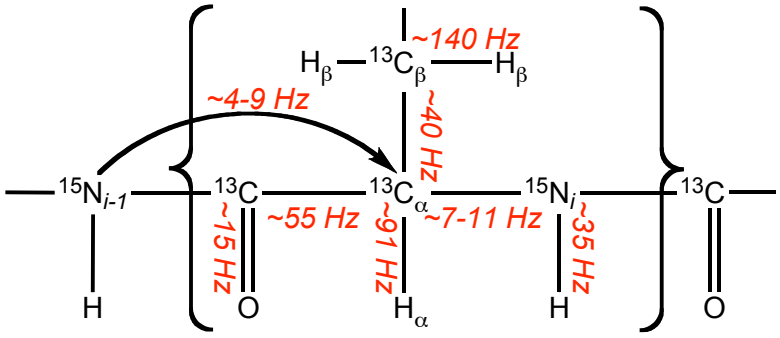


Figure 5-1: Important scalar coupling values between nuclei used in fully ^{13}C carbon and ^{15}N nitrogen labeled protein experiments. Exact values differ for each bond, but the average values indicated are commonly used.

A multitude of scalar interactions can be observed dependent on the number and type of nuclei involved or their frequency difference. Observations of scalar couplings are also possible through multiple bonds. Figure 5-1 shows the most important scalar coupling values used in protein NMR experiments.

Dipolar Couplings

While scalar couplings are indispensable for the resonance assignment of a molecule, dipolar couplings provide crucial information used for their 3D structure determination. Dipolar couplings arise when two nuclear dipoles are spatially close to each other and interact in the absence of a connecting covalent bond. The magnitude of the dipolar coupling depends on the strength of the magnetic field generated by one spin and the size of the magnetic moment of the recipient spin. Isotropic rotation of the molecule in solution is generally fast enough to completely average the angle dependence of the dipolar interaction contribution and consequently no splitting of the resonance lines are observed. This tumbling of the molecule generates a fluctuating magnetic field that stimulates not only single quantum transitions but also the generally forbidden zero- and double quantum transitions, as a mechanism for spin relaxation to the thermal equilibrium. The zero- and double quantum relaxation pathways, also called

cross-relaxation, change the population level of the coupled spins, which can then be detected in form of an enhanced signal. This effect was first observed by Overhauser (*125*) and is referred to as the Nuclear Overhauser Effect (NOE). Dependent on the molecular tumbling rate and therefore molecular size, zero- or double quantum relaxation is preferred, giving rise to either a signal enhancement or signal decrease. In order to give rise to an observable NOE effect, the two spins have to be within ~5 Å of each other. The closer the coupled spins are the more efficient is the cross-relaxation, leading to a stronger enhancement of the signal. The relation between intensity and distance of the coupled spins is proportional to $1/r^6$ and can be used to obtain distances. The NOE effect builds up over time and therefore experiments require a so-called mixing time to allow for sufficient NOE to be formed. For longer mixing times spin diffusion becomes more pronounced and has a considerable impact on the peak intensity. Spin diffusion occurs when two coupled spins have a third partner that is spatially closer to each of the other spins which allows for a more efficient magnetization transfer between the two spins than would be the case for these spins alone. Quantitative analysis of the internuclear distance is best done by a series of NOE measurements at varying mixing times. Most reliable distances are achieved when a full relaxation matrix analysis is used, followed by an exponential fit of the initial data. If a large number of long distance

constraints for a globular protein are available the use of a single mixing time, calibrated to a known internuclear distance, is not uncommon. A full analysis of individual resonances is often impossible due to the spectral complexity and sorting the resonances compiled for the entire macromolecule into classes of distances provides enough information to derive the molecular geometry.

Multi-Pulse Experiments and Gradients

Modern NMR experiments consist of a series of pulses and time increments ingeniously designed to select for the appropriate magnetization pathway and signal encoding of a specific experiment (*126*). Selection of the proper coherences was traditionally accomplished by elaborate phase cycles, which often led to a large number of scans to complete a full phase cycle. With the introduction of high quality gradient coils, coherence editing through pulsed field gradients was made possible. The suppression of unwanted signal, including artifacts, with gradients can often be achieved in a single scan, but in some cases has to be combined with a phase cycle. Regardless, such experiments are always shorter compared to their non-gradient counterparts still yielding equal or better suppression of the unwanted signals (*127, 128*). The reduced

number of scans per increment has a dramatic impact on the experiment time, particularly when more then 2 dimensions are recorded.

A gradient applies a spatially varying magnetic field to produce a characteristic phase shift on all signals. The phase shift depends on the quantum level of the spins. After the initial gradient, the signal of interest can be restored by a second, refocusing gradient, while the undesired signal remains dephased. Gradients play also a major role in the water suppression that is crucial in biomolecular NMR. As will be shown, many protein experiments rely on the detection of amide protons, which are readily exchanged in the ordinarily preferred, deuterated solvent. Therefore 95% H_2O will be used for most protein experiments and the ~100 M water signal has to be suppressed to make optimal use of the dynamic range of the receiver and to detect the protein signals at a level of 50 µM to 1 mM. Many water suppression pulse building blocks such as WATERGATE, "water-flip-back" (*101, 102*), excitation sculpting (*129*) or combinations of them rely on gradients to further improve the solvent suppression.

Indirect Detection

The first proteins were assigned using homonuclear 2D correlation spectroscopy, where the chemical shifts of proton spins were allowed to evolve for a defined time

period (t_1) during the pulse sequence, creating an indirect detection domain. A mixing pulse is used to transfer information from one spin to another as long as they are either coupled by a J-coupling or dipolar coupling. For each of those time increments, the direct acquisition time domain (t_2) is Fourier transformed followed by a second Fourier transformation for the second time domain. Examples for such 2D experiments include many flavors of the COSY (correlated spectroscopy) (*114, 115*), TOCSY (total correlation spectroscopy) (*130*) or the above discussed NOESY spectra (*125*). Resonance overlap posed a limitation for larger proteins. Utilizing the carbon and nitrogen heteroatoms to spread the resonance frequencies was appealing. Because of their low natural abundance (Table 5-1), proteins had to be fully labeled. But even at 100% enrichment, the sensitivity for carbon was 32 fold less then for protons and much worse for nitrogen with a factor of 300 due to its low γ and frequency. Consequently new techniques were developed, where in very general terms, polarization transfers between protons and the heteronuclei were utilized and at the end the proton signals were detected, leading to a relative sensitivity gain in the order of $\gamma_H^{5/2}/\gamma_X^{5/2}$ (*131-134*). The polarization transfer exploits the $^1J_{CH}$ and $^1J_{NH}$ couplings in a series of deliberately placed π and $\pi/2$ pulses on the proton and heteronucleus channel. While the presence of large scalar coupling results in short, well defined delay times, it is also critical for these

experiments to exactly determine a number of carbon and nitrogen pulse lengths. Magnetization transfers are not limited to the proton – heteronuclei but can also be imposed between two heteronuclei just as done for homonuclear correlations. With this in mind, it can be imagined, how the magnetization can be transferred in a well controlled fashion throughout a spin system of multiple nuclei utilizing the coupling constants as shown in Figure 5-1. The evolution for specific heteronuclei can be monitored before the magnetization is returned to the protons for a largest possible sensitivity gain. These are the bases of the common multidimensional experiments. Comprehensive reviews with many more important details for the execution of these experiments have been reported by Sattler (*126*), Bax (*135*) and Griesinger (*136*).

Three-Dimensional (3D) Experiments

The concept of 3D NMR is similar to the one for 2D NMR, where instead of a series of 1D experiments modulated through a second delay time with evolving magnetization, a third evolution time is introduced. This means that multiple 2D experiments with varying evolution times need to be acquired, leading to large data matrix. Processing these data sets initially posed a computational challenge, which is overcome by the speed of modern computers. Limitations imposed by the time required

for the data acquisition remain, but much effort in the field is focused on overcoming them.

The advantage of a higher dimensional spectra lies in the separation of overlapping resonances by correlating them with other, better resolved nuclei and additionally spreading them into other dimensions. Carbon and nitrogen nuclei are utilized to spread the frequencies in protein measurements. A large number of such triple resonance pulse sequences are available and a common nomenclature follows the following scheme: The nuclei are listed in the order the frequency labeling occurs in the pulse sequence. If the magnetization is passed through a spin with no frequency labeling, the spin is listed in parenthesis. Additionally if an experiment begins on the amide proton, the name starts with ‘HN', implying an ‘out-and-back’ experiment, where the magnetization is returned to the amide proton for detection as is the case for the HN(CO)CA. Otherwise the magnetization starts somewhere else and the amide is only involved during detection, such as in the CBCA(CO)NH experiment. A schematic representation of the CBCA(CO)NH experiment is shown in Figure 5-2 and illustrates several principles of a 3D experiment. A 3D spectrum is commonly analyzed by examining individual 2D planes in any dimension of interest. Projections as shown on the right side in Figure 5-2 combine all resonances from a third dimension to a 2D plane.

It is evident from this example that more planes in each dimension increase the resolution of individual correlation peaks and the ^{15}N dimension allows to resolving overlapping carbon resonance frequencies as depicted for planes 2 and 3.

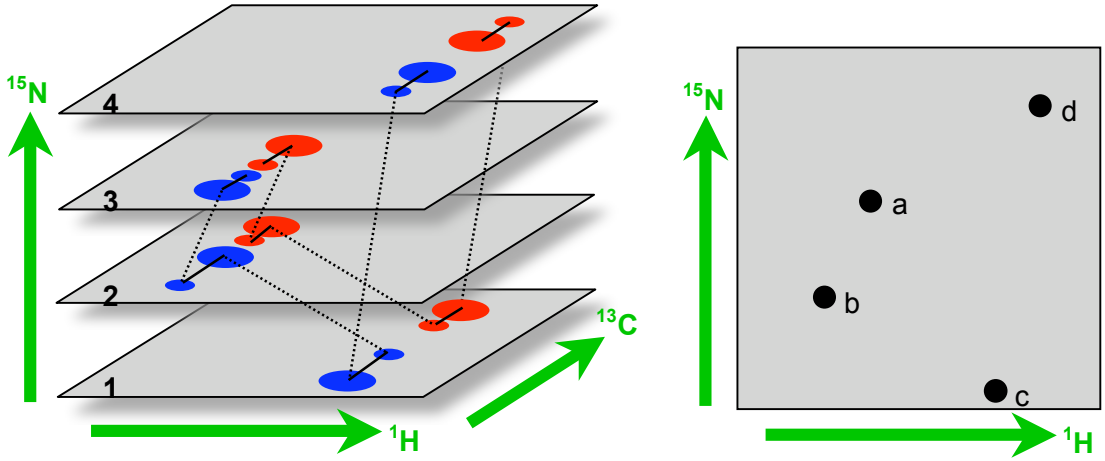


Figure 5-2: Schematic representation of a 3D experiment. The example is based on a CBCA(CO)HN experiment. Left are individual ^{13}C-^{1}H 2D planes with C_α and C_β chemical shifts indicated in blue and red, respectively. Dotted lines denote correlation from either HN_i or HN_{i-1} to the common $C_{\alpha/\beta(i-1)}$ atoms while the full line denotes correlations for a given HN_i to the intra-residue $C_{\alpha/\beta(i)}$ (larger peaks) and inter-residue correlations $C_{\alpha/\beta(i-1)}$ (the smaller peaks). The spectrum on the right side is a ^{15}N-^{1}H 2D projection indicating the position of the four sequential amide peaks.

Materials

NMR Sample Preparation

Uniformly $^{13}C/^{15}N$- and ^{15}N-labeled protein samples were prepared as described in Chapter II. The NMR experiments were performed on Bruker DRX 500, 600 or 800 MHz spectrometers equipped with a cryoprobe, at 20^0C. NMR samples for the free pol X contained 500 µL of 0.2-1 mM pol X, 20 mM PIPES, 20 mM $MgCl_2$, 2-10 mM DTT, 0.02% NaN_3, 300 mM NaCl, and 5% v/v D_2O at pH 6.5 in 5 mm NMR tubes. The sample for the binary complex with dGTP contained 350 µL of 0.8 mM pol X, 20 mM PIPES-d_{18}, 20 mM $MgCl_2$, 2 mM DTT-d_{10}, 0.02% NaN_3, 500 mM NaCl and 4 mM dGTP at pH 6.5 in a 4 mm NMR inner tube, 80 µL D_2O, and 3 mM TSP-d_4 in the 5 mm, outer tube. This dual tube arrangement is described in Chapter VI and allowed measurements on a cryoprobe in the presence of 500 mM NaCl, which greatly stabilized the polymerase in solution.

Results

Oxidized and Reduced Form of ASFV Pol X

A close look at the HSQC data collected on the free pol X indicated that some of the amide resonances were shifting over time. The chemical shifts from a freshly

prepared sample of free pol X were in good agreement with the data deposited by Maciejewski et al (*46, 81*) at the Biological Magnetic Resonance Bank (BMRB). A shift from their original position was observed for some amide resonances over the course of several months. Sixteen peaks with a small chemical shift perturbation, e.g. a change within the linewidth of a peak were identified as followed: Y22, N23, L30, V50, V56, K60, S75, F76, S77, I90, Q98, A106, F116, L130, K131, L163. Larger changes were seen for 29 residues: L52, L53, E58, L62, K63, H64, V65, N68, I69, I71, V78, K79, V80, C81, G82, E83, R84, K85, C86, L88, F89, D100, L101, F102, T103, A104, E158, G164. Mapping these residues on the pol X structure showed an accumulation of the changes in β-sheet 6 and 9-10 as seen in Figure 5-3.

When not used for NMR measurements, the sample was stored without any further manipulations at 4^0C and contained 5 mM DTT. An explanation of the chemical shift perturbation was a slow oxidation of the cysteines to form a disulfide bond. The structure shows that both cysteine residues are within 7 Å and such a reaction can occur. This hypothesis was supported by the second published structure of free pol X by Showalter et al (*47*). That structure contained a disulfide bond even in the presence of 1 mM DTT, which was the most significant difference between this structure and the one reported by Maciejewski et al (*46*).

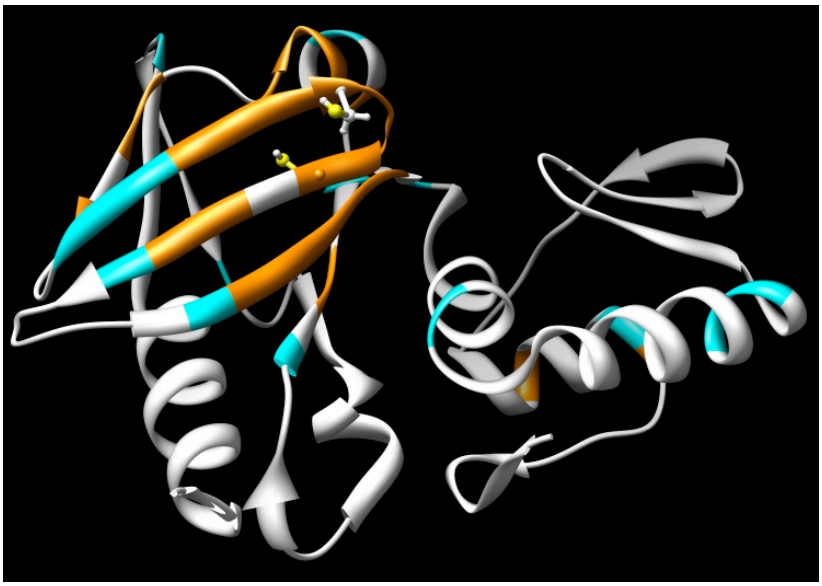


Figure 5-3: Amide chemical shift change on free pol X over time. Observed changes after 8 months in the NMR tube were mapped on the pol X structure (PDB code 1JAJ). Magenta indicates residues whose amide shifts were changing less than 0.01 ppm, orange indicates larger perturbations. Atoms of the cysteine residues 81 and 86 are shown.

The NMR buffers for both systems were different. Maciejewski et al (*46*) used 20 mM PIPES, 10 mM DTT, 0.02% NaN_3, 500 mM NaCl, and 0.5 mM 4-(2-aminoethyl)benzylsulfonyl fluoride (AEBSF), a protease inhibitor, at pH 6.5 and Showalter et al (*47*) described the buffer as 50 mM borate, 50 mM KCl 1mM DTT, and 1 mM EDTA at pH 7.5. The 10-fold difference in the DTT concentration might be an explanation for the observation of the two cysteine forms. Additional effects might have originated in the sample purification and preparation. There is precedence that DNA

viruses possess conserved oxidoreductase systems promoting cytoplasmatic disulfides during viral infection such as reported by Moss et al (*137, 138*). Furthermore the pol X host cell was identified to be the swine macrophage (*56*), a cell type that produces reactive oxygen species (ROS) and presents an oxidizing environment (*139*). Therefore the oxidized, disulfide form of pol X could not be ruled out as the biological relevant species. Another possibility was that both forms would pose an active form with different functions exhibited by the different forms in various parts of the cell as mentioned in the report by Senkevich et al (*137*). Interestingly, discrepancies regarding the fidelity of pol X as discussed earlier were reported in the literature (*9, 59, 61, 63*) and the oxidation form might play a role in that, but the oxidation form was never reported and possibly not considered in any of these studies, although the two different structures were known.

A comparison of HSQC spectra at varying times after sample preparation was done for the pol X - dGTP complex and revealed a similar behavior as described for the free pol X. Amide resonances of the following 32 residues showed chemical shift change: H13, V50, L53, I55, E58, L61, K63, V65, N68, I69, I71, F76, S77, V78, G82, R84, K85, C86, F89, Q98, L99, L101, F102, T103, A104, A106, E108, K109, H115, K144, Q146, E156, E158. They again were located on the β-sheet containing the two cysteines, but some additional changes were seen in the loop connecting β4 with α-helix

D and some changes on the N-terminal side of α-helix F. It was not surprising to see these small deviations since the polymerase was anticipated to undergo some degree of conformational change judged by the perturbations caused upon addition of the deoxynucleotide triphosphate, as shown in Chapter IV.

Figure 5-4 displays a change of the amide correlation in the ^{15}N-HSQC spectra for residues V65 and K85 over time. The shift changes were comparable for the free pol X and the dGTP complex. The amide resonance for the directly involved C81 and C86 also shifted. Maciejewski et al (*46*) reported the following amide shifts for the reduced, free pol X: for $C86_{red}$ ^{15}N=125.6 ppm, H^N=9.19 ppm and $C81_{red}$: ^{15}N=126.5 ppm, H^N=9.03 ppm. The same resonances were present in the reduced form measured, but the peaks, mainly C86, did not superimpose with any peak of the oxidized spectrum, indicating a peak shift between the two forms. C81 was in a cluster with other peaks, making it difficult to analyze it in the HSQC spectrum. The chemical shifts for the oxidized form of $C86_{ox}$ were assigned to: ^{15}N=128.5 ppm, H^N=9.21 ppm, but no assignments were available for C81. The shifts for the pol X-dGTP complex were assigned as $C86_{ox}$ ^{15}N=130.6 ppm, H^N=9.45 ppm and $C86_{red}$: ^{15}N=127.2 ppm, H^N=9.01 ppm.

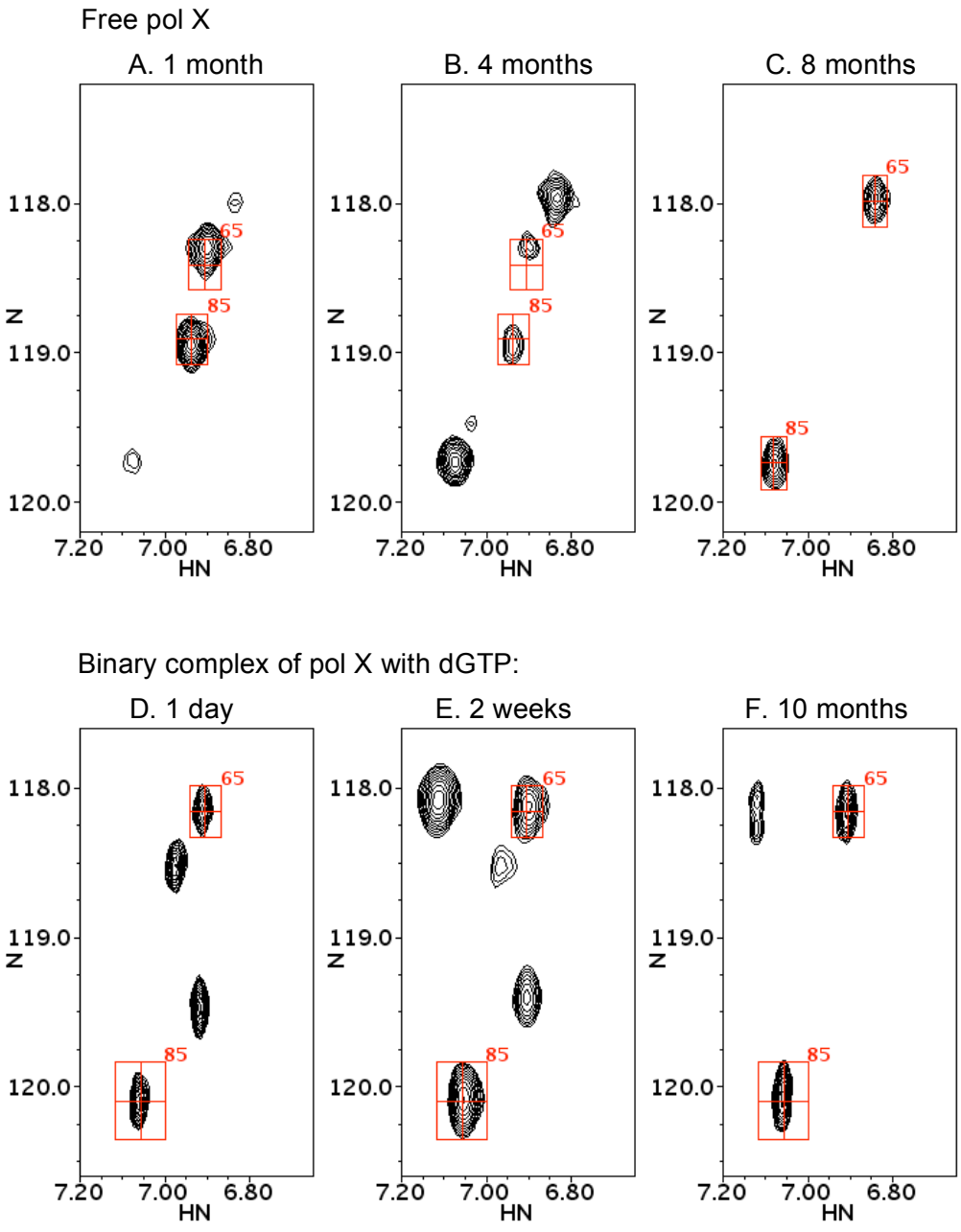


Figure 5-4: Chemical shift change for residues V65 and K85 over time. **A-C:** Series of ^{15}N-HSQC spectra on the same sample after 1 month, 4 month and 8 month, respectively. The peak labels for **A.** and **B.** reflect the shifts reported by Showalter et al (*47*). **D-F:** A series of spectra of the pol X - dGTP complex after 1 day, 2 weeks and 10 month, respectively. The resonance positions reflect assignments for the complex. The additional resonance at $H^N/^{15}N$ of 7.09, 118.08 ppm is a folded in peak.

It appeared that the changes were slower for the free form, but the buffer for that sample contained 5 mM DTT and 300 mM NaCl, while the sample with the complex contained only 2 mM DTT and 500 mM NaCl. If oxidation was the reason for the change, the difference in DTT could readily explain a faster oxidation to occur in the sample with less DTT. Alternatively, the presence of dGTP might have promoted the oxidized form or vice versa.

It was important to validate the hypothesis of a redox reaction as the cause for the peak shifts. To do so, a sample containing the pol X - dGTP complex was used, that exhibited only one set of resonances as depicted in Figure 5-5 D since the sample was 22 month old, it was presumed to represent the fully oxidized form and upon addition of reducing agent, the other resonances were expected to become visible. The reducing agent DTT-d10 was added to the sample to a final concentration of 100 mM. Even with such an excess of DTT, it took more then 32 hours for a full conversion to the other set of peaks described above which is shown in Figure 5-5.

The initial and final resonances line up well with the two previously observed forms, Figure 5-5 (E) and were strong evidence of the hypothesized redox reaction, involving the cysteines 81 and 86 in a reversible disulfide bond formation.

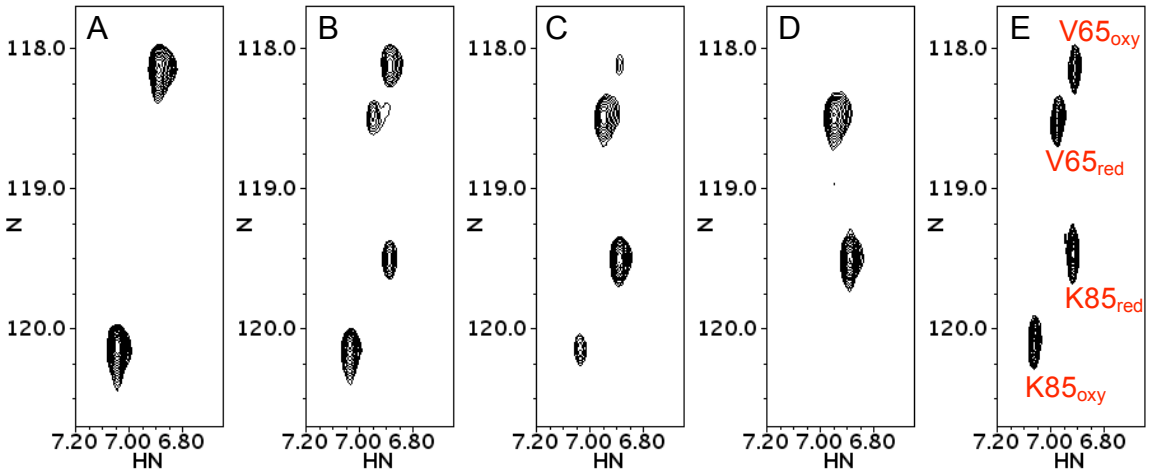


Figure 5-5: Monitoring the resonance shifts after addition of 100 mM DTT-d10 to an NMR sample containing the pol X – dGTP complex. **A.** Spectrum before DTT addition, **B-D:** 10, 32 and 82 h after DTT addition. **E.** Reference spectrum containing the two forms also shown in Figure 5-4 D.

In light of these findings it was essential to examine the biological activity between the oxidized and the reduced form. An NMR sample containing ~0.8 mM polymerase was divided into two fractions. A final concentration of 100 mM DTT-d10 was added to one fraction to reduce the disulfide bond. The reduction was monitored by NMR, and no oxidized components were visible after 4 days. The other fraction was completely oxidized by dialyzing the sample against a DTT free, but O_2 saturated buffer, as observed by NMR.

The nucleotide incorporation assays showed a large difference between the reduced and oxidized forms of pol X. In a first comparison the DTT concentration was

kept at 100 mM to make sure only reduced pol X was present. Even at a four times higher concentration of polymerase, no incorporation was detected after 2 hours as shown in Figure 5-6 (B). The assay was repeated under the same conditions but the DTT concentration for the sample containing the reduced form was reduced from 100 mM to 10 mM. Comparable results were seen for the reduced sample containing 10 mM and 100 mM DTT. At 1 μM polymerase concentration (Figure 5-6, C and D), the oxidized species fully incorporated the nucleotide within the first minute, going on to an extension with a second nucleotide within the two hours period, while the reduced form showed even at 1 μM concentration only signs of incorporation after 5-10 minutes.

The sample containing the reduced form was dialyzed with a N_2 saturated buffer and kept under anaerobic conditions for the duration of the experiment. Figure 5-6 (E) shows that the reduced pol X had very little activity. A semi quantitative analysis of the incorporation shown in Figure 5-6 (A) yielded a k_{cat} of 0.54 min^{-1} for the oxidized form corresponding to the k_{cat} for G:C of 0.34 reported by Lamarche (*59*). Misincorporation activity was tested for both forms. No misincorporation was observed for the reduced form and only a very low intensity band was seen for dATP for the oxidized form of pol X. These assays showed that the activity of pol X depended on their redox state, which might explain some of the activity differences reported and observed for pol X.

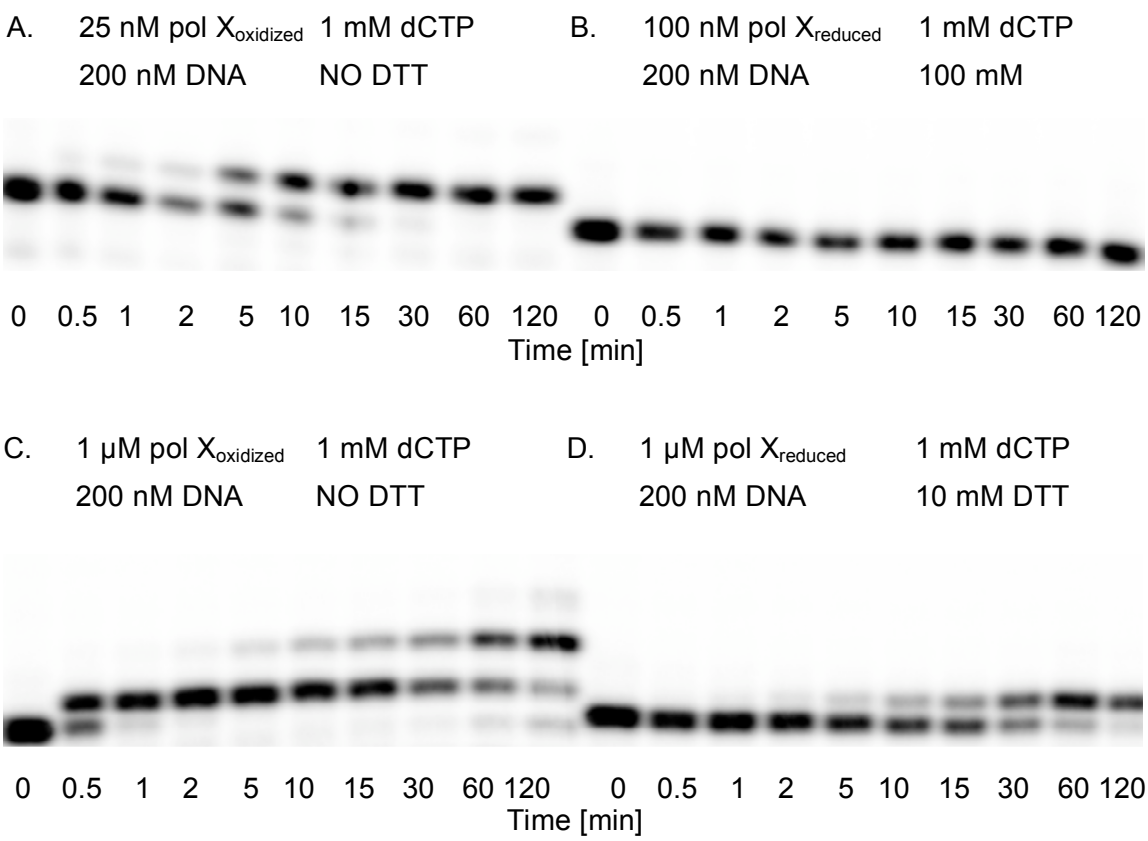


Figure 5-6: Activity tests on ASFV pol X in the oxidized (left panels) and reduced (right panels) form. Concentrations of polymerase, template primer DNA, dCTP, and DTT and incubation times are indicated for each panel. Pol X dilution buffer was 20 mM Tris•HCl, 500 mM NaCl, pH 7.8, reaction buffer: 50 mM Tris•HCl, 150 mM NaCl, 100 µg/mL BSA, 5% glycerol, 20 mM $MgCl_2$, pH 7.8, incubation was performed at 37^0C, 20% PAGE / 8 M urea were used. Template-primer: 5'-TCA TGG AAT CCT TCC CCC-3', 5'-GGG GGA AGG ATT C-3'.

Backbone Assignment

The protein backbone assignments for the carbon, nitrogen and amide proton resonances represent the first step for any structural work on a protein. The resulting knowledge of all amide correlations is an important result of this process and is often all that is needed, especially when the goal is to observe interactions of the protein with any other molecule. As has been shown in Chapter IV, the chemical shifts of the amides are sensitive to any change of the chemical environment for the respective amide and were used for the mapping of changes during the NMR titrations. The amide correlations are also the bases for the resonance assignment. All four experiments that were used to determine the backbone resonances relied on the detection of these amide correlation shifts. Figure 5-7 shows the strategy used for the backbone assignment but combinations of other experiments are possible too.

For a backbone assignment, as used here, the H^N-^{15}N-HSQC plane is common for all four experiments and a 2D ^{15}N-HSQC or the amide correlations separated by the carbonyl in the HNCO serves as a reference spectrum. Several redundant correlations are present and aid in the sequential assignment. The HNCA experiment is identical to the HNCO experiment after interchanging the pulses acting on the C_α and C' nuclei.

Although the information of the HNCA is also available in the HNCACB experiment, the HNCA is acquired because of its ~10-fold higher sensitivity as indicated in Table 5-2.

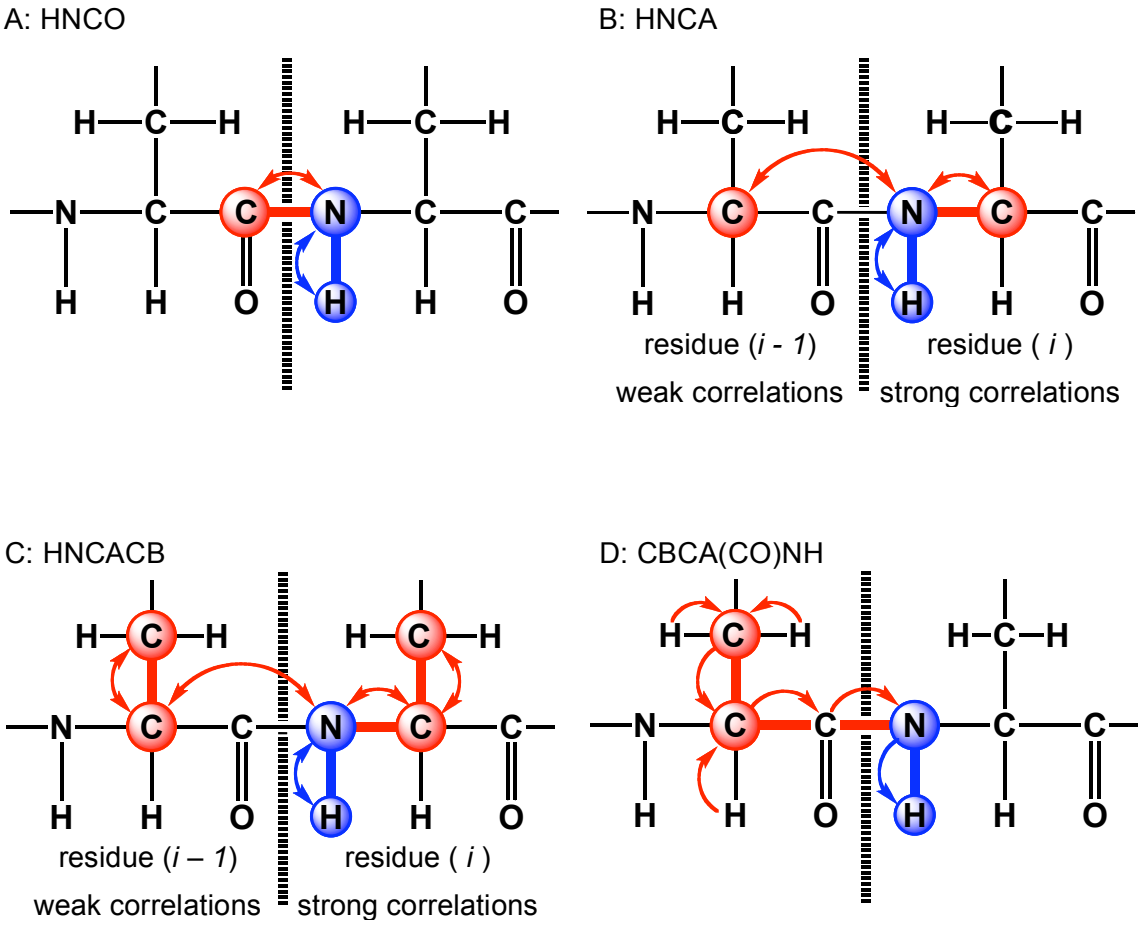


Figure 5-7: Correlations in backbone experiments used, **A.** HNCO, **B.** HNCA, **C.** HNCACB and **D.** CBCA(CO)NH. Arrows indicate the magnetization transfer path, bold bonds indicate the coherence transfer via ^{1}J-coupling and circled nuclei are evolving in one of the dimensions.

This is an advantage for the Cα assignment and important for sequential linking of the *i-1* and *i* residues with each other. A wide range of pulse sequences for each

experiment is available and the specific descriptions given in the text will refer to the version used.

Both, HNCO and HNCA experiments were run in the out-and-back version (*140*), where the magnetization was transferred from the $H^N \rightarrow N \rightarrow C'$ or $C\alpha(t_1) \rightarrow N(t_2) \rightarrow H^N(t_3)$, t_1, t_2 and t_3 indicating the evolution times for each dimension. A constant time (CT) evolution period for the nitrogen dimension was used (*141*) as well as the echo-antiecho gradient enhancement according to Rance and Kay (*142, 143*).

The other pair of experiments, HNCACB and CBCA(CO)NH, include the Cβ carbons providing a further link between the sidechain and the backbone. The Cα and Cβ nuclei for the *i-1* residues are covered in both experiments, but the measurement of the CBCA(CO)NH yields a 10-fold stronger signal for both nuclei (Table 5-2). The HNCACB provides crucial information since it is the only source including both Cα and Cβ for the sequential link. Similar to the HNCA and HNCO, the HNCACB used the out-and-back magnetization transfer $H^N \rightarrow N \rightarrow C\alpha \rightarrow C\alpha(t_1) \rightarrow C\alpha \rightarrow N(t_2) \rightarrow H^N(t_3)$ and echo-antiecho gradient enhancement. A semi-constant time (SCT) evolution (*144*) for nitrogen was implemented where a constant delay that was necessary for the evolution of a coupling was used for chemical shift evolution in addition to the incremented delay.

Compared to the CT version, increased sensitivity and resolution was achieved with the SCT application.

The CBCA(CO)NH experiment belonged to the category of transfer experiments and used a relay-COSY step (*145*) Cβ/Cα →Cα to correlate the chemical shifts. In a first step, the magnetization was transferred from Hα,β via ^{1}J coupling to the respective Cα,β, followed by a CT chemical shift evolution. After the relay-COSY step, couplings were refocused and magnetization was transferred in two INEPT steps via the carbonyl C' to the amide nitrogen. There the second CT-evolution occurred before the final transfer to the amide proton led to the direct detection. In short: Hα,β → Cα,β (t_1) → Cα → C' → N(t_2) → H^N(t_3). Minor modifications of this pulse sequence led to the HBHA(CO)NH experiment. Instead of evolving the carbon chemical shift, the evolution time was placed at the aliphatic protons Hα,β with a semi-constant time implementation. The sensitivity of such an experiment suffers some due to the fast relaxation of the protons during chemical shift evolution in t_1. Although not all of the Hα,β shifts can be unambiguously assigned this way, the experiment allowed for an important link to the amide resonance.

All experiments were acquired on a Bruker DRX 500 MHz spectrometer 4 channel at 20^0C, using XWIN-NMR 6.5 (Bruker Biospin, Rheinstetten Germany). Common acquisition parameters for the transmitter were: ^{1}H at 4.62 ppm, Cα = 54.2

ppm, Cβ = 28.1, Cα/β = 44.6 ppm C' = 173.9 ppm, ^{15}N = 121.4 ppm with the following sweep width: ^{1}H 12.0 ppm, Cα = 34.0 ppm, Cα,β = 66.3 ppm, C' = 14.0 ppm, ^{15}N = 24.0 ppm. To get better resolution, some signals in the nitrogen dimension were folded after careful examination of the final position of the affected resonances.

Table 5-2: List of backbone experiments used in this work, including HBHA(CO)NH for proton Hα and Hβ assignment. Column 2 lists the nuclei in the same order column 3 shows the relative sensitivity as listed by Sattler et al (*126*) for the respective nuclei.
* Values derived from S/N determination between CBCA(CO)NH and HBHA(CO)NH.

Experiment	Nuclei for S/N	Relative S/N [%]	Reference
HNCO	C'(*i-1*)	100	(*146,198*)
HNCA	Cα(*i*), Cα(*i-1*)	50, 15	(*146,198*)
CBCA(CO)NH	Cα(*i-1*), Cβ(*i-1*)	13, 9	(146,147)
HNCACB	Cα(*i*), Cα(*i-1*), Cβ(*i*), Cβ(*i-1*)	4, 1.7, 1.3, 0.5	(147-149)
HBHA(CO)NH	Hα(*i-1*), Hβ(*i-1*)	11, 7*	(143, 150)

Further experimental conditions for the individual measurements are summarized in Table 5-3. Time domain (TD) values are given as complex points, the spectrum data matrix as real points (SI). Data processing was done with NMRPipe/NMRDraw software package (*106*) (Version V 2.5 NMR Science, North Potomac, MD) and the resulting spectra were converted to the NMRView (*107*) (C-Version 5, NMRView J, One Moon Scientific) format for data analysis. NMRPipe processing scripts containing all the

processing parameters are provided in Appendix F, which extends those listed in Table 5-3.

Table 5-3: Select acquisition (A) and processing (B) parameters for backbone experiments. Parameters used were: number of scans (NS), acquisition and spectral matrices were given by the number of points in the time domain (TD) and frequency domain (SI). Linear prediction was applied in the dimension indicted in the LP-row and the abbreviation "apod." is used for apodization. Either a sinebell (SB) or squared sinebell (SSB) was used, the start- and end point of the curve is given as multiples of π.

A. Acquisition Parameters	HNCO	HNCA	CBCA(CO)NH	HNCACB
Pulse sequence	hnco-rk.mv	hnca-rk.mv	cbcaconhz	hncacbTL.mv
NS	4	20	16	32
Relaxation delay [s]	1.2	1.2	1.2	1.5
TD (^{1}H) [pts]	1024	1024	1024	1024
TD (^{13}C) [pts]	96	64	108	96
TD (^{15}N) [pts]	128	96	64	64

B. Processing Parameters				
SI (^{1}H) [pts]	410	410	410	410
SI (^{13}C) [pts]	128	256	256	256
SI (^{15}N) [pts]	256	128	128	128
LP	^{15}N	^{15}N	^{13}C, ^{15}N	^{13}C, ^{15}N
apod. (^{1}H)	SSB 0.35-0.90	SSB 0.50-0.98	SSB 0.45-0.98	SSB 0.50-0.98
apod. (^{13}C)	SB 0.50-0.98	SSB 0.50-0.98	SB 0.50-0.98	SB 0.50-0.98
apod. (^{15}N)	SB 0.50-1.00	SB 0.50-0.98	SSB 0.50-0.98	SB 0.50-0.98

For all amide based detections the region between 5.6-10.4 ppm was retained in the spectral matrix and one time zero filling was applied after linear prediction (LP) where indicated.

The assignment of the backbone resonances was conducted with the program NMRView (*107*) utilizing the NvAssign (*147*) extension. The basic idea was to form clusters of residue pairs based on the overlapping signals obtained by the four experiments discussed. After careful calibration of the spectra, peak picking was performed in NMRView and the peaks were screened for noise- or obvious artifact peaks, which were removed. Corresponding resonances were combined in an arbitrary cluster, describing residue (*i*) including H^N_i, N_i, $C\alpha_i$ and $C\beta_i$ as well as the preceding residue (*i-1*) C'_{i-1}, $C\alpha_{i-1}$ and $C\beta_{i-1}$. Figure 5-8 shows an example of such a cluster including the corresponding peaks in their respective spectra or planes. The green dotted line connects $C\alpha_i$ nuclei between the HNCA and HNCACB spectra, blue dashed line the $C\alpha_{i-1}$ residues that were present in all but the HNCO spectrum. The $C\beta_i$ residue was only present in the HNCACB spectrum and red dash-dot line connects the resonance for the HN-C and N-C plane of that spectrum. $C\beta_{i-1}$ residues were seen in the HNCACB and CBCA(CO)NH spectra and the purple line indicates the correlation. The presence of these resonances varied depending on the residues involved and even the

3D separation did not always prevent areas of overlapping peaks or multiple peaks belonging to several clusters in the same planes. From the 192 identified clusters, several contained weak peaks that were later determined to belong to the reduced form. It was determined, that the about 20% of the protein were in the reduced form when the backbone experiments were acquired, only 5-10% during the sidechain experiments and the final assignment was done for the oxidized complex.

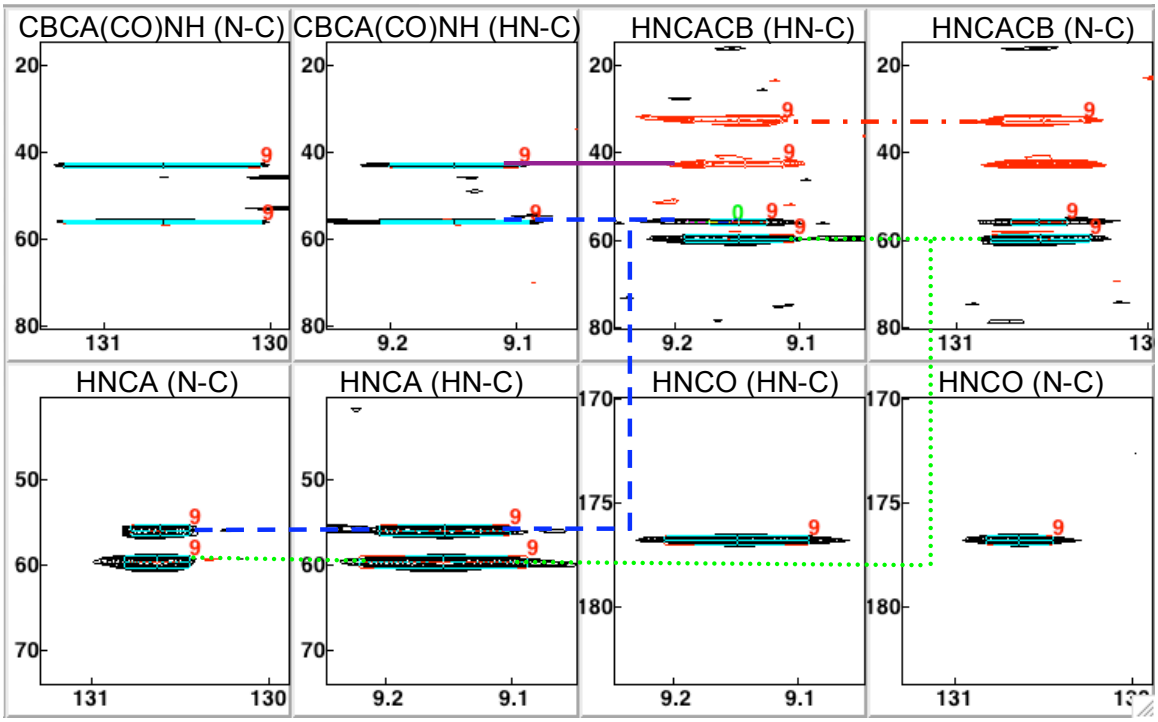


Figure 5-8: NvAssign CBCA layout to assign sequential clusters based on the 4 backbone experiments, showing the 4 HN-C planes in the center and N-C planes on either side. Connecting lines indicate same resonances in different spectra or planes for: $C\alpha_i$ green dotted line, $C\alpha_{i-1}$ blue dashed line, $C\beta_{i-1}$ purple line and $C\beta_i$ is seen only in of the HNCACB spectrum, indicated by a red dash-dot line.

The clusters were then assembled to fragments of multiple, connected residues, leading to a full assignment of the backbone. The software package PACES (Protein Sequential Assignment by Computer-Assisted Exhaustive Search) from Coggins and Zhou (*148*) was an indispensable supplement to the NvAssign tool during the fragment assembly. It is a Microsoft Windows Component Module (COM) based module and can be used as an add-in to Microsoft Excel. With the exception of Cβ for L66, K109 and S121, the proline preceding C' and amide N nuclei and the amide resonance for M1, all backbone resonances were assigned.

Sidechain Assignment

Several of the previously described experiments provided a link between the backbone and sidechain nuclei involving Hβ Cβ. Further experiments to assign the remaining carbon and proton atoms in the sidechain were acquired. The pair of (H)CC(CO)NH and H(CCCO)NH experiments (*105*) are correlating the aliphatic ^{1}H and ^{13}C resonances of a given amino acid with the amide of the following residue. This provides a powerful tool to obtaining nearly complete proton and carbon assignments in the same H_2O containing sample as used for the backbone and relates the resonances to the common amide correlation. This is in contrast to other strategies where HCCH-

COSY (*149, 150*) or HCCH-TOCSY (*151, 152*) correlations are used for the sidechain proton resonance assignment and samples need to be dissolved in D_2O. Additional draw-backs of these experiments are the presence of strong buffer signals typically seen in the aliphatic region which are not suppressed by any INEPT transfer as is the case for the amide based experiments, and a larger degree of spectral overlap. Despite these drawbacks, these experiments are still used to supplement the data.

The pair of sidechain experiments used was closely related to the previously discussed CBCA(CO)NH and HBHA(CO)NH experiments. Instead of a relay-COSY type transfer at the beginning, these experiments used isotropic mixing (*153*) on carbon, transferring the magnetization from the entire sidechain to Cα. Additionally the first step of the (H)CC(CO)NH experiment involved a proton to carbon INEPT transfer to enhance the sensitivity of the experiment. The magnetization traveled the following path for the (H)CC(CO)NH experiment: $H_{aliph} \rightarrow C_{aliph}$ (t_1) $\rightarrow C\alpha \rightarrow C' \rightarrow N(t_2) \rightarrow H^N(t_3)$ and H_{aliph} (t_1) $\rightarrow C_{aliph} \rightarrow C\alpha \rightarrow C' \rightarrow N(t_2) \rightarrow H^N(t_3)$ for H(CCCO)NH. The only difference between the two experiments was the position of the t_1 evolution period, giving rise to either a ^{13}C-^{1}H-^{15}N correlation spectrum or a ^{1}H-^{1}H-^{15}N one. The sensitivity of the HBHA(CO)NH experiment was reduced compared to the CBCA(CO)NH experiment because of the

isotropic mixing time (12-16 ms) and partial ^{13}C magnetization transfer to Cα during that time.

Table 5-4: A. Select acquisition parameters for sidechain experiments. Parameters used were: number of scans (NS), acquisition and spectral matrices given by the number of points in the time domain (TD) and frequency domain (SI). The carbon mixing time conditions and transmitter offset (O2) are listed and the sweep width (SW) covered by the individual evolution periods. H^N indicates the evolution of the amide protons, acquisition time and ^1H the evolution of protons in the indirect dimension.

A. Acquisition Parameters	(H)CC(CO)NH		H(CCCO)NH			HBHA (CO)NH
Field [MHz]	500	600	500	600	800	800
Pulse sequence	cconh.mv	cconh.mv	hcconh.mv	hcconh.mv	hcconh.mv	hbhaconhgp3d
NS	24	32	24	32	24	16
Relaxation delay [s]	1.2	1.0	1.2	0.9	0.9	1.0
TD (H^N) [pts]	1024	1024	1024	1024	1024	1024
TD (^1H) [pts]	-	-	140	140	140	147
TD (^{13}C) [pts]	128	160	-	-	-	-
TD (^{15}N) [pts]	64	64	64	64	64	58
Mixing time [ms]	13.0	13.0	13.0	13.0	13.0	-
Spin lock field [kHz]	9.2	8.4	9.2	8.4	10.0	-
SW (H^N) [ppm]	12	12	12	12	11.0	12
SW (^1H) [ppm]	-	-	7.5	7.0	8.0	5.8
SW (^{13}C) [ppm]	70	67	-	-	-	-
O2 (^{13}C) [ppm]	40	39	40	40	39	44.6

Table 5-4: B. Select processing parameters for sidechain experiments. Linear prediction was applied to the dimension indicted in the LP-row and the abbreviation “apod.” is used for apodization. Either a sinebell (SB) or squared sinebell (SSB) was used, the start- and end point of the curve is given as multiples of π.

B. Processing Parameters						
SI (H^N) [pts]	819	819	410	411	441	410
SI (^1H) [pts]	-	-	256	256	256	256
SI (^{13}C) [pts]	256	256	-	-	-	-
SI (^{15}N) [pts]	256	256	128	256	256	64
LP	^{15}N	^{15}N	^{15}N	^{15}N	^{15}N	^{13}C, ^{15}N
apod (H^N)	SSB 0.45-0.95	SSB 0.45-0.98	SSB 0.45-1.0	SSB 0.45-0.98	SSB 0.45-0.98	SSB 0.45-0.98
apod (^1H)	-	-	SB 0.45-1.0	SB 0.45-0.98	SB 0.45-1.0	SB 0.45-0.98
apod (^{13}C)	SSB 0.45-0.95	SSB 0.5-0.95	-	-	-	-
apod (^{15}N)	SSB 0.5-0.98	SSB 0.5-0.98	SB 0.5-1.0	SB 0.5-1.0	SB 0.45-1.0	SSB 0.45-0.98

Experiments were acquired on Bruker DRX spectrometers at 500 and 600 MHz, at 20 ºC, while additional H(CCCO)NH and HBHA(CO)NH experiments were measured at 800 MHz to benefit from the increased resolution. Parameters specific to the sidechain experiments are listed in Table 5-4, all other experimental conditions were the same as described for the backbone measurements.

The comparison of spectra at different field strength proved beneficial, since recurrent resonances were limited to the carbon and proton α/β resonances only. Using the example of I34, Figure 5-9 shows how the Hα proton was clearly visible in the H(CCCO)NH at 500 and 600 MHz, but missing at 800 MHz, while Hγ_1 had a strong correlation at 600 MHz, a weaker one at 800 MHz but was missing at 500 MHz. Similarly, Cδ or Cβ correlations were well defined at 600 MHz in the (H)CC(CO)NH spectrum, but were not be unambiguously assignable in the 500 MHz spectrum alone. NMRView did not provide a tool to pull out the proper planes for the sidechain assignment of these amide based experiments or the comparison of multiple spectra. Therefore a tcl-script for NMRView C was written to display the proper planes based on the HNCO/HSQC amide resonance selection.

A total of 1864 resonances were assigned and are listed in Appendix G. The assignments describe the oxidized form of the polymerase in complex with the deoxyguanosine triphosphate.

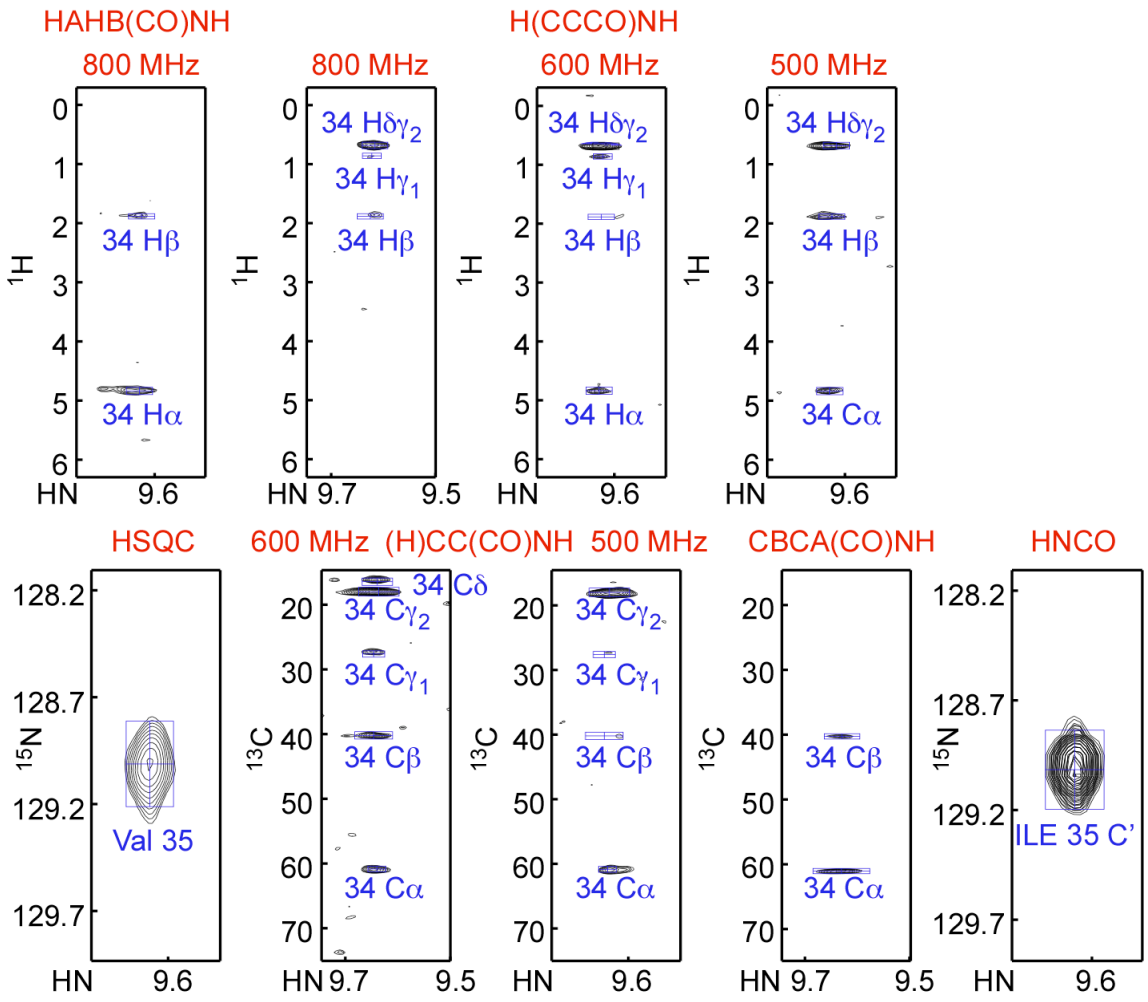


Figure 5-9: Example of a spectral layout for the proton and carbon sidechain assignment. Upper row contains proton-proton correlation spectra, from left to right: HAHB(CO)NH at 800 MHz, H(CCCO)NH at 800, 600, 500 MHz. Lower row shows the reference spectra HSQC and HNCO on either side, (H)CC(CO)NH at 600 and 500 MHz, 2nd and 3rd spectra, respectively and the CBCA(CO)NH spectrum to distinguish Cα and Cβ resonances.

Discussion and Conclusion

Two Conformers of Pol X: Oxidized and Reduced

The initially reduced form of the polymerase was found to be oxidized in the NMR tube over the course of several months, even in the presence of 5 mM DTT. Air will oxidize DTT over time and in addition, its reducing potency is diminished a lower pH. It was shown that the redox process was reversible for pol X. Activity assays showed that the oxidized form was substantially more active, incorporating the proper base, while the reduced form had little activity. A recent study (*51*) of the ASF virus capsid assembly has shown, that a reducing environment during the assembly process of ASFV is important. The assembly involves the cytosolic synthesis of the major capsid protein p73, which binds with the cytoplasmic face of endoplasmic reticulum derived lipids, assembling into the virus capsid (*53*). Once the virus is assembled, it is released (*154*) from the cell into the extracellular space. Doing so, it leaves the reducing environment that is maintained in the cell cytoplasm through high concentrations of reduced glutathione. Cobbold et al (*51*) showed, that this reducing environment is critical during the early stages of the ASFV assembly. Under oxidizing conditions, the capsid precursors folded slowly and incompletely, forming insoluble aggregates. Already newly assembled virus capsid precursors were disassembled under oxidizing conditions. However, those virus particles

that were released from the cells were resistant towards oxidized glutathione. It appears, that as the virus matures, it becomes resistant towards oxidation, possibly in preparation for the release into the extracellular space (*51*).

This extracellular space is rich on oxygen, presenting an oxidizing environment for the virus. Preparation to enter this environment has many forms. It is known from the similar poxviruses, that a series of virally encoded thiol oxoreductases are able to oxidize virus proteins exposed to the cytosol (*137, 138, 155*). Two examples are the G4L and E10R proteins (*138*), that are specifically involved in the disulfide bond formation in intracellular virion membrane proteins. The ASFV genome encodes for the gene 9GL (*56*), which has homology to oxoreductase family genes (*156*). Lewis et al (*156*) showed that the presence of 9GL was crucial for viral functionality, but a direct link to an oxidative behavior was not presented.

The host cell for ASFV has been described to be the swine macrophage, which also is an oxidizing environment, producing and releasing reactive oxygen species (*56, 63, 139*).

Other members of the polymerase X Family were examined for the presence of a disulfide bond. Pol β, contains three cysteines which are spaced too far apart in the molecule to form a disulfide bond and neither has one been reported so far. When

comparing the pol β structure (*41*) with either of the pol X structures, striking analogies were seen. The β-sheet (β4, 6, 8-10) containing the disulfide bond between C81 and C86 in the pol X structure matched the one from pol β, while the β-sheet of the pol X structure in the reduced form was moved towards the C-terminal subdomain, as shown in Figure 5-10. This analysis was done using Chimera extension MatchMaker (*157, 158*). This software provides a best match of two structures based on a pair-wise sequence alignment including an algorithm looking for best alignments of chain pairs.

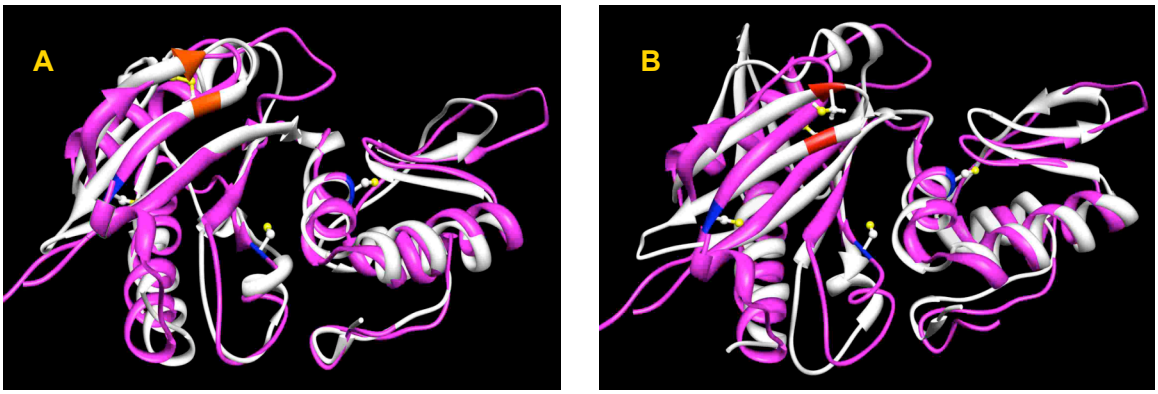


Figure 5-10: Alignment of the two ASFV pol X structures in white with the rat pol β structure omitting residues 1-142 in magenta (Sawaya et al (*41*), PDB code 1BPD). The positions of the cysteines are highlighted in red or blue, respectively. **A.** Comparison of pol β with the oxidized form of pol X (Tsai and coworkers (*47*)), PDB code 1JQR) and **B.** shows the best alignment with the reduced form of pol X (Mullen and coworkers (*46*), 1JAJ).

DNA polymerase λ, μ and the terminal deoxynucleotidyl transferase (TdT) all have seven cysteines. For the murine TdT (Delarue et al (*159*), PDB code 1JMS), four of the seven cysteines are in the palm and thumb subdomain. Interestingly, two of them, C378 and C404, are adjacent to each other on the β-sheet, with a distance of 6.4 Å between the sulfur atoms, sufficiently close to form a disulfide bond. A third cysteine, C438 is located in a loop adjacent to the same β-sheet, but with a slightly larger distance of 13 Å. This β-sheet and the position of the cysteines correspond to pol X but all cysteines are shown in the reduced form. So far, no report of an oxidized form was found in the literature. Similarly, four of the seven cysteines in the human pol λ structure (Garcia-Diaz et al (*160*), PDB code 1RZT) were located in the palm - thumb domain. Three of them (C412, C415 and C425) are again located on the same β-sheet with sulfur-sulfur distances of ~10.4 Å and are also represented in the reduced form. It was reported (*160*), that C543, residing in the finger domain, has a tendency to form intermolecular disulfide bonds and therefore several structures of pol λ are reported containing a C543A mutation. Only three cysteines, C353, C392 and C411 are found in the palm-thumb domain of the murine pol μ structure (Moon et al (*22*), PDB code 2IHM). Of those C392 is placed on the corresponding β-sheet and has a distance to C353, located in the α-helix behind the β-sheet, of almost 13 Å. The third cysteine is on a long

loop connector of the β-sheet, but even further away then C353. Without any major conformational changes, the presence of a disulfide bond similar to the one found in pol X is unlikely. It can be concluded, that the presence of a disulfide bond in the discussed β-sheet is unique for pol X and could play an important role in its activity. The presence of pol X in an oxidative environment like the macrophage and the good agreement of the oxidized form with the pol β structure all indicate, that this form might be biologically relevant and represent an adaptation of this enzyme to the host.

Chemical Shift Comparison Between Free and dGTP-Complexed Pol X

A comparison of the amide resonances of the free pol X with the pol X-dGTP complex ones revealed two regions where substantial chemical shift perturbation for the amide resonances was present. The large number of perturbations indicated, that this was not only a local affect due to the dGTP binding, but most likely the result of an additional conformational change triggered by the dGTP addition.

A comparison of the amide shift differences between the free pol X and the binary complex with dGTP is shown in the graph in Figure 5-11. There are two distinct regions, where the chemical shift perturbations for the amide protons and nitrogen nuclei is substantially larger then for the other residues. G142 is an exception in that it is a

single residue with a large shift for both nuclei. A is misassignment for this residue in either the free or complexed form was the most likely explanation for this difference. Checking the data for the dGTP complex confirmed a proper assignment while the free form could not be verified (*81*). It is possible that such a chemical shift difference occurs for an isolated residue, but they are more likely in unstructured regions of the protein. The location of L142 in β strand 12 as part of a structured β-sheet would make a single chemical shift perturbation less likely. The reason for this perturbation remains elusive.

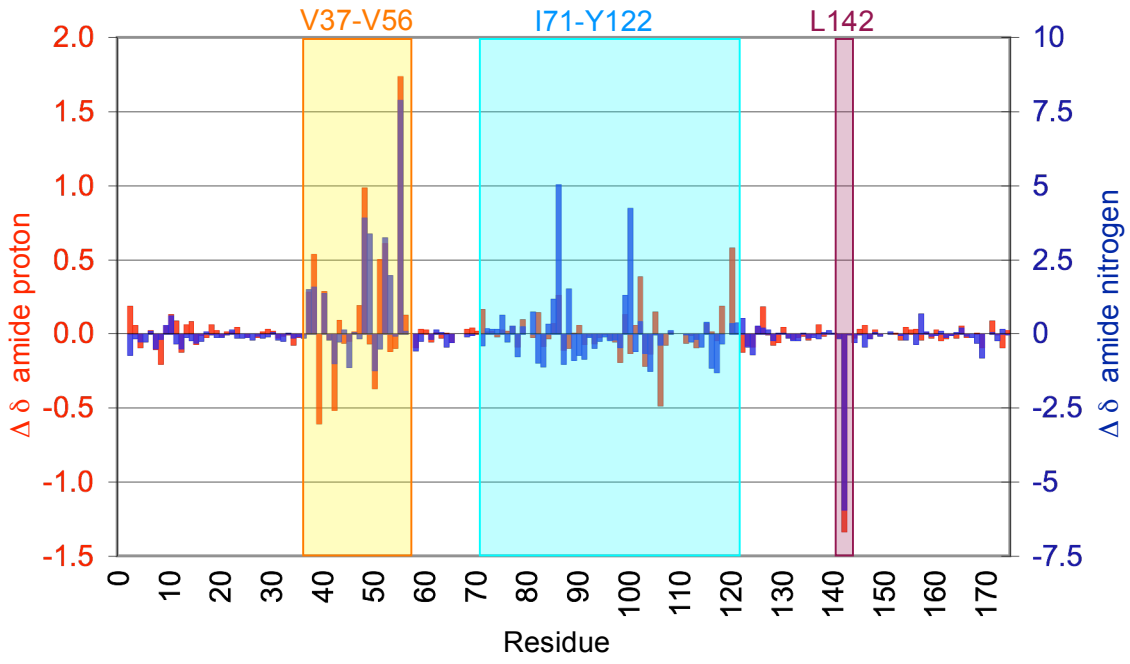


Figure 5-11: Amide chemical shift differences between free and complex pol X. Perturbation between the amide shifts reported by Mullen and coworker (*81*) on the free pol X and determined ones for the pol X-dGTP complex. Chemical shift differences (complex – free) for the amide protons are in red (left Y-scale) and blue bars (right Y-scale) for amide nitrogen.

The two larger regions of perturbations are surprisingly well defined to the subdomain interface as shown in Figure 5-12. Since the assignment was done for the oxidized form and the published shifts were for the reduced form, additional chemical shift differences due to the two forms are expected in parts of the β-sheet 8, 9 and 10. The involvement of β-sheets 4, 5 and 6, and α-helix B have not been seen in that context and are unique to the presence of dGTP.

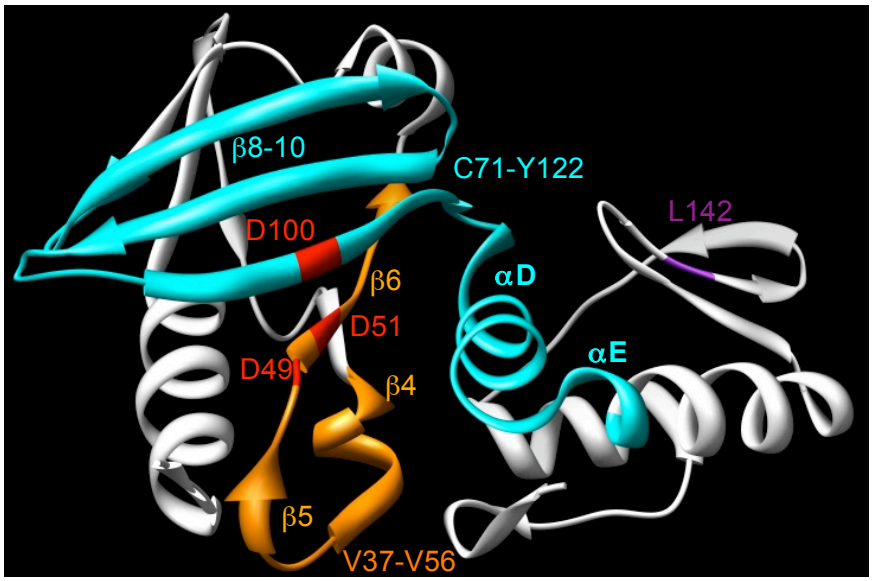


Figure 5-12: Chemical shift changes between amide resonance of free pol X and pol X – dGTP complex mapped to the structure of free pol X (PDB code 1JAJ). Colors correspond to the regions indicated in Figure 5-11, orange region 1, marine region 3 and L142. The three aspartic acid residues 49, 51 and 100 are labeled in red.

One hypothesis was, that the dGTP is positioned in the subdomain interface, affecting those residues. The three aspartic acids 49, 51 and 100, indicated in red, are well positioned for such a putative binding site where they could catalyzed the reaction of the dGTP insertion, all supporting this notion.

Additional changes were observed for α-helix D and the N-terminal end of α-helix E. These helices are on the C-terminal subdomain of the molecule, but lie at the interface with the palm domain. It is conceivable that the insertion of dGTP into that subdomain interface has a pronounced affect on the adjacent helix and alters the amide shift of those residues. But it also might indicate, that there is an additional conformational change between the two subdomains occurring which would affect the residues on both sides of the interface. It is anticipated, that the final structure of the complex will give more precise insight into the actual position of the dGTP and the expected conformational changes that occur upon this binding.

CHAPTER VI

PERFORMANCE OF CRYOGENIC PROBES AS A FUNCTION OF IONIC STRENGTH AND SAMPLE TUBE GEOMETRY [1]

Introduction

The introduction of cryogenic probes has dramatically increased the inherent sensitivity of nuclear magnet resonance experiments (*162-166*). Typically, a sensitivity gain by a factor of three to four relative to conventional room temperature probeheads (*167-170*) can be achieved in low conductivity organic solvents. This gain in performance is reduced for aqueous samples due to the increased conductivity and is further degraded in the presence of moderate amounts of salt (*171, 172*). Consequently, measurements of biomolecules in aqueous solutions are inherently compromised and are generally preferred to be studied under low salt conditions (<50-100 mM). Nature however is not limited to these "ideal" conditions and high salt environments are not uncommon. Elevated ionic conditions are often required for solubility and to prevent aggregation. Additionally, conformational transitions in nucleic acids and proteins, or

[1] This chapter was published by Voehler et al, *Journal of Magnetic Resonance* (2006) ***183***,102-109, (161). Elsevier Limited and is reprinted with minor modifications as permitted by the publisher, license number 1812770606674.

folding studies, often require high ionic strength environments. Titration experiments are another example where a ligand is observed at the beginning of a titration, demanding high sensitivity, and the substrate requires a high salt concentration for solubility.

The requirement of high salt concentrations in biomolecular NMR is often not the only limitation; sample availability and limited solubility are other important considerations. This necessitates the use of methods providing the highest sensitivity offered by cryogenic probes. Unfortunately, the performance gain these probes can offer is negated by high salt concentrations, occasionally to the point where the experiment is not feasible anymore.

Wand has proposed using reversed micelles in organic solvents, where the biomolecule is encapsulated in a water cavity (*173-178*). This lowers the overall dielectric contribution of the sample, which is beneficial for cryoprobe applications and addresses the slow tumbling of larger molecules (*171, 179-181*). The method has been successfully demonstrated but sample preparation remains challenging and its generalization is still under investigation (*182*).

In another approach, Dötsch et al. investigated the effect of different electrolytes on probe performance (*172*). The conductivity of a sample is not only impacted by the concentration and type of buffer, but also by its ion mobility. Comparing a variety of

popular buffers, substantial differences in the sample resistance were observed which result in sensitivity differences by up to a factor of 4. Nevertheless, high conductivity/mobility buffers are often unavoidable and in many cases the choice of buffer is dictated by the behavior of the biomolecule.

It has been recognized for some time that the solubility of proteins can often be maintained and improved in the presence of dipolar ions such as glycine or argenine (*183, 184*). Arumugam et al (*185*) prepared NMR samples of lysozyme and glucose containing either 0.77 M d_5-glycine or up to 0.2 M NaCl. As expected, NaCl containing samples exhibited much longer pulse widths and reduced signal to noise ratios compared to the same samples in presence of high amounts of glycine.

NMR vendors have introduced a rectangular sample tube design taking advantage of an optimized electric field (E-field) distribution (*186*) inside the sample (*187-189*). A sensitivity gain of 20% to 35% has been demonstrated for moderately salty solutions. However, specialized hardware such as probes and tubes are required which may not be available.

With respect to the sample distribution in the probe, alternative distributions are possible. The effect of sample radius on probe sensitivity under high salt conditions will be demonstrated. It was shown that 3 mm or 4 mm NMR tubes have substantial

advantages over a 5 mm tube when salty samples are measured. The susceptibility matched NMR tubes (Shigemi tubes) have also been included in this study since they are widely used and require typically only half the sample volume as a regular 5mm tube. A systematic investigation on a protein sample in 0.5 M NaCl is presented, exploring two commonly occurring situations. In the first case, the sample is soluble in abundance leading to a comparison between the different NMR tube arrangements at the same sample concentrations. The second case represents the often encountered situation where the amount of sample is limited. Therefore a comparison was done using different tube arrangements at the same, low amount of sample, e.g. varying concentration.

In addition, this NMR tube arrangement allows the use of samples with up to 4 M NaCl in the cryogenic probe while still obtaining a sensitivity advantage over the conventional room temperature probes. Salt effects on spin systems will also be shown, which are only visible under these extreme high concentrations and now are observable even on cryogenic probes.

Material and Methods

An AVANCE spectrometer with a triple resonance inverse cryogenic probe $^{1}H\{^{13}C,^{15}N\}$ operated at 500.13 MHz from Bruker BioSpin (Rheinstetten, Germany) was used for NMR investigation. 3 mm, 4 mm and 5 mm NMR tubes, part numbers 335-PP-8, 427-PP-8 and 535-PP-7, respectively were purchased from Wilmad-LabGlass (Buena, NJ). Advanced 5mm micro tubes, susceptibility matched for D_2O (Shigemi tubes), part number Z529451 were purchased from Aldrich (St. Louis, MO).

Sucrose Test Samples

10mM samples were prepared in D_2O in the presence or absence of 4 M NaCl. Additional 1 mM and 3 mM sucrose samples (4 M NaCl, D_2O) were also prepared. For sucrose samples in small diameter tubes the Bruker Match^{TM} system was utilized.

Protein Samples

Purified and lyophilized pol X polymerase from the African Swine Fever Virus (*9, 47*) was dissolved in a buffer solution containing 20mM PIPES, 20 mM $MgCl_2$, 10 mM DTT 0.02% NaN_3, 0.5 M NaCl and adjusted to pH 6.5. For the constant concentration experiments, aliquots of 475 µL, 285 µL 152 µL and 237.5 µL of a 0.5mM stock solution

were pipetted into the respective NMR tubes (5 mm, 4 mm, 3 mm and Shigemi). 25 µL of D_2O was added to the 5mm tube and 12.5 µL D_2O to the Shigemi tube to provide 5% D_2O for locking purposes while buffer solution was added to the 3 mm and 4 mm tubes to achieve the final volume. In the later arrangement, 250 µL and 180 µL D_2O, respectively, containing 20 µM TSP-d4 was added to a 5 mm outer tube for locking and referencing purposes. The smaller tubes were inserted into the 5 mm tube. The inner tube was centered by utilizing the concave bottom of the 5 mm tube and using adhesive tape for the top portion of the 3 mm tube as shown in Figure 6-1. This was not necessary for the 4 mm tube. Equal protein amount samples were prepared by adding a constant volume (160 µL) of the stock solution to each tube. Again 25 µL D_2O was added to the 5mm tube and 12.5 µL D_2O to the Shigemi tube for lock purposes and buffer was added to obtain the final volume, 315 µL, 140 µL 0 µL and 77.5 µL into 5 mm, 4 mm, 3 mm and Shigemi tubes, respectively.

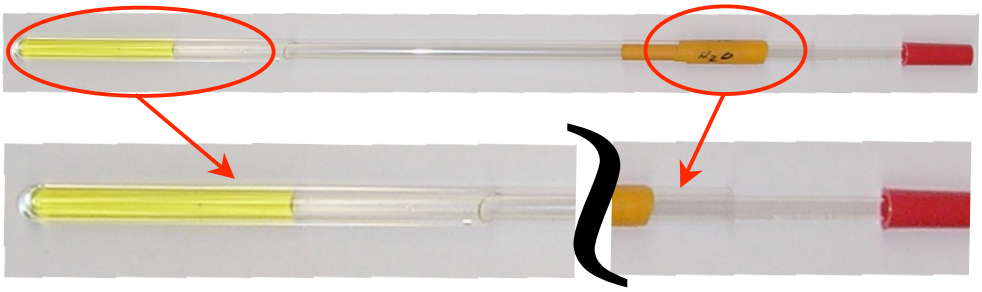


Figure 6-1: Dual NMR sample tube arrangement. Top: the 3 mm (or 4 mm) NMR tube contains the sample, a tape provides stability for the upper tube part. Bottom: the smaller tube is inserted into a 5 mm NMR tube, containing deuterated solvent.

A standard ^{15}N-HSQC measurement with flip-back pulse and watergate sequence for water suppression was used. Further parameters were: spectral width of 7003 Hz, 1024 complex points in the ^{1}H dimension, 1622 Hz with 128 complex points in the ^{15}N dimension and a relaxation delay of 1.2 s. The data was processed and analyzed using NMRPipe and NMRDraw software (*106*). The data was zero-filled in the acquisition dimension and linear predicted followed by zero filling in the indirect dimension for a resulting data matrix of 2k x 512. Peak picking and peak intensity measurements were performed using the same processing software package.

Results and Discussion

Factors Affecting the Signal-To-Noise Ratio

The two main factors that impact the signal-to-noise ratio (S/N) are 1) the overall resistance/geometry of the coil and 2) the environment of the sample (*190-192*). The latter is generally the only component under user control. For a cryogenically cooled probe the coil and sample temperature differ drastically which results in adjusted quality factor and temperature. Following the analysis given by Kelly et al (*172*) the following relation can be obtained:

$$S/N \propto \frac{1}{\sqrt{R_S(T_S + T_A) + R_C(T_C + T_A)}} \qquad [6.1]$$

where T denotes temperatures for RF-coil (C), amplifier (A) and sample (S), while R stands for resistance of coil (C) and sample (S) respectively. In cryogenic probes the term $R_C(T_C+T_A)$ has been reduced by lowering the RF-coil temperature, use of low resistant coil material and cooling the preamplifier, directly affecting the S/N ratio. In the remaining term, the sample temperature is dictated by the measuring condition, which leaves the sample resistance as a parameter to be optimized:

$$R_S \propto \omega^2 \sigma r_S^4 \qquad [6.2]$$

where ω is the angular frequency, σ represents the conductivity, given by $\Sigma c_i q_i \lambda_i$, with c as the concentration, q the charge, and λ the mobility of the different ionic species in the sample, and r_S is the sample radius. As was shown previously, choosing buffers with low conductivity and low mobility is preferred (*172, 185*).

This work addresses the other variable in Eq. (6.2), where a change in the sample radius is expected to have a major impact on the sample resistance and hence the S/N in high salt solutions.

NMR Spectroscopy in High Salt Solution Using Cryogenic Probes

Probe Tuning

The salt tolerance is dependent on probe design and differs somewhat between probes. On the cryogenic probe $^{1}H\{^{13}C,^{15}N\}$ used, the probe could not be tuned and matched for standard 5 mm sample tubes and solutions containing 1 to 4 M NaCl. By using micro tubes the total amount of salt in the probe is reduced while maintaining high ionic strength conditions. Small diameter NMR tubes were shown previously to enable high salt experiments on biomolecules utilizing regular probes (*193*). With a 3 mm tube, tuning/matching characteristics were obtainable for 0.5-4.5 M NaCl samples and are

identical to samples that did not contain any salt. Even higher salt concentrations, approaching the solubility limit, could readily be tuned/matched for 2 mm tubes.

90^0 Degree Pulse

The dependence of the $\pi/2$ pulse on very salty sucrose samples with sample diameters of 2 and 3 mm versus a regular 5 mm tube is shown in Figure 1 A. For samples in 5 mm tubes the $\pi/2$ pulse rose very rapidly, more than doubling before reaching salt concentrations of 1 M; beyond which the probe could not be tuned. This severely degrades the performance of most experiments, especially those containing spin lock and decoupling elements. As anticipated, the increase of the pulse length is much less dramatic for smaller tube diameters. In the case of the 3 mm tube the pulse length increases from 8.5 μs at 0 M NaCl to 12.5 μs at 4 M NaCl, representing an increase of only 51%. An even more modest increase of only 18% is observed for the pulse length is seen for 0-0.5 M NaCl while much smaller increases are observed from 1-2 M and 2-4 M, respectively. Therefore the high salt behavior cannot be linearly extrapolated from data at 0-1 M NaCl. These results clearly demonstrate that it is not the salt concentration alone that determines the pulse length and signal-to-noise degradation.

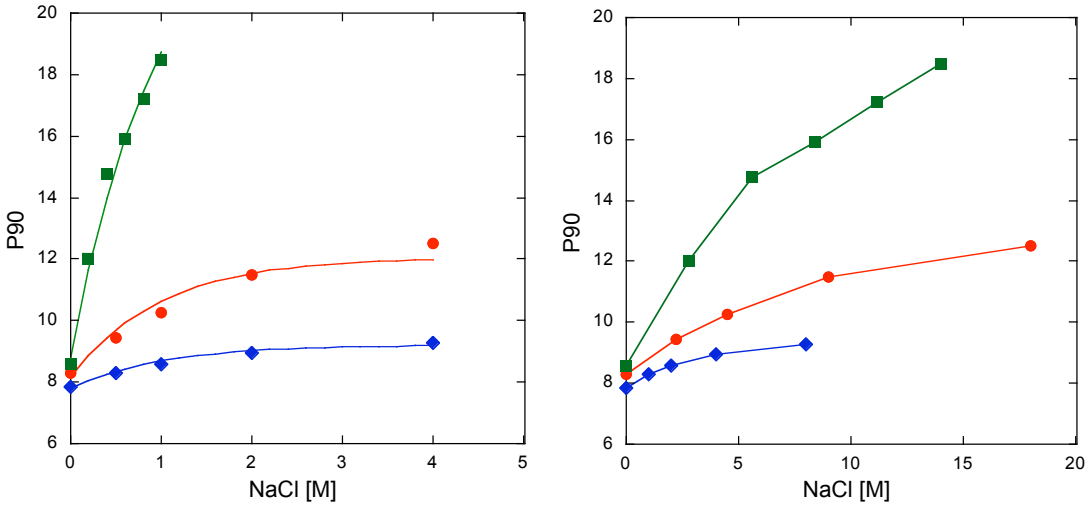


Figure 6-2: **A.** $\pi/2$ pulse length of a 10 mM sucrose test samples containing NaCl concentrations varying between 0 and 4M. Data was recorded at 295 K using a Bruker Avance 500 for 2 [✦], 3 [●] and 5mm [■] NMR tubes. The solid lines correspond to the empirically determined function: P90=P90(no salt) + $0.434d^{2.5}(1\text{-}e^{-[NaCl]})$, where d is the inner diameter (ID) in mm and [NaCl] is the molar salt concentration. (IDs were 1.6, 2.4 and 4.2 mm for 2, 3 and 5 mm NMR tubes). The 5mm data was not extrapolated past 1M because the probe head could not be properly tuned/matched. **B.** Pulse length dependence on amount of NaCl in the active volume. The amount of salt is represented as sample tube area * salt concentration.

If the same data is plotted (Figure 6-1 B) to reflect the amount of salt present in the probe the curves are closer together. However, the larger diameter tubes still exhibit a markedly poorer performance demonstrating that the location of the salt with respect to the receiver coil is dominating. This arises from the dependence of *Rs* on the tube diameter in Eq. (6.2) and provides a rationale for the experimental observation that in very low ionic strength solutions the tube diameter has only a small effect on the signal-

to-noise ratio while under very high salt conditions the size of the NMR tube is absolutely crucial.

Signal-to-Noise Comparisons for a Sucrose Test Sample

The signal-to-noise behavior for a sucrose test sample in a 3 mm NMR tube was investigated in more detail. Figure 6-2 summarizes the comparative results of the S/N determination of either a 1D ^{1}H NMR spectrum and the 1D trace of $^{1}H/^{13}C$-HSQC spectrum recorded at different salt concentrations. In both cases the S/N dependence on the salt concentration is similar. The drop in S/N is most pronounced between 0 and 1 M NaCl, while further increases in the salt concentration up to 4 M NaCl resulted in only a modest drop. In the case of the ^{1}H measurement the signal-to-noise ratio was reduced by a factor of 2.55 going from 0 to 4 M NaCl solutions. A similar result was also obtained from the HSQC experiments where a reduction by a factor of 2.35 was observed, despite the larger number of pulses applied and the use of ^{13}C decoupling. As anticipated, the ^{13}C pulses are virtually unaffected by the salt concentration. The fact that essentially the same reduction in S/N for both experiments were observed, indicates that at the end of the HSQC sequence the same amount of transverse magnetization is created; regardless of the salt concentration.

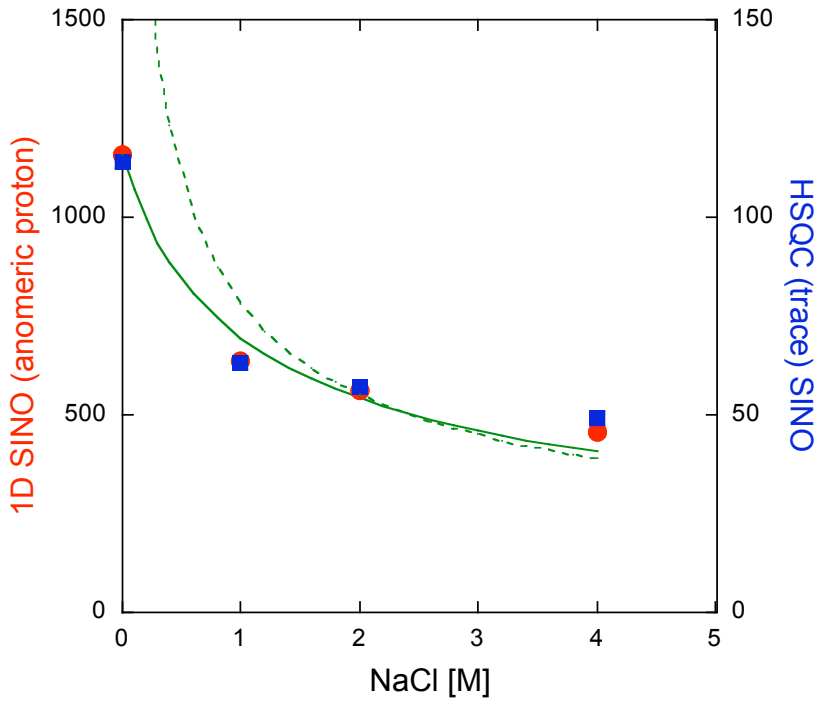


Figure 6-3: Signal to noise data of 10 mM Sucrose, 3 mm NMR tube as a function of the NaCl concentration. S/N analysis of 1D ^{1}H spectra (left ordinate, ●) and HSQC trace (right ordinate, ■) for the anomeric proton (0.5 ppm noise region). The solid line depicts the predicted salt dependence for the 1D data based on Eq. (6.1). $S/N = \frac{1}{\sqrt{cr_S^4[NaCl] + R_C(T_C + T_A)/f}}$, where [NaCl] is the salt concentration [M] and a *f* is a proportionality constant. The sample resistance is expressed as $cr^4[NaCl]$. c incorporates several constants and r_S^4 is the sample radius [mm]. The following values were obtained for the radical: $\sqrt{6.41 \bullet 10^{-7} r_S^4[NaCl] + 7.43 \bullet 10^{-7}}$. If no salt is present the S/N is primarily determined by the second term in Eq. (6.1), while at 2M NaCl it is assumed that the first term dominates, therefore the S/N curve may be approximated as k* [NaCl]$^{-0.5}$ where k = 792, for 3mm tubes (dotted line). This fit provides a simple estimation of the S/N dependency at high salt concentrations. In either case, the conductivity is assumed to be proportional to the salt concentration, which for high salt solutions is no longer accurate.

This is supported by the observation that on the cryoprobe used, for 3mm NMR tubes and 4 M NaCl the intensity of the 450° pulse is >97% of the 90° ^{1}H pulse. It is instructive to compare the observed salt dependence of the S/N ratio to the predicted behavior (Figure 6-2). As a first approximation only the sample contribution in Eq. 6.1 is considered. This approximation, although reasonable at high salt concentration, fails as expect under low salt conditions. A more realistic treatment includes both probe and sample contributions; this approach produces an acceptable agreement over the entire salt range shown on Figure 6.2. Additionally, HSQC spectra for 3 and 1 mM sucrose samples, all at 4 M NaCl, have been recorded and represent more realistic sample concentrations. The measured signal-to-noise ratios correspond well with the expected values for each concentration even at very high salt concentrations (data not shown).

For cryogenic probes and aqueous solvents in absence of salt the sensitivity gains are typically in the range of 2.0-3.2 relative to a conventional probe. The data obtained for 3mm tubes shows that a sensitivity loss of a factor of 2.35-2.55 compared to the same sample containing no salt is incurred. Therefore, by using 3mm tubes cryogenic probe used exhibits, even at 4 M NaCl, comparable sensitivity to a conventional probe for a sample without salt.

Based on the assumptions inherent in the fit of the S/N data shown in Figure 6-2 rough guidelines can be obtained for tube diameter selection and constant concentration samples. For this it was defined that when the S/N advantage offered by the larger diameter tubes drops to 20-25% of the value in absence of NaCl a smaller tube would be advantageous. A switch from 5 to 4 mm tubes should be considered at >0.2 M NaCl, from 4 to 3 mm at >0.6 M and from 3 to 2 mm at salt concentrations >2.2 M. This roughly corresponds to the salt concentrations where the 90° pulse is 40% longer than in absence of salt, providing a handy rule of thumb. It should be noted that the limits are dependent on the NMR experiment and the level of S/N loss one is willing to tolerate. Clearly, if the sample amount is limiting and the sample is soluble at higher concentration smaller tubes are to be preferred at elevated salt concentrations.

Systematic Comparison of Signal-To-Noise Data on a Protein Sample

To compare the signal-to-noise effect on a larger biomolecule several HSQC experiments were recorded on a 21 kDa pol X polymerase. The ^{15}N-HSQC is a key experiment in protein studies and employs a large number of $\pi/2$ and π pulses on the ^{1}H and ^{15}N channels, particularly for the sensitivity enhanced (*143, 194, 195*) echo/antiecho version (*196*). The recorded spectra exhibit well-resolved peaks for most residues. The

aim was to compare the 3, 4 and 5 mm tubes including 5 mm Shigemi tubes at either constant concentration or constant protein mass. The former situation is encountered when the protein is available in abundance; the concentration is ultimately limited by the protein solubility, which may be increased through the addition of salt. This concentration is used to assure the highest possible protein mass in the RF-coil region. It is however quite common, that only a limited amount of the biomolecule is available. Under these conditions the protein mass remains constant but the sample concentration depends on the tube diameter.

Constant Protein Concentration

Experiments were performed on a 0.48 mM protein sample in 0.5 M NaCl buffer. All samples were measured under identical acquisition and processing conditions. Experimental setup and results are summarized in Table 6-1 A. This table shows, that the reduced volume and therefore the reduced amount of protein in the 3, 4 mm and Shigemi tubes does not translate into a proportional reduction of signal intensity. The intensity loss is substantially smaller than expected for the 3 and 4 mm tubes, but for the 5mm Shigemi tube a signal increase of 15.2% is observed. The gain matches the added volume of 13.6% provided by the advanced Shigemi tube through a larger inner

diameter of the sample area. It also explains the slightly larger $\pi/2$ pulse length, since the coil to sample distance is reduced, which has a major impact on the $\pi/2$ pulse length as seen above.

Overall, the Shigemi tube offers the greatest sensitivity gain in this series and is advantageous if one can tolerate the longer pulses and additional radiation damping. The 4mm tube performs nearly as well and is easy to handle, particularly during titrations. At higher salt concentration, smaller diameter tubes are clearly preferred.

Constant Amount of Protein

Analogous to the previous measurements, results from the second case are summarized in Table 6-1 B. Here the total protein amount in the NMR tube was held constant. Samples were prepared by taking equal volumes from a protein stock solution, diluting them with the proper amount of buffer and/or D_2O as indicated. The behavior of the $\pi/2$ pulses are analogous to the previous experiment, which was expected since the sample was diluted with the sample buffer and the amount of ions in the solution did not change. In contrast to the previous experiment, the protein concentrations increase as the diameter decreases and the sample volume becomes smaller. This results in an increased sensitivity. Comparing constant sample amounts in the active volume rather

than the entire sample shows that the estimated sensitivity increase of 51.3% by the 5mm Shigemi tube is comparable to that of the 4 mm tube, however less then half of the sample amount is required. The 3 mm tube shows nearly as good a sensitivity gain as the Shigemi tube although the active volume is smaller than the total volume for this tube. This is a consequence of the reduced sample resistance using small diameter tubes. Clearly, a 3 mm Shigemi tube would combine both of these advantages. A regular 3 mm tube has several practical advantages; it is easy to handle and shim, has short $\pi/2$ pulse and contains only a small amount of solvent. This is beneficial for water suppression because of reduced radiation damping; and in a dual tube arrangement (Materials and Methods) D_2O or reference materials are in the outer tube, thereby separating them from the sample of interest. In summary, this comparison shows that when a limited amount of protein is available, one is served best by either a regular 3 mm or a 5 mm Shigemi tube.

Table 1. Summary of experimental conditions and results comparing ^{15}N-HSQC spectra measured on a 21 kDa pol X polymerase in different NMR tubes. The inner diameters of the tubes were 4.24 mm, 3.24 mm, 2.42 mm and 4.52 mm, which with a coil length of 16 mm results in active volumes of 226 µL, 132 µL, 73,6 µL and 250 µL for the 5mm, 4mm, 3mm and 5mm Shigemi tubes, respectively. **A.** Constant concentrations at 0.475 mM and **B.** constant protein amount (total available sample 80 nmol) resulting in concentrations of 160, 267, 500 and 320 µM for the 5, 4, 3 and 5mm Shigemi tubes, respectively. Since the active volume in a Shigemi tube is equivalent to the total volume, the amount of sample in Table (B) of 80 nmol is more than twice of that in the other samples. The estimated sensitivity gain for a sample with the same amount in the active volume would be 51.3%, which is comparable to the gain in a 4mm tube. The estimated sensitivity gain is based on the sample amount in the active volume with the 5mm tube as a reference. The noise estimate was calculated using the NMRDraw tool. 143 peaks were analyzed.

(A) Tube	5mm	4mm	3mm	5mm Shigemi
Total sample Volume (µL)	500	300	160	250
Sample Amount in active volume (nmol)	107.4	62.7	35.0	118.8
Estimated Sensitivity gain/loss based on the sample amount in the active volume (%)	-	- 41.6	- 67.4	+10.6
π /2 ^{1}H pulse (µs)	16.32	12.65	10.10	16.74
Noise estimate	233,540	198,630	156,174	229,692
Median of sensitivity change (%)	-	-7.5	-37.0	+15.2

(B) Tube	5mm	4mm	3mm	5mm Shigemi
Total sample Volume (µL)	500	300	160	250
Sample amount in active volume (nmol)	36.2	35.2 nmol	36.8	80
Estimated Sensitivity gain/loss based on the sample amount in the active volume (%)	-	-2.7	1.6	+120
π / 2 ^{1}H pulse (µs)	16.25	12.23	10.09	16.83
Noise estimate	207,687	176,990	151,863	214,687
Median of sensitivity change (%)	-	+52.3	+104.8	+114.1

Conclusion

These results demonstrate that it is possible to measure samples in cryogenically cooled NMR probes at very high salt concentrations when using smaller diameter tubes. It has been shown previously that the $\pi/2$ pulse increases with more salty samples. Specifically it was found that 1) the increase is not linear at very high salt concentrations and 2) the pulse length can be reduced by using the smaller 2, 3 or 4 mm tubes. By doing so, the overall amount of salt is reduced while keeping the salt concentration in the sample at a high level and the salt is moved away from the RF coil. As a general rule it can be concluded that as the salt concentration is increased, smaller tube diameters are advantageous but due to the reduced volume, solubility considerations may lead to a compromise between desired diameter size and available volume.

NMR experiments under high salt conditions can result in changes of the spectra. The ability to conduct experiments under high salt conditions is intrinsically relevant as biomolecular conformations and properties are dependent on it, often necessitating high salt concentrations. Approaches have been presented to carry out NMR experiments utilizing smaller sample diameter tubes to take advantage of the cryogenically cooled probes without any further adjustments.

The comparison of protein samples in medium salt concentration (0.5 M NaCl) at either constant protein concentration or constant amount addressed a commonly encountered situation. It was found that the sensitivity of high salt samples is disproportionally enhanced in tubes with a reduced diameter compared to the 5 mm tube and these should be given preference. The Shigemi NMR tube performed well in terms of sensitivity, but requires long pulses at moderate salt concentrations and is expected to suffer the same limitations as a regular 5mm tube at high salt concentrations. As might be expected, the 4 mm tube is most advantageous when solubility is an issue, while the 3 mm tube performs best with low sample amounts. The ease of use is another advantage offered by this arrangement for which no special hardware is required. Under either constant protein concentration or mass, substantial benefits of using the smaller diameter tubes are observed compared to the 5 mm NMR tube. In addition to an increasing signal and shorter $\pi/2$ pulses H/D-exchange of the exchangeable protons can be avoided by the dual tube arrangement when using those smaller tubes.

CHAPTER VII

CONCLUDING REMARKS

There were many reasons to have high hopes that pol X would serve as an ideal model to study nucleotide insertion reactions on a structural level. The main two reasons were the small molecular size of the polymerase and its simplified functionality, representing only the core catalytic site and the thumb domain. NMR studies would be well suitable to study such a small system, even in the presence of an oligonucleotide. Although pol X poses a few distinct challenges, a number of important new insights were obtained through these studies and are summarized below.

Solubility of Pol X at High Concentrations

Pol X appears to be very selective towards its buffer conditions. At the low concentrations typically used for biochemical assays, this was not as evident. But using the same buffers for NMR samples in high micromolar concentrations revealed limited long-term sample stability. Solubility conditions were worked out and sodium acetate and cacodylic acid based buffers provided a good environment for pol X. Once the structure

of free pol X was published, the conditions used by Mullen and coworkers (*46*) were adapted to allow for data comparison of the free and complexed pol X.

Pol X proved stable for months in these 0.5 M NaCl containing buffer conditions and even after more than a year the NMR spectra had the same appearance. While the sample was very stable under these conditions, the high salt concentration corresponded not to a physiological environment. Incorporation assays showed, that the enzyme retained activity at 0.5 M NaCl. High salt concentrations can lead to reduced binding when forming a complex but K_d values of pol X-nucleotide and pol X-DNA complexes at varying salt concentrations showed only a small dependency. Initial incompatibilities between high salt buffers and cryogenically cooled NMR probes were overcome by reducing the sample diameter. This arrangement allowed the use of these probes and utilize their superior performance for as presented in Chapter VI (*161*).

Oxidized and Reduced Form of Pol X

NMR measurements revealed a reversible conversion between the reduced and the oxidized form of pol X. These two forms can be distinguished and the ratio between them can be monitored by ^{15}N-HSQC spectra. Although the structures of both forms were published (*46, 47*) and speculations about a different behavior between them were

reported (*197*), the biochemical difference had not been investigated. The presence of an oxidized and reduced form is a unique feature for pol X compared to pol β and has not been reported for other polymerases from the Family X. This feature has been widely ignored in reports about the biochemical functions or activities of pol X. Although the difference became apparent towards the conclusion of this work, it was shown that the assignment of the pol X - dGTP complex was done for the oxidized form. The samples for the titration work contained a ratio of oxidized/reduced from greater then 1:1 with the exception of the dTTP titration, where it was 1:2. Fully oxidized samples were used for the following titrations: 36-mer, dGTP, $dC(pdC)_6$, $dG(pdG)_6$, $dC(pdC)_8$ and $dA(pdA)_{11}dC(pdC)_6$. The samples for the later four titrations were taken from the same stock solution.

It was shown, that only the oxidized pol X has incorporation and extension activity, while the reduced form appears to be mostly inactive. Pure oxidized form was obtained after removal of DTT and exposure to oxygen, while the reduction was performed under 100 mM DTT. To verify whether DTT had an impact on the activity, the reducing agent was removed under nitrogen and the assay run with 10 mM DTT. The results were the same. Misincorporation studies showed incorporation of the proper nucleotide for the oxidized form, while the reduced pol X did not show any activity on any

nucleotide. These are interesting findings and further experiments will be required to explain to what extent these two forms are responsible for some of the differences reported for the biological activity of pol X.

Complex Formation

Three biophysical methods, fluorescence, ITC and NMR, were utilized to increase the knowledge on the binding interaction between pol X and nucleotides or small oligonucleotides. A distinct difference between the purines and the pyrimidines was observed. While the purines yielded consistent results between the three methods and corroborate the literature values, the pyrimidines did not. It appeared that differences of the methods used and their respective biophysical parameters had an affect on the results.

The NMR spectra for the dNTP complexes revealed a large number of substantial chemical shift perturbations between the free and the dNTP complexes. The system appeared to be in slow exchange and the chemical shift change was not gradual, but a new peak rose at an unknown position in the spectrum, while the intensity of the original peaks was reduced. These large shift changes made it necessary to conduct a full backbone and sidechain assignment project for the dGTP complex. These

assignments will also assist in the future structural work. A number of pol X-oligonucleotide complexes were studied, with the goal to find a system that was suitable to pursue structural work. Adding ds-DNA to pol X always resulted in exchange broadening of the majority of resonances. For the 36- and 44-mer studied, resonance peaks were lost with as little as 0.2 mf of DNA. The spectrum was not recoverable even at a large excess of 5 to 12-fold DNA or by changing the temperature. This behavior was distinctly different then the one seen when adding nucleotides. Depending on the length, ss-DNA showed both, exchange broadening and chemical shift perturbation. For the spectra of the shorter 7-mer and 9-mer less then five resonances were completely gone at stoichiometric concentrations. The 12-mer and 18-mer ss-DNA showed predominantly exchange-broadened resonances. By mapping the changes of these short oligonucleotides to the free pol X structure, it was possible to identify the initial ss-DNA binding site as αE, β11-12 and its loop. The longer oligonucleotides, and ds-DNA used showed an additional affect on αC and part of β8, plus the subdomain interface with αD and part of β4, 6, 9 and10.

Three possible binding models have been discussed. In the first model, the DNA binds to pol X connecting αE, β11-12 with αC. The perturbations observed for the subdomain interface would originate from some conformational changes. The second

model would position the DNA in an L-shape into the subdomain interface, accounting for those perturbations, making a sharp turn in the β11-12 loop region to leave the polymerase at αE, or vice versa. Because pol X is missing the finger and 8 kDa domain compared to pol β, no obvious structure element is available to hold the DNA in the position described for model 1. To cause the interactions observed in both, the subdomain and αC regions, the DNA might move between the two positions. Either position would place the DNA close to the nucleotide, which is proposed to be positioned around β4, 6 and 10. The DNA bent required for the DNA positioning in both models is known and has been reported for the pol β ternary structure (*27*). The third model would allow the DNA to move between the two positions while at the same time undergo some conformational change to accommodate the complexes. These motions also concur with the exchange broadening observed in the NMR spectrum.

Although pol X appeared to be an ideal candidate for structural NMR studies, this and other groups have experienced the challenges pol X poses in terms of its solubility and exchange broadening when forming complexes. Several studies over the last years have focused on biochemical aspects of pol X, others have used fluorescence to explain some binding interactions. Since the structure was published, no further NMR work was reported on pol X. This study showed that despite of the challenges, NMR proved to be

a useful tool to identify and map interaction sites of pol X with nucleotides and oligonucleotides on a residue level and further provided the bases for structural work on the binary dGTP complex.

Conclusion

In Summary, the following seven points were addressed during this work and resulted in novel insights on pol X and its interaction with other biomolecules: 1) NMR methods provide a tool to monitor the presence of the oxidized and reduced form of pol X and biochemical studies of each individual form revealed, that the oxidized form carried the main activity during template-primer extension. 2) Backbone and sidechain assignments were done on a binary complex with dGTP enabling a chemical shift perturbation mapping for the amide resonances in comparison to the free pol X resonances. Addition of dGTP causes chemical shift changes in the subdomain interface, including β-sheet 4-6, 8-10 and α-helices B and D. This interaction surface is lager then expected for a nucleotide binding only and implies that an additional conformational change between the subdomains is likely. 3) ITC and NMR studies indicated a distinct difference in the complex formation between purines and pyrimidines. The NMR study showed a 20-fold lower binding of the pyrimidines. 4) A series of K_d

values in the high nanomolar, low micromolar range for nucleotides and ds-DNA were reported, all indicating good complex formation. 5) NMR studies indicated that the nucleotide complexes were in a slow exchange time regime, ds-DNA in an intermediate exchange and ss-DNA was in intermediate to fast exchange, resulting in massive exchange broadening for the ss-DNA and ds-DNA at sub-stoichiometric ratios. 6) The residue specific amide resonance changes, as monitored by NMR, enabled the identification of binding interaction sites for ss-DNA and ds-DNA, locating them to the C-terminal end of αE, αC, β 11-12 and the subdomain interface. A model was derived from these interactions to explain likely complex structures of pol X with DNA. 7) A new arrangement of smaller diameter NMR tubes made these NMR measurements possible in cryoprobes and the presence of at high salt. 8) Other buffer systems for long term stability of pol X were identified.

APPENDIX A

Cleavage of the Leader Peptide

Cleaving off the leader peptide on the pol X_{orig} protein was desired to reduce the overall molecular size for the NMR studies. A recombinant enterokinase (rEK) cleavage capture kit (Novagen) was used, and the procedure as described by the vendor was followed. Pilot digests with small amounts of pol X_{orig} exposed to the rEK were possible as shown in Figure A-1. Cleavage proceeded well in most of the pilot tests as shown in a representative SDS-PAGE in Figure A-1. Complete digestion was achieved before 22 h. It is notable though, that the protein bands weaken considerably over time. A series of differing rEK concentrations revealed that 100 to 150 µg pol X_{orig} can be cleaved with 1 U of rEK within 16 h.

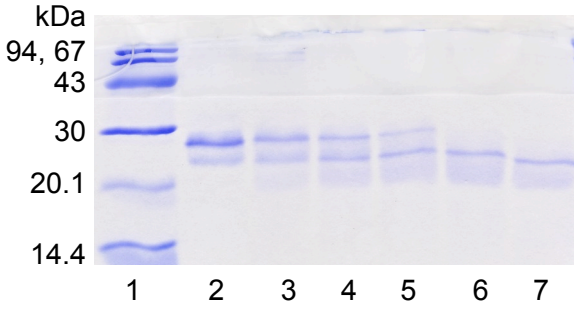


Figure A-1: SDS-PAGE of a pilot cleavage reactions with increasing incubation time; digestion of 23 µg pol X_{orig} digested with 0.4 U rEK, lane 1 low molecular weight (LMW) marker (Amersham Pharmacia Biotech, Piscataway, NJ, 14.4, 20.1, 30, 43, 67 and 94 kDa), lanes 2 to 7 increasing incubation for cleavage: 1, 2, 5, 8, 22 and 26 h, respectively.

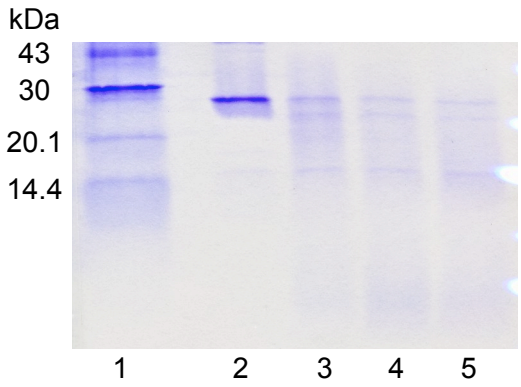


Figure A-2: SDS-PAGE of an NMR sample digestion. 7.5 mg pol X_{orig} in 34 mL cleavage buffer pH 6.5, 75 U rEK added. Lane 1 low molecular weight standard, 2 is mock sample, lanes 3 to 5 after 3, 18 and 24 h of digestion at RT.

Larger amounts of pol X_{orig}, as required for NMR samples, were digested under the same conditions. As shown by the SDS-PAGE in Figure A-2, a strong protein band was seen before digestion. Upon addition of the rEK, the initial protein bands weakened and only very weak bands of the cleaved protein remained. This behavior was generally observed for these digestions and very small amounts of digested protein were obtained. Modifying the pH and sample concentration led to minimal improvements. As will be shown later, these very low yields might be attributed to the limited solubility of the native pol X without the leader peptide. The presence of insoluble protein was an even larger issue for the large-scale samples, where the rEK was captured by EKapture Agarose and removed by spin-filtration. Any insoluble protein was removed during this process as well.

The inability to obtain larger quantities of cleaved protein and the uncertainty of long-term sample integrity after exposure to the enterokinase diminished the usability of this procedure. The availability of a new construct with a short C-terminal polyhistidine tag (6x His) was opportune. With today's knowledge about the poor and selective solubility of the native ASFV pol X protein, the cleavage conditions could most likely be worked out also for larger samples.

Activity of Cleaved Pol X_{orig}

The cleaved pol X_{orig} protein clearly showed activity in this assay as shown in Figure A-3, but it is much reduced compared to the activity of pol β under the same conditions. Tsai et al (*61*) have since reported that pol X shows an average catalytic efficiency of 1/5000 compared to that of pol β.

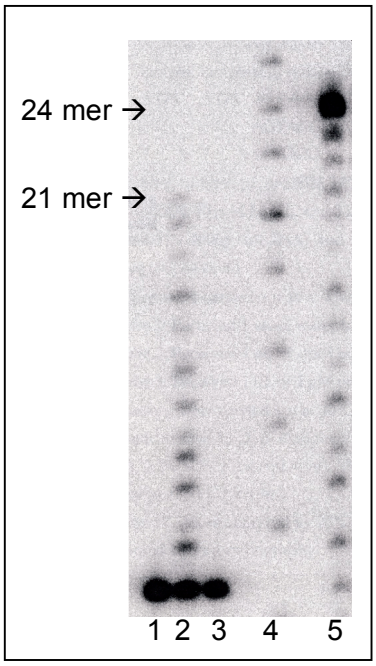


Figure A-3: Primer extension assay on pol X_{orig} after cleavage with rEK. The assay was carried out in the presence of 100 μM dNTP's, 8 nM template-primer DNA (5'-GCAGTATCGCGCGCGGCATGAGCT-3', ^{32}P-5'-AGCTCATGCCG) and 50 mM MOPS instead of Tris•HCl. Other experimental conditions were as described in the text. Lane 1 was a blank without polymerase, 2: 1 μM pol X_{orig}, 3: 1 nM pol X_{orig}, 4: a marker, 8-32 oligomer in 2 base increments, 5: 5nM pol β.

APPENDIX B

Solubility Screen for pol X_{iz}

Tables of Appendix B include the observations of the solubility screen performed with pol X_{iz}.

All buffers added were from 200 mM stock solutions unless indicated otherwise. The solubility screen was performed at room temperature.
Column 1) indicates the well number, 2) the base buffer, 500 µL stock solution added to the well 3) final concentration of additional salts or stabilizer, 4) amount of additional salts or stabilizer added to the well 5) additional water added to a final volume of 1 mL. All additional columns contain the scoring information at the indicated time interval.

Scoring:

0 = clear solution
0.5 = very slight cloudiness throughout the drop
1 = stronger precipitation in the center of the drop
2 = growing precipitation covering ~50% of the drop
3 = growing precipitation covering ~75% of the drop
4 = precipitation completely covering the drop

c = formation of crystals
r = precipitation forms a ring (aggregation)
d = spotty precipitation on the glass cover
dry = drop dried out (seal on cover was not tight enough)

Table 1: Main Buffers and pH

Experimental conditions: 1 mM protein solution, 1 drop on cover slide, pipette 0.5 mL of buffer stock solution, 200 mM into reservoir and add 0.5 mL H_2O.

	Buffer, pH	Score			
Well #	100 mM concentration, unless stated otherwise	2 days	5 days	19 days	70 days
1	Potassium Phosphate pH 4.2	4	4	4	4 d
2	Potassium Phosphate pH 5.0	4	4	4	4 d
3	Potassium Phosphate pH 6.0	4	4	4	4 d
4	Potassium Phosphate pH 7.0	4	4	4	4 d
5	Sodium Phosphate pH 4.2	4	4	4	4 d
6	Sodium Phosphate pH 5.0	4	4	4	4 d
7	Sodium Phosphate pH 6.0	4	4	4	4d
8	Sodium Phosphate pH 7.0	4	4	4	4
9	CH_3COONa pH 4.5	0	0	0	0 d
10	CH_3COONa pH 5.5	0	0	0	0 d
11	CH_3COONa pH 6.5	0	0	0	0.5 d
12	Sodium Citrate pH 4.5	4	4	4	4
13	Sodium Citrate pH 5.5	4	4	4	4
14	Sodium Citrate pH 6.2	4	4	4	4
15	PIPES pH 6.5	1 r	3	3	3 d
16	41 mM HEPES-Na pH 6.6	0	0	2	0 d
17	Tris pH 7.0	1 r	2 c	3 c	2 d
18	Tris pH 8.0	1	3 c	3 c	3 c d
19	MES pH 5.2	0	0	0.5	1
20	MES pH 6.0	0	0 c	3	4 d
21	MES pH 7.0	0	1 c	3 c	3 d
22	Na-Cacodylate pH 5.0	0	0 c	4 c	4
23	Na-Cacodylate pH 6.5	0	0 c	0 c	1 c d
24	H_2O	0	1	1	2 d

Table 2: Main Buffers and pH, 1 mM Protein Solution

Experimental conditions:, 1 drop on cover slide, pipette 500 µL of buffer stock solution, 200 mM, 100 µL of the salt additive and 400 µL H_2O into reservoir.

	Salt additive	Buffer, pH	Score				
Well#	Final conc.	100 mM concentration,	1 day	2 days	5 days	10 days	61 days
1	20 mM $CaCl_2$	H_2O	0	0 d	2	1	1
2	20 mM $MgCl_2$	H_2O	0	0 d	1	1	1
3	20 mM NH_4SO_4	H_2O	0	0 d	1	4	4 d
4	20 mM NH_4Ac	H_2O	0	0 d	0	4	4 d
5	20 mM $CaCl_2$	Sodium Phosphate pH 6.0	-	-	-	-	-
6	20 mM $MgCl_2$	Sodium Phosphate pH 6.0	0	0 c	1	1	1
7	20 mM NH_4SO_4	Sodium Phosphate pH 6.0	0	0	1	1	1
8	20 mM NH_4Ac	Sodium Phosphate pH 6.0	0	0	1	1	1
9	20 mM $CaCl_2$	CH_3COONa pH 5.5	0	0	0	0	0
10	20 mM $MgCl_2$	CH_3COONa pH 5.5	0	0	0	0	0
11	20 mM NH_4SO_4	CH_3COONa pH 5.5	0	0	0	0	0
12	20 mM NH_4Ac	CH_3COONa pH 5.5	0	0	0	0	0
13	20 mM $CaCl_2$	Tris pH 7.0	0	0	0.5	0.5	1
14	20 mM $MgCl_2$	Tris pH 7.0	0	0	0	0	0
15	20 mM NH_4SO_4	Tris pH 7.0	1	1	1	2	3
16	20 mM NH_4Ac	Tris pH 7.0	0.5	0.5	0 c	0.5	2
17	20 mM $CaCl_2$	Potassium Phosphate pH 6.0	-	-	-	-	-
18	20 mM $MgCl_2$	Potassium Phosphate pH 6.0	0	0	0 c	1	2 c
19	20 mM NH_4SO_4	Potassium Phosphate pH 6.0	0.5	0.5	2 c	1	1 c
20	20 mM NH_4Ac	Potassium Phosphate pH 6.0	0.5	0.5	1 c	1	1 c
21	20 mM $CaCl_2$	Na-Cacodylate pH 5.5	0	0	0	0	0
22	20 mM $MgCl_2$	Na-Cacodylate pH 5.5	0	0	0	0	0
23	20 mM NH_4SO_4	Na-Cacodylate pH 5.5	0	0	0	2	3
24	20 mM NH_4Ac	Na-Cacodylate pH 5.5	0	0	0	0	0

Table 3: Main Buffers and pH, 2.5 mM Protein Solution

Experimental conditions: 2.5 mM protein solution, 2 drops on cover slide, pipette 500 µL of buffer stock solution, 200 mM, 100 µL of the salt additive and 400 µL H_2O into reservoir.

	Salt additive	Buffer, pH	Score						
Well #	Final conc.	100 mM concentration, unless stated otherwise	1 day	9 days	18 days	23 days	26 days	40 days	91 days
1	20 mM $CaCl_2$	Sodium Phosphate, pH 5.0	1	1 c	1 c	1 c	1 c	2 c	4 c
2	20 mM $MgCl_2$	Sodium Phosphate pH 5.0	0	0 c	0 c	0 c	0 c	0 c	0 c
3	20 mM NH_4SO_4	Sodium Phosphate pH 5.0	0	1	3	3	3	3	3
4	20 mM NH_4Ac	Sodium Phosphate pH 5.0	0	0	2	3	3	4	4
5	20 mM $CaCl_2$	CH_3COONa pH 5.5	0	0	0	0	0	0	0
6	20 mM $MgCl_2$	CH_3COONa pH 5.5	0	0	0	0	0	0	0
7	20 mM NH_4SO_4	CH_3COONa pH 5.5	0	0	0	0	0	0	4
8	20 mM NH_4Ac	CH_3COONa pH 5.5	0	0	0	0	0	0	0
9	20 mM $CaCl_2$	Tris pH 7.0	0	1 c	2 c	1 c	1 c	1 c	1 c
10	20 mM $MgCl_2$	Tris pH 7.0	0	1 c	2 c	0	0	0.5	0
11	20 mM NH_4SO_4	Tris pH 7.0	4	4	4	4	4	4	4
12	20 mM NH_4Ac	Tris pH 7.0	0	0	4	4	4	4	4 c
13	20 mM $CaCl_2$	41 mM HEPES-Na pH 6.6	0	2	2	2	2	2	1
14	20 mM $MgCl_2$	41 mM HEPES-Na pH 6.6	0	3	3	2	3	3	1
15	20 mM NH_4SO_4	41 mM HEPES-Na pH 6.6	4	4	4	4	4	4	4
16	20 mM NH_4Ac	41 mM HEPES-Na pH 6.6	0	2	2	4	4	4	2
17	20 mM $CaCl_2$	Na-Cacodylate pH 5.0	0	0 c	0	0	0	0	0
18	20 mM $MgCl_2$	Na-Cacodylate pH 5.0	0	0	0	0	0	0	0
19	20 mM NH_4SO_4	Na-Cacodylate pH 5.0	0	0	0	0	0	0	0
20	20 mM NH_4Ac	Na-Cacodylate pH 5.0	0	0	0	0	0	1	4
21	20 mM $CaCl_2$	MES pH 5.2	0	0.5	1	0.5	0.5	0.5	0.5
22	20 mM $MgCl_2$	MES pH 5.2	0	0.5	1	0.5	0.5	0.5	0.5
23	20 mM NH_4SO_4	MES pH 5.2	0	0.5	1	0.5	0.5	0.5	3
24	20 mM NH_4Ac	MES pH 5.2	0	0.5	0.5	1	0.5	0.5	0

Table 4: NaCl Addition, 2.5 mM Protein Solution

Experimental conditions: 2.5 mM protein solution, 1 drop on cover slide; pipette the following solutions into reservoir: 500 µL of buffer stock solution, 200 mM; 25, 50 or 100 µL NaCl stock solution 1M and 475, 450 or 400 µL H_2O, to achieve final NaCl concentrations of 25, 50 and 100 mM, respectively

	Salt additive	Buffer, pH	Score						
Well #	Final conc.	100 mM concentration, unless stated otherwise	1 day	9 days	18 days	23 days	26 days	40 days	91 days
1	25 mM NaCl	Potassium Phosphate pH 6.0	2	3	3	3	3	3	4
2	50 mM NaCl	Potassium Phosphate pH 6.0	3	3	3	3	3	3	4
3	100 mM NaCl	Potassium Phosphate pH 6.0	3	4	4	3	3	4	4
4	25 mM NaCl	Sodium Phosphate pH 6.0	4	4	4	4	4	4	4 d
5	50 mM NaCl	Sodium Phosphate pH 6.0	3	3	3	3	3	4	4
6	100 mM NaCl	Sodium Phosphate pH 6.0	3	3	3	3	3	4	4
7	25 mM NaCl	CH_3COONa pH 5.5	0	0	0	0	0	0	0 d
8	50 mM NaCl	CH_3COONa pH 5.5	0	0 c	0 c	0 c	0 c	0 c	0 d
9	100 mM NaCl	CH_3COONa pH 5.5	0	0	0 c	0 c	0 c	0	0
10	25 mM NaCl	41 mM HEPES-Na pH 6.6	3	3	4	4	4	4	0
11	50 mM NaCl	41 mM HEPES-Na pH 6.6	4	4	4	4	4	4	4
12	100 mM NaCl	41 mM HEPES-Na pH 6.6	4	4	4	4	4	4	4
13	25 mM NaCl	PIPES pH 6.5	0	0.5	0.5	0.5	1	2	2
14	50 mM NaCl	PIPES pH 6.5	0	2	2	3	2	2	3
15	100 mM NaCl	PIPES pH 6.5	0	2	3	4	4	3	4
16	25 mM NaCl	MES pH 6.0	0	0.5	1	0.5	0	0	0.5
17	50 mM NaCl	MES pH 6.0	0	0.5	1	0.5	0	0.5	0.5
18	100 mM NaCl	MES pH 6.0	0	0	1	0	0	0	0.5
19	25 mM NaCl	Na-Cacodylate pH 5.0	0	0	0	0	0	0	0 d
20	50 mM NaCl	Na-Cacodylate pH 5.0	0	0	0	0	0	0	0 d
21	100 mM NaCl	Na-Cacodylate pH 5.0	0	0	0 c	0	0 r	1	0
22	25 mM NaCl	Tris pH 7.0	0	0	0	0	0	0	0 d
23	50 mM NaCl	Tris pH 7.0	0	0.5	0	2	3	4	4
24	100 mM NaCl	Tris pH 7.0	1	2	3	3	4	4	4

Table 5: Stabilizers, 2.5 mM Protein Solution

Experimental conditions: 2.5 mM protein solution, 1 drop on cover slide; pipette 500 µL of buffer stock solution, 200 mM, salt additive and H_2O as indicated in columns 4 and 5 into reservoir. CH_3COONa at pH 5.5, Tris at pH 7.0, Sodium Phosphate at pH 5.0

	Salt additive	Buffer			Score				
well #	Final conc.	100 mM concentration	Salt additive	add. H_2O [µL]	1 day	9 days	18 days	23 days	40 days
1	5 % DMSO	CH_3COONa	50 µL	450	0	0	0	0	0
2	10 % DMSO	CH_3COONa	100 µL	400	0	0	0	0	0
3	30 % DMSO	CH_3COONa	300 µL	200	2	4	4	4	4
4	5 % Glycerol	CH_3COONa	100 µL, 50 %	400	0	0	0	0	0
5	10 % Glycerol	CH_3COONa	200 µL, 50 %	300	0	0	0	0	0
6	30 % Glycerol	CH_3COONa	300 µL, neat	200	0	0	0	0	0
7	5 mM ATP	CH_3COONa	250 µL, 20 mM	250	4	4	4 c	4 c	4 c
8	2 mM CHAPS	CH_3COONa	100 µL, 20 mM	400	0	0	0	0	0
9	5 % DMSO	Tris	50 µL	450	1	2	0 r	0 r	0 r
10	10 % DMSO	Tris	100 µL	400	2	2	3	3	3
11	30 % DMSO	Tris	300 µL	200	4	4	4 c	4	4
12	5 % Glycerol	Tris	100 µL, 50 %	400	0	0	0 r	0 r	0 r
13	10 % Glycerol	Tris	200 µL, 50 %	300	0	0 r	0 r	0	0
14	30 % Glycerol	Tris	300 µL, neat	200	0	0	0 r	0 r	0 r
15	5 mM ATP	Tris	250 µL, 20 mM	250	4	4	4 c	4	4
16	2 mM CHAPS	Tris	100 µL, 20 mM	400	0	1	1 r	1 r	1
17	5 % DMSO	Sodium Phosphate	50 µL	450	3	4	4 r	4	4
18	10 % DMSO	Sodium Phosphate	100 µL	400	3	4	4 c	4	4
19	30 % DMSO	Sodium Phosphate	300 µL	200	4	4	4 c	4	4
20	5 % Glycerol	Sodium Phosphate	100 µL, 50 %	400	1	3	3	3	3
21	10 % Glycerol	Sodium Phosphate	200 µL, 50 %	300	2	2 r	2 r	2 r	2 r
22	30 % Glycerol	Sodium Phosphate	300 µL, neat	200	2	1 r	1 r	1 r	2 r
23	5 mM ATP	Sodium Phosphate	250 µL, 20 mM	250	4	4	4	4 c	4
24	2 mM CHAPS	Sodium Phosphate	100 µL, 20 mM	400	2	2	2 r	2 r	2 r

Table 6: Stabilizers, 1mM Protein Solution

Experimental conditions: 1mM mM protein solution, 2 drops on cover slide; pipette 500 µL of buffer stock solution, 200 mM, salt additive and H_2O as indicated in columns 4 and 5 into reservoir.

	Salt additive	Buffer, pH			Score				
well #	Final conc.	100 mM concentration	Salt additive	add. H_2O [µL]	2 days	5 days	10 days	19 days	70 days
1	10 % DMSO	CH3COONa pH 5.5	100 µL	400	0	0	0	4	4
2	20 % DMSO	CH3COONa pH 5.5	200 µL	300	0	0	0	3	4
3	10 % Glycerol	CH3COONa pH 5.5	200 µL (50 %	300	0	0	0	0	0
4	20 % Glycerol	CH3COONa pH 5.5	400 µL (50 %	100	0	0	0	0	0
5	5 mM ATP	CH3COONa pH 5.5	250 µL (20 mM	250	4	4	4	4	4
6	2 mM CHAPS	CH3COONa pH 5.5	100 µL (20 mM	100	0	0	0	0	0
7	0.1 % BME	CH3COONa pH 5.5	10 µL	490	0	0	0	0	0
8	0.5 % BME	CH3COONa pH 5.5	50 µL	450	0	0	0	0	0
9	1 mM DTT	CH3COONa pH 5.5	2 µL (0.5 M	500	0	0	0	0	0
10	3 mM DTT	CH3COONa pH 5.5	6 µL (0.5 M	500	0	0	0	0	0
11	1 mM SDS	CH3COONa pH 5.5	20 µL (50 mM	480	0	0	0	0	0
12	3 mM SDS	CH3COONa pH 5.5	60 µL (50 mM	440	1	1	1	1	0.5
13	6 % MeOH	CH3COONa pH 5.5	60 µL	440	0	0	0	0	0
14	6 % EtOH	CH3COONa pH 5.5	60 µL	440	0	0	0	0	0
15	NaN_3	CH3COONa pH 5.5	20 µL (50 mM	480	0	0	0	0	0
16	20 mM $MgCl_2$	CH3COONa pH 5.5	100 µL (0.2 M	400	0	0	0	0	0
17	10 % DMSO	CH3COONa pH 6.0	100 µL	400	0	0	0	1	3
18	20 % DMSO	CH3COONa pH 6.0	200 µL	300	0	0	3	4	4
19	10 % Glycerol	CH3COONa pH 6.0	200 µL (50 %	300	0	0	0	0	0.5
20	20 % Glycerol	CH3COONa pH 6.0	400 µL (50 %	100	0	0	0	0	0
21	5 mM ATP	CH3COONa pH 6.0	250 µL (20 mM	250	4	4	4	4	4
22	2 mM CHAPS	CH3COONa pH 6.0	100 µL (20 mM	100	0	0	0	0	0
23	0.1 % BME	CH3COONa pH 6.0	10 µL	490	0	0	0	0	0
24	0.5 % BME	CH3COONa pH 6.0	50 µL	450	0	0	0	0	0

Table 7: Stabilizers, 1mM Protein Solution

Experimental conditions: 1mM mM protein solution, 2 drops on cover slide; pipette 500 µL of buffer stock solution, 200 mM, salt additive and H_2O as indicated in columns 4 and 5 into reservoir.

	Salt additive	Buffer, pH			Score		
well #	Final conc.	41 mM concentration	Salt additive	add. H_2O [µL]	1 day	10 days	61 days
1		HEPES-Na pH 6.6	-	500	0	0	3
2	20 mM $CaCl_2$	HEPES-Na pH 6.6	100 µL	400	0	0.5	2
3	20 mM $MgCl_2$	HEPES-Na pH 6.6	100 µL	400	0	0.5	0.5
4	20 mM NH_4SO_4	HEPES-Na pH 6.6	100 µL	400	1	1	4
5	20 mM NH_4Ac	HEPES-Na pH 6.6	100 µL	400	0	0	4
6	25 mM NaCl	HEPES-Na pH 6.6	25 µL	475	0	0	3
7	50 mM NaCl	HEPES-Na pH 6.6	50 µL	450	0	0	3
8	100 mM NaCl	HEPES-Na pH 6.6	100 µL	400	0	0	4
9	10 % DMSO	HEPES-Na pH 6.6	100 µL	400	0	0	0
10	20 % DMSO	HEPES-Na pH 6.6	200 µL	300	0	0	3
11	10 % Glycerol	HEPES-Na pH 6.6	200 µL (50 %	300	0	0	0
12	20 % Glycerol	HEPES-Na pH 6.6	400 µL (50 %	100	0	0	0
13	5 mM ATP	HEPES-Na pH 6.6	250 µL (20 mM	250	4	4	4
14	2 mM CHAPS	HEPES-Na pH 6.6	100 µL (20 mM	400	0	0	2
15	0.1 % BME	HEPES-Na pH 6.6	10 µL	490	0	0	2 d
16	0.5 % BME	HEPES-Na pH 6.6	50 µL	450	0	0	2 d
17	1 mM DTT	HEPES-Na pH 6.6	2 µL (0.5 M	498	0	0	2
18	3 mM DTT	HEPES-Na pH 6.6	6 µL (0.5 M	494	0	0	3
19	1 mM SDS	HEPES-Na pH 6.6	20 µL (50 mM	480	1	2	2 d
20	3 mM SDS	HEPES-Na pH 6.6	60 µL (50 mM	440	4	4	4 d
21	6 % MeOH	HEPES-Na pH 6.6	60 µL	440	0	0	1 d
22	6 % EtOH	HEPES-Na pH 6.6	60 µL	440	0	0	1 d
23	1 mM EDTA	HEPES-Na pH 6.6	10 µL (100 mM	490	0	0	2 d
24	NaN_3	HEPES-Na pH 6.6	20 µL (50 mM	480	0	0	3 d

APPENDIX C

Expression and Purification of ASFV Pol X_{gpm}

Solutions for protein expression and purification

Solution "O" (dissolve in 100 mL with distilled water, sterile filter through 0.22 micron filter.)
5.76 g $MgCl_2$ * 6 H_20
2 mL $FeCl_2$ solution

$FeCl_2$ solution: (dissolve in 100 mL distilled, sterile water)
8 mL concentrated HCl
5 g $FeCl_2$ * 4 H_2O
184 mg $CaCl_2$ * 2 H_2O
64 mg H_3BO_3
40 mg $MnCl_2$ * 4 H_2O
18 mg $CoCl_2$ * 6 H_2O
4 mg $CuCl_2$ * 2 H_2O
340 mg $ZnCl_2$
605 mg Na_2MoO_4 * 2 H_2O

Vitamin Solution 10x (dissolve in 100 mL ethanol 50%)
2.2 mg biotin
2.2 mg folic acid
220 mg PABA
220 mg riboflavin
440 mg pantothenic acid
440 mg pyridoxine HCl
440 mg thiamine HCl
440 mg niacinamide

SBMX Solution (25x) (dissolve in 500 mL distilled, sterile water. Adjust to pH = 7.5)
16.15 g monobasic potassium phosphate (KH_2PO_4)
87.5 g dibasic potassium phosphate (K_2HPO_4)
18.25 g NaCl

Solution "S" (dissolve in 100 mL distilled water, sterile filter)
4.8 g K_2SO_4

Thiamine Solution (dissolve in 25 mL distilled water, sterile filter)
1 mg/mL

Buffer A (20x) (chelex the buffer solutions to avoid metal contamination)
400 mM PIPES (adjust to pH 6.5)
0.4% NaN_3
40 mM BME
100 mM EDTA

Antibiotics: (store at -20^{0}C)
Ampicillin 200 mg/mL in H_2O, sterile filtered
Chloramphenicol 68 mg/mL in ethanol

Expression and purification protocol

(Adapted with few modifications from G. Mullen)

Overnight culture

Dissolve 2 LB broth tablets (Sigma) in a 1 L Erlenmayer flask with 100 mL milli-Q water and sterilize it in the autoclave. Allow flask to cool, add 40 µL Ampicillin (200 mg/mL), 22 µL Chloramphenicol (68 mg/mL) and a flake of bacterial colony on a loop from the -80 ºC glycerol stock. Grow culture over night (~14 h) in the incubator at 37 ºC and shake at a speed of 250 rpm. Determine OD_{600} absorbance of the final culture that should be around 2.

Main culture

Prepare main culture in 2.8 L Fernbach flask. For a non-labeled growth use 16 LB broth tablets, add 800 mL milli-Q water and sterilize it in the autoclave. Allow flask to cool, add 320 µL Ampicillin, 176 µL Chloramphenicol and 10 to 20 mL of the overnight culture.

For a labeled growth in 800 mL M9 medium add 32 mL of the SBMX solution, 0.8 mL of solution S and 780 mL milli-Q water to a 2.8 L Fernbach flask and sterilize it in the autoclave. After the initial cool down, place the flasks together with the overnight cultures into the incubator to equilibrate the solution to 37 ºC overnight. Prepare the nutrient solution by dissolving 0.8 g $^{15}NH_4Cl$ and 3.2 g $^{12}C_6$-D-glucose or $^{13}C_6$-D-glucose in 10 mL milli-Q water, add 0.8 mL Vitamin solution 10x, 1.6 mL Solution O and 0.8 mL Thiamine solution and sterile filter the final nutrient solution through a 0.2 micron syringe filter. Add the nutrient solution, 320 µL Ampicillin and 176 µL Chloramphenicol to the Fernbach flask. To remove the non-labeled nutrients from the overnight culture, place it into two 50 mL Falcon tubes and gently spin them in a JA12 rotor at 700xg for 15 min. Remove the supernatant and wash the Falcons two times with sterile SBMX buffer 1x, not disturbing the pellets. Disperse the pellets in 10 mL SBMX buffer 1x and add them to the Fernbach flask.

For either culture, place the Fernbach flask into the incubator at 37 ºC and shake at a speed of 250 rpm. Determine the turbidity reading the OD_{600} absorbance. When the growth reaches an OD_{600} of at least 0.8, induce the over-expression by adding 0.8 mL IPTG 1 M and continue to incubate the cells for 4 hours. Transfer the cultures into a 1 L centrifuge bottle and spin the pallets down at 4 ºC with a JLA 9.1000 rotor at 10,000x g (7000 rpm) for 30 min and store pellets at -80 ºC.

Purification

All purification steps are performed at 4 ºC.

Transfer the pallets into a 50 mL Falcon tube and add 30 mL Buffer A, 1x that additionally includes 500 mM NaCl, 20 µL Sigma Protease Inhibitor and 0.01 % w/v lysozyme. Sonicate the solution on ice with 3 to 6 30 s bursts, allowing 30 s of cool time in-between. Once no solid cell material was visible anymore, the cell lysate was spun down for 25 min at 18,000 g using a JA-12 rotor and the cell free extract was collected in

a 50 mL baker on a magnetic stirrer plate. Slowly add ammonium sulfate to a final concentration of 45 % saturation, as listed in Table B-1 and stir the solution for 15 min.

1st Preciptiation		2nd Preciptiation	
Ammonium Sulfate	Total Volume	Ammonium Sulfate	Total Volume
13.35 g	50 mL	11.4 g	50 mL
12.02 g	45 mL	10.26 g	45 mL
10.68 g	40 mL	9.12 g	40 mL
9.35 g	35 mL	7.98 g	35 mL
8.01 g	30 mL	6.84 g	30 mL

Table B-1: Amount of ammonium sulfate required to achieve 45 % saturation during the first precipitation and 77 % for the second precipitation dependent on the supernatant volume.

Spin down for 25 min at 18,000 g using the JA-12 rotor and collect the supernatant again in the 50 mL baker. Slowly add ammonium sulfate to a final concentration of 77 % saturation, as listed in Table B-1 and stir the solution for 15 min. Centrifuge the dispersion for 25 min at 18,000 g using the JA-12 rotor but collect the pellet this time.

Dissolve the pellet in 5 to 10 mL Buffer B, which is the same as Buffer A 1x but includes 500 mM NaCl and 0.5 mM PMSF. Transfer the solution into a 15 mL Slide-A-Lyzer dialysis cassette, 10 kDa MWCO and dialyze against 2 L of Buffer B for 2 hours. The dialyzed sample is diluted with an equal volume of Buffer A, including only 0.5 mM PMSF and no salt. The lower salt concentration might cause some precipitation.

The mixture is loaded on an 20 cm CM-Sephadex column that has been equilibrated with Buffer A, 1x including 250 mM NaCl and 0.5 mM PMSF. The column is washed with at least 10 column volumes using the equilibration buffer, removing some proteins that was not precipitated during the ammonium sulfate treatment and is of no interest. ASFV pol X_{gpm} elutes when increasing the NaCl concentration from 250 mM to 1 M. Individual fractions are collected analyzed by gel chromatography using a 4-12 % Tris gel and Chomassie blue staining.

If the protein purity is not sufficient, further purification is attained by HPLC using a Phenomex Luna C-8, 25cm X 4.6 mm column. The protein was first concentrated with an Amicon Ultra-15, 10 kDa MWCO spin filter to about 1.5 mL and injected in 100 to 150 µL increments. Solvents were A = 0.1% TFA in H_2O and B = 0.1% TFA in acetonitrile. Reverse phase chromatography was run with a gradient and a flow rate of 2 mL / min on

a Beckmann HPLC System. Starting conditions were 10% B, increasing the organic phase during 30 minutes to 70 % B, stay isocratic for the following 10 minutes, reverse back to 10% B during the next 10 minutes and stay with this composition for another 10 min. The retention time for pol X_{gpm} under these conditions was 22.3 min, corresponding to an elution at 55 % B. Fractions were subsequently dried in a CentriVap console and stored dry in the refrigerator until used. The protein was re-dissolved in small amounts of 20 mM sodium cacodylate pH 6.5 followed by a buffer exchange into the final conditions using Amicon Ultra-15 spin filters.

Equipment list:

Centrifuge:	Beckmann Avanti J-25 (Beckmann Instruments, Fullerton, CA)
	Rotors: JA-12, conical, fixed angle, 50 mL
	JA-25.15, conical, fixed angle, 25 mL
	JS-7.5, conical, swing bucket, 50 mL
	JLA-9.1000, fixed angle, 1000 mL
Incubator / Shaker:	Gyrotory Shaker, (New Brunswick Scientific Co, New Brunswick, NJ)
UV-spectrophotometer:	Varian Cary 4E
Sonicator:	Branson Sonifier 450 (Branson Ultrasonics, Danbury, CT)
LP Chromatography:	BioLogic-LP Chromatography System (Bio-Rad Laboratories, Hercules, CA)
SDS-PAGE:	Bio-Rad Power Pac 3000 (Bio-Rad Laboratories, Hercules, CA)
	XCell SureLock MiniCell (Invitrogen, Carlsbad, CA)
HPLC:	Beckmann System Gold (Beckmann Instruments, Fullerton, CA)
	- diode array detector 168
	- solvent module126
Lyophylizer:	CentiVap Console (Labconco, Kansas City, MO)

APPENDIX D

ANALYSIS OF THE NMR-TITRATION EXPERIENTS

NMRPipe

Special thanks to Dr. Frank Delaglio who kindly provided the titration analysis module of NMRPipe. Since it is otherwise part of a commercial package, it might only be used for this project.

1. In the data home directory create an *in.list* of files that contain the HSQC titration data.
2. The *setC.com* file contains the protein and ligand concentration information, which will be written into each spectra header file during execution. This file is closely related to the in.list and the automv.com files and all the data entered have to be consistent.
3. Change directory to the first HSQC data set and process the experiment with NMRPipe setting all parameters correctly in the files *fid.com* and *proc.com*
4. Back in the data home directory create and run the *automv.com*, which is based on the *fid.com* and *proc.com* script. This script performs the conversion and processing for the HSQC series. It is important to write the data out as a 3D spectrum, otherwise the peak picking will not work properly.
5. Two new directories will be written, the fid directory with the converted fid data and the ft directory containing the processed spectra in the form of a pseudo-3D spectrum.
6. Making peaklists:
   The regular NMRDraw peak picking tool can be used to create consistent peaklists for the titration series HSQC spectra.
   a. In the ft directory, open the spectra in nmrDraw as a 3D template (test%03d.dat)
   b. Adjust spectrum intensity on the first plane
   c. Under “Peak Detection” set the “Peak Table” to:
   d. tab_000/test%03d.tab

e. This forces nmrDraw to perform a 2D peak picking for each plane. One peak lists for each 2D spectrum will be stored in the tab- sub-directory as part of the spectra data in ft.
f. The first peak list is used as a reference list for the following titration analysis. Peaks can be modified, deleted or assigned using the regular NMRPipe tools

If an assigned peaklist is available, the first peaklist in the newly created series can be replaced with that. Make sure the peak position match the current spectrum. It is quite often necessary to adjust the peaks to fit the spectrum well. This can be done with the help of the *assD.tcl* script. Make sure the header information fits the current definition and the appropriate filenames for the peaklists and axis definitions are given. Run the assD.tcl script and verify each individual peak using the interactive tool that is part of the assD.tcl script. Replace the first peaklist of the titration series with this newly created one. Make sure there are only tab files in this specific folder that belong to the series, otherwise the following analysis will fail. The number of peak entries for each file might vary and does not affect the analysis.

7. Visualize the titration series:

The *titrView* script sets up the peak matrix, where for each individual residue the HSQC peak is extracted for each titration point. For our 160+ assigned residues, and ten different concentration measurements, this yields a matrix of 1600+ spectra that can be visually inspected and further analyzed. Adjust the threshold value and path to the spectra and peak lists of the titration series in this script.

Starting the script with ./titrView runs the view2D module and displays first a window with all the spectra names, positive and negative color selection. Upon start, all HSQC spectra will be displayed in the overlay mode. Here one can readily pick out some of the more dramatic shift changes occurring, but since the entire spectrum is shown, this serves mainly for overview and checking purposes. Limited functions in this spectrum include zoom and expand, peak identification, assign or delete, threshold and color adjustments.

The Export menu selection draws the individual peak matrix. The first pane displays an overlay of all titration spectra for the specific peak and if selected, its assignment. From here one quickly recognizes, whether there have been changes during the titration, which then can be further inspected in the individual panes for each spectrum. If the data is used for further analysis with the module modelTitr, the crosshair can be adjusted in each individual spectrum to the optimal peak position. This titration analysis tool is geared towards titration systems with fast exchange as the primary mode. Since our case represents the tight binding case, we will analyze

the changes in with external programs such as Prizm or XCRVFIT. We will use this spectrum matrix to visually classify peak changes into the following groups:

G: Good, free standing peaks, that disappears / appear.

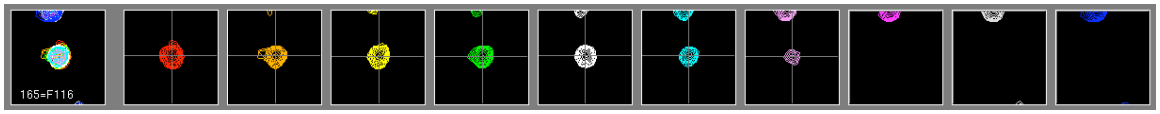
M: Medium, peak disappears / appears, but is close to other peaks, so that there is a possibility for the wrong intensity being picked during the peak analysis in the following step.

B: Bad, one sees peaks disappear / appear, but the peak is severely overlapped with other peaks.

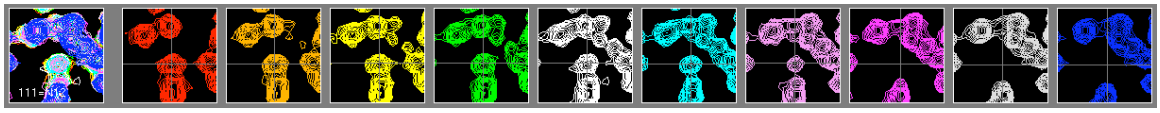
S: Slight shift of peak, more like a change in the peak shoulder. There are two cases, 1) where one peak clearly disappears / appears very close by, or 2) where a second, overlapping peak that looks like a shoulder disappears, leaving a clear, single peak or vice versa.

Examples from the titration of pol X_{gpm} with dGTP at 600 MHz, titration spectra series starting with no ligand.

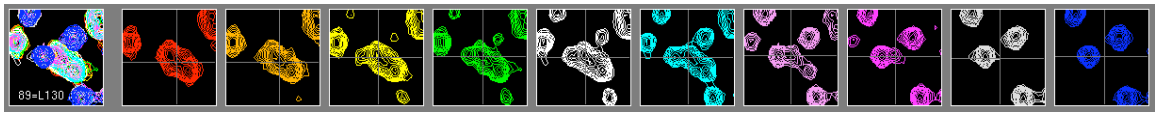
G: F116



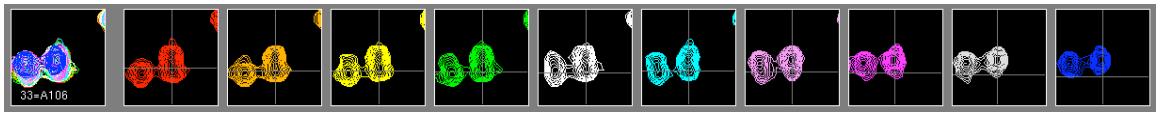
M: N12



B: V50, L130



S: A106



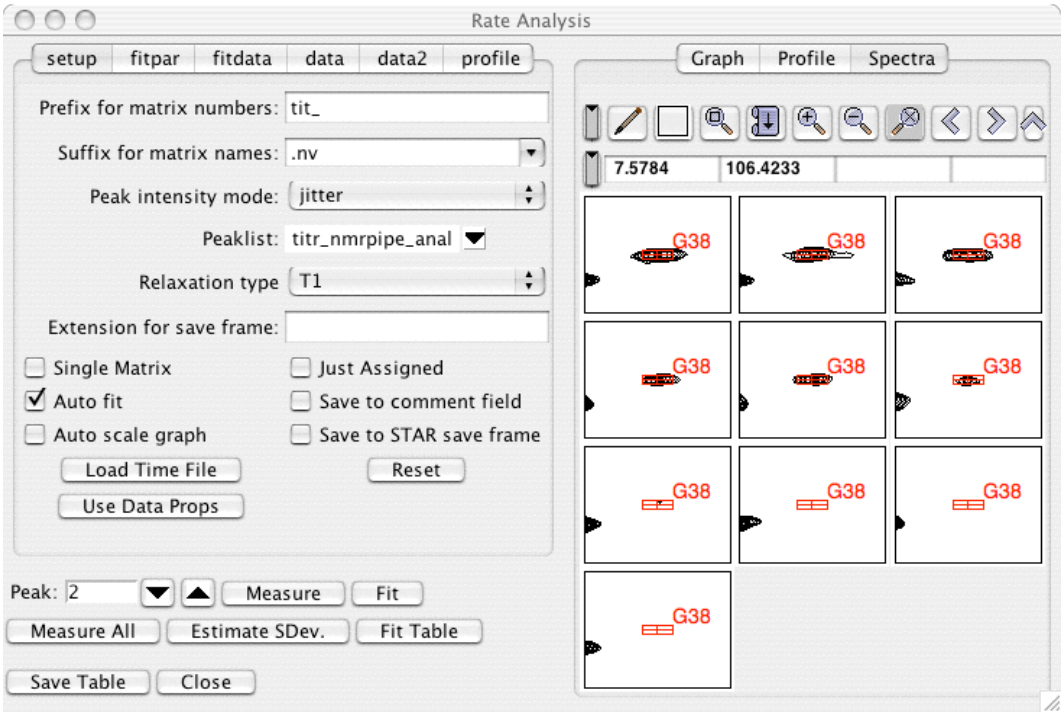
8. Determine peak intensity for curve fit analysis:
   a. Since the NMRPipe module is geared towards fast exchange analysis, only the chemical shift information is contained in the output files of that interface and the peak intensity information has to be extracted somewhere else. We chose to employ the Relaxation Tool in NMRView, which extracts peaks form a given peak

list on any number of spectra. A number of peak quantification options can be chosen from. The most suitable for our purpose was the jitter option. Starting at the peak center, this routine determines the height of the most intense point found within a ±25% range of the peak bounds. This allows for a small degree of variation of the peak from one spectrum to the other compared to the initial position in the peak list.

b. The initial peak list is a copy of the assigned peaklist but omitting all peaks that are not part of group G and M.

c. In the directory containing the NMRView spectra, create a *description* file containing a list of mol fractions for each titration point.

d. Setup of NMRView Relaxation Analysis tool:



Prefix: first part of the spectrum series file name, e.g. ***titr***_1.nv, where titr_ is the prefix, 1 is series number and .nv the extension (Suffix) fro NMRView spectra.

Peak intensity mode: Typically jitter

Peaklist: Define peaklist containing the peaks information for this analysis

Load Time File: browse for the description file

Measure All: identifies peaks and measures all intensities

Save Table: Save output data containing Fit Table with intensities and standard deviation as requested.

Peak: select individual peaks to be examined

e. Spectra of the individual peaks can be displayed as shown above or the points can be shown in an XY-plot. Both representations are very helpful during the fitting procedure to identify outlaying peaks that might need to be omitted from the fit.

9. Import the peak information into either Prizm or xcrvfit and perform the nonlinear square fit analysis with these programs. In some cases, bad data points have to be omitted in order to achieve a proper curve fit. Inspections of the NMRView tile plot and XY- data graphs aid in the process of bad data point elimination. The data format for xcrvfit is shown in the example file *xcrvfit.in*.

APPENDIX E

The following tables list the classification of 15N-HSQC correlation peaks for individual residues resulting from the analysis of the individual titration experiments performed to achieve a binary or ternary complex between ASFV pol X, oligonucleotides and nucleotides, respectively.

Table E-1: Titration Experiment of ASFV Pol X with 36-mer DNA

List of residues whose ^{15}N-HSQC signal has completely disappeared upon addition of the following mol fractions (mf) of 36-mer double hairpin DNA with a CpC-gap: 0.25 mf, 0.35 mf and 0.65 mf in column 1,2 and 3, respectively. The last column lists the residues whose peak intensities have not changed at 0.65 mf.

no peak intensity at 0.20 mf	no peak intensity at 0.35 mf	no peak intensity at 0.65 mf	Peak unchanged at 0.65 mf
		V 37	I 5
	G 38	G 38	Q 6
	S 39	S 39	G 7
		L 40	K 8
		R41	K 9
		R 42	I 10
	N 48	N 48	N 12
		D 49	H 13
		V 50	R 15
	D 51	D 51	S 16
		L 52	R 17
		I 55	L 18
		V 56	A 19
		D 58	F 20
		L 62	Y 22
		K 63	G 24
R 70	R 70	R 70	Q 25
		V 79	L 26
		G 82	K 28
	E 83	E 83	S 31
R 84	R 84	R 84	K 32
K 85	K 85	K 85	N 33
		C 86	I 34
		L 88	E 43
		Y 97	E 44

		Q 98	K 45
		L 99	M 46
		D 100	L 61
		L 101	L 74
		F 102	E 91
		T 103	W 92
	A 104	A 104	K 94
		L 105	K 95
		E 108	E 107
		Y 111	I 153
		A 112	E 162
	I 113	I 113	F 165
		F 114	T 166
	H 115	H 115	
	F 116	F 116	
	T 117	T 117	
		G 118	
	Y 122	Y 122	
	I 124	I 124	
	R 125	R 125	
	I 126	I 126	
		R 127	
		A 128	
		A 129	
	L 130	L 130	
	K 131	K 131	
		K 132	
		K 133	
		N 134	
		Y 135	
	K 136	K 136	
	L 137	L 137	
	N 138	N 138	
	Q 139	Q 139	
	Y 140	Y 140	
		G 141	
	L 142	L 142	
		K 144	
Q 146	Q 146	Q 146	
		V 149	
		K 152	
		E 156	
		K 157	
		E 158	
	K 161	K 161	
		L 163	
		R 168	
		I 169	
		K 171	

Table E-2: Titration Experiment of ASFV Pol X with 44-mer DNA

List of residues whose ^{15}N-HSQC signal has completely disappeared upon addition of the following mol fractions (mf) of 44-mer double hairpin DNA with a CpC-gap: 0.40 mf, and 0.60 mf in column 1 and 2, respectively. The third and fourth columns show the residues whose peak intensities have not changed at o.60 or 0.80 mf, respectively. Residues in red indicate peaks that were not absent in spectra of 36-mer at comparable mol fractions of DNA. Residue peaks that were absent in the 36-mer but present here at 0.60 mf are: V79, Y97, A129, L137, K161, L163, I169, K171.

No peak intensity at 0.40 mf	No peak intensity at 0.60 mf	Unchanged peaks at 0.60 mf	Unchanged peaks at 0.60 mf
L 2	L 2	I 5	I 5
V 37	N 23	R 15	S 16
G 38	V 37	S 16	G 24
S 39	G 38	F 20	Q 25
L 40	S 39	G 24	L 26
N 48	L 40	Q 25	K 28
D 51	R 41	L 26	S 31
L 52	R 42	K 28	K 32
I 55	L 47	S 31	N 33
V 56	N 48	K 32	K 45
L 62	D 49	N 33	M 46
K 63	V 50	I 34	L 61
R 70	D 51	K 45	I 153
G 82	L 52	M 46	E 162
E 83	L 53	L 61	F 165
R 84	I 55	L 74	T 166
K 85	V 56	W 92	
C 86	D 58	K 95	
L 88	L 62	E 107	
Q 98	K 63	I 153	
L 99	H 64		
D 100	V 65		
L 101	R 70		
F 102	I 71		
T 103	V 78		
A 104	G 82		
L 105	E 83		
I 113	R 84		
H 115	K 85		
F 116	C 86		
T 117	L 88		

Y 122	I 90		
I 124	Q 98		
R 125	L 99		
I 126	D 100		
K 136	L 101		
L 137	F 102		
N 138	T 103		
Y 140	A 104		
L 142	L 105		
N 145	E 108		
Q 146	Y 111		
	A 112		
	I 113		
	F 114		
	H 115		
	F 116		
	T 117		
	G 118		
	V 120		
	S 121		
	Y 122		
	I 124		
	R 125		
	I 126		
	R 127		
	A 128		
	L 130		
	K 131		
	K 132		
	K 133		
	N 134		
	Y 135		
	K 136		
	N 138		
	Q 139		
	Y 140		
	G 141		
	L 142		
	K 144		
	Q 146		
	V 149		
	K 152		
	E 156		
	K 157		
	E 158		
	T 166		
	Y 167		
	R 168		

Table E-3: Titration Experiment of ASFV Pol X with ss-DNA

The following table contains the information on ^{15}N-HSQC-peak classification for each residue as determined by the analysis of the titration spectra of the ss-DNA experiments. Classification occurred after visual inspection of each peak according to the following criteria:

g = no peak intensity observed at this position
s = some chemical shift perturbation observed for peak
ss = strong chemical shift perturbation observed for peak
w = reduced intensity for this peak
u = no appreciable change in peak shift and intensity

Residue	$dC(pdC)_8$, 5mf added	$dA(pdA)_{11}$, 5mf added	$dA(pdA)_{11}$, 5mf + 2mf ddTTP added	$dA(pdA)_{11}dC(pdC)_5$ 5mf added
M1				
L2	u	u		s,w
T3	u	u		s,w
L4	u	u		g
I5	u	u		s,w
Q6	u	u		s,w
G7	u	u		g
K8	u	u	u	s,w
K9	u	u		g
I10	u	u		s,w
V11	u	u		g
N12	u	u		s,w
H13	u	u		g
L14	u	u		g
R15	u	u		s,w
S16	u	u	u	u
R17	u	u	u	w
L18	u	u		s,w
A19	u	u		s,w
F20	u	u	u	s,w
E21	ss	u	u	s,w
Y22	u	u		w
N23	u	u		s,w
G24	u	u	u	u
Q25	u	u	u	w

L26	u	u	u	w
I27	u	u	u	w
K28	u	u	u	w
I29	u	u	u	g
L30	u	s		w
S31	u	u	u	s,w
K32	u	u	u	s,w
N33	u	u	u	g
I34	u	u	u	s,w
V35	u	u		g
A36	u	u	u	s,w
V37	u	s		g
G38	u	ss		g
S39				g
L40	ss	u		g
R41	ss	u		w
R42	u	u		ss
E43	u	u		w
E44	u	u		w
K45	u	u		w
M46	u	u		w
L47	s	u		g
N48	u	u		g
D49	u	u		g
V50	u	s		g
D51	ss	u		g
L52	ss	u		g
L53	u	u		g
I54	u	u		g
I55	u	u		g
V56	u	u		g
P57	u	u		
E58	u	ss		ss
K59	u	u		w
K60	u	s		w
L61	u	u		w
L62	u	s		g
K63	u	s		g
H64	u	u		s,w
V65	u	u	u	g
L66				w
P67				
N68	u	u		s,w
I69	u	u	u	g
R70	s	ss		g
I71	u	u		s,w
K72	u	u		g
G73	s	u		w

L74	u	u		u
S75	u	u		w
F76	u	u	u	s,w
S77	u	u		g
V78	u	u		w
K79	u	s		g
V80	u	u		s,w
C81	u	u		g
G82	u	u		g
E83	s	ss		g
R84	g	g		g
K85	g	g		g
C86	s	u		w
V87				
L88	s	u		g
F89	u	u		s,w
I90	u	u		g
E91	u	u		w
W92	u	u	u	w
E93	u	u		w
K94	u	u	u	w
K95	u	u		w
T96	u	u	u	w
Y97	u	u		w
Q98	u	u		w
L99	u	u		g
D100	g	g		g
L101	s	u		g
F102	u	u		g
T103	u	u		w
A104	u	u		g
L105	ss	u		g
A106	ss	u		g
E107	u	u		w
E108	s	s		g
K109				w
P110				
Y111	u	s		ss
A112	u	s		g
I113	s	s		g
F114	u	u		g
H115	u	s		g
F116	s	ss		g
T117	ss	s		g
G118	u	s		g
P119				
V120	ss	u		g
S121	u	s		g

Y122	u	s		g
L123	u	u		w
I124	ss	ss		g
R125	s	ss		g
I126	ss	ss		g
R127	s	s		g
A128	ss	ss		g
A129	s	s		g
L130	s	s		g
K131	s	ss		g
K132	s	ss		g
K133	u	s		g
N134	u	s		ss
Y135	s	s		g
K136	u	ss		g
L137	u	ss		g
N138	s	g		g
Q139	ss	ss		g
Y140	u	ss		g
G141	u	s		g
L142	ss	u		g
F143	ss	s		ss
K144	u	s		ss
N145	u	s		w
Q146	g	s		g
T147	u	u		s,w
L148	u	u		g
V149	s	s		g
P150				
L151	u	u		s,w
K152	ss	u		g
I153	u	u		s,w
T154	ss	u		g
T155	u	u		w
E156	s	u		w
K157	u	u		s,w
E158	ss	s		g
L159	u	s		g
I160	u	s		g
K161	u	s		w
E162	u	u		w
L163	u	u		g
G164	u	u		g
F165	u	u		w
T166	u	u		w
Y167	u	u		g
R168	u	u		g
I169	u	u		g

P170				
K171	u	u		w
K172	u	u		s,w
R173	u	u		w
L174	s	s		s,w

APPENDIX F

NMR Data Processing Scripts for NMRPipe

HSQC

```
#!/bin/csh -f
      nmrPipe -in test.fid                                  \
      nmrPipe  -fn SOL                                      \
      nmrPipe  -fn SP -off 0.40 -end 0.9 -pow 2 -c 1        \
#nmrPipe  -fn GM -g1 15 -g2 20 -c 0.5                       \
      nmrPipe  -fn ZF -size 2048                            \
      nmrPipe  -fn FT -verb                                 \
      nmrPipe  -fn PS -p0 -53.8 -p1 20.8 -di                \
      nmrPipe  -fn POLY -auto -verb                         \
      nmrPipe  -fn EXT -x1 5.5ppm -xn 11ppm -sw             \
      nmrPipe  -fn TP                                       \
#nmrPipe  -fn LP -xn 512 -fb -ord 8                         \
      nmrPipe  -fn SP -off 0.40 -end 0.9 -pow 2 -c 1        \
      nmrPipe  -fn ZF -size 512                             \
      nmrPipe  -fn FT -verb                                 \
      nmrPipe  -fn PS -p0 -76.0 -p1 158.2 -di               \
      nmrPipe  -fn POLY -auto -verb                         \
      nmrPipe  -fn TP                                       \
  -ov -out <file>.dat

#Conversion to NMRVIEW:
nmrPipe -in <file>.dat | pipe2xyz -out <file>.nv -nv -verb
```

HNCO

```
#!/bin/csh
      xyz2pipe -in fid/test%03d.fid -x –verb                \
      | nmrPipe -fn SOL                                     \
      | nmrPipe -fn SP -off 0.35 -end 0.90 -pow 2 -c 0.5    \
      | nmrPipe -fn ZF –auto                                \
      | nmrPipe -fn FT                                      \
      | nmrPipe -fn PS -p0 -11.2 -p1 -19.0 –di              \
```

```
      | nmrPipe  -fn POLY -ord 2 –auto                  \
      | nmrPipe -fn EXT -x1 5.6ppm -xn 10.5ppm –sw      \
      | pipe2xyz -out ft/test%03d.ft3 -x

# 13C dimension                                         \
      xyz2pipe -in ft/test%03d.ft3 -z –verb             \
      | nmrPipe -fn SP -off 0.5 -end 0.98 -pow 1 -c 0.5 \
      | nmrPipe -fn ZF –auto                            \
      | nmrPipe -fn FT                                  \
      | nmrPipe -fn PS -p0 59.5 -p1 180 –di             \
      | nmrPipe  -fn POLY -ord 1 –auto                  \
      | pipe2xyz -out ft/test%03d.ft2 -y
#exit

# 15N dimension                                         \
      xyz2pipe -in ft/test%03d.ft2 -z –verb             \
      | nmrPipe -fn LP –fb                              \
      | nmrPipe -fn SP -off 0.5 -end 1.0 -pow 1 -c 0.5  \
      | nmrPipe -fn ZF –auto                            \
      | nmrPipe -fn FT                                  \
      | nmrPipe -fn PS -p0 -2.0 -p1 4.0 –di             \
      | pipe2xyz -out ft/test%03d.ft3 -z -inPlace

#Conversion to NMRVIEW:
      xyz2pipe -in ft/test%03d.dat | pipe2xyz -out hnco_comp.nv -nv -verb
```

HNCA

```
#!/bin/csh
      xyz2pipe -in fid/test%03d.fid -x -verb            \
      | nmrPipe  -fn SOL                                \
      | nmrPipe  -fn SP -off 0.5 -end 0.98 -pow 2 -c 0.5 \
      | nmrPipe  -fn ZF -auto                           \
      | nmrPipe  -fn FT                                 \
      | nmrPipe  -fn PS -p0 14.0 -p1 0 -di              \
      | nmrPipe  -fn EXT -x1 5.6ppm -xn 10.5ppm -sw     \
      #| nmrPipe  -fn POLY -ord 2 -auto                 \
      | pipe2xyz -out ft/test%03d.fth -x

# 13C dimension
      xyz2pipe -in ft/test%03d.fth -z -verb             \
      | nmrPipe -fn ZF  -pad 1                          \
      | nmrPipe -fn RS  -rs 1 -sw                       \
      | nmrPipe -fn LP  -before -pred 1                 \
```

```
      | nmrPipe  -fn SP -off 0.5 -end 0.98 -pow 2 -c 1.0        \
      | nmrPipe  -fn ZF -size 256                               \
      | nmrPipe  -fn FT                                         \
      | nmrPipe  -fn PS -p0 115.0 -p1 187.0 -di                 \
      | nmrPipe  -fn POLY -ord 1 -auto                          \
      | pipe2xyz -out ft/test%03d.ftc -y
#exit

# 15N dimension
      xyz2pipe -in ft/test%03d.ftc -z  -verb                    \
      | nmrPipe  -fn LP -fb -auto                               \
      | nmrPipe  -fn SP -off 0.5 -end 0.98 -pow 1 -c 0.5        \
      | nmrPipe  -fn ZF -auto                                   \
      | nmrPipe  -fn FT                                         \
      | nmrPipe  -fn PS -p0 -1.9 -p1 -1.0 -di                   \
      | pipe2xyz -out ft/test%03d.dat -z                        \
#exit

#Conversion to NMRVIEW:
xyz2pipe -in ft/test%03d.dat | pipe2xyz -out hnca_comp.nv -nv -verb
```

CBCA(CO)NH

```
#!/bin/csh
      xyz2pipe -in fid/test%03d.fid -x -verb                    \
      | nmrPipe -fn SOL                                         \
      | nmrPipe -fn SP -off 0.45 -end 0.98 -pow 2 -c 0.5        \
      | nmrPipe -fn ZF -auto                                    \
      | nmrPipe -fn FT                                          \
      | nmrPipe -fn PS -p0 72.7 -p1 -5.0 -di                    \
      | nmrPipe -fn EXT -x1 5.6ppm -xn 10.5ppm -sw              \
      | nmrPipe  -fn POLY -ord 0 -auto                          \
      | pipe2xyz -out ft/test%03d.ft3 -x

# 15N dimension:
      xyz2pipe -in ft/test%03d.ft3 -z -verb                     \
      | nmrPipe -fn SP -off 0.5 -end 0.98 -pow 1 -c 0.5         \
      | nmrPipe -fn ZF -auto                                    \
      | nmrPipe -fn FT -neg                                     \
      | nmrPipe -fn PS -p0 1.2 -p1 1.0 -di                      \
      | pipe2xyz -out ft/test%03d.ft3 -z -inPlace
#exit
```

```
# 13C dimension:
      xyz2pipe -in ft/test%03d.ft3 -y -verb                \
      | nmrPipe -fn LP -ps0-0                              \
      | nmrPipe -fn SP -off 0.45 -end 1.0 -pow 2 -c 0.5    \
      | nmrPipe -fn ZF -auto                               \
      | nmrPipe -fn FT                                     \
      | nmrPipe -fn PS -p0 31.1 -p1 -11.0 -di              \
      | pipe2xyz -out ft/test%03d.ft3 -y -inPlace
#exit

# 15N dimension linear prediction:
      xyz2pipe -in ft/test%03d.ft3 -z -verb                \
      | nmrPipe -fn HT                                     \
      | nmrPipe -fn PS -inv -hdr                           \
      | nmrPipe -fn FT -inv                                \
      | nmrPipe -fn ZF -inv                                \
      | nmrPipe -fn SP -inv -hdr                           \
      | nmrPipe -fn LP -ps0-0                              \
      | nmrPipe -fn SP -off 0.5 -end 0.98 -pow 2 -c 0.5    \
      | nmrPipe -fn ZF -auto                               \
      | nmrPipe -fn FT                                             \
      | nmrPipe -fn PS -p0 1.2 -p1 1.0 -di                 \
      | pipe2xyz -out ft/test%03d.dat -z
#exit

#Conversion to NMRVIEW:
xyz2pipe -in ft/test%03d.dat | pipe2xyz -out cbcaconh_comp.nv -nv -verb
```

HNCACB

```
#!/bin/csh
      xyz2pipe -in fid/test%03d.fid -x -verb               \
      | nmrPipe -fn SOL                                    \
      | nmrPipe -fn SP -off 0.5 -end 0.98 -pow 2 -c 0.5    \
      | nmrPipe -fn ZF -auto                               \
      | nmrPipe -fn FT                                     \
      | nmrPipe -fn PS -p0 125.8 -p1 -12.0 -di             \
      | nmrPipe  -fn POLY -ord 2 -auto                     \
      | nmrPipe -fn EXT -x1 5.6ppm -xn 10.5ppm -sw         \
      | pipe2xyz -out ftx/test%03d.ft -x

# 13C dimension
      xyz2pipe -in ftx/test%03d.ft -z -verb                \
      | nmrPipe -fn SP -off 0.5 -end 0.98 -pow 1 -c 0.5    \
```

```
      | nmrPipe -fn ZF -auto                                \
      | nmrPipe -fn FT                                    \
      | nmrPipe -fn PS -p0 -0 -p1 0 -di                   \
      | pipe2xyz -out ftxy/test%03d.ft -y
#exit

# 15N dimension
      xyz2pipe -in ftxy/test%03d.ft -z -verb              \
      | nmrPipe -fn LP -fb                                \
      | nmrPipe -fn SP -off 0.5 -end 1.0 -pow 1 -c 0.5    \
      | nmrPipe -fn ZF -auto                              \
      | nmrPipe -fn FT                                    \
      | nmrPipe -fn PS -p0 -0 -p1 -0.0 -di                \
      | pipe2xyz -out ftxyz/test%03d.ft -z
#exit

# 13C dimension linear prediction
      xyz2pipe -in ftxyz/test%03d.ft -y -verb             \
      | nmrPipe -fn HT                                                \
      | nmrPipe -fn PS -inv -hdr                          \
      | nmrPipe -fn FT -inv                               \
      | nmrPipe -fn ZF -inv                               \
      | nmrPipe -fn SP -inv -hdr                          \
      | nmrPipe -fn LP -ps90-180                          \
      | nmrPipe -fn SP -off 0.5 -end 0.98 -pow 1 -c 0.5   \
      | nmrPipe -fn ZF -auto                              \
      | nmrPipe -fn FT                                    \
      | nmrPipe -fn PS -p0 0 -p1 0.0 -di                  \
      | pipe2xyz -out ftxyz/test%03d.dat -y
#exit

#Conversion to NMRVIEW:
xyz2pipe -in ftxyz/test%03d.dat  | pipe2xyz -out hncacb_comp_lp.nv -nv -verb
```

HBHA(CO)NH

```
#!/bin/csh
      xyz2pipe -in fid/test%03d.fid -x -verb              \
      | nmrPipe -fn SOL                                   \
      | nmrPipe -fn SP -off 0.45 -end 0.98 -pow 2 -c 0.5  \
      | nmrPipe -fn ZF -auto                              \
      | nmrPipe -fn FT                                    \
      | nmrPipe -fn PS -p0 0 -p1 0 -di                    \
      | nmrPipe -fn EXT -x1 5.6ppm -xn 10.5ppm -sw        \
```

```
      | nmrPipe  -fn POLY -ord 0 -auto                \
      | pipe2xyz -out ft/test%03d.ft1 -x

# 15N dimension:
      xyz2pipe -in ft/test%03d.ft1 -y -verb         \
      | nmrPipe -fn SP -off 0.45 -end 0.98 -pow 2    \
      | nmrPipe -fn ZF -auto                         \
      | nmrPipe -fn FT                               \
      | nmrPipe -fn PS -p0 159.6 -p1 -36.0 -di       \
      | pipe2xyz -out ft/test%03d.ft2 -z

# 1H dimension:
      xyz2pipe -in ft/test%03d.ft2 -y -verb         \
      | nmrPipe -fn SP -off 0.5 -end 0.98 -pow 1 -c 0.5  \
      | nmrPipe -fn ZF -auto                         \
      | nmrPipe -fn FT -alt                          \
      | nmrPipe -fn PS -p0 0 -p1 0 -di               \
      | pipe2xyz -out ft/test%03d.ft3 -y
#exit

#Conversion to NMRVIEW:
xyz2pipe -in ft/test%03d.dat | pipe2xyz -out hbhaconh_comp.nv -nv –verb
```

(H)CC(CO)NH

```
#!/bin/csh

xyz2pipe -in fid/test%03d.fid -x -verb          \
      #| nmrPipe  -fn POLY -time                \
      | nmrPipe  -fn SOL                        \
      | nmrPipe  -fn SP -off 0.5 -end 0.98 -pow 2 -c 1.0  \
      | nmrPipe  -fn ZF -size 2048              \
      | nmrPipe  -fn FT -auto                   \
      | nmrPipe  -fn PS -p0 -20.9 -p1 -7.0 -di  \
      | nmrPipe  -fn EXT -x1 5.6ppm -xn 10.4ppm -sw  \
      | nmrPipe  -fn POLY -ord 1 -auto          \
      | pipe2xyz -out ft/test%03d.ftx -x

# 13C dimension
xyz2pipe -in ft/test%03d.ftx -z -verb           \
      | nmrPipe  -fn SP -off 0.5 -end 0.95 -pow 2 -c 1  \
      | nmrPipe  -fn ZF -size 256               \
      | nmrPipe  -fn FT -neg                    \
      | nmrPipe  -fn POLY -ord 1 -auto          \
```

```
      | nmrPipe  -fn PS -p0 -2.80 -p1 -14.0 -di                \
      | pipe2xyz -out ft/test%03d.ftxy -y
#exit

# 15N dimension
xyz2pipe -in ft/test%03d.ftxy -z -verb                         \
      | nmrPipe  -fn LP -ps0-0 -auto                           \
      | nmrPipe  -fn SP -off 0.5 -end 0.98 -pow 2 -c 0.5       \
      | nmrPipe  -fn ZF -size 256                              \
      | nmrPipe  -fn FT                                        \
      | nmrPipe  -fn PS -p0 0 -p1 0.0 -di                      \
      | pipe2xyz -out ft/test%03d.ftxyz -z
#exit

#Conversion to NMRVIEW:
xyz2pipe -in ft/test%03d.ftxyz | pipe2xyz -out <file>.nv -nv -verb
```

H(CCCO)NH

```
#!/bin/csh
xyz2pipe -in fid/test%03d.fid -x -verb                         \
      | nmrPipe  -fn SOL                                       \
      | nmrPipe  -fn SP -off 0.45 -end 1.00 -pow 2 -c 0.5      \
      | nmrPipe  -fn ZF -auto                                  \
      | nmrPipe  -fn FT -auto                                  \
      | nmrPipe  -fn PS -p0 -150.0 -p1 -5.0 -di                \
      | nmrPipe  -fn EXT -x1 5.6ppm -xn 10.4ppm -sw            \
      | nmrPipe  -fn POLY -ord 1 -auto                         \
      | pipe2xyz -out ftx/test%03d.ft -x
#exit
# 1H dimension
xyz2pipe -in ftx/test%03d.ft -z -verb                          \
      | nmrPipe  -fn SP -off 0.45 -end 1 -pow 1 -c 1           \
      | nmrPipe  -fn ZF -size 256                              \
      | nmrPipe  -fn FT                                        \
      | nmrPipe  -fn PS -p0 -0 -p1 00.0 -di                    \
      | nmrPipe  -fn POLY -ord 1 -auto                         \
      | pipe2xyz -out ftxz/test%03d.ft2 -y
#exit
# 15N dimension
xyz2pipe -in ftxz/test%03d.ft2 -z -verb                        \
      | nmrPipe  -fn LP -ps0-0 -auto                           \
      | nmrPipe  -fn SP -off 0.5 -end 1.00 -pow 1 -c 0.5       \
      | nmrPipe  -fn ZF -size 128                              \
```

```
      | nmrPipe  -fn FT                                          \
      | nmrPipe  -fn PS -p0 -0.0 -p1 0.0 -di                     \
      | pipe2xyz -out ftxyz/test%03d.ft3 -z
#exit
# apply circular shift:
xyz2pipe -in ftxyz/test%03d.ft3 -y -verb                         \
| nmrPipe -fn CS -ls 2.0ppm -sw                                  \
| nmrPipe  -fn POLY -ord 1 -auto                                 \
| pipe2xyz -out ftcs/test%03d.ft3 -y
#Conversion to NMRVIEW:
xyz2pipe -in ftcs/test%03d.ft3 | pipe2xyz -out <file>.nv -nv -verb
```

APPENDIX G

NMR Assignment of ^{1}H, ^{13}C and ^{15}N Resonances of ASFV Pol X_{gpm} in a Complex with Deoxyguanosine-Triphosphate.

Buffer: 20 mM PIPES, 20 mM $MgCl_2$, 0.02 % NaN_3, 5 mM DTT, 500 mM NaCl, 20 ºC, reference TSP-d4

1.CA	55.052	4.N	121.943	6.HB2	1.865	9.HB2	2.039
1.HA	4.475	4.HN	7.742	6.HB1	2.344	9.HB1	2.039
1.CB	34.069	4.CA	58.622	6.CG	34.154	9.CG	25.840
1.HB2	2.170	4.HA	4.061	6.HG2	2.456	9.HG2	1.560
1.HB1	2.170	4.CB	40.519	6.HG1	2.456	9.HG1	1.560
1.CG	31.711	4.HB2	1.897	6.C	178.662	9.CD	29.834
1.HG2	2.602	4.HB1	1.897	7.N	107.006	9.HD2	1.823
1.HG1	2.772	4.CG	25.500	7.HN	8.699	9.HD1	1.823
1.C	171.830	4.CD1	23.817	7.CA	47.579	9.CE	42.279
2.N	121.002	4.HD11	0.836	7.HA2	4.014	9.C	179.960
2.HN	9.177	4.CD2	23.817	7.HA1	4.014	10.N	121.805
2.CA	54.832	4.HD21	1.209	7.C	175.836	10.HN	8.032
2.HA	4.946	4.C	178.300	8.N	120.547	10.CA	65.585
2.CB	43.639	5.N	115.153	8.HN	8.451	10.HA	2.793
2.HB2	1.588	5.HN	7.781	8.CA	60.592	10.CB	37.404
2.HB1	1.588	5.CA	64.282	8.HA	3.817	10.HB	1.484
2.CG	26.503	5.HA	3.764	8.CB	32.613	10.CG1	28.805
2.HG	1.721	5.CB	38.026	8.HB2	1.956	10.CD1	14.015
2.CD1	25.454	5.HB	1.869	8.HB1	1.956	10.HD11	0.384
2.HD11	0.370	5.CG1	29.125	8.CG	26.632	10.CG2	16.653
2.CD2	25.454	5.HG12	1.649	8.HG2	1.688	10.HG21	0.384
2.HD21	0.370	5.HG11	1.338	8.HG1	1.688	10.C	177.853
2.C	175.476	5.CD1	13.092	8.CD	29.986	11.N	119.675
3.N	112.467	5.HD11	0.982	8.HD2	1.890	11.HN	8.170
3.HN	8.667	5.CG2	17.617	8.HD1	1.890	11.CA	68.343
3.CA	61.298	5.HG21	0.982	8.CE	42.279	11.HA	3.213
3.HA	5.008	5.C	178.438	8.C	179.634	11.CB	31.256
3.CB	71.377	6.N	121.501	9.N	120.598	11.HB	2.025
3.HB	4.837	6.HN	7.843	9.HN	7.940	11.CG2	21.379
3.CG2	22.382	6.CA	58.826	9.CA	60.001	11.HG21	0.880
3.HG21	1.434	6.HA	4.086	9.HA	4.078	11.CG1	25.721
3.C	175.683	6.CB	29.300	9.CB	32.673	11.HG11	0.880

11.C	177.205	16.HN	7.055	20.HB1	3.074	26.CA	54.441
12.N	115.822	16.CA	57.542	20.C	172.576	26.HA	4.482
12.HN	7.972	16.HA	4.165	21.N	122.289	26.CB	42.045
12.CA	56.676	16.CB	64.278	21.HN	7.913	26.HB2	1.477
12.HA	4.331	16.HB2	3.820	21.CA	55.407	26.HB1	1.477
12.CB	38.499	16.HB1	3.820	21.HA	4.954	26.CG	27.775
12.HB2	2.885	16.C	176.213	21.CB	31.532	26.HG	1.304
12.HB1	2.885	17.N	122.665	21.HB2	1.784	26.CD1	24.608
12.C	177.278	17.HN	7.891	21.HB1	2.039	26.HD11	0.750
13.N	119.299	17.CA	56.971	21.CG	36.622	26.CD2	24.608
13.HN	7.911	17.HA	4.563	21.C	174.867	26.HD21	0.750
13.CA	59.540	17.CB	33.147	22.N	126.130	26.C	176.008
13.HA	4.622	17.HB2	1.255	22.HN	8.956	27.N	130.353
13.CB	31.786	17.HB1	1.294	22.CA	55.902	27.HN	8.960
13.HB2	3.227	17.CG	29.083	22.HA	2.955	27.CA	59.788
13.HB1	3.227	17.HG2	1.808	22.CB	40.997	27.HA	4.006
13.C	177.655	17.HG1	1.808	22.HB2	2.637	27.CB	36.844
14.N	115.900	17.CD	42.511	22.HB1	2.637	27.HB	0.644
14.HN	8.453	17.HD2	2.942	22.C	174.809	27.CG1	27.141
14.CA	57.185	17.HD1	2.456	23.N	128.085	27.HG12	1.105
14.HA	3.858	17.C	174.533	23.HN	9.144	27.HG11	0.975
14.CB	42.213	18.N	122.094	23.CA	53.412	27.CD1	12.224
14.HB2	1.788	18.HN	8.911	23.HA	2.758	27.HD11	0.714
14.HB1	1.788	18.CA	55.061	23.CB	36.933	27.CG2	17.565
14.CG	27.796	18.HA	5.067	23.HB2	1.868	27.HG21	0.714
14.HG	1.223	18.CB	44.436	23.HB1	1.868	27.C	175.136
14.CD1	23.484	18.HB2	1.942	23.C	175.469	28.N	125.884
14.HD11	0.733	18.HB1	1.942	24.N	103.479	28.HN	8.335
14.CD2	23.484	18.CG	25.112	24.HN	8.682	28.CA	57.412
14.HD21	0.653	18.HG	1.289	24.CA	45.477	28.HA	4.134
14.C	179.148	18.HD11	0.728	24.HA2	3.626	28.CB	32.878
15.N	115.299	18.HD21	0.728	24.HA1	4.086	28.HB2	1.741
15.HN	8.001	18.C	176.354	24.C	173.684	28.HB1	1.741
15.CA	60.330	19.N	125.897	25.N	119.503	28.CG	25.303
15.HA	3.907	19.HN	8.765	25.HN	7.679	28.HG2	1.352
15.CB	30.110	19.CA	51.386	25.CA	53.697	28.HG1	1.352
15.HB2	1.841	19.HA	5.125	25.HA	4.578	28.CD	29.262
15.HB1	1.997	19.CB	23.990	25.CB	31.834	28.HD2	1.541
15.CG	28.095	19.HB1	1.226	25.HB2	1.985	28.HD1	1.541
15.HG2	1.728	19.C	173.834	25.HB1	1.985	28.CE	42.692
15.HG1	1.728	20.N	113.911	25.CG	33.654	28.HE2	3.037
15.CD	43.324	20.HN	8.899	25.HG2	2.261	28.HE1	3.037
15.HD2	2.490	20.CA	56.072	25.HG1	2.261	28.C	176.743
15.HD1	2.490	20.HA	4.210	25.C	173.606	29.N	126.129
15.C	176.237	20.CB	42.655	26.N	124.407	29.HN	8.183
16.N	108.480	20.HB2	3.185	26.HN	8.550	29.CA	60.754

29.HA	4.232	33.HN	7.999	37.CG1	23.062	42.HB2	1.890
29.CB	38.207	33.CA	52.426	37.HG11	0.851	42.HB1	1.890
29.HB	1.754	33.HA	4.994	37.C	175.012	42.CG	28.561
29.CG1	28.915	33.CB	38.992	38.N	108.488	42.C	172.815
29.CD1	14.264	33.HB2	2.860	38.HN	8.228	43.N	112.643
29.CG2	17.436	33.HB1	2.860	38.CA	46.356	43.HN	7.793
29.HG21	0.981	33.C	172.333	38.HA2	4.855	43.CA	57.105
29.C	175.859	34.N	119.375	38.HA1	4.495	43.HA	3.730
30.N	129.059	34.HN	7.509	38.C	174.591	43.CB	26.179
30.HN	9.157	34.CA	61.049	39.N	116.273	43.HB2	2.128
30.CA	54.922	34.HA	4.837	39.HN	8.298	43.HB1	2.128
30.HA	4.328	34.CB	40.261	39.CA	61.364	43.CG	36.629
30.CB	41.989	34.HB	1.885	39.HA	4.346	43.HG2	2.227
30.HB2	1.769	34.CG1	27.616	39.CB	62.821	43.HG1	2.227
30.HB1	1.769	34.HG12	0.684	39.HB2	3.955	43.C	175.999
30.CG	25.710	34.HG11	0.859	39.HB1	3.837	44.N	119.346
30.HG	1.769	34.CD1	16.449	39.C	175.931	44.HN	7.350
30.CD1	22.405	34.HD11	0.684	40.N	122.778	44.CA	57.655
30.HD11	0.984	34.CG2	17.997	40.HN	8.505	44.HA	4.179
30.CD2	22.405	34.HG21	0.684	40.CA	59.634	44.CB	28.932
30.HD21	0.984	34.C	175.669	40.HA	4.137	44.HB2	2.253
30.C	178.702	35.N	129.016	40.CB	42.986	44.HB1	2.448
31.N	117.512	35.HN	9.632	40.HB2	1.799	44.CG	38.075
31.HN	8.728	35.CA	60.583	40.HB1	2.051	44.HG2	1.537
31.CA	61.749	35.HA	4.430	40.HG	1.799	44.HG1	1.784
31.HA	4.199	35.CB	35.964	40.CD1	26.466	44.C	174.860
31.CB	63.010	35.HB	1.947	40.HD11	0.823	45.N	115.796
31.HB2	3.989	35.CG2	20.245	40.CD2	26.466	45.HN	8.044
31.HB1	3.989	35.HG21	0.882	40.HD21	0.823	45.CA	59.880
31.C	177.388	35.CG1	20.245	40.C	181.047	45.HA	4.350
32.N	115.374	35.HG11	0.882	41.N	121.608	45.CB	33.280
32.HN	8.208	35.C	175.119	41.HN	6.954	45.HB2	1.892
32.CA	57.776	36.N	130.283	41.CA	59.112	45.HB1	1.892
32.HA	4.293	36.HN	8.867	41.HA	4.013	45.CG	25.615
32.CB	31.230	36.CA	53.190	41.CB	29.549	45.HG2	1.557
32.HB2	1.826	36.HA	4.276	41.HB2	1.619	45.HG1	1.557
32.HB1	1.957	36.CB	19.085	41.HB1	1.619	45.CD	29.214
32.CG	23.271	36.HB1	1.402	41.CD	42.534	45.HD2	1.655
32.HG2	1.412	36.C	176.860	41.HD2	3.493	45.HD1	1.774
32.HG1	1.412	37.N	116.091	41.HD1	3.493	45.CE	42.277
32.CD	29.369	37.HN	8.462	41.C	176.972	45.C	177.680
32.HD2	1.709	37.CA	59.873	42.N	112.821	46.N	115.254
32.HD1	1.709	37.HA	4.585	42.HN	7.007	46.HN	7.734
32.CE	42.092	37.CB	33.197	42.CA	57.575	46.CA	52.361
32.C	176.133	37.CG2	17.884	42.HA	3.665	46.HA	5.188
33.N	118.840	37.HG21	0.423	42.CB	30.436	46.CB	34.348

46.HB2	2.401	51.CB	43.397	55.CG1	27.254	59.HE2	3.033
46.HB1	2.401	51.HB2	2.418	55.CD1	16.474	59.HE1	3.033
46.CG	32.295	51.HB1	2.418	55.CG2	17.718	59.C	179.123
46.HG2	2.681	51.C	173.639	55.HG21	0.750	60.N	117.905
46.HG1	2.681	52.N	124.847	55.C	175.668	60.HN	9.068
46.C	174.159	52.HN	9.599	56.N	129.600	60.CA	58.650
47.N	120.751	52.CA	53.678	56.HN	9.249	60.HA	4.268
47.HN	9.611	52.HA	4.822	56.CA	59.275	60.CB	31.680
47.CA	52.204	52.CB	46.656	56.CB	33.724	60.HB2	1.918
47.HA	4.699	52.HB2	1.352	57.CA	65.194	60.HB1	2.044
47.CB	44.187	52.HB1	1.505	57.HA	4.401	60.CG	24.675
47.CG	27.615	52.HG	1.352	57.CB	33.071	60.HG2	1.455
47.HD11	0.789	52.CD1	24.905	57.HB2	2.117	60.HG1	1.275
47.HD21	0.855	52.HD11	0.673	57.HB1	2.460	60.CD	29.508
47.C	177.566	52.CD2	24.905	57.CG	27.835	60.HD2	1.763
48.N	123.223	52.HD21	0.673	57.HG2	2.117	60.HD1	1.763
48.HN	10.195	52.C	172.860	57.HG1	2.117	60.CE	41.769
48.CA	55.783	53.N	125.865	57.CD	51.599	60.HE2	3.653
48.HA	4.627	53.HN	9.126	57.HD2	4.001	60.HE1	2.994
48.CB	38.499	53.CA	53.375	57.HD1	4.001	60.C	176.989
48.HB2	2.864	53.HA	4.866	57.C	176.585	61.N	115.334
48.HB1	2.864	53.CB	45.645	58.N	113.753	61.HN	7.382
48.C	173.929	53.HB2	2.142	58.HN	6.743	61.CA	55.271
49.N	119.669	53.HB1	2.142	58.CA	54.747	61.HA	4.234
49.HN	7.815	53.CG	27.665	58.HA	4.831	61.CB	42.129
49.CA	51.701	53.CD1	23.771	58.CB	32.770	61.HB2	1.300
49.HA	4.311	53.HD11	0.714	58.HB2	2.169	61.HB1	1.674
49.CB	41.470	53.CD2	23.771	58.HB1	2.169	61.CG	25.906
49.HB2	2.805	53.HD21	0.960	58.CG	36.103	61.HG	0.957
49.HB1	2.805	53.C	173.580	58.HG2	2.303	61.CD1	21.672
49.C	174.875	54.N	127.350	58.HG1	2.303	61.HD11	0.448
50.N	117.687	54.HN	9.305	58.C	175.910	61.CD2	21.672
50.HN	7.917	54.CA	59.328	59.N	126.470	61.HD21	0.448
50.CA	62.740	54.HA	4.863	59.HN	9.044	61.C	178.645
50.HA	4.137	54.CB	39.869	59.CA	60.296	62.N	119.729
50.CB	33.346	54.CG1	27.777	59.HA	3.905	62.HN	7.480
50.HB	1.794	54.CD1	14.526	59.CB	32.275	62.CA	59.339
50.CG2	21.416	54.CG2	17.589	59.HB2	1.971	62.HA	3.853
50.HG21	0.715	54.HG21	0.723	59.HB1	1.971	62.CB	41.843
50.CG1	21.416	54.C	175.040	59.CG	25.115	62.HB2	1.826
50.HG11	0.583	55.N	127.384	59.HG2	1.526	62.HB1	1.826
50.C	172.912	55.HN	9.365	59.HG1	1.526	62.CG	27.061
51.N	126.605	55.CA	58.055	59.CD	29.556	62.HG	1.612
51.HN	8.673	55.HA	4.775	59.HD2	1.760	62.CD1	24.165
51.CA	53.370	55.CB	35.581	59.HD1	1.760	62.HD11	0.843
51.HA	5.552	55.HB	2.025	59.CE	42.110	62.CD2	24.165

62.HD21 0.843
62.C 177.768
63.N 114.045
63.HN 8.128
63.CA 58.094
63.HA 3.943
63.CB 31.857
63.HB2 1.481
63.HB1 1.594
63.CG 23.617
63.HG2 0.739
63.HG1 0.739
63.CD 29.109
63.HD2 1.149
63.HD1 1.149
63.CE 41.779
63.HE2 2.955
63.HE1 2.955
63.C 175.375
64.N 114.282
64.HN 7.695
64.CA 56.266
64.HA 4.540
64.CB 33.063
64.HB2 2.909
64.HB1 3.463
64.C 176.173
65.N 118.120
65.HN 6.867
65.CA 66.631
65.HA 3.013
65.CB 30.895
65.CG2 20.412
65.HG21 0.664
65.CG1 20.412
65.HG11 0.379
65.C 174.722
66.N 114.488
66.HN 9.134
66.CA 58.554
67.CA 65.152
67.HA 4.291
67.CB 31.973
67.HB2 2.413
67.HB1 2.413
67.CG 28.446
67.HG2 1.638
67.HG1 1.638
67.CD 51.503
67.C 176.540
68.N 111.622
68.HN 7.150
68.CA 51.370
68.HA 5.357
68.CB 41.042
68.HB2 3.418
68.HB1 3.418
68.C 174.947
69.N 122.956
69.HN 7.309
69.CA 59.986
69.HA 4.625
69.CB 39.399
69.HB 1.920
69.CG1 19.278
69.HG12 1.154
69.HG11 1.472
69.CD1 12.638
69.HD11 0.799
69.CG2 13.992
69.HG21 0.799
69.C 174.300
70.N 126.641
70.HN 9.041
70.CA 53.689
70.HA 4.831
70.CB 33.534
70.HB2 1.884
70.HB1 1.884
70.CG 26.676
70.HG2 1.524
70.HG1 1.524
70.C 174.629
71.N 119.114
71.HN 7.827
71.CA 60.295
71.HA 4.702
71.CB 39.190
71.HB 1.774
71.CG1 26.484
71.CD1 13.397
71.HD11 0.127
71.CG2 18.907
71.HG21 0.721
71.C 175.546
72.N 128.488
72.HN 8.714
72.CA 57.857
72.HA 4.280
72.CB 32.583
72.HB2 1.766
72.HB1 1.865
72.CG 24.595
72.HG2 1.449
72.HG1 1.449
72.CD 29.481
72.HD2 1.481
72.HD1 1.481
72.CE 42.228
72.HE2 2.723
72.HE1 2.994
72.C 177.730
73.N 113.667
73.HN 8.939
73.CA 46.487
73.HA2 3.869
73.HA1 4.232
73.C 173.806
74.N 120.151
74.HN 7.486
74.CA 54.119
74.HA 4.977
74.CB 45.156
74.HB2 1.597
74.HB1 1.917
74.CG 25.650
74.HG 1.148
74.C 175.851
75.N 120.130
75.HN 9.077
75.CA 58.356
75.HA 4.738
75.CB 63.858
75.HB2 3.894
75.HB1 3.894
75.C 173.066
76.N 118.650
76.HN 8.171
76.CA 55.847
76.HA 5.745
76.CB 43.078
76.HB2 2.989
76.HB1 3.143
76.C 173.489
77.N 114.863
77.HN 9.063
77.CA 56.366
77.HA 4.873
77.CB 66.144
77.HB2 3.892
77.HB1 3.892
77.C 173.132
78.N 123.959
78.HN 8.965
78.CA 62.704
78.HA 4.400
78.CB 33.447
78.HB 2.103
78.CG2 21.314
78.HG21 1.150
78.CG1 21.314
78.HG11 0.923
78.C 176.294
79.N 130.037
79.HN 9.277
79.CA 57.008
79.HA 4.466
79.CB 32.519
79.HB2 1.604
79.HB1 1.734
79.CG 24.508
79.HG2 1.308
79.HG1 1.308
79.CD 28.654
79.HD2 1.497
79.HD1 1.497
79.CE 42.308
79.C 175.902
80.N 116.305
80.HN 7.342
80.CA 61.389
80.HA 4.334
80.CB 33.568
80.HB 2.161

80.CG2	19.702	85.CA	55.605	90.N	119.097	94.CA	57.906
80.HG21	0.944	85.HA	5.144	90.HN	9.457	94.HA	3.786
80.CG1	20.828	85.CB	35.309	90.CA	59.161	94.CB	29.851
80.HG11	0.944	85.HB2	1.406	90.HA	5.156	94.HB2	2.019
80.C	173.046	85.HB1	1.406	90.CB	41.114	94.HB1	2.164
81.N	127.253	85.CG	26.099	90.HB	1.932	94.CG	25.400
81.HN	9.012	85.CD	29.178	90.CG1	26.216	94.HG2	1.322
81.CA	58.042	85.C	174.522	90.HG12	1.040	94.HG1	1.322
81.HA	5.067	86.N	130.646	90.HG11	1.826	94.CD	29.236
81.CB	47.166	86.HN	9.452	90.CD1	14.225	94.HD2	1.651
81.HB2	2.695	86.CA	55.814	90.HD11	0.806	94.HD1	1.651
81.HB1	3.187	86.CB	31.252	90.CG2	18.673	94.CE	42.243
81.C	172.224	86.C	170.718	90.HG21	0.908	94.HE2	2.975
82.N	117.247	87.N	125.495	90.C	173.680	94.HE1	2.975
82.HN	8.353	87.HN	8.683	91.N	122.859	94.C	175.475
82.CA	44.434	87.CA	62.217	91.HN	8.241	95.N	121.333
82.HA2	4.004	87.HA	4.660	91.CA	54.946	95.HN	7.908
82.HA1	4.391	87.CB	33.526	91.HA	5.198	95.CA	54.798
82.C	171.272	87.HB	2.045	91.CB	30.660	95.HA	4.627
83.N	115.710	87.CG2	19.984	91.HB2	2.088	95.CB	35.621
83.HN	8.342	87.HG21	0.847	91.HB1	2.088	95.HB2	1.781
83.CA	59.515	87.CG1	19.984	91.CG	36.353	95.HB1	1.781
83.HA	3.988	87.HG11	0.847	91.HG2	2.088	95.CG	25.328
83.CB	30.920	87.C	173.708	91.HG1	2.088	95.HG2	1.520
83.HB2	1.978	88.N	128.727	91.C	175.670	95.HG1	1.520
83.HB1	1.978	88.HN	9.455	92.N	126.698	95.CD	28.713
83.CG	36.995	88.CA	53.290	92.HN	9.359	95.HD2	1.415
83.HG2	2.256	88.HA	5.377	92.CA	56.532	95.HD1	1.415
83.HG1	2.256	88.CB	45.619	92.HA	4.904	95.CE	42.463
83.C	179.025	88.HB2	1.542	92.CB	31.555	95.HE2	4.021
84.N	114.128	88.HB1	1.722	92.HB2	2.928	95.HE1	4.021
84.HN	8.483	88.CG	29.371	92.HB1	3.344	95.C	174.339
84.CA	54.691	88.HG	1.722	92.C	174.961	96.N	115.794
84.HA	4.981	88.CD1	26.700	93.N	126.944	96.HN	7.898
84.CB	32.326	88.HD11	0.569	93.HN	8.885	96.CA	61.575
84.HB2	1.991	88.CD2	24.848	93.CA	57.318	96.HA	5.338
84.HB1	2.217	88.HD21	0.569	93.HA	3.556	96.CB	70.215
84.CG	26.847	88.C	175.962	93.CB	27.002	96.HB	3.727
84.HG2	1.815	89.N	119.132	93.HB2	1.663	96.CG2	22.235
84.HG1	1.815	89.HN	9.046	93.HB1	1.870	96.HG21	1.115
84.CD	44.960	89.CA	55.976	93.CG	35.372	96.C	174.300
84.HD2	3.711	89.HA	5.590	93.HG2	1.417	97.N	123.713
84.HD1	3.711	89.CB	39.666	93.HG1	1.124	97.HN	9.098
84.C	176.416	89.HB2	2.810	93.C	175.719	97.CA	56.020
85.N	120.078	89.HB1	3.015	94.N	108.567	97.HA	4.854
85.HN	7.023	89.C	177.220	94.HN	8.601	97.CB	42.457

97.HB2	2.964	101.HD11	0.725	106.CB	18.347	112.HN	8.694
97.HB1	2.339	101.CD2	24.597	106.HB1	1.520	112.CA	55.448
97.C	173.232	101.HD21	1.039	106.C	181.142	112.HA	4.001
98.N	122.865	101.C	174.587	107.N	115.901	112.CB	19.357
98.HN	8.812	102.N	119.509	107.HN	9.668	112.HB1	1.678
98.CA	55.750	102.HN	8.755	107.CA	59.003	112.C	178.962
98.HA	4.660	102.CA	55.106	107.HA	4.327	113.N	119.032
98.CB	30.709	102.HA	5.338	107.CB	29.007	113.HN	8.733
98.HB2	2.103	102.CB	42.620	107.HB2	2.242	113.CA	67.002
98.HB1	2.103	102.HB2	2.168	107.HB1	2.242	113.HA	3.396
98.CG	34.989	102.HB1	2.474	107.CG	36.533	113.CB	38.369
98.HG2	1.890	102.C	176.464	107.HG2	2.435	113.HB	1.704
98.HG1	1.890	103.N	107.355	107.HG1	2.435	113.CG1	30.640
98.C	172.479	103.HN	7.796	107.C	177.691	113.HG12	1.835
99.N	131.294	103.CA	58.178	108.N	115.816	113.HG11	1.835
99.HN	9.330	103.HA	5.784	108.HN	7.713	113.CD1	14.551
99.CA	52.870	103.CB	72.461	108.CA	55.481	113.HD11	0.863
99.HA	4.970	103.HB	3.909	108.HA	4.574	113.CG2	16.558
99.CB	45.761	103.CG2	23.214	108.CB	30.705	113.HG21	0.863
99.HB2	1.735	103.HG21	1.082	108.HB2	1.551	113.C	178.380
99.HB1	1.735	103.C	174.867	108.HB1	1.551	114.N	122.168
99.CG	27.188	104.N	121.355	108.CG	35.655	114.HN	7.862
99.HG	1.909	104.HN	8.658	108.HG2	2.180	114.CA	62.299
99.CD1	23.212	104.CA	50.427	108.HG1	2.820	114.HA	4.660
99.HD11	0.915	104.HA	3.379	108.C	177.746	114.CB	40.544
99.CD2	23.212	104.CB	23.720	109.N	119.456	114.HB2	3.013
99.HD21	0.614	104.HB1	1.423	109.HN	7.609	114.HB1	3.343
99.C	172.722	104.C	174.834	109.CA	63.454	114.C	178.225
100.N	129.631	105.N	120.492	110.CA	66.668	115.N	118.089
100.HN	9.154	105.HN	8.479	110.HA	4.020	115.HN	8.312
100.CA	52.409	105.CA	53.782	110.CB	30.626	115.CA	60.093
100.HA	4.970	105.HA	5.054	110.HB2	1.425	115.HA	3.982
100.CB	42.646	105.CB	42.880	110.HB1	1.425	115.CB	32.519
100.HB2	2.239	105.HB2	1.431	110.CG	28.196	115.HB2	2.936
100.HB1	3.026	105.HB1	1.431	110.HG1	0.950	115.HB1	2.529
100.C	174.477	105.CG	25.625	110.CD	48.424	115.C	177.023
101.N	120.123	105.HG	2.044	110.C	177.125	116.N	110.073
101.HN	8.606	105.CD1	23.111	111.N	114.480	116.HN	8.735
101.CA	52.356	105.HD11	1.029	111.HN	6.461	116.CA	60.098
101.HA	5.222	105.CD2	23.111	111.CA	59.500	116.CB	37.814
101.CB	45.194	105.HD21	0.894	111.HA	3.905	116.HB2	2.593
101.HB2	1.702	105.C	179.632	111.CB	36.416	116.HB1	2.897
101.HB1	1.702	106.N	124.253	111.HB2	2.924	116.C	177.622
101.CG	27.252	106.HN	8.058	111.HB1	3.479	117.N	117.223
101.HG	1.702	106.CA	56.843	111.C	179.462	117.HN	7.698
101.CD1	24.597	106.HA	3.869	112.N	124.558	117.CA	67.375

117.HA	4.366	123.HA	3.586	126.CD1	8.778	130.C	179.936
117.CB	68.065	123.CB	40.818	126.HD11	-0.222	131.N	121.973
117.HB	4.040	123.HB2	1.705	126.CG2	17.414	131.HN	8.456
117.CG2	20.821	123.HB1	1.705	126.HG21	-0.222	131.CA	59.529
117.HG21	1.115	123.CG	25.913	126.C	178.420	131.HA	4.335
117.C	173.732	123.HG	1.156	127.N	119.141	131.CB	32.392
118.N	106.975	123.CD1	20.903	127.HN	8.876	131.HB2	2.067
118.HN	6.949	123.HD11	0.525	127.CA	60.403	131.HB1	2.165
118.CA	44.456	123.CD2	20.903	127.HA	3.977	131.CG	25.041
119.CA	63.162	123.HD21	0.401	127.CB	30.492	131.HG2	1.634
119.HA	4.725	123.C	179.518	127.HB2	2.080	131.HG1	1.634
119.CB	34.992	124.N	117.914	127.HB1	2.001	131.CD	29.410
119.HB2	2.200	124.HN	8.505	127.CG	28.905	131.HD2	1.818
119.HB1	2.452	124.CA	65.773	127.HG2	1.884	131.HD1	1.818
119.CG	26.423	124.HA	3.571	127.HG1	1.884	131.CE	42.298
119.HG2	2.200	124.CB	38.114	127.CD	44.384	131.C	180.277
119.HG1	2.200	124.HB	1.779	127.HD2	3.082	132.N	118.086
119.CD	50.349	124.CG1	29.563	127.HD1	3.082	132.HN	7.397
119.C	177.395	124.CD1	13.589	127.C	179.737	132.CA	58.955
120.N	122.455	124.HD11	0.767	128.N	122.402	132.HA	4.224
120.HN	9.501	124.CG2	16.904	128.HN	8.039	132.CB	32.468
120.CA	67.589	124.HG21	0.887	128.CA	55.266	132.HB2	2.036
120.HA	3.086	124.C	178.748	128.HA	4.253	132.HB1	2.036
120.CB	31.258	125.N	120.976	128.CB	17.968	132.CG	25.479
120.HB	1.681	125.HN	7.149	128.HB1	1.584	132.HG2	1.618
120.CG2	20.519	125.CA	59.672	128.C	180.328	132.HG1	1.618
120.HG21	0.466	125.HA	4.162	129.N	121.438	132.CD	29.254
120.CG1	21.905	125.CB	30.171	129.HN	7.595	132.HD2	1.812
120.HG11	0.774	125.HB2	1.933	129.CA	55.030	132.HD1	1.812
120.C	176.804	125.HB1	1.933	129.HA	4.335	132.CE	42.336
121.N	113.382	125.CG	27.711	129.CB	18.461	132.HE2	3.110
121.HN	8.467	125.HG2	1.609	129.HB1	1.811	132.HE1	3.110
121.CA	61.391	125.HG1	1.792	129.C	181.607	132.C	177.462
121.HA	3.870	125.CD	43.590	130.N	118.831	133.N	116.592
121.C	176.003	125.HD2	3.341	130.HN	8.477	133.HN	7.432
122.N	125.029	125.HD1	3.341	130.CA	57.588	133.CA	55.223
122.HN	6.492	125.C	179.574	130.HA	4.275	133.HA	4.484
122.CA	59.430	126.N	122.504	130.CB	41.513	133.CB	33.365
122.HA	4.234	126.HN	7.921	130.HB2	2.031	133.HB2	1.758
122.CB	39.640	126.CA	62.671	130.HB1	2.031	133.HB1	2.080
122.HB2	2.676	126.HA	3.468	130.CG	28.243	133.CG	25.236
122.HB1	2.486	126.CB	36.319	130.HG	1.696	133.HG2	1.758
122.C	175.941	126.HB	1.754	130.CD1	23.034	133.HG1	1.758
123.N	117.074	126.CG1	26.484	130.HD11	0.985	133.CD	29.734
123.HN	7.474	126.HG12	0.535	130.CD2	23.034	133.HD2	1.491
123.CA	57.511	126.HG11	0.535	130.HD21	0.985	133.HD1	1.491

133.CE 42.216
133.HE2 3.013
133.HE1 3.013
133.C 174.470
134.N 112.340
134.HN 8.027
134.CA 54.724
134.HA 4.408
134.CB 36.964
134.HB2 2.807
134.HB1 3.156
134.C 173.840
135.N 114.974
135.HN 7.821
135.CA 56.078
135.HA 5.764
135.CB 42.655
135.HB2 2.933
135.HB1 2.933
135.C 176.357
136.N 122.212
136.HN 9.386
136.CA 55.288
136.HA 4.702
136.CB 35.941
136.HB2 1.813
136.HB1 1.813
136.CG 25.132
136.HG2 1.261
136.HG1 1.261
136.CD 29.236
136.HD2 1.567
136.HD1 1.567
136.CE 42.483
136.C 173.411
137.N 127.065
137.HN 9.034
137.CA 53.529
137.HA 5.454
137.CB 46.860
137.HB2 1.770
137.HB1 1.770
137.CG 27.289
137.HG 1.770
137.CD1 23.833
137.HD11 0.902
137.CD2 23.833
137.HD21 1.250
137.C 175.409
138.N 125.347
138.HN 8.931
138.CA 52.602
138.HA 4.877
138.CB 38.283
138.HB2 3.032
138.HB1 3.032
138.C 175.205
139.N 116.615
139.HN 9.421
139.CA 57.338
139.HA 3.595
139.CB 26.599
139.CG 30.661
139.HG2 0.786
139.HG1 0.786
139.C 174.087
140.N 120.339
140.HN 8.694
140.CA 57.328
140.HA 4.583
140.CB 38.309
140.HB2 3.351
140.HB1 2.970
140.C 175.252
141.N 105.852
141.HN 7.310
141.CA 44.375
141.HA2 3.176
141.HA1 3.176
141.C 177.748
142.N 119.481
142.HN 7.609
142.CA 52.937
142.HA 5.009
142.CB 45.399
142.HB2 1.948
142.HB1 1.948
142.CG 25.968
142.HG 1.089
142.CD1 24.307
142.HD11 0.900
142.CD2 24.307
142.HD21 0.560
142.C 174.062
143.N 122.686
143.HN 9.516
143.CA 56.429
143.HA 5.158
143.CB 43.313
143.HB2 2.581
143.HB1 2.801
143.C 175.169
144.N 122.734
144.HN 8.977
144.CA 55.556
144.HA 4.794
144.CB 34.710
144.HB2 1.486
144.HB1 1.768
144.CG 24.492
144.HG2 1.230
144.HG1 1.230
144.CD 29.152
144.HD2 1.486
144.HD1 1.486
144.C 177.045
145.N 127.183
145.HN 9.424
145.CA 55.482
145.HA 4.260
145.CB 37.103
145.HB2 2.965
145.HB1 3.093
145.C 174.921
146.N 112.167
146.HN 9.109
146.CA 57.269
146.HA 3.993
146.CB 27.587
146.HB2 2.290
146.HB1 2.290
146.CG 34.597
146.HG2 2.290
146.HG1 2.290
146.C 175.229
147.N 117.054
147.HN 8.041
147.CA 62.021
147.HA 4.669
147.CB 70.988
147.HB 4.221
147.CG2 21.374
147.HG21 1.367
147.C 173.168
148.N 130.255
148.HN 8.827
148.CA 55.940
148.HA 3.730
148.CB 42.861
148.HB2 1.389
148.HB1 1.500
148.CG 25.713
148.HG 1.134
148.CD1 23.932
148.HD11 0.635
148.CD2 23.932
148.HD21 0.436
148.C 176.780
149.N 130.637
149.HN 9.149
149.CA 59.700
149.CB 32.535
150.CA 62.417
150.HA 4.613
150.CB 30.006
150.HB2 2.108
150.HB1 2.039
150.CG 27.464
150.HG2 2.039
150.HG1 2.039
150.CD 50.772
150.HD2 4.001
150.HD1 3.733
150.C 176.089
151.N 123.772
151.HN 8.052
151.CA 54.482
151.HA 4.452
151.CB 45.392
151.HB2 1.784
151.HB1 1.784
151.CG 25.522
151.HG 1.242
151.HD11 0.880

151.CD2	23.020	155.HA	4.973	159.HA	3.724	162.HB1	2.107
151.HD21	0.880	155.CB	73.104	159.CB	41.322	162.CG	35.439
151.C	177.656	155.HB	4.764	159.CG	26.211	162.HG2	2.107
152.N	126.504	155.CG2	21.489	159.HG	0.989	162.HG1	2.359
152.HN	9.096	155.HG21	1.231	159.CD1	22.635	162.C	178.736
152.CA	55.110	155.C	174.884	159.HD11	0.631	163.N	116.261
152.HA	4.555	156.N	123.955	159.CD2	22.635	163.HN	7.904
152.CB	30.266	156.HN	9.524	159.HD21	0.359	163.CA	55.425
152.HB2	1.705	156.CA	59.490	159.C	176.413	163.HA	4.022
152.HB1	1.877	156.HA	3.809	160.N	117.959	163.CB	42.272
152.CG	24.449	156.CB	29.540	160.HN	7.919	163.HB2	1.504
152.HG2	1.412	156.HB2	2.394	160.CA	65.726	163.HB1	1.921
152.HG1	1.412	156.HB1	2.231	160.HA	3.239	163.CG	26.268
152.CD	29.063	156.CG	36.547	160.CB	38.780	163.CD1	22.530
152.HD2	1.705	156.HG2	2.597	160.HB	1.869	163.HD11	0.730
152.HD1	1.705	156.HG1	2.476	160.CG1	30.549	163.CD2	22.530
152.CE	42.315	156.C	177.640	160.HG12	1.089	163.HD21	0.631
152.HE2	3.023	157.N	119.768	160.HG11	1.561	163.C	176.642
152.HE1	3.023	157.HN	8.324	160.CD1	14.522	164.N	104.354
152.C	175.374	157.CA	60.597	160.HD11	0.958	164.HN	7.668
153.N	114.176	157.HA	3.860	160.CG2	17.936	164.CA	45.139
153.HN	7.244	157.CB	32.533	160.HG21	0.958	164.HA2	3.638
153.CA	59.115	157.HB2	1.582	160.C	177.135	164.HA1	3.908
153.HA	4.641	157.HB1	1.459	161.N	117.591	164.C	174.391
153.CB	42.781	157.CG	24.985	161.HN	7.529	165.N	118.797
153.HB	1.900	157.HG2	1.148	161.CA	59.285	165.HN	7.394
153.CG1	25.522	157.HG1	1.148	161.HA	4.290	165.CA	56.250
153.HG12	1.251	157.CD	29.363	161.CB	32.145	165.HA	3.475
153.HG11	1.251	157.HD2	0.805	161.HB2	1.706	165.CB	41.327
153.CD1	14.936	157.HD1	0.892	161.HB1	1.921	165.HB2	2.656
153.HD11	0.510	157.CE	41.238	161.CG	24.825	165.HB1	2.656
153.CG2	20.277	157.C	178.632	161.HG2	1.449	165.C	174.646
153.HG21	0.764	158.N	117.779	161.HG1	1.449	166.N	117.435
153.C	175.874	158.HN	7.680	161.CD	29.167	166.HN	8.640
154.N	111.316	158.CA	58.993	161.HD2	1.586	166.CA	62.839
154.HN	9.453	158.HA	3.903	161.HD1	1.586	166.HA	4.211
154.CA	62.545	158.CB	30.252	161.CE	41.814	166.CB	70.041
154.HA	4.621	158.HB2	2.064	161.HE2	3.365	166.HB	4.045
154.CB	71.276	158.HB1	2.064	161.HE1	3.365	166.CG2	21.682
154.HB	4.347	158.CG	37.097	161.C	181.123	166.HG21	1.268
154.CG2	21.849	158.HG2	2.310	162.N	121.794	166.C	174.268
154.HG21	1.231	158.HG1	2.310	162.HN	8.351	167.N	128.497
154.C	174.919	158.C	179.102	162.CA	58.551	167.HN	9.080
155.N	111.420	159.N	120.250	162.HA	4.023	167.CA	61.016
155.HN	7.756	159.HN	7.688	162.CB	28.821	167.HA	4.105
155.CA	59.251	159.CA	58.606	162.HB2	2.012	167.CB	38.039

167.HB2	2.884	169.CB	35.466	171.CD	29.821	172.HE1	2.441
167.HB1	3.093	170.CA	67.117	171.HD2	1.413	172.C	175.701
167.C	174.041	170.HA	3.451	171.HD1	1.200	173.N	121.551
168.N	126.913	170.CB	32.121	171.CE	42.265	173.HN	6.87
168.HN	7.296	170.HB2	2.131	171.C	176.136	173.CA	57.310
168.CA	52.590	170.HB1	1.700	172.N	116.887	173.HA	3.471
168.HA	4.611	170.CG	23.200	172.HN	7.878	173.CB	29.738
168.CB	32.999	170.HG2	0.798	172.CA	54.664	173.HB2	0.319
168.HB2	1.102	170.HG1	0.798	172.HA	4.366	173.HB1	0.060
168.HB1	1.542	170.C	177.500	172.CB	32.893	173.CG	27.550
168.CG	26.954	171.N	113.361	172.HB2	1.530	173.HG2	0.563
168.HG2	1.421	171.HN	8.499	172.HB1	1.710	173.HG1	0.563
168.HG1	1.421	171.CA	58.112	172.CG	25.640	173.CD	43.784
168.CD	43.663	171.HA	4.353	172.HG2	1.172	173.HD2	1.311
168.HD2	3.283	171.CB	32.948	172.HG1	1.172	173.HD1	1.311
168.HD1	3.283	171.HB2	1.468	172.CD	28.534	173.C	172.982
168.C	175.721	171.HB1	2.022	172.HD2	1.378	174.N	128.577
169.N	120.418	171.CG	23.985	172.HD1	1.378	174.HN	7.425
169.HN	8.401	171.HG2	1.671	172.CE	42.650	174.CA	57.326
169.CA	61.035	171.HG1	1.671	172.HE2	2.070	174.CB	45.286

REFERENCES

1. Human Genome Sequencing, C. (2004) Finishing the euchromatic sequence of the human genome. *Nature 431*, 931.

2. Pennisi, E. (2007) GENOMICS: DNA Study Forces Rethink of What It Means to Be a Gene. *Science 316*, 1556-1557.

3. Kolodner, R. D., Putnam, C. D., and Myung, K. (2002) Maintenance of genome stability in Saccharomyces cerevisiae. *Science 297*, 552-557.

4. Lindahl, T. (1993) Instability and Decay of the Primary Structure of DNA. *Nature 362*, 709-715.

5. Ramadan, K., Shevelev, I., and Hubscher, U. (2004) The DNA-polymerase-X family: controllers of DNA quality? *Nat Rev Mol Cell Biol 5*, 1038.

6. De Bont, R., and van Larebeke, N. (2004) Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis 19*, 169-185.

7. Hansen, W. K., and Kelley, M. R. (2000) Review Of Mammalian DNA Repair And Translational Implications. *J Pharmacol Exp Ther 295*, 1-9.

8. Hoeijmakers, J. H. J. (2001) Genome Maintenance Mechanisms For Preventing Cancer. *Nature 411*, 366.

9. Oliveros, M., Yanez, R. J., Salas, M. L., Salas, J., Vinuela, E., and Blanco, L. (1997) Characterization Of An African Swine Fever Virus 20-kDa DNA Polymerase Involved In DNA Repair. *J Biol Chem 272*, 30899-910.

10. Prasad, R., Beard, W. A., and Wilson, S. H. (1994) Studies Of Gapped DNA Substrate Binding By Mammalian DNA Polymerase β. Dependence on 5'-phosphate group. *J. Biol. Chem. 269*, 18096-18101.

11. Prasad, R., Widen, S. G., Singhal, R. K., Watkins, J., Prakash, L., and Wilson, S. H. (1993) Yeast open reading frame YCR14C encodes a DNA β-polymerase-like enzyme. *Nucl. Acids Res. 21*, 5301-5307.

12. Wagner, J., and Nohmi, T. (2000) Escherichia Coli DNA Polymerase IV Mutator Activity: Genetic Requirements And Mutational Specificity. *J Bacteriol 182*, 4587-95.

13. Bernad, A., Blanco, L., Lazaro, J. M., Martin, G., and Salas, M. (1989) A Conserved 3'----5' Exonuclease Active Site In Prokaryotic And Eukaryotic DNA Polymerases. *Cell 59*, 219-28.

14. Kornberg, A., and Baker, T. A. (1992) *DNA replication*, 2nd / ed., W.H. Freeman, New York.

15. Beard, W. A., and Wilson, S. H. (2000) Structural Design Of A Eukaryotic DNA Repair Polymerase: DNA Polymerase β. *Mutation Research/DNA Repair 460*, 231.

16. Hubscher, U., Maga, G., and Spadari, S. (2002) Eukaryotic DNA Polymerases. *Annual Review of Biochemistry 71*, 133-163.

17. Bebenek, K., Kunkel, T. A., and Wei, Y. (2004) Functions of DNA Polymerases, in *Advances in Protein Chemistry* pp 137, Academic Press.

18. Burgers, P. M. (1998) Eukaryotic DNA polymerases in DNA replication and DNA repair. *Chromosoma 107*, 218-27.

19. Hubscher, U., Nasheuer, H.-P., and Syvaoja, J. E. (2000) Eukaryotic DNA polymerases, a growing family. *Trends in Biochemical Sciences 25*, 143.

20. Idriss, H. T., Al-Assar, O., and Wilson, S. H. (2002) DNA polymerase β. *Int J Biochem Cell Biol 34*, 321-4.

21. Wilson, S. H. (1998) Mammalian Base Excision Repair And DNA Polymerase β. *Mutat Res 407*, 203-15.

22. Moon, a. F., Garcia-Diaz, M., Bebenek, K., Davis, B. J., Zhong, X. J., Ramsden, D. A., Kunkel, T. A., and Pedersen, L. C. (2007) Structural insight into the substrate specificity of DNA Polymerase mu. *Nature Structural & Molecular Biology 14*, 45-53.

23. Woodgate, R. (1999) A Plethora Of Lesion-Replicating DNA Polymerases. *Genes Dev. 13*, 2191-2195.

24. Fuchs, R. P., Fujii, S., and Wagner, J. (2004) Properties And Functions Of Escherichia Coli: Pol IV and Pol V. *Adv Protein Chem 69*, 229-64.

25. Ohmori, H., Friedberg, E. C., Fuchs, R. P. P., Goodman, M. F., Hanaoka, F., Hinkle, D., Kunkel, T. A., Lawrence, C. W., Livneh, Z., Nohmi, T., Prakash, L., Prakash, S., Todo, T., Walker, G. C., Wang, Z., and Woodgate, R. (2001) The Y-Family of DNA Polymerases. *Molecular Cell 8*, 7.

26. Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A., and Steitz, T. A. (1992) Crystal Structure At 3.5 A Resolution Of HIV-1 Reverse Transcriptase Complexed With An Inhibitor. *Science 256*, 1783-90.

27. Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J., and Pelletier, H. (1997) Crystal Structures Of Human DNA polymerase β Complexed With Gapped And Nicked DNA: Evidence For An Induced Fit Mechanism. *Biochemistry 36*, 11205-15.

28. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature 391*, 251.

29. Steitz, T. A. (1999) DNA Polymerases: Structural Diversity And Common Mechanisms. *J Biol Chem 274*, 17395-8.

30. Beese, L. S., and Steitz, T. A. (1991) Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *Embo J 10*, 25-33.

31. Brautigam, C. A., and Steitz, T. A. (1998) Structural principles for the inhibition of the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I by phosphorothioates. *J Mol Biol 277*, 363-77.

32. Polesky, A. H., Dahlberg, M. E., Benkovic, S. J., Grindley, N. D., and Joyce, C. M. (1992) Side Chains Involved In Catalysis Of The Polymerase Reaction Of DNA Polymerase I From Escherichia Coli. *J. Biol. Chem. 267*, 8417-8428.

33. Steitz, T. A. (1993) DNA- and RNA-dependent DNA polymerases. *Current Opinion in Structural Biology 3*, 31.

34. Eom, S. H., Wang, J., and Steitz, T. A. (1996) Structure Of Taq Polymerase With DNA At The Polymerase Active Site. *Nature 382*, 278-81.

35. Beese, L. S., Derbyshire, V., and Steitz, T. A. (1993) Structure Of DNA Polymerase I Klenow Fragment Bound To Duplex DNA. *Science 260*, 352-5.

36. 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*, 1222-6.

37. Sousa, R., Rose, J., and Wang, B. C. (1994) The Thumb's Knuckle. Flexibility In The Thumb Subdomain Of T7 RNA Polymerase Is Revealed By The Structure Of A Chimeric T7/T3 RNA Polymerase. *J Mol Biol 244*, 6-12.

38. Pelletier, H. (1994) Polymerase Structures And Mechanism. *Science 266*, 2025-6.

39. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H., and Kraut, J. (1994) Structures Of Ternary Complexes Of Rat DNA Polymerase β, A DNA Template-Primer, And ddCTP. *Science 264*, 1891-903.

40. Pelletier, H., Sawaya, M. R., Wolfle, W., Wilson, S. H., and Kraut, J. (1996) Crystal Structures Of Human DNA Polymerase β Complexed With DNA: Implications For Catalytic Mechanism, Processivity, And Fidelity. *Biochemistry 35*, 12742-61.

41. Sawaya, M. R., Pelletier, H., Kumar, A., Wilson, S. H., and Kraut, J. (1994) Crystal Structure Of Rat DNA Polymerase β: Evidence For A Common Polymerase Mechanism. *Science 264*, 1930-5.

42. Johnson, K. A. (1993) Conformational Coupling In DNA Polymerase Fidelity. *Annu Rev Biochem 62*, 685-713.

43. Beard, W. A., Osheroff, W. P., Prasad, R., Sawaya, M. R., Jaju, M., Wood, T. G., Kraut, J., Kunkel, T. A., and Wilson, S. H. (1996) Enzyme-DNA Interactions

Required For Efficient Nucleotide Incorporation And Discrimination In Human DNA Polymerase β. *J Biol Chem 271*, 12141-4.

44. Batra, V. K., Beard, W. A., Shock, D. D., Krahn, J. M., Pedersen, L. C., and Wilson, S. H. (2006) Magnesium-Induced Assembly Of A Complete DNA Polymerase Catalytic Complex. *Structure 14*, 757-66.

45. Lin, P., Pedersen, L. C., Batra, V. K., Beard, W. A., Wilson, S. H., and Pedersen, L. G. (2006) Energy Analysis Of Chemistry For Correct Insertion By DNA Polymerase β. *Proc Natl Acad Sci U S A 103*, 13294-9.

46. Maciejewski, M. W., Shin, R., Pan, B., Marintchev, A., Denninger, A., Mullen, M. A., Chen, K., Gryk, M. R., and Mullen, G. P. (2001) Solution Structure Of A Viral DNA Repair Polymerase. *Nat Struct Biol 8*, 936-41.

47. Showalter, A. K., Byeon, I. J., Su, M. I., and Tsai, M. D. (2001) Solution Structure Of A Viral DNA Polymerase X And Evidence For A Mutagenic Function. *Nat Struct Biol 8*, 942-6.

48. Boinas, F. S., Hutchings, G. H., Dixon, L. K., and Wilkinson, P. J. (2004) Characterization Of Pathogenic And Non-Pathogenic African Swine Fever Virus Isolates From Ornithodoros Erraticus Inhabiting Pig Premises In Portugal. *Journal of General Virology 85*, 2177-2187.

49. Vinuela, E. (1985) African Swine Fever Virus. *Current Topics in Microbiology and Immunology 116*, 151-170.

50. DEFRA. (current), The British Department for Environment, Food & Rural Affairs (DEFRA).

51. Cobbold, C., Windsor, M., Parsley, J., Baldwin, B., and Wileman, T. (2007) Reduced Redox Potential Of The Cytosol Is Important For African Swine Fever Virus Capsid Assembly And Maturation. *J Gen Virol 88*, 77-85.

52. Andres, G., Simon-Mateo, C., and Vinuela, E. (1997) Assembly Of African Swine Fever Virus: Role Of Polyprotein pp220. *J Virol 71*, 2331-41.

53. Cobbold, C., Whittle, J. T., and Wileman, T. (1996) Involvement Of The Endoplasmic Reticulum In The Assembly And Envelopment Of African Swine Fever Virus. *J Virol 70*, 8382-90.

54. Breese, S. S., and Deboer, C. J. (1966) Electron Microscope Observations of African Swine Fever Virus in Tissue Culture Cells. *Virology 28*, 420-&.

55. Brookes, S. M., Hyatt, A. D., Wise, T., and Parkhouse, R. M. E. (1998) Intracellular Virus DNA Distribution and the Acquisition of the Nucleoprotein Core during African Swine Fever Virus Particle Assembly: Ultrastructuralin SituHybridisation and DNase-Gold Labelling. *Virology 249*, 175.

56. Yanez, R. J., Rodriguez, J. M., Nogal, M. L., Yuste, L., Enriquez, C., Rodriguez, J. F., and Vinuela, E. (1995) Analysis Of The Complete Nucleotide Sequence Of African Swine Fever Virus. *Virology 208*, 249.

57. Carrascosa, A. L., del Val, M., Santaren, J. F., and Vinuela, E. (1985) Purification And Properties Of African Swine Fever Virus. *J. Virol. 54*, 337-344.

58. Rodriguez, J. M., Yanez, R. J., Rodriguez, J. F., Vinuela, E., and Salas, M. L. (1993) The DNA Polymerase-Encoding Gene Of African Swine Fever Virus: Sequence And Transcriptional Analysis. *Gene 136*, 103-10.

59. Lamarche, B. J., Kumar, S., and Tsai, M. D. (2006) ASFV DNA Polymerase X Is Extremely Error-Prone Under Diverse Assay Conditions And Within Multiple DNA Sequence Contexts. *Biochemistry 45*, 14826-33.

60. Prasad, R., Beard, W. A., Strauss, P. R., and Wilson, S. H. (1998) Human DNA Polymerase β Deoxyribose Phosphate Lyase. Substrate Specificity And Catalytic Mechanism. *J Biol Chem 273*, 15263-70.

61. Showalter, A. K., and Tsai, M. D. (2001) A DNA Polymerase With Specificity For Five Base Pairs. *J Am Chem Soc 123*, 1776-7.

62. Matsuda, T., Bebenek, K., Masutani, C., Hanaoka, F., and Kunkel, T. A. (2000) Low Fidelity DNA Synthesis By Human DNA Polymerase-β. *Nature 404*, 1011-3.

63. Garcia-Escudero, R., Garcia-Diaz, M., Salas, M. L., Blanco, L., and Salas, J. (2003) DNA Polymerase X Of African Swine Fever Virus: Insertion Fidelity On

Gapped DNA Substrates And AP Lyase Activity Support A Role In Base Excision Repair Of Viral DNA. *J Mol Biol 326*, 1403-12.

64. Showalter, A. K., and Tsai, M. D. (2002) A reexamination of the nucleotide incorporation fidelity of DNA polymerases. *Biochemistry 41*, 10571-6.

65. Dixon, L. K., and Wilkinson, P. J. (1988) Genetic diversity of African swine fever virus isolates from soft ticks (Ornithodoros moubata) inhabiting warthog burrows in Zambia. *J Gen Virol 69 (Pt 12)*, 2981-93.

66. Garcia-Barreno, B., Sanz, A., Nogal, M. L., Vinuela, E., and Enjuanes, L. (1986) Monoclonal Antibodies Of African Swine Fever Virus: Antigenic Differences Among Field Virus Isolates And Viruses Passaged In Cell Culture. *J Virol 58*, 385-92.

67. Yang, L., Beard, W. A., Wilson, S. H., Broyde, S., and Schlick, T. (2004) Highly Organized but Pliant Active Site of DNA Polymerase β: Compensatory Mechanisms in Mutant Enzymes Revealed by Dynamics Simulations and Energy Analyses. *Biophys. J. 86*, 3392-3408.

68. Vande Berg, B. J., Beard, W. A., and Wilson, S. H. (2001) DNA Structure And Aspartate 276 Influence Nucleotide Binding To Human DNA Polymerase β. Implication For The Identity Of The Rate-Limiting Conformational Change. *J Biol Chem 276*, 3408-16.

69. Prasad, R., Beard, W. A., and Wilson, S. H. (1994) Studies of gapped DNA substrate binding by mammalian DNA polymerase β. Dependence on 5'-phosphate group. *J Biol Chem 269*, 18096-101.

70. Jezewska, M. J., Marcinowicz, A., Lucius, A. L., and Bujalowski, W. (2006) DNA Polymerase X From African Swine Fever Virus: Quantitative Analysis Of The Enzyme-Ssdna Interactions And The Functional Structure Of The Complex. *J Mol Biol 356*, 121-41.

71. Jezewska, M. J., Bujalowski, P. J., and Bujalowski, W. (2007) Interactions of the DNA Polymerase X of African Swine Fever Virus with Double-stranded DNA. Functional Structure of the Complex. *Journal of Molecular Biology 373*, 75.

72. Jezewska, M. J., Bujalowski, P. J., and Bujalowski, W. (2007) Interactions of the DNA Polymerase X from African Swine Fever Virus with Gapped DNA Substrates. Quantitative Analysis of Functional Structures of the Formed Complexes. *Biochemistry*.

73. Ito, J., and Braithwaite, D. K. (1991) Compilation And Alignment Of DNA Polymerase Sequences. *Nucleic Acids Res 19*, 4045-57.

74. Kroll, D. J., Abdel-Malek Abdel-Hafiz, H., Marcell, T., Simpson, S., Chen, C. Y., Gutierrez-Hartmann, A., Lustbader, J. W., and Hoeffler, J. P. (1993) A Multifunctional Prokaryotic Protein Expression System: Overproduction, Affinity Purification, And Selective Detection. *DNA Cell Biol 12*, 441-53.

75. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

76. Marley, J., Lu, M., and Bracken, C. (2001) A Method For Efficient Isotopic Labeling Of Recombinant Proteins. *Journal of Biomolecular NMR 20*, 71.

77. Cai, M., Huang, Y., Sakaguchi, K., Clore, G. M., Gronenborn, A. M., and Craigie, R. (1998) An Efficient And Cost-Effective Isotope Labeling Protocol For Proteins Expressed In Shape Escherichia Coli. *Journal of Biomolecular NMR 11*, 97.

78. Weber, D. J., Gittis, A. G., Mullen, G. P., Abeygunawardana, C., Lattman, E. E., and Mildvan, A. S. (1992) NMR Docking Of A Substrate Into The X-Ray Structure Of Staphylococcal Nuclease. *Proteins 13*, 275-287.

79. Bagby, S., Tong, K. I., Liu, D., Alattia, J. R., and Ikura, M. (1997) The Button Test: A Small Scale Method Using Microdialysis Cells For Assessing Protein Solubility At Concentrations Suitable For NMR. *J Biomol NMR 10*, 279-82.

80. Lepre, C. A., and Moore, J. M. (1998) Microdrop Screening: A Rapid Method To Optimize Solvent Conditions For NMR Spectroscopy Of Proteins. *J Biomol NMR 12*, 493-9.

81. Maciejewski, M. W., Pan, B., Shin, R., Denninger, A., and Mullen, G. P. (2001) 1H, 15N, and 13C Resonance Assignments for a 20 kDa DNA Polymerase from African Swine Fever Virus. *J Biomol NMR 21*, 177-8.

82. Weber, D. J., Gittis, A. G., Mullen, G. P., Abeygunawardana, C., Lattman, E. E., and Mildvan, A. S. (1992) NMR docking of a substrate into the X-ray structure of staphylococcal nuclease. *Proteins 13*, 275-87.

83. Pain, R. H. (2004) Determining the Fluorescence Spectrum of a Protein in *Chapter 7 Characterization of Recombinant Proteins* pp 20, John Wiley & Sons, Inc.

84. Mach, H., Middaugh, C. R., and Denslow, N. (1995) Determining the Identity and Purity of Recombinant Proteins by UV Absorption Spectroscopy, in *Chapter 7 Characterization of Recombinant Proteins* pp 21, John Wiley & Sons, Inc.

85. Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. N. (1989) Rapid Measurement Of Binding Constants And Heats Of Binding Using A New Titration Calorimeter. *Anal Biochem 179*, 131-7.

86. Doyle, M. L. (1999) Titration Microcalorimetry, in *Quantitation of Protein Interactions* pp 24, John Wiley & Sons, Inc.

87. Leavitt, S., and Freire, E. (2001) Direct Measurement Of Protein Binding Energetics By Isothermal Titration Calorimetry. *Current Opinion in Structural Biology 11*, 560.

88. Wimley, W. C., and White, S. H. (2004) Reversible Unfolding of β-Sheets in Membranes: A Calorimetric Study. *Journal of Molecular Biology 342*, 703.

89. Weber, P. C., and Salemme, F. R. (2003) Applications Of Calorimetric Methods To Drug Discovery And The Study Of Protein Interactions. *Current Opinion in Structural Biology 13*, 115.

90. Sigurskjold, B. W. (2000) Exact Analysis of Competition Ligand Binding by Displacement Isothermal Titration Calorimetry. *Analytical Biochemistry 277*, 260.

91. Goldberg, R. N., Kishore, N., and Lennen, R. M. (2002) Thermodynamic Quantities for the Ionization Reactions of Buffers. *Journal of Physical and Chemical Reference Data 31*, 231.

92. Tellinghuisen, J. (2003) A Study Of Statistical Error In Isothermal Titration Calorimetry. *Anal Biochem 321*, 79-88.

93. Cantor, C. R., Warshaw, M. M., and Shapiro, H. (1970) Oligonucleotide Interactions. 3. Circular Dichroism Studies Of The Conformation Of Deoxyoligonucleotides. *Biopolymers 9*, 1059-77.

94. Varani, G. (1995) Exceptionally Stable Nucleic Acid Hairpins. *Annu Rev Biophys Biomol Struct 24*, 379-404.

95. Woese, C. R., Winker, S., and Gutell, R. R. (1990) Architecture Of Ribosomal RNA: Constraints On The Sequence Of "tetra-loops". *Proc Natl Acad Sci U S A 87*, 8467-71.

96. Williams, H. E., Colgrave, M. L., and Searle, M. S. (2002) Drug Recognition Of A DNA Single Strand Break: Nogalamycin Intercalation Between Coaxially Stacked Hairpins. *Eur J Biochem 269*, 1726-33.

97. MicroCal, L. (2007) *VP-ITC MicroCalorimeter User's Manual*, Vol. Rev. E, MicoCal, LLC, Northampton, MA.

98. MicroCal, L. (2004) *ITC Data Analysis in Origin®, Tutorial Guide*, Vol. Version 7.0, MicoCal, LLC, Northampton, MA.

99. Hirao, I., Kawai, G., Yoshizawa, S., Nishimura, Y., Ishido, Y., Watanabe, K., and Miura, K. (1994) Most Compact Hairpin-Turn Structure Exerted By A Short DNA Fragment, d(GCGAAGC) In Solution: An Extraordinarily Stable Structure Resistant To Nucleases And Heat. *Nucleic Acids Res 22*, 576-82.

100. Gill, S. C., and von Hippel, P. H. (1989) Calculation Of Protein Extinction Coefficients From Amino Acid Sequence Data. *Anal Biochem 182*, 319-26.

101. Piotto, M., Saudek, V., and Sklenar, V. (1992) Gradient-Tailored Excitation for Single-Quantum NMR-Spectroscopy of Aqueous-Solutions. *Journal of Biomolecular NMR 2*, 661-665.

102. Sklenar, V., Piotto, M., Leppik, R., and Saudek, V. (1993) Gradient-Tailored Water Suppression for H-1-N-15 Hsqc Experiments Optimized to Retain Full Sensitivity. *Journal of Magnetic Resonance Series A 102*, 241-245.

103. Garwood, M., Nease, B., Ke, Y., Degraaf, R. A., and Merkle, H. (1995) Simultaneous Compensation for B1 Inhomogeneity and Resonance Offsets by a

Multiple-Quantum NMR Sequence Using Adiabatic Pulses. *Journal of Magnetic Resonance, Series A 112*, 272.

104. Davis, A. L., Keeler, J., Laue, E. D., and Moskau, D. (1992) Experiments For Recording Pure-Absorption Heteronuclear Correlation Spectra Using Pulsed Field Gradients. *Journal of Magnetic Resonance (1969) 98*, 207.

105. Grzesiek, S., Anglister, J., and Bax, A. (1993) Correlation of Backbone Amide and Aliphatic Side-Chain Resonances in 13C/15N-Enriched Proteins by Isotropic Mixing of 13C Magnetization. *Journal of Magnetic Resonance, Series B 101*, 114.

106. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A Multidimensional Spectral Processing System Based On UNIX Pipes. *J Biomol NMR 6*, 277-93.

107. Johnson, B. A., and Blevins, R. A. (1994) NMRView: A Computer Program For The Visualization And Analysis Of NMR Data. *Journal of Biomolecular NMR 4*, 603.

108. Kay, L. E., Torchia, D. A., and Bax, A. (1989) Backbone Dynamics Of Proteins As Studied By 15N Inverse Detected Heteronuclear NMR Spectroscopy: application to staphylococcal nuclease. *Biochemistry 28*, 8972-9.

109. States, D. J., Haberkorn, R. a., and Ruben, D. J. (1982) A Two-Dimensional Nuclear Overhauser Experiment with Pure Absorption Phase in 4 Quadrants. *Journal of Magnetic Resonance 48*, 286-292.

110. Rabi, I. I., Zacharias, J. R., Millman, S., and Kusch, P. (1938) A New Method of Measuring Nuclear Magnetic Moment. *Physical Review 53*, 318.

111. Bloch, F. (1946) Nuclear Induction. *Physical Review 70*, 460.

112. Purcell, E. M., Torrey, H. C., and Pound, R. V. (1946) Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review 69*, 37.

113. Ernst, R. R., and Anderson, W. a. (1965) Sensitivity Enhancement in Magnetic Resonance.2. Investigation of Intermediate Passage Conditions. *Review of Scientific Instruments 36*, 1696-&.

114. Jenner, J. (1971) in *Ampere International Summer School*, Basko Poljie, Yugoslavia.

115. Aue, W. P., Bartholdi, E., and Ernst, R. R. (1976) 2-Dimensional Spectroscopy - Application to Nuclear Magnetic-Resonance. *Journal of Chemical Physics 64*, 2229-2246.

116. Griesinger, C., Sorensen, O. W., and Ernst, R. R. (1987) A Practical Approach To Three-Dimensional NMR Spectroscopy. *Journal of Magnetic Resonance (1969) 73*, 574.

117. Wuthrich, K., Wider, G., Wagner, G., and Braun, W. (1982) Sequential Resonance Assignments As A Basis For Determination Of Spatial Protein Structures By High Resolution Proton Nuclear Magnetic Resonance. *Journal of Molecular Biology 155*, 311.

118. Wüthrich, K. (1986) *NMR of proteins and nucleic acids*, Wiley, New York.

119. Cavanagh, J. (1996) *Protein NMR spectroscopy: principles and practice*, Academic Press, San Diego.

120. Claridge, T. D. W. (1999) *High-resolution NMR techniques in organic chemistry*, 1st ed., Pergamon, Amsterdam; New York.

121. Derome, A. E. (1987) *Modern NMR techniques for chemistry research*, 1st ed., Pergamon Press, Oxford [Oxfordshire]; New York.

122. Evans, J. N. S. (1995) *Biomolecular NMR spectroscopy*, Oxford University Press, Oxford; New York.

123. Rule, G. S., and Hitchens, K. T. (2006) *Fundamentals of Protein NMR Spectroscopy*, Vol. 5, 1 ed., Springer, Dordrecht.

124. Teng, Q. (2005) *Structural biology: practical NMR applications*, Springer, New York.

125. Overhauser, A. W. (1953) Paramagnetic Relaxation in Metals. *Physical Review 89*, 689.

126. Sattler, M., Schleucher, J., and Griesinger, C. (1999) Heteronuclear Multidimensional NMR Experiments For The Structure Determination Of Proteins In Solution Employing Pulsed Field Gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy 34*, 93.

127. Kay, L. E. (1995) Field Gradient Techniques In NMR Spectroscopy. *Curr Opin Struct Biol 5*, 674-81.

128. Kay, L. E. (1995) Pulsed Field Gradient Multi-Dimensional NMR Methods For The Study Of Protein Structure And Dynamics In Solution. *Prog Biophys Mol Biol 63*, 277-99.

129. Hwang, T. L., and Shaka, A. J. (1995) Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *Journal of Magnetic Resonance, Series A 112*, 275.

130. Braunschweiler, L., and Ernst, R. R. (1983) Coherence Transfer By Isotropic Mixing: Application To Proton Correlation Spectroscopy. *Journal of Magnetic Resonance (1969) 53*, 521.

131. Bax, A., Griffey, R. H., and Hawkins, B. L. (1983) Correlation Of Proton And Nitrogen-15 Chemical Shifts By Multiple Quantum NMR. *Journal of Magnetic Resonance (1969) 55*, 301.

132. Bodenhausen, G., and Ruben, D. J. (1980) Natural Abundance Nitrogen-15 NMR By Enhanced Heteronuclear Spectroscopy. *Chemical Physics Letters 69*, 185-9.

133. Mueller, L. (1979) Sensitivity Enhanced Detection Of Weak Nuclei Using Heteronuclear Multiple Quantum Coherence. *J. Am. Chem. Soc. 101*, 4481-4484.

134. Royden, V. (1954) Measurement of the Spin and Gyromagnetic Ratio of C^{13} by the Collapse of Spin-Spin Splitting. *Physical Review 96*, 543.

135. Bax, A., and Grzesiek, S. (1993) Methodological advances in protein NMR. *Acc. Chem. Res. 26*, 131-138.

136. Griesinger, C., Sorensen, O. W., and Ernst, R. R. (1989) 3-Dimensional Fourier Spectroscopy - Application to High-Resolution NMR. *Journal of Magnetic Resonance 84*, 14-63.

137. Senkevich, T. G., White, C. L., Koonin, E. V., and Moss, B. (2000) A Viral Member Of The ERV1/ALR Protein Family Participates In A Cytoplasmic Pathway Of Disulfide Bond Formation. *Proceedings of the National Academy of Sciences 97*, 12068-12073.

138. Senkevich, T. G., White, C. L., Koonin, E. V., and Moss, B. (2002) Complete Pathway For Protein Disulfide Bond Formation Encoded By Poxviruses. *Proceedings of the National Academy of Sciences 99*, 6667-6672.

139. Forman, H. J., and Torres, M. (2001) Redox Signaling In Macrophages. *Mol Aspects Med 22*, 189-216.

140. Archer, S. J., Ikura, M., Torchia, D. A., and Bax, A. (1991) An Alternative 3D NMR Technique For Correlating Backbone 15N With Side Chain H β Resonances In Larger Proteins. *Journal of Magnetic Resonance (1969) 95*, 636.

141. Powers, R., Gronenborn, A. M., Marius Clore, G., and Bax, A. (1991) Three-Dimensional Triple-Resonance NMR Of 13C/15N-Enriched Proteins Using Constant-Time Evolution. *Journal of Magnetic Resonance (1969) 94*, 209.

142. Cavanagh, J., and Rance, M. (1990) Sensitivity Improvement in Isotropic Mixing (Tocsy) Experiments. *Journal of Magnetic Resonance 88*, 72-85.

143. KAY, L. E., Keifer, P., and Saarinen, T. (1992) Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *Journal of the American Chemical Society 114*, 10663-10665.

144. Grzesiek, S., and BAX, A. (1993) Amino-Acid Type Determination in the Sequential Assignment Procedure of Uniformly C-13/N-15-Enriched Proteins. *Journal of Biomolecular NMR 3*, 185-204.

145. Ernst, R. R., Bodenhausen, G., and Wokaun, A. (1987) *Principles of nuclear magnetic resonance in one and two dimensions*, Clarendon Press; Oxford University Press, Oxford, New York.

146. Kay, L. E., Ikura, M., Tschudin, R., and Bax, A. (1990) Three-Dimensional Triple-Resonance NMR Spectroscopy Of Isotopically Enriched Proteins. *Journal of Magnetic Resonance (1969) 89*, 496.

147. Kirby, N. I., DeRose, E. F., London, R. E., and Mueller, G. A. (2004) Nvassign: Protein NMR Spectral Assignment With NMRView. *Bioinformatics 20*, 1201-3.

148. Coggins, B. E., and Zhou, P. (2003) PACES: Protein sequential assignment by computer-assisted exhaustive search. *Journal of Biomolecular NMR 26*, 93-111.

149. BAX, A., Clore, G. M., Driscoll, P. C., Gronenborn, a. M., Ikura, M., and Kay, L. E. (1990) Practical Aspects of Proton Carbon Carbon Proton 3-Dimensional Correlation Spectroscopy of C-13-Labeled Proteins. *Journal of Magnetic Resonance 87*, 620-627.

150. KAY, L. E., Ikura, M., and Bax, a. (1990) Proton Proton Correlation Via Carbon Carbon Couplings - a 3-Dimensional NMR Approach for the Assignment of Aliphatic Resonances in Proteins Labeled with C-13. *Journal of the American Chemical Society 112*, 888-889.

151. BAX, A., Clore, G. M., and Gronenborn, a. M. (1990) H-1-H-1 Correlation Via Isotropic Mixing of C-13 Magnetization, a New 3-Dimensional Approach for Assigning H-1 and C-13 Spectra of C-13-Enriched Proteins. *Journal of Magnetic Resonance 88*, 425-431.

152. Fesik, S. W., Eaton, H. L., Olejniczak, E. T., Zuiderweg, E. R. P., Mcintosh, L. P., and Dahlquist, F. W. (1990) 2D and 3D NMR-Spectroscopy Employing C-13-C-13 Magnetization Transfer by Isotropic Mixing - Spin System-Identification in Large Proteins. *Journal of the American Chemical Society 112*, 886-888.

153. BAX, A., and Davis, D. G. (1985) Mlev-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy. *Journal of Magnetic Resonance 65*, 355-360.

154. Jouvenet, N., Windsor, M., Rietdorf, J., Hawes, P., Monaghan, P., Way, M., and Wileman, T. (2006) African Swine Fever Virus Induces Filopodia-Like Projections At The Plasma Membrane. *Cellular Microbiology 8*, 1803-1811.

155. Locker, J. K., and Griffiths, G. (1999) An Unconventional Role for Cytoplasmic Disulfide Bonds in Vaccinia Virus Proteins. *J. Cell Biol. 144*, 267-279.

156. Lewis, T., Zsak, L., Burrage, T. G., Lu, Z., Kutish, G. F., Neilan, J. G., and Rock, D. L. (2000) An African Swine Fever Virus ERV1-ALR Homologue, 9GL, Affects Virion Maturation and Viral Growth in Macrophages and Viral Virulence in Swine. *J. Virol. 74*, 1275-1285.

157. Meng, E. C., Pettersen, E. F., Couch, G. S., Huang, C. C., and Ferrin, T. E. (2006) Tools For Integrated Sequence-Structure Analysis With UCSF Chimera. *Bmc Bioinformatics 7*, -.

158. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF chimera - A Visualization System For Exploratory Research And Analysis. *Journal of Computational Chemistry 25*, 1605-1612.

159. Delarue, M., Boule, J. B., Lescar, J., Expert-Bezancon, N., Jourdan, N., Sukumar, N., Rougeon, F., and Papanicolaou, C. (2002) Crystal Structures Of A Template-Independent DNA Polymerase: Murine Terminal Deoxynucleotidyltransferase. *Embo J 21*, 427-39.

160. Garcia-Diaz, M., Bebenek, K., Krahn, J. M., Blanco, L., Kunkel, T. A., and Pedersen, L. C. (2004) A Structural Solution for the DNA Polymerase λ-Dependent Repair of DNA Gaps with Minimal Homology. *Molecular Cell 13*, 561.

161. Voehler, M. W., Collier, G., Young, J. K., Stone, M. P., and Germann, M. W. (2006) Performance Of Cryogenic Probes As A Function Of Ionic Strength And Sample Tube Geometry. *Journal of Magnetic Resonance 183*, 102.

162. Yao, J., and Yoon, J. (2000) Low-Noise Electrometer And Its Low-Noise Cryogenic Probe With Completely Guarded Sample Chamber. *Review of Scientific Instruments 71*, 1776-1780.

163. Styles, P., Soffe, N. F., Scott, C. A., Crag, D. A., Row, F., White, D. J., and White, P. C. J. (1984) A High-Resolution NMR Probe In Which The Coil And Preamplifier Are Cooled With Liquid Helium. *Journal of Magnetic Resonance (1969) 60*, 397-404.

164. Styles, P., Soffe, N. F., and Scott, C. a. (1989) An Improved Cryogenically Cooled Probe for High-Resolution NMR. *Journal of Magnetic Resonance 84*, 376-378.

165. Hill, H. D. W. (1997) Improved Sensitivity Of NMR Spectroscopy Probes By Use Of High-Temperature Superconductive Detection Coils. *IEEE Transactions on Applied Superconductivity 7*, 3750-5.

166. Withers, R. S., Liang, G. C., Cole, B. F., and Johansson, M. (1993) Thin-Film HTS Probe Coils For Magnetic-Resonance Imaging. *IEEE Transactions on Applied Superconductivity 3*, 3037-42.

167. Hajduk, P. J., Gerfin, T., Boehlen, J. M., Haberli, M., Marek, D., and Fesik, S. W. (1999) High-Throughput Nuclear Magnetic Resonance-Based Screening. *Journal of Medicinal Chemistry 42*, 2315-2317.

168. Serber, Z., Richter, C., Moskau, D., Bohlen, J. M., Gerfin, T., Marek, D., Haberli, M., Baselgia, L., Laukien, F., Stern, A. S., Hoch, J. C., and Dotsch, V. (2000) New Carbon-Detected Protein NMR Experiments Using Cryoprobes. *Journal of the American Chemical Society 122*, 3554-3555.

169. Serber, Z., Richter, C., and Dotsch, V. (2001) Carbon-Detected NMR Experiments To Investigate Structure And Dynamics Of Biological Macromolecules. *Chembiochem 2*, 247-51.

170. Triebe, R., Nast, R., Marek, D., Withers, R., Baselgia, L., Haberli, M. G., T., and Calderon, P. (1999) A User-Friendly System for the Routine Application of Cryogenic NMR Probes: Technology and Results. *40th Experimental Nuclear Magnetic Resonance Conference; Orlando, FL*, 198.

171. Flynn, P. F., Mattiello, D. L., Hill, H. D. W., and Wand, A. J. (2000) Optimal Use Of Cryogenic Probe Technology In NMR Studies Of Proteins. *Journal of the American Chemical Society 122*, 4823-4824.

172. Kelly, A. E., Ou, H. D., Withers, R., and Dotsch, V. (2002) Low-Conductivity Buffers For High-Sensitivity NMR Measurements. *J Am Chem Soc 124*, 12013-9.

173. Babu, C. R., Flynn, P. F., and Wand, A. J. (2003) Preparation, Characterization, And NMR Spectroscopy Of Encapsulated Proteins Dissolved In Low Viscosity Fluids. *J Biomol NMR 25*, 313-23.

174. Doussin, S., Birlirakis, N., Georgin, D., Taran, F., and Berthault, P. (2006) Novel Zwitterionic Reverse Micelles for Encapsulation of Proteins in Low-Viscosity Media. *Chemistry*.

175. Peterson, R. W., Lefebvre, B. G., and Wand, A. J. (2005) High-Resolution NMR Studies Of Encapsulated Proteins In Liquid Ethane. *J Am Chem Soc 127*, 10176-7.

176. Peterson, R. W., Pometun, M. S., Shi, Z., and Wand, A. J. (2005) Novel Surfactant Mixtures For NMR Spectroscopy Of Encapsulated Proteins Dissolved In Low-Viscosity Fluids. *Protein Sci 14*, 2919-21.

177. Van Horn, W. D., Simorellis, A. K., and Flynn, P. F. (2005) Low-Temperature Studies Of Encapsulated Proteins. *J Am Chem Soc 127*, 13553-60.

178. Shi, Z., Peterson, R. W., and Wand, A. J. (2005) New Reverse Micelle Surfactant Systems Optimized For High-Resolution NMR Spectroscopy Of Encapsulated Proteins. *Langmuir 21*, 10632-7.

179. Ehrhardt, M. R., Flynn, P. F., and Wand, A. J. (1999) Preparation Of Encapsulated Proteins Dissolved In Low Viscosity Fluids. *J Biomol NMR 14*, 75-8.

180. Babu, C. R., Flynn, P. F., and Wand, A. J. (2001) Validation Of Protein Structure From Preparations Of Encapsulated Proteins Dissolved In Low Viscosity Fluids. *J Am Chem Soc 123*, 2691-2.

181. Wand, A. J., Ehrhardt, M. R., and Flynn, P. F. (1998) High-Resolution NMR Of Encapsulated Proteins Dissolved In Low-Viscosity Fluids. *Proc Natl Acad Sci U S A 95*, 15299-302.

182. Wu, W. J., Vidugiris, G., Mooberry, E. S., Westler, W. M., and Markley, J. L. (2003) Mixing Apparatus For Preparing NMR Samples Under Pressure. *J Magn Reson 164*, 84-91.

183. Cohn, E. J., and Edsall, J. T. (1943) *Proteins, Amino Acids And Peptides As Ions And Dipolar Ions*, Reinhold publishing corporation, New York.

184. Edsall, J. T. (1951) *Enzymes And Enzyme Systems, Their State In Nature*, Harvard University Press, Cambridge.

185. Lane, a. N., and Arumugam, S. (2005) Improving NMR Sensitivity In Room Temperature And Cooled Probes With Dipolar Ions. *Journal of Magnetic Resonance 173*, 339-343.

186. de Swiet, T. M. (2005) Optimal Electric Fields For Different Sample Shapes In High Resolution NMR Spectroscopy. *J Magn Reson 174*, 331-4.

187. Guclu, C. C. (2001) Effect Of Body Coil Electric Field Distribution On Receive-Only Surface Coil Heating. *J Magn Reson Imaging 14*, 484-7.

188. Mansfield, P., Bowley, R. M., and Haywood, B. (2003) Controlled E-field gradient coils. *Magma 16*, 113-20.

189. Taylor, H. C., Burl, M., and Hand, J. W. (1997) Experimental verification of numerically predicted electric field distributions produced by a radiofrequency coil. *Phys Med Biol 42*, 1395-402.

190. Gadian, D. G., and Robinson, F. N. H. (1979) Radiofrequency Losses in NMR Experiments on Electrically Conducting Samples. *Journal of Magnetic Resonance 34*, 449-455.

191. Hoult, D. I., and Lauterbur, P. C. (1979) Sensitivity of the Zeugmatographic Experiment Involving Human Samples. *Journal of Magnetic Resonance 34*, 425-433.

192. Hoult, D. I., and Richards, R. E. (1976) Signal-to-Noise Ratio of Nuclear Magnetic-Resonance Experiment. *Journal of Magnetic Resonance 24*, 71-85.

193. Binbuga, B., and Young, J. K. (2005) 1H, 13C and 15N backbone and side chain resonance assignments of Haloferax volcanii DHFR1. *J Biomol NMR 33*, 281.

194. Cavanagh, J., Palmer, a. G., Wright, P. E., and Rance, M. (1991) Sensitivity Improvement in Proton-Detected 2-Dimensional Heteronuclear Relay Spectroscopy. *Journal of Magnetic Resonance 91*, 429-436.

195. Palmer, a. G., Cavanagh, J., Wright, P. E., and Rance, M. (1991) Sensitivity Improvement in Proton-Detected 2-Dimensional Heteronuclear Correlation NMR-Spectroscopy. *Journal of Magnetic Resonance 93*, 151-170.

196. Schleucher, J., Schwendinger, M., Sattler, M., Schmidt, P., Schedletzky, O., Glaser, S. J., Sorensen, O. W., and Griesinger, C. (1994) A General Enhancement Scheme in Heteronuclear Multidimensional NMR Employing Pulsed-Field Gradients. *Journal of Biomolecular NMR 4*, 301-306.

197. Beard, W. A., and Wilson, S. H. (2001) DNA polymerases lose their grip. *Nat Struct Biol 8*, 915-7.

198. Grzesiek, S., and Bax, A. (1992) Improved 3D triple-resonance NMR techniques applied to a 31 kDa protein. *Journal of Magnetic Resonance (1969) 96*, 432.