STRUCTURE OF AN ANTIGEN-BINDING FRAGMENT BOUND TO STEM-LOOP DNA AND CRYSTALLIZATION OF RECOMBINANT HAEMOPHILUS INFLUENZAE e(P4) ACID PHOSPHATASE

A Thesis presented to the Faculty of the Graduate School
University of Missouri-Columbia

In Partial Fulfillment
Of the Requirements of the Degree

Master of Science

---

by
ZHONGHUI OU

Dr. John J. Tanner, Thesis Supervisor

MAY 2006

The undersigned, appointed by the Dean of the Graduate School, have examined the thesis entitled

STRUCTURE OF AN ANTIGEN-BINDING FRAGMENT BOUND TO STEM-LOOP DNA AND CRYSTALLIZATION OF RECOMBINANT HAEMOPHILUS INFLUENZAE e(P4) ACID PHOSPHATASE

Presented by Zhonghui Ou

A candidate for the degree of Master of Science

And hereby certify that in their opinion it is worthy of acceptance.

Professor John J. Tanner

Professor Lesa J. Beamer

Professor Thomas J. Reilly

ACKNOWLEDGEMENTS

I would like to express my deep appreciation to the following people. Without anyone of them, this piece of work would not have come into being.

Firstly, I would like to thank my advisor Dr. John J. Tanner for his patient instruction on my research. From showing me how to crystallize lysozyme in my rotation to helping me prepare research manuscripts, he is the one who has led me into the beautiful world of protein crystallography.

Secondary, I would like to thank Dr. Thomas Reilly, Dr. Lesa Beamer, Dr Judy Wall and Dr. Linda Randall for their valuable advice on my research and studies.

I also extend my heartfelt thankfulness to the whole Tanner group for their support and friendship. It was them who have helped me go through all the hardships in the last three years. Specifically, I would like to thank Mr. Min Zhang, who worked with me on my first E.coli culture preparation. I would like to thank Dr. Christopher Bottoms, my kindest colleague and friend, who has taught me a lot of computational techniques and who kept helping me again and again when I was in trouble, and Mr. Jermaine Jenkins, who was always there to give me a hand whenever I had questions. The happy times I have spent with my group members will forever remain in my mind.

Most of all, I would like to thank my father Jingqi Ou, my mother Yueping Xuan , my cousin Shaomeng Xuan and my uncle Liangshen Liu for their support in my oversea studies. It is their love that always encourages me to keep going in spite of all the difficulties.

TABLES OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter 1

Remodeling of a DNA stem-loop by an antigen-binding fragment

Abstract

Anti-DNA antibodies play important roles in the pathogenesis of autoimmune diseases and they also serve as unique models for the study of protein-DNA recognition. DNA-1 and 11F8 are anti-ss DNA antibodies derived from autoimmune lupus-prone mice. They are very similar to each other in terms of CDR sequence and preference for binding T-rich ssDNA. Here, we present the 1.95 Å resolution structure of DNA-1 complexed with a stem-loop DNA ligand, denoted G5-14. G5-14 is a synthetic oligonucleotide with the ten-nucleotide sequence identical to the stem-loop portion above the bulge of G1-17, which is an oligonucleotide identified by *in vitro* selection experiments and binds with high affinity and specificity to Fab 11F8. 11F8 localizes to kidney tissue by binding to DNA adherent to the GBM and eventually leads to renal damage in a mouse model. The DNA-1/G5-14 structure shows that the two DNA strands dimerize to form a double-stranded DNA dumbbell and have a large conformational change including the breaking and reformation of hydrogen bonds. The most striking feature of the Fab/DNA interactions is the use of extensive π-π stacking of the DNA bases and the protein side chains to form base-base and base-aromatic stacking interactions. DNA-1 seems to bind to the stem loop ligand in a way different from 11F8. These results provide insights into the specific recognition model of anti-DNA Abs and the potential challenges in structure based drug design to treat autoimmune diseases.

Keywords: anti-DNA antibodies; X-ray crystallography; systemic lupus erythematosus; stem-loop DNA; single-stranded DNA, DNA dumbbell; structure based drug design

Abbreviations used

SLE: systemic lupus erythematosus

GBM: glomerular basement membrane

CDR: complementarity-determining region

ss: single stranded

ds: double stranded

Fab: antigen binding fragment

Ab: antibody

NDB: Nucleic Acid Data Base

PDB: Protein Data Bank

RMSD: root-mean-square difference

V: variable

C: conserved

H: heavy

L: light

ITC: Isothermal Titration Calorimetry

Introduction

Systemic lupus erythematosus (SLE) is an autoimmune disease with clinical manifestations of arthritis, glomerulonephritis, hemolytic anemia and thrombocytopenia.[1] The major group affected by SLE is women between 20 and 60 years old. Although the exact mechanism of SLE pathogenesis is yet to be clarified, numerous studies have suggested that its development could be divided into two stages.[2] The first stage is the binding of an anti-DNA Ab to its antigen and the succeeding deposition of excessive amounts of autoimmune complexes. The second stage, which is mediated by the Fc region of the complexed antibody, is the recruitment and activation of inflammatory cells. The current medications used for SLE treatment include nonsteroidal anti-inflammatory agents, anti-malarial agents, corticosteroids, and cytotoxic immunosuppressive agents.[3] They all target at the second stage and have many side effects due to their non-specific effects. This underscores the intriguing idea of inventing more specific drugs for SLE. Presumably, molecules that disrupt the interactions between anti-DNA antibodies and their antigens may be potential drug candidates. Understanding the molecular basis of anti-DNA·DNA interactions by protein crystallography sets a primary platform for this approach.

The focus of our present work is the crystallization and structural studies of an anti-ssDNA Fab, denoted DNA-1 complexed with a sequence specific stem-loop DNA ligand, denoted as G5-14 (Figure1-1b). Our initial motivation for this project came from the results from two different groups who studied the molecular basis of SLE. DNA-1, the antibody Fab in our structure, was initially isolated by Deutscher's group from a combinatiorial bacteriophage display library of IgG fragments derived from the immunoglobulin repertoire of an autoimmune SLE-like MRL/lpr mouse. [4] Fab

DNA-1 was found to have a marked preference for binding oligo(dT), especially with oligo(dT) of a length over 15 nucleotides. [5] Results from CDR region switch experiments[6] and additional mutation experiments [7; 8] showed that HCDR1 and HCDR3 of Fab DNA-1 are implicated in the binding. In the same decade, Glick's group isolated 11F8 from a panel of eight anti-DNA mAbs from an antoimmune MRL MpJ-lpr/lpr mouse. [9] 11F8 has high affinity and specificity for oligo(dT) in vitro [9; 10]and was found in mouse studies to localize to kidney tissue by binding to DNA adherent to the GBM and eventually leads to renal damage [11]. In vitro evolution experiments were performed and identified a seventeen-nucleotide stem-loop DNA oligo (Figure 1-1a, denoted G1-17 in this manuscript) that binds to 11F8 with high affinity and specificity.[10] The binding mechanisms of 11F8 to the selected stem loop ligand were further studied by thermodynamics and kinetics approach. [12; 13; 14] These studies suggested that Fab-DNA interactions are mediated by T10, T11, C12 of the DNA and HCDR1, HCDR3 and LCDR1 of Fab 11F8. DNA-1 and 11F8 have 95% amino acid sequence identity in their light chain variable domain and 77.5% identity in their heavy chain variable domain(Figure 1-2). However, there is not a crystal structure of 11F8 bound to the selected stem loop DNA. While it remains to be seen whether DNA-1 also displays pathogenic behavior as 11F8, it is still very tempting for us to determine the structure of DNA-1 complexed with a stem loop DNA oligo to obtain insights into the specific recognition model of anti-DNA Abs. This is our initial motivation of crystallizing DNA-1 with G5-14.

The 1.95 Å resolution structure of DNA-1/G5-14 complex shows that the two DNA molecules dimerize to form a nearly two-fold symmetric DNA dumbbell and undergo a large conformational change when bound by Fab DNA-1 (Figure 1-3, Figure 1-4). In each DNA molecule, the three Watson-Crick base pairs at the stem are completely

lost. The DNA moves to its 5' terminal to interact with the protein and forms one new Watson-Crick base pair. The most striking feature of the Fab/DNA interactions is the use of an extensive π system of the DNA bases and the protein side chains to form base-base and base-aromatic stacking interactions. This underlies the negative entropy and enthalpy in the ITC experiment and suggests that desolvation of hydrophobic side chains are the major force in the complex formation. Structural analysis showed that besides maintaining the same two thymine recognition sites as the previous structure of DNA-1, Fab DNA-1/G5-14 complex also have other sequence-specific interactions in their interface. These extra interactions may contribute to the 2 to 4 fold higher affinity of G5-14 compared to dT_5 in the binding of DNA-1. Compared to the previously determined ligand-free structure of DNA-1, [15] large conformational changes also occurred to HCDR3 and LCDR3 of the protein. In summary, the protein and the DNA both significantly change their shape upon binding. Although the dimerization of the two DNA molecules in the crystal asymmetric unit may result from crystal packing effects, the conformational change of each DNA molecule likely reflects the real solution situation as similar "mutually induced fit" cofolding processes have been also observed for stem-loop RNA/ protein complexes. [16; 17]

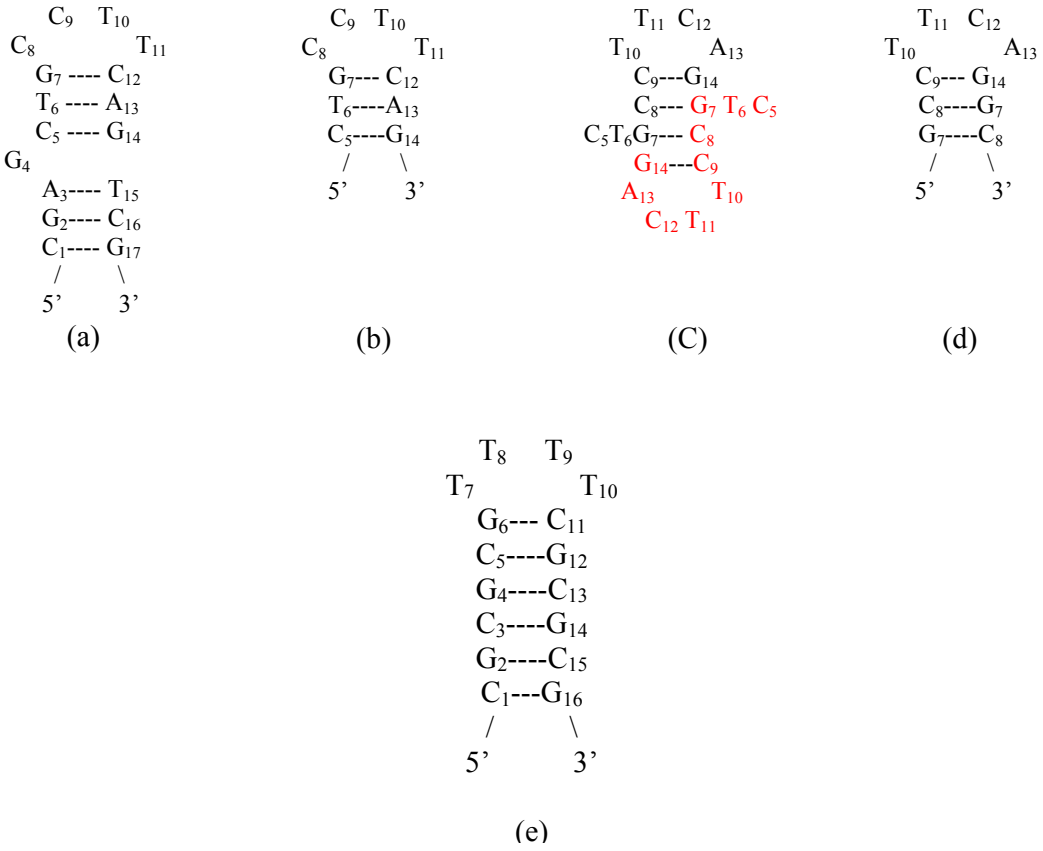


Figure 1-1. The five DNA ligands discussed in this manuscript

(a) G1-17: the 17-nucleotide stem-loop DNA ligand that binds with high affinity and specificity to Fab 11F8. (b) Anticipated G5-14 secondary structure in solution. G5-14 was used for co-crystallization with Fab DNA-1. (c) The observed secondary structure of G5-14 in the crystal structure. (d) Secondary structure of a proposed stem-loop DNA ligand that would bind to DNA-1 with high affinity and specificity. (e) Secondary structure of the stem loop DNA (PDB entry:1D16). 1D16 was used as a template for building the G5-14 ligand in our structure.

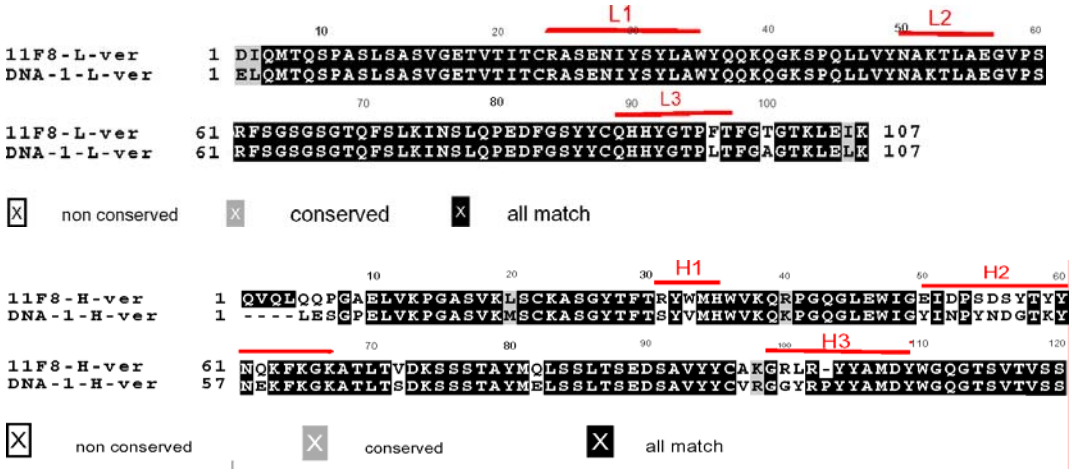


Figure 1-2 Sequence alignment of the variable domains of DNA-1 and 11F8. The CDRs are marked with horizontal bars and correspond to the following residuals: L1,24-34; L2, 50-56; L3,89-97;H1, 31-35;H2, 50-65; H3, 95-102. The alignment was performed using GENESTREAM.

Results

Description of the Overall Structure and Nomenclature

The asymmetric unit consists of two Fab molecules (Fab 1 and 2), two G5-14 DNA molecules, one PEG fragment and 400 water molecules (Figure 1-3 and Table 1-1). The light/heavy chains are denoted L/H in Fab 1 and A/B in Fab 2. The two DNA strands have chain identifiers M and N. The residue numbering scheme and definition of complementarity-determining region (CDR) follow the standard Kabat conventions. [18; 19] The R-factor for the refined model is 0.207, with an R-free[20] (5% test set) of 0.256. A Luzzati plot suggests a mean positional error of 0.180 Å, [21] while the σ_A error estimate is 0.143 Å .[22] The model exhibits good stereochemical quality that is consistent with 1.95 Å resolution data (Table 1-1), and it meets or exceeds all main-chain and side-chain tests of PROCHECK. [23]

As expected, both Fabs display the well-characterized immunoglobulin fold[24] in which each of the four homology subunits, denoted V_H, V_L, C_{H1} and $C_{L,}$ consists of two twisted, antiparallel ß sheets packed tightly against each other (Figure 1-3). The resulting electron density maps were continuous throughout both Fab molecules, except for residues 127-132 of heavy chain H and 128-133 of heavy chain B. These areas are disordered in many Fab structures. Residues L1, L214, A214, H214-223, and B214-223 were omitted from the model due to the unclear electron density at the N- and C- terminal of the two Fabs. Side chains from 26 surface residues, most of which are Asp, Glu, Asn, Gln, Lys and Arg, were truncated due to weak electron density.

DNA strand N primarily interacts with Fab1, while strand M interacts with Fab 2. Superimposition of Fab1/N with Fab2/M yields a root-mean-square difference

(RMSD) of 1.1 Å for the main chain and 1.2 Å for all atoms. The RMSD between the two Fab molecules alone after superimposition is 0.7 Å for the main chain and 1.0 Å for all atoms. If the CDRs are excluded, the RMSD is 0.75 Å for the main chain and 1.1 Å for all atoms. The RMSD for the V_H/V_L superdomain is 0.4 Å for the main chain and 0.8 Å for all atoms, while the analogous values for the C_{H1}/C_L superdomain are 0.55 Å for the main chain and 1.0 Å for all atoms. As far as the local differences between the two Fabs are concerned, the conformations of all six CDRs are similar in the two Fabs. The RMSDs for these CDRs after superimposition are less than 0.3 Å for the main chain and less than 1.1 Å for all atoms.

Compared with the two Fabs RMSDs in the DNA-1/dT_5 structure [25] , the two Fabs in the DNA-1/G5-14 structure appear to adopt a much more similar conformation. The V_H/V_L domains are highly ordered with an RMSD even smaller than that of C_{H1}/C_L.

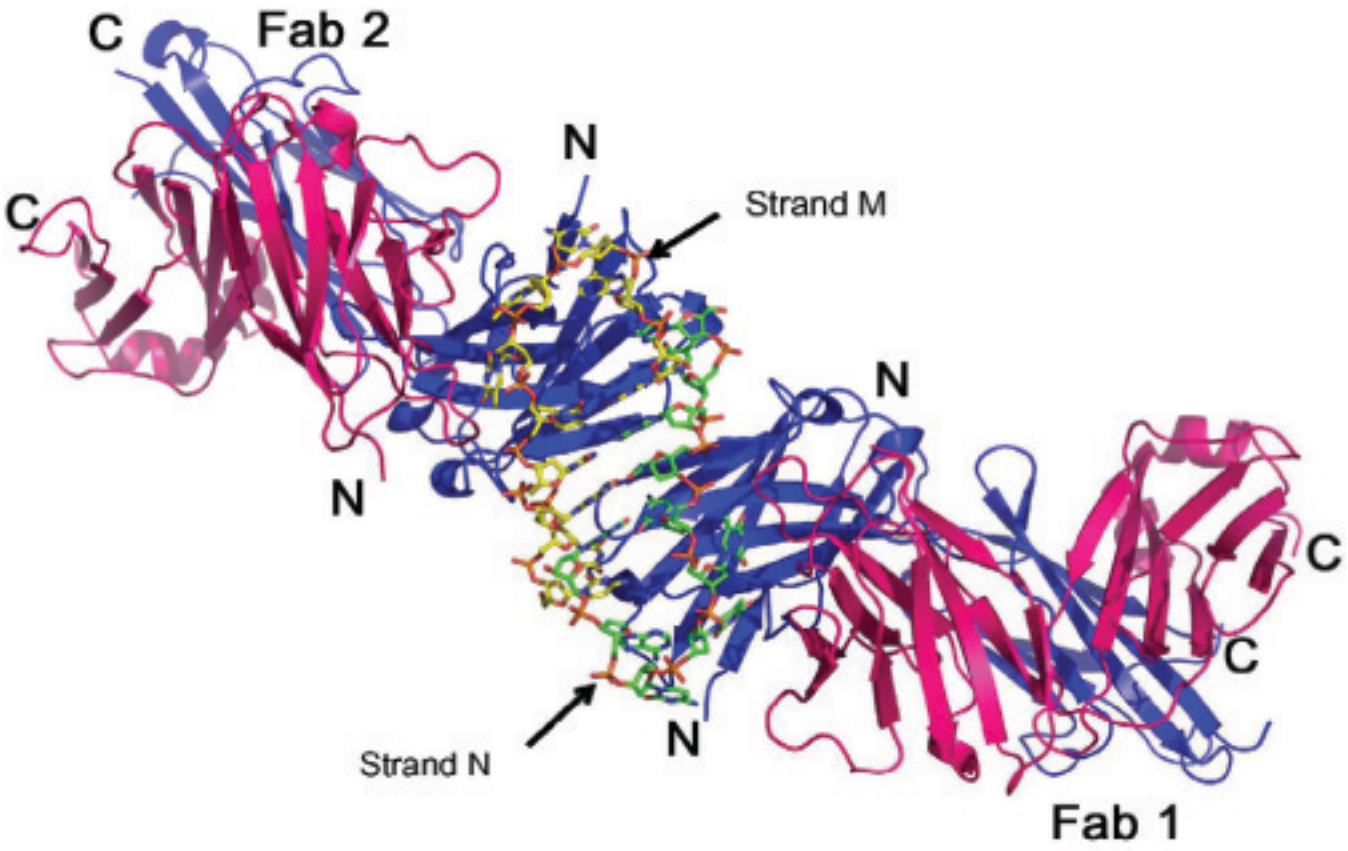
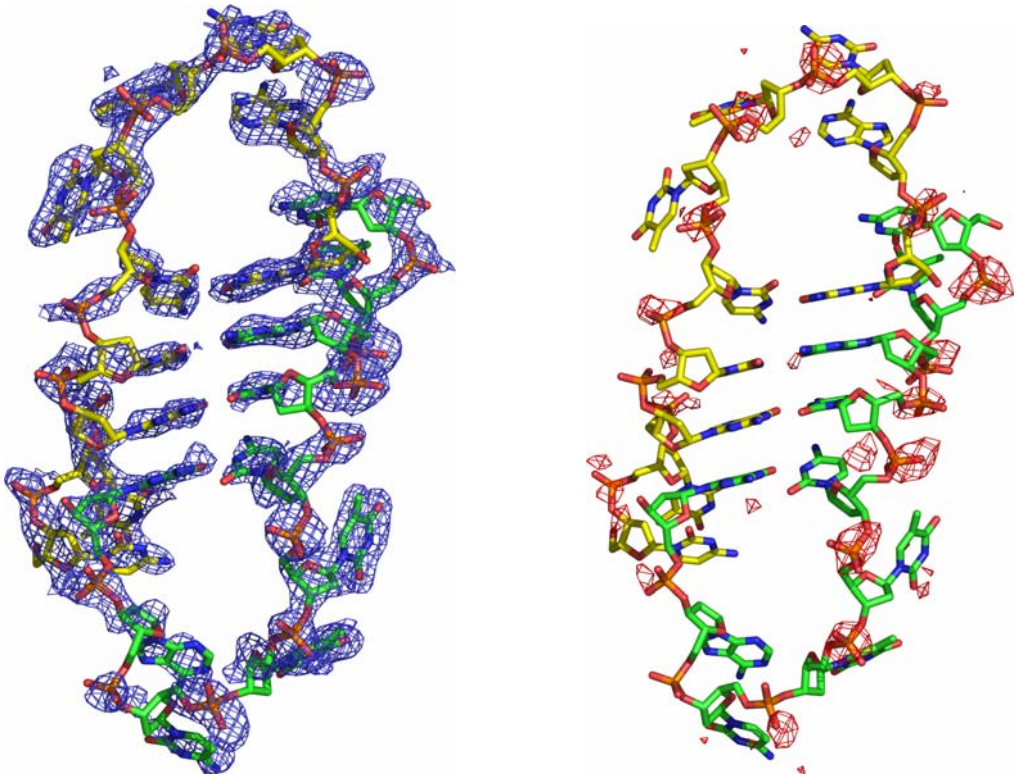


Figure 1-3. Ribbon drawing of the asymmetric unit of the DNA-1/G5-14 structure. The light chains are colored pink and the heavy chains are colored blue. The DNA appears in ball-and-stick mode. Strand M carbon atoms are colored yellow. Strand N carbon atoms are colored green. This figure was prepared with PYMOL [26].

G5-14 conformation

The anticipated secondary structure of G5-14 in solution is a hairpin with three Watson-Crick base pairs(C_5-G_{14}, T_6-A_{13} and G_7-C_{12}) at the stem and four bases at the loop (Figure 1-1b). However, in the DNA-1/G5-14 crystal structure, the DNA molecule adopts a completely different conformation. Two DNA molecules dimerize to form a double-stranded DNA dumbbell. This dumbbell structure has nearly perfect two-fold symmetry (Figure 1-4). The RMSD between the two DNA strands is 0.9 Å for all atoms, which indicates that the two strands have virtually identical conformations. The three base pairs at the stem of G5-14's secondary structure were completely lost and were replaced by one new base pair (C_9-G_{14}) in each DNA molecule in the crystal structure. Four G-C base pairs stack upon each other, forming an extensive π-π system. The four bases (G_{14}, G_7, C_8, C_9) on each side of the stem are approximately coplanar (Figure 1-4). C_5 and T_6 flip out away from the stem as two arms, with their coplanar bases stacking upon each other. The loops are formed by T_{10}, T_{11}, C_{12}, and A_{13}, indicating the moving of G5-14 towards its 5'-terminal. These nucleotides swing out of the DNA loop and interact with the protein and solvent. T_{10} and T_{11} point to different orientations while C_{12} and A_{13} form coplanar stacking interactions with C_5 and T_6 from the other DNA strand. The individual nucleotides, C_5, T_6, T_{10}, and T_{11} have relatively lower *B*-factor, an average of 32.4 $Å^2$ while *B*-factors for G_7, C_8, C_9, C_{12}, A_{13} and G_{14} are relatively high with a 40.4 $Å^2$ average. The nucleotides with lower *B*-factor all have extensive interactions with the Fabs and the corresponding electron density is very strong. In comparison, G_7, C_8, C_9 and G_{14}, which form the four base pairs in the stem, have relatively fewer interactions with the

protein. The density for A13 and G14 was poor at the beginning of the structure refinement, reflecting the flexibility of these two nucleotides.



(a) **(b)**

Figure 1-4 The final, refined model of the G5-14 ligands bound to DNA-1 in the asymmetric unit. This figure was prepared with PYMOL[26].(a) The map colored in blue is a σA-weighted, simulated annealing, Fo-Fc omit electron density map (2σ) generated by a simulated annealing procedure (1000 K) in which both DNA strands had been removed in the final model. (b)The maps colored in red are the 2.5 σ anomalous difference Fourier maps calculated from the model phases and anomalous differences from the low energy data sets. These maps were used for locating DNA phosphate atoms in the model building.

Table 1-1. Data Collection and Refinement Statistics[a]

Data Collection	**High energy**	**Low energy 1**	**Low energy 2**
Wavelength (Å)	0.9786	1.7400	1.7400
Resolution limits (Å)	50 - 1.95	50-2.45	50-2.45
	(2.02 - 1.95)	(2.54-2.45)	(2.54-2.45)
Observations	285,148	239174	254697
Unique reflections	70,761	36631	37300
Completeness (%)	97.4 (97.1)	98.1 (91.9)	99.9 (99.9)
Mean I/σ (I)	23.8 (2.8)	25.3 (10.1)	28.1 (4.5)
R_{merge}	0.052 (0.420)	0.061 (0.170)	0.063 (0.434)
Refinement			
Space group		$P2_12_12_1$	
Unit cell dimensions (Å)		a = 84.76, b=90.45, c = 128.12	
Protein atoms		6479	
DNA atoms		396	
Water molecules		400	
PEG fragments		1	
R_{cryst}		0.207 (0.267)	
R_{free}[b]		0.256 (0.297)	
RMSD[c]			
Bond lengths (Å)		0.013	
Bond angles (°)		1.54	
Ramachandran plot[d]			
Favored (%)		90.5	
Allowed (%)		8.8	
Generous (%)		0.3	
Disallowed (%)		0.4	
Average B-factors (Å^2)			
Protein		40.5	
DNA		37.4	
Water		41.1	
PEG		57.6	

[a]Values for the outer resolution shell of data are given in parenthesis.
[b]5% R_{free} test set.
[c]Compared to the Engh and Huber force field.[27]
[d]The Ramachandran plot was generated with PROCHECK.[28]

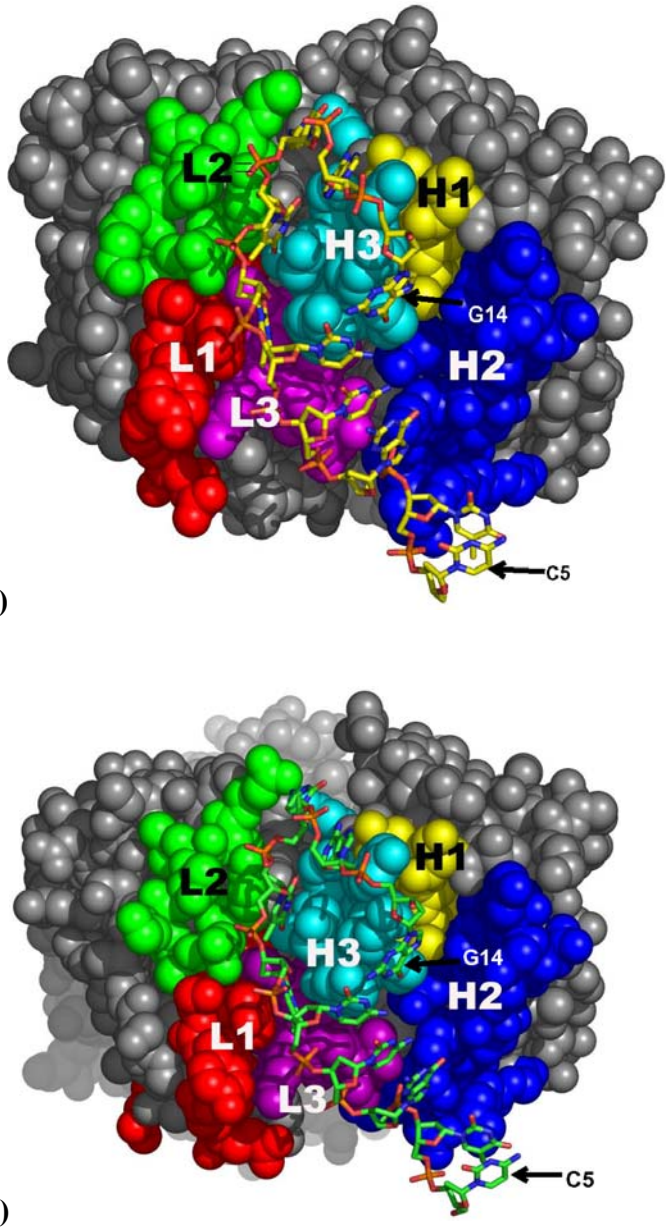


Figure 1-5. Space filling models of the two Fabs in the asymmetric unit of DNA-1/G5-14. (a) A view looking into the combining site of Fab 2.(b) A view looking into the combining site of Fab 1. The two molecules have been rotated to a common orientation to facilitate their comparison. The DNA appears in ball-and stick mode, with strand M carbon atoms colored by yellow and strand N carbon atoms colored by green. The CDRs are colored coded as follows: L1, red; L2, green; L3, purple; H1 yellow; H2, blue; and H3, cyan. This figure was prepared with PYMOL[26].

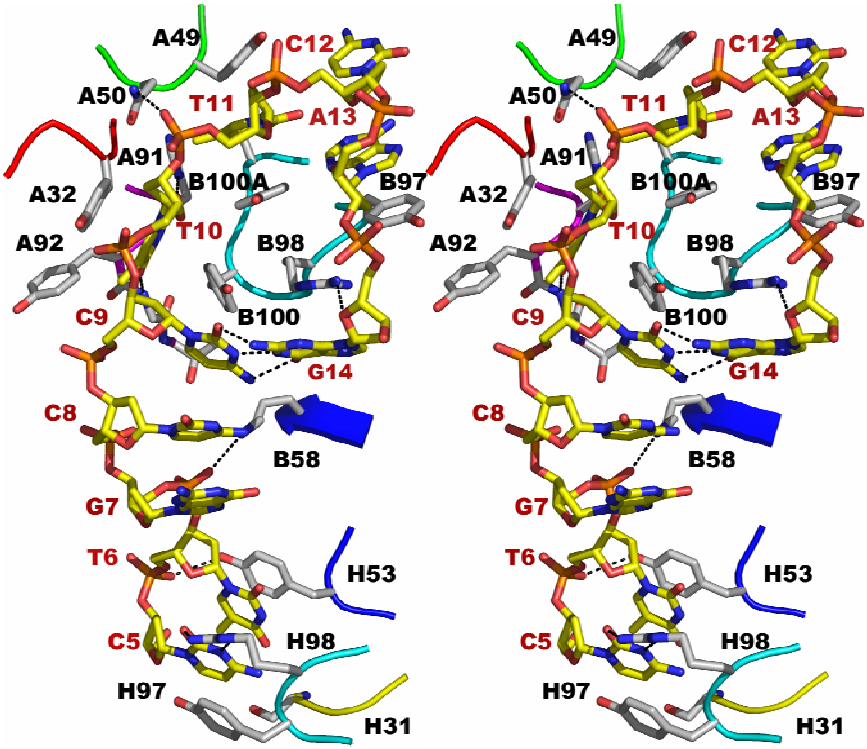


Figure 1-6. Cross-eyed stereo-view of Fab-strand M DNA interface in the DNA-1/G5-14 structure. Protein side chains are colored white, while DNA is shown in yellow. The CDRs are color coded as in Figure 1-5: L1, red; L2, green; L3, purple; H1 yellow; H2, blue; and H3, cyan. This figure was prepared with PYMOL[26].

Fab-DNA interactions

Examination of the Fab-DNA interactions in the crystal structure reveals some interesting features. All the CDRs of both Fabs, as demonstrated by the space-filling models in Figure 1-5, are involved in the Fab/DNA interactions. Nucleotides C_9, T_{10}, T_{11}, C_{12}, A_{13} and G_{14} of strand N interact with Fab 1 while nucleotides C_5 and T_6 of strand N, which form the arm region, interact with Fab 2. The remaining nucleotides of strand N, consisting of G_7 and C_8, do not interact with the protein. The Fab-DNA interactions exhibit nearly perfect two-fold symmetry. Thus, C_9, T_{10}, T_{11}, C_{12}, A_{13} and G_{14} of strand M interact with Fab 2, C_5 and T_6 of strand M interact with Fab 1, and G_7 and C_8 of strand M do not interact with the protein. The only differences between strand M and N in their interactions with the Fab CDRs are at nucleotides T_6, G_7 and C_{12}. The phosphate backbone of T_6 in strand M forms a hydrogen bond with Tyr H53 OH. The phosphate backbone of strand M G_7 forms a 4.0 Å ion pair with Lys B58 NZ. This is the only protein-DNA ion pair observed in this structure, albeit the bond distance is somewhat longer than usual. Compared with an average B-factor of 40.5 $Å^2$ of the protein, Lys B58 has a B-factor of 44.6 $Å^2$, indicating the highly flexibility of this residue. As far as strand N is concerned, the two interactions mentioned above in strand M are both absent. Instead, there is a hydrogen bond between C_{12} base N4 and Glu L56 OE2. This interaction is unavailable in strand M due to the side chain truncation of Glu A56. And a water molecule was modeled at 2.7 Å from C_{12} N4 to stabilize chain M base C_{12}.

The most striking feature of the Fab/DNA interactions is the extensive use of π-π stacking. Six Tyr and one Arg of each Fab and all the ten nucleotides of each DNA

molecule contribute to stacking. In summary, there are eight areas in the structure that have stacking interactions. Area 1: M/C_{12}, M/A_{13}, Tyr B97, N/C_5, N/T_6,Tyr B53 (Figure 1-7-1,a); Area 2: N/C_{12}, N/A_{13},Tyr H97, M/C_5, M/T_6,Tyr H53 (Figure 1-7-1, b); Area 3: Arg H98, N/G_{14}, M/G_7, M/C_8, M/C_9, Tyr B100 (Figure 1-7-2); Area 4: Arg B98, M/G_{14}, N/G_7, N/C_8, N/C_9, Tyr H100 (Figure 1-7-2); Area 5: Tyr A32, M/T_{10}, Tyr B100 (Figure 1-7-3,a); Area 6: Tyr L32, N/T_{10}, Tyr H100 (Figure 1-7-3, b); Area 7: Tyr B100A, M/T_{11}, Tyr A49 (Figure 1-7-4, a); Area 8: Tyr H100A, N/T_{11}, Tyr L49 (Figure 1-7-4, b). Note that the stacking interactions of Area 1, 3, 5, 7 are almost identical to the interactions of Area 2, 4, 6, 8.

More specifically, in area 1 or 2 (Figure 1-7-1), A_{13} base, Tyr 97 side chain, C_5 base, T_6 base and Tyr 53 side chain stack sequentially upon one another. The involvement in stacking interactions of the heavy chain Tyr 97 was also observed in Fab 2 of the DNA-1/dT5 structure. [25] In addition to the stacking interactions, the base of C_5 forms four hydrogen bonds with Ser 31 (O), Tyr 97(O), and Arg 98 (NH1, NH2) of the heavy chain. The large number of interactions between C_5 and the protein explains its low *B*-factors (28.0 $Å^2$ for M/C_5 and 30.5 $Å^2$ for N/C_5). As mentioned before, there is a hydrogen bond between the phosphate backbone of strand M/T_6 and Tyr H53. A_{13} does not form any hydrogen bond with the protein in either M or N, explaining why it has relatively high *B*-factors (37.6 $Å^2$ for M/A_{13}, 44.2 $Å^2$ for N/A_{13}). Another interaction in Area 2 is the hydrogen bond between Glu L56 and C_{12} base and it has been mentioned before.

Areas3/4 (Figure 1-7-2) is an interesting region as it is also the region where the one new Watson-Crick base pair forms in each DNA molecule. Bases of G14, G7, C8 and C9 stack parallel with one another and are sandwiched by the side chains of heavy chain Tyr 100 and heavy chain Arg 98. All the aromatic rings from the DNA bases

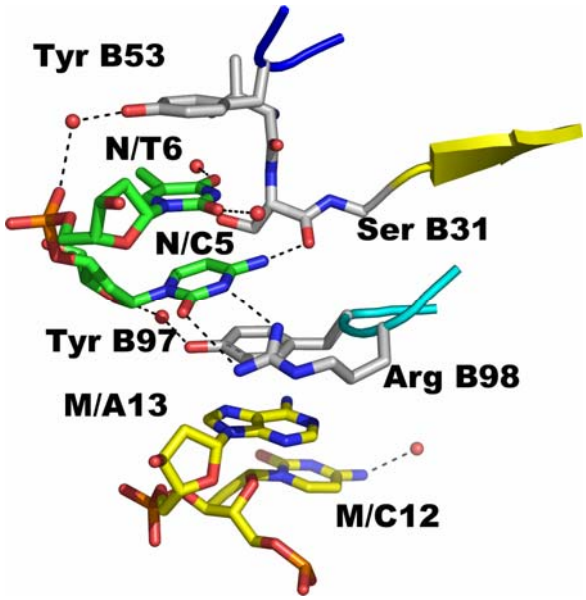
and the Tyr side chains are in the same orientation while the guanidinium of heavy chain Arg 98 point to an opposite orientation. There is a small angle between Tyr 100 and rest of the stacking system. The involvement of Arg in the stacking of ss-DNA binding proteins has been previously observed.[25; 29; 30; 31] As Arg 98 also forms two hydrogen bonds with the C_5 base and the G_{14} ribose in this structure and it is the only Arg that directly interacts with the DNA ligand in the available DNA-1 structures, it should play an important role in the specific function of DNA-1.

Areas 5/6 (Figure 1-7-3) and areas 7/8 (Figure 1-7-4) are the regions where the two sequential thymines (T_{10} and T_{11}) of the G5-14 ligand are located. These areas had the best electron density of DNA bases at the very beginning of the refinement and they were used as the starting point to build the DNA into the model. In the DNA-1/G5-14 crystal structure, the interactions between Fab DNA-1 and T_{10}/T_{11} closely resemble those from the T_2/T_3 sites of the DNA-1/dT_5 structure [25] and the T_1/T_2 sites of the DNA-1/dT3 structure [15]. In general, the thymine of T_{10} is sandwiched by the light chain Tyr 32 and the heavy chain Tyr 100 and it also makes hydrogen bonds with His L91 and Tyr H100. Similarly, the thymine of T_{11} is sandwiched by Tyr L49 and Tyr H100A and it makes a hydrogen bond with heavy chain Ala 100B. The phosphate backbone between T_{10} and T_{11} also makes a hydrogen bond with light chain Asn 50. This is the only hydrogen bond formed between the G5-14 DNA phosphate backbone and the protein which is identical in both Fabs. The only difference between T_{10} of G5-14 and T_2 of dT_5 in their interaction with Fab DNA-1 is that the latter has a hydrogen bond between its 5' phosphate backbone and the light chain Ser208 whereas this interaction is absent in the DNA-1/G5-14 structure. This result is significant as it is the first time for the " ssDNA-antibody recognition module" (D-ARM) hypothesis

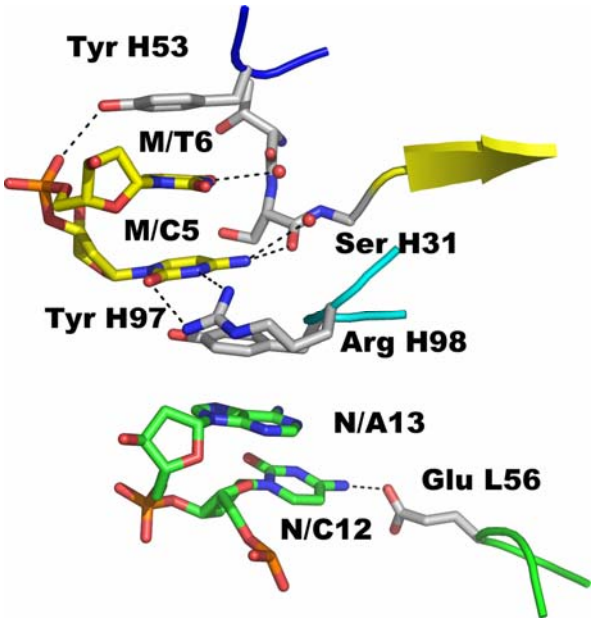
[25] to be confirmed by a crystal structure of a Fab in complex with a sequence-variable DNA ligand.

Twelve water molecules directly bind to the DNA, with ten bound to the DNA loop and the C_5-T_6 arm region and two bound to the four base-pair region in the middle of the dumbbell structure. In particular, ten of the twelve waters bind to the DNA bases while the remaining two waters bind to the DNA ribose or the phosphate backbone. These water molecules help to further stabilize the DNA- protein complex.

In conclusion, the unique recognition between DNA-1 and G5-14 is mediated by the predominant stacking interactions of hydrophobic side-chains against DNA bases and the hydrogen bonding to the bases. All these interactions are shown in Figure 1-6 represented by the Fab-strand M DNA interface and are summarized in the schematic diagram of Figure 1-8.



(a)



(b)

Figure 1-7-1. Stacking interactions in the Fab DNA-1/G5-14 structure. (a) Area 1 stacking interactions (b) Area 2 stacking interactions. Protein side chains are colored white, while DNA is shown in yellow. The CDRs are color coded as in Figure 1-4: L1, red; L2, green; L3, purple; H1 yellow; H2, blue; and H3, cyan. This figure was prepared with PYMOL[26].

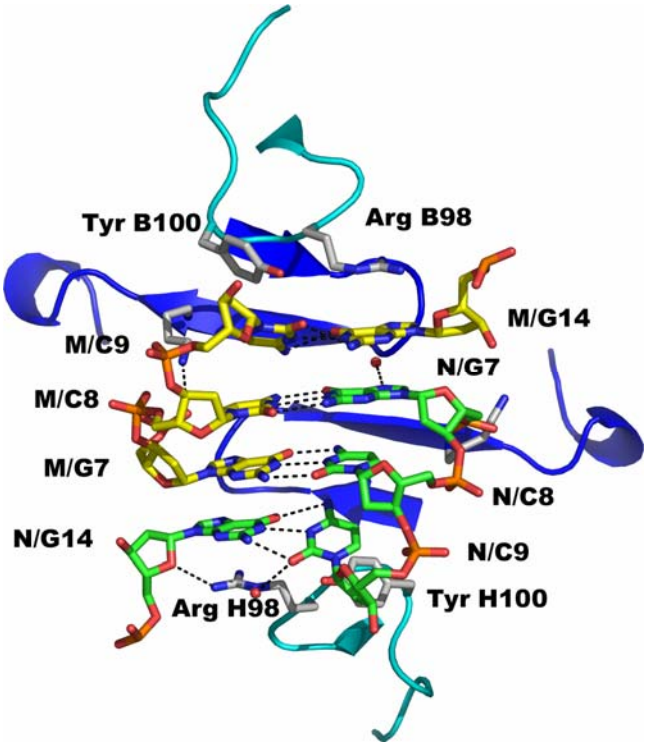
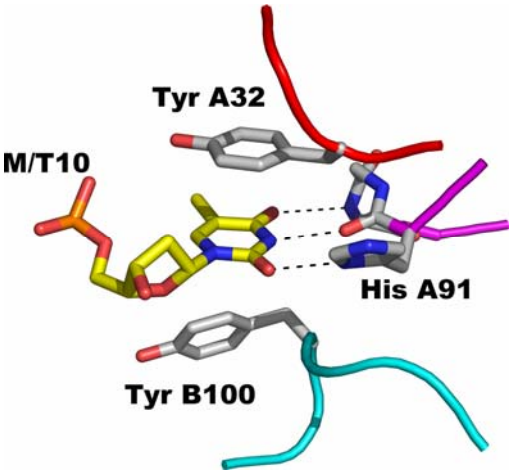
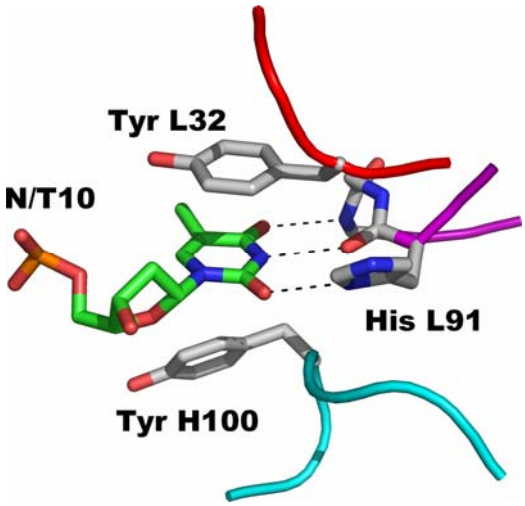


Figure 1-7-2: Stacking interactions in the Fab DNA-1/G5-14 structure Area 3/4. This is also the area where the one new hydrogen bonds form in each of the DNA molecule. Protein side chains are colored white, while DNA is shown in yellow. The CDRs are color coded as in Figure 1-4: L1, red; L2, green; L3, purple; H1 yellow; H2, blue; and H3, cyan. This figure was prepared with PYMOL[26].

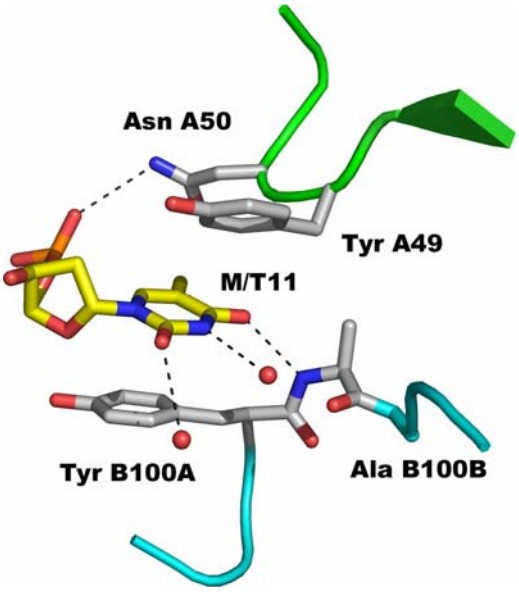


(a)

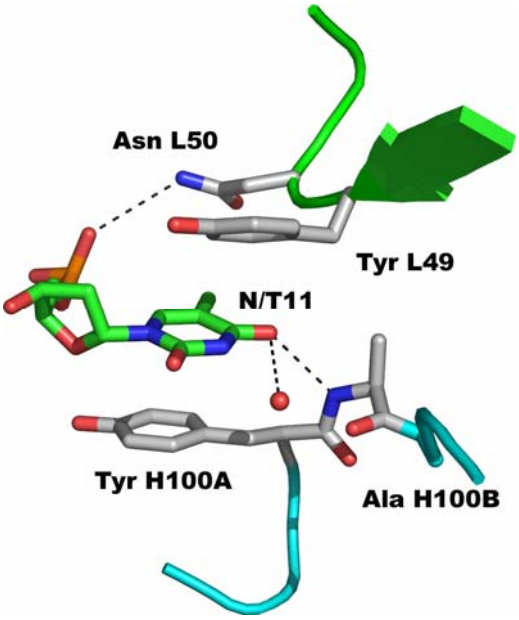


(b)

Figure 1-7-3: Stacking interactions of T10 in the Fab DNA-1/G5-14 structure, referred to Area 5/6 in the text. This element is also regarded as the " ssDNA-antibody recognition module" (D-ARM). (a) Strand M/ T10 stacking interactions (b) Strand N/T10 stacking interactions. This Figure was prepared with PYMOL. Protein side chains are colored white, while DNA is shown in yellow. The CDRs are color coded as in Figure 1-4: L1, red; L2, green; L3, purple; H1 yellow; H2, blue; and H3, cyan. This figure was prepared with PYMOL[26].



(a)



(b)

Figure 1-7-4. Stacking interactions of T11 in the Fab DNA-1/G5-14 structure, referred to Area 7/8 in the text. (a) Strand M/ T11 stacking interactions (b) Strand N/T11 stacking interactions. Protein side chains are colored white, while DNA is shown in yellow. The CDRs are color coded as in Figure 1-4: L1, red; L2, green; L3, purple; H1 yellow; H2, blue; and H3, cyan. This figure was prepared with PYMOL[26].

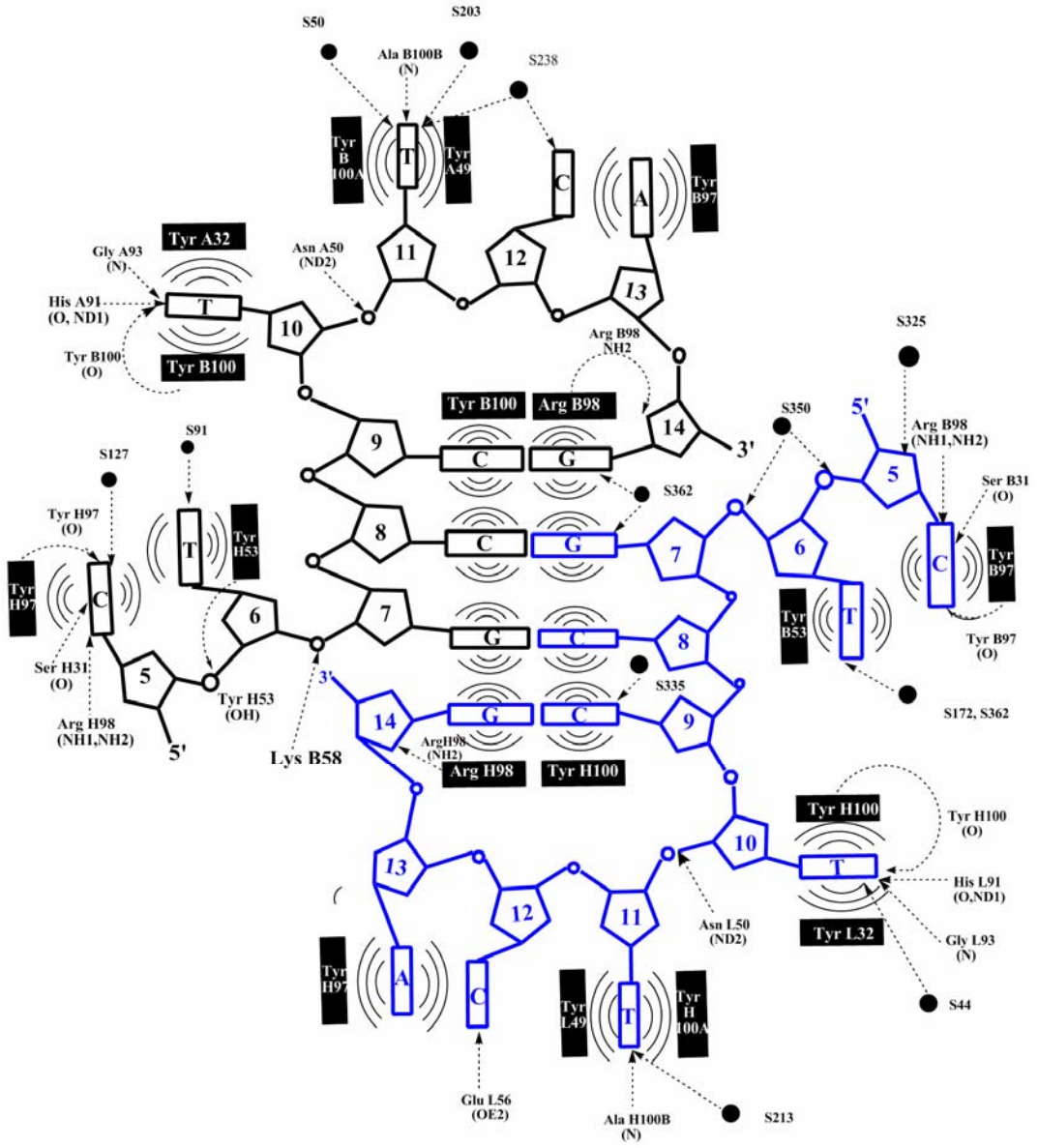


Figure 1-8. Schematic diagram of the Fab DNA-1/G5-14 interactions. The broken lines denote hydrogen bonds and ion pairs, while the arcs denote stacking interactions. Water molecules are illustrated in solid circles.

Isothermal Titration Calorimetry

As shown in the crystal structure (Figure 1-4), the two DNA ligands dimerize and have a large conformational change upon protein binding. This unique structural information prompted us to survey the behavior of G5-14 in solution when it binds to DNA-1 using ITC. G5-14 has the same ten-nucleotide sequence as the stem-loop portion above the bulge of G1-17. G1-17 is an oligonucleotide identified by an *in vitro* evolution experiment and was shown to bind with high affinity and specificity to 11F8[25], an autoantibody very similar to DNA-1 [7; 8; 9; 32] in terms of CDR sequence and base preference. For this reason, analysis of the binding behavior of G1-17 to DNA-1 in solution was also included in the ITC experiment to compare with the previous thermodynamics results of G1-17 binding to 11F8. Both titration experiments were performed in 0.01 M imidazole, 0.1 M NaCl, pH 7.0. The results for the titration of G5-14 and G1-17 were shown in Figure 1-9(a) and Figure 1-9(b) respectively. Thermodynamics data were obtained after fitting the integrated heats of injection to the one-site binding model (Table 1-2). Both ligands bind to DNA-1 with a 1:1 stoichiometry. The apparent association constant is $K_A = 2.5 \times 10^5\ M^{-1}$ for the G5-14 titration and $K_A = 1.4 \times 10^5\ M^{-1}$ for the G1-17 titration. The binding of G5-14 and G1-17 to DNA-1 are both exothermic under our buffer solution conditions, with the apparent ΔH_{cal}=-15.1 kcal/mol and ΔH_{cal}=-13.2 kcal/mol respectively. In this regard, G5-14 and G-17 seem to bind to DNA-1 with similar affinity, which is 2 to 4 fold higher than the binding of dT_5 under the same conditions ($K_A = 6.0 \times 10^4\ M^{-1}$). [25]

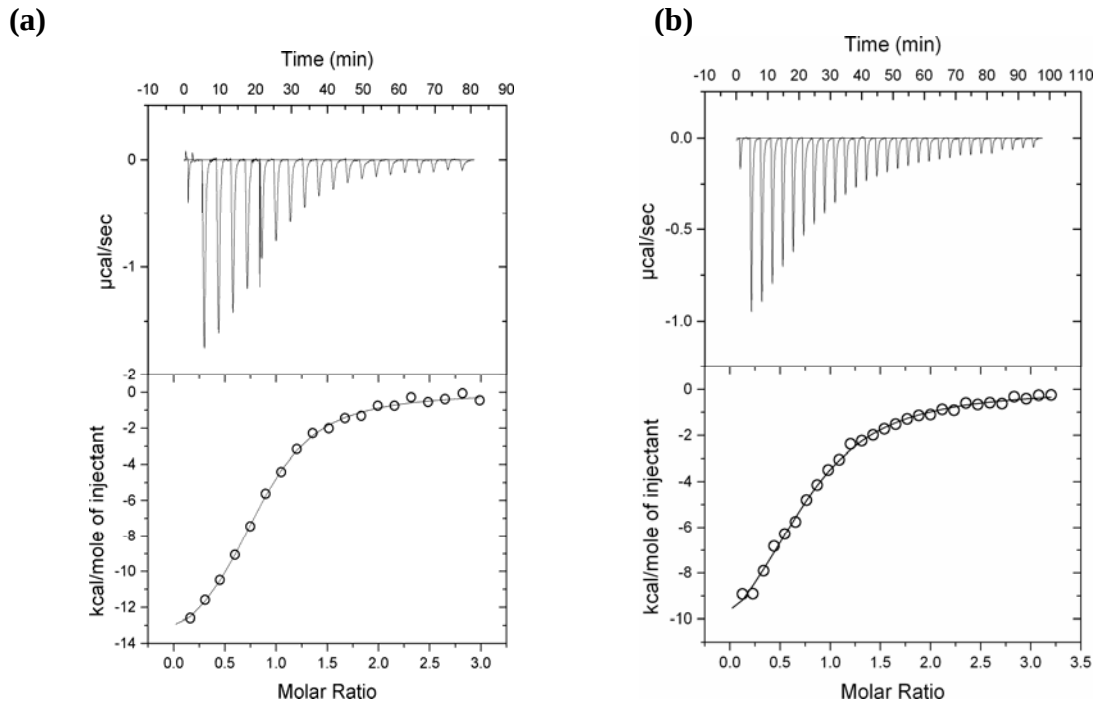


Figure 1-9. ITC analysis of G5-14 and G1-17 binding to Fab DNA-1. The upper panel of each figure shows the raw calorimetric data, and the lower panel shows the corresponding integrated injection heats, corrected for the heat of dilution. The curves in the lower halves of each figure represent the best least-squares fit to the one-site binding model. Prior to titration, G5-14, G1-17 and DNA-1 were dialyzed separately into 0.01 M imidazole, 0.1 M NaCl, pH 7.0. (a) ITC analysis of G5-14 binding to Fab DNA-1. The apparent thermodynamic parameters from the best fit curve are $K=2.5\times10^5\pm0.2\times10^5$ M^{-1}, $\Delta H=-15.1\pm0.3$ kcal/mol. (b) ITC analysis of G1-17 binding to Fab DNA-1. The apparent thermodynamic parameters from the best fit curve are $K=1.4\times10^5\pm0.1\times10^5$ M-1, $\Delta H=-13.2\pm0.5$ kcal/mol.

Table 1-2. Isothermal Titration Calorimetry Data for DNA Ligands Binding to Fab DNA-1 at 298K

Ligand	$K(M^{-1})$	ΔH (kcal/mol)	-TΔS(kcal/mol)
dT_5	6.0×10^4	-16.8	10.3
G5-14	2.5×10^5	-15.1	7.8
G1-17	1.4×10^5	-13.2	6.2

Discussion

Comparison to the DNA-1/dT_5 and DNA-1/dT_3 structures

DNA-1 represents a well-characterized anti-ss DNA antibody in terms of the available data of its genetics, biochemistry, binding thermodynamics and structural information. [4; 5; 6; 7; 8; 15; 25; 33; 34] In contrast with over 800 structures of protein-ds DNA in the NDB, there are less than 100 protein-ss DNA structures. [35] And only four of these are complexes with an antibody, including the 2.66 Å structure of BV04-01/dT_3[36], the 2.1 Å structure of DNA-1/dT_5 [25], the 2.3 Å structure of DNA-1/dT_3[15] and the 1.95 Å structure of DNA-1/G5-14 presented here. It was proposed by comparing the DNA binding site of the BV04-01/dT_3 with that of the DNA-1/dT_5 structure that a "ssDNA-antibody recognition module"(D-ARM) was important for ss-DNA recognition. [25] In the case of DNA-1, the conserved recognition locus consists of a narrow grove formed by the tyrosine side-chains of L32, H49, H100, and H100A, with hydrogen bonding provided by the side-chains of His L91, Asn L50, and the main chain of HCDR3 and LCDR3. This locus corresponds to the T_2-T_3 binding site of DNA-1/dT_5, the T_1-T_2 binding site of DNA-1/dT_3 and the T_{10}-T_{11} binding site of DNA-1/G5-14. This again agrees with the data that DNA-1 prefers to bind to thymines. [4; 5] The structure of DNA-1/G5-14 is important in this perspective as it is the first time for the conserved binding motif of DNA-1 to be structurally confirmed by a sequence-variable ssDNA ligand.

Another similarity of the three ss-DNA ligands binding to DNA-1 is that the majority of the hydrogen bonds formed between the protein and the DNA are base-specific. Compared with BV04-01/dT_3[36], which has most of the hydrogen bonds

formed to the DNA phosphate backbone, there are only six non-specific hydrogen bonds among the seventeen hydrogen bonds in the G 5-14 structure, four of which are formed with the phosphate backbone and the other two are formed with the ribose oxygens.

While it shares some similar features with the other two DNA-1/oligo(dT) structures, the structure of DNA-1/G5-14 possesses some unique features. Globally, the two Fabs are much more symmetric to each other in their DNA binding interface. Both ligands have a complete sequence of the nucleotides used for crystallization. However, in the DNA-1/dT_5 and DNA-1/dT_3 structures, some nucleotides were disordered in the second Fab. The similarity of the DNA-binding interfaces between the two Fabs is advantageous to our structural analysis in inferring common properties to the binding behavior of Fab DNA-1. In addition, since the G5-14 DNA is a much larger ligand than dT_5 and dT_3, its interactions with the protein appears to be much more extensive and specific. Sixteen of the twenty nucleotides of G5-14 in the structure are involved in the interactions with the protein by either stacking or hydrogen bonding.

Turning to the local interactions, in addition to the conserved binding sites of the two contiguous thymines, C_5, M/T_6, M/G_7, N/C_{12}, A_{13} and G_{14} also participate in protein interactions. In this regard, DNA-1 seems to be not only a thymine specific antibody, which is also true for 11F8. [12] The similar properties shared by DNA-1 and 11F8 and the initial motivation of this project inspired us to compare the current structure of DNA-1/G5-14 with the previous fluorescence quenching and foot-printing data of 11F8 bound by G1-17 and several DNA mutants sequences. Those data showed that T_{10}, T_{11}, and C_{12} directly contacted the protein while C_8 and C_9 may not recognize 11F8 by direct contact. [12] In the DNA-1/G5-14 structure, however, N/C_{12} only forms

one hydrogen bond with Glu L56. The hydrogen bond distance is reasonable (2.9 Å) but the density is not good, consistent with a high B-factor (45Å^2) of Glu L56. This interaction is absent in the other Fab. So it is hard to conclude if C_{12} contacts Fab DNA-1 or not. C_9 in the DNA-1/G5-14 structure is stacked by heavy chain Tyr100 and C_8.

When looking at the structure from the protein perspective, all CDRs of DNA-1 are involved in the interactions with G5-14 whereas HCDR1 and HCDR2 do not contribute to interactions in the DNA-1/dT_5 and DNA-1/dT_3 structure. Specifically, Ser 31 in HCDR1 makes a base-specific hydrogen bond with C_5. The involvement of HCDR1 in DNA-1 recognition helped to explain the HCDR region switch variants data [6], indicating that HCDR1 is also important in DNA-1 recognition. On the other hand, Ser 31 in DNA-1 is replaced by Arg 31 in 11F8. And mutational analysis showed that Arg 31 in 11F8 was very critical in the specific binding of the stem loop ligand to 11F8. [13]

The Arg content of the hypervariable loops is thought to be critical for ssDNA binding and correlated to antibody pathogenicity. [37; 38; 39; 40; 41; 42] DNA-1 has Arg residues in HCDR3 at positions 94 and 98 (Figure 1-2). Similar to the DNA-1/oligo dT structures[15; 25], Arg 94 of HCDR3 in the DNA-1/G5-14 structure forms a salt link across the base of HCDR3 to Asp 101 in both Fabs. This salt link presumably helps in DNA recognition by helping to maintain the structural integrity of HCDR3. On the other hand, Arg 98, which did not contribute significantly to the binding of DNA-1 to dT_3 or dT_5, interacts with the DNA both through hydrogen bonds and stacking in the DNA-1/G5-14 structure. The direct contact of the DNA by Arg 98 in DNA-1 is in

agreement with the mutational data of 11F8 [13], which suggests that Arg 98 may play important role in the binding of the stem loop ligand to both DNA-1 and 11F8.

Comparison of DNA-1/G5-14 structure to the ligand-free DNA-1 structure

Cross-reactivity with non-DNA antigens is a hallmark of anti-DNA antibodies and it is thought to be relative to their pathogenicity in autoimmune diseases. [34; 43; 44; 45; 46; 47] Structural plasticity of the hypervariable loops of the antibodies, particularly in HCDR3, was considered to underlie the cross-reactivity of anti-ssDNA antibodies. [15; 34; 36] Induced fit was previously suggested to the CDR motion of Fab BV04-01 and Fab DNA-1, the only two anti-ss DNA antibodies so far that had structures of both their ligand-free form and the DNA-bound form. [15; 36] Local conformational changes of the CDRs have occurred to both BV04-01 and DNA-1. But Fab DNA-1 is somewhat different from BV04-01 in that HCDR3 adopts multiple conformations in the ligand-free state, and the binding of the ligand has organized the disordered HCDR3 and LCDR3. To examine if the same conformational changes happen to DNA-1 when it binds to the much larger DNA ligand G5-14, the same analysis was performed for the hypervariable loops of the DNA-1/G5-14 structure. As expected, the upper half of HCDR3 (^{97}Y^{98}R^{99}P^{100}Y^{100A}Y) and Tyr L92, which were disordered in the ligand-free DNA-1 structure, are highly ordered and have great density in the DNA-1/G5-14 structure. Specifically, the main chain RMSDs, obtained by superimposing the six hypervariable loops of DNA-1/G5-14 with those of the ligand-free DNA-1 structure, are below 0.4 Å for LCDR1, LCDR2, HCDR1 and HCDR2. In contrast, the mainchain RMSDs are approximately 1.0 Å for LCDR3 and HCDR3. Thus, the major conformational changes of DNA-1 upon G5-14 binding still reside in LCDR3 and HCDR3, the recognition site for the two thymines and the same region of

the Fab DNA-1 upon dT_5 or dT_3 binding. These results support the hypothesis that different conformations adopted by LCDR3 and HCDR3 underlines the plasticity of Fab DNA-1 in the recognition of different ligands. [15] This type of plasticity may enable anti-DNA antibodies to bind diverse host ligands, and therefore contribute to pathogenicity.

ITC data and comparison to the thermodynamics data of 11F8

Results from ITC showed that the binding of G5-14 to DNA-1 in solution followed a 1:1 stoichiometry. This is consistent with the Fab/G5-14 structure where two DNA ligands bind to two Fabs in the asymmetric unit. The favorable enthalpy term associated with the unfavorable entropy term agree with the highly organized Fab/G5-14 structure, which is stabilized by tremendous π-π stacking and hydrogen bonding interactions. The "non-classical" hydrophobic effect has been previously suggested to DNA-1 and 11F8 recognition. [12; 25] G5-14 and G-17 bind to DNA-1 with similar affinity, which is 2 to 4 fold higher than dT_5 binding in the same condition ($K_A=6.0\times10^4$ M^{-1}). [25] Also, the binding entropies are smaller for the two stem loop DNAs than dT_5. This is probably due to the preorganization of the two stem loop DNAs, which reduces the conformational entropy penalty of binding relative to dT_5. A similar result was found in the thermodynamics studies of 11F8, that the DNA mutant T7, which had nucleotides on the 8-14 positions of the WT stem loop replaced by seven contiguous thymines, bound to 11F8 with 10 fold less affinity.[12] Given that two contiguous thymines are the recognition site of both DNA-1 and 11F8 [12; 15; 25], statistical effects should also be considered in the ligand binding. In this regard, DNA-1 and 11F8 seem to favor the binding of a stem-loop DNA more than poly dT. This result is interpretable as our structure indicates that G5-14 binds to Fab DNA-1 with much more specific interactions.

In terms of general thermodynamic behavior, DNA-1 and 11F8 seem to be similar to each other, however, several differences also exist in their binding. Firstly, 11F8 binds to the full-length stem loop DNA, namely G1-17 here, with a K_A=1.0 $\times 10^8$. [12] Assuming that the binding constant obtained by fluorescence quenching (performed at 0.02M Tris, 0.15M NaCl, 298 K, pH 8.0) is comparable to our current ITC result (performed at 0.01 M imidazole, 0.1 M NaCl, 298 K, pH 7.0), this is a nearly 10^3-fold higher affinity than DNA-1 bound to G1-17. Although DNA-1 and 11F8 have high sequence similarity in their CDR loops, results from previous studies also showed that slight sequence changes of the antibodies could have significant influence on their binding affinities. Examples include the binding affinity differences between DNA-1 and D5[4; 5; 6], and between 11F8 and its different mutants [13]. For these reasons, sequence alignment of DNA-1 and 11F8 might shed light on the differences of their binding affinities. In the previous mutagenesis and fluorescence quenching experiments, R31, W33, L97, R98, Y100, and Y32L in 11F8 were found to contribute to 80% of the free energy for 11F8 bound to the WT stem loop. [13] In DNA-1 sequence, R31, W33 and L97 of 11F8 are replaced by S31, V33 and Y97 while R98, Y32L and Y100 are the same. Note that R98, Y32L and Y100 are the sites for the two thymine recognition. Ser 31 has a hydrogen bond with C_5 base and Y97 provides stacking interactions with C_5 and A_{13} in the DNA-1/G5-14 structure. However, C_5 interactions are probably due to crystal packing effects (to be discussed later). Without a crystal structure of 11F8/G1-17, it is hard to interpret why the variations of these sequences affect the binding affinity of 11F8 and DNA-1.

A second difference in the thermodynamic behavior between DNA-1 and 11F8 also merits discussion. DNA-1 binds to G5-14 and G1-17 with similar affinity while the binding affinity of 11F8 decreased over 10 fold when A_3 and C_{16} were replaced with

two thymines to preclude secondary structure formation at the stem under the bulge in the WT DNA [12]. The latter data was interpreted by Ackroyd et.al. that preorganization of the recognition site was favorable for the entropy term and was important for high-affinity binding. This statement was tenable for 11F8's binding as CD spectra showed that no large-scale structural changes occurred upon 11F8 binding of the DNA sequence. Nevertheless, in the DNA-1/G5-14 structure, there are large conformational changes in both the protein and the DNA. The anticipated three base pairs at the stem that hold the secondary structure of G5-14 no longer exist upon protein binding. This may help to explain why G1-17, with a longer intact stem, did not help to bind to the Fab DNA-1 better. However, this assertion is not conclusive without a crystal structure of DNA-1/G1-17 since it is unknown whether the G1-17 ligand still adopts a similar conformation as G5-14 when it binds to DNA-1.

Large conformational change of G5-14 upon protein binding and the suggested recognition mechanism

Presumably, G5-14 adopts a hairpin conformation with three Watson-Crick base pairs at the stem and four bases at the loop in solution (Figure 1-1b). This assumption is justified as results from previous NMR and chemical foot-printing experiments showed that the selected high affinity ligand to 11F8, denoted G1-17 here was a hairpin DNA. [10] Moreover, the stem-loop conformation (Figure 1-1b) is theoretically more stable than the G5-14 conformation in the crystal structure (Figure 1-1c) in terms of the energy provided by base stacking and base pairing(Table 3-1) [48; 49]. In the crystal structure, however, the three Watson-Crick base pairs at the stem of the anticipated G5-14 secondary structure are completely lost and are replaced by one new base pair in each DNA molecule. This significant conformational change of the

DNA is unlikely to happen without the involvement of the protein. Shown in the structure, there are many hydrogen bonding and stacking interactions in the protein-DNA interface. These interactions likely offset some of the enthalpic loss due to the disruption of the stem-loop upon binding.

As discussed previously, the protein also undergoes a large conformational change on its binding surface. Referring to the investigations of Fab DNA-1 and 11F8's recognition mechanism in the last decade [14; 15], the whole story of how the Fab DNA-1 recognizes the stem-loop DNA G5-14 in solution could be proposed as follows. Initially, G5-14 adopts the expected stem-loop secondary conformation while the Fab presents its canonical ß-sheet structure with the six hypervariable loops where HCDR3 and LCDR3 are highly disordered. The ligand-free Fab exhibits multiple conformations as suggested by the " preexisting isomeric equilibrium" theory. [15; 50; 51; 52] As the Fab encounters the DNA by recognizing its two contiguous thymines, "induced fit"[34; 51; 53; 54; 55; 56; 57] happens where the Fab modifies its loop conformation to make base-specific hydrogen bonding and π-π stacking interactions with the DNA. The large energy released by the hydrophobic desolvation forces the DNA to open up its original three base pairs at the stem, moving towards its 5'-terminal by two bases, and continue to form more interactions with the protein. When maximal interactions are optimized in the observed complex, both the DNA nucleotides and the protein loops have become highly ordered. DNA is converted into a new conformation with one extra C9-G14 base pair. In summary, the protein and the DNA both undergo significant conformational changes in their recognition. Similar cofolding processes were previously reported for two stem-loop RNA/ protein complexes. [16; 17] In the NMR structure of two RNA-binding domains of Nucleolin bound to a Pre-rRNA target, the Watson-Crick U9·A14 and the non-canonical C10·G13 base-pairs are

disrupted to take part in specific protein-RNA interactions and the disordered loops in the ligand-free protein have become structured upon the RNA binding. [17] Likewise, in the structure of the Alfalfa mosaic virus RNA and its coat protein complex, the two RNA hairpins are extended by two base pairs between the AUGC repeats accompanied by a conformational change of the N' terminal peptide from α helix to extended chain. [16] This mode of protein-DNA binding is an example of "mutually induced fit", which is a common feature of protein-RNA recognition. [16; 17; 58; 59; 60].

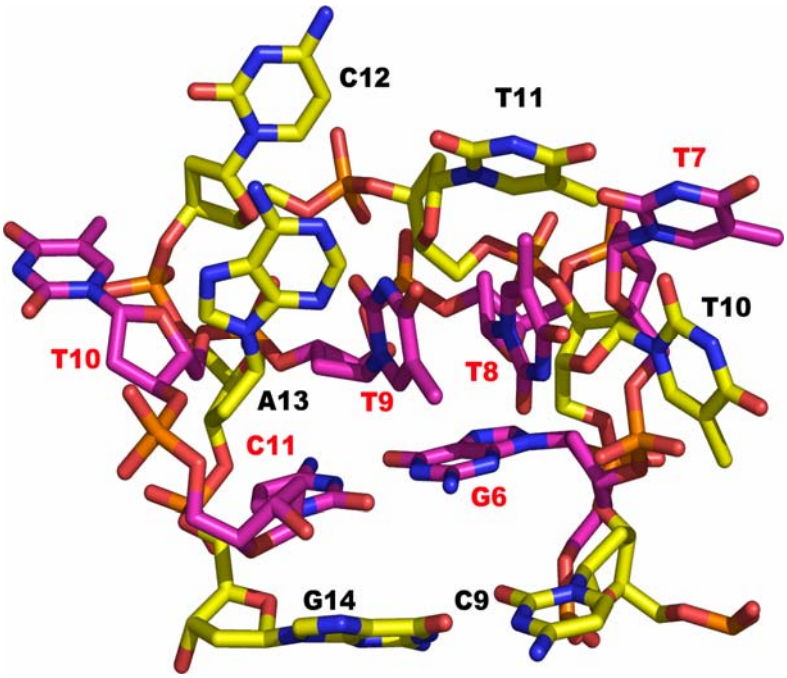


Figure 1-10 Superimposition of the loops from G-14 and 1D16 in their crystal structures, indicating G5-14 conformational changes upon protein binding. G5-14 is colored yellow and 1D16 is colored magenta. 1D16 bases all point inward while G5-14 bases flip outward to make interactions with the protein. This figure was prepared by PYMOL.[26]

Crystallographic artifacts versus solution behaviour

Macromolecular behavior in the crystal is sometimes different from their solution behavior due crystal packing interactions. DNA-1 and G 5-14 concentrations in ITC were 0.03 mM and 0.4 mM respectively while their crystallization concentrations were 0.24 mM and 1.4 mM respectively. Additional intramolecular or intermolecular interactions, referred as crystal contacts, might occur solely as a result of protein crystallization. [61] As only the contacts formed in solution, such as the specific antigen-antibody complexes and oligomer intefaces, could provide relevant insights to these molecules' biological function, it is important for us to distinguish the real solution interactions from the crystallographic artifact interactions.

In the asymmetric unit of the crystal structure, the highly ordered DNA-1/G5-14 complex appears to be unusual. Our ITC data could not explain the DNA dimerization behaviour as the 1:1 stoichiometry could result from either one protein binding to one DNA or two proteins binding to two DNAs. Previous dynamic light scattering data (unpublished data from Season Prewitt, Table 3-2) suggested that either Fab DNA-1 alone or the Fab DNA-1/dT_3 complex was an monomer in solution. In this regard, it seems impossible for one G5-14 molecule to interact with two Fabs in solution. In the DNA-1/G5-14 crystal structure, however, the arm region comprising C_5 and T_6 interacts with one Fab whereas the rest nucleotides interact with the other Fab. Therefore, the interactions between Fab and the DNA C_5 and T_6 bases are probably due to crystallographic packing. And so is the dimerization of the two DNA molecules. In contrast, the DNA-protein interactions formed in nucleotides C_9 to G_{14} should reflect the real solution behavior in the DNA-1/G5-14 recognition. Similar

crystallographic packing effects were previously discussed in the DNA-1/ dT_5 structure. [25] As we've already mentioned, heavy chain Tyr 97 is also involved in the Fab 2 stacking interactions in the DNA-1/dT_5 structure, which is also probably caused by crystallographic packing effects.

Comparison of the protein-ssDNA interactions with protein-RNA interactions

Unlike RNAs, which are usually single stranded and exhibit a variety of conformations, cellular DNAs exist primarily as a long double helix. Single stranded DNAs are observed only occasionally when the cell is subject to environmental stresses such as replication defects and apoptosis. As a unique group of ssDNAs, stem loop DNAs are found to mediate some critical cellular processes, such as regulation of DNA synthesis[62; 63] and the genome evolution [64]. In our present discussion, the stem-loop ligand G1-17 binds to 11F8 with high affinity and specificity[10], and 11F8 was found to localize to kidney tissue by binding to DNA adherent to the GBM and eventually lead to renal damage[11]. DNA antigens present in human anti-DNA complexes are found to have similar sequences and structures as G1-17[63; 65]. In this perspective, it is also plausible that stem loop DNAs may be implicated in the deposition process of autoimmune complexes and further lead to lupus development. It would therefore be interesting to generate a recognition profile of these stem loop DNAs to give insights to their cellular roles. In the PDB [35], besides our structure, there are only two other published structures of stem-loop DNA/protein complexes (PDB entry: 1OMH and 2A6O) and about 200 structures of stem-loop RNA/protein complexes. Results from comparisons of all these deposited structures demonstrated that protein-ssDNA interactions closely resemble protein-RNA interactions. In

addition to the classical helix recognition modules of dsDNA, ssDNA and RNA also interact with the protein loops and ß sheets. Hydrogen bonds are formed both specifically with the nucleotide bases and nonspecifically with the phosphate backbones. Different from DNA, RNA could also use its 2'-OH to make hydrogen bonds with the protein. [66]. Stacking interactions between nucleotide bases and protein planar side chains are both observed. Furthermore, cofolding of the protein and nucleotide sequences occur in both protein/ssDNA and protein/RNA recognition. The structural conformational changes coupled with the cofolding events could be as small as some domain residue variations or as large as breaking and reforming new hydrogen bonds.

Structure-based drug design of the autoimmune disease systemic lupus erythematosus

The approach of nucleic acid structure based drug design has been developed as a new exciting research area in recent years. Stem-loop RNA aptamers selected by in vitro evolution experiments were found to inhibit the HIV-1 reverse transcriptase. [67; 68] Small molecules that mimic the interactions of RNA were chemically synthesized to target the decoding site of the ribosome to regulate the translational machinery. [69; 70; 71]

However, based on the information provided by the only two anti-ss DNA antibodies BV04-01 and DNA-1's structures, anti-DNA Abs represent a challenging ligand design target. Firstly, the antigen binding site is highly flexible and seems to be able to modify its conformation to different ligands. Secondly, comparing the binding affinity and structure of G5-14 to DNA-1 with those of G1-17 to 11F8, it appears that the same DNA sequence would have different affinity and different conformation

specificity to different anti-DNA antoantibodies, although these antibodies have only minimal sequence difference. In our present work, small molecules that resemble the shape and specific interactions provided by G1-17 (Figure 1-1a) are likely to inhibit 11F8 best whereas for DNA-1, another stem-loop shown in Figure 1-1d may work better. As we have mentioned in the discussion, the only common area for DNA-1 and 11F8 recognition are the two continuous thymines while the interactions provided by other parts of the DNA are different for the two antibodies.

Materials and methods

Expression and purification of DNA-1

Expression and purification of the recombinant Fab DNA-1 were performed as previously described.[72] Briefly, a single recombinant E. coli colony was picked from a previously streaked DNA-1 bacterial plate and placed into 15 mL of Super Broth with 20 mM $MgCl_2$ and 50 μg/mL each of carbenicillin and streptomycin. The starter culture was allowed to grow for 12 h at 310 K and expanded to a culture of 1.5 L. After growing for another 6-9 h, the expanded culture was induced with IPTG at a final concentration of 1 mM for 12 h. Cells were then harvested and disrupted by sonication. DNA-1 was purified by Ni-NTA affinity (Figure 3-1) and cation exchange chromatography (Figure 3-2) using an Akta FPLC. All columns were purchased from Amersham Biosciences. Purity of the protein was analyzed by SDS-PAGE(Figure 3-3).

The cation exchange chromatography step was notable because DNA-1 eluted in two peaks from a linear gradient consisting of 0-74 mM NaCl in a 100 mL 20 mM Na_2HPO_4 pH 6.5. SDS-PAGE showed that proteins from both peaks ran to the same

position on the gel corresponding to the ~48 kDa mass of DNA-1. To further identify their difference, protein samples from both peaks were subjected to MALDI-TOF mass spectral analysis at the University of Missouri-Columbia Proteomics Center (Figure 3-4). Protein from peak 1 contained two species having masses 48 kDa and 48.5 kDa. In contrast, protein from peak 2 contained a single species with mass of 48 kDa. DNA-1 samples from the two peaks were separately dialyzed into 10 mM Tris, 50 mM NaCl, pH 7.0 and concentrated to 16 mg/ml using a stirred Amicon ultrafiltration cell (30 kDa MW cutoff). Protein concentration was determined spectrophotometrically using a theoretically determined optical density value (λ=280 nm) of 1.6 mg^{-1} cm^{-1} ml. [73] Protein from cation exchange peak 2 was used for crystallization.

Crystallization and X-ray data collection

The stem-loop DNA oligonucleotide 5'-CpTpGpCpCpTpTpCpApG-3' was purchased from Integrated DNA Technologies, Inc. (standard desalting preparation). This oligonucleotide corresponds to nucleotides 5-14 of the 17-nucleotide stem-loop DNA ligand identified that binds with high affinity and specificity to Fab 11F8. [10] Prior to crystallization, DNA was dissolved in 10 mM Tris, 50 mM NaCl, pH 7.5 to achieve a 4.7 mM stock solution. The DNA/Fab solution used for crystallization studies was formed by mixing DNA in excess with protein at a molar ratio of [DNA]/[Fab] = 3-5. Crystallization experiments were performed in hanging drops by mixing 3 µl of the DNA-1/G5-14 complex solution with 3 µl of the reservoir at 295 K. Initial screening experiments using Emerald Biostructures Wizard kits produced crystals grown in Wizard I Formulations 10 and 19. These crystals were in the shape of chunks of bars having dimensions 0.6 mm ×0.2 mm × 0.1 mm (Figure 3-5). After several rounds of

optimization, reservoir solutions containing 0.1 M Tris pH 7.0, 12-26% (w/v) PEG 2000 MME were found to yield the best crystals. The crystals occupy space group $P2_12_12_1$ with cell dimensions a =84.76 Å, b = 90.45 Å, c = 128.12 Å. The asymmetric unit contains two Fab molecules and two complete G5-14 DNA molecules, with solvent content of 51% and V_M = 2.5 Å3 Da-1. [74]

Several crystals were prepared for cryogenic X-ray data collection by replacing the mother liquor with 0.1 M Tris-HCl pH 7.0, 25% (w/v) PEG 2000 MME and 10% (v/v) PEG 200. The crystals were then plunged into liquid nitrogen.

Three X-ray diffraction data sets were collected at Advanced Photon Source beamline 19-ID. The 1.95 Å resolution data set used for refinement consisted of a 100° scan with oscillation width of 0.5° per frame, exposure time of 10 s per frame, detector distance of 160 mm, and detector 2θ angle of zero. Two data sets were collected at low energy (l= 1.740) for use in anomalous difference Fourier analysis. Low energy data set 1 consisted of a 171° scan with oscillation width 0.5°, exposure time of 5 s per frame, detector distance of 120 mm and detector 2θ of -20°. Low energy data set 2 consisted of a 219° scan collected with oscillation width 0.5°, exposure time of 10 s per frame, detector distance of 120 mm and detector 2θ of -10°. All three data sets were integrated and merged with HKL2000.[75] See Table 1-1 for data collection statistics.

Model building, refinement, and analysis

Molecular replacement calculations were performed with MOLREP[76] using conserved and variable superdomain search models obtained from the DNA-1/dT_5 structure

(PDB entry 1I8M). The best solution, which had correlation coefficient of 0.3, was obtained for 2 Fabs per asymmetic unit in space group P212121.

Initially, O[77] was used for model building and CNS[78] was used for structure refinement. Simulated annealing was used in the first round of refinement and conjugate gradient refinement was used thereafer. Non-crystallographic symmetry restraints were not used. Cross validation was used with a randomly chosen test set of 3596 reflections and a working set of 63,514 reflections. After the protein model was nearly complete, 54 water molecules were added using the water picking algorithm of CNS. The inclusion of these waters in the model helped to improve the map quality and decrease R and R_{free} to 0.297 and 0.336, respectively. At this stage, $2F_o - F_c$ and $F_o - F_c$ maps showed the presence of DNA base pairs.

COOT[79] was used to model DNA and for all subsequent model building sessions. Models built with COOT were refined with REFMAC5.[80] Modeling of DNA phosphate atoms was guided by anomalous difference Fourier maps calculated from the model phases and anomalous differences from the low energy data sets. Maps were calculated to a resolution of 3.0 Å from each of the two low energy data sets and then averaged together using CNS. It was found that the averaged map was superior to either of the individual maps for locating P atoms.

Based on sequence and structure similarity, a synthetic DNA hairpin molecule (5'- Cp Gp Cp Gp Cp Gp Tp Tp Tp Tp Cp Gp Cp Gp Cp G -3')(PDB entry 1D16) was used as a template by removing its first and the last three bases and substituting the remaining bases with the corresponding G5-14 DNA bases. As the stacking interactions between the DNA thymine and the DNA-1 tyrosine were proved to be important in the anti-ssDNA Fab recognition in the previous DNA-1/ dT5 (PDB entry

1I8M) and DNA-1/dT3 structures (PDB entry 1XF2), the two thymines of the tenth and eleventh nucleotides of the G5-14 sequence were used as the starting point to build up the DNA model. Basically, dpT10 was fit into the planar density between Tyr H100 and Tyr L32 and dpT11 was fit into the planar density between Tyr H100A and Tyr L49. Other bases of the G5-14 ligand were built up subsequently.

REFMAC5(including TLS) [81; 82; 83] was used for maximum likelihood refinements of the model. The groups for TLS refinement were as follows: group 1 A1-A108, group 2 A109-A213, group 3 L1-L108, group 4 L109-L213, group 5 B1-B113, group 6 B114-B213, group 7, H1-H113, group 8 H114-H213. Hydrogens were added in the riding positions in the refinement. At early stage of refinements, nucleotides that didn't occupy clear 0.5σ $2F_o - F_c$ density were excluded. As the quality of the map improved, the occupancies of these nucleotides were set back to 1 in the refinement. Both ends of the DNA ligand were hard to locate. Trial and error experiments were performed. The other G5-14 ligand in the asymmetric unit was built in a similar manner. Superposition of the two G5-14 ligand molecules was performed to help locating the unclear nucleotides. After G5-14 ligand building was completed, additional water molecules and one PEG fragment were incorporated to the model. The two DNA strands were included as two separate groups in the final stage of the TLS refinement. The final model had R_{cryst} =0.207 and R_{free}=0.256. See Table 1 for refinement statistics. Structure analysis was performed with COOT [79] and CCP4[81].

Isothermal Titration Calorimetry

ITC experiments were performed in a MicroCal VP-ITC at 298 K to obtain further information about the thermodynamic basis of the DNA ligand binding. The two oligonucleotides G5-14(5'-CpTpGp CpCpTp TpCpAp G -3') and G1-17(5'-

CpGpAp GpCpTp GpCpCp TpTpCp ApGpTp CpG -3') were purchased from Integrated DNA Technologies, Inc. (standard desalting preparation). Prior to ITC experiments, solutions of the DNA ligand and Fab DNA-1 were dialyzed separately into 0.01 M imidazole, 0.1 M NaCl (pH 7.0) buffer. For both titrations, Fab DNA-1 (0.03 mM) was placed in the calorimeter cell, and the oligonucleotide (0.4 mM) was added to the cell via the rotating stirrer-syringe. Blank titrations were performed by injecting the ligand (0.4 mM) into the buffer devoid of protein to correct for the heat of mixing and ligand dilution. Values of the association constant, enthalpy and entropy were obtained by fitting the integrated heats of association to the one-site binding model using the software supplied with the instrument.

PDB accession code

Atomic coordinates and structure factor amplitudes have been deposited in the PDB[84] as entry 2FR4.

1. Lahita, R. G. *Systemic lupus erythematosus*. 3rd edit, Academic Press, San Diego, CA; London, UK.

2. Abbas, A. K., Lichtman, A. H. & Pober, J. S. (2000). *Cellular and molecular immunology.* Fourth Edition edit, W.B. Saunders Company, Philadelphia,London,New York, St. Louis, Sydney, Toronto.

3. Tahir, H. & Isenberg, D. A. (2005). Novel therapies in lupus nephritis. *Lupus* 14, 77-82.

4. Calcutt, M. J., Kremer, M. T., Giblin, M. F., Quinn, T. P. & Deutscher, S. L. (1993). Isolation and characterization of nucleic acid-binding antibody fragments from autoimmune mice-derived bacteriophage display libraries. *Gene* 137, 77-83.

5. Komissarov, A. A., Calcutt, M. J., Marchbank, M. T., Peletskaya, E. N. & Deutsher, S. L. (1996). Equilibrium binding studies of recombinant anti-single-stranded DNA Fab. Role of heavy chain complementarity-determining regions. *J Biol Chem* 271, 12241-6.

6. Calcutt, M. J., Komissarov, A. A., Marchbank, M. T. & Deutscher, S. L. (1996). Analysis of a nucleic-acid-binding antibody fragment: Construction and characterization of heavy-chain complementarity-determining region switch variants. *Gene* 168, 9-14.

7. Komissarov, A. A., Marchbank, M. T., Calcutt, M. J., Quinn, T. P. & Deutscher, S. L. (1997). Site-specific mutagenesis of a recombinant anti-single-stranded DNA Fab. Role of heavy chain complementarity-determining region 3 residues in antigen interaction. *J Biol Chem* 272, 26864-70.

8. Komissarov, A. A. & Deutscher, S. L. (1999). Thermodynamics of Fab-ssDNA interactions: contribution of heavy chain complementarity determining region 3. *Biochemistry* 38, 14631-7.

9. Swanson, P. C., Ackroyd, C. & Glick, G. D. (1996). Ligand recognition by anti-DNA autoantibodies. Affinity, specificity, and mode of binding. *Biochemistry* 35, 1624-33.

10. Stevens, S. Y. & Glick, G. D. (1999). Evidence for sequence-specific recognition of DNA by anti-single-stranded DNA autoantibodies. *Biochemistry* 38, 560-8.

11. Swanson, P. C., Yung, R. L., Blatt, N. B., Eagan, M. A., Norris, J. M., Richardson, B. C., Johnson, K. J. & Glick, G. D. (1996). Ligand recognition by murine anti-DNA autoantibodies. II. Genetic analysis and pathogenicity. *J Clin Invest* 97, 1748-60.

12. Ackroyd, P. C., Cleary, J. & Glick, G. D. (2001). Thermodynamic basis for sequence-specific recognition of ssDNA by an autoantibody. *Biochemistry* 40, 2911-22.

13. Cleary, J. & Glick, G. D. (2003). Mutational analysis of a sequence-specific ssDNA binding lupus autoantibody. *Biochemistry* 42, 30-41.

14. Beckingham, J. A., Cleary, J., Bobeck, M. & Glick, G. D. (2003). Kinetic analysis of sequence-specific recognition of ssDNA by an autoantibody. *Biochemistry* 42, 4118-26.

15. Schuermann, J. P., Prewitt, S. P., Davies, C., Deutscher, S. L. & Tanner, J. J. (2005). Evidence for structural plasticity of heavy chain complementarity-determining region 3 in antibody-ssDNA recognition. *J Mol Biol* 347, 965-78.

16. Guogas, L. M., Filman, D. J., Hogle, J. M. & Gehrke, L. (2004). Cofolding organizes alfalfa mosaic virus RNA and coat protein for replication. *Science* 306, 2108-11.

17. Johansson, C., Finger, L. D., Trantirek, L., Mueller, T. D., Kim, S., Laird-Offringa, I. A. & Feigon, J. (2004). Solution structure of the complex formed by the two N-terminal RNA-binding domains of nucleolin and a pre-rRNA target. *J Mol Biol* 337, 799-816.

18. Martin, A. C. (1996). Accessing the Kabat antibody sequence database by computer. *Proteins* 25, 130-3.

19. Kabat, E. A., Wu, T. T., Perry, H. M., Gottesmann, K. S. & Foeller, C. (1991). *Sequences of Proteins of Immunological Interest, 5th edit.*, National Institutes of Health, Bethesda, MD.

20. Brunger, A. T. (1993). Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Crystallogr D Biol Crystallogr* 49, 24-36.

21. Luzzati, P. V. (1952). Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Crystallog.* 5, 802-810.

22. Read, R. J. (1986). Improved fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallog.sec. A* 42, 140-149.

23. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26, 283-291.

24. Amzel, L. M. & Poljak, R. J. (1979). Three-dimensional structure of immunoglobulins. *Annu Rev Biochem* 48, 961-97.

25. Tanner, J. J., Komissarov, A. A. & Deutscher, S. L. (2001). Crystal structure of an antigen-binding fragment bound to single-stranded DNA. *J Mol Biol* 314, 807-22.

26. DeLano, W. L. (2002). The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos,CA, USA.

27. Huber, R. A. E. a. R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta crystallographica A* 47, 392-400.

28. R. A. Laskowski, M. W. M., D. S. Moss and J. M. Thornton. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* 26, 283-291.

29. Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R. & Xu, R. M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev* 13, 1102-15.

30. Horvath, M. P., Schweiker, V. L., Bevilacqua, J. M., Ruggles, J. A. & Schultz, S. C. (1998). Crystal structure of the Oxytricha nova telomere end binding protein complexed with single strand DNA. *Cell* 95, 963-74.

31. Raghunathan, S., Kozlov, A. G., Lohman, T. M. & Waksman, G. (2000). Structure of the DNA binding domain of E. coli SSB bound to ssDNA. *Nat Struct Biol* 7, 648-52.

32. Lefkowith, J. B., Di Valerio, R., Norris, J., Glick, G. D., Alexander, A. L., Jackson, L. & Gilkeson, G. S. (1996). Murine glomerulotropic monoclonal antibodies are highly oligoclonal and exhibit distinctive molecular features. *J Immunol* 157, 1297-305.

33. Deutscher, S. L., Crider, M. E., Ringbauer, J. A., Komissarov, A. A. & Quinn, T. P. (1996). Stability studies of nucleic acid-binding Fab isolated from combinatorial bacteriophage display libraries. *Arch Biochem Biophys* 333, 207-13.

34. Schuermann, J. P., Henzl, M. T., Deutscher, S. L. & Tanner, J. J. (2004). Structure of an anti-DNA fab complexed with a non-DNA ligand provides insights into cross-reactivity and molecular mimicry. *Proteins* 57, 269-78.

35. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S. H., Srinivasan, A. R. & Schneider, B. (1992). The nucleic acid database. A comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys J* 63, 751-9.

36. Herron, J. N., He, X. M., Ballard, D. W., Blier, P. R., Pace, P. E., Bothwell, A. L., Voss, E. W., Jr. & Edmundson, A. B. (1991). An autoantibody to single-stranded DNA: comparison of the three-dimensional structures of the unliganded Fab and a deoxynucleotide-Fab complex. *Proteins* 11, 159-75.

37. Seal, S. N., Hoet, R. M., Raats, J. M. & Radic, M. Z. (2000). Analysis of autoimmune bone marrow by antibody-phage display: somatic mutations and third complementarity-determining region arginines in anti-DNA gamma and kappa V genes. *Arthritis Rheum* 43, 2132-8.

38. Li, Z., Schettino, E. W., Padlan, E. A., Ikematsu, H. & Casali, P. (2000). Structure-function analysis of a lupus anti-DNA autoantibody: central role of the heavy chain complementarity-determining region 3 Arg in binding of double- and single-stranded DNA. *Eur J Immunol* 30, 2015-26.

39. Mitamura, K., Suenaga, R., Wilson, K. B. & Abdou, N. I. (1996). V gene sequences of human anti-ssDNA antibodies secreted by lupus-derived CD5-negative B cell hybridomas. *Clin Immunol Immunopathol* 78, 152-60.

40. Van Es, J. H., Aanstoot, H., Gmelig-Meyling, F. H., Derksen, R. H. & Logtenberg, T. (1992). A human systemic lupus erythematosus-related anti-cardiolipin/single-stranded DNA autoantibody is encoded by a somatically mutated variant of the developmentally restricted 51P1 VH gene. *J Immunol* 149, 2234-40.

41. Jang, Y. J., Sanford, D., Chung, H. Y., Baek, S. Y. & Stollar, B. D. (1998). The structural basis for DNA binding by an anti-DNA autoantibody. *Mol Immunol* 35, 1207-17.

42. Radic, M. Z., Mackle, J., Erikson, J., Mol, C., Anderson, W. F. & Weigert, M. (1993). Residues that mediate DNA binding of autoimmune antibodies. *J Immunol* 150, 4966-77.

43. Kowal, C., Weinstein, A. & Diamond, B. (1999). Molecular mimicry between bacterial and self antigen in a patient with systemic lupus erythematosus. *Eur J Immunol* 29, 1901-11.

44. Sharma, A., Isenberg, D. A. & Diamond, B. (2001). Crossreactivity of human anti-dsDNA antibodies to phosphorylcholine: clues to their origin. *J Autoimmun* 16, 479-84.

45. Mostoslavsky, G., Fischel, R., Yachimovich, N., Yarkoni, Y., Rosenmann, E., Monestier, M., Baniyash, M. & Eilat, D. (2001). Lupus anti-DNA autoantibodies cross-react with a glomerular structural protein: a case for tissue injury by molecular mimicry. *Eur J Immunol* 31, 1221-7.

46. DeGiorgio, L. A., Konstantinov, K. N., Lee, S. C., Hardin, J. A., Volpe, B. T. & Diamond, B. (2001). A subset of lupus anti-DNA antibodies cross-reacts with the NR2 glutamate receptor in systemic lupus erythematosus. *Nat Med* 7, 1189-93.

47. Ray, S. K., Putterman, C. & Diamond, B. (1996). Pathogenic autoantibodies are routinely generated during the response to foreign antigen: a paradigm for autoimmune disease. *Proc Natl Acad Sci U S A* 93, 2019-24.

48. Van Holde, K. E., Johnson, W. C. & Ho, P. S. (1998). *Principles of Physical Biochemistry*, Prentice-Hall, Inc., New Jersey.

49. Friedman, R. A. & Honig, B. (1995). A free energy analysis of nucleic acid base stacking in aqueous solution. *Biophys J* 69, 1528-35.

50. Foote, J. (2003). Immunology. Isomeric antibodies. *Science* 299, 1327-8.

51. Foote, J. & Milstein, C. (1994). Conformational isomerism and the diversity of antibodies. *Proc Natl Acad Sci U S A* 91, 10370-4.

52. James, L. C., Roversi, P. & Tawfik, D. S. (2003). Antibody multispecificity mediated by conformational diversity. *Science* 299, 1362-7.

53. Edmundson, A. B., Ely, K. R., Girling, R. L., Abola, E. E., Schiffer, M., Westholm, F. A., Fausch, M. D. & Deutsch, H. F. (1974). Binding of 2,4-dinitrophenyl compounds and other small molecules to a crystalline lambda-type Bence-Jones dimer. *Biochemistry* 13, 3816-27.

54. Davies, D. R. & Padlan, E. A. (1992). Twisting into shape. *Curr Biol* 2, 254-6.

55. Schulze-Gahmen, U., Rini, J. M. & Wilson, I. A. (1993). Detailed analysis of the free and bound conformations of an antibody. X-ray structures of Fab 17/9 and three different Fab-peptide complexes. *J Mol Biol* 234, 1098-118.

56. Rini, J. M., Schulze-Gahmen, U. & Wilson, I. A. (1992). Structural evidence for induced fit as a mechanism for antibody-antigen recognition. *Science* 255, 959-65.

57. Wilson, I. A. & Stanfield, R. L. (1994). Antibody-antigen interactions: new structures and new conformational changes. *Curr Opin Struct Biol* 4, 857-67.

58. Allain, F. H., Howe, P. W., Neuhaus, D. & Varani, G. (1997). Structural basis of the RNA-binding specificity of human U1A protein. *Embo J* 16, 5764-72.

59. Leulliot, N. & Varani, G. (2001). Current topics in RNA-protein recognition: control of specificity and biological function through induced fit and conformational capture. *Biochemistry* 40, 7947-56.

60. Williamson, J. R. (2000). Induced fit in RNA-protein recognition. *Nat Struct Biol* 7, 834-7.

61. Dasgupta, S., Iyer, G. H., Bryant, S. H., Lawrence, C. E. & Bell, J. A. (1997). Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins* 28, 494-514.

62. Sakai, H., Hiasa, H., Iwamoto, K., Horimoto, T., Komano, T. & Godson, G. (1988). Role of the potential secondary structures in phage G4 origin of complementary DNA strand synthesis. *Gene* 71, 323-30.

63. Ohshima, K. & Wells, R. D. (1997). Hairpin formation during DNA synthesis primer realignment in vitro in triplet repeat sequences from human hereditary disease genes. *J Biol Chem* 272, 16798-806.

64. Ohshima, A., Inouye, S. & Inouye, M. (1992). In vivo duplication of genetic elements by the formation of stem-loop DNA without an RNA intermediate. *Proc Natl Acad Sci U S A* 89, 1016-20.

65. Terada, K., Okuhara, E., Kawarada, Y. & Hirose, S. (1991). Demonstration of extrinsic DNA from immune complexes in plasma of a patient with systemic lupus erythematosus. *Biochem Biophys Res Commun* 174, 323-30.

66. Messias, A. C. & Sattler, M. (2004). Structural basis of single-stranded RNA recognition. *Acc Chem Res* 37, 279-87.

67. Nickens, D. G., Patterson, J. T. & Burke, D. H. (2003). Inhibition of HIV-1 reverse transcriptase by RNA aptamers in Escherichia coli. *Rna* 9, 1029-33.

68. Held, D. M., Kissel, J. D., Patterson, J. T., Nickens, D. G. & Burke, D. H. (2006). HIV-1 inactivation by nucleic acid aptamers. *Front Biosci* 11, 89-112.

69. Hermann, T. (2005). Drugs targeting the ribosome. *Curr Opin Struct Biol* 15, 355-66.

70. Shandrick, S., Zhao, Q., Han, Q., Ayida, B. K., Takahashi, M., Winters, G. C., Simonsen, K. B., Vourloumis, D. & Hermann, T. (2004). Monitoring molecular recognition of the ribosomal decoding site. *Angew Chem Int Ed Engl* 43, 3177-82.

71. Han, Q., Zhao, Q., Fish, S., Simonsen, K. B., Vourloumis, D., Froelich, J. M., Wall, D. & Hermann, T. (2005). Molecular recognition by glycoside pseudo base pairs and triples in an apramycin-RNA complex. *Angew Chem Int Ed Engl* 44, 2694-700.

72. Prewitt, S. P. (2003). Structural studies of an antigen-binding fragment (FAB) complexed with single-stranded DNA., University of Missouri-Columbia.

73. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31, 3784-8.

74. Matthews, B. W. (1968). Solvent content of protein crystals. *J Mol Biol* 33, 491-7.

75. Minor, Z. O. W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods in Enzymology* 276, 307-326.

76. Vagin, A. & Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* 30, 1022-1025.

77. Jones, T. A. Z., J Y; Cowan, S W; Kjeldgaard. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallographica. Section A* 47, 110-119.

78. Brunger AT, A. P., Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta crystallographica. Section D* 54, 905-921.

79. Cowtan, P. E. a. K. (2004). Coot: Model-Building Tools for Molecular Graphics. *Acta Crystallographica Section D* 60, 2126-2132.

80. Murshudov, G. N., A.A.Vagin & E.J.Dodson. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Cryst.* D, 240-255.

81. Collaborative Computational Project, N. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallographica Section D* 50, 760-763.

82. Winn, M. D. M., Garib N; Papiz, Miroslav Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods In Enzymology* 374, 300-321.

83. Winn, M. D. I., M N; Murshudov, G N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallographica. Section D* 57, 122-133.

84. Berman, H. M. W., J; Feng, Z; Gilliland, G; Bhat, T N; Weissig, H; Shindyalov, I N; Bourne, P E. (2000). The Protein Data Bank. *Nucleic Acids Research* 28, 235-242.

Chapter 2

Crystallization of recombinant *Haemophilus influenzae* *e* (P4) acid phosphatise

Abstract

Haemophilus influenzae infects the upper respiratory tract of humans and can cause infections of the middle ear, sinuses, and bronchi. The virulence of the pathogen is thought to involve a group of surface-localized macromolecular components that mediate interactions at the host-pathogen interface. One of these components is lipoprotein *e* (P4), which is a class C acid phosphatase and potential vaccine candidate for nontypeable *H. influenzae* infections. This paper reports crystallization of recombinant *e* (P4) and acquisition of a 1.7 Å resolution native X-ray diffraction data set. The space group is $P4_22_12$ with a = 65.6, c = 101.4 Å, one protein molecule per asymmetric unit and 37 % solvent content. This is the first report of crystallization of a class C acid phosphatase.

Keywords: *Haemophilus influenzae*; Lipoprotein *e* (P4); Class C acid phosphatase; Crystallization; X-ray diffraction

Introduction

The bacterium *Haemophilus influenzae* is a Gram-negative facultative anaerobic coccobacillus and common commensal of the upper human respiratory tract. [1] The organism is the etiologic agent of a variety of local and invasive infections in humans. Isolates of *H. influenzae* are separated into two groups based on the presence or absence

of capsular carbohydrate. Encapsulated organisms are distinguished serologically into 7 different serotypes designated a through f, whereas nonencapsulated strains are designated as nontypeable. These latter strains account for a majority of mucosal diseases including infections of the middle ear, sinuses, and bronchi.

A group of surface-localized macromolecules of *H. influenzae*, including six abundant proteins designated P1 to P6 in order of decreasing molecular weight, are thought to be involved in pathogenesis. [2; 3] Among these proteins, the 28 kDa cationic lipoprotein *e* (P4) is one of the best characterized. This protein is an acid phosphatase encoded by the *hel* gene and it is conserved in size and antigenicity within and between strains[4; 5]. Its optimum phosphomonoesterase activity is achieved at pH 5.0 in the presence of divalent copper with arylphosphate substrates.[6] The enzyme is resistant to tartrate, inorganic phosphate, fluoride and *p*-hydroxymercuriphenylsulfonate, but it is inhibited by orthovanadate, molybdate and EDTA. [6] Sequence alignments suggest that *e* (P4) contains the DDDD motif that is characteristic of class C bacterial non-specific acid phosphatases. [7]

The recombinant form of *e* (P4), designated rP4, has been previously expressed in *E. coli*, purified and characterized.[8] Physicochemical characterization showed that rP4 maintained similar features as the wild-type enzyme. [8] Enzymatically inactive mutants of rP4 were recently shown to be potential vaccine candidates for *H. influenzae*. [9]

Crystal structure determination of rP4 is an important step in elucidating the molecular basis for the biochemical roles of *e* (P4). The structure will also assist in the design of rP4 mutants for use in a *H. influenzae* vaccine. Here, we report the crystallization and

preliminary X-ray crystallographic analysis of rP4. To our knowledge, this is the first report of crystallization of a bacterial class C acid phosphatase.

Methods and results

Protein purification

Recombinant *e* (P4) was expressed in *E. coli* strain BL21(DE3) as described previously[8]. The harvested bacteria were pelleted at 5000*g*, resuspended in 10 mM Na_2HPO_4 pH 7.2 and frozen at 253 K in preparation for protein purification. All purification procedures were conducted at 277 K. All chromatography steps were performed with an Akta FPLC using columns purchased from Amersham Biosciences (GE Healthcare).

Frozen cells were thawed, supplemented with the protease inhibitor PMSF (1 mM final concentration) and disrupted with two passes through a French pressure cell at 110 MPa. Cell debris was removed by centrifugation at 5000*g* for 10 min. NaCl was added to the supernatant to a final concentration of 1 M and the mixture was stirred for 1 h. This step was necessary to release cationic rP4 from negatively-charged phospholipid head groups of membranes. The sample was then subjected to a low speed centrifugation step (5000*g*, 10 min.) followed by ultracentrifugation at 184,000*g* for 1 h in order to pellet bacterial membranes. The resulting supernatant contained the majority of acid phosphatase activity. The supernatant from the ultracentrifugation step was dialyzed against 10 mM Na_2HPO_4 pH 7.2, 300 mM NaCl (buffer A) and filtered through a 0.45 μm filter. The sample was loaded onto a 5 mL HiTrap Chelating HP column that had been charged with 2.5 mL of 0.1 M $CuSO_4$ and equilibrated with 25 mL of buffer A.

Bound proteins were eluted with a 150 mL linear gradient of 0 – 50 mM imidazole in buffer A. We note that rP4 does not have a polyhistidine affinity tag and thus the protein presumably binds to the Cu-affinity column via an endogenous metal-binding site. Phosphomonoesterase activities of the eluted fractions were measured using a discontinuous colorimetric assay with *p*-nitrophenylphosphate as the substrate[6]. Highly active fractions were pooled and dialyzed into 50 mM sodium acetate pH 6.0 (buffer B) containing 50 mM NaCl, and loaded onto a 5 mL HiTrap SP Sepharose cation exchange column that had been equilibrated with 25 mL of buffer B. Elution of rP4 was achieved with a 150 mL linear gradient of 0.05 - 2 M NaCl in buffer B.

Results from SDS-PAGE indicated that eluted rP4 fractions exhibited minor but unacceptable levels of contaminating proteins, so the active fractions were pooled and dialyzed back into buffer A for further purification. The metal affinity and cation exchange steps were repeated using the procedures described above except that a shallower linear gradient consisting of 0.05 - 2 M NaCl in 500 mL buffer B was used in the cation exchange step. The shallower gradient allowed resolution of two main protein peaks, which eluted in the range 140 – 380 mM NaCl (Figure 2-1a). Both protein peaks displayed phosphomonoesterase activity, but they differed markedly in the level of protein purity. The larger peak (fractions 8 – 14) clearly had a shoulder visible on the high ionic strength side (Figure 2-1a) and SDS-PAGE confirmed the presence of multiple protein species in these fractions (Figure 2-1b). In contrast, protein samples taken from smaller peak (fractions 15 – 18) showed a single band in SDS-PAGE (Fig. 2-1b). We note that the results shown in Fig. 1 were reproducible, but the relative areas under the two chromatogram peaks varied from preparation to preparation. Fractions 8 – 14 (pool

A) and 15 – 18 (pool B) were pooled separately for crystallization trials. The two pools of protein were dialyzed into buffer B and concentrated to 13 mg/mL using an Amicon ultrafiltration cell with a 10-kDa molecular-weight cutoff. The protein concentration was determined by the Coomassie Plus assay (Pierce).

Mass-spectrometric analysis revealed two components with apparent molecular masses of 28378 Da and 28509 Da, which differ by the mass equivalent of one methionine. The 28378 Da protein was the major component of pool A, and the 28509 Da protein was the major component of pool B. The species with apparent mass of 28509 Da likely corresponds to full-length rp4, which has theoretical molecular mass of 28569 Da, while the 28378 Da protein possibly represents rP4 devoid of the N-terminal Met. These results suggest the possibility of proteolytic degradation of the N-terminal Met during expression and purification.

Crystallization

All crystallization experiments were performed at 295 K using Cryschem 24-well sitting drop plates (Hampton Research) with reservoir volume of 0.75 ml. Drops were formed by mixing equal volumes of the reservoir (2 μl) and protein solutions (2 μl). Commercially available crystal screens were used to identify initial crystallization conditions. Both protein samples yielded small crystals in these initial screens, but the most promising crystals were obtained with pool B protein and Hampton Research Crystal Screen reagent 6 (0.2 M $MgCl_2$, 0.1 M Tris-HCl pH 8.5, 30 % w/v PEG 4000). Upon optimization, the best crystals were grown using pool B protein (13 mg/ml in Buffer B) and reservoir solutions of 0.2 M $MgCl_2$, 0.1 M Tris-HCl pH 8.1 - 8.5, and 28 -

36 % (w/v) PEG 4000. These crystals appeared as rectangular blocks with dimensions 0.06 mm x 0.06 mm x 0.16 mm (Fig. 2-2).

Data collection and processing

The crystals were prepared for cryogenic data collection by soaking them in 0.2 M $MgCl_2$, 0.1 M Tris-HCl pH 8.5, 40 % (w/v) PEG 4000 and 15 % (v/v) PEG 200. The cryoprotected crystals were picked up with Hampton mounting loops and plunged into liquid nitrogen. Diffraction to 2.6 Å resolution was observed using an in-house Cu rotating anode system and autoindexing calculations suggested a primitive tetragonal lattice with unit cell dimensions a = 66 Å, c = 101 Å.

We note that attempts to grow this crystal form using pool A protein resulted in clusters of fused crystals rather than single crystals, and the resulting diffraction patterns could not be reliably indexed. Thus, isolation of the 28509 Da form of rP4 was critical for growth of high quality single crystals.

Native X-ray diffraction data sets were collected at beamline 4.2.2 of the Advanced Light Source at Lawrence Berkeley National Laboratory using a NOIR-1 CCD detector. The best data set consisted of a 90° wedge of data collected with an oscillation angle of 1°, exposure time of 10 s per degree of oscillation, detector distance of 125 mm and detector 2θ angle of zero. The data were integrated with MOSFLM[10] and scaled with SCALA[11]. The data exhibited acceptable processing statistics to 1.7 Å resolution (Table 2-1). The space group is $P4_22_12$ and the refined unit cell parameters are a = 65.6, c = 101.4 Å. Matthews calculations suggested that this crystal form has 1 molecule in the asymmetric unit, 37 % solvent content and a Matthews coefficient of 1.95 $Å^3/Da$[12].

Since the apparent Laue class is 4/mmm, the possibility of merohedral twinning was investigated. The cumulative intensity distribution for acentric reflections did not display the sigmoidal shape that is characteristic of twinned data[13]. The average value of $\langle I^2(h) \rangle / \langle I(h) \rangle^2$ was 2.0 for acentric reflections, which equals the expected value for nontwinned data. Note that the expected value of $\langle I^2(h) \rangle / \langle I(h) \rangle^2$ for the case of perfect hemihedral twinning is 1.5.[14] Thus, difficulties due to twinning are not anticipated.

Analysis of the rP4 amino acid sequence with BLAST [15]failed to detect a homolog in the Protein Data Bank [16]that could serve as a suitable search model for molecular replacement phasing. Therefore, the structure of rP4 is being determined with single isomorphous replacement with anomalous scattering.

Acknowledgements

This research was supported by National Institutes of Health grant U54 AI057160 to the Midwest Regional Center of Excellence for Biodefense and Emerging Infectious Diseases Research (MRCE, to J.J.T and T.J.R.) and the University of Missouri Research Board (to J.J.T and T.J.R.). The ALS is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory.

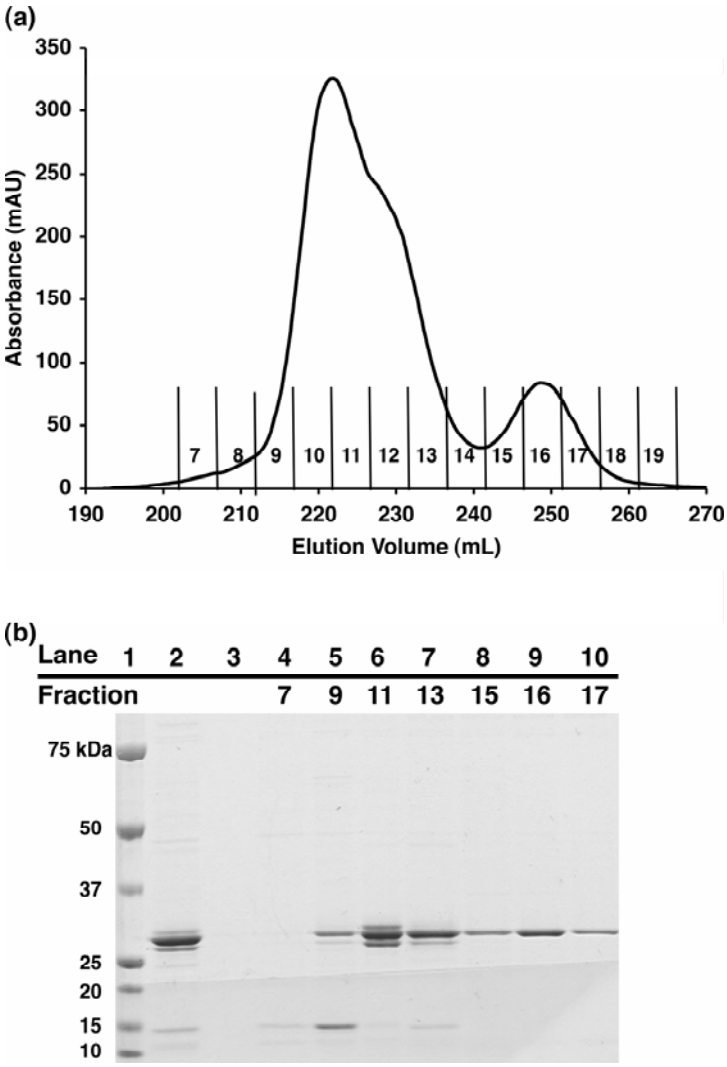


Figure 2-1 Purification of rP4 by cation exchange chromatography. (a) The chromatogram is plot of absorbance (λ = 280 nm) as a function of elution volume obtained with a HiTrap SP Sepharose cation exchange column and Akta FPLC protein purification system. Fractions eluted by a linear NaCl gradient are labelled 7 – 19, with fractions 7 and 18 corresponding to NaCl concentrations of 140 mM and 380 mM, respectively. (b) SDS-PAGE analysis (coomassie stain) of the chromatogram in panel (a). Lanes 1, 2 and 3 correspond to molecular weight marker, protein prior to this cation exchange chromatography step, and flow-through, respectively. Lanes 4 – 10 correspond to elution fractions as indicated above the gel.

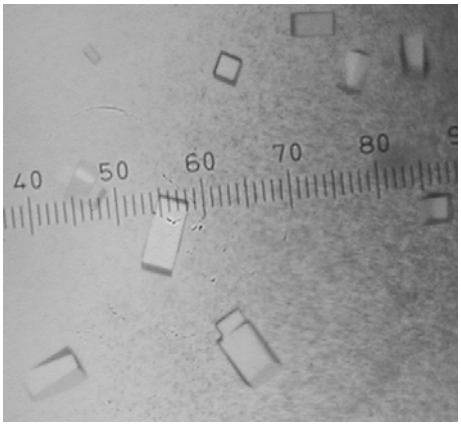


Figure 2-2 Crystals of *Haemophilus influenzae* rP4. The smallest division of the ruler corresponds to 0.02 mm.

Table 2-1

Data collection and processing statistics.

Values for the outer resolution shell of data are given in parentheses.

Beamline	ALS 4.2.2
Wavelength (Å)	1.0359
Space group	$P4_22_12$
Unit cell dimensions (Å)	a = 65.6, c = 101.4
Resolution limits (Å)	42 - 1.70 (1.79 - 1.70)
Number of observations	158442
Unique reflections	25076
Average redundancy	6.3 (3.6)
Completeness (%)	99.8 (98.3)
Average I/σ (I)	17.5 (2.6)
$R_{merge}(I)$	0.071 (0.410)

1. Rao, V. K., Krasan, G. P., Hendrixson, D. R., Dawid, S. & St Geme, J. W., 3rd. (1999). Molecular determinants of the pathogenesis of disease due to non-typable Haemophilus influenzae. *FEMS Microbiol Rev* 23, 99-129.

2. Loeb, M. R. & Smith, D. H. (1980). Outer membrane protein composition in disease isolates of Haemophilus influenzae: pathogenic and epidemiological implications. *Infect Immun* 30, 709-17.

3. Vachon, V., Lyew, D. J. & Coulton, J. W. (1985). Transmembrane permeability channels across the outer membrane of Haemophilus influenzae type b. *J Bacteriol* 162, 918-24.

4. Granoff, D. M. & Munson, R. S., Jr. (1986). Prospects for prevention of Haemophilus influenzae type b disease by immunization. *J Infect Dis* 153, 448-61.

5. Green, B. A., Farley, J. E., Quinn-Dey, T., Deich, R. A. & Zlotnick, G. W. (1991). The e (P4) outer membrane protein of Haemophilus influenzae: biologic activity of anti-e serum and cloning and sequencing of the structural gene. *Infect Immun* 59, 3191-8.

6. Reilly, T. J., Chance, D. L. & Smith, A. L. (1999). Outer membrane lipoprotein e (P4) of Haemophilus influenzae is a novel phosphomonoesterase. *J Bacteriol* 181, 6797-805.

7. Rossolini, G. M., Schippa, S., Riccio, M. L., Berlutti, F., Macaskie, L. E. & Thaller, M. C. (1998). Bacterial nonspecific acid phosphohydrolases: physiology, evolution and use as tools in microbial biotechnology. *Cell Mol Life Sci* 54, 833-50.

8. Reilly, T. J. & Smith, A. L. (1999). Purification and characterization of a recombinant Haemophilus influenzae outer membrane phosphomonoesterase e (P4). *Protein Expr Purif* 17, 401-9.

9. Green, B. A., Baranyi, E., Reilly, T. J., Smith, A. L. & Zlotnick, G. W. (2005). Certain site-directed, nonenzymatically active mutants of the Haemophilus influenzae P4 lipoprotein are able to elicit bactericidal antibodies. *Infect Immun* 73, 4454-7.

10. Leslie, A. G. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* 62, 48-57.

11. Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62, 72-82.

12. Matthews, B. W. (1968). Solvent content of protein crystals. *J Mol Biol* 33, 491-7.

13. Rees, D. C. (1980). The influence of Twinning by Merohedry on Intensity Statistics. *Acta Cryst.* A, 578-581.

14. Redinbo, M. R. & Yeates, T. O. (1993). Structure determination of plastocyanin from a specimen with a hemihedral twinning fraction of one-half. *Acta Crystallogr D Biol Crystallogr* 49, 375-80.

15. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol* 215, 403-10.

16. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res* 28, 235-42.

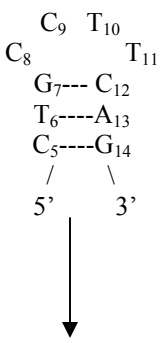
3. Appendix

Table 3-1. Energy Provided by Base-paring and Base-stacking in the DNA.

Interactions	**ΔG^o (298 K, kcal/mol)**
G • C base pair	-2.39
A • T base pair	-0.24
base-stacking ↑ T----A ↓ C----G	-9.68
base-stacking ↑ G----C ↓ T----A	-10.37

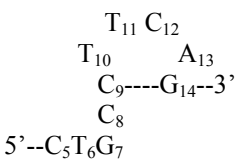
Van Holde et al . Principles of Physical Biochemistry

a



G 5-14 alone:
Energy provided by base pairs: (-2.39)×2+ (-0.24)= -5.02 (kcal/mol)
Energy provided by base stacking: (-9.68)+(-10.37)=-20.05 (kcal/mol)
Total stabilization energy: (-5.02)+(-20.05)=-25.07(kcal/mol)

b



G5-14 in the DNA-1/G5-14 complex:
Energy provided by the base pair: -2.39 (kcal/mol)

Energy difference between State a DNA and State b DNA:

ΔG^o=(-2.39)-(-25.07)=22.68 (kcal/mol)

Energy difference upon protein binding:

$\Delta G^o=\Delta H-T\Delta S$=(-15.1)+7.8=-7.3(kcal/mol)

Energy provided by the specific interactions in the DNA-protein interface:
(-7.3)-22.68=-29.98(kcal/mol)

Table 3-2 Dynamics Light Scattering Data from DNA-1.

Solution Identity	RH (nm)	MW (kDa)	CP (nm)	CP/RH (%)	SOS Error
DNA-1	**3.14**	**46.4**	**.2649**	**8.4**	**1.83**
DNA-1/dT3	**3.15**	**46.9**	**.4026**	**12.6**	**2.66**
DNA-1/dT3 pH 5.0	**3.20**	**48.2**	**.3655**	**11.5**	**2.23**
DNA-1/dT3 1.9 M AS*; pH 5.0	**3.71**	**69.9**	**.9054**	**24.3**	**10.1**
DNA-1/dT3 1.9 M AS*; pH 7.0	**3.61**	**65.0**	**.2115**	**5.9**	**8.61**
DNA-1/dT3 NaCl; pH 5.0	**3.66**	**67.5**	**.5158**	**14.1**	**2.75**
DNA-1/dT3 NaCl; pH 7.0	**2.96**	**40.3**	**.1352**	**4.6**	**3.28**

*** Ammonium Sulfate** **Season Prewitt**

Table 3-3 Summary of the truncated or omitted side chains in the DNA-1/G5-14 structure

	Truncated	**Omitted**
Fab 1(excluding CDRs)	L10, L60, L70, L147,L169, L188, L199, H13, H173, H209	L1, L214, H127-H132, H214-H223
LCDR1 (Fab 1)		
LCDR2 (Fab 1)		
LCDR3 (Fab 1)		
HCDR1 (Fab 1)		
HCDR2 (Fab 1)		
HCDR3 (Fab 1)		
Fab 2(excluding CDRs)	A3, A39, A67, A77, A103, A107, A147, A195, A199, A212, B1, B3, B115, B191	A214, B128- B133, B214-B223
LCDR1 (Fab 2)	A24	
LCDR2 (Fab 2)	A56	
LCDR3 (Fab 2)		
HCDR1 (Fab 2)		
HCDR2 (Fab 2)		
HCDR3 (Fab 2)		

Table 3-4 Summary of the RMSD values when two Fabs in the DNA-1/G5-14 structure were superimposed

	RMSD for Main Chain Atoms (Å)	**RMSD for All Atoms (Å)**
Two Fabs	0.716	1.036
Two Fabs plus DNA	1.087	1.207
Two Fabs excluding CDRs	0.746	1.055
DNA Strand M and Stand N	/	0.901
VL/VH	0.429	0.797
CL/CH	0.546	1.012
L1	0.280	0.652
L2	0.158	1.022
L3	0.192	0.649
H1	0.081	0.152
H2	0.205	1.074
H3	0.146	0.225

* The high RMSDs of L2 and H2 all atoms are due to motions of long side chains.

Table 3-5 Summary of the RMSD values when DNA-1/G5-14 structure CDRs were superimposed with the ligand-free DNA-1 $P6_5$ structure CDRs:

	RMSD for Main Chain Atoms (Å)	**RMSD for All Atoms (Å)**
L1		
Fab1, Fab1	0.245	0.754
Fab1, Fab2	0.204	1.188
Fab2, Fab1	0.295	0.700
Fab2, Fab2	0.240	0.544
L2		
Fab1, Fab1	0.203	0.773
Fab1, Fab2	0.165	0.891
Fab2, Fab1	0.166	0.461
Fab2, Fab2	0.199	0.955
L3		
Fab1, Fab1	1.000	1.204
Fab1, Fab2	0.974	1.145
Fab2, Fab1	1.052	1.364
Fab2, Fab2	1.027	1.318
H1		
Fab1, Fab1	0.230	0.236
Fab1, Fab2	0.125	0.764
Fab2, Fab1	0.179	0.194
Fab2, Fab2	0.158	0.755
H2		
Fab1, Fab1	0.394	1.165
Fab1, Fab2	0.204	1.075
Fab2, Fab1	0.375	0.838
Fab2, Fab2	0.260	0.760
H3		
Fab1, Fab1	1.188	1.269
Fab1, Fab2	0.996	1.308
Fab2, Fab1	0.936	1.102
Fab2, Fab2	0.999	1.319

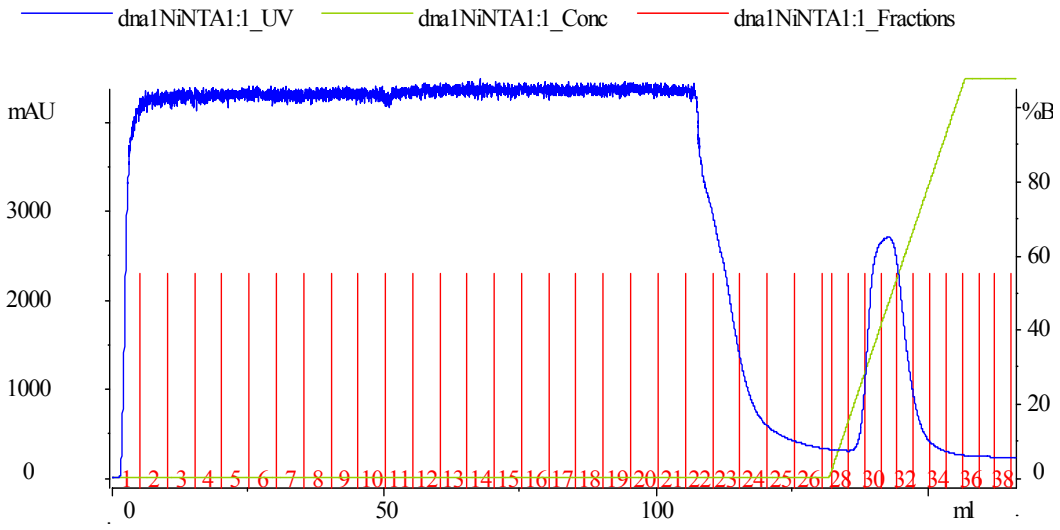


Figure 3-1. Ni-NTA affinity chromatogram for DNA-1. The start buffer consisted of 100 mM Na_2HPO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0. The elution buffer consisted of 100 mM Na_2HPO_4, 300 mM NaCl, 500 mM imidazole, pH 8.0.

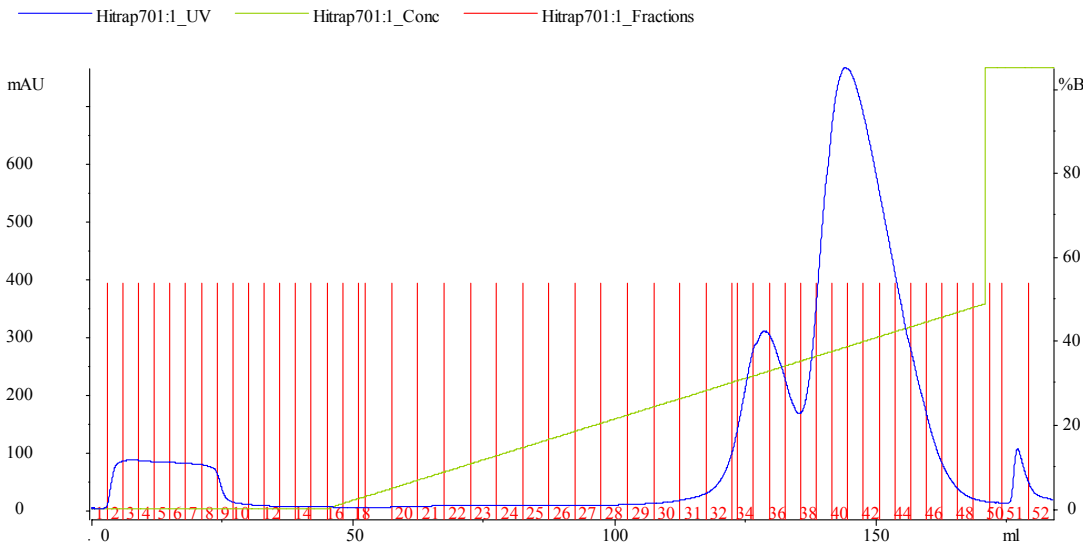


Figure 3-2. Cation exchange chromatogram for DNA-1. The column was a HiTrap SP. The start buffer consisted of 20 mM Na_2HPO_4, pH6.5. DNA-1 eluted in two peaks from a linear gradient consisting of 0-74 mM NaCl in a 100 mL 20 mM Na_2HPO_4 pH 6.5.

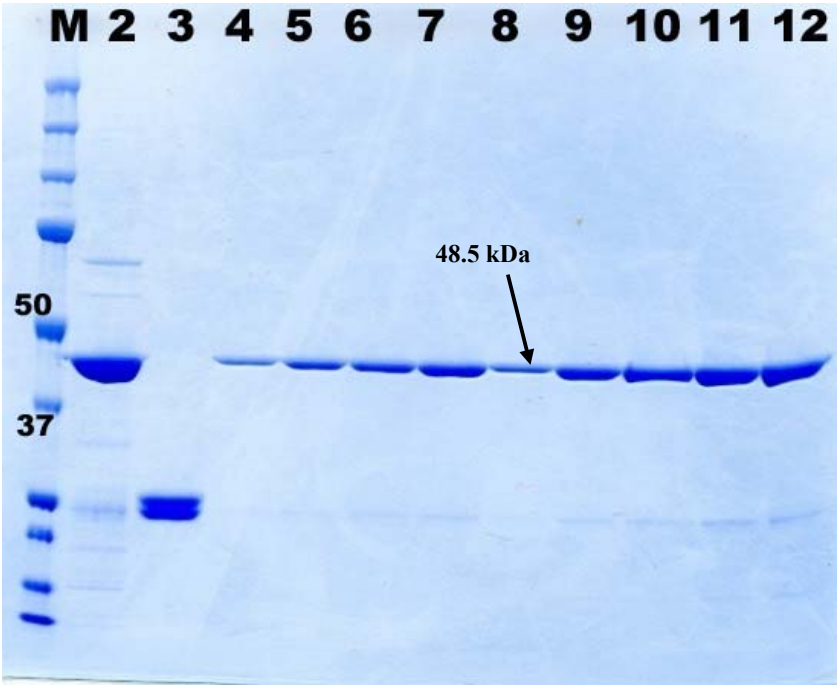
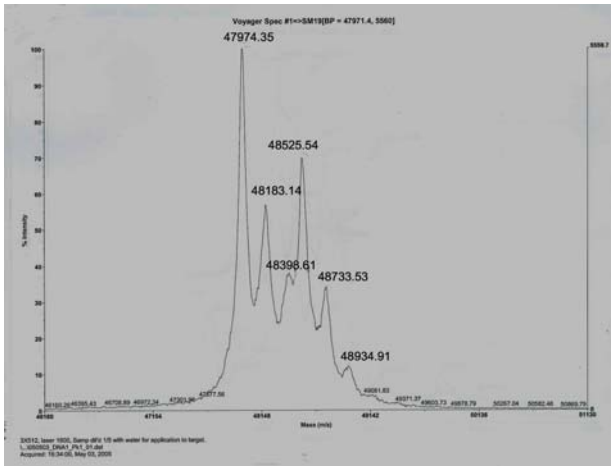


Figure 3-3. SDS-PAGE results of DNA-1 after the two chromatographic steps. The lanes contain protein samples from (1) marker (2) an affinity chromatography elution fraction (3) pure DNA-1+DTT, DTT is at a concentration of 0.05 M. (4-6) protein samples from the first peak of the cation-exchange step (7-12) protein samples from the second peak of the cation-exchange step.

(a)



(b)

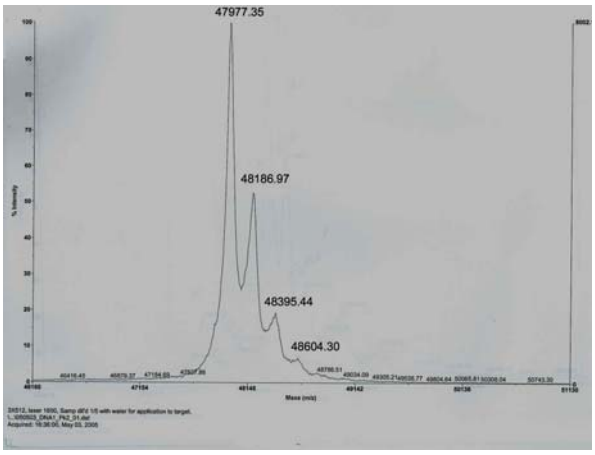
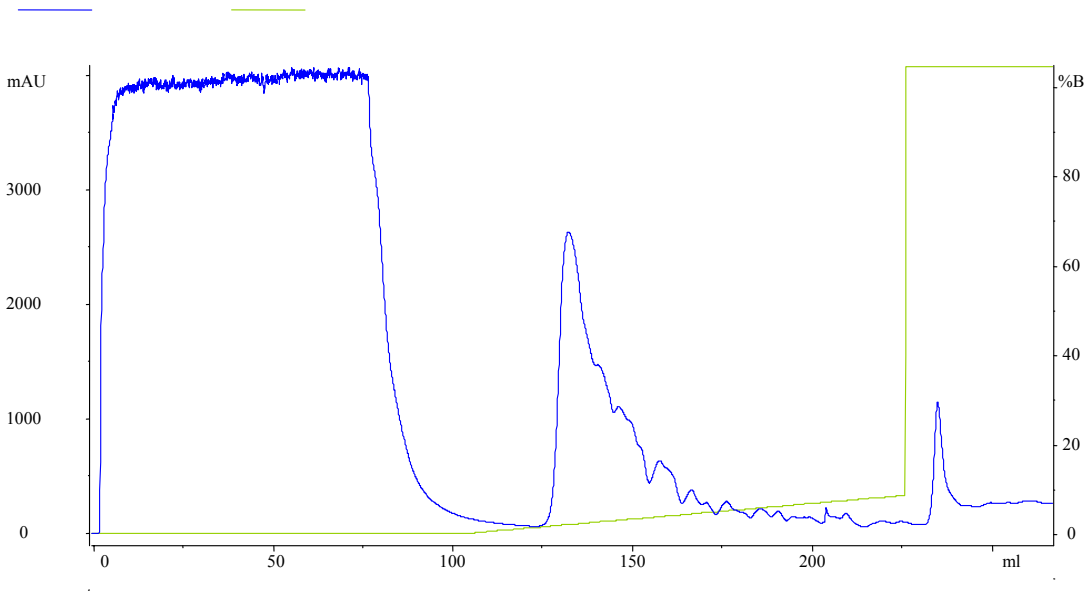
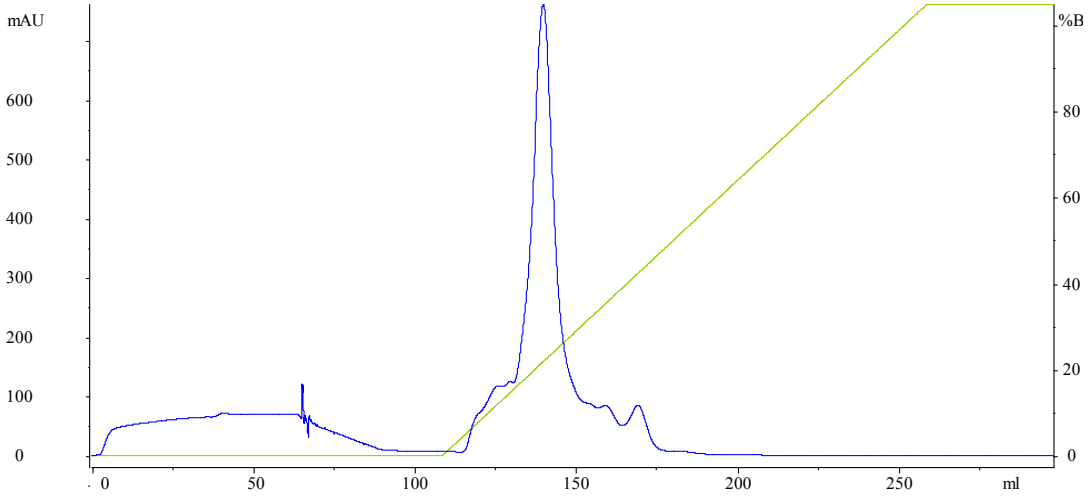


Figure 3-4. MALDI-TOF mass spectroscopy of DNA-1 from protein samples of the two peaks in the cation-exchange step. The expected mass of DNA-1 is 48.5 k Da. (1) Protein from peak 1 contained two species having masses 48 kDa and 48.5 kDa. (2) Protein from peak 2 contained a single species with mass of 48 kDa.

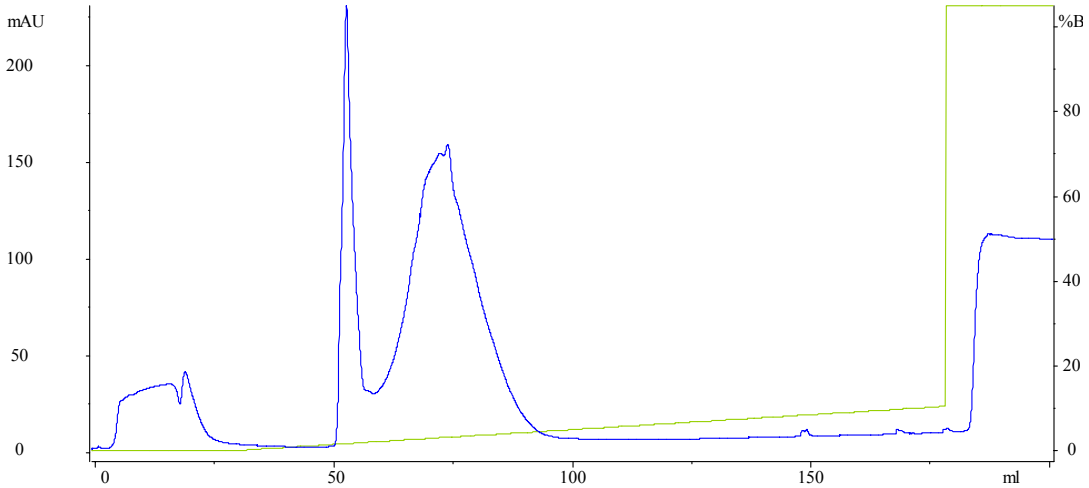
Figure 3-5.Crystals of DNA-1/G5-14 complex. These crystals were in the shape of chunks of bars having dimensions 0.6 mm ×0.2 mm × 0.1 mm. The reservoir solutions contained 0.1 M Tris pH 7.0, 12-26% (w/v) PEG 2000 MME.



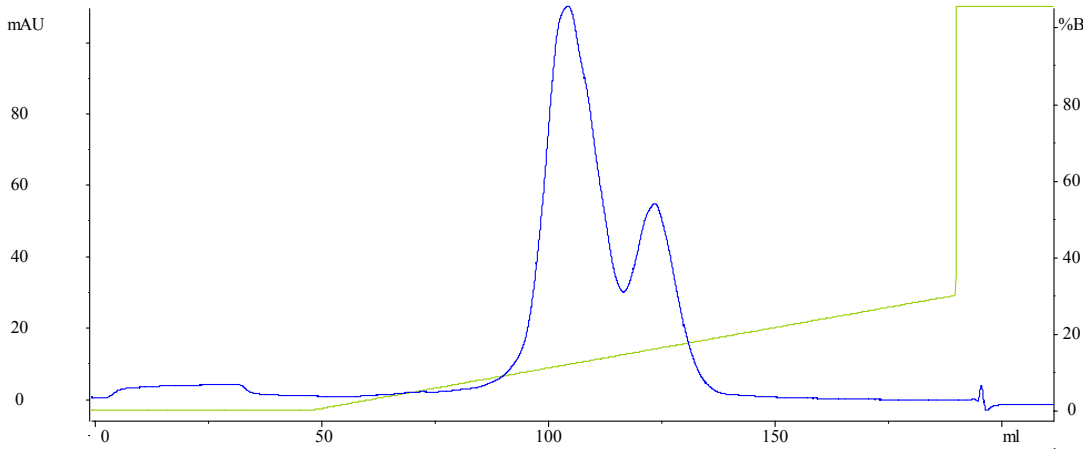
(a) Chromatogram from the first Cu^{2+} chelating column for rP4. The start buffer was 10 mM Na2HPO4 pH 7.2, 300 mM NaCl (buffer A). Bound proteins were eluted with a 150 mL linear gradient of 0 – 50 mM imidazole in buffer A.



(b) Chromatogram from a cation-exchange column in the second chromatographic step for rP4. The start buffer is 50 mM sodium acetate pH 6.0 (buffer B) containing 50 mM NaCl. Elution of rP4 was achieved with a 150 mL linear gradient of 0.05 - 2 M NaCl in buffer B.

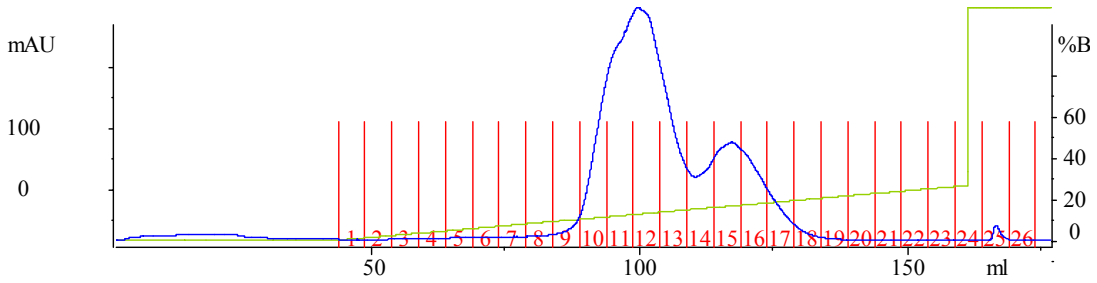


(c) Chromatogram form the third chromatographic step. The same column and conditions in (a) were used again.

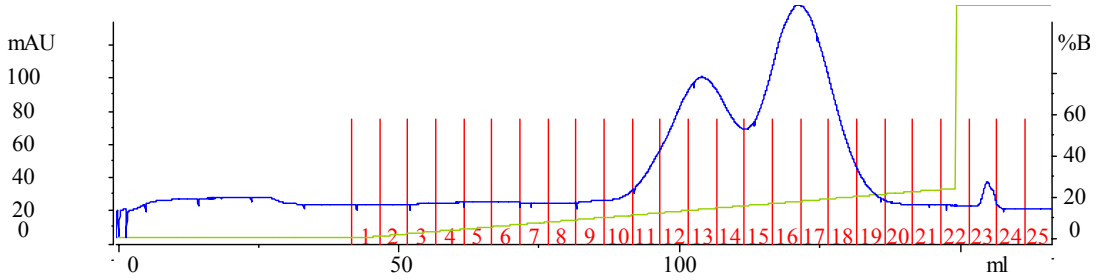


(b) Chromatogram of the last cation-exchange column. The start buffer is 50 mM sodium acetate pH 6.0 (buffer B) containing 50 mM NaCl. Elution of rP4 was achieved with a 500 mL linear gradient of 0.05 - 2 M NaCl in buffer B. The shallower gradient allowed resolution of two main protein peaks, which eluted in the range 140 – 380 mM NaCl.

Figure 3-6 Purification chromatograms for rP4 through four columns.



(a)



(b)

Figure 3-7 Chromatograms obtained from the last cation-exchange step in two different rP4 preparations. The relative areas under the two chromatogram peaks varied from preparation to preparation.

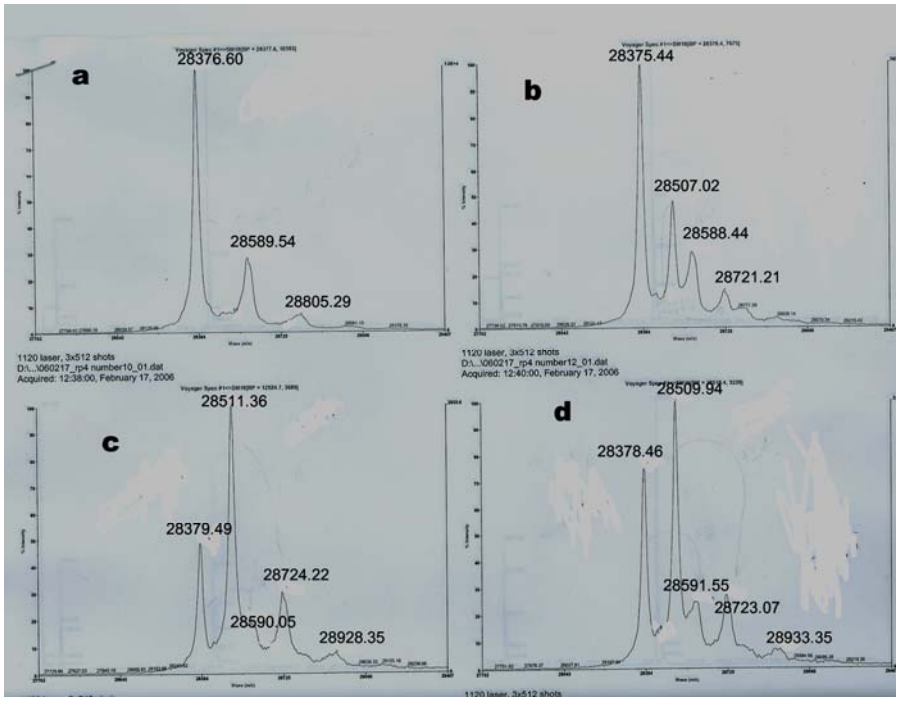


Figure 3-8 MALDI-TOF mass spectroscopy of rP4 in the last cation-exchange run. Figures a, b ,c and d are the results for samples from Figure 3-7(a) tube number 10, 12, 14 and 16 respectively. Mass spectral analysis revealed two components with apparent molecular masses of 28378 Da and 28509 Da.