3D DOMAIN SWAPPING: STRUCTURAL CHARACTERIZATIONS OF DOMAIN-SWAPPED DIMER PROTEINS FVE AND RHODOCETIN

PALASINGAM PAAVENTHAN, M.Sc

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
DEPARTMENT OF BIOLOGICAL SCIENCES
NATIONAL UNIVERSITY OF SINGAPORE
2005

ACKNOWLEDGMENTS

This thesis was only possible because of the support of Dr Prasanna R. Kolatkar, my supervisor, to whom I am indebted not just for his scientific contribution but also for his motivating words, day after day, his help and his friendship.

I thank Professor Hew Choy Leong, Dr Manjunatha Kini and Dr Terje Dokland for their help and advice.

I also wish to thank Dr Howard Robinson for assisting with the data collection. Financial support for data collection comes principally from the National Center for Research Resources of the National Institute of Health, and from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the US Department of Energy. Some of the computation for solving the structures was also performed within Stanford Synchrotron Radiation Laboratory's Collaboratory environment.

In particular I am grateful to Dr Jeremiah S. Joseph for his scientific but also emotional and moral support.

Finally, I would like to pay tribute to the constant support of my family and my friends, without their love over the many months none of this would have been possible and whose sacrifice I can never repay.

LIST OF PUBLICATIONS

1. **Paaventhan, P**., Joseph, J.S., Nirthanan, S., Rajaseger, G., Gopalakrishnakone, P., Kini, M.R. & Kolatkar, P.R. (2003). Crystallization and preliminary X-ray analysis of candoxin, a novel reversible neurotoxin from the Malayan krait Bungarus candidus. Acta Crystallogr D. **59**, 584-586.
2. Seow, S.V., Kuo, I.C., **Paaventhan, P**., Kolatkar, P.R. & Chua, K.Y. (2003). Crystallization and preliminary X-ray crystallographic studies on the fungal immunomodulatory protein Fve from the golden needle mushroom (Flammulina velutipes). Acta Crystallogr D. **59**, 1487-1489.
3. **Paaventhan, P**., Joseph, J.S., Seow, S.V., Vaday, S., Robinson, H., Chua, K.Y. & Kolatkar, P.R. (2003). A 1.7A structure of Fve, a member of the new fungal immunomodulatory protein family. J Mol Biol. **332**, 461-470.
4. **Paaventhan, P**., Kong, C., Joseph, J.S., Chung, M.C.M. & Kolatkar, P.R. Structure of rhodocetin reveals non-covalently bound heterodimer interface. (Submitted).

TABLE OF CONTENTS

***CHAPTER 1**: Introduction* 1

3D DOMAIN SWAPPING 2

Background of 3D domain swapping 2

3D domain swapping definition 4

Helpful definitions 6

History of 3D domain swapping Diphtheria toxin 6

More than one domain swapping 7

Examples of 3D domain swapping 9

Single mutation induce 3D domain swapping 11

Design of 3D domain-swapped molecule 19

Human cystatin C dimerizes through 3D domain swapping 22

Hinge loop role in 3D domain swapping 26

PROTEIN X-RAY CRYSTALLOGRAPHY 28

Protein crystallization 29

Crystal systems and symmetry 29

X-Ray diffraction and Bragg's Law 30

Ewald construction 31

The structure factor 31

Fourier transform and phase problem 33

Model building 35

Refinement 35

AIM AND SCOPE OF THE THESIS 36

***CHAPTER 2**: Structural characterizations of fungal immunomodulatory protein: Fve* 37

MATERIALS AND METHODS 39

Protein purification 39

Protein crystallization and data collection 40

Structure solution and refinement 40

RESULTS AND DISCUSSION 43

Overall fold 43

Topology of FNIII fold in Fve 48

Determinants of the Ig-like fold 51

Structure-function relationships 56

Dimerization by 3D domain swapping 57

***CHAPTER 3**: Structural characterizations of venom of the Malayan pit viper: Rhodocetin* 67

MATERIALS AND METHODS 69

Protein purification 69

Protein crystallization and data collection 69

Structure solution and refinement 70

RESULTS AND DISCUSSION 72

Overall fold 72

Structure comparison with C-type lectin 76

Structure-function relationships 80

Dimerization by 3-D domain swapping 84

***CHAPTER 4**: Conclusion* 88

Bibliography 93

Appendix 112

SUMMARY

Fve, a major fruiting body protein from *Flammulina velutipes*, a mushroom possessing immunomodulatory activity, stimulates lymphocyte mitogenesis, suppresses systemic anaphylaxis reactions and edema, enhances transcription of IL-2, IFN-γ and TNF-α, and hemagglutinates RBCs. It appears to be a lectin with specificity for complex cell surface carbohydrates. Fve is a non-covalently linked homodimer containing no Cys, His and Met. It shares sequence similarity only to the other Fungal Immunomodulatory Proteins (FIPs) LZ-8, Gts, Vvo and Vvl, all of unknown structure. The 1.7 Å structure of Fve solved by Single Anomalous Diffraction of NaBr-soaked crystals is novel: each monomer consists of an N-terminal α-helix followed by a fibronectin III (FNIII) fold. The FNIII fold is the first instance of "pseudo-h-type" topology – a transition between the seven β-stranded s-type and the eight β-stranded h-type topologies. The structure suggests that dimerization, critical for the activity of FIPs, occurs by 3-D domain swapping of the N-terminal helices and is stabilized predominantly by hydrophobic interactions. The structure of Fve is the first in this lectin family, and the first of an FNIII domain-containing protein of fungal origin.

Rhodocetin is a unique heterodimer consisting of α and β subunits of 133 and 129 residues respectively. The molecule, purified from the crude venom of the Malayan pit viper, *Calloselasma rhodostoma*, functions as an inhibitor of collagen induced platelet aggregation. Rhodocetin has been shown to have activity only when present as a dimer. The dimer is formed without an inter-subunit disulfide bridge as observed with all the

other Ca^{2+}- dependent lectin-like proteins (CLPs). The 1.9 Å resolution structure of rhodocetin is determined by molecular replacement. The structure reveals the inter-subunit interface which has compensatory interactions for forming the dimer in the absence of the disulfide bridge. This is the first structure of a CLP without a disulfide connecting the subunits and thus represents a novel molecule which can help to understand a new set of protein-protein interactions. Further, unlike other CLPs, rhodocetin does not require metal ions for its functional activity. However, like other CLPs, rhodocetin also forms the heterodimer by domain swapping, in which the central looped region is swapped.

CHAPTER 1

Introduction

3D DOMAIN SWAPPING

Protein oligomers have evolved because of their advantages over their monomers. These advantages include the possibility of allosteric control, higher local concentration of active sites, larger binding surfaces, new active sites at subunit interfaces, and economic ways to produce large protein interaction networks and molecular machines. However, the mechanisms for the evolution of oligomeric interfaces and for the assembly of oligomers during protein synthesis or refolding remain unclear. Different mechanisms have been proposed for the evolution of protein oligomers, among which is three-dimensional (3D) domain swapping (Liu and Eisenberg, 2002).

Background of 3D domain swapping

Experimentally, the existence of 3D domain swapping was established, and the term introduced, recently, in 1994, when Eisenberg and coworkers observed it for the first time by X-ray crystallography in diphtheria toxin (Bennett *et al.*, 1994). This structure led to a series of elegant theoretical papers by Eisenberg and coworkers that proposed how and why domain swapping might occur and the potential biological implications. However, the concept of 3D domain swapping can be traced back 40 years. Bovine pancreatic ribonuclease (RNase A) forms dimers during lyophilization in acetic acid. Based on elegant chemical modification experiments, Crestfield *et al.*, 1962 proposed that the dimer forms by exchanging the N-terminal fragments (Figure 1.1).This mechanism is essentially identical to what is now called 3D domain swapping.

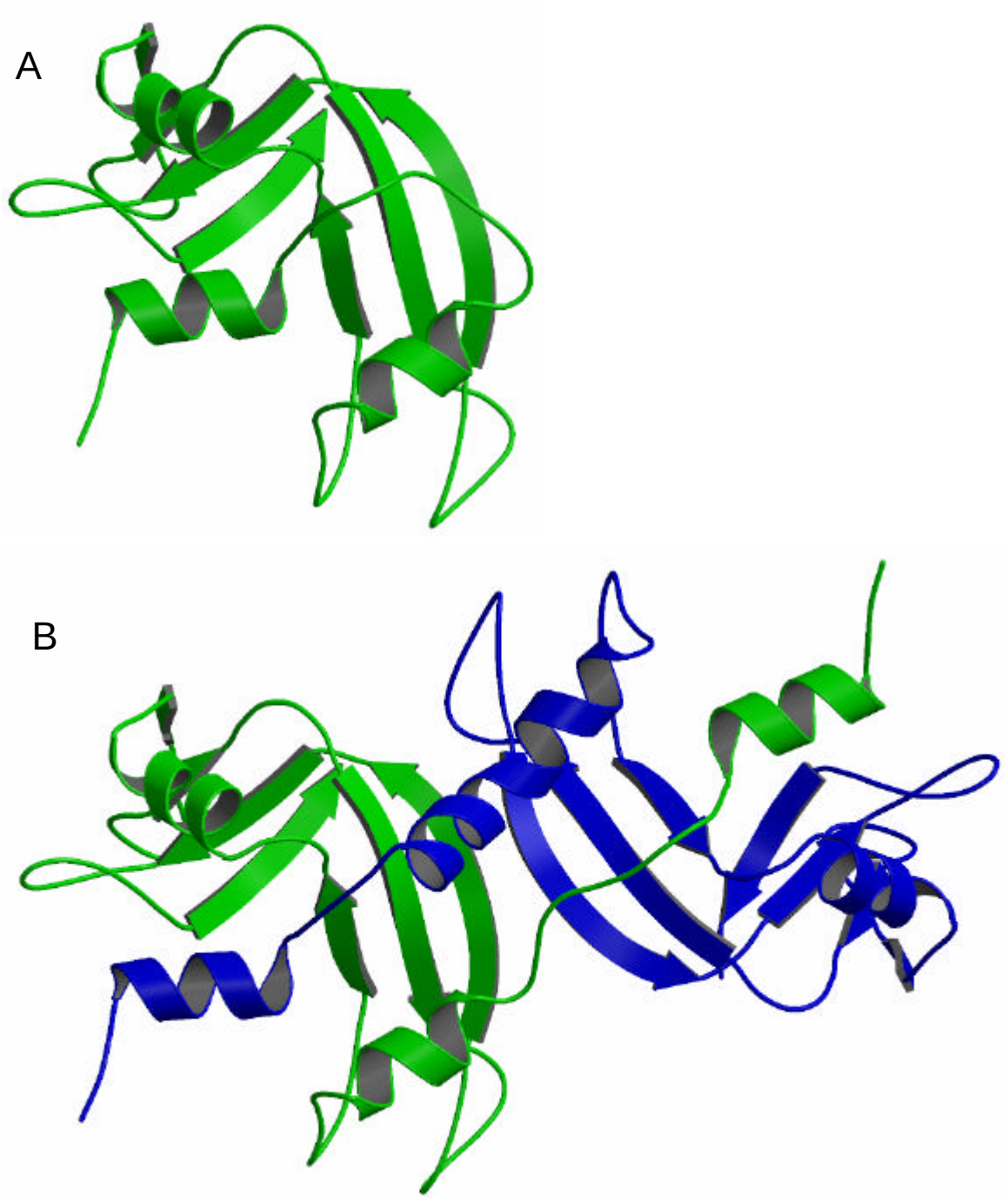


Figure 1.1. (A) RNase A monomer. (B) 3D domain swapped dimer formed by exchanging the N-terminal fragment.

3D domain swapping definition

3D domain swapping is a mechanism for forming oligomeric proteins from their monomers. In 3D domain swapping one domain of a multidomain, monomeric protein is replaced by the same domain from an identical protein chain. The result is an intertwined dimer or higher oligomer, with one domain of each subunit replaced by the identical domain from another subunit. The swapped "domain" can be as large as an entire tertiary, globular domain, or as small as an alpha-helix or a strand of a beta-sheet (Bennett *et al.*, 1994) (Figure 1.2).

Domain-swapped proteins have a C-interface, generally with many specific interactions. The C-interface in the monomeric and dimeric states formed between domains is linked by a hinge loop. The length of the swapped domain varies greatly from a 150-residues globular domain in DT (Diphtheria toxin) to a single α-helix of only 15 residues in BS-Rnase. 3D domain swapping could also create a new open interface in the domain-swapped protein that is not found in the monomeric form (Bennett *et al.*, 1995).

Proteins must have a flexible linker or hinge region to undergo domain swapping. These hinge regions allow conformational changes within a molecule which partially unfold and then find another similar open monomer. Obviously, the hinge region is the only element that has a different structure in the monomeric and 3D domain-swapped forms (Jaskolski *et al*., 2001).

The phenomenon of 3D domain swapping has been studied by examining: (1) *bona fide* 3D domain-swapped proteins, the structures of whose monomeric and oligomeric forms have been characterized by X-ray diffraction or NMR, (2) pairs of proteins whose structures form intertwined 3D domain-swapped oligomers without a

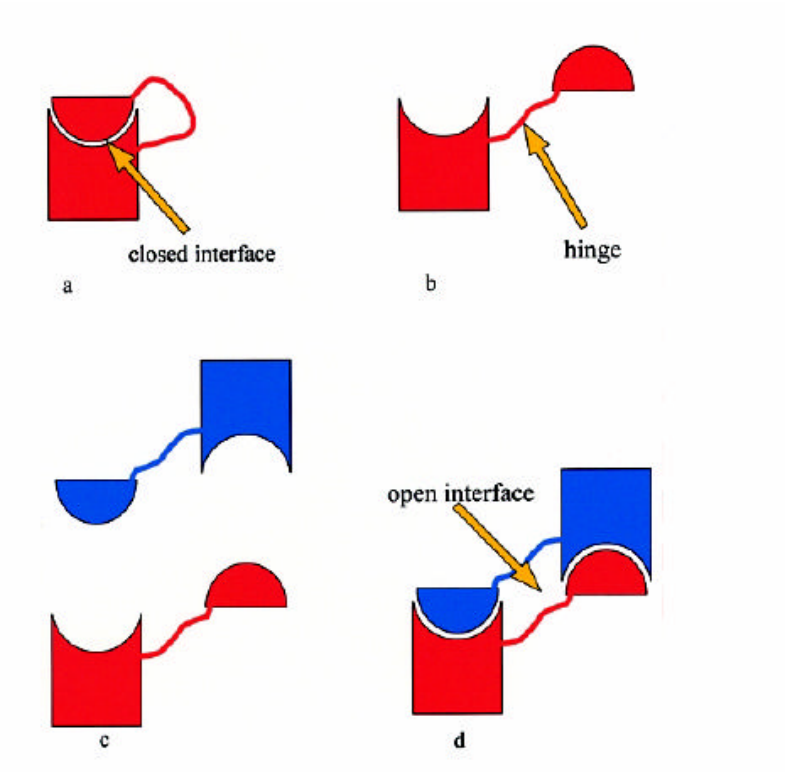


Figure 1.2. Reproduced from Jaskolski *et al.*, 2001. Cartoon illustration of dimer formation *via* 3D domain swapping.

The compact globular fold (a) is partially unfolded (b) through a conformational change at a flexible hinge region. The unfolding temporarily disrupts and exposes the closed interface, i.e. the contact area between the two domains. If sufficiently long-lived, and if present in sufficiently high concentration (c), the unfolded chains will mutually recognize their complementary interfaces and will recreate those contacts in a symmetrical, dimeric fashion (d). Through the closed interfaces, two monomeric folds are reconstructed. However, the dimer is not a simple sum of two monomeric molecules. The hinge regions in the new conformation form a new intermolecular interface that was not present in the monomer. This is the open interface.

known closed monomer. If these proteins have homologs which form a closed monomer, these oligomers considered to be quasi-domain-swapped, and (3) intertwined oligomers that are reminiscent of 3D domain-swapped proteins, but for which no monomeric form is known (Schlunegger *et al*., 1997).

Helpful definitions

***Swapped domain*:** A swapped domain in a protein oligomer is a globular domain that is intertwined with an identical protein chain, with the swapped domain having an environment essentially identical to that of the same domain in a protein.

***Hinge loop*:** A segment of a polypeptide chain that links the swapped domain to the rest of its subunit is a hinge loop.

***C-interface*:** A C-interface occurs between domains in monomeric subunits.

***3D domain-swapped*:** A dimer with a two C-interface between two different subunits is a 3D domain-swapped dimer.

History of 3D domain swapping Diphtheria toxin

The comparison of DT monomer with dimer reveals a mode for protein dimerization which is being called domain swapping. Diphtheria toxin (DT) is a 533-residues protein toxin secreted from a bacterium that causes diphtheria (Collier *et al*., 1975). DT has three domains: a catalytic domain (C), a transmembrane domain (T) and a receptor-binding domain (R) (Bennett *et al*., 1994).

The dimeric form of DT does not form spontaneously even at higher concentrations. However, DT can be induced to form dimers by freezing the protein in phosphate buffer, which causes a drop in pH from the neutral value to 3.6 (Van den berg *et al*., 1959). The decrease in pH converts monomeric DT into open monomers, which form a DT dimer at high concentration by swapping the globular domain (R domain) (Figure 1.3) (Bennett *et al*., 1994).

The structure of monomer DT revealed the mechanism by which low pH could trigger changes in monomer DT and thereby forms an open monomer. The interdomain interface between the R domain and C domain is charged: (1) nine basic and three acidic residues on the R domain interface surface and (2) seven acidic residues on the C domain interface surface. There are three salt bridges stabilizing the interface at neutral pH. The decrease in pH causes protonation of acidic residues and disrupts the salt bridges. Furthermore, buried positive charges in the interface favor the formation of the open monomer. During the dimerization of DT, high oligomers were observed by size-exclusion HPLC. These oligomers include trimers, tetramers and pentamers (Bennett *et al*., 1994).

More than one domain swapping

The most common feature for domain-swapped proteins is that they can exchange either the N-terminal or C-terminal domains. Recent results indicate that RNase A can form two types of dimers where one dimer is more abundant than the other one (Libonati *et al*., 1996 and Gotte *et al*., 1999). The Structures of both dimers revealed that RNase A can have either N-terminal or C-terminal swapped dimers. The less abundant dimer is called

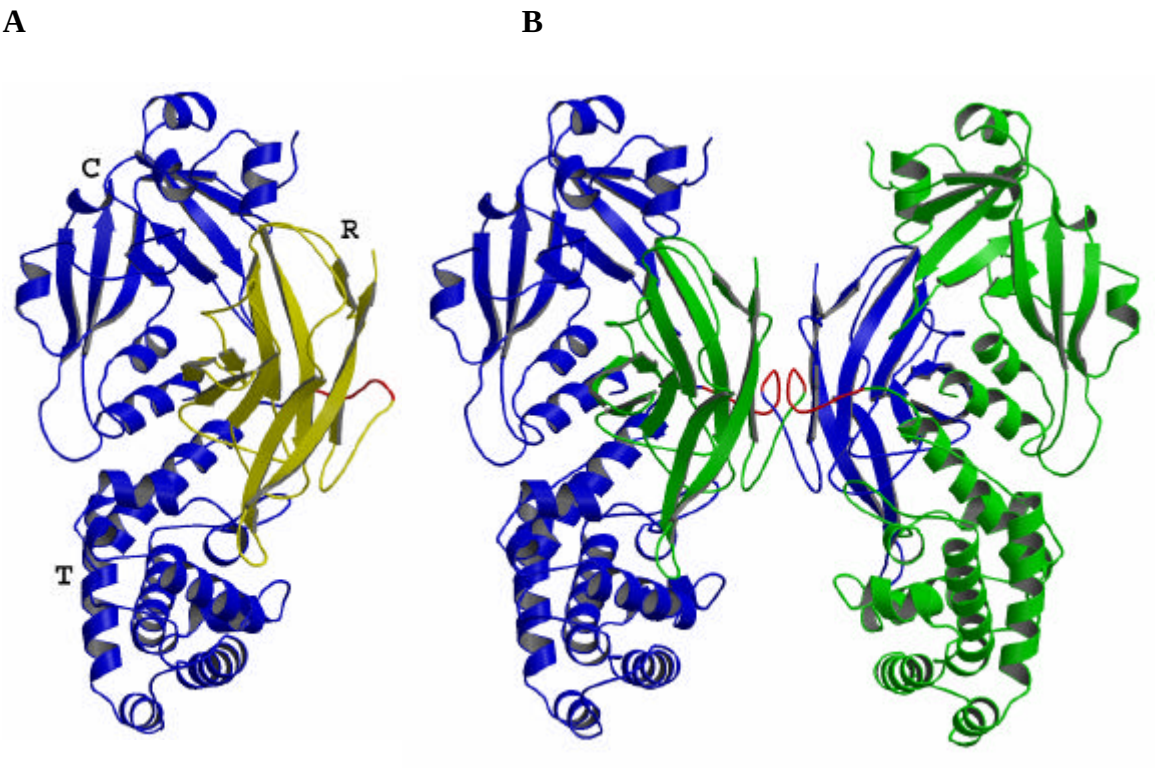


Figure 1.3. (A) DT monomer. The domain that can be swapped is highlighted in yellow. (B) DT dimer. The two subunits are blue and green. The residues having different conformations between monomer and dimer are in the hinge loops (red).

N-terminal swapped dimer (Liu *et al*., 1998) (Figure 1.4B) and the more abundant dimer is called C-terminal swapped dimer (Liu *et al*., 2001) (Figure 1.4C).

Two types of RNase A trimers were also observed: (1) major trimer (Figure 1.4D) and (2) minor trimer (Figure 1.4E) (Liu *et al*., 2002). The major trimer is thought to form a linear trimer that is formed by simultaneously swapping its N-terminal helix with a second molecule and its C-terminal strand with a third molecule, while the minor trimer forms a cyclic trimer by swapping the C-terminal strands only. These hypotheses were initially tested by dissociation studies of trimer RNase A. The dissociation result was consistent with the prediction (Liu *et al*., 2002). Furthermore, the crystal structure of the minor form revealed that the minor trimer is indeed cyclic and 3D domain-swapped at the C-terminal strand (Liu *et al*., 2002).

RNase A can also form tetramers (Gotte *et al*., 1999). The structure of dimers and trimers allows hypothesizing the structure of tetramers. Two of these models are linear: (1) with two C-terminals and one N-terminal portions are swapped and (2) with two N-terminals and one C-terminal portions are swapped. Another model is a combination of cyclic and linear trimers where both types of swapping occur. The last one is a cyclic tetramer where only the C-terminus is swapped (Liu and Eisenberg, 2002).

Examples of 3D domain swapping

Domain-swapped proteins are diverse in sequence, size, function and the way their domains are swapped. These domain-swapped proteins are grouped into three categorizes: (1) *Bona fide*, (2) quasi and (3) candidate for domain swapping (Table 1.1-1.3). In a sequence comparison study of about 40 domain-swapped proteins, Liu and Eisenberg (2002) concluded that a domain swapping protein cannot be predicted based

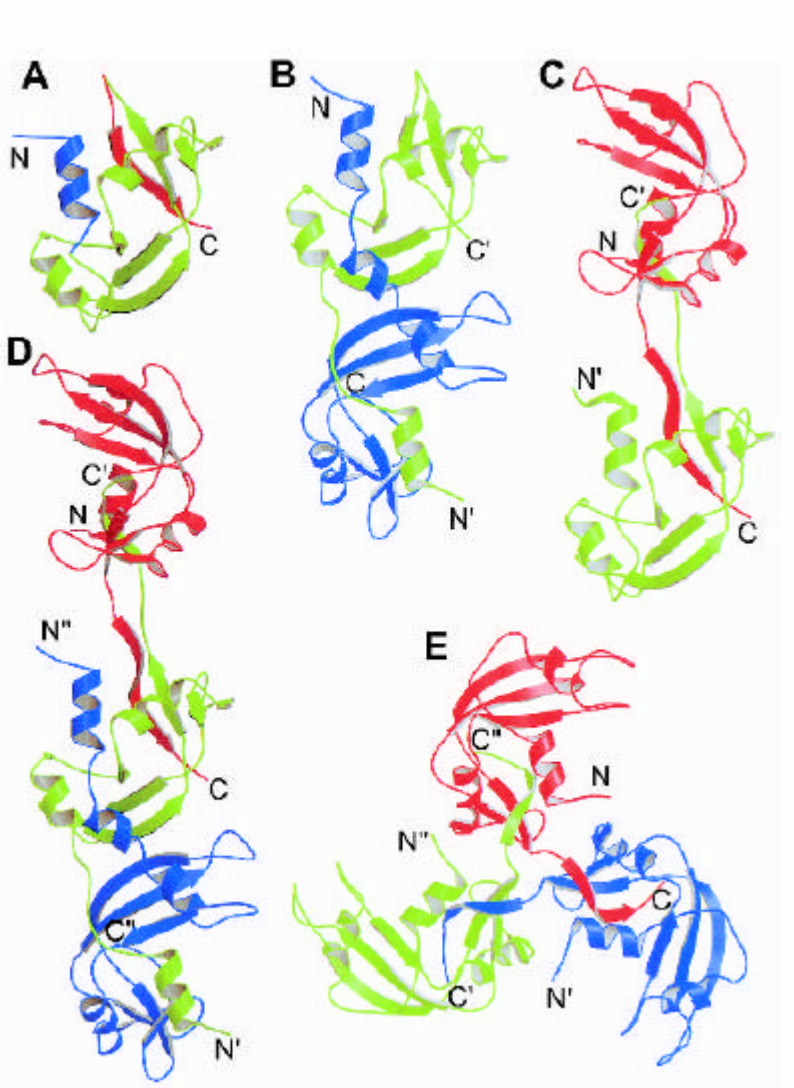


Figure 1.4. Reproduced from Liu and Eisenberg, 2002. Ribbon diagrams of the structures of the RNase A monomer (A), the minor dimer (B), the major dimer (C), the major trimer model (D), and the minor trimer (E) . The N- and C-termini are labeled. The N-terminal helix and C-terminal strand that are swapped in the oligomers are colored in blue and red, respectively, in the monomer (A). Both types of swapping take in the major trimer model (D): The green subunit swaps the C-terminal strand with the red subunit and swaps the N-terminal helix with blue subunit.

on the sequence of the protein. In addition, their studies showed that domain swapped proteins form various interactions at the C-interface such as hydrophobic interactions, hydrogen-bonding, electrostatic interaction including disulfide bridge interactions (Diederichs *et al*., 1991; Milburn *et al*., 1993; Knaus *et al*., 2001).

The swapped domains also have diverse secondary structures as reported by Liu and Eisenberg (2002): Domain-swapped proteins can have a swapped domain with one α-helix (BS-RNase, RNase A N-terminal swapped dimer), one β-strand (CksHs2 dimer, cro dimer), several α-helices (calbindin D_{9k}, barnase), several β-strands (β-B2 crystallin, diphtheria toxin dimer), or a mixture of α-helixes and β-strandes (T7 gp 4 ring helicase, human cystatin C). The observed data shows that 3D domain swapping does not require or prefer certain types of secondary structures (Table 1.1-1.3).

Single mutation induce 3D domain swapping

Manipulation of a protein sequence has given insight into the factors governing 3D domain swapping. The studies of O'Neill *et al*., 2001a showed that conformational strain imposed by mutation in a 64 residue domain of Protein L (Ppl) causes 3D domain swapping.

The 64 residue B1 domain of Protein L (Ppl) from *Peptostreptococcus magnus,* consists of a central α-helix packed against a four stranded β sheet formed by two β hairpins. The pseudo-wild-type Ppl structure (WT*) revealed that the first β turn with residue G15 has a positive ϕ angle while the second β turn contains three residues (D53, K54 and G55) with positive ϕ angles (O'Neill *et al*., 2001b). Positive ϕ angles can induce strain when the C_β atom of a residue makes close contacts with its backbone oxygen.

Table 1.1. Examples of bona fide 3D domain-swapped proteins[a]

Protein (Reference)	PDB code	Number of residues per subunit	Number of residues in swapped domain	Residues in hinge loop	Function	Structure of swapped domain
Barnase monomer (Buckle and Fersht, 1994)	1BRN	110			Ribonuclease	
Barnase trimer (Zegers et al. 1999)	1YVS	110	36	37…41	Ribonuclease	N-terminal helices
Calbindin D9k wild-type monomer (Svensson et al. 1992)	4ICB	76			Transcellular calcium transport and magnesium uptake in the intestine	
Calbindin D9k mutant dimer (Hakansson et al. 2001)	1HT9	76	27	38…47	Unknown	C-terminal helices
Soluble domain CD2 monomer (Jones et al. 1992)[b]	1HNG	177			Binding domain of lymphocyte adhesion protein	
N-terminal domain of CD2 dimer (Murray et al. 1995)[b]	1CDC	99	43	44…50	Unknown	N-terminal α-strands
Insertion mutant cro monomer (Albright et al. 1996)	1ORC	71		55…56d	Unknown	
Wild-type cro dimer (Anderson et al. 1981)[b]	1CRO	66	11	55	DNA repressor	C-terminal β-strand
Cyanovirin-N monomer (Bewley et al. 1998)	2EZM	101			HIV inactivation	
Cyanovirin-N dimer (Yang et al. 1999)	3EZM	101	48	50…53	Unknown	N- or C-terminal globular domain
DT monomer (Bennett and Eisenberg 1994b)[b]	1MDT	535			ADP-ribosylating toxin	
DT dimer (Bennett et al. 1994a)[b]	1DDT	535	148	379…387	Speculative receptor binding	C-terminal globular domain
Human prion monomer (Zahn et al. 2000)	1QLX	108			Unknown	
Human prion dimer (Knaus et al. 2001)	1I4M	108	28	188…198	Unknown	C-terminal helices
Protein L B1 domain monomer (O'Neill et al. 2001)	1HZ5	62			Binding of Ig G	
Protein L B1 domain dimer (Kuhlman et al. 2001)	1JML	61	7	52…55	Unknown	C-terminal β-strand
RNase A monomer (Wlodawer et al. 1982)	5RSA	124			Ribonuclease	
RNase A N-terminal swapped dimer (Liu et al. 1998)	1A2W	124	14	15…22	Ribonuclease	N-terminal α-helix
RNase A C-terminal swapped dimer (Liu et al. 2001)	1F0V	124	9	112…115	Ribonuclease	C-terminal β-strand
RNase A cyclic C-terminal swapped trimer (Liu et al. 2002)	1JS0	124	9	112…115	Ribonuclease	C-terminal β-strand
BS-RNase dimer (Mazzarella et al. 1993)[b]	1BSR	124	14	15…22	Ribonuclease	N-terminal α-helix
Single-chain Fv monomer of antibody NC10 (Malby et al. 1998)	1NMC	246		113…127	Antigen binding	
Diabody (Perisic et al. 1994)	1LMK	243	122 or 116	123…127	Antigen binding	N- or C-terminal globular domain
Triabody (Pei et al. 1997)	1NQB	236	120 or 116	0	Antigen binding	N- or C-terminal globular domain
Phosphorylated N-Spo0A (Lewis et al. 1999)	1QMP	129			Response regulator related to sporulation	
N-domain of Spo0A dimer (Lewis et al. 2000)	1DZ3	129	21	103…109	Response regulator related to sporulation	C-terminal α-helix
Wild-type staphylococcal nuclease monomer (Loll and Lattman 1989)[b]	1SNC	149		112…120	Nuclease	
Deletion mutant of staphylococcal nuclease dimer (Green et al. 1995)	1SND	143	29	112…114	Unknown	C-terminal α-helix
suc1 monomer (Endicott et al. 1995)	1SCE	113			Cell cyle regulation	
suc1 dimer (Khazanovich et al. 1996)[b]	1PUC	113	22	85…91	Cell cycle regulation	C-terminal β-stand
TrkA-d5 and NGF complex (Wiesmann et al. 1999)	1WWW	120			Binding of nerve growth factor	
TrkA-d4 dimer	1WWA	109	13	297…299	Unknown	N-terminal β-strand
TrkC-d5 dimer	1WWB	103	14	299…301		
TrkC-d5 dimer (Ultsch et al. 1999)	1WWC	118	14	317…319		

[a] The oligomers of these proteins are domain-swapped and their monomers adopt a closed conformation.
[b] These entries are taken from the previous review (Schlunegger et al. 1997).

Table 1.2. Examples of proteins that exhibit quasi-domain swapping[a]

Protein (Reference)	PDB code	Number of residues per subunit	Number of residues in swapped domain	Residues in hinge loop	Function	Structure of swapped domain
CksHs1 monomer (Arvai et al. 1995)[b]	1DKS	79			Cell cycle regulation	
CksHs2 dimer (Parge et al. 1993)[b]	1CKS	79	14	60…65	Cell cycle regulation	C-terminal -strand
γ-crystallinII monomer (Blundell et al. 1981)	4GCR	174			Eye lens protein	
βB2-crystallindimer (Bax et al. 1990)	1BLB	204	97	79…87	Eye lens protein	N- or C-terminal globular domain
Chicken cystatin monomer (Bode et al. 1988)	1CEW	108			Cysteine protease inhibitor	
Human cystatin C dimer (Janowski et al. 2001)	1G96	120	54	55…59	Unknown	N-terminal helix and strand
E. coli glyoxalase I dimer (He et al. 2000)	1FA5	135			Interconversion of glutathione thiohemiacetal of methylglyoxal and S-D-lactoyl-glutathione	
Human glyoxalase I dimer (Cameron et al. 1997)	1BH5	183	19	20…32	Interconversion of glutathione thiohemiacetal of methylglyoxal and S-D-lactoyl-glutathione	N-terminal α-helix
GM-CSF monomer (Diederichs et al. 1991)[b]	1GMF	127		87…99	Granulocyte macrophage growth factor	
IL-5 dimer (Milburn et al. 1993)[b]	1HUL	113	26	82…89	B and T cell growth factor	C-terminal strand and helix
IFN- βmonomer(Senda et al. 1992)[b]	1RM1	160		97…114	Fibroblast interferon	
IL-10 dimer (Zdanov et al. 1995)[b]	1ILK	160	46	108…118	Cytokine inhibitory synthesis factor	C-terminal helices
Mannose binding protein (Weis et al. 1991)	1MSB	115			Mannose binding	
IX/X-binding protein (Mizuno et al. 1997)	1IXX	Heterodimer 129 and 123	17	Loop1: 72…75 Loop2: 93…98	Anticoagulation	Middle loop
Major urinary protein monomer (Bocskei et al. 1992)[b]	1MUP	166		126…130	Rodent pheromone transporter	
Odorant binding protein dimer (Tegoni et al. 1996)[b]	1OBP	159	35	121…124	Odorant binding and transport	C-terminal stand and helix
Human pancreatic ribonuclease chimera (Canals et al. 2001)	1H8X	128	15	16…23	RNA digestion (homolog of RNase A and BS-RNase)	N-terminal α-helix
Grb2 adaptor (SH2 + SH3) (Maignan et al. 1995)	1GR1	217			Signal transduction	
Grb2-SH2 domain dimer (Schiering et al. 2000)	1FYR	93	27	121…123	Binding phosphorylated peptide	C-terminal α-helix
Fyn-SH3 monomer (Musacchio et al. 1994)[b]	1FYN	62		112…118	Signal transduction	
SH3 domain of Eps8 (Kishan et al. 1997)	1AOJ	65	26	34…39	Recognition of proline-rich sequences	C-terminal β-strands
SigE dimer (Luo et al. 2001)	1K3S	113			Molecular chaperone in type III secretion system	
CesT dimer (Luo et al. 2001)	1K3E	156	32	33…36	Molecular chaperone in type III secretion system	N-terminal α-helix and β-strand
Monomer of two repeats of α-spectrin(Grum et al. 1999)	1CUN	211			Cytoskeletal protein	
Dimer of one repeat of α-spectrin (Yan et al. 1993)[b]	2SPC	107	32	72…75	Cytoskeletal protein	C-terminal α-helix

[a] These proteins show oligomeric structures with 3D domain swapping, and their homologs adopt the closed monomeric conformation.
[b] These entries are taken from the previous review (Schlunegger et al. 1997).

Table 1 . 3 . Candidates for 3D domain swapping[a]

Protein (Reference)	PDB code	Number of residues per subunit	Number of residues in swapped domain	Residues in hinge loop	Function	Structure of swapped domain
Bleomycin resistance protein dimer (Dumas et al. 1994)	1BYL	122	8	9…11 (?)	Bleomycin resistance	N-terminal β-strand
BTB domain of PLZF (Ahmad et al. 1998)	1BUO	120	12	13…23 (?)	Evolutionarily conserved protein-protein interaction motif	N-terminal β-strand
Cab-type β class carbonic anhydrase (Strop et al. 2001)	1G5C	170	12	13…23 (?)	Catalyzing reversible hydration of CO_2	N-terminal α-helix
Catalase dimer (beef liver) (Fita and Rossmann 1985)	7CAT	500	65	66…70 (?)	H_2O_2 hydrolysis	N-terminal helices
Citrate synthase dimer from chicken heart (Remington et al. 1982)	1CTS	433	10	417…423 (?)	Citrate synthesis	C-terminal helix
Designed coiled coil dimer (Ogihara et al. 2001)	1G6U	48	14	33…34 (?)	Unknown	C-terminal α-helix
dUTPase trimer (Larsson et al. 1996)	1DUD	136	10	125…126 (?)	dUTP hydrolysis	C-terminal strand
Heat shock protein 33 dimer (Vijayalakshmi et al. 2001)	1HW7	255	75	178…184 (?)	Molecular chaperone	C-terminal helices and strands
Phosphoenolpyruvate mutase dimer (Huang et al. 1999)	1PGM	295	35	240…260 (?)	Converting phosphoenolpyruvate to phosphonopyruvate	C-terminal helices
T4 endonuclease VII dimer (Raaijmakers et al. 1999)	1EN7	157	60	63…72 (?)	Mismatch repair, Resolving branchpoint before phage package	N-terminal helix and strands
T7 gene 4-ring helicase hexamer fragment (Singleton et al. 2000)	1E0J	326	40	283…305 (?)	Separation of nucleic acid duplexes into strands	N-terminal α-helix and -strand
T7 gene 4-ring helicase fragment (Sawaya et al. 1999)	1CR0	296	12	283…305 (?)	Unknown	N-terminal α-helix
RecA hexamer from *E. coli* (Story et al. 1992)	2REB	352	26	27…39 (?)	DNA recombination and repair	N-terminal α-helix
Simian virus 40 oligomer (Stehle et al. 1996)	1SVA	361	61	296…300 (?)	Virus coat protein	C-terminal helix and strands

[a] These proteins show oligomeric structures with exchanging domains, but have not been shown to have monomers or monomeric homologs with a similar structure. Therefore, the residues forming hinge loops in these proteins are speculative, and are indicated here with a question mark.

Table1.1-1.3. Reproduced from Liu and Eisenberg, 2002.

Structures of four mutants (V94A, G55A, K54G and G15A) were determined to study the effect of β turn mutants G55A, K54G and G15A, as well as core mutant V94A (O'Neill *et al*., 2001a).

Core mutant V94A. The structure of Ppl (V94A) was determined to a resolution of 1.8 Å. The structure revealed that the asymmetric unit contained two monomers and one β-strand swapped dimer (Figure 1.5). While the mutant monomers maintain the wild type fold, the molecules B and D form a dimer through the exchange of their N-terminal strands (O'Neill *et al*., 2001a). In the V94A monomer, residues D53, K54 and G55 retained the strain as in the WT*, however all the strains are released in one of the dimeric molecule (B) (Figure 1.5).

The hinge region (52-55) of the dimeric molecule B begins its rotation at A52 and completes its rotation at G55, while the hinge loop of the molecule D began it rotation at K54 and completed its rotation at G55. This back bone rotation allowed D53 of B to take a negative ϕ angle while the molecule D retained the positive ϕ angle of D53.

The V94A mutation in Ppl was not expected to have a significant effect on the overall fold because the mutation was in the core region of the protein. However, structural determination of the protein revealed that V94A mutation formed an intertwined dimer by exchanging the N-terminal strand (O'Neill *et al*., 2001a). Furthermore, the structure of the domain-swapped dimer helps to uncover the domain swapping mechanism. In the second β-turn in Ppl, there are three consecutive positive ϕ angles, in which two are non-glycines. Generally, non-glycine amino acids with positive ϕ angles are energetically unfavorable. Because of this conformational strain, the N-terminal β-strand

Figure 1.5. The asymmetric unit contains two monomers (cyan and brown) and a domain-swapped dimer (orange and magenta).

springs open and releases the strain when the interaction is weakened in the close interface (O'Neill *et al*., 2001a).

***Mutant G55A*.** The V49A mutant structure suggests that there are two opposing components of free energy: (1) the free energy component that favors domain swapping as deserved in the strain of the second β-turn and (2) the free energy component that disfavors domain swapping as in the loss of entrophy upon dimer formation. Based on this observation, a domain-swapped dimer was predicted in G55A mutant because the mutation would increase the strain in the second β-turn.

The structure of G55A mutant shows that the fourth β-strand is swapped in a similar manner as in V49A (Figure 1.6). All the angles in the hinge region (53-56) have a negative ϕ value in mutant G55A, unlike molecule D of V49A mutant where it retains the positive ϕ angle of D53. The C_β atom of A55 and its symmetry mate in the dimer are 3.9 Å apart and tilted toward each other to make stabilizing van der Walls contacts. Furthermore, the C_β atom of A55 does not clash with its backbone carbonyl oxygen (O'Neill *et al*., 2001a).

***Mutant K54G*.** The K54G mutant was crystallized to see if relieving the strain caused by the clash between the C_β atom and backbone carbonyl oxygen in K54 would have any effect on the conformation of the second β-turn. The structure revealed little main chain movement in the second β-turn. This confirmed that stability of the protein was increased by eliminating the clash in K54. The ϕ angle for D53, G54 and G55 remained positive. This suggests that increased stability resulted due to the removal of the close contact

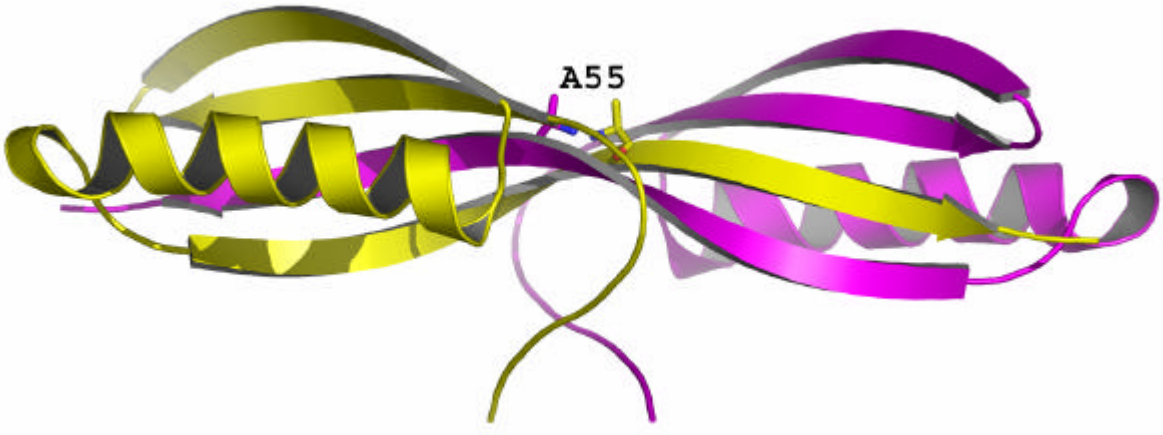


Figure 1.6. The G55A domain-swapped dimer. The C_β atom of A55 and its symmetry mate in the dimer are 3.9 Å apart.

between the C_β atom and backbone carbonyl oxygen in K54 (O'Neill *et al*., 2001a).

***Mutant G15A*.** A mutation was introduced in the first β-turn to test whether increasing the strain in a β-turn would be sufficient to induce domain swapping. Unlike the second β-turn, this β-turn contains only one residue, G15, with the positive ϕ angle. The mutant structure was seen to be a monomer. Furthermore, the A15 residue retained the positive ϕ angle but now is in the allowed region of Ramachandran plot. This observation suggests that strain alone in the β-turn is not sufficient to induce domain swapping (O'Neill *et al*., 2001a).

Free energy estimation shows that the mutant has a higher transition-state free-energy barrier because the cost of removing the first β-strand from the β-sheet is considerably higher. Moreover, the free energy for domain swapping is higher for the first β-strand compared to the fourth β-strand since the energy gain from the release of strain in the second β-turn is much more higher than the strain release in the first β-turn (O'Neill *et al*., 2001a).

Design of 3D domain-swapped molecule

The several designed dimeric protein molecules and their well studied motifs have given us knowledge to design protein structures of open and closed monomers. These designed molecules can lead to a domain-swapped dimer or to domain-swapped oligomers depending on topology (Ogihara *et al*., 2001).

To study the designing of domain-swapped entities, monomeric 3-helix bundles

were utilized. The bundles are variants of designed 3-α-helical bundle called coil-Ser, whose design was based on the heptad repeat sequence of an α-helical coiled coil (O'Neil and DeGrado, 1990 and Lovejoy *et al*., 1993). The coil-Ser α-helical bundle has an antiparellel packing arrangement, in which each α-helix is made of four heptad repeats. The structure of coil-Ser served as a template for designing monomeric 3-helix bundles with up and down topology (Bryson *et al*., 1998 and Walsh *et al*., 1999). In these designs, the helices were shortened to a length of three heptads, and interhelical electrostatic interactions (Lumb & Kim, 1995) were used to define a unique topology. To simplify the study of the domain swapping, the helices were further shortened to a length of two heptads, resulting in a three helix bundle with 14 residues per helix. The helices were connected by loops to provide two different topologies of helix bundles: (1) up-down-up topology (Mon1) and (2) up-down-down topology (Mon2) (Ogihara *et al*., 2001) (Figure 1.7). Loop deletion is a common mechanism for 3D domain swapping (Green *et al*., 1995, Dickason and Huston, 1996 and Pei *et al*., 1997). Therefore, deleting the second loop from the helix bundle would lead to an open monomer that could form a domain-swapped dimer.

Mon1 was crystallized and its structure was determined. The structure revealed that Mon1 formed a domain-swapped dimer. Residues 1a-14a, 19a-48a (monomer A), 1b-13b, 19b-48b (monomer B) are α-helical. The hairpin loop residues 15a-18a and 14b-18b form turns connecting domain I and domain II of each monomer (Ogihara *et al*., 2001) (Figure 1.7).

Mon2 was characterized by negative staining electron microscopy, dynamic light

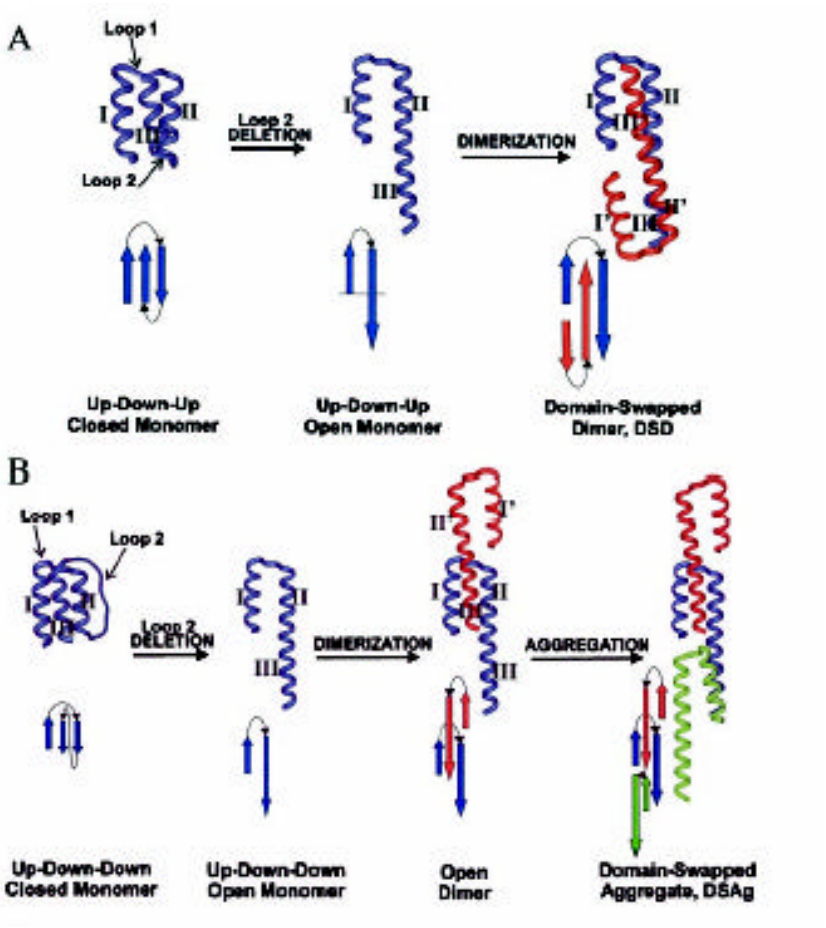


Figure 1.7. Reproduced from Ogihara *et al*., 2001. (A) Design of Mon1 (up-down-up topology). (B) Design of Mon2 (up-down-down topology).

scattering, CD spectroscopy and Fourier transform-IR spectroscopy because of the difficulties in crystallization. At both acidic and neutral pH, electron micrographs shows that Mon2 form long fibrous aggregates. The fibrils are 40-70 nm in width and up to several thousand nm in length. These fibrils are composed of several protofibrils, corresponding approximately to the thickness of one to three triple stranded α-helical bundles. The structure of Mon2 oligomer is not known, but all the results strongly indicates that 3D domain swapping is the mechanism for Mon2 oligomer formation (Ogihara *et al*., 2001) (Figure 1.7).

Human cystatin C dimerizes through 3D domain swapping

Human cystatin C is a cysteine protease inhibitor belong to the papain and legumain families (Turk and Bode, 1991 and Grubb, 2000). Three types of cystatins are present in higher animals: type 1, without signal peptides (cystatins A and B); the secretory type 2 cystatins (C, D, E, F, S, SN, SA) and the multi-domain type 3 cystatins (high and low molecular weight kininogens). Human cystatin C (HCC) is composed of 120 amino acids (Grubb, 2000) and contains, as do other type 2 cystatins, four Cys residues forming two characteristic disulfides (Janowski *et al*., 2001). Wild type HCC forms part of the amyloid deposits in brain arteries of elderly patients suffering from cerebral amyloid angiopathy (Grubb, 2000). In hereditary cystatin C amyloid angiopathy (HCCAA), occurring in the Icelandic population, a natural variant of HCC (Leu68Gln) forms massive amyloid deposits in brain arteries of young adults leading to lethal cerebral hemorrhage (Olafsson & Grubb, 2000).

Crystallographic and NMR studies of three cysteine protease inhibitor, chicken

cystatin (Bode *et al*., 1988, Dieckmann *et al*., 1993 and Engh *et al*., 1993), cystatin B in complex with papain (Stubbs *et al*., 1990) and cystatin A (Martin *et al*., 1995), provide the information about the overall fold of cysteine protease inhibitor. The canonical features of this fold includes the long α1 helix running across a large five-stranded antiparellel β-sheet. The antiparellel β-sheet has the following connectivity: (N)-β1-α1-β2-L1-β3-(AS)-β4-L2-β5-(C) (Figure 1.8A), where AS is a broad 'appending structure' and positioned on the opposite site of the β-sheet relative to the N-terminus and loops L1 and L2 (Janowski *et al*., 2001).

Three regions of cystatin are important in interaction with cysteine protease. These regions include the N-terminal segment and loops L1 and L2 (Janowski *et al*., 2001). The cubic crystal form of HCC was obatained from a solution containing monermeric protein. The crystal structure revealed that HCC forms a dimer with two identical domains contributed by both molecules. The molecules are formed by two fold symmetry. Furthermore, the HCC dimer is formed *via* 3D domain swapping (Janowski *et al*., 2001). The domain-swapped domains consist of an α-helix and two β-strands, β1 and β2 (Figure 1.8B).

Each domain of the HCC is composed of the general fold of the chicken cystatin (Bode *et al*., 1988). The N-terminal subdomain consists of a short β1 strand followed by the α1 helix. After forming a loop at residue Asn 39, the chain forms the long β2 strand. The β-strands, β3-β5, form the β-subdomain through the linker region βL (former L1). The HCC dimer forms an open interface through the β-sheet interaction in the βL region, in addition to the closed interface, α-β interface (Figure 1.8B). The open interface is

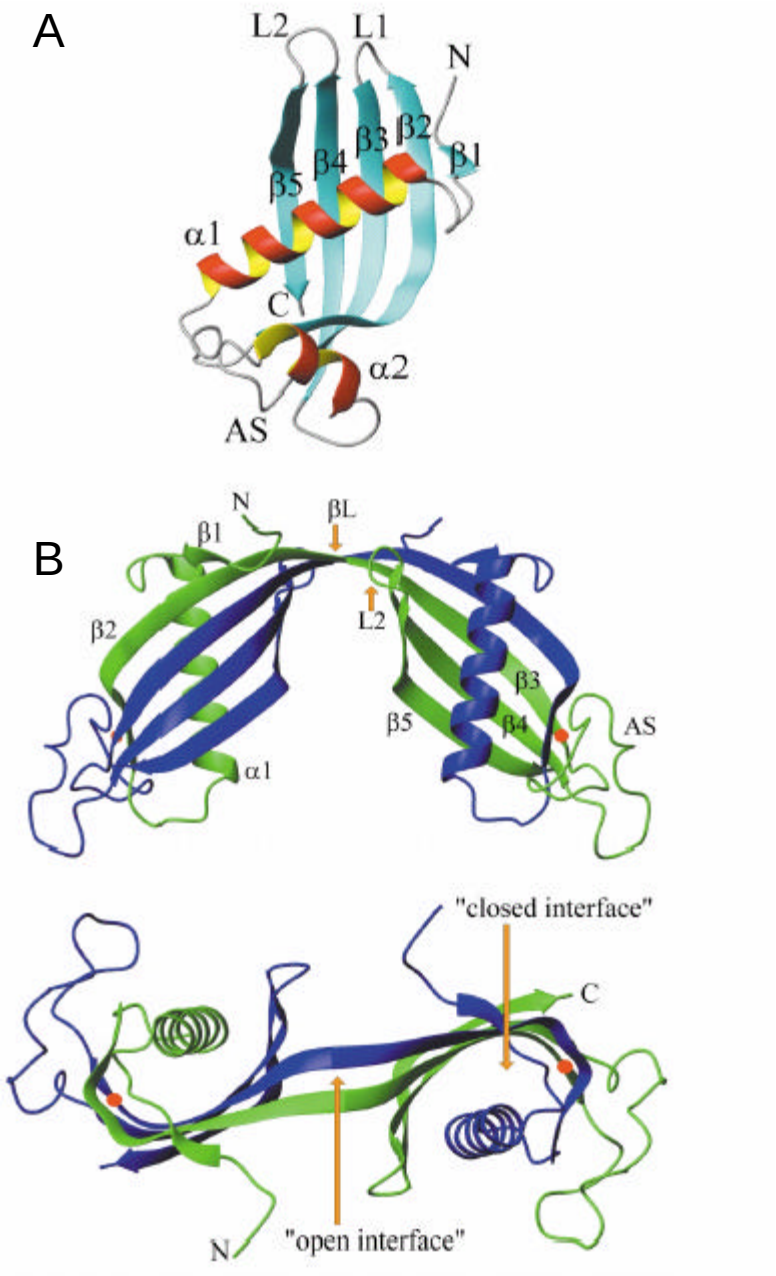


Figure 1.8. Reproduced from Janowski *et al.*, 2000.A. The fold of chicken cystatin. B. Domain swapped dimer HCC in two different view angles. The red dot is the site of the L68Q mutation.

composed of an unusually long antiparellel β-sheet formed by two copies of strand (β2-βL-β3), which cross from one domain to another and are involved in as many as 34 main chain hydrogen bonds.

The βL strands corresponds to the L1 inhibitory loop in the canonical cystatin fold. The disappearance of the loop L1 in the HCC dimer structure and the consequent disruption of the HCC functional element agree with the observation that the HCC dimer has absolutely no inhibitory effect on C1 type proteases (Abrahamson and Grubb, 1994). However, loop 39-41, which connects helix α1 with strand β2 and contains Asn 39 that is crucial for HCC inhibition of mammalian legumain, is not affected by dimerization. This agrees with the observation that HCC dimer is an active inhibitor of porcine legumain like the monomeric protein (Alvarez-Fernandez *et al*., 1999). Leu 68 is located in the central strand (β3) of the β-sheet, on its concave face and covered by helix α1 (Figure 1.8B). In the hydrophobic core of the protein, this residue occupies a pocket formed by the surrounding residues of the β-sheet and the hydrophobic face of the helix. Leu 68 is surrounded by Val 66 and Phe 99 of the same monomer, as well as Leu 27, Val 31, Tyr 34 (in the α-helix) and Ala 46 (strand β2) of the complementary monomer. These residues make an enclosed hydrophobic environment around Leu 68. Replacement of the Leu side chain by the longer Gln side chain, as in the naturally occurring L68Q variant, would not only form tight contacts but would also place the mutated hydrophilic chain in a hydrophobic environment. This would most probably destabilize the hydrophobic pocket. The hydrophilic side chain exerts a repulsive force on the α-helix, together with the strand β2, from the compact molecular core and forces the molecule to unfold into an open monomer. NMR spectroscopy shows that there is an increase in the dynamic nature of the

L68Q mutant compared to wild type HCC (Ekiel *et al*., 1997 and Gerhartz *et al*., 1998). It is conceivable that the refolded dimer recreates the topology of monomeric HCC, whereby the destabilizing effect would be similar in both cases. However, the dimeric structure may be more resistant to destruction because of the extra stabilizing contribution from the open interface, β-sheet interaction in the βL region (Janowski *et al*., 200).

Human cystatin C dimers swap only the N-terminal domain (Janowski *et al*., 200). Higher oligomer may also arise by domain swapping of the N-terminal domain of the HCC which forms part of the amyloid deposits in brain arteries of elderly patients suffering from cerebral amyloid angiopathy (Figure 1.9).

Hinge loop role in 3D domain swapping

A hinge loop has the intrinsic flexibility to adopt different conformations in the monomer and in the domain-swapped oligomer. The flexibility is evident in RNase A, BS-RNase and human pancreatic ribonuclease (hRNase) chimera. RNase A and BS-RNase show 80% sequence identity, and BS-RNase and hRNase chimera share the common hinge loop. All three of these proteins swap the N-terminal helix. However, the relative orientations of the subunits in their dimers are different and resulting in different conformations for the three hinge loops (Mazzarella *et al*., 1993, Liu *et al*., 1998 and Canals *et al*., 2001).

Great flexibility is also exhibited in the C-terminal hinge loop of RNase A: the C-terminal strand of RNase A is swapped in both the C-terminal swapped dimer and the cyclic C-terminal swapped trimer of RNase A. The same hinge loop adopts different conformation in monomer, dimer and trimer.

Hinge loops show a variety of secondary structures in domain-swapped proteins.

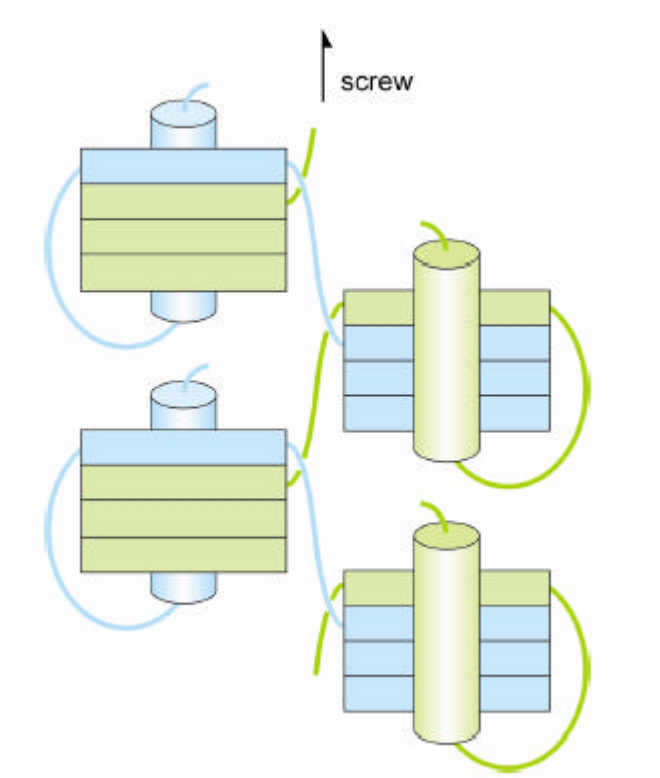


Figure 1.9. Reproduced from Janowski *et al.*, 2000. In this diagram, the cystatin fold is represented by a α-helix (cylinder) running across the concave face the β-sheet (stripes). In a screw operation, new components are added by rotation followed by a translation along the screw axis.

Some hinge loops are coils, some are β-strand and others are α-helix. In the RNase, the N-terminal swapped dimer shows that one hinge loop is a coil and the other one is an α-helix (Mazzarella *et al*., 1993 and Canals *et al*., 2001). A common feature is that when the hinge loop forms an α-helix or a β-strand, the oligomer is favored over the monomer. These proteins usually exist as dimers *in vivo* or have dimeric forms more stable than the monomeric forms (Liu *et al*., 1998). Other domain-swapped oligomers are more stable than their monomers are those for which hinge loop is shortened by a deletion, then the closed monomer is no longer possible (Bennett *et al*., 1995).

PROTEIN X-RAY CRYSTALLOGRAPHY

In 1947 Perutz, along with Kendrew, founded the Medical Research Council Unit for Molecular Biology at Cambridge. There, the two men continued their investigation of hemoproteins, with Kendrew trying to determine the molecular structure of myoglobin (muscular hemoglobin) and Perutz concentrating on the hemoglobin molecule itself. In 1960 Perutz showed that the hemoglobin molecule is composed of four separate polypeptide chains that form a tetrameric structure, with four heme groups near the molecule's surface. However, he did not determine the first three-dimensional structure of a protein molecule. The first three-dimensional structure was determined by John Kendrew. By 1957 Kendrew obtained an electron density map at 6Å resolution which allowed him to build a rough molecular model of myoglobin, and two years later he extended the resolution to 2.0Å, allowing him to build an atomic model. For their work, Perutz and Kendrew were awarded the Nobel Prize for Chemistry in 1962. This

pioneering work led to steady increase in the number of proteins structure determined by using X-ray diffraction. The numbers of protein structures determined by X-ray crystallography is approximately 20000 as of 2004.

Protein crystallization

Obtaining a suitable single crystal is the least understood step in the X-ray structural analysis of a protein. Protein crystallization is mainly a trial and error process. The crystallization of a protein involves four important steps.

1. The purity of the protein is important.
2. The protein should be dissolved in a suitable solvent. The solubility of proteins in water depends on properties such as temperature, pH, and the presence of other solution components as well as amino acid composition.
3. The concentration of a protein solution is brought above its solubility limit, until the solution becomes supersaturated.
4. When a protein solution is brought to the supersaturated point, the protein begins to aggregate and forms nuclei. Once nuclei have formed, actual grystal growth can begin.

Crystal systems and symmetry

The key to unlocking the structure of crystals is finding the symmetry of the crystal's unit cell in the X-ray pattern. A unit cell is a set combination of atoms or molecules which forms a repeating pattern throughout the crystal. Like bricks in a wall the entire structure can be created by simple translations of the fundamental unit. Unit cells exist in three dimensions and thus have three sides (a, b and c) with three angles between each of the

combinations of two sides (α between b and c, β between a and c and γ between a and b). Based on the relations between a, b, c, α, β and γ, all crystals are placed into one of seven crystal systems. These systems are further divided into 32 crystal classes based on the symmetry point groups which they exhibit. These point groups defined by their operations: inversion through a point, rotation about an axis, reflection through a mirror plane and inversion through a point after rotation about an axis. The 32 crystal classes are in turn further broken into 230 space groups, based on symmetries derived from one of two basic translational operations: screw axes and glide planes. Since proteins are asymmetric objects and occur only in the L-form they cannot involved in symmetry element requiring inversion centers, mirrors or glide planes. This limits the possible space groups to 65 out of the 230 mathematically possible space groups.

X-ray diffraction and Bragg's Law

Protein crystallography is a widespread technique for the determination of the three-dimensional structure of protein molecules. It is based on the study of X-ray diffraction patterns by crystals of a protein. Atoms diffract X-rays in a pattern which is dependent on their location in three-dimensional space. To detect diffracted X-rays with high sensitivity, it is essential that many atoms contribute to the diffraction pattern obtained. This means that the molecule under investigation must be present in the ordered three-dimensional array within the crystal so that many equivalent atoms in different molecules contribute to the diffraction pattern.

The Bragg equation relates to the spacing between crystal planes, d, to the particular Bragg angle, ?, at which reflections from these planes are observed. Bragg’s law

indicates that diffraction is only observed when a set of planes makes a very specific angle with the incoming X-ray beam. This angle depends on the inter-plane spacing d. Bragg's Law refers to the simple equation: nλ = 2d sin ?.

Ewald construction

A most useful mean to understand the occurrence of diffraction spots is the Ewald construction. We draw a sphere of radius 1/?, in the center of which we imagine the real crystal. The origin of the reciprocal lattice lies in the transmitted beam, at the edge of the Ewald sphere. Whenever the end point of scattering vectors, within length 1/d, fall on the sphere, Brgg's law is satisfied and reflection occurs. Furthermore, the end point of scattering vectors must be reciprocal lattice points (Figure 1.10). Crystal lattice planes for which the reciprocal lattice points do not lie on the sphere and thus are not in reflecting position can be brought to refection by rotating the crystal.

The structure factor

The structure factor can be calculated as

$$F(hkl) = \sum_{j=1}^{atom} f_j \exp[2pi(hx_j + ky_j + lz_j)] \qquad (1.1)$$

f_j is the scattering factor of atom *j*. *x, y, z* are the fractional coordinates of each atom in the summation, and *h, k, l* are the three indices of the corresponding reflection (1.1).

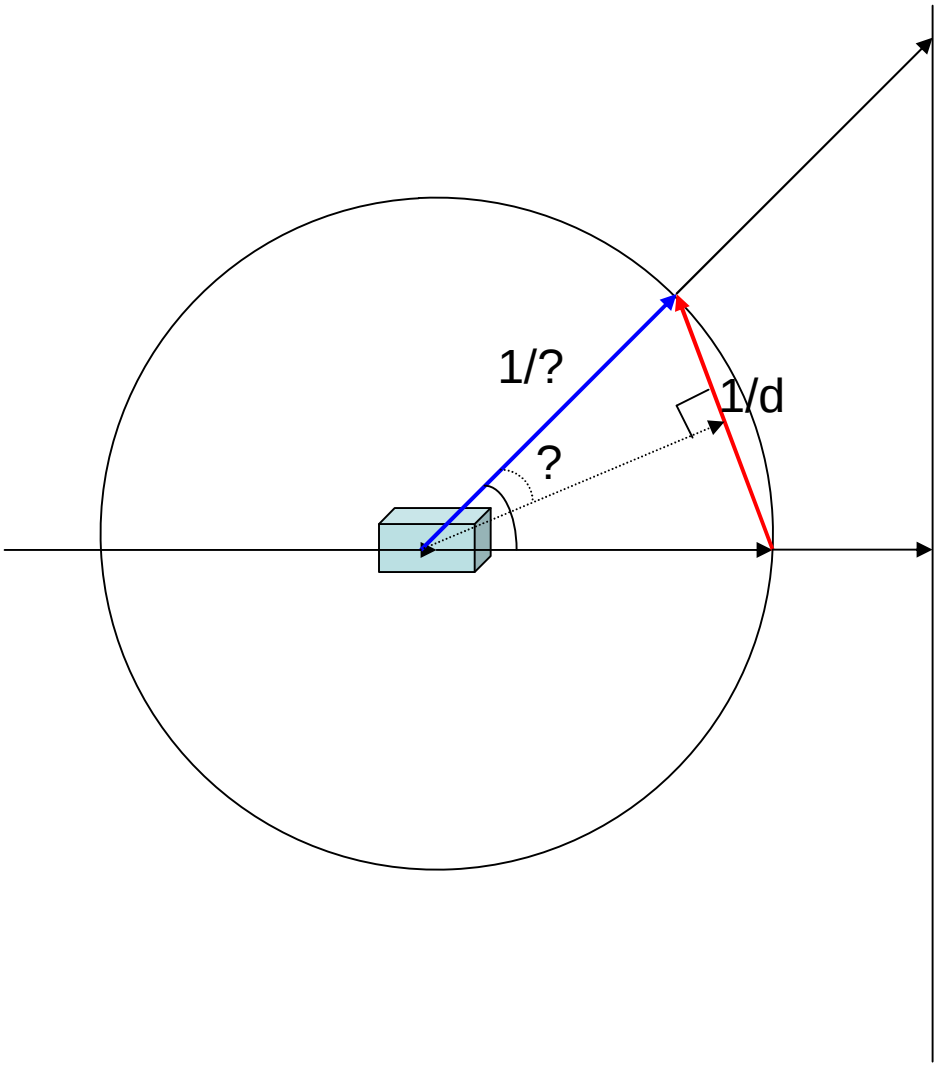


Figure 1.10. Ewald construction. Scattering vectors has length 1/d (red) and the sphere of radius 1/? (blue). Whenever scattering vectors have their end points on the sphere, Brgg's law is satisfied and reflection occurs.

Fourier transform and phase problem

The electron density in a crystal can be obtained by calculating Fourier summation

$$s_{(x,y,z)} = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l} | F_{(h,k,l)} | \exp[-2pi(hx + ky + lz - a_{(h,k,l)})] \qquad (1.2)$$

$|F_{(h\,k\,l)}|$ is the structure factor amplitude of reflection *(h k l)* and $a_{(h\,k\,l)}$ is the phase angle. *x, y, z* are the coordinates to what in the unit cell (1.2). From *I(h k l),* amplitude $|F(h\,k\,l)|$ can be calculated but no information is available on the phase angle. This is called phase problem. In principle, four techniques exist for solving the phase problem in protein X-ray crystallography.

If we already have the coordinates of a similar protein we can try to solve the structure using a process called Molecular Replacement which involves taking this model and rotating and translating it into our new crystal system until we get a good match to our experimental data. If we are successful then we can calculate the amplitudes and phases from this solution which can then be combined with our data to produce an electron density map.

If we have no starting model, then we can use Isomorphous Replacement methods whereby one or more heavy atoms are introduced into specific sites within the unit cell without perturbing the crystal lattice. This is another trial-and-error procedure and often it is not apparent whether it has worked until more X-ray data have been collected. Heavy atoms are electron dense and give rise to measurable differences in the intensities of the spots in the diffraction pattern. By measuring these differences for each reflection, it is

possible to derive some estimate of the phase angle using vector summation methods. In practice data from one or more heavy atom derivatives is required to get good enough phases - hence Multiple Isomorphous Replacement (MIR). Again we use a Fourier transform to calculate a map.

In some cases we can make use of the anomalous scattering behavior of certain atoms at or near their X-ray absorption edges to gain useful phase information. Many of the atoms used in isomorphous replacement are also useful in this respect. This additional information can enhance the structure solution. Multiwavelength Anomalous Dispersion (MAD) is an elegant and often very effective method that relies entirely on the measurement of the anomalous differences produced by one or more anomalously scattering atoms in the crystal. In practice three or more consecutive data sets are recorded from the same crystal at different wavelengths around the X-ray absorption edge of the anomalous scatterer. As this method requires a tuneable X-ray source, it can only be performed at a synchrotron. The resultant phase information can often produce very high quality electron density maps, thereby simplifying the subsequent interpretation. Selenium is a particularly good anomalous scatterer and it can be incorporated into proteins by over-expressing them in strains of *E. coli* that are auxotrophic for methionine. The host cells are then grown on minimal media supplemented with amino acids using selenomethionine in place of methionine.

Direct Methods are commonly used by crystallographers to determine relatively small crystal structures. However, traditional approaches do not appear to be extensible to larger structures. Hauptman and Jerome Karle won a Nobel prize in 1985 for their work in the late 1940s and early 1950s on the direct method, a mathematical approach that makes it possible to glean phase data from the diffraction intensities.

Model building

This is the process where the electron density map is interpreted in terms of a set of atomic coordinates. This is more straightforward in the molecular replacement case because we already have a coordinate set to work with. In the case of isomorphous replacement we simply have the map. The normal procedure is to fit a protein backbone first then if the resolution permits, we insert the sequence. The amount of detail that is visible is dependent on the resolution and the quality of the phases. Often regions of high flexibility are not visible at all due to static disorder, where the structure varies from one molecule to the next within the crystal.

Refinement

Once we have a preliminary model we can refine it against our data. This will have the effect of improving the phases which results in clearer maps and therefore better models. We would typically go round this cycle several times until we get little or no further improvements. At this stage we would expect a crystallographic R-factor of below 25%. This is a measure of the agreement between the model and the data – the lower the value the better the model. Nevertheless, the final model must make chemical sense and there must be no large regions of electron density unaccounted for.

AIM AND SCOPE OF THE THESIS

The fungal immunomodulatory protein from the edible golden needle mushroom (*Flammulina velutipes*), designated Fve, is a single polypeptide consisting of 114 amino-acid residues. It is believed to trigger the mitogenic proliferation of T lymphocytes and Th1 cytokine production. Fve forms a homodimer in nature. On the other hand, rhodocetin is purified from the crude venom of the Malayan pit viper, *Calloselasma rhodostoma*, functions as a novel platelet aggregation inhibitor induced by collagen. The molecule, rhodocetin has been shown to have activity only when present as a dimer. It is a unique heterodimer consisting of α and β subunits of 133 and 129 residues respectively. Structural studies of these proteins are expected to contribute significantly towards understanding structure-function relationships of the proteins. Furthermore, both dimers were believed to be formed by a mechanism called 3D domain swapping. Therefore, functional and structural studies of these proteins were initiated to better understand the mechanisms. Explicitly, the objectives of the study were:

1. To crystallize the protein.
2. To collect native and derivative data.
3. Find appropriate methods to solve the structures.
4. Build the models.
5. Use the structural information to study the structure-function relationships and domain swapping mechanism of the proteins.

CHAPTER 2

Structural characterizations of fungal immunomodulatory protein: Fve

For centuries, the higher fungi, Basidiomyceta - more popularly, mushrooms - have figured strongly in traditional medicine, particularly in the Chinese and Japanese cultures, to promote good health and vitality and to enhance body's adaptive capabilities. But though their culinary and nutritional merits are widely accepted, the utility in modern medicine of mushrooms and their component biomolecules, is only now being appreciated. In just the last few years, the literature has burgeoned with reports of proteins, glycans and small molecules isolated from these fungi, possessing a variety of activities: immunopotentiating, anti-tumor, genoprotective, cholesterol-lowering, hypoglycemia-inducing and appetite-suppressing, to name a few (Wasser *et al*., 1999). Ancient wisdom is vindicated since mushrooms are a rich source of pharmacologically important biomolecules, promising much both as tools to study biological processes and as leads for the development of therapeutics.

The golden needle mushroom, *Flammulina velutipes*, and extracts thereof, possess immunomodulatory, anti-tumor, antiviral and cholesterol-lowering activities (Wasser *et al*., 1999; Otagiri *et al*., 1983; Ikekawa *et al*., 1995; Wang *et al*., 2000; Fukushima *et al*., 2001). A major fruiting body protein, designated Fve or FIP-fve, was isolated and very likely plays a significant role in the mushroom's immunomodulating effects (Ko *et al*., 1995). It stimulates mitogenesis of human peripheral lymphocytes, suppresses systemic anaphylaxis reactions and local swelling of mouse foot pads, and enhances the transcription of interleukin-2 (IL-2), interferon-γ (IFN-γ) (Ko *et al*., 1995) and tumor necrosis factor-α (TNF-α; Seow *et al.*, 2003). In addition, it hemagglutinates human red blood cells (Ko *et al*., 1995). Several simple sugars tested did not inhibit Fve-induced hemagglutination (Ko *et al*., 1995). Fve is a non-covalently linked homodimer (Seow *et*

al., 2003), each subunit 114 amino acids in length. It contains no Cys, Met or His residues and its N-terminal Ser residues are acetylated (Ko *et al*., 1995). It shows sequence similarity only to other mushroom proteins with similar functional characteristics: LZ-8 (*Ganoderma lucidum*, Kino *et al*., 1989), Gts (*Ganoderma tsugae,* Lin *et al*., 1997), Vvo (*Volvariella volvacea*, Hsu *et al*., 1997) and Vvl (*Volvariella volvacea*, She *et al*., 1998). All these proteins form the novel Fungal Immunomodulatory Protein (FIP) family (Ko *et al*., 1995) (See appendix).

To better understand structure-function relationships in this family of proteins, we determined its three-dimensional structure.

MATERIALS AND METHODS

Protein purification

Fve was purified from *Flammulina velutipes* as previously described (Ko *et al*., 1995) with some modifications (Seow *et al*., 2003). Essentially, proteins from a 5% acetic acid, 50 mM 2-mercaptoethanol extract of fruiting bodies were precipitated with 95% ammonium sulfate. The precipitate was dialyzed against 10 mM Tris-HCl, pH 8.5. The dialysate was applied to a Q-Sepharose FF column equilibrated with 10 mM Tris-HCl, pH 8.5. The unbound proteins were then loaded on an SP Sepharose FF column equilibrated with 10 mM sodium acetate pH 5.0. Fractionation with a 0 – 0.5 M NaCl gradient yielded pure Fve. All steps of the purification were performed at 4° C and hemagglutination assays were used to follow the purification. The purity of Fve was assessed by SDS-PAGE and electrospray ionization mass spectrometry.

Protein crystallization and data collection

Well-diffracting Fve crystals were grown by vapor diffusion from hanging drops at 2.5% PEG 400, 2.0 M $(NH_4)_2SO_4$, 0.1 M Tris-base, pH 8.5 at 21°C over 3-5 days (Seow *et al*., 2003). Heavy atom derivatives were prepared by soaking the crystals in mother liquor containing 25% glycerol and 1 M NaBr for 1 min, after which the crystals were flash-frozen at 100 K. SAD data from a derivatized crystal were collected at the National Synchrotron Light Source (NSLS) beam line X12C at one wavelength (0.92020 Å) around the Br absorption edge. The crystal diffracted to 1.7 Å. It belonged to the tetragonal space group $P4_32_12$ and had unit cell dimensions $a = b = 97.12$, $c = 61.41$ Å and $\alpha = \beta = \gamma = 90.0^o$. This gave a Matthew's co-efficient (Matthews *et al*., 1968) of 2.9 $Å^3.Da^{-1}$ and a solvent content of 56.5%, with one molecule (dimer) per asymmetric unit. The SAD data were processed and scaled using DENZO and SCALEPACK from the HKL2000 suite of programs (Otwinowski *et al*., 1997).

Structure solution and refinement

All of the bromine heavy atom positions were located and refined by the program SOLVE (Terwilliger *et al*., 1999) at 1.7 Å and a solvent flattened map was calculated using RESOLVE (Terwilliger *et al*., 2001). The resulting electron density map revealed secondary structure elements and side chains (Figure 2.1). The phases obtained from RESOLVE were directly used in ARP/wARP (Morris *et al*., 2002) for automated main chain tracing, resulted in the building of four continuous fragments that contained 97% of the model. The rest of the model and side chains were fitted manually using XtalView (McRee *et al*., 1999) (Figure 2.2). Refinement was carried out with REFMAC 5

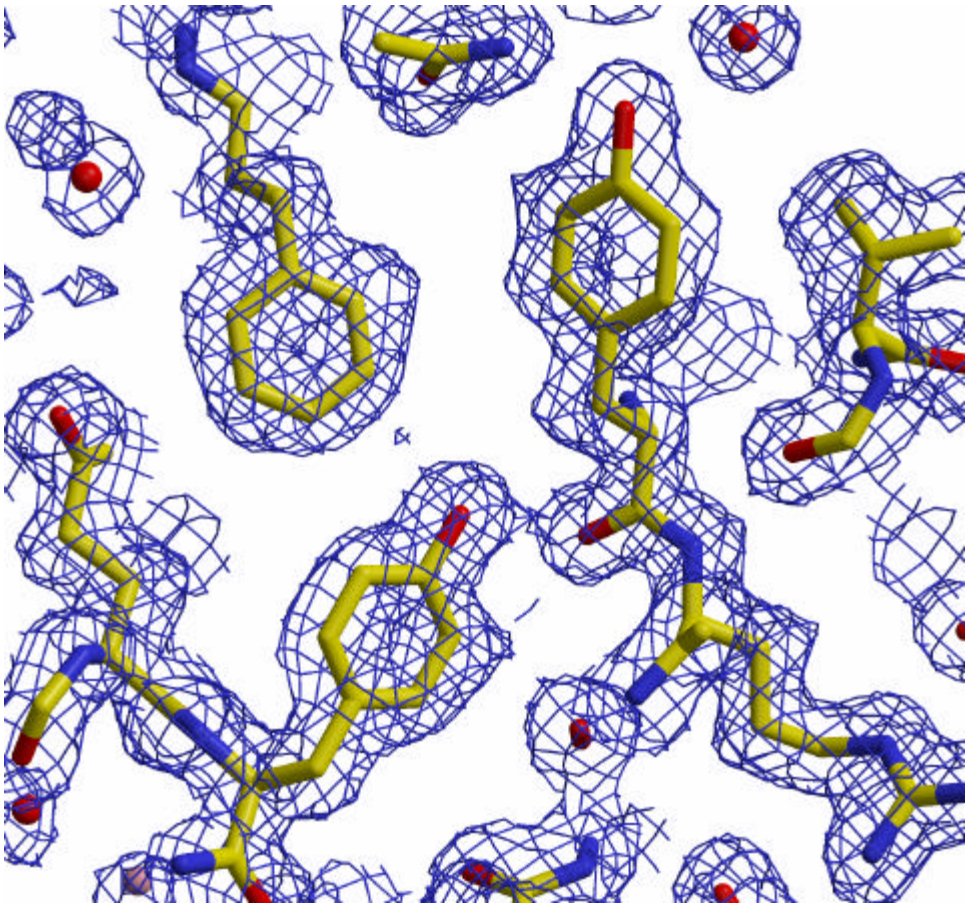


Figure 2.1. Portion of experimental electron density from RESOLVE contoured for 2σ demonstrating the quality of the obtained phases from SAD.

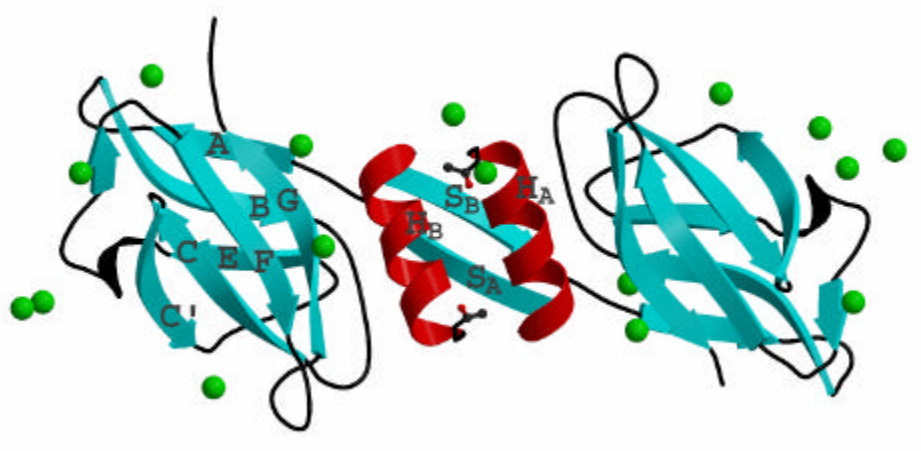


Fig 2.2. Overall structure of the Fve dimer solved by SAD of an NaBr-soaked crystal. Each monomer consists of an FNIII-like domain with an N-terminal α-helix and β-strand. The α-helices (H_A and H_B) interact via hydrophobic interactions while the β-strands (S_A and S_B) form a β-sheet. The FNIII domain consists of a sandwich of two sheets, I and II, formed by β-strands E-B-A and G-F-C-C', respectively (labels on monomer A). The Br atoms are colored green.

(Murshudov *et al*., 1999) using data in the resolution range of 30.02 Å - 1.7 Å and water molecules were identified using ARP/WARP later in the refinement. The data collection and refinement statistics are shown in Table 2.1. Poor electron density in corresponding regions necessitated the omission of the side chain of Thr 114 of chain A and residues Lys 113 and Thr 114 of chain B from the model. The quality of the final model was verified with PROCHECK (Laskowski *et al*., 1993). Of note is Lys 14 in both A and B chains, which, though fitting very well in the 2fo-fc map, falls in the disallowed region in the Ramachandran plot (Figure 2.3). Interestingly, Lys14 participates in both the N-terminal helix (Ala 2 to Lys 14) by the formation of a H-bond between its backbone N atom and the C=O group of Ala10, as well as the ensuing β-sheet (Lys 14 to Asp 19) by a H-bond between its C=O group and the backbone N atom of Tyr20 of the other monomer (Figure 2.4). It is likely that both these H-bonds introduce strain in the peptide backbone of Lys14 causing a departure from the allowed phi/psi angles for both α-helices and β-sheets. The atomic co-ordinates and structure factors have been deposited in the Protein Data Bank (Berman *et al*., 2000) (PDB) as entry 1OSY. MolScript was used for molecular drawing.

RESULTS AND DISCUSSION

Overall fold

The Fve dimer is a dumb-bell shaped molecule, with approximate dimensions of 80 Å × 40 Å × 32 Å (Figure 2.1). The conformations of the two monomers (A and B) are nearly identical as they superimpose almost perfectly: the RMSD between their first 112 residues is 0.29 Å (C_α atoms alone) and 0.87 Å (all atoms). The N-terminal Ser residues are

Table 2.1. Data collection, phasing and refinement statistics.

Data collection	
Space group	$P4_32_12$
Unit cell parameters (Å, °)	$a = b = 97.12$, $c = 61.41$, $\alpha = \beta = \gamma = 90.0$
Resolution range (Å)	30.0 - 1.7
Number of observed reflections	464964
Number of unique reflections	32880 *(3249)*
Redundancy	14.1
Completeness (%)	100 *(100)*
$*R_{merge}$ (%; all reflections)	0.045 *(0.187)*
Matthews co-efficient ($Å3.Da^{-1}$)	2.8
Solvent content (%)	56.4
Number of molecules per asymmetric unit	2
$I/\sigma(I)$	28.1 (*16.7*)
Phasing statistics	
Mean figure of merit	0.41
Model refinement	
Number of protein atoms	1788
Number of bromine atoms	16
Number of water molecules	136
R-factor	0.182
Rfree (test set of 5%)	0.210
Data cutoff in σ units	none
RMS bond length	0.010
RMS bond angle	1.466
RMS chiral	0.231

Values for parameters in the highest resolution shell, 1.76 – 1.70 Å are parenthesized and in italics.

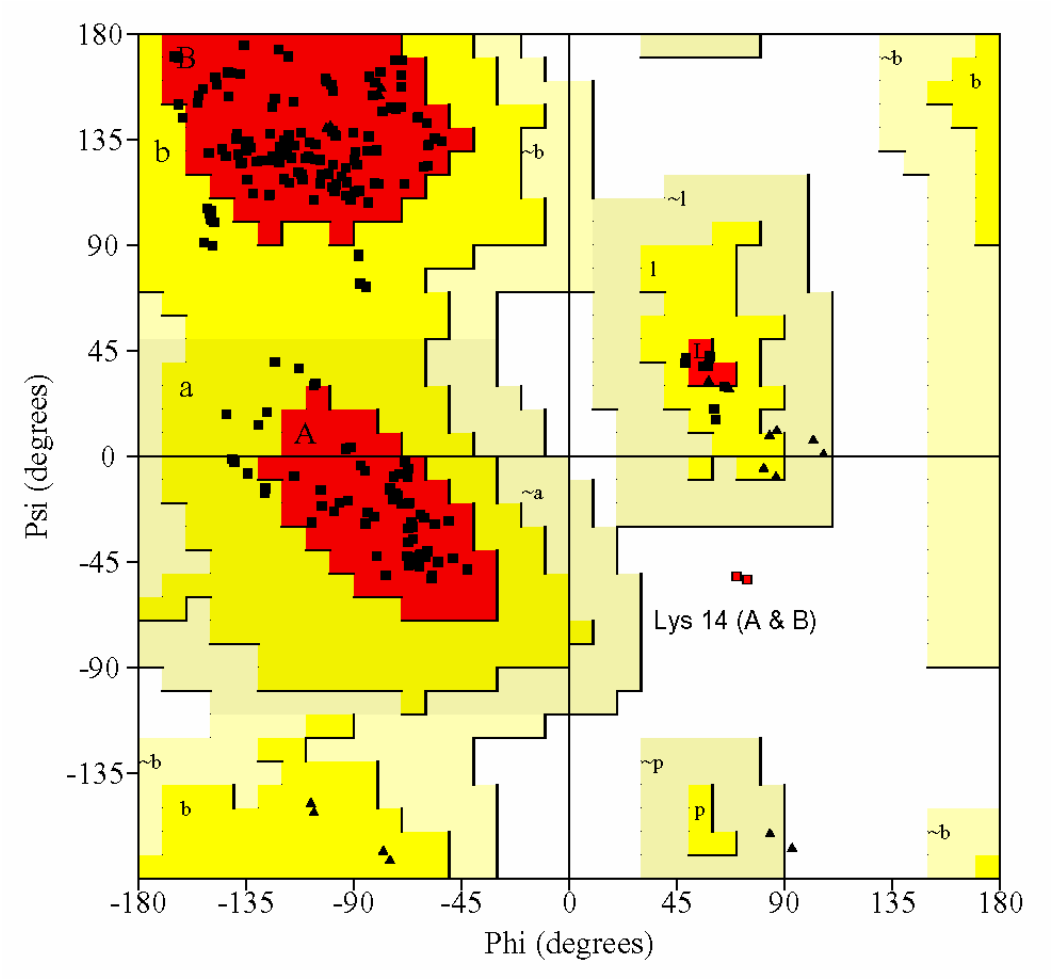


Figure 2.3. In Fve, Lys14 in both A and B chains falls in the disallowed region of the Ramachandran plot.

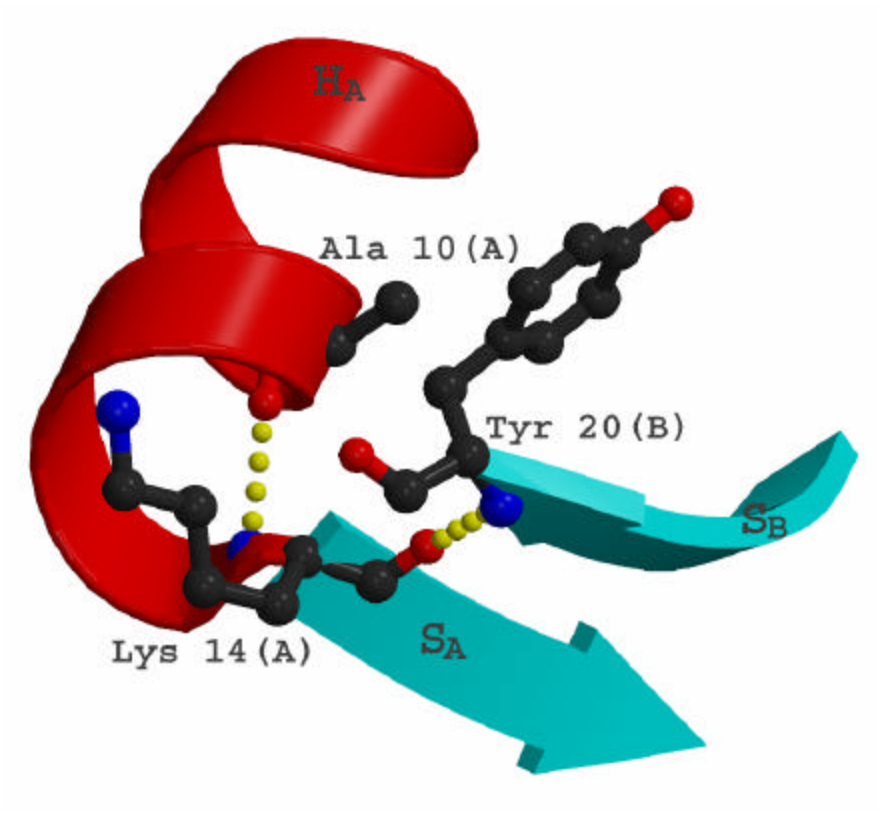


Figure 2.4. The departure from allowed phi-psi angles is likely due to the strain imposed when the backbone N and C=O of Lys14 H-bonds with the C=O of Ala10 and the backbone N of Tyr20, respectively, while forming part of the N-terminal α-helix and β-sheet.

blocked by acetyl groups, the density for which is clearly observed in both subunits. Each subunit comprises of a pair of N-terminal secondary structural elements, an α-helix followed by a β-strand, linked to a domain consisting almost exclusively of β–sheets adopting an Ig-like fold (Figure 2.1). The N-terminal α-helices in both subunits (designated H_A and H_B) span 13 residues from Ala 2 to Lys 14, and the β-strands (S_A and S_B), which run antiparallel to each other form a two-stranded β-sheet, from Lys 14 to Tyr 20. These secondary structure elements play a critical role in dimerization, and are important for the activity of the protein (see below). The rest of the subunit (residues Tyr 20 to Thr 114), folds into an Ig-like domain (Figure 2.1) consisting of seven major antiparallel β–strands arranged into two sheets of three (β-strands A, B and E) and four (β-strands C, C', F and G) packed against each other (for nomenclature of β-strands, see below). The two sheets are rotated with respect to each other by an angle of ~41°. β-strand A stretches from Pro22 - Gly25, B from Ser31 - Thr37, C from Tyr47 - Val52, C' from Gly55 - Val60, E from Gly71 - Ile74, F from Thr90 - Val95 and G from Glu104 - Glu110. The strands forming the two main β-sheets are connected by loops. The A-B, C-C' and F-G hairpins are each connected by type II β-turns: Gly27-Thr-Pro-Ser30, Val53-Asn-Gly-Ser56 and Thr100-Gly-Asn-Ser103, respectively. A type I β-turn (Thr67-Pro-Ser-Gly70) is found in the C'-E inter-sheet crossover loop, while two β-turns, a type I (Ala86-Asp-Thr-Lys89) and a type II (Asn80-Lys-Gly-Tyr83), are found in the E-F crossover loop, in addition to a 3_{10} helical turn (Phe76 – Tyr79) and a short β-strand (Gly84 - Val85).

Structural comparison of the Fve monomer against all structures in the PDB database using the program Dali (Holm *et al.*, 1993) (which maximizes a geometrical similarity score calculated from intramolecular distances in the common core) revealed

140 significant hits with a Z-score > 2.0. The top 14 hits (Z-score > 5.0) were all Ig-like β-sandwich domains belonging to either FNIII-type or Ig-like SCOP (Murzin *et al*., 1995) superfamilies (C1 and V sets) (data not shown). [In SCOP (Structural Classification of Proteins) (Murzin *et al*., 1995), proteins that have low sequence identities, but whose structural and functional features suggest a probable common evolutionary origin are placed together in superfamilies]. Structural alignment of these 14 domains with Fve using the program Topp (Lu, 1996) indicated that Fve had the highest 3D structural similarity (i.e., lowest topology diversity) to domains with the FNIII-type fold.

Topology of FNIII fold in Fve

A more rigorous examination of the β-strand arrangement in the Ig-like domain of Fve reveals some interesting features. As mentioned earlier, the domain consists of a sandwich of two β-sheets possessing three (A, B and E) and four (C, C', F and G) antiparallel β–strands, respectively (Figure 2.1). This arrangement is characteristic of the strand-switched type (s-type; sheet I: E-B-A, sheet II G-F-C-C') topology of Ig domains (Figure 2.5A; see [Bork *et al*., 1994] for classification), and is the hallmark of the FNIII domain. Strands B, C, E and F form the common core of the generic Ig-like fold (Bork *et al*., 1994). The polypeptide chain segment between strands C and E has to change direction by at least 180° and cross over to the other sheet; this is achieved via the C' strand in domains with s-type topology (Figure 2.5A). It has been shown that, within the Ig-like fold, there is a strong correlation between the number of residues between strands C and E, and the topological subtype (Bork *et al*., 1994). Ig-like folds with s-type topology characteristically contain seven to ten residues between a reference point C* (the residue

at the beginning of strand C' in the s-type topology (Bork *et al*., 1994)) and the beginning of the E strand. Interestingly, in Fve, there are 17 residues between C* (Val53) and the beginning of strand E (Gly71), much more reminiscent of domains with h-type topology (Figure 2.5C) wherein sheet crossover from common core strands C to E occurs via a long kinked strand C'/D (Bork *et al*., 1994). In the h-type topology, the N-terminal residues of this segment form hydrogen bonds with strand C (sheet II) whereas the C-terminal residues belong to sheet I (Figure 2.5C). In Fve, the segment from Gly55 to Pro68 bears a striking resemblance to h-type topology, strand C'/D (Figure 2.5B). It consists of two distinct segments demarcated by a kink at Asn63. The N-terminal segment corresponds to the canonical C' strand in s-type or h-type topology. However, unlike in h-type topology, the C-terminal segment traverses alongside strand E in the plane of sheet I, but not quite close enough to form hydrogen bonds with strand E in a manner required to participate in the formation of the sheet (the exception being the N atom of Thr67 H-bonding with the C=O of Gly71). This non-participation may be compounded by the fact that asparagine (in position 63) does not have a high propensity to participate in β-sheet formation (Chou *et al*., 1978). Hence, it appears that while Fve, with strict consideration of its seven β-stranded fold, has s-type topology, in reality it has a pseudo-h-type topology, wherein there is a "pseudo-D" polypeptide segment in an equivalent 3Dimensional position to h-type strand D, enabling strand C to make the crossover to strand E, but which is not close enough to strand E to facilitate H-bonding with it (Figure 2.5B). There are a few other domains that are classified under s-type topology and have extended C*-E regions (Bork *et al*., 1994). Among these, domain 1 of the human growth hormone receptor (PDB entry 3hhr) has the closest similarity to the pseudo-h-type fold of Fve. However, it is still

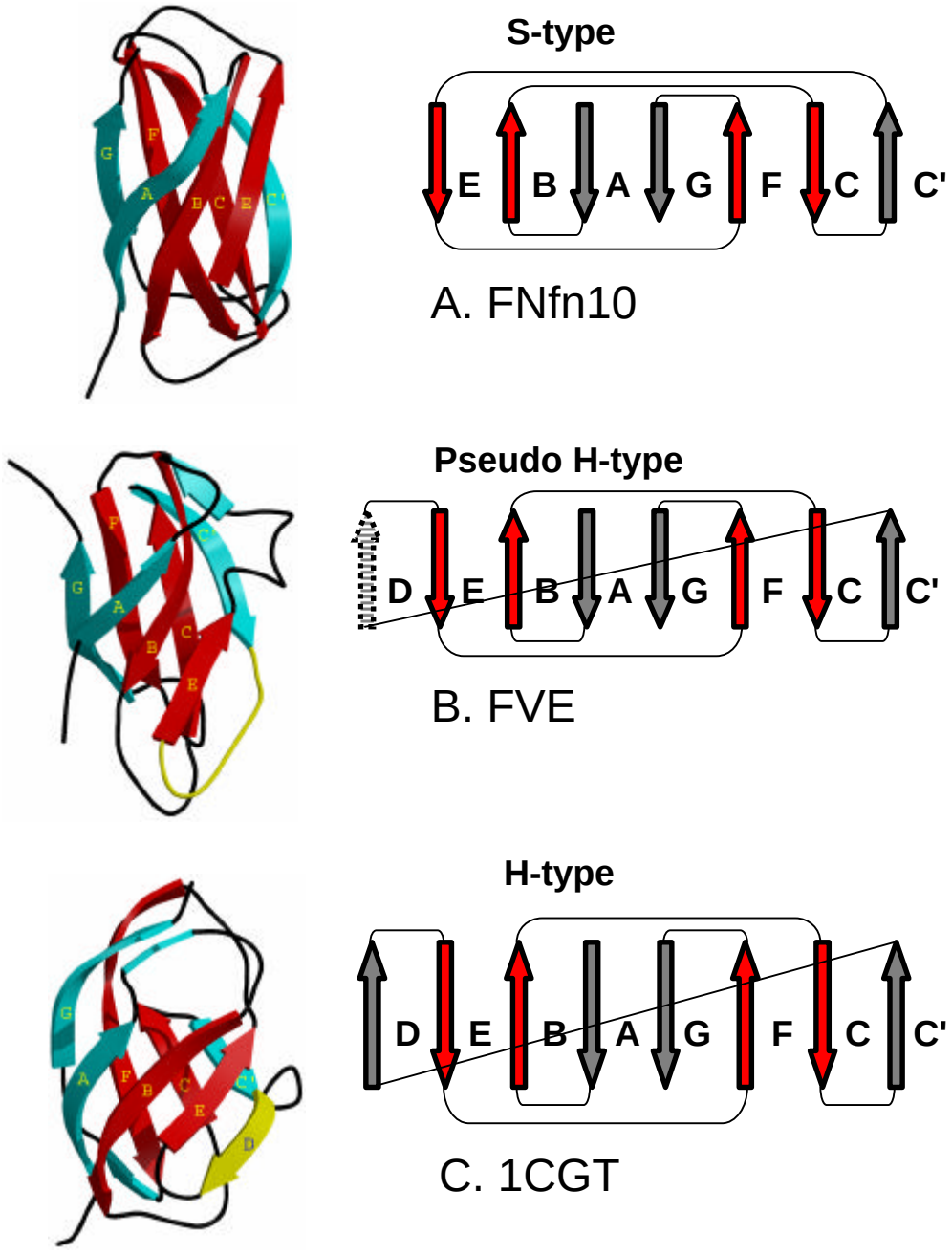


Figure 2.5. S-type topology (**A**), as in FNfn10, is adopted when 7 β-strands form a sandwich of 2 sheets. While the Ig-like fold in Fve resembles S-type topology in that it has 7 strands forming 2 sheets, it has additional segment (shown in yellow) that resembles strand D (yellow) in H-type topology (**C**). The pseudo-D strand is shown with a dashed arrow (**B**). The four common core β-strands, B, C, E and F are shown in red.

clearly s-type, since the polypeptide segment corresponding to the D strand is not in the plane of sheet I and is at right angles to it.

Determinants of the Ig-like fold

Common hydrophobic core. It is important to note that although Fve possesses a FNIII-type fold, it has no significant sequence similarity to any Ig-like fold-containing protein, and therefore a common evolutionary origin appears unlikely. The structural similarity perhaps arises from the chemistry of proteins favoring certain packing arrangements and chain topologies. A significant amount of work has been done to delineate the structural determinants of the Ig-like fold (see Bork *et al*., 1994, Clarke *et al*., 1999, Hamill *et al*., 2000 and Cota *et al*., 2001 and references therein). The Ig-like fold is one of ten proposed "superfolds" (Orengo *et al*., 1994 and Salem *et al*., 1999) – protein folds that are adopted by unrelated sequences and appear frequently in the database of known structures. The Ig-like fold is adopted by different superfamilies such as the Igs (Williams *et al*., 1988 and Chothia *et al*., 1998), FNIII modules (Baron *et al*., 1991 and Doolittle *et al*., 1993) and other more diverse families (Bork *et al*., 1994). Clarke *et al*., 1999 performed a structural and folding comparison of six Ig-like proteins with no significant sequence similarity and identified two trends that are independent of residue identity in the core: a clear correlation between domain stability and rate of folding; and a correlation between domain stability and the population of folding intermediates. The most likely explanation proposed was that the main determinants of the folding pathway are found in the common elements that define the fold, i.e., a common folding nucleus. The folding nucleus of an FNIII-type domain is composed of hydrophobic residues in the central strands (B, C, E and F) of the β-sandwich that form a large core of interactions in the transition state (Cota

et al., 2001). Although Fve has no sequence similarity to other FNIII-type domains, it does have a significant hydrophobic core (Figure 2.6) comprised of residues with <10% relative surface exposure: Ile33, Leu36, Val51 and Phe38 (B strand); Tyr47, Tyr49 and Val52 (C strand); Gly71, Gly72 and Ile74 (E strand); Ile91, Val93 and Val95 (strand F).

Position of proline residues. Important for the elaboration of the fold, and the common hydrophobic core, is the presence of proline residues in FNIII-type domains, that are found mainly in loop regions or at the start or end of β–strands (Steward *et al*., 2002). Fve also has several prolines in these regions (Figure 2.7). Pro22 is located at the start of strand A and appears to hold a key position in the domain swapping mechanism of dimerization (see discussion below). Pro29 is part of the type II β-turn linking strands A and B, while Pro39 is at the end of strand B, introducing a bend in the polypeptide chain facilitating its turn into strand C. Pro68 is within the type I β-turn within the C'-E inter-sheet crossover loop, while Pro98 is in the loop between the F and G strands.

Absence of tyrosine corner. Another structural feature of the FNIII domain (and Ig-like domains in general) is the "tyrosine corner" motif (Hemmingsen *et al*., 1994), that is found in >90% of (even sequence-unrelated) FNIII and Ig variable domains (Hamill *et al*., 2000). It consists of a tyrosine, equivalent to residue Lys89 in Fve, residue in the F strand that forms an H-bond with the local protein backbone in the adjacent E-F loop, which crosses from one sheet to the other (Figure 2.8). The tyrosine corner thus acts as a structural restraint of this loop connecting the β-sheets. The high conservation of this motif within the fold argues strongly for the specific necessity of an H-bond to the

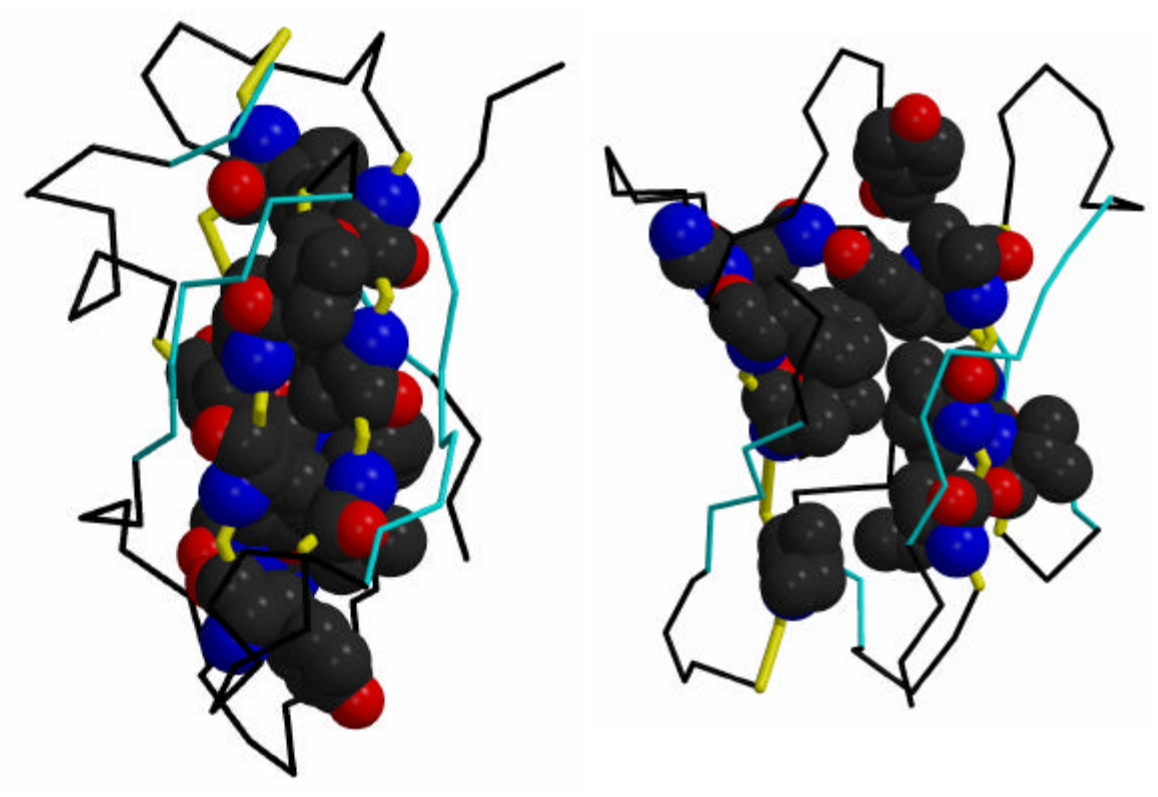


Figure 2.6. Although Fve has no sequence similarity to other FNIII-type domains, it has the characteristic hydrophobic core required to nucleate folding in FNIII domains. The hydrophobic residues in the central strands (B, C, E and F) of the β-sandwich are Ile33, Leu36, Val51 and Phe38 (B strand); Tyr47, Tyr49 and Val52 (C strand); Gly71, Gly72 and Ile74 (E strand); Ile91, Val93 and Val95 (strand F). Side-on (left) and face-on (right) views are shown.

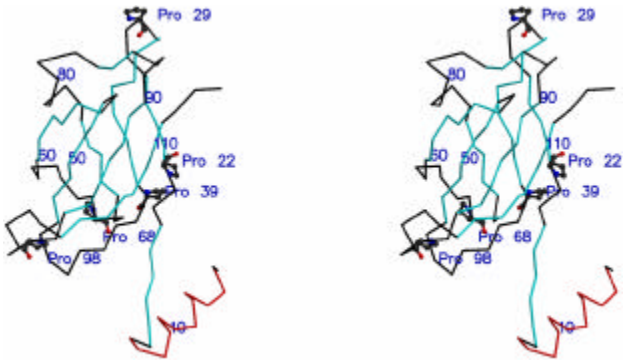


Figure 2.7. Stereo view showing the presence of prolines in loop regions or at the start or end of β–strands is important for stability of the FNIII fold. Fve has several prolines in these regions. Importantly, Pro22, located at the start of strand A, appears to hold a key position in the domain swapping mechanism of dimerization (see text).

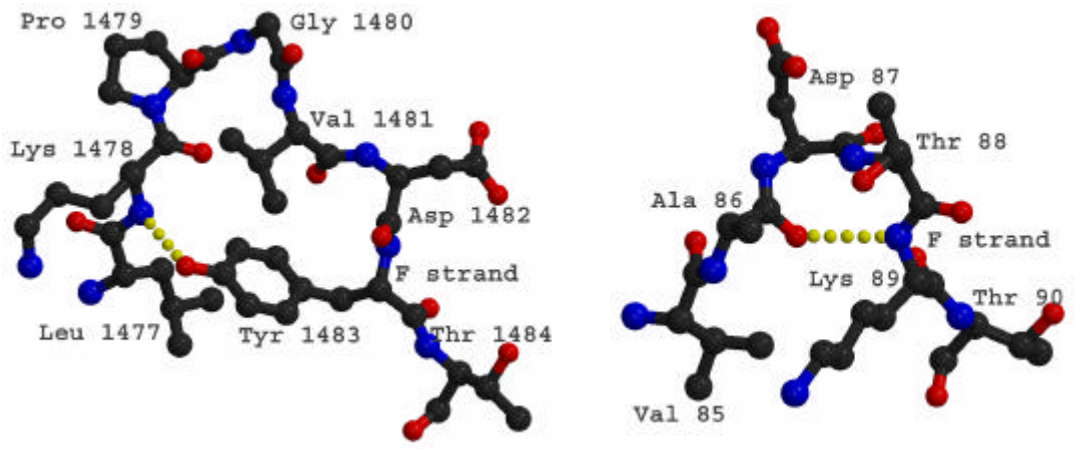


Figure 2.8. Another characteristic of the FNIII fold is the tyrosine corner at the junction of the E and F strands (left panel), which is missing in Fve, but is replaced by a β-turn (right panel).

backbone at this position (Hamill *et al*., 2000). Fve does not contain this tyrosine corner, but interestingly, this critical need is satisfied by the type I β-turn (Ala86-Asp-Thr-Lys89), with the backbone C=O of Ala86 forming an H-bond with the backbone N atom of Lys89 (Figure 2.8).

Structure-function relationships

Carbohydrate recognition. Fve hemagglutinates human red blood cell, and has a stimulatory effect on human peripheral blood lymphocytes (hPBLs) (Ko *et al*., 1995). It also suppresses systemic anaphylaxis reaction and local swelling of mouse footpads, and enhances the transcription of IL-2, IFN-? (Ko *et al*., 1995) and TNF-α (Seow *et al*., 2003). The hemagglutinating activity of Fve was not inhibited by simple sugars such as D-glucose, D-galactose, D-fucose, N-acetyl-D-galactosamine and D-lactose. However, the highly glycosylated porcine thyroglobulin was able to inhibit hemagglutination at a molar ratio of 1:4 (Fve: thyroglobulin), suggesting that Fve is a lectin with specificity for complex cell surface carbohydrates (unpublished observations). Another family of lectins containing an Ig fold is the sialic acid binding Ig-like lectins (Siglecs). These proteins bind specifically to sialylated glycoconjugates, and their carbohydrate binding site and ligand-binding determinants have been largely worked out (for example, see [Alphey *et al*., 2003]). However, since the binding specificity of Fve has not yet been established, comparative studies between Fve and Siglecs, to delineate its carbohydrate binding site, are presently of limited utility.

While there is no single template for recognition of carbohydrate-binding pockets it has been shown that certain amino acids show strong propensity to be in the

carbohydrate binding site of proteins (Taroni *et al*., 2000). They are the residues with aromatic rings (Phe, Trp and Tyr) that can pack against the hydrophobic face of the sugar residues, and Arg, Asp and Glu residues, and can form bidentate interactions with adjacent hydroxyls on the sugar (Taroni *et al*., 2000). Searching for surface patches which have a local combination of these residue types provides a good first diagnostic for a sugar-binding site (Taroni *et al*., 2000). Applying this system to Fve, we found two sets of patches that offer the potential to bind sugars. The first patch, consisting of residues Tyr11, Arg50, Asp 57, Phe94, Glu104, Glu105, Tyr 106 and Glu 110 is found on the same side of the dimer (Figure 2.9) The second patch, continuous with the first (and hence shares common residues), consists of Tyr 11, Trp24, Arg26, Asp87, Glu110 and Trp111. This pocket lies on opposite sides of the dimer (Figure 2.10).

Preliminary data indicates that this second pocket may be involved in sugar-binding as single-residue chemical modification of one of the two tryptophans, Trp24 and Trp111, (which are closely apposed to each other in the 3D structure) using 2,4-dihydroxybenzyl bromide significantly diminished hemagglutination activity (unpublished observation). Further experiments, including site-directed mutagenesis, are underway to further delineate the carbohydrate binding site and the ligand for Fve.

Dimerization by 3D domain swapping

The structure of Fve clearly shows an exchange of the N-terminal α-helices, H_A and H_B, between the symmetry-related monomers suggesting that dimerization may occur by 3D swapping. 3D domain swapping is a mechanism for forming oligomeric proteins from their monomers (for reviews, see [Bennett *et al*., 1995 and Schlunegger *et al*., 1997]).

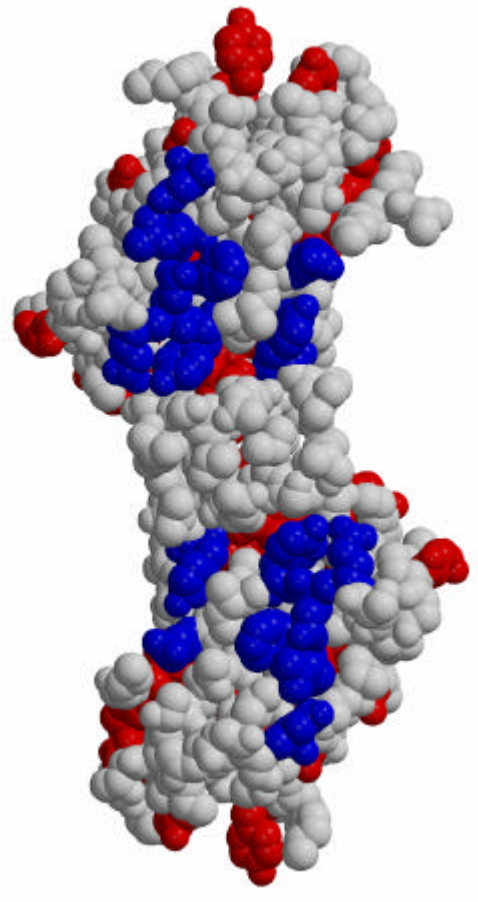


Figure 2.9. Potential carbohydrate-binding sites in Fve. All residues with strong propensity to be in the carbohydrate-binding site of proteins (see text) are shown in red. The first patch (colored blue), consisting of Tyr11, Arg50, Asp 57, Phe94, Glu104, Glu105, Tyr 106 and Glu 110 are found on the same side of the dimer.

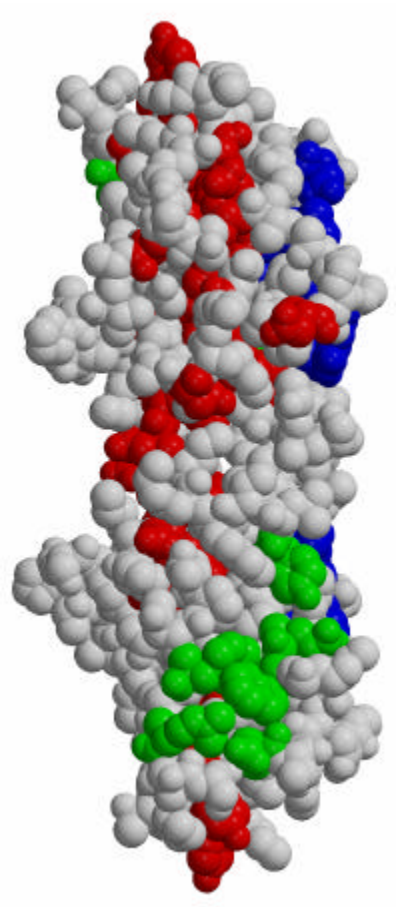


Figure 2.10. Potential carbohydrate-binding sites in Fve. All residues with strong propensity to be in the carbohydrate-binding site of proteins (see text) are shown in red. The second patch (colored green), which is continuous with the first (and hence shares common residues), consists of Tyr 11, Trp24, Arg26, Asp87, Glu110 and Trp111. This pocket lies on opposite sides of the dimer.

Since the structure of the Fve monomer is not known, it falls into the category of candidate for domain swapping.

There are three extended regions of subunit-subunit interactions within the dimer: (a) between H_A and H_B (residues 2 to 14), (b) between the N-terminal β-strands (S_A and $S_{B;}$ residues 14 to 19), and (c) between the N-terminal helices and the FNIII modules of the opposite monomer. Helices H_A and H_B are amphipathic; each has hydrophobic residues Ala 2, Leu 5, Leu 9, Leu 12 and Val 13 located on one face. The side chains of the residues on the hydrophobic face of one helix pack well against those of the other helix enabling them to bind strongly to each other via hydrophobic interactions (Figure 2.11). In Gts, another FIP with high sequence similarity to Fve, N-terminal residues 1-13 were predicted to form an α-helix (Lin *et al*., 1997). Recombinant mutants of FIP-gts with residues 1-13 deleted were incapable of dimerization (Lin *et al*., 1997). Further, triple mutants with Leu 5, Phe 7 and Leu 9 deleted lost the amphipathic characteristics of its putative N-terminal helix, the ability to form dimers with itself and the wild-type FIP-gts, as well as its immunomodulatory activity (Lin *et al*., 1997). The structure of Fve corroborates these predictions and emphasizes the important functional role that the N-terminal helix plays in dimerization and hence the functional lectin and immunomodulatory activity of FIPs. In addition to the hydrophobic interaction between helices H_A and H_B, the β-strands S_A and S_B run antiparallel to each other, forming a two-stranded β-sheet, further strengthening the interaction between the monomers (Figure 2.12). Both the inter-helix hydrophobic interaction and the hydrogen bonding between the sheets would not be found in the putative Fve monomer, but stabilize the dimer state. The N-terminal helices (H_A and H_B) interact with the FNIII modules of the opposite

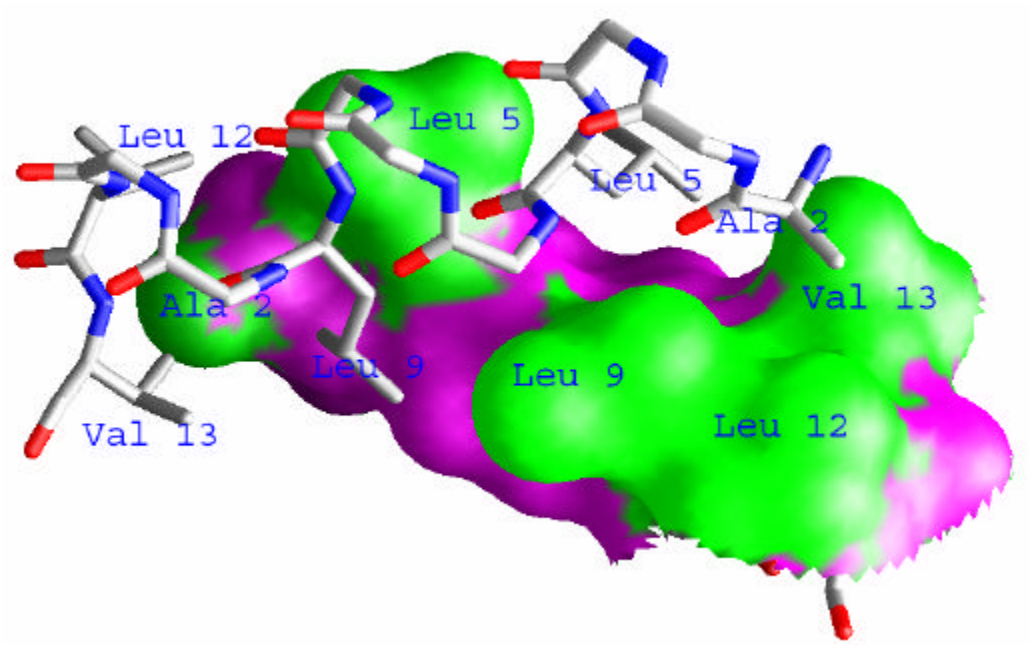


Figure 2.11. Extensive hydrophobic interactions occur between the N-terminal α-helices (H_A and H_B). The hydrophobic residues Ala 2, Leu 5, Leu 9, Leu 12 and Val 13 located on one face (green) pack well against the other helix.

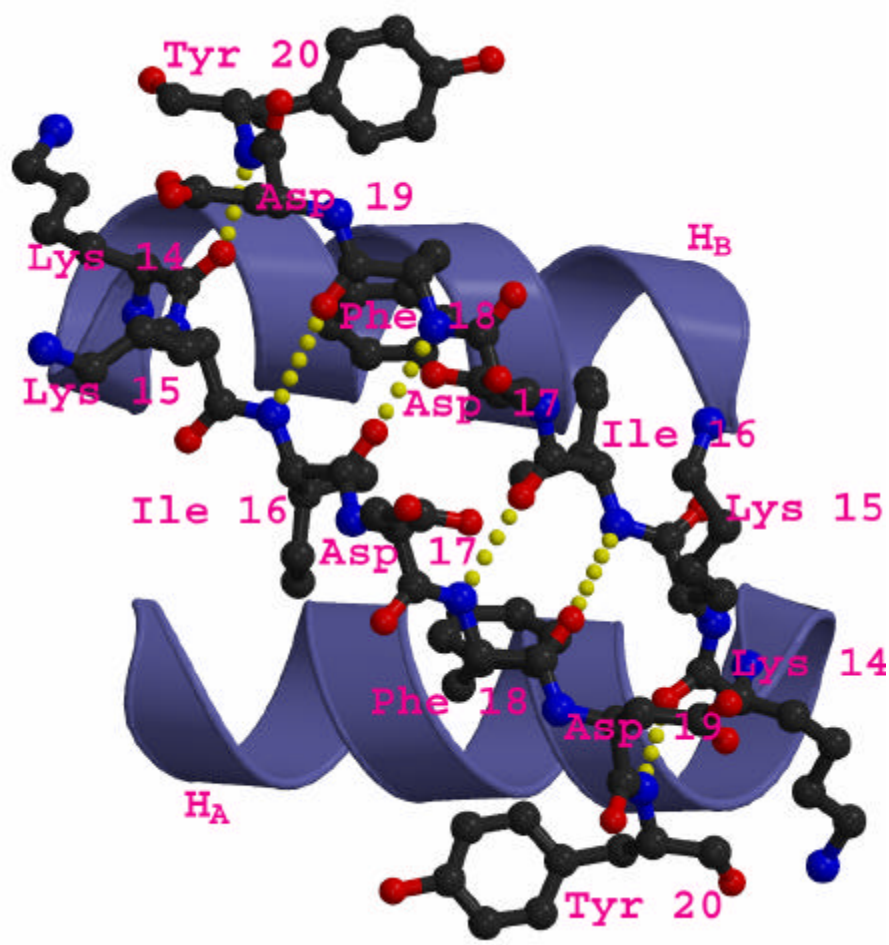


Figure 2.12. The N-terminal β-strands (S_A and S_B) also form a β-sheet (hydrogen bonds shown in yellow) further stabilizing the dimer.

monomers, by both direct hydrogen-bonding interactions (for example, between Thr 3 and Ser 4 of monomer A and Lys 45 and Tyr 106 of FNIII module B, respectively) as well as interactions via ordered water molecules (for example, between Ser 1 and Lys 14 of monomer A, and Glu 104 and Ile 108 of FNIII module B, respectively). In addition to these, hydrophobic interactions also exist, involving Phe 7, Ala 10 and Tyr 11 of monomer A and Tyr 106, Ile 107 and Ile 108 of FNIII module B (similar structural features occurring in another C-interface between swapped domain and the FNIII module A) (Figure 2.13). In a domain-swapped dimer, interaction between the N-terminal helix and the FNIII module would likely be representative of the interaction in the putative monomer prior to dimerization. This interaction cannot be too strong, otherwise displacement of the helix by that of another monomer during dimerization would be energetically unfavorable. Monomer Fve is modeled based upon the crystallographic determined dimer (Figure 2.14).

In the Fve monomer, the N-terminal α-helix (Ala2 to Lys14) is linked to the rest of the subunit by a hinge loop eight residues long, stretching from residue Lys15 to Pro22. (The average length of hinge loops in *bona fide* 3D domain-swapped proteins is nine residues). The distance of the C-α of Lys 15 to the C-α of Pro22 in the open monomer is 22.1 Å, and hence satisfies the distance criterion for the length of the hinge loop (< 29.7 Å) (Bennett *et al*., 1995). Finally, Pro22 can conceivably play a key role in the 3D domain swapping process as it occupies a critical position at the end of the hinge region (Figure 2.7). Prolines are frequently found in the hinge region of domain-swapped proteins particularly at the point where the arm leaves the protomer (Bergdoll *et al*., 1997). Sequence alignment and site-directed mutagenesis confirm the importance of these

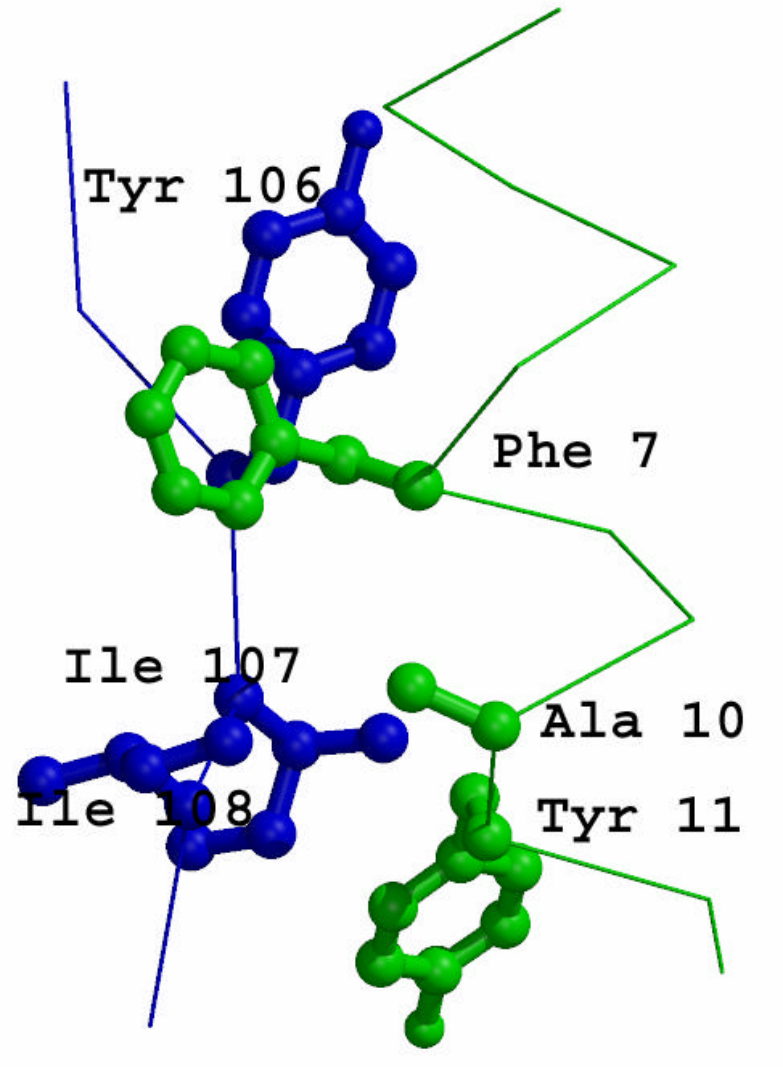


Figure 2.13. Hydrophobic interactions involve Phe 7, Ala 10 and Tyr 11 of the swapped domain (green) and Tyr 106, Ile 107 and Ile 108 of FNIII module (blue).

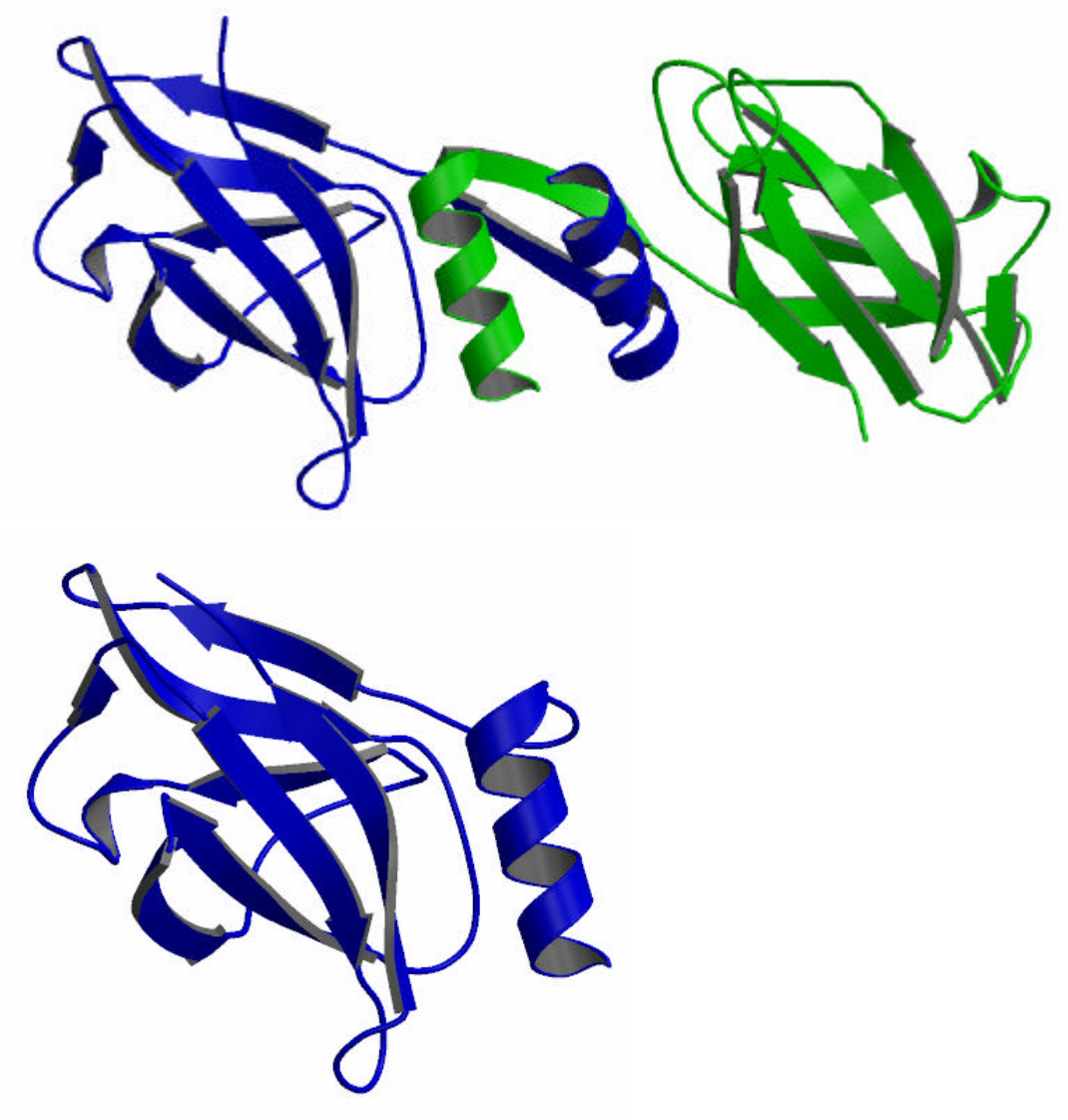


Figure 2.14. Crystallographic determined dimer and the model Fve monomer.

conserved prolines (Bergdoll *et al*., 1997). Their occurrence at the hinge regions can be explained by the constraints they impose on polypeptide conformation and dynamics: rigidifying the main chain, favoring extended conformations of arms and prevention of interaction of the arms with the protomer core (Bergdoll *et al*., 1997). The strain introduced by prolines in the hinge region enables them to behave as "molecular springs" that release tension by adopting alternative conformations while respecting the overall structure of the protein (Rousseau *et al*., 2001).

Dimerization between Ig-fold domains by domain-swapping mechanisms has been proposed in other proteins such as rat CD2 (Murray *et al*., 1995) and the human high affinity IgE receptor (Vangelista *et al*., 2002). However, in these cases, dimerization differs from that in Fve, in that it occurs by intertwining of the β-strands themselves that comprise the Ig-fold.

CHAPTER 3

Structural characterizations of venom of the Malayan pit viper: Rhodocetin

Many proteins with platelet aggregation and blood coagulation effects have been purified from snake venoms. They are a diverse group of molecules which differ in size and the mechanism of action. A group of non-enzymatic proteins from this set have structural homology to CLPs (Ca^{2+}-dependent lectin-like proteins). These proteins are generally heterodimers of MW ~30000 Daltons containing a disulfide bridge between the 2 subunits. Although they share significant sequence and structural similarity, they display varying effects on blood coagulation and platelet aggregation. Some of these proteins, for example, bind coagulation factors X and/or IX such as ECLV IX/X-bp from *Echis carinatus leucogaster* venom (Chan *et al*., 1996), jararaca IX/X-bp from *Bothrops jararaca* venom (Sekiya *et al*., 1993), and habu IX/X-bp and IX-bp from *Trimeresurus flavoviridis* venom (Atoda *et al*., 1991 and Atoda *et al*., 1995). In contrast, bothrajaracin from *Bothrops jararaca* venom (Zingali *et al*., 1993) is a specific inhibitor of thrombin.

A novel platelet aggregation inhibitor, rhodocetin, was purified from the crude venom of *Calloselasma rhodostoma.* It inhibited collagen-induced platelet aggregation in a dose-dependent manner, with an IC_{50} of 41 nM (Wang *et al*., 1999). The complete amino acid sequences of the α (15956.16 Da, 133 residues) and β (15185.10 Da, 129 residues) subunits share a high degree of homology with each other (49% identity) and with the Ca^{2+}-dependent lectin-like proteins (CLPs) (typically 29-46%) isolated from other snake venoms. Rhodocetin is a heterodimer, and individually both α and β subunits fail to show significant inhibition of aggregation (Wang *et al*., 1999). Unlike the other members of the CLP family in which the subunits are held together by an inter-chain disulfide bond, the subunits of rhodocetin are held together only by non-covalent interactions. The cysteinyl residues forming the inter-subunit disulfide bridge in all other

known CLPs are replaced by Ser 79 and Arg 75 in the α and β subunits of rhodocetin, respectively (Wang *et al*., 1999).

We determined the three-dimensional structure of rhodocetin to understand structure-function relationships in this family of proteins.

MATERIALS AND METHODS

Protein purification

Calloselasma rhodostoma crude venom (100 mg) was dissolved in 3.0 mL of 0.1 M ammonium hydrogen carbonate (pH 8.0) and centrifuged at 12000 rpm for 10 min at 4 ºC to remove particulate material. (a) The supernatant was then fractionated by a HiLoad Superdex 75 column (2.6 × 60 cm) equilibrated with the same buffer using a FPLC system. (b) The fractions from peak III (containing proteins of interest) were pooled and loaded directly onto a Mono Q HR 5/5 anion exchange column pre-equilibrated with 20 mM Tris-HCl, pH 8.2. (c) Fractions from peak 2 of Mono Q were buffer exchanged with 50 mM potassium phosphate, pH 7.0, containing 1.7 M ammonium sulfate (NAP-25 column) before being loaded onto a Phenyl-Superose HR 5/5 column equilibrated with the same buffer. Elution was performed at a flow rate of 0.5 mL/min with a linear gradient of 1.7-0.85 M ammonium sulfate in 50 mM potassium phosphate, pH 7.0, over 30 min.

Protein crystallization and data collection

Well-diffracting rhodocetin crystals were grown by vapor diffusion from hanging drops at 0.7 M NaH_2PO_4, 0.7 M KH_2PO_4, 0.1 M HEPES-Na, pH 7.5 at 21 °C . Native crystals

were prepared by soaking the crystals in mother liquor containing 25% glycerol after which the crystals were flash-frozen at 100 K. Native data from a crystal was collected at the National Synchrotron Light Source (NSLS) beam line X8C at the wavelength 0.978569 Å. The crystal diffracted to 1.9 Å. It belonged to the orthorhombic space group $P2_12_12_1$ and had unit cell dimensions a = 46.875, b = 65.935, c = 118.841 Å and $\alpha = \beta = \gamma = 90.0^{o}$. The data were processed and scaled using DENZO and SCALEPACK from the HKL2000 suite of programs (Otwinowski *et al*., 1997).

Structure solution and refinement

The structure was solved by the molecular replacement progam MOLREP (Vagin *et al*., 1997). The A-chain of the coagulation factor IX-bp protein was chosen as a search model based on its high sequence identity to rhodocetin. The six N-terminal residues as well as residues 69 - 102 were deleted from the search model and two monomers were searched for in the asymmetric unit. A top peak was obtained after translation, with a correlation coefficient of 27.5%. The resulting electron density map revealed distinct secondary structure elements. The phases obtained by molecular replacement were directly used in ARP/wARP (Morris *et al*., 2002) for automated main chain tracing, which resulted in the building of six continuous fragments with connectivity index 0.95 that contained 245 residues. The rest of the model and side chains were fitted manually using XtalView (McRee *et al*., 1999). Refinement was carried out with REFMAC5 (Murshudov *et al*., 1999) using data in the resolution range of 30.00 Å- 1.9 Å and water molecules were identified using ARP/WARP later in the refinement. Data collection and refinement statistics are given in Table 3.1. The quality of the final model was verified with

Table 3.1. Data collection and refinement statistics.

Data collection

Space group	$P2_12_12_1$
Unit cell parameters (Å, °)	a = 46.87, b = 65.93, c = 118.84, $\alpha = \beta = \gamma$ = 90.0
Resolution range (Å)	30.0 - 1.90
Number of observed reflections	184500
Number of unique reflections	28995 (*2304*)
Redundancy	6.36
Completeness (%)	97.20 (*78.7*)
R_{sym} (%; all reflections)	0.060 (*0.307*)

Model refinement

Number of protein atoms	2104
Number of water molecules	167
R-factor	0.189
Rfree (test set of 5%)	0.231
Data cutoff in σ units	none
RMS bond length	0.020
RMS bond angle	1.672
RMS chiral	0.175

Values for parameters in the highest resolution shell, 1.97 – 1.90 Å are parenthesized and in italics

PROCHECK (Laskowski *et al*., 1993) (Figure 3.1). The Ramachandran plot shows that 92.1% of the residues fall in the most favored region and the rest in the additional allowed region except the Thr 7 in the β subunit, which falls in the generously allowed region (Figure 3.2). In total, two subunits: (1) α subunit with 132 residues (2) β subunit with 124 residues and 168 solvent molecules were built into the high quality electron density map. The structure was refined at 1.9 Å to an R-factor of 18.9% and an R-free of 23.1%. The atomic co-ordinates and structure factors have been deposited in the Protein Data Bank (Berman *et al*., 2000) (PDB) as entry 1SB2. MolScript was used for molecular drawing.

RESULT AND DISCUSSION

Overall fold

The rhodocetin dimer consists of two subunits, α and β containing 133 and 124 residues, respectively and shows 49 % identity with each other. Interestingly, the structure of the two subunits (α and β) superimpose almost perfectly except for the extended loop and the loop region following the first α-helix: the RMSD between their C_α atoms is 1.04 Å. In the α subunit, the loop region following the first α-helix forms a ?-turn, Gln 33 – Lys 35. In contrast to the α subunit, the loop region in the β subunit forms a β-turn, Ala 36 – Gly 39. Each subunit contains 3 intra-subunit disulfide bonds: Cys 2 – Cys 13, Cys 30 – Cys 127 and Cys 102 – Cys 119 in the α subunit; and Cys 4 – Cys 15, Cys 32 – Cys 123 and Cys 98 – Cys 115 in the β subunit. The disulfides in each subunit are found in similar 3D positions in both subunits when the structures are superimposed (Figure 3.3).

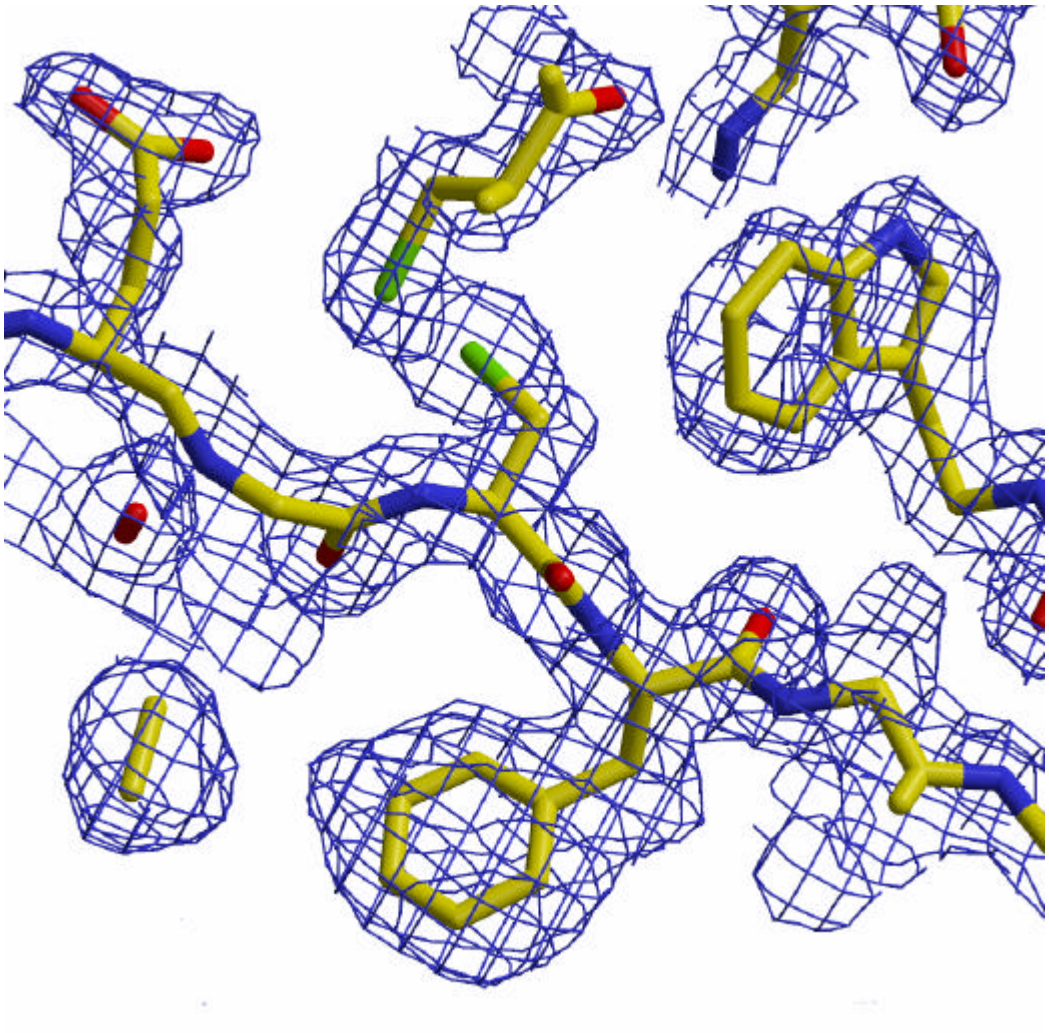


Figure 3.1. Electron density map. A portion of 2Fo-Fc map after the final refinement demonstrating the quality of the structure.

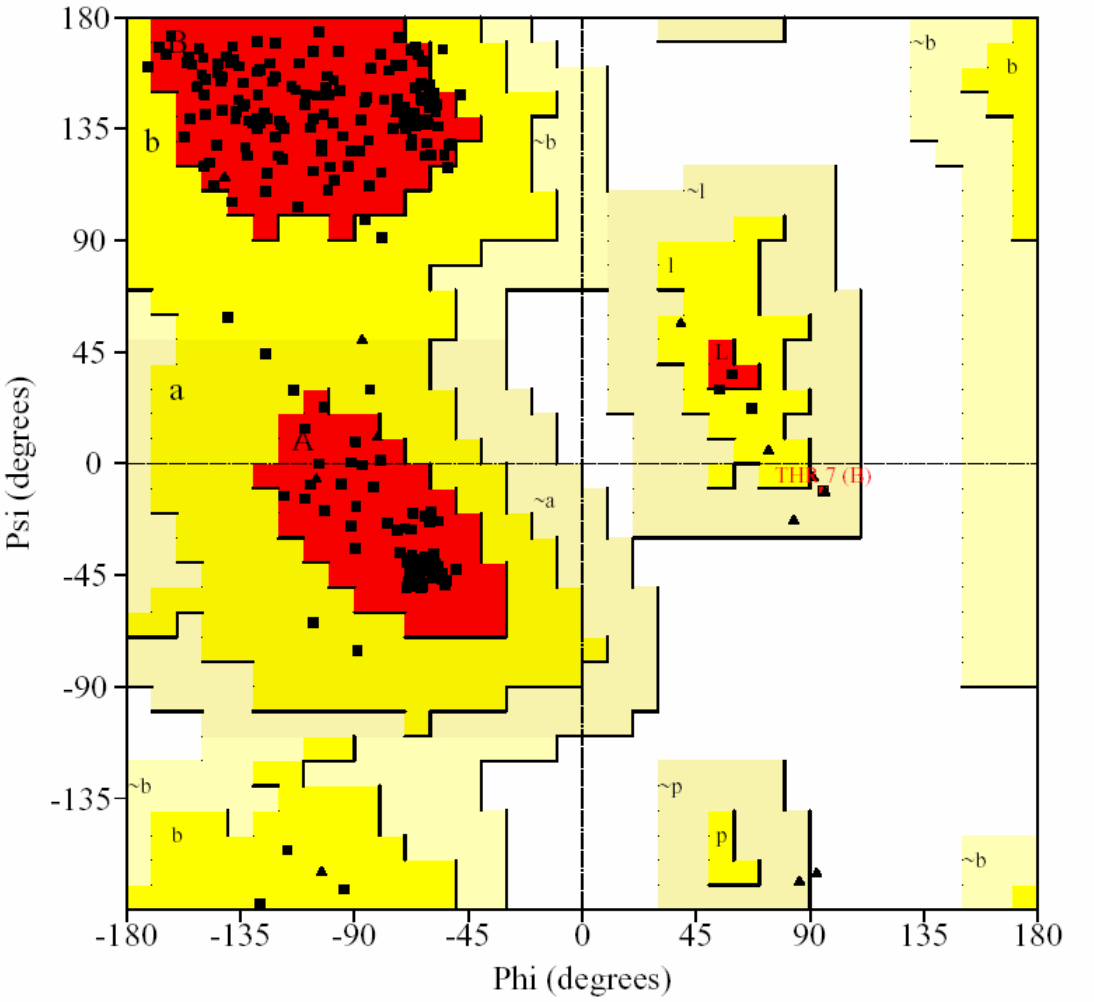


Figure 3.2. Thr 7 in the β subunit falls in the generously allowed region of the Ramachandran plot.

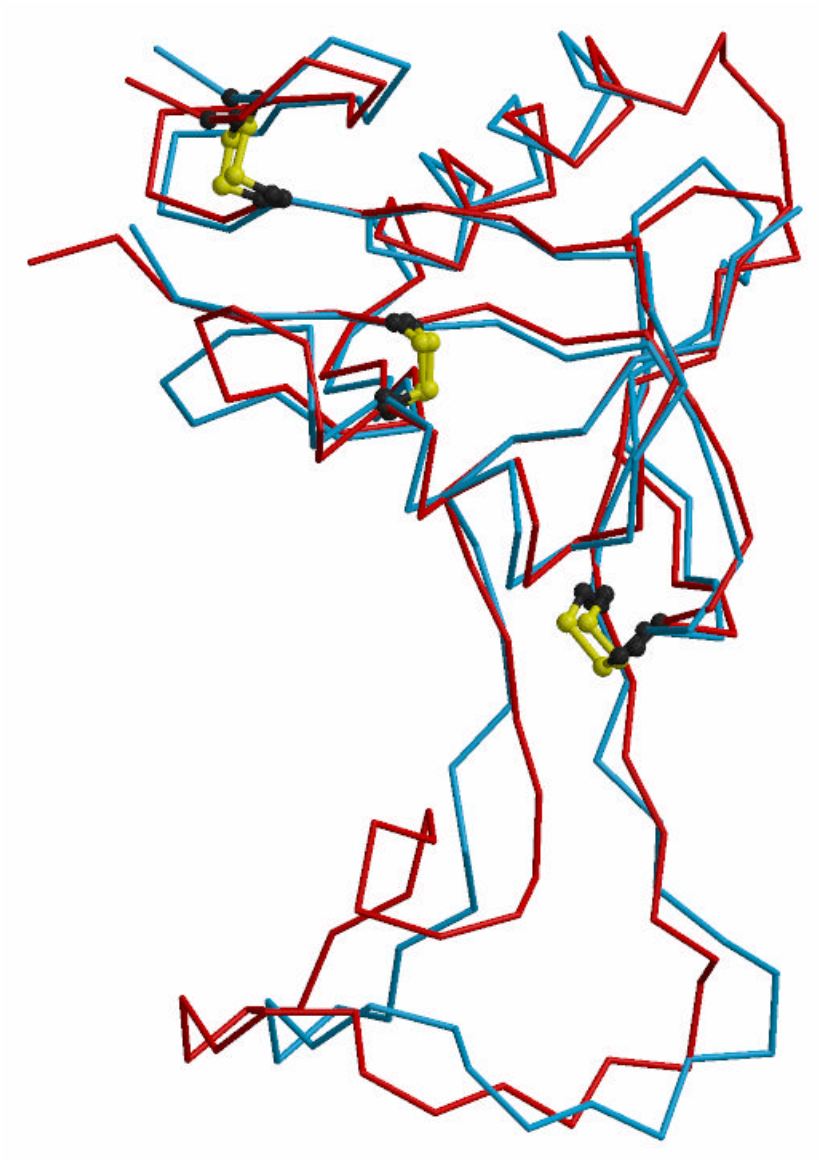


Figure 3.3. Superposition of subunits. The disulfides in each subunit are found in similar positions in the 3D structure. The α subunit is in red and β subunit in blue. Disulfide bonds are colored yellow.

The structure of rhodocetin comprises of a globular unit with an extended loop and forms a CLP fold. The α subunit contains eight β-strands and two α-helices, with seven strands forming two antiparallel β-sheets, while the β subunit contains seven β-strands, six forming two antiparallel β-sheets, and two α-helices. It is important to note that the β-strand spanning four residues from Asn 80 to Ser 84 of α subunit runs antiparallel to the β-strand spanning four residues from Gly 71 to Arg 75 of β subunit to form a two-stranded β-sheet which is a critical secondary structure element for dimerization (Figure 3.4). In addition to this β-sheet, the extended loops also form interactions. The backbone carbonyls of Gly 69, Asn 93 and Trp 114 in the α subunit form hydrogen bonds with the backbone amides of Thr 80, Trp 110 and Tyr 91 in the β subunit, while the backbone amides of Tyr 95 and Trp 114 in the α subunit form hydrogen bonds with the backbone carbonyls of Trp 110 and Asn 89 in the β subunit.

Structure comparison with C-type lectins

A structure similarity search using the DALI (Holm *et al*., 1993) server indicated that the α and β subunits of rhodocetin belong to the C-type lectin super family. Eighteen known structures showed similarity to rhodocetin with a Z score > 2.0. In particular, IX/X-bp (Mizuno *et al*., 1997), lithostathine (Bertrand *et al*., 1996), CRD-4 (Feinberg *et al*., 2000) and tetranectine (Nielsen *et al*., 1997) were most similar, with Z scores > 13.0. Superposition of IX/X-bp, lithostathine, CRD-4 and tetranectine with rhodocetin revealed that major differences in their structure reside in the central loop region. The lithostathine and tetranectine are monomeric C-type lectin whereas IX/X-bp, CRD-4 and rhodocetin are

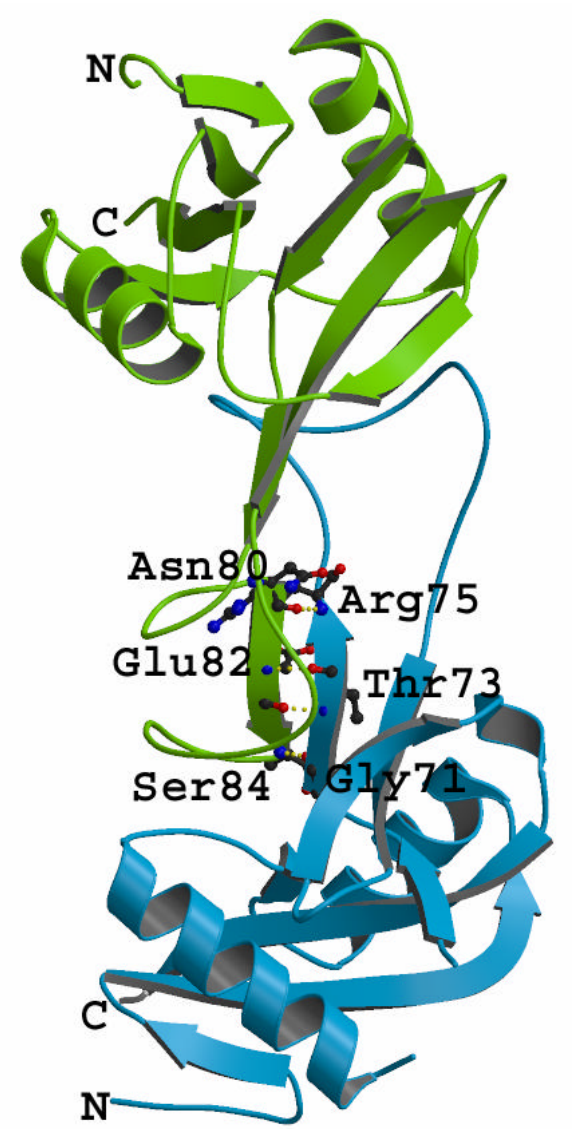


Figure 3.4. Novel interface of rhodocetin. Residues from Asn 80 to Ser 84 of the α subunit run antiparallel to the β-strand spanning four residues from Gly 71 to Arg 75 of the β subunit to form a two-stranded β-sheet. This is a critical secondary structure element for dimerization. Hydrogen bonds are shown in yellow. α subunit is green and β subunit in blue.

dimers. In a monomeric C-type lectin, the cental loop folds back into the body while in rhodocetin it forms an extended central loop from the body as seen in IX/X-bp and CRD-4 (Mizuno *et al*., 1997). Furthermore, superposition of IX/X-bp of subunit A with rhodocetin α subunit shows that the significant deviation in the structure is due to the rotation of the loop by approximately 90º. The rotated loop spans 27 residues, from residue 73 to residue 100 (Figure 3.5).

A comparison of the 3-Dimensional structures of the heterodimers, coagulation factor IX/X-bp, coagulation factor X-bp and botrocetin shows that only local differences exist between them, that are mainly due to insertions/deletions of residues. Besides, all the heterodimers are linked by a disulfide bond (Mizuno *et al*., 1999). In both coagulation factor IX/X-bp and X-bp, the inter-chain disulfide bond occurs between Cys 79A and Cys 75B, while in botrocetin, it occurs between Cys 80A and Cys 75B. The residues in homologous positions to these Cys residues are Ser 79 and Arg 75 in the α and β subunits of rhodocetin, respectively (Wang *et al*., 1999). Moreover, rhodocetin has a unique interface as it lacks the canonical inter-subunit disulfide bridge. This novel interface of rhodocetin allows non-covalent interactions to compensate for the missing disulfide bond. There is an ordering of the loop regions from Asn 80 to Ser 84 of the α subunit and from Gly 71 to Arg 75 of the β subunit to create an extra β-sheet not previously observed in other CLPs at the interface. The backbone carbonyls of Asn 80A and Glu 82A form hydrogen bonds with the backbone amides of Arg 75B and Thr 73B, whereas the backbone amides of Glu 82A and Ser 84A participate in forming hydrogen bonds with the backbone carbonyls of Thr 73B and Gly 71B. It is most likely that this interface has

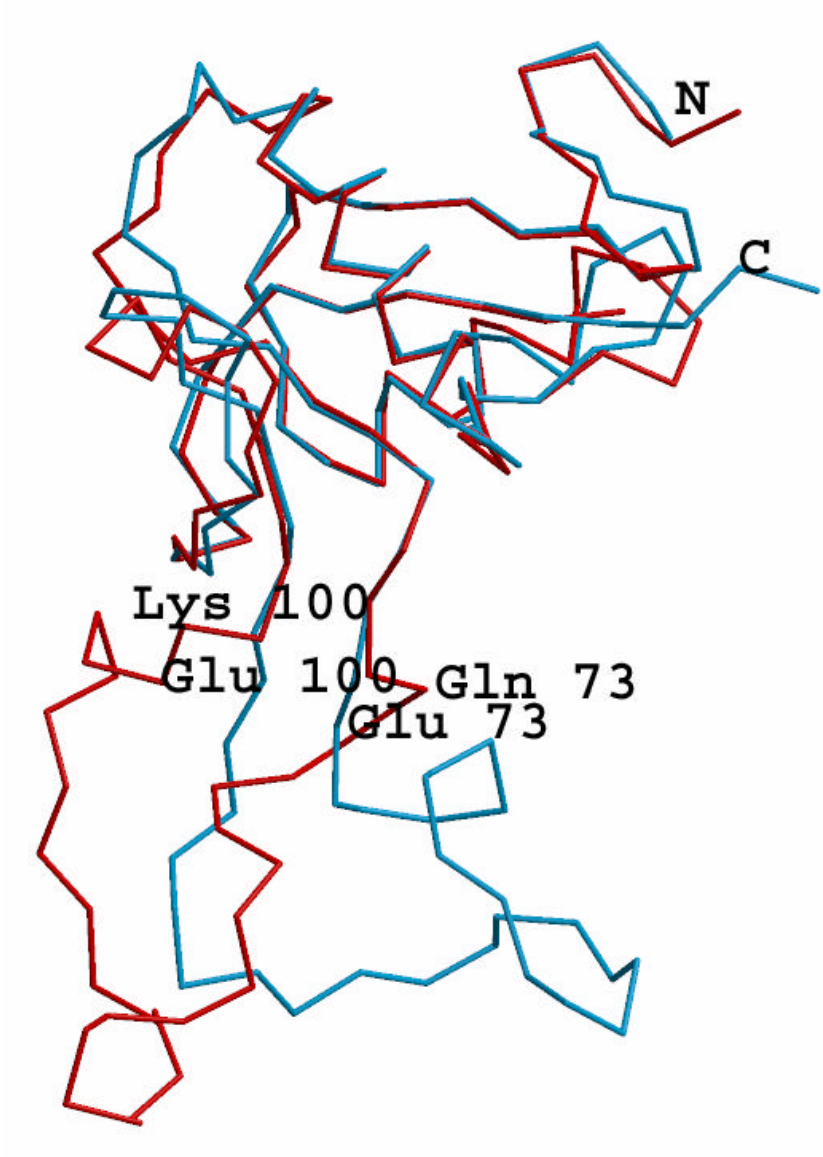


Figure 3.5. Superposition of IX/X-bp of subunit A with rhodocetin α subunit shows that the loop in rhodocetin spanning 27 residues is rotated by ~90º with respect to the similar loop in IX/X-bp loop. IX/X-bp is in red and rhodocetin is in blue.

disordered residues that create an ordered interface after interacting and subsequently present a surface that may then play a role in inhibiting platelet aggregation.

Structure-function relationships

Lack of metal binding site. Calcium ions are essential to the function and the crystallization of IX-bp protein. There are only a few examples of proteins crystallized without calcium ions that require Ca^{2+} to function. Intrinsic fluorescence analysis of IX/X-bp in the presence of calcium ions indicated that Ca^{2+}-binding induces conformational changes which stabilize the structure of the protein (Mizuno *et al*., 1999). In contrast inhibition of platelet aggregation by rhodocetin was shown to be independent of any metal ligand (Kong, 2002). Moreover, this protein can crystallize without metal ions. This indicates that rhodocetin does not need metal ligands to fold into an ordered structure in order to preserve its function.

The structural comparison of IX-bp, IX/X-bp and botrocetin Ca^{2+}-binding residues with rhodocetin revealed some interesting features. The IX-bp and IX/X-bp showed that each subunit of these proteins had a Ca^{2+}-binding site. In botrocetin, only the subunit B was bound to the divalent metal. Interestingly, both subunits of rhodocetin lost the ability to bind calcium ion. The loss of the metal ion binding site in A-subunit of botrocetin is due to the replacement of residues Glu 43 and Glu 128 in IX/X-bp and IX-bp with Lys 43 and Lys 128 (Sen *et al*., 2001). In IX-bp and IX/X-bp, residues Ser 41, Glu 43, Glu 47 and Glu 128 in subunit A, and residues Ser 41, Gln 43, Glu 47 and Glu 120 in subunit B coordinate the binding of the calcium ions (Mizuno *et al*., 1999). In subunit α of rhodocetin residues Ser 41, Glu 43 and Glu 47 occupy the similar coordinate position of

IX/X-bp subunit A. However, the rhodocetin α subunit has Lys 128 instead of Glu 128 of IX/X-bp subunit A. In a similar manner, rhodocetin β subunit residues Gly 45 and Lys 124 replace Gln 43 and Glu 120 of IX/X-bp subunit B, respectively (Figure 3.6). These substitution of Lys in both subunits of rhodocetin give raise to a positive charge instead of a negative charge and thus prevents the metal ion from binding rhodocetin by charge-charge repulsion (Figure 3.7). Furthermore, the Lys 128 residue of α subunit is involved in the formation of a salt bridge with Glu 49, while the residue Lys 124 of the β subunit forms a salt bridge with Glu 47. These salt bridges could stabilize the fold in the absent of Ca^{2+}-binding site, and hence form a functional CLP heterodimer protein.

Differences to Eble rhodocetin. We have also clearly established that rhodocetin mentioned in this thesis as well as that of Wang *et al*., 1999 is a distinct entity from the rhodocetin reported by Eble *et al*., 2001. Rhodocetin reported by Eble *et al*. was in fact shown to be a dimer under nonreducing condition in SDS-PAGE, but fell upon reducing condition in SDS-PAGE. The Wang *et al*. rhodocetin protein exhibited two subunits in SDS-PAGE in the presence or absence of a reducing agent. We additionally note that the N-terminal amino acid sequence determined by Eble et al is different with respect to our sequence. We checked the electron density, in particular, for residue 10 in the β subunit, and found that it corresponded to an Ala (consistent with our rhodocetin sequence), and not Met (in rhodocetin reported by Eble *et al*.). The Eble rhodocetin protein had integrin binding activity which was not observed with our rhodocetin. Therefore these molecules should be regarded as distinct entities. Structural comparison of the two similar rhodocetin proteins may help to pinpoint regions in these proteins critical for their different activities.

A

```
             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                 5          15         25         35         45         55         65
IX/X-bp      DCLSGWSSYE GHCYKAFEKY KTWEDAERVC TEQAKGAHLV SIES-SGEAD FVAQLVTQNM KRLDFYIWIG
IX-bp        DCPSGWSSYE GHCYKPFKLY KTWDDAERFC TEQAKGGHLV SIES-AGEAD FVAQLVTENI QNTKSYVWIG
Botrocetin   DCPSGWSSYE GNCYKFFQQK MNWADAERFC SEQAKGGHLV SIKIYSKEKD FVGDLVTKNI QSSDLYAWIG
Rhodocetin   DCPDGWSSTK SYCYRPFKEK KTWEEAERFC TEQEKEAHLV SMEN-RLEAV FVDMVMENNF ENKIYRSWIG

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....
                 75         85         95        105        115        125
IX/X-bp      LRVQGKVKQC NSEWSDGSSV SYENWIEAES KTCLGLEKET DFRKWVNIYC GQQNPFVCEA ----
IX-bp        LRVQGKEKQC SSEWSDGSSV SYENWIEAES KTCLGLEKET GFRKWVNIYC GQQNPFVCEA ----
Botrocetin   LRVENKEKQC SSEWSDGSSV SYENVVERTV KKCFALEKDL GFVLWINLYC AQKNPFVCKS PPP-
Rhodocetin   LKIENKGQRS NLEWSDGSSI SYENLYEPYM EKCFLMDHQS GLPKWHTADC EEKNVFMCKF QLPR
```

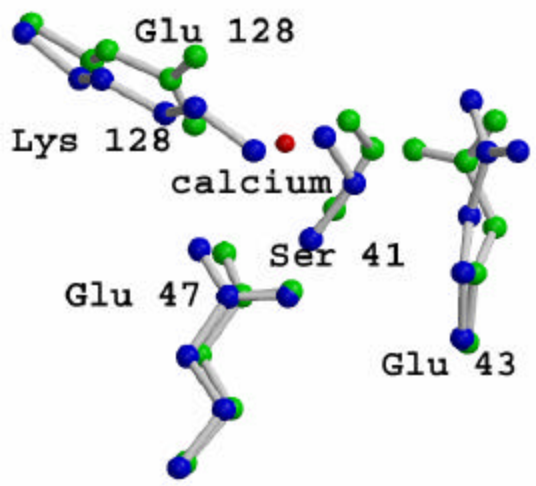
B

```
                 5          15         25         35         45         55         65
IX/X-bp      --DCPSDWSS YEGHCYKPFS EPKNWADAEN FCTQQHAGGH LVSFQSSEEA DFVVKLAFQT FGHS--IFWM
Ix-Bp        --DCPSDWSS YEGHCYKPFS EPKNWADAEN FCTQQHAGGH LVSFQSSEEA DFVVKLAFQT FGHS--IFWM
Botrocetin   --DCPPDWSS YEGHCYRFFK EWMHWDDAEE FCTEQQTGAH LVSFQSKEEA DFVRSLTSEM LKGD--VVWI
Rhodocetin   DFRCPTTWSA SKLYCYKPFK EKKTWIEAER FCAKQAENGH LVSIGSAAEA DFLDLVIVVN FDKQRYRAWT

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ...
                 75         85         95        105        115        125
IX/X-bp      GLSNVWNQCN WQWSNAAMLR YKAW--AEES YCVYFKSTN- -NKWRSRACR MMAQFVCEFQ A--
Ix-Bp        GLSNVWNQCN WQWSNAAMLR YKAW--AEES YCVYFKSTN- -NKWRSRACR MMAQFVCEFQ A--
Botrocetin   GLSDVWNKCR FEWTDGMEFD YDDYYLIAEY ECVASKPTN- -NKWWIIPCT RFKNFVCEFQ A--
Rhodocetin   GLT----ERN LKWTNGASVS YENLYEPYIR KCFVVQPWEG KSKWYKADCE EKNAFLCKFP KPH
```

Figure 3.6. Sequence alignment of IX/X-bp, IX-bp and botrocetin with rhodocetin. (A) Alignment of A-chain. (B) Alignment of B-chain. The potential calcium binding residues are highlighted in red. Non-calcium binding residues in homologous positions are in blue.

A



B

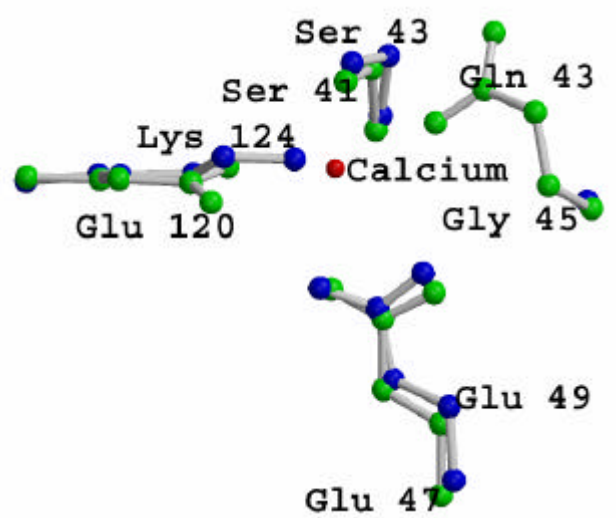


Figure 3.7. Geometry around the calcium binding site in CLPs. (A) Subunit A. (B) Subunit B. Superposition of the calcium ion coordination site in IX/X-bp (green) with the corresponding site in rhodocetin (blue). Glu 128A and Glu 120B of IX/X-bp are replaced by Lys 128 in the α subunit and Lys 124 in the β subunit, respectively. This substitution could prevent the metal ion from binding rhodocetin.

Dimerization by 3D domain swapping.

The structure of rhodocetin clearly shows that the exchange of the extended loop region, between the α and β subunits indicates that dimerization is likely accomplished through 3D swapping. If mannose-binding protein (Weis *et al*., 1991) is taken as the reference monomer as described for IX/X-bp (Mizuno *et al*., 1997), rhodocetin falls into the category of quasidomain-swapped dimer. Domain swapping in the C-type lectin superfamily is dominated by characteristic hydrophobic interactions. These specific interactions between the loop and the body of the adjoining subunit form the C-interface (Sen *et al*., 2001). In rhodocetin, the C-interface involves Leu 81, Trp 83, Ile 89, Tyr 91 and Leu94 on the swapped loop of the α subunit, and Ile 44, Ala 47, Leu 72, Phe 99 and Trp110 on the body side of the β subunit (Figure 3.8). Likewise, the other C-interface consists of Leu 77, Trp 79, Val 85, Tyr 87 and Leu 90 on the swapped loop of the β subunit and the Ala 48, Leu 70, Ile 72, Phe 103 and Trp 114 on the body side of the α subunit. In addition to the hydrophobic interactions, hydrogen-bonding interactions are also observed in the C-interface: the α subunit carbonyl Asn 93 and amide Tyr 95 interact with the amide of Trp 110 and the carbonyl of Trp 110 of the β subunit, respectively. Similar structural features occur between the loop of the β subunit and the body of the α subunit.

The characterization of rhodocetin by gel filtration showed that it is a dimer in solution. Its two subunits are linked by non-covalent interactions. Both subunits are essential for the activity of rhodocetin; neither the α subunit nor the β subunit in isolation inhibits platelet aggregation (Wang *et al*., 1999). Topological comparison of the CLPs

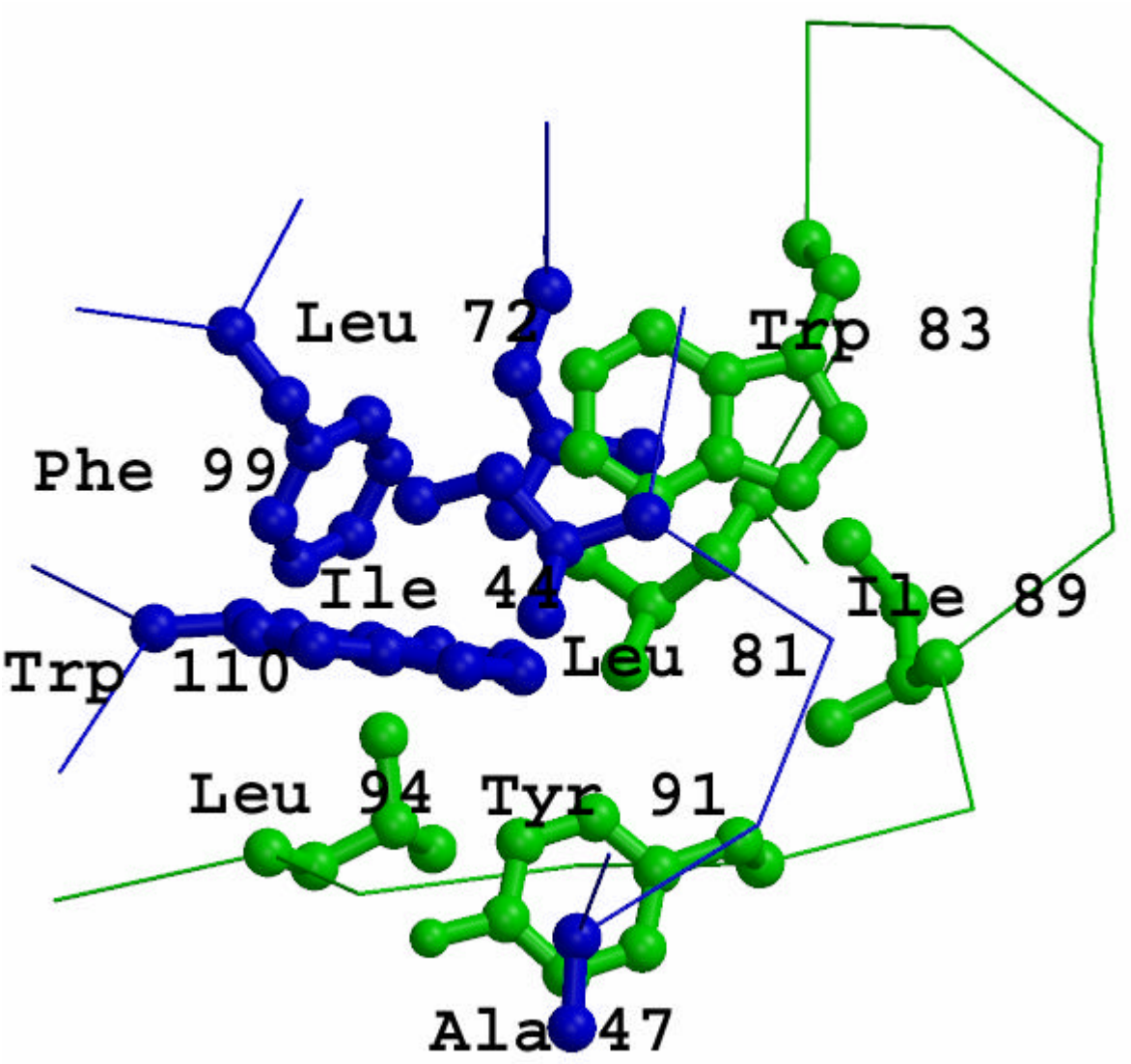


Figure 3.8. One of the C-interfaces of rhodocetin. Residues on the body side are indicated in blue and residues located in the swapped loop are in green.

IX/X-bp, IX-bp, lithostathine, E-selectin, tetranectin and mannose-binding protein, showed that C-terminal side of the hinge loop in the dimers is six or more residues shorter than that of the the monomeric proteins. If a hinge loop is shortened by a deletion, then the closed monomer structure may no longer be possible in geometry and the resultant open monomer may be unstable because of the exposure of residues normally buried in the C-interface. Domain-swapped dimers would then be favored (Mizuno *et al*., 1999 and Bennett *et al*., 1995). Comparison of the rhodocetin C-terminal side hinge loop with the mannose-binding protein revealed that the C-terminal side hinge loop of rhodocetin is six residues shorter than that of the monomeric protein. As a result, the central loop is swapped and subsequently a functionally heterodimer is formed (Figure 3.9).

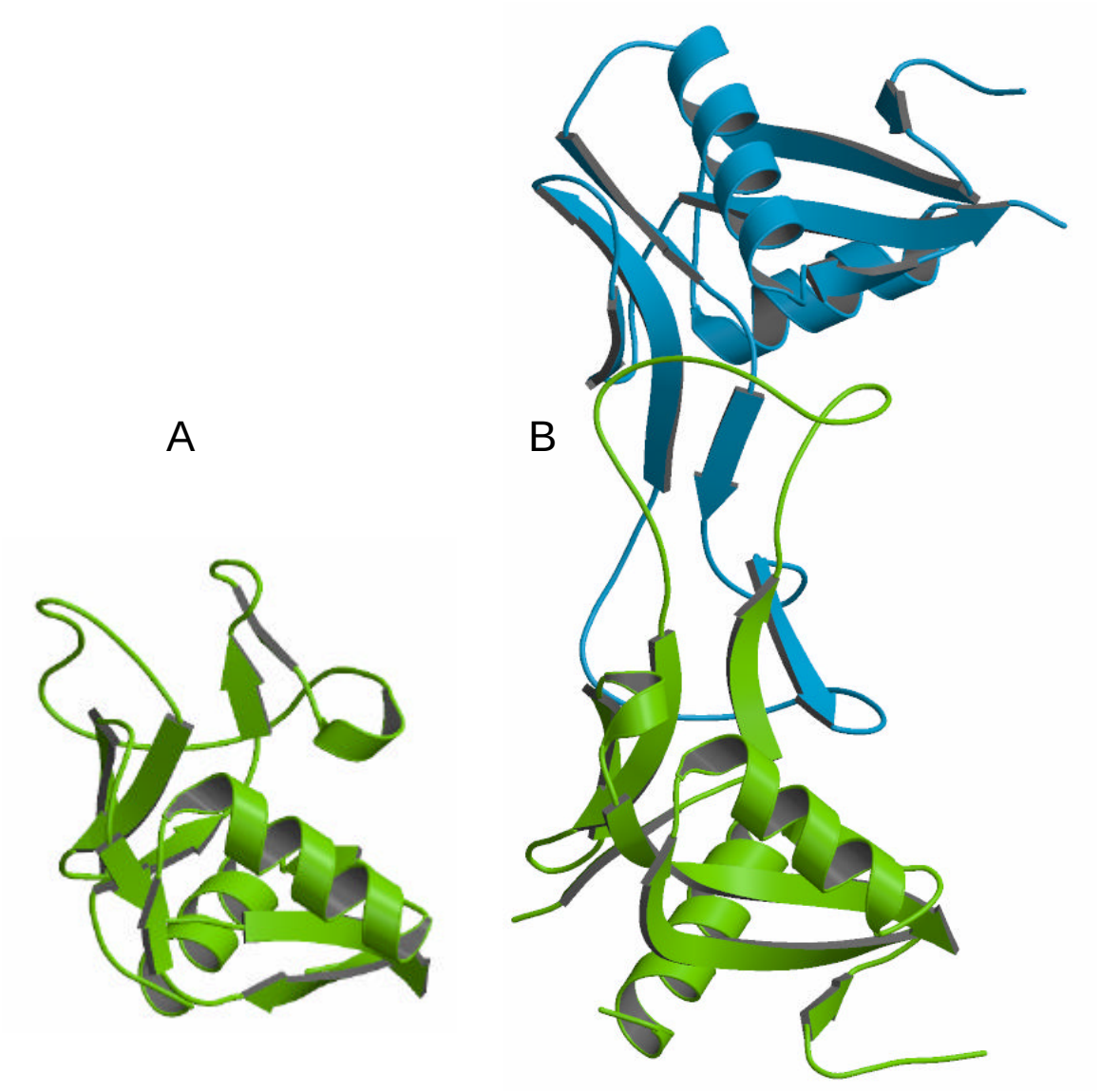


Figure 3.9. Mannose-binding protein is taken as the reference monomer (A). Comparison of the rhodocetin C-terminal side hinge loop with the mannose-binding protein revealed that the C-terminal side hinge loop of rhodocetin is six residues shorter than that of the monomeric protein. As a result, the central loop is swapped and subsequently a functional rhodocetin is formed (B).

CHAPTER 4

Conclusion

The golden needle mushroom, *Flammulina velutipes*, is a popular edible mushroom in the Orient, possessing anti-tumor, antiviral, antifungal and cholesterol-lowering activities. A major fruiting body protein, designated Fve or FIP-fve, was isolated and very likely plays a significant role in the mushroom's immunomodulating effects. It stimulates mitogenesis of human peripheral lymphocytes, suppresses systemic anaphylaxis reactions and local swelling of mouse foot pads, and enhances the transcription of interleukin-2 (IL-2), interferon-γ (IFN-γ) and tumor necrosis factor-α (TNF-α). In addition, it hemagglutinates human red blood cells. Fve is a non-covalently linked homodimer, each subunit 114 amino acids in length. It contains no Cys, Met or His residues and its N-terminal Ser residues are acetylated. It shows sequence similarity only to other mushroom proteins with similar functional characteristics: LZ-8 (*Ganoderma lucidum*), Gts (*Ganoderma tsugae)*, Vvo (*Volvariella volvacea*) and Vvl (*Volvariella volvacea*). All these proteins form the novel Fungal Immunomodulatory Protein (FIP) family. Furthermore, all dimers were thought to be formed by a mechanism called 3D domain swapping.

Many proteins with platelet aggregation and blood coagulation effects have been purified from snake venoms. They are a diverse group of molecules which differ in size and mechanism of action. These proteins are generally heterodimers of MW ~30000 containing a disulfide bridge between the 2 subunits. Although they share significant sequence and structural similarity, they display varying effects on blood coagulation and platelet aggregation. Some of these proteins, for example, bind coagulation factors X and/or IX such as ECLV IX/X-bp from *Echis carinatus leucogaster* venom, jararaca IX/X-bp from *Bothrops jararaca* venom, and habu IX/X-bp and IX-bp from *Trimeresurus flavoviridis* venom. In contrast, bothrajaracin from *Bothrops jararaca* venom is a specific

inhibitor of thrombin. A novel platelet aggregation inhibitor, rhodocetin, was purified from the crude venom of *Calloselasma rhodostoma.* It inhibited collagen-induced platelet aggregation in a dose-dependent manner. The complete amino acid sequences of the α (15956.16 Da, 133 residues) and β (15185.10 Da, 129 residues) subunits share a high degree of homology with each other (49% identity) and with the Ca^{2+}-dependent lectin-like proteins (CLPs) (typically 29-46%) isolated from other snake venoms. Rhodocetin is a heterodimer, and individually both α and β subunits fail to show significant inhibition of aggregation. Unlike the other members of the CLP family in which the subunits are held together by an inter-chain disulfide bond, the subunits of rhodocetin are held together only by non-covalent interactions. Rhodocetin heterodimer was believed to be formed by swapping the central loop.

Although the functional details of both proteins, Fve and rhodocetin, were known prior to this study, structural information was unavailable. The structure of these proteins would be important to permit the understanding of the structure-function relationships.

The structure of Fve, was determined by Single Anomalous Diffraction (SAD) using the anomalous signal of bromide ions present in the crystal for phasing. Fve represents a novel structure, wherein each subunit of the homodimer consists of an N-terminal α-helix followed by a fibronectin III (FNIII) type fold. The FNIII fold is the first instance of "pseudo-h-type" topology – a transition between the seven β-stranded s-type and the eight β-stranded h-type topologies. Fve does not contain tyrosine corner, which is found in >90% of (even sequence-unrelated) FNIII and Ig variable domains, but interestingly, this critical need is satisfied by the type I β-turn (Ala86-Asp-Thr-Lys89), with the backbone C=O of Ala86 forming an H-bond with the backbone N atom of Lys89.

The overall structure suggests that dimerization, critical for the activity of FIP proteins, occurs by 3D domain swapping of the N-terminal helices and is stabilized predominantly by strong hydrophobic interactions. Several simple sugars tested did not inhibit Fve-induced hemagglutination. However, the highly glycosylated protein thyroglobulin was found to inhibit hemagglutination, suggesting that Fve is a lectin with specificity for complex cell surface carbohydrates. We have predicted two pairs of patches that offer the potential to bind sugars. The first pair is found on the same side of the dimer. The second patch is found on opposite sides of the dimer. The structure of Fve is the first representative of this structurally novel family of lectins. It is also, to the best of our knowledge, the first structure of an FNIII domain-containing protein of fungal origin.

The structure of rhodocetin was determined by molecular replacement. This is the first structure of a CLP without a disulfide connecting the subunits and thus represents a novel molecule. The structure of rhodocetin comprises of a globular unit with an extended loop and forms a CLP fold. Superposition of IX/X-bp of subunit A with rhodocetin α subunit shows that the significant deviation in structure is due to the rotation of the loop by approximately 90º. The rotated loop spans 27 residues, from residue 73 to residue 100. The structure reveals the inter-subunit interface which has compensatory interactions for forming the dimer in the absence of the disulfide bridge. There is an ordering of the loop regions from Asn 80 to Ser 84 of the α subunit and from Gly 71 to Arg 75 of the β subunit to create an extra β-sheet not previously observed in other CLPs at the interface. It is most likely that this interface has disordered residues which create an ordered interface after interacting and subsequently present a surface which may then play a role in inhibiting platelet aggregation. Further, unlike other CLPs, rhodocetin does not require metal ions

for its functional activity. However, like other CLPs, rhodocetin also forms the heterodimer by domain swapping, in which the central looped region is swapped.

The structure of Fve and rhodocetin suggest that both dimers are formed by 3D domain swapping process. Both proteins are inactive in monomeric forms and hence the occurrence of dimerization is essential for the biological activity of the proteins by creating new interfaces.

Bibliography

1. Abrahamson, M. & Grubb, A. (1994). Increased body temperature accelerates aggregation of the Leu-68-->Gln mutant cystatin C, the amyloid-forming protein in hereditary cystatin C amyloid angiopathy. Proc Natl Acad Sci U S A, **91**, 1416-1420.

2. Alphey, M.S., Attrill, H., Crocker, P.R. & van Aalten, D.M. (2003). High resolution crystal structures of Siglec-7. Insights into ligand specificity in the Siglec family. J. Biol. Chem. **278**, 3372-3377.

3. Alvarez-Fernandez, M., Barrett, A.J., Gerhartz, B., Dando, P.M., Ni, J. & Abrahamson, M. (1999). Inhibition of mammalian legumain by some cystatins is due to a novel second reactive site. J Biol Chem. **274**, 19195-19203.

4. Atoda, H., Hyuga, M. & Morita, T. (1991). The primary structure of coagulation factor IX/factor X-binding protein isolated from the venom of *Trimeresurus flavoviridis*: homology with asialoglycoprotein receptors, proteoglycan core protein, tetranectin, and lymphocyte Fce receptor for immunoglobulin E. J. Biol. Chem. **266,** 14903-14911.

5. Atoda, H., Ishikawa, M., Yoshihara, E., Sekiya, F. & Morita, T. (1995). Blood coagulation factor IX-binding protein from the venom of *Trimeresurus flavoviridis* purification and characterization. J. Biochem. **118**, 965-973.

6. Baron, M., Norman, D.G. & Campell, I.D. (1991). Protein modules. Trends Biochem. Sci. **16**, 13-17.

7. Bennett, M.J., Schlunegger, M.P. & Eisenberg, D. (1995). 3D domain swapping: a mechanism for oligomer assembly. Protein Sci. **4**, 2455–2468.

8. Bennett, M.J., Choe, S. & Eisenberg, D.S. (1994). Domain swapping: Entangling alliances between proteins. Proc. Natl. Acad. Sci. U.S.A. **91**, 3127–3131.

9. Bergdoll, M., Remy, M.H., Cagnon, C., Masson, J.M. & Dumas, P. (1997). Proline-dependent oligomerization with arm exchange. Structure, **5**, 391-401.

10. 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**, 235-242.

11. Bertrand, J.A., Pignol, D., Bernard, J.P., Verdier, J.M., Dagorn, J.C. & Fontecilla-Camps, J.C. (1996). Crystal structure of human lithostathine, the pancreatic inhibitor of stone formation. EMBO J. **15**, 2678-2684.

12. Bode, W., Engh, R., Musil, D., Thiele, U., Huber, R., Karshikov, A., Brzin, J., Kos, J. & Turk, V. (1988). The 2.0 Å X-ray crystal structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. EMBO J. **7**, 2593-2599.

13. Bork, P., Holm, L. & Sander, C. (1994). The immunoglobulin fold. Structural classification, sequence patterns and common core. J. Mol. Biol. **242**, 309-320.

14. Bryson, J.W., Desjarlais, J.R., Handel, T.M. & DeGrado, W.F. (1998). From coiled coils to small globular proteins: design of a native-like three-helix bundle. Protein Sci. **7**, 1404-1414.

15. Canals, A., Pous, J., Guasch, A., Benito, A., Ribo, M., Vilanova, M. & Coll, M. (2001). The structure of an engineered domain-swapped ribonuclease dimer and its implications for the evolution of proteins toward oligomerization. Structure, **9**, 967–976.

16. Chan, Y.L. & Tsai, I.H. (1996). Functional and sequence characterization of coagulation factor IX/factor X-binding protein from the venom of *Echis carinatus leucogaster*. Biochemistry, **35**, 5264-5271.

17. Chothia, C., Gelfand, I. & Kister, A. (1998). Structural determinants in the sequences of immunoglobulin variable domain. J. Mol. Biol. **278**, 457-479.

18. Chou, P.Y. & Fasman, G.D. (1978). Empirical predictions of protein conformation. Ann. Rev. Biochem. **47**, 251-276.

19. Clarke, J., Cota, E., Fowler, S.B. & Hamill, S.J. (1999). Folding studies of immunoglobulin-like-sandwich proteins suggest that they share a common folding pathway. Structure, **7**, 1145-1153.

20. Coiller, R.J. (1975). Diphtheria toxin: Mode of action and structure. Bacteriol Rev. **39**, 54-85.

21. Cota, E., Steward, A., Fowler, S.B. & Clarke, J. (2001). The folding nucleus of a fibronectin type III domain is composed of core residues of the immunoglobulin-like fold. J. Mol. Biol. **305**, 1185-1194.

22. Crestfield, A.M., Stein, W.H. & Moore, S. (1962).On the aggregation of bovine pancreatic ribonuclease. Arch. Biochem. Biophys. **1**, 217–222.

23. Dauter, Z. & Dauter, M. (1999). Anomalous signal of solvent bromides used for phasing of lysozyme. J. Mol. Biol. **289**, 93-10.

24. Dickason, R.R. & Huston, D.P. (1996). Creation of a biologically active interleukin-5 monomer. Nature, **379**, 652-655.

25. Dieckmann, T., Mitschang, L., Hofmann, M., Kos, J, Turk, V., Auerswald, E.A., Jaenicke, R. & Oschkinat. H. (1993). The structures of native phosphorylated chicken cystatin and of a recombinant unphosphorylated variant in solution. J Mol Biol. **234**, 1048-1059.

26. Diederichs, K., Jacques, S., Boon, T. & Karplus, P.A. (1991). Low-resolution structure of recombinant human granulocyte-macrophage colony stimulating factor. J Mol Biol. **221**, 55-60.

27. Doolittle, R.F. & Bork, P. (1993). Evolutionarily mobile modules in proteins. Sci. Am. **269**, 50-56.

28. Eble, J.A., Beermann, B., Hinz, H.J. & Schmidt-Hederich, A. (2001). Alpha 2beta 1 integrin is not recognized by rhodocytin but is the specific, high affinity target of rhodocetin, an RGD-independent disintegrin and potent inhibitor of cell adhesion to collagen. J Biol Chem. **276**, 12274-12284.

29. Ekiel, I., Abrahamson, M., Fulton, D.B., Lindahl, P., Storer, A.C., Levadoux, W., Lafrance, M., Labelle, S., Pomerleau, Y., Groleau, D., LeSauteur, L. & Gehring, K. (1997). NMR structural studies of human cystatin C dimers and monomers. J Mol Biol. **271**, 266-277.

30. Engh, R.A., Dieckmann, T., Bode, W., Auerswald, E.A., Turk, V., Huber, R. & Oschkinat, H. (1993). Conformational variability of chicken cystatin. Comparison of structures determined by X-ray diffraction and NMR spectroscopy. J Mol Biol. **234**, 1060-1069.

31. Feinberg, H., Park-Snyder, S., Kolatkar, A.R., Heise, C.T., Taylor, M.E. & Weis, W.I. (2000). Structure of a C-type carbohydrate recognition domain from the macrophage mannose receptor. J Biol Chem. **275**, 21539-21548.

32. Fukushima, M., Ohashi, T., Fujiwara, Y., Sonoyama, K. & Nakano, M. (2001). Cholesterol-lowering effects of maitake (*Grifola frondosa*) fiber, shiitake (*Lentinus edodes*) fiber, and enokitake (*Flammulina velutipes*) fiber in rats. Exp Biol Med. **226**, 758-765.

33. Gerhartz, B. & Ekiel, I. (1998). Abrahamson M.Two stable unfolding intermediates of the disease-causing L68Q variant of human cystatin C. Biochemistry, **37**, 17309-17317.

34. Gotte, G., Bertoldi, M. & Libonati, M. (1999). Structural versatility of bovine ribonuclease A. Distinct conformers of trimeric and tetrameric aggregates of the enzyme. Eur J Biochem. **265**, 680-687.

35. Green, S.M., Gittis, A.G., Meeker, A.K. & Lattman, E.E. (1995). One-step evolution of a dimer from a monomeric protein. Nat Struct Biol. **2**, 746-751.

36. Grubb, A.O. (2000). Cystatin C--properties and use as diagnostic marker. Adv Clin Chem. **35**, 63-99.

37. Hamill, S.J., Cota, E., Chothia, C. & Clarke, J. (2000). Conservation of folding and stability within a protein family: the tyrosine corner as an evolutionary cul-de-sac. J. Mol. Biol. **295**, 641-649.

38. Hemmingsen, J.M., Gernert, K.M., Richardson, J.S. & Richardson, D.C. (1994). The tyrosine corner: a feature of most Greek key beta-barrel proteins. Protein Sci. **3**, 1927-1937.

39. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. J. Mol. Biol. **233**, 123-138.

40. Hsu, H.C., Hsu, C.I., Lin, R.H., Kao, C.L. & Lin, J.Y. (1997). Fip-vvo, a new fungal immunomodulatory protein isolated from *Volvariella volvacea*. Biochem. J. **323**, 557-565.

41. Ikekawa, T. (1995). Enokitake, *Flammulina velutipes*: antitumor activity of extracts and polysaccharides. Food Rev Intern. **11**, 203-206.

42. Janowski, R., Kozak, M., Jankowska, E., Grzonka, Z., Grubb, A., Abrahamson, M. & Jaskolski, M. (2001). Cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping. Nat Struct Biol. **8**, 316-320.

43. Jaskolski, M. (2001). 3D domain swapping, protein oligomarization and amyloid formation. Act biochimica polonica. **48,** 807-827.

44. Kino, K., Yamashita, A., Yamaoka, K., Watanabe, J., Tanaka, S., Ko, K., Shimizu, K. & Tsunoo, H. (1989). Isolation and characterization of a new immunomodulatory protein, ling zhi-8 (LZ-8), from *Ganoderma lucidium*. J. Biol. Chem. **264**, 472-478.

45. Knaus, K.J., Morillas, M., Swietnicki, W., Malone, M., Surewicz, W.K. & Yee, V.C. (2001). Crystal structure of the human prion protein reveals a mechanism for oligomerization. Nat Struct Biol. **8**, 770-774.

46. Ko, J.L., Hsu, C.I., Lin, R.H., Kao, C.L. & Lin, J.Y. (1995). A new fungal immunomodulatory protein, FIP-*fve* isolated from the edible mushroom, *Flammulina velutipes* and its complete amino acid sequence. Eur. J. Biochem. **228**, 244-249.

47. Kong, C.G. (2002). Structure and function studies of rhodocetin. PhD thesis. Department of Biochemistry, National University of Singapore, Singapore.

48. Kraulis, P.J., Domaille, P.J., Campbell-Burk, S.L., Van Aken, T. & Laue, E.D. (1997). Solution structure and dynamics of ras p21.GDP determined by heteronuclear three- and four-dimensional NMR spectroscopy. Biochemistry, **33**, 3515-31.

49. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thorton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. **26**, 283-290.

50. Libonati, M., Bertoldi, M. & Sorrentino, S. (1995). The activity on double-stranded RNA of aggregates of ribonuclease A higher than dimers increases as a function of the size of the aggregates. Biochem J. **318**, 287-290.

51. Lin, W.H., Hung, C.H., Hsu, C.I. & Lin, J.Y. (1997). Dimerization of the N-terminal amphipathic alpha-helix domain of the fungal immunomodulatory protein from *Ganoderma tsugae* (Fip-gts) defined by a yeast two-hybrid system and site-directed mutagenesis. J. Biol. Chem. **272**, 20044-20048.

52. Liu, Y. & Eisenberg, D. (2002). 3D domain swapping: as domains continue to swap. Protein Sci. **11**, 1285-1299.

53. Liu, Y., Gotte, G., Libonati, M. & Eisenberg, D. (2001). A domain-swapped RNase A dimer with implications for amyloid formation. Nat Struct Biol. **8**, 211-215.

54. Liu, Y., Gotte, G., Libonati, M. & Eisenberg, D. (2002). Structures of the two 3D domain-swapped RNase A trimers. Protein Sci. **11**, 71-80.

55. Liu, Y., Hart, P.J., Schlunegger, M.P. & Eisenberg, D. (1998). The crystal structure of a 3D domain-swapped dimer of RNase A at a 2.1-A resolution. Proc Natl Acad Sci U S A, **95**, 3437-3442.

56. Lovejoy, B., Choe, S., Cascio, D., McRorie, D.K., DeGrado, W.F. & Eisenberg. D. (1993). Crystal structure of a synthetic triple-stranded alpha-helical bundle. Science, **259**, 1288-1293.

57. Lu, G. (1996). A WWW service system for automatic comparison of protein structures. Protein Data Bank Quarterly Newsletter. **78**, 10-11.

58. Lumb, K.J. & Kim, P.S. (1995). A buried polar interaction imparts structural uniqueness in a designed heterodimeric coiled coil. Biochemistry, **34**, 8642-8648.

59. Martin, J.R., Craven, C.J., Jerala, R., Kroon-Zitko, L., Zerovnik, E., Turk, V. & Waltho, J.P. (1995). The three-dimensional solution structure of human stefin A. J Mol Biol. **246**, 331-343.

60. Matthews, B.W. (1968). Solvent content of protein crystals. J. Mol. Biol. **33**, 491-497.

61. Mazzarella, L., Capasso, S., Demasi, D., Matia, C.A. & Zagari, A. (1993). Bovine seminal ribonuclease: Structure at 1.9 A resolution. Acta. Crystallogr. D. **49,** 389–402.

62. McRee, D.E. (1999). XtalView/Xfit -- A Versatile Program for Manipulating Atomic Coordinates and Electron Density. J. Struct. Biol. **125**, 156-165.

63. Merrit, E.A. & Bacon, D.J. RASTER3D Methods Enzymol. **277**, 505-524.

64. Milburn, M.V., Hassell, A.M., Lambert, M.H., Jordan, S.R., Proudfoot, A.E., Graber, P. & Wells, T.N. (1993). A novel dimer configuration revealed by the crystal structure at 2.4 Å resolution of human interleukin-5. Nature, **363**, 172-178.

65. Mizuno, H., Fujimoto, Z., Koizumi, M., Kano, H., Atoda, H. & Morita, T. (1999). Crystal structure of coagulation factor IX-binding protein from habu snake venom at 2.6 Å: implication of central loop swapping based on deletion in the linker region. J Mol Biol. **289**, 103-112.

66. Mizuno, H., Fujimoto, Z., Koizumi, M., Kano, H., Atoda, H. & Morita, T. (1997). Structure of coagulation factors IX/X-binding protein, a heterodimer of C-type lectin domains. Nat Struct Biol. **4**, 438-441.

67. Morris, R.J., Perrakis, A. & Lamzin, V.S. (2002). ARP/wARP's model-building algorithms. I. The main chain. Acta Crystallogr. D. **58**, 968-975.

68. Murray, A.J., Lewis, S.J., Barclay, A.N. & Brady, R.L. (1995). One sequence, two folds: a metastable structure of CD2. Proc. Natl Acad. Sci. USA, **92**, 7337–7341.

69. Murshudov, G.N., Lebedev, A., Vagin, A.A., Wilson, K.S. & Dodson, E.J. (1999). Efficient anisotropic refinement of Macromolecular structures using FFT. Acta Crystallogr. *D.* **55**, 247-255.

70. Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C. (1995). SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. **247**, 536-540.

71. Newcomer, M.E. (2002). Protein folding and three-dimensional domain swapping: a strained relationship? Curr. Opin. Struct. Biol. **12**, 48-53.

72. Nicholls, A., Sharp, K. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. Proteins Struct. Func. Genet. **11**, 281-286.

73. Nielsen, B.B., Kastrup, J.S., Rasmussen, H., Holtet, T.L., Graversen, J.H., Etzerodt, M., Thøgersen, H.C. & Larsen, I.K. (1997). Crystal structure of tetranectin, a trimeric plasminogen-binding protein with an α-helical coiled coil. FEBS Letters, **412**, 388–396.

74. Ogihara, N.L., Ghirlanda, G., Bryson, J.W., Gingery, M., DeGrado, W.F. & Eisenberg, D. (2001). Design of three-dimensional domain-swapped dimers and fibrous oligomers. Proc Natl Acad Sci U S A, **98**, 1404-1409.

75. Olafsson, I. & Grubb, A. (2000). Hereditary cystatin C amyloid angiopathy. Amyloid, **7**, 70-79.

76. O'Neil, K.T. & DeGrado, W.F. (1990). A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. Science, **250**, 646-651.

77. O'Neill, J.W., Kim, D.E., Baker, D. & Zhang, K.Y. (2001b). Structures of the B1 domain of protein L from Peptostreptococcus magnus with a tyrosine to tryptophan substitution. Acta Crystallogr D. **57**, 480-487.

78. O'Neill, J.W., Kim, D.E., Johnsen, K., Baker, D. & Zhang, K.Y. (2001a). Single-site mutations induce 3D domain swapping in the B1 domain of protein L from Peptostreptococcus magnus. Structure, **9**, 1017-1027.

79. Orengo, C.A., Jones, D.T. & Thornton, J.M. (1994). Protein superfamilies and domain superfolds. Nature, **372**, 631-634.

80. Otagiri, K., Ohkuma, T., Ikekawa, T. & Tanaka, S. (1983). Intensification of antitumor-immunity by protein-bound polysaccharide, EA6, derived from

Flammulina velutipes (Curt. ex Fr.) Sing. combined with murine leukemia L1210 vaccine in animal experiments. J. Pharmacobiodyn. **6**, 96-104.

81. Otwinowski, Z.M. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. **276**, 307-326.

82. Pei, X.Y., Holliger, P., Murzin, A.G. & Williams, R.L. (1997). The 2.0-Å resolution crystal structure of a trimeric antibody fragment with noncognate VH-VL domain pairs shows a rearrangement of VH CDR3. Proc Natl Acad Sci U S A, **94**, 9637-9642.

83. Rousseau, F., Schymkowitz, J.W.H., Wilkinson, H.R. & Itzhaki, L.S. (2001). Three-dimensional domain swapping in p13suc1 occurs in the unfolded and controlled by conserved proline residues. Proc.Natl Acad.Sci.USA, **98**, 5596-5601.

84. Salem, G.M., Hutchinson, E.G., Orengo, C.A. & Thornton, J.M. (1999). Correlation of observed fold frequency with the occurrence of local structural motifs. J. Mol. Biol. **287**, 969-981.

85. Schlunegger, M.P., Bennett, M.J. & Eisenberg, D. (1997). Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. Adv Protein Chem. **50**, 61-122.

86. Sekiya, F., Atoda, H. & Morita, T. (1993). Isolation and characterization of an anticoagulant protein homologous to botrocetin from the venom of Bothrops jararaca. Biochemistry, **32**, 6892-6897.

87. Sen, U., Vasudevan, S., Subbarao, G., McClintock, R.A., Celikel, R., Ruggeri, Z. M. & Varughese, K.I. (2001). Crystal structure of the von Willebrand factor modulator botrocetin. Biochemistry, **40**, 345-352.

88. Seow, S.V., Kuo, I.C., Paaventhan, P., Kolatkar, P.R. & Chua, K.Y. (2003). Crystallization and preliminary X-ray crystallographic studies on the fungal immunomodulatory protein Fve from the golden needle mushroom (*Flammulina velutipes*). Acta Cryst D. **59,** 1487-1496.

89. She, Q.B., Ng, T.B. & Liu, W.K. (1998). A novel lectin with potent immunomodulatory activity isolated from fruiting bodies and cultured mycelia of the edible mushroom *Volvariella volvacea*. Biochem. Biophys. Res. Commun. **247**, 106-111.

90. Steward, A., Adhya, S. & Clarke, J. (2002). Sequence conservation in Ig-like domains: the role of highly conserved proline residues in the fibronectin type III superfamily. J. Mol. Biol. **318**, 935-940.

91. Stubbs, M.T., Laber, B., Bod, W., Huber, R., Jerala, R., Lenarcic, B. & Turk, V. (1990). The refined 2.4 A X-ray crystal structure of recombinant human stefin B in

complex with the cysteine proteinase papain: a novel type of proteinase inhibitor interaction. EMBO J. **9**, 1939-1947.

92. Tanaka, S., Ko, K., Kino, K., Tsuchiya, K., Yamashita, A., Murasugi, A., Sakuma, S. & Tsunoo, H. (1989). Complete amino acid sequence of an immunomodulatory protein, ling zhi-8 (LZ-8). An immunomodulator from a fungus, *Ganoderma lucidium*, having similarity to immunoglobulin variable regions. J. Biol. Chem. **264**, 16372-16377.

93. Taroni, C., Jones, S. & Thornton, J.M. (2000). Analysis and prediction of carbohydrate binding sites. Protein Eng. **13**, 89-98.

94. Terwilliger, T.C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D. **55**, 501-505.

95. Terwilliger, T.C. (2001). Map-likelihood phasing. Acta Crystallogr. D. **57**, 1763-1775.

96. Turk, V. & Bode, W. (1991). The cystatins: protein inhibitors of cysteine proteinases. FEBS Lett. **285**, 213-219.

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

98. van den Berg, L. & Rose, D. (1959). Effect of freezing on the pH and composition of sodium and potassium phosphate solutions; the reciprocal system KH2PO4-Na2-HPO4-H2O. Arch. Biochem. Biophys. **81**, 319-329.

99. Vangelista, L., Cesco-Gaspere, M., Lamba, D. & Burrone, O. (2002). Efficient folding of the FcepsilonRI alpha-chain membrane-proximal domain D2 depends on the presence of the N-terminal domain D1. J. Mol. Biol. **322**, 815-825.

100. Walsh, S.T., Cheng, H., Bryson, J.W., Roder, H. & DeGrado, W.F. (1999). Solution structure and dynamics of a de novo designed three-helix bundle protein. Proc Natl Acad Sci U S A, **96**, 5486-5491.

101. Wang, H.X. & Ng, T.B. (2000). Flammulin: a novel ribosome-inactivating protein from fruiting bodies of the winter mushroom *Flammulina velutipes*. *Biochem.* Cell. Biol. **78**, 699-702.

102. Wang, R., Kini, R.M. & Chung, M.C. (1999). Rhodocetin, a novel platelet aggregation inhibitor from the venom of Calloselasma rhodostoma (Malayan pit viper): synergistic and noncovalent interaction between its subunits. Biochemistry, **38,** 7584-7593.

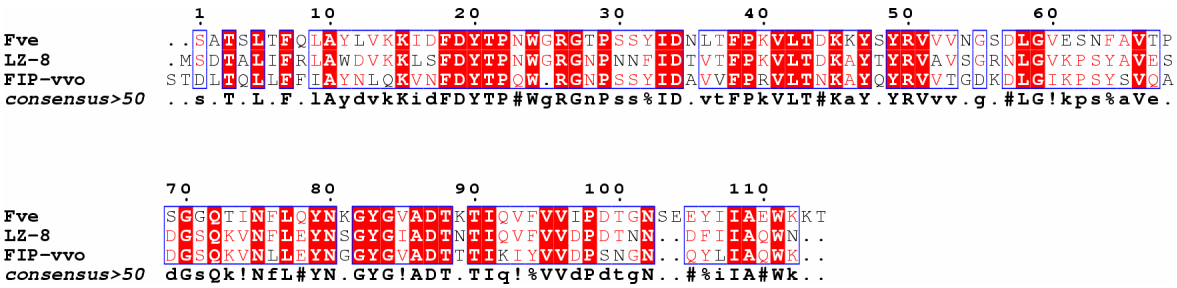
103. Wasser, S.P. & Weis, A.L. (1999). Therapeutic effects of substances occurring in higher Basidiomycetes mushrooms: a modern perspective. Crit Rev Immunol. **19**, 65-96.

104. Weis, W.I., Kahn, R., Fourme, R., Drickamer, K. & Hendrickson, W.A. (1991). Structure of the calcium-dependent lectin domain from a rat mannose-binding protein determined by MAD phasing. Science, **254**, 1608-1615.

105. Williams, A.F. & Barclay, A.N. (1988). The immunoglobulin superfamily – domains for cell surface recognition. Ann. Rev. Immunol. **6**, 381-405.

106. Zingali, R.B., Jandrot-Perrus, M., Guillin, M.C. & Bon, C. (1993). Bothrojaracin, a new thrombin inhibitor isolated from *Bothrops jararaca* venom: characterization and mechanism of thrombin inhibition. Biochemistry, **32**, 10794-10802.

APPENDIX


1 10 20 30 40 50 60
Fve ..SATSLTFQLAYLVKKIDFDYTPNWGRGTPSSYIDNLTFPKVLTDKKYSYRVVVNGSDLGVESNFAVTP
LZ-8 .MSDTALIFRLAWDVKKLSFDYTPNWGRGNPNNFIDTVTFPKVLTDKAYTYRVAVSGRNLGVKPSYAVES
FIP-vvo STDLTQLLFFIAYNLQKVNFDYTPQW.RGNPSSYIDAVVFPRVLTNKAYQYRVVTGDKDLGIKPSYSVQA
consensus>50 ..s.T.L.F.lAydvkKidFDYTP#WgRGnPss%ID.vtFPkVLT#KaY.YRVvv.g.#LG!kps%aVe.
70 80 90 100 110
Fve SGGQTINFLQYNKGYGVADTKTIQVFVVIPDTGNSEEYIIAEWKKT
LZ-8 DGSQKVNFLEYNSGYGIADTNTIQVFVVDPDTNN..DFIIAQWN..
FIP-vvo DGSQKVNLLEYNGGYGVADTTTIKIYVVDPSNGN..QYLIAQWK..
consensus>50 dGsQk!NfL#YN.GYG!ADT.TIq!%VVdPdtgN..#%iIA#Wk..