

# Characterization of the Mechanosensitive Channel of Large Conductance

Thesis by

Pavel Strop

(advisors: D. C. Rees and S. L. Mayo)

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2002

(Submitted April 18, 2002)

©2002

Pavel Strop

All Rights Reserved

## Acknowledgements

I would like to thank my advisors, Doug Rees and Steve Mayo, for their scientific guidance, research ideas and patience. I have really enjoyed the pleasant environment they have created in their labs. They have both given me a large freedom in projects and allowed me to try my own ideas. Apart from allowing me to work on great projects, they also made it possible to enjoy other aspects of life. They were available when I needed advice, encouragement, or guidance, but they have also let me work out problems on my own. Thank you!

I would like to thank Pamela Bjorkman and Carl Parker for their guidance, scientific and career advice as my graduate committee.

I would like to thank Randal Bass for teaching me the basics of membrane protein biochemistry, as well as for vast amount of useful ideas and practical advice. In the same manner, I would like to thank Kaspar Locher for his guidance and influential ideas. I have benefited from numerous discussions with both of them, and without their help the mechanosensitive channel project would not get very far. I would also like to thank to Xavier Ambroggio for his help with molecular biology, Christian Becker for his work on pPEG, and Meg Barclay, Phoebe Ray, and Allen Lee for making the Rees lab run smoothly. Thanks to our collaborator Kerry Smith for the carbonic anhydrase, and Brian Crane and Alex Bilwes for their help with carbonic anhydrase data collection. Also thanks to other members of the Rees group including Kaspar Locher, Randal Bass, Xavier Ambroggio, Andy Yeh, Rob Spencer, Akif Tezcan, Oliver Einsle, Susana Andrade, Mika Yoshida, Liz Borths, Benedikt Schmid, Ivana Hughes, Meg Barclay,

Allen Lee, and Hsiu-Ju Chiu for help along the way, and for making the lab a pleasant work environment.

I would also like to thank Scott Ross for all of his help with nuclear magnetic resonance spectroscopy. His help ranged from collecting the NMR spectra to teaching me the basics of protein NMR. Without his help I would not be able to obtain the structure of rubredoxin. Thanks to Sandy Malakauskas for introducing me to molecular biology, and among other things, teaching me the art of corner cutting. Dan Bolon has helped me not only with computational design, but also with many other aspects of my research. I would like to thank Cynthia Carlson, Rhonda Digiusto, and Monica Breckow for making the Mayo lab run smoothly. Thanks to other people in the Mayo lab that I have interacted with: Arthur Street, Niles Pierce, Possu Huang, Premal Shah, Kirsten Lassila, Jessica Mao, John Love, Geoffrey Hom, Julie Archer, Ben Gordon, Bassil Dahiat, and Cathy Sarisky.

Thanks to Eric Connor for introducing me to whitewater kayaking. Driving with four kayaks on the roof of his overheated Honda, in 110 °F temperatures with heater on full blast to cool the engine was miserable, but paddling the forks of the Kern with Eric, Roger, Ralph, Geno, and Gae was absolutely worth it. “Relaxing” in the rapids has helped me forget about lab, especially when things were not going so well. Thanks to Dan Bolon for cooking lots of pasta, drinking beer and listening to Bob Marley with me. Over the years Dan has convinced me that there are other important things besides work, climbing, and kayaking. Many thanks to Kirk Hansen for showing me how to get through graduate school, for his scientific help, but most importantly for good times during our climbing expeditions. Whether we were climbing on cliffs in France, Canada,

Joshua Tree, Zion, or Yosemite, we always had a great time. Their friendship has made my years at Caltech very pleasant and my best memories come from our outings. Thanks to my roommate J.P. Morgan at the Crary house for his friendship, good parties, and for keeping up Eric's teatime tradition. Thanks to Randal Bass for going surfing with me even in the winter when it was cold. Special thanks to Lisa Welp for her support in my final days at Caltech. I would also like to thank other friends that I made in Southern California: Isaac Carrico, Jennie Stevens, Amy Meyer, Drew Waltman, Zsuzsi Hamburger, Lance Martin, Anne Raney, Akif Tezcan, Derek Debe, Andrei Marinescu, Todd Younkin, David Beckman, Carlo Quinonez, Girish Aakalu, Dan Meulemans, as well as others that I have forgotten to mention.

Last but definitely not least, I would like to thank my family. Thanks to my mom and dad for providing me with everything I needed since I was born, for encouraging me and giving me direction and good advice over the years. They have given me more than I will ever be able to repay. I also need to thank my sister Jitka, her husband Jakub, and my niece Veronika for their invaluable support, friendship, and entertainment.

This work was supported by NSF predoctoral fellowship (P.S.), NIH training grant (P.S.), GM45162 (DCR), and Howard Hughes Medical Institute (D.C.R. and S.L.M.) This work was also based upon research conducted at the Stanford Synchrotron Radiation Laboratory (SSRL), which is funded by the Department of Energy (BES, BER) and the National Institutes of Health (NCRR, NIGMS).

## Abstract

Osmoregulation is an essential process in bacteria and higher organisms regulated by the mechanosensitive ion channels. The mechanosensitive channel of large conductance (MscL) is an integral membrane protein that responds to pressure in an effort to prevent cell lysis during osmotic shock. Conversion of MscL from a membrane-bound form to a water soluble form was attempted by three methods: computational design, random mutagenesis and chemical modification. The water soluble form of MscL was achieved with cysteine modification method. The stability, pH dependence, and C-terminal helix of MscL were also investigated.

The structure of the 'cab'  $\beta$ -class carbonic anhydrase (Cab) has been determined to 2.1 Å resolution. Cab exists as a dimer with a fold similar to 'plant'  $\beta$ -class carbonic anhydrases. The active site zinc is coordinated by Cys32, His87, and Cys90, with the tetrahedral coordination completed by a water molecule. The difference between 'plant' and 'cab'  $\beta$ -class carbonic anhydrases is in the organization of the hydrophobic pocket. The structure reveals a Hepes molecule near the active site, suggesting a proton transfer pathway to the solvent.

The structure of the nitrogenase iron protein in the all-ferrous  $[4\text{Fe-4S}]^0$  form has been determined to 2.2 Å resolution. The structure demonstrates that major conformational changes are not necessary to accommodate cluster reduction to the  $[4\text{Fe-4S}]^0$  state. A survey of  $[4\text{Fe-4S}]$  clusters coordinated by four cysteine ligands reveals that the  $[4\text{Fe-4S}]$  cluster of the iron protein has the largest accessible surface area, suggesting that solvent exposure may be relevant to the capability of existing in three oxidation states.

The role of surface salt bridges in protein stabilization has been investigated. The NMR structure of a rubredoxin variant (PFRD-XC4) and the thermodynamic analysis of two surface salt bridges is presented here. The analysis shows that the surface side chain to side chain salt bridge between does not stabilize PFRD-XC4. The main chain to side chain salt bridge, however, stabilizes PFRD-XC4 by  $1.5 \text{ kcal mol}^{-1}$ . The entropic cost of making a surface salt bridge involving the protein's backbone is reduced, since the backbone has already been immobilized upon protein folding.

## Table of Contents

	Page
Acknowledgements .....	iii
Abstract .....	vi
Table of Contents .....	viii
List of Figures .....	xii
List of Tables .....	xiv
Abbreviations .....	xvi

### **Section I    Mechanosensitive Ion channels**

<b>Chapter 1    Mechanosensitive Ion Channels .....</b>	<b>1</b>
Mechanosensation.....	2
Eukaryotic mechanosensitive channels.....	2
<i>Sodium channels</i> .....	2
<i>Calcium permeable channels</i> .....	4
<i>Potassium channels</i> .....	5
Bacterial mechanosensitive channels.....	5
MscL .....	6
<i>N-terminus</i> .....	7
<i>TM1 helix</i> .....	8
<i>Extra-cellular loop</i> .....	10
<i>TM2 helix</i> .....	11
<i>Cytoplasmic domain</i> .....	12
MscS.....	13
<i>Channel discovery</i> .....	13
<i>Sequences of MscS</i> .....	14
<i>Archaeal MscS-like proteins</i> .....	16
MscM .....	16
References .....	18
<b>Chapter 2    Conversion of Mechanosensitive Channel of Large Conductance to Water Soluble Form.....</b>	<b>20</b>
Introduction.....	21
<i>Higher resolution structure of Tb_MscL</i> .....	25
<i>Open state of MscL</i> .....	25
<i>MscL complexed with channel blockers</i> .....	27
<i>Protein design</i> .....	29



	<i>Cysteine modification</i> .....	30
	<i>Random mutagenesis</i> .....	31
Results	.....	33
	<i>Survey of soluble proteins</i> .....	33
	<i>Computational design</i> .....	34
	<i>Expression and fractionation</i> .....	34
	<i>Refolding</i> .....	36
	<i>MBP fusion</i> .....	38
	<i>Random mutagenesis</i> .....	39
	<i>Expression of fusion proteins</i> .....	39
	<i>Fluorescence of fusion proteins</i> .....	40
	<i>Cysteine modification</i> .....	40
	<i>MscL aggregation</i> .....	43
Discussion	.....	45
Materials and methods	.....	48
	<i>Computational modeling</i> .....	48
	<i>Protein expression</i> .....	50
	<i>Refolding</i> .....	50
	<i>Cysteine modification</i> .....	51
	<i>Random mutagenesis</i> .....	53
	<i>Plasmid</i> .....	53
	<i>Fluorescence</i> .....	53
	<i>Random oligonucleotide design</i> .....	53
Acknowledgements	.....	54
References	.....	55
<b>Chapter 3</b>	<b>Characterization and stability of MscL</b> .....	<b>57</b>
	<i>Cytoplasmic helix</i> .....	58
	<i>MscL stability</i> .....	62
	<i>Higher resolution structure of Tb_MscL</i> .....	66
	<i>Termination mutants</i> .....	66
	<i>COMP fusion</i> .....	67
Materials and methods	.....	70
	<i>Tb_MscL purification</i> .....	70
	<i>pH profiling</i> .....	72
	<i>Circular dichroism</i> .....	72
	<i>C-terminal truncations</i> .....	72
	<i>COMPcc fusion</i> .....	73
Acknowledgements	.....	73
References	.....	74

## **Section II Metalloproteins**

### **Chapter 4 Crystal Structure of the 'cab' Type Beta Class Carbonic Anhydrase from the Archaeon *Methanobacterium thermoautotrophicum* .....75**

Abstract .....	77
Introduction.....	78
Experimental Procedures .....	81
<i>Crystallization</i> .....	81
<i>Data collection and processing</i> .....	81
<i>Phase determination</i> .....	81
<i>Refinement</i> .....	83
Results and Discussion .....	84
<i>Structural organization of Cab</i> .....	84
<i>Cab oligomerization</i> .....	86
<i>Comparison of 'cab' and P.s. and P.p. structures</i> .....	88
<i>Active site</i> .....	88
<i>Comparison of P.s. b-CA and Cab active sites</i> .....	90
<i>Comparison of P.p. b-CA and Cab active sites</i> .....	92
<i>Mechanism</i> .....	92
Acknowledgements .....	95
References .....	96

### **Chapter 5 Crystal Structure of the All-ferrous [4Fe-4S]<sup>0</sup> Form of the Nitrogenase Iron Protein from *Azotobacter vinelandii* .....98**

Abstract .....	100
Introduction.....	101
Methods.....	102
<i>Protein purification and reduction to the [4Fe-4S]<sup>0</sup> state</i> .....	102
<i>Crystallization</i> .....	103
<i>Data collection and structure determination</i> .....	104
Results and Discussion .....	107
<i>Comparison to the other Fe protein structures</i> .....	107
<i>Cluster geometry</i> .....	108
<i>Cluster volumes</i> .....	111
<i>Accessible surface</i> .....	113
<i>N-H-S hydrogen bonds</i> .....	113
Conclusions .....	114
Acknowledgements .....	115
References .....	116

### **Section III Protein Design**

<b>Chapter 6</b>	<b>Rubredoxin Variant Folds without Iron.....</b>	<b>118</b>
	Abstract .....	120
	Introduction.....	121
	Experimental Procedures .....	122
	<i>Computational modeling</i> .....	122
	<i>Mutagenesis and protein purification</i> .....	123
	<i>CD analysis</i> .....	123
	<i>NMR studies</i> .....	124
	Results and Discussion .....	125
	Conclusion .....	127
	Acknowledgements .....	128
	References .....	129
<b>Chapter 7</b>	<b>NMR Structure of Rubredoxin Variant and Contribution of Surface Salt Bridges to Protein Stability .....</b>	<b>132</b>
	Abstract .....	134
	Introduction.....	135
	Materials and methods .....	136
	<i>Mutagenesis and protein purification</i> .....	136
	<i>NMR structure determination</i> .....	136
	<i>CD analysis</i> .....	138
	Results and Discussion .....	139
	<i>Structure determination of PFRD-XC4</i> .....	139
	<i>Stability studies of PFRD-XC4 mutants</i> .....	141
	Conclusions .....	146
	Acknowledgements .....	147
	References .....	148
<b>Chapter 8</b>	<b>Structure of a Protein G helix Variant Suggests the Importance of Helix Propensity and Helix Dipole Interactions in Protein Design. ..</b>	<b>150</b>
	Abstract .....	152
	Introduction.....	153
	Results and discussion .....	154
	Materials and methods .....	156
	<i>Computational design</i> .....	156
	<i>Expression and CD analysis</i> .....	157
	<i>Crystallization, data collection and refinement</i> .....	158
	Acknowledgements .....	159
	References .....	160

## List of Figures

### Chapter 1:

- 1.1 Topology of eukaryotic and bacterial mechanosensitive channels
- 1.2 Structure of mechanosensitive channel of large conductance
- 1.3 Sequence alignment of MscL genes
- 1.4 Sequence alignment of MscS genes

### Chapter 2:

- 2.1 Crystal packing of MscL and KcsA
- 2.2 Model of the open state of MscL
- 2.3 Approaches to creating water soluble form of MscL channel
- 2.4 Correlation of protein solubility and GFP fusion fluorescence
- 2.5 Histogram of protein surface hydrophobicity for soluble and membrane proteins
- 2.6 Ribbon diagram of MscL with residues selected for design
- 2.7 Flowchart of experimental screening of designed proteins
- 2.8 Circular dichroism spectra of MscL and refolded variants
- 2.9 Ribbon diagram of MscL with residues selected for cysteine modification
- 2.10 MscL\_Cys4 gel filtration traces

### Chapter 3:

- 3.1 Ribbon diagram of the structure of MscL and of the cytoplasmic helix
- 3.2 Possible models for cytoplasmic bundle at physiological pH
- 3.3 Wavelength scans of MscL at pH 2-12
- 3.4 Wavelength scans of C-terminal peptides
- 3.5 Wavelength scans of five MscL homologs
- 3.6 Temperature melts of five MscL homologs
- 3.7 Temperature melts of Tb\_MscL and Eco\_MscL at different pH
- 3.8 COMPcc structure
- 3.9 Model of MscL\_COMPcc fusion

### Chapter 4:

- 4.1 Alignment of  $\beta$ -CA sequences
- 4.2 Ribbon diagram of Cab
- 4.3 Stereo view of the crystal packing of Cab
- 4.4 Stereo diagram showing the superimposition of a backbone trace of Cab, *P. s.*  $\beta$ -CA, and *P. p.*  $\beta$ -CA
- 4.5 Stereo view of the zinc environment in Cab
- 4.6 Stereo diagram showing the superposition of the active site of Cab

**Chapter 5:**

- 5.1 Ribbons diagram of the all-ferrous *A. vinelandii* Fe protein dimer
- 5.2 Variation in the volumes of the Fe and sulfur tetrahedron in [4Fe-4S] clusters
- 5.3 Stereodiagram of the [4Fe-4S] cluster environment in the All-ferrous Av2 and 3NIP structures.
- 5.4 Histogram of the number of NH-S hydrogen bonds to the cluster observed in structures of [4Fe-4S] cluster containing proteins

**Chapter 6:**

- 6.1 Ribbon diagram of wild-type mesophilic *Desulfovibrio vulgaris* rubredoxin
- 6.2 Circular dichroism measurements of wild-type *Pyrococcus furiosus* rubredoxin (PFRD) and the PFRD-XC4 variant
- 6.3 NMR assignments summary and NOE connectivities of PFRD-XC4

**Chapter 7:**

- 7.1 Stereoview of the PFRD-XC4 structure ensemble and comparison to the PFRD X-ray structure.
- 7.2 Ribbon diagram of wild type *Pyrococcus furiosus* rubredoxin
- 7.3 Chemical denaturation curves for PFRD-XC4

**Chapter 8:**

- 8.1 Stereoview of the superposition of the G $\beta$ 1 and G $\beta$ 1m2 X-ray structures

## List of Tables

### Chapter 2:

- 2.1 Amino acid sequences selected for the designs
- 2.2 Experimental results for designed proteins
- 2.3 Cysteine modification results

### Chapter 4:

- 4.1 Data Collection Statistics
- 4.2 Refinement Statistics

### Chapter 5:

- 5.1 Data collection and phasing statistics
- 5.2 Fe-Fe and S-S distances in the [4Fe-4S] cluster of different Fe-protein structures
- 5.3 Average values for geometrical parameters for protein bound [4Fe-4S] clusters

### Chapter 6:

- 6.1  $^1\text{H}$  NMR Chemical Shifts for PFRD-XC4

### Chapter 7:

- 7.1 NMR structure statistics
- 7.2 Contributions of salt bridges at 1 °C
- 7.3 Thermal and chemical denaturation data

### Chapter 8:

- 8.1 Refinement and Model Statistics

## Abbreviations

1CP2,	1.90 Å resolution structure of <i>C. pasteurianum</i> Fe protein
1NIP,	2.9 Å resolution structure of <i>A. vinelandii</i> Fe protein;
1pPEG_MscL_Cys4,	MscL_Cys4 mutant modified with single pPEG
2pPEG_MscL_Cys4,	MscL_Cys4 mutant modified with double pPEG
3NIP,	2.2 Å resolution structure of <i>A. vinelandii</i> Fe protein
ADH,	alcohol dehydrogenase
ASIC,	Acid-sensing ion channels
Av1,	<i>A. vinelandii</i> MoFe protein
Av2,	<i>A. vinelandii</i> Fe protein
BME,	beta mercaptoethanol
Bs_MscL,	<i>Bacillus subtilis</i> MscL
BSA,	bovine serum albumin
BtuCD,	<i>Escherichia coli</i> vitamin B12 ABC transporter
CA,	carbonic anhydrase <i>Methanobacterium thermoautotrophicum</i> b-class carbonic anhydrase
Cab,	
CAII,	human alpha carbonic anhydrase II
CD,	circular dichroism
CDR,	Cysteine rich domain
COMP,	cartilage oligomeric matrix protein
COMPcc,	N-terminal heptad repeat of cartilage oligomeric matrix protein
Cp2,	<i>C. pasteurianum</i> Fe protein
Cyt c,	Cytochrome c
D1,	Tb_MscL (E34,K38,Q69,R72,E73,E77,K80,Q81,K84,E87, K88,K91,R94)
D1_GFP,	D1 fusion with green fluorescent protein
D1_T100,	D1 design protein truncated at position 100
D1_T119,	D1 design protein truncated at position 119
D2,	Tb_MscL (Q69,R73,E77,N80,K84,E87,K88,K91)
D2_GFP,	D2 fusion with green fluorescent protein
D2_T100,	D2 design protein truncated at position 100
D2_T119,	D2 design protein truncated at position 119
D3,	Tb_MscL (Q34,Q38,Q69,N72,E73,E77,N80,Q81,K84,E87, S88,R91,N94)

D3_GFP,	D3 fusion with green fluorescent protein
D3_T100,	D3 design protein truncated at position 100
D3_T119,	D3 design protein truncated at position 119
DDM,	n-Dodecyl-beta-D-maltopyranoside
$\Delta G_u$ ,	free energy of unfolding
DM,	n-Decyl-beta-D-maltopyranoside
DTNB,	dithiobisnitrobenzoic acid
EANAL,	energy analysis
Eco_MscL,	<i>Escherichia coli</i> MscL
EDTA,	Ethylenediaminetetra-acetic acid
EGF,	epidermal growth factor
Er_MscL,	<i>Erwinia carotovora</i> MscL
EXAFS,	extended X-ray absorption fine structure spectroscopy
Fab,	Immunoglobulin fragment (light chain and the VH and CH1 of the heavy chain)
FOS12,	N-Dodecylphosphocholine
GdmCl,	guanidinium chloride
GFP,	green fluorescent protein
Hepes,	N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulfonic acid]
Hinf_MscL,	<i>Haemophilus influenzae</i> MscL
HP1,	C-terminal helix peptide (sequence: EVEQPGDTQVVLLTEIR)
HTS,	high throughput screening
IPTG,	Isopropyl-beta-D-thiogalactoside
K <sub>2P</sub> ,	Two-pore domain potassium channels
KcsA,	K <sup>+</sup> channel from <i>Streptomyces lividans</i>
M.W.,	molecular weight
MAD,	multiwavelength anomalous dispersion
MBP,	Maltose binding protein
MBP_D1,	Maltose binding protein fused to D1 design
MBP_D1,	Maltose binding protein fusion to D1
MBP_D2,	Maltose binding protein fused to D2 design
MBP_D2,	Maltose binding protein fusion to D2
MBP_D3,	Maltose binding protein fused to D3 design
MBP_D3,	Maltose binding protein fusion to D3



MBP_Tb_MscL,	Maltose binding protein fusion to Tb_MscL
MPD,	2-Methyl-2,4-pentanediol
MscL ,	Mechanosensitive channel of large conductance
MscL_COMPcc,	Tb_MscL fusion with COMPcc
MscL_Cys4,	MscL variant with L73C, F80C, F84C, and F88C mutations
MscM,	Mechanosensitive channel of mini conductance
MscMJ,	Archael mechanosensitive channel
MscMJLR,	Archael mechanosensitive channel
MscS ,	Mechanosensitive channel of small conductance
m-value,	slope of $\Delta G_u$ versus denaturant concentration
NCS,	non-crystallographic symmetry
NMR,	nuclear magnetic resonance
NOE,	nuclear Overhauser enhancement
OD,	optical density
ORBIT,	optimization of rotamers by iterative techniques
<i>P.p.</i> b-CA,	<i>Porphyridium purpureum</i> b class carbonic anhydrase
<i>P.s.</i> b-CA,	<i>Pisum sativum</i> b class carbonic anhydrase
PCR,	polymerase chain reaction
PDB,	Protein databank
PFRD,	<i>Pyrococcus furiosus</i> rubredoxin
PFRD-XC4,	rubredoxin variant with C5L, C8T, C38A, and C41T mutations
PMSF,	Phenylmethylsulfonyl fluoride
pPEG,	precision polyethylene glycol
RESCLASS,	residue classification
rmsd,	root mean square deviation
rpm,	rotations per minute
SAM,	sterile alpha motif
SDS,	sodium dodecyl sulfate
T100,	Tb_MscL truncation mutant at position 100
T119,	Tb_MscL truncation mutant at position 119
TB,	terrific broth
Tb_MscL,	<i>Mycobacterium tuberculosis</i> MscL
Tb_MscL_GFP,	Tb_MscL fusion with green fluorescent protein
T <sub>m</sub> ,	midpoint of thermal unfolding transition

TM1,	First transmembrane helix in Tb_MscL
TM2,	Second transmembrane helix in Tb_MscL
TRP,	Transient receptor potential
VR-OAC,	Osmotically activated channel related to the vanilloid receptor
WT,	wild type

## **Section I**

### **Chapter 1:**

## **Mechanosensitive Ion Channels**

Pavel Strop, Randal B. Bass, Douglas C. Rees

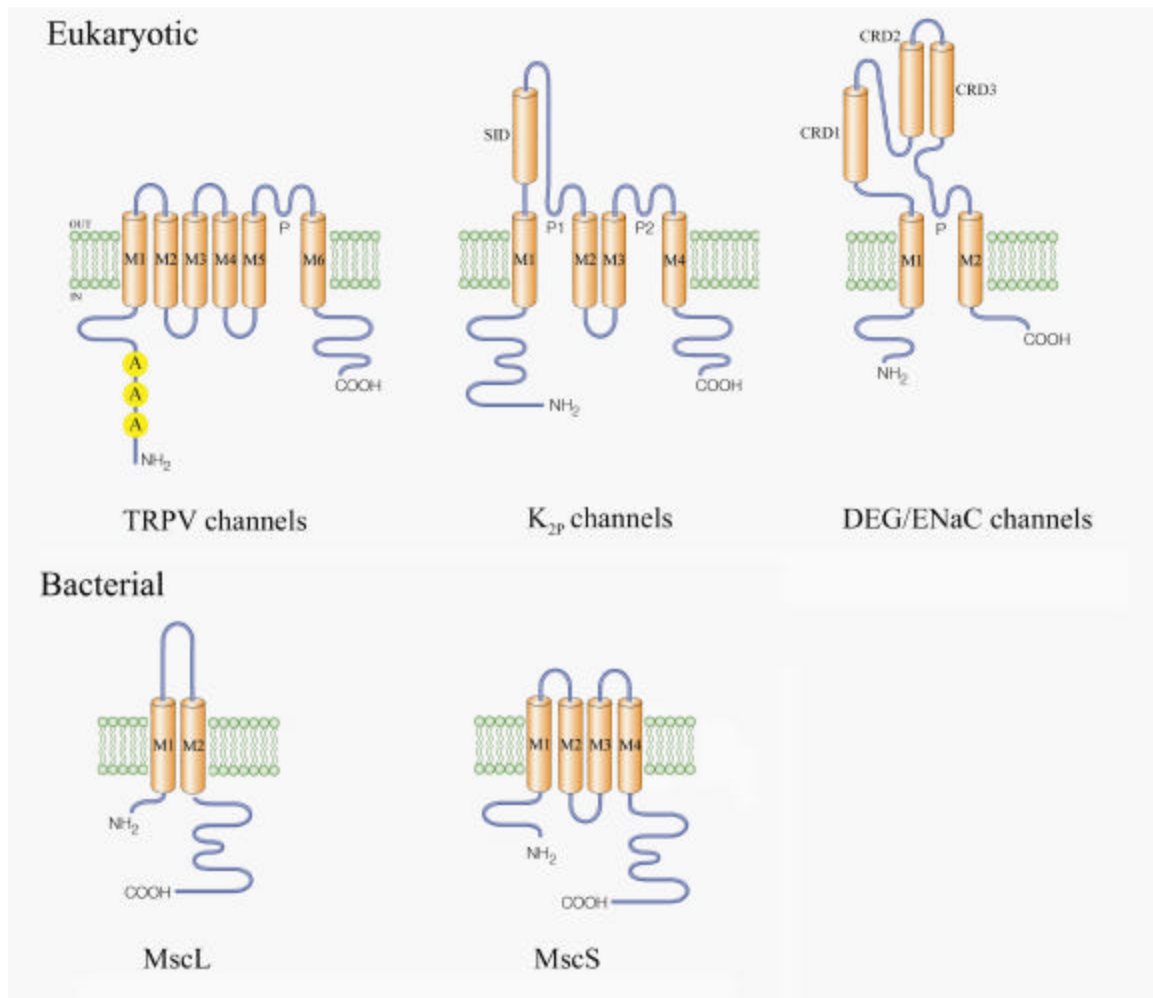
## Mechanosensation

“Imagine a bacterial cell being suddenly transported from the colon of an animal to a pool of water. This situation rapidly becomes life threatening” (1). Osmoregulation is essential in bacteria and higher organisms and is one of many important processes regulated by mechanosensitive ion channels. Mechanosensitive ion channels are integral membrane proteins that not only regulate osmotic pressure, but are also involved in cellular processes such as touch, hearing, cardiovascular tone, detection of gravity, pressure sensation, tissue deformation, swelling, injury, and pain perception (2). Mechanosensitive channels have been discovered in all fundamental branches of the phylogenetic tree, *Eubacteria*, *Eukarya*, and *Archaea* (3).

### Eukaryotic mechanosensitive channels

Although several classes of eukaryotic mechanosensitive channels have recently been identified, the characterization of these channels is just starting to emerge. Currently known eukaryotic channels that display mechanosensitivity can be divided into three main categories: sodium channels, calcium permeable channels, and potassium channels (4).

*Sodium channels:* The DEG/ENaC superfamily of sodium channels participates in diverse biological processes (5). DEG/ENaC channels are made up of two membrane-spanning helices, and several extracellular, cysteine-rich domains (CDRs) (Figure 1). DEG/ENaC channels have both the N- and C- termini on the cytoplasmic side, and most likely form tetramers (5-8). Some members of the DEG/ENaC family (MEC-4, MEC-10, DEG-1, UNC-8, and UNC-105) were proposed to be mechanotransducing channels that



**Figure 1.** Predicted topology for eukaryotic and bacterial mechanosensitive channels. The TRPV channels contain several ankyrin domains at the N-terminus (A), and one pore forming loop (P).  $K_{2P}$  channels have two pore forming loops and a self interacting domain (SID) through which they form dimers. DEG/ENaC sodium channels have a single pore forming loop and three cysteine rich domains (CRDs). The MscL topology and the predicted topology for MscS are also shown for comparison.

mediate touch in neurons and in *C. elegans* (4). The mechanosensitive sodium ion channels are thought to be attached to the extracellular matrix by the MEC-5 and MEC-9 proteins, and to the cytoplasmic microtubule network by the MEC-2 protein (5-8). The DEG/ENaC sodium channel family also includes acid-sensing ion channels (ASIC family) that respond to low pH under chronic inflammatory conditions and tissue damage. ASIC channels are thought to be important in pain perception. Localized

expression of ASIC2 in mechanosensory terminals suggested a role of ASIC2 in touch (9).

*Calcium permeable channels:* The TRP (transient receptor potential) channels are nonselective cationic channels that allow calcium ions into the cell. They are widely distributed in mammals, although, their exact functions are mostly unknown (10, 11). The TRPV (also known as OTRP) subclass of TRP channels contains two known mechanosensitive channels VR-OAC and OSM-9 (4, 11). VR-OAC is an osmotically activated channel related to the vanilloid receptor, and is gated by exposure to hypotonicity. VR-OAC current increases during cell swelling, and decreases during cell shrinkage (12). The OSM-9 channel functions in response to osmotic and nose-touch stimuli in *C. elegans* (13). The TRPV family of channels has a similar topology as the Shaker-related voltage-gated K<sup>+</sup> channels (14), with six transmembrane helices, a pore forming loop between the fifth and sixth transmembrane helices, and the N- and C-termini at the cytoplasmic side of the membrane (Figure 1). The TRPV channels contain several ankyrin domains at the N-terminus (10).

*Drosophila* NOMPC channels have significant sequence similarity in the transmembrane region with the TRP channels, and contain 29 ankyrin repeats at the N-terminus (15). Loss-of-function mutants of NOMPC eliminate mechanoreceptor responses in *Drosophila*. NOMPC was found to be selectively expressed in mechanosensory organs of *Drosophila* (4, 15).

The calcium-permeable, stretch-activated nonselective cation channel Mid1 from yeast was shown to be sensitive to mechanical stress with approximately 30 pS conductance (2). The Mid1 channel is activated in yeast during mating, when the cell

wall is undergoing remodeling, and possibly causes stretch in the membrane. Mid1 has no overall sequence similarity to other known ion channels (2, 4).

*Potassium channels:* The two-pore-domain potassium channel family ( $K_{2P}$ ) constitutes a relatively new family of outward rectifying  $K^+$  channels (14, 16). Two of its members (TREK-1 and TRAAK) are highly flickering stretch activated  $K^+$  channels, with conductances of 100 pS and 45 pS, respectively (17, 18).  $K_{2P}$  channels contain four possible transmembrane helices, an extracellular self interacting domain, and two pore-forming domains (Figure 1). Both the N- and C- termini reside on the cytoplasmic side (4, 16). The C-terminus of TREK-1 is thought to be necessary for mechanosensitivity (17, 19). A charge cluster located at the C-terminus, similar to that of bacterial mechanosensitive channel of large conductance (MscL), appears important in mechanosensitivity of both TREK-1 and MscL (4, 19).  $K_{2P}$  channels associate through extracellular self interacting domains (SID) and form disulfide bridged homodimers (16). TREK-1 and TRAAK channels are activated by cell swelling, shearing, and suction. TREK-1, like ASIC2 sodium channel, is also opened by internal acidification (17, 18).

### **Bacterial mechanosensitive channels**

Mechanosensitive channels are ubiquitous in bacteria and have been proposed to play an important role in managing the transition from high to low osmotic environment. The response to osmotic stress is best characterized in *Escherichia coli* (20). During downshock, when bacteria are shifted from a high osmolarity to a low osmolarity, water enters the cell and causes a large increase in turgor. Bacteria utilize ion channels embedded in their plasma membranes, such as MscL, to respond to large changes in

osmolarity. These channels open under the most stressful conditions and jettison water and solutes from the cytoplasm in an effort to prevent cell lysis during hypo-osmotic shock (21). Small solutes such as monovalent ions and water, as well as large molecules like thioredoxin (12kD), can pass through these nonselective pores (22). Based on their different conductances, three mechanosensitive channels have been identified in *E. coli* (3, 20, 23); mechanosensitive channel of large conductance (MscL), small conductance (MscS), and mini conductance (MscM). These channels are located in the inner membrane of bacteria (24, 25). Mutants lacking both MscL and MscS are severely compromised during osmotic downshock (20). Bacterial mechanosensitive channels might be the simplest mechanosensors, and can serve as good models for more complicated eukaryotic mechanosensitive channels (Figure 1).

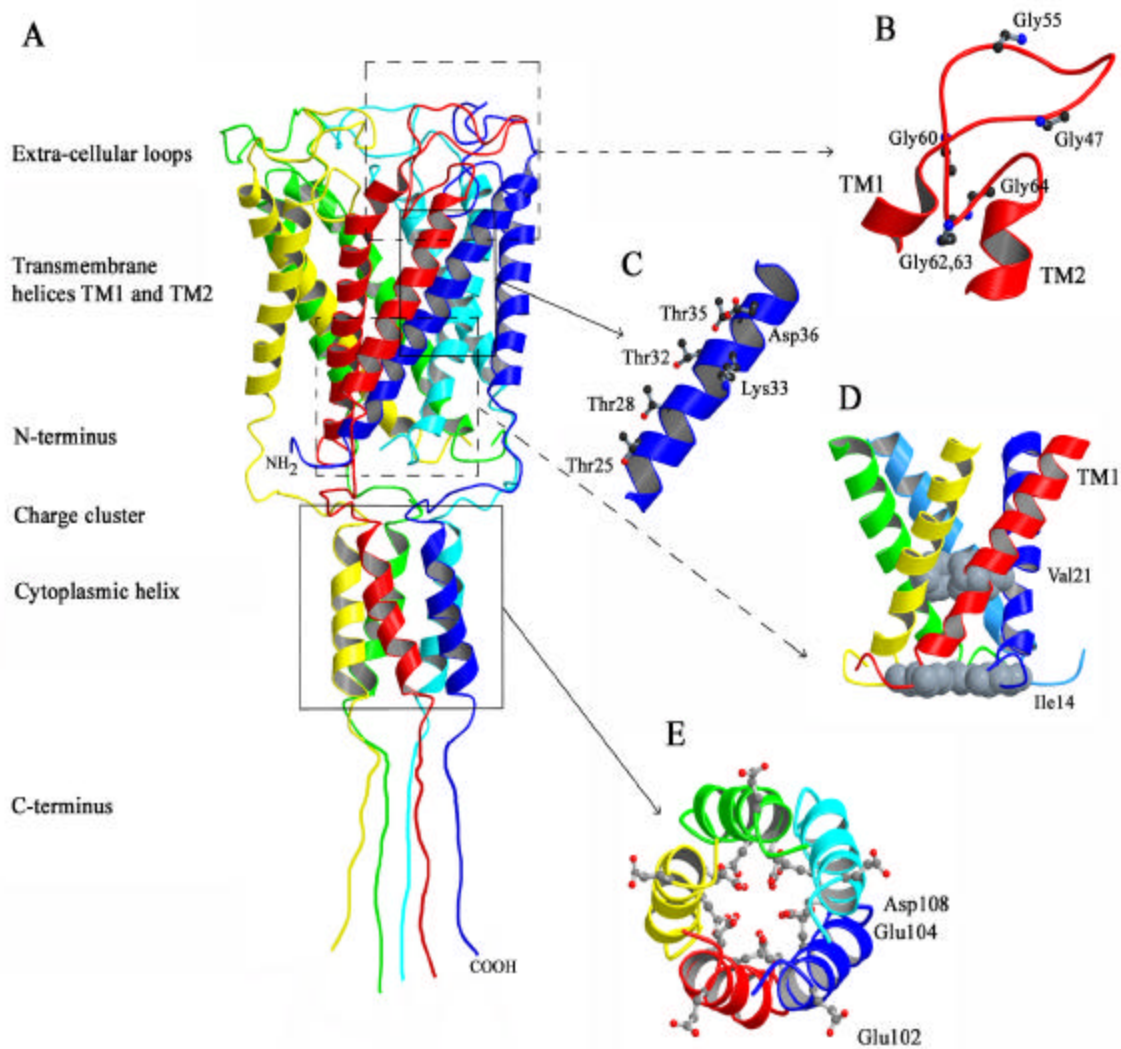
## **MscL**

By far the most extensively studied bacterial mechanosensitive channel is the mechanosensitive channel of large conductance. Electrophysiological measurements suggest that a large pore, similar in size to that found on the extracellular face of the channel (30 Å), would be needed in order to account for the very large conductance (3.5 nS) observed in the open state. The structure of MscL from *Mycobacterium tuberculosis* (Tb\_MscL) provided the first structural insight into mechanosensitive channels (26). Tb\_MscL forms a homopentameric structure organized into two domains, the transmembrane domain and the cytoplasmic domain. The transmembrane domain is made up of ten helices (two per subunit) connected by an extracellular loop, while the cytoplasmic domain consists of five helices that form a left-handed pentameric bundle.



The Tb\_MscL monomer consists of 151 amino acids and can be further subdivided into five segments: N-terminus, the first transmembrane helix (TM1), an extracellular loop, the second transmembrane helix (TM2), and a cytoplasmic domain (Figure 2A). Each of the segments is discussed in more detail below. The pore is aligned along the fivefold symmetry axis and is formed by the first transmembrane helix (TM1) and an extracellular loop from each subunit. The channel has overall dimensions of approximately 85 x 50 x 50 Å and both the N- and C- termini reside on the cytoplasmic side of the membrane. The residue numbering used here (Figure 3) is that of Tb\_MscL, or Tb\_MscL equivalent for experiments that were performed on MscL from *E. coli* (Eco\_MscL).

*N-terminus:* Despite the fact that the first ten residues are not ordered in the crystal structure, the high sequence conservation of the N-terminus together with mutagenesis studies suggests the importance of the N-terminus in channel gating. Deletion of the first two residues ( $\Delta 1-2$ ) or addition of residues at the N-terminus makes little difference in channel function, however, the deletion of the first nine residues ( $\Delta 1-9$ ) is not tolerated (27, 28). More recent N-terminus proteolysis experiments showed that limited N-terminal degradations result in functional channels with increased sensitivity to pressure (29). Disulphide crosslinking studies have demonstrated that by crosslinking the N-terminus to the C-terminal end of the TM2 helix (residues 1 and 91), the channel rarely reaches the fully closed state, and flickers mostly between open state and an intermediately open substate (30). Based on these results, an open state model of MscL was generated, where the disordered N-terminus is proposed to form an additional helix bundle that forms a secondary channel gate (30, 31).



**Figure 2.** A) Structure of the mechanosensitive channel of large conductance. B) The extracellular loops are highly flexible and contain many glycine residues shown as ball-and-sticks. C) The periplasmic side of the channel is lined with hydrophilic mostly threonine residues. D) The crystal structure of MscL is believed to be in the closed state. The constriction of the channel is made up of residues Ile 14 and Val 21, shown in CPK representation. E) Structure of the cytoplasmic helix. The view was rotated 90° towards the reader. The cytoplasmic helix contains a cluster of charged residues which at physiological pH would be expected to repel each other when brought together in the pentameric bundle.

*TM1 Helix:* The first transmembrane helix (residues 15-43) crosses the membrane from the cytoplasm to the periplasm, and together with the other four TM1 helices forms the inside of the pore. The interior of the channel on the periplasmic side is mostly hydrophilic and is formed by the residues Thr25, Thr28, Thr32, Thr35, Lys33 and Asp36 (Figure 2C). The TM1 helix packs against the adjacent TM helix with right-

handed crossing angles of  $-43^\circ$ , and makes little contact with the lipid bilayer. The TM1 helix is tilted approximately  $35^\circ$  with respect to the membrane normal (32), with the channel constriction occurring near the cytoplasmic side of the membrane. The structure of Tb\_MscL is believed to be in the closed or nearly closed state, since the channel constriction formed by hydrophobic residues I14 and V21 is about 2 Å in diameter (Figure 2D). The equivalent residues of the Eco\_MscL also form the constriction site based on numerous mutagenesis studies (33-36). Substitution of all hydrophobic residues at position 20 resulted in fully functional channels (33), however, substitution of polar residues resulted in spontaneous gating, even with no tension applied. In experiments where cysteine residues were substituted for channel constricting residues, reversible modification of hydrophilicity was possible under patch clamp (34). Chemical modifications were consistent with mutagenesis studies, resulting in functional channels when hydrophobic modifying agents were used, and resulting in spontaneous gating when hydrophilic modifying agents were used.

Recently, Perozo et al. (37) performed site-directed spin labeling analysis of Eco\_MscL by EPR spectroscopy under physiological conditions. In these experiments, cysteine mutants were generated for residues in transmembrane helices TM1 and TM2 and modified with methanethiosulfonate spin label. High accessibility of the spin label to molecular oxygen is indicative of membrane exposure, exposure of the spin label to NiEdda is indicative of exposure to aqueous environment, and spectral line shape gives information about spin mobility. The results of this study correlate well with the Tb\_MscL structure. Severely restricted accessibility to NiEdda is observed for residues 19-26, which in the crystal structure form the channel constriction, while strong



**Figure 3.** ClustalX sequence alignment of MscL sequences from *Mycobacterium tuberculosis*, *Escherichia coli*, *Bacillus subtilis*, *Haemophilus influenzae*, and *Erwinia carotovora*. Charge cluster region located at the C-terminus of MscL, important in mechanosensitivity of both MscL and TREK-1 channels, is shown in yellow. Pore constriction is shown in red. Locations of transmembrane helices, and the cytoplasmic helix are shown at the top of the sequence alignment.

periodicity in label mobility is seen for residues 19-38 which indicates a helical structure.

The periodicity of NiEdda and O<sub>2</sub> accessibilities are 180° out of phase, suggesting that one face of the helix is water accessible, while the opposite face of the helix is in a hydrophobic environment. This is in good agreement with the crystal structure, where one face of the TM1 helix forms the water accessible pore and the opposite face is packed against TM2 helix and resides in a hydrophobic environment.

*Extracellular loop:* The connection between the TM1 helix and the TM2 helix is mediated by the extracellular loop formed by residues 44-68. The extracellular loop dips into the channel and lines the extracellular part of the pore. The diameter of the pore is about 20 Å at the extracellular side and narrows down to about 2 Å at the constriction site near the cytoplasmic side of the membrane. The high glycine content of the extracellular loop (Gly47, Gly55, Gly60, Gly62, Gly63, and Gly64) is likely to contribute to the

flexibility in this region seen in the crystal structure (Figure 2B). Mutagenesis studies have shown that mutations in the extracellular loop lead to a gain of function phenotype (38), and proteolytic cleavage experiments have demonstrated that proteolysis of the extracellular loops results in easier opening of the channel (29). The low sequence conservation of the extracellular loop region would appear, however, to be somewhat in disagreement with the functional importance implied by mutagenesis studies.

*TM2 Helix:* The second transmembrane helix (residues 69-89) traverses the lipid bilayer back to the cytoplasmic side on the outside of the TM1 helix bundle, and is responsible for most of the contact with the lipid bilayer. The lipid exposed surface of MscL is composed of approximately 35% TM1 and 65% TM2. The face of TM2 that comes in contact with the lipid bilayer is lined with many hydrophobic residues (Leu69, Leu72, Leu73, Ile77, Phe79, Phe80, Leu81, Phe84, and Phe88) and is more hydrophobic than an average protein core (32, 39, 40). The TM2 helix, like the TM1 helix, is tilted about 35° with respect to the membrane normal; however, unlike the TM1 helices, the TM2 helices from different subunits are separated by ~20 Å and do not make contact with each other.

Site directed spin labeling studies of the TM2 helix (37) revealed that it is slightly less motionally restricted relative to TM1 helix. This is to be expected since TM2 is exposed to the fluid lipid bilayer. As in the TM1 helix, periodicity of oxygen accessibility is in agreement with the crystal structure, where one side of the TM2 faces the lipid bilayer. NiEdda accessibility decreases as the TM2 helix enters the membrane, and sharply increases at residue 100 where it presumably exits the membrane. The spin labeling data deviates from the crystal structure at the C-terminus of the TM2 helix where

the local environment derived from EPR studies does not match the environment derived from the crystal structure. These differences can either be explained by the presence of partially ordered electron density near the C-terminus of the TM2 helix (corresponding to ordered detergent molecules or part of the disordered polypeptide chain), by structural differences between *E. coli* and *M. tuberculosis* homologs, or by differences arising from the use of bilayers and detergent to solubilize Eco\_MscL and Tb\_MscL, respectively. Spin labeling studies of the open state channel will provide important structural information on the environment of the TM1 and the TM2 helices in the open state.

*Cytoplasmic domain:* After the TM2 helix crosses back to the cytoplasm, a second loop connects the TM2 helix to the cytoplasmic helix. Cytoplasmic helices from all five subunits (residues 102-115) associate together to form a left-handed helix bundle. Mutagenesis and proteolysis data have shown that the C-terminus can be removed up to residue 104 without any change in gating properties of MscL (22, 27, 28). A cluster of charged residues (R98, K99, K100, E102, and E104) is present in the loop connecting TM2 to the cytoplasmic helix (Figure 2E and 3) and has been shown to be important for channel gating (4, 19, 23, 41). The overall negative charge of the C-terminal domain, and the relatively low pH (3.7) used for crystal growth have lead to the hypothesis that the cytoplasmic helix bundle is not present at physiological pH; however, there is no experimental evidence for the dissociation of the helix bundle at high pH. The rest of the C-terminus (residues 116-151) does not show any ordered structure in the crystals and correspond to the least conserved region of the protein.

## MscS

### Channel discovery

Both the MscL and MscS channels were functionally identified in 1987, but it was not until 1999 that the *yggB* gene was associated with MscS activity in membrane patches of *E. coli* (20, 42). As shown by a series of *yggB* knockout strains of *E. coli*, the *yggB* gene was required for MscS channel activity. While deletion of either the *mscL* or *yggB* gene appeared to be without a phenotype, mutants lacking both genes become hypersensitive to osmotic downshock. This finding suggested that each of the channels is able to compensate for the other in responding to osmotic stress. KefA, a protein found by homology to contain a MscS domain, also displays MscS-like activity, but was found to be dispensable. Deletion of the *kefA* gene in concert with either the *mscS* gene, or the *mscL* gene was not found to have any phenotype. Although it appears that *kefA* is not required in response to hypo-osmotic shock, it contains a domain highly homologous to *yggB* and does encode a channel of similar conductance to MscS (~1.0 nS).

*E. coli* MscS was extensively analyzed by electrophysiological techniques (41). The channel was shown to have a conductance of approximately 1 nS, roughly one third that of MscL (3.5 nS). The pressure threshold for opening the channel is approximately 60% that of MscL. MscL was found to open in the same patch as MscS at a negative pressure of around 100-200 mmHg for giant spheroplasts or 50-100 mmHg for proteoliposomes. Based on the observed conductance, a pore diameter of approximately 18 Å is estimated for the open state of MscS.

Unlike the MscL channel, MscS exhibits a voltage-dependent gating. Upon reaching its pressure threshold, there is an e-fold increase in open probability per 15 mV

of membrane depolarization (positive voltage) (42). At moderate voltages ( $\pm 20$  mV), MscS has slow kinetics, dwelling in an open or closed state for up to seconds (41). MscS is mostly nonselective with a slight anion selectivity (relative permeability of  $\text{Cl}^- / \text{K}^+ = 3$  (43)).

### Sequence of MscS

Once the *yggB* gene of *E. coli* was identified as encoding the MscS channel, it was possible to search for homologs in other organisms. MscS genes were found to be larger than those found for MscL. While homologs of MscL range from around 130 to 150 amino acids, MscS homologs are typically twice that size (*E. coli* 286aa, *B. subtilis* 371aa, *C. tep* 302aa, *M. jannaschii*, *P. furiosus* 256aa (Figure 4).

MscS is expected to be a homomultimer, although, the exact oligomeric state of MscS remains unknown. An experiment by Sukharev et al. (44) suggested that the functional MscS channel is considerably larger than the pentameric MscL channel. By separating *E. coli* MscS and MscL channels on a size exclusion column and subsequently analyzing the resulting fractions by patch clamp, they were able to show that MscS and MscL activity can be separated. MscS eluted from the column first, indicating that the functional channel is larger than MscL. Based on the size exclusion chromatography, fully assembled MscS channel was estimated to have a molecular weight of 200-400 kDa (44). The analogous estimate of the mass of functional MscL protein was 70 kDa, which was later confirmed to be the molecular weight of the pentamer (75 kDa).





**Figure 4.** ClustalX sequence alignment of MscS sequences from *Escherichia coli*, *Salmonella typhimurium*, *Edwardsiella ictaluri*, *Pyrococcus furiosus*, *Pseudomonas pseudomallei*, *Methanococcus jannaschii*, and *Bacillus subtilis*. Location of putative transmembrane helices predicted by DAS and TMHMM is shown as the top of the sequence alignment.

Hydropathy plots of *E. coli* MscS indicate that the N-terminal half of the protein is highly hydrophobic. Transmembrane helix predictions programs, such as TMHMM, (45-47) and DAS (48), predict either three or four transmembrane helices (the same programs correctly predicted two transmembrane helices for the *M. tuberculosis* MscL). Thus, *E. coli* MscS likely contains three or four transmembrane helices at the N-terminus (Figure 4), followed by a large (160 amino acids) soluble domain. Predictions for the larger MscS proteins from *B. subtilis* and *M. jannaschii* suggest two extra transmembrane helices in addition to the helices predicted for *E. coli* MscS.

### Archaeal MscS-like proteins

Once the *mscL* gene was positively identified, it was possible to search the genomic databases and find homologous channels in a wide array of prokaryotes. Using a similar search criteria as for *mscL* gene, far more homologs of MscS were identified. These results suggest that MscS may be more widely distributed in prokaryotes than MscL. Unlike MscL, MscS homologs appear to be prevalent in *Archaea* as well as the lower eukaryote *S. pombe*. Recently Kloda et al. (49, 50) have identified and characterized two archaeal mechanosensitive channels that appear to be similar to bacterial MscS and MscL. Using the first transmembrane helix of MscL, they identified a homologous channel in the archaea *Methanococcus jannaschii*, termed MscMJ. Using this channel as a search criterion, they found a second mechanosensitive channel MscMJLR. Although the original search was performed using TM1 of MscL, both channels are homologous to MscS. MscMJ was found to have a conductance of approximately 0.3 nS, and was activated over a negative pressure range of 25-80 mmHg. In contrast, MscMJLR was found to have a much larger conductance of 2.0 nS, more similar to MscL, but opened at a much lower and narrower negative pressure range of around 25 mmHg. This indicates that MscMJ and MscMJLR are not strict homologs of MscL and MscS, but instead exhibit mixed channel characteristics.

### MscM

A third mechanosensitive channel of mini conductance (MscM) was functionally identified in *E. coli*, although the encoding gene has not been reported yet. This channel has a conductance less than half that of MscS (0.1-0.4 nS). The channel exhibits slow

opening/closing kinetics relative to those of MscS, and much slower than MscL (51).

Relative to MscL and MscS, MscM has pronounced cation selectivity.

**References:**

1. Blount, P., and Moe, P. C. (1999) *Trends Microbiol* 7, 420-424.
2. Kanzaki, M., Nagasawa, M., Kojima, I., Sato, C., Naruse, K., Sokabe, M., and Iida, H. (1999) *Science* 285, 882-6.
3. Le Dain, A. C., Saint, N., Kloda, A., Ghazi, A., and Martinac, B. (1998) *J Biol Chem* 273, 12116-9.
4. Hamill, O. P., and Martinac, B. (2001) *Physiol Rev* 81, 685-740.
5. Mano, I., and Driscoll, M. (1999) *Bioessays* 21, 568-78.
6. Driscoll, M., and Tavernarakis, N. (1997) *Gravit Space Biol Bull* 10, 33-42.
7. Driscoll, M., and Tavernarakis, N. (2000) *Nat Neurosci* 3, 1232-4.
8. Welsh, M. J., Price, M. P., and Xie, J. (2002) *J Biol Chem* 277, 2369-72.
9. Julius, D., and Basbaum, A. I. (2001) *Nature* 413, 203-10.
10. Harteneck, C., Plant, T. D., and Schultz, G. (2000) *Trends Neurosci* 23, 159-66.
11. Clapham, D. E., Runnels, L. W., and Strubing, C. (2001) *Nat Rev Neurosci* 2, 387-96.
12. Liedtke, W., Choe, Y., Marti-Renom, M. A., Bell, A. M., Denis, C. S., Sali, A., Hudspeth, A. J., Friedman, J. M., and Heller, S. (2000) *Cell* 103, 525-35.
13. Colbert, H. A., Smith, T. L., and Bargmann, C. I. (1997) *J Neurosci* 17, 8259-69.
14. Choe, S. (2002) *Nat Rev Neurosci* 3, 115-21.
15. Walker, R. G., Willingham, A. T., and Zuker, C. S. (2000) *Science* 287, 2229-34.
16. Lesage, F., and Lazdunski, M. (2000) *Am J Physiol Renal Physiol* 279, F793-801.
17. Maingret, F., Patel, A. J., Lesage, F., Lazdunski, M., and Honore, E. (1999) *J Biol Chem* 274, 26691-6.
18. Maingret, F., Fosset, M., Lesage, F., Lazdunski, M., and Honore, E. (1999) *J Biol Chem* 274, 1381-7.
19. Patel, A. J., Honore, E., Maingret, F., Lesage, F., Fink, M., Duprat, F., and Lazdunski, M. (1998) *Embo J* 17, 4283-90.
20. Levina, N., Totemeyer, S., Stokes, N. R., Louis, P., Jones, M. A., and Booth, I. R. (1999) *Embo J* 18, 1730-7.
21. Blount, P., Schroeder, M. J., and Kung, C. (1997) *J Biol Chem* 272, 32150-7.
22. Ajouz, B., Berrier, C., Garrigues, A., Besnard, M., and Ghazi, A. (1998) *J Biol Chem* 273, 26670-4.
23. Sukharev, S. I., Blount, P., Martinac, B., Blattner, F. R., and Kung, C. (1994) *Nature* 368, 265-8.
24. Berrier, C., Coulombe, A., Houssin, C., and Ghazi, A. (1989) *FEBS Lett* 259, 27-32.
25. Oakley, A. J., Martinac, B., and Wilce, M. C. J. (1999) *Protein Science* 8, 1915-1921.
26. Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T., and Rees, D. C. (1998) *Science* 282, 2220-6.
27. Blount, P., Sukharev, S. I., Schroeder, M. J., Nagle, S. K., and Kung, C. (1996) *Proc Natl Acad Sci U S A* 93, 11652-7.
28. Hase, C. C., Le Dain, A. C., and Martinac, B. (1997) *J Membr Biol* 157, 17-25.
29. Ajouz, B., Berrier, C., Besnard, M., Martinac, B., and Ghazi, A. (2000) *J Biol Chem* 275, 1015-22.
30. Sukharev, S., Betanzos, M., Chiang, C. S., and Guy, H. R. (2001) *Nature* 409, 720-4.
31. Sukharev, S., Durell, S. R., and Guy, H. R. (2001) *Biophys J* 81, 917-36.
32. Spencer, R. H., and Rees, D. C. (2001) *Ann. Rev. Bioph. Biomol. Struc.*
33. Yoshimura, K., Batiza, A., and Kung, C. (2001) *Biophys J* 80, 2198-206.
34. Yoshimura, K., Batiza, A., Schroeder, M., Blount, P., and Kung, C. (1999) *Biophys J* 77, 1960-72.
35. Moe, P. C., Levin, G., and Blount, P. (2000) *J Biol Chem* 275, 31121-7.
36. Ou, X., Blount, P., Hoffman, R. J., and Kung, C. (1998) *Proc Natl Acad Sci U S A* 95, 11471-5.
37. Perozo, E., Kloda, A., Cortes, D. M., and Martinac, B. (2001) *J Gen Physiol* 118, 193-206.
38. Maurer, J. A., Elmore, D. E., Lester, H. A., and Dougherty, D. A. (2000) *J Biol Chem* 275, 22238-44.
39. Rees, D. C., DeAntonio, L., and Eisenberg, D. (1989) *Science* 245, 510-3.
40. Wallin, E., Tsukihara, T., Yoshikawa, S., von Heijne, G., and Elofsson, A. (1997) *Protein Sci* 6, 808-15.
41. Sukharev, S. I., Blount, P., Martinac, B., and Kung, C. (1997) *Annu Rev Physiol* 59, 633-57.

42. Martinac, B., Buechner, M., Delcour, A. H., Adler, J., and Kung, C. (1987) *Proc Natl Acad Sci USA* 84, 2297-301.
43. Kloda, A., and Martinac, B. (2002) *Archaea* 1, 35-44.
44. Sukharev, S. I., Martinac, B., Arshavsky, V. Y., and Kung, C. (1993) *Biophys J* 65, 177-83.
45. Sonnhammer, E. L., von Heijne, G., and Krogh, A. (1998) *Proc Int Conf Intell Syst Mol Biol* 6, 175-82.
46. Moller, S., Croning, M. D., and Apweiler, R. (2001) *Bioinformatics* 17, 646-53.
47. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) *J Mol Biol* 305, 567-80.
48. Cserzo, M., Wallin, E., Simon, I., von Heijne, G., and Elofsson, A. (1997) *Protein Eng* 10, 673-6.
49. Kloda, A., and Martinac, B. (2001) *Biophys J* 80, 229-40.
50. Kloda, A., and Martinac, B. (2001) *Embo J* 20, 1888-96.
51. Berrier, C., Coulombe, A., Houssin, C., and Ghazi, A. (1993) *J Membr Biol* 133, 119-27.

**Chapter 2:**  
**Conversion of the Mechanosensitive Channel of Large  
Conductance to a Water-Soluble Form**

## Introduction

Structural characterization of membrane proteins has been extremely slow when compared to their water-soluble counterparts. The total number of “high-resolution” membrane protein structures in Protein Data Bank (PDB) is around 50, while the total number of water-soluble protein structures is near 10,000 (2, 3). Membrane proteins are predicted to compose almost ~30% of the total number of proteins in the human genome, which is higher than the 0.5% seen in PDB (4). Low expression, poor stability in the absence of lipid bilayer, the presence of detergents, and difficulty in forming well-ordered crystals are some of the problems that account for the slow progress (5-7). Respiratory membrane proteins and  $\beta$ -barrel outer membrane proteins compose a large portion of the solved membrane proteins to date. Structural characterization of respiratory membrane proteins has been more successful due to their high natural abundance and presence of cofactors which assist in characterization and phasing. In the case of the mechanosensitive channel of large conductance from *Mycobacterium tuberculosis* (Tb\_MscL), approximately 24,000 crystallization conditions were screened before crystals diffracting to sufficient resolution were obtained (8).

Structural principles of membrane proteins and water-soluble proteins are surprisingly similar given the different environments they reside in. Except for the main difference in surface residue polarity (7, 9), structural differences between membrane proteins and water-soluble proteins include differences in preferred helix-helix interactions and helix-helix packing (10-14). Constrictions imposed by the lipid bilayer and the requirement to span the membrane might be responsible for the increased number of parallel helix orientations in membrane proteins (15). Helix-helix arrangements with a

nearly parallel helix axes are more favored in membrane proteins, contributing to higher helix packing density (10, 14). Interhelical packing in membrane proteins is higher than in soluble proteins and might compensate for the lack of hydrophobic effect driving force in helix-helix association (1, 10, 11, 14, 16). Tighter packing in membrane proteins is due to the smaller side chains present at the helix interfaces relative to side chains found at helix interfaces of water-soluble proteins (10, 11, 14).

Left-handed packing arrangements and anti-parallel helix arrangements are more abundant in membrane proteins than expected from random distribution (1). The distribution of helix packing angles in membrane proteins is different than for soluble proteins (on average +20° for membrane proteins and -35° for soluble proteins) (12). The geometry of side chain packing is, however, similar for left-handed helix-helix pairs embedded in membranes and coiled-coils of soluble proteins (16-18). Transmembrane helices were observed to pack against neighbors in the sequence more often than seen in soluble proteins (12). Insertion of a helical membrane protein into the lipid bilayer has been reported to proceed in pairs of helices (hairpins), which may be responsible for the observed increase in packing against a neighbor in the sequence.

The hydrophobic effect in the membrane spanning region contributes to the formation of secondary structure, but unlike water-soluble proteins, this is not the driving force for the assembly of tertiary and quaternary structure (16, 19). The assembly of transmembrane helices is believed to be driven by van der Waals packing interactions and by interhelical polar interactions (16, 20). In soluble proteins, polar interactions in the core can help folding specificity (21). In membrane proteins, polar interactions can play important roles in helix association; however, due to the hydrophobic environment



of the lipid bilayer and the differences in folding mechanisms of membrane proteins, polar interactions between transmembrane helices can also cause nonspecific interactions and misfolding (20). Unsatisfied interactions of polar side chains in the membrane have unfavorable interactions with lipid bilayer and can increase the probability of formation of unintended hydrogen bonds and helix-helix interactions (20).

The average hydrophobicity of the interior residues in integral membrane proteins is lower than the average hydrophobicity of the exterior residues; however, the average hydrophobicity of interior residues in membrane proteins and soluble proteins is the same (7, 22). In contrast to the “inside-out” hypothesis (23), which suggests that membrane proteins contain polar cores and hydrophobic surfaces, the hydrophobicity of the protein cores in membrane proteins is the same as in water-soluble proteins (7, 22). The accessible surface area and volume of buried residues in the reaction center is similar to those of water-soluble proteins of the same size (7, 9). Protein surfaces are similar to their native environments, polar for water-soluble proteins and hydrophobic for membrane proteins (22).

If the same structural principles apply to both soluble and membrane proteins, it might be possible to convert a membrane protein into a water-soluble protein by changing hydrophobic surface residues that interact with the lipid bilayer to hydrophilic residues seen on surfaces of water-soluble proteins. To date, no integral membrane protein has been successfully converted to a water-soluble form. Some progress was made with phospholamban (24, 25), where a soluble form of the proper oligomeric state was obtained. Phospholamban, however, is a pore-forming transmembrane peptide, and

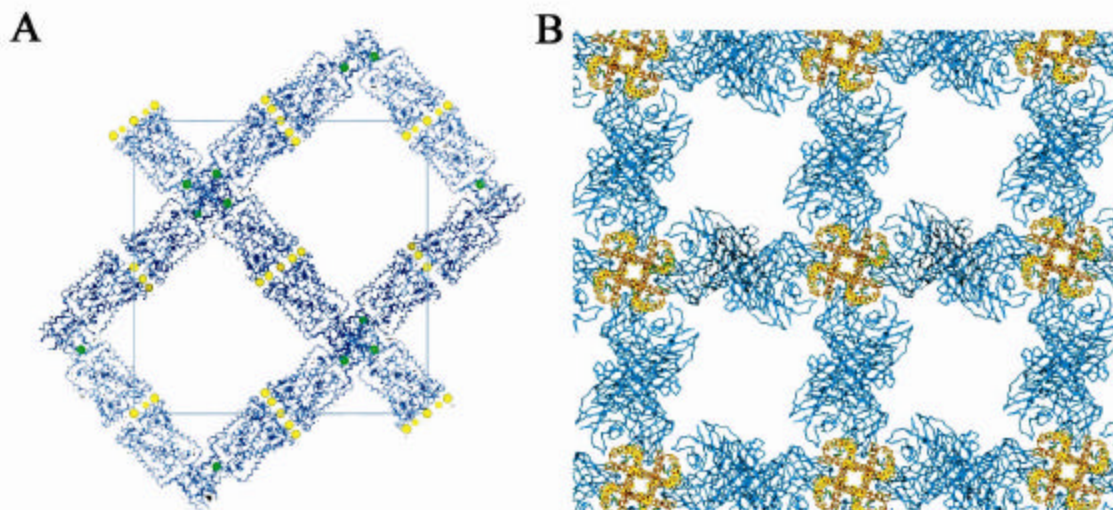
the water-soluble form was a molten globule. Efforts in converting bacteriorhodopsin have also been unsuccessful (25).

The limited number of known membrane protein folds might be one reason why solubilization of membrane proteins has been a difficult task to achieve.  $\beta$ -barrel folds are not suitable for surface redesign because to change the lipid exposed surface, nearly every other residue would need to be mutated, and so the chance of recovering the native-like fold must be very low. On the other hand, many  $\beta$ -barrel outer membrane proteins have been successfully refolded, and it has been postulated that the  $\beta$ -barrel might fold in the periplasm just before insertion to the outer membrane (26-28). Unlike  $\beta$ -barrel folds, fewer residues usually participate in the formation of the lipid interface in  $\alpha$ -helical membrane proteins. The total number of residues that would need to be altered for  $\alpha$ -helical membrane protein, such as bacteriorhodopsin is, however, still relatively large. One possible solution to this problem is to choose a symmetrical molecule such as MscL (homopentamer) (8) or KcsA ( $K^+$  channel from *Streptomyces lividans* which forms a homotetramer) (29). Due to the rotational symmetry of these molecules, relatively few residues have to be changed per monomer. For example, in the case of MscL, the second transmembrane helix accounts for 65% of the total lipid exposed surface. Changing ten residues per monomer (fifty residues in the entire MscL channel) could be enough to approach the hydrophobicity of a water-soluble protein. These properties make MscL one of the simplest model systems with known crystal structure. The disadvantage of choosing multimeric protein is that many subunits have to come together during folding.

Conversion of a membrane protein to a water-soluble protein by redesigning only the surface residues would demonstrate that membrane proteins differ from soluble proteins only by the presence of hydrophobic residues on the surface to interact with the hydrophobic lipid bilayer. A water-soluble version of the MscL would also be useful in overcoming problems associated with structural and thermodynamic analysis of membrane proteins and could conceivably yield a higher-resolution MscL structure, a structure in the open state, or a structure in complex with channel blockers.

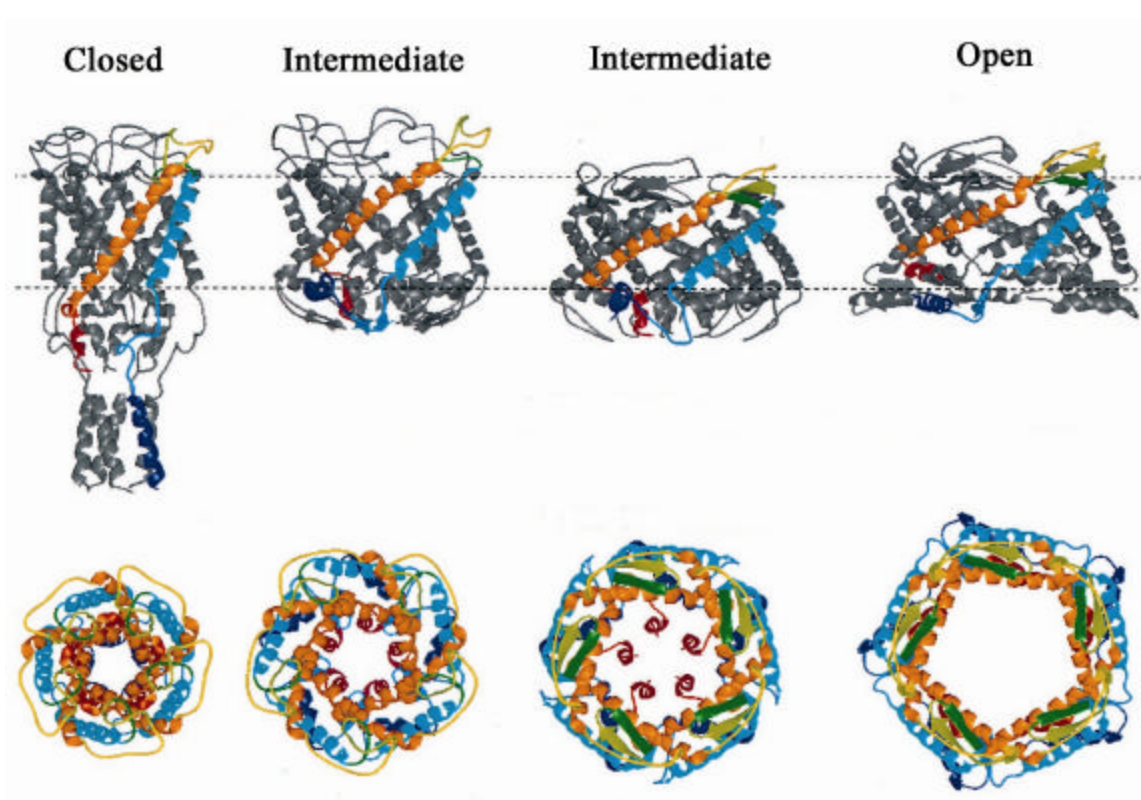
*Higher-resolution structure of Tb\_MscL:* Crystal contacts are generally mediated through the interactions of hydrophilic residues. Since the membrane spanning domains are highly hydrophobic, crystal contacts are mostly limited to water-soluble regions. In the case of KcsA (29) and MscL (8) channels, the hydrophilic surfaces are very small relative to the membrane spanning regions, which results in limited crystal packing. Membrane proteins have detergent present in the crystals, and together with high solvent content, can limit the diffraction resolution (Figure 1A). An increased fraction of hydrophilic surface available for crystal packing has been achieved by co-crystallization of cytochrome-c oxidase and KcsA with antibodies (30-32). In the case of the KcsA crystal structure, Fab fragments form all of the crystal contacts (Figure 1B) and the resolution increased from 3.2 Å to 2.0 Å (29, 32). Obtaining specific antibodies that form a well-ordered crystal lattice is unfortunately time consuming and still remains a great challenge, especially for membrane proteins. A water-soluble variant is expected to crystallize more readily than native MscL, and might diffract to higher-resolution.

*Open state of MscL:* MscL functions in response to osmotic stress by changing between closed and open states, thus acting as a safety valve. Time-resolved patch clamp



**Figure 1.** **A)** Packing diagram of Tb\_MscL channels in the crystal. The crystal contacts are mediated only by the water exposed regions of the protein and by  $\text{Na}_3\text{Au}(\text{S}_2\text{O}_3)_2$ , which is shown as yellow spheres. Adapted from Chang et al. *Science* (1998), 282, 2220-2226. **B)** Crystal packing arrangement in the crystals of KcsA with monoclonal antibody. All of the crystal contacts are mediated by the antibody, resulting in better packing and higher-resolution diffraction. Figure 1B was adapted from Zhou et al. *Nature* (2001), 414, 43-48.

recordings from Eco\_MscL have provided evidence that in the opening process, MscL passes through several intermediate substates before reaching the fully open state (33, 34). Proposed models of the open channel are shown in Figure 2 (35, 36). The crystal structure of MscL (8) is believed to be in the closed or nearly closed state, since the channel constriction is only about 2 Å wide. Determination of the open state structure together with the known structure in the closed state would provide an insight into the motions involved in the opening of the channel, channel gating, mechanism of tension sensing as well as how membranes exert force on proteins. Efforts are in progress to determine the structure of the open state of MscL. Disulphide scanning experiments (R. B. Bass, personal communication) suggest that it might be possible to covalently trap MscL channel in the open state. Recently, disulphide crosslinking experiments have yielded a channel that rarely reaches the fully closed state and that exists mostly between fully open and an intermediate states (35, 36).



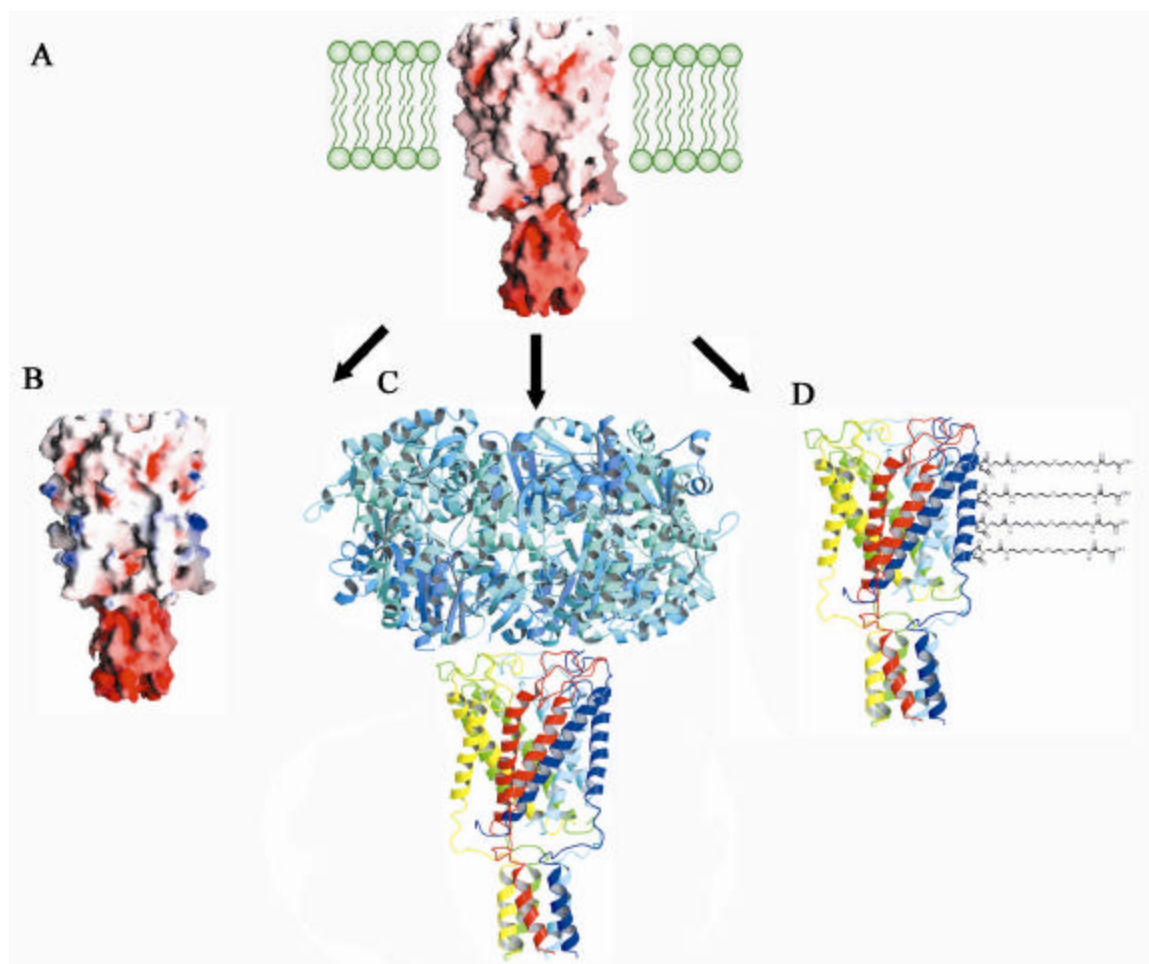
**Figure 2.** Ribbon diagrams of model of Tb\_MscL in the closed, two intermediate, and an open state conformation. Both side view and top view are shown. The channel opening is believed to go through several substates before reaching the fully open state. Adapted from Sukharev et al. *Biophysical J.* (2001), 81, 917-936.

Charged residue substitutions at the channel constriction site (residues 14 and 21) were also found to cause opening of the channel as studied by patch clamp (37-40). Expression of constitutively open channels is difficult, since the open form of MscL would produce large holes in the lipid bilayer. Large quantities of open MscL necessary for crystallization experiments would not be tolerated. If a water-soluble form of MscL could be engineered, it would be expressed in the cytoplasm of the host organism, and would eliminate the problem of creating large pores in cell membranes which are likely toxic to the host.

*MscL complexed with channel blockers:* No structure of a channel complexed with a channel blocker has been reported to date, mainly due to the difficulty in obtaining

high resolution crystals of membrane channels. Routine crystallizations yielding well diffracting crystals make structures of inhibitor complexes more plausible. With water-soluble form of MscL, structures of MscL with various channel blocking reagents can potentially be explored.

In previous unsuccessful attempts to convert a membrane protein to a soluble form, both the position and the identity of altered residues were chosen either by eye, or by “homology” modeling a similar three-dimensional structure of a water-soluble protein (24, 25, 41). It might be necessary to approach this difficult problem more rigorously. Here we report three possible strategies for solubilization of membrane proteins (Figure 3): (i) Change of the membrane protein surface from mostly hydrophobic to more hydrophilic. This approach assumes that the folding pathway of the designed proteins will be similar to that of soluble proteins. We have tried computational protein design and random mutagenesis to determine the positions and identities of residues that require replacement (Figure 3A and 3B). (ii) Attachment of a large soluble protein to either the N- or C- terminus of the insoluble membrane protein. Proteins such as the maltose binding protein (MBP) can be used for the solubilization and increased expression of proteins (Figure 3C). (iii) Chemical modification of introduced cysteine residues. To circumvent the fundamentally different driving forces for folding of soluble proteins and membrane proteins, MscL with cysteine residues introduced on the surface was expressed and solubilized in detergent as wild type MscL protein. The chemically reactive groups were then modified to yield a hydrophilic surface without the need for folding or refolding (Figure 3D). The three approaches have different strengths and weaknesses, and are discussed below.



**Figure 3.** Approaches to creating a water-soluble form of MscL channel. A) Native MscL resides in the lipid bilayer. B) Hydrophobic surface side chains are replaced with more hydrophilic side chains, resulting in a surface that is similar to that of water-soluble proteins. The position and identity of the mutations were selected by a computational algorithm. C) Attachment of large soluble proteins, such as maltose binding protein (MBP), might solubilize the fusion protein. A model of MBP\_MscL is shown in ribbon representation. In reality, the five MBP proteins do not associate, and MBP is fused at the N-terminus of MscL. The relative sizes of five MBPs and MscL are correct. D) Cysteine residues are introduced at the surface of MscL. The surface of MscL is chemically modified with polyethylene glycol at the introduced cysteines. Modifications at only one monomer are shown for clarity.

### Protein design

Protein design algorithms have recently become successful in predicting sequences that fold into a given structure with similar or increased stability when compared to wild type structures (42-47). The same approach that has been used in designing surfaces of water-soluble proteins might be used to rationally redesign the

membrane exposed surface of MscL (with parameters used for water-soluble proteins), thus converting a membrane protein to a water-soluble protein. A non-computational approach has yielded partially soluble bacteriorhodopsin; however, a complete solubilization of an integral membrane protein has not yet been achieved (41). Computational techniques have the advantage of being able to explore very large sequence space, which is experimentally intractable. Disadvantages of this approach include the prerequisite of having a known structure of good quality, and a sensitive potential function that distinguishes between good and bad sequences.

### **Cysteine modification**

Chemical modification of cysteine residues has been used in protein biochemistry for many different purposes, such as probe attachments, solubilization of keratins, cysteine quantification, phasing in crystallography, and even reversible opening of the MscL channel (37, 48-51). Replacement of highly hydrophobic residues of MscL that face the lipid bilayer with cysteines and their subsequent modification with hydrophilic groups should yield less hydrophobic protein surface. Cysteine scanning experiments on MscL (R. B. Bass, personal communication) have provided evidence that cysteine substitutions are tolerated at most positions in the sequence without reducing the expression levels. Both the computational design and random mutagenesis methods require folding and oligomerization in the cytoplasm in the absence of the lipid bilayer. The main advantage of the cysteine modification approach is that the starting point for chemical modification is a well-folded pentameric protein. MscL with introduced cysteine residues, like the native MscL protein, is expressed and inserted into the



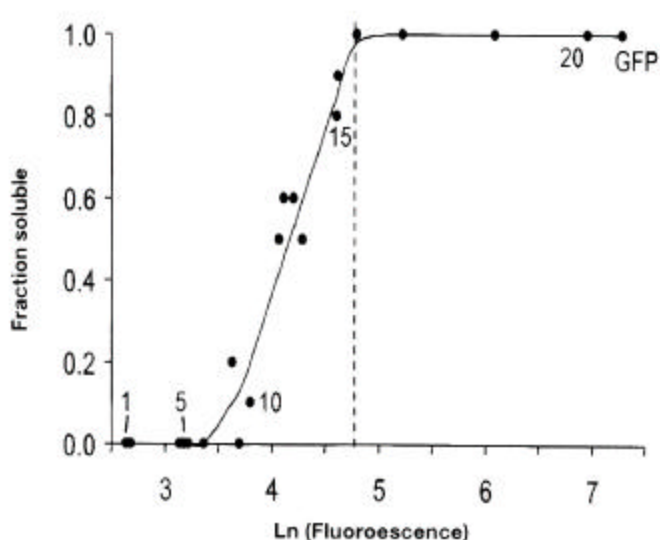
membrane where it forms pentamers, and can be extracted in this native state into detergent. After chemical modification, the detergent can be removed, yielding water-soluble folded protein. One potential disadvantage of this technique is the possibility of unspecific crosslinking between the engineered cysteines. Obtaining monodisperse material, such as required for crystallizations, also limits all cysteine residues to the same modifying groups, however, many different modifying agents (iodoacetamide, iodoacetic acid, bromosuccinic acid) are available and can be tried in parallel.

### **Random mutagenesis**

High throughput screening (HTS) of combinatorial libraries has become an important tool in chemistry and biology in the last decade. Phage display, chemical libraries, and DNA libraries are being used to find targets of desired properties. Drug discovery, study of protein-protein interactions, protein-DNA interactions, and directed evolution are some of the main areas where combinatorial libraries are being used. HTS has two basic prerequisites: large libraries of compounds, proteins, or DNA sequences have to be created which generally contain mostly inactive molecules, but with a few that might have the desired activity. Large libraries have to be screened for desired activity using an appropriate screening assay that is developed to identify positive hits.

Directed random mutagenesis of MscL at selected positions in the second transmembrane helix, together with a screening assay designed to select water-soluble variants could yield the desired water-soluble MscL protein. Since there is no known functional assay for a soluble form of MscL, we have decided to use solubility assays, based on the fluorescence of green fluorescence protein (GFP) developed by Waldo et al

(52). In this assay, DNA containing the randomized protein of interest is fused to the DNA coding for GFP. The intensity of the GFP fusion protein fluorescence signal is correlated with the amount of soluble fusion protein (Figure 4). A library of clones is transformed into bacterial host, plated, and protein expression is induced. The fluorescence of individual bacterial colonies is then measured. High fluorescence clones, coding for soluble GFP fusion proteins, can be then selected.



**Figure 4.** Solubility of proteins expressed in *E. coli* is correlated with fluorescence of the GFP fusion protein. Twenty expressed proteins with different solubilities and fluorescence are shown. Wild type GFP fluorescence is also shown for comparison. Adapted from Waldo et al. *Nature Biotech.* (1999), 691-695.

The three methods used for solubilization of MscL (protein design, random mutagenesis, and cysteine modification) tackle the problem by very different methods and help to maximize the possibility of success in obtaining a water-soluble form of MscL. Obtaining the structures of MscL in the open state, and in complex with channel blocking reagents, is difficult due to the toxicity of the open state to the host organism as well as to the difficulty of obtaining well-diffracting crystals. A water-soluble form of MscL would have many interesting applications including the structure of MscL at higher-resolution, structure of MscL in the open state, structure of MscL in complex with

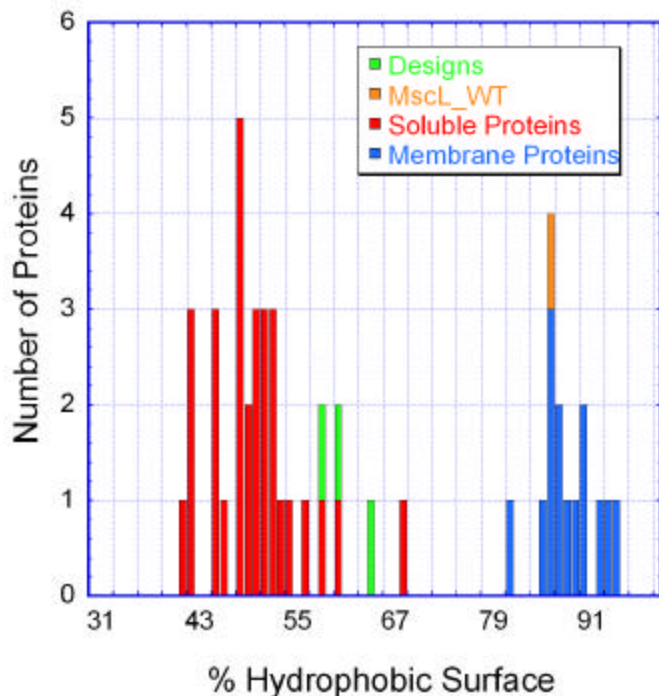
channel blockers, and would serve as a good model system in studying the thermodynamic differences between membrane proteins and water-soluble proteins.

## Results

### Survey of soluble proteins

In order to design a water-soluble form of a membrane protein, it is necessary to estimate how much the surface of a membrane protein must be changed to become water-soluble. Molecular surfaces for a set of 30 single-domain water-soluble proteins were calculated with ORBIT (pdb codes 154l, 1a3d, 1a45, 1ab1, 1adl, 1af7, 1agi, 1agy, 1ah4, 1aii, 1ajz, 1aky, 1bhp, 1bzm, 1enh, 1fel, 1iuz, 1jlm, 1mai, 1mzm, 1omd, 1rcy, 1rec, 1uke, 1who, 1wod, 2chf, 2hcs, 2phy, and 3cbp). The ratio of hydrophobic to hydrophilic surface was analyzed with program EANAL on per atom basis, for the water-soluble set. Hydrophobicity of the surface was also calculated for thirteen membrane proteins (*I*) in the 30 Å thick transmembrane region (pdb codes 1aij, 2prc, 1msl, 1bl8, 1brx, 1occ, 1bgy, 1fum, 1kzu, 1el2, 1f88, 1eul, and 1fx8). The results are shown in Figure 5. On average, the surface of water-soluble protein is 50% hydrophobic, while the transmembrane surfaces were found to be 86% hydrophobic. In this subset of soluble proteins, values as high as 60% were observed for rusticyanin and cutinase demonstrating that proteins with surface hydrophobicity as high as 60% can be water-soluble and do not require detergents for solubilization. The surface of crambin is 68% hydrophobic and requires polar organic solvent, such as ethanol, for solubility in water (53). The membrane spanning region of Tb\_MscL is nearly 85% hydrophobic and 15% hydrophilic. The three calculated designs (D1, D2, and D3) are 64%, 60%, and 58% hydrophobic respectively, in the membrane-

spanning region and 60%, 63% and 60% hydrophobic respectively when the entire surface is used.



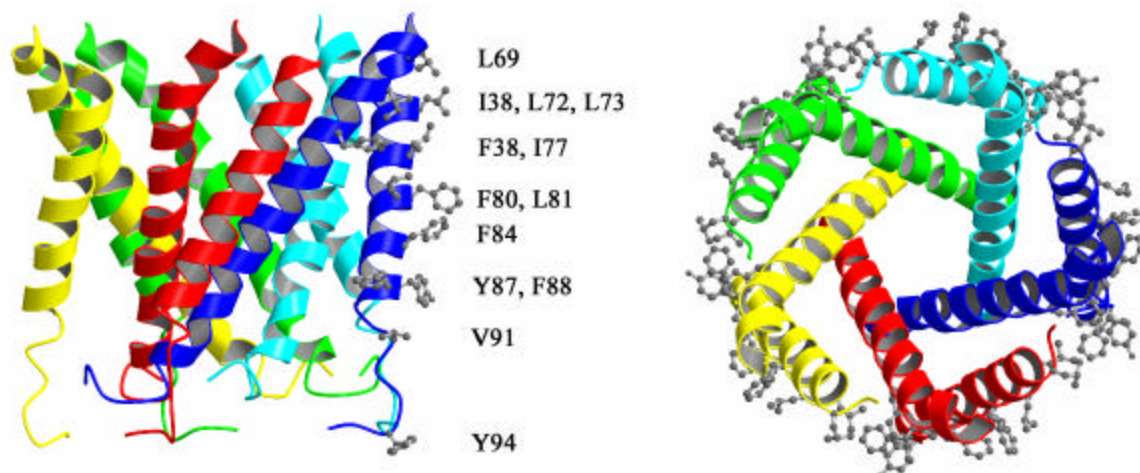
**Figure 5.** Histogram of protein surface hydrophobicity calculated for soluble proteins (red) and membrane proteins (blue). Wild type Tb\_MscL is shown in orange, while the three designs are shown in green. The membrane protein hydrophobicity was calculated in a 30 Å thick transmembrane slab (*I*). The average surface for soluble proteins is 50% hydrophobic, and in membrane proteins is 88% hydrophobic. The redesigned proteins have 58%, 60%, and 64% hydrophobic surfaces, respectively, which is within the range observed for water-soluble proteins. The outlier for water-soluble proteins (68% hydrophobic surface) is crambin. Crambin requires addition of ethanol for solubilization.

### Computational design

*Design:* Details of the computational design are described in the Materials and Methods section. The selected residues for design are shown in Figure 6, and the resulting sequences are shown in Table 1.

*Expression and Fractionation:* The expression of designed proteins was tested by western blots using an anti-His tag antibody. Designs D1 and D3 showed poor expression levels at 37 °C, however, reasonable expression was observed in Terrific Broth media (54) with an 8 hour induction at room temperature. Design D2 did not show any expression under the tested conditions.

In an effort to obtain higher expression and better refolding yields, truncated



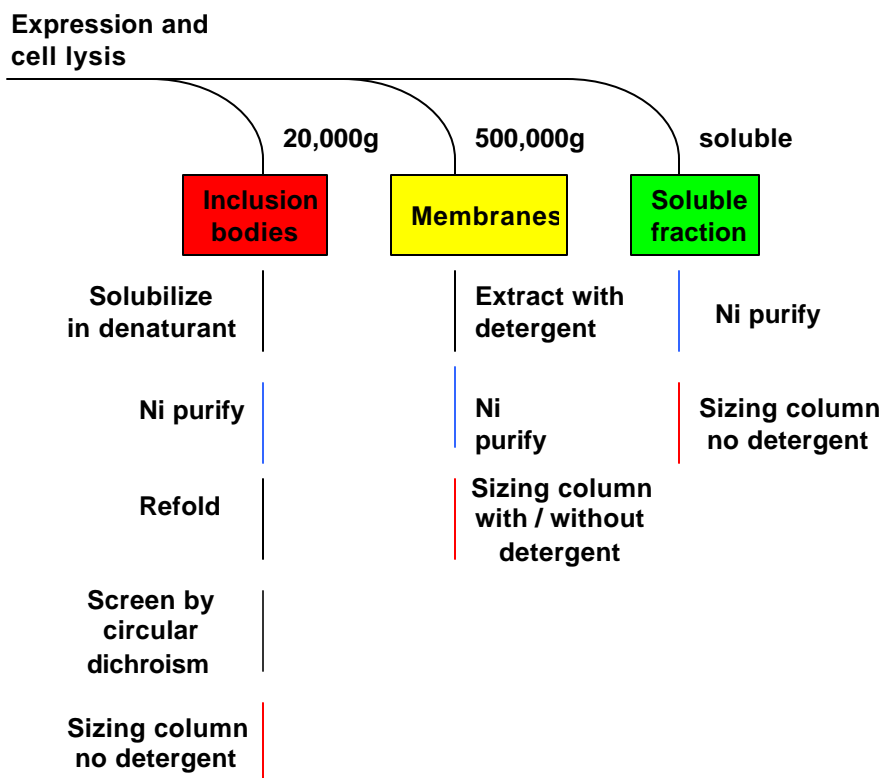
**Figure 6.** Ribbons diagram of the transmembrane region of MscL, with residues chosen for design. The residues were chosen by a combination of visual identification and the program RESCLASS. Residues above and below the region of interest are omitted for clarity. **A)** Side view of MscL with residues shown only on monomer A. **B)** Top view of MscL with selected residues from all five monomers shown in ball-and-stick representation.

versions of the designs were also expressed, in addition to the three designs. The truncations were chosen based on limited proteolysis of wild type Tb\_MscL with trypsin and chymotrypsin (X. Ambroggio, personal communication). Two truncations at positions 100 and 119 were incorporated into the three designs and expressed. These truncations correspond to a transmembrane fragment truncated before the cytoplasmic helix (residue 100) and after the cytoplasmic helix (residue 119). The truncation at position 119 generally expressed better than the full-length designs, except for D2\_T119 where still no expression was observed. Truncation at position 100 did not significantly alter expression levels. None of the expressed designed proteins were found in the soluble fraction after fractionation by ultracentrifugation. All expressed proteins were mostly expressed as inclusion bodies; however, a small fraction of the designed proteins was also found in the membrane. The experimental screening procedure for designed proteins is shown in Figure 7.

*Refolding:* Due to the lack of a functional assay for soluble mechanosensitive channel, refolding conditions were screened by circular dichroism (Figure 8A). In the refolding experiments, ionic strength (0-1M NaCl), pH (4-9), denaturant (Urea/GdmCl) and type of dilution (fast/slow) were systematically varied. Designs D2, D2\_T100, and D2\_T119 did not express and thus refolding could not be attempted. Designs D1 and D3, as well as the D1 and D3 termination mutants (D1\_T100, D1\_T119, D3\_T100, and D3\_T119), all precipitated upon refolding into a low pH buffer (pH 4-6). Refolding of designs D1 and D3 showed no visible aggregation above pH 6.0 independent of salt concentrations, and displayed a significant increase in helical content in their circular dichroism spectra. A circular dichroism wavelength scan of wild type Tb\_MscL in 0.05% DDM is shown in Figure 8A. Minima at 208 nm and 222 nm (typical for helical proteins) are roughly of the same magnitude in wild type Tb\_MscL. Wavelength scans for ten different refolding conditions of D1 are shown in Figure 8B. The protein concentration was the same in all of the samples. The wavelength scans show an increase in helicity at pH 8.0 or above, although, the ratio of the 208 and 222 minima is different than the one observed for wild type Tb\_MscL. Based on the circular dichroism spectra,

	34	38	69	72	73	77	80	81	84	87	88	91	94
<b>WT</b>	<b>PHE</b>	<b>ILE</b>	<b>LEU</b>	<b>LEU</b>	<b>LEU</b>	<b>ILE</b>	<b>PHE</b>	<b>LEU</b>	<b>PHE</b>	<b>TYR</b>	<b>PHE</b>	<b>VAL</b>	<b>TYR</b>
<b>D1</b>	<b>GLU</b>	<b>LYS</b>	<b>GLN</b>	<b>ARG</b>	<b>GLU</b>	<b>GLU</b>	<b>LYS</b>	<b>GLN</b>	<b>LYS</b>	<b>GLU</b>	<b>LYS</b>	<b>LYS</b>	<b>ARG</b>
<b>D2</b>			<b>GLN</b>		<b>ARG</b>	<b>GLU</b>	<b>ASN</b>		<b>LYS</b>	<b>GLU</b>	<b>LYS</b>	<b>LYS</b>	
<b>D3</b>	<b>GLN</b>	<b>GLN</b>	<b>GLN</b>	<b>ASN</b>	<b>GLU</b>	<b>GLU</b>	<b>ASN</b>	<b>GLN</b>	<b>LYS</b>	<b>GLU</b>	<b>SER</b>	<b>ARG</b>	<b>ASN</b>

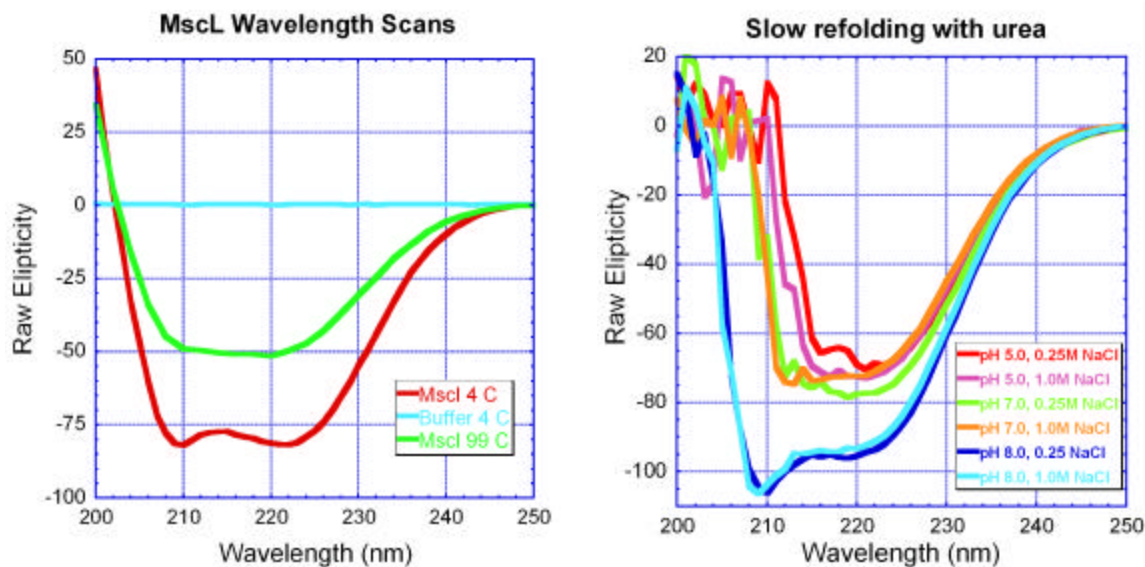
**Table 1.** The amino acid sequences for WT Tb\_MscL, and designs D1, D2, and D3 calculated by ORBIT.



**Figure 7.** Experimental screening of designed proteins. Designed proteins are separated by centrifugation into inclusion bodies, membrane and soluble fractions, purified by nickel affinity chromatography and screened by circular dichroism and by gel filtration.

the designed proteins form some helical structures upon refolding, however, most likely they do not fully refold to the native state of wild type MscL.

Refolded proteins that showed an increase in helicity upon refolding were further analyzed by size exclusion chromatography and dynamic light scattering to determine the oligomeric and aggregation state. An analytical Superdex200 (Pharmacia) was calibrated with dextran,  $\beta$ -amylase, ADH, BSA, CAII, and cytochrome c and refolded variants were passed over the sizing column in the refolding buffer. All samples were observed to be aggregated and eluted at the void volume of the column (Table 2).



**Figure 8.** Circular dichroism measurements of Tb\_MscL. A) Wavelength scans of WT Tb\_MscL at 4 °C, 99 °C, and the buffer baseline scan. B) Wavelength scans of the D1 design refolded from urea by dialysis into pH 5.0, 7.0 and 8.0 with 0.25M and 1.0M NaCl. The increase in helicity for D1 variant refolded at pH 8.0 is shown in blue and turquoise.

Membrane fractions were extracted with 1% DDM and the extracted designs D1 and D3 were purified by nickel affinity chromatography in presence of 0.05% DDM. D1 and D2 solubilized in DDM were found to be aggregated, as observed by analytical sizing chromatography.

*MBP fusion:* Maltose binding protein has been used as a fusion protein to improve expression level and solubility of recombinantly expressed proteins which otherwise form inclusion bodies or show poor expression (55-57). We have made MBP\_MscL N-terminal fusion proteins with the wild type MscL, as well as with the D1, D2, and D3 designs to circumvent the solubility and expression level problems of the three designs (Figure 3C). MBP\_Tb\_MscL was also made as a control. Designs MBP\_D1 and MBP\_D3 showed slightly increased expression level over D1, and D3 proteins, and after fractionation were found mostly in the soluble fraction. A small amount was still found in the inclusion bodies, but most was present in the soluble



	Expression	Fractination			Refolding				CD	Size Exclusion	
		Insoluble	Membrane	Soluble	Slow		Fast			Insoluble	Soluble
					Urea	Gdm	Urea	Gdm			
D1	OK	YES	YES	NONE	12	10	--	6	Some helicity	YES	NO
D1_T100	OK	YES	YES	NONE	4	--	--	--	Some helicity	YES	NO
D1_T119	GREAT	YES	YES	NONE	4	--	--	--	Some helicity	YES	NO
D2	NO	--	--	--	--	--	--	--	--	--	--
D2_T100	NO	--	--	--	--	--	--	--	--	--	--
D2_T119	NO	--	--	--	--	--	--	--	--	--	--
D3	LOW	YES	YES	NONE	4	--	--	--	Low signal	--	--
D3_T100	LOW	--	--	--	--	--	--	--	--	--	--
D3_T119	OK	YES	YES	NONE	4	--	--	--	Some helicity	YES	NO

**Table 2.** Summary of experimental results.

fraction. No protein was found in the membrane fraction, even for the MBP\_Tb\_MscL. MBP\_D2 design showed no expression, just as observed without the MBP fusion protein. MBP\_D1 and MBP\_D3 fusion proteins can be purified on an amylose resin and are soluble in the absence of detergent. Unfortunately, the proper oligomeric state of MscL, a pentamer, was not observed. MBP\_D1 and MBP\_D3 form aggregates of approximately 2,000,000 Da molecular weight as determined by size exclusion chromatography and light scattering. Fusion of MBP to designs D1 and D3 helps the solubility of the D1 and D3 designs, and both fusions are water-soluble without detergent. These fusions, however, do not form monodisperse pentamers, but instead form large soluble aggregates.

### Random mutagenesis

*Expression of fusion proteins:* The expression of fusion proteins was first tested in small cultures by western blots using an anti-His tag antibody. Fusions D1\_GFP and D3\_GFP showed poor expression levels at 37 °C. Reasonable expression levels were

observed, however, in Terrific Broth media (54) with an 8 hour induction at 30 °C or at room temperature. D2\_GFP fusion protein did not show any expression under the tested conditions. In addition to the three design fusions, wild type fusion Tb\_MscL\_GFP was also expressed as a control.

*Fluorescence of fusion proteins:* Surprisingly, Tb\_MscL\_GFP fusion showed no fluorescence at 37 °C, and very low fluorescence at 30 °C or at room temperature when expressed in shaking cultures. MscL fusions D1\_GFP and D3\_GFP unexpectedly showed higher fluorescence than the Tb\_MscL\_GFP fusion. Since the D1 and D3 proteins are insoluble and aggregated in the absence of GFP fusion, no fluorescence was expected. Higher expression levels than wild type, together with help in solubilization by GFP fusion, are probably the cause of the increased fluorescence over wild type. Solubilization of D1 and D3 was also seen in the MBP fusions. No fluorescent colonies were detected at 37 °C, 30 °C or 25 °C for any MscL\_GFP fusion proteins, when the cells were grown on plates. The experimental setup did not allow for measurement of detectable fluorescence signal in bacterial colonies, due to the low expression level seen with these constructs. The GFP protein solubility assay was developed and tested only on water-soluble proteins (52) and the low expression level of membrane proteins make it impractical for soluble MscL screening.

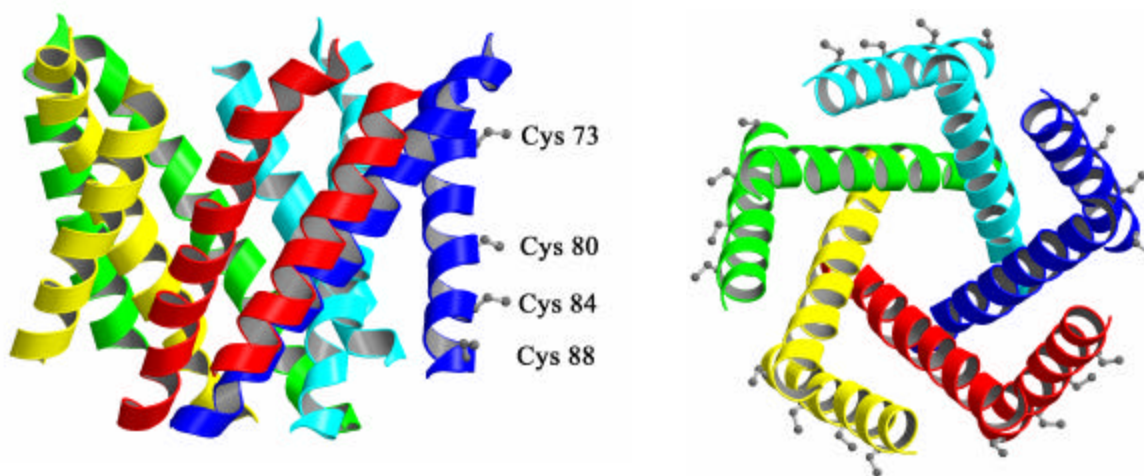
### **Cysteine modification**

Cysteine residues were introduced at positions 73, 80, 84 and 88, resulting in a four-fold mutant MscL\_Cys4. Four cysteines per monomer (20 per pentamer) were chosen as a starting point for chemical modifications (Figure 9). MscL\_Cys4 shows



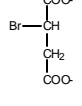
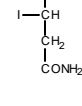
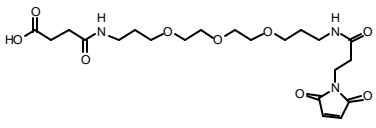
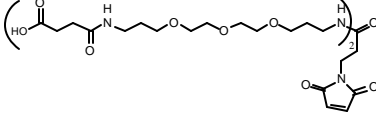
similar expression levels as wild type MscL and behaves in the same way as wild type MscL during purification. MscL\_Cys4 shows a similar CD spectra to Tb\_MscL, and has the same retention time on size exclusion column in the presence of detergent. Removal of detergent resulted in protein aggregation, as seen for the wild type MscL protein. Based on these experiments, MscL\_Cys4 is assumed to form a native-like, pentameric structure.

MscL\_Cys4 was first modified with the fluorescent probe iodoacetamidofluorescein to test whether the selected cysteines are accessible to modification. The modified protein was run on SDS-PAGE gel, and the fluorescence was imaged prior to Coomassie staining. MscL\_Cys4 can be readily modified by iodoacetamidofluorescein, with no further increase in fluorescence after 30 minutes at room temperature.

Modification with iodoacetamide, iodoacetic acid, bromosuccinic acid and



**Figure 9.** Ribbons diagram of the transmembrane region of MscL, with residues chosen for replacement with cysteines for subsequent modification. The residues were chosen by a combination of visual identification and the program RESCLASS. Residues above and below the region of interest are omitted for clarity. **A)** Side view of MscL with residues shown only on monomer A. **B)** Top view of MscL with selected residues from all five monomers shown in ball-and-stick representation.

Compound	Structure	Modification efficiency	State without detergent
Iodoacetamide		50%	Aggregated
Iodoacetic acid		50%	Aggregated
Bromosuccinic acid		50%	Aggregated
Iodosuccinimide		50%	Aggregated
pPEG		100%	Aggregated
2pPEG		100%	Partly soluble

**Table 3.** MscL\_Cys4 modification reagents. The reaction efficiencies of MscL\_Cys4 with iodoacetamide, iodoacetic acid, bromosuccinic acid and iodosuccinimide were approximately 50% as determined by a DTNB assay. Reaction efficiencies of pPEG reagents were 100% as determined by mass spectrometry. MscL\_Cys4 modified with iodoacetamide aggregated upon removal of detergent. Nearly 50% of the protein modified with 2pPEG did not aggregate and remained in the soluble fraction as determined by size exclusion chromatography.

iodosuccinimide (Table 3), in the presence of detergent, resulted in proteins that showed the same retention time on a sizing column as wild type MscL. Removal of detergent from modified proteins, however, caused the protein to aggregate and elute in the void volume of the column. Ellman's reagent (DTNB) was used to test for the presence of free unmodified cysteines, and showed that approximately 25% of cysteines were modified. The reaction efficiency was increased when MscL\_Cys4 was modified on a Ni-column, resulting in approximately 50% modifications with iodoacetic acid.

Maleimide-activated precise polyethylene glycol (pPEG; Gryphon Sciences) was reacted with MscL\_Cys4 on Ni-resin under similar conditions (pH 7.0) as iodoacetic

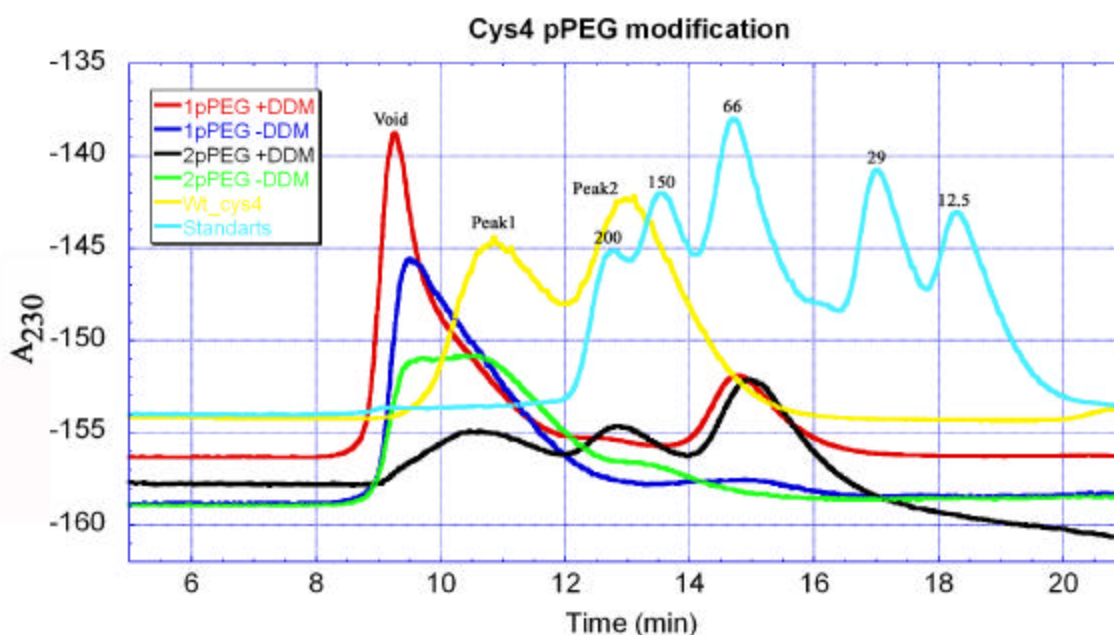
acid. Based on mass spectrometry results, all of the cysteine residues were modified with pPEG (M.W. of MscL\_Cys4 monomer is 18,516 Da, and M.W. of 1pPEG modified monomer is 20,422 Da, corresponding to four pPEG units per monomer). Two different lengths of pPEG (1pPEG and 2pPEG) were used to modify MscL\_Cys4 (table 3). 1pPEG modified MscL contains single pPEG unit at four different sites in each monomer, while 2pPEG modified MscL contains two pPEG units at four sites in each monomer. Both 1pPEG and 2pPEG reaction proceeded to completion, resulting in modification of all cysteines.

Analytical gel filtration chromatography showed that one week after modification, single pPEG modified MscL (1pPEG\_MscL\_Cys4) was mostly aggregated both in the presence and absence of detergent. 2pPEG\_MscL\_Cys4 was tested two days after the modification reaction. 2pPEG\_MscL\_Cys4 did not show aggregation in the presence of detergent, and after detergent removal showed only partial aggregation. Nearly 50% of the protein modified with 2pPEG did not aggregate and remained in the soluble fraction as determined by size exclusion chromatography (Figure 10).

These results suggest that single pPEG modification at four cysteine positions is not enough to solubilize MscL; however, double pPEG modifications are at the borderline of being able to solubilize MscL in aqueous solution in the absence of detergent. Further experiments exploring the time dependence of aggregation, stabilizing conditions that prevent aggregation, and characterization of the soluble 2pPEG\_MscL\_Cys4, will be necessary.

*MscL aggregation:* Tb\_MscL shows a double peak on a gel filtration column. The first peak (peak1) corresponds to approximately 400 kDa and the second peak to

approximately 175 kDa (Figure 10). The second peak (peak2) is believed to be either MscL pentamer with detergent micelle, or a dimer of pentamers. The rod-like shape of MscL and the presence of detergent make it difficult to distinguish between these possibilities. The larger molecular weight of peak1 is not a fully aggregated state, but instead is some well defined higher association state of MscL. Peak1 appears over time, even if the starting material is purely peak2 initially. Low pH (3.5), high pH (10.0), Fos12 detergent and high concentration of glycerol (20%) are some of the factors that reduce association. Both peak1 and peak2 produce crystals under the same crystallization conditions; however, peak2 forms crystals more easily (R.H. Spencer, personal communication). Replacement of the entire C-terminal domain of MscL has no effect on this association, and both peaks are still observed. Modification of the



**Figure 10.** Analytical gel filtration chromatography traces for MscL\_Cys4 and pPEG modified MscL\_Cys4 (Superdex200 column (Pharmacia)). Molecular weight standards are shown in cyan ( $\beta$ -amylase (200kDa), ADH (150kDa), BSA (66kDa), CAII (29kDa), and cytochrome c (12kDa)). MscL\_Cys4 is shown in yellow, 1pPEG with DDM in red, 1pPEG without DDM in blue, 2pPEG with DDM in black, and 2pPEG without DDM in green. Tb\_MscL shows a double peak, corresponding to approximately 400 kDa (peak1) and to 175 kDa (peak2).

transmembrane region with pPEG has also little effect. These observations suggest that the periplasmic loop regions are likely responsible for the observed association. This is consistent with the fact that the periplasmic loops form one of the two major crystal packing interfaces in the  $Tb\_MscL$  crystals. Search for conditions that prevent higher order association is in progress.

## Discussion

The folding of  $\alpha$ -helical integral membrane proteins has been postulated to follow a two-step process, involving (i) formation of individual stable  $\alpha$ -helices and (ii) their association to tertiary and quaternary structures (4, 15, 16, 20). If this model is correct, then one possible problem with the designed water-soluble MscL variant can be the folding and association of the pentamer in aqueous environment. According to the postulated folding pathway, wild type MscL first forms well-folded monomers in the membrane, which then associate to form the functional channel. Since this process happens in the membrane, the driving force of assembly is not the hydrophobic effect, but most likely van der Waals interactions. Assembly of the pentamer might be difficult to accomplish in an aqueous environment since individual MscL monomers contain small hydrophobic cores and outside the membrane environment may not be stabilized enough to fold by themselves. In order to fold the water-soluble MscL variant in an aqueous environment, five monomers might have to come together at the same time and fold together. Calculations with the program QSURF estimate that during monomer folding, approximately  $7800 \text{ \AA}^2$  of hydrophobic area is buried, and approximately  $17500 \text{ \AA}^2$  is buried upon pentamer formation. Due to the difficulty in associating five monomers at

the same time, the cause of the observed nonspecific aggregation is likely the lack of hydrophobic environment of the lipid bilayer.

All three approaches discussed in this study are based on the three-dimensional structure of MscL, so that the position and geometry of the residues interacting with the lipid bilayer can be identified. The structure used in the computational design must be of a good quality and have a well-defined backbone. The computational design algorithm does not consider backbone movement and only identifies the amino acid identity and rotamer conformation at the selected positions. If a protein structure can exist in more than one well-defined state, and the two states are very different, there is an inherent flexibility built into this system which might be problematic for protein design that only considers one state.

GFP solubility screening assay has been tested on both monomeric and multimeric proteins (52), however, it has not been tested previously for membrane proteins. Low expression levels of membrane proteins make this assay difficult to use. Wild type MscL protein is expressed and inserted into membrane where it most likely oligomerizes to its native pentameric form. A GFP fusion protein with Tb\_MscL shows very low fluorescence level, most likely due to low expression level. Computationally designed D1 and D3 proteins form large, nonspecific aggregates. Fusions with computationally designed proteins show higher fluorescence than the Tb\_MscL\_GFP fusion. Since D1 alone is aggregated and D1\_GFP shows higher fluorescence than the Tb\_MscL\_GFP fusion, this assay is not appropriate for screening for a water-soluble version of MscL. The fusion of the designed proteins to either GFP or to MBP causes the fusion proteins to form small aggregates of incorrect oligomeric state.



The cysteine modification of MscL\_Cys4 with 2pPEG showed the most promising results. Modified MscL contained soluble protein in the proper oligomeric form after two days in the absence of detergent. The cysteine modification approach, unlike the computational design and random mutagenesis approaches, does not require experimental folding or refolding of the soluble protein. The starting material for cysteine modifications is a properly folded pentameric MscL, so that no further folding or refolding is necessary. Refolding of membrane proteins has been very difficult, which might explain why the cysteine modification approach appears most promising. The second reason why cysteine modification with 2pPEG has been most encouraging is that the 2pPEG is relatively large, compared to simple amino acid substitutions. The molecular weight of 2pPEG is approximately 800 Da. The 2pPEG\_MscL\_Cys4 contains four 2pPEG units (3200 Da) which accounts for approximately 17% of the total molecular weight of the protein. 1pPEG modified MscL, which corresponds to approximately 10% of the total molecular weight of the protein, was not sufficient to solubilize MscL. At least in the case of polyethylene glycols, it appears that large modifications are necessary to solubilize MscL.

The three different approaches attempted in this study (Figure 3) clearly show that folding or refolding of redesigned membrane proteins in an aqueous solution is problematic. The folding pathways of membrane proteins and soluble proteins are very distinct, and make computational design and high throughput screening impractical. Modification of already folded and functional membrane proteins has been shown in this study to be far more successful. Cysteine residues are tolerated at most positions in a protein sequence, and do not prevent membrane insertion and folding. Active membrane

proteins can be extracted in their native state into detergents, where they can be modified. Manipulating positions of cysteines in the sequence, changing the number of sites available for modifications, and choosing the nature of the modifying reagent, make this a useful method in future attempts for solubilization of membrane proteins.

## Materials and Methods

### Computational Modeling

The computational design algorithm ORBIT (Optimization of Rotamers By Iterative Techniques (42-47)) was used to redesign the surface of the *M. tuberculosis* mechanosensitive channel of large conductance (Tb\_MscL) to make it water-soluble. Three different calculations were run to select both the amino acid identities and side chain conformations for the selected residues. Coordinates for the Tb\_MscL structure used in the calculations were obtained from Brookhaven Protein Data Bank entry 1MSL. In the first design (D1) and third design (D3) (Table 1), thirteen positions on the surface of both transmembrane helices TM1 and TM2 were selected (residues F34, I38, L69, L72, I77, F80, L81, F84, Y87, F88, V91, and Y94). These residues are hydrophobic in the wild type structure and point outward into the membrane (Figure 6). In the second design (D2), only eight residues pointing into the membrane on the surface of the TM2 helix were included in the calculation (Residues L69, L73, I77, F80, F84, Y87, F88, and V91). Ten amino acids (Ala, Arg, Asn, Asp, Gln, Glu, His, Lys, Ser, Thr) were considered at the selected surface positions. At position 69, only amino acids (Ala, Asn, Asp, Gln, Glu, Ser, and Thr) known to favorably interact with N-terminus of helix were allowed. At position 91, only amino acids (Ala, Asn, Arg, Gln, His, Lys, Ser, and Thr)

known to favorably interact with the C-terminus were allowed. Protonated histidine was used in the calculations. In all calculations, the entire pentamer was used, resulting in 40 positions in the D2, and 65 positions in the D1 and D3 calculations. Strict five-fold symmetry was applied, forcing the same amino acid identities and rotamers at the equivalent positions in all five monomers. The remainder of the structure was fixed as a template during the sequence selection calculation.

Due to the large combinatorial complexity of the D1 and D3 calculations (the D1 calculation consists of 65 positions in the pentamer with 13195 rotamers yielding  $2.65 \times 10^{147}$  possible sequences) deterministic methods could not be used and sequence selection was done with Monte-Carlo simulations. The resulting Monte-Carlo sequences were validated by running multiple Monte-Carlo simulations with different random number seeds and obtaining the same lowest energy sequences from three different runs. In the D2 calculation (40 positions instead of 65) a deterministic method (based on dead-end elimination) was used, and resulted in the same sequence as the Monte-Carlo simulations. All calculations were run on R10000 SGI processors, with 1000 cycles of Monte-Carlo with  $10^6$  steps per cycle. Computational details, potential functions, and parameters for van der Waals, solvation, and hydrogen bonding were used as previously described (42-47). For designs D1 and D2, the hydrogen bond depth well was decreased to 5 kcal/mol instead of the 8 kcal/mol used in the D3 calculation. In the D1 and D2 calculations, no distance attenuation was used for the electrostatic component of the force field ( $q_1 q_2 / \epsilon r$ ), as opposed to  $(q_1 q_2 / \epsilon r^2)$  used in the D3 calculation. In the D1 and D2 calculations, a small helix propensity factor of 0.5 was used with the Baldwin helix propensity set (58-65). No helix propensity was used in the D3 calculation. The amino

acid sequences selected by the algorithm for the three designs are D1: F34E, I38K, L69Q, L72R, L73E, I77E, F80K, L81Q, F84K, Y87E, F88K, V91K, and Y94R; D2: L69Q, L73R, I77E, F80N, F84K, Y87E, F88K, V91K; D3: F34Q, I38Q, L69Q, L72N, L73E, I77E, F80N, L81Q, F84K, Y87E, F88S, V91R, Y94N (Table 1).

*Protein Expression:* All three designed proteins (D1, D2, and D3) were constructed by a combination of inverse PCR and USE mutagenesis (66), using a *Mycobacterium tuberculosis* MscL gene in plasmid pet19b (Novagen). All mutants were transformed into BL21-DE3 (Invitrogen) host and verified by sequencing. All designed variants were first grown in small cultures (1L of Terrific Broth (54) with 10 g/L glucose and 100 µg/ml of ampicillin) to an OD of 1.0 at 37 °C, then transferred to 25 °C and induced for 8 hours by IPTG and 20 g/L lactose. Inductions at 37 °C showed poor expression. Pellets were resuspended in the lysis buffer (25 mM Tris pH 7.5, 150 mM NaCl, with ~50 µg/ml of lysozyme and 10 µg/ml DNaseI) and lysed in a continuous flow French press (Emulsiflex 5000) at 20000 psi. Lysed cells were fractionated into soluble, membrane, and insoluble fractions by first centrifugation at 2500g for 5 minutes to remove unbroken cells, then at 20,000g for 45 minutes to remove inclusion bodies and large aggregates, and finally at 500,000g (100,000 rpm) for 30 minutes in a Beckman TL100 bench top analytical ultracentrifuge to isolate membranes. Proteins were considered soluble if they stayed in the supernatant after 30 minutes centrifugation at 500,000g (Figure 7).

*Refolding:* The insoluble fraction was resuspended in 10 M urea or 8 M GdmCl in 100 mM Tris pH 7.5, and left stirring in the cold room for two hours. The insoluble fraction contained large amounts of impurities, so all variants were first purified by nickel

affinity chromatography prior to refolding. Nickel beads were first stripped in 100 mM EDTA, washed, recharged with 100 mM NiSO<sub>4</sub>, and then equilibrated in 30 mM imidazole in either 10 M urea or 8 M GdmCl. The resuspended fractions were passed over nickel beads, washed twice in 100 mM imidazole in 10 M urea or 8 M GdmCl and eluted in 300 mM imidazole in either 10 M urea or 8 M GdmCl. Fast dilution refolding was performed by slowly dripping 50  $\mu$ l of nickel purified protein into a rapidly stirring 25 ml beaker with the given buffer. After dilution, the protein was concentrated by nickel affinity again, without the use of denaturant. Slow dilution refolding was performed by overnight button dialysis of 50  $\mu$ l of nickel purified protein against 25 ml of buffer using 10 kDa cutoff dialysis membrane.

All samples that did not show visible precipitation were screened by circular dichroism (CD). CD data were collected on an Aviv 62DS spectrometer equipped with a thermoelectric unit and using a 0.1 cm path length cell. Data were collected from 250 nm to 200 nm, sampling every 2 nm with an equilibration time of 5 minutes and an averaging time of 1 s. Samples that showed helical content were run on an analytical sizing Superdex200 column (Pharmacia) in the refolding buffer. The exact protein concentration was not determined due to the small volume of protein used in these studies. The estimated protein concentrations in refolding studies from SDS-PAGE gel was 0.5-1.0 mg/ml. The summary of results is shown in Table 2.

### **Cysteine modification**

The four largest hydrophobic side chains were chosen for the initial modification study (residues 73, 80, 84, and 88). Cysteine residues were introduced at these sites in

the same way as for the computational design, resulting in the four-fold mutant denoted MscL\_Cys4. Four cysteines per monomers give rise to twenty cysteines available to modification on the protein surface of the pentamer (Figure 9). The proper balance must be found for the numbers of cysteines introduced in a helix. Too many cysteines can cause undesired crosslinking; however, too few cysteines available for modification will result in aggregation in absence of detergent. The total number of cysteines and their position in the sequence are two parameters that will require further optimization.

MscL\_Cys4 modifications with the thiol-directed reagents iodoacetamide, iodoacetic acid, bromosuccinic acid, and iodosuccinimide, were performed in solution at pH 8.5 (50 mM Tris pH 8.5, 100 mM NaCl, 0.05% DDM) at room temperature for 0.5-2.0 hours. DTNB (dithiobisnitrobenzoic acid) (67) was used to determine the number of free cysteines after modification, relative to the unmodified MscL\_Cys4 (Table 3). Chemical modifications were also performed on the Ni-column. MscL\_Cys4 protein was bound to the nickel resin and washed with wash buffer (50 mM Tris pH 8.5, 2 mM BME, 0.05 % DDM, and 150 mM NaCl) to reduce the cysteines. The BME was removed by washing the nickel resin with ten column volumes of reaction buffer (50 mM Tris pH 8.5, 0.05 % DDM, and 150 mM NaCl). Ni beads were then removed from the hand column and incubated with 1-10mM modifying agent in reaction buffer. Reactions were typically stopped after 0.5-2 hours by removing the modifying agent by washing the resin with wash buffer. Modified proteins were eluted in the presence or absence of detergent with 300 mM imidazole in reaction buffer.

## Random mutagenesis

*Plasmid:* The plasmid was constructed as described by Waldo et al. (52). Briefly, DNA coding for wild type MscL, D1, D2, or D3 was inserted into a pet28 vector, followed by a linker sequence coding for amino acids GSAGSAAGSGEF, followed by the GFP sequence. The same GFP variant used by Waldo (52), which folds properly in *E. coli*, was used in this study.

*Fluorescence:* Colonies were grown at 37 °C on Terrific Broth (TB) media (54) containing plates with 30 µg/ml kanamycin and 10 g/L glucose on nitrocellulose membrane (OSMONICS 142 mm) for ~12 hrs or until ~1 mm large colonies were seen. The membrane was then transferred with colonies onto induction TB plates containing 30 µg/ml kanamycin, 20 g/L lactose, and 2 mM IPTG. The plates were then shifted to room temperature, 30 °C and 37 °C, and protein production was induced for up to 24 hours. Fluorescence was monitored by excitation through a 488 nm filter (Edmund Scientific) and emission was monitored through a 520 nm filter by a digital camera or by eye.

*Random oligonucleotide design:* Deletion of thymidine in the first two positions in each codon of the randomized oligonucleotide will eliminate most hydrophobic residues (Val, Ile, Phe, Leu, Trp) as well as Met, Tyr, and Cys and all undesired stop codons. Incorporation of 38.4% A, 32.0% C and 29.6% G in the first position, 59.9% A, 24.7% C, and 15.4% G in the second position and 30.6% A, 25.5% C, 23.5% G, and 20.4% T in the third position in the randomized triplet will result in a library containing 4.1% of Ser, 8.3% of each of the following amino acids: Ala, Asn, Asp, Gln, Glu, His, Lys, Thr, Pro, and Gly, and 12.3% of Arg.

## **Acknowledgements**

We thank R. Bass and K. Locher for many useful ideas and discussions, M. Meyer for her work on the GFP fusions and MBP fusions, D. Bolon for help with the computational design, R. Spencer for the wild type Tb\_MscL gene used in this study, and C. Becker for his work on pPEG modification.



## References:

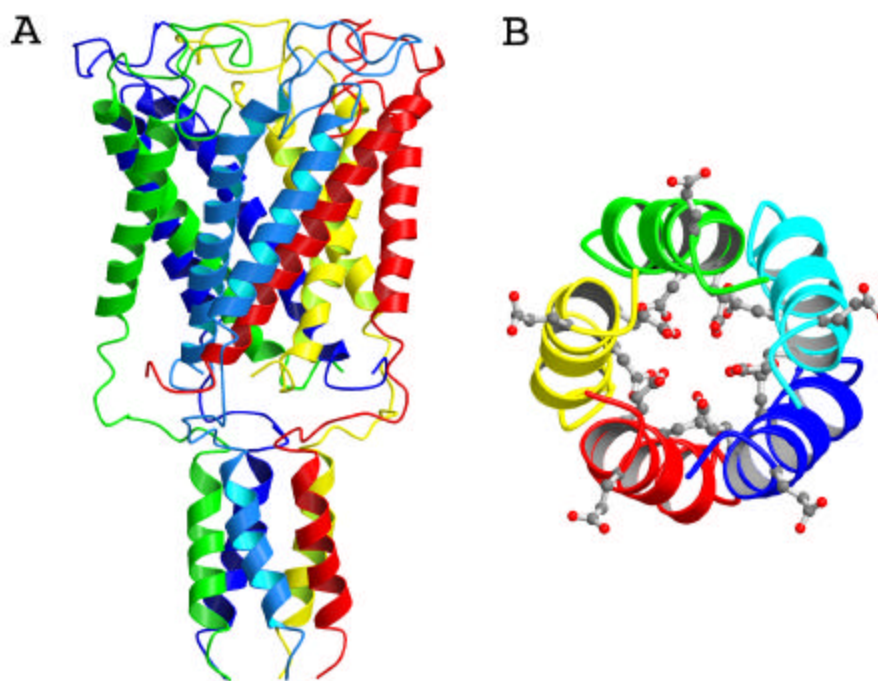
1. Spencer, R. H., and Rees, D. C. (2001) *Ann. Rev. Bioph. Biomol. Struct.*
2. Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O., and Abola, E. E. (1998) *Acta Crystallogr D Biol Crystallogr* 54, 1078-84.
3. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res* 28, 235-42.
4. Popot, J. L., and Engelman, D. M. (2000) *Annu Rev Biochem* 69, 881-922.
5. Kuhlbrandt, W., and Gouaux, E. (1999) *Curr Opin Struct Biol* 9, 445-7.
6. Heijne, G. (1996) *Prog. Biophys. molec. Biol.* 66, 113-139.
7. Cramer, W. A., Engelman, D. M., Von Heijne, G., and Rees, D. C. (1992) *Faseb J* 6, 3397-402.
8. Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T., and Rees, D. C. (1998) *Science* 282, 2220-6.
9. Rees, D. C., Komiya, H., Yeates, T. O., Allen, J. P., and Feher, G. (1989) *Annu Rev Biochem* 58, 607-33.
10. Javadpour, M. M., Eilers, M., Groesbeek, M., and Smith, S. O. (1999) *Biophys J* 77, 1609-18.
11. Eilers, M., Shekar, S. C., Shieh, T., Smith, S. O., and Fleming, P. J. (2000) *Proc Natl Acad Sci USA* 97, 5796-801.
12. Bowie, J. U. (1997) *J Mol Biol* 272, 780-9.
13. Bowie, J. U. (1997) *Nat Struct Biol* 4, 915-7.
14. Russ, W. P., and Engelman, D. M. (2000) *J Mol Biol* 296, 911-9.
15. Ubarretxena-Belandia, I., and Engelman, D. M. (2001) *Curr Opin Struct Biol* 11, 370-6.
16. Lemmon, M. A., and Engelman, D. M. (1994) *FEBS Lett* 346, 17-20.
17. Langosch, D., and Heringa, J. (1998) *Proteins* 31, 150-9.
18. Gurezka, R., Laage, R., Brosig, B., and Langosch, D. (1999) *J Biol Chem* 274, 9265-70.
19. White, S. H., and Wimley, W. C. (1999) *Annu Rev Biophys Biomol Struct* 28, 319-65.
20. Zhou, F. X., Cocco, M. J., Russ, W. P., Brunger, A. T., and Engelman, D. M. (2000) *Nat Struct Biol* 7, 154-60.
21. Bolon, D. N., and Mayo, S. L. (2001) *Biochemistry* 40, 10047-53.
22. Rees, D. C., DeAntonio, L., and Eisenberg, D. (1989) *Science* 245, 510-3.
23. Engelman, D. M., and Zaccai, G. (1980) *Proc Natl Acad Sci USA* 77, 5894-8.
24. Frank, S., Kammerer, R. A., Hellstern, S., Pegoraro, S., Stetefeld, J., Lustig, A., Moroder, L., and Engel, J. (2000) *Biochemistry* 39, 6825-31.
25. Li, H., Cocco, M. J., Steitz, T. A., and Engelman, D. M. (2001) *Biochemistry* 40, 6636-45.
26. Sen, K., and Nikaido, H. (1991) *J Biol Chem* 266, 11295-300.
27. Fourel, D., Mizushima, S., and Pages, J. M. (1992) *Eur J Biochem* 206, 109-14.
28. Freudl, R., Schwarz, H., Stierhof, Y. D., Gamon, K., Hindennach, I., and Henning, U. (1986) *J Biol Chem* 261, 11355-61.
29. Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T., and MacKinnon, R. (1998) *Science* 280, 69-77.
30. Ostermeier, C., Iwata, S., Ludwig, B., and Michel, H. (1995) *Nat Struct Biol* 2, 842-6.
31. Ostermeier, C., Harrenga, A., Ermler, U., and Michel, H. (1997) *Proc Natl Acad Sci USA* 94, 10547-53.
32. Zhou, Y., Morais-Cabral, J. H., Kaufman, A., and MacKinnon, R. (2001) *Nature* 414, 43-8.
33. Biggin, P. C., and Sansom, M. S. (2001) *Curr Biol* 11, R364-6.
34. Cruickshank, C. C., Minchin, R. F., Le Dain, A. C., and Martinac, B. (1997) *Biophys J* 73, 1925-31.
35. Sukharev, S., Durell, S. R., and Guy, H. R. (2001) *Biophys J* 81, 917-36.
36. Sukharev, S., Betanzos, M., Chiang, C. S., and Guy, H. R. (2001) *Nature* 409, 720-4.
37. Yoshimura, K., Batiza, A., and Kung, C. (2001) *Biophys J* 80, 2198-206.
38. Yoshimura, K., Batiza, A., Schroeder, M., Blount, P., and Kung, C. (1999) *Biophys J* 77, 1960-72.
39. Ou, X., Blount, P., Hoffman, R. J., and Kung, C. (1998) *Proc Natl Acad Sci USA* 95, 11471-5.
40. Moe, P. C., Levin, G., and Blount, P. (2000) *J Biol Chem* 275, 31121-7.
41. Chen, G. Q., and Gouaux, E. (1997) *Protein Eng* 10, 1061-6.
42. Dahiyat, B. I., and Mayo, S. L. (1996) *Protein Science* 5, 895-903.

43. Dahiyat, B. I., Gordon, D. B., and Mayo, S. L. (1997) *Protein Sci.* 6, 1333-1337.
44. Dahiyat, B. I., and Mayo, S. L. (1997) *Proc. Natl. Acad. Sci. U.S.A.* 94, 10172-10177.
45. Dahiyat, B. I., Sarisky, C. A., and Mayo, S. L. (1997) *J. Mol. Biol.* 273, 789-796.
46. Dahiyat, B. I., and Mayo, S. L. (1997) *Science* 278, 82-87.
47. Malakauskas, S. M., and Mayo, S. L. (1998) *Nature Structural Biology* 5.
48. Bass, R. B., Coleman, M. D., and Falke, J. J. (1999) *Biochemistry* 38, 9317-27.
49. Bass, R. B., and Falke, J. J. (1999) *Structure Fold Des* 7, 829-40.
50. Schrooyen, P. M., Dijkstra, P. J., Oberthur, R. C., Bantjes, A., and Feijen, J. (2000) *J Agric Food Chem* 48, 4326-34.
51. Schrooyen, P. M., Dijkstra, P. J., Oberthur, R. C., Bantjes, A., and Feijen, J. (2001) *J Agric Food Chem* 49, 221-30.
52. Waldo, G. S., Standish, B. M., Berendzen, J., and Terwilliger, T. C. (1999) *Nat Biotechnol* 17, 691-5.
53. Teeter, M. M., and Hendrickson, W. A. (1979) *J Mol Biol* 127, 219-23.
54. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual 2nd ed.*, Cold Spring Harbor Laboratory Press.
55. Riggs, P. D., and Ausebel, F. M. *Current Protocols in Molecular Biology*, Greene Associates/Wiley Interscience, New York.
56. di Guan, C., Li, P., Riggs, P. D., and Inouye, H. (1988) *Gene* 67, 21-30.
57. Maina, C. V., Riggs, P. D., Grandea, A. G., 3rd, Slatko, B. E., Moran, L. S., Tagliamonte, J. A., McReynolds, L. A., and Guan, C. D. (1988) *Gene* 74, 365-73.
58. Padmanabhan, S., Marqusee, S., Ridgeway, T., Laue, T. M., and Baldwin, R. L. (1990) *Nature* 344, 268-70.
59. Padmanabhan, S., and Baldwin, R. L. (1991) *J Mol Biol* 219, 135-7.
60. Chakrabarty, A., Schellman, J. A., and Baldwin, R. L. (1991) *Nature* 351, 586-8.
61. Fairman, R., Armstrong, K. M., Shoemaker, K. R., York, E. J., Stewart, J. M., and Baldwin, R. L. (1991) *J Mol Biol* 221, 1395-401.
62. Chakrabarty, A., Doig, A. J., and Baldwin, R. L. (1993) *Proc Natl Acad Sci USA* 90, 11332-6.
63. Chakrabarty, A., Kortemme, T., and Baldwin, R. L. (1994) *Protein Sci* 3, 843-52.
64. Padmanabhan, S., York, E. J., Gera, L., Stewart, J. M., and Baldwin, R. L. (1994) *Biochemistry* 33, 8604-9.
65. Baldwin, R. L. (1995) *Biophys Chem* 55, 127-35.
66. Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G., and Galas, D. J. (1989) *Nucleic Acids Res* 17, 6545-51.
67. Riddles, P. W., Blakeley, R. L., and Zerner, B. (1983) *Methods in Enzymology* 91, 49-61.

**Chapter 3:**  
**Characterization and stability of MscL**

## Cytoplasmic helix

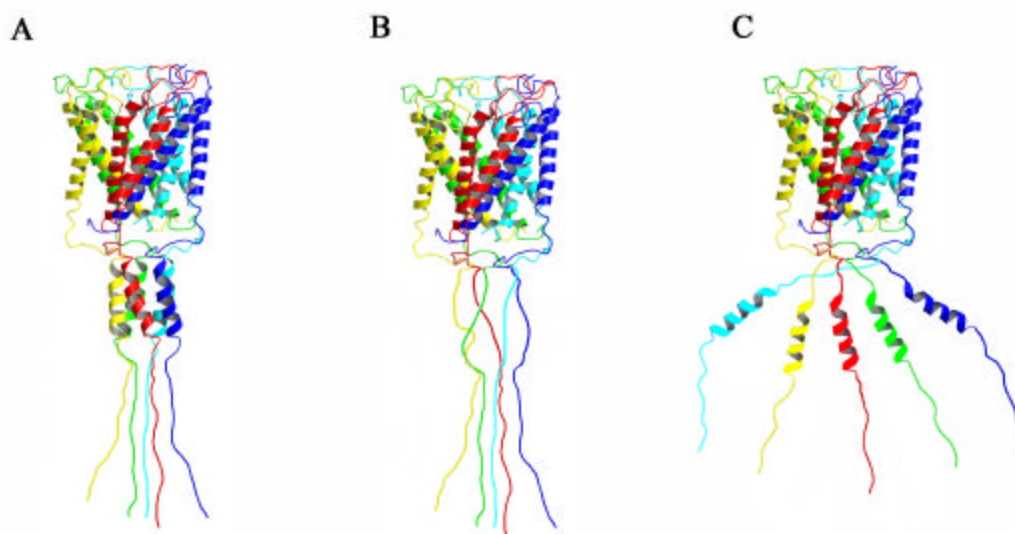
The mechanosensitive channel from *Mycobacterium tuberculosis* forms a homopentameric structure (1) organized into two domains, the transmembrane domain and the cytoplasmic domain (Figure 1A). The transmembrane domain is made up of ten helices, two per subunit, connected by an extracellular loop, while the cytoplasmic domain consists of five helices forming a left-handed pentameric bundle. Mutagenesis and proteolysis data have shown that the C-terminus can be removed up to residue 104 without any change in the gating properties of Eco\_MscL (2-4). A cluster of charged residues (R98, K99, K100, E102, and E104) in the loop connecting TM2 to the cytoplasmic helix has been shown to be important for channel gating. The rest of the C-



**Figure 1.** A) Ribbon diagram of the structure of Tb\_MscL. B) Cytoplasmic helix of Tb\_MscL viewed parallel to the helix bundle axis, showing the charged residues present at the center of the helix bundle.

terminus (residues 116-151) does not show any ordered structure in the crystals and corresponds to the least conserved region of the protein (1). The structure of Tb\_MscL has raised some discussion about the physiological significance of the cytoplasmic helix. The cytoplasmic helix contains a cluster of charged residues which at physiological pH would be expected to repel each other when brought together in the pentameric bundle. The structure was determined at pD 3.5, at which aspartic acid would be protonated (pKa of aspartic acid is 3.9). At pH 7.0, the negative charges would likely repel each other (unless the charge was neutralized by a cation), which would cause the helix bundle to fall apart. However, there is no experimental evidence for the dissociation of the helix bundle at physiological pH.

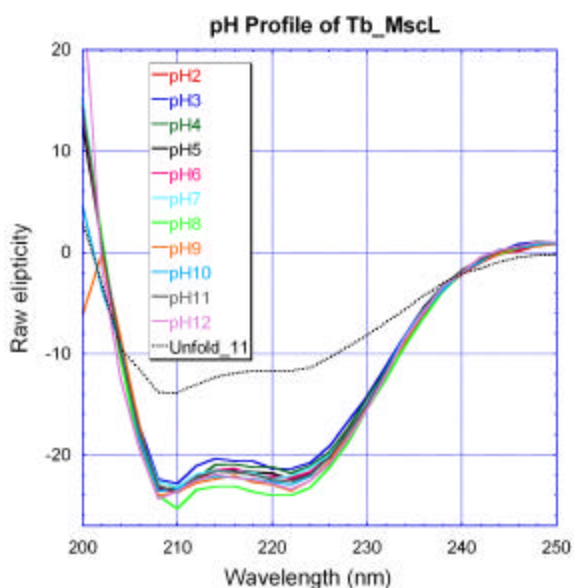
Three possible models of the cytoplasmic helix bundle at physiological pH are shown in Figure 2. The helix bundle can stay unchanged if the charges are somehow neutralized (Figure 2A), or the bundle can dissociate into individual helices (Figure 2C),



**Figure 2.** Possible cytoplasmic helix models at physiological pH. The cytoplasmic helix can stay together as a bundle (A), or it can dissociate and become completely disordered (B), or the bundle can dissociate into individual helices (C).

or the region that forms the helical bundle at lower pH can become completely unstructured (Figure 2B). If the bundle dissociates into individual helices, the individual helices can either stay away from the lipid bilayer or can partition into the bilayer. To test this hypothesis, we have used synthetic peptides corresponding to the cytoplasmic helix and studied their helicity at different pH, detergent, and salt concentrations by circular dichroism. We have also measured the change in helicity of the full-length MscL channel as a function of pH.

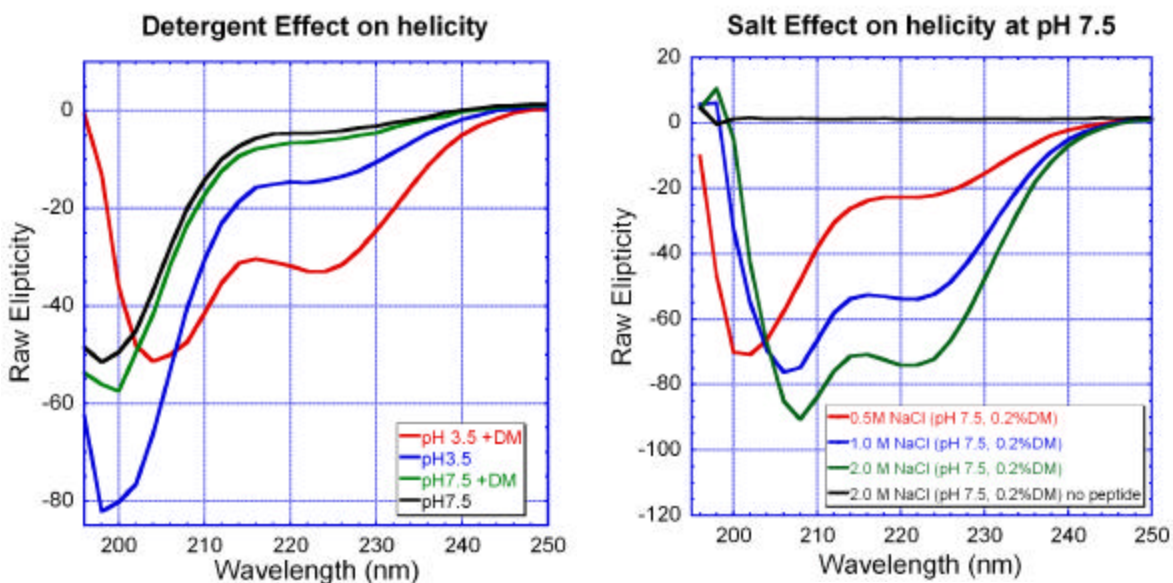
The circular dichroism spectra of the full-length MscL should change if the total helical content of the protein changes. The cytoplasmic helix makes up approximately 22% of the total helical content of MscL, and if it becomes unstructured at physiological pH, a decrease in the CD signal should be observed. Circular dichroism spectra recorded from 200 nm to 250 nm at pH 2-12 are shown in Figure 3. There is no significant change in helicity between pH 2-12 for the full-length MscL protein, suggesting that the helix bundle does not become completely unstructured at physiological pH, but either somehow overcomes the charge repulsion or dissociates into single helices.



**Figure 3.** Circular dichroism wavelength scans of Tb\_MscL at pH 2.0-12.0 at 4 °C. Unfolded protein (pH 11.0, 99 °C) is shown for comparison.

The C-terminal helix peptide HP1 (residues 102-118, sequence: EVEQPGDTQVLLTEIR) was tested for helicity under various conditions by circular dichroism. HP1 showed no helical content at pH 7.5, and very small increase in helicity at pH 3.5 (Figure 4A). Addition of detergent (0.2% DM) caused a strong increase in helicity of HP1 at pH 3.5, however, at pH 7.5, the HP1 helicity remained almost unchanged (Figure 4A). Increasing concentration of salt (0.5-2.0 M NaCl), together with the presence of detergent, showed a strong increase in HP1 helicity even at pH 7.5 (Figure 4B).

Full-length C-termini (residues 102-151) showed no helicity at pH 7.0; however, when attached to a pentameric scaffold, a significant increase in helicity was observed (G. Kochendoerfer, personal communications).



**Figure 4.** The effect of detergent, pH and salt concentration on the C-terminal helix peptides. Circular dichroism wavelength scans were done at 4 °C. (A) The effect of detergent (0.2% DM) on peptide helicity at pH 3.5 and 7.5. (B) The effect of salt on peptide helicity in presence of detergent at pH 7.5.

The formation of the C-terminal helix seems much easier at lower pH (used for crystallization); however, based on the peptide studies, it is possible to form the C-terminal helix at higher pH ~7.5 by either immobilizing the C-terminus on a pentameric scaffold, or by adding salt and detergent. If the helical bundle dissociates into single helices (Figure 2C), the individual helices can partition into the lipid bilayer or at the lipid bilayer interface. Increased helicity of the C-terminal peptide in presence of detergent at both pH 3.5 and 7.5 would support this model. Increased helicity at pH 7.0 of the C-terminus when attached to pentameric scaffold would, however, support the helix bundle model instead. Distinguishing between the C-terminal bundle or free C-terminal helices models at physiological pH (Figure 2A and 2C) will require further investigation.

### **MscL stability**

The structure determination of Tb MscL required tremendous effort before crystals of good quality were obtained. Chang et al. have used the strategy where different homologs of MscL were cloned, overexpressed and purified. The search for crystal yielding conditions was then applied to all homologs, to hopefully identify some that produced crystals. These species were then examined in more detail and optimized in the search for well diffracting crystals. This approach has been successful in producing the structures of several membrane proteins including MscL (1), MsbA (5), and BtuCD (6). After the initial cloning, and even after eliminating the low expressing or misfolded proteins, there could still be many species left for tedious crystallization trials.

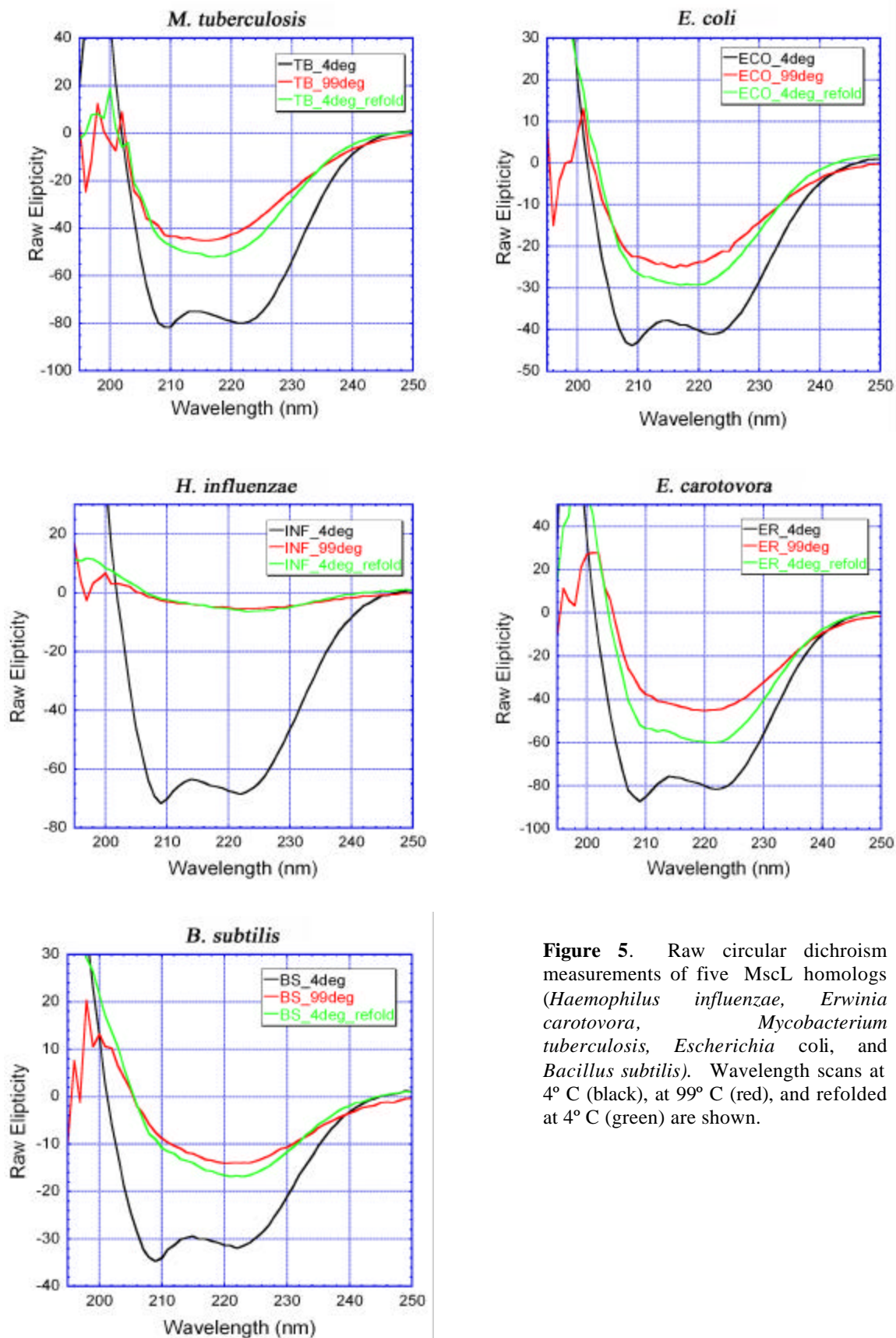


A method that would allow fast identification of species that are more likely to crystallize would save large amounts of tedious work.

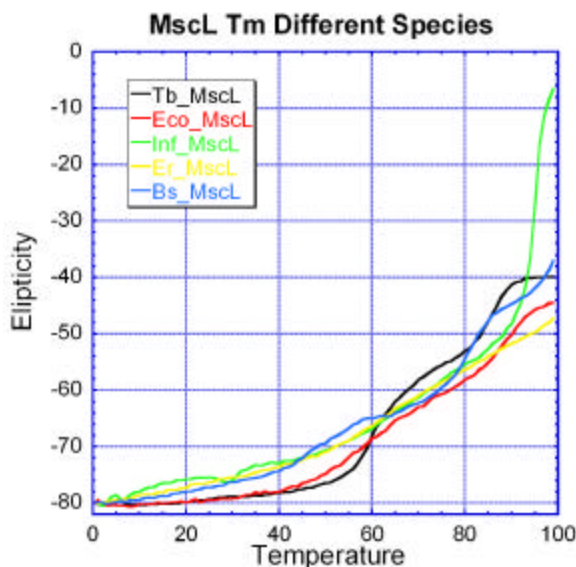
Initially, it was believed that soluble proteins with high thermostability would crystallize easier and form better ordered crystals. For example, the hyperthermophilic *Pyrococcus furiosus* rubredoxin forms crystals that diffract beyond 1.0 Å resolution (7, 8). As more thermophilic proteins and their mesophilic counterparts have been solved, however, this trend seemed to disappear. It became clear, that for water-soluble proteins, hyperthermophilic proteins might crystallize more easily in some cases, but it is not a general trend.

We wanted to test whether more thermostable membrane proteins could crystallize easier or produce crystals of better diffracting quality. Thermostability of membrane proteins was not yet correlated with the ability to crystallize, most likely due to the limiting number of available membrane protein structures. If similar trends seen initially for thermostable soluble proteins prove true for membrane proteins, it could be used as a screening tool to select promising protein species before thousands of crystallization conditions are tried.

Circular dichroism wavelength scans of five MscL homologs are shown in Figure 5. MscL from *Haemophilus influenzae* (Hinf\_MscL), *Erwinia carotovora* (Er\_MscL), *Mycobacterium tuberculosis* (Tb\_MscL), *Escherichia coli* (Eco\_MscL), and *Bacillus subtilis* (Bs\_MscL) show similar, mostly alpha helical CD spectra, with irreversible thermal unfolding. The temperature melts of all species were monitored at 222 nm and are shown in Figure 6.



**Figure 5.** Raw circular dichroism measurements of five MscL homologs (*Haemophilus influenzae*, *Erwinia carotovora*, *Mycobacterium tuberculosis*, *Escherichia coli*, and *Bacillus subtilis*). Wavelength scans at 4° C (black), at 99° C (red), and refolded at 4° C (green) are shown.

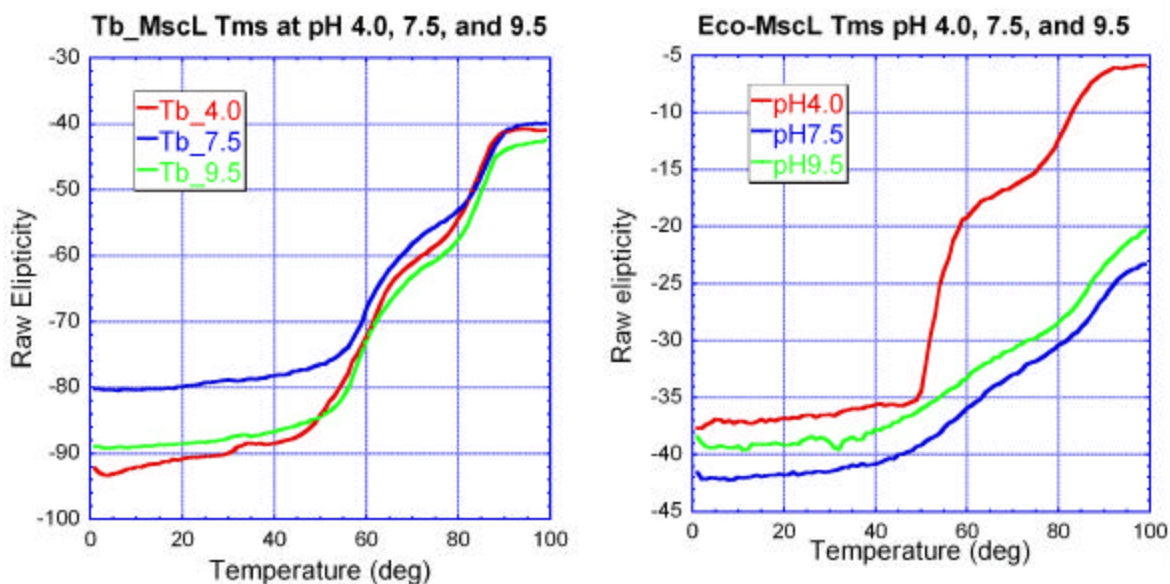


**Figure 6.** Temperature melts monitored by circular dichroism at 222 nm of *Haemophilus influenzae* (green), *Erwinia carotovora* (yellow), *Mycobacterium tuberculosis* (black), *Escherichia coli* (red), and *Bacillus subtilis* (blue) at pH 7.5. The starting ellipticity was adjusted to -80 for all temperature melts.

Tb\_MscL, Eco\_MscL, and Bs\_MscL show two transitions, the first at 65° C, and second at 85° C, while Hinf\_MscL and Er\_MscL show a gradual decrease in ellipticity up to 90° C without any defined transitions. At 90° C, Er\_MscL forms a visible aggregate and falls out of solution which can be seen as large decrease in ellipticity (Figure 6).

The pH dependence of the melting temperature ( $T_m$ ) has been measured for Tb\_MscL and Eco\_MscL (Figure 7). Tb\_MscL shows no pH dependence of  $T_m$ , and the curves at pH 4.0, 7.5, and 9.5 superimpose well. This is in agreement with pH titration data (Figure 3) which show no decrease in helicity between pH 2-12 at 4 °C. At all three pHs (4.0, 7.5, and 9.5), Tb\_MscL exhibits two transitions at 65 °C and 85 °C. Eco\_MscL shows similar trends at pH 7.5 and 9.5, however, at pH 4.0 Eco\_MscL aggregates at ~50 °C (Figure 7). Eco\_MscL also shows a biphasic thermal transition with midpoints at 65 °C and 85 °C, but with less defined inflections than seen for Tb\_MscL.

Tb\_MscL, which produced the best diffracting crystals, is more stable and does not aggregate at low pH, which was used for crystallization. At physiological pH, the



**Figure 7.** Temperature melts monitored by circular dichroism of Tb\_MscL and Eco\_MscL at pH 4.0 (red), 7.5 (blue), and 9.5 (green).

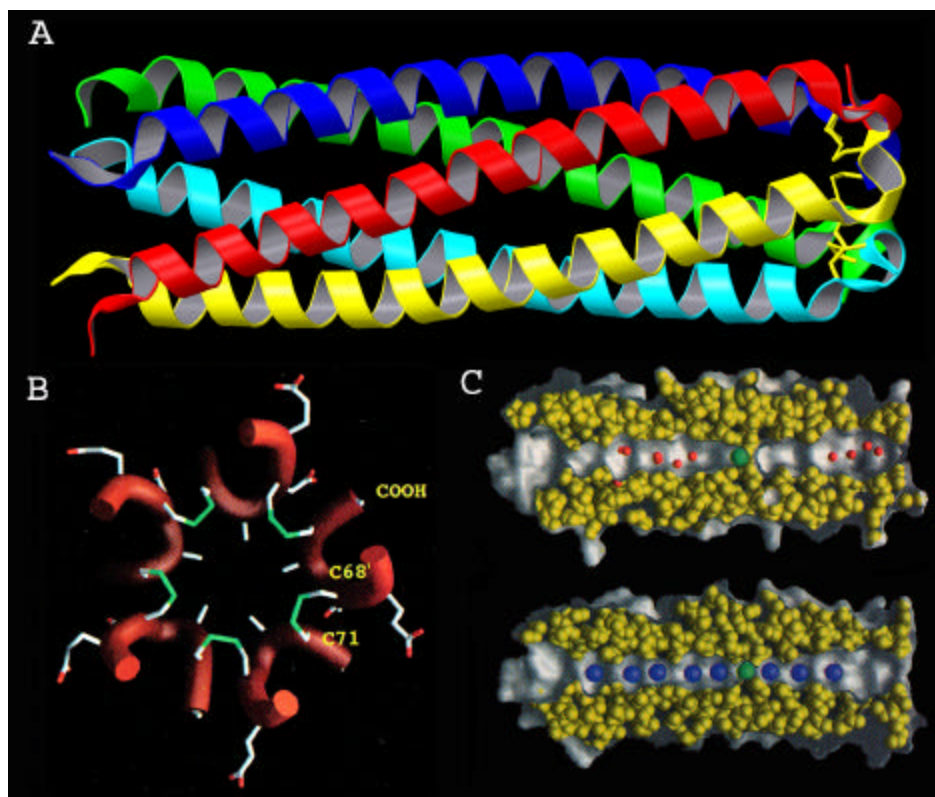
temperature melts of the five MscL homologs generally look similar, and have nearly identical transitions. Apart for the low pH stability of MscL, no clear correlation between melting temperature and ease of crystallization or quality of crystals was observed.

### Higher resolution structure of Tb-MscL

*Termination mutants:* The structure of Tb MscL contains 165 disordered residues (33 per monomer) at the C-terminus following the cytoplasmic helix (Figure 9). These disordered residues at the C-terminus might disrupt the crystal formation and cause disorder in the crystal, resulting in poor diffraction. A truncation mutant that lacks this disordered region could produce better diffracting crystals. Limited proteolysis (X. Ambroggio, personal communication) of Tb\_MscL has identified two C-terminal truncations at positions 100 and 119. Position 100 corresponds to truncation in the loop

between the TM2 helix and the cytoplasmic helix. Position 119 corresponds to truncations right after the cytoplasmic helix. The absence of the C-terminus of MscL causes an inability of the truncated protein to bind to ion exchange resins. Ion exchange columns are used in the purification (due to their high binding capacity) for concentration of the protein without concentrating the detergent. To determine the minimum and maximum pH under which MscL is still stable and can be used for ion exchange chromatography, a pH profile of MscL was performed (Figure 3). As noted previously, Tb\_MscL was found to be extremely stable with no change in circular dichroism spectra between pH 2-12. Even under extreme pHs (pH 11.0 for source Q resin, and pH 2.0 for source S resin), the T119 mutant did not show sufficient binding to the ion exchange resin. Expression levels lower than wild type MscL, together with difficulty in binding to an ion exchange column, has made these mutants impractical for crystallization studies.

*COMP fusion:* Cartilage oligomeric protein is a pentameric glycoprotein found in cartilage and tendon. It belongs to the thrombospondin family and consists of an N-terminal heptad repeat (COMPcc), four epidermal growth factor (EFG) like domains, seven calcium binding domains and a C-terminal globular domain (9-12). The function of COMP is not clear, however. Mutations in the COMP gene are linked to pseudoachondroplasia and increased expression of COMP provides increased strength in muscle and tendon. The structure of the N-terminal domain was solved at 2.0 Å resolution (10) and revealed a homo-pentameric parallel coiled-coil helix bundle (Figure 8A). The fragment that was crystallized is composed of 46 residues per helix which form the only known parallel pentameric bundle with pore of diameter of 2-6 Å (Figure 8C). At the constriction of the channel, a chloride anion was found coordinated by five



**Figure 8** COMPcc structure. (A) Ribbon diagram of the COMPcc structure, showing the parallel pentameric bundle with disulfide bridges at the C-terminus. (B) Structure of the disulfide bridges viewed from the N-terminus. The disulfide bonds are formed between residues 68 and 71 from neighboring subunits. (C) Cross section of the pore of COMPcc. Water molecules are shown as red spheres, chloride ion as green sphere, and xenon atoms in the xenon soak as blue spheres. Adapted from Malashkevich et al. (1996), *Science*, 274, 761-765.

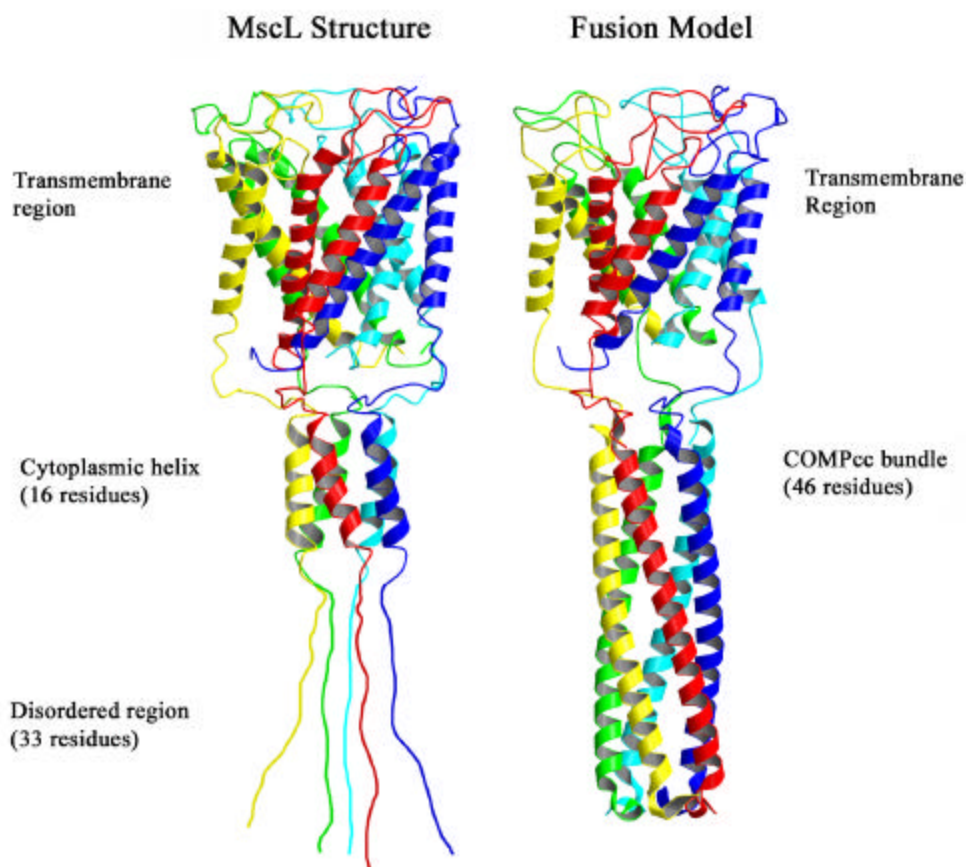
glutamine residues (Figure 8C). The C-terminus of the COMPcc contains two cysteine residues which form disulfide bonds between neighboring subunits, thus covalently crosslinking the bundle together (Figure 8B). The oligomerization of COMPcc is independent of disulfide bond formation. The formation of the disulfide bonds increases the melting temperature of the pentameric bundle from 48 °C to over 100 °C. Several hydrophobic compounds (benzene, cyclohexane, vitamin D<sub>3</sub>, elaidic acid) were shown to bind in the channel and increase the melting temperature by 2-10 °C (13). The melting temperature of oxidized COMPcc in the presence of 4.5 M GdmCl is 60 °C. The parallel

coiled-coil pentameric structure, extremely high thermostability, and the presence of cysteines that can be crosslinked to form covalent pentamers make COMPcc a perfect fusion protein to improve the diffraction of Tb\_MscL crystals, by attachment to the cytoplasmic domain.

The sequence alignment of the C-terminal helix of MscL with COMPcc is shown in Figure 10. The sequences were aligned by a proline residue and the two structures were superimposed with LSQMAN (14). Superposition of the Ca atoms of 60 residues (12 per monomer) resulted in r.m.s.d of 2.0 Å. The MscL\_COMPcc fusion construct consists of residues 1-106 from Tb\_MscL and of residues 32-72 from COMPcc and is only three amino acids shorter than wild type Tb\_MscL. A model of the MscL\_COMPcc fusion protein is shown in Figure 9.

Formation of the disulfide bonds that crosslink the five monomers together happens spontaneously in approximately one week, or can be also achieved by addition of copper phenanthroline. Fusion proteins that did not form covalently linked pentamers contain free cysteines that can be biotinylated and removed on an avidin column.

Previous protein fusion strategies have not been very successful in improving the crystallization properties of proteins. This is mainly due to the fact that to bring the C-terminus of one protein near the N-terminus of another protein, a flexible linker is generally necessary. In this case, since the C-terminal helix bundle is replaced with another helix bundle with very similar helix parameters, no flexible linker is necessary. The MscL\_COMPcc fusion could stabilize the closed state of MscL by forming a stable covalent pentameric bundle at the C-terminus of the channel, which would eliminate 165 disordered residues that are not seen in the crystal structure, and could provide a larger



**Figure 9.** Model of the MscL\_COMpcc fusion. The pentameric wild type MscL structure is composed of the transmembrane domain, cytoplasmic domain and 165 disordered residues at the C-terminus. MscL\_COMpcc fusion would eliminate the disordered region, could potentially stabilize the closed state, and provide larger packing interface for crystal formation.

packing interface for crystal formation. Crystallizations attempts of MscL\_COMpcc fusion are in progress.

## Materials and Methods

*Tb\_MscL purification:* Purifications were done as described previously (1). Cells (BL21DE3 gold, Stratagene) were grown in a 60 L Brunswick Bioflow 5000 fermentor in Terrific Broth media (15) supplemented with 10 g/L of glucose to an OD of 10.0 in the presence of 50  $\mu$ g/ml ampicillin. Induction was done with 20 g/L of lactose and 2 mM IPTG for 2 hours. Cells were then harvested (Pelicon tangential flow ultrafilter,



Millipore) and centrifuged. An average yield of ~800g of wet cell mass was obtained from a 50 L ferment. Cells were frozen and stored at -80 °C. Resuspended cells (100-200g in 150 mM NaCl, 50 mM Tris pH 8.0, PMSF, Roche “complete” protease inhibitors, 50 µg/ml lysozyme, and 10 µg/ml DNaseI) were thawed at room temperature. Two passes through an Emulsiflex French press at a pressure of 15,000 psi - 20,000 psi resulted in complete cell lysis. Tb\_MscL was then extracted into 1 % detergent (either DM or DDM, Anatrace) for 4 hours at 4 °C. After extraction, cell debris was removed by centrifugation at 20,000 g for 30 minutes. The supernatant was supplemented with 30 mM imidazole and loaded onto a 150 ml Ni superflow (Qiagen) column at 5 ml/minute. The column was then washed with 0.5 L of equilibration buffer (30 mM imidazole, 500 mM NaCl, Tris pH 8.0, 0.05 % DDM or 0.2 % DM), and with 0.5 L of wash buffer (100 mM imidazole, 0.05 % DDM or 0.2% DM). Elution was achieved with 300 mM imidazole and 0.05 % DDM or 0.2 % DM. The eluted fractions were pooled and diluted 1:10 with dilution buffer (25 mM Tris pH 8.0, 0.05% DDM and 0.2 % DM). The diluted protein was passed over a 15 mL Source Q (Pharmacia) column (~7.5 ml/min) and washed with buffer A (25 mM Tris pH 8.0, 0.05% DDM or 0.2 % DM, 10 % glycerol) and eluted by buffer B (buffer A with 1.0 M NaCl). The elution was done by slowing the flow to 1.0 ml/minute, 100 % Buffer B and inverting the flow. Protein concentrated on the Source Q column was then passed over a Superdex 200 (Pharmacia) prep sizing column in buffer C (50 mM Tris pH 8.0, 150 mM NaCl, 10 % glycerol, 0.05 % DDM or 0.2 % DM) overnight. Fractions corresponding to a Tb\_MscL pentamer were pooled and diluted 1:3 with dilution buffer and concentrated on the source Q column as described

above. The concentrated fractions were then desalted either by dialysis or on desalting column. All buffers included 0.1 mM PMSF and Roche “complete” protease inhibitors.

*pH profiling:* For pH profiling, the protein concentration in all of the samples had to be exactly the same. Dialysis or regular pipeting introduced too much error into the experiments. Instead, 2.5  $\mu$ L were measured out with Hamilton 10  $\mu$ L syringe and mixed with 250  $\mu$ L of a given buffer (measured with a 300  $\mu$ L Hamilton syringe). Phosphate buffer was used for pH 2.0-4.0, 6.0-7.0, and 10.0-12.0. Tris buffer was used for pH 8.0-10.0 and citrate buffer was used for pH 4.0-6.0.

*Circular dichroism:* CD data were collected on an Aviv 62DS spectrometer equipped with a thermoelectric unit using a 1 mm path length cell. Thermal melts were monitored at 222 nm. Data were collected every 1  $^{\circ}$ C with an equilibration time of 2 min and an averaging time of 30 s. Three wavelength scans were performed at 4  $^{\circ}$ C collecting the data every 1 nm (200-250 nm) with an averaging time of 1 s, and averaged together. C-terminal peptides were obtained from peptide facility and verified by mass spectrometry.

*C-terminal truncations:* The two truncations were made with inverse PCR (16) and expressed in a BL21DE3 host. Both truncation mutations showed much better expression at room temperature than at 37  $^{\circ}$ C. The expression levels for T100 and T119 were lower than seen for wild type MscL, which shows optimal induction at 37  $^{\circ}$ C. T119 was expressed in a 60L Brunswick Bioflow 5000 fermentor in Terrific Broth media (15) supplemented with 10 g/L glucose and induced at 25  $^{\circ}$ C with 2 mM IPTG and 20 g/L lactose. Purification was done as described above.

*COMPcc fusion:* COMPcc DNA was made by recursive PCR (17) from two oligonucleotides (105 bp), which included 30 basepair clamps necessary for integration at a specific site within the MscL vector without the use of restriction enzymes or DNA ligase (18). The designed oligonucleotides also included a stop codon, so no further mutations were necessary. The purification of COMPcc was done as for Tb\_MscL, which is described above.

### **Acknowledgements**

We thank R. Bass and K. Locher for many useful discussions and ideas, X. Ambroggio for help with the making of the MscL\_COMPcc fusion, R. Spencer for the wild type Tb\_MscL gene and purified MscL homologs used in this study, and Suzanna Horwath for peptide synthesis.

## References

1. Chang, G., Spencer, R. H., Lee, A. T., Barclay, M. T., and Rees, D. C. (1998) *Science* 282, 2220-6.
2. Hase, C. C., Le Dain, A. C., and Martinac, B. (1997) *J Membr Biol* 157, 17-25.
3. Ajouz, B., Berrier, C., Garrigues, A., Besnard, M., and Ghazi, A. (1998) *J Biol Chem* 273, 26670-4.
4. Blount, P., Sukharev, S. I., Schroeder, M. J., Nagle, S. K., and Kung, C. (1996) *Proc Natl Acad Sci U S A* 93, 11652-7.
5. Chang, G., and Roth, C. B. (2001) *Science* 293, 1793-800.
6. Locher, K., Lee, A. T., and Rees, D. C. (2002) *Science*.
7. Day, M. W., Hsu, B. T., Joshua-Tor, L., Park, J. B., Zhou, Z. H., Adams, M. W., and Rees, D. C. (1992) *Protein Sci* 1, 1494-507.
8. Bau, R., Rees, D. C., Kurtz, D. M., Scott, R. A., Huang, H., Adams, M. W., and Eidsness, M. K. (1998) *J Biol Inorg Chem* 3, 484.
9. Guo, Y., Kammerer, R. A., and Engel, J. (2000) *Biophys Chem* 85, 179-86.
10. Malashkevich, V. N., Kammerer, R. A., Efimov, V. P., Schulthess, T., and Engel, J. (1996) *Science* 274, 761-5.
11. Efimov, V. P., Lustig, A., and Engel, J. (1994) *FEBS Lett* 341, 54-8.
12. Efimov, V. P., Engel, J., and Malashkevich, V. N. (1996) *Proteins* 24, 259-62.
13. Guo, Y., Bozic, D., Malashkevich, V. N., Kammerer, R. A., Schulthess, T., and Engel, J. (1998) *Embo J* 17, 5265-72.
14. Kleywegt, G. J. (1996) *Acta. Cryst. D* 52 D52, 842-857.
15. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular cloning: a laboratory manual 2nd ed.*, Cold Spring Harbor Laboratory Press.
16. Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G., and Galas, D. J. (1989) *Nucleic Acids Res* 17, 6545-51.
17. Prodromou, C., and Pearl, L. H. (1992) *Protein Eng* 5, 827-9.
18. Geiser, M., Cebe, R., Drewello, D., and Schmitz, R. (2001) *Biotechniques* 31, 88-90, 92.

**Section II**

**Chapter 4:**

**Crystal structure of the 'cab' type beta class carbonic  
anhydrase from the archaeon *Methanobacterium*  
*thermoautotrophicum***

Reprinted from J. Bio. Chem.

**Crystal structure of the 'cab' type beta class carbonic anhydrase from  
the archaeon *Methanobacterium thermoautotrophicum*<sup>†</sup>**

Pavel Strop<sup>‡</sup>, Kerry S. Smith<sup>§</sup>, Tina M. Iverson<sup>¶</sup>, James G. Ferry<sup>§,@</sup>, and Douglas C.

Rees<sup>¶,\*</sup>

<sup>†</sup>This work was supported by grants from the National Institute of Health (J.G.F and D.C.R.), and an NSF predoctoral fellowship (P.S.)

<sup>‡</sup>Biochemistry Option, California Institute of Technology, Mail Code 147-75, Pasadena, California, 91125, USA.

<sup>§</sup>Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, 16802, USA.

<sup>¶</sup>Howard Hughes Medical Institute and Division of Chemistry and Chemical Engineering, California Institute of Technology, Mail Code 147-75CH, Pasadena, California, 91125, USA.

<sup>@</sup>Center for Microbial Structural Biology, The Pennsylvania State University, University Park, PA, 16802, USA

**Abstract**

The structure of the 'cab' type  $\beta$  class carbonic anhydrase from the archaeon *Methanobacterium thermoautotrophicum* (Cab) has been determined to 2.1 Å resolution using the multiwavelength anomalous diffraction (MAD) phasing technique. Cab exists as a dimer with a subunit fold similar to that observed in 'plant' type  $\beta$  class carbonic anhydrases. The active site zinc is coordinated by protein ligands Cys32, His87, and Cys90, with the tetrahedral coordination completed by a water molecule. The major difference between 'plant' and 'cab' type  $\beta$  class carbonic anhydrases is in the organization of the hydrophobic pocket. The structure reveals a Heps buffer molecule bound 8 Å away from the active site zinc, which suggests a possible proton transfer pathway from the active site to the solvent.

Carbonic anhydrases (CAs) are  $\text{Zn}^{2+}$  containing enzymes that catalyze the reversible hydration of  $\text{CO}_2$  (1). With turnover numbers approaching  $10^6 \text{ s}^{-1}$ , CAs are among the fastest known enzymes. CAs have been found in most types of organisms, including mammals, plants, algae, bacteria, and archaea (2). Based on the amino acid sequences, CAs can be assigned to one of three independently evolved classes, designated  $\alpha$ ,  $\beta$  and  $\gamma$  (3).

The  $\alpha$  class contains all mammalian CAs, as well as some CAs from algae and bacteria (3,4).  $\alpha$ -CAs play important roles in respiration, secretion of  $\text{HCO}_3^-$ , pH homeostasis, and ion exchange (5,6). Crystal structures of  $\alpha$ -CA have revealed a monomer organized around a 10-stranded, predominantly anti-parallel  $\beta$ -sheet (7-13). The catalytically active zinc is coordinated by three histidines and one water molecule.

$\gamma$ -CA has thus far been isolated and characterized only from the methanoarchaeon *Methanosarcina thermophila* (14-16), where it is proposed to facilitate the transport of  $\text{CH}_3\text{COO}^-$ , and to convert  $\text{CO}_2$  to  $\text{HCO}_3^-$  outside the cell to assist the removal of excess  $\text{CO}_2$  generated during the growth of this organism on acetate. The *M. thermophila*  $\gamma$ -CA exists as a trimer with the active site located at the interface between two subunits. Each subunit is organized around a left handed  $\beta$ -helix that is completely distinct from the  $\alpha$ -CA fold, although the active site is also coordinated by three histidines, along with two water molecules (17,18).

The  $\beta$  class includes CAs from plants, algae, bacteria, and archaea (2,3). In higher plants,  $\beta$ -CAs play an important role in photosynthesis, by concentrating  $\text{CO}_2$  in the proximity of ribulose biphosphate carboxylase/oxygenase (RuBisCO) for  $\text{CO}_2$

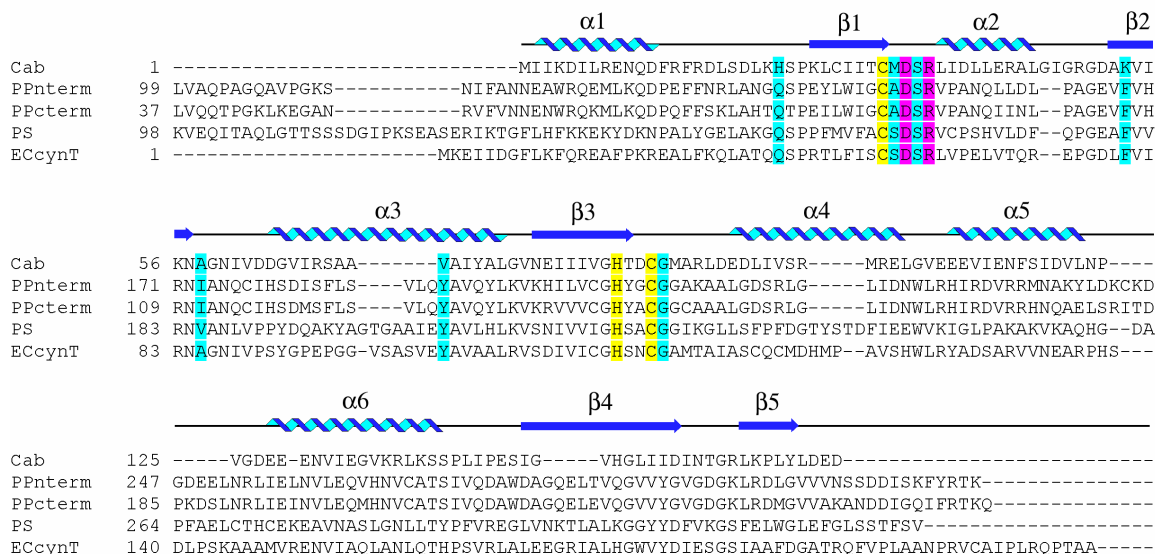


fixation (19). The purification and characterization of Cab from the thermophilic *Methanobacterium thermoautotrophicum* extends this class into the archaea (20). Cab is at the phylogenetic extreme of the beta carbonic anhydrases and forms an exclusively prokaryotic clade consisting primarily of sequences from Gram-positive bacteria (2). In the obligate chemoautolithotroph *M. thermoautotrophicum* Cab converts CO<sub>2</sub> to HCO<sub>3</sub><sup>-</sup>, suggesting the physiological role of this enzyme may be to provide HCO<sub>3</sub><sup>-</sup> to enzymes important in CO<sub>2</sub>-fixation pathways of the microbe (21,22).

β class carbonic anhydrases can be further divided into 'plant' type and 'cab' type, based on the active site sequence conservation (20,23) (Fig. 1). Two crystal structures of 'plant' type β-CA were recently reported from *Porphyridium purpureum* (*P.p.* β-CA) (24) and *Pisum sativum* (*P.s.* β-CA) (23). The basic fold of β-CA consists of a four-stranded, parallel β-sheet core with α-helices forming right handed crossover connections (23,24). The oligomerization state of β-CA is variable, however and *P.p.* β-CA and *P.s.* β-CA exist as a dimer and octamer, respectively, although the dimer of *P.p.* β-CA resembles a tetramer, where two monomers are fused together. In contrast to the protein ligation by three histidines observed in α- and γ-CAs, the active site zinc in β-CAs is coordinated by two cysteines and one histidine, as anticipated from EXAFS studies (21,25,26). The fourth ligand is different in the two β-CAs structures. In the *P.s.* β-CA structure, an acetate molecule is bound to the zinc, whereas in the *P.p.* β-CA structure, the side chain of aspartic acid (Asp151) acts as the fourth ligand. In the *P.s.* β-CA structure, this conserved Asp interacts with a conserved Arg (Fig. 1).

Apart from the conserved zinc ligands and the Asp/Arg pair, the active site of Cab differs significantly from the 'plant' type β-CAs. Cab active site residues His23, Met33,

Fig. 1.



**Fig. 1.** Alignment of  $\beta$ -CA sequences. Abbreviations: Cab (*M. thermoautotrophicum*), *P.p.* Nterm (*P. purpureum* N-terminal domain), *P.p.* Cterm (*P. purpureum* C-terminal domain), *P.s.* (*P. sativum*), ECcynT (*E. coli*). Zinc ligands are colored yellow, the conserved Asp/Arg pair red, and residues differentiating 'cab' type and 'plant' type are colored blue.

Lys53, Ala58 and Val72, are replaced in the 'plant' type  $\beta$ -CA by Gln, Ala, Phe, Val, and Tyr, respectively. Residues that are equivalent to Cab residues Met33, Lys53, Ala58, and Val72 make up the hydrophobic pocket in the *P.s.* and *P.p.*  $\beta$ -CA structures. Substitutions of the two aromatic side chains by Lys and Val in Cab suggest a significant redesign of the hydrophobic pocket in 'cab' type  $\beta$ -carbonic anhydrase. Cab has a  $\text{CO}_2$  hydration activity with a  $k_{\text{cat}}$  of  $1.7 \times 10^4 \text{ s}^{-1}$  and  $K_M$  for  $\text{CO}_2$  of 2.9 mM at pH 8.5 and 25° C (20). Cab is inhibited by iodide, nitrate, and azide; however, in contrast to 'plant' type  $\beta$ -CAs, chloride and sulfate have no effect on Cab activity. These active site substitutions, together with the different effects of inhibitors, imply that there might be mechanistically relevant differences in the organization of the active sites between 'cab'

type and 'plant' type enzymes. Here we present the first structure of the 'cab' type  $\beta$ -carbonic anhydrase from thermophilic methanoarchaeon *Methanobacterium thermoautotrophicum* (Cab), determined at 2.1 Å resolution.

### Experimental Procedures

*Crystallization:* Cab was overexpressed in *E. coli* and purified by a heat denaturation step followed by ion exchange chromatography as previously described (20). Crystals were grown by the hanging drop method at 22 °C using a 5 mg/ml protein solution and a precipitant solution containing 100 mM Hepes, pH 7.5, 35% ethanol, 12% methyl pentanediol (MPD) and 50 mM calcium acetate. The crystals belong to the orthorhombic space group  $P2_12_12_1$  with unit cell dimensions  $a = 54.9 \text{ \AA}$ ,  $b = 113.2 \text{ \AA}$ ,  $c = 156.2 \text{ \AA}$  and three dimers per asymmetric unit. Crystals were transferred stepwise to mother liquor solution containing 30% MPD as a cryoprotectant, and flash frozen.

*Data collection and processing:* A three wavelength MAD (multiwavelength anomalous dispersion) data set was collected at  $-160 \text{ }^\circ\text{C}$  on beamline 9-2 of the Stanford Synchrotron Research Laboratory (SSRL) with an ADSC charge-coupled device detector. The fluorescence spectrum measured around the Zn edge of a single crystal was used to select the inflection point ( $\lambda = 1.2832 \text{ \AA}$ ), the absorption edge ( $\lambda = 1.282 \text{ \AA}$ ) and a high energy remote wavelength ( $\lambda = 1.033 \text{ \AA}$ ) for optimization of the anomalous signal. All data were reduced using DENZO and scaled using SCALEPACK (27) (Table I).

*Phase determination:* The structure was determined by MAD using the signal from only the intrinsic zinc atoms. The program SOLVE (28) was used to find the positions of the heavy atoms using the three wavelength MAD data set using data from

Table I. Data Collection Statistics

Data collection	peak	inflection	remote
	?1	?2	?3
Wavelength (Å)	1.282	1.2832	1.0333
Resolution range (Å)	20-2.3	20-2.4	20-2.1
Unique observations	43450	38663	57495
Total observations	156889	139023	208917
Completeness (%)	97.9/92.0	98.5/96.5	98.8/98.2
R <sub>sym</sub>	5.9(21.8)	5.1(16.7)	6.3(24.3)
<I/s>	24 (7.0)	25 (8.6)	21 (5.7)
R <sub>cullis</sub> (cen/acen/anom)	0.79/0.78/0.86	0.74/0.76/0.91	--/--/0.92
Phasing power (cen/acen)	0.81/1.26	0.94/1.38	
Figure of Merit to 2.4 Å resolution (cen/acen)	0.52/0.57		

Numbers in parentheses indicate values for highest resolution bin (2.38-2.30 Å for ?1, 2.49-2.4 Å for ?2, and 2.17-2.10 Å for ?3)

$$R_{\text{sym}} = \sum |I_{\text{obs}} - I_{\text{avg}}| / \sum I_{\text{obs}}$$

R<sub>cullis</sub>=lack of closure error/iso-ano difference

Phasing power=heavy atom structure factor/ r.m.s. lack of closure error

20 to 2.4 Å resolution. Two zinc ions were identified and when used for phasing yielded a figure of merit (FOM) of 0.44. Four additional Zn sites were located in an anomalous difference Fourier map, yielding a FOM of 0.54. Each of the six zinc sites corresponded to a Cab monomer, which are arranged into three tight dimers in the asymmetric unit. The initial non-crystallographic symmetry (NCS) transformations were established by the relationships between the dimers, and the initial mask was calculated with program NCSMSK (29). The program DM (30) was used for NCS averaging of the electron

density maps, solvent flattening, and phase extension from 2.4 Å to 2.1 Å resolution. The resulting map was of good quality and allowed building of most of the protein.

---

Table II. Refinement Statistics

---

Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimensions	
a (Å)	54.67
b (Å)	113.21
c (Å)	156.18
Resolution (Å)	20-2.1
R <sub>cryst</sub> (%)	21.1
R <sub>free</sub> (%)	24.6
R.m.s.d deviations	
bond lengths (Å)	0.014
bond angles (°)	1.70
torsion angles (°)	22.9
improper torsion angles (°)	1.39

---

$$R_{\text{cryst}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum |F_{\text{obs}}|$$

R<sub>free</sub> = R<sub>cryst</sub> calculated for 5% of reflections omitted from the refinement

---

*Refinement:* Alternate cycles of manual model building using the program O (31) and positional and individual B-factor refinement with program CNS (32) reduced the R and R<sub>free</sub> to 21.1 % and 24.6 %, respectively, where R<sub>free</sub> is calculated for 5% (2875) of the reflections in the resolution range 18-2.1 Å. The model was initially refined with strict NCS restraints, which were released later in the refinement. The rmsd (root mean square deviation) of bond lengths and angles are 0.013 Å and 1.7 °, respectively, with 87.9 % in the most allowed region, and 11.1 % in the additionally allowed region of the

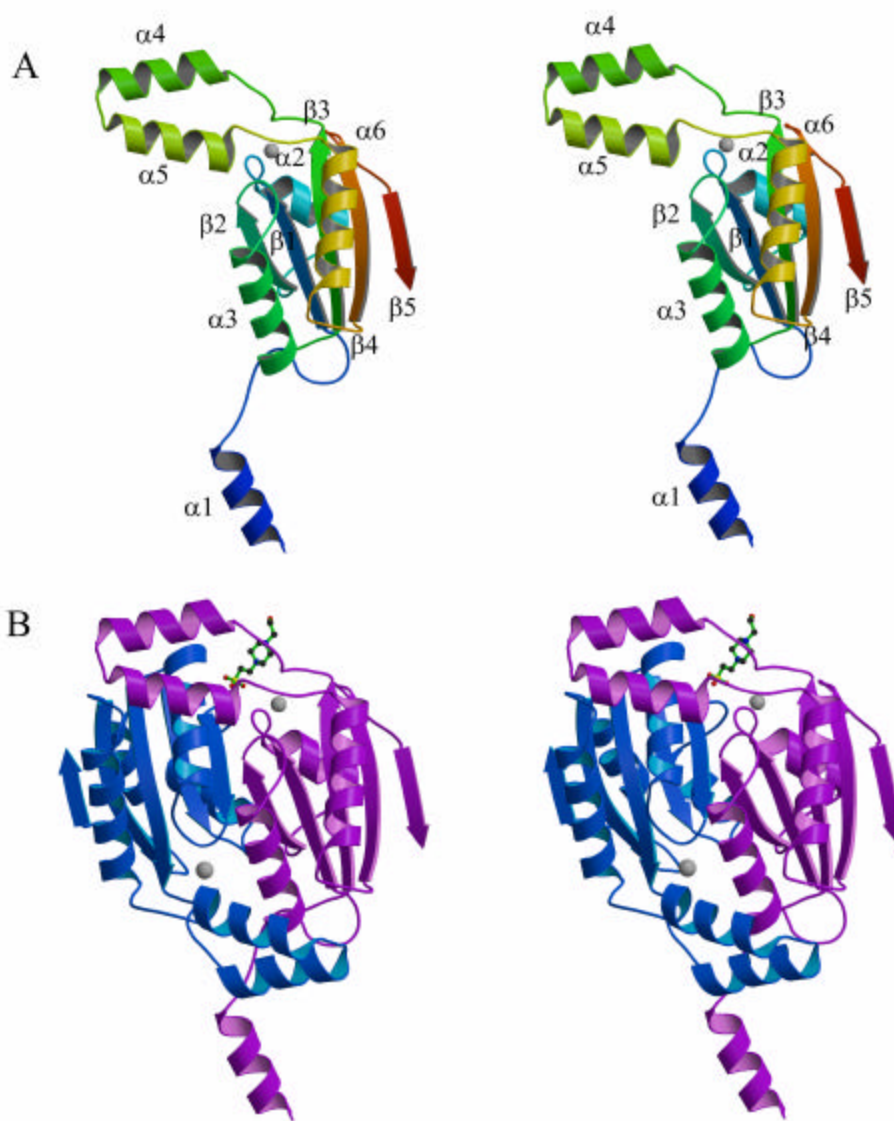
Ramachandran plot. The average B-factors are  $28.6 \text{ \AA}^2$  (main chain),  $32.4 \text{ \AA}^2$  (side chains),  $18.6 \text{ \AA}^2$  (zinc atoms), and  $36.3 \text{ \AA}^2$  (solvent). An average B-factor of  $30.5 \text{ \AA}^2$  is calculated for all protein atoms. The final model contains 7543 protein atoms, 3 Hepes molecules, 6 zinc atoms, 6 calcium atoms, and 409 water molecules, for a total of 8015 atoms (Table II).

## Results and Discussion

*Structural organization of Cab:* The overall fold of the Cab monomer (Fig. 2A) consists of a four-stranded parallel  $\beta$ -sheet core with strand order 2-1-3-4. Monomers in each dimer are related by a two-fold axis centered between strands  $\beta 2$ . The overall dimensions of the dimer are approximately  $40 \text{ \AA} \times 45 \text{ \AA} \times 50 \text{ \AA}$  (Fig. 2B). Three dimers, designated AB, CD, and EF, are present in the asymmetric unit. In the structurally conserved  $\beta$ -sheet region (residues 26-32, 52-57, 80-88, 149-157, and 163-167), the rmsds in  $C_\alpha$  positions between the dimers average  $0.24 \text{ \AA}$ . Using  $C_\alpha$ 's for residues 24-170, the corresponding rmsds are:  $0.34 \text{ \AA}$  between dimers AB and CD,  $1.0 \text{ \AA}$  between dimers AB and EF, and  $1.1 \text{ \AA}$  between dimers CD and EF. The relatively large rmsds for the latter two pairs of dimers reflect the poor ordering for residues 91-126 in monomer E. Excluding these residues, the rmsds drop to  $\sim 0.28 \text{ \AA}$ . The rmsd in the  $\beta$ -sheet region between the A and B, C and D, and E and F subunits of the Cab are approximately  $0.6 \text{ \AA}$ , while the rmsd for all  $C_\alpha$  between the subunits are approximately  $0.8 \text{ \AA}$ .

The regions with the largest conformational variation include the N-terminal residues 1-23 (involved in crystal packing), residues 92-95, and residues 120-125. While the N-terminal residues 1-23 are well defined in monomers A, C, and E, residues 13-23

Fig. 2.



**Fig. 2.** Structure of Cab. **A)** Ribbon diagram of the A monomer of Cab. The color changes from blue (N-terminus) to red (C-terminus). **B)** Dimer of Cab. Monomer A is shown in magenta, monomer B is shown in blue, Zn atoms are shown in grey. Monomer A is shown in the same orientation for both figures 2A and 2B. Hemes bound to subunit A is shown in ball-and-stick representation. The figures were created with BOBSCRIPT (45) and RASTER3D (46).

are disordered in monomers B, D and F. Residues 92-95 and 120-125 form a hinge region for helices  $\alpha 4$  and  $\alpha 5$  and are well ordered in monomers B and D, slightly

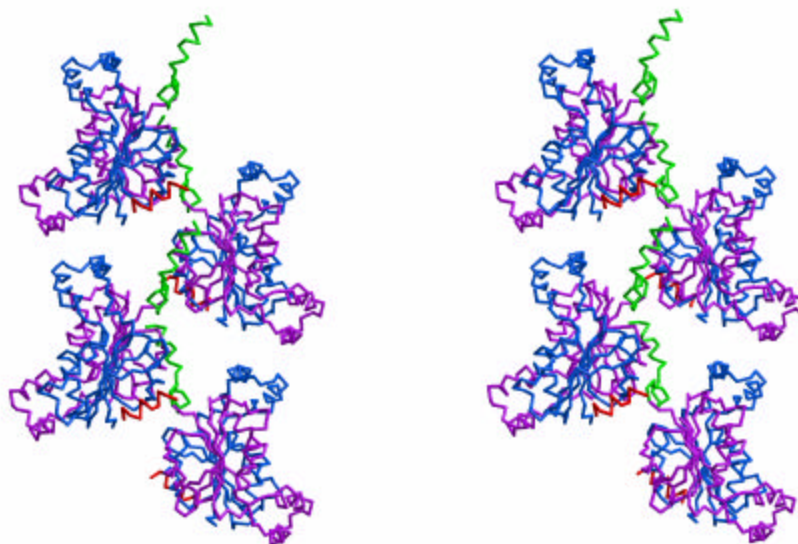
disordered in monomers A, C, and F and disordered in monomer E. The variability in region 90-125 between the six crystallographically independent monomers suggests that these residues are conformationally mobile. Overall, with the exception of the noted regions, the three dimers in the asymmetric unit are very similar. Unless stated otherwise, only dimer AB will be used in the future discussions.

Six calcium ions were located on the surface of Cab and at the crystal packing interfaces between Cab dimers. Most likely, these calcium ions do not play any structural or catalytical role and are a result of the crystallization conditions that contained 50 mM calcium acetate.

*Cab oligomerization:*  $\beta$ -CAs have been found in different oligomeric states ranging from dimers (*O. sativa*, *P. purpureum*), tetramers (*E. coli*), to octamers (*P. sativum*) (20, 23, 24, 33). Although analytical ultracentrifugation results suggest that Cab exists as a homotetramer (21), Cab appears to form a dimer in the crystal. There are numerous interactions stabilizing the dimer involving residues in  $\beta$ -strand  $\beta 2$ , and helices  $\alpha 2$ ,  $\alpha 3$ ,  $\alpha 4$ , and  $\alpha 5$ . Hydrogen bonds between residues 56 and 57 from strand  $\beta 2$  in both subunits result in a formation of a 10-stranded, mainly parallel,  $\beta$  sheet. Helices  $\alpha 4$  and  $\alpha 5$  extend out and make extensive contact with the other monomer in the dimer (Fig. 2B). The interface area between these two subunits was found to be  $\sim 2110 \text{ \AA}^2$  per subunit, which is  $\sim 21\%$  of the total ( $\sim 10000 \text{ \AA}^2$ ) monomer accessible surface area as calculated with GRASP (34). Residues A2-A23 pack against a symmetry related molecule burying  $860 \text{ \AA}^2$ , resulting in the formation of a continuous ribbon through the crystal (Fig. 3). Residues B2-B12 most likely fold back and pack against the B monomer in the same way that A2-A12 packs against the symmetry related molecule. Residues B13-B23 have no



Fig. 3.



**Fig. 3.** Stereo view of the crystal packing of Cab showing the continuous ribbon created by packing of helix  $\alpha 1$  against a crystallographically related molecule. Monomer A is in magenta, with helix  $\alpha 1$  in green. Monomer B is in blue, with helix  $\alpha 1$  in red.

visible electron density and are disordered. A similar type of crystal packing forming a continuous ribbon has been observed in the SAM (sterile alpha motif) domain of the human EphB2 receptor, and together with other evidence was proposed to be functionally important (35). The crystal packing interaction seen in Cab can also be described as a linear (open-ended) domain swapped oligomer, where the swapped domain consists of a twelve residue  $\alpha$ -helix (36,37). It is unclear whether residues 1-24 could facilitate the formation of a higher oligomerization state under physiological conditions, or whether residues A2-A23 also fold back and pack against monomer A. Based on modeling considerations, even though the N-terminal region is highly flexible, formation of a Cab tetramer similar to the one seen in the *P.p.* structure is unlikely due to the steric clashes of

helices  $\alpha 4$  and  $\alpha 5$ . Contacts between Cab dimers in the crystal are relatively small (= 540 Å<sup>2</sup>) and, based on their complementarity and size (38,39), are also unlikely to support formation of stable tetramers.

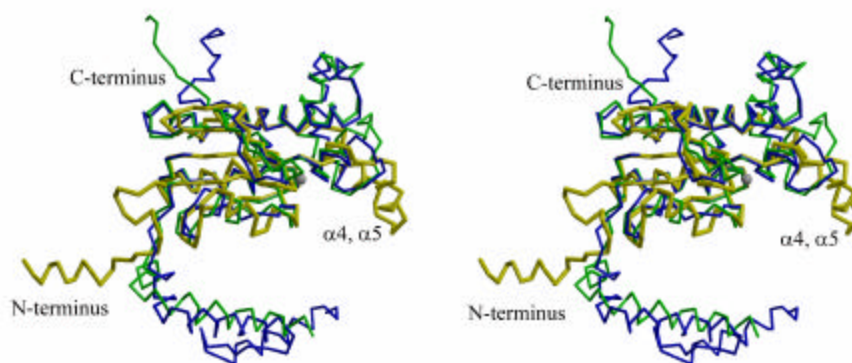
*Comparison of 'cab' and P.s. and P.p. structures:* The fold of Cab is similar to that of the *P.s.* and *P.p.*  $\beta$ -CAs (Fig. 4). While the  $\beta$ -sheet core is conserved, significant secondary structure differences are evident, mostly in the regions at the N-terminus, C-terminus and in the region containing residues 90-125. The N-termini of the *P.s.* and *P.p.*  $\beta$ -CA structures extend out, forming a long helix that packs against a second monomer, making additional dimer interactions. In Cab, the N-terminus is involved in crystal packing and does not adopt the same conformation observed in the *P.s.* and *P.p.*  $\beta$ -CA structures. Cab is one of the smallest  $\beta$ -CAs known, and lacks an extended C-terminus. In the *P.s.*  $\beta$ -CA structure, the C-terminus forms a long  $\beta$ -strand that mediates octamerization. In Cab, residues 90-125 form two helices ( $\alpha 4$ ,  $\alpha 5$ ) that project out to cover the second monomer (Fig. 2B), and fold back to start helix  $\alpha 6$ . In both the *P.s.* and *P.p.*  $\beta$ -CA structures, this segment is longer, forms three helices instead of two, and folds back earlier to create two additional turns of helix  $\alpha 6$ . In the central  $\beta$ -sheet region, the rmsds in C $_{\alpha}$  positions between Cab and *P.s.* and *P.p.* structures are 0.62 Å and 0.56 Å, respectively.

*Active Site:* The active site cleft is located at the C-terminus of the parallel  $\beta$ -sheet and is largely sequestered from solvent. Each Cab subunit contains one zinc atom that resides at the interface of the two monomers (Fig. 2B), although the coordination residues (Cys32, Cys90, and His87) originate from the same monomer. One water molecule completes the tetrahedral coordination sphere of the zinc. Although the

crystallization conditions contained 50 mM calcium acetate, no acetate was found in the active site, unlike in the *P.s.*  $\beta$ -CA structure (Fig. 5A). The average coordination distances from the six active sites are Cys32 S $\gamma$ -Zn ( $2.42 \pm 0.03$  Å), His87 N $\epsilon$ -Zn ( $2.11 \pm 0.04$  Å), Cys 90 S $\gamma$ -Zn ( $2.40 \pm 0.04$  Å), and H $_2$ O-Zn ( $2.15 \pm 0.06$  Å). The four zinc ligands form a number of hydrogen bonds with the surrounding residues. His87 hydrogen bonds with the carbonyl oxygen of residue Thr88; Cys32 S $\gamma$  hydrogen bonds with the amide nitrogens of residues Lys22, Asp34 and Gly59; Cys90 hydrogen bonds with the amide nitrogen of residue Met92; and the coordinating water molecule makes a hydrogen bonds to Asp34. Arg36 forms the conserved Asp/Arg pair with Asp34 and also interacts with Asp89 and a water molecule.

Rather unexpectedly, the two active sites A and B in the Cab dimer exhibit significant differences, which are reflected in the rmsds between monomers in a dimer

Fig. 4.

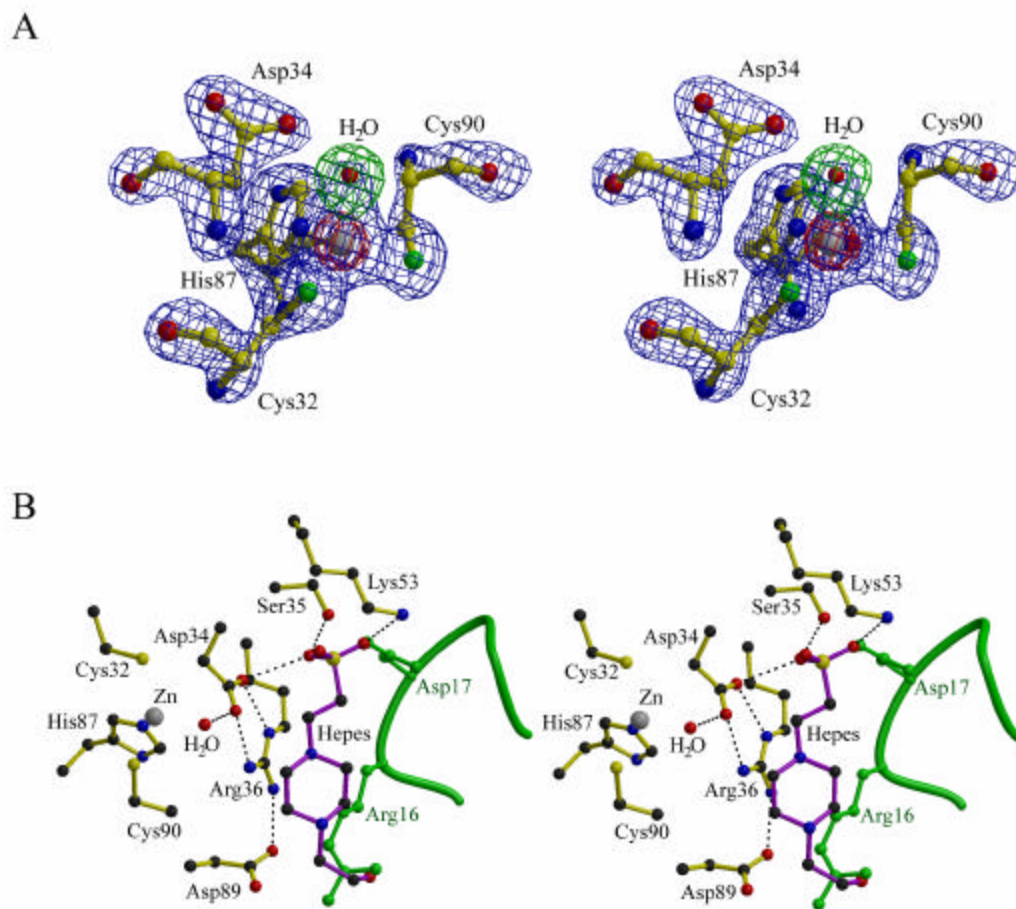


**Fig. 4.** Stereo diagram showing the superimposition of a backbone trace of Cab (yellow), *P. s.*  $\beta$ -CA (green), and *P. p.*  $\beta$ -CA (blue). The main differences are at the N-terminus, C-terminus, and in the  $\alpha 4$ ,  $\alpha 5$  region.

(~0.65 Å) being consistently larger than between the equivalent monomers in different dimers, such as A and C, or B and D (0.24 Å). The active sites in monomers A, C, and E have a Hepes buffer molecule bound near the active site. The sulfate group of the Hepes is located ~8 Å from the zinc atom and the sulfate oxygens hydrogen bond with Lys53 N $\zeta$ , Ser35 O $\gamma$ , and the amide nitrogen of Ser35 (Fig. 5B). The equivalent to Ser35 is present in both 'plant' type  $\beta$ -CAs and in Cab, while Lys53 is unique to the 'cab' type  $\beta$ -CAs. Asp34, that makes a hydrogen bond to the zinc coordinating water molecule, is also within hydrogen bonding distance (~3.0 Å) to the Hepes sulfate group. Another difference between the two active sites is the conformation of residues 13-24. In the active site of subunit A, residues B13-B24 are disordered and have no visible electron density. In the active site of subunit B, however, residues A13-A24 are well defined and extend out to participate in crystal packing. Superposition of residues in the active sites of subunits A and B (Fig. 5B) indicates that the Hepes molecule would sterically clash with residues Arg16 and Asp17, if residues 13-24 adopted the same conformation as in active site B.

*Comparison of P.s.  $\beta$ -CA and Cab active sites:* The zinc coordinating residues (Cys32, His 87, and Cys90) of Cab and P.s.  $\beta$ -CA superimpose closely (Fig. 6A). The water molecule, serving as the fourth zinc ligand in Cab, adopts a position similar to the O1 of the acetate ligand in the P.s.  $\beta$ -CA structure. The conserved residue Asp34 is held in place by forming two hydrogen bonds to Arg36, as do the corresponding residues Asp162 and Arg164 in the P.s.  $\beta$ -CA. The following five active site substitutions distinguish the 'cab' type and 'plant' type  $\beta$ -carbonic anhydrases: H23Q, M33A, K53F, A58V, V72Y. The superposition of Cab and P.s.  $\beta$ -CA structures clearly shows the

Fig. 5.



**Fig. 5.** Stereo view of the zinc environment in Cab. **A)** The simulated annealing  $|F_o - F_c|$  omit map, calculated omitting the four ligands and the zinc, is contoured at  $4\sigma$  (blue) and  $15\sigma$  (red). A simulated annealing omit map calculated with only the coordinating water molecule omitted is shown contoured at  $4\sigma$  (green). **B)** Hepes binding to the active site of subunit A. Hepes is in purple, with hydrogen bonds shown as dotted lines. The loop shown in green would sterically clash with Hepes, if Hepes bound to the active site of subunit B as it is observed to bind to the active site of subunit A.

distinctions in active site organization by these residues. His23, found on a flexible loop, is  $\sim 25$  Å away from the catalytic zinc, and likely is not a part of the active site. On the

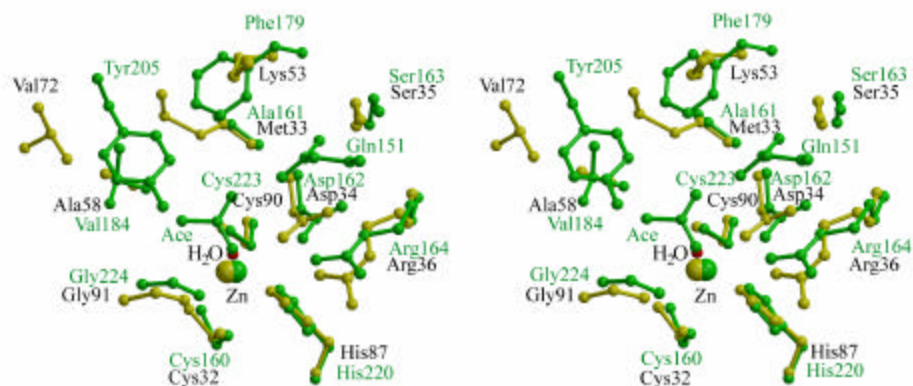
other hand, the equivalent residue Gln151 in the *P.s.*  $\beta$ -CA structure occupies a position similar to the Hepes sulfate group that is  $\sim 8$  Å away from the zinc atom and is possibly involved in ligand binding (23). The hydrophobic pocket of Cab is quite different from that of 'plant' type  $\beta$ -CA (Fig. 6A). In *P.s.*  $\beta$ -CA structure, the hydrophobic pocket is formed by Phe179, Val184, and Tyr205. The corresponding residues in Cab (Lys53, Ala58, and Val72) constitute a more open and less hydrophobic pocket.

*Comparison of P.p. b-CA and Cab active sites:* Again, the protein ligands to the zinc in the two structures superpose closely (Fig 6B). The fourth ligand is different as the side chain of Asp151 in the *P.p.*  $\beta$ -CA structure coordinates the zinc, instead of the water molecule seen in Cab structure. Asp151 of *P.p.*  $\beta$ -CA is equivalent to Asp34 of Cab. As a consequence of the zinc coordination by Asp151 in the *P.p.*  $\beta$ -CA structure, this residue cannot pair with the conserved Arg, and Arg153 has flipped away from the active site. The adjacent Ser152 moved by  $\sim 4$  Å and adopted a different conformation in the *P.p.*  $\beta$ -CA structure (Fig. 6B). The hydrophobic pocket arrangement of *P.p.*  $\beta$ -CA is very similar to that of *P.s.*  $\beta$ -CA and the differences between the 'cab' type and 'plant' type hydrophobic pocket have already been discussed.

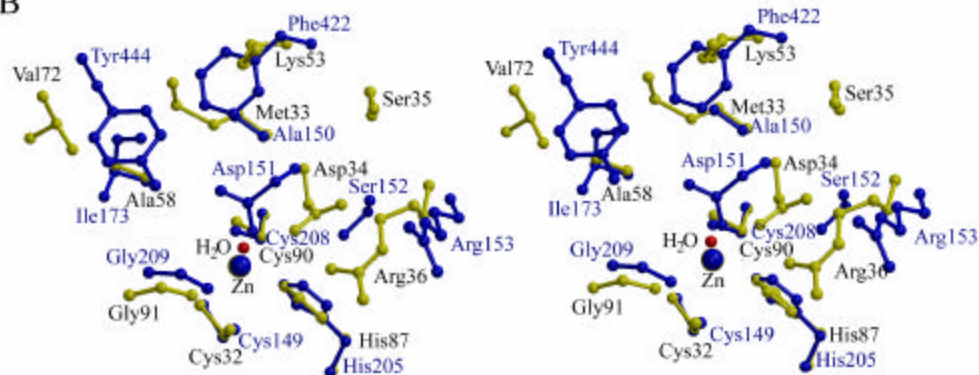
*Mechanism:* The catalytic mechanism of carbonic anhydrases has been most extensively studied for the  $\alpha$ -CA class (5,6,40-42). The zinc hydroxide mechanism established for this class provides an appropriate framework for discussing the catalytic mechanism of Cab. In the first part of the CO<sub>2</sub> hydration reaction, CO<sub>2</sub> binds in the hydrophobic pocket and interacts with the amide nitrogen of the gatekeeper residue. The gatekeeper residue (Thr199) in the  $\alpha$ -CAs plays an important role, together with Glu106,

Fig. 6.

A



B



**Fig. 6.** Stereo diagram showing the superimposition of the active site of Cab, *P. s.*, and *P. p.*  $\beta$ -CAs including zinc, zinc-coordinating residues, as well as the conserved residues differentiating between 'cab' and 'plant'  $\beta$ -CA type. **A)** Cab is shown in yellow, *P. s.*  $\beta$ -CA is shown in green. The coordinating water molecule in Cab is shown in red. **B)** Cab is shown in yellow, *P. p.*  $\beta$ -CA is shown in blue. The coordinating water molecule in Cab is shown in red.

in orienting the  $\text{CO}_2$  molecule for attack by the zinc bound hydroxide. In the *P.s.*  $\beta$ -CA structure, Asp162, Gly224, and Gln151 are thought to play the same role in orienting  $\text{CO}_2$  for this attack (23). In Cab, Asp34 and Gly91 are in the same orientation as Asp162 and Gly224 in *P.s.*  $\beta$ -CA structure and might also help to orient  $\text{CO}_2$ . His23, the

equivalent of *P.s.*  $\beta$ -CA Gln151, is; however, disordered in active sites A, C, and E, and in active sites B, D, and F is at the beginning of a segment of residues that pack against the symmetry related molecule, and lies 25 Å away from the active site zinc.

The second, rate-limiting step in the CO<sub>2</sub> hydration reaction involves the regeneration of a hydroxide ion from the zinc-bound water molecule. In  $\alpha$ -CAII, the zinc ion is located in a deep funnel and requires a proton shuttle to transfer the proton to the bulk solvent. His64 of  $\alpha$ -CAII adopts multiple conformations, which facilitates accepting the proton from the zinc bound water molecule and delivering it to buffer in bulk solution (43). In  $\gamma$ -CA, Glu84 exhibits multiple conformations and has been proposed to participate in a proton shuttle (18,44). Residues with multiple conformation have not been described in the active site of any  $\beta$ -CA structure determined so far. Since the  $\beta$ -CA active site is closer to the surface of the protein than the  $\alpha$ -CA active site, a protein mediated proton shuttle might not be necessary. The  $\beta$ -CA reaction rate depends on buffer concentration, implying that proton transfer can be rate limiting under certain conditions (21). In the Cab structure, a HEPES buffer molecule was found near the active sites A, C, and E. The HEPES sulfate group is located ~8 Å away from the zinc atom, and lies within hydrogen bonding distance of Asp34, which makes a hydrogen bond to the zinc bound water molecule. In the *P.p.*  $\beta$ -CA structure, the equivalent of Asp34 acts as the fourth zinc ligand, and in the proposed mechanism plays a role in the proton transfer (24). Therefore, the most plausible pathway for proton transfer in Cab is from the zinc bound water molecule to Asp34 and then to the sulfate group of the bound HEPES molecule. The conformation of residues 1-25 in active sites B, D, and F is incompatible with HEPES binding and residues 13-25 must adopt different conformations for HEPES to



bind (Fig. 5B). The mobility of residues 1-25 and 92-125 might allow buffer molecules to diffuse into the active site, and serve as the proton acceptor necessary to regenerate the zinc-bound hydroxide.

### **Acknowledgements**

We would like to thank Jessica Wu and Radu Georgescu for help with crystallizations, and Brian R. Crane and Alex M. Bilwes for help with data collection. This work was supported by grants from the National Institute of Health J.G.F (GM44661) and D.C.R. (GM45162), NASA-AMES Cooperative Agreement NCC2-1057 to the Pennsylvania State University Astrobiology Research Center (PSARC), and an NSF predoctoral fellowship (P.S.). This work is based upon research conducted at the Stanford Synchrotron Radiation Laboratory (SSRL), which is funded by the Department of Energy (BES, BER) and the National Institutes of Health (NCRR, NIGMS).

## References

1. Meldrum, H., and Roughton, F. (1933) *J. Physiol.* **80**, 113-142
2. Smith, K. S., Jakubzick, C., Whittam, T. S., and Ferry, J. G. (1999) *Proc. Nat. Acad. Sci. USA* **96**, 15184-15189
3. Hewett-Emmett, D., and Tashian, R. E. (1996) *Mol. Phylogenet. Evol.* **5**, 50-77
4. Soltés-Rak, E., Mulligan, M. E., and Coleman, J. R. (1997) *J. Bacteriol.* **179**, 769-774
5. Coleman, J. E. (1998) *Curr. Opin. Chem. Biol.* **2**, 222-234
6. Sly, W. S., and Hu, P. Y. (1995) *Annu. Rev. Biochem.* **64**, 375-401
7. Kannan, K. K., Notstrand, B., Fridborg, K., Lövgren, S., Ohlsson, A., and Petef, M. (1974) *Proc. Natl. Acad. Sci. USA* **72**, 51-55
8. Kannan, K., Ramanadham, M., and Jones, T. (1984) *Ann. NY Acad. Sci.* **429**, 49-60.
9. Håkansson, K., Carlsson, M., Svensson, L. A., and Liljas, A. (1992) *J. Mol. Biol.* **227**, 1192-1204
10. Eriksson, A. E., and Liljas, A. (1993) *Proteins* **16**, 29-42
11. Boriack-Sjodin, P. A., Heck, R. W., Laipis, P. J., Silverman, D. N., and Christianson, D. W. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 10949-10953
12. Stams, T., Nair, S. K., Okuyama, T., Waheed, A., Sly, W. S., and Christianson, D. W. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 13589-13594
13. Lindskog, S. (1997) *Pharmacol. Thera.* **74**, 1-20
14. Alber, B. E., and Ferry, J. G. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 6909-6913
15. Alber, B. E., and Ferry, J. G. (1996) *J. Bacteriol.* **178**, 3270-3274
16. Alber, B. E., Colangelo, C. M., Dong, J., Stålhandske, C. M. V., Baird, T. T., Tu, C., Fierke, C. A., Silverman, D. N., Scott, R. A., and Ferry, J. G. (1999) *Biochemistry* **38**, 13119-13128
17. Kisker, C., Schindelin, H., Alber, B. E., Ferry, J. G., and Rees, D. (1996) *EMBO J.* **15**, 2323-2330
18. Iverson, T. M., Alber, B. E., Kisker, C., Ferry, J. G., and Rees, D. C. (2000) *Biochemistry* **39**, 9222-9231
19. Badger, M. R., and Price, G. D. (1994) *Ann. Rev. Plant Physiol. Plant Mol. Biol.* **45**, 369-392
20. Smith, K. S., and Ferry, J. G. (1999) *J. Bacteriol.* **181**, 6247-6253
21. Smith, K. S., Cosper, N. J., Stalhandske, C., Scott, R. A., and Ferry, J. G. *J. Bacteriol.*, in press
22. Ferry, J. (1997) *BioFactors* **6**, 25-35
23. Kimber, M. S., and Pai, E. F. (2000) *EMBO J.* **19**, 1407-18
24. Mitsuhashi, S., Mizushima, T., Yamashita, E., Yamamoto, M., Kumasaka, T., Moriyama, H., Ueki, T., Miyachi, S., and Tsukihara, T. (2000) *J. Biol. Chem.* **275**, 5521-5526
25. Bracey, M. H., Christiansen, J., Tovar, P., Cramer, S. P., and Bartlett, S. G. (1994) *Biochemistry* **33**, 13126-13131
26. Rowlett, R. S., Chance, M. R., Wirt, M. D., Sidelinger, D. E., Royal, J. R., Woodroffe, M., Wang, Y. F., Saha, R. P., and Lam, M. G. (1994) *Biochemistry* **33**, 13967-13976
27. Otwinowski, Z., and Minor, W. (1997) *Methods Enzymol.* **276**, 307-326
28. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Cryst. sect. D* **55**, 849-861
29. Bailey, S. (1994) *Acta Cryst. sect. D* **50**, 760-763
30. Cowtan, K. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34-38
31. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Cryst. sect. A* **47**, 110-119
32. Brunger, A. T. (1992) *X-PLOR Manual Version 3.1--A System for X-ray Crystallography and NMR*, Yale University, New Haven, CT
33. Guilloton, M. B., Korte, J. J., Lamblin, A. F., Fuchs, J. A., and Anderson, P. M. (1992) *J. Biol. Chem.* **267**, 3731-3734
34. Nicholls, A., Sharp, K. A., and Honig, B. (1991) *Proteins* **11**, 281-296
35. Thanos, C. D., Goodwill, K. E., and Bowie, J. U. (1999) *Science* **283**, 833-836
36. Bennett, M. J., Choe, S., and Eisenberg, D. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 3127-31
37. Bennett, M. J., Schlunegger, M. P., and Eisenberg, D. (1995) *Protein Sci.* **4**, 2455-2468
38. Jones, S., and Thornton, J. M. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 13-20
39. Janin, J. (1997) *Nat. Struct. Biol.* **4**, 973-974
40. Coleman, J. E. (1967) *J. Biol. Chem.* **242**, 5212-5219
41. Lindskog, S., and Coleman, J. E. (1973) *Proc. Nat. Acad. Sci. USA* **70**, 2505-2508

42. Lindskog, S., and Liljas, A. (1993) *Curr. Op. Struct. Biol.* **3**, 915-920
43. Tu, C. K., Silverman, D. N., Forsman, C., Jonsson, B. H., and Lindskog, S. (1989) *Biochemistry* **28**, 7913-7918
44. Tripp, B., and Ferry, J. (2000) *Biochemistry* **39**, 9232-9240
45. Esnouf, R. (1997) *J. Mol. Graph.* **15**, 133-138
46. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* **277**, 505-524

**Chapter 5:**  
**Crystal Structure of the All-ferrous [4Fe-4S]<sup>0</sup> Form of the  
Nitrogenase Iron Protein from *Azotobacter vinelandii***

Reprinted from Biochemistry

## Crystal Structure of the All-ferrous [4Fe-4S]<sup>0</sup> Form of the Nitrogenase Iron Protein from *Azotobacter vinelandii*<sup>†,‡</sup>

Pavel Strop<sup>§</sup>, Patricia M. Takahara<sup>£</sup>, Hsiu-Ju Chiu<sup>@</sup>, Hayley C. Angove<sup>§</sup>, Barbara K.  
Burgess<sup>§</sup>, and Douglas C. Rees<sup>@</sup>

<sup>†</sup>This work was supported by the National Institute of Health (B.K.B. and D.C.R.), NSF fellowship (P.S.), and NSF postdoctoral fellowship (P.M.T).

<sup>§</sup>Biochemistry Option, California Institute of Technology, Mail Code 147-75, Pasadena, California, 91125, USA.

<sup>£</sup>Division of Chemistry and Chemical Engineering, California Institute of Technology, Mail Code 147-75CH, Pasadena, California, 91125, USA.

<sup>§</sup>Department of Molecular Biology and Biochemistry, University of California, Irvine, California 92697-3900, USA

<sup>@</sup>Howard Hughes Medical Institute and Division of Chemistry and Chemical Engineering, California Institute of Technology, Mail Code 147-75CH, Pasadena, California, 91125, USA.

## **Abstract**

The structure of the nitrogenase iron protein from *Azotobacter vinelandii* in the all-ferrous [4Fe-4S]<sup>0</sup> form has been determined to 2.25 Å resolution using the multiwavelength anomalous diffraction (MAD) phasing technique. The structure demonstrates that major conformational changes are not necessary in either the iron protein or in the cluster to accommodate cluster reduction to the [4Fe-4S]<sup>0</sup> oxidation state. A survey of [4Fe-4S] clusters coordinated by four cysteine ligands in proteins of known structure reveals that the [4Fe-4S] cluster of the iron protein has the largest surface accessible area, suggesting that solvent exposure may be relevant to the capability of the iron protein to exist in three oxidation states.

Biological nitrogen fixation is catalyzed by the nitrogenase system, which consists of two metalloproteins: the iron protein (Fe protein) and the molybdenum-iron protein (MoFe protein) (1-4). The Fe protein is a ~64 kDa  $\gamma_2$  dimer with a single [4Fe-4S] cluster bridging the two subunits, while the MoFe protein is an  $\alpha_2\beta_2$  tetramer that contains two types of metalloclusters, the FeMo-cofactor and the P-cluster. The reduction of dinitrogen to ammonia by nitrogenase involves the reduction of the Fe protein by electron carriers, such as ferredoxin and flavodoxin. The Fe protein subsequently transfers electrons to the MoFe protein in a MgATP-dependent process, followed by electron and proton transfer to the substrate. Each electron transfer step involves the association and dissociation of the Fe protein and MoFe protein complex. This cycle repeats until enough electrons are transferred to the substrate, which is most likely bound to the FeMo-cofactor of the MoFe protein. In the current understanding of the nitrogenase mechanism, the [4Fe-4S] cluster of the Fe protein is believed to undergo a one electron reduction cycle between the [4Fe-4S]<sup>1+</sup> and [4Fe-4S]<sup>2+</sup> oxidation states.

In 1994, Watt et al. (5) reported that the [4Fe-4S]<sup>1+</sup> state of the Fe protein can be further reduced to the all-ferrous [4Fe-4S]<sup>0</sup> state. The properties of an all-ferrous Fe protein produced by reduction with Ti(III) citrate have been extensively characterized by Burgess, Münck and coworkers (6-10). The all-ferrous Fe protein has been shown to bind MgATP, undergo a MgATP induced conformational change, form the productive complex with the MoFe protein, and transfer electrons to the MoFe protein, leaving the cluster of Fe protein in the [4Fe-4S]<sup>2+</sup> oxidation state (7). It is unclear whether the transfer of electrons between proteins takes place one at a time or if both electrons can be transferred together. Starting with the [4Fe-4S]<sup>0</sup> state of the Fe protein, one ATP

molecule can apparently be hydrolyzed per electron transferred (11), rather than the two ATP molecules per electron observed from the [4Fe-4S]<sup>1+</sup> state. Therefore, the ATP utilization process can potentially be twice as efficient through the all-ferrous state of the Fe protein. It has been suggested that both forms of reduced Fe protein might be mechanistically competent (11).

Before discovery of the all-ferrous Fe protein, protein bound [4Fe-4S] clusters had been known to undergo either [4Fe-4S]<sup>3+/2+</sup> or [4Fe-4S]<sup>2+/1+</sup> redox cycles, but not both (12). The [4Fe-4S] cluster of the Fe protein is the only known example of a protein bound [4Fe-4S] cluster that can function in three oxidation states. Furthermore, no other [4Fe-4S] protein clusters or model compounds with thiolate ligands are known to exist in the all-ferrous [4Fe-4S]<sup>0</sup> oxidation state. Here we report a crystal structure of an all-ferrous state of the Fe protein at 2.25 Å resolution. Since there are no model compounds with thiolate ligands in this oxidation state, this work provides the first crystallographic view of an all-ferrous [4Fe-4S]<sup>0</sup> cluster.

## **Methods**

*Protein Purification and Reduction to the [4Fe-4S]<sup>0</sup> state: Azotobacter vinelandii* Fe protein (Av2) was purified as described elsewhere (13). All experimental procedures were carried out in an anaerobic chamber with an oxygen level of less than 1 ppm. Ti(III) citrate and Ti(III) citrate reduced Fe protein were prepared as described previously (6, 7, 14). Ti(III) citrate was removed from the Fe protein on a Sephadex G-25 column, and the protein was concentrated in a Centricon-30 in an anaerobic centrifuge to 30 mg/ml in 20% glycerol, 450 mM NaCl, and 50 mM Tris (pH 8.0). It should be noted that



Table 1. Data Collection Statistics

Data collection	peak	inflection	remote	high- resolution
	?1	?2	?3	
Wavelength (Å)	1.737	1.741	1.378	1.08
Resolution range (Å)	20-3.0	20-3.0	20-3.0	20-2.25
Unique observations	12779	12748	12222	28724
Total observations	66302	64197	55432	154673
Completeness (%)	97(99.6)	97(99.7)	94.4(97.3)	93.2(85.2)
$R_{\text{sym}}$	6.4(15.1)	5.7(20.8)	7.1(23.8)	5.4(30.6)
$\langle I/s \rangle$	7.8(4.0)	8.7(3.1)	8.6(2.8)	21.3(3.8)
$R_{\text{cullis}}$ (cen/acen/anom)	0.72 / 0.59 / 0.85	0.64 / 0.50 / 0.86	./ ./ 0.91	
Phasing power (cen/acen/anom)	2.09 / 3.31 / 1.47	2.33 / 4.83 / 1.48	./ ./ 1.25	
Figure of Merit to 3.0 Å resolution (cen/acen)	0.64 / 0.55			

Numbers in parentheses indicate values for highest resolution bin (3.16-3.00 Å for ?1,?2, and ?3 and 2.33-2.25 Å for high resolution)

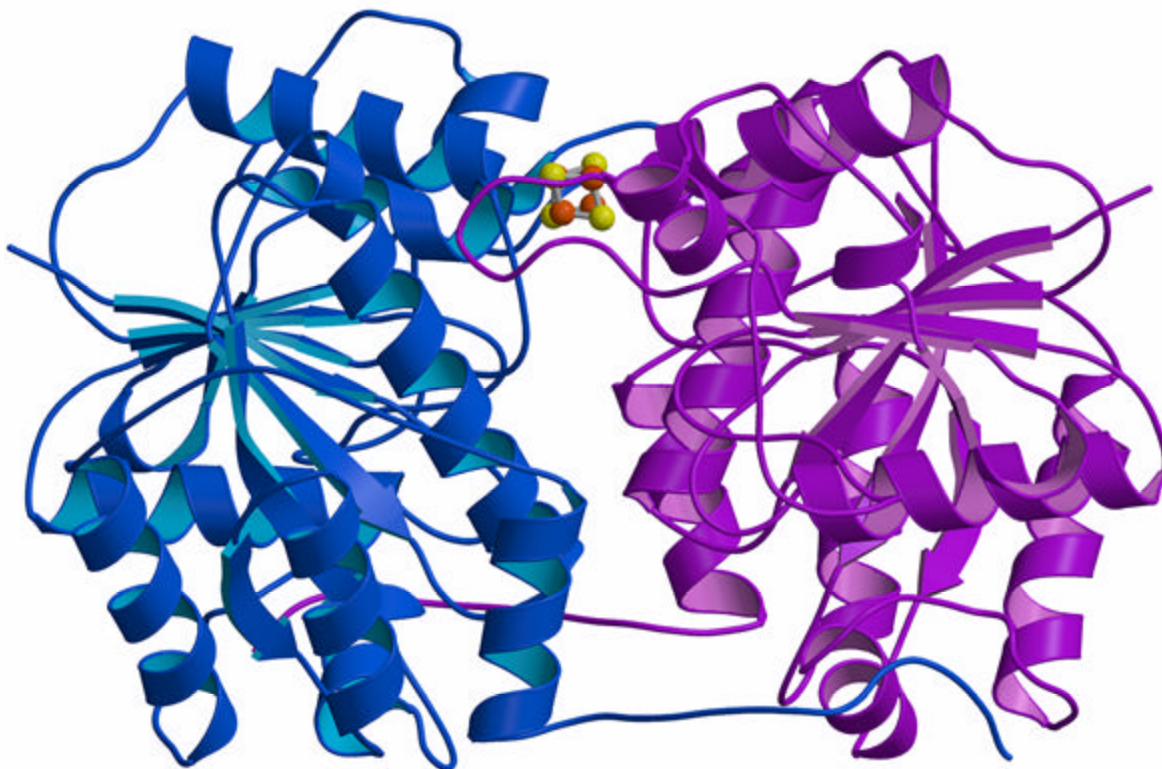
$R_{\text{sym}} = \sum |I_{\text{obs}} - I_{\text{avg}}| / \sum I_{\text{obs}}$

$R_{\text{cullis}}$  = lack of closure error/iso-ano difference

Phasing power = heavy atom structure factor/ r.m.s. lack of closure error

the all-ferrous Av2 is pink whereas the  $[4\text{Fe-4S}]^{+1}$  state is brown. Exposure of the all-ferrous Av2 solution to oxygen immediately leads to oxidation of the cluster and to a dark brown solution

*Crystallization:* Crystals were grown by the capillary microbatch method (15) with 20  $\mu\text{l}$  precipitant solution containing 26-29% PEG 4000, 700-900 mM NaCl, 100 mM Tris (pH 8.0), and 10  $\mu\text{l}$  of protein solution containing 30 mg/ml all-ferrous Av2 in 20% glycerol, 450 mM NaCl, and 50 mM Tris (pH 8.0). Pink crystals were obtained after one



**Figure 1.** Ribbons diagram of the all-ferrous *A. vinelandii* Fe protein dimer. The  $[4\text{Fe-4S}]$  cluster at the dimer interface is represented by the ball-and-stick model. Subunits A and B are colored purple and blue, respectively. Figures 1 and 3 were generated with BOBSCRIPT and RASTER3D (34, 35).

week and belonged to space group  $P2_1$  with cell constants  $a = 57.0 \text{ \AA}$ ,  $b = 93.3 \text{ \AA}$ ,  $c = 62.8 \text{ \AA}$ , and  $\beta = 98.9^\circ$ , which is similar to the crystal form obtained with molybdate described in the original structure determination (15). Molybdate could not be used in these crystallization studies, however, since the presence of this anion led to rapid oxidation of the all-ferrous form of Av2.

*Data Collection and Structure Determination:* Data were collected on frozen crystals, and the color of the crystals was monitored to ensure that data were collected on the all-ferrous state of the protein. At the end of the data collections, there appeared to be slight discoloration on the surfaces of the crystals, however, the crystals always remained

predominantly pink. A high resolution data set was collected at  $-160\text{ }^{\circ}\text{C}$  on beamline 7-1 of the Stanford Synchrotron Research Laboratory (SSRL) (Table 1). The structure was initially solved by molecular replacement with AMORE (16) using the 2NIP structure as the search model. This model, however, could not be refined satisfactorily. At this point, experimental phases were determined to  $3.0\text{ }\text{\AA}$  resolution from a three wavelength MAD data set collected at  $-160\text{ }^{\circ}\text{C}$  on beam line 5.0.2 at the Advanced Light Source (ALS) (Table 1). The fluorescence spectrum measured around the Fe edge of a single crystal was used to select the inflection point ( $\lambda=1.741\text{ }\text{\AA}$ ), the absorption edge ( $\lambda=1.737\text{ }\text{\AA}$ ) and a high-energy remote wavelength ( $\lambda=1.378\text{ }\text{\AA}$ ) for optimization of the anomalous signal. Data were processed with MOSFLM and scaled with the CCP4 program suite (17).

The initial position of the iron-sulfur cluster was taken from the molecular replacement solution and was consistent with the anomalous and dispersive difference Patterson maps. Positions and occupancies of the four cluster irons, as well as  $\Delta f'$  and  $\Delta f''$  values for three wavelengths, were refined with the program SHARP (18). The initial phases were improved by solvent flattening and 2-fold non-crystallographic symmetry averaging (NCS) in DM (19). The model was initially refined to  $3.0\text{ }\text{\AA}$  resolution against SHARP calculated Fourier coefficients based on the two-dimensional centroid of the probability distribution for the native complex structure factor (18). After the  $R_{\text{free}}$  dropped below 35%, the model was refined against the higher resolution ( $2.25\text{ }\text{\AA}$ ) native data set collected at SSRL, incorporating the lower resolution test set reflections into the higher resolution test set. Later in the refinement, individual B factors were refined and the 2-fold NCS restraints were released. Alternate cycles of manual model building using the program O (20), positional and individual B-factor refinement

with programs CNS (21) and REFMAC (17), and addition of water molecules to the model reduced the R and R<sub>free</sub> to 23.8% and 28.9% respectively, where R<sub>free</sub> is calculated for 5% of the reflections. The root mean square deviations (rmsds) of bond lengths and angles from ideal values are 0.012 Å and 1.66° respectively, with 85.9% of the residues in the most allowed region and 14.1% in the additionally allowed region of the Ramachandran plot (22). The average B-factors are: 53.5 Å<sup>2</sup> (mainchain), 55.5 Å<sup>2</sup> (sidechains), 39.8 Å<sup>2</sup> (cluster), and 59.3 Å<sup>2</sup> (solvent). An average B-factor of 55.4 Å<sup>2</sup> is calculated for all protein atoms. The final model contains 4380 protein atoms, one [4Fe-4S] cluster, and 365 water molecules, for a total of 4753 atoms. No Ti(III) citrate or citrate were found bound to the protein.

During cluster refinement, the Fe-S bonds were restrained to 2.33 Å (bond energy of 700 kcal/mol Å<sup>2</sup> in CNS), and the Fe-S-Fe and Fe-SG-CH<sub>2</sub> angles were weakly restrained (angle energy of 100 kcal/mol rad<sup>2</sup> in CNS) to 72° and 109°, respectively. For comparisons between the [4Fe-4S] clusters in different Fe protein structures, the 2NIP structure was refined with the same cluster protocol as the all-ferrous Av2, resulting in the 3NIP model. In the 3NIP refinement, all atoms were held fixed except for the [4Fe-4S] cluster and the cysteine-containing loops (residues 96-98 and 131-133 in each monomer).

In the following discussion, the currently determined structures of Fe proteins will be designated as follows: the present 2.25 Å resolution all-ferrous Av2 structure (all-ferrous Av2), the initial 2.9 Å resolution Av2 structure in a monoclinic crystal form (1NIP), the 2.2 Å resolution Av2 structure in an orthorhombic crystal form (3NIP), the 1.93 Å resolution *C. pasteurianum* Cp2 (1CP2), and the Av1/Av2 complex (1N2C) (15,

23, 24). The oxidation states of the 1NIP, 3NIP and 1CP2 structures have not been unambiguously established. While the initial crystallizations were conducted with the cluster in the dithionite reduced [4Fe-4S]<sup>+1</sup> state, it is possible that the protein oxidizes during the time required for crystallization (25), and could also become photoreduced during X-ray data collection (26). Consequently, the oxidation states of these Fe protein forms are not precisely defined, and will be collectively designated as ‘oxidized’ without assigning formal oxidation states.

## **Results and Discussion**

### **Protein Structure**

*Comparison to other Fe protein structures:* The all-ferrous Fe protein adopts the same overall conformation as the uncomplexed 1NIP and 3NIP Av2 structures (Figure 1). In the structurally conserved  $\beta$ -sheet region, the rmsds in C $_{\alpha}$  positions relative to the all-ferrous Fe protein are 0.41 Å (1NIP), 0.17 Å (3NIP), 0.32 Å (1CP2), and 0.4 Å (1N2C) in monomer A respectively. With all C $_{\alpha}$ 's, the corresponding rmsds are 0.65 Å (1NIP), 1.10 Å (3NIP), and 1.80 Å (1N2C). The structure of the all-ferrous Av2 indicates small deviations from ideal NCS symmetry exist between the two monomers. The rmsd in the  $\beta$ -sheet region between the A and B subunits of the all-ferrous Av2 is 0.3 Å, while the rmsd for all C $_{\alpha}$  between the two subunits is 0.65 Å, a value similar to the 3NIP structure (0.62 Å). The regions with largest conformational variation include the previously reported loop regions containing residues 37-47, 60-70, 90-101, 113-116, 128-136, and 171-175 (24). These loops adopt different conformations in the 1NIP and 3NIP

structures. In the all-ferrous Av2, these loops exhibit conformations that are intermediate between the 1NIP and the 3NIP conformations.

---

Table 2. Individual Cluster Distances for Fe Proteins (Å)

---

Atom	SRFEP	3NIP	1CP2
Fe1-Fe2	2.69	2.70	2.64
Fe1-Fe3	2.79	2.64	2.62
Fe1-Fe4	2.67	2.71	2.68
Fe2-Fe3	2.63	2.71	2.72
Fe2-Fe4	2.52	2.56	2.54
Fe3-Fe4	2.57	2.67	2.60
S1-S2	3.82	3.72	3.72
S1-S3	3.86	3.80	3.77
S1-S4	3.72	3.66	3.64
S2-S3	3.68	3.65	3.67
S2-S4	3.64	3.73	3.69
S3-S4	3.74	3.68	3.74

---

**Table 2:** Fe-Fe and S-S distances (in Å) in the  $[4\text{Fe-4S}]$  cluster of different Fe-protein structures. Fe1, Fe2, Fe3 and Fe4 are coordinated by the S $\gamma$  of Cys B97, Cys A132, Cys A97 and Cys B132, respectively, while S1, S2, S3 and S4 are the sulfurs opposite to Fe3, Fe4, Fe2 and Fe1, respectively, in the  $[4\text{Fe-4S}]$  cubane.

### Structure and Environment of the $[4\text{Fe-4S}]^0$ Cluster

*Cluster Geometry:* This crystallographic analysis of the all-ferrous form of Av2 verifies that the cluster retains the general cubane structure and maintains the same set of cysteine ligands observed in more oxidized forms of Fe protein. Specific Fe-Fe and S-S distances observed in the all-ferrous and other forms of Fe protein are presented in Table 2. The resolution of the all-ferrous Av2 structure (2.25 Å) is not sufficient to derive

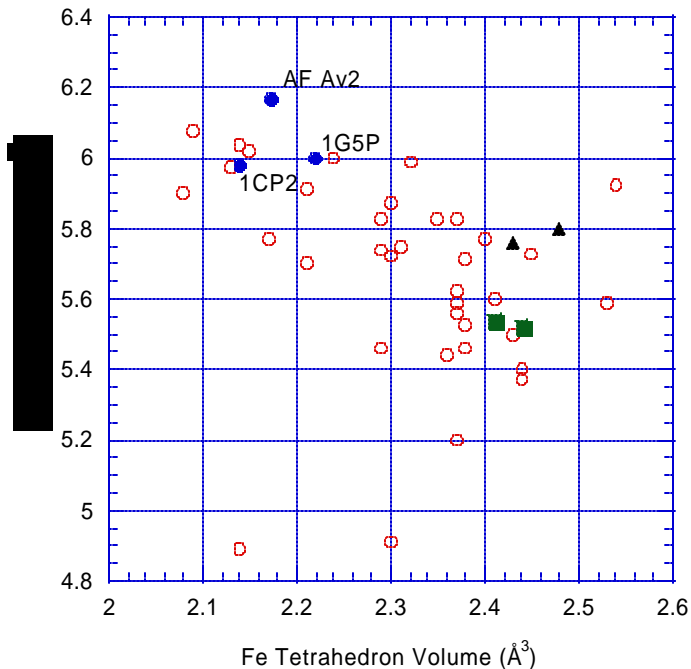
Table 3. Average [Fe<sub>4</sub>S<sub>4</sub>] Cluster Values in PDB<sup>1</sup>

	Å or °	STD	Total # in set	All-ferrous Av2 <sup>2</sup>	3NIP <sup>2</sup>	1CP2 <sup>2</sup>
Fe-SG	2.27	0.06	164	2.31	2.32	2.30
Fe-S	2.29	0.05	492	2.33	2.31	2.30
Fe-Fe	2.68	0.08	246	2.65	2.66	2.63
S-S	3.64	0.08	246	3.74	3.71	3.70
SG-SG	6.36	0.27	246	6.41	6.36	6.37
SG-Fe-S	113.0	6.22	492	111.6	111.3	111.8
S-Fe-S	105.5	2.38	492	107.2	106.6	107.0
Fe-S-Fe	71.8	2.43	492	69.4	70.3	69.7
CB-SG-Fe	108.8	6.78	164	109.0	107.4	110.5
Fe Volume	2.28	0.17	41	2.17	2.22	2.14
S Volume	5.70	0.31	41	6.17	6.00	5.98
SG Volume	29.91	1.40	41	30.82	28.69	30.29
Fe-S Volume	9.27	0.47	41	9.23	9.29	9.06

<sup>1</sup>Values are derived from 41 [Fe<sub>4</sub>S<sub>4</sub>] protein clusters. The selection criteria for the proteins is described in Figure 2 legend

<sup>2</sup>Average values for the indicated Fe-protein structures

accurate [4Fe-4S]<sup>0</sup> cluster geometry parameters, such as the ones available for other oxidation states from model compounds (27). Although the differences are likely within the coordinate uncertainty, the all ferrous Fe-protein shows an overall distribution of four Fe-Fe distances greater than ~2.65 Å and two distances less than this value (Table 2). Since the geometry of the [4Fe-4S] cluster can be sensitive to the restraints used in the refinement of the cluster, the clusters in the all ferrous and more oxidized form of Av2 were refined with the parameter set defined in the Methods section. Comparisons of the cluster geometry refined with the same protocol suggests that, relative to 3NIP, the all-ferrous Av2 has two large S-S distances (~3.80 Å), an increased Fe1-Fe3 distance (2.79

S / Fe Tetrahedron Volumes for  $[4\text{Fe-4S}]$  clusters

**Figure 2.** Variation in the volumes of the Fe and sulfur tetrahedron in  $[4\text{Fe-4S}]$  cluster containing proteins of known structure. The all-ferrous Av2 has the largest sulfur tetrahedron volume, but one of the smallest Fe tetrahedron volumes. The overall all-ferrous Av2 cluster volume is comparable to other  $[4\text{Fe-4S}]$  clusters (Table 3). Only proteins with  $[4\text{Fe-4S}]$  clusters coordinated by four cysteine ligands and refined at resolutions of 2.5 Å or better were used in this analysis. The set contains the following PDB (36, 37) files: All-ferrous (AF) Av2, 3NIP, 1CP2, 1BLU, 2FDN, 1FXR, 2FXB, 1VJW, 7FD1, 1ISU, 1HPI, 1CKU, 2NAP, 1AA6, 1QLA, 1AOP, 2ABK, 1DJN, 1MUU, 1AOR, 1B25, 1BOP, 1C4C, 1H2R. Volumes for model compounds  $[\text{Fe}_4\text{S}_4(\text{SCH}_2\text{Ph})_4]^{2+/1+}$  and  $[\text{Fe}_4\text{S}_4(\text{SPh})_4]^{2+/1+}$  are also shown in the oxidized states (squares) and in the reduced states (triangles) (27). Tetrahedron volumes were calculated from the coordinates of the four defining atoms using equation 20 in section 2.2 of (38).

Å), and a shorter Fe1-Fe4 distance (2.57 Å) (Table 2). For reference, a more general analysis of the geometrical properties of  $[4\text{Fe-4S}]$ -type clusters, including average distances and angles derived from 41  $[4\text{Fe-4S}]$  cluster containing proteins with four cysteine ligands are shown in Table 3.

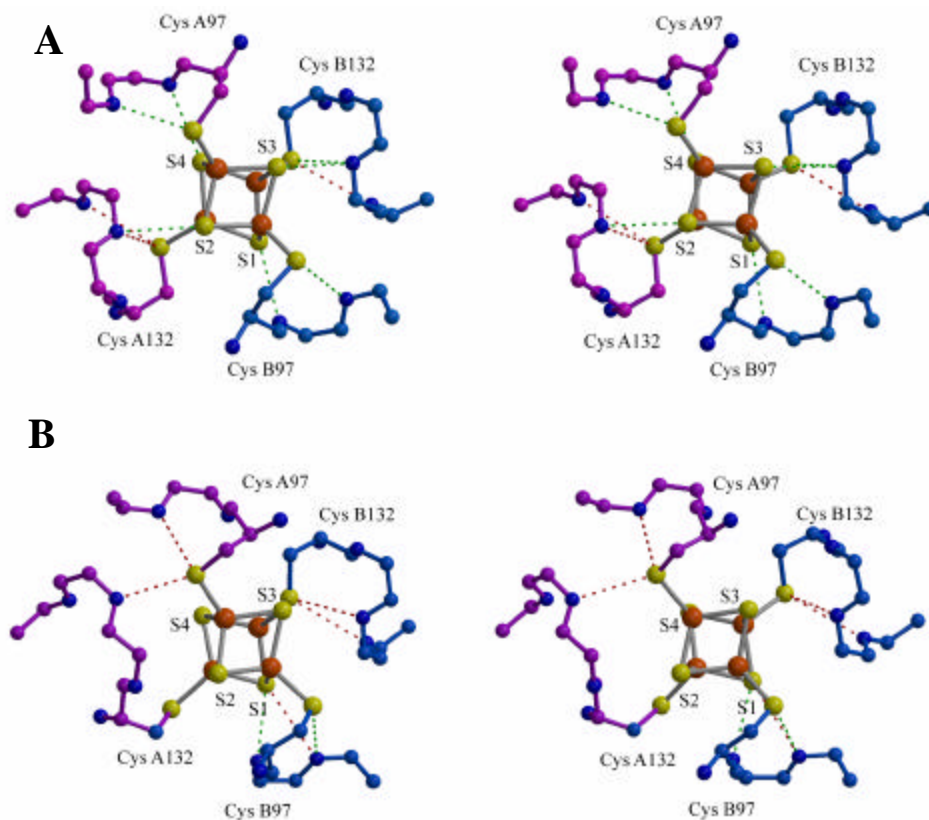
The all-ferrous Fe protein cluster has been studied previously by EXAFS, Mössbauer, and EPR techniques (6, 8-10). The EXAFS curves of the Fe protein were best fit using two short (2.52 Å) Fe-Fe distances and one long (2.77 Å) Fe-Fe distance. A possible model with these distances is that of a compressed  $[4\text{Fe-4S}]$  cluster in which the



vertical Fe-Fe distances are shorter and the horizontal Fe-Fe distances are longer (8). In the Mössbauer studies, one Fe site was found to be unique (10), although the correspondence between this site and the EXAFS derived cluster model is not obvious. It is also not readily apparent how to unambiguously map the EXAFS and Mössbauer derived structural information onto the crystallographic structure, although the coordinate uncertainties in the diffraction analysis make this comparison difficult.

The presence of multiple forms of an iron-sulfur cluster in a crystal can lead to apparent structural distortions following refinement (28). These multiple forms are generally indicated by high B-factors and poor electron density for parts of the cluster. Since the all-ferrous Av2 cluster has a good uniform density with B-factors lower than its average mainchain Bs for the protein, this suggests the cluster exists predominantly in a single form.

*Cluster Volumes:* It has been reported (8) that the volume of the [4Fe-4S] cluster increases upon addition of electrons, although in general, the structural consequences of changes in oxidation state of a cluster will depend on the details of the bonding interactions (29). The Fe proteins have some of the smallest Fe-tetrahedron volumes among [4Fe-4S] proteins, ranging from 2.15-2.20 Å<sup>3</sup> (Figure 2). In contrast, the sulfur tetrahedron in the all-ferrous Av2, due to the longer S-S bond lengths, has one of the largest volumes among [4Fe-4S] proteins (6.15 Å<sup>3</sup>). An increase in volume of the sulfur tetrahedron upon reduction from the [4Fe-4S]<sup>2+</sup> to [4Fe-4S]<sup>1+</sup> state has also been observed in model compounds (27). The overall volume of the all-ferrous Av2 [4Fe-4S] cluster is, however, similar to other [4Fe-4S] protein clusters (Table 3).



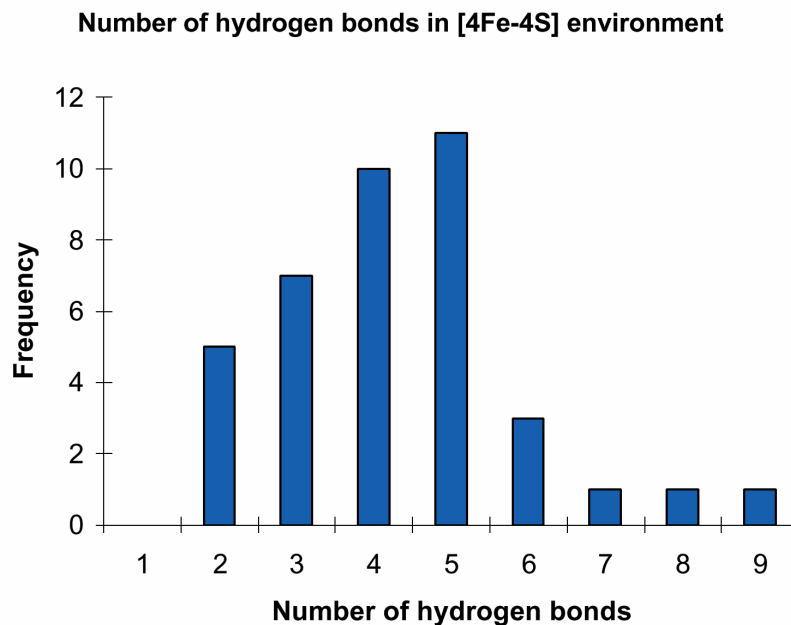
**Figure 3.** Stereo diagram of the  $[4\text{Fe-4S}]$  cluster environment in the **A)** All-ferrous Av2 and **B)** 3NIP structures. NH-S hydrogen bonds were defined when the N-S distance = 3.5 Å and the N-H-S angle  $\geq 120^\circ$ . If the distance is between 3.5 and 3.7 Å, or if the angle is less than  $120^\circ$  then this interaction is termed an incipient bond. NH-S hydrogen bonds are colored green, while the incipient hydrogen bonds are shown in red. Subunits A and B are colored purple and blue, respectively as in Figure 1.

To assess the sensitivity of these volumes to the refinement protocol, the  $[4\text{Fe-4S}]$  clusters of the all-ferrous Av2 and 3NIP were refined with three different restraints. For these refinements, all protein and solvent atoms were fixed, with the exception of the  $[4\text{Fe-4S}]$  cluster and the cysteine-containing loops (residues 96-98, and 131-133 in both monomers). Clusters in all-ferrous Av2 and 3NIP were refined using the following restraint schemes in the program CNS: (1): Fe-S restrained to 2.29 Å, with penalty of 700 kcal mol  $\text{Å}^2$ , (2): Fe-S 2.33 Å, 700 kcal mol  $\text{Å}^2$ , (3): no restraints. The Fe tetrahedron volumes were insensitive to the restraints employed, while the sulfur

tetrahedron was found to fluctuate between 5.9 Å<sup>3</sup> and 6.2 Å<sup>3</sup> in all-ferrous Av2 refinements and to values ~ 0.2 Å<sup>3</sup> smaller in the 3NIP refinements. Although the sulfur volumes appear to have an uncertainty of ~ 0.2 Å<sup>3</sup> depending on the restraints, the sulfur volume of the all-ferrous form of the cluster does appear consistently larger than the more oxidized 3NIP structure.

*Accessible Surface:* The [4Fe-4S] cluster in the Fe protein has the largest accessible surface area among [4Fe-4S] containing proteins of known structure. While most known [4Fe-4S] clusters are completely buried within proteins, the accessible cluster surface areas in the Fe protein structures are between 13-18 Å<sup>2</sup>. For these calculations, the accessible surface area of the clusters was calculated with program SURFACE (17) using radii of 1.8 Å for the Fe and S atoms. Although there is no evidence that surface accessibility correlates with the ability of the [4Fe-4S] cluster to adopt the +2,+1 and 0 oxidation states, solvent exposure has been reported to facilitate reduction of the Fe center in rubredoxin (30).

*N-H-S Hydrogen bonds:* NH-S hydrogen bonds between main chain amide groups and sulfur atoms of the cluster and thiol groups of cysteine ligands (31) are present in all Fe protein structures. For this analysis, NH-S hydrogen bonds were defined as having the N-S distance = 3.5 Å and the N-H-S angle ≥120°. Incipient hydrogen bonds were defined as having N-S distance between 3.5 and 3.7 Å, or having the N-H-S angle between 100° and 120°. The all-ferrous Av2 structure has 7 real and 3 incipient hydrogen bonds (Figure 3A), while the 3NIP structure has only 3 real and 4 incipient hydrogen bonds (Figure 3B), the 1CP2 structure has 5 real and 6 incipient hydrogen bonds, and the 1N2C structure has 4 real and 2 incipient hydrogen bonds. The average



**Figure 4.** Histogram of the number of NH-S hydrogen bonds to the cluster observed in structures of  $[4\text{Fe-4S}]$  cluster containing proteins. Only the hydrogen bonds defined as ‘real’ are shown. The set of proteins used is defined in Figure 2 legend, and the NH-S hydrogen bonds are identified as described in the Figure 3 legend.

number of NH-S hydrogen bonds derived from other known  $[4\text{Fe-4S}]$  cluster containing proteins is 4 real and 3 incipient hydrogen bonds (Figure 4). NH-S hydrogen bonds could help stabilize the more reduced form of the redox centers, as recently detailed for the  $\text{Fe}(\text{CysS})_4$  center in rubredoxin (32). While the larger number of NH-S interactions in the all-ferrous Av2 could help stabilize the reduced  $[4\text{Fe-4S}]^0$  state of the cluster, similar number of NH-S hydrogen bonds have also been observed in other  $[4\text{Fe-4S}]$  containing structures (PDB identifiers 1DJN, 1VJW, 2FDN).

## Conclusions

The  $[4\text{Fe-4S}]$  cluster in the all-ferrous Av2 does not experience any changes in coordination upon reduction to the  $[4\text{Fe-4S}]^0$  oxidation state, unlike the P-cluster of the nitrogenase MoFe protein (33). While Mössbauer spectroscopy suggests one unique Fe site is present in the all-ferrous cluster, such a site could not be identified in the crystal

structure. The distortion of the all-ferrous [4Fe-4S] cluster can not be easily classified; two S-S distances get slightly longer, but the overall volume of the cluster remains nearly constant. The number of NH-S hydrogen bonds to the cluster and thiol groups of cysteine ligands is higher in the all-ferrous Av2 than found in most other [4Fe-4S] proteins. There are, however, a few proteins with similar numbers of cluster hydrogen bonds. One (so far) unique property of the Fe protein cluster is its solvent accessibility. The Fe protein has the most surface accessible cluster of all known [4Fe-4S] proteins, suggesting that the solvent accessibility of the all-ferrous Av2 cluster might play a role in the ability to exist in all three oxidation states. Aside from these correlations, perhaps the most striking aspect of this structural analysis is how closely the all-ferrous Fe protein resembles more oxidized forms. Remarkably, the Fe protein can accommodate three oxidation states of the [4Fe-4S] cluster with only minimal changes in protein or cluster conformation.

### **Acknowledgements**

We thank Tina M. Iverson for assistance with data collection and James B. Howard for discussions. Supported by GM45162 (DCR) and GM43144 (BKB). This work is based upon research conducted at the Stanford Synchrotron Radiation Laboratory (SSRL), which is funded by the Department of Energy (BES, BER) and the National Institutes of Health (NCRR, NIGMS). The Advanced Light Source is supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory.

## References

1. Howard, J. B., and Rees, D. C. (1994) *Annu Rev Biochem* 63, 235-64.
2. Burgess, B. K., and Lowe, D. J. (1996) *Chem. Rev.* 96, 2983-3011.
3. Seefeldt, L. C., and Dean, D. R. (1997) *Acc. Chem. Res.* 30, 260-266.
4. Smith, B. E. (1999) *Adv. Inorg. Chem.* 47, 159-218.
5. Watt, G. D., and Reddy, K. R. N. (1994) *J Inorg Biochem* 53, 281-294.
6. Angove, H. C., Yoo, S. J., Burgess, B. K., and Münck, E. (1997) *J. Am. Chem. Soc.* 119, 8730-8731.
7. Angove, H. C., Yoo, S. J., Münck, E., and Burgess, B. K. (1998) *J Biol Chem* 273, 26330-26337.
8. Musgrave, K. B., Angove, H. C., Burgess, B. K., Hedman, B., and Hodgson, K. O. (1998) *J. Am. Chem. Soc.* 120, 5325-5326.
9. Yoo, S. J., Angove, H. C., Burgess, B. K., Münck, E., and Peterson, J. (1998) *J. Am. Chem. Soc.* 120, 9704-9705.
10. Yoo, S. J., Angove, H. C., Burgess, B. K., Hendrich, M. P., and Münck, E. (1999) *J. Am. Chem. Soc.* 121, 2534-2545.
11. Erickson, J. A., Nyborg, A. C., Johnson, J. L., Truscott, S. M., Gunn, A., Nordmeyer, F. R., and Watt, G. D. (1999) *Biochemistry* 38, 14279-14285.
12. Holm, R., Kennepohl, P., and Solomon, E. (1996) *Chem. Rev.* 96, 2239-2314.
13. Burgess, B. K., Jacobs, D. B., and Stiefel, E. I. (1980) *Biochim Biophys Acta* 614, 196-209.
14. Seefeldt, L. C., and Ensign, S. A. (1994) *Anal. Biochim.* 221, 379-386.
15. Georgiadis, M. M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J. J., and Rees, D. C. (1992) *Science* 257, 1653-9.
16. Navaza, J. (1994) *Acta Crystallogr. A* 50, 157-163.
17. Bailey, S. (1994) *Acta Crystallogr. D* 50, 760-763.
18. de LaFortelle, E., and Bricogne, G. (1997) *Meth. Enzymol.* 276, 472-494.
19. Cowtan, K. (1994) *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* 31, 34-38.
20. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A* 47, 110-119.
21. Brunger, A. T. (1992) *X-PLOR Manual Version 3.1--A System for X-ray Crystallography and NMR*, Yale University, New Haven, CT.
22. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* 26, 283-291.
23. Schindelin, H., Kisker, C., Schlessman, J. L., Howard, J. B., and Rees, D. C. (1997) *Nature* 387, 370-6.
24. Schlessman, J. L., Woo, D., Joshua-Tor, L., Howard, J. B., and Rees, D. C. (1998) *J Mol Biol* 280, 669-85.
25. McKenna, C. E., Gutheil, W. G., and Wei, S. (1991) *Biochim. Biophys. Acta* 1075, 109-117.
26. Nave, C. (1995) *Rad. Phys. Chem.* 45, 483-490.
27. Berg, J. M., and Holm, R. H. (1982) in *Iron-Sulfur Proteins* (Spiro, T. G., Ed.) pp 1-66, John Wiley, New York.
28. Sridhar, V., Prasad, S. G., Burgess, B. K., and Stout, C. D. (1998) *J. Biol. Inorg. Chem.* 3, 140-149.
29. Han, J., Beck, K., Ockwig, N., and Coucouvanis, D. (1999) *J. Am. Chem. Soc.* 121, 10448-10449.
30. Xiao, Z., Maher, M. J., Cross, M., Bond, C. S., Guss, M. J., and Wedd, A. G. (2000) *J. Biol. Inorg. Chem.* 5, 75-84.
31. Adman, E. T., Watenpaugh, K. D., and Jensen, L. H. (1975) *Proc. Natl. Acad. Sci. USA* 72, 4854-4858.
32. Eidsness, M. K., Burden, A. E., Richie, K. A., Kurtz, D. M., Jr., Scott, R. A., Smith, E. T., Ichiye, T., Beard, B., Min, T., and Kang, C. (1999) *Biochemistry* 38, 14803-14809.
33. Peters, J. W., Stowell, M. H., Soltis, S. M., Finnegan, M. G., Johnson, M. K., and Rees, D. C. (1997) *Biochemistry* 36, 1181-1187.
34. Esnouf, R. M. (1997) *J. Molec. Graphics* 15, 133-138.
35. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* 277, 505-524.

36. Sussman, J. L., Lin, D., Jiang, J., Manning, N. O., Prilusky, J., Ritter, O., and Abola, E. E. (1998) *Acta Crystallogr. D Biol. Crystallogr.* 54, 1078-1084.
37. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) *Nucleic Acids Res.* 28, 235-242.
38. Kaspar, J. S., and Lonsdale, K. (1972) *International Tables for X-ray Crystallography*, Vol. II, 3rd ed., D. Reidel, Dordrecht.

**Section III**

**Chapter 6:**

**Rubredoxin Variant Folds without Iron**

Reprinted from J. Amer. Chem. Soc.



## Rubredoxin Variant Folds without Iron<sup>†</sup>

Pavel Strop<sup>‡</sup> and Stephen L. Mayo<sup>§</sup>

<sup>†</sup> This work was supported by the Howard Hughes Medical Institute (S.L.M.) and a NIH fellowship (P.S.).

<sup>‡</sup> Biochemistry Option, California Institute of Technology, mail code 147-75, Pasadena, CA 91125, USA.

<sup>§</sup> Howard Hughes Medical Institute and Division of Biology, California Institute of Technology, mail code 147-75, Pasadena, CA 91125, USA. E-mail: [steve@mayo.caltech.edu](mailto:steve@mayo.caltech.edu); Telephone: 626-396-6408; Fax: 626-568-0934

**Abstract:** *Pyrococcus furiosus* rubredoxin (PFRD), like most studied hyperthermophilic proteins, does not undergo reversible folding. The irreversibility of folding is thought to involve PFRD's iron-binding site. Here we report a PFRD variant (PFRD-XC4) whose iron-binding site was redesigned to eliminate iron binding using a computational design algorithm. PFRD-XC4 folds without iron and exhibits reversible folding with a melting temperature of 82 °C, a thermodynamic stability of 3.2 kcal mol<sup>-1</sup> at 1 °C, and NMR chemical shifts similar to that of the wild-type protein. This variant should provide a tractable model system for studying the thermodynamic origins of protein hyperthermostability.

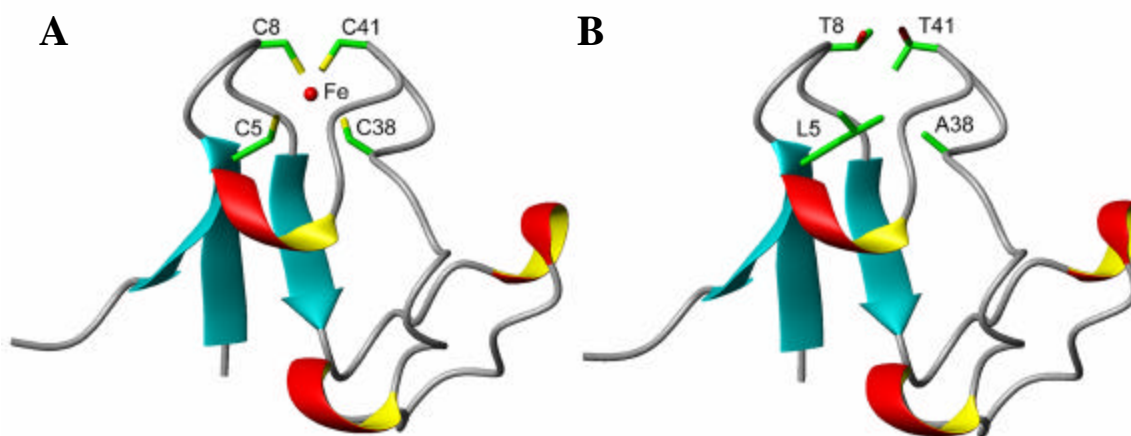
## Introduction

Hyperthermophilic proteins have attracted considerable attention because they can serve as model systems for the determination of factors responsible for protein stability. These include an optimized use of structure elements such as hydrophobic interactions, packing forces, hydrogen bonds, and salt bridges (1-5). In most cases, however, the usual methods for determining thermodynamic stability cannot be used for hyperthermophilic proteins. The measurement of protein stability requires reversible folding, but in almost all studied hyperthermophilic proteins, the denaturation is not reversible (6-8). In the few hyperthermophilic proteins that exhibit reversible folding, there are other elements that complicate the determination of the free energy of folding. These proteins either are dimeric (9,10) or do not fit a two-state denaturation model (10). Irreversible denaturation has been addressed by hydrogen exchange studies (11) and by molecular dynamics simulations (12), but neither of these methods provides an accurate measure of free energy.

Rubredoxins are small (~6 kDa) iron-sulfur proteins with the active site arranged around an iron tetrahedrally coordinated to four cysteinate sulfurs (Figure 1a). Thirteen rubredoxins have been sequenced, and all conserve the four cysteine ligands (residues 5, 8, 38, and 41 in the hyperthermophilic *Pyrococcus furiosus* rubredoxin (PFRD) and residues 6, 9, 39, and 42 in the mesophilic *Desulfovibrio vulgaris* rubredoxin (DVRD)). Since several rubredoxin structures are available from the Brookhaven Protein Data Bank at resolutions of 0.95-1.8 Å (13-19), these proteins are valuable models for the investigation of the origins of thermostability.

It is believed that the iron binding site in rubredoxin is the cause of irreversible

unfolding (20). Here we report a *Pyrococcus furiosus* rubredoxin variant (PFRD-XC4) which was designed to eliminate the iron-binding site using a computational design algorithm. PFRD-XC4 is able to fold in the absence of iron and undergoes reversible denaturation. The variant provides an excellent opportunity for systematic exploration of the factors determining protein thermostability.



**Figure 1.** (A) Ribbon diagram of wild-type mesophilic *Desulfovibrio vulgaris* rubredoxin showing the four cysteine residues that coordinate the iron. (B) Ribbon diagram showing the computed side chain orientations of the four mutations. Structure figures were generated using MOLMOL. (41)

## Experimental Section

*Computational Modeling:* In the calculation using a mesophilic structure, the four cysteine residues that form rubredoxin's iron binding site were optimized in the absence of iron coordination. Two of the cysteines were classified as core residues (Cys<sup>6</sup> and Cys<sup>39</sup>) and two as boundary residues (Cys<sup>9</sup> and Cys<sup>42</sup>). Residues were classified into core, surface, and boundary groups as previously described (21). Seven amino acid types (Ala, Val, Leu, Ile, Phe, Tyr, and Trp) were considered at the two core positions, and 16 amino acid types (Ala, Val, Leu, Ile, Phe, Tyr, Trp, Ser, Thr, Asp, Asn, Glu, Gln, His,

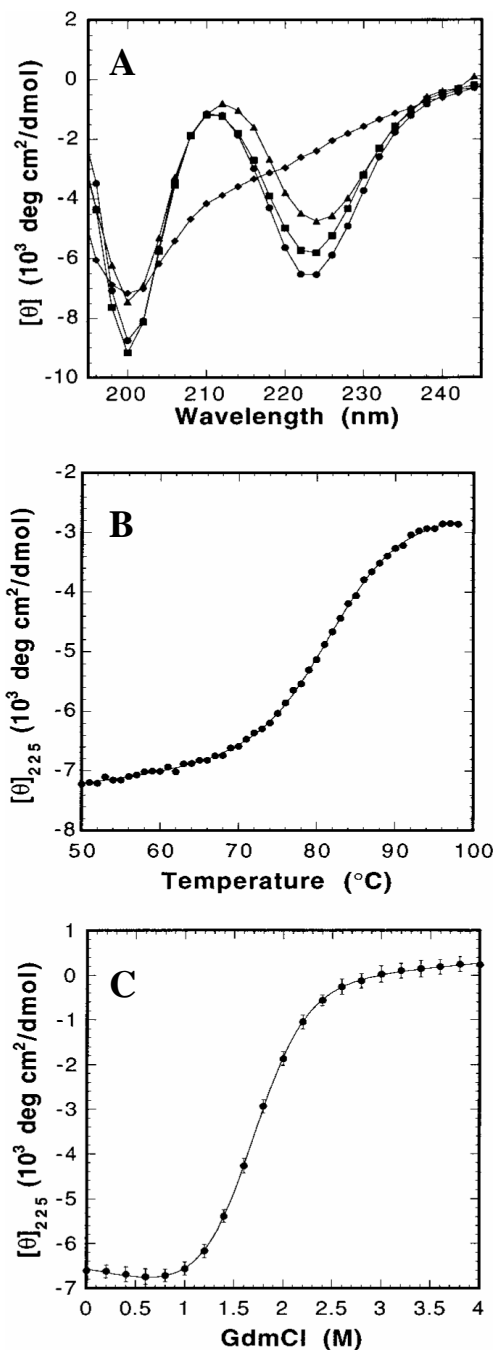
Lys, and Arg) were considered at the two boundary positions. Both amino acid identities and side chain conformations were determined by the algorithm for the four optimized residues. Neighboring residues 7, 10, 40, and 43 were fixed in identity, but their conformations were allowed to change in the calculation. All other residues as well as the backbone were held fixed. Computational details, potential functions, and parameters for van der Waals, solvation, and hydrogen bonding are described in our previous work (21-26).

*Mutagenesis and Protein Purification:* The hyperthermophilic mutant PFRD-XC4 was constructed by inverse PCR (27) using a synthetic *Pyrococcus furiosus* rubredoxin gene in plasmid pt7-7 (4). The mutations for PFRD-XC4 are C5L, C8T, C38A, and C41T. A synthetic gene, based on the sequence of the *Desulfovibrio vulgaris* rubredoxin, was constructed for the mesophilic variant (DVRD-XC4) by recursive PCR(28) and cloned into pt7-7. The mutations for DVRD-XC4 are C6L, C9T, C39A, and C42T. Both mutants were verified by sequencing. Recombinant proteins were expressed by IPTG induction in BL21(DE3) hosts (Invitrogen) as described (29) and isolated using a freeze/thaw method (30). Purification was accomplished by reverse-phase high-performance liquid chromatography using first a linear 1%/min followed by a 0.07%/min acetonitrile/water gradient containing 0.1% TFA. Molecular weights were verified by mass spectrometry.

*CD Analysis:* CD data were collected on an Aviv 62DS spectrometer equipped with a thermoelectric unit and using a 1 mm path length cell. Protein samples were 40  $\mu$ M in 50 mM sodium phosphate buffer at pH 6.3. Concentrations were determined by UV spectrophotometry. Thermal melts were monitored at 225 nm. Data were collected

every 2 °C with an equilibration time of 2 min and an averaging time of 10 s.  $T_m$  was determined by fitting the melting curves to a two-state model as described (31). Guanidinium chloride denaturations were performed at 1 °C.  $\Delta G$ 's,  $m$  values, and error estimates were obtained by fitting the denaturation data to a two-state transition as described (32).

**NMR Studies.** NMR data were collected on a Varian Unity Plus 600 MHz spectrometer equipped with a Nalorac inverse probe with a self-shielded  $z$  gradient. NMR samples were prepared in 90/10 H<sub>2</sub>O/ D<sub>2</sub>O or 99.9% D<sub>2</sub>O with 200 mM NaCl and 25 mM acetate-d<sub>3</sub>, pH 6.3. The sample concentration was approximately 1.5 mM. Sequential assignment of resonances was achieved by the standard homonuclear method (33). Two-dimensional DQF-COSY,(34)



**Figure 2.** Circular dichroism measurements of wild-type *Pyrococcus furiosus* rubredoxin (PFRD) and the PFRD-XC4 variant. (A) Wavelength scans of PFRD-XC4 at 1 °C (circles), 99 °C (diamonds), and refolded at 1 °C (squares). PFRD is also shown at 1 °C for comparison (triangles). (B) Thermal unfolding curve for PFRD-XC4 monitored by CD at 225 nm. (C) Guanidinium chloride denaturation curve for PFRD-XC4 at 1 °C.

TOCSY (35), and NOESY (36) spectra were acquired at 25 °C. The TOCSY spectrum was recorded with an 80 ms mixing time using a clean-MLEV17 mixing sequence. The NOESY spectrum was acquired with a 150 ms mixing time. Water suppression was accomplished either with resaturation during the relaxation delay or with pulsed field gradients (37). Spectra were processed with VNMR (Varian Associates) and were assigned with ANSIG (38).  $^3J_{\text{HNHA}}$  values were derived from NOESY cross-peak fine structure using the INFIT module of XEASY (39).

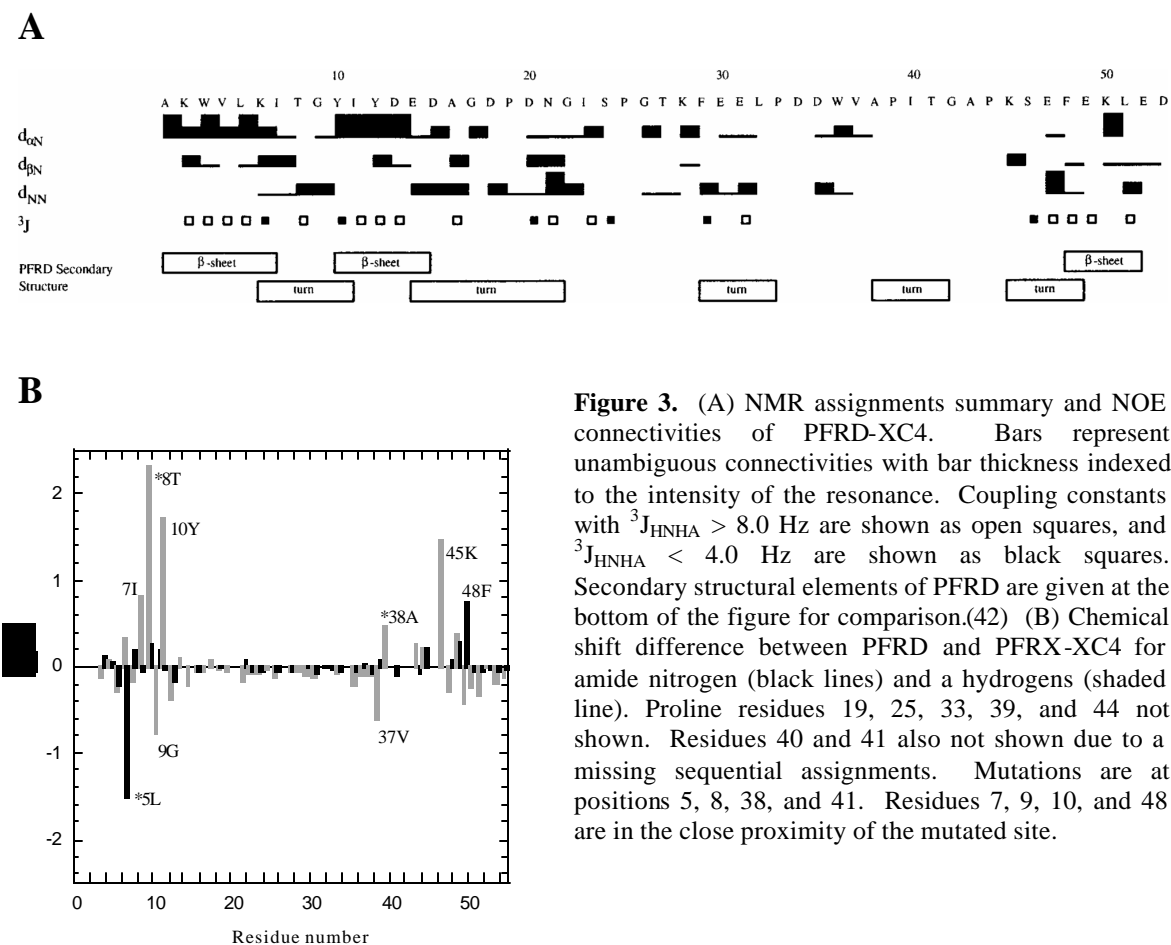
## Results and Discussion

Backbone coordinates for the mesophilic rubredoxin structure used in the calculations were obtained from Brookhaven Protein Data Bank entry 8rxn.(13,15) A radius scale factor of 0.9 for the van der Waals (25) interactions resulted in two threonine residues at the boundary positions and two alanine residues at the core positions. To obtain an alternate sequence that provides greater core volume, selection for the core positions in the calculation was repeated with the radius scale factor decreased to 0.7 and boundary positions 9 and 42 fixed as threonines. The amino acid sequence selected by the algorithm under these conditions contains the following mutations: C6L, C9T, C39A, and C42T (Figure 1b). The equivalent mutations for the hyperthermophilic variant, PFRD-XC4, are C5L, C8T, C38A, and C41T.

The far-ultraviolet circular dichroism (CD) spectrum of PFRD-XC4 is nearly identical to that of the wild-type protein at 1 °C (Figure 2a). Wavelength scans performed at 1 °C, 99 °C, and then again at 1 °C demonstrate reversible folding (Figure 2a). Thermal denaturation of the variant, monitored by CD at 225 nm, shows a

cooperative unfolding transition with a melting temperature ( $T_m$ ) of 82 °C (Figure 2b). Based on the estimated  $T_m$  for the wild-type hyperthermophilic rubredoxin (11), PFRD-XC4 is destabilized by about 80 °C. The mesophilic variant, DVRD-XC4, was also expressed and found to be unfolded at 1 °C. Chemical denaturation of PFRD-XC4 at 1 °C using guanidinium chloride (GdmCl), monitored by CD at 225 nm, yielded a free energy of unfolding ( $\Delta G_u$ ) of  $3.2 \pm 0.1$  kcal mol<sup>-1</sup> and a  $m$  value of 2.0 kcal mol<sup>-1</sup> M<sup>-1</sup> (Figure 2c).

Sequential assignments were obtained for 90% of backbone hydrogens (Table 1).



**Figure 3.** (A) NMR assignments summary and NOE connectivities of PFRD-XC4. Bars represent unambiguous connectivities with bar thickness indexed to the intensity of the resonance. Coupling constants with  $^3J_{\text{HNHA}} > 8.0$  Hz are shown as open squares, and  $^3J_{\text{HNHA}} < 4.0$  Hz are shown as black squares. Secondary structural elements of PFRD are given at the bottom of the figure for comparison.(42) (B) Chemical shift difference between PFRD and PFRX-XC4 for amide nitrogen (black lines) and a hydrogens (shaded line). Proline residues 19, 25, 33, 39, and 44 not shown. Residues 40 and 41 also not shown due to a missing sequential assignments. Mutations are at positions 5, 8, 38, and 41. Residues 7, 9, 10, and 48 are in the close proximity of the mutated site.



On the basis of the chemical shifts for the amide and  $\alpha$  hydrogens, the PFRD-XC4 variant adopts a fold similar to that of PFRD (40). Chemical shift differences between PFRD and PFRD-XC4 are shown in Figure 3b. The chemical shifts of PFRD-XC4 residues that are distal from the site of mutation do not deviate significantly from those of PFRD, suggesting that these residues are in the same chemical environment as in PFRD. The chemical shifts of residues 7, 9, 10, and 48 significantly diverge from the corresponding PFRD chemical shifts. These residues are expected to have different chemical shifts because they are in close proximity to the mutated site. Lysine 45 does not make a direct contact with the mutated site but makes contact with residues 10 and 48. Examination of NOE sequential connections (Figure 3a) reveals strong HN-HA cross-peaks present in the  $\beta$ -sheet regions (residues 1-6, 10-14, and 49) and HN-HN cross-peaks in the turn regions (residues 6-10, 18-22, 29-32, and 46-47). Large coupling constants ( ${}^3J_{\text{HNHA}} > 8.0$  Hz) are also observed for most of the  $\beta$ -sheet residues (Figure 3a). These observations provide additional evidence that PFRD-XC4 adopts a fold similar to that of PFRD.

Given that PFRD-XC4 adopts a fold similar to that of PFRD and undergoes reversible unfolding, it should provide a tractable model system for determination of protein hyperthermostability. PFRD-XC4 should be amenable to mutagenesis studies aimed at addressing the contribution of salt bridges to protein stability using methods such as differential scanning calorimetry and chemical denaturation.

## Conclusion

We have successfully eliminated the iron center in hyperthermophilic rubredoxin and created a small protein that displays reversible folding. The PFRD-XC4 variant has a

melting temperature of 82 °C, a thermodynamic stability of 3.2 kcal mol<sup>-1</sup> at 1 °C, and NMR chemical shifts that are similar to those of the wild type protein. This variant should provide a tractable model system for studying the thermodynamic origins of protein hyperthermostability.

### **Acknowledgements**

We thank M. K. Eidness for the wild type *Pyrococcus furiosus* rubredoxin gene used in this study, S. M. Malakauskas, S. Ross, and C. Sarisky for technical assistance and discussions, and B. I. Dahiyat for discussions. The work was supported by the Howard Hughes Medical Institute (S.L.M.) and a NIH training grant (P.S.).

## References

1. Lee, B., and Vasmatzis, G. (1997) *Curr. Op. Biotech.* **8**, 423-428
2. Straume, M., Murphy, K. P., and Freire, E. (1992) in *Biocatalysis at Extreme Temperatures: Enzyme Systems Near and Above 100 °C* (Adams, M. W. W., and Kelly, R. M., eds) Vol. 498, pp. 122-135, American Chemical Society, Washington, D. C.
3. Rees, D. C., and Adams, M. W. W. (1995) *Structure* **3**, 251-254
4. Eidsness, M. K., Richie, K. A., Burden, A. E., Kurtz, J., D. M., and Scott, R. A. (1997) *Biochemistry* **36**, 10406-10413
5. Jiang, X., Bishop, E. J., and Farid, R. S. (1997) *J. Amer. Chem. Soc.* **119**, 838-839
6. Cavagnero, S., Zhou, Z., Adams, M., and Chan, S. (1998) *Biochemistry* **37**, 3377-3385
7. Klump, H. H., Adams, M. W. W., and Robb, F. T. (1994) *Pure & Applied Chem* **66**, 485-489
8. Wampler, J. E., and Neuhaus, E. B. (1997) *J Protein Chem* **16**, 721-32
9. Li, W., Grayling, A. R., Sandman, K., Edmondson, S., Shriver, W. J., and Reeve, N. J. (1998) *Biochemistry* , 10563-10572
10. MacBeath, G., Kast, P., and Hilvert, D. (1998) *Biochemistry* , 10062-10073
11. Hiller, R., Zhou, Z., Adams, M., and Englander, W. (1997) *Proc. Natl. Acad. Sci* , 11329-11332
12. Bradley, A. E., Stewart, E. D., Adams, W. M., and Wampler, J. E. (1993) *Protein Science* , 650-665
13. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer Jr, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T., and Tasumi, M. (1977) *Journal of Molecular Biology* **112**, 535-542
14. Adman, E. T., Sieker, L. C., and Jensen, L. H. (1991) *J. Mol. Biol* , 337-352
15. Dauter, Z., Sieker, L., and Wilson, K. (1992) *Acta Cryst* , 42-59
16. Day, M., Hsu, B., Joshua-tor, L., Park, J., Zhou, Z., Adams, M., and Rees, D. (1992) *Protein Science* , 1494-1507
17. Frey, M., Sieker, L. C., Payan, F., Haser, R., Bruschi, M., Pepe, G., and LeGall, J. (1987) *J. Mol. Biol* , 525-541
18. Sieker, L. C., Stenkamp, R. E., Jensen, L. H., Prickril, B., and LeGall, J. (1986) *FEBS* , 73-76
19. Watenpugh, K. D., Sieker, L. C., and Jensen, L. H. (1979) *J. Mol. Biol* , 509-522
20. Cavagnero, S., Debe, D., Zhou, Z., Adams, M., and Chan, S. (1998) *Biochemistry* , 3369-3376
21. Dahiyat, B. I., and Mayo, S. L. (1997) *Science* **278**, 82-87
22. Malakauskas, S. M., and Mayo, S. L. (1998) *Nature Structural Biology* **5**
23. Dahiyat, B. I., and Mayo, S. L. (1996) *Protein Science* **5**, 895-903
24. Dahiyat, B. I., Gordon, D. B., and Mayo, S. L. (1997) *Protein Sci.* **6**, 1333-1337
25. Dahiyat, B. I., and Mayo, S. L. (1997) *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10172-10177
26. Dahiyat, B. I., Sarisky, C. A., and Mayo, S. L. (1997) *J. Mol. Biol.* **273**, 789-796
27. Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G., and Galas, D. J. (1989) *Nucl. Acid. Res.* **17**, 6545-6551
28. Prodromou, C., and Pearl, L. H. (1992) *Protein Engineering* **5**, 827-829
29. Alexander, P., Fahnestock, S., Lee, T., Orban, J., and Bryan, P. (1992) *Biochemistry* **31**, 3597-
30. Johnson, B. H., and Hecht, M. H. (1994) *Bio/Technology* **12**, 1357-1360
31. Minor, D. L., and Kim, P. S. (1994) *Nature* **367**, 660-663
32. Santoro, M. M., and Bolen, D. W. (1988) *Biochemistry* **27**, 8063-8068
33. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York
34. Piantini, U., Sorensen, O. W., and Ernst, R. R. (1982) *J. Amer. Chem. Soc.* **104**, 6800-6801
35. Bax, A., and Davis, D. G. (1985) *J. Magn. Reson.* **65**, 355-360
36. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys.* **71**, 4546-4553
37. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR* **2**, 661-665
38. Kraulis, P. J. (1989) *J. Mag. Res.* **24**, 627-633
39. Bartels, C., Xia, T., Billeter, M., Guntert, P., and Wuthrich, K. (1995) *J. Bio. NMR* , 1-10
40. Blake, P. R., Park, J. B., Zhou, Z. H., Hare, D. R., Adams, M. W., and Summers, M. F. (1992) *Protein Sci* **1**, 1508-21.
41. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graph.* **14**, 51-55
42. Blake, P. R., Park, J. B., Bryant, F., Aono, S., Magnuson, J., Eccleston, E., Howard, J., Summers, M., and Adams, M. (1991) *Biochemistry*, 10885-10895

**Table 1.**  $^1\text{H}$  NMR Chemical Shifts for PFRD-XC4, pH 6.3, 25 °C<sup>a</sup>

	Residue	HN	HA	HB	Other
1	Ala				
2	Lys	8.68	5.45	1.62	HE,2.92
3	Trp	9.39	4.95	2.74	HD1,6.87;HE1,9.73;HE3,6.61;HZ2,7.58;HZ3,6.94;HH2,7.21
4	Val	9.97	5.57		HG1,0.78;HG2,0.93
5	Leu	8.85	4.35	0.16,1.91	HG,1.21;HD1,0.53;HD2,0.7
6	Lys	9.09	3.79		HG1,1.41;HG2,1.57;HE1,2.86;HE2,2.95;HD,1.71
7	Ile	8.25	4.09	1.88	
8	Thr	6.92	4.73	4.47	HG2,1.19
9	Gly	8.72	3.47,4.24		
10	Tyr	7.59	4.27	2.70,3.3	HD,7.33;HE,7.2
11	Ile	7.34	4.97	1.49	HG2,0.71;HD1,0.80;HG1,1.79
12	Tyr	9.53	4.64	2.92,3.17	HD,7.10;HE,6.43
13	Asp	8.78	4.92	2.26,2.86	
14	Glu	8.26	3.86	1.94,2.11	HG,3.03
15	Asp	8.03	4.27		
16	Ala	7.2	4.35	1.49	
17	Gly	8.12	3.59,3.98		
18	Asp	8.32	5.17	3.01	
19	Pro				
20	Asp	9.09	4.5	2.75	
21	Asn	7.73	5.19	2.69,3.16	
22	Gly	7.83	3.86,4.15		
23	Ile	7.67	4.38	1.94	HG2,0.91;HD1,0.82;HG1,1.34
24	Ser	8.72	4.61	3.74,3.9	
25	Pro				
26	Gly	8.22	3.74,4.25		
27	Thr	7.3	4.01	3.94	HG1,6.35;HG2,1.04
28	Lys	8.92	4.01	1.50,2.12	HD,1.69;HE,3.09
29	Phe	10.06	3.29		HZ,6.52;HD,5.70;HE,6.12
30	Glu	9.34	3.72	1.92,2.08	HG1,2.21;HG2,2.32
31	Glu	7.41	4.09	2.06,2.2	
32	Leu	6.86	3.88	0.34,0.88	HG,0.70;HD2,-0.14
33	Pro				
34	Asp	8.76	4.25	2.59,2.69	
35	Asp	8.23	4.53	2.58,2.87	
36	Trp	7.7	4.15	2.97	HD1,7.19;HE1,10.41;HE3,6.84;HZ2,7.49;HZ3,6.25;HH2,7.21
37	Val	6.72	4.32		HG1,1.33
38	Ala	8.49	1.28		
39	Pro		4.13	1.70,2.27	HD1,2.76;HD2,3.48;HG,1.9
40	Ile				
41	Thr				
42	Gly	7.62	3.64,3.86		
43	Ala	8.88	4.19		
44	Pro				
45	Lys	6.92		2.19,2.39	HG1,1.00;HG2,1.18;HD,1.7

46	Ser	8.12	4.38	4.03,4.19	
47	Glu	8.04	4.48	1.98,2.32	
48	Phe	8.16		4.94,3.76	
49	Glu	9.34	4.93		
50	Lys	8.55	3.2	1.12,1.28	HG1,0.07;HG2,0.22;HD,0.65;HE,2.68
51	Leu	8.68	4.3	1.29,1.48	HD,0.83
52	Glu	8.1	4.35	1.84	HG1,2.07;HG2,2.19
53	Asp	8.03	4.34	2.49,2.65	

<sup>a</sup>Chemical shifts were referenced to H<sub>2</sub>O (4.77 ppm at 25 °C)

**Chapter 7:**  
**NMR Structure of Rubredoxin Variant and Contribution of**  
**Surface Salt Bridges to Protein Stability**

Reprinted from Biochemistry

## Contribution of Surface Salt Bridges to Protein Stability<sup>†</sup>

Pavel Strop<sup>‡</sup> and Stephen L. Mayo<sup>\*,§</sup>

<sup>†</sup> This work was supported by the Howard Hughes Medical Institute (S.L.M.) and a NSF fellowship (P.S.).

<sup>‡</sup> Biochemistry Option, California Institute of Technology, mail code 147-75, Pasadena, CA 91125, USA.

<sup>§</sup> Howard Hughes Medical Institute and Division of Biology, California Institute of Technology, mail code 147-75, Pasadena, CA 91125, USA. E-mail: [steve@mayo.caltech.edu](mailto:steve@mayo.caltech.edu); Telephone: 626-396-6408; Fax: 626-568-0934

The coordinates for the structure have been deposited in the Protein Data Bank (accession code 1qcv).

**Abstract**

The role of surface salt bridges in protein stabilization has been a source of controversy. Here we present the NMR structure of a hyperthermophilic rubredoxin variant (PFRD-XC4) and the thermodynamic analysis of two surface salt bridges by double mutant cycles. This analysis shows that the surface side chain to side chain salt bridge between Lys 6 and Glu 49 does not stabilize PFRD-XC4. The main chain to side chain salt bridge between the N-terminus and Glu 14 was, however, found to stabilize PFRD-XC4 by 1.5 kcal mol<sup>-1</sup>. The entropic cost of making a surface salt bridge involving the protein's backbone is reduced, since the backbone has already been immobilized upon protein folding.



Proteins from thermophilic organisms offer good model systems to address the origins of thermostability. Several trends were found to be associated with increased thermal stability. These trends include increased packing density, increased core hydrophobicity, decreased length of surface loops, and improved secondary structure stabilization, such as N-terminal and C-terminal helix capping (1). Comparison of many thermophilic proteins also indicates a high occurrence of surface electrostatic interactions relative to mesophilic homologs (1-6). Buried salt bridges, salt bridges at dimer interfaces as well as surface salt bridges have been reported to contribute to the increased stability of proteins (7-11). These findings appear to contradict other thermodynamic evidence that surface ion pairs make little contribution to protein stability (12-14).

It has been difficult to study hyperthermophilic proteins by thermodynamic analysis, since most of the known hyperthermophilic proteins do not unfold reversibly (15-23). We previously designed a *Pyrococcus furiosus* rubredoxin variant (PFRD-XC4) to eliminate the iron-binding site using a computational design algorithm. The iron-binding site in wild type rubredoxin is thought to be the cause of its irreversible unfolding (22). The mutations for PFRD-XC4 are C5L, C8T, C38A, and C41T. PFRD-XC4 is able to fold in the absence of iron, undergoes reversible denaturation and has a melting temperature of 82 °C (24). The variant provides an excellent opportunity for systematic exploration of the factors determining protein thermostability.

The NMR solution structure of PFRD-XC4 reveals a fold similar to wild type *Pyrococcus furiosus* rubredoxin (PFRD). Given that PFRD-XC4 adopts a fold similar to PFRD and undergoes reversible unfolding, the importance of interactions responsible for the extreme stability of PFRD can now be addressed. Two interactions found in PFRD,

but not in mesophilic rubredoxins, have been proposed to contribute to its hyperthermostability (5, 25). These are the salt bridges between the N-terminus and the side chain of Glu 14 and between the side chains of Lys 6 and Glu 49. Here we present the NMR solution structure of PFRD-XC4 and thermodynamic analysis of the two salt bridges by double mutant cycles.

## Materials and Methods

*Mutagenesis and protein purification:* The hyperthermophilic variant PFRD-XC4 and all PFRD-XC4 mutants were constructed by inverse PCR (26) using a synthetic *Pyrococcus furiosus* rubredoxin gene in plasmid pt7-7 (27). The mutations for PFRD-XC4 are C5L, C8T, C38A, and C41T. All mutants were verified by sequencing. Recombinant proteins were expressed by IPTG induction in BL21(DE3) hosts (Invitrogen) as described (28) and isolated using a freeze/thaw method (29). Purification was accomplished by reverse-phase high-performance liquid chromatography using first a linear 1%/minute followed by a 0.07%/minute acetonitrile/water gradient containing 0.1% TFA. Molecular weights were verified by mass spectrometry.

*NMR structure determination:* NMR data were collected on a Varian Unity Plus 600 MHz spectrometer equipped with a Nalorac inverse probe with a self-shielded  $z$  gradient. NMR samples were prepared in 90/10 H<sub>2</sub>O/D<sub>2</sub>O or 99.9 % D<sub>2</sub>O with 200 mM NaCl and 25 mM sodium acetate-d<sub>3</sub> pH 6.3. Sample concentration was approximately

Table 1 NMR structure statistics

<b>Distance restraints</b>		
Intraresidue	162	
Sequentials	96	
Short range ( $ i-j  = 2-5$ residues)	80	
Long range ( $ i-j  > 5$ residues)	114	
Hydrogen bond	20	
Phi	22	
Total	494	
<b>Structural statistics</b>		
R.m.s. deviation	$\langle SA \rangle^1 \pm SD^2$	$(SA)_r^4$
Distance restraints (Å)	$0.031 \pm 0.003$	0.019
Idealized geometry		
Bonds (Å)	$0.0039 \pm 0.0002$	0.0019
Angles (degrees)	$0.64 \pm 0.02$	0.52
Impropers (degrees)	$0.44 \pm 0.02$	0.37
<b>Atomic r.m.s. deviations (Å)</b>		
	$\langle SA \rangle$ vs $SA^3 \pm SD$	
Backbone (1-51)	$1.02 \pm 0.05$	
Heavy atoms (1-51)	$1.42 \pm 0.04$	
Backbone (1-36,47-51)	$0.52 \pm 0.02$	
Heavy atoms (1-36,47-51)	$0.97 \pm 0.03$	
<b>Atomic r.m.s. deviations between PFRD and PFRD-XC4</b>		
Backbone (1-36, 47-51)	1.18	

<sup>1</sup> $\langle SA \rangle$  are 28 simulated annealing structures

<sup>2</sup>SD is the standard deviation

<sup>3</sup>SA is the average structure

<sup>4</sup> $(SA)_r$  is the minimized average structure

1.5 mM. Sequential assignment of resonances was achieved by standard homonuclear methods (30). Two-dimensional DQF-COSY (31), TOCSY (32), and NOESY (33) spectra were acquired at 25 °C. The TOCSY spectrum was recorded with an 80 ms mixing time using a clean-MLEV17 mixing sequence. NOESY spectra were acquired with a 150 ms mixing time for use during resonance assignments and 100 ms to derive distance restraints. Water suppression was accomplished either with presaturation during the relaxation delay or with pulsed field gradients (34). Spectra were processed with

VNMR (Varian Associates), and were assigned with ANSIG (35). NOEs were classified into three distance-bound ranges based on crosspeak intensity: strong (1.8 to 2.7 Å), medium (1.8 to 3.3 Å) and weak (1.8 to 5.0 Å). Upper bounds for restraints involving methyl protons were increased by 0.5 Å to account for the increased intensity of methyl resonances. Partially overlapped NOEs were set to weak restraints. Hydrogen bonds were defined based on the presence of supporting NOE peaks and by association with slowly exchanging backbone amide protons (protection factors  $\geq 100$ ). Distance restraints for the hydrogen bonds were  $r_{\text{NH-O}} = 1.5\text{-}2.8$  Å and  $r_{\text{N-O}} = 2.4\text{-}3.5$  Å. Dihedral restraints were defined based on the coupling constants. Residues with  $^3J_{\text{HNHA}} = 4$  Hz had the Phi torsional angle restricted to be within the interval of  $-80^\circ$  to  $-40^\circ$ . Residues with  $^3J_{\text{HNHA}} = 8$  Hz had the Phi torsion angle restricted to be within the interval of  $-160^\circ$  to  $-80^\circ$ .  $^3J_{\text{HNHA}}$  values were derived from NOESY crosspeak fine structure using the INFIT module of XEASY (36).

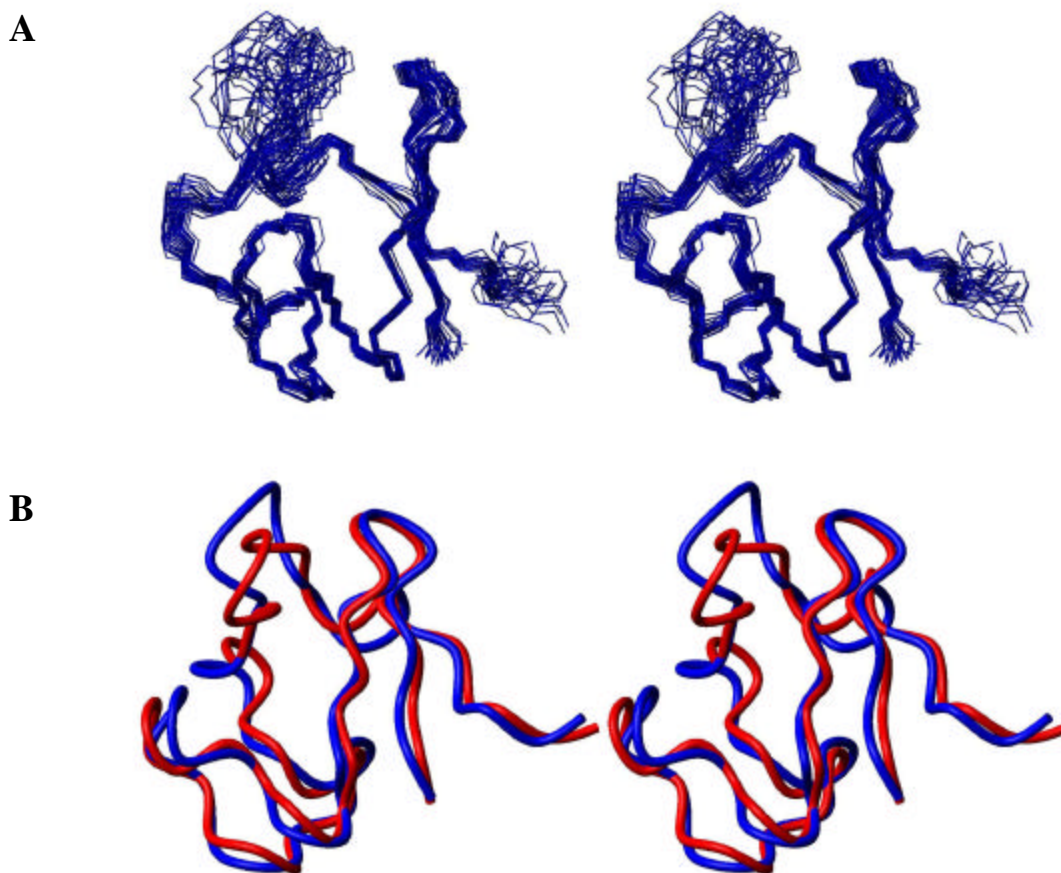
Structure calculations were performed with X-PLOR (37) with the use of standard protocols for hybrid distance geometry-simulated annealing (38-40). Forty structures were generated and following regularization and refinement, resulted in an ensemble of 28 structures with no restraint violations greater than 0.3 Å, r.m.s. deviations from idealized bond lengths less than 0.004 Å and r.m.s. deviations from idealized bond angles and impropers less than 0.65°.

*CD Analysis:* CD data were collected on an Aviv 62DS spectrometer equipped with a thermoelectric unit and using a 1 cm path length cell. Protein samples were 5  $\mu\text{M}$  in 50 mM sodium phosphate buffer at pH 6.3. Concentrations were determined by UV spectrophotometry. Thermal melts were monitored at 225 nm. Data were collected

every 2 °C with an equilibration time of 2 minutes and an averaging time of 10 s.  $T_m$  was determined by fitting the melting curves to a two-state model as described (41). Guanidinium chloride and urea denaturations were performed at 1 °C.  $\Delta G$ 's,  $m$ -values and error estimates were obtained by fitting the denaturation data to a two-state transition as described (42). GdmCl and urea concentrations were measured by refractometry.

## Results and Discussion

*Structure determination of PFRD-XC4:* To ensure that the elimination of the iron center did not change the tertiary fold of PFRD, we determined the solution structure of PFRD-XC4 by standard 2D  $^1\text{H}$  NMR methods (30-33). The structure of PFRD-XC4 was determined from 494 experimental restraints (8.3 restraints per residue) that were non-redundant with covalent structure including 452 NOE distance restraints, 22 dihedral angle restraints, and 20 hydrogen bond restraints (Table 1). Structure calculations were performed using X-PLOR (37) and standard protocols for hybrid distance geometry-simulated annealing (38-40). An ensemble of 28 structures converged with good covalent geometry and no distance restraint violations greater than 0.3 Å. The data reveal a well-ordered structure except for residues 37-46 with a root-mean-square (r.m.s.) deviation from the mean of 0.49 Å for the backbone of residues 1-36 and 47-51. Figure 1a shows the superposition of 28 simulated annealing structures generated using X-PLOR. Residues 37-46 form a loop that is the least ordered region of the structure. In PFRD the metal binding site is between the two loops formed by residues 5-10 and 37-46, each of which donates two cysteine residues to tetrahedrally coordinate the iron atom (Figure 2). The iron covalently links these two loops and prevents loop mobility. In the



**Figure 1:** Stereo view of the PFRD-XC4 structure ensemble and comparison to the PFRD X-ray structure. **a**, Best fit superposition of 28 simulated annealing structures. **b**, Structural comparison of the PFRD (red) and PFRD-XC4 (blue) backbones. The region of greatest deviation is the minimally restrained loop from residues 37-45. Structure figures were generated using MOLMOL (50).

PFRD-XC4 structure, one of these loops (residues 37-46) lacks NOE distance restraints and is probably present in multiple conformations.

The stereochemical quality of the ensemble of structures was examined with PROCHECK (43). Apart from the glycine residues, 64.2% of the residues are in the most favored regions of  $\phi, \psi$  space, 32.1% in the additional allowed regions, 3.3% in the generously allowed regions, and 0.4% in the disallowed regions. The average backbone

structure of PFRD-XC4 superimposes onto PFRD's crystallographic backbone with r.m.s. deviation of 1.2 Å for the backbone of residues 1-36 and 47-51 (Figure 1b).

Table 2 Thermal and chemical denaturation data

Protein	$T_m^1$ (°C)	$\Delta G_{GdmCl}^2$ kcal mol <sup>-1</sup>	$C_m^3$ (M)	m-value <sup>4</sup> kcal mol <sup>-1</sup> M <sup>-1</sup>	$\Delta G_{urea}^5$ kcal mol <sup>-1</sup>	$C_m^6$ (M)	m-value <sup>7</sup> kcal mol <sup>-1</sup> M <sup>-1</sup>
PFRD-XC4	82	3.2	1.68	1.9	4.6	5.7	0.81
PFRD-XC4-1Q	63	2.1	0.91	2.3	3.5	3.8	0.93
PFRD-XC4-14Q	57	1.7	0.63	2.5	2.7	3.0	0.91
PFRD-XC4-1Q14Q	39	1.5	0.50	3.0	2.1	2.0	1.03
PFRD-XC4-14T	38	-	-	-	1.3	1.3	0.98
PFRD-XC4-1Q14T	41	-	-	-	1.7	1.6	1.03
PFRD-XC4-6A	74	3.2	1.60	2.0	4.0	4.9	0.81
PFRD-XC4-49A	77	3.1	1.35	2.3	4.1	4.8	0.86
PFRD-XC4-6A49A	74	2.8	1.27	2.2	3.8	4.5	0.84

<sup>1</sup>Midpoint of thermal unfolding transition determined by CD

<sup>2</sup>Free energy of unfolding at 1 °C determined by GdmCl denaturations

<sup>3</sup>Midpoint of GdmCl unfolding transition

<sup>4</sup>Slope of  $\Delta G$  versus GdmCl concentration

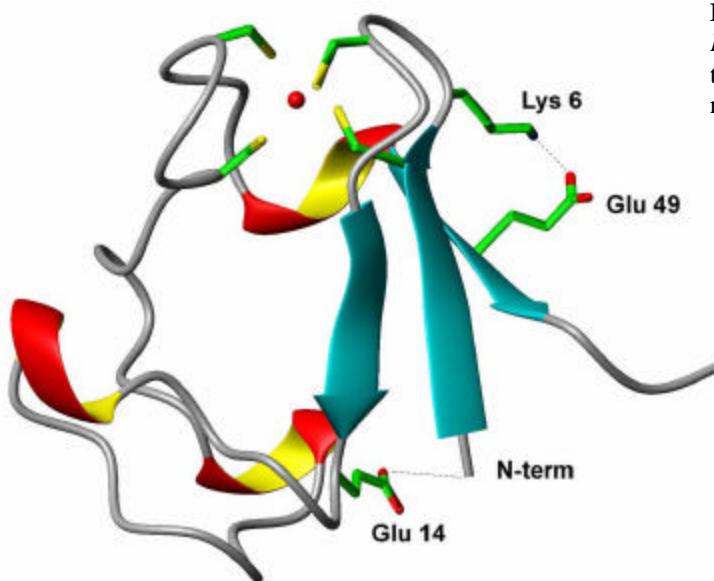
<sup>5</sup>Free energy of unfolding at 1 °C determined by urea denaturations

<sup>6</sup>Midpoint of urea unfolding transition

<sup>7</sup>Slope of  $\Delta G$  versus urea concentration

*Stability studies of PFRD-XC4 mutants:* The salt bridges of interest (between residues 1 and 14 and 6 and 49) are in the well-ordered part of the PFRD-XC4 structure and most likely adopt conformations similar to those in PFRD (Figure 2). A series of mutants were designed to disrupt these salt bridges. The resulting proteins were characterized by both thermal and chemical denaturation (Table 2).

The amino terminus of PFRD makes a main chain to side chain salt bridge with Glu 14. We used double mutant cycles to isolate the stabilization effect of this



**Figure 2:** Ribbon diagram of wild type *Pyrococcus furiosus* rubredoxin showing the two salt bridges and the four cysteine residues that coordinate the iron.

interaction (8, 9). The strength of interaction,  $\Delta G_{\text{interaction}}^{\text{XY}}$ , is calculated according to equation (1), where  $\Delta G^{\text{XY}}$ ,  $\Delta G^{\text{XA}}$ ,  $\Delta G^{\text{AY}}$ , and  $\Delta G^{\text{AA}}$  are the free energies of unfolding for the unmutated protein, the two single mutated proteins, and the doubly mutated protein, respectively.

$$\Delta G_{\text{interaction}}^{\text{XY}} = (\Delta G^{\text{XY}} - \Delta G^{\text{AA}}) - ((\Delta G^{\text{XA}} - \Delta G^{\text{AA}}) + (\Delta G^{\text{AY}} - \Delta G^{\text{AA}})) \quad (1)$$

In mesophilic rubredoxins, methionine is retained at the N-terminus in contrast to the thermophile. The extended N-terminus is too far from the side chain of Glu 14 to form a salt bridge. In order to extend the N-terminus of the hyperthermophilic variant PFRD-XC4, residue 1 was mutated to a large amino acid (Gln), resulting in the retention of the N-terminal methionine. Mutants 14Q and 1Q14Q were also prepared to complete the double mutant cycle. Gln was selected at position 14 to disrupt any ionic interaction with the N-terminus while preserving the potential hydrogen bond to the N-terminus.



The mutants 1Q, 14Q and 1Q14Q were folded at room temperature, with  $T_m$ 's reduced by 20-40 °C relative to PFRD-XC4 (Table 2).

Guanidinium chloride (GdmCl) chemical denaturations yielded free energies of unfolding of  $\Delta G_u(\text{PFRD-XC4}) = 3.2 \text{ kcal mol}^{-1}$ ,  $\Delta G_u(1\text{Q}) = 2.1 \text{ kcal mol}^{-1}$ ,  $\Delta G_u(14\text{Q}) = 1.7 \text{ kcal mol}^{-1}$ , and  $\Delta G_u(1\text{Q}14\text{Q}) = 1.5 \text{ kcal mol}^{-1}$  (Figure 2a). Due to the lack of well-defined pre-transition baselines in the guanidinium chloride denaturation curves for mutants 14Q and 1Q14Q, the denaturation data were fit to a two-state model with a fixed pre-transition slope. The pre-transition slope was determined by averaging the pre-transition slopes from 7 separate PFRD-XC4 mutants. Little variation in the pre-transition slopes was found among the 7 PFRD-XC4 mutants, resulting in  $\sim 0.1 \text{ kcal mol}^{-1}$  error in the final fit of 14Q and 1Q14Q. Based on the guanidinium chloride derived  $\Delta G_u$ 's of PFRD-XC4, 1Q, 14Q, and 1Q14Q (Table 2) the ionic component of the salt bridge between the N-terminus and Glu 14 stabilizes PFRD-XC4 by  $0.9 \text{ kcal mol}^{-1}$  (Table 3).

Urea denaturations yielded free energies of unfolding of  $\Delta G_u(\text{PFRD-XC4}) = 4.6 \text{ kcal mol}^{-1}$ ,  $\Delta G_u(1\text{Q}) = 3.5 \text{ kcal mol}^{-1}$ ,  $\Delta G_u(14\text{Q}) = 2.7 \text{ kcal mol}^{-1}$ , and  $\Delta G_u(1\text{Q}14\text{Q}) = 2.1 \text{ kcal mol}^{-1}$  (Table 2). The urea denaturation data suggest that the ionic component of the N-terminal salt bridge stabilizes PFRD-XC4 by  $0.5 \text{ kcal mol}^{-1}$  (Table 3).

Mutants 14T and 1Q14T were designed to eliminate both the ionic and hydrogen bond components of the salt bridge between the N-terminus and Glu 14. GdmCl denaturations were not possible due to the low stability of these mutants. Urea denaturations yielded free energies of unfolding of  $\Delta G_u(14\text{T}) = 1.3 \text{ kcal mol}^{-1}$ , and  $\Delta G_u(1\text{Q}14\text{T}) = 1.7 \text{ kcal mol}^{-1}$  (Figure 2b). Based on the urea denaturation data, the N-

terminal salt bridge between the N-terminus and Glu 14 stabilizes PFRD-XC4 by 1.5 kcal mol<sup>-1</sup>.

Table 3 Contributions of salt bridges at 1 °C

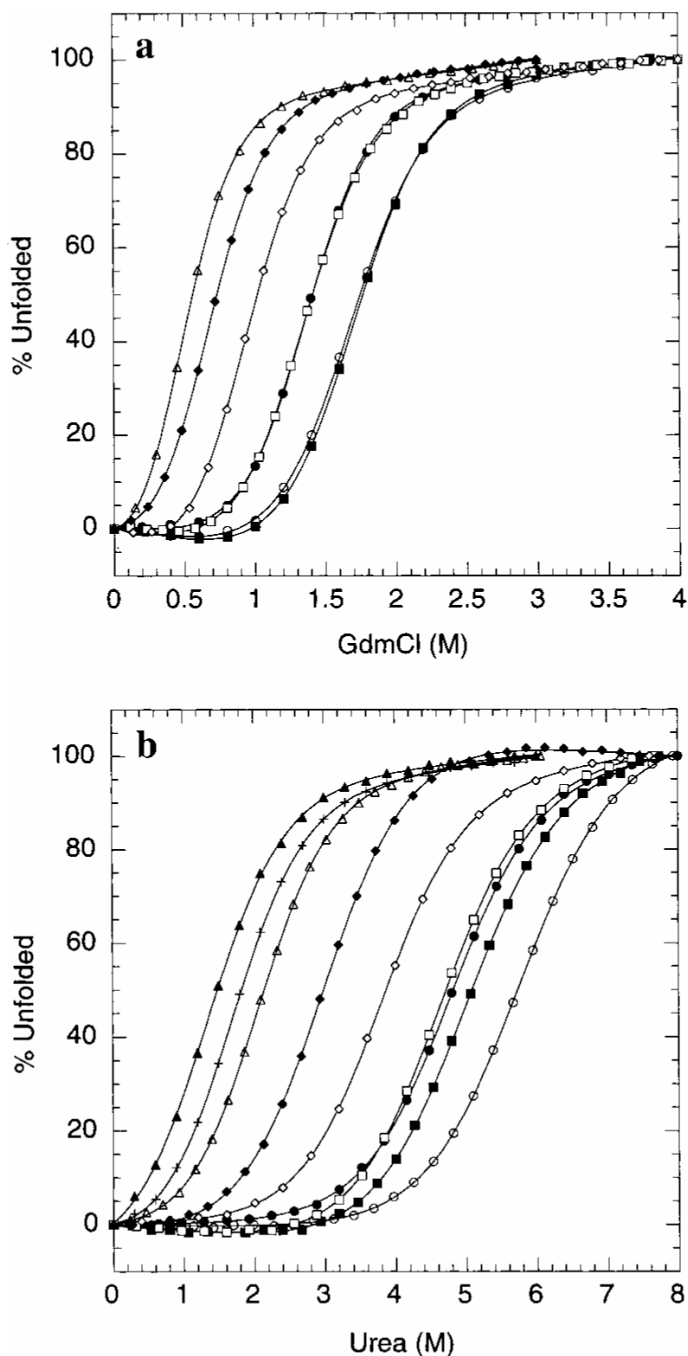
	Nt-E14 (ionic) <sup>1</sup> (kcal mol <sup>-1</sup> )	Nt-E14 (salt bridge) <sup>2</sup> (kcal mol <sup>-1</sup> )	K6-E49 <sup>3</sup> (kcal mol <sup>-1</sup> )
GdmCl	0.9	-	-0.3
Urea	0.5	1.5	0.3
Average	0.7	1.5	0.0

<sup>1</sup>Ionic component of the interaction between the N-terminus and the side chain of residue 14

<sup>2</sup>Total salt bridge interaction between the N-terminus and the side chain of residue 14

<sup>3</sup>Total salt bridge interaction between side chains of residues 6 and 49

The second salt bridge of interest in PFRD is formed by the side chains of Lys 6 and Glu 49. Mutants 6A, 49A and 6A49A were created to disrupt this interaction. 6A, 49A, and 6A49A have melting temperatures lowered by only 5-9 °C relative to PFRD-XC4. The guanidinium chloride denaturation data (Table 2) do not show significant thermodynamic destabilization relative to PFRD-XC4:  $\Delta G_u(\text{PFRD-XC4}) = 3.2 \text{ kcal mol}^{-1}$ ,  $\Delta G_u(6A) = 3.2 \text{ kcal mol}^{-1}$ ,  $\Delta G_u(49A) = 3.1 \text{ kcal mol}^{-1}$ , and  $\Delta G_u(6A49A) = 2.8 \text{ kcal mol}^{-1}$ . The experimentally obtained  $\Delta G_{\text{interaction}}$  for this salt bridge is  $-0.3 \text{ kcal mol}^{-1}$  (Table 3). These data suggest that the side chain to side chain salt bridge between Lys 6 and Glu 49 in rubredoxin might be slightly destabilizing.



**Figure 3:** Chemical denaturation curves for PFRD-XC4 (open circles), 1Q (open diamonds), 14Q (filled diamonds), 1Q14Q (open triangles), 14T (filled triangles), 1Q14T (pluses), 6A (filled squares), 49A (filled circles), and 6A49A (open squares) at 1°C in **a**, guanidinium chloride and **b**, urea. (In order to have the curves appear on the same plot, the initial and final points were arbitrary called 0 % and 100 % unfolded despite the fact, that proteins 14Q, 14T, 1Q14Q, and 1Q14T probably have unfolded population even at 0 % denaturant).

Urea denaturation data yield  $\Delta G_u(\text{PFRD-XC4}) = 4.6 \text{ kcal mol}^{-1}$ ,  $\Delta G_u(6\text{A}) = 4.0 \text{ kcal mol}^{-1}$ ,  $\Delta G_u(49\text{A}) = 4.1 \text{ kcal mol}^{-1}$ , and  $\Delta G_u(6\text{A}49\text{A}) = 3.8 \text{ kcal mol}^{-1}$  (Figure 2b).

The urea denaturation data suggest that the salt bridge between Lys 6 and Glu 49 slightly stabilizes PFRD-XC4 by  $0.3 \text{ kcal mol}^{-1}$ . Taken together, the chemical and thermal

denaturation data suggest that salt bridge between Lys 6 and Glu 49 has little effect on the stability of PFRD-XC4.

## Conclusions

Both theoretical and some experimental results suggest that forming a solvent exposed ion-pair may not be favorable, because in the process two ions have to be desolvated and immobilized (9, 12, 14, 44-47). The surface salt bridge between the side chains of Lys 6 and Glu 49, in which both side chains are immobilized, does not contribute to the overall stability of PFRD-XC4 at 1 °C and thus supports these findings. This salt bridge might play a role in setting the  $\beta$ -sheet register of rubredoxin or might be more stabilizing at higher temperatures (48). Previous reports have suggested that forming a salt bridge network might be stabilizing (2-4, 21, 47, 49). In ion pair network formation after the first two residues are desolvated and immobilized, only one additional side chain needs to be desolvated and immobilized upon salt bridge formation, lowering the energetic cost. Forming a main chain to side chain salt bridge between the N-terminus and the side chain of Glu 14 in PFRD-XC4 stabilizes PFRD-XC4 by 1.5 kcal mol<sup>-1</sup>. Main chain to side chain ion pairs may also have a lower formation cost, because the main chain is already fixed, and only one additional residue must be immobilized. In PFRD-XC4, the N-terminus participates in the  $\beta$ -sheet and is immobilized. The main chain in PFRD-XC4 has already paid the immobilization entropy cost upon protein folding, thus the energetic cost of forming a salt bridge to the N-terminus is significantly reduced in a manner similar to salt bridge formation in a surface ion pair network.

## **Acknowledgements**

We would like to thank M. K. Eidness for the wild-type *Pyrococcus furiosus* rubredoxin gene used in this study, S. Ross for NMR spectroscopy, and S. Ross and C. Sarisky for help with the NMR structure determination.

## References

1. Goldman, A. (1995) *Structure* 3, 1277-1279.
2. Yip, K. S., Stillman, T. J., Britton, K. L., Artymiuk, P. J., Baker, P. J., Sedelnikova, S. E., Engel, P. C., Pasquo, A., Chiaraluce, R., and Consalvi, V. (1995) *Structure* 3, 1147-58.
3. Hennig, M., Darimont, B., Sterner, R., Kirschner, K., and Jansonius, J. N. (1995) *Structure* 3, 1295-306.
4. Korndorfer, I., Steipe, B., Huber, R., Tomschy, A., and Jaenicke, R. (1995) *J. Mol. Biol.* 246, 511-21.
5. Day, M., Hsu, B., Joshua-tor, L., Park, J., Zhou, Z., Adams, M., and Rees, D. (1992) *Protein Science*, 1494-1507.
6. Perutz, M. F., and Raidt, H. (1975) *Nature* 255, 256-9.
7. Anderson, D. E., Becktel, W. J., and Dahlquist, F. W. (1990) *Biochemistry* 29, 2403-8.
8. Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A. R. (1990) *J. Mol. Biol.* 216, 1031-44.
9. Serrano, L., Horovitz, A., Avron, B., Bycroft, M., and Fersht, A. R. (1990) *Biochemistry* 29, 9343-52.
10. Kawamura, S., Tanaka, I., Yamasaki, N., and Kimura, M. (1997) *J. Biochem. (Tokyo)* 121, 448-55.
11. Vetriani, C., Maeder, D. L., Tolliday, N., Yip, K. S., Stillman, T. J., Britton, K. L., Rice, D. W., Klump, H. H., and Robb, F. T. (1998) *Proc. Natl. Acad. Sci.* 95, 12300-5.
12. Hendsch, Z. S., and Tidor, B. (1994) *Protein Science* 3, 211-26.
13. Honig, B., and Nicholls, A. (1995) *Science* 268, 1144-1149.
14. Matthews, B. W. (1993) *Annu Rev Biochem* 62, 139-60.
15. Klump, H., Di Ruggiero, J., Kessel, M., Park, J. B., Adams, M. W., and Robb, F. T. (1992) *J. Biol. Chem.* 267, 22681-5.
16. Wrba, A., Schweiger, A., Schultes, V., Jaenicke, R., and Zavodsky, P. (1990) *Biochemistry* 29, 7584-92.
17. Laderman, K. A., Davis, B. R., Krutzsch, H. C., Lewis, M. S., Griko, Y. V., Privalov, P. L., and Anfinsen, C. B. (1993) *J. Biol. Chem.* 268, 24394-401.
18. DeDecker, B. S., O'Brien, R., Fleming, P. J., Geiger, J. H., Jackson, S. P., and Sigler, P. B. (1996) *J. Mol. Biol.* 264, 1072-84.
19. Pfeil, W., Gesierich, U., Kleemann, G. R., and Sterner, R. (1997) *J. Mol. Biol.* 272, 591-6.
20. Ogasahara, K., Lapshina, E. A., Sakai, M., Izu, Y., Tsunasawa, S., Kato, I., and Yutani, K. (1998) *Biochemistry* 37, 5939-46.
21. Pappenberger, G., Schurig, H., and Jaenicke, R. (1997) *J. Mol. Biol.* 274, 676-83.
22. Cavagnero, S., Zhou, Z., Adams, M., and Chan, S. (1998) *Biochemistry* 37, 3377-3385.
23. Beadle, B. M., Baase, W. A., Wilson, D. B., Gilkes, N. R., and Shoichet, B. K. (1999) *Biochemistry* 38, 2570-6.
24. Strop, P., and Mayo, S. L. (1999) *J. Amer. Chem. Soc.* 121, 2341-2345.
25. Bau, R., Rees, D. C., Kurtz, D. M., Scott, R. A., Huang, H., Adams, M. W. W., and Eidsness, M. K. (1998) *J. Bio. Inor. Chem.* 3, 484-493.
26. Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G., and Galas, D. J. (1989) *Nucl. Acid. Res.* 17, 6545-6551.
27. Eidsness, M. K., Richie, K. A., Burden, A. E., Kurtz, J., D. M., and Scott, R. A. (1997) *Biochemistry* 36, 10406-10413.
28. Alexander, P., Fahnestock, S., Lee, T., Orban, J., and Bryan, P. (1992) *Biochemistry* 31, 3597-3603.
29. Johnson, B. H., and Hecht, M. H. (1994) *Bio/Technology* 12, 1357-1360.
30. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
31. Piantini, U., Sorensen, O. W., and Ernst, R. R. (1982) *J. Amer. Chem. Soc.* 104, 6800-6801.
32. Bax, A., and Davis, D. G. (1985) *J. Magn. Reson.* 65, 355-360.
33. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) *J. Chem. Phys.* 71, 4546-4553.
34. Piotto, M., Saudek, V., and Sklenar, V. (1992) *J. Biomol. NMR* 2, 661-665.

35. Kraulis, P. J. (1989) *J. Mag. Res.* 24, 627-633.
36. Bartels, C., Xia, T., Billeter, M., Guntert, P., and Wuthrich, K. (1995) *J. Bio. NMR*, 1-10.
37. Brünger, A. T. (1992) *X-PLOR, version 3.1, a system for X-ray crystallography and NMR*, Yale Univ. Press, New Haven, CT.
38. Nilges, M., Clore, G. M., and Gronenborn, A. M. (1988) *FEBS Letters* 229, 317-324.
39. Nilges, M., Kuszewski, J., and Brünger, A. T. (1991) in *Computational Aspects of the Study of Biological Macromolecules by NMR* (Hoch, J. C., Poulsen, F. M., and Redfield, C., Eds.) pp 451-457, Plenum Press, New York.
40. Kuszewski, J., Nilges, M., and Brünger, A. T. (1992) *J. Bio. NMR* 2, 33-56.
41. Minor, D. L., and Kim, P. S. (1994) *Nature* 367, 660-663.
42. Santoro, M. M., and Bolen, D. W. (1988) *Biochemistry* 27, 8063-8068.
43. Laskowski, R. A., Macarthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Cryst.* 26, 283.
44. Sun, D. P., Sauer, U., Nicholson, H., and Matthews, B. W. (1991) *Biochemistry* 30, 7142-53.
45. Meeker, A. K., Garcia-Moreno, B., and Shortle, D. (1996) *Biochemistry* 35, 6443-6449.
46. Matthews, B. W. (1993) *Curr. Op. Struct. Biol.* 3, 589-593.
47. Spek, E. J., Bui, A. H., Lu, M., and Kallenbach, N. R. (1998) *Protein Science* 7, 2431-2437.
48. Elcock, A. H. (1998) *J. Mol. Biol.* 284, 489-502.
49. Marqusee, S., and Sauer, R. T. (1994) *Protein Science* 3, 2217-25.
50. Koradi, R., Billeter, M., and Wüthrich, K. (1996) *J. Mol. Graph.* 14, 51-55.

**Chapter 8:**  
**Structure of a Protein G Helix Variant Suggests the**  
**Importance of Helix Propensity and Helix Dipole Interactions**  
**in Protein Design**

Reprinted from Protein Science



**Structure of a Protein G helix variant suggests the importance of helix propensity and helix dipole interactions in protein design.**

Pavel Strop<sup>1</sup>, Andrei M. Marinescu<sup>2</sup>, and Stephen L. Mayo<sup>3,\*</sup>

<sup>1</sup>Biochemistry Option, California Institute of Technology, mail code 147-75, Pasadena, California 91125, USA.

<sup>2</sup>Division of Biology, California Institute of Technology, mail code 147-75, Pasadena, California 91125, USA.

<sup>3</sup>Howard Hughes Medical Institute and Division of Biology, California Institute of Technology, mail code 147-75, Pasadena, California 91125, USA.

\*To whom correspondence should be addressed. E-mail: [steve@mayo.caltech.edu](mailto:steve@mayo.caltech.edu); phone 626-395-6407; fax 626-568-0934.

Running title: Helix surface design

**Abstract:**

Six helix surface positions of protein G (G $\beta$ 1) were redesigned using a computational protein design algorithm, resulting in the five fold mutant G $\beta$ 1m2. G $\beta$ 1m2 is well folded with a circular dichroism spectrum nearly identical to that of G $\beta$ 1, and a melting temperature of 91 °C, approximately 6 °C higher than that of G $\beta$ 1. The crystal structure of G $\beta$ 1m2 was solved to 2.0 Å resolution by molecular replacement. The absence of hydrogen bond or salt bridge interactions between the designed residues in G $\beta$ 1m2 suggests that the increased stability of G $\beta$ 1m2 is due to increased helix propensity and more favorable helix dipole interactions.

While many surface sites in proteins can tolerate a wide variety of side chains, surface residues can still have a significant effect on protein structure and stability (Bowie et al., 1990, Minor & Kim, 1994, O'Neil & DeGrado, 1990, Reidhaar-Olson & Sauer, 1988, Smith et al., 1994, Zhang et al., 1991, Zhao & Arnold, 1999). Previous computational protein design efforts on the surface of GCN4, a homodimeric coiled coil, suggest that helix propensity is the key factor in sequence design for surface helical positions (Dahiyat et al., 1997). The designed GCN4 sequence was randomly scrambled, thereby disrupting the designed hydrogen bond and salt bridge interactions, but not changing the overall helix propensity. Both the designed and the scrambled sequences were 12 °C more stable than the wild type sequence. The increased stability was not due to the designed interactions between the surface residues, but apparently due to increased helical propensity. Results from this GCN4 study suggest that for helix surface design, hydrogen bonding and electrostatic interactions are not as important as helix propensity and helix capping (Blaber et al., 1993, Chakrabartty et al., 1993, Chakrabartty et al., 1994, Doig et al., 1994, Huyghues-Despointes et al., 1993, O'Neil & DeGrado, 1990).

Protein G (G $\beta$ 1) consists of a four-stranded  $\beta$ -sheet traversed by an  $\alpha$ -helix. Protein G is a monomeric globular 56 residue protein that contains no disulfide bonds or cofactors and lacks the repetitive side chain patterns found in coiled coils and helical bundles. These properties make protein G an ideal model system for testing the importance of specific helical surface design in globular proteins.

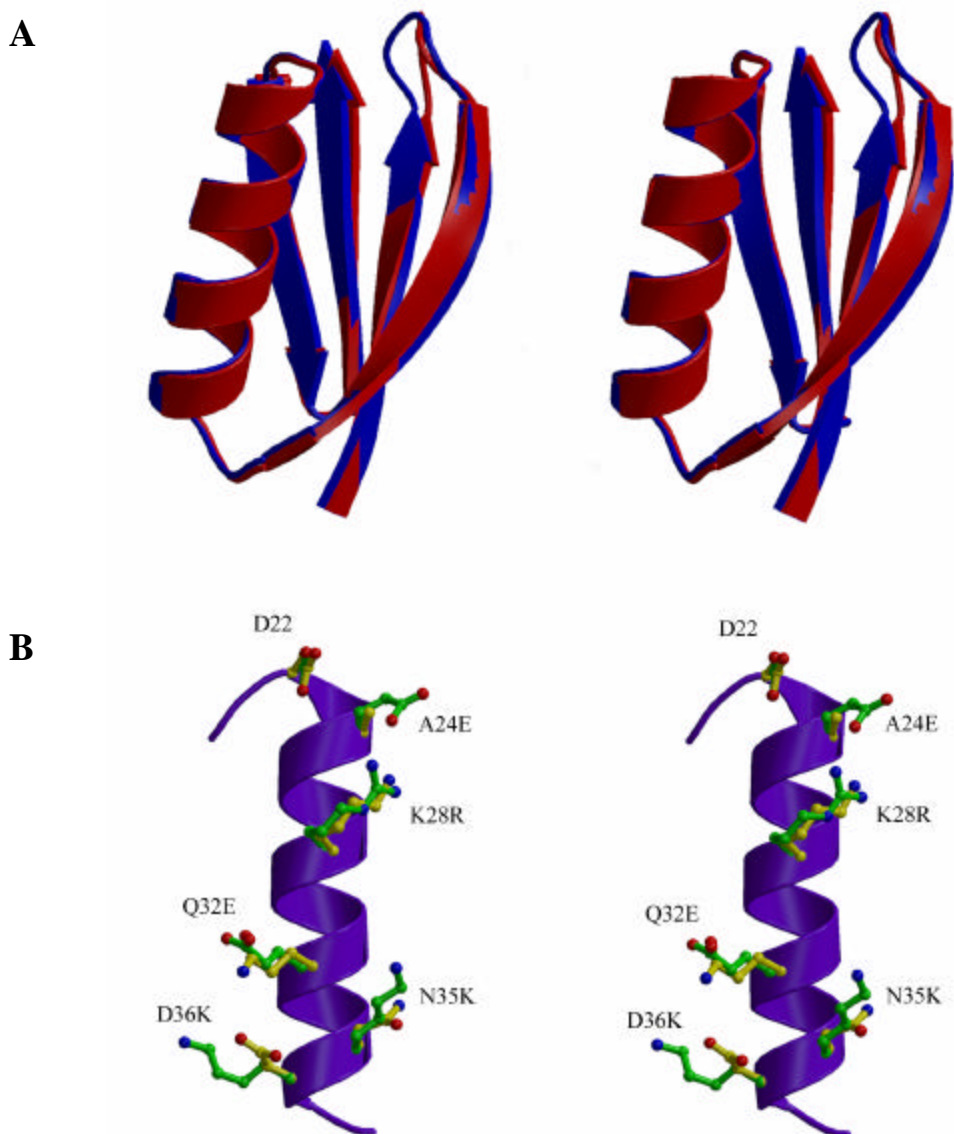
Six residues on the helix surface of protein G were designed using a computational design algorithm called ORBIT (Optimization of Rotamers By Iterative

Techniques) and yielded a variant called G $\beta$ 1m2. The structure of G $\beta$ 1m2 variant was determined by X-ray crystallography in order to assess the performance of the design algorithm and to examine the origins of helix surface stabilization.

## Results and Discussion

G $\beta$ 1m2 contains the following five mutations: A24E, K28R, Q32E, N35K, and D36K (position 22 retained its wild type identity). G $\beta$ 1m2 is well folded with a far-ultraviolet circular dichroism (CD) spectrum nearly identical to the wild type protein G at 1 °C; thermal denaturation of this variant yields a melting temperature ( $T_m$ ) of 91 °C, approximately 6 °C higher than the wild type protein (data not shown).

The backbone of G $\beta$ 1m2, as determined from X-ray crystallographic analysis, superimposes onto the G $\beta$ 1 backbone with an r.m.s. deviation of 0.92 Å (Figure 1a). The backbone of the helix alone (residues 22-37) superimposes onto the helix of G $\beta$ 1 with an r.m.s. deviation of 0.17 Å. Residue D22 is a helix capping position. In the G $\beta$ 1m2 structure, D22 makes hydrogen bonds to both the backbone and side chain of T25, as in the wild type structure (Figure 1b). In the predicted structure of the designed sequence, R28 makes a putative salt bridge to E24. Poor electron density for R28 suggests either low occupancy or the absence of this salt bridge in the G $\beta$ 1m2 structure. Side chains of residues E32 and K35 are mostly disordered, and do not appear to form the predicted E32-K35 interaction. Residues E24 and K36 are well ordered and make hydrogen bonds through water molecules to symmetry related molecules and not to their designed partners (R28 and E32, respectively).



**Figure 1.** Stereoview of the superposition of the C $\beta$ 1 and C $\beta$ 1m2 X-ray structures. (A) Structural comparison of the C $\beta$ 1 (blue) and C $\beta$ 1m2 (red) backbones. (B) Superimposition of the redesigned helix surface residues. Wild type residues shown in yellow and C $\beta$ 1m2 residues shown in green. The side chains of residues R28, E32, and K35 in C $\beta$ 1m2 are mostly disordered. Figures were generated using BOBSCRIPT and RASTER3D (Esnouf, 1997, Merritt & Bacon, 1997).

These results suggest that the stabilization of the G $\beta$ 1m2 mutant is not due to the increased number of hydrogen bonds or salt bridges on the helix surface, as predicted by the design algorithm. However, the amino acids at four of the five helix surface positions

in the designed sequence have higher helical propensity than the wild type sequence. The mutation to a more negatively charged residue at the N-terminus of the helix (A24E), as well as mutations to more positively charged residues at the C-terminus of the helix (N35K and D36K), could also result in more favorable interactions with the helix dipole. Electrostatic interactions with helix dipoles have been shown to be favorable and stabilizing in proteins (Creamer & Rose, 1994, Doig et al., 1994, Huyghues-Despointes et al., 1993, Lockhart & Kim, 1993, Lockhart & Kim, 1992, Nicholson et al., 1991, Nicholson et al., 1988, Walter et al., 1995). The increased stability of G $\beta$ 1m2 relative to G $\beta$ 1, together with the absence of hydrogen bonds or salt bridges between the designed residues, suggests that the stabilization is due to increased helix propensity and more favorable helix dipole interactions.

## Materials and Methods

*Computational design:* In the calculation, protein residues were classified into core, boundary and surface as described (Dahiyat & Mayo, 1997). Six residues were classified as helix surface residues (22, 24, 28, 32, 35, and 36) and were restricted to the following ten amino acids: Ala, Arg, Asn, Asp, Gln, Glu, His, Lys, Ser, Thr. The resulting combinatorial complexity is  $10^6$  possible amino acid sequences. Protonated histidine was used in the calculation. Both amino acid identities and side chain conformations were determined by the design algorithm for the six optimized residues. The remainder of the structure was used as a template for the sequence selection calculation. Computational details, potential functions, and parameters for van der Waals interactions, hydrogen bonding, and electrostatic interactions are described in our

**Table 1.** Refinement and Model Statistics

Space group		P4 <sub>2</sub> 2
Cell dimensions		
a, c (Å)		52.61, 44.59
Resolution (Å)		18.0 – 2.0
Measured reflections		81792
Unique reflections		4020
R <sub>merge</sub>		4.1 (9.1)
Completeness (%)		95.0 (98.0)
Number of p protein atoms	439	
Number of water molecules		29
B factor values (Å <sup>2</sup> )		
Overall		29.8
protein atoms		29.6
solvent		38.6
R <sub>cryst</sub>		0.235
R <sub>free</sub>		0.267
Root mean square deviation		
Bonds		0.016
Angles		1.74
Ramachandran		
most favored regions		96.0
additional allowed regions	4.0	
generously allowed regions		0.0
disallowed regions		0.0

---

$R_{\text{merge}} = \frac{\sum |I_{\text{obs}} - \langle I \rangle|}{\sum \langle I \rangle}$  summed over all observations and reflections  
 $R_{\text{cryst}} = \frac{\sum |F_{\text{obs}} - F_{\text{calc}}|}{\sum |F_{\text{obs}}|}$   
 $R_{\text{free}} = R_{\text{cryst}}$  calculated for 5% of reflections omitted from the refinement

previous work (Dahiyat et al., 1997, Dahiyat & Mayo, 1997, Dahiyat & Mayo, 1997, Dahiyat & Mayo, 1996, Dahiyat et al., 1997).

*Expression and CD analysis:* Variant Gβ1m2 was constructed by inverse PCR (Hemsley et al., 1989) in plasmid pet11c and verified by sequencing. Recombinant protein was expressed by IPTG induction in BL21(DE3) cells (Invitrogen) as described (Alexander et al., 1992) and isolated using a freeze/thaw method (Johnson & Hecht, 1994). Purification was accomplished by reverse-phase high-performance liquid chromatography using an acetonitrile/water gradient containing 0.1 % TFA. Molecular

weight was verified by mass spectrometry. CD data were collected on an Aviv 62DS spectrometer equipped with a thermoelectric unit. Protein samples were studied in 50 mM sodium phosphate buffer at pH 5.0. Wavelength scans were performed at 1 °C. Thermal denaturations were monitored by CD at 222 nm, and fitted to a two-state model as described (Minor & Kim, 1994).

*Crystallization, data collection and refinement:* G $\beta$ 1m2 crystals were grown by the sitting drop method at 4 °C using a precipitant solution containing 100 mM sodium acetate, pH 4.6, 30% polyethylene glycol monomethyl ether 2000, and 200 mM ammonium sulfate. The crystals belong to the space group P4<sub>2</sub>1<sub>2</sub> with unit cell dimensions  $a = b = 52.61 \text{ \AA}$ ,  $c = 44.59 \text{ \AA}$ . Crystals were transferred stepwise to mother solution containing 20 % glycerol and flash frozen. Diffraction data were collected on a R-axis IV image plate detector with CuK $_{\alpha}$  radiation. All data were processed with DENZO and SCALEPACK (Otwinowski & Minor, 1997). Wild type protein G, PDB code 1PGA (Gallagher et al., 1994), was used for molecular replacement with the CCP4 program suit (Bailey, 1994) yielding an initial R factor of 38.3% at 3  $\text{\AA}$ . Alternate cycles of manual model building using the program O (Jones et al., 1991), positional and individual B-factor refinement with the program CNS (Brünger, 1992), and addition of water molecules to the model reduced the R and R-free to 23.5 and 26.7, respectively, for all of the reflections (2.0  $\text{\AA}$  resolution). A total of 96% of the residues in G $\beta$ 1m2 are in the most favored regions of the Ramachandram plot, 4% in the additional allowed regions as calculated with PROCHECK (Laskowski et al., 1993). The coordinates for the structure have been deposited in the Protein Data Bank (pdb code 1EM7).



**Acknowledgements:**

This work was supported by the Howard Hughes Medical Institute (S.L.M.) and a NSF fellowship (P.S.).

## References

- Alexander P, Fahnestock S, Lee T, Orban J, and Bryan P. 1992. Thermodynamic analysis of the folding of the streptococcal protein G IgG-binding domains  $\beta 1$  and  $\beta 2$  -Why small proteins tend to have high denaturation temperatures. *Biochemistry* (31):3597-3603.
- Bailey S. 1994. The Ccp4 Suite - Programs For Protein Crystallography. *Acta Crystallogr (D50)*:760-763.
- Blaber M, Zhang XJ, and Matthews BW. 1993. Structural basis of amino acid  $\alpha$  helix propensity. *Science* (260):1637-1640.
- Bowie JU, Reidhaar-Olson JF, Lim WA, and Sauer RT. 1990. Deciphering the message in protein sequences: tolerance to amino acid substitutions. *Science* (247):1306-1310.
- Brünger AT (1992) *X-PLOR, version 3.1, a system for X-ray crystallography and NMR*, Vol. Yale Univ. Press, New Haven, CT.
- Chakrabartty A, Doig AJ, and Baldwin RL. 1993. Helix capping propensities in peptides parallel those in proteins [see comments]. *Proc Natl Acad Sci U S A* (90):11332-6.
- Chakrabartty A, Kortemme T, and Baldwin RL. 1994. Helix propensities of the amino-acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Science* (3):843-852.
- Creamer TP, and Rose GD. 1994. Alpha-helix-forming propensities in peptides and proteins. *Proteins* (19):85-97.
- Dahiyat BI, Gordon DB, and Mayo SL. 1997. Automated design of the surface positions of protein helices. *Protein Sci* (6):1333-1337.
- Dahiyat BI, and Mayo SL. 1997. De novo protein design: Fully automated sequence selection. *Science* (278):82-87.
- Dahiyat BI, and Mayo SL. 1997. Probing the role of packing specificity in protein design. *Proc Natl Acad Sci USA* (94):10172-10177.
- Dahiyat BI, and Mayo SL. 1996. Protein design automation. *Protein Science* (5):895-903.
- Dahiyat BI, Sarisky CA, and Mayo SL. 1997. De novo protein design: towards fully automated sequence selection. *J Mol Biol* (273):789-796.
- Doig AJ, Chakrabartty A, Klingler TM, and Baldwin RL. 1994. Determination of free energies of N-capping in alpha-helices by modification of the Lifson-Roig helix-coil theory to include N- and C-capping. *Biochemistry* (33):3396-403.
- Esnouf RM. 1997. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J Molec Graphics* (15):133-138.
- Gallagher T, Alexander P, Bryan P, and Gilliland GL. 1994. 2 crystal-structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* (33):4721-4729.
- Hemsley A, Arnheim N, Toney MD, Cortopassi G, and Galas DJ. 1989. A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucl Acid Res* (17):6545-6551.
- Huyghues-Despointes BMP, Scholtz JM, and Baldwin R. 1993. Effect of a single aspartate on helix stability at different positions in a neutral alanine-based peptide. *Protein Science* (2):16604-1611.
- Johnson BH, and Hecht MH. 1994. Recombinant proteins can be isolated from E. coli cells by repeated cycles of freezing and thawing. *Bio/Technology* (12):1357-1360.
- Jones TA, Zou JY, Cowan SW, and Kjeldgaard M. 1991. Improved Methods For Building Protein Models in Electron- Density Maps and the Location of Errors in These Models. *Acta Crystallogr (A47)*:110-119.
- Laskowski RA, Macarthur MW, Moss DS, and Thornton JM. 1993. Procheck - a program to check the stereochemical quality of protein structures. *J Appl Cryst* (26):283.
- Lockhart DJ, and Kim PS. 1993. Electrostatic screening of charge and dipole interactions with the helix backbone. *Science* (260):198-202.
- Lockhart DJ, and Kim PS. 1992. Internal Stark effect measurement of the electric field at the amino terminus of an alpha helix. *Science* (257):947-51.
- Merritt EA, and Bacon DJ. 1997. Raster3D Photorealistic Molecular Graphics. *Methods Enzymol* (277):505-524.

- Minor DL, and Kim PS. 1994. Measurement of the  $\beta$ -sheet-forming propensities of amino acids. *Nature* (367):660-663.
- Nicholson H, Anderson DE, Dao-pin S, and Matthews BW. 1991. Analysis of the interaction between charged side chains and the alpha-helix dipole using designed thermostable mutants of phage T4 lysozyme. *Biochemistry* (30):9816-28.
- Nicholson H, Becktel WJ, and Matthews BW. 1988. Enhanced protein thermostability from designed mutations that interact with alpha-helix dipoles. *Nature* (336):651-656.
- O'Neil KT, and DeGrado WF. 1990. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science* (250):646-651.
- Otwinowski Z, and Minor W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* (276):307-326.
- Reidhaar-Olson JF, and Sauer RT. 1988. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* (241):53-57.
- Smith CK, Withka JM, and Regan L. 1994. A thermodynamic scale for the  $\beta$ -sheet forming tendencies of the amino acids. *Biochemistry* (33):5510-5517.
- Walter S, Hubner B, Hahn U, and Schmid FX. 1995. Destabilization of a protein helix by electrostatic interactions. *J Mol Biol* (252):133-43.
- Zhang XJ, Baase WA, and Matthews BW. 1991. Toward a simplification of the protein folding problem: a stabilizing polyalanine  $\alpha$  helix engineered in T4 lysozyme. *Biochemistry* (30):2012-2017.
- Zhao H, and Arnold FH. 1999. Directed evolution converts subtilisin E into a functional equivalent of thermitase. *Protein Eng* (12):47-53.