Structural Characterization of Proteins from the E. coli β-barrel Assembly Machine using NMR Spectroscopy

Lisa Rose Warner

University of Colorado at Boulder, lisa.warner@colorado.edu

Follow this and additional works at: https://scholar.colorado.edu/chem_gradetds

Part of the Biochemistry Commons

Recommended Citation
https://scholar.colorado.edu/chem_gradetds/128

This Dissertation is brought to you for free and open access by Chemistry & Biochemistry at CU Scholar. It has been accepted for inclusion in Chemistry & Biochemistry Graduate Theses & Dissertations by an authorized administrator of CU Scholar. For more information, please contact cuscholaradmin@colorado.edu.
Structural characterization of proteins from the *E. coli* β-barrel assembly machine using NMR spectroscopy

by

Lisa Rose Warner

B.S., Boise State University, 2002

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy

Biochemistry

2011
This thesis entitled:

Structural characterization of proteins from the *E. coli* β-barrel assembly machine using NMR spectroscopy

Written by Lisa Rose Warner

has been approved for the Department of Biochemistry

__________________________________________

Arthur Pardi

__________________________________________

Marcelo C. Sousa

Date ________________

The final copy of this thesis has been examined by both the signatories, and we find that both the content and form meet the acceptable presentation standards of scholarly work in the above mentioned discipline.
Warner, Lisa Rose (Ph.D. Biochemistry)

“Structural characterization of proteins from the *E. coli* β-barrel assembly machine using NMR spectroscopy”

Thesis directed by Professor Arthur Pardi

The cytosol of Gram-negative bacteria is fortified by an inner membrane and an outer membrane that are separated by ~140 Å of periplasmic space. Spanning the outer membrane are β-barrel proteins that are integral for basic physiological functions, virulence and multidrug resistance. The *E. coli* β-barrel assembly machine is a five protein complex responsible for the folding and insertion of β-barrel proteins into the outer membrane. How this molecular machine works in the absence of ATP or a proton gradient has remained elusive. The prevailing theory is that interactions between proteins in the complex, protein chaperones in the periplasm, and outer membrane protein precursors drive this molecular machine. The work presented here contributes to understanding the mechanics of the β-barrel assembly machine by providing a structural description of two of the components: BamC and BamA.

NMR is a powerful tool to investigate the structure and dynamics of proteins in solution. Here, a novel approach was used to determine the solutions structure of the 27 kDa BamC by combining a limited NMR dataset of chemical shifts, residual dipolar couplings (RDCs) and nuclear Overhauser effect (NOE) distance restraints with the protein fold prediction program, Rosetta. The structure of BamC was determined to consist of two helix-grip type domains connected by an ~18-residue flexible linker. The structure was validated with a supplementary NOE NMR dataset including amide-amide 1H-1H and isoleucine, leucine and valine methyl-
methyl $^1$H-$^1$H NOEs. Interestingly, regions of the structural ensemble that did not converge to a unique conformation also showed increased $^{15}$N backbone amide dynamics.

The domain orientation and flexibility of the periplasmic POlypeptide TRansport Associated (POTRA) domains of BamA were investigated. Solution domain orientation using RDCs validated the orientation of POTRA4–5 in a spliced crystallographic model of POTRA1–5. The flexibility of POTRA1–5 in solution was assessed with $^{15}$N amide backbone dynamic, paramagnetic relaxation enhancement and analysis of RDCs. Previous reports suggested that POTRA1–5 was comprised of two rigid units, POTRA1–2 and POTRA3–5. The results presented here indicate that POTRA1–5 may be more flexible than previously thought.
## CONTENTS

Chapter 1  Introduction .................................................................................................................. 1

1.1  The outer membrane of Gram-negative bacteria ................................................................... 1

1.1.1  The role of the outer membrane in cell homeostasis ......................................................... 1

1.1.2  The β-barrel assembly machine of *E. coli* ....................................................................... 2

1.1.3  Folding OMPs by β-augmentation .................................................................................. 3

1.1.4  The accessory lipoproteins ............................................................................................. 4

1.2  Determining solution structures of proteins with NMR ....................................................... 7

1.3  Determining domain orientation with residual dipolar couplings ...................................... 9

1.4  Overview ............................................................................................................................... 13

Chapter 2  Determining the solution structure of BamC ............................................................... 14

2.1  Introduction .......................................................................................................................... 14

2.1.1  The role of BamC in the BAM complex .......................................................................... 14

2.1.2  Structure determination with Rosetta using a limited NMR dataset ............................... 15

2.1.3  Purpose/Overview .......................................................................................................... 16

2.2  Materials and Methods ......................................................................................................... 17

2.2.1  Expression and purification of uniformly $^{2}$H, $^{15}$N and/or $^{13}$C BamC ...................... 18

2.2.2  NMR experiments on 13C, 15N-labeled BamC101–344 .................................................. 19

2.2.3  Expression and purification of uniformly $^{2}$H, $^{15}$N and methyl ILV $^{1}$H, $^{13}$C BamC101–344 ........................................................................................................... 21

2.2.4  NMR experiments on uniformly $^{2}$H, $^{15}$N and methyl ILV $^{1}$H, $^{13}$C BamC$_{101-344}$ .... 22

2.3  Structure calculation using CS-RDC-NOE Rosetta ............................................................ 24

2.3.1  Structure calculation using XPLOR-NIH ..................................................................... 24

2.3.2  Database search for structural homologues and conserved residues ............................. 25

2.3.3  BamC$_{101-344}$ NMR titration experiments ................................................................. 26

2.3.4  Interaction studies with mature lipoproteins BamC$_{25-344}$ and BamD ....................... 26

2.4  Results and Discussion .......................................................................................................... 27
2.4.1 Chemical shifts and protease assay show the first 75 amino acids of BamC25-344 are disordered. ...............................................................

2.4.2 BamC has two well-defined domains connected by a flexible linker. ...............30

2.4.3 Methyl groups used to search for long-range NOEs in BamC25-344 ...............32

2.4.4 RDCs indicate domains orient independently in solution ..........................36

2.4.5 $^{15}$N relaxation and hydrogen/deuterium exchange show linker region between N- and C-terminal domains is flexible. ........................................39

2.4.6 The N- and C-terminal domains of BamC adopt a helix-grip motif ............45

2.4.7 Validation of BamC$_{101}$–344 structures ................................................46

2.4.8 Helix-grip domains of BamC$_{101}$–344 are remarkably similar to X-ray structures .................................................................51

2.4.9 CS-RDC-NOE Rosetta vs. traditional simulated annealing .......................53

2.4.10 Proposed function of BamC based on structural analysis, binding data and new insights from BamC/D X-ray crystal structure. ...............................56

2.5 Conclusions /Future Directions ..................................................................61

Chapter 3 Refining the structure of BamA POTRA4–5 domains with residual dipolar couplings 63

3.1 Introduction ...............................................................................................63

3.1.1 The role of BamA and POTRA domains in the BAM complex ...............63

3.1.2 Purpose/Overview ..............................................................................65

3.2 Materials and Methods ...........................................................................66

3.2.1 Expression/Purification of uniformly $^{13}$C, $^{15}$N labeled BamA POTRA4–5 66

3.2.2 NMR backbone resonance assignments of BamA POTRA4–5 ..............67

3.2.3 Measurement of $\{^1H\}$-$^{15}$N heteronuclear NOE spectra ....................68

3.2.4 Measurement of Residual Dipolar Couplings .....................................68

3.2.5 Domain orientation of POTRA4–5 in XPLOR-NIH ..............................68

3.3 Results and Discussion ............................................................................70

3.3.1 Backbone resonance assignment of POTRA4–5 ...............................70

3.3.2 Solution domain orientation of POTRA4–5 is consistent with the crystal structure .................................................................78
A higher resolution X-ray crystal structure of POTRA4–5 fits the RDCs with less “structural noise”. ..................................................84

Conclusions/Future directions ..........................................................91

Investigating the flexibility of BamA POTRA 1–5 ................................92

Introduction .......................................................................................92

β-augmentation in the E. coli BamA POTRA domains. ......................92

Crystal structures of POTRA1–4 show two unique conformations with a hinge between POTRA2 and POTRA3...........................................................................93

Purpose/Overview .............................................................................95

Materials and Methods ......................................................................96

Expression/Purification of uniformly $^{13}$C, $^{15}$N labeled BamA POTRA constructs97

NMR backbone resonance assignments of BamA POTRA3–5 ..................98

NMR experiments on BamA POTRA constructs. ....................................98

Results and Discussion ......................................................................101

Backbone resonance assignment of POTRA3–5 ..................................101

Assignment of the $^1$H, $^{15}$N backbone amide resonances of POTRA1–5......108

Backbone $^{15}$N relaxation data indicate flexibility in POTRA1–5 ............109

Paramagnetic relaxation enhancement shows that POTRA1–5 transiently samples a “compact” conformation in solution ......................................................113

Measuring RDCs in the 45.6 kDa POTRA1–5. ......................................115

Analysis of RDCs shows that POTRA4–5 orient independently of POTRA1–2 and POTRA3. .................................................................120

Conclusions/Future directions ..........................................................123

Bibliography ......................................................................................124

Script for measuring methyl-methyl distances in PDB file....................135

XPLOR script for basic simulated annealing of BamC ..............................141

Using XPLOR-NIH for domain orientation using residual dipolar couplings.148

Backbone connectivities for POTRA4–5 ..............................................158

Backbone connectivities for POTRA3–5 ..............................................166
Appendix A6  Filenames and parameters for NMR experiments for POTRA4–5 ..................179
Appendix A7  Filenames and parameters for NMR experiments for POTRA3–5 and POTRA1–5 180
FIGURES

Figure 1.1 The *E. coli* cellular membranes. ................................................................. 2

Figure 1.2 β-augmentation in POTRA3. ........................................................................... 4

Figure 1.3 Model of the β-barrel assembly machine. ......................................................... 5

Figure 1.4 Coordinate system for residual dipolar coupling between one-bonded nuclei *i* and *j*. 12

Figure 2.1 The biosynthetic precursors used “linearized” isotopic labeling strategy. .......... 22

Figure 2.2 Out-and-back magnetization transfer steps in the HMCMCBCA and HMCMCGCBA. ........................................................................................................... 23

Figure 2.3 The predicted order parameter $S^2$, calculated from the chemical shift using the random coil index. ............................................................................................................. 28

Figure 2.4 2D $^1$H-$^{15}$N NMR spectra of mM BamC$_{101-344}$ and BamC$_{25-344}$. ................................................................. 29

Figure 2.5 $^1$H, $^{15}$N HSQC of 1.8 mM BamC$_{101-344}$ at 30 °C at 800 MHz. .................. 30

Figure 2.6 CS-RDC-NOE Rosetta calculation generated BamC$_{101-344}$ structures with clearly defined domains, but orientations did not converge. .................................................. 32

Figure 2.7 $^1$H, $^{13}$C HMQC spectrum of 1.0 mM BamC at 30 °C and 800 MHz. ................. 34

Figure 2.8 Strip plot of HMCMCGCBA spectrum with $^1$H, $^{13}$C HMQC overlaid. ............... 36

Figure 2.9 Histograms of the RDCs measured in BamC. ................................................... 39

Figure 2.10 The $^{15}$N relaxation data on BamC$_{101-344}$ show regions of increased backbone flexibility........................................................................................................... 41

Figure 2.11 Hydrogen deuterium exchange profiles mapped onto a low-energy structure of BamC$_{101-344}$. ........................................................................................................... 42

Figure 2.12 Ensemble of the structures of the N-terminal and C-terminal domains of BamC generated by separate CS-RDC-NOE Rosetta calculations for the two domains........... 44
Figure 2.13 Backbone–backbone $^1$H–$^1$H NOEs confirm the $\beta$-sheet secondary structure in BamC generated by CS-RDC-NOE Rosetta. 

Figure 2.14 A network of methyl-methyl NOEs confirms hydrophobic packing. 

Figure 2.15 The CS-RDC-NOE Rosetta structural ensemble superimposes well with X-ray crystal structure of BamC. 

Figure 2.16 Region of the BamC $\alpha$1 helix with dynamic properties has a zero occupancy in the crystal structure. 

Figure 2.17 CS-RDC-NOE Rosetta modeling is tolerant of misassigned NOEs. 

Figure 2.18 CS-RDC-NOE Rosetta is more tolerant of misassigned NOEs than traditional simulated annealing. 

Figure 2.19 The helix-grip motif is found in the major latex proteins and kinase associated domains. 

Figure 2.20 Electrostatic surface potentials of the N-terminal (a) and C-terminal (b) domains of BamC. 

Figure 2.21 The conserved amino acids in BamC are clustered in sites potentially important for protein–protein interactions. 

Figure 2.22 X-ray crystal structure of BamCD complex. 

Figure 3.1 BamA is comprised of five N-terminal polypeptide transport-associated (POTRA) repeats and a C-terminal $\beta$-barrel domain. 

Figure 3.2 An $^1$H, $^{15}$N HSQC amide spectrum of $^{13}$C, $^{15}$N POTRA4–5. 

Figure 3.3 Slices from an $^{15}$N-edited NOESY spectrum shows crosspeaks between residues Ala363–Arg366. 

Figure 3.4 Slices from an $^{15}$N-edited NOESY spectrum shows crosspeaks between residues Asp380–Glu387. 

Figure 3.5 Residues in the linker between the $\alpha$1 and $\alpha$2 helix of POTRA5 were unassigned. 

Figure 3.6 Peak volumes in the $^1$H, $^{15}$N HSQC of the $^1$H, $^{15}$N amides for POTRA4–5 decrease around unassigned loop.
Figure 3.7 $^1$H-$^{15}$N heteronuclear NOE values plotted as a function of residue number in POTRA4–5. .......................................................... 78

Figure 3.8 Histograms of RDCs collected on POTRA4–5. .......................................................... 80

Figure 3.9 Plot of the experimental and predicted $^1$H-$^{15}$N RDCs of POTRA4–5......................... 81

Figure 3.10 Domain orientation of POTRA4–5 from a starting pool of structures with randomized domain orientation. .......................................................... 82

Figure 3.11 A network of NOEs shows that the backbone around the POTRA4–5 linker is relatively rigid.......................................................... 83

Figure 3.12 Spatial position of hydrogen atoms added to structure coordinates derived from X-ray crystallographic data is dependent on the resolution............................................. 85

Figure 3.13 Plot of the experimental and predicted $^1$H-$^{15}$N RDCs of POTRA4–5 using an RDC data set without filtering out residues with low $^1$H-$^{15}$N heteronuclear NOE values to the low and high-resolution X-ray crystal structures.......................................................... 88

Figure 3.14 Plot of the experimental and predicted $^1$H-$^{15}$N RDCs of POTRA4–5 using an RDC data set after filtering out residues with low $^1$H-$^{15}$N heteronuclear NOE values to the low and high-resolution X-ray crystal structures.............. 89

Figure 3.15 Domain orientation of high resolution POTRA4–5 from a starting pool of structures with randomized domain orientation.......................................................... 90

Figure 4.1 The periplasmic domains of BamA share a common POTRA fold ....................... 93

Figure 4.2 An “extended” and “compact” orientation of POTRA1–5 in crystal structures. ....... 95

Figure 4.3 Schematic description of the NMR resonance assignment strategy................. 96

Figure 4.4 Increasing the temperature and deuterating POTRA1–5 significantly improves spectral quality in the amide region of the $^1$H, $^{15}$N HSQC spectrum.............................................. 103

Figure 4.5 $^1$H, $^{15}$N HSQC spectra POTRA1–2 and POTRA3–5 overlay with POTRA1–5. ...... 105

Figure 4.6 An $^1$H, $^{15}$N HSQC spectrum of $^2$H, $^{13}$C, $^{15}$N POTRA3–5. ......................... 106

Figure 4.7 Residues with weak peak intensity and unassigned residues in the $^1$H, $^{15}$N, HSQC cluster to the POTRA3 domain that interfaces with POTRA2............................................. 107
Figure 4.8 Unassigned POTRA3 residues have zero occupancy in crystal structure. .......... 108

Figure 4.9 $^1\text{H}, ^{15}\text{N}$ HSQC of 0.5 mM $^2\text{H}, ^{13}\text{C}, ^{15}\text{N}$ POTRA1–5 at 30 °C and 800 MHz. ........ 109

Figure 4.10 {$^1\text{H}$}–$^{15}\text{N}$ heteronuclear NOE values mapped onto POTRA1–5 domains. .......... 111

Figure 4.11 Backbone $^{15}\text{N}$ relaxation data indicate flexibility in POTRA1–5. ......................... 112

Figure 4.12 PRE shows that POTRA1–5 transiently adopts a “compact” configuration in solution.......................................................................................................................... 115

Figure 4.13 Schematic representation of couplings measured using three methods for a small and large protein. ........................................................................................................................................ 116

Figure 4.14 $^1\text{H}$–$^{15}\text{N}$ RDCs for POTRA1–5 were measured from the halved couplings. .......... 119

Figure 4.15 ARTSY spectra used to measure $^1\text{D}_{\text{NH}}$ for POTRA1–5. ........................................ 120

Figure 4.16 RDCs measured using ARTSY fit well to RDCs back calculated from crystal structures. ........................................................................................................................................ 123

Figure A4.1 Backbone connectivities for Met263–Ala282. .............................................................. 158

Figure A4.2 Backbone connectivities for Val282–Thr302. ............................................................. 159

Figure A4.3 Backbone connectivities for Thr302–Val322. ............................................................. 160

Figure A4.4 Backbone connectivities for Val322–Asp342. ............................................................. 161

Figure A4.5 Backbone connectivities for Lys333–Leu365. ............................................................. 162

Figure A4.6 Backbone connectivities for Leu365–Leu392. ............................................................. 163

Figure A4.7 Backbone connectivities for Leu392–Val413. ............................................................. 164

Figure A4.8 Backbone connectivities for Asp413–Gly426. ............................................................. 165

Figure A5.1 Backbone connectivities for Met173–Asp191. ............................................................ 166

Figure A5.2 Backbone connectivities for Asp191–Arg237............................................................... 167
Figure A5.3 Backbone connectivities for Arg237–Lys267. .................................................. 168

Figure A5.4 Backbone connectivities for Lys267–Ser281. .................................................. 169

Figure A5.5 Backbone connectivities for Ser281–Tyr296. .................................................. 170

Figure A5.6 Backbone connectivities for Tyr296–Leu311. .................................................. 171

Figure A5.7 Backbone connectivities for Leu311–Ile328. .................................................. 172

Figure A5.8 Backbone connectivities for Ile328–Asp342. .................................................. 173

Figure A5.9 Backbone connectivities for Asp342–Gly356. .................................................. 174

Figure A5.10 Backbone connectivities for Gly356–Glu387. .............................................. 175

Figure A5.11 Backbone connectivities for Glu387–Asp401. .............................................. 176

Figure A5.12 Backbone connectivities for Asp401–Lys417. .............................................. 177

Figure A5.13 Backbone connectivities for Lys417–Gly426. .............................................. 178
### TABLES

Table 2.1 Structural statistics of the BamC N- and C-terminal domains........................................46

Table 3.1 Summary of RDCs fits to high and low resolution structures of POTRA4–5.................90

Table 4.1 Values of $D_{a}$ and $R$ calculated for POTRA1–5..........................................................121

Table A6.1 Filenames and parameters for backbone experiments for BamA POTRA4–5..........179

Table A6.2 Filenames and parameters for experiments used to measure $^{1}D_{NH}$ on POTRA4–5..179

Table A7.1 Filenames and parameters for backbone experiments for BamA POTRA3–5...........180

Table A7.2 Filenames and parameters for experiments used to measure $^{1}D_{NH}$ on POTRA1–5..180

Table A7.3 Filenames and parameters for HSQC spectra collected on POTRA constructs......181
Chapter 1 Introduction

1.1 The outer membrane of Gram-negative bacteria

1.1.1 The role of the outer membrane in cell homeostasis

The cytosol of Gram-negative bacteria is fortified by an inner membrane (IM) and an outer membrane (OM), which are separated by ∼140 Å of periplasmic space (Figure 1.1) (Collins et al., 2007; Graham et al., 1991). The IM is made up of a symmetric phospholipid bilayer and mostly α-helical proteins. The OM is asymmetric, with the inner leaflet containing mostly phospholipids and the outer leaflet containing mostly lipopolysaccharides (Bos et al., 2007a). The OM functions as a selective barrier that protects the bacteria from harmful compounds but allows nutrients to pass through (Gentle et al., 2005). Unlike the IM, the OM does not have access to a proton gradient, ATP or any other known energy source in the periplasm to actively import nutrients or export waste and toxins (Bos et al., 2007a). Instead, a passive selection filter is implemented with a family of outer membrane proteins (OMPs) that span the OM and filter the import and export of compounds to the periplasm from the extracellular environment (Gentle et al., 2005). OMPs are integral for basic physiological functions, virulence and multidrug resistance (Bos et al., 2007a). Disruption of the outer membrane is detrimental to Gram-negative bacteria and makes them susceptible to environmental conditions and is therefore an attractive target for antibiotics (Werner and Misra, 2005; Wu et al., 2005a).
The cytoplasm of Gram-negative bacteria, such as *E. coli*, is surrounded by an inner and an outer membrane. The outer membrane is decorated on the outside with lipopolysaccharides and populated with β-barrel outer membrane proteins (OMPs).

### 1.1.2 The β-barrel assembly machine of *E. coli*

In *Escherichia coli*, OMPs are translated in the cytosol and transported to the periplasm via the SecYEG machinery where they are handed off to chaperones that deliver their cargo to the β-barrel assembly machine (BAM) (Bos et al., 2007a). The BAM is responsible for the folding and insertion of most OMPs (Werner and Misra, 2005; Wu et al., 2005a), making it an interesting target for new antibiotics. In the past 7 years the components of the BAM have been identified as, BamA (“YaeT”), BamB (“YfgL”), BamC (“NlpB”), BamD (“YfiO”) and BamE (“SmpA”). BamA is an essential component of the BAM and is composed of a carboxyl-terminal β-barrel that spans the OM and five periplasmic POTRA (polypeptide transport-associated) domains (Sánchez-Pulido et al., 2003). The POTRA domains have a common fold (~75 amino acids) that comprises a three-stranded β-sheet overlaid by a pair of antiparallel α-helices.
POTRA1–5 are hypothesized to play a direct role in the folding and insertion of β-barrel proteins into the outer membrane (Kim et al., 2007; Gatzeva-Topalova et al., 2008; Bos et al., 2007b; Sánchez-Pulido et al., 2003; Bennion et al., 2010), but the mechanistic details of this function are unclear. Orthologs of BamA are found in the OMs of all Gram-negative bacteria as well as in the mitochondria and chloroplasts of eukaryotes, where it is thought that they are responsible for assembling β-barrels by the same mechanism (Voulhoux et al., 2003; Reumann et al., 1999; Wiedemann et al., 2003).

1.1.3 Folding OMPs by β-augmentation

Crystallographic analyses of the BamA POTRA1–4 suggest that the POTRA domains may function as protein recognition sites, binding the nascent chains of outer membrane proteins by β-augmentation (Kim et al., 2007; Gatzeva-Topalova et al., 2008). The mechanism of β-augmentation involves association of a β-strand from a ligand with a β-strand or a β-sheet in the binding partner (Remaut and Waksman, 2006). In two independent crystal structures, a C-terminal tail of the crystallized constructs was found pushed up against the β-sheet of POTRA3 in a symmetry related molecule of the lattice. The tail adopted a β-strand conformation, “augmenting” the β-sheet in POTRA3 Figure 1.2 (Kim et al., 2007; Gatzeva-Topalova et al., 2008). While this could just be due to a crystallization artifact, the appearance of β-strand pairing in two crystal structures in different space groups and different orientations suggest that this could reflect a novel feature of the system.

Additionally, the POTRA3 domain has an ~30 Å long hydrophobic surface groove that is conserved in similar proteins, such as the cytoplasmic chaperone SecB, which is known to bind outer membrane protein precursors by β-augmentation (Xu et al., 2000; Kim et al., 2007; Gatzeva-Topalova et al., 2008). The length of the β-strands of outer membrane proteins range
from 27–35 Å, making the hydrophobic groove a suitable size to accommodate an outer membrane protein β-strand (Tamm et al., 2001), which may nucleate β-strand conformations in the nascent chains of outer membrane proteins as a step in the formation of the β-barrel formation.

Figure 1.2 β-augmentation in POTRA3.
Evidence for β-augmentation is found in the crystal structures of POTRA1–4 that contained five residues from POTRA5. In the crystal structure 3EFC (a) (Gatzeva-Topalova, et al., 2008), a strand from POTRA5 in a symmetry related molecule at a lattice point appears along side β2 of POTRA3. In the crystal structure 2QCZ (b) (Kim, et al., 2007) a strand from POTRA5 from a second copy of BamA POTRA1–4 in the asymmetric unit that appears along side β2 of POTRA3.

1.1.4 The accessory lipoproteins

Much progress has been made since the discovery of the BAM complex (Onufryk et al., 2005). The key players in the BAM have been identified and recent biochemical reconstitution assays, crystal and NMR structures are laying the groundwork for understanding how the BAM folds and inserts OMPs into the outer membrane (Hagan and Kahne, 2011; Gatzeva-Topalova et al., 2008; Kim et al., 2007; 2011a; Sandoval et al., 2011; Heuck et al., 2011). However, a detailed molecular mechanism for this molecular machine has remained elusive. The accessory proteins BamB, BamC, BamD and BamE are lipoproteins that associate with BamA in the BAM complex. BamD and BamB participate in direct interactions with BamA, while BamC and BamE
affiliate with BamA through interactions mediated by BamD (Figure 1.3) (Malinverni et al., 2006; Vuong et al., 2008; Wu et al., 2005a; Kim et al., 2011b). BamD is an essential component of the BAM and widely represented across Gram-negative bacteria (Sandoval et al., 2011). Crystal structures of the *E. coli* and thermophilic *Rhodothermus marinus* BamD show a protein that contains five α-helical tetratricopeptide repeats (TPRs) (Sandoval et al., 2011; Albrecht and Zeth, 2011) (Figure 1.3, yellow). The three N-terminal TPRs make up a domain that exhibits structural homology with proteins that recognize extended polypeptides (e.g. unfolded OMPs) and the two C-terminal TPRs have been identified in mutant screens as important for binding the BAM components (Sklar et al., 2007; Malinverni et al., 2006). Thus, a model has been proposed in which the N-terminal domain of BamD participates in interaction with OMP substrates and the C-terminal domain provides a scaffold for interaction with BAM components (Sandoval et al., 2011). The requirement of BamA and BamD for viability and the broad phylogenetic distribution of these two proteins suggest that they represent the core of the BAM (Sandoval et al., 2011).

**Figure 1.3 Model of the β-barrel assembly machine.**
Structures of the BAM components include: BamA (POTRA1–4: 3EFC (Gatzeva-Topalova, et al., 2008 5, POTRA4–5: 3OG5 (Gatzeva-Topalova, et al., 2010) green), BamB (3PIL (Kim, et al., 2011), blue), BamD (3QKY (Sandoval, et al., 2011), yellow) and BamE (2KXX, Kim et al., 2011), orange). BamC is represented as a red oval, the structural details are given in this thesis. T
The remaining accessory lipoproteins, BamB, BamC, and BamE, have less phylogenetic distribution than BamA or BamD (Gatsos et al., 2008). Deletion of each gene separately in *E. coli* results in membrane defects, but the bacteria are still viable (Wu et al., 2005b; Sklar et al., 2007). Reconstitution of the BAM in proteoliposomes has shown that including the accessory lipoproteins increases the efficiency of folding and insertion of an outer membrane protein (Hagan et al., 2010; Hagan and Kahne, 2011). The structure of BamB consists of an eight-bladed \( \beta \)-propeller fold (Figure 1.3, blue). Similar to BamA, there are two crystal contacts with BamB symmetry mates in the crystal lattice demonstrating that BamB is capable of \( \beta \)-augmentation (Heuck et al., 2011). Additionally, two mutations in BamB disrupt binding to BamA at a site located somewhere on the POTRA domains 2 to 4. (Vuong et al., 2008). The mutations on BamB are in proximity to one another and suggest an orientation that is shown in Figure 1.3. BamE is the smallest, most conserved of the accessory lipoproteins. Previous studies with BamE null strains of *E. coli* have shown that BamE is important in maintaining outer membrane integrity and normal levels of OMPs (Sklar et al., 2007; Anwari et al., 2010). Structural homology of BamE with a \( \beta \)-lactamase inhibiting protein and the chaperone subunit of subtilisin suggest that BamE may be important for a chaperoning activity in the BAM complex (Kim et al., 2011b). Furthermore, BamE is known to interact with BamD and BamC in the BAM complex, perhaps acting as part of the molecular scaffold that holds the components of the BAM together (Figure 1.3, orange). Similar to BamE, BamC null strains of *E. coli* display an increase in outer membrane permeability and reduced levels of OMPs suggesting that BamC is involved in the formation or integrity of the outer membrane (Wu et al., 2005b; Onufryk et al., 2005), but no structural description was available for BamC before the work presented here. The work presented in this thesis contributes to understanding the mechanics of how the BAM works by providing
a description of the solution NMR structure of BamC and the solution orientation of the periplasmic domains of BamA.

1.2 Determining solution structures of proteins with NMR

NMR spectroscopy is an important technique that can be used to investigate the structure, dynamics and interactions of a wide range of chemical systems; from small organic molecules (Rule and Hitchens, 2006) to proteins and nucleic acids in the 100s of kDa (Riek et al., 2002; Wüthrich, 1986). The field of protein NMR spectroscopy has become popular in recent years because of the useful application to measuring the structure and dynamics of proteins in solution. From determining the structure of the excited states of proteins (Stoner-Ma et al., 2008; Skrynnikov et al., 2002) to transient encounters of protein-nucleic acid complexes (Clore et al., 2007; Tang et al., 2006), NMR is a versatile tool for biophysical characterization of macromolecules.

Since the implementation of the sequential resonance assignment strategy first proposed by Wüthrich and co-workers (Wüthrich, 1986), structure determination of proteins by solution NMR has followed a general procedure in which chemical shift assignments of $^1$Hs are related to the peptide chain by sequential assignments and $^1$H-$^1$H distances are measured with the $^1$H-$^1$H nuclear Overhauser effect (NOE). The NOE has a distance dependence of $r^6$, limiting the distances measured to ~5 Å (Rule and Hitchens, 2006). These distances are then used as restraints in a simulated annealing protocol, where the torsion angles of a polypeptide chain are allowed to rotate freely in space at a high “temperature” (Rule and Hitchens, 2006). The distance restraints are applied as a pseudoenergy force constant (Schwieters et al., 2003). As the temperature is cooled, the force of the distance restraint is increased (Schwieters et al., 2003). As the size of proteins investigated by NMR increased, the spectral overlap increased and the chemical shifts of $^{15}$N amide and $^{13}$C backbone and side chain carbons were introduced into 3D experiments to
separate resonances, with the ultimate goal still to measure $^1$H-$^1$H (Sattler et al., 1999; Cavanagh et al., 2007; Rule and Hitchens, 2006).

Several advances have decreased the reliance on only the NOE for structure determination. For example, the chemical shift of the $^{13}$C and $^{15}$N resonances can be related to the dihedral angles in a peptide bond (Cornilescu et al., 1999; Shen et al., 2009b). Now, dihedral angles estimated from chemical shifts are regularly incorporated into NMR structure determinations (Shen et al., 2009b). Another NMR experimental measurement that has transformed NMR spectroscopy is the residual dipolar coupling (RDC). The RDC is an orientational restraint that is dependent on the angle that a bond vector makes relative to an arbitrary molecular frame (Fischer et al., 1999; Cavanagh et al., 2007; Rule and Hitchens, 2006). The RDC gives information on the local fold of proteins, as well as global orientation information which was previously unattainable with NOE data only (Prestegard et al., 2004). Additional cross-disciplinary techniques can be incorporated into an NMR structure calculation and can give information on domain orientations and overall shape of a molecule. Small Angle X-ray (or Neutron) Scattering (SAXS or SANS) () and solution paramagnetic relaxation enhancement (Madl et al., 2011) are gaining acceptance and are included in a growing list of measurements that can be combined in a solution structure calculation (Gabel et al., 2008).

Along with integrating new types of experimental data into structure calculations, advancements to the structure calculation process itself are requiring fewer experimental restraints (Shen et al., 2008). For example, experimental NMR data can be included in the protein fold prediction program Rosetta to guide the conformational search. The program CS-Rosetta requires only the experimental backbone chemical shifts ($^1$H$^N$, $^1$H$^\alpha$, $^{15}$N$^H$, $^{13}$C$^\alpha$ and $^{13}$C$^\beta$) to accurately generate structures for proteins smaller than $\sim$12 kDa (Shen et al., 2008; 2009a).
However, for larger proteins, the chemical shifts may not provide enough information to guide the conformational search (Raman et al., 2010) and additional experimental data such as RDCs and NOEs can be included and have been shown to increase the ability of Rosetta to generate accurate protein structures (Raman et al., 2010).

1.3 Determining domain orientation with residual dipolar couplings

The relative orientation and dynamics of well-defined domains in proteins or nucleic acids can be determined using RDCs (Fischer et al., 1999). This is an important tool to compensate for the lack of long-range distance restraints inherent in solution NMR datasets, where local distances are well defined by the NOE, but long-range information is missing (Fischer et al., 1999). RDCs can also be used to determine the solution orientation of protein structures solved using X-ray crystallography. In cases such as the SH2-SH3 domain orientation of the Src kinase Lck, the crystal structure trapped a particular orientation of the domains, while the solution RDC domain orientation studies indicated that the domains orient independently in solution and that this flexibility was important for function (Hofmann et al., 2005).

The dipolar coupling is a through-space dipole-dipole interaction between two spin-1/2 nuclei that has a distance and angle dependence as shown in Equation 1.1:

\[
\langle D_{ij}(\sigma) \rangle = \frac{h\gamma_i\gamma_j}{(2\pi r)^3} \left[ \frac{3\cos^2 \sigma - 1}{2} \right]
\]

where \(r\) is the distance between a specific pair of nuclei \(i\) and \(j\), \(\gamma_i, \gamma_j\) are the gyromagnetic ratios for the nuclei, \(h\) is Planck’s constant, \(\sigma\) is the angle between \(ij\) internuclear vector and the static magnetic field (Prestegard et al., 2004). When the \(D_{ij}\) is measured for bonded nuclei, the \(r\)
is considered constant. The largest possible splitting, when $\sigma = 0$, is defined by $D_{\text{max}} = h \gamma_i \gamma_j / (2\pi r)^3$. The value of $D_{ij}$ is in Hertz when all of the parameters are in SI units. For partially aligned molecules, rotational diffusion changes the orientation of the $ij$ bond vector with respect to the field and the measured $D_{ij}$ becomes an average of over time and over the ensemble of orientations of the molecule in the solution, which is indicated by brackets in Equation 1.1. In an isotropic solution, the dipolar coupling between two nuclei averages to zero. However, in partially aligned solutions, such as a liquid crystal, a residual dipolar coupling can be observed. Measuring the residual dipolar coupling in a partially aligned medium requires a coordinate transformation, where the new axis system is related to the molecular frame instead of the magnetic field and is referred to as the alignment tensor. After the coordinate transformation, the $D_{ij}$ is described by Equation 1.2 (Rule and Hitchens, 2006):

**Equation 1.2**

$$D_{ij}(\theta, \phi) = D_{\text{max}} \left[ \frac{1}{2} A_a (3 \cos^2 \theta - 1) + \frac{3}{4} A_r (\sin^2 \theta \cos^2 \phi) \right]$$

Where $A_a$ and $A_r$ are the axial and rhombic components of the alignment tensor and are related to the principal components by:

**Equation 1.3**

$$A_a = \frac{3}{2} A_{zz}$$

**Equation 1.4**

$$A_r = A_{xx} - A_{yy}$$

$A_{xx}$, $A_{yy}$, and $A_{zz}$ are the principal components of the diagonalized Saupe order matrix, which describes the ordering of molecules in a three dimensional space (Rule and Hitchens,
The Saupe order matrix is defined so that $|A_{zz}| \geq |A_{yy}| \geq |A_{xx}|$ and $A_{xx} + A_{yy} + A_{zz} = 0$. There is a six-fold degeneracy to solutions from the diagonalized matrix because $\pm 90^\circ$ rotations about any of the axes merely exchanges the other two axes (Rule and Hitchens, 2006). The Euler angles $\alpha$, $\beta$ and $\gamma$ relate the orientation of the alignment tensor to the magnetic field ($B_0$ in Figure 1.4).

The molecular alignment tensor cannot be measured directly and therefore, the axial component of the alignment tensor, $A_z/2$, is factored out to give Equation 1.5:

**Equation 1.5**

$$D_{ij}(\theta, \phi) = D_a \left[ (3 \cos^2 \theta - 1) + \frac{3}{2} R (\sin^2 \theta \cos 2\phi) \right]$$

where $D_a$ is the magnitude of the RDC alignment tensor and $R$ is defined as the rhombicity, where $R = A_z/A_a$. In proteins, the alignment tensor is commonly determined from the RDC for the amide bond vector, NH. $D_a$ and $R$ can be determined if an adequate sampling of orientations of NHs in a domain of defined structure are collected (Losonczi et al., 1999). In a multi-domain molecule with well-defined domains, if the orientation of the domains is fixed, then the alignment tensor for each domain will be the same (Tolman et al., 1997). If the alignment tensors for each domain are the same, then the orientational data can be used to orient the domains with respect to one another (Fischer et al., 1999).
Figure 1.4 Coordinate system for residual dipolar coupling between one-bonded nuclei $i$ and $j$.

The orientation of the internuclear bond vector $ij$, with radius, $r$, (orange) to the molecular alignment tensor ($S_{zz}$, $S_{yy}$ and $S_{xx}$, green) is described by the polar angles $\theta$ and $\phi$. The Euler angles $\alpha$, $\beta$ and $\gamma$ relate the orientation of the molecular alignment tensor to the direction of the static magnetic field.

Implementations of domain orientation with RDCs have been incorporated in a simulated annealing protocol on multi-domain complexes where the domains are considered as rigid bodies and the torsion angles of the linkers are allowed to vary (Clore et al., 1998b). In this case, the RDCs are included as a pseudoenergy function and an energy penalty is applied to the system for orientations of bond vectors that violate solutions to the Saupe order matrix. Other approaches such as those used in MODULE (Dosset et al., 2001), PALES (Zweckstetter, 2008) and REDCAT (Valafar and Prestegard, 2004) use rigid body modeling of the domains independently.
to determine the tensor eigenvalues and eigenvectors using least-squares minimization (MODULE) or singular value decomposition (PALES and REDCAT) of the target function over all RDCs associated with a given domain (Zweckstetter, 2008; Dosset et al., 2001). The domain orientations are performed in the absence of a linker, whose coordinates must be included later and a standard refinement is performed to minimize the structure (Zweckstetter, 2008; Dosset et al., 2001).

1.4 Overview

The emphasis of this thesis is the structural characterization of two components of the β-barrel assembly machine: BamC and BamA. In Chapter 2, a novel approach to structure calculation of BamC is described. Very few blind CS-RDC-NOE Rosetta structure calculations have been performed to this point. Our choice to employ this method was a bold undertaking. Our confidence in the accuracy of the structures that were produced by CS-RDC-NOE Rosetta was strengthened by validations reported in this thesis. The measurement of methyl-methyl NOEs and reanalysis of a supplementary set of 1H-1H NOEs provided strong support for the validity of the structures. An analysis of the 15N backbone dynamics showed that BamC has a very flexible linker that connects two well-defined domains and that the regions of the structural ensemble that did not converge to a unique conformation also showed increased 15N backbone amide dynamics.

Chapters 3 and 4 detail the domain orientation and flexibility of the POTRA domains of BamA. Solution domain orientation using RDCs validated the orientation of POTRA4–5 in a “spliced” crystallographic model of POTRA1–5. The flexibility of POTRA1–5 in solution was assessed with 15N amide backbone dynamic, paramagnetic relaxation enhancement and analysis of RDCs.
Chapter 2  Determining the solution structure of BamC

2.1  Introduction

2.1.1  The role of BamC in the BAM complex

BamC is a 35 kDa outer membrane lipoprotein of the \( \beta \)-barrel assembly machine (BAM) with unknown function (Bouvier, 1991). Like all lipoproteins, BamC is synthesized in the cytoplasm as a precursor prolipoprotein, with a characteristic 25-amino acid N-terminal periplasmic localization signal sequence. BamC is translocated from the cytosol across the inner membrane into the periplasm by the SecYeg machinery (Hayashi and Wu, 1990). The prolipoprotein signal sequence is then removed by signal peptidase II at the characteristic Leu-X-Y-Cys cleavage site; a glycerol moiety from phosphatidylglycerol (PG) is attached by a thioether linkage to the sulfhydryl group of Cys25 (Chattopadhyay et al., 1979; Chattopadhyay and Wu, 1977; Tokunaga et al., 1984); and finally two ester-linked and one amide linked fatty acid are covalently attached to the new amino terminal glycercylcysteine (Hantke and Braun, 1973). Lipid modification is heterogeneous, but tends to be ester-linked cyclopropyl fatty acids or phospholipids and amide-linked palmitic acid (Inouye et al., 1972). Like most lipoproteins, it is assumed that the lipids anchor BamC to the outer membrane; however, this has not been confirmed experimentally. BamC interacts directly with the C-terminal domain of BamD (Malinverni et al., 2006; Vuong et al., 2008) and with BamE as identified with pull-down assays (Kim et al., 2011b).

The precise role of BamC in the BAM complex is unknown, however BamC null strains display an increase in outer membrane porosity and reduced levels of outer membrane proteins; suggesting that BamC is involved in the formation or integrity of the outer membrane (Wu et al.,
A structural approach was taken here to understand the mechanistic details of the role that BamC plays in the β-barrel assembly machine.

2.1.2 Structure determination with Rosetta using a limited NMR dataset

The traditional method for determining the NMR solution structure of a protein involves three steps: assignment of $^1$H, $^{13}$C, and $^{15}$N resonances; measurement of structural information, such as $^1$H–$^1$H NOE distance data and RDC orientation data; and use of computational methods to generate an ensemble of conformations consistent with the NMR-derived structural restraints (Cavanagh et al., 2007). This approach is routinely applied to small proteins; however, as the size of the protein increases (∼15 kDa), additional methods such as deuteration, relaxation-enhanced experiments (transverse relaxation-optimized spectroscopy (TROSY), cross-correlated relaxation-induced polarization transfer (CRIPT), and cross-correlated relaxation-induced enhanced polarization transfer (CRINEPT)), or specific labeling [such as $^{13}$C labeling of ILV (Ile, Leu, Val) methyl groups] are employed to improve spectral resolution and sensitivity (Riek et al., 2002; Tugarinov and Kay, 2003). A major hurdle in structure determination of larger proteins is the time-consuming task of assigning the NOE resonances of side-chain protons. Various programs such as Cyana (Ikeya et al.) and ARIA (Linge et al., 2001) are being developed to assign NOE resonances automatically as part of the structure refinement process, but these applications are still limited to smaller proteins (< 15 kDa).

An alternate strategy for determining the solution structures of proteins is to include a sparse set of experimental NMR data in the Rosetta protein structure prediction program (Shen et al., 2008; Raman et al., 2010). Rosetta was originally developed as a *de novo* method for generating protein structures from the sequence (Rohl et al., 2004). It has since been extended to a suite of programs that perform homology modeling, protein-protein docking, protein-ligand
docking, protein design, or structure determinations using limited experimental data (Das and Baker, 2008; Kaufmann et al., 2010). For de novo structure determinations, Rosetta starts by assembling fragments derived from a database of known protein structures using a low-resolution energy function, where the side chain atoms are modeled as coarse grained spheres (Rohl et al., 2004). Low-energy conformations are then subjected to further sampling using an all-atom representation. Experimental data can dramatically increase the efficiency with which near-native conformations are sampled. For example, CS-Rosetta incorporates backbone NMR chemical shift data ($^{1}H$, $^{15}N$, $^{13}C'$, $^{13}C^\alpha$, and $^{13}C^\beta$) to help guide fragment selection and conformational searching, enabling atomic-resolution structure determination for small proteins (Shen et al., 2008; 2009a). For proteins larger than $\sim$12 kDa, chemical shifts may not provide enough information to guide the conformational search (Raman et al., 2010). Thus, additional experimental data such as RDCs and NOEs can be incorporated into iterative CS-RDC-NOE Rosetta and have been shown to increase its ability to generate accurate protein structures (Raman et al., 2010). However, Raman et al. demonstrated that for proteins above $\sim$20 kDa, a significant fraction of residues did not converge to well-defined conformations using only a sparse set of amide $^{1}H$–$^{1}H$ NOEs (a total of 21 to 52 NOEs) in CS-RDC-NOE Rosetta (Raman et al., 2010). Thus, improvements in conformational sampling or additional experimental data are required to generate models for even larger proteins using Rosetta.

2.1.3 Purpose/Overview

This chapter describes the application of the iterative CS-RDC-NOE Rosetta program to generate structures of the 27 kDa E. coli periplasmic lipoprotein BamC by including chemical shift data and a limited set of NOE (149 NOEs) and RDC (156 $^{1}H$–$^{15}N$ RDCs) restraints. The CS-RDC-NOE Rosetta calculations show that BamC consists of two helix-grip type domains con-
nected by an ~18-residue linker. Additional backbone-backbone and methyl-methyl NOE data, not used in the structure calculations, were used to validate the Rosetta generated BamC structures. A set of $^{15}\text{N}$ NMR relaxation data demonstrated a high degree of conformational dynamics for the backbone in the linker region of BamC. Analysis of the $^{15}\text{N}$ relaxation data and $^1\text{H}$–$^{15}\text{N}$ RDCs indicates that the two domains in BamC do not have a fixed orientation in solution. The studies here also showed that CS-RDC-NOE Rosetta is quite robust to including inconsistent NOE distance restraints in the experimental data. A search for structural homologues was conducted to identify potential functions. A comparison of the BamC ensemble calculated here with CS-RDC-NOE Rosetta with a recently published X-ray crystal structure provides additional structural validation.

### 2.2 Materials and Methods

These materials and methods document the procedures used to prepare samples and perform experiments for the structural and functional characterization of BamC. Structural characterization consisted of assignment of $^1\text{H}$, $^{13}\text{C}$, and $^{15}\text{N}$ resonances; measurement of $^1\text{H}$–$^1\text{H}$ NOE distance data and RDC orientational data; and generation of an ensemble of conformations consistent with NMR-derived structural restraints. Validation of these structures was accomplished by analysis of an independent set of $^{15}\text{N}$ relaxation, hydrogen/deuterium exchange and NOE distance data not used in the structure calculation. Binding studies were carried out to test the affinity of BamC to lipids and BamD; based on structural homology to lipid-binding proteins and previously reported complex formation of BamC/D in the β-barrel assembly machine.
2.2.1 Expression and purification of uniformly $^2\text{H}$, $^{15}\text{N}$ and/or $^{13}\text{C}$ BamC

Uniformly $^2\text{H}$, $^{15}\text{N}$ and/or $^{13}\text{C}$ BamC samples were prepared by Susan Baker (Sousa Lab) for NMR experiments to assign the $^1\text{H}$, $^{13}\text{C}$, and $^{15}\text{N}$ resonances; measure $^1\text{H}$–$^1\text{H}$ NOE distance, RDC, hydrogen/deuterium exchange and $^{15}\text{N}$ relaxation data; and NMR titrations. The first 24 residues of *E. coli* BamC contain the periplasmic localization signal that is cleaved by signal peptidase II *in vivo* and were not included in the protein constructs studied here. In addition, the N-terminal Cys that is normally modified with a lipid anchor was mutated to Ala to prevent intermolecular disulfide cross-linking. Isotopically labeled BamC constructs used for NMR assignment and relaxation experiments were expressed and purified using the following procedure. Plasmids pMS282 (coding for BamC$_{25-344}$, C25A) and pMS639 (coding for BamC$_{101-344}$) were transformed into *E. coli* Rosetta (DE3) cells (Novagen), and small-scale growths from single colonies were used to inoculate 100 mL of LB supplemented with 50 μg/mL kanamycin. Cultures were grown overnight, spun down, and suspended in 3 L of M9 minimal medium supplemented with 50μg/mL kanamycin, 1.5 g/L $^{[13]}\text{C}$-glucose, and 1 g/L $^{15}\text{NH}_4\text{Cl}$ (Sigma/Isotec). Cultures were grown at 37 °C to an $OD_{600}$ of 0.6 and cooled on ice for 10 min. Expression was induced with 1.0 mM IPTG (Gold Bio Technology, Inc.). Cells were grown overnight at 20 °C and harvested by centrifugation. The cell pellet was suspended in lysis buffer containing 25 mM Tris-Cl, 300 mM NaCl (pH 8.0), and Complete EDTA-free protease inhibitor (Roche), and then sonicated on ice. Cell debris was removed by centrifugation, and the supernatant was applied to a Ni-NTA column (Qiagen) equilibrated with buffer A [25 mM Tris (pH 8) and 150 mM NaCl]. Ni-NTA beads were washed with 2 column volumes of buffer A, followed by 5 column volumes of buffer A containing 25 mM imidazole. The protein was eluted with buffer A containing 200 mM imidazole. Fractions containing the protein were incubated with
His-tagged TEV protease for 24 h at 4 °C by dialysis against buffer A supplemented with 10 mM DTT to cleave the His tag and then dialyzed overnight at 4 °C against buffer A. The TEV protease was removed using Ni-NTA beads, and the protein was loaded onto a size-exclusion column (HiLoad 26/60 Superdex 200; Amersham Pharmacia Biotechnology) equilibrated with an NMR buffer containing 50 mM sodium phosphate (pH 6.0) and 50 mM NaCl, 0.1 mM EDTA (buffer B) and eluted in the same buffer. Protein was then concentrated to 1.6 mM for BamC<sub>25–344</sub> and to 1.8 mM for BamC<sub>101–344</sub>, HALT protease inhibitor cocktail and 10% <sup>2</sup>H<sub>2</sub>O were added to the solution. The protein solutions were stored at −80 °C until further use.

2.2.2 NMR experiments on 13C, 15N-labeled BamC<sub>101–344</sub>

For resonance assignments, the <sup>13</sup>C, <sup>15</sup>N-labeled BamC<sub>101–344</sub> sample was exchanged into NMR buffer containing 10% <sup>2</sup>H<sub>2</sub>O, 50 mM sodium phosphate (pH 6.0), 50 mM NaCl, 0.02% NaN<sub>3</sub>, and a protease inhibitor cocktail (1× HALT; Pierce) to a concentration of ∼1.0 mM. The following 3D spectra were used to generate backbone and side-chain assignments of BamC<sub>101–344</sub>: HNCACB, CBCA(CO)NH, HNCO, HN(CA)CO, HNHAHB, HBHA(CO)NH, H(CCO)NH, (H)C(CO)NH, and (H)CCH total correlated spectroscopy (Cavanagh et al., 2007; Sattler et al., 1999). The backbone assignments for BamC<sub>101–344</sub> are very similar to the previously published assignments for BamC<sub>26–344</sub> (Knowles et al., 2009). The <sup>15</sup>N-edited and <sup>13</sup>C- edited 3D NOESY spectra were collected with a 150-ms mixing time on the <sup>13</sup>C, <sup>15</sup>N-labeled BamC<sub>101–344</sub> sample. For the <sup>13</sup>C-edited 3D NOE, the <sup>13</sup>C, <sup>15</sup>N-labeled BamC<sub>101–344</sub> sample was exchanged into NMR buffer with 99.8% <sup>2</sup>H<sub>2</sub>O. All NMR spectra were collected at 30 °C on VNMRS 900 MHz, VNMRS 800 MHz, or Inova 600 MHz spectrometers equipped with HCN z-axis gradient cold probes. NMR spectra were processed with NMRPipe and analyzed in Sparky (Goddard and
Kneller) or CCPN Analysis (Vranken et al., 2005b). Dr. Geoff Armstrong and Dr. Krisztina Varga collected and analyzed the NMR spectra for backbone resonance assignment of BamC.

The amide $^1$H–$^{15}$N RDCs were measured using a 1.1 mM $^{13}$C, $^{15}$N-labeled BamC$_{101-344}$ sample with no Pf1 phage and a 0.14 mM $^{13}$C, $^{15}$N-labeled BamC$_{101-344}$ sample in 21 mg/mL liquid crystalline Pf1 phage, prepared as described previously (Hansen et al., 1998). Two-dimensional heteronuclear single-quantum coherence sensitivity-enhanced $^{15}$N IPAP spectra (Ding and Gronenborn, 2003) were collected on isotropic (no Pf1 phage) and aligned (with Pf1 phage) samples to measure the $^1$H–$^{15}$N RDCs for orientation restraints.

For the hydrogen/deuterium exchange experiments, 0.2 mM $^{13}$C, $^{15}$N-labeled BamC$_{25-344}$ and 1.0 mM $^{13}$C, $^{15}$N-labeled BamC$_{101-344}$ were exchanged rapidly (~10 minutes) into buffer B with 99.8% $^2$H$_2$O using buffer exchange spin columns (Pierce). The selective optimized flip-angle short transient (SOFAST) $^{15}$N HMQC sequence (Brutscher and Schanda, 2005) was used to collect two-dimensional $^1$H, $^{15}$N spectra, each with a total experiment time of 4 min. Spectra were collected at 600 MHz and 30 °C. Time points were collected every 4 min for the first 4 h, then every 3 h for the following 24 h, then once a day for 7 days. The volumes from readily resolved peaks were then fit to an exponential decay function.

The data for determining $^{15}$N R$_1$ and R$_2$ relaxations rates and $^1$H–$^{15}$N heteronuclear NOE data were collected on a 0.87 mM $^2$H, $^{13}$C, $^{15}$N BamC$_{25-344}$ sample at 600 MHz. Relaxation delays for the R$_1$ and R$_2$ experiments were collected in an interleaved fashion to create a pseudo-3D fid file. For the R$_1$ experiment, the relaxation delays of 0, 100, 200, 400, 500, 600, 700, 800, and 900 ms were acquired, and for the R$_2$ experiment, relaxation delays of 0, 10, 30, 50, 70, 90, 110, 130, 150, 170, and 190 ms were acquired.
Residue-specific $\tau_c$ values were calculated from the $^{15}$N R$_1$ and R$_2$ values for residues with $^{1}$H-$^{15}$N heteronuclear NOE values greater than 0.7 using Equation 2.1 (Kay et al., 1989).

**Equation 2.1**

$$
\tau_c = \frac{1}{4\pi \nu_N} \sqrt{\frac{6}{R_2} - 7}
$$

N-terminal and C-terminal domains $\tau_c$. Estimates for the $\tau_c$ values of the individual domains were made using an empirical relation between molecular weight and $\tau_c$ values (http://www.nmr2.buffalo.edu/nesg.wiki/NMR_determined_Rotational_correlation_time).

### 2.2.3 Expression and purification of uniformly $^2$H, $^{15}$N and methyl ILV $^1$H, $^{13}$C BamC$_{101-344}$

The methyl groups of Ile, Leu and Val were chosen as NMR probes to generate an additional set of NOE distances to search for long-range NOEs and to validate the CS-RDC-NOE Rosetta structures because methyl specific $^{13}$C, $^{15}$N and $^{1}$H NOESY spectra can measure distances up to 8 Å (Mueller et al., 2000). Experiments used to assign the methyl resonances required a protein with a specific “linearized” isotopic labeling strategy, shown in Figure 2.1a and b, where a protein is uniformly $^2$H, $^{15}$N and $^{13}$C labeled, and specifically protonated and $^{13}$C on the terminal methyl group; Ile ($^{13}$CH$_3$, $\delta$1 only), Leu ($\gamma$, $^{13}$CH$_3$ or $^{12}$CD$_3$), Val ($\gamma$, $^{13}$CH$_3$ or $^{12}$CD$_3$) (Tugarinov and Kay, 2003). The pMS282 plasmid was transformed into *E. coli* Rosetta (DE3) cells (Novagen) and plated on LB agarose with 50 μg/mL kanamycin. Single colonies were used to inoculate small cultures that were grown in a rich medium for 3 h at 37 °C. Cells were first conditioned to a minimal medium in solutions containing H$_2$O and finally transferred to 2 L of M9 minimal medium prepared in 99.8% $^2$H$_2$O with 2 g/L [97–99% $^2$H, 99% $^{13}$C]-glucose, and 1 g/L $^{15}$NH$_4$Cl (Isotec). Cells were grown at 37 °C until an OD$_{600}$ of 0.6 had
been reached, at which point 100 mg/L 2-keto-3-(methyl-d₃)-butyric acid-1,2,3,4-¹³C₄,3-d₁ sodium salt (α-ketoisovalerate) and 50 mg/L 2-ketobutyric acid-¹³C₄,3,3-d₂ sodium salt (α-ketoisobutyrate) (Isotec) were added to the media. Cells were cold shocked, induced with 0.4 mM IPTG, and grown overnight at 20 °C. Purification was performed as described for ¹³C/¹⁵N labeled samples.

![Chemical structures](attachment:chemical_structures.png)

**Figure 2.1 The biosynthetic precursors used “linearized” isotopic labeling strategy.**

In this labeling strategy, pioneered by Tugarinov and co-workers (Tugarinov and Kay, 2003), the methyl-protonated Ile (¹³CH₃, δ1 only), Leu (γ, ¹³CH₃ or ¹²CD₃), Val (γ, ¹³CH₃ or ¹²CD₃) methyl labeling patterns result when 2-ketobutyric acid, ¹³C₄, 3,3-d₂ (α-ketoisobuterate) for Ile and 2-keto-3-(methyl-d₃)-butyric acid-1,2,3,4-¹³C₄, 3-d₁ (α-ketoisovalerate) for Leu and Val are supplemented into minimal media in an E. coli expression system. The precursor α-ketoisovalerate that is used is a racemic mixture of [R-¹³CH₃, S-¹²CH₃] and [S-¹³CH₃, R-¹²CH₃] at the C3 site (marked with and asterisk) which causes a labeling scheme where two peaks are observed in an ¹H, ¹³C HMBC spectrum for each of the methyl groups of Leu and Val at ~50% the peak volume relative to 100% labeling.

### 2.2.4 NMR experiments on uniformly ²H, ¹⁵N and methyl ILV ¹H, ¹³C BamC₁₀¹–₃₄₄

A series of NMR experiments were performed to assign the resonances of the methyl ¹H and ¹³C spins. These included a methyl-specific TROSY-HMQC, two experiments that
correlate methyl spins with the aliphatic carbons (methyl-detected HMCMCBCA and HMCMCGCBCA (Tugarinov and Kay, 2003)) and two methyl specific HMQC-NOESY-HMQC experiments ($^{13}$C, $^{13}$C, $^1$H and $^1$H, $^{13}$C, $^1$H). The HMCMCBCA and HMCMCGCBCA pulse sequences relay magnetization from the methyl spins to aliphatic carbons using the $J_{CC}$ ($\sim$30 Hz) in an out-and-back fashion as depicted in Figure 2.2 (Tugarinov and Kay, 2003). Chemical shifts were recorded for the $^1$Hmethyl, $^{13}$Cmethyl, $^{13}$C$^\alpha$, $^{13}$C$^\beta$ and $^{13}$C$^\gamma$ and correlated with $^{13}$C$^\alpha$, $^{13}$C$^\beta$ chemical shifts that were determined using traditional assignment strategies. All spectra were collected at 800 or 900 MHz and 30 °C. The pulse sequences used to collect all methyl spectra were kindly provided by Prof. Lewis Kay (University of Toronto).

![Figure 2.2 Out-and-back magnetization transfer steps in the HMCMCBCA and HMCMCGCBCA.](image)

Magnetization is transferred along the carbon backbone of Ile ($^{13}$CH$_3$, $\delta$1 only), Leu ($\gamma$,$^{13}$CH$_3$ or $^{12}$CD$_3$), Val ($\gamma$,$^{13}$CH$_3$ or $^{12}$CD$_3$) side chains in the HMCMCBCA (dashed arrows) and HMCMCGCBCA (dashed and solid arrows) (Tugarinov and Kay, 2003).
2.3 Structure calculation using CS-RDC-NOE Rosetta

The CS-RDC-NOE Rosetta calculations were performed by Dr. Oliver Lange at the University of Washington. The details of these calculations are reported in (Warner et al., 2011). Briefly, the backbone chemical shifts of the construct BamC<sub>101–344</sub> were used in CS Rosetta 3.X to select 25 and 200 fragments that are nine and three residues in length, respectively (Shen et al., 2008). Regions of the protein that are at least three residues in length or are located at the termini and have TALOS+ predicted order parameters (Shen et al., 2009b) of less than 0.7 (residues 110–114, 218–228, 296–301, 345, and 346, respectively) were excluded from Rosetta score computation, although they were explicitly modeled. Residues 345 and 346 were included to facilitate the cloning of the BamC construct used in the NMR studies, but are not part of wild-type <i>E. coli</i> BamC. Models for the full-length construct (residues 101–346) and for the independent domains (residues 101–212 and 229–346) in the N-terminus and the C-terminus, respectively, were generated with the iterative CS-RDC-NOE Rosetta protocol, as described previously (Warner et al., 2011; Shen et al., 2008). The NOE and RDC data had weights of 5 for the overall scoring in the low-resolution sampling stage and a weight of 0.1 in the all-atom sampling stage. The amide <sup>1</sup>H–<sup>1</sup>H NOE restraints were modeled with a flat-bottom potential, where no score penalty is introduced if the proton–proton distance is between 1.5 and 6.0 Å. Violations of the bounds are penalized quadratically until 0.5 Å above the upper bound, where the potential switches smoothly to a linear function.

2.3.1 Structure calculation using XPLOR-NIH

Structure calculations were performed using XPLOR-NIH to compare the robustness of CS-RDC-NOE Rosetta to incorrectly assigned NOEs with respect to traditional simulated annealing protocol. Structure calculations were performed on the BamC N-terminal domain with
XPLOR-NIH including: 1) the same dataset that was used in CS-RDC-NOE Rosetta, 2) a simulated dataset that included 1684 distance restraints that were measured from a target structure (one of the low energy final CS-RDC-NOE Rosetta structures), and 3) a simulated dataset that included 1684 distance restraints with 16% erroneous distances. The simulated datasets were generated using an in-house tcl script called ILV_methyl_Cdist.tcl (Appendix A1) that reads in atomic coordinates from a PDB file and outputs distances based on selected atomic types and distance thresholds. This set of restraints did not include any intrareidue NOEs and the restraints were input with high precision (± 0.1 Å) using the measured distance in the target structure.

BamC structures were generated with a standard simulated annealing protocol in XPLOR-NIH (Schwieters et al., 2006; 2003). Briefly, a linear polypeptide with randomized coordinates was generated from the sequence and subjected to a three step refinement procedure (Appendix A2): 1) a high temperature (3000 K) loop with NOE restraints but no van der Waals repulsion term, 2) a 100-step cycle with the van der Waals repulsion term turned on, and 3) a cooling loop. The weights for the NOE and van der Waals terms were incrementally raised in the second and third steps from 20 to 30 and 0.4 to 4.0, respectively. NOE violations are penalized quadratically past the upper and lower bounds.

2.3.2 Database search for structural homologues and conserved residues

To assess the phylogenetic conservation of surface residues on BamC, a multiple protein sequence alignment was generated using the program HMMer (jackhammer) (Eddy, 2009) with BamC_{101-344} as bait against the *E. coli* K-12 MG1655 KEGG database. The initial alignment was edited to include only sequences with the required N-terminal cysteine. An E-value threshold of 0.001 was used to generate the final sequence alignment that was submitted to the ConSurf
webserver (Landau et al., 2005; Ashkenazy et al., 2010; Glaser et al., 2003) (http://consurf.tau.ac.il/) to identify conserved residues.

2.3.3 **BamC<sub>101–344</sub> NMR titration experiments**

NMR titration experiments to probe for ligand binding functions of BamC were performed on <sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N BamC<sub>101–344</sub> with concentrations ranging from 0.05 to 0.1 mM at 500 and 600 MHz field strength and 30 °C. Palmitic acid and a mixture of *E. coli* polar lipids were tested based on the functional binding of structural homologues to BamC domains.

Palmitic acid was obtained from Sigma as a powder and suspended in a solution of ethanol, then titrated into a 0.05 mM <sup>1</sup>H, <sup>15</sup>N BamC<sub>101–344</sub> to a final concentration of 90 μM. An *E. coli* total lipid extract was obtained from Avanti Polar Lipids, Inc as a powder and suspended into ethanol, then titrated into 0.1 mM of <sup>1</sup>H, <sup>15</sup>N BamC<sub>101–344</sub> to a final concentration of 120 μM.

2.3.4 **Interaction studies with mature lipoproteins BamC<sub>25–344</sub> and BamD**

To test the interaction between mature lipoproteins, His-tagged, wild type BamC<sub>1–344</sub> and BamD including periplasmic localization sequence (pMS582 and pMS581 respectively) were expressed and subsequently purified from *E. coli* membranes. BamC was prepared with <sup>15</sup>N labeling as detailed in 2.2.1. Susan Baker (Sousa lab) has refined the membrane purification protocol and proteins were obtained from her. The details of the purification protocol are listed here for reference. Expression was performed as described in Section 2.2.1, but purification differed as follows. Cells were passed twice through a French press as 1000 psi. Cell debris was separated by centrifugation in a JA 20 18,900 x G for 20 minutes. The membranes were pelleted by ultracentrifugation at 100,000 x G for 30 mins. The pellet was then suspended in 50 mL of TBS pH 8.0, 2.0 % Triton X 100, 5 mM EDTA and 10 mg/ml lysozyme. A dounce homogenizer was used to break up the membrane. The membrane solution was allowed to rock for 2 h at room
temperature. The remaining debris was removed by ultracentrifugation at 100,000 X G for 30 min. The supernatant was dialyzed overnight in TBS, 0.5% Triton to remove EDTA for NiNTA purification. NiNTA beads (10 mL) pre-equilibrated in TBS and 0.5% Triton were added to the dialyzed membrane solution was allowed to rock for 2 h at room temperature. The beads were collected beads in a column and washed with 50 ml TBS and 0.5% Triton, then TBS and 0.05% DDM and finally eluted in 30 ml TBS, 0.05% DDM, 250 mM Imidazole. The sample was concentrated to 10 ml and run on an S200 16/60 column in TBS and 0.5% Triton. The major peak was pooled and concentrated to 0.5–1.0 mL. BioBeads, which bind detergent, were added to ~½ the volume for 2 h to reduce the concentration of DDM detergent. The BioBeads were filtered out and the protein solution was stored at -80 °C until further use.

2.4 Results and Discussion

2.4.1 Chemical shifts and protease assay show the first 75 amino acids of BamC25-344 are disordered.

The chemical shift is an exquisite probe of the chemical environment of NMR active nuclei and for years, there have been attempts to use the chemical shift to model protein structure and dynamics (Pardi et al., 1983; McDonald and Phillips, 1967; Wagner et al., 1983; Saitô, 1986). The correlation between chemical shift and local structure is well established for small molecules using quantum mechanical approaches (Schindler and Kutzelnigg, 1983; Ditchfield, 1972) but breaks down when applied to larger systems, such as proteins, due to the increase in computational complexity with an increase in the number of atoms (de Dios et al., 1993). However, empirical methodologies, pioneered in the mid 1990’s, use sophisticated search algorithms and reference databases to determine the secondary structure and dynamics of proteins using only the amino acid sequence and the chemical shifts of the backbone and $^{13}$C$^6$ atoms (Wishart et
The program TALOS+ (Torsion Angle Likeliness Obtained from Shift and Sequence Similarity) (Shen et al.)(Shen et al., 2009b) and RCI (Random Coil Index) (Berjanskii and Wishart, 2008) were used here to analyze the predicted secondary structure and backbone dynamics of BamC_{25-344}. The previously reported chemical shifts for BamC_{25-344} (Knowles et al., 2009) were used as input into TALOS+ (Shen et al., 2009b) and RCI (Berjanskii and Wishart, 2008). An analysis of the predicted secondary structure and dynamics of BamC as a function of residue number shows residues 26–75 are characterized as “Loop” by TALOS+ and dynamic by RCI as evidenced by predicted $S^2$ values less than 0.7 (Figure 2.3).

![Figure 2.3](image)

**Figure 2.3** The predicted order parameter $S^2$, calculated from the chemical shift using the random coil index.

The order parameter was predicted using index the backbone chemical shifts (Berjanskii, et al., 2005). Along the top of the schematic, the secondary structure as predicted from TALOS+ using the chemical shifts, is shown to indicate loops (red), strands (blue) and helices (green). The first 75 residues have a predicted $S^2$ of less than 0.7 and predicted loop secondary structure.

In support of this prediction, 35 kDa BamC_{25-344} was subjected to a limited aminopeptidase digestion (Petia Gatzeva-Topolova, unpublished). The resulting stable fragment was ~27 kDa. To test whether the first 75 amino acids alter the structure of the rest of the protein, hydrogen/deuterium (H/D) exchange NMR experiments were carried out on BamC_{25-344} and BamC_{101-}
The amide protons of the 75 N-terminal amino acids in the longer construct exchanged by the end of the first two-dimensional heteronuclear single-quantum coherence (HSQC) experiment (∼11 min), consistent with a disordered N-terminus (Figure 2.4). The exchange profiles of the remaining amide protons were essentially unchanged for the two constructs, indicating that the rest of the protein is unaffected by the presence of the 75-amino-acid N-terminal tail. Thus, a shorter construct (BamC<sub>101–344</sub>) was used for the structure determination and for the majority of the NMR studies. A $^1$H, $^{15}$N HSQC spectrum of BamC<sub>101–344</sub> is shown in Equation 2.5.

![Figure 2.4 2D $^1$H-$^{15}$N NMR spectra of mM BamC<sub>101–344</sub> and BamC<sub>25–344</sub>](image)

Spectra were collected at 30 °C on an INOVA at 600 MHz of (a) 1.0 mM BamC<sub>101–344</sub> 11 mins after exchanging into NMR buffer with 99.8% $^2$H<sub>2</sub>O and (b) 0.2 mM BamC<sub>25–344</sub> 11 mins after exchanging into NMR buffer with 99.8% $^2$H<sub>2</sub>O. Each spectrum was collected with the SOFAST $^{15}$N HMQC sequence (Brutscher et al., 2005) with a total experiment time of 4 min. BamC<sub>25–344</sub> signal intensity is lower due to the 5 fold lower concentration.

The backbone and $^{13}$C<sub>β</sub> chemical shifts of BamC<sub>101–344</sub> were assigned by Dr. Krisztina Varga using conventional heteronuclear triple-resonance NMR experiments, followed by partial assignment of side-chain resonances (Cavanagh et al., 2007). With the exception of residues near the N-terminus, no significant differences were observed when comparing the experimental backbone chemical shifts ($^1$H<sub>N</sub>, $^{15}$N<sub>H</sub>, $^{13}$C<sub>α</sub>, $^{13}$C<sub>β</sub> and $^{13}$C<sup>′</sup>) of BamC<sub>101–344</sub> to the previously assigned
BamC_{25-344} (Knowles et al., 2009). A complete list of the NMR resonance assignments for BamC_{101-344} are archived at the Biological Magnetic Resonance Data Bank under entry 17521.

![Figure 2.5 1H, 15N HSQC of 1.8 mM BamC_{101-344} at 30 °C at 800 MHz.](image)

The fingerprint $^1$H, $^{15}$N HSQC spectrum of BamC has 239 well-resolved peaks that correspond to all the predicted backbone amide peaks for the protein.

### 2.4.2 BamC has two well-defined domains connected by a flexible linker.

CS-RDC-NOE Rosetta calculations on BamC_{101-344} were carried out by Dr. Oliver Lange as described (Warner et al., 2011). The input consisted of the backbone ($^{1}$H$^N$, $^{15}$N, $^{13}$C$'$, $^{13}$C$^\alpha$, and $^{13}$C$^\beta$) chemical shifts for all residues (excluding the terminal and Pro amides), 156 $^1$H–$^{15}$N RDCs, and two different sets of NOE restraints. A preliminary calculation was performed with 62 $^1$H–$^1$H
amide–amide NOE distance restraints, but the low-scoring models consistently violated 10 NOEs. Analysis of the NOE spectra showed that these 10 NOEs were misassigned or ambiguously assigned (discussed in the text below). Thus, a calculation (BamC_I) was carried out with the remaining 52 amide–amide NOE restraints. This greatly improved convergence in both domains and reduced the overall Rosetta energy of the 10 lowest-scoring models (the scores were computed without NOE and RDC restraints) by 60 score units to a range of −556 to −565. An additional set of 97 $^1$H–$^1$H NOE restraints (for a total of 149 NOEs) involving backbone and side-chain protons, which were consistent with the BamC_I structures, was included in the final calculations (BamC_II), where the 10 lowest-energy structures had Rosetta energies ranging from −590 to −596.

The final CS-RDC-NOE Rosetta calculations indicated that BamC has two well-defined domains comprising an N-terminal domain from residues 101–210, a C-terminal domain from residues 229–344 and a linker region from residues 211–228. However, Rosetta was unable to define the relative orientation of these two domains (Figure 2.6). Two explanations for this conformational heterogeneity were explored here: 1) BamC$_{101-344}$ has a well-defined orientation of the two domains in solution, but the global energy minimum was not found during Rosetta calculations due to incomplete sampling of conformational space and 2) the solution orientation is not well-defined, and the residues in the linker region are flexible.
Figure 2.6 CS-RDC-NOE Rosetta calculation generated BamC_{101-344} structures with clearly defined domains, but orientations did not converge.

Three low-energy structures for BamC_{101-344}, shown as superimpositions of (a) their N-terminal domain (cool colors: cyan, marine and blue) and (b) C-terminal domain (warm colors: raspberry, red and gold), illustrating that the folds of the N-terminal and C-terminal domains converged in the Rosetta calculations while the relative orientations of the domains did not. These three structures represent the range of orientations for the two domains in the set of low-energy Rosetta structures.

2.4.3 Methyl groups used to search for long-range NOEs in BamC_{25-344}

We first looked for domain–domain NOEs that would reduce the conformational search. The methyl groups of Ile, Leu and Val were chosen as NMR probes to generate an additional set of NOE distances because, in the context of an otherwise deuterated protein, protonated methyl-groups benefit from a natural TROSY effect when observed using an HMQC-type pulse sequence (Ollerenshaw et al., 2003). This leads to significant enhancement of signal-to-noise and greater peaks resolution when compared to standard methods (Ollerenshaw et al., 2003). Additionally, with all of the aliphatic groups deuterated, the only mode for spin diffusion in a methyl NOESY experiment is limited to protonated methyl groups and backbone amides (in a 90% H2O solution). Spin diffusion can be limited further if the sample is prepared in an ∼100% D2O solution; this sample was prepared in a 90% H2O solution to observe methyl-amide NOEs. Methyl-methyl NOE distances up to 8 Å have been observed in methyl-protonated samples, such
as the 42 kDa maltodextrin binding protein (Mueller et al., 2000). Only a few of the methyl groups in BamC were previously assigned using the traditional side-chain assignment approaches. Therefore, the methyl $^1$H, $^{13}$C chemical shifts of Ile, Leu and Val were assigned using specific labeling and NMR experiments described by Tugarinov and co-workers (Tugarinov and Kay, 2003).

First, a $^1$H, $^{13}$C HMQC spectrum was collected to assess the peak dispersion and signal-to-noise. Due to isotopic labeling with a racemic mixture of 2-ketobutyric acid-$^{13}$C$_4$,3,3-d$_2$ sodium salt (with the methyl groups containing $\sim50\%$ [R-$^{13}$CH$_3$, S-$^{12}$CD$_3$] and $\sim50\%$ [S-$^{13}$CH$_3$, R-$^{12}$CD$_3$] enantiomers, (Figure 2.1b), the $^{13}$CH$_3$ and $^{12}$CD$_3$ groups are incorporated nonstereospecifically into the R and S methyl positions of Leu and Val. Thus, for every Leu or Val, there are two methyl peaks observed with 50% of the peak intensity expected, corresponding to the R and S enantiomers. Figure 2.8 shows a high quality, HMQC spectrum of methyl-protonated ILV BamC$_{25-344}$, with good peak dispersion of 98 of the 115 predicted peaks. During the carbon evolution period, no decoupling sequence is applied and a 30 Hz splitting of the peaks is observed that corresponds to the $J_{cc}$ for methyls to aliphatics. Five peaks that are $\sim3\%$ of the intensity of Ile peaks, marked with an asterisk in Figure 2.7, are observed. These peaks do not show any splitting and have chemical shifts that are consistent with methionine methyl groups. This background labeling is caused by isotopic labeling with $\sim97\%$ $^2$H, $^{13}$C glucose, where 3% background $^1$H, $^{13}$C labeling of sidechain residues is expected; these background peaks did not hinder spectral analysis. Next, HMCMCBCA and HMCMCGCBCA spectra were collected that correlate the methyl spins ($^1$H$_{methyl}$, $^{13}$C$_{methyl}$) to the aliphatic spins ($^{13}$C$_{\alpha}$, $^{13}$C$_{\beta}$ and $^{13}$C$_{\gamma}$) (Tugarinov and Kay, 2003). These out-and-back pulse sequences relay magnetization along a ‘linearized’ spin system as described in 2.2.4 and shown in Figure 2.2. In the HMCMCBCA, each Ile $^{13}$CH$_3$,
δ1 and Leu $^{13}\text{CH}_3$, γ shows correlations for $^{13}\text{C}^\beta$ and $^{13}\text{C}^\gamma$ while Val $^{13}\text{CH}_3$, γ methyl peaks have correlations for $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$. Since the $^{13}\text{C}^\alpha$ resonances for Ile and Leu are not observed in the HMCMMCBA, the spectrum from this pulse sequence are useful to discriminate methyl types in the crowded Leu/Val methyl region.

![Figure 2.7](image)

**Figure 2.7 $^1\text{H},^{13}\text{C}$ HMQC spectrum of 1.0 mM BamC at 30 °C and 800 MHz.**
The $^{13}\text{C}_{\text{methyl}}$ chemical shift of Ile is distinct from Leu and Val. The peaks in this spectrum are coupled to $^{13}\text{C}_{\text{aliph}}$ due to specific isotopic labeling of the sidechain aliphatic carbons with results in an ~30 Hz ($^1J_{\alpha\alpha}$) splitting.

However, the HMCMMCGBCA spectrum was most useful in assigning the methyl spins as demonstrated in Figure 2.8, where strip plots show the $^{13}\text{C}_{\text{aliphatic}}$ region for several Leu and Val residues. The $^1\text{H}_{\text{methyl}}$, $^{13}\text{C}_{\text{methyl}}$ spins were then assigned by correlating the previously assigned $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$ chemical shifts to the experimentally measured ones from the HMCMMCGBCA spectrum. A total of 88/98 of the non-branched methyl groups of Ile, Leu, and Val in the
BamC<sub>25-344</sub> were assigned. The remaining peaks arise from residues in the disordered N-terminal tail, where the $^{13}\text{C}^\alpha$ and $^{13}\text{C}^\beta$ chemical shifts were very similar and ambiguously assigned.

To search for interdomain NOEs that would reduce the conformational search for Rosetta, we acquired 3D ($^{13}\text{C}, ^{13}\text{C}, ^1\text{H}$) and ($^1\text{H}, ^{13}\text{C}, ^1\text{H}$) HMQC-NOESY-HMQC spectra on BamC<sub>25-344</sub>. Although 103 intradomain methyl-methyl NOEs were observed, no NOEs were detected between the N-terminal and C-terminal domains. While the lack interdomain NOEs suggests that the N-terminal and C-terminal domains do not form a stable interaction, but this alone does not prove that there is no interaction or defined orientation between the domains.
Figure 2.8 Strip plot of HMCMCGCBCA spectrum with $^1$H, $^{13}$C HMQC overlaid.

Peaks from five Val and three Leu residues are shown in strips. The methyl peaks (black contours) have carbon chemical shifts $\sim$20 ppm. The peaks with positive phase are shown in blue contours, and negative phase in red contours. The phase changes sign with each magnetization transfer, so $^{13}$C$^\alpha$ and $^{13}$C$^\beta$ peaks have alternating phase for Val and Leu.

2.4.4 RDCs indicate domains orient independently in solution

RDC data are routinely used to determine the relative orientation of well-defined domains in proteins or nucleic acids (Fischer et al., 1999; Mollova et al., 2000). The RDCs used in the
CS-RDC-NOE Rosetta calculations of BamC\textsubscript{101–344} were used primarily to guide the conformational selection and were subsequently de-weighted during the final all-atom refinement step of the calculate as detailed in Section 2.2.5. De-weighting the RDCs during the final refinement abolished any capability of the RDCs to orient the domains. Thus, we wanted to address whether the \textsuperscript{1}H–\textsuperscript{15}N RDCs could be directly used to define the orientation of the N-terminal and C-terminal domains in BamC. The first step in domain orientation using RDC data is to assess whether the two domains have a fixed orientation and, therefore, the same alignment tensor. One indication of a rigid orientation for two domains is if the individual domains have similar values for the magnitude \(D_a\) and the rhombicity \(R\) of their alignment tensors (Fischer et al., 1999). A structure-independent method, which involves an analysis of the shape of the histograms of the RDCs in a molecule, was used here to compare \(D_a\) and \(R\) in the two domains (Clore et al., 1998a).

The positions of three extrema on the histogram of the experimental RDCs (Figure 2.9), give the principal components of the alignment tensor, \(D_{xx}, D_{yy}\) and \(D_{zz}\), where the assumptions are that \(|D_{zz}| > |D_{yy}| > |D_{xx}|\) and \(D_{zz} + D_{yy} + D_{xx} = 0\). \(D_a\) and \(R\) can be calculated using Equations 2.2–2.4 (Clore et al., 1998a) (which are rearrangements of Equation 1.3 and Equation 1.4, taking into account the \(\frac{1}{2} D_{\text{max}}\) prefactor and assumes the extreme case when \(\theta = 90^\circ\) and \(\phi = 0^\circ\)):

\[\text{Equation 2.2}\]
\[D_{zz}^{NH} = 2D_a^{NH}\]

\[\text{Equation 2.3}\]
\[D_{yy}^{NH} = -D_a^{NH}(1 + 1.5R)\]
If the N-terminal and C-terminal domains in BamC had a fixed orientation, the histograms of the individual domains would have similar shapes, indicating similar values for $D_a$ and $R$, assuming that the set of RDCs adequately samples all orientations of bond vectors. The histograms of the 86 and 68 $^1$H–$^{15}$N RDCs for the N-terminal and C-terminal domains have quite different shapes (see Figure 2.9), with similar $D_a$ but very different $R$. This suggests that the domains do not have a fixed orientation in solution. This analysis assumes that the $^1$H–$^{15}$N RDCs represent an adequate distribution of bond orientations, which is challenging using only $^1$H–$^{15}$N RDCs. Analysis of $^1$H$\alpha$–$^{13}$C$\alpha$, $^{13}$C$\alpha$–$^{13}$C', and $^{15}$N$^\text{NH}$–$^{13}$C' RDCs would provide a more rigorous, structure independent determination of the principal values of the alignment tensor. However, the experiments used to measure these additional RDCs are significantly less sensitive. Therefore, an analysis of the backbone $^{15}$N relaxation were carried out to determine if the linker is flexible and the domains tumble independently.

**Equation 2.4**

\[
D_{xx}^{NH} = -D_a^{NH}(1 - 1.5R)
\]
Figure 2.9 Histograms of the RDCs measured in BamC.  
86 and 68 \(^1\text{H}-\text{\textsuperscript{15}N}\) RDCs for the (a) N-terminal and (b) C-terminal domains in BamC\textsubscript{101–344}. The shape of the histogram provides information on the magnitude \(D_a\) and the rhombicity \(R\) of the alignment tensor (Clore et al., 1998). The differences in the shapes of the two histograms indicate that the N-terminal and C-terminal domains have different alignment tensors and, therefore, do not have a fixed orientation in solution (assuming that these RDCs adequately sample all orientations of \(^1\text{H}–\text{\textsuperscript{15}N}\) bond vectors).

### 2.4.5 \(^{\text{15}}\text{N}\) relaxation and hydrogen/deuterium exchange show linker region between N- and C-terminal domains is flexible.

To experimentally test whether the linker is flexible, \(^{\text{15}}\text{N}\) relaxation NMR experiments were performed to directly probe backbone dynamics in BamC. Heteronuclear \({\text{\textsuperscript{1}H}}-{\text{\textsuperscript{15}N}}\) NOEs were measured for the backbone amides in BamC\textsubscript{101–344}. As seen in Fig. Figure 2.10a, residues in the N-terminal and C-terminal domains generally had NOE values above 0.7, whereas residues 214–227 in the linker had significantly lower \({\text{\textsuperscript{1}H}}-{\text{\textsuperscript{15}N}}\) NOE values (0.13–0.42), indicating a
flexible linker. A similar conclusion is obtained by predicting the order parameter of the backbone amide group $S^2$ from the chemical shifts. As seen in Figure 2.10b, residues 217–228 in BamC show lower predicted $S^2$ values, supporting the hypothesis that the linker region is flexible.

H/D exchange experiments provided additional evidence for N-terminal and C-terminal domains connected by a conformationally dynamic linker in BamC$_{101-344}$, where slowly exchanging amide protons were observed for many residues in the N-terminal and C-terminal domains but rapid exchange was observed for all the amide protons in the linker (Figure 2.11). Moreover, it was recently shown that BamC$_{101-344}$ is susceptible to cleavage by subtilisin in the linker region, yielding two stable fragments of 12.2 and 14.5 kDa (Albrecht and Zeth, 2010). These results support a model where BamC$_{101-344}$ contains well-ordered N-terminal and C-terminal domains connected by a flexible linker. Interestingly, the CS-RDC-NOE Rosetta calculations on BamC$_{101-344}$ consistently generated an $\alpha$-helical conformation for the linker region (Figure 2.6 and Figure 2.11). However, the NMR data demonstrate that the linker is flexible. This could be an indication that there is a propensity for this linker to form a helix, but the TALOS+ analysis of the chemical shifts suggested that this region was most likely a loop. Moreover, Rosetta has a tendency to “overfold” dynamic regions, likely resulting from backbone fragments being selected from a database of known well-ordered protein structures. Methods to reduce overfolding of regions experimentally identified as dynamic are currently being developed for Rosetta (Wang et al., 2011).
Figure 2.10 The $^{15}$N relaxation data on BamC$_{101-344}$ show regions of increased backbone flexibility.

Plots of (a) $^{15}$N{$^1$H} NOE, (b) the predicted backbone order parameter $S^2$, (c) the calculated residue-specific rotational correlation time $\tau_c$, (d) the $^{15}$N $R_1$ and (e) the $^{15}$N $R_2$ as a function of residue number. $S^2$ is predicted from experimental chemical shifts using a random-coil index (Berjanskii, et al., 2008). The red circles highlight two regions of the molecule that exhibit increased dynamics (see the text). The regions of regular secondary structure are illustrated at the top of the figure.
Hydrogen deuterium exchange profiles mapped onto a low-energy structure of BamC<sub>101–344</sub>.

Residues in red, white, and blue indicate rapid exchange (<12 min), moderate exchange (12 min to 24 h), and slow exchange (> 24 h) at 30 °C, with gray indicating prolines or residues where exchange could not be determined due to spectral overlap. Black boxes highlight rapid exchange for part of the α1 helix and the linker connecting the domains. Note that Rosetta modeled most of the linker region as a helix even though the <sup>15</sup>N relaxation and H/D exchange data indicate that this region is conformationally dynamic (see text).

Although the {<sup>1</sup>H}<sup>15</sup>N heteronuclear NOE and H/D exchange data show a high level of flexibility for the linker, analysis of RDCs suggests that the domains orient independently, and no interdomain NOEs were observed, this does not rule out a stable interaction between the N-terminal and C-terminal domain in full-length BamC. To address this, the residue-specific rotational correlation times, $\tau_c$, were calculated from <sup>15</sup>N $R_1$ and $R_2$ measurements (Figure 2.10c, d and e respectively) on BamC. If the two domains tumble as a rigid single species in solution, then the rotational correlation times of the individual domains should be the same as that predicted for the full protein. The residues in the N-terminal and C-terminal domains have average $\tau_c$ values of 10.8 and 9.8 ns, respectively, whereas the linker has lower $\tau_c$ values (ranging from 4.6 to 7.4 ns for residues 214–227). The relationship between molecular weight and $\tau_c$ is given by the Stokes-Einstein Equation (Equation 2.5).
where $\eta$ is the viscosity and $r_H$ is the effective hydrodynamic radius of the protein, $k$ is the Boltzmann constant and $T$ is the temperature. The effective hydrodynamic radius can be approximated from the molecular mass of the protein, $M$, assuming a specific volume of the protein is $\bar{V}$ 0.73 cm$^3$/g and a hydration layer, $r_w$ of 1.6 to 3.2 Å (≈ one half to one hydration shell) (Venable and Pastor, 1988) by Equation 2.6 (Cavanagh et al., 2007).

$$\tau_c = \frac{4\pi\eta r_H^3}{3kT}$$

However, an estimate of $\tau_c$ using this method is often inaccurate because the effective hydrodynamic radius assumes a spherical approximation. Thus, an empirical relationship between molecular weight and $\tau_c$ was made using experimentally measured $\tau_c$ from known monomeric NESG (Northeast Structural Genomics Consortium) targets (http://www.nmr2.buffalo.edu/nesgwiki/NMR_determined_Rotational_correlation_time).

The values of $\tau_c$ for the N-terminal and C-terminal domains are $\sim$35% smaller than what would be predicted for a 27 kDa spherical protein ($\sim$16 ns), but larger than what would be predicted if the N-terminal and C-terminal domains were tumbling independently as $\sim$13 kDa spheres ($\sim$7.6 ns). This pattern of $\{^1\text{H}\}^{-15}\text{N}$ heteronuclear NOE and $\tau_c$ values is very similar to what was previously observed in Ca$^{2+}$-loaded calmodulin, which has a flexible linker connecting its N-terminal and C-terminal domains (Barbato et al., 1992). These data are consistent with the
conclusion obtained from the RDC data that N-terminal and C-terminal domains in BamC have unique alignment tensors, do not have a fixed orientation in solution and do not tumble as a rigid species in solution.

The CS-RDC-NOE Rosetta calculations on the full-length BamC assumed a single alignment tensor for the whole molecule, but this is incorrect if domains do not have a fixed orientation. Thus, separate CS-RDC-NOE Rosetta calculations were performed for the N-terminal (residues 101–212) and C-terminal (residues 229–344) domains (Figure 2.12 and Table 2.1). There was no significant difference in the overall folds of the individual domains in the Rosetta calculations performed on the full protein or on separate domains. Thus, all further analyses were performed on the structures generated by the calculations for the individual domains.

Figure 2.12 Ensemble of the structures of the N-terminal and C-terminal domains of BamC generated by separate CS-RDC-NOE Rosetta calculations for the two domains.

Superimposition of nine low-energy structures of the (a) N-terminal (PDB: 2LAF, (Warner et al., 2011) and (b) C-terminal (PDB: 2LAE, (Warner et al., 2011) domains of BamC. Part of the α1 helix in the N-terminal domain that shows conformational flexibility in the $^{15}\text{N}$ relaxation data (see the text) is highlighted by a black box.
2.4.6 The N- and C-terminal domains of BamC adopt a helix-grip motif

Results from the CS-RDC-NOE Rosetta calculation showed that the N-terminal domain of BamC is composed of two α-helices packed against a five-stranded anti-parallel β-sheet (Figure 2.13a), reminiscent of the helix-grip fold. The C-terminal domain closely resembles the N-terminal domain, with two additional structural elements (Figure 2.13b); a short β-strand before β1 and a seven-residue helix inserted between strands β3 and β4, where the corresponding residues in the N-terminal domain form an extended loop. The N-terminal domain superimposes on the C-terminal domain with an average pairwise rmsd of 1.7 Å for the Cα backbone for 60 residues in structurally similar regions identified using LSQMAN (Kleywegt, 1996), indicating a possible gene duplication event.

The boxed region in Figure 2.12a shows a region of BamC that is not uniquely defined in the Rosetta calculations, where some structures show a kink in the helix near Pro117 and others exhibit fraying of the N-terminal residues in this helix. The \( ^1\text{H} - ^{15}\text{N} \) heteronuclear NOE, predicted \( S^2 \), and \( \tau_c \) values for residues 112–117 are lower than those of the residues in the domains (Figure 2.10) and furthermore, the amide protons for these residues exchanged rapidly in the H/D exchange experiments (Figure 2.11). A slow interconversion between cis and trans prolyl imide bond conformers could explain the conformational heterogeneity observed in the α1 helix. However, cis/trans isomerization is generally slow on the NMR timescale, causing amide backbone NH spins to have two peaks (Wüthrich, 1986). However, for all of the peaks in the \( ^1\text{H}, ^{15}\text{N} \) HSQC, only one peak was observed. The \( ^{15}\text{N} \) relaxation and H/D exchange data support a model where the N-terminal part of the α1 helix is conformationally dynamic, consistent with the conformational variation of this region in the Rosetta structures. These results demonstrate
how NMR relaxation data can be used to determine whether non-converged regions observed in the Rosetta calculations reflect true conformational heterogeneity of the protein.

### Table 2.1 Structural statistics of the BamC N- and C-terminal domains

Statistics are given for the 9 lowest energy structures after Rosetta refinement out of 500 calculated.

<table>
<thead>
<tr>
<th></th>
<th>N-terminal domain</th>
<th>C-terminal domain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of residues</strong></td>
<td>112</td>
<td>118</td>
</tr>
<tr>
<td><strong>NOE-based distance restraints</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOE distance restraints (viol. ≥ 0.5 Å)</td>
<td>69 (16 ± 1)</td>
<td>78 (30 ± 2)</td>
</tr>
<tr>
<td>Number of restraints per residue</td>
<td>0.62</td>
<td>0.66</td>
</tr>
<tr>
<td><strong>Other restraints</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>φ + ψ dihedral angle restraints (viol. ≥ 5°)</td>
<td>193 (11 ± 3)</td>
<td>202 (11 ± 4)</td>
</tr>
<tr>
<td>Residual dipolar coupling restraints (viol. ≥ 5 Hz)</td>
<td>60 (24 ± 3)</td>
<td>82 (20 ± 4)</td>
</tr>
<tr>
<td><strong>Average RMSD to the average structure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backbone (Å)</td>
<td>1.39 ± 0.33</td>
<td>0.54 ± 0.06</td>
</tr>
<tr>
<td>Heavy atom (Å)</td>
<td>1.88 ± 0.30</td>
<td>0.95 ± 0.11</td>
</tr>
<tr>
<td><strong>Ramachandran plot</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favored regions</td>
<td>92.2%</td>
<td>94.3%</td>
</tr>
<tr>
<td>Allowed regions</td>
<td>7.8%</td>
<td>4.8%</td>
</tr>
<tr>
<td>Generously allowed regions</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Disallowed regions</td>
<td>0.0%</td>
<td>0.9%</td>
</tr>
</tbody>
</table>

"a All NOEs are long range (|i-j| ≥ 5). Two of the 149 NOEs were between the amino acids in the linker region and N-terminal domain and were not used in the calculations of the individual domains.

"b Torsion angle restraints were derived from TALOS+ (Shen et al., 2009).

"c PROCHECK was used to determine the quality of the structures (Laskowski et al., 1996).

### 2.4.7 Validation of BamC$_{101-344}$ structures

Structure validation is an important part of structure determination regardless of the methods used. The standard quality indicator in X-ray structure determination is the $R$-factor, which is defined as the difference between experimentally observed data and those back-calculated from the structural model. To avoid “overfitting” data, it is standard practice to separate the diffraction data into a “working” and “test” dataset that comprise ~90% and ~10% of the total data respectively. The difference between the test dataset and a dataset back-calculated from the structural model is coined the free $R$-factor (Brunger, 1992). A good X-ray structure will have a low $R$-factor and a free $R$-factor that is close to the $R$-factor. There have
been efforts to apply this same type of validation to NMR structures, which include cross-validation using NOEs and RDCs. In X-ray data, a single diffraction spot contains information about the structure of the entire molecule, whereas a single NOE only gives distance information for a pair of hydrogens. In addition, the intensity of the NOE is not only function of distance, but also of local dynamics, thus complicating its use for cross-validation (Brünger et al., 1993). It has been proposed that a “complete” cross-validation be performed by using a series of “working” and “test” NOE datasets that are randomly selected from the original NOE distance data and combined to calculate an average free $R$-factor, thereby overcoming the limited information contained in a single NOE (Brünger et al., 1993). However, this analysis is limited to very complete datasets, with the analysis breaking down in cases of fewer than two distance restraints per residue (Brünger et al., 1993). Thus, this method could not be applied to the BamC dataset, which has only ~0.6 NOE distance restraints/residue for both domains. The approach has similarly been applied to the RDC and could in principle be applied to the BamC structures. However, the NOEs and RDCs are used simply to guide the selection process and are de-weighting in the final low-resolution and all-atom refinements during the Rosetta calculations. Thus, the structures are not restrained against NOEs and RDCs and the resulting $R$-factor would not be a good reporter of the quality of the BamC structural models.

An analysis of the N- and C-terminal domains of BamC with PROCHECK (Laskowski et al., 1996) indicated that both domains are well within the bounds of the Ramachandran plot (Table 2.1). PROCHECK (Laskowski et al., 1996) compares the backbone and sidechain geometries of the input protein to high resolution crystal structures found in the Protein Data Bank. However, using geometric quality as a standard validation technique is problematic for structures calculated using Rosetta. This is due to the fact that Rosetta queries the same structural database
that is used to by the validation software. This has become an issue in validating structures submitted to the CASD-NMR (Critical Assessment of Automated Structure Determination of Proteins from NMR data), where Rosetta always has high scores for geometric quality, even when the structure is clearly wrong (personal correspondence with Geerten Vuister; http://nmr.cmbi.ru.nl/CASD-NMR/).

It was therefore necessary to validate the BamC101–344 structures using a set of supplementary data not included in the Rosetta calculations for BamC101–344. These data consisted of additional backbone–backbone and methyl–methyl ¹H–¹H NOEs. The input NOE dataset used for the Rosetta calculations comprised 149 readily assigned long-range ¹H–¹H NOEs (between residues i and j, where |i – j| >5). These NOEs were not regularly distributed over the structure, where some regions have clusters of NOEs and others have no NOEs. The gray boxes in Figure 2.13 highlight regions of anti-parallel β-sheet secondary structure where no cross-strand NOEs were included in the Rosetta calculations. For example, β7 and β11 were unrestrained by cross-strand NOEs, which could lead to ambiguity in the register or orientation of the two strands. This lack of NOE restraints provided a valuable opportunity to directly validate the Rosetta models with supplementary data. Further analysis of the original NOESY spectra yielded three cross-strand backbone–backbone NOEs in this region (HN V244 and HN S314, Ha V243 and Ha Ser315, and HN L316 and HN L242), which unambiguously confirmed the strand orientation and register generated by Rosetta between β7 and β11 (Figure 2.13b). The cross-strand Ha–Ha NOE is especially diagnostic of an anti-parallel β-sheet because of the relatively short Ha–Ha distance (~2.2–2.5 Å). Strong Ha–Ha cross-peaks, as well as additional cross-strand ¹H–¹H NOEs, were observed for the two other boxed regions in Figure 2.13a, validating the β-sheet secondary structure and strand register generated by Rosetta.
Figure 2.13 Backbone–backbone $^1$H–$^1$H NOEs confirm the β-sheet secondary structure in BamC generated by CS-RDC-NOE Rosetta.

Schematics for the β-sheet secondary structures in the (a) N-terminal and (b) C-terminal domains showing NOEs included (black) or not included (red) in the CS-RDC-NOE Rosetta calculations. NOEs in red were used to help validate the Rosetta structures. Gray boxes indicate regions of secondary structure that were not restrained by experimental NOEs during CS-RDC-NOE Rosetta calculations.
One of Rosetta's strengths is that it can generate accurate side-chain packing in cases of high backbone convergence (< 2 Å) (Raman et al., 2010; Misura et al., 2006). The ILV methyl-protonated NOESY spectra provide information on the packing of the ILV side chains. Figure 2.15 shows a projection of the methyl specific $^1$H, $^{13}$C, $^1$H and $^{13}$C, $^1$H, $^{13}$C NOESY spectra that were used to measure a set of 103 NOEs observed between the ILV methyl groups in BamC; only three of these were included as restraints in the Rosetta calculations. Figure 2.14 shows the network of experimental methyl NOEs observed in the hydrophobic cores of the N-terminal and C-terminal domains of BamC. In all cases, the range of distances predicted from these methyl–methyl NOEs were consistent with the hydrophobic packing observed in the BamC structures generated by Rosetta.

**Figure 2.14 A network of methyl-methyl NOEs confirms hydrophobic packing.**
A network of 103 experimental methyl NOEs detected with a methyl protonated were observed in the hydrophobic cores of the N-terminal (a) and C-terminal (b) domains of BamC.
2.4.8 Helix-grip domains of BamC101–344 are remarkably similar to X-ray structures

A recent report of the crystal structures of the two domains of BamC (Albrecht and Zeth, 2011) shows remarkable similarity to the CS-RDC-NOE Rosetta structures reported here (Warner et al., 2011). Interestingly, Albrecht and Zeth were unable to crystallize any constructs that contained the linker between the N- and C-terminal domains and found that BamC101–344 is susceptible to cleavage by subtilisin in the linker region, yielding two stable fragments of 12.2 and 14.5 kDa (Albrecht and Zeth, 2010). These fragments were successfully crystallized as BamC101–210 and BamC226–344 (Albrecht and Zeth, 2011). Figure 2.15 shows a superimposition of the crystal structure for each domain with the CS-RDC-NOE Rosetta ensemble. A pairwise superimposition of the crystal structures to low energy Rosetta ensemble results in rmsds of 2.0 ± 0.3 Å for the N-terminal domain and 1.6 ± 0.4 Å for the C-terminal domain. The superimpositions are within the average ensemble rmsd of the structure determined with CS-RDC-NOE Rosetta, also validating our results. Interestingly, the coordinates for residues 111–113 are missing from the crystal structure of BamC (Figure 2.16) (Albrecht and Zeth, 2011). This same region of the α1 helix in the N-terminal domain of BamC showed low backbone convergence in the CS-RDC-NOE Rosetta structures as well as {1H}-{15}N heteronuclear NOE values less than 0.7 (Figure 2.10)
Figure 2.15 The CS-RDC-NOE Rosetta structural ensemble superimposes well with X-ray crystal structure of BamC.

BamC_{101-344} N-terminal (a, blue, PDB: 2LAF (Warner, et al., 2011)) and C-terminal domains (b, blue, 2LAE (Warner et al., 2011)), superimpose well with crystal structures of the BamC_{101-212} (a, orange, PDB: 2YH6 (Zeth and Albrecht, 2011)) and BamC_{202-346} (b, orange, PDB: 2YH5 (Zeth and Albrecht, 2011)). The crystal structures align to the Rosetta ensembles with rmsds of 2.0 ± 0.3 Å for the N-terminal domain and 1.6 ± 0.4 Å for the C-terminal domain.

Figure 2.16 Region of the BamC α1 helix with dynamic properties has a zero occupancy in the crystal structure.

The coordinates from residues 110-114 are missing from the crystal structure of the N-terminal domain of BamC (gray, PDB: 2HY5 (Zeth and Albrecht, 2011)). This same region of the α1 helix showed low backbone convergence for the CS-RDC-NOE Rosetta ensemble (blue) and \{^1H\}-^{15}N heteronuclear NOE values less than 0.7.
2.4.9 CS-RDC-NOE Rosetta vs. traditional simulated annealing

A valuable feature of using CS-RDC-NOE Rosetta to generate protein structures from NMR data is its ability to deal with incorrectly assigned NOEs, which can lead to distance restraints that are inconsistent with the correct structure. This can be a major problem when generating structures from NMR data, where inconsistent distance restraints lead to high energies and can drive the conformation away from the correctly folded structure during refinement. Rosetta is not as susceptible to such problems because, as previously discussed, there is generally a low probability that inaccurate experimental restraints will also yield low-energy structures (Raman et al., 2010). Furthermore, as noted in Materials and Methods, the weightings for the NOE and RDC restraints are reduced from 5.0 to 0.1 in the all-atom refinement and Rosetta uses a linear potential for violations of NOE restraints for large distance violations. The XPLOR-NIH calculations used NOE restraints that increased from 20 to 30 (these are relative weighting factors and should not be compared to weightings used in Rosetta) in the cooling cycle and violations were penalized with a quadratic potential. The results here demonstrate that CS-RDC-NOE Rosetta is robust to inconsistent NOEs. As seen in Figure 2.17, a preliminary calculation that included 10 misassigned NOEs generated similar structures for the N-terminal and C-terminal domains as the final structures, even though eight of the NOEs violated their restraint by ~12 Å.

Rosetta is able to handle inaccurate data because the experimental data are primarily used to increase the efficiency of searching for low-energy conformations (Raman et al., 2010). To test the effect of having this level of misassigned NOEs on a standard structure calculation, we performed XPLOR-NIH (Schwieters et al., 2006) simulations using the N-terminal domain of one of the final Rosetta structures as the target. A set of 1684 NOE restraints was created using
an in-house tcl script (Appendix 1.1) that takes a PDB file an input and generates a list of predicted NOEs based on a threshold distance cutoff. This set of restraints did not include any intraresidue NOEs and the restraints were input with high precision (± 0.1 Å) using the measured distance in the target structure. This procedure generated an ensemble of 100 structures with low NOE energies and small rmsds to the target structures with an average rmsd 0.34 ± 0.14 Å for the 10 lowest energy structures to the target (Figure 2.18a).

![Figure 2.17 CS-RDC-NOE Rosetta modeling is tolerant of misassigned NOEs.](image)

The best-scoring model in a preliminary calculation that included 10 misassigned NOEs is shown in blue. Superimposed in gray are the individual domains of the best-scoring model of the final calculation, BamC_II. Amide H^N–H^N NOEs used as restraints in the preliminary calculation are shown as arrows. Arrows are shown in green if the NOE is within the distance bound of the restraint and in red if it violates the restraint. The red arrows are between residues 53–15, 92–113, 92–191, 101–174, 142–155, 149–194, 183–207, 197–207, 198–207, and 242–237. Figure adapted from O. Lange.

To mimic the same percentage of misassigned NOEs as observed in the preliminary Rosetta calculation (10 out of 62), 16% of the simulated NOEs were misassigned by randomly changing the residue number for one partner in the NOE. Calculations with these misassigned NOEs generated an ensemble of 100 structures with very high NOE energies and large rmsds to the target structure with an average rmsd 15.5 ± 2.0 Å for the 10 lowest energy structures to the
target (Figure 2.18b). The violations to NOE restraints in the XPLOR-NIH calculations were treated with a square potential. These results demonstrate that incorrect NOE assignments have much less influence on structures generated by CS-RDC-NOE Rosetta than a standard NMR structure determination. On the other hand, as seen in Table 2.1, the percentages of NOE and RDC violations are much greater than those observed in standard NMR structure determination. This results from the reduction of the weights of the NOE and RDC restraints during the all-atom refinement. Therefore, the CS-RDC-NOE Rosetta ensemble will generally be of lower resolution than an NMR ensemble generated with thousands of NOE restraints. However, proteins greater than \( \sim \)15 kDa generally require extensive deuteration, which drastically reduces the number of NOEs. Hence, the approach used by CS-RDC-NOE Rosetta, which combines limited experimental data with efficient computational methods for predicting protein folding, will often have a clear advantage when applied to large proteins if the fold is well represented in the database.

(a)  
(b)

**Figure 2.18 CS-RDC-NOE Rosetta is more tolerant of misassigned NOEs than traditional simulated annealing.**  
Superimpositions of the 10 lowest energy structures (blue) from the XPLOR-NIH calculations using simulated NOEs on the target structure (red) are shown with (a) 0% and (b) 16% misassigned NOE restraints.
2.4.10 Proposed function of BamC based on structural analysis, binding data and new insights from BamC/D X-ray crystal structure.

The helix-grip motif of the Bet v 1 superfamily was identified as a structural homologue for both the N-terminal and C-terminal domains of BamC from independent searches of the Dali protein structural database (Holm and Rosenström, 2010). Figure 2.19a shows the superimposition of the N-terminal domain of BamC with the major latex protein At1g24000.1 (Lytle et al., 2009). Proteins with the Bet v 1 motif are known to bind a diverse set of hydrophobic ligands such as membrane lipids, plant hormones, and steroids (Radauer et al., 2008).

Figure 2.19 The helix-grip motif is found in the major latex proteins and kinase associated domains.

The N-terminal domain of BamC (gray) is shown superimposed with structurally similar proteins (blue) identified in a Dali search. (a) The major latex protein At1g24000.1 (PDB 1VJH (Song, et al., 2004)), (b) the yeast septin-associated Kcc4p (PDB 3OSM (Moravcevic, 2010)), and (c) the helix–grip domain of the α-subunit of AMPK (PDB 2V8Q (Xiao et al., 2007)) all share similar helix–grip motifs with the BamC domains. The superimposition in (c) is rotated to highlight the protein–protein interaction surface between the α-subunit and the β/γ-subunits of AMPK (cyan; in surface representation). The central helix in the α-subunit of AMPK is important for protein–protein interaction. The residues used to align the proteins were defined by LSQMAN (Kleywegt, 1996).

Structurally, these proteins are characterized by a ligand-binding hydrophobic cavity located between the β-sheet and a long α-helix. However, this cavity is not present in the Rosetta
structures of either domain of BamC. Consistent with this observation, preliminary lipid binding studies performed with \(^{1}\text{H},^{15}\text{N}\) BamC\(_{101-344}\) and palmitic acid (40 \(\mu\text{M}\)) or an extract of \(E.\text{coli}\) polar lipids (120 \(\mu\text{M}\)) showed no changes in the amide chemical shifts in of a 100 \(\mu\text{M}\) sample of \(^{1}\text{H},^{15}\text{N}\) BamC (data not shown).

The Dali search also identified KA-1 (kinase-associated) domains are structurally related to both the N-terminal domains and the C-terminal domains of BamC (Figure 2.19b and c). It was recently shown that, in some kinases such as human MARK (microtubule affinity regulating kinase)/Par1 (partitioning-defective 1 kinase) (Xiao et al., 2007) and the yeast septin-associated Kcc4p (Moravcevic et al., 2010), the KA-1 domains bind acidic phospholipids and are responsible for targeting the kinase to the plasma membrane (Moravcevic et al., 2010). Phospholipid binding is not accommodated by a hydrophobic cavity as in Bet v 1 proteins. Instead, binding is mediated by electrostatic interactions between positively charged residues on the protein surface and negatively charged phosphates in the phospholipid head groups, consistent with their membrane-targeting role. The electrostatic surface potentials for both domains of BamC shows a more negative character less conducive to binding negatively charged phosphates (Figure 2.20). These results, together with the lack of changes in chemical shifts upon addition of \(E.\text{coli}\) phospholipids, suggest that BamC is not directly involved in membrane binding.

In multisubunit kinases such as AMP-activated protein kinase (AMPK) and Kcc4p, the KA-1 domains mediate intersubunit contacts that are important for the integrity of the kinase complex. The long C-terminal \(\alpha\)-helix in these KA-1 domains contains the key residues involved in intersubunit contacts (Figure 2.19c). A superimposition of the BamC N-terminal domain with AMPK, illustrating how this region in BamC could be involved in binding to other BAM components, is shown in Figure 2.19c. Residues on the surface of the corresponding helix (\(\alpha2\)) of the
N-terminal BamC domain have a relatively high phylogenetic conservation (Figure 2.21a). A second region of surface-exposed highly conserved residues is also observed on the α3 helix in the C-terminal domain of BamC (Figure 2.21b). Previous studies have shown that BamC interacts directly with the C-terminus of BamD and helps stabilize the BAM complex. One possibility is that, analogous to the KA-1 domain of AMPK, each domain of BamC mediates intersubunit contacts. The presence of two structurally similar domains in BamC could thus serve as a scaffold to stabilize the structure of the multisubunit BAM complex.

Figure 2.20 Electrostatic surface potentials of the N-terminal (a) and C-terminal (b) domains of BamC.

Previous NMR interaction studies, which probed for binding “hot spots” on BamC_{101–344} in the presence of BamD (performed by Dr. Kristina Varga) were carried out with proteins purified from the cytosol. In these experiments, BamC_{101–344} and BamD were prepared from constructs that lacked the periplasmic localization sequence and were therefore harvested from the cytosol without any post-translational lipid modifications. Despite the fact that a stable BamC/D complex can be purified from *E. coli* membranes by overexpressing both wild-type
proteins (Sousa, unpublished), no chemical shift perturbations were seen in BamC<sub>101–344</sub> in the presence of BamD. Thus, the BamC/D interaction is mediated by 1) the disordered N-terminal 75-amino acid region, 2) the lipid modified N-terminus played some role in the BamC/D interaction and/or 3) an unknown membrane component. A recent crystal structure of the BamC<sub>26–217</sub>/BamD complex (Kim et al., 2011a) shows that the predominant interaction surface is within the unstructured region of BamC as shown in Figure 2.22.

![Image](image.png)

**Figure 2.21 The conserved amino acids in BamC are clustered in sites potentially important for protein–protein interactions.**
Conserved sequences are mapped onto surface representations of the (a) N-terminal and (b) C-terminal domains of BamC (the insets show the secondary structure representation of the domain in the same orientation). Color scale: dark blue indicates a higher-than-average level of sequence variability, dark red indicates a higher-than-average level of sequence conservation, and white indicates an average level of sequence changes.
With the exception of hydrogen bonds between the amide nitrogen of Leu151 and sidechain oxygen of Asp210 in BamC with BamD residues, the remaining 20 hydrogen bonds identified in the BamC/D complex are between the disordered region of the BamC and the N-terminus of BamD (Kim et al., 2011a). In the crystal structure, part of the disordered of BamC occupies a region of BamD that is a proposed binding site for the C-terminal targeting sequences of OMP, suggesting a potential regulatory function of BamC. Interestingly, only \( \sim 20\% \) of the BamC by residue is involved in BamD interactions, leaving the function of the helix-grip domains still unknown. The BamC/D complex that was crystalized did not contain the C-terminal domain of BamC, so it remains to be seen if the C-terminal domain is also involved in binding to BamD. Perhaps, consistent with our original analysis, the disordered region together with the N- and C-terminal domains act as a scaffold to bring BamD and other BAM components or OMP chaperones in proximity.

**Figure 2.22 X-ray crystal structure of BamCD complex.**
The disordered 75 residues N-terminal of BamC (blue) are essential for protein-protein interactions with BamD (green) (PDB: 3TGO, Kim, et al., 2011a).
2.5 Conclusions /Future Directions

The lipoprotein BamC is one of five proteins in the β-barrel assembly machine in *E. coli*, but its specific role in outer membrane protein folding and insertion is not known. The CS-RDC-NOE Rosetta calculations using a limited set of NMR data showed that BamC is made up of two well-defined helix-grip domains. This helix-grip motif has been previously observed in the Bet v 1 superfamily, where this fold serves either as a ligand-binding domain or as a protein-protein interaction domain. The helix-grip domains in BamC do not have the hydrophobic binding pocket observed in Bet v 1 or the positively charged loops used to interact with phospholipids observed in Kcc4p. The X-ray crystal structure of BamC in complex with BamD and interaction studies reported by Kim and co-workers illustrate that the disordered region of BamC is important for interactions with BamD. BamC pulls down in a complex with BamD and BamE. It could be that the helix-grip motifs in BamC functions as protein-protein interaction domains to help stabilize interactions between the BamE and BamD. Future protein interaction studies should shed light on this. The solution NMR structure of BamE has been determined (Kim et al., 2011a), which opens the door for a variety of experiments to test for interactions between BamC and BamE directly.

The studies here also showed that CS-RDC-NOE Rosetta tolerates the inclusion of some incorrect NOE assignments in the NMR restraints, even those with large distance violations. For some structure generation programs, this level of inconsistent NOE restraints can trap molecules in high-energy conformations, making it difficult to refine the structures. Rosetta is quite robust to such inconsistencies because it identifies correctly folded structures based on low energies for its scoring function, where the experimental data have low weights and serve primarily to guide the conformational search. The limited dataset used here illustrates the usefulness of Rosetta to
determined NMR structures of larger proteins where perdeuteration is necessary to combat relaxation from $^1\text{H}^-^1\text{H}$ dipole-dipole interactions, but limits the number of observable distance restraints to $^1\text{H}^\text{N}^-^1\text{H}^\text{N}$ NOEs.
Chapter 3  Refining the structure of BamA POTRA4–5 domains with residual dipolar couplings

3.1  Introduction

3.1.1  The role of BamA and POTRA domains in the BAM complex

The protein BamA is found in all Gram-negative bacteria and an essential component of the BAM complex (Wu et al., 2005b; Voulhoux et al., 2003). It is comprised of five N-terminal polypeptide transport-associated (POTRA) repeats that extend into the periplasm and a C-terminal β-barrel domain that spans the outer membrane (Gentle et al., 2005; Sánchez-Pulido et al., 2003) as depicted schematically in Figure 3.1. While the mechanistic details of the function of POTRA domains remain unclear, the POTRA domains of E. coli and S. cerevisiae have been shown to interact with C-terminal peptides from outer membrane proteins, protein components of the BAM and chaperones (Sánchez-Pulido et al., 2003; Kim et al., 2007; Knowles et al., 2008).

Each of the five POTRA domains in E. coli BamA are ∼75 residues in length (∼8.5 kDa) and are separated by ∼3–4 residue linkers. The domains have a similar structure with two α-helices packed against a three stranded β-sheet (Kim et al., 2007; Gatzeva-Topalova et al., 2008). Even with this structural homology, the functional significance of each of the domains appears to be varied. Kim et al. investigated the functional significance of the POTRA domains by designing a series of experiments where POTRA domains were systematically deleted and introduced into E. coli with a complementation assay (Kim et al., 2007). POTRA1 and POTRA2 deletion mutants are viable, but grow slowly when compared to the wild type (Kim et al., 2007). The POTRA3–4 domains are essential and the cells did not survive complementation. POTRA5 deletion mutants were toxic to E. coli even in the presence of wild-type BamA, which suggested
that the POTRA5 deletion mutant somehow interferes with BAM function, perhaps by sequestering nascent OMP proteins, or inducing a misfolded or aggregated state in the periplasm (Kim et al., 2007). From these deletion studies, it was determined that BamD interacts with POTRA3, and BamC, BamD and BamE interact with POTRA4 and possibly POTRA5 (Kim et al., 2007). Similarly, in *Neisseria meningitides*, POTRA5 was the only POTRA domain essential for viability (Bos et al., 2007a). Thus, the C-terminal β-barrel and POTRA5 are thought to make up the functional unit of BamA (Bos et al., 2007b).

![Diagram of BamA](image)

**Figure 3.1** BamA is comprised of five N-terminal polypeptide transport-associated (POTRA) repeats and a C-terminal β-barrel domain.

The POTRA domains extend into the periplasm and the C-terminal β-barrel domain spans the outer membrane (Gentle et al., 2005; Sanchez-Pulido et al., 2003). The inner membrane space is ~140 Å, a length that is possible for the POTRA domains to span (Gatzeva-Topalova et al., 2010).

Two groups have reported independent X-ray crystal structures of POTRA1–4 (Kim et al., 2007; Gatzeva-Topalova et al., 2008); however, efforts from both labs to crystalize the entire
POTRA1–5 periplasmic portion of BamA were unsuccessful. Consequently, the orientation of the fifth POTRA domain in relation to the rest of the POTRA domains has been unknown. Gatzeva-Topalova et al., proposed a spliced model of a crystal structure of the forth and fifth POTRA domains (POTRA4–5) of BamA and POTRA1–4 (Gatzeva-Topalova et al., 2010). In this model, the relative orientation of PORTA4 and POTRA5 contradicted a previous report where a model based on SAXS data and a homology model of POTRA5 suggested that the POTRA5 domain folds sharply back onto POTRA4, resulting in a stacked arrangement of these two POTRA domains (Knowles et al., 2008).

Thus, an independent set of SAXS data was collected on POTRA4–5 (Gatzeva-Topalova et al., 2010). These data suggested that POTRA4–5 behaves as a rigid tumbling molecule, with a tight dispersion of the radius of gyration ($R_g$) and a calculated molecular envelope that is consistent with the crystal structure of POTRA4–5 (Gatzeva-Topalova et al., 2010). However, SAXS is a low resolution, model-dependent technique that breaks down in non-ideal solution conditions such as oligomerization or aggregation, which may have biased either the Knowles et al. or the Gatzeva-Topalova structure. Therefore, an independent method for validation was necessary.

3.1.2 Purpose/Overview

It is the aim of this chapter is to describe the use of RDCs to characterize the solution domain orientation of POTRA4–5. The backbone ($^{1}H$, $^{15}N$, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$) chemical shifts were assigned using NMR resonance assignment techniques and validated using a $^{15}N$-edited NOESY HSQC. Analysis of the RDCs showed that the solution domain orientation of BamA POTRA4–5 is consistent with the crystal structure and recent SAXS data, thus validating the domain orientation of POTRA4–5 in the spliced model. A subset of the experimental RDCs deviated from those
calculated from crystal structure and were assumed to arise from “structural noise” in the crystal structure. Further analysis of the RDCs with a higher resolution crystal structure confirmed this hypothesis for most of the experimental RDCs. Most of the remaining experimental RDCs that deviated from the values back-calculated from the crystal structure were located on regions of the protein that were identified by $^{15}$N relaxation analysis to be prone to fast timescale motions.

3.2 Materials and Methods

3.2.1 Expression/Purification of uniformly $^{13}$C, $^{15}$N labeled BamA POTRA4–5

Plasmid pMS487 (BamA POTRA4–5; BamA$_{264-424}$) was transformed into E. coli Rosetta (DE3) cells (Novagen). An overnight culture from a single colony containing 50 μg/mL kanamycin was used to inoculate 250 ml of M9 minimal media supplemented with 50 μg/mL kanamycin. Cultures were grown for 7 h to an OD$_{600}$ of ~0.7, spun down again to remove unlabeled M9 minimal medium, then resuspended in 2 liters of M9 minimal medium supplemented with 50 μg/mL kanamycin, 1.5 g/liter $^{13}$C-glucose and 1 g/liter $^{15}$NH$_4$Cl (Sigma/Isotec). Cultures were grown at 37 °C to OD$_{600}$ of 0.6. Expression was induced with 0.4 mM isopropyl-$\beta$-D-thiogalactopyranoside (IPTG, Gold Bio Technology Inc.) and cultures were allowed to grow an additional 2 h at 37 °C. Cells were harvested by centrifugation, and the cell pellet resuspended in lysis buffer containing 25 mM Tris-Cl pH 8.0 and Complete EDTA-free protease inhibitor (Roche). Cells were lysed on ice by sonication and then sodium chloride was added to final concentration of 0.3 M. Removal of cell debris was achieved by centrifugation at 31,000 x G for 30 min at 4 °C. The supernatant was applied to a Ni-NTA column (Qiagen) pre-equilibrated with buffer A (25 mM Tris, pH 8, 150 mM NaCl). The protein bound to Ni-NTA beads was washed with 2 column volumes buffer A, followed by a wash with 20 column volumes of buffer A containing 25 mM imidazole. The protein was eluted with buffer A supplemented with 200 mM
imidazole. Fractions containing the protein were incubated with TEV protease overnight at 4 °C to achieve cleavage of the His tag. Protein was then dialyzed overnight at 4 °C against buffer A. After removal of the tag and the TEV protease (which is also 6His tagged) using Ni-NTA beads, the protein was loaded on a size exclusion (HiLoad 26/60 Superdex 200, Amersham Pharmacia Biotech) column pre-equilibrated with buffer NMR buffer (50 mM MES [pH 6.5], 50 mM NaCl, 0.1 mM EDTA) and eluted in the same buffer. The protein was concentrated to 1 mM; HALT protease inhibitor, NaN₃, and TSP were added to 1X, 0.01% and 0.14 mM respectively and the sample was stored at –80 °C until further use.

### 3.2.2 NMR backbone resonance assignments of BamA POTRA4–5.

Spectra were collected on a ¹H, ¹³C, ¹⁵N POTRA4–5 sample concentrated to 1 mM in NMR buffer (50 mM MES [pH 6.5], 50 mM NaCl, 0.1 mM EDTA, 0.01% NaN₃, 0.15 mM TSP, complete EDTA-free protease inhibitor cocktail (Roche) 1 tablet/100 ml, 5% (v/v) ²H₂O). NMR experiments for backbone assignments were collected using the TROSY-based Varian Biopack Suite: 2D ¹H, ¹⁵N HSQC, 3D HNCACB, 3D CBCA(CO)NH, and 3D ¹⁵N-edited NOESY-HSQC. Experiments were collected at 30 °C on a VNMRS 800 MHz spectrometer equipped with a HCN z-axis gradient cold probe. Spectra were processed with NMRPipe (Delaglio et al., 1995) and analyzed with CCPNMR Analysis (Vranken et al., 2005a). Peak lists from the ¹H, ¹⁵N HSQC, HNCACB, and CBCA(CO)NH spectra were used as input to the PINE server for automated assignment (Shen et al., 2009a) (http://pine.nmrfam.wisc.edu/). Peaks from regions of the protein with low probability assignments were analyzed after submission and errors were found ranging from misassigned spins to noise peaks that were unintentionally picked. The peak lists were corrected and resubmitted to the PINE server. This was iterated three times before the assignment probabilities were greater than 0.9 for 136 of the amino acids. Additional assignments were
made and verified manually using backbone connectivities from the HNCACB and CBCA(CO)NH and NOEs from the 3D $^{15}\text{N}$-edited NOESY-HSQC in CCPNMR Analysis. A list of the filenames and parameters is shown in Appendix A6.

### 3.2.3 Measurement of $^{1}\text{H}$-$^{15}\text{N}$ heteronuclear NOE spectra

$^{1}\text{H}$-$^{15}\text{N}$ heteronuclear NOE data were collected on a 0.75 mM $^{13}\text{C}$, $^{15}\text{N}$ POTRA4–5 sample in 50 mM MES (pH 6.5), 250 mM NaCl, 0.1 mM EDTA, 0.01% NaN$_3$, 1X HALT protease inhibitor, 0.14 mM TSP and 10% $^2$H$_2$O at 800 MHz at 30 °C. A presaturation period of 3 s and a relaxation delay of 0.5 s were used in the saturated spectrum and a 3.5 second relaxation delay was used in the reference spectrum. The spectra were collected sequentially as opposed to interleaved as suggested by Renner and co-workers (Renner et al., 2002).

### 3.2.4 Measurement of Residual Dipolar Couplings

The amide $^{1}\text{H}$-$^{15}\text{N}$ RDCs were measured on a 0.75 mM $^{15}\text{N}$, $^{13}\text{C}$-labeled POTRA4–5 sample (in NMR buffer with an additional 250 mM NaCl) in the presence and absence of 8.8 mg/ml liquid crystalline Pf1 phage medium prepared as described (Hansen et al., 1998). Spectra were collected at 30 °C on a Varian Inova 600 MHz spectrometer equipped with a cryogenically cooled HCN z-axis gradient probe. 2D HSQC sensitivity-enhanced $^{15}\text{N}$-IPAP spectra (Ding and Gronenborn, 2003) were collected on the isotropic (no Pf1 phage) and aligned (with Pf1 phage) samples. The $^{1}\text{H}$-$^{15}\text{N}$ couplings were measured using CCPNMR Analysis (Vranken et al., 2005a). A list of the filenames and parameters is shown in Appendix A6.

### 3.2.5 Domain orientation of POTRA4–5 in XPLOR-NIH

A histogram analysis of the RDCs was used to estimate the magnitude ($D_a$) and rhombicity ($R$) of the alignment tensor as described in 2.4.4. A grid search approach was then implemented in XPLOR-NIH to more accurately determine $D_a$ and $R$ of the alignment tensors for
the individual POTRA domains using the experimental RDCs and crystal coordinates for POTRA4–5 with hydrogen atoms added to the coordinate file (Appendix A.1.2). A CNS routine was used that adds the hydrogen atoms and then performs a basic energy minimization (Appendix A.1.3). The RDCs restraints are then included as a pseudoenergy function in XPLOR-NIH, where the domain is held rigid and rotated until the energy of the system is minimized, thus orienting the domain relative to an arbitrary axis system. This is repeated while performing a grid search of $D_a$ and $R$. The $D_a$ and $R$ from minimizations with the lowest energy are used in all of the following calculations.

Molecular dynamics calculations were performed using the program XPLOR-NIH version 2.25. A pool of 100 structures with random orientations for POTRA4 and 5 was generated starting with the crystal coordinates and performing a high temperature simulated annealing where only the torsion angles of the 3-residue linker (G$_{344}$N$_{345}$R$_{346}$) were varied. Each structure in the pool was subjected to simulated annealing torsion angle molecular dynamics calculations using $^1$H-$^{15}$N amide RDCs that fit well ($\pm$ 5 Hz) to the RDCs predicted from the individual domains. Two $H^N$-$H^N$ NOE distance constraints (between residues 344–321 and 345–319) for the 3-residue linker region were included in the refinement to help position the domains. Only the torsion angles of the 3-residue linker were allowed to vary in the simulated annealing period and the POTRA4 and five domains were held rigid. The RDC ($\pm$ 2.5 Hz) and NOE (ranges of 2.0–5.0 Å) constraints were included as pseudoenergy functions with a harmonic potential. The weighting of the RDCs was increased linearly from 0.15 to 2.0 and NOEs were held constant during the simulated annealing period. The set of the 20 lowest energy structures were used to represent the set of orientations for POTRA4–5 consistent with the RDC data.
3.3 Results and Discussion

3.3.1 Backbone resonance assignment of POTRA4–5

The $^1$H, $^{15}$N bond vector was chosen as a probe for measuring RDCs in POTRA4–5. The $^1$D$_{\text{NH}}$ is easily measured with conventional experiments that require the assignment of the backbone amide $^1$H, $^{15}$N chemical shifts. The amide $^1$H, $^{15}$N HSQC spectrum of a 1 mM $^{13}$C, $^{15}$N POTRA4–5 sample collected at 800 MHz at 30 °C has 146 of 160 predicted backbone amide peaks (Figure 3.2). The backbone amide $^1$H, $^{15}$N chemical shifts can be assigned by collecting and analyzing a set of standard NMR experiments. The general approach to assign the backbone of a protein is to establish backbone connectivities, which include correlating the $^{13}$C$^\alpha$ and $^{13}$C$^\beta$ chemical shifts of the $i$ and the $i$-1 residues to the backbone $^1$H, $^{15}$N of the $i$ residue. Three NMR experiments were used here to correlate the $^{13}$C$^\alpha$ and $^{13}$C$^\beta$ chemical shifts to the backbone $^1$H, $^{15}$N: the $^1$H, $^{15}$N HSQC; the 3D HNCACB, which detects $^{13}$C$^\alpha$ and $^{13}$C$^\beta$ chemical shifts for the $i$ and the $i$-1 residues; and 3D CBCA(CO)NH, which detects $^{15}$C$^\alpha$ and $^{15}$C$^\beta$ chemical shifts for the $i$-1 residue. A spin system is defined as the $^1$H, $^{15}$N, $^{13}$C$^\alpha$ and $^{13}$C$^\beta$ chemical shifts for the $i$ peaks. Certain amino acids have distinct $^{15}$N, $^{13}$C$^\alpha$ and $^{13}$C$^\beta$ chemical shifts, such as glycine, serine, threonine and alanine, and can be assigned to a spin system type. By determining the spin system types and linking them together with the $i$ and the $i$-1 peaks from the HNCACB and CBCA(CO)NH, resonance assignments can be made by matching these connections into the protein sequence.

This approach to assigning a protein backbone is well-suited for automation and several programs, such as PINE (Shen et al., 2009a) and MARS (Jung and Zweckstetter, 2004), have been developed to use the list of peaks from backbone assignment spectra and the protein sequence to automatically assign NMR spectra. The success of automated assignment hinges on
the quality of the spectra and the accuracy of the peak lists that are input into the program. In the case of POTRA4–5, the signal-to-noise and resolution of the 3D spectra were quite good and PINE was used to aid in the assignment of POTRA4–5. To do this, peaks were picked from the HSQC, HNCACB and CBCA(CO)NH spectra, along with the amino acid sequence and submitted to the PINE server. The PINE output includes a list of proposed peak assignments and a probability associated with each assignment. In the first few attempts, a low assignment probability coincided with inaccuracies in peak picking, which included noise peaks or missing peaks. Thus, peak lists were modified and resubmitted to the PINE server. Two iterations were performed before 136 of 160 non-proline residues had resonance assignment probabilities greater than 0.9. The backbone connectivities for these assignments were manually inspected (A4.1 to A4.7).
Figure 3.2 An $^1$H, $^{15}$N HSQC amide spectrum of $^{13}$C, $^{15}$N POTRA4–5.

The spectrum was collected at 800-MHz on a cryogenically cooled z-axis probe at 30 °C. The spectrum shows 146 well-resolved peaks.

An additional eight assignments, for a total of 144 of 160 non-proline residues, were made by manually searching for backbone connectivities. The peaks in the 3D spectra from these residues were weak and not picked for the automated PINE assignments. For example, the residues from Ala363–Arg366 and Asp380–Glu387 had very weak or missing peaks in the HNCACB and CBCA(CO)NH spectra (Figure A4.6 and A4.8), thus additional confirmation of the assignments were sought. Through-space interactions, measured with an $^{15}$N-edited NOESY
HSQC spectrum, were used as an additional confirmation of peak assignments. A network of NOEs links the residues in Ala363–Arg366; however no links between Val 364 and Leu365 could definitively established because the $^1$H amide chemical shifts of these two residues are so similar (Figure 3.3). The network of NOEs connecting Asp380–Glu387 was more substantial, with Val382 and Asp383 showing strong NOEs from $i$-2 to $i$+2 (Figure 3.4). The unique chemical shifts of $^{13}C^\alpha$ and $^{13}C^\beta$ of serines, threonines and prolines enabled the manual assignment of two serines. While the $i$+1 peak for Ser360 not observed, Ser360 was assigned unambiguously because the $i$-1 residue was a threonine and this residue pair is unique in the sequence (Figure A4.8). Similarly, Ser408 is $i$+1 from a glycine, which is $i$+1 from a proline. This tripeptide is unique in the sequence and so Ser408 and Gly407 were assigned unambiguously (Figure A4.7). In total, 146/160 non-proline residues and 100% of the non-sidechain peaks in the $^1$H, $^{15}$N HSQC spectrum was assigned. The remaining unassigned residues were coincident with breaks in the backbone connectivity between Asn357–Ala363, Arg366–Gly374 and Leu377–Asp380. These regions are found in loops that connect strand $\beta$1 to helix $\alpha$1 and helix $\alpha$1 to strand $\beta$2 as shown in Figure 3.5. Three lines of evidence support that these regions are undergoing conformational exchange. First, the breaks in the backbone connectivity were bracketed on either side with a decrease in peak intensity in the HNCACB and the CBCA(CO)NH. Second, all of the peaks in the $^1$H, $^{15}$N HSQC accounted for. Third, the $^1$H, $^{15}$N HSQC peak volumes decrease around the unassigned residues as shown in Figure 3.6. Interestingly, an analysis of $^1$H-$^{15}$N heteronuclear NOE data shows that two loops in POTRA4, encompassing residues 275–279, which link the $\beta$1 strand to the $\alpha$1 helix, and residues 286–295, which link the $\alpha$1 helix to the $\alpha$2 helix are undergoing fast timescale (ps-ns) dynamics (Figure 3.7). The corresponding regions of POTRA5 are the residues that are either weak or unassigned (likely due to intermediate conformational
exchange). The residues that are assigned in the loops do show increased backbone dynamics. Thus, it appears that the linkers in both POTRA domains are experiencing heightened dynamics with respect to the ordered regions of the protein. Whether or not there is a functional consequence of this is unknown.
Figure 3.3 Slices from an $^{15}$N-edited NOESY spectrum shows crosspeaks between residues Ala363–Arg366.

The peaks for residues Ala363–Arg366 in the HNCACB and CBCA(CO)NH spectra were weak; a network of NOESY crosspeaks connects residues Ala363–Arg366 and supports backbone resonance assignments.
Figure 3.4 Slices from an $^{15}$N-edited NOESY spectrum shows crosspeaks between residues Asp380–Glu387.

The peaks for residues Ala363–Arg366 in the HNCACB and CBCA(CO)NH spectra were weak; a network of NOESY crosspeaks connects residues Ala363–Arg366 and supports backbone resonance assignments.
Figure 3.5 Residues in the linker between the α1 and α2 helix of POTRA5 were unassigned. X-ray structure of POTRA4–5 (PDB: 3OG5 (Gatzeva-Topalova et al., 2010)) with assigned residues shown in grey and unassigned residues in gold. Regions from 357–363, 366–374 and 377–380 were unassigned. The intensities of the peaks of residues on either side of the unassigned regions were weak, suggesting that these regions are undergoing conformational exchange broadening.

Figure 3.6 Peak volumes in the $^1$H, $^{15}$N HSQC of the $^1$H, $^{15}$N amides for POTRA4–5 decrease around unassigned loop.

Regions from 357–363, 366–374 and 377–380 were unassigned and the volumes of the peaks of residues on either side of the breaks are smaller relative to the rest of the protein, suggesting that these regions are undergoing intermediate conformational exchange broadening.
Figure 3.7 \(^{1}H\)-\(^{15}N\) heteronuclear NOE values plotted as a function of residue number in POTRA4–5.

A dashed line indicates a 0.7 NOE threshold, where values below 0.7 indicate fast (ps-ns) timescale dynamics usually indicative of disordered regions of the protein and values above 0.7 indicate highly ordered regions. Two loops in POTRA4 and POTRA5 show increased dynamics: residues 275–279 link the \(\beta1\) strand to the \(\alpha1\) helix and residues 286–295 link the \(\alpha1\) helix to the \(\alpha2\) helix in POTRA4. Residues 357, 360, 363, and 364 link the \(\beta1\) strand to the \(\alpha1\) helix none of the corresponding residues that link the \(\alpha1\) helix to the \(\alpha2\) helix in POTRA5 were assigned likely due to conformational exchange broadening (discussed in text).

3.3.2 Solution domain orientation of POTRA4–5 is consistent with the crystal structure.

Solution NMR was used as an independent method for assessing the orientation of the two POTRA domains in BamA POTRA4–5. Residual dipolar couplings provide information on the orientation of individual bond vectors in a molecule and can be used to determine the flexibility and relative orientation of rigid domains in a macromolecule (Fischer et al., 1999; Bax et al., 2001). The crystal structure coordinates for POTRA4 and POTRA5 (PDB 3OG5) were used to define two rigid domains connected by the 3 amino acid linker (G\(_{344}\)N\(_{345}\)R\(_{346}\)). The \(^1H\)-\(^{15}N\) amide backbone RDCs were measured by analysis of 2D \(^1H\), \(^{15}N\) IPAP HSQC spectra (Ding and Gronenborn, 2003) obtained under isotropic and partially aligned conditions (Bax et al., 2001). A total of 73 and 59 \(^1H\)-\(^{15}N\) amide RDCs from POTRA4 and POTRA5, respectively, were measured.
The first step in domain orientation using RDC data is to assess whether the two domains have a fixed orientation and, therefore, the same alignment tensor. One indication of a rigid orientation for two domains is if the individual domains have similar values for the magnitude \( D_a \) and the rhombicity, \( R \), of their alignment tensors (Fischer et al., 1999). A structure-independent method, discussed in Section 1.2, that involves an analysis of the shape of the histograms of the RDCs in a molecule was used here to compare \( D_a \) and \( R \) in the two domains (Clore et al., 1998a). Figure 3.8 shows the distribution of RDCs for POTRA4 and POTRA5. Although using only \(^1\text{H},\ ^{15}\text{N}\) RDCs is a coarse method for determining \( D_a \) and \( R \), initial estimates of 15 Hz and 0.5 respectively were made. Next, a grid search was employed to find more accurate values of \( D_a \) and \( R \) for the individual domains using the crystal coordinates and the experimental RDCs for each domain (Fischer et al., 1999). No explicit hydrogens were modeled into the X-ray structure and hence, the hydrogens were added using the program CNS. The RDCs can then be predicted from the alignment tensors and a plot of the experimental and predicted RDCs for the individual POTRA4 and 5 domains give an indication of the agreement between experimental RDCs and those predicted from the crystal structure (Figure 3.9a and c). The \(^1\text{H},^{15}\text{N}\) RDCs are a sensitive function of the angle of the bond vector, so even slight differences in the orientations of the HN bond in the crystal and in solution leads to “structural noise” (Zweckstetter and Bax, 2002). Thus, only the subset of 70 RDCs that showed good agreement (within ± 5 Hz) between the experimental and predicted RDCs were used (Figure 3.9b and d). The values of the magnitude of the principal component, \( D_a \) and rhombicity, \( R \), of the alignment tensors for POTRA4 and 5, were consistent with the two domains behaving as a single rigid species (Fischer et al., 1999), thus the next step is to determine the relative orientation of the two domains in an unbiased manner.
Figure 3.8 Histograms of RDCs collected on POTRA4–5.

78 and 54 $^1$H–$^{15}$N RDCs for the (a) POTRA4 and (b) POTRA5 domains of BamA. The shape of the histogram provides structure independent information on the magnitude $D_a$ and the rhombicity $R$ of the alignment tensor (Clore et al., 1998). The similarity in the shapes of the two histograms indicate that the POTRA4 and POTRA5 have the same alignment tensor and, therefore, a fixed orientation in solution (assuming that the RDCs adequately sample all orientations of $^1$H–$^{15}$N bond vectors).

To do this, a pool of 100 structures with randomized orientations between the rigid POTRA4 and POTRA5 domains was generated by varying the torsion angles in the 3-residue linker (Figure 3.10a). This pool of structures was subjected to constrained molecular dynamics calculations (Schwieters et al., 2006), where the experimental RDCs were included in the energy function as constraints and only the torsion angles in the linker were varied. The 20 lowest energy structures (average rmsd of 2.56 Hz for all RDCs) are in good agreement with one another and with the crystal structure (average rmsd of 3.40 Å to the crystal structure over 160 Cα) as shown in Figure 3.10b, further supporting that the crystal structure is a good representation of the domain orientation in solution and not a result of crystal lattice contacts.

The RDC data showed that individual POTRA4 and 5 domains have similar magnitudes of their alignment tensors. Thus, a single alignment tensor was used to find orientations of the two domains that fit well to the RDCs. The resulting ensemble of minimized structures has
similar domain orientation to that of the crystal. Having similar magnitudes for the alignment tensor of individual domains normally provides strong evidence for rigidity in solution (Fischer et al., 1999).

Figure 3.9 Plot of the experimental and predicted $^1$H-$^{15}$N RDCs of POTRA4–5.

The RMSD for the fit is in the lower right hand corner of each panel. (a) The experimental and predicted RDCs for POTRA4. The predicted RDCs were obtained with XPLOR using the crystal structure coordinates, a $D_{\alpha}^{\text{NH}} = 12.75$ Hz and an $R = 0.55$ (the latter two parameters determined using a grid search). (b) The subset of RDCs used for POTRA4, which includes only RDCs with experimental and predicted values within $\pm 5$ Hz. (c) The experimental and predicted RDCs for POTRA5. The predicted RDCs were obtained with XPLOR using the crystal structure coordinates, a $D_{\alpha} = 12.75$ Hz, and an $R = 0.55$ (both determined using a grid search). (d) The subset of RDCs used for POTRA5, which includes only RDCs with experimental and predicted values within $\pm 5$ Hz.
Figure 3.10 Domain orientation of POTRA4–5 from a starting pool of structures with randomized domain orientation.

A representative subset of the starting pool of structures with randomized domain orientation used to fit the experimental RDCs with POTRA4 (salmon) domains of all structures superimposed to show the extent of structural variability in the starting pool (a). Family of low energy NMR structures (in gray) superimposed with the X-ray structure (POTRA4, salmon; POTRA5, sage) showing good agreement between solution and crystal structures (b).

However, since the two domains are the same size, they could show similar alignment tensors even with extensive conformational flexibility in the linker. Thus, amide $^1$H-$^1$H NOEs in the linker region were analyzed to further examine the similarity of the crystal and solution structures. Figure 3.11 shows the observed H$^N$-H$^N$ NOEs for residues 343–349 mapped onto the crystal structure. The pattern of NOEs provides strong evidence that: 1) the backbone around the linker region is in a similar conformation in the crystal and in solution and 2) this backbone is relatively rigid in solution. The NOE data rule out significant conformational flexibility of the backbone in the linker region. Backbone dynamics would reduce the intensities of these amide proton NOEs due to conformational averaging and increased flexibility would reduce the efficiency of spin-spin exchange that gives rise to the NOE.

Additionally, $^1$H-$^{15}$N heteronuclear NOE data support that the linker region is not undergoing fast timescale (ps-ns) dynamics. $^1$H-$^{15}$N heteronuclear NOE values less of than 0.7
can indicate a residue is disordered, generally observed for residues found in linkers, loops or at the termini. \( ^1\text{H}-^{15}\text{N} \) heteronuclear NOE values were measured for residues 343–345, 347, 348 and 358 all were above 0.7 as shown in Figure 3.7. Thus, the size and patterns of the observed \( ^\text{H}^\text{N}-^\text{H}^\text{N} \) NOEs and the lack of fast timescale dynamics strongly supports that the backbone in the linker is well ordered and is in a conformation similar to that in the crystal. The resulting information from this study supported the use of POTRA4–5 in generating a spliced model of POTRA1–5.

\[ \text{Figure 3.11} \, \text{A network of NOEs shows that the backbone around the POTRA4–5 linker is relatively rigid.} \]

NOE restraints mapped onto crystal structure of POTRA4–5 depicting 11 amide \( ^\text{H}^\text{N}-^\text{H}^\text{N} \) NOEs (yellow dashes) measured around the linker region (cyan) linking POTRA4 (salmon) and POTRA5 (sage). NOEs were measured on a 1 mM \( ^{13}\text{C}, ^{15}\text{N} \) POTRA4–5 sample at 30 °C at 800 MHz using a gradient selected \( ^{15}\text{N} \)-edited NOESY-HSQC with a 150 ms mixing time.
3.3.3 A higher resolution X-ray crystal structure of POTRA4–5 fits the RDCs with less “structural noise”.

After the previous study to orient the domains of POTRA4–5, questions remained about the RDCs that did not fit well (> ±5 Hz difference between experimental RDCs and those calculated from the crystal structure). In X-ray crystal structures, the hydrogen atoms do not have enough electron density to reflect X-rays. Thus, in order to analyze $^1$H, $^{15}$N RDCs, hydrogen atoms are added to the coordinate file, where the location is interpolated from the peptide plane. It was assumed in the previous study, using the 2.69 Å resolution 3OG5 crystal structure (Gatzeva-Topalova et al., 2010) that the angle of HN bond vector could not be defined well enough in all of the residues, which resulted in “structural noise”. However, without a higher resolution structure to compare to, this was merely speculation. A recent POTRA4–5 crystal structure at 1.50 Å (PDB: 3Q6B (Zhang et al., 2011)) resolution provided an excellent opportunity to test this hypothesis. The new crystal structure superimposes to 3OG5 (Gatzeva-Topalova et al., 2010) with a 0.78 Å pairwise rmsd to C$\alpha$, with exception of terminal residues 262–266 and 421–426, whose coordinates were not defined in the construct for 3Q6B (Zhang et al., 2011). The addition of hydrogen atoms to a lower resolution crystal structure (3OG5, (Gatzeva-Topalova et al., 2010)) relative to a higher resolution one (3Q6B, (Zhang et al., 2011)) can result in inconsistencies in the position of the hydrogen atom as demonstrated in Figure 3.12. Thus, 3Q6B was fit to the unpruned RDC dataset from 3.3.2 to determine if there was better agreement between the experimental and predicted RDCs (Figure 3.13).
Figure 3.12 Spatial position of hydrogen atoms added to structure coordinates derived from X-ray crystallographic data is dependent on the resolution. On the right is a superimposition of coordinates of 3OG5 after addition of hydrogen atoms using CNS (green) and REDUCE via the MolProbity software (cyan). On the left is a superimposition of coordinates of 3Q6B after addition of hydrogen atoms using CNS (raspberry) and REDUCE via the MolProbity software (salmon). Nitrogens and hydrogens are colored in blue and white, respectively. Structures were aligned in a pairwise manner to the α1 helix of POTRA4, residues 280–296.

The rmsd and the number of experimental RDCs greater than ±5 Hz of the calculated RDCs decreased for both POTRA4 and POTRA5 from the lower resolution structure to the higher resolution structure (Table 3.1). There was a decrease in the rmsd from 4.53 Hz to 3.30 Hz for POTRA4 and 7.96 Hz to 7.63 Hz for POTRA5 when comparing from 3OG5 and 3Q6B respectively. The number of experimental RDCs that were greater than ±5 Hz of the calculated RDCs decreased from 17 to 4 for POTRA4 and from 33 to 26 for POTRA5. While it is clear that the improved structural resolution of the crystal structure decreased the RMS of the calculated to
the experimental RDCs and the number of experimental RDCs that deviated from the calculated RDCs; however, the agreement was not as substantial as anticipated. Therefore, two criteria were used to reevaluate the measured RDCs; 1) peaks that overlapped in the IPAP spectra were disregarded and 2) RDCs from residues that exhibited fast timescale dynamics as measured by heteronuclear \(^1\)H-\(^1\)\(^{15}\)N heteronuclear NOE were disregarded. Any peak overlap in the IPAP spectra resulted in difficulties determining what the peak positions were. This in turn creates errors in the RDC calculation. Thus, 15 RDCs that were initially measured from peaks that were slightly overlapped were removed from the dataset. The second criteria for removing an RDC from the dataset was a low \(^1\)H-\(^1\)\(^{15}\)N heteronuclear NOE, where a value of less than 0.7 was used to validate excluding a measured RDC from the final dataset. The RDCs were then predicted from the alignment tensor and the coordinates for low resolution (PDB: 3OG5 (Gatzeva-Topalova et al., 2010)) and high resolution (PDB: 3Q6B (Zhang et al., 2011)) crystal structures and compared to the experimental RDCs from the updated dataset. A plot of the experimental and predicted RDCs for the individual POTRA4 and 5 domains from the two crystal structures shows that there is better correlation between the experimental RDCs and the higher resolution crystal structure with an rmsd of 2.53 Hz and 3.99 Hz for POTRA4 and POTRA5 (Figure 3.14).

The updated RDC dataset and the higher resolution crystal structure were then used to orient the domains as described in section 3.3.2. A pool of 100 structures with randomized orientations between the rigid POTRA4 and POTRA5 domains was generated from 3Q6B by varying the torsion angles in the 3-residue linker (Figure 3.15). This pool of structures was subjected to constrained molecular dynamics calculations (Schwieters et al., 2006), where the updated list of experimental RDCs were included in the energy function as constraints and only the torsion angles in the linker were varied. The 20 lowest energy structures are in good agree-
ment with one another and with the crystal structure (average rmsd of 1.28 Å to the crystal structure over 160 Cα) as shown in Figure 3.15b. The results of this domain orientation supported the orientation of the crystal structure more than with the lower resolutions structure, where the low energy NMR structures were in good agreement each other and superimpose with the crystal structure with a 1.28 Å rmsd. This is a marked improvement over the domain orientation with the lower resolution structure, where the low energy structures superimposed with the crystal structure with a 3.40 Å rmsd. This improvement is likely due to a better definition of the HN bond vector with the higher resolution structure.

It would stand to reason when a lower resolution crystal structure must be used for domain orientation calculations, that measuring the amide-carbonyl (NC’) bond vector could yield results that are more consistently accurate. This is because both nitrogen and carbon atoms diffract X-rays and are thus generally well represented in the electron density map. The challenge comes in measuring the NC’ dipolar coupling, where spectra are generally less sensitive relative to 1H, 15N IPAP or HSQC spectra, might not be feasible in larger proteins.

Another approach would be to refine lower resolution crystal structure with the main objective to produce structures with hydrogens accurately represented. It turns out that simple geometric addition of hydrogens to the backbone amide in particular is challenging because the geometry around nitrogens is difficult to predict, even when it is incorporated in to the peptide plane (Cooper et al., 2010). Several algorithms already exist to add hydrogens or to refine crystal structure with “riding hydrogens” (Cooper et al., 2010). The high resolution and lower resolution crystal structures of POTRA4–5 would make an ideal test system for analyzing which programs are best suited for adding hydrogens for future studies.
Figure 3.13 Plot of the experimental and predicted $^1$H-$^{15}$N RDCs of POTRA4–5 using an RDC data set without filtering out residues with low $^1$H-$^{15}$N heteronuclear NOE values to the low and high-resolution X-ray crystal structures.

The crystal structure coordinates for a low resolution POTRA4–5 (PDB: 3OG5, Gatzeva-Topalova, et al., 2010) and a high resolution POTRA4–5 (PDB: 3Q6B, (Zhang, et al., 2011) and experimental RDCs from a data set that had not been filtered based on low $^1$H-$^{15}$N heteronuclear NOE values were used to back calculate RDCs. These are compared for the low-resolution structure for POTRA4 (a) and POTRA5 (c) and the high-resolution structure for POTRA4 (b) and POTRA5 (d). The RMS for the fit is boxed in the lower right hand corner of each panel. The predicted RDCs were obtained with XPLOR using the crystal structure coordinates of PDB 3OG5, a $D_{a\text{NH}} = 12.75$ Hz and an $R = 0.55$ (the latter two parameters determined using a grid search). The experimental and predicted RDCs for POTRA4 (boxes). The predicted RDCs were obtained with XPLOR using the crystal structure coordinates of PDB 3Q6B, a $D_{a\text{NH}} = 14.10$ Hz and an $R = 0.59$ (the latter two parameters determined using a grid search).
Figure 3.14 Plot of the experimental and predicted $^1$H-$^{15}$N RDCs of POTRA4–5 using an RDC data set after filtering out residues with low $^1$H-$^{15}$N heteronuclear NOE values to the low and high-resolution X-ray crystal structures.

Plot of the experimental and predicted $^1$H-$^{15}$N RDCs of POTRA4–5 using an updated RDC dataset and crystal structure coordinates for PDB 3OG5 (Gatzeva-Topalova et al., 2010) (a and c) and PDB 3Q6B (Zhang et al., 2011) (b and d). The RMS for the fit is boxed in the lower right hand corner of each panel. (a) The experimental and predicted RDCs for POTRA4 (boxes). The predicted RDCs were obtained with XPLOR using the crystal structure coordinates of PDB 3OG5, a $^{\text{D}_{\alpha}}_{\text{NH}}$ = 12.75 Hz and an $R$ = 0.55 (the latter two parameters determined using a grid search). (b) The experimental and predicted RDCs for POTRA4 (boxes). The predicted RDCs were obtained with XPLOR using the crystal structure coordinates of PDB 3Q6B, a $^{\text{D}_{\alpha}}_{\text{NH}}$ = 14.10 Hz and an $R$ = 0.59 (the latter two parameters determined using a grid search). (c) The experimental and predicted RDCs from PDB 3OG5 (as in (a)) for POTRA5 (circles). (d) The experimental and predicted RDCs from PDB 3Q6B (as in (b)) for POTRA5 (circles).
Figure 3.15 Domain orientation of high resolution POTRA4–5 from a starting pool of structures with randomized domain orientation.

Solution domain orientation using 3Q6B (Zhang et al., 2011) (a) A representative subset of the starting pool of structures with randomized domain orientation used to fit the experimental RDCs with POTRA4 (salmon) domains of all structures are superimposed to show the extent of structural variability in the starting pool. (b) Family of low energy NMR structures (in gray) superimposed with the X-ray structure (sage) showing good agreement between solution and crystal structures. rmsd of 1.58 Å to the crystal structure.

Table 3.1 Summary of RDCs fits to high and low resolution structures of POTRA4–5

Summary of results from calculating $D_a$ and $R$ using the low resolution (3OG5, Gatcheva-Topalova, 2010) and the high resolution (3Q6B, Zhang et al., 2011) with the original RDC data set and a dataset updated with RDCs removed for residues with overlap and low $^{1}H$-$^{15}N$ heteronuclear NOE values.

<table>
<thead>
<tr>
<th>RDC violations $\geq$ / Total of RDCs</th>
<th>Original dataset</th>
<th>Updated dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rmsd (Hz)</td>
<td>rmsd (Hz)</td>
</tr>
<tr>
<td>3OG5 POTRA4</td>
<td>17/73</td>
<td>18/68</td>
</tr>
<tr>
<td>3OG5 POTRA5</td>
<td>33/59</td>
<td>16/44</td>
</tr>
<tr>
<td>3Q6B POTRA4</td>
<td>4/73</td>
<td>2/68</td>
</tr>
<tr>
<td>3Q6B POTRA5</td>
<td>26/59</td>
<td>5/44</td>
</tr>
</tbody>
</table>


3.4 Conclusions/Future directions

RDCs were successfully used to orient the POTRA4–5 domains for the analysis of the complete periplasmic domain of BamA. The solution domain orientation and a network of NOEs along the linker of POTRA4–5 support a rigid orientation of the domains. Interestingly, the loops that connect the α1 and α2 helices in POTRA4 showed increased ps-ns timescale dynamics and in POTRA5 appear to be undergoing conformational exchange broadening, which in the intermediate exchange regime the timescale of the chemical shift could be on the μs-ms timescale. This could just be fluctuations in the local structure that have nothing to do with function, but it would be interesting to see if the dynamics change in the presence of an outer membrane substrate. In the absence of a chemical energy source in the periplasmic space, it’s thought that reactions must proceed with simply the energy of protein-protein binding. Thus, the enthalpy of binding and changes in protein dynamics could set up a thermodynamic gradient that could direct the folding and insertion of outer membrane proteins.
Chapter 4  Investigating the flexibility of BamA POTRA 1–5

4.1  Introduction

4.1.1  β-augmentation in the *E. coli* BamA POTRA domains.

The *E. coli* BamA is an essential component of the outer membrane of Gram-negative bacteria, comprised of a transmembrane β-barrel domain and five periplasmic domains (Werner and Misra, 2005). The five periplasmic domains share a common POTRA fold with two α helices packed against a three-stranded β-sheet (Figure 4.1). POTRA1–5 are hypothesized to play a direct role in the folding and insertion of β-barrel proteins into the outer membrane, but the mechanistic details of this function are unclear (Rossiter et al., 2011; Hagan et al., 2011). A crystallographic analysis of the POTRA1–4 suggests that the POTRA domains may function as protein recognition sites, by binding the nascent chains of outer membrane proteins via β-augmentation (Kim et al., 2007; Gatzeva-Topalova et al., 2008). In one crystal structure, a short C-terminal β-strand of POTRA5 from a second molecule in the asymmetric unit was found packed against the β-sheet of POTRA3 (Kim et al., 2007) (Figure 1.2a). In a second structure, under completely different crystallization conditions and crystal packing, the short C-terminal β-strand of POTRA5 of a symmetry-related partner was packed against the β-sheet of POTRA3 (Gatzeva-Topalova et al., 2008) (Figure 1.2b). Additionally, all of the POTRA domains have an ~30 Å long hydrophobic surface groove that is conserved in structurally similar proteins, such as the cytoplasmic chaperone SecB, which is known to bind outer membrane protein precursors by β-augmentation (Xu et al., 2000; Kim et al., 2007; Gatzeva-Topalova et al., 2008). The length of β-strands of outer membrane proteins range from 27–35 Å, making the hydrophobic groove a suitable size to accommodate an outer membrane protein β-strand (Tamm et al., 2001).
The POTRA domains of BamA share a common fold characterized by two α helices packed against a three strand β-sheet. The structures for POTRA1, 2 and 3 (a, red; b, orange; and c, green) were extracted from PDB 3EFC (Gatzeva-Topalova, et al., 2008) and the structures for POTRA4 (d, cyan) and 5 (e, blue) were extracted from PDB 3OG5 (Gatzeva-Topalova, et al., 2010).

4.1.2 Crystal structures of POTRA1–4 show two unique conformations with a hinge between POTRA2 and POTRA3.

In two crystal structures of POTRA1–4 (Gatzeva-Topalova et al., 2008; Kim et al., 2007), the orientation of the domains between POTRA2 and POTRA3 domains is different, resulting in structures that adopt either an “extended” or a “compact” conformation (Figure 4.2 shows a spliced model of the two structures of POTRA1–4 (PDB: 3EFC (Gatzeva-Topalova et al., 2008) and 2Q2CZ (Kim et al., 2007)) with POTRA4-5 (PDB: 3OG5, (Gatzeva-Topalova et al., 2010)). In the “extended” conformation (Figure 4.2a, PDB: 3EFC and 3OG5 (Gatzeva-Topalova et al., 2008; 2010)) the angle between POTRA2 and POTRA3 is ~130°, while in the “compact” conformation (Figure 4.2b, PDB: 2Q2CZ and 3OG5) it is ~100° (Kim et al., 2007; Gatzeva-Topalova et al., 2008). The POTRA1–2 and POTRA3–4 domains superimpose well between the two structures, with pairwise rmsds of 1.0 and 0.6 Å respectively, suggesting that this fragment consists of two relatively rigid arms formed by POTRA1-2 and POTRA3-4 with a hinge point in the linker between POTRA2 and 3 (Gatzeva-Topalova et al., 2008). The structure of BamA
POTRA4-5 is also relatively rigid, as shown in an NMR study that used RDCs and NOEs to determine the solution domain orientation of POTRA4–5 (see Chapter 3 and (Gatzeva-Topalova et al., 2010)). Moreover, an analysis of SAXS of POTRA4–5 scattering curves reinforced the conclusion that POTRA4–5 tumbles with a fixed orientation of the domains (Gatzeva-Topalova et al., 2010). Spliced models of the complete periplasmic domain were built (Gatzeva-Topalova et al., 2010) using a crystal structure of POTRA4–5 (PDB 3OG5) and the POTRA1–4 structures (Kim et al., 2007; Gatzeva-Topalova et al., 2008) (Figure 4.2). These models were built upon the assumption POTRA1–2 and 3–4 were rigid bodies, with a flexible hinge between domains 2 and 3 (Gatzeva-Topalova et al., 2010). SAXS scattering curves from BamA POTRA3-5 support the assumption that POTRA3–5 tumbles as a rigid body and an EPR study corroborates the notion that the orientation of POTRA1 with respect to POTRA2 is fixed in solution (Gatzeva-Topalova et al., 2010; Ward et al., 2009). An analysis of the complete periplasmic domain of BamA, POTRA1–5, by SAXS solution scattering suggested that the protein exists in solution as a mixture of “compact” and “extended” molecules, consistent with conformational flexibility between POTRA2 and 3 (Gatzeva-Topalova et al., 2010).

One theory suggests that the nascent chain of an outer membrane protein binds to the POTRA domains in the “extended” conformation and that a β-hairpin is formed in the nascent chain during the process of BamA adopting the “compact” conformation (Gatzeva-Topalova et al., 2010). If this process were repeated and coupled somehow to insertion of the β-strands into the outer membrane, this could explain the biological role of BamA in the assembly and insertion of outer membrane proteins. To assess the biological role of the putative flexibility of POTRA1–5, double cysteine mutations between POTRA2 and 3, designed to “lock” POTRA1–5 into the “compact” and “extended” conformations, were complemented into E. coli and defects in the
outer membrane were observed (Sousa, unpublished). These preliminary results suggest that hindering the flexibility of POTRA1–5 impedes its function.

![Figure 4.2](image-url) An “extended” and “compact” orientation of POTRA1–5 in crystal structures.

Two crystal structures of POTRA1–4 spliced (a, PDB: 3EFC (Gatzeva-Topalova et al., 2010) and b, PDB: 2QCZ (Kim et al., 2007)) with POTRA4–5 (PDB: 3OG5 (Gatzeva-Topalova et al., 2010)). The crystal structures of POTRA1–4 show an “extended” (a) and “compact” (b) conformation, where the difference between the two structures is in the angle between POTRA2 and 3. In the “extended” conformation, the angle is 130° and in the “compact” form is reduced to 100°.

### 4.1.3 Purpose/Overview

Here, an in vitro approach to investigate the flexibility of POTRA1–5 in solution using NMR spectroscopy was taken. First, the NMR resonances of the backbone amides of POTRA1–5 were assigned. Building on assignments that were made to POTRA4–5 (Chapter 3 and (Gatzeva-Topalova et al., 2010)) and previously published assignments to POTRA1–2 (Knowles et al., 2008), the strategy for assignments were to assign the resonances for POTRA3 and transfer
all assignments to POTRA1–5 as shown in Figure 4.3. This strategy was successful in the assignment of 256 of the 317 peaks in the $^1$H, $^{15}$N HSQC spectrum of $^2$H, $^{13}$C, $^{15}$N POTRA1–5.

Figure 4.3 Schematic description of the NMR resonance assignment strategy.

The backbone chemical shifts of POTRA1–2 have been published (Knowles, et al., 2008) and POTRA4–5 backbone chemical shifts for POTRA4-5 were reported in Chapter 3 and (Gatzeva-Topalova et al., 2010). The assignment strategy is to 1) transfer the chemical shift assignments from POTRA4–5 to POTRA3–5, 2) assign POTRA3 then 3) transfer the assignments for POTRA1–2 and POTRA3–5 to POTRA1–5.

Three experimental approaches were taken to investigate the flexibility of POTRA1–5 using in solution. The $^{15}$N relaxation experiments indicate that POTRA1–2 and POTRA3–5 tumble partially independently. Paramagnetic relaxation enhancement suggests that POTRA1–5 transiently assumes a “compact” conformation. An analysis of RDCs shows that POTRA4–5 orients independently of POTRA1–2 and POTRA3.

4.2 Materials and Methods

The following materials and methods document the procedures used to prepare samples and perform NMR experiments to characterize the backbone dynamics, paramagnetic relaxation enhancement and domain orientation of BamA POTRA1–5. A detailed account of the NMR experiments used to assign the backbone ($^1$H$^N$, $^{15}$N$^H$, $^{13}$C$^\alpha$, $^{13}$C$^\beta$) resonances of POTRA3–5 are given. The RDCs were measured by two techniques that are detailed here: 1) the halved cou-
plings from a decoupled $^1$H, $^{15}$N HSQC and a TROSY $^1$H, $^{15}$N HSQC and 2) the amide RDCs by TROSY (ARTSY) method.

### 4.2.1 Expression/Purification of uniformly $^{13}$C, $^{15}$N labeled BamA POTRA constructs

Plasmids pMS487 (BamA POTRA4–5; BamA$_{264-424}$), pMS488 (BamA POTRA3–5; BamA$_{172-426}$), pMS679, or (BamA POTRA1–5; BamA$_{25-424}$) or were transformed into *E. coli* Rosetta (DE3) cells (Novagen). An overnight culture from a single colony containing 50 μg/mL kanamycin was used to inoculate 250 ml of M9 minimal media supplemented with 50 μg/mL kanamycin. Cultures were grown for 7 h to an OD$_{600}$ of ~0.7, spun down again to remove unlabeled M9 minimal medium, then resuspended in 2 liters of M9 minimal medium supplemented with 50 μg/mL kanamycin, 1.5 g/liter $^{13}$C-glucose and 1 g/liter $^{15}$NH$_4$Cl (Sigma/Isotec). Cultures were grown at 37 °C to OD$_{600}$ of 0.6. Expression was induced with 0.4 mM isopropyl-$\beta$-D-thiogalactopyranoside (IPTG, Gold Bio Technology Inc.) and cultures were allowed to grow an additional 2 h at 37 °C. Cells were harvested by centrifugation, and the cell pellet resuspended in lysis buffer containing 25 mM Tris-Cl pH 8.0 and Complete EDTA-free protease inhibitor (Roche). Cells were lysed on ice by sonication and then sodium chloride was added to final concentration of 0.3 M. Removal of cell debris was achieved by centrifugation at 16,000 rpm for 30 min at 4 °C. The supernatant was applied to a Ni-NTA column (Qiagen) pre-equilibrated with buffer A (25 mM Tris, pH 8, 150 mM NaCl). The protein bound to Ni-NTA beads was washed with 2 column volumes buffer A, followed by a wash with 20 column volumes of buffer A containing 25 mM imidazole. The protein was eluted with buffer A supplemented with 200 mM imidazole. Fractions containing the protein were incubated with TEV protease overnight at 4 °C to achieve cleavage of the 6His tag. Protein was then dialyzed overnight at 4 °C against buffer A. After removal of the tag and the TEV protease (which is also 6His tagged) using Ni-NTA beads,
the protein was loaded on a size exclusion (HiLoad 26/60 Superdex 200, Amersham Pharmacia Biotech) column pre-equilibrated with NMR buffer (50 mM MES [pH 6.5], 50 mM NaCl, 0.1 mM EDTA) and eluted in the same buffer. The protein was concentrated to 1 mM; HALT protease inhibitor, NaN₃ and TSP were added to 1X, 0.01% and 0.14 mM respectively. The sample was stored at –80 °C until further use.

4.2.2 NMR backbone resonance assignments of BamA POTRA3–5.

NMR spectra for backbone (¹H, ¹⁵N, ¹³Cα, ¹³Cβ) resonance assignments were collected on a ²H, ¹³C, ¹⁵N POTRA3–5 sample concentrated to 1 mM in NMR buffer (50 mM MES [pH 6.5], 50 mM NaCl, 0.1 mM EDTA, 0.01% NaN₃, 0.15 mM TSP, 1X HALT, 10% (v/v) ²H₂O). NMR experiments for backbone assignments were collected using the TROSY-based: 2D ¹H, ¹⁵N HSQC, 3D HNCACB, 3D HN(CO)CA, 3D HN(COCA)CB, and 3D ¹⁵N-edited NOESY-HSQC. Experiments were collected at 30 °C on VNMRS 900 MHz and 800 MHz spectrometers equipped with HCN z-axis gradient cold probes. Spectra were processed with NMRPipe (Delaiglio et al., 1995) and analyzed with CCPNMR Analysis (Vranken et al., 2005a). Filenames and parameters used for backbone resonance assignments of POTRA3–5 are given in Table A7.1

4.2.3 NMR experiments on BamA POTRA constructs.

The amide ¹H–¹⁵N RDCs were measured on a 0.5 mM ¹H, ¹³C, ¹⁵N-labeled POTRA1–5 sample with no Pf1 phage and a 0.5 mM ¹³C, ¹⁵N-labeled POTRA1–5 sample in 10 mg/mL liquid crystalline Pf1 phage, prepared as described previously (Hansen et al., 1998). Two methods were used to measure RDCs. First, TROSY and decoupled ¹H, ¹⁵N HSQC spectra were collected on isotropic (no Pf1 phage) and aligned (with Pf1 phage) samples to measure the ¹H–¹⁵N RDCs for orientation restraints (Kontaxis et al., 2000). The difference in chemical shifts for peaks in the decoupled and TROSY experiments are ½ ¹JNH in the isotropic and ½ (¹JNH+¹DNH) in the aligned
spectra. Thus, the $^1\text{D}_{\text{NH}}$ can be calculated from the difference in the splitting in both spectra. Second, a method for measuring amide RDCs by TROSY (ARTSY) (Fitzkee and Bax, 2010) which implements a modification of a quantitative J-correlation experiment was used to measure $^1\text{D}_{\text{NH}}$ on an aligned (with Pf1 phage) sample of POTRA1–5. In this case, a “reference” and an “attenuated” spectrum are collected. The $^1\text{D}_{\text{NH}}$ is a function of the phase and amplitude of peaks in an attenuated spectrum, relative to the reference spectrum as shown in Equation 4.1 (Fitzkee and Bax, 2010), where Q is the intensity ratio (attenuated intensity over the reference intensity), $^1\text{J}_{\text{NH}} \approx -93$ Hz and $T = 10.7$ ms. Filenames and parameters used for measuring $^1\text{D}_{\text{NH}}$ of POTRA1–5 are given in Table A7.3.

**Equation 4.1**

$$^1\text{D}_{\text{NH}} = -\frac{1}{T} - ^1\text{J}_{\text{NH}} + \left(\frac{2}{\pi T}\right)\sin^{-1}\left(\frac{Q}{2}\right)$$

The data for determining $\{^1\text{H}\}$-$^1\text{H}$ heteronuclear NOE data and $^{15}\text{N}$ $R_I$ and $R_{\rho}$ relaxations rates were collected on a 0.5 mM $^2\text{H}, ^{13}\text{C}, ^{15}\text{N}$ POTRA1–5 sample at 800 MHz at 30 °C. A low power $^1\text{H}$ presaturation period of 3 s was applied in the $\{^1\text{H}\}$-$^1\text{H}$ heteronuclear NOE experiment. Relaxation delays for the $R_I$ and $R_{\rho}$ experiments were collected in an interleaved fashion to create a pseudo-3D fid file. For the $R_I$ experiment, spectra were collected with relaxation delays of 0, 105, 210, 315, 525*, 1050, 1573* and 2622 ms with duplicates indicated with an asterisk. For the $R_{\rho}$ experiment, spectra were collected with relaxation delays of 0*, 5, 10*, 20* 40, 60, and 80 ms with duplicates indicated with an asterisk with a $^{15}\text{N}$ spin lock field of 1908 Hz. The volumes from readily resolved peaks were fit to an exponential decay function to determine $R_I$ and $R_{\rho}$ values. $R_2$ values were estimated from $R_I$ and $R_{\rho}$, neglecting relaxation from chemical exchange, using Equation 4.2 and Equation 4.3 (Kempf and Loria, 2003):
Residue-specific $\tau_c$ values were calculated from the $^{15}$N $R_1$ and $R_2$ values for residues with $^1$H-$^{15}$N heteronuclear NOE values greater than 0.7 using Equation 4.4 (Simplification of Eq. 8 in (Kay et al., 1989)).

$$\tau_c = \frac{1}{4\pi n} \sqrt{\frac{6R_2}{R_1} - 7}$$

Paramagnetic relaxation enhancement was used to qualitatively measure the distance between the POTRA1–2 and POTRA3–5 domains. A mutation of Ser143 to cysteine was introduced to PORTA1–5 using a Stratagene QuickChange® kit. The manufacturer protocols were followed and the resulting mutation was confirmed by DNA sequencing. The nitrooxide paramagnetic tag MTSL (S-(2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate, Toronto Research Chemicals, Inc) or the diamagnetic control Ac-MTSL (1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl methanethiosulfonate, Toronto Research Chemicals, Inc) were coupled to the protein through a disulfide bond to Cys143. $^1$H, $^{15}$N HSQC spectra were collected on 0.075 mM samples of $^2$H, $^{15}$N BamA POTRA1–5 S143C coupled with MTSL and Ac-MTSL respectively. Coupling was performed as previously described (Su and Otting, 2010). Briefly, $^3$H, $^{15}$N BamA POTRA1–5 S143C in 50 mM MES, pH 6.5, 50 mM NaCl, 0.1 mM
EDTA, 2 mM DTT was concentrated and diluted into a buffer of a 50 mM MES, pH 6.5, 50 mM NaCl, 0.1 mM EDTA, to a final protein concentration of ~200 μM. MTSL or Ac-MTSL was added to ~2 mM and the solutions were allowed to react at room temperature for 2 h and then overnight at 4 °C. Unreacted MTSL or Ac-MTSL was removed the following morning by passing the solution over a protein buffer exchange column (Pierce). 1H, 15N HSQC spectra were then collected at 30 °C on the tagged proteins. The protein concentration was reduced to 0.075 mM to ensure that the measured PREs were due to intramolecular and not intermolecular interactions. The effect of the paramagnetic relaxation was quantified as the ratio of the intensity of the paramagnetic and diamagnetic spectra. The filenames and experimental conditions are listed in. Several 1H, 15N HSQC TROSY and decoupled spectra were collected on samples of POTRA3–4, POTRA1–2 and POTRA1–5 mutants. The filenames and experimental conditions for these experiments are listed in Table A7.3.

4.3 Results and Discussion

4.3.1 Backbone resonance assignment of POTRA3–5

POTRA1–5 is a large protein (46.5 kDa) by NMR standards and shows low peak intensity and broad lines in the 1H, 15N HSQC. A comparison of 1H, 15N spectra from a fully protonated POTRA1–5 sample at 30 °C, 35 °C and 40 °C (Figure 4.4a-c) and a perdeuterated sample at 30 °C (Figure 4.4d) demonstrates that increasing temperature and perdeuteration dramatically improve spectra quality. However, even with these improvements to the 1H, 15N HSQC spectra, de novo sequential resonance assignment of proteins with molecular mass greater than ~20 kDa suffer from two serious challenges (Arrowsmith and Wu, 1998; Salzmann et al., 1998; Riek et al., 2002; Cavanagh et al., 2007). First, as the molecular mass of a protein increases, the rotational correlation time increases, which ultimately results in larger linewidths of peaks; a problem
that is exacerbated in the multidimensional spectra needed for backbone resonance assignments (Riek et al., 2002; Cavanagh et al., 2007). Second, in larger proteins, a greater the number of resonances crowd spectra, making assignments difficult (Cavanagh et al., 2007). With the introduction of deuteration and TROSY techniques (Salzmann et al., 1998; Pervushin et al., 1997) that effectively decrease the linewidth of peaks in larger proteins, assigning the backbone of proteins from \( \sim 25-40 \) kDa is becoming readily achievable (Cavanagh et al., 2007). For proteins greater than \( \sim 40 \) kDa, the spectral overlap becomes the greater challenge. Several strategies exist to overcome this obstacle including specific isotope and segmental labeling (Muona et al., 2010; Skrisovska et al., 2010; Ohki and Kainosho, 2008). Amino acid specific isotope labeling, where proteins produced in an \( E. \ col i \) or cell-free expression system can be amino acid type-selectively \( ^2\)H-, \( ^{13}\)C- and/or \( ^{15}\)N-labeled which significantly simplify NMR spectra (Ohki and Kainosho, 2008). Segmental isotope labeling uses a post-translational chemical modification with the peptide splicing strategy to join domains together (Muona et al., 2010; Skrisovska et al., 2010). Using this strategy in a multi-domain protein, specific isotope labeling of domains can be achieved (Muona et al., 2010; Skrisovska et al., 2010). One simple solution is to break the protein into smaller units, assign the resonances in the smaller units and then transfer the assignments to the spectra of the full-length protein. The multi-domain architecture of POTRA1–5 lends itself to examination of the domains as independent units and this simplified strategy was employed here. Previous NMR studies on POTRA1–2 and POTRA4–5 were successful in assigning 100% and 95% of the backbone resonances of each of these domains. Thus, a divide and conquer approach was implemented to assign the \( ^1\)H, \( ^{15}\)N amide spectrum of POTRA1–5.
Figure 4.4 Increasing the temperature and deuterating POTRA1–5 significantly improves spectral quality in the amide region of the $^1$H, $^{15}$N HSQC spectrum.

Increased temperature and deuteration increase peak dispersion and sharpens linewidth in TROSY $^1$H, $^{15}$N HSQC spectra of a $^1$H, $^{15}$N POTRA1–5 at 30 °C (a), 35 °C (b), and 40 °C (c) and a $^2$H, $^{15}$N POTRA1–5 sample at 30 °C (d) all collected at 800 MHz.

Since POTRA3 was the only domain unassigned, initial attempts were made to produce a sample with the third domain alone or in tandem with POTRA1–2. Interestingly, while constructs of POTRA1–5, POTRA1–4, POTRA1–2, POTRA3–5, and POTRA4–5 can all be
expressed as soluble, well-behaved proteins, any construct that exposed the C-terminus of the third POTRA domain, such as POTRA1–3, and 3 alone, were expressed in inclusion bodies and not amenable to refolding (data not shown, Susan Baker). Thus, a construct of POTRA3–5 was chosen for further assignment experiments. The $^1$H, $^{15}$N HSQC spectra of POTRA1–2 and POTRA3–5 overlay well with the spectrum of the full length POTRA1–5 as seen in Figure 4.5. The amide $^1$H, $^{15}$N HSQC spectrum of a 1 mM $^2$H, $^{13}$C, $^{15}$N POTRA3–5 sample collected at 900 MHz at 30 °C has 210 of the 248 predicted backbone amide peaks (Figure 4.6).

To assign the backbone amide $^1$H, $^{15}$N chemical shifts of POTRA3–5, assignments were first transferred from POTRA4–5, and then verified by analyzing the backbone connectivities. Backbone connectivities were established for the $^{13}$C$^\alpha$ and $^{13}$C$^\beta$ chemical shifts of the $i$ and the $i$-1 residues to the backbone $^1$H, $^{15}$N of the $i$ residue. Four NMR experiments were used here to correlate the $^{13}$C$^\alpha$ and $^{13}$C$^\beta$ chemical shifts to the backbone $^1$H, $^{15}$N: the $^1$H, $^{15}$N HSQC; the 3D HNCACB, which detects $^{13}$C$^\alpha$ and $^{13}$C$^\beta$ chemical shifts for the $i$ and the $i$-1 residues; the 3D HN(CO)CA and the 3D HN(COCA)CB, which detects $^{13}$C$^\alpha$ and $^{13}$C$^\beta$ chemical shifts for the $i$-1 residue respectively. As previously reported, the backbone resonance assignments for POTRA4–5 were readily assigned using the automated program, PINE (Gatzeva-Topalova et al., 2010). The success of automated assignment programs hinges the accuracy of the peak lists that are input into the program, which depend on spectra with high signal-to-noise and well-resolved peaks. The signal-to-noise and resolution of the 3D spectra of POTRA3–5 were not high enough to use an automated assignment program and thus backbone connectivities were manually constructed (Figure A5.1 to A5.13).
Figure 4.5 $^1$H, $^{15}$N HSQC spectra POTRA1–2 and POTRA3–5 overlay with POTRA1–5.

$^1$H, $^{15}$N HSQC spectra of $^2$H, $^{13}$C, $^{15}$N POTRA1–2 (green) and $^2$H, $^{13}$C, $^{15}$N POTRA3–5 (purple) (a) reproduce the pattern of amide backbone peaks observed in $^2$H, $^{13}$C, $^{15}$N POTRA1–5 (black) (b). An overlay of $^1$H, $^{15}$N HSQC spectra of $^2$H, $^{13}$C, $^{15}$N POTRA1–2 (green), $^2$H, $^{13}$C, $^{15}$N POTRA3–5 (purple), and $^2$H, $^{13}$C, $^{15}$N POTRA1–5 (black) (c) demonstrates that most peaks show little to no chemical shift perturbations between the multi-domain constructs and the full-length POTRA1–5. Exceptions included residues located at domain-domain interfaces such as Ser148 (d).
Figure 4.6 An $^1$H, $^{15}$N HSQC spectrum of $^2$H, $^{13}$C, $^{15}$N POTRA3–5.

$^1$H, $^{15}$N HSQC spectrum of $^2$H, $^{13}$C, $^{15}$N POTRA3–5 collected at 800 MHz at 30 °C. The spectrum has 210 well-resolved non-side chain NH peaks. The backbone amide $^1$H, $^{15}$N assignments for 177 of the 210 peaks were made.

Backbone, $^1$H$^N$, $^{15}$N$^H$, $^{13}$C$^\alpha$, and $^{13}$C$^\beta$, chemical shift assignments for 177 of the 210 peaks in the $^1$H, $^{15}$N spectra of POTRA3–5 were made. The remaining 33 residues could not be assigned due to line broadening and overlap of resonances in the HNCACB, HN(CO)CA and HN(COCA)CB spectra. Breaks in the backbone connectivities in POTRA3 for residues 177–180, 195–226, and 243–258 prevented chemical shift assignments for these residues. The $^1$H, $^{15}$N
HSQC peak intensities decrease in the residues neighboring the unassigned residues as shown in Figure 4.7a. Additionally, the unassigned residues cluster to a region of POTRA3 that is at the POTRA2–3 interface Figure 4.7b. The region encompassing residues 195–226 in POTRA3 was found in different conformations in two crystal structures of POTRA1–4, as seen in a superposition of POTRA3 from 3EFC (Gatzeva-Topalova et al., 2008) and 2QCZ (Kim et al., 2007) (Figure 4.8). Moreover, residues 200–216 in the loop connecting α-helices 1 and 2 had a zero occupancy in the 2QCZ crystal structure (Kim et al., 2007), presumably due to conformational flexibility. However, this region is involved in a crystal contact that might affect its conformation. The clustering of residues with a loss of signal intensity, coupled with the conformational flexibility indicated by the crystal structure in the same region, suggests that there is conformational heterogeneity in POTRA3. However, spectra were collected on the $^3$H, $^{13}$C, $^{15}$N POTRA3–5 at 30 °C and 35 °C at 600 MHz, 800 MHz and 900 MHz without any noticeable changes in the peak intensities or appearance of new peaks, suggesting the dynamics are not on the chemical shift timescale (μs-ms).

Figure 4.7 Residues with weak peak intensity and unassigned residues in the $^1$H, $^{15}$N, HSQC cluster to the POTRA3 domain that interfaces with POTRA2.

(a) Weak intensity of peaks in the $^1$H, $^{15}$N HSQC is observed for residues neighboring unassigned regions of the protein, suggesting increased dynamics. (b) Residues of POTRA1–5 for which $^1$H, $^{15}$N amide resonances were unassigned (red) cluster to a region of POTRA3 that is located in the putative hinge region between POTRA2 and 3.
Figure 4.8 Unassigned POTRA3 residues have zero occupancy in crystal structure.
A superposition of the POTRA3 domain from 3EFC (Gatzeva-Topalova et al., 2008) (orange) and 2QCZ (Kim et al., 2007) (yellow and purple). Residues 200–216 had a zero occupancy in 2QCZ (Kim et al., 2007) crystal structure and are depicted as a dashed purple line. These same residues were unassigned in the POTRA3–5 $^1$H, $^{15}$N spectrum.

4.3.2 Assignment of the $^1$H, $^{15}$N backbone amide resonances of POTRA1–5

The backbone $^1$H, $^{15}$N amide peak assignments for POTRA1–2 and POTRA3–5 that showed minimal chemical shift perturbations when compared to the POTRA1–5 $^1$H, $^{15}$N HSQC spectrum, were transferred to the $^1$H, $^{15}$N HSQC spectrum of POTRA1–5. A total of 317 chemical shift assignments were transferred to POTRA1–5, with all 144 assignments from POTRA1–2 and 173 of the 177 assignments from POTRA3–5. The subsequent RDC and backbone $^{15}$N relaxation experiments on POTRA1–5 required the analysis of only well-resolved peaks; therefore, peaks with any overlap were disregarded. Limiting the transferred assignments to peaks that showed minimal chemical shift perturbations and no overlap resulted in assignment of 256 of 317 peaks in the $^1$H, $^{15}$N HSQC spectrum of POTRA1–5 (Figure 4.9).
Figure 4.9 $^1$H, $^{15}$N HSQC of 0.5 mM $^2$H, $^{13}$C, $^{15}$N POTRA1–5 at 30 °C and 800 MHz.

The chemical shifts assignments for 317 of the amide $^1$H, $^{15}$N peaks were transferred from POTRA1–2 and POTRA3–5 to POTRA1–5.

4.3.3 Backbone $^{15}$N relaxation data indicate flexibility in POTRA1–5.

A set of $^{15}$N relaxation experiments was collected on a 0.5 mM $^2$H, $^{13}$C, $^{15}$N POTRA1–5 sample to assess the fast timescale (ps-ns) backbone dynamics. A \{$^1$H\}-$^{15}$N heteronuclear NOE experiment was performed to determine if there were regions of POTRA1–5 with above average fast timescale dynamics, which would indicate a highly flexible linker. The representation of \{$^1$H\}-$^{15}$N heteronuclear NOE values was typical of a protein with regular secondary structural element. Decreased \{$^1$H\}-$^{15}$N heteronuclear NOE values were observed at the end of strands and
helices and in loop regions (Figure 4.10 and Figure 4.11). One exception is the \(^{1}\text{H}\)-\(^{15}\text{N}\) heteronuclear NOE values for the \(\alpha 1\) helix of POTRA4 measured on POTRA1–5, which demonstrated decreased values compared to those measured on a construct containing only POTRA4–5 (f and Section 3.3.1). The loop connecting the \(\alpha 1\) and \(\alpha 2\) helices shows decreased \(^{1}\text{H}\)-\(^{15}\text{N}\) heteronuclear NOE values in measurements made on both the POTRA1–5 and the POTRA4–5 constructs. There are no unusually low \(^{1}\text{H}\)-\(^{15}\text{N}\) heteronuclear NOE values that might indicate a flexible linker, such as those reported for the proteins BamC and Ca\(^{2+}\)-loaded calmodulin (Warner et al., 2011; Barbato et al., 1992). In these proteins, the linkers were at least 10 residues in length and displayed \(^{1}\text{H}\)-\(^{15}\text{N}\) heteronuclear NOE values that were similar to those observed at the termini (well below 0.5). The POTRA domains are connected with shorter 3–4 amino acid linkers, where interdomain motions would likely be subtle (if present at all).

The extent of the relative flexibility between domains in a multi-domain protein can be measured with the apparent rotational correlation times, \(\tau_\text{c, app}\), or the average of the residue specific rotational correlation time over a region of the protein. If the POTRA domains tumble as a single rigid species in solution, then the residue specific rotational correlation times should be constant across all the domains. Excluding residues with \(^{1}\text{H}\)-\(^{15}\text{N}\) heteronuclear NOE values less than 0.7, the residues in the POTRA1–2 have a \(\tau_\text{c, app}\) value of 24 ± 4 ns, whereas POTRA3–5 has a longer \(\tau_\text{c, app}\) value of 36 ± 4 ns in seen in Figure 4.11b. The differences in \(\tau_\text{c, app}\) could arise from POTRA1–2 forming one rigid arm and POTRA3–5 forming a second rigid arm that tumble independently in solution. However, the differences in \(\tau_\text{c, app}\) could also be due to the anisotropic tumbling of POTRA1–5 with a rigid orientation of POTRA1–2 relative to POTRA3–5. To accurately evaluate the anisotropic diffusion tensor using a model-free approach would require an extensive field dependence study of the \(^{15}\text{N}\) relaxation (Lipari and Szabo, 1982; Clore et al.,
Additionally, because the explicit solution orientation of POTRA1–5 domains is unknown, a “sample and select” ensemble approach would have to be taken in order to determine conformational ensembles that were consistent with the NMR $^{15}$N relaxation data (Chen et al., 2007).

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure4.png}
\caption{$^1$H-$^{15}$N heteronuclear NOE values mapped onto POTRA1–5 domains. $^1$H-$^{15}$N heteronuclear NOE values measured on 0.5 mM $^2$H, $^{13}$C, $^{15}$N BamA POTRA1–5 at 30 \degree C and 800 MHz (a-e) and 0.5 mM $^1$H, $^{13}$C, $^{15}$N BamA POTRA4–5 at 30 \degree C and 800 MHz (f and g) mapped onto the POTRA domains. Values of $^1$H-$^{15}$N heteronuclear NOE less than 0.7 (red, amide hydrogens are shown as white sticks for clarity), greater than 0.7 (blue) and residues that were unassigned (light blue) are shown on POTRA1 (a), POTRA2 (b), and POTRA3 (c) domains from 3EFC (Gatzeva-Topalova et al., 2008) and POTRA4 (d) and POTRA5 (e) domains from 3OG5 (Gatzeva-Topalova et al., 2010). Low values indicate increased fast timescale (ps-ns).}
dynamics and are typically found at the end of strands, ends of helices, and in loops as seen in domains 1, 2, 3 and 5. One exception is the α1 helix of POTRA4 in POTRA1–5, which demonstrated increased fast timescale dynamics compared to a construct containing only POTRA4–5 (f and Section 3.3.1). The α1 helix is connected to the α2 helix by a loop that shows decreased $\{^{1}\text{H}\}$-$^{15}\text{N}$ heteronuclear NOE values in both the POTRA1–5 and the POTRA4–5 constructs.

**Figure 4.11** Backbone $^{15}\text{N}$ relaxation data indicate flexibility in POTRA1–5.

Plots of (a) $\{^{1}\text{H}\}$-$^{15}\text{N}$ heteronuclear NOE, (b) the calculated residue-specific rotational correlation time $\tau_c$, (c) the $^{15}\text{N} R_1$ and (d) the $^{15}\text{N} R_2$ as a function of residue number. A $\{^{1}\text{H}\}$-$^{15}\text{N}$ heteronuclear NOE threshold value of 0.7 is shown as a dashed black line in (a). Average
rotational correlation time $\tau_c$ values of 24 ns and 36 ns are shown in panel (b) as red and blue dashed lines, respectively. The domain architecture is shown at the top.

4.3.4 **Paramagnetic relaxation enhancement shows that POTRA1–5 transiently samples a “compact” conformation in solution.**

Paramagnetic relaxation enhancement (PRE) is the result of magnetic dipolar interactions between an NMR active nucleus (e.g. $^1$H or $^{15}$N) and unpaired electrons. This interaction causes an increase in the relaxation rate of the NMR active nucleus, which can be observed by a loss of signal intensity of the NMR active nuclei in a spectrum relative to one collected without the paramagnetic agent in solution (Bloembergen and Morgan, 1961). The magnitude of the interaction is dictated by the electron–nucleus distance $r$, and is proportional to $r^{-6}$, similar to the dependence of the NOE on $^1$H-$^1$H distances (Kleckner and Foster, 2011; Clore et al., 2007). However, because the magnetic moment of the unpaired electron is large, the PRE effects are large and can provide long-range distance information (~25 Å) (Kleckner and Foster, 2011; Clore et al., 2007), whereas the $^1$H-$^1$H NOE is generally limited to less than ~5 Å or in cases of selective isotope labeling up to ~8 Å (Mueller et al., 2000). A paramagnetic tag can be readily incorporated into a protein by covalently attaching a “spin label”, such as a nitroxide group or an EDTA-Mn$^{2+}$ compound, both of which contain a pair of free electrons, onto a cysteine moiety (Kosen, 1989). PREs can be used to quantitate distances of protons in the protein to the spin label by a relationship of the intensity ratio of peaks in a spectrum collected with the paramagnetic tag ($I_{para}$) to peaks in a spectrum collected with a diamagnetic tag ($I_{dia}$) (Battiste and Wagner, 2000). However, for transient interactions between the electron pair and a nucleus in the fast exchange time regime, where the rate constant for the conformational exchange between two conformations is much greater than the difference in PRE relaxation rates ($k_{ex} \gg \Delta I$), the measured PRE is a population-weighted average of the contributions from all distances (Kleckner and
Foster, 2011). Qualitative information about whether a distance threshold between an electron pair and a nucleus was crossed is still included in the measured PRE and can therefore be used to determine if the pair is ever within ~25 Å from one another. Here, qualitative PRE was used to provide evidence that POTRA1–5 can transiently adopt a “compact” configuration in solution.

A S148C mutation was designed into POTRA1–5 to attach MTSL, a paramagnetic nitrooxide spin label, or an acetylated diamagnetic form of MTSL (Ac-MTSL) used as the reference. The location of the label was chosen based on the “extended” (PDB:3EFC (Gatzeva-Topalova et al., 2008)) and “compact” (PDB:2QCZ (Kim et al., 2007)) crystal structures, such that a measurable PRE effect could be observed for residues in the α1 and α2 helix of POTRA3 only if the protein adopted a “compact” conformation (Figure 4.12a and b). Experimental PREs as measured from the intensity ratio of peaks in an $^1$H, $^{15}$N HSQC from a POTRA1–5 sample prepared with MTSL and with Ac-MTSL. PREs ranging from 0.6 to 0.8 for assigned peaks were observed in domains 1, 2, and 3. In particular, PREs were observed in the α1 and α2 helices of POTRA3 supporting the theoretical PREs based on the “compact” (PDB:2QCZ (Kim et al., 2007)) crystal structure (Figure 4.12a and c). Interestingly, PREs were measured in POTRA1, which were not predicted based on the crystal structures (Figure 4.12a, b and c) but support a model in which POTRA1–5 samples a variety of “compact” orientations in solution, perhaps including some flexibility between domains 1 and 2.
Figure 4.12 PRE shows that POTRA1–5 transiently adopts a “compact” configuration in solution.

Theoretical PREs based distances from the terminal oxygen of Ser148 (pink) were used to approximate MTSL with a 25 Å threshold are shown in blue for POTRA1–3 fragments of the “extended” 3EFC (Gatzeva-Topalova, et al., 2008) crystal structure (a) and the “compact” 2QCZ (Kim, et al., 2007) crystal structure (b). Differences in the theoretical PREs are noted in the α1 and α2 helices of POTRA3. Experimental PREs from 0.6 to 1.0 (blue to yellow) where 0.8 (transition to yellow) is the upper threshold for PRE were observed in domains 1, 2, and 3 (c, unassigned residues are shown in black). Of note are PREs that were observed in the α1 and α2 helices of POTRA3 as well as in POTRA1 which were not predicted based on the crystal structures, but support POTRA1–5 sampling a variety of “compact” orientations in solution.

4.3.5 Measuring RDCs in the 45.6 kDa POTRA1–5.

RDCs can provide information on the orientation of individual bond vectors in a molecule and can be used to determine the flexibility and relative orientation of rigid domains in a macromolecule (Bax et al., 2001; Fischer et al., 1999). The RDC of a protein backbone amide, $^{1}{D}_{NH}$, can be obtained by measuring the one bond $^{15}$N-$^{1}$H couplings in a partially aligned (with Pf1 phage) and unaligned (no Pf1 phage) solution. The $^{15}$N-$^{1}$H couplings can be measured in a
2D NMR spectrum by removing $^1$H decoupling during $^{15}$N evolution (Tolman et al., 1995) (Figure 4.13a and d); however, the increase in peak overlap due to the doubling of resonances becomes prohibitive for proteins with more than $\sim$75 residues. The two components of the split peak can be separated into separate spectra using the IPAP method (Ding and Gronenborn, 2003) or a spin-state selective polarization transfer (Lerche et al., 1999) method (Figure 4.13b and d). In large proteins (> 20 kDa), the components of the split peaks have very different relaxation properties, with the upfield component suffering from adverse relaxation properties that result in severe line broadening in the case of large proteins, making it difficult if not impossible to measure the couplings accurately (Fitzkee and Bax, 2010).

![Figure 4.13 Schematic representation of couplings measured using three methods for a small and large protein.](image)

Couplings measured from the peak splittings in small proteins (< 20 kDa a-c) and large proteins (> 20 kDa, d-f) observed by removing $^1$H decoupling in $^{15}$N chemical shift evolution (Tolman, et al., 1995) (a) and (d), IPAP (Ding and Gronenborn, 2003) or spin-state selective polarization transfer (Lerche et al., 1999) (b) and (e), and halved couplings between decoupled and TROSY $^1$H, $^{15}$N HSQC (Kontaxis et al., 2000) (c) and (f). Red represents peaks collected in a separate spectrum from those in black.
Another method utilizes the halved couplings that result from the splitting between peak positions in spectra collected with conventional decoupling or TROSY peak selection in the $^{15}$N evolution period (Kontaxis et al., 2000). In this method, a $^1$H, $^{15}$N HSQC spectrum is collected on an unaligned sample with $^1$H decoupling during $^{15}$N evolution and a second spectrum is collected with TROSY peak selection (Figure 4.13c and f). The difference in peak position between the two spectra is equal to $1/2(J_{\text{NH}})$. The spectra are collected again, but this time on a partially aligned sample, where the difference in peak position is equal to $1/2(J_{\text{NH}} + D_{\text{NH}})$. The $D_{\text{NH}}$ can then be calculated. This method also becomes difficult with larger proteins because the peaks in the decoupled $^1$H, $^{15}$N HSQC spectrum become broad due to the mixing of the broad upfield component with the narrow downfield component.

Recently, a pulse sequence was developed to measure amide RDCs by TROSY Spectroscopy (ARTSY) that incorporates a TROSY detection scheme with a quantitative J experiment, where the size and magnitude of the splitting are a function of the relative intensities of spectra collected at two different J modulation time points (Fitzkee and Bax, 2010). In the reference spectrum, the intensity of the peaks is modulated by $\sin[\pi(J_{\text{NH}} + D_{\text{NH}})T/2]$, where $J_{\text{NH}}$ is the J-coupling between $^1$H and $^{15}$N, $D_{\text{NH}}$ is the residual dipolar coupling between $^1$H and $^{15}$N and $T$ is the time of $^1$H dephasing duration of the first INEPT transfer module. When $T$ is set to $(|J_{\text{NH}}|)^{-1}$, which is assumed as $J_{\text{NH}} = -93$ Hz, then the intensity of the peaks at a maximum is used as a reference. For the attenuated spectrum, the intensity of the peaks is modulated by $\sin[\pi(J_{\text{NH}} + D_{\text{NH}})T]$. In this case, for an unaligned sample, where $D_{\text{NH}}$ averages to zero, the intensity of the peaks is zero (this is assuming that the $J_{\text{NH}}$ for all backbone amides is $-93$ Hz). The consequence of this assumption is that for peaks with a $J_{\text{NH}} > -93$ Hz, the value of the calculated $D_{\text{NH}}$ will be larger than the true $D_{\text{NH}}$, and for peaks with a $J_{\text{NH}} < -93$ Hz, the value of...
the calculated $^{1}D_{\text{NH}}$ will be smaller than the true $^{1}D_{\text{NH}}$. A deviation from the true $^{1}J_{\text{NH}}$ is such that a peak with a $^{1}J_{\text{NH}}$ of −92 Hz will have a calculated $^{1}D_{\text{NH}}$ that is 1 Hz greater than the true value. In practice, the $^{1}J_{\text{NH}}$ has been observed to vary between −92 and −94 Hz. Thus, the error from this assumption is far less than that associated with the error in measuring the peak position in a decoupled HSQC for large proteins, where at a magnetic field strength of 18.7 Tesla (800 MHz \(^{1}\text{H}\) Larmor frequency), an error in peak position of 0.05 ppm in the \(^{15}\text{N}\) dimension (80 MHz \(^{15}\text{N}\) Larmor frequency) results in an error in the measured $^{1}D_{\text{NH}}$ of 4 Hz. Another benefit of using the ARTSY over other methods is that the $^{1}D_{\text{NH}}$ is measured from just the partially aligned sample. While other methods rely on measuring splittings on a partially aligned and unaligned sample, the ARTSY requires only the partially aligned sample. Here, RDCs for POTRA1–5 were measured from the halved couplings and the ARTSY method.

As expected, the peaks in the decoupled HSQC spectra of \(^{2}\text{H},\ \^{13}\text{C},\ \^{15}\text{N}\) POTRA1–5 were significantly broadened, especially in the partially aligned sample which suffers increased from \(^{1}\text{H}-^{1}\text{H}\) dipolar couplings increase line broadening in the \(^{1}\text{H}\) dimension (Figure 4.14a and b, red). Only 108 of the 256 well resolved, assigned peaks from the TROSY spectrum were able to be distinguished in the decoupled spectrum. For disperse peaks, the halved couplings were measured for an unaligned (no Pf1 phage) and partially aligned (10 mg/ml Pf1 phage) of 0.5 mM \(^{2}\text{H},\ \^{13}\text{C},\ \^{15}\text{N}\) POTRA1–5. The spectra in Figure 4.14 illustrate the problems associated with line broadening in these experiments. This region of the spectra had relatively well resolved peaks. While the peak positions were well defined in the TROSY spectra of the aligned and unaligned samples, the peak positions in the decoupled spectra were difficult to determine precisely. For example, the peaks of Val26 and Lys152 had well defined maxima in all spectra, but Ala343 and Val415 peaks were broadened in the decoupled HSQC spectra to the point that the precision of
determining the peak maximum was limited to \( \pm 0.1 \text{ ppm} (\pm 8 \text{ Hz}) \) in the \( ^{15}\text{N} \) dimension. Only 108 of the 264 assigned peaks were well resolved enough in the decoupled \( ^{1}\text{H}, ^{15}\text{N} \) HSQC to precisely measure splittings.

![Figure 4.14](image)

**Figure 4.14 \( ^{1}\text{H}-^{15}\text{N} \) RDCs for POTRA1–5 were measured from the halved couplings.**

The \( ^{1}\text{H}-^{15}\text{N} \) RDCs for POTRA1–5 were measured from the halved couplings that result from the difference in peak positions in a TROSY (black) and a conventional decoupled (red) \( ^{1}\text{H}, ^{15}\text{N} \) spectrum. The RDC is then calculated by the difference in couplings measured from an unaligned (a) and partially aligned in 10 mg/ml Pf1 phage (b) in 0.5 mM \( ^{2}\text{H}, ^{13}\text{C}, ^{15}\text{N} \) POTRA1–5 measured at 30º C at 800 MHz.

The ARTSY method offers a major advantage for large proteins because the spectra are collected using only the sharp TROSY component of the \( ^{1}\text{H}, ^{15}\text{N} \) HSQC, thus removing the adverse relaxation effects that causes line broadening in the decoupled HSQC. Additionally, the RDC is calculated from spectra collected on only one sample, the partially aligned sample, which is a clear advantage for proteins that are difficult to express and purify. Here, RDCs were measured for 264 well resolved peaks in the \( ^{1}\text{H}, ^{15}\text{N} \) HSQC. The reference and attenuated spectra were collected on a partially aligned (10 mg/ml Pf1 phage) 0.5 mM \( ^{2}\text{H}, ^{13}\text{C}, ^{15}\text{N} \) POTRA1–5 sample and are shown in Figure 4.15 with the same spectral region as in Figure 4.14 for comparison. The 96 RDCs measured from the halved couplings were compared with the corresponding RDCs measured with the ARTSY method directly. Figure 4.29a shows the correlation between the
methods is relatively rmsd of 6.80 Hz between the datasets. The 264 RDCs measured with the ARTSY method were used for further analysis.

![ARTSY spectra](image)

**Figure 4.15 ARTSY spectra used to measure $^1$D$_{\text{NH}}$ for POTRA1–5.**

The RDCs are calculated from a ratio of intensities of peaks in the reference spectrum (a) and the attenuated spectrum (b). Positive and negative intensity contours are blue and red, respectively. In the attenuated spectrum, positive (negative) peaks correspond to resonances with $^1$D$_{\text{NH}}$ greater than (less than) 0 Hz.

**4.3.6 Analysis of RDCs shows that POTRA4–5 orient independently of POTRA1–2 and POTRA3.**

Based on crystallographic and EPR evidence from POTRA1–4 and POTRA1–2, respectively, (Kim et al., 2007; Gatzeva-Topalova et al., 2008; Ward et al., 2009), and SAXS data from POTRA1–5 (Gatzeva-Topalova et al., 2010), it is suspected that flexibility in POTRA1–5 is limited to the POTRA2–3 hinge region. Here, the flexibility was assessed by an analysis of the alignment tensors for the individual domains of POTRA1–5. The values of $D_a$ and $R$ for alignment tensors of the individual domains give an indication of the flexibility of domains. If the alignment tensors of the individual domains have the same $D_a$ and $R$, this is consistent with the domains having a rigid orientation and orienting as a single unit. Conversely, if the values of $D_a$ and $R$ are very different, this indicates that the orientations of domains are not fixed. The RDCs measured from the ARTSY method (183 RDCs after pruning to include only residues that had $^1$H-$^{15}$N heteronuclear NOE values greater than 0.7) and the crystal structures of POTRA1–4
(PDB: 2QCZ (Kim et al., 2007)) and POTRA4–5 (PDB: 3Q6B (Zhang et al., 2011)) were used to calculate the $D_a$ and $R$ for each of the domains in POTRA1–5 using calcTensor (Xu et al., 2000; Schwieters et al., 2006). A plot of the experimental and predicted RDCs for the individual domains shows quite good agreement between experimental RDCs and those predicted from the crystal structures for all five POTRA domains (Figure 4.16b-f). POTRA4 and 5 have $D_a$ and $R$ values that are nearly identical, suggesting a rigid orientation between the domains. This is consistent with RDC measurements made on a construct of containing only POTRA4–5, where the $D_a$ and $R$ and a network of NOEs from the residues in the linker to residues in the domains showed that the domains orient together (Chapter 3 and (Gatzeva-Topalova et al., 2010)). PORTA1, 2 and 3 have $D_a$ and $R$ values that are different suggesting that they do not have a fixed orientation in solution. POTRA3 has $D_a$ and $R$ values that are very different from POTRA1 and 2 or POTRA4 and 5, which could be due to POTRA3 having a unique orientation from the other POTRA domains. However, the sampling of bond vector orientations in POTRA3 was limited due to the incomplete peak assignments (Figure 4.7), which could skew the estimated values of $D_a$ and $R$.

**Table 4.1 Values of $D_a$ and $R$ calculated for POTRA1–5.**

The values of $D_a$ and $R$ were calculated from an SVD fit of RDCs measured on 0.5 mM $^2$H, $^{13}$C, $^{15}$N POTRA1–5 in 10 mg/ml Pf1 phage and crystal structures of POTRA1–4 (PDB: 2QCZ (Kim et al., 2007)) and POTRA4–5 (PDB: 3Q6B (Zhang et al., 2011)).

<table>
<thead>
<tr>
<th></th>
<th>POTRA1</th>
<th>POTRA2</th>
<th>POTRA3</th>
<th>POTRA4</th>
<th>POTRA5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_a$</td>
<td>12.9</td>
<td>15.5</td>
<td>-4.1</td>
<td>-9.4</td>
<td>-9.0</td>
</tr>
<tr>
<td>$R$</td>
<td>0.62</td>
<td>0.53</td>
<td>0.33</td>
<td>0.16</td>
<td>0.18</td>
</tr>
<tr>
<td># of RDCs</td>
<td>31</td>
<td>41</td>
<td>17</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>
Graphs (a) to (f) show the comparison between calculated RDCs and RDCs measured from ARTSY (Hz) and experimental RDCs (Hz). The RMSD values for each graph are as follows:

- Graph (a): RMSD = 6.80 Hz
- Graph (b): RMSD = 3.57 Hz
- Graph (c): RMSD = 2.98 Hz
- Graph (d): RMSD = 2.98 Hz
- Graph (e): RMSD = 2.09 Hz
- Graph (f): RMSD = 2.97 Hz
Figure 4.16 RDCs measured using ARTSY fit well to RDCs back calculated from crystal structures.

RDCs were measured from a 0.5 mM $^2$H, $^{13}$C $^{15}$N POTRA1–5 sample with 10 mg/ml Pf1 phage at 800 MHz. The 96 RDCs measured with the halved couplings correlate well with those measured with the ARTSY method (a). The RDCs measured with the ARTSY method and the crystal structures of POTRA1–4 (PDB: 2QCZ (Kim et al., 2007)) and POTRA4–5 (PDB: 3Q6B (Zhang et al., 2011)) were used to estimate the $D_a$ and $R$ for the individual POTRA domains (b-f).

4.3.7 Conclusions/Future directions

The results here indicate that POTRA1–5 is a flexible protein with POTRA4–5 forming a rigid arm, while POTRA1–3 are relatively more flexible. Previous studies indicated that the POTRA1 and 2 are rigid (Ward et al., 2009), but the PRE data suggests there is flexibility between the domains. Moreover, an analysis of the alignment tensors for the individual domains indicates flexibility between POTRA1 and 2. The original hypothesis was that a flexible “hinge” was formed between POTRA2 and 3. The $^{15}$N relaxation data suggest that POTRA3–5 tumble in solution slower than POTRA1–2, supporting this hypothesis. The clustering of unassigned residues to region of POTRA3 that interfaces with POTRA2 and the conformational heterogeneity observed in the crystal structures suggest that the “hinge” is not limited to a linker, but is in fact part of the POTRA3 domain. Thus, the data presented here modify the hypothesis to include the POTRA3 domain as an integral part of the “hinge”. This system provides an interesting case study of a multi-domain protein with rigid arm (POTRA4–5) and three flexible domains where for example, further investigations using an “sample and select” approach could be used to generate an ensemble of structures that fit the $^{15}$N relaxation and RDC data.
Bibliography


Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A.

Lipoprotein SmpA is a component of the YaeT complex that assembles outer membrane proteins in Escherichia coli. Proc Natl Acad Sci USA 104, 6400–6405.


Appendix A1  Script for measuring methyl-methyl distances in PDB file

ILV_methyl_dist.tcl

#------------------------------------------
#The distance between methyl groups to other protons
#is calculated from the center of the three protons.
#For the amino group, first the distances from/to each
#proton in the group is calculated, and the distance between the two
#amino protons (Daa) is added to the longest
#distance among the amino protons to give the upper bound for the NOE
#and Daa is substracted from the shortest distance to give the
#lower bound of the NOE distance. If the lower bound
#is smaller than 1.8 Å, it is reset to 1.8 Å. The final
#distance used in the NOE restraint statement is the average of the
#of the upper and lower bounds, and the "substract/add" values
#are readjusted. For the methyls, wild card atom selection is used (H7*).
#For the amino protons, explicit selection of the two atoms
#are used to exclude other undesirable atoms. -- HJ Zhou 8/31/99
#------------------------------------------

#------ need modifications -------------------
#pdb file name (inside current folder).
set pdbfile 1erkH.pdb

#output NOE restraint file name
set outfile ILV_methyl_Cdist_once.dat

#NOE distance cutoff
set noecut 7.0

#------------------------------------------
#DO NOT CHANGE BELOW #---------------------
#------------------------------------------

#delete all list items with a known value
proc ldelete {list value} {
    set ix [lsearch -exact $list $value]
    while {$ix >= 0} {
        set list [lreplace $list $ix $ix]
        set ix [lsearch -exact $list $value]
    }
    return $list
}

#------------------------------------------
puts stdout "#-----------------------------------#"
puts stdout "#--- Input pdb file: $pdbfile"

set idpdb [open $pdbfile r]
while {[gets $idpdb line] >=0} {
    set plist [split [string trim $line]]
    set plist [ldelete $plist {}]
    if {[[lindex $plist 0] == "ATOM"]} {
        set hask [string range [lindex $plist 2] 0 0]
        set res [lindex $plist 3]
        set atm [lindex $plist 2]
        if {$hask=="C"} {
            foreach atom $mlist {
                if {$res == "LEU"} {
                    if {$atm == "CD1" || $atm == "CD2"} {
                        lappend matm [lindex $plist 2]
                        lappend mres [lindex $plist 3]
                        lappend mseq [lindex $plist 5]
                        lappend mxcor [lindex $plist 6]
                        lappend mycor [lindex $plist 7]
                        lappend mzcor [lindex $plist 8]
                        #puts stdout "reslist=$mres "
                    }
                }
                if {$res == "VAL"} {
                    if {$atm == "CG1" || $atm == "CG2"} {
                        lappend matm [lindex $plist 2]
                        lappend mres [lindex $plist 3]
                        lappend mseq [lindex $plist 5]
                        lappend mxcor [lindex $plist 6]
                        lappend mycor [lindex $plist 7]
                        lappend mzcor [lindex $plist 8]
                        #puts stdout "reslist=$mres "
                    }
                }
                if {$res == "ILE"} {
                    if {$atm == "CD1"} {
                        lappend matm [lindex $plist 2]
                        lappend mres [lindex $plist 3]
                        lappend mseq [lindex $plist 5]
                        lappend mxcor [lindex $plist 6]
                        lappend mycor [lindex $plist 7]
                        lappend mzcor [lindex $plist 8]
                        #puts stdout "reslist=$mres "
                    }
                }
            }
        }
    }
}
if {$res == "VAL"} {
    if {$atm == "CG1" || $atm == "CG2"} {
        lappend matm [lindex $plist 2]
        lappend mres [lindex $plist 3]
        lappend mseq [lindex $plist 5]
        lappend mxcor [lindex $plist 6]
        lappend mycor [lindex $plist 7]
        lappend mzcor [lindex $plist 8]
        #puts stdout "reslist=$mres "
    }
}
if {$res == "ILE"} {
    if {$atm == "CD1"} {
        lappend matm [lindex $plist 2]
        lappend mres [lindex $plist 3]
        lappend mseq [lindex $plist 5]
        lappend mxcor [lindex $plist 6]
        lappend mycor [lindex $plist 7]
        lappend mzcor [lindex $plist 8]
        #puts stdout "reslist=$mres "
    }
}
close $idpdb

puts stdout "Read in a total of [llength $matm] ILV methyls"
#--------------------------
#read amino group protons
#--------------------------
#set idpdb [open $pdbfile r]
#while {
  #   gets $idpdb line
  #}
  #   puts stdout "atmlist=$atom [lindex $plist 2]"
#}
#close $idpdb
#puts stdout "Read in a total of [llength $iatm] amino protons"
#} # this is end of comment 1
puts stdout "DONE reading co-
ordinates."
#puts stdout "calculating the geometric center coordinates of CH3's ..."
#
#foreach {x1 x2 x3} $mxcor {
#      lappend xav [expr ($x1+$x2+$x3)/3.0]
#      puts stdout "$x1 $x2 $x3 
#      }
#
#foreach {y1 y2 y3} $mycor {
#      lappend yav [expr ($y1+$y2+$y3)/3.0]
#      }
#
##foreach {z1 z2 z3} $mzcor {
#      lappend zav [expr ($z1+$z2+$z3)/3.0]
#      }
#
#foreach {at1 at2 at3} $matm {s1 s2 s3} $mres {r1 r2 r3} $mseq {
#puts stdout "atom=$at1"
#      set last [string range $at1 end end]
#      set at1 [string trimright $at1 $last]
#      lappend ma $at1
#      lappend ms $s1
#      lappend mr $r1
#      }
#
#puts stdout "atom=[llength $ma] ms=[llength $ms] xav=[llength $xav] yav=[llength $yav] zav=[llength $zav]"
#puts stdout "$xav $yav $zav"
#--------------------------------------------------------
#calculating ..
set idout [open $outfile w]

#write some parameters into the noe file
puts $idout "! -- pdb file: $pdbfile"
puts $idout "! -- max. distance: $maxdis A"
puts $idout "! -- distance cutoff: $noecut A"
puts $idout "! -- noe substract: $noesub A"
puts $idout "! -- noe add: $noeadd A"
puts $idout "! -- methyl protons: $mlist"
puts $idout "! -- amino protons: $ilist"
puts $idout "!-----------------------------------------------"
#distance between regular atoms:
set count 0
set temp 0
#--------------------------------------------------------
#--------ILV methyl to IVL methyl proton NOEs
puts stdout "calculating methyl to methyl distances "
set mmcount 0
set mtotal 0
set mnum [llength $mres]
puts $idout "!
--------------
ILV methyl to ILV methyl distances < $noecut A"
for {set i 0} {$i <= [expr $mnum - 1]} {incr i} {
    set x1 [lindex $mxcor $i]
    set y1 [lindex $mycor $i]
    set z1 [lindex $mzcor $i]
    set m1 [lindex $mres $i]
    set m2 [lindex $mseq $i]
    set m3 [lindex $matm $i]
    #puts stdout "$m1 $m2 $m3 $x1 $y1 $z1"
    for {set j [expr $i+1]} {$j <= [expr $mnum - 1]} {incr j} {
        set x2 [lindex $mxcor $j]
        set y2 [lindex $mycor $j]
        set z2 [lindex $mzcor $j]
        set dis [expr sqrt(($x1 -$x2)*($x1 -$x2)+($y1 -$y2)*($y1 -$y2)+($z1 -$z2)*($z1 -$z2))]
        if {$dis < $noecut} {
            incr count
            incr mmcount
            incr mtotal
            set dis [expr $dis]
            if {{[lindex $mseq $i] != [lindex $mseq $j]}} {
                set nm1 [lindex $matm $i]
                set nm2 [lindex $matm $j]
                puts $idout "[lindex $mseq $i] [lindex $mres $i] [string range $nm1 0 2] to "
                puts $idout "[lindex $mseq $j] [lindex $mres $j] [string range $nm2 0 2] [string range $dis 0 3] "
                #puts $idout "[lindex $ms $i] [lindex $mr $i] [lindex $ma $i] to \n
                # [lindex $ms $j] [lindex $mr $j] [lindex $ma $j] [string range $dis 0 3] "
                # puts stdout "dist=[lindex $ma $i] [lindex $ma $j] $dis"
            }
        }
    }
    #puts stdout "found a total of $mmcount ILV methyl-IVL methyl distances < $noecut A"
    #puts stdout "found a total of $mnum ILV methyls"
    puts $idout "!--------- Total methyl to methyl NOEs: $mmcount"
#----------------------amino protons
#
#puts $idout "!---------------------------------------------------------------"
puts $idout "!------------- Total : $mmcount ILV methyl-methyl distance < $noecut"
#close $idout
#---------------------------------------------------------------
puts stdout "#--------------------------------------------------------------"
puts stdout "Check NOE restraint file: $outfile"
Appendix A2  XPLOR script for basic simulated annealing of BamC

# slow cooling protocol in torsion angle space BamC. Uses # NOE, RDC # # CDS 5/14/03 #

# this checks for typos on the command-line. User-customized arguments could # also be specified.
#
xplor.parseArguments()

outFilename = "SCRIPT_STRUCTURE.pdb"
numberOfStructures=100
seed=3421
simWorld.setRandomSeed(seed)  #set random seed

command = xplor.command

# # import necessary modules # # from xplorPot import XplorPot from rdcPotTools import create_RDCPot from varTensorTools import create_VarTensor import varTensorTools from ivm import IVM from potList import PotList import protocol

# # generate PSF data from sequence and initialize the correct parameters. # # from psfGen import seqToPSF seqToPSF( open("BamC.seq").read(), startResid=98)

# # generate a random extended structure with correct covalent geometry # protocol.genExtendedStructure("BamC_extended_%d.pdb" % seed)

#
# a PotList contains a list of potential terms. This is used to specify which
# terms are active during refinement.
#
potList = PotList()

# orientation Tensor - used with the dipolar coupling term
#
oTensor = create_VarTensor("oTensor")
oTensor.setDa(15.36)
oTensor.setRh(0.63)

# dipolar coupling restraints for protein amide NH.
#
rdc = create_RDCPot(name="rdc",
oTensor=oTensor,
file="BamC_pf1.tbl")
rdc.setThreshold(0.5)
print rdc.info()
potList.append(rdc)

# set up NOE potential
from noePotTools import create_NOEPot
noe = create_NOEPot("noe",file="BamC_NOE.tbl")
noe.setPotType( "soft" )
noe.setRSwitch( 0.5 )
noe.setAsympSlope( 1. )
noe.setSoftExp(1.)
noe.setThreshold(0.5)
print noe.info()
potList.append(noe)

# set up J coupling
#from jCoupPotTools import create_JCoupPot
#jCoup = create_JCoupPot("jcoup",file="jna_coup.tbl",
#                      A=6.98,B=-1.38,C=1.72,phase=-60)
#jCoup.setThreshold(1.0)
#print jCoup.info()
potList.append(jCoup)

# Set up dihedral angles
protocol.initDihedrals("BamC_dih.tbl",
        scale=5,  #initial force constant
        useDefaults=0)
potList.append( XplorPot('CDIH') )
# radius of gyration term - not used here.
#
#protocol.initCollapse("resid 4:20",
#   scale=25.0,
#   Rtarget=14.0)
#potList.append( XplorPot('COLL') )

#
# setup parameters for atom-atom repulsive term. (van der Waals-like term)
#
protocol.initNBond(nbxmod=4,  # Can use 4 here, due to IVM dynamic
   repel=0.5) # initial effective atom radius
potList.append( XplorPot('VDW') )

#
# annealing settings
#
init_t  = 3500.     # Need high temp and slow annealing to converge
final_t = 100
cool_steps = 15000  # slow annealing

#
# initial force constant settings
#
ini_rad  = 0.4
ini_con = 0.004
ini_ang = 0.4
ini_imp = 0.4  # was 0.1
ini_noe = 1   # was 150
ini_sani = 0.01

potList.add( XplorPot("BOND") )
potList.add( XplorPot("ANGL") )
potList.add( XplorPot("IMPR") )

#
# potential terms used for high-temp dynamics
#
hitemp_potList = PotList()
hitemp_potList.add( XplorPot("BOND") )
hitemp_potList.add( XplorPot("ANGL") )
hitemp_potList.add( XplorPot("IMPR") )
hitemp_potList.add( XplorPot("CDIH") )
hitemp_potList.add( noe   )
#hitemp_potList.add( jCoup )
hitemp_potList.add( rdc   )

# Give atoms uniform weights, except for the anisotropy axis
from atomAction import SetProperty
AtomSel("not resname ANI").apply( SetProperty("mass",100. ) )
AtomSel("resname ANI    ").apply( SetProperty("mass",300. ) )
AtomSel("all            ").apply( SetProperty("fric",10. ) )

#
# IVM setup
#

dyn = IVM()

dyn.potList().add( XplorPot("BOND") )
dyn.potList().add( XplorPot("ANGL") )
dyn.potList().add( XplorPot("IMPR") )

dyn.breakAllBondsIn("not resname ANI")
oTensor.setFreedom("fix")
varTensorTools.topologySetup(dyn,oTensor)

protocol.initMinimize(dyn,numSteps=1000)
dyn.run()

#
# reset ivm topology for torsion-angle dynamics
#
dyn.reset()
dyn.potList().removeAll()

oTensor.setFreedom("fixDa, fixRh")
varTensorTools.topologySetup(dyn,oTensor)
protocol.torsionTopology(dyn)
# minc used for final cartesian minimization
m = IVM()
protocol.initMinimize(min)
from selectTools import IVM_groupRigidSidechain
IVM_groupRigidSidechain(min)
m.breakAllBondsIn("not resname ANI")

oTensor.setFreedom("varyDa, varyRh")
varTensorTools.topologySetup(min, oTensor)

def setConstraints(k_ang, k_imp):
    command(""
    constraints
        interaction (not resname ANI) (not resname ANI)
        weights * 1 angl %f impr %f
    end
    end""" % (k_ang, k_imp))
    return

def structLoopAction(loopInfo):
    #
    # this function calculates a single structure.
    #
    # set some high-temp force constants
    #
    noe.setScale(20)  # use large scale factor initially
    rdc.setScale(ini_sani)
    command("parameters nbonds repel %f end end" % ini_rad)
    command("parameters nbonds rcon %f end end" % ini_con)
    setConstraints(ini_ang, ini_imp)
    # Initial weight--modified later
    command("restraints dihedral scale=5. end")

    #
    # high temp dynamics
    # note no Van der Waals term
    #

    init_t = 3500
    ini_timestep = 0.010
bath = init_t
timestep = ini_timestep

protocol.initDynamics(dyn,  
    potList=hitemp_potList, # potential terms to use  
    bathTemp=bath,  
    initVelocities=1,  
    finalTime=20,  
    printInterval=100)

dyn.setETolerance(bath/100)  # used to det. stepsize. default: bath/1000
dyn.run()

# increase dihedral term
command("restraints dihedral scale=200. end")

#
# cooling and ramping parameters
#

global k_ang, k_imp

# MultRamp ramps the scale factors over the specified range for the  
# simulated annealing.
from simulationTools import MultRamp
k_noe = MultRamp(ini_noe,30.,"noe.setScale( VALUE )")
k_rdc = MultRamp(ini_sani,1.."rdc.setScale( VALUE )")
radius = MultRamp(ini_rad,0.8,  
    "command('param nbonds repel VALUE end end')")
k_vdw = MultRamp(ini_con,4, "command('param nbonds rcon VALUE end end')")
k_ang = MultRamp(ini_ang,1.0)
k_imp = MultRamp(ini_imp,1.0)

protocol.initDynamics(dyn,  
    potList=potList,  
    bathTemp=bath,  
    initVelocities=1,  
    finalTime=2,  
    printInterval=100,  
    stepsize=timestep)

from simulationTools import AnnealIVM
AnnealIVM(initTemp =init_t,  
    finalTemp=100,  
    tempStep =25,  
    .ivm=dyn,
rampedParams = [k_noe, k_rdc, radius, k_vdw, k_ang, k_imp],
extrCommands = lambda notUsed:
    setConstraints(k_ang.value(),
    k_imp.value())).run()

# final torsion angle minimization
# protocol.initMinimize(dyn,
#     printInterval=50)
dyn.run()

# final all atom minimization
# protocol.initMinimize(minc,
#     potList=potList,
#     printInterval=100)
minc.run()

# analyze and write out structure
# loopInfo.writeStructure(potList)
pass

from simulationTools import StructureLoop
StructureLoop(numStructures=numberOfStructures,
pdbTemplate=outFilename,
structLoopAction=structLoopAction,
genViolationStats=1,
averagePotList=potList).run()
Appendix A3 Using XPLOR-NIH for domain orientation using residual dipolar couplings

In order to orient domains using RDC data, first collect IPAP or decoupled/trosy HSQC or whatever your favorite experiment is to measure J coupling on an isotropic and a weakly aligned, anisotropic sample. The splitting in the isotropic spectra will equal the $J_{NH}$, while the splitting in the anisotropic spectra is equal to the $J_{NH} + D_{NH}$. Don’t forget that the $J_{NH}$ is negative. The residual dipolar coupling is then the splitting in the anisotropic spectra minus the splitting in the isotropic spectra. An initial estimate of the magnitude and rhombicity of the alignment tensor ($D_a$ and $R$) can be done by creating a histogram plot of the RDCs to estimate the magnitude of the alignment tensor ($D_a$) and the rhombicity ($R$) and using Equations 2.2–2.4.

A more accurate estimation of $D_a$ and $R$ can be done by performing a grid search of $D_a$ and $R$. For this, you will need a coordinate file of your molecule with hydrogens added (pdb file) and the corresponding structure file (psf file) as input. Be sure that the axis.pdb does not interfere with your molecule of interest. Open them both in Pymol, move the axis far away from you molecule and save the coordinates. At the command line, run the following

```bash
> xplor loopDaR.inp > loopDaR.log
```

{*This NIH-XPLOR script performs a grid search of Da (aa) and Rh (ar). This script should be run and the output file searched for "SANI energy". Use values of Da and Rh that minimize the SANI energy. Written March 2010, Lisa Warner*}

! Read in structure and parameter files

```bash
structure reset @3Q6B_1H.psf
    @axis.psf end
parameter reset @protein-allhdg.param
```
{*define the initial, final, and stepsize (increment) for Aa,Ar loops*}

\texttt{evaluate ($aa\_init=13.0$) evaluate ($aa\_final=15$)}
\texttt{evaluate ($ar\_init=0.5$) evaluate ($ar\_final=0.65$)}
\texttt{evaluate ($aa\_incr=0.1$) evaluate ($ar\_incr=0.01$)}


\texttt{!***Fix the origin of the floating axis}
\texttt{constraints fix (resid 500 and name OO) end}  \texttt{!FIXES THE OO of axis}

\texttt{vector do (mass = 100.0) (not resid 500)} \texttt{! uniform mass for all atoms}
\texttt{vector do (fbeta = 10.0) (all)} \texttt{! coupling to heat bath}
\texttt{vector do (mass = 30.0) (resid 500)} \texttt{! for axis}

\texttt{eval ($init\_t=500.01$)}
\texttt{eval ($final\_t=100.01$)}
\texttt{eval ($cool\_steps=25000$)}
\texttt{eval ($tempstep=100$)}

\texttt{eval ($ncycle=($init\_t-$final\_t)/tempstep$)}
\texttt{eval ($nstep=\text{int}($cool\_steps/$ncycle$))}$

\texttt{eval ($ini\_timestep=0.010$)} \texttt{!reduced from 0.015}
\texttt{eval ($fin\_timestep=0.003$)}
\texttt{eval ($timestep\_fac=(fin\_timestep/ini\_timestep)^{1/ncycle}$)}

\texttt{evaluate ($ener\_min=1.0e+20$)}
\texttt{evaluate ($aa=aa\_init$)}

\texttt{coor @3Q6B_1H.pdb}
\texttt{coor @axis.pdb}

{*outer loop Aa*}
\texttt{evaluate ($aa=aa\_init$)}
\texttt{while ($aa <$ aa\_final) loop aa_loop}

{*inner loop Ar*}
\texttt{evaluate ($ar=ar\_init$)}
\texttt{while ($ar <$ ar\_final) loop ar_loop}

{*for each Ar,Ar*}
sani
reset
nres=4000                   ! > # of dp constraints in this group

class p4                   !designates the variable for the sani force p4 (my name)
potential square
@p4RDC_101811.tbl
coeff 0.0 $aa $ar
force 0

class p5                   !designates the variable for the sani force p4 (my name)        potential
square
@p5RDC_101811.tbl
coeff 0.0 $aa $ar
force 0

end

{*fitting the axis (resid 500) and potra 4 to the RDCs*}

sani class p4 force 2 end
!  sani class p5 force 2 end
  flags exclude * include sani end

minimize rigid
nstep=1000   drop=10.
group=(residue 262:421)
group=(residue 500:501)
nprint=4
end

eval ($bath = $init_t)
eval ($timestep = $ini_timestep)

if ($ener < $ener_min) then
  evaluate ($ener_min=$ener)
evaluate ($aa_min=$aa)
evaluate ($ar_min=$ar)
remark Current Aa $aa
remark Best Aa $aa_min
remark Current Ar $ar
remark Best Ar $ar_min
end if

{*exit ar_loop*}
evaluate ($ar=$ar+$ar_incr)
end loop ar_loop

{*exit aa_loop*}
evaluate ($aa=$aa+$aa_incr)
end loop aa_loop

remarks Best Aa=$aa_min Ar=$ar_min
stop

If you designate the out file as *.log, you can open it in the MacOSX program Console. Just double click on it in Finder and it will automatically open. Once the program has finished running, you can grep “AA_MIN and AR_MIN” from the .log file, where AA = Da, AR = R. Run it first with a coarse grained search of $D_a$ and $R$, then do a fine-grained search. As the program is written, you must run it 2 times, once for each domain (comment out by adding “!” to sani force definition of each name).

Next create pdb files with structures where the orientation of the domains is randomized using the script below, changing files that are in bold font:

! This xplor file creates a series of pdbs with a 3 resid linker
! region randomized using high tempurature torsion angle
! dynamics. 3/31/30 lrw

structure reset @3Q6B_1H.psf
   @axis.psf end

parameter reset @protein-allhdg.param
   @para_axis.pro end

!***Fix the origin of the floating axis
constraints fix (resid 500 and name OO) end  !FIXES THE OO of axis

vector do (mass = 50.0) (not resid 500)  ! uniform mass for all atoms
    vector do (fbeta = 5.0) (all)  ! coupling to heat bath
    vector do (mass = 30.0) (resid 500)  ! for axis
evaluate ($count = 0)
evaluate ($end_count=100)  \{*Loop through a family of structures.*\}
while ($count < $end_count ) loop main
  evaluate ($count=$count+1)
  coor @3Q6B_1H.pdb
  coor @axis.pdb

\{*generating random orientations*\}

eval ($randomSeed = 21)  !random seed
  eval ($seed = $randomSeed+$count)

!  ! Be sure different structures start with different seeds
!  set seed $seed end

\{====> Define input/output pdb file names\}

!
! annealing settings
!

eval ($high_t = 6500.01)

!  ! setting up the groups
  dynamics internal
    group (resid 266:343 )
    group (resid 347:420 )
    cloop=false
    auto torsion
    maxe 10000
end

!  ! setting high temp parameters

eval ($bath = $high_t)

flags
  !exclude bonds * include vdw angl impr
  exclude * include vdw bonds impr angl
end

  vector do (vx = maxwell($bath)) (all)
  vector do (vy = maxwell($bath)) (all)
vector do (vz = maxwell($bath)) (all)
{ * Set initial velocities to fit a temperature of 3500K. *}
{ * High temperature to promote convergence.    *}

!  ! high temp dynamics
!

evaluate ($tol = $bath/1000)  
dynamics internal  
nstep 50000  
tbath $bath  
nprint 10000  
etol $tol  
end

remarks This file was made with the xplor script  
remarks create_random.inp. This pdb contains coordinates  
remarks for POTRA45 in which the torsion angles for residues 344-348  
remarks were allowed to randomly orient, resulting in a structure with  
remarks domains 4 and 5 randomly oriented with respect to one another.

evaluate ($filename="random_"+encode($count)+".pdb")  
   write coordinates output =$filename end  
end loop main  
stop  
~

Finally, orient the domains using the following script, once again changing files that are  
in bold font:

structure reset @pot45_all.psf  
   @axis.psf end  
parameter reset @protein-allhdg.param  
   @para_axis.pro end

noe  
{=====}  
nres=1000  
class all  
{=====}  
   @noe1.noe
end
noe
  ceiling=1000
  averaging * cent
  potential * square
  sqconstant * 1.
  sqexponent * 2
  scale * 50.          {*Constant NOE scale throughout the protocol.*}
end

{*define the initial and final constants for SANI*}
evaluate ($ini_sani=  0.1)       evaluate ($fin_sani=  2)

!
!Setting up the RDC restraint tables
!
sani
  reset
  nres=4000                   ! > # of dp constraints in this group
  class  p4                 !Not sure what this does but we should check
  potential square
      @p4RDCcalc_030310.tbl
      coeff 0.0 -12.5 0.550    {*Da and Rh for the input molecule: calculated first in an outside
                               program...*}
      force 2
  class p45          !Not sure what this does but we should check
      potential square
      @p45RDCcalc_030310.tbl
      coeff 0.0 -12.5 0.550    {*Da and Rh for the input molecule: calculated first in an outside
                               program...*}
      force 0
end

!***Fix the origin of the floating axis
constraints fix (resid 500 and name OO) end       !FIXES THE OO of axis

  vector do (mass  = 50.0) (not resid 500)       ! uniform mass for all atoms
  vector do (fbeta = 10.0) (all)                 ! coupling to heat bath
vector do (mass = 30.0) (resid 500) ! for axis

evaluate ($count = 0)
evaluate ($end_count=100)  {*Loop through a family of structures.*}
while ($count < $end_count ) loop main
  evaluate ($count=$count+1)
  coor @@$filename="random_"+encode($count)+".pdb"

  coor @@axis.pdb

  coor @@filename

  eval ($init_t = 1500.01)
  eval ($final_t = 100.01)
  eval ($cool_steps = 25000)
  eval ($tempstep = 100)
  eval ($ncycle = ($init_t-$final_t)/$tempstep)
  eval ($nstep = int($cool_steps/$ncycle))
  eval ($ini_timestep = 0.010)  !reduced from 0.015
  eval ($fin_timestep = 0.003)
  eval ($timestep_fac = ($fin_timestep/$ini_timestep)^(1/$ncycle))

  {*fitting the axis (resid 500) and potra 4 to the RDCs*}

  sani class p4 force 2 end
  sani class p45 force 0 end
  flags exclude *

  minimize rigid
  nstep=500  drop=10.
  group=(residue 262:421 and segid A)
  group=(residue 500:501)
  nprint=1
  end

  eval ($bath = $init_t)
  eval ($timestep = $ini_timestep)

  {*fitting potra 5 and fixing potra4 and axis *}

! setting up the groups
dynamics internal
   group ((resid 262:343 and segid A) or resid 500:501 )
   group (resid 347:421 and segid A )
   cloop=false
   auto torsion
   maxe 10000
end

evaluate ($ksani=$ini_sani)
evaluate ($i_cool = 0)
while ($i_cool < 13) loop cool
   evaluate ($i_cool=$i_cool+1)
   evaluate ($bath=$bath-100)

sani class p4 force 0 end
sani class p45 force $ksani end

flags exclude * include bond angle impr vdw sani noe end

vector do (vx = maxwell($bath)) (all)
vector do (vy = maxwell($bath)) (all)
vector do (vz = maxwell($bath)) (all)
{* Set initial velocities to fit a temperature of 1500K. *}

!
! high temp dynamics
!

evaluate ($tol = $bath/1000)
dynamics internal
   nstep 0
   endt 15
   timestep $timestep
   tbath $bath
   response 20
   nprint 1000
   etol $tol
evaluate ($ksani=$ksani + 0.15)
end loop cool

!************************************************************************************************
print thres=0.05 bonds    {*Print deviations from ideal geometry.*}
evaluate ($rms_bonds=$result)
print thres=5. angles
evaluate ($rms_angles=$result)
print thres=5. impropers
evaluate ($rms_impropers=$result)
sani print threshold = 0.5 end
evaluate($rms_sani=$result)
evaluate($viol_sani=$violations)
end

remarks This file was generated from an input randomized pdb (from
remarks create_rand.inp) and has been oriented with RDCs and
remarks the fitRDCs.inp xplor program.
remarks ==============================================================
remarks overall,bonds,angles,impr,vdw, sani, noe
remarks energies: $ener, $bond, $angl, $impr, $vdw, $sani, $noe
remarks ==============================================================
remarks bonds, angles, impropers, sani,
remarks rms-d: $rms_bonds,$rms_angles,$rms_impropers, $rms_sani
remarks ==============================================================
remarks sani(0.0)
remarks violations.: $viol_sani
remarks

evaluate ($filename="fitallRDCaddnoe_"+encode($count)+"_041210.pdb")
write coordinates output =$filename end
end loop main

stop
~
~
Figure A4.1 Backbone connectivities for Met263–Ala282.

The Cα and Cβ resonances for the i and i-1 residues are observed in the HNCACB spectrum (positive contours, blue, correspond to Cα resonances and negative contours, red, to Cβ resonances). The Cα and Cβ resonances for the i-1 residues are observed in the CBCA(CO)NH spectrum (positive contours are green and negative contours are purple, all positive contoured peaks are Cα and Cβ resonances).
Figure A4.2 Backbone connectivities for Val282–Thr302.

The $C^\alpha$ and $C^\beta$ resonances for $i$ and $i-1$ residues are observed in the HNCACB spectrum (positive contours, blue, correspond to $C^\alpha$ resonances and negative contours, red, to $C^\beta$ resonances). The $C^\alpha$ and $C^\beta$ resonances for the $i-1$ residues are observed in the CBCA(CO)NH spectrum (positive contours are green and negative contours are purple, all positive contoured peaks are $C^\alpha$ and $C^\beta$ resonances). A break in the backbone connectivity occurs between Glu291 and Gly293. The $C^\alpha$ and $C^\beta$ for Pro292 are observed in the $i-1$ peaks at the Gly 293 $^1$H, $^{15}$N chemical shift.
Figure A4.3 Backbone connectivities for Thr302–Val322.

The C\textsuperscript{\alpha} and C\textsuperscript{\beta} resonances for \textit{i} and \textit{i}-1 residues are observed in the HNCACB spectrum (positive contours, blue, correspond to C\textsuperscript{\alpha} resonances and negative contours, red, to C\textsuperscript{\beta} resonances). The C\textsuperscript{\alpha} and C\textsuperscript{\beta} resonances for the \textit{i}-1 residues are observed in the CBCA(CO)NH spectrum (positive contours are green and negative contours are purple, all positive contoured peaks are C\textsuperscript{\alpha} and C\textsuperscript{\beta} resonances). A break in the backbone connectivity occurs between Tyr319 and Arg321. The C\textsuperscript{\alpha} and C\textsuperscript{\beta} for Pro320 are observed in the \textit{i}-1 peaks at the Arg321 \textsuperscript{1}H, \textsuperscript{15}N chemical shift.
Figure A4.4 Backbone connectivities for Val322–Asp342.

The $C^\alpha$ and $C^\beta$ resonances for $i$ and $i-1$ residues are observed in the HNCACB spectrum (positive contours, blue, correspond to $C^\alpha$ resonances and negative contours, red, to $C^\beta$ resonances). The $C^\alpha$ and $C^\beta$ resonances for the $i-1$ residues are observed in the CBCA(CO)NH spectrum (positive contours are green and negative contours are purple, all positive contoured peaks are $C^\alpha$ and $C^\beta$ resonances). A break in the backbone connectivity occurs between Met325 and Glu327. The $C^\alpha$ and $C^\beta$ for Pro326 are observed in the $i-1$ peaks at the Glue327 $^1$H, $^{15}$N chemical shift.
Figure A4.5 Backbone connectivities for Lys333–Leu365.

The $\mathrm{C}^a$ and $\mathrm{C}^\beta$ resonances for $i$ and $i$-1 residues are observed in the HNCACB spectrum (positive contours, blue, correspond to $\mathrm{C}^a$ resonances and negative contours, red, to $\mathrm{C}^\beta$ resonances). The $\mathrm{C}^a$ and $\mathrm{C}^\beta$ resonances for the $i$-1 residues are observed in the CBCA(CO)NH spectrum (positive contours are green and negative contours are purple, all positive contoured peaks are $\mathrm{C}^a$ and $\mathrm{C}^\beta$ resonances). Peaks from Gly344 were folded spectroscopically causing the peak phases to opposite, where $\mathrm{C}^a$ and $\mathrm{C}^\beta$ resonances are blue and red in the HNCACB and $\mathrm{C}^a$ and $\mathrm{C}^\beta$ resonances are purple in the CBCA(CO)NH. A break occurs in the backbone connectivity between Asn357 and Ala363. Ser360 was assigned based on the chemical shift of the $i$ and $i$-1 resonances, which correspond to a ThrSer dipeptide. There is only one occurrence of this in the POTRA4–5 sequence and was therefore attributed to Ser360. No corresponding resonances were observed for Thr359.
Figure A4.6 Backbone connectivities for Leu365–Leu392.

The C\(^\alpha\) and C\(^\beta\) resonances for the \(i\) and \(i-1\) residues are observed in the HNCACB spectrum (positive contours, blue, correspond to C\(^\alpha\) resonances and negative contours, red, to C\(^\beta\) resonances). The C\(^\alpha\) and C\(^\beta\) resonances for the \(i-1\) residues are observed in the CBCA(CO)NH spectrum (positive contours are green and negative contours are purple, all positive contoured peaks are C\(^\alpha\) and C\(^\beta\) resonances). Two breaks in the backbone connectivity occur between Arg366–Gly374 and Leu377–Asp380 are bracketed on either side with a concomitant decrease in peak intensity, suggesting that these regions are undergoing conformational exchange broadening.
Figure A4.7 Backbone connectivities for Leu392–Val413.

The C\textsuperscript{a} and C\textsuperscript{\beta} resonances for the i and i-1 residues are observed in the HNCACB spectrum (positive contours, blue, correspond to C\textsuperscript{a} resonances and negative contours, red, to C\textsuperscript{\beta} resonances). The C\textsuperscript{a} and C\textsuperscript{\beta} resonances for the i-1 residues are observed in the CBCA(CO)NH spectrum (positive contours are green and negative contours are purple, all positive contoured peaks are C\textsuperscript{a} and C\textsuperscript{\beta} resonances). A break in the backbone connectivity occurs between Val405 and Gly407. The C\textsuperscript{a} and C\textsuperscript{\beta} for Pro406 are observed in the i-1 peaks at the Gly407 \textsuperscript{1}H, \textsuperscript{15}N chemical shift. Ser408 was assigned based on the i-1 residue being a proline, which is the only occurrence of this in the sequence.
Figure A4.8 Backbone connectivities for Asp413–Gly426.

The $\alpha$ and $\beta$ resonances for the $i$ and $i$-1 residues are observed in the HNCACB spectrum (positive contours, blue, correspond to $\alpha$ resonances and negative contours, red, to $\beta$ resonances). The $\alpha$ and $\beta$ resonances for the $i$-1 residues are observed in the CBCA(CO)NH spectrum (positive contours are green and negative contours are purple, all positive contoured peaks are $\alpha$ and $\beta$ resonances). A break in the backbone connectivity occurs between Gly424 and Gly426. The $\alpha$ and $\beta$ for Pro425 are observed in the $i$-1 peaks at the Gly426 $^1$H, $^{15}$N chemical shift.
Appendix A5  Backbone connectivities for POTRA3–5

Figure A5.1 Backbone connectivities for Met173–Asp 191.
The Cα and Cβ resonances for the i and i-1 residues are observed in the HNCACB spectrum (positive contours, blue, correspond to Cα resonances and negative contours, red, to Cβ resonances). The Cα and Cβ resonances for the i-1 residues are observed in the HN(CO)CA and HN(COCA)CB spectra (positive contours are green and negative contours are purple, all positive contoured peaks are Cα and Cβ resonances). A break is observed between Glu176 and Asn181.
The C^\alpha and C^\beta resonances for the \(i\) and \(i-1\) residues are observed in the HNCACB spectrum (positive contours, blue, correspond to C^\alpha resonances and negative contours, red, to C^\beta resonances). The C^\alpha and C^\beta resonances for the \(i-1\) residues are observed in the HN(CO)CA and HN(COCA)CB spectra (positive contours are green and negative contours are purple, all positive contoured peaks are C^\alpha and C^\beta resonances).
Figure A5.3 Backbone connectivities for Arg237–Lys267.

The Cα and Cβ resonances for the i and i-1 residues are observed in the HNCACB spectrum (positive contours, blue, correspond to Cα resonances and negative contours, red, to Cβ resonances). The Cα and Cβ resonances for the i-1 residues are observed in the HN(CO)CA and HN(COCA)CB spectra (positive contours are green and negative contours are purple, all positive contoured peaks are Cα and Cβ resonances).
**Figure A5.4 Backbone connectivities for Lys267–Ser281.**

The C$^\alpha$ and C$^\beta$ resonances for the $i$ and $i-1$ residues are observed in the HNCACB spectrum (positive contours, blue, correspond to C$^\alpha$ resonances and negative contours, red, to C$^\beta$ resonances). The C$^\alpha$ and C$^\beta$ resonances for the $i-1$ residues are observed in the HN(CO)CA and HN(COCA)CB spectra (positive contours are green and negative contours are purple, all positive contoured peaks are C$^\alpha$ and C$^\beta$ resonances).
The Cα and Cβ resonances for the i and i-1 residues are observed in the HNCACB spectrum (positive contours, blue, correspond to Cα resonances and negative contours, red, to Cβ resonances). The Cα and Cβ resonances for the i-1 residues are observed in the HN(CO)CA and HN(COCA)CB spectra (positive contours are green and negative contours are purple, all positive contoured peaks are Cα and Cβ resonances).
**Figure A5.6 Backbone connectivities for Tyr296–Leu311.**

The C$^\alpha$ and C$^\beta$ resonances for the $i$ and $i-1$ residues are observed in the HNCACB spectrum (positive contours, blue, correspond to C$^\alpha$ resonances and negative contours, red, to C$^\beta$ resonances). The C$^\alpha$ and C$^\beta$ resonances for the $i-1$ residues are observed in the HN(CO)CA and HN(COCA)CB spectra (positive contours are green and negative contours are purple, all positive contoured peaks are C$^\alpha$ and C$^\beta$ resonances).
Figure A5.7 Backbone connectivities for Leu311–Ile328.

The $C^\alpha$ and $C^\beta$ resonances for the $i$ and $i-1$ residues are observed in the HNCACB spectrum (positive contours, blue, correspond to $C^\alpha$ resonances and negative contours, red, to $C^\beta$ resonances). The $C^\alpha$ and $C^\beta$ resonances for the $i-1$ residues are observed in the HN(CO)CA and HN(COCA)CB spectra (positive contours are green and negative contours are purple, all positive contoured peaks are $C^\alpha$ and $C^\beta$ resonances).
**Figure A5.8 Backbone connectivities for Ile328–Asp342.**

The $\mathrm{C}^\alpha$ and $\mathrm{C}^\beta$ resonances for the $i$ and $i-1$ residues are observed in the HNCACB spectrum (positive contours, blue, correspond to $\mathrm{C}^\alpha$ resonances and negative contours, red, to $\mathrm{C}^\beta$ resonances). The $\mathrm{C}^\alpha$ and $\mathrm{C}^\beta$ resonances for the $i-1$ residues are observed in the HN(CO)CA and HN(COCA)CB spectra (positive contours are green and negative contours are purple, all positive contoured peaks are $\mathrm{C}^\alpha$ and $\mathrm{C}^\beta$ resonances).
Figure A5.9 Backbone connectivities for Asp342–Gly356.

The $\alpha$ and $\beta$ resonances for the $i$ and $i-1$ residues are observed in the HNCACB spectrum (positive contours, blue, correspond to $\alpha$ resonances and negative contours, red, to $\beta$ resonances). The $\alpha$ and $\beta$ resonances for the $i-1$ residues are observed in the HN(CO)CA and HN(COCA)CB spectra (positive contours are green and negative contours are purple, all positive contoured peaks are $\alpha$ and $\beta$ resonances).
Figure A5.10 Backbone connectivities for Gly356–Glu387.
The C$^\alpha$ and C$^\beta$ resonances for the $i$ and $i-1$ residues are observed in the HNCA(CB) spectrum (positive contours, blue, correspond to C$^\alpha$ resonances and negative contours, red, to C$^\beta$ resonances). The C$^\alpha$ and C$^\beta$ resonances for the $i-1$ residues are observed in the HN(CO)CA and HN(COCA)CB spectra (positive contours are green and negative contours are purple, all positive contoured peaks are C$^\alpha$ and C$^\beta$ resonances).
**Figure A5.11 Backbone connectivities for Glu387–Asp401.**

The C$^\alpha$ and C$^\beta$ resonances for the $i$ and $i-1$ residues are observed in the HNCACB spectrum (positive contours, blue, correspond to C$^\alpha$ resonances and negative contours, red, to C$^\beta$ resonances). The C$^\alpha$ and C$^\beta$ resonances for the $i-1$ residues are observed in the HN(CO)CA and HN(COCA)CB spectra (positive contours are green and negative contours are purple, all positive contoured peaks are C$^\alpha$ and C$^\beta$ resonances).
The $\mathrm{C}^{\alpha}$ and $\mathrm{C}^{\beta}$ resonances for the $i$ and $i-1$ residues are observed in the HNCACB spectrum (positive contours, blue, correspond to $\mathrm{C}^{\alpha}$ resonances and negative contours, red, to $\mathrm{C}^{\beta}$ resonances). The $\mathrm{C}^{\alpha}$ and $\mathrm{C}^{\beta}$ resonances for the $i-1$ residues are observed in the HN(CO)CA and HN(COCA)CB spectra (positive contours are green and negative contours are purple, all positive contoured peaks are $\mathrm{C}^{\alpha}$ and $\mathrm{C}^{\beta}$ resonances).
Figure A5.13 Backbone connectivities for Lys417–Gly426.
The $\alpha$ and $\beta$ resonances for the $i$ and $i-1$ residues are observed in the HNCACB spectrum (positive contours, blue, correspond to $\alpha$ resonances and negative contours, red, to $\beta$ resonances). The $\alpha$ and $\beta$ resonances for the $i-1$ residues are observed in the HN(CO)CA and HN(COCA)CB spectra (positive contours are green and negative contours are purple, all positive
## Appendix A6 Filenames and parameters for NMR experiments for POTRA4–5

### Table A6.1 Filenames and parameters for backbone experiments for BamA POTRA4–5.

<table>
<thead>
<tr>
<th>Filename</th>
<th>B&lt;sub&gt;0&lt;/sub&gt;</th>
<th>sw</th>
<th>sw1</th>
<th>ni</th>
<th>ni2</th>
<th>np</th>
<th>nt</th>
<th>d1</th>
</tr>
</thead>
<tbody>
<tr>
<td>cbcacnh_f1f2f3_13c15nP45_30c_113109.fid</td>
<td>800</td>
<td>11261</td>
<td>14000</td>
<td>2950</td>
<td>64</td>
<td>61</td>
<td>2048</td>
<td>16</td>
</tr>
<tr>
<td>hncacb_f1f2f3_13c15nP45_30c_112209.fid</td>
<td>800</td>
<td>11261</td>
<td>14000</td>
<td>2950</td>
<td>53</td>
<td>44</td>
<td>2048</td>
<td>16</td>
</tr>
<tr>
<td>gNhsqc_13c15nP45_30c_8h_112009.fid</td>
<td>800</td>
<td>11261</td>
<td>2950</td>
<td>N/A</td>
<td>128</td>
<td>N/A</td>
<td>2048</td>
<td>64</td>
</tr>
</tbody>
</table>

B<sub>0</sub> – Static magnetic field strength in MHz
ni – number of complex points in t<sub>1</sub>
sw – acquisition sweep width (Hz)
sw1 – sweep width in t<sub>1</sub> (Hz)
sw2 – sweep width in t<sub>2</sub> (Hz)
nt – number of scans per FID
d1 – Interscan delay

### Table A6.2 Filenames and parameters for experiments used to measure ¹D<sub>NH</sub> on POTRA4–5.

<table>
<thead>
<tr>
<th>Filename</th>
<th>Field</th>
<th>sw</th>
<th>sw1</th>
<th>ni</th>
<th>np</th>
<th>nt</th>
<th>d1</th>
</tr>
</thead>
<tbody>
<tr>
<td>pf1_15N_YaeTP45250mMNaCl_IPAP13h_012610.fid</td>
<td>600</td>
<td>11990</td>
<td>2500</td>
<td>200</td>
<td>3300</td>
<td>32</td>
<td>1.7</td>
</tr>
<tr>
<td>15N_YaeTP45_15Nipap_13hr_250mMNacl012710.fid</td>
<td>600</td>
<td>11990</td>
<td>2500</td>
<td>200</td>
<td>3300</td>
<td>32</td>
<td>1.7</td>
</tr>
</tbody>
</table>

B<sub>0</sub> – Static magnetic field strength in MHz
ni – number of complex points in t<sub>1</sub>
sw – acquisition sweep width (Hz)
sw1 – sweep width in t<sub>1</sub> (Hz)
nt – number of scans per FID
d1 – Interscan delay
## Appendix A7  Filenames and parameters for NMR experiments for POTRA3–5 and POTRA1–5

### Table A7.1 Filenames and parameters for backbone experiments for BamA POTRA3–5.

<table>
<thead>
<tr>
<th>Filename</th>
<th>$B_0$</th>
<th>sw</th>
<th>sw1</th>
<th>sw2</th>
<th>ni</th>
<th>ni2</th>
<th>np</th>
<th>nt</th>
<th>d1</th>
</tr>
</thead>
<tbody>
<tr>
<td>HNcoCACB_F1F2F3_2h13c15nBamAP35_052002.fid</td>
<td>800</td>
<td>11261</td>
<td>13267</td>
<td>2845</td>
<td>80</td>
<td>44</td>
<td>2048</td>
<td>8</td>
<td>1.7</td>
</tr>
<tr>
<td>HNcoCAF1F2F3_2h13c15nP35_051510.fid</td>
<td>800</td>
<td>11261</td>
<td>6031</td>
<td>2900</td>
<td>46</td>
<td>44</td>
<td>2048</td>
<td>16</td>
<td>1.5</td>
</tr>
<tr>
<td>HSQC_TROSY_30C-041910.fid</td>
<td>900</td>
<td>14535</td>
<td>3200</td>
<td>N/A</td>
<td>128</td>
<td>N/A</td>
<td>2048</td>
<td>8</td>
<td>1.7</td>
</tr>
<tr>
<td>HNCACB_TROSY_30C-042310.fid</td>
<td>900</td>
<td>14535</td>
<td>3200</td>
<td>N/A</td>
<td>128</td>
<td>N/A</td>
<td>2048</td>
<td>8</td>
<td>1.7</td>
</tr>
</tbody>
</table>

$B_0$ – Static magnetic field strength in MHz  
np – number of complex points in $t_2$ (or $t_3$)  
ni – number of complex points in $t_1$  
sw – acquisition sweep width (Hz)  
sw1 – sweep width in $t_1$ (Hz)  
sw2 – sweep width in $t_2$ (Hz)  
nt – number of scans per FID  
d1 – Interscan delay

### Table A7.2 Filenames and parameters for experiments used to measure $^1$D$_{NH}$ on POTRA1–5.

<table>
<thead>
<tr>
<th>Filename</th>
<th>Field</th>
<th>sw</th>
<th>sw1</th>
<th>ni</th>
<th>np</th>
<th>nt</th>
<th>d1</th>
</tr>
</thead>
<tbody>
<tr>
<td>gNhsqc_decoup_BamAP15_10mgmlpf1_100711.fid</td>
<td>800</td>
<td>14045</td>
<td>3000</td>
<td>450</td>
<td>2048</td>
<td>16</td>
<td>1.5</td>
</tr>
<tr>
<td>gNhsqc_trosy_BamAP15_10mgmlpf1_100711.fid</td>
<td>800</td>
<td>14045</td>
<td>3000</td>
<td>450</td>
<td>2048</td>
<td>16</td>
<td>1.5</td>
</tr>
<tr>
<td>gNhsqc_decoup_BamAP15_100811.fid</td>
<td>800</td>
<td>14045</td>
<td>3000</td>
<td>450</td>
<td>2048</td>
<td>16</td>
<td>1.5</td>
</tr>
<tr>
<td>gNhsqc_trosy_BamAP15_100811.fid</td>
<td>800</td>
<td>14045</td>
<td>3000</td>
<td>450</td>
<td>2048</td>
<td>16</td>
<td>1.5</td>
</tr>
<tr>
<td>artsy_y_n_BamAP15_10mgmlPf1_111411.fid</td>
<td>800</td>
<td>12019</td>
<td>3000</td>
<td>128</td>
<td>2048</td>
<td>144</td>
<td>1.6</td>
</tr>
</tbody>
</table>

$B_0$ – Static magnetic field strength in MHz  
np – number of complex points in $t_2$ (or $t_3$)  
ni – number of complex points in $t_1$  
sw – acquisition sweep width (Hz)  
sw1 – sweep width in $t_1$ (Hz)  
nt – number of scans per FID  
d1 – Interscan delay
Table A7.3 Filenames and parameters for HSQC spectra collected on POTRA constructs.

<table>
<thead>
<tr>
<th>Filename</th>
<th>$B_0$</th>
<th>sw</th>
<th>np</th>
<th>sw1</th>
<th>ni</th>
<th>nt</th>
<th>d1</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>gNhsqc_2h15nP15_S143C_AcMTSL_0604117hour.fid</td>
<td>800</td>
<td>12019</td>
<td>2048</td>
<td>3000</td>
<td>120</td>
<td>56</td>
<td>1.7</td>
<td>30</td>
</tr>
<tr>
<td>gNhsqc_2h15nP15_S143C_MTLS_0604117hour.fid</td>
<td>800</td>
<td>12019</td>
<td>2048</td>
<td>3000</td>
<td>120</td>
<td>56</td>
<td>1.7</td>
<td>30</td>
</tr>
<tr>
<td>gNhsqc_2H13C15N_YaeTP12_decop_062510.fid</td>
<td>600</td>
<td>11990</td>
<td>2048</td>
<td>1944</td>
<td>90</td>
<td>8</td>
<td>1.7</td>
<td>30</td>
</tr>
<tr>
<td>gNhsqc_yaetp15_062009.fid</td>
<td>800</td>
<td>11261</td>
<td>2048</td>
<td>2950</td>
<td>80</td>
<td>48</td>
<td>3.0</td>
<td>30</td>
</tr>
<tr>
<td>15N_YateTP15gNhsq_35c_062009.fid</td>
<td>800</td>
<td>11261</td>
<td>2048</td>
<td>2950</td>
<td>144</td>
<td>128</td>
<td>1.5</td>
<td>35</td>
</tr>
<tr>
<td>15N_YateTP15gNhsq_40c_103009.fid</td>
<td>800</td>
<td>11261</td>
<td>2048</td>
<td>2950</td>
<td>144</td>
<td>128</td>
<td>1.5</td>
<td>40</td>
</tr>
</tbody>
</table>

$B_0$ – Static magnetic field strength in MHz  
np – number of complex points in $t_2$ (or $t_3$)  
ni – number of complex points in $t_1$  
sw – acquisition sweep width (Hz)  
sw1 – sweep width in $t_1$ (Hz)  
nt – number of scans per FID  
d1 – Interscan delay  
T – temperature (°C)