



Structure of the BamC Two-Domain Protein Obtained by Rosetta with a Limited NMR Data Set

Lisa R. Warner^{1†}, Krisztina Varga^{1†}, Oliver F. Lange², Susan L. Baker¹, David Baker^{2,3}, Marcelo C. Sousa¹ and Arthur Pardi^{1*}

¹Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA

²Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

³Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA

Received 12 March 2011;
received in revised form
13 May 2011;
accepted 16 May 2011
Available online
23 May 2011

Edited by M. F. Summers

Keywords:

β-barrel assembly machine;
helix–grip motif;
NMR structure;
Rosetta;
outer membrane protein

The CS-RDC-NOE Rosetta program was used to generate the solution structure of a 27-kDa fragment of the *Escherichia coli* BamC protein from a limited set of NMR data. The BamC protein is a component of the essential five-protein β-barrel assembly machine in *E. coli*. The first 100 residues in BamC were disordered in solution. The Rosetta calculations showed that BamC_{101–344} forms two well-defined domains connected by an ~18-residue linker, where the relative orientation of the domains was not defined. Both domains adopt a helix–grip fold previously observed in the Bet v 1 superfamily. ¹⁵N relaxation data indicated a high degree of conformational flexibility for the linker connecting the N-terminal domain and the C-terminal domain in BamC. The results here show that CS-RDC-NOE Rosetta is robust and has a high tolerance for misassigned nuclear Overhauser effect restraints, greatly simplifying NMR structure determinations.

© 2011 Elsevier Ltd. All rights reserved.

Introduction

The traditional method for determining the NMR solution structure of a protein involves three steps: assignment of ¹H, ¹³C, and ¹⁵N resonances; measurement of structural information such as ¹H–¹H

nuclear Overhauser effect (NOE) distance data and residual dipolar coupling (RDC) orientational data; and use of computational methods to generate an ensemble of conformations consistent with NMR-derived structural restraints.¹ 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 (such as transverse relaxation-optimized spectroscopy (TROSY), cross-correlated relaxation-induced polarization transfer (CRIPT), and cross-correlated relaxation-enhanced polarization transfer (CRINEPT)), or specific labeling [such as ¹³C labeling of ILV (Ile, Leu, Val) methyl groups] are employed to improve spectral resolution and sensitivity.^{2,3} A major challenge in the structure determination of larger proteins is the time-consuming resonance assignment of NOEs involving side-chain protons. Thus, various procedures are being developed to assign NOEs automatically as part of the structure refinement process.^{4–8}

*Corresponding author. E-mail address:
arthur.pardi@colorado.edu.

† L.R.W. and K.V. contributed equally to this work.

Present address: K. Varga, Department of Chemistry, University of Wyoming, Laramie, WY 82071, USA.

Abbreviations used: NOE, nuclear Overhauser effect; RDC, residual dipolar coupling; BAM, β-barrel assembly machine; H/D, hydrogen/deuterium; 3D, three-dimensional; HMQC, heteronuclear multiple-quantum coherence; NOESY, nuclear Overhauser enhancement spectroscopy; AMPK, AMP-activated protein kinase; TEV, tobacco etch virus; PDB, Protein Data Bank; NIH, National Institutes of Health.

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.^{9,10} Rosetta was originally developed as a *de novo* method for generating protein structures from the sequence.¹¹ 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.^{12,13} For *de novo* structure determinations, Rosetta starts by assembling fragments derived from a database of known protein structures using a low-resolution energy function.¹¹ 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 to help guide fragment selection and conformational searching, enabling atomic-resolution structure determination for small proteins.^{9,14} However, for proteins larger than ~12 kDa, chemical shifts may not provide enough information to guide the conformational search.¹⁰ Thus, additional experimental data such as RDCs and NOEs have been incorporated into iterative CS-RDC-NOE Rosetta and have been shown to increase its ability to generate accurate protein structures.¹⁰ For proteins above ~20 kDa, a significant fraction of residues did not converge to well-defined conformations using only a sparse set of amide ¹H-¹H NOEs (a total of 21 to 52 NOEs) in CS-RDC-NOE Rosetta.¹⁰ Thus, improvements in conformational sampling or additional experimental data are required to generate models for even larger proteins using Rosetta.

This iterative CS-RDC-NOE Rosetta program was used here to generate structures of the 27-kDa *Escherichia coli* periplasmic protein BamC by including chemical shift data and a limited set of NOE restraints (149 NOEs). The lipoprotein BamC is a component of the essential five-protein β -barrel assembly machine (BAM).¹⁵ This complex is involved in the folding and insertion of β -barrel proteins into the outer membranes of Gram-negative bacteria. Although *bamC*-null *E. coli* strains are viable, the mutants display outer membrane permeability defects and reduced levels of β -barrels in the outer membrane.^{16,17} The CS-RDC-NOE Rosetta calculations show that BamC consists of two helix-grip-type domains connected by an ~18-amino-acid linker. Additional backbone-backbone and methyl-methyl NOE data not used in the structure calculations validated the BamC structures generated by Rosetta. A set of ¹⁵N NMR relaxation data was collected for BamC and demonstrated a high degree of conformational dynamics for the backbone in the linker region.

Analysis of the ¹⁵N relaxation data and ¹H-¹⁵N 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.

Results and Discussion

The first 100 N-terminal residues are not structured in BamC

The first 24 residues of *E. coli* BamC contain a periplasmic localization signal cleaved *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. The previously reported chemical shifts for BamC₂₆₋₃₄₄ indicated that the next ~75 N-terminal residues had no regular secondary structure.¹⁸ To confirm this hypothesis, we subjected the 35-kDa BamC₂₆₋₃₄₄ to limited aminopeptidase digestion, resulting in an ~27-kDa stable fragment (data not shown). To test whether the first 75 amino acids alter the structure of the rest of the protein, we carried out hydrogen/deuterium (H/D) exchange NMR experiments on BamC₂₆₋₃₄₄ and BamC₁₀₁₋₃₄₄. 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 experiment (~11 min), consistent with a disordered N-terminus (Fig. S1). 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 (BamC₁₀₁₋₃₄₄) was used for the structure determination and for most of the NMR studies here (Fig. S2). The backbone and C ^{β} chemical shifts of BamC₁₀₁₋₃₄₄ were then assigned using conventional heteronuclear triple-resonance NMR experiments, followed by partial assignment of side-chain resonances, as described in [Materials and Methods](#).¹ Except for residues near the N-terminus, no significant differences were observed when comparing the backbone chemical shifts in the previously assigned BamC₂₆₋₃₄₄.¹⁸

The structure of BamC generated by CS-RDC-NOE Rosetta reveals a two-domain protein connected by a flexible linker

CS-RDC-NOE Rosetta calculations¹⁰ on BamC₁₀₁₋₃₄₄ were carried out as described in [Materials and Methods](#). The input consisted of the backbone (¹H^N, N, C', C ^{α} , and C ^{β}) chemical shifts for all residues

(excluding the terminal and Pro amides), 156 ^1H - ^{15}N RDCs, and two different sets of NOE restraints. A preliminary calculation was performed with 62 ^1H - ^1H 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 ^1H - ^1H 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 calculations indicated that BamC has two well-defined domains (an N-terminal domain from residues 101 to 210 and a C-terminal domain from residues 229 to 346) and a linker from residues 211 to 228. However, the relative orientation of the two domains was not defined by Rosetta (Fig. 1). One possibility for this conformational heterogeneity is that the N-terminal and C-terminal domains do not interact, and the residues between these regions are flexible. A second possibility is that BamC₁₀₁₋₃₄₄ does have a well-defined orientation of the two domains in solution, but the calculations did not find the global minimum due to incomplete sampling of conformational space. To try to observe long-range NOEs between the N-terminal domains and the C-terminal domains, we acquired a three-dimensional (3D) (^{13}C , ^{13}C , and ^1H) HMQC-NOESY-HMQC spectrum on an ILV methyl-protonated ^2H , ^{15}N , ^{13}C -labeled sample of BamC₂₆₋₃₄₄. Previous studies

have shown that methyl-methyl distances over 8 Å can be observed in NOESY spectra.¹⁹ A total of 100 new ILV methyl-methyl NOEs were observed, but none was found between residues in the N-terminal domain and residues in the C-terminal domain. The absence of interdomain NOEs alone does not prove that the N-terminal and C-terminal domains are not interacting, but it is a strong indication that there is no stable interaction between the two domains in isolated BamC.

Since the size of ^1H - ^1H NOE is a function of both the distance and the dynamics between nuclei, ^{15}N relaxation NMR experiments were performed to directly probe backbone dynamics in BamC. Heteronuclear $^{15}\text{N}\{^1\text{H}\}$ NOEs were measured for the backbone amides in BamC₁₀₁₋₃₄₄. As seen in Fig. 2a, 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 $^{15}\text{N}\{^1\text{H}\}$ 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.²⁰ As seen in Fig. 2b, residues 217-228 in BamC show lower predicted S^2 values, supporting flexibility for the linker region.

H/D exchange experiments provided additional evidence for N-terminal and C-terminal domains connected by a conformationally dynamic linker in BamC₁₀₁₋₃₄₄, 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 (Fig. 3). Moreover, it was recently shown that BamC₁₀₁₋₃₄₄ is susceptible to cleavage by subtilisin in the linker region, yielding two stable fragments of 12.2 and 14.5 kDa.²¹ All these results support a model where BamC₁₀₁₋₃₄₄ contains well-ordered N-terminal and C-terminal domains connected by a flexible linker. The CS-RDC-NOE

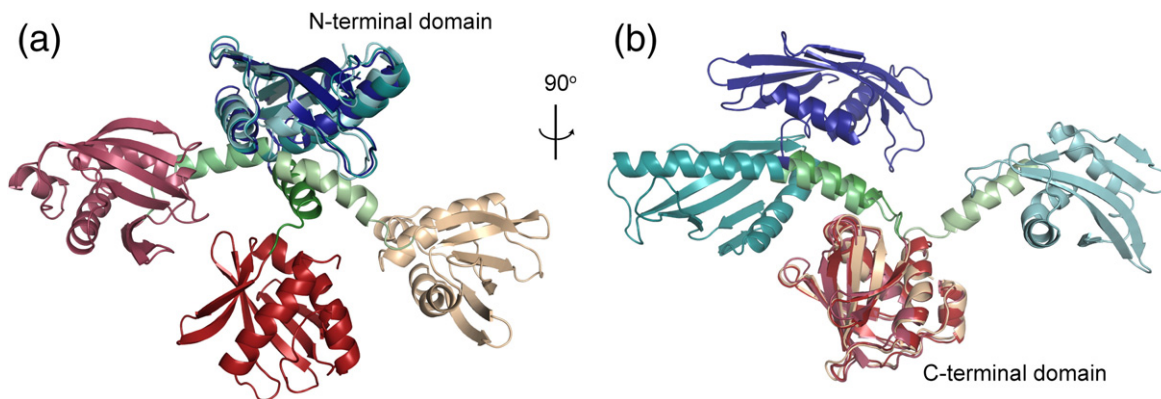


Fig. 1. Three low-energy structures for BamC₁₀₁₋₃₄₄, shown as superimpositions of (a) their N-terminal domain (cool colors) and (b) C-terminal domain (warm colors), illustrating that the folds of the N-terminal and C-terminal domains converged in the Rosetta calculations while the relative orientations 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.

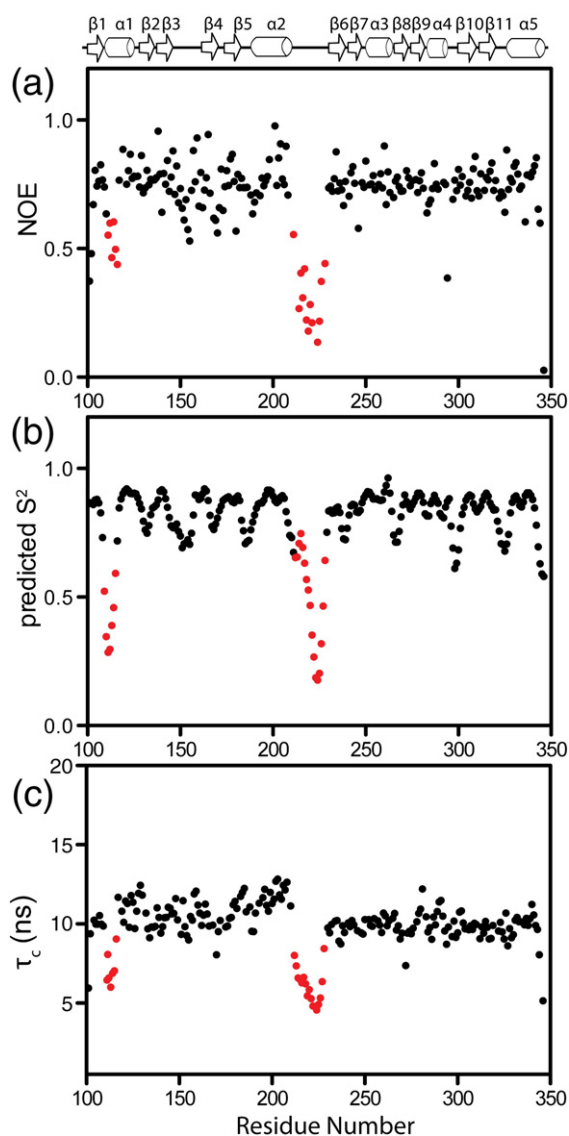


Fig. 2. The ^{15}N relaxation data on BamC₁₀₁₋₃₄₄ show regions of increased backbone flexibility. Plots of (a) ^{15}N { ^1H } NOE, (b) the predicted backbone order parameter S^2 , and (c) the calculated residue-specific rotational correlation time τ_c as a function of residue number. S^2 is predicted from experimental chemical shifts using a random-coil index.²⁰ 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.

Rosetta calculations on BamC₁₀₁₋₃₄₄ consistently generated an α -helical conformation for this linker region (Figs. 1 and 3). However, the NMR data demonstrate that the linker is flexible. This indicates that 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 (D.B., unpublished results).

A high level of flexibility for the linker does not rule out a stable interaction between the N-terminal domain and the C-terminal domain in full-length BamC. However, if the two domains are not tumbling as a rigid single species in solution, then the individual domains should have smaller rotational correlation times than predicted for the full protein. Residue-specific rotational correlation times τ_c were calculated from ^{15}N R_1 and R_2 measurements (data not shown) on BamC (Fig. 2c).^{1,22} The residues in the N-terminal and C-terminal domains have average τ_c values of 10.8 and 9.8 ns, respectively, whereas the linker has lower τ_c values (ranging from 4.6 to 7.4 ns for residues 214–227). The values for the N-terminal and C-terminal domains are $\sim 35\%$ smaller than what would be predicted for a 27-kDa spherical protein, but larger than what would be predicted if the N-terminal and C-terminal domains were tumbling independently as ~ 13 -kDa spheres (~ 7.7 ns). This pattern of ^{15}N { ^1H } NOE and τ_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.²³

RDC data are routinely used to determine the relative orientation of well-defined domains in proteins or nucleic acids.^{24,25} Thus, we wanted to address whether the ^1H - ^{15}N RDCs could be 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 the two domains is if the individual domains have similar values for the magnitude D_a and the rhombicity R of their alignment tensors.²⁴ A structure-independent method, which involves an analysis of the shape of the histograms of the RDCs in a molecule, was used here for comparing D_a and R in the two domains.²⁶ 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 ^1H - ^{15}N RDCs for the N-terminal and C-terminal domains have quite different shapes (see Fig. 4). These data are consistent with the conclusion obtained from the τ_c data that N-terminal and C-terminal domains in BamC do not have a fixed orientation in solution and are not tumbling as a rigid species.

The CS-RDC-NOE Rosetta calculations on the full-length BamC assumed a single alignment tensor for the whole molecule, but this is not correct if domains do not have a fixed orientation. Thus, separate CS-RDC-NOE Rosetta calculations

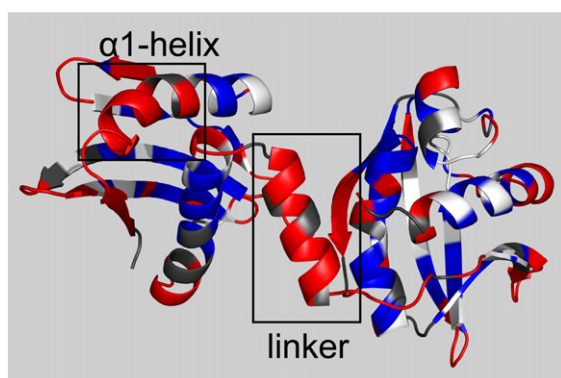


Fig. 3. H/D exchange profiles are mapped onto a low-energy structure of BamC_{101–344}, with residues in red, white, and blue indicating rapid exchange (<12 min), moderate exchange (12 min to 24 h), and slow exchange (>24 h) at 30 °C, with gray indicating prolines or residues where exchange could not be determined due to spectral overlap. Black boxes highlight rapid exchange for part of the $\alpha 1$ helix and the linker connecting the domains. Note that Rosetta modeled most of the linker region as a helix even though the ^{15}N relaxation and deuterium exchange data indicate that this region is conformationally dynamic (see the text).

were performed for the N-terminal (residues 101–212) and C-terminal (residues 229–344) domains (Fig. 5 and Table 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.

The N-terminal domain of BamC is composed of two α -helices packed against a five-stranded anti-parallel β -sheet (Fig. 5a), reminiscent of the helix-grip fold.²⁹ The C-terminal domain closely resembles the N-terminal domain, with two additional structural elements (Fig. 5b): (i) a short β -strand before $\beta 1$ and (ii) a seven-residue helix inserted between strands $\beta 3$ and $\beta 4$ (where the corresponding residues in the N-terminal domain form an extended loop). The N-terminal domain superimposes on the C-terminal domain with a pairwise RMSD of 1.7 Å for the C $^\alpha$ backbone for 60 residues in structurally similar regions identified using LSQMAN.³⁰

Figure 2 indicates that there is another region of flexibility in the N-terminal domain where residues 112–117 show $^{15}\text{N}\{^1\text{H}\}$ NOE-predicted S^2 and τ_c values lower than those of the residues in the domains. The amide protons for these residues exchanged rapidly in the H/D exchange experiments (Fig. 3). This region 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 (boxed region in Fig. 5a). Thus, the ^{15}N relaxation and H/D exchange data support a model where the N-terminal part of the $\alpha 1$ 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 nonconverged regions observed in the Rosetta calculations reflect true conformational heterogeneity of the protein.

Validation of the BamC structure generated by Rosetta

Previous studies on a number of known structures showed that iterative CS-RDC-NOE Rosetta yields accurate conformations if the calculations converge.¹⁰ The N-terminal and C-terminal regions in BamC converged well (Fig. 5); thus, based on previous experience, Rosetta is expected to generate

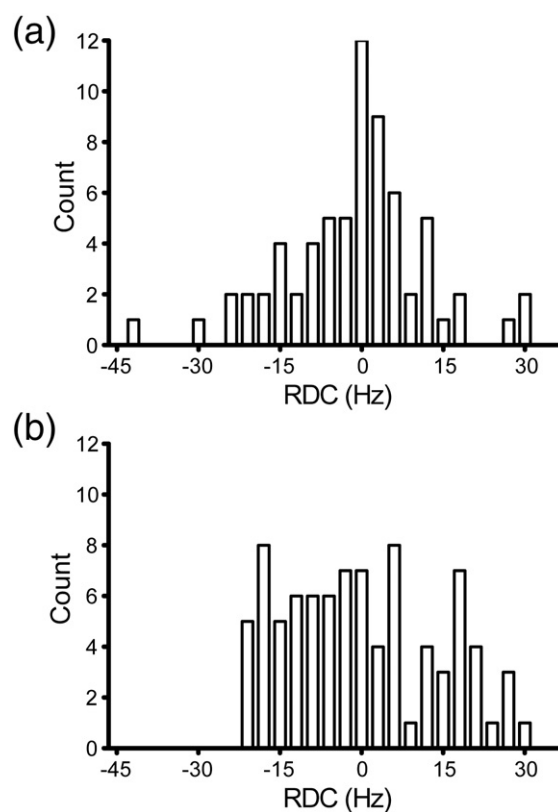


Fig. 4. Histograms of the 86 and 68 ^1H - ^{15}N RDCs for the (a) N-terminal and (b) C-terminal domains in BamC_{101–344}. The shape of the histogram provides information on the magnitude D_a and the rhombicity R of the alignment tensor.²⁶ The differences in the shapes of the two histograms indicate 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 ^1H - ^{15}N bond vectors).

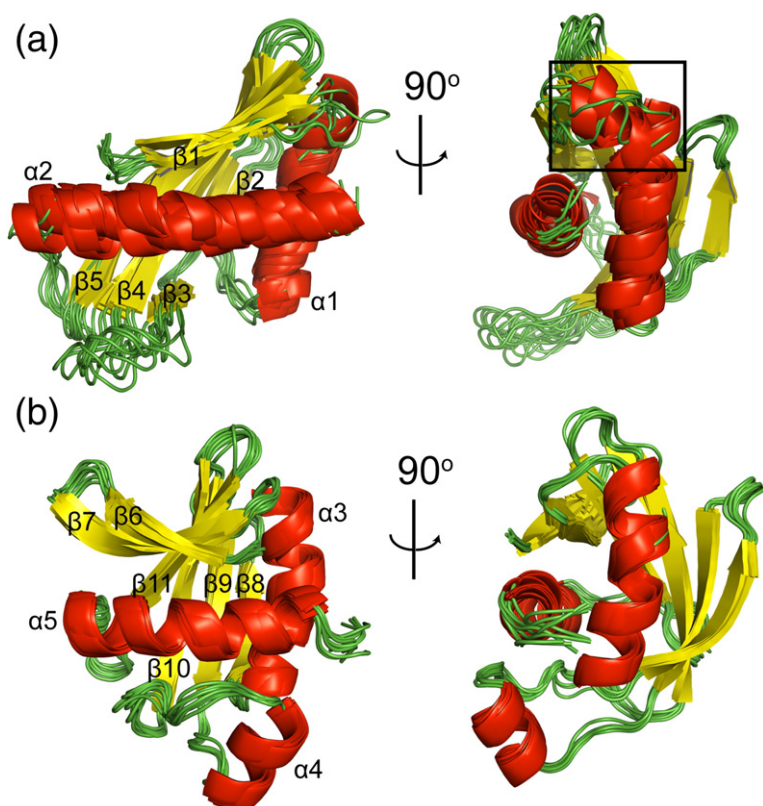


Fig. 5. Ensemble of the structures of the N-terminal and C-terminal domains of BamC generated by separate CS-RDC-NOE Rosetta calculations for the two domains. Superimposition of nine low-energy structures of the (a) N-terminal and (b) C-terminal domains of BamC. Part of the $\alpha 1$ helix in the N-terminal domain that shows conformational flexibility in the ^{15}N relaxation data (see the text) is highlighted by a black box.

accurate structures for these domains. Nevertheless, it is advisable to validate the Rosetta structures with independent experimental data. Thus, we

Table 1. Structural statistics for the BamC N-terminal and C-terminal domains

	N-terminal domain	C-terminal domain
Number of residues	112	118
NOE-based distance restraints		
NOE distance restraints (violations ≥ 0.5 Å) ^a	69 (16 $\pm$ 1)	78 (30 $\pm$ 2)
Number of restraints per residue	0.62	0.66
Other restraints		
$\phi + \psi$ dihedral-angle restraints (violations $\geq 5^\circ$) ^b	193 (11 $\pm$ 3)	202 (11 $\pm$ 4)
RDC restraints (violations ≥ 5 Hz)	60 (24 $\pm$ 3)	82 (20 $\pm$ 4)
Average RMSD to the average structure		
Backbone (Å)	1.39 $\pm$ 0.33	0.54 $\pm$ 0.06
Heavy atom (Å)	1.88 $\pm$ 0.30	0.95 $\pm$ 0.11
Ramachandran plot ^c		
Most favored regions (%)	92.2	94.3
Allowed regions (%)	7.8	4.8
Generously allowed regions (%)	0.0	0.0
Disallowed regions (%)	0.0	0.9

Statistics are given for the nine lowest-energy structures after an all-atom refinement with CS-RDC-NOE Rosetta.

^a All NOEs are long range ($|i-j| \geq 5$). Two of the 149 NOEs were found between the amino acids in the linker region and the N-terminal domain, and were not included in the calculations of the individual domains.

^b Torsion-angle restraints were derived from TALOS+.²⁷

^c Procheck was used to calculate these data.²⁸

analyzed how additional backbone-backbone and methyl-methyl ^1H - ^1H NOEs not included in the calculations fit the Rosetta models for BamC₁₀₁₋₃₄₄. The input NOE data set 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 evenly distributed over the structure, where regions have clusters of NOEs and others have no NOEs. The gray boxes in Fig. 6 highlight regions of anti-parallel β -sheet secondary structure where no cross-strand NOEs were included in the Rosetta calculations. For example, $\beta 7$ and $\beta 11$ had no cross-strand NOE restraints, 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. Further analysis of the original NOESY spectra yielded three cross-strand backbone-backbone NOEs in this region (H^{N} V244 and H^{N} S314, H^{α} V243 and H^{α} Ser315, and H^{N} L316 and H^{N} L242), which unambiguously confirmed the strand orientation and register generated by Rosetta between $\beta 7$ and $\beta 11$ (Fig. 6). The cross-strand H^{α} - H^{α} NOE is especially diagnostic of an anti-parallel β -sheet because the H^{α} - H^{α} distance is ~ 2.2 - 2.5 Å. Strong H^{α} - H^{α} cross-peaks, as well as other cross-strand ^1H - ^1H NOEs, were also observed for the two other boxed regions in Fig. 6, validating the β -sheet secondary structure generated by Rosetta.

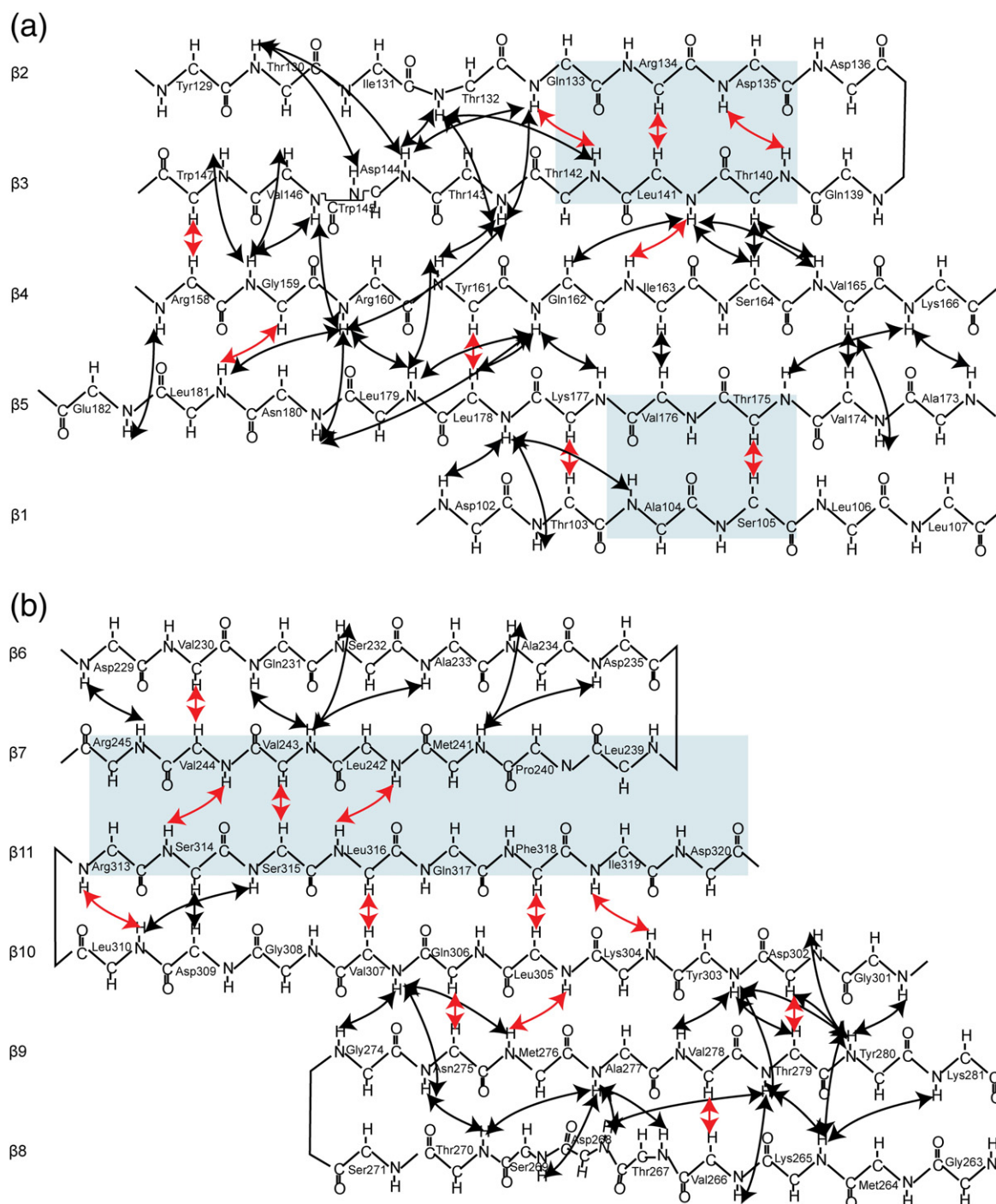


Fig. 6. Backbone-backbone ^1H - ^1H NOEs confirm the β -sheet secondary structure in BamC generated by CS-RDC-NOE Rosetta. Schematics for the β -sheet secondary structures in the (a) N-terminal and (b) C-terminal domains showing NOEs included (black) or not included (red) in the CS-RDC-NOE Rosetta calculations. NOEs in red were used to help validate the Rosetta structures.

One of Rosetta's strengths is that it can generate accurate side-chain packing in cases of high backbone convergence (<2 Å).¹⁰ The ILV methyl-protonated NOESY spectra on BamC were initially collected to search for long-range interdomain methyl-methyl NOEs. However, these spectra also

provide data on the packing of the ILV side chains. A set of 103 NOEs was observed between the ILV methyl groups in BamC, but only three of these were included as restraints in the Rosetta calculations. In all cases, the range of distances predicted from these methyl-methyl NOEs was consistent with the

hydrophobic packing observed in the BamC structures generated by Rosetta.

CS-RDC-NOE Rosetta is very robust to errors in an NOE restraint list

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.¹⁰ 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 for large distance violations Rosetta uses a linear potential as opposed to the quadratic potential commonly used for NOE restraints. The results here demonstrate that CS-RDC-NOE Rosetta is very robust in dealing with inconsistent NOEs. As seen in [Fig. 7](#), 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.¹⁰ To test the effect of having this level of misassigned NOEs on a standard structure calculation, we performed XPLOR-NIH³¹ simulations using the N-terminal domain of one of the final Rosetta structures as the target ([Supplementary Materials](#)). As expected, when no misassignments were introduced in the simulated data, the calculations generated well-defined structures with low NOE energies and small RMSDs to the target structures (average RMSD, 0.34 ± 0.14 Å). However, when 16% of ~1700 simulated NOEs were misassigned, mimicking the 10 of 62 misassignments in the preliminary CS-RDC-NOE Rosetta calculations, the ensemble of structures had huge NOE energies and did not converge to the target (average RMSD, 15.5 ± 2.0 Å) (see [Supplementary Materials](#)). 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 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

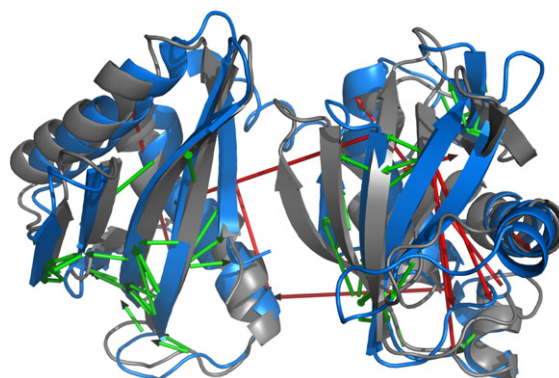


Fig. 7. CS-RDC-NOE Rosetta modeling is tolerant of misassigned NOEs. The best-scoring model in a preliminary calculation that included 10 misassigned NOEs is shown in blue. Superimposed in gray are the individual domains of the best-scoring model of the final calculation, BamC_{II}. Amide H^N–H^N NOEs 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.

RDC restraints during the all-atom refinement (see [Materials and Methods](#)). 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 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 have a clear advantage when applied to large proteins.

Comparison of BamC with other helix–grip motifs

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 from independent searches of the Dali protein structural database.³² [Figure 8a](#) shows the superimposition of the N-terminal domain with the major latex protein At1g24000.1.²⁹ 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.³³ Structurally, these proteins are characterized by a hydrophobic cavity between the β -sheet and a long α -helix that accommodates the ligand. 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 ¹H,¹⁵N BamC_{101–344} and

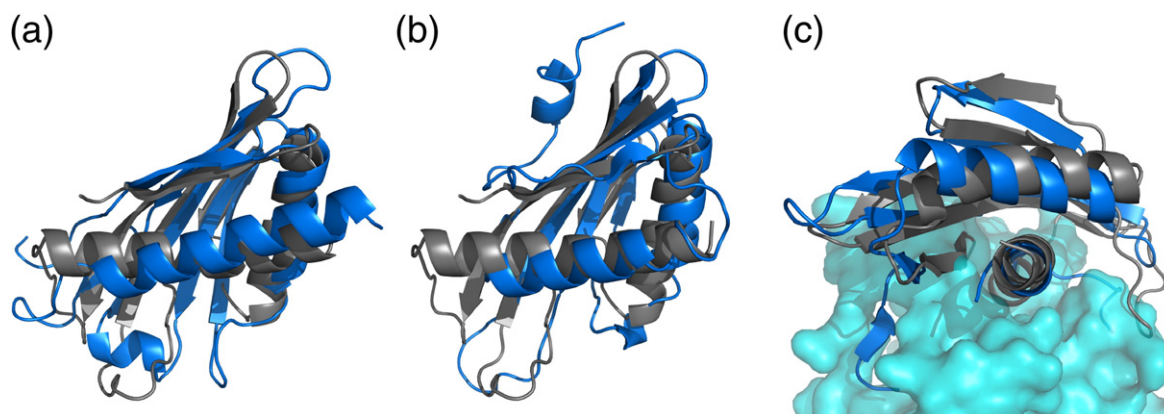


Fig. 8. The N-terminal domain of BamC (gray) is shown superimposed with structurally similar proteins (blue) identified in a Dali search. (a) The major latex protein At1g24000.1 (PDB ID: 1VJH), (b) the yeast septin-associated Kcc4p (PDB ID: 3OSM), and (c) the helix–grip domain of the α -subunit of AMPK (PDB ID: 2V8Q) 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.³⁰

palmitic acid (40 μ M) or an extract of *E. coli* polar lipids (120 μ M) showed no changes in the amide chemical shifts in BamC (data not shown).

The Dali search also identified KA-1 (kinase-associated) domains as being structurally related to both the N-terminal domains and the C-terminal domains of BamC (Fig. 8b). It was recently shown that, in some kinases such as human MARK (microtubule affinity regulating kinase)/Par1 (partitioning-defective 1 kinase) and the yeast septin-associated Kcc4p, the KA-1 domains bind acidic phospholipids and are responsible for targeting the kinase to the plasma membrane.³⁴ 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 display a more negative character than the KA-1 domains in these kinases (data not shown). These results, together with the lack of changes in chemical shifts upon addition of *E. coli* phospholipids, suggest that BamC is not 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.^{35,36} The long C-terminal α -helix in these KA-1 domains contains most of the residues that participate in intersubunit contacts (see Fig. 8c). 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 Fig. 8c. Residues on the surface of the corresponding helix ($\alpha 2$) of the N-

terminal BamC domain have a relatively high phylogenetic conservation (Fig. 9). A second region of surface-exposed highly conserved residues is also observed on the $\alpha 3$ helix in the C-terminal domain of BamC. Previous studies have shown that BamC interacts directly with the C-terminus of BamD and helps stabilize the BAM complex.^{37,38} 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.

Conclusions

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 loop used to interact with phospholipids observed in Kcc4p. Therefore, we speculate that helix–grip motifs in BamC function as protein–protein interaction domains to help stabilize interactions between the subunits of the BAM complex. Specifically, the long helix in the helix–grip domains may mediate interactions with the other components of the BAM complex, similar to the protein–protein interaction domain observed in AMPK.³⁶

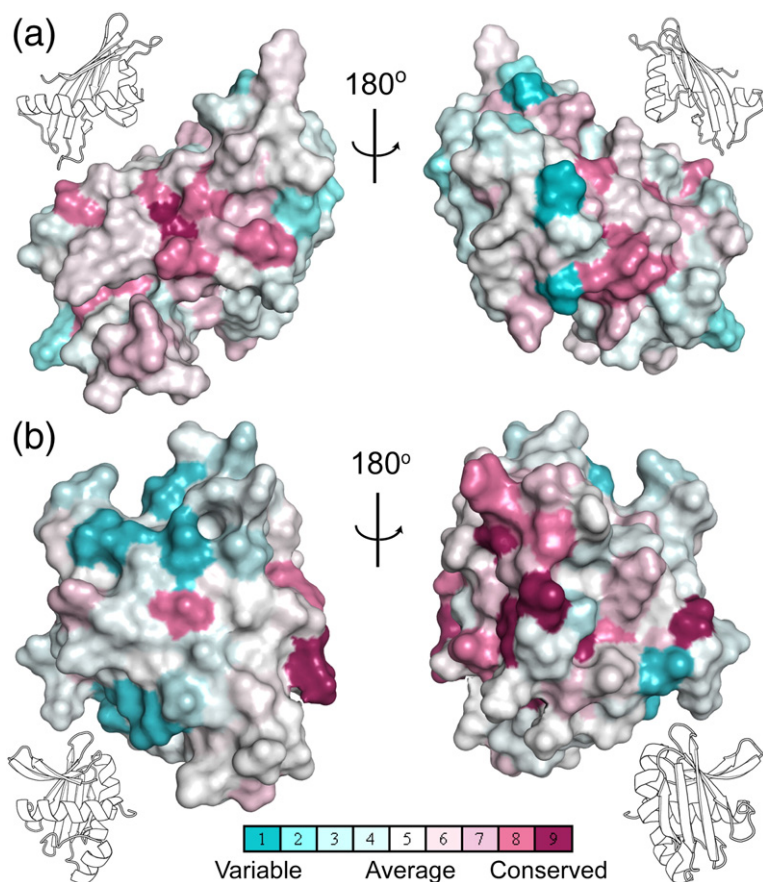


Fig. 9. The conserved amino acids in BamC are clustered in sites potentially important for protein–protein interactions. Conserved sequences are mapped onto surface representations of the (a) N-terminal and (b) C-terminal domains of BamC (the insets show the secondary structure representation of the 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.

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 could 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. This greatly simplifies the analysis of NOE spectra and makes CS-RDC-NOE Rosetta an attractive alternative to restrained molecular mechanics/dynamics methods for generating solution structures of proteins from NMR data.

Materials and Methods

Cloning of BamC constructs

The gene for *E. coli* BamC (residues 26–344) was obtained by PCR from genomic *E. coli* DNA using primers that introduce unique NcoI and XmaI restriction sites. This gene was introduced into the pMS174 vector [an engi-

neered variant of the pET28 vector that generates an N-terminal His-tag fusion that can be specifically cleaved with tobacco etch virus (TEV) protease] to yield pMS282. BamC_{101–344}, a second construct that is shorter by 75 amino acids at the N-terminus, was PCR amplified from pMS282 and ligated into pMS174 with the NcoI/XmaI restriction site to yield pMS639. Plasmids pMS282 (BamC_{26–344}) and pMS639 (BamC_{101–344}) were sequenced to confirm the absence of mutations. To facilitate cloning, we included three nonnative amino acids at the N-terminus (AGM) and two nonnative amino acids at the C-terminus (PG) in BamC_{26–344} and BamC_{101–344} constructs, but these amino acids are not included in the numbering system used to refer to the constructs.

Protein expression and purification of BamC constructs

Plasmids pMS282 and pMS639 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 resuspended in 3 L of M9 minimal medium supplemented with 50 µg/mL kanamycin, 1.5 g/L [¹³C]glucose, and 1 g/L ¹⁵NH₄Cl (Sigma/Isotec). Cultures were grown at 37 °C to an OD₆₀₀ 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 resuspended 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) preequilibrated 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) preequilibrated with buffer A and eluted in the same buffer. Protein was concentrated to 1.6 mM for BamC_{26–344} and to 1.8 mM for BamC_{101–344}, and stored at –70 °C.

For the ILV methyl-protonated labeled sample, pMS282 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 grown in a rich medium for 3 h at 37 °C. Cells were first conditioned to a minimal medium in solutions containing H₂O and finally transferred to 2 L of M9 minimal medium prepared in 99.8% ²H₂O with 2 g/L [²H,¹³C]glucose and 1 g/L ¹⁵NH₄Cl (Isotec). Cells were grown at 37 °C until an OD₆₀₀ of 0.6 had been reached, at which point 100 mg/L 2-keto-3-(methyl-*d*₃)-butyric acid-1,2,3,4-¹³C₄,3-*d*₁ sodium salt and 50 mg/L 2-ketobutyric acid-¹³C₄,3,3-*d*₂ sodium salt (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 above.

NMR spectroscopy

For resonance assignments, the ¹³C,¹⁵N-labeled BamC_{101–344} sample was exchanged into NMR buffer containing 10% ²H₂O, 50 mM sodium phosphate (pH 6.0), 50 mM NaCl, 0.02% NaN₃, 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_{101–344}: 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.¹ The backbone assignments for BamC_{101–344} are very similar to the previously published assignments for BamC_{26–344}.¹⁸ The ¹⁵N-edited and ¹³C-edited 3D NOESY spectra were collected with a 150-ms mixing time on the ¹³C,¹⁵N-labeled BamC_{101–344} sample. For the ¹³C-edited 3D NOESY, the ¹³C,¹⁵N-labeled BamC_{101–344} sample was exchanged into NMR buffer with 99.8% ²H₂O. 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.

The amide ¹H–¹⁵N RDCs were measured using a 1.1 mM ¹³C,¹⁵N-labeled BamC_{101–344} sample with no Pf1 phage and a 0.14 mM ¹³C,¹⁵N-labeled BamC_{101–344} sample in 21 mg/mL liquid crystalline Pf1 phage, prepared as

described previously.³⁹ Two-dimensional heteronuclear single-quantum coherence sensitivity-enhanced ¹⁵N IPAP spectra⁴⁰ were collected on isotropic (no Pf1 phage) and aligned (with Pf1 phage) samples. For H/D exchange experiments, 0.2 mM ¹³C,¹⁵N-labeled BamC_{26–344} and 1.0 mM ¹³C,¹⁵N-labeled BamC_{101–344} were rapidly exchanged into NMR buffer with 99.8% ²H₂O using buffer-exchange spin columns (Pierce). The selective optimized flip-angle short transient ¹⁵N HMQC sequence⁴¹ was used to collect two-dimensional ¹H, ¹⁵N spectra, each with a total experiment time of 4 min. 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 peak volumes from the spectra were fitted to an exponential decay function. The data for determining ¹⁵N *R*₁ and *R*₂ relaxations rates and ¹⁵N{¹H} heteronuclear NOE data were collected as described previously.^{1,22} Residue-specific *τ*_c values were calculated from the ¹⁵N *R*₁ and *R*₂ relaxations rates using $\tau_c = \frac{1}{4\pi\nu_N} \sqrt{6\frac{R_2}{R_1} - 7}$.⁴² The *τ*_c values for the N-terminal and C-terminal domains were the average of the *τ*_c values for residues with ¹⁵N{¹H} heteronuclear NOE values greater than 0.7. Estimates for the *τ*_c values of the individual domains were made using an empirical relation between molecular weight and *τ*_c values[†]. The methyl groups were assigned in the ILV methyl-protonated ²H,¹³C,¹⁵N-labeled sample of BamC_{26–344} using a spectrum collected with an HMMCGCBCA pulse sequence.³ Methyl–methyl NOEs were measured using a 3D ¹³C,¹³C,¹H NOESY pulse sequence⁴³ recorded with a 240-ms mixing time. 2,2,3,3-Tetradeutero-3-(trimethylsilyl)-propionic acid was used as internal chemical shift reference, and spectra were referenced to 2,2-dimethyl-2-silapentane-5-sulfonate by correcting for the pH dependence of the 2,2,3,3-tetradeutero-3-(trimethylsilyl)-propionic acid chemical shift.⁴⁴ All spectra were processed with NMRPipe⁴⁵ and analyzed with either SPARKY⁴⁶ or CCPNMR analysis.⁴⁷

Structure generation with iterative CS-RDC-NOE Rosetta

The backbone (¹H^N, N, C', C^α, and C^β) chemical shifts of the construct BamC_{101–344} were used in CS Rosetta 3.X to select 25 and 200 fragments that are nine and three residues in length, respectively.⁹ Since TALOS+²⁷ predicted dynamic or disordered regions at the termini of the BamC_{101–344} construct to be less than five residues in length, no tails were removed for the modeling procedure. However, sequence regions that are at least three residues in length or are located at the termini (residues 110–114, 218–228, 296–301, 345, and 346, respectively) and have TALOS+–predicted order parameters²⁷ of less than 0.7 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 *E. coli* 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-

[†] http://www.nmr2.buffalo.edu/negs.wiki/NMR_determined_Rotational_correlation_time

NOE Rosetta protocol, as described previously.¹⁰ A set of 50,000 structures was produced in the initial low-resolution sampling, and 15,000 of these were chosen for all-atom refinement. 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. To select the pool of structures that are iterated further, we used a weight of 5 for the NOE and RDC data for the low-resolution and all-atom structures. The standard Rosetta all-atom energy function was used with the weight set score13_env_hb.⁴⁸ We note that this weight set has recently been found to hold no improvement over the more established score12,⁴⁸ which is therefore recommended for use in the future. The amide ¹H–¹H 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. The iterative CS-RDC-NOE Rosetta calculations on the full-length BamC required ~40 h on 512 nodes of a Blue Gene/P computer (~20,000 node hours).

Accession numbers

Atomic coordinates have been deposited in the Protein Data Bank (PDB) under accession codes 2LAF (N-terminal domain) and 2LAE (C-terminal domain). Chemical shifts and structural data have been deposited in the BioMagRes-Bank under accession code 17521.

Acknowledgements

This research was supported by National Institutes of Health (NIH) grants AI080709 (M.C.S.) and GM092802 (D.B.), and NIH training grant T32 GM08759 (L.R.W.). NMR instrumentation was purchased with partial support from NIH grants GM068928, RR11969, and RR16649; National Science Foundation grants 9602941 and 0230966; and the W. M. Keck Foundation. We also thank Dr. Geoff Armstrong for help with NMR data collection.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2011.05.022](https://doi.org/10.1016/j.jmb.2011.05.022)

References

- Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III, Rance, M. & Skelton, N. J. (2007). *Protein NMR Spectroscopy: Principles and Practice*, 2nd edit. Academic Press, San Diego, CA.
- Wider, G. (2005). NMR techniques used with very large biological macromolecules in solution. *Methods Enzymol.* **394**, 382–398.
- Tugarinov, V. & Kay, L. E. (2003). Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J. Am. Chem. Soc.* **125**, 13868–13878.
- Lemak, A., Gutmanas, A., Chitayat, S., Karra, M., Fares, C., Sunnerhagen, M. & Arrowsmith, C. H. (2011). A novel strategy for NMR resonance assignment and protein structure determination. *J. Biomol. NMR*, **49**, 27–38.
- Guntert, P. (2009). Automated structure determination from NMR spectra. *Eur. Biophys. J. Biophys.* **38**, 129–143.
- Altieri, A. S. & Byrd, R. A. (2004). Automation of NMR structure determination of proteins. *Curr. Opin. Struct. Biol.* **14**, 547–553.
- Raman, S., Huang, Y. J. P., Mao, B. C., Rossi, P., Aramini, J. M., Liu, G. H. *et al.* (2010). Accurate automated protein NMR structure determination using unassigned NOESY data. *J. Am. Chem. Soc.* **132**, 202–207.
- Tang, Y., Schneider, W. M., Shen, Y., Raman, S., Inouye, M., Baker, D. *et al.* (2010). Fully automated high-quality NMR structure determination of small (2)H-enriched proteins. *J. Struct. Funct. Genomics*, **11**, 223–232.
- Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J. M., Liu, G. *et al.* (2008). Consistent blind protein structure generation from NMR chemical shift data. *Proc. Natl Acad. Sci. USA*, **105**, 4685–4690.
- Raman, S., Lange, O. F., Rossi, P., Tyka, M., Wang, X., Aramini, J. *et al.* (2010). NMR structure determination for larger proteins using backbone-only data. *Science*, **327**, 1014–1018.
- Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. (2004). Protein structure prediction using Rosetta. *Methods Enzymol.* **383**, 66–93.
- Das, R. & Baker, D. (2008). Macromolecular modeling with Rosetta. *Annu. Rev. Biochem.* **77**, 363–382.
- Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H. & Meiler, J. (2010). Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry*, **49**, 2987–2998.
- Shen, Y., Vernon, R., Baker, D. & Bax, A. (2009). *De novo* protein structure generation from incomplete chemical shift assignments. *J. Biomol. NMR*, **43**, 63–78.
- Knowles, T. J., Scott-Tucker, A., Overduin, M. & Henderson, I. R. (2009). Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. *Nat. Rev. Microbiol.* **7**, 206–214.
- Wu, T., Malinverni, J., Ruiz, N., Kim, S., Silhavy, T. J. & Kahne, D. (2005). Identification of a multicomponent complex required for outer membrane biogenesis in *Escherichia coli*. *Cell*, **121**, 235–245.
- Onufryk, C., Crouch, M. L., Fang, F. C. & Gross, C. A. (2005). Characterization of six lipoproteins in the sigmaE regulon. *J. Bacteriol.* **187**, 4552–4561.
- Knowles, T. J., McClelland, D. M., Rajesh, S., Henderson, I. R. & Overduin, M. (2009). Secondary structure and (1)H, (13)C and (15)N backbone resonance assignments of BamC, a component of the outer membrane protein assembly machinery in *Escherichia coli*. *Biomol. NMR Assign.* **3**, 203–206.
- Mueller, G. A., Choy, W. Y., Yang, D., Forman-Kay, J. D., Venters, R. A. & Kay, L. E. (2000). Global folds of proteins with low densities of NOEs using residual

- dipolar couplings: application to the 370-residue maltodextrin-binding protein. *J. Mol. Biol.* **300**, 197–212.
20. Berjanskii, M. V. & Wishart, D. S. (2005). A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc.* **127**, 14970–14971.
 21. Albrecht, R. & Zeth, K. (2010). Crystallization and preliminary X-ray data collection of the *Escherichia coli* lipoproteins BamC, BamD and BamE. *Acta Crystallogr. Sect. F*, **66**, 1586–1590.
 22. Farrow, N. A., Zhang, O. W., Formankay, J. D. & Kay, L. E. (1995). Comparison of the backbone dynamics of a folded and an unfolded SH3 domain existing in equilibrium in aqueous buffer. *Biochemistry*, **34**, 868–878.
 23. Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W. & Bax, A. (1992). Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry*, **31**, 5269–5278.
 24. Fischer, M. W., Losonczi, J. A., Weaver, J. L. & Prestegard, J. H. (1999). Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. *Biochemistry*, **38**, 9013–9022.
 25. Molloy, E. T., Hansen, M. R. & Pardi, A. (2000). Global structure of RNA determined with residual dipolar couplings. *J. Am. Chem. Soc.* **122**, 11561–11562.
 26. Clore, G. M., Gronenborn, A. M. & Bax, A. (1998). A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J. Magn. Res.* **133**, 216–221.
 27. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. (2009). TALOS plus: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, **44**, 213–223.
 28. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). Procheck—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291.
 29. Lytle, B. L., Song, J., de la Cruz, N. B., Peterson, F. C., Johnson, K. A., Bingman, C. A. *et al.* (2009). Structures of two *Arabidopsis thaliana* major latex proteins represent novel helix–grip folds. *Proteins*, **76**, 237–243.
 30. Kleywegt, G. J. & Jones, T. A. (1997). Detecting folding motifs and similarities in protein structures. *Methods Enzymol.* **277**, 525–545.
 31. Schwieters, C. D., Kuszewski, J. J. & Clore, G. M. (2006). Using XPLOR-NIH for NMR molecular structure determination. *Prog. NMR Spectrosc.* **48**, 47–62.
 32. Holm, L. & Rosenstrom, P. (2010). Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549.
 33. Radauer, C., Lackner, P. & Breiteneder, H. (2008). The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *BMC Evol. Biol.* **8**, 286.
 34. Moravcevic, K., Mendrola, J. M., Schmitz, K. R., Wang, Y. H., Slochow, D., Janmey, P. A. & Lemmon, M. A. (2010). Kinase associated-1 domains drive MARK/PAR1 kinases to membrane targets by binding acidic phospholipids. *Cell*, **143**, 966–977.
 35. Xiao, B., Heath, R., Saiu, P., Leiper, F. C., Leone, P., Jing, C. *et al.* (2007). Structural basis for AMP binding to mammalian AMP-activated protein kinase. *Nature*, **449**, 496; U14.
 36. Oakhill, J. S., Scott, J. W. & Kemp, B. E. (2009). Structure and function of AMP-activated protein kinase. *Acta Physiol.* **196**, 3–14.
 37. Sklar, J. G., Wu, T., Gronenberg, L. S., Malinverni, J. C., Kahne, D. & Silhavy, T. J. (2007). 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.
 38. Malinverni, J. C., Werner, J., Kim, S., Sklar, J. G., Kahne, D., Misra, R. & Silhavy, T. J. (2006). YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in *Escherichia coli*. *Mol. Microbiol.* **61**, 151–164.
 39. Hansen, M. R., Mueller, L. & Pardi, A. (1998). Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.* **5**, 1065–1074.
 40. Ding, K. Y. & Gronenborn, A. M. (2003). Sensitivity-enhanced 2D IPAP, TROSY–anti-TROSY, and E.COSY experiments: alternatives for measuring dipolar N–H–H–N couplings. *J. Magn. Res.* **163**, 208–214.
 41. Schanda, P., Kupce, E. & Brutscher, B. (2005). SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J. Biomol. NMR*, **33**, 199–211.
 42. Kay, L. E., Torchia, D. A. & Bax, A. (1989). Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, **28**, 8972–8979.
 43. Zwahlen, C., Gardner, K. H., Sarma, S. P., Horita, D. A., Byrd, R. A. & Kay, L. E. (1998). An NMR experiment for measuring methyl–methyl NOEs in C-13-labeled proteins with high resolution. *J. Am. Chem. Soc.* **120**, 7617–7625.
 44. Tynkkynen, T., Tiainen, M., Soininen, P. & Laatikainen, R. (2009). From proton nuclear magnetic resonance spectra to pH. Assessment of ^1H NMR pH indicator compound set for deuterium oxide solutions. *Anal. Chim. Acta*, **648**, 105–112.
 45. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe—a multidimensional spectral processing system based on Unix pipes. *J. Biomol. NMR*, **6**, 277–293.
 46. Goddard, T. D. & Kneller, D. G. SPARKY 3. University of California, San Francisco, CA.
 47. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M. *et al.* (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59**, 687–696.
 48. Tyka, M. D., Keedy, D. A., Andre, I., Dimaio, F., Song, Y., Richardson, D. C. *et al.* (2011). Alternate states of proteins revealed by detailed energy landscape mapping. *J. Mol. Biol.* **405**, 607–618.

Note added in proof: After this manuscript was accepted, X-ray structures of the individual N-terminal and C-terminal domains of BamC were reported [Albrecht, R. & Zeth, K. (2011). *J. Biol. Chem.* on-line; <http://www.jbc.org/cgi/doi/10.1074/jbc.M111.238931>]. The X-ray structures are essentially identical to those determined here, further validating the accuracy of the models generated by CS-RDC-NOE Rosetta.