Analysing structural data to explore the function of an essential bacterial protein foldase

Samuel F. Haysom
(University of Leeds, UK)

Structural biology, or the study of how protein structures dictate their function, is a fundamental part of life science research, allowing the mechanisms underpinning life to be unravelled at the molecular level. Due to the complexity of 3D data, researchers often use special visualization methods to extract useful information from protein structures. This article uses the most common of these visualisation methods to examine different structures of the β-barrel assembly machinery complex (BAM), an essential protein that folds other proteins into the outer-membranes of Gram-negative bacteria. By exploring how BAM’s 3D shape changes as it interacts with its substrates throughout the folding process, it is possible to reconstruct a potential mechanism for this molecular machine that can be used to drive further research.

What is structural biology?
At the microscopic level, most of life’s processes are carried out by molecular machines. These machines, predominantly made of protein, enable essential activities, for instance nutrient acquisition, energy production and cellular repair/maintenance, making the understanding of how they work a fundamental biological question. Structural biologists attempt to answer this question by determining the 3D shape of these molecular machines. By seeing how their shape changes, or how the molecule ‘moves’ as it performs its function, we can often work out the underlying mechanism.

To illustrate how this works, imagine trying to understand how a human hand can pick up an object. If we had enough ‘snapshots’ of the process occurring we might see that the hand can be closed in a fist, that it can also be open and flat and finally that the fingers and thumb can wrap around an object. Intuitively we could work out that the hand starts as a fist, opens and then grasps the object. From the snapshots we can piece together the whole process.

While molecules cannot be seen with the naked eye, structural biologists use techniques such as crystallography and cryo-electron microscopy (see further reading 1, 2 and 3) to determine the 3D position of atoms in biomolecules, resulting in molecular ‘snapshots’ commonly called structures. In crystallography, structures are obtained from the diffraction pattern formed by interaction of X-rays with a crystal formed from many copies of a biomolecule. In cryoEM, structures are obtained by averaging thousands of images of individual biomolecules frozen in non-vitreous ice, with the images obtained by passing a beam of electrons through this ice onto a specialized detector.

Biomolecular structures can be incredibly complex, containing thousands of atoms, and require special visualization approaches to extract useful information. These approaches are now easily accessible using programs such as PyMOL and UCSF Chimera/ChimeraX, which allow viewing and manipulation of structures computationally. In this article we will examine the most popular approaches in detail and see how they can be used to reconstruct a mechanism of action for one particular molecular machine known as the β-barrel assembly machinery complex (BAM).

The BAM complex is a molecular machine that assembles other proteins
BAM is found in the Gram-negative bacterial outer-membrane (OM), a physical barrier that separates and protects the cell interior from the surrounding environment. Gram-negative bacteria need many proteins, known as outer-membrane proteins (OMPs), in their OM. Some build and maintain the membrane itself, others acquire food and in certain species some assist in causing disease. These proteins do not just appear in the membrane ready to do their job. Instead bacteria rely on the BAM
complex to take the OMPs, fold them into the correct shape for their function and insert them into the membrane. This is an essential process and inhibiting it could lead to new antibiotic therapies for bacterial diseases. Structural biologists have therefore solved structures of the BAM complex at various stages throughout the folding cycle, allowing them to start to piece together this mechanism.

**Cartoon representations and secondary structure**

When scientists look at a structure for the first time, they often use a cartoon representation of the protein. This takes advantage of two aspects of protein structure to give a simplified representation that still provides a lot of information. First, proteins are composed of amino acids bonded together in a long chain. Instead of showing the whole of each amino acid, we can just show a trace through an equivalent position of each amino acid (the Cα carbon), revealing how the protein chain is folded without any added complexity (Figure 1a). Second, protein chains mostly fold into the same patterns, known as secondary structure motifs, as these allow an optimum arrangement of hydrogen bonds (h-bonds, Table 1) between amino acids. Where a protein chain adopts a particular secondary structure, this can be represented by making the trace a certain shape. The two most common secondary structures are α-helices, where the chain forms a right-handed helical corkscrew with h-bonds between every fourth amino acid, and β-sheets, where strands of the protein chain line up next to each other to form a sheet, and h-bonds link amino acids in neighbouring strands. In cartoon representations,

![Figure 1](http://portlandpress.com/biochemist/article-pdf/43/5/90/922653/bio_2021_166.pdf)

**Figure 1.** The cartoon representation simplifies a protein's structure. (a) Top shows part of a protein in atomic representation (showing all atoms). To simplify this view, a trace (purple line) can be taken through the Cα atom of each amino acid residue in the chain (purple circles). Bottom shows the resulting cartoon representation. (b) Protein chains commonly fold into α-helices, where the chain forms a corkscrew with h-bonds between amino acids at different levels in the chain (top). In cartoon representations α-helices are shown as cylindrically spiralling ribbons (bottom). (c) Protein chains also commonly form β-sheets, where strands (known as β-strands) of the protein chain line up parallel to each other, with hydrogen bonds between the β-strands keeping them in place. In cartoon representations, β-strands are shown as flattened arrows which line up next to each other to form the sheet.
α-helical stretches are shown as cylindrically spiralled ribbons (Figure 1b), while each strand in a β-sheet (known as a β-strand) is shown as a flattened arrow (Figure 1c).

By combining the backbone trace with these secondary structure patterns, a scientist can discern a large amount about a protein structure at a glance.

We can see this in practice in a cartoon representation of a BAM structure, determined by crystallography (Figure 2a). BAM has five protein subunits, known as BamA–E, with BamA forming a ‘core’ to which all other subunits are bound. The subunits differ markedly in how their secondary structure elements fold together into the proteins’ overall shape, commonly known as their tertiary structure. Different tertiary structures are associated with different functions, suggesting distinct roles for each subunit. Currently, BamA is thought to perform the actual primary folding process, while the other subunits regulate how BamA behaves.

So, what does BamA’s tertiary structure tell us about its function? One half of the protein (C-terminal) forms a structure called a β-barrel (Figure 2b), which is simply a β-sheet bent to form a cylinder shape (Figure 2c). This motif is very common and allows proteins to sit within the OM because the chemical groups on the barrel exterior like to be in contact with membrane lipids rather than the water molecules (they are hydrophobic, see Table 1). Additionally, a ‘girdle’ of aromatic residues (Table 1) at the meeting point between membrane and surrounding solution helps to stabilize the barrel in the membrane (Figure 2e). The β-barrel is the defining feature of OMPs, with BAM itself being the machinery that allows them to be inserted into the membrane. Notably, BamA itself is itself an OMP, and new copies of BamA can be folded into the membrane by existing BAM complexes. The other half of BamA (N-terminal) is composed of five POTRA domains, which corkscrew down from the membrane into the cell and act as binding sites for BAM’s other subunits. Overall, BamA’s location in the membrane and role as a focal point for the other complex components suggests it plays an important role in BAM’s folding mechanism.

### Surface representations and the BamA lateral-gate

The OMPs that BAM folds into the membrane are known as its substrates and are delivered to the complex on the POTRA domain side as an unfolded chain. Upon receipt, BAM must both fold this chain into a β-barrel and transfer it into the membrane. An obvious component to mediate this transfer into the membrane is the BamA β-barrel, since it is also embedded in the membrane.

When scientists solved the first structures of BAM, they noticed that the BamA β-barrel was not always the same shape. The most striking difference was at the location where the two ends of the β-sheet that are curved to form the β-barrel came together to close it into a cylinder, known as the β-seam. In other OMPs, this seam is normally completely zipped shut by h-bonds (Figure 2d) but for some BAM structures it is completely open, with no h-bonds between the two ends of the β-sheet. As a result of this opening and closing behaviour (Figure 3a), the BamA β-seam is often called a lateral-gate.

To better show the implications of this for BAM’s mechanism, it helps to use a surface representation of the protein. Here we are interested in which parts of the protein are in contact with the surrounding solvent molecules, either water, or for the BamA β-barrel the lipid molecules that make up the membrane into which it is embedded. By rolling a virtual solvent molecule over the protein, we can calculate and display a solvent-accessible surface. This surface is of interest to structural biologists as features at the solvent–protein boundary determine how a protein interacts with the outside world, for instance in determining how it binds to other proteins.

By viewing the two conformations of BamA as a surface (Figure 3c), we can see that the opening of the lateral-gate leads to a large gap in the side of the BamA barrel. As this gap faces the OM it was hypothesized that it could act as an exit for OMPs folding on BAM to pass into the membrane.

### Substrate folding occurs at the BamA lateral-gate

For BAM to fold an OMP’s β-barrel, it must first mediate formation of a β-sheet. When the lateral-gate opens, this breaks the h-bonds linking the first and last β-strands in the membrane (Figure 2e).

<table>
<thead>
<tr>
<th>Table 1. Key Definitions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gram-negative bacteria</strong></td>
</tr>
<tr>
<td><strong>Outer-membrane (OM)</strong></td>
</tr>
<tr>
<td><strong>Outer-membrane proteins (OMPs)</strong></td>
</tr>
<tr>
<td><strong>Hydrogen bonds (h-bonds)</strong></td>
</tr>
<tr>
<td><strong>Cα carbon</strong></td>
</tr>
<tr>
<td><strong>Hydrophobic</strong></td>
</tr>
<tr>
<td><strong>Aromatic</strong></td>
</tr>
</tbody>
</table>
BamA barrel (β1 and β16 respectively). Scientists suggested that the exposed β1 strand could then act as a template for the formation of the substrate β-sheet. Here, β1 forms h-bonds to one end of the substrate chain, creating the substrate’s first β-strand. This strand can then h-bond to the next part of the chain to create the next strand and so on, until the substrate β-sheet is fully formed and can be closed off to bud into the membrane.

To test this prediction, a structure of BAM stalled during folding was obtained by engineering a covalent linkage between BamA in the intact BAM complex (BamA_{machine}) and a BamA substrate (BamA_{substrate}). Amazingly, when the structure was solved by cryoEM, an almost entirely folded copy of BamA_{substrate} was observed trapped on BamA_{machine} (Figure 4a), with the β-sheet almost complete and beginning to close into a complete barrel. Crucially, the β-sheet of...
BamA substrate was continuous with the β1 of BamA machine (Figure 4b), just as predicted by templating.

The BAM:BamA structure showed the templating hypothesis could occur; however, the structure seemed to represent a late stage in the folding process, just before BamA substrate is released into the membrane. To see how templating starts, structures of BAM with a single substrate β-strand bound were solved by crystallography (Figure 4c). Surprisingly, the lateral-gate was closed, but β1 was still able to bind to the first substrate β-strand (Figure 4d). It transpired that even with the gate closed, the last BamA β-strand only forms a few h-bonds to β1, leaving the rest free for binding substrate.

**A mechanism for BAM**

A key goal of structural biology is to reconstruct and understand how proteins move and change shape to carry out their designated function. For BAM, the structures above serve as snapshots throughout the process of folding an OMP and based on these we can propose a mechanism for exactly how this occurs.

Before the substrate arrives, BAM resembles our structures in Figures 2 and 3. The lateral-gate may be predominantly open, closed or dynamically switching between the two. When unfolded substrate arrives at BAM, the lateral-gate closes (if it was open), and the first substrate β-strand forms by binding to β1 of BamA, resulting in the structure in Figure 4c. At some point during formation, the lateral-gate must open, as our structure depicting the late stage of folding (Figure 4a) has an open lateral-gate conformation. This opening would then flip the partially folded barrel into the membrane where the two ends of the β-sheet close to complete the
substrate barrel. The fully folded substrate is then released into the membrane.

This mechanism is also supported by a surplus of non-structural data from other disciplines. For instance, it is possible to form cross-links between folding substrates and BamA’s β1 in living cells, providing evidence for substrate templating by β1 taking place in nature (see further reading 11). However, there are still many unknowns, for instance the starting state of the BAM lateral-gate, or when exactly the substrate is flipped into the membrane and how it is released. Yet precisely by pointing out gaps in current knowledge, such models are immensely useful to suggest future experiments and hypotheses to increase our understanding of a protein’s mechanism. To conclude, through the use of standard visualization approaches for structural data we can greatly enhance our understanding of the cellular machinery that underpins life.

Figure 4. BAM binds to folding substrates at the lateral-gate via β1. (a) Cartoon representation of a BAM:BamA substrate complex solved by cryoEM, stalled at a late stage of folding. The substrate is a second copy of BamA (BamA_{substrate}), which is almost completely folded. The copy of BamA within the BAM complex (BamA_{machine}) is lateral-open. (b) In this stalled complex, β1 of BamA_{machine} is joined to a β-strand from BamA_{substrate}, forming a continuous β-sheet with one unified h-bond network. (c) Cartoon representation of a crystal structure of the BAM complex stalled at an early stage of folding with a single β-strand of a substrate bound to β1. Unlike the late stage of folding, the lateral-gate of BamA is closed. (d) Close-up of the interface between the substrate β-strand and BamA β1, showing the h-bond network linking them into a continuous β-sheet.
Further Reading


Samuel F. Haysom is a postgraduate research student in the laboratories of Neil Ranson and Sheena Radford at the University of Leeds. His research focuses on unravelling the role of the BAM complex in outer-membrane protein biogenesis, principally using cryo-electron microscopy. Email bssfh@leeds.ac.uk