Understanding membrane protein association through molecular modeling and evolution

By

Samson Gerald Funakoshi Condon

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN-MADISON

2020

Date of final oral examination: 2020/05/01

This dissertation is approved by the following members of the Final Oral Committee:
Alessandro Senes, Associate Professor, Biochemistry
Ivan Rayment, Professor, Biochemistry
John Markley, Professor, Biochemistry
Bret Payseur, Professor, Genetics
Aaron Hoskins, Associate Professor, Biochemistry

Table of Contents

Acknowledgments

This work would not have been possible had I not had the good fortune of finding so many excellent mentors and friends who helped me on the way. I had long been fascinated by science in school but never considered a career in research. A long and pointed conversation with my best friends' parents right before I started college at Iowa State University set me on my path today by giving me the courage to pursue my interests. My undergraduate advisor Desi Gunning, research mentor Marna Yandeau-Nelson, and REU internship advisor Janice Pata (Wadsworth Center) introduced me to the laboratory, which was a lot less intimidating and a lot more exciting than I originally thought. They were formative in my decision to pursue a PhD.

I would like to thank my advisor Alessandro Senes for his constant mentorship and guidance. Alessandro has let me pursue creative directions for my projects and always provides encouragement and helpful advice whenever I get stuck in the weeds. One of his great strengths is explaining complicated topics very clearly and I hope I have taken his lessons to heart. Special thanks to members of the Senes lab both past and present for their great ideas and for being such amazing colleagues. Sabareesh and Ben helped me learn how to code and taught me plenty of fun tricks you can do on a terminal. Loren and Rika gave me lots of support when I needed it and introduced me to the world of bacterial cell division. Deena and Claire helped me out a ton when I started working on FtsLB and are always great to talk to. Kai and Beth were very generous with their experience and helped me learn a lot about protein purification and labeling. Samuel’s grasp of the literature is incredible and he is always down to play board games and EDH. He’s an excellent travel companion to boot! I could not have asked for a better bay-mate than Samantha (sorry for the clutter). Her perspective has always been useful and her empathy was a great

comfort to me when we were dealing with similar family tragedies. Gilbert and Josh are always great to bounce ideas off of and/or go deep into video game strategy. Finally, I can't wait to see where the research endeavors of our newest students Tyler and Samridhi (a fourth Sam for the Senes lab!) take them.

I have also been lucky to mentor a few undergraduate students on various computational and experimental projects, namely Praisten, Marley, Xiangyang, Yudong, and Raven. These students were fantastic and I hope they got as much from me as I got from them.

I am grateful to my collaborators, including Dave Pagliarini and Aaron Hoskins here at the University of Wisconsin-Madison, as well as with Jialing Lin at the University of Oklahoma, Jimmy Feix at the Medical College of Wisconsin, and Kevin MacKenzie at Baylor College of Medicine. These groups have expanded my scientific horizons and working with them has been a real pleasure.

A big thank you to the current and former members of my thesis committee: Ivan Rayment, John Markley, Bret Payseur, Aaron Hoskins, Brian Fox, and Julie Mitchell. Their probing questions and helpful suggestions have made me a better scientist. I have also been fortunate to be supported by various sources of funding, including the Computation and Informatics in Biology and Medicine Training Grant, the William R. and Dorothy E. Sullivan Wisconsin Distinguished Graduate Fellowship, and the Arthur B. Michael Departmental Fellowship.

The friends I've made in Madison have made life here better than I could have hoped. Plenty of bike rides, brunches, potlucks, camping trips, and long nights full of board games and good company have been a great distraction from the lab. In particular, I would like to thank Evgenia, Belal, Ian, Murshid, Willey, Stefanie, Vincent, and Jon. I would also like to thank all of my old

friends, some of whom I have known since elementary school. Jake, Jordan, and Dillon, your companionship has meant so much to me over the years. Ellen, thanks for sharing your adventures in cooking, it's always wonderful to get recipe ideas from you. Alana and Norah, it was so good to catch up with you when I was out west. I'll have to get out there more often! Joy and Ryan, I'm heartbroken that a pandemic prevented us from eating Hickory Park barbecue together at the Stupka Symposium, but it will just make next year's event even better.

My coworker, biking buddy, confidant and partner, Gladys Díaz Vázquez, deserves special mention. She has inspired me to be a better person and I am so grateful she has let me be a part of her life. Te amo mucho.

Finally, I would like to thank my family. My aunts, uncles, cousins, and grandparents have supported me immensely and have been pleasantly curious about what I do for a living. My parents David and Cindy always encouraged me to work hard and try my best in everything I do, and without their support I would not be where I am today. My mom taught me how to study and focus, and my dad taught me to embrace trying new things in the face of failure as well as how to effectively solve problems. I recently helped my dad fix my car and he quipped that it must be a lot different than the work I usually do. But the literature review, testing kit with controls, and the many hypotheses we came up with suggested that the only real difference was the weather. My little sister Ayoko has never once stopped believing in me and I will never stop believing in her. As soon as it's safe to travel I'm going home to give them all a big hug.

Abstract

The gap between known membrane protein sequences and solved membrane protein structures is quite large and only continues to grow. Predicting protein structures entirely from first principles remains challenging, but we can take advantage of the plethora of available sequences to identify evolutionarily coupled residues and infer the structures and functions of their evolutionary ancestors. In this thesis, I combine computational modeling techniques with evolutionary analysis as well as biochemical and biophysical experiments to better understand the structures and functions of membrane protein interactions.

The FtsLB complex plays an essential role in mediating bacterial cell division. However, structural and mechanistic understanding of its role is lacking. To understand the structural organization of the FtsLB complex, we demonstrated that the likely oligomeric state is a 2:2 heterotetramer. I then used restraints derived from evolutionary couplings analysis to identify the contacting residues in the FtsLB interface, which were used to computationally model its structure. Features of the predicted structure, including the transmembrane FtsLB interface, the continuous helix spanning the FtsL transmembrane and periplasmic domains, and the conserved polar and charged residues within the core of the FtsLB coiled coil were shown to be important for properly mediating cell division *in vivo*. These features give structural insight into how FtsLB plays an active role within the bacterial divisome.

I also combined *ab initio* modeling with ancestral sequence reconstruction to explore the energetic and sequence landscape of a membrane protein homodimer throughout its evolutionary history. BNIP3 dimerization is important for its activity, but the predicted interface is shifted in some homologues. How was the dimer interface shifted during evolution? Modeling BNIP3

homologues and inferred ancestors indicated that dimerization was indeed preserved and that the shift was mediated via the addition of a stabilizing interaction. Interfacial combinations of distant homologues indicate that the location of the crossing point within the membrane is an important determinant of dimer stability. These results contribute to our understanding of how individual transmembrane helices can associate so strongly and how such association may be preserved and changed through evolution.

Chapter 1: Introduction

1.1 Protein structure prediction

It has long been understood that protein structure and function are intimately linked, and structure-focused paradigms are quite common in biochemical investigations of proteins. However, many structure determining techniques such as X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy and cryogenic electron microscopy (cryo-EM) are costly and labor intensive. Additionally, many proteins have properties that make them difficult to characterize biochemically, such as transmembrane domains, multiple conformational states, or regions prone to aggregation. As such, only a small percentage of known protein sequences have had their structures experimentally determined, and the majority of protein structures have come from thermostable, water-soluble proteins, which are relatively easy to work with. Major advancements in genetic sequencing technology has led to an abundance of protein sequences without corresponding structures, and this gap is only growing.

A compelling alternative to determining protein structures experimentally is to predict their structures computationally, which can help close this gap and generate structure-based hypotheses on a much faster time scale. Predicting protein structures from their sequences is a long-standing challenge but recent advancements have made the task much less formidable. In this chapter, I summarize the problem of protein structure prediction and illustrate the many challenges that make it so difficult. I also describe methods to integrate various forms of low-resolution experimental data into protein structure prediction to improve models and guide the exploration of the conformational landscape. Finally, I discuss recent exciting developments that use explosion of sequence data as an asset to identify evolutionary constraints in protein families and infer their contacts and structures with very high accuracy.

1.1.1 Anfinsen's hypothesis

Predicting the structure that a string of amino acids folds into implies that there is a correspondence between the two. Anfinsen's hypothesis states that for a protein in its physiological environment, its native structure corresponds to the free energy minimum of its conformational landscape (Anfinsen, 1973). Put another way, a protein sequence contains all of the information necessary for it to fold into its final structure. Naturally, there are exceptions—some proteins require chaperones to properly fold (Kerner et al., 2005), and there are other proteins whose active state is kinetically trapped in a thermodynamically less favorable conformation (Baker et al., 1992). Additionally, intrinsically disordered proteins have no “native structure” in physiological conditions (Dunker et al., 2001). Nonetheless, to a first approximation Anfinsen's hypothesis appears to be correct. This statement is of profound importance for the field of protein structure prediction: if we have a sufficient physical and mathematical understanding of protein thermodynamics, we would have an equation whose solution would correspond to the native structure of a protein. In other words, it should be possible to predict the structure of a protein using only its sequence as input.

The mathematical functions used in protein structure prediction to evaluate the potential energy of a particular protein conformation are computationally inexpensive approximations of atomic interactions known as force fields. These force fields correspond to terms which capture various features of covalently bonded atoms (to preserve ideal bond, angle, and torsional geometries) as well as longer-range interactions between nonbonded atoms (such as van der Waals and electrostatic forces). Additional terms to implicitly model the protein’s interaction with its solvent are frequently included (Lazaridis, 2003; Lazaridis and Karplus, 1999; Weinstein et al., 2019). The function may also include statistical or knowledge-based potentials

that are derived from analyses of known structures and may or may not have a direct physical interpretation (Fleishman and Ben-Tal, 2002; Miyazawa and Jernigan, 1985; Zhang and Liu, 2006). Computing the potential energy of a particular protein conformation with these force fields is a simple summation of all of these various terms for each relevant group of atoms. By limiting each nonbonded interaction and most bonded interactions to one or two atoms, the total number of interactions is tractable and evaluation of the total score is relatively rapid.

Unfortunately, even if these force fields were perfectly able to predict the stability of a particular protein conformation, we would still be unable to conclusively identify the most stable structure a sequence could form. There are two major obstacles, the first of which is the extraordinarily large conformational landscape. For example, consider a relatively small protein with 100 amino acids. There are 99 peptide bonds in this protein, which we will assume are fixed and immobile. Predicting the backbone structure of the protein can then be defined as correctly predicting the 198 φ and ψ dihedral angles for the polypeptide. If we generously assume that each φ and ψ angle can only take on one of three possible values, then we already have 3^{198} combinations to sample, an astronomically large number. The problem is further compounded by the second obstacle, which is that the energy function is extremely rugged and nonconvex. Even slightly different conformations can have extremely different stabilities. The native structure may be highly optimal, but a nearly identical conformation where two nonbonded atoms overlap is so unstable as to be considered forbidden. These challenges place protein structure prediction in a class of computational problems designated as NP-Hard (Pierce and Winfree, 2002). Put simply, this means protein structure prediction is a problem where evaluating the stability of a particular backbone conformation is trivial, but a fast algorithm to conclusively identify the most stable conformation does not exist. Levinthal noted this issue in the context of protein folding

(Levinthal, 1969). Even if one combination was sampled every nanosecond, the protein would not fold into its native state within the age of the universe. Nonetheless, protein folding rates generally range from microseconds to seconds (Kubelka et al., 2004; Zou et al., 2014). With such a large conformational landscape, how do proteins reliably fold to their native states so quickly?

Solutions to Levinthal's paradox are highly important for protein structure prediction, as they help guide navigation of the rugged conformational landscape. First, since proteins fold quickly, they must not need to exhaustively sample every conformation. Second, since proteins can fold into their native structure from an ensemble of disordered conformational states, there must be many different pathways that lead to the optimal solution. Finally, proteins first fold via local interactions to form secondary structure elements before the tertiary contacts form. This means that proteins fold from their unfolded state by sampling nearby conformations. Unlike a safe, where there is one correct combination and every other combination is equally incorrect, partially folded intermediates are more stable than completely unfolded ones and the optimal solution can be reached by iteratively exploring the local landscape.

1.1.2 Navigating rugged energy landscapes

While no algorithm exists to solve NP-Hard problems, heuristic solutions have been developed that can efficiently find solutions useful enough for practical purposes. By analogy, our task is similar to finding the deepest valley in a large mountain range while blindfolded. Safety concerns aside, there are a number of strategies available to solve this problem. A brute-force strategy where every inch of the landscape is measured will identify the deepest point at the cost of much wasted time and energy. Alternately, you may try to start at a random location in the range and always move in the steepest direction downhill. This strategy (also known as gradient descent) will converge at a local minimum, but there is no guarantee that it is the lowest point in

the landscape. Once this local minimum is found, the process can be repeated from a different starting point to find a new minimum. Trials can be iterated until an acceptable solution is identified, but since you can get stuck in even very shallow divots the process is quite inefficient.

An improvement on this strategy is to iteratively make random movements that go downhill *most of the time*, but occasionally choose a direction that goes uphill. In this way, you will still seek the minimum but be less likely to get trapped in shallow local minima. By sometimes hiking over hills and climbing over boulders in the range, you can more effectively explore the landscape in pursuit of the lowest valley. Additionally, as deeper and deeper points are identified over the course of the hike, you may want to more fully explore the local region by walking uphill less often. This strategy, known as simulated annealing (Kirkpatrick et al., 1983), belongs to a class of methods known as Markov Chain Monte Carlo (MCMC) methods (Hansmann and Okamoto, 1999). Markov Chains are probabilistic explorations of states where transitions between states are mediated with certain probabilities. The Monte Carlo name designation simply refers to the fact that we are using random sampling of the energy landscape to find an optimal solution. Trial moves away from the current state are evaluated and accepted according to a scoring criterion. Often, the decision of whether to accept a transition or not is based on the Metropolis-Hastings criterion (Hastings, 1970; Przytycka, 2004):

$$p_{accept} = min\left(1, e^{\frac{\Delta E}{RT}}\right) \qquad \text{Equation (1)}$$

Where ΔE is the difference in energy score between the previous state and the current state, R is the gas constant, and T is the temperature factor of the simulation. If the solution is an improvement over the previous state, it is always accepted. Otherwise, the probability of

acceptance is based on the Boltzmann distribution. A common strategy is to start with a high temperature factor which will lead to more frequent acceptance of poor transitions and a more broad traversal of the energy landscape. As the simulation continues the temperature factor is decreased to ensure better sampling of the local minimum.

Given infinite time, MCMC methods are guaranteed to find the optimal solution within the landscape, whether the task is protein structure prediction or blindfolded hiking. However, the same is true of a method in which conformations are sampled completely at random or the entire space is exhaustively examined. MCMC methods are generally much more efficient than complete randomness or brute-forcing, but are still liable to get trapped in local minima. Thus, it is important to run many of these simulations with different initial conditions such that the entire space is adequately covered. Then, the best-scoring solutions can be compared to one another to get a sense of the global or near-global minimum. If multiple trials points have converged on a particular conformation, it lends more confidence that the model is correct. This can be visualized in funnel plots, where a particular model's energy score is plotted against the similarity of that model to either the ground truth (i.e. the known native structure) or the best model if no ground truth is available (Figure 1.1). Common similarity measures include the root mean square deviation (RMSD) between atoms after alignment of the models (Coutsias et al., 2004), fraction of common contacts (Rodrigues et al., 2012), and the template-modeling score (TM-score) (Zhang and Skolnick, 2004). If sampling has been insufficient, there will be no correlation between the energy of a particular solution and its similarity to the best conformation. Conversely, a funnel of well-scoring models near the ground truth with more divergent models scoring poorly suggests that the many trials have converged on a good solution.

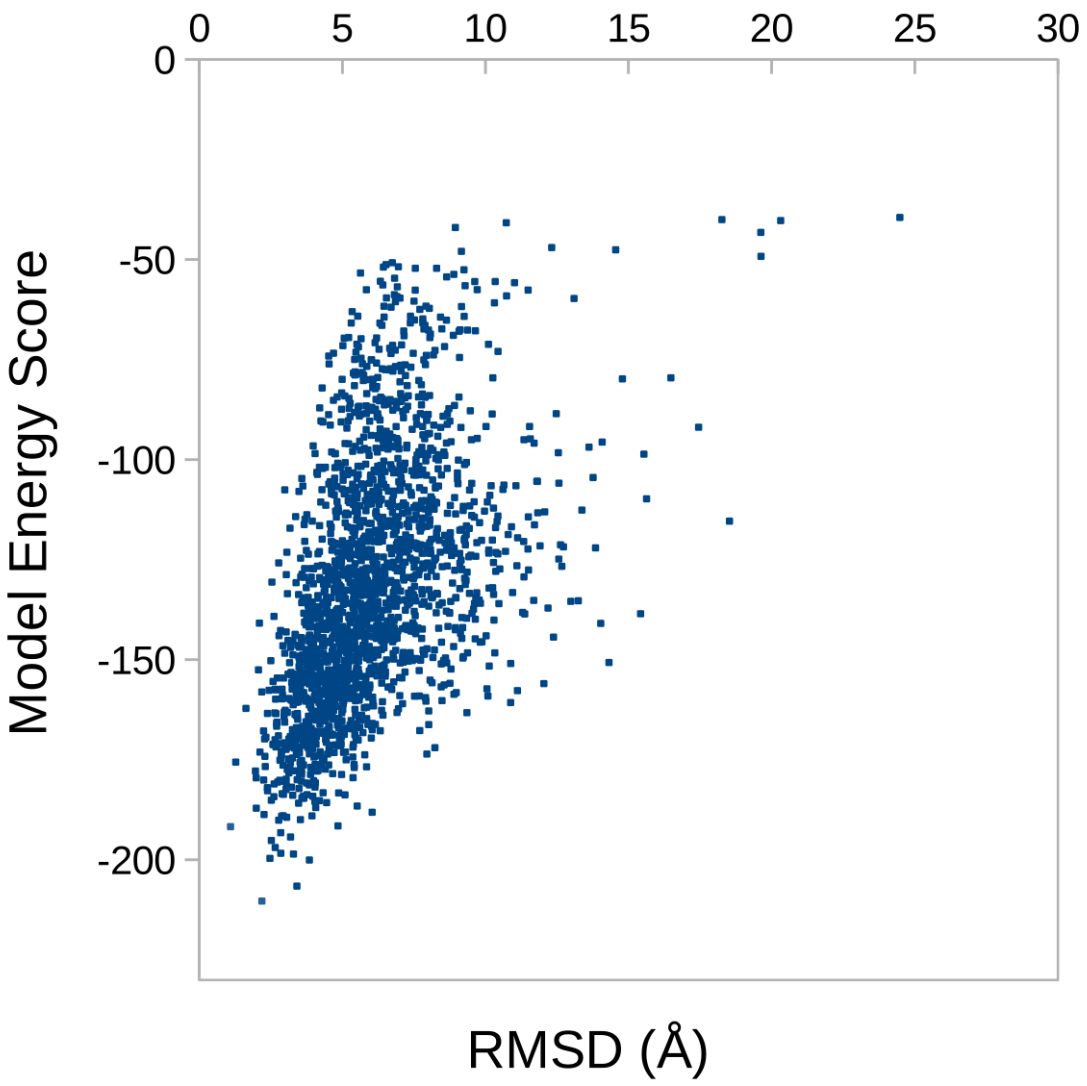

Figure 1.1: Example of a funnel plot used to assess convergence in protein structure prediction. Each model generated from separate trials is aligned to the ground truth and the deviation is plotted against its energy score.

1.1.3 Movements to navigate energy landscapes

In the hiking analogy, a movement is simply a step in a particular direction. The possible movements in the conformational landscape of protein structure prediction can more precisely be

defined as the degrees of freedom that are optimized according to the scoring function. Each degree of freedom is an independent parameter that has to be optimized in the context of all the others. In the most extreme case, a protein's structure could be modeled by optimizing the Cartesian coordinates of each individual atom separately and independently of all the others. Such a strategy would be extremely computationally expensive and many simplifications can be made based on reasonable assumptions. One example of these simplifications is restricting the lengths and angles of atomic bonds to fixed canonical values. While this idealization is not completely correct, particularly in the case of bond angles (Berkholz et al., 2009), fixing bonds in this way dramatically reduces the number of degrees of freedom with relatively little cost. The conformation of the protein backbone is then largely defined by the φ and ψ torsional angles between peptide bonds (Ramachandran et al., 1963). Of course, proteins are much more than backbones. For the specific problem of side chain optimization, the side chain atoms of residues in a protein are each usually restricted to a few discrete states known as “rotamers” (in which the bond lengths and angles are fixed to optimal values while the torsion angles vary) or “conformers” (discretized side chain conformations from high-resolution structures that do not have restricted bond geometries) (Subramaniam and Senes, 2012). This simplification from a continuous set to a discrete space dramatically reduces the degrees of freedom that need to be sampled.

Another attractive simplification is coarse-graining. Instead of including all of the atoms in the protein, groups of atoms are combined. For example, the methyl side chain of alanine could be reduced to one larger group, collapsing four atoms into one. Likewise, other, larger side chains can be grouped into one or a few centroids. This simplification is frequently used in molecular dynamics simulations where inclusion of all the atoms including water and lipids

would make long trajectories intractable (Monticelli et al., 2008; Neria et al., 1996). However, it also finds much use in protein structure prediction (Kmiecik et al., 2016). Coarse-graining the problem in this way can more rapidly sample the landscape and remove obviously bad conformations. Subsequently, the coarse-grained solutions can be refined to incorporate all the atoms (Badaczewska-Dawid et al., 2020).

1.2 Combining *ab initio* force fields with empirical data

Computing the structure of a protein entirely from first principles, or *ab initio*, is still a large technical challenge. Instead, researchers typically integrate *ab initio* methods with sources of biochemical, biophysical, and bioinformatic data relevant to their proteins of interest. These external data can be used as a compass to guide sampling in the conformational landscape, excluding broad swaths of unlikely or conflicting structures and providing important clues towards the native state. Of course, the *ab initio* energy functions themselves are parameterized such that they reproduce experimental or high-quality theoretical data. For example, the energy functions in the molecular modeling suite Rosetta have been parameterized to recaptiulate hydrogen bond distances in known structures and accurately reproduce known thermodynamic properties of small molecules (Park et al., 2016). The forcefields utilized in the molecular dynamics program CHARMM are routinely updated to better conform to quantum mechanics calculations on model systems and known properties of macromolecules (Mackerell, 2004; Zhu et al., 2012). The molecular modeling program SCWRL4 was parameterized to accurately model side chain conformation from electron density maps of protein crystal structures (Krivov et al., 2009). As such, the distinction between *ab initio* protein structure prediction and structure prediction guided by experimental data is not as clear cut. Nonetheless, it is useful to distinguish

methods and data that are useful for protein structure prediction *in general* versus experimental data that researchers collect that is *specific* to their protein of interest.

The sources of empirical data useful for protein structure prediction are myriad. They can include crosslinking experiments, comparisons with homologous proteins, secondary structure predictions from amino acid propensities, and analyses of conserved and coevolving residues. Distance constraints can be derived from crosslinking studies, NMR, Electron Paramagnetic Resonance (EPR), and various bioinformatic analyses. Methods like H/D exchange and radical footprinting can determine solvent-exposed portions of the protein, both of which can be used to help determine whether a protein model is accurate (Aprahamian et al., 2018; Liu et al., 2012). Empirical methods to solve the structures of proteins like Cryo-EM, X-ray crystallography, and NMR spectroscopy can be thought of as particularly rich sources of data, where the methods used to build the models are primarily guided by the observations, with *ab initio* force fields playing a more supporting role. These methods and their applications to protein structure prediction are discussed below.

1.2.1 Homology modeling

If the structure of a protein is unknown, the next best thing is the structure of a closely related one. Homology modeling is a technique in which a previously determined or modeled structure of a protein is used as a template to predict the structure of a homologous protein of unknown structure. Having such a template available is quite valuable, as homologous proteins are highly likely to share the same core structure (Illergård et al., 2009). The shared fold may lend insight into a common evolutionary history or regulatory mechanism. Homology models can also be compared with the templates to identify differences which may lead to variations in function (Lukk et al., 2012).

The first step in homology modeling is to obtain a suitable template. This can be as simple as performing sequence homology searches on the database of solved protein structures, which works well if the target protein is very similar to one that has been solved. However, when the target protein is more distantly related and does not have a high level of sequence identity to known structures, it can become difficult to identify and align a correct template, if one exists at all. Improvements in this case can come from better template identification strategies such as the use of profile-based queries (Lam et al., 2017) or the use of multiple distant templates. The use of multiple independent templates can be very advantageous, particularly if the templates cover different regions of the target sequence (Larsson et al., 2008). Inclusion of multiple templates can also increase the robustness to the inclusion of divergent templates (Meier and Söding, 2015).

Once a suitable template or set of templates has been selected and aligned to the target sequence, structural models of the target are generated. A variety of algorithms are available for generating these models based on the target and templates. These vary in the level of fidelity to the template, from rigid-body extraction of matching pieces from the template (Schwede, 2003) to generation of softer spatial restraints that the program attempts to satisfy (Šali and Blundell, 1993).

Many subtleties complicate homology modeling, most of which unsurprisingly surround situations where the template and target do not match. These may include insertion-deletion events (indels), divergent loops, and challenges that arise from using a template with low homology to the target. Loop sequences are highly variable and typically poorly conserved even among homologues, so the template may provide little useful information in these segments. A survey of homologous structures with insertion or deletion events (indels) indicated that the overall structural impact of the indel tended to be low, particularly within loop regions (Kim and

Guo, 2010). However, if the indel occurs within a secondary structural element, loops upstream or downstream of the indel were altered to preserve the core structure, potentially meaning the homology model template would be less useful for those regions. Finally, while significant advances have been made both in detecting homology (Chen et al., 2018), using distant templates for modeling (Yang et al., 2015), and structural genomic initiatives (Schwede, 2013), many proteins and protein complexes still lack an acceptable template for homology modeling.

1.2.2 The importance of secondary structure restraints

Predictions and measurements of protein secondary structure at the residue level are an additional, extremely useful, source of information to restrict the conformational landscape by dramatically narrowing the range of the backbone dihedral angles to sample. Secondary structural elements such as α helices and β strands have very well defined φ and ψ backbone torsional conformations for a stretch of residues. Instead of needing to sample all regions of allowed torsional space for each residue, a much narrower set of conformations for the entire length of the element can be explored. Secondary structure restraints can be incorporated using rigid, idealized geometries for the elements (Grigoryan and DeGrado, 2011; Karakaş et al., 2012; Mueller et al., 2014) or as restraints (Brunger, 2007; Raval et al., 2016) that penalize but can still tolerate deviations from the assigned states.

A variety of tools have been developed to assign secondary structure to residues in a protein sequence. NMR is an effective empirical tool to obtain residue-level information on the protein's secondary structure. Backbone chemical shift indices for ^{1}Hα and ^{13}Cα have a tendency to vary based on their secondary structure. ^{1}Hα chemical shifts for backbone atoms within alpha helices are shifted upfield relative to a random coil while those within beta strands are shifted down,

while $^{13}C\alpha$ shifts have the opposite response. The relative changes versus random coil orientations can be displayed on a graph to identify stretches of helical and stranded regions (Wishart et al., 1992) and this process has been automated to classify secondary structure with high accuracy (Hafsa et al., 2015). NMR chemical shifts can also be used to predict the backbone and even side chain χ1 dihedral angles (Shen and Bax, 2013). Patterns of cross-peaks in Nuclear Overhauser Effect Spectroscopy (NOESY) can also be used to identify secondary structure elements (Clore and Gronenborn, 1989). Additionally, many computational algorithms have been developed to predict protein secondary structure from sequence, with tools such as PSIPRED (Buchan and Jones, 2019) and DeepCNF (Wang et al., 2016) approaching theoretical limits of accuracy (Yang et al., 2016). Other computational tools are available to predict more specific structural elements, such as coiled-coils (Ludwiczak et al., 2019; McDonnell et al., 2006), transmembrane segments (Petersen et al., 2011; Sonnhammer et al., 1998; Wimley and White, 1996), and intrinsically disordered regions (Dosztányi, 2018). Prediction of transmembrane segments in particular is quite useful, as the cell membrane forces a particular topology for the protein to take. Trasnmembrane helices are embedded within a relatively thin lipid bilayer, and the tilt angles of transmembrane helices relative to the membrane normal are relatively restricted (Weinstein et al., 2019). Furthermore, sequential transmembrane domains within a protein have antiparallel topologies. These properties all limit the conformational space that is necessary to explore for membrane protein structure prediction, though the lack of data to parameterize membrane protein scoring functions remains a hinderance (Koehler Leman and Bonneau, 2018).

1.2.3 Fragment assembly

Restraining the conformational space by secondary structure elements can be further generalized using fragments of previously solved structures (Das and Baker, 2008; Rohl et al.,

2004). A stretch of amino acids with a particular sequence is likely to adopt a similar conformation across a variety of protein structures (Bowie and Eisenberg, 1994; Simons et al., 1997). Based on this intuition, fragment assembly algorithms such as Rosetta identify a set of backbone conformations that each n-mer of residues (typically 3-mers and 9-mers) within the given protein sequence form in known high-resolution structures. In many ways this is similar to homology modeling, except these “templates” are typically unrelated to the target sequence by an evolutionary mechanism. These fragment libraries are then combinatorially explored using a MCMC algorithm as described above to generate candidate structures. For better navigation, this phase of the protocol employs a smoother, low-resolution scoring function that replaces side-chain atoms with single-body centroids which do not severely penalize small clashes (Rohl et al., 2004). After the low-resolution protocol generates a large set of structures with local minima, an all-atom refinement procedure takes place where the best models from the low-resolution protocol are further refined to generate the final ensemble of structures.

Similarly, statistical analyses of structures in the Protein Data Bank are a powerful tool to optimize the possible orientations of protein side chains. Since side chains contain many rotatable bonds, there are a staggering number of degrees of freedom that would need to be optimized to predict their conformations. Fortunately, in crystal structures of proteins only a limited set of possible orientations are commonly observed (Dunbrack, 2002). This allows the continuous space of side chain conformation to be discretized into a much more tractable set of orientations per residue. Optimization can be further improved by incorporating the frequency of observing a particular conformation in crystal structures (Shapovalov and Dunbrack, 2011). Another strategy to improve sampling uses a library of rotamers that have been pre-ranked according to the number of structural environments those side chains can be placed in to without

significant steric occlusion (Subramaniam and Senes, 2012). In this way, rotamers can be chosen such that side-chain optimization can rely on the top few performers rather than exhaustively searching the space of potential side chain conformations and combinations and sampling can be tuned for different residues and levels of granularity.

1.2.4 Shape and size restraints

Structure prediction space can also be restrained through the shape or size of a protein. Multiple experimental approaches exist to measure general size and shape features of the protein of interest. For example, small-angle X-ray scattering (SAXS) provides information about the protein's overall size and shape, including estimations of oligomeric state. Low-resolution, 3D models of proteins can be fit to SAXS data to obtain a reliable estimate of the protein's overall shape in solution (Petoukhov and Svergun, 2013). Typically, this modeling step uses coarse-grained beads, modeling them into a compact, interconnected shape that generates a theoretical scattering profile similar to the experimental data (Franke and Svergun, 2009). For protein complexes where the structures of the subunits are known, rigid-body docking can be guided by the SAXS data (Petoukhov and Svergun, 2013). However, it is also possible to directly incorporate SAXS data into conformational sampling during docking (Schindler et al., 2016) and protein structure prediction (Förster et al., 2008; Zheng and Tekpinar, 2014). However, fitting the protein structure to SAXS data can be complicated by multiple oligomeric states in the protein sample and intrinsically disordered regions. Both of these phenomena can distort the SAXS envelope and lead to worse models (Ogorzalek et al., 2018).

Low resolution maps from cryogenic electron microscopy (cryo-EM) can also be used to bias conformational sampling algorithms in a manner analagous to SAXS, penalizing candidate models that do not conform to the density (DiMaio et al., 2009). The resolution of negative-stain

EM is considerably lower than cryo-EM, but the density map has also been successfully used to dock subunits of protein complexes (Gallagher et al., 2019).

These shape and size restraints are extremely useful for modeling large macromolecular complexes with data of varying quality. Integrative modeling is a procedure in which experimental data are combined with known structures of monomeric proteins to build reasonable models of large macromolecular complexes. A popular tool for modeling large macromolecular complexes with the help of empirical data is the Integrative Modeling Platform (Webb et al., 2018). Data pertinent to the protein complex are first collected, including any available protein structures or models, as well as diverse experimental data that can be converted into spatial restraints. The restraints and data are converted into a scoring function and the proteins are modeled at different resolutions according to the available data. Conformations are sampled through a MCMC procedure and the data are finally aggregated and analyzed to select high-quality models. One advantage of the Integrative Modeling Platform is that it can simultaneously model portions of the complex at multiple different resolutions. For example, if one subunit has a known, high-resolution structure, but another only has descriptions for size and shape, the former can be modeled using all atoms while the latter can be represented with a string of beads that is restrained to have a particular shape. Of course, limits in the amount of known data place limitations on the efficacy of integrative modeling, so it will be of little use if only the sequence of the interacting proteins is known.

1.2.5 Distance restraints help to constrain sampling of tertiary and quaternary contacts

Information about residue-residue contacts within the protein derived from biochemical or bioinformatic data is typically integrated into protein structure prediction in the form of a

parameterized interaction within the energy scoring function (Bender et al., 2016; Brunger, 2007; Li et al., 2004). For example, a spring or sigmoidal function can be used between two residues of interest so that predicted structures where the two residues are far apart have an energy penalty or a bonus when they are closer together. These distance restraints are effective in helping guide the random walk to the global minimum, provided that the researcher is confident of the extra information.

Distance restraints work very well to constrain the search space when both of the interacting residues are known, but many experiments only report an interaction of one of the residues in the pair. For example, Chemical Shift Perturbation mapping is an NMR technique that can identify residues of an isotopically labeled protein that undergo a change in environment upon titration with an unlabeled ligand (Gao et al., 2004). The technique can identify residues that may form the binding site of the labeled protein, but no information is collected for the ligand interface. Likewise, mutagenesis experiments may reveal sensitive residues in a protein-protein interface, but the other residues participating in that interface are unknown. These data can still be very informative for protein structure prediction but a more sophisticated approach than pairwise distance restraints is required. The docking program HADDOCK incorporates this information by using ambiguous interaction restraints (AIRs), which calculate an effective distance measured between an “active” residue (that is, a residue presumed to lie at the interface) on protein A and the set of active and “passive” residues (surface-exposed residues that may or may not be within the interface) on protein B (van Zundert et al., 2016). The effective distance is always at least as short as the shortest distance in the set and can be shorter. This effective distance is then converted to a pseudo-energetic term which is added to the overall scoring function. Incorporating the ambiguity in this way ensures that the sensitive residues identified from

experiments are guided towards the interface without forcing the researcher to guess the location of the opposite binding site.

1.3 Protein structures from multiple sequence alignments

1.3.1 Molecular coevolution

One of the most exciting breakthroughs in the field of protein structure prediction has been the development of methods to infer protein tertiary contacts and structure from multiple sequence alignments. Natural selection has generated many diverse, functional sequences from a relative handful of folds (Schaeffer and Daggett, 2011), and these folds have placed certain constraints on sequence evolution. By disentangling how natural sequences evolve it is now possible to infer these constraints and predict the fold.

Coevolution is the selection of a trait in response to another trait. This phenomena is often thought of in the context of species-species interactions such as pollinators and flowers or predator-prey relationships. However, it can also occur at smaller evolutionary scales like between males and females of the same species, different genes within an organism, or different parts of a protein. Coevolution within a protein sequence intuitively might occur between physically interacting positions (Figure 1.2). For example, if a mutation occurs that changes an aspartate-lysine salt bridge to a potentially unfavorable aspartate-leucine interaction, certain additional mutations might be favored and selected for over time. A simple reversion back to aspartate-lysine could occur, or the leucine might be mutated to another residue like arginine to restore the salt bridge. Alternately, mutations from aspartate to a more hydrophobic residue like isoleucine would introduce a favorable hydrophobic interaction. This process would repeat throughout the protein's evolutionary history, and an alignment of these homologues would show

that mutations in these two positions are highly correlated with one another. We might then ask the inverse problem: if highly correlated pairs of positions are observed in a protein family's multiple sequence alignment, do those positions form tertiary contacts in the protein family's native structure?

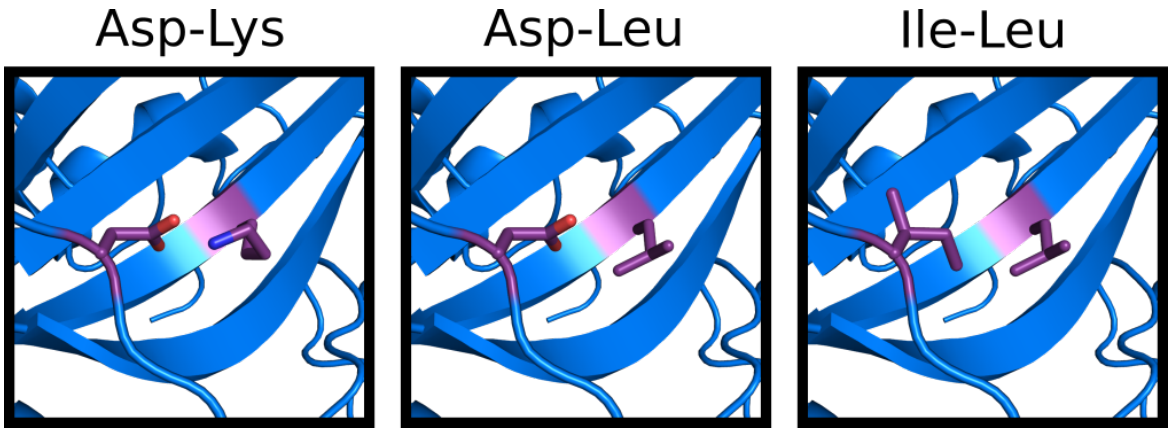

Figure 1.2: Toy example of how coevolution might occur in protein tertiary interactions. In this example, an Asp-Lys salt bridge (purple) is destabilized when the lysine is randomly mutated to leucine. Compensatory mutations such as Asp→Ile might be favored during the course of natural selection. Mutations generated in PyMOL *on* the structure of HcaII (PDB ID: 2VVB).

This inverse problem is simple to describe and its solution has been pursued for years, but many difficulties led to relatively little success (Clarke, 1995; Dunn et al., 2007; Korber et al., 1993; Lapedes et al., 1998). Simple examination of aligned sets of sequences for correlated pairs of columns has its pitfalls. A common technique is to calculate the mutual information (MI) for column *i* and column *j* in an MSA, which is a measure of how correlated the two are:

$$MI_{ij} = \sum_{\sigma,\omega} f_{ij}(\sigma,\omega) \ln\left(\frac{f_{ij}(\sigma,\omega)}{f_i(\sigma) f_j(\omega)}\right) \qquad \text{Equation (2)}$$

Where f describes the frequencies of amino acids σ and ω within each column and the joint frequency of observing both σ in column *c* and ω in column *d*. This score is calculated for each

pair of columns in the MSA and the pairs of columns with the highest MI are assumed to be the most coevolving. Unfortunately, this technique is of somewhat limited use for two reasons. The first is that the protein's phylogenetic history is ignored, and highly conserved columns will tend to have high MI simply because they happen to share the same history. The second reason is that indirect correlations can have a profound impact on the MI due to transitivity. Imagine a network of residues where position A interacts with position B, and position B interacts with both position A and C. Position C is far from position A, but because the A-B pair and the B-C pair are correlated, the A-C pair would also have an indirect, transitive correlation. Indeed, this pattern can continue in the network, and some indirect correlations stemming from this problem can appear stronger than direct ones (Burger and van Nimwegen, 2010; Lapedes et al., 1998; Marks et al., 2011).

1.3.2 Maximum entropy models of protein multiple sequence alignments

In contrast to MI methods that examine each pair of positions independently from one another, maximum-entropy methods simultaneously model the entire set of correlations between all pairs of positions (Lapedes et al., 1998). By performing this global fit to the MSA, the problem of indirect correlations stemming from multiple weak correlations from distant residues is largely eliminated. This dramatically improves prediction accuracy in terms of identifying close contacts in proteins (Kamisetty et al., 2013; Marks et al., 2011; Monastyrskyy et al., 2016).

Briefly, these maximum-entropy models of evolutionary couplings accept the MSA of a protein family as input and then identify a probability distribution that models the frequencies of each amino acid in each column, subject to the constraints that the frequencies in the model for each column and each pair of columns match those observed in the MSA. Unfortunately, there

are infinitely many probability distributions that would be consistent with the computed frequencies. Given these constraints, we can optimize another function to identify a unique probability distribution with desired properties. For example, we could maximize the likelihood of the data, or add a regularization term to impose smoothness or sparseness. However, there are several desirable properties that arise if we identify the probability distribution with the maximum information entropy:

$$\textit{Maximize} \quad S = -\sum_{A_i} P(A_1, \ldots, A_L) \ln P(A_1, \ldots, A_L) \qquad \text{Equation (3)}$$

where $P(A_1, \ldots, A_L)$ is the probability of an amino acid sequence with L residues to be a member of the protein family in the multiple sequence alignment. By maximizing the entropy, we identify the model with the fewest features imposed by the model itself. Choosing the maximum-entropy distribution is useful to prevent overfitting, as this distribution adds the least amount of additional external information. The resulting global model of the sequence alignment takes the following form:

$$P(A_1 \ldots A_L) = \frac{1}{Z} \exp\left\{ \sum_{1 \le i \le j \le L} e_{ij}(A_i, A_j) + \sum_{1 \le i \le L} h_i(A_i) \right\} \qquad \text{Equation (4)}$$

where A_i and A_j are specific amino acids in columns *i* and *j* of the MSA, Z is a normalization constant to make the probability distribution sum to 1, and h_i and e_{ij} are Lagrange multipliers constraining the probability distribution to the observed individual frequencies of amino acids in column *i* and pairwise frequencies of amino acids in columns *i* and *j*. The matrix of parameters in e_{ij} describe the direct coupling of each pair of possible residues in columns *i* and *j*, and this information can be summarized to quantify the likelihood of a direct interaction between them.

Unfortunately, explicitly determining each of the parameters in this model is computationally prohibitive. This presents a challenge similar in scale to protein structure prediction, where any particular solution can be calculated efficiently, but identifying the global optimum is extremely challenging. Fortunately, approximate solutions produce reasonable results. Multiple approaches to approximate a solution include the use of MCMC methods to stochastically search the probability distribution space (Lapedes et al., 2012), as well as approximation methods such as belief propagation (Weigt et al., 2009), mean-field approximation (Marks et al., 2011) and pseudolikelihood maximization (Ekeberg et al., 2013; Kamisetty et al., 2013).

1.3.3 Use of evolutionary couplings analysis in protein structure prediction

The vast majority of residue pairs with significant evolutionary couplings tend to be relatively short-range contacts in which the coupled residues physically interact (Anishchenko et al., 2017). The most strongly coupled pairs of positions also tend to be distributed throughout the protein instead of localized in one particular domain, as can be the case for local methods like MI (Marks et al., 2011). As such, these distant contacts can be used to guide structure prediction algorithms with excellent results (Kamisetty et al., 2013; Marks et al., 2011; Monastyrskyy et al., 2016). Protein families considered to be refractory to structural characterization, including membrane proteins and large proteins, can have residue-residue coupling signals strong enough for accurate contact map and structural prediction (Hopf et al., 2012). This method can also be extended to protein-protein contacts as long as orthologous pairs of proteins are identified (Hopf et al., 2014; Ovchinnikov et al., 2014).

As most covariation algorithms identify a set of pairs of positions that are in physical contact with one another, it is fairly straightforward to incorporate them in the form of distance restraints

between the two positions. Typically, these restraints take the form of distance restraints between the Cα or Cβ atoms of each residue. The equations used to incorporate distance restraints into the minimization calculations can take a variety of forms, including a “soft” harmonic potential (Wang and Barth, 2015) “soft-square” potential (van Ingen and Bonvin, 2014; Marks et al., 2011), or a sigmoidal function (Braun et al., 2015; Ovchinnikov et al., 2014). The sigmoidal function is particularly useful for cases of high ambiguity, as extreme outliers that are inconsistent with the rest of the data do not strongly penalize the overall score. There is additional nuance to be considered when choosing the parameters of the distance restraint. GREMLIN, for example, explicitly considers the wild-type residues in the pair and adjusts the optimal distance accordingly (Ovchinnikov et al., 2014). Large, bulky residues are given a longer optimal distance than shorter residues, which keeps the calculation relatively simple while making sure the two residues are approximately contacting one another without clashes.

An exciting new development has been the use of deep learning techniques to infer residue-residue distances instead of binary contact predictions from a protein family's multiple sequence alignment (Senior et al., 2020; Xu, 2019). For AlphaFold, a statistical potential is learned for a specific protein and its associated MSA. The network was trained to predict residue-residue distances and torsion angles solely from MSAs of proteins with known structure. By predicting these distances simultaneously, features of the MSA including covariation and local structure are preserved. From these distance distributions, a novel algorithm was used to fold the protein. Instead of adding distance potentials to an energetic forcefield, AlphaFold instead constructs a target-specific statistical potential. This statistical potential is an equally weighted sum of distance and torsion distributions inferred from the deep learning with an added van der Waals term. Surprisingly, this protein-specific potential is extremely smooth with a minimum that

aligns with the native structure, meaning it can be trivially searched using gradient descent to identify the optimal solution (Senior et al., 2020). This strategy lies in stark contrast with other approaches to protein structure prediction, which typically rely on expensive Monte Carlo sampling to navigate a highly rugged, universal energy landscape. Nevertheless, deep learning to predict residue-residue distances followed by simple minimization vastly outperformed more traditional approaches in the CASP13 free modeling competition (Kryshtafovych et al., 2019).

While evolutionary methods of identifying residue-residue contacts are extremely powerful, they are not without limits. The major drawback of maximum-entropy models of coevolution is the number of sequences they require. Typically, good predictions are achieved when there is at least one nonredundant sequence for every column in the multiple sequence alignment (Marks et al., 2011), though performance tends to increase with larger numbers of sequences. It is not unusual for hundreds or thousands of sequences to be required. Fortunately, there has been such an explosion of sequencing data in the past decade that for many protein families this is not an issue. A promising area of research is using functional sequences generated from experimental evolution as input for the evolutionary couplings analysis, which may be useful to study proteins with poor conservation and few homologues (Stiffler et al., 2020). Another major problem is that the contacts inferred from multiple sequence alignments invariably correspond to average results across sequences and structures in the MSA. Structures inferred from these contacts may be quite adequate models of the protein family but the subtleties for individual sequences are lost. Additionally, structural diversity within the MSA may lead to “false positive” contacts that are nonetheless correct for a subset of the proteins (Anishchenko et al., 2017). To counteract this, additional distance restraints for the protein of interest derived from NMR or other experiments may be used to further refine the structure (Tang et al., 2015).

A problem specific to inter-protein evolution is the requirement for proper pairing of cognate pairs of interacting proteins, also known as interologs. If the proteins of interest have undergone a duplication event throughout their evolutionary history, it is possible that the paralogs of one protein will only interact with their cognate partner instead of the other paralogs. This can be a major issue if it is unknown which paralogs interact with each other in a species, as mis-matching proteins that presumably interact with one another will dilute the covariation signal (Hopf et al., 2014; Ovchinnikov et al., 2014). In prokaryotic species, assignment of interologs can be inferred based on genetic distance. Interacting pairs of genes are typically lie close together in the genome (Yellaboina et al., 2007). As such, simply picking the closest pair of the proteins of interest within a genome is a reasonable approach to identify interologs, provided the distance is conserved across taxa (Ovchinnikov et al., 2014). This strategy is particularly useful for the bacterial two-component signalling system. The sensor histidine kinases and the response regulator proteins that make up the two-component signalling system are highly paralogous in bacterial genomes, but most cognate pairs lie adjacent to one another on the genome (Stock et al., 2000). Unfortunately, metrics to identify interologs based on intergenic distance limit the use of evolutionary coupling analysis to complexes within operons. Fortunately, other strategies can be used to confidently, albeit conservatively, identify orthologs, enabling the discovery of evolutionary couplings across the proteome. In this case, when there are multiple paralogs in a genome to choose from, the paralog that is closest to the query sequence is selected. This can be based simply on identity (Green et al., 2019; Zeng et al., 2018), or by a “reciprocal best hits” procedure (Cong et al., 2019). In this procedure, the query *E. coli* sequence is used to identify homologues in another proteome via a BLAST search. Hits from the BLAST search are then

used as query sequences on the *E. coli* proteome. If the original query is among the best hits, then that sequence is selected as a candidate ortholog.

There is considerable debate over the nature of the signal that evolutionary couplings analysis measures (Aurell, 2016; Meyer et al., 2019; van Nimwegen, 2016; Talavera et al., 2015). Coupled pairs of columns in an MSA could be related to one another by a coevolutionary mechanism, where an unfavorable mutation in one position is compensated for by another mutation at a different position. However, they could also be related by shared evolutionary history, where a pair of changes have diverged into their own phylogenetic branch. Both of these events appear identical in a MSA but are easily distinguished with the use of a phylogenetic tree, with coevolutionary events occurring multiple independent times on separate branches (Figure 1.3). A comparison of maximum-entropy models with highly curated and well-established phylogenetic trees suggests that these models are not measuring coevolution *per se*, as the correlated pairs of mutations do not typically occur on the same branch of the tree (Talavera et al., 2015). On the other hand, tree-aware methods assume the input phylogenetic tree is the ground truth, but these trees are generally constructed using algorithms that assume each site evolves independently of the others (Caporaso et al., 2008; Meyer et al., 2019; Talavera et al., 2015). In other words, many tree-aware methods to detect coevolution may be starting from the assumption that coevolution does not occur! Indeed, a Bayesian tree-building algorithm that explicitly accounts for coevolution between sites identified over 200 coevolving pairs of nucleotides in 16S RNA, the vast majority of which were contacting one another and were consistent with Watson-Crick-Franklin base pairing (Meyer et al., 2019). While this algorithm was developed to examine coevolution between nucleotides, it may be possible to use the same framework for protein sequences.

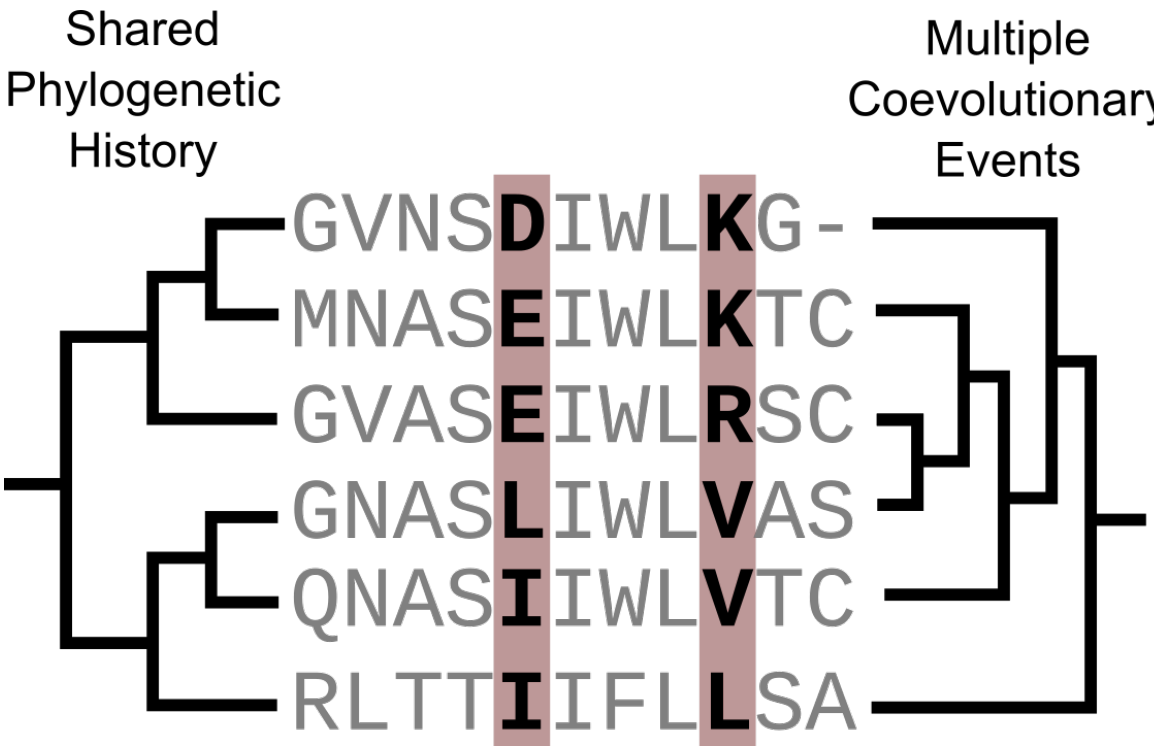

Figure 1.3: Correlated columns in a multiple sequence alignment may arise from shared phylogeny or coevolution. Left: example tree topology that would generate a correlated pair of mutations (hilighted in red) from a single event. Right: the same alignment could have been generated from a different phylogenetic tree, with multiple independent changes at these positions. This topology is more consistent with the pairs of positions coevolving with one another.

A recent analysis of weighting sequences for evolutionary direct coupling analysis further demonstrates that coevolution may not be exactly what these models capture (Hockenberry and Wilke, 2019). Direct coupling analysis requires large numbers of representative homologues to successfully infer contacting residues, but sequence databases are highly biased for particular taxa. The usual method employed in these coupling analyses is to assign a weight to each input

sequence that is inversely proportional to the number of similar sequences in the alignment and filter any completely redundant ones (Hopf et al., 2014, 2017; Marks et al., 2011). Such a weighting scheme loosely attempts to both account for biased sampling as well as shared phylogeny, as similar sequences presumably diverged more recently in their evolutionary history. However, alternative weights could be chosen for the sequences that more explicitly take the phylogenetic tree and evolution into account. If maximum-entropy models are measuring coevolution, such a phylogenetic reweighting should be critical. However, these tree-aware reweighting schemes performed worse than the standard, tree-independent method. Even using the best weighting scheme did little better than not using any weights at all (Hockenberry and Wilke, 2019). Since coevolution and shared phylogeny can only be distinguished through a phylogenetic tree which evolutionary coupling analysis is completely blind to, another mechanism may be at work. These metrics may in fact be detecting slowly evolving residues within the core of the protein, where acceptable mutations are highly constrained (Talavera et al., 2015).

Nonetheless, even though *genuine* coevolution can only be detected by incorporating phylogeny, tree-independent methods have been shown to match or outperform phylogenetic methods in terms of reconstructing protein tertiary contacts (Caporaso et al., 2008; Dunn et al., 2007). This gap has only gotten wider with the advent of maximum-entropy models of evolutionary couplings. It is likely that better accounting of phylogeny will improve the performance of these models (Stein et al., 2015), but they are already clearly capturing some form of the protein's evolutionary history in a manner useful for structure prediction.

1.4 Outline of thesis

In this thesis I combine data from a variety of *in vitro*, *in vivo*, and bioinformatic sources with protein structure prediction to better understand protein conformational changes, interpret mutagenesis results, and predict the structures of membrane proteins. These molecular models are then used to help interpret the experimental results and generate hypotheses for further experiments.

In Chapter 2, I used maximum-entropy statistical models of MSAs of the bacterial cell division proteins FtsB and FtsL to identify evolutionarily coupled positions between the two proteins. These positions were consistent with a parallel helical interaction spanning the transmembrane and periplasmic coiled coil domains, with a shift in the FtsB interacting face between the transmembrane domain and coiled coils. We copurified these proteins and estimated their oligomeric state using FRET, determining the most likely stoichiometry to be a $FtsB_2$:$FtsL_2$ heterotetramer. I combined this information with the evolutionary analysis to predict the structure of the membrane-embedded FtsLB subcomplex. Structural features of the complex, such as a continuous helix in FtsL spanning the transmembrane and coiled coil regions, a juxtamembrane flexible linker in FtsB, and the FtsB:FtsL transmembrane interface were investigated via *in vivo* mutagenesis and were found to be important for properly maintaining cell division.

In Chapter 3, I further investigate the coiled coil domain of FtsL and FtsB, particularly in relation to the number of highly polar and charged residues at interfacial positions, which are hydrophobic in canonical coiled-coils. We found that some of these positions appear to be critical for maintaining proper cell division, but the complex seems to be tuned for instability. Replacing multiple polar and charged residues with hydrophobic ones increases helicity at elevated

temperatures but results in a loss of cooperativity during unfolding. These results also suggest a possible alternate arrangement of the FtsLB coiled coils in order to better expose the charged and polar residues to the external solvent.

In Chapter 4, I combine *ab initio* structure prediction with ancestral sequence reconstruction and combinatorial interface swapping to better understand the dimerization landscape of the transmembrane domain of the mitophagy receptor BNIP3. BNIP3 dimerization is mediated by a GxxxG dimerization motif as well as interhelical hydrogen bonds from membrane-embedded polar residues. Mutations that disrupt these important dimerization features have been shown to lead to defects in mitophagy, suggesting that dimerization is a functionally important feature. Based on MSAs, homologues of BNIP3 have shifted their GxxxG motifs and hydrogen bonding nodes by four residues within the TMD. Structures predicted for these homologues are generally very similar to the human BNIP3 NMR structure, but the helix-helix crossing point is shifted. These homologues dimerize with similar affinities in a genetic reporter assay, leaving the evolutionary mechanism by which these features shifted position while maintaining dimerization unclear. Ancestral sequence reconstruction suggested that dimerization was preserved throughout BNIP3's evolutionary history via a tandem, stabilizing GxxxGxxxG motif. Predicted BNIP3 sequences lacking a GxxxG motif generally had worse energy scores than those with a GxxxG motif, indicating its criticality, while the energy scores of interfaces without a hydrogen bonding node were similar to those with one. As the latter result somewhat conflicts with experimental evidence, it indicates that the structure prediction algorithm accurately recapitulates the structure of BNIP3, but underestimates the importance of hydrogen bonding.

In Chapter 5, I describe the use of cysteine crosslinking experiments to model a conformational change in the pro-apoptotic protein Bax. Activated Bax induces apoptosis

through oligomerization, though the mechanism is unclear. Chemical labeling suggested helix α5 partially inserts into the mitochondrial outer membrane during activation. After this occurred, pairs of cysteine variants in helices α2, α3, and α4 that are distant in known structures formed strong crosslinks. Short molecular dynamics simulations suggested some of these pairs could contact each other after insertion of helix α5, but the trajectories did not converge. To explore alternative conformations more rapidly, I used protein structure prediction guided by the cysteine crosslinking distance restraints and found that the helices could be rearranged into a well-packed conformation that satisfied the experimental data.

In Chapter 6, I performed *in silico* mutagenesis to better understand the regulatory mechanism of Cit1p phosphorylation. Mass spectrometry identified the citrate synthase Cit1p as a key substrate of the phosphatate Ptc7p in yeast. Phosphorylation of Cit1p eliminates its activity and disrupts the oligomeric state, though the mechanism was unclear. Mutating the phosphorylated serine in Cit1p structures to glutamate and phosphoserine *in silico* dramatically increased the energy score due to clashes, indicating that the active dimeric state of Cit1p cannot accommodate a phosphorylated or phosphomimetic residue at that position.

Finally, in Appendix I, I describe the development and implementation of a step-detection algorithm for single-molecule Total Internal Reflection Fluorescence (TIRF) microscopy experiments. Individual fluorophores stochastically undergo photobleaching when excited, and these stepwise changes in fluorescence can be used to count the number of fluorophores per particle and infer the oligomeric state of a fluorescently labeled protein complex. Assessment of photobleaching traces can be performed manually, but this is tedious and prone to subjective biases. I used a step-detection algorithm commonly used in image processing to identify stepwise changes in fluorescence traces and found that it outperforms manual detection while being

significantly faster and somewhat less prone to bias. The algorithm was integrated into a software package specifically designed for TIRF microscopy analyses.

1.5 References

Anfinsen, C.B. (1973). Principles that Govern the Folding of Protein Chains. Science *181*, 223–230.

Anishchenko, I., Ovchinnikov, S., Kamisetty, H., and Baker, D. (2017). Origins of coevolution between residues distant in protein 3D structures. Proc. Natl. Acad. Sci. *114*, 9122–9127.

Aprahamian, M.L., Chea, E.E., Jones, L.M., and Lindert, S. (2018). Rosetta Protein Structure Prediction from Hydroxyl Radical Protein Footprinting Mass Spectrometry Data. Anal. Chem. *90*, 7721–7729.

Aurell, E. (2016). The Maximum Entropy Fallacy Redux? PLOS Comput. Biol. *12*, e1004777.

Badaczewska-Dawid, A.E., Kolinski, A., and Kmiecik, S. (2020). Computational reconstruction of atomistic protein structures from coarse-grained models. Comput. Struct. Biotechnol. J. *18*, 162–176.

Baker, D., Sohl, J.L., and Agard, D.A. (1992). A protein-folding reaction under kinetic control. Nature *356*, 263–265.

Bender, B.J., Cisneros, A., Duran, A.M., Finn, J.A., Fu, D., Lokits, A.D., Mueller, B.K., Sangha, A.K., Sauer, M.F., Sevy, A.M., et al. (2016). Protocols for Molecular Modeling with Rosetta3 and RosettaScripts. Biochemistry *55*, 4748–4763.

Berkholz, D.S., Shapovalov, M.V., Dunbrack, R.L., and Karplus, P.A. (2009). Conformation Dependence of Backbone Geometry in Proteins. Structure *17*, 1316–1325.

Bowie, J.U., and Eisenberg, D. (1994). An evolutionary approach to folding small alpha-helical proteins that uses sequence information and an empirical guiding fitness function. Proc. Natl. Acad. Sci. *91*, 4436–4440.

Braun, T., Koehler Leman, J., and Lange, O.F. (2015). Combining Evolutionary Information and an Iterative Sampling Strategy for Accurate Protein Structure Prediction. PLOS Comput. Biol. *11*, e1004661.

Brunger, A.T. (2007). Version 1.2 of the Crystallography and NMR system. Nat. Protoc. *2*, 2728–2733.

Buchan, D.W.A., and Jones, D.T. (2019). The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Res. *47*, W402–W407.

Burger, L., and van Nimwegen, E. (2010). Disentangling Direct from Indirect Co-Evolution of Residues in Protein Alignments. PLoS Comput. Biol. *6*, e1000633.

Caporaso, J.G., Smit, S., Easton, B.C., Hunter, L., Huttley, G.A., and Knight, R. (2008). Detecting coevolution without phylogenetic trees? Tree-ignorant metrics of coevolution perform as well as tree-aware metrics. BMC Evol. Biol. *8*, 327.

Chen, J., Guo, M., Wang, X., and Liu, B. (2018). A comprehensive review and comparison of different computational methods for protein remote homology detection. Brief. Bioinform. *19*, 231–244.

Clarke, N.D. (1995). Covariation of residues in the homeodomain sequence family. Protein Sci. *4*, 2269–2278.

Clore, G.M., and Gronenborn, A.M. (1989). Determination of Three-Dimensional Structures of Proteins and Nucleic Acids in Solution by Nuclear Magnetic Resonance Spectroscop. Crit. Rev. Biochem. Mol. Biol. *24*, 479–564.

Cong, Q., Anishchenko, I., Ovchinnikov, S., and Baker, D. (2019). Protein interaction networks revealed by proteome coevolution. Science *365*, 185–189.

Coutsias, E.A., Seok, C., and Dill, K.A. (2004). Using quaternions to calculate RMSD. J. Comput. Chem. *25*, 1849–1857.

Das, R., and Baker, D. (2008). Macromolecular Modeling with Rosetta. Annu. Rev. Biochem. *77*, 363–382.

DiMaio, F., Tyka, M.D., Baker, M.L., Chiu, W., and Baker, D. (2009). Refinement of Protein Structures into Low-Resolution Density Maps Using Rosetta. J. Mol. Biol. *392*, 181–190.

Dosztányi, Z. (2018). Prediction of protein disorder based on IUPred: Prediction of Protein Disorder Based on IUPred. Protein Sci. *27*, 331–340.

Dunbrack, R.L. (2002). Rotamer Libraries in the 21st Century. Curr. Opin. Struct. Biol. *12*, 431–440.

Dunker, A.K., Lawson, J.D., Brown, C.J., Williams, R.M., Romero, P., Oh, J.S., Oldfield, C.J., Campen, A.M., Ratliff, C.M., Hipps, K.W., et al. (2001). Intrinsically disordered protein. J. Mol. Graph. Model. *19*, 26–59.

Dunn, S.D., Wahl, L.M., and Gloor, G.B. (2007). Mutual information without the influence of phylogeny or entropy dramatically improves residue contact prediction. Bioinformatics *24*, 333–340.

Ekeberg, M., Lövkvist, C., Lan, Y., Weigt, M., and Aurell, E. (2013). Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models. Phys. Rev. E *87*.

Fleishman, S.J., and Ben-Tal, N. (2002). A Novel Scoring Function for Predicting the Conformations of Tightly Packed Pairs of Transmembrane α-Helices. J. Mol. Biol. *321*, 363–378.

Förster, F., Webb, B., Krukenberg, K.A., Tsuruta, H., Agard, D.A., and Sali, A. (2008). Integration of Small-Angle X-Ray Scattering Data into Structural Modeling of Proteins and Their Assemblies. J. Mol. Biol. *382*, 1089–1106.

Franke, D., and Svergun, D.I. (2009). *DAMMIF* , a program for rapid *ab-initio* shape determination in small-angle scattering. J. Appl. Crystallogr. *42*, 342–346.

Gallagher, J.R., Kim, A.J., Gulati, N.M., and Harris, A.K. (2019). Negative-Stain Transmission Electron Microscopy of Molecular Complexes for Image Analysis by 2D Class Averaging. Curr. Protoc. Microbiol. *54*.

Gao, G., Williams, J.G., and Campbell, S.L. (2004). Protein–Protein Interaction Analysis by Nuclear Magnetic Resonance Spectroscopy. In Protein-Protein Interactions, (New Jersey: Humana Press), pp. 079–092.

Green, A.G., Elhabashy, H., Brock, K.P., Maddamsetti, R., Kohlbacher, O., and Marks, D.S. (2019). Proteome-scale discovery of protein interactions with residue-level resolution using sequence coevolution (Bioinformatics).

Grigoryan, G., and DeGrado, W.F. (2011). Probing Designability via a Generalized Model of Helical Bundle Geometry. J. Mol. Biol. *405*, 1079–1100.

Hafsa, N.E., Arndt, D., and Wishart, D.S. (2015). CSI 3.0: a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts. Nucleic Acids Res. *43*, W370–W377.

Hansmann, U.H.E., and Okamoto, Y. (1999). New Monte Carlo algorithms for protein folding. Curr. Opin. Struct. Biol. *9*, 177–183.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. Biometrika *57*, 97–109.

Hockenberry, A.J., and Wilke, C.O. (2019). Phylogenetic Weighting Does Little to Improve the Accuracy of Evolutionary Coupling Analyses. Entropy *21*, 1000.

Hopf, T.A., Colwell, L.J., Sheridan, R., Rost, B., Sander, C., and Marks, D.S. (2012). Three-Dimensional Structures of Membrane Proteins from Genomic Sequencing. Cell *149*, 1607–1621.

Hopf, T.A., Schärfe, C.P.I., Rodrigues, J.P.G.L.M., Green, A.G., Kohlbacher, O., Sander, C., Bonvin, A.M.J.J., and Marks, D.S. (2014). Sequence co-evolution gives 3D contacts and structures of protein complexes. ELife *3*.

Hopf, T.A., Ingraham, J.B., Poelwijk, F.J., Schärfe, C.P.I., Springer, M., Sander, C., and Marks, D.S. (2017). Mutation effects predicted from sequence co-variation. Nat. Biotechnol. *35*, 128–135.

Illergård, K., Ardell, D.H., and Elofsson, A. (2009). Structure is three to ten times more conserved than sequence-A study of structural response in protein cores. Proteins Struct. Funct. Bioinforma. *77*, 499–508.

van Ingen, H., and Bonvin, A.M.J.J. (2014). Information-driven modeling of large macromolecular assemblies using NMR data. J. Magn. Reson. *241*, 103–114.

Kamisetty, H., Ovchinnikov, S., and Baker, D. (2013). Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. Proc. Natl. Acad. Sci. *110*, 15674–15679.

Karakaş, M., Woetzel, N., Staritzbichler, R., Alexander, N., Weiner, B.E., and Meiler, J. (2012). BCL::Fold - De Novo Prediction of Complex and Large Protein Topologies by Assembly of Secondary Structure Elements. PLoS ONE *7*, e49240.

Kerner, M.J., Naylor, D.J., Ishihama, Y., Maier, T., Chang, H.-C., Stines, A.P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M., et al. (2005). Proteome-wide Analysis of Chaperonin-Dependent Protein Folding in Escherichia coli. Cell *122*, 209–220.

Kim, R., and Guo, J. (2010). Systematic analysis of short internal indels and their impact on protein folding. BMC Struct. Biol. *10*, 24.

Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P. (1983). Optimization by Simulated Annealing. Science *220*, 671–680.

Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A.E., and Kolinski, A. (2016). Coarse-Grained Protein Models and Their Applications. Chem. Rev. *116*, 7898–7936.

Koehler Leman, J., and Bonneau, R. (2018). A Novel Domain Assembly Routine for Creating Full-Length Models of Membrane Proteins from Known Domain Structures. Biochemistry *57*, 1939–1944.

Korber, B.T., Farber, R.M., Wolpert, D.H., and Lapedes, A.S. (1993). Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: an information theoretic analysis. Proc. Natl. Acad. Sci. *90*, 7176–7180.

Krivov, G.G., Shapovalov, M.V., and Dunbrack, R.L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. Proteins Struct. Funct. Bioinforma. *77*, 778–795.

Kryshtafovych, A., Schwede, T., Topf, M., Fidelis, K., and Moult, J. (2019). Critical assessment of methods of protein structure prediction (CASP)—Round XIII. Proteins Struct. Funct. Bioinforma. *87*, 1011–1020.

Kubelka, J., Hofrichter, J., and Eaton, W.A. (2004). The protein folding 'speed limit.' Curr. Opin. Struct. Biol. *14*, 76–88.

Lam, S.D., Das, S., Sillitoe, I., and Orengo, C. (2017). An overview of comparative modelling and resources dedicated to large-scale modelling of genome sequences. Acta Crystallogr. Sect. Struct. Biol. *73*, 628–640.

Lapedes, A., Giraud, B., and Jarzynski, C. (2012). Using Sequence Alignments to Predict Protein Structure and Stability With High Accuracy. ArXiv12072484 Q-Bio.

Lapedes, A.S., Giraud, B.G., Liu, L.C., and Stormo, G.D. (1998). A Maximum Entropy Formalism for Disentangling Chains of Correlated Sequence Positions.

Larsson, P., Wallner, B., Lindahl, E., and Elofsson, A. (2008). Using multiple templates to improve quality of homology models in automated homology modeling. Protein Sci. *17*, 990–1002.

Lazaridis, T. (2003). Effective energy function for proteins in lipid membranes. Proteins Struct. Funct. Genet. *52*, 176–192.

Lazaridis, T., and Karplus, M. (1999). Effective energy function for proteins in solution. Proteins *35*, 133–152.

Levinthal, C. (1969). How To Fold Graciously. Mössbaun Spectrosc. Biol. Syst. Proc. *67*, 22–24.

Li, W., Zhang, Y., and Skolnick, J. (2004). Application of Sparse NMR Restraints to Large-Scale Protein Structure Prediction. Biophys. J. *87*, 1241–1248.

Liu, T., Pantazatos, D., Li, S., Hamuro, Y., Hilser, V.J., and Woods, V.L. (2012). Quantitative Assessment of Protein Structural Models by Comparison of H/D Exchange MS Data with Exchange Behavior Accurately Predicted by DXCOREX. J. Am. Soc. Mass Spectrom. *23*, 43–56.

Ludwiczak, J., Winski, A., Szczepaniak, K., Alva, V., and Dunin-Horkawicz, S. (2019). DeepCoil—a fast and accurate prediction of coiled-coil domains in protein sequences. Bioinformatics *35*, 2790–2795.

Lukk, T., Sakai, A., Kalyanaraman, C., Brown, S.D., Imker, H.J., Song, L., Fedorov, A.A., Fedorov, E.V., Toro, R., Hillerich, B., et al. (2012). Homology models guide discovery of diverse enzyme specificities among dipeptide epimerases in the enolase superfamily. Proc. Natl. Acad. Sci. *109*, 4122–4127.

Mackerell, A.D. (2004). Empirical force fields for biological macromolecules: Overview and issues. J. Comput. Chem. *25*, 1584–1604.

Marks, D.S., Colwell, L.J., Sheridan, R., Hopf, T.A., Pagnani, A., Zecchina, R., and Sander, C. (2011). Protein 3D Structure Computed from Evolutionary Sequence Variation. PLoS ONE *6*, e28766.

McDonnell, A.V., Jiang, T., Keating, A.E., and Berger, B. (2006). Paircoil2: improved prediction of coiled coils from sequence. Bioinformatics *22*, 356–358.

Meier, A., and Söding, J. (2015). Automatic Prediction of Protein 3D Structures by Probabilistic Multi-template Homology Modeling. PLOS Comput. Biol. *11*, e1004343.

Meyer, X., Dib, L., Silvestro, D., and Salamin, N. (2019). Simultaneous Bayesian inference of phylogeny and molecular coevolution. Proc. Natl. Acad. Sci. *116*, 5027–5036.

Miyazawa, S., and Jernigan, R.L. (1985). Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules *18*, 534–552.

Monastyrskyy, B., D'Andrea, D., Fidelis, K., Tramontano, A., and Kryshtafovych, A. (2016). New encouraging developments in contact prediction: Assessment of the CASP11 results: Contact Assessment. Proteins Struct. Funct. Bioinforma. *84*, 131–144.

Monticelli, L., Kandasamy, S.K., Periole, X., Larson, R.G., Tieleman, D.P., and Marrink, S.-J. (2008). The MARTINI Coarse-Grained Force Field: Extension to Proteins. J. Chem. Theory Comput. *4*, 819–834.

Mueller, B.K., Subramaniam, S., and Senes, A. (2014). A frequent, GxxxG-mediated, transmembrane association motif is optimized for the formation of interhelical C -H hydrogen bonds. Proc. Natl. Acad. Sci.

Neria, E., Fischer, S., and Karplus, M. (1996). Simulation of activation free energies in molecular systems. J. Chem. Phys. *105*, 1902–1921.

van Nimwegen, E. (2016). Inferring Contacting Residues within and between Proteins: What Do the Probabilities Mean? PLOS Comput. Biol. *12*, e1004726.

Ogorzalek, T.L., Hura, G.L., Belsom, A., Burnett, K.H., Kryshtafovych, A., Tainer, J.A., Rappsilber, J., Tsutakawa, S.E., and Fidelis, K. (2018). Small angle X-ray scattering and cross-linking for data assisted protein structure prediction in CASP 12 with prospects for improved accuracy. Proteins Struct. Funct. Bioinforma. *86*, 202–214.

Ovchinnikov, S., Kamisetty, H., and Baker, D. (2014). Robust and accurate prediction of residue–residue interactions across protein interfaces using evolutionary information. ELife *3*.

Park, H., Bradley, P., Greisen, P., Liu, Y., Mulligan, V.K., Kim, D.E., Baker, D., and DiMaio, F. (2016). Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. J. Chem. Theory Comput. *12*, 6201–6212.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods *8*, 785–786.

Petoukhov, M.V., and Svergun, D.I. (2013). Applications of small-angle X-ray scattering to biomacromolecular solutions. Int. J. Biochem. Cell Biol. *45*, 429–437.

Pierce, N.A., and Winfree, E. (2002). Protein Design is NP-hard. Protein Eng. Des. Sel. *15*, 779–782.

Przytycka, T. (2004). Significance of conformational biases in Monte Carlo simulations of protein folding: Lessons from Metropolis-Hastings approach. Proteins Struct. Funct. Bioinforma. *57*, 338–344.

Ramachandran, G.N., Ramakrishnan, C., and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. J. Mol. Biol. *7*, 95–99.

Raval, A., Piana, S., Eastwood, M.P., and Shaw, D.E. (2016). Assessment of the utility of contact-based restraints in accelerating the prediction of protein structure using molecular dynamics simulations: Contact Restraints in Protein Simulations. Protein Sci. *25*, 19–29.

Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastritis, P., Karaca, E., Melquiond, A.S.J., and Bonvin, A.M.J.J. (2012). Clustering biomolecular complexes by residue contacts similarity. Proteins Struct. Funct. Bioinforma. n/a-n/a.

Rohl, C.A., Strauss, C.E.M., Misura, K.M.S., and Baker, D. (2004). Protein Structure Prediction Using Rosetta. In Methods in Enzymology, (Elsevier), pp. 66–93.

Šali, A., and Blundell, T.L. (1993). Comparative Protein Modelling by Satisfaction of Spatial Restraints. J. Mol. Biol. *234*, 779–815.

Schaeffer, R.D., and Daggett, V. (2011). Protein folds and protein folding. Protein Eng. Des. Sel. *24*, 11–19.

Schindler, C.E.M., de Vries, S.J., Sasse, A., and Zacharias, M. (2016). SAXS Data Alone can Generate High-Quality Models of Protein-Protein Complexes. Structure *24*, 1387–1397.

Schwede, T. (2003). SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Res. *31*, 3381–3385.

Schwede, T. (2013). Protein Modeling: What Happened to the “Protein Structure Gap”? Structure *21*, 1531–1540.

Senior, A.W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A.W.R., Bridgland, A., et al. (2020). Improved protein structure prediction using potentials from deep learning. Nature *577*, 706–710.

Shapovalov, M.V., and Dunbrack, R.L. (2011). A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. Structure *19*, 844–858.

Shen, Y., and Bax, A. (2013). Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J. Biomol. NMR *56*, 227–241.

Simons, K.T., Kooperberg, C., Huang, E., and Baker, D. (1997). Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. J. Mol. Biol. *268*, 209–225.

Sonnhammer, E.L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. ISMB Int. Conf. Intell. Syst. Mol. Biol. *6*, 175–182.

Stein, R.R., Marks, D.S., and Sander, C. (2015). Inferring Pairwise Interactions from Biological Data Using Maximum-Entropy Probability Models. PLOS Comput. Biol. *11*, e1004182.

Stiffler, M.A., Poelwijk, F.J., Brock, K.P., Stein, R.R., Riesselman, A., Teyra, J., Sidhu, S.S., Marks, D.S., Gauthier, N.P., and Sander, C. (2020). Protein Structure from Experimental Evolution. Cell Syst. *10*, 15-24.e5.

Stock, A.M., Robinson, V.L., and Goudreau, P.N. (2000). Two-Component Signal Transduction. Annu. Rev. Biochem. *69*, 183–215.

Subramaniam, S., and Senes, A. (2012). An energy-based conformer library for side chain optimization: Improved prediction and adjustable sampling. Proteins Struct. Funct. Bioinforma. *80*, 2218–2234.

Talavera, D., Lovell, S.C., and Whelan, S. (2015). Covariation Is a Poor Measure of Molecular Coevolution. Mol. Biol. Evol. *32*, 2456–2468.

Tang, Y., Huang, Y.J., Hopf, T.A., Sander, C., Marks, D.S., and Montelione, G.T. (2015). Protein structure determination by combining sparse NMR data with evolutionary couplings. Nat. Methods *12*, 751–754.

Wang, Y., and Barth, P. (2015). Evolutionary-guided de novo structure prediction of self-associated transmembrane helical proteins with near-atomic accuracy. Nat. Commun. *6*.

Wang, S., Peng, J., Ma, J., and Xu, J. (2016). Protein Secondary Structure Prediction Using Deep Convolutional Neural Fields. Sci. Rep. *6*.

Webb, B., Viswanath, S., Bonomi, M., Pellarin, R., Greenberg, C.H., Saltzberg, D., and Sali, A. (2018). Integrative structure modeling with the Integrative Modeling Platform: Integrative Structure Modeling with IMP. Protein Sci. *27*, 245–258.

Weigt, M., White, R.A., Szurmant, H., Hoch, J.A., and Hwa, T. (2009). Identification of direct residue contacts in protein-protein interaction by message passing. Proc. Natl. Acad. Sci. *106*, 67–72.

Weinstein, J.Y., Elazar, A., and Fleishman, S.J. (2019). A lipophilicity-based energy function for membrane-protein modelling and design. PLOS Comput. Biol. *15*, e1007318.

Wimley, W.C., and White, S.H. (1996). Experimentally determined hydrophobicity scale for proteins at membrane interfaces. Nat. Struct. Biol. *3*, 842–848.

Wishart, D.S., Sykes, B.D., and Richards, F.M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry *31*, 1647–1651.

Xu, J. (2019). Distance-based protein folding powered by deep learning. Proc. Natl. Acad. Sci. *116*, 16856–16865.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2015). The I-TASSER Suite: protein structure and function prediction. Nat. Methods *12*, 7–8.

Yang, Y., Gao, J., Wang, J., Heffernan, R., Hanson, J., Paliwal, K., and Zhou, Y. (2016). Sixty-five years of the long march in protein secondary structure prediction: the final stretch? Brief. Bioinform. bbw129.

Yellaboina, S., Goyal, K., and Mande, S.C. (2007). Inferring genome-wide functional linkages in E. coli by combining improved genome context methods: Comparison with high-throughput experimental data. Genome Res. *17*, 527–535.

Zeng, H., Wang, S., Zhou, T., Zhao, F., Li, X., Wu, Q., and Xu, J. (2018). ComplexContact: a web server for inter-protein contact prediction using deep learning. Nucleic Acids Res. *46*, W432–W437.

Zhang, J., and Liu, J.S. (2006). On Side-Chain Conformational Entropy of Proteins. PLoS Comput. Biol. *2*, e168.

Zhang, Y., and Skolnick, J. (2004). Scoring function for automated assessment of protein structure template quality. Proteins Struct. Funct. Bioinforma. *57*, 702–710.

Zheng, W., and Tekpinar, M. (2014). High-Resolution Modeling of Protein Structures Based on Flexible Fitting of Low-Resolution Structural Data. In Advances in Protein Chemistry and Structural Biology, (Elsevier), pp. 267–284.

Zhu, X., Lopes, P.E.M., and MacKerell, A.D. (2012). Recent developments and applications of the CHARMM force fields. WIREs Comput. Mol. Sci. *2*, 167–185.

Zou, T., Williams, N., Ozkan, S.B., and Ghosh, K. (2014). Proteome Folding Kinetics Is Limited by Protein Halflife. PLoS ONE *9*, e112701.

van Zundert, G.C.P., Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastritis, P.L., Karaca, E., Melquiond, A.S.J., van Dijk, M., de Vries, S.J., and Bonvin, A.M.J.J. (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J. Mol. Biol. *428*, 720–725.

Chapter 2: The FtsLB subcomplex of the bacterial divisome is a tetramer with an uninterrupted FtsL helix linking the transmembrane and periplasmic regions

This chapter was prepared for publication as:

Condon S.G.F.*, Mahbuba D.-A.*, Armstrong C.R., Diaz Vazquez G., Craven S.J., LaPointe L.M., Khadria A.S., Chadda R., Crooks J.A., Rangarajan N., Weibel D.B., Hoskins A.A., Robertston J.L., Cui Q., Senes A. 2018. “The FtsLB subcomplex of the bacterial divisome is a tetramer with an uninterrupted FtsL helix linking the transmembrane and periplasmic regions.” *J. Biol. Chem. 293*, 1623–1641.

**Both myself and Deena-Al Mahbuba contributed to this work equally.*

Statement of contribution

I performed the evolutionary couplings analysis and modeled the FtsLB complex. I also helped carry out the FRET experiments, Western blots, and data analysis.

2.1 Abstract

In *Escherichia coli*, FtsLB plays a central role in the initiation of cell division, possibly transducing a signal that will eventually lead to the activation of peptidoglycan remodeling at the forming septum. The molecular mechanisms by which FtsLB operates in the divisome, however, are not understood. Here, we present a structural analysis of the FtsLB complex, performed with biophysical, computational, and *in vivo* methods, that establishes the organization of the transmembrane region and proximal coiled coil of the complex. FRET analysis *in vitro* is consistent with formation of a tetramer composed of two FtsL and two FtsB subunits. We predicted subunit contacts through co-evolutionary analysis and used them to compute a structural model of the complex. The transmembrane region of FtsLB is stabilized by hydrophobic packing and by a complex network of hydrogen bonds. The coiled coil domain probably terminates near the critical constriction control domain, which might correspond to a structural transition. The presence of strongly polar amino acids within the core of the tetrameric coiled coil suggests that the coil may split into two independent FtsQ-binding domains. The helix of FtsB is interrupted between the transmembrane and coiled coil regions by a flexible Gly-rich linker. Conversely, the data suggest that FtsL forms an uninterrupted helix across the two regions and that the integrity of this helix is indispensable for the function of the complex. The FtsL helix is thus a candidate for acting as a potential mechanical connection to communicate conformational changes between periplasmic, membrane, and cytoplasmic regions.

2.2 Introduction

The cell envelope of Gram-negative bacteria, such as *Escherichia coli*, consists of three layers: a cytoplasmic membrane, an outer membrane, and a rigid cell wall formed by a mesh of

peptidoglycan located in the intervening periplasmic space. Cell division requires mechanisms for the separation of these three layers. The first step is the establishment of a division site and the segregation of the duplicated chromosome. This allows for the beginning of constriction, during which the coordinated activities of numerous peptidoglycan synthases and hydrolases remodel the cell wall. Finally, the process leads to membrane fusion and to the separation of two daughter cells. The large multiprotein complex that supports a majority of these functions is called the divisome.

At least 3 dozen proteins are known to participate in cell division in *E. coli*, but the essential components of the divisome consist of a core of 12 proteins. Their recruitment to mid-cell follows a hierarchical order of dependence (FtsZ → FtsA-ZipA → FtsE-FtsX → FtsK → FtsQ → FtsL-FtsB → FtsW → FtsI → FtsN; Figure 2.1*a*), which also reflects, in part, the timing and interactions that occur in the complex assembly (Buddelmeijer and Beckwith, 2004; Buddelmeijer et al., 2002; Chen and Beckwith, 2001; Ghigo et al., 1999; Goehring, 2005; Goehring et al., 2006; Hale and de Boer, 1999, 2002; Liu et al., 1999; Mercer and Weiss, 2002; Schmidt et al., 2004; Wang et al., 1998). The divisome assembles around the tubulin homolog FtsZ, which forms a ringlike structure at mid-cell (the Z-ring) (de Boer, 2010; Erickson et al., 2010; Lutkenhaus et al., 2012). FtsZ protofilaments treadmill around the circumference of the cell, providing a scaffold for the recruitment and movement of the components around the cell and ultimately leading to incorporation of new peptidoglycan around the forming septum (Bisson-Filho et al., 2017; Yang et al., 2017). FtsZ is tethered to the plasma membrane by the cooperative action of ZipA, a single-pass transmembrane (TM) protein (Haney et al., 2001; Ortiz et al., 2016), and of FtsA, an actin homolog and peripheral membrane protein able to form protofilaments (Szwedziak et al., 2012). Other early components of the divisome include FtsEX,

an ABC transporter-like complex that controls cell wall hydrolysis and possibly divisome assembly (Du et al., 2016; Yang et al., 2011), as well as FtsK, a DNA translocase important for chromosome segregation (Männik et al., 2017).

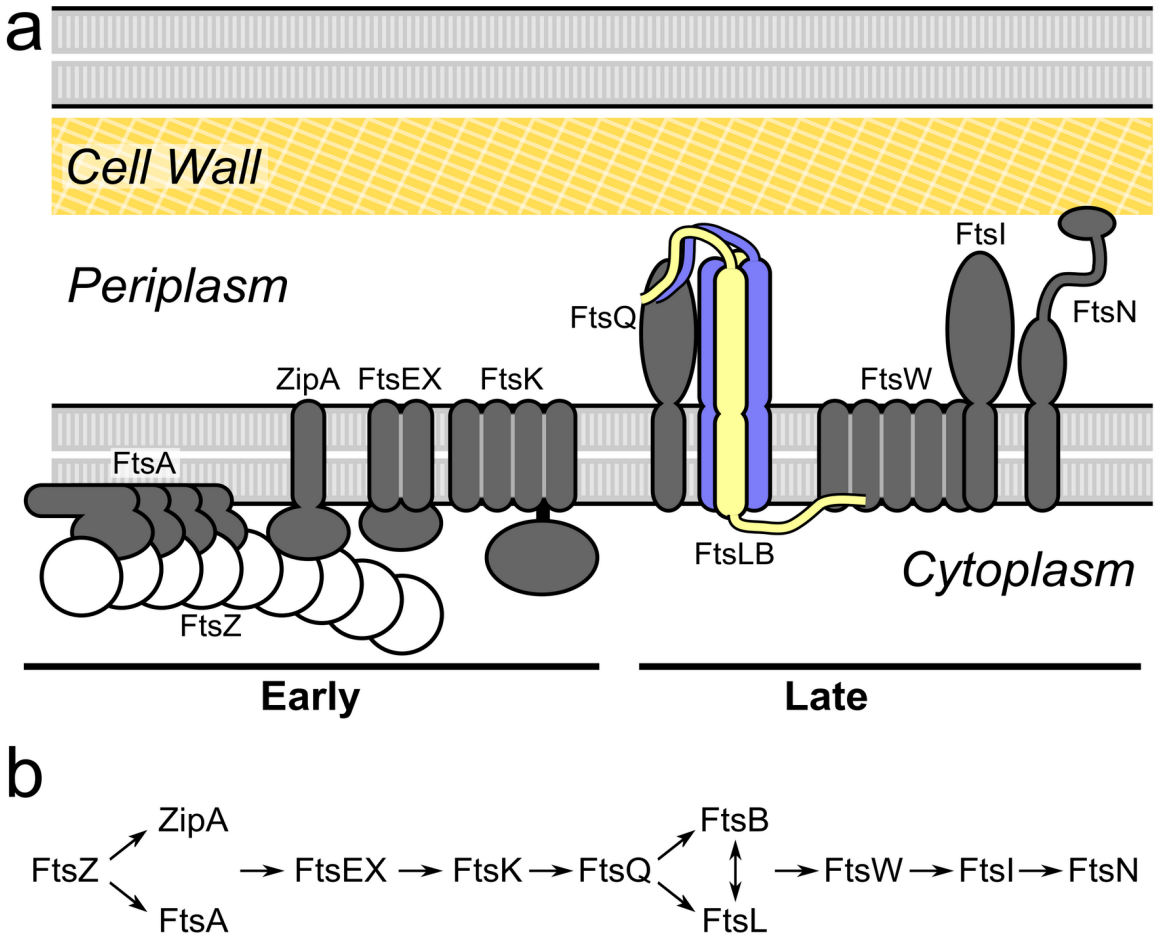

Figure 2.1: **The essential proteins of the divisome.** *a*, schematic representation of the divisome of *E. coli*. The complex assembles around a polymeric scaffold formed by FtsZ. With a few exceptions (FtsZ, FtsA, and FtsE), all essential components are integral membrane proteins. As reported in this work, the FtsLB complex is a heterotetramer formed by two FtsL subunits (*yellow*) and two FtsB subunits (*blue*). The complex forms an extended helical bundle that comprises the transmembrane helices and the periplasmic coiled coil. The C-terminal periplasmic tails of FtsLB mediate the binding to FtsQ. The cytoplasmic tail of FtsL has been hypothesized to bind to FtsW. Description of the

function of the various components is provided in the Introduction. *b*, the divisome of *E. coli* displays a characteristic hierarchy of recruitment at the division site, with the *arrows* indicating a dependence of a component on the one that precedes it in the sequence. The hierarchy also roughly corresponds to the order of recruitment, which can be subdivided into early and late components.

The late components of the divisome (FtsQ, FtsB, FtsL, FtsW, FtsI, and FtsN) localize approximately at the beginning of constriction (Aarsman et al., 2005). FtsQ, FtsB, and FtsL are three bitopic (single-pass) membrane proteins that form a complex whose exact function is not well understood (Buddelmeijer and Beckwith, 2004). The FtsQLB complex is required for the recruitment of FtsW, a large multispan membrane protein, and FtsI, a bitopic protein. FtsW and FtsI work in coordination to synthesize septal cell wall (Mercer and Weiss, 2002); FtsI is a penicillin-binding protein (PBP3) with transpeptidase activity (Vollmer and Bertsche, 2008), and FtsW is most likely its cognate glycosyltransferase (Cho et al., 2016; Emami et al., 2017; Meeske et al., 2016) (FtsW has also been proposed to be a flippase for peptidoglycan precursors (Mohammadi et al., 2011, 2014)). The last protein in the recruitment hierarchy is FtsN, a bitopic protein with an N-terminal TM domain, a predicted disordered region, and a C-terminal domain capable of recognizing septal peptidoglycan (Yang et al., 2004). The precise role of FtsN is not understood, but its accumulation at the septal ring represents a key event for triggering constriction. This activation may be mediated by some of the interactions that have been postulated for FtsN, which include FtsA, the peptidoglycan synthase complex (FtsW, FtsI, and

PBP1B), and the FtsQLB complex (Alexeeva et al., 2010; Busiek et al., 2012; Di Lallo, 2003; Karimova et al., 2005; Müller et al., 2007).

Because the divisome consists primarily of integral membrane proteins, its structural characterization has been lagging and limited to fragments of water-soluble domains (van den Ent et al., 2008; LaPointe et al., 2013; Massey et al., 2006; Mosyak, 2000; Yang et al., 2004). Here we focus on the structural organization of the FtsL-FtsB subcomplex (FtsLB), which has been implicated as a critical player in triggering constriction (Liu et al., 2015; Tsang and Bernhardt, 2015). Topologically, FtsL and FtsB are both small bitopic proteins with a nearly identical domain organization, which suggests that they may derive from a common evolutionary ancestor. Both proteins have short (or absent) N-terminal cytoplasmic tails, one TM domain, a juxtamembrane coiled coil, and C-terminal tails in the periplasm (Figure 2.1*a*). Peptides corresponding to the TM helices form a stable higher-order oligomer *in vitro*, with an equal number of FtsL and FtsB subunits, indicating that the TM region is an important contributor to the stability of the complex (Khadria and Senes, 2013). This is consistent with the observation that the TM domains are biologically important (Buddelmeijer and Beckwith, 2004; Gonzalez and Beckwith, 2009; Guzman et al., 1997). FtsL and FtsB associate *in vivo* even in the absence of FtsQ (Goehring et al., 2006; Robichon et al., 2008), although FtsQ is required for their recruitment to mid-cell (Buddelmeijer and Beckwith, 2004). The association with FtsQ to form the FtsQLB complex is primarily mediated by the C-terminal tails of FtsLB, which bind to the C-terminal end of FtsQ, as evidenced by truncation functional analysis (Gonzalez and Beckwith, 2009; Gonzalez et al., 2010) as well as cross-linking performed *in vivo (van den Berg van Saparoea et al., 2013)*. The periplasmic domains of the three proteins are sufficient to form soluble complexes with submicromolar binding affinities, as established for *E. coli* (Glas et al.,

2015) and *Streptococcus pneumoniae* (Masson et al., 2009; Noirclerc-Savoye et al., 2004) proteins. Notably, these studies have been performed using solubilized FtsLB constructs that were fused to a stable heterodimeric coiled coil and thus forced into a dimeric state. However, FtsLB is likely to be a higher-order oligomer (such as a tetramer), as evidenced by biophysical analysis of their TM helices in isolation (Khadria and Senes, 2013). The tight binding affinity of these 1:1:1 FtsQLB soluble constructs therefore suggests that independent FtsQ-binding sites exist in FtsLB, each formed by one FtsL and one FtsB subunit.

The precise function of FtsLB has not yet been determined. It was originally hypothesized that FtsLB could have a structural role in stabilizing the divisome (Gonzalez and Beckwith, 2009; Gonzalez et al., 2010). The finding that FtsL is subject to active degradation unless it is stabilized by interaction with FtsB in *Bacillus subtilis* (Bramkamp et al., 2006; Daniel and Errington, 2000; Daniel et al., 2006) and *E. coli* (Gonzalez and Beckwith, 2009) suggested that FtsL levels may be rate-limiting for division (Bramkamp et al., 2006). More recent evidence indicates that FtsLB is an active participant in the decision making that controls the cell division process, playing a central role in triggering of septal peptidoglycan synthesis. This hypothesis is supported by observations that point mutations in a particular region of FtsL and FtsB alter the tightly regulated cell division process, allowing it to proceed even in situations in which normally it would not occur (Liu et al., 2015; Tsang and Bernhardt, 2015). Tsang and Bernhardt (Tsang and Bernhardt, 2015) discovered that a single point mutation in FtsL (E88K) allows the cells to bypass the normally strict requirements for other division proteins, namely FtsK, ZipA, FtsN, and FtsA (although residual levels of FtsA expression appear to be still necessary). Independently, De Boer and colleagues (Liu et al., 2015) identified a series of mutants that bypass the need for FtsN. These mutants map to two short regions of the coiled coil of FtsL

(residues 88–94) and FtsB (residues 55–59). These regions, which are located ~30 residues past the TM domain, were named the “constriction control domain” (CCD) (Liu et al., 2015).

The observed properties of these CCD mutants are consistent with a conformational change in FtsLB that has become deregulated. In other words, an OFF/ON structural transition in FtsLB may be part of the events that control the beginning of cell constriction. This transition may be triggered allosterically by FtsN itself, as suggested by the observation that the essential region of FtsN (^{E}FtsN, a short stretch of sequence that is separated by ~20 amino acids from the C-terminal side of the TM domain) is in a position that is topologically equivalent with the location of the CCD in FtsLB (Liu et al., 2015; Tsang and Bernhardt, 2015).

To elucidate the effect of the CCD mutants and the overall function of the FtsQLB complex in molecular detail, it is necessary to understand its structural organization. So far, two partial structures have been obtained. The structure of the periplasmic domain of FtsQ has been solved by X-ray crystallography, without the TM and cytoplasmic domains (van den Ent et al., 2008). Additionally, we previously determined the crystal structure of the coiled coil domain of FtsB and also produced a computational model of the TM domain of the same protein, based on mutagenesis data (LaPointe et al., 2013). Both structures were obtained in the absence of FtsL and in a homodimeric form. In addition, computational models of the FtsQLB periplasmic region (both in the trimeric and hexameric state) have also been reported (Villanelo et al., 2011). These models were based on available experimental data regarding contacts between FtsQ and FtsLB, whereas the FtsLB component was predicted *ab initio*. Additionally, these models lacked the TM region.

Here, we present a structural analysis of the TM domain and periplasmic coiled coil of FtsLB, based on a combination of biophysical and computational methods with *in vivo* validation. The work includes *in vitro* FRET to determine the oligomerization state of FtsLB; a co-evolutionary analysis of sequence alignments to provide an extensive set of predicted contacts between positions in FtsL and FtsB; molecular modeling to compute a three-dimensional structural model of the complex; molecular dynamics in lipid bilayers to evaluate stability and dynamics of the model; and a functional analysis of mutants *in vivo* to experimentally test the model and investigate the biological relevance of the predicted structural features.

Our results confirm that FtsLB is a higher oligomer and point to an L_2B_2 tetramer. Overall, the phenotypic analysis is in good agreement with structural features identified computationally. The distribution of deleterious phenotypes among the TM region mutants is consistent with the predicted interface. We confirm that the helix of FtsB breaks between the TM and periplasmic regions, with the formation of a flexible linker, as hypothesized previously (LaPointe et al., 2013). In contrast, we show that FtsL forms a continuous helix and that the integrity of this helix is crucial for function. We also found indications that the coiled coil might not be built for structural stability, at least not in the form of a canonical tetrameric helical bundle assembly, suggesting that the coil may either serve as a dynamic structural unit or else split into two independent domains.

2.3 Results and discussion

2.3.1 Co-evolutionary analysis identifies potential quaternary contacts consistent with an extended helical bundle

To predict the positions that mediate the association of the FtsLB complex, we analyzed a paired alignment of FtsL and FtsB sequences from proteobacteria species using the EV-

Couplings algorithm (Marks et al., 2011). EV-Couplings uses a maximum-entropy model of the evolutionary history of a protein (or protein complex) to infer potential tertiary (Marks et al., 2011) and quaternary (Hopf et al., 2014) contacts in its structure. It is based on the notion that, to maintain complementary interactions, amino acid changes influence the variation of positions that are in close proximity. The results of our EV-Couplings analysis are shown in Figure 2.2a. The figure displays the top 95 pairs of positions identified between FtsL and FtsB. The complete ranked list is provided in Table S2.1. FtsB-FtsB and FtsL-FtsL pairs are displayed in Figure S2:1.

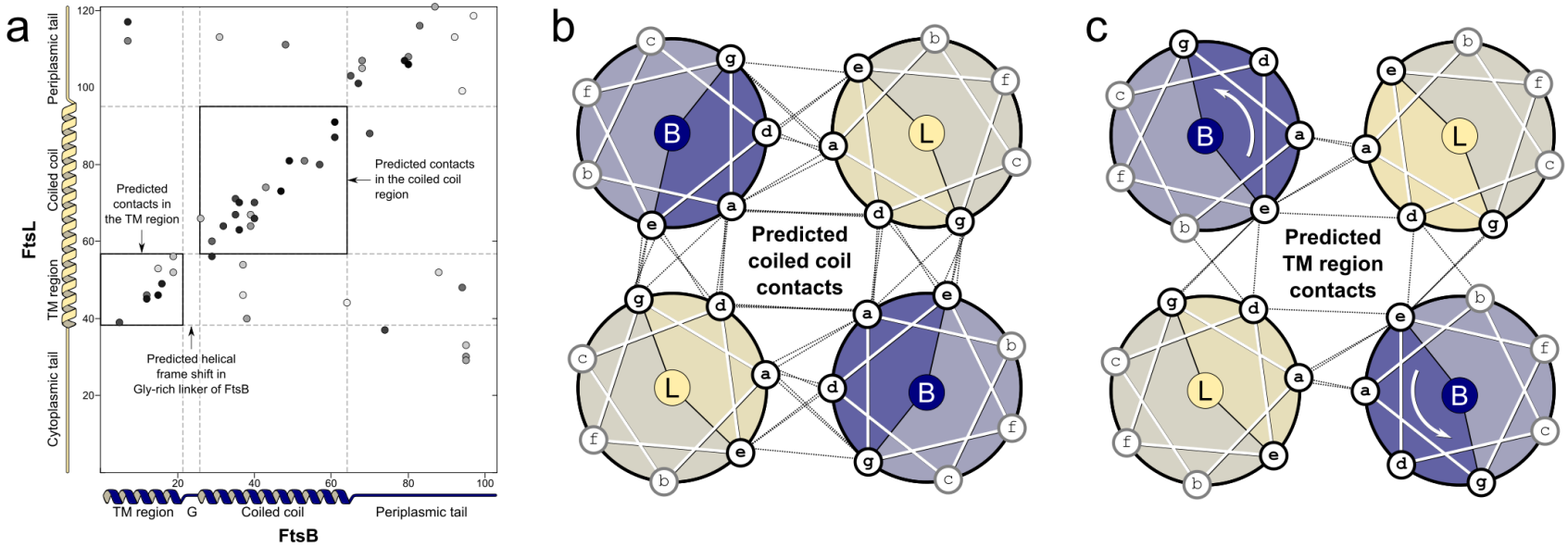

Figure 2.2: Co-evolutionary analysis identifies a self-consistent map of contacts in the FtsLB helical bundle. a, map of potential contacts between positions in the sequence in FtsB (x axis) and FtsL (y axis), inferred by co-evolutionary analysis. Displayed are the top 95 pairs identified by the analysis (the darkness of the shading reflects their ranking). The complete ranked list is provided in Table S2.1. The TM region and coiled coil are highlighted by boxes. Marked in the sequence of FtsB is also the position of a Gly-rich region between the TM and coil domains (G). b, predicted contacts in the coiled coil domain between FtsB (blue) and FtsL (yellow) displayed using a helical-wheel projection. The contacts are in excellent agreement with the expected contacts for a canonical coiled

coil. c, similar map for the TM region. The contacts are also consistent with the formation of a helical bundle. The interfacial positions of FtsL are on the same face of the interfacial position in the coiled coil, suggesting that FtsL forms a continuous helix. The positions in FtsB are rotated by ~60° with respect to those of the coiled coil (white arrow), indicating that a discontinuity is probably present between the two helical domains.

A notable diagonal pattern of co-evolving positions starts in the TM region of both proteins and continues in the coiled coil region, which is consistent with the contact map expected for an extended parallel helical bundle. As shown in the helical wheel diagram of Figure 2.2*b*, the co-evolving positions in the periplasmic region are also consistent with the expected geometry of a coiled coil (Buddelmeijer et al., 2002; Ghigo and Beckwith, 2000; Masson et al., 2009), occurring at positions of the "heptad repeat" (*abcdefg*) that are interfacial, at either the buried *a* and *d* positions or at the partially solvent exposed *e* and *g* positions. Moreover, these predicted contacts occur between pairs that are in proximity in a canonical coiled coil (*e.g. a-a*, *a-d*, and *g-e*, but not *a-e*). The pattern of connectivity remains in good agreement with a coiled coil contact map until positions 61 in FtsB and 91 in FtsL, after which it loses this periodicity. This range covers approximately five heptad repeats, which for FtsB is ~10–20 amino acids shorter than the consensus of the sequence-based predictors Paircoil (McDonnell et al., 2006), Multicoil (Wolf et al., 1997), Marcoil (Delorenzi and Speed, 2002), and Coils (Alva et al., 2016; Lupas et al., 1991) (inferring the length of the coiled coil of FtsL with the same methods is difficult because it is poorly predicted (Ghigo and Beckwith, 2000)). The coiled coil predicted by the evolutionary analysis approximately reaches the CCD (55–59 in FtsB, 88–94 in FtsL, *highlighted* in *yellow* in

Figure 2.2, *a* and *d*), suggesting that it is possible that this critical regulatory region occurs toward the end of the coil and thus in proximity of a structural transition.

2.3.2 A continuous FtsL helix and a discontinuous FtsB

As illustrated in the helical wheel diagram of Figure 2.2*c*, the TM region also displays a pattern of predicted contacts consistent with a bundle of parallel helices. The co-evolving positions are clustered toward the C-terminal side of the TM domains. Remarkably, the interfacial positions of the TM domain FtsB are in excellent agreement with those identified previously from their sensitivity to mutagenesis in a self-association assay (LaPointe et al., 2013).

Comparison of the contact maps (Figure 2.2, *b* and *c*) indicates that, in FtsL, the predicted interfaces of coiled coil and TM domains occur on the same face of the helix (involving the *a*, *d*, and *g* positions in the TM domain and the *a*, *d*, *e*, and *g* positions in the coil). This pattern indicates that FtsL forms a continuous helix across the membrane and the periplasmic region. A different outcome is observed for FtsB, for which the TM domain interface involves *a* and *e* but also *b* positions (*versus a*, *d*, *e*, and *g* positions in the coil) and thus appears rotated by ~60° with respect to the periplasmic region (as indicated by the *arrow* in Figure 2.2*c*). This rotation suggests the presence of a helical break. Rearrangements of the periodicity from the canonical heptad repeat are not uncommon in long coils. The 60° shift of the relative orientation of the interface between the TM and coiled coil regions of FtsB would correspond to the insertion of four amino acids (*abcdabcdefg*), which is designated as a “stutter” (Brown et al., 1996; Hicks et al., 2002). Such mismatches influence the coil's local structure and can possibly facilitate conformational changes necessary for function (Schmidt et al., 2017; Strelkov et al., 2004). In

this particular case, however, the presence of a conserved Gly-rich region (^{22}GKNG^{25} in *E. coli* and ^{22}GKGG^{25} in the consensus sequence of α-, β-, and γ-proteobacteria (LaPointe et al., 2013)) suggests that the juxtamembrane linker of FtsB is more likely to adopt a flexible and extended conformation instead of a local distortion of the helix.

2.3.3 FtsLB is a higher oligomer, probably an L_2B_2 tetrameric complex

To build a structural model for the FtsLB complex based on the evolutionary constraints, it was necessary to establish its oligomeric state. In previous work, we demonstrated that the isolated TM region of the FtsLB complex assembles to form a higher-order oligomer consisting of an equal number of FtsL and FtsB subunits, but we were unable to distinguish between heterotetrameric (L_2B_2), heterohexameric (L_3B_3), or even higher oligomeric forms (Khadria and Senes, 2013). Here, we assessed the stoichiometry of a construct that includes both periplasmic and TM domains by FRET analysis *in vitro*.

The FtsLB complex was overexpressed in *E. coli*, consisting of an N-terminally His-tagged FtsB and an N-terminally Strep-tagged FtsL. FtsL and FtsB were co-expressed. The proteins co-purify over sequential Ni-NTA and streptavidin columns in DM and DDM detergents, indicating that they form a stable complex with a density over Coomassie staining compatible with a 1:1 stoichiometric ratio (Figure S2:2*a*). A C-terminal Cys residue was introduced in either FtsB or FtsL, and protein samples were separately labeled with either donor (Cyanine 3; Cy3) or acceptor (Cyanine 5; Cy5) fluorophores. Because the stoichiometric analysis is performed separately for the two proteins, to further confirm interaction between FtsL and FtsB in the

experimental conditions, we mixed and equilibrated $FtsL_{unlabeled}$-$FtsB_{Cy3}$ and $FtsL_{Cy5}$-$FtsB_{unlabeled}$ samples, which produced a distinct FRET signal (Figure 2:2*b*).

We analyzed the stoichiometry of FtsLB by adopting a FRET method based on changing the relative fraction of donor- and acceptor-labeled molecules while the total protein concentration was maintained constant (Adair and Engelman, 1994). In these conditions, the theoretical variation of donor quenching (Q) depends on the oligomeric state of the complex, being proportional to $(1 - {P_D}^{n-1})$, where P_D is the relative fraction of donor-labeled protein, and n is the number of subunits in the complex (see Equation (8) under "Experimental procedures"). Due to the power law, no variation of donor quenching is expected for monomers as the donor fraction decreases, a linear increase of Q is expected for dimers, and a curved relationship is expected for trimers and higher oligomers. We performed two independent experiments, labeling only one of the two proteins with donor and acceptor fluorophores while leaving the other unlabeled. By labeling only either the FtsL or the FtsB moieties of the FtsLB complex, the procedure allowed us to calculate the number of subunits of FtsL and of FtsB individually.

As illustrated in Figure 2.3 (*left*), donor quenching increased linearly for FtsL-labeled samples when the acceptor fraction was increased. We used the sum of residuals between the experimental data and the models to compare the fits. The residual of the two-subunit linear model (*dashed line*, residual (r) = 0.001) is 10-fold smaller than the residuals of the three-subunit model (*continuous line*, r = 0.01), indicating that two FtsL molecules are present in the FtsLB complex. The quenching data of FtsB (Figure 2.3, *right*) also display a progressive increase of donor quenching, establishing that at least two FtsB subunits are present in the FtsLB complex. However, the fit to two subunits (r = 0.0055) is only marginally better than the fit to three

subunits (r = 0.0064); therefore, the data cannot differentiate between these two models, as in the case of FtsL. The fit, however, can reasonably rule out a four-subunit model (*dotted line*, r = 0.02).

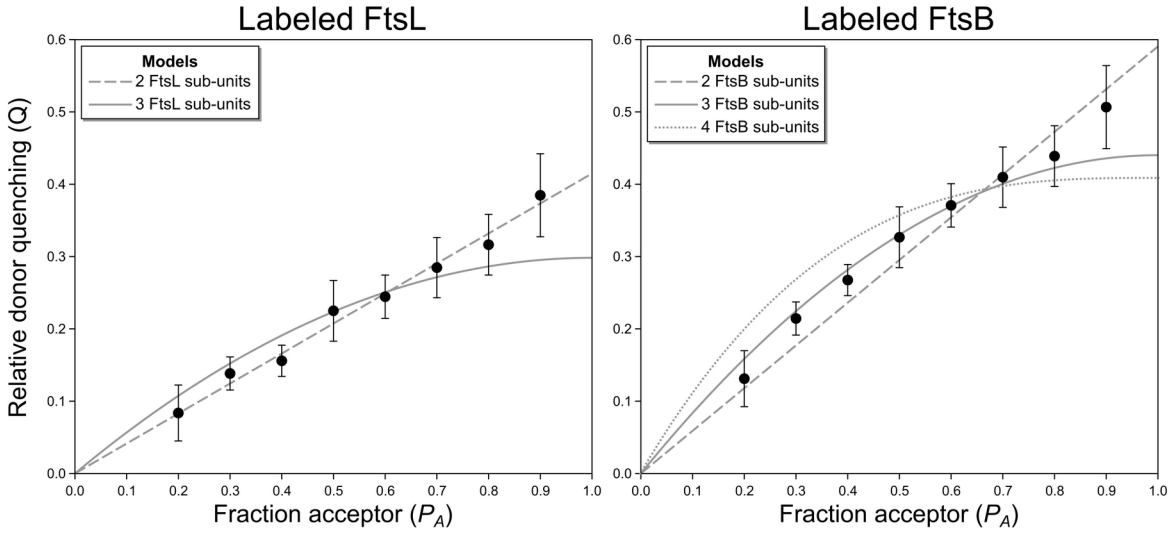

Figure 2.3: FRET analysis indicates FtsLB is a 2:2 tetramer. Analysis of the stoichiometry of Cy3- and Cy5-labeled FtsLB complexes was performed in DDM detergent. Two independent analyses were performed to count the subunits of FtsL (left) and FtsB (right) by labeling only one of the two proteins of the complex. Left, the FRET data of FtsL fit well to a two-subunit stoichiometry (dashed line, sum of residuals r = 0.0010) with respect to a three-subunit model (continuous line, r = 0.0111). Right, the fit for FtsB establishes that there are at least two subunits but does not differentiate between two- and three-subunit models (r = 0.0055 and 0.0064, respectively). A four-subunit model can be excluded (dotted line, r = 0.0224). Considering prior evidence and likely symmetry, the most likely model for FtsLB is a 2:2 heterotetramer.

Overall, the analysis confirms that the FtsLB complex is a higher-order oligomer. Considering the entire body of available evidence, the most likely model is an L_2B_2 heterotetramer. Because FtsL fits well to a model containing two subunits, the possible states supported by the data are

L_2B_2 or L_2B_3 stoichiometries. However, a 2:3 stoichiometry is in disagreement with previous FRET data that indicated that the FtsLB complex has an equivalent number of FtsL and FtsB subunits (Khadria and Senes, 2013). A 2:3 pentamer is also a less common oligomeric form in nature than a 2:2 heterotetramer (Levy et al., 2006). In addition, a 2:3 pentamer would be necessarily asymmetrical, whereas the co-evolutionary contacts are strongly consistent with a standard symmetrical organization (Figure 2.2, *b* and *c*). Finally, 2:2 would also be consistent with a previous estimate of the stoichiometry of the divisome based on ribosome profiling data (Egan and Vollmer, 2015).

2.3.4 Molecular modeling of the FtsLB helical bundle

We used a Monte Carlo docking procedure guided by the evolutionarily based distance constraints to create an all-atom computational model of the structure of the TM and coiled coil domains of the tetrameric complex (residues 1–61 of FtsB and 35–91 of FtsL). The model, which is illustrated in Figure 2.4*a*, consists of a four-helix bundle that spans the membrane and projects into the periplasmic region for ~55 Å. The structure is well packed, with the exception of a small void (~12 $Å^3$) present at the level of the juxtamembrane linker region. A majority of the co-evolutionary distance constraints used to derive the model (*dashed lines* in the figure, listed in Table S2.1) are satisfied; of the 27 side chain pairs involved, 22 are in contact (minimal distance between heavy atoms (d_{min}) < 5 Å), four are in proximity (d_{min} < 10 Å), and only one pair is separated by over 10 Å (FtsL Gln-66 with FtsB Ile-26). Because an L_3B_3 hexamer could not be entirely excluded, we also modeled this stoichiometry. The co-evolutionary restraints were well satisfied by the resulting model, but the model is significantly underpacked, with the presence of

an incomplete pore spanning parts of the TM and coiled coil domains (Figure S2:3). For this reason, the L_3B_3 model was no longer pursued.

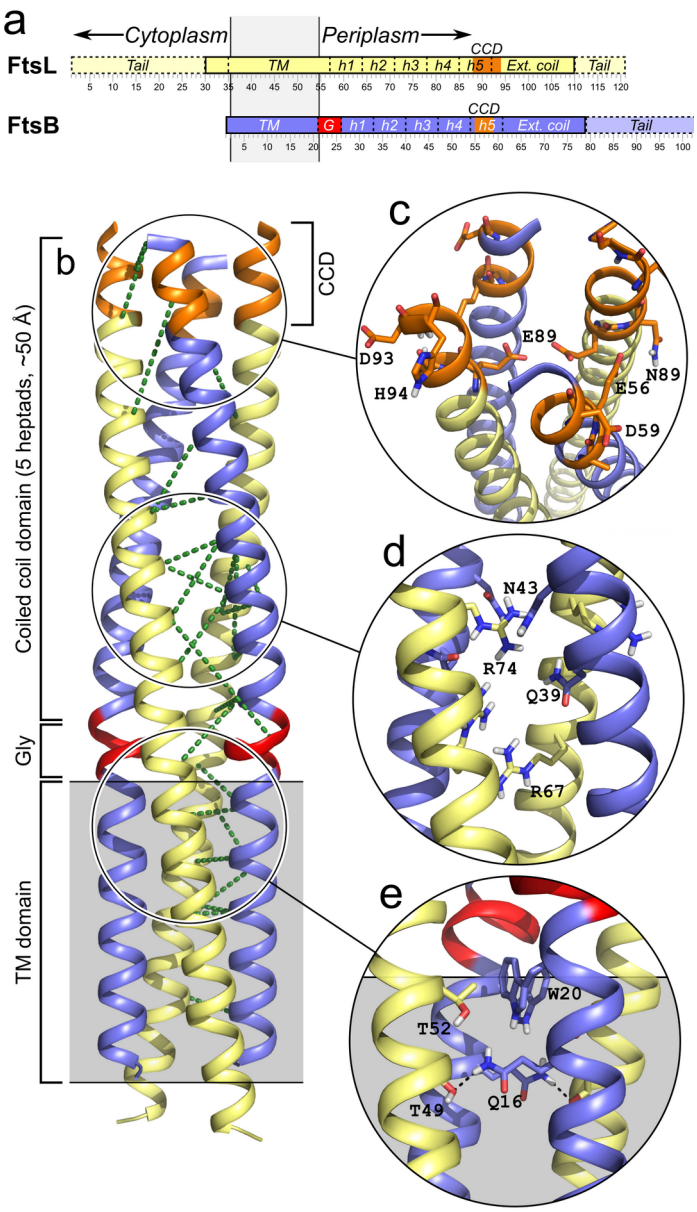

Figure 2.4: Computational model of FtsLB: A bundle with a helical break in FtsB. a, schematic representation of the domains of FtsL and FtsB as discussed in this work. The five heptad repeats in the coiled coil are designated as h1, h2, etc. Ext. coil, extended coil region, as defined for the MD analysis. G (red), Gly-rich linker in FtsB. The tails that are

absent in the modeling are represented with dashed lines. The position of the CCD region is marked in orange. b, ribbon representation of the model of the FtsLB complex. The dashed lines represent the co-evolutionary constraints applied to obtain the model. FtsL (yellow) consists of continuous helices that traverse the membrane into the periplasmic region. The helix of FtsB (blue) is modeled in an unwound conformation in the juxtamembrane Gly-rich region (red). c, detail of the CCD region. If the coiled coil conformation persists beyond this region, the polar side chain of Glu-88 would be buried in the coil's core. d, the lower coiled coil also contains a cluster of very polar amino acids buried in the core (Gln-39 and Asn-41 from FtsB; Arg-67 and Arg-74 from FtsL). These amino acids are likely to be destabilizing if buried in a canonical coiled coil, as represented by the model. e, the TM region contains a cluster of amino acids able to form hydrogen bonds (Gln-16, Tyr-17, and Trp-20 from FtsB; Thr-49, Thr-52, and Thr-56 from FtsL), which may contribute to stabilize this domain.

The TM region of the L_2B_2 model forms a helical bundle characterized by a left-handed crossing angle (7° for FtsB and 10° for FtsL, tilt angle of the helices with respect to the membrane normal). The two TM helices of FtsB are in closer proximity to each other compared with the two helices of FtsL (interhelical distance of 13.6 and 15.2 Å, respectively). Remarkably, the conformation of FtsB is similar to a model of a FtsB dimer that we obtained previously (RMSD of 2.5 Å; Figure S2:4). This previous model was obtained from the effect of point mutations on the homodimerization of the TM domain of FtsB in the absence of FtsL, as assayed with TOXCAT (LaPointe et al., 2013). The convergence of two completely independent sets of

data, evolutionary information and experimental mutagenesis, to a similar model is a strong indication that the conformation and interface of the TM helices of FtsB are correctly predicted.

The C-terminal side of the TM region contains a number of side chains that can form hydrogen bonds, including the polar Gln-16. In the lowest-energy model, Gln-16 acts as a hydrogen bond donor to the side chain hydroxyl group of Thr-49 of FtsL (Figure 2.4*b*). Small changes in conformation would allow alternative interactions of Gln-16 with a number of other donor and acceptor groups. In particular, Gln-16 could form self-interactions with Gln-16 from the opposing chain as well as potential hydrogen bonds with Thr-52 of FtsL and Trp-20 of FtsB.

As expected, FtsL was modeled as a continuous helix across the TM and coiled coil domains. Also as expected, the change of orientation of the interface between the same two domains of FtsB required the introduction of a break in the α-helix in the juxtamembrane region (*red* in Figure 2.4). We opted to model the linker in the least perturbing conformation (an unwound helix), although the Gly-rich linker is likely to adopt a more extended conformation. The relative flexibility of the linker regions of FtsL and FtsB will be addressed later using molecular dynamics.

The periplasmic region (residues 27–61 in FtsB and 58–91 in FtsL) produced a well-packed canonical coiled coil. The coil is approximately five heptad repeats long (~35 amino acids). Interestingly, the domain is unusually rich in polar amino acids occurring at *a* and *d* buried positions, with a total of 12. These polar amino acids are equally contributed by both proteins (three per subunit), although FtsB contains only neutral side chains (Gln-39, Asn-43, and Asn-50), whereas FtsL contains amino acids that are normally charged (Arg-67, Arg-74, and Glu-80). A region around the second and third heptad repeats of the coiled coil is particularly polar, where

Arg-67 and Arg-74 from FtsL and Gln-39 and Asn-43 from FtsB occur in close proximity. The presence of so many buried hydrophilic side chains is interesting because they are likely to destabilize the coiled coil. In particular, the four Arg residues contributed by FtsL would be charged even if buried inside the protein core and thus very costly to desolvate (Harms et al., 2011). Another potentially charged side chain, Glu-80 in FtsL, is in the core in the modeled complex, but it is placed toward the end of the predicted coiled coil, near the CCD region, and thus it may be solvent-accessible.

2.3.5 Molecular dynamics suggest a stable TM region with an intricate network of hydrogen bonding

Two features, the presence of a flexible linker and a potentially destabilized coiled coil, raise questions about the dynamic properties of the FtsLB complex. To address them, we performed molecular dynamics (MD) simulations of the FtsLB model in explicit POPE bilayers. For this simulation, we extended the coiled coil conformation by ~20 amino acids beyond the region predicted by the co-evolutionary contacts, to avoid end effects and to test the coil boundaries. We will refer to this added region as the “extended coil” (residues 92–110 for FtsL and 62–79 for FtsB) and to the section predicted by the co-evolutionary analysis as the “lower coil” (residues 29–61 of FtsB and 57–91 of FtsL). Three replica MD simulations were run for 260 ns (run 1; Figure 2.5) and for 200 ns (runs 2 and 3; Figure S2:5).

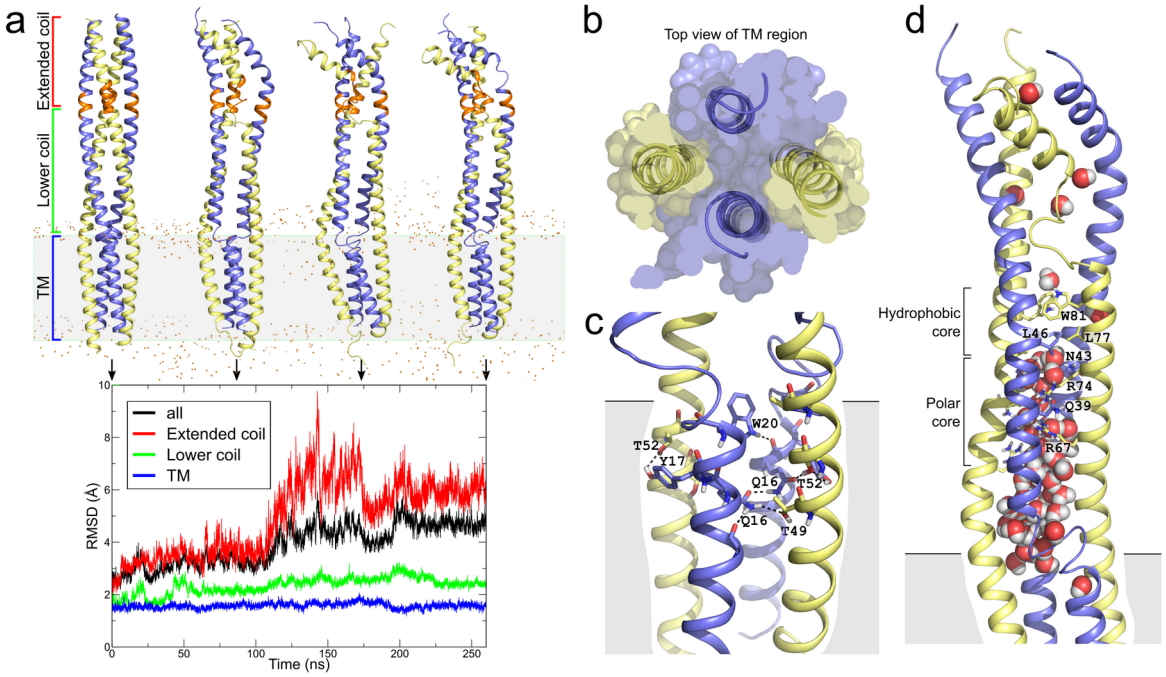

Figure 2.5: A stable TM region, a dynamic coiled coil. a, trajectory of the 260-ns run, with displayed conformations at 0, 87, 173, and 260 ns. The RMSD analysis is shown for the entire complex (black), the TM region (blue), the lower coil (green; 29–61 of FtsB and 57–91 of FtsL), and the extended coil (red, 62–79 of FtsB and 92–110 of FtsL). The CCD, which separates the lower from the extended coil, is in orange. As evident from the RMSD traces, the TM and lower coil are stable during the simulation, whereas the upper coil diverges, although it retains some helicity and interaction between pairs of FtsL and FtsB chains. The unfolding of the Gly-rich linker in the juxtamembrane region of FtsB is also evident. b, the TM region seen from above the membrane; the region rearranges during the simulation to bring the two FtsB helices (blue) closer to each other and in contact, whereas the FtsL helices are now separated from each other. c, a configuration of the extended

hydrogen bonding network present in the C-terminal side of the TM region, involving Gln-16, Tyr-17, and Trp-20 of FtsB and three threonine residues (Thr-49, -52, and -56) of FtsL. d, a view of the water that invades the core of the lower coiled coil, hydrating otherwise buried side chains or Arg-67, Arg-74, Gln-39, and Asn-43.

During the simulation, the RMSD from the initial structure increases to ~4, 5, and 6 Å in the three replica runs (*black traces* in Figure 2.5*a* and Figure S2:5). The majority of these changes are localized to the "extended coil" section (*red traces*), which quickly separates and partially unfolds. The lower coil remains relatively stable during the run (final RMSD around 2.3, 3.4, and 2.9 Å; *green traces*). The most stable region is the TM domain, which remains stable for the entire run across all three simulations, with average RMSDs around 1.6, 1.9, and 2.3 Å during the three runs (*blue traces*).

A slight rearrangement of the relative orientation of the TM helices is observed, which brings the two FtsB helices closer to each other by ~1–2 Å and, consequently, further separates the FtsL helices by a similar distance. This rearrangement gives the bundle a less "square" and more "rhomboid" configuration, one in which extensive packing occurs not only between FtsL and FtsB but also between the two FtsB helices, whereas the contacts between the two FtsL helices become reduced (Figure 2.5*b*). With the two helices of FtsB in closer proximity, their conformation becomes even more similar to our previous model of an FtsB dimer (RMSD ~2 Å; *Figure S2.4b*) (LaPointe et al., 2013).

A second important change that occurs in the TM region during the MD runs is a rearrangement and expansion of the hydrogen-bonding network. In the C-terminal section of

FtsLB, there are 12 side chains in close proximity that are able to hydrogen-bond (three from each helix: Gln-16, Tyr-17, and Trp-20 in FtsB and three threonines, Thr-49, -52, and -56, in FtsL). With the exception of Trp-20, all of these side chains have both donor and acceptor groups. In addition, Gln-16 is also very flexible. Therefore, the network can rearrange in multiple configurations. Figure 2.5*c* illustrates one of the configurations observed, which displays a total of seven interhelical hydrogen bonds. The configuration of the hydrogen-bonding network varies over the simulations, but some interactions are predominant during the runs (Table S2.2). The most persistent interaction is between Tyr-17 of FtsB subunit A and the side chain of Thr-52 of FtsL subunit C (Tyr-17 of subunit B prefers to interact with the backbone carbonyl of Leu-48). Gln-16 interacts primarily with the side chain hydroxyl groups of Thr-49 and Thr-52 from FtsL and with Gln-16 from the opposed FtsB helix. Trp-20 primarily donates to the side chain hydroxyl group of FtsL Thr-56, but it can also donate to the carbonyl group of Gln-16 side chain. This extended network of hydrogen bonds is likely to contribute significantly to the association of the TM region, which can form a stable oligomer in isolation *in vitro* (Khadria and Senes, 2013).

As a control, we also performed a 160-ns simulation of a structural model that has comparably low energy but does not satisfy the evolutionary constraints in the TM region (“bad” model; *Figure S2:5d*). The TM region rearranges away from the initial model relatively quickly, reaching an RMSD of >3 Å in the first 60 ns. Its average (2.8 Å) and maximum (3.5 Å) are higher than the RMSD of the TM region in the three replica runs of the “good” model (1.6, 1.9 and 2.3 Å, with maxima of 2.1, 2.4, and 2.7 Å, respectively).

2.3.6 A continuous FtsL helix and a dynamic coiled coil

Different sections of the coiled coil behave differently during the MD runs. These are notable in the helicity analysis presented in Figure S2:6. The “extended coil” (the region that extends beyond the pattern of co-evolutionary contacts consistent with a coiled coil (Figure 2.2)) unfolds partially as a bundle during the runs, although the region retains substantial helicity, and interactions still occur between pairs of FtsL and FtsB helices. The likely occurrence of a breakage of the helix between the “lower” and the “extended” coil regions was also hypothesized by a previous modeling analysis of the periplasmic region of the FtsLBQ complex (Villanelo et al., 2011). Interestingly, the location of the transition between these two regions roughly corresponds to the CCD.

The lower coil remains more stable during the run (final RMSD around 2.3, 3.4, and 2.9 Å). However, notable changes occur even in this region, providing further indication that the FtsLB complex may not be built to form a rigid, canonical, tetrameric coiled coil. Partial unfolding of the FtsL helix is observed in subunit C, and occasional unfolding is also notable in both FtsL subunits in the third replica run.

In the transition between the TM and coiled coil domains, the segment of 5–10 amino acids centered around the Gly-rich section of FtsB rapidly unfolds from the “unwound helix” conformation imposed by the modeling, unlinking the two domains and adopting a flexible and extended conformation (Figure 2.5*a*). Conversely, the juxtamembrane region of FtsL remains stable as a continuous helix throughout the duration of all three replica runs. The unfolding of the Gly-rich linker of FtsB creates an opening that allows water to access the core of the four-helix bundle (Figure 2.5*d*). Water molecules penetrate deeply up into the core of the coil, solvating the strongly polar side chains that would otherwise be buried in the structure (Arg-67 and Arg-74

from FtsL and Gln-39 and Asn-43 from FtsB, “polar core” in Figure 2.5*d*). In this region near the membrane, the coil becomes essentially separated by the water into two two-helix bundles. Water is excluded from the above layer, where the four-helix bundle becomes compact again around a hydrophobic core consisting of Leu-46 from FtsB and Leu-77 and Trp-81 from FtsL (“hydrophobic core” in Figure 2.5*d*).

With all caution in drawing conclusions from a theoretical model, it appears unlikely that a core so enriched in strongly polar amino acids would produce a very stable coiled coil. A possibility is that the stability of the coil is purposely “detuned” because the function of the FtsLB complex requires a weak coil, possibly to allow for a conformational change. A second hypothesis is that the region is actually designed to split into two separate two-helix coils. This split is most evident, as illustrated in the last frame of MD run number 3 (Figure S2:5). A “split” coiled coil would be consistent with the observation that a solubilized version of the periplasmic region of FtsLB forced into a heterodimeric form binds to FtsQ with high affinity (Glas et al., 2015; Masson et al., 2009; Noirclerc-Savoye et al., 2004), suggesting that association with FtsQ may be mediated by two independent binding domains of FtsLB. Such an organization would designate the TM region as the major factor driving tetramerization.

2.3.7 Functional analysis: The effect of mutations in the TM region is consistent with the predicted interface

To investigate how the model's structural features support the function of FtsLB, a series of rationally designed variants of the complex were tested *in vivo* for their ability to support cell division. Because FtsL and FtsB are essential proteins, the mutant proteins were introduced into strains in which a chromosomal copy of the wildtype protein is under the control of a repressible promoter (Gonzalez and Beckwith, 2009; Gonzalez et al., 2010). This allows for the depletion of

the wildtype copy and the induction of the mutant version to reveal its phenotype. The expectation is that the most severe mutations will produce cells that elongate but are unable to divide, resulting in the formation of very long filaments, whereas less severe mutations will allow the cells to divide but will produce subpopulations of elongated cells. To assess each mutant, we measured the distribution of cell lengths, as exemplified in Figure 2.6*a* for the L15A mutant of FtsB. The data for each individual mutant are reported in Figure S2:7.

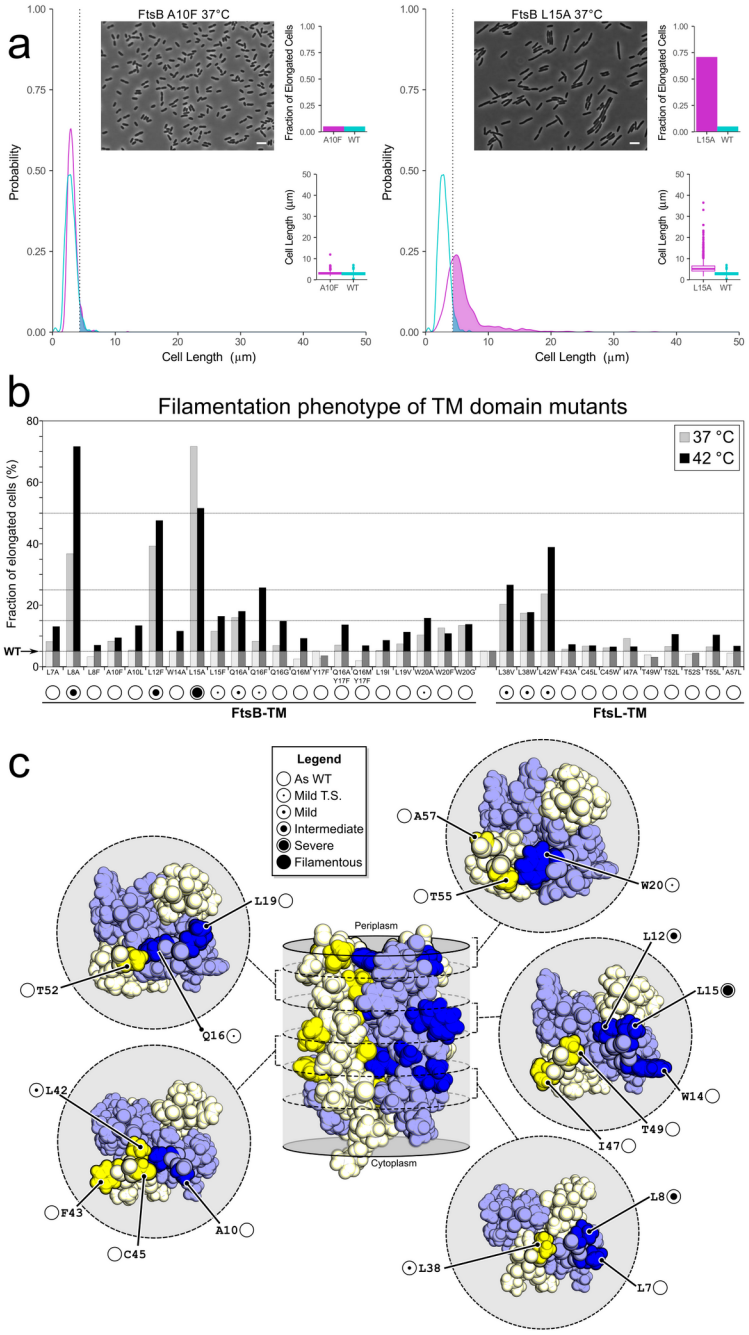

Figure 2.6: Mutations at the interface of the TM domain cause mild phenotypes in vivo. a, examples of in vivo analysis of point mutants with wildtype-like (A10F) and defective

(L15A) phenotypes. The distribution of cells lengths is compared between the wildtype (aqua) and the mutant (magenta). L15A displays 72% of cells that are longer than the 95th percentile in the wildtype distribution (colored areas, past the dotted line). As such, it is classified as a “severe” mutation. Graphs for each individual mutant are provided and explained in detail in Figure S2:7. Scale bar, 5 μm. b, phenotypes of TM domain mutants at 37 and 42 °C growth conditions. Classification is indicated using filled circles below. Filamentous, all cells are elongated at 37 °C; Severe, >50% of the cells are elongated; Intermediate, >25% of cells are elongated at 42 °C; Mild, >15% of cells are elongated at 42 °C; mild temperature-sensitive (Mild T.-S.), >15% of cells are elongated at 42 °C. c, location of the mutations within the structure of the TM domain. All mutations that display a cell division phenotype map within the interface of the helical bundle. All mutations on the outer surface of the bundle display As WT phenotypes.

Although none of the point mutations caused filamentous phenotypes, many produced elongated cells. To determine whether the difference between each mutant's and the wildtype's length distributions was significant, we used the Mann–Whitney *U* test (Mann and Whitney, 1947). However, as a consequence of the very large sample size of each experiment (500 to >1,000 cells), minuscule differences can be statistically significant even if they are not biologically relevant. To overcome this issue, we adopted an operative classification based on the fraction of elongated cells observed for each variant. We defined as “elongated” those cells that are longer than the 95th percentile in the distribution of wildtype cell length (Figure 2.6*a*). Based on this threshold, we defined phenotypes as “mild temperature–sensitive” (*Mild T.S.*) when

>15% (*i.e.* >3-fold compared with the wildtype) of cells were elongated only at the less permissive temperature of 42 °C; as “mild” when >15% of cells were elongated at 37 °C; as “intermediate” when >25% of cells were elongated (>5-fold); as “severe” when over 50% of the cells were elongated (>10 fold); and, finally, as “filamentous” when all cells were filamentous. The example of Figure 2.6*a*, which has 72% elongated cells at 37 °C, is therefore classified as “severe.”

The fractions of elongated cells for each variant in the TM region of FtsL and FtsB are reported in Figure 2.6*b*, with the relative classification reported at the *bottom*. The variants include a variety of drastic small-to-large amino acid changes (such as A10F in FtsB and L38W in FtsL), large-to-small changes (such as W14A and L8A in FtsB), and some conservative mutations (such as L19I in FtsB). We expected to observe division phenotypes when mutations affected the packing or hydrogen bonding at positions that mediate interaction between the helices. The results are in good agreement with this prediction. First, all positions that displayed impaired cell division phenotypes (Leu-8, Leu-12, Leu-15, Gln-16, and Trp-20 in FtsB; Leu-38 and Leu-42 in FtsL (Figure 2.6*b*)) occur at the helix-helix interfaces, as illustrated schematically in Figure 2.6*c*. In addition, all positions predicted to be away from the interface and exposed solely to lipids are classified as indistinguishable from wildtype (Leu-7, Ala-10, and Trp-14 in FtsB; Phe-43 and Ile-47 in FtsL). However, some of the TM region mutations that were predicted to be detrimental displayed little or no effect (*e.g.* Q16M, W20A, C45W, and T49W). This is particularly noticeable for the C-terminal side of the TM bundle, a section characterized by an extensive hydrogen bonding network (Figure 2.5*c*). Mutations of Thr-49, Thr-52, and Thr-55 of FtsL and Tyr-17 of FtsB have no effect, and those of Gln-16 and Trp-20 of FtsB have little or no

phenotype. Even a double mutation (Q16M/Y17F or Q16A/Y17F) that removes two hydrogen bonding groups is completely tolerated.

The finding is surprising because the C-terminal side of the TM bundle is the region with the highest number of co-evolutionary "connections," suggesting that these positions would be structurally or functionally important. Moreover, Gln-16 and Trp-20 of FtsB, are nearly absolutely conserved in proteobacteria (LaPointe et al., 2013). A potential explanation is that the hydrogen bonding network may be sufficiently robust, extensive, and plastic to accommodate changes. Because of the large network, the removal of some donors and acceptors may not be sufficiently detrimental to destabilize the complex. Conversely, the N-terminal region, which is mediated exclusively by packing of hydrophobic residues, could be more sensitive to disruption of the complementary side chain packing. It is also possible that thermodynamic stability of the TM region is not strictly required for function, at least in the conditions tested.

To investigate whether there is an overall correspondence between the observed biological phenotypes and predictions based on the structural model, we calculated mutational energies *in silico* for all TM mutants and compared them with the fraction of elongated cells observed for each variant (using the 42 °C data to maximize the dynamic range of the experimental observations). We calculated the energies using a rigid-body model without backbone movement as well as with FoldX, a method specifically calibrated to calculate the folding stability of protein mutants (Schymkowitz et al., 2005). The data are plotted in Figure S2:8. The energies produced by FoldX statistically correlate with the severity of the phenotypes by rank order Spearman correlation coefficient analysis ($r = 0.4631$, $n = 29$, $p < 0.01$) (Spearman, 1904). We conclude that perturbation of stability estimated on the basis of the structural model is a reasonably good predictor of the functional state of FtsLB.

2.3.8 The juxtamembrane and coiled coil of FtsB are tolerant to mutation

We applied a similar mutational approach to investigate the functional importance of the coiled coil and of the juxtamembrane region. The results are reported in Figure 2.7 (and Figure S2:7). We first tested whether the evolutionarily conserved Gly-rich juxtamembrane linker of FtsB (^{22}GKNG^{25}) is essential. A potential role for Gly is to provide structural flexibility, as suggested by our MD simulations. Alternatively, because of glycine's less restrictive Ramachandran distribution, this amino acid can also be required in a rigid structure to enable conformations of the backbone that are forbidden to all other amino acids.

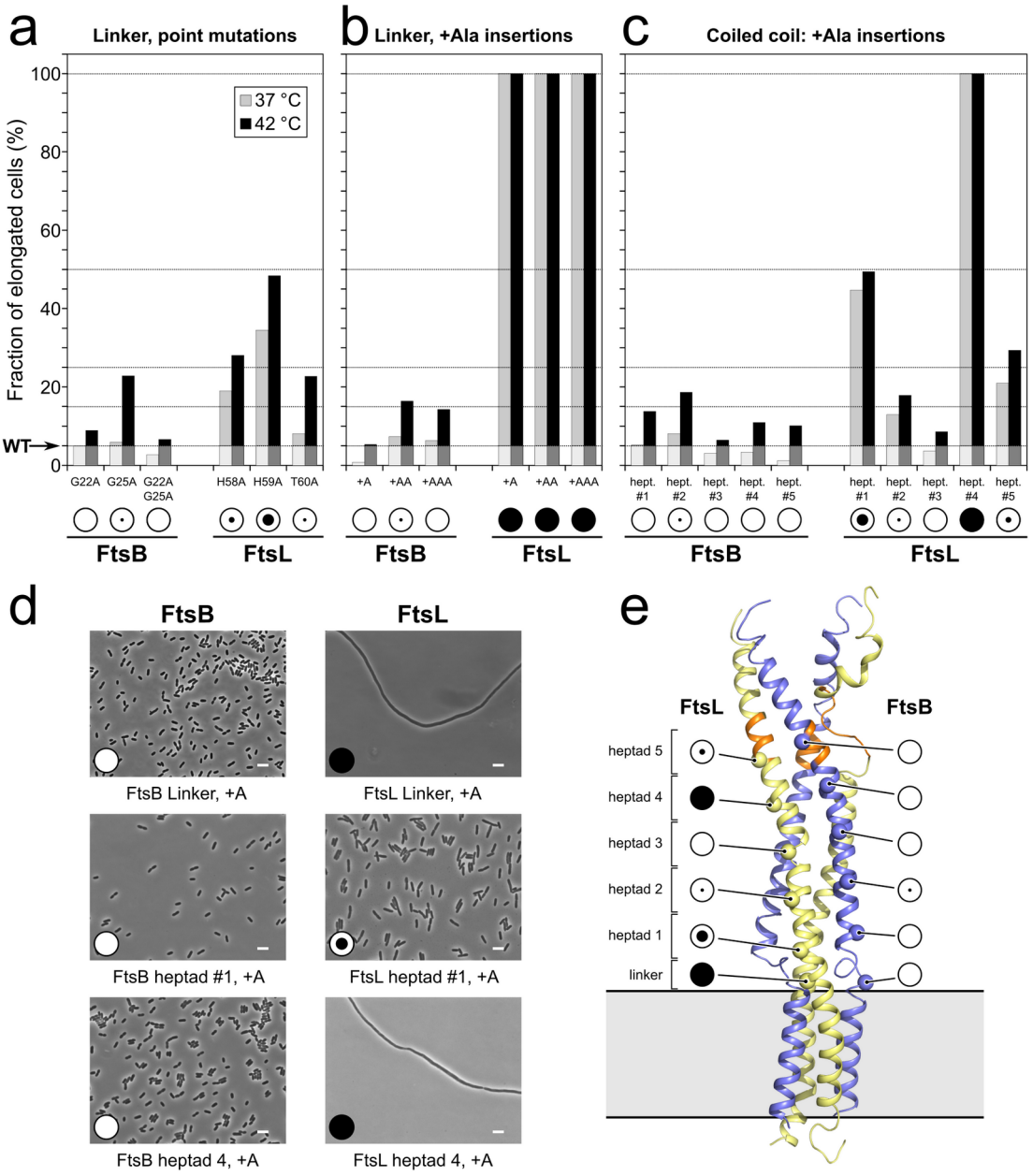

Figure 2.7: The juxta-membrane and coiled coil regions of FtsL are sensitive to mutation. a) Phenotypic analysis of juxta-membrane ("linker" region) point mutations of FtsL and FtsB, classified as detailed in Figure 2.6. Graphs for each individual mutant are provided in Figure S2:7. b) Alanine insertion mutation (one to three extra residues) in the same

region. c) Ala insertion mutations in the coiled coil region (operated between positions “b” and “c” in the heptad repeats).The evidence is consistent with a flexible linker, with the presence of a helical break, in FtsB, and with an uninterrupted helix that runs through the membrane and periplasmic regions for FtsL. Comparable expression levels of all variants displaying a phenotype were verified by Western blotting (Figure S2:9).

We tested individual and double Gly-to-Ala mutations at positions 22 and 25. They presented either no defects (G22A and G22A/G25A) or a mild TS phenotype (G25A) (Figure 2.7*a*). A possible explanation is that the structure of the linker retains sufficient flexibility even when Gly-22 and Gly-25 are substituted by Ala. To further test the flexibility of the linker, we inserted a series of Ala residues (up to three) between the TM domain and the coiled coil region (between positions Phe-21 and Gly-22), with the rationale that insertions should be better tolerated in a flexible region. The cells appear largely unaffected by the changes (only the two-Ala insertion mutation displays a mild TS phenotype; Figure 2.7*b*). The outcome is therefore consistent with a flexible FtsB linker and confirms the presence of a helical break between the membrane and periplasmic domains of FtsB. These observations are not consistent with the alternative hypothesis that Gly is required to enable a rigid backbone conformation.

We also applied this Ala insertion strategy to test the coiled coil region of FtsB. A single additional Ala residue was inserted between the *b* and *c* positions (*i.e.* in a position that is solvent-exposed) of each of the five heptad repeats. Most of these five insertion mutants did not display any impairment, with the exception of a mild TS phenotype for the second heptad

insertion after position Ala-37 (Figure 2.7*c*). Overall, the analysis suggests that the coiled coil of FtsB is likely to be quite plastic and able to tolerate insertion mutations all along its length.

2.3.9 Integrity of the juxtamembrane and coiled coil regions of FtsL is essential for function

Whereas the juxtamembrane and coiled coil domains of FtsB appeared tolerant to insertion, the opposite was observed for FtsL. We first introduced single, double, and triple Ala insertion mutations in the juxtamembrane region of FtsL (between positions Ala-57 and His-58). If the helix of FtsL is uninterrupted, the insertions should introduce strain into the structure because they would either produce a 100° rotation of the helix for each additional Ala or, more likely, introduce distortions into the helix. We observed that all insertion mutants, even the single Ala insertion, produced cells that were completely filamentous (Figure 2.7, *b* and cell images in *c*). These are the most dramatic phenotypes observed in the whole study. The results are consistent with the hypothesis that FtsL forms an uninterrupted helix that runs through the membrane and periplasmic regions.

We also tested the juxtamembrane linker of FtsL with a series of point substitutions (H58A, H59A, and T60A; Figure 2.7*a*). Interestingly, we observed some degree of functional disruption, particularly with the H59A mutation, which displays a 7-fold increase in the number of elongated cells compared with wildtype at 37 °C. In our model, His-59 is solvent-exposed and engaged only in minor interactions with the FtsB helix. It is possible that this position is important for interactions with other components of the divisome, possibly with the periplasmic domain of FtsQ, which was shown to cross-link with this region of FtsL (van den Berg van Saparoea et al., 2013).

Finally, we scanned the entire coiled coil region of FtsL by Ala insertion (Figure 2.7*c*), as we did with FtsB. We observed a severe phenotype for the first heptad insertion, a mild phenotype for the second heptad insertion, and a wildtype phenotype for the third heptad insertion. A completely filamentous phenotype is then observed when Ala is inserted in the fourth heptad repeat (after position Arg-82). It is difficult to speculate what might cause such a dramatic defect, but we note that the mutation is in proximity to the CCD region of FtsL (residues 88–94, highlighted in *orange* in Figure 2.7*e*). Insertion at position 89, which is within the CCD and which would correspond to the fifth heptad, if the coiled coil persisted in this region, produced a mild phenotype (~4-fold increase in elongated cells). Overall, the periplasmic coiled coil of FtsL is sensitive to mutation, whereas the same domain of FtsB is tolerant, as schematically illustrated in Figure 2.7*e*. Interestingly, this outcome is the opposite of that observed for the mutagenesis of the TM region, where FtsB was the sensitive subunit, compared with the much milder phenotype displayed by FtsL (Figure 2.6).

2.4 Conclusions

The topology of the FtsLB complex, a helical bundle spanning the membrane and periplasmic space, with terminal tails available for binding other components, is widely conserved evolutionarily (Buddelmeijer and Beckwith, 2004; Masson et al., 2009). The widespread occurrence of FtsLB across a broad variety of bacterial species with different cellular envelopes implies that this specific domain organization fulfills some important functional purpose, which is still unknown. To investigate this question, we have analyzed the structure-function relationship of the extended helical region of FtsLB, revealing a number of important features regarding the organization of the complex.

This study confirms that FtsLB is a higher-order oligomer and provides evidence that the complex consists of a heterotetramer. It identifies with a high degree of confidence the interface of the helical bundle region of the complex. It also produces a structural model of FtsLB validated experimentally through functional analysis. Further validation is also provided by a notable convergence of features in FtsB (the conformation of the TM helices and the presence of a flexible linker) that were also predicted in a prior model of FtsB based on a completely orthogonal set of experimental data (LaPointe et al., 2013).

The number of strongly polar, and even charged, amino acids that are predicted to be buried in the core of the coiled coil is a surprising and probably important finding, suggesting that the coil region may not be built for enhanced structural stability. A structural alternative to a monolithic four-helix coil could be the formation of two independent “dimeric” branches formed by one FtsL and one FtsB subunit. This possibility is consistent with previous observations that FtsLB fragments can bind to FtsQ with high affinity even when forced in a heterodimeric state (Glas et al., 2015; Masson et al., 2009; Noirclerc-Savoye et al., 2004).

We found that the integrity of the extended helix of FtsL is essential for the complex. A continuous helix is probably important for structural stability; given that the juxtamembrane linker of FtsB is disordered, a stable FtsL helix is probably necessary for nucleating the coiled coil. This hypothesis is in good agreement with the observation that amino acid insertions are most detrimental at the base of this domain. The helix of FtsL is also an interesting candidate for acting as a mechanical connection that could propagate a postulated allosteric conformational change (Liu et al., 2015; Tsang and Bernhardt, 2015) across the periplasmic, TM, and cytoplasmic regions. For example, the cytoplasmic region of FtsL is important for the recruitment of FtsW (Gonzalez et al., 2010); therefore, it is possible that FtsLB could control the

activation of the FtsWI peptidoglycan synthase complex through direct interaction by coordinating a signal from the periplasm (*i.e.* interaction with FtsN). This study provides the structural groundwork necessary for investigating this and other hypotheses with coordinated biophysical and functional studies, which are necessary to finally clarify the precise role and the molecular mechanisms of the FtsLB complex in bacterial cell division.

2.5 Experimental Procedures

2.5.1 Co-evolutionary analysis

Sequences of FtsL and FtsB homologues were collected using the DELTA-BLAST algorithm on the RefSeq database (O'Leary et al., 2016). These sequences were filtered to include only proteobacterial species and then concatenated by matching the organism name in each record. The paired sequences were then aligned using the ClustalW algorithm with five guide tree and five hidden Markov model iterations (Li et al., 2015). Columns in the alignment with a gap fraction higher than 0.3 were masked. This paired multiple sequence alignment was analyzed with the EV-Couplings algorithm (Marks et al., 2011) using default parameters and ignoring membrane topology, which would otherwise assume that sequential TM domains are anti-parallel. The top 95 (the number of unmasked columns in the multisequence alignment divided by 2) evolutionary constraints (*i.e.* the co-varying residues) predicted by EV-Fold between FtsL and FtsB were used for analysis and molecular modeling. The paired alignment is provided as a supplementary FASTA file. A total of 1291 sequences were used in the alignment. The number of effective sequences after reweighting for similarity is 883.1, corresponding ~4.6 sequences/residue. From this, ~40% of the top 95 contacts are estimated to be false positive pairs in the co-evolutionary analysis (Kamisetty et al., 2013).

2.5.2 Modeling the TM region of FtsLB

All modeling was performed using programs written in the Molecular Software Library (MSL) (Kulp et al., 2012). For the TM domain, ideal helices corresponding to residues 1–21 of FtsB and residues 35–58 of FtsL were generated. C_2 rotational symmetry (or C_3, in the case of the hexamer) was preserved around the *z* axis for FtsB helices and FtsL helices. The geometry of the FtsL and FtsB bundles was defined by the following parameters: interhelical distance (*d*); rotation around the helical axis (ω); crossing angle (θ), and position of the crossing point (*s*). Additionally, the orientation of the FtsL and FtsB bundles were changed by operating a rotation (Ω) and a translation (*S*) of each helix bundle about the *z* axis.

Starting from an initial random assignment of the parameters, the FtsLB TM helix complex was optimized using a Monte Carlo procedure that altered the interhelical geometry. The conformation of the side chains was periodically optimized with a 5% probability after each move. Side chain optimization was performed with a greedy trials algorithm using the backbone-dependent Energy-Based Conformer Library (bEBL) applied at the SL80 level (Subramaniam and Senes, 2012, 2014). Energies were calculated using the CHARMM 22 van der Waals function (MacKerell et al., 1998) and the hydrogen-bonding function of SCWRL 4 (Krivov et al., 2009) as implemented in MSL. Additionally, sigmoidal distance restraints were placed between the Cα atoms of each pair of top co-evolving positions (Ovchinnikov et al., 2015), using the following form:

$$E_{sigmoid} = \frac{\omega}{1+e^{-a(r-r_0)}} + C \qquad \text{Equation (5)}$$

Where $E_{sigmoid}$ is the additional energy term, *w* is the weight, *a* is the slope, *C* is the intercept, *r* is the distance between the atoms, and r_0 is the distance cutoff. For these experiments, *w* was set to 10 kcal/mol, *a* was set to 0.5 $Å^{-1}$, *C* was set to −2.5 kcal/mol, and r_0 was set to 10 Å.

For each pair of co-evolving residues, there are multiple pairs in the complex corresponding to the same residues on different chains. Only the restraints with the lowest energies for each pair were added to the energy score, whereas the remaining ones were masked. The evolutionary constraints are listed in Table S2.1. Models were sorted by energy and clustered using a greedy algorithm and a Cα RMSD threshold of 2.5.

As a negative control for the molecular dynamics simulations, a second model of the TM domain was produced by identifying a conformation with comparable energy of the initial model but with poor agreement with the co-evolutionary data (bad model). In this model, only one pair of co-evolving positions had a minimum heavy atom distance below 5 Å, and only two pairs had Cα distances below 10 Å, whereas all pairs satisfy these conditions in the good model.

In silico mutational energies for the mutants in the transmembrane region were calculated in two ways. Rigid-backbone repacking was performed using MSL. Point mutations of interest were generated, and neighboring side chains were repacked using 100 rounds of a greedy trials algorithm. Mutated residues were repacked at the SL99 conformer level of the bEBL library; residues within 8, 16, and 20 Å were repacked at the SL95, SL90, and SL80 levels, respectively. van der Waals radii were scaled by 0.8. Energies of the mutants were subtracted from that of the wildtype to calculate their $\Delta\Delta E$. Mutations were also analyzed using the FoldX Suite (Schymkowitz et al., 2005). The models were first energy-minimized using the *RepairPDB*

command with the membrane parameter set to true. Mutant $\Delta\Delta G$ calculations were performed using the *buildModel* command with the membrane parameter set to true.

2.5.3 Modeling the coiled coil domains of FtsL and FtsB

Supercoiled helices corresponding to residues 52–94 of FtsL and 21–63 of FtsB were generated by using a coiled coil generator based on a coiled coil parameterization described previously (North et al., 2001). The superhelical radius (r_1), superhelical pitch (P), helical rotation (Φ_1), and z-shift (s) of both FtsL and FtsB were freely altered, whereas the rise per residue (h) and helical radius (r_0) were kept constant. Additionally, the orientation of the FtsL and FtsB bundles were changed by operating a rotation (Ω) and a translation (S) of each helix bundle about the z axis. C_2 symmetry (C_3 for the hexamer) was preserved within the FtsL backbones and the FtsB backbones.

The coiled coils were optimized using a Monte Carlo procedure, changing their superhelical parameters starting from an initial assignment of random parameters. Side-chain conformational sampling was performed with a variable number of conformers with the positions that participate at a canonical coiled coil interface receiving higher sampling; the *a* and *d* positions were sampled at the SL80 conformer level of the bEBL library, *e* and *g* positions at the SL75 level, *b* and *c* at the SL70 level, and *f* at the SL60 level. Energies were calculated based on CHARMM 22 van der Waals and CHARMM 22 electrostatic terms. Additionally, sigmoidal restraints for each co-evolving pair in the coiled coil region were added, as described above.

2.5.4 Modeling the juxtamembrane regions of FtsL and FtsB

Top models of the TM and coiled coil domains were connected by aligning the helical residues 52–58 of FtsL, which were present in the models of both regions. The RMSD between the Cα atoms of these residues was minimized while keeping the main axis of both domains parallel to and centered on the *z* axis. The juxtamembrane regions of FtsL and FtsB were then replaced with loops corresponding to fragments from the PDB, as described previously (LaPointe et al., 2013). For FtsB, six-residue loops, corresponding to positions 21–26, with four flanking helical residues on each side, were used, with an additional sequence requirement that the fragment contain at least one glycine. For FtsL, 15-residue fragments with four flanking helical residues on each side were used with the requirement that the loop have helical secondary structure. Long helix fragments were used to better distribute minor deviation in alignment between the transmembrane and coiled coil helices. The connecting regions were optimized based on a greedy trials algorithm to minimize steric clashing, and the final model was minimized using BFGS constrained optimization using CHARMM (Brooks et al., 1983). Structural voids were analyzed using BetaCavityWeb (Kim et al., 2015). For the creation of the bad model, the TM model that did not satisfy the co-evolutionary constraints was aligned and connected to the same model of the coiled coil domain using an identical procedure.

2.5.5 All-atom molecular dynamics simulations

For the molecular dynamics simulations, the model's coiled coil region was extended, to avoid edge effects, to residues 110 (FtsL) and 79 (FtsB). The cytoplasmic side of FtsL was also extended to include residues 30–34, modeled in ideal α-helix. Four all-atom MD simulations (a 260-ns run, two 200-ns replica runs, and a 160-ns control run on the control bad model) were performed using the CHARMM 36 force field and NAMD version 2.10 software (Klauda et al.,

2010; Phillips et al., 2005). CHARMM-GUI membrane builder (Jo et al., 2008) was used to prepare systems composed of a POPE bilayer consisting of 301 lipids, the FtsLB tetramer, an ionic concentration of 0.150 m NaCl, and 59,034 TIP3P water molecules for hydration. The sizes of the boxes at the beginning of the simulation were ~97 × 97 × 242 $Å^3$ for runs 1, 2, and 3 and 97 × 97 × 245 $Å^3$ for the control run. The simulations were initially minimized and equilibrated for 75 ps at an integration time of 1 fs/step and for 600 ps at an integration time of 2 fs/step. The integration time step for the production runs of each of the systems was 2.0 fs/step. The simulations were carried out in the NPT ensemble at a pressure of 1 atmosphere and a temperature of 310.15 K, using the Nose–Hoover Langevin piston and Langevin dynamics method. Particle mesh Ewald was used for electrostatic interactions, and a 12-Å cutoff was applied to Lennard–Jones interactions with a switching function from 10 to 12 Å. The RMSD analysis was performed using the RMSD trajectory tool in VMD (Humphrey et al., 1996). Hydrogen-bonding analysis was performed with an in-house script.

Helicity analysis was performed by measuring the backbone dihedrals and the distance between O_i and N_{i+4} for each residue in each selected frame of the simulation. If the O_i–N_{i+4} distance was between 2.0 and 4.2 Å and the backbone dihedral angles were within the favored α-helical region as defined in PROCHECK (Laskowski et al., 1996), the residue was classified as helical. If the O_i–N_{i+4} distance was within 5 Å and the backbone dihedral angles were within either the favored or allowed α-helical region, the residue was classified as near-helical. Otherwise, the residue was classified as non-helical.

2.5.6 Cloning, expression, purification, and labeling of FtsLB constructs for FRET measurements

The His-tagged FtsB and Strep-tagged Cys-less (C41A and C45A) FtsL(35–121) (Table S2.3) were introduced into a modified pETDuet-1 vector at restriction sites NcoI/HindIII and NdeI/XhoI, respectively. For fluorophore labeling, cysteine mutations were introduced either to FtsB (S97C) or to Cys-less FtsL (I100C) via QuikChange mutagenesis (Novagene). All constructs were confirmed by DNA sequencing (Quintarabio).

The plasmids were transformed into BL21 (DE3) cells. Individual colonies were picked and grown overnight in 3 ml of LB broth containing 100 μg/ml of ampicillin before being diluted 1:500 in 1 liter of ZYP-5052 autoinduction medium as described (Studier, 2005), and grown at 37 °C until reaching an A_{600} of 0.8, after which it was incubated for an additional 20 h at 25 °C. The cells were then lysed by sonication in 10 ml/g lysis buffer (50 mm NaCl, 50 mm HEPES, 10 mm TCEP, brought to pH 7.0 with NaOH) supplemented with 0.25 mg/ml lysozyme, 5 mm β-mercaptoethanol, and 1 mm phenylmethylsulfonyl fluoride. The inclusion body fraction was separated by centrifugation at 10,000 × *g* for 20 min, followed by ultracentrifugation of the supernatant at 130,000 × *g* for 30 min to isolate the cell membranes. The FtsLB complex was then extracted from the membrane fraction with lysis buffer supplemented with 18 mm *n*-decyl-β-d-maltopyranoside (DM; Anatrace) and 10 mm TCEP, rocking at 4 °C overnight. Solubilized protein was added to 3 ml of Ni-NTA-agarose resin (Qiagen) and rocked for 2 h at 4 °C before the resin was washed and the complex was labeled on-column with either Cy3 or Cy5 maleimide (Lumiprobe) for FtsB or FtsL. On-column fluorophore labeling was performed by running 3 column volumes of Ni wash buffer (300 mm NaCl, 25 mm HEPES, pH 8.0, 50 mm imidazole, 10 mm TCEP, brought to pH 7.0 with NaOH) supplemented with 5.4 mm DM, 3 column

volumes Cy3 or Cy5 labeling buffer (100 μm Cy3 or Cy5 maleimide, 5.4 mm DM, 50 mm NaCl, 50 mm HEPES, pH 8.0, brought to pH 7.0 with NaOH), 6 column volumes of Ni wash buffer supplemented with 450 μm *n*-dodecyl-β-d-maltoside (DDM; Avanti Polar Lipids), and 2 column volumes of elution buffer (300 mm NaCl, 25 mm HEPES, pH 8.0, 450 μm DDM, 300 mm imidazole, brought to pH 7.0 with NaOH). For quantification, the elution fractions were dialyzed overnight at 4 °C against FRET buffer (300 mm NaCl, 25 mm HEPES pH 8.0, 450 μm DDM, brought to pH 7.0 with NaOH).

Labeling efficiency of each fraction was quantified by UV-visible spectroscopy, using absorbance at 280 nm (protein), 550 nm (Cy3), and 650 nm (Cy5). First, the concentration of the protein was calculated, taking into account the absorbance of the fluorophore:

$$[Protein] = \frac{A_{280} - (A_{Cy} \times CF_{Cy})}{\epsilon_{280}} \qquad \text{Equation (6)}$$

Where A_{280} is the absorbance at 280 nm, A_{Cy} is the absorbance at 550 nm (Cy3-labeled samples) or 650 nm (Cy5-labeled samples), ϵ_{280} is the molar extinction coefficient of FtsLB at 280 nm (32,430 $\mathrm{m^{-1}\ cm^{-1}}$), and CF_{Cy} is the appropriate correction factor (CF_{Cy3} = 0.11 and CF_{Cy5} = 0.05) to subtract the contribution of the fluorophore to absorbance at 280 nm. Then the labeling efficiency P_{Cy} was calculated according to the following:

$$P_{Cy} = \frac{A_{Cy}}{[Protein] \times \epsilon_{Cy}} \qquad \text{Equation (7)}$$

Where ϵ_{Cy} is the molar extinction coefficient of the fluorophore (ϵ_{Cy3} = 150,000 m^{-1} cm^{-1} at 550 nm, ϵ_{Cy5} = 250,000 m^{-1} cm^{-1} at 650 nm). Cysteine-less versions of the FtsL and FtsB constructs were also purified and labeled three times with the same protocol to determine background labeling, which was negligible.

A second round of purification could be performed by loading the Ni-NTA–purified FtsLB complex over a streptavidin column. The most concentrated Ni-NTA elution fractions were added to 3 ml of streptavidin resin and rocked for 2 h at 4 °C before being washed with 3 column volumes of streptavidin buffer (20 mm Tris, 300 mm NaCl, 0.1 mm EDTA, 0.5 mm TCEP, brought to pH 8.0 with NaOH) with 450 μm DDM, followed by elution with 3 column volumes of streptavidin buffer plus 2.5 mm desthiobiotin (initially dissolved in 1 ml of DMSO) and 450 μm DDM. Samples from each fraction of the nickel and streptavidin column were boiled in 4× SDS-PAGE loading buffer before being run on NuPage 4–12% BisTris protein gels at 150 V for 1 h (Thermo Fisher Scientific). Because the FtsLB complex obtained after Ni-NTA had sufficient purity for FRET studies, this second chromatography was not routinely performed, but it was used initially to confirm the stability of the FtsLB complex.

2.5.7 FRET measurements

In all experiments, the FtsLB complex was diluted to 1.35 μm in FRET buffer, which contains 450 μm DDM, for a final protein/detergent molar ratio of 1:333. Experiments were performed by labeling either the FtsL moiety (FtsL-C41A/C45A/I100C + FtsB) or the FtsB moiety (FtsL-C41A/C45A + FtsB-S97C) in the protein complex. All FRET analysis was performed on an M1000 Tecan plate reader, with excitation set at 550 nm and emission recorded from 560 to 800 nm in 1-nm increments.

The formation of FtsL-FtsB oligomers *in vitro* was assessed by measuring FRET on a 50:50 mix of $FtsL_{Cy3}$-$FtsB_{unlabeled}$ and $FtsL_{unlabeled}$-$FtsB_{Cy5}$ complexes, which were compared with a 50:50 mix of $FtsL_{Cy3}$-$FtsB_{unlabeled}$ and $FtsL_{unlabeled}$-$FtsB_{unlabeled}$ (donor-only sample) complexes and a 50:50 mix of $FtsL_{unlabeled}$-$FtsB_{Cy5}$ and $FtsL_{unlabeled}$-$FtsB_{unlabeled}$ (acceptor-only sample) complexes.

The stoichiometric analysis of FtsL was performed by mixing $FtsL_{Cy3}$-$FtsB_{unlabeled}$ (donor) and $FtsL_{Cy5}$-$FtsB_{unlabeled}$ (acceptor) samples in different ratios, from 20:80 donor/acceptor up to 90:10, in 10% increments. To determine donor fluorescence in the absence of the acceptor, equivalent samples were produced by mixing $FtsL_{Cy3}$-$FtsB_{unlabeled}$ (donor) and $FtsL_{unlabeled}$-$FtsB_{unlabeled}$ (unlabeled) in the same molar ratios. The stoichiometric analysis of FtsB was performed similarly. Donor (Cy3) fluorescence was recorded using its peak emission at 570 nm. Quenching (Q) of the donor fluorophore was calculated, as explained below, according to Equation (10), from donor/acceptor-labeled protein sample mixes (F) and donor/unlabeled protein sample mixes (F_0). Each experiment was independently replicated at least five times.

2.5.8 Fitting of experimental FRET data to obtain the number of subunits

Fitting for different oligomeric states was performed as described by Adair and Engelman (Adair and Engelman, 1994):

$$Q=\left(1-\frac{f_Q}{f_D}\right)\left(1-P_D^{n-1}\right) \qquad \text{Equation (8)}$$

Where Q is related to the molar fluorescence of the quenched donor (f_Q) in the presence of the acceptor, the molar fluorescence of the donor f_D in the absence of acceptor, the number of subunits in the oligomer n, and the molar fraction of donor-labeled protein P_D.

P_D was calculated as follows:

$$P_D = \frac{[D]}{[D]+[A]} \qquad \text{Equation (9)}$$

Where [D] is the molar concentration of donor-labeled protein and [A] is the molar concentration of acceptor-labeled protein.

Relative quenching was calculated as follows,

$$Q = 1 - \frac{F}{F_0} \qquad \text{Equation (10)}$$

Where F is the experimentally measured quenched fluorescence for a certain donor/acceptor molar fraction and F_0 is the experimentally measured unquenched fluorescence of the same amount of donor, obtained in the absence of acceptor and in the presence of an equivalent amount of unlabeled protein. In Equation (8), the quantity related to the molar fluorescence can be treated as an overall unknown constant k.

$$\left(1 - \frac{f_Q}{f_D}\right) = k \qquad \text{Equation (11)}$$

The parameter k was thus fit using a least square procedure to the experimental data Q as a function of donor fraction P_D according to Equation (12) for the different oligomeric states n.

$$Q(P_D) = k(1 - P_D^{n-1}) \qquad \text{Equation (12)}$$

The relative quality of the fits to different oligomeric states was assessed by comparing the sum of the squared residuals between the experimental data and the models.

2.5.9 Bacterial strains, plasmids, and media for in vivo experiments

The phenotypic analysis was performed using depletion strain NB946 for FtsB (Buddelmeijer et al., 2002) and MDG277 for FtsL (Gonzalez and Beckwith, 2009). For all experiments described, bacterial cells were grown in LB medium supplemented with 100 μg/ml spectinomycin (Dot Scientific) and the appropriate carbon source. Medium was supplemented with 0.2% (w/v) l-arabinose (Sigma) or 0.2% (w/v) d-glucose (Sigma) to induce or repress, respectively, the expression of chromosomal copies of the wildtype genes regulated by the PBAD promoter. 20 μm isopropyl-β-d-1-thiogalactoside was added to the media to induce the expression of mutant genes regulated by the pTrc promoter in the plasmid. Point mutants and insertion mutants of FtsL and FtsB were constructed in plasmid pNG162 (Gonzalez and Beckwith, 2009) using either standard QuikChange mutagenesis or inverse PCR mutagenesis.

2.5.10 Depletion strain experiments

The protocol for the depletion strain experiment was adapted from Gonzalez and Beckwith (Gonzalez and Beckwith, 2009). In short, a mutated copy of FtsB or FtsL was transformed into its respective depletion strain. Strains were grown overnight at 37 °C on an LB plate supplemented with arabinose and spectinomycin. A single colony from the plate was grown

overnight at 37 °C in 3 ml of LB medium supplemented with arabinose and spectinomycin. The overnight culture was then diluted 1:100 into fresh LB medium containing the same supplement and grown to an A_{600} of ~0.3. An aliquot of 1 ml of culture was washed twice with LB medium lacking any sugar and then diluted 1:100 into 3 ml of fresh LB medium supplemented with glucose and isopropyl-β-d-1-thiogalactoside to induce the expression of the mutated gene and the repression of the wildtype gene. The cells were then grown at 37 or 42 °C for 3.5 h before microscopy, the approximate time necessary to deplete the cells of the wildtype chromosomal copy (Gonzalez and Beckwith, 2009). Depletion strains provided with their respective wildtype copy of the protein in the plasmid were tested as positive controls, and, similarly, depletion strains with no protein in the plasmid (empty vector) were tested as negative controls.

2.5.11 Microscopy and cell length measurement

10 µl of cell samples were mounted on a number 1.5, 24 × 50-mm (0.16–0.19-mm thickness) coverglass slide (Fisher). Cells were cushioned with a 3% (w/v) agarose gel pad to restrict the movement of the live cells. Cells were optically imaged using a Nikon Eclipse Ti inverted microscope equipped with crossed polarizers and a Photometrics CoolSNAP HQ2 CCD camera using a Nikon ×100 oil objective lens. Phase-contrast images of bacterial cells were recorded with a 70-ms exposure time using Nikon NIS Elements software. Multiple snapshots were collected for each experiment. All images were analyzed to measure the cell length in Oufti (Paintdakhi et al., 2016) using one single optimized parameter set.

2.5.12 Whole-cell lysate preparation and Western blotting

Expression level across all variants was assessed by Western blot analysis (Figure S2:9). 3.0 ml of cells were pelleted and resuspended in 500 µl of sonication buffer (25 mm Tris-HCl, 2 mm

EDTA, pH 8.0). The cells were sonicated and centrifuged at 16,000 × *g* for 5 min before collecting the supernatant. Total protein concentration was determined by BCA assay (Pierce). 150 µl of lysates were mixed with 50 µl of 4× LDS sample buffer (Novex, Life Technologies) and boiled at 95 °C for 5 min. For each sample, the equivalent of 10 µg of total protein was separated by SDS-PAGE (Invitrogen) and transferred to polyvinylidene difluoride membrane (VWR). Horseradish peroxidase-tagged anti-FLAG (M2) antibodies (Sigma; 1:1,000) were used for immunoblotting analysis.

2.6 Author contributions

S.G.C., D.-A.M., C.R.A., G.D.-V., and A.S. conceptualization; S.G.C., D.-A.M., and A.S. data curation; S.G.C. and A.S. software; S.G.C., D.-A.M., C.R.A., G.D.-V., and A.S. formal analysis; S.G.C., D.-A. M., C.R.A., G.D.-V., S.J.C., and A.S. investigation; S.G.C., D.-A.M., C.R.A., L.M.L., A.S.K., R.C., J.A.C., N.R., D.B.W., A.A.H., J.L.R., Q.C., and A.S. methodology; S.G.C., D.-A.M., C.R.A., G.D.-V., and A.S. writing—original draft; S.G.C., D.-A.M., C.R.A., G.D.-V., S.J.C., and A.S. writing—review and editing; C.R.A. validation; D.B.W., A.A.H., J.L.R., Q.C., and A.S. supervision; A.S. funding acquisition; A.S. project administration.

2.7 Acknowledgments

We thank Drs. Jon Beckwith and Mark Gonzalez for kindly providing plasmids and strains for the in vivo analysis and for guidance in performing the experiments. Computational resources from the Extreme Science and Engineering Discovery Environment (XSEDE), supported by National Science Foundation Grant OCI-1053575, are greatly appreciated. We thank Dr. Kai Cai for assistance during the biophysical studies and Dr. Elizabeth Caselle for critical reading of the manuscript.

This work was supported in part by National Institutes of Health (NIH) Grant R01-GM099752, National Science Foundation (NSF) Grant CHE-1415910 (to A. S.), and NSF Grant DMS-1661900 (to Q. C.). *The authors declare that they have no conflicts of interest with the contents of this article*. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

2.8 Supplementary Figures and Tables

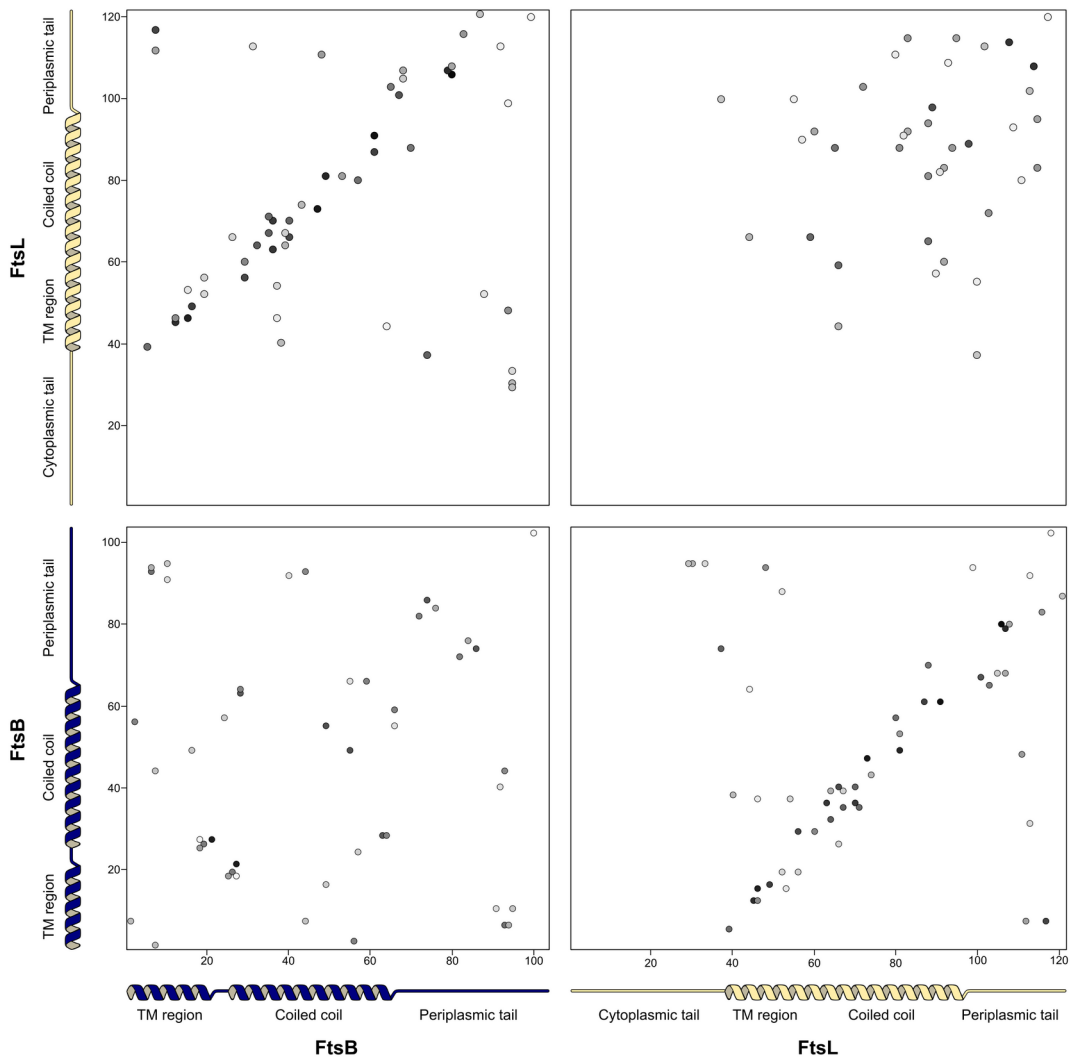

Figure S2:1: Co-evolutionary analysis of the paired alignment of FtsB and FtsL sequences. Map of potential contacts between positions in the sequence in FtsB and FtsL, inferred by co-evolutionary analysis. The top-left and bottom-right quadrants correspond to the FtsB-FtsL potential contacts reported in Figure 2.2. In addition, this figure reports the potential intra-subunit contacts, within FtsB (bottom-right) and within FtsL (top-right).

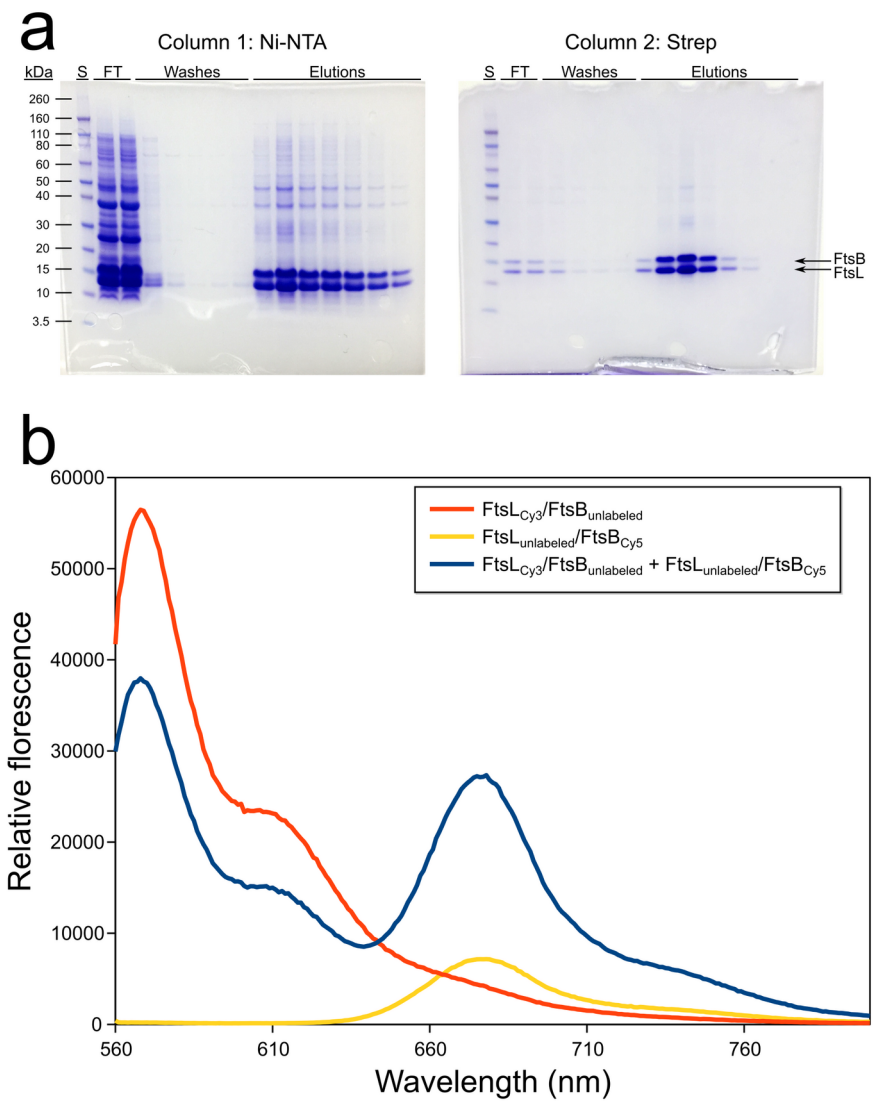

Figure S2:2: Determination of FtsL and FtsB association in vitro. a) Co-expressed strep-tagged FtsL and His-tagged FtsB proteins co-elute when run over consecutive Ni-NTA and streptavidin columns, demonstrating that FtsLB a stable complex in detergent. The density of the bands is consistent with a 1:1 stoichiometric ratio. S: standard ladder; FT: flow-through fractions. b) Acceptor-labeled FtsB ($FtsL_{unlabeled}$ /$FtsB_{Cy5}$) and donor-labeled FtsL ($FtsL_{Cy3}$ /$FtsB_{unlabeled}$) were mixed in a 50:50 ratio and equilibrated and emission spectrum

was recorded (blue line, excitation wavelength 550 nm). The mixed sample shows a substantial increase of acceptor emission around 680 nm, indicating proximity between $FtsB_{Cy5}$ and $FtsL_{Cy3}$. The emission scan is compared to those of an acceptor-only sample (red, 50:50 mix of $FtsL_{unlabeled}$ /$FtsB_{Cy5}$ and $FtsL_{unlabeled}$ /$FtsB_{unlabeled}$) and of a donor-only sample (yellow, 50:50 mix of $FtsL_{Cy3}$ /$FtsB_{unlabeled}$ and $FtsL_{unlabeled}$ /$FtsB_{unlabeled}$). For these two control samples, mixing with $FtsL_{unlabeled}$ /$FtsB_{unlabeled}$ was performed to maintain identical total protein concentration and protein:detergent ratios. All other conditions are identical to the stoichiometric analysis of Figure 2.3.

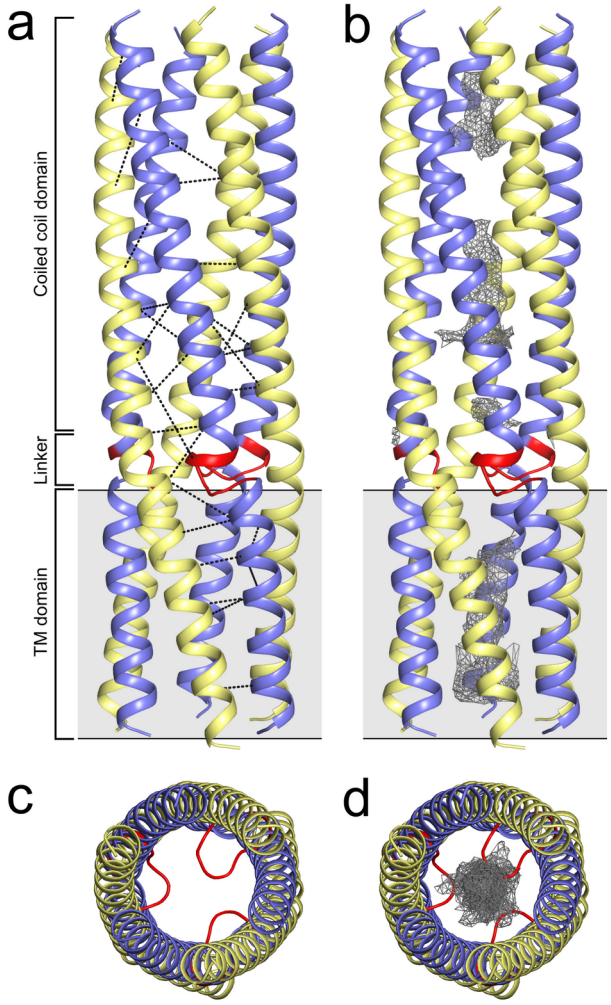

Figure S2:3: Hexameric model of the FtsLB complex. a) Model of the L3B3 hexamer with the co-evolutionary restraints (Table S2.1) shown as black dashes. Most of the evolutionary constraints are fit by this model. FtsB rendered in blue, FtsL rendered in yellow, and the juxtamembrane linker in FtsB highlighted in red. b) Solvent-accessible cavities rendered in gray wireframe. These cavities do not form a complete channel and are instead occluded by bulky side chains and the juxtamembrane linker of FtsB. c) and d) Top-down views of the hexameric model presented in a and b.

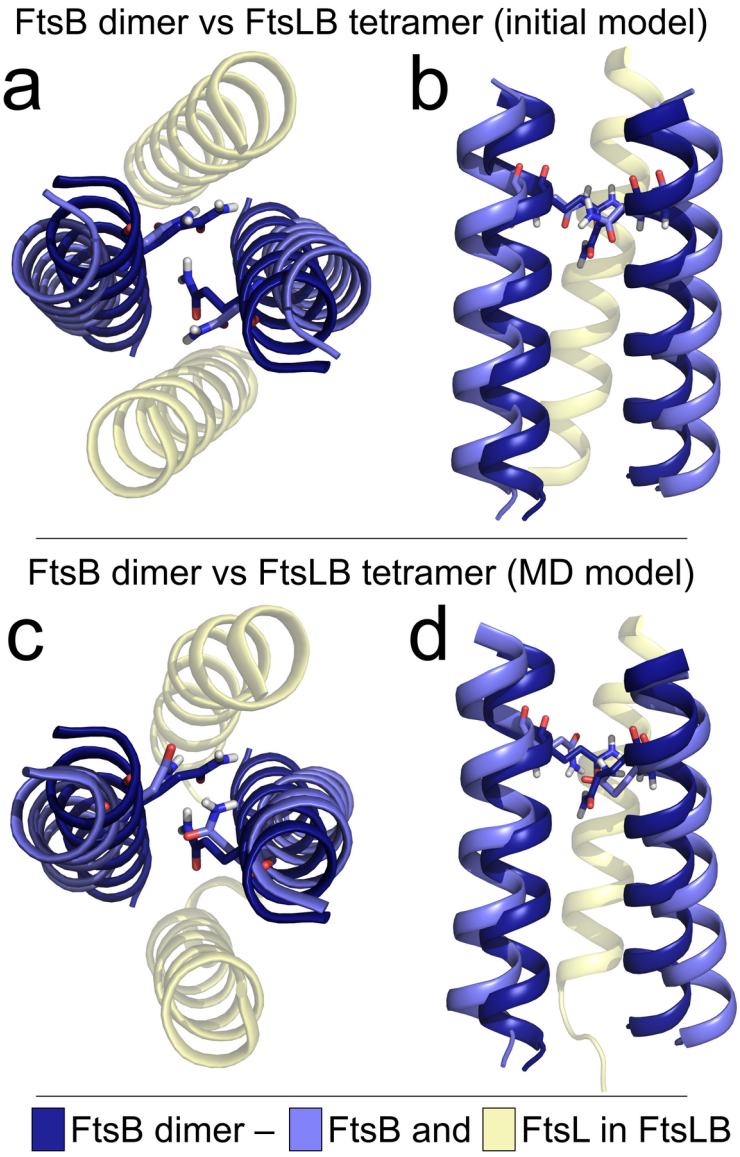

Figure S2:4: The configuration of the TM region of FtsB is similar to a previous model obtained in the absence of FtsL. Structural alignment of the L_2B_2 complex to a model of the FtsB homodimer. a) Top-down and b) side views of an alignment with the L_2B_2 model (blue, yellow) with the model of the homodimer (dark blue), which was previously generated using mutagenesis data based on a genetic reporter assay of self-association

(LaPointe et al., 2013). Residue Q16 is shown in sticks. The interface of FtsB is the same in both models, but the helices are farther apart in the L_2B_2 tetramer. c) and d) Top-down and side views of the FtsB homodimer model aligned with the last frame of the 260ns MD simulation. The FtsB helices have moved closer together, improving the RMSD of the alignment from 2.5 Å to ~2 Å.

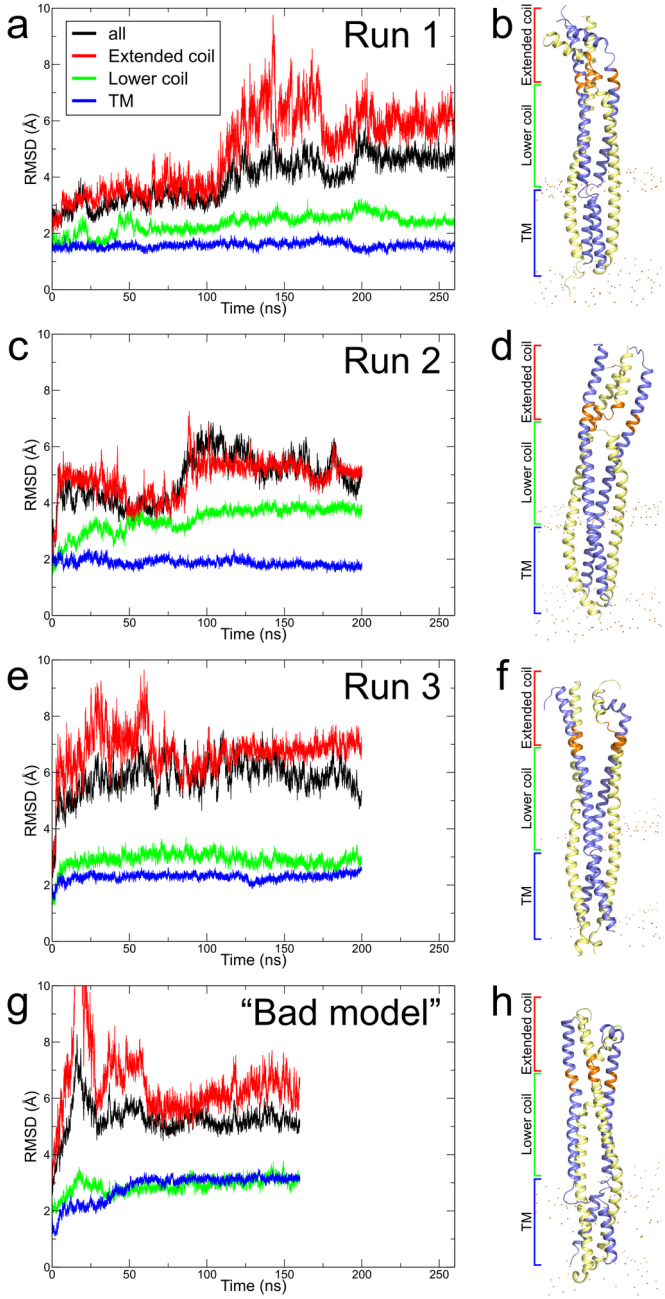

Figure S2:5: RMSD analysis of the three replica MD runs and of a "bad model". The figure is similar to Figure 2.5a (Run 1). The RMSD analysis is shown for the entire complex (black), the TM region (blue), the lower coil (green, 29-61 of FtsB and 57-91 of

FtsL) and the extended coil (red, 62-79 of FtsB and 92-110 of FtsL). The CCD, which separates the lower from the extended coil is in orange. a) Trajectory of the 260 ns Run 1. b) Conformation of the last step of the trajectory. c and d) Trajectory of the 200 ns Run 2. e and f) Trajectory of the 200 ns Run 3. g and h) Trajectory of 160 ns run of a “bad model”, an alternate low energy model in which the conformation of the TM region does not satisfy most of the evolutionary constraints. It is notable how the RMSD of the TM region quickly rises to a higher value (approximately 3 Å, compared to the same variable in Runs 1-3, which is around 2 Å).

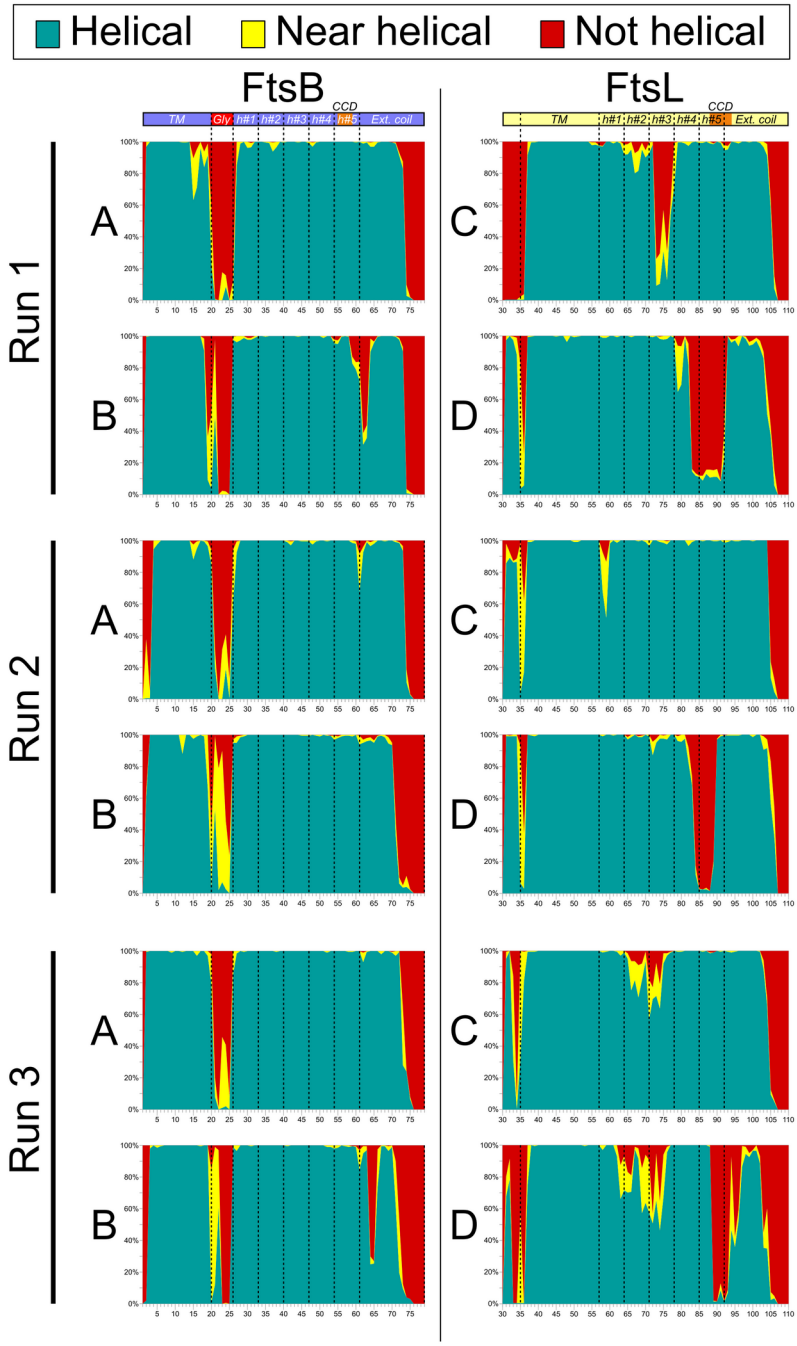

Figure S2:6: Analysis of helicity in the three replica MD runs. The graphs display the fraction of time each position adopted a helical conformation (green), nearly helical

(yellow) or non-helical (red). A and B are the two individual FtsB chains (left panels) and C and D are the two individual FtsL chains (right panels).

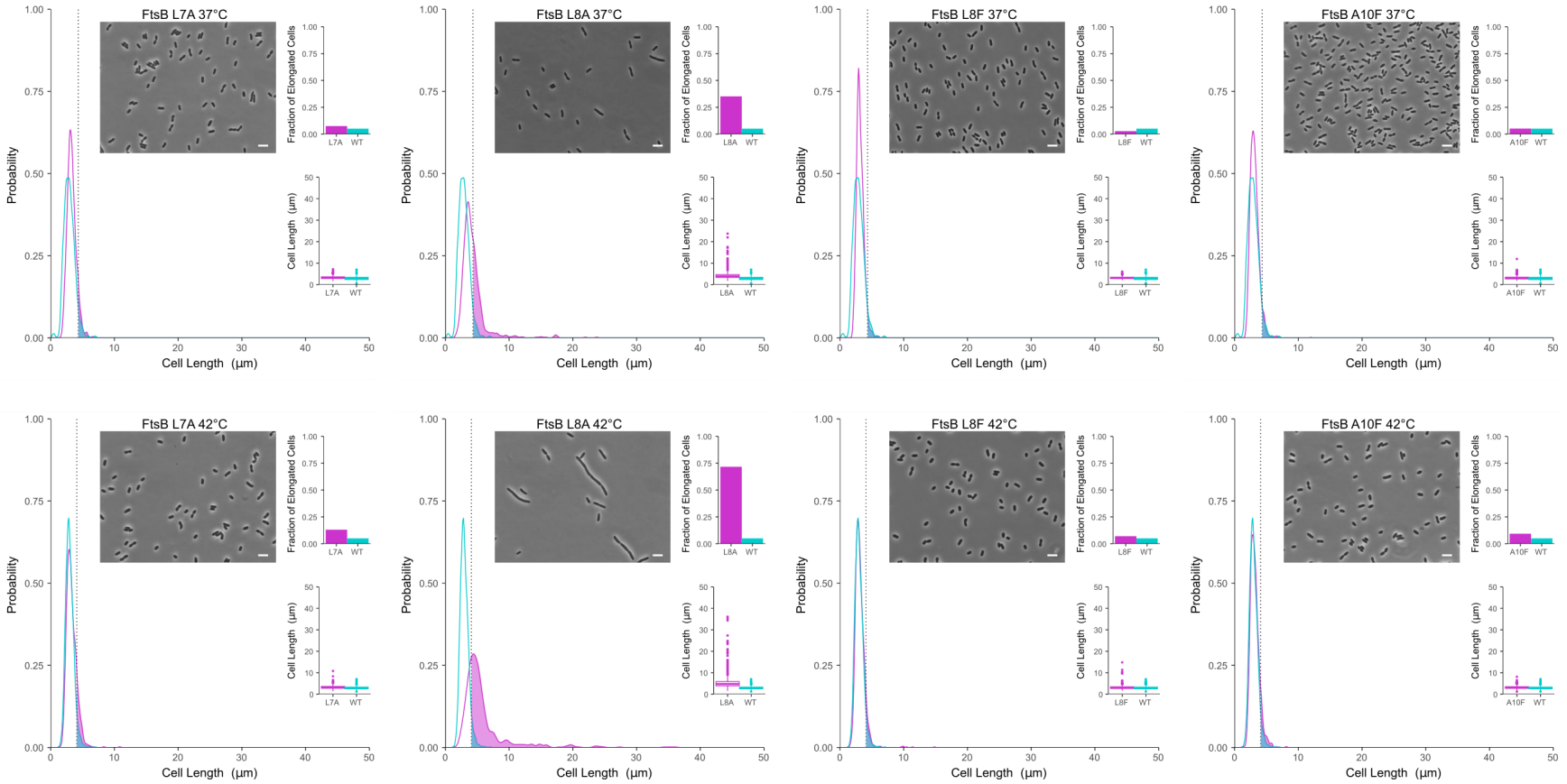

Figure S2:7: Full analysis of cell division phenotypes of point and alanine insertions mutants reported in Figure 2.6 and Fig 7. The distribution of cell lengths for each point substitution or insertion mutant was compared between wild type (aqua) and mutant

(magenta) cell population. Center: kernel density functions of mutant and wild-type cell lengths. The area above the 95th percentile of the wild-type distribution (dotted line) is shaded. Insert: representative photo of mutation. Scale bar corresponds to 5μm. Right, top: fraction of cells above the 95th percentile of wild-type cell length. Right, bottom: box and whisker plots of cell lengths for wild-type and mutant. Whiskers represent the interquartile range multiplied by 1.5.

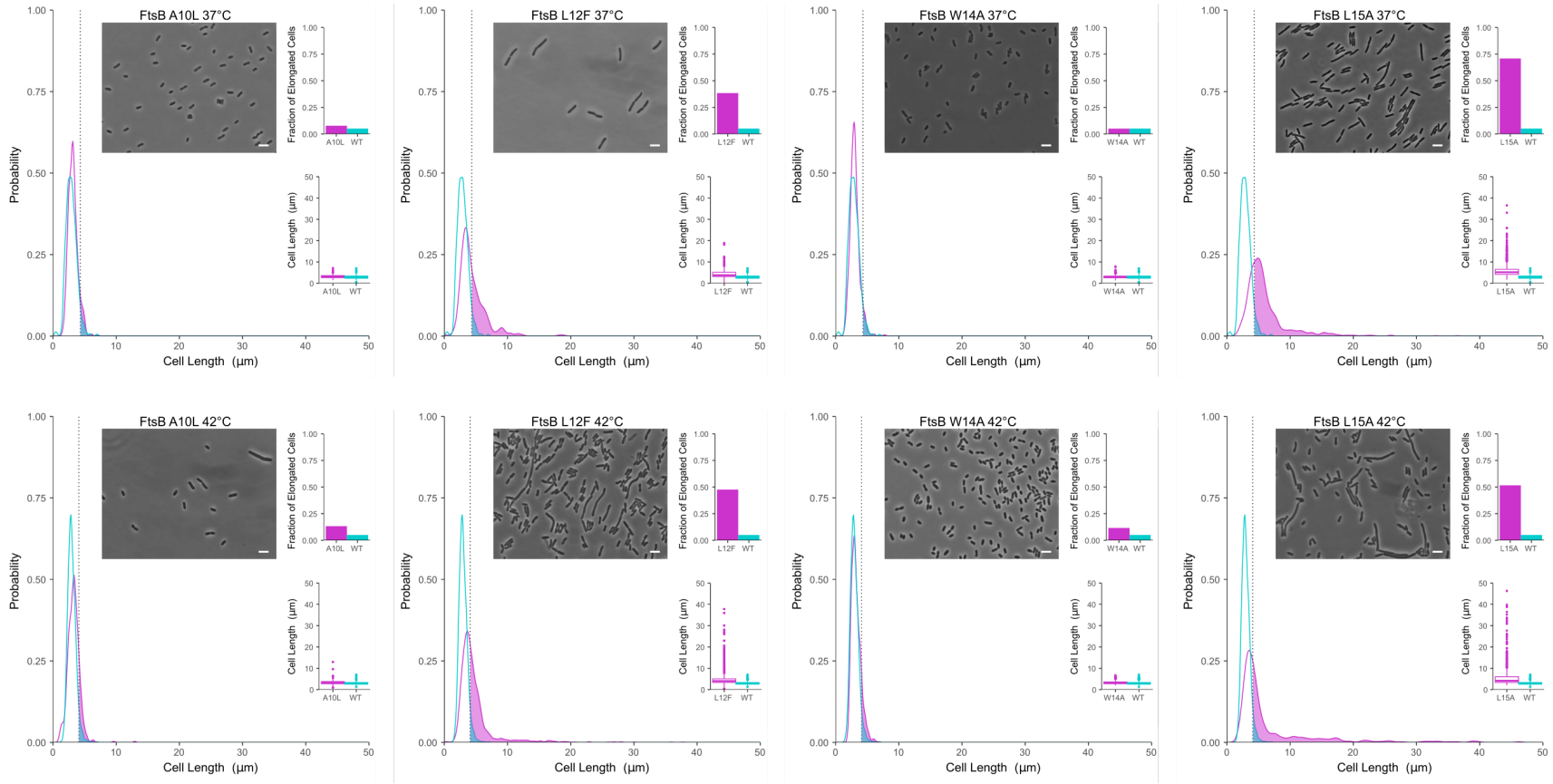

Figure S2:7, continued

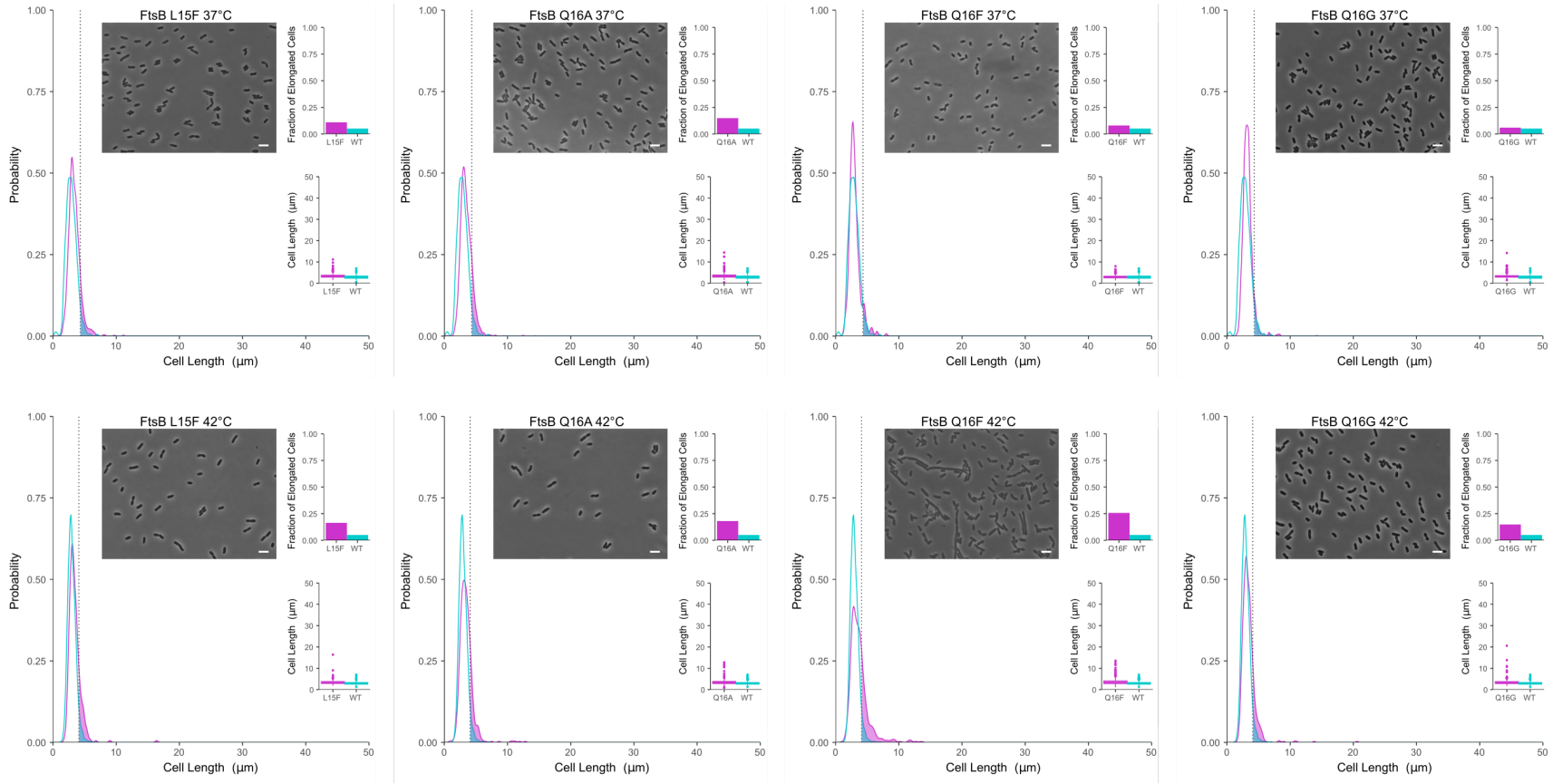

Figure S2:7, continued

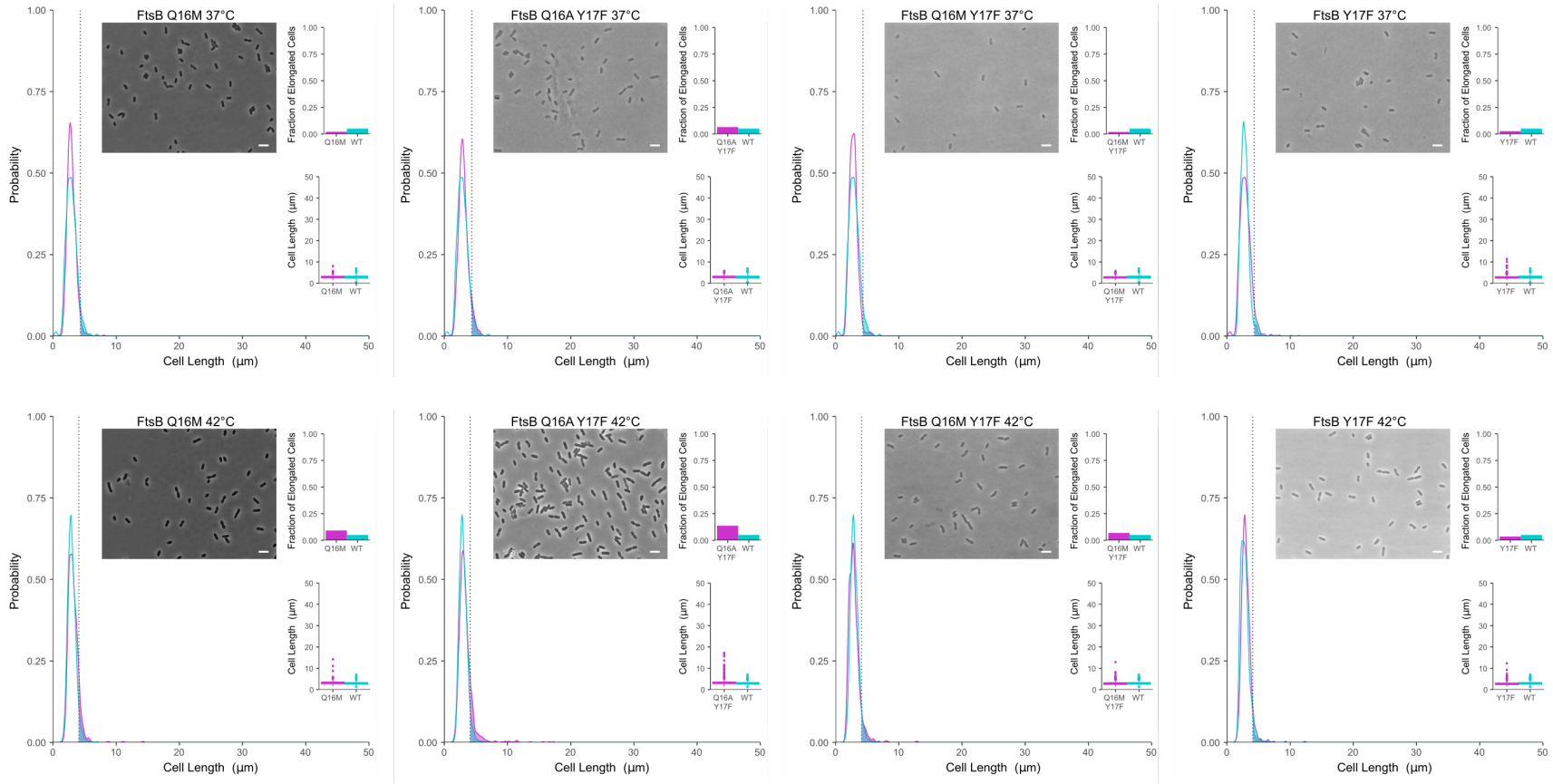

Figure S2:7, continued

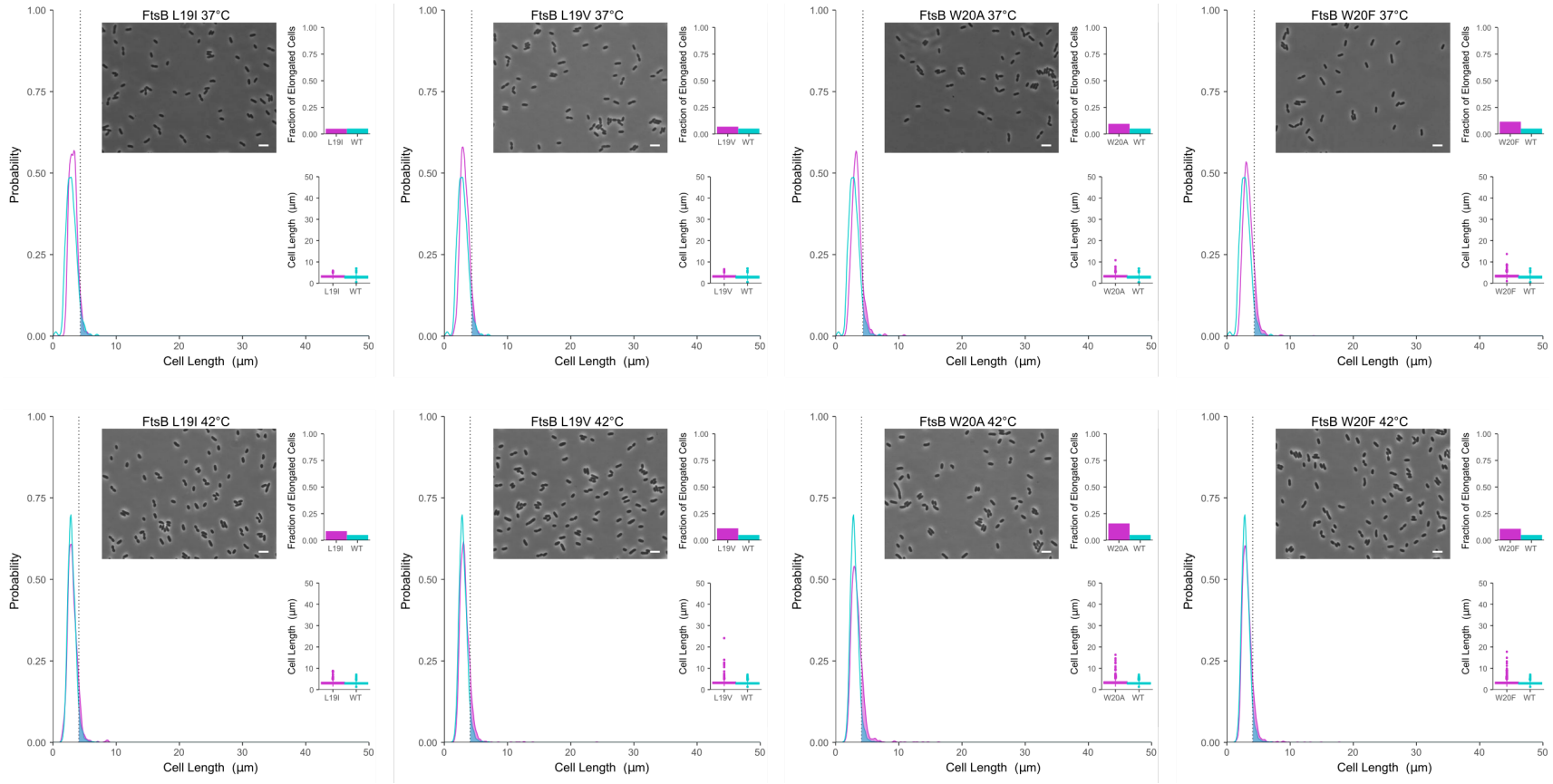

Figure S2:7, continued

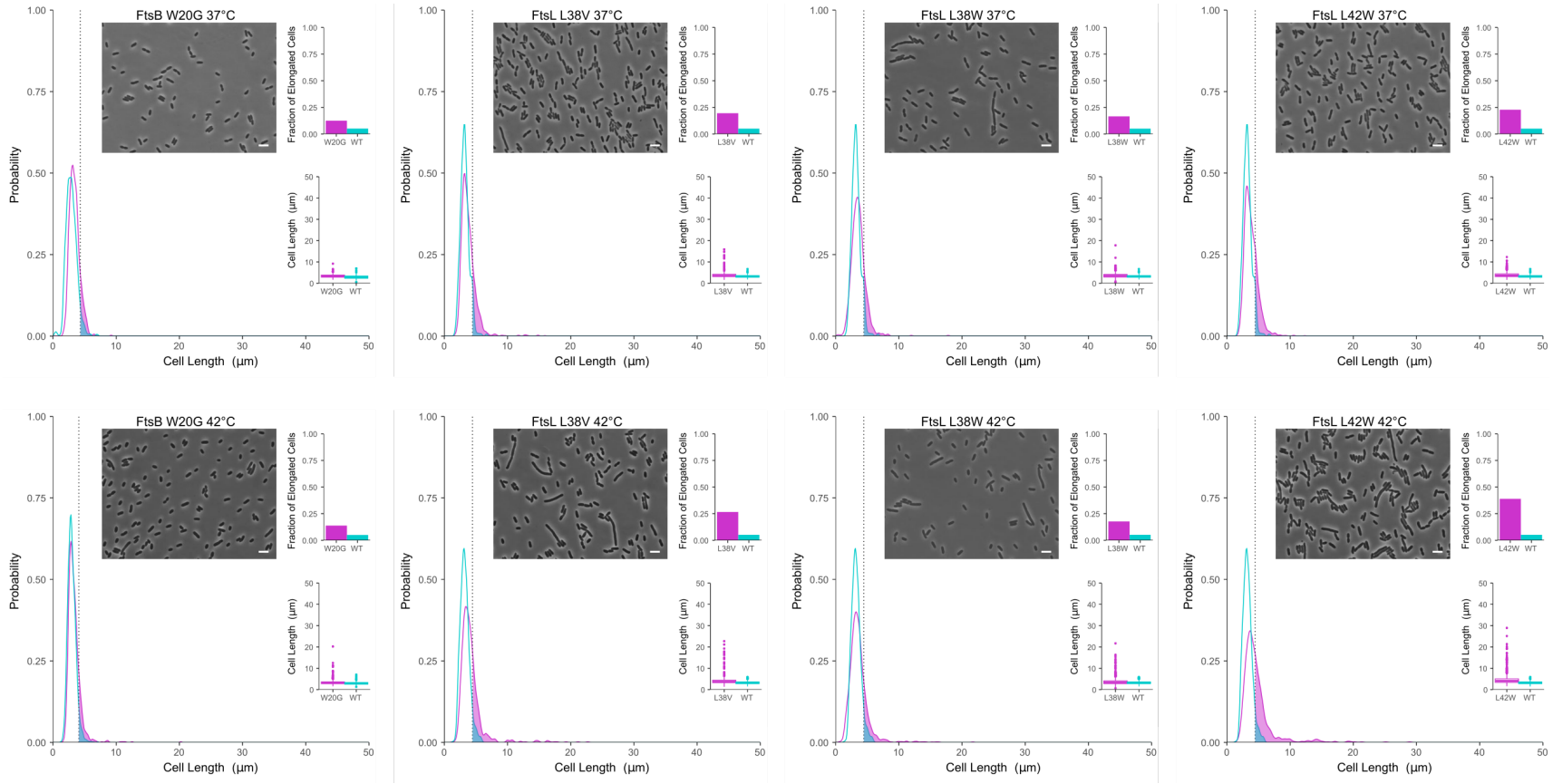

Figure S2:7, continued

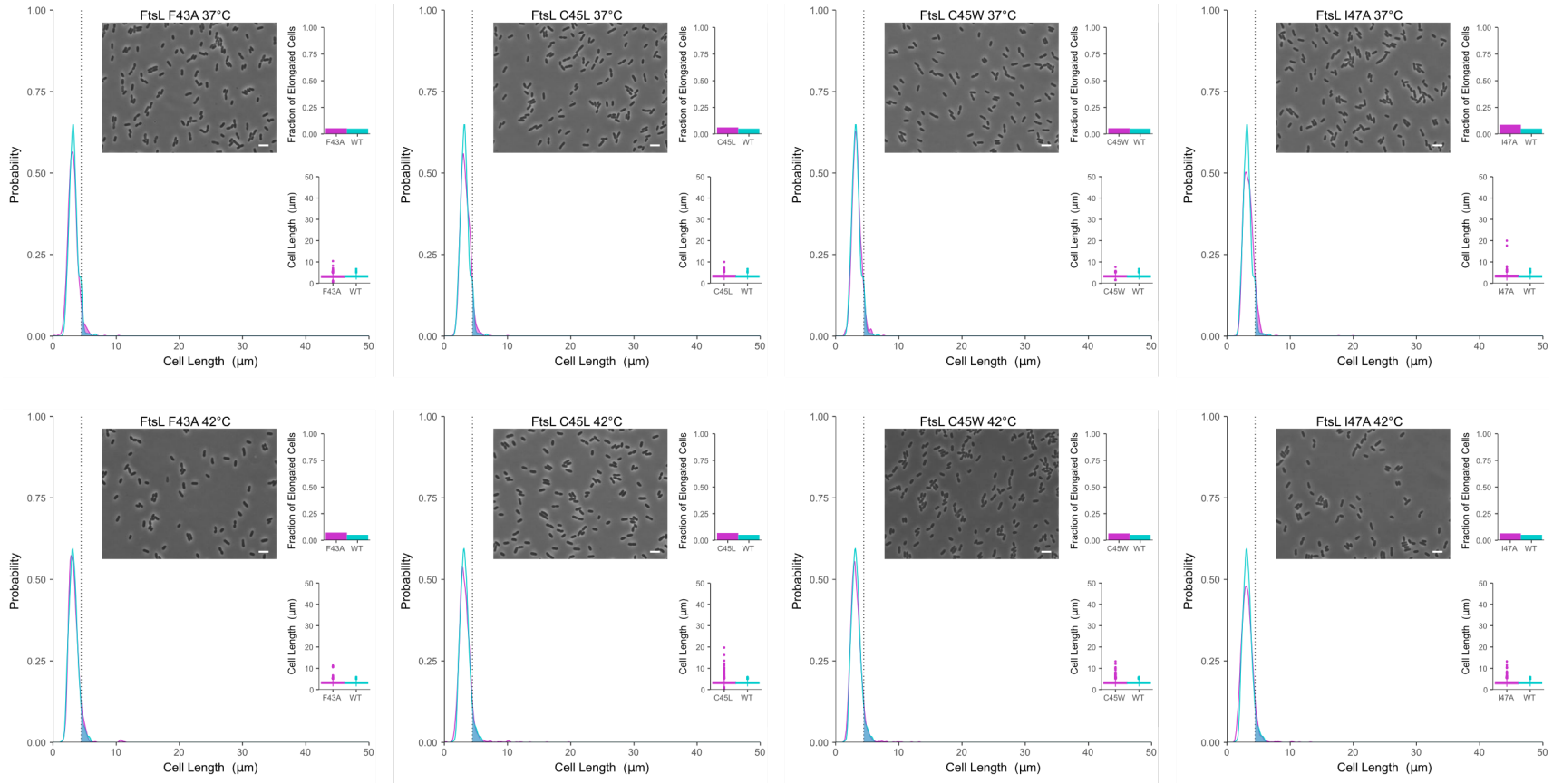

Figure S2:7, continued

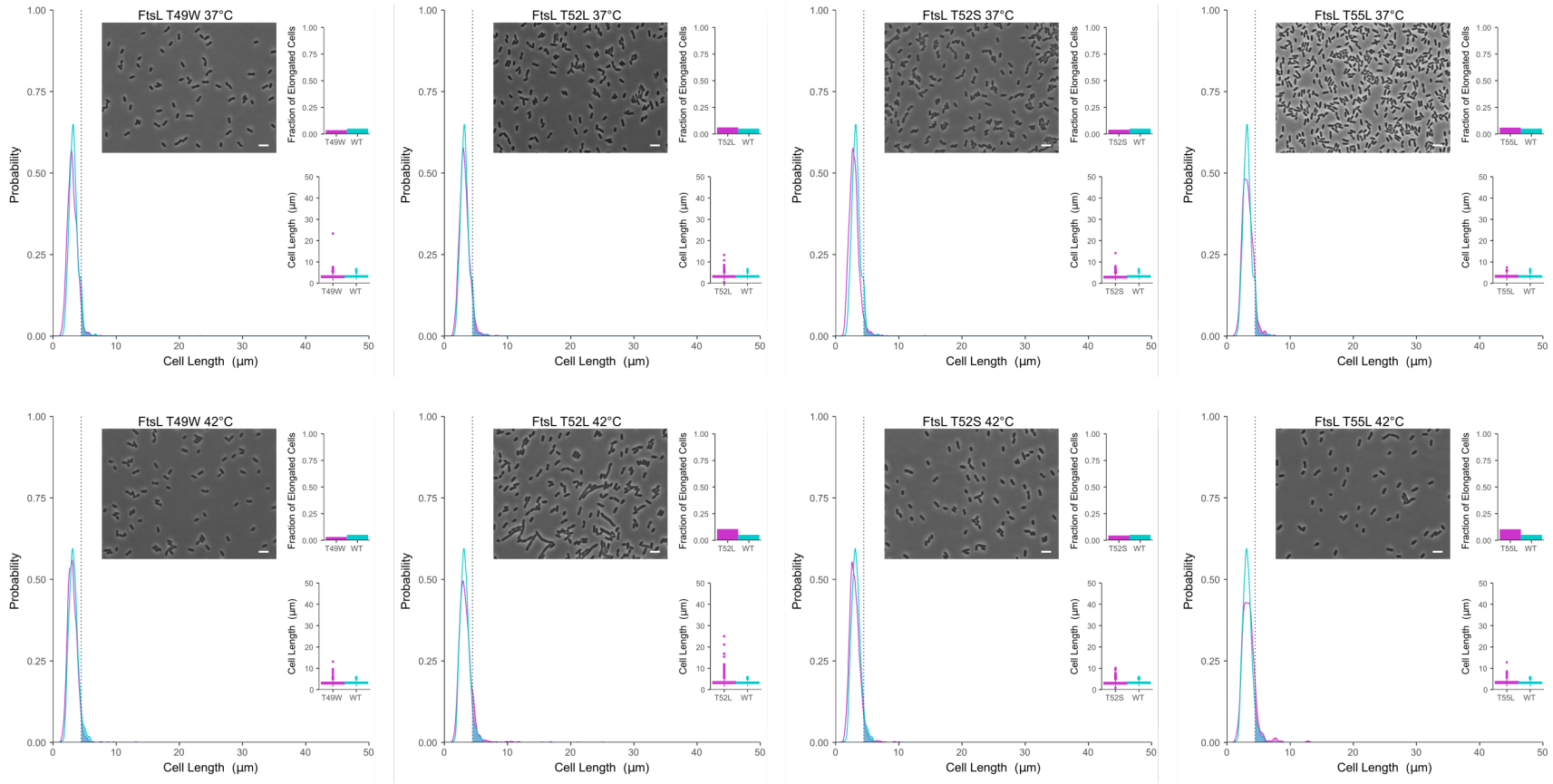

Figure S2:7, continued

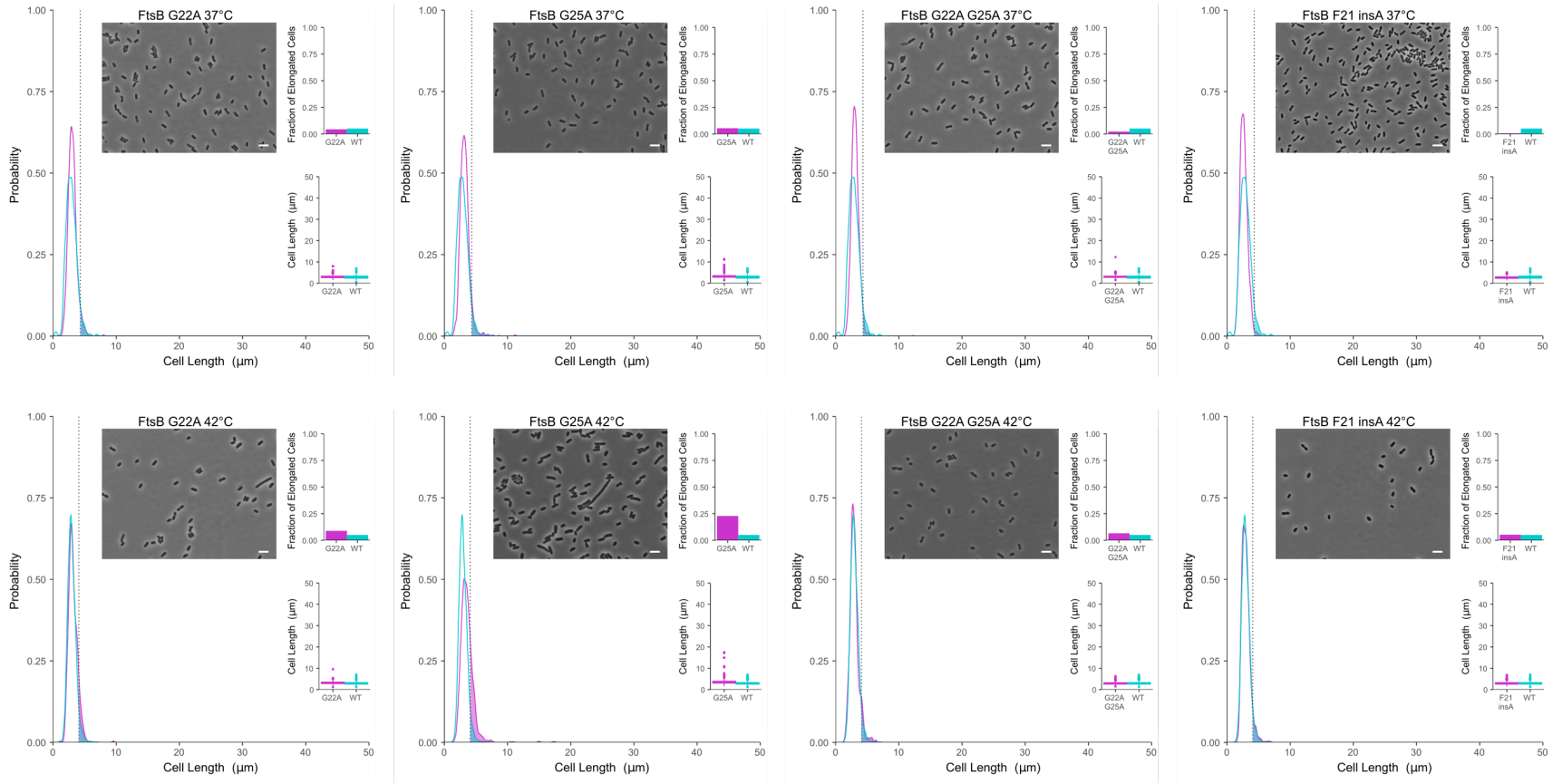

Figure S2:7, continued

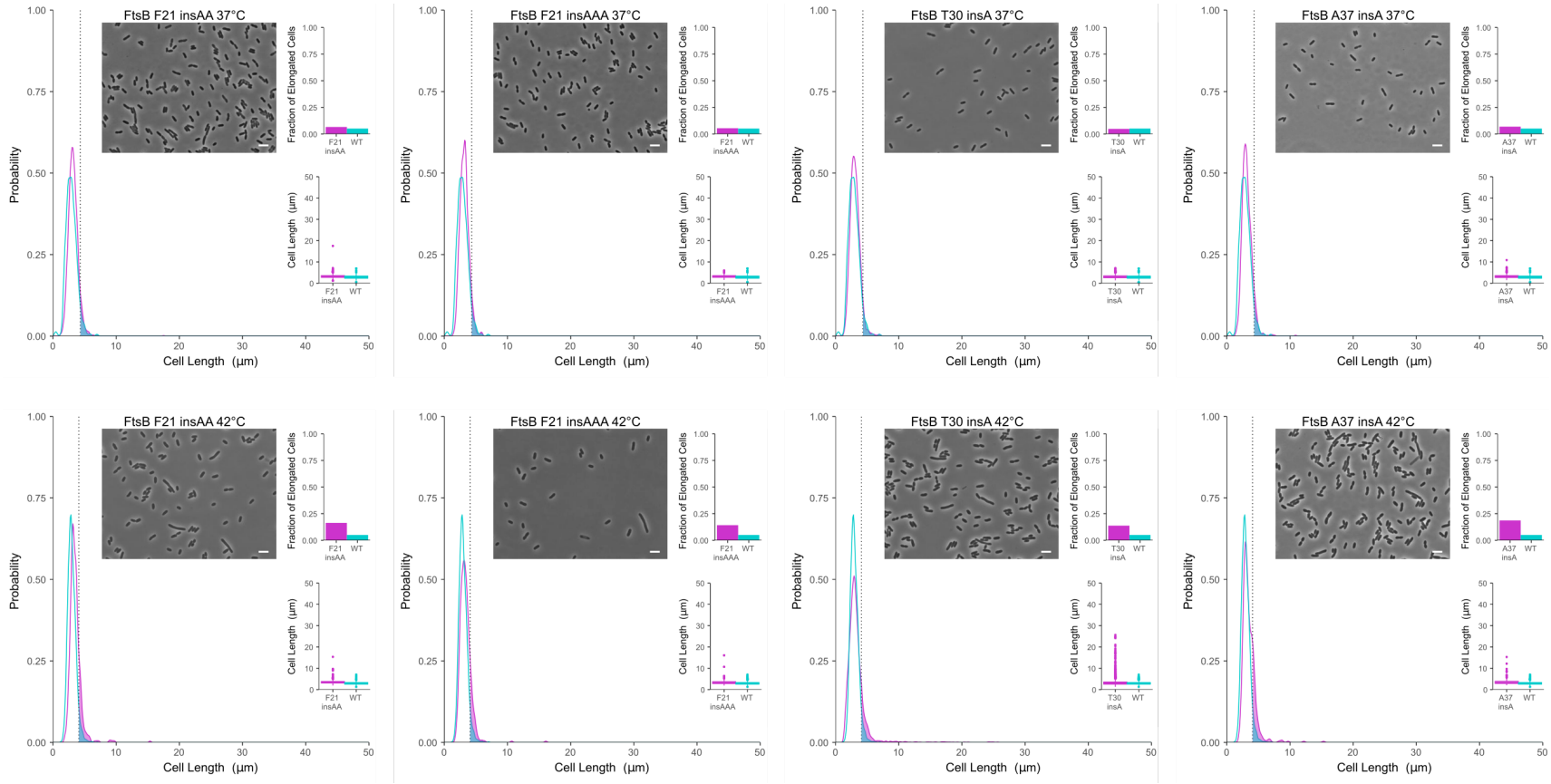

Figure S2:7, continued

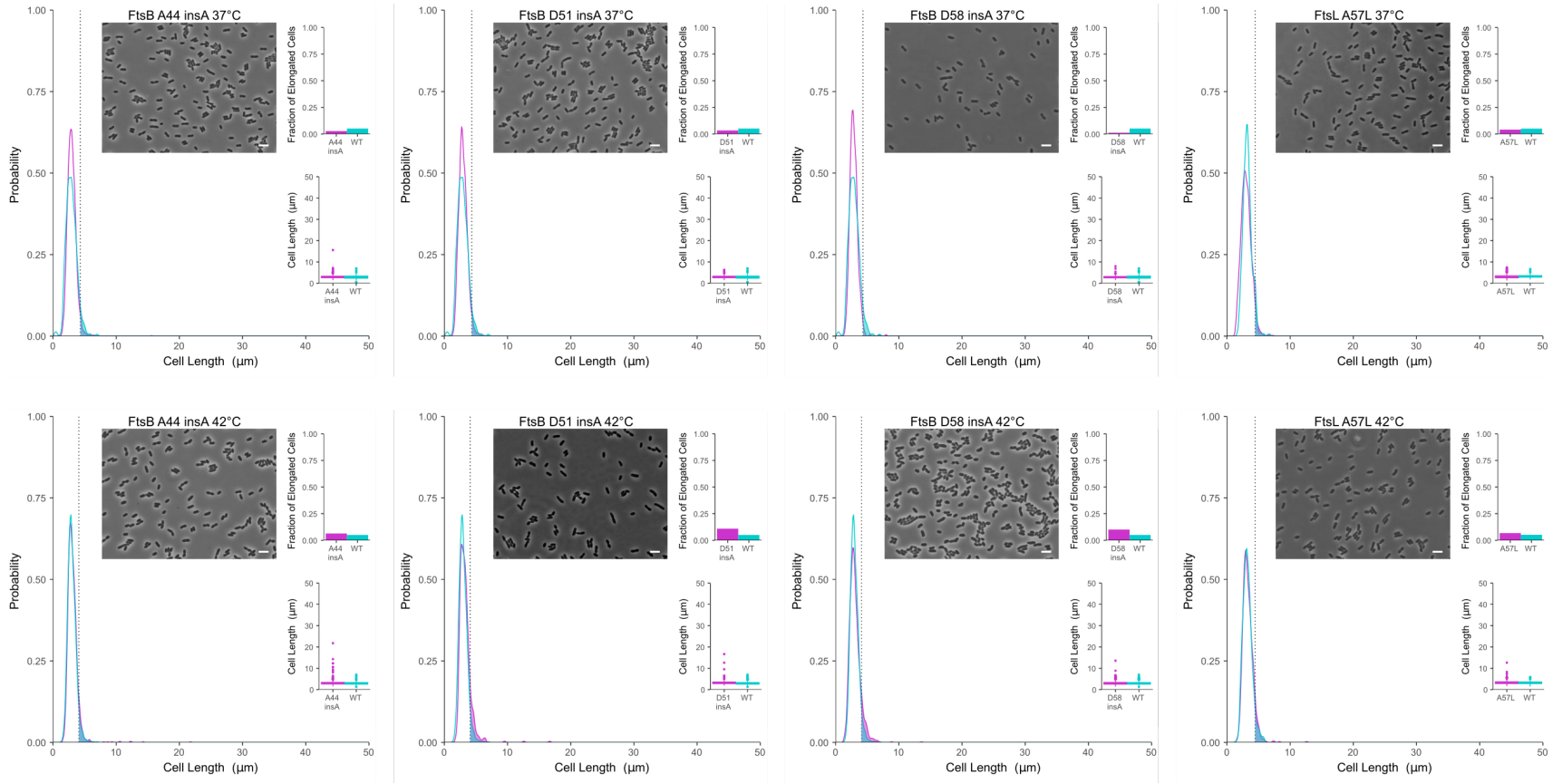

Figure S2:7, continued

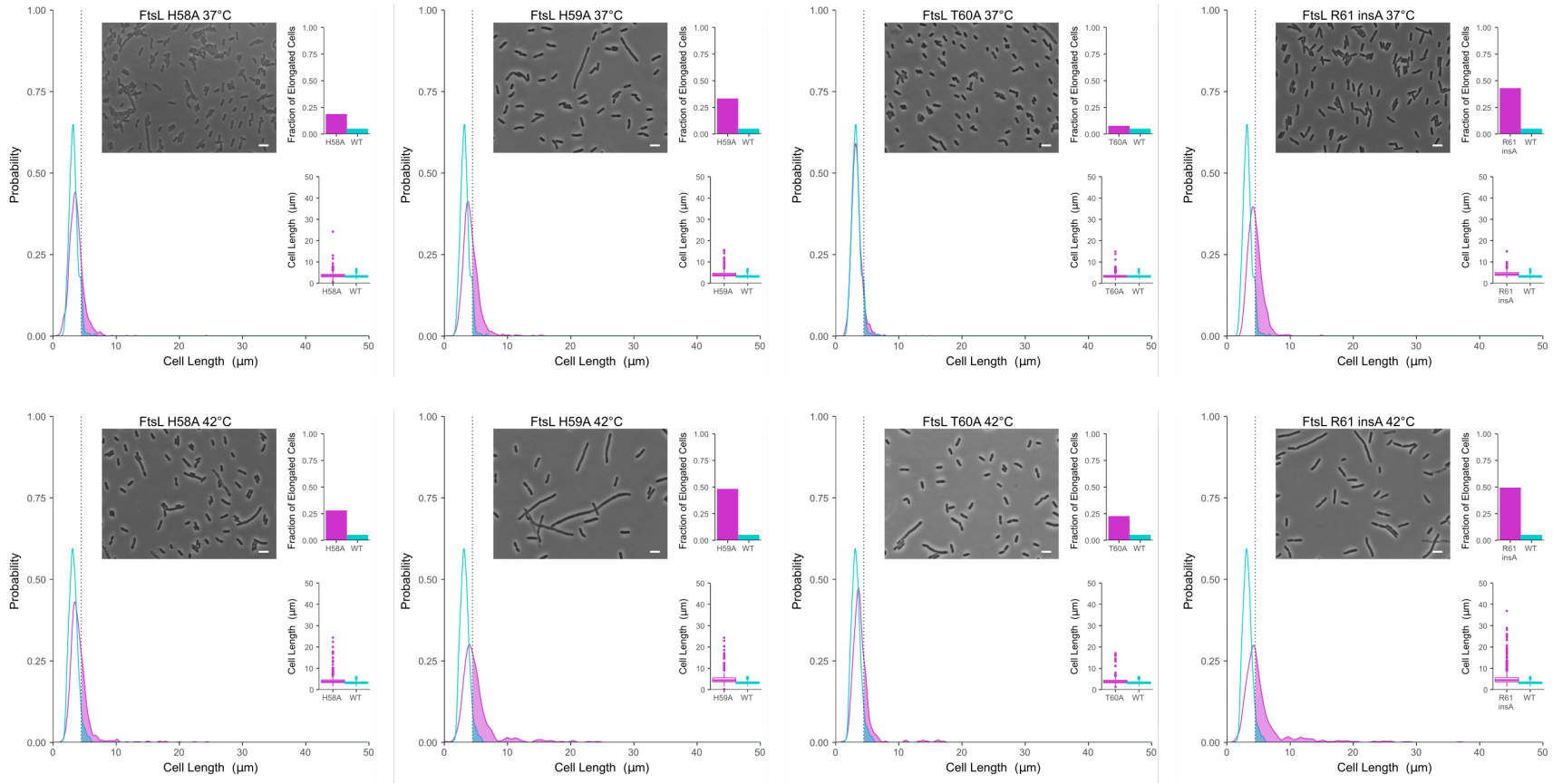

Figure S2:7, continued

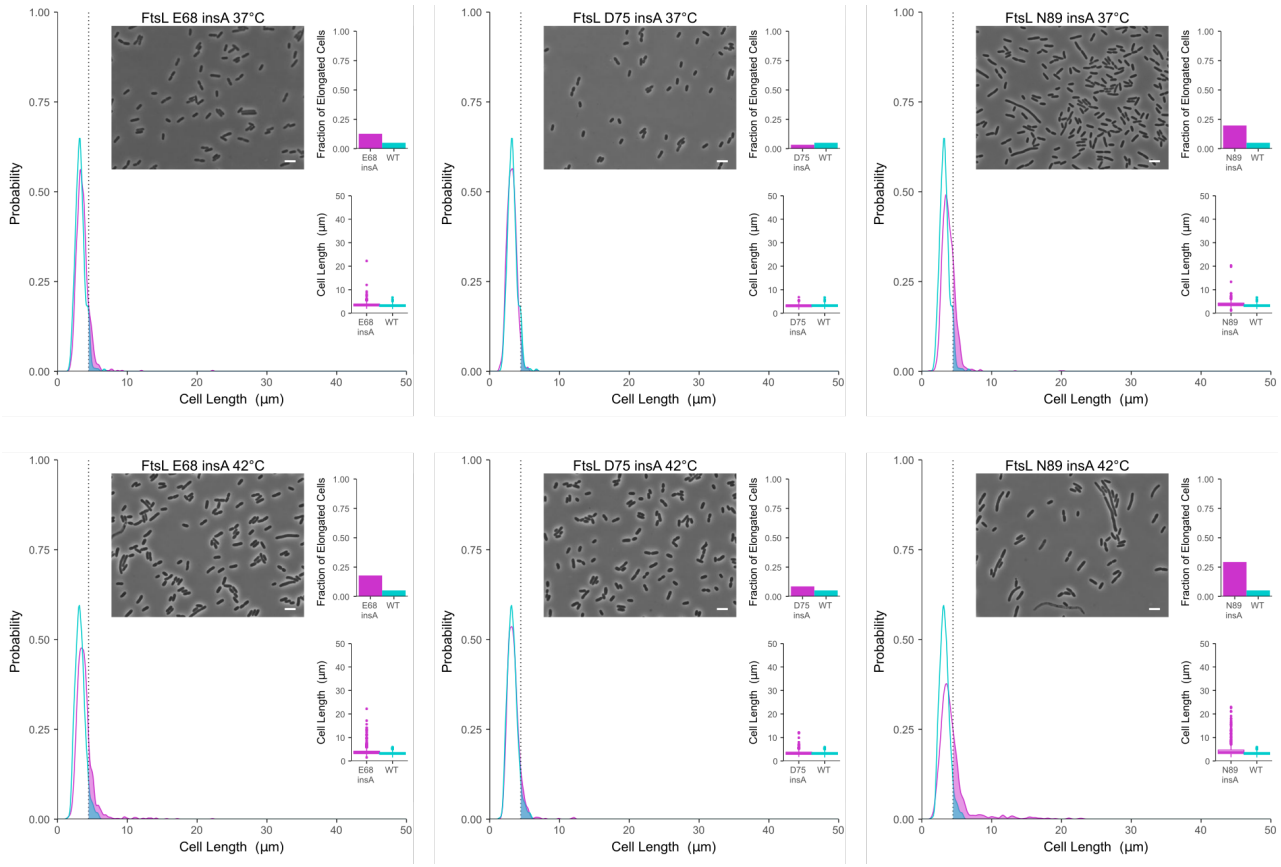

Figure S2:7, continued

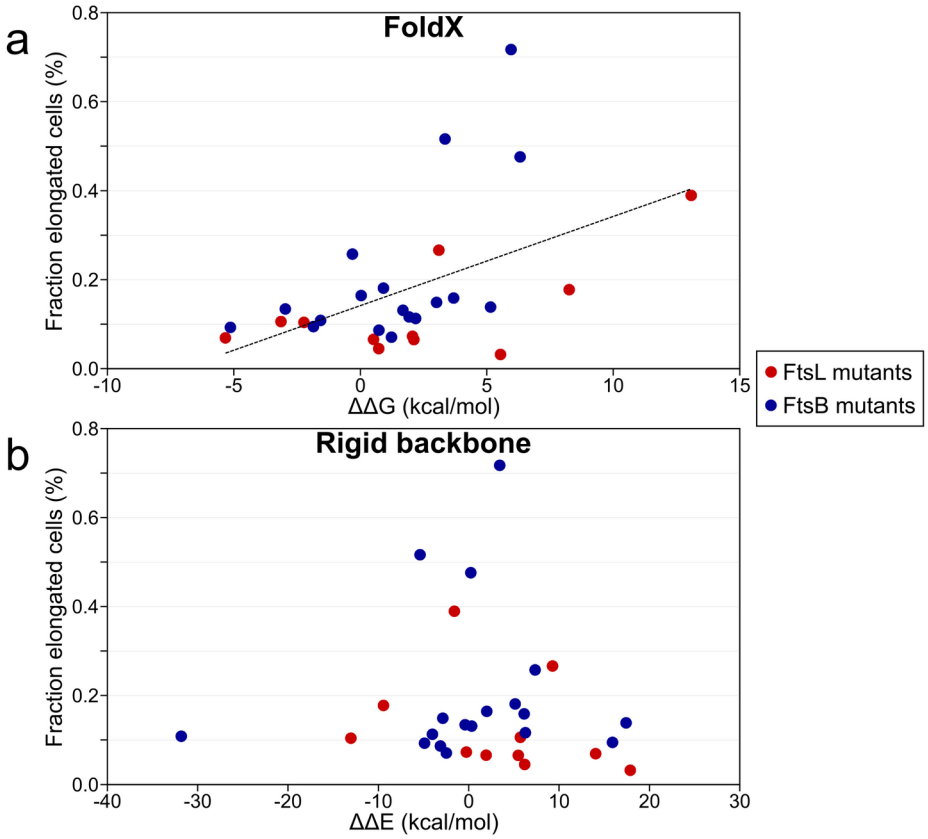

Figure S2:8: Comparison of the predicted mutational energy and the biological phenotypes of the mutants in the transmembrane region. Biological phenotype (expressed as the fraction of elongated cells, as defined in Figure 2.6) compared with the energy of destabilization of each mutant calculated in silico using a) FoldX or b) rigid body repacking in which the mutation was modeled without relaxing the backbone. The energies predicted by FoldX correlate statistically with the severity of the phenotypes by Rank Order Spearman Correlation Coefficient analysis, which does not assume a linear model ($r = 0.4631$, $N = 29$, $p < 0.01$), and by linear regression ($R^2 = 0.25$, $p < 0.01$, dashed line). Correlation with rigid backbone energies is not statistically significant.

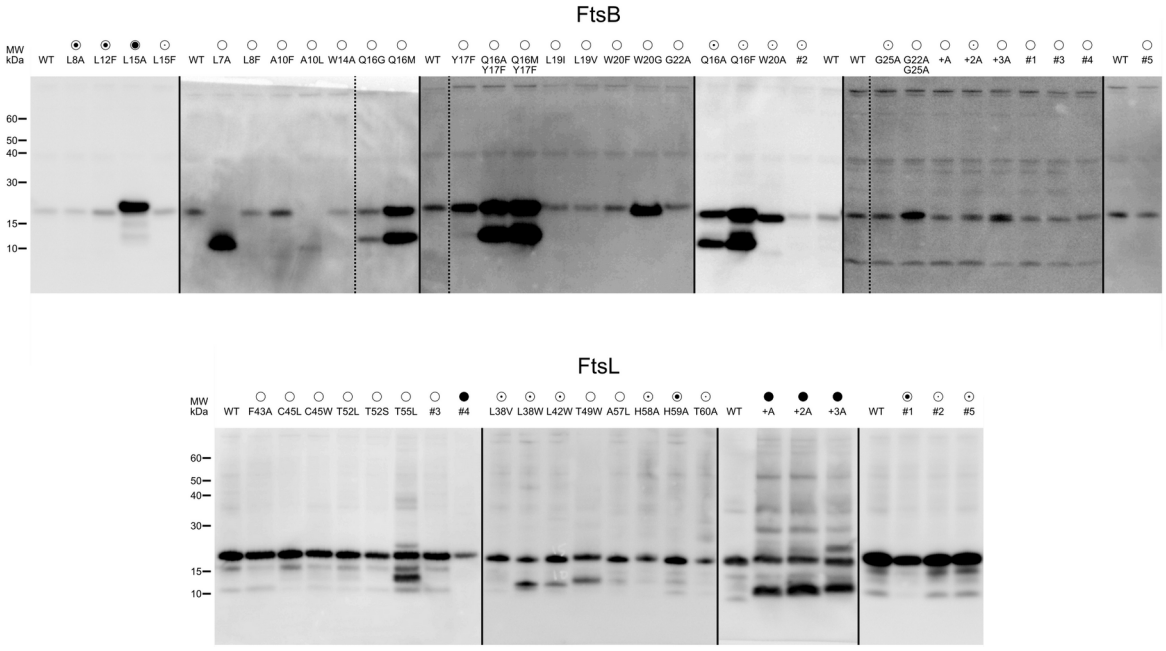

Figure S2:9: Expression level of FtsB and FtsL mutants assessed by Western Blot analysis. Image of all western blots of the FtsB and FtsL mutants tested in this work. The circles indicate the observed phenotypes, as defined in Figure 2.6. The FtsB wild type and its mutants display a notably fainter band with respect to the FtsL variants. Protein expression level of the FtsB and FtsL mutants with defective phenotypes are generally comparable and to the respective wild type. There are cases of FtsB mutants with increased protein level. These correspond to mutations in a specific region of the TM domain of FtsB, involving L15, Q16, Y17 or W20. These mutants with increased protein level correspond to interfacial positions that have a wild-type-like (Q16M, Y17F, Q16M/Y17F, Q16A/Y17F, W20A), mild temperature- ensitive (Q16F), mild (Q16A) phenotypes, or, in one case, a severe phenotype (L15A). It is possible that these apparent increases in protein levels are due to reduced protein degradation, possibly because the mutations mask a site

recognized by a membrane protease. Additionally, two FtsB mutants (L7A and A10L) present faint full-length bands and significant degradation. However, these mutants are viable and are classified as wild-type-like. Individual gels are separated by solid lines. Spliced lanes within the same gel are indicated by a dashed line.

FtsB		FtsL		Co-evo ranking		Structural model	
Position	**Heptad[1]**	**Position**	**Heptad[1]**	**Rank[2]**	**Score[3]**	**Min distance (Å)[4]**	**Cα distance (Å)[5]**
			Other				
65		103		33	0.18404		
67		101		19	0.21003		
68		105		72	0.16013		
68		107		55	0.16661		
79		107		7	0.24894		
80		106		1	0.41824		
80		108		53	0.16726		
83		116		45	0.17608		
87		121		67	0.16224		
7		112		51	0.16889		
7		117		14	0.21928		
31		113		80	0.15693		
37		46		89	0.15313		
37		54		77	0.1583		
38		40		63	0.16335		
48		111		48	0.17405		
64		44		94	0.15196		
70		88		32	0.18635		
74		37		24	0.19954		
88		52		84	0.15532		
92		113		91	0.15285		
94		48		42	0.17769		
94		99		93	0.1522		
95		29		66	0.16249		
95		30		60	0.16504		
95		33		79	0.15708		

Table S2.1: Top 95 predicted co-evolutionary contacts between FtsL and FtsB

[1]Residue assignment to a canonical coiled-coil heptad repeat, assuming a continuous transmembrane-coiled-coil helix.

[2]Pair ranking originated from the EV-Couplings algorithm.

[3]Pseudo-likelihood maximization (PLM) score computed for the pair of positions from the EV-Couplings algorithm.

FtsB		FtsL		Co-evo ranking		Structural model	
Position	**Heptad[1]**	**Position**	**Heptad[1]**	**Rank[2]**	**Score[3]**	**Min distance (Å)[4]**	**Cα distance (Å)[5]**
			TM Domain				
T 5	e	P 39	a	28	0.18971	4.76	6.43
L 12	e	C 45	g	8	0.23962	3.61	7.56
L 12	e	I 46	a	41	0.17785	3.71	5.86
L 15	a	I 46	a	5	0.27006	4.01	8.11
L 15	a	V 53	a	86	0.15394	4.61	8.4
Q 16	b	T 49	d	12	0.23363	3.08	6.86
L 19	e	T 52	g	78	0.15754	3.58	8.4
L 19	e	T 56	d	75	0.15967	3.49	7.66
			Coiled coil				
I 26	e	Q 66	g	76	0.15946	11.22**	16.52
Y 29	a	T 56	d	15	0.21667	7.97*	10.12
Y 29	a	T 60	a	39	0.17999	5.91*	8.69
V 32	d	T 64	e	22	0.20568	4.54	7.28
D 35	g	R 67	a	23	0.20543	4.03	8.08
D 35	g	V 71	e	27	0.19052	4.46	7.78
V 36	a	L 63	d	10	0.23707	5.38*	9.22
V 36	a	L 70	d	11	0.23583	3.78	7.7
Q 39	d	T 64	e	59	0.16512	9.17*	14.52
Q 39	d	R 67	a	74	0.15968	3.37	10.34
Q 40	e	Q 66	g	16	0.21574	3.54	9.58
Q 40	e	L 70	d	25	0.19574	3.48	6.94
N 43	a	R 74	a	62	0.16375	3.33	7.96
K 47	e	E 73	g	3	0.3737	3.88	9.46
R 49	g	W 81	a	6	0.25888	2.78	8.18
L 53	d	W 81	a	49	0.1721	3.52	10.29
I 57	a	E 80	g	29	0.1897	8.62*	13.43
N 61	e	E 87	g	13	0.22554	3.78	9.25
N 61	e	L 91	d	2	0.41222	4.06	6.67

Table S2.1, continued

[4]Minimum distance between heavy atoms of the residues in the computational model of the L_2B_2 tetramer. Asterisks denote outliers not well satisfied by the computational model. Single asterisk: distances above 5 Å; double asterisks: distances above 10 Å.

[5]Minimum distance between alpha carbons of the residues, which were used to compute sigmoidal distance restraints during the modeling procedure.

Chain	Position	Chain	Position[2]	Occupancy
A	Y17	C	T52	92.00%
A	W20	C	T56	81.00%
A	Q16	C	T49	81.00%
A	Q16	B	Q16	62.00%
B	Q16	B	W20	39.00%
B	Y17	D	L48*b*	31.00%
A	Q16	C	T52	26.00%
B	Q16	C	T49	18.00%
A	Y17	C	L48*b*	13.00%
A	Q16	A	Y17	11.00%
B	Y17	D	T52	10.00%
B	W20	A	Q16*b*	8.00%
B	Q16	D	T49	2.00%
A	Q16	A	V13*b*	1.00%

Table S2.2: Network of hydrogen bonds observed in TM domain of FtsLB during molecular dynamics simulation, run 1[1]

[1]The most frequently observed side-chain to side-chain or side-chain to backbone hydrogen bonding interactions in the molecular dynamics simulations are listed. Chains A and B correspond to FtsB while chains C and D correspond to FtsL. The percentage is cumulative of all possible combinations of donor-acceptor interactions between the two residues. Co-occurring interactions between different donors and acceptors within the same pair of residues are not double-counted.

[2]The letter “b” indicates that the hydrogen bonding involves a backbone group

Name	Sequence[1]
His_6-FtsB	*MGSHHHHHHHHDYDIPTSENLYFQGGSG*MGKLTLLLLAILVWLQYSLWF GKNGIHDYTRVNDDVAAQQATNAKLKARNDQLFAEIDDLNGGQEALEER ARNELSMTRPGETFYRLVPDASKRAQSAGQNNR
His_6-FtsB-S97C	*MGSHHHHHHHHDYDIPTSENLYFQGGSG*MGKLTLLLLAILVWLQYSLWF GKNGIHDYTRVNDDVAAQQATNAKLKARNDQLFAEIDDLNGGQEALEER ARNELSMTRPGETFYRLVPDASKRAQ**C**AGQNNR
Strep-$FtsL_{35-121}$-C41A-C45A (Cys-less FtsL)	*MSNWSHPQFEKDYDIPTSENLYFQGGSG*FGKLPL**A**LFI**A**IILTAVTVVT TAHHTRLLTAQREQLVLERDALDIEWRNLILEENALGDHSRVERIATEK LQMQHVDPSQENIVVQK
Strep-$FtsL_{35-121}$-C41A-C45A-I100C	*MSNWSHPQFEKDYDIPTSENLYFQGGSG*FGKLPL**A**LFI**A**IILTAVTVVT TAHHTRLLTAQREQLVLERDALDIEWRNLILEENALGDHSRVER**C**ATEK LQMQHVDPSQENIVVQK

Table S2.3: Sequences of FtsB and FtsL used for *in vitro* FRET experiments.

1The sequences of the added purification tags (His_6 or Strep) are in italics. Mutations are hilighted in bold and underlined.

2.9 References

Aarsman, M.E.G., Piette, A., Fraipont, C., Vinkenvleugel, T.M.F., Nguyen-Distèche, M., and den Blaauwen, T. (2005). Maturation of the *Escherichia coli* divisome occurs in two steps: Maturation of the divisome. Mol. Microbiol. *55*, 1631–1645.

Adair, B.D., and Engelman, D.M. (1994). Glycophorin A Helical Transmembrane Domains Dimerize in Phospholipid Bilayers: A Resonance Energy Transfer Study. Biochemistry *33*, 5539–5544.

Alexeeva, S., Gadella, T.W.J., Verheul, J., Verhoeven, G.S., and Den Blaauwen, T. (2010). Direct interactions of early and late assembling division proteins in Escherichia coli cells resolved by FRET: Bacterial division studied by spectral FRET. Mol. Microbiol. *77*, 384–398.

Alva, V., Nam, S.-Z., Söding, J., and Lupas, A.N. (2016). The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. Nucleic Acids Res. *44*, W410–W415.

van den Berg van Saparoea, H.B., Glas, M., Vernooij, I.G.W.H., Bitter, W., den Blaauwen, T., and Luirink, J. (2013). Fine-mapping the Contact Sites of the *Escherichia coli* Cell Division Proteins FtsB and FtsL on the FtsQ Protein. J. Biol. Chem. *288*, 24340–24350.

Bisson-Filho, A.W., Hsu, Y.-P., Squyres, G.R., Kuru, E., Wu, F., Jukes, C., Sun, Y., Dekker, C., Holden, S., VanNieuwenhze, M.S., et al. (2017). Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division. Science *355*, 739–743.

de Boer, P.A. (2010). Advances in understanding E. coli cell fission. Curr. Opin. Microbiol. *13*, 730–737.

Bramkamp, M., Weston, L., Daniel, R.A., and Errington, J. (2006). Regulated intramembrane proteolysis of FtsL protein and the control of cell division in *Bacillus subtilis*: FtsL proteolysis in *Bacillus subtilis*. Mol. Microbiol. *62*, 580–591.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. *4*, 187–217.

Brown, J.H., Cohen, C., and Parry, D.A. (1996). Heptad breaks in alpha-helical coiled coils: stutters and stammers. Proteins *26*, 134–145.

Buddelmeijer, N., and Beckwith, J. (2004). A complex of the Escherichia coli cell division proteins FtsL, FtsB and FtsQ forms independently of its localization to the septal region: Assembly of a subcomplex of cell division proteins in E. coli. Mol. Microbiol. *52*, 1315–1327.

Buddelmeijer, N., Judson, N., Boyd, D., Mekalanos, J.J., and Beckwith, J. (2002). YgbQ, a cell division protein in Escherichia coli and Vibrio cholerae, localizes in codependent fashion with FtsL to the division site. Proc. Natl. Acad. Sci. *99*, 6316–6321.

Busiek, K.K., Eraso, J.M., Wang, Y., and Margolin, W. (2012). The Early Divisome Protein FtsA Interacts Directly through Its 1c Subdomain with the Cytoplasmic Domain of the Late Divisome Protein FtsN. J. Bacteriol. *194*, 1989–2000.

Chen, J.C., and Beckwith, J. (2001). FtsQ, FtsL and FtsI require FtsK, but not FtsN, for co-localization with FtsZ during Escherichia coli cell division. Mol. Microbiol. *42*, 395–413.

Cho, H., Wivagg, C.N., Kapoor, M., Barry, Z., Rohs, P.D.A., Suh, H., Marto, J.A., Garner, E.C., and Bernhardt, T.G. (2016). Bacterial cell wall biogenesis is mediated by SEDS and PBP polymerase families functioning semi-autonomously. Nat. Microbiol. *1*, 16172.

Daniel, R.A., and Errington, J. (2000). Intrinsic instability of the essential cell division protein FtsL of Bacillus subtilis and a role for DivIB protein in FtsL turnover. Mol. Microbiol. *36*, 278–289.

Daniel, R.A., Noirot-Gros, M.-F., Noirot, P., and Errington, J. (2006). Multiple Interactions between the Transmembrane Division Proteins of Bacillus subtilis and the Role of FtsL Instability in Divisome Assembly. J. Bacteriol. *188*, 7396–7404.

Delorenzi, M., and Speed, T. (2002). An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics *18*, 617–625.

Di Lallo, G. (2003). Use of a two-hybrid assay to study the assembly of a complex multicomponent protein machinery: bacterial septosome differentiation. Microbiology *149*, 3353–3359.

Du, S., Pichoff, S., and Lutkenhaus, J. (2016). FtsEX acts on FtsA to regulate divisome assembly and activity. Proc. Natl. Acad. Sci. *113*, E5052–E5061.

Egan, A.J.F., and Vollmer, W. (2015). The stoichiometric divisome: a hypothesis. Front. Microbiol. *6*.

Emami, K., Guyet, A., Kawai, Y., Devi, J., Wu, L.J., Allenby, N., Daniel, R.A., and Errington, J. (2017). RodA as the missing glycosyltransferase in Bacillus subtilis and antibiotic discovery for the peptidoglycan polymerase pathway. Nat. Microbiol. *2*, 16253.

van den Ent, F., Vinkenvleugel, T.M.F., Ind, A., West, P., Veprintsev, D., Nanninga, N., den Blaauwen, T., and Löwe, J. (2008). Structural and mutational analysis of the cell division protein FtsQ. Mol. Microbiol. *68*, 110–123.

Erickson, H.P., Anderson, D.E., and Osawa, M. (2010). FtsZ in Bacterial Cytokinesis: Cytoskeleton and Force Generator All in One. Microbiol. Mol. Biol. Rev. *74*, 504–528.

Ghigo, J.-M., and Beckwith, J. (2000). Cell Division in Escherichia coli: Role of FtsL Domains in Septal Localization, Function, and Oligomerization. J. Bacteriol. *182*, 116–129.

Ghigo, J.-M., Weiss, D.S., Chen, J.C., Yarrow, J.C., and Beckwith, J. (1999). Localization of FtsL to the Escherichia coli septal ring. Mol. Microbiol. *31*, 725–737.

Glas, M., van den Berg van Saparoea, H.B., McLaughlin, S.H., Roseboom, W., Liu, F., Koningstein, G.M., Fish, A., den Blaauwen, T., Heck, A.J.R., de Jong, L., et al. (2015). The Soluble Periplasmic Domains of *Escherichia coli* Cell Division Proteins FtsQ/FtsB/FtsL Form a Trimeric Complex with Submicromolar Affinity. J. Biol. Chem. *290*, 21498–21509.

Goehring, N.W. (2005). Premature targeting of a cell division protein to midcell allows dissection of divisome assembly in Escherichia coli. Genes Dev. *19*, 127–137.

Goehring, N.W., Gonzalez, M.D., and Beckwith, J. (2006). Premature targeting of cell division proteins to midcell reveals hierarchies of protein interactions involved in divisome assembly: Assembly of *Escherichia coli* division apparatus. Mol. Microbiol. *61*, 33–45.

Gonzalez, M.D., and Beckwith, J. (2009). Divisome under Construction: Distinct Domains of the Small Membrane Protein FtsB Are Necessary for Interaction with Multiple Cell Division Proteins. J. Bacteriol. *191*, 2815–2825.

Gonzalez, M.D., Akbay, E.A., Boyd, D., and Beckwith, J. (2010). Multiple Interaction Domains in FtsL, a Protein Component of the Widely Conserved Bacterial FtsLBQ Cell Division Complex. J. Bacteriol. *192*, 2757–2768.

Guzman, L.M., Weiss, D.S., and Beckwith, J. (1997). Domain-swapping analysis of FtsI, FtsL, and FtsQ, bitopic membrane proteins essential for cell division in Escherichia coli. J. Bacteriol. *179*, 5094–5103.

Hale, C.A., and de Boer, P.A. (1999). Recruitment of ZipA to the septal ring of Escherichia coli is dependent on FtsZ and independent of FtsA. J. Bacteriol. *181*, 167–176.

Hale, C.A., and de Boer, P.A.J. (2002). ZipA Is Required for Recruitment of FtsK, FtsQ, FtsL, and FtsN to the Septal Ring in Escherichia coli. J. Bacteriol. *184*, 2552–2556.

Haney, S.A., Glasfeld, E., Hale, C., Keeney, D., He, Z., and de Boer, P. (2001). Genetic Analysis of the *Escherichia coli* FtsZ·ZipA Interaction in the Yeast Two-hybrid System: CHARACTERIZATION OF FtsZ RESIDUES ESSENTIAL FOR THE INTERACTIONS WITH ZipA AND WITH FtsA. J. Biol. Chem. *276*, 11980–11987.

Harms, M.J., Schlessman, J.L., Sue, G.R., and Garcia-Moreno E., B. (2011). Arginine residues at internal positions in a protein are always charged. Proc. Natl. Acad. Sci. *108*, 18954–18959.

Hicks, M.R., Walshaw, J., and Woolfson, D.N. (2002). Investigating the Tolerance of Coiled-Coil Peptides to Nonheptad Sequence Inserts. J. Struct. Biol. *137*, 73–81.

Hopf, T.A., Schärfe, C.P.I., Rodrigues, J.P.G.L.M., Green, A.G., Kohlbacher, O., Sander, C., Bonvin, A.M.J.J., and Marks, D.S. (2014). Sequence co-evolution gives 3D contacts and structures of protein complexes. ELife *3*.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. J. Mol. Graph. *14*, 33–38.

Jo, S., Kim, T., Iyer, V.G., and Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. J. Comput. Chem. *29*, 1859–1865.

Kamisetty, H., Ovchinnikov, S., and Baker, D. (2013). Assessing the utility of coevolution-based residue-residue contact predictions in a sequence- and structure-rich era. Proc. Natl. Acad. Sci. *110*, 15674–15679.

Karimova, G., Dautin, N., and Ladant, D. (2005). Interaction Network among Escherichia coli Membrane Proteins Involved in Cell Division as Revealed by Bacterial Two-Hybrid Analysis. J. Bacteriol. *187*, 2233–2243.

Khadria, A.S., and Senes, A. (2013). The transmembrane domains of the bacterial cell division proteins FtsB and FtsL form a stable high-order oligomer. Biochemistry *52*, 7542–7550.

Kim, J.-K., Cho, Y., Lee, M., Laskowski, R.A., Ryu, S.E., Sugihara, K., and Kim, D.-S. (2015). BetaCavityWeb: a webserver for molecular voids and channels. Nucleic Acids Res. *43*, W413–W418.

Klauda, J.B., Venable, R.M., Freites, J.A., O'Connor, J.W., Tobias, D.J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A.D., and Pastor, R.W. (2010). Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. J. Phys. Chem. B *114*, 7830–7843.

Krivov, G.G., Shapovalov, M.V., and Dunbrack, R.L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. Proteins Struct. Funct. Bioinforma. *77*, 778–795.

Kulp, D.W., Subramaniam, S., Donald, J.E., Hannigan, B.T., Mueller, B.K., Grigoryan, G., and Senes, A. (2012). Structural informatics, modeling, and design with an open-source Molecular Software Library (MSL). J. Comput. Chem. *33*, 1645–1661.

LaPointe, L.M., Taylor, K.C., Subramaniam, S., Khadria, A., Rayment, I., and Senes, A. (2013). Structural organization of FtsB, a transmembrane protein of the bacterial divisome. Biochemistry *52*, 2574–2585.

Laskowski, RomanA., Rullmann, J.AntoonC., MacArthur, MalcolmW., Kaptein, R., and Thornton, JanetM. (1996). AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR *8*.

Levy, E.D., Pereira-Leal, J.B., Chothia, C., and Teichmann, S.A. (2006). 3D Complex: A Structural Classification of Protein Complexes. PLoS Comput. Biol. *2*, e155.

Li, W., Cowley, A., Uludag, M., Gur, T., McWilliam, H., Squizzato, S., Park, Y.M., Buso, N., and Lopez, R. (2015). The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res. *43*, W580–W584.

Liu, B., Persons, L., Lee, L., and de Boer, P.A.J. (2015). Roles for both FtsA and the FtsBLQ subcomplex in FtsN-stimulated cell constriction in *E scherichia coli*: Cell fission regulation by FtsN, FtsA and FtsBLQ. Mol. Microbiol. *95*, 945–970.

Liu, Z., Mukherjee, A., and Lutkenhaus, J. (1999). Recruitment of ZipA to the division site by interaction with FtsZ. Mol. Microbiol. *31*, 1853–1861.

Lupas, A., Van Dyke, M., and Stock, J. (1991). Predicting coiled coils from protein sequences. Science *252*, 1162–1164.

Lutkenhaus, J., Pichoff, S., and Du, S. (2012). Bacterial cytokinesis: From Z ring to divisome. Cytoskeleton *69*, 778–790.

MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †. J. Phys. Chem. B *102*, 3586–3616.

Mann, H.B., and Whitney, D.R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. Ann. Math. Stat. *18*, 50–60.

Männik, J., Bailey, M.W., O'Neill, J.C., and Männik, J. (2017). Kinetics of large-scale chromosomal movement during asymmetric cell division in Escherichia coli. PLOS Genet. *13*, e1006638.

Marks, D.S., Colwell, L.J., Sheridan, R., Hopf, T.A., Pagnani, A., Zecchina, R., and Sander, C. (2011). Protein 3D Structure Computed from Evolutionary Sequence Variation. PLoS ONE *6*, e28766.

Massey, T.H., Mercogliano, C.P., Yates, J., Sherratt, D.J., and Löwe, J. (2006). Double-Stranded DNA Translocation: Structure and Mechanism of Hexameric FtsK. Mol. Cell *23*, 457–469.

Masson, S., Kern, T., Le Gouëllec, A., Giustini, C., Simorre, J.-P., Callow, P., Vernet, T., Gabel, F., and Zapun, A. (2009). Central Domain of DivIB Caps the C-terminal Regions of the FtsL/DivIC Coiled-coil Rod. J. Biol. Chem. *284*, 27687–27700.

McDonnell, A.V., Jiang, T., Keating, A.E., and Berger, B. (2006). Paircoil2: improved prediction of coiled coils from sequence. Bioinformatics *22*, 356–358.

Meeske, A.J., Riley, E.P., Robins, W.P., Uehara, T., Mekalanos, J.J., Kahne, D., Walker, S., Kruse, A.C., Bernhardt, T.G., and Rudner, D.Z. (2016). SEDS proteins are a widespread family of bacterial cell wall polymerases. Nature *537*, 634–638.

Mercer, K.L.N., and Weiss, D.S. (2002). The Escherichia coli Cell Division Protein FtsW Is Required To Recruit Its Cognate Transpeptidase, FtsI (PBP3), to the Division Site. J. Bacteriol. *184*, 904–912.

Mohammadi, T., van Dam, V., Sijbrandi, R., Vernet, T., Zapun, A., Bouhss, A., Diepeveen-de Bruin, M., Nguyen-Distèche, M., de Kruijff, B., and Breukink, E. (2011). Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane: FtsW transports Lipid II across the membrane. EMBO J. *30*, 1425–1432.

Mohammadi, T., Sijbrandi, R., Lutters, M., Verheul, J., Martin, N.I., den Blaauwen, T., de Kruijff, B., and Breukink, E. (2014). Specificity of the Transport of Lipid II by FtsW in *Escherichia coli*. J. Biol. Chem. *289*, 14707–14718.

Mosyak, L. (2000). The bacterial cell-division protein ZipA and its interaction with an FtsZ fragment revealed by X-ray crystallography. EMBO J. *19*, 3179–3191.

Müller, P., Ewers, C., Bertsche, U., Anstett, M., Kallis, T., Breukink, E., Fraipont, C., Terrak, M., Nguyen-Distèche, M., and Vollmer, W. (2007). The Essential Cell Division Protein FtsN Interacts with the Murein (Peptidoglycan) Synthase PBP1B in *Escherichia coli*. J. Biol. Chem. *282*, 36394–36402.

Noirclerc-Savoye, M., Le Gouëllec, A., Morlot, C., Dideberg, O., Vernet, T., and Zapun, A. (2004). In vitro reconstitution of a trimeric complex of DivIB, DivIC and FtsL, and their transient co-localization at the division site in Streptococcus pneumoniae: Trimeric complex of DivIB, DivIC and FtsL. Mol. Microbiol. *55*, 413–424.

North, B., Summa, C.M., Ghirlanda, G., and DeGrado, W.F. (2001). D(n)-symmetrical tertiary templates for the design of tubular proteins. J. Mol. Biol. *311*, 1081–1090.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745.

Ortiz, C., Natale, P., Cueto, L., and Vicente, M. (2016). The keepers of the ring: regulators of FtsZ assembly. FEMS Microbiol. Rev. *40*, 57–67.

Ovchinnikov, S., Kinch, L., Park, H., Liao, Y., Pei, J., Kim, D.E., Kamisetty, H., Grishin, N.V., and Baker, D. (2015). Large-scale determination of previously unsolved protein structures using evolutionary information. ELife *4*, e09248.

Paintdakhi, A., Parry, B., Campos, M., Irnov, I., Elf, J., Surovtsev, I., and Jacobs-Wagner, C. (2016). Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis: Oufti: image analysis software. Mol. Microbiol. *99*, 767–777.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. J. Comput. Chem. *26*, 1781–1802.

Robichon, C., King, G.F., Goehring, N.W., and Beckwith, J. (2008). Artificial Septal Targeting of Bacillus subtilis Cell Division Proteins in Escherichia coli: an Interspecies Approach to the Study of Protein-Protein Interactions in Multiprotein Complexes. J. Bacteriol. *190*, 6048–6059.

Schmidt, K.L., Peterson, N.D., Kustusch, R.J., Wissel, M.C., Graham, B., Phillips, G.J., and Weiss, D.S. (2004). A Predicted ABC Transporter, FtsEX, Is Needed for Cell Division in Escherichia coli. J. Bacteriol. *186*, 785–793.

Schmidt, N.W., Grigoryan, G., and DeGrado, W.F. (2017). The accommodation index measures the perturbation associated with insertions and deletions in coiled-coils: Application to understand signaling in histidine kinases: Accommodation Index for Insertions in Coiled-Coils. Protein Sci. *26*, 414–435.

Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., and Serrano, L. (2005). The FoldX web server: an online force field. Nucleic Acids Res. *33*, W382–W388.

Spearman, C. (1904). The Proof and Measurement of Association between Two Things. Am. J. Psychol. *15*, 72.

Strelkov, S.V., Schumacher, J., Burkhard, P., Aebi, U., and Herrmann, H. (2004). Crystal Structure of the Human Lamin A Coil 2B Dimer: Implications for the Head-to-tail Association of Nuclear Lamins. J. Mol. Biol. *343*, 1067–1080.

Studier, F.W. (2005). Protein production by auto-induction in high-density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Subramaniam, S., and Senes, A. (2012). An energy-based conformer library for side chain optimization: Improved prediction and adjustable sampling. Proteins Struct. Funct. Bioinforma. *80*, 2218–2234.

Subramaniam, S., and Senes, A. (2014). Backbone dependency further improves side chain prediction efficiency in the Energy-based Conformer Library (bEBL): A Backbone-Dependent Energy-Based Library. Proteins Struct. Funct. Bioinforma. *82*, 3177–3187.

Szwedziak, P., Wang, Q., Freund, S.M., and Löwe, J. (2012). FtsA forms actin-like protofilaments: FtsA forms actin-like protofilaments. EMBO J. *31*, 2249–2260.

Tsang, M.-J., and Bernhardt, T.G. (2015). A role for the FtsQLB complex in cytokinetic ring activation revealed by an *fts L* allele that accelerates division: FtsL and divisome function. Mol. Microbiol. *95*, 925–944.

Villanelo, F., Ordenes, A., Brunet, J., Lagos, R., and Monasterio, O. (2011). A model for the Escherichia coli FtsB/FtsL/FtsQ cell division complex. BMC Struct. Biol. *11*, 28.

Vollmer, W., and Bertsche, U. (2008). Murein (peptidoglycan) structure, architecture and biosynthesis in Escherichia coli. Biochim. Biophys. Acta BBA - Biomembr. *1778*, 1714–1734.

Wang, L., Khattar, M.K., Donachie, W.D., and Lutkenhaus, J. (1998). FtsI and FtsW are localized to the septum in Escherichia coli. J. Bacteriol. *180*, 2810–2816.

Wolf, E., Kim, P.S., and Berger, B. (1997). MultiCoil: A program for predicting two-and three-stranded coiled coils: MultiCoil. Protein Sci. *6*, 1179–1189.

Yang, D.C., Peters, N.T., Parzych, K.R., Uehara, T., Markovski, M., and Bernhardt, T.G. (2011). An ATP-binding cassette transporter-like complex governs cell-wall hydrolysis at the bacterial cytokinetic ring. Proc. Natl. Acad. Sci. *108*, E1052–E1060.

Yang, J.-C., Van Den Ent, F., Neuhaus, D., Brevier, J., and Löwe, J. (2004). Solution structure and domain architecture of the divisome protein FtsN: Solution structure of FtsN. Mol. Microbiol. *52*, 651–660.

Yang, X., Lyu, Z., Miguel, A., McQuillen, R., Huang, K.C., and Xiao, J. (2017). GTPase activity–coupled treadmilling of the bacterial tubulin FtsZ organizes septal cell wall synthesis. Science *355*, 744–747.

Chapter 3: Functional analysis of the *E. coli* FtsLB periplasmic domain

This chapter is based on a manuscript which is currently being prepared for submission:

Running title: Functional analysis of the *E. coli* FtsLB periplasmic domain

Samuel J. Craven, Samson G.F. Condon, Gladys Diaz Vazquez, and Alessandro Senes

Statement of contribution

I performed the bioinformatic analyses of the FtsLB coiled coil domains and the molecular modeling of the rearranged FtsLB complex. I also modeled the structural effects of deletions within the juxtamembrane linker of FtsB.

3.1 Abstract

Cell division in bacteria is mediated by a multiprotein machine called the divisome, which acts at the division site to direct constrictive force and remodel the cell wall into a septum. In *Escherichia coli*, an essential component of the divisome is the FtsLB complex, which is necessary for recruiting other proteins to the division site and is involved with triggering septum formation. However, the molecular mechanisms behind FtsLB function and the overall structure of the complex are not well understood. Here, we provide evidence to support functional relevance for distinct regions of the FtsLB complex – namely, a marginally-stable FtsLB coiled coil and a constrained FtsB linker region. Multiple sequence alignments and *in vivo* complementation experiments demonstrate that nonideal interfacial residues within the FtsLB coiled coil have functional importance, possibly to enable conformational flexibility within the complex and/or to promote a dual-dimeric arrangement for periplasmic domain. We model such an arrangement *in silico* and present molecular dynamics simulations that complement our previous FtsLB models. We also provide *in vivo* evidence that the FtsB flexible, juxtamembrane linker region is constrained for length. Together, these results suggest that key regions throughout the FtsLB complex play an important role in cell division regulation by tuning FtsLB to the proper relative flexibility needed for signaling cell wall synthase activities.

3.2 Introduction

Cell division in bacteria is a complex process involving intricate coordination between numerous cellular components. Central to this coordination is the divisome, a multiprotein complex that in the Gram-negative bacterium *Escherichia coli* consists of 12 essential proteins (FtsZ, FtsA, ZipA, FtsE, FtsX, FtsK, FtsQ, FtsL, FtsB, FtsW, FtsI, and FtsN; Figure 3.1a) and a suite of nonessential, conditionally essential, and redundant proteins (den Blaauwen et al., 2017; Du and Lutkenhaus, 2017). These proteins mediate the various functions necessary for cell division to occur, including establishing the site of division, coordinating invagination of the inner and outer membranes during cytokinesis (the physical process of cell division), and remodeling the cell wall at mid-cell into a septum to compartmentalize the nascent daughter cells. If any of these functions is abrogated through deletion or mutation of essential proteins, the bacteria can continue to elongate and replicate their DNA, but they will be unable to divide and form distinct daughter cells, resulting in the formation of long filaments and eventual cell lysis and death.

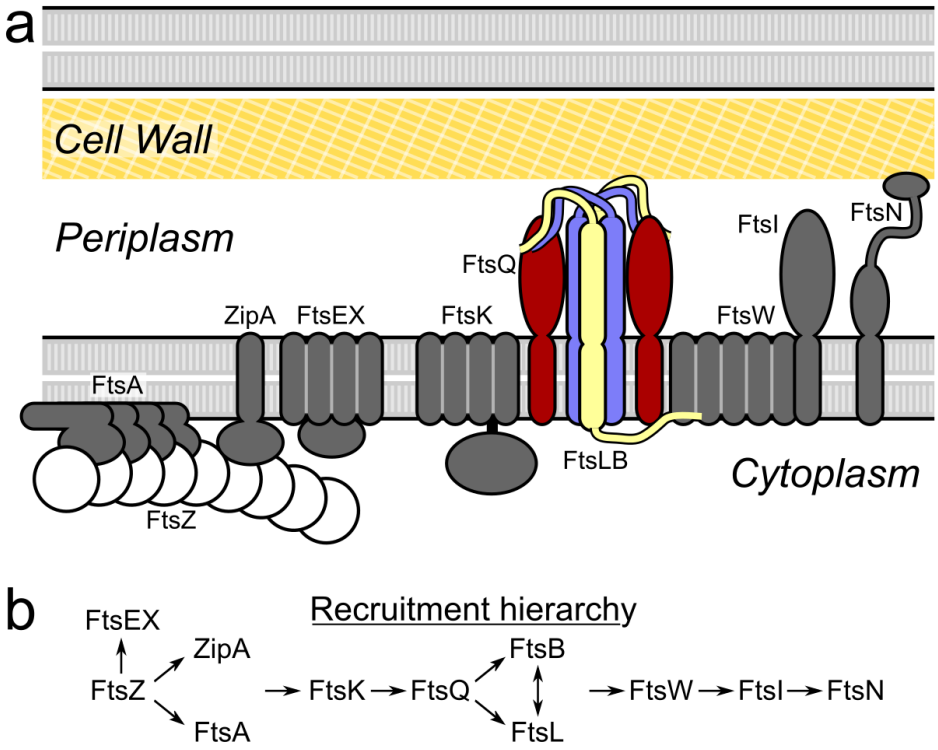

Figure 3.1: The *E. coli* divisome. a) Schematic representation of the essential early and late components of the *E. coli* divisome. Special emphasis is placed on the subcomplex formed by FtsQ (red), FtsL (yellow), and FtsB (blue). The FtsQLB subcomplex bridges the early components to the cell wall synthases. The stoichiometries presented are not necessarily representative of the actual divisome. b) The divisome assembles in a roughly hierarchical manner, with arrows representing order of dependence for localization to the septum. This hierarchy generally corresponds to the timing of recruitment to the division site, which divides the proteins into the early and late components.

The essential divisome proteins are recruited to the division site in a roughly hierarchical manner (Figure 3.1b) through interactions with a ring-like arrangement of FtsZ, a tubulin

homologue that readily polymerizes in the presence of GTP (Bi and Lutkenhaus, 1991; Fu et al., 2010; Löwe and Amos, 1998). This Z-ring (or proto-ring) is not a continuous structure, but instead consists of numerous laterally-overlapping FtsZ protofilaments that are constantly being turned over due to FtsZ GTPase activity. The concomitant polymerization and depolymerization from opposite ends results in directional movement of the entire protofilament without movement of the individual FtsZ monomers. This “treadmilling” activity somehow coordinates the spatiotemporal localization of other divisome components like the machinery necessary for remodeling the cell wall (Bisson-Filho et al., 2017; McCausland et al., 2019; Yang et al., 2017, 2019). The tethering of the FtsZ protofilaments to the cytoplasmic face of the inner membrane and the lateral lateral interactions of the individual filaments is regulated by FtsA, ZipA, and various other proteins (Hale and Boer, 2002; Hale and de Boer, 1997, 1999; Krupka et al., 2017; Pazos et al., 2013; Pichoff and Lutkenhaus, 2002, 2005; RayChaudhuri, 1999). FtsA also plays a critical role in regulating divisome assembly in a manner dependent on its oligomerization state, which is also influenced by ZipA and other downstream divisome proteins such as FtsEX (Pichoff et al., 2012; Shiomi and Margolin, 2007). FtsEX also activates peptidoglycan (PG) hydrolases in the periplasm, which may enable to coordinate divisome assembly with septal PG remodeling (Du et al., 2016; de Leeuw et al., 1999; Schmidt et al., 2004; Yang et al., 2011), the final critical step of cell division.

PG remodeling involves breakdown of old PG at the division site and the synthesis of new material to form first a septum and, eventually, the poles of the nascent daughter cells. Numerous nonessential enzymes are involved in this process (Booth and Lewis, 2019; Egan and Vollmer, 2013), but the synthesis of new material is performed primarily by the FtsWI complex (Cho et al., 2016; Fraipont et al., 2011; Mercer and Weiss, 2002). However, the mere presence of the

FtsWI complex at midcell is necessary but not sufficient for completion of cytokinesis, which must be first activated. The signal for activation of PG reconstruction coincides with the accumulation at midcell of FtsN (Addinall et al., 1997). Although the precise mechanism by which FtsN triggers PG recontruction is not known, it is thought to occur through of a positive-feedback loop. Interactions between the FtsN cytoplasmic tail and FtsA, along with interactions with other components in the periplasm, promote an “on” state that activates septal PG remodeling (Liu et al., 2015; Tsang and Bernhardt, 2015; Pichoff et al., 2015). The nascent PG that is formed as a result serves to recruit more FtsN to midcell via its SPOR domain (Gerding et al., 2009), and the increased FtsN localization, further activating the process.

Recent evidence indicates that the process associated with accumulation of FtsN that leads to the triggering of PG remodeling involves the participation of the FtsLB complex (Liu et al., 2015; Tsang and Bernhardt, 2015). In the absence of FtsN, the divisome assembles but PG reconstruction is not triggered. However, a series of ΔftsN-suppressing mutations were identified in FtsLB. These suppressing mutations are concentrated within a small region of the periplasmic domain named the Constriction Control Doman (CCD) (Liu et al., 2015; Tsang and Bernhardt, 2015). CCD mutants effectively rescue a ΔftsN phenotype, suggesting that changes in the FtsLB complex can mimic the signal given by FtsN. The CCD mutants are likely able to undergo through some structural changes – normally directly or indirectly induced by FtsN – that cause the FtsLB complex to switch from an “off” to “on” state. This switch, in turn, would signal to downstream components to begin the large-scale septal PG reconstruction that lead to completion of cell division.

FtsL and FtsB share a similar topology, with N-terminal transmembrane domains followed by periplasmic coiled coils. FtsB also contains a short, Gly-rich linker between the TM and coiled-

coil regions, which may provide flexibility between these regions and enable proper interactions with FtsL (Condon et al., 2018; LaPointe et al., 2013). Formation of the FtsLB complex is thought to be mediated largely through the TM and coiled-coil regions of both proteins (Buddelmeijer and Beckwith, 2004; Khadria and Senes, 2013; LaPointe et al., 2013; Robichon et al., 2011). However, the TM regions seem to be especially important as they can associate on their own (Khadria and Senes, 2013), whereas the coiled-coil regions alone are unable to interact within a soluble FtsLB construct lacking the TM regions (Glas et al., 2015; Kureisaite-Ciziene et al., 2018). Although partial crystal structures of FtsB have been solved either alone (LaPointe et al., 2013) or in complex with FtsQ (Choi et al., 2018; Kureisaite-Ciziene et al., 2018), the full structural arrangement of FtsLB has not been determined. We recently proposed a computational model for a heterotetrameric 2:2 arrangement (Condon et al., 2018), but this still awaits structural validation.

Our understanding of the function of the FtsLB subcomplex has developed much in the recent decades, evolving from an assumption that these small proteins merely act as scaffolds to aid in divisome assembly to a more nuanced role in regulating division. FtsL and FtsB depend on each other for stability (Buddelmeijer et al., 2002; Gonzalez and Beckwith, 2009), and experiments performed in *Bacillus subtilis* have indicated that the intrinsic instability of FtsL may play a regulatory role in cell division (Bramkamp et al., 2006; Daniel and Errington, 2000; Daniel et al., 2006). Detailed mechanisms for how FtsLB achieves regulation of cytokinesis are still forthcoming, but they are likely to necessitate structural changes within the complex. As discussed in our previous work (Condon et al., 2018; LaPointe et al., 2013), the Gly-rich juxtamembrane linker region in FtsB may provide flexibility to the complex and allow such structural rearrangements of FtsLB. On the other hand, the rigid FtsL may provide a backbone

enabling signal transduction within the complex in response to FtsN accumulation. Presumably, this signal transduction would initiate in the periplasm where both the essential domain of FtsN and the CCD of FtsLB are located (Liu et al., 2015; Tsang and Bernhardt, 2015). From there, the signal may be transmitted directly to the periplasmic regions of the PG synthesis machinery to relieve inhibition (Boes et al., 2019), or it may travel across the membrane into the cytoplasm, where the primary FtsL-FtsW interaction is located (Gonzalez et al., 2010). Either case would likely involve conformational changes in the regions near the CCD of FtsLB, meaning the FtsLB coiled coils and the FtsB juxtamembrane linker are particularly attractive targets of study.

In this chapter, we show that replacing polar and charged interfacial residues in the FtsLB coiled-coil interface with more canonical hydrophobic results in cell division defects *in vivo*. We also show that the charge identity of the interfacial residues in FtsL is important, indicating those residues may play a more nuanced role than simply destabilizing the hydrophobic coiled-coil interface. We also present evidence that the flexible juxtamembrane linker region of FtsB is constrained for length. We incorporate the *in vivo* results into computational models and molecular dynamics simulations that propose a splitting of the periplasmic domain of FtsLB into two coiled-coil pairs instead of a single heterotetramer as originally proposed (Condon et al., 2018).

3.3 Results and discussion

3.3.1 The coiled coil interfaces of FtsB and FtsL are enriched in polar and charged residues

The coiled-coil domain of *E. coli* FtsB contains multiple polar residues at interfacial *a* and *d* positions of the heptad repeat (LaPointe et al., 2013). The coiled-coil domain of *E. coli* FtsL also contains multiple charged residues at putatively interfacial positions and somewhat surprisingly,

coevolutionary analysis and molecular modeling suggest that the polar interfacial residues in FtsB do not fully complement the charged interfacial residues of FtsL. This could lead to a destabilized FtsLB complex (Condon et al., 2018). The presence of these non-optimal amino acids at the core of the FtsLB coiled coil raises the possibility that the complex evolved as a marginally-stable coiled coil, potentially to accommodate dynamics of the complex (e.g. during FtsN-mediated signaling of cytokinesis). Marginally-stable coiled-coil domains have been shown to have functional roles in other proteins including families of sensor kinases (Albanesi et al., 2009; Bhate et al., 2018; Schmidt et al., 2017), and in a specific case, the degree of hydrophobicity of the coiled-coil interface was shown to impact the rigidity of that region and cause a shift from kinase to phosphatase activity(Lesne et al., 2016). The presence of these non-optimal interfacial residues in FtsLB raises the possibility that similar functional relevance may be associated with a marginally stable FtsLB coiled coil.

We investigated the overall conservation of the polar and charged residues within the FtsB and FtsL core interface by aligning homologues from diverse bacterial taxa. There was substantial variation in the total number of polar and charged amino acids (defined here as Aspartate, Glutamate, Histidine, Arginine, Lysine, Glutamine, and Asparagine) at *a* and *d* positions in both FtsB and FtsL, but most homologues each had 3 of these residues spread across five coiled-coil heptads (Figure 3.2). This is an approximately threefold enrichment of hydrophillic residues at these positions compared to a database of known coiled coil structures (Testa et al., 2009).

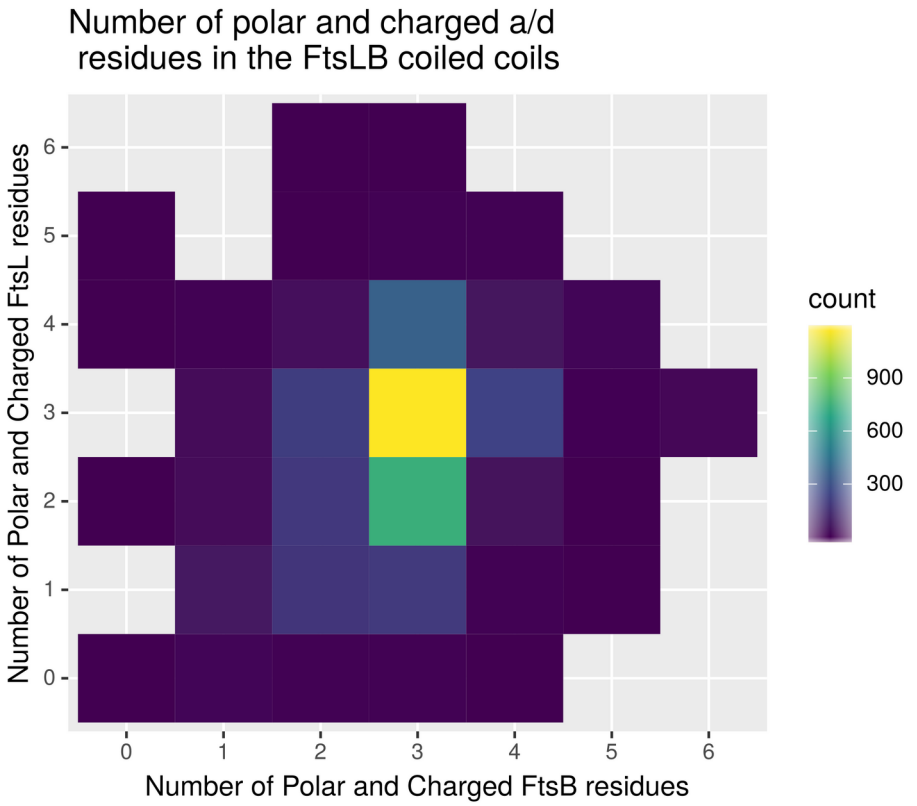

Figure 3.2: Distribution of polar and charged residues in homologues of FtsB and FtsL. Pairs of FtsB and FtsL sequences from individual bacterial taxa were assessed for the number of polar and charged residues at a and d positions within the heptad repeat. Most FtsL sequences contained between two and three polar and charged residues while most FtsB sequences contained three.

We identified strong patterns of conservation of highly polar and charged residues across the coiled-coil heptads. FtsB homologues had polar and charged residues at positions 2*d*, 3*a*, and 4*a* (corresponding to Q39, N43, and N50 in *E. coli* FtsB, respectively) with the vast majority of proteobacterial FtsL homologues having polar and charged residues at the *a* positions of heptad 3 and heptad 5 (corresponding to R74 and E88 in *E. coli* FtsL) (Figure 3.3). Intriguingly, for most

FtsL homologues only two positions are highly conserved to be polar and charged, while the location of the third polar or charged residue is somewhat variable (Figure 3.3, bottom). In contrast, the location of polar and charged residues in FtsB is almost completely conserved. Thus, both the position and total number of polar and charged residues is conserved in FtsB, while in FtsL the total number is more conserved than their positions.

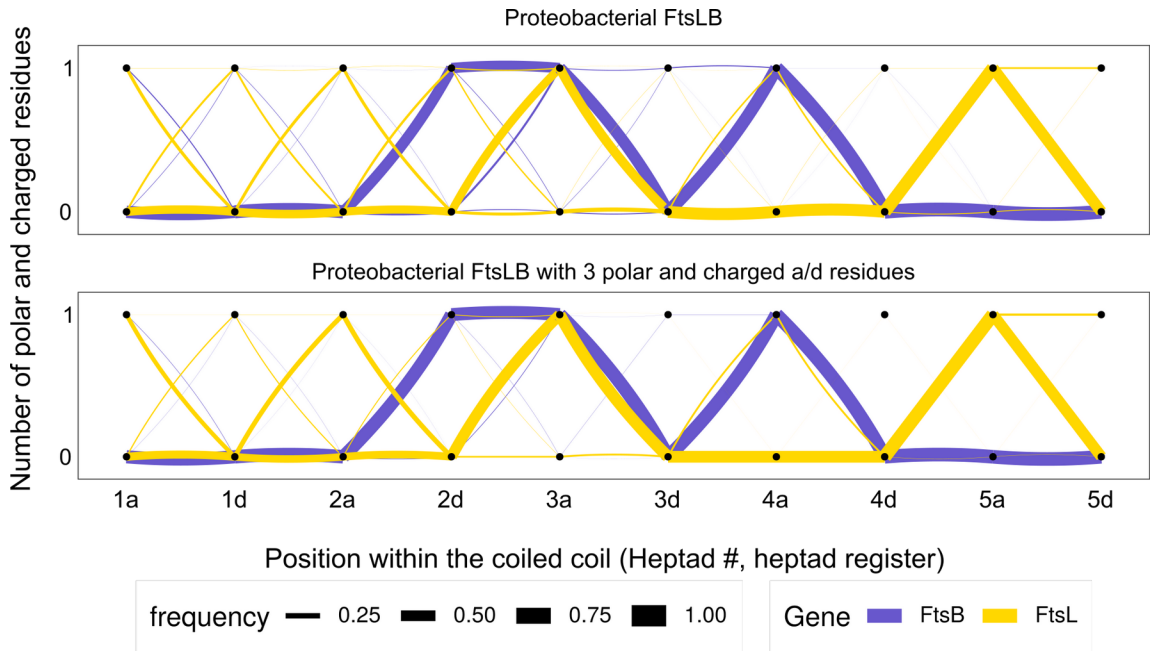

Figure 3.3: The positions of polar and charged residues at a/d registers in FtsLB homologues do not completely overlap. Line widths correspond to the frequency of observed patterns of polar and charged residues within the coiled coils of FtsL and FtsB. Top: pattern frequencies for proteobacterial FtsLB homologues. Bottom: pattern frequencies for pairs of proteobacterial FtsLB homologues that have 3 polar and charged a/d residues in each protein. Most FtsL homologues have polar and charged residues at positions 3a and 5a in the heptad repeat and the distribution of the third polar *or* charged residue is more variable.

3.3.2 The non-optimal residues at the coiled-coil interface are functionally important

In order to test whether these non-optimal amino acids are critical for function, we performed *in vivo* complementation experiments with mutant forms of FtsL and FtsB (Figure 3.4). We selected five of the polar and charged interfacial residues (FtsL R67/R74 and FtsB

Q39/N43/N50) for mutational analysis, changing residues at the *a* position to isoleucine and residues at the *d* position to leucine to mimic canonical coiled coils (Woolfson, 2005). We also mutated FtsL W81 (which is at the *a* position in heptad 4), since aromatic residues also tend to be excluded from canonical coiled-coil interfaces (Testa et al., 2009; Woolfson, 2005). However, we excluded FtsL E88 from further consideration as this residue lies within the FtsL CCD and may be involved in functions besides tuning the stability of the FtsLB interaction (Liu et al., 2015). Severe mutations in either protein produced cells that were unable to divide but continued elongating, resulting in long filaments. Less severe mutations gave rise to cells that retained a reduced ability to divide, therefore resulting in an average increase in cell length or subpopulations of elongated or filamentous cells. We compared the cell length distributions of the mutant and WT cells to assess the severity of each mutation. This was accomplished using the same classification scheme as previously described (Condon et al., 2018). In short, cells with lengths longer than the 95th percentile in the WT length distribution were considered “elongated”. Length distributions for the various mutations were compared to this threshold, and phenotypes with >15% of total cells being elongated were classified as “mild,” >25% as “intermediate,” >50% as “severe,” and “filamentous” when complete filamentation occurred. Four of the six mutations tested (FtsL R67I and W81I; FtsB Q39L and N50I) resulted in “mild” to “moderate” phenotypes when mutated to the hydrophobic amino acids leucine or isoleucine, whereas the remaining two mutations (FtsL R74I and FtsB N43I) were similar to WT. To exclude that the division defects were due to changes in protein expression levels, we performed western blot analyses (Figure S3:1), which indicated that each mutant was expressed to similar levels as WT (except for FtsB N43I, which showed increased expression). The effects of the arginine to isoleucine mutations in FtsL are particularly intriguing given that the effects contrast with their

conservation. The R74I mutation (corresponding to heptad 3*a*, where polar and charged residues are highly conserved) did not produce filamentation defects *in vivo* but the R67I mutation (corresponding to heptad 2*a*, which is a hydrophobic residue in most proteobacterial homologues) leads to moderate defects. These results suggest that the coiled coil of FtsLB has evolved to a marginal stability by including multiple non-optimal residues within the coiled coil interface, possibly to enable signaling during the ending stages of cell division.

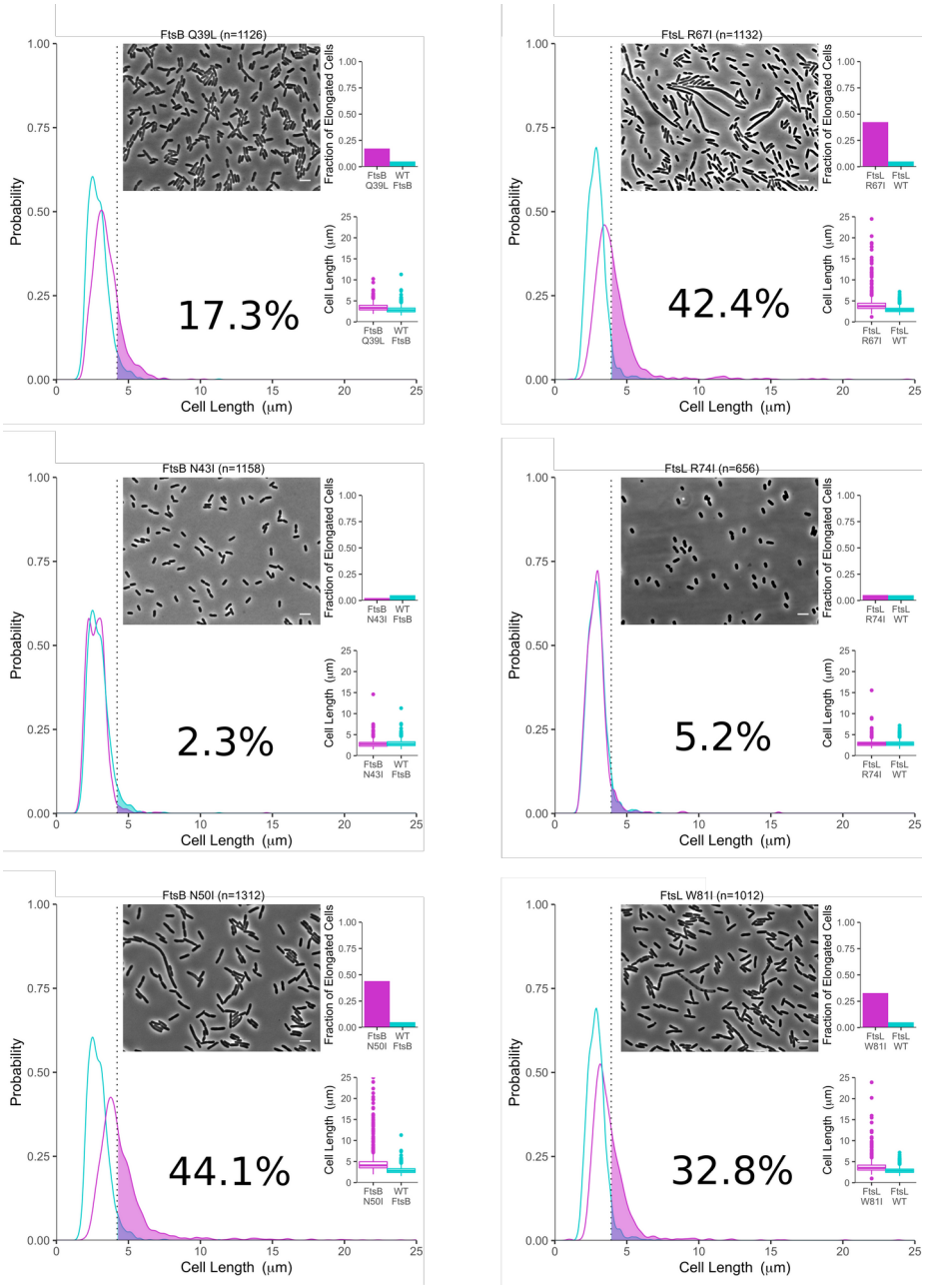

Figure 3.4: Mutations in the non-optimal coiled-coil residues cause defects in vivo. Cell length distributions (represented by kernel density estimation and a box plot) between wild-type (WT) cells (cyan; n=1296) and FtsB or FtsL mutants (magenta) to convert non-

optimal a and d coiled-coil residues to hydrophobic ones (Ile or Leu, respectively). Percentages correspond to the fraction of cells in the mutant population with a length greater than the 95%th percentile of the WT sample. Representative phase-contrast images for each mutant are included; scale bar 5μm.

If the primary function of the FtsL residues R67 and R74 is to provide instability to the coiled coil, then it is possible that the identity of the charged amino acid (either positive or negative) may be unimportant. In order to test this, we performed a double charge reversal on these residues (R67E+R74E) and tested for cell division defects *in vivo* (Table 1 and Figure S3:2). We found that the double charge reversal resulted in WT-like cells, indeed suggesting that a charge at those positions is sufficient, regardless of identity. Similarly, both R67K and R74K mutations also resulted in WT-like cells (Figure S3:3), which strengthens this hypothesis.

	R74 WT	R74I	R74E
R67 WT	5	42	38
R67I	5	39	16
R67E	2	3	9

Table 1: Summary of the effects charge reversal and removal mutations in FtsL R67 and R74 in the *in vivo* complementation experiments. Data represent the percent of elongated cells for each set of single and double mutations. Cell length distributions, summary statistics, and representative images shown in Figure 3.4 and Figure S3:2.

Interestingly, individual charge reversals told a different story. The R74E mutation resulted in cells that appeared largely WT-like or even a bit small, which is unsurprising considering the R74I mutation was also WT-like. The R67E mutation, on the other hand, produced a “moderate” phenotype, on par with the R67I mutation (38% and 42%, respectively). This finding is surprising, because it suggests that the R74E mutation somehow suppresses the R67E defect to reestablish WT-like cells in the double mutant.

There are a few possible explanations for the defective R67E mutant compared to the WT-like R67E+R74E mutant. First, the R67E mutation could lead to the formation of a salt bridge between the negative E67 residue and the positive R74 residue. Such an arrangement is more likely to promote coiled-coil stability compared to the WT R67/R74 arrangement, in which the unpaired positive charges would disfavor their burial in the hydrophobic coiled-coil interface. The formation of this salt bridge by the R67E mutation could therefore have similar functional consequences to hydrophobic mutations at that site (i.e. R67I), and the subsequent breakage of

the salt bridge by the R67E+R74E mutation could reestablish a WT phenotype. To test this hypothesis, we combined the R67E mutation with the R74I mutation. This double mutation would leave the negative charge on the E67 residue unpaired, and we expected the resulting phenotype to be WT-like since the R74I mutation alone causes no defect. However, the resulting R67E+R74I phenotype was “mild” (Table 1 and Figure S3:2), which argues against the formation of a salt bridge as the sole cause of the R67E division defect. An alternate explanation is that the two mutations are acting somewhat independently. As noted previously, the cells resulting from R74E alone appeared slightly smaller than WT, suggesting the mutation may cause early cell division. In this case, the suppression of R67E by R74E could be due to a simple competition between delayed cell division (R67E) and early cell division (R74E). Other such gain-of-function mutations in FtsLB have been identified before (Liu et al., 2015; Tsang and Bernhardt, 2015), predominantly within the CCD, and a similar effect may be seen with R74E. To test this, we examined if R74E could also suppress the “moderately” elongated phenotype of R67I. Like the case for R67E+R74E, the double R67I+R74E mutation resulted in WT-like cells, suggesting that R74E may indeed cause early cell division. Regardless of how FtsL positions 67 and 74 relate to each other, the overall phenotypes observed in this analysis are consistent with an important functional role for the polar amino acids of the FtsLB coiled coil – possibly in creating a marginal stability necessary for proper signaling to occur throughout the complex.

3.3.3 The coiled coil domains of the FtsLB complex can be rearranged into two heterodimers

We originally modeled the tetrameric complex of FtsB and FtsL as a parallel four-helix bundle spanning the transmembrane and coiled coil domains, with the two chains of each protein on a diagonal from each other. This arrangement oriented FtsB in a way that preserved the

homodimeric interface observed in its crystal structure (LaPointe et al., 2013). However, this arrangement buries the charged *a* and *d* positions of FtsL within the core, potentially destabilizing the complex. Indeed, molecular dynamics simulations of this model showed that water infiltrated the lower coiled coil of the complex while the membrane-embedded segment of the FtsLB complex remained relatively static (Condon et al., 2018). This suggested that the complex might be tuned for instability, but it is also possible that the arrangement of the coiled coils in the model is incorrect. For example, rearranging the four-helix bundle of the coiled coil domain into two heterodimeric FtsLB coiled coils would be consistent with *in vitro* evidence that these domains heterodimerize in the absence of their TMs (Glas et al., 2015).

We wondered whether an alternate arrangement of the coiled-coils could more favorably accommodate these polar and charged interfacial residues. Therefore, we modeled a FtsLB coiled coil heterodimer guided by the pairs of evolutionarily coupled residues (Hopf et al., 2014) between FtsL and FtsB used as constraints in the original modeling process. We connected this model of the coiled coils to the tetrameric model of the tetrameric FtsLB transmembrane domains and found that they could be accomodated without clashes or severe distortions (Figure 3.5a-b).

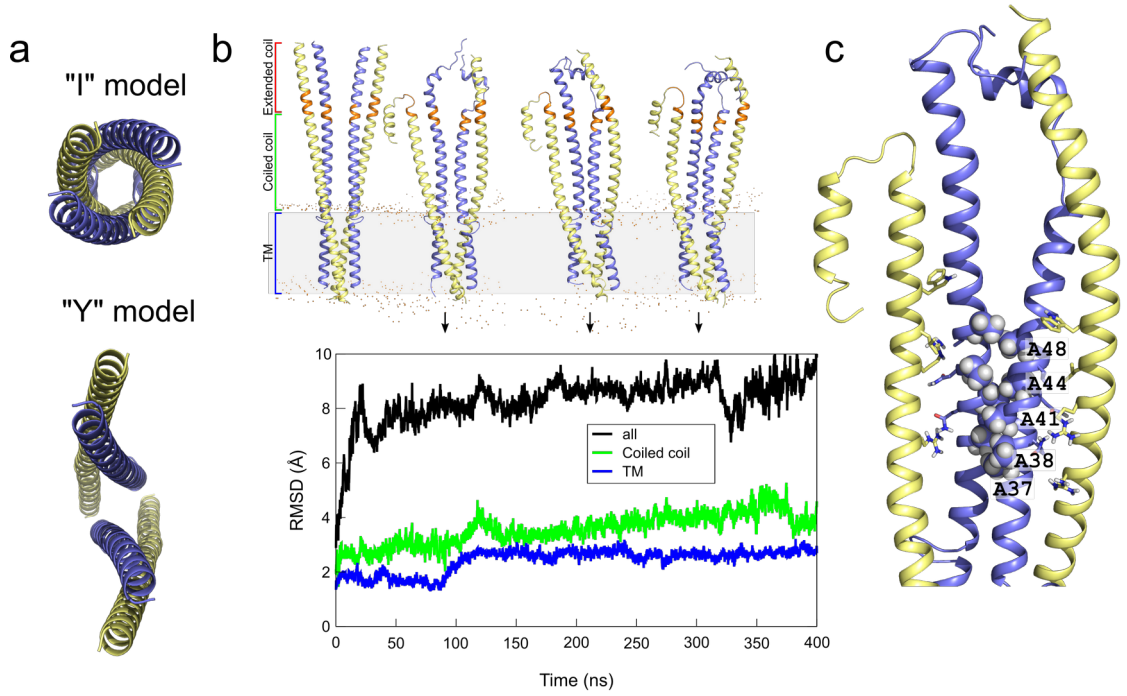

Figure 3.5: Modeling alternate arrangements of the FtsLB coiled coil. a) Top-down views from the periplasmic face of the FtsLB "I" model (Condon et al., 2018) and the rearranged "Y" model. In both models the transmembrane domains have the same parallel four-helix bundle arrangement. b) Molecular dynamics simulations of the rearranged FtsLB coiled coil. Top: Side views of representative snapshots of the trajectory. The CCD of FtsB and FtsL is colored orange. The model before equilibration is shown at the left. Bottom: RMSD values for the transmembrane domain (blue), the coiled coil domain (green), and the full model (black) over time. c) Over the course of the simulation, a cluster of alanine residues on FtsB (spheres) interact to form a secondary interface. The polar and charged side chains at a and d positions are shown as sticks.

3.3.4 Molecular dynamics simulations of the rearranged FtsLB complex

Molecular dynamics (MD) simulations of the rearranged FtsLB complex, which we designated the “Y” model, showed slightly higher divergence from the starting structure compared to the original “I” model (Figure 3.5b). However, water did not infiltrate the heterodimeric coiled coil interfaces as was previously observed (Condon et al., 2018). Over the course of the MD simulation of the “Y” model, a cluster of alanines (positions 37, 38, 41, 44, and 48) on the solvent-exposed faces of the FtsB coiled-coil domains began to interact, forming a secondary FtsB-FtsB interface (Figure 3.5c). To test if these residues are important for function, we mutated all five alanine residues to either glutamate or aspartate, since the introduction of so many negative charges so close together would likely disrupt any interface that formed around these positions. Unexpectedly, even with all five alanine residues mutated at once (AAAAA→DDDEE or AAAAA→EEEDD), no increase in the fraction of elongated cells was observed (Figure S3:4), indicating that these residues neither form a structurally important interface nor have any other essential function.

3.3.5 Circular dichroism spectra reveal structural effects of the non-optimal coiled-coil interface

Based on the *in vivo* complementation experiments, we hypothesized that increasing the hydrophobicity of the FtsLB coiled-coil interface causes cell division defects by over-rigidifying the complex. To obtain direct structural evidence for this hypothesis, we analyzed the stability of purified FtsLB complex *in vitro* using circular dichroism (CD) experiments. For these experiments, we used a version of FtsL lacking both the N-terminal tail ($FtsL_{35\text{-}121}$) as well as both native cysteine residues (C41A+C45A). The N-terminal tail of FtsL is likely unstructured

and would therefore reduce the average helical content of the complex. It is also unnecessary for the interaction between FtsL and FtsB (Gonzalez et al., 2010), so we rationalized its removal would be unlikely to impact the structure of the TM and coiled-coil regions of the complex. We have previously shown that the native FtsL cysteines are not important for cell division (Condon et al., 2018), so we elected to use a Cys-less version of FtsL to avoid using reducing agents that have strong absorbance in the UV region where CD is performed.

We obtained CD spectra comparing purified His-FtsB/Strep-FtsL$_{35-121}$ C41A-C45A (termed "WT" for these experiments) to His-FtsB Q39L-N43I/Strep-FtsL$_{35-121}$ C41A-C45A-R67I-R74I (termed "4x-mutant") at 4°C in detergent (Figure 3.6). The resulting spectra indicated little difference between the two samples, suggesting that either the mutations do not cause any structural effects or that these effects are minimal at low temperature. We then obtained spectra comparing the samples at increased temperatures (Figure S3:5). These initial data indicated that the 4x-mutant retains more helical content than WT at higher temperatures as well as after cooling back to the initial low temperature, which suggested a higher melting temperature for the 4x-mutant. To better compare the WT to the 4x-mutant samples, we performed thermal melt CD, monitoring the signal at 224 nm. Unexpectedly, the resulting melts did not indicate an increased melting temperature for the 4x-mutant. Instead, the increased helicity of the 4x-mutant at higher temperatures seems to be due to a transition from the sigmoidal curve seen in the WT sample to a much more linear curve, which precluded confident melting temperature determination. This transition in the shape of the melting curve could indicate a loss of cooperativity of unfolding for the 4x-mutant. However, it is also possible that multiple states of folding are involved. Further biophysical analyses would be needed to better explain what structural changes are caused by the mutations and whether the effects seen are due to the presence of detergent instead of the native

lipid environment, but some sort of structural difference between the WT and 4x-mutant complex is evident.

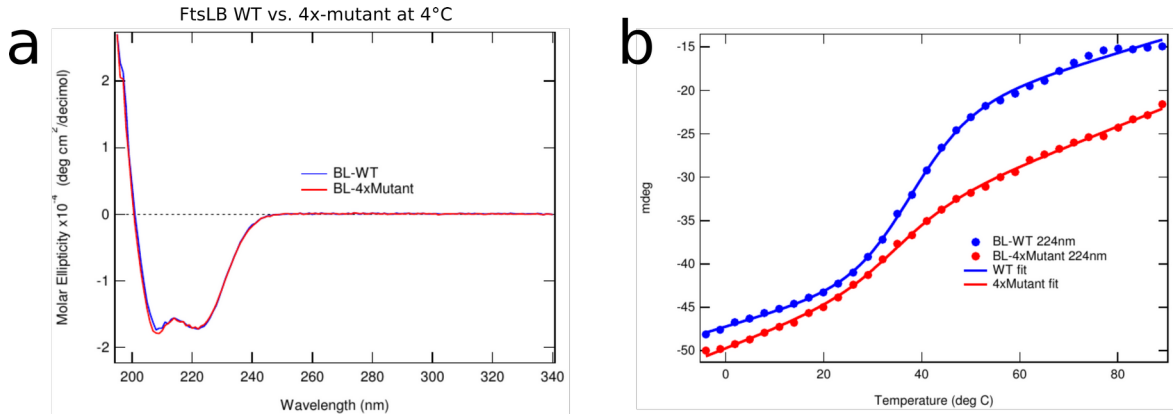

Figure 3.6: a) Far-UV CD spectra of WT FtsLB (blue) compared to the 4x-mutant (red). Essentially no difference is seen between the two at 4°C. b) Representative CD melt comparing WT FtsLB (blue) to the 4x-mutant (red). The 4x-mutant curve is less sigmoidal than WT, possibly indicating a loss of cooperativity of unfolding or the introduction of multiple states of unfolding. Scans were monitored at 224 nm from -4°C to 89°C.

3.3.6 The FtsB flexible juxtamembrane linker is constrained for length *in vivo* but tolerant of deletions *in silico*

In our previous work (Condon et al., 2018; LaPointe et al., 2013), we concluded that a juxtamembrane linker region in FtsB containing a pair of highly-conserved glycine residues (G22 and G25) is likely to adopt a flexible structure. This was demonstrated by inserting 1-3 alanine residues into the FtsB linker and the corresponding region in FtsL (which lacks such a Glycine-rich region), which resulted in minor elongation phenotypes for FtsB and complete filamentation for FtsL. We rationalized that the flexible structure of the FtsB linker region provides tolerance to residue insertion, whereas FtsL is likely a continuous helix and is unable to tolerate residue insertions.

If the FtsB linker allows residue insertion, then a logical follow-up question is whether it can tolerate residue deletion. To test this, we individually deleted residues K23, N24, and I26 within the FtsB linker and monitored cell division defects using the same *in vivo* complementation experiments as described earlier. As shown in Figure 3.7a, deletion of each of these residues caused “moderate” to “severe” elongation phenotypes without impacting expression levels, as analyzed via Western blot (Figure S3:1). The deletions were noticeably more severe than the corresponding alanine mutations, suggesting that it is the loss of linker length and not the loss of the specific residue identity that caused the defect. Interestingly, we were able to completely complement the deletion of either K23 or N24 by combining the deletion with a single alanine insertion following F21, which again suggests that it is the loss of length and not the loss of the specific residue that caused the division defect. Surprisingly, the single alanine insertion did not complement the deletion of Ile26.

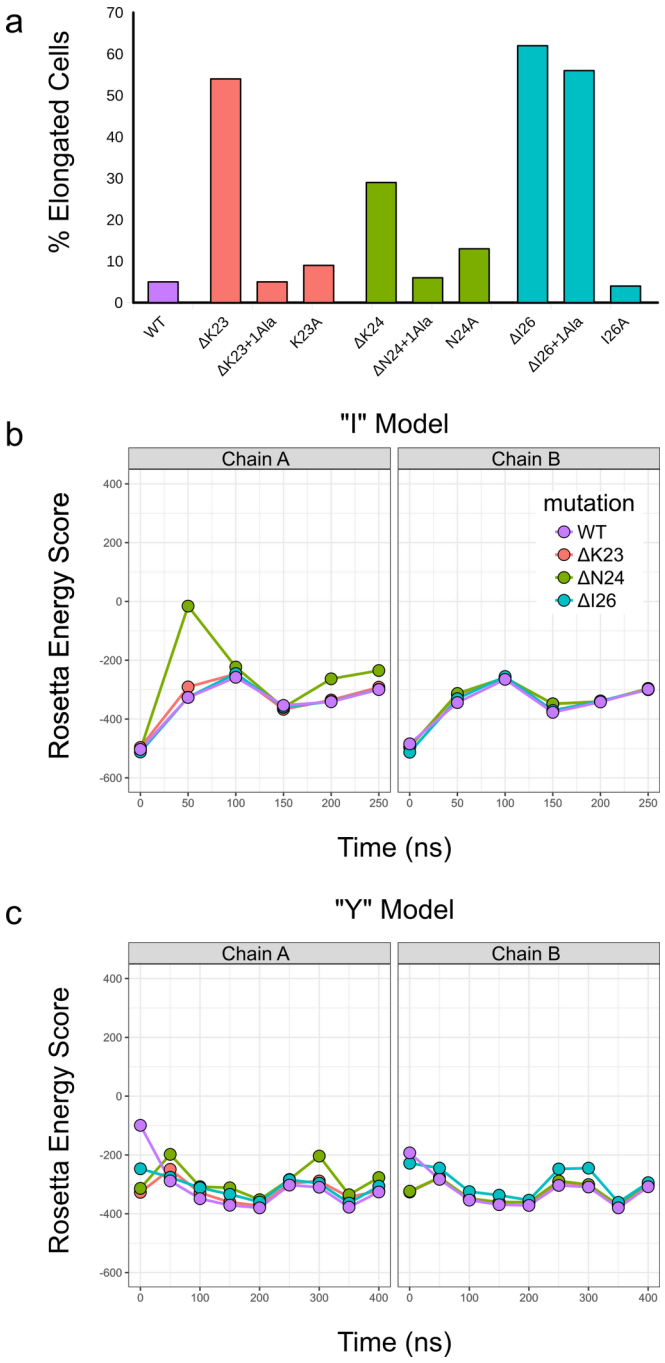

Figure 3.7: Deletions within the juxtamembrane linker of FtsB disrupt cell division *in vivo* but do not destabilize predicted models of the FtsLB complex. *a)* Fractions of elongated

cells measured from complementation experiments using deletion mutants show that the juxtamembrane linker is constrained for length. Snapshots of the all-atom molecular dynamics trajectories for both the “Y” model (*b*) and “I” model (*c*) were relaxed and subjected to a loop remodeling procedure to shorten the juxtamembrane linker of FtsB. Energy scores for the deletions are generally similar to when the full-length loop is remodeled (purple).

To investigate whether the *in vivo* impact of loop deletions could be recapitulated in our computational models, we used the Rosetta remodel application (Huang et al., 2011) to shorten the juxtamembrane linker of FtsB by individually deleting residues 23, 24, and 26. Since the linker deletions lead to pronounced filamentation defects, we would expect the models to be destabilized or for the remodeling protocol to fail in reconnecting the TM and coiled-coil domains of FtsB. However, reconstructing the shorter linker was achieved in all cases and the Rosetta energy score did not increase compared to remodeling the wild-type loop in either chain of the “I” model or the “Y” model (Figure 3.7). To investigate this further, we performed the same remodeling on snapshots of the molecular dynamics trajectories to investigate whether the FtsB linker explores a conformation that is sensitive to deletions. However, no conformation or set of conformations were identified that led to drastic changes in the energy score of the deletion mutants compared to the wild-type. Since the computational remodeling does not explain the effect of these deletions *in vivo*, we suspect that these deletions alter the conformation or dyamics of FtsB in a way our computational models have not sampled.

3.4 Conclusions and future directions

In this chapter, we identify various mutations within the FtsLB periplasmic region that cause defects in cell division, and use computational models to interpret these results. We show that the coiled coils of FtsL and FtsB contain interfacial residues that are normally disfavored in canonical coiled-coil sequences but are nevertheless crucial for function in this complex. It is likely that these residues impart a marginal stability to the FtsLB coiled coil and may also drive the complex to favor a dual-dimeric coiled-coil arrangement instead of a tetrameric one. The functional implications of these structural effects are largely hypothetical at the moment. Perhaps a highly stable FtsLB coiled coil would be too rigid to allow proper signaling of cytokinesis upon accumulation of FtsN. Alternately, a tetrameric arrangement might disfavor FtsQ binding, whereas the dual-dimeric form can more easily bind an FtsQ pair, as speculated for the 2:2:2 FtsQLB complex (van den Berg van Saparoea et al., 2013; Choi et al., 2018; Villanelo et al., 2011). It is even tempting to envision a scenario in which a switch between the dual-dimeric and tetrameric arrangement is crucial to FtsLB-mediated signaling – a conformational change that might not be possible were the coiled coil overly stable. Whether any of these scenarios actually represent reality remains to be seen, but the work presented here lays a foundation for further analysis into this essential cell division complex. It is likely that a combination of biophysical, biochemical, genetic, and computational experimentation will be required to directly address these questions, but the payoff will be worth the efforts. As essential components of a fundamental process within the bacterial life cycle, a better understanding of these proteins has important implications for any research involving bacterial growth and division. More specifically, as bacterial-specific proteins that lack clear human homologues, FtsL and FtsB represent excellent targets for drug design, and a better understanding of the structure of the

FtsLB complex is clearly needed is this regard. Moreover, as the periplasmic region of a protein is generally more accessible to drugs than cytoplasmic regions (which are separated from the surrounding environment by an extra membrane barrier), a detailed understanding of the structural arrangement of the FtsLB coiled coil is especially applicable in regards to drug design.

3.5 Methods

3.5.1 Plasmid cloning

For the *in vivo* complementation experiments, mutant variants of FtsB or FtsL were cloned via standard QuikChange mutagenesis or inverse PCR into pMDG7 (Gonzalez and Beckwith, 2009) (flag3-FtsB) or pMDG29 (Gonzalez et al., 2010) (flag3-FtsL), respectively. For the CD experiments, the His-tagged FtsB and Strep-tagged Cys-less (C41A and C45A) $FtsL_{35–121}$ were ligated into a modified pETDuet-1 vector at restriction sites NcoI/HindIII and NdeI/XhoI, respectively. Point mutations were introduced using standard QuikChange mutagenesis. All constructs were confirmed by DNA sequencing (Quintarabio).

3.5.2 Bacterial strains, plasmids, and media for in vivo experiments

The phenotypic analyses were performed using depletion strains NB946 (Buddelmeijer et al., 2002) for FtsB and MDG277 (Gonzalez and Beckwith, 2009) for FtsL (both obtained from Jon Beckwith and associates) in which the WT copy of the protein of interest is under control of a repressible P_{BAD} promoter within the chromosome. These strains were transformed with plasmids containing either WT protein (positive control), empty vector (negative control), or a mutant version of the protein to test for defect in cell division as evidenced by an increase in cell length. For all experiments described, bacterial cells were grown in LB medium supplemented with 100 μg/mL spectinomycin (Dot Scientific) and the appropriate carbon source. Medium was

supplemented with 0.2% (w/v) L-arabinose (Sigma) or 0.2% (w/v) D-glucose (Sigma) to induce or repress, respectively, the expression of chromosomal copies of the WT genes regulated by the P_{BAD} promoter. 20 μM isopropyl-β-D-1-thiogalactoside (IPTG) was added to the medium to induce the expression of mutant genes regulated by the P_{trc} promoter in the plasmid.

3.5.3 Depletion strain experiments

The protocol for the depletion strain experiments was adapted from Gonzalez and Beckwith (Gonzalez and Beckwith, 2009). In short, a mutated copy of FtsB or FtsL was transformed into its respective depletion strain. Strains were grown overnight at 37°C on an LB plate supplemented with arabinose and spectinomycin. A single colony from the plate was grown overnight at 37°C in 3 mL of LB medium supplemented with arabinose and spectinomycin. The overnight culture was then diluted 1:100 into fresh LB medium containing the same supplement and grown to an OD600 of ~0.3. An aliquot of 1 mL of culture was washed twice with LB medium lacking any sugar and then diluted 1:100 into 3 mL of fresh LB medium supplemented with glucose, IPTG, and antibiotic to induce the expression of the mutated gene and the repression of the WT gene. The cells were then grown at 37°C or 42°C for 3.5 hr, the approximate time necessary to deplete the cells of the WT chromosomal copy (Gonzalez and Beckwith, 2009). The cells were then placed on ice to stop growth before imaging. Depletion strains provided with their respective WT copy of the protein in the plasmid were tested as positive controls, and, similarly, depletion strains with no protein in the plasmid (empty vector) were tested as negative controls.

3.5.4 Microscopy and cell length measurements

10 μl of cell samples were mounted on a number 1.5, 24 X 50 mm (0.16 – 0.19 mm thickness) cover glass slide (Fisher or VWR). Cells were cushioned with a 3% (w/v) agarose gel pad to

restrict the movement of the live cells. Cells were optically imaged using a Nikon Eclipse Ti inverted microscope equipped with crossed polarizers and a Photometrics CoolSNAP HQ2 CCD camera using a Nikon X100 oil objective lens. Phase-contrast images of bacterial cells were recorded with a 50 ms exposure time using Nikon NIS Elements software. Multiple snapshots were collected for each experiment. All images were analyzed to measure the cell length in Oufti (Paintdakhi et al., 2016) using one single optimized parameter set and manual verification.

3.5.5 Western blots

Expression level across all variants was assessed by Western blot analysis (Figure S3:1). 3.0 mL of cells were pelleted and resuspended in 300 µl of lysis buffer (50 mM HEPES pH 8.0, 50 mM NaCl) with 5 mM βME. The cells were sonicated and centrifuged at 21,000g for 10 min before collecting the supernatant. Total protein concentration was determined by BCA assay (Pierce). 120 µl of lysates was mixed with 40 µl of 4x LDS sample buffer (Novex, Life Technologies) with βME and boiled at 98 °C for 3 min. For each FtsL or FtsB sample, the equivalent of 7 µg or 15 µg, respectively, of total protein was separated by SDS-PAGE (Invitrogen) and transferred to PVDF membrane (VWR). Horseradish peroxidase-tagged anti-FLAG (M2) antibodies (Sigma; 1:1,000) were used for immunoblotting analysis.

3.5.6 Protein expression and purification for CD

Plasmids were transformed into BL21(DE3) cells (NEB) and plated overnight at 37ºC on LB agar with 100 µg/mL ampicillin. Cells were washed off the plates with 1mL LB broth and inoculated into 1 L of ZYP-5052 autoinduction medium as described (Studier, 2005) and grown at 37°C until reaching an OD600 of ~0.8, after which they were incubated overnight at 22°C. Following expression, cells were pelleted, resuspended in cell wash buffer (100 mM NaCl + 10 mM HEPES pH 8.0), pelleted again, flash frozen, and stored at -80°C for future use. The cells

were then lysed by sonication in 10 mL/g lysis buffer (50 mM NaCl, 50 mM HEPES pH 8.0) supplemented with 0.5 mg/mL lysozyme, 5 mM βME, 1 mM phenylmethylsulfonyl fluoride, 1 mM EDTA, and a protease inhibitor cocktail providing (final concentrations) 8 μM leupeptin (Peptides International), 11.2 μM E-64 (Peptides International), 0.32 μM aprotinin (ProSpec), and 0.32 mM 4-(2-aminoethyl)benzenesulfonyl fluoride (Gold BioTechnology). The inclusion body fraction was separated by centrifugation at 10,000g for 20 min, followed by ultracentrifugation of the supernatant at 180,000g for 30 min to isolate the cell membranes. The FtsLB complex was then extracted from the membrane fraction with lysis buffer supplemented with 18 mM n-decyl-β-D-maltopyranoside (DM; Anatrace) and 5 mM βME, rocking at room temperature overnight. Non-resuspended debris was separated from the solubilized protein via centrifugation at 10,000g for 20 min. The supernatant was added to ~3 mL of Ni-NTA-agarose resin (Qiagen) and rocked for 2 h at 4°C before performing gravity-flow purification. Purification was performed by running 10 column volumes of Ni wash buffer (300 mM NaCl, 25 mM HEPES pH 8.0, 50 mM imidazole, 1 mM βME) supplemented with 510 μM n-dodecyl-β-D-maltopyranoside (DDM; Avanti Polar Lipids) and 10 column volumes of elution buffer (300 mM NaCl, 25 mM HEPES pH 8.0, 300 mM imidazole) also supplemented with 510 μM DDM. Protein purity was assessed via SDS-PAGE (Invitrogen).

3.5.7 CD experiments

Purified FtsLB protein was dialyzed twice at room temperature for at least 2 h into 1 L CD buffer (10 mM phosphate buffer pH 7.4, 100 mM NaF) supplemented with 170 mM DDM (1x critical micelle concentration to prevent detergent exchange), then overnight at 4°C in 1 L CD buffer supplemented with 510 mM DDM. Samples were kept at 4°C or on ice from this point forward. Protein concentration was determined against the final dialysis buffer using A280 and

an extinction coefficient of 32,430 M-cm- for the FtsLB complex (calculated via ExPASy). Protein was diluted to ~140 μM, then filtered with 0.22 μm (13 mm diameter) PVDF syringe filters (CELLTREAT) before redetermining the final protein concentration. Samples were degassed in a vacuum chamber for at least 30 min, then centrifuged for 20 min at 21,000g. The final dialysis buffer was also filtered and degassed in the same manner to use as a blank in the CD experiments. CD spectra were obtained using an Aviv model 420 CD spectrometer and quartz cuvettes with a 0.1 cm pathlength. All spectra were recorded in 1 nm increments, with either a 10 s or 20 s averaging time, and after a 5 min equilibration time upon reaching a 0.3ºC deadband. The spectra were baseline corrected by buffer subtraction. For the CD-monitored thermal melting experiments, the samples were heated at 3ºC intervals with a 10 min equilibration time and 20 s averaging time. Because the transitions were not reversible, detailed thermodynamic analyses were not carried out, and the curves were only fitted to sigmoidal transitions to calculate their temperature midpoints (Tm).

3.5.8 Bioinformatic analysis

Homologues of FtsB and FtsL were collected using the DELTA-BLAST algorithm (Boratyn et al., 2012) on the RefSeq database (O'Leary et al., 2016). FtsB-FtsL pairs were selected by the NCBI taxonomic identifier. In the case of multiple sequences per taxa, the one with the lowest E-value to the query FtsB or FtsL sequence was selected. Sequences were aligned using the MAFFT algorithm (Katoh and Standley, 2013). Statistical analyses were performed in R (R Core Team, 2019) with the aid of the following packages: tidyverse (Wickham, 2017), Biostrings (Pagès et al., 2019), zoo (Zeileis and Grothendieck, 2005), taxize (Chamberlain et al., 2019), rentrez (Winter, 2017), and tidygraph (Pedersen, 2019).

3.5.9 Modeling alternate arrangements of the FtsLB complex

Modeling of the rearranged FtsLB complex was performed as described previously (Condon et al., 2018). Briefly, the FtsLB heterodimer was modeled using a Monte Carlo procedure to model supercoiled helical bundles (Grigoryan and DeGrado, 2011). The superhelical radius (r_1), superhelical pitch (P), helical rotation (Φ_1), and z-shift (s) of both $FtsL_{52\text{-}94}$ and $FtsB_{21\text{-}63}$ were freely altered, whereas the rise per residue (h) and helical radius (r_0) were kept constant. Energies were calculated based on CHARMM 22 van der Waals and CHARMM 22 electrostatic terms with additional sigmoidal distance restraints for each pair of evolutionary couplings in the coiled coil region were added (Chapter 2, Table S2.1).

The heterodimeric FtsLB coiled coil was then aligned with one half of the previously modeled heterotetrameric transmembrane domain (Condon et al., 2018) using residues 52-58 of FtsL, which were present in both models. Both domains were kept parallel to the Z-axis. The juxtamembrane regions of FtsL and FtsB were then replaced with loops corresponding to fragments from the PDB. For FtsB, six-residue loops, corresponding to positions 21–26, with four flanking helical residues on each side, were used, with an additional sequence requirement that the fragment contain at least one glycine. For FtsL, 15-residue fragments with four flanking helical residues on each side were used with the requirement that the loop have helical secondary structure. This arrangement was made C_2 symmetric to generate the “Y” model. Finally, the side chains were repacked using a greedy trials algorithm and the model was minimized using BFGS constrained optimization in CHARMM (Brooks et al., 1983).

3.5.10 Molecular Dynamics Simulation

All-atom MD simulations were carried out as described previously (Condon et al., 2018). The “Y” model's coiled coil region was extended, to avoid edge effects, to residues 110 (FtsL) and 79 (FtsB). The cytoplasmic side of FtsL was also extended to include residues 30–34, modeled in ideal α-helix. A 400ns run was performed using the CHARMM 36 force field and NAMD version 2.10 software (Klauda et al., 2010; Phillips et al., 2005). CHARMM-GUI membrane builder (Jo et al., 2008) was used to prepare systems composed of a POPE bilayer consisting of 301 lipids, the FtsLB tetramer, an ionic concentration of 0.150 M NaCl, and 59,163 TIP3P water molecules for hydration. The sizes of the boxes at the beginning of the simulation was ~97 × 97 × 242 $Å^3$. The simulation was initially minimized and equilibrated for 75 ps at an integration time of 1 fs/step and for 600 ps at an integration time of 2 fs/step. The integration time step for the production run was 2.0 fs/step. The simulation was carried out in the NPT ensemble at a pressure of 1 atmosphere and a temperature of 310.15 K, using the Nose–Hoover Langevin piston and Langevin dynamics method. Particle mesh Ewald was used for electrostatic interactions, and a 12-Å cutoff was applied to Lennard–Jones interactions with a switching function from 10 to 12 Å. The RMSD analysis was performed using the RMSD trajectory tool in VMD (Humphrey et al., 1996).

3.5.11 Modeling juxtamembrane linker deletions in FtsB

Modeling the effect of deletions within the FtsB juxtamembrane linker was done using the remodel application in ROSETTA (Huang et al., 2011). To determine if there was a particular conformation of the linker region that was sensitive to deletions, template structures corresponding to 50ns snapshots of the MD trajectories for the “I” model and “Y” model were used. Each model was relaxed using the ROSETTA relax protocol with the relax:thorough flag

set to true and with constraints to the starting backbone and side chain coordinates (Nivón et al., 2013). One hundred trajectories were attempted for each deletion on each chain of FtsB with the remodel::quick_and_dirty set to true. Two flanking residues on either side of the deletion were allowed to be flexible during the remodeling procedure. The total energy score of the best trajectory for each run was selected for analysis. For comparison, this process was repeated for a “wild-type” remodel, where no residue was deleted and residues 22-25 of FtsB were allowed to be flexible.

3.6 Supplementary Figures

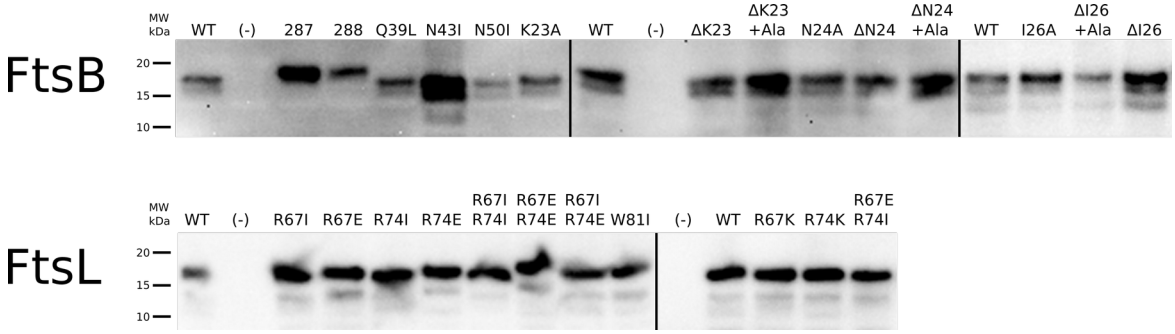

Figure S3:1: Representative Western blots of FtsB (top) and FtsL (bottom) mutants used for in vivo complementation experiments. Black bars represent separate blots.

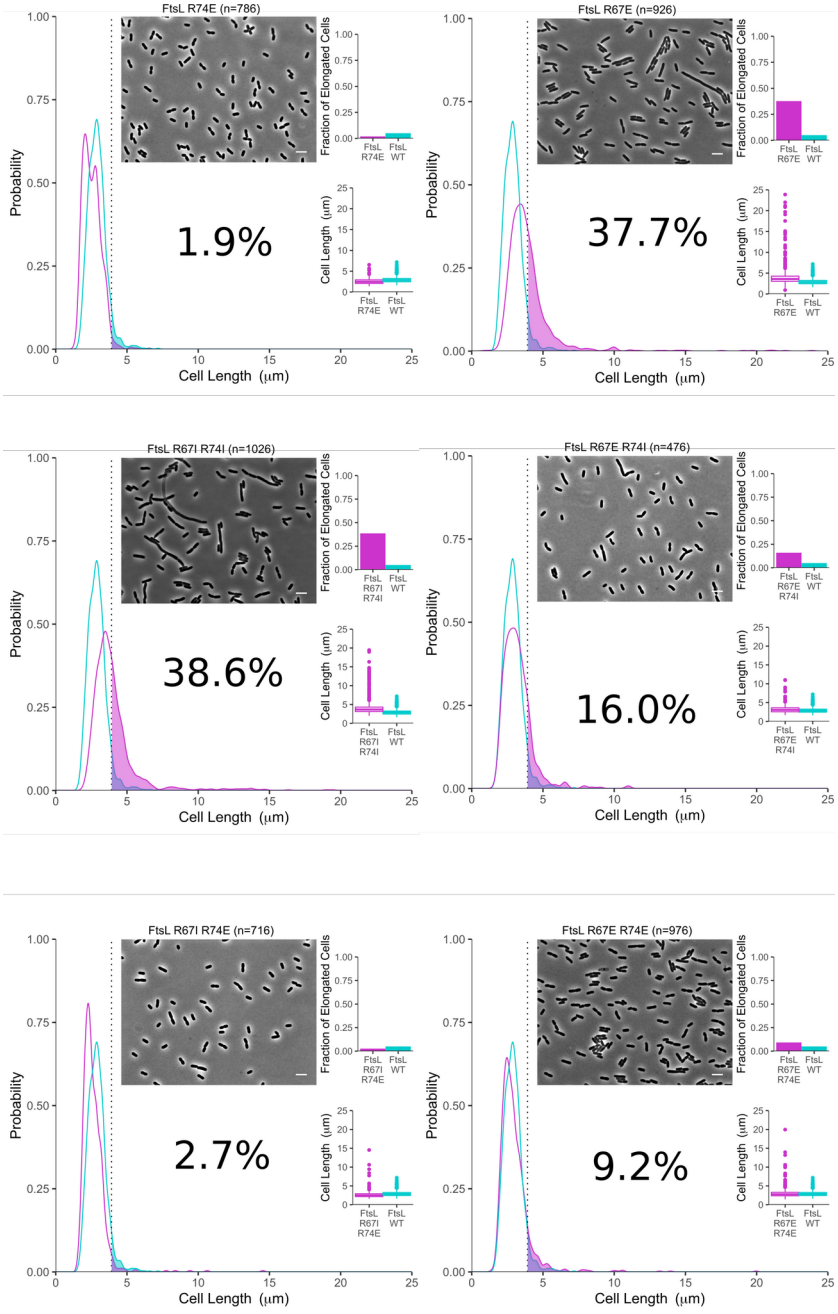

Figure S3:2: Single and double charge reversal and removal experiments for FtsL residues 67 and 74. Data presented in the same style as Figure 3.4.

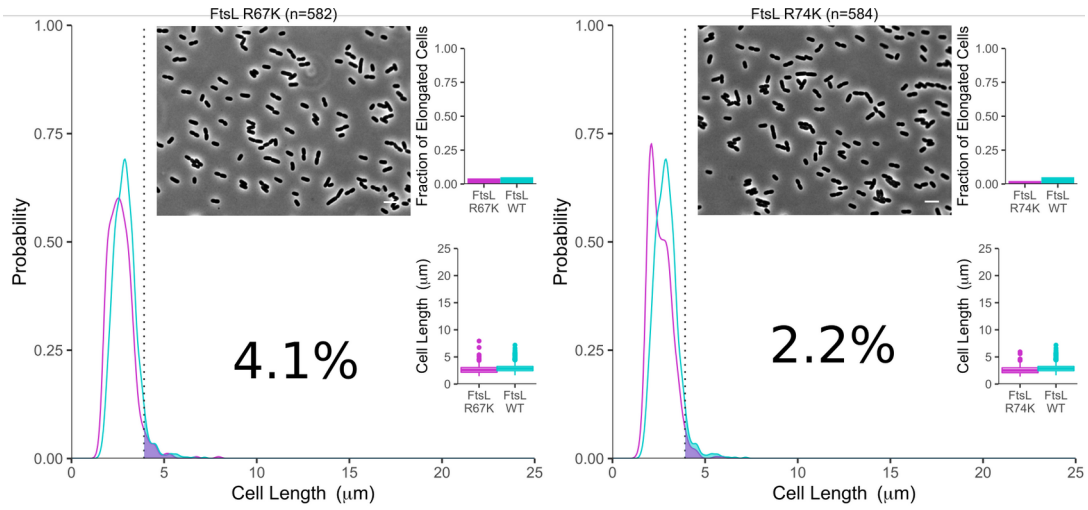

Figure S3:3: Arginine to Lysine mutations for FtsL residues 67 and 74. Data presented in the same style as Figure 3.4.

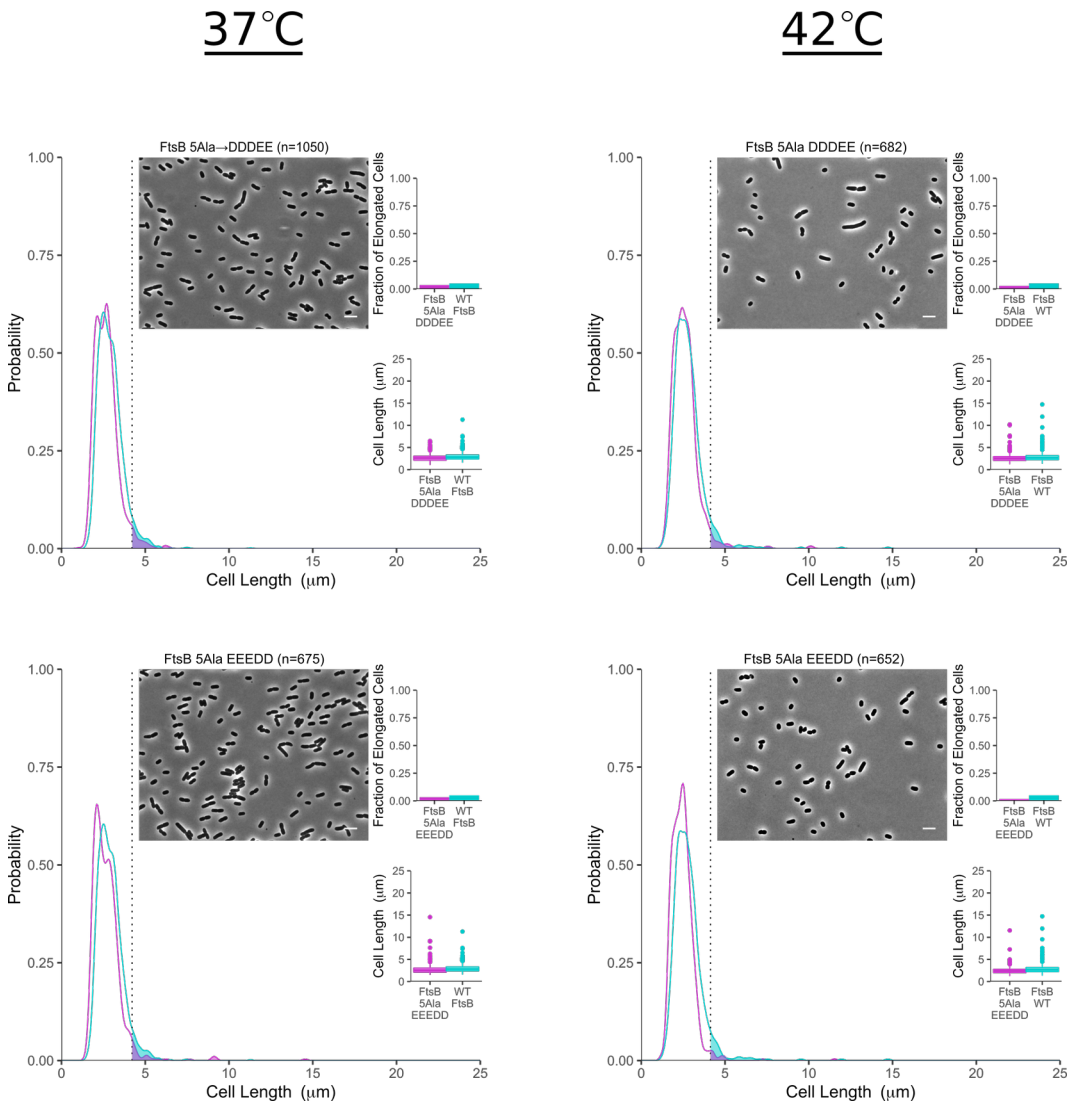

Figure S3:4: Mutations of alanines in the back face of the FtsB coiled coil to negatively charged residues do not impact cell division. Data presented in the same style as Figure 3.4. Data collected at 37° C (left) and 42° C (right) to increase phenotype severity. Scale bar 5μm.

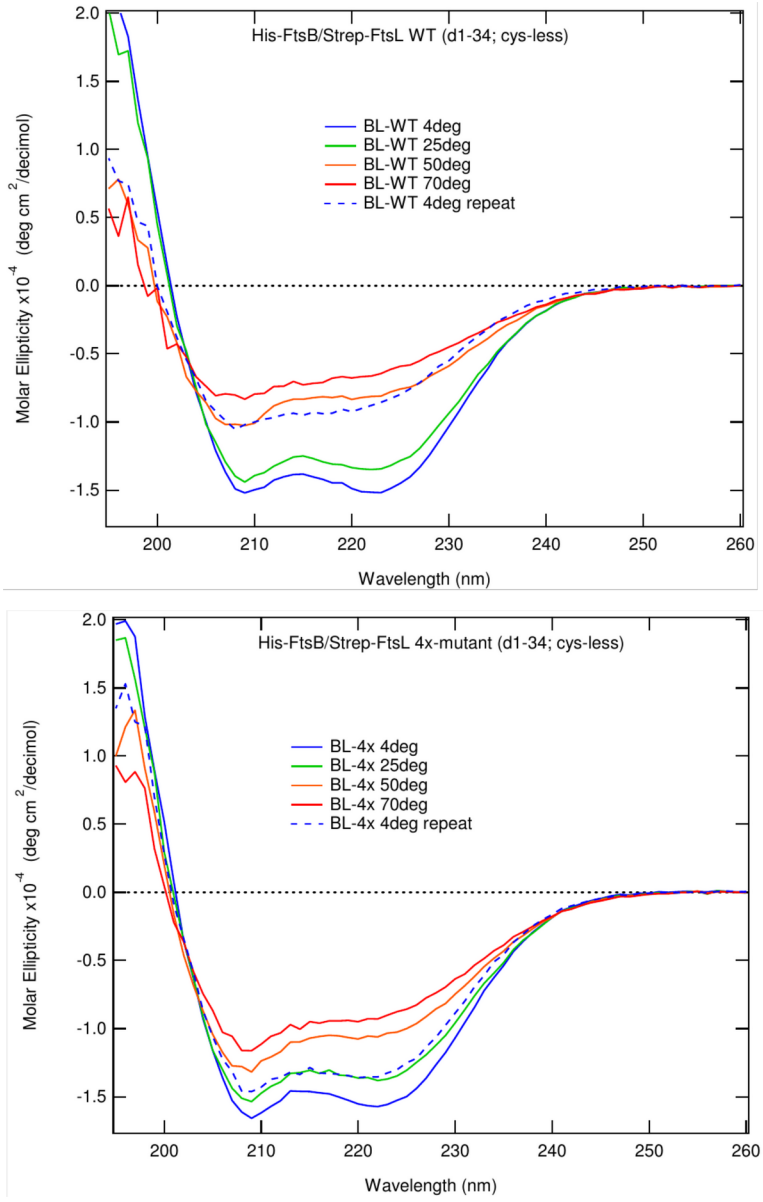

Figure S3:5: Far-UV CD Spectra of WT FtsLB (top) and the 4x-mutant (bottom) at different temperatures.

3.7 References

Addinall, S.G., Cao, C., and Lutkenhaus, J. (1997). FtsN, a late recruit to the septum in Escherichia coli. Mol. Microbiol. *25*, 303–309.

Albanesi, D., Martín, M., Trajtenberg, F., Mansilla, M.C., Haouz, A., Alzari, P.M., de Mendoza, D., and Buschiazzo, A. (2009). Structural plasticity and catalysis regulation of a thermosensor histidine kinase. Proc. Natl. Acad. Sci. U. S. A. *106*, 16185–16190.

van den Berg van Saparoea, H.B., Glas, M., Vernooij, I.G.W.H., Bitter, W., den Blaauwen, T., and Luirink, J. (2013). Fine-mapping the contact sites of the Escherichia coli cell division proteins FtsB and FtsL on the FtsQ protein. J. Biol. Chem. *288*, 24340–24350.

Bhate, M.P., Lemmin, T., Kuenze, G., Mensa, B., Ganguly, S., Peters, J.M., Schmidt, N., Pelton, J.G., Gross, C.A., Meiler, J., et al. (2018). Structure and Function of the Transmembrane Domain of NsaS, an Antibiotic Sensing Histidine Kinase in Staphylococcus aureus. J. Am. Chem. Soc. *140*, 7471–7485.

Bi, E.F., and Lutkenhaus, J. (1991). FtsZ ring structure associated with division in Escherichia coli. Nature *354*, 161–164.

Bisson-Filho, A.W., Hsu, Y.-P., Squyres, G.R., Kuru, E., Wu, F., Jukes, C., Sun, Y., Dekker, C., Holden, S., VanNieuwenhze, M.S., et al. (2017). Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division. Science *355*, 739–743.

den Blaauwen, T., Hamoen, L.W., and Levin, P.A. (2017). The divisome at 25: the road ahead. Curr. Opin. Microbiol. *36*, 85–94.

Boes, A., Olatunji, S., Breukink, E., and Terrak, M. (2019). Regulation of the Peptidoglycan Polymerase Activity of PBP1b by Antagonist Actions of the Core Divisome Proteins FtsBLQ and FtsN. MBio *10*.

Booth, S., and Lewis, R.J. (2019). Structural basis for the coordination of cell division with the synthesis of the bacterial cell envelope. Protein Sci. Publ. Protein Soc. *28*, 2042–2054.

Boratyn, G.M., Schäffer, A.A., Agarwala, R., Altschul, S.F., Lipman, D.J., and Madden, T.L. (2012). Domain enhanced lookup time accelerated BLAST. Biol. Direct *7*, 12.

Bramkamp, M., Weston, L., Daniel, R.A., and Errington, J. (2006). Regulated intramembrane proteolysis of FtsL protein and the control of cell division in Bacillus subtilis. Mol. Microbiol. *62*, 580–591.

Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. *4*, 187–217.

Buddelmeijer, N., and Beckwith, J. (2004). A complex of the Escherichia coli cell division proteins FtsL, FtsB and FtsQ forms independently of its localization to the septal region. Mol. Microbiol. *52*, 1315–1327.

Buddelmeijer, N., Judson, N., Boyd, D., Mekalanos, J.J., and Beckwith, J. (2002). YgbQ, a cell division protein in Escherichia coli and Vibrio cholerae, localizes in codependent fashion with FtsL to the division site. Proc. Natl. Acad. Sci. U. S. A. *99*, 6316–6321.

Chamberlain, S., Szoecs, E., Foster, Z., Arendsee, Z., Boettiger, C., Ram, K., Bartomeus, I., Baumgartner, J., O'Donnell, J., Oksanen, J., et al. (2019). taxize: Taxonomic information from around the web.

Cho, H., Wivagg, C.N., Kapoor, M., Barry, Z., Rohs, P.D.A., Suh, H., Marto, J.A., Garner, E.C., and Bernhardt, T.G. (2016). Bacterial cell wall biogenesis is mediated by SEDS and PBP polymerase families functioning semi-autonomously. Nat. Microbiol. *1*, 16172.

Choi, Y., Kim, J., Yoon, H.-J., Jin, K.S., Ryu, S., and Lee, H.H. (2018). Structural Insights into the FtsQ/FtsB/FtsL Complex, a Key Component of the Divisome. Sci. Rep. *8*, 18061.

Condon, S.G.F., Mahbuba, D.-A., Armstrong, C.R., Diaz-Vazquez, G., Craven, S.J., LaPointe, L.M., Khadria, A.S., Chadda, R., Crooks, J.A., Rangarajan, N., et al. (2018). The FtsLB subcomplex of the bacterial divisome is a tetramer with an uninterrupted FtsL helix linking the transmembrane and periplasmic regions. J. Biol. Chem. *293*, 1623–1641.

Daniel, R.A., and Errington, J. (2000). Intrinsic instability of the essential cell division protein FtsL of Bacillus subtilis and a role for DivIB protein in FtsL turnover. Mol. Microbiol. *36*, 278–289.

Daniel, R.A., Noirot-Gros, M.-F., Noirot, P., and Errington, J. (2006). Multiple interactions between the transmembrane division proteins of Bacillus subtilis and the role of FtsL instability in divisome assembly. J. Bacteriol. *188*, 7396–7404.

Du, S., and Lutkenhaus, J. (2017). Assembly and activation of the Escherichia coli divisome. Mol. Microbiol. *105*, 177–187.

Du, S., Pichoff, S., and Lutkenhaus, J. (2016). FtsEX acts on FtsA to regulate divisome assembly and activity. Proc. Natl. Acad. Sci. U. S. A. *113*, E5052-5061.

Egan, A.J.F., and Vollmer, W. (2013). The physiology of bacterial cell division. Ann. N. Y. Acad. Sci. *1277*, 8–28.

Fraipont, C., Alexeeva, S., Wolf, B., van der Ploeg, R., Schloesser, M., den Blaauwen, T., and Nguyen-Distèche, M. (2011). The integral membrane FtsW protein and peptidoglycan synthase PBP3 form a subcomplex in Escherichia coli. Microbiol. Read. Engl. *157*, 251–259.

Fu, G., Huang, T., Buss, J., Coltharp, C., Hensel, Z., and Xiao, J. (2010). In vivo structure of the E. coli FtsZ-ring revealed by photoactivated localization microscopy (PALM). PloS One *5*, e12682.

Gerding, M.A., Liu, B., Bendezú, F.O., Hale, C.A., Bernhardt, T.G., and de Boer, P.A.J. (2009). Self-enhanced accumulation of FtsN at Division Sites and Roles for Other Proteins with a SPOR domain (DamX, DedD, and RlpA) in Escherichia coli cell constriction. J. Bacteriol. *191*, 7383–7401.

Glas, M., van den Berg van Saparoea, H.B., McLaughlin, S.H., Roseboom, W., Liu, F., Koningstein, G.M., Fish, A., den Blaauwen, T., Heck, A.J.R., de Jong, L., et al. (2015). The Soluble Periplasmic Domains of Escherichia coli Cell Division Proteins FtsQ/FtsB/FtsL Form a Trimeric Complex with Submicromolar Affinity. J. Biol. Chem. *290*, 21498–21509.

Gonzalez, M.D., and Beckwith, J. (2009). Divisome under construction: distinct domains of the small membrane protein FtsB are necessary for interaction with multiple cell division proteins. J. Bacteriol. *191*, 2815–2825.

Gonzalez, M.D., Akbay, E.A., Boyd, D., and Beckwith, J. (2010). Multiple interaction domains in FtsL, a protein component of the widely conserved bacterial FtsLBQ cell division complex. J. Bacteriol. *192*, 2757–2768.

Grigoryan, G., and DeGrado, W.F. (2011). Probing Designability via a Generalized Model of Helical Bundle Geometry. J. Mol. Biol. *405*, 1079–1100.

Hale, C.A., and de Boer, P.A. (1997). Direct binding of FtsZ to ZipA, an essential component of the septal ring structure that mediates cell division in E. coli. Cell *88*, 175–185.

Hale, C.A., and de Boer, P.A. (1999). Recruitment of ZipA to the septal ring of Escherichia coli is dependent on FtsZ and independent of FtsA. J. Bacteriol. *181*, 167–176.

Hale, C.A., and Boer, P.A.J. de (2002). ZipA Is Required for Recruitment of FtsK, FtsQ, FtsL, and FtsN to the Septal Ring in Escherichia coli. J. Bacteriol. *184*, 2552–2556.

Hopf, T.A., Schärfe, C.P.I., Rodrigues, J.P.G.L.M., Green, A.G., Kohlbacher, O., Sander, C., Bonvin, A.M.J.J., and Marks, D.S. (2014). Sequence co-evolution gives 3D contacts and structures of protein complexes. ELife *3*.

Huang, P.-S., Ban, Y.-E.A., Richter, F., Andre, I., Vernon, R., Schief, W.R., and Baker, D. (2011). RosettaRemodel: A Generalized Framework for Flexible Backbone Protein Design. PLoS ONE *6*, e24109.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. J. Mol. Graph. *14*, 33–38.

Jo, S., Kim, T., Iyer, V.G., and Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. J. Comput. Chem. *29*, 1859–1865.

Katoh, K., and Standley, D.M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol. Biol. Evol. *30*, 772–780.

Khadria, A.S., and Senes, A. (2013). The transmembrane domains of the bacterial cell division proteins FtsB and FtsL form a stable high-order oligomer. Biochemistry *52*, 7542–7550.

Klauda, J.B., Venable, R.M., Freites, J.A., O'Connor, J.W., Tobias, D.J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A.D., and Pastor, R.W. (2010). Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. J. Phys. Chem. B *114*, 7830–7843.

Krupka, M., Rowlett, V.W., Morado, D., Vitrac, H., Schoenemann, K., Liu, J., and Margolin, W. (2017). Escherichia coli FtsA forms lipid-bound minirings that antagonize lateral interactions between FtsZ protofilaments. Nat. Commun. *8*, 15957.

Kureisaite-Ciziene, D., Varadajan, A., McLaughlin, S.H., Glas, M., Montón Silva, A., Luirink, R., Mueller, C., den Blaauwen, T., Grossmann, T.N., Luirink, J., et al. (2018). Structural Analysis of the Interaction between the Bacterial Cell Division Proteins FtsQ and FtsB. MBio *9*.

LaPointe, L.M., Taylor, K.C., Subramaniam, S., Khadria, A., Rayment, I., and Senes, A. (2013). Structural organization of FtsB, a transmembrane protein of the bacterial divisome. Biochemistry *52*, 2574–2585.

de Leeuw, E., Graham, B., Phillips, G.J., ten Hagen-Jongman, C.M., Oudega, B., and Luirink, J. (1999). Molecular characterization of Escherichia coli FtsE and FtsX. Mol. Microbiol. *31*, 983–993.

Lesne, E., Krammer, E.-M., Dupre, E., Locht, C., Lensink, M.F., Antoine, R., and Jacob-Dubuisson, F. (2016). Balance between Coiled-Coil Stability and Dynamics Regulates Activity of BvgS Sensor Kinase in Bordetella. MBio *7*, e02089.

Liu, B., Persons, L., Lee, L., and de Boer, P.A.J. (2015). Roles for both FtsA and the FtsBLQ subcomplex in FtsN-stimulated cell constriction in Escherichia coli. Mol. Microbiol. *95*, 945–970.

Löwe, J., and Amos, L.A. (1998). Crystal structure of the bacterial cell-division protein FtsZ. Nature *391*, 203–206.

McCausland, J.W., Yang, X., Lyu, Z., Söderström, B., Xiao, J., and Liu, J. (2019). Treadmilling FtsZ polymers drive the directional movement of sPG-synthesis enzymes via a Brownian ratchet mechanism. BioRxiv 857813.

Mercer, K.L.N., and Weiss, D.S. (2002). The Escherichia coli cell division protein FtsW is required to recruit its cognate transpeptidase, FtsI (PBP3), to the division site. J. Bacteriol. *184*, 904–912.

Nivón, L.G., Moretti, R., and Baker, D. (2013). A Pareto-Optimal Refinement Method for Protein Design Scaffolds. PLoS ONE *8*, e59004.

O'Leary, N.A., Wright, M.W., Brister, J.R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res. *44*, D733–D745.

Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S. (2019). Biostrings: Efficient manipulation of biological strings.

Paintdakhi, A., Parry, B., Campos, M., Irnov, I., Elf, J., Surovtsev, I., and Jacobs-Wagner, C. (2016). Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis. Mol. Microbiol. *99*, 767–777.

Pazos, M., Natale, P., and Vicente, M. (2013). A specific role for the ZipA protein in cell division: stabilization of the FtsZ protein. J. Biol. Chem. *288*, 3219–3226.

Pedersen, T.L. (2019). tidygraph: A Tidy API for Graph Manipulation.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. J. Comput. Chem. *26*, 1781–1802.

Pichoff, S., and Lutkenhaus, J. (2002). Unique and overlapping roles for ZipA and FtsA in septal ring assembly in Escherichia coli. EMBO J. *21*, 685–693.

Pichoff, S., and Lutkenhaus, J. (2005). Tethering the Z ring to the membrane through a conserved membrane targeting sequence in FtsA. Mol. Microbiol. *55*, 1722–1734.

Pichoff, S., Shen, B., Sullivan, B., and Lutkenhaus, J. (2012). FtsA mutants impaired for self-interaction bypass ZipA suggesting a model in which FtsA's self-interaction competes with its ability to recruit downstream division proteins. Mol. Microbiol. *83*, 151–167.

Pichoff, S., Du, S., and Lutkenhaus, J. (2015). The bypass of ZipA by overexpression of FtsN requires a previously unknown conserved FtsN motif essential for FtsA-FtsN interaction supporting a model in which FtsA monomers recruit late cell division proteins to the Z ring. Mol. Microbiol. *95*, 971–987.

R Core Team (2019). R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing).

RayChaudhuri, D. (1999). ZipA is a MAP-Tau homolog and is essential for structural integrity of the cytokinetic FtsZ ring during bacterial cell division. EMBO J. *18*, 2372–2383.

Robichon, C., Karimova, G., Beckwith, J., and Ladant, D. (2011). Role of leucine zipper motifs in association of the Escherichia coli cell division proteins FtsL and FtsB. J. Bacteriol. *193*, 4988–4992.

Schmidt, K.L., Peterson, N.D., Kustusch, R.J., Wissel, M.C., Graham, B., Phillips, G.J., and Weiss, D.S. (2004). A predicted ABC transporter, FtsEX, is needed for cell division in Escherichia coli. J. Bacteriol. *186*, 785–793.

Schmidt, N.W., Grigoryan, G., and DeGrado, W.F. (2017). The accommodation index measures the perturbation associated with insertions and deletions in coiled-coils: Application to understand signaling in histidine kinases. Protein Sci. Publ. Protein Soc. *26*, 414–435.

Shiomi, D., and Margolin, W. (2007). Dimerization or oligomerization of the actin-like FtsA protein enhances the integrity of the cytokinetic Z ring. Mol. Microbiol. *66*, 1396–1415.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Testa, O.D., Moutevelis, E., and Woolfson, D.N. (2009). CC+: a relational database of coiled-coil structures. Nucleic Acids Res. *37*, D315–D322.

Tsang, M.-J., and Bernhardt, T.G. (2015). A role for the FtsQLB complex in cytokinetic ring activation revealed by an ftsL allele that accelerates division. Mol. Microbiol. *95*, 925–944.

Villanelo, F., Ordenes, A., Brunet, J., Lagos, R., and Monasterio, O. (2011). A model for the Escherichia coli FtsB/FtsL/FtsQ cell division complex. BMC Struct. Biol. *11*, 28.

Wickham, H. (2017). tidyverse: Easily Install and Load the “Tidyverse.”

Winter, D.J. (2017). rentrez: an R package for the NCBI eUtils API. R J. *9*, 520–526.

Woolfson, D.N. (2005). The Design of Coiled-Coil Structures and Assemblies. In Advances in Protein Chemistry, (Elsevier), pp. 79–112.

Yang, D.C., Peters, N.T., Parzych, K.R., Uehara, T., Markovski, M., and Bernhardt, T.G. (2011). An ATP-binding cassette transporter-like complex governs cell-wall hydrolysis at the bacterial cytokinetic ring. Proc. Natl. Acad. Sci. U. S. A. *108*, E1052-1060.

Yang, X., Lyu, Z., Miguel, A., McQuillen, R., Huang, K.C., and Xiao, J. (2017). GTPase activity-coupled treadmilling of the bacterial tubulin FtsZ organizes septal cell wall synthesis. Science *355*, 744–747.

Yang, X., McQuillen, R., Lyu, Z., Phillips-Mason, P., Cruz, A.D.L., McCausland, J.W., Liang, H., DeMeester, K.E., Grimes, C.L., Boer, P. de, et al. (2019). FtsW exhibits distinct processive movements driven by either septal cell wall synthesis or FtsZ treadmilling in E. coli. BioRxiv 850073.

Zeileis, A., and Grothendieck, G. (2005). zoo: S3 Infrastructure for Regular and Irregular Time Series. J. Stat. Softw. *14*, 1–27.

Chapter 4: Evolutionary and combinatorial investigations of the BNIP3 transmembrane helix dimerization landscape

Statement of Contribution

I designed the ancestral BNIP3 sequences and interfacial combinations. I also predicted the structures of the BNIP3 variants and analyzed the computational data. Todd Jaszewski and Professor Kevin MacKenzie collected the TOXCAT data of BNIP3 homologues.

4.1 Abstract

Transmembrane helix oligomerization is a notable and often functionally important feature of many membrane proteins. Better understanding of the sequence requirements for oligomerization can lead to insight into membrane protein folding as well as its biological role. In this chapter, I explored the sequence landscape of the transmembrane domain of the mitophagy receptor BNIP3, examining what features of its interface are critical for maintaining its particularly high dimerization propensity. Using ancestral sequence reconstruction, I demonstrated that the helix-helix crossing point of the BNIP3 dimer shifted during its evolution, but this shift occurred in a manner that preserved its dimerization. Additionally, I explored shuffled interfacial sequences between human and nematode BNIP3 to better understand the relative importance of the GxxxG motif and hydrogen bonding node on BNIP3 oligomerization. Structure and energetic prediction of these mutants indicated that the location of the helix-helix crossing point may have a strong influence on self-association. Dimers that cross near the middle of the membrane bury more surface area than dimers with C-terminal crossing points, resulting in more favorable energy scores. These results contributed to our understanding of how individual transmembrane helices can associate so strongly and how such association may be preserved and changed through evolution.

4.2 Introduction

Lateral interactions between membrane proteins are a common structural and functional feature in diverse biological processes. Natural selection tunes these interactions to suit the functional context. The underlying sequence landscape places structural and functional constraints on the evolutionary process, and understanding how membrane protein interactions are preserved and changed can give insight into how they fold and their biological role.

A notable feature of the hypoxia-inducible mitochondrial protein BNIP3 is its behavior on SDS-PAGE gels, where the protein migrates as dimeric and monomeric species under both reducing and nonreducing conditions. BNIP3 truncations without the transmembrane domain migrate as a monomer, suggesting the SDS-resistant dimerization was solely determined through its TMD via noncovalent interactions (Chen et al., 1997). Further studies probed the self-association of the BNIP3 TMD in bacterial membranes, where it was also found to dimerize quite strongly (Lawrie et al., 2010; Sulistijo and MacKenzie, 2006; Sulistijo et al., 2003). While the wild-type BNIP3 TMD dimerizes with high propensity, dimerization is quite sensitive to mutagenesis along one face of the helix. Individual point mutations are sufficient to completely abolish association in both bacterial membranes as well as detergent.

Solution NMR structures for the BNIP3 dimer have been solved in dodecylphosphocholine (DPC) micelles (Sulistijo and MacKenzie, 2009) as well as lipid bicelles (Bocharov et al., 2007). The two structures are in close agreement for the general arrangement and interface of the BNIP3 dimer. The dimer interface contains a GxxxG sequence motif that associates in a GAS_{right} structural motif, a common interaction motif in transmembrane helices (Walters and DeGrado, 2006). The small amino acids in BNIP3 allow for close interhelical packing. The right-handed

crossing angle is optimized for the formation of intermolecular, noncanonical Cα-H hydrogen bonds (Mueller et al., 2014; Sulistijo and MacKenzie, 2009). The dimer is further stabilized via interhelical hydrogen bonds between Ser172 and His173. Disruption of either the GxxxG motif or the His-Ser hydrogen bond via mutagenesis strongly destabilizes the association on SDS-PAGE and in TOXCAT (Lawrie et al., 2010; Sulistijo and MacKenzie, 2006), indicating that neither is sufficient alone to drive the observed strong association. In summary, the combination of canonical and alpha carbon hydrogen bonds along with many van der Waals interactions due to the close interhelical packing in BNIP3 lead to a very stable homodimer.

The BNIP3 transmembrane domain is essential for both its pro-apoptotic (Chen et al., 1997) and pro-autophagic (Hanna et al., 2012) activities, but there is conflicting evidence on the role of dimerization in mediating cell death. BNIP3 is a tail-anchored membrane protein, and its TMD is sufficient to target proteins exclusively to mitochondria (Yasuda et al., 1998). Removing the TMD entirely prevents its cell death activity (Chen et al., 1997), but the TMD of BNIP3 could be replaced with an unrelated one and still induce cell death despite mis-localization (Ray et al., 2000). This same study identified mutations in the BNIP3 TMD that disrupt association on SDS-PAGE gels but have little impact on inducing apoptosis, which suggested that BNIP3 homodimerization is inessential for its apoptotic activity. However, mutations that only slightly weaken BNIP3 dimerization in bacterial membranes can have much more destabilizing effects on SDS-PAGE gels (Lawrie et al., 2010), so it is possible that these variants associate sufficiently in mitochondrial membranes to carry out BNIP3's function *in vivo*. In support of this, a later study found that the BNIP3 H173A mutant, which significantly reduces dimerization in TOXCAT, does not induce cell death as effectively as wild-type (Kubli et al., 2008). In contrast, homodimerization of the BNIP3 TMD seems to be important to carry out autophagy. BNIP3

G180F and BNIP3 H173A severely decrease dimerization in *E. coli* membranes as well as in SDS-PAGE (Lawrie et al., 2010) and these mutations failed to induce autophagy when expressed in HeLa cells. Furthermore, these mutations weaken the interaction between BNIP3 and the autophagosome receptor LC3, suggesting that the BNIP3 dimer is necessary for the interaction (Hanna et al., 2012). Since BNIP3 has both pro-apoptotic functions as well as anti-apoptotic, pro-autophagy functions, these findings may simply indicate that the role of the BNIP3 TMD is directed towards induction of autophagy instead of apoptosis.

Dimerization of BNIP3 appears to play an important role in its overall biological function, and this property appears to be conserved in homologues. The *C. elegans* BNIP3 protein (ceBNIP3) also dimerizes on SDS-PAGE gels and contains a putative interhelical hydrogen bonding node as well as a GxxxG motif, critical features of human BNIP3 dimerization (Cizeau et al., 2000). However, alignments of the sequences indicate the locations of these features are shifted between them. Since the dimer interface of BNIP3 is highly sensitive to mutagenesis, it is unclear how the diverging locations of these features originated through natural selection. Was BNIP3 dimerization lost and then regained at some point in its evolutionary history, or was there a pathway of individual mutations in the BNIP3 sequence landscape that preserved dimerization during the interfacial shift? To better understand how BNIP3 dimerization changed through the course of its evolution, we combined ancestral sequence reconstruction with computational structure prediction to examine BNIP3 dimerization structure and propensity throughout its evolutionary history. We also conducted a horizontal analysis of the BNIP3 dimerization sequence landscape by combinatorially shuffling the interfacial residues of distantly related BNIP3 homologues. These combinations of different interfaces helped us better understand the relative importance of the hydrogen bonding node and GxxxG motif, as well as the importance

of their shifted locations within the membrane. Together, the combined analyses probe how the sequence context can facilitate subtle structural changes while preserving very strong self-association of a single transmembrane helix.

4.3 Results and Discussion

4.3.1 BNIP3 homologs associate with similar affinities and interfaces

To understand the conservation of BNIP3 TMD self-association, the TMDs of hsBNIP3 and four other homologs from model organisms were analyzed using the TOXCAT assay (Russ and Engelman, 1999) (Figure 4.1a). The selected BNIP3 homologues came from *Xenopus laevis* (xlBNIP3), *Tetraodon nigroviridis* (tnBNIP3), *Caenorhabditis elegans* (ceBNIP3), *and Drosophila melanogaster* (dmBNIP3). While all of the sequences have a GxxxG motif and a strongly polar residue, the positions of these essential features are shifted between the vertebrates and invertebrates. As expected from previous studies (Lawrie et al., 2010; Sulistijo and MacKenzie, 2006), the TOXCAT signal for hsBNIP3 was ~2-3 fold higher than the signal for Glycophorin A, another SDS-resistant transmembrane helix homodimer. This high level of signal from the TOXCAT assay may actually be near the limit of saturation (Lawrie et al., 2010). Interestingly, the signal for the other BNIP3 homologs was almost exactly the same, suggesting they all associate at least as strongly as hsBNIP3 (Figure 4.1b).

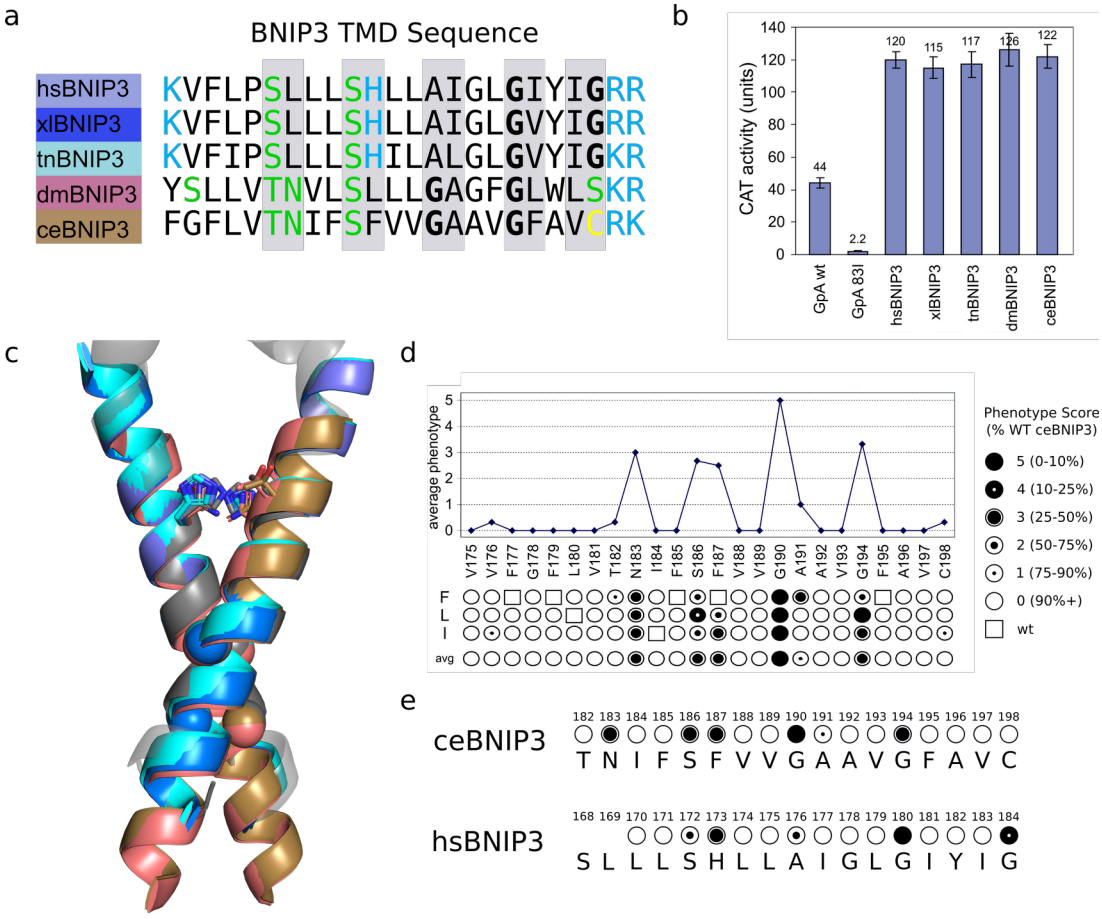

Figure 4.1: BNIP3 homologues self-associate with similar propensities despite shifts in critical interfacial features. *a)* multiple sequence alignment of representative BNIP3 TMDs from *Homo sapiens* (hsBNIP3, slate), *Xenopus laevis* (xlBNIP3, marine), Tetraodon *nigroviridis* (tnBNIP3, cyan), *Drosophila melanogaster* (dmBNIP3, pink), and *Caenorhabditis elegans* (ceBNIP3, brown). Positions involved in the BNIP3 dimer interface of hsBNIP3 are hilighted in gray boxes and GxxxG motifs shown in bold. *b)* TOXCAT measurements of BNIP3 TMDs expressed as chloramphenicol acetyltransferase

activity. The associated activity of the BNIP3 homologues is much higher than the strongly dimeric glycophorin A (GpA) TMD as well as a monomeric mutant (GpA G83I). Despite sequence differences at the interface, all homologues give similar levels of CAT activity. c) predicted structures of BNIP3 homologues aligned by their GxxxG motifs (spheres) and hydrogen bonding nodes (sticks), colored as in a. The NMR structure of hsBNIP3 is shown in gray. The crossing angle and interfacial residues of all models are very similar, but the helix-helix crossing point of dmBNIP3 and ceBNIP3 is shifted relative to the other models by one helical turn. d) Mutational analysis of ceBNIP3 self-association. Each position of ceBNIP3 was mutated to various hydrophobic residues to infer its effect on self-association in the TOXCAT assay. Phenotype scores are nonlinearly based on the level of dimerization relative to wild-type. e) Comparison of average mutational sensitivities of ceBNIP3 in TOXCAT with previously published sensitivities of hsBNIP3 (Lawrie et al., 2010). In both sequences, the GxxxG motif and the hydrogen bonding node are highly sensitive to mutagenesis, while the solvent-exposed residues are not.

To determine whether the predicted BNIP3 dimer interface is the same in both sets of homologues, we employed site-directed mutagenesis on the TMD of ceBNIP3, mutating each position to various hydrophobic residues. Such a mutagenesis strategy assumes that mutations at the helix-helix interface are more likely to be disruptive and reduce the signal in TOXCAT compared to mutations outside the interface (Sulistijo and MacKenzie, 2006). The mutagenesis revealed that the GxxxG motif is highly sensitive to substitution, and mutation of N183 in

ceBNIP3 to hydrophobic residues almost completely eliminated association (Figure 4.1d). Residues S186 and F187 (corresponding to hsBNIP3 residues S172 and H173, respectively) were also somewhat sensitive to mutagenesis, though not to the same extent. Comparison of the mutagenesis results in ceBNIP3 to hsBNIP3 shows that while the most sensitive positions differ between the two sequences, the importance of the GxxxG motif and the polar hydrogen bonding node is preserved (Figure 4.1e). This pattern matches the shift in the interfacial features as well as the change in interhelical crossing point observed in the predicted structures.

4.3.2 The hydrogen bonding node and GxxxG motif are conserved features of the BNIP3 TMD, but their locations are not

To determine if the shift in the GxxxG motif, hydrogen bonding node, and crossing point were similar in a broader set of BNIP3 homologues, we examined BNIP3 sequences from a recently published phylogenetic analysis of hypoxia receptors (Wu et al., 2017) (Figure 4.2, left).

As expected from the alignments of model organisms, certain features known to be important for self-association were almost completely conserved but their locations were not (Figure 4.2, center). The BNIP3 sequences in this set all had GxxxG motifs and a potential hydrogen bonding node that were separated by 8 residues. However, there were two possible locations for these features in the multiple sequence alignment (MSA) which were separated by four residues from one another. In invertebrate sequences, the location of the GxxxG motif and the hydrogen bonding node are shifted by four residues towards the N-terminus, with the i^{th} glycine in hsBNIP3 being the $i+4^{th}$ glycine in the invertebrate BNIP3 homologues.

An alternative hypothesis is that the multiple sequence alignment is incorrect and that the shift of critical features was generated by an insertion-deletion event. A four-residue deletion at the BNIP3 C-terminus coupled to a four-residue insertion at the N-terminus could generate the

observed pattern. However, we feel that such a hypothesis is unlikely for the following reasons. First, the multiple sequence alignment places the transmembrane domain and flanking regions within a highly conserved, gap-free block of residues. Transmembrane domains are confined to the membrane plane and insertion/deletion events within them are very rare (Taylor et al., 2004). Second, position 182 of the multiple sequence alignment (based on the sequence of hsBNIP3) contains a highly conserved aromatic residue, while position 178 typically contains a small residue such as alanine or glycine in both vertebrate and invertebrate sequences. An insertion/deletion event would have also created a split at these positions between vertebrates and invertebrates. Finally, a BLAST search identified some BNIP3 homologues with a tandem GxxxGxxxG motif, merging the individual GxxxG motifs found in the other sequences (Figure S4:1). We cannot exclude the possibility of a C-terminal deletion of four residues followed by an N-terminal insertion of four residues, but the likelihood is remote.

We next wondered whether these shifts would lead to shifts in the helix-helix crossing point. As expected, the structures predicted by CATM for these homologues are remarkably similar to one another. Global alignments of the top models to those generated for hsBNIP3 and ceBNIP3 show very little divergence. These models fall into two clusters, matching the shift in crossing point observed in the homologues selected for TOXCAT (Figure 4.2, right). The clusters of models seem to be taxonomically related, with deuterostomal sequences clustering with the hsBNIP3 model and protostomal sequences clustering with the ceBNIP3 model. The location of the GxxxG motif in the sequences is almost completely correlated with the corresponding location of the crossing point. Interestingly, these similarities persist even though not all features of the BNIP3 sequence are perfectly conserved. For example, mollusc BNIP3 TMDs in the tree do not contain a GxxxG motif at all, instead having an A/T/SxxxGxxxM interfacial sequence.

However, the top models from these sequences are more similar to the ceBNIP3 model than the hsBNIP3 one. Methionine is a very bulky residue and would not be accommodated at the C1 position in a GAS_{right} homodimer (Mueller et al., 2014). Thus, the only favorable dimer models in these sequences use the Sm-xxx-G sequence as the interface.

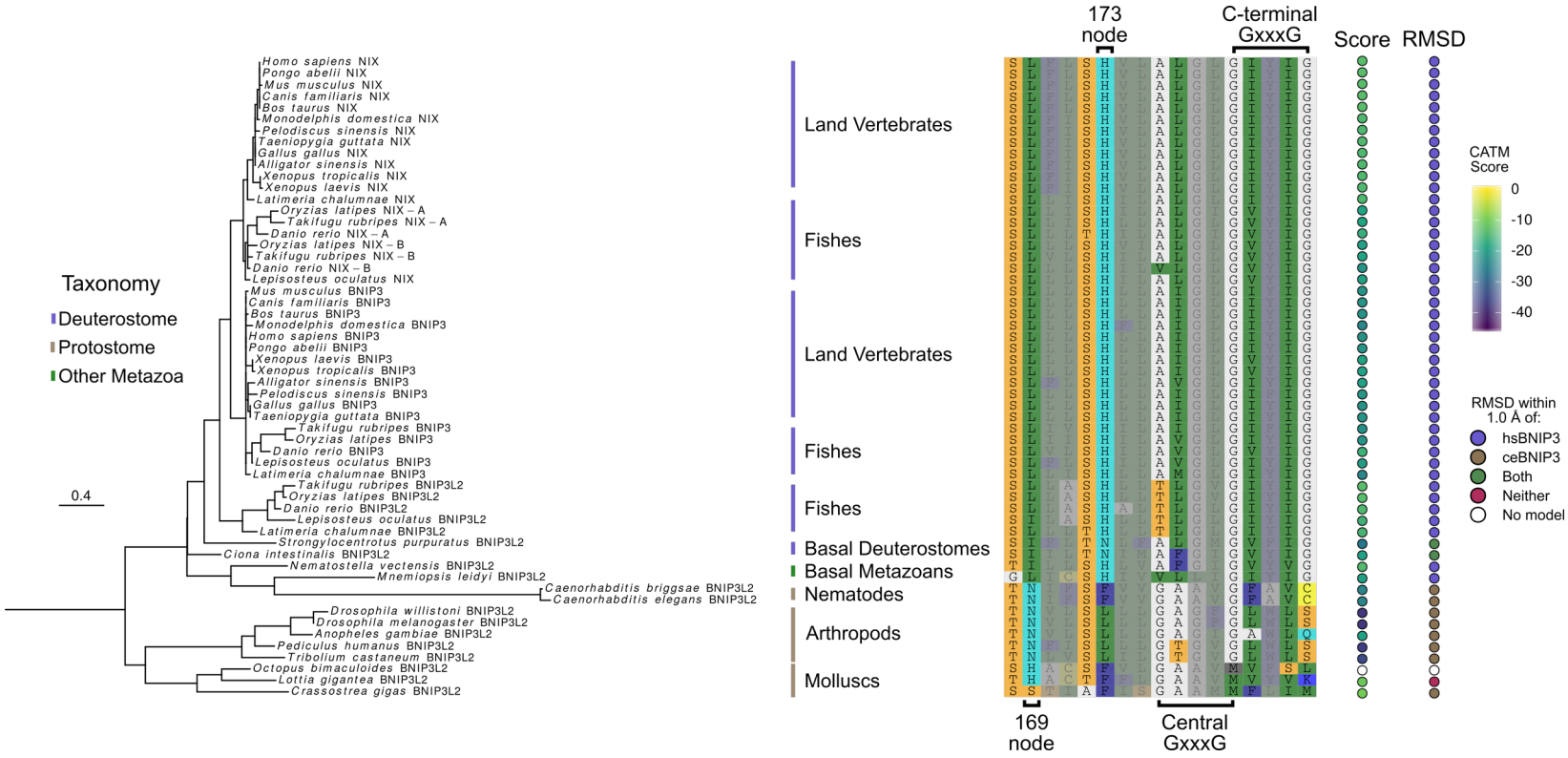

Figure 4.2: BNIP3 dimerization is highly conserved. Left: phylogenetic tree of BNIP3 homologues (Wu et al., 2017). BNIP3 is broadly conserved in animals. Center: Multiple sequence alignment of BNIP3 transmembrane domains. Noninterfacial residues are faded. Locations of the hydrogen bonding node and GxxxG motif are highlighted for deuterostomal (top) and protostomal (bottom) homologues. These features are almost completely conserved, but their positions differ. Right: Energy scores and structural properties of BNIP3 dimers predicted by CATM. All but one of the sequences produced favorable dimer models.

CATM predicts that the overall structure of the BNIP3 TMD homodimer is highly conserved, except for the change in the helix-helix crossing point. How such a shift was mediated through evolution is unclear, since the dimeric interface of BNIP3 is highly sensitive to mutations. Was BNIP3 dimerization lost and subsequently regained across phyla, or was there a path that preserved BNIP3 association throughout evolution? In addition, how where the critical self-association features shifted? To answer these questions we turned to ancestral sequence reconstruction (Ashkenazy et al., 2012), a technique by which the ancestor of a protein family is estimated from the phylogenetic tree.

We estimated the ancestors of all nodes in the tree using a maximum-likelihood method (Ashkenazy et al., 2012) (Figure 4.3a). To understand how sequence changes in the BNIP3 ancestors led to structural changes, we predicted their dimeric structures and propensities using CATM. As expected, the CATM algorithm predicted that the ancestral BNIP3 sequences form GAS_{right} homodimers with very similar structures as the extant sequences, suggesting that dimerization was preserved throughout its evolutionary history. The only exceptions were some of the inferred sequences corresponding to ancestral mollusc BNIP3 sequences, which did not return favorable dimeric models. As mentioned, these BNIP3 variants do not contain a GxxxG motif, instead having a SmxxxGxxxM sequence, and those sequences that did produce favorable models had much higher energy scores compared to BNIP3 sequences with a GxxxG motif. It is possible that mollusc BNIP3 TMDs do not dimerize to the same extent as others. However, since CATM only models sequences in GAS_{right} conformations, it is also possible that mollusc BNIP3

dimerizes in a way that CATM does not sample. Experimental characterization of these sequences will be critical to determining whether BNIP3 dimerization is conserved in molluscs.

Regardless of whether the mollusc BNIP3 TMDs dimerize, it is clear that BNIP3 dimerization was maintained throughout its evolutionary history rather than it being lost and regained. How was the BNIP3 dimer structure maintained through its sequence? Examining the posterior probabilities at the root node reveals a high probability of a tandem GxxxGxxxG sequence (Figure 4.3b), which can drive self-interaction of TM helices, likely by increasing packing and providing an extended network of Ca-H hydrogen bonds (Mueller et al., 2014). Thus, the shift between GxxxG motifs was likely mediated through a tandem sequence rather than a loss and subsequent regaining of the motif. As the phylogenetic tree was not time-calibrated, it is unclear when this divergence occurred or what the true root ancestral BNIP3 sequence looked like. Nonetheless, this finding demonstrates that BNIP3 was able to maintain dimerization throughout its evolutionary history and the transition *between* a central and a C-terminal GxxxG motif was likely mediated through the addition of a tandem, stabilizing interaction.

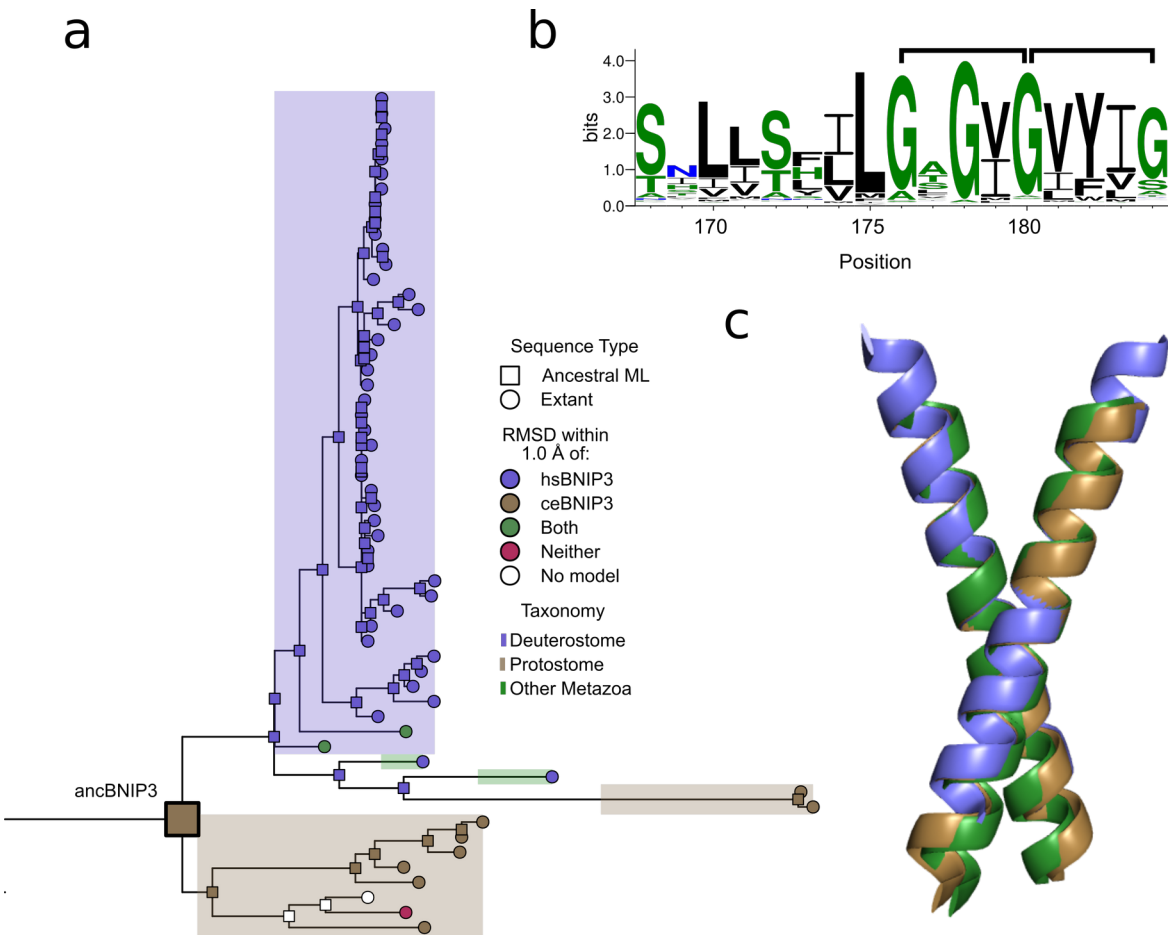

Figure 4.3: Ancestral sequence reconstruction of the BNIP3 TMD. a) Maximum-likelihood sequences from each node of the phylogenetic tree were analyzed with CATM to predict their dimer structures. The RMSD-based classification of each sequence is represented by coloring the node. b) Sequence logo from the root ancestor corresponding to the marginal probabilities of each residue at each position. The GxxxG motifs are *highlighted*. c) Alignment of dimer models for hsBNIP3 (slate), ceBNIP3 (brown) and the maximum-likelihood ancestral BNIP3 (green) at their crossing points.

The predicted structure of the root node ancestral sequence is much closer to the ceBNIP3 model than hsBNIP3 (Figure 4.3c), despite the sequence having two GxxxG motifs to potentially cross at. This suggests that a central crossing point is more favorable than a C-terminal one. Indeed, the energy scores for most extant protostomal sequences are generally more favorable than the deuterostome sequences (Figure 4.2, right). This may be explained by the shift in the interhelical crossing point. The crossing point of ceBNIP3 is more embedded in the membrane than hsBNIP3, whose crossing point is closer to the membrane edge.

4.3.3 Horizontal analysis of human and nematode BNIP3 interfaces

A more central crossing point increases the number of contacts between the two helices and buries the hydrogen bonding groups more deeply in the membrane, which should promote association. Indeed, shifting the location of a GxxxG motif in model TM helices farther into the membrane has been shown to stabilize self-association (Johnson et al., 2006). To further investigate the sequence landscape of BNIP3 dimerization, particularly in regards to the effect of crossing point on stability, we carried out a horizontal shuffling analysis between hsBNIP3 and ceBNIP3. Alignment of hsBNIP3 with ceBNIP3 indicates that there are 7 sequence differences at the interface (Figure 4.1a, gray boxes). These sequence differences somehow mediate a shift in the interhelical crossing point while still maintaining strong dimerization, but the specific determinants are unknown. To better understand the sequence rules governing the mode of BNIP3 association, we generated a library of shuffled interfacial sequences between hsBNIP3 and ceBNIP3. Combinatorial shuffling of these residues generates 256 possible sequences of the form (S/T)(L/N)xxS(H/F)xx(A/G)(I/A)xxG(I/F)x(I/V)(G/C). To account for the effect that noninterfacial differences might play between the two sequences, these interfacial combinations were grafted onto either the hsBNIP3 or the ceBNIP3 background sequence for a total of 512

combinations. These sequences were analyzed in CATM to predict their GAS_{right} structures and dimerization propensities. The vast majority of the shuffled interfaces produced models that did not differ significantly in terms of energy scores or RMSD between the two sequence backgrounds (Figure S4:2), suggesting that the chosen interfacial residues or a subset thereof are the sole determinants of BNIP3 dimerization.

4.3.4 The location of the GxxxG motif has a large impact on the CATM energy score and dimeric structure

The energy score in CATM is composed of three terms: a van der Waals term borrowed from CHARMM (Brooks et al., 2009), a hydrogen bonding term from SCWRL4 (Krivov et al., 2009) that has been parameterized to include Cα -H hydrogen bonds, and an implicit solvation term (Lazaridis, 2003). These terms alone are sufficient to predict the structure of known GAS_{right} dimers, including BNIP3, to near-atomic accuracy (Mueller et al., 2014). The energy scores also correlate well with observed TOXCAT signals for model systems (Anderson et al., 2017). Inspired by these results, we examined differences in the CATM energy scores between the hs-ceBNIP3 swaps. All of the sequences returned dimer models, though the energy scores varied wildly. We found that the clearest correspondence between energy score and sequence differences primarily involved the presence and location of the GxxxG motif (Figure 4.4a). Sequences with a tandem GxxxGxxxG motif produced models with the lowest energy scores. Sequences with either a C-terminal GxxxG motif or no GxxxG motif were scored similarly to one another and much less favorably, while those sequences with a central GxxxG motif returned models with energy scores between the other groups. Within these shuffled combinations, it is important to note that the sequences with no GxxxG motif still contain small interfacial residues,

which are amenable to forming GAS_{right} interfaces (Mueller et al., 2014; Walters and DeGrado, 2006).

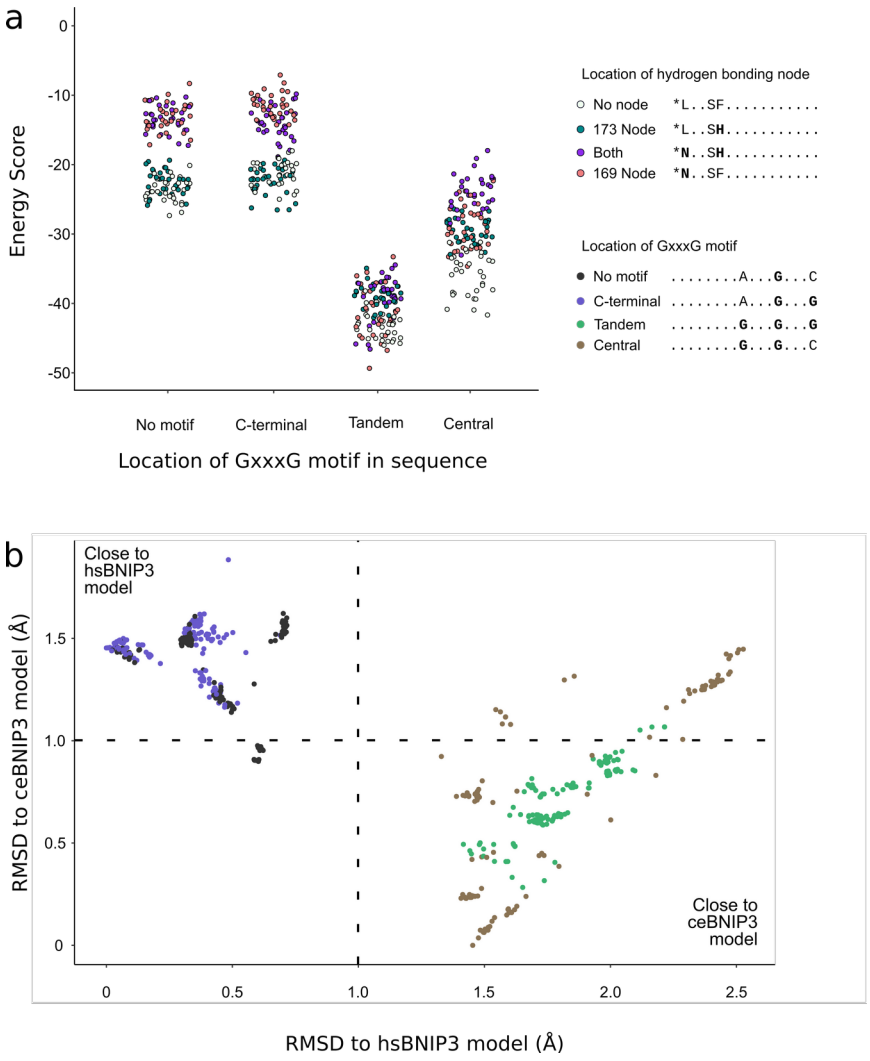

Figure 4.4: Energies and structures of horizontal interface combinations between hsBNIP3 and ceBNIP3. a) CATM energy scores of top dimer models for each sequence classified by the location of the GxxxG motif (X-axis) and the location of the hydrogen bonding node (color). b) RMSD of each interface combination to the reference hsBNIP3 dimer model and ceBNIP3 dimer model. Colors correspond to the location of the GxxxG motif within the sequence.

In contrast, grouping the sequences solely by the location and presence of a hydrogen bonding node led to less segregation and even some paradoxical effects. For example, within the set of sequences with a C-terminal GxxxG motif, sequences with no hydrogen bonding node (white) had roughly equivalent energy scores as those with the histidine motif (teal). However, these scores were much more favorable than sequences with a C-terminal GxxxG motif and an asparagine residue (magenta and purple), which formed very weak dimers. The introduction of an asparagine so far from the crossing point may be destabilizing because the asparagines are too far apart to interact with the opposite helix. Furthermore, within the group of sequences with a central GxxxG motif, those with no histidine or asparagine at all had the best energy scores. These observations contrast heavily with the TOXCAT mutagenesis data, which showed that disruption of the hydrogen bonding node dramatically reduced levels of association for both hsBNIP3 and ceBNIP3 (Figure 4.1e). These findings suggest that CATM may be capturing the packing and Cα-H hydrogen bonding of GAS_{right} dimers but is underestimating the driving force of canonical hydrogen bonds in the membrane. Indeed, it was previously observed that the energy scores in CATM were able to capture TOXCAT association strengths for a set of diverse membrane interfaces, but these interfaces did not contain any polar or charged residues (Anderson et al., 2017).

We next wondered how models of these sequences compared to the ceBNIP3 and hsBNIP3 reference models (Figure 4.4b). Alignments of these shuffled sequences to hsBNIP3 and ceBNIP3 revealed there were primarily two clusters of models, with most models falling within 1Å of either the hsBNIP3 or ceBNIP3 dimer. Classifying the sequences revealed that the presence of a Central GxxxG motif was a key determinant of the resulting structure. Sequences with either a C-terminal or absent GxxxG motif exclusively produced models that aligned quite

closely to the hsBNIP3 dimer, but sequences with a central or tandem GxxxG motif aligned more closely to ceBNIP3 due to a more central helix-helix crossing point. This finding suggests that a central crossing point was more favorable, but it does not exclude the possibility that the sequences with tandem or central GxxxG motifs were forbidden from forming dimers close to hsBNIP3. Since CATM also returns alternate conformations, which it predicts are more stable than the monomeric state, we can identify alternate models for each sequence that align more closely to the ceBNIP3 and hsBNIP3 reference models. We found that sequences with a C-terminal or absent GxxxG motif did not produce any alternate models that aligned within 0.5 Å of the ceBNIP3 reference model, suggesting such sequences are completely restricted to forming models with a C-terminal crossing point. In contrast, sequences with a tandem or central GxxxG motif readily formed alternate conformations with a C-terminal crossing point, albeit with worse energy scores than their best overall conformation. Therefore, CATM predicts that BNIP3 variants with a central crossing point have some conformational plasticity.

4.3.5 Comparison of GxxxG swaps suggest BNIP3 dimers with a central crossing point bury more surface area

We next wondered why dimer models with a central crossing point were generally more favorable than models with a C-terminal crossing point. Shifting the interhelical crossing point into the center of the membrane might stabilize a GAS_{right} interaction, as more helix-helix contacts are possible (Johnson et al., 2006). However, a set of designed GAS_{right} dimers indicated that the interactions can be stabilized through better packing interactions without an increase in the buried surface area (Anderson et al., 2017). We compared the interfaces and CATM energy scores between sequences with C-terminal GxxxG motif and the equivalent sequence with a central GxxxG motif, an example of which is shown in Figure Error: Reference source not

founda. Models with the central GxxxG motif tended to have a larger buried surface area than their equivalent sequence with a C-terminal motif (Figure Error: Reference source not foundb). The energy score of the best model for each of these sequence pairs was also always more favorable for the sequence with the central GxxxG motif. The largest change in the individual energy terms for these swaps was the van der Waals interaction. In contrast, the total hydrogen bonding energy, interhelical hydrogen bonding energy, and solvation terms were more mixed and had an average ΔΔE closer to 0. Thus, the improvement of CATM energy scores for dimers with a central crossing point is largely due to an increase in the buried surface area and more van der Waals interactions.

4.3.6 Comparison with known mutagenesis data

To better understand the overall performance of the CATM algorithm in predicting the dimerization strength of BNIP3 variants, we compared the energy scores of hsBNIP3 variants with a previously published TOXCAT mutagenesis study (Lawrie et al., 2010) (Figure 4.5a, small circles). Though the predictive value of any one mutation is low, there is a clear correlation between disruptive mutations and a decrease in the CATM energy score. Comparing the average TOXCAT sensitivity with the average CATM score for each position led to a much better correlation (Figure 4.5a, large circles). While *in silico* mutagenesis does identify that mutations of the interfacial residues of hsBNIP3 are more destabilizing than mutations of the lipid-facing residues, it predicts that most mutations at the hydrogen bonding node are either neutral or only slightly destabilizing (Figure 4.5b). This result contrasts with TOXCAT results for mutagenesis of the BNIP3 TMD, in which the H173F mutation reduced the signal by approximately 50% and mutations to other hydrophobic residues disrupted it even more. While a 50% decrease of the wild-type BNIP3 signal is still higher than the total signal for the very stable Glycophorin A

homodimer, it underscores the hypothesis above that CATM underestimates the importance of the hydrogen bonding node in BNIP3 dimerization, despite converging on the correct structure. Examining the individual energy scores indicates that a loss of hydrogen bonding energy is compensated by more favorable van der Waals interactions. While this result indicates that CATM's predictive value may falter when examining sequences with canonical hydrogen bonds, it presents a useful opportunity for training and improvement of the CATM energy score.

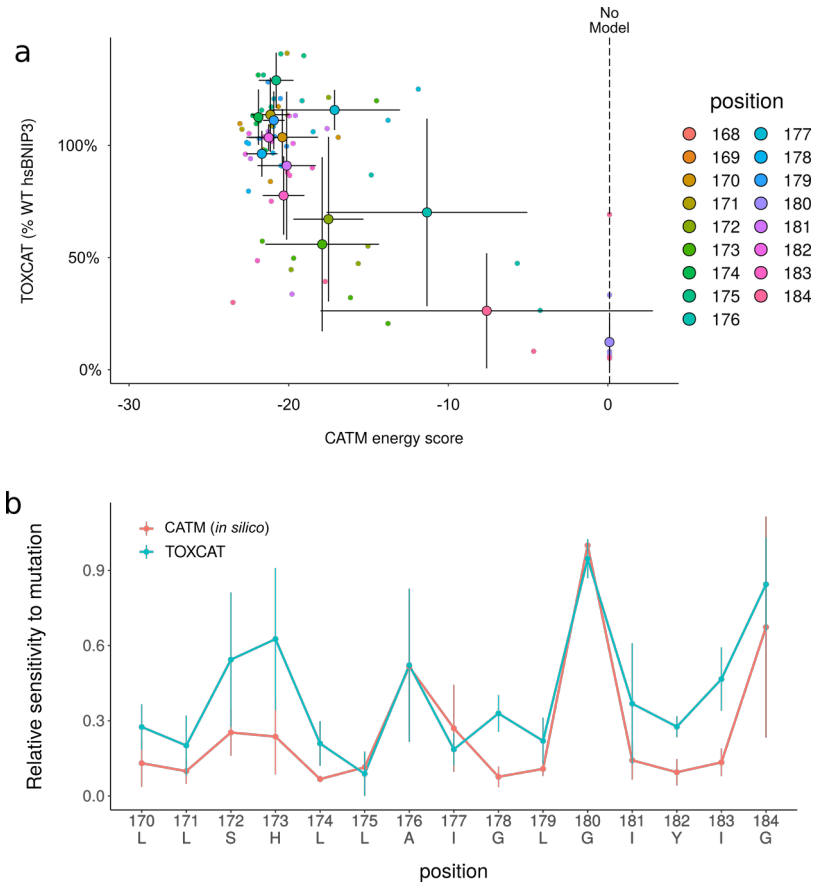

Figure 4.5: Comparison of TOXCAT and *in silico* mutagenesis of hsBNIP3. a) TOXCAT mutagenesis data for hsBNIP3 from (Lawrie et al., 2010) was compared to energetic predictions for the same mutants by CATM. Individual mutations are colored by position and shown as small circles and position-averaged results are showed as large circles. Error bars represent standard deviation for all mutations within a position. b) Relative sensitivities for each position in TOXCAT and *in silico* show a similar pattern of disruption, but S172 and H173 are much more sensitive to mutation experimentally rather than computationally.

4.3.7 Conclusions and future directions

The dimerization of BNIP3 appears to be important for its autophagic activity, and dimerization of BNIP3 has been broadly conserved, though the exact interface has subtly shifted over the course of evolution. This study examined putative ancestors of the BNIP3 TMD and has established that, although extant sequences have markedly different interfaces, dimerization may have been preserved throughout its evolutionary history. Preservation of the self-association despite changes in the interface indicate that dimerization is indeed important for BNIP3's biological role. We have also established that the structure of the BNIP3 TMD likely shifts slightly between protostomal and deuterostomal BNIP3, suggesting that a single transition led to the differences in crossing points observed in models of extant sequences. The protostomal sequences produce dimers with the crossing point shifted more into the center of the membrane. This may strengthen the association, as placing the crossing point further into the membrane increases the potential interhelical contacts as well as buries the GxxxG more deeply in the membrane, which has been shown to increase association in model systems. Whether this increased stability was selected in protostomes or was simply a consequence of its stochastic, path-dependent evolutionary history remains unclear. Unfortunately, dimerization of hsBNIP3 and ceBNIP3 approach saturation of the TOXCAT assay (Lawrie et al., 2010) and so these are not seen experimentally in the wild-type sequences. However, it may be possible to rigorously analyze the dimerization propensities of weakened interfaces with central and C-terminal crossing points.

The horizontal analysis of hsBNIP3 and ceBNIP3 interfacial swaps indicated that the GAS_{right} dimerization of BNIP3 was primarily mediated by the GxxxG motif with the hydrogen bonding node playing a supporting role. Generally, CATM predicts that models with a more central

crossing point (as opposed to the C-terminal crossing point observed in the hsBNIP3 structure and its homologous models) tend to be more stable due to an increase in the buried surface area. The ancestral sequence reconstruction provides an orthogonal measure of the importance of a hydrogen bonding node in BNIP3 self-association. Within these positions, the ASR is relatively ambiguous as to whether positions 173 and 169 contain an aliphatic or polar residue, while the hydroxyl residues at positions 168 and 172 are predicted with high confidence. This is consistent with the observation that the GxxxG motif is the primary determinant of BNIP3 self-association but that the intermolecular hydrogen bonds play a significant supporting role (Lawrie et al., 2010).

The strength of hydrogen bonds within proteins can vary dramatically (Bowie, 2011). Even in the membrane, where the apolar environment should stabilize these interactions, there is significant competition from the protein backbone and water, which makes hydrogen bonds in the membrane overall less stabilizing than might naively be expected. Of course, functionally important hydrogen bonds selected during evolution can be much more favorable. The GxxxG motif in BNIP3 aligns H173 with S172 on the opposing helix such that they are in a favorable position to form hydrogen bonds, which would reduce competing interactions from other donors and acceptors. When this alignment through the GAS_{right} motif is disrupted, the self-association of BNIP3 is dramatically reduced. The entire network of interactions in BNIP3 therefore work together in concert. The requirement of the two helices to be pre-aligned in this way explains why dimerization is nearly abolished when the Sm-xxx-Sm motif is removed in either ceBNIP3 or hsBNIP3 (Figure 4.1e). Without such pre-positioning, the dimeric structure faces too much competition. This evolutionary pressure would place constraints on where the hydrogen bonding node would be relative to the GxxxG motif. The ancestral sequence reconstruction predicted a

tandem GxxxGxxxG motif with little ambiguity while the presence of polar residues was more uncertain. This suggests that the shift in the polar residue followed the shift of the tandem GxxxGxxxG motif.

The precise mechanistic role that BNIP3's strong transmembrane dimerization plays in promoting autophagy or apoptosis remains somewhat unclear. It is possible that the BNIP3 transmembrane domain oligomerizes with such high affinity to reduce its surface area in the crowded membrane, allowing further increases in protein density. In this case, BNIP3 dimerization via its TMD would not be directly involved in autophagy, but it would be involved in promoting the overall integrity of the mitochondrial outer membrane. Alternately, BNIP3 dimerization may be important to interact with the autophagosome. Indeed, BNIP3 variants with monomerizing mutations in the TMD did not interact as strongly with LC3 (a component of the autophagosome) in a co-immumoprecipitation assay (Hanna et al., 2012).

An important unaddressed question is whether and how the BNIP3 oligomeric state is regulated *in vivo*. Depending on the sample source and choice of antibody used for detection, BNIP3 can appear as multiple different bands in Western blots, predominately as a 60kDa dimer or a 30kDa monomer (Kubli et al., 2008). Notably, *in vitro* synthesized BNIP3 migrates predominately as a homodimer (Chen et al., 1997; Ray et al., 2000), suggesting that post-translational modifications lead to the appearance of the other bands. Treating cell fractions with phosphatase led to the appearance of multiple BNIP3 bands that migrated faster than the 30kDa “monomeric” species, one of which was close to the BNIP3 molecular weight of 21kDa (Graham et al., 2007). Conversely, inhibiting phosphatase activity in the cell extracts with okadaic acid eliminated the rapidly migrating species in favor of the 30kDa band. Unfortunately, this study did not report the effect of phosphatase treatment or inhibition on the BNIP3 dimer band, but this

study clearly demonstrated that BNIP3 is phosphorylated *in vivo* and that the “monomer” band corresponds to phosphorylated BNIP3. The juxtamembrane C-terminus of BNIP3 contains multiple known phosphorylation sites, and it is tempting to speculate that positioning multiple negatively charged moieties so closely together in the dimeric state may be destabilizing. However, phosphomimetic mutants in the BNIP3 C-terminus prevented cell death while having no effect on autophagy (Liu and Frazier, 2015). Of course, SDS-PAGE is highly denaturing and the “monomeric” BNIP3 species may still self-associate in the cell. A more thorough examination of BNIP3's oligomeric state *in vivo* and the functional relevance of these different forms must be undertaken to better understand why such strong dimerization might be required in certain conditions and not in others.

Future work will experimentally test the predictions of whether the ancestral variants of BNIP3 and the shuffled sequences dimerize using a sort-seq assay (Anderson, 2019). A library of constructs corresponding to the ancestral sequence reconstruction and the horizontal interface combinations analyzed in this study has been designed and is currently being cloned into the pcc-GFP-Kan plasmid for TOXGREEN analysis (Anderson, 2019; Armstrong and Senes, 2016). Once this library is constructed, it will be transformed into *E. coli* cells, and the dimerization strength of each BNIP3 variant will be assayed via a flow cytometer readout of GFP expression. This large library of BNIP3 variants will help evaluate the predictions made from the protein structure prediction and will lead to further understanding of how transmembrane helices associate.

4.4 Materials and methods

4.4.1 Cloning

BNIP3 homolog TMD constructs were generated by PCR to contain Xba I and Bam HI ends and cloned in-frame into the Nhe I and Bam HI sites of the vector pccKAN (Russ and Engelman, 1999). DNA for ceBNIP3 point mutants was generated by site-directed mutagenesis using the QuickChange mutagenesis kit (Stratagene). All sequences were confirmed using automated dideoxynucleotide sequencing.

4.4.2 TOXCAT

Cultures and controls were performed as previously described (Lawrie et al., 2010; Sulistijo and MacKenzie, 2006), using 3.0 OD420 of malE NT326 cells. All constructs conferred the ability to grow on maltose plates, indicating that proper membrane insertion of the ToxR-TMD-MBP fusion protein occurs (Russ and Engelman, 1999), and all constructs show similar levels of ToxR-TMD-MBP fusion protein by western blot using an anti-MBP antibody. Data for wild-type BNIP3 homologues represent average raw CAT activity scores for three independent cultures. Data for each point mutant represent at least three independent cultures, and CAT activities are scaled (relative to wild-type ceBNIP3 processed in parallel) prior to averaging. Bins were assigned as previously described (Lawrie et al., 2010), based on three-fold steps in the dimer-to-monomer ratio implied by the fractional saturation of the TOXCAT signal (bin boundaries of 90%, 75%, 50%, 25%, and 10% of wild type CAT activity).

4.4.3 Ancestral sequence reconstruction

BNIP3 sequences analyzed in this study originated from a multiple sequence alignment and maximum-likelihood phylogenetic tree generated by (Wu et al., 2017). This tree contains a set of

56 BNIP3 sequences from a diverse set of metazoans including vertebrates, arthropods, nematodes, and some basal metazoans. Ancestral sequences for the TMD of BNIP3 (corresponding to residues 168-184 of hsBNIP3) were inferred using the FastML webserver (Ashkenazy et al., 2012).

4.4.4 Sequence design

BNIP3 TMD variants corresponding to residues 168-184 of human BNIP3 were synthesized as an oligo pool (Twist Biosciences). For each node in the tree, the maximum-likelihood ancestral sequence and the “alt-all” sequence were generated. The maximum-likelihood sequence consists of the most probable residue at each position, independent of the other positions. The “alt-all” sequence accounts for sequence uncertainty during reconstruction by incorporating the second most-likely residue at each position, provided that this character has a posterior probability greater than 0.2 (Eick et al., 2016).

The sequence of *Octopus bimaculoides* BNIP3 used in the phylogenetic tree (NCBI RefSeq ID XP_014784636) contains an ambiguous or unknown residue at position 157 of the protein sequence. A BLAST search of this sequence identified a close homolog in *Octopus vulgaris* (RefSeq ID XP_029656842.1) where this position codes for methionine. Therefore, the unknown residue in Obi_BNIP3 was changed to methionine for this work.

To test the importance of the GxxxG motif, positions 176, 180, and 184 of each extant and inferred ancestral BNIP3 sequence were mutated to isoleucine. The maximum-likelihood sequence and the “alt-all” sequence corresponding to the root BNIP3 node in the phylogenetic tree were further subjected to hydrophobic scannning mutagenesis. Each position was mutated to

alanine, leucine, and isoleucine. If the wild-type residue at this position was one of those three, an additional phenylalanine mutation was generated.

For the horizontal analysis of human and *C. elegans* BNIP3, each possible combination of the 8 positions that differ at the interface between the two sequences were generated both for the human and *C. elegans* backgrounds, for a total of 512 sequences.

NcoI and DpnII restriction sites were added to the N- and C-termini of the transmembrane domains. To facilitate cloning, all BNIP3 TMDs were constructed with an additional Leucine at the C-terminus. The gene chip was amplified using qPCR followed by regular PCR. The product was digested with NcoI and DpnII and ligated into the ToxGREEN vector (Anderson et al., 2017; Armstrong and Senes, 2016).

4.4.5 Structural prediction of BNIP3 homologues and variants

Sequences corresponding to the TMD sequence plus flanking residues of the TOXR construct (GNRAS-BNIP3$_{TMD}$-ILIN) were submitted to CATM version 24 for GAS$_{right}$ structure prediction (Anderson et al., 2017; Mueller et al., 2014). Models generated by CATM were aligned to each other via their backbone atoms using a quaternion strategy (Coutsias et al., 2004; Kulp et al., 2012). The top CATM-generated model for the wild-type *Homo sapiens* BNIP3 was used as the hsBNIP3 reference and the top wild-type *Caenorhabditis elegans* model was used as the ceBNIP3 reference. Since the sequences are all the same length and have identical secondary structure, all positions were used for the alignment. When performing the alignment by the crossing point of each structure, the maximum number of residues spanning the CATM-predicted crossing point shared by the two structures were used. This accounts for the shift of the crossing point from one Sm-xxx-Sm motif to another.

4.4.6 Comparison of CATM energy scores with mutagenesis

Mutations from a previous TOXCAT-based mutagenesis of the hsBNIP3 TMD were compared with energy scores of the same variants in CATM-predicted dimers. To perform the *in silico* mutagenesis, the full CATM protocol was performed on the point mutants.

4.4.7 Data analysis

Data were analyzed using scripts written in R (R Core Team, 2019), using the tidyverse (Wickham, 2017), ape (Paradis and Schliep, 2018), ggraph (Pedersen, 2019), and ggseqlogo (Wagih, 2017) libraries.

4.5 Supplementary Figures

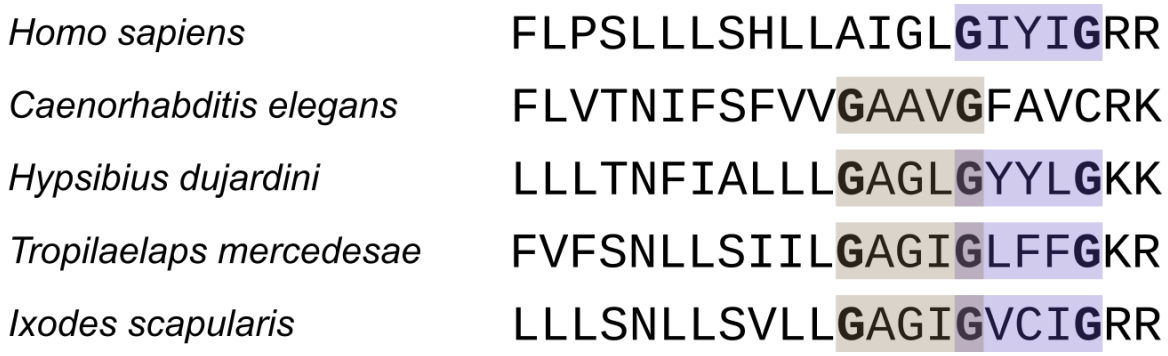

Figure S4:1: Extant BNIP3 homologues with a tandem GxxxGxxxG motif. Alignment of BNIP3 TMDs from *H. sapiens, C. elegans, Hypsibus dujardini* (tardigrade), *Tropilaelaps mercedesae* (mite), and *Ixodes scapularis* (deer tick). GxxxG motifs are highlighted in slate (C-terminal) and brown (Central).

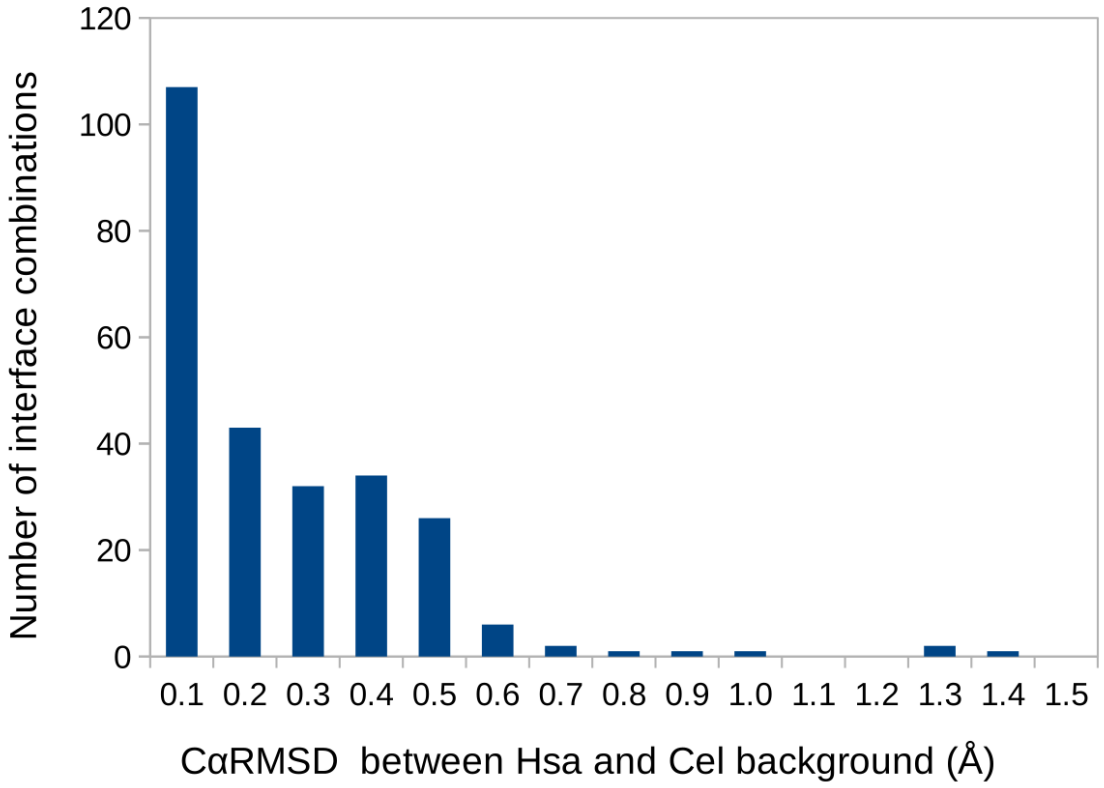

Figure S4:2: The background sequence does not substantially alter the BNIP3 dimer interface structures predicted by CATM. Structural alignments between the top models of each interface combination in the human or nematode noninterfacial BNIP3 background shows very little deviation for the vast majority of combinations.

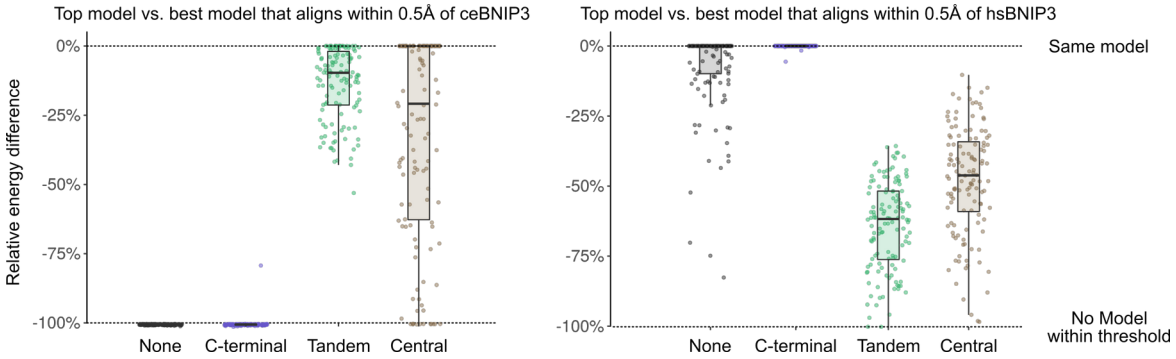

Figure S4:3: Structural plasticity of BNIP3 sequences with a central or tandem GxxxG motif. Energy scores for the top model and the top model that aligned within 0.5A of the ceBNIP3 (left) or hsBNIP3 (right) dimer. Sequences with no GxxxG motif (grey) or a C-terminal motif (slate) did not produce any lower-ranked models that closely aligned to the ceBNIP3 dimer, but sequences with a tandem (green) or central (brown) GxxxG motif produced lower-ranked models that aligned closely to the hsBNIP3 dimer.

4.6 References

Anderson, S.M. (2019). Understanding the GAS-right Motif: Sequence, Structure, and Stability. PhD Thesis.

Anderson, S.M., Mueller, B.K., Lange, E.J., and Senes, A. (2017). Combination of Cα–H Hydrogen Bonds and van der Waals Packing Modulates the Stability of GxxxG-Mediated Dimers in Membranes. J. Am. Chem. Soc. *139*, 15774–15783.

Armstrong, C.R., and Senes, A. (2016). Screening for transmembrane association in divisome proteins using TOXGREEN, a high-throughput variant of the TOXCAT assay. Biochim. Biophys. Acta BBA - Biomembr. *1858*, 2573–2583.

Ashkenazy, H., Penn, O., Doron-Faigenboim, A., Cohen, O., Cannarozzi, G., Zomer, O., and Pupko, T. (2012). FastML: a web server for probabilistic reconstruction of ancestral sequences. Nucleic Acids Res. *40*, W580–W584.

Bocharov, E.V., Pustovalova, Y.E., Pavlov, K.V., Volynsky, P.E., Goncharuk, M.V., Ermolyuk, Y.S., Karpunin, D.V., Schulga, A.A., Kirpichnikov, M.P., Efremov, R.G., et al. (2007). Unique dimeric structure of BNip3 transmembrane domain suggests membrane permeabilization as a cell death trigger. J. Biol. Chem. *282*, 16256–16266.

Bowie, J.U. (2011). Membrane protein folding: how important are hydrogen bonds? Curr. Opin. Struct. Biol. *21*, 42–49.

Brooks, B.R., Brooks, C.L., Mackerell, A.D., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., et al. (2009). CHARMM: the biomolecular simulation program. J. Comput. Chem. *30*, 1545–1614.

Chen, G., Ray, R., Dubik, D., Shi, L., Cizeau, J., Bleackley, R.C., Saxena, S., Gietz, R.D., and Greenberg, A.H. (1997). The E1B 19K/Bcl-2-binding protein Nip3 is a dimeric mitochondrial protein that activates apoptosis. J. Exp. Med. *186*, 1975–1983.

Cizeau, J., Ray, R., Chen, G., Gietz, R.D., and Greenberg, A.H. (2000). The C. elegans orthologue ceBNIP3 interacts with CED-9 and CED-3 but kills through a BH3- and caspase-independent mechanism. Oncogene *19*, 5453–5463.

Coutsias, E.A., Seok, C., and Dill, K.A. (2004). Using quaternions to calculate RMSD. J. Comput. Chem. *25*, 1849–1857.

Eick, G.N., Bridgham, J.T., Anderson, D.P., Harms, M.J., and Thornton, J.W. (2016). Robustness of Reconstructed Ancestral Protein Functions to Statistical Uncertainty. Mol. Biol. Evol. msw223.

Graham, R.M., Thompson, J.W., Wei, J., Bishopric, N.H., and Webster, K.A. (2007). Regulation of Bnip3 Death Pathways by Calcium, Phosphorylation, and Hypoxia–Reoxygenation. Antioxid. Redox Signal. *9*, 1309–1316.

Hanna, R.A., Quinsay, M.N., Orogo, A.M., Giang, K., Rikka, S., and Gustafsson, Å.B. (2012). Microtubule-associated Protein 1 Light Chain 3 (LC3) Interacts with Bnip3 Protein to Selectively Remove Endoplasmic Reticulum and Mitochondria via Autophagy. J. Biol. Chem. *287*, 19094–19104.

Johnson, R.M., Rath, A., and Deber, C.M. (2006). The position of the Gly-xxx-Gly motif in transmembrane segments modulates dimer affinity. Biochem. Cell Biol. Biochim. Biol. Cell. *84*, 1006–1012.

Krivov, G.G., Shapovalov, M.V., and Dunbrack, R.L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. Proteins Struct. Funct. Bioinforma. *77*, 778–795.

Kubli, D.A., Quinsay, M.N., Huang, C., Lee, Y., and Gustafsson, Å.B. (2008). Bnip3 functions as a mitochondrial sensor of oxidative stress during myocardial ischemia and reperfusion. Am. J. Physiol.-Heart Circ. Physiol. *295*, H2025–H2031.

Kulp, D.W., Subramaniam, S., Donald, J.E., Hannigan, B.T., Mueller, B.K., Grigoryan, G., and Senes, A. (2012). Structural informatics, modeling, and design with an open-source Molecular Software Library (MSL). J. Comput. Chem. *33*, 1645–1661.

Lawrie, C.M., Sulistijo, E.S., and MacKenzie, K.R. (2010). Intermonomer Hydrogen Bonds Enhance GxxxG-Driven Dimerization of the BNIP3 Transmembrane Domain: Roles for Sequence Context in Helix–Helix Association in Membranes. J. Mol. Biol. *396*, 924–936.

Lazaridis, T. (2003). Effective energy function for proteins in lipid membranes. Proteins Struct. Funct. Genet. *52*, 176–192.

Liu, K.E., and Frazier, W.A. (2015). Phosphorylation of the BNIP3 C-Terminus Inhibits Mitochondrial Damage and Cell Death without Blocking Autophagy. PLOS ONE *10*, e0129667.

Mueller, B.K., Subramaniam, S., and Senes, A. (2014). A frequent, GxxxG-mediated, transmembrane association motif is optimized for the formation of interhelical C -H hydrogen bonds. Proc. Natl. Acad. Sci.

Paradis, E., and Schliep, K. (2018). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528.

Pedersen, T.L. (2019). ggraph: An Implementation of Grammar of Graphics for Graphs and Networks.

R Core Team (2019). R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing).

Ray, R., Chen, G., Vande Velde, C., Cizeau, J., Park, J.H., Reed, J.C., Gietz, R.D., and Greenberg, A.H. (2000). BNIP3 Heterodimerizes with Bcl-2/Bcl-XL and Induces Cell Death Independent of a Bcl-2 Homology 3 (BH3) Domain at Both Mitochondrial and Nonmitochondrial Sites. J. Biol. Chem. *275*, 1439–1448.

Russ, W.P., and Engelman, D.M. (1999). TOXCAT: a measure of transmembrane helix association in a biological membrane. Proc. Natl. Acad. Sci. U. S. A. *96*, 863–868.

Sulistijo, E.S., and MacKenzie, K.R. (2006). Sequence Dependence of BNIP3 Transmembrane Domain Dimerization Implicates Side-chain Hydrogen Bonding and a Tandem GxxxG Motif in Specific Helix–Helix Interactions. J. Mol. Biol. *364*, 974–990.

Sulistijo, E.S., and MacKenzie, K.R. (2009). Structural Basis for Dimerization of the BNIP3 Transmembrane Domain ⁺. Biochemistry *48*, 5106–5120.

Sulistijo, E.S., Jaszewski, T.M., and MacKenzie, K.R. (2003). Sequence-specific Dimerization of the Transmembrane Domain of the “BH3-only” Protein BNIP3 in Membranes and Detergent. J. Biol. Chem. *278*, 51950–51956.

Taylor, M.S., Ponting, C.P., and Copley, R.R. (2004). Occurrence and consequences of coding sequence insertions and deletions in Mammalian genomes. Genome Res. *14*, 555–566.

Wagih, O. (2017). ggseqlogo: A “ggplot2” Extension for Drawing Publication-Ready Sequence Logos.

Walters, R.F.S., and DeGrado, W.F. (2006). Helix-packing motifs in membrane proteins. Proc. Natl. Acad. Sci. *103*, 13658–13663.

Wickham, H. (2017). tidyverse: Easily Install and Load the “Tidyverse.”

Wu, X., Wu, F.-H., Wu, Q., Zhang, S., Chen, S., and Sima, M. (2017). Phylogenetic and Molecular Evolutionary Analysis of Mitophagy Receptors under Hypoxic Conditions. Front. Physiol. *8*.

Yasuda, M., Theodorakis, P., Subramanian, T., and Chinnadurai, G. (1998). Adenovirus E1B-19K/BCL-2 Interacting Protein BNIP3 Contains a BH3 Domain and a Mitochondrial Targeting Sequence. J. Biol. Chem. *273*, 12415–12421.

Chapter 5: BH3-in-groove dimerization initiates and helix 9 dimerization expands Bax pore assembly in membranes

This chapter was prepared for publication as:

Zhang Z., Subramaniam S., Kale J., Liao C., Huang B., Brahmbhatt H., Condon S.G., Lapolla S.M., Hays F.A., Ding J., He F., Zhang X.C., Li J., Senes A., Andrews D.W., Lin J. 2016. "BH3-in-groove dimerization initiates and helix 9 dimerization expands Bax pore assembly in membranes." *EMBO J.* 35(2):208-36

Statement of contribution

I generated the alternate model of the BH3 α2-α3-α4 homodimer based on the disulfide crosslinking results after activation by tBid. I incorporated the crosslinking data into the energy score using harmonic distance restraints on the separated helices, then rebuilt the loops to form the final rearranged model.

5.1 Abstract

Pro-apoptotic Bax induces mitochondrial outer membrane permeabilization (MOMP) by forming oligomers through a largely undefined process. Using site-specific disulfide crosslinking, compartment-specific chemical labeling, and mutational analysis, we found that activated integral membrane Bax proteins form a BH3-in-groove dimer interface on the MOM surface similar to that observed in crystals. However, after the α5 helix was released into the MOM, the remaining interface with α2, α3, and α4 helices was rearranged. Another dimer interface was formed inside the MOM by two intersected or parallel α9 helices. Combinations of these interfaces generated oligomers in the MOM. Oligomerization was initiated by BH3-in-groove dimerization, without which neither the other dimerizations nor MOMP occurred. In contrast, α9 dimerization occurred downstream and was required for release of large but not small proteins from mitochondria. Moreover, the release of large proteins was facilitated by α9 insertion into the MOM and localization to the pore rim. Therefore, the BH3-in-groove dimerization on the MOM nucleates the assembly of an oligomeric Bax pore that is enlarged by α9 dimerization at the rim.

5.2 Introduction

In most cells, Bax is a monomeric protein in the cytosol. During apoptosis initiation, Bax is activated and targeted to the mitochondrial outer membrane (MOM). The active Bax undergoes a series of conformational changes and eventually forms oligomeric pores in the MOM that release cytochrome c and other mitochondrial intermembrane space proteins to activate caspases and nucleases, thereby dismantling the cells (Borner and Andrews, 2014; Chi et al., 2014; Moldoveanu et al., 2014; Volkmann et al., 2014; Westphal et al., 2014a).

Previous studies revealed a multi-step process that transforms Bax from a dormant soluble protein to an active integral membrane protein. Bax cycles on and off membranes by a process called retrotranslocation (Edlich et al., 2011; Schellenberg et al., 2013). The Bax is activated and bound to mitochondria by BH3-only proteins such as tBid (Lovell et al., 2008) or by previously activated Bax (Tan et al., 2006). The activated Bax partially embeds helices α5, α6, and α9 into the MOM as cysteines positioned in these helices become inaccessible to a membrane-impermeant sulfhydryl-specific modifying agent when Bax inserts into membranes (Annis et al., 2005; Westphal et al., 2014b). Inaccessibility was observed for Bax bound to the mitochondria isolated from etoposide-treated Myc-null cells where Bax did not oligomerize, and from Myc-expressing cells where Bax did oligomerize, suggesting that the helices are inserted into the MOM before oligomerization (Annis et al., 2005). Consistent with this model, residue inaccessibility was also observed before the tBid-induced oligomerization of Bax mutants that were constitutively bound to mitochondria due to a mutation in α9 (Westphal et al., 2014b). The α5 insertion was also supported by an increase of fluorescence of an environment-sensing fluorophore attached to the α5 of Bax after it was activated by membrane-bound tBid (Lovell et al., 2008). Kinetic analysis of the fluorescence changes associated with α5 insertion and Bax oligomerization indicated that the insertion occurred earlier than the oligomerization. However, while the early chemical labeling study concluded that the α5, α6, and α9 are deeply inserted into the lipid bilayer, the later study concluded that the α5 and α6 are shallowly inserted into the cytosolic leaflet of the bilayer with some α5 residues buried in the cytosolic domain of Bax oligomer after Bax is activated by tBid. Therefore, a more rigorous topology survey is required to differentiate the models for the three membrane-embedded helices, and to ascertain the topology of the other regions, particularly the BH3 region and its binding groove, which are

critical to Bax interaction and function (Bleicken et al., 2010; Dewson et al., 2012; Zhang et al., 2010).

The structure of the oligomeric Bax pore was largely unexplored until recently. Our photocrosslinking study revealed two interdependent interfaces in the Bax oligomer formed in detergent micelles (Zhang et al., 2010). A double electron–electron resonance (DEER) study of Bax oligomer formed in detergent micelle and liposomal membrane suggested an antiparallel helical dimer interface formed by α2-α3 region of neighboring Bax molecules in the oligomer (Bleicken et al., 2010). A disulfide-crosslinking study indicated that the antiparallel α2-α3 interface was extended to include α4 which binds to the other side of α2, and this interface together with a parallel α6 interface could generate Bax oligomers in the MOM (Dewson et al., 2012). Another disulfide-crosslinking study detected an α9 interface formed by a Bax mutant constitutively bound to mitochondria (Iyer et al., 2015). Whether wild-type Bax can form this α9 interface was unknown. Moreover, the contribution of all of these interfaces to the oligomeric pore assembly has not been assessed.A crystallographic study revealed structures of three Bax complexes (Czabotar et al., 2013). The first is a domain-swapped dimer in which two Bax polypeptides lacking the C-terminal α9 helix (BaxΔα9) swap their α6-α8 helices, resulting in two globular units, each comprising α1-α5 helices from one monomer (the core domain), plus α6-α8 helices from the other monomer (the latch domain), that are bridged by two extended antiparallel α5-α6 helices. The second is the domain-swapped BaxΔα9 dimer with a BH3 peptide of tBid bound to each globular unit via a hydrophobic groove that is occupied by α9 in full-length Bax monomer (Suzuki et al., 2000). The third was formed by a GFP fusion protein containing Bax α2-α5 helices. A symmetric dimer interface exists in this complex, in which the BH3 region or

α2 of one Bax fragment engages a groove in the other Bax fragment and vice versa, resulting in two reciprocal BH3-in-groove interfaces (Figure 5.1A).

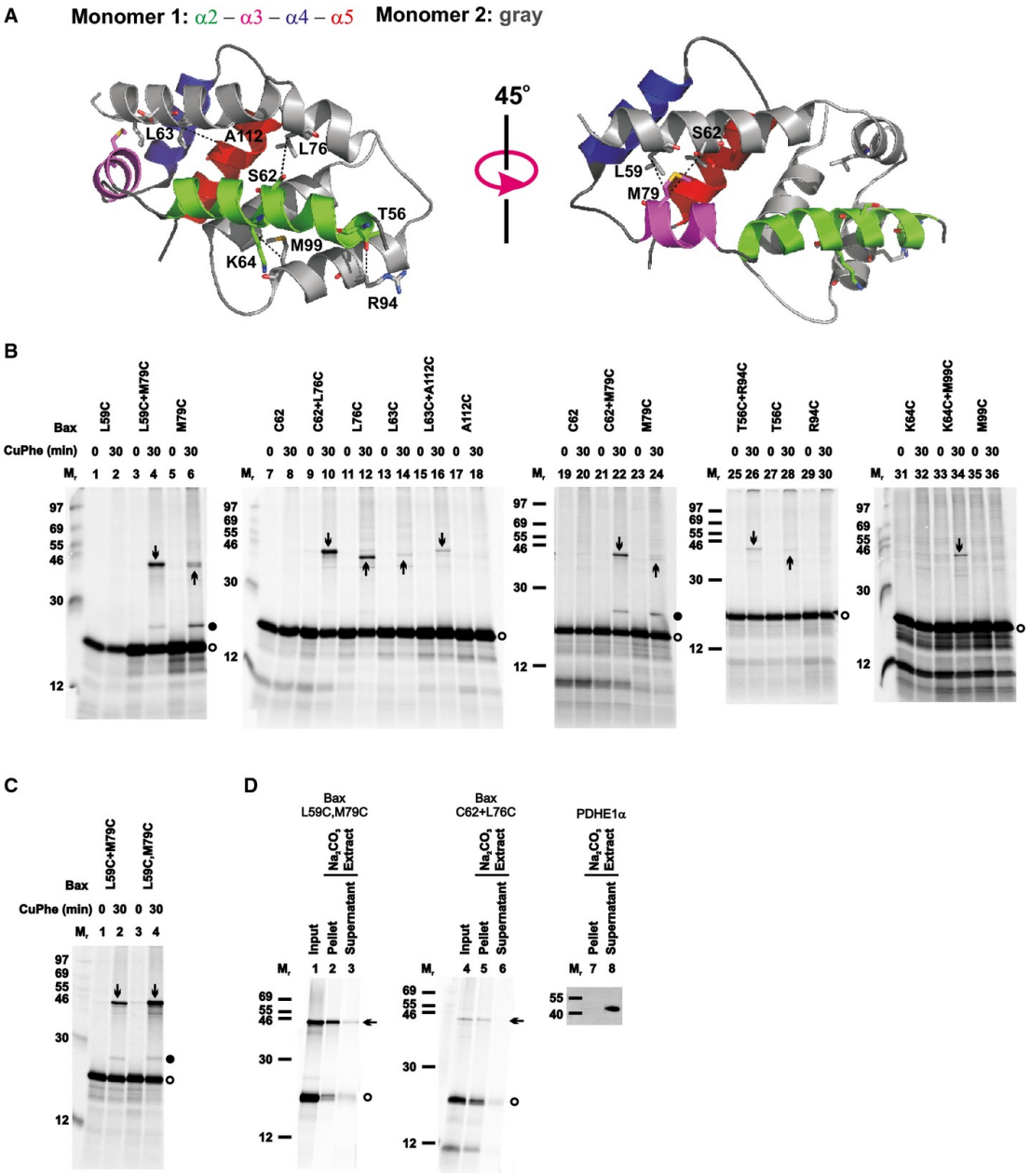

Figure 5.1: Activated Bax proteins integrate into the MOM and dimerize via a BH3-in-groove interface

A) The crystal structure of the BH3-in-groove Bax homodimer (PDB entry 4BDU) is shown with one monomer colored gray and the other colored green, magenta, blue,

and red for its α2, α3, α4, and α5 helices, respectively, as indicated. The residue pairs that were replaced with cysteine pairs in (B) are presented in stick form, and their β-carbon atoms linked by dashed lines with the distances ranging from 5.0 to 6.0 Å.

B) The *in vitro* synthesized [^{35}S]Met-labeled single-cysteine Bax proteins were activated and targeted to the mitochondria that were pretreated with NEM to block the sulfhydryls of mitochondrial proteins. The resulting mitochondria were isolated and oxidized by CuPhe for 30 min. NEM and EDTA were then added to stop the oxidation. For the "0 min" controls, NEM and EDTA were added before the addition of CuPhe. The resulting samples were analyzed by phosphorimaging after non-reducing or reducing SDS–PAGE (see Figure S5:2A).

C) Oxidized mitochondria with the radioactive single-cysteine Bax protein pair or double-cysteine Bax protein were prepared and analyzed as in (B).

D) Oxidized mitochondria with the radioactive single-cysteine Bax protein pair or double-cysteine Bax protein were prepared as in (B). After an aliquot was withdrawn as input, another aliquot was extracted by Na_2CO_3 (pH 11.5) and centrifuged through a sucrose cushion to separate the integral proteins in the membrane pellet from the soluble and peripheral proteins in the supernatant. The input, pellet, and supernatant were analyzed by non-reducing SDS–PAGE and phosphorimaging. In a parallel control experiment, the pellet and supernatant were

analyzed by reducing SDS–PAGE and immunoblotting with an antibody specific to PDHE1α, a soluble mitochondrial matrix protein.

Based on these crystal structures, the following model was proposed for Bax activation and oligomerization (Czabotar et al., 2013). Binding of a BH3-only protein to the hydrophobic groove of Bax after α9 is released from the groove and inserted into the MOM triggers the release of the latch domain from the core domain and the exposure of the BH3 region. The exposed BH3 region of one Bax replaces the BH3-only protein from the groove of other Bax and vice versa, resulting in a Bax homodimer with the BH3-in-groove interface that nucleates the oligomerization process. In addition, a hydrophobic patch consisting of aromatic residues from α4 to α5 that is located on one side of the BH3-in-groove dimer engages the MOM to promote MOMP (Figure 5.2A, top-left). However, these inferences were made based on the structure of a domain-swapped dimer that is acknowledged to be an off mechanism dead-end complex and the structures of Bax deletion mutants, some of them fused with GFP and all of them formed without membranes. Although the core–latch separation and the BH3-in-groove dimerization were confirmed recently by a DEER study that measured the intra- and intermolecular distances of tBid-activated liposome-bound Bax molecules with spin-labeled cysteines (Bleicken et al., 2014), the broad distance distributions implied that the Bax oligomer structure might be dynamic and flexible. Moreover, only three intramolecular distances were obtained from mitochondrion-bound Bax. Thus, it was uncertain whether the domain separation and dimerization actually occurred for the Bax at mitochondria.

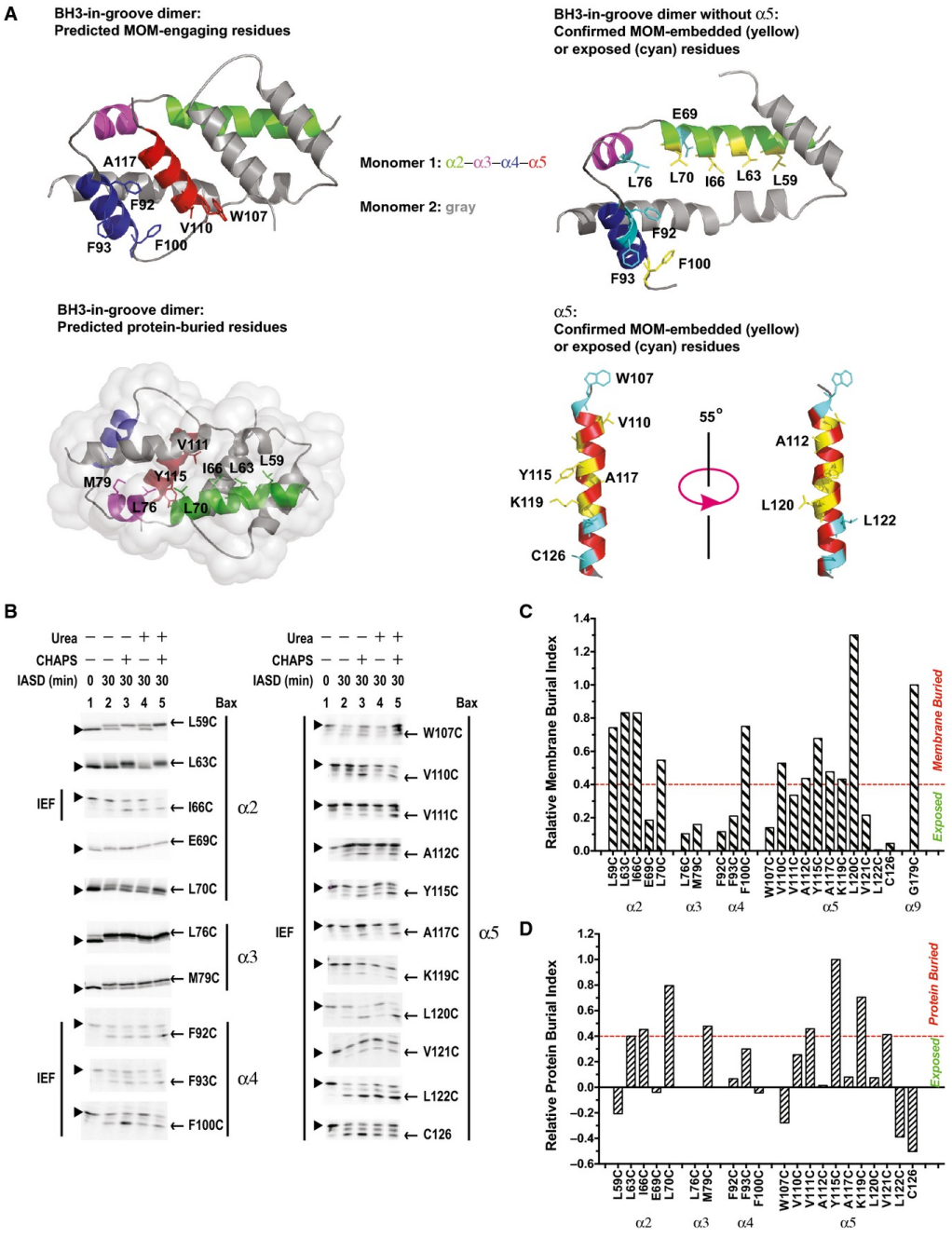

Figure 5.2: The BH3 region and the groove are partially embedded in the MOM

A) Top-left panel, structure of the BH3-in-groove dimer (PDB entry 4BDU) with the predicted MOM-engaging residues presented in stick form. Bottom-left panel,

structure of the BH3-in-groove dimer with the predicted protein-buried residues presented in stick form. Top-right panel, structure of the BH3-in-groove dimer without helix α5 (modified from PDB entry 4BDU), and bottom-right panel, structure of α5 (extracted from the NMR structure of Bax monomer, PDB entry 1F16). In both structures in the right panels, the residues that are buried in the MOM or exposed to the aqueous milieu as the IASD-labeling data in (B–D) confirmed are presented in yellow or cyan stick form, respectively. In all panels, the color codes for the two monomers are the same as that in Figure 5.1A, and indicated.

B) The *in vitro* synthesized radioactive Bax proteins, each with a single cysteine positioned in helix α2, α3, α4, or α5, were activated and targeted to the mitochondria. The resulting mitochondria were isolated and treated with IASD in the absence or presence of CHAPS, urea, or both. After 30 min, the labeling reactions were stopped by mercaptoethanol. For the “0 min” controls, the samples were pretreated with mercaptoethanol before addition of IASD. The IASD-labeled radioactive Bax proteins were resolved from the unlabeled ones using either isoelectric focusing (IEF; as indicated) or gradient SDS–PAGE and detected by phosphorimaging. Triangles and arrows indicate the unlabeled and IASD-labeled Bax proteins, respectively. $n = 3$ for V111C, and K119C; 4 for W107C, and A117C; 2 for other mutants.

C) The phosphorimaging data for IASD labeling of Bax mutants in (B) and the similar data from the independent replicates were quantified to derive the membrane or protein burial indices as described and shown in Figure S5:3. The membrane burial indices were normalized by that of G179C in α9, a reference for the membrane-buried residues. The residues with the relative membrane burial index ≥40% of that of Gly^{179} are considered as buried in the MOM.

D) Similarly, the relative protein burial indices were obtained by using Y115C in α5 as a reference for the protein-buried residues. The residues with the relative protein burial index ≥40% of that of Tyr^{115} are considered as buried in the protein or its complex.

Förster resonance energy transfer (FRET) was used to measure intramolecular distances within a Bax molecule and intermolecular distances between two Bax molecules in the cytosol or the mitochondria of live and apoptotic cells (Gahl et al., 2014). The intramolecular distance from the donor dye-labeled α2 to the acceptor dye-labeled α7 was increased after Bax translocation from the cytosol to the mitochondria, consistent with the core–latch separation model. However, a significant decrease in FRET efficiency between the donor at α2 and the acceptor at α5 suggested a large conformational change within the core that separates the two helices further apart, inconsistent with the subtle distance changes between the two sites during a transition from the monomer NMR structure to the domain-swapped and the BH3-in-groove dimer crystal structures (Czabotar et al., 2013; Suzuki et al., 2000). In addition, the FRET distance between α9 and α2 indicated that α9 was distant from the canonical groove in the cytosolic Bax, thereby

exposing the groove and BH3 region for potential homo- and hetero-interactions. After migrating to the mitochondria, while the intramolecular distance between α9 and α2 remained large and in accordance with the proposed core–latch separation, the intermolecular distance measured by the FRET between a donor-labeled α9 in one Bax molecule and an acceptor-labeled α9 in other Bax molecule was comparable with a homodimerization between the two α9 helices, in line with a model proposed in the DEER and crosslinking studies (Bleicken et al., 2014; Iyer et al., 2015). However, the DEER study proposed an antiparallel α9 dimer model, different from the parallel dimer model proposed by the other studies. Furthermore, intermolecular FRET measurements supported a Bax oligomer model in which in addition to the α9 dimer interface, α2 and α3 form the other dimer interface that unlike the crystallographic BH3-in-groove interface does not involve α5.

To further refine the mechanism of Bax activation and oligomerization, several important questions must be addressed. Does reciprocal binding of the BH3 region of one Bax to the groove of other Bax indeed occur at the MOM resulting in a BH3-in-groove dimer interface as revealed by crystallography? Does formation of the BH3-in-groove interface nucleate Bax oligomerization? For an oligomer to form, there must be additional dimer interfaces, but what regions of Bax are involved? In particular, does α9 form an additional interface, which together with the BH3-in-groove interface mediates Bax oligomerization? Are these dimer interfaces located above, on or in the MOM? If the BH3-in-groove dimer interface is located on the membrane surface with the hydrophobic patch engaging the membrane as the crystallography study proposed and the DEER and one chemical labeling study concluded (Bleicken et al., 2014; Czabotar et al., 2013; Westphal et al., 2014b), could that drive the hydrophobic patch, particularly the hydrophobic α5 helix, more deeply into the membrane as the other chemical

labeling study suggested (Annis et al., 2005)? If the α5 inserts into the membrane and thereby separates from the rest of the BH3-in-groove dimer as the FRET study indicated (Gahl et al., 2014), would the remnant interface rearrange to another conformation that is more stable? In other words, does the BH3-in-groove dimer represent a transient intermediate state? Most importantly, how do Bax oligomers form pores in the MOM? And how flexible is the pore structure? Accumulating evidence suggests that the pores are proteolipidic with some regions of Bax embedded into one leaflet of the bilayer to increase the membrane tension to a point that the bilayer would fuse to a highly curved monolayer, resulting in a toroidal pore with polar and charged lipid head groups lining the rim (García-Sáez, 2012). Other regions of Bax might localize to the rim to decrease the line tension, thereby stabilizing the lipidic pore. However, what are the Bax regions that induce, and that stabilize the lipidic pore? Finally, the size of Bax pore is tunable (Bleicken et al., 2013), but does higher order oligomerization expand the pore?

To answer these questions, we used disulfide crosslinking to map the dimer interfaces in the active mitochondrial Bax oligomer, and compartment-specific chemical labeling to determine the membrane topology. We built structural models and conducted molecular dynamics (MD) simulations to fit the experimental data, and generated mutations to test the functional relevance of the models. Based on these data and models, we propose a molecular scheme for how active Bax proteins are assembled into oligomers to induce and expand lipidic pores in the MOM.

5.3 Results

5.3.1 Bax mutants for interface and topology mapping are functional

Bax mutants with single, double, or triple cysteines located at specific positions were generated from a Bax cysteine-null mutant (Figure EV5.1A). The mutant proteins were

synthesized by using a coupled *in vitro* transcription and translation (TNT) system, and their tBid-dependent MOMP activity was measured in an *in vitro* cytochrome c release assay (Ding et al., 2014) using Bax and Bak double deficient mitochondria (Bax−/−/Bak−/− mitochondria) (Billen et al., 2008). Addition of both tBid protein and the wild-type (WT) Bax protein-producing TNT mixture resulted in a synergistic increase in cytochrome c released above the protein-independent background release (Figure EV5.1B, compare the open bars from "+Bax WT, +tBid" and "+Vector" samples). The releases above the background, indicated as the "corrected" releases (hatched bars) calculated from the "raw" releases (open bars) as described in the legend, showed the protein-dependent releases of cytochrome c by tBid (~20%), Bax (~0%), and both (~50%). With the background release common to these samples subtracted, the "corrected" data in Figure EV5.1C demonstrated that the cytochrome c release due to Bax alone is ~10% or lower for the wild-type Bax and all the mutants, except for L76C and V110C. These two mutants are "autoactive" as they released ~30–60% of cytochrome c in the absence of tBid. As expected, addition of tBid to the "non-autoactive" Bax mutants increased the release to ~30–60%, comparable to the wild-type Bax. Even though some of the mitochondria are somewhat leaky in our *in vitro* lysate-based system (Figure EV5.1B, "Mito-only" sample), the intact mitochondria still respond to the tBid and Bax proteins appropriately.

To relate this *in vitro* MOMP activity to apoptotic activity in live cells, we expressed the two single- and two double-cysteine mutants that were most frequently used in this study transiently as Venus fusion proteins in *bax/bak* double-knockout baby mouse kidney (*bax/bak* DKO BMK) cells (Figure EV5.2). We compared their intracellular location and apoptotic activity before and after staurosporine (STS) treatment to that of wild-type Bax and the cysteine-null mutant. Average Venus fluorescence per cell was measured and correlated to the protein expression. All

mutant constructs were expressed at similar levels compared to Venus-WT Bax (Figure EV5.2A). Expression of Venus-WT Bax increased apoptosis compared to the Venus-only control, and the STS treatment further increased apoptosis (Figure EV5.2B). Consistent with the results from the *in vitro* MOMP assay, the cysteine-null (C0), the single-cysteine (A178C and A183C), and the double-cysteine (L59C,M79C and L59C,L76C) mutants significantly elevated apoptosis in response to STS like the wild-type protein. In addition, the intracellular localizations of these mutants and the Venus-WT Bax are similar, mostly in the cytoplasm but partially at the mitochondria in the untreated cells (Figure EV5.2C and D). The only exception is Venus-Bax L59C,L76C, which is mostly localized to the mitochondria. As expected, these intracellular localization data are in line with the mitochondrial binding data obtained *in vitro* (Figure S5:1, and below).

In addition, previous studies showed that the following Bax mutants, T56C, E69C, R94C, L122C, C126, I175C, V177C, A178C, G179C, and V180C, were active in *bax/bak* double-knockout mouse embryo fibroblast cells, inducing apoptosis after etoposide treatment like WT Bax (Dewson et al., 2012; Iyer et al., 2015; Westphal et al., 2014b). Thus, the cytochrome c release by these mutants *in vitro* (Figure EV5.1C) is consistent with their apoptotic activity in cells. Therefore, the Bax mutants used here to map the dimer interface and membrane topology are functionally similar to the wild-type protein, ensuring that structural information obtained from them is relevant to the functional MOM-bound Bax.

5.3.2 The BH3-in-groove dimer interface exists in the MOM-bound Bax complex

To determine whether the BH3-in-groove dimer interface observed in crystals exists in the MOM-bound Bax complex, we generated Bax mutants with single cysteines located throughout

the BH3 region and the groove. Some of the cysteines are located in the known dimer interface and thus expected to form a disulfide-linked dimer (Figure 5.1A). Other cysteines are located farther away in the dimer structure and thus not expected to form a disulfide. We synthesized the [35S]Met-labeled Bax mutants in an *in vitro* translation system, activated them with a Bax BH3 peptide, and targeted them to the Bax−/−/Bak−/− mitochondria. The mitochondria-bound proteins were separated from the soluble ones and oxidized with copper(II)(1,10-phenanthroline)3 (CuPhe). The resulting radioactive proteins and the potential disulfide-linked protein complexes were analyzed using non-reducing SDS–PAGE and phosphorimaging. A radioactive product of an apparent molecular mass (Mr) close to that of a Bax homodimer (Figure 5.1B, indicated by downward arrow) was detected with each of the indicated single-cysteine Bax pairs predicted to form a disulfide-linked Bax homodimer according to the BH3-in-groove dimer structure (Figure 5.1A).

The following eight lines of evidence demonstrated that the products indicated by downward arrows in Figure 5.1B are the disulfide-linked dimers of the corresponding single-cysteine Bax pairs after they were activated and integrated into the MOM.

1. These products appeared on the non-reducing gel (Figure 5.1B) and disappeared on the reducing gel (Figure S5:2A) and were not formed by the cysteine-null Bax (C0) (Figure EV5.3A) or when NEM or EDTA was added to block the sulfhydryl and chelate the Cu2+, respectively, before the addition of CuPhe to catalyze the oxidation (Figure 5.1B, CuPhe 0 versus 30 min). Therefore, they are disulfide-linked products generated from the oxidized single-cysteine Bax pairs.

2. Bax L59C, C62, R94C, M99C, or A112C did not form a disulfide-linked homodimer (Figure 5.1B, lane 2, 8, 30, 36, or 18). However, when they were paired with another Bax mutant, M79C, L76C, S56C, K64C, or L63C, respectively, a disulfide-linked dimer was formed (lane 4, 10, 26, 34, or 16, indicated by a downward arrow). Therefore, each of these single-cysteine Bax pairs formed a disulfide-linked heterodimer.

3. Although Bax L63C, L76C, or M79C formed a disulfide-linked homodimer (Figure 5.1B, lane 14, 12, or 24, indicated by upward arrow), a new disulfide-linked dimer of a slightly higher Mr was formed when they were paired with Bax A112C, C62, or C62, respectively (lane 16, 10, or 22, indicated by downward arrow), suggesting that the new dimer is a heterodimer formed by the respective single-cysteine Bax pair.

4. Bax T56C or M79C also formed a disulfide-linked homodimer (Figure 5.1B, lane 28 or 6, indicated by upward arrow), and the disulfide-linked dimer formed when they were paired with Bax R94C or L59C, respectively (lane 26 or 4, indicated by downward arrow), had a similar Mr as the corresponding homodimer. However, the yield of the dimer formed in the presence of the paired mutants was higher than that of the homodimer formed by the single mutant, suggesting that some of the dimers formed in the presence of the paired mutants are the disulfide-linked heterodimers. Consistent with this hypothesis, a disulfide-linked dimer was formed in a reaction containing 35S-labeled Bax M79C and unlabeled but six-histidine (6H)-tagged Bax L59C, and detected by phosphorimaging in the Ni2+-bound fraction (Figure EV5.3B, lane 2, indicated by arrow), demonstrating that the dimer contains both the 35S-Bax and the 6H-Bax mutants. In a parallel reaction containing only the 35S-Bax mutant, the dimer was not detected in the Ni2+-bound fraction (lane 3). Therefore, the dimer detected in the reaction containing

both the 35S-Bax and the 6H-Bax mutants must be the disulfide-linked heterodimer of these two mutants. A similar “epitope tagging” experiment to prove the disulfide linkage between T56C and R94C was unnecessary because this disulfide-linked heterodimer was detected before in mitochondria isolated from apoptotic cells expressing the two mutants with different tags (Dewson et al., 2012).

5. The yield of the disulfide-linked dimer formed by the double-cysteine mutant Bax L59C,M79C was doubled compared to that formed by the two corresponding single-cysteine mutants L59C and M79C when mixed together (Figure 5.1C, lane 4 versus 2, indicated by arrow). Therefore, this mutant was examined in more detail.

6. A disulfide-linked dimer of the double-cysteine mutant was not detected in the absence of the Bax−/−Bak−/− mitochondria, the BH3 peptide, or both (Figure EV5.3C, lanes 1, 2, and 3 versus 4, indicated by arrow). Most of the Bax L59C,M79C monomer and the disulfide-linked dimer remained in the mitochondrial pellet fraction after Na2CO3 (pH 11.5) extraction as expected for integral membrane proteins (Figure 5.1D, lane 2 versus 3, indicated by open circle and arrow, respectively), whereas a soluble mitochondrial matrix protein, pyruvate dehydrogenase E1-alpha subunit (PDHE1α), was extracted to the supernatant fraction (lane 7 versus 8). Similar data were obtained for the single-cysteine Bax C62 and L76C pair after the Na2CO3 extraction (lane 5 versus 6).

7. In all the experiments above, Bax was activated by the Bax BH3 peptide that mostly stayed in the soluble fraction and hence would not interfere with the homodimerization of Bax proteins at the mitochondria via the same BH3-binding groove. To verify the results obtained from the BH3 peptide, we used tBid or caspase-cleaved Bid (cBid), a BH3

protein known to activate Bax in cells. Both tBid and cBid activations induced Bax homodimerization at the mitochondria as detected by the disulfide crosslinking of the L59C,M79C mutant, although the intensity of the homodimer band was less than that induced by the BH3 peptide (Figure EV5.3D, indicated by arrow), a result expected from the competition between Bax homo- and Bax-tBid hetero-dimerization.

8. The disulfide-linked homodimer of Bax L59C,M79C was clearly detected in the mitochondrial pellet fraction only after activation by cBid (Figure EV5.3E, indicated by arrow), consistent with the result from the BH3 peptide-activated Bax mutant (Figure EV5.3C). In the soluble supernatant fraction, a band of the same Mr as the Bax homodimer was detected in the CuPhe-induced disulfide-crosslinking samples (Figure EV5.3E, indicated by open triangle). However, because it was also formed before the CuPhe induction, it was not a disulfide-linked product, just like most of the other bands detected in these supernatant samples. Further, the specificity for where and when the Bax homodimer-specific disulfide linkage can form suggests that the introduction of cysteine into Bax by itself does not enhance the dimerization. Instead, our data indicate that disulfide linkage of these cysteine mutants can be used to probe the dimer formation by active Bax molecules at the mitochondria.

Together, these results demonstrate that Bax dimers, as detected by the disulfide crosslinking, were formed only after the Bax mutants were activated by the BH3 peptide or protein and inserted into the MOM. This result is consistent with reports suggesting that Bax oligomerization occurs after the protein inserts into membranes (Annis et al., 2005; Lovell et al., 2008). Moreover, our data are consistent with a twofold symmetry of the BH3-in-groove structure for

the core domain of Bax, further validating that the integral Bax proteins in the MOM form a symmetric dimer interface similar to that observed by crystallography.

The crystal structure of the Bax core dimer shows that two α5 helices are aligned in antiparallel forming the last part of the BH3-in-groove dimer interface (Czabotar et al., 2013). If the MOM-bound Bax proteins formed a dimer interface exactly like that in the crystal, cysteine substitution of Val111 in one Bax and Tyr115 in other Bax would generate a disulfide-linked dimer after the two Bax mutants were activated and integrated into the MOM. However, such a dimer was not detected (Figure EV5.3F), suggesting that the two α5 helices in the MOM-bound Bax dimer may not interact exactly like those in the crystal. To further probe the proximity of the α5 helices, we used BMH, a chemical crosslinker with a spacer of 13 Å between the two sulfhydryl-reactive maleimides, to crosslink the MOM-bound Bax mutants with single cysteine located in α5. Two W107C or Y115C monomers were crosslinked by BMH forming a homodimer (Figure EV5.3G). Similarly when the single cysteine was positioned in α2 in the I66C or L70C mutant, a BMH-crosslinked homodimer was also detected. Because the distance between the β-carbons of each of these residues in the BH3-in-groove dimer is from 8 to 18 Å (Figure EV5.3H), the cysteine replacements would be located within the range that can be linked by BMH, if one considers the flexibility of the BMH spacer and the cysteine side chain. The BMH crosslinking experiments thus revealed that the two α5 and the two α2 helices are in close proximity in the MOM-bound Bax dimer. Taken together with the disulfide-crosslinking data above, these chemical crosslinking results corroborate the existence of a BH3-in-groove interface in the active mitochondrial Bax complex, which is similar but not identical to the crystal structure of the Bax core dimer.

5.3.3 Helix α5 is embedded partially in the MOM, whereas helices α2, α3, and α4 are located on the surface

A hydrophobic patch was observed on one side of the BH3-in-groove dimer crystal that includes the aromatic residues from both α4 (Phe92, Phe93, and Phe100) and α5 (Trp107) (Figure 5.2A, top-left). If this hydrophobic patch engaged the MOM as proposed in the crystallographic study, it might lead to the integration of these helices into the MOM, particularly the more hydrophobic α5. To determine whether the hydrophobic patch is embedded in the membrane and the rest of the BH3-in-groove dimer structure is located on the surface, we used IASD, a membrane-impermeant sulfhydryl-specific labeling agent (Annis et al., 2005), to label Bax mutants with single cysteine positioned throughout the BH3-in-groove structure after they were activated by the BH3 peptide and targeted to Bax−/−/Bak−/− mitochondria. In this procedure, surface-accessible cysteine residues are labeled by IASD directly; residues buried inside proteins are more labeled when the reactions include urea to loosen protein structures; and residues buried in the membrane-embedded protein domains are most efficiently labeled in the reactions containing both urea and the detergent CHAPS to unfold the protein and solubilize the membrane. The amino acid sequence surrounding the cysteine residue might also influence the labeling efficiency but would not be affected by urea or CHAPS (Annis et al., 2005). Therefore, it is the relative changes in accessibility to IASD resulting from added urea and CHAPS that enable assignment of the environment of individual cysteine residues.

Some of the Bax mutants with a single cysteine in α2, α3, and α4, including F92C and F93C with the aromatic residues in α4 replaced by cysteine, were labeled by IASD in the absence of CHAPS and urea, and the labeling was not increased in the presence of CHAPS, urea, or both (Figure 5.2B). Thus, these residues are surface exposed. Other Bax mutants, such as L59C,

L63C, I66C, and L70C with the cysteine positioned on the hydrophobic side of the amphipathic α2, and F100C with the cysteine replacing the last aromatic residue in α4, were partially labeled by IASD in the absence of CHAPS and urea, and addition of CHAPS, urea, or both resulted in an increase of IASD labeling. To quantify the extent of burial in the MOM, we measured the fraction of IASD labeling for each cysteine in the presence of urea and in the presence of both urea and CHAPS, and then subtracted the former from the latter to get a membrane burial index, as described in Figure S5:3A. The resulting membrane burial indices of all the cysteines shown in Figure S5:3B were then normalized to that of G179C, a residue in α9 that is mostly embedded in the MOM as shown by the data in Figure 5.5 and our previous study (Annis et al., 2005). The resulting relative membrane burial indices shown in Figure 5.2C indicate that at least for a fraction of the mitochondrial Bax proteins, their α2, α3, and α4 are located on the MOM surface with some residues exposed to aqueous milieu as their membrane burial indices are < 40% of that for the membrane-buried reference Gly179, and they are also not buried in the protein (see below). Other residues, particularly the four hydrophobic residues on one side of α2 and the one aromatic residue at the C-terminus of α4, are buried in the MOM as their membrane burial indices are more than 40% of that for Gly179 (Figure 5.2A, top-right).

The seven residues in three consecutive helical turns of α5 were buried in the MOM, as evidenced by their IASD-labeling profile (Figure 5.2B) and relative membrane burial indices (Figure 5.2C). These results indicate that about three-fifths of the five-turn α5 helix is embedded in the MOM for at least a fraction of the mitochondrial Bax proteins (Figure 5.2A, bottom-right), consistent with the conclusion about the α5 topology of mitochondrial Bax from apoptotic cells in our previous study (Annis et al., 2005) and a recent study by others (Westphal et al., 2014b).

The BH3-in-groove crystal structure predicts that many non-polar and a few polar residues, such as the eight residues shown in Figure 5.2A, bottom-left, would be buried in the dimer interface. To examine this, we subtracted the fraction of IASD labeling of each single-cysteine mutant in the presence CHAPS from that in the presence of both urea and CHAPS to derive a protein burial index indicating the extent of burial of the cysteine in the protein complex. The resulting protein burial indices of all the cysteines shown in Figure S5:3C were then normalized to that of Y115C, a residue that is buried in the hydrophobic core of BH3-in-groove dimer crystal structure (PDB entry 4BDU), and that is also the most protein-buried residue in mitochondria-bound Bax according to the protein burial index. The resulting relative protein burial indices shown in Figure 5.2D suggest that six of the eight predicted protein-buried residues, including Leu63, Ile66, and Leu70 in α2, Met79 in α3, and Val111 and Tyr115 in α5, are indeed buried in the protein complex for at least a fraction of the mitochondrial Bax proteins, as their protein burial indices are more than 40% of that for the protein-buried reference residue Tyr115 (Figure 5.2D). Of the other two residues, Leu59 is buried in the MOM, whereas Leu76 is exposed to the aqueous milieu. In addition, Lys119 and Val121 in α5 are buried in the protein complex at mitochondria, although they are not buried in the BH3-in-groove dimer crystal structure, likely due to the unfolding of the last two helical turns in the crystal.

Based on all the data presented so far, we conclude that a fraction of mitochondrial Bax proteins formed a dimer interface similar to that shown in the BH3-in-groove dimer crystal structure. However, in another fraction of mitochondrial Bax, α5 is partially embedded in the MOM, separating from α2, α3, and α4 on the MOM surface. This conclusion is also consistent with the observation that some of the protein-buried residues such as Leu63, Leu66, and Leu70 in α2, and Tyr115 and Lys119 in α5 are also buried in the membrane. Thus, the protein burial of

these residues occurs in one fraction of Bax molecules that is in the BH3-in-groove dimer conformation on the membrane surface, while the membrane burial of the same residues takes place in the rest of the Bax molecules that are more deeply embedded in the membrane.

5.3.4 After the departure of α5, the remnant BH3-in-groove dimer interface is rearranged to another conformation

Insertion of α5 into the MOM would peel it off from the rest of the BH3-in-groove dimer structure remaining on the membrane surface. This conformational change might destabilize the dimer structure and initiate additional conformational changes to rearrange the structure to a more stable conformation. To test this hypothesis, we performed a MD simulation, starting with the BH3-in-groove dimer structure on the surface of a lipid bilayer comprised of MOM characteristic phospholipids (Kuwana et al., 2002), and then moving the two α5 helices into the bilayer as described in Supplementary Methods and Table S5.1. The simulation disclosed a dynamic dimer structure in which the α2-α3-α4 regions remain on the bilayer surface (Figure S5:4A). Several residues that were initially distal in the interface became proximal during the simulation (Figure S5:4B). To determine whether the simulated conformational changes indeed happen in the Bax proteins bound to mitochondria, we made more Bax mutants with single or double cysteines replacing the interfacial residues that either approached one another or remained separated during the simulation. All of these Bax mutants generated disulfide-linked Bax dimers after they were activated and integrated into the Bax−/−/Bak−/− MOM (Figure 5.3A–C and EV4; Figure S5:2B), suggesting that the BH3-in-groove dimer interface remained on the membrane surface was altering such that the respective cysteines were close enough to form a disulfide, as the MD simulation predicted. Note that the C62 and S72C mutants also generated a disulfide-linked dimer even though the two residues never reached a disulfide-

linkable distance during the 135-ns simulation. These results suggest that after binding to membranes, the Bax dimer interface seen in the crystal rearranges, and the scale of the conformational changes is larger than that can be sampled within this short simulation time.

To search for a rearranged BH3-in-groove dimer interface that fits all the crosslinking data in Figures 5.3 and EV5.4, we performed a global computational search for a model that simultaneously optimized for the distance between the crosslinked positions and the chemical geometry of the interfacial residues. In particular, we allowed the three helices α2, α3, and α4 to move freely in space driven by the optimization of distance constraints derived from the crosslinking data. The search produced a model in which helices α2 and α3 are merged into a single helix forming a homodimer in a conformation that resembles an antiparallel coiled coil, whereas α4 folds back after a loop capping the end of the coil (Figure 5.3D). This α2-α3-α4 dimer model has excellent packing and satisfies all imposed experimental constraints, with the β-carbon distances near 5 Å for all the disulfide-linkable residue pairs. A similarly extended α2-α3 single helix was observed before in the structure of a Bax BH3 peptide (residues 54–73 or 54–79) in complex with either a domain-swapped Bax dimer (PDB entry 4BD6) (Czabotar et al., 2013) or a Bcl-XL monomer (PDB entry 3PL7) (Czabotar et al., 2011). Moreover, one side of the dimer can be embedded in the MOM, whereas the other side of the dimer is exposed to the aqueous milieu (Figure 5.3E), consistent with the conclusion drawn from the IASD-labeling data (Figure 5.2).

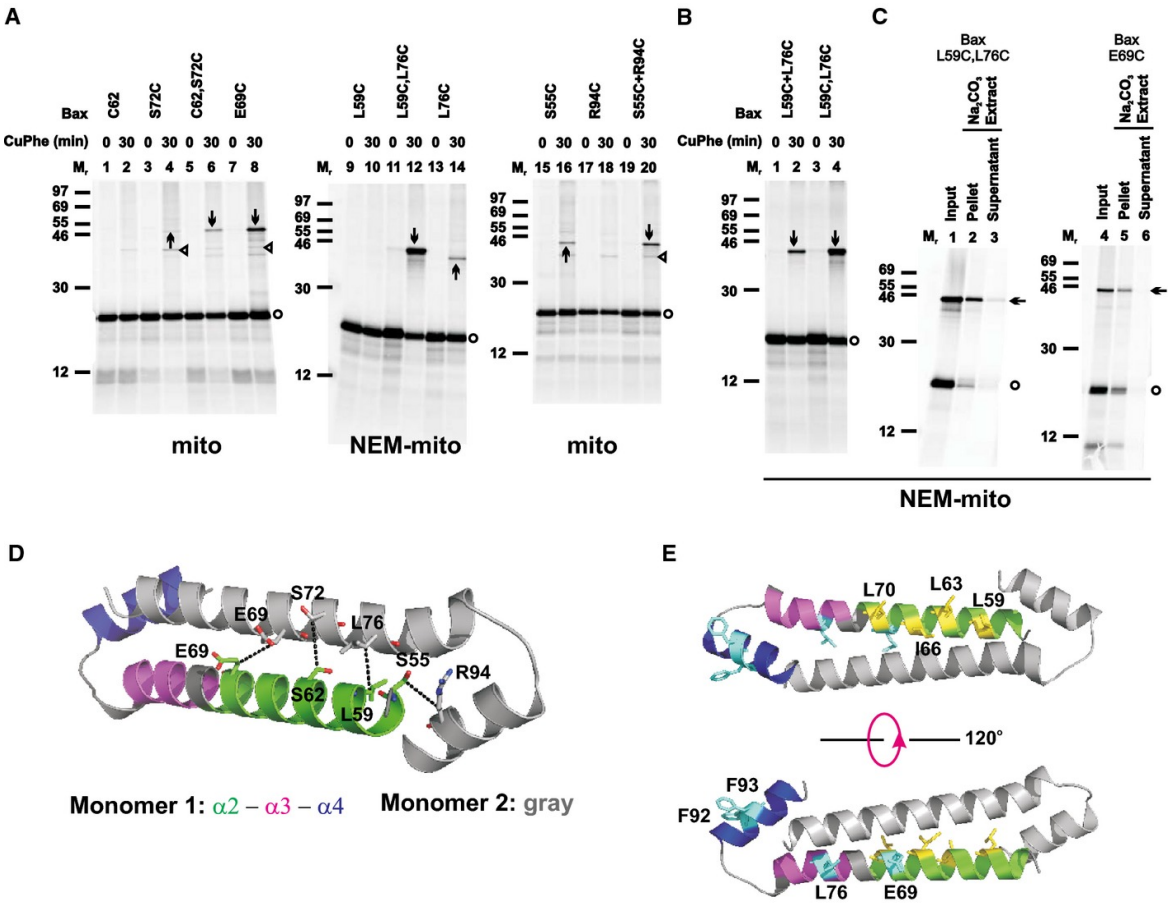

Figure 5.3: The BH3-in-groove dimer interface rearranges to form a helices α2-α3-α4 dimer interface

A) Oxidized mitochondria with the radioactive single- or double-cysteine Bax proteins were prepared and analyzed as in Figure 5.1B.

B) Oxidized mitochondria with the single-cysteine Bax protein pairs were prepared and analyzed as in Figure 5.1B.

C) Oxidized mitochondria with the radioactive double-cysteine Bax protein were prepared, subjected to Na_2CO_3 extraction, and analyzed as in Figure 5.1D.

D) The best structural model for the helices α2-α3-α4 dimer is shown with the residue pairs replaced by cysteine pairs in (B) presented in stick form. The β-carbon atoms of each residue pair are linked by a dashed line with the distance ranging from 4.7 to 5.2 Å.

E) The structural model for the α2-α3-α4 dimer is shown with the residues that are buried in the MOM or exposed to the aqueous milieu as determined in Figure 5.2 presented in yellow or cyan stick form, respectively.

Data information: In (D–E), the color codes for the two monomers are indicated. In (A–C), protein standards, Bax monomers, and disulfide-linked Bax dimers are indicated as in Figure 5.1B. In some targeting reactions, the mitochondria were pretreated with NEM (NEM-mito), while in others the mitochondria were untreated (mito). In (A and B), $n = 2$ for C62,S72C, R94C, and L59C+L76C; 3 for S72C, S55C, and S55C+R94C; 4 for C62; 6 for E69C, L59C, and L76C; 8 for L59C,L76C. In (C), $n = 2$ for each mutant.

The triangle-indicated bands generated from the untreated mitochondria in (A) are disulfide-linked heterodimers between the Bax cysteine mutants and unknown mitochondrial proteins, as they disappeared if the mitochondria were pretreated with NEM to block the sulfhydryls of the mitochondrial proteins before adding the Bax mutants (e.g., compare the products from E69C at the untreated mitochondria in lane 8 of Figure 5.3A to those from the same mutant at the NEM-pretreated mitochondria in lane 4 of Figure 5.3C, or the products from S72C at the untreated mitochondria in lane 4 of Figure 5.3A to those from the same mutant at the NEM-pretreated mitochondria in lane 4 of Figure EV5.4A).

In contrast, the disulfide-linked Bax homodimers were still formed in the NEM-pretreated mitochondria (e.g., the arrow-marked band in lane 4 of Figure 5.3C and lane 4 of Figure EV5.4A).

5.3.5 Formation of the α2-α3-α4 interface depends on prior formation of the BH3-in-groove interface

According to the BH3-in-groove dimer structure, Gly108 in the groove makes van der Waals' contacts with Gly67 and Asp71 in the BH3 region, thereby contributing to the overall stability of the dimer (Figure 5.4A, left). Gly108 to glutamate mutation (G108E) would not only eliminate these favorable interactions, but also generate steric clashes with the Gly67, Leu70, and Asp71, and electrostatic repulsions with the Asp68 and Asp71 (Figure 5.4A, right). In fact, a previous study showed that the G108E mutation inhibited Bax oligomerization and MOMP activity even in the presence of another mutation, Ser184 to valine (S184V), that targeted the G108E mutant to mitochondria (Kim et al., 2009). To determine whether the G108E mutation disrupts the BH3-in-groove interface formed by the mitochondria-bound Bax, we put it together with the S184V mutation into the single- or double-cysteine Bax mutants that could generate disulfide-linked dimers according to the BH3-in-groove dimer structure. The experiments with these mutants showed that the G108E mutation inhibited the disulfide formation along the entire BH3-in-groove interface (Figure 5.4B), suggesting that the mutation disrupts the dimer interface.

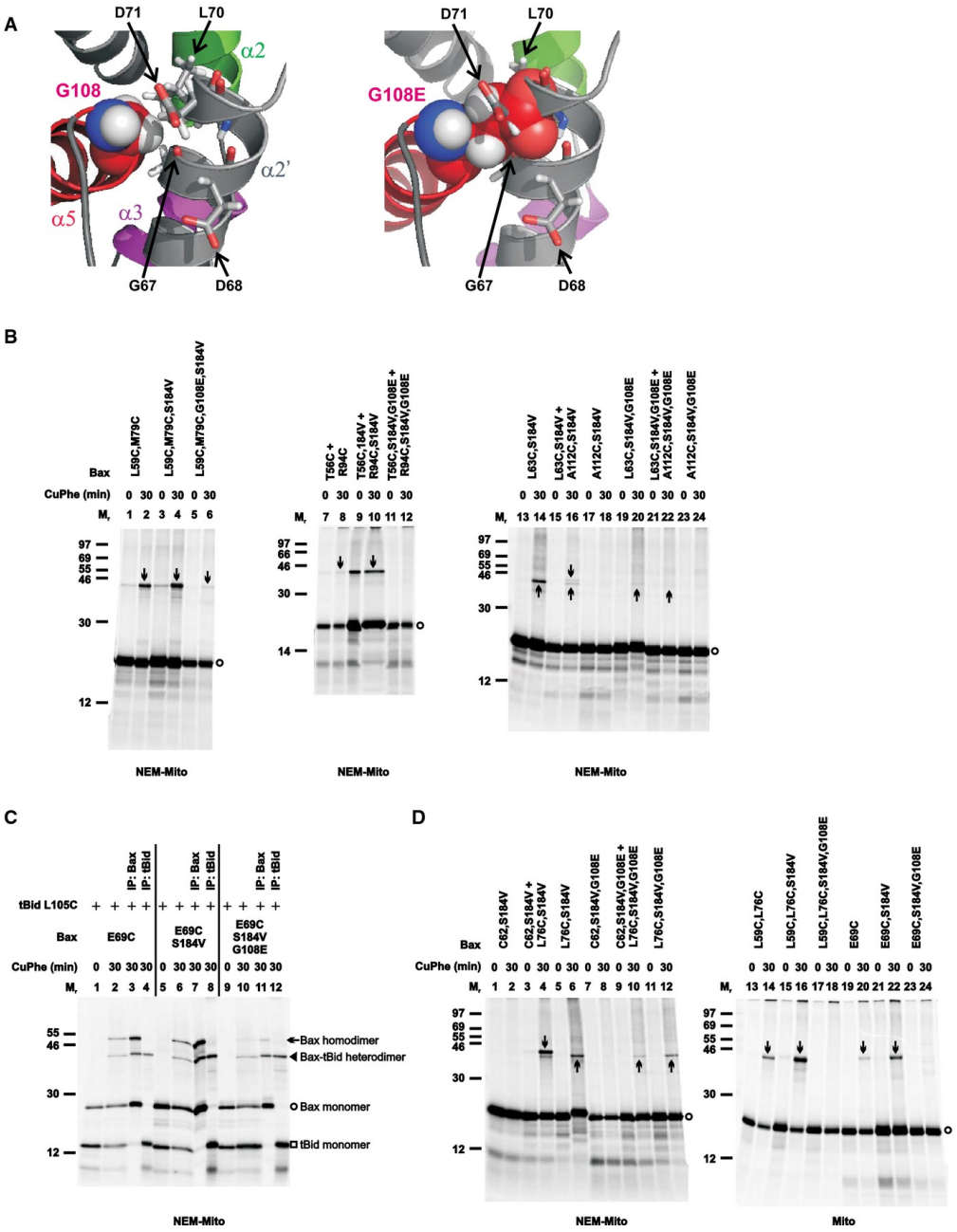

Figure 5.4: G108E mutation in the groove disrupts the BH3-in-groove as well as the helices α2-α3-α4 Bax homodimer interface, but not tBid–Bax heterodimer interface

A) Left panel, part of the BH3-in-groove dimer structure is shown with Gly^{108} from one monomer presented in sphere form, and Gly^{67} and Leu^{70} from the other monomer that have van der Waals' contacts with the Gly^{108} presented in stick form. Asp^{68} and Asp^{71} from the other monomer are also presented in stick form. Right panel, the same structure is shown with the Gly^{108} changed to glutamate (G108E), which results in steric clashes with the Gly^{67}, Leu^{70}, and Asp^{71}, and electrostatic repulsions with the Asp^{68} and Asp^{71}.

B) The radioactive single- or double-cysteine Bax proteins with or without the indicated mutations were activated and targeted to the mitochondria that were either untreated (mito) or pretreated with NEM (NEM-mito), then oxidized and analyzed as in Figure 5.1B. Protein standards, and monomers and disulfide-linked dimers of the Bax proteins are indicated as in Figure 5.1B. $n = 2$ for all mutants.

C) The *in vitro* synthesized [^{35}S]Met-labeled Bax E69C protein with or without the additional mutations (G108E and/or S184V) were activated by the *in vitro* synthesized [^{35}S]Met-labeled tBid L105C protein and targeted to the $Bax^{-/-}$/$Bak^{-/-}$ mitochondria that were pretreated with NEM. The resulting mitochondria were isolated and oxidized by CuPhe for 0 or 30 min. The resulting “0 min” samples (1 equivalent each) and “30 min” samples (1 equivalent each) were analyzed by non-reducing SDS–PAGE and phosphorimaging. The remaining “30 min” samples (4 equivalent each) were immunoprecipitated (IP) by either Bax-

or tBid-specific antibody, and then analyzed by non-reducing SDS–PAGE and phosphorimaging. The identities of the four major products, indicated on the right side of the image, were based on their M_r and recognition by the respective antibody. $n = 2$.

D) The disulfide crosslinking data were obtained from the indicated Bax mutants and presented as in (B). $n = 2$ for all mutants.

The previous study also showed that a Bax with G108E and L63E double mutations did not bind to tBid, but Bax with an L63E single mutation did (Kim et al., 2009). Because the L63E mutation also targeted Bax to mitochondria even in the presence of the G108E mutation, the authors concluded that the G108E mutation in the groove disrupted tBid binding to mitochondrial Bax, a step that would be upstream of Bax oligomerization. Therefore, it is conceivable that the G108E mutation does not directly disrupt the BH3-in-groove Bax homodimerization but does so indirectly by blocking the tBid–Bax heterodimerization that induces Bax homodimerization. To test this hypothesis, we performed disulfide crosslinking of tBid L105C, a mutant with the single cysteine located near the BH3 region, with Bax E69C, a mutant with the single cysteine in the BH3-binding groove. As expected from the Bid BH3 peptide–BaxΔα9 complex structure (PDB entry 4BD2) (Czabotar et al., 2013), and hence a positive control, a disulfide-linked dimer with a Mr expected for the tBid–Bax heterodimer was detected in the mitochondria together with tBid and Bax monomers and a disulfide-linked Bax homodimer (Figure 5.4C, lane 2). The identity of the tBid–Bax heterodimer was confirmed by

reciprocal immunoprecipitation with tBid- or Bax-specific antibodies (lanes 3 and 4). Surprisingly, when the G108E mutation was introduced into the Bax E69C together with the S184V mutation that targets Bax to the mitochondria, the disulfide-linked tBid–Bax heterodimer was still formed (lanes 10–12), just like the other control, Bax E69C with the S184V mutation (lanes 6–8). As expected, the formation of the disulfide-linked Bax homodimer was inhibited by the G108E mutation. These data demonstrate that the G108E mutation in Bax does not disrupt the interaction with tBid, and hence, the inhibitory effect of this mutation on the BH3-in-groove Bax homodimerization is direct. The mutation also does not interfere with the overall folding of Bax, because the mutant Bax retains the capacity to bind tBid. The previous observation that the G108E mutation disrupts Bax-tBid interaction is likely due to an additional impact from the second mutation, L63E.

To determine whether the α2-α3-α4 Bax homodimer can still form even when the BH3-in-groove homodimer is disrupted, we tested the effect of the G108E mutation on the disulfide crosslinking via the α2-α3-α4 interface. The results (Figure 5.4D) suggest that the G108E mutation entirely abolishes the formation of the α2-α3-α4 interface. Because Gly108 is the first residue of α5 that is not part of the α2-α3-α4 dimer (Figure 5.3D), one does not expect the G108E mutation to disrupt this dimer directly. Further, the mutation does not disrupt the tBid–Bax interaction on which the α2-α3-α4 dimerization may depend. Therefore, elimination of α2-α3-α4 dimerization by the G108E mutation suggests that the α2-α3-α4 dimerization depends on the BH3-in-groove dimerization, a prediction from the model in which the BH3-in-groove dimerization is upstream of, and required for, other interactions that further oligomerize Bax at the MOM (Czabotar et al., 2013).

5.3.6 Helix α9 inside the MOM forms a dimer interface that has two conformations

Previous studies suggested that α9 is inserted into the MOM after Bax is activated (Annis et al., 2005; Westphal et al., 2014b). Whether the insertion results in a partially or fully embedded helix in the MOM is unknown. To determine the membrane topology of α9, we constructed ten Bax mutants with single cysteines positioned throughout α9 and did the IASD-labeling experiment after activating and targeting the Bax mutants to the Bax−/−/Bak−/− mitochondria. The results (Figure 5.5A; Figure S5:5) show that seven residues in the middle three turns of α9 helix are embedded in the MOM, as evidenced by their relative membrane burial indices being 40% or more of that of the membrane-buried reference Gly179 (Figure 5.5B). In contrast, the relative membrane and protein burial indices of two residues in the N-terminal two turns and one residue in the C-terminal turn are < 40% of that of the respective reference Gly179 and Tyr115 (Figure 5.5B and C), indicating that these parts of α9 are exposed to aqueous milieu. Therefore, like α5, α9 is partially embedded in the MOM in a fraction of the activated mitochondrial Bax molecules (Figure 5.5D).

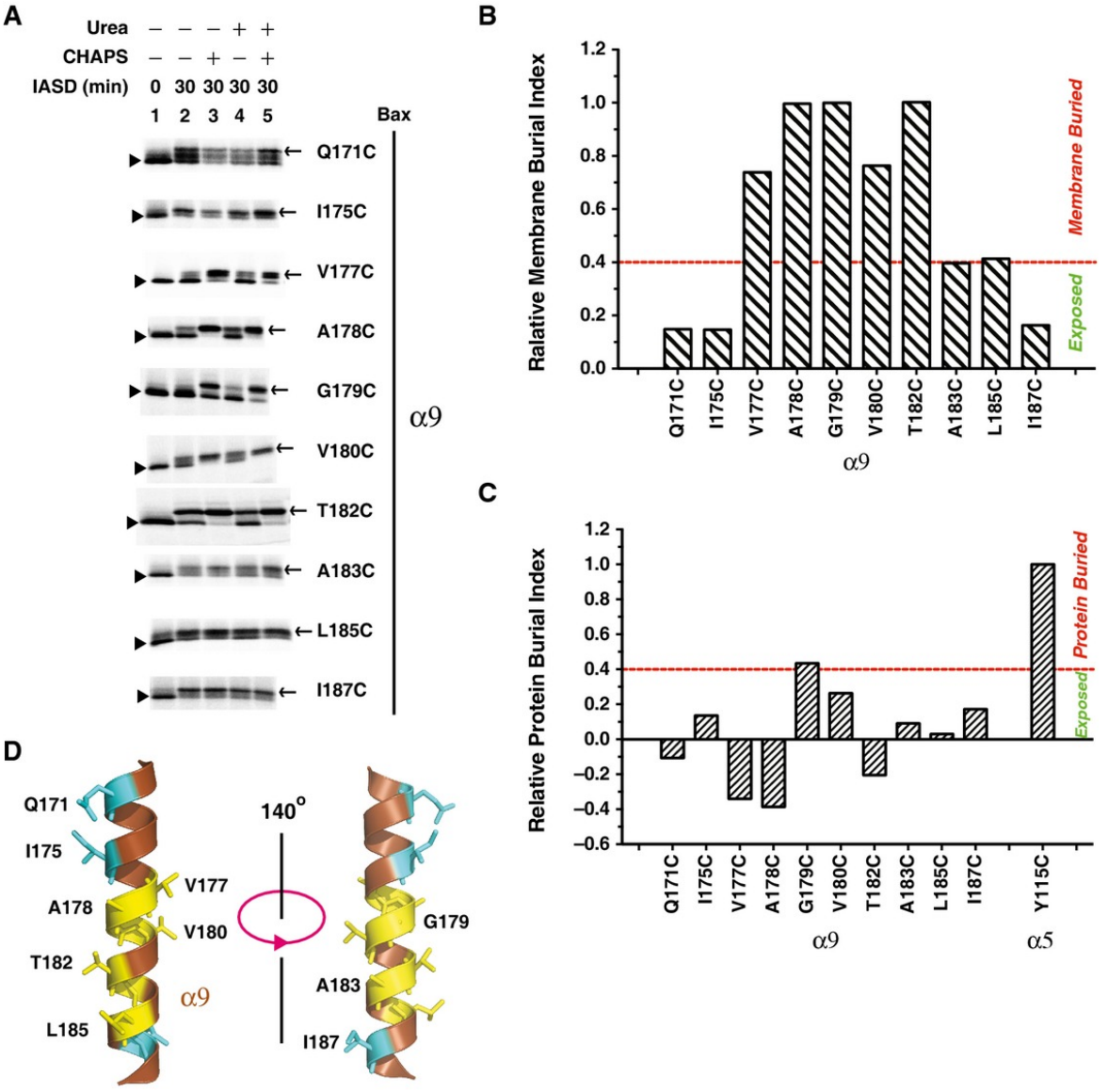

Figure 5.5: Helix α9 is partially embedded in the MOM

A) The *in vitro* synthesized radioactive Bax proteins, each with a single cysteine positioned in α9, were activated and targeted to the mitochondria, and labeled with IASD as in Figure 5.2B. The IASD-labeled and unlabeled Bax proteins were resolved using gradient SDS–PAGE, detected by phosphorimaging and indicated

by triangles and arrows, respectively. For all mutants, $n = 2$, except for A183C, $n = 3$.

B) Relative membrane burial index for single-cysteine variants of α9, quantified as described in Figure 5.2C.

C) Relative protein burial indices obtained as described in Figure 5.2D.

D) The structure of α9 (extracted from the NMR structure of Bax monomer, PDB entry 1F16) is shown with the residues that are buried in the MOM or exposed to the aqueous milieu, as the IASD-labeling data in (A–C) suggested, presented in yellow or cyan stick form, respectively.

Sequence analysis of α9 revealed a GxxxA motif in the C-terminal half (Gly179-Val-Leu-Thr-Ala183; Figure EV5.1A). GxxxG and GxxxG-like motifs are often found in transmembrane domains and are known to promote dimerization (Russ and Engelman, 2000; Senes et al., 2000). In particular, the GxxxG-like motifs are involved in homodimerization of transmembrane helices from glycophorin A and a BH3-only protein Bnip-3 (MacKenzie, 1997; Sulistijo and MacKenzie, 2009). To investigate the hypothesis that the GxxxA motif mediates homodimerization of Bax α9 in the MOM as previously observed (Bleicken et al., 2014; Gahl et al., 2014; Iyer et al., 2015), we built a homodimer model for the α9 using the program CATM (Mueller et al., 2014). We tested the model by generating single-cysteine Bax mutants with the cysteines positioned in the dimer interface (Figure 5.6A), and induced the disulfide crosslinking after activating and targeting these mutants to the Bax−/−/Bak−/− mitochondria. As predicted by the model, the integral Bax I175C, G179C, A183C, and I187C mutants formed disulfide-linked homodimers (Figures 5.6B and C, and EV5.5). Particularly, the cysteine that replaced Ala183 of the GxxxA

motif yielded the strongest homodimer band among the four cysteines, comparable to the dimers captured by the cysteines in the BH3-in-groove and α2-α3-α4 interfaces (e.g., the disulfide-linked L59C-M79C or C62-L76C dimer in Figure 5.1B, and E69C-E69C or L59C-L76C dimer in Figure 5.3A). These data suggest that like the other two dimerizations outside the MOM, GxxxA-mediated α9 dimerization inside the MOM occurs between activated mitochondrial Bax proteins.

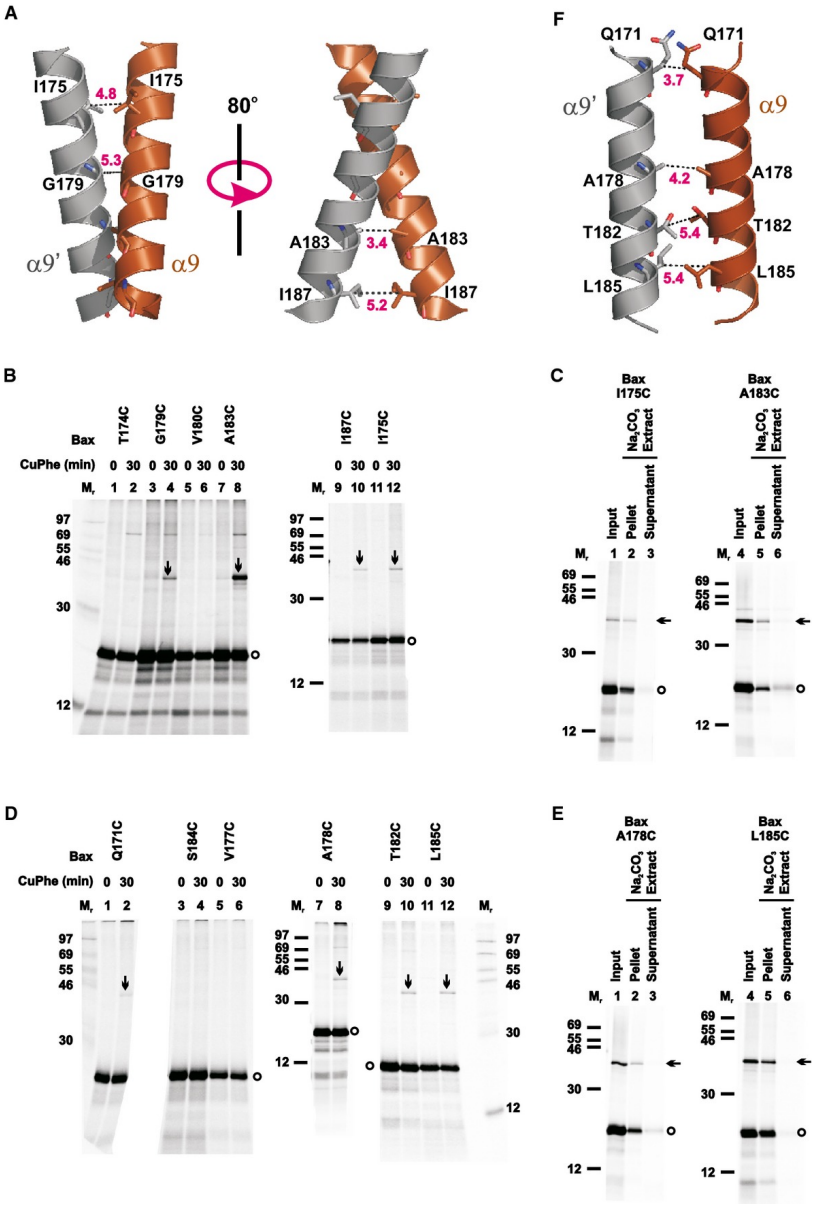

Figure 5.6: Helix α9 homodimerizes with two interfacial conformations

A) Structural model for intersected α9 homodimer is shown with one monomer colored brown (α9) and the other gray (α9′). The residue pairs that formed disulfide bonds after they were replaced with cysteine pairs in (B) are presented in stick form, and their β-carbon atoms linked by dashed lines with the distances

indicated in Å, except that the Gly^{179} pair is linked by a dashed line via their α carbon atoms with the distance indicated.

B) Oxidized mitochondria with radioactive single-cysteine Bax proteins testing the α9 intersected model.

C) Na_2CO_3 extraction of oxidized mitochondria with radioactive Bax I175C or A183C.

D) Oxidized mitochondria with radioactive single-cysteine Bax proteins testing the α9 parallel model.

E) Na_2CO_3 extraction of oxidized mitochondria with radioactive Bax A178C or L185C.

F) Structural model for parallel a9 homodimer is shown with the color codes described in (A). The residue pairs that formed disulfide bonds after they were replaced with cysteine pairs in (D) are presented in stick form, and their b carbon atoms linked by dashed lines with the distances indicated in Å.

Data information: In (B–E), protein standards, Bax monomers and disulfide-linked Bax dimers are indicated as in Figure 5.1B. *n* = 3 for A183C, and A178C; 2 for other mutants.

To further test the GxxxA α9 dimer model, disulfide crosslinking was performed with other single-cysteine Bax mutants with the cysteines located at nearby positions that would have low propensity to form disulfides according to the model. Four of the Bax mutants (T174C, V177C, V180C, and S184C) did not generate disulfide-linked homodimers, as predicted by the GxxxA model (Figure 5.6B and D). Unexpectedly, the other four mutants (Q171C, A178C, T182C, and

L185C) generated homodimers (Figues 5.6D and E, and EV5.5). We therefore identified another dimer model from our computational search that could fit these positive disulfide-crosslinking data. In this alternative dimer model, the two helices are more parallel (thereby termed the parallel α9 dimer model, Figure 5.6F) in contrast to the GxxxA dimer model, in which the two helices intersect with a right-handed crossing angle of approximately −40° (thereby termed the intersected α9 dimer model, Figure 5.6A). The parallel α9 dimer model could also accommodate the negative disulfide-crosslinking data from the other four Bax mutants. However, to account for all of the crosslinking data, both models must be correct, suggesting that the disulfide-crosslinking approach identified specific, non-random, yet flexible interactions between the two α9 helices.

A recent FRET study also found a α9 dimer interface after translocation of Bax to the mitochondria in apoptotic cells (Gahl et al., 2014). In accordance with the above-mentioned study, our crosslinking data show that both α9 dimers are formed by the Bax proteins that are activated by BH3 protein or peptide and integrated into the MOM (Figures 5.6C and E, and EV5.5B–D). To assess the effect of lipids on the dimer structure and stability, we carried out MD simulations, described in Supplementary Methods and Table S5.1, for the α9 dimers in solution containing counterions and in a lipid bilayer comprised of the MOM characteristic lipids, particularly those negatively charged ones known to facilitate Bax activation, insertion, oligomerization, and pore formation (Kuwana et al., 2002; Lucken-Ardjomande et al., 2008; Shamas-Din et al., 2015). We compared the initial conformations of both intersected and parallel α9 dimers (Figure 5.6A and F), which were generated by the computational modeling and fit the crosslinking data, to the final conformations after 175-ns simulation. The results indicate that both dimer conformations in the lipid bilayer remained stable, with small backbone root-mean-

square deviation (RMSD) fluctuations ~1 Å (Figure S5:6A, membrane). In contrast, both dimer conformations in the solution became unstable resulting in large backbone RMSD fluctuations ~5 Å and loss of helicity from both termini (Figure S5:6A, solution). Therefore, the MD simulation data support an important role for the membrane in stabilizing the α9 dimer structures. Consistent with this prediction, the α9 dimer-specific crosslinking was observed in the mitochondrial fraction after Bax proteins were activated by cBid, but not in the soluble fraction even when cBid was present (Figure EV5.5D).

To determine the relative stability of the intersected and parallel α9 dimer conformations in the membrane, we did free-energy simulation of α9 dimer transition between the two conformational states. The results suggest that the free energy of the intersected dimer is ~3.5 kcal/mol lower than the parallel one (Figure S5:6B). In addition, several intermediate conformational states with different free energy were revealed. In summary, in membranes the intersected α9 dimer is more stable than the parallel dimer, and they may coexist in equilibria with other intermediate dimeric conformations.

5.3.7 The two α9 dimer conformations are formed interdependently, and the formation of both depends on the BH3-in-groove dimerization

The MD simulation data suggest that the two extreme α9 dimer states may be linked through the intermediate states and transitions between them are possible if the energy barriers between them can be surpassed. To determine whether one α9 dimer state could be reached in the absence of the other, we designed one mutation for each interface that would be disruptive to that interface but not the other based on the respective “static” models (Figure 5.6A and F). The intersected dimer model predicts van der Waals' interactions between Gly179 of one α9 and Phe176 and Val180 of the other α9 (Figure 5.7A, intersected α9 dimer, left). Thus, a Gly179 to

isoleucine mutation (G179I) would not only eliminate these favorable interactions but also introduce steric clashes between Ile179 of one α9 and Phe176, Ile179 and Val180 of the other α9, thereby disrupting the intersected dimer (Figure 5.7A, intersected α9 dimer, right). The parallel dimer model predicts favorable van der Waals' interactions between Gln171, Ile175, Ala178, Thr182, and Leu186 in the two α9 helices. Mutating Thr182 to isoleucine (T182I) would remove the favorable interaction between the two Thr182 (Figure 5.7A, parallel α9 dimer, left) and introduce a steric clash between the two Ile182 (Figure 5.7A, parallel α9 dimer, right), thereby disrupting the parallel dimer. Confirming these predictions, the G179I or T182I mutation greatly inhibited the disulfide crosslinking of Bax A183C or A178C that was specific to the intersected or parallel dimer, respectively (Figure 5.7B, arrow-indicated bands in lane 4 versus lane 2, or lane 8 versus lane 6).

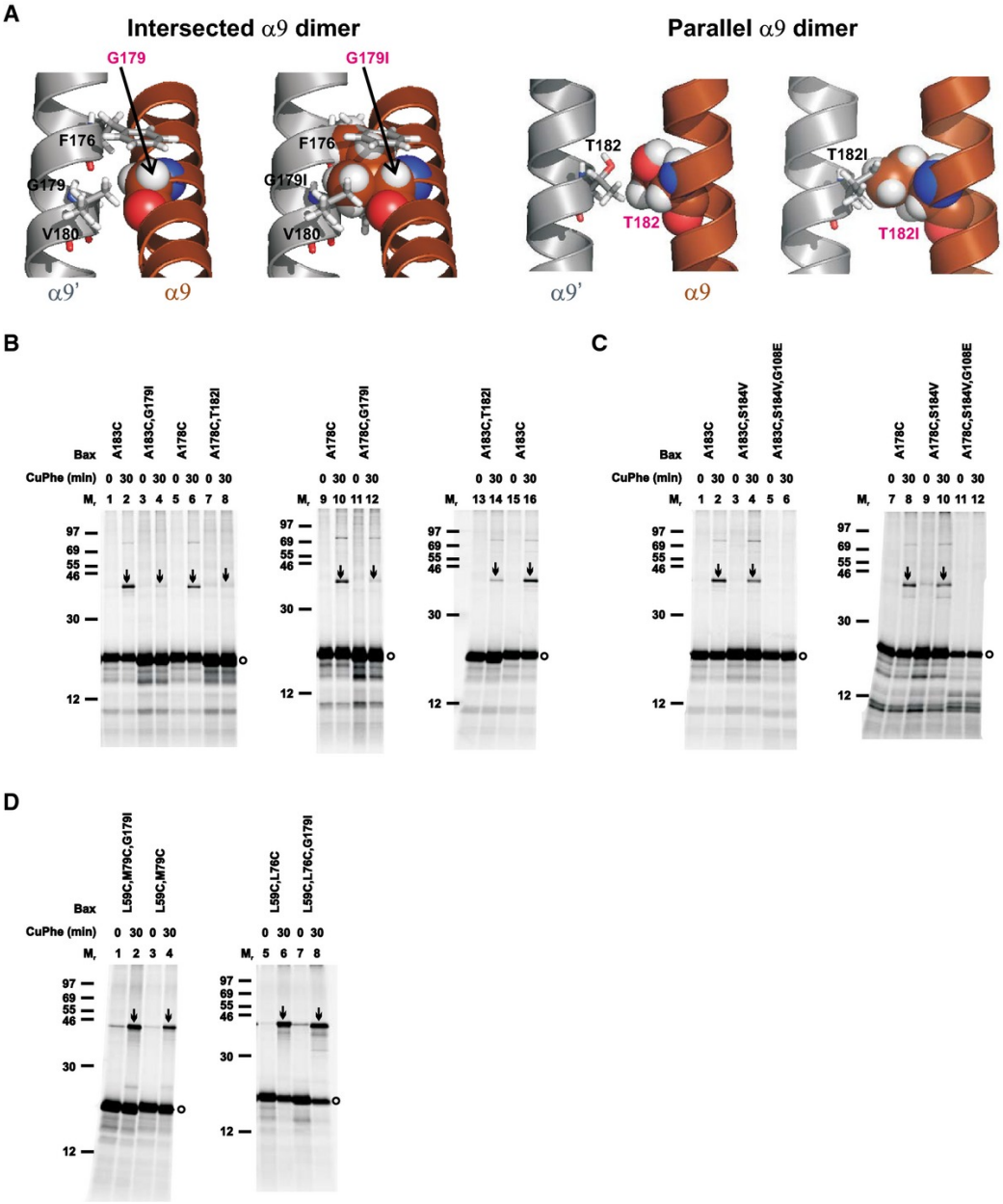

Figure 5.7: G179I or T182I mutation in α9 disrupts the α9 dimer but not the BH3-in-groove and α2-α3-α4 dimers, whereas G108E mutation in the groove disrupts the α9 dimer

A) Intersected α9 dimer: the structural model on the left depicts one monomer with Gly^{179} in sphere form, and the other monomer with Phe^{176}, Gly^{179}, and Val^{180} that have van der Waals' contacts with the first Gly^{179} in stick form. The same model is shown on the right, except that the two Gly^{179} are changed to isoleucine (G179I), which results in steric clashes between the Ile^{179} in one monomer and the Phe^{176}, Ile^{179} and Val^{180} in the other monomer, and vice versa. Parallel α9 dimer: the structural model on the left depicts the two interfacial Thr^{182} that have van der Waals' contacts in sphere or stick form. The same model is shown on the right, except that the two Thr^{182} are changed to isoleucine (T182I), which results in a steric clash with each other.

B) The *in vitro* synthesized radioactive single-cysteine Bax proteins with or without the indicated mutations were activated and targeted to the mitochondria, oxidized, and analyzed as in Figure 5.1B.

C) Oxidative crosslinking of radioactive Bax A183C or A178C in the presence of the S184V and G108E mutations.

D) The *in vitro* synthesized radioactive double-cysteine Bax proteins with or without the G179I mutation were activated and targeted to the mitochondria, oxidized, and analyzed as in Figure 5.1B. Protein standards, and monomers and disulfide-linked dimers of the Bax proteins are indicated as in Figure 5.1B. $n = 2$ for all mutants.

The G179I or T182I mutation also greatly inhibited the disulfide crosslinking of Bax A178C or A183C, respectively (Figure 5.7B, arrow-indicated bands in lane 12 versus lane 10, or lane 14 versus lane 16), suggesting that the disruptive mutation in the intersected dimer is also disruptive to the parallel dimer and vice versa. This result was not predicted by the static dimer models, suggesting that the two α9 dimer conformations are formed in an all-or-none fashion or interdependently.

We carried out MD simulations with the α9 dimers containing the G179I or T182I mutation in the membrane to assess the effects of these mutations in a dynamic setting. Comparison of the initial and final structures of the dimers suggests that the G179I mutation disrupts both intersected and parallel dimers, whereas the T182I mutation only disrupts the parallel dimer (Figure S5:6C). To compare with the effect of each mutation observed in the crosslinking experiment with the A183C or A178C single-cysteine mutant that captured the intersected or parallel dimer, respectively (Figure 5.7B), we measured the β-carbon distance between the two Ala183 residues or between the two Ala178 residues in the two α9 dimers with the G179I or T182I mutation in the initial and final structures. The results indicate that the G179I mutation increases the β-carbon distance for both residue pairs (Figure S5:6C), thereby explaining its inhibitory effect on the disulfide crosslinking of both A183C and A178C mutants. In contrast, the T182I mutation only increases the β-carbon distance for the Ala178 residue pair significantly, thereby only explaining its inhibition of the A178C crosslinking. While the inhibition of the A183C crosslinking or the intersected α9 dimerization by the T182I mutation can be explained by neither the MD simulation nor the static modeling, it remains possible that the intersected α9

dimer with the lowest free energy may be evolved from the parallel dimer through the intermediate conformational states revealed by the free-energy simulation of the transition (Figure S5:6B). And if so, blocking the upstream parallel dimerization by the T182I mutation may also block the downstream intersected dimerization.

The model that the BH3-in-groove dimerization is a nucleation event that initiates other dimerizations to further oligomerize Bax predicts that the α9 dimerization would depend on the BH3-in-groove dimerization (Czabotar et al., 2013). To verify this prediction, we tested the effect of the G108E mutation that disrupted the BH3-in-groove dimerization on the α9 dimerization. The G108E mutation completely abolished the homo-disulfide crosslinking of both Bax A183C and A178C that specifically captured the two conformations of the α9 dimer, respectively, even though both Bax mutants were targeted to the mitochondria by the S184V mutation in α9, which by itself did not abolish the homo-disulfide crosslinking (Figure 5.7C, disappearance of the arrow-indicated bands from lanes 6 and 12, but not lanes 4 and 10). These data suggest that α9 dimerization does not occur in the absence of the BH3-in-groove dimerization, thereby placing the α9 dimerization downstream of, and dependent on, the BH3-in-groove dimerization.

In contrast, the G179I mutation that disrupted the α9 dimerization via both conformations did not affect the BH3-in-groove dimerization as captured by the disulfide linkage of Bax L59C,M79C (Figure 5.7D, arrow-indicated bands in lane 2 versus lane 4). The G179I mutation also did not reduce the α2-α3-α4 dimerization as detected by the disulfide linkage of Bax L59C,L76C (Figure 5.7D, arrow-indicated bands in lane 8 versus lane 6). Thus, both BH3-in-groove and α2-α3-α4 dimer interfaces are formed in the absence of α9 dimerization, as expected

from the model that places the BH3-in-groove and the α2-α3-α4 dimerizations upstream of the α9 dimerization.

5.3.8 The α9 dimerization links the BH3-in-groove or the α2-α3-α4 dimer into higher order oligomers

The oligomerization of Bax requires at least two separate dimer interfaces. To determine whether the BH3-in-groove or the α2-α3-α4 dimer interface located on the MOM and the intersected or parallel α9 dimer interface located in the MOM could mediate oligomerization of Bax dimers, we generated Bax mutants with three cysteines. The first two cysteines, L59C and M79C or L59C and L76C, were positioned in the BH3-in-groove or the α2-α3-α4 dimer interface, and the third cysteine, A183C or A178C, was positioned in the intersected or the parallel α9 dimer interface, respectively. After activating and targeting to the Bax−/−/Bak−/− mitochondria, and oxidizing by CuPhe, each of these triple-cysteine Bax mutants formed multiple products with Mr higher than the monomers (Figure 5.8A, lanes 4, 10, 16, and 22, indicated by arrows, and open and closed triangles). The products with the Mr below 46 are the disulfide-linked Bax homodimers, because they were also formed by the corresponding double- or single-cysteine Bax mutant (lanes 2, 6, 8, 12, 14, 18, 20, and 24, indicated by arrows). The product with the Mr above 69 formed by the triple-cysteine Bax mutants containing A178C or A183C is not a disulfide-linked dimer, because it could not be reduced (compare open triangle-indicated band in lanes 4 and 10 of non-reducing gel to that in the respective lanes of reducing gel). The remaining products (lanes 4, 10, 16, and 22) indicated by closed triangles) are the disulfide-linked Bax oligomers, as judged by their Mr and disappearance on the reducing gel. These products were detected only in the samples containing the triple-cysteine mutants but not those containing the double- or single-cysteine mutants, suggesting that the two disulfides in the

BH3-in-groove or the α2-α3-α4 dimer interface together with the one disulfide in the intersected or the parallel α9 dimer interface linked Bax monomers into oligomers. Together, these results suggest that the α2, α3, and α4 with or without α5 formed a dimer interface that was separated from another dimer interface formed by the α9, and that the two dimer interfaces together mediated Bax oligomerization.

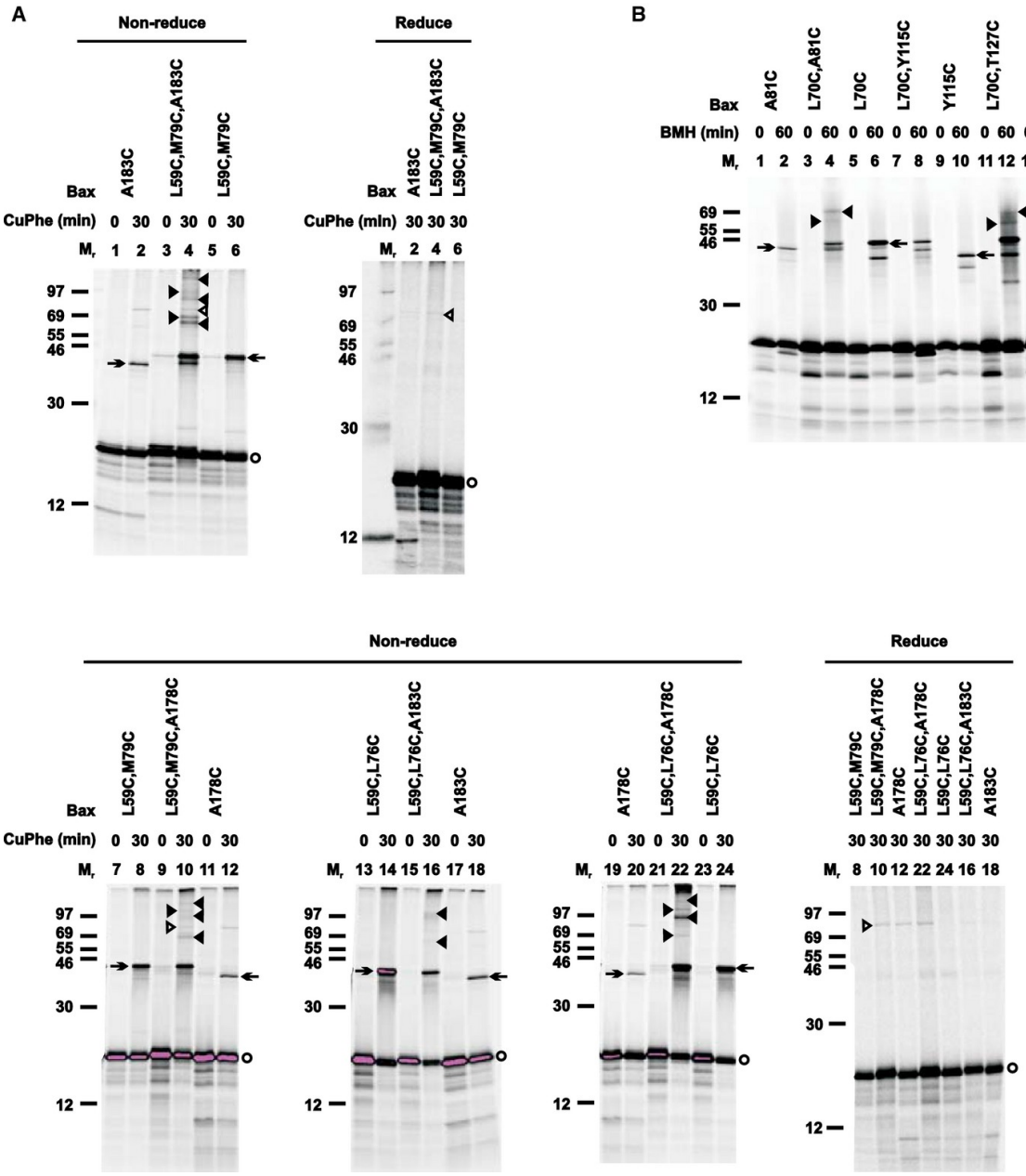

Figure 5.8: Combinations of either the BH3-in-groove or the α2–α3–α4 dimer interface with the α9 dimer interface mediate Bax oligomerization

A) The oxidized mitochondria with radioactive single-, double-, or triple-cysteine Bax proteins were prepared and analyzed by non-reducing or reducing SDS–PAGE as

in Figure 5.1B. Protein standards, Bax monomers, and disulfide-linked Bax homodimers are also indicated as in Figure 5.1. Closed triangles indicate disulfide-linked higher order Bax homo-oligomers. Open triangles indicate a background band that could not be reduced, and hence is not a disulfide-linked complex.

B) The BMH-treated mitochondria with the radioactive single- and double-cysteine Bax proteins were prepared and analyzed as in Figure EV5.3G, except that closed triangles indicate the BMH-linked Bax oligomers.

Data information: In (A, B), $n = 2$ for all mutants.

A DEER study of a liposome-bound Bak complex suggested proximity between two BH3-in-groove dimers, particularly the C-termini of the α3 and α5 helices (Aluvila et al., 2014). To determine whether the mitochondrion-bound Bax complex adapts a similar configuration, we used BMH to crosslink the active mitochondrial Bax mutant A81C or T127C with the single cysteine located at the C-terminus of either α3 or α5, respectively. The results show that BMH could link each Bax mutant to a homodimer (Figure 5.8B, lanes 2 and 14, indicated by arrow). Because the distance between the two Ala81 or the two Thr127 is longer than 35 Å according to the BH3-in-groove dimer structure (Czabotar et al., 2013) and the DEER-derived Bax dimer model (Bleicken et al., 2014), the BMH linkage would not form intradimerically. Instead, the BMH linkage should form interdimerically, if the neighboring BH3-in-groove dimers are in proximity. If this is true, one of these interdimer BMH linkages plus an intradimer BMH linkage (e.g., via the L70C shown in Figure EV5.3G and H) should produce higher order oligomers. This prediction was confirmed by the BMH crosslinking of two double-cysteine mutants with one cysteine (L70C) forming the intradimer linkage and the other (A81C or T127C) forming the

interdimer linkage, and both linkages together forming high order oligomers (Figure 5.8B, lanes 4 and 12, indicated by closed triangles). As a control, another double-cysteine mutant L70C,Y115C did not form higher order BMH-linked oligomers (Figure 5.8B, lane 8), because the two BMH linkages are in the same BH3-in-groove dimer (Figure EV5.3G and H). Taken together, the disulfide and the BMH crosslinking data support a Bax oligomer model in which the BH3-in-groove or α2-α3-α4 dimers are close to each other and linked to higher order oligomers by intersected or parallel α9 dimerization.

5.3.9 The BH3-in-groove dimerization is required for Bax to permeabilize the MOM

To determine which Bax dimer interface is required for assembly of a functional pore in the MOM, we monitored the effect of the interface-disrupting mutations on the MOMP activity of Bax. As shown in Figure 5.9, the G108E mutation in the BH3-in-groove dimer interface inhibited the tBid-induced and Bax-mediated cytochrome c release from the Bax−/−/Bak−/− mitochondria. This inhibition was not due to the deficiency of Bax binding to mitochondria caused by the G108E mutation, because it occurred in the presence of the S184V mutation that rescued the mitochondrial binding (Figure S5:1). This inhibition was caused by the deficiency of Bax dimerization via all of the four interfaces, because the G108E mutation inhibited the formation of all of them as determined by the interface-specific disulfide crosslinking (Figures 5.4 and 5.7C) This inhibition is independent of the locations of the cysteines in the Bax mutants (Figure 5.9) that were used in the disulfide-crosslinking experiments. Some of the Bax mutants (e.g., A112C,S184V and L59C,L76C,S184V) were autoactive, as they released cytochrome c in the absence of tBid. The G108E mutation inhibited the release of cytochrome c by these autoactive Bax mutants and the other Bax mutants that require tBid to activate,

suggesting that the inhibition occurred at a step downstream of Bax activation. Consistent with this hypothesis, the G108E mutation did not inhibit Bax interaction with tBid (Figure 5.4C). Therefore, Bax dimerization via the BH3-in-groove interface that is directly disrupted by the G108E mutation is not only required for dimerizations of Bax via other interfaces including the α2-α3-α4 and the α9 interfaces, but also for the MOMP by Bax.

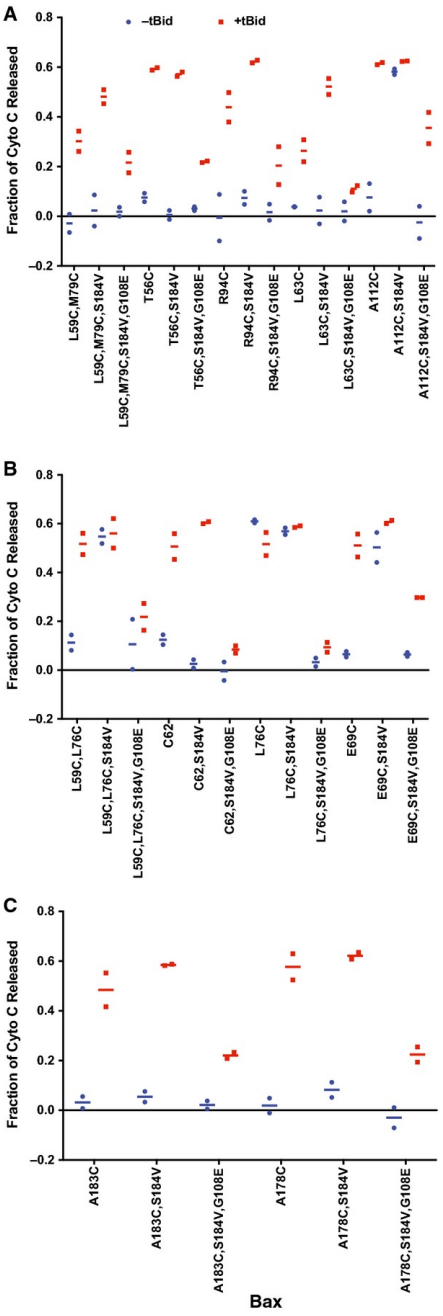

Figure 5.9: The BH3-in-groove dimerization is required for Bax to permeabilize the MOM

The single- or double-cysteine Bax proteins with or without the indicated mutations were synthesized *in vitro*, activated by tBid protein and targeted to the $Bax^{-/-}/Bak^{-/-}$

mitochondria. Cytochrome c release from the mitochondria was measured using ELISA. The dots are the fractions of cytochrome c release from two independent replicates after they were corrected as described in Figure EV5.1B, and the lines are the averages. The amount of the Bax mutants bound to the mitochondria in the assay was determined and the results are shown in Figure S5:1.

5.3.10 The α9 dimerization is not required for the formation of small pores in MOM, but facilitates the formation of large pores

To determine the effect of deficient α9 dimerization on Bax MOMP activity, we used the G179I and T182I mutations that abolished the α9 dimerization, but not the BH3-in-groove and α2-α3-α4 dimerizations as shown by the interface-specific disulfide crosslinking (Figure 5.7B and D; data not shown). These mutations did not reduce the tBid-induced cytochrome c release by Bax with cysteine(s) in any of the four interfaces (Figure 5.10A). In fact, the G179I mutation made some of these Bax cysteine mutants autoactive, thereby releasing cytochrome c in the absence of tBid.

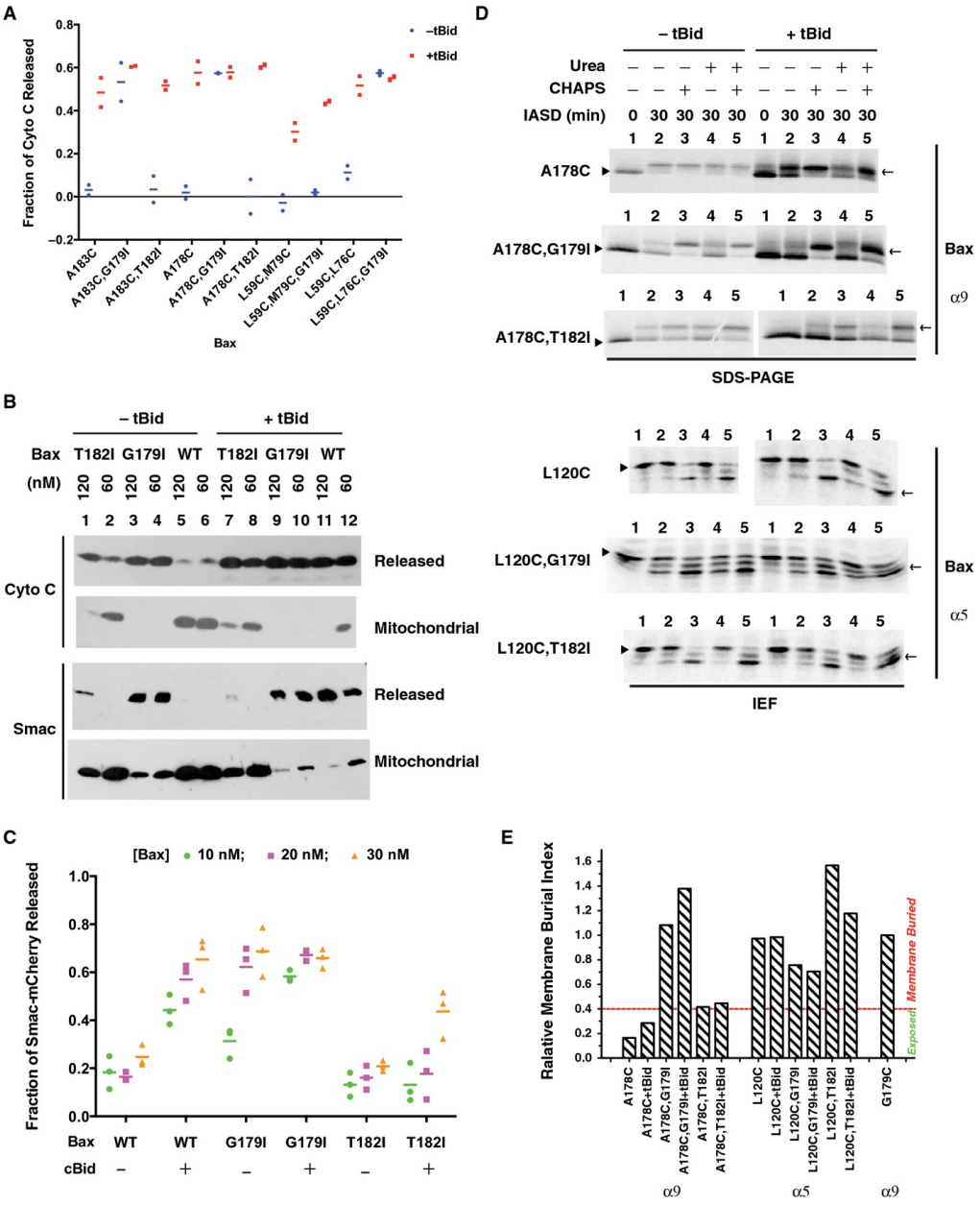

Figure 5.10: The α9 dimerization is not required for the release of cytochrome c from mitochondria, but facilitates the release of Smac and larger proteins

The cytochrome c release by the indicated Bax mutants was measured, corrected and shown as in Figure 5.9. The dots are the fractions of cytochrome c release from two

independent replicates, and the lines are the means. The amount of the Bax mutants bound to the mitochondria in the assay was determined and the results are shown in Figure S5:1.

A) Purified wild-type (WT) or mutant (G179I or T182I) Bax protein of the indicated concentration was incubated with the $Bax^{-/-}/Bak^{-/-}$ mitochondria in the absence or presence of tBid protein. The proteins released from and those remaining in the mitochondria were separated and analyzed by SDS–PAGE and immunoblotting with antibody specific to either cytochrome c or Smac. $n = 2$.

B) Purified wild-type or mutant Bax protein of the indicated concentration alone or together with purified cBid was incubated with the $Bax^{-/-}/Bak^{-/-}$ mitochondria containing Smac-mCherry protein. Release of Smac-mCherry was detected by measuring the fluorescence in the supernatant after the mitochondria were pelleted by centrifugation. The dots are the fractions of Smac-mCherry release from three independent replicates, and the lines are the means.

C) The *in vitro* synthesized radioactive Bax proteins with indicated mutations were targeted to the mitochondria in the absence or presence of purified tBid protein, and then labeled with IASD, as in Figure 5.2B. The IASD-labeled and unlabeled Bax proteins were resolved using either gradient SDS–PAGE or IEF, detected by phosphor-imaging, and indicated by arrows and triangles, respectively. $n = 3$ for A178C,G179I; 2 for other mutants.

D) The IASD-labeling data in (D) and the similar data from the independent replicates were quantified to derive the membrane burial indices shown in Figure S5:7A,

from which the relative membrane burial indices were obtained as described in Figure 5.2C, and shown.

To verify these results, we purified wild-type Bax protein and mutant Bax protein with either G179I or T182I as the only mutation. We incubated each protein with the Bax−/−/Bak−/− mitochondria in the absence or presence of purified tBid protein and monitored the cytochrome c release using immunoblotting. Consistent with the results obtained from the *in vitro* synthesized Bax proteins using the ELISA-based cytochrome c release assay (Figure 5.10A), the immunoblots clearly show a tBid-induced cytochrome c release by the purified wild-type Bax protein (Figure 5.10B, cytochrome c blots, lanes 5 and 6 versus 11 and 12). The result from the T182I mutant at 60 nM was similar to the wild-type protein, but at 120 nM it released cytochrome c in the absence of tBid, and hence was autoactive (lanes 1 and 2 versus 7 and 8). In contrast, the G179I mutant at both concentrations was autoactive, releasing cytochrome c in the absence of tBid (lanes 3 and 4 versus 9 and 10). Therefore, the dimerization of α9 is not required for Bax to form small pores in the MOM that release cytochrome c, a 12-kDa globular protein with diameter ~30 Å (PDB entry 2AIU) (Liu et al., 2006).

We next assayed the release of Smac, a 27-kDa cylinder-like intermembrane space protein with diameter ~25 Å and height ~100 Å (PDB entry 1FEW) (Chai et al., 2000) that is released from mitochondria during apoptosis (Du et al., 2000; Verhagen et al., 2000). As expected, the wild-type Bax released Smac dependent on tBid (Figure 5.10B, Smac blots, lanes 5 and 6 versus 11 and 12), whereas the G179I mutant released Smac independent of tBid (lanes 3 and 4 versus 9

and 10). Surprisingly, the T182I mutant only released traces of Smac, even at the high concentration and in the presence of tBid (lanes 1 and 2 versus 7 and 8). Therefore, at least one of the α9 dimer-disruptive mutations inhibited the formation of Smac-releasing pores that might be larger than the cytochrome c-releasing pores.

To further test this hypothesis, we monitored the release of Smac-mCherry, a 37-kDa fluorescent fusion protein that contains a cylinder-like GFP fold with diameter ~35 Å and height ~45 Å (PDB entry 1EMA) (Ormö et al., 1996) and is localized to the intermembrane space of the mitochondria isolated from the *bax/bak* DKO BMK cells (Shamas-Din et al., 2014). By comparing the fluorescence intensities in the supernatant and the mitochondrial pellet, we found that the wild-type Bax released Smac-mCherry in a dose- and cBid-dependent manner (Figure 5.10C). The G179I mutant also released Smac-mCherry, but independent of cBid, particularly at higher concentrations. In contrast, the T182I mutant released significantly less Smac-mCherry, even at the highest concentration and in the presence of cBid. These data and those from the Smac release assay (Figure 5.10B) are consistent with the conclusion that α9 dimerization facilitates the formation of large pores in the MOM that release large proteins.

To determine why the G179I and T182I mutations that inhibited the α9 dimerization to the same extent displayed different phenotypes in the large mitochondrial protein release assays, we performed IASD-labeling experiment to examine whether the mutations affect the insertion of α9 and α5 into the MOM. The labeling profiles (Figure 5.10D) and the membrane burial indices (Figure 5.10E; Figure S5:7A) of wild-type Bax and the two mutants that contain a single-cysteine (A178C) in α9 indicate that the G179I mutation greatly enhanced the α9 insertion into the MOM compared to the wild-type protein, whereas the T182I mutation did not. In contrast,

the T182I mutation enhanced the α5 insertion, whereas the G179I mutation did not, as shown by the data from Bax L120C with the single cysteine in α5 (Figure 5.10D and E; Figure S5:7A).

We used disulfide crosslinking to assess the effect of G179I mutation on the BH3-in-groove and α2-α3-α4 dimerizations at the mitochondria before and after Bax activation by cBid. The L59C,M79C and the L59C,L76C with the G179I mutation formed the BH3-in-groove and the α2-α3-α4-specific disulfide-linked dimer, respectively, in the mitochondrial pellet fraction in the absence of cBid (Figure EV5.5E). Dimer formation without cBid is consistent with our data showing that this mutation autoactivates Bax such that it permeabilized the MOM in the absence of tBid. In comparison, dimer formation by the respective cysteine mutants without the G179I mutation in the mitochondrial fraction was cBid dependent (Figures EV5.3E and EV5.4D). Intriguingly, a significant dimerization of the L59C,L76C occurred in the post-mitochondrial supernatant fraction (Figure EV5.4D), which was dramatically enhanced by the G179I mutation but not by cBid or mitochondria or both (Figure EV5.5E). Whether this dimerization in the soluble fraction contributes to the MOMP activity is uncertain. Nonetheless, increased α9, but not α5, insertion into the MOM can compensate for a deficiency of α9 dimerization, promoting the dimerization on the membrane surface and large pore formation in the membrane. When α9 insertion is at the wild-type level, α9 dimerization facilitates the formation of large pores.

5.3.11 Small pore formation in the MOM is necessary and sufficient for Bax to kill cells

To determine whether the *in vitro* activities of these Bax mutants are relevant to the biological activity of Bax in cells, we expressed wild-type Bax, and the G108E, S184V, G108E/S184V, G179I, and T182I mutants as N-terminal Venus fusion proteins in *bax/bak* DKO BMK cells, and assessed their intracellular localization and apoptotic activity in the absence and presence of

STS. Expression levels of these mutants were similar to that of Venus-WT Bax, except for Venus-Bax G108E that had an increased expression (Figure EV5.2A), which was likely tolerated by the cells since this mutant induced a significantly lower level of apoptosis in the absence and presence of STS when compared to Venus-WT Bax (Figure EV5.2B). In untreated cells, this mutant was mainly localized to the cytoplasm similar to the wild-type protein (Figure EV5.2C and D).

Addition of the S184V mutation to Venus-Bax G108E increased the mitochondrial localization resulting in apoptotic activity comparable to wild-type Bax and inducible by STS (Figure EV5.2B–D). However, the apoptotic activity of the G108E/S184V double mutant was significantly lower than the S184V single mutant (Figure EV5.2B), even though they had similar mitochondrial localization (Figure EV5.2C and D). Therefore, the G108E mutation inhibited an activation step downstream of mitochondrial localization of Bax, such as the dimerizations detected by the crosslinking assay above. Furthermore, the residual apoptotic activity of the G108E mutant that could be amplified by the S184V mutation is consistent with the residual dimerizations and cytochrome c release observed *in vitro* (Figures 5.4 and 5.9). Together, the results from these mutants in our cell-free and cell-based systems suggest that the G108E mutation incompletely blocks Bax dimerization, MOMP and apoptotic activities especially in the presence of the S184V mutation that targets Bax to the mitochondria.

In contrast, the G179I mutant was constitutively active and spontaneously killed cells (Figure EV5.2B), an observation that is consistent with those from the *in vitro* MOMP assays (Figure 5.10A–C). Due to the extensive cell death caused by the G179I mutant, accurate localization data was not obtainable. However, because the cells were dying, Venus-Bax G179I was likely

targeted to the mitochondria constitutively, a notion supported by the mitochondrial association of the G179I-containing Bax cysteine mutants observed *in vitro* (Figure EV5.5E).

When expressed in cells, the T182I mutant had similar localization and slightly higher apoptotic activity compared to that of WT Bax (Figure EV5.2B–D). While the apoptosis data from the cells expressing the T182I mutant are mostly in line with the cytochrome c release data from the isolated mitochondria reconstituted with the *in vitro* synthesized or purified T182I mutant (Figure 5.10A and B), the defect of the T182I mutant detected in the *in vitro* Smac and Smac-mCherry release assays (Figure 5.10B and C) does not translate into a defect in the cell-based apoptosis assay. The most simple explanation is that the cytochrome c release caused by the T182I mutant in the cells is sufficient to induce apoptosis, and the reduction of the Smac release caused by the mutation in the cells is not sufficient to block apoptosis. Therefore, the small pore-forming activity of Bax, which was largely inhibited by the G108E, and enhanced by the G179I, but not altered by the T182I mutation, mirrors the cellular apoptotic activity. The large pore-forming activity, which was inhibited by the T182I mutation, is not required at least for Bax to kill the *bax/bak* DKO BMK cells.

5.4 Discussion

In this study, we made several major observations about the conformational changes of activated Bax at the MOM, which are accompanied by a series of protein–protein and protein–membrane interactions, resulting in oligomeric pores that permeabilize the MOM. These observations, together with those published previously, allowed us to propose a mechanistic model by which the activated Bax induces MOMP, as illustrated in Figure 5.11.

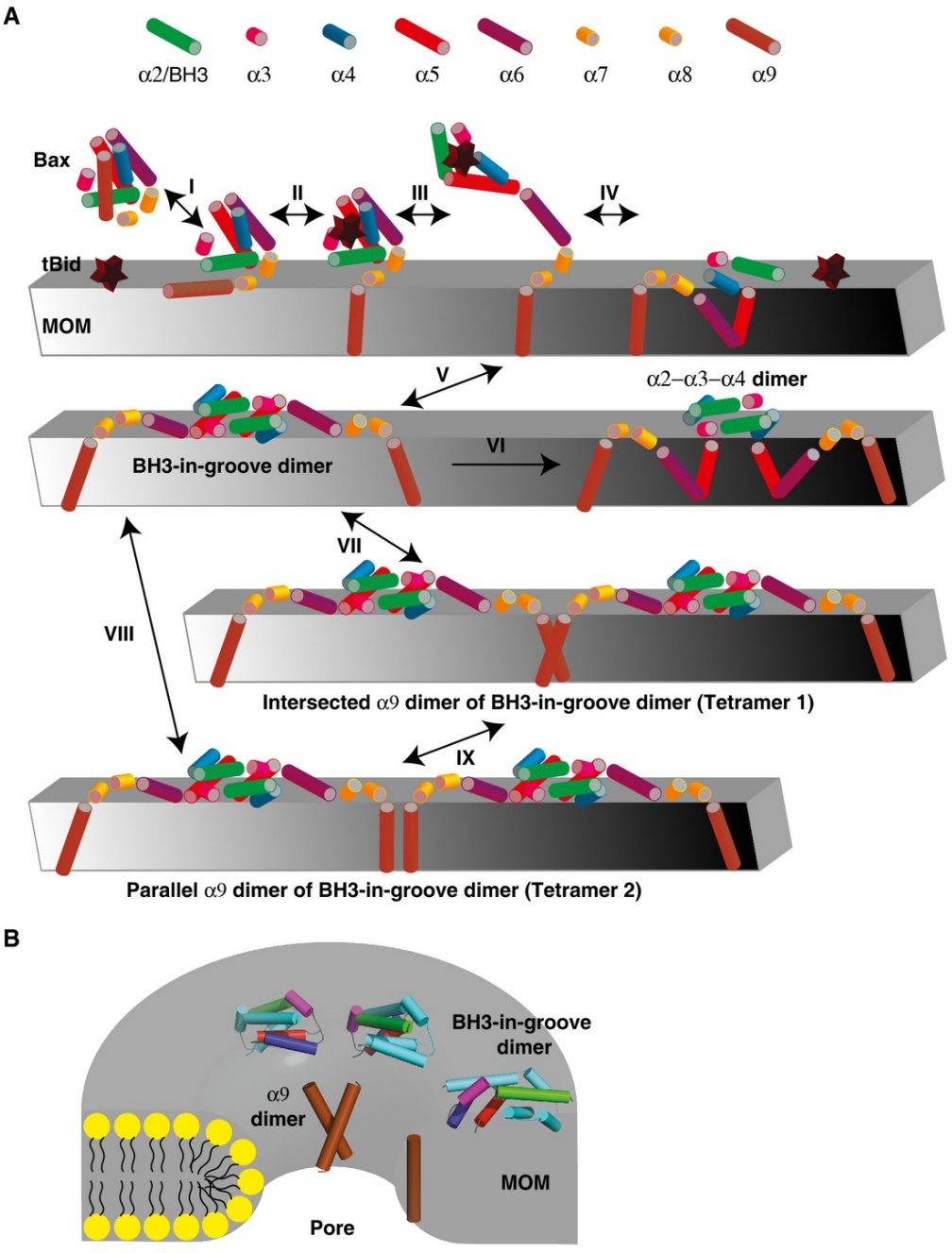

Figure 5.11: A model for conformational changes and interactions of Bax at the MOM, and the resulting pore

A) Schematics illustrate the soluble Bax monomer and the MOM-bound Bax monomers, dimers and tetramers. The helices α2-α9 are depicted as cylinders with the indicated colors, and the helix α1 is omitted for simplicity. The MOM-bound tBid, depicted as a star, enhances cytosolic Bax interaction with the membrane (step I) and binds to the groove of Bax, thereby releasing α9 from the groove to the membrane (step II). The tBid binding also unfolds the Bax separating the latch (α6-α8) from the core (α2-α5) (step III). After tBid dissociates, the activated Bax monomer either embeds more hydrophobic regions such as α5 and α6 into the membrane (step IV) or binds other activated Bax monomer forming a BH3-in-groove dimer on the membrane surface (step V). After α5 is released into the membrane, the remnant BH3-in-groove dimer is rearranged to a α2-α3-α4 dimer (step VI). A Bax tetramer is then formed by two BH3-in-groove dimers via an additional dimer interface formed by two membrane-embedded α9 helices either intersected (step VII) or in parallel (step VIII). These Bax complexes may be in equilibrium with one another, but the BH3-in-groove dimerization is a prerequisite for the other dimerizations. In principle, combinations of these dimer interfaces would be sufficient to generate higher order oligomers, although other dimer interfaces have been discovered.

B) A lipidic pore in the MOM is illustrated with BH3-in-groove dimers located on the cytosolic surface of the membrane and α9 monomer and dimer at the rim. Note that the BH3-in-groove dimers are partially embedded in the cytosolic monolayer,

generating membrane tension that induces the pore, while the α9 monomer and dimer are integrated into the highly curved monolayer to reduce the line tension, thereby stabilizing and expanding the pore.

After soluble Bax is activated by a BH3 activator such as tBid and targeted to the MOM (Figure 5.11A, step I), it inserts into the MOM with either α9 helix alone (step II) or together with α5 and α6 helices (step IV) embedded in the MOM at least partially (Figures 5.2 and 5.5) (Annis et al., 2005; Gavathiotis et al., 2010; Lovell et al., 2008; Westphal et al., 2014b). Insertion of α9 is favored when the mitochondrial Bax binds to the activator via the canonical BH3-binding groove (step II; Figure 5.4C) (Czabotar et al., 2013; Kim et al., 2009), because the release of α9 from the groove empties the groove for the BH3 binding, and the BH3 binding to the groove prevents α9 from reoccupying the groove, thereby keeping α9 in the MOM. In line with this scenario, disruption of α9-groove interaction by the S184V mutation (Suzuki et al., 2000) or increase of α9 hydrophobicity by the G179I mutation enhances Bax mitochondrial targeting, MOMP and apoptotic activities (Figures 5.9, 5.10, EV5.2 and EV5.5E). When the BH3 activator dissociates from the groove, insertion of α5 and α6 is favored (step IV), because the empty groove exposes a large hydrophobic surface to aqueous milieu, making the cytosolic domain of the tail-anchored Bax less stable, thereby forcing a conformational change to bury the hydrophobic residues in the MOM, particularly those in the more hydrophobic α5 and those on the hydrophobic side of the amphipathic α2 (Figure 5.2). Consistently, our previous study suggested that the mitochondrial Bax monomer could embed α5, α6, and α9 in the MOM before oligomerization (Annis et al., 2005). A recent study also showed that the S184L mutation in α9 inserted not only α9 but also α5 and α6 into the MOM in the absence of tBid (Westphal et al.,

2014b). Alternatively, α9 could be released from the membrane and reoccupy the empty groove, thereby refolding the Bax monomer, which could then return to the cytosol through retrotranslocation, a cellular process regulated by both anti- and pro-apoptotic family members (Billen et al., 2008; Edlich et al., 2011; Schellenberg et al., 2013; Todt et al., 2013, 2015). The G108E mutation adds a strong hydrogen donor to the edge of the groove, which increases the propensity for hydrogen bonding with the α9 residues in vicinity, and hence could stabilize the α9 in the groove, thereby keeping Bax soluble in the cytosol (Suzuki et al., 2000). The S184V mutation apparently overwrites the inhibitory effect of G108E on the mitochondrial targeting reactivating the MOMP and apoptotic activity (Figures 5.9 and EV5.2) (Kim et al., 2009). In the absence of membranes, a domain-swapped Bax dimerization could bury some of these hydrophobic residues as shown by the crystal structure of BaxΔα9 dimer (Czabotar et al., 2013). However, such dimerization would unlikely to occur in cells because activation of Bax takes place at membranes (Lovell et al., 2008), and the domain-swapped BaxΔα9 dimer is a detergent-induced complex that does not form by active Bax at membranes, as demonstrated in a recent DEER study (Bleicken et al., 2014). Here, we also saw majority of the Bax dimerizations in the mitochondrial, but not the soluble fraction (Figures EV5.3E, EV5.4D, and EV5.5D).

The BH3-in-groove dimer interface is formed by two tail-anchored Bax proteins (step V) after departure of the BH3 activator from the groove that readies both the BH3 region and the groove of Bax for their engagement (Figure 5.1). The crystal structure of tBid BH3 peptide in complex with the domain-swapped BaxΔα9 dimer displays a cavity in the BH3-binding groove (Czabotar et al., 2013) that may destabilize the complex, thereby facilitating the departure of tBid BH3 peptide and the exposure of Bax BH3 region. This complex structure also shows separation of the latch (α6-α8) from the core (α1-α5) of Bax, implying that the binding of tBid BH3 peptide to

Bax groove induces the separation (step III). This “unlatching” event would occur before the BH3-in-groove dimerization required for Bax function (Figures 5.4 and 5.9), because disulfide linkage of α5 and α6 blocked tBid-induced MOMP by Bax (Czabotar et al., 2013), and opening of the α5-α6 hairpin in membrane-bound Bax is consistent with a DEER-derived distance between the spin pair located at the beginning and the end of the hairpin (Bleicken et al., 2014). The core–latch separation was also observed through FRET imaging in apoptotic cells (Gahl et al., 2014).

The BH3-in-groove dimerization exposes the hydrophobic residues in α4 and α5 to aqueous milieu (Czabotar et al., 2013), but only transiently because this high energetic state is quickly relaxed to a low energetic one either by burial of these hydrophobic residues in the cytosolic leaflet of the MOM bilayer (step V) or by release of the more hydrophobic α5, a part of the groove that houses the BH3 region, into the membrane (step VI; Figure 5.2). In contrast, the more amphipathic α4 does not insert into the membrane. Instead, α4 adjusts its position in the remnant dimer to form a α2-α3-α4 dimer on the membrane to hide some of the hydrophobic residues from aqueous milieu (step VI; Figures 5.2 and 5.3). These conformational changes would separate α4 from α5, an implication somewhat different than that from the crystallography study, which concluded that α4 and α5 were not separated in the activated mitochondrial Bax (Czabotar et al., 2013). Most likely, α4 and α5 become separated in a fraction of, but not all, Bax proteins when they are activated and bound to mitochondria.

After the departure of α5, the remnant BH3-in-groove dimer rearranges to the helices α2-α3-α4 dimer (step VI), in which the α2 and α3 from one Bax are merged into a single helix that forms an antiparallel coiled-coil dimer on the membrane with its counterpart from the other Bax (Figure 5.3). This rearrangement re-stabilizes this dimer interface on the MOM by shielding the

hydrophobic residues either in the new interface or in the cytosolic leaflet of the bilayer. It is conceivable that the α2-α3-α4 dimer could be formed directly by the Bax monomer that has the α5, α6, and α9 embedded in the MOM, but the α2, α3, and α4 exposed to the cytosol (step IV product) and hence available for dimerization. However, inhibition of the α2-α3-α4 dimerization by the G108E mutation that should not directly disrupt the α2-α3-α4 dimer, but rather the BH3-in-groove dimer (Figure 5.4), suggests that the α2-α3-α4 dimerization depends on, and thus occurs downstream of, the BH3-in-groove dimerization. The MD simulation (Figure S5:4) indicates the α2-α3-α4 dimer structure we show here based on the static modeling represents a cohort of different, yet related, conformations that likely coexist among the membrane-bound Bax complexes.

The α9 from one BH3-in-groove dimer binds its counterpart from the other dimer (step VII and VIII), resulting in an intersected or parallel α9 dimer interface in the membrane (Figures 5.5 and 5.6). As shown by the G108E mutant (Figure 5.7C), the α9 dimerization also depends on the BH3-in-groove dimerization, hence occurring downstream, to link the BH3-in-groove dimers to tetramers and higher order oligomers (Figure 5.8).

Numerous studies suggest that active Bax embeds into the cytosolic monolayer to generate membrane tension that induces lipidic pore in the MOM bilayer (Basañez et al., 2002; Bleicken et al., 2013; Schafer et al., 2009; Terrones et al., 2004). The membrane tension is proportional to the local concentration of Bax that is increased initially by the BH3-in-groove dimerization and further by the α9 dimerization. The membrane tension is also proportional to the regions of Bax that can partition into the cytosolic monolayer, including α9, because α9 only has three helical turns embedded in the MOM (Figure 5.5), too short to span the entire bilayer. Once the lipidic pore is formed, α9 would slide into the rim, where the bilayer is fused to a highly curved

monolayer. The α9 at the rim would reduce the line tension, stabilizing the lipidic pore. The α9 dimerization at the rim might recruit more Bax proteins to the vicinity, expanding the pore. Therefore, both membrane insertion and dimerization of α9 would induce, stabilize, and expand the lipidic pore. As shown by the G179I and T182I mutants (Figures 5.10 and EV5.5E), more α9 insertion and the consequently more BH3-in-groove dimerization could compensate for less α9 dimerization to enlarge the pore. However, at a normal α9 insertion level, we propose that α9 dimerization becomes critical to the pore expansion. Thus, dimerization of α9 is not required for the small pore formation that releases cytochrome c, but facilitates large pore formation that releases Smac and larger proteins. The membrane insertion of α5 may play a similar role as it also has three helical turns embedded in the membrane (Figure 5.2), and thus prefers the highly curved monolayer at the pore rim. In fact, α5 peptide can form a toroidal pore by localizing to the rim (García-Sáez et al., 2007; Qian et al., 2008), although α9 peptide can oligomerize in membranes and may form a barrel-stave proteinous pore (Garg et al., 2013).

We built a model to illustrate the architecture of Bax pore in the MOM (Figure 5.11B). In this model, the BH3-in-groove dimers are placed on the cytosolic surface of the MOM, forming the outer ring of the pore. The α9 dimers are placed at the edge of the bilayer, forming the rim of the pore with lipid head groups in the highly curved monolayer. The α9 dimer of BH3-in-groove dimer is one of the multiple tetrameric units that are assembled into an oligomer. The number of Bax molecules in the oligomer dictates the pore size. In order to transport hydrated cytochrome c across the MOM, the minimal pore diameter must be ~50 Å. This model represents one conformation of a dynamic Bax pore, variations from which are possible. For example, the α5 may leave the BH-in-groove dimer, inserting into the bilayer or joining the α9 at the pore rim,

whereby the remnant of the dimer rearranges to the α2-α3-α4 dimer that remains on the cytosolic surface of the MOM.

The DEER study proposed another conformation for the pore (Bleicken et al., 2014), in which the BH3-in-groove dimer is located at the rim with the hydrophobic side of α4 and α5 embedded in the curved lipid monolayer, while the α9 is inserted into the bilayer farther from the rim and linked to the α5 by α6 located on the surface of the bilayer. Due to the symmetry of the BH3-in-groove dimer, this conformation requires two α6 helices, one from each monomer, to locate on the opposite surfaces of the MOM, and two α9 helices to span the MOM in an antiparallel orientation. While this model fits some of the intra- and intermolecular distances derived from the DEER data, the broad distance distribution indicates a dynamic localization of the α6 and α9 relative to the BH3-in-groove dimer that requires other models to fit. Moreover, it is difficult to imagine how a large part of the Bax molecule stretching from α6 to α9 can cross to the intermembrane space side from the cytosolic side of the MOM where the initial binding occurs to drive the pore formation (Andrews, 2014). Our model, on the other hand, would allow the BH3-in-groove dimer to slide into the rim after pore formation, which might further stabilize and expand the lipidic pore. Moreover, while the localization of α6 on the MOM surface is supported by previous IASD-labeling studies (Annis et al., 2005; Westphal et al., 2014b), the transmembrane topology of α9 requires at least five helical turns to be buried in the bilayer, which is inconsistent with our IASD-labeling data (Figure 5.5) unless α9 also accesses the aqueous interior of the pore or IASD can label the residues located in the bilayer but near the surfaces such as the lipid head group region. Finally, the antiparallel orientation of two α9 helices is at odds with our disulfide-crosslinking data (Figure 5.6) and others' intermolecular FRET imaging data (Gahl et al., 2014) that support a parallel or intersected α9 dimer model.

By monitoring the conformation and interaction of active Bax at mitochondria in a biologically relevant cell-free system supplemented with a cell system, we have demonstrated the formation of two separate dimer interfaces. The BH3-in-groove interface located on the membrane surface is formed earlier than and required for the formation of the α9 interface inside the membrane. Together, both interfaces assemble Bax proteins into oligomers that induce membrane permeabilization by forming lipidic pores. While it appears that the BH3-in-groove dimerization is required for pore formation, α9 dimerization may be primarily involved in pore expansion. The concerted action of this series of protein–protein interactions and the subsequent membrane destabilization results in the release of not only cytochrome c, which induces caspase activation, but also larger proteins like Smac, which neutralizes caspase inhibitors, thereby culminating in a full-scale apoptotic reaction that dismantles the cell. The protein complexes and conformations that we have characterized here may be potential targets for the future development of apoptotic modulators to eradicate cancer or to prevent ischemia reperfusion injury (Brahmbhatt et al., 2015; Czabotar et al., 2014; Moldoveanu et al., 2014).

5.5 Materials and Methods

5.5.1 Reagents

Single-cysteine Bax mutant plasmids for *in vitro* transcription and translation were constructed from pSPUTK vector (Stratagen) as described earlier (Ding et al., 2014). Additional mutations, as described in Results, were introduced into some of the single-cysteine mutants or wild-type Bax using QuickChange mutagenesis (Agilent technologies). The single-cysteine mutants were designated as a letter and a number that indicate the wild-type residue and its position follow by C that indicates the change to a cysteine. The other mutants were similarly

designated. The single-cysteine tBid mutant pladmid was constructed from pSPUTK to include the same 5′ untranslated region as that in the Bax plasmid.

Human Bax protein with or without G179I or T182I mutation, and murine tBid protein were expressed and purified as described elsewhere (Yethon et al., 2003; Zha, 2000; Zhang et al., 2010). The Bax BH3 peptide was synthesized by Abgent as described (Tan et al., 2006).

The rabbit polyclonal antibody to Bid (catalog No. ab77815) and the mouse monoclonal antibody to pyruvate dehydrogenase E1-alpha subunit (catalog No. ab110334) were from Abcam. The rabbit polyclonal antibody to Smac (catalog No. sc-22766) was from Santa Cruz Biotechnology. The rabbit polyclonal antibody to Bax and the sheep polyclonal antibody to cytochrome c were produced in David Andrews's laboratory.

5.5.2 MOMP assays

The ELISA-based cytochrome c release assay was performed as described earlier (Ding et al., 2014), except that we used 2 μl Bax protein synthesized in transcription/translation (TNT)-coupled SP6 RNA polymerase/reticulocyte lysate system (Promega), 17 nM purified tBid protein, and 0.6 mg/ml Bax−/−/Bak−/− mitochondria.

The immunoblot-based cytochrome c and Smac release assay began with incubation of 60 or 120 nM purified Bax protein with 1 mg/ml Bax−/−/Bak−/− mitochondria in the absence or presence of 17 nM purified tBid protein at 37°C for 2 h. The resulting sample was centrifuged at 13,000 g and 4°C for 5 min. The supernatant and pellet were separated and analyzed by reducing SDS–PAGE and immunoblotting with the antibody to cytochrome c (dilution 1:1,500) or to Smac (dilution 1:750).

The fluorescence-based Smac-mCherry release assay was performed as described (Shamas-Din et al., 2014) with 10, 20, or 40 nM purified Bax protein, and where indicated, 2 nM purified caspase-cleaved Bid protein (cBid).

5.5.3 Live cell imaging and analysis

Bax/Bak double-knockout (DKO) baby mouse kidney (BMK) cells were cultured in DMEM supplemented with 10% FBS. For transfections, Bax/Bak DKO cells were seeded at 1,000 cells/ well in a 384-well imaging plate (PerkinElmer Cell Carrier Ultra) and transiently transfected 24 h later with Venus-Bax constructs (ratio of 1:6 DNA:Fugene HD transfection reagent from Promega). Cells were incubated for 14 h post-transfection of Venus-Bax constructs and then treated with or without 50 nM staurosporine (STS) for 8 h. Cells were stained 30 min prior to imaging with 5 μM Draq5 (BioStatus) and either 10 nM TMRE (Invitrogen) for activity assays or 100 nM Mitotracker Red (ThermoFischer) for colocalization studies. Cells were imaged using the Opera high-content screening system (PerkinElmer) with either a 20× air objective (for activity assays) or a 40× water immersion objective (for colocalization studies). Three channels were captured: Draq5, Venus, and TMRE or Mitotracker Red. Images were analyzed as in (Shamas-Din et al., 2013). Briefly, a script written for Acapella high-content imaging and analysis software (available for download from dwalab.ca) was used to segment and detect only Venus-positive cells. The same script was used to extract intensity and morphology features for each segmented cell as well as Pearson's correlation coefficient (for colocalization experiments). Dead cells were classified as small rounded up TMRE-negative cells. The percentage of apoptosis was calculated by taking total number of Venus-positive cells classified as dead and dividing by the total number of Venus-positive cells.

To minimize the effects of subjective bias in sample allocation and result analysis, a fully automated microscope was used to image cells based on a predetermined pattern of locations within the wells of an imaging plate. Therefore, the selection of cells analyzed was unbiased. Images acquired by the automated microscope were then analyzed numerically based on automated image analysis of computer-measured features in the images rather than a subjective interpretation.

Justifications for the statistical test used are that points are generally averages of ~150 randomly selected automatically analyzed cell images from three biological replicates. Standard deviation is reported to provide error bars on the histograms. However, to determine significance of differences between mutants, one-way ANOVA was used and determined that the means are not identical. *Post-hoc* multiple comparisons were made using a *t*-test with a *P*-value < 0.05 being significant. To correct for type I error, the Bonferroni correction was applied to the *P*-value (0.05/24(number of measurements)), and any *P*-value < 0.002083 is deemed significant. For expression levels, *post-hoc* multiple comparisons were made using a *t*-test with a *P*-value < 0.05 being significant. To correct for type I error, the Bonferroni correction was applied to the *P*-value (0.05/12(number of measurements)), and any *P*-value < 0.00417 is deemed significant. *t*-tests were used to compare the Venus intensity of each mutant construct to that of WT Bax in order to compare expression levels. Raw data are plotted and means were compared to medians. Data is close enough to normal distribution that standard deviation is a reasonable estimate.

5.5.4 Disulfide crosslinking

Disulfide crosslinking of mitochondria-bound Bax proteins was performed and analyzed as described (Ding et al., 2014) with the following modifications. We incubated 15 µl *in vitro* synthesized [35S]Met-labeled Bax protein, 17 µM Bax BH3 peptide, and 0.6 mg/ml

Bax−/−/Bak−/− mitochondria at 37°C for 1 h. The mitochondria with activated Bax were pelleted by centrifugation at 13,000 g and 4°C for 5 min. The mitochondria were resuspended in high salt buffer (400 mM KOAc, 4 mM Mg(OAc)2, and 25 mM HEPES, pH 7.5) to remove loosely bound proteins before adding redox catalyst CuPhe to induce disulfide crosslinking. For some of the mitochondrial samples, low salt buffer (110 mM KOAc, 1 mM Mg(OAc)2, and 25 mM HEPES, pH 7.5) was used, and the crosslinking results were similar to the “high salt” controls. The disulfide-crosslinked samples were precipitated by Cl3CCOOH. The resulting protein pellets were resuspended in quenching buffer (20 mM N-ethylmaleimide (NEM), 100 mM EDTA, 1 mM NaAsO2), solubilized in non-reducing SDS sample buffer (12.5% SDS, 500 mM Tris, 25% glycerol, 250 mM EDTA, and a trace of bromophenol blue) at 65°C for 30 min, and analyzed using SDS–15% PAGE and phosphorimaging.

Some of the disulfide-crosslinked samples were treated with Na2CO3 (pH 11.5) as described earlier (Billen et al., 2008), except that DTT was omitted from the 200-mM Na2CO3 solution and 0.5-M sucrose cushion. The resulting pellet and supernatant fractions were analyzed by SDS–PAGE and phosphorimaging as describe above, or by immunoblotting with mouse monoclonal antibody to pyruvate dehydrogenase E1-alpha subunit (dilution 1:2,500).

5.5.5 Immunoprecipitation

To identify the disulfide-linked products from the reactions containing single-cysteine Bax and tBid mutants, the samples were divided into two equal portions, and diluted approximately 1:7 in IP buffer containing 100 mM Tris pH 8.0, 100 mM NaCl, 10 mM EDTA, 1 mM PMSF, 1% (v/v) Triton X-100 and incubated with Bax (dilution 1:100) or Bid (dilution 1:1,600) antibody. Reactions were rotated overnight at 4°C. The resulting immuno-complexes were bound to Sepharose protein G beads (6–8 μl of 50% bead suspension in IP buffer) for 2 h at 4°C. After

centrifugation at 2,000 *g* for 30 s at 4°C, the beads were washed three times with IP buffer and twice with 100 mM Tris pH 8.0 and 100 mM NaCl. The bound proteins were eluted from the beads with non-reducing SDS sample buffer and analyzed by non-reducing SDS–PAGE and phosphorimaging.

5.5.6 4-acetamido-4′-((iodoacetyl)amino)stilbene-2,2′-disulfonic acid (IASD) labeling

IASD labeling of mitochondria-bound Bax proteins was performed and analyzed as described (Ding et al., 2014) with the following modifications. We incubated 2.5 µl Bax protein synthesized *in vitro* with 17 µM Bax BH3 peptide or 17 nM tBid protein, and 0.8 mg/ml Bax−/−/ Bak−/− mitochondria at 37°C for 1 h. The mitochondria with activated Bax were pelleted by centrifugation as described above, and resuspended in medium salt buffer (200 mM KOAc, 2 mM Mg(OAc)2, and 10 mM HEPES, pH 7.5) to increase IASD-labeling efficiency (Westphal et al., 2014b). For some mitochondrial samples, low salt buffer was used, and the IASD-labeling results were similar to the “medium salt” controls. The mitochondrial suspension was split equally to five, and 0.5 mM IASD was added to each aliquot alongside 230 mM Tris–HCl, pH 7.5, 100 ng/ml each of leupeptin, antipain, chymostatin, and pepstatin, 150 ng/ml aprotinin, 0.05 mM DTT, and, if indicated, 2% 3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonate (CHAPS), 4 M urea, or both. The labeling reactions were carried out in dark at 22°C for 30 min and then quenched by 50 mM 2-mercaptoethanol. For the “0 min” labeling control, 2-mercaptoethanol was added prior to the addition of IASD to prevent the labeling. After the concentration of CHAPS and urea in each sample was brought to 2% and 4 M, respectively, the proteins were precipitated by Cl3CCOOH.

For reducing SDS-gradient (15–16%) PAGE analysis, the protein pellet was solubilized at 65°C for 30 min in the SDS sample buffer plus 1 M 2-mercaptoethanol. For IEF analysis, the protein pellet was sonicated at 22°C for 20 min in IEF sample buffer (10 mM Tris–HCl, pH 7.5, 8 M urea, 4% CHAPS, 4% Tergitol NP-40, 0.4% Triton X-100, 30% glycerol, 2% 5–8 ampholyte, 1 M 2-mercaptoethanol, and a trace of bromophenol blue), loaded on a pH 5–8 gel (1.5% 5–8 ampholyte, 0.5% 3–10 ampholyte, 7% acrylamide (acrylamide:methylenebisacrylamide = 37.5:1), 7 M urea, 2% CHAPS, 4% Tergitol NP-40, 0.08% tetramethylethylenediamine, and 0.09% ammonium persulfate, pre-run at 100 V for 30 min), and run with anode buffer (50 mM H3PO4, pH ~4) and cathode buffer (20 mM lysine, 20 mM arginine, pH ~10) at 100 V for 1 h, 250 V for 1.5 h, and 400 V for 14 h 45 min. The radioactive proteins in the resulting gels were detected by phosphorimaging, and quantified using Multi Gauge program (Fujifilm) as described in Figure S5:3A.

5.5.7 Computational modeling of Bax dimer interfaces

α2-α3-α4 dimer interface

We modeled this dimer based on the published structure of Bax monomer (PDB entry: 1F16) and the distance restraints provided by the disulfide crosslinking between the specific interfacial residues after they were replaced by cysteines (Figure 5.3A). Briefly, the partial Bax structure formed by residues 54–103 including α2, α3, and α4 helices and the inter-helical loops were extracted from the whole structure. The loops were removed and a global search was performed to optimize the distance restraints and the protein geometry according to the following procedures.

First, helices α2 (residues 54–71), α3 (residues 75–81), and α4 (residues 88–99) were converted to poly-alanine and allowed to independently move in space as rigid bodies, using a

Monte Carlo procedure to minimize the energy of springs functions introduced to represent the linkages from crosslinking experiments. The springs were places between the Cβ atoms of each crosslinked position in the two chains (S55-R94′, L59-L76′, S62-S72′, and E69-E69′) using a spring constant of 200 kcal/mol/Å^2 and an energy minimum distance of 5 Å. Spring penalties were not calculated for distances of the crosslinked Cβ atoms below 5 Å. In addition, the helical termini of α2 and α3, and of α3 and α4 were also constrained to remain within 6 Å of each other. Once these distance constraints were minimized, the original side chains were reintroduced and the same procedure was continued—except that each cycle included a round of side chain optimization—until the energy converged. The α2, α3, and α4 helices were then connected together by the residues with favorable geometry using fragments extracted from the PDB database according to a procedure previously described (LaPointe et al., 2013). Briefly, protein fragments with pattern *hhha$_x$hhh* were extracted from high-resolution X-ray structures, where *h* are positions in a helical conformation and a_x are positions in any conformation corresponding to the number needed for the loop connecting the helices (x = 3 for α2-α3 loop; x = 6 for the α3-α4 loop). The termini of the inter-helical fragments were then aligned with the termini of the helices to be connected to obtain favorable geometries and the helices and fragments were fused. Only the N, C, CA, and O atoms were considered for the alignment and the fragments with the lowest root-mean-square deviation (RMSD) were selected. The side chains on the fragment were replaced with the one corresponding to the native Bax sequence and their conformation was optimized.

The procedure yielded a well-packed model with favorable energies, in which a distorted helical region connected helices α2 and α3. To investigate whether the geometry of the α2-α3 region was compatible with canonical helical conformation, the procedure was repeated as

described except that helices α2 and α3 were fused into a single canonical helical fragment. This second computation produced a model closely related to the original that satisfied the distance constraints in similar fashion but had slightly lower energies. This model is illustrated in Figure 5.3D.

In the procedure, the energies were calculated using the CHARMM 22 van der Waals' function (MacKerell et al., 1998) and the hydrogen bonding function of SCWRL 4 (Krivov et al., 2009), as implemented in MSL C++ libraries (Kulp et al., 2012). All side chain optimization procedures were performed using the Energy-Based Conformer Library applied at the 95% level (Subramaniam and Senes, 2012, 2014) with a greedy trials algorithm (Xiang and Honig, 2001) as implemented in MSL.

α9 dimer interface

Two models were predicted for the membrane-embedded α9 dimer. The first model was predicted *ab initio* by CATM program (Mueller et al., 2014) from the sequence of α9, $G^{166}TP^{168}TWQTVTIFVAGVLTASLTIWKKM^{191}$, except that Pro^{168} was replaced by alanine to increase the helical propensity. The program predicted a homo-dimeric structure adopting a GAS_{right} conformation (Walters and DeGrado, 2006) mediated by the GxxxA sequence motif (Russ and Engelman, 2000; Senes et al., 2000). The resulting intersected α9 dimer model (Figure 5.6A) fits the disulfide-crosslinking data from four single-cysteine Bax mutants (Figure 5.6B).

The second model (the parallel α9 dimer) was obtained through a systematic search of the dimer conformational space directed by the disulfide-crosslinking data from the other four single-cysteine Bax mutants (Figure 5.6D). The search program generated two standard helices for α9 and performed a systematic search of the space of all possible C2 symmetrical dimer geometries, varying four degrees of freedom, (i) the rotation around the helical axis, (ii) the inter-

helical crossing angle, (iii) the position of crossing point along the axis of symmetry, and (iv) the inter-helical distance. The distance between the β-carbon atoms of Q171, A178, T182, and L185 was evaluated for each geometry. If all four distances fell below a 6 Å, the side chains were optimized. The energy of association of this optimized dimer was computed as the difference between the energy of the dimer and the energy of a side chain optimized monomer, and all dimers with a favorable energy of association were stored. From this set of structures, a structure with good packing was selected by visual inspection (Figure 5.6F).

Disruptive mutations were predicted from both α9 dimer models by a computational mutagenesis protocol. Each position was mutated into glycine, alanine, leucine, valine, isoleucine, and phenylalanine residues. For each mutation, side chains were optimized and the energy of association was computed for the dimer using hydrogen bonding and van der Waals' functions as reported previously (LaPointe et al., 2013). Each mutation was classified into one of four categories based on the difference in the energy (ΔE) between the wild-type and the mutated dimer. The classification is as follows: (i) silent: $\Delta E \leq 2$ units, (ii) significant: $2 < \Delta E \leq 4$ units, (iii) mildly disruptive: $4 < \Delta E \leq 8$ units, and (iv) disruptive: $\Delta E > 8$ units. An average score was computed for each position, and used to identify disruptive mutations. All programs were implemented using the MSL molecular modeling software library (Kulp et al., 2012).

5.6 Supplementary Methods

5.6.1 Molecular dynamics simulations

All molecular dynamics simulations were performed with the NAMD package (Phillips et al., 2005) with CHARMM force field (Best et al., 2012; Klauda et al., 2010): CHARMM36 for the lipid parameters, CHARMM27-cmap for the rest of the systems. Both equilibration and

production runs were performed in the NPT ensemble (310 K, 1 bar, Nose-Hoover coupling scheme) with a time step of 2 fs. The particle mesh Ewald technique was used for the electrostatic calculations. The van der Waals and short-range electrostatic interactions were cut off at 12.0 Å with switch at 10.0 Å. CHARMM-GUI (Jo et al., 2008) was used to build the protein-membrane and protein-solvent systems. The lipid bilayer consist of the MOM characteristic phospholipids (Kuwana et al., 2002), including 46.5% L-α-phosphatidylcholine (PC), 28.4% L-α-phosphatidylethanolamine (PE), 8.9% L-α-phosphatidylinositol (PI), 8.9% L-phosphatidylserine (PS) and 7.3% cardiolipin. All systems were solvated with explicit water in TIP3P model in a periodic box. Counterions (Na+ and Cl-) were added to the systems. The system setups are listed in Table S5.1. We have performed two equivalent metadynamics simulations (Iannuzzi et al., 2003; Laio and Parrinello, 2002), each in 30 ns, with NAMD package, for free energy change from the intersected to parallel α9 dimer, with the assumption that the transition between two conformations is reversible. The sampling bias was applied to two collective variables – the angles of rotation (α, β) around the principle axis of the monomeric helix. The results from the two replicates are consistent.

5.6.2 IASD Quantification

The IASD labeling data from Bax L59C were shown to illustrate quantification procedures using the Multi Gauge program (Figure S3). To create a quantifying region, a rectangle was drawn to enclose a typical band, such as the rectangle c that encloses the IASD-labeled band in lane 3. Rectangles of the same size (a, b, d-j) were then used to enclose the other nine bands (labeled and unlabeled) in lanes 1-5, and to enclose a region in which there is no band (rectangle-k), creating other quantifying regions. The amount of photo-stimulated luminescence (PSL) from each of the 11 quantifying regions was measured. The PSL value is in proportion to the amount

of radiation emitted from the radioactive protein band and exposed to phosphor-imaging plate. To correct for the uneven sample recovery and loading, the total PSL from the labeled and unlabeled bands in each lane was divided by the total PSL from the labeled and unlabeled bands in lane 1 to obtain the normalization factor for each sample. The PSL from each of the 10 labeled and unlabeled bands was normalized by the corresponding normalization factor. Using the IASD-labeling sample in lane 2 as an example,

$$NF = \frac{PSL(b) + PSL(g)}{PSL(a) + PSL(f)} \qquad \text{Equation (13)}$$

$$PSL_{norm}(b) = \frac{PSL(b)}{(NF)} \qquad \text{Equation (14)}$$

$$PSL_{norm}(g) = \frac{PSL(g)}{(NF)} \qquad \text{Equation (15)}$$

To calculate the net PSL of each of the IASD-labeled bands, subtract the normalized PSL of the labeled band in lane 1 from the normalized PSL of the corresponding labeled band. This is based on an assumption that the IASD labeling should not occur in the sample in lane 1, as IASD should be quenched by β-mecaptoethanol that was already present in the sample before the addition of IASD. Any PSL from rectangle-a would result from an IASD-independent modification of the Bax protein that would also occur in the other samples, and hence, must be subtracted. To calculate the net PSL of each of the unlabeled bands, subtract the PSL of rectangle-k from the normalized PSL of the corresponding unlabeled band. This is because there is no band in rectangle-k; any PSL from there must be from the background that would also be present in other rectangles enclosing the unlabeled bands, and hence, must be deducted. Using the IASD-labeling sample in lane 2 as an example,

$$PSL_{net}(b) = PSL_{norm}(b) - PSL_{norm}(a) \quad \text{Equation (16)}$$

$$PSL_{net}(g) = PSL_{norm}(b) - PSL(k) \quad \text{Equation (17)}$$

To calculate the fraction of IASD-labeling, divide the net PSL of a labeled band by the sum of the net PSL's of the labeled band and the corresponding unlabeled band. Using the IASD-labeling sample in lane 2 as an example,

$$F_{IASD} = \frac{PSL_{net}(b)}{PSL_{net}(b) + PSL_{net}(g)} \quad \text{Equation (18)}$$

To calculate the membrane burial index, subtract the fraction of IASD-labeling from the reaction performed in the presence of urea (lane 4) from the fraction of IASD-labeling from the reaction performed in the presence of both urea and CHAPS (lane 5). This is because the urea in both samples should denature the protein or the protein complex, exposing the protein-buried cysteine to IASD. The difference between the fractions of IASD-labeling from the two reactions must be due to the CHAPS in the second reaction that solubilizes the MOM, exposing the MOM-buried cysteine to IASD. Thus,

$$MBI_{Bax\,L59C} = F_{IASD}(5) - F_{IASD}(4) \quad \text{Equation (19)}$$

To calculate the protein burial index, subtract the fraction of IASD-labeling from the reaction performed in the presence of CHAPS (lane 3) from the fraction of IASD- labeling from the reaction performed in the presence of both urea and CHAPS (lane 5). This is because CHAPS in both samples should solubilize the MOM, exposing the MOM-buried cysteine to IASD. The difference between the fractions of IASD labeling from the two reactions must be due to the urea in the second reaction that denatures the protein or the protein complex, exposing the protein-buried cysteine to IASD. Thus,

$$PBI_{Bax\,L\,59\,C} = F_{IASD}(5) - F_{IASD}(3)$$ Equation (20)

5.7 Author contributions

JLin and DWA conceptualized the study, designed experiments, analyzed data, and wrote the paper. ZZ, JK, HB, SML, JD, and FH conducted experiments and analyzed data. SS, BH, SGFC, FAH, XCZ, and AS performed computational modeling and edited the paper. CL and JLi performed molecular dynamics simulations and edited the paper.

5.8 Conflict of interest

The authors declare that they have no conflict of interest.

5.9 Supplementary Figures and Tables

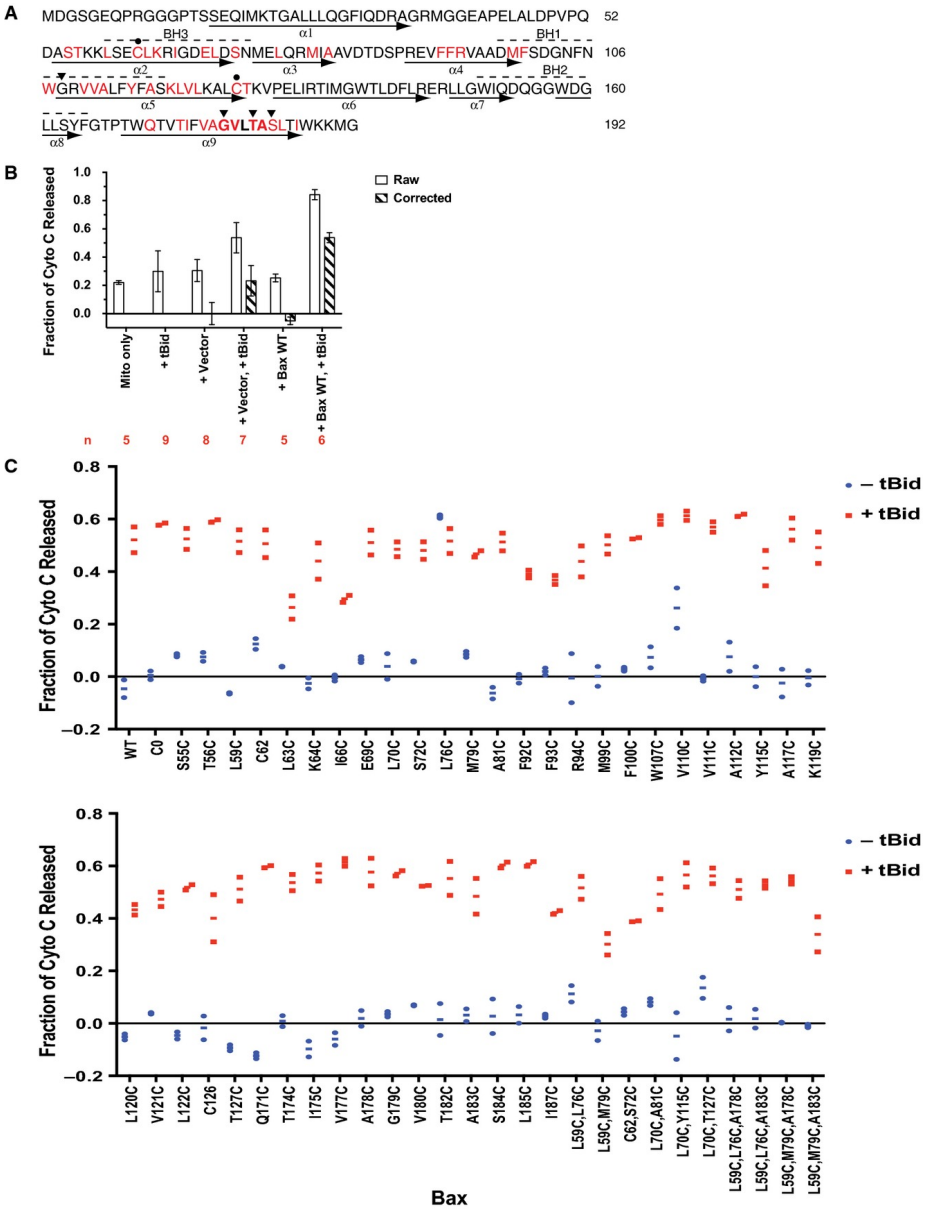

Figure EV5.1: Sequence and in vitro MOMP activity of Bax mutants

A) Bax sequence is shown with BH motifs highlighted by dashed lines above and helices identified by arrows below. The native cysteines (closed circle-indicated

Cs) were changed to alanine to create the cysteine-null mutant. Single-cysteine Bax mutants were created from the cysteine-null mutant by individually replacing the red-colored residues with cysteine. Triangles indicate Gly^{108}, Gly^{179}, Thr^{182}, and Ser^{184} of Bax that were changed to glutamate, isoleucine, isoleucine, and valine in the G108E, G179I, T182I, and S184V mutants, respectively. The GxxxA (GVLTA) motif in α9 is in bold font.

B) The cytochrome c release from the $Bax^{-/-}$/$Bak^{-/-}$ mitochondria-only sample (Mito only), and from the mitochondrial plus purified recombinant tBid protein (+tBid), the TNT reaction programmed by pSPUTK vector (+Vector) or wild-type Bax gene inserted after SP6 promoter in the vector (+Bax WT), or their combinations was measured in *n* independent replicates as indicated. The raw data shown as open bars are the means with the standard deviations (s.d.). A background release ˜20% was observed in the mitochondria-only sample, which might be due to the mitochondria that were frozen and thawed once before used in the assay according to an established protocol (Yamaguchi et al., 2007). Addition of bacterial expressed and purified tBid protein increased the release slightly. The TNT mixture containing the vector plasmid, or the WT Bax plasmid that produced the WT Bax protein (see Figure S5:1) also increased the release slightly. While addition of both tBid protein and the vector-containing TNT mixture showed a marginally additive release, addition of both tBid protein and the WT Bax protein-producing TNT mixture resulted in a synergistic increase in cytochrome c release. After the “raw”

release of “+Vector” control was subtracted from the “raw” releases of “+Vector, +tBid”, “+Bax WT”, and “+Bax WT, +tBid” samples, which all contained the vector (with or without the WT Bax coding region) and the TNT mixture, the “corrected” cytochrome releases were obtained and shown as hatched bars for the means ± s.d.

C) Cytochrome c release by the wild-type (WT), cysteine-null (C0), or the indicated single-, double-, or triple-cysteine Bax mutants synthesized in the TNT reaction in the absence or presence of the tBid protein was measured in two independent replicates. The raw data were corrected as described in (B). The corrected data are shown as the dots with the lines for the averages. The amount of Bax protein that was bound to the mitochondria in each cytochrome c release assay above was determined and the results are shown in Figure S5:1.

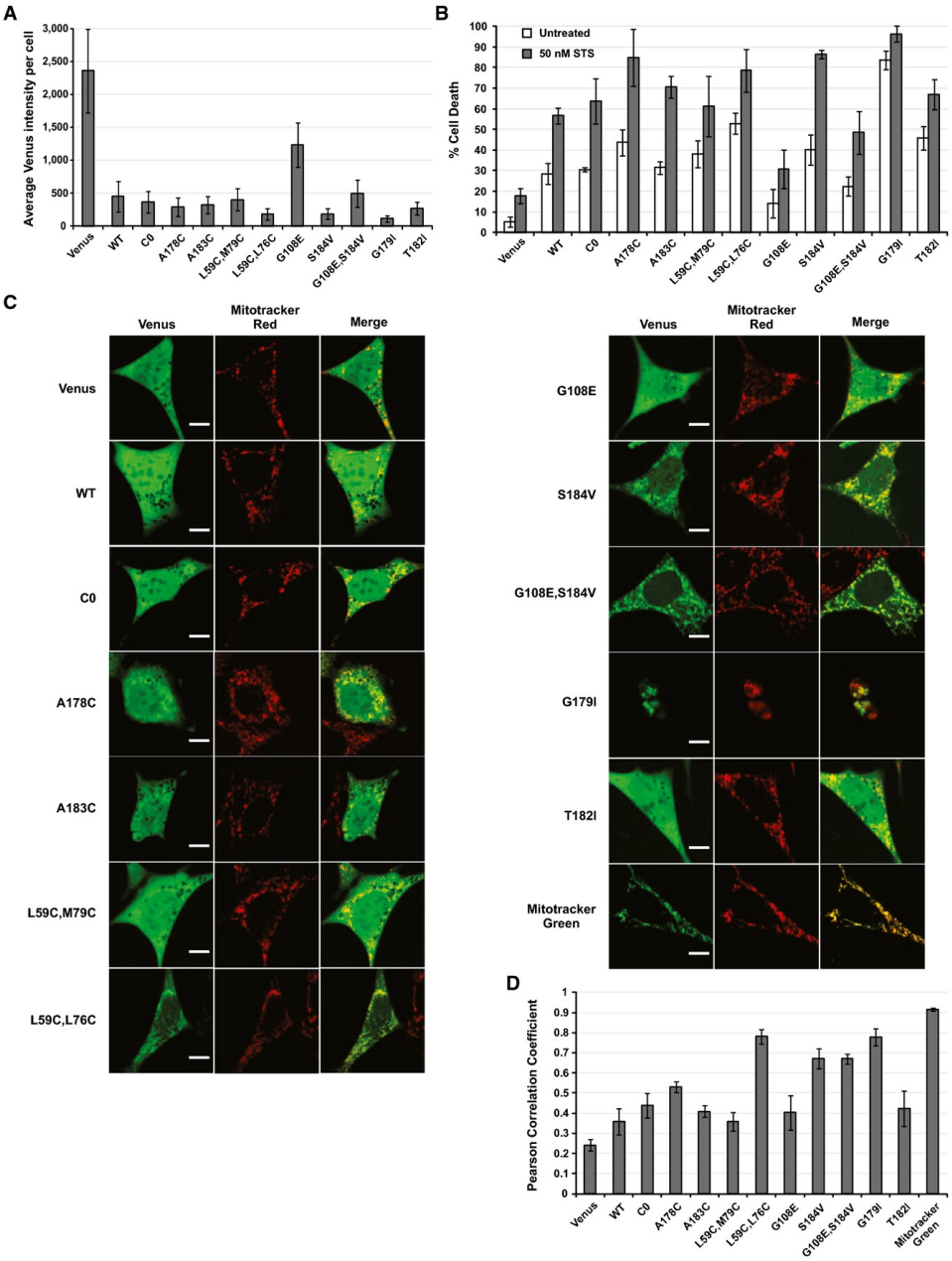

Figure EV5.2: Intracellular localization and apoptotic activity of Bax mutants

A) Expression of Venus-Bax fusion proteins. The average Venus intensity per cell was determined as an estimate of relative Bax expression.

B) Cell death induced by Venus-Bax constructs. The cells were treated with or without STS, imaged, and analyzed. Only Venus-positive cells were used to calculate the percent cell death using an automated image analysis routine. Small rounded up TMRE-negative cells were automatically classified as dead.

C) Colocalization of Venus-Bax constructs with mitochondria. Representative fluorescence images of cells show the localization of Venus-Bax proteins expressed in cells and were compared to the localization of Mitrotracker Red. Scale bar represents 10 μm.

D) Quantification of mitochondrial localization of Venus-Bax constructs. Pearson's correlation coefficient (PCC) was calculated using the Venus and Mitrotracker channels for each Venus-positive cell and averaged. Images of cells expressing Venus only were used as a control representing the PCC for a diffuse localization, and cells stained with Mitotracker Green were used as a control representing the PCC for the highest attainable mitochondrial colocalization.

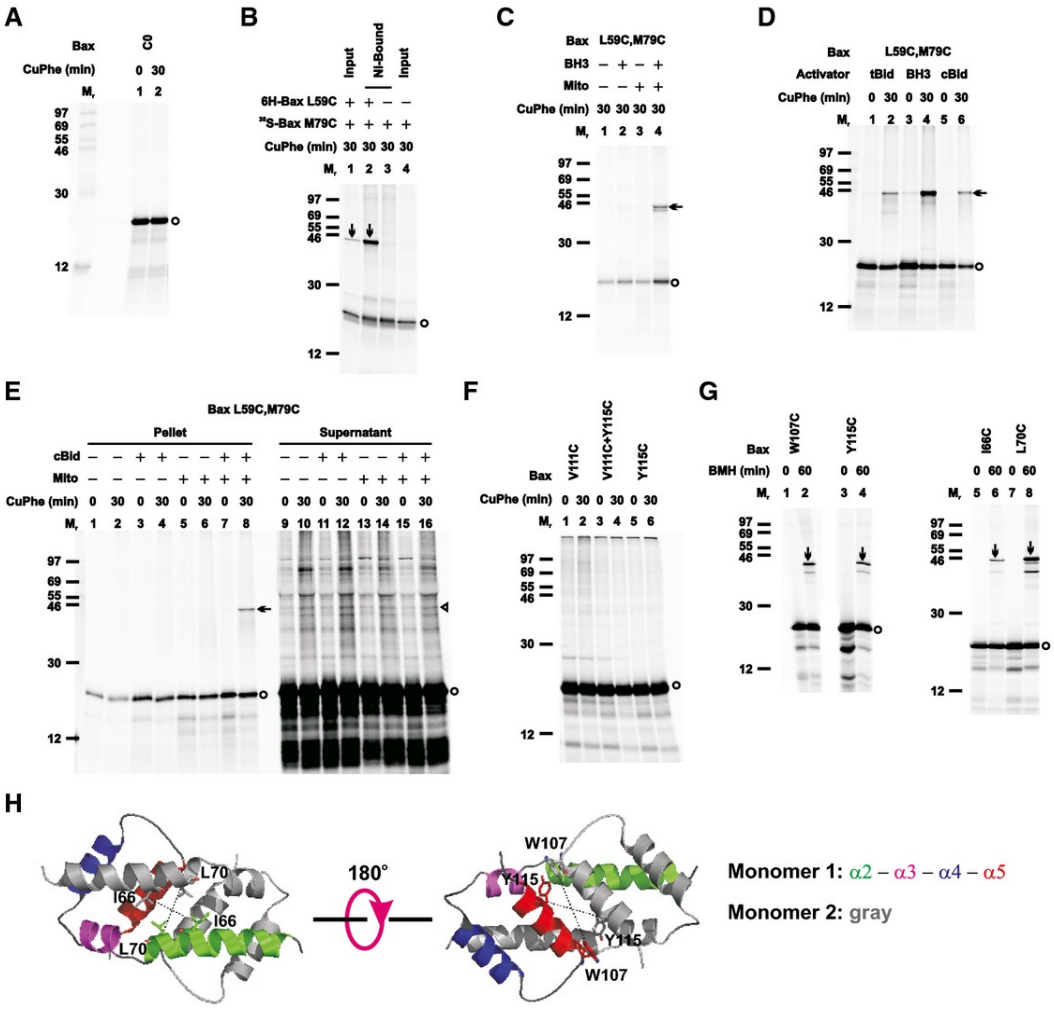

Figure EV5.3: Control disulfide-crosslinking data for the BH3-in-groove dimer interface, related to Figure 5.1

A) The oxidized mitochondria with radioactive cysteine-null Bax protein (C0) were prepared and analyzed as in Figure 5.1B.

B) The *in vitro* synthesized [^{35}S]Met-labeled single-cysteine Bax protein (^{35}S-Bax M79C) was activated and targeted to the Bax$^{-/-}$/Bak$^{-/-}$ mitochondria in the absence or presence of the *in vitro* synthesized non-radioactive hexahistidine-

tagged single-cysteine Bax protein (6H-Bax L59C). The resulting mitochondria were oxidized, and an aliquot (4 µl) was saved as input. The remaining sample (56 µl) was solubilized with buffer A (the high salt buffer plus 1% Triton X-100 and 2.5 mM imidazole), and incubated with 15 µl of 50% Ni-NTA agarose suspension. After washing the beads with the buffer A and phosphate-buffered saline (pH 7.5), the Ni-bound proteins were eluted with the non-reducing SDS sample buffer plus the quenching buffer, and analyzed alongside with the input, as in Figure 5.1B.

C) The radioactive double-cysteine Bax protein was incubated in the absence or presence of Bax BH3 peptide (BH3), the $Bax^{-/-}/Bak^{-/-}$ mitochondria (mito), or both. The "mitochondrial" fraction was isolated, oxidized, and analyzed as in Figure 5.1B.

D) The radioactive double-cysteine Bax protein was activated by Bax BH3 peptide (BH3), tBid or cBid protein, targeted to the mitochondria, oxidized, and analyzed as in Figure 5.1B.

E) The radioactive double-cysteine Bax protein was incubated in the absence or presence of cBid protein, the $Bax^{-/-}/Bak^{-/-}$ mitochondria (mito), or both. The "mitochondrial" pellet was separated from the "post-mitochondrial" supernatant. Both fractions were oxidized and analyzed as in Figure 5.1B.

F) The oxidized mitochondria with the radioactive single-cysteine Bax protein pair were prepared and analyzed as in Figure 5.1B.

G) The mitochondria with the radioactive single-cysteine Bax proteins were prepared as in Figure 5.1B, and crosslinked by 100 μM bismaleimidohexane (BMH). After 60 min, the reactions were stopped by the addition of 50 mM mercaptoethanol. The products were analyzed by reducing SDS–PAGE and phosphorimaging. For the "0 min" control, mercaptoethanol was added before the addition of BMH.

H) The structure of the BH3-in-groove dimer is shown with the residue pairs replaced with cysteine pairs in (G) presented in stick form, and their β-carbon atoms linked by dashed lines with the distances ranging from 8.5 to 18.1 Å.

Data information: In (A–G), the protein standards are indicated on the side of phosphor images by their M_r, the Bax monomers by open circles, the disulfide- or BMH-linked Bax dimers by arrows, and non-disulfide-linked products in the supernatant of (E) that have the same M_r as the disulfide-linked Bax dimer in the pellet are indicated by open triangle. $n = 2$.

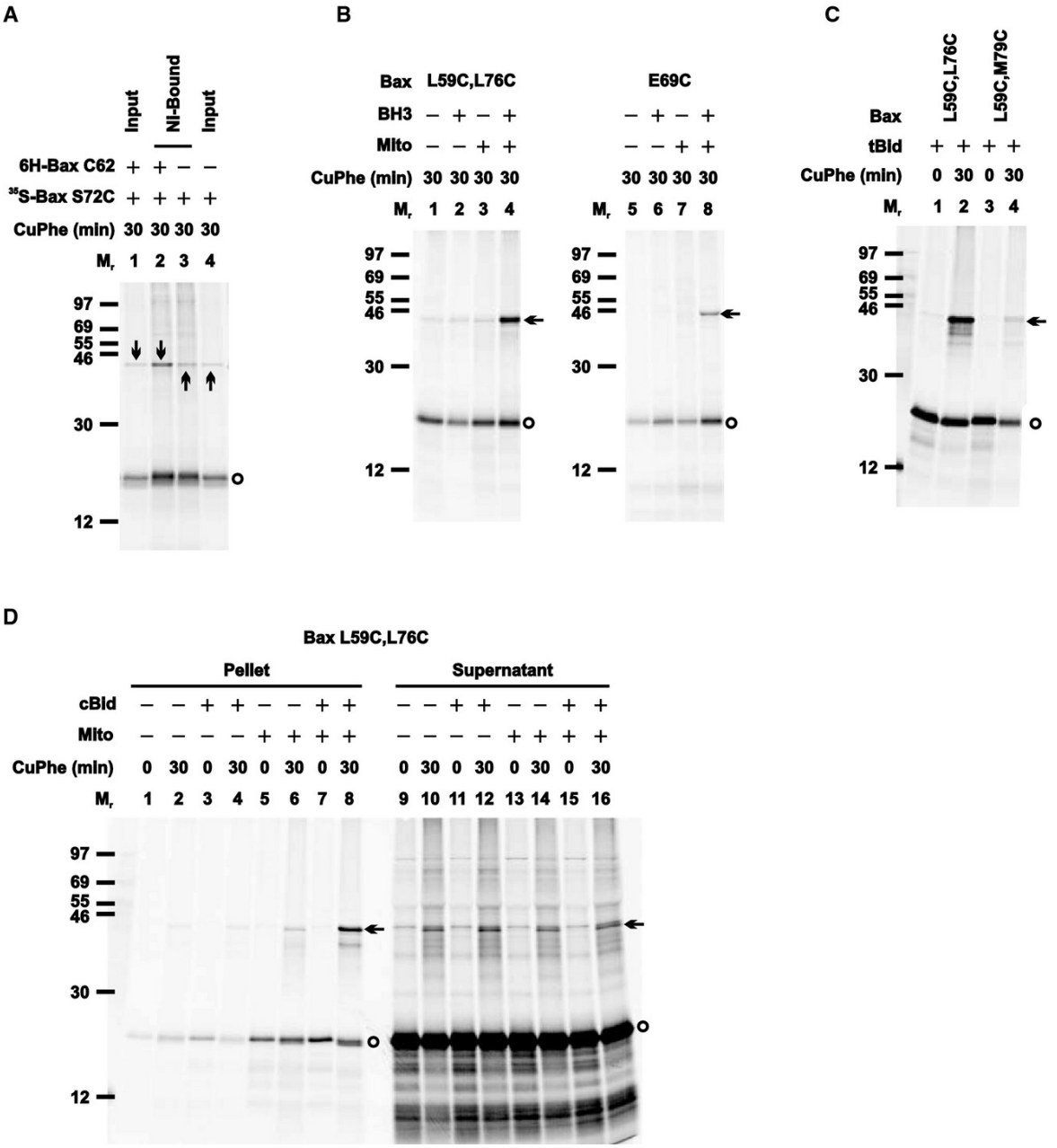

Figure EV5.4: Control disulfide-crosslinking data for the helices α2-α3-α4 dimer interface, related to Figure 5.3

The Ni-bound proteins from the oxidized mitochondria containing the non-radioactive 6H-Bax C62 and/or the radioactive ^{35}S-Bax S72C were prepared and analyzed alongside with the inputs, as in Figure EV5.3B.

A) The radioactive single- or double-cysteine Bax protein was incubated in the absence or presence of Bax BH3 peptide, the $Bax^{-/-}/Bak^{-/-}$ mitochondria, or both. The "mitochondrial" fraction was isolated, oxidized, and analyzed as in Figure 5.1B.

B) The radioactive double-cysteine Bax protein was activated by tBid protein, targeted to the mitochondria, oxidized and analyzed as in Figure 5.1B.

C) The radioactive double-cysteine Bax protein was incubated in the absence or presence of cBid protein, the $Bax^{-/-}/Bak^{-/-}$ mitochondria (mito), or both. The "mitochondrial" pellet was separated from the "post-mitochondrial" supernatant. Both fractions were oxidized and analyzed as in Figure 5.1B.

Data information: In (A–D), the protein standards are indicated on the side of phosphor images by their M_r, the Bax monomers by open circles, and the disulfide-linked Bax dimers by arrows. $n = 2$.

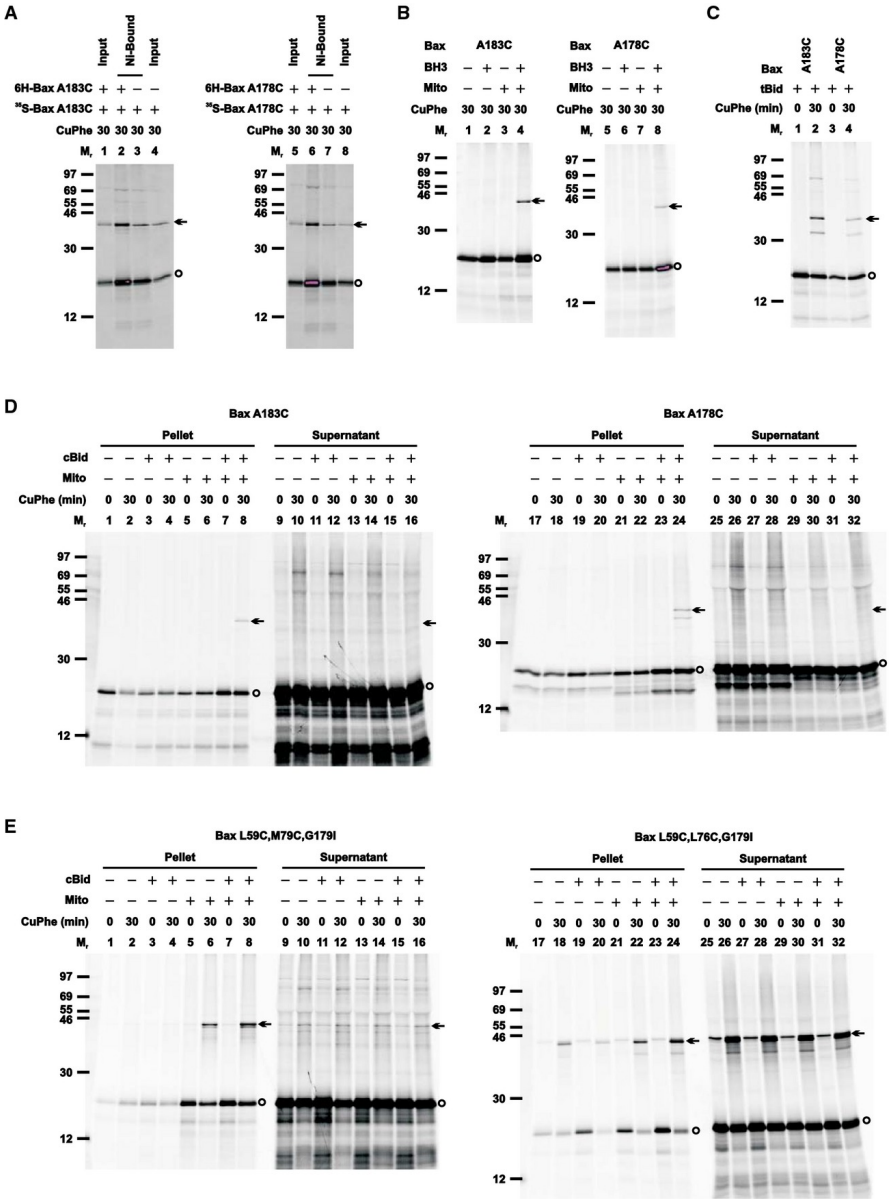

Figure EV5.5: Control disulfide-crosslinking data for the helix α9 dimer interface, related to Figure 5.6; and for the G179I mutants, related to Figure 5.10

A) The Ni-bound proteins from the oxidized mitochondria containing the non-radioactive 6H-tagged and/or the [^{35}S]Met-labeled single-cysteine Bax protein were prepared and analyzed alongside with the inputs, as in Figure EV5.3B.

B) B. The radioactive single-cysteine Bax protein was incubated in the absence or presence of Bax BH3 peptide, the Bax$^{-/-}$/Bak$^{-/-}$ mitochondria, or both. The "mitochondrial" fraction was isolated, oxidized, and analyzed as in Figure 5.1B.

C) C. The radioactive single-cysteine Bax protein was activated by tBid protein, targeted to the mitochondria, oxidized, and analyzed as in Figure 5.1B.

D) The radioactive single-cysteine Bax protein was incubated in the absence or presence of cBid protein, the Bax$^{-/-}$/Bak$^{-/-}$ mitochondria (mito), or both. The "mitochondrial" pellet was separated from the "post-mitochondrial" supernatant. Both fractions were oxidized and analyzed as in Figure 5.1B.

E) Same as Figure EV5.5D but with the double-cysteine Bax protein with G179I mutation

Data information: In (A–E), the protein standards are indicated on the side of phosphor images by their M_r, the Bax monomers by open circles, and the disulfide-linked Bax dimers by arrows. $n = 2$.

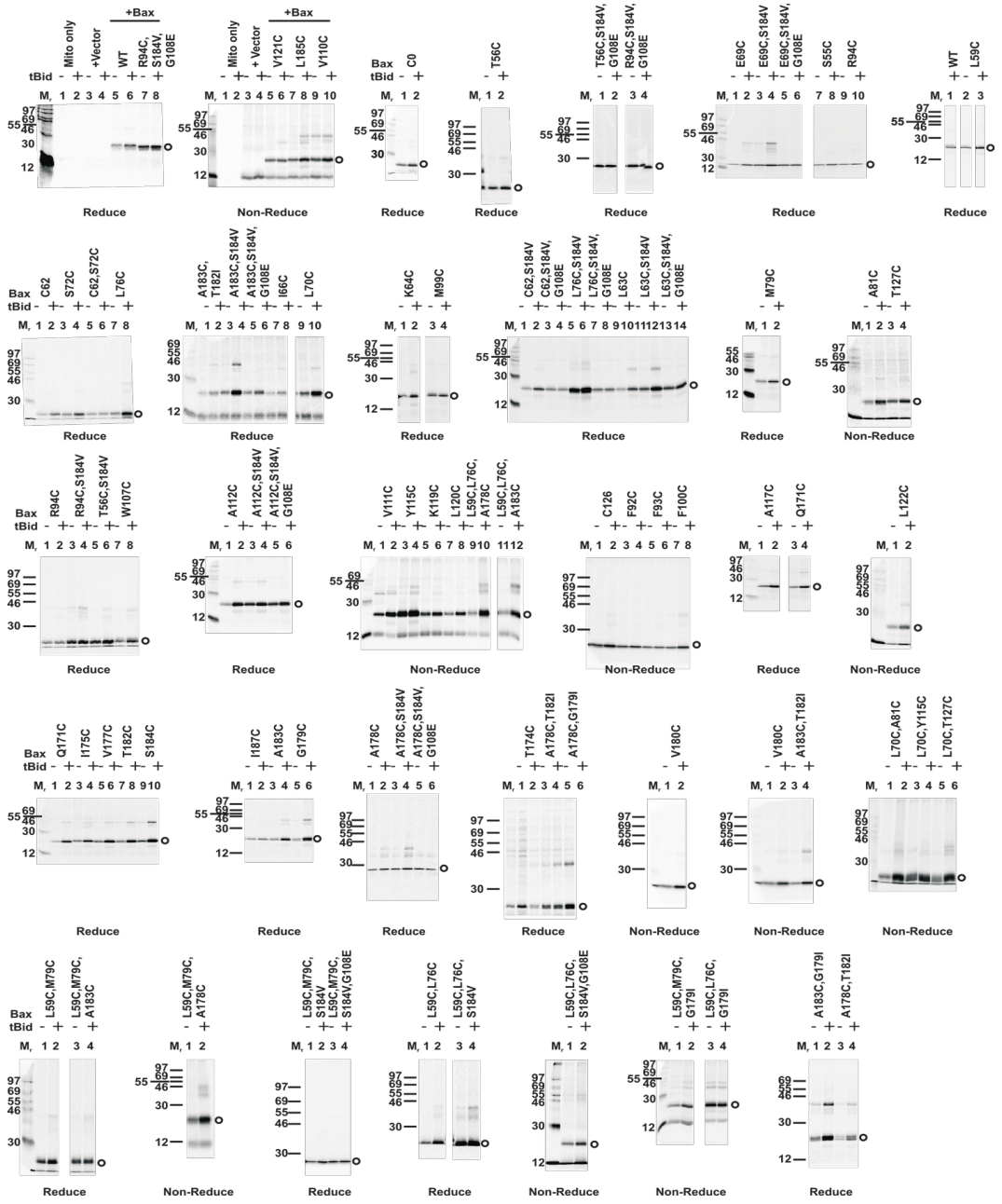

Figure S5:1: Mitochondria-bound Bax proteins from the cytochrome c release assay.

Aliquots of the mitochondrial pellet fractions from the cytochrome c release assay were analyzed by reducing or non-reducing SDS-PAGE. The mitochondria-bound radioactive wild-type Bax protein or the indicated mutants that mediated the tBid-induced cytochrome

c release shown in Figures EV5.1B-C, and Figures 5.9 and 5.10A were detected by phosphor-imaging, and indicated by circles on one side of the images. The protein standards are indicated on the other side with their Mr. The other minor radioactive protein bands are backgrounds produced by the TNT reactions. As expected, addition of tBid increased the binding of most of the Bax proteins to the mitochondria. The amount of Bax proteins bound to the mitochondria varied among the mutants; however, it was not correlated with the extent of cytochrome c release by the respective mutants in Figure EV5.1C and Figures 5.9 and 5.10A, suggesting that even the lower amounts of Bax were sufficient for the MOMP unless the protein was inactivated by mutations.

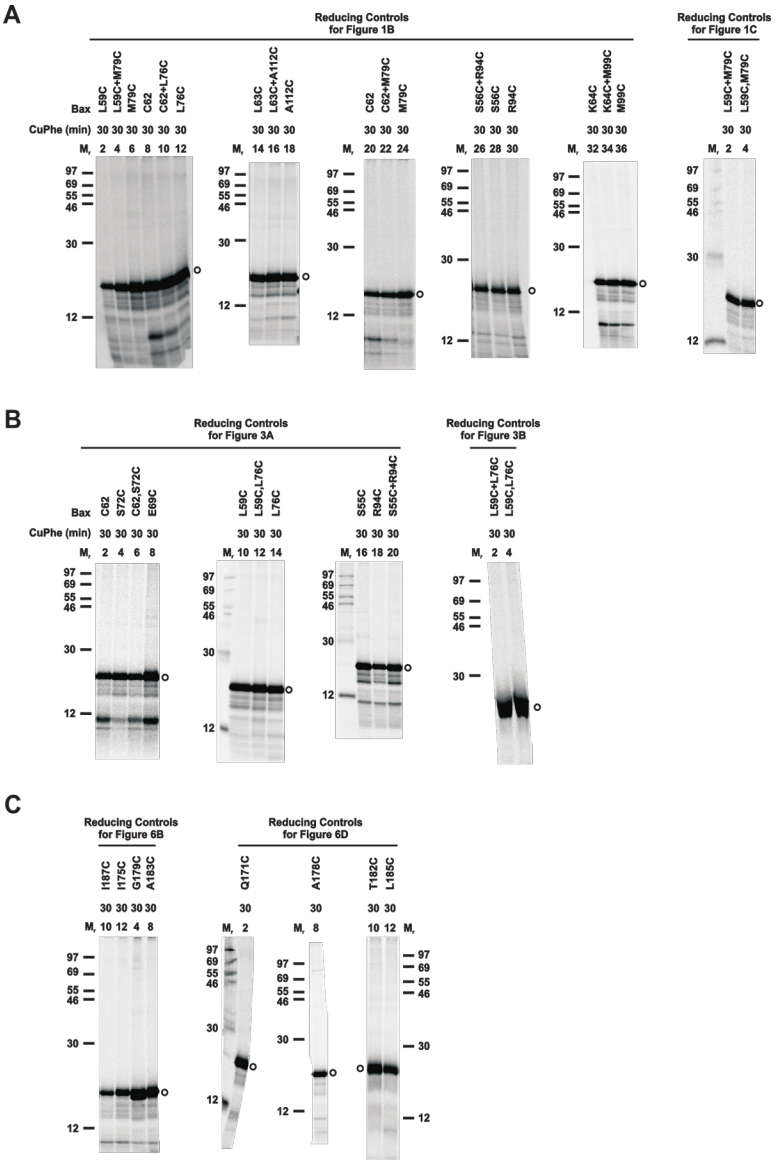

Figure S5:2: Reducing SDS-PAGE analysis of disulfide-crosslinked Bax proteins.

A) Reducing SDS-PAGE of crosslinked samples from Figure 5.1B,C.

B) Reducing SDS-PAGE of crosslinked samples from Figure 5.3A,B.

C) Reducing SDS-PAGE of crosslinked samples from Figure 5.6B,D.

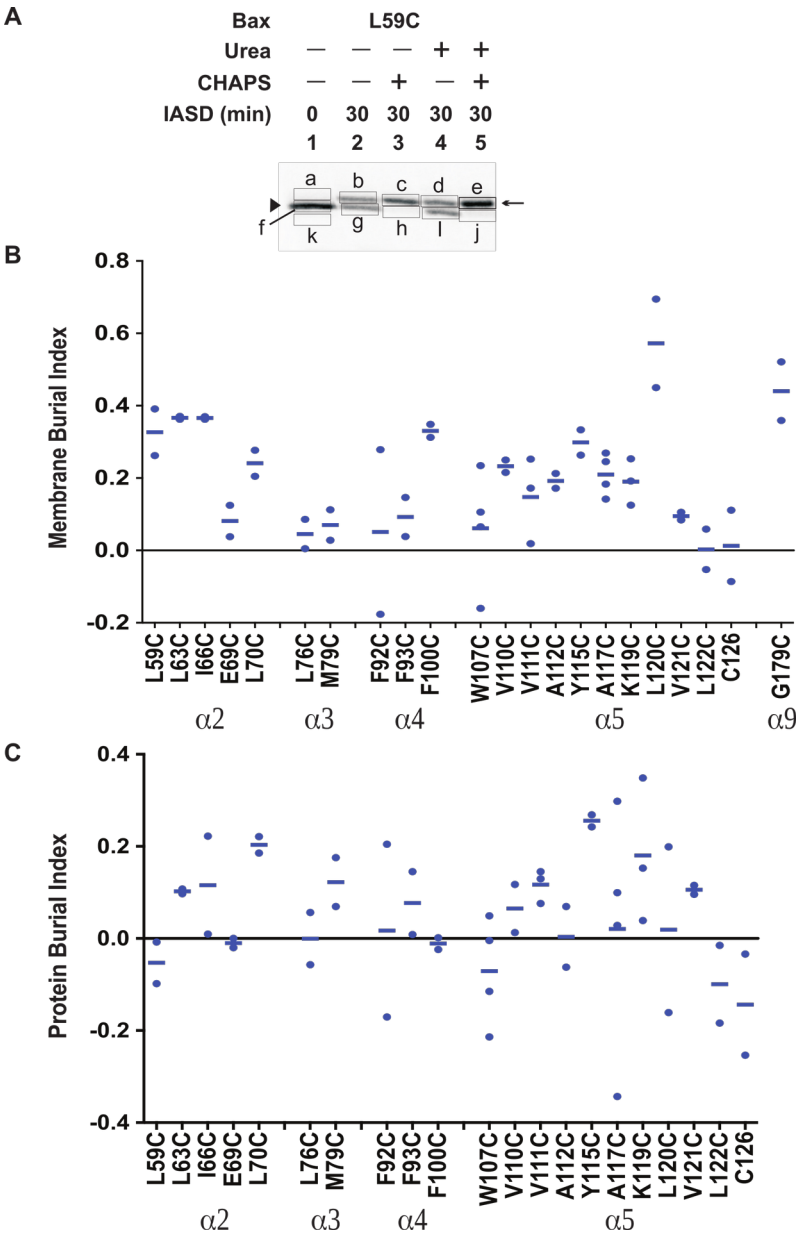

Figure S5:3: Quantification of IASD labeling data, related to Figure 5.2

A) Example data from Bax L59C for an illustration of the quantification procedure (Supplementary Methods)

B) Membrane burial indices for the analyzed Bax cysteine variants. Dots corespond to individual measurements and lines represent means of each variant.

C) Protein burial indices for the analyzed cysteine variants. Dots corespond to individual measurements and lines represent means of each variant.

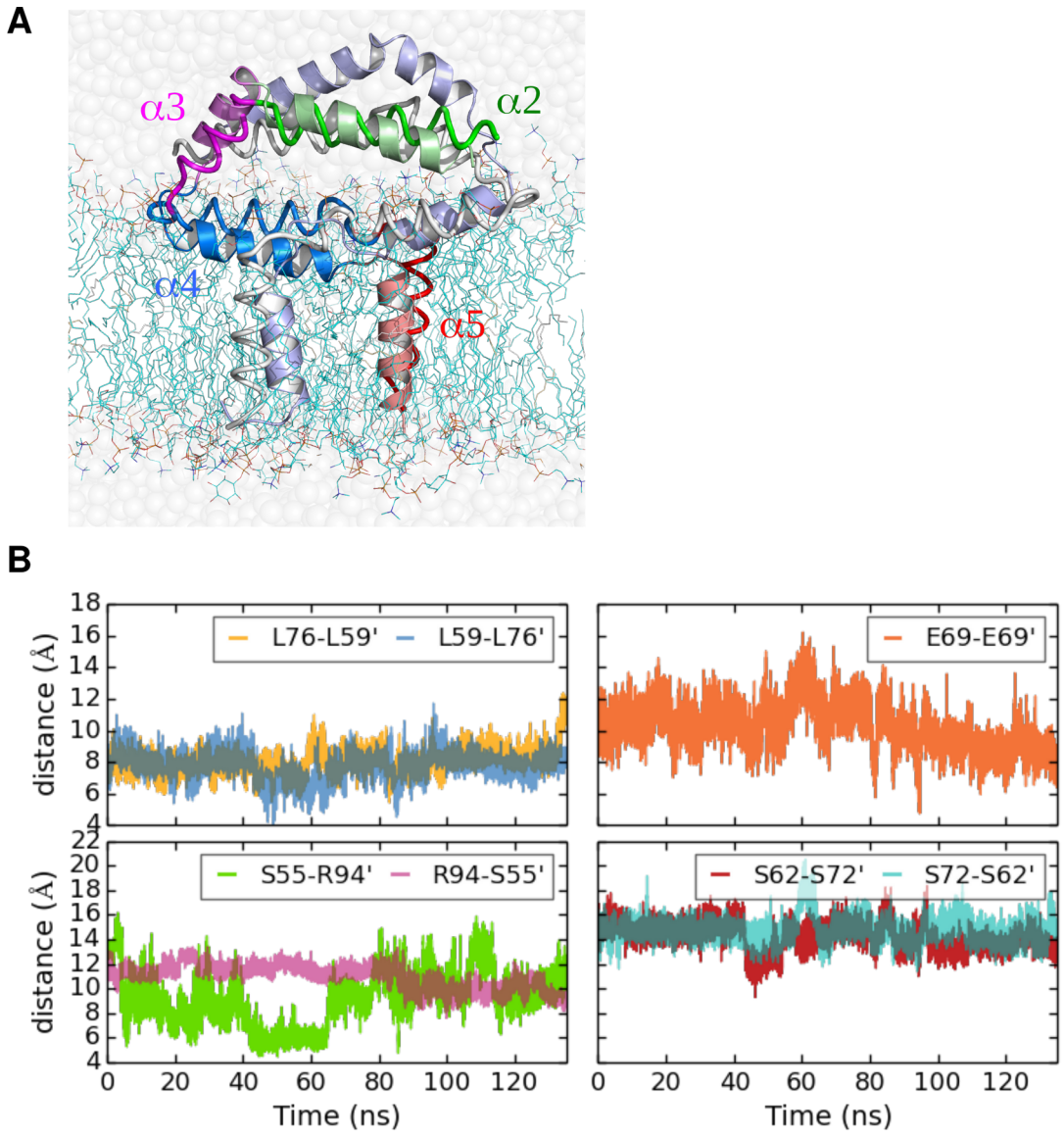

Figure S5:4: Molecular dynamics simulation of the BH3-in-groove dimer on the membrane after helix α5 inserts into the membrane.

A) The initial structure in tube representation was generated from the BH3-in-groove dimer structure (PDB entry 4BDU) with the α2-α3-α4 part remained on the membrane surface and the α5 part released into the lipid bilayer. The final structure

at t = 135 ns is shown in cartoon representation. One monomer is shown with α2 in green, α3 in magenta, α4 in blue, and α5 in red, while the other monomer is colored in gray, consistent with the color codes assigned in Figure 5.1A. The lipids in the membrane and the surrounding waters are presented as lines and van der Waals' beads, respectively.

B) Time evolutions of the β carbon distances between the indicated residue pairs in the BH3-in-groove dimer during the molecular dynamics simulation described in (A). Within 135 ns the β carbon distance between L76 in one monomer and L59' in the other monomer reached below 6 Å constantly, whereas the distance between E69 and E69' or between S55 and R94' reached below 6 Å occasionally. However, the distance between S62 and S72' remained above 10 Å.

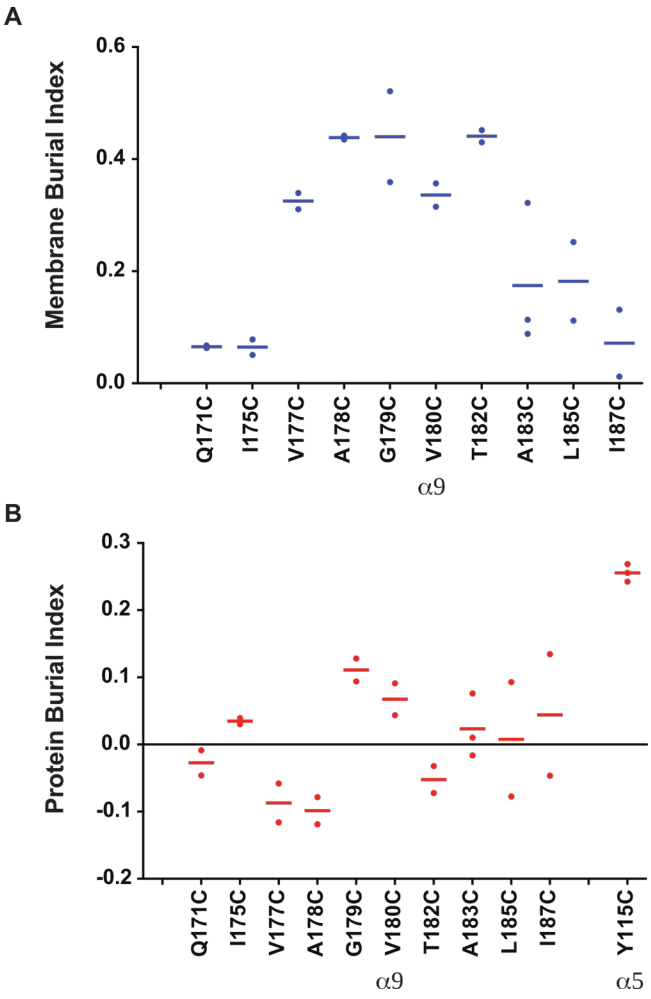

Figure S5:5: Quantification of IASD labeling data from Bax mutants with single cysteine in helix α9, related to Figure 5.5.

A) The membrane burial indices were derived and shown as dots with lines for the means for IASD labeling of Bax mutants in Figure 5.5A.

B) The protein burial indices were derived and shown as dots with lines for the means for IASD labeling of Bax mutants in Figure 5.5A.

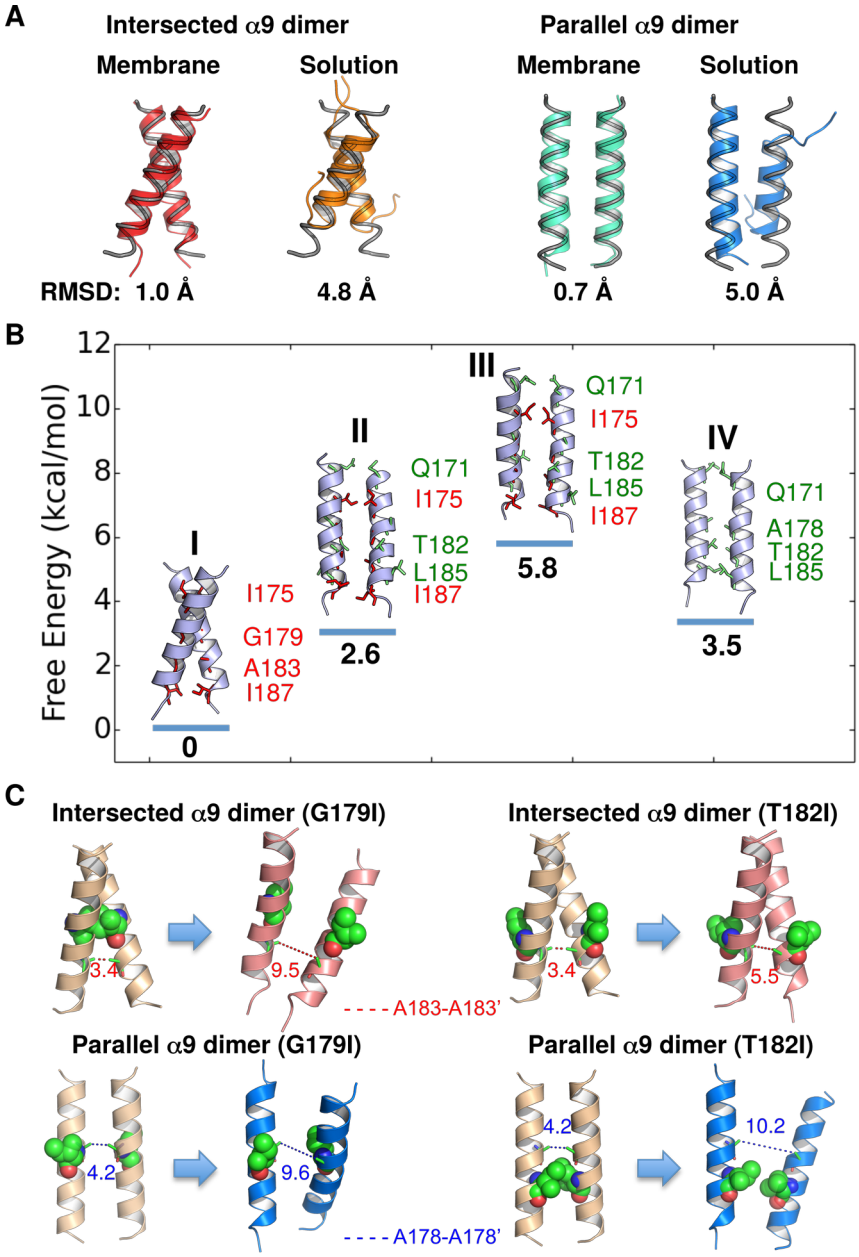

Figure S5:6: Molecular dynamics simulations of helix α9 dimers without or with G179I or T182I mutation.

A) Metadynamics simulation of free-energy of α9 dimer transition between parallel (I) and intersected (IV) conformational states. II and III are intermediate states. The

intersected state has the lowest relative free-energy, which is set to be zero. The relative free-energy (kcal/mol) of the other states are indicated under the structures. The contact residues in the intersected dimer observed in the crosslinking experiments are shown as red sticks, while those in the parallel dimer are depicted as green sticks. The backbone RMSD of conformations I, II, III, and IV, if comparing with I, are 0, 2.8, 3.3, and 5.2 Å; and if comparing with IV, are 5.2, 3.2, 3.3 and 0 Å, respectively.

B) The initial intersected and parallel α9 dimer structures at t = 0 ns are shown on the left side of each panel as light orange cartoons with the mutated residues G179I and T182I highlighted as spheres. The final structures at t = 45 ns are shown on the right side as pink and blue cartoons. Within the simulation time, the G179I mutation disrupts both intersected and parallel dimer conformations, while T182I mutation only disrupts the parallel dimer conformation, as indicated by the > 2-fold increase of β carbon distance (Å) between A183 and A183' (linked by a red dashed line with the distance value) or between A178 and A178' (linked by a blue dashed line with the distance value) from the initial to the final structure.

C) Comparison of initial and final conformations of two α9 dimers simulated in the membrane and solution. Initial conformations at t = 0 ns are shown in tube representation, to which the final conformations at t = 175 ns in cartoon representation are superimposed. As indicated by the backbone RMSD between the initial and final conformations, both intersected and parallel α9 dimers display

relatively high stability in the membrane, but they are unstable in the solution and lose helicity from both termini.

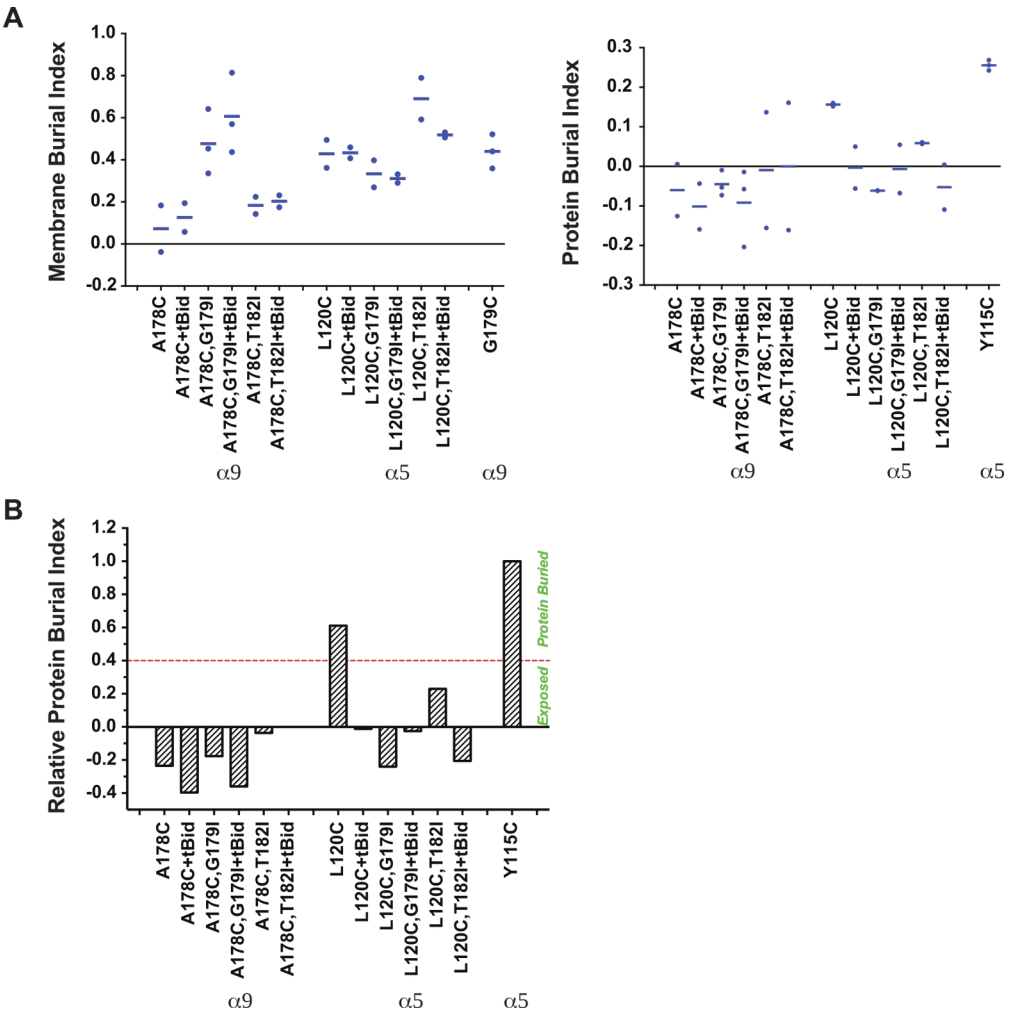

Figure S5:7: G179I mutation enhances membrane insertion of α9, related to Figure 5.10. The phosphor-imaging data for IASD labeling of Bax mutants in Figure 5.10D and the similar data from the independent replicates were quantified as described in Figure S5:3A. The membrane or protein burial indices were derived and shown in (A) as the dots with the lines for the means. The relative protein burial indices were then obtained as described in Figure 5.2D and shown in (B).

Bax dimer	Box dimensions (Å)	No. of atoms	No. of lipids	No. of water molecules	Simulation Length (ns)
Intersected α9 in membrane	77×77×69	37165	160	4711	175
Parallel α9 in membrane	77×77×69	37051	160	4673	175
Intersected to parallel α9 transition by metadynamics	77×77×69	37165	160	4711	30×2*
Intersected α9 in solvent	51×51×51	12264	0	3852	174
Parallel α9 in solvent	55×55×55	15504	0	490	174
Intersected α9-G179I in membrane	77×77×69	37018	160	4654	70
Intersected α9-T182I in membrane	77×77×69	37046	160	4668	70
Parallel α9-G179I in membrane	77×77×69	37096	160	4680	70
Parallel α9 -T182I in membrane	77×77×69	37088	160	4682	70
α2-α3-α4-α5 at membrane	73×73×97	47630	141	8557	135

* Two equivalent metadynamics simulations, each for 30 ns.

Table S5.1: Description of molecular dynamics simulations used in this study

5.10 References

Aluvila, S., Mandal, T., Hustedt, E., Fajer, P., Choe, J.Y., and Oh, K.J. (2014). Organization of the Mitochondrial Apoptotic BAK Pore: OLIGOMERIZATION OF THE BAK HOMODIMERS. J. Biol. Chem. *289*, 2537–2551.

Andrews, D.W. (2014). Pores of No Return. Mol. Cell *56*, 465–466.

Annis, M.G., Soucie, E.L., Dlugosz, P.J., Cruz-Aguado, J.A., Penn, L.Z., Leber, B., and Andrews, D.W. (2005). Bax forms multispanning monomers that oligomerize to permeabilize membranes during apoptosis. EMBO J. *24*, 2096–2103.

Basañez, G., Sharpe, J.C., Galanis, J., Brandt, T.B., Hardwick, J.M., and Zimmerberg, J. (2002). Bax-type Apoptotic Proteins Porate Pure Lipid Bilayers through a Mechanism Sensitive to Intrinsic Monolayer Curvature. J. Biol. Chem. *277*, 49360–49365.

Best, R.B., Zhu, X., Shim, J., Lopes, P.E.M., Mittal, J., Feig, M., and MacKerell, A.D. (2012). Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. J. Chem. Theory Comput. *8*, 3257–3273.

Billen, L.P., Kokoski, C.L., Lovell, J.F., Leber, B., and Andrews, D.W. (2008). Bcl-XL Inhibits Membrane Permeabilization by Competing with Bax. PLoS Biol. *6*, e147.

Bleicken, S., Classen, M., Padmavathi, P.V.L., Ishikawa, T., Zeth, K., Steinhoff, H.-J., and Bordignon, E. (2010). Molecular Details of Bax Activation, Oligomerization, and Membrane Insertion. J. Biol. Chem. *285*, 6636–6647.

Bleicken, S., Landeta, O., Landajuela, A., Basañez, G., and García-Sáez, A.J. (2013). Proapoptotic Bax and Bak Proteins Form Stable Protein-permeable Pores of Tunable Size. J. Biol. Chem. *288*, 33241–33252.

Bleicken, S., Jeschke, G., Stegmueller, C., Salvador-Gallego, R., García-Sáez, A.J., and Bordignon, E. (2014). Structural Model of Active Bax at the Membrane. Mol. Cell *56*, 496–505.

Borner, C., and Andrews, D.W. (2014). The apoptotic pore on mitochondria: are we breaking through or still stuck? Cell Death Differ. *21*, 187–191.

Brahmbhatt, H., Oppermann, S., Osterlund, E.J., Leber, B., and Andrews, D.W. (2015). Molecular Pathways: Leveraging the BCL-2 Interactome to Kill Cancer Cells--Mitochondrial Outer Membrane Permeabilization and Beyond. Clin. Cancer Res. *21*, 2671–2676.

Chai, J., Du, C., Wu, J.-W., Kyin, S., Wang, X., and Shi, Y. (2000). Structural and biochemical basis of apoptotic activation by Smac/DIABLO. Nature *406*, 855–862.

Chi, X., Kale, J., Leber, B., and Andrews, D.W. (2014). Regulating cell death at, on, and in membranes. Biochim. Biophys. Acta BBA - Mol. Cell Res. *1843*, 2100–2113.

Czabotar, P.E., Lee, E.F., Thompson, G.V., Wardak, A.Z., Fairlie, W.D., and Colman, P.M. (2011). Mutation to Bax beyond the BH3 Domain Disrupts Interactions with Pro-survival Proteins and Promotes Apoptosis. J. Biol. Chem. *286*, 7123–7131.

Czabotar, P.E., Westphal, D., Dewson, G., Ma, S., Hockings, C., Fairlie, W.D., Lee, E.F., Yao, S., Robin, A.Y., Smith, B.J., et al. (2013). Bax Crystal Structures Reveal How BH3 Domains Activate Bax and Nucleate Its Oligomerization to Induce Apoptosis. Cell *152*, 519–531.

Czabotar, P.E., Lessene, G., Strasser, A., and Adams, J.M. (2014). Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. Nat. Rev. Mol. Cell Biol. *15*, 49–63.

Dewson, G., Ma, S., Frederick, P., Hockings, C., Tan, I., Kratina, T., and Kluck, R.M. (2012). Bax dimerizes via a symmetric BH3:groove interface during apoptosis. Cell Death Differ. *19*, 661–670.

Ding, J., Mooers, B.H.M., Zhang, Z., Kale, J., Falcone, D., McNichol, J., Huang, B., Zhang, X.C., Xing, C., Andrews, D.W., et al. (2014). After Embedding in Membranes Antiapoptotic Bcl-XL Protein Binds Both Bcl-2 Homology Region 3 and Helix 1 of Proapoptotic Bax Protein to Inhibit Apoptotic Mitochondrial Permeabilization. J. Biol. Chem. *289*, 11873–11896.

Du, C., Fang, M., Li, Y., Li, L., and Wang, X. (2000). Smac, a Mitochondrial Protein that Promotes Cytochrome c–Dependent Caspase Activation by Eliminating IAP Inhibition. Cell *102*, 33–42.

Edlich, F., Banerjee, S., Suzuki, M., Cleland, M.M., Arnoult, D., Wang, C., Neutzner, A., Tjandra, N., and Youle, R.J. (2011). Bcl-xL Retrotranslocates Bax from the Mitochondria into the Cytosol. Cell *145*, 104–116.

Gahl, R.F., He, Y., Yu, S., and Tjandra, N. (2014). Conformational Rearrangements in the Pro-apoptotic Protein, Bax, as It Inserts into Mitochondria: A CELLULAR DEATH SWITCH. J. Biol. Chem. *289*, 32871–32882.

García-Sáez, A.J. (2012). The secrets of the Bcl-2 family. Cell Death Differ. *19*, 1733–1740.

García-Sáez, A.J., Chiantia, S., Salgado, J., and Schwille, P. (2007). Pore Formation by a Bax-Derived Peptide: Effect on the Line Tension of the Membrane Probed by AFM. Biophys. J. *93*, 103–112.

Garg, P., Nemec, K.N., Khaled, A.R., and Tatulian, S.A. (2013). Transmembrane pore formation by the carboxyl terminus of Bax protein. Biochim. Biophys. Acta BBA - Biomembr. *1828*, 732–742.

Gavathiotis, E., Reyna, D.E., Davis, M.L., Bird, G.H., and Walensky, L.D. (2010). BH3-Triggered Structural Reorganization Drives the Activation of Proapoptotic BAX. Mol. Cell *40*, 481–492.

Iannuzzi, M., Laio, A., and Parrinello, M. (2003). Efficient Exploration of Reactive Potential Energy Surfaces Using Car-Parrinello Molecular Dynamics. Phys. Rev. Lett. *90*.

Iyer, S., Bell, F., Westphal, D., Anwari, K., Gulbis, J., Smith, B.J., Dewson, G., and Kluck, R.M. (2015). Bak apoptotic pores involve a flexible C-terminal region and juxtaposition of the C-terminal transmembrane domains. Cell Death Differ. *22*, 1665–1675.

Jo, S., Kim, T., Iyer, V.G., and Im, W. (2008). CHARMM-GUI: A web-based graphical user interface for CHARMM. J. Comput. Chem. *29*, 1859–1865.

Kim, H., Tu, H.-C., Ren, D., Takeuchi, O., Jeffers, J.R., Zambetti, G.P., Hsieh, J.J.-D., and Cheng, E.H.-Y. (2009). Stepwise Activation of BAX and BAK by tBID, BIM, and PUMA Initiates Mitochondrial Apoptosis. Mol. Cell *36*, 487–499.

Klauda, J.B., Venable, R.M., Freites, J.A., O'Connor, J.W., Tobias, D.J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A.D., and Pastor, R.W. (2010). Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. J. Phys. Chem. B *114*, 7830–7843.

Krivov, G.G., Shapovalov, M.V., and Dunbrack, R.L. (2009). Improved prediction of protein side-chain conformations with SCWRL4. Proteins Struct. Funct. Bioinforma. *77*, 778–795.

Kulp, D.W., Subramaniam, S., Donald, J.E., Hannigan, B.T., Mueller, B.K., Grigoryan, G., and Senes, A. (2012). Structural informatics, modeling, and design with an open-source Molecular Software Library (MSL). J. Comput. Chem. *33*, 1645–1661.

Kuwana, T., Mackey, M.R., Perkins, G., Ellisman, M.H., Latterich, M., Schneiter, R., Green, D.R., and Newmeyer, D.D. (2002). Bid, Bax, and Lipids Cooperate to Form Supramolecular Openings in the Outer Mitochondrial Membrane. Cell *111*, 331–342.

Laio, A., and Parrinello, M. (2002). Escaping free-energy minima. Proc. Natl. Acad. Sci. *99*, 12562–12566.

LaPointe, L.M., Taylor, K.C., Subramaniam, S., Khadria, A., Rayment, I., and Senes, A. (2013). Structural organization of FtsB, a transmembrane protein of the bacterial divisome. Biochemistry *52*, 2574–2585.

Liu, Z., Lin, H., Ye, S., Liu, Q. -y., Meng, Z., Zhang, C. -m., Xia, Y., Margoliash, E., Rao, Z., and Liu, X. -j. (2006). Remarkably high activities of testicular cytochrome c in destroying reactive oxygen species and in triggering apoptosis. Proc. Natl. Acad. Sci. *103*, 8965–8970.

Lovell, J.F., Billen, L.P., Bindner, S., Shamas-Din, A., Fradin, C., Leber, B., and Andrews, D.W. (2008). Membrane Binding by tBid Initiates an Ordered Series of Events Culminating in Membrane Permeabilization by Bax. Cell *135*, 1074–1084.

Lucken-Ardjomande, S., Montessuit, S., and Martinou, J.-C. (2008). Contributions to Bax insertion and oligomerization of lipids of the mitochondrial outer membrane. Cell Death Differ. *15*, 929–937.

MacKenzie, K.R. (1997). A Transmembrane Helix Dimer: Structure and Implications. Science *276*, 131–133.

MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. (1998). All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †. J. Phys. Chem. B *102*, 3586–3616.

Moldoveanu, T., Follis, A.V., Kriwacki, R.W., and Green, D.R. (2014). Many players in BCL-2 family affairs. Trends Biochem. Sci. *39*, 101–111.

Mueller, B.K., Subramaniam, S., and Senes, A. (2014). A frequent, GxxxG-mediated, transmembrane association motif is optimized for the formation of interhelical C -H hydrogen bonds. Proc. Natl. Acad. Sci.

Ormö, M., Cubitt, A.B., Kallio, K., Gross, L.A., Tsien, R.Y., and Remington, S.J. (1996). Crystal Structure of the Aequorea victoria Green Fluorescent Protein. Science *273*, 1392–1395.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. J. Comput. Chem. *26*, 1781–1802.

Qian, S., Wang, W., Yang, L., and Huang, H.W. (2008). Structure of transmembrane pore induced by Bax-derived peptide: Evidence for lipidic pores. Proc. Natl. Acad. Sci. *105*, 17379–17383.

Russ, W.P., and Engelman, D.M. (2000). The GxxxG motif: a framework for transmembrane helix-helix association. J. Mol. Biol. *296*, 911–919.

Schafer, B., Quispe, J., Choudhary, V., Chipuk, J.E., Ajero, T.G., Du, H., Schneiter, R., and Kuwana, T. (2009). Mitochondrial Outer Membrane Proteins Assist Bid in Bax-mediated Lipidic Pore Formation. Mol. Biol. Cell *20*, 2276–2285.

Schellenberg, B., Wang, P., Keeble, J.A., Rodriguez-Enriquez, R., Walker, S., Owens, T.W., Foster, F., Tanianis-Hughes, J., Brennan, K., Streuli, C.H., et al. (2013). Bax Exists in a Dynamic Equilibrium between the Cytosol and Mitochondria to Control Apoptotic Priming. Mol. Cell *49*, 959–971.

Senes, A., Gerstein, M., and Engelman, D.M. (2000). Statistical analysis of amino acid patterns in transmembrane helices: the GxxxG motif occurs frequently and in association with beta-branched residues at neighboring positions. J. Mol. Biol. *296*, 921–936.

Shamas-Din, A., Bindner, S., Zhu, W., Zaltsman, Y., Campbell, C., Gross, A., Leber, B., Andrews, D.W., and Fradin, C. (2013). tBid Undergoes Multiple Conformational Changes at the Membrane Required for Bax Activation. J. Biol. Chem. *288*, 22111–22127.

Shamas-Din, A., Satsoura, D., Khan, O., Zhu, W., Leber, B., Fradin, C., and Andrews, D.W. (2014). Multiple partners can kiss-and-run: Bax transfers between multiple membranes and permeabilizes those primed by tBid. Cell Death Dis. *5*, e1277–e1277.

Shamas-Din, A., Bindner, S., Chi, X., Leber, B., Andrews, D.W., and Fradin, C. (2015). Distinct lipid effects on tBid and Bim activation of membrane permeabilization by pro-apoptotic Bax. Biochem. J. *467*, 495–505.

Subramaniam, S., and Senes, A. (2012). An energy-based conformer library for side chain optimization: Improved prediction and adjustable sampling. Proteins Struct. Funct. Bioinforma. *80*, 2218–2234.

Subramaniam, S., and Senes, A. (2014). Backbone dependency further improves side chain prediction efficiency in the Energy-based Conformer Library (bEBL): A Backbone-Dependent Energy-Based Library. Proteins Struct. Funct. Bioinforma. *82*, 3177–3187.

Sulistijo, E.S., and MacKenzie, K.R. (2009). Structural Basis for Dimerization of the BNIP3 Transmembrane Domain ˙. Biochemistry *48*, 5106–5120.

Suzuki, M., Youle, R.J., and Tjandra, N. (2000). Structure of Bax. Cell *103*, 645–654.

Tan, C., Dlugosz, P.J., Peng, J., Zhang, Z., Lapolla, S.M., Plafker, S.M., Andrews, D.W., and Lin, J. (2006). Auto-activation of the Apoptosis Protein Bax Increases Mitochondrial Membrane Permeability and Is Inhibited by Bcl-2. J. Biol. Chem. *281*, 14764–14775.

Terrones, O., Antonsson, B., Yamaguchi, H., Wang, H.-G., Liu, J., Lee, R.M., Herrmann, A., and Basañez, G. (2004). Lipidic Pore Formation by the Concerted Action of Proapoptotic BAX and tBID. J. Biol. Chem. *279*, 30081–30091.

Todt, F., Cakir, Z., Reichenbach, F., Youle, R.J., and Edlich, F. (2013). The C-terminal helix of Bcl-xL mediates Bax retrotranslocation from the mitochondria. Cell Death Differ. *20*, 333–342.

Todt, F., Cakir, Z., Reichenbach, F., Emschermann, F., Lauterwasser, J., Kaiser, A., Ichim, G., Tait, S.W., Frank, S., Langer, H.F., et al. (2015). Differential retrotranslocation of mitochondrial Bax and Bak. EMBO J. *34*, 67–80.

Verhagen, A.M., Ekert, P.G., Pakusch, M., Silke, J., Connolly, L.M., Reid, G.E., Moritz, R.L., Simpson, R.J., and Vaux, D.L. (2000). Identification of DIABLO, a Mammalian Protein that Promotes Apoptosis by Binding to and Antagonizing IAP Proteins. Cell *102*, 43–53.

Volkmann, N., Marassi, F.M., Newmeyer, D.D., and Hanein, D. (2014). The rheostat in the membrane: BCL-2 family proteins and apoptosis. Cell Death Differ. *21*, 206–215.

Walters, R.F.S., and DeGrado, W.F. (2006). Helix-packing motifs in membrane proteins. Proc. Natl. Acad. Sci. *103*, 13658–13663.

Westphal, D., Kluck, R.M., and Dewson, G. (2014a). Building blocks of the apoptotic pore: how Bax and Bak are activated and oligomerize during apoptosis. Cell Death Differ. *21*, 196–205.

Westphal, D., Dewson, G., Menard, M., Frederick, P., Iyer, S., Bartolo, R., Gibson, L., Czabotar, P.E., Smith, B.J., Adams, J.M., et al. (2014b). Apoptotic pore formation is associated with in-plane insertion of Bak or Bax central helices into the mitochondrial outer membrane. Proc. Natl. Acad. Sci. *111*, E4076–E4085.

Xiang, Z., and Honig, B. (2001). Extending the accuracy limits of prediction for side-chain conformations. J. Mol. Biol. *311*, 421–430.

Yamaguchi, R., Andreyev, A., Murphy, A.N., Perkins, G.A., Ellisman, M.H., and Newmeyer, D.D. (2007). Mitochondria frozen with trehalose retain a number of biological functions and preserve outer membrane integrity. Cell Death Differ. *14*, 616–624.

Yethon, J.A., Epand, R.F., Leber, B., Epand, R.M., and Andrews, D.W. (2003). Interaction with a Membrane Surface Triggers a Reversible Conformational Change in Bax Normally Associated with Induction of Apoptosis. J. Biol. Chem. *278*, 48935–48941.

Zha, J. (2000). Posttranslational N-Myristoylation of BID as a Molecular Switch for Targeting Mitochondria and Apoptosis. Science *290*, 1761–1765.

Zhang, Z., Zhu, W., Lapolla, S.M., Miao, Y., Shao, Y., Falcone, M., Boreham, D., McFarlane, N., Ding, J., Johnson, A.E., et al. (2010). Bax Forms an Oligomer via Separate, Yet Interdependent, Surfaces. J. Biol. Chem. *285*, 17614–17627.

Chapter 6: Ptc7p dephosphorylates select mitochondrial proteins to enhance metabolic function

This chapter was prepared for publication as:

Guo X., Niemi N.M., Hutchins P.D., Condon S.G., Jochem A., Ulbrich A., Higbee A., Russell J.D., Senes A., Coon J.J., Pagliarini D.J. 2017. “Ptc7p dephosphorylates select mitochondrial proteins to enhance metabolic function.” *Cell Reports* 18(2):307-313.

Statement of contribution:

I generated a rotamer library for phosphoserine and used a side chain repacking algorithm to model mutations of the citrate synthase Cit1p in order to better understand why phosphorylation of S462 inactivates the enzyme.

6.1 Abstract

Proper maintenance of mitochondrial activity is essential for metabolic homeostasis. Widespread phosphorylation of mitochondrial proteins may be an important element of this process; yet little is known about which enzymes control mitochondrial phosphorylation, or which phosphosites have functional impact. We investigate these issues by disrupting Ptc7p—a conserved but largely uncharacterized mitochondrial matrix PP2C-type phosphatase. Loss of Ptc7p causes respiratory growth defects concomitant with elevated phosphorylation of select matrix proteins. Among these, Δ*ptc7* yeast exhibit an increase in phosphorylation of Cit1p—the canonical citrate synthase of the tricarboxylic acid cycle—that diminishes its activity. We find that phosphorylation of S462 can eliminate Cit1p enzymatic activity likely by disrupting its proper dimerization, and that Ptc7p-driven dephosphorylation rescues Cit1p activity. Collectively, our work connects Ptc7p to an essential TCA cycle function and to additional phosphorylation events that may affect mitochondrial activity inadvertently or in a regulatory manner.

6.2 Introduction

Nearly all cells rely on mitochondria for their core metabolic needs. As such, it is essential that cells are able to appropriately maintain the content, function, and activity of these organelles. Mitochondrial adaptation can involve transcriptional reprogramming, protease-based proteome remodeling, altered mitochondrial dynamics, and the use of post-translational modifications (PTMs) to regulate protein activity (Lin et al., 2005; Nunnari and Suomalainen, 2012; Pagliarini and Dixon, 2006; Quirós et al., 2015). The importance of PTMs in mitochondrial metabolism was recognized ~50 years ago with the discovery that reversible phosphorylation modulates the

activity of pyruvate dehydrogenase (PDH) (Linn et al., 1969). More recently, roles for mitochondrial acylation (He et al., 2012), nitrosylation (Doulias et al., 2013), and redox modifications (Mailloux et al., 2014) have been identified, and a more exhaustive cataloging of protein phosphorylation sites has been achieved (Grimsrud et al., 2012). From this collective work, it is clear that PTMs are abundant, dynamic, and can affect enzymatic function.

Despite this progress, various aspects of PTM-driven regulation of mitochondrial function are unclear. First, it remains largely uncertain which mitochondrial PTMs—from among the thousands identified—might affect protein function, and which may instead be inconsequential. Second, the enzymes that regulate these modifications are largely unidentified or uncharacterized. To explore these issues, we focused on Ptc7p, a poorly characterized mitochondrial PP2C phosphatase conserved from yeast through humans. Our work connects this phosphatase to a select set of ~20 mitochondrial phosphoproteins from amongst thousands of observed phosphoisoforms that span distinct pathways. In particular, we demonstrate that Ptc7p-driven dephosphorylation of mitochondrial citrate synthase may be an important process for maintaining optimal mitochondrial activity.

6.3 Results

6.3.1 Ptc7p phosphatase activity supports respiratory function

To further elucidate the nature of mitochondrial protein phosphorylation, we focused on identifying phosphatases that could directly dephosphorylate matrix phosphoproteins. The MitoCarta compendium includes 11 proteins with predicted PFAM phosphatase domains (Calvo et al., 2016; Pagliarini et al., 2008), and two additional dual specificity phosphatases have also been localized to mitochondria (Rardin et al., 2008). Among these, we prioritized PPTC7 for

further study based on its matrix localization (Rhee et al., 2013), conservation through *Saccharomyces cerevisiae* (Ptc7p), and canonical PP2C protein phosphatase domain (Figure 6.1A).

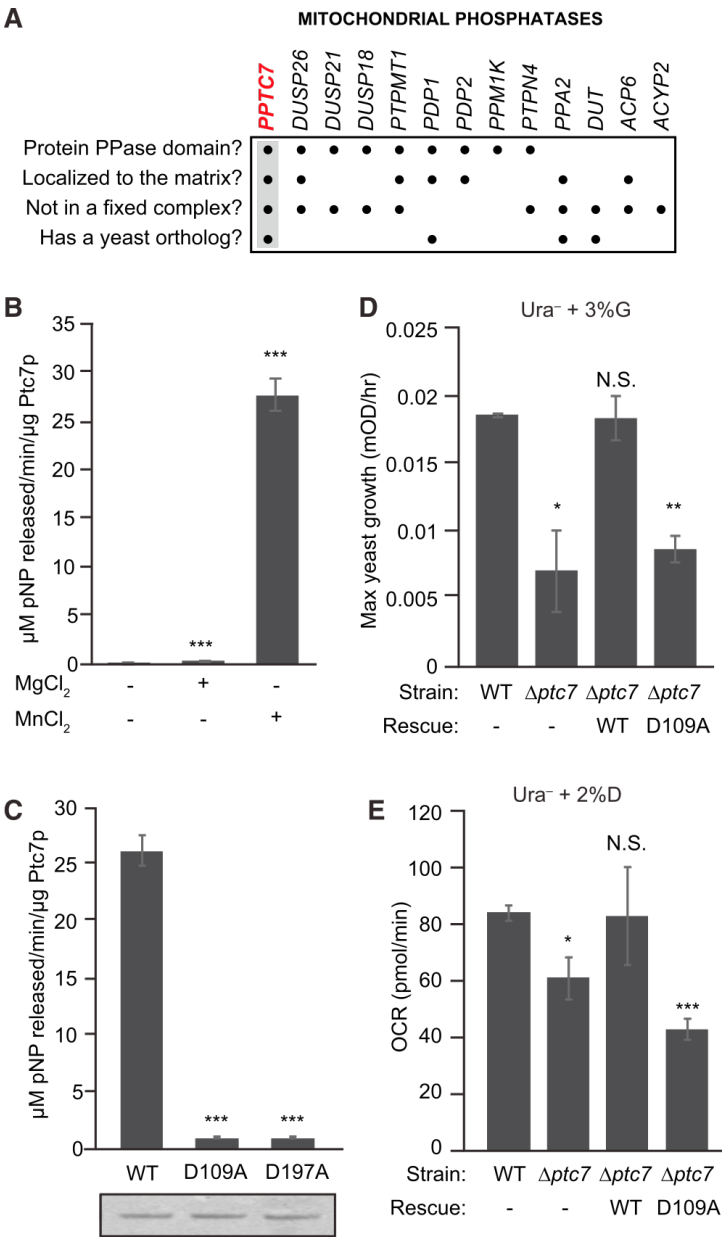

Figure 6.1: Ptc7p phosphatase activity supports respiratory function

A) Summary table of mitochondrial phosphatases, analyzed for protein phosphatase domains, matrix localization, non-association with a protein complex (e.g. PDH), and

presence of a yeast ortholog. PPTC7 (highlighted) is the only phosphatase to meet all four criteria. **B)** *In vitro* phosphatase activity assay of Ptc7p against pNPP with divalent cations Mn^{2+} and Mg^{2+} (mean ± SD, n=3). **C)** *In vitro* pNPP phosphatase activity assay of WT Ptc7p and two mutants predicted to disrupt metal binding (mean ± SD, n=3). Coomassie staining (below) demonstrates comparable protein concentration and purity. **D)** Maximum growth rate of WT and Δ*ptc7* yeast in Ura^- media containing 3% glycerol (G) (mean ± SD, n=4). Rescue strains express a plasmid containing *ptc7* (WT: wild type *ptc7*; D109A: catalytically inactive mutant of *ptc7*). **E)** OCR of same cultures as in (D) grown in Ura^- media containing 2% dextrose (D) (mean ± SD, n=3).

* p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; N.S., not significant.

We expressed and purified recombinant Ptc7p from *E. coli* and found that it dephosphorylates *para*-nitrophenyl phosphate (pNPP) *in vitro* with kinetics consistent with known protein phosphatases (Figure S6:1A) (McAvoy and Nairn, 2010), and that it strongly prefers Mn^{2+} to Mg^{2+} for catalysis (Figure 6.1B). Mutation of the conserved aspartic acid residues that coordinate these metal ions completely abrogated activity (Figure 6.1C), consistent with the established catalytic mechanism of PP2C-type phosphatases (Das et al., 1996).

We next examined phenotypes related to mitochondrial function in Δ*ptc7* yeast. We found no difference in growth between wild type (WT) and Δ*ptc7* yeast in dextrose-based media ((Figure S6:1B), but a marked decrease in Δ*ptc7* growth rate in non-fermentable glycerol-based media (Figure 6.1D). This growth deficiency was rescued by expression of WT but not the catalytically

inactive D109A Ptc7p mutant (Figure 6.1D). Consistently, Δ*ptc7* yeast also exhibited a decreased oxygen consumption rate (OCR) that was likewise rescued only by WT Ptc7p (Figure 6.1E). In contrast to a recent report (Martín-Montalvo et al., 2013), we did not observe a defect in coenzyme Q (CoQ) levels in Δ*ptc7* yeast (Figure S6:1C), and thus hypothesize that the growth impairment we observe is caused by distinct mitochondrial defects. Collectively, these results suggest that Ptc7p is an active protein phosphatase poised to dephosphorylate mitochondrial matrix proteins to promote mitochondrial function.

6.3.2 Quantitative phosphoproteomics identifies potential Ptc7p substrates

To identify potential substrates of Ptc7p, we performed quantitative phosphoproteomic analysis on WT, Δ*ptc7*, and Δ*ptc7* + *ptc7* (rescue) yeast strains (Figure 6.2A). Yeast were cultured in media containing 0.1% glucose and 3% glycerol to enable a short fermentation phase (allowing the growth of sufficient yeast for proteomics analysis), followed by a longer respiration phase during which the Δ*ptc7* growth defect is evident (Figure S6:2A). Total proteomic and phosphoproteomic analyses were performed using 8-plex tandem mass tags (TMTs) for isobaric quantification, resulting in the quantification of 4271 proteins (886 mitochondrial) and 6647 phosphoisoforms (1182 mitochondrial) (Figure 6.2B).

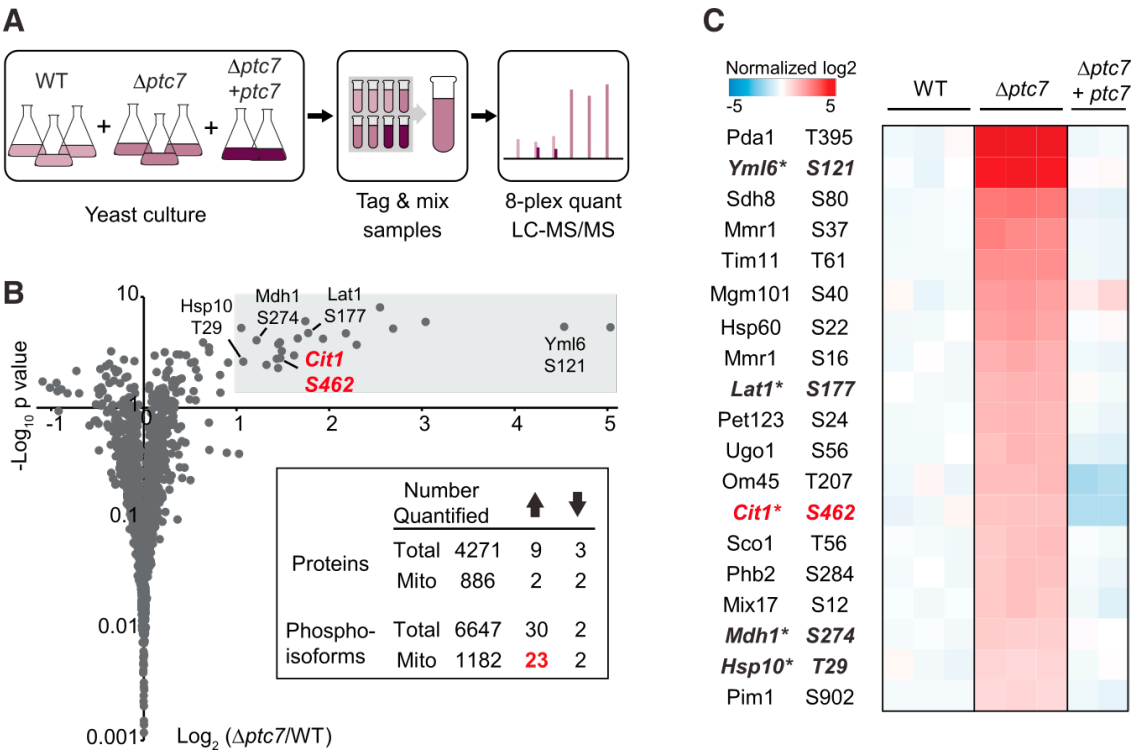

Figure 6.2: Quantitative phosphoproteomics identifies potential Ptc7p substrates

A) Experimental workflow of phosphoproteomics. Peptides from three strains (WT, Δ*ptc7*, and Δ*ptc7* + *ptc7*) were tagged with 8-plex TMT for isobaric quantification. **B)** Fold changes in mitochondrial phosphoisoform abundances ($\log_2$(Δ*ptc7*/WT), normalized to total protein abundance, n=3) versus significance ($-\log_{10}$(p-value)). Grey area indicates significance threshold ($\log_2$(Δ*ptc7*/WT) > 1 and p-value < 0.05). Five highlighted phosphoisoforms are prioritized candidate Ptc7p substrates. Inset table summarizes quantified total and mitochondrial proteins or phosphoisoforms; up arrow indicates significantly increased changes; down arrow indicates significantly decreased changes. **C)**

Heat map of 19 mitochondrial phosphoisoforms whose abundances were significantly increased in Δ*ptc7*, and were restored to WT level or below in the rescue strain (Δ*ptc7* + *ptc7*). * denotes mitochondrial matrix phosphosites conserved in higher eukaryotes.

Given that Ptc7p is an active protein phosphatase, we hypothesized that Δ*ptc7* yeast would exhibit increased phosphorylation levels of its substrate(s). Indeed, the deletion of *ptc7* caused a marked increase of 23 mitochondrial phosphoisoforms (Figure 6.2B). In contrast, only 7 non-mitochondrial phosphoisoforms (~0.1%) were significantly increased (Figures 6.2B, S6:2B). Furthermore, 19 of the 23 upregulated phosphoisoforms returned to WT levels (or below) when we rescued the Δ*ptc7* strain with plasmid-based *ptc7* expression (Figure 6.2C). Coupled with the general lack of proteome alterations in Δ*ptc7* yeast (Figure 6.2C), these data suggest that these 19 phosphoproteins may be direct substrates of Ptc7p. None of these phosphorylation sites were associated with enzymes in CoQ biosynthesis, again suggesting our observed mitochondrial defect in Δ*ptc7* yeast is distinct from this pathway.

To prioritize these likely Ptc7p substrates, we considered the submitochondrial localization of each protein and the conservation of each phosphosite. Five of the 19 phosphoproteins noted above—Yml6p, Lat1p, Cit1p, Mdh1p, and Hsp10p—share matrix localization with Ptc7p and have conserved phosphorylatable residues at the positions identified in our analyses (Figure S6:2D-H). Notably, three of these proteins (Lat1p, Cit1p, and Mdh1p) are enzymes associated with the pyruvate dehydrogenase complex/TCA cycle. From among these, we focused on Cit1p, the mitochondrial citrate synthase in yeast whose activity is generally not known to be affected by PTMs.

Phosphorylation of Cit1p at S462 disrupts enzyme function

The mitochondrial dysfunction of Δ*ptc7* yeast (Figure 6.1) suggests that the phosphorylation of putative Ptc7p substrates likely diminishes their function. To test this hypothesis, we focused on phosphorylation of S462 on Cit1p (Figure S6:3A). We made non-phosphorylatable (S462A) and phosphomimetic (S462E) mutants of Cit1p and assessed their functional effects. First, we tested the ability of these mutants to rescue the growth of Δ*cit1* yeast on acetate-based media. Consistent with our hypothesis, S462E was unable to rescue the Δ*cit1* growth deficiency, while WT and S462A fully restored growth (Figure 6.3A). Second, we measured the activity of each construct from yeast lysate and found that Δ*cit1* yeast expressing S462E had no Cit1p activity, while lysate containing a S462A mutant had a slight increase in activity versus WT (Figure 6.3B). Further, expression of S462E failed to restore total cellular citrate levels in Δ*cit1* yeast (Figure S6:3B), and resulted in an altered metabolic profile more similar to Δ*cit1* yeast expressing empty vector (EV) than those expressing WT or S462A (Figure S6:3C). Finally, we purified recombinant WT and Cit1p mutants and measured their kinetics with oxaloacetate (OAA). Consistent with our *in vivo* data, the phosphomimetic S462E completely disrupted activity, while, surprisingly, S462A showed a significantly higher V_{max} for OAA compared to WT (Figure 6.3C).

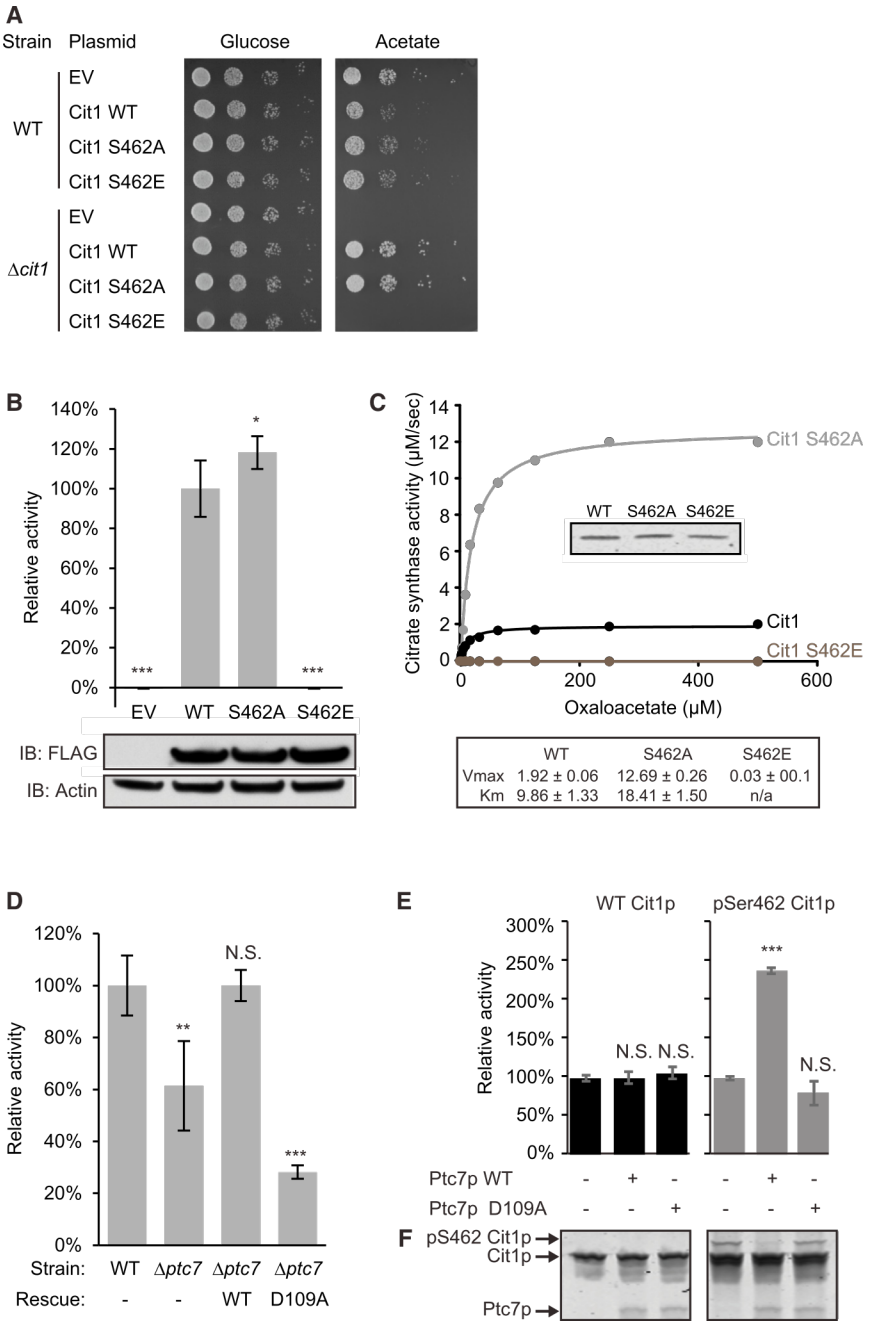

Figure 6.3: Phosphorylation of Cit1p at S462 disrupts enzyme function

A) Serial dilutions of WT and Δ*cit1* yeast expressing various plasmids grown on glucose- or acetate-containing Ura^- plates. **B)** Citrate synthase activity of Δ*cit1* lysate expressing EV, WT, S462A, or S462E Cit1p (mean ± SD, n=3). Statistics are relative to WT activity (lane #2). Inset shows immunoblot against FLAG (Cit1p-FLAG), or actin (loading control). **C)** Kinetic curve of recombinant WT, S462A, or S462E Cit1p. Citrate synthase activity (μM/sec) is plotted versus concentration of OAA (μM). Inset shows comparable loading (Coomassie staining). Table shows calculated V_{max} and K_m for each Cit1p mutant. **D)** Citrate synthase activity of lysate from WT or Δ*ptc7* (mean ± SD, n=4). Rescue strains express a plasmid containing *ptc7* (WT: wild type *ptc7*; D109A: catalytically inactive mutant of *ptc7*). **E)** Citrate synthase activity of recombinant WT or phospho-S462 (pSer462) Cit1p treated with WT or D109A Ptc7p (mean ± SD, n=3). **F)** Coomassie staining of a PhosTag gel loaded with samples in (E).

* p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; N.S., not significant.

To further assess whether Cit1p is a *bona fide* substrate of Ptc7p, we tested citrate synthase activity in Δ*ptc7* yeast. Citrate synthase activity was decreased by ~40% in this strain, and was restored by WT Ptc7p, but not by the D109A mutant (Figure 6.3D). These data are consistent with our hypothesis that phosphorylation of Cit1p S462 can decrease its enzymatic activity, especially given that Cit1p protein levels are largely unchanged in Δ*ptc7* yeast (Figure S6:3D). *Ptc7* can be alternatively spliced to create mRNAs that encode Ptc7p isoforms that translocate to the nuclear envelope (the unspliced form) or to mitochondria (the spliced form) (Juneau et al., 2009). To ensure that the change in citrate synthase activity was due to the mitochondrial form of

Ptc7p, we mutated endogenous *ptc7* to disrupt its splicing. This strain (uns-*ptc7*), which produces only the non-mitochondrial form of Ptc7p, exhibited the same loss of citrate synthase activity as did Δ*ptc7* yeast (Figure S6:3E,F).

To confirm the ability of Ptc7p to dephosphorylate Cit1p, we performed phosphatase assays with phospho-S462 Cit1p at the peptide and protein level. Ptc7p can efficiently dephosphorylate the S462 phosphopeptide and exhibits a more favorable K_m for this substrate than for pNPP (Figure S6:3G). Ptc7p also robustly dephosphorylated recombinant Cit1p phosphorylated specifically at S462, which caused a band shift on a PhosTag gel and resulted in a marked increase in Cit1p catalytic activity (Figure 6.3E,F). Notably, Ptc7p did not affect the activity of unphosphorylated Cit1p, demonstrating that the differences in activity are due specifically to phosphorylation. These data confirm that Ptc7p can directly dephosphorylate Cit1p at S462 to enhance its catalytic activity.

6.3.3 Phosphorylation of Cit1p occurs at the dimer interface

We next explored how a single phosphorylation event could cause a dramatic decrease in Cit1p activity. As the region surrounding S462 is highly conserved (Figure S6:4A), we used the porcine (PDB:3ENJ) and chicken (PDB:5CSC) citrate synthase structures to investigate the potential effects of S462 phosphorylation. These structures have revealed that eukaryotic citrate synthases function as obligate homodimers, with each subunit contributing essential residues for activity (Figure S6:4B) (Larson et al., 2009; Liao et al., 1991). Interestingly, S462 lies at this dimer interface, where it produces inter-subunit contacts (Figure S6:4C) and packs tightly with the L89 side chain of the opposite monomer (Figure 6.4A). To assess the effects of disrupting S462, we mutated this residue *in silico* (Kulp et al., 2012) and found that neither phosphorylation

of S462 nor the phosphomimetic substitution (S462E) is accommodated due to significant clashes that arise with L89, while the S462A substitution is readily accommodated (Figure 6.4B-E). Further, our analyses reveal that S462 is proximal to a salt-bridge between K461 and E277 on the opposite subunit (Figure S6:4D), suggesting that phospho-S462 may disrupt inter-subunit interaction through competition with E277 to destabilize this salt-bridge. Collectively, these data suggest that S462 likely mediates important interactions between citrate synthase monomers, leading us to hypothesize that phosphorylation of S462 may interfere with proper dimer formation or subunit orientation.

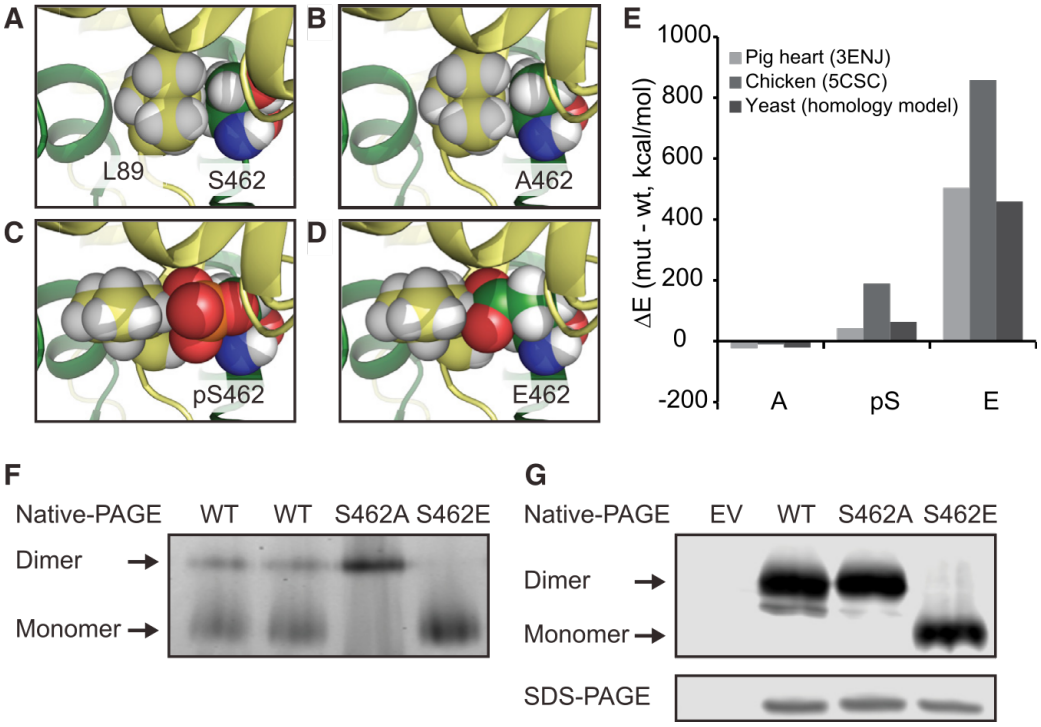

Figure 6.4: Phosphorylation of Cit1p occurs at the dimer interface

A-D) Structural interaction of residue 462 with L89 in the opposite chain with residue 462 modeled as (A) serine (WT), (B) alanine, (C) phospho-serine, and (D) glutamate (based on structure PDB code 3ENJ). **E)** Energy of computed models of alanine, phosphoserine, and glutamate mutations at position 462, compared to WT, with analysis of pig and chicken structures and a yeast homology model. **F)** Coomassie staining of native-PAGE loaded with recombinant WT (2x biological replicates), S462A, and S462E Cit1p. **G)** FLAG immunoblot of native- and SDS-PAGE resolved lysates of Δ*cit1* expressing EV, or c-terminal FLAG-tagged WT, S462A, or S462E Cit1p.

To test this hypothesis, we determined the dimerization states of recombinant WT and mutant Cit1p purified from *E. coli* using native-PAGE. WT Cit1p from two separate purifications

displayed a mixed proportion of dimeric and monomeric states, with the majority of Cit1p existing as a monomer (Figure 6.4F). Mutation of S462 disrupted this equilibrium: S462A purified almost exclusively as a dimer, while S462E was almost exclusively monomeric (Figure 6.4F). The increased proportion of the S462A dimer likely explains the increased V_{max} we observe for this variant (Figure 6.3C). To confirm that the S462E mutation is sufficient to disrupt dimerization *in vivo*, we determined the native states of WT, S462A, and S462E mutants in yeast lysate. Consistent with our *in vitro* data, WT and S462A Cit1p formed dimers *in vivo*, whereas S462E was almost exclusively monomeric (Figure 6.4G). Collectively, these experiments highlight the importance of a single conserved non-catalytic residue (S462) in enabling Cit1p function, and support a model whereby the proper dimerization state—and thus activity—of Cit1p can be influenced by Ptc7p-dependent dephosphorylation *in vivo*.

6.4 Discussion

Numerous recent investigations have revealed a vast landscape of mitochondrial PTMs associated with a diverse array of activities. From among these, select phosphorylation sites have already been implicated in apoptosis, oxidative phosphorylation, mitochondrial protein import, ketogenesis, and mitochondrial biogenesis (O'Rourke et al., 2011; Pagliarini and Dixon, 2006; Schmidt et al., 2011). Despite these findings, the nature and importance of most protein phosphorylation events in mitochondria remain unclear, and there is uncertainty regarding which enzymes are responsible for modulating the levels of this PTM.

To begin addressing these issues, we focused on Ptc7p, one of multiple likely protein phosphatases in the core mitochondrial proteome (Pagliarini et al., 2008; Rardin et al., 2008). We rationalized that disruption of Ptc7p would result in higher levels of the phosphosites that it

dephosphorylates. Indeed, the knockout of *ptc7* resulted in significantly elevated levels of just 23 mitochondrial phosphosites—less than 2% of those observed in mitochondria. Moreover, groups of these sites are on proteins are of related function. We focused on one of these proteins, Cit1p —the mitochondrial citrate synthase in *S. cerevisiae*—because of its central function in mitochondrial and cellular metabolism. We discovered that phosphorylation of S462 can inhibit Cit1p activity, likely through disruption of proper dimerization, and that Ptc7p can dephosphorylate Cit1p *in vitro* and *in vivo* to reverse this inhibition.

Further studies will be required to determine whether the phosphorylation of Cit1p and other Ptc7p-related targets is part of *bona fide* regulatory processes. Given the general absence of established kinases for these events, and the fact that mitochondrial enzymes can exhibit autophosphorylation in a potentially spurious manner (Phillips et al., 2011), it is possible that Ptc7p serves to reverse unintended and damaging protein phosphorylation. However, multiple lines of evidence suggest a more discrete regulatory process is also a possibility. First, it is curious that evolution has selected and maintained a phosphorylatable residue for Cit1p at its dimer interface (Figure S6:2H), given our demonstration that substitution of an alanine markedly enhanced its activity. This suggests that a balance between the monomeric and dimeric forms of Cit1p may play a role in its regulation, although this remains to be proven. Second, the thioredoxin-mediated redox regulation of citrate synthase in *Arabidopsis* and the regulation of the glyoxylate cycle Cit2p by ubiquitinylation in yeast suggest that post-translational modification of citrate synthase can be beneficial (Nakatsukasa et al., 2015; Schmidtmann et al., 2014). Finally, acute regulation of citrate synthase could have cellular benefits. In addition to its function in the TCA cycle, citrate acts allosterically to regulate glycolytic flux and fatty acid

synthesis, is the source of extra-mitochondrial acetyl-CoA used for *de novo* fatty acid synthesis, and its overabundance can contribute to iron accumulation-induced mitochondrial dysfunction (Chen et al., 2002).

Beyond citrate synthase, our data suggest that additional mitochondrial proteins from diverse pathways and processes may likewise be affected by Ptc7p (Figure 6.2B) and contribute to the Δ*ptc7* phenotype. Moreover, *ptc7* is alternatively spliced in response to cellular nutrient status, with the protein isoform encoded by the spliced message translocating to the mitochondrion (Juneau et al., 2009). Collectively, these data suggest that Ptc7p may participate in a nutrient-responsive adaptive process aimed at fine tuning mitochondrial function. Determining the origin of these PTMs and their full effects on enzyme and organellar function will be critical toward understanding the broader role of PTMs in maintaining mitochondrial metabolic homeostasis.

6.5 Experimental Procedures

6.5.1 Yeast Strains and Cultures

BY4741 yeast (Open Biosystems) were transformed with vector (p416-GPD) or *ptc7*-expressing plasmids. Knockout (Δ*ptc7*) and rescue (Δ*ptc7* + *ptc7*) strains were generated by replacing *ptc7* with His3MX in cells expressing vector or *ptc7* plasmids, respectively. 48 hours after transformation, individual colonies were picked directly into uracil dropout (Ura^-) media containing 0.1% glucose and 3% glycerol. Yeast were grown at 30°C for 16–24 hours to an OD ~1 before collection, and were snap frozen for proteomics analysis.

6.5.2 OCR and Growth Curve Analysis

OCR was analyzed using a Seahorse analyzer as previously described (He et al., 2013). Growth rates of yeast were calculated based on the logarithmic growth phase (ΔOD600/minute) in strains grown in 3% glycerol in a 96-well plate.

6.5.3 Ptc7p Recombinant Protein Purification and Phosphatase assay

N-terminally truncated *ptc7* (Nd38) was expressed in RIPL *E. coli* and purified using methods previously described (Stefely et al., 2015). Ptc7p and associated mutants were assayed for phosphatase activity using pNPP absorbance at 402 nm (A402).

6.5.4 LC-MS/MS Proteomics

Yeast were lysed, digested and labeled with 8-plex TMT as previously described (Hebert et al., 2014) with modification. Phosphopeptides were enriched by immobilized metal affinity chromatography (IMAC). High pH reversed phase chromatography was used to fractionate both nonphospho- and phospho-fraction after IMAC. Nano LC-MS/MS analysis was performed on Thermo Orbitrap Fusion, and raw MS data searching was done by our custom software COMPASS with 1% false discovery rate (FDR) for both peptide and protein group.

6.5.5 Citrate Synthase Activity Assay

A citrate synthase enzyme assay was performed using 100 mM Tris (pH=7.4), 0.3 mM acetyl CoA, 0.1 mM 5,5′-dithio-bis(2-nitrobenzoic acid) (DTNB), and 0.5 mM OAA (all final concentrations). Cell lysate or recombinant Cit1p protein were added as source of enzyme. Absorbance at 412 nm (A412) was monitored for reaction rate calculation.

6.5.6 Cit1p Recombinant Protein Purification

N-terminally truncated *cit1* (Nd48) was cloned into pGEX-6P-1, expressed in RIPL *E. coli*, and purified as described (Fuhs et al., 2015). Phosphoserine-incorporated Cit1p was generated using the Rinehart lab reagents for recombinant protein expression (Addgene). Briefly, Nd48-Cit1p with S462 mutated to a TAG codon as well as WT protein (both in pGEX) were co-expressed with SepOTSλ (a gift from Jesse Rinehart; Addgene # 68292) in C321.ΔA *E. coli* (a gift from Jesse Rinehart; Addgene # 68306), and purified to generate pSer462 Cit1p and WT Cit1p. Phospho-incorporation was confirmed using PhosTag acrylamide (Wako Chemical).

6.5.7 Molecular modeling

Molecular modeling of S462 variants (A, E and pS) was based on the structures of citrate synthase from pig heart (3ENJ) and chicken (5CSC). Side chains were optimized using conformers from the Energy-Based Conformer Library (Subramaniam and Senes, 2012) . The calculations were performed with programs written with the Molecular Software Library (Kulp et al., 2012).

6.5.8 Native-PAGE

Samples were loaded on a 4–16% Native-PAGE Bis-Tris gel and run for 2 hours according to manufactural protocol (Life Technologies). Protein was visualized via western blotting or Coomassie staining.

6.5.9 Statistical Analysis

Statistical analysis was performed using Excel with a two-tailed Student’s t-test.

6.6 Supplemental Experimental Procedures

6.6.1 Cloning of Ptc7 constructs and mutants

The full-length open reading form of *ptc7* (systematic name: YHR076w) was obtained from yeast genomic DNA and cloned into p416-GPD using EcoRI and HindIII restriction sites. To generate the spliced version of *ptc7*, 93 base pairs corresponding to amino acids 19-50 were eliminated using a PstI restriction-based method. The resulting ORF is 31 amino acids shorter than full length and corresponds to the *ptc7* spliced sequence as previously reported (Juneau et al., 2009). Catalytically inactive mutants were generated through the mutation of highly conserved aspartate residues required for metal binding in PP2C phosphatases (Bork et al., 1996). These two aspartates lie at positions D140/D109 (full length/spliced) and D228/D197 (full length/spliced). Each aspartate was mutated to an alanine using standard site directed mutagenesis (SDM) technique using the following primer sets:

D109A fwd: GCCGGTGTCGCAGCTGGTGTTGGAGGATG

D109A rev: CATCCTCCAACACCAGCTGCGACACCGGC

D197A fwd: GCCAACTTGGGTGCTTCTTGGTGTGG

D197A rev: CCACACCAAGAAGCACCCAAGTTGGC

For recombinant protein expression, an N-terminal deletion of 38 amino acids (Nd38) was made to the *ptc7* ORF, corresponding to the probable mitochondrial targeting sequence (MTS) of Ptc7p as predicted by MitoProt II (Claros and Vincens, 1996). Nd38-*ptc7* was cloned into the pVP68K vector as previously described (Blommel et al., 2009) creating a fusion protein of the *ptc7* ORF (C-terminal) with 8xHis and MBP tags (N-terminal) separated a tobacco etch virus (TEV) cleavage site (8His-MBP-TEV- d38-*ptc7*). Catalytically inactive mutants of Nd38-Ptc7p

in pVP68K were generated using SDM and the primer sets for D109A and D197A shown above. All cloning products and mutants were verified by Sanger sequencing.

6.6.2 Ptc7p recombinant protein generation

pVP68K plasmids encoding wild type, D109A, and D197A *ptc7* were transformed into the BL21-CodonPlus(DE3) RIPL *E. coli* strain (Agilent Technologies) and selected on 15 µg/mL chloramphenicol (cam) and 50 µg/mL kanamycin (kan). Single colonies were picked into 6 mL starter cultures (LB media supplemented with cam/kan) and grown overnight at 37°C; these starter cultures were then diluted into 500 mL TB+G autoinduction medium and allowed to grow at 37°C for 2 hours as previously described (Fox and Blommel, 2009). Bacteria were then shifted to 25°C and grown overnight (~18 hours). Cells were collected at 6000 x g and either flash frozen or immediately lysed for protein prep. Recombinant proteins were purified as previously described, with minor changes (Stefely et al., 2015). Briefly, cells were resuspended in 25 mL lysis buffer (50 mM HEPES (pH 7.2), 150 mM NaCl, 5% glycerol, 5 mM BME, 0.25 mM phenylmethylsulfonyl fluoride (PMSF), pH 7.5) and sonicated (30% amplitude, on ice; 20 s on, 40 s rest x 4 cycles per prep). Lysates were clarified by centrifugation (15,000 x g, 30 min, 4°C). TALON resin was equilibrated in ES buffer (50 mM HEPES (pH 7.5), 150 mM NaCl, 5% glycerol, 5 mM BME, 0.25 mM PMSF, pH 7.2). Lysates were added to equilibrated buffer (1 mL resin/0.5 L culture) and incubated on a nutator at 4oC for 1 hour. Protein-bound resin was washed 3x with ES buffer, followed by 3x washes in wash buffer (50 mM HEPES (pH 7.2), 150 mM NaCl, 5% glycerol, 5 mM BME, 0.25 mM PMSF, 10 mM imidazole, pH 7.2). His-tagged protein was eluted by incubating TALON resin in Elution buffer (50 mM HEPES (pH 7.2), 150 mM NaCl, 5% glycerol, 5 mM BME, 0.25 mM PMSF, 100 mM imidazole, pH 7.2) for 10 min at

room temperature on a nutator. Beads were spun (1,000 x g) and supernatant was transferred to a 30 kDa MW spin-column (Amicon) to concentrate protein. A TEV cleavage reaction was performed overnight at 4°C with constant rotation. His-MBP was subsequently purified from the reaction through reverse IMAC (TALON resin added and incubated with post-TEV reaction for 1 hour at 4°C). Supernatant of this reaction was again concentrated using 30 kDa MW spin-olumns, aliquoted, flash frozen in liquid nitrogen, and stored at -80°C. Final elutions were quantified via Nanodrop (A280) and quantification verified through equal loading of ~100 ng protein on a 10% SDS-PAGE gel, followed by Coomassie staining.

6.6.3 Ptc7p-based phosphatase assays

Nd38-Ptc7p recombinant proteins and catalytically inactive mutants (D109A and D197A) were assayed for activity using the generic phosphatase substrate para-nitrophenyl phosphate (pNPP) (New England Biolabs). For pNPP-based phosphatase assays, 100 ng of phosphatase was incubated with 10 mM pNPP in the presence of 5 mM divalent cation (either $MgCl_2$ or $MnCl_2$) diluted in a final volume of 100 µl of 50 mM Tris (pH = 8.0). Addition of pNPP was used to initiate the reaction, and reactions lacking enzyme or divalent cations were set up as negative controls. The dephosphorylation of pNPP to para-nitrophenol (pNP) produces a colorimetric reaction which was followed by monitoring absorbance at 405 nm on a Cytation 3 plate reader (BioTek). A standard curve of pNP (Sigma-Aldrich) was created for normalization of A405 to µM pNP released. To determine biochemical parameters, various concentrations of pNPP were assayed with limiting amounts of Nd38-Ptc7p or mutants to generate linear curves. The slopes of these curves (minimum 5 points) were calculated and the substrate concentration/catalysis rate pairs were used to generate Michaelis-Menten curves using the “ligand binding – one site

saturation" fit model on SigmaPlot 13 software. For phosphopeptide-based phosphatase assays, an 11 amino acid residue containing phosphopeptide corresponding to Cit1 pS462 (amino acid sequence: IERPKpSerFSTEK) was obtained from Biomatik at greater than 90% purity. Phosphopeptide was reconstituted in water at a concentration of 1 mM, aliquoted, and stored at –80°C until use. For the dephosphorylation reaction, 25 ng of Nd38-Ptc7p (wild type or D109A) was incubated with phosphopeptide (concentrations ranging from 0.4 to 400 μM), 5 mM $MnCl_2$, and 50 mM Tris (pH = 8.0). Reactions were allowed to proceed for 15 min at room temperature before determining free phosphate release through malachite green assays (Geladopoulos et al., 1991). Briefly, malachite green reagent (prepared as described in (Baykov et al., 1988)) was added at a volume of 20 μl to 80 μl phosphatase reaction (100 μl final volume). Malachite green absorbance was read in a 96 well plate at 621 nm on the Cytation 3 plate reader (BioTek). A standard curve of KH_2PO_4 was made to normalize A621 to μM free phosphate released. Kinetics were determined using the SigmaPlot software as described above.

6.6.4 Yeast strains used in this study

Haploid strains BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) and BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) were obtained from Open Biosystems and were used throughout this study. For our phosphoprotoemic analysis, we generated acute knockout and rescue strains for analysis from the parental yeast strain BY4741. The plasmids p416-GPD (vector only), *ptc7* in p416-GPD (wild type), and *ptc7* D109A in p416-GPD (catalytically inactive) were transformed into wild type BY4741 yeast as previously described (Daniel Gietz and Woods, 2002). Each strain was made null for *ptc7* through homologous recombination-based replacement of the *ptc7* ORF with a His3MX6 cassette (Longtine et al., 1998). For all experiments, Δ*ptc7* strains were

freshly generated or streaked, grown on plates for 48 hours at 30°C, stored at 4°C and used within 1 week to prevent compensatory signaling processes. The Δ*cit1* strain used in this study was generated in the Stanford Genome Project deletion library (Winzeler, 1999) and was obtained from the BY4742 library commercially available through Open Biosystems. Wild type BY4742 yeast were used as controls for these experiments. To generate rescue strains, the *cit1* ORF including ~500 base pairs of its endogenous promoter was cloned from yeast genomic DNA. This fragment was ligated into a p416-based vector containing a C-terminal 6xHis/3xHA (a generous gift from Jared Rutter) using the restriction sites SacI and SpeI. The vector was modified to replace the C-terminal 6xHis/3xHA tag with a C-terminal FLAG tag through PIPE cloning (Klock and Lesley, 2009). S462 mutants (alanine or glutamate) were generated using SDM using the following primers:

S462A fwd: AAAGGCCAAAAGCGTTCTCCACCGA

S462A rev: TCGGTGGAGAACGCTTTTGGCCTTT

S462E fwd: AAAGGCCAAAAGAGTTCTCCACCGA

S462E rev: TCGGTGGAGAACTCTTTTGGCCTTT

Empty vector, wild type cit1, S462A or S462E were all transformed into wild type BY4742 or Δcit1 BY4742 as previously described (Daniel Gietz and Woods, 2002). The marker-free unspliceable *ptc7* strain was made with the mutations and the methodology described previously (Juneau et al., 2009; Storici et al., 2001). In brief, the CORE cassette (KIURA3-KanMX4) was integrated into the genome to replace the intron of *ptc7*. A double stranded DNA containing the desired mutations in the intron region, was homologously recombined to replace the CORE cassette. The mutations were verified by sequencing. An endogenous three tandem

hemagglutinin (3xHA) tag was introduced into both WT and unspliceable *ptc7* strains on the carboxyl terminus of Ptc7p by using pFA6a-3HA-His3MX6 plasmid and the method described before (Longtine et al., 1998).

6.6.5 Yeast growth assays

For all growth assays, yeast were transformed or freshly streaked onto uracil dropout (Ura^-) agar plates containing 2% dextrose (D, w/v). Individual colonies were selected into Ura^-2%D liquid media (~3-4 mL) and incubated overnight (~14-16 hours, 30°C, 230 RPM). Cells were quantified by measurement of OD600 using a Nanodrop after a 1:20 dilution into a disposable cuvette. For drop assays, cells were serially diluted and plated at 10^4 , 10^3 , 10^2 , and 10^1 on Ura^- agar plates containing either 2%D, 3% glycerol (G, w/v), or 2% sodium acetate (w/v), and incubated at 30°C for 2 to 4 days. For growth curves, yeast were diluted to a final concentration of 5 x 10^6 cells/mL in either Ura^-2%D, Ura^-3%G or Ura^-0.1%D+3%G. 100 µl (final [c] cells = 5 x 10^5 cells/well) were transferred to a round bottom 96 well plate and sealed with a Breathe-Easy gas permeable sealing membrane (Sigma-Aldrich). Plates were transferred to a Cytation 3 plate reader (BioTek) and incubated at a constant temperature of 30°C with constant linear shaking (1096 cpm). OD600 were taken every 10 min for 16-24 hours. To quantify maximum yeast growth, the MaxV function of at least 5 points of the linear region of growth (~4 hours for glucose-containing cultures, ~12 hours for glycerol containing cultures) was calculated using the Gen5 2.09 plate reader software. This corresponds to the maximal slope of growth across these time points, and is reported in mOD/hr. At least 3 replicates were quantified, and statistical significance was calculated using a Student's t-test on the growth rates quantified from these replicates.

6.6.6 Oxygen consumption assays

Oxygen consumption assays for yeast were performed using a Seahorse analyzer as previously described (He et al., 2013). Briefly, cells were grown in overnight cultures containing Ura$^-$2%D as described above (see Yeast growth assays). Cells were normalized to 4 x 10^6 cells/mL in fresh Ura$^-$2%D media, and 100 µl of these cells (final [c] cells = 4 x 10^5 cells/well) were transferred to a poly-D-lysine coated (25 µg/mL) 96 well Seahorse plate (Seahorse Biosciences). Plate was spun for 1 min at 1000 x g to sediment yeast, which were then incubated at 30°C for 30 min before being placed in the Seahorse XF-96 Extracellular Flux Analyzer. Oxygen consumption rate (OCR) was measured with a 1 min mix step and 3 min measure step for 3 rounds, with the final round of OCR reported. At least 3 replicates were quantified, and statistical significance was calculated using a Student's t-test on the growth rates quantified from these replicates.

6.6.7 Yeast growth for phosphoproteomics study

Yeast strains were generated as acute knockout and rescue strains as detailed (yeast strains used for this study) above. Importantly, 48 hours after transformation, individual colonies representing true biological replicates (independently derived knockouts) were selected directly into 100 mL Ura$^-$0.1%D 3%G media. For this study, 3x wild type colonies (BY4741 + p416-GPD), 3x Δ*ptc7* knockout colonies (BY4741 *ptc7*:His3MX + p416-GPD), and 2x rescue colonies (BY4741 *ptc7*:His3MX + full-length *ptc7* in p416-GPD) were cultured. Each strain was grown to an OD~1 at 30°C with constant shaking (230 RPM) for ~16-24 hours, depending on the genotype of the strain (see Figure S6:2A). 10 mL of sample was saved for analysis of CoQ levels (see Coenzyme Q6 and HAB analysis), 10 mL of sample was saved for genomic DNA isolation

and subsequent genotyping, and the remaining 80 mL was saved for proteomic and phosphoproteomic analysis. All samples were spun down at 1000 x g and the pellet flash frozen in liquid nitrogen. Samples were stored at -80°C until processed for mass spectrometry analysis. All samples were confirmed as proper genotypes at both the *ptc7* genomic locus, as well as for plasmid expression (empty vector or *ptc7*-expressing).

6.6.8 Sample preparation for LC-MS/MS

Samples were thawed on ice and resuspended in lysis buffer (8 M urea, 100 mM Tris (pH = 8.0), 10 mM tris(2- carboxyethyl)phosphine (TCEP), and 40 mM chloroacetamide (CAA)), followed by addition of methanol to 90% as final volume percentage. The mix was vortexed for ~30 s, and protein precipitate were collected after centrifugation (12,000 x g for 5 min). The protein pellet was resuspended in lysis buffer, vortexed for 30 min, and further diluted to 1.5 M urea with 50 mM Tris (pH = 8.0). Trypsin was added according to ~1:50 mass ratio to protein, and incubated overnight at room temperature. The digestion was quenched by adding 10% trifluoroacetic acid (TFA) to bring the pH below 2.0. The peptides were desalted using Sep-Pak Vac 1cc tC18 cartridges (Waters). The C18 columns were equilibrated in sequential order with 3 mL acetonitrile (ACN), 1 mL 70% ACN, 1 mL 40% ACN, 1 mL 20% ACN, and 3 mL 0.1% TFA. Samples were loaded, and followed by 3 mL 0.1% TFA wash. Peptides were eluted off by adding 1 mL 40% ACN, and then 0.75 mL 70% ACN. The elutions were dried in a speed vac, and resuspended in 0.5 mL 0.2% formic acid (FA). Pierce™ quantitative colorimetric peptide assay (Thermo) was used to determine peptide concentration. A total amount of 0.5 mg peptide from each sample was aliquoted out, and followed by 8-plex TMT labeling. Each sample was resuspended in 100 μL of 200 mM triethyl ammonium bicarbonate (TEAB, pH = 8.0), and each

tag was resuspended in 50 µL ACN. After mixing sample with tag, incubation was done at room temperature for 2 hours while shaking. 8 µL of 5% hydroxylamine was added to each sample, and incubated at room temperature for 15 min while shaking to quench the reaction. A test mixture was analyzed by LC-MS/MS to confirm larger than 95% labeling efficiency before combining all samples. The mixture was dried down and ready for phosphopeptide enrichment.

6.6.9 Phosphopeptide enrichment and high pH reverse phase fractionation

Immobilized metal affinity chromatography (IMAC) with magnetic agarose beads (Qiagen) was used to enrich for phosphopeptide (Phanstiel et al., 2011). Dried samples were resuspended in 1 mL 80% ACN/0.1% TFA. The magnetic beads were washed three times with water, incubated in 40 mM (pH = 8.0) ethylenediaminetetraacetic acid (EDTA) for 30 min while shaking. The beads were then washed three times with water again, followed with incubating in 100 mM $FeCl_3$ for 30 min while shaking, and 4 times wash with 80% ACN/0.1% TFA. Resuspended samples were added into beads and incubated for 30 min while shaking. Flow-through was saved for protein analysis. The beads were washed 3 times with 80% ACN/0.1% TFA. 400 µL 50% ACN/0.7% NH_4OH was added and vortexed for 1 min to elute phosphopeptides out of beads. The elution procedure was done twice, and combined to dry. Both IMAC flow- through and elution were further fractionated by a Gemini C18 reversed phase column (4.6 mm x 250 mm; Phenomenex). Mobile phase A (20 mM ammonium formate, pH = 10) and mobile phase B (20 mM ammonium formate in 80% ACN, pH = 10) were used to make gradient. Surveyor LC quaternary pump (Thermo) was applied to generate gradient with flow rate 0.8 mL/min. Samples were resuspended in 500 µL mobile phase A and injected into column.

Fractions were collected every minute, dried out, and further combined to make a total of 10 fractions for both phosphorylation and protein analysis.

6.6.10 LC-MS/MS

Samples were analyzed by reverse phase liquid chromatography on a Dionex UltiMate UPLC system (Thermo) coupled to a Thermo Orbitrap Fusion (Thermo) (Hebert et al., 2014; Richards et al., 2015). A 75-360 μm inner-outer diameter silica capillary was packed with 1.7 μm diameter Bridged Ethylene Hybrid C18 particles (Waters). Mobile phase A (0.1% formic acid, and 5% dimethyl sulfoxide (DMSO) in water), and mobile phase B (0.1% formic acid, and 5% DMSO in acetonitrile) were made for a 90 min gradient running at 600C. MS1 scan with 60 K resolution and 5 x 10^5 AGC target was performed from 300 to 1250 m/z, followed with data dependent MS2 scans in a cycle time of 3 s. The precursors were isolated in 1.2 Th window, followed by HCD fragmentation with normalized collision energy of 35. The MS2 scan was set at resolution of 60 K with 5 x 10^4 AGC target. The maximum injection time was set up at 100 ms for MS1 and 300 ms for MS2. 60 s of dynamic exclusion time were applied, and precursor ions with +2 to +8 charge states are included for MS2.

6.6.11 Database searching and protein/phosphoprotein analysis

Coon OMSSA Proteomics Software Suite (COMPASS) was applied to analyze MS raw files (Wenger et al., 2011). MS raw files were first converted into .dta text files using DTA Generator, and searched against a target-decoy data base of both canonical and isoforms of saccharomyces cerevisiae proteome (.fasta file of database was downloaded from UniProt). A precursor mass tolerance of 150 ppm, a product ion mass tolerance of 0.01 Da, and up to 3 missed cleavages with trypsin were applied to searching based on Open Mass Spectrometry Search Algorithm

(OMSSA) (Geer et al., 2004). Fixed modifications included carbamidomethylation of cysteines, and 8-plex TMT labeling on N- terminus and lysine residues. Variable modifications included oxidation of methionines and 8-plex TMT on tyrosines. Additional variable modification including phosphorylation with neutral loss on serine and threonine residues, and intact phosphorylation on tyrosine residues were applied to phosphopeptide searching. FDR Optimizer was utilized to filter peptides to 1% false discovery rate (FDR) with maximum 25 ppm mass error. TagQuant was used to assign TMT reporter ion intensities, with isotope purity corrections and normalization to total intensities in each channel. Protein Hoarder grouped peptides grouped into parsimonious protein groups at 1% FDR at the unique protein group level, and total reporter ion intensities of all PSMs within one protein group was used for protein intensity (Nesvizhskii and Aebersold, 2005). Phosphorylation localization was further decided by Phosphinator with ambiguity score threshold set at 13. Total reporter ion intensities of all localized phosphopeptides represented the intensity of one phosphoisoform. The intensity of phosphoisoform was normalized to the corresponding protein intensity. A database of yeast mitochondrial proteome (http://mitominer.mrc-mbu.cam.ac.uk/release-3.0/begin.do) was utilized to classify proteins as mitochondrial or non-mitochondrial.

6.6.12 Coenzyme Q6 and HAB analysis

A frozen pellet of yeast (1 x 10^8 yeast cells) generated from identical samples used for phosphoproteomic analysis (see *Yeast growth for phosphoproteomic study*) was thawed on ice and mixed with 200 µL phosphate buffered saline and 100 µL glass beads (0.5 mm diameter), followed by 30 s vortexing to lyse. 10 µL of 10 µM Coenzyme Q10 (CoQ10) was added as an internal standard, and the lysate was vortexed for 30 s again. 500 µL hexanes/2-propanol (10:1,

v/v) was added and vortexed (2 x 30 s). The samples were centrifuged at 3,000 g for 1 min at 4 °C, and 400 µL of the organic phase was transferred to a clean tube and dried under N 2 (g). The residue was reconstituted in ACN/IPA/H2O (65:30:5, v/v/v) (100 µL) by vortexing (30 s) and transferred to a glass vial for LC-MS analysis. LC- MS analysis of CoQ6 and HAB was done as previously described (Stefely et al., 2015).

6.6.13 Citrate quantification by GC/MS

$1x10^7$ yeast cells were inoculated in to 50 mL Ura^-0.1%D 3%G media. After 21 h, OD600 was measured to determine cell density (OD600 of 1 corresponds to 10^7 cells/mL). 2 x 10^8 yeast cells were rapidly isolated by vacuum filtration onto 1 nylon filter membrane (0.45 µm pore size, Millipore). The cells were then washed by 1 mL of phosphate buffered saline, and immediately submerged into 1.5 mL pre-cooled extraction solvent (ACN/MeOH/H_2O = 2/2/1 (v/v/v)) in a 2 mL plastic tube. The tubes were stored at -80°C before analysis. Yeast extract (50 µL aliquot) and internal standard (10 µL d4-citrate, 80 µM) were aliquoted into glass vials and dried by vacuum centrifuge (1 hr). The dried extracts were resuspended in pyridine (25 µL) and vortexed. 25 µL of N-methyl-N-trimethylsilyl trifluoroacetamide (MSTFA) with 1% trimethylchlorosilane (TMCS) was added, and the sample was vortexed and incubated (60°C, 60 min). Samples were then analyzed using a GC/MS instrument comprising a Trace 1310 GC coupled to a Q Exactive Orbitrap mass spectrometer. The thermal gradient started at 150°C for 21 min, and then was ramped by 100°C/min to 320°C over 1.7 min and then held for 10 min. Analytes were injected onto a 30 m TraceGOLD TG-5SILMS column (Thermo Scientific) using a 1:10 split at a temperature of 275 °C and ionized using electron ionization (EI). The mass spectrometer was

operated in SIM mode monitoring 273 m/z and 276 m/z using a resolution of 30,000 (m/Δm) relative to 200 m/z.

6.6.14 Citrate synthase activity assay

Yeast pellets were collected by centrifugation at 4000 rpm (3220 x g) for 5 min, washed with 0.5 mL H_2O, and transferred to 1.5 mL Eppendorf tube. The rest of the procedure was done either on ice or at 4°C. 300 µL lysis buffer (100 mM Tris (pH = 7.4), 1% (v/v) Triton X-100, 1 mM EDTA, 1 mM benzamidine hydrochloride, 1 mM PMSF, 1 x Protease Inhibitor Cocktail (Roche), 1 x PhosSTOP (Roche)), and 200 µL glass beads were added to the cell pellet containing tube. Five rounds of 1 min beating and 30 s resting were done to lyse the cells. The clarified cell lysate was collected after centrifugation at 16,000 g for 10 min. Bicinchoninic acid (BCA) assay (Thermo) was done to quantify the protein concentration. Cell lysates were normalized to the same concentration. Cytation 3 plate reader (BioTek) and a 96 well plate were used to read the colorimetric citrate synthase assay. 40 µL of 500 mM Tris (pH = 7.4), 2 µL of 30 mM acetyl CoA, 2 µL of 10 mM 5,5'- dithio-bis(2-nitrobenzoic acid) (DTNB) , 96 µL H_2O, and 50 µL cell lysate containing 2.5 µg total protein, were added into each well (Srere, 1969). 10 µL of 10 mM oxaloacetic acid (OAA) were added per well, and mixed by pipetting up and down. Absorbance at 412 nm (A412) was measured every 50 s. The initial slope was calculated by using data from the first 12 min, and used as enzyme reaction rate. For recombinant protein citrate synthase assays, 100 ng protein was incubated in the same reaction mixture as described above, but the concentration of OAA was adjusted (range of 0.5 to 500 µM) to facilitate kinetic measurements, as assayed by linear curves generated with limiting enzyme. The slopes of these curves (minimum 5 points) were calculated and the substrate concentration/catalysis rate pairs were

used to generate Michaelis-Menten curves using the “ligand binding – one site saturation” fit model on SigmaPlot 13 software.

6.6.15 Homolog sequence alignment

Protein sequences were downloaded from the Uniprot (http://www.uniprot.org/) or NCBI (http:// ww.ncbi.nlm.nih.gov/) website. MacVector was used to alignment multiple protein sequences. ClustalW alignment algorithm was used with default settings.

6.6.16 Blue-Native gel and immunoblotting

Yeast cells were collected, lysed, and quantified the same way as described (see *Citrate synthase activity assay*). The following protocol is adopted from NativePAGE Bis-Tris gel electrophoresis protocol (Life Technologies). Briefly, 15 µL cell lysate was mixed with 5 µL 4 x NativePAGE sample buffer, and 1 µL of 5% NativePAGE G-250 sample additive. 10 µL of each sample were loaded to 4-16% NativePAGE Bis-Tris gel. 1 x NativePAGE anode buffer and 1 x NativePAGE dark blue cathode buffer were added into anode, or cathode chamber, respectively. The gel was run at 150 V for 1 h, then the dark blue cathode buffer was replaced with 1x light blue cathode buffer. The voltage was increased to 250 V until the dye front reached the bottom of the gel.

A standard western blotting protocol was done for regular SDS-PAGE. For Blue NativePAGE, after the transfer, the membrane was incubated in 20 mL 8% acetic acid for 15 min before a water wash, air dry, and rewet with methanol. Antibodies against FLAG (M2, Sigma-Aldrich with 1:5,000 dilution), HA (Thermo, 26183 with 1:10,000 dilution), and actin (Abcam, ab8224 with 1:1,000 dilution) were each used and incubated overnight at 4°C. Secondary antibodies

diluted to 1:10,000 were used, and the LiCor Odyssey CLx imaging system was used to visualize the protein band.

6.6.17 Recombinant Cit1p protein prep

For recombinant protein expression, an N-terminal deletion of 48 amino acids (Nd48) was made to the *cit1* ORF, corresponding to the previously determined MTS of Cit1p (Vögtle et al., 2009). Nd48-cit1 was cloned into the pGEX-6P-1 using BamHI and XhoI restriction sites to create a GST-Prescission Protease-Nd48 Cit1p fusion protein. S462A and S462E mutants were generated using SDM using the same primers pairs as described above (see *Yeast strains used in this study*). All cloning and mutagenesis was confirmed using Sanger sequencing. pGEX-6P-1 plasmids encoding wild type, S462A, and S462E *cit1* were transformed into the BL21-CodonPlus(DE3) RIPL *E. coli* strain (Agilent Technologies) and selected on 15 µg/mL chloramphenicol (cam) and 100 µg/mL ampicillin (amp). Single colonies were picked into 50 mL starter cultures (LB media supplemented with cam/amp) and grown overnight at 37°C; these starter cultures were then diluted into 500 mL LB media supplemented with cam/amp and allowed to grow at 37°C until bacteria reached log phase (OD600 ~ 0.7). Bacteria were then induced with 1mM final [c] IPTG, shifted to 25°C, and grown overnight (~18 hours). Cells were collected at 6000 x g and either flash frozen or immediately lysed for protein prep. Recombinant proteins were purified as previously described, with minor changes (Fuhs et al., 2015). Briefly, cells were resuspended in 1 mL GST lysis buffer (PBS, pH 8.0, 1% Triton X-100, 5% glycerol, 1 mM DTT) per 50 mL total culture and sonicated (30% amplitude, on ice; 20 s on, 40 s rest x 4 cycles per prep). Lysates were clarified by centrifugation (15,000 x g, 30 min, 4°C). Glutathione resin was washed with GST lysis buffer clarified lysates were added to equilibrated buffer (1.5

mL resin/0.5 L culture) and incubated on a nutator at 4°C for 1.5 hour. Protein-bound resin was washed 3x with GST lysis buffer, followed by resuspension in Prescission Protease Buffer (20 mM Tris pH=7.0, 150 mM NaCl, 1 mM DTT, 0.5 mM EDTA). 10U of Prescission Protease (GE Healthcare) was added to each reaction and incubated overnight at 4°C with constant rotation. After this overnight cleavage reaction, glutathione resin was spun (1000 x g) and supernatant was transferred to a 30 kDa MW spin-column (Amicon) to concentrate protein. Concentrated protein was then aliquoted, flash frozen in liquid nitrogen, and stored at -80°C. Final elutions were quantified via Nanodrop (A280) and quantification verified through equal loading of ~100 ng protein on a 10% SDS- PAGE gel, followed by Coomassie staining.

6.6.18 Phosphoserine-incorporated Cit1p protein prep

To generate recombinant Cit1p with phosphoserine specifically incorporated at S462, we used a phosphoserine-incorporation system previously described (Pirman et al., 2015) with reagents provided as a kind gift from the Rinehart laboratory. N48-Cit1p in pGEX-6P-1, as described above, was used as the backbone expression vector for phosphoserine-incorporated protein. To generate “wild type” (non-phosphoserine incorporated) protein in the phosphoserine-expression system, the wild type sequence was used. To generate phosphoserine incorporated Nd48- Cit1p phosphoS462, the S462 codon was mutated to a “TAG” using the following primer pairs:

S462TAG fwd: CTCCAATCGAAAGGCCAAAATAGTTCTCCACCGAAAAATAC

S462TAG rev: GTATTTTTCGGTGGAGAACTATTTTGGCCTTTCGATTGGAG

Either wild type or S462TAG Nd48-Cit1p in pGEX-6P-1 were co-transformed into C321.ΔA bacteria (a gift from Jesse Rinehart; Addgene # 68306) with the SepOTSλ (a gift from Jesse Rinehart; Addgene # 68292) and selected on LB agar plates supplemented with amp and kan.

Transformed C321.ΔA bacteria were grown at 30°C. Single colonies were selected into 50 mL starter cultures (LB media supplemented with amp/kan) and grown overnight at 30°C; these starter cultures were then diluted into 1L LB media supplemented with amp/kan, 0.08% glucose, and 1.25 mM O-phospho-serine (Sigma-Aldrich). These cultures were allowed to grow at 30°C until bacteria reached log phase (OD600 ~ 0.7), and then were induced with 1mM final [c] IPTG, shifted to 23°C, and grown overnight (~18 hours). Proteins were purified using the same techniques as above (see *Recombinant Cit1p protein prep*).

6.6.19 PhosTag gel

Phosphoserine incorporation was validated by running samples on PhosTag acrylamide gel (Wako Biosciences) according to manufacturer's instructions. Before loading onto the PhosTag gel, both wild type and phosphoserine incorporated Nd48-Cit1p were subjected to phosphatase treatment using wild type or catalytically inactive (D109A) Nd38-Ptc7p. 2 µl Nanodrop-normalized Cit1p was incubated with 50 mM $MnCl_2$ in 50 mM Tris (pH=8.0). No enzyme, Ptc7p, or Ptc7p D109A were added (100 ng/reaction) to start the phosphatase reaction. Reactions were incubated at room temperature for 15 min before quenching the reaction with 4x EDTA-free Laemmli buffer and boiling the samples at 95°C for 5 min. Samples (12.5 µl total) were then run on a 10% acrylamide gel supplemented with 100 µM PhosTag acrylamide (Wako Biosciences) according to manufacturer's instructions. PhosphoS462-Cit1p was visualized through a significant gel shift that was diminished by treatment with wild type Ptc7p, but not catalytically inactive protein.

6.6.20 Molecular modeling

A homology-based model of dimeric yeast Cit1p in the open state was generated from the pig heart structure (PDB code 3ENJ) using SWISS-MODEL (Biasini et al., 2014). This homology model as well as the pig heart (3ENJ) and chicken (5CSC) structures were then minimized using BFGS constrained optimization. Serine 462 (424 in 3ENJ and 5CSC) was replaced with alanine, glutamate, and phosphoserine (using the SP2 CHARMM patch), and side chains were optimized using conformers from the Energy-Based Conformer Library (Subramaniam and Senes, 2012) with multiple iterations of a greedy algorithm using the CHARMM 22 van der Waals and electrostatic functions (Brooks et al., 2009). A conformer library for dianionic phosphoserine was generated by using the serine conformers in the Energy-Based Rotamer Library as a base (Subramaniam and Senes, 2012), and by sampling the χ3 rotation (dihedral rotation about the Oγ-P bond) in increments of 10 degrees. All calculations were performed with programs written with the Molecular Software Library (Kulp et al., 2012).

6.7 Supplementary Figures

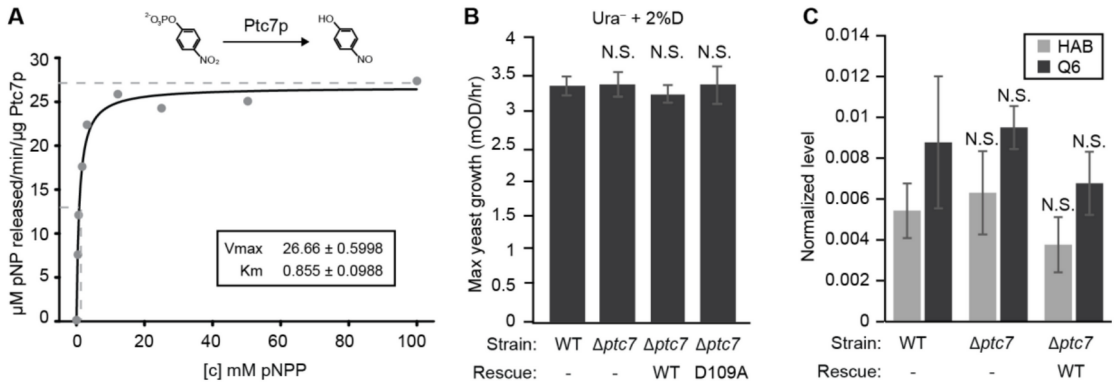

Figure S6:1: Ptc7p is an active phosphatase in vitro. Δptc7 has normal growth rate in dextrose and normal coenzyme Q levels.

A) Kinetic curve of in vitro phosphatase assay of recombinant Ptc7p on para-nitrophenyl phosphate (pNPP). The inset shows V_{max} and K_m values. **B)** Maximum growth rate of WT and Δ*ptc7* in Ura - synthetic media containing 2% dextrose (mean ± SD, n=10). Rescue indicates if the strain is expressing a plasmid containing Ptc7p (WT: wild type Ptc7p; D109A:catalytically inactive mutant of Ptc7p). **C)** Normalized level of 3-hexaprenyl-4-aminobenzoic acid (HAB), and coenzyme Q (Q6) in three strains (WT, Δ*ptc7*, and Δ*ptc7* expressing WT Ptc7p, mean ± SD, n=3).

* p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; N.S., not significant.

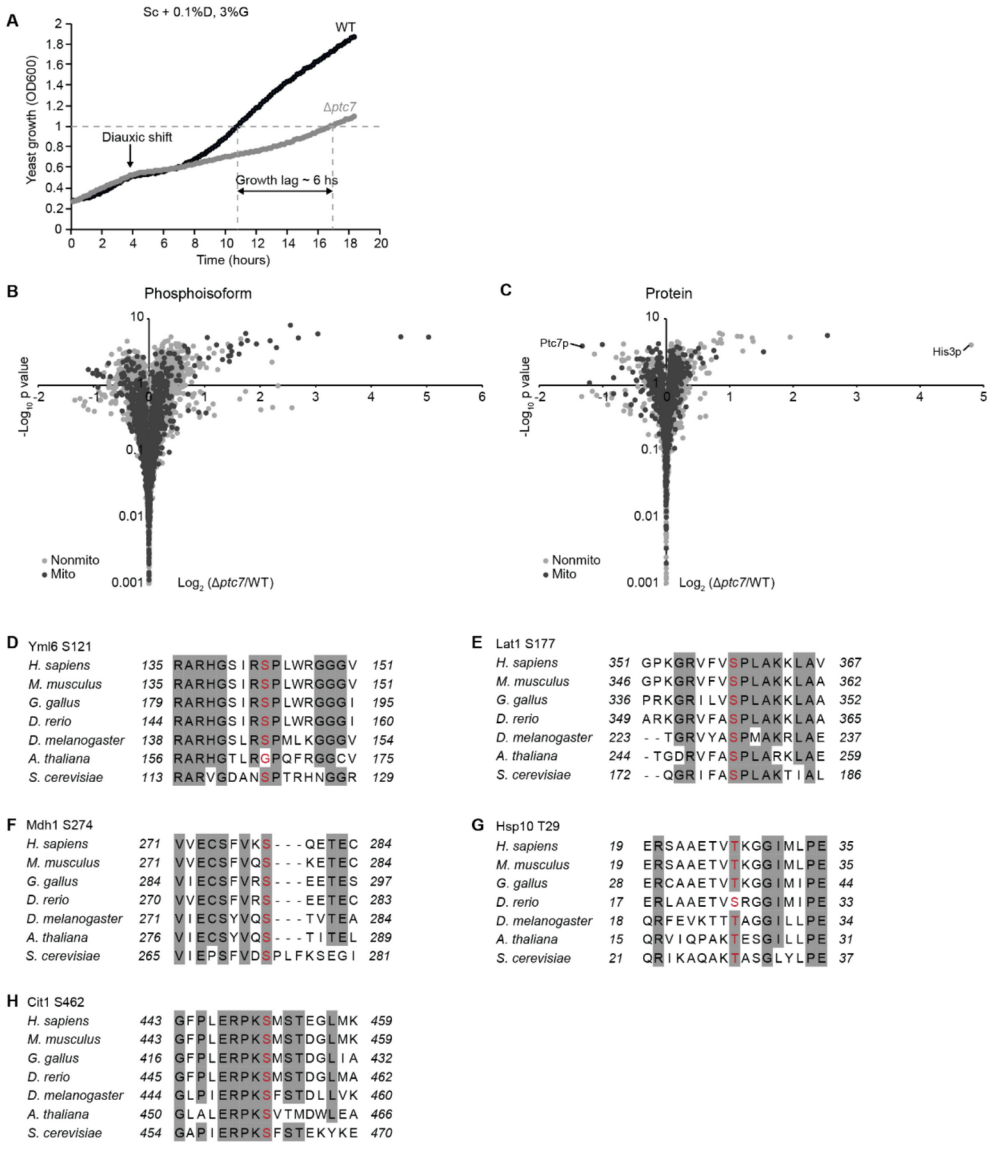

Figure S6:2: Proteomics and phosphoproteomics results.

A) Growth curve of WT and Δ*ptc7* in synthetic complete (SC) media with 0.1% dextrose and 3% glycerol. Diauxic shift and growth lag are indicated on the plot. **B)** Fold changes in phosphoisoforms abundances ($\log_2$ (Δ*ptc7*/WT), n=3) versus statistical significance (-

$\log_{10}$(p-value)). Both mitochondrial (mito) and non-mitochondrial (nonmito) proteins are plotted. **C)** Fold changes in protein abundances ($\log_2$(Δ*ptc7*/WT), n=3) versus statistical significance (-$\log_{10}$(p-value)). Ptc7p and His3p are highlighted. **D**-**H)** Homology sequence alignments of Yml6, Lat1, Mdh1, Hsp10, and Cit1. Phosphosites are highlighted in red.

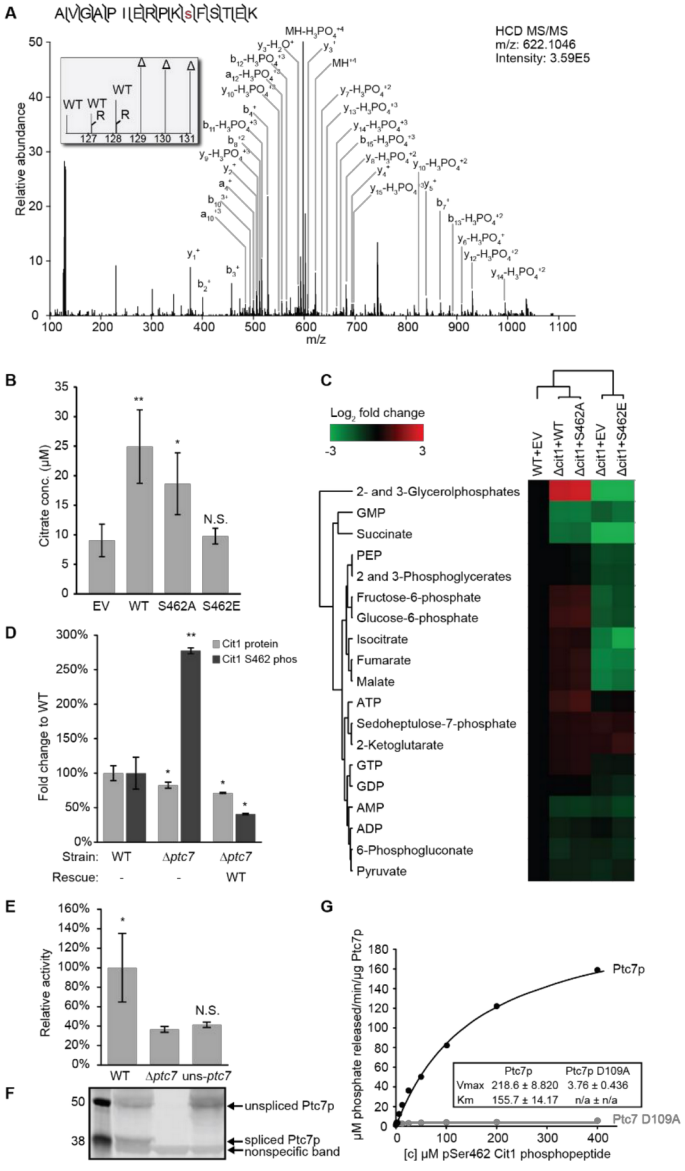

Figure S6:3: A phosphomimetic mutant of Cit1p exhibits a distinguished metabolite profile.

A) MS/MS spectrum identifies phosphorylation of serine 462 (S462) on Cit1p. The inset shows TMT reporter ions from eight channels representing wild type (WT), Δ*ptc7* (Δ) and

Δptc7 rescued with Ptc7p (R). **B)** Citrate concentrate of Δ*cit1* cell lysate expressing empty vector (EV), wild type (WT), S462A, or S462E Cit1p (mean ± SD, n=4). **C)** Hierarchical clustering of WT strain expressing empty vector, Δ*cit1* strain expressing EV, WT, S462A, and S462E Cit1p. Values are $\log_2$ fold change of corresponding strain compared to WT strain expressing empty vector (mean ± SD, n=3). GMP, guanosine monophosphate; PEP, phosphoenolpyruvate; ATP, adenosine triphosphate; GTP, guanosine triphosphate; GDP, guanosine diphosphate; AMP, adenosine monophosphate; ADP, adenosine diphosphate. **D)** Fold change of both Cit1 protein and Cit1 S462 phosphorylation from MS dataset (mean ± SD, n=3 or 2). Rescue indicates if the strain is expressing a plasmid containing Ptc7p (WT: wild type Ptc7p). **E)** Citrate synthase activity of cell lysate from WT, Δptc7, or unspliceable *ptc7* (uns-*ptc7*) strains (mean ± SD, n=4). The p-values are calculated with comparison to the activity from Δ*ptc7* strain. **F)** Immunoblotting against HA (Ptc7p-HA) of cell lysate from WT, Δ*ptc7*, or uns-*ptc7* strains. Both WT and uns-*ptc7* have an endogenous HA tag on the C-terminus of Ptc7p. The molecular weights (kDa) are labeled (left). Both spliced and unspliced Ptc7p, as well as a non-specific band are labeled (right). **G)** Kinetic curve of *in vitro* phosphatase assay of recombinant Ptc7p or Ptc7p D109A on synthetic Cit1p phosphopeptide. The inset shows V_{max} and K_m values.

* p-value < 0.05; ** p-value < 0.01; *** p-value < 0.001; N.S., not significant.

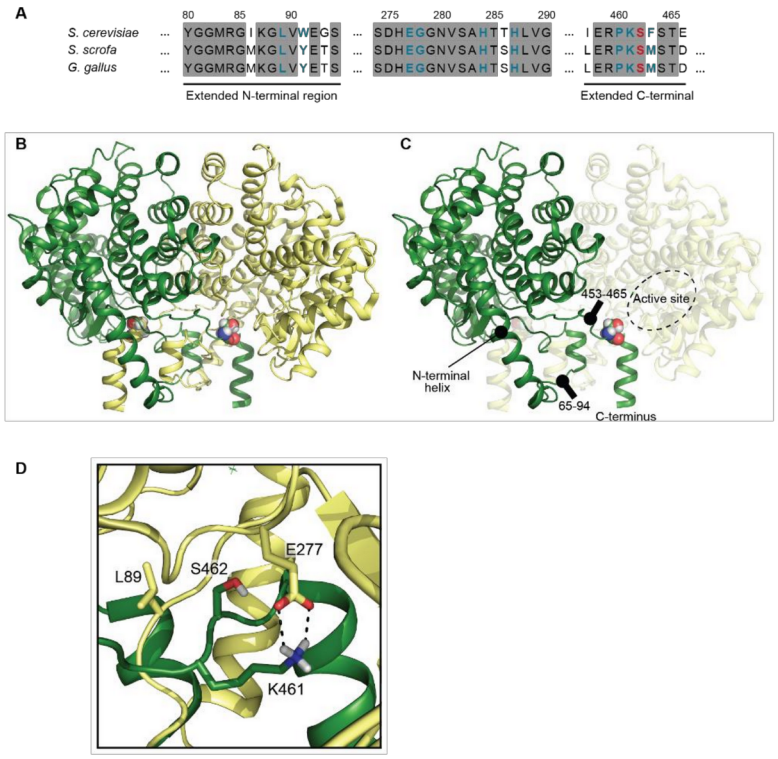

Figure S6:4: Serine 462 belongs to an extended region of the dimer interface.

A) The amino acid positions (highlighted in blue) that contact S462 (highlighted in red) are highly conserved among the yeast, pig and chicken proteins. **B)** Structure of dimeric citrate synthase. S462 is shown in spheres. Figure based on the structure of pig heart synthase (pdb code: 3ENJ). **C)** Same structure in which subunit A (green) has been enhanced, showing that part of the dimer interface consists of two extended and highly interdigitated regions of contact between the two monomers. These consist of the N-terminal helix (residues 46-64) followed by long unstructured region (65-94), and of a C-terminal unstructured region (453-465) followed by a C-terminal helix. Figure based on

the structure of pig heart synthase (pdb code: 3ENJ), using yeast sequence numbering. **D)** Highlight of the region surrounding S462. The position is in proximity of an inter-subunit salt bridge (E277 and K461). S462 is in contact also with L89, in the extended N-terminal region of the opposite subunit.

6.7.1 References

Baykov, A.A., Evtushenko, O.A., and Avaeva, S.M. (1988). A malachite green procedure for orthophosphate determination and its use in alkaline phosphatase-based enzyme immunoassay. Anal. Biochem. *171*, 266–270.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., et al. (2014). SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. *42*, W252–W258.

Blommel, P.G., Martin, P.A., Seder, K.D., Wrobel, R.L., and Fox, B.G. (2009). Flexi Vector Cloning. In High Throughput Protein Expression and Purification, S.A. Doyle, ed. (Totowa, NJ: Humana Press), pp. 55–73.

Bork, P., Brown, N.P., Hegyi, H., and Schultz, J. (1996). The protein phosphatase 2C (PP2C) superfamily: Detection of bacterial homologues. Protein Sci. *5*, 1421–1425.

Brooks, B.R., Brooks, C.L., Mackerell, A.D., Nilsson, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., et al. (2009). CHARMM: the biomolecular simulation program. J. Comput. Chem. *30*, 1545–1614.

Calvo, S.E., Clauser, K.R., and Mootha, V.K. (2016). MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res. *44*, D1251–D1257.

Chen, O.S., Hemenway, S., and Kaplan, J. (2002). Genetic analysis of iron citrate toxicity in yeast: Implications for mammalian iron homeostasis. Proc. Natl. Acad. Sci. *99*, 16922–16927.

Claros, M.G., and Vincens, P. (1996). Computational Method to Predict Mitochondrially Imported Proteins and their Targeting Sequences. Eur. J. Biochem. *241*, 779–786.

Daniel Gietz, R., and Woods, R.A. (2002). Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. In Methods in Enzymology, (Elsevier), pp. 87–96.

Das, A.K., Helps, N.R., Cohen, P.T., and Barford, D. (1996). Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 A resolution. EMBO J. *15*, 6798–6809.

Doulias, P.-T., Tenopoulou, M., Greene, J.L., Raju, K., and Ischiropoulos, H. (2013). Nitric Oxide Regulates Mitochondrial Fatty Acid Metabolism Through Reversible Protein S-Nitrosylation. Sci. Signal. *6*, rs1–rs1.

Fox, B.G., and Blommel, P.G. (2009). Autoinduction of Protein Expression. In Current Protocols in Protein Science, J.E. Coligan, B.M. Dunn, D.W. Speicher, and P.T. Wingfield, eds. (Hoboken, NJ, USA: John Wiley & Sons, Inc.), pp. 5.23.1-5.23.18.

Fuhs, S.R., Meisenhelder, J., Aslanian, A., Ma, L., Zagorska, A., Stankova, M., Binnie, A., Al-Obeidi, F., Mauger, J., Lemke, G., et al. (2015). Monoclonal 1- and 3-Phosphohistidine Antibodies: New Tools to Study Histidine Phosphorylation. Cell *162*, 198–210.

Geer, L.Y., Markey, S.P., Kowalak, J.A., Wagner, L., Xu, M., Maynard, D.M., Yang, X., Shi, W., and Bryant, S.H. (2004). Open Mass Spectrometry Search Algorithm. J. Proteome Res. *3*, 958–964.

Geladopoulos, T.P., Sotiroudis, T.G., and Evangelopoulos, A.E. (1991). A malachite green colorimetric assay for protein phosphatase activity. Anal. Biochem. *192*, 112–116.

Grimsrud, P.A., Carson, J.J., Hebert, A.S., Hubler, S.L., Niemi, N.M., Bailey, D.J., Jochem, A., Stapleton, D.S., Keller, M.P., Westphall, M.S., et al. (2012). A Quantitative Map of the Liver Mitochondrial Phosphoproteome Reveals Posttranslational Control of Ketogenesis. Cell Metab. *16*, 672–683.

He, W., Newman, J.C., Wang, M.Z., Ho, L., and Verdin, E. (2012). Mitochondrial sirtuins: regulators of protein acylation and metabolism. Trends Endocrinol. Metab. *23*, 467–476.

He, X., Zhu, X., Wang, X., Wang, W., Dai, Y., and Yan, Q. (2013). Nuclear Modifier MTO2 Modulates the Aminoglycoside-Sensitivity of Mitochondrial 15S rRNA C1477G Mutation in Saccharomyces cerevisiae. PLoS ONE *8*, e81490.

Hebert, A.S., Richards, A.L., Bailey, D.J., Ulbrich, A., Coughlin, E.E., Westphall, M.S., and Coon, J.J. (2014). The One Hour Yeast Proteome. Mol. Cell. Proteomics *13*, 339–347.

Juneau, K., Nislow, C., and Davis, R.W. (2009). Alternative Splicing of PTC7 in *Saccharomyces cerevisiae* Determines Protein Localization. Genetics *183*, 185–194.

Klock, H.E., and Lesley, S.A. (2009). The Polymerase Incomplete Primer Extension (PIPE) Method Applied to High-Throughput Cloning and Site-Directed Mutagenesis. In High Throughput Protein Expression and Purification, S.A. Doyle, ed. (Totowa, NJ: Humana Press), pp. 91–103.

Kulp, D.W., Subramaniam, S., Donald, J.E., Hannigan, B.T., Mueller, B.K., Grigoryan, G., and Senes, A. (2012). Structural informatics, modeling, and design with an open-source Molecular Software Library (MSL). J. Comput. Chem. *33*, 1645–1661.

Larson, S.B., Day, J.S., Nguyen, C., Cudney, R., and McPherson, A. (2009). Structure of pig heart citrate synthase at 1.78 Å resolution. Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun. *65*, 430–434.

Liao, D.I., Karpusas, M., and Remington, S.J. (1991). Crystal structure of an open conformation of citrate synthase from chicken heart at 2.8-.ANG. resolution. Biochemistry *30*, 6031–6036.

Lin, J., Handschin, C., and Spiegelman, B.M. (2005). Metabolic control through the PGC-1 family of transcription coactivators. Cell Metab. *1*, 361–370.

Linn, T.C., Pettit, F.H., Hucho, F., and Reed, L.J. (1969). -KETO ACID DEHYDROGENASE COMPLEXES, XI. COMPARATIVE STUDIES OF REGULATORY PROPERTIES OF THE PYRUVATE DEHYDROGENASE COMPLEXES FROM KIDNEY, HEART, AND LIVER MITOCHONDRIA. Proc. Natl. Acad. Sci. *64*, 227–234.

Longtine, M.S., Mckenzie III, A., Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. (1998). Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. Yeast *14*, 953–961.

Mailloux, R.J., Jin, X., and Willmore, W.G. (2014). Redox regulation of mitochondrial function with emphasis on cysteine oxidation reactions. Redox Biol. *2*, 123–139.

Martín-Montalvo, A., González-Mariscal, I., Pomares-Viciana, T., Padilla-López, S., Ballesteros, M., Vazquez-Fonseca, L., Gandolfo, P., Brautigan, D.L., Navas, P., and Santos-Ocaña, C. (2013). The Phosphatase Ptc7 Induces Coenzyme Q Biosynthesis by Activating the Hydroxylase Coq7 in Yeast. J. Biol. Chem. *288*, 28126–28137.

McAvoy, T., and Nairn, A.C. (2010). Serine/Threonine Protein Phosphatase Assays. In Current Protocols in Molecular Biology, F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, and K. Struhl, eds. (Hoboken, NJ, USA: John Wiley & Sons, Inc.), p.

Nakatsukasa, K., Nishimura, T., Byrne, S.D., Okamoto, M., Takahashi-Nakaguchi, A., Chibana, H., Okumura, F., and Kamura, T. (2015). The Ubiquitin Ligase SCFUcc1 Acts as a Metabolic Switch for the Glyoxylate Cycle. Mol. Cell *59*, 22–34.

Nesvizhskii, A.I., and Aebersold, R. (2005). Interpretation of Shotgun Proteomic Data: The Protein Inference Problem. Mol. Cell. Proteomics *4*, 1419–1440.

Nunnari, J., and Suomalainen, A. (2012). Mitochondria: In Sickness and in Health. Cell *148*, 1145–1159.

O’Rourke, B., Van Eyk, J.E., and Foster, D.B. (2011). Mitochondrial Protein Phosphorylation as a Regulatory Modality: Implications for Mitochondrial Dysfunction in Heart Failure: mitochondrial protein phosphorylation. Congest. Heart Fail. *17*, 269–282.

Pagliarini, D.J., and Dixon, J.E. (2006). Mitochondrial modulation: reversible phosphorylation takes center stage? Trends Biochem. Sci. *31*, 26–34.

Pagliarini, D.J., Calvo, S.E., Chang, B., Sheth, S.A., Vafai, S.B., Ong, S.-E., Walford, G.A., Sugiana, C., Boneh, A., Chen, W.K., et al. (2008). A Mitochondrial Protein Compendium Elucidates Complex I Disease Biology. Cell *134*, 112–123.

Phanstiel, D.H., Brumbaugh, J., Wenger, C.D., Tian, S., Probasco, M.D., Bailey, D.J., Swaney, D.L., Tervo, M.A., Bolin, J.M., Ruotti, V., et al. (2011). Proteomic and phosphoproteomic comparison of human ES and iPS cells. Nat. Methods *8*, 821–827.

Phillips, D., Aponte, A.M., Covian, R., and Balaban, R.S. (2011). Intrinsic Protein Kinase Activity in Mitochondrial Oxidative Phosphorylation Complexes. Biochemistry *50*, 2515–2529.

Pirman, N.L., Barber, K.W., Aerni, H.R., Ma, N.J., Haimovich, A.D., Rogulina, S., Isaacs, F.J., and Rinehart, J. (2015). A flexible codon in genomically recoded Escherichia coli permits programmable protein phosphorylation. Nat. Commun. *6*.

Quirós, P.M., Langer, T., and López-Otín, C. (2015). New roles for mitochondrial proteases in health, ageing and disease. Nat. Rev. Mol. Cell Biol. *16*, 345–359.

Rardin, M.J., Wiley, S.E., Murphy, A.N., Pagliarini, D.J., and Dixon, J.E. (2008). Dual Specificity Phosphatases 18 and 21 Target to Opposing Sides of the Mitochondrial Inner Membrane. J. Biol. Chem. *283*, 15440–15450.

Rhee, H.-W., Zou, P., Udeshi, N.D., Martell, J.D., Mootha, V.K., Carr, S.A., and Ting, A.Y. (2013). Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. Science *339*, 1328–1331.

Richards, A.L., Hebert, A.S., Ulbrich, A., Bailey, D.J., Coughlin, E.E., Westphall, M.S., and Coon, J.J. (2015). One-hour proteome analysis in yeast. Nat. Protoc. *10*, 701–714.

Schmidt, O., Harbauer, A.B., Rao, S., Eyrich, B., Zahedi, R.P., Stojanovski, D., Schönfisch, B., Guiard, B., Sickmann, A., Pfanner, N., et al. (2011). Regulation of Mitochondrial Protein Import by Cytosolic Kinases. Cell *144*, 227–239.

Schmidtmann, E., König, A.-C., Orwat, A., Leister, D., Hartl, M., and Finkemeier, I. (2014). Redox Regulation of Arabidopsis Mitochondrial Citrate Synthase. Mol. Plant *7*, 156–169.

Srere, P.A. (1969). [1] Citrate synthase. In Methods in Enzymology, (Elsevier), pp. 3–11.

Stefely, J.A., Reidenbach, A.G., Ulbrich, A., Oruganty, K., Floyd, B.J., Jochem, A., Saunders, J.M., Johnson, I.E., Minogue, C.E., Wrobel, R.L., et al. (2015). Mitochondrial ADCK3 Employs an Atypical Protein Kinase-like Fold to Enable Coenzyme Q Biosynthesis. Mol. Cell *57*, 83–94.

Storici, F., Lewis, L.K., and Resnick, M.A. (2001). In vivo site-directed mutagenesis using oligonucleotides. Nat. Biotechnol. *19*, 773–776.

Subramaniam, S., and Senes, A. (2012). An energy-based conformer library for side chain optimization: Improved prediction and adjustable sampling. Proteins Struct. Funct. Bioinforma. *80*, 2218–2234.

Vögtle, F.-N., Wortelkamp, S., Zahedi, R.P., Becker, D., Leidhold, C., Gevaert, K., Kellermann, J., Voos, W., Sickmann, A., Pfanner, N., et al. (2009). Global Analysis of the Mitochondrial N-Proteome Identifies a Processing Peptidase Critical for Protein Stability. Cell *139*, 428–439.

Wenger, C.D., Phanstiel, D.H., Lee, M.V., Bailey, D.J., and Coon, J.J. (2011). COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA. PROTEOMICS *11*, 1064–1074.

Winzeler, E.A. (1999). Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. Science *285*, 901–906.

Appendix I: A comparison of manual and computational methods for detecting photobleaching steps in single-molecule fluorescence traces

In this appendix, I describe the testing and development of a step-detection algorithm to facilitate single-particle photobleaching analysis. The software was included in the AGATHA package, which is described in the following publication:

Kaur H., Jamalidinan F., Condon S.G.F., Senes A., Hoskins A.A., 2019. “Analysis of spliceosome dynamics by maximum likelihood fitting of dwell time distributions.” *Methods* 15;153:13-21

Statement of contribution

I generated the simulated photobleaching traces, designed the benchmarking experiments, performed the data analysis, and wrote MATLAB scripts to automate step detection from single-molecule TIRF microscopy experiments.

Abstract

Biochemical measurements are increasingly performed at the level of individual molecules or macromolecular complexes to better understand the heterogeneous subsets that make up an entire bulk population. It has recently become possible to directly count the number of fluorescently labeled molecules in a single particle through the use of photobleaching. Analysis of this type of data involves detecting stepwise changes in intensity, a common problem across many fields. This process is sometimes performed by eye, but this may introduce subjective biases and is impractical for high-throughput analysis. In this work I use simulated photobleaching data to show that manual counting may be impacted by subjective biases, which may lead to an incorrect assessment of the correct distribution of steps. Fortunately, benchmarking of different step-detection algorithms on both simulated and experimental data demonstrate that better performance in higher throughput can be achieved.

Introduction

Biochemical measurements of enzyme kinetics, protein structure determination, and protein-protein interaction have long relied on bulk techniques that report the average state of an ensemble of particles. These techniques are useful but ignore the possible underlying distributions making up that average. For example, alternate conformational states, fractions of unfolded or inactive protein, and different oligomeric states may make up subpopulations of the whole set, skewing the results. Ion channels have been studied at the single-molecule level for quite some time (Hamill et al., 1981), but recent advances in fluorescense microscopy have enabled the analysis of other individual macromolecules.

One application of single-molecule fluorescence microscopy is the analysis of the oligomeric states of molecules, such as the cell division proteins FtsB and FtsL. Bulk measurements of their interactions have suggested that FtsB forms a weak homodimer (LaPointe et al., 2013) and that FtsB and FtsL cooperatively associate as a 1:1 higher-order oligomer (Khadria and Senes, 2013). However, it may be possible that fractions of higher-order oligomers and protein aggregates are skewing the observed results. Counting the number of subunits through single-molecule techniques will help determine how FtsB and FtsL interact with one another and potentially provide insight as to how these membrane proteins participate in cell division.

To count subunits in a macromolecular complex through photobleaching, the molecules of interest are labeled with a fluorescent molecule and bound to a glass coverslip at a low enough dilution that single particles can be resolved as diffraction-limited spots on a Total Internal Reflection Fluorescense (TIRF) microscope (Chen et al., 2014). Excitation of the fluorophores over time will lead to photobleaching and determining the number of times a spot undergoes this process is a direct measure of the number of fluorophores in that spot (Arant and Ulbrich, 2014). Photobleaching of individual fluorescent molecules is a stochastic, all-or-nothing process, so measuring the fluorescence intensity of a spot over time generates a characteristic piecewise constant function with some level of noise (Figure 1).

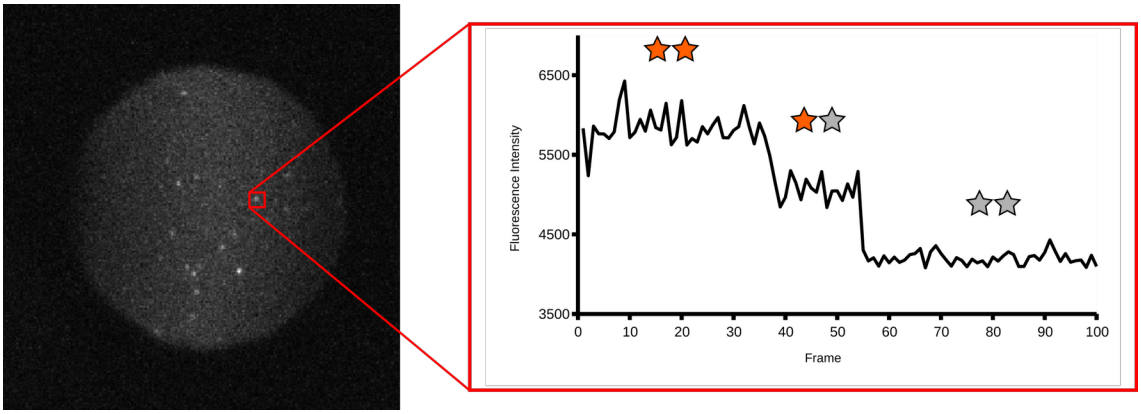

Figure 1, left: A typical field of view collected from a TIRF microscopy single-molecule photobleaching experiment, in this case analyzing the cell division proteins FtsB and FtsL in liposomes. Bright spots in the field of view correspond to single fluorescent particles. Right: exciting the fluorophores (orange stars) over time leads to stochastic photobleaching (gray stars), producing a characteristic step curve. Two steps are clearly observed in this trace, indicating the presence of two fluorescently labeled molecules.

The simplest method to detect steps in these noisy photobleaching traces is to examine the fluorescence intensity traces by eye (Carter et al., 2008). This is a straightforward procedure and useful for identifying patterns in these complex data, but there are significant limitations. For example, it is possible that there are subjective biases that lead to miscounting the number of steps, which would lead to an incorrect analysis of the molecule's oligomeric state. Counting steps by hand is also extremely tedious and time-consuming, preventing analysis of more than a few hundred traces per experiment. Fortunately, step detection is a common challenge across a wide variety of disciplines like image processing (Shen and Castan, 1992) and motion detection (Brajdic and Harle, 2013). Step detection is also important for other biophysical processes like

analyzing the movement of molecular motors (Carter et al., 2008). As such, a large number of computational algorithms have been developed to accurately and objectively automate this process.

One simple way to count steps computationally is velocity thresholding. As steps are characterized by relatively flat segments with steep changes between them, it is possible to identify steps by using a filter to reduce noise and then find peaks corresponding to steep regions of change above some threshold (Carter et al., 2008). At low signal to noise ratios, velocity thresholding performs quite well. However, it is easy to imagine scenarios where velocity thresholding would fail, such as when steps are closely spaced or when steps can have a wide range of sizes. Additionally, this protocol is likely to miscount steps when the signal to noise ratio is high because the actual step may be less prominent.

The tDetector2 algorithm was developed to count photobleaching steps for cellulose synthesis complex proteins tagged with green fluorescent protein (Chen et al., 2014). In this method, a two-sample t-test is used to distinguish plateaus between steps, which are generated by iteratively searching through the trace for statistically significant changes. One feature of tDetector2 is that it does not assume equal noise variance across the trace, as it has been observed that the observed noise in a photobleaching signal increases with mean fluorescense intensity (Chen et al., 2014).

GLS-STEPS was developed to count steps from molecular motors collected by optical tweezers (Arunajadai and Cheng, 2013). Arunajadai and Cheng demonstrated that the noise present in this type of data has significant autocorrelation, which can lead to dramatic overfitting or underfitting of the data to a step curve. Explicitly modeling the noise as dependent on the

previous values with a constant mean allowed for the estimation of step sizes based on least-squares regression. To avoid over-fitting, the number of steps in the trace are evaluated based on an information criterion parameter.

Pottslab is a MATLAB toolbox used to perform image segmentation (Weinmann et al., 2015). It is also used for reconstructing step traces via jump-sparse minimization and is highly robust to different sources of noise, including salt and pepper noise, Laplacian noise, and Gaussian noise.

To determine if step-detection algorithms are more accurate than manual detection, I generated simulated photobleaching traces to benchmark the methods. These traces were analyzed both manually by multiple experimenters in the Senes Lab as well as by the step-detection algorithms velocityThreshold, tDetector2, and GLS-STEPS. I also analyzed the performance of these algorithms on real photobleaching data collected from the bacterial cell division proteins FtsB and FtsL.

Methods

Generation of simulated photobleaching trace data

A Perl script was written to generate simulated data that qualitatively reflect photobleaching traces from a typical experiment. For a trace, 0-5 steps were placed at randomly selected points within a 100 frame window. For simplicity, the size of each step within a trace was kept constant. Next, 100 Gaussian random values were generated with a mean of zero and a standard deviation proportional to 0.4 times the step size, which was subjectively determined to be the highest noise ratio that steps could still be readily distinguished by eye. These values were added to the steps to produce the final noisy trace. The sizes of the steps for a trace and the initial value for the first step were randomly assigned to traces to obscure the axes. The trace ID along with the number

and location of steps was written to a separate file. 100 of these traces were generated for benchmarking. A representative trace generated by this script is shown in Figure 2.

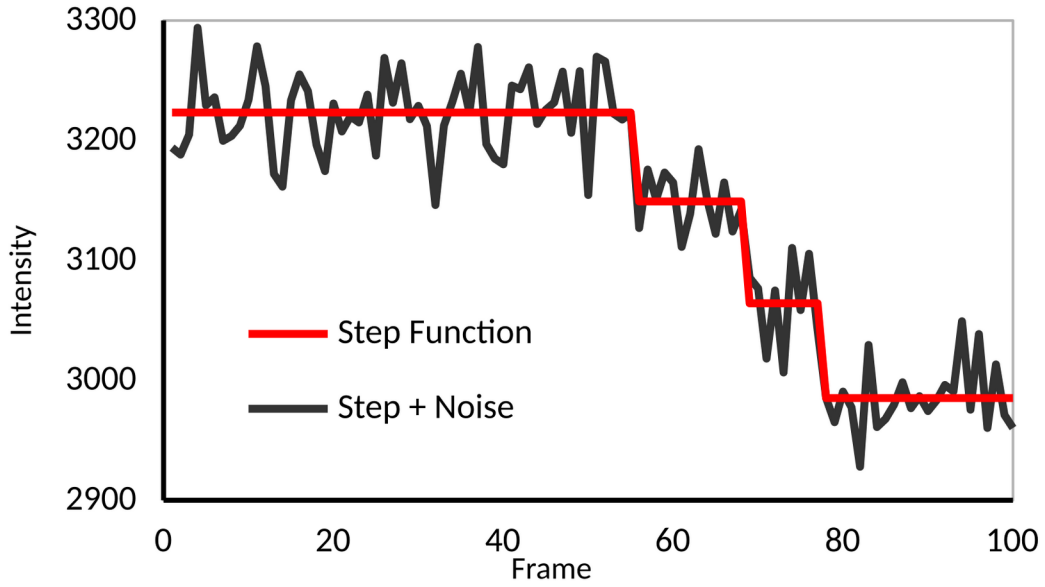

Figure 2: Representative simulated photobleaching trace. The ideal step function without noise is shown in red and the final trace with Gaussian noise added is shown in red.

Acquisition of FtsB-FtsL photobleaching traces

Briefly, the cell division proteins FtsB and FtsL were co-overexpressed in an inducible strain of E. coli and copurified via affinity chromatography. FtsB was selectively labeled with the fluorescent dye Alexa Fluor 633. The mixture of labeled FtsB and unlabeled FtsL was lyophilized and mixed in a 1:10000 protein:lipid ratio with a 2:1 mixture of the phospholipids POPE and POPG, which was then sonicated to produce liposomes with membrane-embedded FtsB and FtsL molecules. These liposomes were then bound to a glass coverslip and placed under a TIRF microscope. The fluorescent label was excited at 633nm and the emission intensity of a field of view at 647nm was recorded with a CCD detector for 1s in 10ms intervals. Photobleaching traces for areas of interest within the field of view were extracted using the

GLIMPSE-IMSCROLL program (Friedman and Gelles, 2015). Five fields of view were analyzed and from these, 85 traces were collected.

Step detection

Manual: Photobleaching traces were plotted using the “plot” function in MATLAB with default parameters and saved as PNG images. Steps were identified manually by experimenters from these images. For the FtsB-FtsL photobleaching traces, one experimenter (SC) counted steps. For the simulated traces, two different researchers experienced in single-molecule photobleaching analysis (identified here as CA and AK) counted steps independently from one another. These researchers were not informed of the methods used to generate the simulated traces prior to counting.

VelocityThreshold: the velocityThreshold (vT) algorithm as described in (Carter et al., 2008) was implemented as a MATLAB script for this work. Traces were filtered with a second-order Savitzky-Golay filter with a window size of 5 before calculating the absolute velocity. 50 evenly distributed candidate velocity thresholds were compared between the average velocity for the trace and the maximum velocity. Carter et al. used a subjective heuristic to identify the optimum threshold from these values by identifying where the plot of threshold vs number of steps “flattens out”. In this work, this optimal threshold was chosen to be based on the mode value of the number of steps detected within those 50 candidate thresholds.

GLS-STEPS: The GLS-STEPS (GLS) algorithm was used with valid windows ranging from 2-98 frames in increments of 1. The script was modified to not crash when zero steps were detected in a trace and to print PNG files of the data with a superimposed ideal step function instead of PDF files.

tDetector2: the tDetector2 (tD2) algorithm accepts a vector of doubles corresponding to the trace and a parameter to either assume equal variance in the noise (1, default) or allow it to vary (2). For this work the variance was allowed to vary.

Pottslab: The fluorescence traces were first normalized so that all intensity values were between 0 and 1. The minL1Potts function was then used to fit a step function to the trace, with the sensitivity parameter set to 0.6.

Results

GLS-STEPS produces poor fits of traces when not assuming Gaussian noise

One of the supposed key advantages of the GLS-STEPS algorithm is that it does not necessarily assume that the noise in a trace is uncorrelated and Gaussian. However, when run using the default parameters on experimental data, it tended to vastly overcount the number of steps in ways that did not seem to follow the photobleaching traces at all (Figure 3). An extra parameter had to be set to force the algorithm assume Gaussian noise. This did lead to somewhat improved agreement between the step function generated by GLS-STEPS and the observed data for both the simulated and experimental traces. As such, the distribution of steps calculated by GLS-STEPS when assuming autocorrelated noise was not determined.

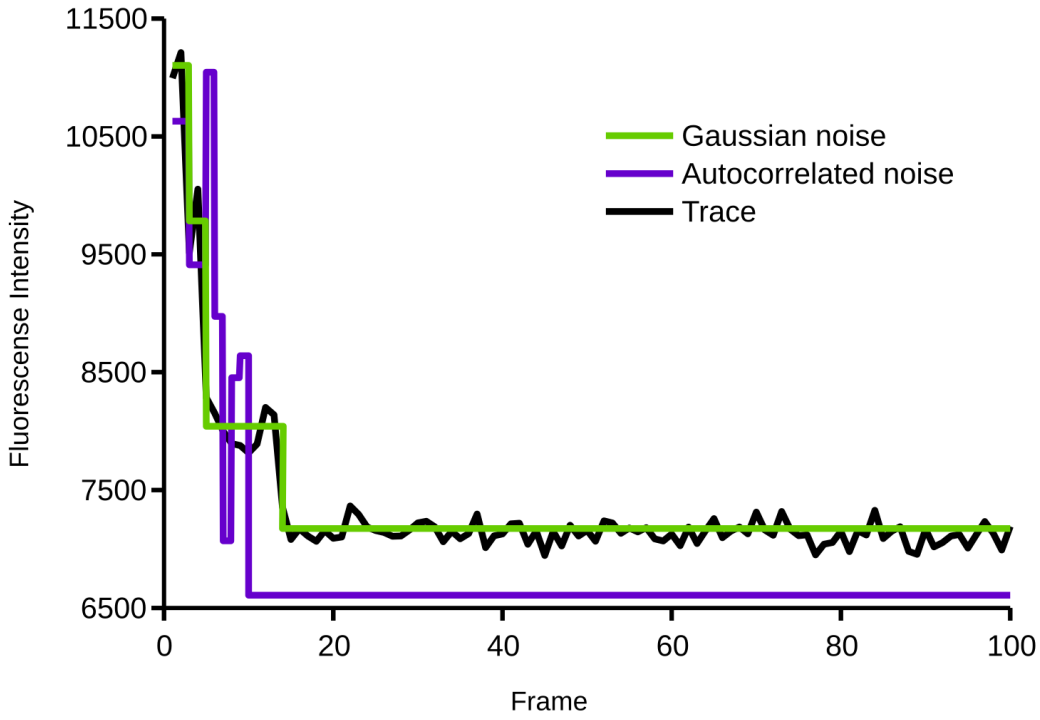

Figure 3: Fits of a representative experimental photobleaching trace (black) by the GLS-STEPS algorithm under the assumption of Gaussian noise (green) or autocorrelated noise (purple).

The three algorithms counted a different distribution of steps from manual counting of the experimental traces

From the photobleaching analysis of the FtsB-FtsL complex, 85 spots were identified for step counting. Manual counting identified between 0 and 4 steps for each of the traces, while counting using the three algorithms identified anywhere from 0 to 10 steps per trace (Figure 4). After combining the counts for four or more steps into a single category, a Chi-Square test showed that the distributions identified from each algorithm were statistically different from the distribution that was calculated from manual step counting (Table 1). If the goal is to use

algorithms to replicate the results of manual counting simply to save time for researchers, clearly none of these methods would be suitable for the job. However, it is entirely possible that the manual counts are not completely accurate, as subjective biases might skew the results. To address this and to have a “ground truth” that all of the step-detection methods could be compared to, simulated photobleaching traces were generated and analyzed both by eye and with algorithms.

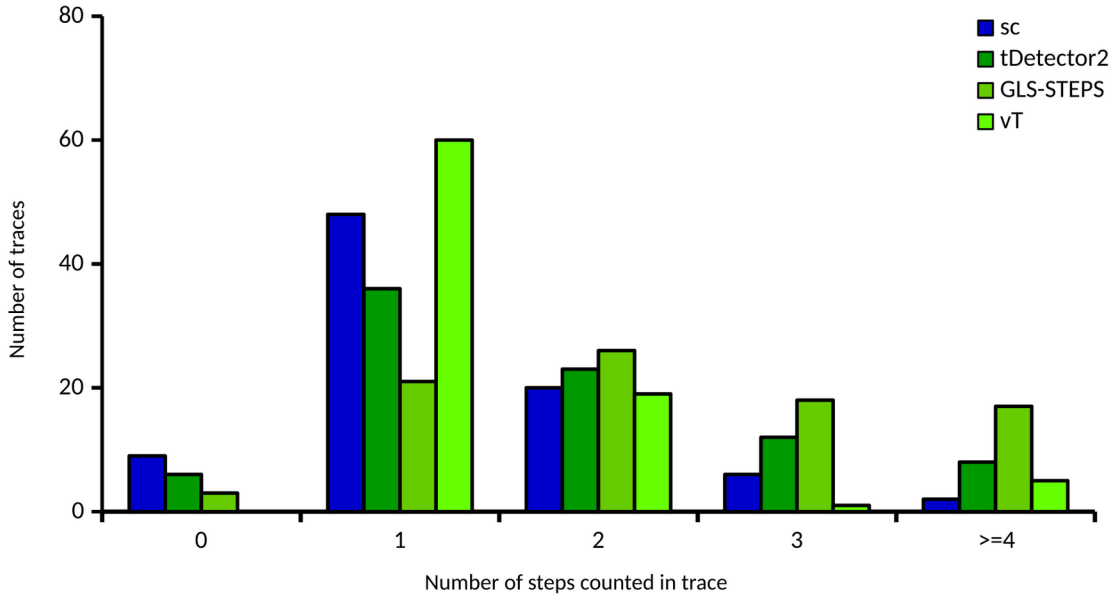

Figure 4: Different step-detection methods counted different distributions of steps from a set of single-molecule FtsB-FtsL photobleaching traces. Step detection algorithms are shown in green while manual counting of steps is shown in blue.

	tD2	GLS	vT
Comparison with manual count	p < 0.0001	p < 0.0001	p = 0.0004

Table 1 : Chi-Square analysis of the observed step counts from three different step-detection algorithms compared to the expected counts from manual counting as shown in Figure 4. DF = 4, N = 85.

Clear differences observed between manual and algorithmic counting

Comparisons made using simulated data should be more useful to benchmark each step counting method, as the actual expected distribution is known. None of the methods were able to count steps with perfect accuracy, though there were large differences between them (Figure 5). Both of the experimenters had a clear tendency to overestimate the number of steps in a trace. Interestingly, CA was trained by AK to count steps. This indicates that having multiple users count steps (Chadda et al., 2016) will not necessarily improve accuracy because training may also pass on biases between researchers. However, since none of the step counters were trained based on known data, it might be possible to improve manual performance by having a ground truth that can be objectively compared to.

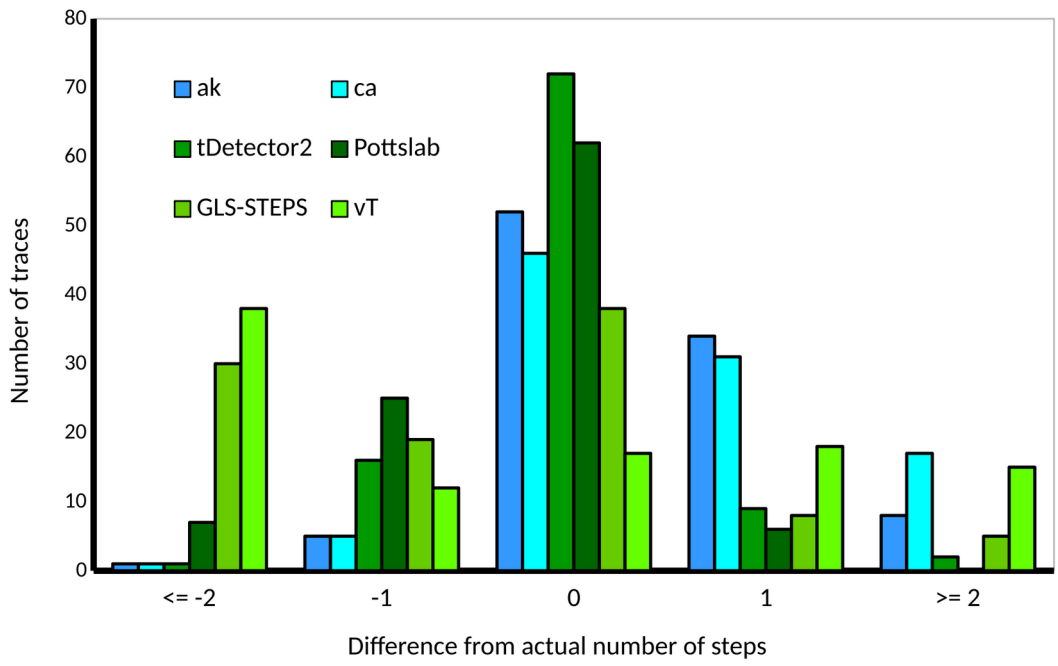

Figure 5: Comparison of steps detected to the known number of steps in simulated photobleaching traces. Computational methods are shown in green while the results of manual counts from different experimenters are shown in blue.

Two of the tested algorithms outperformed manual counting

Of the computational methods tested, velocity thresholding by far performed the worst, identifying the correct number of steps only a fifth of the time and often miscounting the number of steps by two or more. Most of this seems to stem from the algorithm being unable to identify traces with zero steps, as the chosen thresholds to examine are all within the data instead of going outside. Also, the choice of optimal threshold seems to be problematic, as steps can be missed if they are present but have a reduced slope (possibly due to filtering) compared to other steps in the trace.

GLS-STEPS accurately counted steps in twice as many traces as velocityThreshold, though this was still lower than the results obtained from manual counting. In the simulated traces, there was a tendency to undercount, with 30 traces underestimated by two steps or more (Figure 5). Interestingly, this somewhat contrasts with the analysis of experimental traces, as GLS-STEPS identified the highest proportion of traces with 4 or more steps (Figure 4). One likely explanation for these results is that the noise in the experimental traces is not necessarily constant and Gaussian, as was assumed in the simulated traces.

TDetector2 and Pottslab performed similarly to one another and were both more accurate than the manual counts, with TDetector2 coming closest to matching the known distribution of steps. Since TDetector2 assumes Gaussian noise in the trace, the fact that it performed the best in this experiment is not surprising. However, Pottslab is highly robust to non-Gaussian noise as well as blurring (Weinmann et al., 2015). Additionally, the Pottslab toolbox is available on GitHub with extensive documentation and is continuously updated. In contrast, the TDetector2 algorithm must be downloaded separately as a static text file with limited documentation, making it less appealing for general use.

Performance decreases across methods with increasing numbers of steps.

When expecting to count many subunits in a particle, care should be taken to use an algorithm that performs well within that particular distribution. All of the methods except for velocityThresholding were able to approach 60% accuracy when analyzing traces with either 0 or 1 step, but for all of the methods, the fraction of traces that were correctly counted tended to decrease as the number of steps in the trace went up (Figure 6). As the number of data points in each trace was kept constant, the average number of frames between each step went down. Consequently, a trace with many steps looks more like a smooth line rather than a step function and it becomes more difficult to identify the separated plateaus. In order to identify multiple steps in a trace, it is extremely important to have high-quality data with adequate sampling of plateaus.

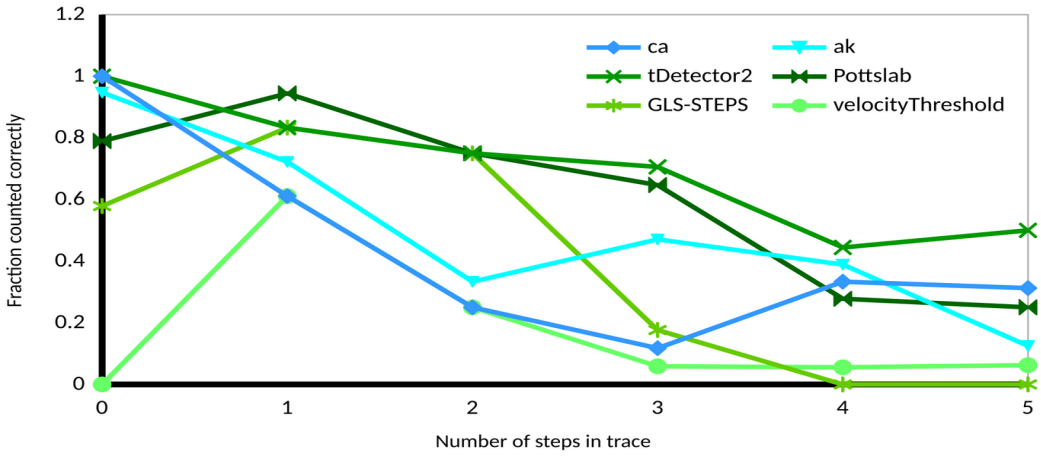

Figure 6: Performance decreases with more steps in a trace. Step detection algorithms are shown in green while results from different researchers counting steps manually is shown in blue.

Development of a step-detection GUI for single-molecule photobleaching analysis

Inspired by these results, we created a graphical user interface (GUI) to use Pottslab to detect photobleaching events in single-particle TIRF microscopy photobleaching experiments (Figure 7). The program loads a series of .dat files generated from the GLIMPSE-Imscroll program (Friedman and Gelles, 2015), each of which corresponds to particles from a field of view during a photobleaching experiment. A step function is fit to each fluorescense trace and the results are output in a series of annotated images for manual verification. Summary statistics are written separately to a TSV file. Additionally, simple heuristics have been added to distinguish photobleaching events from alternate changes of fluorescence intensity. For example, fluorophore blinking is a common phenomenon in which the fluorophore randomly switches between bright and dark states. This behavior can be filtered by only counting stepwise decreases

without corresponding increases in the fluorescence trace (Figure 8). This step-detection GUI was published in a package of other tools for single-molecule analysis called AGATHA (Kaur et al., 2019) and is available on GitHub (https://github.com/hoskinslab/AGATHA).

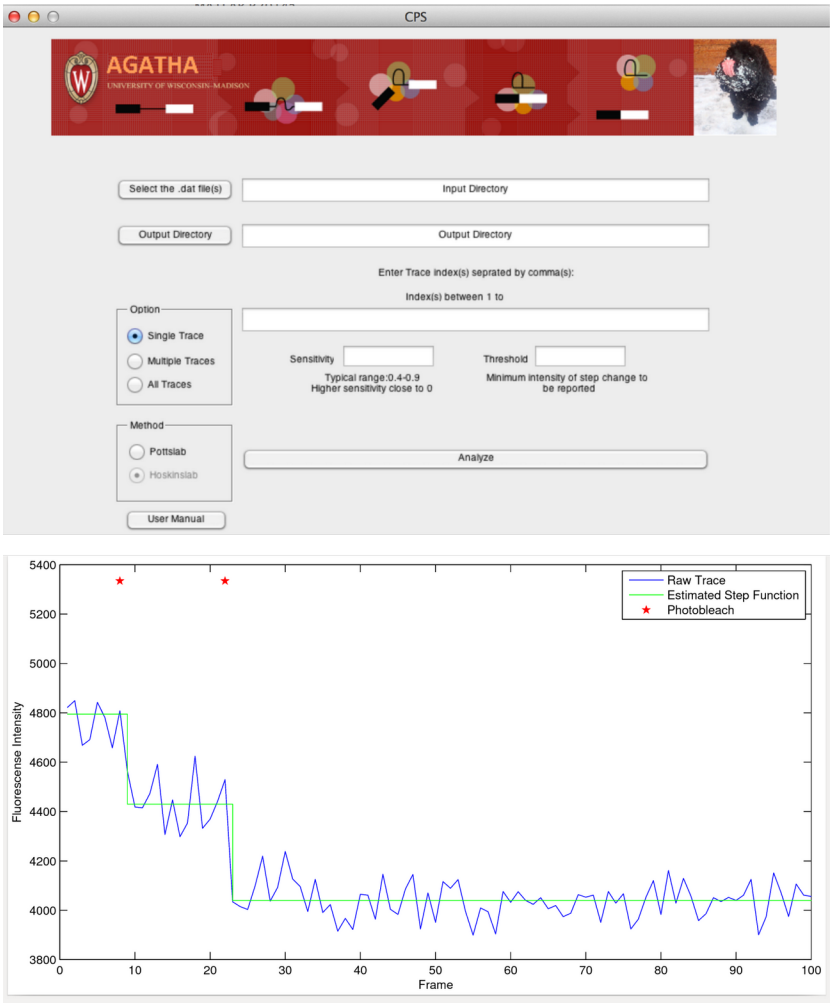

Figure 7: Step-detection GUI for analysis of single-particle photobleaching. Top: interface for data input and parameter selection. Bottom: example output for a single trace. Estimated step function is shown in green and photobleaching events are annotated as red asterisks.

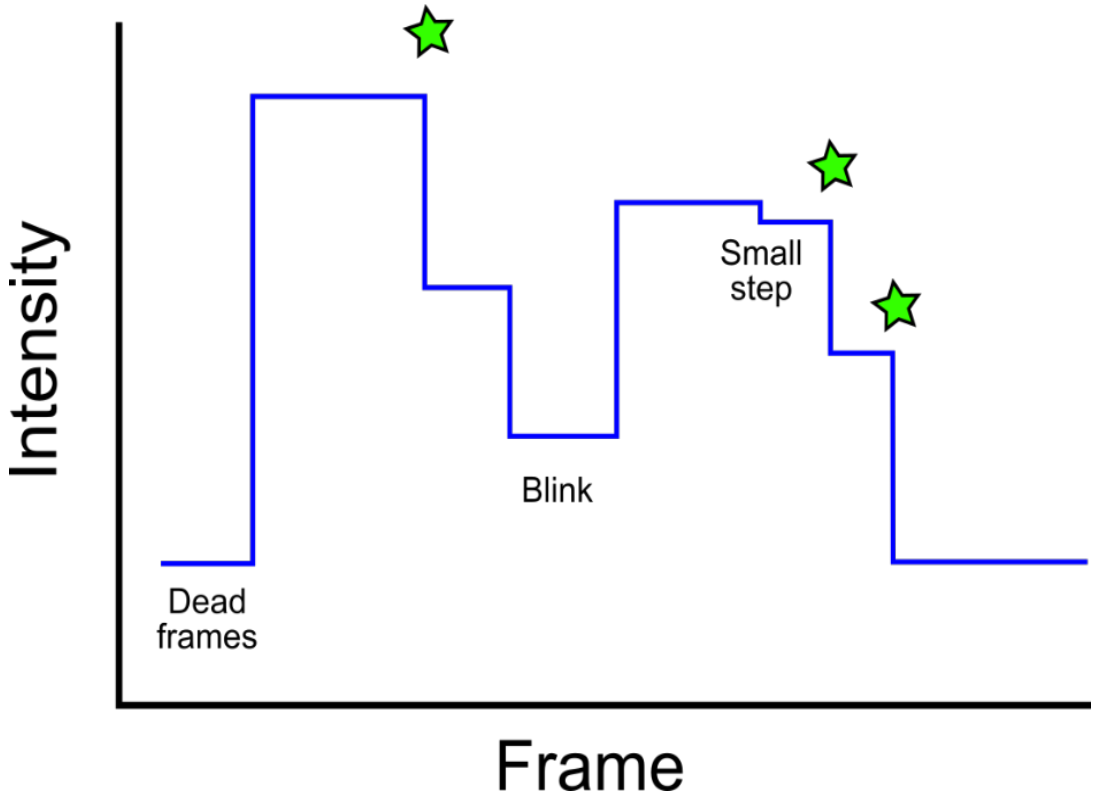

Figure 8: Heuristics to identify photobleaching events. Intensity decreases that are accompanied by corresponding increases are labeled as blinking events. Additionally, dead frames and very small steps (corresponding to background fluctuations) are ignored. Remaining stepwise decreases are identified as photobleadching events (stars).

Discussion

Counting steps to determine the oligomeric state of a molecule requires an accurate and rapid step-detection method. This work clearly demonstrates the importance of method validation and points to some best practices when counting steps in photobleaching traces. This work demonstrates that algorithmic step-detection methods can improve on accuracy while also being much less labor intensive.

The driving force behind the development of the GLS-STEPS algorithm was to identify steps in the presence of correlated noise. This advantage appears to be wasted on the photobleaching traces, as detecting steps when assuming Gaussian noise improved step counts both for the simulated and the experimental photobleaching traces. It also seems likely that GLS-STEPS requires more data to accurately fit steps and model the correlated noise than was provided in these traces. In the original publication, simulated traces with 60,000 data points with an average of 33 steps in each trace were used to benchmark the GLS-STEPS algorithm (Arunajadai and Cheng, 2013). However, the experimental and simulated traces used in this work contain just 100 data points with ~0-5 steps in each trace. Unfortunately, collecting more time points for photobleaching at the same laser power would reduce the observed signal in each frame, as is the case in high-speed photography. Increasing the laser power to account for this would simply lead to more rapid photobleaching. Using oxygen scavengers would likely improve signal quality by decreasing the probability of photobleaching, though this would come at the cost of increased sampling time.

Both Tdetector2 and Pottslab were able to outperform manual counts both in speed and accuracy, but neither algorithm perfectly reconstructed all of the steps. Since the step-detection step is very fast in comparison to data collection, it might also be useful to analyze fluorescense traces with multiple algorithms in order to identify a consensus set of steps. Such a strategy has proven useful for automatically annotating transmembrane domains in proteins (Bernsel et al., 2009) as well as identifying deleterious single nucleotide polymorphisms (Boyle et al., 2012). Future work should also examine the effect of non-Gaussian sources of noise in the traces, as well as methods to better filter small changes in the baseline from genuine photobleaching steps.

A more realistic set of simulated data is likely to be more difficult for each of the methods to analyze, but it is important to have an accurate measure of their capabilities.

Acknowledgments

Special thanks to Ambalika Khadria and Claire Armstrong for counting steps in the simulated photobleaching traces. Additionally, I would like to thank Audra Amasino, Janice Robertson, and members of the Senes and Hoskins laboratories for helpful discussions while working on this project.

References

Arant, R.J., and Ulbrich, M.H. (2014). Deciphering the Subunit Composition of Multimeric Proteins by Counting Photobleaching Steps. ChemPhysChem *15*, 600–605.

Arunajadai, S.G., and Cheng, W. (2013). Step Detection in Single-Molecule Real Time Trajectories Embedded in Correlated Noise. PLoS ONE *8*, e59279.

Bernsel, A., Viklund, H., Hennerdal, A., and Elofsson, A. (2009). TOPCONS: consensus prediction of membrane protein topology. Nucleic Acids Res. *37*, W465–W468.

Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. *22*, 1790–1797.

Brajdic, A., and Harle, R. (2013). Walk detection and step counting on unconstrained smartphones. (ACM Press), p. 225.

Carter, B.C., Vershinin, M., and Gross, S.P. (2008). A Comparison of Step-Detection Methods: How Well Can You Do? Biophys. J. *94*, 306–319.

Chadda, R., Krishnamani, V., Mersch, K., Wong, J., Brimberry, M., Chadda, A., Kolmakova-Partensky, L., Friedman, L.J., Gelles, J., and Robertson, J.L. (2016). The dimerization equilibrium of a ClC Cl^-/H^+ antiporter in lipid bilayers. ELife *5*.

Chen, Y., Deffenbaugh, N.C., Anderson, C.T., and Hancock, W.O. (2014). Molecular counting by photobleaching in protein complexes with many subunits: best practices and application to the cellulose synthesis complex. Mol. Biol. Cell *25*, 3630–3642.

Friedman, L.J., and Gelles, J. (2015). Multi-wavelength single-molecule fluorescence analysis of transcription mechanisms. Methods *86*, 27–36.

Hamill, O.P., Marty, A., Neher, E., Sakmann, B., and Sigworth, F.J. (1981). Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. Pflüg. Arch. Eur. J. Physiol. *391*, 85–100.

Kaur, H., Jamalidinan, F., Condon, S.G.F., Senes, A., and Hoskins, A.A. (2019). Analysis of spliceosome dynamics by maximum likelihood fitting of dwell time distributions. Methods *153*, 13–21.

Khadria, A.S., and Senes, A. (2013). The transmembrane domains of the bacterial cell division proteins FtsB and FtsL form a stable high-order oligomer. Biochemistry *52*, 7542–7550.

LaPointe, L.M., Taylor, K.C., Subramaniam, S., Khadria, A., Rayment, I., and Senes, A. (2013). Structural organization of FtsB, a transmembrane protein of the bacterial divisome. Biochemistry *52*, 2574–2585.

Shen, J., and Castan, S. (1992). An optimal linear operator for step edge detection. CVGIP Graph. Models Image Process. *54*, 112–133.

Weinmann, A., Storath, M., and Demaret, L. (2015). The L^1-Potts Functional for Robust Jump-Sparse Reconstruction. SIAM J. Numer. Anal. *53*, 644–673.