The Behavior of Spin-Labeled Mutants of the *Escherichia coli* Cobalamin Transporter BtuB Probed by EPR Spectroscopy in Native Environments

David Alexander Nyenhuis

Bachelor of Science in Chemistry, University of Virginia, 2014

A Dissertation presented to the Graduate Faculty of the University of Virginia in Candidacy for the Degree of Doctor of Philosophy

Department of Chemistry

University of Virginia
April, 2019

Abstract

Gram-negative bacteria rely on active transport proteins on their outer surface for the import of large, scarce nutrients across the outer membrane. These TonB dependent transporters have two main domains, with a cylindrical β-barrel surrounding a globular hatch domain. The hatch domain must undergo substantial conformational changes to facilitate substrate, but the mechanistic details of the transport process are still unknown. Much of the work to date has focused on the study of these proteins *in vitro*, but the work presented here focuses on the use of CW and pulsed-EPR techniques to study them in native systems. To this end, EPR results are presented for the *E. coli* cobalamin transporter BtuB in both whole cells and isolated outer membranes. These results demonstrate conformational responses to substrate binding on the extracellular face of BtuB, and potentially substrate independent extension of the N-terminal ton-box sequence on the periplasmic side. Additionally, pulsed-EPR measurements on BtuB in native systems display novel long-distance components that appear to indicate crowding and potentially organization in the outer membrane. Finally, CW EPR spectra for a set of sites across the extracellular face of BtuB are presented, from which substrate dependent conformational shifts can begin to be mapped out in the native environment. These sites may also help to resolve the nature of the protein's organization on the OM. These EPR results are supported by the presentation of a method for denoising CW EPR spectra through the stationary wavelet transform.

Acknowledgements

None of this would have been possible without the contributions of many others.

Linda and Carol, thank you. Without Biochem lab, I would never have considered research, and would not have learned the skills to feel confident attempting it.

Dave, thank you for providing me with the freedom to grow and succeed. You have taught me almost everything I know about EPR, and I am deeply grateful.

Bob, for your infinite patience with us, and providing us with the biochemical knowledge to keep pressing forward. Also, a special thanks to Bob's fever dreams, for providing us with the reason to attempt Dsb null mutants and securing the future of the whole cell project.

Thushani, for being the driving force behind our success. You never wavered, even when we had gone months without success, and always kept us moving forward. You were exactly the person that this project needed, at just the right time.

Sarah, for your continued love and support, and for being the one-woman support system that keeps the lab running.

The Tamm lab, for providing so many opportunities to expand my horizons through collaborations and discussion. A special thanks to Jinwoo, for your hearty dose of reality, Alex, for your grand visions, Volker, for your help and discussions, Binyong, for your unending enthusiasm.

My committee members, Dr. Linda Columbus, Dr. Ken Hsu, Dr. Andreas Gahlmann, and Dr. Robert Nakamoto, thank you for agreeing to devote your time towards this thesis.

Susie Marshall, thank you for your patience throughout this process, and for always finding a way to help us all, no matter the situation.

And for the many others that helped me along the way, I am deeply indebted to you all.

Abstract iii
Acknowledgements iv
Table of Contents v

Chapter 1: Introduction
1.1 Overview 1
Figure 1. 2d projection of cyanocobalamin 1
1.2 Cobalamin in eukaryotes and humans
Figure 2. Crystal Structures of human methylmalonyl-CoA mutase 4
And *E. coli* methionine synthase
1.3 Cobalamin in Prokaryotes 4
1.4 Structural Features of the E. coli Membrane Environments 6
1.5 The Inner Membrane 7
Figure 3. The structure of the BtuCD complex 8
1.6 The Periplasm 8
Figure 4. Crystal structure of BtuF 11
1.7 The Outer Membrane 11
Figure 5. Simulation model of bacterial OM 13
1.8 Outer Membrane Proteins 13
1.9 Folding and Insertion of Outer Membrane Proteins 16
1.10 Organization of the Outer Membrane 21
1.11 TonB dependent Transporters and BtuB 25
Figure 6. Crystal structures of BtuB and FhuA 26
1.12 The ExbB, ExbD, TonB Complex 27
Figure 7. Crystal structure of the FhuA-TonB complex 30
Figure 8. Models of the TonB C-terminal dimer interface 31
1.13 Models for TonB Dependent Transport 32
1.14 Summary of Remaining Chapters 37
1.15 References 39

Chapter 2: Theory and Methods
2.1 EPR in Native Systems 48
2.2 The Electron in the Magnetic Field 48
Figure 1. Electron Zeeman splitting 49
Figure 2. The behavior of the bulk magnetization 50
in the applied field
2.3 Choice of Reference Frame 50
From Atoms to Molecules
2.4 Anisotropy in g 51
2.5 Hyperfine Interactions 52
2.6 Electron-Electron Interactions 53
2.7 Relaxation 54
2.8 Continuous Wave Experiments, Broadening, and Lineshape 55
Figure 3. Examples of CW EPR spectra 55
Pulsed EPR
2.9 Introduction 58
2.10 Relaxation II 61
2.11 Homogeneous and Inhomogeneous Line Broadening 63
2.12 Pulse Sequences and Echos 63
Figure 4. The bulk magnetization in the presence of applied pulses 64
2.13 The DEER Experiment 65

Figure 5. The pulse sequence of the 4-pulse DEER experiment 65
2.14 Data Processing and Extraction of the Intramolecular Distances 67
Figure 6. Examples of DEER data and distance extraction 67
2.15 Preprocessing 67
2.16 Model-Free Analysis 68
2.17 Model-Based Analysis 70
Methods
2.18 Plasmids and Strains 71
2.19 Mutants 71
2.20 Standard Materials 72
2.21 Nonstandard Materials 72
2.22 Preparation of Whole Cell Samples 72
2.23 Preparation of Isolated Outer Membranes 73
2.24 Preparation of EPR Samples 74
2.25 CW Experiments 74
2.26 DEER Experiments 75
2.27 Figure Generation 76
2.28 References 77

Chapter 3: Wavelet Denoising of Slow-Motional CW EPR Spectra
3.1 Background and Transforms 78
Figure 1. The wavelet and scaling functions used in this work 80
Figure 2. Results of an eight-level decomposition of an EPR signal 81
3.2 Wavelet Thresholds 83
3.3 Wavelet Thresholding Methods 85
3.4 Description of Method 86
Figure 3. The six test signals used for testing the denoising method 87
Figure 4. Noisy test signals generated for various target SNRs 88
Table 1. The actual SNR values of the noisy signals 88
3.5 Performance of Method 89
Figure 5. SWT denoising performance of various methods on the blocks test signal at decomposition level 5 91
Figure 6. SWT denoising performance of various methods on the bumps test signal at decomposition level 5 92
Figure 7. SWT denoising performance of various methods on the heavisine test signal at decomposition level 5 93
Figure 8. SWT denoising performance of various methods on the doppler test signal at decomposition level 5 94
Figure 9. SWT denoising performance of various methods on the first EPR test signal at decomposition level 5 95
Figure 10. SWT denoising performance of various methods on the second EPR signal at decomposition level 5 96
Figure 11. SWT performance for three metrics across SNR for the first EPR test signal 98
Figure 12. SWT performance for three metrics across SNR for the second EPR test signal 99
3.6 Conclusion 100
3.7 References 101
3.8 Supplemental Information 101
Figure S1. SWT denoising performance of various methods on the blocks test signal at decomposition level 5 102

Figure S2. SWT denoising performance of various methods on the bumps test signal at decomposition level 5 103
Figure S3. SWT denoising performance of various methods on the heavisine test signal at decomposition level 5 104
Figure S4. SWT denoising performance of various methods on the doppler test signal at decomposition level 5 105
Figure S5. SWT denoising performance of various methods on the first EPR test signal at decomposition level 5 106
Figure S6. SWT denoising performance of various methods on the second EPR signal at decomposition level 5 107
Table S1. SWT performance for three metrics at levels 5 and 6 for varying SNR and method with the coif3 wavelet. 108
Table S2. SWT performance for three metrics at levels 5 and 6 for varying SNR and method with the coif4 wavelet. 109
Table S3. SWT performance for three metrics at levels 5 and 6 for varying SNR and method with the coif5 wavelet. 110
Table S4. SWT performance for three metrics at levels 5 and 6 for varying SNR and method with the bior2.2 wavelet. 111
Table S5. SWT performance for three metrics at levels 5 and 6 for varying SNR and method with the bior2.4 wavelet. 112
Table S6. SWT performance for three metrics at levels 5 and 6 for varying SNR and method with the bior2.6 wavelet. 113

Chapter 4: Accessing BtuB hatch-barrel information in the native system through modulation of the *E. coli* Dsb system and DEER spectroscopy
4.1 Introduction 114
Figure 1. Visualization of the 90-188 pair in BtuB 115
4.2 Additional Methods Information by Figure 116
Results
4.3 Development of a method for labeling hatch-barrel double mutants in whole cells 119
Figure 2. DEER data for BtuB 90-188 in RK5016 OM samples 119
Figure 3. Apo and cobalamin bound crystal structures of BtuB 120
Figure 4. CW EPR spectra of BtuB 90 and BtuB 188 in RK5016 cells 121
Figure 5. CW EPR spectra obtained for BtuB 90-188 in RK5016 cells with varying amounts of MTSL label 123
Figure 6. DEER data for BtuB 90-188 in RK5016 cells 124
Figure 7. CW EPR spectra obtained for BtuB 90-188 in DsbA$^-$ cells 125
Figure 8. DEER data for BtuB 90-188 in DsbA$^-$ cells 126
Figure 9. DEER data for BtuB 90-188 in DsbA$^-$ OMs 127
Figure 10. CW EPR spectra for WT BtuB and BtuB 90-188 in Dsb, DsbA$^-$, DsbB$^-$, and DsbC$^-$ strains 128
Figure 11. Comparison of DEER data for BtuB 90-188 between DsbA$^-$ and DsbB$^-$ cells 129
Figure 12. DEER data for BtuB 90-188 in DsbC$^-$ cells 130
4.4 Dose-Dependent Response of the BtuB 90-188 pair in DsbA$^-$ cells 130
Figure 13. Dose dependence of the BtuB 90-188 DEER in DsbA$^-$ cells Across a range of cobalamin concentrations 132
4.5 Attempted Modulation of the 90-188 Conformational Change 133
Figure 14. Comparison of DEER data for BtuB 90-188 with and without The I8P mutation in DsbA$^-$ cells 134

Figure 15. Visualization of the 188-399 pair in BtuB 135
Figure 16. CW EPR spectra for BtuB 188-399 in DsbA$^-$ cells 136
Figure 17. DEER data for BtuB 188-399 in DsbA$^-$ cells 136
Figure 18. Visualization of the 74-188 pair in BtuB 137
Figure 19. CW EPR spectra for BtuB 74-188 in DsbA$^-$ cells 137
Figure 20. DEER data for BtuB 74-188 in DsbA$^-$ cells 138
Figure 21. Addition of glucose does not alter DEER data Obtained for BtuB 90-188 in DsbA$^-$ cells 139
4.6 Extension of New method into the Periplasmic Face of BtuB 139
Figure 22. Visualization of the 6-510 pair in BtuB 140
Figure 23. CW EPR spectra for BtuB 6-510 in RK5016 OMs 141
Figure 24. DEER data for BtuB 6-510 in RK5016 OMs 141
Figure 25. DEER data for BtuB 6-510 in DsbA$^-$ cells 142
4.7 Discussion 143
Figure 26. Model of the Dsb system and BtuB insertion 143
4.8 References 146

Chapter 5: Crowding and Organization of overexpressed BtuB On the *E. coli* cell surface
5.1 Introduction 148
Figure 1. Recurrent long-distance components observed in prior DEER data from BtuB OM samples 150
5.2 Additional Methods Information by Figure 151
5.3 Characterizing Recurrent Long-Distance Components in Native Systems 153
Figure 2. Simulated DEER data for distances of 3 and 5.5 nm 153
Figure 3. DEER data for BtuB 90-188 in RK5016 OMs analyzed using DeerAnalysis 154
Figure 4. DEER data for BtuB 188 and a spin-labeled substrate analyzed using DeerAnalysis 156
Figure 5. DEER data for BtuB 188 and a spin-labeled substrate analyzed using DD 159
Figure 6. DEER data for BtuB 90 and BtuB 188 in RK5016 cells analyzed using DD 160
Figure 7. DEER data for BtuB 188 in RK5016 OMs showing the effect of diamagnetic dilution on the long-distance peak 161
Figure 8. DEER data for BtuB in RK5016 cells showing the effect of growth time on the long-distance peak 163
Figure 9. Summary of distance distributions for a set of 23 samples Of BtuB 188 in RK5016 cells analyzed using DD 164
Figure 10. DEER data for BtuB 188 grown in RK5016 cells with deuterated media and carbon source 165
Figure 11. Models of BtuB pairwise interactions 167
Figure 12. Comparisons of BtuB 188 DEER data to pairwise interaction models 168
Figure 13. Comparisons of various background subtraction models on The deuterated BtuB 188 data shown in Fig. 10 171
Figure 14. A comparison of the quadratic background subtraction of the deuterated BtuB 188 data shown in Fig. 10 with the pairwise interaction models 172

Figure 15. Testing background models and data truncation for the suppression of the long-distance component 174
5.4 Case Study: BtuB 6-510 in Cells and Outer Membranes 174
Figure 16. Comparisons of the BtuB 6-510 DEER in OMs and DsbA- cells with variable dimension or quadratic background models 176
5.5 References 178
5.6 Supplementary Figures 180
Figure S1. Dipolar traces corresponding to the distance distributions shown in Fig. 8 180
Figure S2. Time-domain data corresponding to the distance distributions shown in Fig. 9 181
Figure S3. Dipolar traces corresponding to the distance distributions shown in Fig. 9 182
Table S1. Summary of the distribution and fit statistics for the distance distributions shown in Fig. 9 183

Chapter 6: Future Directions
6.1 Introduction 184
6.2 Additional Methods 184
Results and Discussion
6.3 Additional Hatch Sites 184
Figure 1. The position of new mutant sites in the BtuB hatch domain 185
Figure 2. The CW spectra of the new mutant sites in the BtuB hatch domain 186
Figure 3. The CW spectra shown in Fig. 2 arrayed around a Visualization of the BtuB hatch domain 188
6.4 Additional Barrel Sites 188
Figure 4. The CW spectra of the new mutant sites around the extracellular loops of BtuB 189
Figure 5. The CW spectra shown in Fig. 4 arrayed around two Visualizations of the BtuB barrel 191
6.5 Future Work 192
6.6 References 195

Chapter 1: Introduction

1.1 Overview

Vitamin B_{12} is a general name for several members of the cobalamin (Cbl) family of corrinoid compounds. It is the most complex natural cofactor, having a three-part structure consisting of a planar corrin ring, a flavin nucleotide derived 5,6-dimethylbenzimidazole (DMBI) group, and an exchangeable active group surrounding a central cobalt(III) (1). The octahedral coordination of the cobalamin center is fulfilled by four nitrogens from the corrin ring, an additional nitrogen from DMBI, and a carbon from the active group. The latter Co-C bond was the first biological organometallic bond to be identified (2). The common structural features of cobalamins are shown in Figure 1, where the exchangeable ligand (X) can be one of a variety of adenosyl or methyl donors, leading to: 5'-deoxyadenosyl, methyl, hydroxo, and cyanocobalamins, among others (1, 2). Of particular relevance is cyanocobalamin (CnCbl), which is light stable and the principal commercially produced cobalamin derivative, at a rate of more than 10 tons a year (2). Additionally, adenosylcobalamin (AdeCbl), which is also known as coenzyme B_{12}, and methylcobalamin (MeCbl) are widely used in the human body as cofactors of the enzymes methylmalonyl-CoA mutase and methionine synthase, respectively.

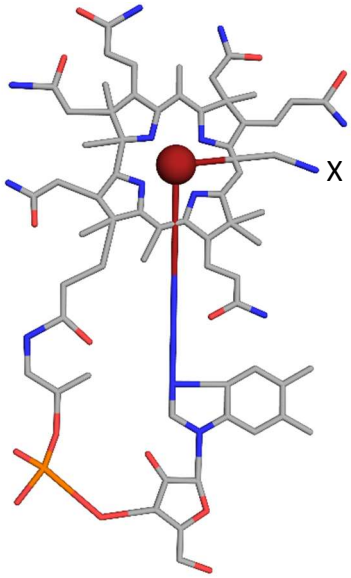


Figure 1. 2d projection of cyanocobalamin structure

While vitamin B_{12} is required in vertebrates, some invertebrates, half of all algae, and most bacteria and archaea, it is produced only by a subset of bacterial and archaeal species (3). Interest in vitamin B_{12} has continued now for almost a century, beginning with its implicit discovery in the 1920s, when it was discovered that eating liver cured pernicious anemia, which led to a 1934 Nobel Prize for Minot, Murphy, and Whipple (2). By the late 1940s the red compound

responsible for this clinical effect had been isolated from liver by the corporations Merck and Glaxo, and termed vitamin B_{12}. The first biologically active coenzymes of B_{12} were produced in 1958 and the three dimensional structures of CnCbl and later AdeCbl were determined by the group of Dorothy Hodgkin in the latter half of the 1950s (2, 4). Hodgkin's work led into the second Nobel prize related to B_{12}, this time recognizing her work in crystallographic structure determination. The means of producing this important cofactor quickly became a topic of interest. A synthetic route to vitamin B_{12} was finally completed in the early 1970s, with two variant approaches by Robert Woodward, who had won a Nobel of his own in 1965 for his achievements in organic synthesis, at Harvard and Albert Eschenmoser, who would later go on to be a pioneer in origin of life and artificial RNA research, at ETH. These synthetic routes contained nearly 70 steps, however, and were not efficient or cost-effective enough to be used in industrial production (2). Instead, the identification of the aerobic microbial pathway to B_{12} biosynthesis, described first in the organism *Pseudomonas denitrificans* in 1993, combined with subsequent mutagenic selection and genetic engineering, has led to today's production of vitamin B_{12} being exclusively via bacterial fermentation (2). More details are provided in the following sections.

1.2 Cobalamin in Eukaryotes and Humans

Humans and other animals that are dependent on vitamin B_{12} lack biosynthetic pathways and are thus reliant on consuming sufficient quantities in their diet. The recommended intake varies between countries, but is on the order of 1-3 ug, whereas the body can store several mg of B_{12}, often delaying by years the onset of adverse effects from B_{12} deficiency (5). Dietary B_{12} is not reserved for the human host, however, as more than 80% of the trillions of gut bacteria that reside in the colon are also dependent on B_{12} uptake (3, 6). Thus, the human system for B_{12} uptake and storage must be tightly regulated and efficient to ensure that the bodies' needs are met.

The first step in cobalamin uptake is the release of the glycoprotein haptocorrin (HC) by the salivary glands in the mouth. This protein binds tightly to most cobalamin derivatives but initial binding in the mouth is low, as the cobalamin is still tightly bound to proteins and molecules within the food. Instead, the majority of initial binding to HC occurs after the low pH environment and proteases of the stomach release the rest of the bound cobalamin (1). Progressing into the small intestine, the pH shift and action of pancreatic enzymes result in the HC being cleaved and releasing most of its bound cobalamin load (1, 5). They are then transferred to intrinsic factor (IF), which binds preferentially to the human usable derivatives, and functions as a selectivity filter in the system (1). The IF-cobalamin complex is then absorbed via receptor-mediated endocytosis using the cubilin receptor in the ileum (1). In the blood, the cobalamin load is split between HC and another binder, transcobalamin, although the majority of tissue cobalamin is stored in complex with HC (5). Unused cobalamin accumulates in the liver, where as noted above it can reach milliGram levels, sufficient for several years.

In humans and other eukaryotes, the primary uses of cobalamin involve its use as a catalytic cofactor in the enzymes methylmalonyl-CoA mutase and methionine synthase. The former is located in the mitochondria, where it uses AdeCbl as a cofactor to convert methylmalonic acid to succinate (7). This reaction is critical in the catabolism of branched or odd-numbered fatty acids (1, 7). The latter is instead located in the cytoplasm, where it uses MeCbl as a cofactor in the transfer of methyl groups from methyl tetrahydrofolate to homocysteine, remaking methionine. Thus, methionine synthase is a principal controller of methylation reactions in the body (1, 7). The

structures of these proteins are shown in Figure 2, with the cobalamin cofactors highlighted in red.

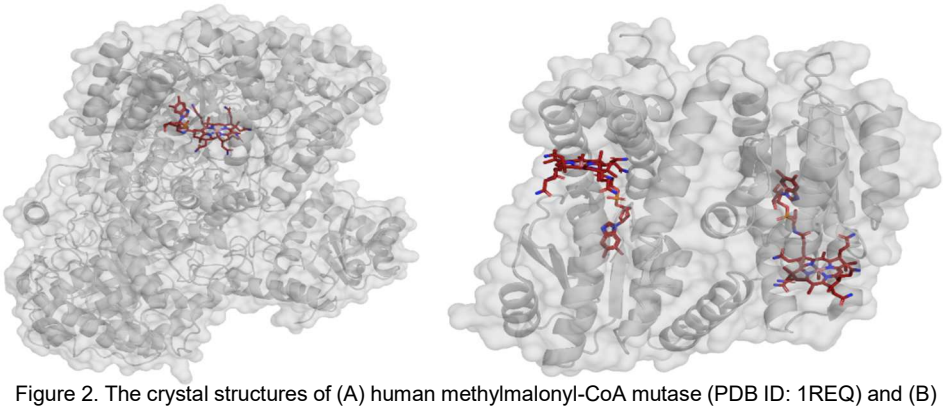


Figure 2. The crystal structures of (A) human methylmalonyl-CoA mutase (PDB ID: 1REQ) and (B) the B_{12} binding domains of *Escherichia coli* methionine synthase (PDB ID: 1BMT)

1.3 Cobalamin in Prokaryotes

Vitamin B_{12} may be the oldest biological cofactor, with potential roots in the RNA world (2, 7). Its initial role was probably in anaerobic fermentation of small molecules, but as atmospheric oxygen became more available, new offshoots of its biosynthesis were created that eventually led to siroheme and later to heme and chlorophyll that support today's eukaryotic organisms (2). Biosynthesis of B_{12} is exclusive to certain bacteria and archaea, with the elucidation of the aerobic bacterial pathway representing a 25 year journey led by several groups at Rhone-Poulenc (RP, currently part of Aventis) (2). They were joined by Sir Allen Battersby at Cambridge, who often alluded to B_{12} synthesis in terms of grand feats of mountaineering, and a host of other notable scientists (2). To accomplish this, the groups at RP checked for complementation between over 150 mutants of known B_{12} biosynthetic genes in *Agrobacterium tumefaciens and Pseudomonas putida* against more than three and a half thousand *E. coli* strains, each containing a piece of the *Pseudomonas* denitrificans genome (2). This let them identify a cluster of 22 biosynthetic genes in *P. denitrificans*, but it would be several more years of work on sequencing, purification, and isolation of intermediates until the full pathway was solved (2). Bacteria have two different pathways towards the synthesis of cobalamin, one anaerobic and one aerobic (8). They differ in

the point at which the cobalt is incorporated, with the aerobic pathway, found in *P. denitrificans*, inserting the cobalt much later (8). This was beneficial to the researchers, as the cobalt containing intermediates tended to be unstable. In both cases, most bacterial syntheses start from a glutamate-tRNA molecule with progression to the key cyclic intermediate uroporphyrinogen III, from which the pathways for chlorophyll, bacteriochlorophyll, heme, siroheme, and cobalamin diverge (2). In total, the biosynthesis requires more than 30 genes but has only 20 steps (9). Thus, the bacterial pathway is more efficient than the synthetic pathways of Woodward and Eschenmoser.

The majority of the world's B_{12} is still produced by RP (Aventis) using derivatives of *P. denitrificans*. The research groups at RP continued their work with the organism for another ten years, using random mutagenesis and selection for B_{12} production to increase the yield by two orders of magnitude (2). Since then, production strains have been enhanced with plasmids containing more efficient genes from other bacterial strains. The process of growth and extraction involves several days of aerobic fermentation, after which the vitamin B_{12} is extracted through superheating and the stable cyanocobalamin complex is formed by addition of cyanide. The solution can then be filtered and crystallized until the pure vitamin is isolated (2).

Within the human gut, however, the bacterial community does not have the luxury of relying on mutant strains engineered specifically for B_{12} production. Indeed, a review of 303 species present in gut flora found that more than 80% utilized cobalamin, while only about 20% had a full biosynthesis pathway (6). As their human host possesses a highly efficient system to harvest most of the dietary cobalamin, these organisms must possess their own, highly effective means of vitamin B_{12} uptake. The uptake machinery in Gram-negative bacterium, including those of phylum Bacteroidetes, and from family Enterobacteriaceae, such as the model organism *Escherichia coli*, have been studied for decades. These Gram-negative bacteria rely on a family of proteins that form the BtuBFCD transport system to move cobalamin derivatives across their

two membranes into the cytoplasm (3). The details of this system are discussed in the following sections. The first protein, BtuB, which sits in the outer membrane of the bacterium, is responsible for the binding of cobalamin derivatives in the medium. Competition in the gut is sufficiently fierce that most members of Bacteroidetes encode not one, but two, three, or even four analogues of the BtuB protein (6). As an example, *Bacteroides thetaiotaomicron*, the second most common bacterium in the adult gut, encodes three distinct homologs of BtuB, with two binding to a broad range of variants and one only to cobalamins with adenine or benzimidazole ligands (6, 10) This competition and specialization into variants of cobalamin among gut residents has even led to proposals for targeted remodeling of gut species by selective supplementation with specific cobalamin analogues .

The experiments presented in this work focus on the role of the BtuB transport protein in the organism *E. coli*, and so the unique structural elements of Gram-negative bacteria, and the details of the BtuBFCD transport system are presented in the following sections.

1.4 Structural features of the *E. coli* membrane environments.

Like other Gram-negative bacteria, the cytoplasm of *E. coli* is surrounded by both an inner and outer membrane (IM, OM), with the periplasm between them. While the inner membrane is comprised of phospholipids and is filled with proteins comprised mainly of helical bundles, the outer membrane has a unique, asymmetric composition. The inner leaflet again consists of phospholipids, while the outer leaflet is populated by large lipopolysaccharides (LPS) interspersed with protein β-barrels. The gel-like periplasm that stretches between them is filled with chaperones, energy transduction complexes, and structural elements that link the outer membrane to the semi-rigid peptidoglycan (PG) matrix that helps the cell maintain its shape.

1.5 The Inner Membrane

The inner membrane of Gram-negative bacteria is a phospholipid bilayer made primarily of phosphatidylethanolamine (PE), phosphatidylglycerol (PG), and cardiolipin (CL). The ratios of these components are approximately 75% PE to 25% PG to 5% CL, although this varies depending on the strain, growth conditions, environment, and point in the cell cycle (11). PE is a zwitterionic lipid, whereas both PG and CL carry a net negative charge between the pH values of 5.5 and 7.0 used for most bacterial growths presented here (12). In the rod-shaped *E. coli* cell, PG and CL are enriched towards the higher curvature region of the cell poles due to their higher intrinsic curvature (13). This is particularly apparent for CL, which has a relatively small glycerol head-group and 4 fatty acid tails.

In addition to phospholipids, the inner membrane is filled with both integral and peripheral membrane proteins and protein complexes. In fact, more than 20% of the open reading frames in the *E. coli* genome may encode for IM proteins, which generally are comprised of α-helical bundles (14). These inner membrane proteins are responsible for generation of energy for the cell via the electron transport chain, and thus to creation of the proton motive force (PMF). The PMF is the sum of both the electrical potential created across the IM, and the chemical potential created by the pH difference between the cytoplasm and periplasm, and in *E. coli* can be generated through both aerobic and anaerobic respiration (15). The PMF is essential to many IM and OM processes, including active transporters and the generation of ATP through ATP synthase. Inner membrane proteins are also critical to cell division, signal transduction, and the efflux of toxic or antimicrobial compounds.

Regarding the BtuBFCD system for cobalamin uptake, the IM contains the BtuCD protein complex. BtuCD is a heterotetrameric complex, with the stoichiometry $BtuC_2D_2$. It belongs to the ABC transporter family and couples the binding and hydrolysis of ATP to ADP with import of cobalamin across the IM. The structure of a closed conformation of BtuCD, bound to the

periplasmic element of the BtuBFCD system, BtuF, is shown in Figure 3. Each BtuC monomer contributes ten transmembrane (TM) domains, which together form a 20 helix bundle structure, which is capable of creating an alternating access pore large enough for the 1.5 kDa cobalamin substrate (16). The BtuD monomers, meanwhile, are involved in nucleotide binding. There are 2 ATP-binding cassettes in the overall complex, with each site containing elements of both BtuD monomers (16). The action of ATP binding and hydrolysis is coupled to motion in these nucleotide binding domains (NBDs), which in turn cause rearrangements of the TMDs that alter the facing direction of the pore (16). In bacteria, most ABC transporters are involved in the import of rare nutrients, as with BtuCD (17). Some are instead used for the export of hydrophobic molecules, however, which is the predominant mode in eukaryotic cells (18). For a bacterium, these hydrophobic molecules are most commonly toxins or drugs, giving ABC transporters an important role in bacterial drug resistance.

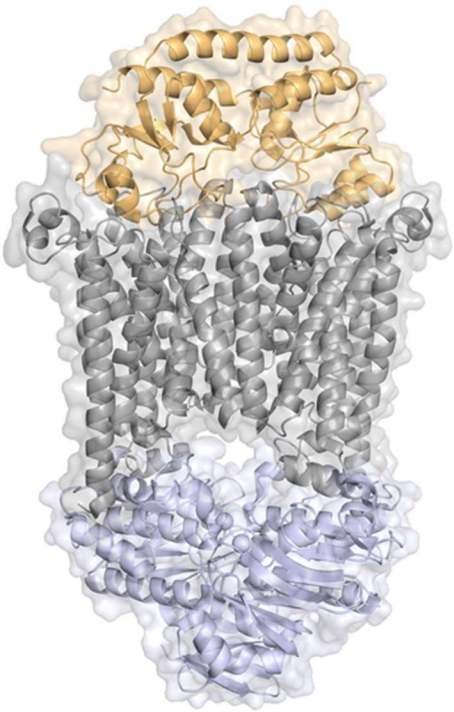


Figure 3. The structure of the BtuCD complex with bound BtuF (PDB ID: 4FI3). The TM domains of BtuC are colored gray, and are attached to the two nucleotide binding domains of BtuD in blue. The periplasmic transport protein BtuF, shown in orange, shuttles cobalamin between the OM and IM.

1.6 The Periplasm

Between the IM and OM lies the periplasm, or alternatively periplasmic space. This aqueous compartment is filled with chaperones and substrate binding proteins and is critical in the folding and insertion pathways for OM proteins. The periplasm possesses no ATP, and so the only source of energy is the PMF across the IM. The periplasm is thus also home to a variety of energy transducing complexes that link the IM and OM, including the mot, tol, and ton systems. In addition to protein elements, the middle of the periplasmic space contains the peptidoglycan

(PG) matrix, a hybrid protein and sugar structure with glycan strands cross-linked by short peptide chains (19). The PG confers structure to the cell but it is also highly flexible, permitting the cell to respond to a wide variety of mechanical stressors and changes in osmotic strength.

The diameter of the periplasm is about 15-20 nm, although there is still substantial ambiguity regarding its size and it is possible that the diameter varies across the cell (20, 21). The distance between the inner leaflet of the OM and the PG matrix is controlled by Braun's lipoprotein (BLP), one of the most common *E. coli* proteins (22, 23). BLP is an elongated, helical protein that forms a stable homotrimer (24). The trimer is stabilized by a three-helix coiled-coil arrangement with an alanine-zipper, affording it a high degree of strength and stability (24). The N-terminus of BLP is covalently attached to the phospholipids of the OM inner leaflet, while the C-terminus is covalently linked to the PG matrix, directly coupling the two (23). The solution length of a BLP trimer is approximately 9 nm, but molecular dynamics simulations found that in membrane it is able to bend slightly, and also adopts a substantial tilt, resulting in an overall separation between OM and PG of about 7 nm (23). Less is known about what controls the distance between the PG and the IM, but the lengths of periplasm spanning proteins, including LpoA and TonB, are consistently between 14 and 15 nm (21, 25). For a periplasm of length 15 nm, consistent with the length of these proteins, this would place the PG in the middle of the aqueous compartment.

The peptidoglycan matrix itself is highly complex, being a mixed polymer of between 1.5 and 3 nm in thickness with both sugar and peptide substituents (26). The glycan strands consist of alternating N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc) sugar units linked by a B-(1,4) glycosidic bond (19). The glycan strands average 25 to 35 disaccharides in length and are cross linked by short peptides consisting of alternating D and L-amino acids (27). The peptides are short, about five amino acids in length, and the alternating structure allows for interpeptide crosslinks between L and D constituents (27). The combination of flexible peptides

with rigid glycans allows the PG to maintain the shape of the cellular envelope while accommodating changes in the extracellular environment. The matrix is expandable, with the surface area increasing up to three-fold before rupture and the pores stretching to accommodate proteins of between 20 and 100 kDa in size (28-31).

This flexibility in pore size also has implications in PG synthesis, which must be tightly controlled to maintain the intended rod-like shape. *E. coli* cells have two distinct modes of PG synthesis. During growth, the cell grows mostly through elongation, and PG synthesis occurs throughout the side walls of the cell at elongasomes (32). Prior to division, however, a different cluster of proteins is recruited to form a divisome and synthesize the components of the two new cell poles in the resulting daughter cells (32). The two main GTase/TPases responsible for PG polymerization and peptide crosslinking in *E. coli* are PBP1A and PBP1B (32). These proteins associate with structural elements, cytoskeletal proteins, and hydrolases for recycling of the old PG components at the elongosomes, and divisomes, respectively (32). Yet, if the complexes were formed stochastically, the cell might lose its shape. Instead, it appears that interactions between PBP1A/PBP1B and the lipoproteins LopA/LpoB, which contain relatively bulky domains that must pass through the PG matrix, allow the synthesis of PG to targeted only to regions of a particular pore size (33).

The periplasm is also home to the periplasmic member of the BtuBCDF system, BtuF. This substrate binding protein shuttles cobalamin across the periplasm from BtuB in the OM to the periplasmic binding pocket of the BtuCD complex. The structure of BtuF is shown in Figure 4. The cobalamin is surrounded by two binding lobes, each with fingerlike helices around a central β-sheet, joined by a helical backbone (34). The interaction between BtuF and BtuB at the OM is poorly understood, but the interaction between BtuCD and BtuF has been revealed through

several crystal structures, one of which was shown in Fig. 3. There are two negatively charged regions on BtuF, one on each binding lobe, that fit into positively charged pockets on the TM domains of BtuC (34). This creates a closed cavity where the vitamin can be transferred to the BtuC pore. Concurrently, binding of BtuF to BtuCD opens the two binding lobes on the former, which together with insertion of loop residues on BtuC reduces the affinity of BtuF towards cobalamin, allowing it to enter the cavity (35). The BtuCD transporter is then free to complete the cycle, moving the cobalamin substrate into the cytoplasm.

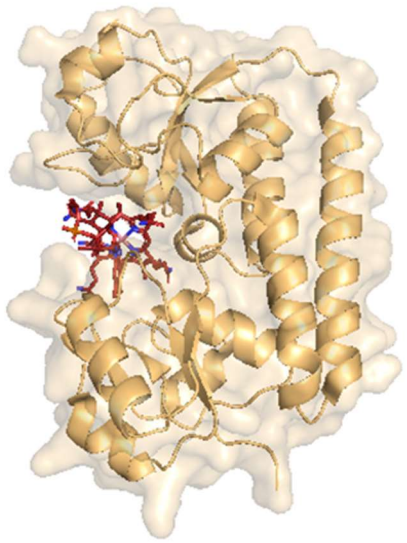


Figure 4. Crystal structure of BtuF (PDB ID: 1N2Z) with bound cobalamin

1.7 The Outer Membrane

Like the IM, the OM of *E. coli* contains a similar ratio of PE, PG, and CL phospholipids in its inner leaflet. The outer leaflet, in contrast, is comprised of unique lipopolysaccharide (LPS) molecules, which consist of lipid A linked to a polysaccharide that projects up out of the bilayer. Lipid A is unique in that its headgroup consists of a glucosamine disaccharide linked by a β-1',6 glycosidic bond (36). The sugars are linked to 3-hydroxymyristic acids at the 2, 3, 2', and 3' positions, and on one of the sugars the two fatty acids are acylated again off their hydroxy groups (36). This produces a total of six tails, which are typically fully saturated. The saturation and number of tails leads to an extremely low fluidity for the outer leaflet hydrocarbon. In isolation, LPS can withstand temperatures of up to 75° C before melting, ensuring that the outer leaflet is always in an ordered, gel-like state *in vivo* (37).

The lipid A is linked to a polysaccharide core that extends up and away from the bilayer surface. K-strain possesses only the core-oligosaccharide, whose general sugar scheme is given below, but in virulent strains a further set of repeating O-antigen sugars is present (36).

```
                Kdo (4-2) Kdo        Hep        αGal                Hep
                (4-2)                (7-1)      (6-1)               (2-1)
Lipid A (6-2) Kdo (5-1) Hep (3-1) Hep (3-1) αGlc (3-1) αGlc (2-1) αGlc
```

Here αGlc is α-D-glucose, Hep is L-glycero-D-manno-heptose, αGal is α-D-galactose, and Kdo is 3-deoxy-D-manno-oct-2-ulosonic acid. Portions of the core-oligosaccharide are essential, as truncation mutants lead to defects in OM protein biogenesis and insertion (38).

LPS also has a high anionic charge density, with a phosphate group on each sugar, carboxy groups on the Kdo species, and mono or di-phosphates linked to the heptose sugars (36). This gives it a charge density per acyl chain twice that of most anionic phospholipids. To stabilize these charges, LPS typically contains a 4:1:1 ratio of stabilizing Mg^{2+} and Ca^{2+} counterions per LPS (39). This charge network also creates strong lateral bridging between LPS molecules, which is supplemented by extensive hydrogen bonding between neighboring sugars. In combination with the extreme ordering of the acyl chains, domains of LPS can even remain stably isolated in phospholipid bilayers for days at a time (40). The strength of these lateral associations also means that LPS is a potent diffusion barrier, blocking the passage of most small solutes and particularly hydrophobic ones, which diffuse about two orders slower than through a typical phospholipid bilayer (41, 42). The structure of a model OM bilayer produced using the CHARMM-GUI tool with a 100 Angstrom box and 18:1:1 PE:PG:CL ratio in the inner leaflet is shown below in Figure 5 (43, 44).

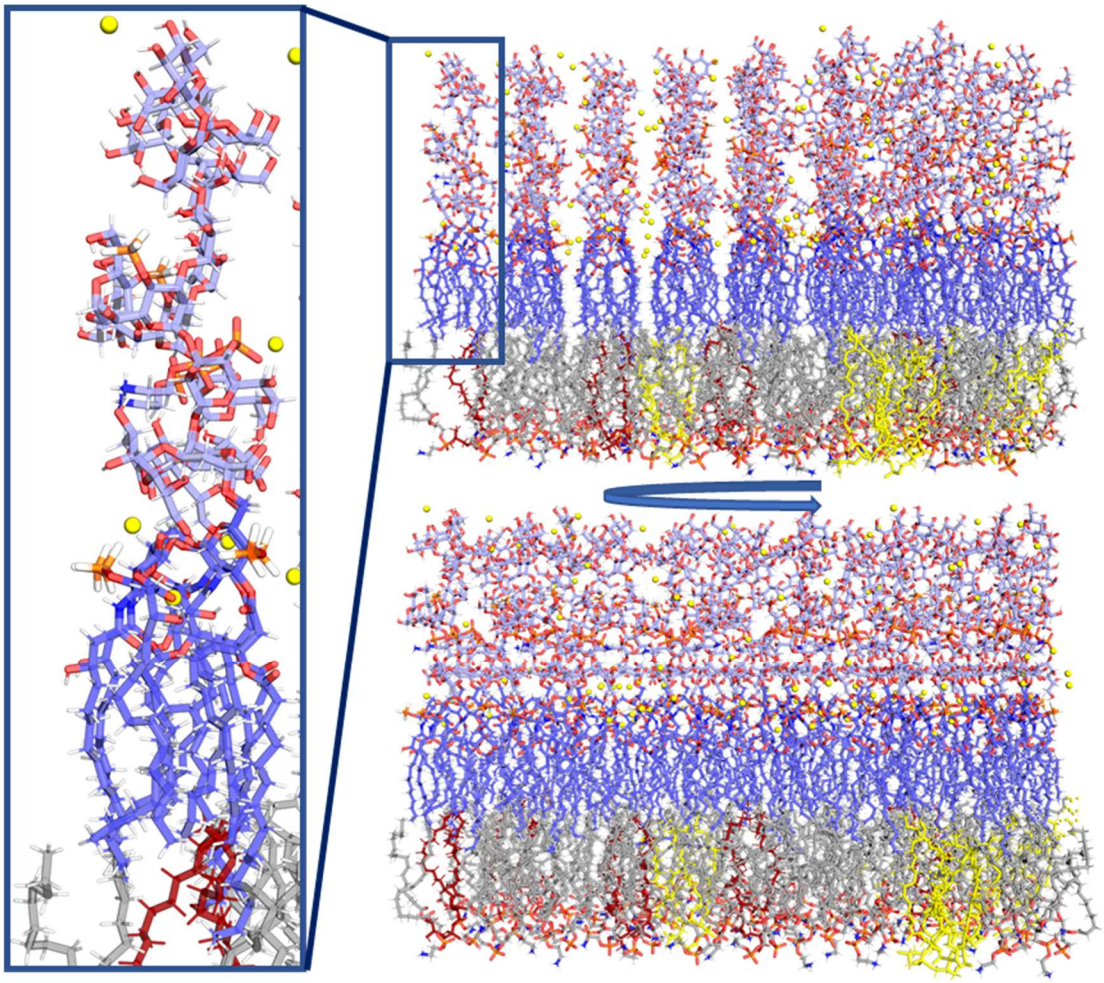


Figure 5. Model bacterial OM. The outer leaflet is composed solely of LPS, with lipid A (darker blue) and the core-polysaccharide (lighter blue). The inset shows an enlarged view of one LPS molecule. The phospholipid inner leaflet is composed of PE (gray), PG (red), and cardiolipin (yellow). Membrane generated using the CHARMM-GUI Membrane Builder tool.

The Tip3 water model atoms have been omitted in Fig. 5 for clarity. The simulated bilayer highlights the difference in order between the acyl chains of the inner leaflet phospholipids when compared to the outer leaflet lipid A. It is also immediately apparent that the core polysaccharide is stabilized by an extensive network of calcium counter ions (yellow spheres); magnesium was omitted for simplicity.

1.8 Outer Membrane Proteins

Proteins of the OM constitute a variety of porins, transporters, and sensory systems. They are composed of 8 to 26-strand β-barrels, and range in copy number from dozens to thousands.

Together, they cover the majority of the cell surface, ensuring that important compounds and ions are exchanged between the extracellular medium and the periplasm.

The most common proteins on the surface of *E. coli* K-strain are the major trimeric porins, OmpF, and OmpC. In some cases, these two proteins alone can constitute 80% of OM protein content, dominating the membrane landscape (45). These OM proteins are responsible for the passive movement of small molecules (<600 Da), and are highly efficient ion conductors with rates of 10^8-10^9 ions/s in physiological conditions (46). Compared to OmpC, OmpF produces a larger channel and thus leads to greater membrane permeability. This change in permeability can be exploited by the cell, with environmental stressors including: high osmotic stress, high temperature, antibiotics, bile acids, and acidic pH tending to favor incorporation of OmpC over OmpF (36).

Other major surface proteins include the porin OmpA and several specific transporters. OmpA is a two-conformer protein, able to fold into both an 8-stranded β-barrel with a C-terminal globular domain, and a larger 16-stranded β-barrel (47). In the two-domain form, OmpA lacks a continuous, water accessible channel in its barrel, but possesses the ability to anchor to the PG matrix via the C-terminal domain (47). The single domain form, in contrast, appears to be mobile in the membrane but has channel forming properties (47). Early studies on the propensity of each form found that only 2-3% of total OmpA appeared to be in the larger conformation while later studies at elevated temperatures have instead shown that the larger conformer is greatly enhanced and may even represent the native fold at physiological temperatures (47). Other porins include the recently discovered OmpL and PhoE, which is expressed in phosphate starvation conditions and has some selectivity towards anions, instead of the cation-selectivity seen in OmpF and OmpC (48).

Much of the remaining surface is taken up by various specific transporters, which move substrates that are too large or too rare to be taken up via nonspecific diffusion through the major

porins. Specific transporters can be passive, as with the maltose and sucrose transporters LamB and ScrY, or active, as with the TonB dependent transporters (TBDTs) responsible for the movement of iron siderophores and cobalamin. The first passive transporter, LamB, is responsible for enhancing the rate of diffusion of maltose and oligosaccharides across the OM, which would otherwise diffuse very slowly through the nonspecific OmpF and OmpC porins. In the intestine, maltose is the principal breakdown product of dietary starch, and so efficient uptake is crucial for a cell's survival (36). Structurally, LamB is another trimeric protein, with each subunit contributing an 18-stranded β-barrel (36). Inside the barrel is a narrow channel with a diameter of half a nm, which is lined with a string of aromatic residues called the greasy slide (49). The residues of the greasy slide interact with the less polar regions of maltose and other sugars, facilitating their movement (49). The center of the channel also contains several polar residues that can hydrogen bond to the hydroxyl groups of the translocating sugars, although the orientation of this region has the unintended consequence of preventing the translocation of sucrose (49). Thus, growth in the presence of sucrose requires the expression of a different channel, ScrY. Like LamB, ScrY is also trimeric, but it has a larger pore with a diameter of about 0.8 to 1.1 nm, and lacks the bulky residues of LamB that prevent the diffusion of sucrose (36). Other selective transporters that still rely on passive diffusion include the nucleoside transporter Tsx and OmpW, which enhances the diffusion of small hydrophobic compounds (36, 50).

To capture even rarer substrates, the cell must turn to active transport processes. The outer membrane portion of the BtuFCD system, BtuB, is an example of an OM active transporter. It and the various BtuB homologs are again responsible for the uptake of various cobalamin derivatives. The other principal target for active transport in the gut is iron, whose free concentration in the human host is as low as 10^{-24} M (51). As with cobalamin, nearly all of the free iron is tightly bound to large siderophores, and so the cell must use a variety of active transporters to capture one or more of these iron complexes to survive. Other active transporters

might capture large polysaccharides or other metals, but in most cases they belong to the family of TonB dependent transporters, which are discussed in Section 1.11.

Finally, the OM of *E. coli* is also home to proteins and protein complexes that function in extracellular export, motility, or adhesion. Examples of the former functions include the six major extraction systems that move proteins into the extracellular medium, and the bacterial flagellum that permits movement. Adhesion, meanwhile, is modulated by various proteins, but the 8-stranded B-barrel OmpX plays a key role in both biofilm formation with neighboring bacteria, and the invasion of eukaryotic cells (52, 53). These proteins and complexes tend to be minimally expressed or absent in K-strain *E. coli* and are further repressed in the Dsb knockout strains used in much of this work, limiting their impact on the present results.

1.9 Folding and Insertion of Outer Membrane Proteins

Getting nascent OM proteins from the cytosol where transcripts are produced to the membrane where they can be folded and inserted requires the coordinated work of several protein complexes. In *E. coli*, there are a total of eight systems that can serve to export proteins targeted to the IM, periplasm, OM , or extracellular medium, although not all present in K-strain derivatives (54). For OM insertion, the first step is to translocate an unfolded peptide across the IM, which is handled primarily by the Sec system. Proteins targeted for Sec translocation typically have an N-terminal signal sequence of hydrophobic and/or helical character, which is recognized by the signal recognition particle (SRP) that handles membrane targeting of the nascent protein (54). In the case of IM proteins, anchor sequences are also possible, which play a role in targeting but also halt translocation, leaving the protein in the IM with the anchor as a TM helix (55).

Sec translocation is used for unfolded peptides. It takes place co-translationally, although only once a majority (about 80%) of the sequence has been translated by the ribosome (54). In addition to the SRP, the unfolded protein is bound to the tetrameric chaperone protein SecB. This

chaperone has a high affinity for the motor protein SecA, which is an ATPase and provides the energy necessary for translocation (54). SecA and SecB in turn form a complex with the SecYEG translocon, which provides the channel through which the substrate is translocated (54). The Sec system can also interact with the YidC insertase, which can insert some smaller IM proteins on its own and is critical in the insertion of subunits of ATP synthase and some cytochrome oxidases (56).

While the Sec complex mediates the passage of unfolded substrates across the IM, folded proteins are translocated via the twin-arginine translocation (Tat) pathway. The Tat pathway is named for the dual arginine motif in its signal sequence, S-R-R-x-F-L-K, and tends to move proteins that are already folded and sometimes already oligomerized (57). Its substrates can be destined for excretion, but also include redox proteins used in anaerobic respiration and those used for production and maintenance of the cell membranes (58). The Tat system consists of oligomers of a six-pass protein, TatC, and the single-pass protein TatB interspersed with a few single-pass TatA monomers. The TatBC complex forms a partially curved, wall-like structure with an upper gate, and several appear to work together to form the translocon. While the Tat complex moves a few substrates bound for the extracellular medium, the primary means of excretion to the exterior are the type I-VI secretion pathways. Periplasmic proteins are excreted via the type II and V systems, whereas movement directly from the cytoplasm to the exterior is handled by the type I , III, IV, and VI systems (54).

Both lipoproteins and OM targeted β-barrels enter the periplasm primarily through the Sec system. After translocation, lipoproteins are first anchored to the IM through attachment of a diacylglyerol to their N-terminal cysteine, which in *E. coli* is further modified with a third acyl chain (59). The +2 position relative to this cysteine then determines their final destination; an apartate residue causes them to be retained in the IM, while any other residue results in transfer to the periplasmic leaflet of the OM via the lipoprotein outer membrane localization (Lol) pathway (59).

Transfer is initiated by the ABC transporter LolCDE complex, which releases the lipoprotein complexed to the LolA chaperone (59). The complex then moves to the LolB receptor, which reattaches the lipoprotein to the periplasmic OM leaflet (59).

In contrast to lipoproteins, OM barrels are targeted through a C-terminal signal sequence, which is recognized by part of the Bam complex that handles integral OM insertion. Since the periplasm lacks ATP, the sorting and insertion events are driven by the energy sink of the folding event, which can exceed -20 kcal/mol (60). Together with the high tendency of unfolded β-barrels to aggregate, driven by their great number of hydrophobic residues, it is clear that periplasmic chaperones must be present to ensure that OM barrels reach their intended destination. These chaperones can employ one of two strategies to prevent folding and aggregation of their substrates.

The first involves several domains capable of forming a cage around the nascent protein, with the cage restricting the available conformers. This strategy is pursued by the seventeen-kilodalton protein Skp, which is a helical trimer with three flexible arms that give the protein a form reminiscent of a jellyfish or octopus (61). The flexibility allows for the substrate to sample an extended, ovular conformational space while protecting it from significant outside contacts. The chaperone and serine endoprotease DegP also follows this strategy, although it does so by constructing huge circular or near spherical complexes with nearly 20 nm diameters that can span the width of the periplasm (62). The resting state of DegP appears to be as a hexamer, but it assembles into larger 12-mers and 24-mers both to act as a chaperone and to enhance its proteolytic activity towards misfolded targets by 15-fold (62). Both the 12-mer and 24-mer forms have gaps between monomers of a size sufficient for the free diffusion of unfolded substrates. The internal cavity, meanwhile, is about 8 nm in diameter making it just large enough to fit a folded porin such as OmpF or OmpC (62). The addition of proteolytic activity to degrade substrates that are trapped in off-pathway conformers makes DegP a powerful solution to OM protein folding.

Alternatively, the two periplasmic prolyl-isomerases FkpB binding protein A (FkpA) and survival protein A (SurA) opt to cradle their substrates between two binding domains, separated by an extended helix (63, 64). This helix can rotate relatively freely, allowing the two domains to flex and compress with the movements of the nascent protein (63, 64). There is some redundancy between chaperones, with Skp and DegP able to rescue mutant strains deficient in SurA (65).

Prior to insertion, the structural disulfide bonds (DSBs) of some OM proteins must also be formed and checked for accuracy. DSBs are extremely uncommon in the cytoplasm of most organisms, including *E. coli* due to the reducing environment. They become possible in the significantly more oxidizing environment of the periplasm, however, and are handled by the aptly named Dsb system. These proteins are also localized to the IM, and disulfide formation happens in-line with extrusion of the nascent protein through the SecYEG translocon. This means that initial disulfides are formed in sequence order, and additional proteins are required to function as disulfide isomerases, to ensure that in the case of multiple cysteine residues only the correct disulfides are maintained (66).

Mechanically, cysteine residues on the new OM protein are contacted first by the periplasmic protein DsbA, which is part of the thioredoxin-family and contains an internal C-x-x-C motif (67). These cysteines are initially disulfide bonded, but interaction with the substrate cysteine causes a disulfide bond transfer, resulting in a covalent link between the substrate and DsbA with consequent reduction of one of the DsbA catalytic cysteines (67). The appearance of a second substrate cysteine results in a second transfer reaction, which forms a disulfide bond within the substrate and reduces the remaining catalytic DsbA cysteine.

The means to recycle the catalytic motif of DsbA is provided by the integral IM protein DsbB (67). Inside this protein is a quinone cofactor, with the type being dependent on growth conditions. Ubiquinones are generally present in aerobic growth, while menaquinones are found in anaerobic conditions (68). As with the formation of the substrate disulfide, the catalytic

cysteines of DsbA are recharged by passing the electrons to a matching C-x-x-C motif on a periplasmic loop in DsbB (67). This loop passes the electrons in turn to another, more membrane proximal loop, which finally passes them to the quinone cofactor (67). Finally, the quinone passes the electrons to terminal oxidases in the respiratory chain, completing the cycle (67).

Since disulfide formation occurs sequentially during translocation, the disulfide isomerase DsbC is used to ensure that only native disulfides are retained. DsbC is a homodimeric, periplasmic protein with a C-x-x-C motif on each monomer (67). The catalytic domains of DsbC begin in a reduced state, allowing for nucleophilic attack by each cysteine on improperly formed DSBs. Creation of the native DSB can follow a concerted isomerization or separated rounds of DSB dissolution and formation (67). DsbC is regenerated through reduction by DsbD, another integral IM protein that is itself reduced by thioredoxins in the cytoplasm (67, 69).

Null mutants of DsbA and DsbB prevent the formation of most disulfide bonds in the periplasm. A null mutant of DsbC, meanwhile, causes the accumulation of misfolded substrates if more than two cysteines are present in the template. Interestingly, a null mutant of DsbD results in the accumulation of oxidized DsbC, which can replace DsbA in certain cases, while overexpression of DsbD creates an excess of reductive potential in the periplasm and presents the same phenotypically as DsbA and DsbB null mutants (69, 70).

After translocation, chaperone binding, and possible disulfide bond formation, the new OM barrel can proceed to the Bam complex for insertion. The Bam complex has four scaffold lipoproteins arrayed around a central BamA barrel (71). These lipoproteins associate through a set of five periplasmic polypeptide transport-associated (POTRA) domains (71). The fifth POTRA domain is responsible for binding to the BamCDE complex, which contains the essential protein BamD (71). This lipoprotein appears to be responsible for binding to the C-terminal signal sequence and passing the substrate to the BamA protein (71). BamD may also regulate the function of BamA, as it remains associated to the fifth POTRA domain until the substrate has been

completely inserted (71). BamC and E appear to modulate the activity or structure of BamD. Finally, BamB also interacts with BamA through the third POTRA domain, but its role is not well understood.

There is still some controversy regarding the functional mode of the Bam complex, although recent work on the mitochondrial homologue of BamA, Sam50, has provided a highly plausible model. *In* vivo crosslinking of the yeast Sam50 determined that the first and final β-strands of the Sam50 barrel can separate, opening a lateral gate in the barrel (72). The first two strands of the new OM protein appear to form a β-hairpin and are guided into the gate by an overhead loop conserved in Sam50 and BamA (72). The substrate hairpin then inserts, forming 2 new sheets in the now mixed-protein β-barrel. This sequence of opening and insertion of another β-hairpin continues until the new OM protein is fully formed (72).

1.10 Organization of the Outer Membrane

While LPS possesses unique properties that make it an ideal permeability barrier and provide a great deal of stability to the outer envelope of the cell, it is not the only constituent of the bacterial OM. The surface of a Gram-negative bacterial cell is densely packed with β-barrel proteins covering the majority of the membrane envelope. The major trimeric porins, OmpF and OmpC, make up most of the protein content, but they are interspersed with active transporters, specific channels, and other barrels of varying oligomeric state. The emerging picture of the membrane is not one of a fluid, lipid driven bilayer, but that of a molecular sieve with an extensive network of protein pores thinly separated by LPS molecules.

Recent developments in atomic force microscopy have begun to produce images of native bacterial OM and intact cells sufficient to directly resolve the protein complexes on their surfaces. Work on OM sheets obtained from the marine picoplankton *Roseobacter denitrificans* determined that complexes of the *R. denitrificans* trimeric porin covered most of the surface and were

organized into constricted and relaxed forms based on local density (45). In the tightest packing regions, these complexes covered more than 75% of the available surface area (45). Compared to *E. coli, R. denitrificans* has a higher percentage of major porins as a function of total protein, at 90% instead of 80%, but the *E. coli* OM is still likely to be a densely packed network of trimeric porin complexes (45). Atomic models of the porin assemblies determined that the distances between them were consistent with interactions between aromatic residues. Such interactions have been identified before, and aromatic girdles are present in almost all β-barrels near the lipid headgroups (73). Additionally, the individual trimers were found to have little directional correlation, indicating a lack of specific interaction faces between the complexes (45). Similar results were also seen for an AFM study of the magnetotactic *Magnetospirillum magneticum*, where the surface of the protein was revealed to consist of an interlocking net-like structure (74). Individual holes or pores in the net were found to display limited, bounded diffusion and were consistent with trimeric porins moving on the surface (74).

These AFM results corroborate a large body of evidence stemming from fluorescence diffusion and single-particle tracking studies, which has indicated that most OM proteins display highly anomalous, confined diffusion behavior. An extensive review of this work is available, but some of the details are summarized here (75). Fluorescence recovery after photobleaching (FRAP) experiments on the porin OmpA showed that photobleaching persisted without recovery for a full 15 minutes (76). OmpA has a noncovalent PG interaction domain, but recovery was absent even when this domain was removed (76). Similar experiments using colicins E9 and Ia, bacteriocins that bind specifically to OM proteins BtuB and Cir, respectively, found no recovery after three minutes (77). These results indicate that long range diffusion on the order of 100 nm or more is entirely absent for these OM proteins. Single-particle tracking (SPT) of the maltose channel, LamB, using microspheres and gold particles for tracking has also demonstrated local Brownian diffusion on very short time scales, but increasingly confined behavior in longer ones.

The former study found that the LamB receptor appeared to diffuse within domains with a radius of about 25 nm, while the latter identified a slow moving population within a 20-50 nm compartment on the surface (78, 79). A variant approach using fluorescent antibodies to BtuB and OmpF failed to show confinement of BtuB but did determine that its diffusion was slowed five-fold in the presence of its periplasmic binding partner, TonB. This study did determine that the trimeric porin OmpF was confined, this time with domains of about 100 nm in size (80). Overall, these studies point to the existence of confined diffusion for many elements of the OM.

Various fluorescence approaches have also demonstrated that specifically labeled proteins rarely display diffuse surface fluorescence, but instead appear in distinct surface puncta, or patches. This patch-like fluorescence has been observed for LamB, and for BtuB and Cir bound to fluorescently tagged colicin molecules (77, 79). The latter study also determined that these fluorescent patches migrated towards the cell poles, and the authors proposed that these puncta represented islands of particular OM proteins (77). The mechanism of formation for these OMP islands involves the insertion machinery. Since all OM barrels must be inserted across the membrane by the Bam complex, if the unfolded precursors of a single protein were localized to a single Bam complex, it would become locally enriched on the cell surface. The presence of a very high protein density and confined diffusion would then prevent the members of this island from spreading significantly, ensuring that it continued to increase in density. Analysis of the fluorescent patches determined that the islands had a likely diameter of about half a micron (77).

Protein-protein contacts and inter-protein distances within these domains would be mediated through a number of weak driving forces. The aromatic interactions of the girdles that surround each protein at the headgroup interface were described previously. In the case of BtuB, simulations with high densities of BtuB found that these aromatic residues, in conjunction with other hydrophobic residues buried in the lipid bilayer contributed to extensive promiscuous protein interactions (PPIs) between neighboring BtuB molecules (77). In addition to protein-protein

contacts, most OM proteins display an asymmetric height distribution around their circumference. Given that the lipid bilayer is of relatively even thickness, this results in an asymmetric lipid mismatch around the protein. For BtuB, the variation is about 11 Å in membrane thickness from the shortest to the longest β-strand (81). Lipid mismatch has been observed to drive aggregation of proteins in several molecular dynamics studies, even when protein-protein interactions are entirely omitted (82, 83).

In addition to being confined to patches, domains, or islands on the membrane, these concentrated regions of protein are in constant, slow motion towards the cell poles. In the case of LamB, this movement was dependent on the cell cycle, and was halted when cells were treated with the RNA polymerase inhibitor rifampicin (84). This indicated that the motion was entirely due to insertion events, creating a conveyor-belt like motion of the cell surface driven by stochastic insertion events which are heavily biased towards the mid-cell. This trend towards the pole is greatly accelerated at each division event, where parts of the mid-cell become new poles for each daughter cell. Since the net direction of motion is towards the poles, once proteins enter, they will tend to remain.

Additionally, the large folding free energies for OM barrels provided an excellent energy sink to drive their insertion, but also make it impossible for the cell to recycle them. The eventual effect is then a slow accumulation of old OM proteins at the cell pole. Each division event creates daughter cells with one new pole and one old one, and over time generations can be assigned to the pole contents, and by extension to the cells themselves. As the environment that the cells are subjected to changes over time, and the outer membrane is filled with porins and transport proteins that must respond to the environment, cells with older poles will become increasingly unable to compete. Phenotypically, these cells display many of the hallmarks of aging and senescence in more complex organisms, including decreased metabolic efficacy which leads to smaller offspring, and even an increased chance of death (85).

1.11 TonB dependent Transporters and BtuB

In addition to the cobalamin transporter BtuB, *E. coli encodes several* TonB dependent transporters (TBDTs) primarily for the transport of iron siderophores across the OM. In the laboratory K-strain derivatives, the most common is the ferric enterobactin transporter, FepA. Other *E. coli* iron TBDTs transport ferrichrome (FhuA), coprogen (FhuE), citrate (FhuE), catecholates (Cir), and dihydroxybenzoylserine (Fiu) (36). Beyond cobalamin and iron siderophores, other organisms encode TBDTs that transport: copper, nickel, chito-oligosaccharides, maltodextrin, thiamine, and sucrose (86). TBDTs are even critical to digestion. The starch utilization system (SUS) in many gut bacteria that helps break down the hosts dietary starches relies on a TBDT, SusC (87). Diversity in available glycans can lead some gut bacteria to express a wide variety of SUS TBDTs, with *B. thetaiotaomicron* exceeding 120 TBDT paralogs (87). Structurally, TBDTs are separated into a two-domain structure where an internal, globular hatch domain is surrounded by a 22-strand β-barrel. Crystal structures are currently available for several of these proteins, and a selection is shown below in Figure 6. The extracellular face of the transporter has a central depression for substrate binding, surrounded by a series of long, flexible loops. On the periplasmic side, the loops take the form of short turns. The hatch domains of these proteins are extensively solvated within the barrel, and some are stable on their own (88). Interactions between the barrel and hatch-domain are mediated by a series of ion-pairs, and the two domains are connected by a short linker. The N-terminal region of the hatch domain contains a short sequence known as the ton-box, which interacts with the partner protein TonB. In FhuA, there is also a switch-helix upstream of the ton-box, which changes conformation in response to substrate binding.

Several elements of the TBDT structure are known to be involved in the transport process. On the extracellular side, the long, flexible loops are capable of gating by closing over the bound substrate. This can be seen directly in the crystal structures of citrate bound and unbound FecA and has also been demonstrated via cross-linking of FepA to neighboring OMPs, which blocked substrate binding (89). Additionally, EPR studies on BtuB detected a significant reduction in the distance between loops 8 and 10 of BtuB after binding of cobalamin (90). On the periplasmic side, the ton-box is known to interact with TonB to form a mixed β-sheet. In BtuB, the ton-box spans residues 6-12 and has sequence DTLVVTA (91). Proline mutations at even numbered

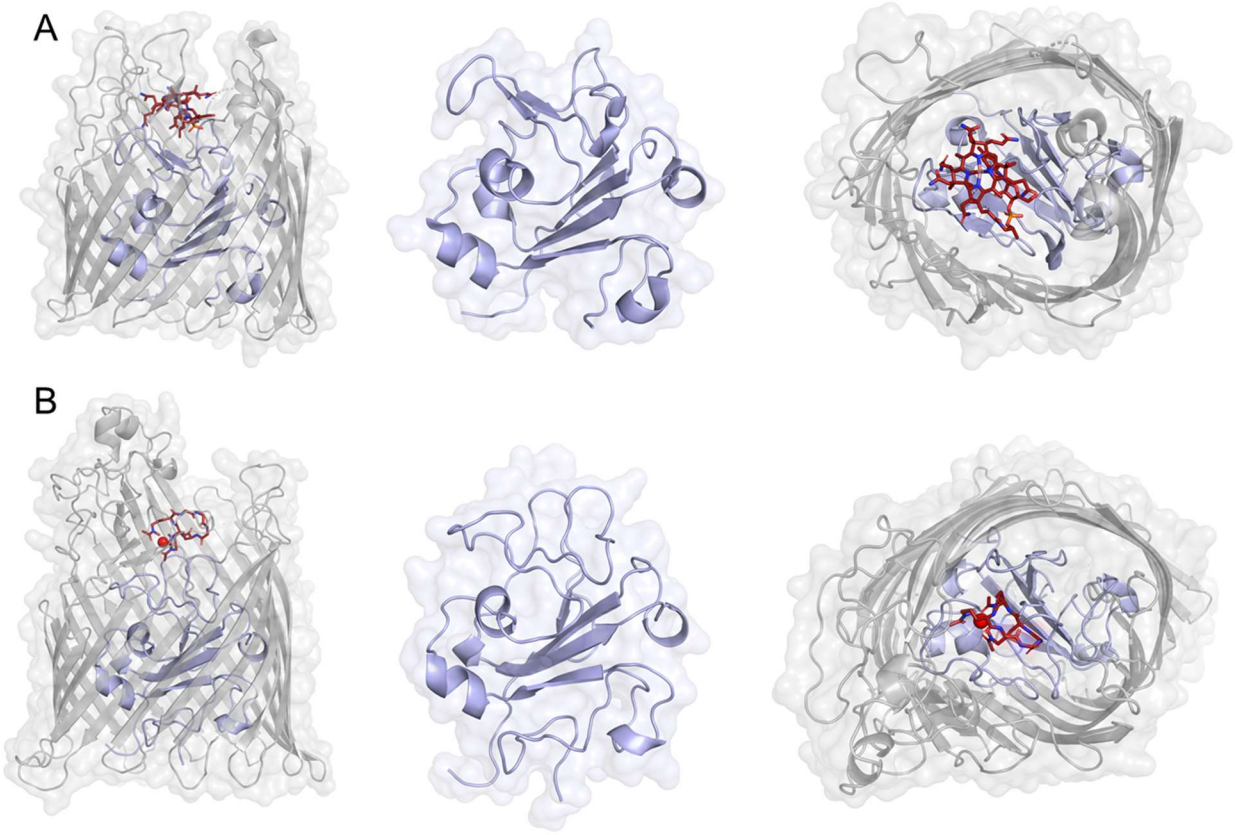


Figure 6. Structures of the TonB dependent transporters BtuB (A, pdb ID: 1NQH) and FhuA (B, PDB ID: 1BY5). The transporters are shown in gray from the side (left) and top (right), while the hatch domain is shown in blue (center). The cobalamin and ferrichrome ligands of BtuB and FhuA, respectively, are shown in dark red.

sites in the ton-box disrupt β-strand hydrogen bonding with TonB and result in a transport deficient phenotype (91). The most widely used of these transport defective mutants in BtuB are the L8P and V10P mutations. In the crystal structures of BtuB, the ton-box region displays a minor

conformational change upon cobalamin binding, flipping over inside the transporter. In EPR studies, however, lineshape analysis and pulsed-EPR measurements show a substrate dependent extension of the ton-box, potentially by several nm, into the periplasm (92). Such an extension would greatly facilitate the interaction with TonB. It is now assumed that changes in the ton-box are passed allosterically to the extracellular side of the transport, but the mechanism of these conformational changes is not currently understood (93).

Expression levels of BtuB are regulated by AdeCbl responsive riboswitches in the btuB mRNA (94). In the gut where competition for cobalamin is fierce, BtuB paralogs might be the dominant TBDTs on the OM surface. In contrast, *E. coli* natively expresses BtuB at a modest 200 copies per cell (95). The low levels may stem from its ability to utilize a cobalamin independent methionine synthase, MetE, aerobically, reducing its reliance on cobalamin uptake (96). When overexpressed with the high copy pAG1 plasmid used in this work copy numbers are instead expected at 40-50% of the native porins (95, 97). At least at native copy numbers, TBDTs are currently thought to exist as monomers *in vivo*, although early work on FepA noted that it purified as a trimer while FhuA clearly purified monomerically (98, 99). The uptake of both Ferric enterobactin and micron E492 from *Klebsiella pneumoniae* by FepA also show cooperativity with a Hill coefficient of about 3, indicative of either 3 sites per monomer or a trimeric organization (100, 101). When grown in 2d crystals, FhuA can also be induced to form dimers, and may have been detected in the dimeric form in AFM images of the surface of *R. denitrificans* (45, 102). Overall, however, most evidence points to TBDTs functioning monomerically on the OM surface.

1.12 The ExbB, ExbD, TonB complex

The Ton complex, composed of the inner membrane proteins ExbB and ExbD together with the trans-periplasmic protein TonB, is responsible for the transduction of energy from the IM to TBDTs in the OM. ExbB is generally believed to serve as a scaffolding element, while ExbD is somehow sensitive to the PMF and drives conformational changes that energize TonB. There is

extensive homology between the Ton systems and other periplasmic systems for energy transduction, namely the Ton and Mot systems. In the case of the Ton system, TolQ and TolR can even partially replace ExbB and ExbD function *in vivo*, preventing a total loss of ton system function in null mutants (103).

The first member of the energy transduction complex, ExbB, is a 26 kDa integral IM protein (104). Structures of ExbB are only available for assembled complexes with ExbD, and unfortunately come in a variety of oligomeric states. Still, the general features of each ExbB monomer are consistent. The protein forms seven helices, with two of them forming an extended, kinked helix for a total of six TM passes (105). Very little of the protein is unstructured, and each monomer has a line of six lysine residues that usually point inward toward a central core (105). Two residues have been implicated to be critical to the Ton complexes energy transduction, T148 and T181 (106).

ExbD is a smaller 17 kDa protein, with 141 residues and a single TM helix (107). The first 22 residues are cytoplasmic and may form a small structured domain (107). The TM helix extends from residues 23-42 and is followed by the 98-residue periplasmic domain (107). The TM helix contains a protonatable aspartate, D25, that is conserved in the homologous Tol and Mot system proteins and is essential for the complex to respond to the PMF (108). A deletion scanning approach with 10-residue units determined that the periplasmic domain can be further subdivided into residues 42-61, which appear to modulate PMF dependent interactions with TonB, and residues 62-141, which engaged in TonB contacts independent of PMF (109). Thus, these two regions may play different roles in the energy transduction process. These deletion scanning results agree favorably with an NMR solution structure of the periplasmic domain, which found a flexible region between residues 43-63 followed by an extended folded domain from 64-133 and finally an unstructured tail from 134-141 (110). The structured domain has two, kinked helices next to a 5-stranded sheet.

It has been frequently suggested that the periplasmic domain of ExbD may play a role in binding to the PG matrix. The domain is only about half the size of the comparable periplasmic domain of the homologous MotB, however, and has no sequence similarity in the PG binding region (109). Further, the topology of the ExbD periplasmic domain is almost identical to some periplasmic substrate binding proteins, such as FhuD, that receives ferrichrome siderophores from the FhuA TBDT (110).

Finally, TonB is also a 26 kDa protein with a single TM pass and 239 residues (104). Its first 33 residues contain the TM helix which is followed by a proline rich region from residues 66-102, a flexible linker from 103-149, and a globular C-terminal domain from 150-239 (104). TonB is an integral IM protein, but its C-terminus must interact with TBDTs on the periplasmic surface of the OM. To this end, distance measurements using pulsed-EPR revealed that the total length of the protein was about 14.5 nm, enough to span the shorter end of estimates for the width of the periplasm (21). Also, this distance was most consistent with the presence of a type II polyproline helix, where all of the proline residues are oriented in their trans isomer, for the proline-rich region from residues 66-102 (21). An extended conformation for the proline-rich region was also observed in solution NMR (111)

Crystal structures of the globular C-terminal domain of TonB show a structure containing two helices in front of a 3-stranded β-sheet. This sheet interacts with the ton-box of TBDTs,

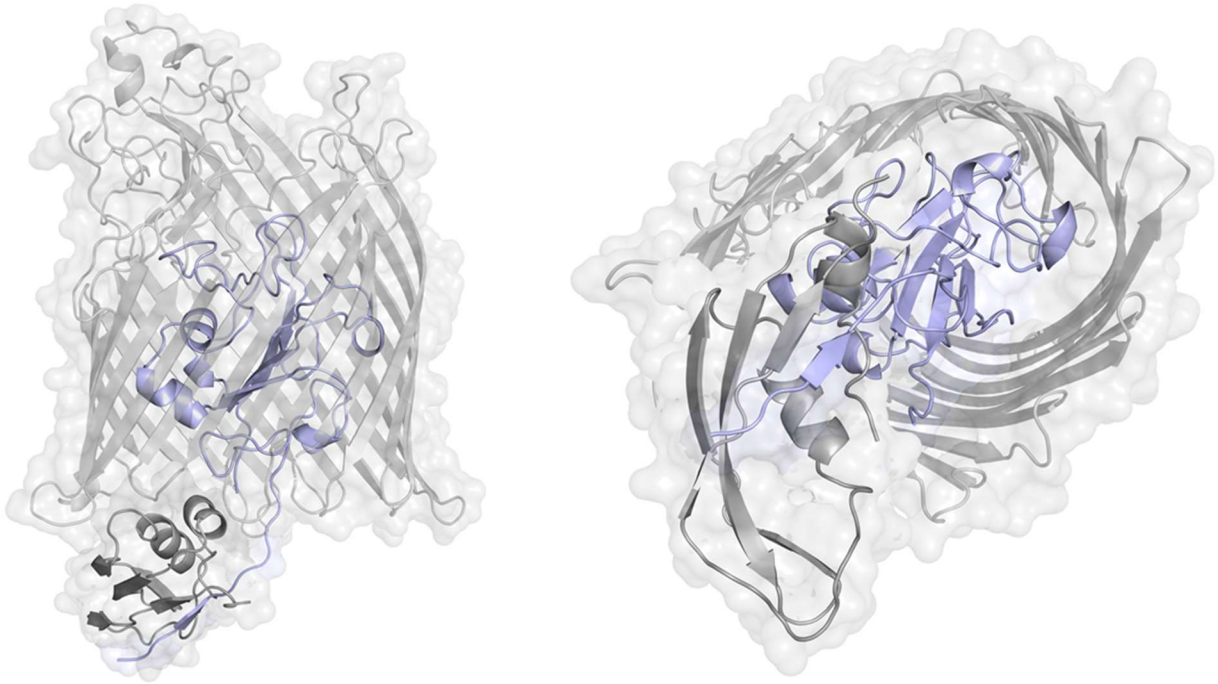


Figure 7. Structure of the interaction between FhuA and TonB (PDB ID: 2GRX) from the side (A) and bottom (B). FhuA interacts with a monomeric TonB C-terminal domain, forming a mixed 4-strand β-sheet through incorporation of the FhuA ton-box residues.

forming a stable 4-stranded sheet through which force is putatively applied to the transport protein. Binding to the TBDT has a 1:1 stoichiometry, and an example is shown in Figure 7, for the interaction of TonB with the TBDT FhuA. EPR studies of the C-terminal domain binding to the FhuA ton-box indicate that even after binding the complex remains dynamic, sampling several nm in pulsed-EPR distance measurements (112).

In addition to forming complexes with TBDTs, TonB can also form homodimers through its C-terminal domain under certain conditions (113). Two models of the dimer interface are shown in Figure 8. The initial dimeric structure, shown in Fig. 8A, used a truncated C-terminal construct with residues 164-239 (114). This construct formed a stable dimer with β-strand exchange leading to substantial contacts between the two monomers (114). Later attempts with a longer construct consisting of residues 148-239, however, produced a protein that was monomeric in solution but with dimeric crystals (Fig. 8B) (115). Here, the dimer interface is made primarily through a less extensive exchange of the C-terminal strand and the orientation is markedly different from the short construct. Follow up work with analytical ultracentrifugation

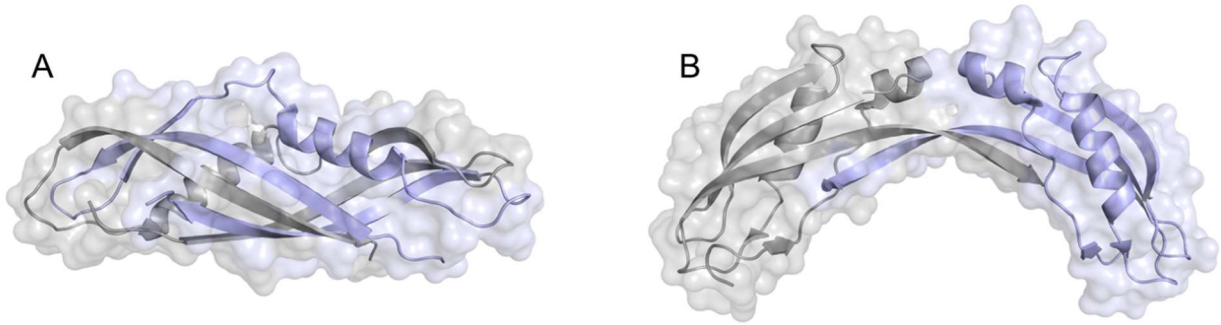


Figure 8. Models of the TonB C-terminal dimer interface. (A) The original dimer interface (PDB ID: 1IHR) obtained for shorter constructs (residues 164-239) of the TonB C-terminal domain. Here, the two domains form a stable dimer through strand exchange. (B) The dimer interface (PDB ID: 1U0) obtained for a slightly longer construct (residues 148-239) of the TonB C-terminal domain. In this case, the construct is monomeric in solution, but dimeric in the crystal.

confirmed that the shorter construct produced solution dimers, while the longer construct produced solution monomers (116, 117). Currently, there is still a great degree of controversy over the role of the dimeric form in the transport process.

In vivo, the levels of each member of the Ton complex are tightly regulated and varied primarily in response to changes in iron concentration. In cells grown in abundant iron, the WT levels of TonB are around 300 copies per cell, with a 2-fold excess of ExbD and 7-fold excess of ExbB. Iron starvation conditions or deregulation, however, can lead to the production of more than 1300 TonBs, with the ratios of ExbB and ExbD to TonB maintained at approximately 2:1 and 7:1. Thus, in the cell ExbD is more abundant than TonB, and ExbB is three or four times more abundant than ExbD (118). Various attempts have been made to determine the stoichiometry of the actual functional complex, with chromatography and *in vivo* cross-linking indicating a wide range of 2 to 6 ExbB monomers per complex (119, 120). Later work with cryo-EM has been used to generate structures of $ExbB_4/ExbD_2$ complexes, as well as heterodimeric complexes of the form $ExBB_4/ExbD_1/TonB_1$ (121). Subsequent studies using crystallography and cryo-EM, meanwhile identified several pentameric and hexameric complexes of ExbB (105, 122). Electron densities consistent with a single ExbD are found in the ExbB pentameric structures, while three ExbD helices were found within the ExbB hexamer (105, 122). It has been speculated that the complex stoichiometry may change with pH, but it nevertheless appears to be heterogeneous in these *in vitro* preparations (122).

1.13 Models for TonB Dependent Transport

Since Hancock and Brown first proposed in 1976 that the energy for TBD transport must be transduced across the IM to the transport proteins in the OM, various models have been proposed for how this might work, and the subsequent conformational changes that would be induced in TBDTs to effect transport (123). Shortly after, in 1980, it was proposed that TonB might use the PMF to generate a soluble messenger capable of dissociating cobalamin from BtuB (124). The model also proposed that BtuC would use the energy of acetyl phosphate moieties to drive cobalamin import across the IM. At the time, it was not known that BtuC formed part of the relatively new family of ABC transport proteins, and acetyl phosphate had been proposed as the energy source for active transport across the IM (125). Other models of the time included TonB functioning to bring the IM and OM into close apposition, or TonB as a regulator or protease that modulates the activities of BtuB and the IM importer (126). These early models were based primarily on null mutant phenotypes and the effects of phage and colicin binding events.

By 1990, knowledge of the system had progressed to the point that Robert Kadner proposed an alternative idea, that the TonB protein could be energized and transition between the IM and OM to provide energy to the OM receptor proteins (127). This was later expanded on by the group of Kathleen Postle into the shuttle model, which first appeared in 1997 when the Postle group determined that TonB was able to associate with both the IM and the OM (128). In the shuttle model, TonB is locked in a dynamic cycle, beginning in the inner membrane where it is energized via the PMF (129). The TM anchor of energized TonB would then dissociate in some manner, and the protein would travel to the OM, where it would release stored potential energy to the OM transporter (129). Finally, TonB would return to the IM, with its TM helix reinserting into the bilayer (129). The model was based on a variety of biochemical evidence. Stoichiometric analysis and chemical cross-linking out of the Postle group identified that there were multiple ExbB and ExbD proteins interacting with TonB as the functional energy transduction complex

(129). These interactions were also determined to involve the N-terminus of TonB, where the TM domain is located (129). The authors then speculated that the ExbB/ExbD complex might form the majority of contacts with the TonB TM helix, reducing the energy barrier for its removal from the IM (129).

Structures of TonB, showing it to behave both as a monomer and a dimer dependent on construct length and experimental conditions, also became available in the years following the shuttle model's introduction (104). This led to speculation that the dimeric form of TonB might be responsible for storing the potential energy derived from ExbB/ExbD. Early evidence for the actual shuttling process came from the determination that the N-terminus of TonB could be fluorophore labeled, and so it appeared to be accessible to the periplasm (130). The amount of labeling increased when ExbB or D were deleted, and so it appeared that they were responsible for localizing TonB to the IM (130). Unfortunately, in 2011 the shuttle model was withdrawn. A fusion of the *V.* cholerae ToxR cytoplasmic domain to the N-terminus of TonB was found to produce a protein that was active, but was unable to be dissociated from the IM fraction (131). Thus, it appeared that TonB function was not dependent on whole protein movement between the two membranes. Curiously, several years after the shuttle model was retracted a structure was published which displayed a pentameric ExbB complex, with an internal protein pore that appeared to accommodate a TM helix that was completely isolated from membrane contacts and that may be able to move up and down within the pore (105). While unlikely to facilitate actual shuttling, such a complex would agree with the main biochemical observations that led to the formation of the shuttle model.

The appearance of dimeric structures of the TonB C-terminus also inspired the propeller, or rotational model of TonB dependent transport (114). This model was further based on the extensive homology between the Tol, Mot, and Ton systems. In particular, both the ExbB/D and TolQ/R proteins are highly homologous to the MotB/A proteins (132). The latter are known to use

the PMF to drive rotation of the bacterial flagellum, and so it is reasonable to assume that the Ton and Tol complexes may act in the same way (132). The initial homology study depicted TonB as a monomer in the ExbB/ExbD/TonB complex, but most subsequent depictions of the propeller model show TonB as a dimer, with the TonB monomers either inside or outside a ring-like ExbB/ExbD complex (104, 132). Direct evidence for rotation of the TonB complex comes primarily from a single study involving a GFP-fusion to the TonB N-terminus. Using fluorescence anisotropy, it was found that there was a PMF dependent motion of the fluorophore, which was assumed to correlate to rotational motion of TonB in the IM (133). This motion would be about two orders of magnitude faster than observed in the flagellum, but the mechanical details of how it is produced are unclear (133). Only ExbD is sensitive to the PMF, and it does display a high degree of structural similarity with the ion channels formed in the Mot complex (134). So, the ExbB/ExbD complex would be assumed to form the stator of the rotational complex. In both the Tol and Mot complexes there are PG binding domains in at least one member protein, which would appear to stabilize the outer, stator portion of the complex and permit efficient rotation of the inner elements (104, 132, 134). Both ExbB and ExbD lack a known PG binding domain or motif, however, and so it would appear that the entire complex might rotate, perhaps even nonproductively, in the presence of a proton gradient (104). Still, there is some evidence that the TonB protein can itself interact with PG, based on homology with the PG binding motif of LdtC, although this too would likely lead to nonproductive rotation of the ExbB/ExbD stator instead of TonB motion (135).

The propeller model has since been refined into the rotational surveillance and energy transfer (ROSET) model (136). In this model, it is speculated that the dimeric form of TonB possesses the PG affinity, and is responsible for keeping the C-terminus in proximity to the OM (136). The monomeric form of the protein, meanwhile, is responsible for binding events to TBDTs. Building on the realization that ExbB/ExbD lack PG binding motifs, the ROSET model posits the

movement of the stator complex as a feature, which permits a single ExbB/ExbD/TonB complex to sample wide areas of the OM above (136). This is supported indirectly by the copy number disparity between TonB and TBDTs, with iron starvation conditions able to produce a more than 12-fold excess of the latter on the OM surface (137). The ROSET model paints a convenient picture of the energy transduction complex, but it still lacks direct physical evidence for most of its component pieces.

An alternative to rotational motion driving TonB dependent transport is the pulling model. In this scheme, TonB interacts as a monomer with the ton box of the TBDT, and some downward force is created and mediated through TonB, which tugs on the ton box and releases the substrate. The pulling model was likely inspired by single molecule pulling experiments to study the unfolding of IgG and ubiquitin (138, 139). As in the structures of IgG and ubiquitin, a strong 4-stranded β-sheet is formed during TonB dependent transport between the ton box of a TBDT and TonB, and this linkage appears to be stable enough for sustained force transduction. Initial evidence for the pulling model came from steered molecular dynamics simulations, where a constant loading force was applied to the ton box of BtuB, and the N-terminal hatch domain was slowly pulled out, with applied forces ranging between 50 and 350 pN (140). Unfortunately, the creation of a channel large enough to permit cobalamin passage required a total unfolding of about 20 nm (140). This is comparable to the size of the periplasm, and it is unlikely that such an extension would be possible *in vivo*. Steered molecular dynamics was repeated more recently, again finding that a channel sufficient for translocation was formed after 20 nm of pulling and unfolding of about 50 N-terminal residues of the hatch domain (141). In this case, however, single-particle AFM experiments were also performed, where a TonB C-terminus covalently attached to an AFM probe tip was used to pull on immobilized BtuB and FhuA proteins (141). These AFM experiments confirmed that the beta sheet interaction is strong enough to sustain the force load from pulling events, and that 6 to 8 nm pulling events were tolerated before rupture of

the complex was observed (141). Support for the pulling model has also been provided through cryo-EM work on the ExbB/ExbD/TonB complex, where the Coulton group has isolated several different forms of the complex that differ in the position of ExbD and TonB monomers, and in the extent of their heterodimerization (121). In one of the complexes, a downward projection of electron density was observed that may be a piston-like motion of ExbD in response to proton translocation, which could be transferred to TonB through heterodimer formation and used to drive a pulling motion at the TBDT (121).

Alternatively, the pulling motion could be provided by the structure of TonB. As noted previously, EPR measurements on TonB found that its length was consistent with a type II polyproline helix, an extended arrangement consisting of all trans-isomer proline residues (21). By exchanging environment polarity, e.g. by solvent exchange, polyproline type II helices have been observed to interconvert to the all cis type I form, which results in a 40 to 50% reduction in length (21, 142, 143). Such a conversion would provide a similar pulling distance to that tolerated by the TonB/ton-box pair in AFM experiments and could be induced by rotational motion imparted to the polyproline region of TonB once it became immobilized via ton-box binding. Alternatively, it could be induced in the pulling model by extensive heterodimerization between the TonB polyproline region and the extended domain of ExbD, potentially with subsequent motion from the piston-like rearrangement of ExbD in the ExbB/ExbD complex.

While the means by which the ExbB/ExbD/TonB complex may function to transduce energy from the PMF at the IM to the TBDTs in the OM are still controversial, even less is known about how TonB binding might drive rearrangements in the TBDT to create substrate movement across the OM. TBDTs move very large substrates that often have cross-sections almost as large as their globular hatch domains. Creation of a pore capable of substrate passage would necessitate a rearrangement of the hatch domain, or its partial or complete dissociation from the barrel. Speculating on the nature of conformational changes in the hatch domain is difficult, although some work has been done using unfolding titrations with urea and EPR spin-label

reporters to attempt to observe hatch transitions in reconstituted TBDTs (144). Partial or complete removal, meanwhile, has been frequently proposed on the basis of the high-resolution crystal structures of TBDTs, where the hatch-barrel interfaces are highly solvated by so-called lubricating water molecules (88). In combination with the ability to express stable, isolated hatch domains of some TBDTs, the waters would facilitate a total removal of the hatch from the barrel. One problem with complete removal, however, is that the linker between the hatch domain and the barrel is only a few residues long, and it is unclear how it would accommodate the large motion required for hatch removal (88). Most recently, pulling, AFM, and molecular dynamics have suggested that the N-terminal approximately 50 residues may form a structurally weak domain, which could be selectively unfolded and removed from the barrel without disturbing the remaining hatch elements (141). Still, much more work will be required in order to understand the details of how these proteins are able to move such complex substrates.

1.14 Remaining Chapters in This Work

Chapter 2 gives an overview of EPR theory and provides the methods for both whole cell and OM sample preparations of spin-labeled BtuB.

Chapter 3 demonstrates the applicability of the stationary wavelet transform to the recovery of CW EPR samples. The methodology presented in this chapter facilitates the comparative analysis of CW lineshapes for very low signal-to-noise ratio spectra in the later sections.

Chapter 4 demonstrates the applicability of CW and pulsed-EPR to the study of BtuB in its native environment, the whole cell. The role of the *E. coli* Dsb system is examined as it relates to the labeling of BtuB cysteine double-mutants, and the means of successfully producing DEER spectra in the living cell are established. DEER results are provided for both the extracellular and periplasmic faces of the protein in living cells, with both displaying potentially novel conformational changes.

Chapter 5 explores the both fortuitous and frustrating discovery of additional DEER peaks in the cellular and OM samples. These peaks are indicative of crowding and potentially of organization on the cell surface, which has been an emerging area of interest via fluorescence and AFM. The results presented in this chapter indicate a role for DEER in these explorations, but also highlight limitations in the analysis of DEER data with relation to background subtraction.

Finally, **chapter 6** extends the previous chapters by introducing almost twenty more sites across the extracellular face of the protein. Preliminary CW results are shown for these sites, some of which display significant changes with the addition of cobalamin, and their future potential for mapping out conformational motions in the protein and for further localizing its interactions with neighboring OM proteins are highlighted.

1.15 References

1. Fedosov, S. N. 2012. Physiological and Molecular Aspects of Cobalamin Transport. Water Soluble Vitamins: Clinical Research and Future Application. O. Stanger, editor. Springer Netherlands, Dordrecht, pp. 347-367.
2. Martens, J. H., H. Barg, M. Warren, and D. Jahn. 2002. Microbial production of vitamin B12. Applied Microbiology and Biotechnology 58(3):275-285.
3. Degnan, P. H., M. E. Taga, and A. L. Goodman. 2014. Vitamin B12 as a modulator of gut microbial ecology. Cell metabolism 20(5):769-778.
4. Barker, H. A., H. Weissbach, and R. D. Smyth. 1958. A COENZYME CONTAINING PSEUDOVITAMIN B(12). Proceedings of the National Academy of Sciences of the United States of America 44(11):1093-1097.
5. Hunt, A., D. Harrington, and S. Robinson. 2014. Vitamin B₁₂ deficiency. BMJ : British Medical Journal 349:g5226.
6. Degnan, P. H., N. A. Barry, K. C. Mok, M. E. Taga, and A. L. Goodman. 2014. Human Gut Microbes Use Multiple Transporters to Distinguish Vitamin B-12 Analogs and Compete in the Gut. Cell Host & Microbe 15(1):47-57.
7. Banerjee, R. 1997. The Yin-Yang of cobalamin biochemistry. Chemistry & Biology 4(3):175-186.
8. Roth, J. R., J. G. Lawrence, and T. A. Bobik. 1996. Cobalamin (coenzyme B-12): Synthesis and biological significance. Annual Review of Microbiology 50:137-181.
9. LeBlanc, J. G., C. Milani, G. S. de Giori, F. Sesma, D. van Sinderen, and M. Ventura. 2013. Bacteria as vitamin suppliers to their host: a gut microbiota perspective. Current Opinion in Biotechnology 24(2):160-168.
10. Snydman, D. R., N. V. Jacobus, L. A. McDermott, Y. Golan, D. W. Hecht, E. J. C. Goldstein, L. Harrell, S. Jenkins, D. Newton, C. Pierson, J. D. Rihs, V. L. Yu, R. Venezia, S. M. Finegold, J. E. Rosenblatt, and S. L. Gorbach. 2010. Lessons Learned from the Anaerobe Survey: Historical Perspective and Review of the Most Recent Data (2005-2007). Clinical Infectious Diseases 50:S26-S33.
11. Cronan, J. E. 2003. Bacterial Membrane Lipids: Where Do We Stand? Annual Review of Microbiology 57(1):203-224.
12. Olofsson, G., and E. Sparr. 2013. Ionization Constants pKa of Cardiolipin. PLOS ONE 8(9):e73040.
13. Oliver, P. M., J. A. Crooks, M. Leidl, E. J. Yoon, A. Saghatelian, and D. B. Weibel. 2014. Localization of anionic phospholipids in Escherichia coli cells. Journal of bacteriology 196(19):3386-3398.
14. Luirink, J., Z. Yu, S. Wagner, and J.-W. de Gier. 2012. Biogenesis of inner membrane proteins in Escherichia coli. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1817(6):965-976.
15. Farha, Maya A., Chris P. Verschoor, D. Bowdish, and Eric D. Brown. 2013. Collapsing the Proton Motive Force to Identify Synergistic Combinations against Staphylococcus aureus. Chemistry & Biology 20(9):1168-1178.
16. Locher, K. P., and E. Borths. 2004. ABC transporter architecture and mechanism: implications from the crystal structures of BtuCD and BtuF. FEBS Letters 564(3):264-268.
17. Nikaido, H., and J. A. Hall. 1998. [1] Overview of bacterial ABC transporters. Methods in Enzymology. Academic Press, pp. 3-20.
18. Dean, M., and R. Allikmets. 2001. Complete characterization of the human ABC gene family. Journal of Bioenergetics and Biomembranes 33(6):475-479.
19. Schleifer, K. H., and O. Kandler. 1972. Peptidoglycan types of bacterial cell walls and their taxonomic implications. Bacteriological reviews 36(4):407-477.

20. Graham, L. L., R. Harris, W. Villiger, and T. J. Beveridge. 1991. Freeze-substitution of Gram-negative eubacteria: general cell morphology and envelope profiles. Journal of bacteriology 173(5):1623-1633.
21. Köhler, S. D., A. Weber, S. P. Howard, W. Welte, and M. Drescher. 2010. The proline-rich domain of TonB possesses an extended polyproline II-like conformation of sufficient length to span the periplasm of Gram-negative bacteria. Protein science : a publication of the Protein Society 19(4):625-630.
22. Miller, S. I., and N. R. Salama. 2018. The Gram-negative bacterial periplasm: Size matters. PLOS Biology 16(1):e2004935.
23. Samsudin, F., A. Boags, T. J. Piggot, and S. Khalid. 2017. Braun's Lipoprotein Facilitates OmpA Interaction with the Escherichia coli Cell Wall. Biophysical Journal 113(7):1496-1504.
24. Shu, W., J. Liu, H. Ji, and M. Lu. 2000. Core structure of the outer membrane lipoprotein from Escherichia coli at 1.9 å resolution11Edited by D. Rees. Journal of Molecular Biology 299(4):1101-1112.
25. Jean, Nicolas L., Catherine M. Bougault, A. Lodge, A. Derouaux, G. Callens, Alexander J. F. Egan, I. Ayala, Richard J. Lewis, W. Vollmer, and J.-P. Simorre. 2014. Elongated Structure of the Outer-Membrane Activator of Peptidoglycan Synthesis LpoA: Implications for PBP1A Stimulation. Structure 22(7):1047-1054.
26. Nanninga, N. 1970. Ultrastructure of the Cell Envelope of Escherichia coli B After Freeze-Etching. Journal of Bacteriology 101(1):297.
27. Vollmer, W., and J.-V. Höltje. 2004. The architecture of the murein (peptidoglycan) in Gram-negative bacteria: vertical scaffold or horizontal layer(s)? Journal of bacteriology 186(18):5978-5987.
28. Barnickel, G., H. Bradaczek, H. Labischinski, and P. Giesbrecht. 1979. Conformational Energy Calculation on the Peptide Part of Murein. European Journal of Biochemistry 95(1):157-165.
29. Yao, X., M. Jericho, D. Pink, and T. Beveridge. 1999. Thickness and elasticity of Gram-negative murein sacculi measured by atomic force microscopy. Journal of bacteriology 181(22):6865-6875.
30. Demchick, P., and A. L. Koch. 1996. The permeability of the wall fabric of Escherichia coli and Bacillus subtilis. Journal of Bacteriology 178(3):768.
31. Vázquez-Laslop, N., H. Lee, R. Hu, and A. A. Neyfakh. 2001. Molecular sieve mechanism of selective release of cytoplasmic proteins by osmotically shocked Escherichia coli. Journal of bacteriology 183(8):2399-2404.
32. Typas, A., M. Banzhaf, B. van den Berg van Saparoea, J. Verheul, J. Biboy, R. J. Nichols, M. Zietek, K. Beilharz, K. Kannenberg, M. von Rechenberg, E. Breukink, T. den Blaauwen, C. A. Gross, and W. Vollmer. 2010. Regulation of peptidoglycan synthesis by outer-membrane proteins. Cell 143(7):1097-1109.
33. Paradis-Bleau, C., M. Markovski, T. Uehara, T. J. Lupoli, S. Walker, D. E. Kahne, and T. G. Bernhardt. 2010. Lipoprotein cofactors located in the outer membrane activate bacterial cell wall polymerases. Cell 143(7):1110-1120.
34. Borths, E. L., K. P. Locher, A. T. Lee, and D. C. Rees. 2002. The structure of Escherichia coli BtuF and binding to its cognate ATP binding cassette transporter. Proceedings of the National Academy of Sciences of the United States of America 99(26):16642-16647.
35. Hvorup, R. N., B. A. Goetz, M. Niederer, K. Hollenstein, E. Perozo, and K. P. Locher. 2007. Asymmetry in the Structure of the ABC Transporter-Binding Protein Complex BtuCD-BtuF. Science 317(5843):1387.
36. Nikaido, H. 2003. Molecular Basis of Bacterial Outer Membrane Permeability Revisited. Microbiology and Molecular Biology Reviews 67(4):593.

37. Nikaido, H., Y. Takeuchi, S.-I. Ohnishi, and T. Nakae. 1977. Outer membrane of Salmonella typhimurium. Electron spin resonance studies. Biochimica et Biophysica Acta (BBA) - Biomembranes 465(1):152-164.
38. Nikaido, H., and M. Vaara. 1985. Molecular basis of bacterial outer membrane permeability. Microbiological reviews 49(1):1-32.
39. Coughlin, R. T., S. Tonsager, and E. J. McGroarty. 1983. Quantitation of metal cations bound to membranes and extracted lipopolysaccharide of Escherichia coli. Biochemistry 22(8):2002-2007.
40. Takeuchi, Y., and H. Nikaido. 1981. Persistence of segregated phospholipid domains in phospholipid-lipopolysaccharide mixed bilayers: studies with spin-labeled phospholipids. Biochemistry 20(3):523-529.
41. Plesiat, P., J. R. Aires, C. Godard, and T. Köhler. 1997. Use of steroids to monitor alterations in the outer membrane of Pseudomonas aeruginosa. Journal of bacteriology 179(22):7004-7010.
42. Plésiat, P., and H. Nikaido. 1992. Outer membranes of Gram-negative bacteria are permeable to steroid probes. Molecular Microbiology 6(10):1323-1333.
43. Jo, S., J. B. Lim, J. B. Klauda, and W. Im. 2009. CHARMM-GUI Membrane Builder for mixed bilayers and its application to yeast membranes. Biophysical journal 97(1):50-58.
44. Hsu, P.-C., B. M. H. Bruininks, D. Jefferies, P. Cesar Telles de Souza, J. Lee, D. S. Patel, S. J. Marrink, Y. Qi, S. Khalid, and W. Im. 2017. CHARMM-GUI Martini Maker for modeling and simulation of complex bacterial membranes with lipopolysaccharides. Journal of Computational Chemistry 38(27):2354-2363.
45. Jaroslawski, S., K. Duquesne, J. N. Sturgis, and S. Scheuring. 2009. High-resolution architecture of the outer membrane of the Gram-negative bacteria Roseobacter denitrificans. Molecular Microbiology 74(5):1211-1222.
46. Danelon, C., A. Suenaga, M. Winterhalter, and I. Yamato. 2003. Molecular origin of the cation selectivity in OmpF porin: single channel conductances vs. free energy calculation. Biophysical Chemistry 104(3):591-603.
47. Negoda, A., E. Negoda, and R. N. Reusch. 2010. Resolving the native conformation of Escherichia coli OmpA. The FEBS journal 277(21):4427-4437.
48. Meyer, S. E., S. Granett, J. U. Jung, and M. R. Villarejo. 1990. Osmotic regulation of PhoE porin synthesis in Escherichia coli. Journal of bacteriology 172(9):5501-5502.
49. Dutzler, R., Y. F. Wang, P. J. Rizkallah, J. P. Rosenbusch, and T. Schirmer. 1996. Crystal structures of various maltooligosaccharides bound to maltoporin reveal a specific sugar translocation pathway. Structure 4(2):127-134.
50. Hong, H. D., D. R. Patel, L. K. Tamm, and B. van den Berg. 2006. The outer membrane protein OmpW forms an eight-stranded beta-barrel with a hydrophobic channel. Journal of Biological Chemistry 281(11):7568-7577. Article.
51. Miethke, M., and M. A. Marahiel. 2007. Siderophore-based iron acquisition and pathogen control. Microbiology and molecular biology reviews : MMBR 71(3):413-451.
52. Meng, X., X. Liu, L. Zhang, B. Hou, B. Li, C. Tan, Z. Li, R. Zhou, and S. Li. 2016. Virulence characteristics of extraintestinal pathogenic Escherichia coli deletion of gene encoding the outer membrane protein X. The Journal of veterinary medical science 78(8):1261-1267.
53. Otto, K., and M. Hermansson. 2004. Inactivation of ompX causes increased interactions of type 1 fimbriated Escherichia coli with abiotic surfaces. Journal of bacteriology 186(1):226-234.
54. Crane, J. M., and L. L. Randall. 2017. The Sec System: Protein Export in Escherichia coli. EcoSal Plus 7(2).
55. Sakaguchi, M., R. Tomiyoshi, T. Kuroiwa, K. Mihara, and T. Omura. 1992. Functions of signal and signal-anchor sequences are determined by the balance between the

hydrophobic segment and the N-terminal charge. Proceedings of the National Academy of Sciences of the United States of America 89(1):16-19.
56. Celebi, N., and R. E. Dalbey. 2007. 4 - YidC: A Protein with Multiple Functions in Bacterial Membrane Biogenesis. The Enzymes. R. E. Dalbey, C. M. Koehler, and F. Tamanoi, editors. Academic Press, pp. 93-109.
57. Berks, B. C. 1996. A common export pathway for proteins binding complex redox cofactors? Molecular Microbiology 22(3):393-404.
58. Kudva, R., K. Denks, P. Kuhn, A. Vogt, M. Müller, and H.-G. Koch. 2013. Protein translocation across the inner membrane of Gram-negative bacteria: the Sec and Tat dependent protein transport pathways. Research in Microbiology 164(6):505-534.
59. Zückert, W. R. 2014. Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. Biochimica et biophysica acta 1843(8):1509-1516.
60. Moon, C. P., N. R. Zaccai, P. J. Fleming, D. Gessmann, and K. G. Fleming. 2013. Membrane protein thermodynamic stability may serve as the energy sink for sorting in the periplasm. Proceedings of the National Academy of Sciences of the United States of America 110(11):4285-4290.
61. Plummer, A. M., and K. G. Fleming. 2016. From Chaperones to the Membrane with a BAM! Trends in Biochemical Sciences 41(10):872-882.
62. Krojer, T., J. Sawa, E. Schäfer, H. R. Saibil, M. Ehrmann, and T. Clausen. 2008. Structural basis for the regulated protease and chaperone function of DegP. Nature 453:885. Article.
63. Bitto, E., and D. B. McKay. 2002. Crystallographic Structure of SurA, a Molecular Chaperone that Facilitates Folding of Outer Membrane Porins. Structure 10(11):1489-1498.
64. Hu, K., V. Galius, and K. Pervushin. 2006. Structural Plasticity of Peptidyl−Prolyl Isomerase sFkpA Is a Key to Its Chaperone Function As Revealed by Solution NMR. Biochemistry 45(39):11983-11991.
65. Sklar, J. G., T. Wu, D. Kahne, and T. J. Silhavy. 2007. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli. Genes & development 21(19):2473-2484.
66. Kadokura, H., and J. Beckwith. 2009. Detecting folding intermediates of a protein as it passes through the bacterial translocation channel. Cell 138(6):1164-1173.
67. Hatahet, F., D. Boyd, and J. Beckwith. 2014. Disulfide bond formation in prokaryotes: history, diversity and design. Biochimica et biophysica acta 1844(8):1402-1414.
68. Bader, M., W. Muse, D. P. Ballou, C. Gassner, and J. C. A. Bardwell. 1999. Oxidative Protein Folding Is Driven by the Electron Transport System. Cell 98(2):217-227.
69. Missiakas, D., F. Schwager, and S. Raina. 1995. Identification and characterization of a new disulfide isomerase-like protein (DsbD) in Escherichia coli. The EMBO journal 14(14):3415-3424.
70. Rietsch, A., D. Belin, N. Martin, and J. Beckwith. 1996. An in vivo pathway for disulfide bond isomerization in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America 93(23):13048-13053.
71. Konovalova, A., D. E. Kahne, and T. J. Silhavy. 2017. Outer Membrane Biogenesis. Annual Review of Microbiology 71(1):539-556.
72. Höhr, A. I. C., C. Lindau, C. Wirth, J. Qiu, D. A. Stroud, S. Kutik, B. Guiard, C. Hunte, T. Becker, N. Pfanner, and N. Wiedemann. 2018. Membrane protein insertion through a mitochondrial β-barrel gate. Science 359(6373):eaah6834.
73. Burley, S. K., and G. A. Petsko. 1985. Aromatic-aromatic interaction: a mechanism of protein structure stabilization. Science 229(4708):23.

74. Yamashita, H., A. Taoka, T. Uchihashi, T. Asano, T. Ando, and Y. Fukumori. 2012. Single-Molecule Imaging on Living Bacterial Cell Surface by High-Speed AFM. Journal of Molecular Biology 422(2):300-309.
75. Ritchie, K., Y. Lill, C. Sood, H. Lee, and S. Zhang. Single-molecule imaging in live bacteria cells. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 368(1611):20120355-20120355.
76. Verhoeven, G. S., M. Dogterom, and T. den Blaauwen. 2013. Absence of long-range diffusion of OmpA in E. coli is not caused by its peptidoglycan binding domain. BMC microbiology 13:66-66.
77. Rassam, P., N. A. Copeland, O. Birkholz, C. Toth, M. Chavent, A. L. Duncan, S. J. Cross, N. G. Housden, R. Kaminska, U. Seger, D. M. Quinn, T. J. Garrod, M. S. P. Sansom, J. Piehler, C. G. Baumann, and C. Kleanthous. 2015. Supramolecular assemblies underpin turnover of outer membrane proteins in bacteria. Nature 523(7560):333-+. Article.
78. Oddershede, L., J. K. Dreyer, S. Grego, S. Brown, and K. Berg-Sørensen. 2002. The motion of a single molecule, the lambda-receptor, in the bacterial outer membrane. Biophysical journal 83(6):3152-3161.
79. Gibbs, K. A., D. D. Isaac, J. Xu, R. W. Hendrix, T. J. Silhavy, and J. A. Theriot. 2004. Complex spatial distribution and dynamics of an abundant Escherichia coli outer membrane protein, LamB. Molecular Microbiology 53(6):1771-1783.
80. Spector, J., S. Zakharov, Y. Lill, O. Sharma, W. A. Cramer, and K. Ritchie. 2010. Mobility of BtuB and OmpF in the Escherichia coli Outer Membrane: Implications for Dynamic Formation of a Translocon Complex. Biophysical Journal 99(12):3880-3886.
81. Ellena, J. F., P. Lackowicz, H. Mongomery, and D. S. Cafiso. 2011. Membrane Thickness Varies Around the Circumference of the Transmembrane Protein BtuB. Biophysical Journal 100(5):1280-1287.
82. Periole, X., T. Huber, S. J. Marrink, and T. P. Sakmar. 2007. G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. Journal of the American Chemical Society 129(33):10126-10132.
83. Parton, D. L., J. W. Klingelhoefer, and M. S. P. Sansom. 2011. Aggregation of Model Membrane Proteins, Modulated by Hydrophobic Mismatch, Membrane Curvature, and Protein Class. Biophysical Journal 101(3):691-699.
84. Ursell, T. S., E. H. Trepagnier, K. C. Huang, and J. A. Theriot. 2012. Analysis of surface protein expression reveals the growth pattern of the Gram-negative outer membrane. PLoS computational biology 8(9):e1002680-e1002680.
85. Stewart, E. J., R. Madden, G. Paul, and F. Taddei. 2005. Aging and Death in an Organism That Reproduces by Morphologically Symmetric Division. PLOS Biology 3(2):e45.
86. Schauer, K., D. A. Rodionov, and H. de Reuse. 2008. New substrates for TonB-dependent transport: do we only see the 'tip of the iceberg'? Trends in Biochemical Sciences 33(7):330-338.
87. Bolam, D. N., and B. van den Berg. 2018. TonB-dependent transport by the gut microbiota: novel aspects of an old problem. Current Opinion in Structural Biology 51:35-43.
88. Wiener, M. C. 2005. TonB-dependent outer membrane transport: going for Baroque? Current Opinion in Structural Biology 15(4):394-400.
89. Scott, D. C., S. M. C. Newton, and P. E. Klebba. 2002. Surface Loop Motion in FepA. Journal of Bacteriology 184(17):4906.
90. Joseph, B., A. Sikora, and D. S. Cafiso. 2016. Ligand Induced Conformational Changes of a Membrane Transporter in E. coli Cells Observed with DEER/PELDOR. Journal of the American Chemical Society 138(6):1844-1847.

91. Cadieux, N., C. Bradbeer, and R. J. Kadner. 2000. Sequence changes in the ton box region of BtuB affect its transport activities and interaction with TonB protein. Journal of bacteriology 182(21):5954-5961.
92. Kim, M., G. E. Fanucci, and D. S. Cafiso. 2007. Substrate-dependent transmembrane signaling in TonB-dependent transporters is not conserved. Proceedings of the National Academy of Sciences of the United States of America 104(29):11975-11980.
93. Sikora, A., B. Joseph, M. Matson, J. R. Staley, and D. S. Cafiso. 2016. Allosteric Signaling Is Bidirectional in an Outer-Membrane Transport Protein. Biophysical journal 111(9):1908-1918.
94. Nahvi, A., J. E. Barrick, and R. R. Breaker. 2004. Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. Nucleic acids research 32(1):143-150.
95. Fuller-Schaefer, C. A., and R. J. Kadner. 2005. Multiple extracellular loops contribute to substrate binding and transport by the Escherichia coli cobalamin transporter BtuB. Journal of Bacteriology 187(5):1732-1739.
96. Pejchal, R., and M. L. Ludwig. 2005. Cobalamin-Independent Methionine Synthase (MetE): A Face-to-Face Double Barrel That Evolved by Gene Duplication. PLOS Biology 3(2):e31.
97. Di Masi, D. R., J. C. White, C. A. Schnaitman, and C. Bradbeer. 1973. Transport of vitamin B12 in Escherichia coli: common receptor sites for vitamin B12 and the E colicins on the outer membrane of the cell envelope. Journal of bacteriology 115(2):506-513.
98. Boulanger, P., M. le Maire, M. Bonhivers, S. Dubois, M. Desmadril, and L. Letellier. 1996. Purification and Structural and Functional Characterization of FhuA, a Transporter of the Escherichia coli Outer Membrane. Biochemistry 35(45):14216-14224.
99. Liu, J., J. M. Rutz, J. B. Feix, and P. E. Klebba. 1993. Permeability properties of a large gated channel within the ferric enterobactin receptor, FepA. Proceedings of the National Academy of Sciences of the United States of America 90(22):10653-10657.
100. Thulasiraman, P., S. M. Newton, J. Xu, K. N. Raymond, C. Mai, A. Hall, M. A. Montague, and P. E. Klebba. 1998. Selectivity of ferric enterobactin binding and cooperativity of transport in Gram-negative bacteria. Journal of bacteriology 180(24):6689-6696.
101. Strahsburger, E., M. Baeza, O. Monasterio, and R. Lagos. 2005. Cooperative uptake of microcin E492 by receptors FepA, Fiu, and Cir and inhibition by the siderophore enterochelin and its dimeric and trimeric hydrolysis products. Antimicrobial agents and chemotherapy 49(7):3083-3086.
102. Lévy, D., G. Mosser, O. Lambert, G. S. Moeck, D. Bald, and J.-L. Rigaud. 1999. Two-Dimensional Crystallization on Lipid Layer: A Successful Approach for Membrane Proteins. Journal of Structural Biology 127(1):44-52.
103. Braun, V. 1989. The structurally related exbB and tolQ genes are interchangeable in conferring tonB-dependent colicin, bacteriophage, and albomycin sensitivity. Journal of Bacteriology 171(11):6387.
104. Krewulak, K. D., and H. J. Vogel. 2011. TonB or not TonB: is that the question? Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire 89(2):87-97.
105. Celia, H., N. Noinaj, S. D. Zakharov, E. Bordignon, I. Botos, M. Santamaria, T. J. Barnard, W. A. Cramer, R. Lloubes, and S. K. Buchanan. 2016. Structural insight into the role of the Ton complex in energy transduction. Nature 538:60. Article.
106. Braun, V., and C. Herrmann. 2004. Point mutations in transmembrane helices 2 and 3 of ExbB and TolQ affect their activities in Escherichia coli K-12. Journal of bacteriology 186(13):4402-4406.
107. Kampfenkel, K., and V. Braun. 1992. Membrane topology of the Escherichia coli ExbD protein. Journal of bacteriology 174(16):5485-5487.

108. Braun, V., S. Gaisser, C. Herrmann, K. Kampfenkel, H. Killmann, and I. Traub. 1996. Energy-coupled transport across the outer membrane of Escherichia coli: ExbB binds ExbD and TonB in vitro, and leucine 132 in the periplasmic region and aspartate 25 in the transmembrane region are important for ExbD activity. Journal of bacteriology 178(10):2836-2845.
109. Ollis, A. A., A. Kumar, and K. Postle. 2012. The ExbD periplasmic domain contains distinct functional regions for two stages in TonB energization. Journal of bacteriology 194(12):3069-3077.
110. Garcia-Herrero, A., R. S. Peacock, S. P. Howard, and H. J. Vogel. 2007. The solution structure of the periplasmic domain of the TonB system ExbD protein reveals an unexpected structural homology with siderophore-binding proteins. Molecular Microbiology 66(4):872-889.
111. Brewer, S., M. Tolley, I. P. Trayer, G. C. Barr, C. J. Dorman, K. Hannavy, C. F. Higgins, J. S. Evans, B. A. Levine, and M. R. Wormald. 1990. Structure and function of X-Pro dipeptide repeats in the TonB proteins of Salmonella typhimurium and Escherichia coli. Journal of Molecular Biology 216(4):883-895.
112. Sarver, J. L., M. Zhang, L. S. Liu, D. Nyenhuis, and D. S. Cafiso. 2018. A Dynamic Protein-Protein Coupling between the TonB-Dependent Transporter FhuA and TonB. Biochemistry 57(6):1045-1053. Article.
113. Mills, A., H.-T. Le, J. W. Coulton, and F. Duong. 2014. FhuA interactions in a detergent-free nanodisc environment. Biochimica et Biophysica Acta (BBA) - Biomembranes 1838(1, Part B):364-371.
114. Chang, C. S., A. Mooser, A. Pluckthun, and A. Wlodawer. 2001. Crystal structure of the dimeric C-terminal domain of TonB reveals a novel fold. Journal of Biological Chemistry 276(29):27535-27540.
115. Ködding, J., F. Killig, P. Polzer, S. Peter Howard, K. Diederichs, and W. Welte. 2005. Crystal Structure of a 92-Residue C-terminal Fragment of TonB from Escherichia coli Reveals Significant Conformational Changes Compared to Structures of Smaller TonB Fragments.
116. Koedding, J., P. Howard, L. Kaufmann, P. Polzer, A. Lustig, and W. Welte. 2004. Dimerization of TonB Is Not Essential for Its Binding to the Outer Membrane Siderophore Receptor FhuA of Escherichia coli.
117. Khursigara, C. M., G. De Crescenzo, P. D. Pawelek, and J. W. Coulton. 2004. Enhanced binding of TonB to a ligand-loaded outer membrane receptor - Role of the oligomeric state of TonB in formation of a functional FhuA center dot TonB complex. Journal of Biological Chemistry 279(9):7405-7412.
118. Ahmer, B. M., M. G. Thomas, R. A. Larsen, and K. Postle. 1995. Characterization of the exbBD operon of Escherichia coli and the role of ExbB and ExbD in TonB function and stability. Journal of bacteriology 177(16):4742-4747.
119. Higgs, P. I., P. S. Myers, and K. Postle. 1998. Interactions in the TonB-dependent energy transduction complex: ExbB and ExbD form homomultimers. Journal of bacteriology 180(22):6031-6038.
120. Pramanik, A., F. Zhang, H. Schwarz, F. Schreiber, and V. Braun. 2010. ExbB Protein in the Cytoplasmic Membrane of Escherichia coli Forms a Stable Oligomer. Biochemistry 49(40):8721-8728.
121. Sverzhinsky, A., J. W. Chung, J. C. Deme, L. Fabre, K. T. Levey, M. Plesa, D. M. Carter, P. Lypaczewski, and J. W. Coulton. 2015. Membrane Protein Complex ExbB4-ExbD1-TonB1 from Escherichia coli Demonstrates Conformational Plasticity. Journal of bacteriology 197(11):1873-1885.

122. Maki-Yonekura, S., R. Matsuoka, Y. Yamashita, H. Shimizu, M. Tanaka, F. Iwabuki, and K. Yonekura. 2018. Hexameric and pentameric complexes of the ExbBD energizer in the Ton system. eLife 7:e35419.
123. Hancock, R. W., and V. Braun. 1976. Nature of the energy requirement for the irreversible adsorption of bacteriophages T1 and phi80 to Escherichia coli. Journal of bacteriology 125(2):409-415.
124. Reynolds, P. R., G. P. Mottur, and C. Bradbeer. 1980. Transport of vitamin B12 in Escherichia coli. Some observations on the roles of the gene products of BtuC and TonB. J Biol Chem 255(9):4313-4319.
125. Hong, J. S., A. G. Hunt, P. S. Masters, and M. A. Lieberman. 1979. Requirements of acetyl phosphate for the binding protein-dependent transport systems in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America 76(3):1213-1217.
126. Wookey, P. 1982. The tonB gene product in Escherichia coli: Energy-coupling or molecular processing of permeases? FEBS Letters 139(2):145-154.
127. Kadner, R. J. 1990. Vitamin B12 transport in Escherichia coli: energy coupling between membranes. Molecular Microbiology 4(12):2027-2033.
128. Letain, T. E., and K. Postle. 1997. TonB protein appears to transduce energy by shuttling between the cytoplasmic membrane and the outer membrane in Escherichia coli. Mol Microbiol 24(2):271-283.
129. Postle, K., and R. A. Larsen. 2007. TonB-dependent energy transduction between outer and cytoplasmic membranes. Biometals 20(3-4):453-465.
130. Larsen, R. A., T. E. Letain, and K. Postle. 2003. In vivo evidence of TonB shuttling between the cytoplasmic and outer membrane in Escherichia coli. Molecular Microbiology 49(1):211-218.
131. Gresock, M. G., M. I. Savenkova, R. A. Larsen, A. A. Ollis, and K. Postle. 2011. Death of the TonB Shuttle Hypothesis. Frontiers in microbiology 2:206-206.
132. Cascales, E., R. Lloubès, and J. N. Sturgis. 2001. The TolQ–TolR proteins energize TolA and share homologies with the flagellar motor proteins MotA–MotB. Molecular Microbiology 42(3):795-807.
133. Jordan, L. D., Y. Zhou, C. R. Smallwood, Y. Lill, K. Ritchie, W. T. Yip, S. M. Newton, and P. E. Klebba. 2013. Energy-dependent motion of TonB in the Gram-negative bacterial inner membrane. Proceedings of the National Academy of Sciences of the United States of America 110(28):11553-11558.
134. Zhai, Y. F., W. Heijne, and M. H. Saier. 2003. Molecular modeling of the bacterial outer membrane receptor energizer, ExbBD/TonB, based on homology with the flagellar motor, MotAB. Biochimica Et Biophysica Acta-Biomembranes 1614(2):201-210.
135. Kaserer, W. A., X. Jiang, Q. Xiao, D. C. Scott, M. Bauler, D. Copeland, S. M. C. Newton, and P. E. Klebba. 2008. Insight from TonB hybrid proteins into the mechanism of iron transport through the outer membrane. Journal of Bacteriology 190(11):4001-4016.
136. Klebba, P. E. 2016. ROSET Model of TonB Action in Gram-Negative Bacterial Iron Acquisition. Journal of Bacteriology 198(7):1013.
137. Higgs, P. I., R. A. Larsen, and K. Postle. 2002. Quantification of known components of the Escherichia coli TonB energy transduction system: TonB, ExbB, ExbD and FepA. Molecular Microbiology 44(1):271-281.
138. Brockwell, D. J., E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, and S. E. Radford. 2003. Pulling geometry defines the mechanical resistance of a β-sheet protein. Nature Structural Biology 10:731. Article.
139. Carrion-Vazquez, M., H. Li, H. Lu, P. E. Marszalek, A. F. Oberhauser, and J. M. Fernandez. 2003. The mechanical stability of ubiquitin is linkage dependent. Nature Structural Biology 10:738. Article.

140. Gumbart, J., M. C. Wiener, and E. Tajkhorshid. 2007. Mechanics of force propagation in TonB-dependent outer membrane transport. Biophysical journal 93(2):496-504.
141. Hickman, S. J., R. E. M. Cooper, L. Bellucci, E. Paci, and D. J. Brockwell. 2017. Gating of TonB-dependent transporters by substrate-specific forced remodelling. Nature Communications 8:14804. Article.
142. Shi, L., A. E. Holliday, H. Shi, F. Zhu, M. A. Ewing, D. H. Russell, and D. E. Clemmer. 2014. Characterizing Intermediates Along the Transition from Polyproline I to Polyproline II Using Ion Mobility Spectrometry-Mass Spectrometry. Journal of the American Chemical Society 136(36):12702-12711.
143. Sachiro, K., Y. Hirano, and M. Oka. 2005. On the Stability of Polyproline-I and II Structures of Proline Oligopeptides.
144. Flores Jiménez, R. H., and D. S. Cafiso. 2012. The N-Terminal Domain of a TonB-Dependent Transporter Undergoes a Reversible Stepwise Denaturation. Biochemistry 51(17):3642-3650.

Chapter 2: Theory and Methods

2.1 EPR in Native Systems

The first true EPR result was obtained by Zavoisky in Tatarstan, Russia for hydrates of manganese sulfate and copper sulfate and was included as part of his doctoral thesis in 1944 and in publication in 1945 (1). He built the instrument himself, leveraging experience in amateur radio and in a military laboratory during WWII (1). By 1946, his discoveries had spread to the US and Europe, and by 1952 EPR was being applied to studies of biological materials. The first work on bacterial cells by EPR was likely performed in 1952 for samples of freeze-dried *pseudomonas fluorscens* by Barry Commoner, Jonathan Townsend, and George Pake at Washington University in St. Louis (2). They followed up on this work in 1954, publishing early evidence for EPR active metabolites in lyophilized plant and animal tissues (2). In the 1960s, several groups identified iron and copper compounds in *E. coli*, as well as T1 and T2 bacteriophages (3). Work by Isenberg and Baird in 1961 also identified free radical spectra in *E. coli*, which disappeared when the cells were killed via cyanide (3).

These early studies have grown into a variety of subfields, spanning EPR imaging, oximetry, cancer detection, the study of iron/sulfur complexes, and others (4) Here, the focus will be on continuous wave (CW) and pulsed measurements of bacterial outer membrane proteins in whole *E. coli* cultures, or in native OM isolate.

2.2 The Electron in the Magnetic Field

Generally, EPR deals with the detection of systems that have discrete Zeeman levels induced by the presence of a static, external magnetic field upon their unpaired electrons. For the simplest case, of an electron orbiting an isolated atom in the static field, the Zeeman energies are:

$$H_z = \beta_e g_e H S = \beta_e g_e H_0 S_z \quad (1)$$

where β_e is the Bohr magneton, g_e is the electron's spectroscopic splitting factor, H is the external field, H_0 its magnitude, and S is the spin angular momentum (5). In the case of a single unpaired electron, which will apply for all nitroxide spin labels, the allowed values of S are m_s = -½ and + ½. The difference in energy between the two spin states in this system would then be $\beta_e\, g\, H_0$ (Figure 1).

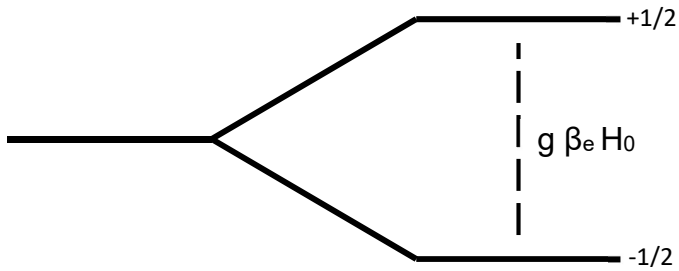


Figure 1: Zeeman splitting for an electron

Following from above, if the energy separation is $\beta_e\, g\, H_0$ then application of microwave energy can induce a transition between the two energy levels when the field strength, H_0, is equal to $\frac{h\nu_0}{g\beta_e}$.

This simple behavior applies to a single atom, but in an EPR experiment there are typically micromolar level spins within the resonant cavity of the instrument. In the aggregate, the magnetic moments of the electrons will become aligned parallel to the applied magnetic field. If said field is considered to be applied along the positive x-axis, then the bulk magnetization, M, will lie along this same axis (Fig. 2A). Further, the static field exerts a torque on the electrons due to their angular momenta, resulting in a precession about the field vector. The frequency at which this occurs is called the Larmor frequency, and is given by:

$$\omega_L = -\gamma H_0 \quad (2)$$

where γ is the gyromagnetic ratio. The presence of the field will induce the Zeeman splitting described above, with the parallel, ground state being initially more populated. In the EPR spectrometer, however, there is a second, microwave field applied perpendicular to the static field. This applied field, which will be abbreviated H_1, can induce transitions from the parallel to

the antiparallel state (Fig. 2B). This causes the bulk magnetization to no longer be fully aligned with the external field, and effectively tips it out of the positive z-axis. The bulk magnetization vector now has non-zero magnitude in the transverse axes, which allows for the detection of EPR signals by the instrument (Fig. 2C).

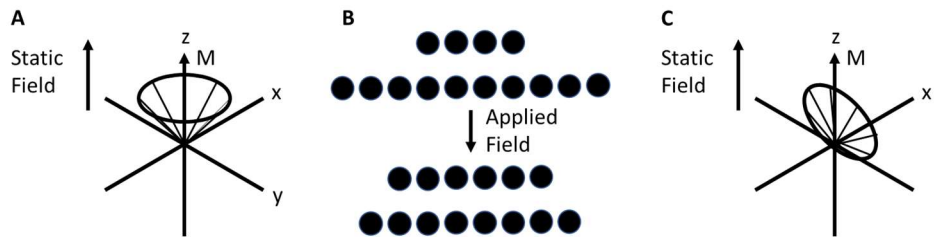


Figure 2. The behavior of the bulk magnetization in the presence of applied fields. (A) in the presence of a static field along the z-axis, the bulk magnetization aligns parallel to the static field. The electrons precess about the z-axis with random orientations. (B) The ground state parallel to the static field is initially more populated, but the addition of a perpendicular field supplied by the EPR instrument can induce transitions. (C) These have the effect of causing the bulk magnetization to no longer be fully aligned, effectively "tipping" it out of alignment.

Before proceeding from this simple picture, it is necessary to cover several reference frames, as well as the deviations from equation 1 that arise when moving from atoms to molecules. Relaxation effects are discussed next, followed by continuous wave experiments, and finally pulsed-EPR and the 4P Double-Electron-Electron-Resonance (DEER) experiment.

2.3 Choice of reference frame

1.) The laboratory frame, with axes: x, y, z. In this coordinate system, the static field is aligned along the positive z-axis, having the benefit of simplifying it to a scalar quantity in most representations.

2.) The molecular axis system, with axes: p, q, r. This coordinate system is aligned to the molecular principle axes, having the benefit of simplifying tensor variables related to molecular orientation to their trace.

3.) The rotating frame, with axes: x, y, z. This is a special modification of the laboratory frame. At resonance, where the frequency of the applied field equals the Larmor frequency, fixing

the frame of reference to said frequency renders it stationary, greatly simplifying many expressions in pulsed-EPR.

From Atoms to molecules

2.4 Anisotropy in g

For an isolated atom, the value of g_e approaches the expectation value of 2.00232 (5). In the presence of an external magnetic field and considering spin-orbit coupling, however, deviations are seen for the value of g from the expected value. The magnitude of these difference depends on the orientation of the applied magnetic field with respect to the molecule and results in g being better represented as a second-rank tensor. In the laboratory frame, the tensor can be written as:

$$g = \begin{pmatrix} g_{xx} & g_{xy} & g_{xz} \\ g_{yx} & g_{yy} & g_{yz} \\ g_{zx} & g_{zy} & g_{zz} \end{pmatrix} \text{-} \rightarrow g = \begin{pmatrix} & & \\ g_{zx} & g_{zy} & g_{zz} \end{pmatrix} \quad (3)$$

In the laboratory axis system, equation 1 can then be rewritten taking into account the g-anisotropy as:

$$H_z = \beta_e H_0 \left(g_{zx} S_x + g_{zy} S_y + g_{zz} S_z \right) \quad (4)$$

where the simplification results from the magnetic field being aligned along the +z axis (5). Likewise, in the molecular axis system, the tensor is:

$$g = \begin{pmatrix} g_{pp} & g_{pq} & g_{pr} \\ g_{qp} & g_{qq} & g_{qr} \\ g_{rp} & g_{rq} & g_{rr} \end{pmatrix} \text{-} \rightarrow g = \begin{pmatrix} g_{pp} & & \\ & g_{qq} & \\ & & g_{rr} \end{pmatrix} \quad (5)$$

The Hamiltonian in the molecular axis system now becomes:

$$H_z = \beta_e \left(g_{pp} H_p S_p + g_{qq} H_q S_q + g_{rr} H_r S_r \right) \quad (6)$$

where the extra terms H_p, H_q, H_r, reflect the fact that the vector orientation of the external field no longer coincides with one of the system axes (5). Equation 6 can also be written in terms of the direction cosines of the external field as:

$$g = \left(g_{pp}^2 l^2 + g_{qq}^2 m^2 + g_{rr}^2 n^2\right)^{\frac{1}{2}} \quad (7)$$

Finally, Equation 6 can be written in terms of the Euler angles (a, B, y) to rotate from the laboratory axis system into the molecular one. In this form, g can be represented as:

$$g = g_{pp}^2 \sin^2\beta \cos^2\gamma + g_{qq}^2 \sin^2\beta \sin^2\gamma + g_{rr}^2 \cos^2\beta \quad (8)$$

Where l, m, and n are the direction cosines that relate the external magnetic field to the molecular axis system (5). In this case, the Euler rotations are intrinsic, with each subsequent turn occurring in the new axis system created by the previous rotation. The order is an initial rotation α about the z-axis, followed by a rotation β about the effective y-axis, and finally a rotation γ about the effective z-axis, which now corresponds to the r axis in the molecular frame. Equations 7 and 8 are useful when considering simulations of EPR spectra.

2.5 Hyperfine Interactions

Within a nitroxide spin-label, the unpaired electron is also subject to interactions with nearby nuclei, particularly that of the nitrogen atom. The magnitude of these interactions is much greater than those of the g-tensor deviations for nitroxides. To account for them, equation 1 can be extended as follows:

$$H_z = -\beta_e g_e \beta_n g_n \left(\frac{(I\cdot S)r^2 - 3(I\cdot r)(S\cdot r)}{r^5} - \frac{8\pi}{3}(I\cdot S)\delta(r)\right) \quad (9)$$

where the subscript e refers to the electron, N to the nucleus, I is now the nuclear spin, r is the vector between the electron and nucleus, and δ denotes the dirac delta function (5). The left term describes the dipolar interaction, which again depends on the orientation of the molecule in the

magnetic field. The right term gives rise to Fermi contact coupling, and is dependent only on finite, nuclear electron density. Together, these two effects are frequently represented by the hyperfine interaction tensor, or matrix, A. As with the g-tensor, the representation for the A-tensor in the molecule's principal axis system is:

$$A = \begin{pmatrix} A_{pp} & A_{pq} & A_{pr} \\ A_{qp} & A_{qq} & A_{qr} \\ A_{rp} & A_{rq} & A_{rr} \end{pmatrix} \text{-} \rightarrow A = \begin{pmatrix} A_{pp} & & \\ & A_{qq} & \\ & & A_{rr} \end{pmatrix} \quad (10)$$

Allowing the hyperfine Hamiltonian to be written more simply as:

$$H_{\text{hyperfine}} = I \cdot A \cdot S \text{ -} \rightarrow \quad H_{\text{hyperfine}} = \left(A_{pp}I_pS_p + A_{qq}I_qS_q + A_{rr}I_rS_r\right) \quad (11)$$

with the A-tensor now containing components representing both the dipolar and fermi contact interactions (5).

2.6 Electron-Electron interactions

While nitroxide spin labels have only a single unpaired electron, the presence of several labels within a single protein, or of labels on nearby proteins, can lead to interactions between electrons. As with the hyperfine interactions there are two main components to electron-electron interactions, an isotropic exchange interaction and a through-space, orientation dependent dipolar interaction. The former will occur in cases where there is significant orbital overlap between the two electrons, which for spin labels would require a very close separation of only a few Angstroms. The resulting exchange coupling is frequently termed Heisenberg exchange, and has the following form:

$$H_{HE} = JS_1 \cdot S_2 = \frac{1}{2}J(S^2\text{-}S_1^2\text{-}S_2^2) = \frac{1}{2}J(S^2\text{-}\frac{3}{2}) \quad (11)$$

where S_1 and S_2 are the spin angular momentum of the first and second spins, respectively, and J is now the exchange energy (5). The result of this exchange function is singlet and triplet states for the coupled spins, with positive J leading to antiparallel orientations of the spins.

The through-space dipolar interaction has a much longer range, and for two electrons has the form:

$$H_d = g_e^2 \beta_e^2 \frac{((S_1 \cdot S_2) r^2 - 3(S_1 \cdot r)(S_2 \cdot r))}{r^5} \quad (12)$$

where S_1 and S_2 are again the spin angular momentum of the first and second spins, and r the distance between them (5). The dipolar interaction will be dealt with in more detail in the discussion of the 4P-DEER experiment.

2.7 Relaxation

The effect of the static magnetic field on the bulk magnetization was discussed in section 2.2. The presence of the static field induces a separation of the Zeeman levels for the electron, resulting in a thermal equilibrium state with more spins in the lower level parallel to the static field. The applied microwave field can then induce transitions between these levels, moving the system out of equilibrium. Of course, after the disturbance stops the system will trend back towards equilibrium. This is governed by two main rates, with corresponding constants T_1 and T_2. The former, which is also called the spin-lattice or longitudinal relaxation time, describes the return of the z-axis component of the magnetization to equilibrium, and has the following relation:

$$\frac{dM_z}{dt} = \frac{-(M_z - M_{eq})}{T_1}, \; M_{eq} = \frac{N g_e^2 \beta_e^2 S(S+1) H_0}{3kT} \quad (13, 14)$$

where the second expression for M_{eq} is taken from the Curie law (5). The spin-lattice relaxation is usually a mono-exponential decay.

The spin-lattice relaxation has the effect of shortening the time spent in the antiparallel state, returning them to the lower energy parallel condition. It thus has a direct effect on the Zeeman energy for the system. In contrast, T_2, which is also called the transverse relaxation, does not affect the Zeeman energy but instead describes the loss of coherence in the x and y axes of the magnetization. The transverse relaxation has the relation:

$$\frac{dM_{x,y}}{dt} = \frac{-M_{x,y}}{T_2} \quad (15)$$

Equations 13 and 15 are also known as the Bloch equations for the macroscopic magnetization; introduced by Felix Bloch in 1946 (5).

2.8 Continuous wave Experiments, Broadening, and Lineshapes

To continue, it is useful to consider the conditions of an actual EPR experiment. Continuous wave, or CW, experiments are widely used in the field, and involve recording characteristic lineshapes of the paramagnetic species being studied that are highly sensitive to their local environment. Two examples of CW spectra for nitroxide labels are given in Figure 3.

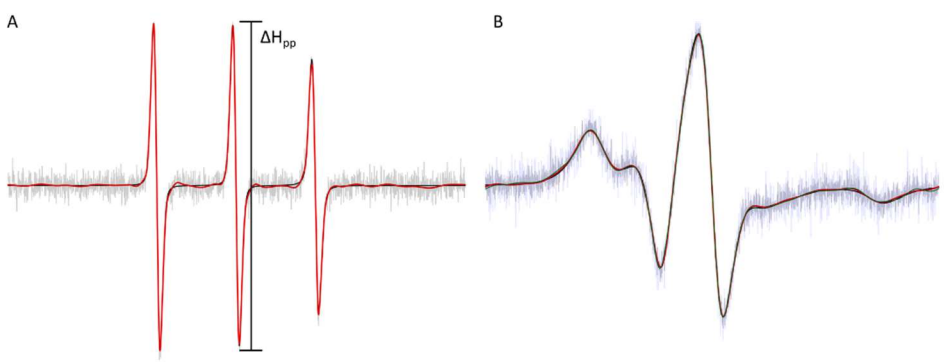


Figure 3. Examples of CW EPR spectra for (A) fast, and (B) slow motional regimes.

Starting with the spectra in Fig. 3A, it is immediately apparent that the spectrum has three main features, or peaks. In the CW experiment, the applied field H_1 is held constant, and the magnetic field is swept. As noted in section 2.2, an absorbance will occur when the magnitude of the field is equal to $\frac{h\nu_0}{g\beta_e}$. The presence of the unpaired electron leads to two-state splitting of the Zeeman levels, but this would give a single absorbance peak. The nitroxide spin label has most of its electron density in the p-orbitals of the nitrogen atom, however, which has nuclear spin I=1. This leads to further hyperfine splitting of 2I+1 levels, giving 3 total peaks as seen in this spectrum. Also, the features themselves are derivative lines rather than absorbance peaks. This stems from the use of 100 kHz field modulation in the CW EPR

instrument coupled with phase sensitive detection, which has the effect of recording the derivative of the absorbance phenomena.

Next, the shape of the lines in Fig. 3A is characteristic of a derivative Lorentzian. This indicates the dominance of the transverse relaxation in their decay. In practice, however, there is also a Gaussian contribution from inhomogeneous broadening of the spectra due to unresolved couplings, particularly with other nuclei, that makes the lineshapes a combination of both Lorentzian and Gaussian features (6). This combination is termed a derivative Voigt profile (6). The Lorentzian contribution from the transverse relaxation can be read off the spectra quite easily, as it is inversely proportional to the peak-to-peak height, ΔH_{pp}, of the central line:

$$\frac{1}{T_2} = (\frac{\sqrt{3}}{2})\Delta H_{pp} \quad (16)$$

where ΔH_{pp}, is the peak-to-peak magnitude (5). Comparing the spectrum in Fig. 3B to that of Fig. 3A, it is apparent that the lines of the spectra in Fig. 3A are sharp and narrow, while those of Fig. 3B are broadened and of greatly reduced magnitude. Even in the sharp spectrum, however, the peaks are not the same height, with the rightmost peak having significantly reduced amplitude. For lines corresponding to Lorentzian derivative functions, the height scales with the inverse of the width squared, and thus the different peaks, and by extension the two example spectra, must be differentially broadened.

To deal with this, it is necessary to add in the contributions of motion for the spin labels and their covalently bound proteins. EPR experiments are sensitive to motions on the nanosecond timescale, which captures tumbling of the spin probe as well as local conformational rearrangement of the protein backbone, and even protein tumbling for smaller proteins. As several of the effects described in the previous sections were orientation dependent, changes in orientation on the EPR timescale will result in fluctuations in the energy levels and subsequent positions of the transition frequencies for the individual spins. On the bulk level, this would be

expected to lead to broadening of the absorbance envelope, and subsequently of the observed derivative spectrum.

The degree of reorientation can be represented by the rotational correlation time, which has the form:

$$\tau_c = (6D)^{-1}, D = \frac{kT}{8\pi\eta R^3} \quad (17)$$

with τ_c representing the rotational correlation time, D the rotational diffusion constant, and η the viscosity (5). Small values of τ_c (< 1 ns) reflect a short residence time in any given conformation and are common for solution spectra with a high degree of isotropic averaging. This is true for the spectra in Fig. 3A, which is a free MTSL spin label in aqueous conditions. As the rotational correlation time increases, in contrast, there is progressive broadening of the lines. This broadening is described by the following equation:

$$(T_2(M))^{-1} = A + BM + CM^2 \quad (18)$$

where A, B, and C modulate the linewidths of the spectral lines, and M specifies the transition (5). The A term acts upon all three lines in a nitroxide spectra, and accounts for non-anisotropic factors, including paramagnetic broadening agents in the solution and instrumental uncertainty. The B and C terms, however, must act differently upon each line to give rise to the unequal line heights observed in the sample spectra in Figure 3. For a nitroxide in solution, they have the following relation:

$$(T_2(0)^{-1}\left(\frac{T_2(0)}{T_2(\pm 1)}\text{-}1\right)) = C \pm B \quad (19)$$

where $T_2(0)^{-1}$ is the width of the central spectral line, which has transition number M=0 (5). The B and C terms themselves are derivative of the anisotropy in the g-tensor and hyperfine interactions discussed earlier, but for this discussion it is sufficient to consider two simplified cases. The three lines in the nitroxide spectrum, from left to right, have M values of 1, 0, and -1,

respectively. At X-band, the value of B is approximately equal to -C, and so the contributions to the left (lowfield) and center peaks are approximately cancelled out (5). In contrast, the right peak (highfield) will have an additive effect from both terms and will appear to be substantially broader and shorter than the other two peaks. At Q-band, the B term increases in magnitude, and so while the highfield line remains the broadest, the lowfield line will begin to diverge in height from the center line (5).

These highly simplified relations work only for quickly tumbling species in solution. For cases of intermediate exchange, or highly anisotropic or solid systems it is necessary to use more rigorous treatments. The reader is referred to treatments of the Microscopic Order Macroscopic Disorder (MOMD) model, and to a review chapter by Schneider and Freed (7, 8).

Pulsed-EPR

2.9 Intro

In a pulsed-EPR experiment, the magnet again creates a static magnetic field, which is aligned with the z-axis in the laboratory frame. The microwave source, meanwhile, produces linearly polarized microwaves that create a field perpendicular to the static field, along either the x or y-axis. In contrast to the CW experiment, however, the second field does not remain on, but is instead pulsed for discrete time intervals. The behavior of the bulk magnetization, M, in the presence of a static magnetic field was presented in section 2.2. Here, the magnetization will be defined in terms of the magnetic induction, B, with the following relation:

$$M_0 = \frac{B_0}{\mu_0} - H_0 \qquad (20)$$

where H_0 is the magnitude of the static field, and μ_0 the sum of the individual spin magnetic moments (9). The magnetic induction is used in place of the magnetic field for the duration of this pulse EPR section. The static and applied fields are now denoted B_0 and B_1, respectively.

The presence of the static field creates a separation between the Zeeman levels of the affected spins, and transitions can again be induced between them when the applied RF radiation from the B_1 field is matched to the energy separation of the two levels. With B_0 aligned along the +z-axis, the spins will begin an initial precession about the same axis with frequency ω_0, also termed the Larmor frequency. To simplify the treatment of pulse experiments, the rotating frame, described in section 2.3 is used, where the frame of reference is fixed to a particular frequency. There are two choices for the reference frequency, the Larmor frequency and the frequency of B_1. The former causes the bulk magnetization to be initially stationary in the rotating frame, but the latter makes B_1 time independent. Here, the rotating frame is assumed to be fixed to the B_1 frequency, denoted $\omega_{applied}$.

With this choice of reference, any deviation between the larmor frequency of a group of spins and the B_1 frequency will cause the magnetization to again rotate with a frequency offset, ω_{off}, equal to the difference between the two. Additionally, while the intensity of B_1 is nonzero, it will induce an additional precession about the transverse axis along which the B_1 field is applied. Control over the direction of the B_1 field, and thus over the axis of precession for the magnetization, is produced by changing the phase of the linearly polarized microwaves. In the rotating frame, the direction of B_1 as a function of a phase angle, Φ, is given by:

$$\mathrm{B_{1x}} = \mathrm{B_1}\cos\Phi\,, \mathrm{B_{1y}} = \mathrm{B_1}\sin\Phi\,, \mathrm{B_{1z}} = 0 \qquad (21)$$

If B_1 is held on, the magnetization will thus continue to precess about the transverse axis along which B_1 is applied. But if the field is applied for only a brief period, on the order of tens of nanoseconds, then the magnetization will not have time to fully rotate and the length of such a pulse can be used to control its tip angle as:

$$\angle = \omega_1 \mathrm{t_p} \qquad (22)$$

where ∠ is the angle of rotation for the magnetization, ω_1 is the frequency of precession about B_1, and t_p is the pulse length. Pulses are typically named for their angle of rotation, with the most common being the π/2 pulse, producing a 90° rotation, and the π pulse, which produces a full 180° inversion.

In the simplest case, of the application of a single pulse of some angle ∠ which is ≠ 180°, the magnetization will be rotated about the chosen transverse axis. After the pulse is finished, the system will continue to precess about B_0, and will eventually relax back to equilibrium. In the rotating frame, the Bloch equations describing the relaxation are:

$$\frac{dM_x}{dt} = -\omega_{off}M_y - \frac{M_x}{T_2}, \quad \frac{dM_y}{dt} = \omega_{off}M_x - \omega_1 M_z - \frac{M_y}{T_2}, \quad \frac{dM_z}{dt} = \omega_1 M_y - \frac{M_z - M_0}{T_1} \quad (23)$$

where T_1 and T_2 again represent the spin-lattice and transverse relaxation times, respectively (9). The components of the bulk magnetization will then be described by:

$$M_x t = M_0 \sin(\angle) \sin(\omega_{off} t)\, e^{-\frac{1}{T_2}}, \; M_x t = -M_0 \sin(\angle) \cos(\omega_{off} t)\, e^{-\frac{1}{T_2}} \quad (24)$$

where it should be noted that in pulsed-EPR, the signal is normally detected in quadrature. Similar to the use of 100 kHz field modulation in CW EPR detection, the output of the microwave resonator in the pulse experiment is mixed with the microwave source, and so the detected signal is technically the offset frequency, ω_{off}. As described in equation 21, in the rotating frame, direction is represented by a phase shift, and the signal components in the x and y-axis are separated by 90°. Generally, this means the overall signal is encoded as a complex number, with the phase adjusted to maximize the signal intensity in the real. This can also be seen above in equation 24, with the exchange of the sin term for cos when moving between the M_x and M_y components. The resulting complex signal, which in this single pulse experiment decays as a pseudo mono-exponential, is called a free induction decay or FID and has the form:

$$V(t) \propto e^{i\omega_{off}t} e^{-\frac{1}{T_2}} \quad (25)$$

where V(t) will be used for all time-domain signals moving forward (9). The FID is the basis of the majority of detected quantities in pulse-EPR, including the echo, which will be explored in section 2.12.

Compared to CW EPR experiments, which for biological systems are usually performed in aqueous solution and at RT, pulsed experiments typically use frozen, solid phase samples. This presents a variety of additional problems related to relaxation and broadening effects that are explored in more detail in the following two sections.

2.10 Relaxation II

The spin-lattice or longitudinal relaxation process, which is described by T_1, was briefly described in Section 2.7. Longitudinal relaxation changes the overall system energy by changing the spin state of individual electrons, and has three main component processes: spontaneous emission, direct absorption of lattice vibrational energy, and Raman transitions.

The rate of spontaneous emission is dependent on the frequency cubed, with the form:

$$W = \frac{2}{3}\frac{\mu_0 g^2 \beta_e^2 \omega^3}{4\pi\hbar c^3} \quad (26)$$

but even in the GHz region of X and Q-band, it would still take many years for the system to decay back to thermal equilibrium through emission alone (9). Instead, the longitudinal relaxation is dominated in liquids by Brownian motion, which can be described by Redfield theory, where molecular reorientation adjusts the hyperfine, exchange, and spin-spin interaction energies, causing fluctuating local fields, which will induce transitions in spins if they match the Larmor frequency. In solids, meanwhile, longitudinal relaxation is dominated by the transfer of energy from vibrations in the lattice to the spins. This can be direct, where a unit of lattice energy, or phonon, is absorbed by the system, or it can be through a virtual Raman process. Direct absorption dominates at the extremely low temperatures used in liquid helium experiments,

typically below 5-10K, and in the intermediate regime used in the DEER experiment, of between 40-80 K, the relaxation will be primarily through Raman processes (9).

The transverse relaxation process, described by T_2, does not change the total system energy, but instead involves exchange of spin-states between different members of the system or ensemble. This is particularly relevant to pulse experiments, where pulses have limited bandwidth and both detection and excitation will affect only a subpopulation of spins in the system. Thus, transfer of the magnetization to unaffected spins will have the affect of making it effectively invisible or decreasing the measured signal intensity.

In liquids, transverse relaxation through spin flip-flops is again dependent on the rate of reorientation and can be described through Redfield Theory where in general, $T_2 \leq T_1$. In solids, rearrangement is much more limited, and many spins may be coupled, with an infinitely coupled approximation sometimes employed. Nuclear spin flip-flops are also possible, causing changes in the hyperfine interaction that again contribute to transverse dephasing. For pulse experiments in frozen samples, nuclear flip-flops via matrix protons can come to dominate longer time relaxation. The limited mobility inherent in frozen samples also brings in additional factors that influence the decay of the transverse magnetization, including spectral and instantaneous diffusion. As a result, T_2 is usually replaced with the phase memory time, T_m, which is related to the decay of spin echos described in section 2.12.

Spectral diffusion is the formal term for the situation described above, where the entire spectrum is not excited by a pulse in an experiment, creating separate pools of affected A spins and unaffected B spins. Energy can be exchanged between the A and B spin populations, and if only the A spins are detectable, this will manifest as an enhancement of the signal decay.

In contrast, instantaneous diffusion describes the fact that in a frozen, solid-phase sample, different A spins affected by a particular pulse are differentially oriented and spaced relative to

each other. This means that the application of the pulse changes the spin-spin dipolar interaction between them and has a subtle resultant affect on the local field, creating additional offsets between spin-packets and resulting in further decay of the transverse signal. The name comes from the very short timeframe of the pulses, which is near instantaneous relative to periods where the pulse is not applied, and the spin system is allowed to precess or evolve.

2.11 Homogeneous and Inhomogeneous Line Broadening

Within a spin system or ensemble, some spins will be similar enough in environment to be considered as sharing the same effective field. In pulse experiments, these groups of spins are termed packets or spin-packets. As described in section 2.8, the base shape of an EPR line for each of these spin-packets is Lorentzian and the lines will be homogeneously broadened, as they all share the same effective field. The overall EPR spectrum, in comparison, is composed of a multitude of these individual packets at different frequencies. An additive set of Lorentzian peaks has an overall Gaussian shape, and is inhomogeneously broadened. Inhomogeneous broadening is much more apparent in transition metal systems and systems where the electron is delocalized, both of which enhance anisotropic effects of the electron and nuclear interactions and the inability to resolve hyperfine couplings in the requisite spectra.

2.12 Pulse Sequences and Echos

Building from the basic concepts of pulse EPR described in Section 2.10, a simple 2-pulse experiment and the behavior of the bulk magnetization is shown in Figure 4.

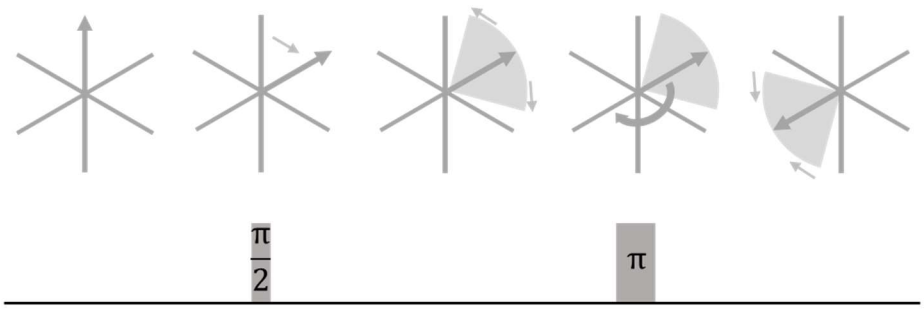


Figure 4. The behavior of the bulk magnetization in the presence of applied microwave pulses.

Initially, the bulk magnetization vector, M, is aligned with the z-axis parallel with the static field. At some time, t, a π/2 pulse is then applied, and M is rotated 90° into the transverse xy plane. When the pulse stops, the system continues to precess about the z-axis, although the individual spin packets will carry unique frequencies and thus will begin to travel at different rates in the rotating frame. As time continues, M will begin to lose coherence in the transverse plane and dephase. At this point, a π pulse is applied to rotate M 180°, effectively flipping it. This pulse can also be called a refocusing pulse, as the rate differences of the various spin packets remain constant, meaning that after the flip they now act to bring the various vectors back together, restoring the coherence of M in the xy plane. After coherence is reestablished, the spin packets will continue to move and the overall signal will again decay as they did after the application of the initial π/2 pulse. The combination of the refocusing and subsequent loss of focus period produces a quasi-exponential rise and fall in the signal, which is called an echo, spin-echo, or Hahn echo. Echos can also be considered as back-to-back FIDs.

In practice, all pulse experiments are created by stringing together sequences of pulses in one or more frequencies. The detectable quantity is either a FID or an echo, but this simple set of tools can result in a multitude of detectable quantities, ranging from relaxation times to local motion to nuclear states to the distance between pairs of excited spins. The following sections describe the Double Electron-Electron Resonance (DEER) experiment, which detects the dipolar

correlation between pairs of electron spins, and the factors that affect the experiment's performance.

2.13 The DEER Experiment

The pulse sequence for the 4-pulse DEER experiment used in the majority of the experiments in this work is shown in Figure 5. The 4-pulse experiment is derivative of an original, 3-pulse sequence, with the aim of improving the originals resolvable distance range and signal to noise ratio (SNR). The standard DEER experiment employs pulses in two different frequencies, the observe frequency and the pump frequency. A variant of the technique, the 2+1 experiment,

Figure 5. The pulse sequences and echo positions of the (A) 4-pulse DEER experiment and (B) 5-pulse DEER experiment.

requires only a single frequency but is not presented here. In the observe frequency, a series of 3 pulses with fixed separation are applied to an initial population of A-spins. The initial π/2 pulse rotates the magnetization into the xy plane, after which it is allowed to precess for a time interval t_1. A refocusing π pulse is then applied, which produces a spin-echo whose maximum occurs at a time t_1 after this first π pulse ends. The final refocusing π pulse in the observe frequency is separated from the first by a time t_2, and thus an additional refocused echo appears after this third observe pulse with a maximum at t_2-t_1.

During the second evolution, after the maximum point occurs in the first spin-echo, the dipolar coupling of the A-spins is modulated by applying a pump π pulse in the second frequency. The pump pulse excites a separate group of B-spins, with the change in magnitude of the dipolar interaction influencing the height of the second refocused echo. The magnitude of this effect is time dependent, and so the experiment is repeated many times, with the position of the pump

pulse being stepped through the available time interval t_2-t_1 during which the spins precess before the second refocusing pulse is applied in the observe frequency.

The time-domain signal obtained from the DEER experiment has a decaying oscillatory form, and is given by the following expression:

$$V(t) = (1\text{-}\lambda(1\text{-}\int_0^1 \cos(\omega_{ee}(1\text{-}3\cos^2\theta_i)t)d\cos\theta))B(t),\ \omega_{ee} = \frac{\mu_0\gamma_e^2\hbar}{4\pi r^3} \qquad (27)$$

where V(t) is the signal, λ is the inversion frequency, being the fraction of B spins that are coupled to A spins and successfully inverted by the pump pulse, ω_{ee} is the dipolar coupling frequency, θ is the angle between the dipolar coupling vector and the static field, r its distance, and γ_e is the electron gyromagnetic ratio (10, 11). The term B(t) represents the background function contributing to additional signal decay. Typically, the DEER experiment is performed for biradicals, or for proteins with two spin labels attached at different locations. Thus, the desired information is the intramolecular dipolar coupling. The presence of other molecules leads to a significant contribution from intermolecular couplings, however, and so the signal is often represented as a product of the intramolecular form factor F(t) and a background function B(t) as given in equation 28:

$$V(t) = F(t)B(t) \qquad (28)$$

Determination of the background term is reliant on assumptions regarding the nature of the molecular level organization of a sample. The most common choice involves assuming a homogeneous distribution of hard-sphere particles according to the shell-factorization model, where the distribution is assumed to have fractal dimension d. Such a background creates a stretched-exponential signal decay with the form:

$$B(t) = \exp(\text{-}k_{id}t^{\frac{d}{3}}) \qquad (29)$$

where k_{id} is the rate of instantaneous diffusion, and makes the degree of background decay dependent on the spin concentration (9). The value of the fractal dimension d loosely maps to the geometry of common sample conditions, with d≈3 for aqueous or detergent-solubilized proteins and d≈2 for proteins on cellular surfaces or in lipid vesicles. In practice, the dimension varies between values of slightly below 2 up to 4, with values greater than 3 typically indicating violations of the shell-factorization model in the form of crowding or excluded-volume effects (12, 13).

2.14 Data Processing and Extraction of the Intramolecular Distances

An example of the data produced in a DEER experiment, and its subsequent analysis is presented in figure 6. The oscillatory decay form of the time-domain signal is immediately apparent in Fig. 6A, but as mentioned above, it contains information on both the intramolecular interaction of interest and the unwanted intermolecular interactions. Separation of the background signal from the desired intramolecular form factor, and conversion from the time-

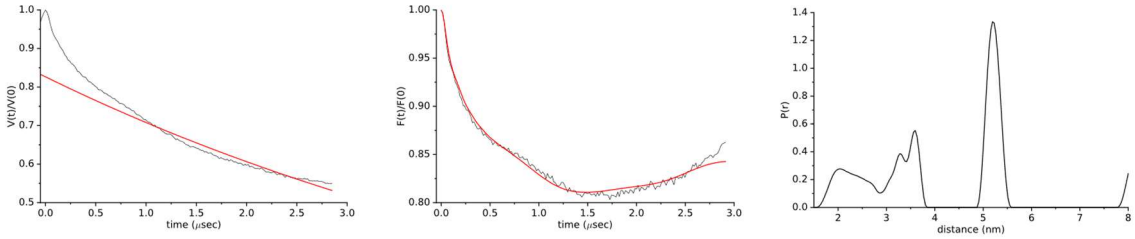


Figure 6. Examples of (A) time-domain data V(t), (B) the intramolecular form-factor F(t), and (C) the resulting distance distribution for a 4-pulse DEER experiment.

domain data to a distance distribution can be handled with either a model-free, or model-based approach.

2.15 Preprocessing

The original time-domain signal requires several adjustments before either method can be employed for extraction of the pair-wise distances. First, the signal-acquisition time usually begins before the maximum of the spin-echo is reached, and so the zero-time must be determined. This

typically involves either a Gaussian-fit to the initial data points, or a simple selection of the greatest magnitude value in the dataset. Additionally, the end of time-traces for 4-pulse DEER experiments using rectangular pulses typically contains an artifact resulting from the imperfect separation of the observe and pump pulses. This artifact is the result of the first two observe pulses and the overlapping portion of the pump pulse producing an analogue of the 2+1 experiment mentioned briefly at the beginning of this section (14). The 2+1 experiment is a single-frequency variant for the detection of electron-electron coupling, and it manifests as additional oscillations in the data that become more apparent at the end of the time-trace. Fortunately, the effect of the 2+1 artifact can be minimized by subtraction of the final 5-10 points of the trace, or by using Gaussian or other shaped pulses in place of the traditional rectangular pulses during the experiment. Finally, the time-trace is recorded in quadrature, and the data is typically phased using a Hilbert transform to move all of the signal intensity into the real component.

2.16 Model-Free Analysis

Changing between the initial time-trace and the parwise distance distribution is an ill-posed problem, requiring some method to solve a Fredholm equation of the first kind, with varied possible solutions. The most common method used to deal with this problem is Tikhonov regularization, which has been implemented in software available from the Acert Group at Cornell, and in the widely used DeerAnalysis package by the Jeschke group at ETH Zurich (15, 16). Prior to extraction of the distance distribution, a background function of the stretched-exponential form given in Equation 29 is fit to the data with the fractal dimension being varied, and the resulting equation is subtracted from the time-trace to yield the intramolecular form factor shown in Figure 6B. This can then be used for extraction of the pairwise-distance distribution, as in Figure 6C.

For this case, the Tikhonov functional has the following form:

$$\Phi[P] = \|KP\text{-}S\|^2 + \lambda^2\|LP\|^2 \quad (30)$$

where || || is the Euclidean norm, S is the time-domain data vector with size N, P is the distance-domain solution vector with size M, L is an identity or second-derivative matrix operator, and K is an MxN matrix whose elements contain the discrete approximation for the dipolar interaction between spin-pairs (16). The left-hand side of equation 30 is the residual norm, which is minimized in standard linear regression. Unfortunately, in this case minimizing the residual norm alone produces a solution that is derived mainly from noise and imprecision in the data. Thus, the solution norm in the second term is preceded by a regularization factor, λ, which enforces a degree of smoothness on the solution. The solution to equation 30 can be expressed in terms of the value of the regularization factor as:

$$P_\lambda = (K^TK + \lambda^2L^TL)^{-1}K^TS \quad (31)$$

but some method must still be employed to select the optimal value of λ (16). The most widely used when fitting DEER data is the L-curve, which is a plot of the solution norm on the y-axis against the residual norm on the x-axis for all tested values of λ. The shape of the curve is always somewhat L shaped, and the corner of the curve is usually selected as the optimum value of λ. Qualitatively, this location represents the point where the agreement between the solution and the data no longer changes rapidly with increasing smoothness of the solution.

Recently, a variety of improvements have been developed for fitting using Tikhonov and related model-free approaches. The program LongDistances by Christian Altenbach at UCLA uses a variant smoothing-factor in place of the classical regularization parameter, which may increase the quality of the resulting solutions. The L-curve method for selection of λ can also be replaced, with Edwards et al. finding that information criterion values, such as the Akaikake information criterion (AIC) or Bayesian information criterion (BIC), as well as cross-validation methods, can all outperform the L-curve in many situations (17). More recently, several

alternatives to Tikhonov regularization were tried, with Ibanez et al. determining that Osher's Bregman-iterated regularization (OBIR) typically performed the best across a dataset of more than 620,000 noisy signals derived from a T4 lysozyme crystal structure (18). Finally, Worswick et al. have produced a feed-forward neural network, DEERNet, which functions as an alternative, general Fredholm solver for these problems (19).

2.17 Model-Based Analysis

In contrast to the model-free methods, DEER distance distributions can also be modeled as a sum of Gaussian components. This approach allows for the background function to be cofit with the distribution. Under this scheme, the form-factor in Figure 6B is produced at the same time as the distance distribution in Figure 6C. This is particularly beneficial in cases where the time-trace is relatively short, making background determination and subtraction before fitting difficult. A simple approach for fitting of n Gaussian components and a homogeneous background is present in the DeerAnalysis package mentioned above, but more rigorous implementations are available in LongDistances and particularly in the proGram DD by Eric Hustedt at Vanderbilt (20). The latter package also allows for adjustment of the background factor to one that accounts for deviations related to excluded volume effects, where the spin-labels are only able to approach to a certain non-zero minimum distance.

As the distribution is modeled as a set of discrete Gaussian components, the selection of the number of peaks can be handled easily by finding the minimum of the AIC or BIC values. This is in sharp contrast to the difficulty in determining the optimal value of λ in the Tikhonov regularization methods (20). Overall fit quality is generally assessed through the value of χ^2after fitting, with a value near 1 being an ideal fit where the error is explained solely by data noise (20).

Methods

2.18 Plasmids and Strains

Plasmid pAG1 (Figure 7) containing the *btuB* gene was originally provided by the Kadner group at the University of Virginia. *E. coli* K-12 strain RK5016 (*-argH, -btuB, -metE*), and the following strains from the Coli Genetic Stock Center at Yale: strain RI89 (*araD139* Δ(*araABC-leu*)*7679 galU galK* Δ*(lac)X74 rpsl thi phoR* Δ*ara714 leu*$^{+}$*),* strain RI90 (RI89 *dsbA:: Kan*^{r}), strain RI317 (RI89 *dsbA:: Kan*^{r}), and strain RI179 (RI89 *dsbC:: Kan*^{r}), were provided by the Nakamoto lab at the University of Virginia. Top10 cells were purchased from Thermo Fisher Scientific (Waltham, MA).

Figure 7. Sequence map of the pAG1 vector used for BtuB expression in this work. NGS sequencing performed by Applied Biological Materials Inc. (Richmond, BC). Image generated in SnapGene.

2.19 Mutants

BtuB mutants based on the following sites: 6, 74, 90, 188, 288, 384, 510, were generated in the Cafiso lab by Dr. Thushani Nilaweera or previous members. Primers were purchased through Integrated DNA Technologies (Coralville, IA).

BtuB mutants based on the following sites: 55, 57, 62, 63, 65, 67, 72, 91, 93, 98, 237, 304, 330, 403, 404, 450, 451, 491, 493, 494, were purchased from Applied Biological Materials Inc. (Richmond, BC, Canada).

2.20 Standard Materials

Mono and disodium phosphate were obtained from Chem-Impex (Wood Dale, IL). Ampicillin and AEBSF were obtained from Cayman Chemical (Ann Arbor, MI). HEPES was obtained from Bio-Basic (Toronto, Ontario, Canada). MES was obtained from GoldBio (St. Louis, MO). Glucose was obtained from Sigma-Aldrich (St. Louis, MO). Ammonium sulfate, sodium citrate, thiamine, methionine, arginine, calcium chloride, magnesium sulfate, and sarkosyl were obtained from Fisher Scientific (Waltham, MA). DNAse was obtained from Applied Biological Materials Inc. (Richmond, Britich Columbia, Canada).

2.21 Nonstandard Materials

The tempo-labeled cyanocobalamin analogue (tempo-B12) was provided by Dr. Benesh Joseph (University of Frankfurt, Germany). The MTSSL spin label was purchased from Cayman Chemical (Ann Arbor, MI). CW EPR and pulse experiments used glass capillaries purchased from Vitrocom (Mountain Lakes, NJ) of the following dimensions: 0.84 mm O.D. x 0.6 mm I.D. for CW experiments, and 1.6 mm O.D. x 1.1 mm I.D. for pulse experiments.

2.22 Preparation of Whole Cell Samples

Preparation of whole *E. coli* cells with spin-labeled, overexpressed BtuB followed a general schema, with a variety of substitutions for: growth time, buffer composition, spin label concentration, number of washes, and length of wash steps. The ranges for each are given in the following text, and specific values will be noted as needed in the following chapters.

Plasmids containing the desired single or double-mutation in the *btuB* gene were transformed into one of the strains above. A single colony was then used to inoculate minimal media (MM, 100 mM phosphate buffer, 8 mM (NH_4SO_4), 2 mM sodium citrate, 200 µg/mL ampicillin, 0.2% w/v glucose, 150 µM thiamine, 3 mM $MgS0_4$, 300 µM $CaCl_2$, 0.01% w/v methionine, and 0.01% w/v arginine) precultures, which were grown overnight at 34°C. Precultures were then used to inoculate MM maincultures, which were grown to an optical density (OD) of either 0.3 or 0.6, corresponding to early and mid-log phase growth, respectively. Cells were then collected in 50 mL aliquots and centrifuged at 3260 x g for 10 minutes at 4° C. The resulting cell pellet was resuspended in 10 mL of wash buffer (100 mM HEPES, pH 7 or 100 mM MES, pH 5.5) and MTSSL was added for spin labeling at concentrations ranging from 0.01 to 5 mg / mL. Labeling proceeded in the dark for between 15 minutes and 1 hr, after which the cells were resuspended in wash buffer and incubated either on-ice or at RT for up to 1 hour. Final collection of the cell pellets involved centrifugation at 3260 x g for 6 minutes and subsequent surface washes prior to use. The particular conditions for each experiment will be noted as they appear in later chapters.

2.23 Preparation of Isolated Outer Membranes

Plasmids containing the desired single or double-mutation in the *btuB* gene were transformed into one of the strains above. A single colony was again used to inoculate 10 mL MM precultures, which were grown overnight at 34°C. Precultures were used to ioculate MM maincultures, which were grown for 8 hours. Cells were pelleted at 6,080 x g for 10 minutes and the pellets were resuspended in 10-15 mL of 0.1 M HEPES, pH 8.0 (HEPES buffer), supplemented with 20 µL of 20 mM AEBSF and 1 µL of DNase. Cells were disrupted using a French pressure cell, and the cell debris was pelleted out through centrifugation with an SS-34 rotor at 17,210 x g for 20 minutes. The resulting supernatant was then spun at 118,730 x g in a floorstand ultracentrifuge for 60 minutes to pellet the membrane fractions, after which the pellets

were resuspended in HEPES buffer supplemented with 1% sarkosyl. The resuspension was then incubated at 37° C for 30 minutes, followed by a second 60 minute spin at 118,730 x g. The pellet was resuspended in HEPES buffer to a final volume of 5 mL, and 1-2 mg of MTSSL was added for spin labeling, which proceeded for 2 hours in the dark. Following spin labeling, the sample was spun at 125,750 x g for 20 minutes. The supernatant was discarded, and the pellet surface washed 3 times with HEPES buffer. The pellet was resuspended and spun in a tabletop-ultracentrifuge at 156,424 x g for 20 minutes. The pellet was surface washed twice with HEPES buffer, and manually resuspended into the same buffer using a pipette tip. The tabletop spin-wash-resuspend cycle was repeated two or three additional times, after which the pellets were resuspended in minimal volume HEPES buffer and flash frozen in liquid nitrogen for storage.

2.24 Preparation of EPR Samples

For CW samples, approximately 6 μL of concentrated cell pellet or minimally resuspended OM was loaded into 0.84 mm x 0.6 mm capillaries. The pellet was spun to the bottom of the capillary using a hand-crank centrifuge, and the tubes were loaded into an ER 4123D dielectric resonator.

For DEER experiments, 16 μL of concentrated cell pellet or minimally resuspended OM was combined with 4 μL of deuterated glycerol. To test the effect of substrate, some also contained cyanocobalamin, although the total sample volume was kept between 20 and 22 μL. Samples were loaded into 1.6 mm x 1.1 mm capillaries and flash-frozen either in liquid nitrogen or a dry-ice and isopropanol mixture. The flash-frozen samples were then stored in a -80 freezer until use.

2.25 CW Experiments

All CW experiments were performed on a Bruker EMX spectrometer. Experiments were conducted at RT, with a 100 Gauss sweep-width, 1 Gauss modulation, and 2 mW incident power.

Spectra were recorded either as an additive sum of n scans or as a time-scan, where each successive scan was added row-wise to form a 2d matrix. The separation between scans in each time-scan was marginally greater than the per-scan time of 21 seconds. Phasing and normalization of the data used my own software written in python and leveraging either the matplotlib package or the DASH and Plotly.py libraries by Plotly. Subtraction of free-spin from the spectra used LabVIEW software provided by Dr. Christian Altenbach at UCLA.

2.26 DEER Experiments

All DEER experiments were performed on a Bruker E580 spectrometer operating at Q-band with an EN5107D2 dielectric resonator. For experiments detailed in Chapters 2 and 3, the instrument was equipped with a SpinJet AWG, a 10 W solid state amplifier, and experiments were run at 80 K. For the experiments detailed in Chapter 5, the instrument was instead equipped with a 300 W TWT amplifier and a cryogen-free cooling system, which allowed experiments to be performed at 60 K. Most experiments used the standard dead-time free 4-pulse DEER experiment, with an 18 ns π/2 pulse and 36 ns π pulses. The separation between observe and pump frequencies was 75 MHz. In chapters 3 and 4, all pulses were rectangular.

Data analysis for DEER experiments used one of DeerAnalysis2015, DD version 7.b, or LongDistances version 785 depending on the application. In DeerAnalysis, data were loaded and the phase and start time of the decay were determined automatically in the proGram. The 2+1 pulse artifact was minimized by removing the number of data points required to remove any nonzero oscillations in the imaginary data channel. Typically, this resulted in the removal of 5 or 6 data points. The background start time was determined automatically, but the background dimension was determined manually between values of 2.0 and 3.0 by examining the curvature of the resulting form factor after subtraction. Distance distributions were fit using the standard Tikhonov-regularization options, and the value of the regularization parameter was determined

by selecting either the corner of the L-curve, or one point lower if it significantly improved the visual quality of the fit.

In DD, data were loaded and the initial preprocessing was handled by the software. A total of approximately 50 ns of data was removed to compensate for the 2+1 pulse artifact, which is equivalent to the removal of 5 or 6 data points as in DeerAnalysis. The position, width, and percentage contribution of each Gaussian component were free components in the fit. Additional free components were the background parameters, the modulation depth, and a scaling factor. The number of Gaussian components was picked by manually refitting for 1 to 4 components and selecting the fit with the lowest value of the AIC and BIC.

In LongDistances, data were loaded and the phase and start time were determined automatically in the proGram. 5 to 6 data points were truncated from the end to suppress the 2+1 artifact, as in the DeerAnalysis procedure. The background was fit to a stretched exponential with a quadratic dimension expansion of $\mathrm{B(t)} = \exp(\mathrm{b_1} + \mathrm{b_2 t} + \mathrm{b_3 t^2})$. Distances were then fit one of two ways. 1. A model free approach where the smoothing factor was picked based on the corner of the L-curve as in DeerAnalysis. 2. A model based approach using Gaussian components. In the model based approach, the initial background subtracting was refined by background cofitting, as in the DD protocol.

2.27 Figure Generation

Figures for CW experiments were generated using my own software written in python, matplotlib, QT 5, and the PYQT python bindings to the QT framework. Figures for DEER experiments were generated in the same way using a different version of the CW proGram, or with one of my web-dashboards running on the Heroku cloud platform, again written in python and the DASH and plotly.py libraries by Plotly. Structure figures were generated using Pymol or UCSF Chimera.

2.28 References

1. Ptushenko, V. V., and N. E. Zavoiskaya. 2017. EPR in the USSR: the thorny path from birth to biological and chemical applications. Photosynthesis Research 134(2):133-147.
2. Commoner, B., J. Townsend, and G. E. Pake. 1954. Free Radicals in Biological Materials. Nature 174:689.
3. Whiffen, R. J. C. a. D. H. 1962. Electron spin resonance and its applications. Physics in Medicine & Biology 7(3):277.
4. Berliner, L. J. 2003. In vivo EPR (ESR) : theory and application. Kluwer Academic/Plenum Publishers, New York.
5. Berliner, L. J. 1976. Spin labeling : theory and applications. Academic Press, New York; London.
6. Robinson, B. H., C. Mailer, and A. W. Reese. 1999. Linewidth Analysis of Spin Labels in Liquids: I. Theory and Data Analysis. Journal of Magnetic Resonance 138(2):199-209.
7. Schneider, D. J., and J. H. Freed. 1989. Calculating Slow Motional Magnetic Resonance Spectra.1-76.
8. Meirovitch, E. 1984. Molecular configuration, intermolecular interactions and dynamics in smectic liquid crystals from slow motional ESR lineshapes. Journal of physical chemistry (1952) 88:4995.
9. Schweiger, A., and G. Jeschke. 2001. Principles of pulse electron paramagnetic resonance. Oxford University Press, Oxford, UK; New York.
10. Jeschke, G. 2012. DEER Distance Measurements on Proteins. Annual Review of Physical Chemistry 63(1):419-446.
11. Borbat, P., E. Georgieva, and J. H Freed. 2013. Improved Sensitivity for Long-Distance Measurements in Biomolecules: Five-Pulse Double Electron–Electron Resonance.
12. Brandon, S., A. H. Beth, and E. J. Hustedt. 2012. The global analysis of DEER data. Journal of Magnetic Resonance 218:93-104.
13. Kattnig, D. R., J. Reichenwallner, and D. Hinderberger. 2013. Modeling Excluded Volume Effects for the Faithful Description of the Background Signal in Double Electron–Electron Resonance. The Journal of Physical Chemistry B 117(51):16542-16557.
14. Teucher, M., and E. Bordignon. 2018. Improved signal fidelity in 4-pulse DEER with Gaussian pulses. Journal of Magnetic Resonance 296:103-111.
15. Jeschke, G., V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C. R. Timmel, D. Hilger, and H. Jung. 2006. DeerAnalysis2006 - a comprehensive software package for analyzing pulsed ELDOR data. Applied Magnetic Resonance 30(3-4):473-498.
16. Chiang, Y. W., P. P. Borbat, and J. H. Freed. 2005. The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. Journal of Magnetic Resonance 172(2):279-295.
17. Edwards, T. H., and S. Stoll. 2018. Optimal Tikhonov regularization for DEER spectroscopy. Journal of Magnetic Resonance 288:58-68.
18. Fábregas Ibáñez, L., and G. Jeschke. 2019. General regularization framework for DEER spectroscopy. Journal of Magnetic Resonance 300:28-40.
19. Worswick, S. G., J. A. Spencer, G. Jeschke, and I. Kuprov. 2018. Deep neural network processing of DEER data. Science Advances 4(8):eaat5218.
20. Stein, R. A., A. H. Beth, and E. J. Hustedt. 2015. A Straightforward Approach to the Analysis of Double Electron-Electron Resonance Data. Electron Paramagnetic Resonance Investigations of Biological Systems by Using Spin Labels, Spin Probes, and Intrinsic Metal Ions, Pt A 563:531-567.

Wavelet Denoising of Slow-Motional CW EPR Spectra

3.1 Background and Transforms

CW EPR spectra are commonly used for both qualitative lineshape analysis and quantitative comparisons in titrations or other series evaluations. In the former case, the shape of the signal is crucial in determining the environment and thus location of the paramagnetic label. In the latter case, reliable measures of peak height and width are critical to the determination of accurate binding, dissociation, or other constants. In practice, however, all EPR data is convolved with noise, which for many protein EPR samples dominates the spectra. Traditionally, the signal-to-noise ratio (SNR) is enhanced by additive averaging, with successive scans being added together to achieve a square-root n improvement in SNR. Still, the relatively long time-per-scan on commercial hardware (about 21 seconds for the Bruker EMX), and low gain through additive averaging can result in some samples being scanned for many hundreds of scans and thus many hours. Additionally, there is usually a residual underlying contribution to the signal from the resonator cavity in which the sample is held, and its contribution increases with the number of scans. Thus, the determination of alternative ways to improve SNR is highly desirable.

The approach presented here uses wavelet threshold denoising to recover underlying signal traces from noisy EPR data. Wavelet denoising builds upon earlier work in the field of Fourier denoising. Where Fourier transforms provide information on the frequency content of a given signal, the finite basis of wavelet transforms allows the recovery of both frequency and spatial information. Generally, wavelets are basis functions in Hilbert space, which exist in families defined by their mother wavelet, ψ(t). The continuous wavelet transform can be written:

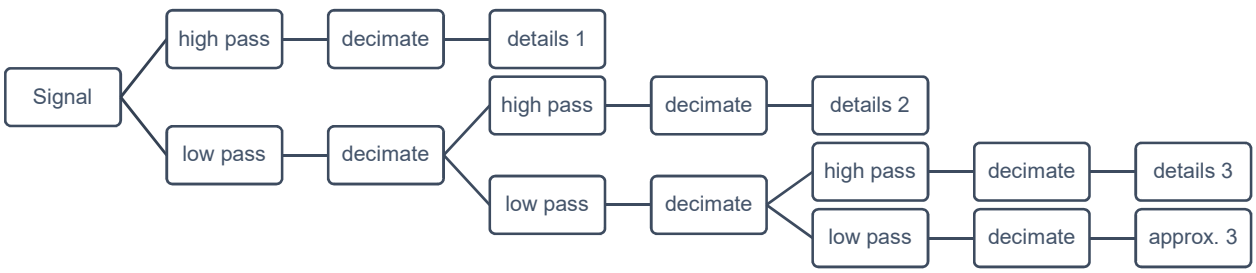
$$\mathrm{CWT} = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}\psi\frac{t\text{-}b}{a}x(t)dt \qquad (1)$$

where a denotes dilation and is thus related to frequency, and b denotes translation is thus related to position. For computational purposes, determination of the inner products in (1) is usually

handled using the discrete wavelet transform (DWT), first proposed by Mallat in the 1980s (1). In the DWT, the dilation (a) and translation (b) parameters have the form:

$$a = 2^{-j}, b = k*2^{-j} \quad (2)$$

forming a discrete grid over which the DWT can be computed for integer j and k. The method determined by Mallat for the DWT is a pyramidal algorithm which decomposes the signal into successive approximation and detail arrays with index k for j decomposition levels (1). A schematic representation of the DWT is shown below:



For each decomposition level, the signal is passed through both a low-pass filter and a high-pass filter. The low-pass filter leaves behind low frequency elements, and thus produces an approximation of the underlying signal. The high-pass filter leaves behind high frequency elements that provide the details of the signal and is initially dominated by noise. The signals are then decimated, and the procedure is repeated starting from the low-pass output / approximation from the previous level. This is repeated for j levels, creating a filter bank where the detail coefficients from the high-pass outputs at each level are retained, along with the approximation coefficients from the final transform level.

These filters are in turn derived from transformations of the parent wavelet. The low-pass filters are derived from shifting and scaling of the mother wavelet (wavelet function). The high-pass filters are derived from a corresponding father wavelet (scaling function). The choice of mother wavelet, and thus on the corresponding filter bank, is dependent on the form of a given signal and is often selected through experimentation. The wavelets used in this work are derivatives of the coiflet and biorthogonal families and are shown below in Figure 1.

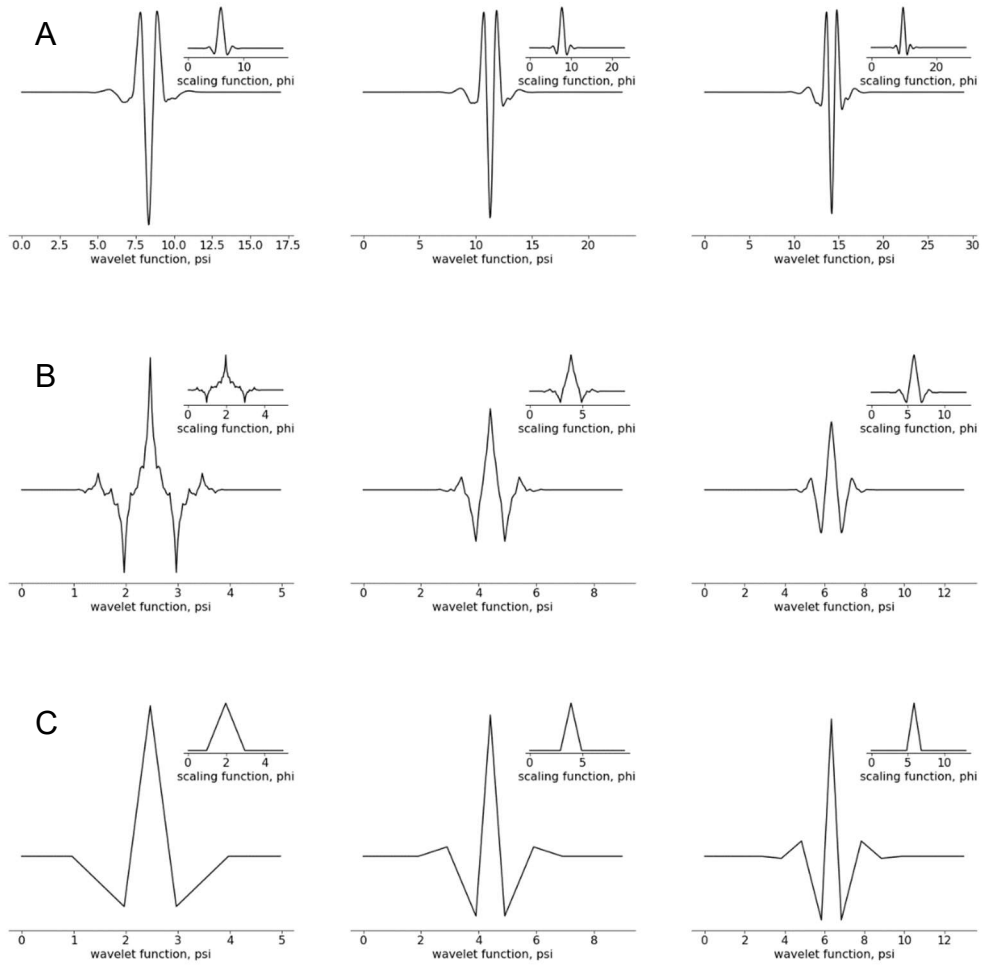


Figure 1. The wavelet and scaling functions for the wavelets used in this work. (A) The coif3, coif4, and coif5 coiflets. The numbers refer to the number of vanishing moments in the scaling functions. (B and C) the bior2.2, bior2.4, and bior 2.6 biorthogonal wavelets. In this case, different wavelet and scaling functions are used for decomposition and reconstruction of the signal. The numbers refer to the number of vanishing moments in the reconstruction and decomposition scaling functions, respectively.

To illustrate the noise versus signal content of successive levels of detail and approximation coefficients, the results of decomposing a test EPR signal with the DWT and the coif3 wavelet to the j=8 level are shown in Figure 2. The detail coefficients are shown in the left panel, and the approximations are shown in the right. Initially, the details are dominated by high frequency noise, with the approximation becoming successively less noisy as j increases.

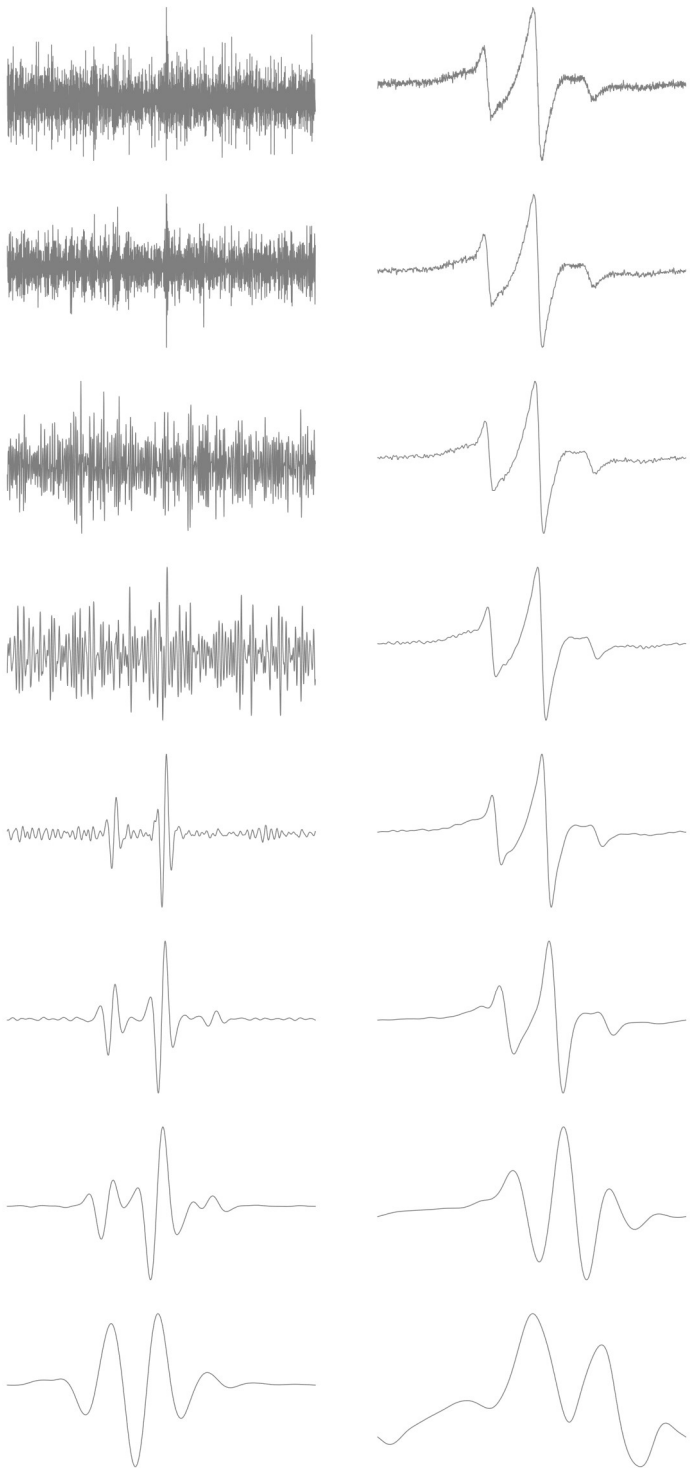


Figure 2. The results of decomposing a test EPR signal to j=8 levels with the DWT and coif3 wavelet. The detail coefficients are shown on the left, and approximation coefficients on the right. The level of decomposition increases going down. Initially, the details are dominated by noise, but by the 5th level begin to contain some elements of the underlying signal. By levels 7 and 8, the signal is split between approximations and details

Transitioning from levels 4 to 5, however, the noise begins to vanish from the details, and they instead begin to represent elements of the signal. By levels 7 and 8, the signal is split between the details and approximation. Manipulation of the detail coefficients, such as to remove noise, would thus be most optimal up to level 5 or 6 for this test signal.

While decomposing the signal using the DWT already affords the ability to see the noise content of a signal, the original can also be reconstructed using an inverse transform. For orthogonal wavelets, including the coiflet family, the same wavelet and scaling function can be used for signal decomposition, and subsequent reconstruction. For biorthogonal wavelets, different wavelets are used in decomposition and reconstruction. The numbers refer to the number of vanishing moments in the scaling function and determine the smoothness of the frequency response of the resulting filter bank.

In the simple case where a signal is decomposed for a given wavelet to a level j and then reconstructed using the resulting detail coefficients and j^{th} approximation coefficients, the signal will retain all of its initial noise. In order to achieve signal denoising, it is necessary to determine which detail coefficients correspond to noise and either remove (kill) or reduce (shrink) them prior to signal reconstruction. Typically, this is handled by the selection of a threshold value. Below the threshold value, coefficients are typically shrunk or killed, while above the threshold they may be shrunk or kept. The threshold can either be selected once and used for all decomposition levels, or a new threshold can be selected at each level. Several methods for applying thresholds are given in section 3.2, and for their selection in section 3.3. When detail coefficients are manipulated in this way, their contributions will be absent in the reconstructed signal, and thus removal of all noise-related coefficients would theoretically produce a noiseless output signal.

While the DWT is a powerful tool for signal analysis, it is not translation-invariant. When used for signal denoising, jump discontinuities in the original signal can lead to oscillatory pseudo-Gibbs artifacts which contaminate the reconstructed signal (2). These artifacts are akin to the Gibbs artifacts in Fourier denoising, although reduced in amplitude (2). Jump discontinuities

occur frequently in CW EPR spectra, where they are concentrated at transitions between mobile and immobile components and in the sharp, central line. To deal with these features, complicated thresholding methods are sometimes developed, which was the case for an earlier attempt at signal denoising of primarily fast-motional CW EPR using the DWT (3).

Alternatively, one of several variants of the DWT can be employed, which introduce varying degrees of translation-invariance. This is realized by shifting the data by one point, repeating the denoising process, and then unshifting. The approach is repeated for a given number of shifts, and the results are averaged. In addition to the time domain, the shift/denoise/unshift approach can also be applied to frequency modulations. Averaging of several shifted transforms reduces or removes oscillations at the jump discontinuities and smooths the reconstructed signal, but at the cost of increased computational complexity. Balancing these two factors is achieved by picking DWT variants with a given degree of signal shifting. The simplest is the dual-tree complex wavelet transform (DTCWT), which applies the DWT twice in a dual-tree filter bank (4). If that is insufficient, the use of about 10 to 20 shifts falls under the cycle-spinning designation (2). Finally, the computation of all possible shifts results in a fully translation-invariant transform termed the stationary wavelet transform (SWT) (2). The SWT has a large hit in computation time relative to the DWT, although it has been observed to permit the use of simple thresholding functions and selection algorithms (2). The present work considers only the SWT.

3.2 Wavelet Thresholds

A variety of estimators have been proposed for use as thresholds for the detail coefficients of wavelet transformed signals. The initial, universal threshold estimator is dependent only on the length of the data n, having no dependence on the form of the data. It has the form:

$$t_{\mathrm{universal}} = \sqrt{2\log(n)} \quad (3)$$

The universal threshold is a conservative estimator, which for most data sizes suppresses greater than 99% of all noise related coefficients (5). This comes at the cost of potentially suppressing signal coefficients, particularly in sharply varying regions such as narrow peaks. Due to this drawback, a variety of softer threshold selection methods have been developed. This work tests the heuristic Stein's unbiased risk estimate (heursure or heurSURE) and MINIMAX thresholds.

The heursure method combines the universal threshold with a second option derived from SURE (6). This allows the threshold to be softer in regions of high SNR, and more conservative in low SNR regions. The SURE threshold, t_{SURE}, is selected from a set of thresholds, t, by:

$$t_{SURE} = \min(SURE(t;x) = \min(n\text{-}2\#\{i: |x_i| \leq t\} + \sum_{i=1}^{n}(|x_i| \wedge t)^2) \quad (4)$$

Where n is the length of the data, # is the cardinality operator, and ∧ gives the minimum of two elements. Then, the criteria for threshold selection is given by the following:

$$t = \left\{ \begin{matrix} t_{universal} & A < B \\ \min(t_{universal}, t_{SURE} & A \geq B \end{matrix} \right\} \quad (5)$$

Where $A = \frac{(s\text{-}n)}{n}$ with s equal to the sum of squares of the detail coefficients and n to the length of the detail coefficients (6). The value of B is given by $(\log_2(n))^{\frac{3}{2}}\sqrt{n}$ with n again being the length of the detail coefficients. As A is dependent on the squares of the coefficients, it is small for levels that contain primarily noise with low SNR. Thus, low SNR conditions use the more conservative universal estimator. SURE and heurSURE thresholds are recomputed at each decomposition level.

Finally, several precomputed thresholds are used that are designed to adhere to MINIMAX risk minimization principles (7). These are typically provided in look-up tables as a function of the length of the data vector, but the simple case of a single threshold is often approximated as:

$$t_{minimax} = 0.3936 + 0.1829\log_2(n) \quad (6)$$

Where n is the length of the data vector, and $t_{minimax}$ is approximately 2.4 for the signals used here with length n=2048.

3.3 Wavelet Thresholding Methods

The simplest way to remove unwanted detail coefficients in the wavelet transformed signal is through a keep or kill rule, where coefficients below the threshold value are set equal to 0, and above the threshold value are left alone. This scheme is referred to as hard thresholding:

$$d_{jk} = \begin{cases} 0 & \text{if } |d_{jk}| < t \\ d_{jk} & \text{if } |d_{jk}| \geq t \end{cases} \qquad (7)$$

where d_{jk} is the detail coefficient at position k and decomposition level j, and t is the threshold value. Here, hard thresholding is used with the heuristic SURE, MINIMAX, and universal estimator derived threshold values. Alternatively, a shrink or kill rule can be used, where the wavelet coefficients are set to 0 below the threshold and shrunk by the threshold value above the threshold. This is termed soft thresholding, and has the following form:

$$d_{jk} = \begin{cases} 0 & \text{if } d_{jk} \leq t \\ d_{jk}\text{-}t & \text{if } d_{jk} > t \\ d_{jk} + t & \text{if } d_{jk} < \text{-}t \end{cases} \qquad (8)$$

Soft thresholding was used only with the universal estimator for the threshold value. The choice between soft and hard thresholding is a balance between bias and variance in the denoised signal. As hard thresholding keeps values above the threshold, it tends to have better performance in signal peaks at the cost of greater variance in response to large fluctuations in the data. These fluctuations are suppressed using soft thresholding, but at the cost of reduced amplitude in the signal peaks.

A variety of methods have been created in an attempt to balance the features of both soft and hard thresholding. Two of these methods, non-negative garrote thresholding and firm thresholding, are presented here. The non-negative garrote thresholding is similar to soft thresholding, with the exception that the shrinkage decreases with the size of the coefficient (8). This is intended to increase its performance in sharp data peaks. Its form is:

$$d_{jk} = \begin{cases} 0 & \text{if } |d_{jk}| \leq t \\ d_{jk}\text{-}\frac{t^2}{d_{jk}} & \text{if } |d_{jk}| > t \end{cases} \qquad (9)$$

This method was used only with the universal estimator for the threshold value. Alternatively, firm shrinkage introduces a 3-part keep, shrink, or kill rule with 2 thresholds (9). It has the form:

$$d_{jk} = \begin{cases} 0 & \text{if } |d_{jk}| \leq t_1 \\ \text{sgn}(d_{jk}) \frac{t_2(|d_{jk}|-t_1)}{t_2-t_1} & \text{if } t_1 < |d_{jk}| \leq t_2 \\ d_{jk} & \text{if } |d_{jk}| < -t_2 \end{cases} \quad (10)$$

Where t_1 and t_2 are now the lower and upper threshold values, respectively. Here, the values of t_1 and t_2 were taken from a literature table for the MINIMAX condition, and had values of 2.737 and 6.939, respectively for n=2048 data points (9).

Finally, as the threshold values presented in section 3.2 are often dependent only on the size of the data, it is necessary to scale them when applying a thresholding method. This requires an estimate of the standard deviation of the noise, which can be determined in several ways. The most common in wavelet thresholding is to use a median-absolute-deviation (MAD) estimator with a scaling factor for the normal distribution of white-noise values (5). It has form:

$$\sigma = \frac{\text{MAD}}{0.6745} \quad (11)$$

where σ can be calculated level independently using just the first level detail coefficients that contain the most noise, or level dependently at each of the j decomposition levels. Here, σ was calculated level independently for each method. It was then multiplied by the threshold values prior to scale them prior to thresholding.

3.4 Description of Method

A total of six test signals were used for this analysis. The four original signals (blocks, bumps, heavisine, and doppler) were generated using the pywavelets python package, which in turn uses the code available in the proGram wavelab (10). The remaining EPR test signals were generated using the multicomponent software by Christian Altenbach, UCLA. The EPR test signals do not represent physical spectra but were selected to mimic multicomponent ensembles with smoothly varying elements that are difficult denoising targets. All analysis used the test

signals at n=2048 data points. The original signals were generated with n=2048 points, but the output of multicomponent has n=512 data points. Thus, the EPR test signals had to be upscaled to n=2048 points, which was accomplished through cubic spline interpolation. The test signals are shown in Figure 3.

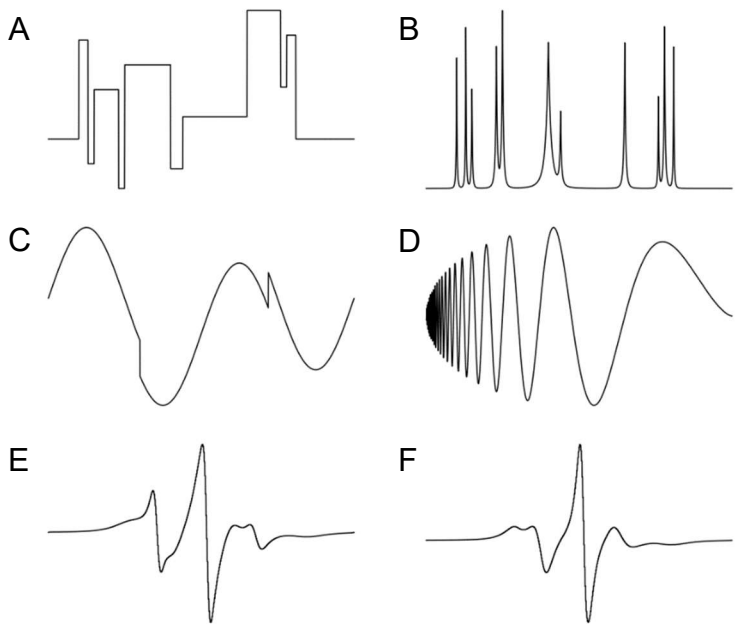


Figure 3. Test Signals for Denoising. The signals are (A) blocks, (B) bumps, (C) heavisine, (D) doppler, (E) 1st EPR signal, and (F) 2nd EPR signal. Signals A-D were generated using pywavelets. Signals E and F were generated using Multicomponent.

To generate noisy signals at varying SNR ratios, the initial noise-free test signals were combined with normally distributed arrays of n=2048 points generated using the numpy software package random.normal function. Noise arrays had a mean of 0, and the variance was determined as:

$$\text{variance} = \text{abs}\left(\frac{\sum_1^{\text{n}} \text{signal}^2}{\text{SNR}_{\text{target}} * \text{n}} \right) \quad (12)$$

where signal is the magnitude of the test signal at point n, and the target SNR is the desired signal-to-noise ratio. The resulting noise arrays were added to the initial data to generate noisy traces for each SNR. This was repeated for 10 trials, resulting in 40 total noisy signals for each of the six test signals. The resulting noisy signals were used for all subsequent wavelet denoising and are shown in Figure 4 for the SNR=200 and 10 cases

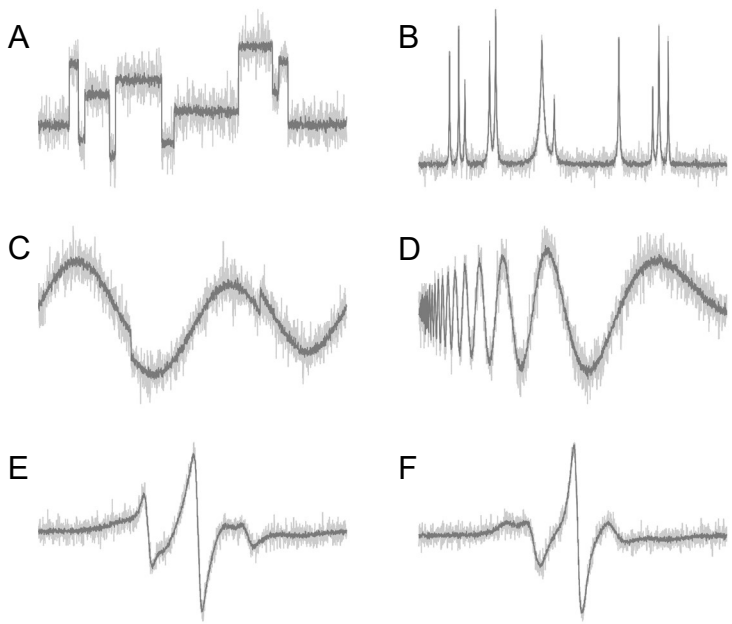


Figure 4. Noisy Test Signals for Denoising. The signals are (A) blocks, (B) bumps, (C) heavisine, (D) doppler, (E) 1st EPR signal, and (F) 2nd EPR signal. The lighter gray shows the noise present at SNR=10, while the darker gray shows the noise present at SNR=200. Each spectrum is a representative trace of 10 noisy spectrum generated for each signal and SNR level.

The actual average SNRs of these noisy signals are given in Table 1. There is some deviation from the target SNRs of 200, 100, 50, and 10, but the overall agreement with the targets is good. When performance metrics are presented, they represent the average of the 10 traces for a given combination of parameters.

	Signal					
Target SNR	**Bumps**	**Blocks**	**Heavisine**	**Doppler**	**EPR 1**	**EPR 2**
200	201.034	200.790	199.879	201.013	202.536	204.160
100	100.713	102.649	101.062	101.807	101.446	100.429
50	50.474	51.045	50.983	51.013	51.439	51.884
10	11.167	10.998	11.137	10.877	10.947	10.808

Table 1. The actual SNR values for the noisy test signals at target SNRs of 200, 100, 50, and 10. Each value is the average of 10 trials adding white Gaussian noise to the same initial test signal.

The general method for wavelet denoising of CW EPR spectra in the slow-motional regime involved the use of the SWT with a particular wavelet, level of decomposition, threshold value, and thresholding method. Signal deconstruction and reconstruction using the SWT were performed using the python pywavelets package. All data presented here is the top subset of a much larger parameter search. Three coiflets (coif3, coif4, coif5) and three biorthogonal wavelets

(bior2.2, bior2.4, bior2.6) are featured in this analysis. Decomposition levels of 5 and 6 were determined to be most useful based on decomposition of the 1st EPR test signal in Fig. 4. The methods used for threshold determination were the universal estimator, the heuristic SURE estimator, and MINIMAX determined estimators. Thresholding methods employed were hard, soft, non-negative garrote, and firm thresholding. Threshold determination and thresholding was handled in python scripts that made use of the numpy and scipy python packages. Plotting used the matplotlib software library.

3.5 Performance of Method

To evaluate the performance of the SWT with a given combination of transform parameters, the SNR, structural similarity index (SSIM, and root-mean-square error (RMSE) were determined for each combination. The SNR in this case is defined as follows:

$$\mathrm{SNR} = \frac{\sum_{1}^{n}(x_{denoised}(n))^2}{\sum_{i}^{n}(x_{denoised}(n)\text{-}x(n))^2} \quad (13)$$

where $x_{denoised}(n)$ is the wavelet transformed signal, and x(n) is the test signal. The SNR and RMSE were computed using functions written in python. The SSIM was computed using a built-in function in the scikit-image module of scipy. The resulting performance metrics are shown in supplementary data tables 1-6. They are organized by wavelet, starting with coif3, and by increasing SNR. The data were first assessed to determine the optimal wavelet for each test signal. For the four abstract signals (blocks, bumps, heavisine, doppler), this involved finding the wavelet corresponding to the combination with the best SNR/SSIM/RMSE. Fortunately, the three parameters show the same trend at higher SNRs, and so the optimal wavelet for each signal was determined based on SNR=200. The chosen wavelets were bior2.4 for bumps and blocks, coif3 for heavisine, and coif5 for doppler. For the EPR signals, it would be more ideal to use a single wavelet for all future data, and so the wavelet was selected that gave the best average performance across both spectra. The selected wavelet for the EPR test signals was coif3.

Next, the combinations of threshold selection and thresholding methods were visually analyzed for each signal and the best performing wavelet. These results are shown for decomposition level j=5 in Figures 5-10 on the following pages, and for decomposition level j=6 in Figures S1-6. The figures are arranged vertically in order of increasing SNR moving down, from SNR=10 to SNR=200. The left side shows performance across the whole spectra, while the right side shows performance of a single feature so that the better performing methods can be visually separated. The methods are color coded with the following combinations: hard thresholding with the universal estimator is blue, soft thresholding with the universal estimator is green, non-negative garrote thresholding with minimax thresholds is red, hard thresholding with the heuristic SURE thresholds is orange, hard thresholding with a minimax threshold is brown, and firm thresholding with the universal estimator is purple.

For the blocks signal (Figure 5), the best performance at the single feature level is obtained for the hard+MINIMAX combination, although this comes at the expense of additional oscillations in the flat portions of the spectrum. The next best feature performance is for the hard+universal combination, which has markedly reduced oscillation in the flat portions, and probably represents the best balance of performance. The results are similar for the Bumps signal (Figure 6), although now the firm+universal and hard+universal combinations are very similarly performant. The hard+heurSURE combination falls behind on the bumps signal, strongly underestimating peak heights compared to the other methods. The results for bumps are almost identical to those observed for the heavisine and doppler signals (Figures 7 and 8). In both bases hard+MINIMAX gives the best feature performance at the cost of not fully denoising the signal. Again, the next best performance comes from the firm+universal and hard+universal combinations, with hard+heurSURE being the least performant.

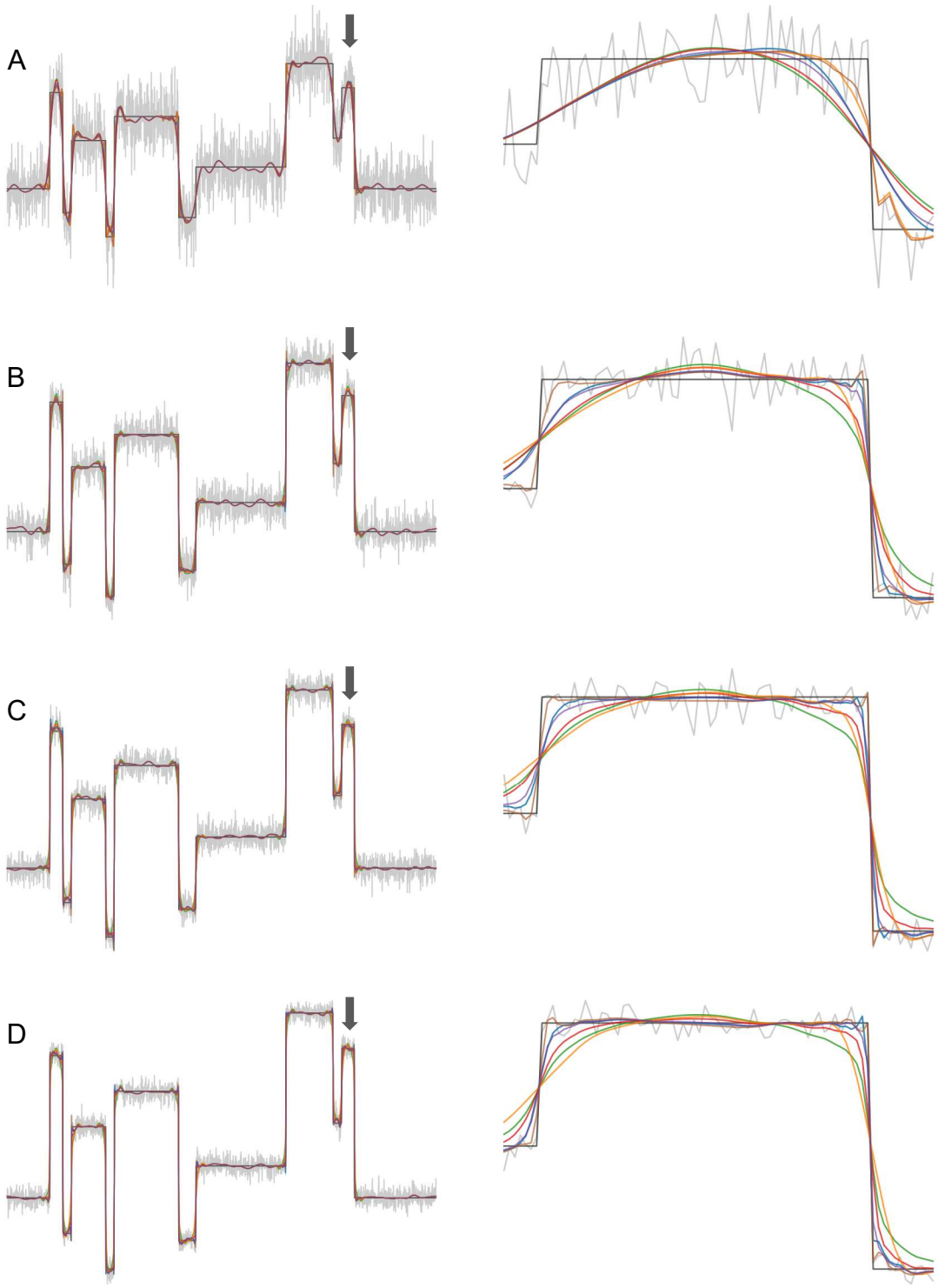


Figure 5. Performance of the SWT at level 5 across 6 different methods for the blocks test signal at SNR of 10 (A), 50 (B), 100 (C) or 200 (D). Methods are color coded as follows (threshold selection + method): universal+hard, universal+soft, universal+garrote, heuristic SURE+hard, MINIMAX+hard, universal+firm. The left panel shows whole trace performance while the right shows a zoomed region indicated by a black arrow. At low SNR, the best performance at feature boundaries is obtained with MINIMAX+hard, but at the cost of a high degree of spurious noise in flat regions. By the SNR 100 trace, most methods catch up, with the universal+hard combination now having the best average performance across the signal.

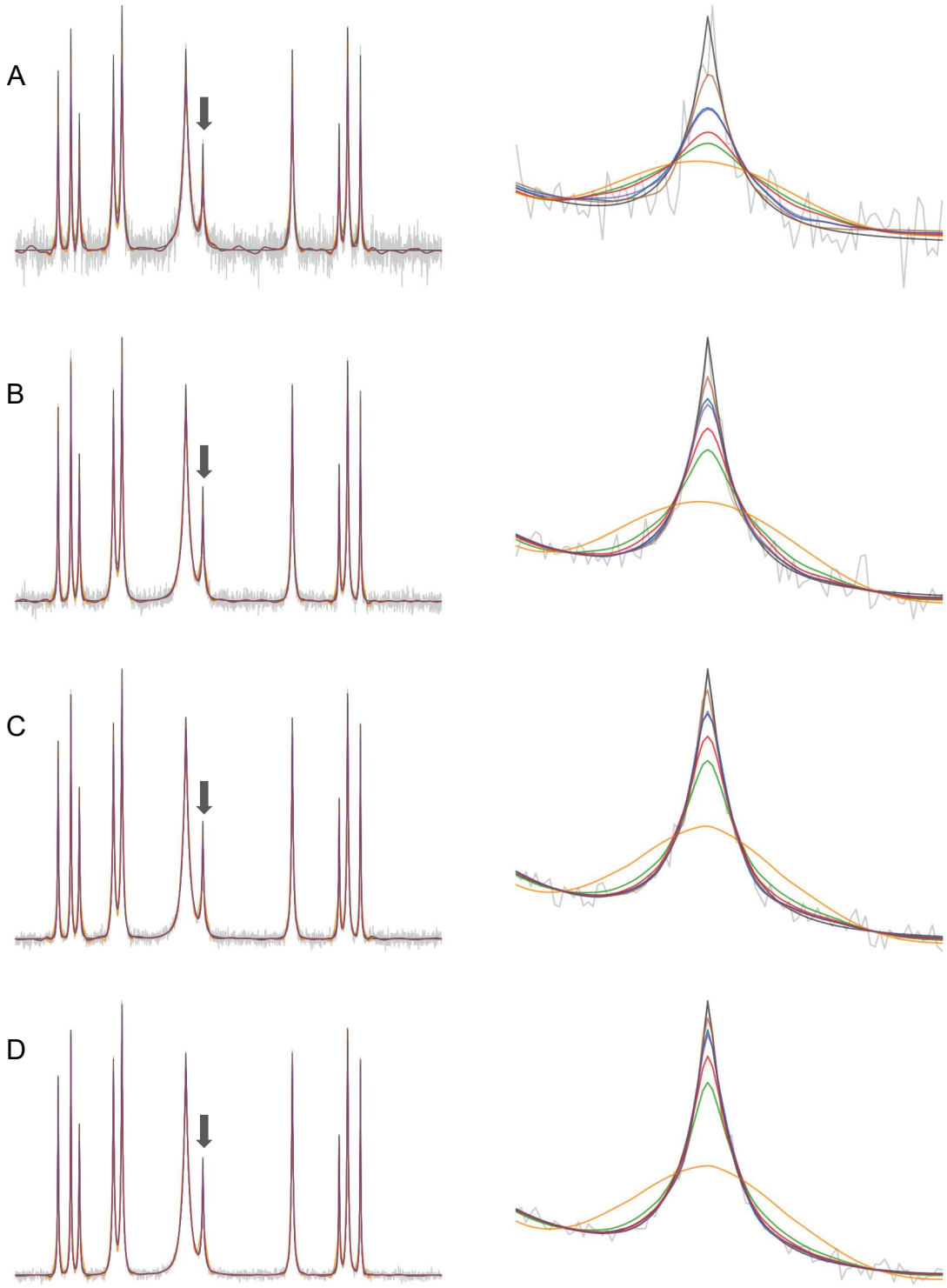


Figure 6. Performance of the SWT at level 5 across 6 different methods for the bumps test signal at SNR of 10 (A), 50 (B), 100 (C) or 200 (D). Methods are color coded as follows (threshold selection + method): universal+hard, universal+soft, universal+garrote, heuristic SURE+hard, MINIMAX+hard, universal+firm. The left panel shows whole trace performance while the right shows a zoomed region indicated by a black arrow. At low SNR, the best performance at feature boundaries is obtained with MINIMAX+hard, but at the cost of a high degree of spurious noise in flat regions. By the SNR 100 trace, most methods catch up, with the universal+hard and universal+firm combinations now having the best average performance across the signal.

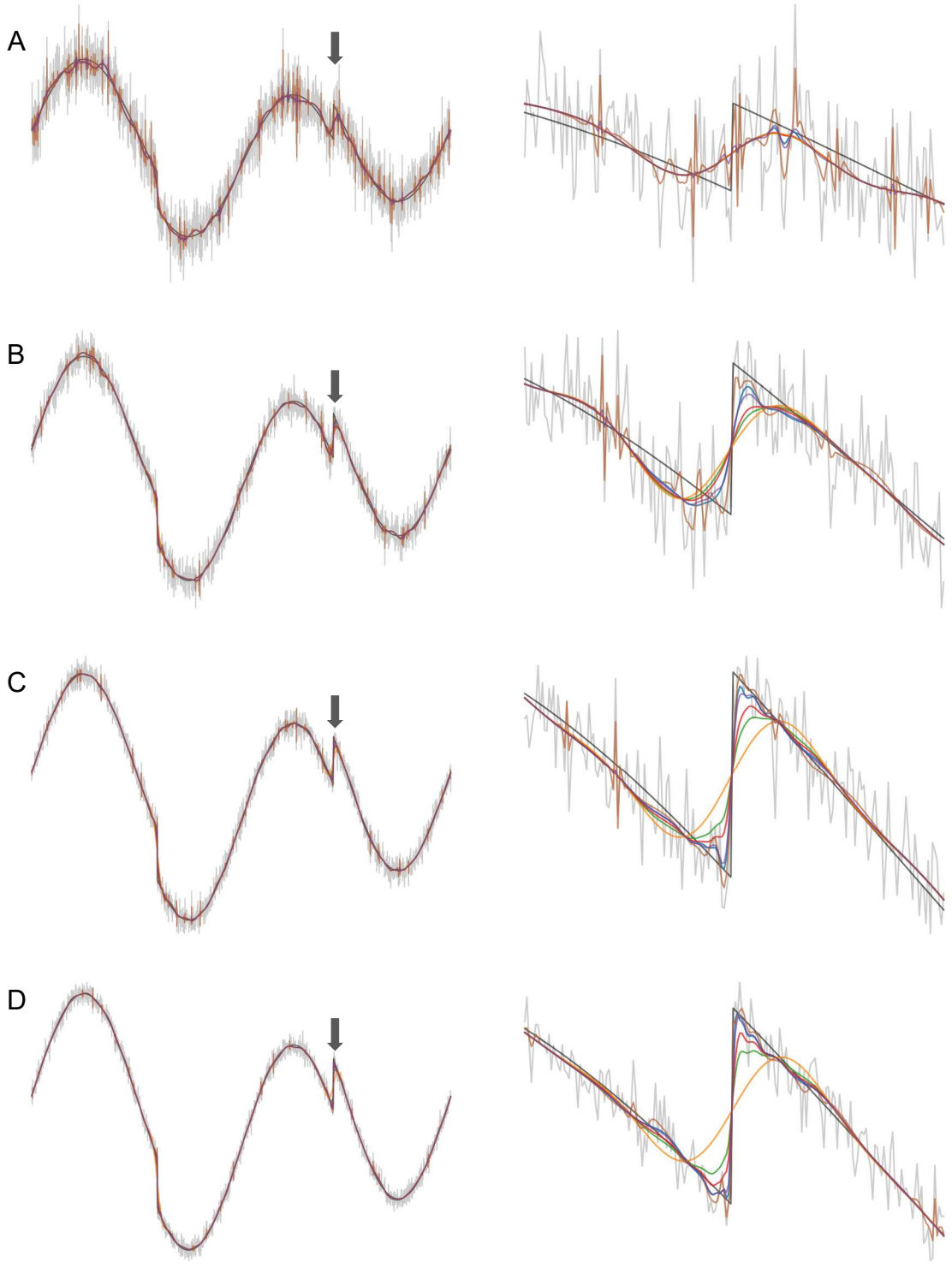


Figure 7. Performance of the SWT at level 5 across 6 different methods for the heavisine test signal at SNR of 10 (A), 50 (B), 100 (C) or 200 (D). Methods are color coded as follows (threshold selection + method): universal+hard, universal+soft, universal+garrote, heuristic SURE+hard, MINIMAX+hard, universal+firm. The left panel shows whole trace performance while the right shows a zoomed region indicated by a black arrow. At low SNR, the best performance at feature boundaries is obtained with MINIMAX+hard, but at the cost of a high degree of spurious noise in flat regions. By the SNR 100 trace, most methods catch up, with the universal+hard and universal+firm combinations now having the best average performance across the signal.

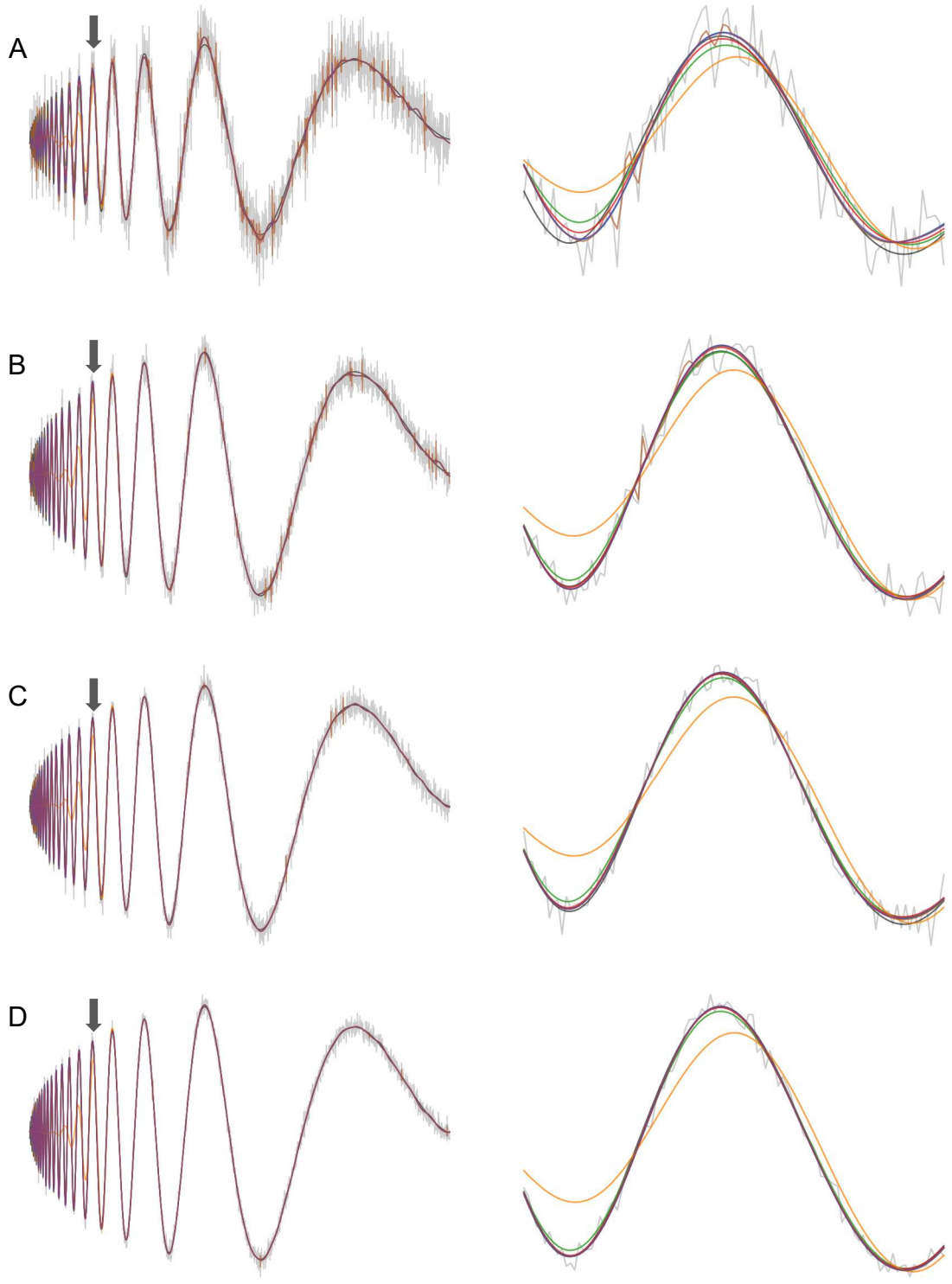


Figure 8. Performance of the SWT at level 5 across 6 different methods for the doppler test signal at SNR of 10 (A), 50 (B), 100 (C) or 200 (D). Methods are color coded as follows (threshold selection + method): universal+hard, universal+soft, universal+garrote, heuristic SURE+hard, MINIMAX+hard, universal+firm. The left panel shows whole trace performance while the right shows a zoomed region indicated by a black arrow. At low SNR, the best performance at feature boundaries is obtained with MINIMAX+hard, but at the cost of a high degree of spurious noise in flat regions. By the SNR 100 trace, most methods catch up, with the universal+hard and universal+firm combinations now having the best average performance across the signal.

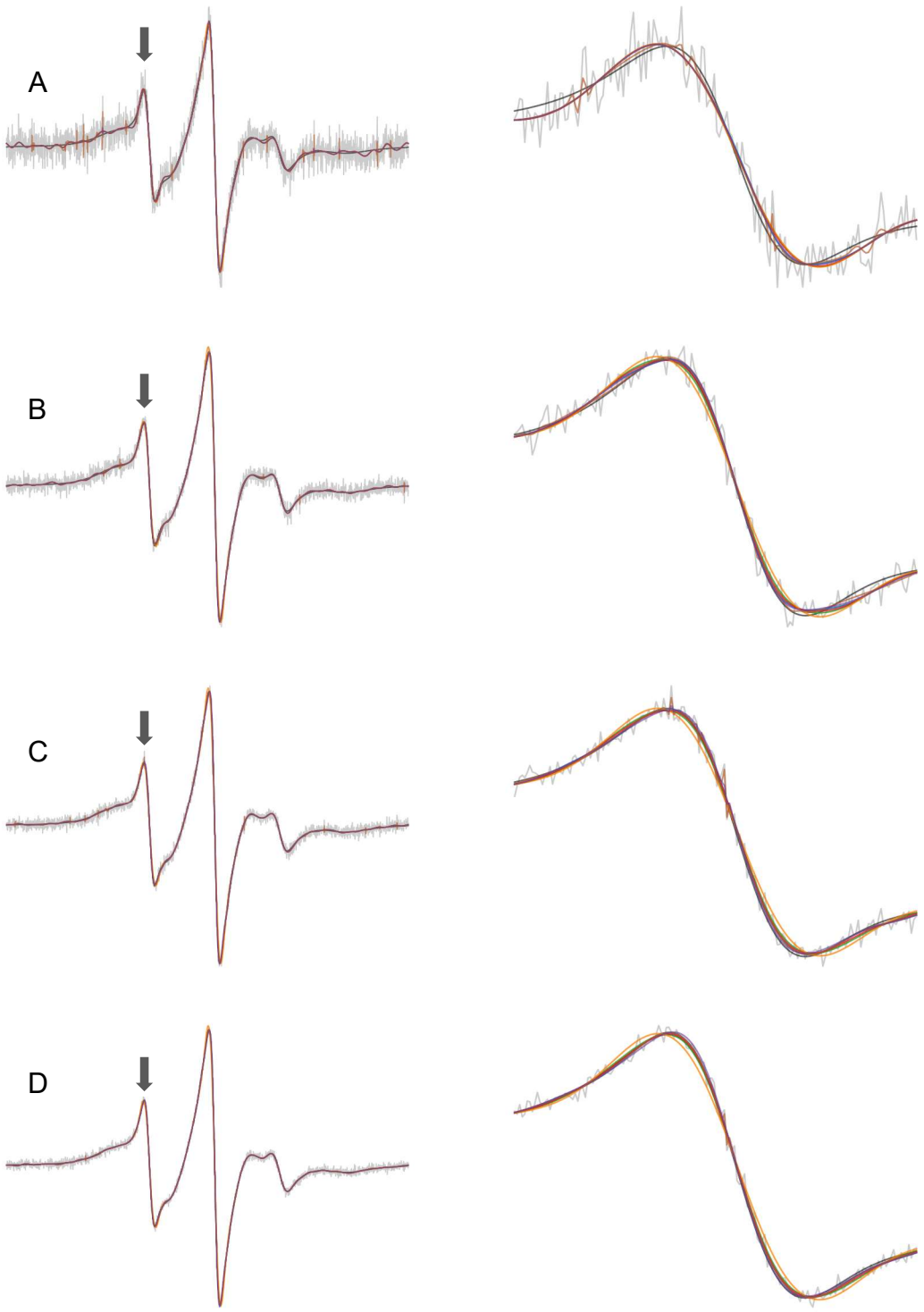


Figure 9. Performance of the SWT at level 5 across 6 different methods for the 1st EPR test signal at SNR of 10 (A), 50 (B), 100 (C) or 200 (D). Methods are color coded as follows (threshold selection + method): universal+hard, universal+soft, universal+garrote, heuristic SURE+hard, MINIMAX+hard, universal+firm. The left panel shows whole trace performance while the right shows a zoomed region indicated by a black arrow. . Across all SNR values, the universal+hard combination displays the best signal performance, lacking the noisy oscillations still present in the MINIMAX+hard and universal+firm conditions.

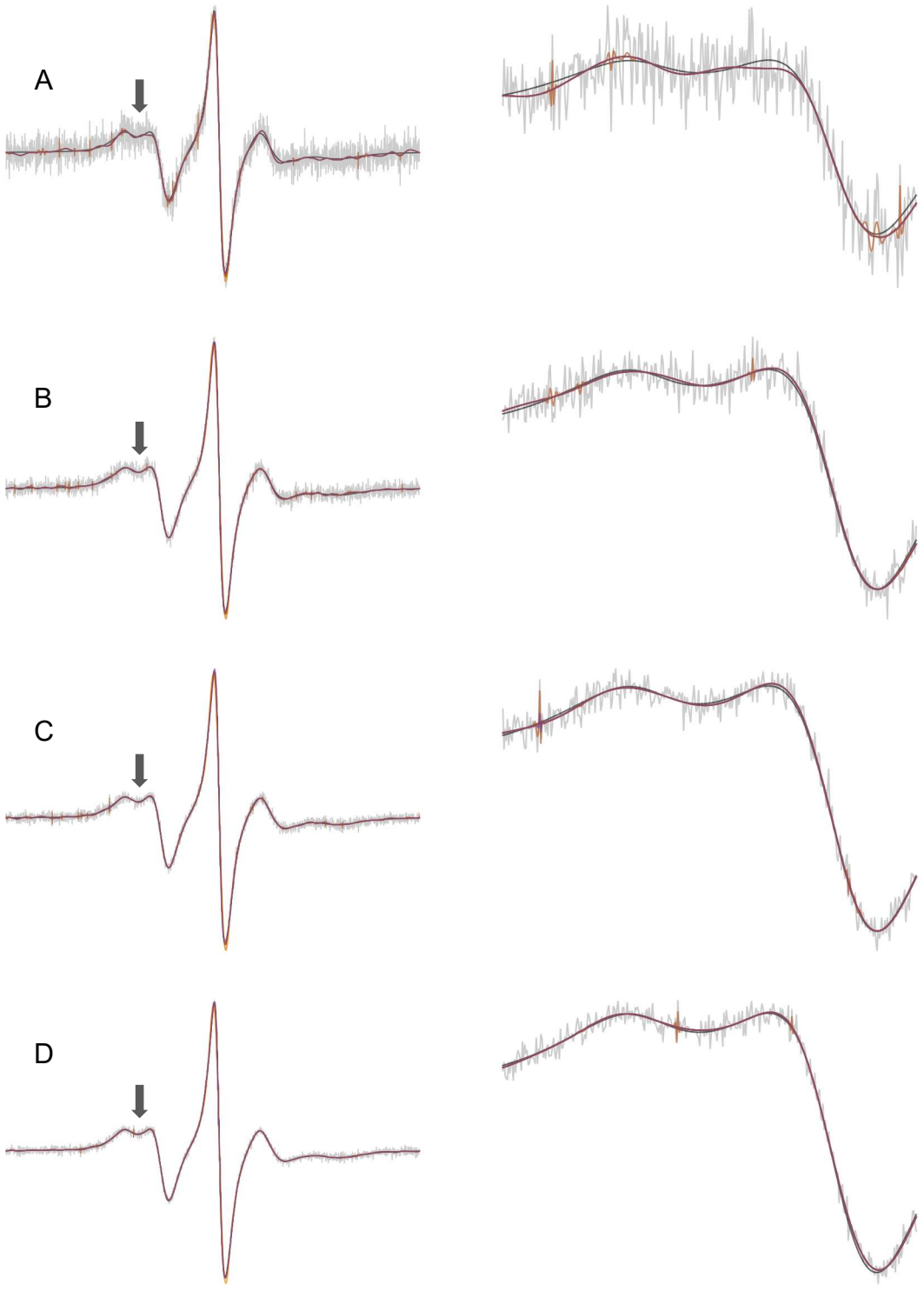


Figure 10. Performance of the SWT at level 5 across 6 different methods for the 2nd EPR test signal at SNR of 10 (A), 50 (B), 100 (C) or 200 (D). Methods are color coded as follows (threshold selection + method): universal+hard, universal+soft, universal+garrote, heuristic SURE+hard, MINIMAX+hard, universal+firm. The left panel shows whole trace performance while the right shows a zoomed region indicated by a black arrow. . Across all SNR values, the universal+hard combination displays the best signal performance, lacking the noisy oscillations still present in the MINIMAX+hard and universal+firm conditions.

The doppler signal also provides useful validation of all of the methods except for hard+heurSURE. This test signal increases in frequency from right to left. The zoomed region of Figure 8 shows the 5th oscillation from the right, which is higher in frequency and thus narrower than that expected for a free spin component in a typical CW EPR spectrum. For this oscillation, all methods except for hard+heurSURE perform well at capturing the signal features, and so it is unlikely that any tested EPR signals will exceed the frequency response of the SWT methods investigated here.

Moving to the two EPR test signals (Figures 9 and 10), it is apparent that all methods do well at decomposition level j=5 across all tested SNR values. By SNR=100, all except the hard+heurSURE combination produce an almost complete reconstruction of the test signal. The hard+MINIMAX, firm+universal, and hard+universal combinations get slightly closer to the reference signal, but the former two combinations display some residual noise left in the trace. This is true across both EPR test signals, and so the universal+hard combination appears to be the best for use with unknown CW EPR data.

Next, to confirm the conclusion from the visual comparisons and assess which level of decomposition should be selected, the three performance metrics were examined as a function of signal SNR for the EPR test signals and the coif3 wavelet. The results are shown in Figures 11 and 12. Generally, the performance at level 5 is slightly better than at level 6 when measured numerically, although the level 6 results better smooth the baseline, and so the level may be chosen on a case-by-case basis (see Figs S1-5). For both EPR test signals, the greatest performance is obtained for the hard+MINIMAX condition, although as noted above this combination was too soft and did not fully denoise the spectrum. The next best performance, as in the visual comparisons, was given by the firm+universal and hard+universal combinations. Again, the firm+universal combination retained a small degree of noise, and so the hard+universal combination holds as the best combination of numeric performance and full denoising.

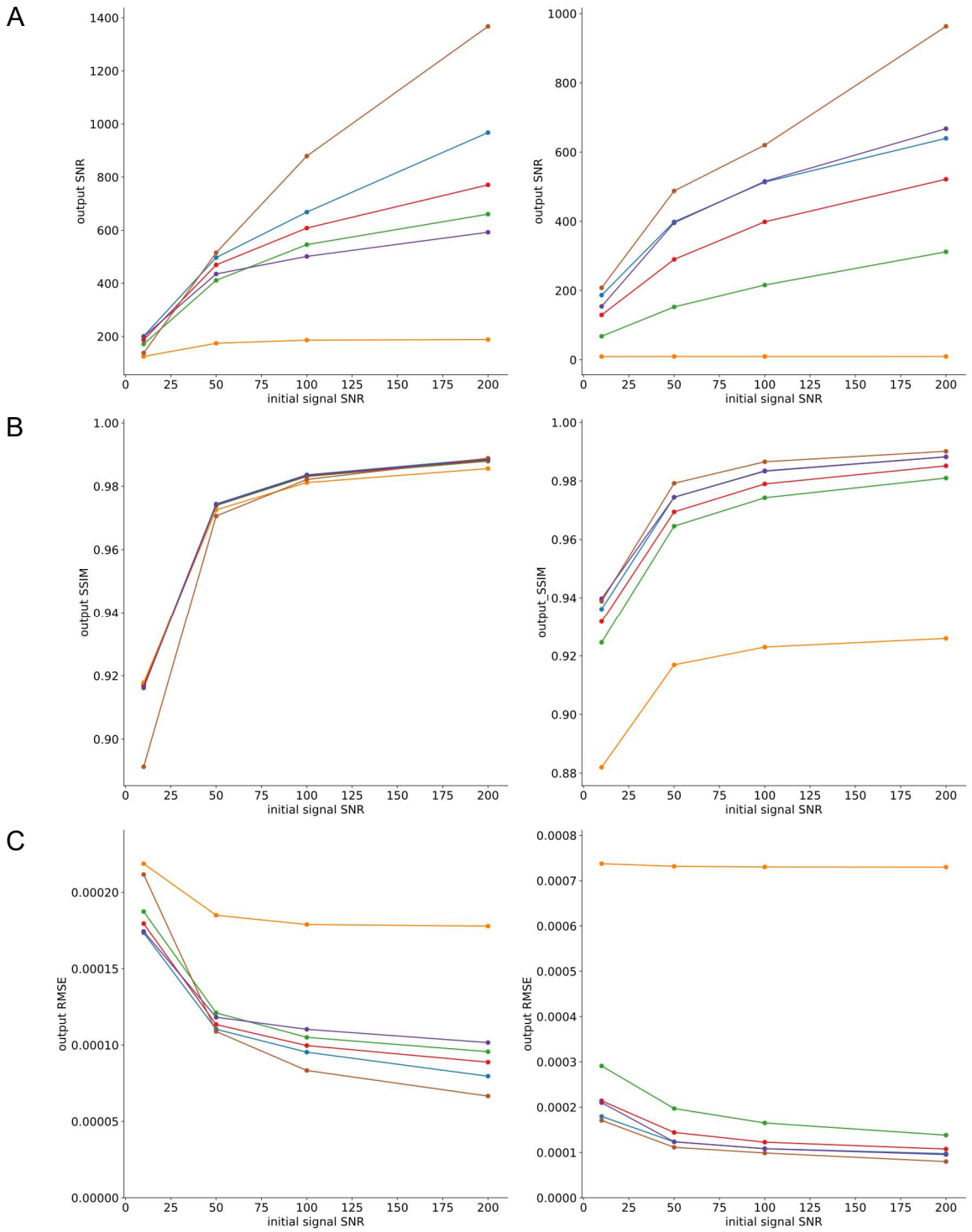


Figure 11. Performance of the SWT of the 1st EPR test signal at level 5 (left) and 6 (right) as a function of SNR measured by (A) SNR, (B) SSIM, and (C) RMSE. Methods are color coded as follows (threshold selection + method): universal+hard, universal+soft, universal+garrote, heuristic SURE+hard, MINIMAX+hard, universal+firm. Overall, the level 5 results are better numerically than those obtained for level 6. The performance of the heuristic SURE+hard combination is particularly poor when measured in terms of SNR and RMSE, with the remaining methods clustering for the low SNR conditions (10 and 50). The MINIMAX+firm condition performs disproportionally well in the SNR metric at high starting SNR, but as noted in the visual comparison figures, this comes at the expense of incomplete signal denoising. This was also the case for the universal+firm threshold. The next best performing combination in all metrics was universal+hard, which was also observed visually.

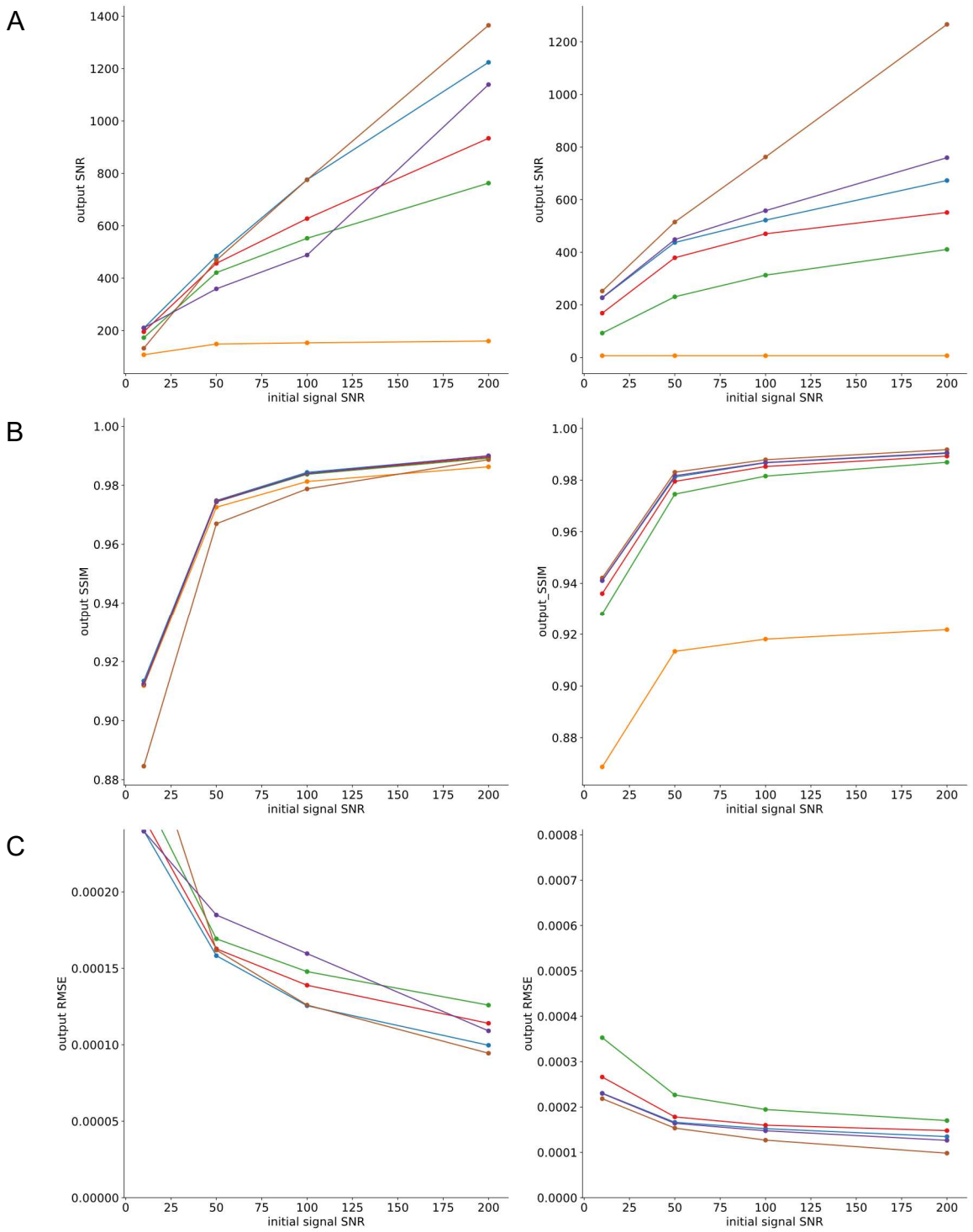


Figure 12. Performance of the SWT of the 2nd EPR test signal at level 5 (left) and 6 (right) as a function of SNR measured by (A) SNR, (B) SSIM, and (C) RMSE. Methods are color coded as follows (threshold selection + method): universal+hard, universal+soft, universal+garrote, heuristic SURE+hard, MINIMAX+hard, universal+firm. Overall, the level 5 results are better numerically than those obtained for level 6. The performance of the heuristic SURE+hard combination is particularly poor when measured in terms of SNR and RMSE, with the remaining methods clustering for the low SNR conditions (10 and 50). The MINIMAX+firm condition performs disproportionally well in the SNR metric at high starting SNR, but as noted in the visual comparison figures, this comes at the expense of incomplete signal denoising. This was also the case for the universal+firm threshold. The next best performing combination in all metrics was universal+hard, which was also observed visually.

3.6 Conclusion

Overall, the results presented in this chapter demonstrate that wavelet threshold denoising is applicable to test EPR signals that represent MTSL spin labels in the slow-motion regime. The combination of hard thresholding with a universal estimator of the threshold value was found to best balance performance in terms of signal recovery with full removal of noise coefficients in the transform. Decomposition to level 5 was found to behave the best numerically for test signals with n=2048 points, but level 6 was quite close in performance and may be beneficial when more smoothing of the baseline is desired. Additionally, the coif3 wavelet was found to give very good performance for both tested EPR signals and is likely the best choice for general analysis of EPR signals of this type. This wavelet was also determined to be the optimal choice in an earlier analysis of EPR spectra in the fast-motional regime using the DWT and a custom thresholding function (3). For the remaining chapters of this work, the noisy spectra are presented as an overlay with a denoised signal representing the SWT transform with the coif3 wavelet at level 5 or 6 with hard thresholding and the universal estimator.

3.7 References

1. Mallat, S. G. 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Transactions on Pattern Analysis and Machine Intelligence 11(7):674-693.
2. Coifman, R. R., and D. L. Donoho. 1995. Translation-Invariant De-Noising.125-150.
3. Srivastava, M., C. L. Anderson, and J. H. Freed. 2016. A New Wavelet Denoising Method for Selecting Decomposition Levels and Noise Thresholds. IEEE Access 4:3862-3877.
4. Selesnick, I. W., R. G. Baraniuk, and N. C. Kingsbury. 2005. The dual-tree complex wavelet transform. IEEE Signal Processing Magazine 22(6):123-151.
5. Johnstone, I. M., and B. W. Silverman. 1997. Wavelet Threshold Estimators for Data with Correlated Noise. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 59(2):319-351.
6. Donoho, D. L., and I. M. Johnstone. 1995. Adapting to Unknown Smoothness via Wavelet Shrinkage. Journal of the American Statistical Association 90(432):1200-1224.
7. Donoho, D., and I. M. Johnstone. 1998. Minimax estimation via wavelet Shrinkage.
8. Gao, H.-Y. 1998. Wavelet Shrinkage Denoising Using the Non-Negative Garrote. Journal of Computational and Graphical Statistics 7(4):469-488.
9. Gao, H.-Y., and A. G Bruce. 1997. WaveShrink with firm shrinkage.
10. Buckheit, J. B., and D. L. Donoho. 1995. WaveLab and Reproducible Research. Wavelets and Statistics. A. Antoniadis and G. Oppenheim, editors. Springer New York, New York, NY, pp. 55-81.

3.8 Supplemental Information

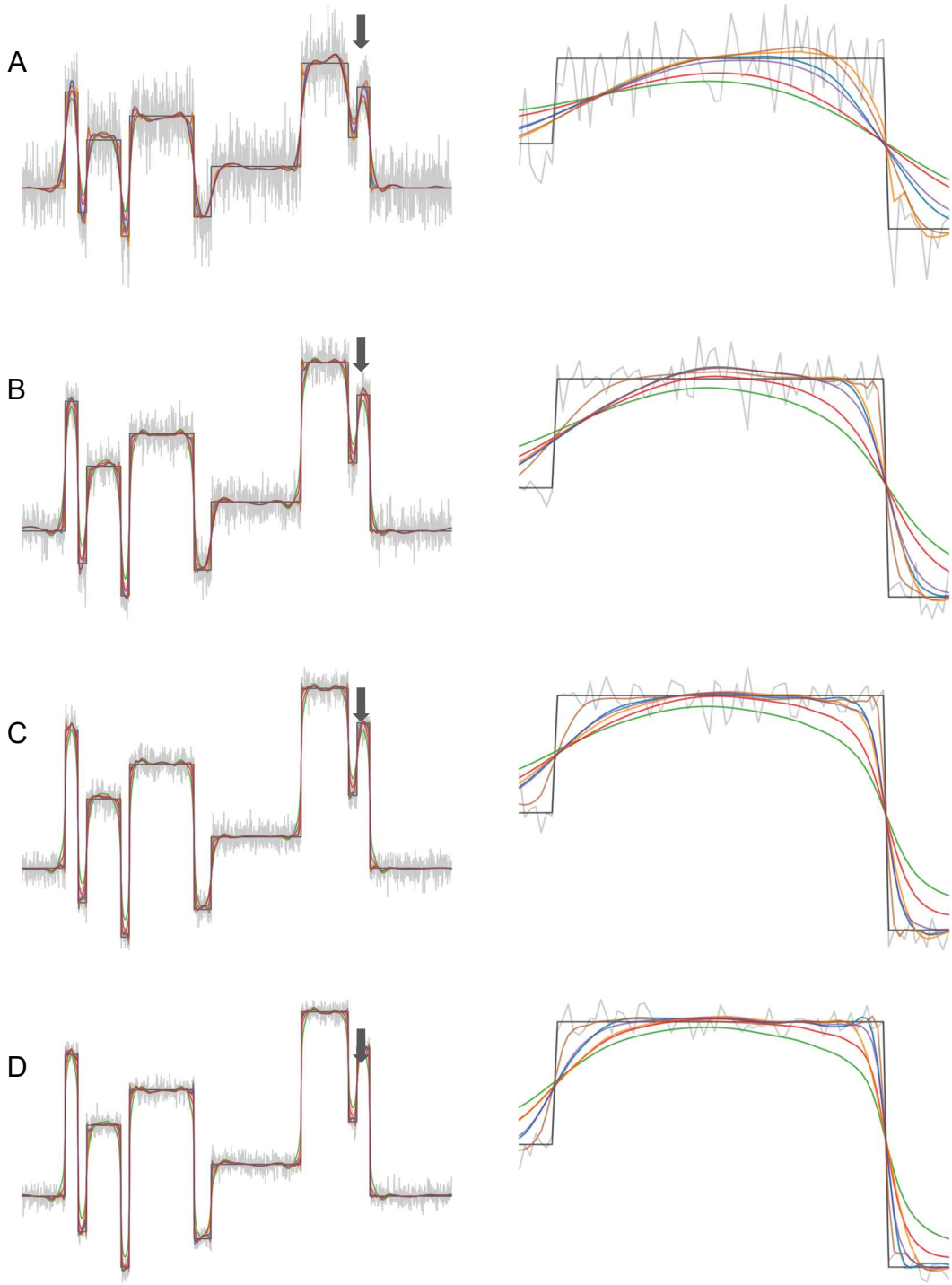


Figure S1. Performance of the SWT at level 6 across 6 different methods for the blocks test signal at SNR of 10 (A), 50 (B), 100 (C) or 200 (D). Methods are color coded as follows (threshold selection + method): universal+hard, universal+soft, universal+garotte, heuristic SURE+hard, MINIMAX+hard, universal+firm. The left panel shows whole trace performance while the right shows a zoomed region indicated by a black arrow. At low SNR, the best performance at feature boundaries is still obtained with MINIMAX+hard at the expense of incomplete denoising, and universal+hard still provides the best balance. Compared to level 5, the level 6 result stays flatter in broad regions, but tends to induce curvature in more narrow regions.

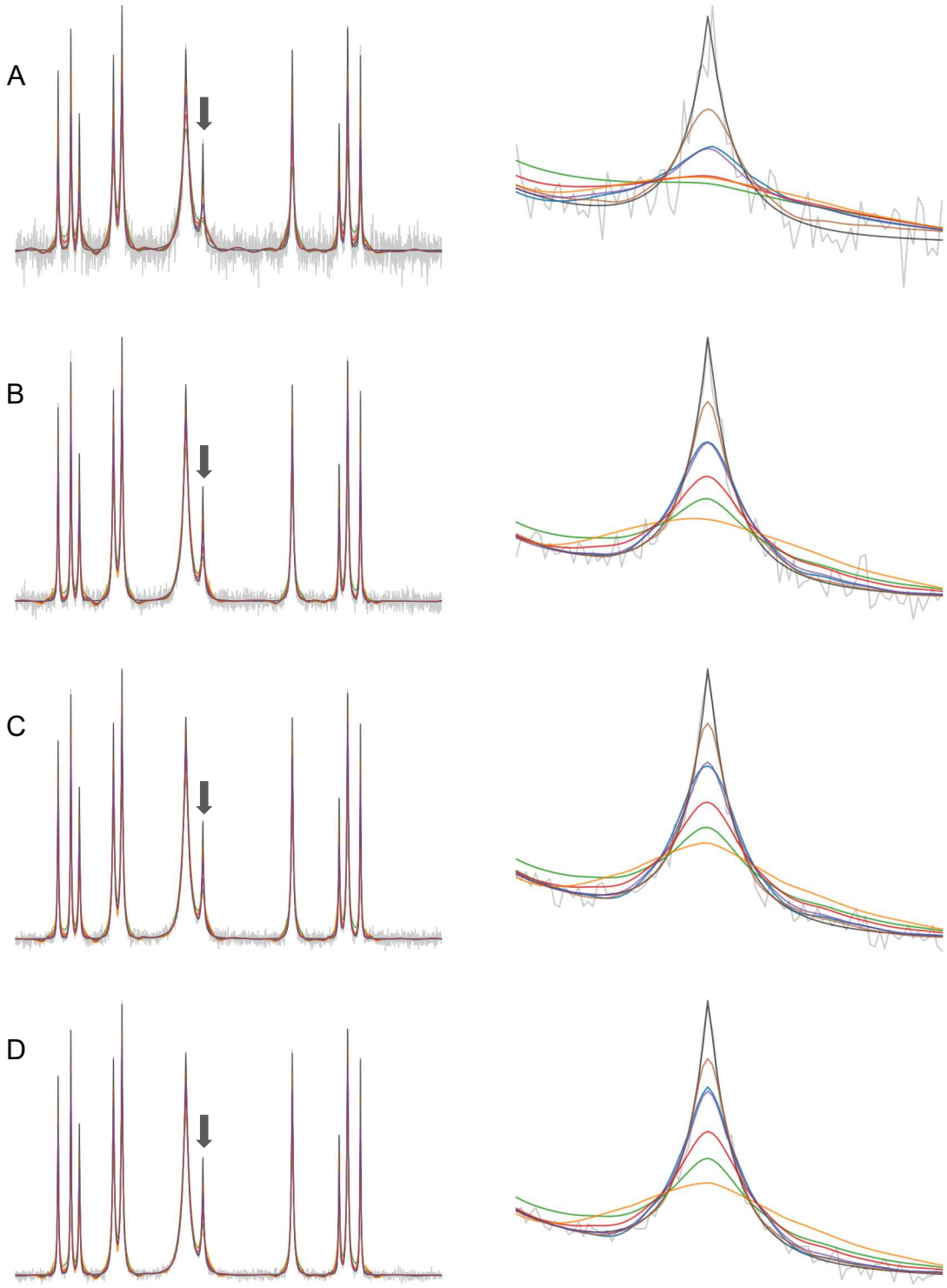


Figure S2. Performance of the SWT at level 6 across 6 different methods for the bumps test signal at SNR of 10 (A), 50 (B), 100 (C) or 200 (D). Methods are color coded as follows (threshold selection + method): universal+hard, universal+soft, universal+garotte, heuristic SURE+hard, MINIMAX+hard, universal+firm. The left panel shows whole trace performance while the right shows a zoomed region indicated by a black arrow. At low SNR, the best performance at feature boundaries is still obtained with MINIMAX+hard at the expense of incomplete denoising, and universal+hard still provides the best balance. Compared to level 5, the level 6 result excellently captures the baseline, but all methods fail to capture the signal peaks.

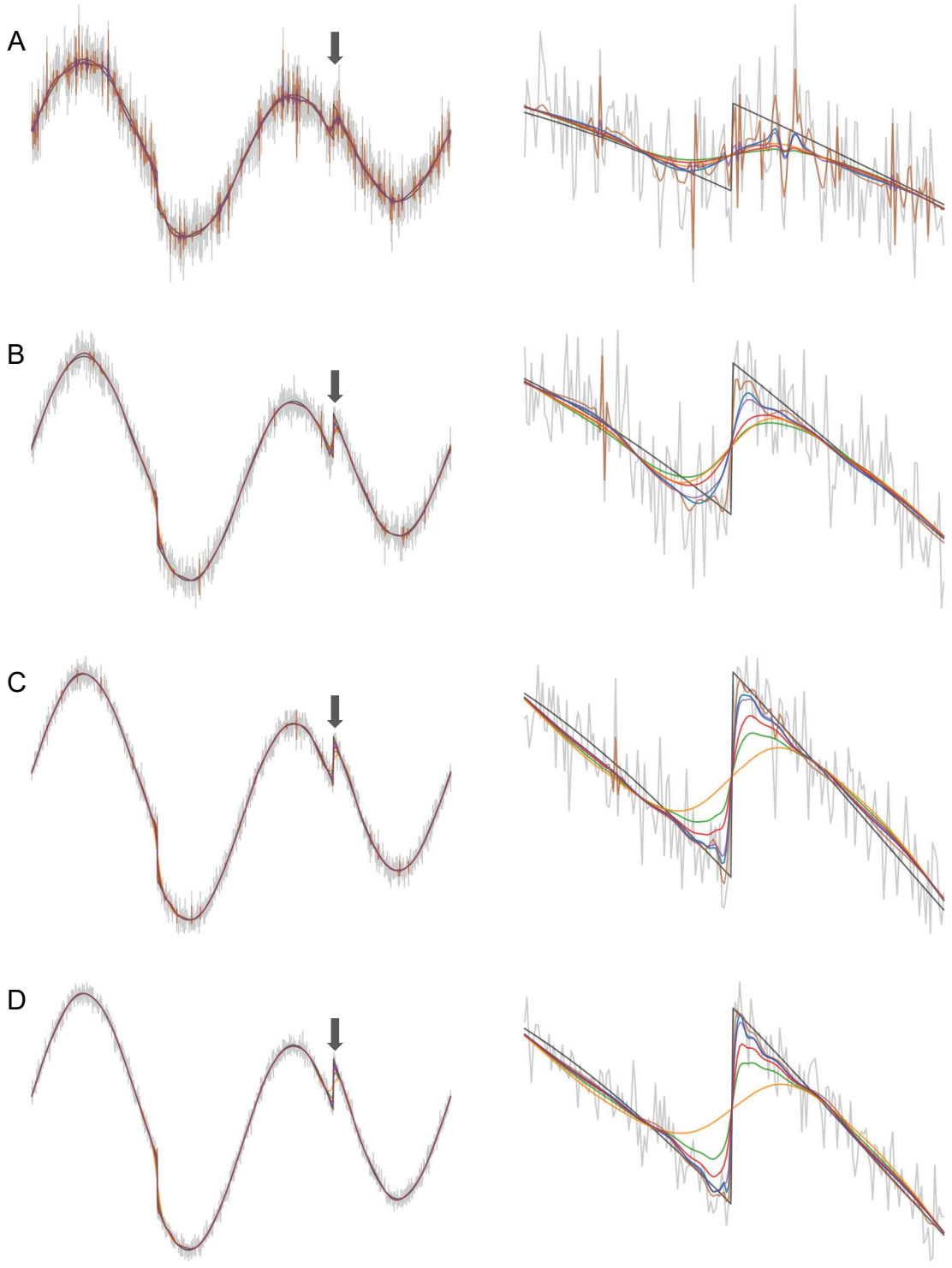


Figure S3. Performance of the SWT at level 6 across 6 different methods for the heavisine test signal at SNR of 10 (A), 50 (B), 100 (C) or 200 (D). Methods are color coded as follows (threshold selection + method): universal+hard, universal+soft, universal+garotte, heuristic SURE+hard, MINIMAX+hard, universal+firm. The left panel shows whole trace performance while the right shows a zoomed region indicated by a black arrow. At low SNR, the best performance at feature boundaries is still obtained with MINIMAX+hard at the expense of incomplete denoising, and universal+hard still provides the best balance. For this low frequency signal, the level 6 transform provides a generally smoother fit through the data with better visual agreement.

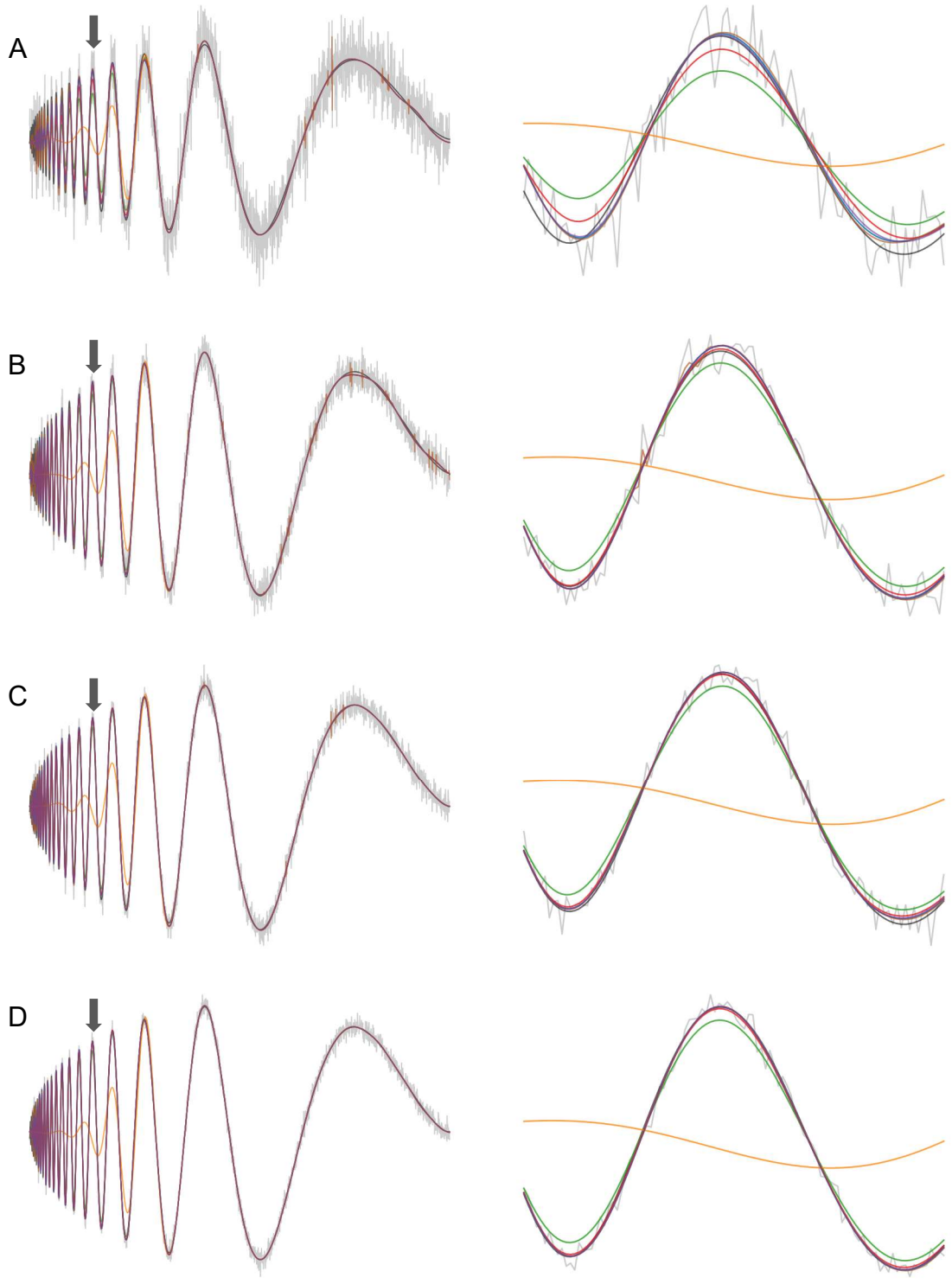


Figure S4. Performance of the SWT at level 6 across 6 different methods for the heavisine test signal at SNR of 10 (A), 50 (B), 100 (C) or 200 (D). Methods are color coded as follows (threshold selection + method): universal+hard, universal+soft, universal+garotte, heuristic SURE+hard, MINIMAX+hard, universal+firm. The left panel shows whole trace performance while the right shows a zoomed region indicated by a black arrow. At low SNR, the best performance at feature boundaries is still obtained with MINIMAX+hard at the expense of incomplete denoising, and universal+hard still provides the best balance. Even at level 6, most methods still capture the 5th oscillation (right side), but heuristic SURE+hard undershoots it completely.

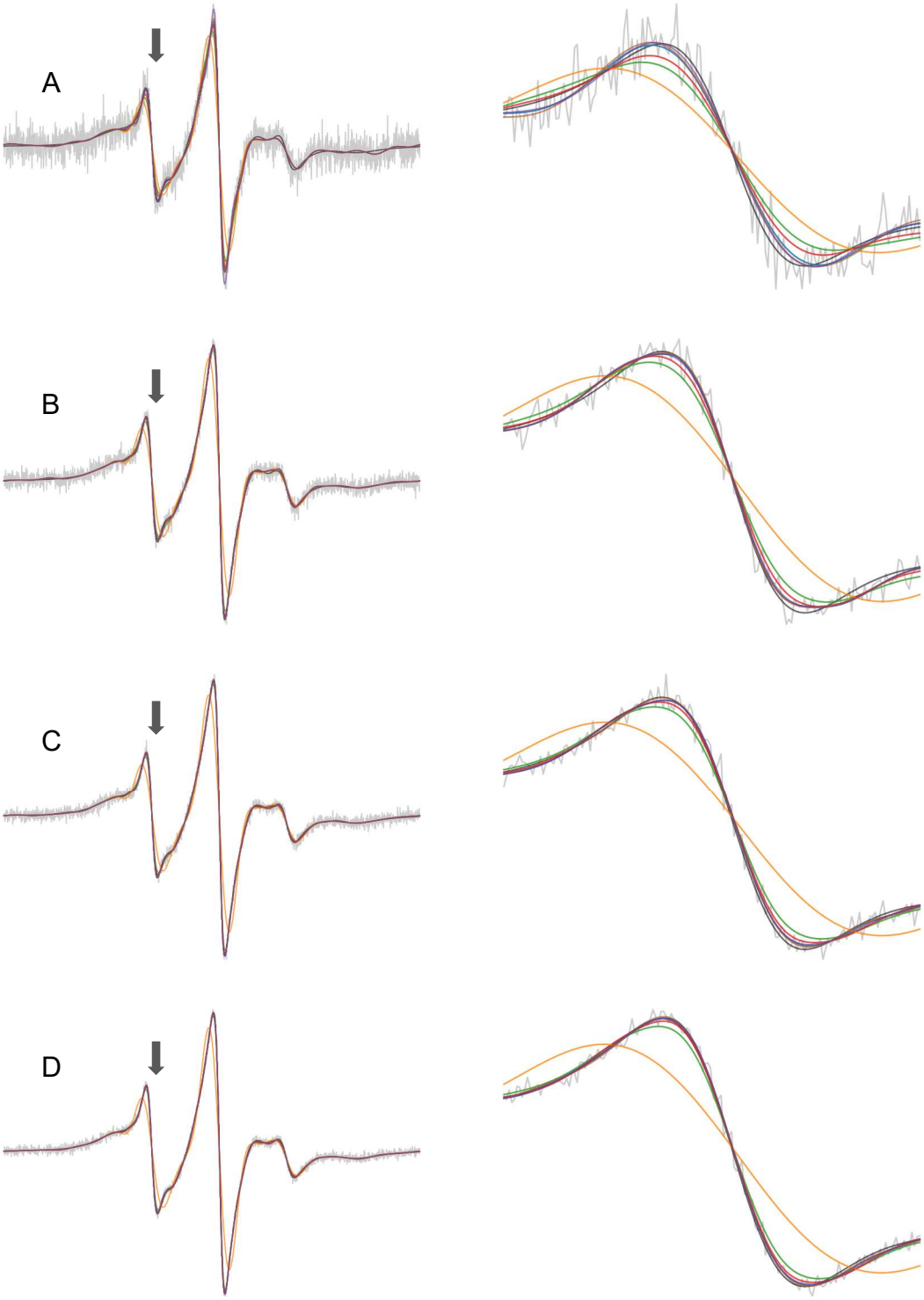


Figure S5. Performance of the SWT at level 6 across 6 different methods for the 1st EPR test signal at SNR of 10 (A), 50 (B), 100 (C) or 200 (D). Methods are color coded as follows (threshold selection + method): universal+hard, universal+soft, universal+garotte, heuristic SURE+hard, MINIMAX+hard, universal+firm. The left panel shows whole trace performance while the right shows a zoomed region indicated by a black arrow. . Across all SNR values, the universal+hard combination displays the best signal performance. At level 6, the baseline is almost oscillation free, although this comes at the expense of a small amount of undershoot in the right, highfield line.

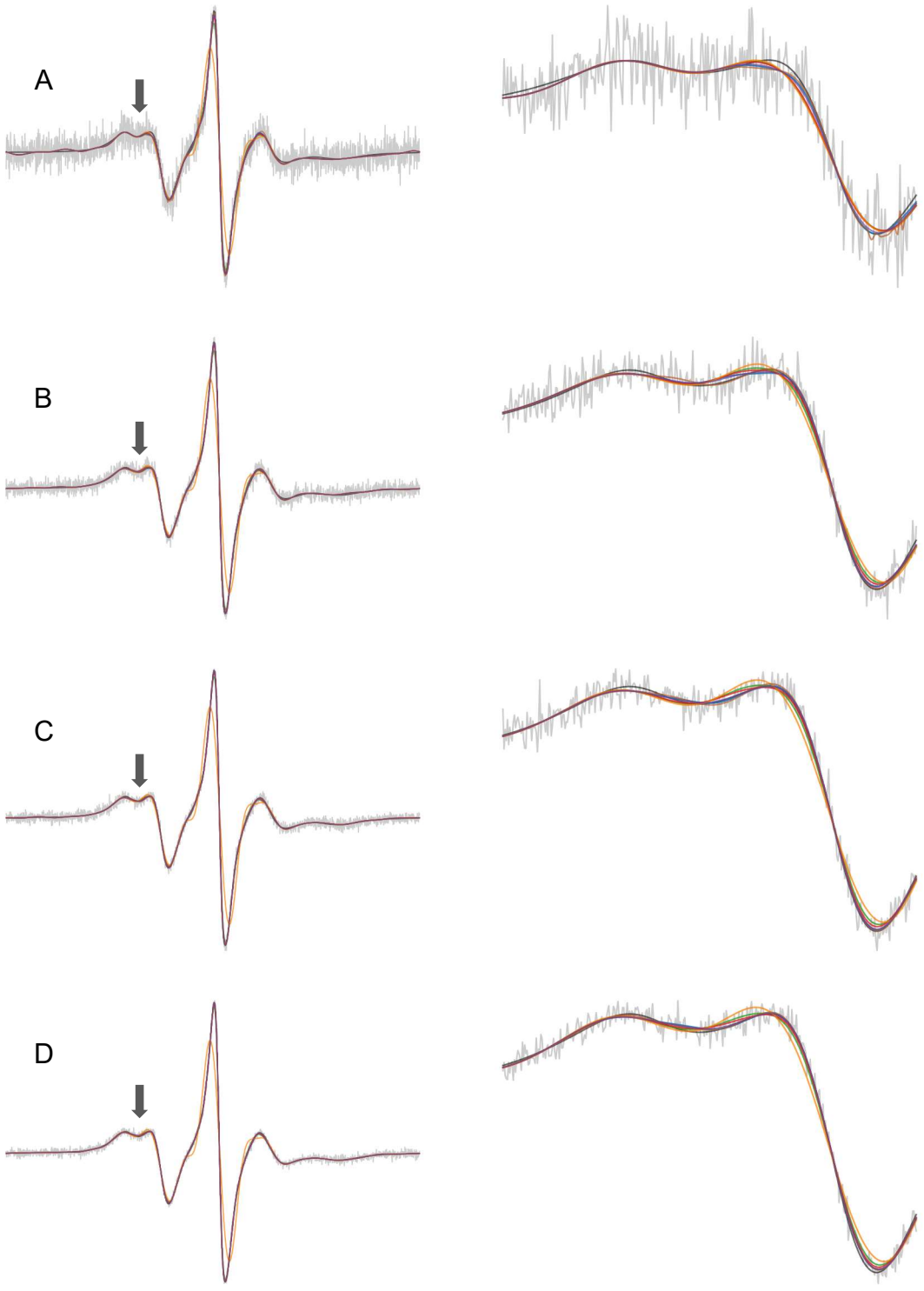


Figure S6. Performance of the SWT at level 6 across 6 different methods for the 2nd EPR test signal at SNR of 10 (A), 50 (B), 100 (C) or 200 (D). Methods are color coded as follows (threshold selection + method): universal+hard, universal+soft, universal+garotte, heuristic SURE+hard, MINIMAX+hard, universal+firm. The left panel shows whole trace performance while the right shows a zoomed region indicated by a black arrow. . Across all SNR values, the universal+hard combination displays the best signal performance. At level 6, the baseline is almost oscillation free, although this comes at the expense of a small amount of undershoot in the left, lowfield line.

				Signal																	
	COIF3	SNR=10		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	39.2904	0.7457	0.3876	38.2562	0.7335	0.1128	251.7343	0.9514	0.1953	126.2126	0.9205	0.0263	200.8590	0.9169	0.0002	209.3502	0.9134	0.0002
			6	17.3107	0.7162	0.5657	9.8761	0.6725	0.2054	292.5546	0.9508	0.1813	119.3653	0.9233	0.0268	186.7972	0.9361	0.0002	226.9048	0.9411	0.0002
	Soft	Universal	5	26.0395	0.7154	0.4708	7.0448	0.6219	0.2357	249.0657	0.9525	0.1964	51.2226	0.8888	0.0403	171.3662	0.9171	0.0002	172.6110	0.9122	0.0003
			6	10.9310	0.6824	0.6984	2.3331	0.5050	0.3628	260.7873	0.9498	0.1914	36.0471	0.8821	0.0472	68.0115	0.9246	0.0003	92.6128	0.9281	0.0004
	Garotte	Universal	5	27.9997	0.7212	0.4548	13.3578	0.6779	0.1792	249.3753	0.9525	0.1963	74.9511	0.9041	0.0336	187.1382	0.9168	0.0002	195.4268	0.9124	0.0002
			6	12.3579	0.6917	0.6596	3.5665	0.5671	0.3073	267.1845	0.9501	0.1892	63.7807	0.9039	0.0361	129.4550	0.9318	0.0002	168.5798	0.9360	0.0003
	Hard	Heur. SURE	5	50.9655	0.7606	0.3405	6.0418	0.5917	0.2557	250.9329	0.9520	0.1958	21.4232	0.8422	0.0616	124.6502	0.9178	0.0002	107.5135	0.9118	0.0003
			6	52.3782	0.7906	0.3344	4.7302	0.6029	0.2759	305.7495	0.9529	0.1775	8.8285	0.7798	0.0922	9.4271	0.8820	0.0007	6.8258	0.8687	0.0012
	Hard	MINIMAX	5	78.4501	0.7784	0.2779	75.7177	0.7679	0.0820	135.7663	0.8576	0.2807	145.2872	0.9000	0.0247	137.8284	0.8913	0.0002	132.2939	0.8846	0.0003
			6	28.4495	0.7538	0.4492	31.7795	0.7529	0.1226	185.7297	0.8680	0.2530	195.1075	0.9302	0.0211	208.0117	0.9388	0.0002	252.6938	0.9421	0.0002
	Firm	MINIMAX	5	39.8081	0.7462	0.3841	34.2403	0.7346	0.1180	255.0305	0.9481	0.1944	123.2121	0.9188	0.0265	198.9044	0.9162	0.0002	210.1505	0.9122	0.0002
			6	17.2309	0.7162	0.5657	8.9518	0.6695	0.2123	300.1422	0.9465	0.1790	120.0135	0.9229	0.0267	154.2569	0.9398	0.0002	227.7062	0.9411	0.0002

				Signal																	
	COIF3	SNR=50		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	117.6766	0.8614	0.2260	132.3755	0.8706	0.0621	967.0361	0.9776	0.0998	654.3880	0.9733	0.0115	496.9912	0.9744	0.0001	484.4573	0.9748	0.0002
			6	21.0114	0.7738	0.5177	14.0210	0.7515	0.1754	1106.2114	0.9793	0.0940	644.5249	0.9750	0.0116	398.5812	0.9744	0.0001	437.2156	0.9812	0.0002
	Soft	Universal	5	43.1152	0.8104	0.3682	17.3518	0.7874	0.1572	672.8937	0.9721	0.1194	169.8575	0.9515	0.0224	411.2515	0.9739	0.0001	421.0379	0.9744	0.0002
			6	12.0844	0.7195	0.6679	2.7190	0.5685	0.3391	549.0000	0.9675	0.1322	118.3611	0.9460	0.0266	152.7073	0.9645	0.0002	230.9166	0.9745	0.0002
	Garotte	Universal	5	58.7037	0.8308	0.3171	46.1187	0.8406	0.1016	750.6759	0.9745	0.1132	360.3592	0.9647	0.0155	469.3934	0.9741	0.0001	456.4432	0.9746	0.0002
			6	14.2290	0.7312	0.6190	4.5185	0.6366	0.2776	700.2866	0.9719	0.1177	293.9365	0.9629	0.0172	290.0510	0.9694	0.0001	379.2423	0.9795	0.0002
	Hard	Heur. SURE	5	59.9384	0.8346	0.3138	6.2061	0.6709	0.2520	564.8370	0.9677	0.1301	22.7118	0.8636	0.0598	174.2614	0.9726	0.0002	148.4155	0.9726	0.0003
			6	58.5596	0.8414	0.3163	4.8149	0.6414	0.2731	495.5619	0.9665	0.1394	8.9327	0.7930	0.0918	9.5827	0.9169	0.0007	6.9394	0.9134	0.0012
	Hard	MINIMAX	5	285.0594	0.8784	0.1459	267.4548	0.8961	0.0440	656.8990	0.9395	0.1243	700.6193	0.9634	0.0113	515.2282	0.9706	0.0001	468.2160	0.9670	0.0002
			6	38.3098	0.8176	0.3895	47.2658	0.8264	0.1011	1109.6473	0.9643	0.0969	899.4165	0.9759	0.0098	487.8101	0.9792	0.0001	515.1486	0.9830	0.0002
	Firm	MINIMAX	5	114.2561	0.8616	0.2290	127.9073	0.8706	0.0630	952.9107	0.9768	0.1006	646.9738	0.9733	0.0116	435.3151	0.9741	0.0001	359.1518	0.9744	0.0002
			6	20.9722	0.7694	0.5173	12.2868	0.7458	0.1841	1081.2067	0.9785	0.0949	619.2778	0.9747	0.0119	395.6048	0.9744	0.0001	448.4639	0.9817	0.0002

				Signal																	
	COIF3	SNR=100		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	220.3258	0.9077	0.1674	213.6559	0.9082	0.0491	2022.6848	0.9887	0.0694	1445.0967	0.9866	0.0077	668.1910	0.9836	0.0001	775.6943	0.9844	0.0001
			6	20.8872	0.7849	0.5179	13.2258	0.7620	0.1792	2128.1522	0.9893	0.0675	1339.6524	0.9859	0.0081	513.2964	0.9833	0.0001	522.1809	0.9867	0.0002
	Soft	Universal	5	57.4016	0.8569	0.3200	24.4045	0.8422	0.1347	1072.3945	0.9818	0.0944	316.0377	0.9709	0.0164	545.6898	0.9831	0.0001	552.0390	0.9837	0.0001
			6	11.9577	0.7271	0.6702	2.6288	0.5796	0.3433	798.3356	0.9741	0.1096	205.8504	0.9656	0.0202	216.0126	0.9743	0.0002	313.0371	0.9815	0.0002
	Garotte	Universal	5	89.7255	0.8809	0.2583	73.2369	0.8864	0.0819	1365.8412	0.9853	0.0839	790.2273	0.9807	0.0105	608.0694	0.9833	0.0001	627.2660	0.9840	0.0001
			6	14.0322	0.7386	0.6218	4.2942	0.6463	0.2828	1175.0474	0.9822	0.0907	611.0427	0.9783	0.0119	398.5573	0.9790	0.0001	470.3285	0.9852	0.0002
	Hard	Heur. SURE	5	60.8600	0.8662	0.3108	6.1961	0.7031	0.2517	670.1873	0.9728	0.1191	22.7148	0.8679	0.0598	185.9973	0.9812	0.0002	153.0335	0.9813	0.0003
			6	59.0785	0.8551	0.3144	4.7928	0.6553	0.2732	545.4891	0.9696	0.1323	8.9476	0.7955	0.0917	9.6011	0.9230	0.0007	6.9417	0.9181	0.0012
	Hard	MINIMAX	5	622.9315	0.9188	0.0998	428.1358	0.9377	0.0349	1353.1443	0.9635	0.0861	1472.3381	0.9809	0.0078	879.0543	0.9821	0.0001	775.5827	0.9788	0.0001
			6	38.1791	0.8305	0.3889	45.2654	0.8368	0.1029	2448.0742	0.9816	0.0647	2013.1140	0.9889	0.0066	620.1223	0.9865	0.0001	762.3302	0.9879	0.0001
	Firm	MINIMAX	5	211.0884	0.9087	0.1705	200.6697	0.9115	0.0507	2015.0713	0.9885	0.0692	1445.7176	0.9866	0.0077	501.3063	0.9835	0.0001	488.0483	0.9840	0.0002
			6	20.6287	0.7760	0.5201	11.5819	0.7553	0.1882	2106.5750	0.9893	0.0678	1355.9751	0.9863	0.0080	515.3938	0.9834	0.0001	558.1464	0.9868	0.0001

				Signal																	
	COIF3	SNR=200		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	435.9799	0.9323	0.1187	292.3541	0.9291	0.0421	3965.4110	0.9935	0.0494	3046.4523	0.9919	0.0053	967.6530	0.9887	0.0001	1223.8032	0.9899	0.0001
			6	21.3155	0.7922	0.5138	14.2678	0.7706	0.1738	4006.7168	0.9931	0.0496	2353.8329	0.9909	0.0061	639.6839	0.9882	0.0001	672.9855	0.9903	0.0001
	Soft	Universal	5	77.6021	0.8841	0.2754	32.0694	0.8690	0.1184	1632.9167	0.9882	0.0765	535.1242	0.9803	0.0126	660.4859	0.9879	0.0001	762.6194	0.9892	0.0001
			6	12.1343	0.7323	0.6665	2.7315	0.5910	0.3381	1187.9494	0.9812	0.0897	293.3397	0.9736	0.0170	311.8605	0.9810	0.0001	410.9174	0.9869	0.0002
	Garotte	Universal	5	140.2059	0.9105	0.2063	104.9249	0.9070	0.0686	2389.3626	0.9914	0.0635	1586.7429	0.9881	0.0074	770.7496	0.9882	0.0001	933.8030	0.9895	0.0001
			6	14.2994	0.7439	0.6172	4.5561	0.6583	0.2763	2033.6369	0.9894	0.0688	955.7837	0.9844	0.0095	521.6119	0.9852	0.0001	551.2631	0.9893	0.0001
	Hard	Heur. SURE	5	61.7689	0.8811	0.3090	6.2119	0.7145	0.2514	732.9354	0.9755	0.1138	22.8159	0.8687	0.0597	188.3633	0.9856	0.0002	159.7469	0.9863	0.0003
			6	59.6630	0.8613	0.3133	4.8083	0.6614	0.2728	542.0173	0.9696	0.1327	8.9632	0.7961	0.0916	9.6334	0.9259	0.0007	6.9445	0.9218	0.0012
	Hard	MINIMAX	5	1130.4263	0.9401	0.0734	638.7912	0.9574	0.0285	3027.7771	0.9828	0.0566	3004.7534	0.9890	0.0054	1367.4628	0.9889	0.0001	1365.1770	0.9887	0.0001
			6	39.4177	0.8392	0.3837	49.7865	0.8480	0.0985	4584.1827	0.9893	0.0466	3976.8059	0.9945	0.0047	963.0752	0.9901	0.0001	1266.4364	0.9918	0.0001
	Firm	MINIMAX	5	407.6570	0.9354	0.1219	293.4173	0.9335	0.0419	3848.6653	0.9933	0.0502	3041.6173	0.9920	0.0053	592.3201	0.9885	0.0001	1138.8663	0.9900	0.0001
			6	21.0941	0.7823	0.5154	12.4834	0.7674	0.1826	3954.4179	0.9934	0.0498	2391.3118	0.9916	0.0060	667.5651	0.9883	0.0001	759.6951	0.9906	0.0001

Table S1. Performance of the SWT at decomposition levels 5 and 6 with the coif 3 wavelet for 6 thresholding combinations. The tables give the performance of each combination in terms of SNR, SSIM, and RMSE in order of increasing starting SNR (10, 50, 100, 200) of the noisy test signal. Results are presented for the following test signals: blocks, bumps, heavisine, doppler, 1st EPR test signal, and 2nd EPR test signal. Method refers to the thresholding form, while selection refers to the means of calculating the threshold value.

				Signal																	
	COIF4	SNR=10		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	35.9052	0.7388	0.4046	31.4528	0.7140	0.1238	248.2443	0.9506	0.1968	129.0582	0.9207	0.0261	195.8470	0.9164	0.0002	206.0534	0.9123	0.0002
			6	16.1175	0.7046	0.5848	7.3690	0.6365	0.2331	282.0648	0.9493	0.1847	120.5438	0.9227	0.0267	162.2021	0.9305	0.0002	216.1024	0.9397	0.0002
	Soft	Universal	5	25.2813	0.7124	0.4780	6.3290	0.6018	0.2475	246.0888	0.9521	0.1976	52.7927	0.8895	0.0397	165.3036	0.9168	0.0002	165.0094	0.9112	0.0003
			6	10.3700	0.6727	0.7171	2.0690	0.4785	0.3824	256.2397	0.9490	0.1931	36.9173	0.8823	0.0467	58.8070	0.9175	0.0003	85.4192	0.9239	0.0004
	Garotte	Universal	5	26.7425	0.7170	0.4652	11.3819	0.6543	0.1929	246.3365	0.9521	0.1975	77.4289	0.9046	0.0331	180.3249	0.9164	0.0002	187.8255	0.9114	0.0003
			6	11.5310	0.6818	0.6822	2.9413	0.5348	0.3332	261.7692	0.9491	0.1911	65.1905	0.9037	0.0358	108.8544	0.9258	0.0002	152.3667	0.9329	0.0003
	Hard	Heur. SURE	5	48.4654	0.7562	0.3489	5.7521	0.5765	0.2621	247.2566	0.9516	0.1972	21.2484	0.8410	0.0619	125.6078	0.9174	0.0002	108.8837	0.9110	0.0003
			6	49.2777	0.7828	0.3443	4.4971	0.5899	0.2826	296.4051	0.9521	0.1802	8.7779	0.7786	0.0926	9.3010	0.8730	0.0007	6.8460	0.8619	0.0012
	Hard	MINIMAX	5	72.5885	0.7730	0.2888	67.6204	0.7565	0.0865	134.6387	0.8534	0.2823	143.2566	0.8968	0.0249	138.0696	0.8907	0.0002	131.8611	0.8820	0.0003
			6	24.9435	0.7408	0.4778	24.5838	0.7275	0.1384	183.9663	0.8676	0.2544	192.3958	0.9286	0.0213	191.7769	0.9355	0.0002	250.2723	0.9414	0.0002
	Firm	MINIMAX	5	36.7563	0.7400	0.3990	28.2232	0.7154	0.1293	251.0934	0.9471	0.1959	126.0477	0.9189	0.0262	195.7465	0.9157	0.0002	207.6875	0.9111	0.0002
			6	15.7982	0.7046	0.5891	6.7572	0.6309	0.2395	289.9108	0.9447	0.1823	121.8164	0.9223	0.0266	133.6359	0.9356	0.0002	190.1030	0.9404	0.0003

				Signal																	
	COIF4	SNR=50		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	105.7130	0.8540	0.2386	97.4161	0.8483	0.0723	928.5443	0.9770	0.1018	676.4126	0.9734	0.0113	486.5547	0.9740	0.0001	479.8274	0.9745	0.0002
			6	19.0621	0.7511	0.5425	8.7343	0.7006	0.2161	1023.3733	0.9781	0.0975	677.3955	0.9754	0.0113	349.0700	0.9722	0.0001	428.3581	0.9806	0.0002
	Soft	Universal	5	40.4096	0.8037	0.3805	13.9661	0.7534	0.1738	655.0390	0.9712	0.1210	177.4126	0.9520	0.0219	394.0117	0.9736	0.0001	412.2868	0.9740	0.0002
			6	11.3890	0.7029	0.6878	2.2433	0.5299	0.3682	521.1115	0.9664	0.1357	125.3222	0.9468	0.0260	125.6510	0.9587	0.0002	220.1285	0.9731	0.0002
	Garotte	Universal	5	53.5251	0.8225	0.3322	34.0819	0.8114	0.1172	725.5065	0.9736	0.1151	377.4478	0.9650	0.0152	454.3813	0.9738	0.0001	451.4551	0.9742	0.0002
			6	13.1786	0.7139	0.6424	3.3440	0.5901	0.3145	652.1424	0.9705	0.1218	313.0257	0.9633	0.0167	234.5317	0.9643	0.0002	368.5970	0.9791	0.0002
	Hard	Heur. SURE	5	57.0754	0.8289	0.3215	5.8131	0.6506	0.2601	555.2683	0.9666	0.1312	22.5119	0.8623	0.0601	176.8102	0.9724	0.0002	151.0636	0.9723	0.0003
			6	55.3164	0.8279	0.3252	4.4967	0.6254	0.2818	467.1477	0.9653	0.1435	8.8855	0.7919	0.0922	9.4507	0.9087	0.0007	6.9545	0.9062	0.0012
	Hard	MINIMAX	5	242.9689	0.8705	0.1583	213.2484	0.8810	0.0492	647.0876	0.9383	0.1251	698.9193	0.9610	0.0113	507.4947	0.9702	0.0001	455.3316	0.9658	0.0002
			6	31.7054	0.7951	0.4271	30.4385	0.7898	0.1243	1064.0682	0.9644	0.0982	911.0589	0.9748	0.0098	471.2778	0.9795	0.0001	510.2059	0.9828	0.0002
	Firm	MINIMAX	5	101.4251	0.8536	0.2433	93.6644	0.8495	0.0733	916.3671	0.9760	0.1025	671.1986	0.9734	0.0114	439.9303	0.9737	0.0001	369.3515	0.9740	0.0002
			6	18.8816	0.7466	0.5436	8.1230	0.6921	0.2202	999.0261	0.9772	0.0987	655.7505	0.9752	0.0115	343.9621	0.9718	0.0001	442.9016	0.9816	0.0002

				Signal																	
	COIF4	SNR=100		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	165.0189	0.8969	0.1928	141.8517	0.8864	0.0603	1918.3577	0.9882	0.0711	1549.5580	0.9871	0.0075	665.2106	0.9834	0.0001	769.7876	0.9842	0.0001
			6	18.5700	0.7570	0.5480	8.6582	0.7147	0.2164	1951.9919	0.9886	0.0704	1530.7926	0.9871	0.0076	496.2153	0.9842	0.0001	515.8071	0.9867	0.0002
	Soft	Universal	5	50.5561	0.8468	0.3408	18.2711	0.8044	0.1542	1025.8273	0.9808	0.0964	338.2191	0.9717	0.0159	528.7430	0.9829	0.0001	542.0270	0.9835	0.0001
			6	11.2151	0.7084	0.6918	2.2336	0.5450	0.3683	758.1462	0.9732	0.1124	229.4200	0.9671	0.0192	180.6361	0.9700	0.0002	299.7449	0.9807	0.0002
	Garotte	Universal	5	74.6356	0.8692	0.2828	49.0371	0.8565	0.0994	1289.1846	0.9844	0.0862	848.5424	0.9811	0.0101	598.9911	0.9831	0.0001	621.2706	0.9838	0.0001
			6	12.8957	0.7193	0.6477	3.3257	0.6044	0.3148	1095.7672	0.9811	0.0938	694.5931	0.9794	0.0112	338.7550	0.9763	0.0001	460.9681	0.9852	0.0002
	Hard	Heur. SURE	5	57.2058	0.8595	0.3205	5.7995	0.6789	0.2600	663.2329	0.9721	0.1197	22.5270	0.8667	0.0601	189.0535	0.9811	0.0002	156.1583	0.9811	0.0003
			6	54.9785	0.8393	0.3256	4.4739	0.6392	0.2820	513.6680	0.9684	0.1364	8.8999	0.7944	0.0921	9.4665	0.9145	0.0007	6.9581	0.9108	0.0012
	Hard	MINIMAX	5	479.7317	0.9106	0.1135	305.0808	0.9213	0.0414	1334.6661	0.9639	0.0863	1476.0339	0.9790	0.0078	866.0724	0.9816	0.0001	762.7388	0.9782	0.0001
			6	30.7671	0.8030	0.4315	30.5932	0.8044	0.1238	2226.6253	0.9796	0.0679	2033.2418	0.9882	0.0066	607.8531	0.9865	0.0001	743.2160	0.9877	0.0001
	Firm	MINIMAX	5	163.9634	0.8986	0.1932	136.5921	0.8884	0.0612	1911.0495	0.9880	0.0710	1516.7946	0.9869	0.0076	519.7587	0.9833	0.0001	501.3694	0.9838	0.0002
			6	18.4386	0.7509	0.5482	8.0906	0.7067	0.2202	1951.8110	0.9887	0.0704	1505.0257	0.9871	0.0076	477.6022	0.9830	0.0001	550.7523	0.9867	0.0001

				Signal																	
	COIF4	SNR=200		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	266.1629	0.9200	0.1510	179.5064	0.9014	0.0534	3586.3914	0.9927	0.0522	3291.0501	0.9924	0.0051	946.5564	0.9886	0.0001	1242.7769	0.9899	0.0001
			6	19.0007	0.7651	0.5429	8.6574	0.7199	0.2164	3469.7197	0.9920	0.0530	2969.5079	0.9924	0.0055	619.3798	0.9884	0.0001	637.5076	0.9901	0.0001
	Soft	Universal	5	62.8888	0.8703	0.3056	21.0388	0.8260	0.1439	1533.5635	0.9873	0.0789	584.0959	0.9810	0.0121	644.5811	0.9878	0.0001	758.7242	0.9892	0.0001
			6	11.3790	0.7144	0.6880	2.2401	0.5513	0.3678	1088.6471	0.9797	0.0937	338.3716	0.9754	0.0159	265.2068	0.9783	0.0002	393.7376	0.9863	0.0002
	Garotte	Universal	5	102.5536	0.8958	0.2406	59.3675	0.8746	0.0900	2203.7026	0.9906	0.0661	1742.3873	0.9886	0.0071	761.7111	0.9881	0.0001	940.2266	0.9895	0.0001
			6	13.1446	0.7254	0.6427	3.3399	0.6108	0.3141	1806.4823	0.9880	0.0730	1144.4846	0.9857	0.0088	468.0640	0.9842	0.0001	536.7975	0.9892	0.0001
	Hard	Heur. SURE	5	58.0444	0.8736	0.3186	5.8083	0.6896	0.2598	726.7183	0.9748	0.1143	22.6263	0.8675	0.0600	191.9570	0.9855	0.0002	162.7804	0.9863	0.0003
			6	55.5274	0.8460	0.3244	4.4763	0.6443	0.2819	506.9860	0.9682	0.1373	8.9155	0.7951	0.0920	9.4962	0.9174	0.0007	6.9598	0.9144	0.0012
	Hard	MINIMAX	5	841.7390	0.9316	0.0848	397.1334	0.9391	0.0362	2994.2741	0.9829	0.0569	3065.8261	0.9881	0.0054	1370.8620	0.9888	0.0001	1340.3287	0.9884	0.0001
			6	31.8306	0.8126	0.4252	31.0093	0.8120	0.1228	4294.2586	0.9892	0.0479	4346.2393	0.9945	0.0045	872.8757	0.9898	0.0001	1233.0313	0.9917	0.0001
	Firm	MINIMAX	5	262.4699	0.9237	0.1518	169.4106	0.9056	0.0548	3564.2700	0.9928	0.0522	3224.1386	0.9923	0.0052	613.3731	0.9884	0.0001	1150.1662	0.9900	0.0001
			6	18.8438	0.7580	0.5435	8.1359	0.7136	0.2194	3484.5952	0.9926	0.0530	2811.3139	0.9924	0.0056	630.7379	0.9882	0.0001	733.3395	0.9904	0.0001

Table S2. Performance of the SWT at decomposition levels 5 and 6 with the coif 4 wavelet for 6 thresholding combinations. The tables give the performance of each combination in terms of SNR, SSIM, and RMSE in order of increasing starting SNR (10, 50, 100, 200) of the noisy test signal. Results are presented for the following test signals: blocks, bumps, heavisine, doppler, 1st EPR test signal, and 2nd EPR test signal. Method refers to the thresholding form, while selection refers to the means of calculating the threshold value.

				Signal																	
	COIF5	SNR=10		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	34.2653	0.7347	0.4137	26.8285	0.6993	0.1335	245.9187	0.9502	0.1977	128.7084	0.9203	0.0261	191.2331	0.9158	0.0002	203.0394	0.9115	0.0002
			6	15.0199	0.6943	0.6044	6.4241	0.6158	0.2470	275.1030	0.9489	0.1872	118.9055	0.9216	0.0269	142.0215	0.9259	0.0002	203.5528	0.9377	0.0002
	Soft	Universal	5	24.8930	0.7107	0.4818	5.8059	0.5851	0.2575	244.0497	0.9517	0.1984	53.5512	0.8896	0.0395	160.7549	0.9165	0.0002	158.7573	0.9105	0.0003
			6	9.9325	0.6649	0.7326	1.9378	0.4627	0.3937	252.6133	0.9485	0.1945	37.0567	0.8817	0.0467	52.0234	0.9116	0.0003	79.7137	0.9203	0.0004
	Garotte	Universal	5	26.1201	0.7147	0.4708	9.9980	0.6346	0.2047	244.2450	0.9517	0.1984	78.3631	0.9044	0.0329	175.1535	0.9161	0.0002	181.1042	0.9107	0.0003
			6	10.8641	0.6732	0.7021	2.6420	0.5149	0.3488	257.4440	0.9487	0.1928	65.0523	0.9029	0.0359	94.4455	0.9207	0.0003	140.7925	0.9304	0.0003
	Hard	Heur. SURE	5	46.9546	0.7533	0.3544	5.5259	0.5642	0.2673	245.0180	0.9512	0.1982	21.1133	0.8401	0.0621	125.6039	0.9171	0.0002	109.2518	0.9103	0.0003
			6	47.2757	0.7765	0.3513	4.2865	0.5795	0.2890	290.4076	0.9513	0.1820	8.7356	0.7778	0.0930	9.1625	0.8657	0.0008	6.8465	0.8567	0.0012
	Hard	MINIMAX	5	69.2139	0.7688	0.2954	60.7254	0.7438	0.0911	133.4758	0.8534	0.2826	141.6631	0.8957	0.0250	136.3864	0.8902	0.0002	133.8481	0.8825	0.0003
			6	22.7022	0.7308	0.4993	20.3301	0.7048	0.1512	193.8519	0.8764	0.2443	193.5711	0.9281	0.0213	190.3413	0.9361	0.0002	251.1743	0.9416	0.0002
	Firm	MINIMAX	5	35.2917	0.7365	0.4070	24.0394	0.6994	0.1394	248.5419	0.9468	0.1969	126.7758	0.9186	0.0261	192.3381	0.9153	0.0002	204.5309	0.9104	0.0002
			6	14.6904	0.6949	0.6094	5.7426	0.6080	0.2568	284.4572	0.9456	0.1838	121.0155	0.9214	0.0267	103.4059	0.9344	0.0003	136.0136	0.9420	0.0004

				Signal																	
	COIF5	SNR=50		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	97.2119	0.8483	0.2485	75.8932	0.8289	0.0815	904.2523	0.9767	0.1031	680.0136	0.9732	0.0113	477.4955	0.9737	0.0001	470.7685	0.9741	0.0002
			6	17.7688	0.7335	0.5605	7.4277	0.6762	0.2312	969.5799	0.9773	0.1003	669.9741	0.9749	0.0114	312.0494	0.9706	0.0001	430.7919	0.9812	0.0002
	Soft	Universal	5	38.6942	0.7994	0.3888	11.5997	0.7247	0.1891	642.5065	0.9705	0.1221	179.4286	0.9517	0.0218	378.2160	0.9733	0.0001	402.0949	0.9737	0.0002
			6	10.8805	0.6908	0.7034	2.0773	0.5124	0.3808	505.0368	0.9658	0.1378	123.0495	0.9455	0.0262	108.4300	0.9546	0.0002	209.7058	0.9716	0.0002
	Garotte	Universal	5	50.2765	0.8170	0.3426	26.2768	0.7861	0.1323	707.6754	0.9730	0.1165	380.3814	0.9647	0.0151	439.9448	0.9735	0.0001	444.5478	0.9739	0.0002
			6	12.4091	0.7012	0.6610	2.9566	0.5683	0.3312	624.6830	0.9695	0.1245	304.8056	0.9620	0.0170	202.3280	0.9613	0.0002	358.5973	0.9786	0.0002
	Hard	Heur. SURE	5	55.1914	0.8244	0.3269	5.5970	0.6361	0.2651	549.3318	0.9660	0.1319	22.3543	0.8613	0.0604	177.2800	0.9722	0.0002	151.7936	0.9720	0.0003
			6	53.0417	0.8170	0.3319	4.3225	0.6161	0.2873	450.5415	0.9643	0.1463	8.8456	0.7912	0.0925	9.3063	0.9018	0.0007	6.9509	0.9008	0.0012
	Hard	MINIMAX	5	219.8947	0.8658	0.1664	173.0495	0.8658	0.0545	641.7667	0.9378	0.1259	708.8202	0.9603	0.0113	505.0913	0.9703	0.0001	463.3897	0.9659	0.0002
			6	28.0883	0.7797	0.4518	24.2321	0.7673	0.1384	1042.0788	0.9640	0.0991	942.9292	0.9753	0.0096	451.6910	0.9791	0.0001	497.8094	0.9823	0.0002
	Firm	MINIMAX	5	93.2296	0.8477	0.2537	70.9052	0.8305	0.0839	890.7495	0.9754	0.1040	675.1045	0.9733	0.0113	441.8430	0.9734	0.0001	377.4904	0.9737	0.0002
			6	17.4738	0.7300	0.5633	6.7536	0.6652	0.2383	949.1975	0.9763	0.1012	643.7166	0.9745	0.0117	314.4057	0.9704	0.0001	435.6191	0.9813	0.0002

				Signal																	
	COIF5	SNR=100		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	138.5960	0.8895	0.2098	99.0651	0.8632	0.0718	1833.9204	0.9876	0.0728	1576.0468	0.9872	0.0074	657.0315	0.9833	0.0001	770.7123	0.9841	0.0001
			6	17.6176	0.7383	0.5618	7.5524	0.6927	0.2294	1845.3986	0.9880	0.0723	1534.1732	0.9871	0.0076	458.5849	0.9828	0.0001	501.2712	0.9862	0.0002
	Soft	Universal	5	46.3927	0.8398	0.3557	13.8861	0.7671	0.1747	994.1497	0.9800	0.0979	346.0324	0.9717	0.0158	512.8109	0.9827	0.0001	532.9110	0.9833	0.0002
			6	10.7929	0.6953	0.7052	2.0960	0.5299	0.3788	730.2146	0.9725	0.1145	228.3804	0.9664	0.0192	156.5728	0.9669	0.0002	282.5089	0.9796	0.0002
	Garotte	Universal	5	65.9442	0.8608	0.3007	33.7368	0.8261	0.1184	1237.5945	0.9836	0.0879	865.2308	0.9811	0.0100	589.1419	0.9829	0.0001	617.0429	0.9836	0.0001
			6	12.2687	0.7059	0.6636	3.0027	0.5857	0.3287	1039.2808	0.9803	0.0963	687.3909	0.9788	0.0112	300.7501	0.9745	0.0001	448.1095	0.9848	0.0002
	Hard	Heur. SURE	5	55.0778	0.8549	0.3265	5.5730	0.6610	0.2652	658.1572	0.9716	0.1202	22.3787	0.8658	0.0603	189.7536	0.9809	0.0002	157.1578	0.9809	0.0003
			6	52.6838	0.8280	0.3324	4.2975	0.6283	0.2876	493.7265	0.9677	0.1391	8.8594	0.7937	0.0924	9.3197	0.9074	0.0007	6.9554	0.9053	0.0012
	Hard	MINIMAX	5	405.9475	0.9048	0.1235	219.8949	0.9016	0.0485	1325.9118	0.9636	0.0867	1481.9069	0.9782	0.0078	859.7482	0.9816	0.0001	757.7529	0.9779	0.0001
			6	27.6555	0.7883	0.4546	25.7596	0.7850	0.1343	2179.5845	0.9796	0.0681	2074.5236	0.9885	0.0065	593.1300	0.9861	0.0001	719.1125	0.9874	0.0001
	Firm	MINIMAX	5	137.6594	0.8909	0.2107	93.9431	0.8654	0.0736	1832.7230	0.9876	0.0725	1533.4767	0.9869	0.0075	532.0710	0.9831	0.0001	511.2195	0.9836	0.0002
			6	17.2810	0.7346	0.5653	6.9172	0.6822	0.2356	1842.4255	0.9881	0.0724	1489.8284	0.9871	0.0077	445.4389	0.9821	0.0001	535.7472	0.9864	0.0002

				Signal																	
	COIF5	SNR=200		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	208.2155	0.9134	0.1700	111.9661	0.8790	0.0672	3365.2407	0.9921	0.0538	3348.3633	0.9925	0.0051	938.6651	0.9885	0.0001	1235.7235	0.9899	0.0001
			6	17.8817	0.7459	0.5587	7.4515	0.6972	0.2307	3210.7574	0.9914	0.0553	3001.9109	0.9925	0.0054	578.5280	0.9878	0.0001	616.5406	0.9899	0.0001
	Soft	Universal	5	55.8884	0.8617	0.3239	15.0849	0.7859	0.1676	1476.6719	0.9867	0.0804	603.1572	0.9811	0.0119	628.3121	0.9877	0.0001	752.0269	0.9891	0.0001
			6	10.9178	0.7016	0.7024	2.0909	0.5351	0.3792	1021.0333	0.9785	0.0969	347.0005	0.9753	0.0156	228.6387	0.9758	0.0002	377.1889	0.9856	0.0002
	Garotte	Universal	5	86.2422	0.8859	0.2619	37.7844	0.8431	0.1114	2096.9345	0.9901	0.0677	1794.9318	0.9886	0.0070	750.3304	0.9880	0.0001	940.3162	0.9895	0.0001
			6	12.4543	0.7122	0.6598	2.9911	0.5910	0.3292	1658.1665	0.9868	0.0764	1173.6201	0.9857	0.0086	420.7774	0.9829	0.0001	523.8939	0.9888	0.0002
	Hard	Heur. SURE	5	55.9466	0.8680	0.3244	5.6148	0.6718	0.2643	722.0750	0.9743	0.1147	22.4769	0.8666	0.0602	192.9850	0.9854	0.0002	163.6533	0.9862	0.0003
			6	53.3230	0.8350	0.3308	4.3158	0.6338	0.2870	483.3397	0.9673	0.1406	8.8749	0.7944	0.0923	9.3476	0.9102	0.0007	6.9562	0.9088	0.0012
	Hard	MINIMAX	5	692.8110	0.9261	0.0934	266.1092	0.9146	0.0441	2958.4999	0.9828	0.0572	2994.5495	0.9873	0.0055	1362.3670	0.9888	0.0001	1325.9304	0.9882	0.0001
			6	28.3074	0.7960	0.4501	25.5206	0.7906	0.1348	4180.0672	0.9894	0.0483	4488.4772	0.9946	0.0044	800.5943	0.9895	0.0001	1216.9505	0.9917	0.0001
	Firm	MINIMAX	5	203.5541	0.9155	0.1717	107.1882	0.8798	0.0685	3392.7314	0.9923	0.0535	3289.9461	0.9923	0.0051	629.3615	0.9883	0.0001	1152.3854	0.9899	0.0001
			6	17.5801	0.7417	0.5616	6.8746	0.6878	0.2361	3173.3330	0.9919	0.0555	2892.3908	0.9926	0.0055	588.5430	0.9876	0.0001	708.2885	0.9902	0.0001

Table S3. Performance of the SWT at decomposition levels 5 and 6 with the coif 5 wavelet for 6 thresholding combinations. The tables give the performance of each combination in terms of SNR, SSIM, and RMSE in order of increasing starting SNR (10, 50, 100, 200) of the noisy test signal. Results are presented for the following test signals: blocks, bumps, heavisine, doppler, 1st EPR test signal, and 2nd EPR test signal. Method refers to the thresholding form, while selection refers to the means of calculating the threshold value.

				Signal																	
	BIOR22	SNR=10		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	41.4970	0.7579	0.3763	39.6093	0.7775	0.1101	271.0727	0.9558	0.1882	72.8114	0.9074	0.0337	198.0704	0.9220	0.0002	214.7371	0.9202	0.0002
			6	24.6391	0.7600	0.4802	19.6650	0.7932	0.1517	291.8265	0.9548	0.1811	47.0702	0.8983	0.0412	166.2818	0.9427	0.0002	195.5132	0.9434	0.0002
	Soft	Universal	5	26.0574	0.7250	0.4669	7.4475	0.6636	0.2289	267.9706	0.9549	0.1893	33.1572	0.8706	0.0493	139.4033	0.9200	0.0002	142.4853	0.9168	0.0003
			6	12.8509	0.7143	0.6422	3.5059	0.6032	0.3047	267.9607	0.9531	0.1887	17.0146	0.8432	0.0661	48.9434	0.9332	0.0003	53.5314	0.9309	0.0005
	Garotte	Universal	5	28.3341	0.7316	0.4489	14.1162	0.7184	0.1744	268.2973	0.9549	0.1892	42.4920	0.8841	0.0437	164.9758	0.9206	0.0002	180.4965	0.9182	0.0003
			6	15.3871	0.7288	0.5919	6.5146	0.6818	0.2393	270.1197	0.9532	0.1880	24.2176	0.8671	0.0560	96.8012	0.9392	0.0002	109.2636	0.9371	0.0003
	Hard	Heur. SURE	5	73.3307	0.7855	0.2847	9.4641	0.6671	0.2106	276.0162	0.9490	0.1872	21.4388	0.8459	0.0610	93.1264	0.9189	0.0002	77.3256	0.9140	0.0004
			6	77.2462	0.8177	0.2770	8.5795	0.7072	0.2161	374.7538	0.9557	0.1615	8.6592	0.7853	0.0914	8.2940	0.9014	0.0007	5.8051	0.8924	0.0012
	Hard	MINIMAX	5	84.0595	0.7918	0.2677	80.4162	0.8038	0.0791	250.4107	0.9334	0.1983	120.5231	0.9197	0.0264	197.1193	0.9154	0.0002	201.4585	0.9135	0.0002
			6	54.6881	0.8030	0.3292	54.6515	0.8381	0.0945	363.5897	0.9490	0.1651	94.4392	0.9228	0.0295	220.3703	0.9460	0.0002	260.6859	0.9484	0.0002
	Firm	MINIMAX	5	41.6570	0.7567	0.3742	34.9487	0.7720	0.1165	276.2110	0.9560	0.1865	72.7321	0.9062	0.0337	199.2842	0.9218	0.0002	220.1270	0.9203	0.0002
			6	25.5099	0.7656	0.4709	17.5898	0.7854	0.1581	305.8133	0.9555	0.1769	45.5987	0.8979	0.0416	159.2962	0.9455	0.0002	188.7215	0.9436	0.0003

				Signal																	
	BIOR22	SNR=50		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	151.0134	0.8851	0.1999	196.8082	0.9320	0.0509	926.0523	0.9789	0.1020	250.7820	0.9596	0.0184	484.8865	0.9756	0.0001	431.3728	0.9761	0.0002
			6	79.4510	0.8917	0.2741	68.6446	0.9307	0.0849	1039.1477	0.9805	0.0973	117.6024	0.9451	0.0264	352.9374	0.9764	0.0001	411.5234	0.9819	0.0002
	Soft	Universal	5	50.3246	0.8348	0.3396	23.7794	0.8764	0.1360	665.6600	0.9726	0.1201	68.3388	0.9258	0.0346	331.9310	0.9747	0.0001	345.7118	0.9755	0.0002
			6	24.0500	0.8143	0.4794	8.1789	0.7739	0.2144	512.6012	0.9677	0.1367	30.3908	0.8957	0.0503	106.7514	0.9700	0.0002	130.6239	0.9736	0.0003
	Garotte	Universal	5	73.6250	0.8562	0.2829	66.4503	0.9147	0.0853	723.0526	0.9746	0.1154	113.9026	0.9424	0.0270	429.8003	0.9753	0.0001	416.0130	0.9760	0.0002
			6	37.2149	0.8464	0.3919	20.8983	0.8656	0.1452	632.6386	0.9709	0.1236	53.5881	0.9216	0.0384	220.1926	0.9730	0.0002	278.0957	0.9778	0.0002
	Hard	Heur. SURE	5	87.1714	0.8659	0.2615	9.5081	0.7739	0.2096	673.2948	0.9732	0.1197	22.6322	0.8681	0.0594	117.4861	0.9719	0.0002	96.9654	0.9722	0.0003
			6	86.9797	0.8950	0.2615	8.5281	0.7625	0.2163	721.8025	0.9740	0.1163	8.7464	0.7977	0.0911	8.4263	0.9372	0.0007	5.9120	0.9373	0.0012
	Hard	MINIMAX	5	312.1474	0.9012	0.1401	399.8178	0.9391	0.0361	1062.1718	0.9747	0.0957	490.0415	0.9703	0.0132	504.4275	0.9748	0.0001	492.5760	0.9749	0.0002
			6	187.2176	0.9207	0.1797	182.0533	0.9485	0.0531	1525.5232	0.9819	0.0800	278.1575	0.9636	0.0175	471.3278	0.9805	0.0001	452.9541	0.9832	0.0002
	Firm	MINIMAX	5	145.0570	0.8834	0.2038	181.6408	0.9309	0.0530	934.2656	0.9787	0.1016	235.7614	0.9591	0.0189	439.6543	0.9759	0.0001	334.0593	0.9766	0.0002
			6	76.1504	0.8886	0.2792	62.7433	0.9280	0.0886	1004.1832	0.9795	0.0986	111.6603	0.9446	0.0271	346.4846	0.9767	0.0001	399.5213	0.9815	0.0002

				Signal																	
	BIOR22	SNR=100		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	302.5134	0.9324	0.1422	359.1325	0.9663	0.0382	1941.8538	0.9889	0.0706	506.6253	0.9766	0.0130	581.3284	0.9837	0.0001	530.6969	0.9841	0.0002
			6	128.3804	0.9256	0.2169	103.6527	0.9623	0.0696	1870.2313	0.9890	0.0721	165.2425	0.9555	0.0224	473.8570	0.9843	0.0001	464.1594	0.9864	0.0002
	Soft	Universal	5	76.7110	0.8850	0.2762	37.1811	0.9315	0.1104	1033.6593	0.9821	0.0961	101.1128	0.9456	0.0285	440.0385	0.9833	0.0001	419.3689	0.9837	0.0002
			6	32.1897	0.8495	0.4168	10.9900	0.8308	0.1881	720.0404	0.9731	0.1154	37.6373	0.9106	0.0454	145.5568	0.9774	0.0002	187.4941	0.9805	0.0002
	Garotte	Universal	5	132.5825	0.9100	0.2122	119.6250	0.9574	0.0643	1284.4093	0.9850	0.0864	200.1970	0.9624	0.0205	547.6186	0.9837	0.0001	493.9512	0.9840	0.0002
			6	55.5109	0.8848	0.3232	30.6463	0.9140	0.1214	1022.5786	0.9800	0.0974	71.5805	0.9359	0.0334	300.7634	0.9803	0.0001	370.0135	0.9841	0.0002
	Hard	Heur. SURE	5	85.7986	0.8962	0.2628	9.4540	0.8138	0.2098	758.6546	0.9768	0.1121	22.5990	0.8715	0.0594	123.1603	0.9802	0.0002	97.8830	0.9803	0.0003
			6	84.6403	0.9109	0.2643	8.4519	0.7833	0.2167	761.9088	0.9771	0.1119	8.7569	0.7998	0.0910	8.4374	0.9433	0.0007	5.9097	0.9424	0.0012
	Hard	MINIMAX	5	958.4361	0.9430	0.0811	807.2213	0.9704	0.0255	2283.4756	0.9872	0.0651	886.7644	0.9818	0.0098	731.4712	0.9840	0.0001	786.0141	0.9847	0.0001
			6	311.0608	0.9461	0.1397	274.0819	0.9719	0.0432	3235.4991	0.9923	0.0548	443.8720	0.9757	0.0138	569.7662	0.9861	0.0001	570.2064	0.9876	0.0001
	Firm	MINIMAX	5	307.8672	0.9312	0.1411	354.2006	0.9664	0.0382	1928.4188	0.9889	0.0707	459.1724	0.9760	0.0136	467.0613	0.9841	0.0001	415.9858	0.9844	0.0002
			6	120.8063	0.9239	0.2225	95.6455	0.9598	0.0720	1851.7938	0.9890	0.0723	164.5700	0.9576	0.0224	454.1600	0.9839	0.0001	476.6276	0.9865	0.0002

				Signal																	
	BIOR22	SNR=200		Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	694.6432	0.9593	0.0946	722.4458	0.9825	0.0267	3563.8823	0.9939	0.0519	806.2839	0.9829	0.0103	708.7095	0.9881	0.0001	727.2110	0.9893	0.0001
			6	242.2129	0.9505	0.1587	166.7632	0.9771	0.0556	3243.0239	0.9938	0.0545	213.4945	0.9623	0.0198	581.7004	0.9881	0.0001	510.0676	0.9893	0.0002
	Soft	Universal	5	115.1432	0.9172	0.2264	61.5627	0.9617	0.0870	1524.4889	0.9883	0.0792	139.5169	0.9572	0.0244	515.3936	0.9876	0.0001	517.6145	0.9886	0.0002
			6	48.8145	0.8791	0.3417	15.1070	0.8702	0.1639	940.8218	0.9781	0.1006	44.0760	0.9197	0.0421	211.1037	0.9829	0.0002	254.5368	0.9852	0.0002
	Garotte	Universal	5	234.4160	0.9426	0.1602	235.2659	0.9778	0.0462	2165.1882	0.9913	0.0665	309.9027	0.9726	0.0165	629.2516	0.9879	0.0001	598.7811	0.9888	0.0001
			6	96.0851	0.9166	0.2480	46.5696	0.9450	0.1004	1497.4779	0.9873	0.0799	87.5617	0.9435	0.0304	419.9459	0.9858	0.0001	441.6324	0.9880	0.0002
	Hard	Heur. SURE	5	85.5984	0.9135	0.2632	9.4880	0.8364	0.2093	783.1875	0.9786	0.1103	22.7113	0.8723	0.0593	123.4859	0.9844	0.0002	101.5968	0.9852	0.0003
			6	83.9953	0.9188	0.2654	8.4738	0.7888	0.2163	736.3482	0.9770	0.1138	8.7706	0.8003	0.0909	8.4741	0.9466	0.0007	5.9139	0.9462	0.0012
	Hard	MINIMAX	5	2129.9909	0.9661	0.0536	1469.9861	0.9836	0.0188	4434.7390	0.9931	0.0467	1450.9790	0.9873	0.0077	1090.4735	0.9891	0.0001	1277.0210	0.9902	0.0001
			6	827.7577	0.9648	0.0881	458.2951	0.9837	0.0339	5239.3475	0.9955	0.0432	585.1196	0.9807	0.0121	729.1466	0.9893	0.0001	805.3058	0.9907	0.0001
	Firm	MINIMAX	5	730.5684	0.9589	0.0916	732.1787	0.9821	0.0265	3449.0884	0.9937	0.0527	737.0599	0.9825	0.0108	578.4583	0.9883	0.0001	813.9672	0.9895	0.0001
			6	227.7371	0.9490	0.1633	148.1099	0.9759	0.0585	2989.0921	0.9936	0.0566	209.0008	0.9640	0.0199	577.3558	0.9881	0.0001	553.2546	0.9895	0.0001

Table S4. Performance of the SWT at decomposition levels 5 and 6 with the bior2,2 wavelet for 6 thresholding combinations. The tables give the performance of each combination in terms of SNR, SSIM, and RMSE in order of increasing starting SNR (10, 50, 100, 200) of the noisy test signal. Results are presented for the following test signals: blocks, bumps, heavisine, doppler, 1st EPR test signal, and 2nd EPR test signal. Method refers to the thresholding form, while selection refers to the means of calculating the threshold value.

Signal

BIOR24	SNR=10			Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	46.5057	0.7609	0.3573	48.5285	0.7775	0.1005	263.5782	0.9549	0.1910	91.3612	0.9132	0.0303	210.8588	0.9204	0.0002	219.1190	0.9178	0.0002
			6	27.5225	0.7627	0.4575	19.4970	0.7778	0.1531	309.9567	0.9560	0.1760	73.0578	0.9121	0.0336	196.5427	0.9431	0.0002	243.6321	0.9455	0.0002
	Soft	Universal	5	27.4816	0.7238	0.4570	8.4339	0.6659	0.2174	259.5529	0.9539	0.1924	39.2097	0.8778	0.0457	170.2713	0.9192	0.0002	171.8807	0.9155	0.0003
			6	13.6970	0.7119	0.6256	3.3848	0.5859	0.3103	268.3856	0.9521	0.1887	22.2265	0.8582	0.0587	67.0162	0.9348	0.0003	76.4914	0.9333	0.0004
	Garotte	Universal	5	30.5305	0.7320	0.4350	16.9048	0.7255	0.1614	259.9769	0.9540	0.1922	53.4624	0.8928	0.0394	192.8349	0.9195	0.0002	203.4115	0.9165	0.0002
			6	16.8405	0.7285	0.5695	6.2503	0.6656	0.2445	274.0844	0.9525	0.1867	34.7351	0.8826	0.0476	131.1982	0.9399	0.0002	157.3984	0.9399	0.0003
	Hard	Heur. SURE	5	74.0756	0.7845	0.2836	9.8289	0.6630	0.2079	273.1135	0.9504	0.1881	21.6943	0.8452	0.0609	111.4474	0.9186	0.0002	94.4266	0.9135	0.0004
			6	78.3572	0.8167	0.2753	8.9718	0.7063	0.2128	377.5119	0.9578	0.1605	8.8492	0.7835	0.0913	9.1445	0.9004	0.0007	6.4515	0.8861	0.0012
	Hard	MINIMAX	5	91.3278	0.7918	0.2584	92.7938	0.8001	0.0741	222.9178	0.9194	0.2116	138.8789	0.9198	0.0248	182.9861	0.9091	0.0002	193.7320	0.9088	0.0002
			6	58.4856	0.8032	0.3193	56.5581	0.8265	0.0935	314.9986	0.9315	0.1804	130.6104	0.9296	0.0254	230.4127	0.9437	0.0002	270.7126	0.9462	0.0002
	Firm	MINIMAX	5	45.9741	0.7590	0.3583	42.6750	0.7748	0.1065	269.0971	0.9546	0.1890	89.5800	0.9120	0.0306	206.1864	0.9200	0.0002	216.8272	0.9178	0.0002
			6	28.4869	0.7662	0.4489	17.2957	0.7715	0.1602	317.4938	0.9552	0.1738	68.2772	0.9099	0.0346	177.7018	0.9450	0.0002	237.3486	0.9452	0.0002

Signal

BIOR24	SNR=50			Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	162.3575	0.8830	0.1935	237.1333	0.9281	0.0467	998.4312	0.9792	0.0981	380.3219	0.9661	0.0150	498.8530	0.9751	0.0001	439.8185	0.9755	0.0002
			6	73.9978	0.8819	0.2860	61.3896	0.9141	0.0894	1127.1018	0.9809	0.0931	179.6432	0.9523	0.0217	437.5078	0.9782	0.0001	445.8799	0.9826	0.0002
	Soft	Universal	5	53.3718	0.8318	0.3310	27.8743	0.8780	0.1267	689.4479	0.9733	0.1180	91.2508	0.9348	0.0302	413.0300	0.9748	0.0001	408.3364	0.9753	0.0002
			6	23.6750	0.8035	0.4854	6.9648	0.7367	0.2295	549.7399	0.9681	0.1322	43.3192	0.9114	0.0428	145.2064	0.9717	0.0002	191.3154	0.9762	0.0002
	Garotte	Universal	5	79.1023	0.8547	0.2739	82.7753	0.9152	0.0772	764.1955	0.9756	0.1123	167.0381	0.9504	0.0225	484.1802	0.9751	0.0001	441.7916	0.9755	0.0002
			6	36.0533	0.8360	0.4004	17.2990	0.8337	0.1578	697.7693	0.9724	0.1180	84.0135	0.9352	0.0312	291.2591	0.9747	0.0001	367.3591	0.9801	0.0002
	Hard	Heur. SURE	5	92.0969	0.8639	0.2547	10.0116	0.7706	0.2056	671.1148	0.9730	0.1199	22.9789	0.8671	0.0592	148.4501	0.9726	0.0002	124.8375	0.9727	0.0003
			6	91.6669	0.8914	0.2550	9.0417	0.7584	0.2116	746.6795	0.9745	0.1141	8.9436	0.7964	0.0910	9.2966	0.9345	0.0007	6.5690	0.9310	0.0012
	Hard	MINIMAX	5	314.8996	0.8966	0.1396	447.4117	0.9343	0.0341	981.9046	0.9690	0.0998	605.3142	0.9732	0.0119	509.3342	0.9738	0.0001	500.6672	0.9742	0.0002
			6	173.8049	0.9118	0.1875	164.6313	0.9352	0.0559	1493.6221	0.9800	0.0808	454.5650	0.9713	0.0137	508.8838	0.9807	0.0001	466.8390	0.9834	0.0002
	Firm	MINIMAX	5	155.2147	0.8817	0.1976	223.2248	0.9278	0.0481	977.4594	0.9790	0.0993	348.6904	0.9650	0.0157	404.3591	0.9752	0.0001	316.5486	0.9756	0.0002
			6	71.7986	0.8794	0.2895	53.3518	0.9089	0.0956	1096.7996	0.9802	0.0943	184.6406	0.9544	0.0214	423.2466	0.9782	0.0001	446.5288	0.9825	0.0002

Signal

BIOR24	SNR=100			Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	316.3572	0.9302	0.1398	431.2842	0.9621	0.0350	2041.8418	0.9892	0.0690	749.1271	0.9803	0.0107	601.2888	0.9835	0.0001	560.9492	0.9840	0.0001
			6	117.8724	0.9157	0.2273	78.3269	0.9411	0.0794	1994.8557	0.9895	0.0696	293.2134	0.9663	0.0171	557.2437	0.9856	0.0001	501.8069	0.9870	0.0002
	Soft	Universal	5	81.4734	0.8844	0.2688	45.6103	0.9331	0.1008	1098.5701	0.9826	0.0933	145.2900	0.9551	0.0240	529.6500	0.9834	0.0001	493.5934	0.9838	0.0002
			6	31.1940	0.8377	0.4250	8.6091	0.7844	0.2091	765.4946	0.9738	0.1120	54.2388	0.9250	0.0384	196.9947	0.9793	0.0002	268.8277	0.9827	0.0002
	Garotte	Universal	5	142.2422	0.9096	0.2055	157.4018	0.9566	0.0568	1388.2620	0.9857	0.0832	323.8502	0.9703	0.0162	592.1742	0.9836	0.0001	531.8295	0.9840	0.0002
			6	52.4701	0.8736	0.3338	22.9949	0.8774	0.1384	1104.6980	0.9815	0.0937	114.7821	0.9475	0.0268	392.1839	0.9821	0.0001	455.8287	0.9856	0.0002
	Hard	Heur. SURE	5	92.3863	0.8963	0.2534	9.9014	0.8119	0.2063	800.7370	0.9770	0.1092	22.9548	0.8709	0.0592	157.1476	0.9809	0.0002	127.1961	0.9810	0.0003
			6	91.0072	0.9073	0.2550	8.9159	0.7800	0.2127	820.3453	0.9782	0.1081	8.9564	0.7986	0.0909	9.3152	0.9409	0.0007	6.5677	0.9360	0.0012
	Hard	MINIMAX	5	917.9456	0.9385	0.0827	888.9350	0.9674	0.0243	2179.6339	0.9852	0.0664	1174.1633	0.9846	0.0086	770.9487	0.9837	0.0001	804.4988	0.9840	0.0001
			6	279.6157	0.9374	0.1480	239.5332	0.9582	0.0463	3227.1873	0.9917	0.0548	704.3728	0.9804	0.0111	599.0050	0.9863	0.0001	617.2707	0.9879	0.0001
	Firm	MINIMAX	5	322.2700	0.9291	0.1383	451.3934	0.9633	0.0341	2042.4828	0.9893	0.0688	709.8079	0.9801	0.0110	438.2359	0.9836	0.0001	580.0236	0.9843	0.0001
			6	111.1729	0.9142	0.2331	73.1108	0.9389	0.0822	1963.2975	0.9896	0.0701	282.7801	0.9670	0.0173	531.0351	0.9850	0.0001	524.7500	0.9872	0.0002

Signal

BIOR24	SNR=200			Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	697.7270	0.9560	0.0954	855.2380	0.9794	0.0247	3717.0172	0.9938	0.0509	1115.5558	0.9846	0.0088	738.3267	0.9881	0.0001	824.2442	0.9894	0.0001
			6	202.7430	0.9387	0.1738	90.2786	0.9550	0.0743	3532.3055	0.9938	0.0523	392.3793	0.9735	0.0148	626.0736	0.9886	0.0001	559.3632	0.9901	0.0001
	Soft	Universal	5	120.6366	0.9162	0.2219	73.0066	0.9604	0.0807	1688.5348	0.9890	0.0754	205.8278	0.9651	0.0202	603.9420	0.9878	0.0001	598.7140	0.9888	0.0001
			6	44.5778	0.8653	0.3575	9.3598	0.8027	0.2016	1030.6722	0.9792	0.0961	63.4949	0.9326	0.0356	286.4301	0.9846	0.0001	350.4759	0.9873	0.0002
	Garotte	Universal	5	244.8698	0.9413	0.1573	295.7592	0.9751	0.0417	2429.3543	0.9918	0.0629	510.7738	0.9781	0.0129	675.2571	0.9880	0.0001	657.4185	0.9890	0.0001
			6	83.8600	0.9032	0.2653	25.7977	0.8954	0.1312	1683.1395	0.9882	0.0753	140.6621	0.9539	0.0243	524.9386	0.9870	0.0001	510.9414	0.9893	0.0002
	Hard	Heur. SURE	5	92.8947	0.9138	0.2530	10.0963	0.8341	0.2045	842.9094	0.9797	0.1067	23.0640	0.8715	0.0591	157.9007	0.9852	0.0002	132.8477	0.9859	0.0003
			6	90.8215	0.9141	0.2556	9.0910	0.7880	0.2109	805.9618	0.9789	0.1093	8.9713	0.7992	0.0908	9.3527	0.9440	0.0007	6.5721	0.9399	0.0012
	Hard	MINIMAX	5	1958.3518	0.9633	0.0559	1637.5570	0.9821	0.0179	4087.3566	0.9907	0.0488	1910.6945	0.9893	0.0067	1140.6878	0.9890	0.0001	1326.1252	0.9900	0.0001
			6	579.8509	0.9584	0.1064	267.6269	0.9690	0.0437	5259.8157	0.9952	0.0431	898.8342	0.9844	0.0098	812.4889	0.9895	0.0001	926.6422	0.9911	0.0001
	Firm	MINIMAX	5	742.1036	0.9565	0.0913	873.3774	0.9793	0.0245	3698.0715	0.9937	0.0510	1117.2395	0.9852	0.0088	684.8821	0.9883	0.0001	905.9233	0.9896	0.0001
			6	189.8985	0.9388	0.1787	84.0091	0.9523	0.0767	3268.4399	0.9937	0.0542	356.5246	0.9725	0.0154	645.8709	0.9887	0.0001	613.7425	0.9903	0.0001

Table S5. Performance of the SWT at decomposition levels 5 and 6 with the bior2,4 wavelet for 6 thresholding combinations. The tables give the performance of each combination in terms of SNR, SSIM, and RMSE in order of increasing starting SNR (10, 50, 100, 200) of the noisy test signal. Results are presented for the following test signals: blocks, bumps, heavisine, doppler, 1st EPR test signal, and 2nd EPR test signal. Method refers to the thresholding form, while selection refers to the means of calculating the threshold value.

Signal

BIOR26	SNR=10			Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	46.6752	0.7593	0.3572	48.6925	0.7729	0.1007	258.7202	0.9544	0.1928	99.3534	0.9151	0.0292	211.0833	0.9193	0.0002	217.4147	0.9162	0.0002
			6	26.2913	0.7577	0.4672	16.8658	0.7540	0.1640	307.1511	0.9559	0.1769	83.7679	0.9149	0.0315	209.9651	0.9432	0.0002	259.5044	0.9460	0.0002
	Soft	Universal	5	27.5220	0.7216	0.4576	8.3901	0.6580	0.2184	254.9443	0.9533	0.1941	42.3314	0.8808	0.0441	180.0375	0.9185	0.0002	181.1171	0.9145	0.0003
			6	13.4892	0.7070	0.6310	3.0698	0.5648	0.3237	265.4175	0.9513	0.1898	24.6078	0.8629	0.0562	74.3442	0.9349	0.0003	88.4703	0.9343	0.0004
	Garotte	Universal	5	30.5539	0.7297	0.4357	16.8613	0.7198	0.1620	255.3111	0.9534	0.1940	59.1884	0.8962	0.0376	199.6378	0.9187	0.0002	208.1547	0.9153	0.0002
			6	16.5163	0.7233	0.5752	5.4199	0.6403	0.2597	271.4189	0.9517	0.1877	39.6821	0.8870	0.0449	144.7322	0.9399	0.0002	180.3623	0.9410	0.0003
	Hard	Heur. SURE	5	73.8510	0.7836	0.2840	9.9973	0.6591	0.2069	265.6862	0.9489	0.1906	21.6536	0.8442	0.0611	118.9662	0.9184	0.0002	101.6289	0.9129	0.0003
			6	77.8029	0.8150	0.2762	9.1126	0.6968	0.2122	364.7661	0.9559	0.1629	8.8749	0.7820	0.0916	9.4013	0.8948	0.0007	6.6813	0.8799	0.0012
	Hard	MINIMAX	5	89.0713	0.7881	0.2614	92.0422	0.7949	0.0745	217.2932	0.9164	0.2144	143.7177	0.9173	0.0244	178.7654	0.9077	0.0002	190.4567	0.9070	0.0003
			6	54.9703	0.7985	0.3288	50.6026	0.8115	0.0988	307.8142	0.9300	0.1819	142.0037	0.9306	0.0244	233.9031	0.9427	0.0002	269.7234	0.9447	0.0002
	Firm	MINIMAX	5	45.8852	0.7570	0.3591	42.8962	0.7701	0.1066	264.6277	0.9538	0.1906	97.4025	0.9140	0.0295	204.4299	0.9189	0.0002	211.8493	0.9161	0.0002
			6	27.4100	0.7613	0.4570	14.8464	0.7498	0.1717	315.2443	0.9551	0.1745	77.6126	0.9127	0.0326	204.1519	0.9437	0.0002	253.4129	0.9456	0.0002

Signal

BIOR26	SNR=50			Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	155.8103	0.8788	0.1976	231.1077	0.9214	0.0473	991.3281	0.9789	0.0985	434.5537	0.9677	0.0141	497.9014	0.9748	0.0001	436.8189	0.9751	0.0002
			6	62.1347	0.8699	0.3129	42.1042	0.8787	0.1073	1124.6757	0.9806	0.0932	222.3264	0.9560	0.0197	466.8725	0.9788	0.0001	449.3501	0.9826	0.0002
	Soft	Universal	5	51.9546	0.8271	0.3359	27.8006	0.8708	0.1269	687.0793	0.9730	0.1182	102.9697	0.9379	0.0285	439.7567	0.9746	0.0001	424.6085	0.9750	0.0002
			6	21.6797	0.7899	0.5076	5.1562	0.6866	0.2606	554.6952	0.9679	0.1316	51.0346	0.9171	0.0397	166.0232	0.9725	0.0002	218.3386	0.9771	0.0002
	Garotte	Universal	5	76.0754	0.8503	0.2797	83.1164	0.9091	0.0770	764.2064	0.9754	0.1123	194.9379	0.9531	0.0209	495.4483	0.9749	0.0001	443.2941	0.9752	0.0002
			6	31.9353	0.8220	0.4259	11.4858	0.7802	0.1889	707.1879	0.9725	0.1173	102.3821	0.9394	0.0285	329.2534	0.9755	0.0001	395.1867	0.9808	0.0002
	Hard	Heur. SURE	5	95.0584	0.8628	0.2509	10.1181	0.7669	0.2052	676.6237	0.9728	0.1195	22.9581	0.8659	0.0594	162.4887	0.9727	0.0002	137.5213	0.9727	0.0003
			6	94.2569	0.8872	0.2517	9.1273	0.7430	0.2116	764.6798	0.9746	0.1131	8.9733	0.7951	0.0912	9.5589	0.9287	0.0007	6.8000	0.9248	0.0012
	Hard	MINIMAX	5	293.2406	0.8919	0.1447	432.0617	0.9293	0.0348	958.8900	0.9677	0.1010	627.8837	0.9734	0.0117	503.1900	0.9732	0.0001	496.5145	0.9735	0.0002
			6	149.0472	0.9027	0.2035	113.2635	0.9140	0.0671	1460.3418	0.9791	0.0818	539.7529	0.9733	0.0126	516.5735	0.9807	0.0001	467.1842	0.9833	0.0002
	Firm	MINIMAX	5	148.3976	0.8780	0.2023	220.7731	0.9213	0.0484	972.1746	0.9788	0.0995	401.6730	0.9667	0.0146	389.8671	0.9748	0.0001	306.8108	0.9751	0.0002
			6	61.6784	0.8668	0.3134	35.4075	0.8710	0.1159	1101.7092	0.9798	0.0941	228.4886	0.9577	0.0194	453.8732	0.9787	0.0001	452.8321	0.9827	0.0002

Signal

BIOR26	SNR=100			Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	302.9551	0.9260	0.1427	418.6107	0.9558	0.0358	2029.8474	0.9891	0.0691	830.4000	0.9807	0.0102	601.7693	0.9834	0.0001	566.9782	0.9839	0.0001
			6	97.7455	0.9033	0.2492	48.5016	0.9023	0.0997	2015.2432	0.9896	0.0692	383.4721	0.9705	0.0150	574.8497	0.9858	0.0001	507.5687	0.9871	0.0002
	Soft	Universal	5	79.4055	0.8816	0.2725	45.3480	0.9262	0.1015	1101.8452	0.9824	0.0931	169.9817	0.9584	0.0222	555.1471	0.9833	0.0001	511.6007	0.9837	0.0002
			6	28.0015	0.8235	0.4476	5.6184	0.7181	0.2505	767.3933	0.9736	0.1120	64.7796	0.9306	0.0353	223.3263	0.9801	0.0002	304.5431	0.9834	0.0002
	Garotte	Universal	5	137.2276	0.9068	0.2092	156.4542	0.9501	0.0572	1394.5120	0.9856	0.0830	393.7634	0.9725	0.0147	599.1075	0.9834	0.0001	536.5783	0.9838	0.0001
			6	45.1499	0.8592	0.3587	12.9659	0.8129	0.1780	1107.1479	0.9815	0.0937	143.7524	0.9519	0.0241	433.9684	0.9829	0.0001	479.1500	0.9860	0.0002
	Hard	Heur. SURE	5	94.8206	0.8963	0.2502	10.0288	0.8063	0.2058	816.5070	0.9769	0.1083	22.9428	0.8699	0.0594	172.7061	0.9811	0.0002	140.9057	0.9812	0.0003
			6	93.0020	0.9029	0.2524	9.0363	0.7646	0.2124	845.7042	0.9786	0.1066	8.9874	0.7974	0.0911	9.5793	0.9350	0.0007	6.8000	0.9298	0.0012
	Hard	MINIMAX	5	814.9282	0.9348	0.0877	842.5343	0.9643	0.0250	2124.9382	0.9843	0.0673	1228.5398	0.9851	0.0084	765.8856	0.9835	0.0001	804.7227	0.9836	0.0001
			6	240.8853	0.9291	0.1593	130.4695	0.9335	0.0626	3167.5023	0.9914	0.0553	798.2082	0.9811	0.0104	607.1773	0.9864	0.0001	628.0676	0.9878	0.0001
	Firm	MINIMAX	5	303.8457	0.9258	0.1423	434.6282	0.9574	0.0348	2030.6721	0.9891	0.0690	817.3358	0.9811	0.0103	428.8261	0.9833	0.0001	547.7358	0.9840	0.0002
			6	92.5220	0.9017	0.2548	40.6069	0.8966	0.1079	1949.3939	0.9894	0.0704	360.6326	0.9704	0.0154	556.0264	0.9854	0.0001	533.4063	0.9873	0.0002

Signal

BIOR26	SNR=200			Blocks			Bumps			Heavisine			Doppler			Epr 1			Epr 2		
	Method	Selection	Level	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE	SNR	SSIM	RMSE
SWT	Hard	Universal	5	570.4563	0.9515	0.1045	784.3059	0.9756	0.0259	3693.5765	0.9934	0.0510	1205.9519	0.9847	0.0085	740.2639	0.9881	0.0001	829.9270	0.9893	0.0001
			6	131.2387	0.9210	0.2149	49.3317	0.9107	0.0988	3586.7497	0.9937	0.0519	528.4988	0.9772	0.0128	635.0006	0.9886	0.0001	565.5057	0.9901	0.0001
	Soft	Universal	5	114.8077	0.9129	0.2274	72.1631	0.9545	0.0813	1691.2403	0.9887	0.0753	246.8249	0.9679	0.0185	629.6063	0.9878	0.0001	616.1310	0.9888	0.0001
			6	33.9541	0.8412	0.4081	5.5683	0.7244	0.2513	1044.6376	0.9791	0.0954	77.8314	0.9385	0.0323	322.4789	0.9852	0.0001	387.3388	0.9879	0.0002
	Garotte	Universal	5	226.1353	0.9377	0.1634	289.1478	0.9696	0.0422	2431.6438	0.9916	0.0629	629.4125	0.9796	0.0117	682.1657	0.9879	0.0001	666.5885	0.9890	0.0001
			6	58.1641	0.8786	0.3172	12.8103	0.8196	0.1788	1713.2062	0.9882	0.0746	183.2220	0.9589	0.0214	563.3187	0.9875	0.0001	526.5293	0.9896	0.0002
	Hard	Heur. SURE	5	95.7539	0.9137	0.2493	10.1883	0.8263	0.2042	862.9306	0.9798	0.1058	23.0485	0.8706	0.0592	174.0262	0.9855	0.0002	147.3062	0.9861	0.0003
			6	93.0933	0.9101	0.2525	9.1477	0.7719	0.2111	831.1565	0.9794	0.1080	9.0027	0.7980	0.0910	9.6154	0.9381	0.0007	6.8042	0.9336	0.0012
	Hard	MINIMAX	5	1750.1691	0.9608	0.0592	1580.6431	0.9803	0.0182	4014.1880	0.9904	0.0492	2094.5528	0.9900	0.0064	1133.6803	0.9889	0.0001	1326.3942	0.9897	0.0001
			6	326.7668	0.9478	0.1369	134.5106	0.9413	0.0614	5125.7262	0.9946	0.0436	1048.5716	0.9853	0.0091	831.0163	0.9896	0.0001	942.4059	0.9911	0.0001
	Firm	MINIMAX	5	660.3647	0.9535	0.0966	826.3895	0.9757	0.0252	3656.4594	0.9935	0.0513	1283.8979	0.9859	0.0082	718.8349	0.9882	0.0001	916.1859	0.9895	0.0001
			6	125.5106	0.9207	0.2194	40.6155	0.9053	0.1078	3289.4069	0.9935	0.0540	470.5048	0.9758	0.0135	664.1622	0.9888	0.0001	621.5024	0.9904	0.0001

Table S6. Performance of the SWT at decomposition levels 5 and 6 with the bior2,6 wavelet for 6 thresholding combinations. The tables give the performance of each combination in terms of SNR, SSIM, and RMSE in order of increasing starting SNR (10, 50, 100, 200) of the noisy test signal. Results are presented for the following test signals: blocks, bumps, heavisine, doppler, 1st EPR test signal, and 2nd EPR test signal. Method refers to the thresholding form, while selection refers to the means of calculating the threshold value.

Chapter 4: Accessing BtuB hatch-barrel information in the native system through modulation of the *E. coli* Dsb system and DEER spectroscopy.

4.1 Introduction

As with other TonB dependent transporters, the *E. coli* cobalamin transporter BtuB must coordinate with the trans-periplasmic protein TonB to move cobalamin across the outer membrane. The size of the cobalamin substrate means that substantial rearrangement is likely to be required in the hatch domain of BtuB before a sufficiently large translocation channel is formed. Additionally, the loops of BtuB have been observed to be in dynamic conformational equilibria, which can be shifted in the presence and absence of calcium, substrate, or TonB (1). To function productively, these motions must also be coordinated allosterically to TonB binding events on the periplasmic face. Thus, it is critical to identify methods with which conformational changes can be determined both in the extracellular loops and the hatch domain.

Distance measurements using pulsed-EPR are an attractive means of obtaining this information. DEER requires much smaller labels than comparable fluorescence techniques while providing sufficient resolution to identify even small conformational shifts. Additionally, recent work in our lab had involved the translation of EPR techniques into the native environments of outer membrane isolates and intact cells. For this project, the aim was to develop a strategy to observe changes between a relatively static site in the barrel, and a potentially mobile one in the hatch.

To do this, site 188 in the extracellular loops was selected as the initial barrel residue. This site had previously been characterized in isolated outer membranes and appeared to label in intact cells (1, 2). DEER data was also available in both systems from site 188 to site 399 in an opposing loop. In Outer membranes, this pair underwent a disorder-to-order transition with the addition of calcium and displayed a change in width but not position with the subsequent addition of substrate (1). *In vivo*, calcium will always be present, and so this pair appeared to present an attractive target as semi-static reference sites in the barrel.

In the hatch domain, site 90 was selected as the initial site as it had also been shown to label in the same intact cell experiments (2). Further, it occupies the most extracellular position in the hatch domain, is near the substrate, and points upward into the extracellular environment, hopefully increasing its accessibility towards the label. Having selected the sites, the 90-188 pair, which is shown in Figure 1, was developed both in the OM and whole cell systems. More details of how the current protocols presented in this work were obtained are available in the recent thesis of Thushani Nilaweera. Here, a brief summary of the method's development is presented in non-chronological order with a subsequent focus on its application to conformational changes observed in BtuB.

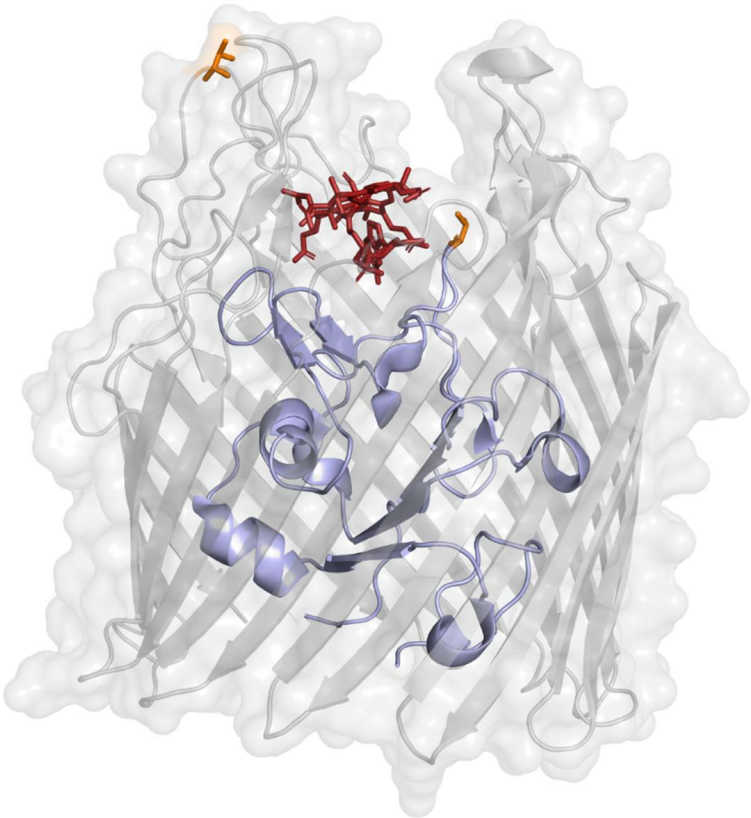


Figure 1. A side view of BtuB showing the sites 188 and 90 (orange sticks) in the extracellular loops and upper hatch, respectively. The cobalamin substrate is shown in dark red. This pair was used to check for movement of the upper hatch region in response to substrate binding.

4.2 Additional Methods Information by Figure

Fig. 2 Growth and isolation of OMs from RK5016 cells overexpressing BtuB 90C-188C proceeded as written in Chapter 2.

Fig. 4 Growth of RK5016 cells overexpressing BtuB 90C or BtuB 188C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 1 hr with 1 mg of MTSL and was followed by 2x 30-minute wash steps on ice with 100 mM HEPES, pH 7.0 and 100 mM MES, pH 5.5, respectively.

Fig. 5 (A, B) Growth of RK5016 cells overexpressing BtuB 90C-188C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0 with or without 100 mM DTT. Incubation with DTT proceeded for 10 minutes at 37° C after which Labeling occurred for 1 hr with 3 mg of MTSL and was followed by 2x 30-minute wash steps on ice with 100 mM HEPES, pH 7.0 and 100 mM MES, pH 5.5, respectively. **(C, D)** Growth of RK5016 cells overexpressing BtuB 90C-188C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 1 hr with 0.1 or 0.2 mg of MTSL and was followed by 2x 30-minute wash steps on ice with 100 mM HEPES, pH 7.0 and 100 mM MES, pH 5.5, respectively.

Fig. 6 Growth of RK5016 cells overexpressing BtuB 90C-188C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 1 hr with 3 mg of MTSL and was followed by 2x 30-minute wash steps on ice with 100 mM HEPES, pH 7.0 and 100 mM MES, pH 5.5, respectively.

Fig. 7 Growth of DsbA$^-$ cells overexpressing BtuB 90C-188C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 1 hr with 0.05, 0.1, 0.25, or 0.5 mg of MTSL and was followed by 2x 30-minute wash steps on ice with 100 mM HEPES, pH 7.0 and 100 mM MES, pH 5.5, respectively.

Fig. 8 Growth of DsbA$^-$ cells overexpressing BtuB 90C-188C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 1 hr with 0.1 mg of MTSL and

was followed by 2x 30-minute wash steps at RT with 100 mM HEPES, pH 7.0 and 100 mM MES, pH 5.5, respectively. The +cobalamin condition used 100 μM B_{12}.

Fig. 9 Growth and isolation of OMs from RK5016 cells overexpressing BtuB 90C-188C proceeded as written in Chapter 2. 100 μM cobalamin was added for the + cobalamin condition.

Fig. 10 (A-D) Growth of DsbA$^-$ cells overexpressing WT BtuB to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 1 hr with 0.2 mg of MTSL and was followed by 2x 30-minute wash steps on ice with 100 mM HEPES, pH 7.0 and 100 mM MES, pH 5.5, respectively.

(E-H) Growth of DsbA$^-$ cells overexpressing BtuB 90C-188C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 1 hr with 0.1 mg of MTSL and was followed by 2x 30-minute wash steps at RT with 100 mM HEPES, pH 7.0 and 100 mM MES, pH 5.5, respectively.

Fig. 11 Growth of DsbA$^-$ cells overexpressing BtuB 90C-188C was the same as in Fig. 8**.** Growth of DsbB$^-$ cells overexpressing BtuB 90C-188C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 1 hr with 0.1 mg of MTSL and was followed by 2x 30-minute wash steps on ice with 100 mM HEPES, pH 7.0 and 100 mM MES, pH 5.5, respectively.

Fig. 12 Growth of DsbC$^-$ cells overexpressing BtuB 90C-188C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 1 hr with 0.1 mg of MTSL and was followed by 2x 30-minute wash steps on ice with 100 mM HEPES, pH 7.0 and 100 mM MES, pH 5.5, respectively.

Fig. 13 Growth of DsbA$^-$ cells overexpressing BtuB 90C-188C was the same as in Fig. 8**.** 0, 5, 10, 20, 30, 60, or 100 μM B_{12} was added to + substrate samples.

Fig. 14 Growth of DsbA$^-$ cells overexpressing BtuB 90C-188C was the same as in Fig. 8**.** Growth of DsbA$^-$ cells overexpressing BtuB L8P-90C-188C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 1 hr with 0.1 mg of MTSL and was followed by

2x 30-minute wash steps at RT with 100 mM HEPES, pH 7.0 and 100 mM MES, pH 5.5, respectively. The +cobalamin condition used 100 μM B_{12}.

Fig. 16 Growth of DsbA$^-$ cells overexpressing BtuB 188C-399C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 1 hr with 0.1 mg of MTSL and was followed by 2x 30-minute wash steps on ice with 100 mM HEPES, pH 7.0 and 100 mM MES, pH 5.5, respectively.

Fig. 17 Growth of DsbA$^-$ cells overexpressing BtuB 188C-399C was the same as in Fig. 16. The +cobalamin condition used 100 μM B_{12}.

Fig. 19 Growth of DsbA$^-$ cells overexpressing BtuB 74C-188C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 30 minutes with 0.1 mg of MTSL and was followed by 2x 15-minute wash steps at RT with 100 mM HEPES, pH 7.0. All buffers now contained 2.5% glucose.

Fig. 20 Growth of DsbA$^-$ cells overexpressing BtuB 74C-188C was the same in Fig. 19. The +cobalamin condition used 100 μM B_{12}.

Fig. 21 Growth of DsbA$^-$ cells overexpressing BtuB 90C-188C or L8P-90C-188C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 30 minutes with 0.1 mg of MTSL and was followed by 2x 15-minute wash steps on ice with 100 mM HEPES, pH 7.0. All buffers now contained 2.5% glucose. The +cobalamin condition used 100 μM B_{12}.

Fig. 23 Growth and isolation of OMs from RK5016 cells overexpressing BtuB 6C-510C proceeded as written in Chapter 2.

Fig. 24 Growth and isolation of OMs from RK5016 cells overexpressing BtuB 6C-510C proceeded as written in Chapter 2. The +cobalamin condition used 32 μM B_{12}.

Fig. 25 Growth of DsbA$^-$ cells overexpressing BtuB 6C-510C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 30 minutes with 0.1 mg of MTSL and was followed by 1x minimum time wash step at RT with 100 mM HEPES, pH 7.0. All buffers now contained 2.5% glucose. The +cobalamin condition used 100 μM B_{12}.

Results:

Development of a method for labeling hatch-barrel double mutants in whole cells:

As indicated above, sites 90 and 188 had been previously shown to label in OM isolates and apparently intact cells. Still, they had not been combined and used in pulsed-EPR experiments. So, initial work involved acquiring DEER data for the pair in OMs and characterization of the single-site mutants in RK5016 cells. The 90-188 pair was successfully isolated in OMs, and the resulting DEER spectra are shown in Figure 2.

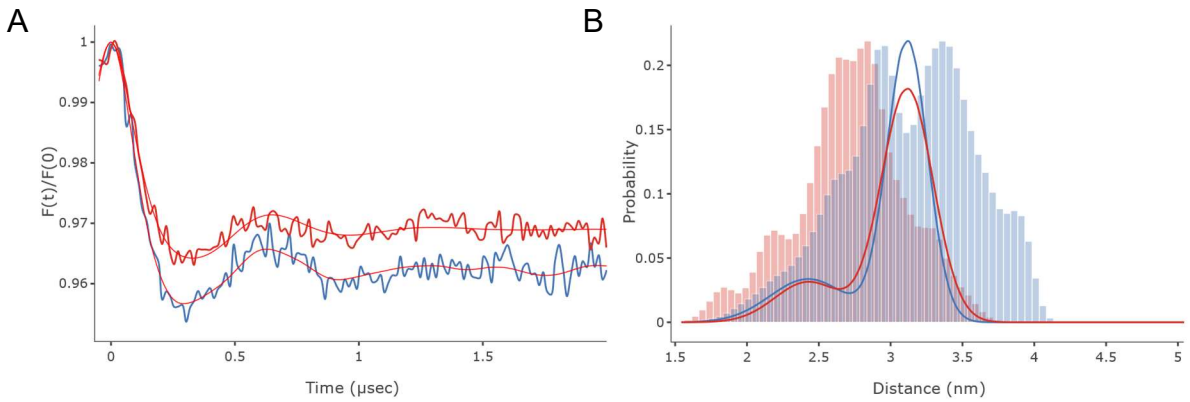


Figure 2. The (A) dipolar evolution, and (B) Distance distribution for the BtuB 90-188 pair in RK5016 Outer Membranes. For the time domain and dipolar evolution data, the apo condition is shown in blue, and a +100 μM cobalamin condition is shown in red. For the distance data, the results of 2-Gaussian fitting with a quadratic background function in the program LongDistances are shown. Two peaks are observed in the distribution, with a longer component at 3.2 nm and a shorter component at 2.4 nm. The histograms give the predicted distances from the apo (PDB ID: 1NQG) and substrate-bound (PDB ID:1NQH) crystal structures using the program MMM. The longer distance aligns with the apo prediction, while the shorter distance aligns with +substrate.

The DEER traces were cut back from their original length to a set time of 2 μs and subtracted using a quadratic background function in the proGram LongDistances. The distances were then extracted using 2-Gaussian fitting to the dipolar evolution functions. Unless otherwise noted, the other traces presented in this chapter were analyzed in the same way. The details of why this method of analysis was selected are explained in the following chapter on crowding effects in the bacterial OM. Looking at the dipolar traces shown in Fig. 2A, it is apparent that both the apo (blue) and +cobalamin (red) traces display oscillations at identical time-positions, differing only in their oscillatory amplitudes. This is borne out in the distance distributions recovered in Fig. 2B, where the positions of the two components are identical, and the increase in amplitude results in

slightly narrower features in the apo distribution.

The positional difference of the two components was 8 Å, larger than the expectation length of the MTSL spin-label used in this experiment, and thus unlikely to arise solely through rotameric rearrangement of the label. To help explain this shift, the predicted DEER distributions are shown for in-silico labels attached at sites 90 and 188 in the apo and substrate bound crystal structures of BtuB (PDB ID: 1NQG and PDB ID 1NQH). These predictions were generated using the software package MMM (3). The apo prediction (blue) shows a split distribution, whose average aligns well with the longer distance component in the DEER distribution. The +substrate prediction (red), meanwhile, displays a shorter peak centered at 2.7 nm, which was still longer than the shorter data peak. Still, the shift direction aligned with the idea that the long-distance component corresponded to the apo structure, and the short-distance component to the +substrate structure. An overlay of these structures from the front and side is shown in Figure 3.

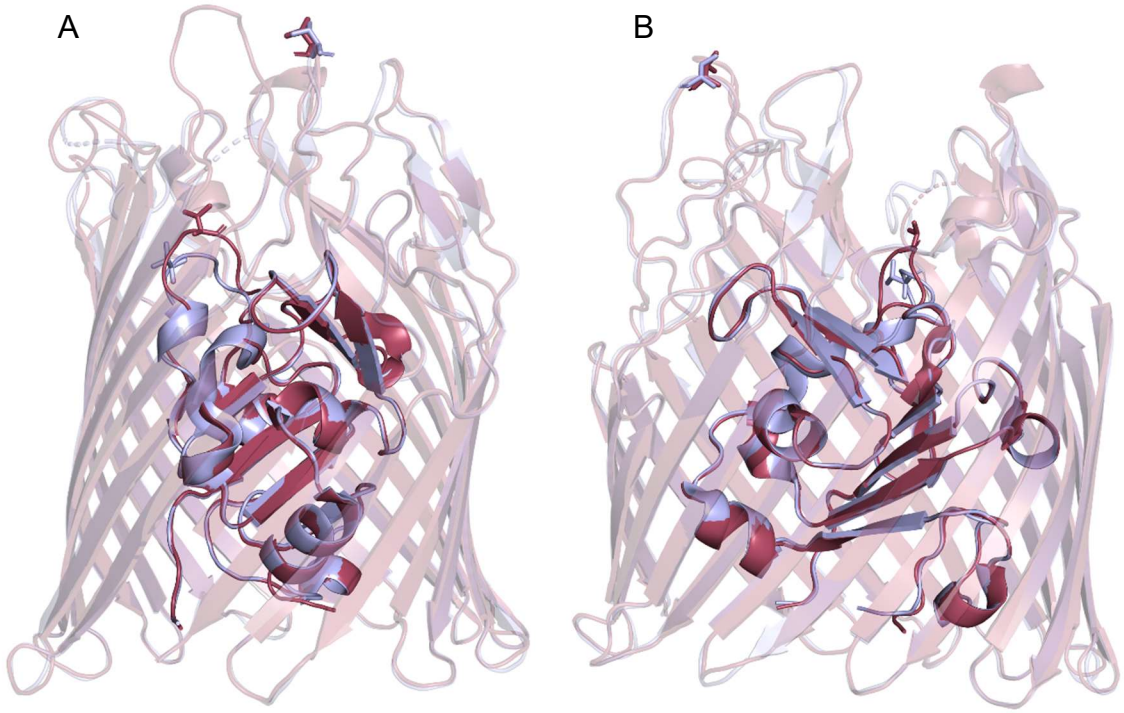


Figure 3. Front (A) and Side (B) views of the apo (PDB ID: 1NQG) and +cobalamin (PDB ID: 1NQH) crystal structures of BtuB. The apo and +substrate structures are shown in blue and raspberry, respectively. The labels used in this section at sites 90 and 188 are shown in stick representation, with 90 in the hatch domain and 188 in the extracellular loops. It is apparent from the structures that the hatch loop containing site 90 is partially helical in the apo structure, drawing it down and increasing the inter-label distance. In the +cobalamin structure, the helix becomes disordered, raising site 90 and decreasing the inter-label distance.

In both the apo (blue) and +cobalamin (raspberry) structures, the position of site 188 in the extracellular loops is invariant. In contrast, the hatch loop containing site 90 displays helical structure at its base in the apo structure. This pulls site 90 towards the hatch, increasing the observed distance for the 90-188 pair. In the +cobalamin structure, the helical element is disordered, and the loop is free to extend upward, reducing the expected interspin distance. Thus, the two components may correspond to a conformational change in a hatch element at the substrate binding site. Observing shifts in these components would then be an excellent start to mapping out the conformational environment of the BtuB hatch domain.

Unfortunately, no such shift was observed here in the OM preparations, and so the project moved into whole cells. There, the initial work involved spin-labeling single mutants of 90 and 188 for BtuB overexpressed in RK5016 cell cultures. Representative spectra obtained for these sites are shown in Figure 4. The dominant component in site 90 is highly immobile, and reflects

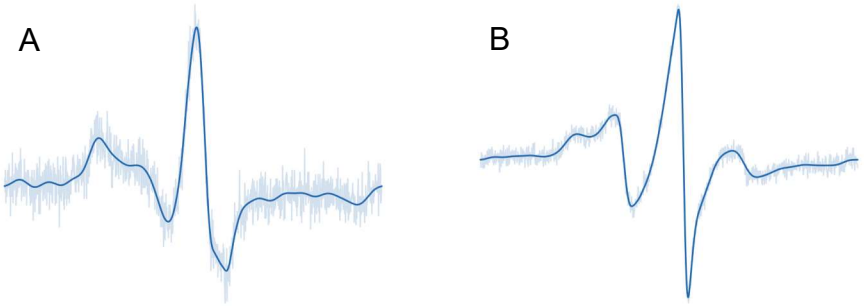


Figure 4. Representative spectra of site 90 (A) and 188 (B) in BtuB overexpressed in RK5016 cells. The cultures were grown to early log (OD 0.3) and spin labeled with 1 mg of MTSL. Spectra are not normalized.

its position in the interior hatch domain. Site 188, by contrast, shows a much more mobile lineshape that reflects its elevated position on an extracellular loop. The contrast in splitting between these elements means that both should be discriminable in composite CW spectra of the 90-188 pair. After confirming the viability of single mutants of 90 and 188 in RK 5016 cells, the double mutant was attempted several times without success. No significant labeling was observed in the RK5016 cells for the 90-188 pair despite BtuB appearing on protein gels. Additionally, the ability to produce DEER spectra of 90-188 in isolated OMs indicated that the

protein was produced and stably folded in the cells.

Something appeared to be altering the ability to label the protein in live cells. The simplest explanation was that the redox environment was different between the cells and their isolated membranes. As discussed in the introductory chapter, cysteine chemistry in the periplasm is tightly controlled by members of the Dsb system. Additionally, the periplasm had consistently been observed to reduce MTSL label after entry, indicating a significant reductive potential. These observations provided two explanations for the failure to label 90-188. First, the protein may have gained a disulfide bond between site 90 and site 188, with the formation of this bond occurring at one of three points. During translocation, disulfide bonds are nucleated by the DsbA protein, which would produce a linkage prior to protein folding. Absent disulfide bond formation at this stage, it was possible that the disulfide bond was acquired during insertion by the Bam complex, as consecutive β-hairpins are added around the central hatch domain. Finally, a sufficiently close approach distance would have permitted formation of the disulfide after insertion on the cell surface. This last possibility was unlikely, as both distance components observed in the OM DEER were too far for efficient disulfide bond formation post-insertion.

Still, all three points of formation were expected to lead to the formation of stable disulfide-bonded BtuB on the cell surface. To perturb these potential bonds, excess MTSL label was used in combination with the reducing agent DTT. The results are shown in Figure 5 A and B. Excess label produced two distinct spectral components. The major component was highly mobile, and indicative of free MTSL reagent. The second, minor component was highly immobile and may correspond to adsorption of label to the cell surface or interactions with the LPS glycans. Addition of DTT increased the percentage of the immobile component, with the shift likely indicating an increased affinity of off-target DTT-MTSL species for the surface. Neither case produced the expected lineshape for 90-188.

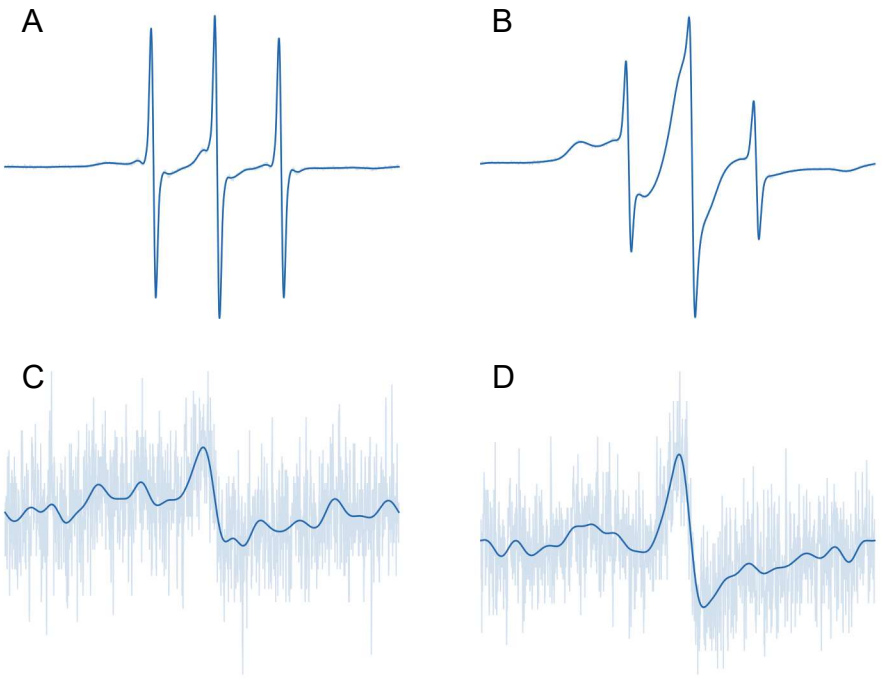


Figure 5. Results of labeling Btub 90-188 in RK5016 cells at OD 0.3 with (A) excess MTSL, (B) excess MTSL and DTT, (C) 0.1 mg MTSL, and (D) 0.2 mg MTSL. The excess condition produces a small population of immobile label that may correspond to adsorption onto the cell surface and a large population of highly mobile free MTSL. Addition of DTT causes a sharp increase in the immobile population, which may now be comprised of adsorbed DTT-MTSL species. Both fail to produce the expected lineshape for 90-188. Low MTSL concentrations produce similarly poor results, failing to produce significant labeling above baseline at both 0.1 and 0.2 mg of MTSL. Spectra are not normalized.

An alternative explanation to the labeling problem stemmed from the rapid reduction observed for MTSL labels entering the periplasm. Perhaps this reductive environment extended to the cell surface, and the BtuB cysteines existed as highly reduced sulfhydryls. In this scheme, addition of excess MTSL, itself a weak reducing agent, would serve to further prevent formation of protein-label disulfide bonds. To test this, much lower concentrations of MTSL were employed, at 0.1 or 0.2 mg of MTSL instead of the typical 1 mg quantity that was sufficient for single-site experiments. The results of these trials are shown in Figure 5. C and D. Again, they failed to display the expected lineshape for the 90-188 pair. In both cases, there was no indication of labeling above the resonator baseline.

Despite the unappealing CW results, DEER experiments were still attempted in the hopes that the desired spectra were hidden beneath the free-spin and adsorption components. Representative apo and B12 traces for 90-188 in RK5016 cells are shown in Figure 6. Both in

the absence and presence of DTT and with and without substrate, these early DEER traces could be cleanly subtracted with a 3-dimensional background function, leaving a trace with 0 depth and no apparent oscillation. The lack of oscillation was consistent with no intramolecular dipolar coupling, and thus with a lack of doubly-labeled transporters. The 3D background, meanwhile, indicated that the spins were homogeneously distributed in the cell buffer, and thus also did not correspond to singly-labeled proteins on the 2d cell surface. So, the DEER results at this stage confirmed that the protein was not being sufficiently labeled.

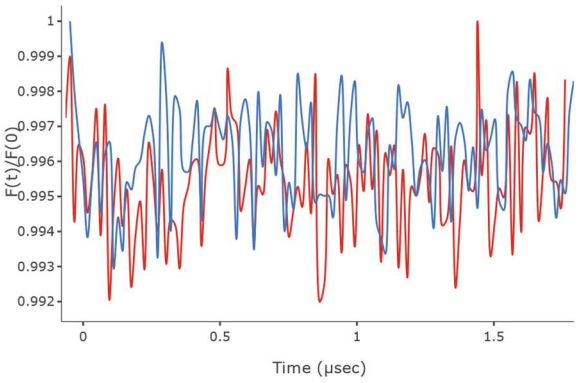


Figure 6. Dipolar Evolutions obtained for BtuB 90-188 in RK5016 cells at OD 0.3 with excess MTSL with (red) and without (blue) cobalamin. The data cleanly subtracts with a 3-dimensional background to leave no apparent oscillations. This indicates a total lack of intramolecular coupling in the sample.

It was clear that a different approach was required. The second line of reasoning, with reduced BtuB sulfhydryls, could have been explored further by treating the cells with weak oxidizing agents, but western blots became available that showed that the protein appeared to be quantitatively cross-linked (currently unpublished data). This was also consistent with another observation from the OM isolates. In those preparations, a large excess of MTSL was used in combination with very long labeling times of 2 hours. Unlike other systems in the lab, where labeling could be cut back to as short as 10-15 minutes and small quantities of MTSL could be employed, in the OM preparations any deviation led to poor labeling. In light of the western result, this was now consistent with the MTSL acting as both label and reducing agent, slowly breaking apart protein disulfide bonds before forming its own protein-label linkages.

All attention was now focused on ways to disrupt disulfide formation in the protein. The Dsb system could be modulated through mutants of *dsbA*, *dsbB*, and *dsbC*. If bonds were formed during insertion by the Bam complex, however, little could be done to block their formation and so it was hoped that the disulfide bond was formed during periplasmic translocation. To test this

hypothesis, the *E. coli* K-12 strain RI89 strain was obtained along with corresponding *dsbA*$^-$, *B*$^-$, and *C*$^-$ mutations. These knockouts are referred to by the affected protein for the remainder of the work, as DsbA$^-$, DsbB$^-$, and DsbC$^-$, while the WT control RI89 strain is referred to simply as Dsb. The DsbA$^-$ strain was selected for initial experiments, as it interacts directly with the protein during disulfide-bond formation and was considered the most likely to disrupt the process. High levels of MTSL again displayed no evidence of labeling for BtuB 90-188. However, the situation at low concentrations was entirely different, with spectra obtained for DsbA$^-$ cells with 0.05 to 0.25 mg of MTSL shown in Figure 7. At low levels of spin-label, there is now a distinct lineshape that

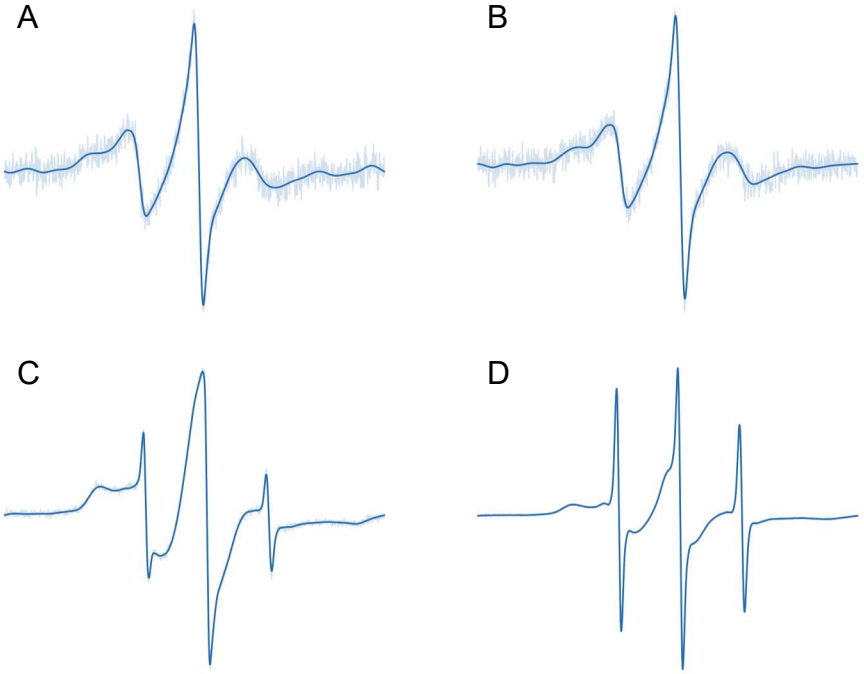


Figure 7. CW spectra obtained for BtuB 90-188 in DsbA$^-$ cells at OD 0.3 with (A) 0.05, (B) 0.1, (C) 0.25, and (D) 0.5 mg of MTSL. A lineshape consistent with an additive combination of sites 90 and 188 is observed in the 0.05 mg MTSL condition and is enhanced in the 0.1 mg MTSL condition. For the higher quantities of label, the spectra become overwhelmed by background adsorption, and in the 0.5 mg MTSL case is completely hidden.

corresponds to an additive combination of the 90 and 188 lineshapes observed in the singly-labeled RK5016 cultures. This lineshape is present with only 0.05 mg of MTSL but is clearly enhanced with an increase to 0.1 mg of MTSL. Further increases instead contribute to the background, with 0.25 mg of MTSL displaying a dominant background adsorption and significant

free MTSL. By 0.5 mg of MTSL, any trace of BtuB 90-188 is hidden. Thus, 0.1 mg of MTSL was settled upon as the optimal spin label quantity.

Now, with the observation of an apparently double-labeled protein in the DsbA$^-$ cells, DEER experiments were possible. The DEER results for apo and +cobalamin BtuB 90-188 in DsbA$^-$ cells are presented in Figure 8. The peak positions are unchanged from the OM data

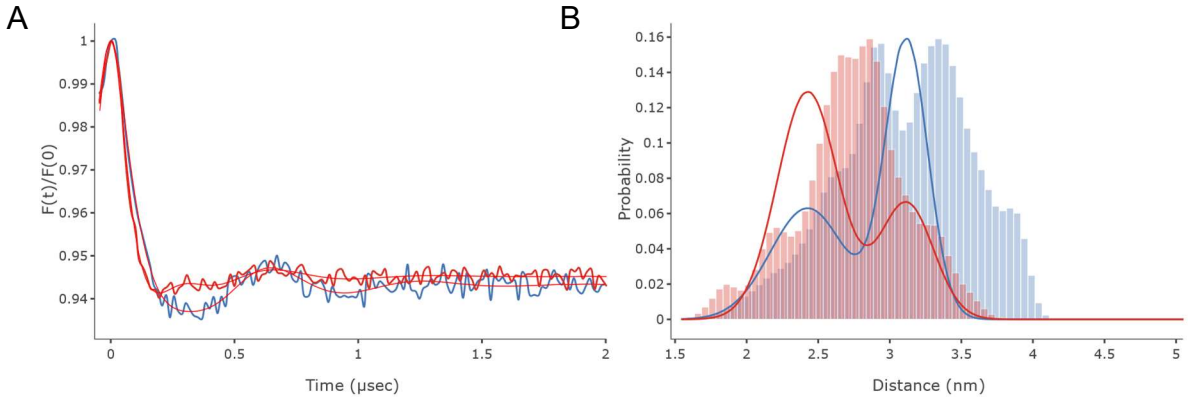


Figure 8. The (A) dipolar evolution, and (B) Distance distribution for the BtuB 90-188 pair in DsbA$^-$ cells at OD 0.3. For the time domain and dipolar evolution data, the apo condition is shown in blue, and a +100 μM cobalamin condition is shown in red. For the distance data, the results of 2-Gaussian fitting with a quadratic background function in the program LongDistances are shown. Two peaks are observed in the distribution, with a longer component at 3.2 nm and a shorter component at 2.4 nm. Histograms again give the predicted distances from the crystal structures. Compared to the RK5016 OM data, there is now a marked shift from the longer distance component in the apo state to the shorter distance component in the +cobalamin state.

presented in Fig. 2 for the RK5016 cell line, but the peak areas now shift dramatically from the longer component in the apo state to the shorter component in the +cobalamin state. This appeared to be evidence of the desired conformational shift, but it was necessary to determine if it was inherent to the membrane environment of the DsbA$^-$ cell line, or unique to the whole cells.

To test this, the OM isolate control was repeated using the BtuB 90-188 pair and the DsbA$^-$ cell line. The resulting DEER data is shown in Figure 9. As with the RK5016 OM data, the two dipolar traces are nearly identical, with the apo trace again displaying enhanced oscillation amplitude consistent with narrower distance components. In the distance distributions, the data now fit best to 3 Gaussians, with the shorter distance component observed in previous results being split into a peak at 2.3 nm and one at 2.7 nm. There is still minimal shifting between the two conditions in the OM, with only a slight increase in the shortest component in the +cobalamin

condition. Thus, most of the observed conformational change for BtuB 90-188 appears to be unique to the whole cells.

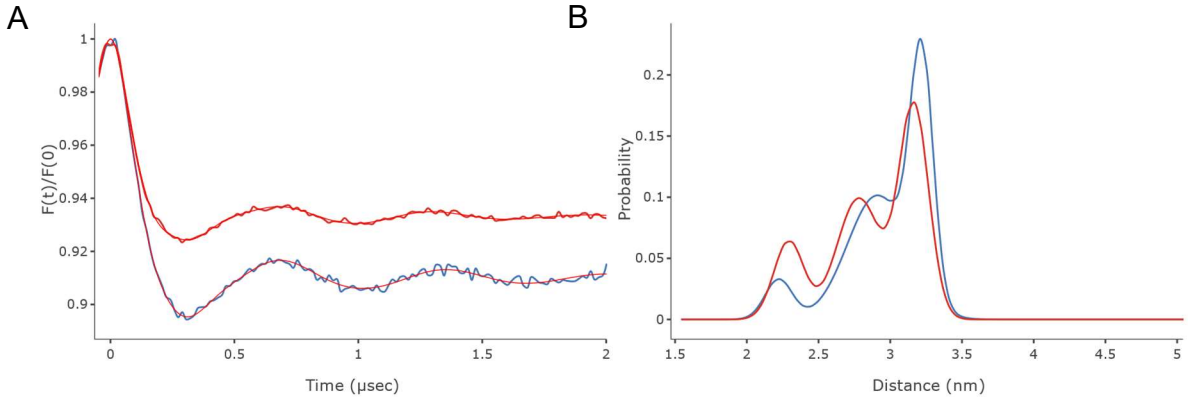


Figure 9. The (A) dipolar evolution, and (B) Distance distribution for the BtuB 90-188 pair in RK5016 Outer Membranes. For the time domain and dipolar evolution data, the apo condition is shown in blue, and a +100 μM cobalamin condition is shown in red. For the distance data, the fit is now improved with a 3rd Gaussian component that splits the shorter distance observed in previous results. Despite this, it is apparent that there is still minimal difference between the apo and +cobalamin conditions.

To further verify this result, and to confirm that the DsbA mutant functioned by disruption of disulfide bond formation in BtuB, the experiments were repeated in Dsb, DsbB$^-$, and DsbC$^-$ strains. The results of labeling all four Dsb strains at OD 0.3 and with low levels of MTSL are shown in Figure 10. The left side of Fig. 10 shows the results of attempted labeling of WT BtuB with 0.2 mg of MTSL, while the right side shows the results of labeling BtuB 90-188 with 0.1 mg of MTSL. For WT BtuB, no evidence of significant labeling is observed in any Dsb strain (Fig. 10 A-D), even at twice the MTSL quantity employed for labeling of BtuB 90-188 in DsbA$^-$ cells. Meanwhile, for BtuB 90-188, there is again no evidence of labeling in the Dsb strain (Fig. 10 E). This matches the result for RK5016. Conversely, both DsbA$^-$ and DsbB$^-$ display significant labeling with nearly identical lineshape (Fig. 10 F, G), containing an immobile component contributed by site 90 and a more mobile component from site 188. That these two mutations are mutually replaceable is consistent with their functional role in the cell. DsbA is responsible for formation of intra-protein disulfide bonds, but it is dependent on DsbB to recharge it for further rounds of disulfide formation. Without either protein the system is arrested after at most one round of bond formation. In contrast, the DsbC$^-$ strain again displays no significant labeling.

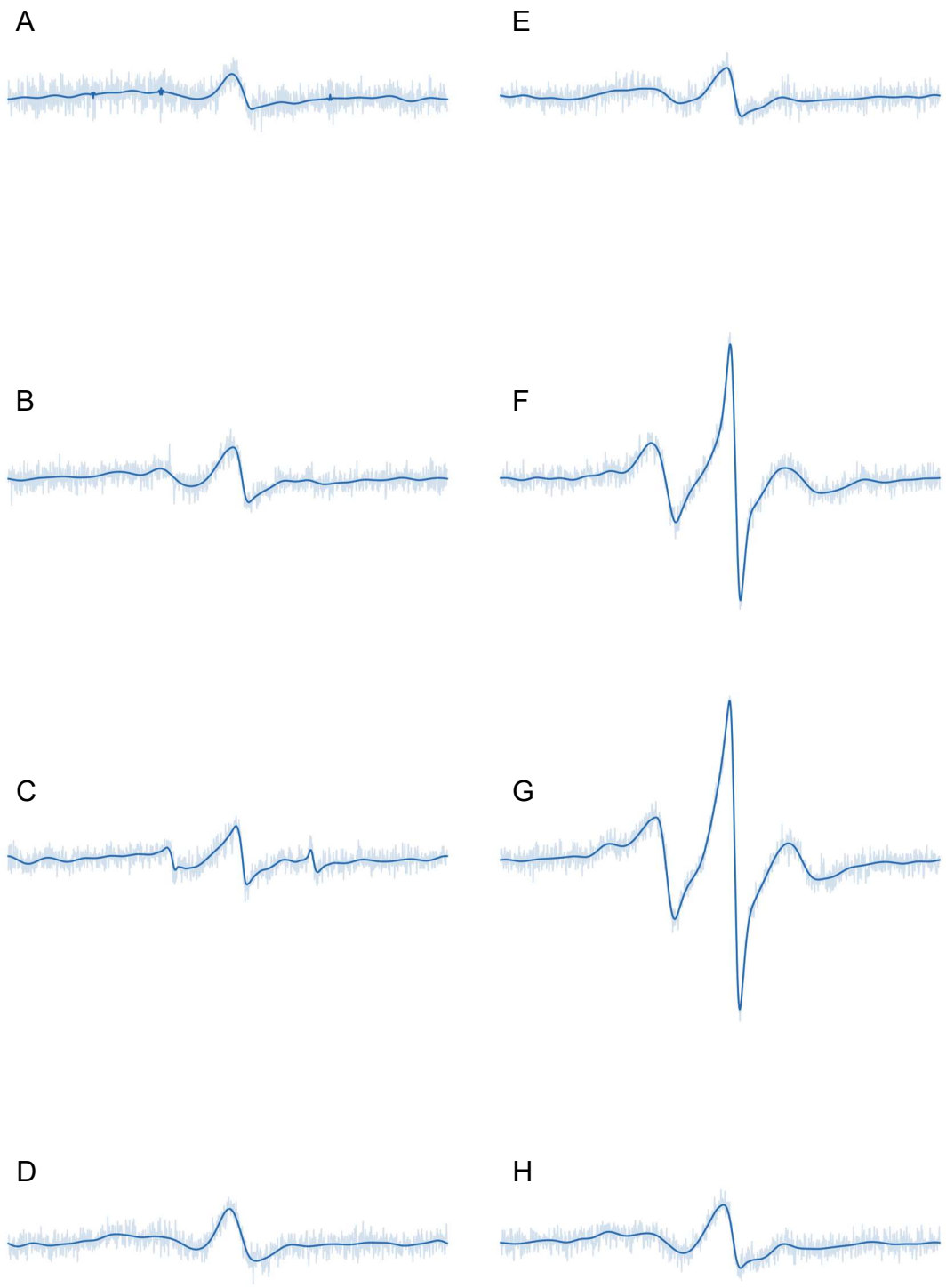


Figure 10. CW Spectra obtained for WT BtuB (Left) or BtuB 90-188 (Right) in the presence of *dsb* mutant strains at OD 0.3. For WT BtuB, 0.2 mg of MTSL was used for labeling. The spectra correspond to (A) Dsb, (B) DsbA$^-$, (C) DsbB$^-$, and (D) DsbC$^-$. In all four cases there is no evidence of significant labeling above baseline, implying that there is no significant background labeling in dilute MTSL for this protocol. The only differences are the presence of small free spin components in the DsbA$^-$ and particularly DsbB$^-$ traces. For BtuB 90-188, 0.1 mg of MTSL was used for labeling. The Spectra correspond to (E) Dsb, (F) DsbA$^-$, (G) DsbB$^-$, and (D) DsbC$^-$. Both Dsb and DsbC$^-$ display no significant labeling, while the spectra obtained in DsbA$^-$ and DsbB$^-$ are nearly identical. This was the expected result, as DsbA and DsbB are directly involved in disulfide bond formation. DsbC is a disulfide isomerase, but as the protein has only a single possible disulfide, knocking out DsbC should play little role in the formation of this bond.

The DsbC protein functions as an isomerase for proteins with multiple disulfides. The BtuB 90-188 used here contains only a single possible disulfide, and so the presence of an isomerase was not expected to have a major effect on formation of the unwanted 90-188 disulfide bond.

Building on the CW results, DEER data were also collected for BtuB 90-188 in the four Dsb strains. A comparison of the results for DsbA$^-$ and DsbB$^-$ are shown in Figure 11.

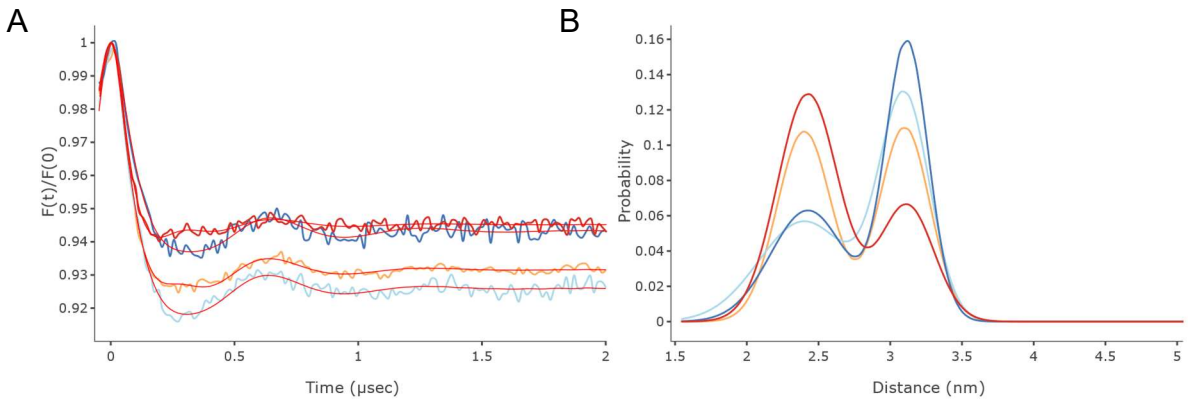


Figure 11. The (A) dipolar evolution, and (B) Distance distribution for the BtuB 90-188 pair in DsbA$^-$ and DsbB$^-$ cells at OD 0.3. For the time domain and dipolar evolution data, the DsbA$^-$ data is shown in blue (apo) and red (+cobalamin), while the DsbB- data is shown in light blue (apo) and light orange (+cobalamin). Aside from differences in modulation depth of the dipolar evolution functions, the DsbA$^-$ and DsbB$^-$ DEER are nearly identical. Both fit well to 2 Gaussians, with similar shifts from the longer distance component in the apo state to the shorter distance component in the +cobalamin state.

The DsbA$^-$ traces for BtuB 90-188 shown in Figure 8 are replotted here in blue and red for the apo and +cobalamin conditions, respectively, while the DsbB$^-$ traces are shown in light blue and orange. As expected from the CW spectra, the DEER data for the two strains are very similar. Both fit well with 2 Gaussian components, have nearly identical peak locations, and display a shift from the longer component to the shorter component with addition of cobalamin. To finish confirming the trend observed in the CW, DEER results for DsbC$^-$ are shown in Figure 12. While the CW results did not appear markedly different from the background labeling, there is a small DEER signal. Still, the modulation depth and thus quantity of interacting spin-pairs is reduced with respect to DsbA$^-$ and DsbB$^-$ and the distances are poorly resolved, fitting to a single Gaussian component. Thus, DsbC$^-$ appears to allow some leak of unbonded BtuB 90-188 into the OM but is not nearly as efficient as DsbA$^-$ or DsbB$^-$.

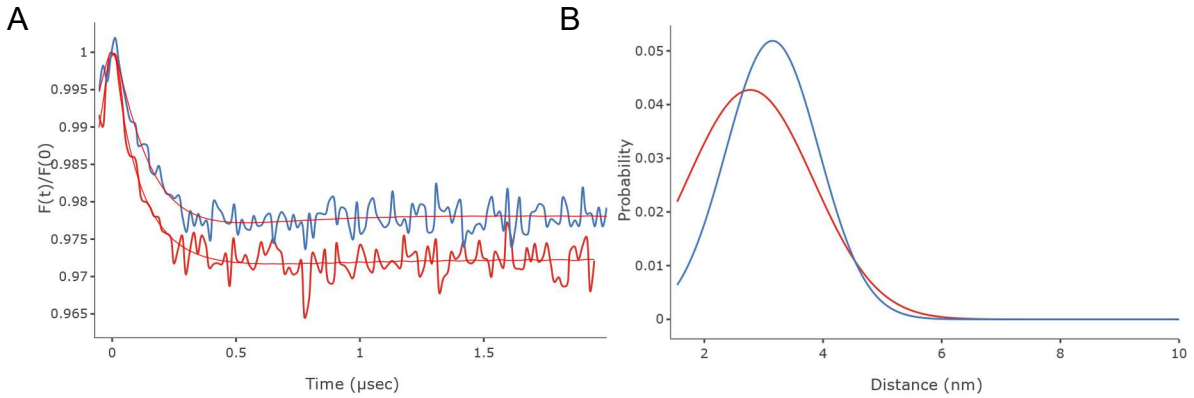


Figure 12. The (A) dipolar evolution, and (B) Distance distribution for the BtuB 90-188 pair in DsbC$^-$ cells at OD 0.3. For the time domain and dipolar evolution data, the apo condition is shown in blue, and a +100 μM cobalamin condition is shown in red. The modulation depth for the DsbC$^-$ traces is significantly smaller, and the distributions are poorly resolved, fitting only to single Gaussian distributions. There may still be a slight shift between apo and +cobalamin conditions, but the DsbC$^-$ strain is clearly inferior to data produced in DsbA$^-$ or DsbB$^-$.

Taken together, the preceding results clearly demonstrate that while single mutants of 90 and 188 were accessible to MTSL, the double mutant was inaccessible. Labeling was only possible in *dsbA* and *dsbB* null strains, which are directly involved in the formation of protein disulfide bonds during translocation into the periplasm. This was also found in the DEER results, where dipolar correlation was significant when DsbA or DsbB were eliminated, but greatly reduced when only the isomerase DsbC was removed. Further, no DEER could be detected for RK5016 cells having a WT Dsb system. Having thus determined a way to produce efficient double-mutants between hatch and barrel in whole cells, focus shifted to exploiting this method.

4.4 Dose-Dependent Response of the BtuB 90-188 Pair in DsbA$^-$ Cells:

The DEER data for the BtuB 90-188 pair, first shown in Figure 8, displayed a strong shift from a distance component at 3.2 nm to a shorter component at 2.4 nm upon addition of the cobalamin substrate. As mentioned in the beginning of this chapter, this site may serve as a reporter for secondary structure changes and potential domain movement in the upper hatch domain. Still, this preliminary result raised a number of questions. Was this shift a local response to substrate binding, or did it reflect larger changes in the hatch domain? Did the shift result from motion of 188, 90, or both? Could the change be tied to TonB binding? Many more were posed, and so several experiments were designed to begin to unravel the nature of this result.

First, the shift was reexamined at a range of cobalamin concentrations. Figure 13 A presents the distance distributions obtained for BtuB 90-188 overexpressed in DsbA$^-$ cells at cobalamin concentrations of 0, 5, 10, 20, 30, 60, and 100 μM. The apo condition is shown in blue, and the color transitions to red with increasing concentration of the cobalamin substrate. The data were fit to 2 Gaussians. To determine the positions of the peaks, the data were fit individually, and the average of each component was used as the peak position in subsequent fitting. The widths were allowed to vary although they displayed minimal variance, further indicating that this was a 2-state process. Since the peak widths were not fixed, height is not a direct reporter on the percentage of each component, but it can still be seen from the plot that there is a clear dose-dependent response to cobalamin at the ensemble level. The area occupied by the shorter component was then plotted against the cobalamin concentration in Figure 13 B and C. The first shows a hill fit to the full dataset, with a hill coefficient of 3.29 and R^2 of 0.992. The second shows a linear fit to the 0-30 uM concentration range, with an R^2 of 0.983.

Titration-like or globally analyzed plots using DEER data have only recently been introduced, with two literature examples involving the BmrCD ABC transporter and the multidrug transporter LmrP (4, 5). Both are large proteins that exhibit several nm conformational changes, and the studies used lipid nanodiscs or detergent solubilization. The data shown here should represent the first example in a native system. Still, there are a several caveats to producing this type of data with DEER. First, EPR spectroscopy works well in concentration ranges between tens and hundreds of μM. This is well above the substrate affinity of BtuB and many other proteins, and so a dilute condition is impossible. This limits the interpretation of any binding or dissociation constants obtained through such analysis.

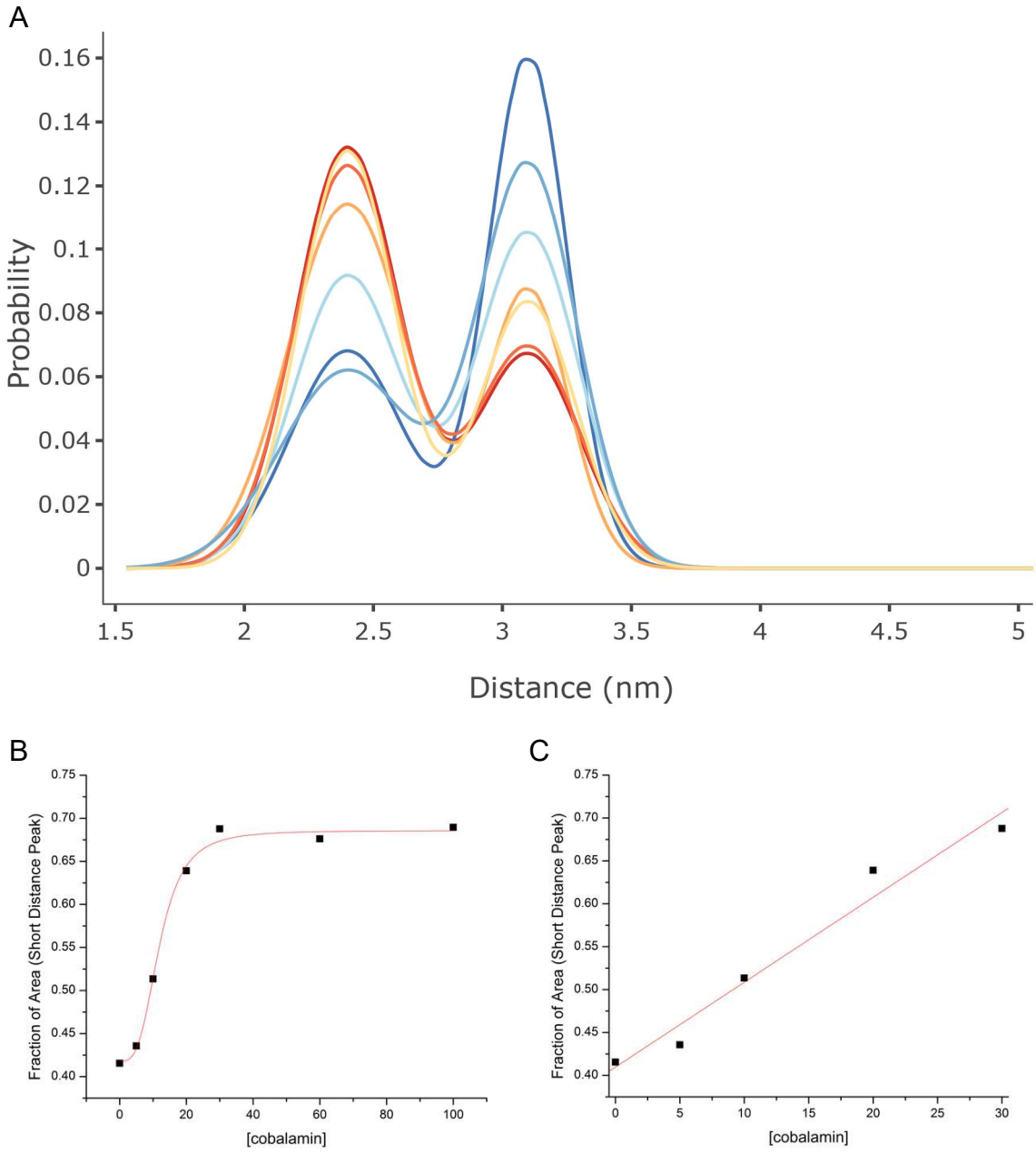


Figure 13. Analysis of DEER data for BtuB 90-188 over a range of cobalamin concentrations. The distance distributions obtained through fitting the data for cobalamin concentrations of 0, 5, 10, 20, 30, 60 and 100 μM to 2 Gaussian components are shown in (A). The color shifts from blue in the apo state to red with increasing cobalamin concentration. The areas of the shorter component were extracted and plotted against the cobalamin concentration to construct the plots shown in (B) and (C). The hill fit shown in (B) has a hill coefficient of 3.29 ± 1.19 and a coefficient of correlation of 0.992. The linear fit shown in (C) has a coefficient of correlation of 0.983. The meaning of other fit parameters is suspect, as both the protein and substrate are in excess in these experiments.

Additionally, these previously published studies on BmrCD and LmrP relied on a single software package, GLADD, by Eric Hustedt at Vanderbilt (6). In order to produce estimates of error from the set of related fits, the data were globally analyzed with linked peak positions and

widths, allowing for the determination of 95% confidence intervals at each point. Here, the GLADD software could not be employed due to the presence of an anomalous background form likely resulting from protein crowding in the OM, the details of which are described in the next chapter. This background contribution can currently only be removed using a different proGram, LongDistances by Christian Altenbach, but this prevents the determination of error in the individual values. By extension, it cannot be statistically determined if the data is best represented by the sigmoidal fits. At the time of writing, the author was in contact with the creators of both packages, and it may soon be possible to produce error analysis of this dataset.

In the absence of definitive error analysis, the simplest explanation is that the data are described by the piecewise linear fit and represents a sequential loading of BtuB by increasing amounts of cobalamin. This may provide a future method of determining the labeled, accessible concentration of BtuB on the cell surface. Alternatively, if the data are found to be best fit via the hill equation with a high coefficient value, the simplest explanation is that this value reflects the protein organization on the surface. While most TBDTs are thought to act monomerically, many surface proteins are trimeric in Gram-negative bacteria. The FepA TBDT has also been shown to have sigmoidal binding kinetics with a Hill-coefficient of 3, and this was posited to come from a trimeric organization of the FepA protein (7, 8). In support of this, that protein was originally present in a trimeric form in native gels (9). Still, the most likely explanation is that a conformational change occurs in response to substrate, and this experiment follows the percentage of proteins that have undergone said shift.

4.5 Attempted Modulation of the 90-188 Conformational Change

Next, the dependence of this structural transition on the trans-periplasmic energy transduction protein TonB was examined using a transport-defective mutation in the ton-box region, L8P. The C-terminal domain of TonB and the N-terminal ton-box in BtuB form a 4-stranded mixed B-sheet, which is critical to the transport of cobalamin across the OM. This interaction is dependent on hydrogen bonding though the even-numbered residues in the ton-

box, and proline mutations at these sites disrupt sheet formation and interactions between the two proteins (10). A comparison of the results obtained for BtuB 90-188 in DsbA$^-$ cells with and without the L8P mutation is shown in Figure 14.

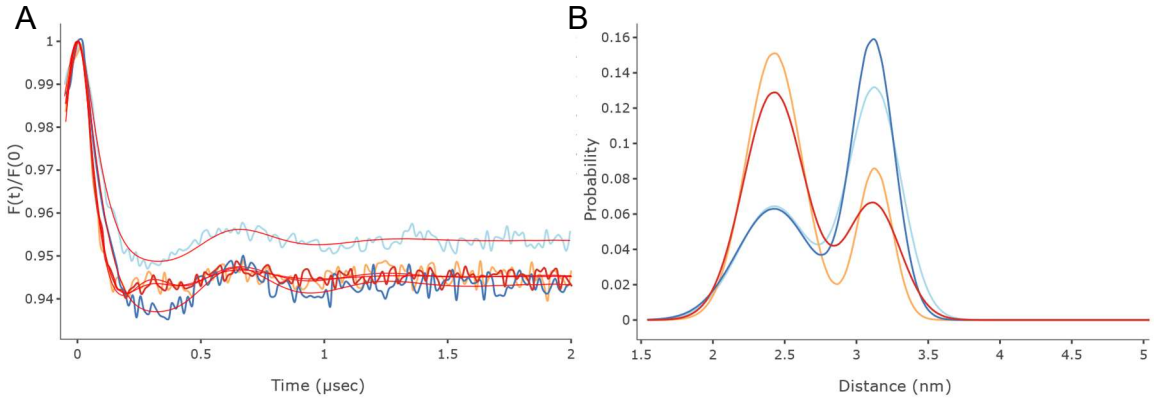


Figure 14. The (A) dipolar evolution, and (B) Distance distribution for the BtuB 90-188 pair in DsbA$^-$ cells with and without the L8P mutation. For the time domain and dipolar evolution data, the 90-188 data is shown in blue (apo) and red (+cobalamin), while the L8P-90-188 data is shown in light blue (apo) and light orange (+cobalamin). The traces are nearly identical, and the presence of the substrate defective mutation L8P does not appear to alter the peak positions or the extent of the conformational shift.

The darker blue and red colors show the original 90-188 data while the lighter colors give the results for L8P-90-188. It is immediately clear from the +cobalamin dipolar traces that the 90-188 and L8P-90-188 traces overlay almost exactly. This is also seen in the distance distributions, where L8P alters neither the position of the components nor the magnitude of the shift between them. This result could indicate that the conformational change is dependent only on the substrate binding event. Alternatively, it could reflect a decoupling of TonB binding from the response. TonB interactions are known to provide the energy for transport, but if energization by TonB precedes substrate binding, then a change might not be detected. Finally, recent work with AFM to probe interactions between the TonB C-terminus and the ton-box determined that the primary effect of L8P was to decrease the number of productive association events (11). Still, a substantial percentage of events were productive, and at the ensemble level seen in the DEER experiment they may still be sufficient to create the observed shift (11).

Next, the identity of which element of the 90-188 pair was moving during the shift was examined using two additional pairs, 188-399 and 74-188. The first pair had been previously

characterized in OM and an earlier attempt at EPR in intact cells, and was repeated here using the new protocol (2). This pair is shown in Figure 15. Site 399 is positioned at the base of the

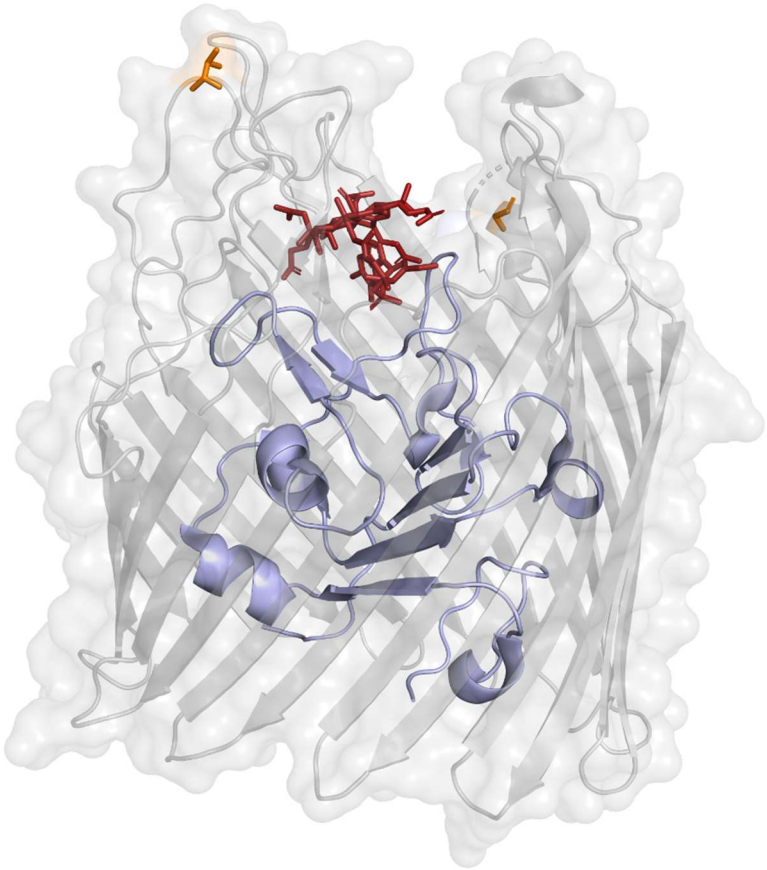


Figure 15. A side view of BtuB showing the sites 188 and 399 (orange sticks) in the extracellular loops. The cobalamin substrate is shown in dark red. This pair was used to check for motion of the 188 site in response to substrate binding

extracellular loop above site 90 and was previously found to change width but not shift in position in response to substrate binding (1). The CW spectrum of 188-399 in DsbA- cells is shown in Figure 16. Again, it displays an immobile component coming from the interior facing 399 site and a more mobile component from site 188. The DEER data obtained for this site in the apo and +cobalamin conditions, meanwhile, is presented in Figure 17. This time, the data were better fit using a quadratic background function and model-free distance determination in LongDistances, resulting in more apparent distance components.

In both the apo and +cobalamin conditions, the major distance is at 3 nm with a shoulder at 2.5 nm. This element does not shift in response to substrate addition. Both cases do show a

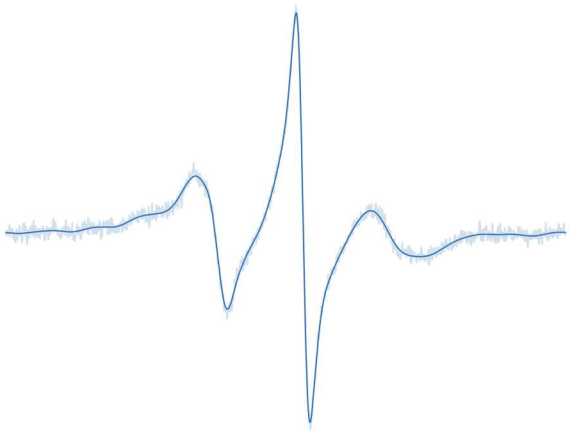


Figure 16. CW spectrum of the 188-399 pair in DsbA$^-$ cells at OD 0.3

unique secondary peak, however, with a poorly resolved short peak in the apo case and a small peak at 4 nm in the +cobalamin condition. Thus, while the majority of the population does not shift in response to substrate binding, a subpopulation does move, either from <2 nm to 4 nm, or in and out of the major population at 3 nm. The goal of this data was to determine if 188 was static in response to substrate binding, and so an additional pair was needed.

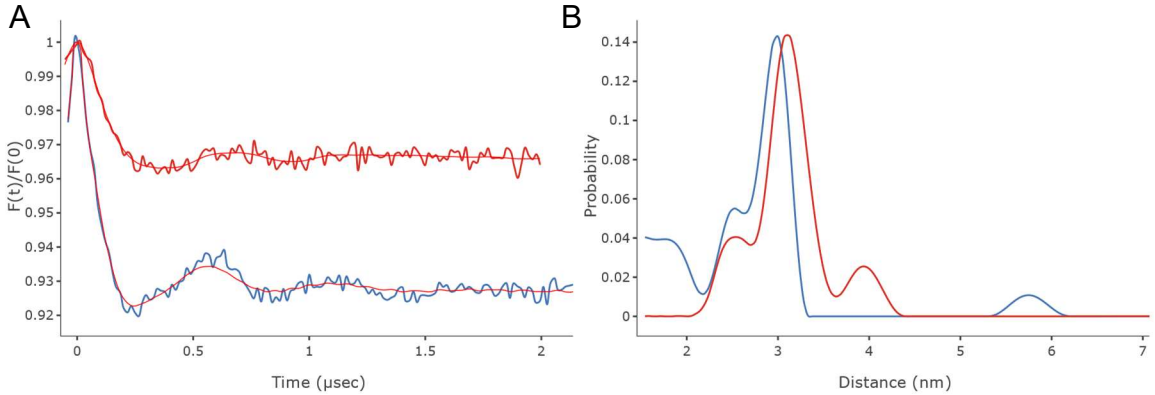


Figure 17. The (A) dipolar evolution, and (B) Distance distribution for the BtuB 188-399 pair in DsbA$^-$ cells at OD 0.3. For the time domain and dipolar evolution data, the apo condition is shown in blue, and a +100 μM cobalamin condition is shown in red. The major distant component at 3nm with a 2.5 nm shoulder does not shift significantly, but the apo condition displays a unique very short distance, which is poorly resolved by the experiment. Likewise, the +cobalamin condition displays a small population of a longer distance at 4 nm. Thus, this trace indicates that a subpopulation of 188-399 moves, and cannot confirm that 188 is static in response to substrate binding. Here, the data were fit using the model-free approach in LongDistances, with a smoothing factor of 10.

The other hatch site that had been previously identified for OM and intact cells was 74, and this was combined with 188 to make a 74-188 pair. The location of this pair is shown in Figure 18. Compared to site 90, 74 is located lower in the hatch domain and has proven more difficult to label due to its reduced accessibility. The CW spectrum of BtuB 74-188 in DsbA$^-$ cells is in turn shown in Figure 19, displaying the same general pattern of immobile hatch site and more mobile 188 site. Here, evidence of the struggle to label this site is apparent in the reduced

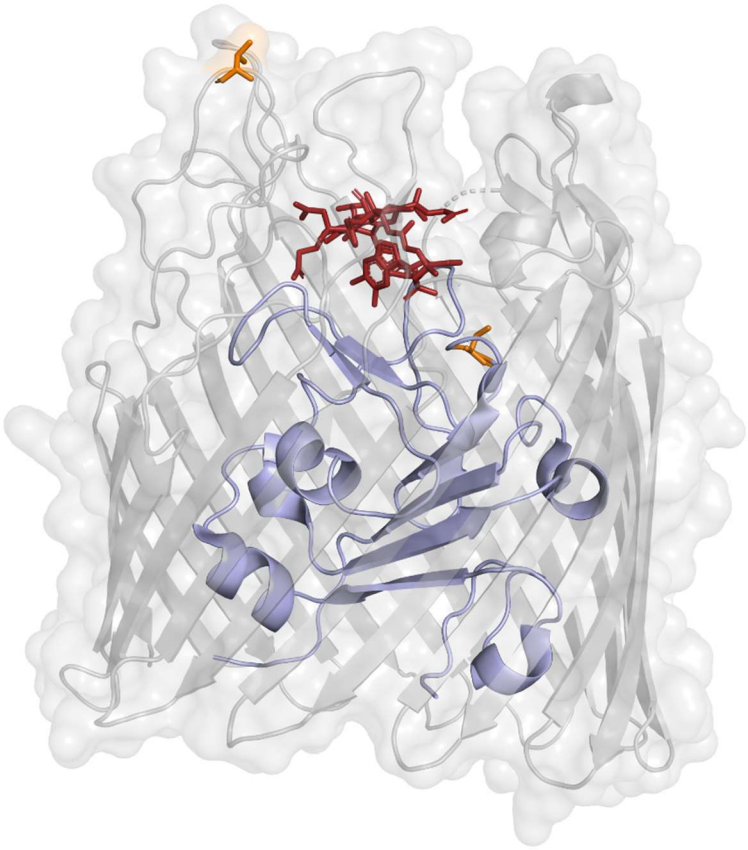


Figure 18. A side view of BtuB showing the sites 188 and 74 (orange sticks) in the extracellular loops and hatch domain, respectively. The cobalamin substrate is shown in dark red. This pair was used to check for motion of the 188 site or for additional conformational changes in the hatch, depending on the outcome of the DEER experiment.

contribution of the immobile component. After several attempts, DEER data were obtained for BtuB 74-188 in DsbA$^-$ cells and are shown below in Figure 20. The apo and +cobalamin traces show the same, sharp oscillation. Both produce a similar distribution to that observed with 90-188, with a main distance at 3.2 nm and a minor distance around 2.5 nm. Unlike 90-188, these distances do not interconvert in response to substrate. Thus, this pair does not display any additional conformational changes of the hatch domain. However, it does indicate that site 188 is

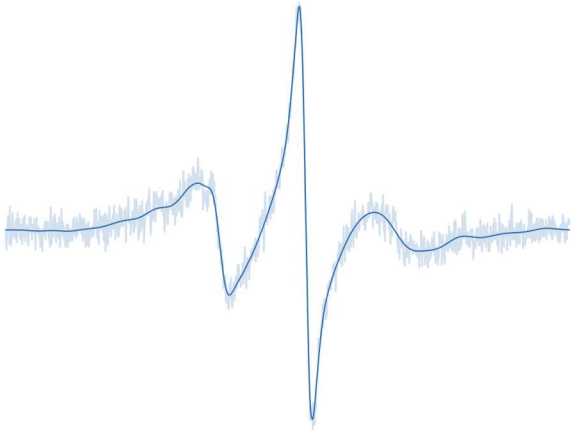


Figure 19. CW spectrum of the 74-188 pair in DsbA$^-$ cells at OD 0.3

likely to remain stationary in response to substrate addition, and that the conformational shift in 90-188 occurs in the hatch and not in the extracellular loops.

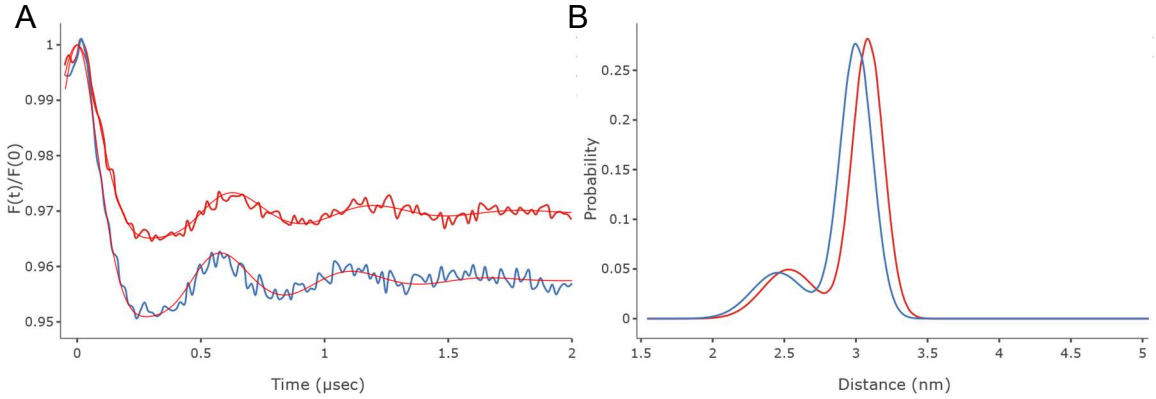


Figure 20. The (A) dipolar evolution, and (B) Distance distribution for the BtuB 74-188 pair in DsbA$^-$ cells at OD 0.3. For the time domain and dipolar evolution data, the apo condition is shown in blue, and a +100 μM cobalamin condition is shown in red. The data are once again fit to 2 Gaussian components in LongDistances. There is no shift in response to cobalamin addition, indicating that sites 74 and 188 are immobile. The amplitudes are also very sharp, with correspondingly narrow peaks. This may reflect rotameric restriction as a result of a more confined environment for site 74, which is deeper in the hatch domain than site 90.

Most of the data described here had been obtained using cell pellets that were collected through centrifugation, and the cell pellet was subsequently loaded directly into capillaries for EPR analysis. There was concern regarding the state of the PMF in these cells, as there was not sufficient buffer for aerobic processes, and the previous wash buffers did not contain glucose for metabolism and subsequent generation of an anaerobic PMF. Thus, the protocol was modified to include 2.5% glucose in all buffers and to keep buffer present with the cells at all steps. This made sample preparation more difficult, as the more dilute cell suspensions had reduced signal in the instruments. Still, data were recollected for both BtuB 90-188 and L8P-90-188 to see if a present or more robust PMF impacted the conformational shift in the 90-188 pair. The DEER data for this protocol variant are shown in Figure 21.

There is no apparent change in the results with the addition of glucose to all buffers. There is still a shift from the longer component to the shorter one with cobalamin, and the L8P mutant fails to change this behavior. This lack of difference implies that the transition is either not associated with energization via TonB interaction, or that the energization precedes conformational shifts that accompany substrate binding.

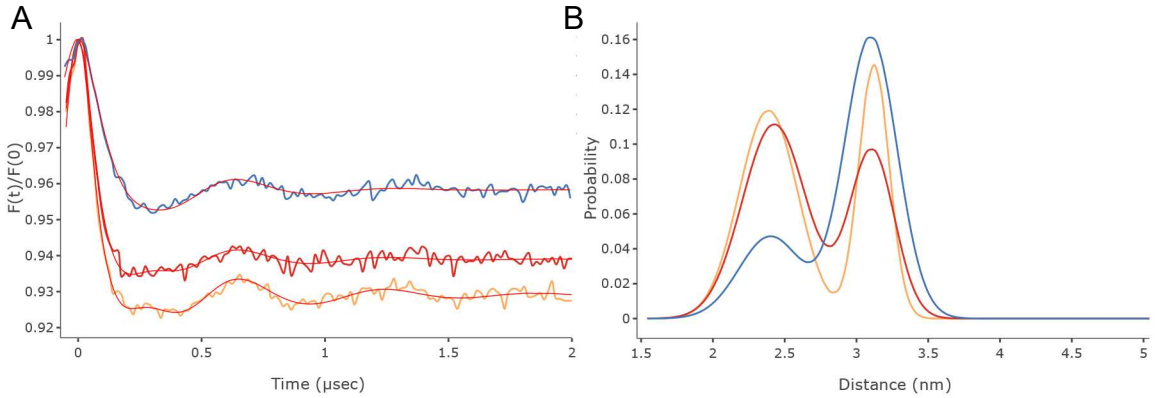


Figure 21. The (A) dipolar evolution, and (B) Distance distribution for the BtuB 90-188 pair in DsbA$^-$ cells at OD 0.3 and with 2.5% glucose. For the time domain and dipolar evolution data, the apo condition is shown in blue, and a +100 μM cobalamin condition is shown in red. The L8P-90-188 pair is shown for the +cobalamin condition in orange. The data are once again fit to 2 Gaussian components in LongDistances. The results are the same as before, with no apparent differences from the addition of glucose.

The easiest way to further characterize the extracellular face of the barrel would be through additional sites, and some progress is reported in the final chapter of this work. Several sites have been identified in both the hatch and barrel domain that label efficiently in DsbA$^-$ cells. Combinations of these new residues with sites 188 and 90 that were presented here will hopefully allow for more information to be determined regarding the nature of this conformational change, and if it is indicative of wider shifts in the overall hatch domain.

4.6 Extension of New Method into the Periplasmic Face of BtuB

During the transition towards using glucose in all cell culture buffers, it became apparent that extra metabolite led to a decreased reduction rate of the BtuB spin-labels. This may correlate with a reduced rate of cell death, as several cytoplasmic species are capable of spin-label reduction. Though regardless of the cause it made labeling of the periplasmic side of BtuB in whole cells seem plausible. Work from previous members of the lab had found that no signals could be detected for BtuB sites on the periplasmic surface, and early work leading up to the present method supported this. Still, the periplasm is an extremely attractive target for in-cell studies of BtuB. The periplasmic barrel surface contains only short, static turns instead of the elongated loops present on the extracellular side, which provide excellent reference sites.

Further, the Cafiso group had previously identified extension of a number of residues around the ton-box region of BtuB in response to substrate (12).

To both test if periplasmic labeling was possible and search for ton-box extension within the cellular environment, the 6-510 pair was selected. These sites are shown in orange in Figure 22. Residue 510 is located in one of the periplasmic turns that connect B-strands in the BtuB barrel, while residue 6 is part of the ton-box region of the hatch domain. Unlike proline mutations, cysteine mutations in the ton-box do not abrogate TonB binding.

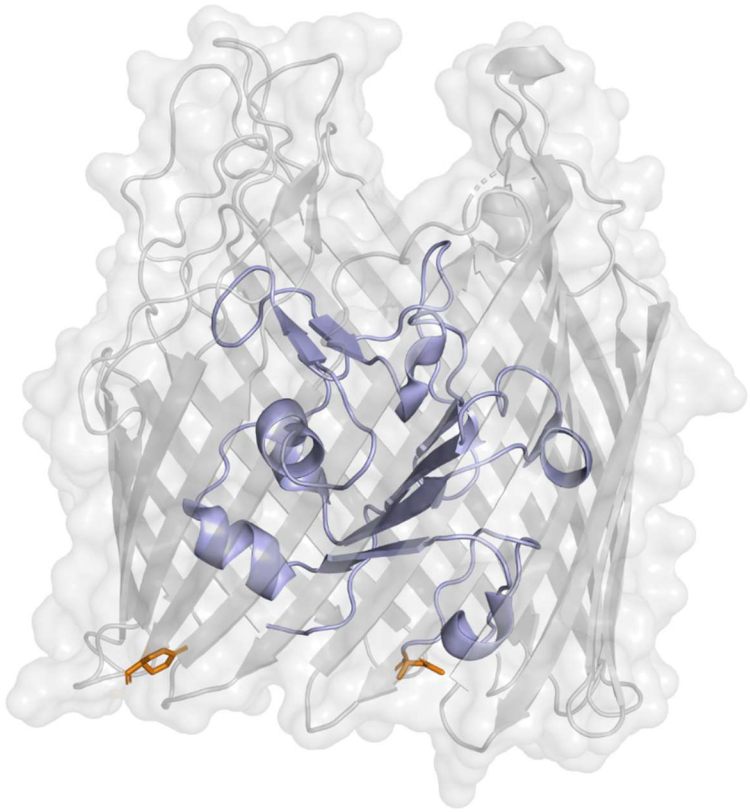


Figure 22. A side view of BtuB showing the sites 510 and 6 (orange sticks) in the barrel and ton-box region, respectively. This pair can be used to probe for extension of the ton-box into the periplasm.

This pair was previously characterized using both CW and DEER for BtuB reconstituted into POPC vesicles (12). *In vitro*, the apo distance for the 6-510 pair was about 2.5 nm (12). Addition of cobalamin created a second peak at 3.5 nm, consistent with extension of the ton-box away from the barrel (12). To test for this in native systems, the 6-510 pair was first isolated in

RK5016 OM. The CW spectra for this pair in the OM is shown in Figure 23. Ordinarily, there is significant background labeling in the OM isolates, which makes determination of unique spectral features difficult. Fortunately, site 6 has a characteristic bump in the high field line, which makes it easy to identify. The DEER data for this pair in RK5016 OMs is shown in Figure 24. The Dipolar traces show a significant bowl-shape partial oscillation that is particularly evident in the apo trace. This partial oscillation is likely indicative of crowding and correlates with the small peaks at 5-6 nm in the distance distributions. Otherwise, the apo distribution shows a main peak centered around 2.5 nm, which aligns with the substrate bound predicted distance, but not with the apo structure. Still, the apo predicted distance is less than 2 nm, and firmly in a region that is poorly resolved in DEER experiments. It thus cannot be ruled out that some population samples this distance.

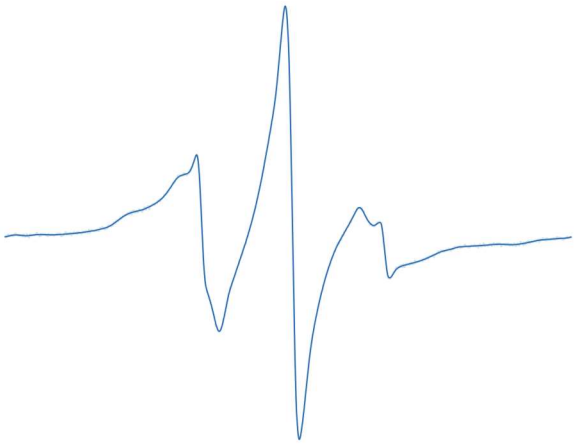


Figure 23. CW spectra for the 6-510 pair in RK5016 Outer Membranes. The contribution from site 6 is best seen in the highfield line, where it contributes a distinct central bump.

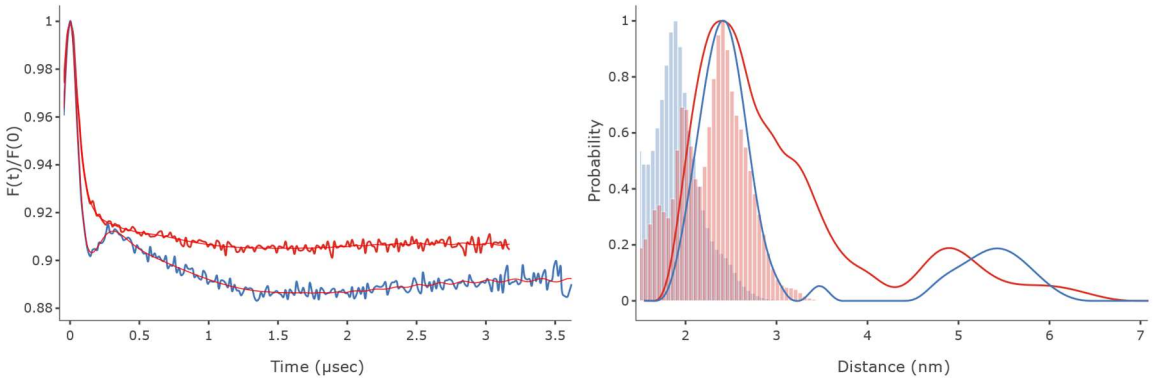


Figure 24. The (A) time domain, (B) dipolar evolution, and (C) Distance distribution for the BtuB 6-510 pair in RK5016 outer membranes analyzed using DeerAnalysis with a variable dimension background. Apo is shown in blue, and +cobalamin is shown in red. The shortest distance aligns in each condition and is similar to the predicted distances for the substrate bound structure (red histogram, PDB ID: 1NQH) but not for the apo structure (blue histogram, PDB ID: 1NQG). Substrate addition creates an intermediate population consistent with ton-box extension. The longest distances around 5-6 nm likely result from crowding.

The +cobalamin distribution has the same main peak at 2.5 nm, but now displays a shoulder around 3.2 nm consistent with ton-box extension. This aligns reasonably well with the *in vitro*

results, where the peaks were centered at 2.5 and 3.5 nm, although more of the population shifted into the longer component *in vitro*.

For the whole cell experiments, the 6-510 pair was grown in DsbA$^-$ cells to OD 0.3 in the presence of glucose. Surprisingly, it labeled, although the label disappeared very quickly, and all washing steps had to be cut out of the protocol in order to retain enough signal to obtain DEER data. Still, the resulting traces and distributions are shown in Figure 24 for the 6-510 pair in whole cells. Most of the partial crowding oscillation from the OM samples has disappeared, owing to the sparse distribution of labels that have escaped reduction in the periplasmic space. The distributions also still show the same main peak, which aligns with a folded ton-box observed in the +cobalamin structure and also includes an additional short component that may represent the apo structure. Now, however, there is evidence for several conformations that represent ton-box extension at distances of about 3.2, 3.6, and 4.6 nm. These extended peaks are present in both the apo and +cobalamin conditions, with no real change between the two. It is unclear why substrate dependent extension is not observed in the whole cells. One simple explanation is that only a small fraction of proteins are labeled, and are spared from reduction by their inactivity. Alternatively, it may be that the natural state of the ton-box *in vivo* is extended, to better search for the TonB C-terminus. Clearly, more experiments are needed to understand this phenomenon.

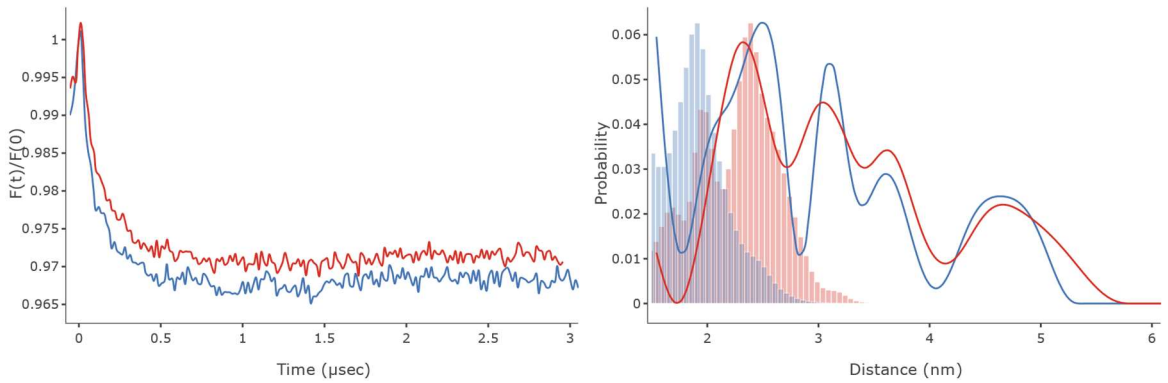


Figure 25. The (A) time domain, (B) dipolar evolution, and (C) Distance distribution for the BtuB 6-510 pair in dsbA$^-$ whole cells. Apo is shown in blue, and +cobalamin is shown in red. The major short distance still aligns in each condition and remains similar to the predicted distances for the substrate bound structure, but not for the apo structure. There is now also an extremely short component (<20 A) which is poorly captured but may correspond to the apo structure. The intermediate components that are consistent with ton-box extension are greatly enhanced in whole cells and are present in both apo and +cobalamin conditions.

4.7 Discussion

The results shown here indicate the dependence of disulfide linkages between hatch and barrel cysteine mutant pairs in BtuB upon the bacterial Dsb system. Initial results in a strain having WT Dsb activity (RK5016) showed that it was possible to label single cysteine mutants in the hatch and barrel domains. Subsequent attempts at DEER using the BtuB 90-188 pair in the hatch and barrel, respectively, resulted in the successful acquisition of DEER data in the isolated OM, while the whole cells showed a total absence of DEER (Fig. 2 and 6). The ability to successfully double-label the protein in isolated OMs indicated that the protein was present on the surface and hypothetically possessed a crosslink between sites 90 and 188. This crosslink may have formed at several points during formation and insertion of BtuB (Figure 26).

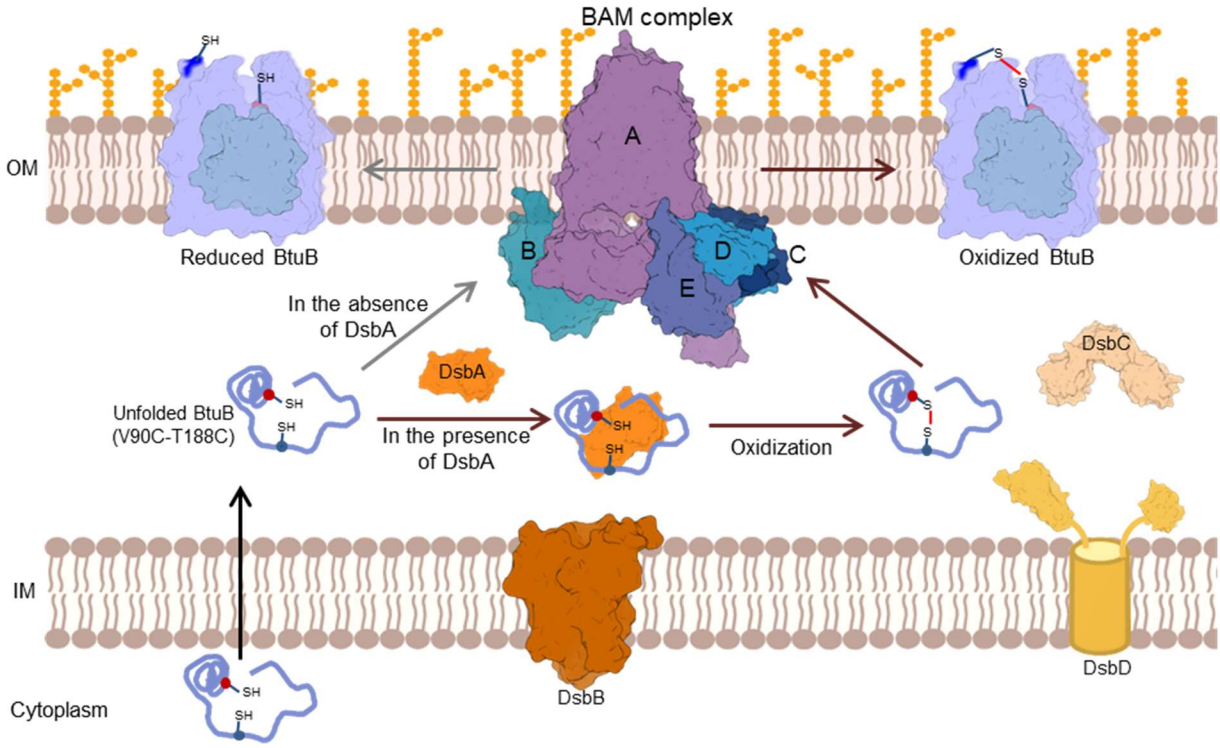


Figure 26. A model of the disulfide bond formation and OM insertion steps for BtuB. The putative crosslink between sites 90 and 188 in BtuB could have been formed by the proteins of the Dsb system (DsbA, DsbB, DsbC, DsbD in figure), during membrane insertion via the BAM complex, or after insertion on the cell surface. Within the Dsb system, DsbA is involved in the direct formation of template crosslinks and is reoxidized by the IM protein DsbB. DsbC is an isomerase which is recharged by the IM protein DsbD. Figure created by Dr. Thushani Nilaweera using BioRender.

The surface crosslink was probed by addition of excess label and pre-treatment with DTT, neither of which resulted in the formation of double-labeled protein. This may indicate that

the crosslinked conformation involved a folding over of the 188 loop into a form that was not readily accessible to DTT. Any perturbation of the BAM complex was likely to lead to folding and insertion defects, and thus attention was focused on the Dsb system. Disulfide formation is tightly regulated in Gram-negative bacteria, with initial crosslinks being formed by the protein DsbA during or after translocation of the protein template into the periplasm (13). This occurs through the initial formation of a mixed-protein disulfide between the catalytic Cys-X-X-Cys motif in DsbA and the first template cysteine, with subsequent formation of an additional mixed-disulfide with the second template cysteine and isomerization to give the template disulfide bond (13). The resulting reduced cysteines in DsbA are reoxidized by the IM protein DsbB (13). While this reoxidation of the DsbA catalytic motif is performed by two similar Cys-X-X-Cys motifs in DsbB, the latter protein does not participate directly in template disulfide formation, even when DsbB is decoupled from the IM and the Dsb system is moved to the bacterial cytoplasm (14). That the formation of template disulfides is dependent on both proteins was also observed here, as null mutations in both *dsbA* and *dsbB* resulted in the successful double-labeling of BtuB 90-188 (Fig. 10A and B) and the acquisition of significant DEER data (Fig. 11).

The removal of DsbC through a null mutant of *dsbC*, in contrast, produced no significant increase in labeling relative to a WT BtuB control (Fig. 10C). Further, it produced a much smaller DEER trace having only about 2% modulation depth in the apo condition (Fig. 12). This was consistent with the primary role of DsbC being a disulfide isomerase, as this function is unnecessary for a protein template having only a single putative crosslink. Still, the removal of DsbC did produce evidence of some double-labeled protein existing on the cell surface, and this implied that its removal was altering the extent of periplasmic crosslinking when compared to the WT Dsb and RK5016 strains. Recently, a secondary function has been demonstrated for DsbC and another periplasmic component of the Dsb system, DsbG. *In vivo*, these proteins serve not only as isomerases but also in the protection of free sulfenic acids in proteins

possessing only a single, cysteine residue, such as AraF (15). In the case of AraF, this occurs through the reduction of an inter-protein disulfide, which forms a nonfunctional AraF dimer (15). If DsbC was acting in a protecting role for the BtuB 90-188 template, then its removal would not be expected to contribute to successful double-labeling of the BtuB 90-188 target. Both DsbC and DsbG are reduced by the same IM protein, DsbD, however, and the two proteins have differing substrate specificity (16). Thus, the removal of DsbC may contribute to upregulating the reductive or protective actions of DsbG, although this would require significant future work to fully identify.

The direct role of DsbA in the formation of template disulfides made its removal a logical choice for future work requiring modulation of the Dsb system, and so the remaining experiments were conducted with the *dsbA* null mutant strain. For the 90-188 pair on the extracellular surface, the short-distance component was generally consistent with that seen in the OM data and also with the crystal structure predicted distances, although in the whole cells it readily split into two components with a substrate dependent inversion (Fig. 11). This inversion may represent an order-to-disorder transition in the hatch loop containing site 90, which can be seen in the apo and substrate bound crystal structures (PDB IDs: 1NQG and 1NQH). The change was also dose-dependent (Fig. 13) with a clear proportional response in addition to increasing quantities of substrate. That the substrate-dependent inversion in the short distances was not affected by the addition of the L8P mutation, which blocks TonB interactions, indicated that the change may not be energy-dependent (Fig. 14). Although, this observation would also be consistent with energization by TonB preceding substrate binding events. DEER spectra in cells lacking DsbA were also obtained for the 74-188 and 188-399 pairs, which showed minimal substrate-dependent shifts and indicated that the 188 loop was a suitable control for looking at movement in the hatch domain (Fig. 17 and 20).

Alteration of the Dsb system also made it possible to double-label sites on the periplasmic face of BtuB, including the 6-510 pair. In Isolated OMs, the DEER spectra for this pair showed the presence of a major short-distance element between 2 and 3 nm which was consistent with a folded ton-box region (Fig. 24). The addition of substrate, however, resulted in the formation of a second, shoulder peak past 3 nm which would be indicative of substrate-dependent extension of the ton-box. This substrate-dependence on the state of the ton-box was also observed in previous *in vitro* work from the Cafiso Lab (12). Through a combination of reducing the time spent in the labeling and processing of the cells, and the probable alterations to the periplasmic redox state due to the removal of DsbA, it was also possible to collect the same DEER data in whole cells (Fig. 25). Curiously, the whole cell data displayed peaks consistent with both a folded and unfolded ton-box in the apo state, with no change upon substrate addition. Thus, it appears that some percentage of BtuB possess constitutively unfolded ton-box elements in the cellular environment, which aligns to *in vitro* data collected for the ferrichrome transporter FhuA, and may have strong future implications on the nature of the interaction between BtuB and TonB (17).

4.8 References

1. Sikora, A., B. Joseph, M. Matson, J. R. Staley, and D. S. Cafiso. 2016. Allosteric Signaling Is Bidirectional in an Outer-Membrane Transport Protein. Biophysical journal 111(9):1908-1918.
2. Joseph, B., A. Sikora, and D. S. Cafiso. 2016. Ligand Induced Conformational Changes of a Membrane Transporter in E. coli Cells Observed with DEER/PELDOR. Journal of the American Chemical Society 138(6):1844-1847.
3. Jeschke, G. 2018. MMM: A toolbox for integrative structure modeling. Protein Science 27(1):76-85.
4. Stein, R. A., A. H. Beth, and E. J. Hustedt. 2015. A Straightforward Approach to the Analysis of Double Electron-Electron Resonance Data. Electron Paramagnetic Resonance Investigations of Biological Systems by Using Spin Labels, Spin Probes, and Intrinsic Metal Ions, Pt A 563:531-567.
5. Mishra, S., B. Verhalen, R. A. Stein, P.-C. Wen, E. Tajkhorshid, and H. S. McHaourab. 2014. Conformational dynamics of the nucleotide binding domains and the power stroke of a heterodimeric ABC transporter. eLife 3:e02740-e02740.
6. Brandon, S., A. H. Beth, and E. J. Hustedt. 2012. The global analysis of DEER data. Journal of Magnetic Resonance 218:93-104.

7. Thulasiraman, P., S. M. Newton, J. Xu, K. N. Raymond, C. Mai, A. Hall, M. A. Montague, and P. E. Klebba. 1998. Selectivity of ferric enterobactin binding and cooperativity of transport in Gram-negative bacteria. Journal of bacteriology 180(24):6689-6696.
8. Strahsburger, E., M. Baeza, O. Monasterio, and R. Lagos. 2005. Cooperative uptake of microcin E492 by receptors FepA, Fiu, and Cir and inhibition by the siderophore enterochelin and its dimeric and trimeric hydrolysis products. Antimicrobial agents and chemotherapy 49(7):3083-3086.
9. Liu, J., J. M. Rutz, J. B. Feix, and P. E. Klebba. 1993. Permeability properties of a large gated channel within the ferric enterobactin receptor, FepA. Proceedings of the National Academy of Sciences of the United States of America 90(22):10653-10657.
10. Cadieux, N., C. Bradbeer, and R. J. Kadner. 2000. Sequence changes in the ton box region of BtuB affect its transport activities and interaction with TonB protein. Journal of bacteriology 182(21):5954-5961.
11. Hickman, S. J., R. E. M. Cooper, L. Bellucci, E. Paci, and D. J. Brockwell. 2017. Gating of TonB-dependent transporters by substrate-specific forced remodelling. Nature Communications 8:14804. Article.
12. Xu, Q., J. F. Ellena, M. Kim, and D. S. Cafiso. 2006. Substrate-dependent unfolding of the energy coupling motif of a membrane transport protein determined by double electron-electron resonance. Biochemistry 45(36):10847-10854.
13. Hatahet, F., D. Boyd, and J. Beckwith. 2014. Disulfide bond formation in prokaryotes: history, diversity and design. Biochimica et biophysica acta 1844(8):1402-1414.
14. Mizrachi, D., M.-P. Robinson, G. Ren, N. Ke, M. Berkmen, and M. P. DeLisa. 2017. A water-soluble DsbB variant that catalyzes disulfide-bond formation in vivo. Nature Chemical Biology 13:1022. Article.
15. Denoncin, K., D. Vertommen, I. S. Arts, C. V. Goemans, S. Rahuel-Clermont, J. Messens, and J.-F. Collet. 2014. A new role for Escherichia coli DsbC protein in protection against oxidative stress. The Journal of biological chemistry 289(18):12356-12364.
16. Depuydt, M., S. E. Leonard, D. Vertommen, K. Denoncin, P. Morsomme, K. Wahni, J. Messens, K. S. Carroll, and J.-F. Collet. 2009. A Periplasmic Reducing System Protects Single Cysteine Residues from Oxidation. Science 326(5956):1109.
17. Kim, M., G. E. Fanucci, and D. S. Cafiso. 2007. Substrate-dependent transmembrane signaling in TonB-dependent transporters is not conserved. Proceedings of the National Academy of Sciences of the United States of America 104(29):11975-11980.

Chapter 5: Crowding and Organization of overexpressed BtuB on the *E. coli* cell surface

5.1 Introduction

The surface of Gram-negative bacteria is comprised of an OM whose makeup differs drastically from the phospholipid bilayer below. It is filled with transporters and trimeric porins, which form the majority of its surface. In between them sit LPS molecules, whose strong adjacent interactions reduce both diffusion across the membrane and lateral movement along it. Together, they form a constrained envelope around the cell, where diffusion is typically local and movement is driven stochastically by synthesis and insertion of new elements.

It is into these crowded environments that BtuB mutants are inserted in the whole cell experiments presented in this work. In WT cells, BtuB has a copy number of only several hundred, with haploid cells producing about 200 per cell (1). Yet, when overexpressed with the pAG1 plasmid, the level of BtuB may reach 40% of the native porins (1). There can be many tens of thousands of OmpF and OmpC porins on the cell surface, and the addition of a great number of BtuB transporters can be expected to further constrain this tightly packed surface. Additionally, recent work has highlighted the sparse nature of insertion events, with fluorescently tagged TBDTs being observed to appear together in organized patches, or islands, on the OM (2). These patches reached half a micron in size, containing hundreds of proteins (2). Additionally, the proteins comigrated across the cell surface, in agreement with previous measurements of confined diffusion in most OM proteins (3).

Coarse-grained simulation studies designed to mimic BtuB in these islands found that the proteins interacted through two weak interfaces formed from hydrophobic and aromatic residues on the barrel surface (2). An extension of these studies to the mesoscale found the formation of extensive, string-like aggregates of protein driven by these weak interfaces (4). Branching aggregates have also been observed in coarse-grained studies of visual rhodopsin, and even in simplified B-barrel mimetics containing only three amino acids: leucine, serine, and tryptophan (5, 6). For these systems, the aggregation was driven by lipid mismatch effects, which have also

been demonstrated in BtuB. Yet, these protein strings have not been detected experimentally on the cell surface. Fluorescence techniques lack the resolution, and while AFM has been used to probe the supramolecular organization and interactions of the major porins due to their trimeric oligomers, it would be difficult to characterize a mixed-population of smaller, monomeric TBDTs interacting only through weak protein-protein interactions. This presents another niche for EPR spectroscopy, where the DEER experiment can fill in distance information at the molecular level, covering the 20-80 Angstrom range, and does so only for the mutant protein of interest, without interference from other surface species.

Viewed from the top, a single BtuB molecule is an ellipsoid with major and minor axes of 4.2 and 3.7 nm, respectively (7). A pair-wise interaction between BtuB molecules would then produce a distance between about 4 and 8 nm depending on the orientation and radial-offset of the labeled site. For non-random cases, such interactions should appear as discrete peaks in the distribution, while for random interactions created only by the crowded environment, they may only affect the intermolecular background term. Background functions in DEER are typically modeled as simple, decaying exponential functions but this holds only for homogeneously distributed systems. Deviations due to excluded volume, radial offset, and crowded environments all significantly alter the decay of the time-domain data (8). Thus, it should be theoretically possible to discern the mode of interaction for BtuB overexpressed into a highly crowded native environment.

The *in vitro* experiments used by the Cafiso lab to study BtuB for many years have not produced evidence of crowding or protein-protein interactions. Yet, these experiments used BtuB reconstituted into excess POPC vesicles, producing low density incorporation without the LPS or other proteins that define the native surface. During the transition from reconstituted systems towards OM isolates, however, long-distance features began to consistently appear across seemingly unrelated traces. An example is shown in Figure 1, taken from a distance measurement between two extracellular loop sites, 449 and 553, in OM isolate (9). The short

distance components align well with the predicted crystal structure distance (histoGram), but in both the apo and +B12 conditions there are unlinked peaks between 5 and 6 nm. This is particularly apparent in the +B12 trace, where the longer distance component dominates the data, manifesting as a bowl-shaped partial oscillation between 0.5 and 3.5 μs. These peaks were observed in many, if not most, of the data collected in OM isolates, with a distance between 5 and 6 nm. The peak's position changed more with runtime than identity of the sample, suggesting that it might be inherent to BtuB rather than a product of any given label-pair. Yet 5-6 nm is only slightly greater than both the width and height of the protein, and so for most distributions it could be reasonably argued that this long-distance component was an intramolecular feature. Still, when these distance components began to appear again and again in the data used for this thesis, it became apparent that the identity of this peak needed to be investigated.

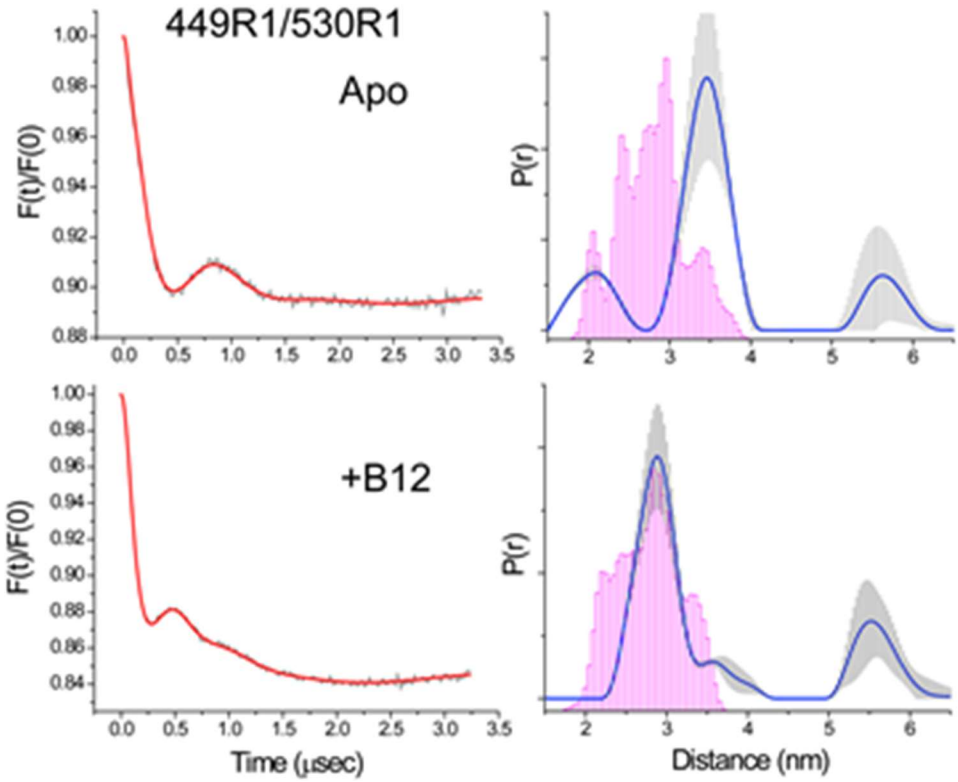


Figure 1. The long-distance component observed in prior OM isolate data. The figure is taken from (9) for DEER data obtained between sites 449 and 530 in BtuB overexpressed in OM isolates. The peak between 5 and 6 nm is disjoint from the rest of the distribution and is the dominant distance element in the +b12 trace.

5.2 Additional Methods Information by Figure

The modifications to the protocols in Chapter 2 are given below as a function of the experiments reported in each figure.

Fig. 3 Growth and isolation of OMs from RK5016 cells overexpressing BtuB 90C-188C proceeded as written in Chapter 2. 100 μM cobalamin was added for the + cobalamin condition.

Fig. 4 9/20/17 Growth of RK5016 cells overexpressing BtuB 188C to OD 0.6 was followed by resuspension in 50 mM HEPES, pH 8.0. Labeling proceeded for 1 hr with 1 mg of MTSL and was followed by 1x surface wash ice with 50 mM HEPES, pH 8.0. 32 μM spin-labeled substrate was added for the +substrate condition.

Fig. 6 6/30/17 and 4/21/18 Growth of RK5016 cells overexpressing BtuB 188C to 6.5 hrs post inoculation was followed by resuspension in 50 mM HEPES, pH 8.0. Labeling proceeded for 1 hr with 1 mg of MTSL and was followed by 1x surface wash with 50 mM HEPES, pH 8.0. Growth of RK5016 cells overexpressing BtuB 90C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 1 hr with 1 mg of MTSL and was followed by 2x 30-minute wash steps on ice with 100 mM HEPES, pH 7.0 and 100 mM MES, pH 5.5, respectively.

Fig. 7 5/24/18 and 5/25/18 Growth and isolation of OMs from RK5016 cells overexpressing BtuB188C proceeded as written in Chapter 2. 1 mg of MTSL was used for the full labeling condition, and 0.5 mg MTSL was combined with 0.5 mg of 1-acetyl-2,2,5,5-tetramethyl-Δ3-(pyrroline-15N)-3-methyl methanethiosulfonate for the dilution condition.

Fig. 8 6/30/18 and 7/1-4/18 and 7/25/18 and 7/29/18 Growth of RK5016 cells overexpressing BtuB 188C to 6.5, 8, 9.5, 11, 12.5, 14, or 28 hrs post inoculation was followed by resuspension in 50 mM HEPES, pH 8.0. Labeling proceeded for 1 hr with 1 mg of MTSL and was followed by 1x surface wash with 50 mM HEPES, pH 8.0.

Fig. 9 Summarizes results obtained for RK5016 cells at various OD values overexpressing BtuB 188C in the presence and absence of the spin-labeled substrate. The samples were resuspended in 50 or 100 mM HEPES buffer at pH 7.0 or 8.0 with 0-2 30-minute wash steps. All samples were

incubated on ice and 20-60 uM of spin-labeled substrate was added for samples having 2 peaks in the distribution.

Fig. 10, 12, 13, 14 Growth of RK5016 cells overexpressing BtuB 188C OD 0.5 in deuterated media with 0.2% deuterated glycerol as carbon source was followed by resuspension in 100 mM HEPES, pH 8.0. Labeling proceeded for 1 hr with 1 mg of MTSL and was followed by 2x 30-minute wash steps on ice with 100 mM HEPES, pH 7.0 and 100 mM MES, pH 5.5, respectively. All buffers were made in D2O.

Fig. 15 Growth and isolation of OMs from RK5016 cells overexpressing BtuB 90C-188C proceeded as written in Chapter 2.

Fig. 16 (A) Growth and isolation of OMs from RK5016 cells overexpressing BtuB 6C-510C proceeded as written in Chapter 2. 32 μM spin-labeled substrate was added for the +substrate condition. **(B and C)** Growth of DsbA$^-$ cells overexpressing BtuB 6C-510C to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 30 minutes with 0.1 mg of MTSL and was followed by 1x minimum time wash step at RT with 100 mM HEPES, pH 7.0. All buffers now contained 2.5% glucose. 100 μM cobalamin was added for the + cobalamin condition.

Results and Discussion

5.3 Characterizing Recurrent Long-Distance Components in Native Systems

Before experimental data showing similar long-distance components to that in Figure 1 are presented, Figure 2 gives simulated examples of DEER data showing a single distance at 3 nm or at the 5.5 nm position indicated in Figure 1. These simulated examples have no background from intermolecular interactions, to show what the target trace would look like after an ideal background model is removed from the experimental signal. In the short distance case, a complete oscillation is obtained before 1 μs, and the trace has decayed almost completely by 2 μs. Most of the DEER measurements taken in native systems have runtimes between 3 and 3.5 μs, and so if only short distances are present it would be relatively simple to account for this component. The simulated 5.5 nm trace, meanwhile, completes a full oscillation at more than 4 μs. For traces between 3 and 3.5 μs only a partial oscillation will be present, and this will make background subtraction difficult, since the asymptotic behavior of the signal will not be visible.

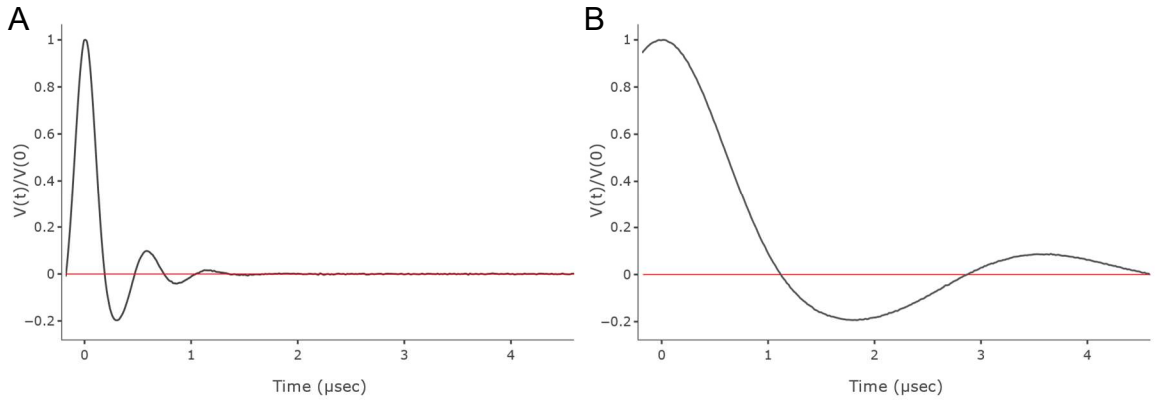


Figure 2. Simulated time domain data for DEER traces with distances at 3 nm (A) and 5.5 nm (B) and widths of 4 and 8 Å, respectively. The data were generated using LongDistances with modulation depth of 0.5 and no intermolecular background. In the absence of intermolecular interactions, the DEER traces decay asymptotically towards 0. For the short distance example, which emulates the 90-188 pair discussed in the previous chapter, most of the oscillation is present before 1 μs. For the longer distance, which mimics the long-distance component in the OM isolate data, a full oscillation is not observed until 4 μs. Much of the data in native systems could only be run to 3 or 3.5 μs, and so the longer distance component will appear as a bowl-like shape, making background subtraction difficult.

As a data example, the DEER data for the 90-188 pair in RK5016 OM isolates from Chapter 4 is shown again in Figure 3. Here, the analysis was performed using Tikhonov regularization in DeerAnalysis. The trace was subtracted with a stretched exponential function of

dimension 2.4. The initial portion of the dipolar oscillation shown in Fig. 3A is raised above the latter part of the trace, which stems from convoluting the higher frequency short-distance oscillation with the broader decay from the long-distance component. The fit to the data is shown in red, and while sufficient for the early portion of the trace, becomes increasingly poor through the partial oscillation at the end. This is evident in the residuals plot shown in the inset, where an oscillation is still present in the residuals.

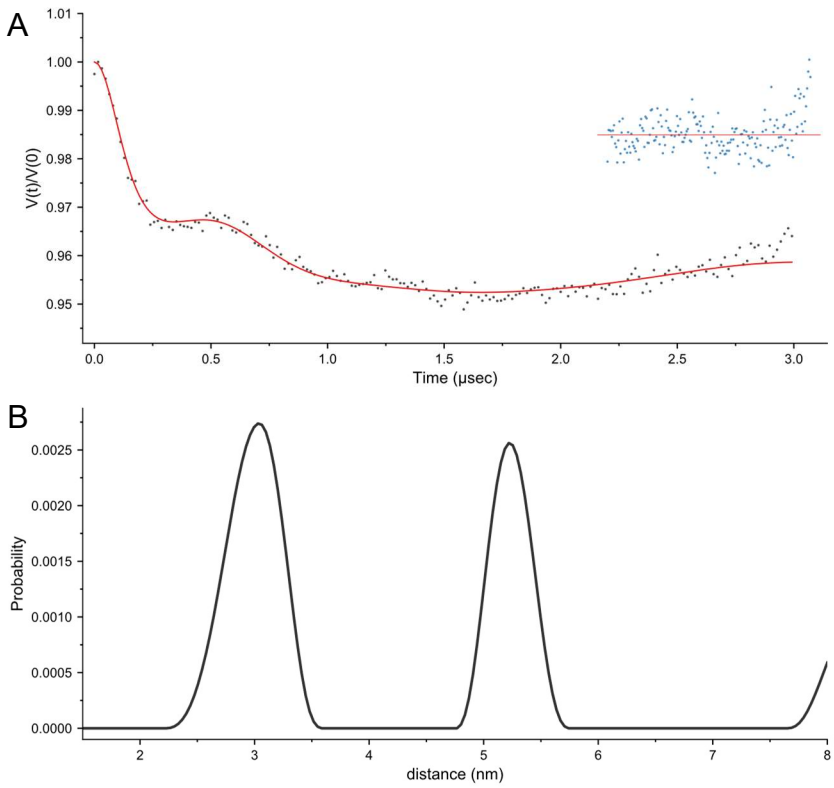


Figure 3. The (A) dipolar evolution, and (B) Distance distribution for the BtuB 90-188 pair in RK5016 OM isolates analyzed using Tikhonov regularization in DeerAnalysis2015. The background dimension was 2.4 and the value of the regularization parameter was 100. The dipolar trace is dominated not by the expected 90-188 distance, but by a partial long-distance oscillation from 0.5 – 3.0 μs that appears as the peak at 5.2 nm in the distance distribution.

The distance distribution in Figure 3. B contains two peaks. The first is the expected distance at 3 nm, which is not split into two components at 2.4 and 3.2 nm due to the relatively high value (100) of the regularization parameter. The second is at 5.2 nm, and its low width is largely responsible for the poor fit at the end of the trace. Tikhonov regularization is a global optimization procedure, with one regularization parameter that controls the smoothness and by extension peak widths of all components in the distribution. The width of the longer distance peak

is thus constrained by the relatively narrow short distance peak, creating the spike in the residual plot at the end of the trace. Still, the background model is also nonideal, which is manifested in the smooth oscillation present throughout the fit residuals.

The expected distances for the 90-188 pair from the apo and +cobalamin crystal structures were shown in Chapter 4. In both cases, there is no evidence for a peak at 5-6 nm, and if this peak came from an intramolecular interaction, it would indicate substantial rearrangement of the hatch domain consistent with the hatch being partially removed from the barrel. The alternative explanation is that the peak comes from lateral, intermolecular interactions between spins on neighboring BtuB proteins.

To help clarify, data for a second pair is shown in Figure 4. Here, the data were obtained for BtuB 188 in RK5016 cells and a spin-labeled cobalamin analogue, which was provided by Dr. Benesh Joseph (University of Frankfurt) and has been previously characterized (10, 11). The data were again analyzed with DeerAnalysis, this time with a background dimension of 3.0 and a regularization parameter of 10. The Dipolar trace shown in Fig. 4A shows the same features as were seen for 90-188 in Fig. 3. There is a short-time oscillation that correlates to the peaks between 2 and 3 nm in Fig. 4B and a partial oscillation observed through the rest of the trace, which again appears between 5 and 6 nm in the distance distribution. Here, the experiment could be run for slightly longer, allowing more information to be collected on the longer oscillation. Compared to the peak in Fig. 3 with a center at 5.2 nm, the increased runtime has shifted the current peak to 5.6 nm. This suggests that the real distance is longer still, and simply cannot be resolved by the given runtime. Again, the fit is poor at the end of the trace, with a notable spike in the residuals indicating that the peak is overly narrow. Also, there is still a full oscillation across the residuals plot, indicating that the background subtraction and fit still do not fully describe the data.

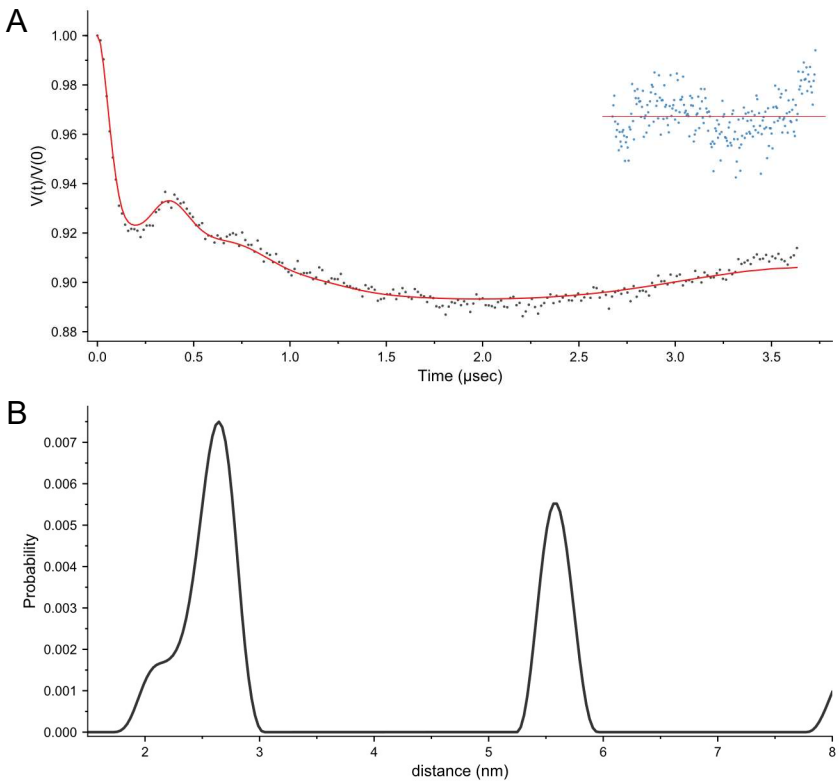


Figure 4. The (A) dipolar evolution, and (B) Distance distribution for RK5016 BtuB 188 and a spin-labeled cobalamin substrate analyzed using Tikhonov regularization in DeerAnalysis2015. The background dimension was 3.0 and the value of the regularization parameter was 10. The expected background dimension for proteins on a membrane is 2, and so a value of 3 for the background is abnormal. This could reflect the presence of free spin-labeled substrate, but the concentration of cobalamin (32 uM) was too low for free cobalamin to be the dominant background element. Increases in apparent dimensionality have also been seen for crowded systems and in the presence of excluded volume.

For this trace, an intramolecular distance of 5-6 nm would indicate that the substrate had moved to an additional binding site on the periplasmic surface of BtuB. Such a site would almost certainly be accessible to the periplasm, where the rapid reduction would mean that the long-distance peak should disappear after a short incubation. Yet, these traces were taken after much longer incubation times than had to be employed for the 6-510 data shown in Chapter 4, and the long-distance component was relatively stable. If the distance was instead intermolecular, then it would also help explain the other unexpected element of this dataset, the background dimension.

Usually, a dimensional value of 3 would be expected for soluble proteins or detergent solubilized samples. The second spin in this experiment was donated by the spin-labeled

substrate which would be three-dimensionally distributed, but the ligand concentration was only 32 μM in this experiment. Recalling the dose-dependent response of the BtuB 90-188 pair in Chapter 4, if the data were analyzed in a piece-wise linear form consistent with a measure of substrate loading then the transition point was between 30 and 60 μM. If that is taken as the typical concentration range for BtuB in these experiments, then the substrate is not present here in excess. BtuB also has a very high substrate affinity, in the nm range. Taken together, there should not be sufficient free, spin-labeled substrate to boost the apparent background dimension.

Alternatively, elevated background dimensions are the hallmark of excluded volume effects. The quasi-exponential decay observed in the DEER experiment, and which is removed through background subtraction, arises mainly from instantaneous diffusion (8). Due to the limited bandwidth of the pulses that can be applied in the experiment, only a fraction of the total spins can be flipped by the observe or pump pulses, and thus contribute to the experiment. If spin flip-flops occur between these experimentally relevant spins and other spins which are not affected by the experiment, then the signal is effectively lost via instantaneous diffusion. For this loss to be modeled as an idealized stretched exponential, the surrounding spins must be homogeneously distributed and all interspin distances must be available. For small proteins this is possible, but as the protein diameter gets larger the minimum interlabel distance also increases and the corresponding high frequency decay at the beginning of the trace is reduced.

Even for moderate protein sizes with diameters of 4 to 6 nm, this reduction in the signal decay can persist for several μs and long enough to affect the entirety of the traces in Figures 3 and 4 (8). Literature examples of excluded volume effects are limited to soluble proteins and one-dimensional polymers, but generally find that the background decay occurs with a higher apparent dimension. For human serum albumin (HSA), affected DEER traces had to be fit with a background dimension of 3.74 (12, 13). HSA has a diameter of about 5 nm, slightly larger than a direct BtuB-BtuB contact. Still, AFM studies on the surface of *R. denitrificans* found that the OM porins were separated either by the protruding residues of their aromatic girdles or by single

molecules of LPS. The width of the latter, taken from the bilayer simulation shown in the introduction, was about 1.5 nm. Thus, the total inter-protein contact distance would be about 4-6 nm and the increase of 1.0 in dimensionality for the trace shown in Figure 4 is consistent with that seen for HSA.

One approach to dealing with excluded volume effects is to switch the analysis method to a model-based approach with Gaussian distance components. This has the advantage of allowing the background to be cofit with the distance distribution, assisting with its determination. This approach has been applied using the GLADD and DD software packages from Eric Hustedt (14, 15). These packages also contain a specialized background function meant to model excluded volume effects, although this variant was developed in (14) and (8) for three-dimensional systems, and was not found to benefit the traces shown here. A model-free approach has an additional benefit, which is the ability to have unique widths for each peak. Both presented datasets showed poor fit agreement at the end of the trace, which appeared to arise from the constrained, narrow width of the long-distance component. In a Gaussian approach, these peaks are decoupled from the narrow, short-distance components and a better estimate of the peak width may be possible.

The results of refitting the data shown in Figure 4 for BtuB 188 and the spin-labeled cobalamin using a model-based approach in DD are shown in Figure 5. A single-Gaussian fit is shown in blue, and a two-Gaussian fit is shown in black. With one component, the fit had an apparent background dimension of 2.4 and has noticeably poor agreement with the data, with the residuals (Fig. 5A inset top) still containing a strong oscillation. The two-component fit, meanwhile, shows no oscillation in the residuals (Fig. 5A inset bottom), but fit with an apparent dimensionality of 3.3. Both fits show the same short-distance peak, which is consistent with the crystal structure prediction (pink histoGram). For the two-Gaussian fit, the long-distance peak moves relative to the DeerAnalysis fit, with a new position of 6.5 nm. The width also increases, with a σ of 11.5 Å.

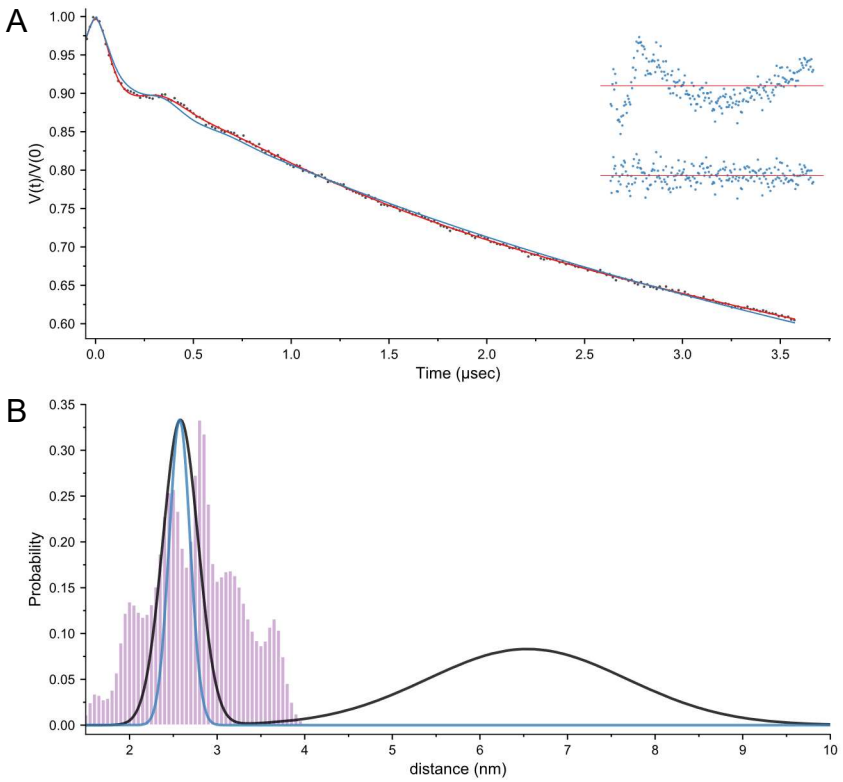


Figure 5. The (A) dipolar evolution, and (B) Distance distribution for RK5016 BtuB 188 and a spin-labeled cobalamin substrate analyzed using model-based fitting with Gaussian peaks in DD. The background dimension was 3.3. The elevated background dimension again may indicate a crowded system with excluded volume effects. The fit with a single Gaussian component is shown in blue, and with two Gaussian components in black. The single-Gaussian fit shows a clear oscillation in the residuals (inset top) which is not apparent in the two-Gaussian fit (inset bottom). The two-Gaussian fit also shows a long-distance peak at a similar position to that observed in the DeerAnalysis fit, but it is now much broader.

The substantial increase in width would suggest that this peak is likely an intermolecular effect, with heterogeneous contacts or aggregation of BtuB monomers. If this is the case, then the effect should be observed for single mutants of the protein as well.

To this end, DEER data for singly labeled BtuB 90 and BtuB 188 in RK5016 cells are shown in Figure 6. Cofitting in DD with a variable dimension background and a single Gaussian component produces similar results for both 188 (teal) and 90 (purple). Both fits have apparent background dimensions of 3.3 or 3.4, and the peak positions are at 6.2 nm ± 10 Å. It can be seen from the residual inset that both fits have good agreement with the data.

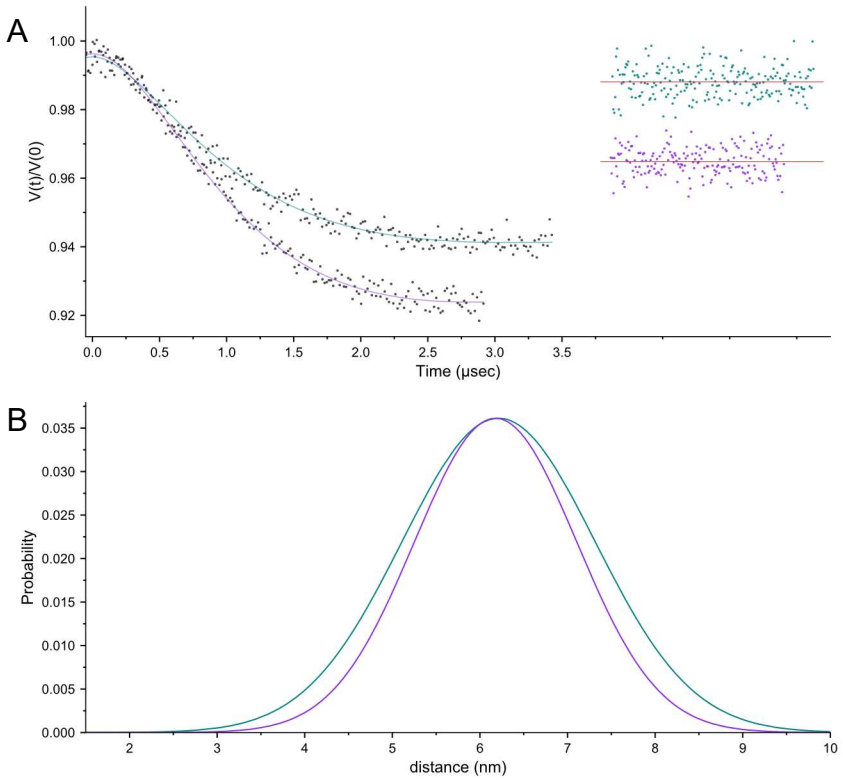


Figure 6. The (A) dipolar evolution, and (B) Distance distribution for BtuB 188 (teal) and BtuB 90 (purple) in RK5016 cells using model-based fitting with Gaussian peaks in DD. The background dimensions were 3.3 and 3.4, respectively. The dipolar evolution is shown at the top in (A) and the distance distribution is shown in the lower panel in (B). Both sites show a similar long-distance peak with position around 6.2 nm and a σ of one nm. The residuals shown in the plot inset indicate that both fits have good agreement with the data.

There are, again, several possible explanations for this result. Sites 90 and 188 are on opposite sides of the protein, and so if the peaks are taken to represent a particular interaction, then the proteins must be aligned head-to-tail. Any significant twists out of such an alignment would produce markedly shorter distances for one site over the other. Alternatively, the peaks may represent the contact distance for fully nonspecific crowding of the protein. If the proteins show no preferential alignment during close contact, then no site-specific alterations in peak position would be expected. Finally, the coincidence of the two peaks may reflect insufficient time to resolve the true distance.

One way to separate these effects would be to modulate the degree of crowding and organization on the surface. This could be done experimentally, with dilution of the spin-label with a nonparamagnetic reagent, or physically by changing the point at which the cells are harvested,

and by extension the amount of time for protein expression and surface incorporation. For the former case, the results of 50/50 dilution of MTSL label with the diamagnetic analogue are shown in Figure 7. The MTSL condition is shown in purple, and the 50/50 dilution is shown in teal, where the dilution cuts the modulation depth in half. The modulation depth is typically expressed as scaling with the square of the labeling efficiency, and so the modulation depth would be expected to drop to a quarter of the original value. The actual reduction was by half, indicating that the diamagnetic analogue is not as efficiently incorporated as the MTSL, and that the reduction in labeling efficiency was less than 50%.

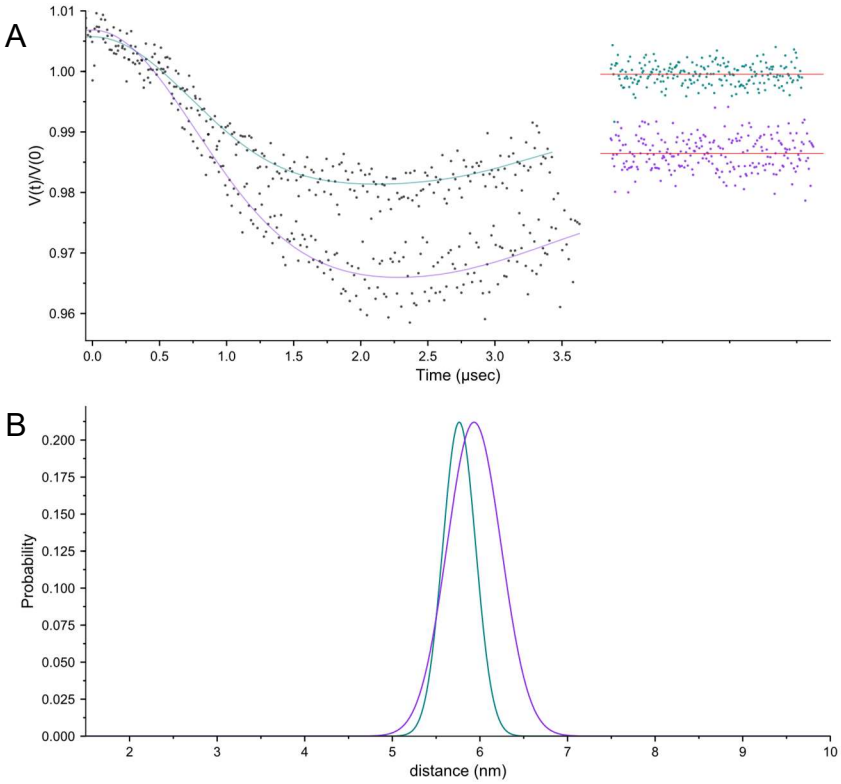


Figure 7. The (A) dipolar evolution, and (B) Distance distribution for BtuB 188 in RK5016 OMs with MTSL (purple) and with a 50/50 dilution of MTSL and a diamagnetic analogue (teal) using model-based fitting with Gaussian peaks in DD. The background dimensions were 2.4 and 2.3, respectively. The reduction in apparent background dimension likely reflects a reduction in protein density or aggregation, although the reason for this is not apparent. Both peaks are around 6 nm as before, but the σ values in this case have reduced to half a nm. The modulation depth decreases in half with a 50% reduction in paramagnetic label. The modulation depth scales with the square of the labeling efficiency, and so a 50% dilution with diamagnetic analogue may not produce a 50% reduction in labeling efficiency.

Interestingly, while the positions of the long-distance peaks at 5.9 nm in the full MTSL condition and 5.7 nm in the dilute condition are similar to the previous traces, the widths are markedly narrower. The overall modulation depth for the full MTSL condition is also lower in this

OM growth than was observed in the whole cell traces shown in Figure 6, suggesting that perhaps the growth conditions can affect this distance component. A change in width might suggest a change in protein density, and since BtuB is overexpressed via a leaky promoter this could be controlled by harvesting the cells at different time points during the growth cycle.

A comparison of the long-distance components obtained for RK5016 cells overexpressing BtuB at a series of time points is shown in Figure 8, with the corresponding dipolar traces shown in Figure S1. The times were 6.5, 8, 9.5, 11, 12.5, 14, and 28 hours after the start of growth. The distances bunch together, with a range of 6.2 to 6.7 nm. There may be a small trend towards longer distances with increasing expression time (inset), but the effect is very small with changes of 1 or 2 Å per time point. Given the length of these traces, there is significant error in the position of peaks around 6 nm, and it is unlikely that these differences are statistically significant. The widths of each peak are encoded as the inset error bars, and they too are decoupled from time with all traces showing a similar σ of 13-15 Å.

While the long-distance component did not appear to change as a function of the growth cycle or the length of protein expression, other factors could still influence its form. Over the course of optimizing the whole cell methods presented in this work, a great number of spectra were collected for BtuB 188 in RK5016 cells. Between them, there are variations in buffer pH, growth time, temperature, expression levels, and the presence or absence of the spin-labeled cobalamin. Applying this model-based approach with background cofitting across the larger dataset would then allow for testing of the long-distance components' general stability in response to experimental manipulations.

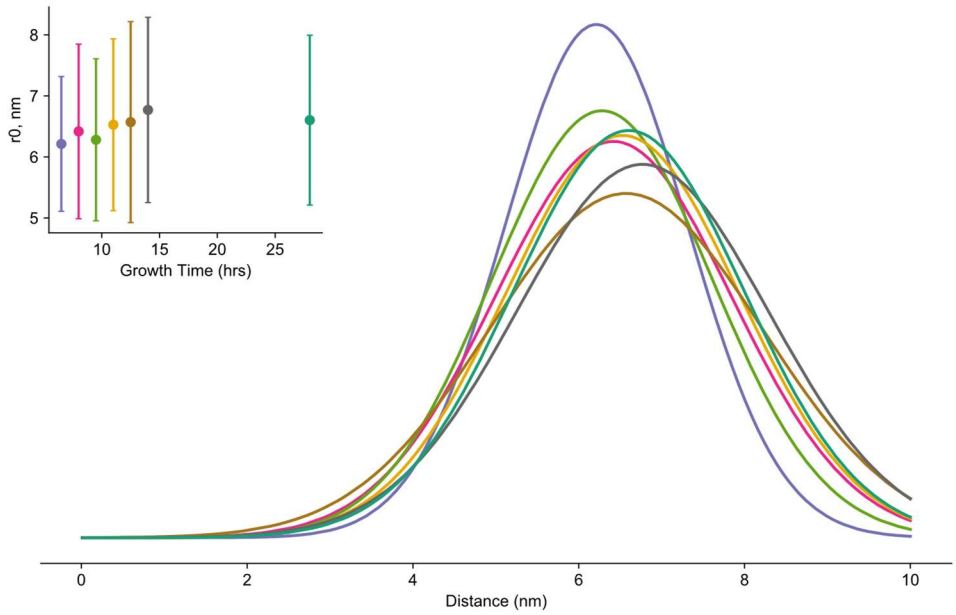


Figure 8. The distance distribution for BtuB 188 in RK5016 harvested at varying times post-induction analyzed using model-based fitting with Gaussian peaks in DD. The growth times were 6.5, 8, 9.5, 11, 12.5, 14, and 28 hours. The position of the peaks are plotted as a function of time in the inset, with the error bars giving the σ values of each peak. There is an apparent trend towards longer distances with time, but this is within error. Overall, the position and form of the long-distance component appears to be independent of growth duration, at least for cells harvested beginning in mid-log phase.

A total of 23 traces were found to fit easily using the model-based cofitting approach in DD. Of these, 16 are single-site controls with BtuB 188 and 7 have BtuB 188 in the presence of spin-labeled cobalamin. The time domain data and dipolar traces for these datasets are shown in supplementary figures 2 and 3, while the distance data are shown below in Figure 9. From the distributions, it is apparent that the position of the long-distance component is nearly constant across growths. There is more variance in the width, but only 2 of the 23 traces show a significantly narrower peak than was observed in the dilution controls. A summary of the fit statistics is given in supplementary Table 1. The apparent background dimension was 3.1 ± 0.1 across all 23 fits, with all fits showing an increase in dimension consistent with excluded volume effects. The long-distance component had an average position of 6.4 nm, and an average σ of 12 Å.

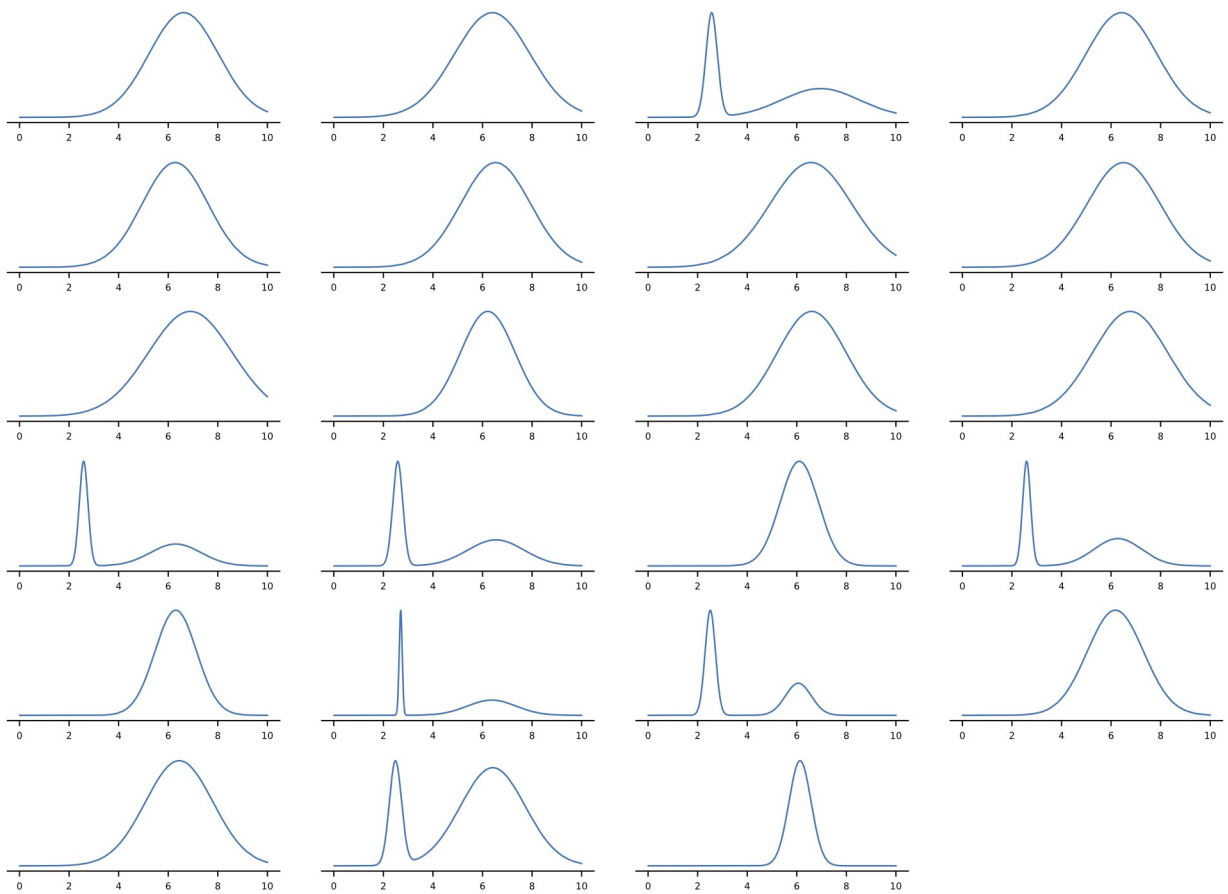


Figure 9. Distance distributions obtained for a total of n=23 distributions representing samples of BtuB 188 grown in RK5016 cells. All distributions were obtained using model-based fitting in DD. Single peak distributions are control samples with BtuB 188 while distributions with two peaks also contained 20-60 μM spin-labeled cobalamin. The position of the long-distance component is highly stable across conditions and growths. A few traces show notably narrower peak widths, but the majority show a consistent σ of 1 to 1.5 nm indicative of broad, heterogeneous interactions between proteins.

With limited variance across growth conditions, the remaining explanations for the constant position and width of the long-distance peak were limited. Either the peak represented the protein-protein contact distance, or the real distance was much longer and the short runtimes prevented it from being properly resolved. In DEER, positional accuracy depends on how much of an oscillation is captured in the time-domain data. This is dependent on the experimental runtime which is in turn dependent on the relaxation rates of the sample and the spectrometer hardware. To reasonably measure distances in the 6-7 nm range, the runtimes needed to nearly double from 3-4 μs to around 7 μs. Modest improvements in runtime can be achieved by using D20 in the solvent buffer, which reduces relaxation due to nearby matrix protons. Almost doubling the runtime required more drastic measures, however, and so BtuB 188 was

grown in RK5016 cells using D_20 and deuterated glycerol as a carbon source.

The resulting data are shown in Figure 10. The target of 7 us was achieved, although only with 48 hours of averaging, and there is still a single, long-distance peak. The apparent background dimension for the fit was 2.7, indicating that the potential excluded volume driven enhancement is still present when the cells are fully deuterated. The position of the peak changed only slightly with this significant increase in runtime, moving to a new value of 6.8 nm, and the σ increased to 18 A. With the increase in width, the peak now covers most of the distance window accessible to the DEER experiment. Such a broad distribution would be consistent with heterogeneous interactions of BtuB across a variety of inter-protein distances and angles.

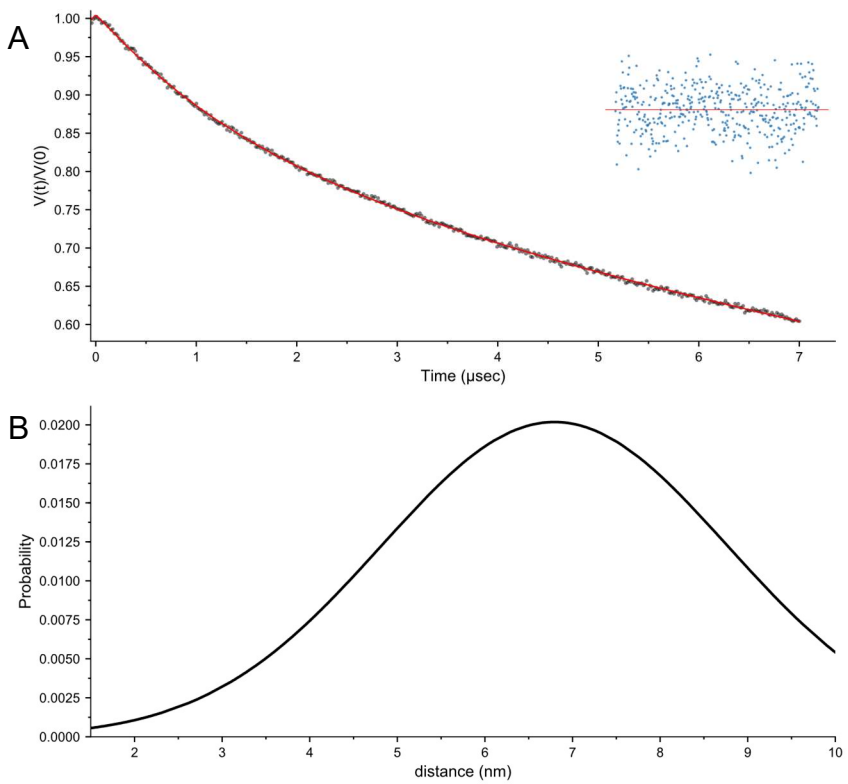


Figure 10. The (A) dipolar evolution, and (B) Distance distribution for RK5016 BtuB grown in deuterated media and analyzed using model-based fitting with Gaussian peaks in DD. The background dimension was 2.7. The peak position is now 6.8 nm, with a large increase in width that leads to the peak covering most of the observable distance window (2-10 nm). The extremely broad peak would be consistent with heterogeneous interactions covering a variety of contact distances and orientations.

With several datasets to work from, it became possible to examine some of the models for aggregation in the OM. As mentioned in the introduction, lipidic (hydrophobic) mismatch has

been observed to drive aggregation of model β-barrel proteins in coarse-grained simulation. Further, this aggregation had the form of linear, string-like aggregates even when the model proteins were composed of highly simplified sequences with only three amino acids (6). This pattern of linear aggregation was also observed in coarse-grained simulations of BtuB at densities determined from fluorescence experiments on BtuB in native surfaces (2) In this latter study, the protein was observed to interact through two hydrophobic interfaces, which form vertically aligned strips of hydrophobic contacts at either end of the protein. In a follow-up experiment these interfaces were used to drive larger scale aggregation of the protein at the mesoscale, and successfully recreated most of the experimentally observed qualities of the protein islands (4).

Fluorescence experiments lack the resolution to examine direct BtuB-BtuB interactions, and so these simulation derived interfaces were the best available comparison for the previously shown datasets. For a string-like aggregation of BtuBs with a defined contact interface, the chain can be thought of as a set of repeating dimers, from which expected distances could be generated in MMM and used to compare to the experimental data. The protein models were generated in pymol from the apo structure of BtuB (PDB id: 1NQG) with the separation distance being the minimum distance that resulted in no overlap of the protein van der Waals surfaces. The BtuB monomers were then rotated to achieve alignment of the hydrophobic strips identified in (6), and the resulting BtuB pairs are shown in Figure 11. With two interfaces per protein, this created three possible orientations, which are defined relative to the 188 site. In the head-to-head interaction (Fig. 11A), the two sites that sit directly below 188 (orange) are aligned together, resulting in a short expected distance. The head-to-tail interaction involves the rotation of one of the two proteins by 180 degrees, creating a mixed interaction with the 188 sites aligned to one side of the pair. This interaction should have an intermediate distance that may align with the experimental results. Finally, the tail-to-tail interaction involves a further rotation of 180 degrees for the remaining protein, resulting in the interaction of the two sites opposite residue 188 (red). This

interaction creates the maximum possible distances for direct BtuB-BtuB contact and may be difficult to resolve in DEER.

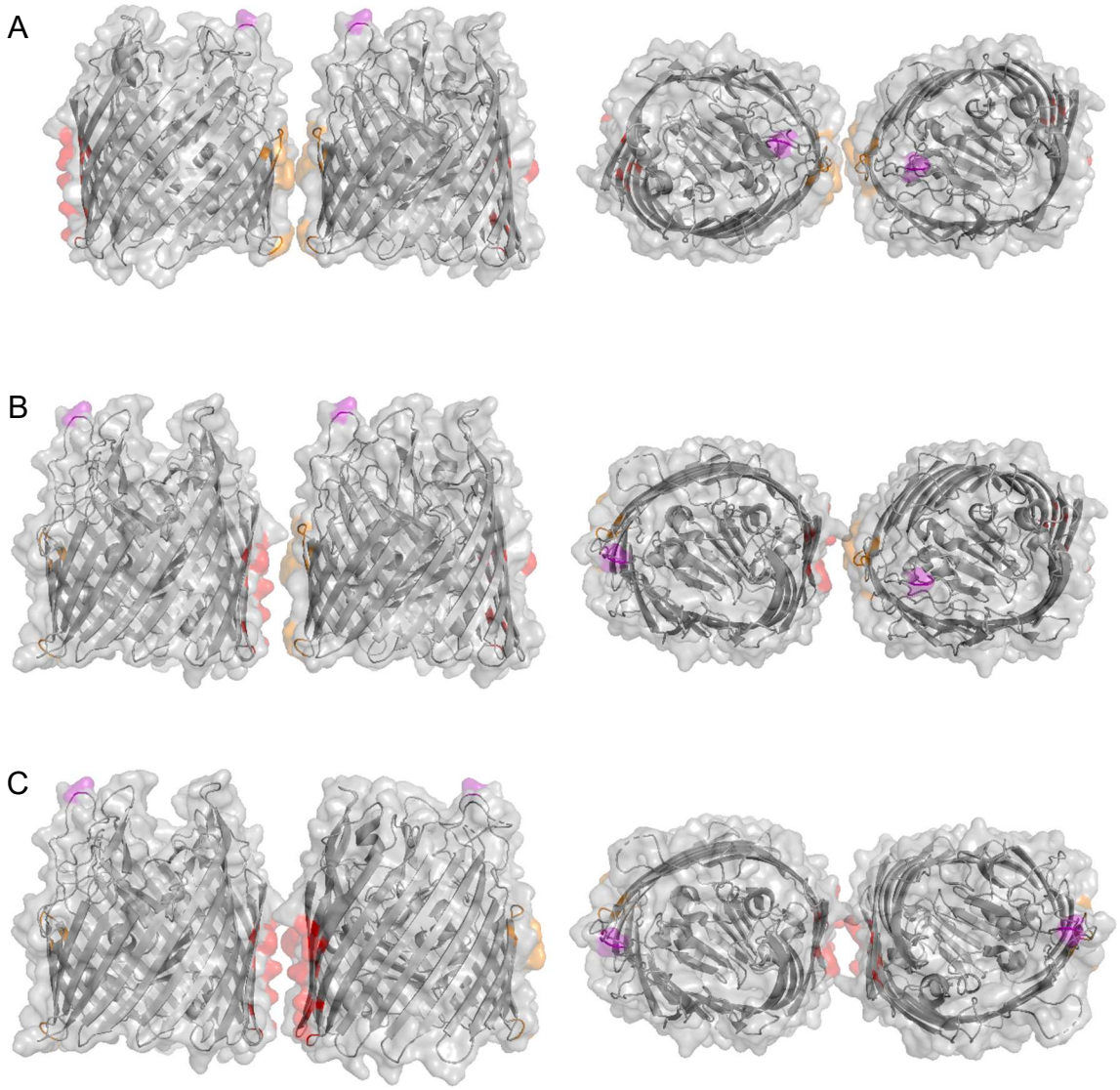


Figure 11. Models of the BtuB-BtuB interaction from the two interaction interfaces derived from coarse-grained simulations in (2). The apo structure (PDB ID: 1NQG) was used to create the interacting BtuB pairs, and the two interfaces are shown in orange and red. The 188 site (purple) sits above the orange interface. The interfaces were aligned in (A) head-to-head, (B) head-to-tail, or (C) tail-to-tail orientations, with the names based on the facing direction of the 188 site. The inter-protein contact distance was set as the minimum separation that did not produce overlap of the protein surfaces.

From these models of the BtuB-BtuB interactions, predicted distance distributions were generated for 188-188 pairs and are shown as histoGrams in Figure 12. A comparison of these predictions with the deuterated 188 trace from Fig. 10 is shown in Fig. 12A. The center of the peak aligns reasonably well with the head-to-tail predicted distances (orange), while the peak

shoulders stretch into the head-to-head (green) and tail-to-tail (purple) distributions. This would indicate that the protein may interact promiscuously with other BtuB monomers, sampling most possible orientations with a weak preference for the head-to-tail interaction. Relative to the expected string-like aggregates predicted in the coarse-grain simulations, this distribution would be consistent with a mixed population of various sized aggregates and distributed monomers.

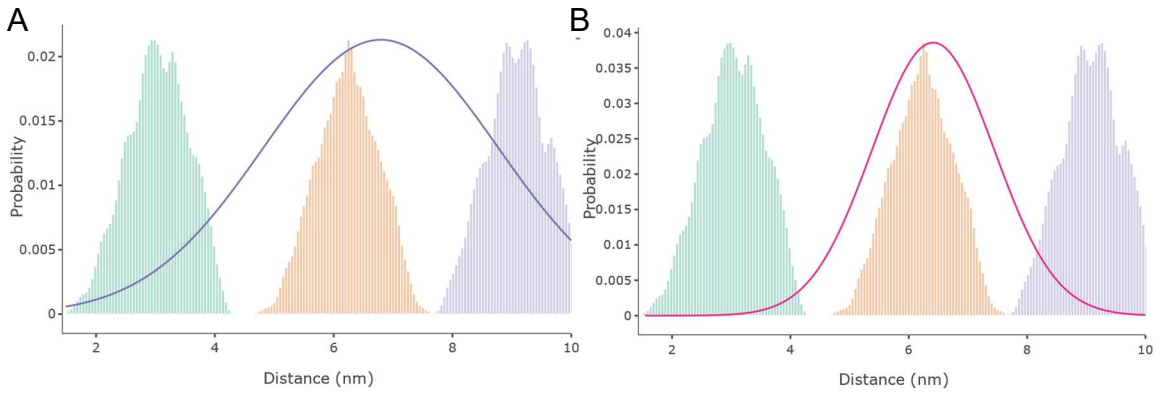


Figure 12. Comparisons of the predicted distances from the three BtuB-BtuB interaction modes shown in Fig. 11 with experimental data. The head-to-head distribution is green, the head-to-tail conformation is pale orange, and the tail-to-tail orientation is purple. (A) a comparison of the three orientations with the deuterated 188 trace. While the peak is broad enough to have statistically significant probability in all 3 modes, the center is close to the head-to-tail orientation. (B) A comparison of a simulated distance distribution representing the average position and width of the 23 experimental distributions shown in Fig. 9 against the 3 orientations. This average distribution also closely aligns with the head-to-tail case, and the decrease in width would imply a strong favoritism towards this orientation.

Similarly, Fig. 12B compares the expected interaction distances with a simulated DEER distribution taken from the average peak position and width for the 23 traces shown in Figure 9. It thus represents the average form of the peak for the shorter length traces that were not deuterated. Again, there is good agreement with the expected distances for the head-to-tail orientation, although this averaged distribution has a reduced width that implies greatly reduced sampling of other rotational states. Overall, this method of analyzing the single site distributions, where DD is used with a single Gaussian component and a variable dimension background, is consistent with the model of head-to-tail aggregates. This orientation also explains the observation that sites 90 and 188 showed the same peak position in Fig. 6, as the repeating nature of this interaction mode creates an identical peak for all sites on the protein. Given the widths of the distributions, there is still substantial variability in the actual contact angle, and this

may reflect either fraying of the aggregate chains, or perhaps more simply a mixture of aggregated and randomly distributed proteins.

The previous discussion centered on what type of distribution might be observed for BtuB overexpressed on the cell surface. This perspective centers on what remains after a traditional background subtraction. Still, for most of the data shown in the previous chapter, these additional, varied long-distance peaks are a hindrance to the extraction of accurate information about changes in the short-distance peaks. In these cases, the ideal would be to determine a different background model or analysis scheme that permits the subtraction of all of these intermolecular interactions in the crowded cell surface, to facilitate the extraction of the intramolecular distances of interest. Again, no fully realized model exists for DEER traces showing excluded volume effects derived from proteins on 2d surfaces, but the conclusions of the 3d treatment should serve as a useful guide.

Figure 13 gives the results of four different background subtractions of the deuterated 188 trace first shown in Fig. 10. The time-domain data are shown in the main plot, with a fit line from the corresponding background model, while the plot insets give the subtracted dipolar traces. The first subtraction used a stretched exponential with the expected dimension of 2.0 for proteins on a flat membrane surface (Fig. 10A). Here, the data decays sharply until about 2 us, after which it increases constantly until the end of the trace. This initially appears to be a partial oscillation, as seen with the earlier traces, but closer inspection shows that the late time behavior is nearly linear. Excluded volume effects cause a reduction in decay rate that should pervade most or all of the traces shown here. Crowding is noted in (8), however, to repopulate distances equal to or greater than the contact distance. In this way, it should restore intermediate to long-time decay of the background, and at high enough concentrations cause a return to near-homogenous behavior in the later portions of the time-data (8). The effects of crowding were noted to become most apparent after 2 us, and so this was tested in Fig. 13B by shifting the start time of the subtraction

to 2 us (8). It is apparent that while a small upward turn exists in the final us of the data, overall the latter portion of the trace is quite well represented by the 2d background.

In addition to crowding, radial offset of the labeled site is also expected to alter the background decay rate. Site 188 is on the barrel, and as such is on the edge of the protein with a near maximal offset from the protein center. This edge position partially repopulates the short label-label distances that were lost due to excluded volume, causing restoration of the homogeneous background behavior. For radial offset, this effect is most apparent early in the trace although it continues to increase the apparent decay rate until 5 or 6 us. Thus, the overall background behavior should consistent of a competition between excluded volume and radial offset effects in the short-time region leading up to 2 us, and a combination of all three effects in the later portions. The exact contributions of the radial offset and crowding contributions will vary with the site and the protein density, and thus each trace may show differences in their early-time versus late-time behavior.

Still, an attempt to observe this is shown in Fig. 13C. The trace was separated into two sections from 0-2 us and 2-7 us, and two stretched exponential backgrounds with dimension 2 were fit piecewise to the data. It can be seen from the recovered dipolar trace, shown in orange and magenta, that while there is still a small oscillation this piecewise treatment succeeds in removing most of the intermolecular oscillation. The difference between the decay constants for the 2 elements was 17%, which may indicate the relative importance of repopulation due to radial offset versus crowding effects. This approach is limited in that it would require considerable manual effort with each individual trace, but it shows that the same processes governing 3 dimensionally distributed particles with excluded volume effects appear to be operating here.

Piecewise background fitting cannot currently be implemented in analysis software for DEER software, and the only currently available alternative to the stretched exponential model is

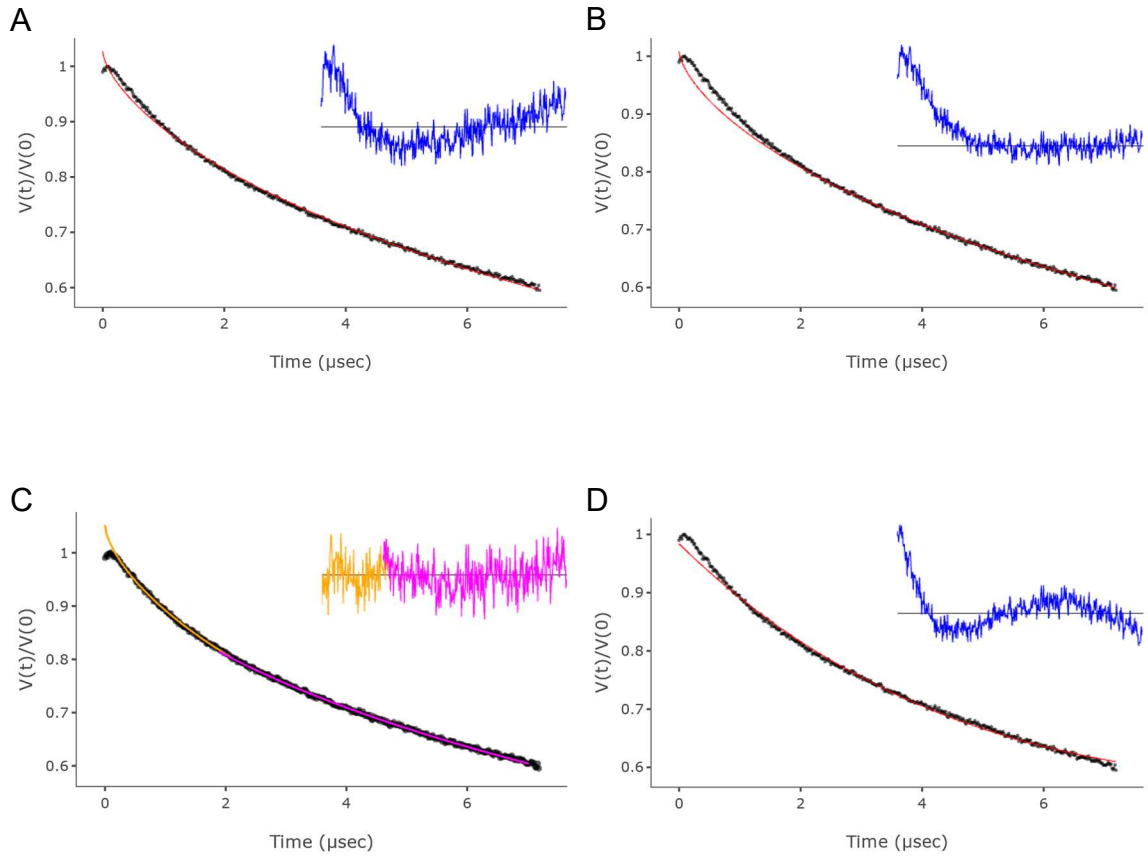


Figure 13. The results of applying different background models to the deuterated 188 trace first shown in Fig. 10. (A) Fitting the expected 2-dimensional stretched exponential to the data starting from the maximum position. (B) Shifting the start time of the background fit to 2 us indicates that the long-time behavior can be approximated by a 2-dimensional stretched exponential. (C) Fitting the short (0-2 us) and long (2-7 us) portions of the trace to separate 2-dimensional backgrounds suppresses most of the apparent oscillation. (D) A quadratic background function with 2 background terms having apparent dimensions of 3 and 6 replaces the partial oscillation seen in A with a full oscillation, indicating that the resulting distance should be decreased.

a quadratic model available in LongDistances. This has the form $e^{b_1+b_2*t+b_3*t^2}$ where b_1, b_2, and b_3 are decay constants with generally negative values. DEER backgrounds are typically quoted as the fractal dimension of the time dependence with 2d backgrounds, for example, having $\frac{2}{3}*t$ dependence. The two time-dependent decay terms in the quadratic background then have fractal dimensions 3 and 6 respectively, and so this model loosely emulates the increase in apparent dimension observed for excluded volume effects.

A fit to the deuterated 188 trace with a quadratic background is shown in Fig. 13D. This background model appears to force a complete oscillation in the data, having a similar shape to the original 2d subtraction in Fig. 13A until 4 or 5 us, where the dipolar trace now inverts to

complete the oscillation instead of continuing upward in the 2d subtraction. Fitting this background corrected trace required 2-Gaussian components, and the result is shown in Figure 14 against the predicted distances for the 3 interaction modes from Fig. 12.

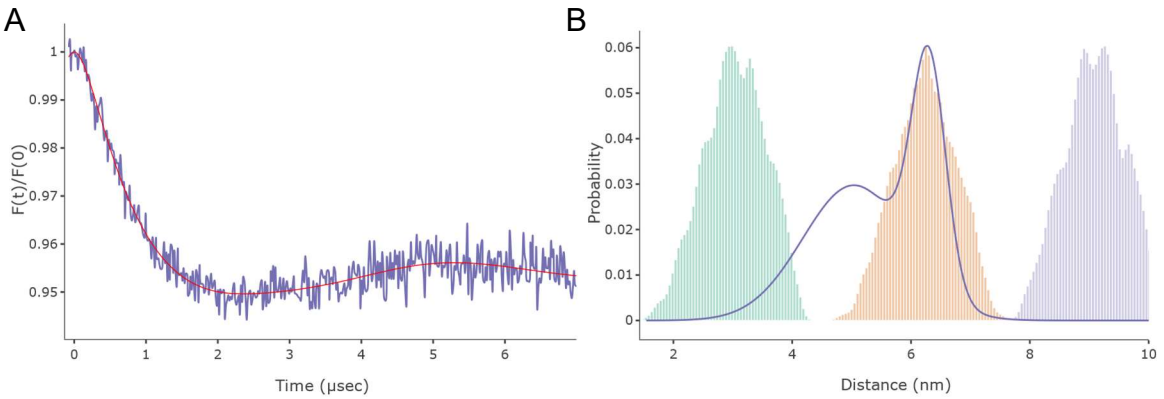


Figure 14. Comparison of the deuterated 188 trace subtracted with a quadratic background function to the three interaction modes presented in Fig. 11. (A) The dipolar trace obtained after subtraction with the quadratic background shows a complete oscillation. (B) The resulting distance distribution fits best to two components, with a short main peak at 6 nm that aligns well with the head-to-tail interaction mode. A second peak is now present at 5 nm, which may indicate poor performance of the quadratic background in the intermediate distance range, selective variance in orientation angle between the head-to-tail and head-to-head modes, or an entirely different form of bimodal associations.

The quadratic fit has good numeric agreement with the dipolar data, and the distances are shifted left with respect to the variable dimension data. There is now a relatively narrow peak that coincides with the expected distances for the head-to-tail interaction mode, and a satellite peak at 5 nm. Theoretically, this satellite peak could be interpreted as indicating that other orientations sampled by the proteins tend only towards the head-to-head interaction type. Still, for data of this length simply adding another term to the stretched exponential may not be enough. Truly dealing with these effects may require radial distribution functions derived either from molecular dynamics or Monte Carlo simulations. The former would be useful if the protein is expected to exist primarily in ordered aggregates, while the latter would provide an approximation of a crowded cell surface with non-interacting particles.

Until such simulations or alternative experimental evidence are available, it would still be useful to have an approximation that aids in the analysis of short-distance components. In Figure 13, it was shown that the decay of the deuterated trace can be split into two portions, with one

ranging from 0-2 us and another from 2-7 us. Also, adding an additional term to the stretched exponential background to create a quadratic background did not appear to fully suppress this difference. With the very short distances observed both for BtuB 90-188 and BtuB 188 with the spin-labeled cobalamin, however, less than 2 us is required to observe a full oscillation. It should then be possible to truncate the traces to 2 us and deal only with one of the two decay modes.

This is applied in Figure 15, which shows the results of applying variable or quadratic background subtractions with and without data truncation in LongDistances. Cofitting of the background dimension was enabled and the data were fit to the minimal number of possible Gaussian components. The dataset is an OM sample from RK5016 cells overexpressing BtuB 90-188. The variable background model with the full dataset is shown in purple, which fit with an apparent dimension of 2.36. The resulting dipolar trace required a minimum of 2 Gaussian components, with one being the expected long-distance component. Truncating the data to 2 us, shown in pink, raises the background dimension to 2.57, but the data still require 2 Gaussian components to fit. In both cases, the dipolar traces are dominated by the partial oscillation corresponding to the long-distance component, and the background dimension is ambiguous. The quadratic background model with the full dataset is shown in orange and can be fit with a single Gaussian component. With the full-time data, the quadratic background does a good job of flattening the trace to produce the expected asymptotic decay behavior, but does so by inverting the long-time partial oscillation. This results in a slight shift of the short component, and a strong mismatch between the fit and the data through the partial oscillation. Finally, truncating prior to using the quadratic background is shown in green. Truncation removes the inverted oscillation present in the full-length case, and the short-distance is now recovered at the same position as in the variable fits. The shape of the dipolar data now matches very closely to its expected asymptotic decay behavior, with the oscillations decaying about a straight horizontal line drawn through the end of the data, as seen for the simulated no-background traces shown in Fig. 2.

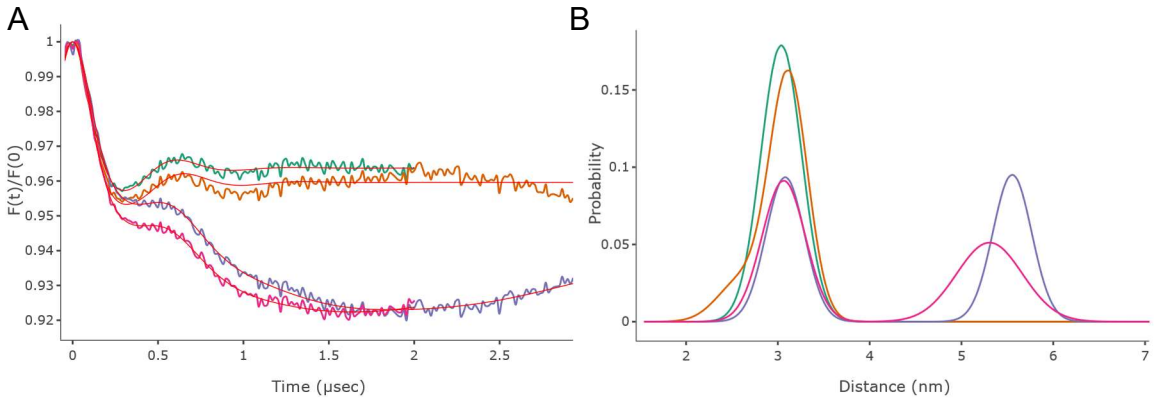


Figure 15. Suppressing the long-distance component through background alteration and data truncation. The dataset was an RK5016 OM sample for BtuB 90-188. The (A) dipolar and (B) distance data are shown for this trace for two background functions and two data lengths. The variable dimension fits with the full-length data (purple) and the data truncated to 2 μs (pink) both require 2 Gaussian components in the fit and display long-distance peaks. Conversely, the quadratic fits to the full-length (orange) and 2 μs (green) data both fit with a single Gaussian component. The full-length result is very poor, however, with the background subtraction producing an inverted partial oscillation. This is removed in the truncated dataset. The poor performance in the later time data indicates that the quadratic background can still only treat the first background decay mode, characterized by excluded volume effects and radial-offset contributions.

With the true nature of the long-distance component still unknown, this makes the quadratic background model in combination with a truncation to remove the second decay regime very attractive. This combination effectively removes the long-distance component while producing subtracted data that has the expected shape and behavior for dipolar decays functions. This method was employed throughout the previous chapter, where the expected distances for the extracellular pairs never exceeded 35 angstroms, and the oscillations could be fully observed even at 2 us.

5.4 Case Study: BtuB 6-510 in cells and Outer Membranes

To conclude this chapter and to illustrate the importance of characterizing the protein organization on the membrane surface, the BtuB 6-510 pair on the periplasmic side of the OM is reconsidered here in Figure 16. The DEER data for BtuB 6-510 in isolated OMs is reproduced in Fig. 16A. The apo and +cobalamin traces were analyzed using DeerAnalysis and had background dimensions of 2.3 and 2.2, respectively. The smaller dimensional enhancement when compared to the 90-188 pair may stem from the location of site 6, which is centrally positioned in the protein without radial offset. Both conditions show the long-distance peak at 5-6 nm discussed in this chapter. For the outer membrane samples, where protein densities and

labeling efficiency are high, this peak is easily explained as a result of intermolecular crowding or aggregation. These physical assumptions also translate to the outer surface of the whole cell preparations, where the labels are still relatively stable. Moving to the periplasmic side of the protein, however, these idealistic assumptions break down. On the periplasmic face the labels are quickly reduced, and the modulation depths of the DEER traces shrink. This implies a significant reduction in labeling efficiency, and the spatial distribution of the still labeled proteins will determine whether or not a long-distance crowding peak emerges. If the labeled proteins are sparsely distributed through the islands, then no peak should be visible. If instead the label remains only on some isolated subpopulation of proteins, say at the cell poles, then the peak may still appear. Thus, the interpretation of distances > 4 nm on the periplasmic face becomes even more ambiguous.

To illustrate this further, the DEER data for the BtuB 6-510 pair in DsbA$^-$ cells are shown with a quadratic background subtraction in Fig. 16B, and with a variable dimension background subtraction in Fig. 16C. LongDistances was used in both cases with model-free fitting, and the variable background dimensions were 2.35 for the apo condition and 2.45 for +cobalamin. Both background models show the same three major peaks between 2 and 4 nm, with no apparent change after substrate addition. The longest peak shifts by a nm between the quadratic and variable fits, however, with a center at 4.7 nm in the former case and 5.6 nm in the latter. The position in the variable dimension fit aligns to the long-distance peak in the OM sample, although the huge difference in modulation depth, from 12% to just over 2%, would imply that the crowding peak should be minimized in the whole cells. Instead, this peak is even more dominant as a percentage of area in the live cell sample. This final peak would set the limit of ton-box extension from the BtuB hatch domain and has strong implications on the transport mechanism.

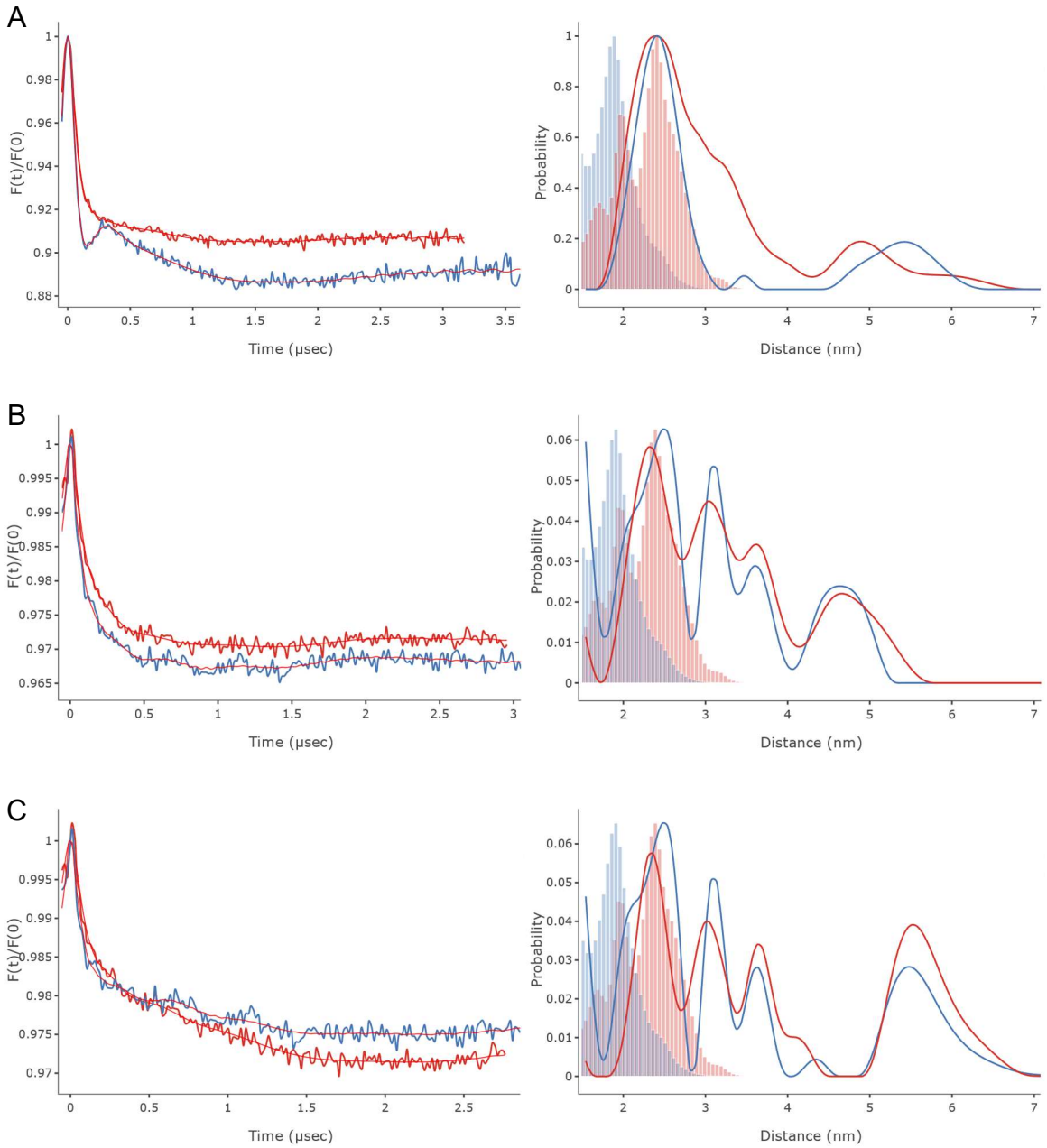


Figure 16. The dipolar (left) and distance (right) data for BtuB 6-510 in (A) RK5016 OM isolates, (B) DsbA$^-$ cells with a quadratic background, and (C) DsbA$^-$ cells with a variable dimension background. The apo condition is shown in blue while the +cobalamin condition is shown in red. The main peak at 2.5 nm is the same across all three, with the +cobalamin condition in the OM showing another peak between 3 and 4 nm that may represent ton-box extension. Both background subtractions for the pair in DsbA$^-$ cells show two peaks that are consistent with extension between 3 and 4 nm. The differences between all three conditions appear in the long-distance component. This is between 5 and 6 nm for the OM sample and for the whole cell sample analyzed with the quadratic background, but the rapid reduction of periplasmic labels in the whole cell call the size of this peak into question. In the quadratic subtraction, the long-distance component moves back to between 4 and 5 nm and is decreased in relative area.

In the *E. coli* ferrichrome transporter, FhuA, DEER distances between site 13 in the FhuA ton-box and similar locations to BtuB 510 in the FhuA barrel displayed peaks out to 4.5 nm, with the most probably distances between 3 and 4 nm. Modeling of these distances using simulated

annealing determined that movement of the first 29 residues of FhuA was required to satisfy the most probable distances, with many more being required to satisfy the edges of the distribution (16). The difference in labeled ton-box position, with site 13 in FhuA versus site 6 in BtuB, means that the BtuB distributions should be shifted to the right, but peaks at 4.5 or 5.5 nm should require the movement of dozens of N-terminal residues in the protein. One model for pore formation in BtuB involves the unfolding of the first 50 residues of the hatch domain, and a peak distance of 5.5 nm in particular would be consistent with this theory.

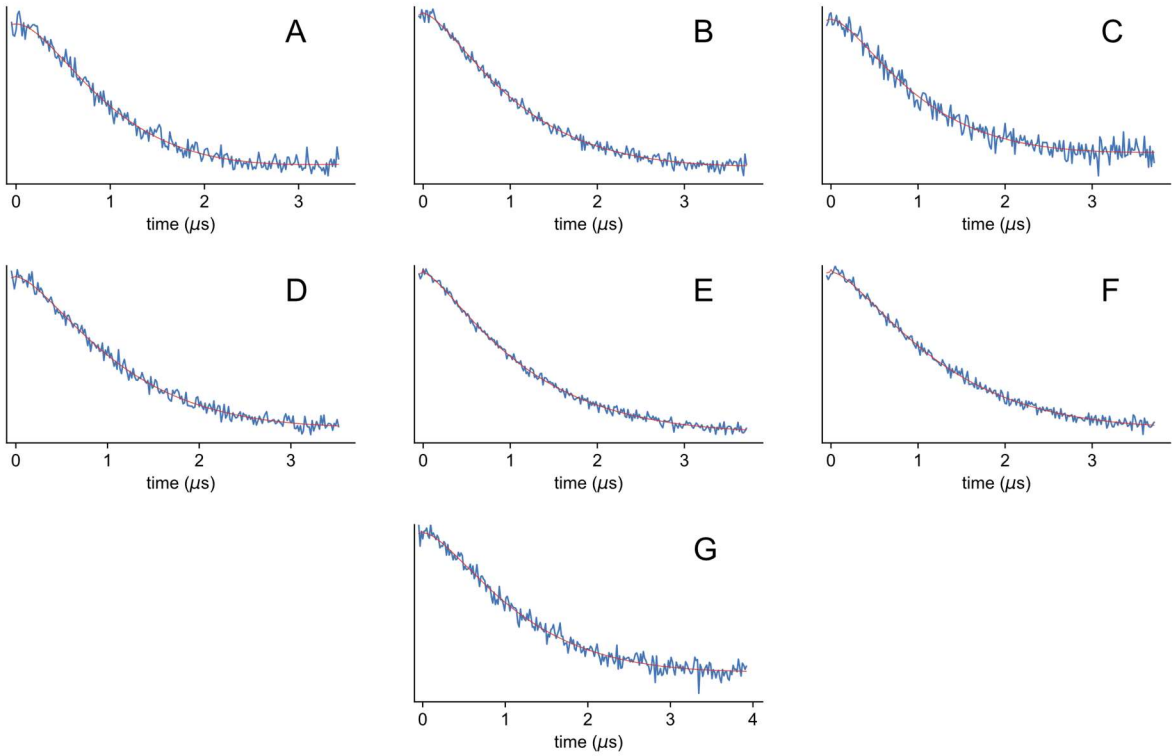
Thus, being able to determine when peaks in the 5-6 nm range arise from intramolecular distances or intermolecular crowding may become critical in the near future. One way around this problem is the use of many sites around the barrel surface, with the hope of only using distances of < 4 or 5 nm. Still, the end goal of a set of periplasmic distance restraints would likely be to model the state of the extended ton-box and its implications on unfolding in the hatch domain, as was done with FhuA (16). This will require several restraints for triangulation of the ton-box segment in space, and it may not be possible to find several pairs with only short distances. Alternatively, dilution of the protein at the growth level through the introduction of WT BtuB plasmids or harvesting the cells even earlier may drop the protein concentrations low enough to minimize the intermolecular peak. Finally, future progress in simulation and modeling may allow for the identification of a radial distribution function for this mode of crowding and potential string-like aggregation, from which background subtraction would be possible. Regardless, future progress will be required both for the understanding of the protein organization on the bacterial cell surface, and for its suppression when attempting to draw quantitative conclusions regarding the transport mode of BtuB and other TBDTs.

5.5 References

1. Fuller-Schaefer, C. A., and R. J. Kadner. 2005. Multiple extracellular loops contribute to substrate binding and transport by the Escherichia coli cobalamin transporter BtuB. Journal of Bacteriology 187(5):1732-1739.
2. Rassam, P., N. A. Copeland, O. Birkholz, C. Toth, M. Chavent, A. L. Duncan, S. J. Cross, N. G. Housden, R. Kaminska, U. Seger, D. M. Quinn, T. J. Garrod, M. S. P. Sansom, J. Piehler, C. G. Baumann, and C. Kleanthous. 2015. Supramolecular assemblies underpin turnover of outer membrane proteins in bacteria. Nature 523(7560):333-+. Article.
3. Ritchie, K., Y. Lill, C. Sood, H. Lee, and S. Zhang. Single-molecule imaging in live bacteria cells. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 368(1611):20120355-20120355.
4. Chavent, M., A. L. Duncan, P. Rassam, O. Birkholz, J. Hélie, T. Reddy, D. Beliaev, B. Hambly, J. Piehler, C. Kleanthous, and M. S. P. Sansom. 2018. How nanoscale protein interactions determine the mesoscale dynamic organisation of bacterial outer membrane proteins. Nature Communications 9(1):2846.
5. Periole, X., T. Huber, S. J. Marrink, and T. P. Sakmar. 2007. G protein-coupled receptors self-assemble in dynamics simulations of model bilayers. Journal of the American Chemical Society 129(33):10126-10132.
6. Parton, D. L., J. W. Klingelhoefer, and M. S. P. Sansom. 2011. Aggregation of Model Membrane Proteins, Modulated by Hydrophobic Mismatch, Membrane Curvature, and Protein Class. Biophysical Journal 101(3):691-699.
7. Cherezov, V., E. Yamashita, W. Liu, M. Zhalnina, W. A. Cramer, and M. Caffrey. 2006. In meso structure of the cobalamin transporter, BtuB, at 1.95 angstrom resolution. Journal of Molecular Biology 364(4):716-734.
8. Kattnig, D. R., J. Reichenwallner, and D. Hinderberger. 2013. Modeling Excluded Volume Effects for the Faithful Description of the Background Signal in Double Electron–Electron Resonance. The Journal of Physical Chemistry B 117(51):16542-16557.
9. Sikora, A., B. Joseph, M. Matson, J. R. Staley, and D. S. Cafiso. 2016. Allosteric Signaling Is Bidirectional in an Outer-Membrane Transport Protein. Biophysical journal 111(9):1908-1918.
10. Joseph, B., A. Sikora, E. Bordignon, G. Jeschke, D. S. Cafiso, and T. F. Prisner. 2015. Distance Measurement on an Endogenous Membrane Transporter in E-coli Cells and Native Membranes Using EPR Spectroscopy. Angewandte Chemie-International Edition 54(21):6196-6199.
11. Joseph, B., A. Sikora, and D. S. Cafiso. 2016. Ligand Induced Conformational Changes of a Membrane Transporter in E. coli Cells Observed with DEER/PELDOR. Journal of the American Chemical Society 138(6):1844-1847.
12. Junk, M. J. N., H. W. Spiess, and D. Hinderberger. 2011. DEER in biological multispin-systems: A case study on the fatty acid binding to human serum albumin. Journal of Magnetic Resonance 210(2):210-217.
13. Junk, M. J. N., H. W. Spiess, and D. Hinderberger. 2010. The Distribution of Fatty Acids Reveals the Functional Structure of Human Serum Albumin. Angewandte Chemie International Edition 49(46):8755-8759.
14. Brandon, S., A. H. Beth, and E. J. Hustedt. 2012. The global analysis of DEER data. Journal of Magnetic Resonance 218:93-104.
15. Stein, R. A., A. H. Beth, and E. J. Hustedt. 2015. A Straightforward Approach to the Analysis of Double Electron-Electron Resonance Data. Electron Paramagnetic Resonance Investigations of Biological Systems by Using Spin Labels, Spin Probes, and Intrinsic Metal Ions, Pt A 563:531-567.

16. Sarver, J. L., M. Zhang, L. S. Liu, D. Nyenhuis, and D. S. Cafiso. 2018. A Dynamic Protein-Protein Coupling between the TonB-Dependent Transporter FhuA and TonB. Biochemistry 57(6):1045-1053. Article.

5.6 Supplementary Figures



Figures S1. Summary of the dipolar traces obtained from samples of BtuB labeled at site 188 from different points in the cell growth cycle. All samples were incubated at RT and pH 8.0. Cells were harvested at the following times after inoculation of the main culture: (A) 6.5 hours (B) 8 hours (C) 9.5 hours (D) 11 hours (E) 12.5 hours (F) 14 hours (G) 28 hours. An increase in growth time, or transition from log phase to stationary phase in the growth cycle, does not appear to change the nature of the observed distance distribution.

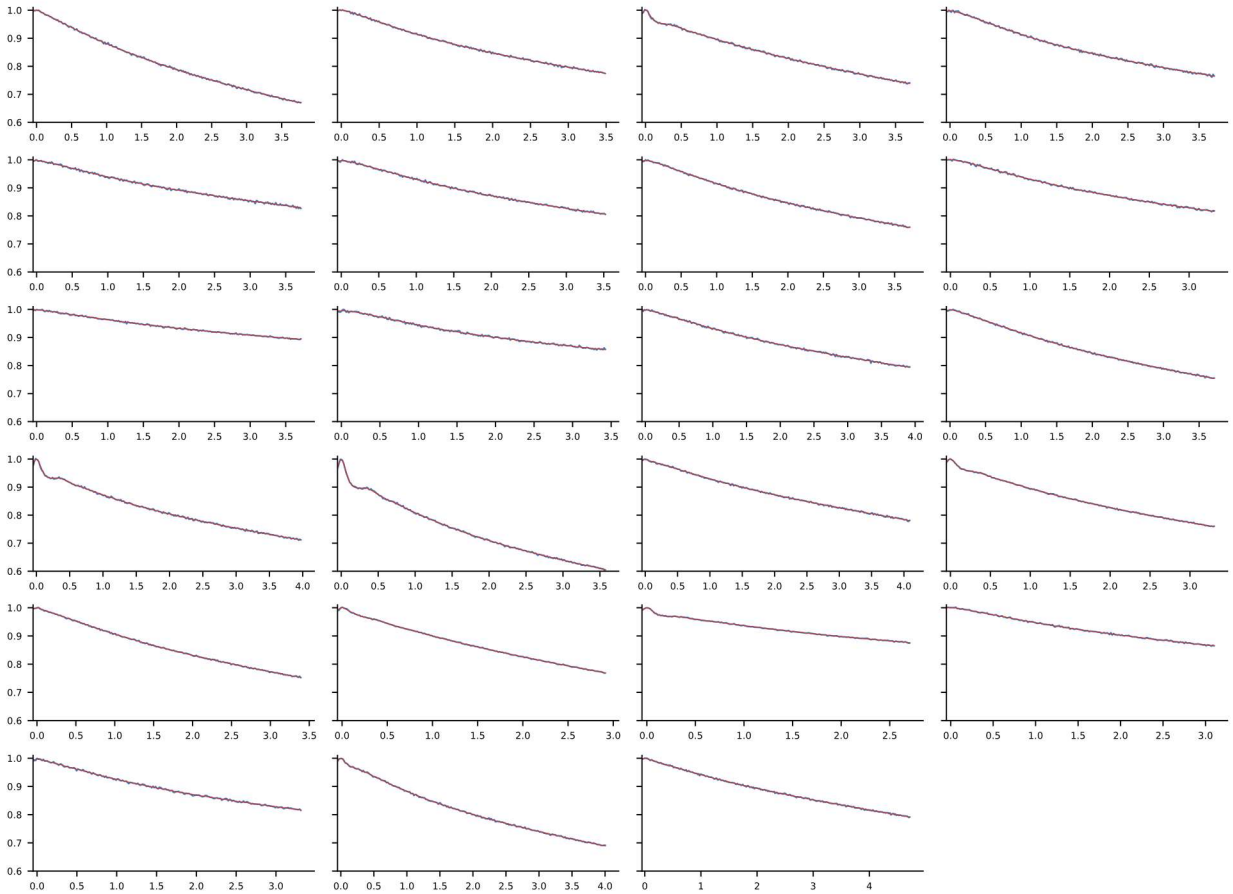


Figure S2. Time-domain data obtained for a total of n=23 distributions representing samples of BtuB 188 grown in RK5016 cells. The data correlates to the distance distributions in Fig. 19 in the main text.

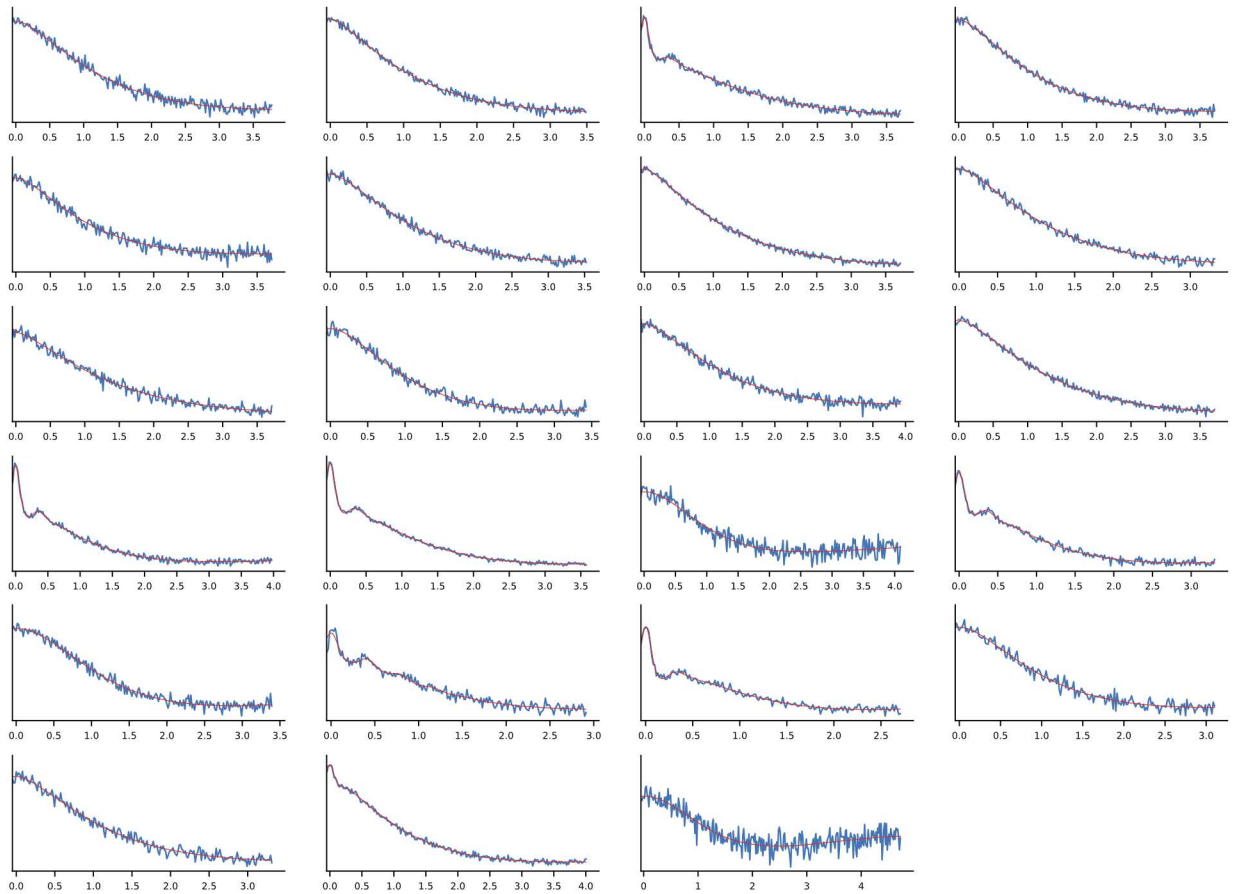


Figure S3. Dipolar traces obtained for a total of n=23 distributions representing samples of BtuB 188 grown in RK5016 cells. The data correlates to the distance distributions in Fig. 19 in the main text.

	λ	d	r_0	σ_0	r_1	σ_1	χ^2
$\bar{x}$	4.73 ± 0.029	3.13 ± 0.064	25.74 ± 0.26	1.83 ± 0.23	64.30 ± 0.50	12.17 ± 0.71	0.68 ± 0.032

Table S1. A summary of the DD fit statistics for 23 PELDOR data sets for BtuB labeled at site 188 in the presence or absence of the spin-labeled cobalamin**.** Values are reported plus or minus the standard error of the mean. λ and d represent the leading factor and dimensionality of the exponential background used in DD. r_0 and σ_0 are the center and standard deviation of the short-distance component for the 7 samples with labeled cobalamin while r_1 and σ_1 are the center and standard deviation of the long-distance component for all 23 samples having BtuB labeled at site 188. The average chi squared of the fits is less than 1, in accordance with guidelines for fit selection in DD.

Chapter 6: Future Directions

6.1 Introduction

One of the major limitations of the previous two chapters was the number of sites in the hatch and barrel domains of BtuB. For the hatch, the data was presented primarily for the 74 and 90 positions, while the barrel used residues 188 and 399. The long timeframe of development of the whole cell EPR method presented in this work meant that dozens of measurements were collected for sites 90 and 188, both together and separately, but these cannot substitute for the benefits of having additional sites throughout the protein. Once the method had been standardized, the next direction was thus to obtain a mix of sites in the hatch and barrel. The CW spectra could be used to look for regions of change in response to cobalamin binding, and the DEER data could triangulate the motion observed between 90 and 188, or potentially indicate additional sites of substrate-dependent conformational changes. A total of 10 additional hatch sites and 9 additional barrel positions for grew and labeled successfully, and preliminary results are explored in the following sections.

6.2 Additional Methods

Growth of DsbA$^-$ cells overexpressing all new mutants to OD 0.3 was followed by resuspension in 100 mM HEPES, pH 7.0. Labeling proceeded for 30 minutes with 0.1 mg of MTSL and was followed by 2x 15-minute wash steps at RT with 100 mM HEPES, pH 7.0. All buffers now contained 2.5% glucose. 140 μM cobalamin was added for the + cobalamin conditions.

Results and Discussion

6.3 Additional Hatch Sites

In addition to sites 74 and 90, the following sites produced CW spectra in DsbA$^-$ cells: 55, 57, 62, 63, 65, 67, 72, 91, 93, and 98. The locations of these new sites are shown in Figure 1. The 90 position shown frequently throughout chapters 4 and 5 is shown in orange, and the 10 new positions are shown in blue. Together, they surround the cobalamin substrate (Fig. 1A) and make up almost the entire upper surface of the hatch domain (Fig. 1B). Based on the position of

the sidechain β-carbon, all but sites 55 and 98 may be in direct substrate contact upon binding. Additionally, all reside in crowded environments and should produce immobilized components in the resulting CW spectra.

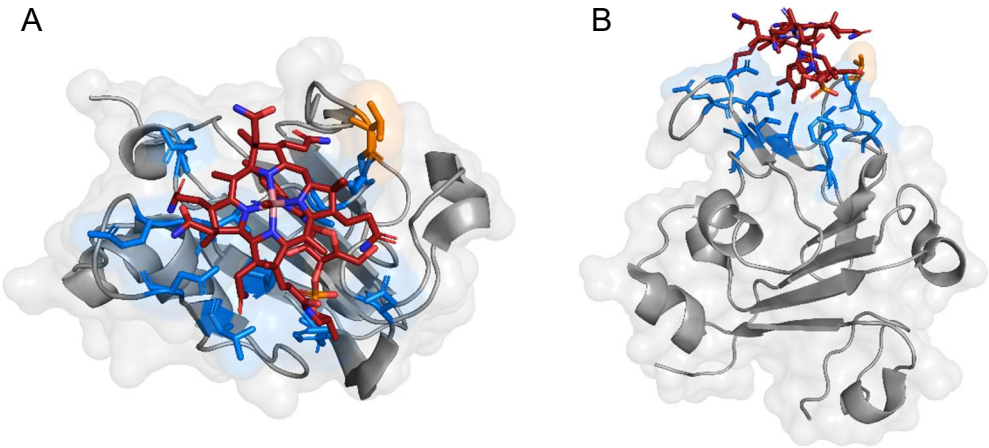


Figure 1. Locations of the new mutant sites in the BtuB hatch domain. The original 90 location is shown in orange, and the new positions are shown in blue. The cobalamin substrate is colored dark red. (A) a top view of the hatch domain shows that the new sites fully surround the substrate, allowing future DEER experiments to isolate movement in any direction. (B) a side view showing that the new sites comprise most of the upper hatch domain. These positions were selected based on sidechain directionality, with sidechains pointing up and out of the protein being predicted to be more easily labeled.

Spectra were obtained in the absence (blue) and presence (red) of cobalamin for each site in DsbA$^-$ cells and are shown in Figure 2 alongside the previous sites 74 and 90. All spectra are normalized and represent 10 scan averages with the wavelet denoised transforms plotted overtop the noisy data. Based on noise level, the spectra can be broken into two groups with sites 55, 65, 67, and 98 labeling poorly, and the remainder having similar performance to the previously characterized sites 74 and 90. All locations have multicomponent spectra, with one component having a nearly 70 Gauss splitting indicative of rigid limit behavior, and another component being somewhat more mobile. This reflects the crowded environment, with most of the possible MTSL movement blocked by adjacent protein sidechain and backbone atoms. There is some variance in the mobile component, with sites 63, 72, and 91 displaying different motional schemes. Addition of cobalamin causes a shift towards the immobile component in several spectra, notably at sites 63, 72, 74, 91, and 93. The +cobalamin spectra are very similar at all sites, indicating that

the lack of change in site 90, for example, may simply reflect a lack of sufficient mobility in the apo state to resolve further immobilization with the addition of substrate.

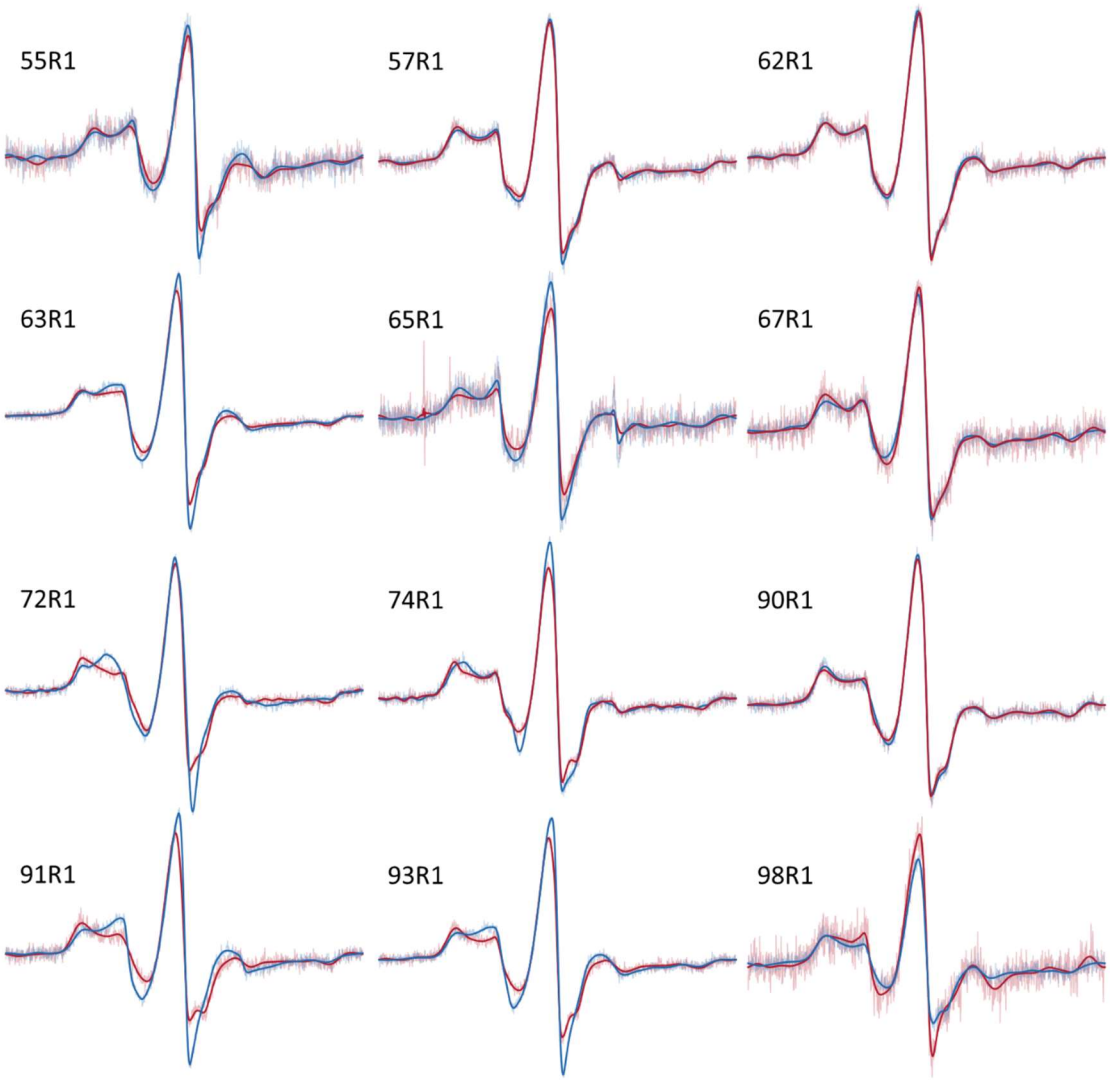


Figure 2. CW spectra obtained for sites in the BtuB hatch domain in DsbA$^-$ cells. All spectra are multicomponent, indicating that at least 2 motional regimes are available to the spin label, and all contain a highly immobile component indicative of the crowded environment inside the hatch. Changes are observed at several sites upon substrate addition, with the mobile component transitioning towards the immobile one. For sites that already strongly favor the immobile component in the apo condition, such as site 90, no change is observed. This implies that much of the additional ordering comes from rotameric restriction from the bulky substrate. All spectra are normalized averages of 10 scans and the noisy data are overlayed with the wavelet transform.

To look at the structural significance of the substrate induced shifts, the spectra are replotted around the hatch domain in Figure 3. Site 90 is again shown in orange and was noted in Chapter 4 to be located on a loop that undergoes a conformational shift between a partially helical character in the apo crystal structure, and a fully unstructured conformation in the +cobalamin structure. The new sites 91 and 93 are on the opposite side of the loop to the residues that participate in this helix (85-88) but do display substrate dependent shifts in the CW spectra. This may reflect increased motion further along the loop but given the lack of change in site 90 may simply point to the establishment of direct contact between the substrate and label. Moving clockwise, Sites 72 and 67 sit at opposite ends of a 4-residue turn that lies alongside the cobalamin ribose sugar and phosphate group. Both moieties are bulky and would be expected to reduce the rotamers available to the side chain at these locations, although the change in 67 is much less apparent than at the 72 location. The large shift in the spectrum of 72 is interesting on account of the lack of change in the DEER distribution for the 74-188 pair and could be an excellent follow-up to see if more local shifts are occurring in the hatch domain.

Sites 67, 65, and 55 show similarly poor labeling. Together with site 98, these residues are the furthest towards the hatch interior, and indicate that the center of the hatch may be inaccessible in the native environment. Finally, moving towards the opposite side of the hatch from site 90, only site 63 shows a significant shift in response to substrate binding. The sidechain of residue 63 points directly towards the 5,6-dimethylbenzimidazole (DMBI) ligand in the cobalamin substrate, which likely explains its rigidization. Sites 57 and 62, contrastingly, point away from the substrate and sit on a loop at the top of the hatch domain. Site 98 was the most interior location to label, being a nm or more away from the cobalamin binding site and may show a slight shift towards the mobile component with cobalamin addition. This would make an interesting target for DEER experiments, but it may not be possible to get sufficient labeling efficiency for DEER at this location.

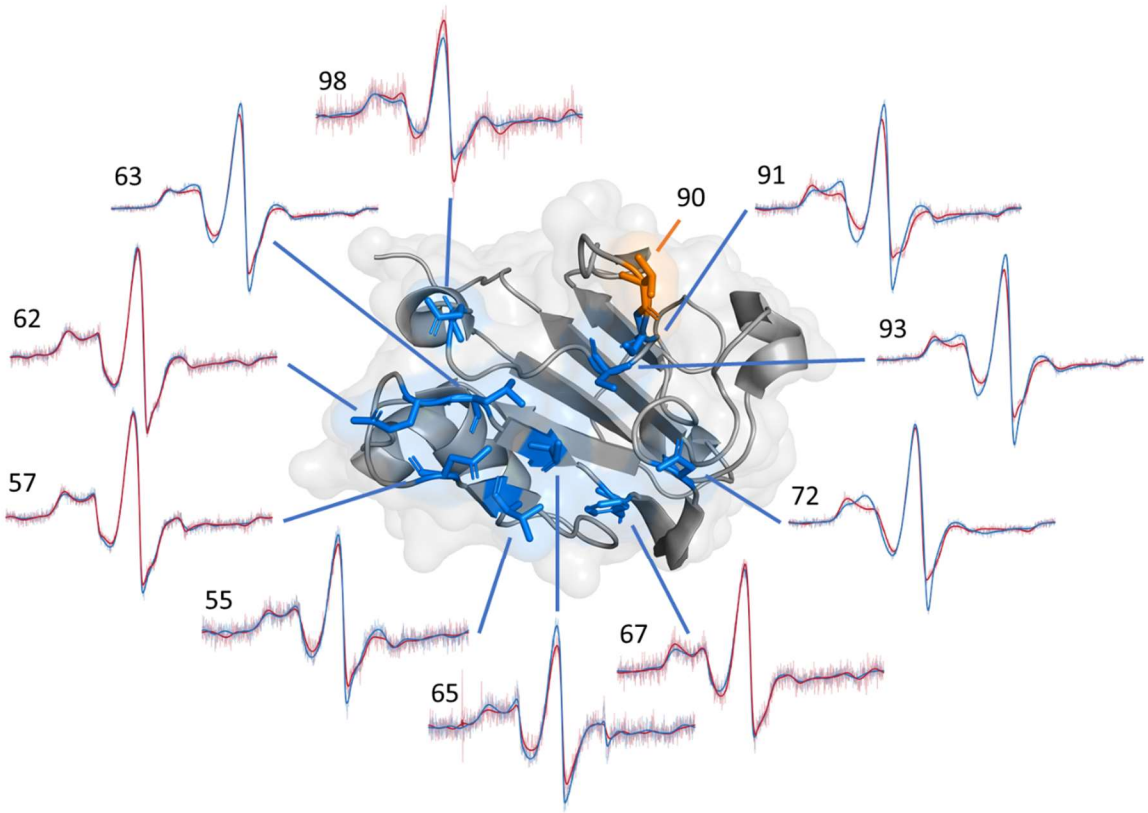


Figure 3. The hatch domain CW spectra as a function of their lateral position in the protein. From this perspective, changes are predominantly seen in the right side of the hatch domain. This is the location of the bulky phosphate, ribose, and DMBI groups, and direct contact between the substrate and label may be producing much of the immobilization. Sites 91, and 93 lie on the same loop as site 90, which was observed to shift in the 90-188 DEER data in chapter 4. Sites 72 and 74 (not shown) are on one end of a 4-residue turn next to the substrate phosphate, with site 67 on the other. The large shift in site 72 was not corroborated by the DEER data for 74-188, which showed no changes in either distance component, suggesting that all of the CW change comes from interaction with the nearby ribose sugar. Sites on the left side of this view, which are typically further away from the substrate show little change upon substrate addition. The spectra also fall into two distinct labeling regimes, with more distal sites (63,91,93,72) showing less noise and better labeling efficiency than the remaining residues, which sit lower in the domain. All spectra are normalized averages of 10 scans and the noisy data is overlayed with the wavelet transform.

6.4 Additional Barrel Sites

In addition to new hatch locations, nine barrel locations around the extracellular face of the protein were labeled successfully in DsbA$^-$ cells, and the CW spectra are shown in Figure 4. There were two main goals with these new barrel sites. The first was to identify additional positions that like 188, did not appear to shift in response to substrate. This would allow them to be used as controls for triangulation of motion in the hatch domain. The second was to identify sites that showed strong substrate-dependent response, and that could be combined with the control sites for measurements of gating motions in the loops. These measurements have been

performed previously in OMs, but not in the whole cells (1). Looking at Fig. 4, sites 237, 330, 404, and 534 show minimal substrate response. This should make them ideal starting points for DEER experiments with hatch domain residues. Of course, there is the possibility that these sites are in constant, unaltered motion, but this should merely increase the widths of the resulting DEER distributions without creating differences in position. Conversely, sites 403, 451, 491, and 493 show strong immobilization upon substrate addition and may provide additional information on loop motions in the native environment.

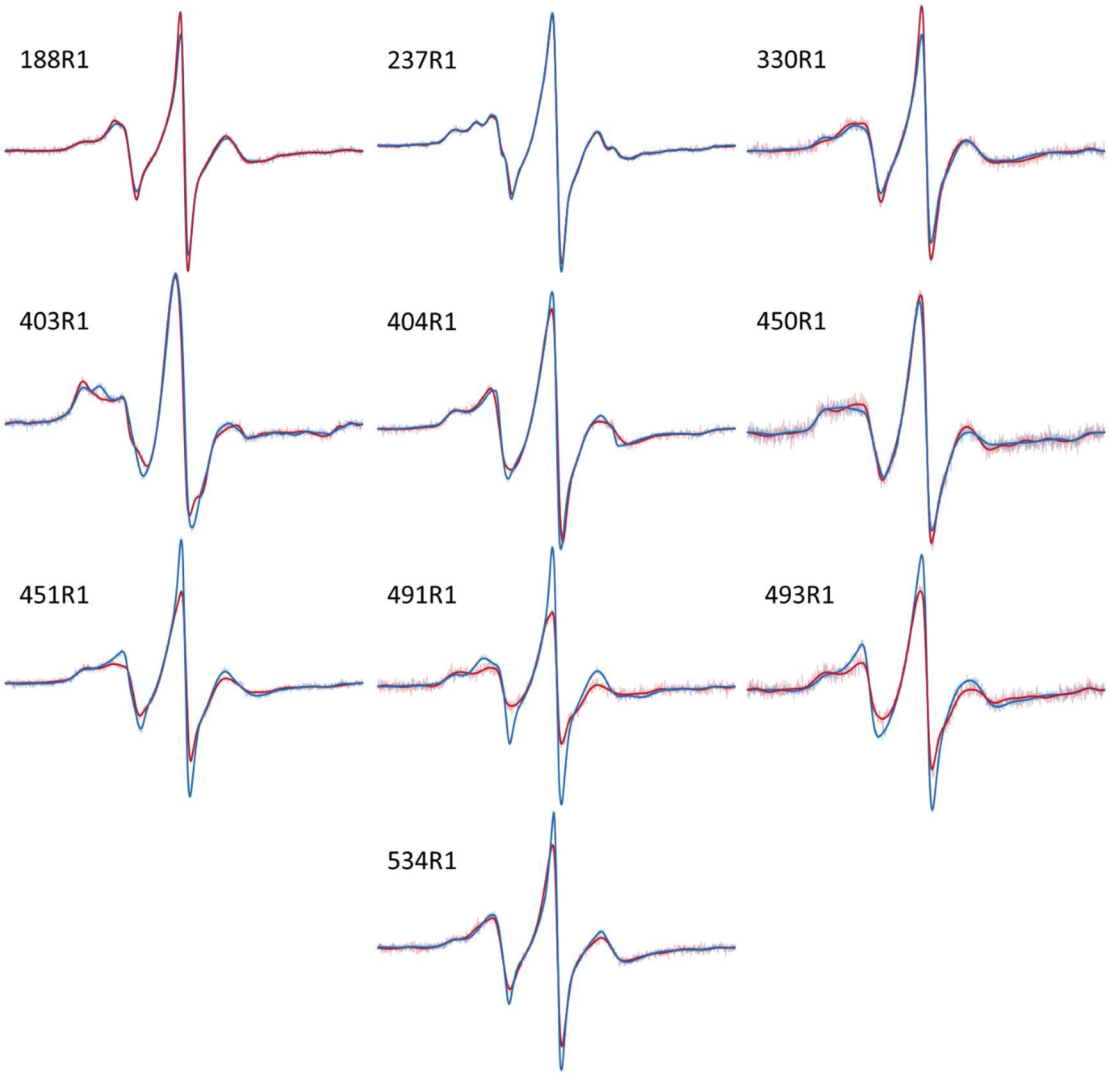


Figure 4. CW spectra obtained for sites in the BtuB barrel domain in DsbA$^-$ cells. All spectra are multicomponent, owing to the ability of the loops to contact each other, other proteins, and the LPS, all of which may lead to distinct motional modes. For DEER spectroscopy to the hatch domain, ideal sites would show no movement with the addition of substrate. This applies to sites 237, 330, 404, and 534 shown here. The loops that show strong changes in response to substrate may instead be ideal for interloop measurements. All spectra are normalized averages of 10 scans overlaid with the wavelet transform.

The barrel spectra are plotted as a function of their position in the protein in Figure 5. Starting from site 188 and working clockwise, residue 237 is on the adjacent loop to 188 and also shows no apparent change in the CW spectra upon substrate binding. This loop is unresolved in the apo structure and fully unstructured in the substrate-bound structure. The spin label sits at the apex of the loop and may form transient interactions with other elements of the protein or the surroundings. Still, the flexibility indicates that its motions are likely uncoupled from substrate binding and capture. This site may work for DEER measurements, but the length and motion may produce extremely broad distributions that could hinder its use as a control site. Next, site 330 sits nearly opposite the 188 site, and is part of series of 3 associated loops containing sites 330, 403, 404, 450, and 451. The 330 spectrum shows only a small change upon substrate addition, although this residue points up and slightly away from the rest of the protein, minimizing any changes that would be observed with loop motion.

The adjacent loop contains residues 403 and 404, with the former showing a strong immobilization upon cobalamin addition and the latter remaining static. This loop is highly structured, with a 2-turn helix element and strong lateral contacts to the 450/451 loop. 404 sits at the loop apex and points upward, while 403 points towards the 403 loop. Thus, the change in 403 may indicate proximity of the 330 loop. A similar relationship is observed for sites 450 and 451, where 450 is static and 451 immobilizes with substrate. This loop contains one of the extracellular calcium binding sites, which likely immobilizes the backbone, and residue 450 points outward toward the calcium ion, likely contributing to its comparative immobility. Residue 451, in contrast, points inwards towards the phenylalanine and tyrosine residues at positions 404 and 405, respectively, and may form differential contacts with them in response to substrate binding.

Residues 491 and 493 lie on the left side of Fig. 5A and are situated on a short loop that projects outward from the barrel. This loop also contains residue 488 which was previously shown to be static with respect to substrate binding in DEER measurements to site 188 in OM isolates (1). Both 491 and 493 show immobilizations with substrate binding, and the simplest explanation

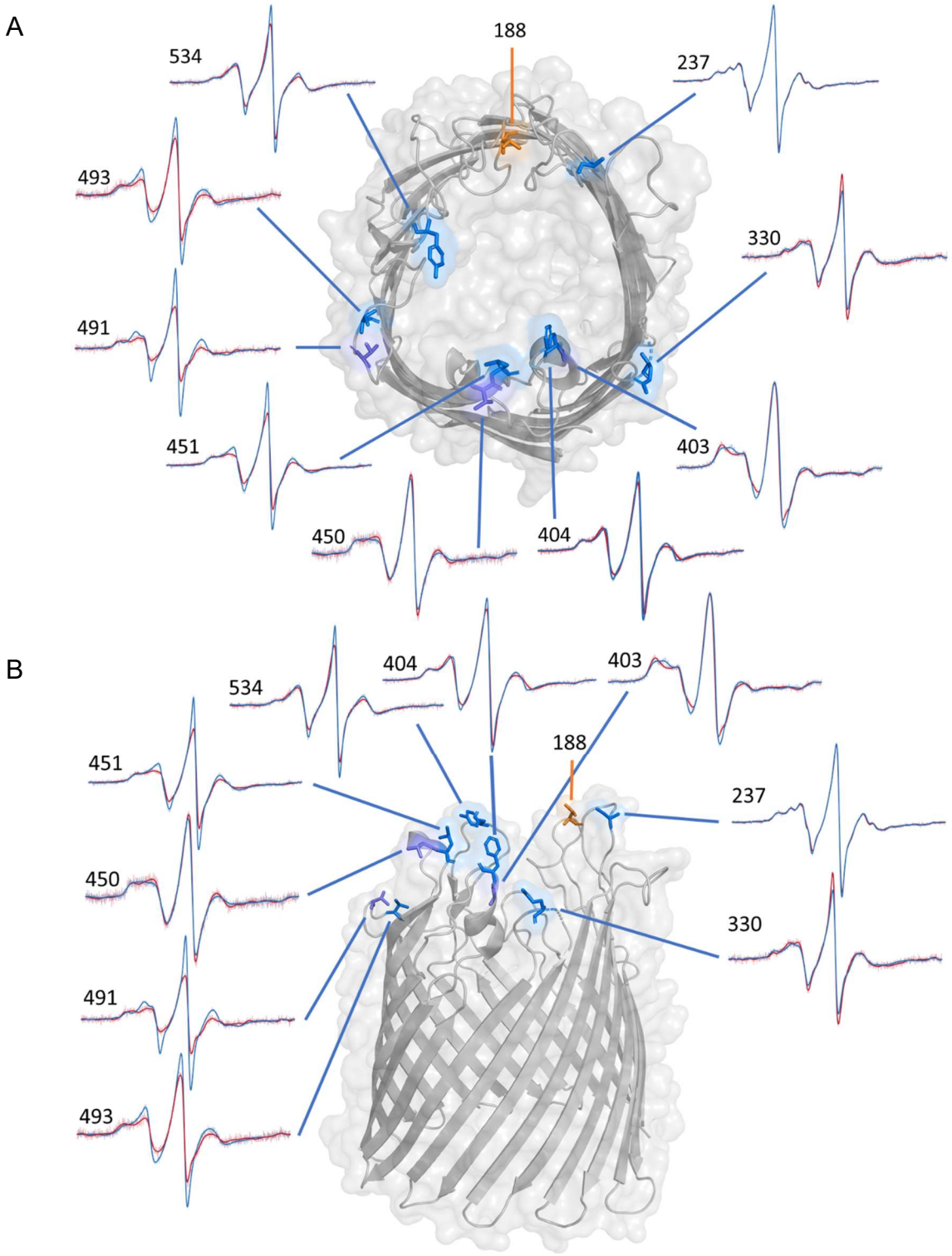


Figure 5. The barrel CW spectra as a function of their lateral position in the protein, viewed from the (A) top and (B) side. Moving clockwise around the protein, site 237 sits on the adjacent loop to site 188. Both sites show no change with addition of substrate, and so site 237 is expected to provide another static control site for hatch-barrel DEER. Measurements to the opposite side of the protein from 188 will be provided by sites 330, 403, 404, 450, and 451, with 404 showing the least substrate dependent change. Finally, sites 491, 493, and 534 are perpendicular to the 188 position and may prove ideal for determining the potential mode of aggregation in the crowded environments. All spectra are normalized averages of 10 scans and the noisy data is overlayed with the wavelet transform.

would be that this loop moves inward towards the rest of the barrel, establishing new label-protein contacts. Finally, site 534 is the apical site on another loop in steric contact with the 188 loop. This position has no nearby bulky sidechains, and so the lack of change with respect to cobalamin addition was expected. Residues 491, 493, and 534 may also be useful in the organization and crowding experiments, as they sit a quarter turn away from the 188 site. These sites would produce the same peak distance as 188 only in the head-to-tail aggregation case, or where there is no preference for contact interface between proteins and all rotations are sampled. Residues 491 and 493 have the additional benefit of being positioned for easy accessibility to the LPS, and it may be possible to use them to further probe the environment around the protein.

6.5 Future Work

The CW experiments shown here establish a set of new hatch and barrel mutants from which a variety of future DEER experiments can be derived. In chapter 4, it was shown that the 90-188 pair responds to substrate in a dose-dependent manner, with quantitative shifts in the population of two distance components that may correspond to an unfolding event in a helical hatch element. To fully explain this potential conformational change, additional triangulation of sites 90 and 188 would be useful. Thus, additional barrel sites will be paired with site 90 and additional hatch sites with residue 188 to fully ascertain whether the motion is confined to the 90 loop and confirm that this effect can be observed at nearby residues, such as sites 91 and 93. Next, further pairs can be constructed to look for additional sites of conformational change in response to substrate binding. The barrel sites can also be combined into loop-loop pairs to observe gating motions in response to substrate binding in the extracellular face, with the expectation of shrinking interlabel distances with substrate.

For the periplasmic face, site 6-510 was shown in Chapter 4 to produce small but significant DEER traces in the DsbA$^-$ whole cells. This pair contains site 6 in the ton-box and can be used to follow ton-box extension into the periplasmic environment. In addition to site 510, residue 6 was previously paired with sites 157 and 384 in a reconstituted DEER study by the

Cafiso lab (2). Reobtaining the 6-157 and 6-384 results in the whole cells will allow for triangulation of the ton-box position and provide a comparison with the *in vitro* system that should highlight the differences in the native environment. The resulting DEER distributions can then be used as restraints in XPLOR-NIH for molecular modeling to produce a structure or structures of BtuB with an extended ton-box, and this should provide information on the number of N-terminal residues that must reorient in response to extension. This, work should mirror that performed in FhuA, and will provide information on the validity of hatch formation via N-terminal unfolding (3, 4).

For the crowding work presented in Chapter 5, there are several ways to further resolve the nature of the OM BtuB organization. The first would be through increases in the sample runtime and apparent modulation depth of the traces, as the distances were too long and broad to properly resolve even in the 7 us deuterated trace for BtuB 188. At the time of writing, the Q-band instrument used in these experiments was being upgraded with a higher power TWT amplifier, larger volume resonator, and lower temperature cryostat, which should combine to boost both the runtime and modulation depth of subsequently run samples. This experimental setup should also allow the use of refined DEER experiments, such as 5-pulse and echo-train DEER, which leverage additional pulse-trains in the observe frequency and shaped pulses to greatly increase runtime (5, 6). Alternatively, a 7-pulse refocused out of phase (ROOPh) DEER experiment has been devised, which minimizes all non-intermolecular contributions to the background signal and should permit a more accurate determination of the background model at the same runtime as in 4-pulse DEER (7).

From a sample standpoint, the use of other sites around the barrel that were identified in this chapter should also be able to determine whether there is head-to-tail protein aggregation, or the BtuB proteins are simply forced together with no angular dependence. If the aggregation is not head-to-tail, then the sites that are roughly a quarter-turn around the barrel from BtuB, namely 491, 493, and 534, may give substantially shorter distances in single-label DEER experiments. If

the same distance is observed, then the aggregation is either nonspecific, head-to-tail, or the distance is still unresolved and will have to be proved through the experimental improvements described above. Minimization of the crowding effect can also be achieved through the addition of plasmids encoding WT BtuB to dilute the labeled proteins on the surface, or through the introduction of the WT gene into the same pAG1 plasmid for direct coexpression. The DsbA-mutant also has been seen to suppress the size of the crowding peak and harvesting even earlier in the cell cycle may be sufficient to suppress the island formation. This will have to be in conjunction with the upgraded experimental setup, as many of the DsbA$^-$ samples are at the lower end of the DEER concentration range.

Moving to the longer term, the method explored in this work should also be extensible from BtuB to other TBDTs and OM proteins. Ideal initial extensions would be previously studied TBDTs, such as FhuA, FecA, and ShuA, for which prior EPR data from the Cafiso lab is available (8, 9). FhuA could be used for comparisons in ton-box extension between the OM and whole cell environments, while ShuA is structurally highly similar to BtuB and may provide insights into the conservation of motions observed in the BtuB hatch domain (8, 10). From there, labels on OmpF could be used for comparison with the trimeric porin organization observed using AFM, and may provide a reference for comparison with BtuB organization as well (11, 12).

Eventually, work in the native system may progress to the point that the transport pathways can be identified, and work on BtuB can transition towards the development of targeted molecules or peptides that modulate its function for the maintenance of gut ecology or the development of novel antibiotics. This will require further novel extensions of experiment and methodology from those presented here, but the ability to create targeted control of not just *E. coli* but potentially the entire gut microbiota make this system a tantalizing target for future Research.

6.6 References

1. Sikora, A., B. Joseph, M. Matson, J. R. Staley, and D. S. Cafiso. 2016. Allosteric Signaling Is Bidirectional in an Outer-Membrane Transport Protein. Biophysical journal 111(9):1908-1918.
2. Xu, Q., J. F. Ellena, M. Kim, and D. S. Cafiso. 2006. Substrate-dependent unfolding of the energy coupling motif of a membrane transport protein determined by double electron-electron resonance. Biochemistry 45(36):10847-10854.
3. Sarver, J. L., M. Zhang, L. S. Liu, D. Nyenhuis, and D. S. Cafiso. 2018. A Dynamic Protein-Protein Coupling between the TonB-Dependent Transporter FhuA and TonB. Biochemistry 57(6):1045-1053. Article.
4. Hickman, S. J., R. E. M. Cooper, L. Bellucci, E. Paci, and D. J. Brockwell. 2017. Gating of TonB-dependent transporters by substrate-specific forced remodelling. Nature Communications 8:14804. Article.
5. Borbat, P., E. Georgieva, and J. H Freed. 2013. Improved Sensitivity for Long-Distance Measurements in Biomolecules: Five-Pulse Double Electron–Electron Resonance.
6. Mentink-Vigier, F., A. Collauto, A. Feintuch, I. Kaminker, V. Tarle, and D. Goldfarb. 2013. Increasing sensitivity of pulse EPR experiments using echo train detection schemes. Journal of Magnetic Resonance 236:117-125.
7. Milikisiyants, S., M. A. Voinov, and A. I. Smirnov. 2018. Refocused Out-Of-Phase (ROOPh) DEER: A pulse scheme for suppressing an unmodulated background in double electron-electron resonance experiments. Journal of Magnetic Resonance 293:9-18.
8. Sarver, J. L., M. Zhang, L. Liu, D. Nyenhuis, and D. S. Cafiso. 2018. A Dynamic Protein–Protein Coupling between the TonB-Dependent Transporter FhuA and TonB. Biochemistry 57(6):1045-1053.
9. Kim, M., G. E. Fanucci, and D. S. Cafiso. 2007. Substrate-dependent transmembrane signaling in TonB-dependent transporters is not conserved. Proceedings of the National Academy of Sciences of the United States of America 104(29):11975-11980.
10. Cobessi, D., A. Meksem, and K. Brillet. 2010. Structure of the heme/hemoglobin outer membrane receptor ShuA from Shigella dysenteriae: Heme binding by an induced fit mechanism. Proteins: Structure, Function, and Bioinformatics 78(2):286-294.
11. Jarosławski, S., K. Duquesne, J. N. Sturgis, and S. Scheuring. 2009. High-resolution architecture of the outer membrane of the Gram-negative bacteria Roseobacter denitrificans. Molecular Microbiology 74(5):1211-1222.
12. Yamashita, H., A. Taoka, T. Uchihashi, T. Asano, T. Ando, and Y. Fukumori. 2012. Single-Molecule Imaging on Living Bacterial Cell Surface by High-Speed AFM. Journal of Molecular Biology 422(2):300-309.