

Simulated Docking Reveals Putative Channels for the Transport of Long Chain Fatty Acids in *Vibrio Cholera*

Andrew Turgeson*

Department of Chemical and Computational Engineering, University of Tennessee, Chattanooga, USA

*Corresponding Author: Andrew Turgeson, Department of Chemical and Computational Engineering, University of Tennessee, Chattanooga, USA;

E-mail: andrew-turgeson@utc.edu

Received date: December 02, 2021; Accepted date: December 16, 2021; Published date: December 23, 2021

Citation: Turgeson A (2021) Simulated Docking Reveals Putative Channels for the Transport of Long Chain

Fatty Acids in *Vibrio Cholera* J Mol Microbiol Vol.5 No.3.

Abstract

Fatty acids (FA) play an important role in biological functions, such as membrane homeostasis, metabolism, and as signaling molecules. FadL is the only known protein that uptakes long chain fatty acids in bacteria, and until recently it was believed that only FA of 18 carbons could be uptaken by bacteria. To investigate the uptake of FA by *Vibrio cholerae*, the causative agent of cholera, sequences of the homologs of FadL were found and folded into tertiary structures. These structures were placed in membrane systems and simulated using molecular dynamic software. The trajectories of each simulation were then docked to find FA binding sites and mapped to find the FA transport pathways of the selected homologs.

carbons (many plants are still capable of monounsaturated and unsaturated FAs for waxes and seed storage lipids)³. However, plants are generally the only producers of n-3 (ω -3) and n-6 (ω -6) where the first unsaturated carbon starts on the 3rd or 6th carbon from the tail methyl group.

Oddly, there are some heterotrophic bacteria (*Vibrio* and *Pseudomonas*) that can also produce the typically plant based n-3 PUFAs⁴. Mammalian cells possess cytoplasmic fatty acid synthase (FAS) a major producer of 16-18 carbon atoms (which are also the most common cellular FAs in mammals) [3]. Typically, plants and animals do not create the higher order (> 20 carbons) unsaturated fatty acids; instead, these longer chain FAs are commonly produced by marine protists and microalgae³⁻⁵. It's widely known that fish, mollusks, and crustaceans tend to have high concentrations of the longer chain FAs such as eicosapentaenoic acid (EPA, 20:5) and docosahexaenoic acid (DHA, 22:6). It is thought that all PUFA in food webs originate from primary producers, where organisms further up the food chain have only the ability to modify the FA by bioconversion and elongation as they pass through the food web (i.e., trophic upgrading).

Thus fish, mollusks, and crustaceans which have a diet of microalga and protist have higher concentrations of the longer chain PUFAs, but have a lessened ability for FA conversion to long PUFAs than freshwater fish.

In bacteria, FAs are primarily used as components for the phospholipid bilayer of the membrane. These membrane phospholipids are constantly being synthesized, modified, recycled, and degraded to maintain membrane homeostasis and to respond to environmental stressors [6-7].

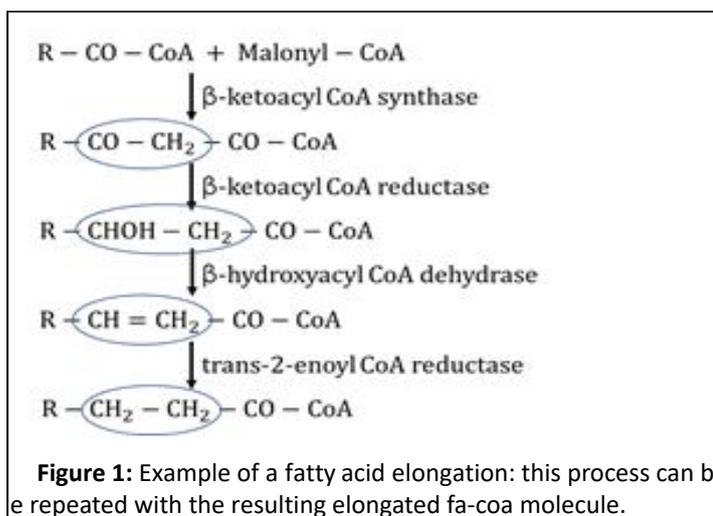
Free FAs are released during these processes, constituting important sources of metabolic energy. Fatty acid biosynthesis involves stepwise carbon elongation and unsaturation as needed that contribute to membrane homeostasis. Further maintenance of membrane dynamics can be mediated by enzymes acting on constructed phospholipids, such as desaturases, cis/trans isomerases, and cyclopropane synthases [8].

Introduction

In Gram negative bacteria a single protein is responsible for the recognition and transport of long chain fatty acids (LCFA) across the outer membrane leaflet; this protein is known as FadL. FadL's ability acquire LCFAs gives bacterial versatility in carbon sources utilization, which provides selective advantages for survival. In the case of *Vibrio cholerae*, the causative agent of cholera, this bacterial robustness may have ecological and medical implications. In this paper we will discuss the importance of bacterial FA synthesis and uptake. Then, we will compare novel structural models of *V. cholerae* FadL homologs with that of *E. coli*.

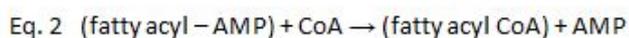
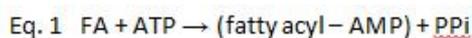
Fatty acids (FA) are molecules with a carboxylic acid head group and an aliphatic tail group of varying length and saturation. FAs are used primarily as building blocks for cell membranes, but also supply energy, and can be used as signaling molecules^[1]. Fatty acids can be acquired from exogenous sources as well as being synthesized de novo. However, many organisms (such as *Homo sapiens*) require specific exogenous sources of FAs for specific metabolic functions¹. In humans, this can be immune system regulation, blood clotting, neurotransmitter biosynthesis, cholesterol metabolism, and phospholipids for the brain and the retina [2].

In nature, plants typically have a limited synthesizing capacity that produces polyunsaturated fatty acids (PUFA) up to only 18



Fatty acid synthesis pathways are highly conserved between bacteria and eukaryotes, the differences being the resulting fatty acids synthesized by bacteria tend to be slightly shorter, generally lack poly-unsaturation, and the monoenoic C18 acids have different double bond positions [9]. Type II fatty acid synthesis (FASII) is the process used by bacteria to generate the fatty acid components of phospholipids [10]. The first step of the reaction is performed by the acetyl-CoA carboxylase complex (ACC) where the biotin-dependent enzyme that catalyzes an irreversible carboxylation of acetyl-CoA to produce malonyl-CoA. The resulting malonyl-CoA is used to for the elongation cycle which extends the growing fatty acid with consecutive reduction, dehydration, reduction and condensation reactions by various fatty acid biosynthesis (Fab) enzymes such as the ones seen in Figure 1.

The pathway for long chain FA uptake in Gram-negative bacteria begins with the trans membrane protein FadL[11-12] to transport the FA into to periplasmic space, where it is then delivered through the inner membrane to FadD (acyl-CoA synthase or fatty acid-CoA ligase). FadD uses adenosine triphosphate (ATP) along with a FA, producing adenosine monophosphate (AMP) and $P_2O_7^{4-}$ (PPi) and a FA bonded to a coenzyme A (CoA) (shown in Equations 1 and 2 below)[13]. The FA-CoA can then be shortened in β oxidation producing a shorter FA and generating some energy and/or it can be modified in a FAS cycle to meet the specific needs of the cell.



It was previously believed that enteric bacteria, such as *E. coli* and *V. cholerae*, were only able to acquire up to 18 carbon length FAs[14]. However, over the past decade several Gram-negative pathogens have been shown to assimilate and respond to exogenous PUFAs [15-18].

In the case of *V. cholerae*, this increased uptake is likely due to its natural ecosystem of tropical climates where marine algae and protist are a common food source. The uptake of PUFAs allows the incorporation of this long chain FA into the cell envelope, and this incorporation has been shown to affect the

membrane permeability, motility, biofilm formation, and antimicrobial resistance of the bacterium [19].

With increasing attention towards FAs and their effects in biology, the study of a species that exhibits broader capacity for the uptake and use of FAs presents an opportunity for comparison and elucidation of the uptake dynamics of the transmembrane protein FadL. In this paper we will study the structure and functions of several *V. cholerae* FadL homologs using molecular dynamics and perform comparisons to the *E. coli* FadL homolog.

Methods

Generating the FadL Membrane Systems

The most prevalent putative FadL homologs in *Vibrio cholerae* were identified via a bioinformatic search using the NCBI's Basic Local Alignment Search Tool (BLAST) [20]. Briefly, the protein sequence of *E. coli* M1655 FadL (accession number: NP_416846) was used as input for homolog searches against all available sequenced *V. cholerae* strains of the pathogenic O1 and O139 serogroups. The search algorithm settings were set at 100 max target sequences, short queries, an expect threshold of 10, a BLOSUM64 scoring matrix, with gap costs determined by Existence: 11 Extension: 1 and a Conditional composition score matrix adjustment. No filters or masks were used to analyze the results.

Each identified protein identified was presented with metrics that measured the similarity between the identified protein and *E. coli*'s FadL. These metrics are:

- % identity - is the percent of two protein/nucleotide sequences matching the same residues at the same positions in alignment [20].
- Expect value (E value) - represents the number of different alignments that are better than or equal to the protein/nucleotide sequence that are expected to appear in the database by chance (in this case, the lower the score, the better aligned the sequence)[20].
- Max and Total Score - the score (max/total) of an alignment is calculated as the sum of the substitution and gap scores – determined by the algorithm used [20].
- Query Coverage % - is the percent of the query sequence that overlaps the subject sequence [20]

The BLAST search originally found many proteins; however, many were identical in sequence.

Removing the repeated proteins lead to 19 unique proteins. The resulting BLAST algorithm metrics were all similar for all identified proteins. Briefly, the % identity ranged from about 30-38%, a max/total score range of 225-293, and the E-values were between 6.62E-96 and 1.62E-69.

These proteins are all acceptable as homologs of FadL, see figure 1.

Protein Accession #	Percent Identity (%)	Expect Value	Max/Total Score	Query Coverage	Serogroups*			# Identical Proteins	# Unique Strains
					O1	O139	O395		
EMP90254.1	38.53	6.62E-06	298	0.98	X			153	77
EMO65983.1	38.53	8.84E-06	295	0.98	X			3	1
EMP95771.1	38.307	2.31E-05	292	0.98	X			19	9
EMO59550.1	38.307	2.34E-05	292	0.98	X			142	100
NP_233248.1	38.307	4.97E-05	291	0.98	X	X	X	1386	770
ACC62431.1	38.053	8.19E-05	291	0.99	X			4	4
KJX78191.1	37.671	3.6E-02	284	0.96	X			7	4
EMP87839.1	36.056	1.17E-04	265	0.95	X			26	14
EMO00450.1	36.136	1.56E-03	262	0.95	X		X	168	97
NP_230688.1	35.455	1.61E-03	262	0.95	X	X		1405	811
EMP95211.1	35.973	1.91E-03	261	0.95	X			7	3
EMO66441.1	34.292	2.26E-05	241	0.95	X			10	5
EMO33484.1	34.071	8.72E-05	239	0.95	X			3	1
NP_230687.1	30.538	1.6E-00	228	0.99	X	X		1557	891
EMO69440.1	30.538	2.7E-00	228	0.99	X			9	4
EMP87811.1	31.169	2.8E-00	228	0.99	X			119	58
EMO00449.1	30.323	4E-00	228	0.99	X		X	62	32
EMP95212.1	30.67	9.26E-00	226	0.99	X			3	1
EMO33483.1	30.752	1.62E-09	225	0.97	X			2	1

*Proteins found in the corresponding serogroup are marked with an X.
**Proteins found in the O395 serotype are marked with an X.

Figure 1: BLAST search result between the two serogroups and one serotype tested (o1, o139, and o395).

The BLAST reports were conglomerated and analyzed for the number of repeated identical proteins and the number of unique strains each protein was found in. Three of the nineteen sequences were chosen accession numbers: NP_233248, NP_230688, NP_230687 (henceforth described as NP233248, NP230688, and NP230687). These homologs were chosen based on virulent serotypes and prevalence (Table 1). The selected *V. cholerae* sequences were folded using the I-TASSER [21] standalone version, and compared to the known crystal structure of *E. coli*'s FadL from the RCSB database (PDB ID 1T16) [22]. The resulting structures can be seen in the supplementary material, Figure A1.

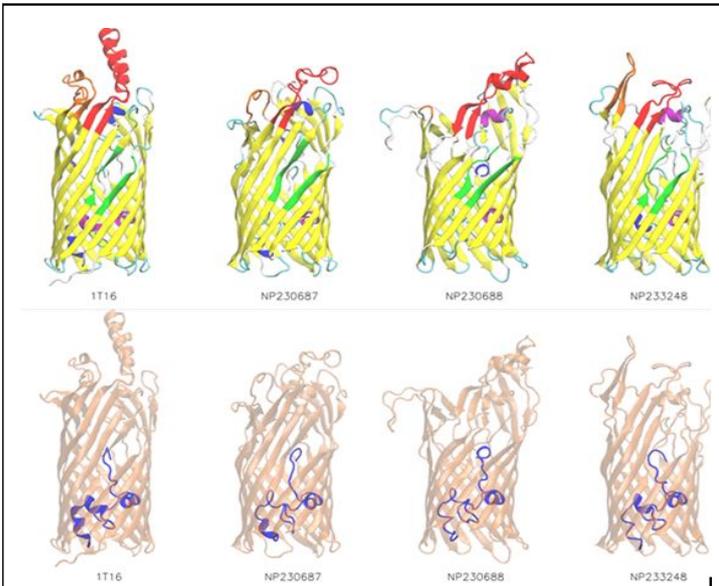


Figure A1: Folding results of fadL from i-tasser. top - the coloration of fadL is secondary structure, with the exception of the I3 loop (red), I4 loop (orange), and S3 kink (green) for reference. the β barrel strands are lined up well, and the S3 kinks are apparent in all the homologs generated. the I3 and I4 loops are somewhat formed, but the *V. cholerae* homologs' extra cellular loops do not mirror the original 1t16 crystal structure. bottom - the fadL homologs internal n-terminus structure (blue) all appear to follow a similar folding scheme. the secondary structure detection algorithms tend to detect α and 310 helices in some homologs but then detects turns and coils in others - even with very similar backbone shape.

The resulting FadL structures (including an equilibrated *E. coli* 1T16 structure annotated as *E. coli*-MD henceforth) were then each placed into a membrane using CHARMM-GUI Membrane Builder [23]. The *V. cholerae* homologs' membrane had an outer leaflet of *V. cholerae* type 1 Lipid A, Core A, and 15 O1 O-antigen units.

The *E. coli* outer leaflet was composed of *E. coli* type 1 Lipid A, Core R1, and 3 *E. coli* O1 O-antigen units (with 5 sugars per O unit). Both types had an inner leaflet of 67% phosphatidylethanolamine (PE) and 33% phosphatidylglycerol (PG). The structure of each of these molecules can be seen in the supplementary material, Figure A2.

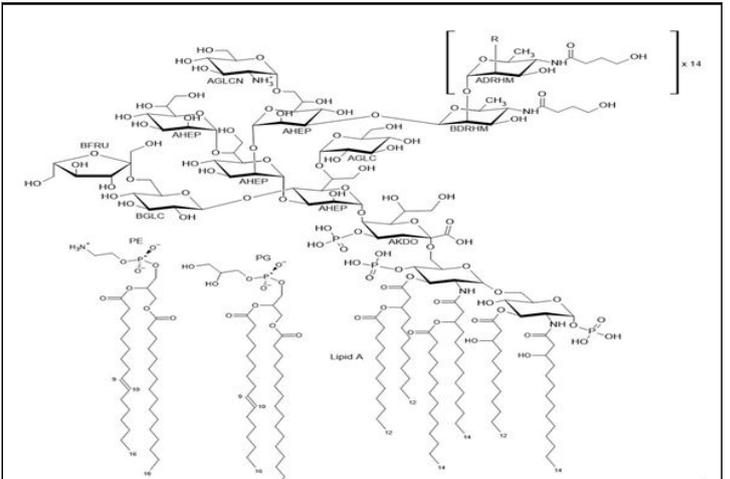
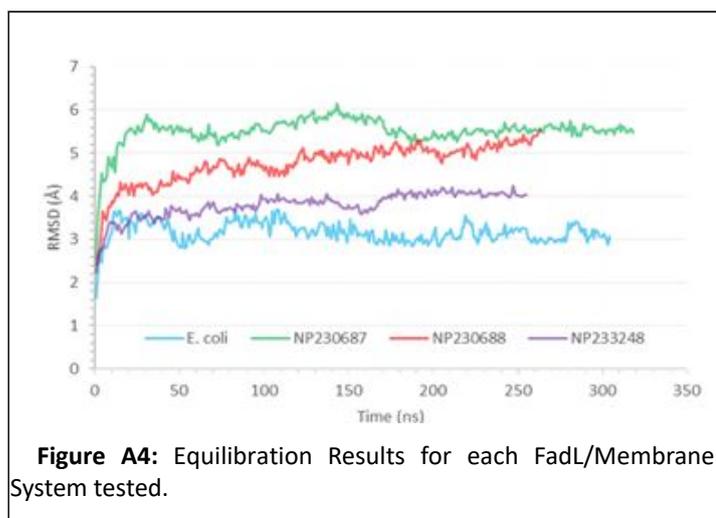


Figure A2: 2D representation of the skeletal formula for phosphatidylethanolamine (pe), phosphatidylglycerol (pg), and lipopolysaccharides (lps). these molecules make the membrane the fadL proteins were set in during equilibration. the core region carbohydrate abbreviations are as follows: α -keto-deoxyoctulosonate (akdo), α -D-mannoheptose (ahep), α -D-glucose (aglc), α -D-glucosamine (aglcN), β -D-glucose (bglc), and β -D-fructofuranose (bfuru). the o-antigen region abbreviations are α -D-rhamnose (adrhm) and β -D-rhamnose (bdrhm). both α and β rhamnoses are modified with a (s)-2,4-dihydroxybutanoyl off of the 4' carbon position. the final r- group for the o-antigen is a hydroxyl group (oh).

Equilibrating the Membrane Systems

The resulting simulation constraints generated by the CHARMM-GUI were then used in conjunction with NAMD [24] and CHARMM36 force fields [25]. During simulations, Langevin dynamics were used to maintain constant temperature (310 K) and pressure (1 atm). The simulations were sized as 80Å x 80Å x 140Å and a flexible cell boundary was chosen for an anisotropic membrane system. A cutoff of 12Å was used along with a particle mesh Ewald [26] for electrostatic interactions. All equilibrations used a timestep of 2 fs and nonbonded frequency and full electrostatics calculated at every step.

Each trial of the four trials was equilibrated for a minimum of 250 ns. The RMSD of the equilibration run for each protein tested can be seen in Figure A4. 50 DCD trajectory frames were taken from each of the equilibrated runs, taking one frame every 1 ns starting from 201 ns and ending at 250 ns.

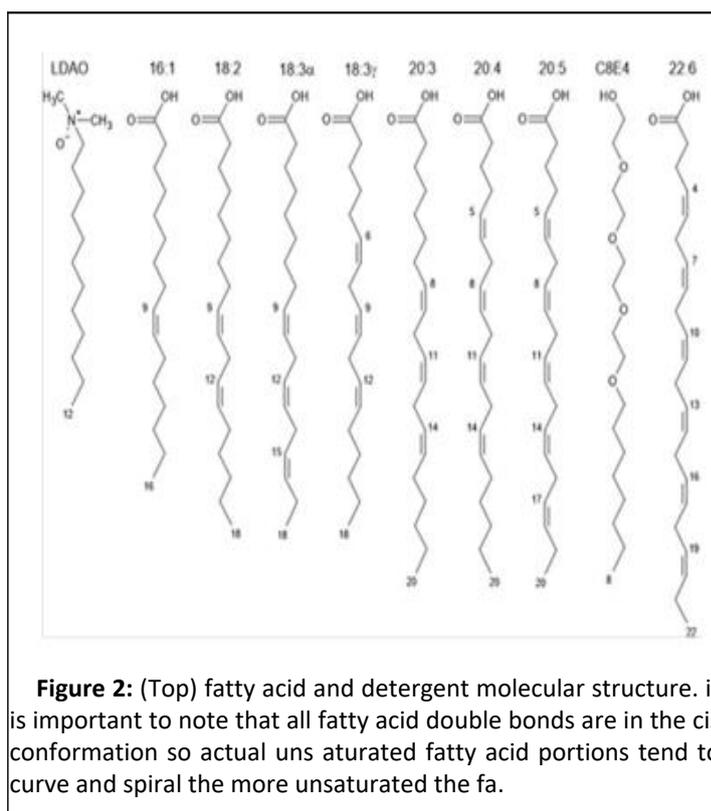


Docking of FadL

The 50 DCD trajectory frames for each of the tested FadL proteins were then aligned to the original 1T16 structure and the protein was isolated from the rest of the membrane system. The 200 isolated FadL frames as well as 50 instances of the original unmoving 1T16 FadL structure were then docked using AutoDock4 and AutoGrid4 using AutoDockTools (ADT) as a GUI [36]. A variety of LCFAs and the original two ligands from the 1T16 PDB resulting in 10 different ligands were tested Table 2 and Figure 2.

Table 2: Fatty acids tested for docking.

Fatty Acid	Abbreviation	Biological Significance
Lauryldimethylamine oxide	LDAO	Antimicrobial zwitterionic surfactant (C ₁₄ H ₃₁ NO)[27]
Palmitoleic acid	16:1	A common component of bacterial lipids[18,28]
Linoleic acid	18:2	Found in VC membrane phospholipids [14] and bile[29]
α-Linolenic acid	18:3α	Found in microalgae, cyanobacteria, and fish [30-32]
γ-Linolenic acid	18:3γ	An intermediate in the conversion of linoleic to arachidonic acid [33-34]
Dihomo-γ-linolenic acid	20:3	Found in microalgae and fish [31-32,35]
Arachidonic acid	20:4	Key cellular signaling molecule and inflammatory intermediate[34]
Eicosapentaenoic acid	20:5	Found in microalgae and fish[31-32,35]
Tetraethylene glycol mono-octyl ether	C8E4	A membrane solubilizing detergent (C ₁₆ H ₃₄ O ₅). As an ether it is 21 heavy atoms in length.
Docosahexaenoic acid	22:6	Found in microalgae and fish[31-32, 35]



The 80x80x120 AutoDockTools gridbox binding region (closer to a 40Å x 40Å x 60Å box) was restricted to the upper extracellular region of the FadL proteins encasing the majority of the FadL proteins. To maintain the cis structures of the FAs, the

unsaturated double bonds of the ligands were kept rigid during docking.

Each docking used a genetic algorithm with a population size of 150, a maximum number of evaluations of 2,500,000, and a maximum of 27,000 generations.

Results

Docking Viability

To test the viability of the docking procedure, the FA library was docked with E.

coli FadL (1T16) and specifically comparing the bindings of LDAO and C8E4 with the proximal bindings of the native LDAO and C8E4 molecules attached to the 1T16 PDB crystal structure.

Figure 3 examines one frame of the resulting dockings showing a preference of the AutoDock binding's sites to primarily be locations that were bound experimentally.

However, the selected docking of the S3 kink binding site (residues highlighted in green) contains C8E4 molecules where in the original crystal structure C8E4 molecules were restricted to the low affinity binding site in the L3 and L4 extracellular loops.

This may be due to during the methodology of van den Berg LDAO and C8E4 competed for binding during the protein washing phase[22], while in docking there was no binding competition.

Docking Nodal Analysis Overview

Figure 4 shows the original 1T16 crystal structure with the native detergents outlining the low affinity binding site, the high affinity binding site, and the S3 kink [22].

The AutoDock binding within the 1T16 crystal structure in conjunction with a cluster analysis using a mean shift algorithm found the location of the low affinity binding site (Node 1), the high affinity binding site (Node 2), and the S3 kink (Node 3) by using the clustering of the docked FAs.



Figure 3: Docking of 1t16 with ldao and c8e4 (transparent) compared with the original ldao and c8e4 (opaque). while this is only one frame, the ldao tends to be strongly correlated to the high affinity binding site while the c8e4 tends remain in the low affinity binding site or it bypass the transport channels and appears in the s3 kink (this occurrence is likely an effect of autodock's ligand placement algorithm).

The cluster analysis was performed for each of the FadL and V. cholerae and E. coli-MD homologs structures, resulting in the nodal locations seen in Figure 5. The nodes from the E. coli-MD (the equilibrated 1T16 structure), Figure 5A, shows almost identical nodal locations as the unequilibrated 1T16 structure (with a small exception of the high affinity binding site – Node 2 which tended to be closer to the center of the FadL β barrel). Exception aside, this demonstrates the NAMD equilibration E. coli FadL structure tended to retain important structures during the simulation.

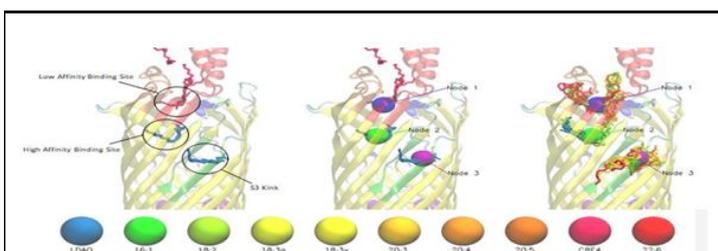
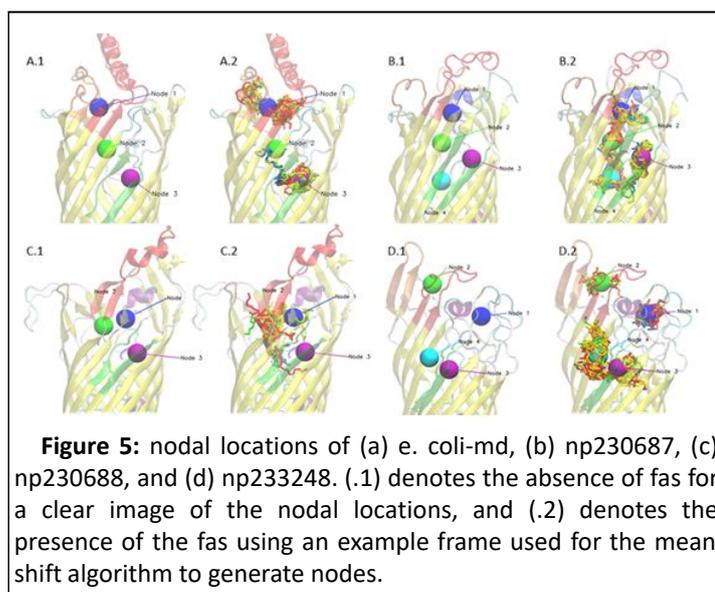


Figure 4: Mean shift based nodal analysis of the 1t16 docking. (left) the original detergents and locations of the low affinity binding site, the high affinity binding site, and the s3 kink. (Center) the location of the nodes found by the mean shift algorithm. (right) a frame of the fa clusters from docking that the mean shift algorithm used to generate the nodes. (bottom) color coding of each fa tested to show clustering by type.



The FAs were then categorized based on proximity, with each docked FA being prescribed one node per frame. With 10 FAs per type per frame and 50 frames, 5,000 FAs were assigned to each FadL homolog - giving a reasonable statistical model. The resulting docked locations were summarized in Figure 6.

For the 1T16 test case, LDAO had a strong affinity for the high affinity binding pocket (Node 2) with 86.0% of the LDAO molecules docked appearing in or around the high affinity binding site.

The other small molecules such as 16:1 and C18 FAs also showed some clustering in the Node 2 region (42.2%, 17.8%, 23.4%, and 23.2%). This is reasonable due to the original 1T16 structure having a LDAO molecule bound to the high affinity binding site [22] (the Node 2 locus), where other similarly shorter chained FAs could also fit into the open pocket.

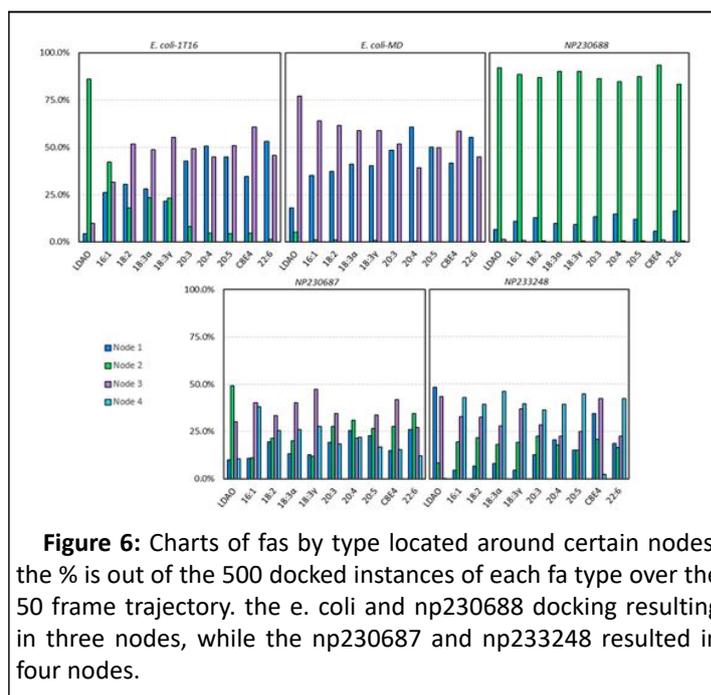
Interestingly, the S3 kink (Node 3) tended to have more docked FAs than the low affinity binding site (Node 1) which may be due to the tubular cavity of the S3 kink region, providing more surface area for FAs to bind to than the more open low affinity binding site.

Amongst all the FAs tested, the average docking binding energy for Nodes 1 and 3 of 1T16 were -8.975 and

8.976 kcal/mol respectively, indicating a very close average binding energy. Examining 18:2 in specific, the binding energy of 18:2 with Node 1 was better (-9.29 versus -8.97 kcal/mol), but AutoDock propagated more 18:2 FAs on Node 3.

This is likely due to the AutoDock algorithm finding it more difficult to dock the Node 1 area due to a smaller binding channel, even if the binding channel has a better binding energy.

Another example of this phenomena is 22:6 having the best overall binding energy when it was found in the high affinity binding site (-12.65 versus 9.73 and -10.11 kcal/mol), however, this only occurred with 1.4% of dockings because the high affinity binding site was originally bound to LDAO – a shorter chain FA.



The *E. coli-MD* docking revealed that the high affinity binding site had fewer dockings than the other sites. This indicates that due to the vacancy of FAs during simulation, that the high affinity binding site was closed and did not dock many FAs. Further investigation revealed that small positional changes in the high affinity binding pocket residues - particularly ALA153, ILE155, and LEU200 impeded the binding pocket channel, and greatly reducing the ability for FAs to fit in the binding pocket. Unlike the high affinity binding site, the S3 kink did appear to have substantial binding, indicating that there may not be a conformational shift during FA transport, but instead a shift in the gated channel between the high affinity site and S3 kink. The size and saturation of the FA did have an effect on the docking. Typically, the longer the FA carbon chains and more unsaturated, the affinity for Node 1 was increased and the affinity for Node 3 was decreased - mirroring the 1T16 dockings.

The NP230687 docking revealed a visible main channel. It is predicted that the FAs move from Node 1 to Node 2 to Node 4 and then to Node 3, the S3 kink Figure 5B.2. The clustering of FAs did not show much preference for any one of the four nodes with the exception of Node 3, where the totaled percent FAs located at Nodes 1, 2, 3, and 4 were 17.5%, 26.2%, 35.0%, and 21.3% respectively. As expected of a *V. cholerae* homolog there was no discernible difference in FA tail length or saturation.

In NP230688's docking, Nodes 1 and 2 were in close proximity to one another as seen in Figure 5C.1, the major difference between the two beings Node 2 is the locus of the high affinity binding site in the *E. coli* homolog. NP230688 showed a strong favoritism for Node 2 with a total of 88.3% of all FAs appearing in the Node 2 region. Figure 5C.2 illustrates the FA's tendency to funnel around Node 2. Only 0.6% of FAs were found in the S3 kink region Node 3, alluding to a conformational mechanism to allow passage of the FA.

The docking of NP230688 showed a high affinity for the outer portion of the S3 kink Node 4. This is unexpected based on the premise that FAs travel through the β barrel in *E. coli*. Node 4

does tend to have a more pronounced indentation making docking more ideal than some other locations; however, the docking did not factor in the LPS which encompassed the outer perimeter of the FadL β barrel, which would leave little room for FAs. Autodock's current atom limit prevents a system with LPS included. These results indicate that the NP233248 β barrel was not in an open conformation for the 50 frames used for docking and suggests that there is a very large conformational change that possibly starts with FA binding to Nodes 1 or 2 in the extracellular loop region.

E. coli FA Transport

To determine any important residues in the transport of FAs, the docked homologs were searched for any residues within 3 Å of each of the docked FAs. These residues were agglomerated, and each residue found was counted for recurrences. The resulting Table 3 shows the twenty residues that were found to interact with the docked FAs most often.

For the E. coli homolog dockings, the residues found most frequently were those of the low affinity binding site and the S3 kink.

This was expected, as the nodal analysis determined that the majority of FAs were docked in the Node 1 and 3 regions. While not in the same proportions, many of the same residues were found for both the E.

coli 1T16 and the E. coli-MD structures. Residues PRO253, ILE254, PRO255, and PHE315 have a reoccurring presence in the low affinity binding sites for both structures Figure 8A,C. Residues GLY2, LEU5, PRO54, VAL56, ALA74, GLY103, LEU104, ALA105, PRO362, and ARG366 are commonly found in the S3 kink region.

The majority of these residues are nonpolar except for the polar glycines GLY2 and GLY103 and the positively charged arginine, ARG366.

The arginine headgroup faces towards the S3 kink pocket indicating an affinity for carboxyl groups of FAs which is confirmed by the number of FA carboxyl headgroups in the proximity of ARG366 during docking.

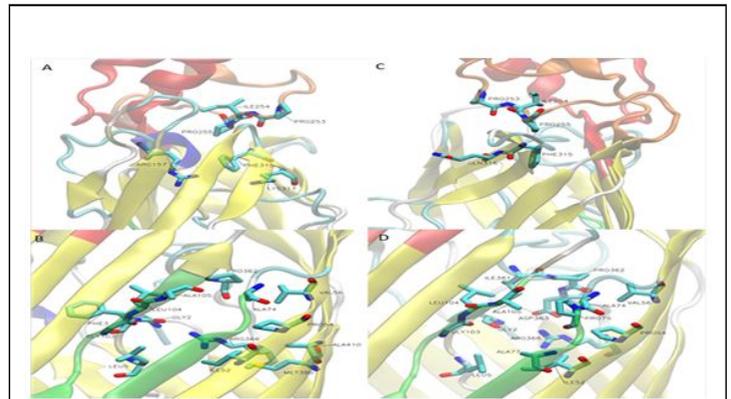


Figure 7: E. coli FadL binding residues of (A) 1T16 low affinity binding site residues, (B) 1T16 S3 kink, (C) E. coli-MD low affinity binding site, and (D) E. coli-MD S3 kink. The L3 loops is colored red, the L4 loop is colored orange, and the S3 kink is colored green for reference. Perspective angles differ for easier observation of residues.

This could indicate an orientation of FA with the tail group facing the outlet before egress of FA through the S3 kink. The RMSD for the heavy atoms of these residues tends to be between 1.1 and 1.7, alluding to a stable S3 kink structure even with the difference of a bonded LDAO in the S3 kink of 1T16.

V. cholerae NP230687 FA Transport

The V. cholerae homolog NP230687 residues were primarily centered around the predicted transport channel Figure 8. This channel tends to start from between the 5th and 6th extracellular loops Figure 5. The FA is expected to continue to the base of the N-terminal hatch (residues 1 through 5), and then past the N-terminal hatch through the S3 kink opening. The N-terminal hatch in the docking does not restrict transport as predicted in the E. coli homolog. This is somewhat unexpected as generally a FA transport protein would have some selection mechanism specific to FAs. The residues that line the channel are primarily hydrophobic, with a few exceptions GLN4, HSD79, THR244, THR331, GLU336, and ARG399 which are all hydrophilic (ARG339 also having a positive charge). These residues are placed periodically throughout the channel in such a way that it could be the FA headgroup's attraction to these residues that guide the movement of the FA through the channel in a specific orientation. The channel seems to end at the S3 kink as with the E. coli homolog. A similar experiment with the docking of the bottom half of the homologs showed that there was a discontinuity from the main channel to any docking channels found in the bottom of the protein reinforcing the hypothesis that the S3 kink is the FA egress point. Oddly, the channel shares a similar overlap of the E. coli homolog's high affinity binding site location and N terminal hatch domain. While the new pathway bypasses this predicted pathway in E. coli, it is interesting that the original pathway (through the high affinity binding site location, and then through a tunnel created by a conformational mechanism occurring with the N terminal hatch domain) may still exist somehow in the V. cholerae homolog. Whether or not this N terminal hatch pathway is vestigial or is functional has not been determined at this time.

Rank	E. coli 1T16		E. coli-MD		NP230687		NP230688		NP233248	
	Count	Residue	Count	Residue	Count	Residue	Count	Residue	Count	Residue
1	2958	LEU104	2852	LEU104	2654	PHE3	4253	PHE308	2627	ALA102
2	2156	ARG366**	2705	ALA105	2488	GLN4*	3751	PHE3	2522	GLN74*
3	2083	PHE3	2683	ARG366**	2075	LEU5	3586	TRP300	1761	PRO72
4	2044	PRO54	2651	ALA74	1981	MET108	3415	TYR296*	1732	SER119*
5	1977	ALA74	2624	PRO54	1797	ARG339**	3342	GLN4*	1716	LYS151**
6	1943	ALA105	2331	PRO362	1709	PHE77	3249	LEU238	1501	VAL5
7	1910	PRO362	2274	ASP368*	1646	PRO54	3182	TRP268	1498	VAL101
8	1849	LEU5	2178	GLY2*	1826	VAL52	3074	ARG163**	1456	VAL52
9	1789	VAL56	2154	LEU5	1568	VAL396	2858	MET161	1417	VAL76
10	1680	PHE315	2087	ILE361	1538	PRO335	2504	ILE274	1401	PRO54
11	1673	PRO255	2078	ILE52	1523	VAL56	2277	PRO241	1396	ILE121
12	1639	PRO253	2058	GLY103*	1503	THR331*	2187	VAL129	1237	ALA71
13	1626	GLY2*	2049	PRO255	1491	TRP272	1697	VAL339	1168	ASP327*
14	1623	ILE52	2006	GLN316*	1462	VAL243	1588	LYS296**	1137	ILE75
15	1594	GLY103*	1965	PHE315	1418	THR244*	1567	PHE276	1095	ARG330**
16	1588	ARG157**	1923	VAL56	1385	GLU336*	1413	ILE266	1019	GLY100*
17	1580	ILE254	1870	ILE254	1384	ALA1	1319	LEU240	1013	GLY2*
18	1526	LYS317**	1756	PRO75	1258	HS079**	1293	THR127*	1004	THR103*
19	1440	MET390	1673	PRO253	1231	ILE334	1272	SER338*	977	ASP117*
20	1426	ALA410	1643	ALA77	1187	MET241	1256	ALA330	948	ALA379

* Polar, but neutrally charged residue.

** Positively charged residues (pH 7).

over the 50 frames, giving a maximum residue count of 5,000.

Figure 7: Residue count for residues found within 3 Å of each fa for each frame for each fadL homolog, 5,000 fas were docked

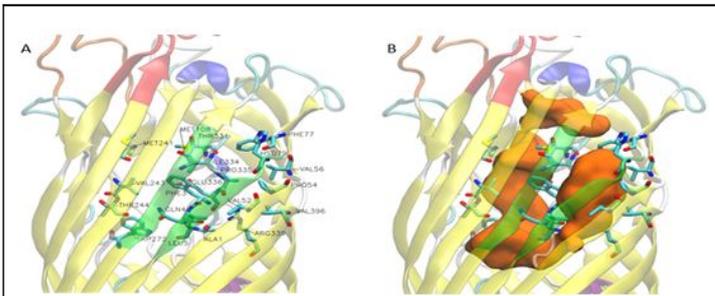


Figure 8: *V. cholerae* NP230687 FA interacting residues; (A) residue names and locations, (B) transport channel (orange) shown with its relation to the displayed residues. The L3 loops is colored red, the L4 loop is colored orange, and the S3 kink is colored green for reference.

V. cholerae NP230688 FA Transport

V. cholerae homolog NP230688 has a very large flat channel that seems to funnel FAs into the N-terminal hatch domain. The entrance point can be seen in the appendix (Figure A6) where the undefined low affinity binding region is found between the base of the extracellular loops. The channel leads to the N-terminal hatch opposite the S3 kink. The channel itself may have vestigial links to the *E. coli* homolog, where the N terminal hatch domain rests in the same position as the *E. coli* 1T16 structure and the channel overlaps the general area of the high affinity binding.



Figure A6: Docked NP230688 surface view; the purple is the docked fa showing the entrance of the fa transport channel.

This could indicate the removal of the high affinity binding site in favor of a more direct pathway but leaving the mechanisms of the N terminal hatch domain which would play the same role for the NP230688 homolog as it does for the *E. coli* homolog. Alternatively, the NP230688 channel, in comparison to the *E. coli* homolog, is the NP230688 channel does appear to go further down than the transposed high affinity binding site location above the N terminal hatch domain which ends at the N terminal hatch opposite the S3 kink. This could be an alternate pathway than the predicted high affinity site transposed gateway, but this proposed pathway has yet to be substantiated. The docked FAs did not appear in the S3 kink pore, likely due to LYS130 from the fourth β strand, S4, positioned parallel to the S3 kink that appears to be attracted to GLU50, SER106, and ASN107 as well as the backbone Oxygens of the S3 kink residue GLY109. This attraction causes LYS130 to fill the S3 transport pore and prevent docking (and possibly FA transport). This could be a selection mechanism that may determine the resulting FA position or FA type. The channel of NP230688 was found to be composed generally of hydrophobic residues. The exceptions to this are GLN4, THR127, TYR298, and SER338 which are hydrophilic, and ARG163 and LYS296 which are positively charged. Previously, it was postulated that the hydrophilic residues in NP230687 guided the FAs head-first through the channel, but the location of these hydrophilic residues in the NP230688 homolog are located on the top of the conical channel. Additionally, there are no remarkable hydrophilic residues present in the transport channel to guide the FA headgroup through the channel. It could be predicted that the hydrophilic head groups bind to the hydrophilic residues at the top of the channel for alignment further into the channel. Directional positioning of the FA is yet to be determined, but the direction may play an important role with the positively charged LYS130 residue blocking the S3 kink pore.

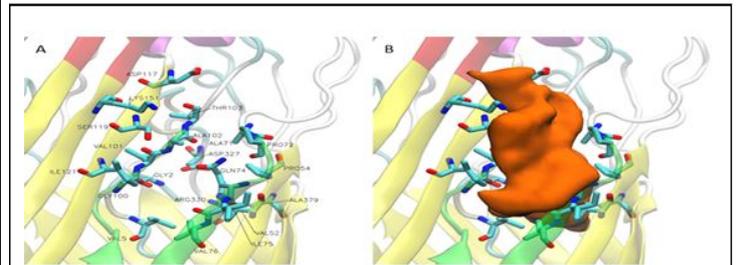


Figure 10: *V. cholerae* NP233248 FA interacting residues; (a) residue names and locations, (b) transport channel (orange) shown weaving through the displayed residues. the L3 loops is colored red, the L4 loop is colored orange, and the S3 kink is colored green for reference.

V. cholerae NP233248 FA Transport

The docking of *V. cholerae* homolog NP233248 revealed that the majority of docking sites did not occur within the β barrel structure of the FadL protein, but rather along the outer barrel primarily around the S3 kink Figure 10. This appears to be due to the protein did not have substantial open space for FAs to be docked on the inner portion of the β barrel. Oddly, there seems to be a pathway from between the L3 and L4 loops that goes down the side of the protein and to the outside of the S3 kink as

shown in Figure 11. Normally, this would essentially mean that any molecule would be able to make its way through the channel unless there was some interplay with the interface of the LPS bilayer to create some sort of selectivity mechanism - which is possible, but the channel between the outer portion of the S3 kink and the predicted initial binding sites between the L3 and L4 extracellular loops tends to close off depending on the L3 and L4 conformations. These L3 and L4 conformations may be the selectivity mechanism that this homolog uses to ensure the uptake of FAs instead of bactericidal compounds. Many of the docked FA were found within the S3 kink, where the internal cavity of the S3 kink would be vestigial if the FAs are transported to the predicted egress point without entry of the FA into the FadL β barrel structure. This vestigial S3 kink cavity agrees with the secondary bottom half docking analysis, where no FA pathways were found from the S3 kink to the bottom of the FadL protein.

S3 Kink FA Egress Point

To verify that the S3 kink is the egress point first suggested by Hearn et al.[37] the membrane layer location after equilibration was checked for the possibility of membrane diffusion. The resulting LPS bilayer head groups or the polar heavy atoms of the LPS were shown in relation to the D3 kink pore Figure 12. This pore was typically found at the upper portion of the LPS polar region indicating a strong affinity for the polar head groups of the FA with the polar LPS atoms, indicating a good possibility for assimilation into the LPS bilayer and diffusion into the periplasmic space.

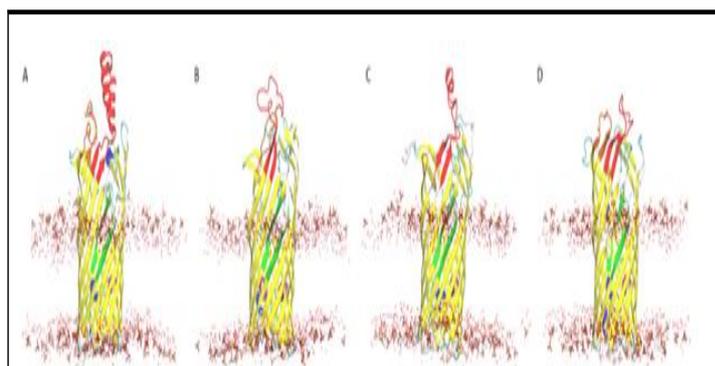
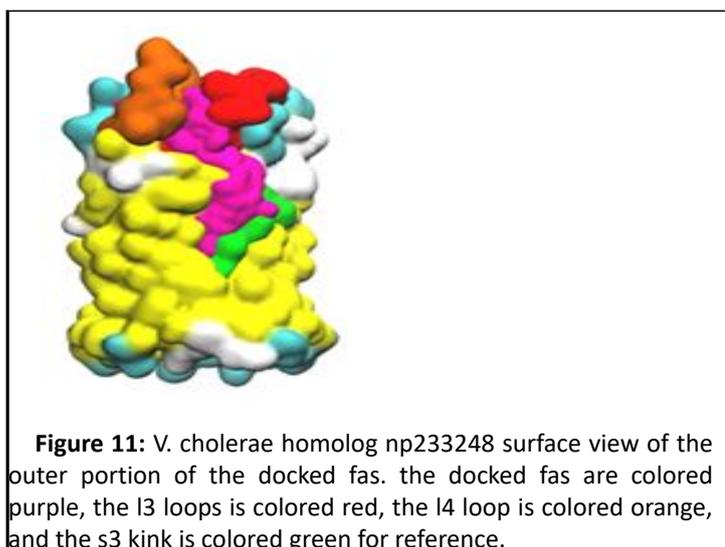


Figure 12: Lps bilayer polar atom locations with respect to the e. coli (a) np230687 (b) np230688 (c) np233248 (d) fadl proteins after equilibration.

Docking Energies

10 docking conformations were produced per ligand creating a total of 100 docking conformations per FadL homolog frame.

With 250 frames being docked, 25,000 conformations overall were generated. The best conforming (lowest energy) are shown in Table 4.

Similarly, the averaged docking energies by FA are given in Table 5.

Table 4: Best conformation energies from docking (energy units in kcal/mol).

Homolog	LDAO	16:1	18:2	18:3 α	18:3 γ	20:3	20:4	20:5	C8E4	22:6
E.coli 1T16	-10.11	-12.41	-13.53	-13.72	-13.33	-14.33	-14.58	-14.5	-10.96	-14.25
E.coli-MD	-9.15	-9.91	-10.36	-10.13	-10.45	-10.61	-10.71	-9.88	-9.92	-11.72
NP230687	-8.81	-10.37	-10.12	-9.9	-10.28	-11	-11.44	-10.55	-9.55	-11.24

NP23068 8	-9.92	-10.48	-11.48	-11.12	-11.17	-12.19	-12.37	-11.93	-10.22	-12.4
NP23324 8	-9.28	-10.95	-10.62	-9.93	-10.36	-10.84	-10.32	-9.96	-9.57	-10.95

Table 5: Overall average conformation energy from docking (energy units in kcal/mol).

Homolog	LDAO	16:1	18:2	18:3 α	18:3 γ	20:3	20:4	20:5	C8E4	22:6
E.coli 1T16	-9.17	-9.62	-9.55	-9.32	-9.74	-9.83	-9.46	-9.44	-8.15	-9.95
E.coli-MD	-7.39	-7.82	-8.09	-7.87	-8.18	-8.42	-8.34	-8.2	-7.47	-8.69
NP23068 7	-7.86	-8.33	-8.56	-8.36	-8.62	-9.01	-8.77	-8.7	-7.77	-9.24
NP23068 8	-8.53	-9.07	-9.42	-9.1	-9.38	-9.82	-9.61	-9.42	-8.17	-10.11
NP23324 8	-7.57	-7.81	-7.97	-7.6	-7.94	-8.22	-7.94	-7.83	-7.14	-8.21

Simulated docking results indicate that the original crystal structure E. coli 1T16 tended to have the most energetically favorable docking with respect to overall average as well as the best individual FA docking conformations. This is likely because the 1T16 structure was generated with the FadL protein bound with LDAO and C8E4 in the structure when the PDB was generated, giving it the specific conformation needed for strong binding. It is also apparent that the docking energies are more favorable for the longer chain FAs (with exception to C8E4 which has a total length of 21 heavy atoms). This result is likely due to the fact that longer FA chains provide more surface area for binding. However, many of the longer chained polyunsaturated FAs tend to have a hairpin tail due to the cis unsaturated portions. The uptake of these longer chained FAs would likely require some internal mechanisms for FA uptake that compensates for these rigid sections of the FAs, although these compensation mechanisms have not been found computationally.

Conclusions

Three V. cholerae FadL homologs were folded, docked, and analyzed. The E. coli FadL control dockings showed van den Berg's predicted docking²² in both the 1T16 crystal and NAMD equilibrated structures. It is also apparent that all the V. cholerae and E. coli FadL homologs did appear to typically bind at the base of the L3 extracellular loop, as van den Berg predicted in the 1T16 crystal structure²². The exception to this is V. cholerae homolog NP233248, where two binding sites appear around the S3 kink itself, as well as the S8 and S9 β strands.

The N terminus sequence of residues ALA, GLY, PHE, and GLN were conserved throughout the homologs, and were also

located in a similar area near the S3 kink regardless of N terminus features. These residues may play a part in the conformational changes that allow the selective passage of FAs through into the cell²². However, the investigation of the V. cholerae homologs may suggest this conserved sequence may be conserved for a signal peptide lysis function³⁸ where the cleavage site retain similar sequences AGFQ.

It was found from the equilibration of 1T16 that in E. coli, there are conformational changes that adapt to the uptake of the FAs creating open channels and binding sites for FAs that would otherwise be closed. This has not been determined for any of the V. cholerae homologs, and future simulations where the FadL system includes a range of FAs may reveal more of the channel properties and mechanics that cause selection and transport of the larger range of FAs and possibly other compounds.

These novel dockings of the wide range of FAs onto the V. cholerae FadL homologs shows where attention should be focused towards binding sites. Some of the more interesting results appear with the appearance of pores opposite the S3 kink as well as secondary binding sites. More investigation will be needed to further define the mechanics and get a clearer picture of the circumstances surrounding the formation of these docking sites.

References

1. De Carvalho C, Caramujo MJ (2018) The Various Roles of Fatty Acid s. *Molecules* 23:2583
2. Abedi E, Sahari MA (2014) Longchain polyunsaturated fatty acid sources and evaluation of their nutritional and functional properties. *Food Sci Nutr* 2:443-463.

3. Leonard AE, Pereira SL, Sprecher H, Huang YS (2004) Elongation of long-chain fatty acids. *Prog Lipid Res* 43:36-54.
4. Berge JP, Barnathan G (2005) Fatty acids from lipids of marine organisms: molecular biodiversity, roles as biomarkers, biologically active compounds, and economical aspects. *Adv Biochem Eng Biotechnol* 96:49-125.
5. Harwood JL, Guschina IA (2009) The versatility of algae and their lipid metabolism. *Biochimie* 91: 79-684.
6. Ernst R, Ejsing CS, Antonny B (2016) Homeoviscous Adaptation and the Regulation of Membrane Lipids. *J Mol Biol* 428:4776-4791.
7. Zhang YM, Rock CO (2008) Membrane lipid homeostasis in bacteria. *Nat Rev Microbiol* 6:222-233.
8. Cronan JE, Thomas J (2009) Bacterial Fatty Acid Synthesis and its Relationships with Polyketide Synthetic Pathways. *Methods Enzymol* 459: 395-433.
10. Parsons JB, Rock CO (2011) Is bacterial fatty acid synthesis a valid target for antibacterial drug discovery? *Curr Opin Microbiol* 14:544-549.
11. Nunn WD, Simons RW (1978) Transport of long-chain fatty acids by *Escherichia coli*: mapping and characterization of mutants in the *fadL* gene. *Proc Natl Acad Sci U S A* 75:3377-3381.
12. Ginsburgh CL, Black PN, Nunn WD (1984) Transport of long chain fatty acids in *Escherichia coli*. Identification of a membrane protein associated with the *fadL* gene. *J Biol Chem* 259:8437-8443.
13. Weimar JD, DiRusso CC, Delio R, Black PN (2002) Functional role of fatty acyl-coenzyme A synthetase in the transmembrane movement and activation of exogenous long-chain fatty acids. Amino acid residues within the ATP/AMP signature motif of *Escherichia coli* *FadD* are required for enzyme activity and fatty acid transport. *J Biol Chem* 277:29369-29376.
14. Giles DK, Hankins JV, Guan Z, Trent MS (2011) Remodelling of the *Vibrio cholera* membrane by incorporation of exogenous fatty acids from host and aquatic environments. *Mol Microbiol* 79:716-728.
15. Hobby CR, Herndon JL, Morrow CA, Peters RE, Symes SJK, Giles DK (2019) Exogenous fatty acids alter phospholipid composition, membrane permeability, capacity for biofilm formation, and antimicrobial peptide susceptibility in *Klebsiella pneumoniae*. *Microbiol open* 8:e00635.
16. Eder AE, Munir SA, Hobby CR, Anderson DM, Herndon JL, Symes SJK, Giles DK et.al(2017) Exogenous polyunsaturated fatty acids (PUFAs) alter phospholipid composition membrane permeability, biofilm formation and motility in *Acinetobacter baumannii*. *Microbiol (Reading)* 163:1626-1636.
17. Baker LY, Hobby CR, Siv AW, Bible WC, Glennon MS, Anderson DM, Symes SJ, Giles DK (2018) *Pseudomonas aeruginosa* responds to exogenous polyunsaturated fatty acids (PUFAs) by modifying phospholipid composition, membrane permeability, and phenotypes associated with virulence. *BMC Microbiol* 18:117.
18. Moravec AR, Siv AW, Hobby CR, Lindsay EN, Norbush LV, Shults DJ, Symes SJK, Giles DK et.al(2017) Exogenous Polyunsaturated Fatty Acids Impact Membrane Remodeling and Affect Virulence Phenotypes among Pathogenic *Vibrio* Species. *Appl Environ Microbiol* 83:256
19. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y (2015) The I-TASSER Suite: protein structure and function prediction. *Nat Meth* 12:7-8.
20. Van den Berg B, Black PN, Clemons WM Jr, Rapoport TA (2004) Crystal structure of the long-chain fatty acid transporter *FadL*. *Science* 304:1506-1509.
21. Wu EL, Cheng X, Jo S, Rui H, Song KC, Davila-Contreras EM, Qi Y, Lee J, MonjeGalvan V, Venable RM, Klauda JB, Im W et.al(2014) CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J Comput Chem* 35:1997-2004.
22. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K et.al (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26:1781-1802.
23. Brooks BR, Brooks CL, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu J Z, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M et.al (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30:1545-1614.
24. Birnie CR, Malamud D, Schnaare RL (2000) Antimicrobial evaluation of N-alkyl betaines and N-alkyl-N, N-dimethylamine oxides with variations in chain length. *Antimicrob Agents Chemother* 44:2514-5217.
25. Hood MA, Guckert JB, White DC, Deck F (1986) Effect of nutrient deprivation on lipid, carbohydrate, DNA, RNA, and protein levels in *Vibrio cholerae*. *Appl Environ Microbiol* 52 :788-793.
26. Van Berge Henegouwen GP, Van der Werf SDJ, Ruben AT (1987) Fatty acid composition of phospholipids in bile in man: Promoting effect of deoxycholate on arachidonate. *Clinica Chimica Acta* 165:27-37.
27. Guedes AC, Amaro HM, Barbosa CR, Pereira RD, Malcata FX (2011) Fatty acid composition of several wild microalgae and cyanobacteria, with a focus on eicosapentaenoic, docosahexaenoic and α -linolenic acids for eventual dietary uses. *Food Res Int* 44:2721-2729.
28. Ozogul Y, Ozogul F, Alagoz S (2007) Fatty acid profiles and fat contents of commercially important seawater and freshwater fish species of Turkey: A comparative study. *Food Chem* 103:217-223.
29. Durmuş M (2019) Fish oil for human health: omega-3 fatty acid profiles of marine seafood species. *Food Sci Technol* 39:454-461.
30. Gunstone FD (1992) Gamma linolenic acid-occurrence and physical and chemical properties. *Progress in Lipid Research* 31:145-161.
31. Hanna VS, Hafez EAA (2018) Synopsis of arachidonic acid metabolism: A review. *J Adv Res* 11:23-32.
32. Pereira H, Barreira L, Figueiredo F, Custodio L, VizettoDuarte C, Polo C, Resek E, Engelen A, Varela J et.al (2012) Polyunsaturated Fatty acids of marine macroalgae: potential for nutritional and pharmaceutical applications. *Mar Drugs* 10:1920-1935.
33. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ et.al (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem* 30:2785-2791.
34. Hearn EM, Patel DR, Lepore BW, Indic M, VandenBerg B (2009) Transmembrane passage of hydrophobic compounds through a protein channel wall. *Nature* 458:367-370.
35. Black PN (1991) Primary sequence of the *Escherichia coli* *fadL* gene encoding an outer membrane protein required for long-chain fatty acid transport. *J Bacteriol* 173: 435-42.