Investigations of Ion Channel Structure-Function Relationships Using Molecular Modeling and Experimental Biochemistry

Thesis by

Donald Eugene Elmore, Jr.

In Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

California Institute of Technology

Pasadena, California

2004

(Defended April 22, 2004)
2004

Donald Eugene Elmore, Jr.

All Rights Reserved
Chapter 2: Confirmation of the *M. tuberculosis* MscL Crystal Structure and Comparisons of MscL Homologues
The M. tuberculosis MscL Crystal Structure: Unique Opportunities and Ambiguity

Mechanosensation is a central part of numerous biological processes ranging from circulation and hearing in higher animals to maintaining proper osmotic conditions in bacteria (Wood, 1999). One family of bacterial channels gated by membrane tension was first identified and characterized electrophysiologically as the mechanosensitive channels of large, small, and mini conductance (MscL, MscS, and MscM, respectively) (Berrier et al., 1996; Martinac et al., 1987; Sukharev et al., 1993). The first of these to be cloned, E. coli MscL (Ec-MscL) (Sukharev et al., 1994), has been the subject of extensive studies including site-directed and random mutagenesis (Blount et al., 1997; Blount et al., 1996b; Ou et al., 1998; Yoshimura et al., 1999), covalent cross-linking (Blount et al., 1996a; Sukharev et al., 1999b), structural probing with IR and CD (Arkin et al., 1998), and extended electrophysiology (Sukharev et al., 1999c). Additionally, homologues from seven varied bacterial species were cloned and found to have mechanosensitive activity analogous to Ec-MscL (Moe et al., 1998). The ability of MscL to rescue an osmotically sensitive bacterium, V. alginolyticus, from osmotic downshock supports proposals that MscL acts as a “release valve” to reduce membrane tension during osmotic stress (Nakamaru et al., 1999).

The crystal structure of the M. tuberculosis MscL (Tb-MscL) homologue (Fig. 2.1) solved by the Rees group (Chang et al., 1998) gives further insight into previous studies while provoking new questions about MscL. The crystal structure gives unique opportunities for structure-function studies employing a variety of computational and experimental techniques. The relatively small size of MscL and the ability to produce
reasonably large quantities of it from bacterial cultures also makes it a promising model system for studying general principles of mechanosensitive ion channels.

**Figure 2.1:** The Tb-MscL crystal structure with the five identical subunits shown in different colors. The interaction between Gln 51 and Arg 45 residues on adjacent subunits is highlighted.

However, there is some ambiguity in the crystal structure of Tb-MscL. The crystallization was performed under non-physiological conditions, at low pH and in the presence of heavy metal ions (Chang et al., 1998), which has caused some concern about the relevance of the structure to physiological conditions (Oakley et al., 1999). Additionally, the pentameric structure was at odds with several studies claiming MscL was a homohexamer, including biochemical cross-linking (Blount et al., 1996a; Sukharev
et al., 1999b), estimates of pore size (Cruickshank et al., 1997), and a two-dimensional crystal structure (Saint et al., 1998).

Concerns over Tb-MscL multimerization could be addressed by designing an intersubunit crosslink using the crystal structure. These designed reactions would give a more definitive answer to the multimerization than the use of non-specific crosslinkers. As well, the success of a designed reaction would support the validity of the structure in the region of the design. This would be particularly desirable since the electron density for some regions, such as the loop region, was more ambiguous than for other regions of Tb-MscL, such as the transmembrane helices. Also, as discussed below, the extracellular loop region exhibits interesting sequence diversity among MscL homologues.

An apparent intersubunit hydrogen-bond in the crystal structure between R45 and Q51 can be ideally exploited for designed cross-linking reactions (Fig. 2.1). To this end, an R45K/Q51E mutant was made to allow intersubunit amide bond formation mediated by the peptide bond forming reagents EDC or DCC. However, EDC is known to cause some background cross-linking in wild-type EcMscL (Sukharev et al., 1999b). This led to the production of an R45C/Q51C mutant which can be crosslinked through disulfide bond formation with Cu(phen)$_3$ or with bifunctional bismaleimide crosslinkers of differing lengths: BMOE, BNDB, and BMH. Structures of these crosslinkers are shown in Fig. 2.2. The absence of native cysteines in TbMscL will preclude background reactions for these cases.
Production of MscL Proteins and Cross-linking Protocols

All mutations were generated using the QuikChange Method (Stratagene) on a pET 19b (Novagen) construct containing the *M. tuberculosis* MscL open reading frame (Chang et al., 1998). Mutations were confirmed by enzymatic digest and sequencing. Expression was carried out in a MscL-null strain of BL21(DE3) *E. coli* (Chang et al., 1998). All bacterial growth was done in the presence of 100 µg/mL ampicillin.

Protein expression was performed by growing cells at 37 °C to the midpoint of log phase (OD$_{600} \approx 0.6$-0.8) and inducing with 0.1% IPTG and 1% lactose. Following induction, cells were grown for an additional 2 hours, harvested, and solubilized in 1% DDM, 10 mM TRIS, and 10 mM NaCl. Protein was purified on a nickel-chelation column (Qiagen) in the presence of 0.05% DDM. The resulting proteins were identified by MALDI-TOF mass spectral analysis.

Wild-type or R45K/Q51E protein solubilized in DDM micelles was diluted to a concentration of approximately 25 µg/mL and cross-linked at 4 °C for 2 hours using 10 mM EDC, 10 mM DCC, 10 mM EDC/10 mM Sulfo-NHS, or 10 mM DCC/10 mM NHS. All cross-linking reactions were quenched with SDS-PAGE loading buffer containing D-mercaptoethanol. Reaction products were run on 4-15% gradient polyacrylamide gels.
and visualized by Western blotting with either a 6-His Antibody (Amersham) or INDIA HisProbe-HRP (Pierce). Cysteine cross-linking reactions were performed and assayed in a similar manner on wild-type and R45C/Q51C Tb-MscL. In these reactions, thioethers were formed with bis-maleimide reagents (Pierce) or disulfide bond formation was mediated by 3 mM copper phenanthroline. For the disulfide formation reactions with copper phenanthroline, D-mercaptoethanol was omitted from the loading buffer.

**Results of Cross-linking Studies**

A typical SDS-PAGE Western blot of cross-linking products is shown in Fig. 2.3. All reagents showed at least a weak pentameric band in the mutant with slight background cross-linking in wild-type. The background cross-linking is most likely due to cross-linking in the carboxy terminus of the protein, which contains a number of glutamates, aspartates and lysines.

The most interesting cross-linking results were seen with EDC and Sulfo-NHS. This combination gives mainly pentamer and tetramer for cross-linked products (Fig. 2.3). The strong pentameric band in this designed system provides the best evidence to date that Tb-MscL is pentameric under physiological conditions. Other cross-linking studies typically show progressively weaker band intensities on going from monomer to dimer to trimer, etc., analogous to our results with just EDC and other non-optimal conditions (Fig. 2.3) (Blount et al., 1996a; Hase et al., 1997; Sukharev et al., 1999a). Such observations always leave open the possibility that a hexamer band is present, but is too weak to be seen as the intensity progressively falls off with higher oligomerization. In fact, under some conditions, a weak band assigned to hexamer is occasionally seen in
our reactions. However, with the designed double mutant under appropriate conditions (EDC/Sulfo-NHS), very strong tetramer and pentamer bands are seen, but no hexamer band is visible. This provides compelling evidence that no significant fraction of Tb-MscL is present in hexameric (or higher oligomerization) states when reconstituted in DDM micelles. Most likely, crosslinked complexes with more than five subunits in previous work resulted from cross-linking between subunits in nearby but distinct channels or cross-linking of malformed channels.

Figure 2.3: Cross-linking of the R45K/Q51E mutant of M. tuberculosis MscL. Purified wild-type and R45K/Q51E M. tuberculosis MscL proteins were cross-linked for 2 hours at 4 °C using EDC, DCC, EDC with sulfo-NHS, and DCC with NHS. The reactions were quenched with β-mercaptoethanol, run on a 4-15% SDS-polyacrylamide gel and visualized by Western blotting with 6-His antibody.
Cross-linking studies using R45C/Q51C Tb-MscL did not produce the quantitative results observed with R45K/Q51E Tb-MscL. However, high molecular weight bands were observed upon cross-linking, either by the formation of disulfide bonds between subunits or by reactions with bis-malimide reagents (Fig. 2.4). Since there are no cysteines other than the designed mutations in MscL, the occurrence of significant cross-linking further supports the crystal structure conformation of the Tb-MscL extracellular loop region. Interestingly, when R45C/Q51C Tb-MscL was cross-linked using bis-malimide reagents of varying tether lengths, no consistent distance dependence was observed for cross-linking efficiency.

**Figure 2.4:** Cross-linking of the R45C/Q51C mutant of *M. tuberculosis* MscL. Purified wild-type and R45C/Q51C *M. tuberculosis* MscL proteins were cross-linked for 2 hours at 4 °C using copper phenanthroline, BMH, BMDB, or BMOE. The reactions were run on a 4-15% SDS-polyacrylamide gel and visualized by Western blotting with 6-His antibody. The copper phenanthroline reactions were run in the absence of β-mercaptoethanol.
Sequence Analysis of the MscL Channel Family

Since the cross-linking studies helped confirm the crystal structure conformation of Tb-MscL, we began to consider how that channel compared to the well-characterized Ec-MscL and other MscL homologues. One way to perform such comparisons is through bioinformatics methods that compare the primary sequences of different channels and determine the extent of differences between members of the MscL family.

As a first step, the sequences of 35 MscL homologues were obtained from BLAST searches, including searches of several sequenced and partially sequenced bacterial genomes. A multiple sequence alignment of the sequences was obtained using AMPS (Alignment of Multiple Sequences) (Barton, 1990; Barton and Sternberg, 1987). Although clearly related, the mechanosensitive channels from various organisms show moderate to low sequence identities. For example, the sequence identity of Tb-MscL compared to Ec-MscL is 37%, while the sequence identity of *B. bronchiseptica* MscL compared to *M. leprae* MscL is 15%. Therefore, development of an optimal alignment was not straightforward. For this reason, we augmented the sequence alignment approach with MEME analysis (Bailey and Elkan, 1994; Bailey and Gribskov, 1998), which identifies conserved regions within a group of sequences. The AMPS multiple sequence alignment was slightly adjusted to align corresponding MEME groups from different sequences. Fig. 2.5 shows the final multiple sequence alignment and MEME group analysis of the 35 putative MscL sequences. The alignment was divided into regions—extracellular loop, carboxy terminus, and transmembrane regions one and two—using the helix definitions of Chang *et al.* (1998). The extracellular loop is defined as the region between the first and second transmembrane domains, and the carboxy
Figure 2.5: MEME consensus group analysis shown on the AMPS multiple sequence alignment of 35 putative MscL homologues. The colored regions on the sequence alignment indicate the MEME consensus groups.
terminus is the region from the end of the second transmembrane domain to the end of
the carboxy helix.

The MEME sequence analysis gives insight into the overall similarity of the
MscL homologues. Not surprisingly, the homologues are most similar in the
transmembrane regions and most divergent in the loop and carboxy terminus regions.
The strong similarities in the transmembrane domains are highlighted by the fully
conserved groups, II and VIII, and the highly conserved group III. Additionally,
members of the MscL family that lack group III in the first transmembrane region tend to
have a similar conserved group IV in this region.

The carboxy terminus and loop region are much less conserved. Despite the
appearance of three consensus groups in the loop region—V, VI and VII—these groups
are clearly not universal. The carboxy terminus is more highly conserved than the loop
region, containing two very highly conserved groups (IX and XIII), but it is clearly not as
well conserved as the transmembrane helices. Mycobacteria do not contain group IX, but
an analogous charged region is evident (group X). Previously, it has been shown that a
large portion of the carboxy terminus in Ec-MscL can be deleted without affecting
protein function (Blount et al., 1996b). Although the same may not be true for all
homologues, this is consistent with the lack of sequence conservation in this region.

To further examine the similarities and differences among MscL homologues, a
pairwise alignment of the various regions, such as the first and second transmembrane
domains, the extracellular loop, and the carboxy terminus, was employed. Pairwise
alignments of the various regions were performed using AMPS, and scores for each pair
were summarized as contour plots. Scores reflect the alignment of sequence A to
sequence B relative to a shuffled sequence B and are therefore corrected for length.
Scores above 5 indicate very good alignment between two protein sequences, scores
between 2 and 5 indicate moderate alignment, and scores below 2 indicate poor
alignment. Contour plots showing scores for the AMPS pairwise alignments of the first
transmembrane domain, the extracellular loop, and the carboxy terminus are shown in
(Fig. 2.6).

The pairwise alignments showed the same general trends observed with MEME
analysis. In general, all regions of the MscL sequence are conserved, however the loop
region has pairs of sequences with poor alignment. To some extent the sequence pairs
within the loop region can be used to group the homologues into subfamilies. The largest
and most obvious subfamily includes *E. coli* MscL and other sequences containing
MEME group VI. Another distinctive subfamily includes the Mycobacteria. Thus, based
on their primary sequences, Ec-MscL and Tb-MscL appear to reside in different
subfamilies.
Figure 2.6: Regional AMPS pairwise alignments for the first transmembrane domain (A), the loop region (B), and the carboxyl terminus (C). Numbers on axes correspond to the sequence numbers in Figure 2.5. The grouping of sequences into two main subfamilies can be seen for the loop region (B), with a large subfamily containing Ec-MscL and a small subfamily containing Tb-MscL.
Analysis of the MscL Channel Family with Circular Dichroism

The secondary structure of MscL homologues was also compared using circular dichroism to measure relative helicity. To do this, protein samples were purified for the nine different MscL homologues that have been characterized by electrophysiology (Moe et al., 1998; Moe et al., 2000). This protein purification was done using the same methods described above to produce MscL proteins for cross-linking reactions. These protein samples were then used to obtain circular dichroism spectra on an Aviv 42a DS circular dichroism spectrometer using a strain-free circular cuvette with a pathlength of 0.1 cm. Spectra were collected between 260 nm and 185 nm and averaged over three scans. All data were collected at room temperature. Concentrations for conversion to molar ellipticity units were obtained using the BioRad DC compatible protein concentration kit and the Pierce BCA protein concentration kit.

The spectra obtained for these homologues are shown in Fig. 2.7, and selected features are listed in Table 2.1. As observed with the sequence alignment of MscL homologues, the homologues fall into two distinct families. However, the families resulting from the circular dichroism studies are different than those observed from the sequence alignment data. Based on the observed circular dichroism spectra, the first family is composed of *E. coli*, *H. influenza*, and *P. fluorescens* and the second family is composed of *M. tuberculosis*, *E. carotovora*, *C. perfringens*, *S. aureus*, *B. subtilis*, and *Synechocytis sp.*
Using the ellipticity at 222 nm observed in the circular dichroism spectra, the percent helicity of each MscL homologue was estimated as previously described for Ec-MscL (Table 2.1) (Arkin et al., 1998). The two sequence families show dramatic differences in their predicted helical content, with the first family exhibiting helical contents of 85-110% and the second family having a helical content between 25% and 55%.

To determine if the relative helicity values were predicted from differences in primary sequence, the amount of helix in each homologue was predicted using Jpred, a program that combines results from several independent secondary structure prediction methods. The predicted helicity values obtained for the MscL homologues using Jpred
are quite similar (Table 2.1), and would not predict two families as observed. It is true that the highest predicted value is for In-MscL, which lies in the Ec-MscL family, and the lowest predicted value is for Clo-MscL, which lies in the Tb-MscL family. However, in general, there is no trend between predicted helical secondary structure and the helical content determined by circular dichroism. In fact, it should be noted that Pf-MscL, which is in the Ec-MscL family with high apparent helicity, exhibits one of the lowest values for predicted helicity. This inconsistency could point to a flaw in Jpred for predicting $\alpha$-helicity for MscL homologues. However, other structural features, such as tertiary structure and quaternary assembly, can affect observed circular dichroism signals. Thus, the differences in CD spectra could result from different folding or assembly of some portion of MscL channels in different subfamilies.

Table 2.1: A summary of the circular dichroism spectra for various homologues of MscL. Helical content for the various MscL homologues was determined from the absorbance in the CD spectrum at 222 nm using the method previously described for Ec-MscL (Arkin et al., 1998). Jpred predictions of helical content are based on sequence analysis.

<table>
<thead>
<tr>
<th>Species</th>
<th>Abbreviation</th>
<th>Sequence number</th>
<th>Helical Content from CD</th>
<th>Helical Content from Jpred</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>P. fluorescens</em></td>
<td>Pf</td>
<td>1</td>
<td>85-51%</td>
<td>48%</td>
</tr>
<tr>
<td><em>H. influenza</em></td>
<td>In</td>
<td>10</td>
<td>95-57%</td>
<td>56%</td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td>Ec</td>
<td>4</td>
<td>110-66%</td>
<td>53%</td>
</tr>
<tr>
<td><em>E. carotovora</em></td>
<td>Er</td>
<td>6</td>
<td>55-33%</td>
<td>49%</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>Tb</td>
<td>31</td>
<td>42-25%</td>
<td>45%</td>
</tr>
<tr>
<td><em>C. perfringens</em></td>
<td>Clo</td>
<td>25</td>
<td>31-18%</td>
<td>35%</td>
</tr>
<tr>
<td><em>S. aureus</em></td>
<td>Sa</td>
<td>20</td>
<td>44-26%</td>
<td>50%</td>
</tr>
<tr>
<td><em>B. subtilis</em></td>
<td>Bs</td>
<td>18</td>
<td>39-23%</td>
<td>52%</td>
</tr>
<tr>
<td><em>Synechocystis sp.</em></td>
<td>Syn</td>
<td>35</td>
<td>25-15%</td>
<td>50%</td>
</tr>
</tbody>
</table>

Since MscL homologues vary considerably in length from 143 amino acids to 175 amino acids, the observed helicity could be related to protein length. Fig. 2.8 shows a
plot of the maximal helicity predicted for each homologue as a function of homologue length. Clearly, the predicted helicity from circular dichroism does not correlate with protein length. The longest and shortest homologues both lie in the Tb-MscL family that displays relatively low helicity.

**Figure 2.8:** Comparison of protein length to the maximal predicted helical content for the various homologues of MscL. No clear trends between protein length and helical content are observed.

Interestingly, although Tb-MscL and Ec-MscL are in different MscL sub-families in both the sequence comparisons and CD analysis, the CD-based sub-families are somewhat different than those in the sequence analysis. As well, on the surface the Tb-MscL family from CD is composed of seemingly poorly related sequences. For example, some of the sequences in the Tb-MscL family contain MEME group V and/or MEME group XII, while other members lack these groups. Nonetheless, all of the members of the Ec-MscL family contain MEME group VI, which is located in the loop region. Similarly, none of the members of the Tb-MscL family, except for Er-MscL, contain
MEME group VI. Additionally, Er-MscL has the largest helicity of the Tb-MscL family. Thus, the structural differences between homologue subfamilies observed by CD may be due to interactions in the second half of the extracellular loop region. Future studies of MscL channels with chimera loop regions could give insight into this possibility.

**Summary**

The absence of crosslinked protein with greater multimerization than pentamer supports the formation of a pentameric MscL structure under physiological conditions. Also, these results provide an important biochemical verification of the crystal structure in a relatively ambiguous region. This implies that the structure can be used as a basis for interpreting previous experimental data on MscL. As well, the success of the designed reactions shows that the Tb-MscL crystal structure could be used as a basis for other design efforts on the channel.

Using the sequence alignment in Figure 2.5, there is no obvious Ec-MscL analogue to the R45•••Q51 hydrogen bond seen in Tb-MscL. Technically, the alignment is L47/D53 (Ec-MscL numbering), which is not a favorable interaction. There is no cationic or hydrogen bond donating residue near L47 that could pair with D53. However, residues on either side of D53 are hydrophobic, suggesting that perhaps the ion pair of Tb-MscL is replaced by a hydrophobic contact such as L47/I52 or L47/F54 in Ec-MscL. Thus, in addition to investigating the physiological role of the R45•••Q51 interaction in Tb-MscL, it would be interesting to investigate the role of potentially analogous interactions in Ec-MscL. Also, these putative hydrophobic residue pairs in the Ec-MscL
loop may provide a site for design of a cross-linking reaction similar to that designed for Tb-MscL.

The differences among the families of MscL homologues obtained from sequence analysis and circular dichroism are quite intriguing. In particular, it is interesting that Tb-MscL and Ec-MscL consistently lie in different subfamilies. Thus, although there are clear overall similarities between those homologues, researchers may need to be careful before assuming results on one channel necessarily reflect on the other. In fact, electrophysiological measurements have shown that Tb-MscL opens at a much greater tension than Ec-MscL in E. coli spheroplasts (Moe et al., 2000). Additional studies comparing members of different MscL subfamilies in detail could give further insights into how differences in sequence and helicity affect channel function.

References


