Chemical ligation of the influenza M2 protein for solid-state NMR characterization of the cytoplasmic domain

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1002/pro.2690">http://dx.doi.org/10.1002/pro.2690</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>John Wiley &amp; Sons, Inc.</td>
</tr>
<tr>
<td>Version</td>
<td>Author’s final manuscript</td>
</tr>
<tr>
<td>Citable link</td>
<td><a href="http://hdl.handle.net/1721.1/107217">http://hdl.handle.net/1721.1/107217</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/4.0/">http://creativecommons.org/licenses/by-nc-sa/4.0/</a></td>
</tr>
</tbody>
</table>
Chemical Ligation of the Influenza M2 Protein for Solid-State NMR Characterization of the Cytoplasmic Domain

Byungsu Kwon, Daniel Tietze, Paul B. White, Shu Y. Liao and Mei Hong*
Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139

Revised for Protein Science

* Corresponding author: Mei Hong, Tel: 617-253-5521, Email: meihong@mit.edu

§ Current address: Eduard-Zintl-Institut fur Anorganische und Physikalische Chemie, Technische Universitat Darmstadt, Alarich-Weiss-Str. 8, 64287, Darmstadt, Germany

Running title: Chemical Ligation of Influenza M2 for Solid-State NMR
Abstract

Solid-state NMR-based structure determination of membrane proteins and large protein complexes faces the challenge of limited spectral resolution when the proteins are uniformly $^{13}$C-labeled. A strategy to meet this challenge is chemical ligation combined with site-specific or segmental labeling. While chemical ligation has been adopted in NMR studies of water-soluble proteins, it has not been demonstrated for membrane proteins. Here we show chemical ligation of the influenza M2 protein, which contains a transmembrane (TM) domain and two extramembrane domains. The cytoplasmic domain, which contains an amphipathic helix (AH) and a cytoplasmic tail, is important for regulating virus assembly, virus budding, and the proton channel activity. A recent study of uniformly $^{13}$C-labeled full-length M2 by spectral simulation suggested that the cytoplasmic tail is unstructured. To further test this hypothesis, we conducted native chemical ligation of the TM segment and part of the cytoplasmic domain. Solid-phase peptide synthesis of the two segments allowed several residues to be labeled in each segment. The post-AH cytoplasmic residues exhibit random-coil chemical shifts, low bond order parameters, and a surface-bound location, thus indicating that this domain is a dynamic random coil on the membrane surface. Interestingly, the protein spectra are similar between a model membrane and a virus-mimetic membrane, suggesting that the structure and dynamics of the post-AH segment is insensitive to the lipid composition. This chemical ligation approach is generally applicable to medium-sized membrane proteins to provide site-specific structural constraints, which complement the information obtained from uniformly $^{13}$C, $^{15}$N-labeled proteins.
Introduction

Multidimensional magic-angle-spinning (MAS) solid-state NMR (SSNMR) spectroscopy is a powerful atomic-resolution technique to elucidate the structure and dynamics of non-crystalline and disordered proteins bound to phospholipid membranes (1-4). However, many membrane proteins exhibit conformational disorder and plasticity, thus giving partially overlapping resonances that are difficult to assign completely, which preclude the determination of their three-dimensional structures. In addition, the high concentration of a few types of hydrophobic residues with relatively uniform (ϕ, ψ) torsion angles in membrane proteins reduces the chemical shift dispersion of the NMR spectra. As a result, multidimensional SSNMR spectra of recombinant and uniformly 13C-labeled membrane proteins are challenging to analyze. Biosynthetic sparse 13C labeling using specifically 13C-labeled glucose or glycerol alleviates this difficulty by skipping the labeling of some of the carbons in the amino acid residues (5-8). However, resonance assignment of such selectively labeled proteins usually requires the same 2D and 3D experiments as those developed for uniformly labeled proteins, thus multiple labeled samples are necessary to piece together the complete assignment. In addition to disordered membrane proteins, many membrane proteins contain large water-soluble domains whose crystal structures have been determined and small hydrophobic domains whose structures are unknown. These small hydrophobic domains can play important roles in the biological function of the intact protein. NMR-based structural studies may provide the most mechanistic insights by focusing on these hydrophobic domains. Recombinant expression of these large proteins not only is challenging but also limits the options of isotopic labeling, making it difficult to selectively detect the hydrophobic domains.

The challenges in assigning the SSNMR spectra of disordered and large membrane proteins can in principle be met by chemical ligation of two or more protein segments (9-12), in which either one of the segments is labeled or both segments are labeled but only at specific residues. The individual segments can be produced by solid-phase peptide synthesis when they are sufficiently small or by recombinant protein expression when they are large. One of the most common ligation mechanisms is thiolate-thioester exchange between an N-terminal cysteine residue of segment 2 and a C-terminal thioester of segment 1 to form the requisite peptide bond (11). This so-called native chemical ligation approach uses unprotected peptide segments and is carried out under neutral aqueous conditions. The same ligation chemistry can also be mediated by inteins, which are naturally occurring proteins that catalyze their own excision from the precursor proteins.

Ligation-based segmental isotopic labeling has been applied to some extent in solution NMR studies of water-soluble multi-domain proteins and glycoproteins (13; 14), and has recently been demonstrated for an amyloid protein (15). However, to our knowledge, ligation-based isotopic labeling has not been demonstrated in SSNMR studies of membrane proteins. This is likely due to the difficulties of producing and solubilizing sufficient quantities of hydrophobic membrane peptides and proteins. For low-resolution biophysical and biochemical characterizations that do not require a large quantity of proteins, chemical ligation has been achieved for several membrane proteins such as potassium channels (16-19), mechanosensitive channels (20), HIV-1 Vpu (21), and the influenza M2 protein (22). Full-length influenza M2 was synthesized by ligating M2(1-49) and M2(50-97), each produced by Boc-chemistry solid-phase peptide synthesis (SPPS) (22). The resulting synthetic full-length M2 was used to investigate the
thermodynamic stability of the tetrameric assembly of the intact protein compared to that of the TM peptide. However, no site-specific structural information was obtained from this study.

In this work, we adopt Fmoc SPPS and native chemical ligation to produce a M2 construct that includes the TM domain and part of the cytoplasmic domain, in order to investigate the site-specific conformation and dynamics of the cytoplasmic tail of this protein by SSNMR. The full cytoplasmic domain of M2 spans residues 47-97, and plays important roles in infectious virus production (23), viral morphogenesis (24), binding to other influenza virus proteins such as the matrix protein 1 (M1) (25; 26), and regulation of the proton-channel activity (27). The first 16-residues of the cytoplasmic domain (residues 47-62) forms a membrane-surface-bound amphipathic helix (AH) (28-33) and is responsible for membrane scission during virus release (34). However, the structure of the post-AH cytoplasmic tail is still elusive. In a recent study of recombinant uniformly $^{13}$C-labeled full-length M2 (35), we reported chemical shift analysis of the cytoplasmic domain without complete resonance assignment. The protein was reconstituted into two lipid membranes, 1,2-dimyristoyl-snglycero-3-phosphocholine (DMPC) and a complex virus-mimetic (VM+) membrane containing cholesterol, sphingomyelin in addition to phospholipids. The DMPC-bound full-length M2 exhibited a significant number of overlapped peaks at random-coil chemical shifts, and spectral simulation by calculating the chemical shifts of various conformational models suggested that the cytoplasmic domain is unstructured. The spectra of the VM+ bound protein have broader linewidths than the DMPC-bound protein, thus no structural conclusion was made. To understand the M2 cytoplasmic structure in a site-specific fashion and how it depends on the membrane composition, we have now conducted total chemical synthesis of M2(22-71), by ligating M2(22-49)-thioester with M2(50-71). We show that the yield of native chemical ligation is sufficiently high to enable SSNMR characterization of the ligated protein in DMPC and VM+ membranes. We find that the post-AH cytoplasmic tail up to residue 71 is unstructured, highly dynamic, and bound to the membrane surface, thus validating the previous chemical-shift analysis. Interestingly, the chemical shifts of the cytoplasmic residues are similar between the two membranes, thus ruling out substantial conformational changes of this post-AH cytoplasmic segment by cholesterol and sphingomyelin.

**Results**

**Native Chemical Ligation of M2(22-71)**

The M2(22-71) construct was chosen because it is short enough to be amenable to synthesis yet long enough to show full proton channel activity (27; 36) and partial virus-assembly activity (25). The protein was synthesized by assembling M2(22-49)-thioester and M2(50-71) using thio-ligation chemistry (Fig. 1b). Although the 50-residue construct may be synthesizable without ligation, the yield would be much less than 1% if we assume an average coupling efficiency of 90% per residue, considering the hydrophobic nature of the N-terminal half of this protein. The chemo-selective ligation reaction between the two peptides was initiated by the N-terminal cysteine residue of M2(50-71) attacking the M2(22-49) thioester, giving rise to a thioester-linked intermediate (9; 11). This thiol-thioester exchange is reversible at neutral pH in the presence of the catalyst MPAA, but is followed by a rapid, irreversible intramolecular S-to-N acyl shift reaction to create the thermodynamically favored peptide bond at the ligation site (37).
One of the major obstacles of membrane protein chemical ligation is solubilizing the hydrophobic segments in the solution for the ligation reaction. We first tested a 48 mM dodecylphosphocholine (DPC) and 8 M urea solution to solubilize the hydrophobic M2(22-49). HPLC and ESI-MS analysis showed that the ligation reaction proceeded in this solution, but DPC was difficult to remove completely from the product. We next tested 100 mM octyl-β-D-glucopyranoside (OG) for solubilizing M2(22-49), but the ligation reaction did not proceed in this solution, as shown by HPLC. A TFE/guanidinium chloride solution similarly did not allow ligation to proceed. Finally, we carried out the ligation reaction in a 60% TFE solution containing 5.4 M urea at 35°C, which enabled both efficient solubilization of the M2(22-49)-thioester and the ligation reaction.

The combined synthesis and purification yields were ~10% for the hydrophobic M2(22-49)-thioester and ~30% for the more polar M2(50-71). Once the two segments were successfully synthesized, as verified by HPLC and ESI-MS (Fig. 2), we monitored the progress of the ligation reaction by analytical HPLC. At 4 hours, the M2(22-49)-thioester peak at 24.3 min was dominant, while the M2(22-71) peak appeared at 20.5 min. The elution time difference is similar to that reported between full-length M2 and M2(1-49)-thioester (22). The M2(22-49)-thioester peak intensity decreased over time while the M2(22-71) intensity increased. The reaction reached completion after 18 h. Longer reaction times did not increase the yield of the ligation product further.

Conformation of membrane-bound M2(50-71) and M2(22-71)

The main purpose of the present study is to measure site-specific conformation of the cytoplasmic tail C-terminal to the amphipathic helix. Previous 2D 13C-13C correlation spectra of uniformly 13C-labeled full-length M2 suggested that the cytoplasmic tail may be a random coil. However, the broad linewidths of the spectra at low temperature precluded full resonance assignment, and chemical shift prediction and spectral simulation were used instead to obtain this information (35; 38). With the ability to synthesize the cytoplasmic peptide, we can now obtain site-specific conformation of the post-AH cytoplasmic residues from chemical shifts. We incorporated four 13C, 15N-labeled residues in the region S64-V68. In addition, several structurally known TM residues were labeled to monitor the conformation of the TM domain.

Fig. 3 shows 1D 13C cross-polarization (CP) and direct-polarization (DP) spectra of M2(50-71) and M2(22-71) in various lipid membranes. The spectra were measured in both the gel phase and liquid-crystalline phase of the membranes to detect the rigid-limit conformation and the dynamics of the peptides, respectively. Resonance assignments were confirmed from 2D 13C-13C correlation spectra (Fig. 4). At high temperature, POPC/POPG-bound M2(50-71) exhibits narrow but weak 13C signals in the CP spectrum and much stronger signals in the DP spectrum, indicating that this cytoplasmic peptide is highly mobile. The peptide 13C signals in the CP spectrum are much weaker than the lipid signals despite 13C labeling, indicating that M2(50-71) motion interferes with dipolar-based CP transfer. The resolved V68 and S64 Cα peaks in the DP spectrum both appear at random-coil chemical shifts, suggesting that this segment is unstructured. In the gel-phase membrane, all backbone Cα signals become severely broadened, indicating a large conformational distribution. The combination of narrow linewidths at high temperature and line broadening at low temperature resembles the behavior of HIV TAT(48-60), a cell-penetrating peptide that also exhibits a random coil conformation (39).
Compared to M2(50-71), DMPC-bound M2(22-71) gave better resolved $^{13}$C CP spectra at low temperature, but all resolved $^{13}$C signals come from the TM residues A30, G34, and I35 (Fig. 3b), while the signals of the labeled cytoplasmic residues remain narrow at high temperature and severely broadened at low temperature. Thus, the post-AH cytoplasmic tail adopts the same disordered conformation in the ligated product as in the M2(50-71) peptide. The TM residues’ Cα and Cβ chemical shifts are similar to those found in the TM peptide (40; 41), and indicate well defined α-helical conformation, confirming that chemical ligation does not perturb the TM helix structure.

To investigate if the lipid composition affects the M2(22-71) conformation, we also reconstituted the ligated protein into the virus-mimetic VM+ membrane, which contains cholesterol and sphingomyelin in addition to phospholipids (42). Recently reported 2D $^{13}$C-$^{13}$C correlation spectra of uniformly $^{13}$C-labeled M2FL showed moderate differences in some of the correlation peaks (35), suggesting that the membrane composition may induce small conformational changes of the protein between a non-cholesterol and a cholesterol-containing membrane. However, Fig. 3c shows that the membrane composition change did not cause noticeable differences to the 1D $^{13}$C spectra compared to the DMPC data, suggesting that the lipid composition may not cause significant conformational changes to the TM domain and the post-AH cytoplasmic tail up to residue 71.

2D $^{13}$C-$^{13}$C correlation spectra (Fig. 4) allowed the $^{13}$C chemical shifts to be fully resolved and assigned. For POPC/POPG-bound M2(50-71) at low temperature, S64 shows a Cα-Cβ cross peak well away from the diagonal, with a chemical shift difference of 5.3 ppm (Table 1), consistent with a random-coil conformation (38). The T65, E66, and V68 Cα and Cβ chemical shifts are generally indicative of coil and sheet secondary structures (38). The low-temperature 2D spectra of DMPC- and VM+ bound M2(22-71) (Fig. 4b, e) highlight the challenges of studying membrane proteins that contains a well-structured domain together with a disordered domain. The three TM residues exhibit relatively narrow linewidths and high intensities while the cytoplasmic residues give much broader and weaker peaks. The $^{13}$C chemical shifts of A30, G34 and I35 are consistent with those of the TM peptide (40; 41; 43) while the cytoplasmic residues’ chemical shifts are identical to those of M2(50-71). The similarity of TM residues’ chemical shifts between the long peptide and the short TM constructs is also verified in 2D $^{15}$N-$^{13}$C correlation spectra (Fig. 5).

The similarity of the chemical shifts of the S64-V68 segment in M2(50-71) and M2(22-71) indicates that membrane anchoring by the TM helix and anionic lipids do not affect the unstructured nature of the post-AH segment, thus the random coil conformation is likely intrinsic to the amino acid sequence of this domain. Increasing the temperature to the liquid-crystalline phase and using direct $^{13}$C polarization gave rise to much narrower peaks for the cytoplasmic residues in the 2D spectra (Fig. 4c, f), consistent with the 1D $^{13}$C DP spectra. The average Cα and Cβ linewidths of the four labeled cytoplasmic residues are 1.5 and 2.2 ppm in the gel-phase DMPC and VM+ membranes, respectively, but decrease to 0.8 ppm in the liquid-crystalline phase of both membranes.
The chemical shifts of DMPC-bound M2(22-71) can also be compared to those of DMPC-bound M2FL (Fig. 4d) (35). It can be seen that the α-helical chemical shifts are well matched between the synthetic M2(22-71) and the recombinant full-length protein, and the S64, T65 and V68 peaks also fit within the Ser, Thr and Val regions of the M2FL spectrum.

The 2D correlation spectra of VM+ bound M2(22-71) mirror those of the DMPC-bound protein, with the same chemical shifts for most sites except for some of the A30 and I35 sidechain peaks (Table 1). Thus, the membrane composition has a minor effect on the TM helix conformation and no effect on the cytoplasmic conformation up to residue 71. The latter is consistent with the notion that the cytoplasmic tail is predominantly influenced by the aqueous environment, and is thus relatively insensitive to changes in the membrane fluidity, lateral pressure and other physical properties.

**Depth of insertion of M2(22-71) in the lipid membrane**

To investigate if the post-AH cytoplasmic segment lies on the membrane surface, we measured the depth of insertion of M2(22-71) in a site-specific manner using the gel-phase 2D $^1$H-$^13$C correlation experiment (44) (Fig. 6a) with $^1$H spin diffusion mixing times of 4 and 25 ms. At 4 ms, A30 and G34 Cα peaks show much higher lipid CH$_2$ cross peaks than water cross peaks (Fig. 6b), indicating the membrane-embedded nature of these residues (45). Viewed along the $^1$H cross sections, S64 and V68 show a dominant water cross peak, followed by the lipid CH$_2$ and headgroup H$_7$ peaks, indicating that these residues are much closer to water than the membrane interior. When the $^1$H cross sections of various $^13$C signals are normalized to the lipid CH$_2$ cross-peak at 1.4 ppm, S64 and V68 exhibit the highest water cross peak, followed by A30 and G34, while the lipid acyl chains have negligible water intensity. Thus, the TM residues have more water contact than lipids, consistent with the fact that they surround a water-filled pore, but the water exposure of the TM residues is much lower than that of S64-V68, indicating that the cytoplasmic tail is exposed to the large aqueous environment on the membrane surface. When the $^1$H spin diffusion mixing time increased to 25 ms, the TM and cytoplasmic residues show the same relative lipid and water cross-peak intensities, indicating that the water magnetization has equilibrated in the protein by this time.

**Mobility of membrane-bound M2(22-71)**

An implication of the random coil conformation is the presence of large-amplitude motion. To quantitatively determine the motional amplitudes of the cytoplasmic residues, we measured the $^{13}$C-$^1$H dipolar order parameters of DMPC-bound M2(22-71) at 303 K using the 2D DIPSHIFT experiment. The TM helix is expected to be immobilized and have large order parameters, while the post-AH cytoplasmic residues are expected to have low order parameters. Fig. 7 shows representative C-H dipolar dephasing curves of the TM and cytoplasmic residues. Since the individual cross sections of residues within each domain do not differ within experimental uncertainty, we summed the relevant cross sections to obtain higher sensitivity. The Cα sites of the TM residues exhibit deep dephasing curves that translate to order parameters of 0.94 – 1.0, while the cytoplasmic residues exhibit shallower dephasing curves that correspond to order parameters of 0.44 – 0.50. Therefore, the cytoplasmic segment is indeed highly dynamic, consistent with its random coil chemical shifts.

**Discussion**
Structural studies of the influenza M2 protein has so far largely focused on the TM domain, to understand the tetrameric assembly, lipid- and detergent-bound structures, the drug binding site, and the proton conduction mechanism (for reviews, see ref (46-50)). In comparison, the N-terminal ectodomain and the C-terminal cytoplasmic domain are structurally mostly unknown. These two domains are involved in a number of functions such as antibody binding, M1 binding, regulation of virus morphology and virus budding (24; 25; 27; 51-53). The best available structural information so far about these extra-membrane regions concerns the amphipathic helix immediately C-terminal to the TM helix. $^{13}$N orientational NMR constraints showed that this AH domain lies on the membrane surface after a tight turn from the TM helix (28; 29). This conserved AH domain causes high membrane curvature, which is necessary for membrane scission and virus release (34; 54). However, the rest of the cytoplasmic domain is structurally unknown.

Recently, a 2D SSNMR study of uniformly $^{13}$C-labeled full-length M2 was conducted that suggested that the post-AH cytoplasmic domain and the ectodomain are mostly unstructured (35). This conclusion came from chemical shift prediction (38) and simulations of overlapped 2D $^{13}$C-$^{13}$C correlation spectra, thus it is limited by the fact that alternative conformational models may also explain the measured spectra. This motivated us to explore chemical ligation to obtain site-specific structural constraints about the cytoplasmic domain. M2 is an excellent target for native chemical ligation: it has a naturally occurring Cys50 at the junction between the TM and cytoplasmic domains; the TM domain, while hydrophobic, is synthetically more accessible than many other membrane-spanning helices; and the cytoplasmic domain is relatively polar and amenable to synthesis. We obtained synthesis and purification yields of ~10% and ~30% for the two segments and a ligation yield of 14%. These values correspond to a total protein yield of ~1.4%. For comparison, a 25-residue hydrophobic membrane peptide typically shows synthesis and purification yields of ~5%. If the same efficiency is assumed for a 50-residue peptide, then the total yield would drop to ~0.2%. Thus, the protein yield from native chemical ligation is not inferior to, and may be comparable to or even better than, the yield of direct synthesis of hydrophobic peptides without ligation. Solubilization of the hydrophobic transmembrane segment, which is usually the limiting factor in chemical ligation, was achieved here using a TFE solution.

The solid-state NMR spectra of both DMPC- and VM+ membrane-bound M2(22-71) provide clear evidence that the post-AH cytoplasmic tail up to residue 71 is unstructured, exposed to the aqueous environment on the membrane surface, and highly mobile. These results significantly validate the previous conformational analysis of the full-length protein (35). The large-amplitude dynamics of the post-AH cytoplasmic segment, with backbone Cα-Hα order parameters of ~0.5, suggests that this segment may only interact weakly with the membrane surface. Comparative studies of the amphipathic helix domain would be interesting to test this hypothesis. The random coil conformation and dynamics of the post-AH ten-residue segment may be functionally relevant, as they may allow this domain to bind to other influenza virus proteins. For example, M2 binding to M1 is implicated in the regulation of virion morphology. The M1 binding site has been variously suggested to lie within residues 45-69 or residues 70-79 of M2 (25; 51). Since M1 is entirely helical (55; 56), if the binding site lies in the post-AH ten-residue segment, then coil-to-helix conformational changes may occur to enable M1 binding. Structural studies of M2 in the absence and presence of M1 will be useful for answering this
question. For such studies, it will be important to extend the cytoplasmic segment towards the C-terminus. The high synthesis yield of M2(50-71) suggests that this should be feasible, since C-terminal extension should not affect the solubility of this polar domain nor the spatial accessibility of the ligation site at Cys50 and the thioester-linked Lys49. In principle, it is possible that the last 26 residues of M2 may contain small regions of hydrogen-bonded secondary structure elements. Total synthesis of the entire cytoplasmic domain will be important for providing more complete structural information of this domain. Analogous chemical ligation approaches, whether by total synthesis or by semi-synthesis involving protein splicing, are expected to be useful for site-resolved structural studies of functionally important residues in membrane proteins as well as complementing 3D structure determination based on uniformly $^{13}$C-labeled proteins.

**Materials and Methods**

**Solid-Phase Peptide Synthesis of M2(22-49)-thioester and M2(50-71)**

Peptides corresponding to residues M2(22-49) and M2(50-71) of the influenza A/Udorn/72 strain of M2 were synthesized by Fmoc SPPS. M2(22-49)-thioester contained $^{13}$C, $^{15}$N-labeled A30, G34, and I35 while M2(50-71) contained $^{13}$C, $^{15}$N-labeled S64, T65, E66, and V68 (Fig. 1a).

M2(22-49) was synthesized on a Lys-preloaded thioester-generating resin, H-Lys(Boc)-sulfamylbutyryl NovaSyn® TG resin. 0.12 mmol resin (0.6 g at 0.2 mmol/g loading size) was swelled for 1 hour in 2 mL N,N-dimethylformamide (DMF). Stepwise addition of Fmoc-protected amino acids (0.48 mmol, 4 eq.) was carried out with double coupling using 1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate (HATU) (0.48 mmol, 4 eq.) and N,N-diisopropylethylamine (DIPEA) (0.96 mmol, 8 eq.) in 2 mL of DMF for 0.5 – 1 h each. Fmoc cleavage was performed twice with 20% piperidine in DMF (2 mL) for 5 min and 15 min. After the last Fmoc cleavage step, the α-nitrogen of Ser22 was Boc-protected by dissolving the resin-bound peptide in 0.5 mL DMF with DIPEA (1.2 mmol, 10 eq.), then adding di-tert-butyl dicarbonate (0.6 mmol, 5 eq.) in DMF (1 mL) to the solution and stirring for 2 h. After Boc protection, the resin-bound M2(22-49) was methylated at the sulfamyl group to ensure that the thiolate acts as a nucleophile rather than a base, which would deactivate the resin towards cleavage. The resin-bound peptide was swelled for 2 h in 2 mL tetrahydrofuran (THF), then a 4 mL solution (1 : 1 v/v THF : hexane) of 1 M (trimethylsilyl)-diazomethane was added and stirred for 2 h at room temperature. Thioester formation and peptide cleavage were induced by adding sodium thiophenolate (0.336 mmol, 2.8 eq.) and 100-fold excess methyl 3-mercaptopropionate (33.6 mmol, 280 eq.) (57). The solution was stirred at room temperature for 24 h, then filtered to remove the cleaved resin. The filtrate containing the unbound peptide was collected and dried using a rotary evaporator.

Sidechain deprotection of M2(22-49)-thioester was carried out by dissolving the peptide in 3 mL of a 88 : 5 : 2 : 5 (by volume) mixture of trifluoroacetic acid (TFA), water, triisopropylsilane (TIPS) and phenol and shaking for 2 h. The peptide was precipitated and triturated three times with cold diethyl ether. The crude peptide was dissolved in a 30% acetonitrile solution containing 0.1% TFA, then purified by preparative reversed-phase (RP) HPLC on a Varian ProStar 210 System using a Vyae C18 column (10 μm particle size, 2.2 cm x 25 cm) with a linear gradient of 80-99% acetonitrile over 60 min at a flow rate of 10 mL/min.
(Fig. 2a). The peaks were assigned by electrospray ionization mass spectrometry (ESI-MS) (Fig. 2d). The observed mass of 3264.9 ± 0.1 Da was in good agreement with the calculated mass of 3265.0 Da. The synthesis and purification yield of M2(22-49)-thioester was ~10%.

M2(50-71) was synthesized on a rink amide MBHA resin. 0.05 mmol resin (0.063 g with a 0.79 mmol/g loading size) was swelled for 2 h in 2 mL DMF and activated by cleaving the terminal Fmoc group using 2 mL of 20% piperidine in DMF. After the last coupling step, the peptide was deprotected and cleaved from the resin by addition of a TFA/TIPS/H2O solution (95 : 2.5 : 2.5 by volume) for 2 h. The resin was filtered off, and the crude peptide was precipitated and triturated three times with cold diethyl ether and dissolved in 30% acetonitrile solution containing 0.1% TFA. Crude peptide was purified by preparative RP-HPLC using a Vydac C18 column with a linear gradient of 20-43% acetonitrile over 60 min at a flow rate of 10 mL/min (Fig. 2b). ESI-MS analysis showed a mass of 2529.3 ± 0.1 Da, in good agreement with the calculated mass of 2528.9 Da (Fig. 2e). The total synthesis and purification yield of M2(50-71) was ~30%.

Native Chemical Ligation of M2(22-71)

M2(22-71) was ligated by adding 50% molar excess of 45.2 mM M2(50-71) in 0.1 mL of 60% TFE (v/v) solution to 0.72 mM M2(22-49)-thioester in 4.2 mL of ligation buffer. The ligation buffer consists of 60% TFE (v/v), 5.4 M urea, 67 mM sodium phosphate, 50 mM 4-mercapto-phenylacetic acid (MPAA) and 20 mM Tris(2-carboxyethyl)phosphine (TCEP) hydrochloride. The pH 4.5 solution was stirred at 35°C, then the reaction was initiated by increasing the pH to 7.0 with 4 M NaOH. The reaction was monitored by analytical RP-HPLC using a Vydac C4 column (5 μm particle size, 0.46 cm x 15 cm) with a 40-99% acetonitrile gradient over 30 min. Unreacted M2(22-49)-thioester and M2(50-71) peptides appeared at 24.3 min and 6 min, respectively, while the ligation product appeared at 20.5 min (Fig. 2c). The ligation was complete after 18 hours. M2(22-71) was purified by preparative RP-HPLC using a linear gradient of 70-99% acetonitrile over 60 min. ESI-MS showed a mass of 5675.2 ± 0.2 Da, in good agreement with the calculated mass of 5675.7 Da (Fig. 2f). About 2.5 mg of purified ligation product was obtained from starting materials of ~10 mg each of the two peptides. The ligation yield, calculated as the ratio of the number of moles of the product with the number of moles of the M2(22-49)-thioester, which is the limiting reagent, is ~14%.

Membrane reconstitution of M2(22-71) and M2(50-71)

Three membrane samples were prepared. M2(22-71) was reconstituted into the DMPC and VM+ membranes using an organic solution mixing protocol. 4.3 mg M2(22-71) in 200 μL TFE was mixed with 12.9 mg DMPC in 500 μL chloroform, while 2 mg of M2(22-71) was mixed with 4 mg VM+ lipid mixture. These correspond to peptide : lipid molar ratios (P : L) of 1 : 25 for the DMPC sample and 1 : 18 for the VM+ sample. The organic solvents were removed by a stream of nitrogen gas and the mixture was lyophilized overnight. The protein-lipid mixture was suspended in a pH 7.5 Tris buffer (10 mM Tris base, 1 mM ethylenediaminetetraacetic acid, and 0.1 mM sodium azide) and centrifuged at 55,000 rpm at 4°C for 4 hours to obtain a homogeneous membrane pellet. The pellet was equilibrated to a hydration level of ~40% by mass, packed into a 3.2 mm MAS rotor and stored at -30°C until the NMR experiments. The VM+ membrane consists of an equimolar mixture of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE), egg
sphingomyelin and cholesterol. The presence of cholesterol and sphingomyelin immobilizes the whole-body uniaxial rotational diffusion of the M2 transmembrane peptide so that protein internal dynamics can be investigated at physiological temperature (50; 58). 5 mg M2(50-71) was reconstituted into 20 mg of a POPC/POPG (3:1) membrane (P : L = 1 : 13). This negatively charged membrane was chosen to maximize the electrostatic attraction with M2(50-71).

**Solid-state NMR experiments**

Most solid-state NMR spectra were measured on a 900 MHz (21.1 Tesla) Bruker NMR spectrometer using a 3.2 mm $^{1}$H/$^{13}$C/$^{15}$N MAS probe. Typical radiofrequency field strengths were 71.4 kHz and 62.5 kHz for $^{1}$H and $^{13}$C, respectively. Spectra were measured in both gel-phase and liquid-crystalline membranes to investigate the conformation and dynamics of the protein. $^{13}$C chemical shifts were externally referenced to the CH$_2$ signal of adamantane at 38.48 ppm on the TMS scale.

2D $^{13}$C-$^{13}$C correlation spectra were measured using a dipolar-assisted rotational resonance (DARR) (59) experiment with a 100 ms mixing time to obtain $^{13}$C chemical shifts. 2D $^{15}$N-$^{13}$C correlation spectra were measured on a 400 MHz (9.4 Tesla) Bruker NMR spectrometer using a REDOR pulse train for polarization transfer (60; 61). Gel-phase $^{13}$C-$^{1}$H heteronuclear correlation (HETCOR) spectra (44) were measured under 7 kHz MAS between 248 and 258 K. The frequency-switched Lee-Goldburg (FSLG) sequence (62) was used to suppress the $^{1}$H-$^{1}$H dipolar coupling during the $^{1}$H evolution period. The $^{1}$H spin diffusion mixing time was 4 ms and 25 ms, and Lee-Goldburg cross polarization (LG-CP) (63-65) was used to transfer the $^{1}$H magnetization to $^{13}$C. $^{13}$C-$^{1}$H dipolar couplings were measured using the 2D dipolar chemical-shift (DIPSHIFT) correlation experiment (66) under 7 kHz MAS at 303 K. The FSLG sequence was used for $^{1}$H homonuclear decoupling during t$_{1}$. The dipolar oscillation was simulated using a Fortran program to obtain the couplings, which were then divided by the FSLG scaling factor of 0.577 to obtain the real couplings. The order parameter was calculated as the ratio of the real couplings to the rigid-limit value of 22.7 kHz.

**Acknowledgement**

This work is funded by NIH grant GM088204 to M. H. The 900 MHz NMR spectrometer at the MIT/Harvard Center for Magnetic Resonance is supported by NIH grant P41-EB-002026. The authors thank Tuo Wang for experimental help.
Table 1. $^{13}$C and $^{15}$N chemical shifts of the labeled residues in membrane-bound M2(22-71) obtained from total chemical synthesis.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Site</th>
<th>Chemical Shifts (ppm) $^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cα</td>
<td>52.6</td>
</tr>
<tr>
<td>A30</td>
<td>Cβ</td>
<td>16.1 (16.7)</td>
</tr>
<tr>
<td></td>
<td>CO</td>
<td>178.1</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>118.6</td>
</tr>
<tr>
<td></td>
<td>Cα</td>
<td>45.4 (44.9)</td>
</tr>
<tr>
<td>G34</td>
<td>CO</td>
<td>Not labeled</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>107.6</td>
</tr>
<tr>
<td></td>
<td>Cα</td>
<td>63.4</td>
</tr>
<tr>
<td></td>
<td>Cβ</td>
<td>35.7</td>
</tr>
<tr>
<td></td>
<td>Cγ1</td>
<td>28.0 (27.0)</td>
</tr>
<tr>
<td>I35</td>
<td>Cγ2</td>
<td>15.6</td>
</tr>
<tr>
<td></td>
<td>Cδ</td>
<td>11.6</td>
</tr>
<tr>
<td></td>
<td>CO</td>
<td>175.3</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>121.7</td>
</tr>
<tr>
<td></td>
<td>Cα</td>
<td>56.2</td>
</tr>
<tr>
<td>S64</td>
<td>Cβ</td>
<td>61.5</td>
</tr>
<tr>
<td></td>
<td>CO</td>
<td>172.9</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cα</td>
<td>60.1</td>
</tr>
<tr>
<td>T65</td>
<td>Cβ</td>
<td>67.8</td>
</tr>
<tr>
<td></td>
<td>Cγ</td>
<td>19.9</td>
</tr>
<tr>
<td></td>
<td>CO</td>
<td>172.9</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Cα</td>
<td>54.8</td>
</tr>
<tr>
<td>E66</td>
<td>Cβ</td>
<td>28.4</td>
</tr>
<tr>
<td></td>
<td>Cγ</td>
<td>33.5</td>
</tr>
<tr>
<td></td>
<td>Cδ</td>
<td>- (180.9)</td>
</tr>
<tr>
<td></td>
<td>CO</td>
<td>174.5</td>
</tr>
<tr>
<td></td>
<td>Cα</td>
<td>57.9</td>
</tr>
<tr>
<td>V68</td>
<td>Cβ</td>
<td>30.8</td>
</tr>
<tr>
<td></td>
<td>Cγ1, γ2</td>
<td>19.6, 18.6</td>
</tr>
<tr>
<td></td>
<td>CO</td>
<td>172.6</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>120.9</td>
</tr>
</tbody>
</table>

$^1$ $^{13}$C chemical shifts are reported on the TMS scale. Values in brackets are those measured in the VM+ membrane that differ by more than 0.5 ppm from the chemical shifts in the DMPC membrane. Chemical shifts of the transmembrane A30, G34 and I35 were obtained from low-temperature 2D correlation spectra whereas those of the cytoplasmic S64, T65, E66 and V68 residues were obtained from high-temperature 2D spectra.
Fig. 1. (a) Amino acid sequence of M2(22-71). The $^{13}$C, $^{15}$N-labeled residues (A30, G34, I35, S64, T65, E66, and V68) are bolded and underlined. The naturally occurring Cys50 provides the ligation site. (b) Native chemical ligation strategy for M2. Synthetic M2(22-49)-thioester and M2(50-71) are coupled together by MPAA-catalyzed transthioesterification and an irreversible amide-bond-forming rearrangement.
Fig. 2. Purification and characterization of (a, d) M2(22-49)-thioester, (b, e) M2(50-71), and (c, f) M2(22-71) by RP-HPLC (a-c) and ESI-MS (d-f). HPLC shows that the ligation product’s concentration increased with time over ~18 hours whereas the M2(22-49)-thioester concentration decreased. The purified product (red trace) eluted as a single peak at ~21 minutes.
Fig. 3. Variable-temperature 1D $^{13}$C MAS spectra of site-specifically labeled M2 peptides in lipid bilayers. (a) POPC/POPG (3:1)-bound M2(50-71). At 298 K, the peptide signals are weak in the CP spectrum but much stronger in the DP spectrum, indicating high mobility. At 253 K, the spectrum is severely broadened, consistent with the unstructured nature of this peptide. (b) DMPC-bound M2(22-71). The low-temperature spectrum shows narrow linewidths for the TM residues but broad linewidths for the cytoplasmic residues. At high temperature, the TM residues give high signals in the CP spectrum but low intensities in the DP spectrum, indicating their relative rigidity. (c) VM+ bound M2(22-71). The chemical shifts and linewidths are similar to the DMPC-bound M2(22-71).
Fig. 4. 2D $^{13}$C-$^{13}$C correlation spectra of various M2 construct at low and high temperatures (T). TM and cytoplasmic residues are assigned in red and blue, respectively. Most spectra were
measured with a spin diffusion mixing time of 100 ms except where otherwise noted. (a) 2D spectrum of POPC/POPG (3 : 1)-bound M2(50-71) measured at 253 K. The labeled residues show random coil chemical shifts. (b) 2D spectrum of DMPC-bound M2(22-71) measured between 245 and 263 K. A30, G34, and I35 exhibit α-helical chemical shifts whereas S64, T65 and V68 show random-coil chemical shifts that are consistent with those in M2(50-71). (c) 2D DP-DARR spectrum of DMPC-bound M2(22-71) at 298 K. The cytoplasmic residues show much sharper signals than at low temperature. (d) 2D spectrum of DMPC-bound full-length M2 with 15 ms mixing, measured between 253 K and 273 K (35). (e) 2D spectrum of VM+ bound M2(22-71) at 253 K. (f) 2D DP-DARR spectrum of VM+ bound M2(22-71) at 298 K. The VM+ spectra are similar to the DMPC spectra at similar temperatures.
Figure 5. 2D $^{15}$N-$^{13}$C correlation spectra of (a) DMPC-bound M2(22-71) and (b) DLPC-bound M2(22-46) (40). The similarity of the $^{15}$N and $^{13}$C chemical shifts for the common TM residues G34 and I35 indicates that the TM domain adopts the same conformation in the ligated protein and in the isolated TM peptide.
**Fig. 6.** Depth of insertion of DMPC-bound M2(22-71) by gel-phase 2D $^1$H-$^{13}$C correlation. (a) 2D spectrum with 4 ms $^1$H spin diffusion, measured at 258 K. (b) $^{13}$C cross sections at the lipid CH$_2$ (black) and water (blue) $^1$H chemical shifts. A30$\alpha$ and G34$\alpha$ exhibit higher lipid/water intensity ratios than S64$\alpha$ and V68$\alpha$. (c) $^1$H cross section with 4 ms mixing. The 1.4-ppm peak of lipid CH$_2$ is resolved from the A30 and V68 sidechain $^1$H signals. The higher water and headgroup H$_\gamma$ cross peaks with S64$\alpha$ and V68$\alpha$ (blue) compared to those with A30$\alpha$ and G34$\alpha$ (red) indicates that the cytoplasmic residues are in closer contact with the membrane-surface water than the TM residues. The lipid CH$_2$ (black) cross section has the lowest water cross peak intensity, as expected. (d) $^1$H cross sections measured with 25 ms $^1$H spin diffusion. The water cross peaks of the TM and cytoplastic residues have equilibrated by this time.
**Fig. 7.** $^{13}$C-$^1$H dipolar couplings of DMPC-bound M2(22-71) at 303 K from the DIPSHIFT experiment. Solid lines are best-fit simulations without $T_2$ decays. (a) $^{13}$C dimensions of the 2D spectrum. (b) Sum of the A30 and I35 Cα cross sections. (c) G34 Cα cross section. The order parameters are close to 1, indicating that the TM domain is immobilized in the membrane at 303 K. (d) Sum of the S64, E66 and V68 Cα cross sections. (e) S64 Cβ cross section. Low C-H order parameters of 0.50 and 0.44 are obtained, indicating that the cytoplasmic residues undergo large-amplitude motion.
References


