# Supporting Information

## Augustyniak and Kay 10.1073/pnas.1721811115

### SI Materials and Methods

**Protein Expression and Purification.** All variants of the *T. acidophilum* VAT protein, including the truncated ΔN construct (residues 183–745), have been obtained by QuikChange site-directed mutagenesis of a codon-optimized gene encoding full-length VAT and inserted into the pProEx expression vector. All constructs included an N-terminal $His_6$-tag followed by a TEV cleavage site (sequence I of Fig. S1*B*). All ILMV-residues in the tag, with the exception of the N-terminal Met, were mutated to Ala to minimize spectral overlap. The proteins were expressed and purified according to a previously described protocol (1). Pure proteins were used for degradation assays or were reacted with Ellman's reagent (Sigma) for chemical cross-linking with substrates as described below. The *T. acidophilum* proteasome CP (20S CP) was expressed and purified as described previously (2).

The GB1 domain of protein G (GB1) was expressed as a His-tagged protein and the expression was carried out in *Escherichia coli* BL21(DE3) cells for 3 h at 37 °C (overnight at 16 °C for ssrA-tagged protein) after induction with 1 mM isopropyl β-D-1-thiogalactopyranoside. Purification was achieved via chromatography with His-trap affinity and HiLoad 16/60 Superdex 75 gel filtration (GE Healthcare) columns, using the same buffers as for purification of VAT. CaM was expressed and purified according to a protocol adapted from ref. 3, where the last purification step involved size-exclusion chromatography as for the GB1 domain. Stabilities of CaM cysteine mutants were followed by thermal shift assays using SYPRO Orange (Life Technologies and ref. 4).

For preparation of NMR samples cells were grown in M9 $D_2O$ media supplemented with $^{14}NH_4Cl$ and [$^2$H,$^{12}$C]glucose as the sole nitrogen and carbon sources, respectively. In the case of GB1 and CaM, protein was produced as Ile-δ1-[$^{13}CH_3$], Leu/Val-[$^{13}CH_3$,$^{12}CD_3$], Met-ε-[$^{13}CH_3$] (ILVM labeling) (5–7). An ILVM-labeled sample of VAT was also prepared for diffusion measurements.

**Degradation and Cross-Linking.** Proteins expressed and purified as described above were used for degradation assays and cross-linking experiments. Substrate degradation was performed at 45 °C by mixing purified 20S CP proteasome from *T. acidophilum* (2 μM) with VAT (1 μM; Fig. 2*A*) and substrate (15 μM) in a buffer containing 50 mM Hepes, 200 mM NaCl, 20 mM $MgCl_2$ (120 mM for the assays with fl VAT), and 5 mM ATP. ATP levels were maintained throughout the duration of the assay by the addition of a phosphoenolpyruvate/pyruvate kinase ATP regenerating system (20 U·mL$^{-1}$ pyruvate kinase and 15 mM phosphoenolpyruvate). The reaction was allowed to proceed for 15 min (or 30 min when less active fl VAT was used), with aliquots taken every 5 min and diluted immediately with SDS/PAGE loading buffer for gel analysis.

Substrate cross-linking for analyses as in Fig. 2 *B* and *C* and to produce stable samples for NMR experiments was achieved in the following way. Variants of VAT (C77A/C679A) (50–100 μM in protomer) containing mutations in one of the pore loops (W541C for gel and NMR studies and Y264C and E300C for gel studies) were initially unfolded into isolated protomers by the addition of 6 M guanidinium hydrochloride (GdnHCl) in a buffer containing Hepes (50 mM), NaCl (200 mM), and Tris(2-carboxyethyl)phosphine (TCEP) (1 mM), pH 7.5. Subsequently, this solution was titrated into a second solution of Ellman's reagent (20 mM) dissolved in the same Hepes/NaCl/GdnHCl buffer. The reacted VAT was then refolded by fast dilution into refolding buffer containing 0.5 M arginine, 50 mM Hepes, and 200 mM NaCl (pH 7.5) and purified on an S200 column. Protein was frozen and stored at −80 °C. The substrate cross-linking reaction was performed by mixing 15-fold excess (relative to the concentration of VAT protomers) of substrate with unfoldase in a buffer containing 100 mM ATP (no regeneration system), 50 mM Hepes, 200 mM NaCl, and 20 mM $MgCl_2$ (120 mM for the assays with the fl VAT), pH 7.5. NMR samples were further purified on an S200 column without nucleotide. Protein was subsequently buffer exchanged from 18 mL to 0.25 mL into 25 mM imidazole (pH 6.8 uncorrected), 100 mM KCl, 1.5 mM EGTA, and 1.5 mM EDTA in 99.9% $D_2O$ three times to eliminate residual nucleotide and $Ca^{2+}$/$Mg^{2+}$. Subsequently 2 mM ADP was added. Before analyzing products with SDS/PAGE 300 mM maleimide was added.

As described in the main text, the cross-linking approach detailed above produces samples with approximately three substrate molecules attached to each ΔN VAT particle (Fig. 2*C* and Fig. S6*A*). In an effort to evaluate whether the conclusions of our study depend on the number of linked substrates we have generated samples with only approximately one single copy of ILVM-labeled substrate attached (Fig. S6*A*). This was achieved by mixing perdeuterated WT and Ellman's reagent-activated W541C ΔN VAT molecules in an 11:1 ratio in the presence of 6 M guanidine hydrochloride; in this manner 59% of the VAT hexamers will be 100% WT, 32% will contain a single W541C protomer, and 7.4% will contain two W541 protomers. Reconstitution of hexameric rings as well as reaction with substrate followed the previously described protocol used for preparation of VAT hexamers bearing six cysteine mutations per ring. The final purification step used a Ni column, exploiting the fact that substrate but not ΔN VAT contained a hexa-his tag, allowing separation of substrate-containing VAT from hexamers with no substrate attached. The majority of signal observed in spectra recorded of samples prepared in this manner derives from single-substrate rings (~66%) with 30% of the signal derived from VAT molecules with two substrates.

**Additional NMR Details.** All NMR experiments were carried out using an 18.8 T (800 MHz) Varian Inova spectrometer with a room-temperature triple resonance probe, with the exception of studies of VAT conjugated with a single substrate that were performed on a Bruker Avance III 18.8 T spectrometer with a z-gradient cryo-probe. NMRPipe (8) and CCPN analysis (9) software were used to process and visualize NMR spectra, respectively.

***Diffusion measurements.*** Translational diffusion coefficients for ILVM-labeled substrates either free or cross-linked to VAT were measured by recording a series of 1D $^{13}$C-edited spectra at 25 °C, using a pulse scheme that is similar to an $^{15}$N-edited experiment published previously with $^{15}$N and $^{13}$C pulses interchanged (10). A 200-ms diffusion delay and encoding/decoding gradient strengths ranging from 5 to 54 G/cm were used. The resulting $^1$H signal was integrated over the methyl $^1$H range of the spectra to obtain intensities as a function of encoding/decoding gradient strength. Diffusion constants were extracted by nonlinear fitting of NMR signal intensities as a function of the square of applied gradient pulses using the relation $I = I_0\exp(-AG^2D)$, where $I$ and $I_0$ are peak intensities in the presence and absence of gradient $G$,

respectively, $D$ is the diffusion coefficient, and $A$ is a constant that depends on experimental conditions.

***Measurement of methyl $R^s_{2,H}$ rates.*** Residue-specific relaxation rates of the slowly relaxing methyl $^1$H single-quantum transitions were measured for both free and cross-linked samples of L4C CaM using the pulse scheme of Tugarinov and Kay (11). A set of 2D $^{13}$C-$^1$H planes was recorded with relaxation delays ($T$) of 1, 3, 5, 7, 10, 13, 16, and 20 ms for the cross-linked sample and 1, 5, 10, 15, 20, 30, 40, 50, 60, 70, and 80 ms for free CaM. $R^s_{2,H}$ rates were obtained from exponential fits of peak intensities, $I$, as a function of relaxation delay, using the equation $I = I_0\exp(-R^s_{2,H}T)$.

## SI Text

**NMR Snapshots of Substrate Are Averages over Different Attachment Positions with only Small Variations Expected for Different Points of Attachment.** A surface representation of the ADP-bound form of ΔN VAT that is used in the cross-linking studies described in this work is illustrated in Fig. 1*C*. Unlike the ATP state of the enzyme where the D1 and D2 rings are stacked (Fig. 1*B*), a split-ring structure prevails for the ADP conformation, placing the pore-1 loops of D2 that form the points of attachment of substrate (discussed in the main text) in a helical configuration. As a result, regions of substrate either inside or extending on either side of the barrel of the unfoldase differ depending on which protomer serves as the substrate attachment site. For example, consider substrate that is threaded from the N terminus by VAT. If the attachment is to protomer 4 then effectively all substrate residues N-terminal to the point of attachment are external to the lumen (Fig. 1*E* distance to the bottom = 0 Å). In contrast, if the substrate is attached to protomer 1 then a distance of 16 Å must be traversed before residues are released from the bottom of the lumen. Fig. 1*E* tabulates the distances from the top and bottom of VAT, as defined in the text, for attachment at the pore-1 D2 loops of each protomer. Thus, the NMR snapshots obtained are of necessity averages over potentially different substrate conformations, reflecting their attachment to protomers 1–6 (Fig. 1) and the fact that the structure of VAT-ADP is not symmetric. The snapshots represent what the substrate "looks like" as it passes pore loop 1 of D2, recognizing that the pore loops are at different positions in the structure depending on the protomer and hence potentially six different substrate conformations are possible assuming simplistically only a single conformer for substrate as it passes a given pore loop. With this in mind, it is of interest to calculate what the expected variability in the region of substrate sequestered in VAT might be in our cross-linking experiments to ascertain how different the structural features might be expected to be depending on which VAT protomer is involved. In what follows we assume that substrate threading is from the N to the C terminus and that when inside the lumen of the unfoldase the substrate forms an extended conformation, as was observed in a previous cryo-EM study where one VAT molecule served as a substrate for a second molecule acting as an unfoldase (12). Thus, the distance between successive residues that are localized in the lumen is ∼3.5 Å and in Fig. 1*E* the "distance from the top" ranges from ∼3–11 residues, while the "distance from the bottom" lies between 5 and 0 residues. In general, for a substrate attached at residue j, the entry(exit) point to(from) VAT is residue j + 3(j − 5) for protomer 1 and residue j + 11(j) for protomer 6. There are, thus, a maximum difference of 8 (= 11 − 3) and 5 (= 5 − 0) residues from the top and bottom, respectively, depending on which protomer is involved in the attachment. Between 8 and 11 residues are sequestered within the lumen of VAT depending on the attachment site and when considering all attachment sites in total 16 residues can be sequestered. In what follows we summarize the secondary structural elements of CaM that are at least partially sequestered within the lumen of VAT, focusing on the extreme cases where cross-linking involves protomers 1 or 6.

| CaM | Protomer of VAT | Secondary structure elements sequestered |
|---|---|---|
| L4C | 1 | H1 |
| L4C | 6 | H1 |
| A15C | 1 | H1 |
| A15C | 6 | Loop between H1/H2 to C terminus of H1 |
| I52C | 1 | H3 |
| I52C | 6 | Loop between H3/H4 to C terminus of H3 |
| M71C | 1 | H4 |
| M71C | 6 | N terminus of H5 to C terminus of H4 |
| A73C | 1 | H4 |
| A73C | 6 | N terminus of H5 to C terminus of H4 |
| D80C | 1 | N terminus of H5 to loop between H4/H5 |
| D80C | 6 | H5 |
| F89C | 1 | H5 |
| F89C | 6 | Loop between H5/H6 to C terminus of H5 |
| V121C | 1 | H7 to N terminus of loop between H6/H7 |
| V121C | 6 | Loop between H7/H8 to H7 |
| M144C | 1 | H8 |
| M144C | 6 | H8 |

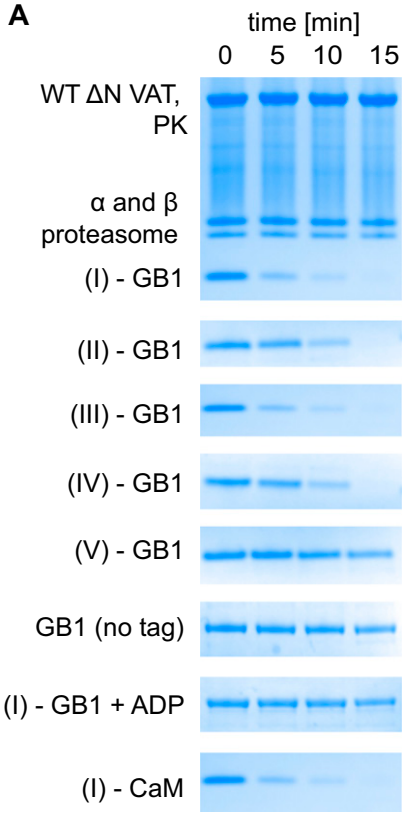
While the residues sequestered in the lumen do vary depending on the protomer, in general the secondary elements that are involved do not. Thus, the structural features of a given substrate as it passes pore loop 1 of D2 are expected to be similar irrespective of the protomer involved and the structural models as reported by the NMR-based cross-linking studies in this work are likely to be reflective of attachment to any of the VAT protomers. This, of course, does not mean that there are no differences in substrate structure between attachment points and it may be the case that a given NMR experiment is more sensitive to one attachment position over another because of differences in substrate dynamics, for example, at the two positions.

**Unfolding of Substrate as It Traverses the Lumen of VAT.** We have shown that upon cross-linking to position 4 all N-lobe peaks of CaM disappear and concomitantly additional peaks appear in the random coil regions of the $^{13}$C-$^1$H HMQC spectrum. Fig. S3*C* shows a superposition of spectral regions from ILVM-labeled samples of L4C CaM cross-linked to VAT (orange) and a fragment of the intact protein comprising residues 1–76 that includes only the NTD that has been denatured through the addition of 8 M urea (teal, single contour). Notably, the spectra of the N-domain of the cross-linked sample and the urea unfolded NTD are very similar, strongly supporting the notion that the N-domain of CaM cooperatively unfolds upon cross-linking to Cys4. Cross-peaks for the C-domain remain in the spectrum of the cross-linked protein, indicating that the C-domain is intact. In a similar manner, the spectrum of ILVM-labeled F89C CaM cross-linked to VAT (red) shows a complete disappearance of C-lobe resonances and a concomitant large increase in the intensities of unfolded peaks. Notably, the unfolded peaks coincide with the correlations obtained in a spectrum recorded of an 8 M urea unfolded CaM C-domain fragment (residues 77–148; purple, single contour), establishing that cross-linking at position 89 leads to unfolding of the CaM C-domain. The N-domain, however, remains folded, as all peaks derived from it are present in the HMQC spectrum.

In the general case, however, the absence of some of the cross-peaks that would be expected for a particular region of a protein does not mean that the corresponding residues are in unfolded conformations. Indeed, as discussed in the main text, the intensities of NMR correlations are sensitive to conformational exchange processes that can lead to severe peak broadening to the point where the peak is not observed, especially in the context of a 400-kDa complex where cross-peaks are relatively weak to begin with. For cross-linking at positions 4 and 89, discussed above, however, it is unlikely that exchange exclusively is responsible for peak disappearance, as it would then be expected that residues would be differentially affected in a manner that depends on chemical shift differences between exchanging states so that not all N (attachment at position 4) or C (attachment at position 89) domain peaks would be eliminated. It is also very unlikely that the disappearance of peaks is the result of slow tumbling resulting from attachment of CaM to a large complex. First, high-quality methyl-TROSY datasets have been recorded of VAT previously, and second, although all "folded" N-domain peaks disappear in the L4C complex, attachment at the corresponding position in the C-domain, D80C, eliminates only a fraction of the "folded" C-domain resonances (Fig. 5).

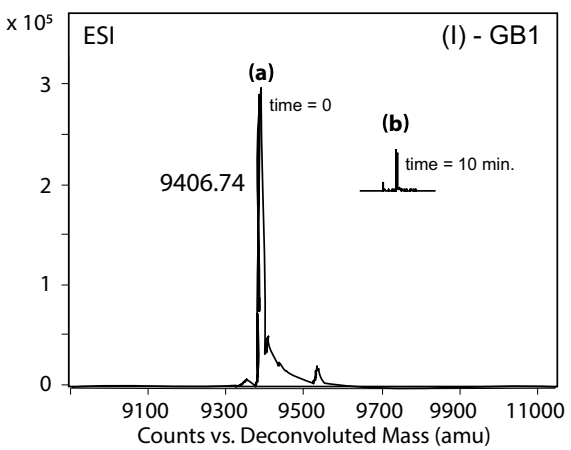
In contrast, the fact that only some cross-peaks are observed in their folded locations for the N-domain of the A73C complex or the C-domain of the D80C complex does suggest that multiple conformers are interconverting that significantly attenuates correlations, with some, but not all, disappearing completely. Of course, in addition there is at least some degree of unfolding as the attachment positions (A73C or D80C) overlap with regions that would normally assume helical structure and these must, at least locally, be unfolded. We indicate this flexibility by denoting secondary structural elements using "wavy" representations in Fig. 6. Thus, to summarize, the absence of at least some cross-peaks does establish that the protein is sampling one or more conformations that are distinct from the folded state, irrespective of what these conformations might be, and in general we have interpreted our NMR data in this context. Finally, it is worth mentioning that spectra may preferentially report on substrate attached to one VAT protomer over the others, due to potential differences in dynamics, or the fact that for one VAT protomer there are larger regions of substrate that reside outside of the lumen, for example, that leads to improved signal.


1. Huang R, et al. (2016) Unfolding the mechanism of the AAA+ unfoldase VAT by a combined cryo-EM, solution NMR study. *Proc Natl Acad Sci USA* 113:E4190–E4199.
2. Velyvis A, Ruschak AM, Kay LE (2012) An economical method for production of (2)H, (13)CH3-threonine for solution NMR studies of large protein complexes: Application to the 670 kDa proteasome. *PLoS One* 7:e43725.
3. Latham MP, Kay LE (2012) Is buffer a good proxy for a crowded cell-like environment? A comparative NMR study of calmodulin side-chain dynamics in buffer and E. coli lysate. *PLoS One* 7:e48226.
4. Lo M-C, et al. (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal Biochem* 332:153–159.
5. Tugarinov V, Kanelis V, Kay LE (2006) Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat Protoc* 1:749–754.
6. Tugarinov V, Kay LE (2004) An isotope labeling strategy for methyl TROSY spectroscopy. *J Biomol NMR* 28:165–172.
7. Gelis I, et al. (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. *Cell* 131:756–769.
8. Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293.
9. Vranken WF, et al. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* 59:687–696.
10. Choy W-Y, et al. (2002) Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. *J Mol Biol* 316:101–112.
11. Tugarinov V, Kay LE (2006) Relaxation rates of degenerate 1H transitions in methyl groups of proteins as reporters of side-chain dynamics. *J Am Chem Soc* 128: 7299–7308.
12. Ripstein ZA, Huang R, Augustyniak R, Kay LE, Rubinstein JL (2017) Structure of a AAA+ unfoldase in the process of unfolding substrate. *eLife* 6:e25754.




**B**

| Tag sequence/substrate | Relative degradation rate | $R^2$ |
|---|---|---|
| **(I)** MKHHHHHHPMSDYDIPTTENLYFQ- (24 residues, N-term.) | 1 | 0.89 |
| **(II)** -QFYLNETTPIDYDSMPHHHHHHHKM (24 residues, C-term.) | 0.91 | 0.98 |
| **(III)** MKIIIIIIPMSDYDIPTTENLYFQ- (24 residues, N-term.) | 0.62 | 0.84 |
| **(IV)** MKEEEEEEPMSDYDIPTTENLYFQ- (24 residues, N-term.) | 0.62 | 0.98 |
| **(V)** MHHHHHHG- (8 residues, N-term.) | 0.13 | 0.96 |
| GB1 (no tag) | 0.02 | 0.85 |
| (I) - GB1 + ADP | 0.01 | 0.89 |
| (I) - CaM | 0.96 | 0.98 |



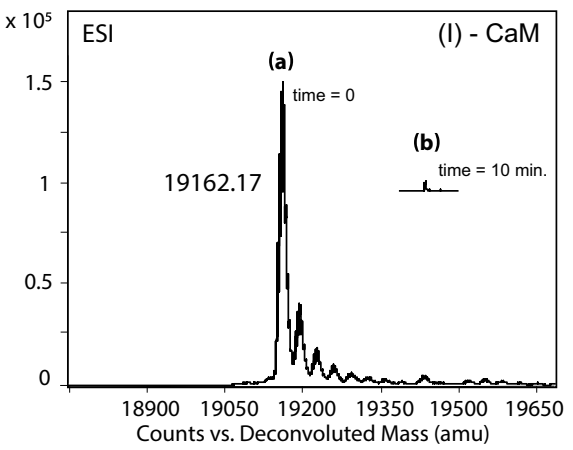




**Fig. S1.** Effect of different tags on substrate unfolding. (*A*) SDS gels showing ATP-dependent degradation assays using the ΔN version of VAT, along with a pyruvate kinase ATP regeneration system. A number of different variants of GB1 have been assayed that include different amino acid sequences for the flexible tags and different tag lengths used to promote degradation [*B*; (x)-GB1 refers to GB1 with tag x, as defined]. Relative degradation rates for each construct were obtained by measuring SDS/PAGE band intensities using ImageJ software (1). The squares of the correlation coefficient ($R^2$) from the linear fits of the degradation data are listed. Controls were performed where GB1 without a flexible tail was used as substrate and where ATP was replaced by ADP. CaM exhibits behavior similar to that of GB1 and can be efficiently degraded by the proteasome. An ATP regeneration system was used involving PK (pyruvate kinase), as described in the text. ΔN VAT and PK migrate similarly on the gel; α and β subunits of the 20S CP proteasome used for degradation migrate separately, as indicated. Mass spectrometry (electrospray ionization) confirms that substrates GB1 (*C*) and CaM (*D*) are degraded upon addition of ATP, PK, ΔN VAT, and the proteasome. Spectra (a) in *C* and *D* were recorded on the reaction mixture immediately before the addition of ATP and spectra (b) were obtained 10 min after addition of ATP, plotted on the same scale as for (a). After 30 min substrates were not detected in the reaction mixture.

1. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9:671–675.

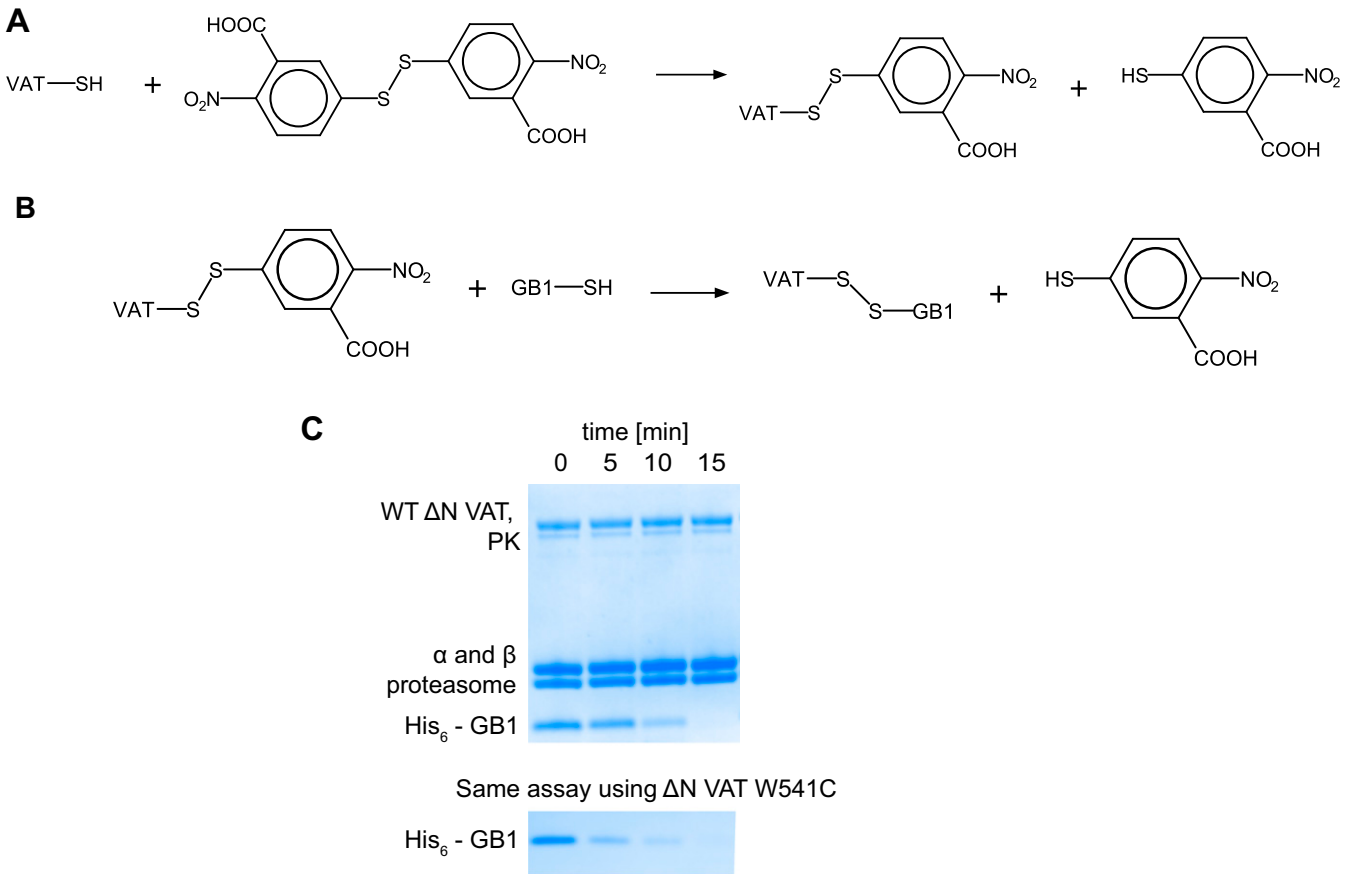



**Fig. S2.** The VAT W541C mutant has unfolding properties similar to those of the WT unfoldase. (*A*) Reaction scheme for the activation of W541C VAT with Ellman's reagent (1) followed by cross-linking with substrate (*B*, GB1 in this example). (*C*) SDS gels showing similar unfoldase activities for WT ΔN VAT and ΔN VAT where Cys residues at position 541 have been modified by Ellman's reagent (ΔN VAT W541C). Assays were carried out in the presence of ATP and an ATP regeneration system involving pyruvate kinase, PK.

1. Ellman GL (1959) Tissue sulfhydryl groups. *Arch Biochem Biophys* 82:70–77.

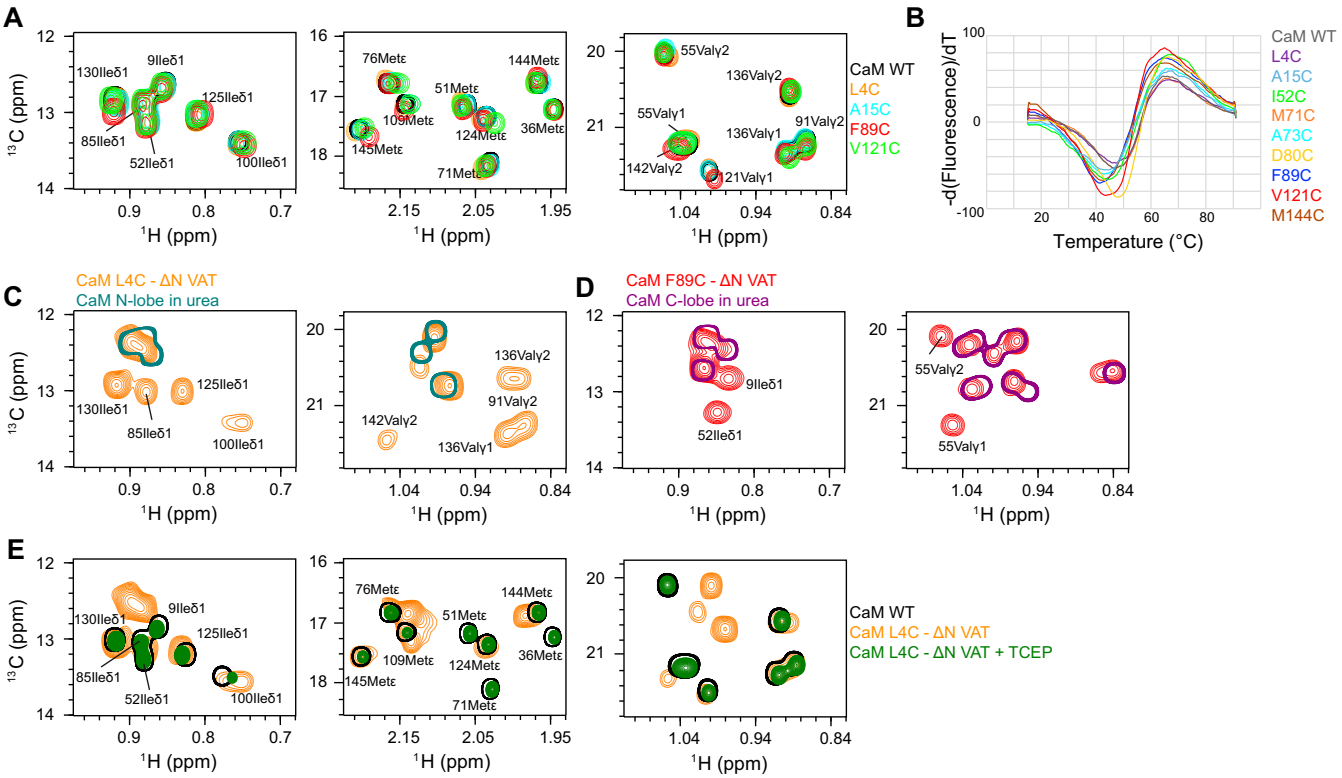



**Fig. S3.** CaM structure and stability are little affected by the mutations considered. (*A*) The similarity of HMQC spectra recorded of WT CaM along with a number of mutants, as listed (far right side of *A*), establishes that cysteine mutations do not affect the structure of CaM. Selected regions of spectra recorded for representative constructs of CaM are shown. (*B*) Thermal stabilities of WT and cysteine mutations of CaM are similar (see minima in curves that define the melting temperature, $T_m$), as established by protein thermal shift assays (Life Technologies). The least stable construct (F89C) has a 4 °C lower $T_m$ than WT CaM. (*C*) Superposition of Ile and Val regions of $^{13}$C-$^{1}$H HMQC spectra of the isolated N-lobe of CaM (residues 1–76) denatured with 8 M urea (teal, single contour) and ILVM-labeled His-tagged CaM L4C cross-linked to perdeuterated ΔN VAT W541C (orange, multiple contours). (*D*) Superposition of Ile and Val regions of $^{13}$C-$^{1}$H HMQC spectra of the isolated C-lobe of CaM (residues 77–148) denatured with 8 M urea (purple, single contour) and ILVM-labeled His-tagged CaM F89C cross-linked to perdeuterated ΔN VAT W541C (red, multiple contours). (*E*) Disulphide cross-linked complexes dissociate and CaM refolds upon addition of the reducing agent (1 mM TCEP) as shown for CaM L4C – VAT.

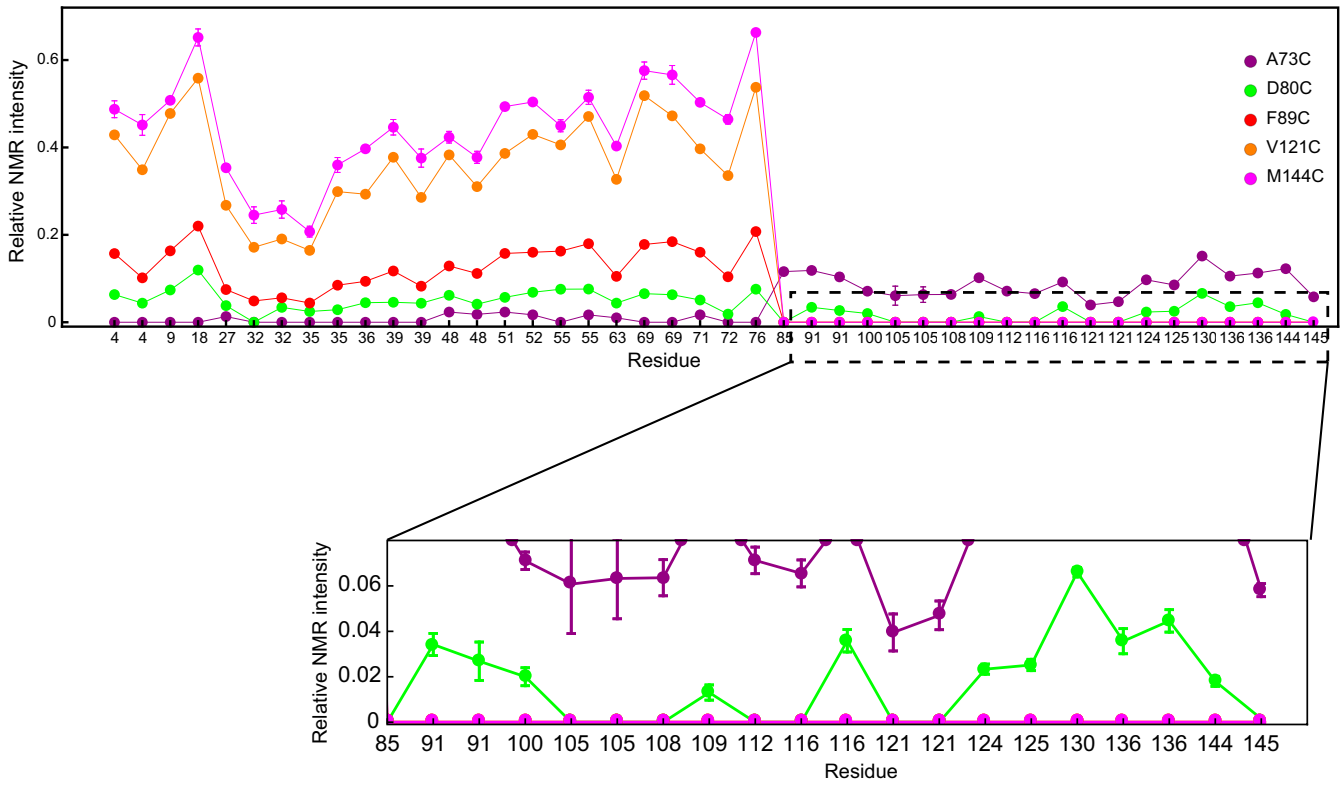



**Fig. S4.** Relative cross-peak intensities from $^{13}C$-$^{1}H$ HMQC spectra of ILVM-labeled CaM cross-linked to VAT at different positions as a function of residue number. A sample of unbound WT CaM was used as a reference.

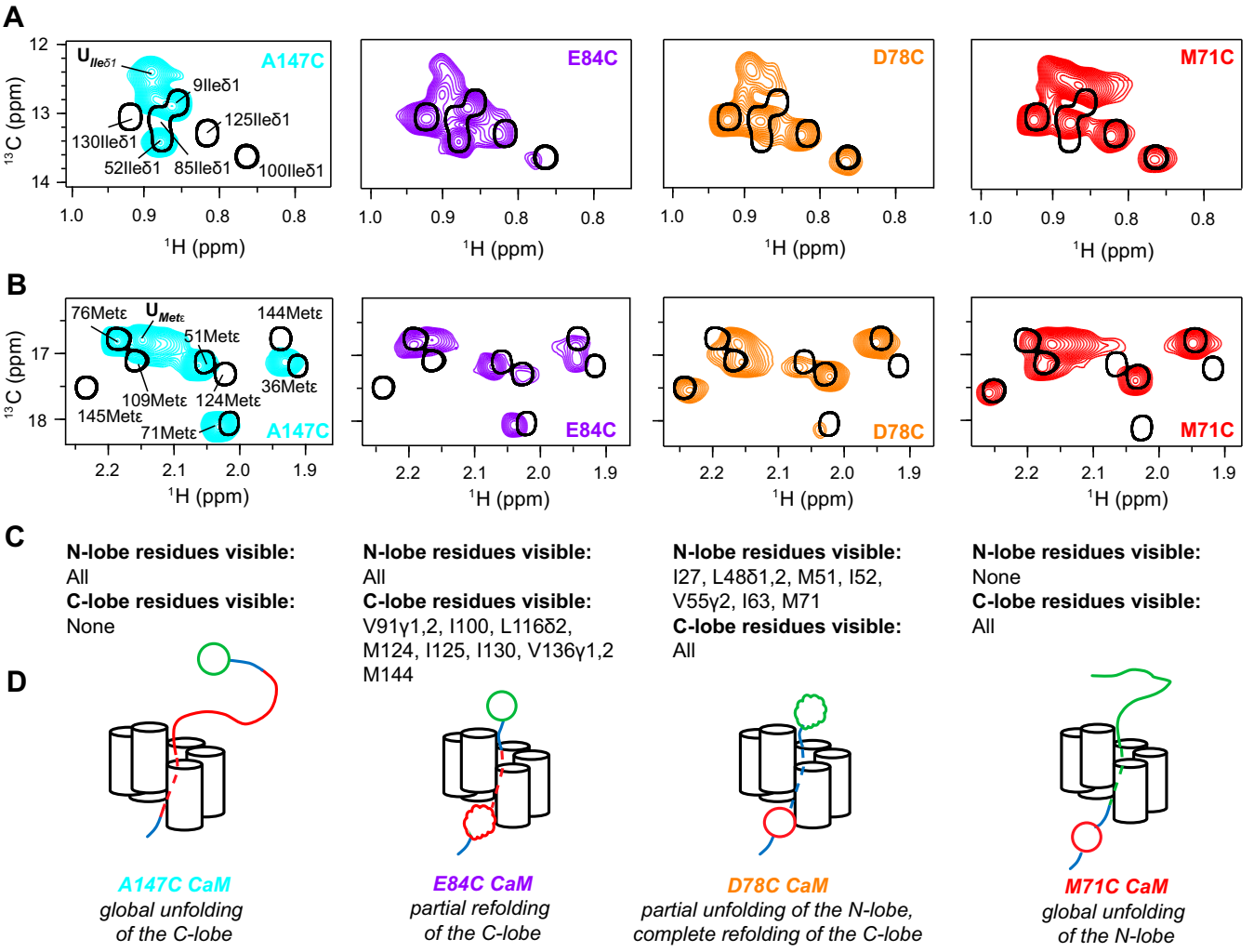



**Fig. S5.** Unfolding of CaM from the C terminus follows a similar path as N-terminal unfolding. Ile (*A*) and Met (*B*) regions of $^{13}$C-$^{1}$H HMQC spectra of cross-linked Cys mutants of CaM and of unbound WT CaM (black single contours). Cross-linking positions 147, 84, 78, and 71, shown above for the C-terminal pulling reaction, correspond to positions 4, 73, 80, and 89 for N-terminal pulling. (*C*) Evolution of structure during C-terminal unfolding, translocation, and refolding as established by the presence (visible) or absence of "native" correlations in spectra, as listed. All of the C-lobe signals disappear upon cross-linking at position 147 (*B*, *Left*), begin to reappear as a result of partial refolding for the E84C-linked CaM sample (*B*, *Middle Left*) and are present for D78C CaM-VAT. In contrast, some of the N-lobe peaks are not at their "folded positions" for D78C CaM-VAT, with complete unfolding of the N-lobe for the CaM M71C-VAT cross-linked sample. Note the increase in intensities of peaks located in the random-coil regions of the spectra, $U_{Ile\delta1}$ and $U_{Met\varepsilon}$. (*D*) Proposed model of unfolding/refolding of CaM domains with pulling from the C terminus of the substrate.

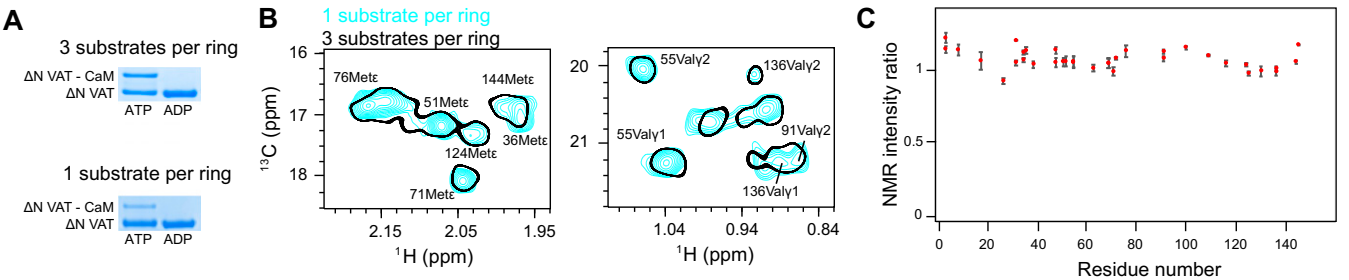



**Fig. S6.** Similar unfolding/refolding profile when one or three substrates on average are attached per VAT. (*A*) Cross-linking reaction using ΔN VAT bearing six W541C mutations per ring results in three substrate molecules (CaM D80C) bound, as judged from SDS/PAGE band intensities (~1:1 ratio). Using mixed hexamers (W541 ΔN VAT and WT ΔN VAT in a 1:11 ratio), followed by purification to select for VAT molecules conjugated with substrate, results in attachment of approximately only one substrate molecule on average (SDS band ratio 1:5). (*B*) Superposition of Met or Val regions of $^{13}$C-$^{1}$H HMQC spectra of ILVM-labeled His-tagged CaM D80C cross-linked to perdeuterated ΔN VAT W541C bearing six cysteine mutations per ring (black; single contour) or one mutation per ring, on average (cyan; multiple contours). (*C*) Intensity ratios of cross-peaks in spectra recorded on samples with approximately one or three substrate molecules cross-linked to ΔN VAT normalized to take into account the relative protein concentrations and the different number of transients used to record the data. Vertical bars indicate errors.

**Table S1. A73C cross-linked CaM-VAT intermediate structure via N-terminal pulling**

| Residue | Location | Present or absent | Comments |
|---|---|---|---|
| Leu4δ1,δ2 | Before helix 1 | Absent | In the native structure Leu4 interacts with Ala73 (C terminus of helix 4) that is mutated to Cys in the A73C-CaM cross-linked complex. |
| Ile9 | Helix 1 | Absent | In the native structure Ile9 contacts helix 4 (Leu69), which is not yet formed. |
| Leu18 | Helix 1 | Ambiguous | Side chain pointing to solvent, but also located within helix 1 that interacts with helix 4 in native structure. δ1 methyl close to random coil region and δ2 methyl overlaps with L112 δ2. |
| Ile27 | β-Sheet between helix 1 and helix 2 | Present | Side chain essential to core of the N-lobe, backbone forms H-bond with Ile63 and stabilizes domain fold. Also interacts with Phe16 from helix 1, which may suggest that the C terminus of helix 1 forms native-like contacts. |
| Leu32δ1,δ2 | Helix 2 | Absent | Interacts with side chains of Ile27, Leu48, and Met71 that are visible in spectra. Its absence may be related to chemical exchange. |
| Val35γ1,γ2 | Helix 2 | Absent | Interacts with Ala15 and Phe19 from helix 1, which is unlikely to be native-like due to missing interactions with helix 4 that is not yet formed. |
| Met36 | Helix 2 | Absent | In the native structure interacts with the C terminus of helix 4, including Met71 and Lys75 that are sequestered in the channel of VAT for this cross-link. |
| Leu39δ1,δ2 | Helix 2 | Absent | Interacts with Phe12 and Ala15 from helix 1, consistent with helix 1 not yet fully formed. |
| Leu48δ1,δ2 | Helix 3 | Present | Interacts with helix 2 (in contact with Gly33) in the native structure. |
| Met51 | Helix 3 | Present | Interacts with other residues from helix 3 (Leu48) and in contact with Met71 (from helix 4). |
| Ile52 | Helix 3 | Present | In contact with Thr29 from helix 2. |
| Val55γ2 | Helix 3 | Present | Interacts with Ile63 (from β-sheet) and Met71 (helix 4). γ1 methyl overlaps with Val142. |
| Ile63 | β-Sheet between helix 3 and helix 4 | Present | Side chain essential to the core of the N-lobe, backbone forms critical H-bond with Ile27 and stabilizes the fold of the N-domain. |
| Leu69δ1,δ2 | Helix 4 | Absent | Interacts with helix 1 (contact with Ile9). |
| Met71 | Helix 4 | Present | Side chain pointing toward the core of the domain, interacts with Val55 and Ile63. |
| Met72 | Helix 4 | Absent | Too close to the cross-linking point (position 73). |
| Met76 | Helix 4 | Absent | Sequestered inside the pore of VAT. |

Pulling from the N terminus leads to the formation of an intermediate structure involving the N-lobe as position 73 in CaM passes the D2 pore-loop 1 of VAT. A summary of the ILVM-methyl peaks reporting on structure is provided along with evidence for the presence or absence of secondary structural elements.

Table summary:

*i*) Little evidence to suggest that helix 1 is formed in the partially folded state (A73C cross-link):

• All nonoverlapping resonances from methyl-containing residues in helix 1 are absent (Ile9 and Leu18) as is Leu4 located prior the beginning of helix 1.

• In the native structure of CaM, helix 1 contacts helix 4 that for the A73C cross-linked protein is (at least) partially unfolded and sequestered inside the channel of VAT, preventing helix 1 from forming native contacts with helix 4.

• Leu4 interacts with Ala73 in native CaM that now serves as the cross-link point (location of inserted Cys residue), precluding native-like contacts.

• Phe16 from the C terminus of helix 1, that contributes to the stability of the domain core, forms native contacts with Ile27 (based on the presence of Ile-27 cross peak).

*ii*) β-strand linking helix 1 and helix 2 is formed:

• Cross-peak for Ile27, a residue that forms critical interactions in the native structure, is present in native position.

*iii*) Helix 2 is not natively formed:

• Several methyl-containing residues from this helix (Val35 and Leu39) contact helix 1 in the native state, while Met36 interacts with the C terminus of helix 4, yet helices 1 and 4 are not natively formed.

• There is some indirect evidence that helix 2 may have some native-like properties. Helix 2 contacts helix 3 that appears formed. Helix 2–3 contacts include Gly33 (helix 2)–Leu48 (helix 3, visible) and Thr29 (helix 2)–Ile52 (helix 3, visible). Contacts between helices 1 and 2 are not formed, yet the interface between helix 2 and helix 3 appears native-like.

*iv*) Helix 3 appears native-like:

• All methyl signals from helix 3 are visible at native locations in the spectrum for A73C CaM cross-linked to VAT, strongly suggesting that native interactions involving helix 3 are at least transiently formed.

*v*) β-Strand linking helices 3 and 4 is formed:

• Cross-peak for Ile63, a residue that forms critical interactions in native structure, including with Ile27 is present at its native position in spectra. The importance of this residue and Ile27 is underscored by the fact that Ala mutations lead to compete unfolding of the apo domain.

*vi*) Little evidence for folding of helix 4:

• The C-terminal helix residues are not available for refolding (at least 73–76) due to placement of cross-link at position 73.

• Notably the NMR signal from Met71 is visible in spectra, suggesting that this residue forms native-like contacts with Val55 (helix 3) and Ile63 (strand between helices 3 and 4).

• Phe68 that is located at the N terminus of the helix 4 contacts Ile27 that is visible in spectra. This suggests that the N-terminal region of helix 4 may be interacting in a native-like manner with other regions of secondary structure and hence partially formed.

When pulling from the C terminus, the intermediate obtained for D78C CaM is very similar to that for A73C highlighted above, based on the presence/absence of correlations in spectra as outlined in Table S1.

**Table S2. D80C cross-linked CaM-VAT intermediate structure via N-terminal pulling**

| Residue | Location | Present or absent | Comments |
|---|---|---|---|
| Ile85 | Helix 5 | Absent | Close to the cross-linking point, within helix 5 |
| Val91γ1,γ2 | Helix 5 | Present | Second-to-last residue of helix 5, side chain pointing toward solvent and interacting with neighboring Phe92. |
| Ile100 | β-Sheet between helix 5 and helix 6 | Present | Side chain forms core of the C-lobe (with Ile125 and Ile130 among others), backbone forms H-bond with Val136 that is critical for stabilization of the fold of the C-domain. |
| Leu105δ1 | Helix 6 | Present | Side chain pointing toward the core, but also in contact with Phe89 from helix 5 (that is likely unfolded; see below). |
| Leu105δ2 | Helix 6 | Absent | Side chain pointing toward the core, but also in contact with Phe89 from helix 5 (that is likely unfolded; see below). |
| Val108γ1,γ2 | Helix 6 | Absent | Interacts with helix 5 (Phe89), which is unfolded. |
| Met109 | Helix 6 | Ambiguous | A very weak peak is present at the native position; however, there is partial overlap with intense unfolded correlations. Met109 is involved in core formation (close contact with Met124 from helix 7). |
| Leu112δ1 | Loop between helix 6 and helix 7 | Absent | Close contact with the N terminus of helix 5 (Ala88), which is unfolded. Leu112δ2 overlaps with Leu18δ2. |
| Leu116δ1 | Loop between helix 6 and helix 7 | Present | Side chain at the interface between helices 6, 7, and 8, within the core of the domain, in contact with Met124. Leu116δ2 resonates in unfolded region so that quantitation is not possible. |
| Val121γ1,γ2 | Helix 7 | Absent | Side chains pointing toward the core of the domain, contacting Met109. Absence may be related to intrinsic dynamics of the intermediate and/or lower intensities of Val121 resonances in native conformation. |
| Met124 | Helix 7 | Present | Residue at the interface between helices 7 and 8, interacting with Met144 (helix 4). |
| Ile125 | Helix 7 | Present | Close contact with Arg106 (helix 6) and Ile130. |
| Ile130 | Loop between helix 7 and β-sheet | Present | Tightly packed between Ile100 and Ala102 stabilizing interactions between EF3 and EF4. In contact with Ile125 from helix 7. |
| Val136γ1,γ2 | β-Sheet between helix 7 and helix 8 | Present | Side chain forms the core of the C-lobe, backbone forms H-bond with Ile100 and stabilizes the fold of the domain. |
| Met144 | Helix 8 | Present | Side-chain pointing toward helix 7 (Met124 and Ala128). |
| Met145 | Helix 8 | Absent | Interacts with Ile85 from helix 5 (unfolded). |

Pulling from the N terminus leads to the formation of an intermediate structure involving the C-lobe as position 80 in CaM passes the D2 pore-loop 1 of VAT. A summary of the ILVM-methyl peaks reporting on structure is provided along with evidence for the presence or absence of secondary structural elements.

Table summary:

*i*) Absence of helix 5:

• Absence of Ile85 peak in spectra.

• In the native structure helix 5 interacts with helices 6 and 8 (in the N-lobe helix 1 interacts with 2 and 4). Here, all peaks from residues in other regions of the domain that would be in contact with helix 5 are missing from spectra (with the exception of Leu105δ1). This includes Leu105δ2 and Val108 (both from helix 6) that would normally interact with Phe89, Leu112 (located within the loop between helices 6 and 7) that would interact with Ala88, and, finally, Met145 (helix 8) that contacts Ile85 (helix 5).

• Notably peaks from Val91 located on the C terminus of helix 5 are observed. This side chain, however, is oriented toward the outside of the domain (solvent exposed) and its chemical shift would be very little affected by the backbone conformation.

*ii*) β-Strand linking helix 5 and 6 is formed:

• Cross-peak for Ile100, a residue that forms critical interactions in the C-lobe, is present.

*iii*) Helix 6 is unlikely to be natively formed:

• Leu105δ2 and Val108γ1,γ2 are absent in spectra since their native orientation is toward the helix 5 that is unfolded; Met109 is ambiguous.

• Arg106 (helix 6) contacts Ile125 (helix 7) that is visible in the spectra. The interface between helices 6 and 7 may be similar to that in the native state, but interactions between helix 5 and 6 are broken due to the unfolding of helix 5.

*iv*) Helix 7 appears to be native-like:

• Met124 and Ile125 are visible in spectra.

• Methyl groups from loop residues flanking this helix (Leu116 and Ile130) and interacting with helix 7 residues are also at their native locations.

• Val121 is the only helix 7 residue that is not observed; may reflect broadening related to conformational exchange of the unfolding intermediate.

*v*) β-Strand linking helices 7 and 8 is formed:

• Val136, that forms key interactions with Ile100 and is critical for domain stability, is present.

*vi*) Helix 8 is not native but may be partially formed:

• Met144 side chain oriented toward helix 7 is visible in NMR spectra.

• Cross-peak of Met145 is absent due to the loss of interaction with Ile85 (unfolded helix 5).

When pulling from the C terminus the intermediate obtained for E84C CaM is very similar to that for D80C highlighted above, based on the presence/absence of correlations in spectra as outlined in Table S2.