2016

Atomic resolution experimental phase information reveals extensive disorder and bound 2-methyl-2,4-pentanediol in Ca^{2+}-calmodulin

Jiusheng Lin
University of Nebraska-Lincoln, jlin4@unl.edu

Henry van den Bedem
Stanford University

Axel T. Brunger
Stanford University

Mark A. Wilson
University of Nebraska-Lincoln, mwilson13@unl.edu

STRUCTURAL BIOLOGY
ISSN 2059-7983

Atomic resolution experimental phase information reveals extensive disorder and bound 2-methyl-2,4-pentanediol in Ca^{2+}-calmodulin

Jiusheng Lin,[a] Henry van den Bedem,[b] Axel T. Brunger[c] and Mark A. Wilson[a]*

[a]Department of Biochemistry and Redox Biology Center, University of Nebraska, Beadle Center, Lincoln, NE 68588, USA, [b]Biosciences Division, SLAC National Accelerator Laboratory, Stanford University, Menlo Park, CA 94025, USA, and [c]Departments of Molecular and Cellular Physiology, Neurology and Neurological Sciences, Structural Biology, and Photon Science, Stanford University and Howard Hughes Medical Institute, Stanford, CA 94305, USA. *Correspondence e-mail: mwilson13@unl.edu

Received 8 October 2015
Accepted 14 November 2015

Edited by Z. Dauter, Argonne National Laboratory, USA

Keywords: experimental phases; disorder; *qFit*; density modification.

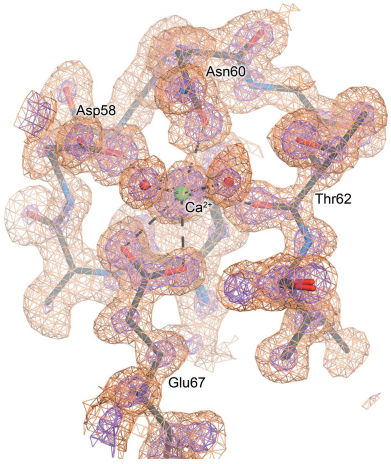
PDB references: Ca^{2+}-calmodulin from *Paramecium tetraurelia*, 5e1p; selenomethionine derivative, 5e1k; 5e1n

Supporting information: this article has supporting information at journals.iucr.org/d

Calmodulin (CaM) is the primary calcium signaling protein in eukaryotes and has been extensively studied using various biophysical techniques. Prior crystal structures have noted the presence of ambiguous electron density in both hydrophobic binding pockets of Ca^{2+}-CaM, but no assignment of these features has been made. In addition, Ca^{2+}-CaM samples many conformational substates in the crystal and accurately modeling the full range of this functionally important disorder is challenging. In order to characterize these features in a minimally biased manner, a 1.0 Å resolution single-wavelength anomalous diffraction data set was measured for selenomethionine-substituted Ca^{2+}-CaM. Density-modified electron-density maps enabled the accurate assignment of Ca^{2+}-CaM main-chain and side-chain disorder. These experimental maps also substantiate complex disorder models that were automatically built using low-contour features of model-phased electron density. Furthermore, experimental electron-density maps reveal that 2-methyl-2,4-pentanediol (MPD) is present in the C-terminal domain, mediates a lattice contact between N-terminal domains and may occupy the N-terminal binding pocket. The majority of the crystal structures of target-free Ca^{2+}-CaM have been derived from crystals grown using MPD as a precipitant, and thus MPD is likely to be bound in functionally critical regions of Ca^{2+}-CaM in most of these structures. The adventitious binding of MPD helps to explain differences between the Ca^{2+}-CaM crystal and solution structures and is likely to favor more open conformations of the EF-hands in the crystal.

1. Introduction

Calmodulin (CaM) is a highly conserved eukaryotic protein that binds to Ca^{2+} and regulates the activities of hundreds of downstream binding partners. CaM contains two similar domains, each of which comprises two EF-hand Ca^{2+}-binding motifs (Babu *et al.*, 1985). These two domains are separated by a flexible linker region that allows the large interdomain mobility needed for CaM to interact with its structurally diverse targets (Ikura, Clore *et al.*, 1992; Meador *et al.*, 1992). Upon binding Ca^{2+}, the EF-hand helices in each of the two domains of CaM sample more open conformations and expose clusters of predominantly hydrophobic residues with an unusually high surface density of phenylalanine and methionine (Kuboniwa *et al.*, 1995; O'Neil & Degrado, 1990). These flexible hydrophobic pockets are then poised to interact with protein targets in a variety of interaction modes that have been structurally characterized over the past two decades (Franco-Echevarría *et al.*, 2014; Lau *et al.*, 2012; Sarhan *et al.*,



2012; Fallon *et al.*, 2009; Meador *et al.*, 1992, 1993; Yamauchi *et al.*, 2003).

Because of its physiological importance, small size (148 amino acids), multidomain architecture and extensive conformational dynamics, CaM has been widely studied using a variety of structural and biophysical techniques. Both solution NMR spectroscopy and X-ray crystallography have demonstrated that CaM is highly flexible, particularly in the central linker region comprising residues 70–90 (Ikura, Barbato *et al.*, 1992; Meador *et al.*, 1993). Despite general agreement about its structure, NMR and X-ray crystallography have provided differing views of several important details of Ca^{2+}-CaM. For example, the central linker region is helical in triclinic crystal structures of Ca^{2+}-CaM (Babu *et al.*, 1988; Chattopadhyaya *et al.*, 1992), but NMR spectroscopy and crystal structures in other space groups have shown that the linker is nonhelical from residues 78 to 81 (Fallon & Quiocho, 2003; Ikura, Barbato *et al.*, 1992). In addition, an NMR spectroscopic study using residual dipolar coupling information showed that both domains of Ca^{2+}-CaM are more closed in solution than in most crystal structures, particularly in the N-terminal domain (Chou *et al.*, 2001). Fundamentally, these differences provide strong evidence for the highly dynamic character of Ca^{2+}-CaM; however, a more specific explanation for these disparities is desirable. Of particular interest is why the EF-hands in the triclinic crystal structure of Ca^{2+}-CaM are more open than they are in solution. While crystal lattice artefacts are often invoked to explain the continuous central helix region observed in triclinic Ca^{2+}-CaM, the more open EF-hands in each domain are not as easily rationalized using crystal-packing arguments.

Previously, we determined an atomic resolution (1.0 Å) X-ray crystal structure of Ca^{2+}-CaM from *Paramecium tetraurelia* (PDB entry 1exr; Wilson & Brunger, 2000). The high resolution of this data set allowed the detailed characterization of side-chain and backbone atomic disorder on multiple length scales, providing evidence for widespread spatially correlated disorder in the protein. However, several features in those electron-density maps remained ambiguous, including the prominent difference electron density located in both hydrophobic binding pockets (Wilson & Brunger, 2000). These features were mentioned in the original publication but were left unmodeled owing to uncertainty in their proper assignment. Similar electron-density features have been described in other triclinic Ca^{2+}-CaM structures (Chattopadhyaya *et al.*, 1992), but no adequate explanation for them has been provided. Because these unexplained electron-density features are located in functionally critical regions of Ca^{2+}-CaM, it is important to convincingly assign them to a molecular species.

To examine these ambiguous electron-density features with minimal structural model phase bias, we collected a 1.0 Å resolution single-wavelength anomalous diffraction (SAD) data set from Ca^{2+}-SeMet CaM. Combined with modern density-modification procedures, the experimental phases produced an electron-density map of comparable quality to the $2mF_o - DF_c$ electron density calculated from the final refined model. The experimentally phased electron-density map significantly reduced the ambiguity in the assignment of correlated main-chain and side-chain disorder in Ca^{2+}-CaM. In addition, this experimental phase information confirms that 2-methyl-2,4-pentanediol (MPD) molecules are located in the C-terminal binding pocket and mediate a crystal-packing contact between adjacent N-terminal domains in the lattice. Electron density in the N-terminal binding pocket is highly suggestive of another bound MPD. MPD has been used at ∼50%(*v*/*v*) concentration as a precipitant in all prior triclinic Ca^{2+}-CaM structures; therefore, its binding to these regions of the protein offers a potentially general explanation for the more open EF-hands of the triclinic Ca^{2+}-CaM crystal structure.

2. Methods

2.1. Protein expression, purification and crystallization

P. tetraurelia CaM was expressed using the previously described pKK233-3 expression construct (Wilson & Brunger, 2000). Selenomethionine (SeMet) was incorporated using the *Escherichia coli* methionine-auxotroph strain B834 grown in M9 defined medium containing 1% glucose, 50 mg l^{-1} of each amino acid except methionine and cysteine, 7.5 mg l^{-1} $FeSO_4$ and 50 mg l^{-1} L-SeMet. Protein expression and purification were performed as described previously (Wilson & Brunger, 2000).

The purified and lyophilized SeMet CaM protein was resuspended at 15 mg ml^{-1} in 5 m*M* $CaCl_2$, 50 m*M* sodium cacodylate pH 5.0 and crystallized using the hanging-drop vapor-equilibration method. A drop consisting of 4 µl 15 mg ml^{-1} SeMet CaM in buffer was mixed with 1 µl 2-methyl-2,4 pentanediol (MPD). The drop was equilibrated against a 1 ml reservoir consisting of 50 m*M* sodium cacodylate pH 5.0, 50% MPD at 4°C. Crystals of the native protein were used for initial microseeding of SeMet CaM crystals using a cat whisker. Subsequent microseeding was performed using SeMet CaM crystals to ensure a uniform population of SeMet-substituted protein.

Ca^{2+}-SeMet CaM crystals grew in 2–5 d and reached dimensions of 0.7 × 0.4 × 0.1 mm. The crystals were blade-shaped and were mounted in nylon loops such that the longest crystal axis was not collinear with the pin. The 50% concentration of MPD in the mother liquor is sufficient to cryoprotect the crystals, which were removed from the drop and immersed into liquid propane held at its melting point in a liquid-nitrogen bath.

2.2. Data collection and processing

SAD data were collected on beamline 9-1 at the Stanford Synchrotron Radiation Laboratory (SSRL) from a single crystal maintained at 100 K. The crystal was illuminated with 12 670 eV incident X-rays (0.9785 Å) and data were measured using a MAR345 image-plate detector with 150 µm pixel size. The X-ray energy is ∼10 eV higher than the peak of the measured Se X-ray fluorescence in similar crystals of

Table 1
Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	Beamline 9-1, SSRL
Data-collection date	7/4/2000
Wavelength (Å)	0.9785
Temperature (K)	100
Detector	MAR345
Space group	*P*1
a, *b*, *c* (Å)	25.04, 29.41, 52.79
α, β, γ (°)	89.45, 86.29, 82.39
Mosaicity (°)	0.45
Resolution range (Å)	25–1.00 (1.02–1.00)
Total No. of reflections	419257
No. of unique reflections	75551
Completeness (%)	94.0 (84.7)
Anomalous completeness (%)	92.6 (78.8)
Multiplicity	5.5 (3.8)
Anomalous multiplicity	2.8 (2.0)
$\langle I/\sigma(I)\rangle$†	11.5 (1.6)
$CC_{1/2}$‡	0.997 (0.294)
R_{meas} (I^+, I^- merged)	0.121 (2.19)
R_{meas} (I^+, I^- separate)	0.099 (2.19)
Overall *B* factor from Wilson plot (Å^2)	8.96

† $\langle I/\sigma(I)\rangle$ decreases to 2.0 at 1.04 Å resolution. ‡ $CC_{1/2}$ (Karplus & Diederichs, 2012) was used to determine the high-resolution cutoff.

Ca^{2+}-SeMet CaM. Data were collected in inverse-beam geometry with 20° wedges in both high-resolution and low-resolution passes to avoid overloaded reflections. In order to collect data in the blind cone, a second high-resolution pass was measured from the crystal after reorientation. Diffraction data were indexed, integrated and scaled in *HKL*-2000 (Otwinowski & Minor, 1997). Merging statistics, anomalous signal and resolution limit were independently analysed using the unmerged, integrated data with *AIMLESS* (Evans & Murshudov, 2013) as implemented in the *CCP*4 suite (Winn *et al.*, 2011). Final data statistics are provided in Table 1.

2.3. SAD phasing and density modification

Anomalously scattering atoms were located using *SHELXC*/*D* (Sheldrick, 2008), compared with the anomalous difference Fourier maps calculated using model phases from PDB entry 1exr and moved to a common origin to facilitate comparison with the previously solved structure. These 12 heavy-atom sites were input into *Phaser* for SAD phasing against the full data set extending to 1.0 Å resolution (Read & McCoy, 2011). The overall figure of merit for the final SAD phases was 0.44.

Density modification was performed using default settings in *RESOLVE* as implemented in *PHENIX* (Terwilliger, 2004), *SOLOMON* (Abrahams & Leslie, 1996), *Pirate* (Cowtan, 2000), *DM* (Cowtan *et al.*, 2001) and *Parrot* (Cowtan, 2010) in the *CCP*4 suite (Winn *et al.*, 2011). Importantly, we used Hendrickson–Lattman coefficients for the anomalous heavy-atom contribution only (HA_{anom}) for density modification wherever possible, as these consistently produced superior density-modified maps. Map quality was judged initially by visual inspection and quantified by calculating the correlation coefficient with the final refined $2mFo - DF_c$ map using *PHENIX* (Adams *et al.*, 2011). Of the density-modification programs tested, *Parrot* (Zhang *et al.*, 1997; Cowtan, 2010) delivered the best maps, which could be further improved slightly by using PDB entry 4zgf (Joint Center for Structural Genomics, unpublished work) as a reference data set for histogram matching rather than the default PDB entry 1tqw (Kuntz *et al.*, 2005).

Table 2
Structure solution and refinement.

n.a., not applicable; this quantity is not calculated by the refinement program used.

Model	Ca^{2+}-SeMet CaM	Ca^{2+}-SeMet CaM, *qFit*2.0	Ca^{2+}-CaM, *qFit*2.0
PDB code	5e1k	5e1n	5e1p
Refinement program	*SHELX*-14	*PHENIX* 1.9	*PHENIX* 1.9
Resolution range (Å)	25–1.00	25–1.00	25–1.00
Completeness (%)	94.0	94.0	97.4
No. of reflections, working set	71779	71737	69345
No. of reflections, test set	3772	3772	7778
R_{cryst}	0.1436	0.1337	0.1249
R_{free}	0.1649	0.1495	0.1424
R_{all}	0.1440	0.1345	0.1267
No. of non-H atoms			
Protein	1636	2942	2607
Ion	5	5	5
Ligand	16	16	16
Water	176	178	180
Total	1833	3141	2808
R.m.s. deviations			
Bonds (Å)	0.013	0.013	0.010
Angles (°)	n.a.	1.489	1.317
Angle (1–3) distances (Å)	0.0294	n.a.	n.a.
Average *B* factors (Å^2)			
Protein	17.5	14.1	13.9
Ion	11.4	10.8	10.3
Ligand	28.2	29.8	21.0
Water	32.3	26.8	25.5
ADP anisotropy†			
Protein	0.388	0.432	0.464
Ion	0.414	0.384	0.368
Ligand	0.358	0.578	0.714
Water	0.335	0.479	0.594
Ramachandran plot			
Most favored (%)	99.6	99.4	98.7
Allowed (%)	0.4	0.6	0.9

† ADP anisotropy is defined as the ratio of the smallest to the largest eigenvalue of the anisotropic ADP tensor.

2.4. Model refinement and validation

An initial model derived from PDB entry 1exr with all amino-acid side chains in a single conformation was used to build the Ca^{2+}-SeMet CaM model. 1exr was used instead of an automatically built model because autobuilding struggled with the weak density of the central helix region, resulting in fragmented models with the two domains separated. Adjustments to the model, including the introduction of ordered solvent molecules and alternate amino-acid conformations, were made in *Coot* (Emsley & Cowtan, 2004) with reference to density-modified experimentally phased $2mF_o - DF_c$ and $mF_o - DF_c$ maps. Cycles of model building and adjustment were alternated with conjugate-gradient least-squares refinement in *SHELX* against an intensity-based residual target function with geometric (DFIX, DANG, FLAT), ADP (DELU, SIMU) and occupancy (PART) restraints (Sheldrick,

2008). Friedel-merged intensities were used. Anisotropic ADPs were introduced into the model after convergence of the isotropic ADP refinement, and riding H atoms were included except for the O^{γ} atom of Ser, the OH atom of Tyr and the $N^{\delta1}$ atom of His, owing to ambiguities in their idealized position or residue ionization state. In the final cycles of refinement of the converged model, the test set of reflections that were previously sequestered for the calculation of the R_{free} value (Brünger, 1992) were included in the refinement. Several close contacts between partially occupied water molecules and disordered protein atoms were identified during model building. These clashes occur between atoms whose occupancies sum to less than unity and are therefore examples of correlated disorder. In order to prevent these regions being incorrectly flagged as van der Waals contact violations, we assigned altloc IDs to the disordered water molecules in the final models.

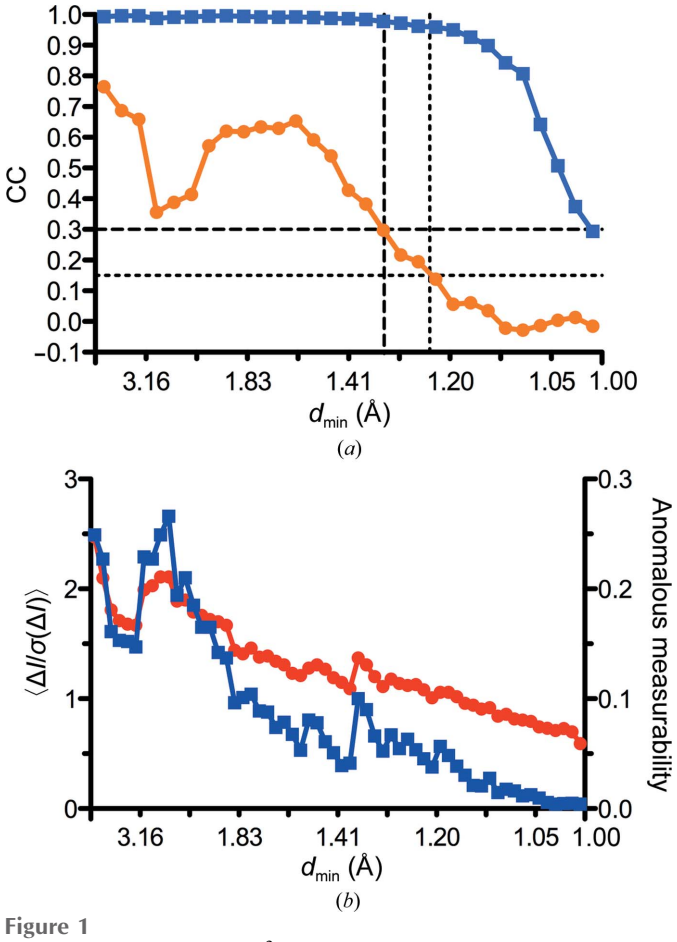


Figure 1
Anomalous signal in the Ca^{2+}-SeMet CaM Se SAD data set. In (*a*), the correlation coefficient for anomalous differences over random halves of the data (CC_{anom}; orange circles) and the correlation coefficient for intensities over random halves of the data ($CC_{1/2}$; blue squares) are shown as a function of resolution. The dashed line shows the CC_{anom} cutoff of 0.3 (1.33 Å resolution) and the dotted line shows a less conservative cutoff of 0.15 (1.23 Å resolution). In (*b*), the extent of anomalous signal is plotted as the ratio of the anomalous difference signal to noise (red circles) and anomalous measurability (blue squares). All three estimates suggest that there is measurable anomalous signal to ~1.3 Å resolution in this data set.

In order to facilitate direct comparison between the *SHELXL*-refined model and the *PHENIX*-refined *qFit*2.0 models (see below), we used the program *phenix.model_vs_data* to calculate the *R* and R_{free} values for the *SHELXL* model using the bulk-solvent and anisotropic scaling routines implemented in *PHENIX* (Afonine *et al.*, 2010). These values are R_{free} = 15.6%, *R* = 13.4% and should be compared with the values reported in Table 2. The model was validated using *Coot* (Emsley & Cowtan, 2004) and *MolProbity* (Chen *et al.*, 2010) and the anisotropic ADPs were analysed using *PARVATI* (Zucker *et al.*, 2010). Refined model statistics are shown in Table 2. Structural figures were produced with *POVScript+* (Fenn *et al.*, 2003).

Automated detection and modeling of alternative side-chain and backbone conformations was performed using *qFit*2.0 (Keedy *et al.*, 2015). Structure-factor amplitudes and refined single-conformer models were used to calculate $2mF_o - DF_c$ and $mF_o - DF_c$ maps. For the Ca^{2+}-CaM *qFit*2.0 model (PDB entry 5e1p), the diffraction data were those deposited with the original PDB code 1exr. Electron density was sampled using standard rotamer libraries for all residues that had anisotropic electron density at side-chain or backbone atoms indicative of a potentially disordered residue. One to four conformations were chosen for each amino acid and the resulting models were refined in *PHENIX* with anisotropic ADPs and riding H atoms against an Friedel-merged amplitude-based maximum-likelihood target with geometric and ADP restraints (Afonine *et al.*, 2012). The ADP and geometric weights were optimized and the occupancies of alternative conformations were refined but constrained to sum to unity. Model statistics are shown in Table 2.

Anisotropic ADP validation was performed using *PARVATI* (Zucker *et al.*, 2010) and *ANISOANL* (Winn, 2001). The anisotropies of the *qFit* models refined using *PHENIX* are systematically higher (indicative of more isotropic character of the ADPs) than those refined against the same data in *SHELX* (Table 2). This disparity reflects differences in the type and application of anisotropic ADP restraints in these two programs.

3. Results

3.1. SAD phasing and density modification

Experimental phases were obtained using single-wavelength anomalous diffraction (SAD) near the Se *K* edge of Ca^{2+}-SeMet CaM crystals. CaM has eight methionines (excluding the N-terminal Met), resulting in the incorporation of eight anomalous scatterers in 148 amino acids. In addition, the five bound Ca^{2+} ions also have a weak anomalous signal (0.58 e^{-}) at this wavelength. The absence of point-group symmetry in these crystals poses particular challenges for anomalous data collection, including balancing the requirements for reasonable data completeness and multiplicity with the need to minimize radiation damage to the sample. Consequently, a SAD data set was collected because a two-wavelength MAD data collection would inflict unacceptable radiation damage

on the sample. Furthermore, the completeness is lower than desired in the highest resolution bin (85%; Table 1) owing to complications arising from the blind cone in space group *P*1 at high resolution. Although the sample was reoriented in order to capture some of the blind-cone data, the range of this reorientation was limited by the beamline geometry.

The SeMet SAD data set contains significant anomalous signal to approximately 1.33 Å resolution based on the anomalous correlation coefficient between random halves of the unmerged data (CC_{anom} = 0.3) as calculated by *AIMLESS* (Evans & Murshudov, 2013; Fig. 1*a*). Some anomalous signal may extend as far as 1.23 Å resolution if a less conservative cutoff of CC_{anom} = 0.15 is used. As additional measures of anomalous data quality, the mean anomalous signal-to-noise $[\langle \Delta I/\sigma(\Delta I)\rangle]$, where ΔI is the Bijvoet difference $I^+ - I^-$] and anomalous measurability (Zwart, 2005) also indicate that anomalous signal is present to 1.3 Å resolution (Fig. 1*b*) using cutoffs of 1.2 and 0.05, respectively. Despite challenges arising from the low symmetry of this triclinic crystal, the anomalous data have an overall completeness of 93% and a multiplicity of 2.8 (Table 1). These values are likely to represent upper limits attainable from this sample, as more extensive exposure of similar crystals to X-rays resulted in an obvious loss of high-resolution reflections. SAD phasing performed using *Phaser* (Read & McCoy, 2011) gives a final figure of merit of 0.44, which indicates good starting SAD phase-angle estimates.

As SAD phase probability distributions are bimodal, phase improvement by density modification was necessary to obtain optimal electron-density maps, although the SAD-phased maps were clearly interpretable without modification. Density modification was performed with default settings as implemented in several current software packages (see §2). While all produced excellent maps, the best-performing program was *Parrot* (Zhang *et al.*, 1997; Cowtan, 2010), which produced an electron-density map (Fig. 2) that was comparable to others in well ordered regions but showed clearer electron density for many disordered or partially occupied residues. We suspected that the superior performance of *Parrot* may be owing in part to its default use of the high-resolution (1.2 Å) reference structure 1tqw for histogram matching. This atomic resolution reference structure should more closely resemble the expected features of the electron density of Ca^{2+}-SeMet CaM at 1.0 Å resolution than would the default settings in other programs. In order to optimally match reference and target electron density, we selected PDB entry 4zgf as a new reference for histogram matching. The 1.0 Å resolution 4zgf structure (Joint Center for Structural Genomics, unpublished work) contains both selenomethionine and Ca^{2+} and thus closely matched the atomic composition and data-set resolution of Ca^{2+}-SeMet CaM. The resulting *Parrot* density-modified map had a correlation coefficient (CC) of 0.88 with the final $2mF_o - DF_c$ electron density, a CC of 0.77 with the final refined Ca^{2+}-SeMet CaM model and was slightly better by visual inspection

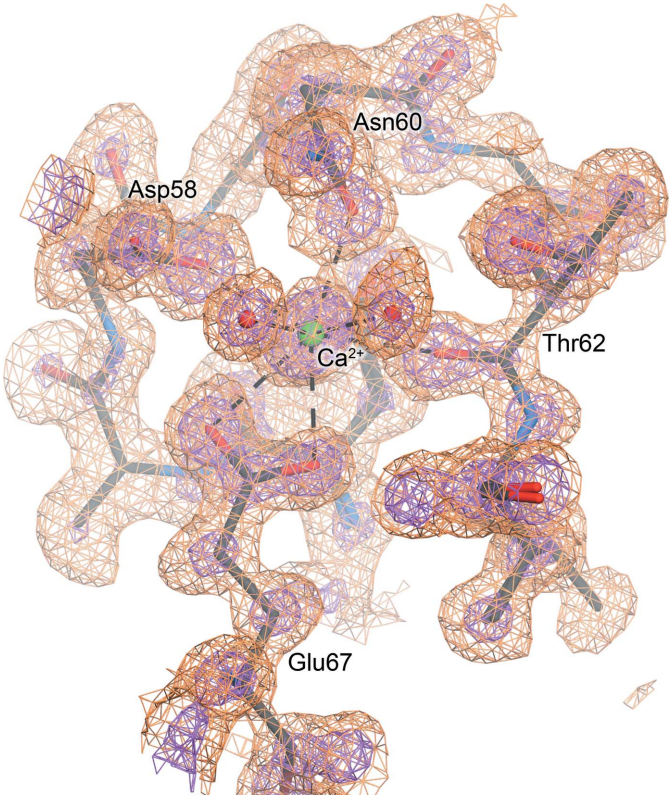


Figure 2
A view of the density-modified SAD electron density around the second Ca^{2+} site. *Parrot*-modified, SAD-phased experimental electron density is shown at 1.0σ (orange) and 4.0σ (magenta) for an EF-hand Ca^{2+}-binding site (site II). The map contoured at 4.0σ shows separation of atomic electron density for Glu67 and other residues.

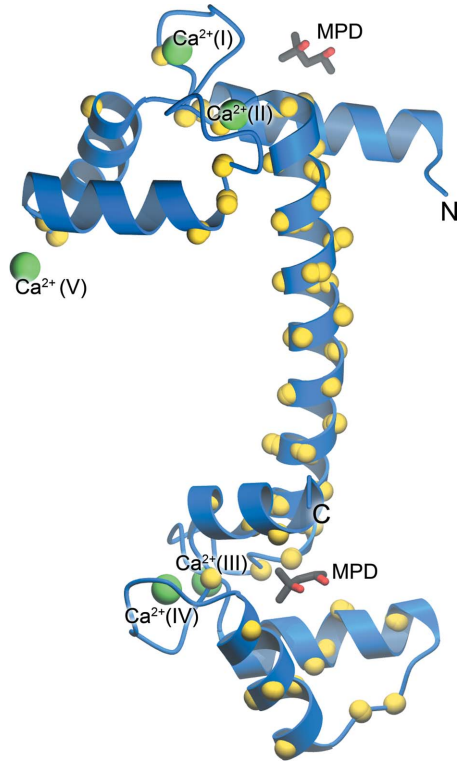


Figure 3
Location of disordered residues and bound MPD in Ca^{2+}-SeMet CaM. The locations of modeled disordered residues (yellow spheres) and of bound MPD are shown on a ribbon diagram of the *SHELXL*-refined model of Ca^{2+}-SeMet CaM (blue). N- and C-termini are labeled.

than the map obtained using PDB entry 1tqw as a reference. Overall, the *Parrot* $2mF_o - DF_c$ map CC was 0.05–0.1 higher than those obtained using maps calculated with other density-modification programs. The sensitivity of the final electron-density map to the particular density-modification algorithm employed indicates that this Ca^{2+}-SeMet CaM SAD data set could be useful for the testing and improvement of density-modification approaches at atomic resolution.

3.2. Modeling conformational disorder and comparison with 1exr

As expected, the final refined Ca^{2+}-SeMet CaM model is nearly identical to the previously refined 1.0 Å resolution crystal structure (PDB entry 1exr; Wilson & Brunger, 2000), with a C^{α} r.m.s.d. of 0.1 Å. The protein adopts the same extended, dumbbell-shaped conformation as seen in all other triclinic Ca^{2+}-CaM structures (Fig. 3). As in PDB entry 1exr, there is a fifth Ca^{2+} ion that mediates a lattice contact between Glu47 and Asp58 of a neighboring molecule.

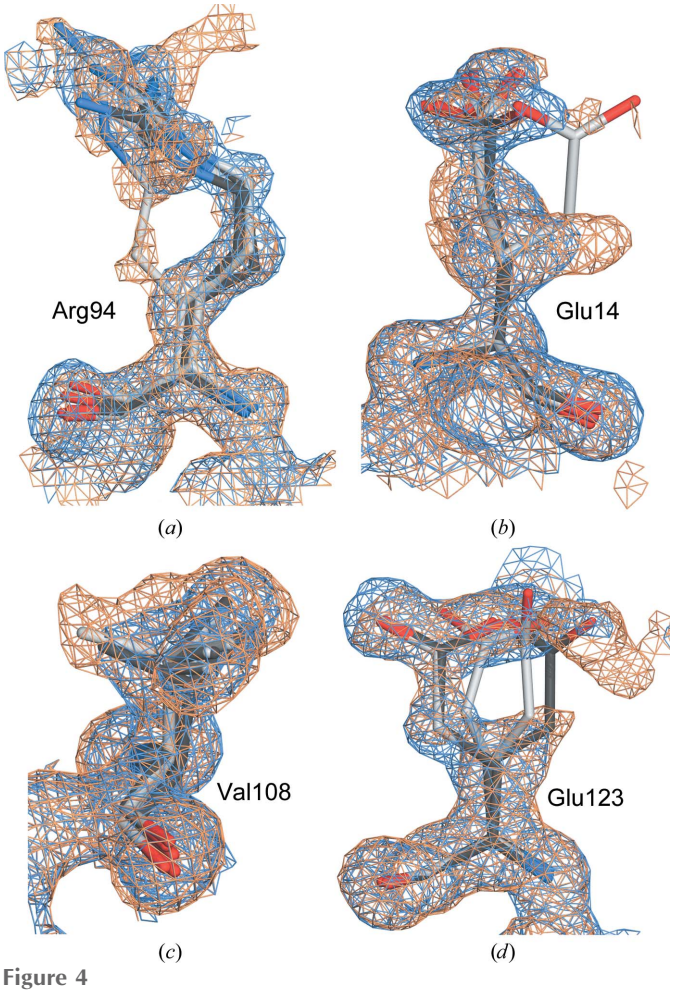


Figure 4
Experimental phase information clarifies cases of ambiguous discrete disorder. In all panels, *Parrot*-modified, SAD-phased electron density at 0.9σ is shown in orange and final $2mF_o - DF_c$ electron density at 0.9σ is shown in blue. The darker-colored atoms are the manually built disordered model and the residues in lighter grey are those automatically built by *qFit*2.0. Arg94 (*a*), Glu14 (*b*), Val108 (*c*) and Glu123 (*d*) all show clearer evidence of discrete disorder in the experimental electron-density map than in the $2mF_o - DF_c$ map at this contour level.

Ca^{2+}-CaM is one of the premier examples of physiologically relevant protein conformational dynamics. Consequently, crystalline Ca^{2+}-CaM is more extensively disordered than most proteins that diffract to atomic resolution. We previously modeled 37 disordered residues in the structural model of Ca^{2+}-CaM (PDB entry 1exr) using conservative criteria that excluded many ambiguously disordered residues (Wilson & Brunger, 2000). For the Ca^{2+}-SeMet CaM model, we relaxed these criteria and manually built a total of 57 residues into alternate conformations, which represents 39% of the protein (Fig. 3). The larger number of alternate conformations in Ca^{2+}-SeMet CaM is solely a consequence of this less conservative approach to identifying disordered residues, as the final $2mF_o - DF_c$ electron-density maps for this structure and the prior 1.0 Å resolution Ca^{2+}-CaM structure are quite similar, with a CC of 0.94. In addition, the disordered residues common to both 1exr and Ca^{2+}-SeMet CaM are similarly modeled, despite the fact that the disorder model in 1exr was not used as a reference or a starting point for the Ca^{2+}-SeMet CaM model. Moreover, the degree of overall atomic disorder between the two models is also similar, with an average ADP for all protein atoms of 17.4 $Å^2$ and an average protein ADP anisotropy of 0.388 for Ca^{2+}-SeMet CaM, which is comparable to the average protein ADP of 16.4 $Å^2$ and protein ADP anisotropy of 0.353 previously reported for 1exr. The consistencies between 1exr and our new model suggest functional conformational disorder, rather then idiosyncrasies resulting from cryocooling (Keedy *et al.*, 2014). Therefore, the increased number of modeled disordered residues in Ca^{2+}-SeMet CaM represents an improved model of conformational disorder compared with the conservative approach taken with 1exr.

As an objective measure of the degree of conformational disorder in these two proteins, *qFit*2.0 was used for automated alternate main-chain and side-chain conformation modeling based on $2mF_o - DF_c$ electron density (Keedy *et al.*, 2015). *qFit*2.0 identifies 138 disordered residues in Ca^{2+}-CaM (data from PDB entry 1exr) and 137 residues in Ca^{2+}-SeMet CaM, indicating that these two proteins are similarly disordered. The exceptionally high percentage of disordered residues identified by *qFit*2.0 (93%) is consistent with the highly flexible nature of Ca^{2+}-CaM, although some overestimation of disorder is also possible, as an earlier implementation of *qFit* had an estimated false-discovery rate of 5–10% (van den Bedem *et al.*, 2009).

Given the large number of disordered residues, it is unsurprising that a significant fraction are difficult to model. Experimental phase information has been shown to improve the identification and modeling of conformational disorder in other systems (Burling *et al.*, 1996; Brodersen *et al.*, 2000) and shows similar benefit in this system. An example is Arg94, which was identified as disordered in both the manually built and *qFit*2.0 models. The manual model features two conformations, but three conformations were identified by *qFit*2.0 based on automated sampling of the model-phased $2mF_o - DF_c$ electron density. The third *qFit*2.0 alternate

side-chain conformation, a rotamer about the χ_1 angle, is not visible in $2mF_o - DF_c$ electron density contoured at 0.9–0.7σ and was therefore missed by manual building into the same maps. However, inspection of the density-modified SAD-phased maps at 0.9σ shows electron density supporting the third conformation (Fig. 4*a*). Importantly, *qFit*2.0 was able to accurately model the disorder at Arg94 using only model phases, but the inclusion of less biased experimental phase information in the density-modified maps makes this case of disorder easier to identify and confidently model. $2mF_o - DF_c$ electron density supporting the third conformation of Arg94 becomes visible at ∼0.5σ, which is below the contour level that most crystallographers typically use for model construction. The utility of such low-contour electron-density map features in characterizing protein conformational heterogeneity has recently been demonstrated (Lang *et al.*, 2010, 2014; Fraser *et al.*, 2009). Additional examples where experimental phase information supports *qFit*2.0 modeling of side-chain disorder in low-contour electron-density features are Glu14, Val108 and Glu123 (Figs. 4*b*, 4*c* and 4*d*).

3.3. Previously unexplained electron-density features are consistent with MPD bound to Ca^{2+}-CaM in multiple locations

Inspection of the density-modified SAD-phased electron density revealed strong evidence for two molecules of MPD: one mediates a lattice contact involving helices I and III of the N-terminal domain and the other is located in the C-terminal hydrophobic binding pocket (Figs. 3, 5*a* and 5*b*). Both of these areas contained residual $mF_o - DF_c$ electron density in the prior structure of Ca^{2+}-CaM (PDB entry 1exr) suggestive of an unidentified bound species. Because this density was weak and thus their assignment was ambiguous, these features were conservatively but incorrectly modeled as water. In this new structure, the density-modified experimental electron density for these MPD molecules is highly consistent with the final $2F_o - DF_c$ map, enhancing the confidence that these two MPD molecules are correctly assigned and minimally influenced by model bias, which can be especially pernicious when modeling partially occupied or otherwise disordered ligands (Pozharski *et al.*, 2013).

Both MPD molecules are bound in regions that bias the two domains of Ca^{2+}-CaM towards more open EF-hand conformations (Fig. 3), likely contributing to the discrepancy between the crystal structure and the more closed solution structure of Ca^{2+}-CaM (Chou *et al.*, 2001; Wilson & Brunger, 2000). For the N-terminal MPD mediating the lattice contact

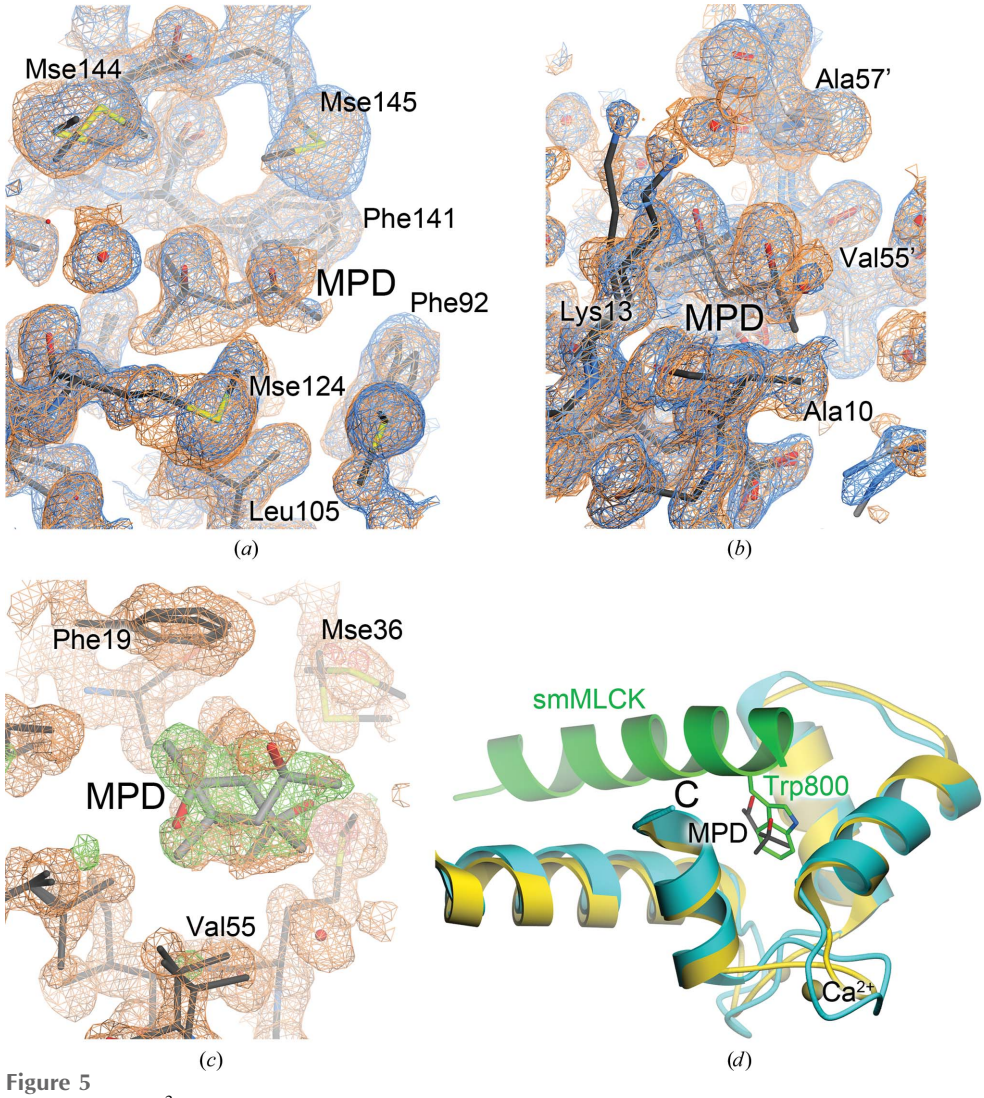


Figure 5
MPD binds to Ca^{2+}-CaM at multiple locations and overlaps with a critical target-binding site. (*a*) shows the density-modified, SAD-phased electron density contoured at 0.9σ (orange) and $2mF_o - DF_c$ electron density at 0.9σ (blue) around the MPD-binding site in the C-terminal hydrophobic binding pocket. Surrounding residues are labeled. (*b*) shows the MPD bound at the N-terminal lattice contact between residues in helix I (black lines) and helix III (background, light gray lines). In both panels, the experimental electron density unambiguously supports a bound MPD molecule. In (*c*) $mF_o - DF_c$ electron density at ±3.0σ (green, red) is shown with the density-modified map contoured at 0.9σ (orange) for a potential MPD site in the N-terminal domain. The disordered MPD molecule (gray lines) is broadly consistent with this density feature but was not included in the final model. In (*d*) the C-terminal lobe of the crystal structure of Ca^{2+}-CaM bound to a peptide from smooth muscle myosin light-chain kinase (smMLCK; PDB entry 1cdl) is shown in ribbon representation, with the peptide in green and CaM in yellow. The C-terminal domain of Ca^{2+}-SeMet CaM (blue) is superimposed, with MPD rendered as sticks. The MPD site overlaps closely with the position of the indole side chain of Trp800 in the smMLCK peptide.

between helices I and III, these intermolecular interactions would not be possible if these helices were shifted into the more closed conformation observed in solution (PDB entry 1j7o; Chou *et al.*, 2001) owing to steric conflicts between multiple residues. In particular, the 26° shift of helix I observed in the solution structure would place Glu14 into conflict with MPD and also disrupt other lattice contacts. Although it is impossible to formally separate the contributions of crystal lattice contacts and MPD binding to the more open conformation of the crystalline N-terminal lobe, we can conclude that MPD could not reside at this location if the more closed solution structure had been crystallized. A similar steric conflict is observed with MPD bound in the C-terminal hydrophobic pocket, where the shifted helix VIII in the solution structure (PDB entry 1j7p) would place residues Met144 and Met145 into steric conflict with the bound MPD. Therefore, binding of MPD both favors more open conformations of the two domains and facilitates triclinic lattice formation.

MPD has a chiral center at the C4 atom (secondary alcohol position) and racemic MPD was used as a precipitant to grow these crystals. Both molecules of bound MPD were modeled in their 4*S* configuration, which is consistent with the electron density. We cannot rule out, however, that both enantiomers might be present at both bound MPD sites. The electron density for MPD bound in the C-terminal pocket of Ca^{2+}-SeMet CaM could support either enantiomer reasonably well, as the density corresponding to the C5 atom of *S*-MPD is weak and could also agree with the O4 atom of *R*-MPD. The environment of this atom is hydrophobic and thus is less consistent with an O atom (in *R*-MPD) than with a C atom (in *S*-MPD), leading to the tentative assignment of *S*-MPD. We note that the C5 atom has the highest ADPs and lowest electron density of any atom in both MPD molecules, consistent with possible averaging over MPD enantiomers.

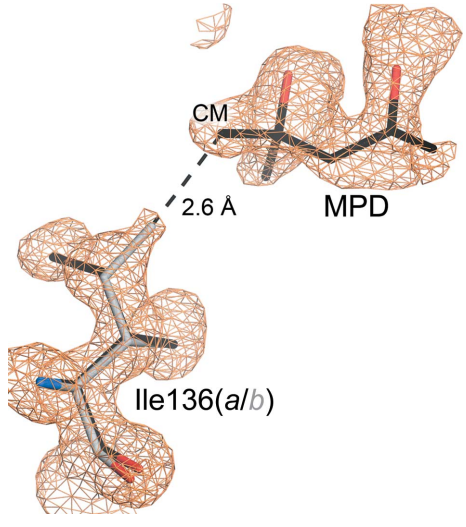


Figure 6
Binding of MPD in the C-terminal pocket is correlated with side-chain disorder at Ile136. The density-modified map contoured at 0.9σ (orange) is shown with the C-terminal MPD molecule and its nearby residue Ile136. Ile136 samples two conformations (black and gray) and one (gray, conformer *b*) sterically conflicts with the bound MPD as shown by the dashed line.

While the C-terminal hydrophobic pocket contains a clear MPD molecule, there is ambiguous but significant (3–6σ) $mF_o - DF_c$ electron density in the N-terminal pocket that overlaps partially with a feature in the density-modified SAD map (Fig. 5*c*). Although this electron density is difficult to definitely assign and was left unmodeled in the final structure, it is consistent with a disordered MPD molecule (Fig. 5*c*). These two conformations make chemically plausible contacts with surrounding residues and are in a highly hydrophobic local environment that would be expected to interact with less polar components of the mother liquor such as MPD.

3.4. MPD binds in a critical region of the C-terminal pocket and is correlated with side-chain disorder

The presence of bound MPD in the C-terminal binding pocket is intriguing since Ca^{2+}-CaM uses these surfaces to bind its many targets. One of the best-characterized examples is the complex between Ca^{2+}-CaM and a peptide derived from smooth muscle myosin light-chain kinase (smMLCK; PDB entry 1cdl; Meador *et al.*, 1992). Superimposing the structure of Ca^{2+}-CaM–smMLCK with Ca^{2+}-SeMet CaM shows that the MPD in the C-terminal binding pocket overlaps with Trp800 from smMLCK (Fig. 5*d*). Trp800 penetrates deeply into the hydrophobic binding pockets of Ca^{2+}-CaM and is a critical binding determinant for the interaction with smMLCK (Meador *et al.*, 1992; Bagchi *et al.*, 1992). Because this site would be partially occluded in the more closed conformations of the EF hands, bound MPD molecules are likely to bias the C-terminal (and possibly N-terminal) domains towards more open, target complex-like conformations in the triclinic crystal structure of Ca^{2+}-CaM.

The C-terminal lobe MPD molecule was refined to an occupancy of 0.52, which is similar to the refined occupancy of the nearer of the two alternate conformations of Ile136. This similarity is significant because the Cδ1 atom of the minor second conformation of Ile136 would clash (2.6 Å C–C atom separation) with the CM atom of MPD if both were fully occupied (Fig. 6). This provides an unequivocal example of side-chain disorder coupled to MPD binding, which imposes a correlation between protein conformational fluctuations and the occupancy of this binding pocket. Although not the only example of spatially correlated disorder in Ca^{2+}-CaM, the adventitiously bound MPD clearly illustrates the importance of Ca^{2+}-CaM conformational heterogeneity for binding to the hydrophobic pockets.

4. Discussion

The 1.0 Å resolution Ca^{2+}-SeMet CaM SAD data set represents one of the higher resolution examples of experimental phase information in the PDB. Because bimodal SAD phase probability distributions require improvement by density modification before optimal maps are obtained, this high-resolution SAD data set provides a potentially useful resource

for optimizing the performance of density-modification algorithms at atomic resolution. Our results show that *Parrot* was the best-performing algorithm tested, although we used default parameters for the tested programs and thus cannot rule out the possibility that other programs may perform better with different parameterization. Although the quality of all density-modified maps was comparable in well ordered regions, *Parrot*-calculated maps showed stronger and more easily interpreted density for disordered residues, which are of particular interest in this system. Experimental phase information is highly valuable when used to clarify these weak electron-density features, as experimentally phased maps are not subject to phase bias introduced from a structural model and thus can be used to critically evaluate the placement of alternative side-chain conformations and ligands into ambiguous electron density. We note that density modification represents the deliberate introduction of particular forms of phase bias into the map; however, this bias does not favor a particular structural model and is thus less problematic than traditional sources of model phase bias when interpreting difficult regions of electron density.

The value of experimental phase information for characterizing conformational heterogeneity has been recognized for some time (Brodersen *et al.*, 2000; Burling *et al.*, 1996). This analysis of Ca^{2+}-SeMet CaM confirms and extends a prior characterization of the extensive side-chain disorder present in crystalline Ca^{2+}-CaM based on model phases (Wilson & Brunger, 2000). Although much of the modeled disorder in PDB entry 1exr is supported by this new Ca^{2+}-SeMet CaM structure, less stringent criteria for introducing alternative conformations as well as the availability of experimental phase information allows a significant expansion of the disordered model. The manually built Ca^{2+}-SeMet CaM model contains ∼39% of the residues in alternate conformations, while the automated disorder-modeling approach implemented in *qFit*2.0 indicates that 93% of the residues sample multiple conformations. Considered in the light of recent results from diverse proteins (Deis *et al.*, 2014; Fraser *et al.*, 2009, 2011; Keedy *et al.*, 2014; Lang *et al.*, 2010, 2014), there is compelling evidence that crystalline proteins are more extensively disordered than previously thought and that traditional criteria for modeling disorder are probably too conservative. Being free of model bias, experimental phase information could aid in the development of new criteria for the detection and modeling of protein conformational disorder. Because experimental phases are observations and thus increase the observation-to-parameter ratio in refinement, they may be fruitfully used in ensemble-refinement approaches, where they could mitigate concerns about potential overfitting of the data (Burnley *et al.*, 2012; Burling & Brünger, 1994), although we did not explore this application here. Moreover, experimental phase information would be a particularly powerful way to characterize conformational heterogeneity when obtained from crystals at room temperature.

These SeMet Ca^{2+}-CaM SAD data have resolved a long-standing question about the proper assignment of ambiguous electron-density features in both hydrophobic binding pockets of Ca^{2+}-CaM, which we have shown to be MPD. The observation of bound MPD at multiple locations helps explain why the EF-hands of Ca^{2+}-CaM are more open in triclinic crystal structures than in solution, as MPD binds in a target-like fashion to at least the C-terminal binding pocket and probably the N-terminal pocket as well. In addition, an MPD molecule mediates a lattice contact that bridges helices I and III of neighboring N-terminal lobes, stabilizing more open conformations of this domain. Ca^{2+}-CaM is a difficult protein to crystallize, and nearly all prior structures have used MPD as a precipitant. Even in CaM structures determined in different space groups, electron-density features suggestive of bound MPD are present (Wilson & Brunger, 2003). Interestingly, one structure of Ca^{2+}-CaM that used PEG rather than MPD as a precipitant (PDB entry 1prw) shows the molecule in a dramatically different, partially closed conformation (Fallon & Quiocho, 2003). Furthermore, a structure of the Ca^{2+}-CaM N-terminal domain alone (PDB entry 3ikf) obtained using PEG as a precipitant shows residues from the central linker region bound in the N-terminal pocket of neighboring molecules, similar to the location of bound MPD in the triclinic structure (O'Donnell *et al.*, 2009). Therefore, we propose that MPD acts as both ligand and precipitant in many crystal structures of Ca^{2+}-CaM, and that the presence of MPD as an adventitious CaM ligand needs to be incorporated into future interpretations of this well known structure.

Acknowledgements

We thank Dr Luke Rice (UT Southwestern) for assistance with data collection and Dr Paul Adams (Lawrence Berkeley National Laboratory) for advice about data analysis. Use of the Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Contract No. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393).

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Adams, P. D. *et al.* (2011). *Methods*, **55**, 94–106.
Afonine, P. V., Grosse-Kunstleve, R. W., Chen, V. B., Headd, J. J., Moriarty, N. W., Richardson, J. S., Richardson, D. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2010). *J. Appl. Cryst.* **43**, 669–676.
Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.
Babu, Y. S., Bugg, C. E. & Cook, W. J. (1988). *J. Mol. Biol.* **204**, 191–204.
Babu, Y. S., Sack, J. S., Greenhough, T. J., Bugg, C. E., Means, A. R. & Cook, W. J. (1985). *Nature (London)*, **315**, 37–40.
Bagchi, I. C., Huang, Q. H. & Means, A. R. (1992). *J. Biol. Chem.* **267**, 3024–3029.
Bedem, H. van den, Dhanik, A., Latombe, J.-C. & Deacon, A. M. (2009). *Acta Cryst.* D**65**, 1107–1117.
Brodersen, D. E., de La Fortelle, E., Vonrhein, C., Bricogne, G., Nyborg, J. & Kjeldgaard, M. (2000). *Acta Cryst.* D**56**, 431–441.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

Burling, F. T. & Brünger, A. T. (1994). *Isr. J. Chem.* **34**, 165–175.
Burling, F. T., Weis, W. I., Flaherty, K. M. & Brünger, A. T. (1996). *Science*, **271**, 72–77.
Burnley, B. T., Afonine, P. V., Adams, P. D. & Gros, P. (2012). *Elife*, **1**, e00311.
Chattopadhyaya, R., Meador, W. E., Means, A. R. & Quiocho, F. A. (1992). *J. Mol. Biol.* **228**, 1177–1192.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Chou, J. J., Li, S., Klee, C. B. & Bax, A. (2001). *Nature Struct. Biol.* **8**, 990–997.
Cowtan, K. (2000). *Acta Cryst.* D**56**, 1612–1621.
Cowtan, K. (2010). *Acta Cryst.* D**66**, 470–478.
Cowtan, K. D., Zhang, K. Y. J. & Main, P. (2001). *International Tables for Crystallography*, Vol. *F*, edited by E. Arnold & M. G. Rossman, ch. 25.2.5. Dordrecht: Kluwer Academic Publishers.
Deis, L. N., Pemble, C. W., Qi, Y., Hagarman, A., Richardson, D. C., Richardson, D. C. & Oas, T. G. (2014). *Structure*, **22**, 1467–1477.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Fallon, J. L., Baker, M. R., Xiong, L., Loy, R. E., Yang, G., Dirksen, R. T., Hamilton, S. L. & Quiocho, F. A. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 5135–5140.
Fallon, J. L. & Quiocho, F. A. (2003). *Structure*, **11**, 1303–1307.
Fenn, T. D., Ringe, D. & Petsko, G. A. (2003). *J. Appl. Cryst.* **36**, 944–947.
Franco-Echevarría, E., Baños-Sanz, J. I., Monterroso, B., Round, A., Sanz-Aparicio, J. & González, B. (2014). *Biochem. J.* **463**, 319–328.
Fraser, J. S., Clarkson, M. W., Degnan, S. C., Erion, R., Kern, D. & Alber, T. (2009). *Nature (London)*, **462**, 669–673.
Fraser, J. S., van den Bedem, H., Samelson, A. J., Lang, P. T., Holton, J. M., Echols, N. & Alber, T. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 16247–16252.
Ikura, M., Barbato, G., Klee, C. B. & Bax, A. (1992). *Cell Calcium*, **13**, 391–400.
Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B. & Bax, A. (1992). *Science*, **256**, 632–638.
Karplus, P. A. & Diederichs, K. (2012). *Science*, **336**, 1030–1033.
Keedy, D. A., Fraser, J. S. & van den Bedem, H. (2015). *PLoS Comput. Biol.* **11**, e1004507.
Keedy, D. A., van den Bedem, H., Sivak, D. A., Petsko, G. A., Ringe, D., Wilson, M. A. & Fraser, J. S. (2014). *Structure*, **22**, 899–910.
Kuboniwa, H., Tjandra, N., Grzesiek, S., Ren, H., Klee, C. B. & Bax, A. (1995). *Nature Struct. Mol. Biol.* **2**, 768–776.
Kuntz, D. A., Ghavami, A., Johnston, B. D., Pinto, B. M. & Rose, D. R. (2005). *Tetrahedron Asymmetry*, **16**, 25–32.
Lang, P. T., Holton, J. M., Fraser, J. S. & Alber, T. (2014). *Proc. Natl Acad. Sci. USA*, **111**, 237–242.
Lang, P. T., Ng, H.-L., Fraser, J. S., Corn, J. E., Echols, N., Sales, M., Holton, J. M. & Alber, T. (2010). *Protein Sci.* **19**, 1420–1431.
Lau, S.-Y., Procko, E. & Gaudet, R. (2012). *J. Gen. Physiol.* **140**, 541–555.
Meador, W. E., Means, A. R. & Quiocho, F. A. (1992). *Science*, **257**, 1251–1255.
Meador, W. E., Means, A. R. & Quiocho, F. A. (1993). *Science*, **262**, 1718–1721.
O'Donnell, S. E., Newman, R. A., Witt, T. J., Hultman, R., Froehlig, J. R., Christensen, A. P. & Shea, M. A. (2009). *Methods Enzymol.* **466**, 503–526.
O'Neil, K. T. & DeGrado, W. F. (1990). *Trends Biochem. Sci.* **15**, 59–64.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pozharski, E., Weichenberger, C. X. & Rupp, B. (2013). *Acta Cryst.* D**69**, 150–167.
Read, R. J. & McCoy, A. J. (2011). *Acta Cryst.* D**67**, 338–344.
Sarhan, M. F., Tung, C.-C., Van Petegem, F. & Ahern, C. A. (2012). *Proc. Natl Acad. Sci. USA*, **109**, 3558–3563.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Terwilliger, T. (2004). *J. Synchrotron Rad.* **11**, 49–52.
Wilson, M. A. & Brunger, A. T. (2000). *J. Mol. Biol.* **301**, 1237–1256.
Wilson, M. A. & Brunger, A. T. (2003). *Acta Cryst.* D**59**, 1782–1792.
Winn, M. D. (2001). *CCP4 Newsl. Protein Crystallogr.* **39**, contribution 3.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Yamauchi, E., Nakatsu, T., Matsubara, M., Kato, H. & Taniguchi, H. (2003). *Nature Struct. Biol.* **10**, 226–231.
Zhang, K. Y. J., Cowtan, K. & Main, P. (1997). *Methods Enzymol.* **277**, 53–64.
Zucker, F., Champ, P. C. & Merritt, E. A. (2010). *Acta Cryst.* D**66**, 889–900.
Zwart, P. H. (2005). *Acta Cryst.* D**61**, 1437–1448.