doi:10.1006/jmbi.2000.4029 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2000) **301**, 1237–1256

JMB

# The 1.0 Å Crystal Structure of $Ca^{2+}$-bound Calmodulin: an Analysis of Disorder and Implications for Functionally Relevant Plasticity

**Mark A. Wilson and Axel T. Brunger***

*The Howard Hughes Medical Institute and, Departments of Molecular and Cellular Physiology, Neurology and Neurological Sciences and Stanford Synchrotron Radiation Laboratory, Stanford University Stanford, CA 94305, USA*

**Corresponding author*

Calmodulin (CaM) is a highly conserved 17 kDa eukaryotic protein that can bind specifically to over 100 protein targets in response to a $Ca^{2+}$ signal. $Ca^{2+}$-CaM requires a considerable degree of structural plasticity to accomplish this physiological role; however, the nature and extent of this plasticity remain poorly characterized. Here, we present the 1.0 Å crystal structure of *Paramecium tetraurelia* $Ca^{2+}$-CaM, including 36 discretely disordered residues and a fifth $Ca^{2+}$ that mediates a crystal contact. The 36 discretely disordered residues are located primarily in the central helix and the two hydrophobic binding pockets, and reveal correlated side-chain disorder that may assist target-specific deformation of the binding pockets. Evidence of domain displacements and discrete backbone disorder is provided by translation-libration-screw (TLS) analysis and multi-conformer models of protein disorder, respectively. In total, the evidence for disorder at every accessible length-scale in $Ca^{2+}$-CaM suggests that the protein occupies a large number of hierarchically arranged conformational substates in the crystalline environment and may sample a quasi-continuous spectrum of conformations in solution. Therefore, we propose that the functionally distinct forms of CaM are less structurally distinct than previously believed, and that the different activities of CaM in response to $Ca^{2+}$ may result primarily from $Ca^{2+}$-mediated alterations in the dynamics of the protein.

© 2000 Academic Press

*Keywords:* calmodulin; X-ray crystallography; *Paramecium tetraurelia*; disorder; atomic resolution

## Introduction

Calmodulin (CaM) is an abundant and highly conserved eukaryotic protein that acts as the primary regulator of intracellular $Ca^{2+}$ signaling. Elevations in intracellular $Ca^{2+}$ levels result in the formation of a $Ca^{2+}$-CaM complex that can interact with and modulate the activity of a large number of proteins, including adenylate cyclase, phosphodiesterases, $Ca^{2+}$ pumps, and an assortment of kinases and phosphatases (Kink *et al.*, 1990). CaM is thus responsible for mediating the initial cellular response to a $Ca^{2+}$ signal.

CaM binds $Ca^{2+}$ with four EF-hand motifs, each of which coordinates a single $Ca^{2+}$ with micromolar affinity. Upon binding $Ca^{2+}$, the relative orientation of the two α-helices that define the EF-hand changes substantially, resulting in a transition from the closed to open conformation of the motif. Nuclear magnetic resonance (NMR) studies have shown that this $Ca^{2+}$-mediated conformational change is associated with complex, multibarrier dynamics that span a six order of magnitude timescale (Evenas *et al.*, 1999). Additional evidence of ''substate-hopping'' dynamics in the EF-hands has been provided by fluorescence lifetime studies of terbium ($Tb^{3+}$)-substituted CaM that showed all four $Ca^{2+}$-binding sites in the protein sample many distinguishable conformations (Austin *et al.*, 1987). These studies reveal that $Ca^{2+}$ binding alters both the structure and the conformational dynamics of the EF-hands.

Structural studies have demonstrated that $Ca^{2+}$ activation of CaM is also accompanied by a global

Abbreviations used: ADP, anisotropic displacement parameter; CaM, calmodulin; ESD, estimated standard deviation; MPD, methyl-2,4-pentanediol; MSD, mean-squared deviation; PDB, Protein Data Bank; TLS, translation-libration-screw.

E-mail address of the corresponding author: axel.brunger@stanford.edu

0022-2836/00/051237–20 $35.00/0

© 2000 Academic Press

conformational change, whereby the compact $Ca^{2+}$-free (apo) form of CaM is converted to a more extended dumbbell-shaped molecule upon binding $Ca^{2+}$ (Zhang *et al*., 1995). The extended conformation of the protein, which is by far the most thoroughly studied, consists of two lobes separated by an eight-turn central α-helix. The $Ca^{2+}$-induced extension of CaM exposes two hydrophobic pockets, one per lobe, that bind target proteins. In addition, NMR and solution X-ray scattering results indicate that the central helix in the $Ca^{2+}$-loaded protein is flexible and allows considerable movement of the two lobes with respect to one another in solution (Barbato *et al*., 1992; Osawa *et al*., 1999a; Trewhella, 1992).

CaM undergoes an additional conformational change upon binding to target proteins in which the two hydrophobic pocket regions of CaM are brought near to one another and wrap around an amphiphilic α-helix in the target. In order to accommodate this large conformational change, the central helix unwinds in the middle and acts as a flexible linker that is capable of allowing the two lobes to adopt variable orientations when bound to the target helix (Ikura *et al*., 1992; Meador *et al*., 1992). This helical flexibility, combined with the inherent deformability of the hydrophobic pockets, is thought to account for the ability of CaM to interact with a variety of different targets in a sequence-independent fashion (Meador *et al*., 1993; Oneil & Degrado, 1990). More generally, the prominent role that conformational dynamics plays in both $Ca^{2+}$ binding and target recognition illustrates the importance of structural plasticity in every aspect of CaM function (Meador *et al*., 1993).

X-ray diffraction data provide a wealth of information concerning atomic displacements in the crystalline environment. Unfortunately, much of this information is not extracted from the data, due to the prevailing use of simplified models that assume an isotropic Gaussian distribution of atomic displacements from a single equilibrium position. Refinement of more detailed and accurate models of protein disorder has traditionally been prohibited by the unfavorable observable to parameter ratio that results from fitting a large number of adjustable parameters against a modest quantity of diffraction data. Fortunately, cryogenic data collection methods, synchrotron radiation sources, and improved methods of data processing have significantly increased the amount of diffraction data that can be obtained from macromolecular crystals. With higher-resolution data, more sophisticated models of protein disorder can be refined. For example, protein conformational substates, principally at the level of side-chain disorder, can be described explicitly by the introduction of alternative side-chain conformations into the structural model. In addition, for diffraction data that extend to atomic resolution (greater than about 1.2 Å), individual atomic anisotropic displacement parameters (ADPs) can also be refined. The ADP is a second-rank tensor that describes the magnitude and directionality of disorder for every atom in the protein (Willis & Pryor, 1975). The directional information contained within the ADP model is especially valuable, since the preferred directions of atomic displacements can potentially be related to structural and hence possible functional aspects of the protein.

Relating the directional information contained within the atomic ADPs to larger-scale displacements within a macromolecule requires the use of models that impose correlations on the displacements of groups of atoms. The translation-libration-screw (TLS) model is one such approach, in which a set of atoms is defined as comprising a rigid group and then harmonic displacements about the rigid body degrees of freedom (and correlations amongst these displacements) for this group are fit to the diffraction data (Schomaker & Trueblood, 1968). Since only translations and rotations about a fixed point (librations) are allowed for a rigid group, the TLS model contains far fewer parameters than an atomic ADP treatment (20 parameters per rigid group in the most general case). This economy of parameters allows the TLS model to be refined against lower-resolution data than is required for the refinement of individual atomic ADP models (Howlin *et al*., 1989). Furthermore, because the rigid group(s) can be of any size, the TLS model has been applied to characterize the disorder in rigid groups ranging from aromatic side-chains to whole unit cells, and can be of particular value in investigating domain disorder in proteins (Howlin *et al*., 1989; Moss *et al*., 1996; Schneider, 1996). It should be borne in mind that larger groups of atoms (i.e. secondary structural elements, domains and whole proteins) are, at best, only approximately rigid, and the application of the TLS model to analyze the disorder of such groups will always overestimate the contribution of rigid body motion (Howlin *et al*., 1989; Schneider, 1996).

Despite high-information content, the ADP representation is limited, in that it provides only a Gaussian (harmonic) model of protein disorder (Willis & Pryor, 1975). Many studies have shown that proteins populate a large number of discrete conformational substates in the crystalline environment, and the Gaussian assumption used in the ADP model does not satisfactorily describe the multimodal nature of this type of disorder (Burling *et al*., 1996; Garcia *et al*., 1997; Ichiye & Karplus, 1988; Rejto & Freer, 1996; Smith *et al*., 1986). When electron density maps provide clear evidence of multimodal disorder, the affected regions of the macromolecular model are typically built into multiple conformations to account for the observed features in the electron density map (Sheldrick & Schneider, 1997). This explicitly accounts for the most obvious cases of anharmonic disorder, but requires direct intervention by the model builder and is not applicable to more ambiguously disordered portions of the protein. An alternative description of atomic fluctuations that includes a

non-Gaussian (anharmonic) contribution for all atoms in the model can be achieved by expanding the crystallographic model to include several independent copies of the protein. The relative displacements of the individual conformers provide an ensemble description of protein disorder that contains far fewer assumptions than the ADP model and thus is capable of describing both anisotropic and anharmonic displacements with no direct intervention from the model builder. This method, called multiple conformer refinement, is a generalization of local modeling of discrete disorder to the entire macromolecule, and has been shown to provide significantly better agreement with diffraction data than traditional single conformer isotropic displacement models, even when the diffraction data do not extend to atomic resolution (Burling & Brunger, 1994; Kuriyan *et al.*, 1991).

The 1.0 Å structure of $Ca^{2+}$-CaM presented here is the highest-resolution structure of CaM yet obtained and the first that allows the inclusion of ADPs and alternative side-chain conformations in the structural model. CaM exhibits disorder on a wide range of length-scales, suggesting that $Ca^{2+}$-CaM occupies a very large number of hierarchically arranged conformational substates in the crystalline environment. The extensive disorder in the protein suggests that $Ca^{2+}$-CaM samples a very large volume of conformation space in solution, and thus may not exist in a single, well-defined structure as previously believed. Lastly, the combination of atomic-resolution diffraction data and pronounced disorder provides a unique opportunity to test various crystallographic models of protein disorder in a well-characterized and biologically important model system.

## Results and Discussion

### Global structural features

Structures of the $Ca^{2+}$-CaM complex from several different organisms have been described elsewhere (Babu *et al.*, 1985; Chattopadhyaya *et al.*, 1992). In brief, $Ca^{2+}$-CaM is an extended, dumbbell-shaped molecule of the approximate dimensions 45 Å × 45 Å × 65 Å (Figure 1(a)). The molecule consists of two roughly globular domains separated by a 28 residue central helix region (helix D/E). The two lobes are structurally similar and are situated *in trans* about the axis of the central helix. Each lobe is comprised of two EF-hand motifs that are spatially related to one another by a pseudo dyad symmetry axis. The helices of two EF-hands in each lobe define a phenylalanine and methionine-rich hydrophobic pocket that is exposed to solvent and involved in target binding. In addition, the EF-hand motifs each contain a hepta-coordinated $Ca^{2+}$, where six ligands are provided by acidic residues in the protein and the seventh ligand is a water molecule.

Although the structure determined in this study is similar to previously solved structures of $Ca^{2+}$-CaM, it has a surprisingly large overall $C^{\alpha}$ root-mean-square deviation (RMSD) of 0.427 Å from the 1.68 Å *Paramecium tetraurelia* CaM model (Ban *et al.*, 1994). Nearly all of this deviation is due to differences in the central helix region, as the individual lobes are nearly identical between the two structures, with a $C^{\alpha}$ RMSD of 0.191 Å for the N-terminal lobes and 0.199 Å for the C-terminal lobes. In contrast, the central helix in the present structure is smoothly bent with a radius of curvature of 35.4 Å, which is significantly smaller than the 40.5 Å radius of curvature found in the 1.68 Å study and results in a more compact protein than seen in the previous structure (see Materials and Methods). This change in helical curvature results in a $C^{\alpha}$ RMSD of 0.330 Å for the central helices in two models. This RMSD is substantially larger than both the estimated coordinate error in this region (0.07 Å) and the RMSD observed for either of the two lobes (see above), confirming the presence of a significant increase in helical bend in the 1.0 Å structure. A likely reason for this increased helical bend is the lower temperature at which the data were collected in this study (100 K *versus* 263 K in the previous work), demonstrating that the central helix region of the protein is flexible and capable of significant temperature-dependent deformation in the crystalline environment.

### Fifth calcium site

A major difference between this and all previously reported $Ca^{2+}$-CaM structures is the presence of a fifth $Ca^{2+}$, which mediates a crystal contact between residues Glu47 and Asp58 of translationally related molecules. The $Ca^{2+}$ has well-defined $F_o - F_c$ difference electron density at 20 σ, clear anomalous difference density at 15 σ, and exhibits octahedral coordination geometry (Figure 2(a)). Electron density for this site was present in the 1.68 Å structure; however, it was interpreted as a water molecule and refined to a *B*-factor of 14 $Å^2$ (Ban *et al.*, 1994). This *B*-factor is reasonable for a well-ordered water molecule and may explain why the site was not identified as a $Ca^{2+}$ in previous studies. The density can be unambiguously identified as a $Ca^{2+}$ in this work, due to the presence of strong anomalous difference density around this site (Figure 2(a)), which rules out the possibility that the site is occupied by any other component of the mother liquor.

$Ca^{2+}$-CaM is known to be monomeric in solution, therefore the two residues observed to apically coordinate this fifth $Ca^{2+}$ in the crystal are almost certainly not close enough in solution to bind $Ca^{2+}$. In the absence of a neighboring molecule, this $Ca^{2+}$ would be solvent exposed, making only one direct contact with the protein, and therefore the physiological relevance of this fifth site is highly dubious. This site is most likely a consequence of the close packing of the crystalline environment and therefore of no biological importance.

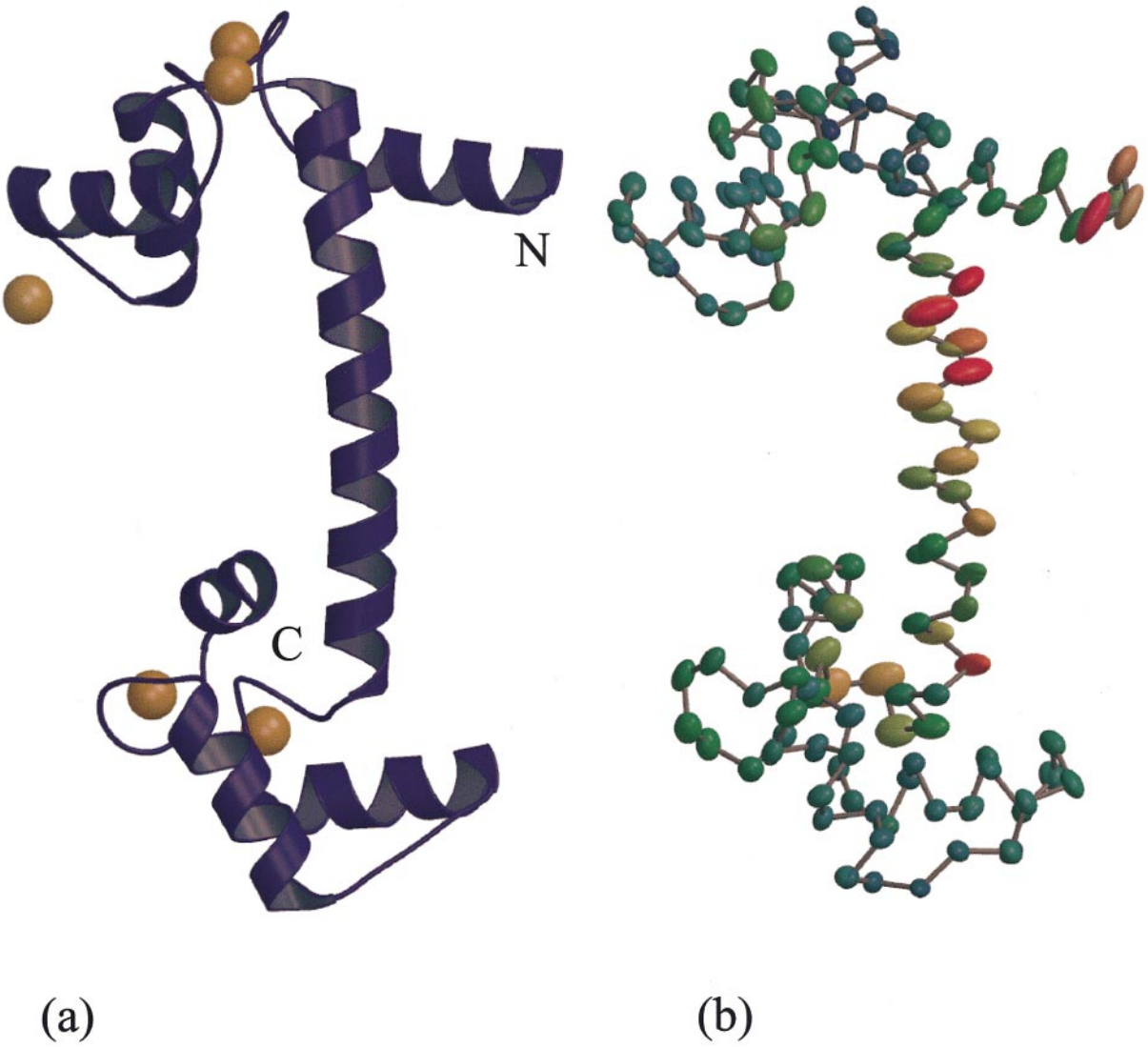


**Figure 1.** Ribbon and ADP representations of CaM. (a) A ribbon representation of CaM, with $Ca^{2+}$ represented as yellow spheres. The protein adopts the extended dumbell conformation seen in previous studies. (b) A $C^{\alpha}$ trace of CaM, with thermal ellipsoids for the $C^{\alpha}$ atoms represented at the 90% probability level. The ellipsoids are colored according to $U_{eq}$, ranging from blue (lowest $U_{eq}$) to red (highest $U_{eq}$). The largest values of $U_{eq}$ are found in the central helix and the hydrophobic binding pockets, as well as the two termini. Panel (a) was made with MOLSCRIPT (Kraulis, 1991) and RENDER (Merritt & Murphy, 1994). Figure (b) made with MOLSCRIPT and RASTEP (Merritt & Murphy, 1994).

## Discrete disorder in calmodulin

The 1.0 Å model contains 36 residues in alternative conformations, which represents 24% of the entire protein. By comparison, most atomic-resolution protein structures contain between 6 and 15% of residues in alternative conformations, indicating that $Ca^{2+}$-CaM is more extensively disordered than other proteins that diffract to comparable resolution (Smith *et al.*, 1986). In all instances of discrete disorder, multiple conformations were assigned only when the $\sigma_A$-weighted $2F_o - F_c$ electron density at a contour level of 1.3 σ was clearly not consistent with a single conformation and there existed density for placement of an alternative conformation in $\sigma_A$-weighted $F_o - F_c$ maps contoured at 3 σ. In addition, only two conformations were modeled for residues that displayed evidence of discrete disorder. These criteria are rather stringent for this system: there exist many residues where the electron density indicated multimodal disorder, although the evidence for this disorder was apparent only at a lower contour level or the placement of alternative conformations was ambiguous. While these more ambiguously disordered residues were not modeled in multiple conformations, they confirm that CaM samples many discrete conformational substates in the crystal and that this disorder is distributed over much of the protein. It is also important to note that,

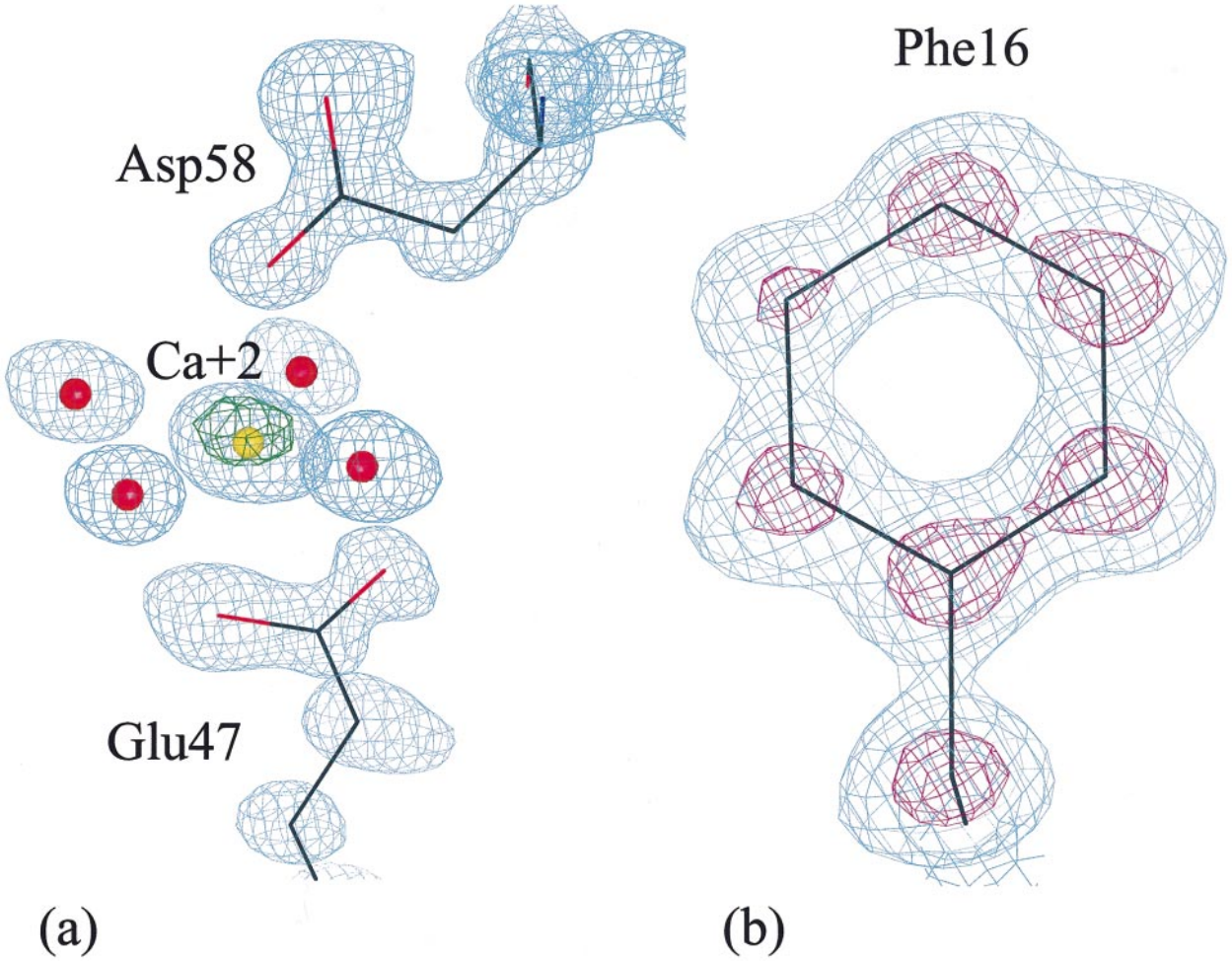


**Figure 2.** Two views of the final $\sigma_A$-weighted $2F_o - F_c$ electron density. The blue electron density is contoured at 1.8 σ, magenta density is contoured at 4 σ. (a) A view of the fifth $Ca^{2+}$ and its environment. The red spheres are water molecules and the yellow sphere represents the $Ca^{2+}$. The green density around the $Ca^{+2}$ is anomalous difference electron density ($F_{hkl+} - F_{hkl-}$) contoured at 5 σ, corroborating the assignment of this density as a $Ca^{2+}$. The coordination geometry for this $Ca^{2+}$ is octahedral, with the apical ligands provided by acidic side-chains on translationally related CaM molecules. (b) The electron density for Phe16 illustrates the quality of the electron density for CaM in well-ordered regions. At 4 σ (magenta), the electron density for individual carbon atoms is well-resolved and atomic. This Figure was generated by OPLOT (Jones *et al.*, 1991).

since the data were collected at 100 K, thermal motion contributes little to the observed displacements. Therefore, frozen conformational substates are responsible for the vast majority of the disorder even where well-resolved alternative conformations cannot be discerned.

The central helix contains ten residues in alternative conformations, with the remainder concentrated primarily in the two hydrophobic binding pockets. The distribution of discrete disorder is informative, since structural plasticity in these regions of the protein is suspected to play an important role in the function of CaM (see Introduction). The electron density for the central helix is of generally poorer quality than the rest of the molecule, and reveals evidence of appreciable disorder for every residue in this region. In addition, several of the discretely disordered residues in $Ca^{2+}$-CaM, including Ser70, Met72, Ser81, Glu87 and Val85 in the central helix, have alternative conformations that flank the single conformation in the previously reported 1.68 Å model (Ban *et al.*, 1994). The existence of flanking alternative conformations indicates that the previous 1.68 Å structure showed these residues in an average conformation that can be resolved into two discrete conformational substates in the present work (Kuriyan *et al.*, 1986). The extensive disorder in this region is not surprising, since the central helix is largely exposed to solvent, making few contacts with other regions of the protein or with translationally related molecules. The scarcity of crystal contacts in this region may be a consequence rather than a cause of this conformational disorder, since the large entropic penalty resulting from lattice constraints on an inherently flexible region of the protein would oppose crystallization of the molecule. In addition, NMR results have demonstrated

that the central helix is poorly structured in solution, and that residues Asp78 through Ser81 sample non-helical conformations (Barbato *et al.*, 1992).

The two hydrophobic pocket regions show dramatic evidence of disorder, with 16 residues in alternative conformations and several others in highly anisotropic or multimodal density. Examples of this disorder are shown in Figure 3(a) and (b). Since the hydrophobic pockets make direct contact with target proteins, it has been speculated that the abundance of phenylalanine and methionine in these regions presents a deformable and polarizable surface that can adapt to the detailed structural and electrostatic features of various targets (Oneil & Degrado, 1990). The discrete phenylalanine disorder (residues 12, 68, 92 and 141) is best modeled by a combination of ring rotation and translation, and the density for these side-chains indicates that more alternative conformations may be justified. Surprisingly, although these side-chains are located predominantly in the interior of the hydrophobic pockets, they are subject to only modest packing interactions, as reflected in the mean occluded surface ratio of 0.49, as determined by the program OS (Pattabiraman *et al.*, 1995). The discretely disordered methionine residues 36, 72, 76 and 144 extend into the solvent-exposed regions around the binding pockets and are consequently quite loosely packed (mean occluded surface ratio, 0.38). Both the phenylalanine and methionine packing values are near the average mean occluded surface ratio of 0.42 found for all 16 of the disordered residues in the hydrophobic pockets. For all the discretely disordered residues in the hydrophobic pockets, most of the packing contacts are with other disordered residues (Figure 4). This is especially noteworthy for the moderately packed residues in the core of the pockets: the combination of discrete disorder and packing constraints on atomic displacements requires spatially correlated side-chain displacements in these regions of the protein in order to avoid close contacts between neighboring disordered residues. Thus, the moderate degree of overall packing in the hydrophobic binding pockets may be evolutionarily optimized to achieve a large number of conformational substates while simultaneously allowing correlated side-chain motion. The partially collective character of these motions would allow efficient sampling of the side-chain conformational substates seen in the present work and thus facilitate target-specific structural deformation of the binding surfaces in

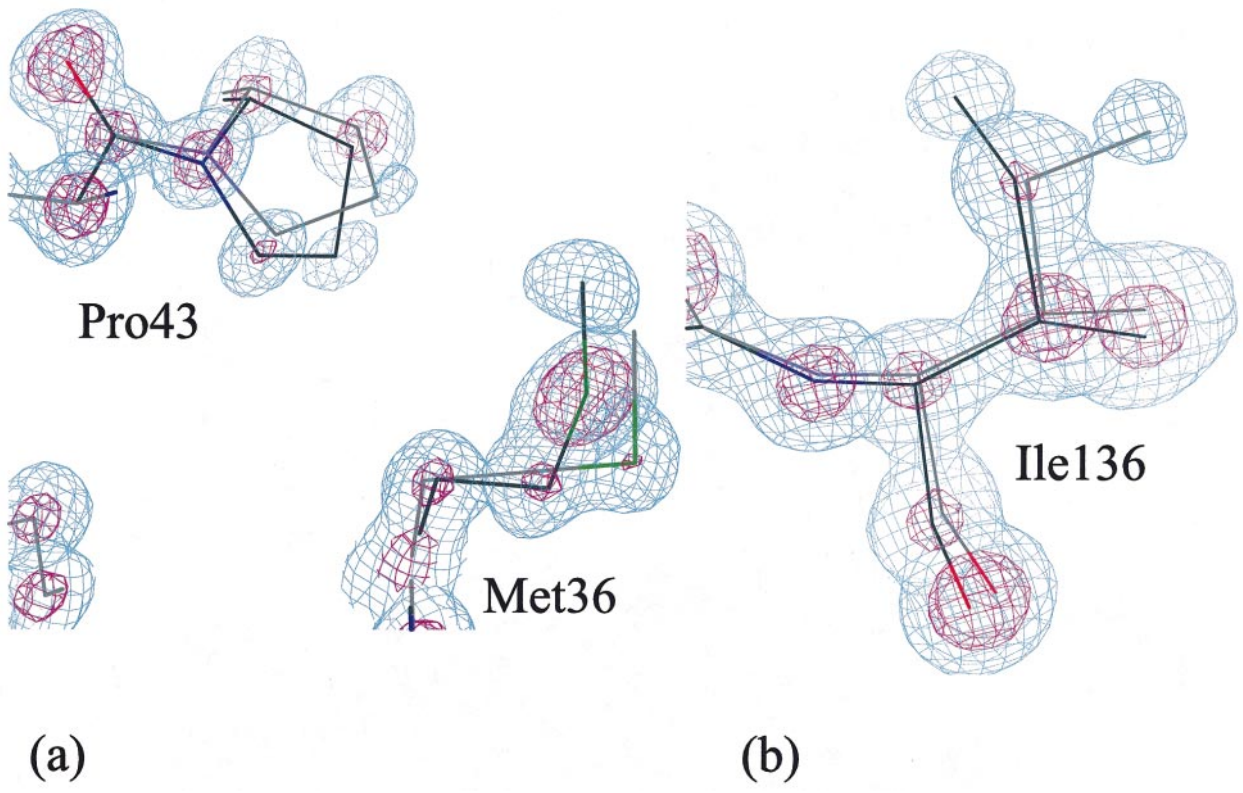


**Figure 3.** Two views of discrete side-chain disorder in the hydrophobic binding pockets of CaM. In (a) and (b), the major conformation is represented with a darker line. (a) Pro43 and Met36 illustrate the clustering of discretely disordered residues that is typical in both hydrophobic binding pockets (here the N-terminal binding pocket). (b) An example of clearly resolved discrete disorder for Ile136 in the C-terminal binding pocket. Approximately 25% of the 36 disordered residues in CaM show comparably well-resolved disorder. This Figure was generated by OPLOT (Jones *et al.*, 1991).

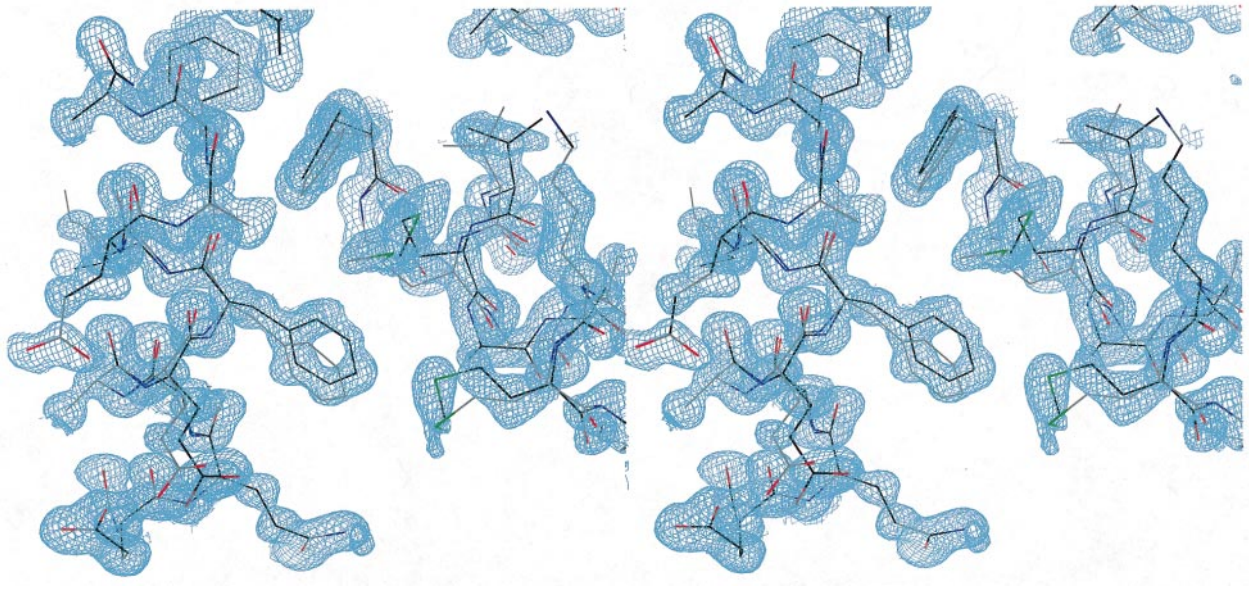

**Figure 4.** A stereo view of clustered discrete disorder in the N-terminal hydrophobic binding pocket. The electron density is contoured at 1.5 σ. The local environments of these hydrophobic residues are dominated by other discrete disordered residues, suggesting that discrete disorder for residues in this region of CaM (as well as the C-terminal hydrophobic binding pocket) is correlated. Here, spatially correlated side-chain disorder is required in order to avoid unreasonably close (3 Å or less) non-bonded contacts between atoms in the major conformation of these residues (represented with a darker line) and atoms in the lower occupancy conformations of surrounding residues. This Figure was generated by OPLOT (Jones *et al.*, 1991).

CaM. Therefore, we propose that correlated disorder in the hydrophobic binding pockets allows CaM to interact with structurally diverse targets by a dynamically driven selection of conformational substates that optimize surface complementarity with targets.

The remaining residues in alternative conformations are either polar or charged and are distributed on the surface of the two lobes. These side-chains protrude into the solvent channels of the lattice, and do not generally participate in direct interactions with other protein atoms in the crystal. The hydrophilic character and exterior location of these residues indicate that they are probably involved primarily in the solvation of the molecule and therefore this disorder is of limited functional relevance.

## The ADP model of $Ca^{2+}$-calmodulin disorder

The ADP model of CaM disorder represents a dramatic improvement over the isotropic model, reducing the $R_{free}$ from 26.23 % ($R = 24.56$ %) to 18.69 % ($R = 15.57$ %) in a model without riding hydrogen atoms or alternative conformations for the 36 discretely disordered residues. The surprisingly large 8 % drop in $R_{free}$ upon the inclusion of ADPs emphatically demonstrates the inadequacy of the isotropic displacement model for this system. As expected, the largest displacement parameters are found in the central helix region and in the two hydrophobic pockets and the smallest displacement parameters are found on the surfaces of the two lobes that participate in crystal contacts (Figure 1(b)). The thermal ellipsoids for the $C^{\alpha}$ atoms in the central helix region are oblate, with the shortest principal axis oriented roughly parallel with the helical axis, indicating that displacements of the central helix backbone are most constrained along the helical axis and least constrained in directions orthogonal to the helix axis. This result is consistent with both the geometry of the molecule and the absence of crystal contacts in the central helix region.

An unexpected feature of the ADP model is the pronounced anisotropy observed for the five $Ca^{2+}$. The mean anisotropy of these ions is 0.346 ($\sigma$=0.118), which deviates substantially from the value of unity expected for perfectly isotropic displacements and is also considerably more anisotropic than the $Ca^{2+}$ present in pike parvalbumin ($\langle A \rangle = 0.647$, $\sigma = 0.014$), which is the only other $Ca^{2+}$-containing EF-hand protein in the PDB that includes anisotropically refined $Ca^{+2}$ (Declercq *et al.*, 1999). However, the mean $Ca^{2+}$ anisotropy in $Ca^{2+}$-CaM is nearly identical to the mean anisotropy of the protein atoms ($\langle A \rangle = 0.353$, $\sigma = 0.125$), suggesting that the displacements of the $Ca^{2+}$ are coupled to overall protein disorder. Additional evidence in favor of this interpretation is provided by the observation that the orientations of the principal axes of the $Ca^{2+}$ ADPs correspond closely to those of nearby atoms, including ordered solvent. The fact that the $Ca^{2+}$, the surrounding protein atoms and the local ordered solvent all display similar ADP magnitudes and orientations suggests

that rigid-body lobe displacements may contribute significantly to the total molecular disorder.

This hypothesis was tested by determining the contribution of these postulated domain displacements to the ADP model using the translation-libration-screw (TLS) description of protein disorder (Schomaker & Trueblood, 1968). The protein was segmented into three domains (residues 2-70, 71-90, and 91-147), and TLS tensors for these three domains were refined individually. Since in reality these domains are not strictly rigid and some non-rigid body disorder will always be compatible with a rigid body description, the refined TLS tensors for each group absorb some of the contribution of smaller length-scale disorder. Therefore, the TLS model provides only an upper limit for the contribution of true rigid-body domain displacements to the observed disorder (Harata *et al*., 1998; Howlin *et al*., 1989; Kidera & Go, 1990; Schneider, 1996). Although many efforts have been made to correct for this effect, none of them is entirely effective at removing the contaminating disorder at smaller length-scales (Schneider, 1996). Therefore, as a control, the entire protein was treated as a single TLS group, so that only overall translations and librations of the molecule were refined. This whole-protein TLS model provides an upper limit for the extent to which overall rigid-body protein motion could account for the observed disorder. Thus, any additional improvement in the $R_{free}$ gained with the use of the segmented domain TLS model is taken as evidence for the existence of disorder that is incompatible with simple rigid-body protein displacements.

Both TLS models represent an improvement over the isotropic *B*-factor model; however, the segmented domain TLS refinement provided a lower $R_{free}$ than the whole-protein TLS model, indicating that the inclusion of relative domain displacements significantly improves the quality of the model (Table 1). The thermal ellipsoids generated from the segmented domain TLS model compare favorably with the individual atomic ADP representation of the disorder in CaM, with the most pronounced disagreements in the central helix region and the N terminus (Figures 1(b) and 5(b)). This result is expected, since the central helix and the termini are the regions least likely to be even approximately rigid on the domain length-scale. In addition, the anisotropy of the $C^{\alpha}$ atoms in the segmented domain TLS model is much closer to the individual atomic ADP model value than is the overall TLS model, further supporting the contention that relative domain displacements are more consistent with the observed magnitude of anisotropy than is rigid-body protein disorder (Table 1).

The principal axes of the libration tensors for the three domains are shown in Figure 5 and the magnitudes of both the translational and librational components of the TLS models are listed in Table 2. The translational tensor Eigenvalues are all approximately equal for each rigid group, indicating nearly isotropic translational disorder. In contrast, the libration axes shown in Figure 5 reveal that the librations of the three domains are of different magnitudes and preferred directions, and that they are all highly anisotropic. Such anisotropic domain librations correspond to relative lobe torsions and lobe closure motions, both of which represent functionally relevant motions of CaM and thus provide the first direct evidence for previously proposed modes of $Ca^{2+}$-CaM deformation in the crystalline environment (Meador *et al*., 1992, 1993; Oneil & Degrado, 1990). Similar lobe motions have been observed in a three nanosecond molecular dynamics simulation of solvated $Ca^{2+}$-CaM (Wriggers *et al*., 1998). In addition, a least-squares minimized superposition of the 1.68 Å model (Ban *et al*., 1994) and the present model display relative displacements between the two structures of precisely these types (data not shown). On a cautionary note, it must be emphasized that the domain displacements observed in the crystal are nearly two orders of magnitude smaller than those that occur in solution, and thus cannot be interpreted as a direct reflection of the solution behavior of the protein.

Two aspects of the TLS refinement deserve special attention. First, although the segmented domain TLS treatment is a much better model than

**Table 1.** Comparison of models of disorder

| Model | $R_{work}$ (%)[a] | $R_{free}$ (%)[a] | $\langle U_{eq} \rangle$ (Å$^2$)[b] | $\langle A_{C\alpha} \rangle$[c] |
|---|---|---|---|---|
| Single conf., isotropic | 24.56 | 26.23 | 0.186 | 1.000 |
| Single conf., anisotropic | 15.52 | 18.55 | 0.207 | 0.373 |
| SHELX final model | 13.45 | 16.35 | 0.205 | 0.373 |
| 8 conf., model A | 28.08 | 30.92 | 0.211 | 0.736 |
| 8 conf., model B | 22.32 | 24.61 | 0.205 | 0.678 |
| 8 conf., model C | 20.59 | 22.14 | 0.216 | 0.687 |
| TLS, rigid protein | 20.50 | 22.98 | 0.212 | 0.620 |
| TLS, rigid domains | 19.01 | 21.44 | 0.220 | 0.508 |

[a] Both $R_{work}$ and $R_{free}$ were calculated with no sigma cutoffs using the full resolution range of the data (50-1·0 Å). $R_{free}$ (Brunger, 1992) was calculated from a test set that consisted of 10% of the data (approximately 7800 reflections).

[b] $\langle U_{eq} \rangle$ is the mean-squared displacement for all protein atoms. For the multiconformer models, $\langle U_{eq} \rangle$ includes the contribution from both relative conformer displacements and the isotropic *B*-factors associated with each conformer.

[c] $\langle A_{C\alpha} \rangle$ is the mean anisotropy of the $C^{\alpha}$ atoms in the model. Anisotropy is defined as the ratio of the lengths of the shortest and longest principal axes of the atomic thermal ellipsoid and was determined using PARVATI (Merritt, 1999).

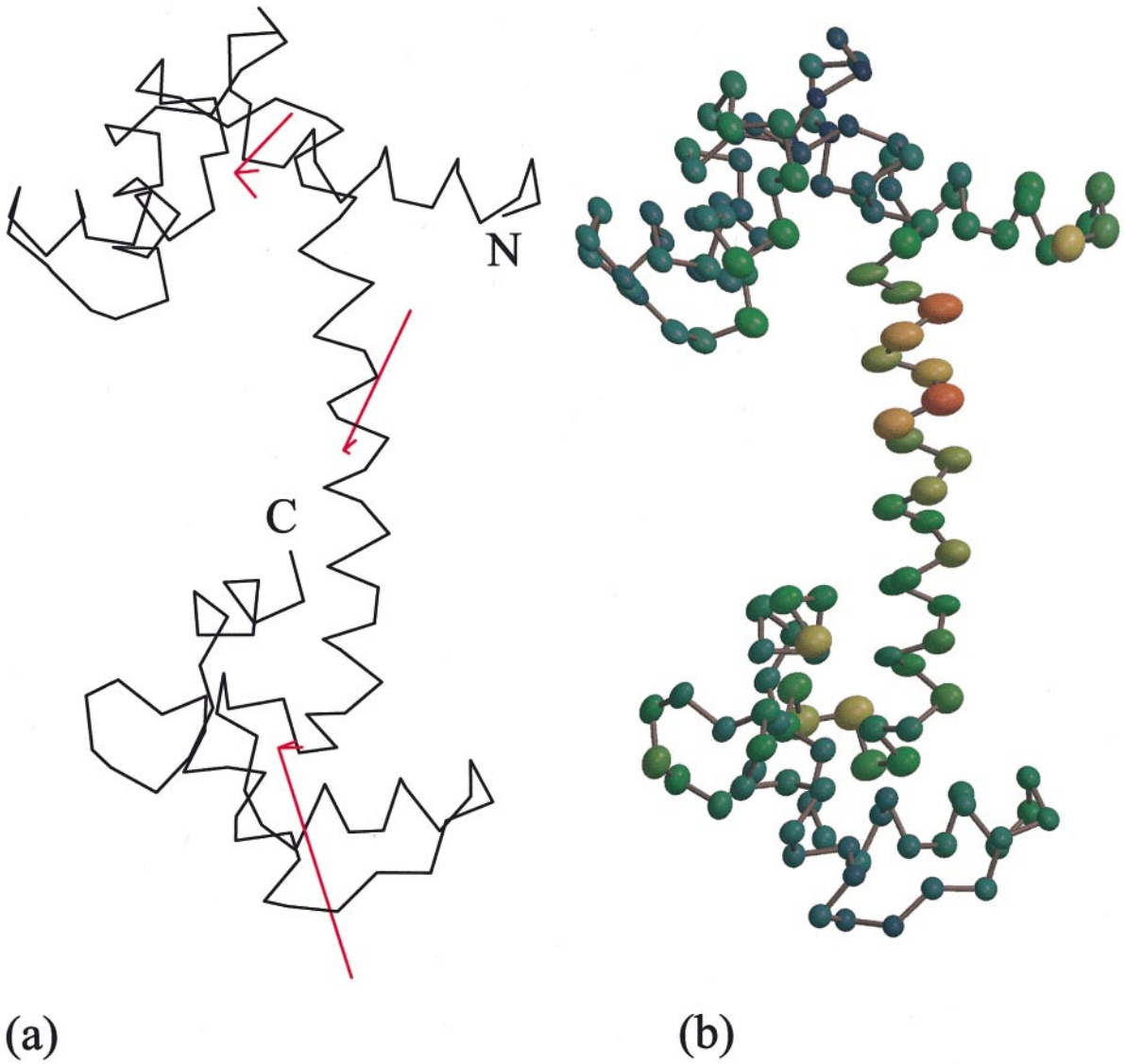


**Figure 5.** The segmented domain TLS model of CaM disorder. (a) The $C^{\alpha}$ trace of CaM (black), with the principal axes of the libration tensors from the segmented domain TLS refinement shown in red. The axes are scaled relative to the mean-squared libration amplitude about each axis and reveal the highly anisotropic character of the domain librations. Rotation of the domains about these axes would produce relative lobe torsion and lobe closure displacements. (b) The thermal ellipsoids represented at the 90% probability level resulting from the segmented domain TLS model. These ellipsoids compare well with those generated from the individual ADP refinement shown in Figure 1(b), suggesting that the gross features of the $C^{\alpha}$ anisotropy in CaM are consistent with the existence of relative domain displacements in the protein. Part (a) was generated with OPLOT (Jones *et al.*, 1991) and (b) was generated with MOLSCRIPT (Kraulis, 1991) and RASTEP (Merritt & Murphy, 1994).

the rigid protein TLS model, the segmented domain TLS model still exhibits substantially less anisotropy than the individual atomic ADP model (Table 1). This discrepancy indicates that a significant portion of the total CaM anisotropy is due to disorder at length-scales smaller than domain displacements. This conclusion is corroborated by a detailed analysis of diffuse X-ray scattering from crystals of a $Ca^{2+}$-CaM-peptide complex, which supported an anisotropic liquid-like model of CaM disorder with unusually small correlation length (Wall *et al.*, 1997). Secondly, TLS refinement is capable of describing only correlated amplitude disorder, not rigid-body disorder in the most rigorous sense of the term. The distinction arises from the inability of Bragg data to discriminate between true rigid body disorder, which involves in-phase motion of all atoms of the group, and motion of the atoms with different relative phases, which implies a more flexible structure. Since both possibilities give rise to identical magnitudes and directions of disorder, they are indistinguishable using conventional diffraction data. This caveat is attendant upon all TLS refinements.

**Table 2.** Tensor Eigenvalues from TLS refinement

| | Libration (deg.$^2$) | Translation (Å$^2$) |
|---|---|---|
| Whole protein | | |
| Eigenvalue 1 | 0.004 | 0.245 |
| Eigenvalue 2 | 0.071 | 0.135 |
| Eigenvalue 3 | 0.175 | 0.213 |
| N-term. lobe | | |
| Eigenvalue 1 | 0.472 | 0.169 |
| Eigenvalue 2 | 0.788 | 0.139 |
| Eigenvalue 3 | 0.309 | 0.214 |
| Central helix | | |
| Eigenvalue 1 | 1.670 | 0.388 |
| Eigenvalue 2 | 0.226 | 0.521 |
| Eigenvalue 3 | 0.054 | 0.342 |
| C-term. lobe | | |
| Eigenvalue 1 | 0.512 | 0.226 |
| Eigenvalue 2 | 0.180 | 0.162 |
| Eigenvalue 3 | 2.218 | 0.168 |

## The multiconformer model of $Ca^{2+}$-calmodulin disorder

A less restrictive alternative to the ADP description is a model that allows for both anisotropic and anharmonic atomic displacements. This can be achieved by simultaneously refining several independent (i.e. non-interacting) copies of the structural model against the data. Apart from stereochemical restraints on the individual structures, the multiconformer model contains no assumption concerning the nature of the relative displacements of the atoms in the model. Analysis of the displacements between individual structures in the final ensemble model therefore provides a detailed and minimally biased view of intramolecular crystalline disorder (Burling & Brunger, 1994; Kuriyan *et al.*, 1991; Wall *et al.*, 1997).

The multiconformer models used here consisted of copies of the entire single conformer $Ca^{2+}$-CaM model, including the ordered solvent and the five $Ca^{2+}$. This was done in order to allow direct comparison with the all-atom ADP model, although single conformer treatments of the ordered solvent model are typically employed in lower-resolution multiconformer refinements (Burling & Brunger, 1994; Rader & Agard, 1997). Each conformer was assigned a fractional occupancy equal to the reciprocal of the number of conformers used in the refinement and these occupancies were not subsequently refined. Trial refinements with two, four, six, eight, and ten copies were performed to determine the largest number of independent models that the data would support without overfitting as judged by the $R_{free}$ value. The $R_{free}$ decreased steadily upon introduction of additional models until ten models were used, at which point both the $R_{free}$ and the difference between the $R$ and $R_{free}$ values increased slightly. An increase in either of these quantities is an indication that the model has begun to overfit the data, and that fewer model parameters should be refined. Therefore, eight copies were judged as optimal for this system.

One contentious aspect of multiconformer refinement is the proper treatment of the atomic isotropic *B*-factors in the individual conformers (Burling & Brunger, 1994). In principle, all disorder should be directly described by the relative displacements of the conformers and thus the atomic *B*-factors should be set to zero. However, this approach has been found to result in both unreasonably high mobilities in the multiconformer model and an increase in the $R_{free}$ value relative to conventional single conformer refinement (unpublished results). In this work, the best results were obtained by using the refined single conformer isotropic *B*-factors as starting values and performing individual *B*-factor refinement after the multiconformer model was subjected to torsion angle simulated annealing and minimization refinements (multiconformer model C). This strategy produced a significantly lower $R_{free}$ value than either fixing all the *B*-factors at 10 Å$^2$ throughout the refinement (multiconformer model A) or by starting the refinement with 10 Å$^2$ *B*-factors and then refining them after annealing and minimization (multiconformer model B) (Table 1). These findings contradict previous conclusions drawn from a multiconformer refinement of α-lytic protease, which proved to be comparatively insensitive to the particular isotropic *B*-factor model employed (Rader & Agard, 1997). The α-lytic protease study used a different solvent model than the present work; however, it is likely that the importance of the isotropic *B*-factor model in multiconformer refinements is dictated principally by the degree and uniformity of the disorder in the macromolecule. Therefore, a well-ordered protein like α-lytic protease would be expected to exhibit less sensitivity to the details of the *B*-factor model than the extensively disordered $Ca^{2+}$-CaM. In addition, large regional variations in protein disorder further enhance the sensitivity of the multiconformer refinement to the isotropic *B*-factor model, as seen in a comparison of the three $Ca^{2+}$-CaM multiconformer models refined using different isotropic *B*-factor models (Figure 6). This comparison reveals that the greatest variability between the mean-squared deviations (MSD) calculated from the three multiconformer ensembles occurs in the central helix, confirming that multiconformer refinement is most sensitive to the details of the isotropic *B*-factor model in the most poorly ordered regions of the protein.

The multiconformer model successfully models residues in alternative conformations when the electron density for the affected residue is continuously spread among the alternative conformer choices. However, the multiconformer ensemble identifies only the dominant conformer in cases where there exists well-resolved density for the alternative conformations, as in Figure 3. Since the starting conformation of the multiconformer ensemble was the refined single conformer model in which all disordered residues are in their dominant conformation, it appears that the multiconformer refinement protocol is not able to drive the

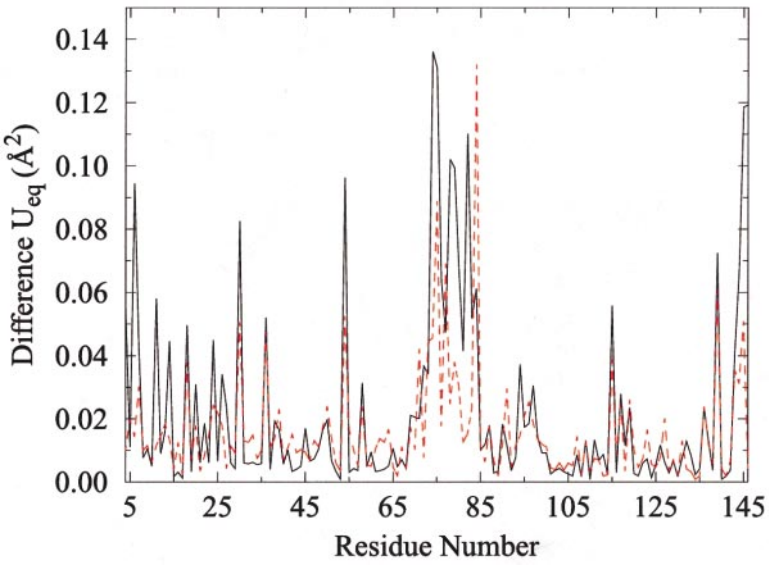


**Figure 6.** Differences in $U_{eq}$ amongst the three multiconformer models. The black trace shows the average difference between $U_{eq}$ calculated from the three different multiconformer models, where $U_{eq}$ includes only the contribution of the ensemble mean-squared deviation (MSD). The red dotted trace is the average difference between $U_{eq}$ of the three multiconformer models where $U_{eq}$ now includes both the ensemble MSD and isotropic *B*-factors. In both cases, the largest variability in the multiconformer representations of disorder is in the most disordered regions of the protein (residues 70-90 in particular), demonstrating that multiconformer refinement is most sensitive to the details of the isotropic *B*-factor model in the most disordered regions of the molecule.

ensemble to sample regions far away from the starting minimum for a limited number of disordered residues. This failing was unexpected and reveals an important limitation in the multiconformer approach as it is implemented in this work. Multiconformer refinement provides a sensitive means to diagnose and model more ambiguously disordered regions of the protein, but work remains to be done on improving ensemble sampling for areas exhibiting spatially well-resolved disorder. Recent work has shown that even when the amplitude data extend to atomic resolution, electron density maps made using experimental phase information can identify disordered regions of proteins that are not obviously disordered in electron density maps made using model phases alone (Brodersen *et al.*, 2000). Therefore, the inclusion of experimental phase information into the refinement target function may substantially improve the performance of multiconformer refinement in the problematic regions and is currently being investigated.

Inspection of the eight-conformer ensemble reveals obvious clustering of the models in several regions of the protein, especially in the hydrophobic pockets (Figure 7). Particularly striking is the fact that this clustering is as pronounced for the backbone atoms as it is for the side-chains, indicating that the disorder in the protein is distributed over a wide range of length-scales. This clustering provides direct evidence for the existence of multimodal electron density in these portions of the molecule and demonstrates that $Ca^{2+}$-CaM occupies several discrete backbone conformational substates in the crystal at cryogenic temperatures. In combination with the evidence for relative lobe displacements, discrete side-chain disorder, and atomic disorder, the clustering of both backbone and side-chain atoms reveals an intermediate level of structural disorder that strongly suggests that the organization of conformational substates in $Ca^{2+}$-CaM may be hierarchical, with many side-chain conformational substates nested within a single backbone conformational substate and several backbone conformational substates nested within the single average conformation seen in lower-resolution studies. A hierarchical distribution of disorder in CaM accounts for the wide range of length-scales over which the disorder is observed in this study, and is consistent with previously proposed hierarchical models of protein energy landscapes (Frauenfelder *et al.*, 1991). Although Bragg diffraction data are insensitive to correlations in disorder and thus can not conclusively distinguish between hierarchical and non-hierarchical arrangements of conformational substates, the increase in the number of substates as the length-scale decreases observed here is symptomatic of hierarchical disorder and is difficult to reconcile with alternative interpretations.

## Comparison of the ADP and multiconformer models

The best multiconformer model (model A) represents a significant improvement over the isotropic single conformer model, but does not agree with the data as well as the ADP single conformer

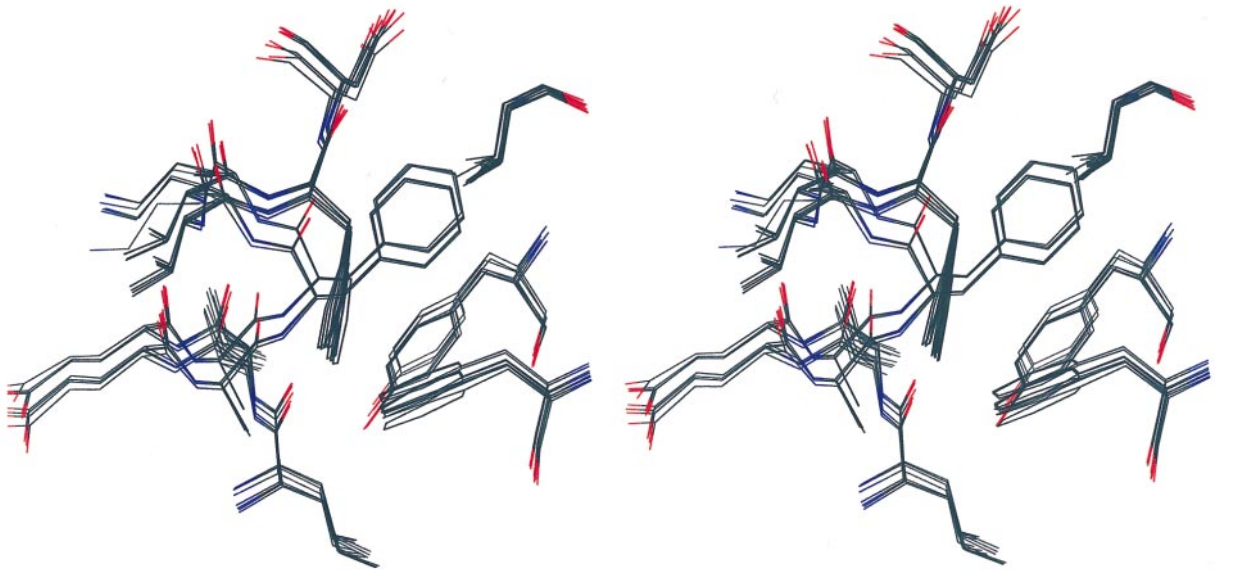

**Figure 7.** A stereo view of the C-terminal hydrophobic binding pocket in the eight-conformer model, showing clustering of the individual conformers into two conformational substates. Both backbone and side-chain atoms are clustered along the length of the α-helix on the left of the Figure, demonstrating that CaM occupies both discrete backbone and side-chain conformational substates at cryogenic temperatures in the crystalline environment. This intermediate level of disorder is consistent with a hierarchical arrangement of substates. The Figure was generated with OPLOT (Jones *et al.*, 1991).

model (Table 1). This result is surprising in view of the harmonic restriction of the ADP model and the compelling evidence for multimodality (anharmonicity) in the electron density for $Ca^{2+}$-CaM. An analysis of the total $U_{eq}$ from the multiconformer models demonstrates that all three multiconformer models contain a significant contribution from the isotropic *B*-factors (Figure 8(a), (b) and (c)). Furthermore, an analysis of ADPs synthesized from the multiconformer models (see Materials and Methods) reveals that the best multiconformer model (model A) gives a mean anisotropy for the $C^{\alpha}$ atoms ($\langle A_{C\alpha} \rangle$) of 0.687 ($\sigma = 0.101$), which is far more isotropic than the value of ($A_{C\alpha}\rangle = 0.373$ ($\sigma = 0.127$) obtained from the single conformer ADP model. The other two multiconformer refinements display a similar degree of isotropy (Table 1), demonstrating that all three multiconformer models provide an overly isotropic depiction of $Ca^{2+}$-CaM disorder.

This isotropic bias is inherent in any multiconformer refinement strategy that employs isotropic *B*-factors and indicates that a model that allows an anisotropic treatment of each conformer would provide a more accurate description of both anisotropic and anharmonic macromolecular disorder. Although a full atomic ADP multiconformer model contains far too many parameters for refinement against typical macromolecular data, reduced variable representations of ADPs for each conformer, such as normal mode or TLS models (Kidera & Go, 1990; Schomaker & Trueblood, 1968), are viable options that warrant further investigation.

## Conclusions

The 1.0 Å crystal structure of $Ca^{2+}$-CaM reveals that the protein exhibits both anisotropic and anharmonic disorder that spans a range of length-scales, from discrete side-chain disorder to domain displacements. The 36 discretely disordered residues are concentrated in the central helix and the hydrophobic pockets, supporting the view that structural plasticity in these regions of the protein is important for target recognition and binding. Furthermore, the number and proximity of discretely disordered residues in the hydrophobic pockets suggests that correlated side-chain displacements may facilitate target specific deformation of the binding surfaces in CaM. In addition, a TLS analysis of the anisotropy of the protein disorder reveals relative lobe displacements in the crystal that coincide with proposed motions of the protein in solution. Lastly, clustering of the individual conformers in multiconformer models of $Ca^{2+}$-CaM disorder provides clear evidence for multimodal disorder for both side-chain and backbone atoms in the protein and reveals an intermediate level of disorder in CaM that is consistent with a hierarchical arrangement of protein conformational substates. In total, the many manifestations of disorder in $Ca^{2+}$-CaM demonstrate that the protein occupies a very large number of conformational substates in the crystalline environment at cryogenic temperatures. It is certain that the protein is more disordered in solution at physiological temperatures, and thus the role of dynamics in CaM function is probably even more prominent than has been pre-

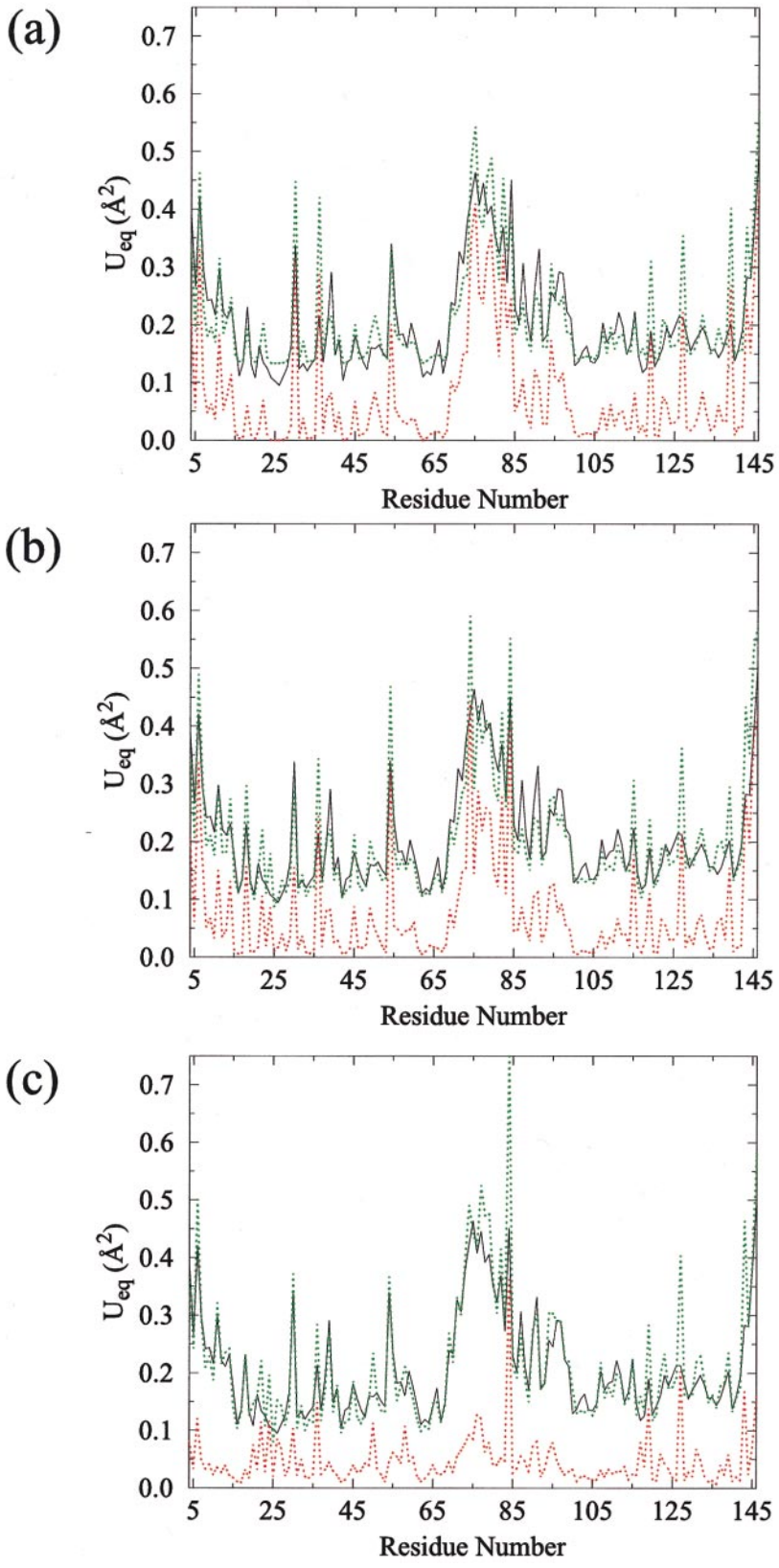


**Figure 8.** Analysis of the disorder in the three multiconformer refinements. Throughout, the black trace represents the $U_{eq}$ of the single conformer ADP model, the red dotted trace is the MSD of the multiconformer ensemble, and the green dotted trace is the $U_{eq}$ of the multiconformer models calculated by summing the contributions of the ensemble MSD and the atomic isotropic *B*-factors in each conformer. (a) Multiconformer model A, in which the isotropic *B*-factors were fixed at 10 Å$^2$ through the refinement. (b) Multiconformer model B, in which initial *B*-factors of 10 Å$^2$ were refined. (c) Multiconformer model C, in which the initial *B*-factors were taken from the refined single conformer model and were then further refined. In all three cases, the isotropic *B*-factors contribute substantially to the total $U_{eq}$ of the multiconformer models and are required to achieve quantitative agreement with the single conformer ADP model.

viously recognized. We propose a model whereby CaM samples a quasi-continuous spectrum of conformational substates, rather than cycling between a few well-defined conformations in response to $Ca^{2+}$ or target proteins. In this model, both apo ($Ca^{2+}$-free) and holo ($Ca^{2+}$-bound) CaM sample very large and partially intersecting volumes of configuration space in solution. Binding of $Ca^{2+}$ to

apo CaM stabilizes certain conformational substates at the expense of others, resulting in structural changes in the EF-hands that reduce the frequency that the most compact conformational substates of the protein are sampled. This dynamical change alters the time-averaged structure of the protein but still allows large fluctuations in the instantaneous structure of $Ca^{2+}$-CaM. This work does not directly address the structure of the apo CaM, and therefore the details of this model are speculative; however, the extensive disorder in crystalline $Ca^{2+}$-CaM observed here, as well as NMR, solution and thermal diffuse X-ray scattering studies of both forms of the protein are consistent with the existence of a spectrum of conformational substates that spans the both the apo and holo forms of the protein (Ikura *et al*., 1992; Meador *et al*., 1993; Osawa *et al*., 1999b; Swindells & Ikura, 1996; Trewhella, 1992; Wall *et al*., 1997). In addition, this model may account for the variety of different binding modes seen in $Ca^{2+}$-CaM-target complexes (Evenas *et al*., 1999). Furthermore, such pronounced flexibility should allow $Ca^{2+}$-CaM to be observed in conformations that differ substantially from the canonical extended structure seen in all previous crystallographic studies. Thus, we propose that the functionally distinct forms (apo and holo) of CaM may be less structurally distinct than previously believed, and the different activities CaM in response to $Ca^{2+}$ may result primarily from $Ca^{2+}$-mediated alterations in the dynamics of the protein.

The nature of the disorder in CaM provides direct structural support for the hierarchical view of protein potential energy landscapes (Frauenfelder *et al*., 1991). We propose that the observation of a hierarchical organization of protein conformational substates can be explained using the notion of stochastic self-similarity in the topology of minima in protein potential energy functions. Strict self-similarity requires invariance under a fixed scale transformation (magnification or contraction by a constant factor), which is unlikely to be obeyed in protein potential energy functions (Frauenfelder, 1995; Glockle & Nonnenmacher, 1995; Leeson & Wiersma, 1995). A less artificial model can be obtained by replacing the fixed scale factor with a random variable, resulting in a stochastically self-similar potential function. Stochastic self-similarity allows for random deviations from strict scale invariance, while preserving much of the qualitative behavior of a self-similar model, including a hierarchical construction, and provides a robust mathematical framework in which to view the geometry of protein potential energy surfaces (Roman, 1997; Veneziano, 1999). One attractive feature of a self-similar model is that it requires a rugged funnel-like structure of the potential energy surface, consistent with current models of the potential energy landscape in the protein folding community (Dill & Chan, 1997). Furthermore, self-similarity naturally gives rise to non-trivial correlations in the heights of energy barriers that separate conformational substates. These correlations are expected to result in power law time-dependence in the non-equilibrium dynamic properties of the protein (including the velocity autocorrelation function) and anomalous diffusion behavior in the time-dependence of conformation space sampling at equilibrium (Hambly, 1992; Roman, 1997; Veneziano, 1999). Evidence for both of these aspects of protein dynamics has been seen in molecular dynamics simulations (Garcia *et al*., 1997; Garcia & Hummer, 1999) and could be experimentally addressed in the $Ca^{2+}$-CaM model system with low-temperature spectroscopic methods using either lanthanide-substituted CaM or the single intrinsic tyrosine fluorophore in the protein. In addition, a stochastically self-similar potential energy surface would produce a spatial correlation function with power law functional form, which can be directly tested using X-ray diffuse scattering. The combination of detailed experimental information on the dynamics of CaM and improved theory based on a stochastically self-similar topology of conformational substates may aid in the development of more quantitative models of protein dynamics. Given the compelling evidence for functionally relevant dynamics in $Ca^{2+}$-CaM and the extensive structural characterization of this protein, we propose that CaM provides an ideal model system for this work.

## Materials and Methods

### Protein expression, purification and crystallization

A culture of *Escherichia coli* strain JM109 (Promega) containing the *Paramecium tetraurelia* CaM gene in the expression vector pKK233-3 (Amersham-Pharamcia) was kindly provided by the laboratory of C. Kung (University of Wisconsin-Madison) and the sequence of the gene was verified by DNA sequencing (W. M. Keck Biotechnology Resource Laboratory, Yale University) using the Sanger method (Sanger *et al*., 1977). Protein was overexpressed by IPTG induction of log phase cells (absorbance at 600 nm = 0.5-0.7) in LB broth and purified according to a modified version of the method by Putkey *et al.* (1985). Briefly, pelleted cells were resuspended in protein buffer (40 mM Tris-HCl, pH 7.5) supplemented with 2.4 M sucrose and 50 μM EDTA, and then lysed by adding four volumes of protein buffer supplemented with 100 μg/ml lysozyme at room temperature. The lysate was then heated to 100 °C for five minutes, rapidly cooled on an ethanol/ice bath, and centrifuged at 100,000 *g* for five hours at 4 °C to pellet cellular debris. The heat treatment procedure results in a five- to tenfold enrichment of CaM in the soluble fraction. The supernatant was decanted and made 5 mM in $Ca^{2+}$ prior to batch-binding to phenyl-Sepharose resin at 4 °C. After binding, the resin was washed with five volumes of protein buffer supplemented with 5 mM $CaCl_2$, followed by a wash of five volumes of protein buffer supplemented with 500 mM NaCl and 5 mM $CaCl_2$, and a final wash of five volumes of protein buffer supplemented with 5 mM $CaCl_2$. Protein was eluted from the resin by washing with a minimal volume of protein buffer supplemented with 10 mM EDTA and 5 mM EGTA. The washes that contained protein were determined by Brad-

ford assay and then pooled. The protein was applied to a Mono Q 16/10 column (Amersham-Pharmacia) and eluted with a 12 column volume gradient of 0-500 mM KCl in protein buffer supplemented with 5 mM EGTA. The peak fractions were pooled, dialyzed for 48 hours against two changes of 100 volumes of water, and lyophilized. The protein was judged homogeneous by silver-staining following SDS-PAGE and by electrospray mass spectrometry. The approximate yield was 5-7 mg of pure protein from 1 g of wet cell mass.

Immediately before crystallization trials were performed, lyophilized protein was resuspended in deionized water to a concentration of 15 mg/ml as determined by absorbance at 280 nm using an $\varepsilon_{280}$ of 0.077 for a 1 mg/ml solution of *P. tetraurelia* CaM. Crystallization was accomplished using the hanging drop vapor diffusion method at 4 °C with the drop consisting of 4 µl of 15 mg/ml CaM, 5 mM $CaCl_2$, 50 mM sodium cacodylate (pH 5.0), and 1 µl of methylpentanediol (MPD). The reservoir contained 1 ml of 50 mM sodium cacodylate (pH 5.0) and 50 % MPD. Crystals appeared after two weeks and were used to microseed other drops, since spontaneous nucleation events were rare. The knife blade-shaped crystals typically grew to a size of 400 µm × 300 µm × 100 µm in an additional week and were in space group *P*1. Since the concentration of MPD in the mother liquor was sufficient to act as a cryoprotectant, the crystals were removed from the drop using a nylon loop and directly immersed in liquid propane for shock-cooling.

## Data collection and processing

*P. tetraurelia* $Ca^{2+}$-CaM crystallized in space group *P*1 with unit cell dimensions $a = 25.02$ Å, $b = 29.42$ Å, $c = 52.76$ Å, $\alpha = 89.54°$, $\beta = 86.10°$, $\gamma = 82.39°$. Diffraction data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) on beamline 9-2 using an ADSC Quantum-4 CCD detector and 14 KeV radiation (0.886 Å) at 100 K. Data were obtained from two unusually thick crystals (500 µm × 400 µm × 300 µm) in different orientations with respect to the spindle axis in order to uniformly sample reciprocal space. Complete coverage of reciprocal space is particularly important for this study, since the omission of blind region data would have resulted in artifacts in the electron density map and would produce systematic errors in the ADP refinement. In addition, merging data collected from two crystals in different orientations with respect to the spindle axis minimizes the effect of crystal anisotropy on the data, which can also be a source of systematic error in the ADP refinement for low symmetry space groups. Separate high (1.8-1.0 Å) and low (50-1.5 Å)-resolution data sets were recorded with variable exposure times, oscillation widths and detector distances in order to eliminate pixel overloads for the most intense reflections. The data were processed and scaled using DENZO and SCALEPACK, respectively (Otwinowski & Minor, 1997). Four data sets collected from two crystals were merged in the final SCALEPACK job to yield a 5.4-fold redundant, 96.9 % complete data set extending to 1.0 Å resolution with 78,247 unique reflections. The data were of high quality, with $\langle I/\sigma(I)\rangle = 17.4$ and $R_{merge}$ of 7.3 %. Data statistics are summarized in Table 3.

## Refinement of the model

### CNS refinement

The 1.0 Å data were phased using a previously solved but unpublished 1.1 Å structure of *P. tetraurelia* CaM and the initial refinement was performed using the Crystallography and NMR System (CNS) (Brunger *et al.*, 1998). Throughout the CNS portion of the refinement, all data from ∞-1.0 Å were used and no sigma cutoff was applied at any time. In addition, overall anisotropic scale factors and a bulk solvent correction were applied to the data and a maximum likelihood amplitude-based target function including stereochemical restraints was employed during refinement (Adams *et al.*, 1997). The progress of the rebuilding and refinement was monitored using $R_{free}$ (Brunger, 1992), which was computed from a randomly chosen test set comprising 10 % of the data (approximately 7800 reflections).

The starting model was optimized by rigid body refinement, followed by molecular dynamics simulated annealing torsion angle refinement (Rice & Brunger, 1994) and individual isotropic *B*-factor refinement as implemented in CNS (Brunger *et al.*, 1998). Manual adjustments to the model (including solvent) were performed in O (Jones *et al.*, 1991) over several alternating

**Table 3.** Data statistics

| Resolution range (Å) | Unique reflections | Completeness (%) | Multiplicity | $\langle I/\sigma(I)\rangle$ | $I > 2\sigma(I)$ (%) | $R_{merge}$ [b] |
|---|---|---|---|---|---|---|
| 52-2.15 | 7758 | 95.9 | 5.9 | 17.0 | 99.7 | 0.068 |
| 2.15-1.71 | 8054 | 100 | 6.0 | 25.5 | 98.9 | 0.057 |
| 1.71-1.49 | 8100 | 100 | 5.9 | 25.9 | 97.3 | 0.064 |
| 1.49-1.36 | 8063 | 99.8 | 5.8 | 22.0 | 94.2 | 0.082 |
| 1.36-1.26 | 8036 | 99.4 | 5.7 | 17.2 | 89.7 | 0.107 |
| 1.26-1.19 | 7957 | 99.0 | 5.7 | 13.0 | 84.6 | 0.136 |
| 1.19-1.13 | 7974 | 98.6 | 5.6 | 10.2 | 79.0 | 0.171 |
| 1.13-1.08 | 7935 | 97.9 | 5.4 | 6.8 | 68.7 | 0.234 |
| 1.08-1.04 | 7613 | 94.7 | 4.8 | 3.7 | 56.1 | 0.356 |
| 1.04-1.00 | 6757 | 83.9 | 3.3 | 2.1 | 49.3 | 0.506 |
| 52-1.00 | 78,247 | 96.9 | 5.4 | 17.4 | 81.8 | 0.073 |

$Ca^{2+}$-CaM crystallizes in space group *P*1 with unit cell dimensions $a = 25.02$ Å, $b = 29.42$ Å, $c = 52.76$ Å, $\alpha = 89.54°$, $\beta = 86.10°$, $\gamma = 82.39°$. There is one molecule in the unit cell and the solvent content of the crystal is 40 % (v/v). The final data set is a composite of four data sets collected from two crystals at 100 K using 0.886 Å X-rays. All data were scaled together and no final merging of separate data sets was performed.

[a] $R_{merge} = \Sigma_h \Sigma_j |I_j(\mathbf{h}) - \langle I(\mathbf{h})\rangle| / \Sigma_h \Sigma_j I_j(\mathbf{h})$, where $\langle I(\mathbf{h})\rangle$ is the mean intensity of a set of equivalent (Friedel related or multiply measured) reflections and $I_j(\mathbf{h})$ is the *j*th observation of the reflection with index $h = (h, k, l)$ or $(-h, -k, -l)$.

cycles of model rebuilding, conjugate direction minimization and map calculation. Throughout the rebuilding process, alterations to the model were guided by consideration of both $F_o - F_c$ and $2F_o - F_c$ $\sigma_A$-weighted composite annealed omit maps in order to minimize the effects of residual model bias. The initial solvent model from the 1.1 Å structure was extended by alternating cycles of automated water picking in CNS, followed by map inspection and rejection of any water molecules that were placed in density that may belong to alternative side-chain conformations. Candidate solvent molecules were retained only if they refined to a *B*-factor less than 55 Å$^2$, made plausible hydrogen bonding interactions and corresponded to a peak of 3.5 σ or greater in $\sigma_A$-weighted $F_o - F_c$ maps.

The electron density maps near the end of the CNS refinement showed clear evidence of alternative conformations for several residues and strong indications of anisotropy in most of the protein. The final CNS model refined to an $R_{free}$ of 26.23 % ($R = 24.56$ %) and included 148 amino acid residues, five $Ca^{2+}$, and 160 water molecules.

### *SHELX97 refinement*

The final CNS model was further refined in SHELX97-2 (Sheldrick & Schneider, 1997) using conjugate gradient least-squares minimization against an intensity-based residual target function that included stereochemical and displacement parameter restraints. A bulk solvent correction (SWAT) was applied to the data, which allowed all data from ∞-1.0 Å to be included in the refinement. In contrast to the CNS refinement, no anisotropic scaling was applied to the data during the SHELX97 portion of the refinement. This was done in order to avoid the introduction of bias into the final ADP model, since anisotropic scaling parameters are highly correlated with atomic ADPs.

Initial conjugate gradient least-squares minimization with isotropic displacement parameters was followed by the refinement of individual ADPs for all atoms in the model (including solvent). The protein atom ADPs were subject to rigid-bond (DELU) and similarity (SIMU) restraints and the ordered solvent ADPs were subject to approximate isotropic restraints (ISOR). The introduction of ADPs into the model resulted in a dramatic improvement in both the $R_{free}$ (8 % drop) and the appearance of the $\sigma_A$-weighted electron density maps. Inspection of both $F_o - F_c$ and $2F_o - F_c$ maps allowed 36 residues to be placed into alternative conformations (see Results and Discussion), although many of these residues were previously identified as disordered in the $\sigma_A$-weighted $2F_o - F_c$ composite annealed omit maps using the CNS model. The occupancies of these alternative conformations were refined, but constrained to sum to unity (FVAR). The ordered solvent model was substantially altered at this stage by manual rebuilding into the improved maps. In addition to the introduction of several fully occupied water sites, nine partially occupied water sites were included with fixed fractional occupancies. Following every alteration to the model, the effected atoms were refined with isotropic displacement parameters for five cycles before the introduction of ADPs in order to achieve rapid convergence in the refinement. The later stages of the SHELX97 refinement consisted of minor manual adjustments to side-chain positions and the solvent model. Riding hydrogen atoms were introduced in the final stages of refinement, resulting in a significant improvement to the model as judged by the $R_{free}$ factor (Table 4).

The final SHELX97 model ($R_{free} = 16.35$ %, $R_{work} = 13.45$ %, $R_1 = 13.68$ %) consists of 146 amino acid residues, 36 of which are in alternative conformations, five $Ca^{2+}$, and 178 water molecules. The behavior of the model at key stages of the refinement is shown in Table 4. The quality of the final $\sigma_A$-weighted $2F_o - F_c$ electron density is excellent and reveals extensive discrete disorder in the protein (Figure 3(a) and (b)). However, despite the generally high quality of the electron density, the N-terminal Ala and the C-terminal Lys residue could not be unambiguously placed in density and thus were not included in the final model.

Although the model was extensively refined, the $R_{free}$ is somewhat higher than expected for 1.0 Å resolution data. There are a variety of possible reasons for the high $R_{free}$ value. First, the final data set was obtained by merging four separate data sets collected from two crystals. Second, the data at high resolution are weak and have a high $R_{merge}$ value (Table 3). Third, the nature and extent of the disorder in $Ca^{2+}$-CaM may not be modeled accurately by currently employed methods of refinement. Lastly, there are several instances of residual density in the final $F_o - F_c$ map that cannot be readily interpreted. The density may correspond to partially ordered solvent or MPD molecules, although neither of these alternatives is completely convincing. These features, which are most pronounced in the two hydrophobic binding pockets, are prominent in maps contoured at 2 σ, but are not visible

**Table 4.** Refinement progress

| Model | Protein atoms | H atoms | Water[a] | $N_{par}$ | $R_{work}$ (%)[b] | $R_{free}$ (%)[b] |
|---|---|---|---|---|---|---|
| Final CNS model | 1165 | 0 | 160 | 5323 | 25.82 | 26.23 |
| CNS model in SHELXL | 1165 | 0 | 160 | 5323 | 25.25 | 27.69 |
| ADPs included | 1165 | 0 | 158 | 11,955 | 15.57 | 18.69 |
| Discrete disorder included | 1165 | 0 | 176 | 14,957 | 14.23 | 17.50 |
| Solvent adjustments | 1165 | 0 | 191 | 15,388 | 14.14 | 17.35 |
| Terminal residues removed | 1144 | 0 | 169 | 15,078 | 14.30 | 17.51 |
| Riding hydrogen added | 1144 | 1089 | 173 | 14,905 | 13.59 | 16.49 |
| Minor adjustments | 1144 | 1089 | 178 | 14,910 | 13.52 | 16.37 |
| Final model | 1144 | 1089 | 178 | 14,975 | 13.45 | 16.35 |

[a] After every refinement, water molecules with $U_{eq}$ greater than 0.6 Å$^2$ were rejected, and any peaks greater than 3.5 σ that made plausible hydrogen bonding interactions were included in the solvent model. Newly placed water molecules (and adjusted protein atoms) were refined with isotropic displacement parameters for ten cycles before introducing ADPs.

[b] Both $R_{work}$ and $R_{free}$ were calculated with no sigma cutoffs using the full resolution range of the data (50-1.0 Å). $R_{free}$ (Brunger, 1992) was calculated from a test set that consisted of 10 % of the data (approximately 7800 reflections).

in maps contoured at 3 σ. The significance of these features is uncertain and therefore no adjustment to the protein model was made in response to this unexplained density.

### Assessment of model quality

The SHELX97 model was subjected to a final cycle of unrestrained full matrix least-squares minimization in order to determine the estimated standard deviations (ESDs) for atomic positions by matrix inversion. The result is shown in Table 5. The average values shown in Table 5 compare favorably with typical values found in other high-resolution structures, indicating a well-refined model. The final model was also subjected to quality assessment using PROCHECK (Laskowski *et al*., 1993). The model has 95.4 % of the residues in core regions of the Ramachadran plot, 3.8 % are located in allowed regions, and one residue (Gln3) is located in a generously allowed region. The backbone atoms in Gln3 are located in well-defined $2F_o - F_c$ density at a contour level of 1.5 σ and cannot be placed into this density for any choice of backbone angles that is consistent with more favored regions of Ramachadran space. No residue is in the disallowed region of the plot. The overall *G*-score for the structure is 0.09.

### Radius of curvature calculation

The radius of curvature for the central helix was calculated using the relationship between the lengths of an arc ($L_{arc}$) and a chord ($L_{chord}$) with common endpoints, $L_{chord}/L_{arc} = \sin\theta/\theta$, where θ is half the angle that subtends the arc, $L_{chord}$ is the measured distance between $C^{\alpha}70$ and $C^{\alpha}90$ in the central helix, and $L_{arc}$ is the length of an idealized α-helix of 21 residues. The equation was solved for θ by Maclaurin series expansion to third order, and the resulting value of θ was used to calculate the radius of curvature for the central helix according to $R_{curvature} = L_{arc}/2\theta$. This method is potentially subject to systematic error that may lead to precise but inaccurate values of $R_{curvature}$. However, the relative difference between values of $R_{curvature}$ for the previous structure and the present model are not effected by this error and thus reflect a significantly more bent helix in this study.

**Table 5.** Model statistics

| | |
|---|---|
| No. protein residues[a] | 146 |
| No. discretely disordered residues | 36 |
| No. $Ca^{2+}$ | 5 |
| No. protein/solvent/hydrogen atoms | 1144/178/1089 |
| $R_{free}$ (%)[b] | 16.35 |
| $R_{work}$ (%)[b] | 13.45 |
| $R_1$ (%)[b] (all reflections) | 13.68 |
| Mean ESD (Å)[c] | |
| Protein main-chain/side-chain atoms | 0.026/0.039 |
| Solvent atom positions | 0.048 |
| $Ca^{2+}$ positions | 0.004 |
| Mean $U_{eq}$ (Å$^2$) | |
| Protein main-chain/side-chain atoms | 0.169/0.240 |
| Solvent atoms | 0.348 |
| $Ca^{2+}$ ions | 0.137 |
| Mean anisotropy[d] | |
| Protein atoms | 0.353 |
| Solvent atoms | 0.371 |
| $Ca^{2+}$ | 0.346 |

[a] The N-terminal Ala and the C-terminal Lys could not be placed into density and thus were omitted from the final model.

[b] $R_{work}$, $R_{free}$ and $R_1$ were calculated with no sigma cutoffs using the full resolution range of the data (50-1.0 Å). $R_{free}$ (Brunger, 1992) was calculated from a test set that consisted of 10 % of the data (approximately 7800 reflections).

[c] Estimated standard deviations (ESDs) were determined from unrestrained full matrix least-squares minimization followed by matrix inversion in SHELX97 (Sheldrick & Schneider, 1997). Only fully occupied non-hydrogen atoms contribute to the averages reported above.

[d] Anisotropy is defined as the ratio of the lengths of the shortest and longest principal axes of the atomic thermal ellipsoid and was determined using PARVATI (Merritt, 1999).

### TLS refinement

TLS refinements were performed in RESTRAIN (Collaborative Computational Project Number 4, 1994) and included both side-chain and backbone atoms (as well as $Ca^{2+}$) in the variously defined TLS groups. As a control, TLS refinements including only main-chain atoms and $Ca^{2+}$ were performed and gave tensors that were comparable to the all-atom TLS refinements, indicating that inclusion of side-chain atoms does not introduce substantial bias into the refined translations or librations for this system. The center of reaction for each TLS tensor was fixed at the center of mass of the TLS group and not refined. However, the screw tensor was allowed to be non-symmetric to compensate for the fact that the center of mass may not correspond to the center of reaction for a rigid group. TLS refinements were carried out for ten cycles using conjugate gradient minimization against a least-squares target function with stereochemical restraints. The resulting tensors were analyzed using TLSANL (Collaborative Computational Project Number 4, 1994).

### Multiconformer refinement

All multiconformer refinements were performed in CNS (Brunger *et al*., 1998) using a maximum likelihood amplitude-based target function that included stereochemical restraints. A bulk-solvent correction was applied to the data, and overall isotropic scaling was used. Scaling the data isotropically is important for multiconformer refinement, as the correlation between anisotropic scaling parameters and the individual anisotropy of the protein atoms will result in an isotropic bias in the final multiconformer model if prior anisotropic scaling is performed. Two tandem cycles of molecular dynamics simulated annealing torsion angle refinement (Rice & Brunger, 1994), conjugate direction minimization, and individual isotropic *B*-factor refinement were performed with a starting temperature for the simulated annealing refinement of 2500 K. The use of multiple tandem cycles of refinement has been observed to improve the convergence of the refinements when isotropic *B*-factors are refined (unpublished results). We believe this is because the individual conformer displacements and the isotropic *B*-factors are correlated and are not refined simultaneously in CNS, although the influence of this postulated correlation on multiconformer refinement has not been thoroughly investigated. Trials using two, four, six, eight, and ten copies of the entire single conformer model with various isotropic *B*-factor models were performed, and the progress of the refinement was monitored using the $R_{free}$ value (see Results).

ADPs were computed for all $C^{\alpha}$ atoms from the multi-conformer models by determining an atomic variance-covariance matrix from the final ensemble model. For the example of a $U_{xy}$ covariance element in the ADP tensor, this element is given by:

$$U_{xy} = 10000\{\langle(\langle x\rangle - x_i)\rangle\langle(\langle y\rangle - y_i)\rangle\} = 10000\sigma_x\sigma_y$$

In this expression, $x_i$ is the $x$ coordinate of a $C^{\alpha}$ atom in the $i$th conformer, $y_i$ is the $y$ coordinate of a $C^{\alpha}$ atom in the $i$th conformer, angle brackets indicate an average over all conformers in the ensemble and $\sigma_x$ is the standard deviation in the $x$ coordinate of a $C^{\alpha}$ atom in the ensemble. The factor of 10000 is included to conform to the PDB convention for the representation of ADPs.

The contribution of the isotropic $B$-factors to the variance-covariance matrix was included by adding the equivalent $U_{eq}$ to each diagonal term in the matrix, where the $U_{xx}$ diagonal (variance) element in the ADP tensor is calculated according to:

$$U_{xx} = 10000\{\langle(\langle x\rangle - x_i)\rangle^2 + U_{eq}\} = 10000(\sigma_x^2 + U_{eq})$$

The resulting ADP model was analyzed using PARVATI (Merritt, 1999).

### Protein Data Bank accession number

The coordinates and structure factors have been deposited in the RCSB Protein Data bank with accession code 1EXR.

## Acknowledgments

We thank Paul Adams, Joel Hyman, Luke Rice, Fred Richards and Robert Rizzo for many helpful discussions and for critical reading of the manuscript. Particular gratitude is extended to Kit-Yin Ling in the laboratory of C. Kung (University of Wisconsin-Madison) for providing us with the clone for *P. tetraurelia* CaM, to Ethan Merritt for many helpful discussions concerning ADP and TLS refinement, and to F. Temple Burling for the suggestion of CaM as a model system. This work is based upon research conducted at the Stanford Synchrotron Radiation laboratory (SSRL), which is funded by the Department of Energy (BES, BER) and the National Institutes of Health (NCRR, NIGMS).

## References

Adams, P. D., Pannu, N. S., Read, R. J. & Brunger, A. T. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. *Proc. Natl Acad. Sci. USA,* **94**, 5018-5023.

Austin, R. H., Stein, D. L. & Wang, J. (1987). Terbium luminescence lifetime heterogeneity and protein equilibrium conformational dynamics. *Proc. Natl Acad. Sci. USA,* **84**, 1541-1545.

Babu, Y. S., Sack, J. S., Greenhough, T. J., Bugg, C. E., Means, A. R. & Cook, W. J. (1985). 3-Dimensional structure of calmodulin. *Nature,* **315**, 37-40.

Ban, C., Ramakrishnan, B., Ling, K. Y., Kung, C. & Sundaralingam, M. (1994). Structure of the recombinant *Paramecium tetraurelia* calmodulin at 1.68 Å resolution. *Acta Crystallog. sect. D,* **50**, 50-63.

Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W. & Bax, A. (1992). Backbone dynamics of calmodulin studied by N-15 relaxation using inverse detected 2-dimensional NMR spectroscopy - the central helix is flexible. *Biochemistry,* **31**, 5269-5278.

Brodersen, D. E., de La Fortelle, E., Vonrhein, C., Bricogne, G., Nyborg, J. & Kjeldgaard, M. (2000). Applications of single-wavelength anomalous dispersion at high and atomic resolution. *Acta Crystallog. sect. D,* **56**, 431-441.

Brunger, A. T. (1992). Free *R*-value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature,* **355**, 472-475.

Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D,* **54**, 905-921.

Burling, F. T. & Brunger, A. T. (1994). Thermal motion and conformational disorder in protein crystal structures: comparison of multi-conformer and time-averaging models. *Isr. J. Chem.* **34**, 165-175.

Burling, F. T., Weis, W. I., Flaherty, K. M. & Brunger, A. T. (1996). Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science,* **271**, 72-77.

Collaborative Computational Project Number 4 (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D,* **50**, 760-763.

Chattopadhyaya, R., Meador, W. E., Means, A. R. & Quiocho, F. A. (1992). Calmodulin structure refined at 1.7 Å resolution. *J. Mol. Biol.* **228**, 1177-1192.

Declercq, J. P., Evrard, C., Lamzin, V. & Parello, J. (1999). Crystal structure of the EF-hand parvalbumin at atomic resolution (0.91 Å and at low temperature (100 K). Evidence for conformational multistates within the hydrophobic core. *Protein Sci.* **8**, 2194-2204.

Dill, K. A. & Chan, H. S. (1997). From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4**, 10-19.

Evenas, J., Forsen, S., Malmendal, A. & Akke, M. (1999). Backbone dynamics and energetics of a calmodulin domain mutant exchanging between closed and open conformations. *J. Mol. Biol.* **289**, 603-617.

Frauenfelder, H. (1995). Complexity in proteins. *Nature Struct. Biol.* **2**, 821-823.

Frauenfelder, H., Sligar, S. G. & Wolynes, P. G. (1991). The energy landscapes and motions of proteins. *Science,* **254**, 1598-1603.

Garcia, A. E. & Hummer, G. (1999). Conformational dynamics of cytochrome *c*: correlation to hydrogen exchange. *Proteins: Struct. Funct. Genet.* **36**, 175-191.

Garcia, A. E., Blumenfeld, R., Hummer, G. & Krumhansl, J. A. (1997). Multi-basin dynamics of a protein in a crystal environment. *Physica D,* **107**, 225-239.

Glockle, W. & Nonnenmacker, T. (1995). A fractional calculus approach to self-similar protein dynamics. *Biophys. J.* **68**, 46-53.

Hambly, B. M. (1992). Brownian motion on a homogeneous random fractal. *Probabil. Theory Relat. Fields,* **94**, 1-38.

Harata, K., Abe, Y. & Muraki, M. (1998). Full-matrix least-squares refinement of lysozymes and analysis

of anisotropic thermal motion. *Proteins: Struct. Funct. Genet.* **30**, 232-243.
Howlin, B., Moss, D. A. & Harris, G. W. (1989). Segmented anisotropic refinement of bovine ribonuclease A by the application of the rigid-body TLS model. *Acta Crystallog. sect. A,* **45**, 851-861.
Ichiye, T. & Karplus, M. (1988). Anisotropy and anharmonicity of atomic fluctuations in proteins: Implications for X-ray analysis. *Biochemistry,* **27**, 3487-3497.
Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B. & Bax, A. (1992). Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science,* **256**, 632-638.
Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A,* **47**, 110-119.
Kidera, A. & Go, N. (1990). Refinement of protein dynamic structure: normal mode refinement. *Proc. Natl Acad. Sci. USA,* **87**, 3718-3722.
Kink, J. A., Maley, M. E., Preston, R. R., Ling, K. Y., Wallenfriedman, M. A., Saimi, Y. & Kung, C. (1990). Mutations in paramecium calmodulin indicate functional differences between the C-terminal and N-terminal lobes *in vivo. Cell,* **62**, 165-174.
Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946-950.
Kuriyan, J., Petsko, G. A., Levy, R. M. & Karplus, M. (1986). Effect of anisotropy and anharmonicity on protein crystallographic refinement: an evaluation by molecular dynamics. *J. Mol. Biol.* **190**, 227-254.
Kuriyan, J., Osapay, K., Burley, S. K., Brunger, A. T., Hendrickson, W. A. & Karplus, M. (1991). Exploration of disorder in protein structures by X-ray restrained molecular dynamics. *Proteins: Struct. Funct. Genet.* **10**, 340-358.
Laskowski, R. A., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283-291.
Leeson, D. T. & Wiersma, D. A. (1995). Looking into the energy landscape of myoglobin. *Nature Struct. Biol.* **2**, 848-851.
Meador, W. E., Means, A. R. & Quiocho, F. A. (1992). Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin-peptide complex. *Science,* **257**, 1251-1255.
Meador, W. E., Means, A. R. & Quiocho, F. A. (1993). Modulation of calmodulin plasticity in molecular recognition on the basis of X-ray structures. *Science,* **262**, 1718-1721.
Merritt, E. A. (1999). Expanding the model: anisotropic displacement parameters in protein structure refinement. *Acta Crystallog. sect. D,* **55**, 1109-1117.
Merritt, E. A. & Murphy, M. E. P. (1994). Raster3d version-2.0: a program for photorealistic molecular graphics. *Acta Crystallog. sect. D,* **50**, 869-873.
Moss, D. S., Tickle, I. J., Theis, O. & Wostrack, A. (1996). X-ray analysis of domain motions in protein crystals. In *Macromolecular Refinement: Proceedings of the CCP4 Study Weekend* (Dodson, E., Moore, M., Ralph, A. & Bailey, S., eds), pp. 105-113, Council for the Central Laboratory of the Research Councils, Daresbury Laboratory, England.
Oneil, K. T. & Degrado, W. F. (1990). How calmodulin binds its targets: sequence independent recognition of amphiphilic alpha-helices. *Trends Biochem. Sci.* **15**, 59-64.
Osawa, M., Kuwamoto, S., Izumi, Y., Yap, K. L., Ikura, M., Shibanuma, T., Yokokura, H., Hidaka, H. & Matsushima, N. (1999a). Evidence for calmodulin inter-domain compaction in solution induced by W-7 binding. *FEBS Letters,* **442**, 173-177.
Osawa, M., Tokumitsu, H., Swindells, M. B., Kurihara, H., Orita, M., Shibanuma, T., Furuya, T. & Ikura, M. (1999b). A novel target recognition revealed by calmodulin in complex with $Ca^{2+}$-calmodulin-dependent kinase kinase. *Nature Struct. Biol.* **6**, 819-824.
Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307-326.
Pattabiraman, N., Ward, K. B. & Fleming, P. J. (1995). Occluded molecular surface: analysis of protein packing. *J. Mol. Recogn.* **8**, 334-344.
Putkey, J. A., Slaughter, G. R. & Means, A. R. (1985). Bacterial expression and characterization of proteins derived from the chicken calmodulin cDNA and a calmodulin processed gene. *J. Biol. Chem.* **260**, 4704-4712.
Rader, S. D. & Agard, D. A. (1997). Conformational substates in enzyme mechanism: The 120 K structure of alpha-lytic protease at 1.5 Å resolution. *Protein Sci.* **6**, 1375-1386.
Rejto, P. A. & Freer, S. T. (1996). Protein conformational substates from X-ray crystallography. *Prog. Biophys. Mol. Biol.* **66**, 167-196.
Rice, L. M. & Brunger, A. T. (1994). Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. *Proteins: Struct. Funct. Genet.* **19**, 277-290.
Roman, H. E. (1997). Diffusion on self-similar structures. *Fractals Interdiscipl. J. Complex Geom. Nature,* **5**, 379-393.
Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA,* **74**, 5463-5467.
Schneider, T. R. (1996). What can we learn from anisotropic temperature factors? In *Macromolecular Refinement: Proceedings of the CCP4 Study Weekend* (Dodson, E., Moore, M., Ralph, A. & Bailey, S., eds), pp. 133-144, Council for the Central Laboratory of the Research Councils, Daresbury Laboratory, England.
Schomaker, V. & Trueblood, K. (1968). On rigid body motion of molecules in crystals. *Acta Crystallog. sect. B,* **24**, 63-76.
Sheldrick, G. M. & Schneider, T. R. (1997). SHELXL: high-resolution refinement. *Methods Enzymol.* **277**, 319-343.
Smith, J. L., Hendrickson, W. A., Honzatko, R. B. & Sheriff, S. (1986). Structural heterogeneity in protein crystals. *Biochemistry,* **25**, 5018-5027.
Swindells, M. B. & Ikura, M. (1996). Pre-formation of the semi-open conformation by the apo calmodulin C-terminal domain and implications for binding IQ-motifs. *Nature Struct. Biol.* **3**, 501-504.
Trewhella, J. (1992). The solution structures of calmodulin and its complexes with synthetic peptides based on target enzyme binding domains. *Cell Calcium,* **13**, 377-390.
Veneziano, D. (1999). Basic properties and characterization of stochastically self-similar processes in $R^d$. *Fractals-Complex Geom. Patterns Scaling Nature Soc.* **7**, 59-78.

Wall, M., Clarage, J. & Phillips, G. (1997). Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering. *Strucutre,* **5**, 1599-1612.

Willis, B. T. M. & Pryor, A. W. (1975). *Thermal Vibrations in Crystallography*, Cambridge University Press, London.

Wriggers, W., Mehler, E., Pitici, F., Weinstein, H. & Schulten, K. (1998). Structure and dynamics of calmodulin in solution. *Biophys. J.* **74**, 1622-1639.

Zhang, M., Tanaka, T. & Ikura, M. (1995). Calcium induced conformational transition revealed by the solution structure of apo calmodulin. *Nature Struct. Biol.* **2**, 758-767.

***Edited by I. Wilson***

*(Received 15 May 2000; received in revised form 7 July 2000; accepted 7 July 2000)*