

# A predicted structure of calmodulin suggests an electrostatic basis for its function

(calmodulin-peptide interaction/protein structure prediction/molecular mechanics/computer graphics)

KARYN T. O'NEIL AND WILLIAM F. DEGRADO\*

E. I. Du Pont de Nemours & Co., Central Research & Development Department, Experimental Station, Wilmington, DE 19898

Communicated by H. E. Simmons, April 3, 1985

**ABSTRACT** By using interactive computer graphics, two models for calmodulin have been constructed based on the structures of two functionally and structurally related proteins, intestinal calcium-binding protein and carp parvalbumin. The two models have been compared and contrasted to the parent proteins with respect to proportion of solvent-exposed hydrophobic residues, solvent-accessible surface area, and side-chain packing. Electrostatic potential surfaces generated for the models suggest a probable binding site for basic amphiphilic  $\alpha$ -helical peptides located between the last E and F helices in the second domain of calmodulin. Both electrostatic and hydrophobic complementarity can contribute to stabilization of a peptide-protein complex in this region.

Proteins with homologous amino acid sequences are known to have similar tertiary structures (1). This has led to the suggestion that if the crystal structure of one protein from a given structural class is known, the structures of other proteins in the same class might be inferred based on their sequence homologies. The feasibility of this concept has been experimentally verified as the crystal structures for many globins, proteases, and cytochromes have been solved, and the structures of several members within these protein families were found to be similar (2-4). Models for proteases, immunoglobulins (for a brief review, see ref. 5), and calcium-binding proteins (6-8) have also been predicted based on homology. The modeling procedures used in these studies involved aligning the sequences of the proteins of known and unknown tertiary structure to maximize homology and building a computer or physical model in which the most homologous regions occupy geometrically equivalent positions (9). Regions that display low levels of sequence homology, including insertions and deletions, are assigned conformations based on intuition, secondary structure prediction schemes (10), and model building. The geometry of the resulting structure is often optimized by energy minimization (5).

While this approach is potentially valuable, it suffers from several pitfalls. It requires precise alignment of sequences or the resulting model will necessarily be incorrect. Secondly, energy minimization of the model structure leads to only minor changes in geometry on the order of 0.5-1.0 Å (5), so it is imperative that the input structure be reasonable. Finally, while homologous proteins have qualitatively similar structures, they also can have major quantitative differences in their geometries. For instance, the position of corresponding helices in globins vary by as much as 7 Å and 30° (2).

With these possible shortcomings in mind, we describe models for calmodulin (11) based on the structures of two calcium-binding proteins, carp parvalbumin (CPV; ref. 12) and intestinal calcium-binding protein (ICB; ref. 13). Calmodulin provides an excellent system for testing this

modeling approach because its structure was unknown while the modeling was in progress, but a crystal structure was expected to be forthcoming from several laboratories (14-16). A wealth of other structural information from NMR and spectroscopic investigations was also available to test the validity of the model (11). Finally, while calmodulin shares with CPV and ICB the function of binding calcium, it is unique in having the additional function of binding peptides, hydrophobic drugs, and target enzymes (11). If a structural basis for this latter function could be inferred from a predicted structure, it would indeed be a success for interactive computer modeling.

Calmodulin contains two internally homologous domains, each of which displays  $\approx 20\%$  exact sequence homology to ICB and CPV (Table 1). The sequences of CPV, ICB, and calmodulin's two domains can be further divided into two internally homologous subdomains. Crystal structures for ICB and CPV indicate that each subdomain of the protein forms a calcium-binding loop flanked on either side by an  $\alpha$ -helix (Fig. 1). This folding motif was termed an EF hand by Kretsinger, who predicted that it would be an important conserved feature in the structures of a variety of calcium-binding proteins (6, 7). Models for troponin C (6) and calmodulin (7) have been proposed from the CPV structure. The more recently determined structure of ICB<sup>†</sup> provides an alternative basis for predicting the structure of calmodulin. While the EF hands in ICB fold neatly into a compact globular structure, the two EF hands in CPV form a more open structure, which packs against two additional helices located at the NH<sub>2</sub> terminus of CPV. These two additional helices have no equivalent in either domain of calmodulin.

In this paper, we compare and contrast ICB- and CPV-based models of calmodulin. Potential peptide-binding sites on calmodulin are proposed based on electrostatic complementarity between the peptide and calmodulin models.

## METHODS

The sequences of calmodulin were aligned with those of ICB and CPV as described in the literature (13). Models were constructed by using the interactive computer graphics program CHEM written by A. Dearing while at University of California at San Francisco and modified by S. Dixon (Lederle) for an Evans and Sutherland Multi Picture System running under the VMS operating system on a VAX 11/750. Side chains were replaced using a molecular model building program FRAG (G. M. Cole, Du Pont) that is interfaced to CHEM. Initial side-chain geometries were taken as those of randomly selected residues found in a highly refined structure

Abbreviations: CPV, carp parvalbumin; ICB, intestinal calcium-binding protein.

\*To whom reprint requests should be addressed.

<sup>†</sup>Coordinates for this structure, which was solved in K. Moffat's laboratory, are available from the current listing of the Brookhaven Protein Data Bank.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. Amino acid sequences of calmodulin, CPV, and ICB

	Helix E	Loop	Helix F		
		* * * † *			
Calmodulin (bovine brain)	ADQLTEEQIAE	FKEAFSLF	DKDNGTITTKE	LGTVMRSL	GQNPT EA (1-46)
	E LQDMINEV	DADGNGTIDFPE	FLTMMARK	MKDTDSEE	(47-83)
	E IREAFRVF	DKDNGYISAAE	LRHVMTNL	GEKLTDE	(84-119)
	EIVDEMIREA	NIDGDGEVNYEE	FVQMMTAK		(120-148)
CPV	ADDVKKAF AII	DQDKSGFIEEDE	LKLFLQNF	KADARALTDG	(40-80)
	EITKTKFLKAG	DSGDGKIGVDE	FTALVKA		(81-108)
ICB	KSPEE	LKGI FEKY	AKEGLPQLSKEE	LKLL LQTE	FPSLLKGPS (1-44)
	TLDLFEEL	DKNGDGEVSPEE	FQVLVKKI	SQ	(45-75)

The sequences of calmodulin, ICB, and CPV are aligned as described in ref. 13. They are organized into regions first hypothesized by Kretsinger to form the helices and calcium binding loops that collectively form an EF hand. The residues aligned with \* indicate the oxygen-containing residues that are ligands for the  $\text{Ca}^{2+}$  ion, and the residues aligned with the † indicate residues that contribute a backbone carbonyl for complexation with calcium. Amino acids are designated by the single-letter code; sequence numbering is shown in parentheses.

of bovine pancreatic trypsin inhibitor. Coordinates for bovine pancreatic trypsin inhibitor, CPV, and ICB (refined to 1.5, 1.85, and 2.0 Å, respectively) were taken from the Brookhaven Protein Data Bank (April 1984 listing). The distance geometry program used was originally described by G. M. Crippin (17), and modified extensively by J. M. Blaney (Du Pont, unpublished program). The potential energies of model structures were minimized using AMBER (18) modified by Z. Wasserman (Du Pont) to run under the VM/CMS operating system of an IBM 3381. Minimization was allowed

to proceed until the root-mean-square (rms) energy gradient was  $<0.5$  kcal per cycle (1 cal = 4.184 J). Accessible surface area was measured by using Connolly's molecular surface program (19). Electrostatic calculations were performed using a program described (20). All nonhydrogen atoms plus all potentially H-bonding hydrogens were included in the electrostatic calculations, and partial charges were those assigned by AMBER (18). Amino acid side chains were assumed to be in their predominant ionization states (31).

## RESULTS

**Construction of Models for Calmodulin.** The initial models for the first and second domains of calmodulin were built by using CHEM to replace the side chains of the parent proteins while maintaining the position of the main chain atoms. The first six residues of the first domain were omitted in the models because there was no equivalent for this region in either parent protein. Construction of the CPV-based model required the deletion of a single three-residue stretch (residues 72-74), which could be accomplished with minimal perturbation of the protein's tertiary structure (8). The construction of the ICB-based model proved to be more difficult because of two insertions at positions 14 and 21 in the first loop of ICB (Table 1). To construct the corresponding loop in calmodulin (sequence, Phe<sup>19</sup>-Asp-Lys-Asp-Gly-Asn-Gly-Thr-Ile-Thr-Thr-Lys-Glu-Leu<sup>32</sup>), the positions of the main chain atoms of residues 19 and 32 were constrained to remain the same as the homologous residues (residues 13 and 28) of ICB, and the geometry of the rest of the loop was allowed to vary. Distance geometry (17) was used to create random conformations for the loop in which the side-chain oxygens of Asp-20, Asp-22, Asn-24, and Glu-31, and the main-chain carbonyl of Thr-26 were fixed in geometries and positions suitable to complex calcium in a manner analogous to the second loop of ICB. The distance geometry algorithm included chiral constraints that ensured planarity of peptide bonds and maintained chirality at the  $\text{C}_\alpha$  positions. Bond lengths and angles were kept rigid, while free rotation was allowed about all other single nonpeptide bonds. Lower bounds for all remaining distances were set to the van der Waals contact distance, while upper bounds were set to the maximum distance for a fully extended chain. Thirty conformers were generated, of which 20 were degenerate by virtue of being superimposable to within 2 Å rms. The geometries of the nondegenerate conformers were improved by 500 cycles of energy refinement using AMBER (18). Four of these structures had energies within 5 kcal/mol of the lowest energy conformer, and these were further refined by

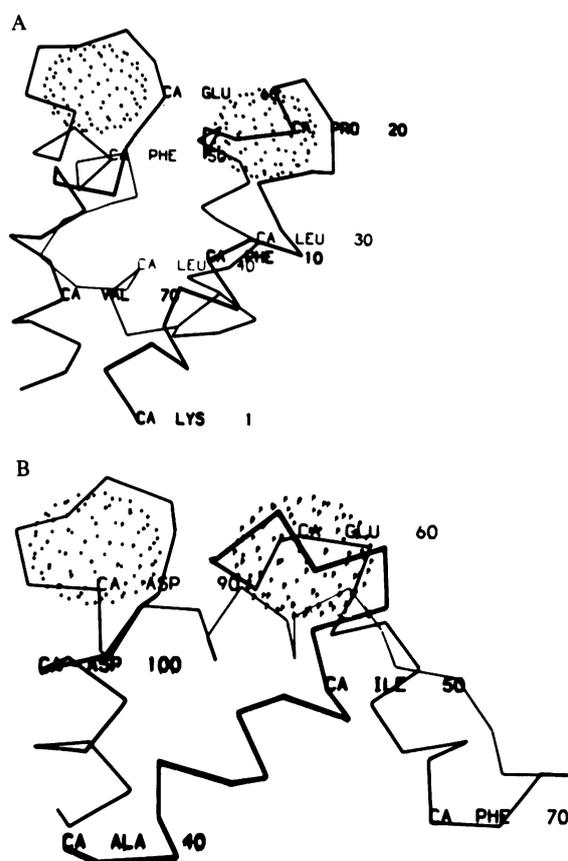


FIG. 1.  $\text{C}_\alpha$  backbone for ICB (A) and CPV (B).  $\text{Ca}^{2+}$  ions are depicted as spheres with a radius of  $1.8\times$  van der Waals radius for  $\text{Ca}^{2+}$ . The two  $\text{NH}_2$ -terminal helices of CPV have been omitted for clarity.

an additional 500 cycles of AMBER. Of these conformers, only one displayed reasonable contacts with the rest of the protein when inserted into the model calmodulin structure. The backbone of this loop was then used as the template for the first loop in the second domain. The completed ICB- and CPV-based models for the individual domains of calmodulin were then energy refined by using AMBER, resulting in a major reduction of bad contacts between side chains but minimal (0.5–1.0 Å) changes in the positions of the main-chain atoms.

The relative orientation of calmodulin's domains in the model of the whole molecule was based on the known physical properties of intact calmodulin. Its anomalously high Stokes radius (ref. 11; unpublished results) suggests that it has an elongated nonspherical geometry. Also, the two isolated domains of calmodulin obtained by limited proteolysis have stable folded conformations and fail to self-associate under a variety of conditions (21). NMR data on intact calmodulin and its isolated domains rule out major interactions between the two domains (8). All of these observations suggest a conformation for calmodulin in which the two domains are held at a distance from one another. A

simple model for accomplishing this is suggested by examination of the amino acid sequence joining the last helix of the first domain and the first helix of the second domain. This sequence has a high potential to form an  $\alpha$ -helix [Chou Fasman  $P_{\alpha} = 1.16$  (10)] and does not contain any helix breaking residues. Due to the position of this sequence between two helices, helix initiation, which is energetically the most costly step in helix formation, has already occurred. Thus, this sequence most likely adopts a helical conformation, and the ICB- and CPV-based models for calmodulin were obtained by joining the two individual domains of calmodulin with an  $\alpha$ -helical linker (Fig. 2). While this work was being reviewed, the structure of troponin C, which is highly homologous to calmodulin, was published (22, 23). The structure of this protein conforms to the general folding pattern described here, with a single stretch of  $\alpha$ -helix connecting the two domains.

The geometries of the ICB- and CPV-based models differ significantly from one another, raising the question of which is the best representation of calmodulin. Distributions of hydrophobic residues, the solvent-accessible surface areas, and side-chain packing densities have recently been shown to

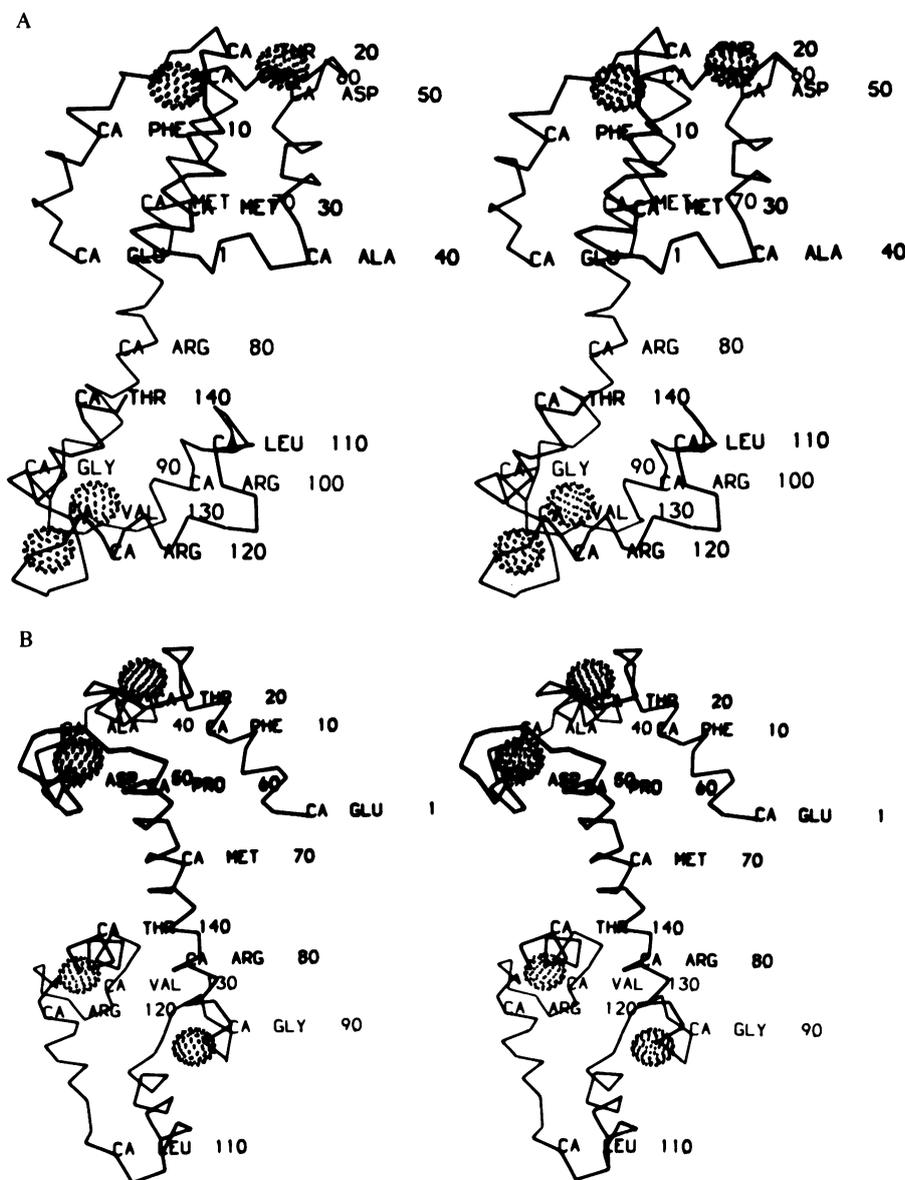


FIG. 2. Stereo drawings of the  $\text{Ca}_2^+$  backbone for the ICB-based model (A) and CPV-based model (B) for calmodulin.  $\text{Ca}^{2+}$  ions were generated as described in Fig. 1. The numbering of residues used in these figures correlates with the actual residue number in calmodulin minus six.

be sensitive and useful criteria for discerning correctly from incorrectly folded proteins (5). When the individual domains of the two models were examined by these criteria, the ICB-based model appeared to be somewhat better than the CPV-based model. Hydrophobic side chains tended to cluster into the interior of the individual domains of the ICB-based model, but they tended to project toward the bottom of the domains of the CPV-based models when viewed as in Fig. 1B. The hydrophobic moments (24) of the helices in the CPV-based model also did not cancel as effectively as in the ICB-based model. These qualitative observations were supported by a quantitative analysis of the solvent-accessible surface areas (25) of the models. The solvent-accessible surface area of a protein has been found to be simply related to the molecular weight of a protein by the equation  $A = 11.1 \times (M_r)^{2/3}$  (26) for single domain proteins. The solvent-accessible surface areas for the ICB- and CPV-based models, when expressed on a per domain basis, were 5096 Å<sup>2</sup> per molecule and 5338 Å<sup>2</sup> per molecule, 11% and 16% higher than expected from the above equation. For comparison, the value calculated for native ICB was within 2% of the value calculated by this equation. The fraction of the solvent-accessible surface area associated with hydrophobic groups (C-H and C-C bonds) followed the same trend, 0.48 for native ICB, 0.52 for the ICB-based model, and 0.54 for the CPV-based model. Qualitatively, the packing density of both models appeared poorer and less uniform than that observed for experimentally determined protein structures. The origin of the high surface areas associated with the models appears to derive from poor packing of the side chains, particularly near the NH<sub>2</sub> termini of the E helices and the COOH termini of the F helices. A second cause for the high surface areas of the models is that the side chains extending from the surface of the models tend to be in more extended conformations than in the actual structures of proteins (see also ref. 5). The above observations suggest that by the above criteria ICB is a better model for predicting the structure of calmodulin than CPV, but that even the ICB-based model does not have all the attributes typically found in crystal structures of proteins.

**Prediction of the Location of Peptide-Binding Sites.** Calmodulin has a site on each of its domains that binds basic hydrophobic drugs and peptides (21, 27). One site has a very high calcium-dependent affinity ( $K_d, \approx 1 \times 10^{-9}$  M) for positively charged amphiphilic  $\alpha$ -helical peptides (28). Binding of peptides to this site blocks calmodulin's ability to activate the target enzymes phosphodiesterase (28) and myosin light chain kinase (27, 29). Tight binding requires multiple positively charged residues on the peptide, so the electrostatic surfaces (20) for the models were examined for regions with a high density of negatively charged residues that might attract such peptides. Three sites with very negative potential were identified, one along the first E helix of the first domain, one along the linker between the two domains, and one along the second E helix of the second domain. The site on the second domain had the highest area associated with a very negative electrostatic potential, so further attention was focused on this site as a probable peptide-binding site. It should also be noted that studies on the isolated domains of calmodulin, obtained by limited proteolysis, showed that only the second domain of calmodulin was capable of binding certain target enzymes (21) and that this domain had the highest affinity for the basic amphiphilic  $\alpha$ -helical peptide mastoparan X (27).

Fig. 3 illustrates a color contour of the electrostatic potential for the second domain of the ICB-based model calculated with a calcium ion positioned in each site. The very negative potential in the vicinity of the calcium binding loop that is only partially neutralized when calcium is included has contributions from both side-chain carboxylates and main-chain carbonyls. This appears to represent the

electrostatic contribution to calcium binding and is probably unrelated to peptide binding because ICB, which does not bind peptides (unpublished results), also has a highly negative field in this region. In contrast, the region of very negative electrostatic potential along the second E helix was not present in ICB and is a likely site for complexing positively charged helical peptides. Similar electrostatic potentials were found in this portion of the protein for both the ICB- and CPV-based calmodulin models.

In addition to the ionic interactions, hydrophobic interactions are also essential for the binding of peptides to calmodulin. For instance, polylysine does not interact strongly with calmodulin (29), but the amphiphilic  $\alpha$ -helical peptide  $\alpha$ -N-fluorenylmethoxycarbonyl-(Leu-Lys-Lys-Leu-Leu-Lys-Leu)<sub>2</sub> binds with a  $K_a$  of  $3 \times 10^{-9}$  M (28). The site along the second E helix of domain 2 has exposed hydrophobic residues available for interaction with the leucyl side chains of this peptide. Val-121 and Ile-125 are partially exposed, as well as Val-142, Met-144, Thr-146, and Ala-147 of the adjacent, very hydrophobic F helix (sequence, Phe<sup>141</sup>-Val-Gln-Met-Met-Thr-Ala-Lys). A reasonable interaction appeared to be possible when the peptide binds with its axis inclined by  $\approx 20^\circ$  or  $-60^\circ$  with respect to the E helix. When the peptide was positioned at  $20^\circ$  and the energy of the complex was minimized using AMBER, the complex illustrated in Fig. 4 was obtained. The possibilities for both hydrophobic and electrostatic interactions are apparent. To determine the extent to which hydrophobic interactions might contribute to the stability of the complex, the solvent exposed surface areas for calmodulin, the calmodulin-peptide complex, and the peptide alone in an extended conformation [the peptide adopts a random conformation in dilute aqueous solution in the absence of calmodulin (28)], were calculated. These calculations indicate a decrease of  $\approx 700$  Å<sup>2</sup> upon complexation. Assuming a decrease of  $-15$  cal/mol in free energy for each square angstrom buried (25), this translates to 10 kcal/mol. For comparison, the experimentally determined free energy of stabilization calculated from the known dissociation constant is  $-11.8$  kcal/mol. Thus, it appears that both electrostatic complementarity and sequestering of hydrophobic residues on complexation contribute to the high binding affinities of basic amphiphilic peptides for calmodulin.

## CONCLUSIONS

Previous attempts at structure prediction based on sequence homology have used only single proteins of known structure as the templates for each protein of unknown structure. A more objective approach to this type of modeling would be to use several different template proteins to create alternative models for a given protein, assuming of course that the crystal structures are known for more than one protein that is homologous to the protein of unknown structure. The degree of similarity of the predicted structure would place limits on the accuracy of the predictions and on the conclusions derived therefrom. The availability of coordinates for ICB and CPV allowed us to take this approach to modeling the structure of calmodulin. The structures of ICB and CPV have qualitatively similar folding patterns but differ in their interhelical packing angles and the distance of closest approach of the F helices. ICB and CPV have approximately equal sequence homology with each other and with calmodulin. Since the structural differences are related to differences in their sequences, it was anticipated that if a common sequence were built into each backbone they should converge onto a single structure upon energy minimization. However, when this was attempted with the sequence of either domain of calmodulin, the resulting models resembled their parent proteins more closely than each other. This

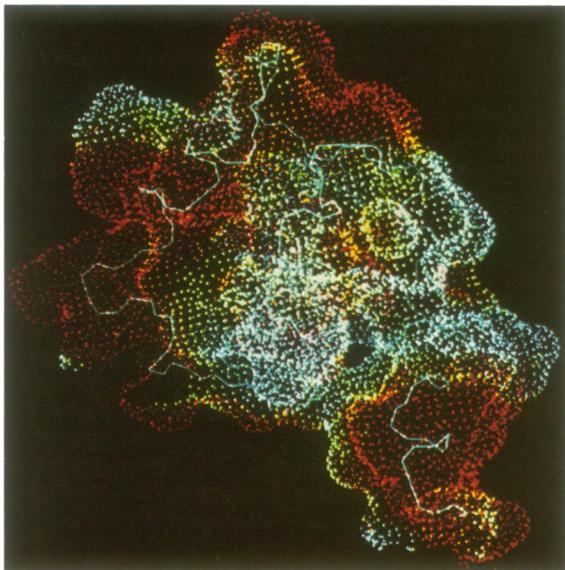


FIG. 3. Electrostatic potential surface (20) for the second domain of the ICB-based model for calmodulin. Contour levels: red,  $V < -10$  kcal/mol; orange,  $-10$  kcal/mol  $< V < -3$  kcal/mol; green,  $-3$  kcal/mol  $< V < +3$  kcal/mol; bluegreen,  $+3$  kcal/mol  $< V < +10$  kcal/mol; blue,  $V > +10$  kcal/mol. The backbone bonds of the structure are white. The second F helix of this domain is in the center with its axis vertically oriented, the second calcium binding loop is in the upper left, and the second E helix runs along the far left of the diagram. Note the very negative potential surrounding the calcium binding loop and the second E helix.

indicates that the methods used in this investigation will yield a reasonably accurate structure only if the structures of the parent protein and the protein to be modeled are very similar.

Given the uncertainties in these models, it is difficult to predict precise molecular details of interactions between basic peptides and calmodulin. However, the general location of the binding sites for a very basic amphiphilic peptide with a large cross-sectional area might be inferred from consideration of the protein's electrostatic surface. Electrostatic fields of functional importance tend to extend over large areas (30, 31). Thus, our conclusions should be resistant to minor differences between the models and the actual structure of calmodulin. Indeed, the electrostatic surfaces of the ICB- and CPV-based models were similar, although their structural details were different. This work suggests that the electrostatic contribution to peptide binding comes from acidic residues along the E helix while hydrophobic contributions to stabilization are made from residues on the E helix and the hydrophobic F helix in the second domain.

We thank Drs. J. Blaney and W. Ripka for many helpful discussions and for the use of their equipment; without their help this work would not have been possible. We also thank Drs. F. R. Salemme, C. Bugg, and S. L. Brenner for critical reading of the paper and for helpful discussions.

- Richardson, J. (1981) *Adv. Protein Chem.* **34**, 168–339.
- Lesk, A. M. & Chothia, C. J. (1980) *J. Mol. Biol.* **136**, 225–270.
- Remington, S. J. & Matthews, B. W. (1980) *J. Mol. Biol.* **140**, 77–99.
- Salemme, F. R. (1977) *Annu. Rev. Biochem.* **46**, 299–329.
- Novotny, J., Brucoleri, R. & Karplus, M. (1984) *J. Mol. Biol.* **177**, 787–818.
- Kretsinger, R. H. & Barry, C. D. (1975) *Biochim. Biophys. Acta* **405**, 40–52.

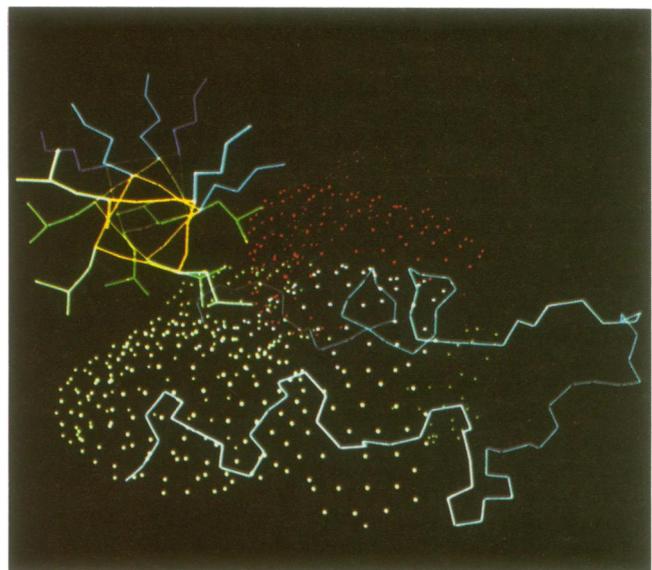


FIG. 4. Proposed docking orientation for the basic amphiphilic peptide (Leu-Lys-Lys-Leu-Leu-Lys-Leu)<sub>2</sub> to the second domain of the ICB-based model for calmodulin. Hydrophobic residues are green, basic residues blue, and acidic residues red for the peptide and the surface. The surface shown is a solvent-accessible surface generated for those protein residues in the binding site that are capable of contributing to stabilization of the peptide-protein complex. In this orientation there is the potential for both hydrophobic and ionic interactions.

- Kretsinger, R. H. (1980) *Ann. N.Y. Acad. Sci.* **356**, 14–19.
- Aulabaugh, A., Niemczura, W. P., Blundell, T. L. & Gibbons, W. A. (1984) *Eur. J. Biochem.* **143**, 409–418.
- Greer, J. (1981) *J. Mol. Biol.* **153**, 1027–1042.
- Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47**, 45–149.
- Klee, C. B. & Vanaman, T. C. (1982) *Adv. Protein Chem.* **35**, 213–321.
- Kretsinger, R. H., Rudnick, S. E., Sneden, D. A. & Schatz, V. B. (1980) *J. Biol. Chem.* **255**, 8154–8156.
- Szebenyi, D. M. E., Obendorf, S. K. & Moffat, K. (1981) *Nature (London)* **294**, 327–332.
- Moews, P. C. & Kretsinger, R. H. (1975) *J. Mol. Biol.* **91**, 201–228.
- Cook, W. J., Dedman, J. R., Means, A. R. & Bugg, C. E. (1980) *J. Biol. Chem.* **255**, 8152–8153.
- Gehrig, L. M. B., Delbaere, L. T. J. & Hickie, R. A. (1984) *J. Mol. Biol.* **177**, 559–561.
- Crippen, G. M. (1981) in *Chemometric Research Studies Series*, ed. Baudan, D. (Wiley, New York), Vol. 1, pp. 1–58.
- Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. & Weiner, P. (1984) *J. Am. Chem. Soc.* **106**, 765–784.
- Connolly, M. L. (1981) *Quantum Chem. Program Exchange Bull.* **1**, 75.
- Weiner, P. K., Langridge, R., Blaney, J. M., Schaefer, R. & Kollman, P. A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 3754–3758.
- Newton, D. L., Oldewurtel, M. D., Krinks, M. H., Shiloach, J. & Klee, C. B. (1984) *J. Biol. Chem.* **259**, 4419–4426.
- Sundralingam, M., Bergstrom, R., Strasburg, G., Rao, S. T., Roychowdhury, P., Greaser, M. & Wang, B. B. (1985) *Science* **227**, 945–948.
- Herzberg, O. & James, M. N. G. (1985) *Nature (London)* **313**, 653–659.
- Eisenberg, D., Weiss, R. M., Terwilliger, T. C. & Wilcox, W. (1982) *Faraday Symp. Chem. Soc.* **17**, 109–120.
- Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng.* **6**, 151–176.
- Chothia, C. (1975) *Nature (London)* **254**, 304–308.
- Malencik, D. A. & Anderson, S. R. (1984) *Biochemistry* **23**, 2420–2428.
- Cox, J. A., Comte, M., Fitton, J. E. & DeGrado, W. F. (1985) *J. Biol. Chem.* **260**, 2527–2534.
- Malencik, D. A. & Anderson, S. R. (1982) *Biochemistry* **21**, 3480–3486.
- Matthew, J. B., Weber, P. C., Salemme, F. R. & Richards, F. M. (1983) *Nature (London)* **301**, 169–171.
- Getzoff, E. D., Tainer, J. A., Weiner, P. K., Kollman, P. A., Richardson, J. S. & Richardson, D. S. (1983) *Nature (London)* **306**, 287–330.