Further Insights into Calmodulin−Myosin Light Chain Kinase Interaction from Solution Scattering and Shape Restoration†

William T. Heller,‡,§ Joanna K. Krueger,‖ and Jill Trewhella*,‡

Bioscience Division, MS M-888, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, and Chemistry Department, University of North Carolina at Charlotte, Charlotte, North Carolina 28223

Received May 22, 2003; Revised Manuscript Received July 21, 2003

ABSTRACT: We have gained new insight into the interactions between the second-messenger protein calmodulin (CaM) and myosin light chain kinase from skeletal muscle (skMLCK) using small-angle solution scattering and shape restoration. Specifically, we explored the nature of a $2Ca^{2+}$-CaM−skMLCK complex and compared it to a $4Ca^{2+}$-CaM−skMLCK complex under the same conditions. The $2Ca^{2+}$ complex has been proposed to be physiologically relevant. To aid in the interpretation of the data, we developed a shape restoration approach, implemented in GA_STRUCT, that combines many of the best features of other available methods into a single, automated package. Importantly, GA_STRUCT explicitly addresses the problem of the existence of multiple solutions to the inverse scattering problem and produces a consensus envelope from a set of shapes that fit the input intensity. Small-angle scattering intensity profiles measured or calculated from known structures were used to test GA_STRUCT, which was then used to generate low-resolution models for three complexes: $2Ca^{2+}$-CaM−skMLCK, $4Ca^{2+}$-CaM−skMLCK, and $4Ca^{2+}$-CaM−skMLCK with a bound substrate. These models were used in conjunction with high-resolution structures of the protein components to better understand the interactions among them. In the case of the $2Ca^{2+}$-CaM−skMLCK complex, the consensus envelope is consistent with CaM in a fully collapsed state with its two globular lobes in close contact with each other while the catalytic cleft of the kinase is open. The consensus envelope for the $4Ca^{2+}$-CaM−skMLCK complex indicates that the collapsed CaM has swung further away from the open catalytic cleft of the skMLCK than in the $2Ca^{2+}$ complex, and further that substrate binding to this complex results in closure of the kinase catalytic cleft, in agreement with previous neutron scattering results. These results indicate that activation of MLCK by CaM can only occur once CaM is fully translocated away from the catalytic cleft, which is presumably linked to full release of the pseudo-substrate/inhibitory sequence. Our scattering data indicate that this step is completed only when all four calcium binding sites are loaded.

Calmodulin (CaM)[1] is a dumbbell-shaped protein that is a major receptor for intracellular calcium signals in eukaryotic systems and is responsible for mediating the activation of a wide variety of enzymes, including a number of kinases. CaM's interaction with myosin light chain kinase (MLCK) serves as a model for the activation of protein kinases by Ca^{2+}-CaM. We have previously studied CaM and its interaction with a catalytically active skeletal MLCK (skMLCK) using small-angle scattering (*1*−*4*). CaM adopts a dumbbell shape in solution (*5*) similar to the crystal structure (*6*), but its two Ca^{2+}-binding lobes are closer together, inferring that the helical connector between them is flexible in solution (*5*, *7*). Subsequent structural studies have affirmed the importance of this flexibility in allowing $4Ca^{2+}$-CaM to form a tightly collapsed structure in which its globular lobes come into close contact about the helical binding domains of its various targets, including MLCK (*8*−*10*).

Skeletal muscle MLCK has a C-terminal autoinhibitory sequence that includes a calmodulin binding sequence and interacts directly with the enzyme's active site to inhibit activation. The autoinhibitory sequence forms an extensive network of interactions with the surface of the kinase (*11*−*13*). The active site sits in a catalytic cleft between a small and a large domain that, based on studies of the CaM−skMLCK complex (*2*) and the homologous cPKA (*14*), can open and close about a glycine hinge. In the absence of Ca^{2+}, skMLCK is autoinhibited and does not bind CaM. Using neutron scattering with contrast variation, we showed that $4Ca^{2+}$-CaM binds to skMLCK at a site distant from the catalytic cleft on the C-terminal lobe of skMLCK, and CaM is in a fully collapsed conformation around the CaM-binding sequence of the kinase (*1*). Our studies also showed that the

† This work was performed under the auspices of the Department of Energy under contract to the University of California (Contract W-7405-ENG-36) and was supported by Department of Energy/BER Project KP1101010 in support of the Oak Ridge National Laboratory Center for Structural Molecular Biology and National Institutes of Health Grant GM40528 (J.T.).

* To whom correspondence should be addressed. Phone: (505) 667-2690. Fax: (505) 667-2891. E-mail: jtrewhella@lanl.gov.

‡ Los Alamos National Laboratory.

§ Present address: Condensed Matter Sciences Division and Center for Structural Molecular Biology, Oak Ridge National Laboratory, Oak Ridge, TN 37831.

‖ University of North Carolina at Charlotte.

[1] Abbreviations: CaM, calmodulin; skMLCK, skeletal muscle myosin light chain kinase; MMO, hydroxylase component of methane monooxygenase; cPKA, catalytic subunit of cAMP-dependent protein kinase A; R_g, radius of gyration; d_{max}, maximum linear dimension.

10.1021/bi0348664 CCC: $25.00
Published on Web 08/22/2003

kinase contracts in response to substrate binding, most likely due to closure of the catalytic cleft, while CaM reorients so that its N-terminal sequence interacts with the surface of the kinase (*2*). Our X-ray scattering data on substoichiometric Ca^{2+}-CaM indicate that complex formation with the skMLCK enzyme begins when <2 mol of Ca^{2+}/mol of CaM is present and is fully complexed somewhere between 2 and 3 mol of Ca^{2+}/mol of CaM (3). We did not attempt to model the $2Ca^{2+}$ complex at the time of these studies. A $2Ca^{2+}$-CaM−skMLCK intermediate has been suggested to be important in the MLCK activation mechanism based on kinetic studies and biophysical studies with peptides and enzymes (*15*−*17*), but no one has yet demonstrated that this complex forms at physiological protein concentrations.

Shape restoration from small-angle scattering data holds the promise of providing more detailed structural information than the two-ellipsoidal modeling approach previously applied to the CaM−skMLCK complex (*1*−*3*) because there are no assumptions about the basic shape. A number of shape restoration methods have become available recently, e.g., using spherical harmonics (*18*−*25*) or aggregates of spheres (*26*−*30*). Many current methods use a large number of degrees of freedom to reconstruct the shape of a scattering object in more detail. The problem of shape restoration for solution scattering data is particularly complex because the rotationally isotropic nature of the samples results in a one-dimensional (1D) scattering intensity profile. For this reason, the uniqueness of a three-dimensional (3D) structure associated with a 1D scattering profile cannot be guaranteed and multiple shapes can result from shape restoration methods.

In this paper, we use shape restoration to evaluate scattering data on the CaM−skMLCK complexes in different states of bound Ca^{2+} and substrate. Our approach to shape restoration uses aggregates of spheres to model the shape of the scattering particle while explicitly addressing the problem of multiple solutions in a single integrated package. We present a number of tests of our approach that demonstrate its utility for producing the general shape of a scattering particle, as well as potential traps for the user. We then apply the method to three different CaM−skMLCK complexes: $2Ca^{2+}$-CaM−skMLCK, $4Ca^{2+}$-CaM−skMLCK, and $4Ca^{2+}$-CaM−skMLCK with substrate. Both Ca^{2+}-saturated structures reveal a collapsed CaM, and the structure in the presence of a substrate strongly suggests that the catalytic cleft of skMLCK is closed, as expected and previously reported on the basis of our neutron contrast variation studies (*1*−*4*). The model resulting from the $2Ca^{2+}$-CaM−skMLCK scattering data indicates that the CaM is also in a fully collapsed conformation, but its position and orientation with respect to the skMLCK differ from those of the Ca^{2+}-saturated complexes such that the autoinhibitory sequence is not fully removed from its binding site on the kinase.

MATERIALS AND METHODS

Sample Preparation and Scattering Experiments. MMO was prepared according to previously described methods (*31*). X-ray scattering data for MMO were collected using a laboratory-based instrument at Los Alamos National Laboratory, which employs a sealed tube Cu Kα source (wavelength of 1.54 Å) (*5*). The instrument has a slit geometry and a one-dimensional position sensitive detector. The data were reduced according to previously described procedures into $I(q)$ versus q [$q = (4\pi \sin \theta)/\lambda$, 2θ is the scattering angle, and λ is the wavelength] (*5*). The CaM−skMLCK data were taken from previously published experiments: ref *3* for the $2Ca^{2+}$-CaM−skMLCK complex, ref *1* for the $4Ca^{2+}$-CaM−skMLCK complex without substrates, and ref *2* for the $4Ca^{2+}$-CaM−skMLCK complex with substrates (myosin regulatory peptide, KKRAARATSNVFS, and a nonhydrolyzable analogue of ATP, AMPPNP). Structural parameters such as the radius of gyration (R_g) and the maximum linear dimension (d_{max}) were calculated from $P(r)$, the probable frequency distribution of vector lengths between all scattering centers in the scattering particle.

Shape Restoration. The shape restoration approach implemented here is, in some respects, an amalgamation of a variety of existing approaches incorporated into a single program called GA_STRUCT.[2] Our approach employs aggregates of spheres to define the volume of the scattering object. It differs from other aggregate approaches (*26*−*30*) in that the spheres are not forced to have a single radius or lie on a grid. Unlike other techniques, the number of spheres is fixed and is related to the expected volume of the scattering particle. The scattered intensity profile is calculated using a Monte Carlo approach implemented previously (*5*) that first calculates $P(r)$. Then, $I(q)$ is calculated by the Fourier transform defined in eq 1.

$$I(q) = \int_0^\infty P(r)\frac{\sin(qr)}{qr}\,dr \qquad (1)$$

This method differs from the Debye formula for calculating the SAS intensity of a collection of nonoverlapping spheres (*32*) because the density of the interior of the scattering object is truly uniform. The spheres used by GA_STRUCT can overlap, thereby eliminating the internal gaps in the particle volume resulting from sphere packing. The Monte Carlo intensity calculation also has an advantage over that of the spherical harmonic shape restorations (*18*−*25*) in that a constraint that the scattering length density be greater than zero is not required for calculation of $I(q)$.

Nonlinear minimization is accomplished through a genetic algorithm (*33*), as employed by Chacón and co-workers (*26*, *27*), but the details of the algorithm differ significantly. Unlike other available methods using aggregates of spheres (*26*−*30*), several independent runs of the minimization process are automatically performed to generate a family of structures. This family is then characterized for similarity, and a consensus envelope is produced from the set of structures that represents the most common structural features of the family. The details of the shape restoration process follow.

The population fed into the genetic algorithm consists of 50 models having the same number of spheres. The average radius of the spheres (r_{ave}) is given by the resolution of the data that have been collected ($r_{ave} = \pi/q_{max}$, where q_{max} is the maximum q value to be fit). The number of spheres is calculated using r_{ave} and either the user-supplied expected volume or the molecular weight, which is related to the

[2] GA_STRUCT is available upon request from the authors, who can be contacted at hellerwt@ornl.gov. Executables have been compiled for the Windows 98/2000/XP, Linux, and SGI Irix platforms.

volume by the partial specific volume [0.74 cm^3/g (*34*)]. Between 5 and 10 spheres are added, depending on the volume, to provide flexibility in the modeling. The initial 50 structures are randomly oriented chains of spheres constrained to lie in a box with sides of lengths equal to d_{max}. The radii of the initial spheres are randomly distributed within 30% of r_{ave}, with an empirically determined minimum of 5 Å, which is enforced throughout the minimization process. The model intensity profiles are calculated from the original population using between 2000 and 8000 points for the Monte Carlo integration. The initial population is then sorted in ascending values of the fitting parameter *F*, given by eq 2.

$$F = \frac{1}{N_{pts}}\left\{\sum_{N_{pts}} \frac{[I(q) - I_m(q)]^2}{\sigma(q)^2}\right\} \quad (2)$$

N_{pts} is the number of points in the experimental intensity, $I(q)$, which has experimental uncertainty $\sigma(q)$. $I_m(q)$ is the model intensity. The fitting parameter is a modification of the reduced χ^2 that does not take the number of parameters used for the fit into account.

The genetic algorithm employs three operations to find a model that fits the experimental intensity profile: mating, mutation, and extinction. In the mating process, pairs of models from the best five are used to make two new models by swapping one-third of the spheres from each. The coordinates of the spheres remained unaltered in the two "daughter" models, but the radii are adjusted such that the final volume is contiguous, as one would expect for a protein. In the mutation events, which happen every 10 generations, one sphere in each model is moved and given a new radius such that it remains in contact with the remainder of the model. The radii of the other spheres are also adjusted such that the volume remains contiguous. The third operation is extinction, in which all but the best model are replaced with new models made exactly as during the population initialization. Additionally, the best model is subjected to 500 perturbations where a randomly chosen sphere is moved slightly to see if the fit to the data can be improved. After every operation in the genetic algorithm, the intensity profiles of new or altered models are calculated and the population is sorted according to the ascending value of *F*. Once a best fit model has been found, GA_STRUCT either starts the genetic algorithm over to find another structure or moves on to characterize the similarity of the family of structures that have been generated (typically 10−25).

The last two steps performed by GA_STRUCT are intended to characterize the reproducibility of the shape restoration and to provide the user with an "average" shape, called the consensus envelope. While similar averaging and filtering approaches have been implemented and employed previously (*26*−*30*, *35*−*39*), this process is an integral part of shape restoration by GA_STRUCT. To check the population for similarity, all models are docked onto one another pairwise. The models are placed on a grid with spacing related to the d_{max} of the particle such that the centers of mass coincide. The rotation angles that maximize the amount of overlap between the two models are found by Monte Carlo sampling. The overlap is calculated as the fraction of grid points occupied by both models divided by the total number of grid points occupied. All possible mirror symmetries are checked because scattering cannot differentiate a particle from its mirror image.

Table 1: Structural Parameters of Consensus Particle Envelopes for Test Structures[a]

test structure	R_g (Å)	d_{max} (Å)	*F* range
lysozyme	14.5 (14.3)	48.3 (45)	0.004−0.010
lysozyme with noise at 100%	14.4 (14.3)	43.6 (45)	0.355−0.395
lysozyme with noise at 500%	14.7 (14.3)	46.7 (45)	7.651−7.668
MMO	38.5 (41.1)	130 (130)	0.118−0.297
MMO experimental data	40.5 (41.5)	133 (130)	1.293−1.360
ribonuclease inhibitor	25.8 (24.3)	76.0 (80)	0.002−0.020
$2Ca^{2+}$-CaM−skMLCK	32.8 (32.9)	115 (105)	0.607−0.649
$4Ca^{2+}$-CaM−skMLCK	34.1 (33.9)	117 (125)	0.846−0.892
$4Ca^{2+}$-CaM−skMLCK with substrate	31.6 (31.2)	108 (105)	0.960−1.000

[a] The *F* range gives the range of values for the family of structures produced by GA_STRUCT (see eq 2 in Materials and Methods). Values in parentheses were calculated from the crystal structures [lysozyme (*40*), MMO (*41*), and ribonuclease inhibitor (*42*)] or determined from experimental data (for MMO and the CaM−MLCK complexes). All calculations were carried out with intensities calculated from the crystal structure coordinates except for the indicated MMO experimental data set.

The model with the highest total docking score, being defined as the sum of its docking scores with the all of the other members of the population, is used as the basis for determining the consensus envelope. The consensus envelope is constructed from the model and 70% of the remaining models having the best overlap with it. The optimal rotation angles from the docking step are used to overlay the retained models on a grid, again having a spacing related to d_{max}. The grid points occupied by more than half of the models that make up the consensus envelope are output in the final step of the program. It is important to note that the consensus envelope is not necessarily the "best fit" to the scattering data. The consensus envelope only represents those features most frequently emerging in the population of best fit models. An evaluation of how well the consensus envelope represents this family is made by reviewing the individual members of the family. For a relatively simple shape, GA_STRUCT can find a family of 10 structures and calculate the consensus envelope in roughly 12 h on a 500 MHz Pentium III Windows personal computer when using 4000 points for the Monte Carlo intensity calculation.

Several tests of the effectiveness of GA_STRUCT were carried out using scattered intensity profiles calculated from known structures using the program PRPDB (*31*). One test used lysozyme, a small protein with a catalytic cleft [PDB entry 6LYZ (*40*)]. Trials were performed on the noise-free simulated intensity profile and with noise added at 100 and 500% of the simulated experimental uncertainty, which is proportional to the square root of the intensity. Other tests employed the hydroxylase component of methane monooxygenase (MMO), a large disklike protein [PDB entry 1MMO (*41*)], and the C-shaped ribonuclease inhibitor protein [PDB entry 2BNH (*42*)]. The MMO tests used a simulated intensity profile and experimental data collected as described in ref *31*. Each test consisted of generation of a family of 10 independent models and the consensus envelope.

GA_STRUCT then was applied to existing experimental intensities collected for the CaM−skMLCK complex under the three calcium and substrate conditions (*1*−*3*). Each

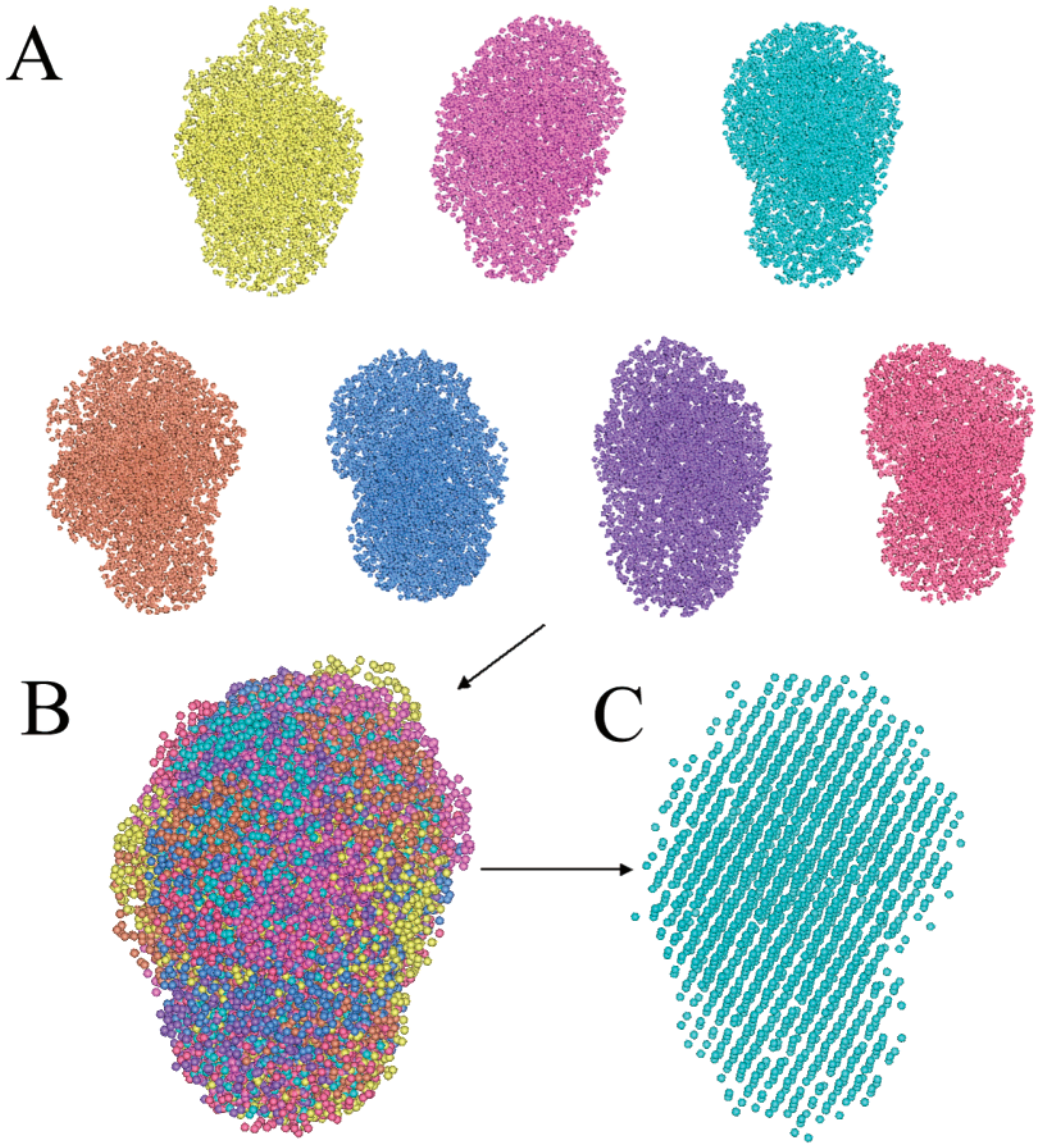

FIGURE 1: Steps involved in the envelope generation process. (A) Seven (of ten) structures from the family of structures that were produced for the lysozyme test found to be most similar. (B) Seven docked structures overlaid onto one another. (C) Final consensus envelope.

consensus envelope was generated from 25 final structures, and 4000 points were used for the Monte Carlo integration. The consensus envelopes were used in conjunction with existing high-resolution structures to develop better models of the interaction. The coordinates in PDB entry 2BBK (*43*) were used to represent the collapsed state of CaM. For the extended state of CaM, 3CLN was used (*11*). There is no crystal structure of skMLCK, so the homologous cPKA [PDB entry 2CPK (*44*)] was substituted for the catalytic core of skMLCK. The core accounts for only 87% of skMLCK, as 32 and 29 residues are "missing" from the model at the N- and C-termini, respectively. Also, this crystal structure of cPKA has a "closed" catalytic cleft, with a peptide pseudosubstrate bound. Our previous experiment found that the "open" state of the kinase can be approximated using 2CPK with the cleft opened 39° about the glycine hinge (*14*). The high-resolution structures were fit into the consensus envelopes manually. Constraints requiring CaM to interact with the large lobe of the cPKA and the catalytic cleft to be directed toward the CaM (*4*) were used as an aid in building these models.

RESULTS

Shape Restoration from Simulated and Experimental Data from Known Structures. Shape restoration tests were performed for lysozyme, MMO, and the ribonuclease inhibitor protein. The structural parameters derived from each test are summarized in Table 1. It is clear that the consensus envelopes reproduce R_g and d_{max} values quite well (within 5−10%), including those tests done with data containing noise, either simulated or from experiment. The set of the most similar structures from the family of 10 generated by GA_STRUCT for the lysozyme intensity profile without noise is shown in Figure 1A. The degree of variability of the individual structures is typical of all tests performed on particles having a roughly globular character. The results of the docking of the seven structures are shown in Figure 1B, and Figure 1C is the consensus envelope that we will see has the aspect ratio and shape characteristic of lysozyme, although the catalytic cleft is not well resolved, as might be expected from the resolution of the scattering data.

Figure 2 shows the fits of the respective model data calculated using GA_STRUCT and either simulated or

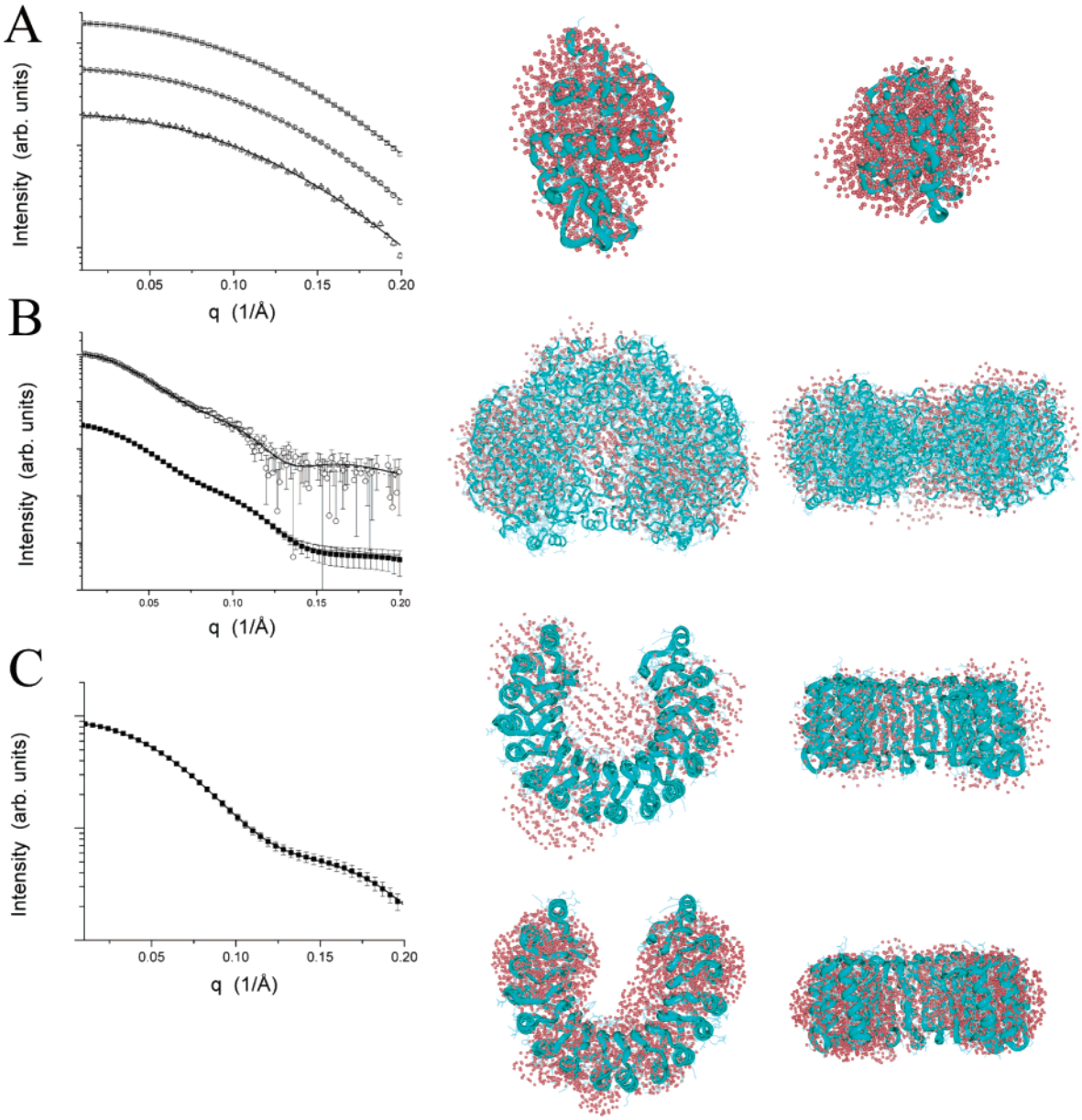

FIGURE 2: Results of the shape restoration tests. (A) Lysozyme. Intensity profiles calculated from the lysozyme crystal structure plotted with sample intensities resulting from the shape restoration. The curves are as follows: (□) noiseless data, (○) noise at 100% uncertainty, and (△) noise at 500% uncertainty. The consensus envelope produced from the noiseless intensity profile overlaid onto the crystal structure (*40*) is shown in two orthogonal views to the right of the plot. (B) MMO. Three sample intensities produced by GA_STRUCT are shown with the input data. The lower data set (■) was produced from the crystal structure (*41*). The upper curve (○) was collected at Los Alamos National Laboratory. Two orthogonal views of the MMO consensus envelope resulting from the simulated intensity are shown in the same row docked into the original structure (*41*). (C) Ribonuclease inhibitor. Sample intensities produced by the program are plotted along with the simulated data. Two orthogonal views of the resulting consensus envelope are shown docked onto the original structure (*42*) in the top pair of images to the right of the plot. The two images below are of the single model resulting from the shape restoration having the correct C shape.

experimental scattering data for lysozyme (Figure 2A), MMO (Figure 2B), and the ribonuclease inhibitor protein (Figure 2C). Each shape restoration gave excellent fits to the intensity profiles shown in the left column of Figure 2, as might be expected from the fidelity with which the R_g and d_{max} were reproduced (Table 1). The middle and right columns show two views of the consensus envelope docked onto the original protein structure. In the case of the ribonuclease inhibitor protein, two structures are shown: the consensus envelope (top) and a single model not included in the consensus envelope (bottom). Both the lysozyme and MMO consensus envelopes well reproduce the shapes of the crystal structures of these enzymes, which are globular and disklike, respectively. In contrast, the consensus envelope generated by GA_STRUCT for the ribonuclease inhibitor protein does not reproduce the expected shape (Figure 2C, top images). Inspection of the individual structures saved by GA_STRUCT (not shown) reveals that only one has the C shape of the crystal structure (Figure 2C, bottom images). A T shape, much like the consensus envelope shown in the views of the top structure of Figure 2C, resulted from eight of the reconstruction runs. Additionally, an intermediate structure with a V shape was found. The C-shaped structure was eliminated by the procedure used to generate the consensus envelope because of a lack of sufficient overlap with the majority of the structures that were found. It is

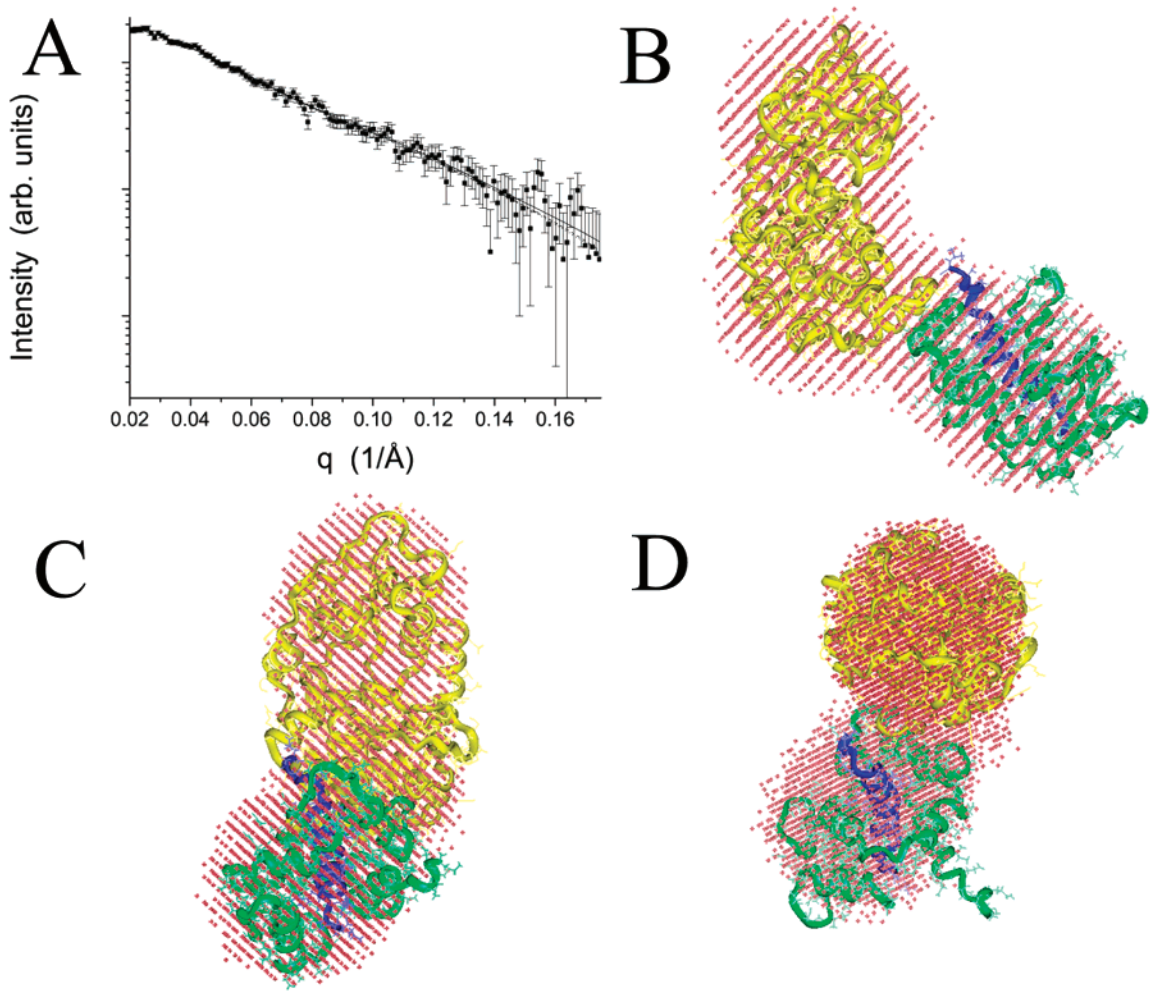

FIGURE 3: Results of shape restoration for the $2Ca^{2+}$-CaM−skMLCK data (*3*). (A) The data are shown with three sample model intensity profiles produced by GA_STRUCT. (B−D) Three orthogonal views of the extended cPKA (*14*, *44*) and collapsed CaM [2BBK (*43*)] structures docked by hand into the consensus envelope from GA_STRUCT. The yellow molecule is the open cPKA structure (*14*, *44*), the green the CaM, and the blue the MLCK-I peptide from the 2BBK structure (*43*).

interesting that each of the structures that failed the overlap test, and hence was not used for the consensus envelope, had *F* values that fall within the lower third of the range shown in Table 1. These tests of GA_STRUCT demonstrate three things. (1) If the shape is sufficiently simple, the probability of the consensus envelope representing the actual shape well is high. (2) For complex shapes, it is possible and even likely that distinct 3D shapes will fit the 1D small-angle scattering intensity profile equally well. (3) *F* values speak to the quality of the fit of the model intensity profile to the data but may not be useful in discriminating between different classes of shapes that share the same overall R_g value.

Shape Restoration for CaM−skMLCK Complexes. Table 1 summarizes the R_g and d_{max} values obtained from GA_STRUCT for the CaM−skMLCK complexes. These values agree well with those obtained from $P(r)$ analyses of the experimental data. Figures 3−5 each show the fit of three representative model intensity profiles to the measured intensity profile for the $2Ca^{2+}$-CaM−skMLCK complex, the $4Ca^{2+}$-CaM−skMLCK complex, and the $4Ca^{2+}$-CaM−skMLCK complex with substrate, respectively, along with three orthogonal views of the respective consensus envelopes derived by GA_STRUCT. The high-resolution models made by fitting the known structures of the components into the consensus envelope are also shown docked into the consensus envelope.

The consensus envelope for the $2Ca^{2+}$ complex is essentially a bent cylinder with one end slightly longer than the other. The length is ∼115 Å, and the width, which is roughly constant along the length, is ∼42 Å. The high-resolution component structures of a collapsed CaM and the open catalytic core fit reasonably well into the consensus envelope. It was not possible to fit both the kinase and an extended CaM structure within the boundary of the consensus envelope (not shown). As can be seen in Figure 3C, the most natural placement for the CaM is one in which it is not directly in front of the catalytic cleft, as we observed for the bound delNCaM mutant that is missing residues from the N-terminal helix of CaM and is deficient in its ability to activate skMLCK (*4*). Rather, in the $2Ca^{2+}$ complex, CaM has swung off more to the side of the catalytic cleft.

For the $4Ca^{2+}$ complex without substrate bound, the length of the consensus envelope is ∼117 Å and the width is ∼44 Å. The bend in the cylinder is less pronounced than that observed for the $2Ca^{2+}$ complex, and the structure is somewhat more symmetrical about the bend. The open catalytic core and collapsed CaM structures fit into the consensus envelope. Again, the extended CaM structure could not be fit into the consensus envelope with the catalytic

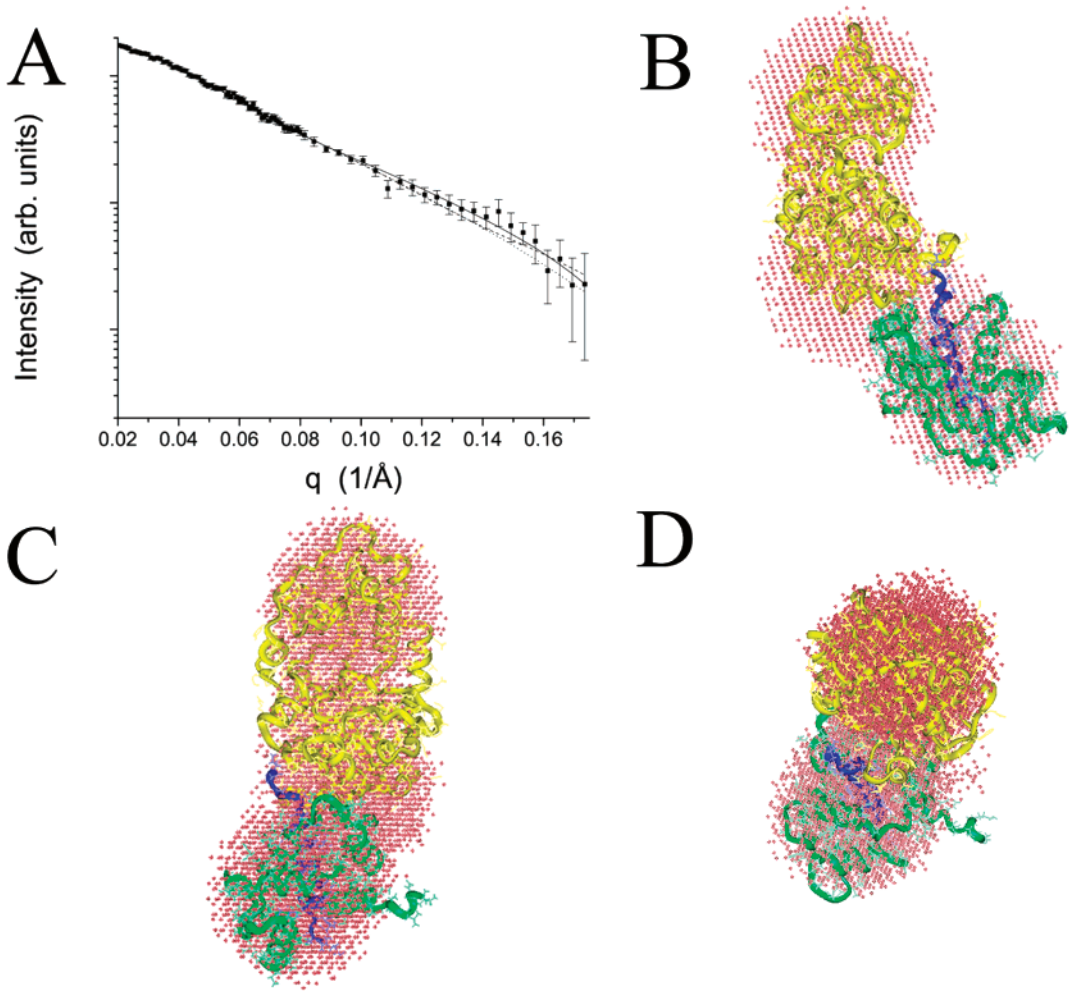

FIGURE 4: Results of shape restoration for the $4Ca^{2+}$-CaM−skMLCK data in the absence of substrate (*1*). (A) The data are shown with three sample model intensity profiles produced by GA_STRUCT. (B−D) Three orthogonal views of the extended cPKA (*14*, *44*) and collapsed CaM structures docked by hand into the consensus envelope from GA_STRUCT. The yellow molecule is the open cPKA structure (*14*, *44*), the green the CaM, and the blue the MLCK-I peptide from the 2BBK structure (*43*).

core. Further, the fit to the open catalytic core was more optimal than to the corresponding closed structure (not shown). As was found for the $2Ca^{2+}$ complex, the CaM is offset from the natural axis defined by the cPKA cleft, but it has translocated further away from the catalytic cleft.

The $4Ca^{2+}$ complex with the bound substrate also gives a consensus envelope that is a bent cylinder, but it is significantly more compact than the other structures, having a length of 108 Å and a width of 42 Å. The angle between the segments is similar to that of the $4Ca^{2+}$-CaM−skMLCK complex without substrate structure. The main difference is that one arm of the bent cylinder is significantly shorter than in the $4Ca^{2+}$-CaM−skMLCK complex without substrate. Fitting the crystal structures into the consensus envelope required that the collapsed CaM (*40*) and the closed 2CPK (*41*) structures be used. The open cPKA crystal structure extends significantly beyond the bounds of the consensus envelope (not shown).

DISCUSSION

Our tests of GA_STRUCT with known structures demonstrate that our algorithm for shape restoration from small-angle scattering data is effective at reconstructing the low-resolution structure of globular proteins. The general dimensions, aspect ratios, and shape of the particle are faithfully reproduced for relatively simple shapes. Details, such as the catalytic cleft of lysozyme, are not resolved, although the general "light bulb" shape of that protein is faithfully reproduced. As the tests using the ribonuclease inhibitor structure (*41*) demonstrate, structurally dissimilar, degenerate solutions are likely for complex shapes. Degenerate solutions were also recently observed (*39*) in tests of the program DAMMIN (*28*). In these cases, additional sources of experimental data are required to provide constraints that can disqualify "false" solutions. Analyzing scattering data for $P(r)$ provides a good indication of the complexity of the particle shape and enables the user to determine the extent upon which shape restoration by GA_STRUCT or any other approach can be relied.

The CaM−skMLCK structures are relatively simple, bent rodlike structures and were therefore reproduced quite well across the family of structures generated for the consensus envelope. We are confident therefore that the resulting particle envelope faithfully represents the low-resolution structures of the complexes. None of the best-fit structures derived by GA_STUCT supported a complex between an extended CaM and skMLCK. The models determined from the data of the $4Ca^{2+}$-CaM−skMLCK complex with sub-

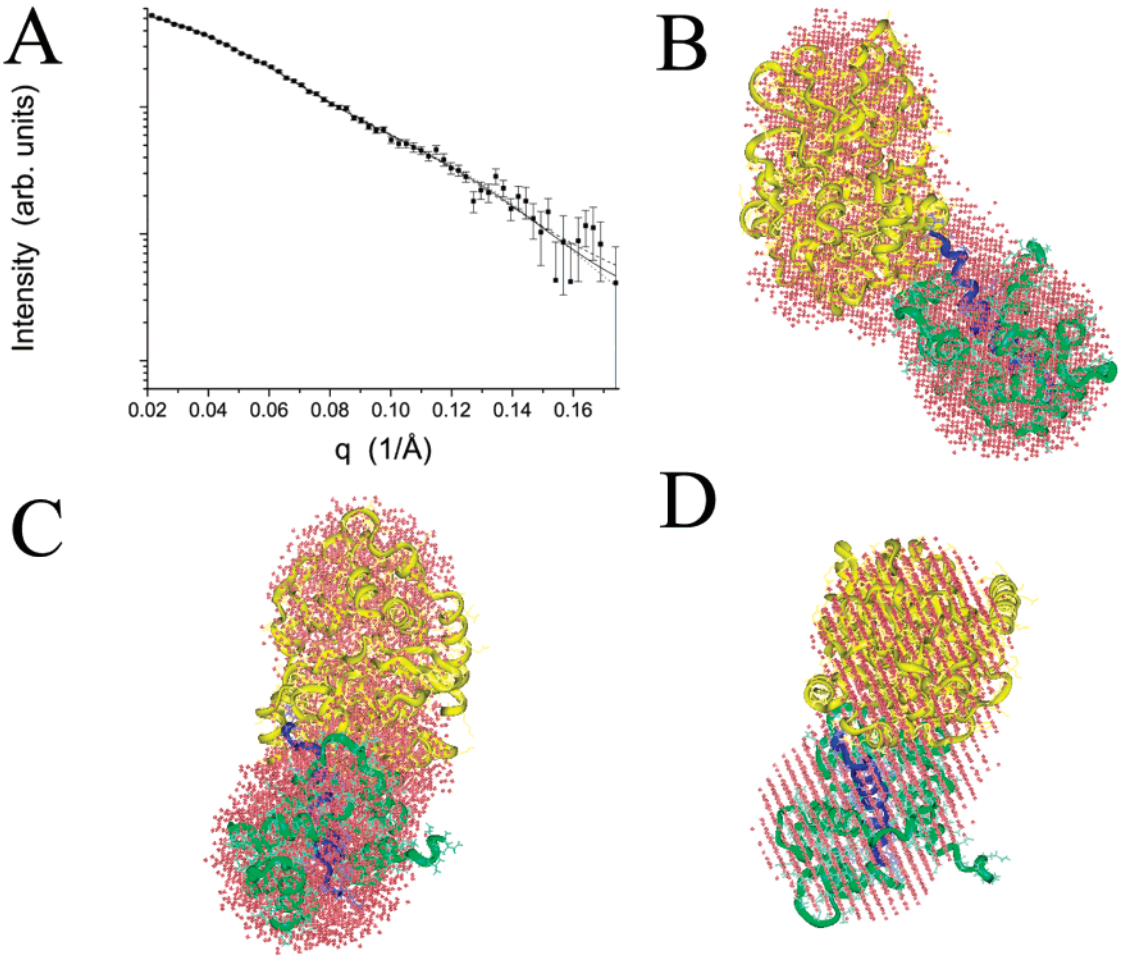

FIGURE 5: Results of shape restoration for the $4Ca^{2+}$-CaM−skMLCK data in the presence of substrate (*2*). (A) The data are shown with three sample model intensity profiles produced by GA_STRUCT. (B−D) Three orthogonal views of the closed cPKA (*44*) and collapsed CaM (*43*) structures docked by hand into the consensus envelope from GA_STRUCT. The yellow molecule is the closed cPKA structure [2CPK (*44*)], the green the CaM, and the blue the MLCK-I peptide from the 2BBK structure (*43*).

strate are similar to the results obtained with the original small-angle X-ray and neutron scattering study (*1*). A detailed comparison with the neutron-derived models was not carried out because the manual fitting of the structures into the low-resolution shapes defined by the scattering does not have the precision to make such a comparison particularly meaningful. Nonetheless, it is impressive that the results of modeling the single X-ray scattering profiles provide as much detail as they do in these studies, including evidence for the closing of the kinase catalytic cleft upon substrate binding. Overall, the neutron contrast variation studies with deuterium labeling of the CaM would be expected to more precisely define the dispositions of the components in the complexes.

The consensus envelope modeled from the $2Ca^{2+}$-CaM−skMLCK data is consistent with CaM adopting the collapsed state in the $2Ca^{2+}$-CaM−skMLCK complex. CaM is located away from the catalytic cleft, but it is not fully translocated to the position occupied by the Ca^{2+}-saturated complex. This can be best seen in Figure 6, where the two hand-produced high-resolution models are overlaid such that the cPKA structures are coincident. The CaM moves further away from the cleft when all four Ca^{2+}-binding sites in CaM are occupied. The fact that the open cleft cPKA structure best fits the $2Ca^{2+}$-CaM−skMLCK complex suggests that the autoinhibitory sequence is at least partially released in this complex. The skMLCK portion of the consensus envelope in the $2Ca^{2+}$ state is less extended than that of the $4Ca^{2+}$ state without substrate, suggesting that the catalytic cleft in the $2Ca^{2+}$ state may not be as open as it is in the Ca^{2+}-saturated complex. The data lack the resolution to quantify this however. Complete removal of the autoinhibitory sequence that permits unhindered access by the substrate to the catalytic cleft may only be achieved in the fully translocated position observed in the Ca^{2+}-saturated complexes. This idea is supported by the fact that only the Ca^{2+}-saturated CaM is capable of fully activating a number of target enzymes (*45*−*48*).

It remains something of an open question in this study as to which Ca^{2+}-binding sites are occupied in the $2Ca^{2+}$-CaM−skMLCK complex. The difference in Ca^{2+} affinities between the four sites in free CaM is relatively small, the N-terminal so-called regulatory sites having approximately 6-fold lower affinities (*49*). However, the binding of different target peptides or proteins has been observed to increase this difference by as much as 2 orders of magnitude and also have an impact on the kinetics of Ca^{2+} binding (*15*, *16*). Whatever the specifics of the binding affinities and kinetics, we know from the forward scattering data that all of the CaM is bound to the kinase with only two Ca^{2+} atoms per CaM, and that independent of which and how many sites are filled, the scattering data show that CaM is fully collapsed, implying that the two globular lobes are interacting with the helical

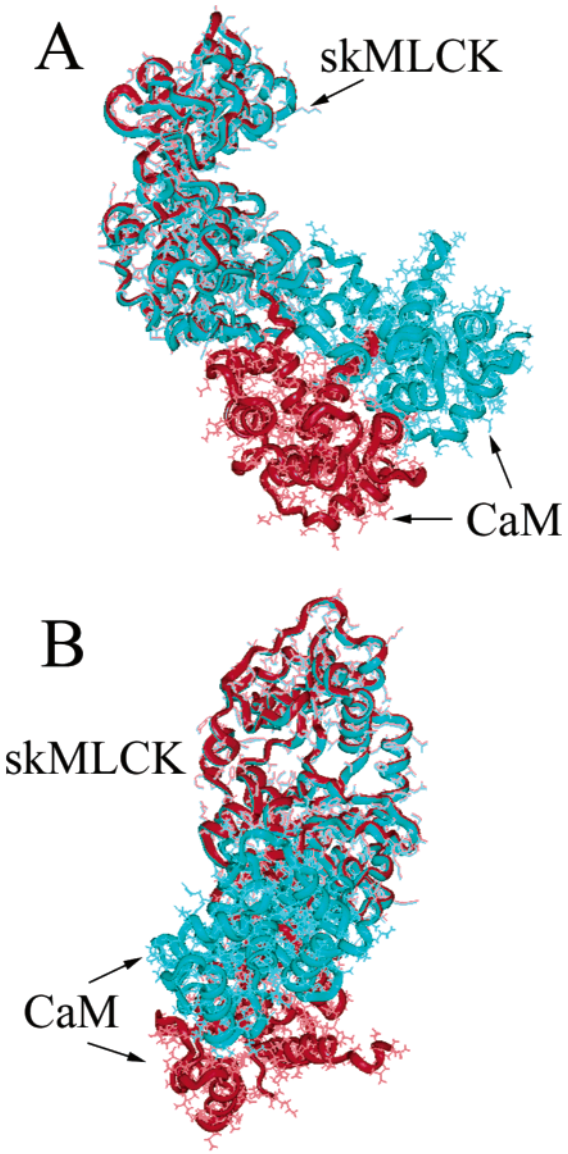

FIGURE 6: High-resolution structures resulting from the docking of the extended cPKA structure (*14*, *44*) and collapsed CaM structure [2BBK (*43*)] are shown in two views [the $2Ca^{2+}$-CaM−skMLCK complex (cyan) and the $4Ca^{2+}$-CaM−skMLCK complex (red)]. The two complexes are overlaid such that the cPKA structures are coincident. The resulting view shows how far the CaM translocates away from the catalytic cleft of skMLCK when all four Ca^{2+}-binding sites are occupied.

CaM-binding domain of the kinase. The previous dogma for CaM binding to its targets inferred sequential steps of Ca^{2+} binding which resulted in an "opening" of the interacting lobe via concerted motions of pairs of opposing helices and then ligand binding. Our recent scattering studies of the N-terminal lobe of CaM (nCaM) in its Ca^{2+}-loaded and Ca^{2+}-free states, combined with molecular dynamics studies on the N-terminal lobe and intact CaM, have shown that the picture is more complex than a simple conformational change (*50*). We demonstrated that apo-nCaM undergoes spontaneous opening and closing motions that would be expected to facilitate binding to targets in the absence of Ca^{2+}. It appears from the $2Ca^{2+}$-CaM results here that once CaM associates with skMLCK, most likely through an interaction between the C-terminal lobe of CaM and the Trp implicated as the initial recognition site for the related CaM-dependent protein kinase I (*51*), the N-terminal lobe interactions occur independent of the number of Ca^{2+} atoms bound. Our data suggest, however, that activation cannot occur until CaM is in the fully translocated position away from the catalytic cleft, which is presumably linked to full release of the pseudosubstrate/inhibitory sequence and that this step is completed only once the CaM is fully loaded. Interestingly, binding of CaM to residues within the autoinhibitory domain, those N-terminal to the traditional CaM-binding sequence, was observed in a recent crystal structure of the complex of CaM with a target sequence of CaM-dependent protein kinase I (*52*). It could be that the complete loading of Ca^{2+} onto CaM facilitates secondary interactions required for kinase activation.

SUPPORTING INFORMATION AVAILABLE

Families of structure used to produce the consensus envelop for the three CaM−skMLCK complexes presented here. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Krueger, J. K., Olah, G. A., Rokop, S. E., Zhi, G., Stull, J. T., and Trewhella, J. (1997) *Biochemistry 36*, 6017−6023.
2. Krueger, J. K., Zhi, G., Stull, J. T., and Trewhella, J. (1998) *Biochemistry 37*, 13997−14004.
3. Krueger, J. K., Bishop, N. A., Blumenthal, D. K., Zhi, G., Beckingham, K., Stull, J. T., and Trewhella, J. (1998) *Biochemistry 37*, 17810−17817.
4. Krueger, J. K., Gallagher, S. C., Zhi, G., Geguchadze, R., Persechini, A., Stull, J. T., and Trewhella, J. (2001) *J. Biol. Chem. 276*, 4535−4538.
5. Heidorn, D. B., and Trewhella, J. (1988) *Biochemistry 27*, 909−915.
6. Babu, Y. S., Bugg, C. E., and Cook, W. J. (1988) *J. Mol. Biol. 204*, 191−204.
7. Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., and Bax, A. (1992) *Biochemistry 31*, 5269−5278.
8. Heidorn, D. B., Seeger, P. A., Rokop, S. E., Blumenthal, D. K., Means, A. R., Crespi, H., and Trewhella, J. (1989) *Biochemistry 28*, 6757−6764.
9. Ikura, M., Clore, G. M., Gronenborn, A. M., Zhu, G., Klee, C. B., and Bax, A. (1992) *Science 256*, 632−638.
10. Meador, W. E., Means, A. R., and Quiocho, F. A. (1992) *Science 257*, 1251−1255.
11. Kemp, B. E., and Pearson, R. B. (1991) *Methods Enzymol. 200*, 121−134.
12. Gallagher, P. J., Herring, B. P., Trafny, A., Sowadski, J., and Stull, J. T. (1993) *J. Biol. Chem. 268*, 26578−26582.
13. Krueger, J. K., Padre, R. C., and Stull, J. T. (1995) *J. Biol. Chem. 270*, 16848−16853.
14. Olah, G. A., Mitchell, R. D., Sosnick, T. R., Walsh, D. A., and Trewhella, J. (1993) *Biochemistry 32*, 3649−3657.
15. Bayley, P. M., Findlay, W. A., and Martin, S. R. (1996) *Protein Sci. 5*, 1215−1228.
16. Peersen, O. B., Madsen, T. S., and Falke, J. J. (1997) *Protein Sci. 6*, 794−807.
17. Persechini, A., Yano, K., and Stemmer, P. M. (2000) *J. Biol. Chem. 275*, 4199−4204.
18. Stuhrmann, H. B. (1970) *Acta Crystallogr. A26*, 297−306.
19. Svergun, D. I., and Stuhrmann, H. B. (1991) *Acta Crystallogr. A47*, 736−744.
20. Grossmann, J. G., Abraham, Z. H. L., Adman, E. T., Neu, M., Eady, R. R., Smith, B. E., and Hasnain, S. S. (1993) *Biochemistry 32*, 7360−7366.
21. Svergun, D. I., Koch, M. H. J., and Serdyuk, I. N. (1994) *J. Mol. Biol. 240*, 66−77.
22. Svergun, D. I., Pedersen, J. S., Serdyuk, I. N., and Koch, M. H. J. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 11826−11830.
23. Svergun, D. I., Volkov, V. V., Kozin, M. B., and Stuhrmann, H. B. (1996) *Acta Crystallogr. A52*, 419−426.

24. Svergun, D. I., Volkov, V. V., Kozin, M. B., Stuhrmann, H. B., Barberato, C., and Koch, M. H. J. (1997) *J. Appl. Crystallogr. 30*, 798−802.
25. Spinozzi, F., Carsughi, F., and Mariani, P. (1998) *J. Chem. Phys. 109*, 10148−10158.
26. Chacón, P., Morán, F., Díaz, J. F., Pantos, E., and Andreu, J. M. (1998) *Biophys. J. 74*, 2760−2775.
27. Chacón, P., Morán, F., Díaz, J. F., and Andreu, J. M. (2001) *J. Mol. Biol. 299*, 1289−1302.
28. Svergun, D. I. (1999) *Biophys. J. 76*, 2879−2886.
29. Walther, D., Cohen, F. E., and Doniach, S. (2000) *J. Appl. Crystallogr. 33*, 350−363.
30. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. J. (2001) *Biophys. J. 80*, 2946−2953.
31. Gallagher, S. C., Callaghan, A. J., Zhao, J., Dalton, H., and Trewhella, J. (1999) *Biochemistry 38*, 6752−6760.
32. Debye, P. (1915) *Ann. Phys. 46*, 809−823.
33. Winter, G., Ed. (1995) *Genetic Algorithms in Engineering and Computer Science*, Wiley, Chichester, U.K.
34. Harpaz, Y., Gerstein, M., and Chothia, C. (1994) *Structure 2*, 641−649.
35. Bada, M., Walter, D., Arcangioli, B., Doniach, S., and Delarue, M. (2000) *J. Mol. Biol. 300*, 563−574.
36. Funari, S. S., Rapp, G., Perbandt, M., Dierks, K., Vallazza, M., Betzel, C., Erdmann, V. A., and Svergun, D. I. (2000) *J. Biol. Chem. 275*, 31283−31288.
37. Kozin, M. B., and Svergun, D. I. (2001) *J. Appl. Crystallogr. 34*, 33−41.
38. Takahashi, Y., Nishikawa, Y., and Fujisawa, T. (2003) *J. Appl. Crystallogr. 36*, 549−552.
39. Volkov, V. V., and Svergun, D. I. (2003) *J. Appl. Crystallogr. 36*, 860−864.
40. Diamond, R., Phillips, D. C., Blake, C. C. F., and North, A. C. T. (1974) *J. Mol. Biol. 82*, 371−391.
41. Rosenzweig, A. C., Frederick, C. A., Lippard, S. J., and Nordlund, P. (1993) *Nature 366*, 537−543.
42. Kobe, B., and Deisenhofer, J. (1996) *J. Mol. Biol. 264*, 1028−1043.
43. Ikura, M., Barbato, G., Klee, C. B., and Bax, A. (1992) *Cell Calcium 13*, 391−400.
44. Knighton, D. R., Zheng, J., Ten Eyck, L. F., Xuong, N.-H., Taylor, S. S., and Sowadski, J. M. (1991) *Science 253*, 407−414.
45. Cox, J. A., Malone, A., and Stein, E. A. (1981) *J. Biol. Chem. 256*, 3218−3222.
46. Cox, J. A., Conte, M., and Stein, E. A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4265−4269.
47. Burger, D., Stein, E. A., and Cox, J. A. (1983) *J. Biol. Chem. 258*, 4733−4739.
48. Blumenthal, D. K., and Stull, J. T. (1982) *Biochemistry 21*, 2386−2391.
49. Linse, S., Hemersson, A., and Forsen, S. (1991) *J. Biol. Chem. 266*, 8050−8054.
50. Vigil, D., Gallagher, S. C., Trewhella, J., and Garcia, A. E. (2001) *Biophys. J. 80*, 2082−2092.
51. Goldberg, J., Nairn, A. C., and Kuriyan, J. (1996) *Cell 84*, 875−887.
52. Clapperton, J. A., Martin, S. R., Smerdon, S. J., Gamblin, S. J., and Bayley, P. M. (2002) *Biochemistry 41*, 14669−14679.

BI0348664