Appendix A

Toward computationally designed calmodulin variants with enhanced peptide binding specificity

Abstract

Calmodulin (CaM) is a second messenger protein that binds a wide variety of natural protein and peptide substrates. Although these interactions are of high affinity, CaM shows little specificity among its native substrates. We used computational design including explicit negative design to engineer CaM variants that preferentially bind one natural peptide substrate over another. By specifically modeling the structures in complex with both the desired and undesired substrates, negative design methods allow us to predict variants that will bind the desired substrate (smMLCK peptide) with high affinity and the undesired substrate (CaMKI peptide) with lower affinity. We found that one of our variants (M51Y) showed a 10-fold preference for $smMLCK_p$ over $CaMKI_p$ in a fluorescence assay, as predicted from the negative design calculation. Additionally, we show that a surface plasmon resonance (SPR)-based method is more effective at determining the dissociation constants for these high affinity interactions than the commonly used fluorescence-based assay. These initial studies provide a strong foundation for future work on the CaM-peptide system.

Introduction

Protein-protein interactions. Protein-protein interactions are necessary for every cellular process: from signal transduction to regulation of gene expression to apoptosis to intracellular transport. The protein-protein interfaces involved in these interactions are, by necessity, specific to the binding partners and exhibit vast diversity in their size, types of interactions, and shape of the interaction surface. To better understand cellular processes and the proteome, several groups have focused on modeling and predicting protein-protein interactions.[1-3] A complementary approach is to design novel or improved protein-protein interactions to test our knowledge and understanding of the forces that are important for both affinity and specificity.

The ability to design protein-protein interfaces with high affinity and specificity could be valuable in developing new or improved protein therapeutics as well as in advancing basic scientific research. Novel interfaces have already been generated to create restriction enzymes with new DNA recognition sequences,[4,5] new Ca^{2+} sensors that can be used *in vivo* to study variations in calcium concentration,[6] and new orthogonal protein-peptide pairs that can provide an alternative for current affinity purification and pull-down assays.[7]

Design of protein-protein interactions. The importance of protein-protein interactions in understanding biological systems has stimulated a vast amount of research aimed at designing and modulating protein-protein interfaces. Various techniques have been employed to optimize or alter protein-protein contacts, including selection schemes to assay libraries of variants,[8-10] structure-based rational design,[11] and computational

methods. Computational design methods have become increasingly valuable for modulating protein-protein interfaces because they allow researchers to screen many more sequences than would be realistic using experimental methods. This technique has led to the success of several designed protein-protein interfaces, including engineered obligate heterodimeric endonucleases that recognize novel DNA targets,[4,5] new orthogonal interacting pairs with calmodulin[6] and PDZ domains,[7] protein-inhibitor interfaces with redesigned specificities,[12-14] and novel protein-protein interfaces from monomeric proteins.[15]

Calmodulin-peptide complexes have proven to be useful systems for studying specificity[16,17] as well as for designing orthogonal interfaces.[6,18,19] Previous work from our group showed that modest specificity enhancements could be obtained by stabilizing one CaM-peptide structure without taking any negative design into account.[16] Other researchers showed that by designing both the peptide and the calmodulin surface, highly specific interactions could be formed.[6] However, not all studies have found calmodulin specificity to be easily modulated. Green *et al.* found that although they could design new calmodulin-peptide pairs that display high affinity for each other, the designed molecules did not show enhanced specificity.[19]

Various strategies for optimizing or modulating interfaces have been employed in computational protein interface design. A “knob-and-hole” strategy in which unfavorable mutations are introduced into one binding partner and the other binding partner is designed with compensating mutations is very common.[13] This is an efficient method for developing novel interfaces, but is less useful for altering specificity among native substrates. Additionally, burial of more hydrophobic surface area[20] or increasing the

amount of surface area in the interface[21] have been valuable approaches for enhancing the binding affinities of protein-protein or protein-peptide interactions; however, they do not necessarily provide specificity of interaction.

Although much work has already been done, we still face a number of challenges in the design of protein-protein interactions. While standard optimization methods can lead to stabilized complexes and potentially slight increases in specificity, designs are often more successful when negative design is included.[22,23] Bolon *et al.* showed that although heterodimers could be designed using only positive design methods, the resulting molecules were more specific if unfavorable structures (homodimers) were also modeled and specifically selected against during the optimization.[23]

In this study, we investigated the specificity of the native calmodulin system. In a continuation of previous results from our group,[16,17] we hoped to modulate the specificity of calmodulin between two native peptide substrates by explicitly modeling both bound states in our calculations. This work differs from previous studies in that our intention was not to generate a new interface, but to impart specificity between two substrates that already bind to the target protein with high affinity.

Computational protein design using ORBIT. ORBIT (Optimization of Rotamers By Iterative Techniques) is a fully automated computational protein design software suite developed in the Mayo laboratory. This method has been used with great success to address a number of protein design problems. ORBIT was effectively used to design a zinc finger motif without zinc, and the predicted sequence was subsequently shown to adopt the expected fold.[24] Many proteins were designed to be more thermodynamically

stable than their wild-type counterparts.[25,26] More recently, ORBIT has been used to modulate the function of proteins. In 2002, Datta *et al*. used ORBIT to design a tRNA-synthetase that allows an unnatural amino acid to be incorporated into proteins *in vivo*.[27] Modest success has been made in the area of enzyme design. ORBIT was used to design an active site onto an inert protein scaffold and generate an "enzyme-like protein" that exhibits catalytic activity,[28] and a more efficient variant was created by optimizing the residues surrounding the active site of a native enzyme.[29]

To reduce the combinatorial complexity inherent in computational design, ORBIT optimizations begin with a fixed backbone taken from a high-resolution crystal structure. In addition, discrete rotamers or conformers are used to model the flexibility of the side chains within the protein. Rotamers are idealized versions of the average conformation of rotamers found in protein structures in the Protein Data Bank (PDB,[30] whereas conformers are side-chain conformations that are the closest to the average conformation within a defined area of phi/psi space.[31] These conformers are actual 3D coordinates from structures and not idealized averages, and therefore may represent more realistic conformations for side chains. A backbone-dependent or backbone-independent rotamer library can be used to limit the computation time and select for more physically relevant side-chain conformations.

The ORBIT scoring function is based on the DREIDING force field[32]; it calculates the energy of a given conformation of a protein sequence, and incorporates four physical forces that are believed to be important in protein design and folding. The first component is a scaled van der Waals interaction energy that includes both a short-range repulsive component as well as a long-range attractive component.[33] Hydrogen

bonding energies are calculated using a distance-, angle-, and hybridization-dependent hydrogen bond scoring function. The third component is a distance-dependent electrostatic potential.[33] Finally, a surface area-based solvation potential is included that applies a benefit for burial of non-polar surface area, a penalty for burial of polar surface area, and a penalty for exposure of non-polar surface area.[34] After the energies are calculated using the scoring function, various deterministic[35-37] and stochastic[38-41] search algorithms are used to find either the best sequence (the global minimum energy conformation or GMEC) or other low energy sequences that are consistent with the design and backbone.

Previous studies using ORBIT, including those on calmodulin,[16,17] were performed using only positive computational design. That is, a sequence is determined using the scoring function and the desired structure in order to stabilize the fold specified. In the case of calmodulin designs, this was a successful strategy for creating a variant that showed specificity between some peptides, but not all of those tested. In this study, we added a negative design component to the computational design in which we explicitly model the undesired state (the protein bound to a different peptide) and select for sequences that favor the desired structure while disfavoring the undesired one. By incorporating this additional information, we hoped to validate our negative design methods as well as develop methods for incorporating specificity into a protein system.

Calmodulin-peptide system. Calmodulin (CaM) is a highly conserved, ubiquitous protein found in all eukaryotes. It is a calcium-dependent regulatory protein that binds to and modulates the function of many cellular proteins including those involved in

phosphorylation, muscle contraction, cell proliferation, and metabolism.[42,43] CaM regulates proteins in multiple ways, the most common of which is by relieving autoinihibition of the target by binding to the autoinhibitory domain (AID).[44] Although CaM binds with broad specificity to a vast array of different proteins, many of these targets contain a basic amphiphilic α-helical peptide to which CaM can bind with high affinity.[45-47] This natural high affinity and broad specificity makes CaM an ideal candidate for studying protein-protein interactions, protein specificity, and negative design.

Currently, over 15 high-resolution structures are available of CaM bound to peptides in the most common conformation in which the two domains of CaM wrap around the helical peptide (Figure A-1). When bound to a target peptide in this manner, CaM forms a compact α-helical structure bound to four Ca^{2+} ions which envelopes the amphiphilic peptide. Within the protein-peptide interface, eight methionine residues and a flexible central helix are thought to provide a great deal of conformational elasticity that may allow the CaM to adjust to many different peptide substrates.[46,48]

For this study, we chose to investigate protein-peptide specificity using two CaM-peptide structures: smMLCK$_p$-CaM[49] and CaMKI$_p$-CaM.[50] These high-resolution crystal structures have very similar backbone orientations, with an RMSD of 1.4 Å. However, the peptides show little sequence homology with the exception of the tryptophan, which anchors the peptide and is common in CaM-binding peptides (see Table A-1). Additionally, it has been shown that CaM can distinguish between these and other peptides by binding them with slight backbone differences to accommodate the peptide side chains.[48,51] These minor differences in structure lead us to believe that we can

computationally design variants to show specificity between the two high affinity substrates $smMLCK_p$ and $CaMKI_p$ by using negative design techniques that explicitly model both complexed structures.

Methods

ORBIT calculations. ORBIT optimization calculations were run on the protein-peptide interface in both the $smMLCK_p$-CaM cocrystal structure (1CDL)[49] and the $CaMKI_p$-CaM cocrystal structure (1MXE).[50] To prepare the structures for calculations, water molecules and calcium ions were removed from the pdb files for each structure and Reduce, a program in the MolProbity suite, was used to add hydrogen atoms.[52] A side-chain placement calculation was run to add side chains for residues 77 and 115, which were missing from the 1CDL structure. In addition, 99 Phe and 143 Thr were placed in the 1CDL structure to mutate the sequence from the human CaM to Drosophila CaM and therefore be consistent with the 1MXE sequence. All four of these residues are distal from the protein-peptide interface and were therefore not expected to have an impact in the ultimate design. 1CDL and 1MXE were minimized for 50 steps using the DREIDING force field.[32] Each residue was characterized as surface, boundary, or core based on the distance of its Cα and Cβ to the solvent-accessible surface.[24]

Positive designs were first conducted on both structures. Residues that were within 4 Å of the peptide in both structures, were defined as core, and were either hydrophobic or glutamate (Glu) in the wild-type protein were allowed to vary in the calculations. Residues 11, 12, 15, 18, 19, 32, 39, 55, 68, 84, 88, 91, 92, 105, 108, 112, and 128 were allowed to sample Ala, Val, Leu, Ile, Phe, Tyr, and Glu amino acids.

Residues 36, 51, 71, 72, 109, 124, 144, and 145 are methionine in the wild-type protein and therefore were allowed to sample conformations of Met in addition to the amino acids listed above. Wild-type CaM contains no tryptophans and Trp was not included in the designs to simplify the binding analysis by Trp fluorescence. The side chains in the peptides were allowed to change conformation but not identity.

Two positive designs were carried out on each structure. In the first set of designs, a backbone-dependent rotamer library in which each rotamer was expanded by 1 standard deviation around $\chi 1$ and $\chi 2$ was used (e2).[30] These results are designated "peptide-e2." In the second set of designs, a backbone-independent conformer library was used.[31] Results from these calculations are designated "peptide-conf."

Standard ORBIT parameters were used for these calculations[33,34] with the exception of a decreased polar burial penalty (0.03 kcal/mol/Å^2) to not overly penalize Glu rotamers, and a 100 kcal/mol cutoff energy. The increased cutoff energy allowed us to keep rotamers that might be necessary in the negative design calculation. To place more emphasis on the interactions between CaM and the peptides, we introduced a scale factor to bias the intermolecular interaction energies.[17] This scale factor resulted in recovery of Glu84, which forms favorable interactions with the peptides but has unfavorable electrostatics with the rest of CaM. The scale factors used were 1.3 (conf designs) and 1.1 (e2 designs). Optimization of the positive designs was accomplished using a modification of the FASTER algorithm, giving a single low-energy sequence, the FMEC.[38,39]

Negative designs were scored using a simple scoring function (Equation A-1), where ΔE_{pos} and ΔE_{neg} are the energies of a given sequence in the context of the desired

complex and the undesired complex, respectively. W is an arbitrary weighting factor used to balance the desire for a good energy for the positive structure and a bad energy for the negative structure. The weighting factor for these calculations was 0.2.

$$Score = \Delta E_{pos} - w * \Delta E_{neg} \quad \text{(A-1)}$$

The negative design calculations were performed using the 1CDL structure (CaM-smMLCK$_p$) as the desired complex and the 1MXE structure (CaM-CaMKI$_p$) as the undesired complex. The energy calculations from the positive designs (described above) were used as input and the score was minimized according to Equation A-1 using FASTER[39] with 200 trajectories. The single sequence with the lowest score was investigated further. This protocol was completed for both the conformer library designs and the standard rotamer library designs.

Construct generation. Wild-type Drosophila CaM and CaM8 (an eight-fold mutant previously shown to have enhanced specificity) constructs in pET15b were obtained from Julia M. Shifman.[16,17] Variant constructs containing a single mutation were created using inverse PCR mutagenesis,[53] and constructs with two or more mutations were generated using the Quikchange multi-site mutagenesis kit (Stratagene). All constructs were verified through DNA sequencing.

Protein expression and purification. Wild-type CaM and all variants were expressed in *E. coli* BL21(DE3). All growth was conducted in LB supplemented with 100 μg/mL ampicillin. Expression was induced at OD_{600} = 0.6 with 1 mM IPTG and cells

were grown for 2-4 hours at 37 ˚C. Cells were harvested at 5,000 × g for 10 minutes in a centrifuge and the cell pellet resuspended in 25 mL of Lysis Buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 5 mM $CaCl_2$). Resuspended cells were frozen at -20 ˚C overnight. After thawing, 1 mM PMSF, DNase, and RNase were added and the cells were lysed using an Emulsiflex (Avestin) for 3-5 minutes. The lysate was clarified by centrifuging for 30 minutes at 21,000 × g and filtered through a 0.22 μm filter before chromatography was conducted. The clarified lysate containing the CaM was first purified on a Phenyl Sepharose HP HiLoad 16/10 column (General Electric) preequilibrated with Buffer A (50 mM Tris-HCl pH 7.5, 1 mM $CaCl_2$), essentially as described.[54] Briefly, the clarified lysate was loaded on the column, washed with 4 CV of Buffer A, 2 CV of Buffer A containing 500 mM NaCl, then an additional 2 CV of Buffer A. CaM was eluted with Buffer C (50 mM Tris-HCl pH 7.5, 1 mM EDTA). The phenyl sepharose elution was concentrated to approximately 0.5 mL using 5,000 MWCO Amicon Ultra concentrators (Millipore), then loaded onto a SDX-75 column (General Electric) preequilibrated with 50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM $CaCl_2$. Pure CaM eluted at approximately 0.53 CV and was collected and stored at 4 ˚C. Wild type (WT) CaM and the variants were purified to > 95% purity as determined by SDS-PAGE. In addition, the molecular weight for all proteins was verified through mass spectrometry.

CaM concentrations were calculated one of two ways. The first method measures the absorbance of tyrosine in denatured CaM. A 1/10 dilution of protein into 8 M GnHCl was incubated for 10 minutes before the absorbance at 278 nm was taken. The extinction coefficient at 278 nm for WT and variants containing one tyrosine was 1400 $M^{-1}cm^{-1}$ and 2800 $M^{-1}cm^{-1}$ for variants containing two tyrosines.[55] The second method was used to

calculate the concentration of protein under native conditions. Typically 1/3 dilutions of CaM were made into 50 mM Tris-HCl pH 7.5, 1 mM $CaCl_2$ and the A_{276} was measured. An extinction coefficient of 1576 $M^{-1}cm^{-1}$ was used for proteins, including WT, that contain only one tyrosine.[56] The concentrations from both methods are typically within 5% of each other.

CaM-binding peptides. Unmodified peptides corresponding to the peptides bound to CaM in the crystal structures 1CDL ($smMLCK_p$) and 1MXE ($CaMKI_p$) were purchased from PeptidoGenic Research and Co., Inc. or from Sigma Genosys at greater than 95% purity. Biotinylated smMLCK and CaMKI peptides were purchased from AnaSpec, Inc. at greater than 95% purity. The N-terminal biotin moiety was separated from the peptide sequence by two 6-aminohexanoic acid linkers (LC) and a Gly-Gly-Ser-Gly-Gly peptide linker. The sequences for all peptides used in this study can be found in Table 5-1.

The molecular weight of each peptide was verified through mass spectrometry, and amino acid analysis was performed on unmodified smMLCK and CaMKI peptides (Jinny Johnson, Texas A&M University) to confirm the composition. The extinction coefficient A_{280} = 5690 $M^{-1}cm^{-1}$ was used to determine the concentration of all peptides under denaturing conditions.[55]

Fluorescence binding assays. Fluorescence assays were conducted at 25 ˚C on a QuantaMaster UV-Vis fluorimeter (Photon Technology International) essentially as described.[16] Samples (1 mL) were prepared with varying amounts of CaM (0 to 2.0 μM)

and 1 μM peptide (unless otherwise noted) in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM $CaCl_2$. Samples were allowed to equilibrate a minimum of 1 hour before measurements were taken. Each sample was excited at 295 nm and the emission spectrum was recorded between 310 nm and 460 nm. A 1 nm step size and 1 second averaging time were used in all assays. The slit width on the excitation source and emission detectors were adjusted to give a maximum signal near the limit of the detector, typically between 0.9 and 1.1 nm.

The fluorescence at 318 nm (unless otherwise noted) as a function of the CaM concentration was plotted in Kaleidograph (Synergy Software). Data were fit to a 1:1 binding model to obtain the dissociation binding constant (K_D) (Equation A-2) where ε_{free} (the extinction coefficient for free peptide), ε_{bound} (the extinction coefficient for bound peptide), and K_D are parameters that are fit during optimization, and [P] and [CaM] are the concentration of peptide and CaM, respectively. In addition, a CaM concentration factor was fit during the curve fitting procedure to correct for inaccuracies in the determination of CaM concentrations. This fit value typically ranged between 0.9 and 1.1, indicating that our CaM concentrations were off by 10% or less.

$$Fluorescence = \varepsilon_{free} * [P] + (\varepsilon_{bound} - \varepsilon_{free}) * 0.5 * (K_D + [P] + [CaM] \pm \sqrt{(K_D + [P] + [CaM])^2 - 4 * [CaM] * [P]}) \quad \text{(A-2)}$$

Surface Plasmon Resonance (SPR) assays. SPR (Biacore) experiments were performed on a T100 instrument (Biacore). Approximately 3000-5000 response units of streptavidin were immobilized to all four flow cells of a CM5 chip through standard amine coupling. Biotinylated peptide was captured by the streptavidin surface on flow cells 2-4. Flow cell 1 was reserved as a control. All assays were conducted in HBS-P

buffer (Biacore) with 1 mM $CaCl_2$ (10 mM HEPES pH 7.4, 150 mM NaCl, 0.0005% v/v Surfactant P20, 1 mM $CaCl_2$). The chip was regenerated with two 30-second injections of 10 mM EGTA unless otherwise noted.

Kinetics experiments for wild-type CaM were conducted on a chip with between 5 and 15 RUs of immobilized peptide. Samples containing serial dilutions between 2 μM and 6.3 pM WT CaM were injected over the surface of the chip for 180 seconds at 100 μL/min then allowed to dissociate for 300 seconds before the surface was regenerated. One sample in the series was repeated to confirm that the immobilized surface was unchanged during the experiment. Sensorgrams were analyzed using the BiaEvaluation software package (Biacore). Various fitting models were used to try to fit the data accurately, including a 1:1 model, a multivalent model, and a mass-transport limited model.

Equilibrium experiments were performed with approximately 20 RUs of peptide immobilized on the streptavidin surface. Samples used to create binding curves were either from a 2X dilution series containing 20 nM to 9.8 pM CaM in running buffer or a 24-point semi-log dilution series with the highest concentration of 50 μM. The flow rate and contact time for each injection was varied in order to reach equilibrium binding. The flow rates varied from 20 to 8 μL/min and the contact times ranged between 420 and 2500 sec. All concentration series were run from low to high concentration without regeneration in order to expedite the equilibrium. After a complete concentration series, the chip was regenerated with 5 pulses of 10 mM EGTA. In addition, a 2.5 nM WT CaM reference sample was run after every two proteins in order to verify the stability of the surface. The data were analyzed by plotting the equilibrium RU value as a function of the

concentration of CaM injected. Data are shown as unfit curves and can only be ranked to determine the relative affinities of the variants due to poor fits to a 1:1 model.

Approximately 120 RUs of biotinylated smMLCK peptide was immobilized to a streptavidin CM5 chip for the competition assays. Samples containing 100 nM CaM (WT or variant) and either no peptide or between 250 nM and 3.3 nM non-biotinylated peptide were injected onto the chip at 50 μL/min with a 30 second association time and no dissociation time. Two samples were repeated during the run to ensure that the streptavidin/peptide surface was not being damaged or degraded. The surface was regenerated after each injection. Data was exported from BiaEvaluation software and the initial rate of signal increase for each sensorgram was determined by fitting the initial seconds of injection to a linear regression in Excel (Microsoft). The dissociation binding constant (K_D) was determined by plotting the initial slopes as a function of the log of the concentration of competitive (non-biotinylated) peptide. Kaleidograph was used to solve for K_D by fitting the data to Equation A-3 where $[A_0]$ is the CaM concentration, R_0 is the initial rate with no competing receptor present, X is the $\log[L_0]$, where $[L_0]$ is the concentration of competitive peptide, and R is the initial rate at $[L_0]$.[57]

$$R = \frac{R_0}{2[A_0]}\left([A_0] - 10^x - K_D\right) + \sqrt{(K_D{}^2 + 2(10^x)(K_D) + (10^x)^2 + 2[A_0]K_D - 2[A_0]10^x + [A_0]^2)} \quad \text{(A-3)}$$

Results

ORBIT calculations. We chose the structures for the smMLCK-CaM complex (pdb ID 1CDL)[49] and the CaMKI-CaM complex (pdb ID 1MXE)[50] to incorporate explicit negative design into our computational protocol. These complexes were chosen because both have high-resolution structures available, both have very high CaM-peptide binding

affinities, and smMLCK$_p$ and CaMKI$_p$ both contain a tryptophan residue. As CaM contains no Trps, the change in fluorescence of the peptide Trp can be monitored to determine binding.[16] In addition, these two structures are very similar: their RMSD is only 1.4 Å and the peptide is bound in a very similar orientation. Previous work in the Mayo lab indicated that CaM can be engineered to have altered specificity without including explicit negative design.[16] We hoped to expand on this work and create a CaM variant with increased specificity for the smMLCK peptide relative to the CaMKI peptide.

The ORBIT protein design software suite was used to both optimize the peptide-protein interface in the two CaM-peptide input structures as well as predict sequences that would stabilize the smMLCK$_p$-bound structure (1CDL) and destabilize the CaMKI$_p$ structure (1MXE). The resulting sequences are shown in Table A-2. The sequences are split into two categories depending on the rotamer library used, as described in the methods.

The positive designs for the smMLCK$_p$-CaM structure (smMLCK-conf and smMLCK-e2) contain four and five mutations, respectively, from wild-type CaM. V55I, V91I, and V108I are found in both designs. These are very conservative mutations with the extra methyl group predicted to fill small spaces in the protein-peptide interface. Positions 18 and 51 were also mutated in one or both of these sequences; these mutations will be discussed in the negative design results as they are also found in the negative design sequences.

The positive designs for the CaMKI$_p$-CaM structure (CaMKI-conf and CaMKI-e2) contain five and three mutations, respectively. Many of the mutations are valine to

phenylalanine or tyrosine. The $CaMKI_p$-CaM structure is not as packed at the protein-peptide interface as the $smMLCK_p$-CaM structure, so there is additional room for larger side chains to pack into the interface at positions 55, 91, and 108. Position 84 is also mutated from a Glu to an Ile or Leu in these designs. In the $smMLCK_p$-CaM structure, Glu 84 makes electrostatic interactions with two arginine residues in the peptide. However, in the CaMKI peptide these arginines have been replaced by an alanine and a histidine that is not within hydrogen bonding distance. Therefore, in the $CaMKI_p$-CaM optimization, the electrostatics interactions with E84 are less stabilizing than the van der Waals interactions between the hydrophobic Ile or Leu and the peptide.

Because the goal of the negative designs was to create a CaM variant that binds $smMLCK_p$ with high affinity and $CaMKI_p$ with reduced affinity, it was no surprise that the sequences resulting from the negative design protocol are very similar to the sequences for the $smMLCK_p$-CaM positive designs. There are, however, two notable exceptions that may convey specificity to the negative design variants. Those mutations are L18I, found in both negative design solutions as well as in smMLCK-e2, and M51Y, which was selected in the negative design calculation from the e2 rotamer library and the positive design sequence smMLCK-conf. Both L18I and M51Y are predicted to make van der Waals clashes with the CaMKI peptide, but can be accommodated in the $smMLCK_p$ structure (Figure A-2). In the case of L18I, the clash between residue 18 and Gln 305 of the peptide cannot be resolved through alternate conformations due to the constraints on the peptide side chain. In the $smMLCK_p$-CaM structure, however, the lysine residue at the equivalent position in the peptide is predicted to sample a different conformation that allows space for the isoleucine. In addition, the isoleucine rotamer

chosen in the smMLCK structure is similar to the crystallographic orientation of the wild-type leucine. M51Y is located in the protein-peptide interface near the C-terminus of the peptide (Figures A-2C and A-2D). In the $smMLCK_p$ design, the tyrosine side chain extends toward the solvent and is not sterically occluded by the serine residue at the end of the peptide. The tyrosine in the $CaMKI_p$ structure is predicted to be blocked from extending toward the solvent by the long arginine side chain near the end of the peptide and therefore must take an alternate conformation in which it sterically clashes with Met 316 of the CaMKI peptide. These computational models suggest that both L18I and M51Y might be good candidates for incorporating specificity into CaM-peptide binding. We therefore decided to construct variants containing either single or double mutations to investigate the effects of these positions: L18I, L18I/M51F, L18I/M51Y, M51F, and M51Y.

Native purification. Native purification of CaM has been reported using hydrophobic interaction chromatography.[54,58-60] In the presence of Ca^{2+}, a large hydrophobic surface on CaM is exposed and interacts with the column. After non-hydrophobic proteins are washed off the column, CaM is eluted with a buffer containing EDTA. The EDTA chelates the Ca^{2+}, causing the CaM to undergo a conformational change that buries the hydrophobic patch. This method has numerous advantages over the non-native purification method previously used in the lab.[16] First, the protein does not need to undergo refolding and is exposed only to gentle, native conditions, resulting in protein preparations that are less likely to be damaged. Second, only the CaM that undergoes the necessary Ca^{2+}-dependent conformation change is purified; damaged

proteins will either be lost in the washes or remain bound to the column after elution. Finally, only proteins with a Ca^{2+} dependency will elute with the buffer containing EDTA, so the protein is greater than 95% pure after a single purification step. We performed an additional gel filtration step to exchange the buffer and ensure that the protein was monodispersive and monomeric.

SDS-PAGE on the purified protein showed that it was more than 98% pure. A double band was often seen when excess EDTA or Ca^{2+} was not added before denaturing the samples. The double band was shown to correspond to Ca^{2+}-bound and Ca^{2+}-free versions of the protein,[61] which has been verified in our laboratory (data not shown).

Tryptophan fluorescence assays. Tryptophan (Trp) fluorescence is a convenient, commonly used technique to determine the affinity of CaM-peptide interactions.[16,17,62-64] The analysis is simplified because CaM contains no Trps. Fluorescence from the single Trp in the peptide can be monitored to indicate movement from solvent-exposure to a hydrophobic environment, giving a direct measurement of peptide binding to CaM.[65] After excitation at 295 nm, the emission spectrum was monitored for samples containing a fixed amount of Trp-containing peptide and varying amounts of CaM (Figure A-3). Water emits a significant amount of fluorescence around 330 nm. When the background was subtracted, the emission spectra clearly showed the expected blue-shift and increase in intensity as CaM was titrated in, which is characteristic of the burial of the Trp (Figure A-3B). To eliminate error due to sampling around the water peak, we determined that monitoring emission at 318 nm gave results similar to those at 326 nm, but with smaller error associated with the dissociation constant (data not shown). We therefore plotted

fluorescence at 318 nm as a function of the CaM/peptide ratio in order to calculate K_D (Figure A-3C).

Although the fluorescence-based assay is simple and straightforward, the low intensity of Trp fluorescence emission makes it impossible to use low concentrations of peptide. This is a problem because the K_Ds we are analyzing are around 1 nM and our peptide concentration is 1 μM. This three orders-of-magnitude difference between the expected K_D and the peptide concentration creates a great deal of error in the K_D extrapolated from the data. We experimented with decreasing the concentration of peptide and found that 0.3 μM peptide gave reasonable signal to noise, but that 0.1 μM peptide gave poor data (data not shown). However, 0.3 μM peptide did not get us significantly closer to the "10-fold above the K_D to 10-fold below the K_D" general rule for concentrations in binding assays, so we continued the experiments with 1 μM peptide to maximize the signal to noise.

CaM variants-peptide binding: Fluorescence assays. Eight computationally designed variants and wild-type CaM were assayed for binding to both $smMLCK_p$ and $CaMKI_p$ as described in the methods. The results are shown in Table A-3. We successfully reproduced the results from Shifman *et al*. for wild-type binding to both $smMLCK_p$ and $CaMKI_p$.[16] The K_D resulting from this experiment was two-fold weaker than that reported by Shifman *et al*., but well within the limitations of the assay. Unfortunately, most of the variants did not affect the specificity of binding as had been anticipated from the designs. Four out of the five variants representing the negative design calculations (L18I, M51F, L18I/M51F, L18I/M51Y) did not show any specificity

for $smMLCK_p$ over $CaMKI_p$. In addition, the positive designs on the $smMLCK_p$-CaM structure did not alter the specificity significantly. These assays did result in some interesting variants, however. First, the M51Y single mutant from the negative design calculations showed a 10-fold preference for the smMLCK peptide over the CaMKI peptide. This preference seems to result from an increase in affinity to $smMLCK_p$ as opposed to a decrease in affinity to $CaMKI_p$, contrary to what was expected from the negative design calculations. The variant corresponding to the positive design calculation performed on the $CaMKI_p$-CaM structure using the e2 library (CaMKI-e2) also showed an approximately 10-fold preference for binding $smMLCK_p$. This result was unexpected, as the calculation should have predicted mutations that stabilize the $CaMKI_p$-CaM interface. Finally, CaMKI-conf appeared to show a moderate specificity switch from a two-fold preference for $smMLCK_p$ for wild-type to a two-fold preference for $CaMKI_p$. Although these results were far from definitive, we decided to investigate the results from the two $CaMKI_p$-CaM structure positive designs further.

Five single mutation variants were created to investigate their effects on peptide binding. We created a variant with each mutation found in either CaMKI-conf or CaMKI-e2 that could have an affect on activity. We chose not to make L39I because it was also seen in the negative design result; similarly, V108I was not made because it was seen in both $smMLCK_p$-CaM designs and both negative design results and had no effect in those proteins. We therefore made V55F, E84I, V91F, V91Y, and V108F constructs and tested them in the Trp fluorescence assay. V55F was chosen during the optimization for both $CaMKI_p$-CaM positive designs. E84I, V91Y, and V108F were found only in the CaMKI-conf positive design, which has higher affinity for $CaMKI_p$ than $smMLCK_p$. The final

mutation, V91F, is derived from the CaMKI-e2 positive design sequence that has a 10-fold preference for $smMLCK_p$. The results from the fluorescence assay show that V55F, E84I, and V108F mutations may create a slight preference for $smMLCK_p$ (Table A-4). These results, however, are inconsistent with their presence in the CaMKI-conf design, which shows a switch toward better binding of $CaMKI_p$. The V55F mutation may provide some of the increased specificity in CaMKI-e2 binding to $smMLCK_p$. Mutations at position 91 had no effect on the binding to either $smMLCK_p$ or $CaMKI_p$, indicating that this mutation is not the cause of the differences seen in the positive designs.

Because these results seemed counter in many ways to our computational designs and had a great deal of error associated with them due to the large difference between the K_Ds of interest and the concentrations of peptide used, we decided to further study these variants using surface plasmon resonance (SPR, Biacore).

Biacore experiments to determine binding constants. Biacore has been used previously to provide both kinetics and steady-state dissociation constants for CaM binding to various peptides.[66-68] We chose to immobilize our two biotinylated peptides through streptavidin interaction. This allowed us to analyze multiple CaM variants at a time without making new chips and oriented the peptides in a consistent way to minimize the error associated with random orientation by amine coupling. Initial experiments indicated that the kinetics for the wild-type CaM interaction with $smMLCK_p$ and $CaMKI_p$ could not be fit to a 1:1 binding model (Figure A-4A). A better fit was obtained using a multivalent model (Figure A-4B). However, these complexes have been shown previously to be 1:1 interactions, and fitting to a multivalent model is not

appropriate.[16,49,50] We therefore decided to obtain dissociation constants from equilibrium experiments instead of analyzing the kinetics of the binding.

Equilibrium Biacore experiments measure the maximal response for a given analyte concentration. No kinetics information is obtained, but when the equilibrium binding response is plotted as a function of the concentration, a sigmoidal curve fit can give a K_D of binding (Figure A-5). When wild-type data was plotted, however, the shape of the curves was not as expected (Figure A-5B). At lower concentrations, the system acted as predicted, but at high concentrations, the response did not level out to reach a maximal value. Even at 50 μM CaM, maximal binding was not observed. This indicates that the system is not an ideal 1:1 binding system and instead shows some self-binding or aggregation at high concentrations.

Although the system did not act in an ideal fashion, we proceeded to evaluate the designed variants via equilibrium binding studies. None of the data fit well to the 1:1 model, and therefore K_Ds could not be determined. Instead, we ranked the variants relative to WT CaM (Figure A-6). The two smMLCK$_p$-CaM structure positive design variants (smMLCK-conf and smMLCK-e2) bound to both smMLCK$_p$ and CaMKI$_p$ with approximately the same affinity as WT. The mutations in these two variants did not affect the binding to either of these peptides. In addition, CaMKI-conf and CaM8, which had previously shown improved specificity to other peptides, bound significantly worse to both peptides as compared to WT. One or more of the mutations in these variants significantly affected either the stability of the protein or the binding in a negative way that was not peptide specific. One of the variants, CaMKI-e2, showed some specificity between the peptides. This variant binds significantly worse than WT to smMLCK$_p$ but

binds similar to WT to $CaMKI_p$. This result is consistent with the purpose of the design: to stabilize the CaMKI-CaM structure. Many of the single point mutants derived from CaMKI-conf and CaMKI-e2 (see *CaM variants-peptide binding: Fluorescence assays*) also bound with higher relative affinity to $CaMKI_p$ than $smMLCK_p$ (Figures A-6C, A-6D). The V91F, V91Y, and V108F mutations appear to convey some specificity toward $CaMKI_p$. The E84I mutation interferes with binding to both peptides, whereas V55F affords some specificity toward $smMLCK_p$.

Finally, because of the difficulty in getting K_D values from the equilibrium binding experiments, we also attempted competition Biacore assays. These assays were a proof-of-concept experiment and were not completed in full. Instead, only the binding between variants and $smMLCK_p$ was investigated and compared to the fluorescence assays and the equilibrium Biacore assays. In the competition Biacore experiments, each sample contains a constant amount of CaM or CaM variant and a variable amount of competing peptide, in this case non-biotinylated $smMLCK_p$. A set of sensorgrams for each variant was collected and the initial rate of binding was plotted as a function of the concentration of inhibiting peptide in the sample (Figure A-7).

The competition assays produced reasonable data with small error. The rankings generally agree well with the rankings for the variants binding to $smMLCK_p$ in the equilibrium experiment (compare Figure A-6E and Figure A-7C). The competition experiment shows that, as expected, both positive designs modeled after the $smMLCK_p$-CaM structure have high affinity for $smMLCK_p$. In contrast to previous experiments and the expectation of the design, CaMKI-e2 binds $smMLCK_p$ as well as WT. These results also confirm the equilibrium experiment conclusion that CaMKI-conf and CaM8 both

bind with significantly lower affinity than WT to $smMLCK_p$. The single variant results indicate that V108F is an unfavorable mutation for binding to $smMLCK_p$, and mutation at position 91 slightly disrupts the interaction.

This proof-of-concept Biacore experiment could be expanded to test the variants against CaMKI to determine whether any specificity changes have occurred and also to test the negative design variants: L18I, M51F, M51Y, L18I/M51F, and L18I/M51Y.

Conclusions

Assaying CaM-peptide complexes. It was shown during the course of this work that assaying CaM-peptide complexes by tryptophan fluorescence is not ideal. The concentration of peptide that must be used to obtain a good signal to noise is too high to evaluate affinities near 1 nM. Instead, we investigated assays using SPR technology (Biacore). The CaM-smMLCK peptide complex does not exhibit 1:1 binding as expected in the kinetics experiment, so we were forced to study the equilibrium binding to obtain relative affinities between the variants. Even with the equilibrium studies, the CaM complexes did not reach saturation even at 50 μM CaM, indicating that aggregation or self-binding was occurring at high concentrations. We were, however, able to rank the variants tested for their affinities to both the smMLCK and CaMKI peptides to test our computational designs. Additionally, we provided proof-of-concept work that a competition-based assay using the Biacore can provide more accurate data including K_Ds for the interactions. Future work for CaM-peptide interactions of high affinity should be investigated using the competition-based method.

Evaluating computational designs. Because of the difficulty in assaying CaM-peptide affinities, it was difficult to accurately assess our computational designs. We were, however, able to predict variants that would destabilize binding to $smMLCK_p$ by using computational design to predict mutations that would stabilize a competing structure, including V108F, which on its own as well as in CaMKI-conf acts to decrease the affinity for $smMLCK_p$. Also, one of our negative design variants, M51Y, shows a 10-fold preference for $smMLCK_p$ over $CaMKI_p$. Further work to characterize the negative and positive design variants may provide more insight into the computational designs. Additional work is also required to optimize the negative design scoring function, cutoff parameters, and energy thresholds to obtain the best negative design results.

CaM-peptide complexes as a model system for specificity. In addition to the difficultly in determining the binding affinities for CaM-peptide complexes, it is unclear whether the CaM system is ideal for investigating protein-protein specificity. While CaM binds many targets with high affinity, making it a seemingly good candidate, this broad promiscuity may incorporate too much flexibility into the protein-protein interface. A previous study from the Mayo lab showed that single mutations to glutamate residues near the interface can significantly affect the binding to $smMLCK_p$.[17] It is likely, however, that these mutations not only affect binding but also protein stability in general. It would therefore be useful to probe the interface residues with a site-saturation mutagenesis library to determine which residues are consistent with binding at each position. This type of experiment would lead to a great deal of data that could be used to inform the computational designs and perhaps greatly increase our understanding of

protein-protein specificity. In addition, by incorporating mutations that destroy the binding, we can gain insight into how much mutation is tolerable to the system. Ideally, a model system for specificity would have some flexibility but also be very sensitive to mutation in order to affect binding significantly. Calmodulin, because it binds so many non-homologous native substrates, may have too much flexibility in the binding to be a good model.

Although perhaps not ideal for protein-protein specificity, this system can be very valuable for studying protein-protein interactions. Recently, Roger Tsien in collaboration with David Baker designed a specific peptide-CaM interface using computational design.[6] In this case, the peptide is a designed variation of smMLCKp that incorporates bulky groups that clash with wild-type CaM, employing a "knobs-and-holes" approach. CaM was then designed around this new peptide, giving a specific interaction. The CaM-peptide system seems ideal for creating new interactions between protein and peptide, taking advantage of the inherently high specificity and large flexibility in the system. Future work to design other non-native peptide interactions with CaM may prove to be quite successful.

Acknowledgements

I want to thank Jost Vielmetter in the Protein Expression Center at Caltech for his invaluable help and advice with the Biacore experiments. I would also like to acknowledge Ben Allen, Heidi Privett, and the rest of the Mayo Lab, without whom the setup and execution of the computational designs would have been much more difficult.

References

1. Schueler-Furman, O., Wang, C., Bradley, P., Misura, K. & Baker, D. (2005). Progress in modeling of protein structures and interactions. *Science* **310**, 638-42.

2. Gomez, S. M., Choi, K. & Wu, Y. (2008). Prediction of protein-protein interaction networks. *Curr Protoc Bioinformatics* **Chapter 8**, Unit 8.2.

3. Fernandez-Ballester, G. & Serrano, L. (2006). Prediction of protein-protein interaction based on structure. *Methods Mol Biol* **340**, 207-34.

4. Chevalier, B. S., Kortemme, T., Chadsey, M. S., Baker, D., Monnat, R. J. & Stoddard, B. L. (2002). Design, activity, and structure of a highly specific artificial endonuclease. *Mol Cell* **10**, 895-905.

5. Fajardo-Sanchez, E., Stricher, F., Paques, F., Isalan, M. & Serrano, L. (2008). Computer design of obligate heterodimer meganucleases allows efficient cutting of custom DNA sequences. *Nucleic Acids Res* **36**, 2163-73.

6. Palmer, A. E., Giacomello, M., Kortemme, T., Hires, S. A., Lev-Ram, V., Baker, D. & Tsien, R. Y. (2006). Ca^{2+} indicators based on computationally redesigned calmodulin-peptide pairs. *Chem Biol* **13**, 521-30.

7. Reina, J., Lacroix, E., Hobson, S. D., Fernandez-Ballester, G., Rybin, V., Schwab, M. S., Serrano, L. & Gonzalez, C. (2002). Computer-aided design of a PDZ domain to recognize new target sequences. *Nat Struct Biol* **9**, 621-7.

8. Huang, J., Koide, A., Makabe, K. & Koide, S. (2008). Design of protein function leaps by directed domain interface evolution. *Proc Natl Acad Sci U S A* **105**, 6578-83.

9. Schneider, S., Buchert, M., Georgiev, O., Catimel, B., Halford, M., Stacker, S. A., Baechi, T., Moelling, K. & Hovens, C. M. (1999). Mutagenesis and selection of PDZ domains that bind new protein targets. *Nat Biotechnol* **17**, 170-5.

10. Sidhu, S. S. & Koide, S. (2007). Phage display for engineering and analyzing protein interaction interfaces. *Curr Opin Struct Biol* **17**, 481-7.

11. Potapov, V., Reichmann, D., Abramovich, R., Filchtinski, D., Zohar, N., Ben Halevy, D., Edelman, M., Sobolev, V. & Schreiber, G. (2008). Computational redesign of a protein-protein interface for high affinity and binding specificity using modular architecture and naturally occurring template fragments. *J Mol Biol* **384**, 109-19.

12. Joachimiak, L. A., Kortemme, T., Stoddard, B. L. & Baker, D. (2006). Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein-protein interface. *J Mol Biol* **361**, 195-208.

13. Kortemme, T., Joachimiak, L. A., Bullock, A. N., Schuler, A. D., Stoddard, B. L. & Baker, D. (2004). Computational redesign of protein-protein interaction specificity. *Nat Struct Mol Biol* **11**, 371-9.

14. Reynolds, K. A., Hanes, M. S., Thomson, J. M., Antczak, A. J., Berger, J. M., Bonomo, R. A., Kirsch, J. F. & Handel, T. M. (2008). Computational redesign of the SHV-1 beta-lactamase/beta-lactamase inhibitor protein interface. *J Mol Biol* **382**, 1265-75.

15. Huang, P. S., Love, J. J. & Mayo, S. L. (2007). A de novo designed protein protein interface. *Protein Sci* **16**, 2770-4.

16. Shifman, J. M. & Mayo, S. L. (2002). Modulating calmodulin binding specificity through computational protein design. *J Mol Biol* **323**, 417-23.

17. Shifman, J. M. & Mayo, S. L. (2003). Exploring the origins of binding specificity through the computational redesign of calmodulin. *Proc Natl Acad Sci U S A* **100**, 13274-9.

18. DeGrado, W. F., Prendergast, F. G., Wolfe, H. R., Jr. & Cox, J. A. (1985). The design, synthesis, and characterization of tight-binding inhibitors of calmodulin. *J Cell Biochem* **29**, 83-93.

19. Green, D. F., Dennis, A. T., Fam, P. S., Tidor, B. & Jasanoff, A. (2006). Rational design of new binding specificity by simultaneous mutagenesis of calmodulin and a target peptide. *Biochemistry* **45**, 12547-59.

20. Sammond, D. W., Eletr, Z. M., Purbeck, C., Kimple, R. J., Siderovski, D. P. & Kuhlman, B. (2007). Structure-based protocol for identifying mutations that enhance protein-protein binding affinities. *J Mol Biol* **371**, 1392-404.

21. Sood, V. D. & Baker, D. (2006). Recapitulation and design of protein binding peptide structures and sequences. *J Mol Biol* **357**, 917-27.

22. Havranek, J. J. & Harbury, P. B. (2003). Automated design of specificity in molecular recognition. *Nat Struct Biol* **10**, 45-52.

23. Bolon, D. N., Grant, R. A., Baker, T. A. & Sauer, R. T. (2005). Specificity versus stability in computational protein design. *Proc Natl Acad Sci U S A* **102**, 12724-9.

24. Dahiyat, B. I. & Mayo, S. L. (1997). De novo protein design: fully automated sequence selection. *Science* **278**, 82-7.

25. Malakauskas, S. M. & Mayo, S. L. (1998). Design, structure and stability of a hyperthermophilic protein variant. *Nat Struct Biol* **5**, 470-5.

26. Marshall, S. A. & Mayo, S. L. (2001). Achieving stability and conformational specificity in designed proteins via binary patterning. *J Mol Biol* **305**, 619-31.

27. Datta, D., Wang, P., Carrico, I. S., Mayo, S. L. & Tirrell, D. A. (2002). A designed phenylalanyl-tRNA synthetase variant allows efficient in vivo incorporation of aryl ketone functionality into proteins. *J Am Chem Soc* **124**, 5652-3.

28. Bolon, D. N. & Mayo, S. L. (2001). Enzyme-like proteins by computational design. *Proc Natl Acad Sci U S A* **98**, 14274-9.

29. Lassila, J. K., Keeffe, J. R., Oelschlaeger, P. & Mayo, S. L. (2005). Computationally designed variants of Escherichia coli chorismate mutase show altered catalytic activity. *Protein Eng Des Sel* **18**, 161-3.

30. Dunbrack, R. L., Jr. & Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci* **6**, 1661-81.

31. Lassila, J. K., Privett, H. K., Allen, B. D. & Mayo, S. L. (2006). Combinatorial methods for small-molecule placement in computational enzyme design. *Proc Natl Acad Sci U S A* **103**, 16710-5.

32. Mayo, S. L., Olafson, B. D. & Goddard, W. A. (1990). Dreiding - a generic force-field for molecular simulations. *J Phys Chem* **94**, 8897-8909.

33. Dahiyat, B. I. & Mayo, S. L. (1997). Probing the role of packing specificity in protein design. *Proc Natl Acad Sci U S A* **94**, 10172-7.

34. Street, A. G. & Mayo, S. L. (1998). Pairwise calculation of protein solvent-accessible surface areas. *Fold Des* **3**, 253-8.

35. Gordon, D. B., Hom, G. K., Mayo, S. L. & Pierce, N. A. (2003). Exact rotamer optimization for protein design. *J Comput Chem* **24**, 232-43.

36. Gordon, D. B. & Mayo, S. L. (1998). Radical performance enhancements for combinatorial optimization algorithms based on the dead-end elimination theorem. *J Comp Chem* **19**, 1505-14.

37. Pierce, N. A., Spriet, J. A., Desmet, J. & Mayo, S. L. (2000). Conformational splitting: A more powerful criterion for dead-end elimination. *J Comp Chem* **21**, 999-1009.

38. Allen, B. D. & Mayo, S. L. (2006). Dramatic performance enhancements for the FASTER optimization algorithm. *J Comput Chem* **27**, 1071-5.

39. Desmet, J., Spriet, J. & Lasters, I. (2002). Fast and accurate side-chain topology and energy refinement (FASTER) as a new method for protein structure optimization. *Proteins* **48**, 31-43.

40. Kirkpatrick, S., Gelatt, C. D., Jr. & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science* **220**, 671-80.

41. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953). Equation of state calculations by fast computing machines. *J Chem Phys* **21**, 1087-92.

42. Stevens, F. C. (1983). Calmodulin: an introduction. *Can J Biochem Cell Biol* **61**, 906-10.

43. Jurado, L. A., Chockalingam, P. S. & Jarrett, H. W. (1999). Apocalmodulin. *Physiol Rev* **79**, 661-82.

44. Hoeflich, K. P. & Ikura, M. (2002). Calmodulin in action: diversity in target recognition and activation mechanisms. *Cell* **108**, 739-42.

45. Degrado, W. F. (1988). Design of peptides and proteins. *Adv Protein Chem* **39**, 51-124.

46. O'Neil, K. T. & DeGrado, W. F. (1990). How calmodulin binds its targets: sequence independent recognition of amphiphilic alpha-helices. *Trends Biochem Sci* **15**, 59-64.

47. Cox, J. A., Comte, M., Fitton, J. E. & DeGrado, W. F. (1985). The interaction of calmodulin with amphiphilic peptides. *J Biol Chem* **260**, 2527-34.

48. Kranz, J. K., Lee, E. K., Nairn, A. C. & Wand, A. J. (2002). A direct test of the reductionist approach to structural studies of calmodulin activity: relevance of peptide models of target proteins. *J Biol Chem* **277**, 16351-4.

49. 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**, 1251-5.

50. Clapperton, J. A., Martin, S. R., Smerdon, S. J., Gamblin, S. J. & Bayley, P. M. (2002). Structure of the complex of calmodulin with the target sequence of calmodulin-dependent protein kinase I: studies of the kinase activation mechanism. *Biochemistry* **41**, 14669-79.

51. Kurokawa, H., Osawa, M., Kurihara, H., Katayama, N., Tokumitsu, H., Swindells, M. B., Kainosho, M. & Ikura, M. (2001). Target-induced conformational adaptation of calmodulin revealed by the crystal structure of a complex with nematode Ca(2+)/calmodulin-dependent kinase kinase peptide. *J Mol Biol* **312**, 59-68.

52. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35**, W375-83.

53. Hemsley, A., Arnheim, N., Toney, M. D., Cortopassi, G. & Galas, D. J. (1989). A simple method for site-directed mutagenesis using the polymerase chain reaction. *Nucleic Acids Res* **17**, 6545-51.

54. Persechini, A., Blumenthal, D. K., Jarrett, H. W., Klee, C. B., Hardy, D. O. & Kretsinger, R. H. (1989). The effects of deletions in the central helix of calmodulin on enzyme activation and peptide binding. *J Biol Chem* **264**, 8052-8.

55. Gill, S. C. & von Hippel, P. H. (1989). Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* **182**, 319-26.

56. Maune, J. F., Klee, C. B. & Beckingham, K. (1992). Ca2+ binding and conformational change in two series of point mutations to the individual Ca(2+)-binding sites of calmodulin. *J Biol Chem* **267**, 5286-95.

57. Lazar, G. A., Dang, W., Karki, S., Vafa, O., Peng, J. S., Hyun, L., Chan, C., Chung, H. S., Eivazi, A., Yoder, S. C., Vielmetter, J., Carmichael, D. F., Hayes, R. J. & Dahiyat, B. I. (2006). Engineered antibody Fc variants with enhanced effector function. *Proc Natl Acad Sci U S A* **103**, 4005-10.

58. Gopalakrishna, R. & Anderson, W. B. (1982). Ca2+-induced hydrophobic site on calmodulin: application for purification of calmodulin by phenyl-Sepharose affinity chromatography. *Biochem Biophys Res Commun* **104**, 830-6.

59. 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**, 4704-12.

60. Brockerhoff, S. E., Edmonds, C. G. & Davis, T. N. (1992). Structural analysis of wild-type and mutant yeast calmodulins by limited proteolysis and electrospray ionization mass spectrometry. *Protein Sci* **1**, 504-16.

61. Rhyner, J. A., Koller, M., Durussel-Gerber, I., Cox, J. A. & Strehler, E. E. (1992). Characterization of the human calmodulin-like protein expressed in Escherichia coli. *Biochemistry* **31**, 12826-32.

62. Findlay, W. A., Martin, S. R., Beckingham, K. & Bayley, P. M. (1995). Recovery of native structure by calcium binding site mutants of calmodulin upon binding of sk-MLCK target peptides. *Biochemistry* **34**, 2087-94.

63. Mirzoeva, S., Weigand, S., Lukas, T. J., Shuvalova, L., Anderson, W. F. & Watterson, D. M. (1999). Analysis of the functional coupling between calmodulin's calcium binding and peptide recognition properties. *Biochemistry* **38**, 3936-47.

64. Martin, S. R. & Bayley, P. M. (2002). Regulatory implications of a novel mode of interaction of calmodulin with a double IQ-motif target sequence from murine dilute myosin V. *Protein Sci* **11**, 2909-23.

65. O'Neil, K. T., Wolfe, H. R., Jr., Erickson-Viitanen, S. & DeGrado, W. F. (1987). Fluorescence properties of calmodulin-binding peptides reflect alpha-helical periodicity. *Science* **236**, 1454-6.

66. Montigiani, S., Neri, G., Neri, P. & Neri, D. (1996). Alanine substitutions in calmodulin-binding peptides result in unexpected affinity enhancement. *J Mol Biol* **258**, 6-13.

67. Sacks, D. B., Mazus, B. & Joyal, J. L. (1995). The activity of calmodulin is altered by phosphorylation: modulation of calmodulin function by the site of phosphate incorporation. *Biochem J* **312 (Pt 1)**, 197-204.

68. Takano, E., Hatanaka, M. & Maki, M. (1994). Real-time-analysis of the calcium-dependent interaction between calmodulin and a synthetic oligopeptide of calcineurin by a surface plasmon resonance biosensor. *FEBS Lett* **352**, 247-50.

Table 5-1. CaM-binding peptides.

Peptide	Sequence[a]	MW (g/mol)	ε_{280} ($M^{-1}cm^{-1}$) [d]
smMLCK	ARRKWQKTGHAVRAIGRLSS[b]	2278.6	5690
CaMKI	IKKNFAKSKWKQAFNATAVVRHMRK[c]	2988.6	5690
biotin-smMLCK	biotin-LC-LC-GGSGG-ARRKWQKTGHAVRAIGRLSS	3047.6	5690
biotin-CaMKI	biotin-LC-LC-GGSGG- IKKNFAKSKWKQAFNATAVVRHMRK	3757.6	5690

[a] LC: 6-aminohexanoic acid linker ("long-chain").
[b] smMLCK peptide sequence corresponds to residues 796-815 of smMLCK.
[c] CaMKI peptide sequence corresponds to residues 294-318 of CaMKI.
[d] Extinction coefficients were theoretically calculated based on amino acid composition.[55]

Table A-2. Sequence alignment of calmodulin designs.

Position	11	12	15	18	19	32	36	39	51	55	68	71	72	84	88	91	92	105	108	109	112	124	128	144	145
WT	E	F	A	L	F	L	M	L	M	V	F	M	M	E	A	V	F	L	V	M	L	M	A	M	M
CaM8[16]	L	Y	-	-	-	-	-	-	-	I	-	-	(E76)[a]	Y	-	I	-	-	-	L	-	-	-	-	I
smMLCK-conf	-	-	-	-	-	-	-	-	Y	I	-	-	-	-	-	I	-	-	I	-	-	-	-	-	-
CaMKI-conf	-	-	-	-	-	-	-	I	-	F	-	-	-	I	-	Y	-	-	F	-	-	-	-	-	-
Conf Negative Design	-	Y	-	I	-	-	-	I	F	I	-	-	-	-	-	I	-	-	I	-	-	-	-	-	-
smMLCK-e2	-	-	-	I	-	-	-	-	F	I	-	-	-	-	-	I	-	-	I	-	-	-	-	-	-
CaMKI-e2	-	-	-	-	-	-	-	-	-	F	-	-	-	-	-	F	-	-	I	-	-	-	-	-	-
e2 Negative Design	L	Y	-	I	-	-	-	-	Y	I	-	-	-	L	-	I	-	-	I	L	-	-	-	-	-

[a] The CaM8 design did not include position 72 in the design, but included position 76, which was mutated to Glu.[16] Position 76 was not included in the current designs.

Table A-3. K_Ds for CaM variants as determined by fluorescence assays.

Variant	K_D (nM)	
	smMLCK$_p$	**CaMKI$_p$**
WT[16]	1.8 ± 1.3	1.7 ± 0.7
WT	1.6 ± 0.9	3.5 ± 1.2
L18I	5.0 ± 3.1	6.2 ± 2.2
M51F	1.2 ± 0.9	1.9 ± 1.4
M51Y	0.4 ± 0.8	4.4 ± 1.5
L18I/M51F	5.8 ± 2.4	3.2 ± 1.3
L18I/M51Y	0.4 ± 0.7	0.9 ± 1.4
smMLCK-conf	1.7 ± 0.6	3.6 ± 0.8
smMLCK-e2	0.6 ± 0.8	0.2 ± 0.5
CaMKI-conf	4.4 ± 1.5	2.0 ± 1.1
CaMKI-e2	0.6 ± 0.6	7.0 ± 1.9

Table A-4. K_Ds for $CaMKI_p$ positive design-derived single mutant variants as determined by fluorescence assays.

Variant	K_D (nM)	
	$smMLCK_p$	**$CaMKI_p$**
V55F	1.4 ± 1.3	4.8 ± 1.3
E84I	0.4 ± 0.7	2.6 ± 1.7
V91F	2.1 ± 1.1	1.6 ± 0.9
V91Y	3.0 ± 1.1	2.3 ± 1.6
V108F	4.4 ± 1.9	7.6 ± 2.6

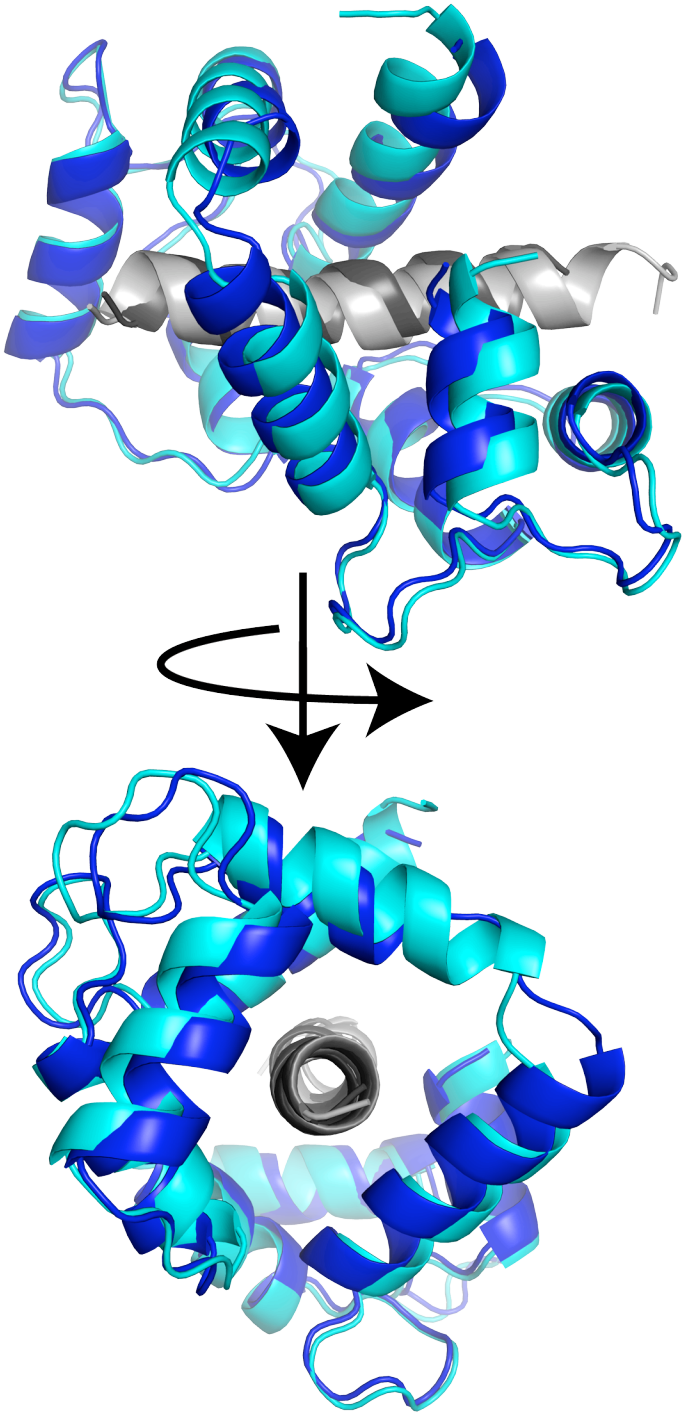


Figure A-1. CaM-peptide structures. The CaM-smMLCK$_p$ structure is shown in dark blue (CaM) and dark gray (peptide), and the CaM-CaMKI$_p$ structure is shown in cyan and light gray. The Cα RMSD between the two structures is 1.4 Å. The peptide is bound between the N- and C-terminal domains of CaM such that the peptide is almost completely buried.

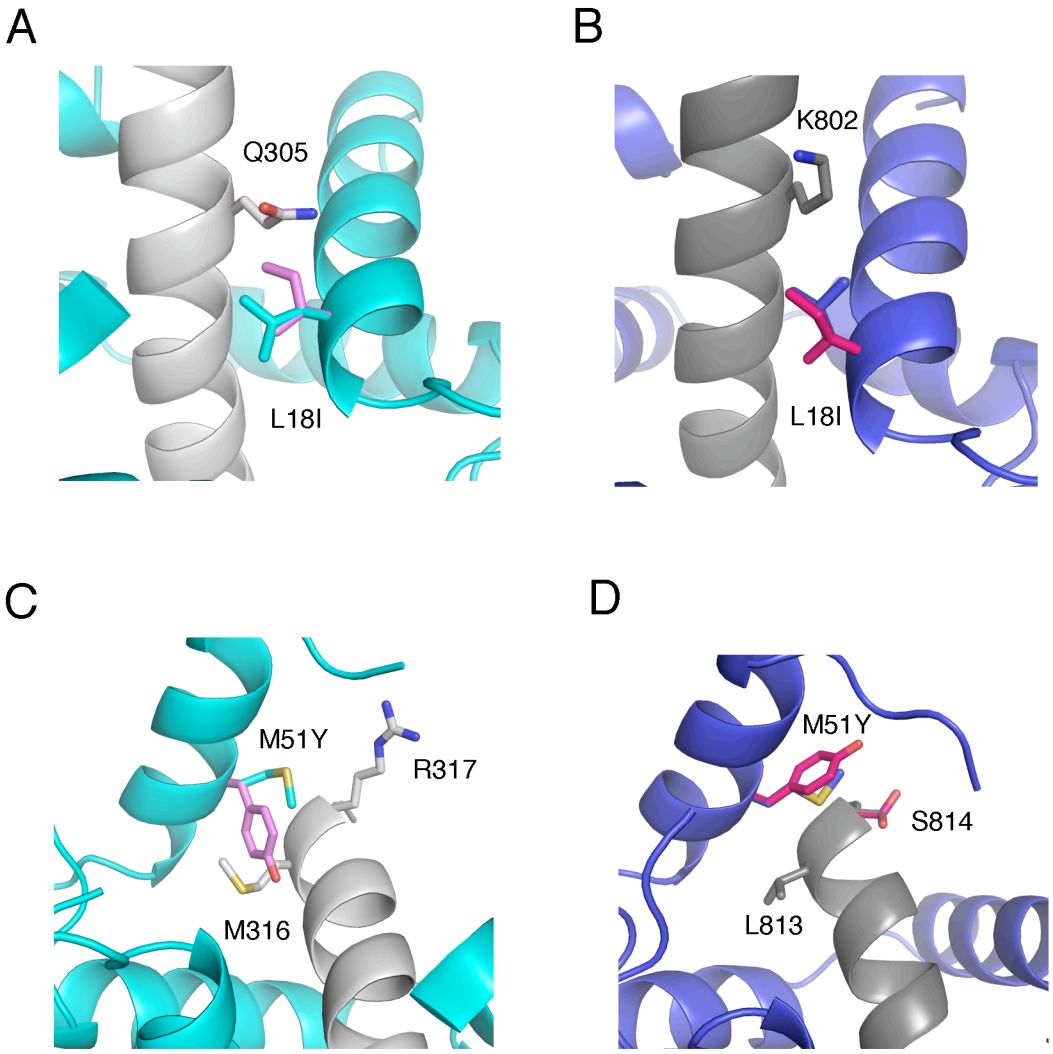


Figure A-2. Negative design result structures. The CaM-CaMKI$_p$ structure (1MXE) is shown in light colors (panels A and C), and the CaM-smMLCK$_p$ structure (1CDL) is shown in dark colors (panels B and D). The peptides are shown in gray, CaM is shown in blue, and the designed residues of interest are shown in magenta. (A) An Ile at position 18 cannot be accommodated in the CaMKI$_p$-bound structure due to Gln 305 in the peptide. This creates an unfavorable van der Waals interaction, which is expected to destabilize the complex. (B) In the smMLCKp structure, however, the Lys in the equivalent position of smMLCK$_p$ takes on a conformation that does not clash with position 18. (C) At position 51, a Tyr mutation in the negative design result from the e2 library clashes with M316 of CaMKI$_p$. The Tyr side chain is pointed inward toward the peptide because it is hindered from pointing toward the solvent by the long side chain of Arg 317. (D) The smaller Ser residue at position 814 (equivalent to R317) in the smMLCK peptide allows the Tyr at position 51 to point toward the solvent, eliminating the unfavorable interactions with the peptide that are seen in the CaMKI$_p$ structure.

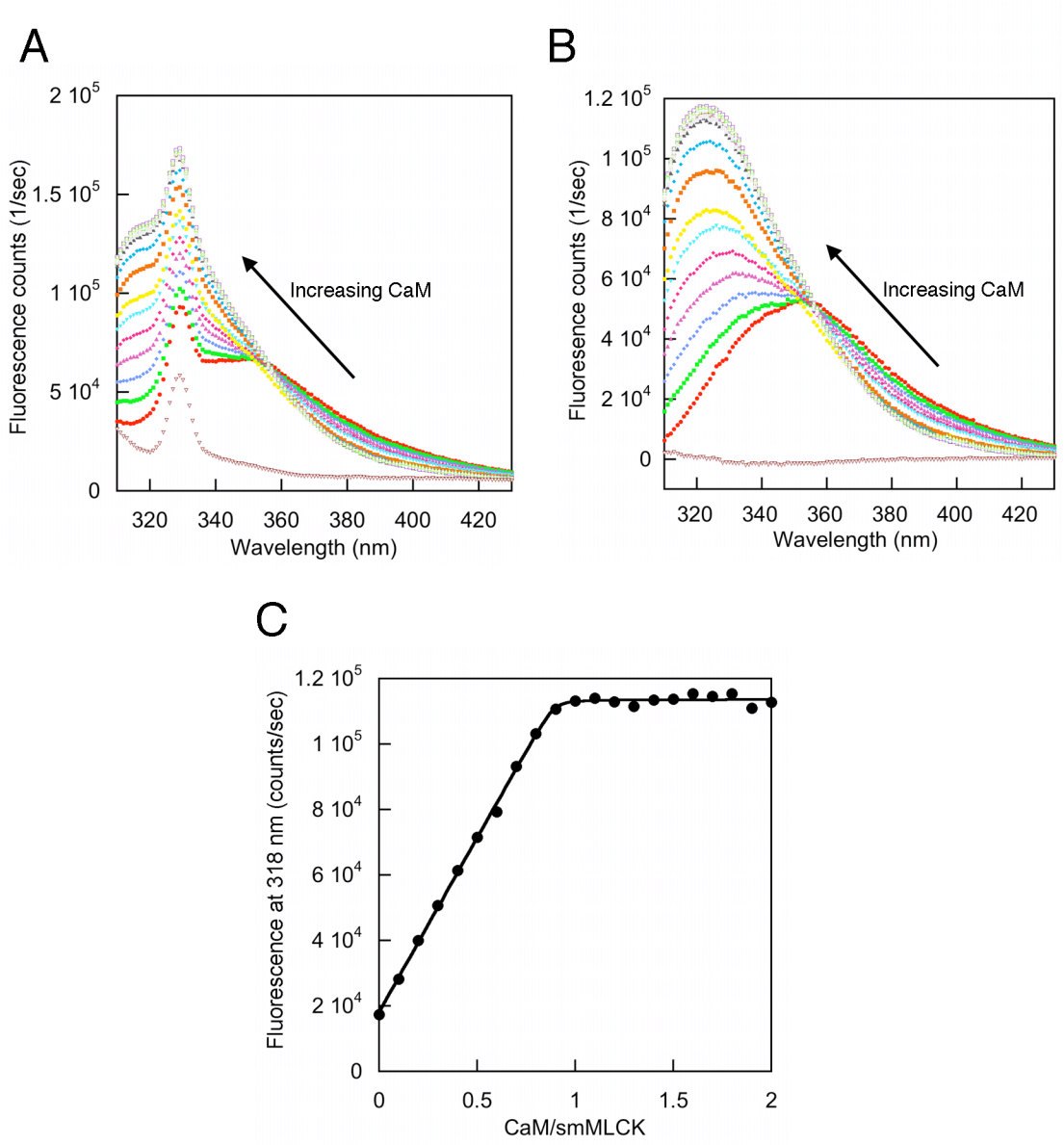


Figure A-3. Tryptophan fluorescence assay data analysis. (A) Samples containing 1 mM peptide and varying amounts of CaM are excited at 295 nm and the emission spectrum from the tryptophan fluorescence is measured. The brown curve at the bottom of both panels (A) and (B) is CaM alone with no peptide present. (B) The buffer is subtracted giving smooth curves that show a blue-shift and increase in intensity, which indicates burial of the tryptophan. (C) The fluorescence intensity at 318 nm is plotted as a function of the ratio of CaM to peptide. This titration curve shows that the binding between CaM and smMLCK$_p$ is 1:1 and the data can be fit to Equation A-2, giving a K_D of interaction.

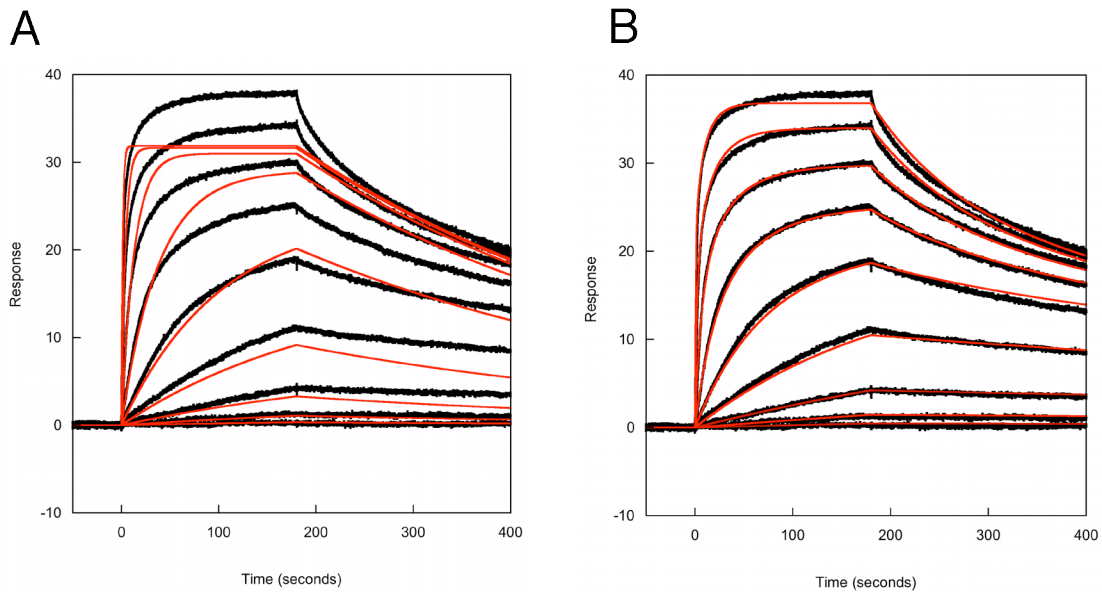


Figure A-4. Biacore kinetics of wild type CaM and smMLCK$_p$. Biacore data are shown in black and model curve fits are shown in red. Approximately 15 response units of biotinylated smMLCK$_p$ was immobilized to streptavidin on a CM5 chip. Various concentrations of wild-type CaM were flowed over the peptide surface and binding was monitored. BiaEvaluation software was used to fit the data to a (A) 1:1 model or a (B) multivalent model. Neither model was a satisfactory fit due to previous work showing that the complex should be 1:1.

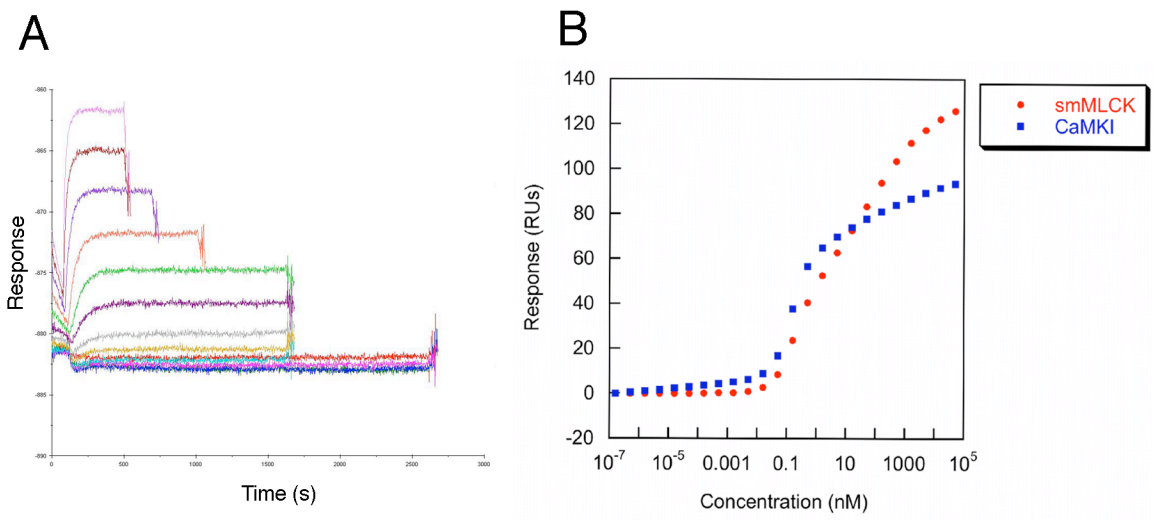
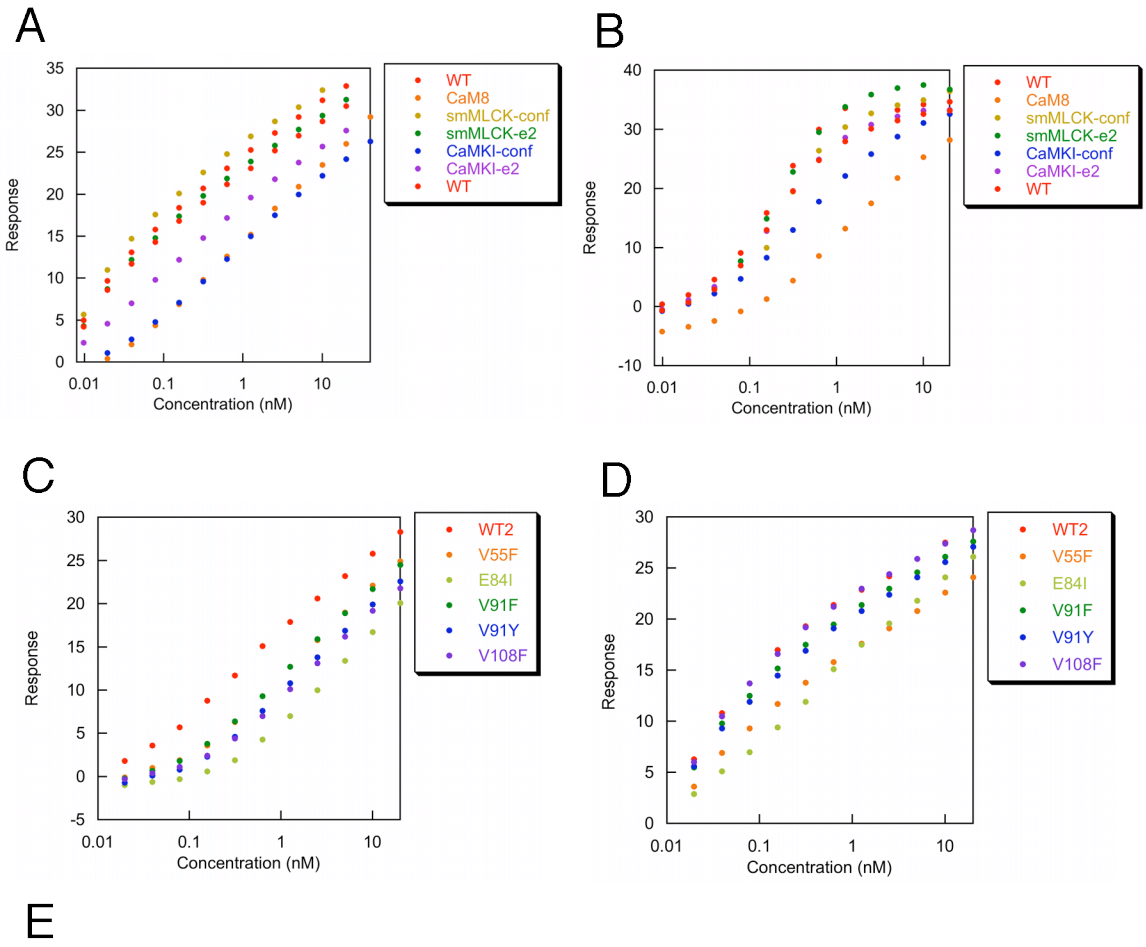


Figure A-5. Biacore equilibrium data. (A) Typical equilibrium data set. Multiple concentrations of CaM are injected over the peptide surface for various amounts of time until equilibrium binding is reached. (B) The equilibrium response value at each concentration is plotted as a function of concentration. For the wild-type interaction with both $smMLCK_p$ and $CaMKI_p$, the interaction does not reach an R_{max}. This indicates that at high concentrations of CaM we are seeing a type of self-binding or aggregation.



For $smMLCK_p$ binding:

WT = smMLCK-conf = smMLCK-e2 < CaMKI-e2 < CaMKI-conf = CaM8

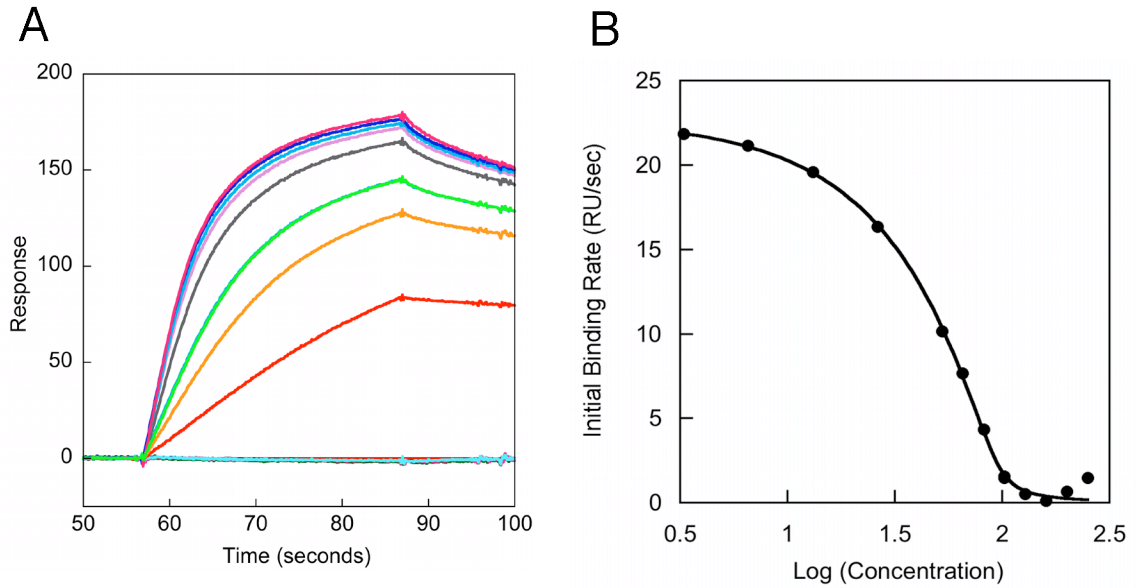
WT < V55F = V91F = V91Y = V108F < E84I

For $CaMKI_p$ binding:

WT = smMLCK-conf = smMLCK-e2 = CaMKI-e2 < CaMKI-conf < CaM8

WT = V91F = V91Y = V108F < V55F = E84I

Figure A-6. CaM variants–peptide equilibrium binding data. (A) Positive design CaM variants binding to $smMLCK_p$. (B) Positive design CaM variants binding to $CaMKI_p$. (C) Single mutant CaM variants derived from CaMKI-CaM positive design constructs binding to $smMLCK_p$. (D) Single mutant CaM variants derived from CaMKI-CaM positive design constructs binding to $CaMKI_p$. (E) Ranking of variants compared to WT based on Biacore equilibrium experiments.



C

Variant	KD (nM)
	$smMLCK_p$
WT	1.2 ± 0.4
CaM8	9.0 ± 1.9
smMLCK-conf	0.4 ± 0.1
smMLCK-e2	0.5 ± 0.2
CaMKI-conf	3.9 ± 1.0
CaMKI-e2	1.3 ± 0.2
V55F	1.2 ± 0.2
E84I	0.9 ± 0.2
V91F	2.5 ± 0.5
V91Y	2.3 ± 0.5
V108F	6.1 ± 1.8
WT	1.0 ± 0.3

Figure A-7. Competition Biacore assay. (A) Sensorgrams showing the binding of CaM to the $smMLCK_p$ surface and including increasing amounts of inhibiting $smMLCK_p$. As competitor is added, the maximal response decreases, as expected. (B) A plot of the initial binding rate as a function of the log of the competitor concentration. The initial rates are calculated from the first few seconds of each sensorgram as seen in (A). The data are fit to Equation A-3 to give a K_D. (C) Results of the competition assay. WT CaM was run both at the beginning of the experiment and the end to confirm the stability of the immobilized surface and for reproducibility.