Calmodulation meta-analysis: Predicting calmodulin binding via canonical motif clustering

Karen Mruk,[1] Brian M. Farley,[1] Alan W. Ritacco,[2] and William R. Kobertz[1]

[1]Department of Biochemistry and Molecular Pharmacology and Programs in Chemical Biology and Neuroscience, and
[2]Department of Scientific and Research Computing, University of Massachusetts Medical School, Worcester, MA 01605

The calcium-binding protein calmodulin (CaM) directly binds to membrane transport proteins to modulate their function in response to changes in intracellular calcium concentrations. Because CaM recognizes and binds to a wide variety of target sequences, identifying CaM-binding sites is difficult, requiring intensive sequence gazing and extensive biochemical analysis. Here, we describe a straightforward computational script that rapidly identifies canonical CaM-binding motifs within an amino acid sequence. Analysis of the target sequences from high resolution CaM–peptide structures using this script revealed that CaM often binds to sequences that have multiple overlapping canonical CaM-binding motifs. The addition of a positive charge discriminator to this meta-analysis resulted in a tool that identifies potential CaM-binding domains within a given sequence. To allow users to search for CaM-binding motifs within a protein of interest, perform the meta-analysis, and then compare the results to target peptide–CaM structures deposited in the Protein Data Bank, we created a website and online database. The availability of these tools and analyses will facilitate the design of CaM-related studies of ion channels and membrane transport proteins.

INTRODUCTION

The Ca^{2+}-binding protein calmodulin (CaM) directly binds to membrane transport proteins to regulate membrane excitability and Ca^{2+}-dependent intracellular signal transduction cascades. CaM communicates changes in intracellular Ca^{2+} levels to channels and transporters by binding to the cytoplasmic domains of these proteins to modulate protein function (calmodulation) (Mruk et al., 2012; Biswas et al., 2013; Ben-Johny and Yue, 2014). The physiological relevance of calmodulation is highlighted by the disease-associated mutations that disrupt CaM–ion channel protein interactions (Weiss et al., 2003; Ghosh et al., 2006; Shamgar et al., 2006; Etxeberria et al., 2008; Alaimo et al., 2009; Hino et al., 2012). Accordingly, many biochemical and structural investigations have focused on determining CaM-binding sites in peptides derived from the water-soluble domains of membrane proteins. Although these concerted efforts have resulted in identifying more proteins that bind to CaM, identifying CaM-binding sites in proteins is still mostly a haphazard exploration, requiring unguided, brute force experimentation.

Part of the challenge of discovering and investigating novel CaM–membrane transport protein interactions is that CaM-binding sites do not contain high sequence similarity. Instead, CaM targets often share common biochemical and biophysical characteristics such as a propensity to form amphipathic α helices, net positive charge, moderate hydrophilicity, and hydrophobic anchor residues (Rhoads and Friedberg, 1997). Because of the lack of a well-defined CaM-binding consensus sequence, CaM-binding motifs have been classified by the spacing between hydrophobic anchor residues that are broadly characterized into two subgroups: Ca^{2+}-independent binding and Ca^{2+}-dependent binding (Table 1). A comparison of proteins that bind to CaM in the absence of Ca^{2+} identified the hallmark IQ sequence motif (Rhoads and Friedberg, 1997). In this motif, amino acids at positions 1, 2, 5, 6, 11, and 14 are highly conserved. Sequences containing different amino acids at these positions are classified as IQ-like motifs, which CaM binds to in the presence or absence of Ca^{2+}. In addition to IQ-like motifs, Ca^{2+} binding to CaM induces a conformational change that promotes binding to additional target proteins, which do not contain easily identifiable motifs. A closer examination of these targets showed that the primary requirement for CaM binding is the presence of bulky hydrophobic residues: Phe, Ile, Leu, Val, and Trp, at the

---

first and last position of the binding region (Rhoads and Friedberg, 1997), which anchor the protein into the two lobes of CaM (LaPorte et al., 1980). The remaining intermediate residues between the anchors are highly variable in both sequence and spacing because the central CaM helix is flexible (Seaton et al., 1985; Ikura et al., 1991; Barbato et al., 1992), enabling CaM to bind to a wide variety of protein targets.

Although most CaM-binding motifs can be grouped into categories based on the spacing between anchor residues, several high resolution structures suggest that CaM also binds atypical sequences. For example, the crystal structures of CaM bound to a peptide from the ryanodine receptor (Maximciuc et al., 2006), the plasma membrane Ca^{2+}-ATPase (PMCA) pump (Juranic et al., 2010), and the NMDA receptor (Ataman et al., 2007) show that CaM can bind to motifs that contain hydrophobic anchors either further apart (16 and 17 residues) or closer together (6 residues) than the classical Ca^{2+}-dependent binding motifs. In addition, helicity of the unbound target peptide is not a strict requirement because CaM binds to the disordered proteins neuromodulin and neurogranin (Kumar et al., 2013). Thus, the repertoire of peptide sequences that CaM binds to is difficult to categorize.

Several algorithms have been developed to predict CaM-binding domains within the proteome (Radivojac et al., 2006; Hamilton, M., A.S.N. Reddy, and A. Ben-Hur. 2011. Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine; Minhas and Ben-Hur, 2012; Wang et al., 2012). The web-accessible algorithm (Calmodulin Target Database) relies on the biochemical and biophysical properties of classic CaM targets to predict a CaM-binding site within a user-inputted sequence (Yap et al., 2000). Thus, CaM-binding sites that do not meet these biophysical criteria or stretches of protein sequence that contain multiple or partially overlapping CaM-binding sites will be missed. Moreover, these algorithms are only predictive; they do not display all of the known canonical CaM-binding motifs, leaving the experimentalist in the dark. Therefore, we developed a script to identify every canonical CaM-binding motif within a given sequence. Analysis of PDB-deposited, CaM-target peptide structures revealed that CaM often binds to regions that have multiple overlapping CaM-binding motifs. Combining this identification method with a simple charge discriminator results in reasonable predictive power (71% true positive [TP]; 78% true negative [TN]) on a test set of biochemically characterized CaM-binding motifs. Because we have found this analysis useful in both the design of experiments and analysis of experimental results, we have made the meta-analysis Perl script available for download as a ZIP file (see Online supplemental material below) and created a database that allows users to search for CaM-binding motifs within a protein of interest and perform the meta-analysis

(Mruk et al., 2014). Future updates to the database and scripts will be available at the Calmodulation Database and Meta-Analysis website (http://cam.umassmed.edu), which also allows users to compare their sequences to target peptide–CaM structures deposited in the Protein Data Bank (PDB).

MATERIALS AND METHODS

Dataset generation
To generate a set of proteins known to bind CaM, we searched the PDB for structures of CaM–peptide complexes deposited through December 2012. For each structure, the full sequence of the protein from which the peptide is derived was retrieved from the Universal Protein Resource Knowledgebase (UniProt). Duplicate sequences were identified on the basis of their UniProt identification numbers and removed from the dataset, yielding a final sample size of 48 different proteins and 52 CaM-binding sequences. The coordinates of CaM-binding proteins within the full sequences were determined by visual inspection of the structures deposited in the PDB.

Motif identification and sequence characterization
To identify the canonical CaM-binding motifs (listed in Table 1) in the protein sequences within our dataset, we used degenerate text pattern matching via regular expressions. We used a custom Perl script that searched the full sequence of each member of our dataset for subsequences that matched any of the motifs listed in Table 1. Matches to motifs were recorded and used to calculate the motif score for each amino acid, which we define as the number of motifs that include that amino acid.

To determine hydrophobicity, we used a rolling window average using multiple window sizes. For a window size N, the Kyte and Doolittle (1982) hydrophobicity values for the first N amino acids were summed and divided by N. This process was repeated for the window spanning amino acid 2 to N + 1, and for each window thereafter. Net charge was determined in a similar fashion, but the sum was reported instead of the average. The meta-analysis Perl script is available for download as a ZIP file as part of the online supplemental material (see below) and at http://cam.umassmed.edu (Mruk et al., 2014).

Parameter optimization
To establish an optimal combination of window size, motif count, hydrophobicity, and net charge for discriminating CaM-binding regions from nonbinding regions, we used our CaM-binding protein dataset to generate both a CaM-binding–negative and CaM-binding–positive test set. For each of the three window sizes we tested (8, 10, and 15 amino acids long), each full protein sequence in the dataset was divided into all possible windows. If a window was wholly contained within the CaM-binding regions we determined by visual inspection of deposited structures above, it was assigned to the positive test set. If a window was wholly outside of a CaM-binding region, it was instead assigned to the negative test set.

For each window in each test set, the motif score for each amino acid was averaged across the window. To determine the set of parameters that most accurately identified windows as CaM-binding regions or non–CaM-binding regions, we measured the percentage of windows correctly assigned to either the CaM-binding–positive test set (TPs) or the CaM-binding–negative test set (TNs) for 1,800 combinations of motif count, charge, and hydrophobicity. We chose the set of parameters that gave the best sum of TP and TN rates, which was as follows: window size of 10 amino acids, average motif count of ≥2, net charge of ≥1, and average hydrophobicity between −3 and 2.5 (Table S1).

Database comparison

To compare our database to the Calmodulin Target Database (Yap et al., 2000), amino acid sequences were culled from the 52 CaM-binding domains and entered into both our script and the Calmodulin Target Database. Results were tabulated (Table S2).

Online supplemental material

Fig. S1 shows that the addition of a positive charge discriminator to the script excludes all KCNQ4 transmembrane domains except the positively charged S4 segment. Table S1 compares the TP/TN values for the optimized parameters, and Table S2 compares the CaM motif identification and binding region predictions of the meta-analysis and the Calmodulin Target Database (Yap et al., 2000). In addition, the Perl script is available as part of the online supplemental material for download as a ZIP file, and it is also available at http://cam.umassmed.edu (Mruk et al., 2014). To run the standalone script, type (with spaces): perl NameofFile AminoAcid-Sequence 0or1 StartingAminoAcidNumber (optional). Two tab-delineated text files will be saved: (1) the meta-analysis results and (2) a list of motifs sorted by amino acid position (Option 0) or motif classification (Option 1). Note that Perl is not included with Windows OS and must be manually downloaded and installed. The program is free of charge from Perl. The online supplemental material is available at http://www.jgp.org/cgi/content/full/jgp.201311140/DC1.

RESULTS

CaM's promiscuous binding and multifarious roles in KCNQ (Kv7.x) channel modulation motivated us to devise a method to identify CaM-binding motifs in the C terminus of these channels. Previous biochemical studies show that CaM binds to helix A and/or helix B of the KCNQ C terminus through an IQ-like motif and two adjacent 1–5–10 motifs, respectively (Yus-Najera et al., 2002) (Fig. 1 A). However, for CaM to interact with full-length KCNQ channels, both target sites must be intact (Gómez-Posada et al., 2011). In addition to the canonical CaM-binding motifs, the recently determined crystal structure of CaM bound to helix B of KCNQ4 identified a noncanonical 1–14 motif in which methionine acts as the first anchoring residue (Fig. 1 A) (Xu et al., 2013). A retrospective glance at the KCNQ C termini readily identifies these published motifs, and further sequence gazing yields additional canonical motifs that are also present in the two helices. Because the unaided scanning for CaM motifs in membrane transport proteins is prone to human bias and error, we initially wrote a script that identifies all of the canonical CaM-binding motifs within a given sequence. Examination of helix A of KCNQ channels using this script identifies the conserved IQ-like motifs. For KCNQ2 channels, the "IQ" isoleucine is also part of an as yet unnoticed 1–12 motif (Fig. 1 B). In contrast to helix A, helix B of the KCNQ family contains multiple (5–16) canonical motifs (Table 2); KCNQ4's helix B has CaM motifs from every subgroup except for IQ (Fig. 1 C). We next ran our script on 48 unique CaM–peptide structures deposited in the PDB, as these structures contain well-annotated CaM-binding sites. This analysis revealed that CaM often binds to target peptides that contain multiple overlapping canonical CaM-binding motifs, suggesting a straightforward method for predicting CaM-binding domains. However, using this criterion alone on full-length channel sequences resulted in hydrophobic stretches (e.g., transmembrane domains) misidentified as potential CaM-binding motifs. Because most confirmed CaM motifs have a net positive charge, we added a simple positive charge discriminator to the script, which increased specificity by ∼50% while having only a modest effect on sensitivity (positive charge discriminator: 71% TP,

TABLE 1
Canonical CaM-binding motifs

Motif	Sequence
Ca²⁺ dependent	
1–10	[FILVW]xxxxxxxx[FILVW]
1–5–10	[FILVW]xxx[FAILVW]xxxx[FILVW]
Basic 1–5–10	[RK][RK][RK][FAILVW]xxx[FILV]xxxx[FILVW]
1–12	[FILVW]xxxxxxxxxx[FILVW]
1–14	[FILVW]xxxxxxxxxxxx[FILVW]
1–8–14	[FILVW]xxxxxx[FAILVW]xxxxx[FILVW]
1–5–8–14	[FILVW]xxx[FAILVW]xx[FAILVW]xxxxx[FILVW]
Basic 1–8–14	[RK][RK][RK][FILVW]xxxxxx[FAILVW]xxxxx[FILVW]
1–16	[FILVW]xxxxxxxxxxxxxx[FILVW]
Ca²⁺ independent	
IQ	[FILV]Qxxx[RK]Gxxx[RK]xx[FILVWY]
IQ-like[a]	[FILV]Qxxx[RK]xxxxxxxx
IQ-2A	[IVL]QxxxRxxxx[VL][KR]xW
IQ-2B	[IL]QxxCxxxxKxRxW
IQ unconventional	[IVL]QxxxRxxxx[RK]xx[FILVWY]

Numbers for the Ca²⁺-dependent motifs indicate the positions that require a hydrophobic residue. Residues in the brackets can substitute for each other; x indicates any amino acid residue.
[a]Some IQ-like motifs require Ca²⁺ for CaM binding.

78% TN; neutral charge: 85% TP, 51% TN) (Table S1). Although the charge discriminator excludes most KCNQ4 transmembrane regions, the meta-analysis does predict CaM binding to the positively charged voltage sensor (Fig. S1). Counterintuitively, the addition of a stringent hydrophobicity parameter (to exclude transmembrane domains) did not significantly improve the accuracy of the meta-analysis for all proteins or membrane proteins (Table S1).

Mapping our meta-analysis (Fig. 1 C) onto the KCNQ4 helix B–CaM crystal structure (PDB accession no. 4GOW) illustrates the utility of this simple tool (Fig. 1 D). The motif score is the number of times (hexadecimal with scores of ≥15 returning a value of "Z") a residue in the amino acid sequence is part of a unique canonical CaM-binding motif. The individual canonical motifs and their locations are shown below the KCNQ4 helix B amino acid sequence. Because both the motif score and positive charge of the highlighted sequence are equal to or greater than the cutoff values determined using the CaM target peptide test set (Materials and methods), the meta-analysis predicts a CaM-binding region in the KCNQ4

B helix. This prediction misses the first anchor in the structure by only one residue, which is expected because methionine is a noncanonical anchor. In contrast, our script identifies potential anchor residues in the distal C-terminal end of helix B that do not form contacts with CaM in the crystal structure. Although these residues are beyond the grasp of CaM in this structure, the presence of multiple canonical CaM-binding motifs within the distal end of helix B hints that CaM has the opportunity to adopt more than one conformation when bound to helix B of KCNQ channels.

Given our success at identifying CaM-binding regions in KCNQ channels, we applied the meta-analysis to the voltage-gated calcium channel, $Ca_V1.2$. $Ca_V1.2$ is an ideal test subject because it contains three well-characterized CaM-binding domains, all of which have been co-crystallized with CaM (Fig. 2 A). We first performed the meta-analysis on the pre-IQ/IQ region of $Ca_V1.2$, which predicted three regions for CaM binding (Fig. 2 B): the two regions in the pre-IQ domain each contained the anchor residue identified in the crystal structure (PDB accession no. 3G43; Fig. 2 A), and the third region
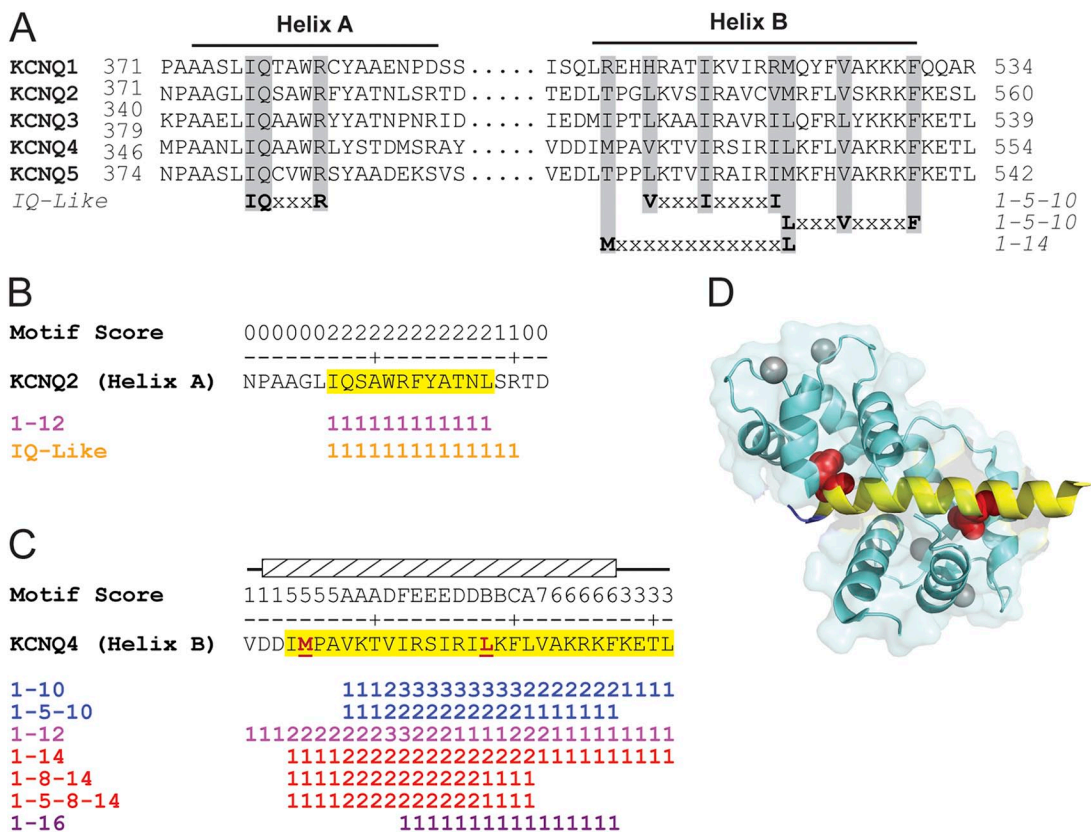

Figure 1. Meta-analysis of KCNQ CaM-binding regions identifies multiple canonical motifs that map to the crystallized structure. (A) Sequence alignment of helices A and B. Gray shading demarcates key residues that define the previously identified CaM-binding motifs. "Wiggle plot" depiction of all of the canonical CaM-binding motifs within (B) KCNQ2 helix A and (C) KCNQ4 helix B. Yellow highlighted residues denote the meta-analysis CaM-binding domain predictions. Hatched box and red underlined residues denote target peptide and anchor residues shown in D. Motif score equals the number of times an amino acid is found in a unique canonical CaM-binding motif. (D) Ribbon diagram of CaM–KCNQ4 helix B (PDB accession no. 4GOW) with the meta-analysis prediction (yellow) mapped onto the target peptide (blue). Red space fill, anchor residues; cyan with gray calcium ions, CaM.

corresponded with the crystallized IQ domain (PDB accession no. 3G43; Fig. 2 A). We next examined the N-terminal spatial Ca²⁺-transforming element (NSCaTE) peptide (Dick et al., 2008) containing the noncanonical motif xWxxx(I/L)xxxx (Taiakina et al., 2013). Currently, high resolution structures with this motif have been determined with either the N- or C-lobes of CaM (Liu and Vogel, 2012), but not with full-length CaM. Although there are canonical CaM motifs within this region of Ca$_V$1.2, they do not substantially overlap with the noncanonical motif that CaM is bound to in the crystal structures (Fig. 2 C, left). Meta-analysis on the NSCaTE target peptide did not predict a CaM-binding domain in the structure (PDB accession no. 2LQC; Fig. 2 A, right). Lastly, we ran the meta-analysis on the calcium-binding protein (CaBP)1-binding domain in Ca$_V$1.2 (Fig. 2 C, right) (Zhou et al., 2005). CaBPs 1–5 are homologous to CaM (Haeseleer et al., 2000) but have been shown to differentially regulate voltage-gated calcium channels and transient receptor potential channels (Lee et al., 2002; Haeseleer et al., 2004; Kinoshita-Kawada et al., 2005; Zhu, 2005; Oz et al., 2011). Interestingly, the N-terminal half of the CaBP1-binding domain is devoid of canonical CaM-binding motifs, whereas the other half contains multiple motifs, resulting in the prediction of a CaM-binding domain in the C-terminal end of the CaBP1-binding domain (Fig. 2 C, right).

We also tested how our meta-analysis fared with target peptides from the voltage-gated sodium channel, Na$_V$1.5. Examination of the IQ domain of the C-terminal region of Na$_V$1.5 channels identified multiple canonical CaM-binding motifs (Fig. 3 A). Similar to KCNQ4 channels, the meta-analysis predicted a larger site for CaM binding than was determined by the NMR structure (Protein Data Base accession no. 4DCK; Fig. 3 B). More recently, Sarhan et al.

(2012) crystallized the structure of CaM bound to the DIII–IV inactivation gate of Na$_V$1.5 channels, which binds using an unorthodox tyrosine anchor. The meta-analysis predicted two CaM-binding domains in the DIII–IV linker (Fig. 3 C); however, these domains do not substantially overlap with residues interacting with CaM in the crystal structure (PDB accession no. 4DJC; Fig. 3 D). In addition, the noncanonical motif, phenylalanine–isoleucine–phenylalanine (Potet et al., 2009), is also missed by the meta-analysis. Given that the meta-analysis relies on canonical CaM-binding motifs, it was not surprising that these noncanonical CaM-binding sites were not identified.

Lastly, we ran the meta-analysis on peptides derived from a membrane transporter: the PMCA isoform 4 (PMCA4). PMCA4 has a splice site that is located in the middle of a CaM-binding domain, which results in two different targets (Strehler, 1991; Strehler et al., 1991). Although CaM does not bind to a canonical motif in either target peptide, our meta-analysis mapped well to the NMR-determined CaM-binding regions of both the C20 and C28 peptides (PDB accession nos. 1CFF and 2KNE; Fig. 4 A). A comparison of the canonical motifs in each peptide (Fig. 4 B) reveals that the C28 peptide contains several 1–14 motifs using the tryptophan anchor at the number 1 position that are absent in the shorter C20 peptide.

These test cases highlight the predictive power and limitations of our canonical motif-clustering meta-analysis. To determine how our simple meta-analysis compares to an algorithm that relies on the biophysical properties of the target peptide, we ran 53 PDB CaM target sequences with one CaM-binding site through the Calmodulin Target Database (Yap et al., 2000). For small target peptides (<100 residues), our meta-analysis (67%) fared better than the Calmodulin Target Database (<50%) at

TABLE 2
Canonical CaM-binding motifs in helix B of KCNQ channels

Motif	KCNQ1		KCNQ2		KCNQ3		KCNQ4		KCNQ5	
	Residues	Sequence	Residues	Sequence	Residues	Sequence	Residues	Sequence	Residues	Sequence
1–10	514–523	IKVIRRMQYF	536–545	LKVSIRAVCV	515–524	LKAAIRAVRI	530–539	VKTVIRSIRI	518–527	LKTVIRAIRI
			540–549	IRAVCVMRFL	525–534	LQFRLYKKKF	533–542	VIRSIRILKF	521–530	VIRAIRIMKF
							534–543	IRSIRILKFL		
							540–549	LKFLVAKRKF		
1–5–10			536–545	LKVSIRAVCV	515–524	LKAAIRAVRI	530–539	VKTVIRSIRI	518–527	LKTVIRAIRI
					525–534	LQFRLYKKKF	533–542	VIRSIRILKF	521–530	VIRAIRIMKF
							540–549	LKFLVAKRKF		
1–12	506–517	LREHHRATIKVI	532–543	LTPGLKVSIRAV			526–537	IMPAVKTVIRSI	514–525	LTPPLKTVIRAI
			538–549	VSIRAVCVMRFL			533–544	VIRSIRILKFLV	521–532	VIRAIRIMKFHV
1–14	516–529	VIRRMQYFVAKKKF	532–545	LTPGLKVSIRAVCV	512–525	IPTLKAAIRAVRIL	526–539	IMPAVKTVIRSIRI	514–527	LTPPLKTVIRAIRI
			536–549	LKVSIRAVCVMRFL			530–543	VKTVIRSIRILKFL		
1–8–14	516–529	VIRRMQYFVAKKKF	536–549	LKVSIRAVCVMRFL	512–525	IPTLKAAIRAVRIL	526–539	IMPAVKTVIRSIRI	514–527	LTPPLKTVIRAIRI
							530–543	VKTVIRSIRILKFL		
1–5–8–14			536–549	LKVSIRAVCVMRFL			526–539	IMPAVKTVIRSIRI	514–527	LTPPLKTVIRAIRI
							530–543	VKTVIRSIRILKFL		
1–16	514–529	IKVIRRMQYFVAKKKF	540–555	IRAVCVMRFLVSKRKF	512–527	IPTLKAAIRAVRILQF	534–549	IRSIRILKFLVAKRKF	522–537	IRAIRIMKFHVAKRKF
					519–534	IRAVRILQFRLYKKKF				

CaM motifs are described in Table 1.

predicting a CaM-binding domain in these structures (Table S2). In contrast, our meta-analysis overestimates the number of CaM-binding domains for targets >100 residues compared with the algorithm used by the Calmodulin Target Database. The imperfection of both computational methods suggests that neither canonical motif clustering nor the biochemical properties of the target are sufficient to flawlessly predict the molecular complexity of CaM target recognition and binding.

Because delineating every canonical CaM-binding motif in the cytoplasmic domains of ion channels and membrane transporters has more utility than computational predictions of CaM-binding sites, we created the Calmodulation Database and Meta-Analysis website: http://cam.umassmed.edu (Mruk et al., 2014). The website allows users to find canonical CaM-binding motifs within an inputted sequence and uses the meta-analysis to predict potential CaM-binding domains within that sequence. In addition to searching a given protein sequence,

the database can be searched to find PDB files that contain a specific canonical CaM-binding motif in the target peptide, which may or may not be anchored by the N- and C-lobes of CaM. Additionally, PDB files that contain target peptides derived from any protein, species, or type of membrane transport protein can be retrieved and subsequently analyzed. The worldwide availability of this simple meta-analysis and database will enable ion channel and membrane transport researchers to readily identify the canonical CaM-binding motifs within their protein of interest, assist in the design of CaM target peptides, and compare CaM–peptide structures to their protein(s) of interest.

DISCUSSION

Given the increasing interest in calmodulation of ion channels and transporters, there existed a need for a simple tool that quickly identifies the potential CaM-binding
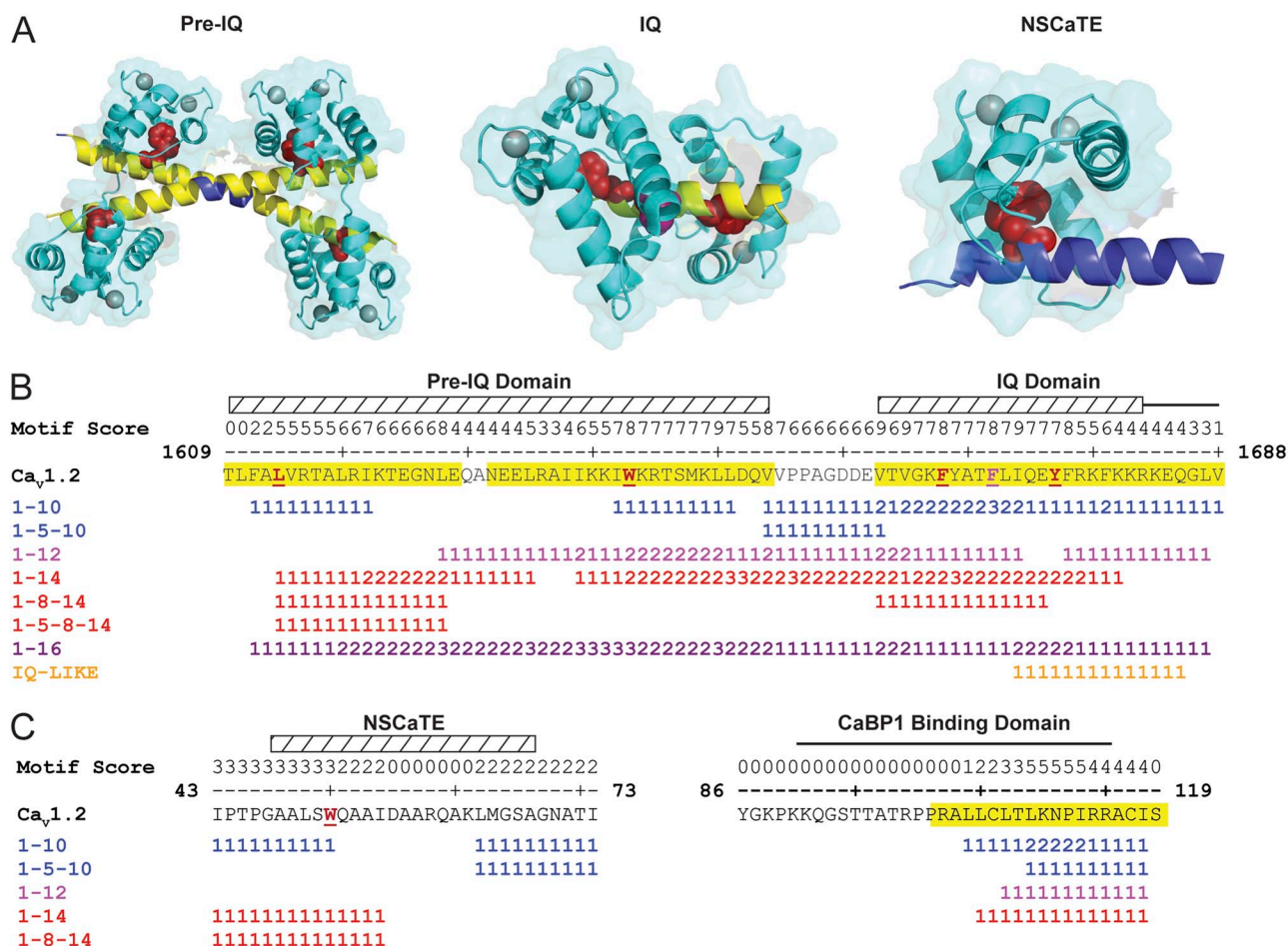

Figure 2. Meta-analyses of cytoplasmic Ca$_V$1.2 CaM-binding domains. (A) Ribbon diagrams of CaM bound to the pre-IQ and IQ domains (PDB accession no. 3G43) and the N-terminal lobe of CaM bound to the NSCaTE peptide (PDB accession no. 2LQC is shown). Predicted residues (yellow) are shown on a blue target peptide. Red space fill, terminal anchor residues; magenta space fill, internal hydrophobic residues; cyan with gray calcium ions, CaM. "Wiggle plot" of the (B) C-terminal pre-IQ and IQ domains and (C) N-terminal NSCaTE- and CaBP1-binding domains. Yellow highlighted residues denote the meta-analysis of CaM-binding domain predictions. Hatched box and red underlined residues denote target peptide and anchor residues identified in the structures shown in A. Motif score equals the number of times an amino acid is found in a unique canonical CaM-binding motif.

regions of these membrane proteins. Therefore, we wrote a computational script to identify every canonical CaM-binding motif within a given protein sequence. Using this script on target sequences from high resolution CaM–peptide structures, we found that CaM often binds to peptide sequences containing multiple overlapping canonical motifs. Combining our motif identification script with a simple charge discriminator yielded a useful tool that can predict CaM-binding regions with both specificity and sensitivity.

To determine the strengths and weaknesses of this simple meta-analysis, we mapped our CaM-binding domain predictions onto a panel of high resolution CaM–peptide structures. For structures in which CaM binds to canonical motifs, such as the IQ domains from Ca$_V$1.2 and Na$_V$1.5, our meta-analysis correctly predicts the region to which CaM binds. In contrast, the binding sites in noncanonical target peptide–CaM structures (Ca$_V$1.2 NSCaTE peptide and Na$_V$1.5 DIII–DIV linker) are missed by the meta-analysis. In fact, the multiple flanking canonical motifs in the Na$_V$1.5 DIII–DIV linker induce our meta-analysis to predict two potential binding domains; however, neither of these regions forms protein–protein interactions with CaM in the crystal structure (Fig. 3 D). Although these noncanonical binding domains are missed by our meta-analysis, CaM-binding domains were correctly predicted for both the C20 and C28 peptides of PMCA4,

which bind to CaM in a noncanonical fashion (Fig. 4). In the shorter C20 peptide structure, only one CaM lobe makes contact with the peptide. Closer examination of the sequence shows that the anchoring tryptophan has the potential to act as the first anchor for canonical motifs (1–5–10, 1–5–8–14), but the shorter C20 peptide lacks the requisite terminal anchors (Fig. 4 B). Indeed, these motifs are completed in the longer C28 peptide, which in combination with a phenylalanine in position 18, anchors both lobes of CaM, wrapping the C28 peptide in an antiparallel manner.

Several proteomic computational methods have been developed to identify CaM-binding proteins and the location of the CaM-binding sites on these proteins (Yap et al., 2000; Radivojac et al., 2006; Hamilton, M., A.S.N. Reddy, and A. Ben-Hur. 2011. Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine; Minhas and Ben-Hur, 2012; Wang et al., 2012). The majority of these prediction methods use structural information to predict whether specific residues may belong to a protein–protein interface. Despite the increasing number of CaM–peptide structures, predicting the exact CaM-binding site on any one protein remains challenging, in part, because CaM's interactions with its targets can be dynamic and not well represented by a single structure. Therefore, prediction methods dependent on sequence information alone
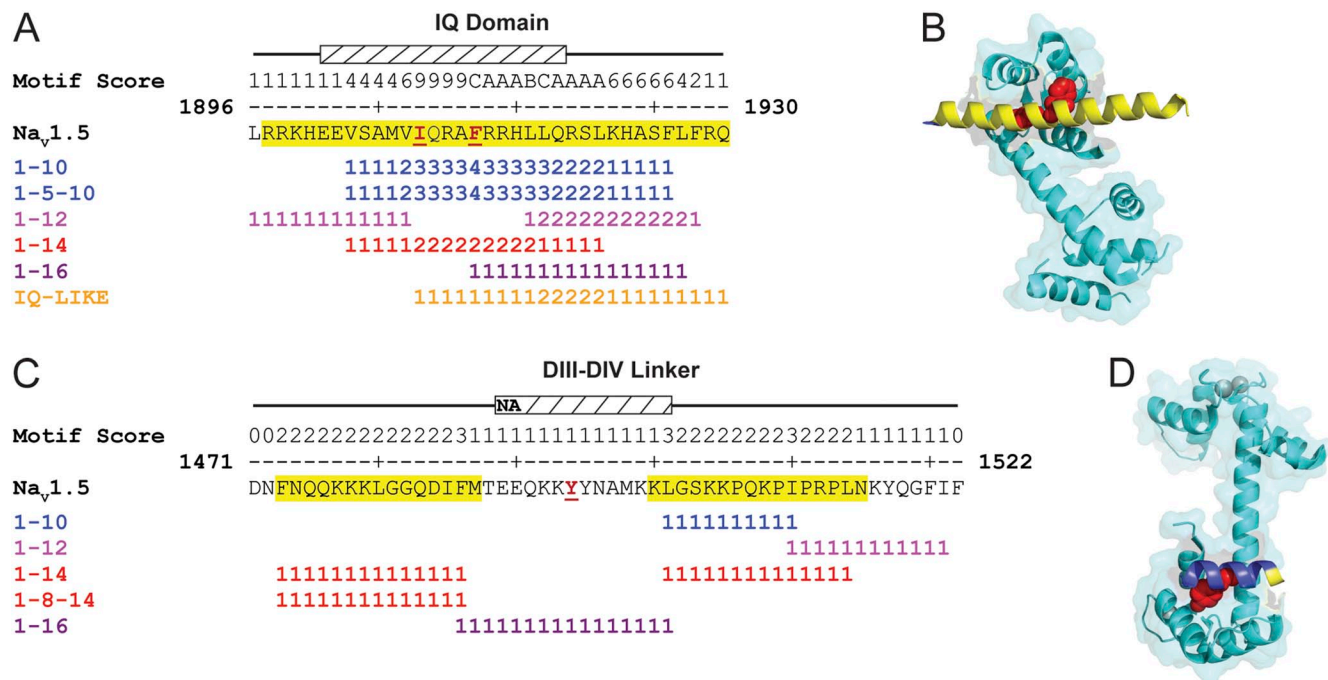

Figure 3. Multiple Na$_V$1.5 CaM-binding regions are predicted by the meta-analysis. Meta-analysis of the Na$_V$1.5 (A) IQ domain and (C) DIII–IV linker. Yellow highlighted residues denote the meta-analysis CaM-binding domain predictions. Hatched box and red underlined residues denote target peptide and anchor residues shown in B and D; NA in the hatched box represents the peptide sequence (NAQKKYYNAMK) that was determined in D. Motif score equals the number of times an amino acid is found in a unique canonical CaM-binding motif. Ribbon diagrams of CaM bound to the Na$_V$1.5 (B) IQ domain (Protein Data Base accession no. 4DCK) and (D) DIII–IV linker (PDB accession no. 4DJC). Predicted residues (yellow) are shown on a blue target peptide. Red space fill, anchor residues; cyan with gray calcium ions, CaM.

could be useful in identifying CaM-binding sites. However, simple sequence alignments of potential partners to known CaM-binding sites afford poor sensitivity (TP ~40%) (Minhas and Ben-Hur, 2012). Accordingly, most protein–protein predictive algorithms that rely on sequence information alone are based on neural networks (Ofran and Rost, 2003; Minhas and Ben-Hur, 2012) or hidden Markov models (Friedrich et al., 2006). Because we optimized the parameters of the meta-analysis on high resolution CaM target peptide structures in the PDB, this limited dataset is not compatible with these computational methods. In spite of this limitation, the meta-analysis TP/TN values are respectable when compared with the Calmodulin Target Database and a published neural
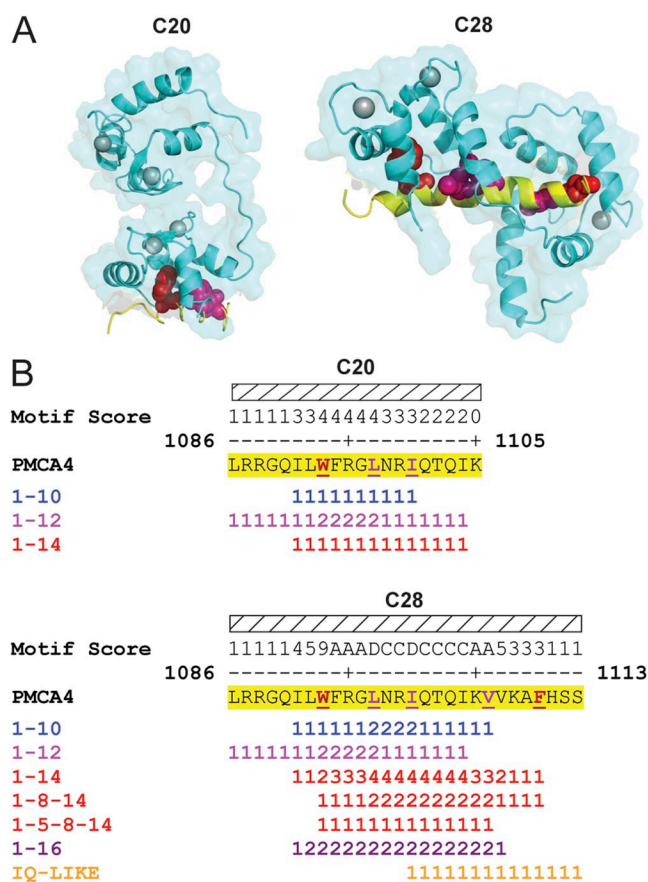

Figure 4. Meta-analysis predicts binding to PMCA4 peptides that bind CaM via noncanonical motifs. (A) Ribbon diagrams of CaM bound to C20 (PDB accession no. 1CFF) and C28 (PDB accession no. 2KNE) peptides. Predicted residues are shown in yellow. Red space fill, terminal anchor residues; magenta space fill, internal hydrophobic residues; cyan with gray calcium ions, CaM. (B) Meta-analysis and wiggle plot of the C20 and C28 peptides. Yellow highlighted residues denote the meta-analysis CaM-binding domain predictions. Hatched box and red underlined residues denote target peptide and terminal anchor residues shown in A; pink underlined residues denote the internal residues that become anchored when both lobes of CaM wrap around the C28 peptide. Motif score equals the number of times an amino acid is found in a unique canonical CaM-binding motif.

network algorithm for predicting CaM-binding regions (Yap et al., 2000; Minhas and Ben-Hur, 2012).

For those interested in the calmodulation of ion channels and membrane transporters, the significant advantage of the web-based meta-analysis is the ability to efficiently visualize all of the canonical motifs within a given sequence. This palatable presentation of every canonical CaM-binding motif highlights the complexity of CaM binding and may explain the discrepancies observed between KCNQ–CaM structural and biochemical studies. For example, our meta-analysis predicts a larger CaM-binding domain that contains several potential anchor residues in the distal C-terminal end of helix B that do not form contacts with CaM in the KCNQ4 crystal structure (Fig. 1, C and D). This prediction, however, is consistent with biochemical and mutational data that demonstrate that this region is important for CaM association with KCNQ channels (Ghosh et al., 2006; Gómez-Posada et al., 2011). Because crystal structures are static snapshots, it is possible that CaM adopts more than one conformation when bound to functioning KCNQ channels or shuttles between the multiple binding motifs within helix B on a single KCNQ C terminus.

Although most attention has been paid to the calmodulation of ion transport proteins, additional CaBPs also regulate ion channel function. Similar to CaM, the CaBP family (CaBPs 1–8) contains paired EF hands that undergo structural rearrangements upon Ca^{2+} binding. CaBP1 has been shown to compete for binding to the IQ- and to the N-terminal domains of $Ca_V1.2$ channels (Lee et al., 2002). Consistent with this finding, our meta-analysis picks up a potential CaM-binding region within the CaBP1-binding domain (Fig. 2 C, right), lending credence to the hypothesis that the CaBPs may recognize their targets similarly to the way CaM binds to its targets.

Our meta-analysis suggests that CaM binds to regions rich with canonical motifs. Using the canonical motif finder, structurally defined CaM-binding domains can be redesigned to determine whether the overlapping CaM-binding sites are necessary for the binding and calmodulation of ion transport proteins. Together with the meta-analysis, these tools make it simpler to find canonical motifs, identify potential CaM-binding regions, and design future biochemical and structural experiments, thereby accelerating our understanding of calmodulation of ion channels and membrane transport proteins.

REFERENCES

Alaimo, A., J.C. Gómez-Posada, P. Aivar, A. Etxeberría, J.A. Rodriguez-Alfaro, P. Areso, and A. Villarroel. 2009. Calmodulin activation limits the rate of KCNQ2 K+ channel exit from the endoplasmic reticulum. *J. Biol. Chem.* 284:20668–20675. http://dx.doi.org/10.1074/jbc.M109.019539

Ataman, Z.A., L. Gakhar, B.R. Sorensen, J.W. Hell, and M.A. Shea. 2007. The NMDA receptor NR1 C1 region bound to calmodulin: Structural insights into functional differences between homologous domains. *Structure.* 15:1603–1617. http://dx.doi.org/10.1016/j.str.2007.10.012

Barbato, G., M. Ikura, L.E. Kay, R.W. Pastor, and A. Bax. 1992. Backbone dynamics of calmodulin studied by nitrogen-15 relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. *Biochemistry.* 31:5269–5278. http://dx.doi.org/10.1021/bi00138a005

Ben-Johny, M., and D.T. Yue. 2014. Calmodulin regulation (calmodulation) of voltage-gated calcium channels. *J. Gen. Physiol.* 143:679–692. http://dx.doi.org/10.1085/jgp.201311153

Biswas, S., D.A. DiSilvestre, P. Dong, and G.F. Tomaselli. 2013. Mechanisms of a human skeletal myotonia produced by mutation in the C-terminus of Na$_V$1.4: Is Ca^{2+} regulation defective? *PLoS ONE.* 8:e81063. http://dx.doi.org/10.1371/journal.pone.0081063

Dick, I.E., M.R. Tadross, H. Liang, L.H. Tay, W. Yang, and D.T. Yue. 2008. A modular switch for spatial Ca^{2+} selectivity in the calmodulin regulation of Ca$_V$ channels. *Nature.* 451:830–834. http://dx.doi.org/10.1038/nature06529

Etxeberria, A., P. Aivar, J.A. Rodriguez-Alfaro, A. Alaimo, P. Villacé, J.C. Gómez-Posada, P. Areso, and A. Villarroel. 2008. Calmodulin regulates the trafficking of KCNQ2 potassium channels. *FASEB J.* 22:1135–1143. http://dx.doi.org/10.1096/fj.07-9712com

Friedrich, T., B. Pils, T. Dandekar, J. Schultz, and T. Müller. 2006. Modelling interaction sites in protein domains with interaction profile hidden Markov models. *Bioinformatics.* 22:2851–2857. http://dx.doi.org/10.1093/bioinformatics/btl486

Ghosh, S., D.A. Nunziato, and G.S. Pitt. 2006. KCNQ1 assembly and function is blocked by long-QT syndrome mutations that disrupt interaction with calmodulin. *Circ. Res.* 98:1048–1054. http://dx.doi.org/10.1161/01.RES.0000218863.44140.f2

Gómez-Posada, J.C., P. Aivar, A. Alberdi, A. Alaimo, A. Etxeberría, J. Fernández-Orth, T. Zamalloa, M. Roura-Ferrer, P. Villace, P. Areso, et al. 2011. Kv7 channels can function without constitutive calmodulin tethering. *PLoS ONE.* 6:e25508. http://dx.doi.org/10.1371/journal.pone.0025508

Haeseleer, F., I. Sokal, C.L. Verlinde, H. Erdjument-Bromage, P. Tempst, A.N. Pronin, J.L. Benovic, R.N. Fariss, and K. Palczewski. 2000. Five members of a novel Ca^{2+}-binding protein (CABP) subfamily with similarity to calmodulin. *J. Biol. Chem.* 275:1247–1260. http://dx.doi.org/10.1074/jbc.275.2.1247

Haeseleer, F., Y. Imanishi, T. Maeda, D.E. Possin, A. Maeda, A. Lee, F. Rieke, and K. Palczewski. 2004. Essential role of Ca^{2+}-binding protein 4, a Ca$_V$1.4 channel regulator, in photoreceptor synaptic function. *Nat. Neurosci.* 7:1079–1087. http://dx.doi.org/10.1038/nn1320

Hino, A., M. Yano, T. Kato, M. Fukuda, T. Suetomi, M. Ono, W. Murakami, T. Susa, S. Okuda, M. Doi, et al. 2012. Enhanced binding of calmodulin to the ryanodine receptor corrects contractile dysfunction in failing hearts. *Cardiovasc. Res.* 96:433–443. http://dx.doi.org/10.1093/cvr/cvs271

Ikura, M., S. Spera, G. Barbato, L.E. Kay, M. Krinks, and A. Bax. 1991. Secondary structure and side-chain proton and carbon-13 resonance assignments of calmodulin in solution by heteronuclear multidimensional NMR spectroscopy. *Biochemistry.* 30:9216–9228. http://dx.doi.org/10.1021/bi00102a013

Juranic, N., E. Atanasova, A.G. Filoteo, S. Macura, F.G. Prendergast, J.T. Penniston, and E.E. Strehler. 2010. Calmodulin wraps around its binding domain in the plasma membrane Ca^{2+} pump anchored by a novel 18-1 motif. *J. Biol. Chem.* 285:4015–4024. http://dx.doi.org/10.1074/jbc.M109.060491

Kinoshita-Kawada, M., J. Tang, R. Xiao, S. Kaneko, J.K. Foskett, and M.X. Zhu. 2005. Inhibition of TRPC5 channels by Ca^{2+}-binding protein 1 in *Xenopus* oocytes. *Pflugers Arch.* 450:345–354. http://dx.doi.org/10.1007/s00424-005-1419-1

Kumar, V., V.P. Chichili, L. Zhong, X. Tang, A. Velazquez-Campoy, F.S. Sheu, J. Seetharaman, N.Z. Gerges, and J. Sivaraman. 2013. Structural basis for the interaction of unstructured neuron specific substrates neuromodulin and neurogranin with calmodulin. *Sci Rep.* 3:1392. http://dx.doi.org/10.1038/srep01392

Kyte, J., and R.F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157:105–132. http://dx.doi.org/10.1016/0022-2836(82)90515-0

LaPorte, D.C., B.M. Wierman, and D.R. Storm. 1980. Calcium-induced exposure of a hydrophobic surface on calmodulin. *Biochemistry.* 19:3814–3819. http://dx.doi.org/10.1021/bi00557a025

Lee, A., R.E. Westenbroek, F. Haeseleer, K. Palczewski, T. Scheuer, and W.A. Catterall. 2002. Differential modulation of Ca$_V$2.1 channels by calmodulin and Ca^{2+}-binding protein 1. *Nat. Neurosci.* 5:210–217. http://dx.doi.org/10.1038/nn805

Liu, Z., and H.J. Vogel. 2012. Structural basis for the regulation of L-type voltage-gated calcium channels: interactions between the N-terminal cytoplasmic domain and Ca^{2+}-calmodulin. *Front Mol Neurosci.* 5:38. http://dx.doi.org/10.3389/fnmol.2012.00038

Maximciuc, A.A., J.A. Putkey, Y. Shamoo, and K.R. Mackenzie. 2006. Complex of calmodulin with a ryanodine receptor target reveals a novel, flexible binding mode. *Structure.* 14:1547–1556. http://dx.doi.org/10.1016/j.str.2006.08.011

Minhas, F., and A. Ben-Hur. 2012. Multiple instance learning of Calmodulin binding sites. *Bioinformatics.* 28:i416–i422. http://dx.doi.org/10.1093/bioinformatics/bts416

Mruk, K., S.M. Shandilya, R.O. Blaustein, C.A. Schiffer, and W.R. Kobertz. 2012. Structural insights into neuronal K+ channel-calmodulin complexes. *Proc. Natl. Acad. Sci. USA.* 109:13579–13583. http://dx.doi.org/10.1073/pnas.1207606109

Mruk, K., B. Farley, W.R. Kobertz, and A. Ritacco. 2014. Calmodulation database and meta-analysis predictor. University of Massachusetts Medical School. http://cam.umassmed.edu/ (accessed May 27, 2014).

Ofran, Y., and B. Rost. 2003. Predicted protein-protein interaction sites from local sequence information. *FEBS Lett.* 544:236–239. http://dx.doi.org/10.1016/S0014-5793(03)00456-3

Oz, S., V. Tsemakhovich, C.J. Christel, A. Lee, and N. Dascal. 2011. CaBP1 regulates voltage-dependent inactivation and activation of Ca$_V$1.2 (L-type) calcium channels. *J. Biol. Chem.* 286:13945–13953. http://dx.doi.org/10.1074/jbc.M110.198424

Potet, F., B. Chagot, M. Anghelescu, P.C. Viswanathan, S.Z. Stepanovic, S. Kupershmidt, W.J. Chazin, and J.R. Balser. 2009. Functional interactions between distinct sodium channel cytoplasmic domains through the action of calmodulin. *J. Biol. Chem.* 284:8846–8854. http://dx.doi.org/10.1074/jbc.M806871200

Radivojac, P., S. Vucetic, T.R. O'Connor, V.N. Uversky, Z. Obradovic, and A.K. Dunker. 2006. Calmodulin signaling: Analysis and prediction of a disorder-dependent molecular recognition. *Proteins.* 63:398–410. http://dx.doi.org/10.1002/prot.20873

Rhoads, A.R., and F. Friedberg. 1997. Sequence motifs for calmodulin recognition. *FASEB J.* 11:331–340.

Sarhan, M.F., C.C. Tung, F. Van Petegem, and C.A. Ahern. 2012. Crystallographic basis for calcium regulation of sodium channels. *Proc. Natl. Acad. Sci. USA.* 109:3558–3563. http://dx.doi.org/10.1073/pnas.1114748109

Seaton, B.A., J.F. Head, D.M. Engelman, and F.M. Richards. 1985. Calcium-induced increase in the radius of gyration and maximum

dimension of calmodulin measured by small-angle x-ray scattering. *Biochemistry*. 24:6740–6743. http://dx.doi.org/10.1021/bi00345a002

Shamgar, L., L. Ma, N. Schmitt, Y. Haitin, A. Peretz, R. Wiener, J. Hirsch, O. Pongs, and B. Attali. 2006. Calmodulin is essential for cardiac I_{KS} channel gating and assembly: Impaired function in long-QT mutations. *Circ. Res.* 98:1055–1063. http://dx.doi.org/10.1161/01.RES.0000218979.40770.69

Strehler, E.E. 1991. Recent advances in the molecular characterization of plasma membrane Ca^{2+} pumps. *J. Membr. Biol.* 120:1–15. http://dx.doi.org/10.1007/BF01868586

Strehler, E.E., R. Heim, and E. Carafoli. 1991. Molecular characterization of plasma membrane calcium pump isoforms. *Adv. Exp. Med. Biol.* 307:251–261. http://dx.doi.org/10.1007/978-1-4684-5985-2_23

Taiakina, V., A.N. Boone, J. Fux, A. Senatore, D. Weber-Adrian, J.G. Guillemette, and J.D. Spafford. 2013. The calmodulin-binding, short linear motif, NSCaTE is conserved in L-type channel ancestors of vertebrate Cav1.2 and Cav1.3 channels. *PLoS ONE.* 8:e61765. http://dx.doi.org/10.1371/journal.pone.0061765

Wang, L., Z.P. Liu, X.S. Zhang, and L. Chen. 2012. Prediction of hot spots in protein interfaces using a random forest model with hybrid features. *Protein Eng. Des. Sel.* 25:119–126. http://dx.doi.org/10.1093/protein/gzr066

Weiss, L.A., A. Escayg, J.A. Kearney, M. Trudeau, B.T. MacDonald, M. Mori, J. Reichert, J.D. Buxbaum, and M.H. Meisler. 2003. Sodium channels SCN1A, SCN2A and SCN3A in familial autism. *Mol. Psychiatry.* 8:186–194. http://dx.doi.org/10.1038/sj.mp.4001241

Xu, Q., A. Chang, A. Tolia, and D.L. Minor Jr. 2013. Structure of a Ca^{2+}/CaM:Kv7.4 (KCNQ4) B-helix complex provides insight into M current modulation. *J. Mol. Biol.* 425:378–394. http://dx.doi.org/10.1016/j.jmb.2012.11.023

Yap, K.L., J. Kim, K. Truong, M. Sherman, T. Yuan, and M. Ikura. 2000. Calmodulin target database. *J. Struct. Funct. Genomics.* 1:8–14. http://dx.doi.org/10.1023/A:1011320027914

Yus-Najera, E., I. Santana-Castro, and A. Villarroel. 2002. The identification and characterization of a noncontinuous calmodulin-binding site in noninactivating voltage-dependent KCNQ potassium channels. *J. Biol. Chem.* 277:28545–28553. http://dx.doi.org/10.1074/jbc.M204130200

Zhou, H., K. Yu, K.L. McCoy, and A. Lee. 2005. Molecular mechanism for divergent regulation of $Ca_v1.2$ Ca^{2+} channels by calmodulin and Ca^{2+}-binding protein-1. *J. Biol. Chem.* 280:29612–29619. http://dx.doi.org/10.1074/jbc.M504167200

Zhu, M.X. 2005. Multiple roles of calmodulin and other Ca^{2+}-binding proteins in the functional regulation of TRP channels. *Pflugers Arch.* 451:105–115. http://dx.doi.org/10.1007/s00424-005-1427-1