**A peer-reviewed version of this preprint was published in PeerJ on 22 September 2015.**

View the peer-reviewed version (peerj.com/articles/1265), which is the preferred citable publication unless you specifically need to cite this preprint.

Permyakov SE, Permyakov EA, Uversky VN. 2015. Intrinsically disordered caldesmon binds calmodulin via the “buttons on a string” mechanism. PeerJ 3:e1265 https://doi.org/10.7717/peerj.1265

# Does intrinsically disordered caldesmon bind calmodulin via the “buttons on a string” mechanism?

Sergei E Permyakov, Eugene A Permyakov, Vladimir N Uversky

We show here that chicken gizzard caldesmon (CaD) and its C-terminal domain (residues 636-771, $CaD_{136}$) are intrinsically disordered proteins. The computational and experimental analyses of the wild type $CaD_{136}$ and series of its single tryptophan mutants (W674A, W707A, and W737A) and a double tryptophan mutant (W674A/W707A) suggested that although the interaction of $CaD_{136}$ with calmodulin (CaM) can be driven by the non-specific electrostatic attraction between these oppositely charged molecules, the specificity of $CaD_{136}$-CaM binding is likely to be determined by the specific packing of important $CaD_{136}$ tryptophan residues at the $CaD_{136}$-CaM interface. It is suggested that this interaction can be described as the “buttons on a charged string” model, where the electrostatic attraction between the intrinsically disordered $CaD_{136}$ and the CaM is solidified in a “snapping buttons” manner by specific packing of the $CaD_{136}$ “pliable buttons” (which are the short segments of fluctuating local structure condensed around the tryptophan residues) at the $CaD_{136}$-CaM interface. Our data also show that all three “buttons” are important for binding, since mutation of any of the tryptophans affects $CaD_{136}$-CaM binding and since $CaD_{136}$ remains CaM-buttoned even when two of the three tryptophans are mutated to alanines.

# Does intrinsically disordered caldesmon bind calmodulin via the "buttons on a string" mechanism?

*Sergei E. Permyakov,* [†] *Eugene A. Permyakov,* [†] *and Vladimir N. Uversky*[†,‡,*]

[†]Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia;

[‡]Department of Molecular Medicine and USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, Florida 33612, USA;

*To whom correspondence should be addressed: Vladimir N. Uversky, Department of Molecular Medicine, College of Medicine, University of South Florida, 12901 Bruce B. Downs Blvd, MDC3540, Tampa, FL 33612, USA; E-mail: vuversky@health.usf.edu



PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1261v1 | CC-BY 4.0 Open Access | rec: 25 Jul 2015, publ: 25 Jul 2015

**ABSTRACT** We show here that chicken gizzard caldesmon (CaD) and its C-terminal domain (residues 636-771, $CaD_{136}$) are intrinsically disordered proteins. The computational and experimental analyses of the wild type $CaD_{136}$ and series of its single tryptophan mutants (W674A, W707A, and W737A) and a double tryptophan mutant (W674A/W707A) suggested that although the interaction of $CaD_{136}$ with calmodulin (CaM) can be driven by the non-specific electrostatic attraction between these oppositely charged molecules, the specificity of $CaD_{136}$-CaM binding is likely to be determined by the specific packing of important $CaD_{136}$ tryptophan residues at the $CaD_{136}$-CaM interface. It is suggested that this interaction can be described as the "buttons on a charged string" model, where the electrostatic attraction between the intrinsically disordered $CaD_{136}$ and the CaM is solidified in a "snapping buttons" manner by specific packing of the $CaD_{136}$ "pliable buttons" (which are the short segments of fluctuating local structure condensed around the tryptophan residues) at the $CaD_{136}$-CaM interface. Our data also show that all three "buttons" are important for binding, since mutation of any of the tryptophans affects $CaD_{136}$-CaM binding and since $CaD_{136}$ remains CaM-buttoned even when two of the three tryptophans are mutated to alanines.

**KEYWORDS**: Intrinsically disordered protein; caldesmon; calmodulin, protein-protein interaction; MoRF.

**ABBREVIATIONS**

AIBS, disorder-based ANCHOR-identified binding site; CaD, caldesmon; $CaD_{136}$, C-terminal domain (636-771) of CaD; CaM, calmodulin; CD, circular dichroism; DSC, differential scanning calorimetry; IDP, intrinsically disordered protein; IDPR, intrinsically disordered protein region; MoRF, molecular recognition feature; PTM, posttranslational modification; UV, ultra violet

## INTRODUCTION

Caldesmon, CaD, is a ubiquitous actin-binding protein of ~770 residues with the molecular mass of 88.75 kDa and *pI* of 5.56, which is almost exclusively localized within the contractile domain of the smooth muscle cells (Mabuchi et al. 1996). CaD is involved in the regulation of smooth muscle contraction, non-muscle motility, and cytoskeleton formation (Czurylo & Kulikova 2012; Gusev 2001; Marston & Redwood 1991; Martson & Huber 1996; Matsumura & Yamashiro 1993; Sobue & Sellers 1991). Particularly, CaD plays a role in a thin-filament-linked regulation of smooth muscle contraction through specific binding to F-actin and F-actin-tropomyosin leading to the inhibition of the actin-stimulated myosin ATPase (Marston & Redwood 1991). The inhibitory action of CaD is reversed by interaction of this protein with various calcium-dependent proteins, such as calmodulin (CaM) and caltropin (Mani & Kay 1996). The functional activity of CaD is further regulated by phosphorylation at multiple sites (Shirinsky et al. 1999). CaD is also engaged in the so-called "flip-flop" interactions, where, depending on the calcium concentration and the availability of $Ca^{2+}$-binding proteins, CaD is alternatively bound either to F-actin or to CaM (Adelstein & Eisenberg 1980; Gusev 2001). These thin filament-based modulatory effects provide additional "fine-tuning" to the well-established, myosin light chain phosphorylation-dependent, thick filament-based regulation of smooth muscle contraction (Adelstein & Eisenberg 1980). CaD is found to form tight complexes with several proteins, such as myosin, actin, tropomyosin, CaM (Marston & Redwood 1991), caltropin (Gusev 2001; Mani & Kay 1996), calcyclin (Kuznicki & Filipek 1987), $S100a_o$, S100a and S100b proteins (Polyakov et al. 1998), and non-muscle tropomyosin (Gusev 2001). It also possesses distinctive phospholipid-binding properties (Bogatcheva & Gusev 1996; Czurylo et al.

1993a; Czurylo et al. 1993b; Makowski et al. 1997; Vorotnikov et al. 1992; Vorotnikov & Gusev 1990a; Vorotnikov & Gusev 1990b).

Sequence of CaD can be divided to four independent functional domains. The first N-terminal domain interacts with myosin and tropomyosin. The second domain is characteristic for smooth muscle CaD and also participates in the tropomyosin binding. The third domain is involved in the CaD interaction of with myosin, tropomyosin, and actin. The fourth C-terminal domain plays the most important role in the function of CaD, interacting with actin, various $Ca^{2+}$-binding proteins, myosin, tropomyosin, and phospholipids (Gusev 2001). Furthermore, interaction of CaD with actin, tropomyosin, and CaM involves multiple sites (Fraser et al. 1997; Gusev 2001; Huber et al. 1996; Medvedeva et al. 1997; Wang et al. 1997), with CaD being wrapped around its partners (Gusev 2001; Permyakov et al. 2003).

CaD exists as two isoforms that are generated by alternative splicing of a single mRNA transcript. These CaD isoforms are differently distributed among tissues (Abrams et al. 2012; Kordowska et al. 2006). The light (or low molecular weight) isoform (l-CaD) is expressed in most cell types, including at low levels in smooth muscle, where it mediates actin and non-muscle myosin interaction in the cortical cytoskeleton (Helfman et al. 1999). The heavy (or high molecular weight) isoform (h-CaD) is expressed specifically in smooth muscle. It is believed that this isoform is capable of simultaneous binding to smooth muscle actin and myosin filaments due to the presence of a peptide spacer domain in the middle of the protein (Wang et al. 1991).

Based on these functional peculiarities (the ability to interact with multiple binding partners, the presence of numerous sites of posttranslational modifications, the capability to be engaged in wrapping interactions, and the presence of multiple alternatively spliced isoforms) one could

conclude that CaD belongs to the realm of the intrinsically disordered proteins (IDPs), which were recognized quite recently (Dunker et al. 2001; Dunker et al. 2008a; Dunker et al. 2008b; Dyson & Wright 2005; Tompa 2002; Uversky 2002a; Uversky 2002b; Uversky 2010; Uversky & Dunker 2010; Uversky et al. 2000; Wright & Dyson 1999) as important biologically active proteins without unique 3D-structures that represent a crucial extension of the protein kingdom (Dunker et al. 2008a; Dyson 2011; Tompa 2012; Turoverov et al. 2010; Uversky 2002a; Uversky 2003; Uversky 2013a; Wright & Dyson 1999). IDPs and hybrid proteins containing both ordered and intrinsically disordered domains/regions (Dunker et al. 2013) are very common in nature (Dunker et al. 2000; Tokuriki et al. 2009; Uversky 2010; Ward et al. 2004; Xue et al. 2012a; Xue et al. 2010b). They constitute significant fractions of all known proteomes, where the overall amount of disorder in proteins increases from bacteria to archaea to eukaryota, and over a half of the eukaryotic proteins are predicted to possess long IDP regions (IDPRs) (Dunker et al. 2000; Oldfield et al. 2005a; Uversky 2010; Ward et al. 2004; Xue et al. 2012b). Due to the lack of unique 3D-structures, IDPs/IDPRs carry out numerous crucial biological functions (such as signaling, regulation, and recognition) (Daughdrill et al. 2005; Dunker et al. 2002a; Dunker et al. 2002b; Dunker et al. 2005; Dunker et al. 1998; Dunker et al. 2001; Dyson & Wright 2005; Tompa 2002; Tompa 2005; Tompa & Csermely 2004; Tompa et al. 2005; Uversky 2002a; Uversky 2002b; Uversky 2003; Uversky 2010; Uversky et al. 2000; Uversky et al. 2005; Vucetic et al. 2007; Wright & Dyson 1999; Xie et al. 2007a; Xie et al. 2007b) that complement functions of ordered proteins.(Vucetic et al. 2007; Xie et al. 2007a; Xie et al. 2007b) Furthermore, many IDPs/IDPRs are associated with the variety of human diseases (Uversky et al. 2014; Uversky et al. 2008).

In our previous study, we showed that the C-terminal domain of chicken gizzard CaD, $CaD_{136}$ (636-771 fragment), is a typical extended IDP characterized by the almost complete lack of secondary structure, absence of a globular core, and a large hydrodynamic volume (Permyakov et al. 2003). Although $CaD_{136}$ can effectively bind to the $Ca^{2+}$-loaded CaM, this protein was shown to remain mostly unfolded within its complex with CaM (Permyakov et al. 2003). In this paper, we first performed comprehensive computational characterization of chicken gizzard CaD to confirm the overall disorder status of this protein. Then, we found that the $CaD_{136}$ has three major disorder-based potential binding sites located around the tryptophan residues W674, W707, and W737. To verify the role of these sites in $CaD_{136}$ binding to CaM, we designed and characterized biophysically three single tryptophan mutants (W674A, W707A, and W737A) and a double tryptophan mutant (W674A/W707A). This analysis suggests that $CaD_{136}$ potentially binds CaM via the "buttons on a charged string" mechanism. Some biological significance of these observations is discussed.

## MATERIALS AND METHODS

### Materials

Samples of CaM, $CaD_{136}$, its single tryptophan mutants (W674A, W707A, and W737A), and a double tryptophan mutant (W674A/W707A) were a kind gift of Dr. Yuji Kobajashi (Department of Physical Chemistry, Institute of Protein Research, Osaka University, Osaka 565, Japan).

All chemicals were of analytical grade from Fisher Chemicals. Concentrations of CaD and CaM were estimated spectrophotometrically. Molar extinction coefficient for CaM was calculated based upon amino acids content according to (Pace et al. 1995): $\varepsilon_{280nm}$=2,980 $M^{-1}cm^{-1}$. For the wild type CaD $\varepsilon_{280nm}$=17,990 $M^{-1}cm^{-1}$ was used, whereas molar extinction coefficients

for single and double tryptophan mutants were taken to be $\varepsilon_{280nm}$=12,490 $M^{-1}cm^{-1}$ and $\varepsilon_{280nm}$=6,990 $M^{-1}cm^{-1}$, respectively.

**Methods**

*Absorption Spectroscopy.* Absorption spectra were measured on a spectrophotometer designed and manufactured in the Institute for Biological Instrumentation (Pushchino, Russia).

*Circular Dichroism Measurements.* Circular dichroism measurements were carried out by means of a AVIV 60DS spectropolarimeter (Lakewood, N. J., USA), using cells with a path length of 0.1 and 10.0 mm for far and near UV CD measurements, respectively. Protein concentration was kept at 0.6-0.8 mg/ml throughout all the experiments.

*Fluorescence Measurements.* Fluorescence measurements were carried out on a lab-made spectrofluorimeter main characteristics of which were described earlier (Permyakov et al. 1977). All spectra were corrected for spectral sensitivity of the instrument and fitted to log-normal curves (Burstein & Emelyanenko 1996) using nonlinear regression analysis (Marquardt 1963). The maximum positions of the spectra were obtained from the fits. The temperature inside the cell was monitored with a copper-constantan thermopile.

*Parameters of CaD136 Binding to CaM.* The apparent binding constants for complexes of calmodulin with the caldesmon mutants were evaluated from a fit of the fluorescence titration data to the specific binding scheme using nonlinear regression analysis (Marquardt 1963). The binding scheme was chosen on the "simplest best fit" basis. The quality of the fit was judged by

149 a randomness of distribution of residuals. Temperature dependence of intrinsic fluorescence was
150 analyzed according to (Permyakov & Burstein 1984).

151

152 *Differential Scanning Microcalorimetry.* Scanning microcalorimetric measurements were
153 carried out on a DASM-4M differential scanning microcalorimeter (Institute for Biological
154 Instrumentation of the Russian Academy of Sciences, Pushchino, Russia) in 0.48 mL cells at a 1
155 K/min heating rate. An extra pressure of 1.5 atm was maintained in order to prevent possible
156 degassing of the solutions on heating. Protein concentrations were in the 0.5 to 0.7 mg/mL range.
157 The heat sorption curves were baseline corrected by heating the measurement cells filled by the
158 solvent only. Specific heat capacities of the proteins were calculated according to (Privalov
159 1979; Privalov & Potekhin 1986).

160

161 *Sequence Analyses.* Amino acid sequences of human and chicken caldesmons (UniProt IDs:
162 P12957 and Q05682, respectively) and human and chicken calmodulins (UniProt IDs: P62149
163 and P62158, respectively) were retrieved from UniProt (http://www.uniprot.org/).

164 The intrinsic disorder propensities of query proteins were evaluated by several per-residues
165 disorder predictors, such as PONDR® VLXT (Dunker et al. 2001), PONDR® VSL2 (Peng et al.
166 2005), PONDR® VL3 (Peng et al. 2006b), and PONDR® FIT (Xue et al. 2010a). Here, scores
167 above 0.5 are considered to correspond to the disordered residues/regions. PONDR® VSL2B is
168 one of the more accurate stand-alone disorder predictors (Fan & Kurgan 2014; Peng et al. 2005;
169 Peng & Kurgan 2012), PONDR® VLXT is known to have high sensitivity to local sequence

peculiarities and can be used for identifying disorder-based interaction sites (Dunker et al. 2001), whereas a metapredictor PONDR-FIT is moderately more accurate than each of the component predictors, PONDR® VLXT (Dunker et al. 2001), PONDR® VSL2 (Peng et al. 2005), PONDR® VL3 (Peng et al. 2006b), FoldIndex (Prilusky et al. 2005), IUPred (Dosztanyi et al. 2005a), and TopIDP (Campen et al. 2008). Disorder propensities of CaD and CaM were further analyzed using the MobiDB database (http://mobidb.bio.unipd.it/) (Di Domenico et al. 2012; Potenza et al. 2015) that generates consensus disorder scores based on the outputs of ten disorder predictors, such as ESpritz in its two flavors (Walsh et al. 2012), IUPred in its two flavors (Dosztanyi et al. 2005a), DisEMBL in two of its flavors (Linding et al. 2003a), GlobPlot (Linding et al. 2003b), PONDR® VSL2B (Obradovic et al. 2005; Peng et al. 2006a), and JRONN (Yang et al. 2005).

For human CaM and CaD proteins, disorder evaluations together with the important disorder-related functional annotations were retrieved from $D^2P^2$ database (http://d2p2.pro/) (Oates et al. 2013). $D^2P^2$ is a database of predicted disorder that represents a community resource for pre-computed disorder predictions on a large library of proteins from completely sequenced genomes (Oates et al. 2013). $D^2P^2$ database uses outputs of PONDR® VLXT (Dunker et al. 2001), IUPred (Dosztanyi et al. 2005a), PONDR® VSL2B (Obradovic et al. 2005; Peng et al. 2006a), PrDOS (Ishida & Kinoshita 2007), ESpritz (Walsh et al. 2012), and PV2 (Oates et al. 2013). This database is further enhanced by information on the curated sites of various posttranslational modifications and on the location of predicted disorder-based potential binding sites.

Interactability of chicken CaD and CaM was evaluated by STRING (Search Tool for the Retrieval of Interacting Genes, http://string-db.org/), which is the online database resource, that provides both experimental and predicted interaction information (Szklarczyk et al. 2011).

193 STRING produces the network of predicted associations for a particular group of proteins. The
194 network nodes are proteins, whereas the edges represent the predicted or known functional
195 associations. An edge may be drawn with up to 7 differently colored lines that represent the
196 existence of the seven types of evidence used in predicting the associations. A red line indicates
197 the presence of fusion evidence; a green line - neighborhood evidence; a blue line – co-
198 occurrence evidence; a purple line - experimental evidence; a yellow line – text mining evidence;
199 a light blue line - database evidence; a black line – co-expression evidence (Szklarczyk et al.
200 2011).

201 Potential disorder-based binding sites in $CaD_{136}$ (which is the C-terminal domain (636-771) of
202 CaD) were found using three computational tools, α-MoRF identifier (Cheng et al. 2007;
203 Oldfield et al. 2005b), ANCHOR (Dosztanyi et al. 2009; Meszaros et al. 2009), and MoRFpred
204 (Disfani et al. 2012). Since IDPs/IDPRs are commonly involved in protein-protein interactions
205 (Daughdrill et al. 2005; Dunker et al. 2002a; Dunker et al. 2002b; Dunker et al. 2001; Dunker et
206 al. 2008b; Dunker & Uversky 2008; Oldfield et al. 2005b; Radivojac et al. 2007; Tompa 2002;
207 Uversky 2011b; Uversky 2012; Uversky 2013b; Uversky & Dunker 2010; Uversky et al. 2005),
208 and since they are able to undergo at least partial disorder-to-order transitions upon binding,
209 which is crucial for recognition, regulation, and signaling (Dunker et al. 2001; Dyson & Wright
210 2002; Dyson & Wright 2005; Mohan et al. 2006; Oldfield et al. 2005b; Uversky 2013b; Uversky
211 2013c; Uversky et al. 2000; Vacic et al. 2007a; Wright & Dyson 1999), these proteins and
212 regions often contain functionally important, short, order-prone motifs within the long disordered
213 regions. Such motifs are known as Molecular Recognition Feature (MoRF), they are able to
214 undergo disorder-to-order transition during the binding to a specific partner, and can be
215 identified computationally (Cheng et al. 2007; Oldfield et al. 2005b). For example, an α-MoRF

predictor indicates the presence of a relatively short, loosely structured region within a largely disordered sequence (Oldfield et al. 2005b), which can gain functionality upon a disorder-to-order transition induced by binding to partners (Mohan et al. 2006; Vacic et al. 2007a). In addition to MoRF identifiers, potential binding sites in disordered regions can be identified by the ANCHOR algorithm (Dosztanyi et al. 2009; Meszaros et al. 2009). This approach relies on the pairwise energy estimation approach developed for the general disorder prediction method IUPred (Dosztanyi et al. 2005a; Dosztanyi et al. 2005b). being based on the hypothesis that long regions of disorder contain localized potential binding sites that cannot form enough favorable intrachain interactions to fold on their own, but are likely to gain stabilizing energy by interacting with a globular protein partner (Dosztanyi et al. 2009; Meszaros et al. 2009). Regions of a protein suggested by the ANCHOR algorithm to have significant potential to be binding sites are the ANCHOR-indicated binding site (AIBS).

## RESULTS AND DISCUSSION

### Characterization of Functional Disorder in Caldesmon and Calmodulin

The amino acid sequences and compositions of IDPs/IDPRs are significantly different from those of ordered proteins and domains. For example, the amino acid compositions of extended IDPs/IDPRs (i.e., highly disordered proteins and regions lacking almost any residual structure (Dunker et al. 2001; Uversky 2002a; Uversky 2002b; Uversky 2003; Uversky 2013a; Uversky 2013c; Uversky & Dunker 2010; Uversky et al. 2000)) are characterized by high mean net charge and low mean hydropathy, being significantly depleted in order-promoting residues C, W, Y, F, H, I, L, V, and N and significantly enriched in disorder-promoting residues A, R, G, Q, S, P, E, and K (Dunker et al. 2001; Radivojac et al. 2007; Romero et al. 2001; Vacic et al. 2007b).

The fractional difference in composition between CaD and a set of ordered proteins from PDB Select 25 (Berman et al. 2000) was calculated as $(C_{CaD}-C_{order})/C_{order}$, where $C_{CaD}$ is the content of a given amino acid in CaD, and $C_{order}$ is the corresponding value for the set of ordered proteins. This analysis revealed that in comparison with typical ordered proteins, CaD is significantly depleted in major order-promoting residues (C, Y, F, H, V, L, and I) and is significantly enriched in major disorder-promoting residues, such as A, R, E, and K. This means that CaD might contain multiple structural and functional signatures typical for the IDPs.

In agreement with this conclusion, Figure 1A represents the results of the disorder predisposition analysis in CaD by a family of PONDR disorder predictors, PONDR® VLXT (Dunker et al. 2001), PONDR® VSL2 (Peng et al. 2005), PONDR® VL3 (Peng et al. 2006b), and PONDR® FIT (Xue et al. 2010a). Since the absolute majority of residues is predicted to have disorder scores above 0.5 and since the mean disorder score for the full-length protein ranges, depending on the predictor, from 0.69 to 0.93, this analysis clearly shows that CaD is expected to be mostly disordered. In agreement with this conclusion, the consensus MobiDB analysis (http://mobidb.bio.unipd.it/entries/P12957) revealed that chicken gizzard CaD contains 98.4% disordered residues. Curiously, the C-terminal domain of this protein, $CaD_{136}$, is predicted to be a bit more predisposed for order than the remaining protein (depending on the predictor, the mean disorder score for this 636-771 fragment of CaD ranges from 0.52 to 0.81). This observation is further illustrated by Figure 1B which represents the PONDR-based disorder profiles of this region.

Curiously, although several X-ray crystal (PDB IDs: 1ahr, 1up5, 2bcx, 2bki, 2o5g, 2o60, 2vb6, 3gog, and 3gp2) and NMR solution structures (PDB IDs: 2kz2 and 2m3s) of CaM are known, Figure 1C shows that this protein is predicted to be rather disordered too. These findings are not

too surprising, since it is known that the CaM structure and folding are strongly dependent on the metal ion binding (Li et al. 2014; Sulmann et al. 2014), and that there is a great variability in the crystal structures of CaM in isolation (i.e., where it is not bound to its protein or peptide partners and exists in the unliganded form) which is considered as an illustration of CaM plasticity in solution (Kursula 2014). Furthermore, several studies on the structure of unliganded CaM in solution using small angle scattering and other methods have indicated the presence of a mixture of conformations (Bertini et al. 2010; Heller 2005; Kursula 2014; Yamada et al. 2012). Also in agreement with these predictions, the analysis of one of the NMR structures of CaM (PDB ID: 2m3s) revealed that this protein might contain up to 50.3% of disordered residues in solution (Moroz et al. 2013). Again, the results of the per-residue predictions by the members of the PONDR family are further supported by the results of the MobiDB analysis, according to which the consensus disorder content of CaM based on the outputs of ten disorder predictors is 18.1%. The corresponding values evaluated by the individual predictors (http://mobidb.bio.unipd.it/entries/P62149) are ranging from 6.0% and 13.4% for the ESpritz-XRay and DisEMBL-465, respectively to 41.6% and 69.1% for the IUPred-long and PONDR® VSL2, respectively. Note that both ESpritz-XRay and DisEMBL-465 are trained based on proteins with known crystal structures and containing regions of missing electron density, whereas IUPred-long and PONDR VSL2 use different criteria for training.

Further information on the functional disorder status of CaD and CaM was retrieved from $D^2P^2$ portal, which represents a database of pre-computed disorder predictions for a large library of proteins from completely sequenced genomes (Oates et al. 2013), which in addition to outputs of nine disorder predictors provides information on the curated sites of various posttranslational modifications and on the location of predicted disorder-based potential binding sites. Since this

database does not include data for chicken, the human homologues of CaD and CaM were used for this analysis. The validity of this approach is justified by the fact that sequences of human and chicken CaMs are identical (100% identity), whereas sequences of human and chicken CaD are highly conserved (61% identity).

Figures 2A and 3A represents the results of this analysis of CaD and CaM, respectively, and provide further support for the abundance and functional importance of intrinsic disorder in these proteins, which are predicted to contain long disordered regions enriched in potential disorder-based binding motifs and containing numerous sites of posttranslational modifications, PTMs. The fact that disordered domains/regions of the human CaD and CaM contain numerous PTM sites is in agreement with the well-known notion that phosphorylation(Iakoucheva et al. 2004) and many other enzymatically catalyzed PTMs are preferentially located within the IDPRs (Pejaver et al. 2014).

The interactivity of chicken CaD and CaM was evaluated by the online database resource, STRING, which provides information on both experimental and predicted interactions (Szklarczyk et al. 2011). Figure 2B and 3B clearly show that both proteins are predicted to have numerous binding partners. Predicted here high levels of connectivity and binding promiscuity indicate that, in the related protein-protein interaction networks (PPI), chicken CaD and CaM serve as hub proteins connecting biological modules to each other. The binding promiscuity of hub proteins is believed to be dependent on intrinsic disorder (Dosztanyi et al. 2006; Ekman et al. 2006; Haynes et al. 2006; Patil & Nakamura 2006; Singh et al. 2006; Uversky et al. 2005), where disorder and related disorder-to-order transitions enable one protein to interact with multiple partners (one-to-many signaling) or enable multiple partners to bind to one protein (many-to-one signaling) (Dunker et al. 1998). In line with these considerations, intrinsically

disordered nature of chicken CaD and CaM provides a plausible explanation for their potential roles as hub proteins. Therefore, data reported in Figures 1, 2 and 3 suggest that both CaD and CaM are expected to contain substantial amounts of functional disorder, which CaD being predicted to be mostly disordered.

Figure 1D shows that the positively charged R and K residues are evenly distributed within the $CaD_{136}$ sequence and that the sequence of CaM contains evenly spread negatively charged residues D and E. Since under the physiologic conditions of neutral pH, the C-terminal interacting domain of CaD and CaM possess charges of opposite sign (+9 for $CaD_{136}$ and -24 for CaM) it is likely that electrostatic interactions play important role in interaction between these two proteins. This hypothesis is further supported by Figure 4, which represents the charge distribution over the CaM surface and shows that negative charges are almost evenly distributed over the entire protein surface. What then defines the specificity of interaction between a highly positively charged IDP ($CaD_{136}$) and a highly negatively charged surface of CaM? Some answers to this important question can be obtained analyzing peculiarities of the disorder distribution in $CaD_{136}$. In fact, many IDPs/IDPRs involved in protein-protein interactions and molecular recognitions are able to undergo at least partial disorder-to-order transitions upon binding (Daughdrill et al. 2005; Dunker et al. 2002a; Dunker et al. 2002b; Dunker et al. 2001; Dunker et al. 2008b; Dunker & Uversky 2008; Dyson & Wright 2002; Dyson & Wright 2005; Mohan et al. 2006; Oldfield et al. 2005b; Radivojac et al. 2007; Tompa 2002; Uversky 2011b; Uversky 2012; Uversky 2013b; Uversky 2013c; Uversky & Dunker 2010; Uversky et al. 2000; Uversky et al. 2005; Vacic et al. 2007a; Wright & Dyson 1999). Such potential disorder-based binding sites are known as Molecular Recognition Feature (MoRF), and they often can be found based on the peculiar shape of a disorder profile (sharp "dips" within the long IDPRs). These observations

serve as a foundation for the corresponding computational tools, e.g., α-MoRF-Pred (Cheng et al. 2007; Oldfield et al. 2005b) or MoRFpred (Disfani et al. 2012). Alternatively, the disorder-based binding sites can be identified by ANCHOR (Dosztanyi et al. 2009; Meszaros et al. 2009) (see Materials and Methods). There is generally a good agreement between the results of binding sites prediction by these two tools.

These analyses revealed that $CaD_{136}$ has several disorder-based potential binding sites and three of them correspond to the major minima in the $CaD_{136}$ disorder plots obtained by both PONDR® VLXT and PONDR-FIT (see Figure 5). Since each of these three dip-centered potential binding sites include a tryptophan residue, we decided to mutate those tryptophans in order to evaluate their roles in the $CaD_{136}$ binding to CaM. At the first stage, the disorder propensities of three single tryptophan mutants (W674A, W707A, and W737A) and a double tryptophan mutant (W674A/W707A) were compared using PONDR® VLXT and PONDR FIT algorithms. Figure 5 represents the results of these analyses and shows that the local disorder propensities were noticeably affected by single mutations W674A and W707A and by the W674A/W707A double mutation, whereas W737A had a very minimal effect on the $CaD_{136}$ disorder profile. Although the depth of corresponding disorder minima was affected by mutations, none of these tryptophan-to-alanine substitutions completely eliminated dips. These data suggested that binding affinity of $CaD_{136}$ can be moderately affected by single substitutions W674A and W707A, and that the W674A/W707A double mutation could have somewhat stronger effect on protein-protein interactions. To check these predictions, we analyzed biophysical properties and binding affinities of three single tryptophan mutants W674A, W707A, and W737A, and a double tryptophan mutant W674A/W707A. Results of these analyses are represented below.

**Effect of tryptophan substitutions on tryptophan fluorescence spectrum of the C-terminal CaD domain**

Figure 6 shows normalized tryptophan fluorescence spectra of $CD_{136}$ and its mutants in solution and in complex with CaM (which does not have tryptophan residues). It is clearly seen that the spectra of all the $CD_{136}$ proteins in solutions are practically the same. They have extremely long wavelength positions and are similar to spectrum of a free tryptophan in water, which shows that in all these proteins, the tryptophan residues are totally exposed to water. The spectra of the complexes with CaM are different. The CaM-complexes W737A mutant has the most blue-shifted spectrum, whereas the W707A mutant in its bound state has the least blue-shifted spectrum. The Table 1 represents the relative fluorescence quantum yields for $CD_{136}$ and its mutants in solution and in the complex with CaM.

**Effect of tryptophan substitutions on far-UV CD spectra of $CaD_{136}$ mutants**

Figure 7 represents the far-UV CD spectra of wild type, W674A, W707A, W737A and W674A/W707A $CaD_{136}$ and shows that all these proteins have far-UV CD spectra typical of the almost completely unfolded polypeptides. In other words, the data are consistent with the conclusion that at physiological conditions none of the $CaD_{136}$ domains has considerable amount of ordered secondary structure; i.e., they belong to the family of so-called natively unfolded proteins, which are the most disordered members of the realm of intrinsically disordered proteins. On the other hand, more detailed analysis of the far-UV CD spectrum shows that the wild type $CaD_{136}$, being mostly disordered, is still far from to be completely unfolded and preserves some residual structure (e.g., $[\theta]_{222} \sim -3{,}000$ deg cm$^2$ dmol$^{-1}$, the minimum is located at 200, rather than at 196-198 nm, see Figure 7A).

Figure 7A shows that all amino acid substitutions affect the far-UV CD spectrum of the C-terminal CaD domain in a similar manner, inducing considerable decrease in the spectrum intensity around 200 nm. Figure 7B represents the difference spectra between the wild type $CaD_{136}$ and muted domains and clearly shows that all the amino acid substitutions induce noticeable additional unfolding of the residual structure in the originally rather disordered protein.

**Effect of tryptophan substitutions on the near-UV CD spectra of $CaD_{136}$ mutants**

Surprisingly, Figure 8 shows that wild type $CaD_{136}$ and all its mutants possess rather intensive and pronounced near-UV CD spectra. This means that tryptophan residues of these proteins are in relatively asymmetric environment. Figure 8 shows that any tryptophan substitution analyzed in this study has a considerable effect on the near-UV CD spectrum of $CaD_{136}$, leading to the substantial decrease in the spectral intensity. It also can be seen that different tryptophan residues have different contributions to the near-UV CD spectrum of protein. In fact, Figure 8A shows that the effect of amino acid substitutions increases in the following order: W707A < W737A < W674A ≤ W674A/W707A. This conclusion is confirmed by the difference spectra shown in (Figure 8B). Therefore, these data suggest that tryptophan residues have noticeable contributions to the residual structure of $CaD_{136}$, likely serving as condensation centers around which the local dynamic structure is formed.

**Conformational stability of $CaD_{136}$ and its mutants analyzed by the effect of temperature on their near- and far-UV CD spectra**

Figure 9 represents near-UV CD spectra of the wild type and mutated $CaD_{136}$ measured at different temperatures. It can be seen that heating affects the near-UV CD spectra of different proteins in different manner. In the case of the wild type protein, some initial decrease in the spectral intensity at 40°C is followed by the increase in spectral intensity at higher temperatures. Interestingly, after the cooling, the near-UV CD spectrum of this variant is somewhat more intensive than spectrum measured before the heating. Spectrum of W674A mutant increases with the temperature and this effect is reversible. Mutants W707A and W737A show reversible decrease in spectral intensity, whereas spectrum of the double W674A/W707A mutant is practically unaffected by temperature. Importantly, Figure 9 shows that even at 90°C all of the protein variants analyzed in this study show pronounced near-UV CD spectra, reflecting the fact that the temperature increase does not destroy completely the asymmetric environment of their aromatic residues.

Temperature had similar effect of the far-UV CD spectra of all the $CaD_{136}$ variants. As an example, Figure 10A represents the far-UV CD spectra of W674A mutant measured at different temperatures. It can be seen that shape and intensity of the spectrum undergo considerable changes with the increase in temperature, reflecting the temperature-induced formation of the more ordered secondary structure. Same spectral changes were observed for several other IDPs and were classified as the "turn-out" response of extended IDPs to changes in their environment (Uversky 2002a; Uversky 2002b; Uversky 2011a; Uversky 2013a; Uversky 2013c; Uversky & Dunker 2010). Figure 10B summarizes the data on the effect of heating on the secondary structure of the $CaD_{136}$ variants as corresponding $[\theta]_{222}$ *vs*. temperature dependences. One can

see that in all cases studied temperature increase was accompanied by the steady increase in the negative ellipticity at 222 nm. It is necessary to emphasize here that this behavior is totally different from the conformational behavior of typical globular proteins, which show temperature-induced reduction in the content of ordered secondary structure.

**Studying the $CaD_{136}$ variants by scanning microcalorimetry**

Figure 11 represents the calorimetric scans obtained for the wild type $CaD_{136}$ and its mutants. The absolute values of the specific heat capacity (ranging from ~2 to 3 J/(g·K)) and the absence of distinct heat absorption peaks within the temperature region from 10 to 100ºC for these proteins suggest that their structure is predominantly unfolded.

**Interactions of the $CaD_{136}$ and its tryptophan mutants with calmodulin studied by intrinsic fluorescence**

Figure 12 represents the results of the spectrofluorimetric titration of $CD_{136}$ and its tryptophan mutants with CaM. The increase in CaM concentration induces an increase in fluorescence quantum yield and a blue shift of the fluorescence spectrum maximum (see also data presented in Figure 6 and Table 1). The points shown in this figure are experimental data, and the curves are theoretical fits. The corresponding curves were computed using the simplest one-site binding scheme by fitting the experimental points varying the binding constant. The values of the binding constants which give the best fits are collected in the Table 1. This analysis revealed that the substitution of the tryptophan residues by alanines resulted in a decrease in the $CaD_{136}$-CaM binding constant in all the cases except W737A, where mutation caused an increase in the

$CaD_{136}$ affinity for CaM. Table 1 also shows that the double W674A/W707A mutation caused the largest reduction in the $CaD_{136}$ binding efficiency.

## CONCLUSIONS

Altogether, data presented in our study suggest that CaD and its C-terminal domain, $CaD_{136}$, are intrinsically disordered proteins. CaD potentially serves as a disordered hub in several important protein-protein interaction networks. It is likely that $CaD_{136}$-CaM interaction is driven by the non-specific electrostatic attraction interactions due to the opposite charges of these two proteins. Specificity of $CaD_{136}$-CaM binding is likely to be determined by the definite packing of important tryptophan residues at the $CaD_{136}$-CaM interface, which is manifested by the dramatic blue shift of the intrinsic $CaD_{136}$ fluorescence. In its non-bound form, $CaD_{136}$ is highly disordered, with the aforementioned tryptophan residues potentially serving as centers of local fluctuating structural elements. Therefore, our bioinformatics and experimental data suggest that the interaction between $CaD_{136}$ and CaM can be described within the "buttons on a charged string" model, where the electrostatic attraction between the positively charged and highly disordered $CaD_{136}$ containing at least three segments of fluctuating local structure ("pliable buttons") and the negatively charged CaM is solidified by the specific packing of three short regions containing tryptophan residues in a "snapping a button" manner. This model is schematically represented in Figure 13. Curiously, it seems that all three "buttons" are important for binding, since mutation of any of the tryptophans affects $CaD_{136}$-CaM binding and since $CaD_{136}$ remains CaM-buttoned even when two of the three tryptophans are mutated to alanines.

464 **ACKNOWLEDGEMENTS**

465 This work was supported by grants from the Programs of the Russian Academy of Sciences

466 «Molecular and Cellular Biology» (P.E.A.) and «Fundamental Science for Medicine» (P.S.E.).

467

## REFERENCES


Abrams J, Davuluri G, Seiler C, and Pack M. 2012. Smooth muscle caldesmon modulates peristalsis in the wild type and non-innervated zebrafish intestine. *Neurogastroenterol Motil* 24:288-299.

Adelstein RS, and Eisenberg E. 1980. Regulation and kinetics of the actin-myosin-ATP interaction. *Annu Rev Biochem* 49:921-956.

Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, and Eddy SR. 2004. The Pfam protein families database. *Nucleic Acids Res* 32:D138-141.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, and Bourne PE. 2000. The Protein Data Bank. *Nucleic Acids Res* 28:235-242.

Bertini I, Giachetti A, Luchinat C, Parigi G, Petoukhov MV, Pierattelli R, Ravera E, and Svergun DI. 2010. Conformational Space of Flexible Biological Macromolecules from Average Data. *Journal of the American Chemical Society* 132:13553-13558.

Bogatcheva NV, and Gusev NB. 1996. Interaction of caldesmon and calponin with phospholipids. *Journal of Muscle Research and Cell Motility* 17:146-146.

Burstein EA, and Emelyanenko VI. 1996. Log-normal description of fluorescence spectra of organic fluorophores. *Photochem Photobiol* 64:316-320.

Campen A, Williams RM, Brown CJ, Meng J, Uversky VN, and Dunker AK. 2008. TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein Pept Lett* 15:956-963.

Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, and Dunker AK. 2007. Mining alpha-helix-forming molecular recognition features with cross species sequence alignments. *Biochemistry* 46:13468-13477.

Czurylo EA, and Kulikova N. 2012. *Anatromy and Physiology of Proteins: Caldesmon*. New York: Nova Science Publishers.

Czurylo EA, Zborowski J, and Dabrowska R. 1993a. Interaction of Caldesmon with Phospholipids. *Journal of Muscle Research and Cell Motility* 14:240-240.

Czurylo EA, Zborowski J, and Dabrowska R. 1993b. Interaction of Caldesmon with Phospholipids. *Biochemical Journal* 291:403-408.

Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS, and Dunker AK. 2005. Natively disordered proteins. In: Buchner J, and Kiefhaber T, eds. *Handbook of Protein Folding*. Weinheim, Germany: Wiley-VCH, Verlag GmbH & Co., 271-353.

Di Domenico T, Walsh I, Martin AJ, and Tosatto SC. 2012. MobiDB: a comprehensive database of intrinsic protein disorder annotations. *Bioinformatics* 28:2080-2081.

Disfani FM, Hsu WL, Mizianty MJ, Oldfield CJ, Xue B, Dunker AK, Uversky VN, and Kurgan L. 2012. MoRFpred, a computational tool for sequence-based prediction and characterization of short disorder-to-order transitioning binding regions in proteins. *Bioinformatics* 28:i75-83.

Dosztanyi Z, Chen J, Dunker AK, Simon I, and Tompa P. 2006. Disorder and sequence repeats in hub proteins and their implications for network evolution. *J Proteome Res* 5:2985-2995.

Dosztanyi Z, Csizmok V, Tompa P, and Simon I. 2005a. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21:3433-3434.

513 Dosztanyi Z, Csizmok V, Tompa P, and Simon I. 2005b. The pairwise energy content estimated
514 from amino acid composition discriminates between folded and intrinsically unstructured
515 proteins. *J Mol Biol* 347:827-839.
516 Dosztanyi Z, Meszaros B, and Simon I. 2009. ANCHOR: web server for predicting protein
517 binding regions in disordered proteins. *Bioinformatics* 25:2745-2746.
518 Dunker AK, Babu M, Barbar E, Blackledge M, Bondos SE, Dosztányi Z, Dyson HJ, Forman-
519 Kay J, Fuxreiter M, Gsponer J, Han K-H, Jones DT, Longhi S, Metallo SJ, Nishikawa K,
520 Nussinov R, Obradovic Z, Pappu R, Rost B, Selenko P, Subramaniam V, Sussman JL,
521 Tompa P, and Uversky VN. 2013. What's in a name? Why these proteins are intrinsically
522 disordered. *Intrinsically Disordered Proteins* 1:e24157.
523 Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, and Obradovic Z. 2002a. Intrinsic disorder
524 and protein function. *Biochemistry* 41:6573-6582.
525 Dunker AK, Brown CJ, and Obradovic Z. 2002b. Identification and functions of usefully
526 disordered proteins. *Adv Protein Chem* 62:25-49.
527 Dunker AK, Cortese MS, Romero P, Iakoucheva LM, and Uversky VN. 2005. Flexible nets. The
528 roles of intrinsic disorder in protein interaction networks. *Febs J* 272:5129-5148.
529 Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, Obradovic Z, Kissinger C, and
530 Villafranca JE. 1998. Protein disorder and the evolution of molecular recognition: theory,
531 predictions and observations. *Pac Symp Biocomput*:473-484.
532 Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM,
533 Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey
534 RW, Griswold MD, Chiu W, Garner EC, and Obradovic Z. 2001. Intrinsically disordered
535 protein. *J Mol Graph Model* 19:26-59.
536 Dunker AK, Obradovic Z, Romero P, Garner EC, and Brown CJ. 2000. Intrinsic protein disorder
537 in complete genomes. *Genome Inform Ser Workshop Genome Inform* 11:161-171.
538 Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, Vacic V, Obradovic Z, and
539 Uversky VN. 2008a. The unfoldomics decade: an update on intrinsically disordered
540 proteins. *BMC Genomics* 9 Suppl 2:S1.
541 Dunker AK, Silman I, Uversky VN, and Sussman JL. 2008b. Function and structure of
542 inherently disordered proteins. *Curr Opin Struct Biol* 18:756-764.
543 Dunker AK, and Uversky VN. 2008. Signal transduction via unstructured protein conduits. *Nat*
544 *Chem Biol* 4:229-230.
545 Dyson HJ. 2011. Expanding the proteome: disordered and alternatively folded proteins. *Q Rev*
546 *Biophys* 44:467-518.
547 Dyson HJ, and Wright PE. 2002. Coupling of folding and binding for unstructured proteins. *Curr*
548 *Opin Struct Biol* 12:54-60.
549 Dyson HJ, and Wright PE. 2005. Intrinsically unstructured proteins and their functions. *Nat Rev*
550 *Mol Cell Biol* 6:197-208.
551 Ekman D, Light S, Bjorklund AK, and Elofsson A. 2006. What properties characterize the hub
552 proteins of the protein-protein interaction network of Saccharomyces cerevisiae? *Genome*
553 *Biol* 7:R45.
554 Fan X, and Kurgan L. 2014. Accurate prediction of disorder in protein chains with a
555 comprehensive and empirically designed consensus. *J Biomol Struct Dyn* 32:448-464.
556 Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S,
557 Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer EL, and Bateman A. 2006.
558 Pfam: clans, web tools and services. *Nucleic Acids Res* 34:D247-251.

559 Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, Forslund K, Eddy SR,
560 Sonnhammer EL, and Bateman A. 2008. The Pfam protein families database. *Nucleic*
561 *Acids Res* 36:D281-288.
562 Fraser IDC, Copeland O, Wu B, and Marston SB. 1997. The inhibitory complex of smooth
563 muscle caldesmon with actin and tropomyosin involves three interacting segments of the
564 C-terminal domain 4. *Biochemistry* 36:5483-5492.
565 Fukuchi S, Sakamoto S, Nobe Y, Murakami SD, Amemiya T, Hosoda K, Koike R, Hiroaki H,
566 and Ota M. 2012. IDEAL: Intrinsically Disordered proteins with Extensive Annotations
567 and Literature. *Nucleic Acids Res* 40:D507-511.
568 Gusev NB. 2001. Some properties of caldesmon and calponin and the participation of these
569 proteins in regulation of smooth muscle contraction and cytoskeleton formation.
570 *Biochemistry-Moscow* 66:1112-1121.
571 Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, and
572 Iakoucheva LM. 2006. Intrinsic disorder is a common feature of hub proteins from four
573 eukaryotic interactomes. *PLoS Comput Biol* 2:e100.
574 Helfman DM, Levy ET, Berthier C, Shtutman M, Riveline D, Grosheva I, Lachish-Zalait A,
575 Elbaum M, and Bershadsky AD. 1999. Caldesmon inhibits nonmuscle cell contractility
576 and interferes with the formation of focal adhesions. *Mol Biol Cell* 10:3097-3112.
577 Heller WT. 2005. Influence of multiple well defined conformations on small-angle scattering of
578 proteins in solution. *Acta Crystallographica Section D-Biological Crystallography* 61:33-
579 44.
580 Huber PAJ, ElMezgueldi M, Grabarek Z, Slatter DA, Levine BA, and Marston SB. 1996.
581 Multiple-sited interaction of caldesmon with Ca2(+)-calmodulin. *Biochemical Journal*
582 316:413-420.
583 Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, and Dunker AK.
584 2004. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res*
585 32:1037-1049.
586 Ishida T, and Kinoshita K. 2007. PrDOS: prediction of disordered protein regions from amino
587 acid sequence. *Nucleic Acids Res* 35:W460-464.
588 Kordowska J, Huang R, and Wang CL. 2006. Phosphorylation of caldesmon during smooth
589 muscle contraction and cell migration or proliferation. *J Biomed Sci* 13:159-172.
590 Kursula P. 2014. The many structural faces of calmodulin: a multitasking molecular jackknife.
591 *Amino Acids* 46:2295-2304.
592 Kuznicki J, and Filipek A. 1987. Purification and Properties of a Novel Ca2+-Binding Protein
593 (10.5 Kda) from Ehrlich-Ascites-Tumor Cells. *Biochemical Journal* 247:663-667.
594 Li W, Wang W, and Takada S. 2014. Energy landscape views for interplays among folding,
595 binding, and allostery of calmodulin domains. *Proc Natl Acad Sci U S A* 111:10550-
596 10555.
597 Linding R, Jensen LJ, Diella F, Bork P, Gibson TJ, and Russell RB. 2003a. Protein disorder
598 prediction: implications for structural proteomics. *Structure* 11:1453-1459.
599 Linding R, Russell RB, Neduva V, and Gibson TJ. 2003b. GlobPlot: exploring protein sequences
600 for globularity and disorder. *Nucleic Acids Res* 31:3701-3708.
601 Mabuchi K, Li Y, Tao T, and Wang CL. 1996. Immunocytochemical localization of caldesmon
602 and calponin in chicken gizzard smooth muscle. *J Muscle Res Cell Motil* 17:243-260.
603 Makowski P, Makuch R, Sikorski AF, Jezierski A, Pikula S, and Dabrowska R. 1997. Interaction
604 of caldesmon with endoplasmic reticulum membrane: effects on the mobility of

phospholipids in the membrane and on the phosphatidylserine base-exchange reaction. *Biochemical Journal* 328:505-509.
Mani RS, and Kay CM. 1996. Calcium binding proteins. In: Barany M, ed. *Biochemistry of Smooth Muscle Contraction*. New York: Academic Press, 105-116.
Marquardt DW. 1963. An algorithm for least-squares estimation of nonlinear parameters. *J Soc Indust Appl Math* 11:431-441.
Marston SB, and Redwood CS. 1991. The molecular anatomy of caldesmon. *Biochem J* 279 ( Pt 1):1-16.
Martson SB, and Huber PAJ. 1996. Caldesmon. In: Barany M, ed. *Biochemistry of Smooth Muscle Contraction*. New York: Academic Press, 70-90.
Matsumura F, and Yamashiro S. 1993. Caldesmon. *Curr Opin Cell Biol* 5:70-76.
Medvedeva MV, Kolobova EA, Huber PAJ, Fraser IDC, Marston SB, and Gusev NB. 1997. Mapping of contact sites in the caldesmon-calmodulin complex. *Biochemical Journal* 324:255-262.
Meszaros B, Simon I, and Dosztanyi Z. 2009. Prediction of protein binding regions in disordered proteins. *PLoS Comput Biol* 5:e1000376.
Mohan A, Oldfield CJ, Radivojac P, Vacic V, Cortese MS, Dunker AK, and Uversky VN. 2006. Analysis of molecular recognition features (MoRFs). *J Mol Biol* 362:1043-1059.
Moroz OV, Moroz YS, Wu Y, Olsen AB, Cheng H, Mack KL, McLaughlin JM, Raymond EA, Zhezherya K, Roder H, and Korendovych IV. 2013. A single mutation in a regulatory protein produces evolvable allosterically regulated catalyst of nonnatural reaction. *Angew Chem Int Ed Engl* 52:6246-6249.
Oates ME, Romero P, Ishida T, Ghalwash M, Mizianty MJ, Xue B, Dosztanyi Z, Uversky VN, Obradovic Z, Kurgan L, Dunker AK, and Gough J. 2013. D(2)P(2): database of disordered protein predictions. *Nucleic Acids Res* 41:D508-516.
Obradovic Z, Peng K, Vucetic S, Radivojac P, and Dunker AK. 2005. Exploiting heterogeneous sequence properties improves prediction of protein disorder. *Proteins* 61 Suppl 7:176-182.
Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, and Dunker AK. 2005a. Comparing and combining predictors of mostly disordered proteins. *Biochemistry* 44:1989-2000.
Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, and Dunker AK. 2005b. Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry* 44:12454-12470.
Pace CN, Vajdos F, Fee L, Grimsley G, and Gray T. 1995. How to Measure and Predict the Molar Absorption-Coefficient of a Protein. *Protein Science* 4:2411-2423.
Patil A, and Nakamura H. 2006. Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett* 580:2041-2045.
Pejaver V, Hsu WL, Xin F, Dunker AK, Uversky VN, and Radivojac P. 2014. The structural and functional signatures of proteins that undergo multiple events of post-translational modification. *Protein Sci* 23:1077-1093.
Peng K, Radivojac P, Vucetic S, Dunker AK, and Obradovic Z. 2006a. Length-dependent prediction of protein intrinsic disorder. *Bmc Bioinformatics* 7.
Peng K, Radivojac P, Vucetic S, Dunker AK, and Obradovic Z. 2006b. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* 7:208.

650 Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, and Obradovic Z. 2005. Optimizing
651 long intrinsic disorder predictors with protein evolutionary information. *J Bioinform*
652 *Comput Biol* 3:35-60.
653 Peng ZL, and Kurgan L. 2012. Comprehensive comparative assessment of in-silico predictors of
654 disordered regions. *Curr Protein Pept Sci* 13:6-18.
655 Permyakov EA, and Burstein EA. 1984. Some aspects of studies of thermal transitions in
656 proteins by means of their intrinsic fluorescence. *Biophys Chem* 19:265-271.
657 Permyakov EA, Burstein EA, Sawada Y, and Yamazaki I. 1977. Luminescence of
658 phenylalamine residues in superoxide dismutase from green pea. *Biochim Biophys Acta*
659 491:149-154.
660 Permyakov SE, Millett IS, Doniach S, Permyakov EA, and Uversky VN. 2003. Natively
661 unfolded C-terminal domain of caldesmon remains substantially unstructured after the
662 effective binding to calmodulin. *Proteins* 53:855-862.
663 Polyakov AA, Huber PAJ, Marston SB, and Gusev NB. 1998. Interaction of isoforms of S100
664 protein with smooth muscle caldesmon. *Febs Letters* 422:235-239.
665 Potenza E, Domenico TD, Walsh I, and Tosatto SC. 2015. MobiDB 2.0: an improved database of
666 intrinsically disordered and mobile proteins. *Nucleic Acids Res*.
667 Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, Silman I, and
668 Sussman JL. 2005. FoldIndex: a simple tool to predict whether a given protein sequence
669 is intrinsically unfolded. *Bioinformatics* 21:3435-3438.
670 Privalov PL. 1979. Stability of proteins: small globular proteins. *Adv Protein Chem* 33:167-241.
671 Privalov PL, and Potekhin SA. 1986. Scanning microcalorimetry in studying temperature-
672 induced changes in proteins. *Methods Enzymol* 131:4-51.
673 Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, and Dunker AK. 2007.
674 Intrinsic disorder and functional proteomics. *Biophys J* 92:1439-1456.
675 Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, and Dunker AK. 2001. Sequence
676 complexity of disordered protein. *Proteins* 42:38-48.
677 Shirinsky VP, Vorotnikov AV, and Gusev NB. 1999. Caldesmon phosphorylation and smooth
678 muscle contraction. In: Kohama K, and Sasaki Y, eds. *Molecular Biology of Smooth*
679 *Muscle Contraction*. Austin, TX, USA: R.G. Landes Company, 59-79.
680 Singh GP, Ganapathi M, Sandhu KS, and Dash D. 2006. Intrinsic unstructuredness and
681 abundance of PEST motifs in eukaryotic proteomes. *Proteins* 62:309-315.
682 Sobue K, and Sellers JR. 1991. Caldesmon, a novel regulatory protein in smooth muscle and
683 nonmuscle actomyosin systems. *J Biol Chem* 266:12115-12118.
684 Sulmann S, Dell'Orco D, Marino V, Behnen P, and Koch KW. 2014. Conformational changes in
685 calcium-sensor proteins under molecular crowding conditions. *Chemistry* 20:6756-6762.
686 Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, Doerks T, Stark M,
687 Muller J, Bork P, Jensen LJ, and von Mering C. 2011. The STRING database in 2011:
688 functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids*
689 *Res* 39:D561-568.
690 Tokuriki N, Oldfield CJ, Uversky VN, Berezovsky IN, and Tawfik DS. 2009. Do viral proteins
691 possess unique biophysical features? *Trends Biochem Sci* 34:53-59.
692 Tompa P. 2002. Intrinsically unstructured proteins. *Trends Biochem Sci* 27:527-533.
693 Tompa P. 2005. The interplay between structure and function in intrinsically unstructured
694 proteins. *FEBS Lett* 579:3346-3354.

695 Tompa P. 2012. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci* 37:509-
696 516.
697 Tompa P, and Csermely P. 2004. The role of structural disorder in the function of RNA and
698 protein chaperones. *FASEB J* 18:1169-1175.
699 Tompa P, Szasz C, and Buday L. 2005. Structural disorder throws new light on moonlighting.
700 *Trends Biochem Sci* 30:484-489.
701 Turoverov KK, Kuznetsova IM, and Uversky VN. 2010. The protein kingdom extended: ordered
702 and intrinsically disordered proteins, their folding, supramolecular complex formation,
703 and aggregation. *Prog Biophys Mol Biol* 102:73-84.
704 Uversky VN. 2002a. Natively unfolded proteins: a point where biology waits for physics.
705 *Protein Sci* 11:739-756.
706 Uversky VN. 2002b. What does it mean to be natively unfolded? *Eur J Biochem* 269:2-12.
707 Uversky VN. 2003. Protein folding revisited. A polypeptide chain at the folding-misfolding-
708 nonfolding cross-roads: which way to go? *Cell Mol Life Sci* 60:1852-1871.
709 Uversky VN. 2010. The mysterious unfoldome: structureless, underappreciated, yet vital part of
710 any given proteome. *J Biomed Biotechnol* 2010:568068.
711 Uversky VN. 2011a. Intrinsically disordered proteins from A to Z. *Int J Biochem Cell Biol*
712 43:1090-1103.
713 Uversky VN. 2011b. Multitude of binding modes attainable by intrinsically disordered proteins:
714 a portrait gallery of disorder-based complexes. *Chem Soc Rev* 40:1623-1634.
715 Uversky VN. 2012. Disordered competitive recruiter: fast and foldable. *J Mol Biol* 418:267-268.
716 Uversky VN. 2013a. A decade and a half of protein intrinsic disorder: Biology still waits for
717 physics. *Protein Sci* 22:693-724.
718 Uversky VN. 2013b. Intrinsic Disorder-based Protein Interactions and their Modulators. *Curr*
719 *Pharm Des* 19:4191-4213.
720 Uversky VN. 2013c. Unusual biophysics of intrinsically disordered proteins. *Biochim Biophys*
721 *Acta* 1834:932-951.
722 Uversky VN, Dave V, Iakoucheva LM, Malaney P, Metallo SJ, Pathak RR, and Joerger AC.
723 2014. Pathological unfoldomics of uncontrolled chaos: intrinsically disordered proteins
724 and human diseases. *Chem Rev* 114:6844-6879.
725 Uversky VN, and Dunker AK. 2010. Understanding protein non-folding. *Biochim Biophys Acta*
726 1804:1231-1264.
727 Uversky VN, Gillespie JR, and Fink AL. 2000. Why are "natively unfolded" proteins
728 unstructured under physiologic conditions? *Proteins* 41:415-427.
729 Uversky VN, Oldfield CJ, and Dunker AK. 2005. Showing your ID: intrinsic disorder as an ID
730 for recognition, regulation and cell signaling. *J Mol Recognit* 18:343-384.
731 Uversky VN, Oldfield CJ, and Dunker AK. 2008. Intrinsically disordered proteins in human
732 diseases: introducing the D2 concept. *Annu Rev Biophys* 37:215-246.
733 Vacic V, Oldfield CJ, Mohan A, Radivojac P, Cortese MS, Uversky VN, and Dunker AK.
734 2007a. Characterization of molecular recognition features, MoRFs, and their binding
735 partners. *J Proteome Res* 6:2351-2366.
736 Vacic V, Uversky VN, Dunker AK, and Lonardi S. 2007b. Composition Profiler: a tool for
737 discovery and visualization of amino acid composition differences. *BMC Bioinformatics*
738 8:211.

739 Vorotnikov AV, Bogatcheva NV, and Gusev NB. 1992. Caldesmon Phospholipid Interaction -
740 Effect of Protein-Kinase-C Phosphorylation and Sequence Similarity with Other
741 Phospholipid-Binding Proteins. *Biochemical Journal* 284:911-916.
742 Vorotnikov AV, and Gusev NB. 1990a. Interaction of Duck Gizzard Caldesmon with
743 Phospholipids. *Journal of Muscle Research and Cell Motility* 11:442-442.
744 Vorotnikov AV, and Gusev NB. 1990b. Interaction of Smooth-Muscle Caldesmon with
745 Phospholipids. *Febs Letters* 277:134-136.
746 Vucetic S, Xie H, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, and Uversky VN.
747 2007. Functional anthology of intrinsic disorder. 2. Cellular components, domains,
748 technical terms, developmental processes, and coding sequence diversities correlated
749 with long disordered regions. *J Proteome Res* 6:1899-1916.
750 Walsh I, Martin AJ, Di Domenico T, and Tosatto SC. 2012. ESpritz: accurate and fast prediction
751 of protein disorder. *Bioinformatics* 28:503-509.
752 Wang CL, Chalovich JM, Graceffa P, Lu RC, Mabuchi K, and Stafford WF. 1991. A long helix
753 from the central region of smooth muscle caldesmon. *J Biol Chem* 266:13958-13963.
754 Wang EZ, Zhuang SB, Kordowska J, Grabarek Z, and Wang CLA. 1997. Calmodulin binds to
755 caldesmon in an antiparallel manner. *Biochemistry* 36:15026-15034.
756 Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, and Jones DT. 2004. Prediction and functional
757 analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol*
758 337:635-645.
759 Wright PE, and Dyson HJ. 1999. Intrinsically unstructured proteins: re-assessing the protein
760 structure-function paradigm. *J Mol Biol* 293:321-331.
761 Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, and Uversky VN.
762 2007a. Functional anthology of intrinsic disorder. 3. Ligands, post-translational
763 modifications, and diseases associated with intrinsically disordered proteins. *J Proteome*
764 *Res* 6:1917-1932.
765 Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, and Obradovic Z.
766 2007b. Functional anthology of intrinsic disorder. 1. Biological processes and functions
767 of proteins with long disordered regions. *J Proteome Res* 6:1882-1898.
768 Xue B, Dunbrack RL, Williams RW, Dunker AK, and Uversky VN. 2010a. PONDR-FIT: a
769 meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* 1804:996-
770 1010.
771 Xue B, Dunker AK, and Uversky VN. 2012a. Orderly order in protein intrinsic disorder
772 distribution: disorder in 3500 proteomes from viruses and the three domains of life. *J*
773 *Biomol Struct Dyn* 30:137-149.
774 Xue B, Dunker AK, and Uversky VN. 2012b. Orderly order in protein intrinsic disorder
775 distribution: Disorder in thirty five hundred proteomes from viruses and the three
776 domains of life. *Journal of Biomolecular Structure and Dynamics*:In press.
777 Xue B, Williams RW, Oldfield CJ, Dunker AK, and Uversky VN. 2010b. Archaic chaos:
778 intrinsically disordered proteins in Archaea. *BMC Syst Biol* 4 Suppl 1:S1.
779 Yamada Y, Matsuo T, Iwamoto H, and Yagi N. 2012. A Compact Intermediate State of
780 Calmodulin in the Process of Target Binding. *Biochemistry* 51:3963-3970.
781 Yang ZR, Thomson R, McNeil P, and Esnouf RM. 2005. RONN: the bio-basis function neural
782 network technique applied to the detection of natively disordered regions in proteins.
783 *Bioinformatics* 21:3369-3376.
784

## Table 1(on next page)

Table 1

Equilibrium association constants for complexes between CaM and wild type $CaD_{136}$ and its mutants and their relative fluorescence quantum yields in the free and CaM-bound states.

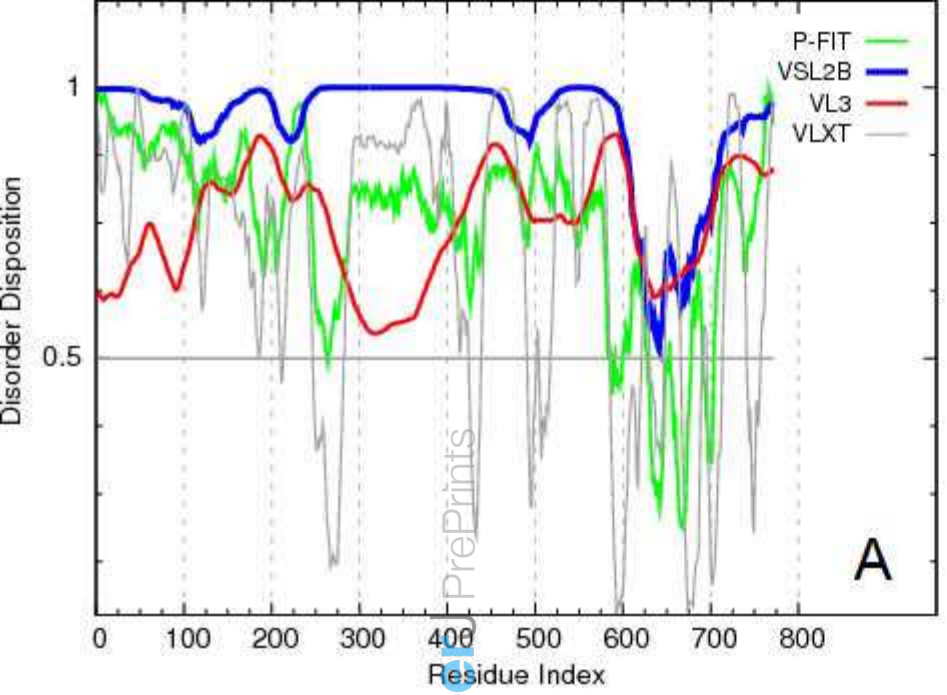
**Table 1**. Equilibrium association constants for complexes between CaM and wild type $CaD_{136}$ and its mutants and their relative fluorescence quantum yields in the free and CaM-bound states.

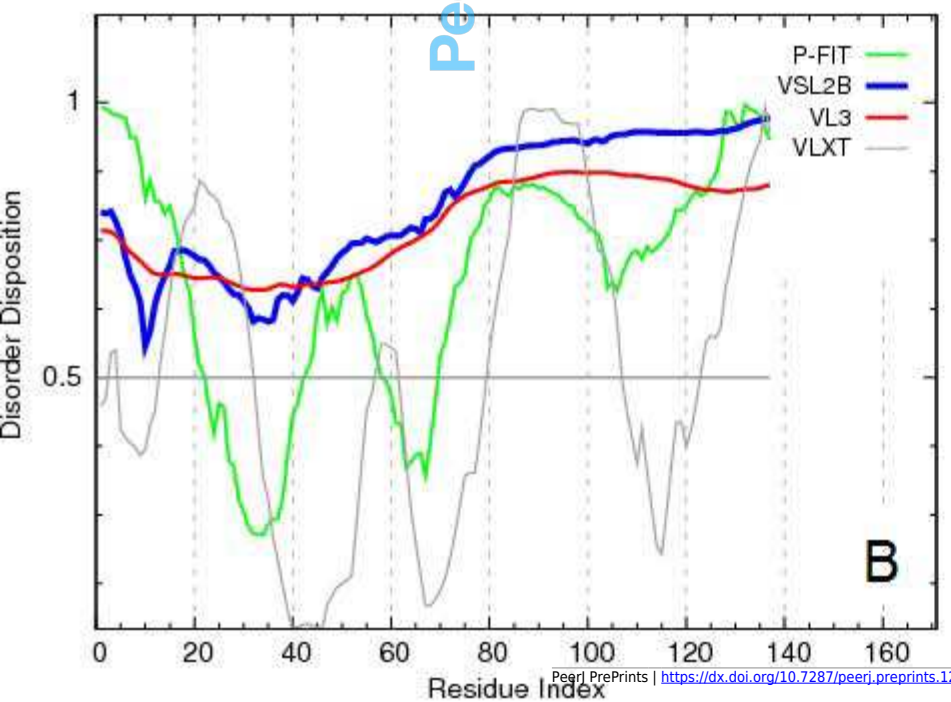
| Protein | $K_{CaM}$ | $Q/Q_{trp}$ (in solution) | $Q/Q_{trp}$ (in complex with calmodulin) |
|---|---|---|---|
| WT | $6.5x10^5$ | 1.25 | 2.40 |
| W674A | $2.2x10^5$ | 1.25 | 2.72 |
| W707A | $3.0x10^5$ | 1.50 | 2.55 |
| W737A | $1.8x10^6$ | 1.49 | 2.95 |
| Double mutant | $4.4x10^4$ | 1.19 | 2.64 |

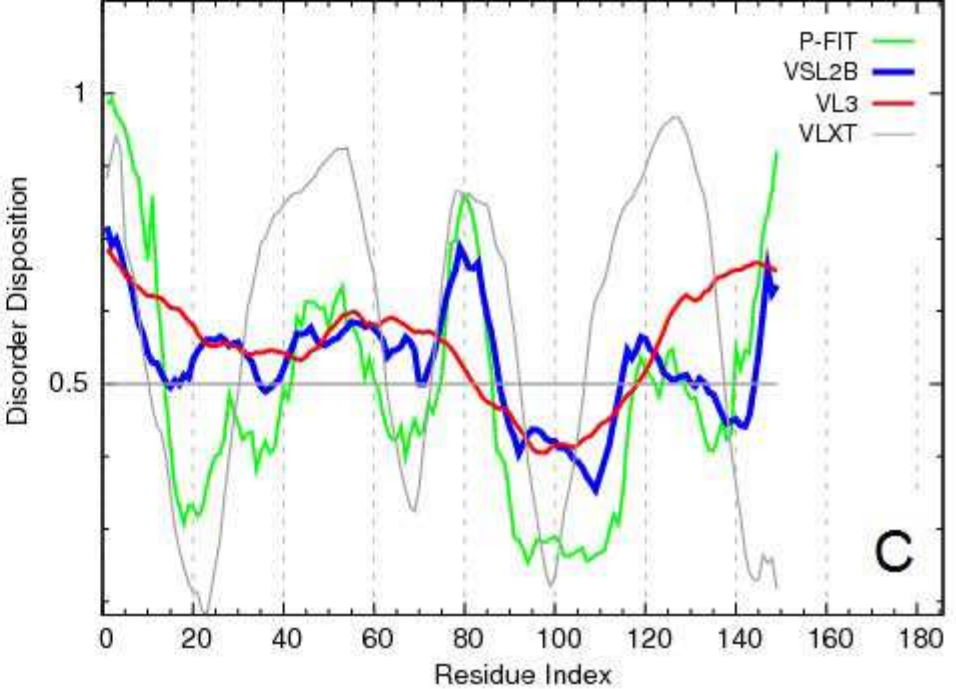
## Figure 1(on next page)

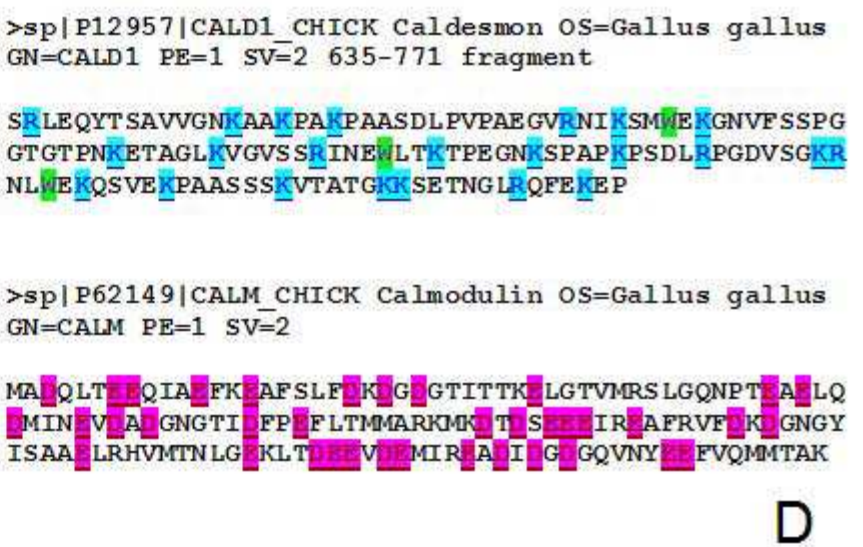
Evaluating the intrinsic disorder propensities of proteins.

**Figure 1.** Evaluating the intrinsic disorder propensities of chicken CaD (**A**), $CaD_{136}$ (**B**), and chicken CaM (**C**) by the family of PONDR predictors. A disorder threshold is indicated as a thin line (at score = 0.5) in all plots to show a boundary between disorder (>0.5) and order (<0.5). Plot **D** represents the amino acid sequences of $CaD_{136}$ and CaM, for which the positively and negatively charged residues are highlighted. The positions of tryptophan residues within the $CaD_{136}$ sequence are also indicated.



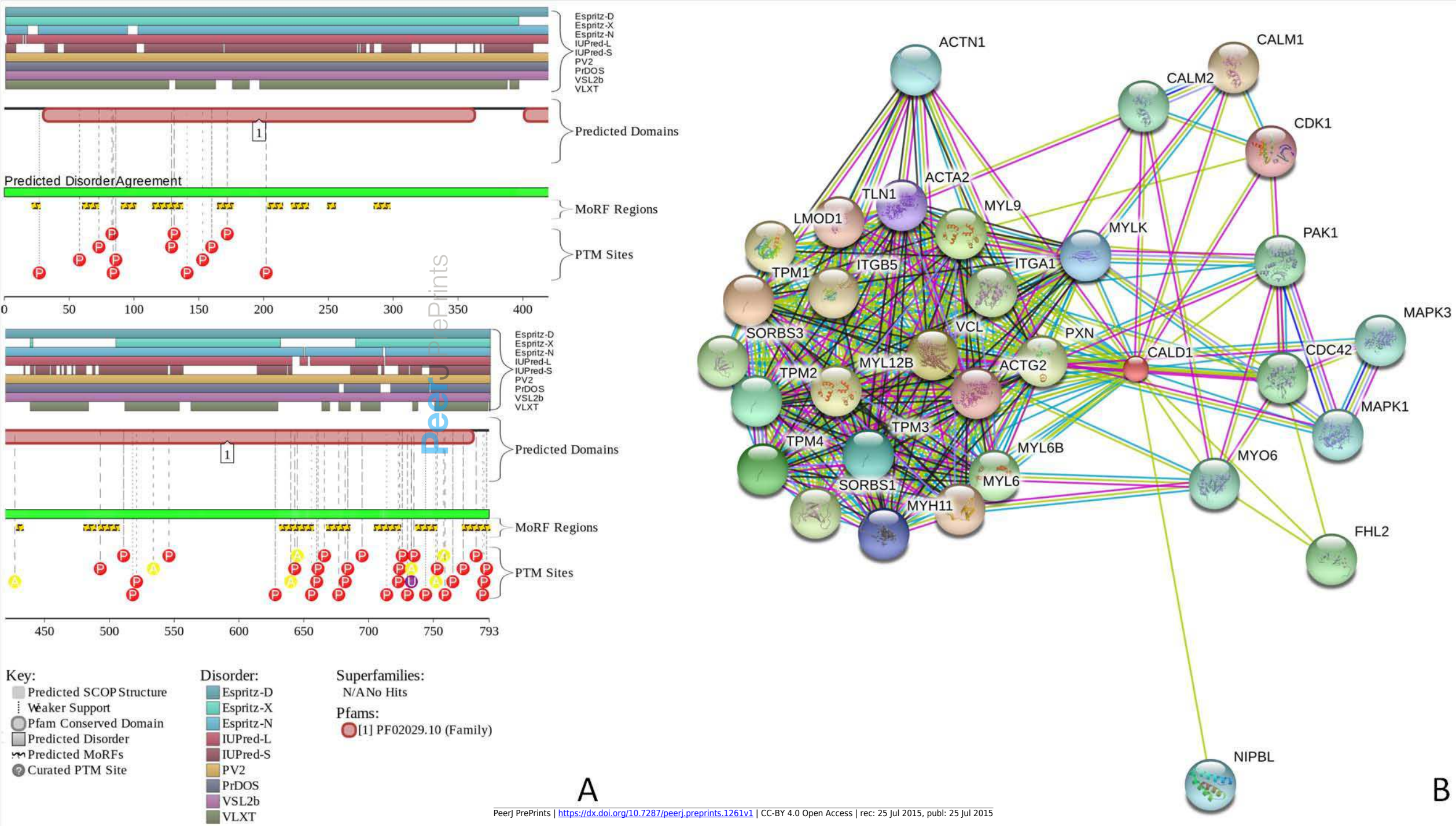



PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1261v1 | CC-BY 4.0 Open Access | rec: 25 Jul 2015, publ: 25 Jul 2015

## Figure 2(on next page)

Evaluation of the functional intrinsic disorder propensity of human CaD.

**Figure 2.** Evaluation of the functional intrinsic disorder propensity of the human CaD (UniProt ID: Q05682) by the $D^2P^2$ platform ( http://d2p2.pro/ ) (Oates et al. 2013) . In this plot, top nine colored bars represent location of disordered regions predicted by different computational tools (Espritz-D, Espritz-N, Espritz-X, IUPred-L, IUPred-S, PV2, PrDOS, PONDR® VSL2b, and PONDR® VLXT, see keys for the corresponding color codes). Dark red bar shows the location of the functional domain found by the Pfam platform, which is a database ofprotein familiesthat includes their annotations andmultiple sequence alignmentsgenerated usinghidden Markov models (Bateman et al. 2004; Finn et al. 2006; Finn et al. 2008) . Green-and-white bar in the middle of the plot shows the predicted disorder agreement between these nine predictors, with green parts corresponding to disordered regions by consensus. Red, yellow and purple circles at the bottom of the plot show the locations of phosphorylation, acetylation and ubiquitination sites, respectively. **B**. Analysis of the interactivity of the chicken gizzard CaD (UniProt ID: P12957) by STRING (Szklarczyk et al. 2011) . STRING produces the network of predicted associations for a particular group of proteins. The network nodes are proteins, whereas the edges represent the predicted or known functional associations. An edge may be drawn with up to 7 differently colored lines that represent the existence of the seven types of evidence used in predicting the associations. A red line indicates the presence of fusion evidence; a green line - neighborhood evidence; a blue line – co-occurrence evidence; a purple line - experimental evidence; a yellow line – text mining evidence; a light blue line - database evidence; a black line – co-expression evidence (Szklarczyk et al. 2011) .



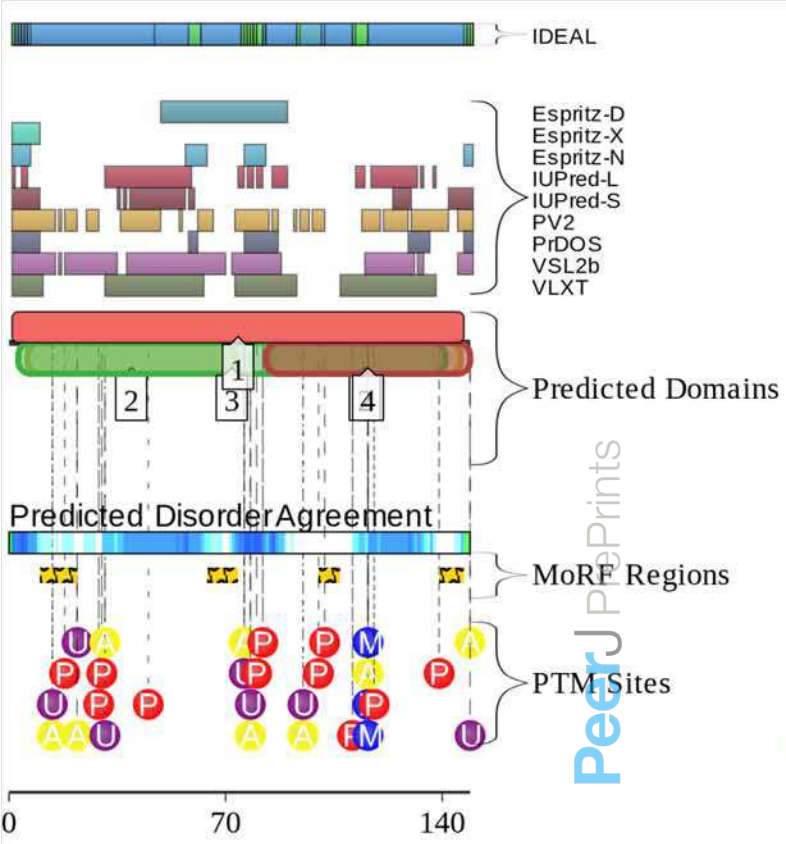



PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1261v1 | CC-BY 4.0 Open Access | rec: 25 Jul 2015, publ: 25 Jul 2015

## Figure 3(on next page)

Evaluation of the functional intrinsic disorder propensity of human CaM.

**Figure 3.** Evaluation of the functional intrinsic disorder propensity of human CaM (UniProt ID: P62158) by $D^2P^2$ database ( http://d2p2.pro/ ) (Oates et al. 2013) . In this plot, top dark blue bar with green stripes shows the localization of disordered region annotated in the IDEAL database (Fukuchi et al. 2012) for this protein. Next nine colored bars represent location of disordered regions predicted by different disorder predictors (Espritz-D, Espritz-N, Espritz-X, IUPred-L, IUPred-S, PV2, PrDOS, PONDR® VSL2b, and PONDR® VLXT, see keys for the corresponding color codes). Dark red bar shows the location of the functional domain found by the Pfam platform, which is a database ofprotein familiesthat includes their annotations andmultiple sequence alignmentsgenerated usinghidden Markov models. (Bateman et al. 2004; Finn et al. 2006; Finn et al. 2008) Blue-and-white bar in the middle of the plot shows the predicted disorder agreement between these nine predictors, with green parts corresponding to disordered regions by consensus. Red, yellow, purple and blue circles at the bottom of the plot show the location of phosphorylation, acetylation, ubiquitination, and methylation sites, respectively. **B**. Analysis of the interactivity of the chicken CaM (UniProt ID: P62149) by STRING (Szklarczyk et al. 2011) .



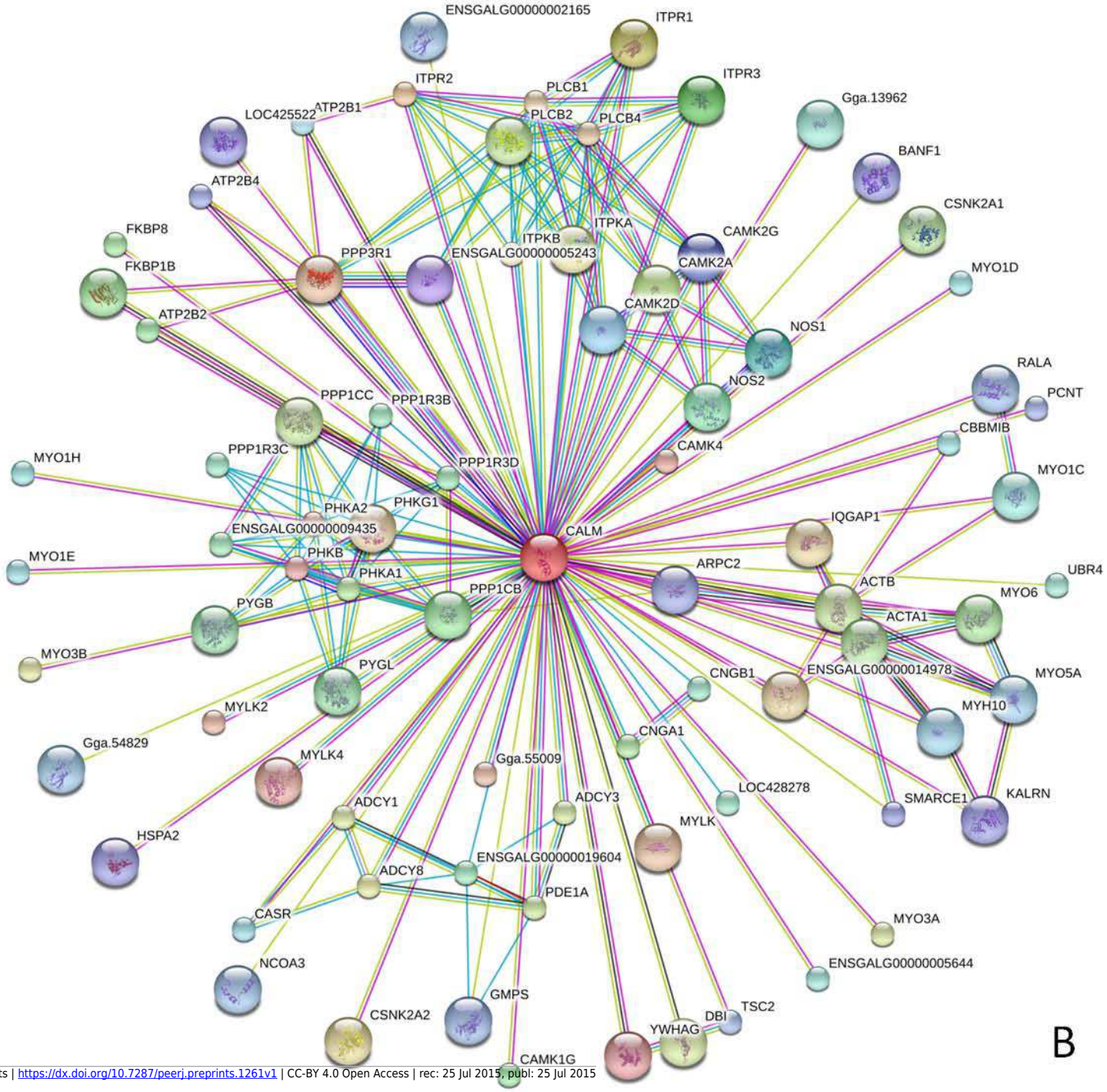



Key:
- Predicted SCOP Structure
- Weaker Support
- Pfam Conserved Domain
- Predicted Disorder
- Predicted MoRFs
- Curated PTM Site

Superfamilies:
- [1] EF-hand

Pfams:
- [2] EF-hand domain pair
- [3] PB005796 (Pfam-B)
- [4] PB016454 (Pfam-B)

Disorder:
- Espritz-D
- Espritz-X
- Espritz-N
- IUPred-L
- IUPred-S
- PV2
- PrDOS
- VSL2b
- VLXT




A

B

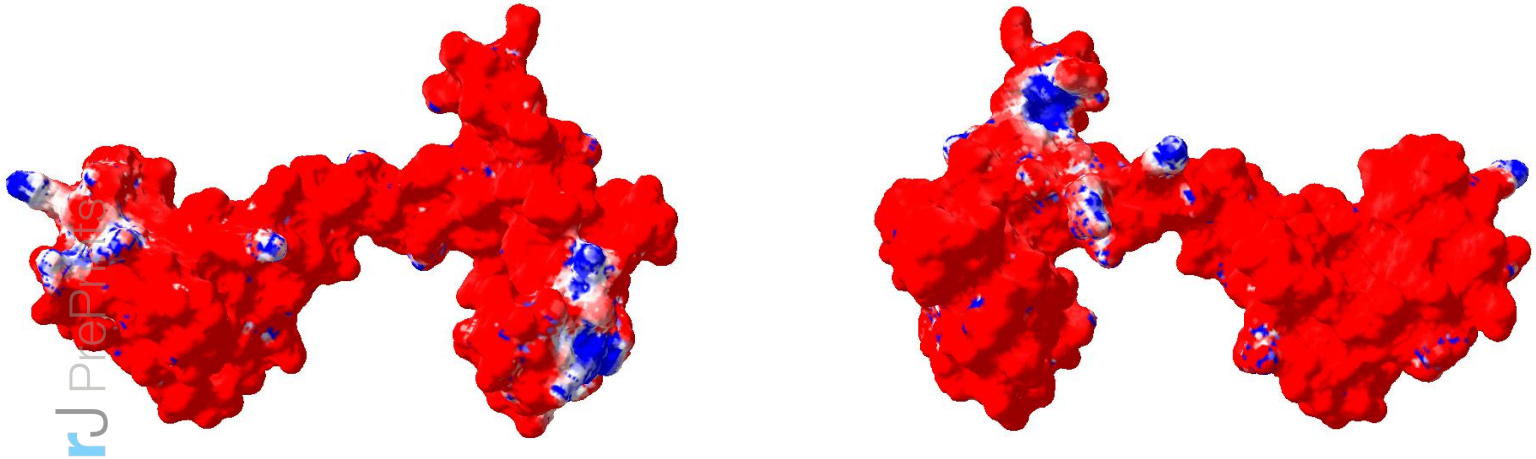
PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1261v1 | CC-BY 4.0 Open Access | rec: 25 Jul 2015, publ: 25 Jul 2015

## Figure 4(on next page)

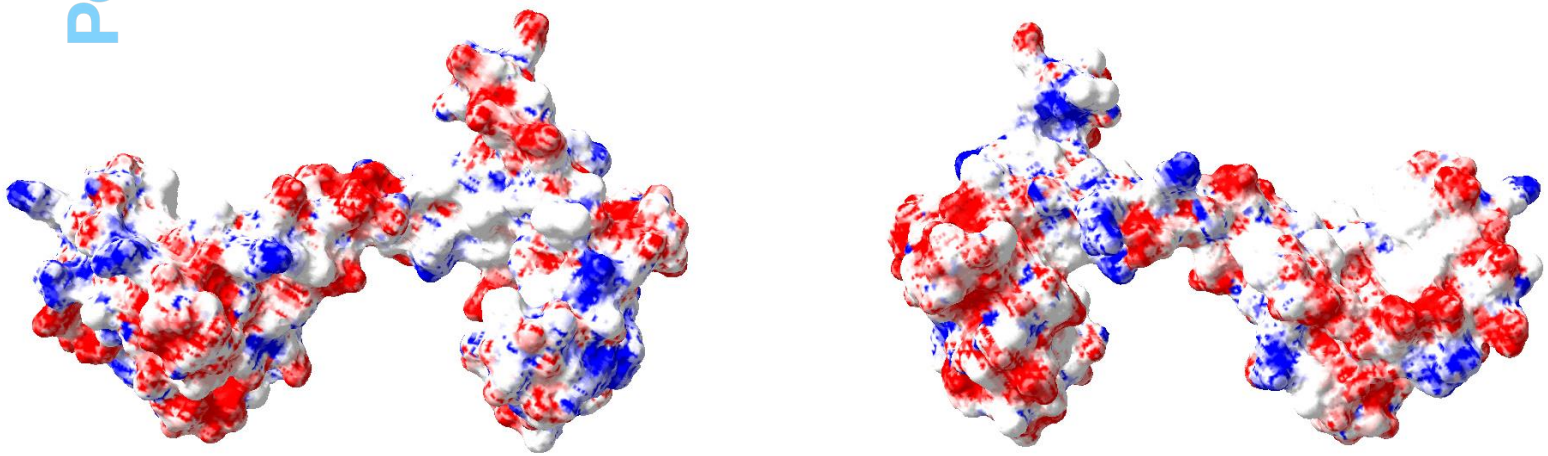
Charge distribution on the surface of CaM molecule.

**Figure 4.** Analysis of the charge distribution on the surface of CaM molecule. PDB file: 1CLM. Analyzed protein: calmodulin, $Ca^{2+}$-form (1 chain, 4 Ca ions), without first 3 residues Ala, Gln, and Glu and without a last residue Lys. $Ca^{2+}$ ions and water molecules were removed, absent hydrogen atoms were added. Calculations were done using the Swiss-PdbViewer v3.7b2 program. Method of calculation: Poisson-Boltzmann, using partial atom charges, ionic strength 0M or 0.05M, dielectric constant of solvent 80, for protein - 4. Colors: Red - potential value is NEGATIVE, -1.8 kT/e; White - potential value is ZERO; Blue - potential value is POSITIVE, 1.8 kT/e.

**I=0 M**



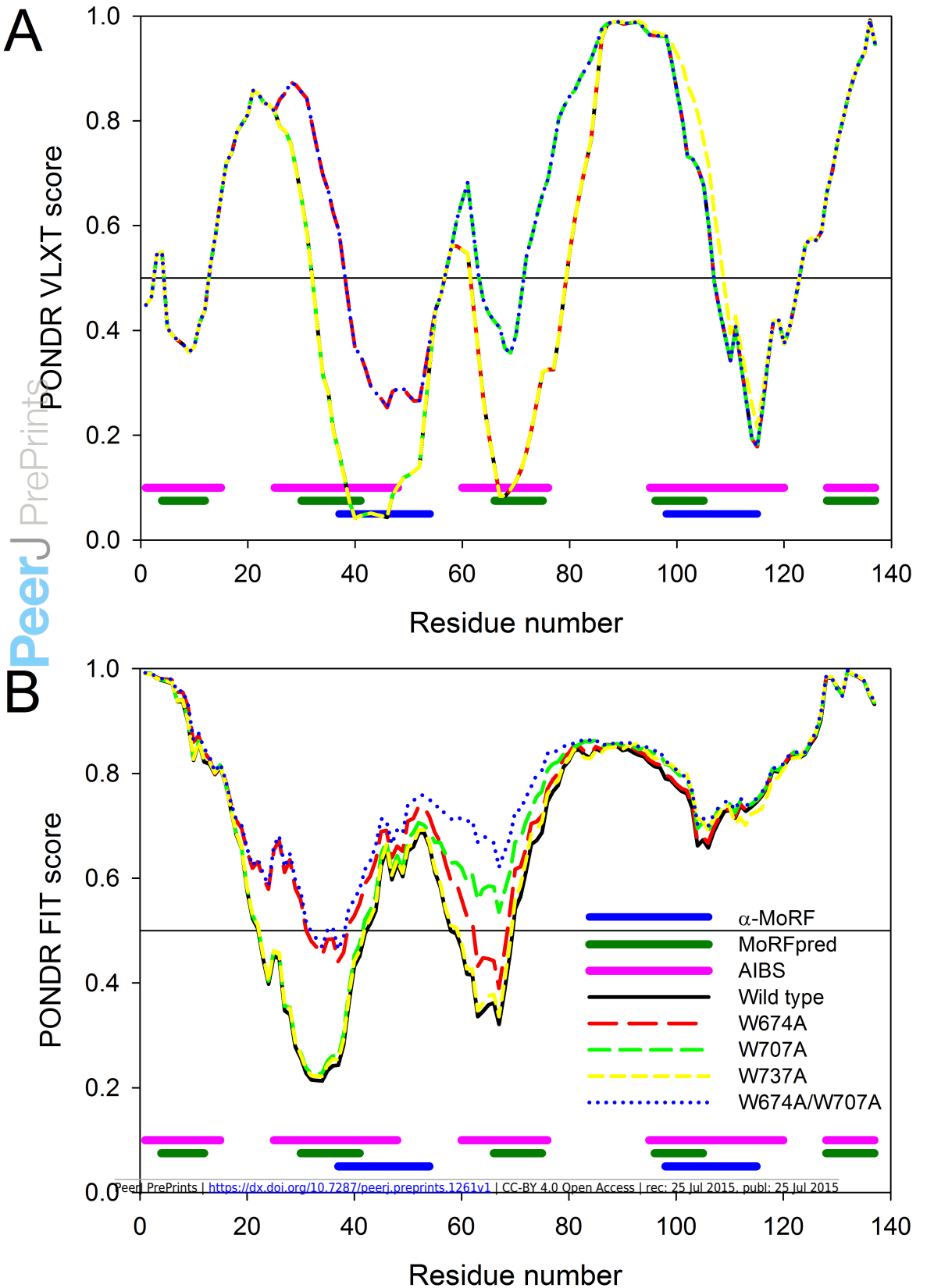
**I=0.05 M**



## Figure 5(on next page)

Effect of tryptophan mutations on the disorder propensity of CaD136.

**Figure 5.** Computational analysis of the effect of tryptophan mutations on the disorder propensity of CaD136 evaluated by PONDR® VLXT (**A**) and PONDR-FIT (**B**). Locations of the predicted disorder-based binding sites are shown at the bottom of plots as pink (AIBSs), dark green (MoRFpreds), and dark blue (a-MoRFs) bars, respectively.



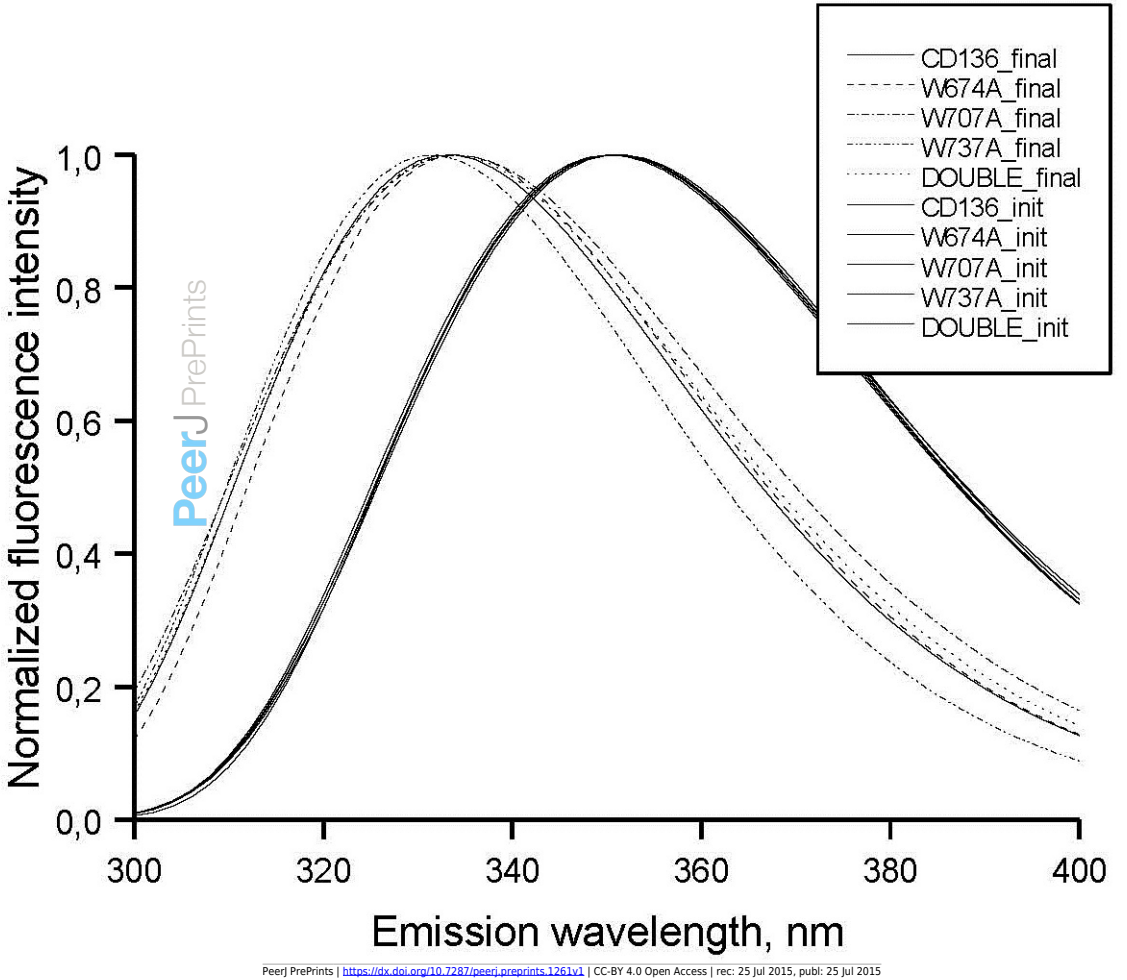



PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1261v1 | CC-BY 4.0 Open Access | rec: 25 Jul 2015, publ: 25 Jul 2015

## Figure 6(on next page)

Tryptophan fluorescence spectra of wild type $CaD_{136}$ and its mutants in the free and CaM-bound states.

**Figure 6.** Tryptophan fluorescence spectra of wild type $CaD_{136}$ and its mutants in the free and CaM-bound states.

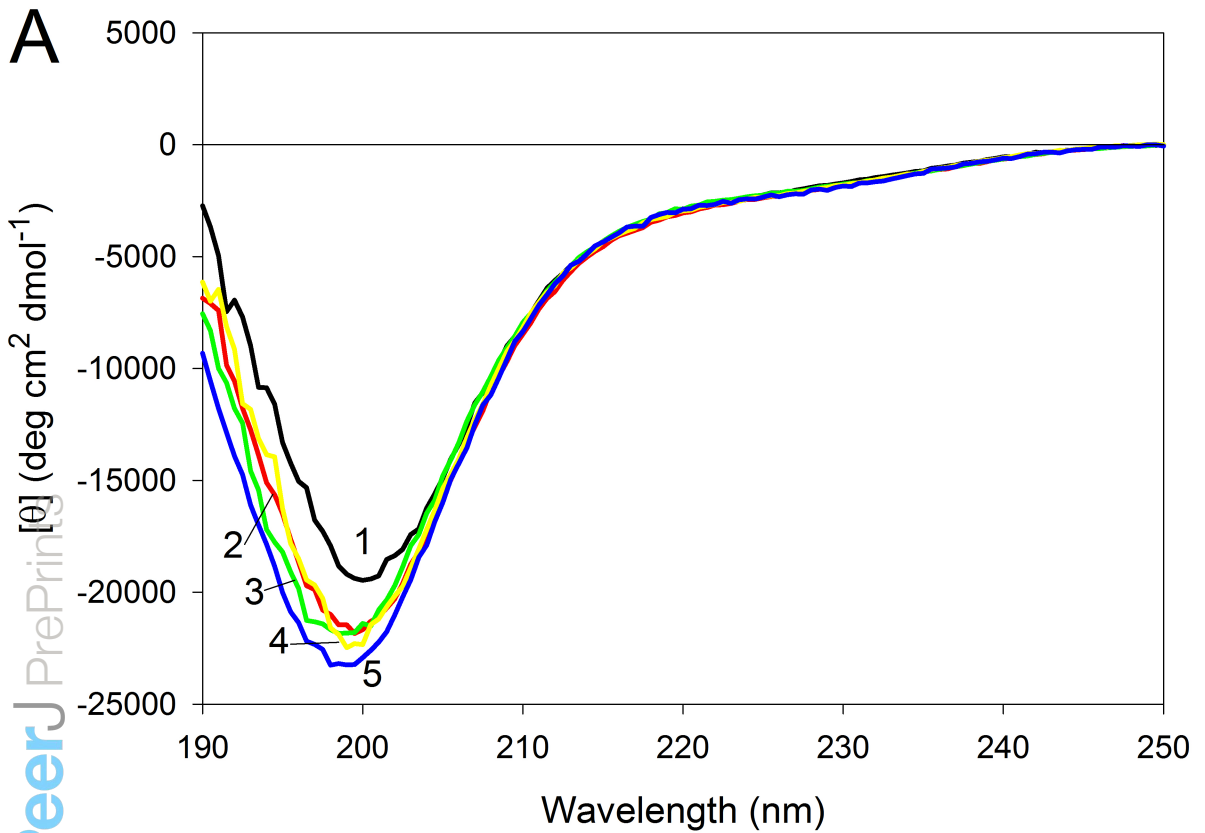


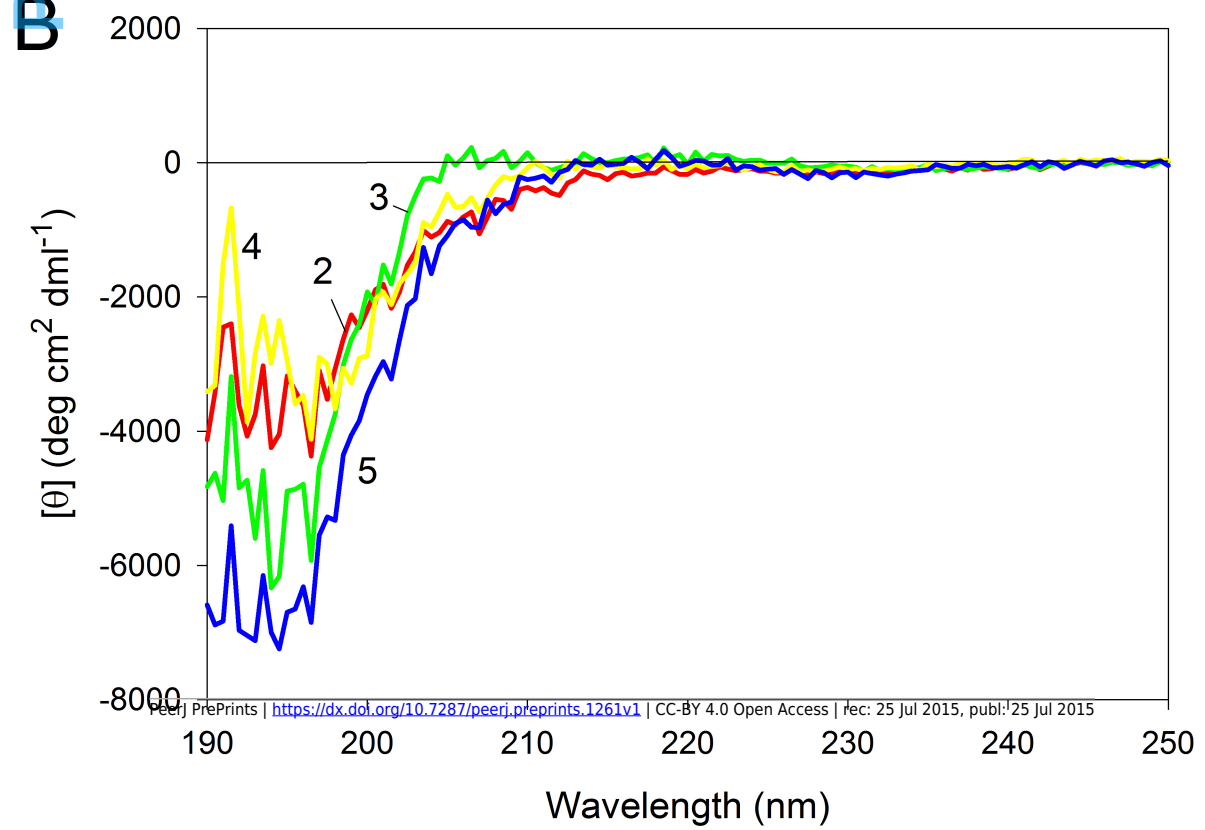

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1261v1 | CC-BY 4.0 Open Access | rec: 25 Jul 2015, publ: 25 Jul 2015

## Figure 7 (on next page)

Effect of mutations on the far-UV CD spectra of CaD.

**Figure 7. A**. Far-UV CD spectra of wild type (**1**), W674A (**2**), W707A (**3**), W737A (**4**) and W674A/W707A (**5**) $CaD_{136}$. All measurements were carried out at a protein concentration of 0.6-0.8 mg/ml, cell pathlength 0.1 mm, 15ºC. **B.** Difference spectra determined by the subtraction from the far-UV CD spectrum of the wild type $CaD_{136}$ the far-UV CD spectrum of: W674A (**2**), W707A (**3**), W737A (**4**) and W674A/W707A (**5**).



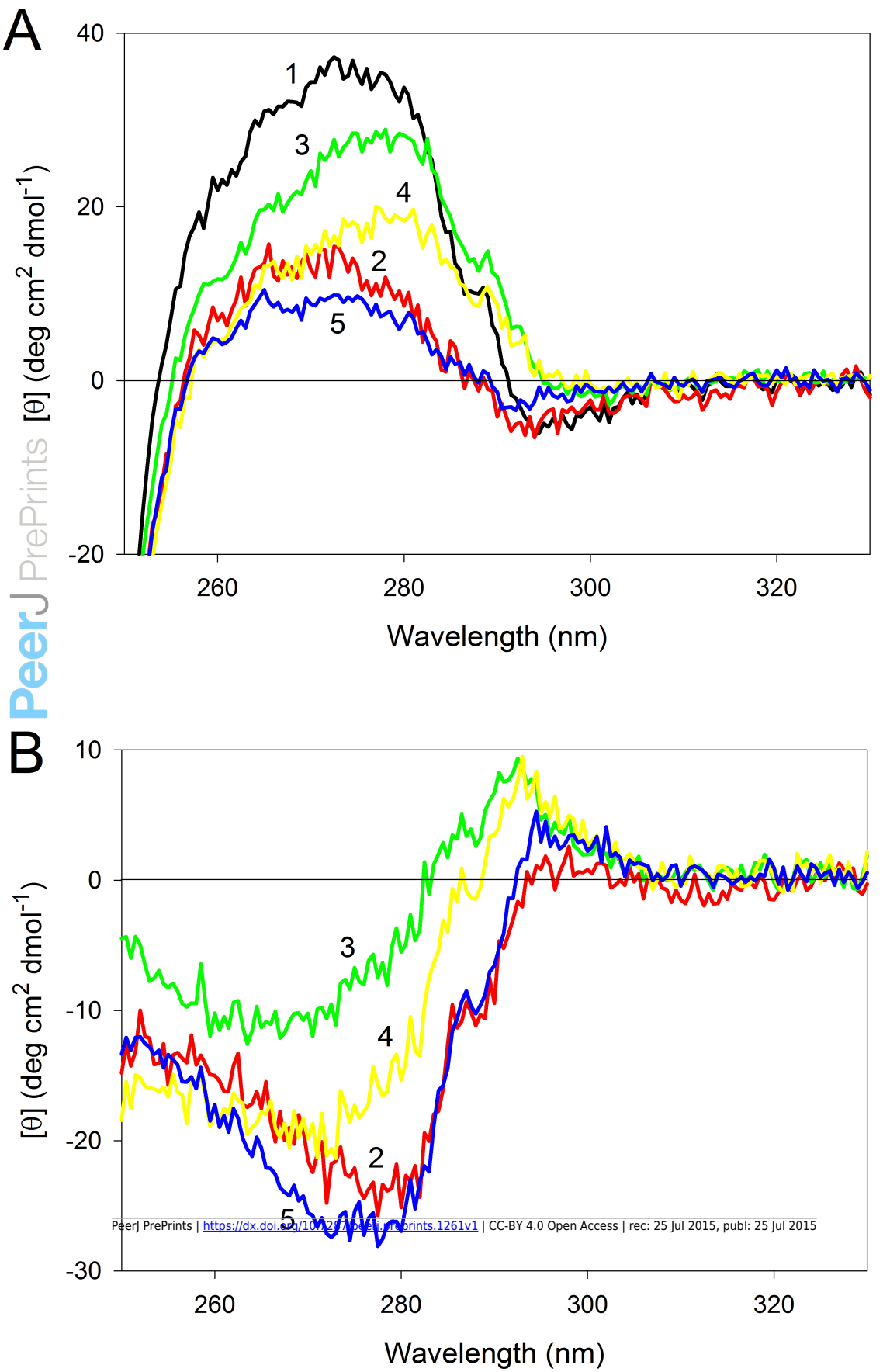



PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1261v1 | CC-BY 4.0 Open Access | rec: 25 Jul 2015, publ: 25 Jul 2015

## Figure 8(on next page)

Effect of mutation on near-UV CD spectra of CaD.

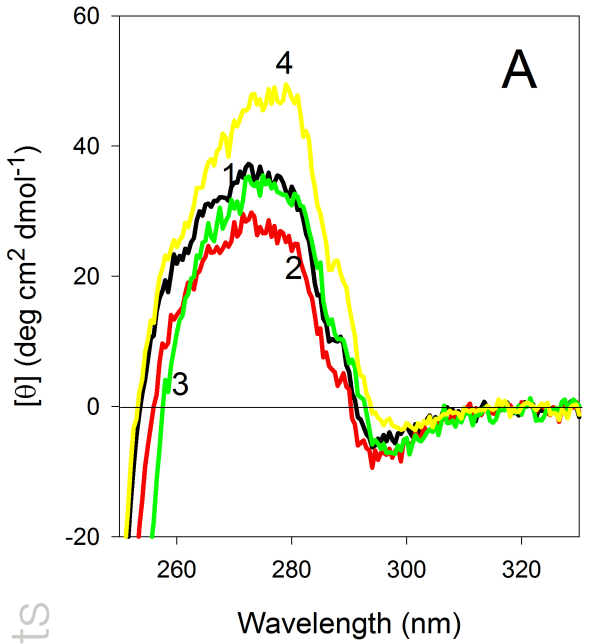
**Figure 8. A**. Near-UV CD spectra of wild type (**1**), W674A (**2**), W707A (**3**), W737A (**4**), and W674A/W707A (**5**) $CaD_{136}$. All measurements were carried out at a protein concentration of 0.6-0.8 mg/ml, cell pathlength 10 mm, 15°C. **B.** Difference spectra determined by the subtraction from the near-UV CD spectrum of the $CaD_{136}$ the near-UV CD spectrum of: W674A (**2**), W707A (**3**), W737A (**4**) and W674A/W707A (**5**).

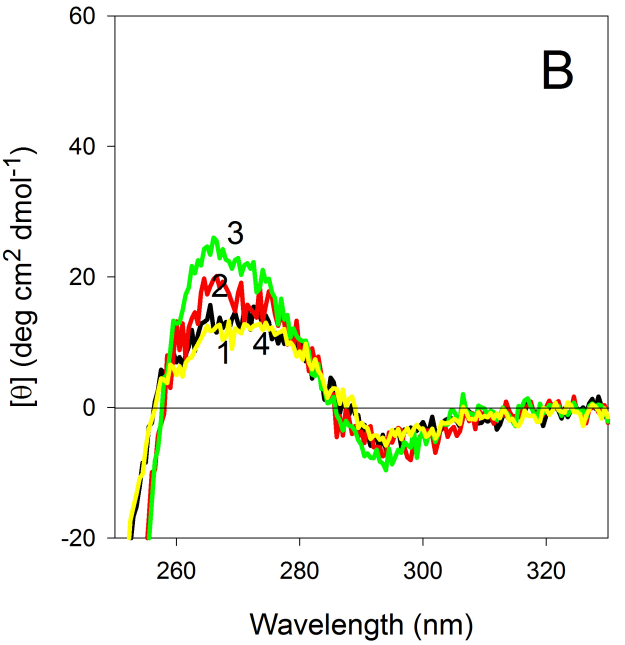


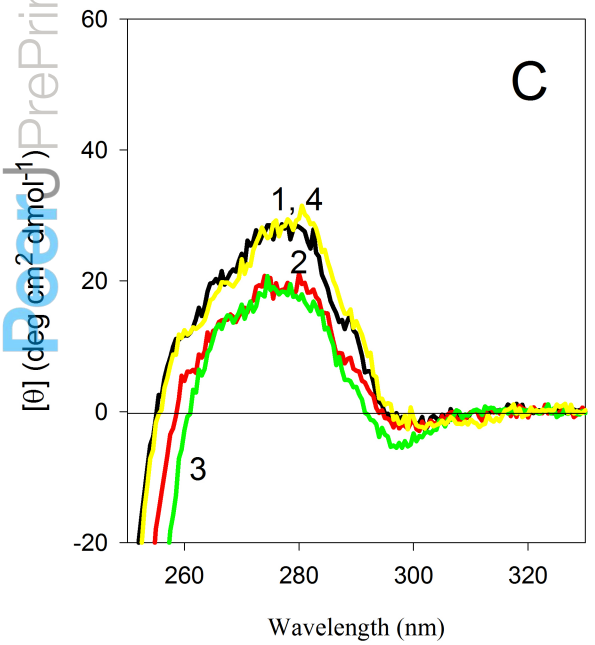
## Figure 9(on next page)

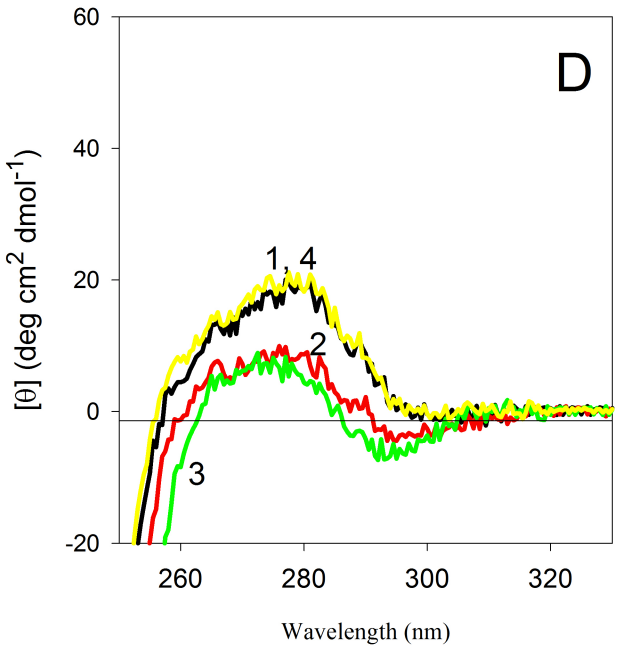
Effect of temperature on the near-UV CD spectra of CaD mutants.

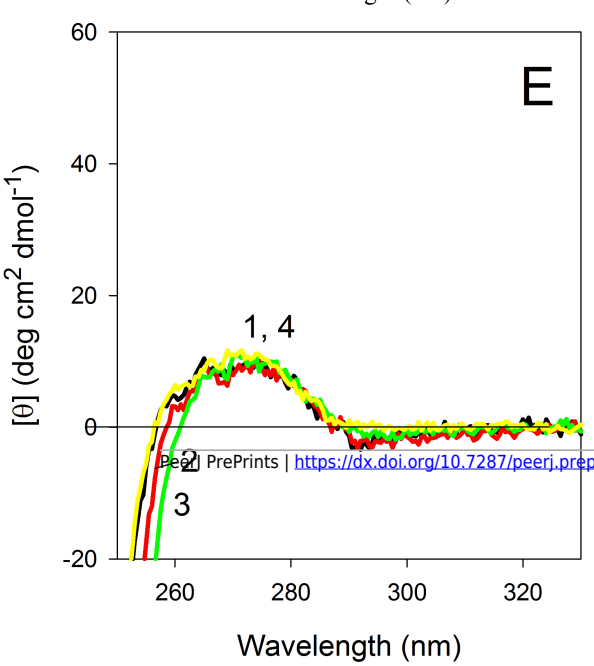
**Figure 9.** Near-UV CD spectra of the wild type (**A**), W674A (**B**), W707A (**C**), W737A (**D**) and W674A/W707A (**E**) $CaD_{136}$ measured at different temperatures: 15°C (**1**); 40°C (**2**), 90°C (**3**) and 15°C after the cooling (**4**). All measurements were carried out at a protein concentration of 0.6-0.8 mg/ml, cell pathlength 10 mm.



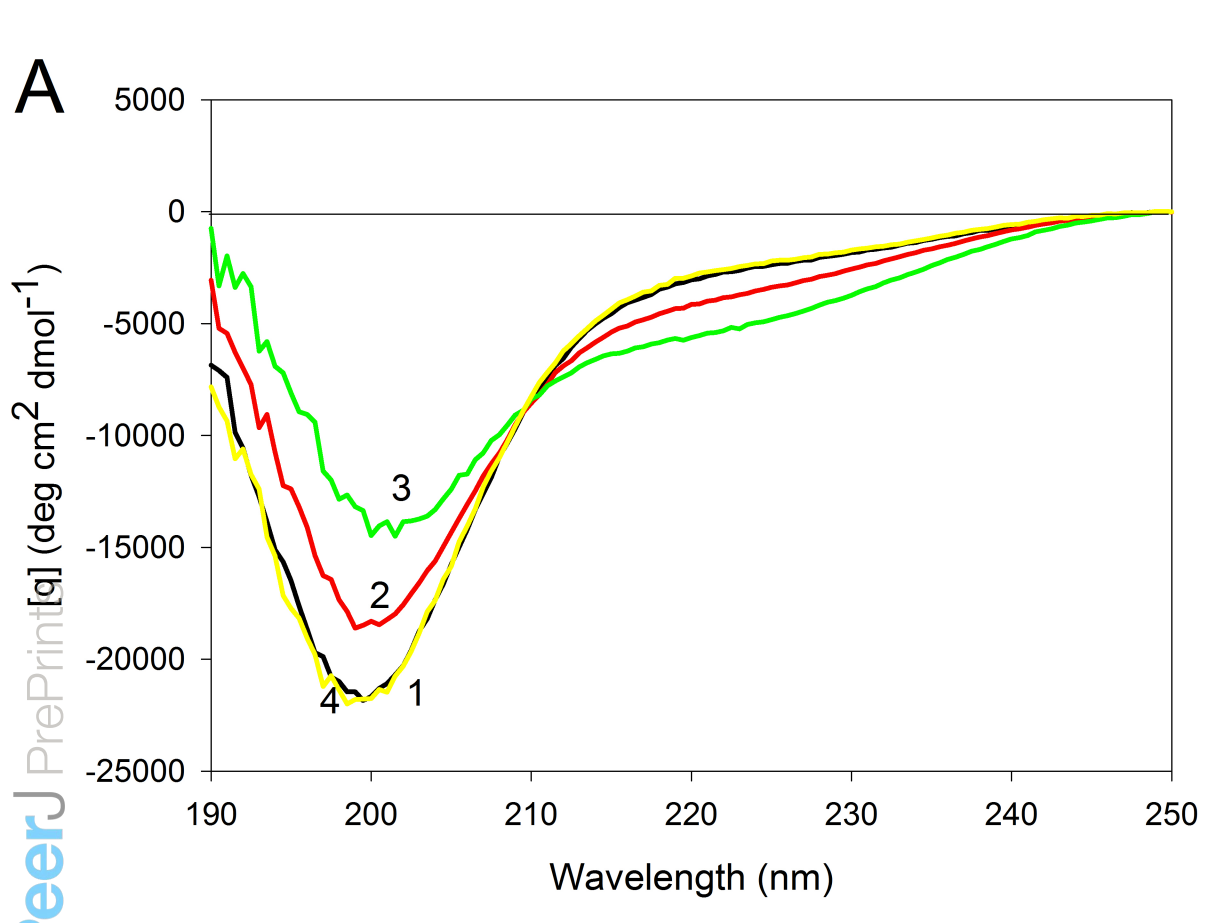


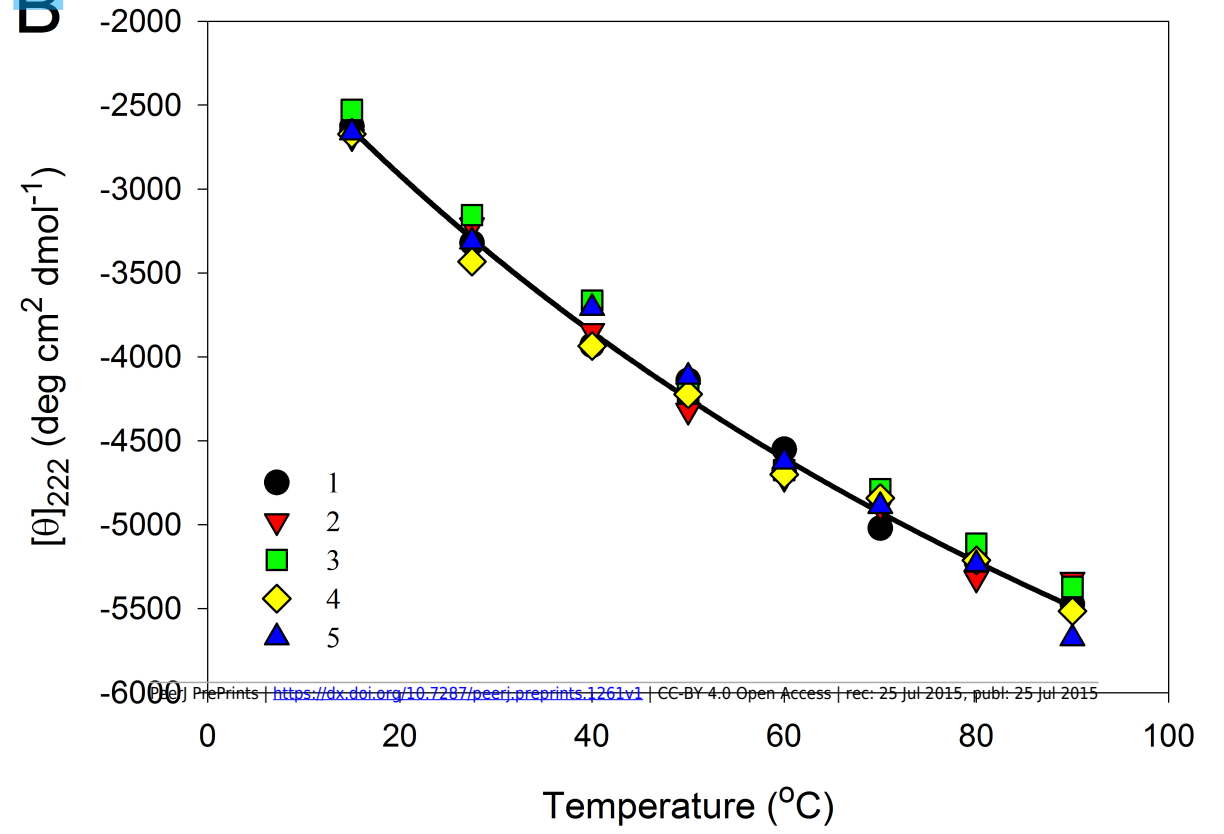

PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1261v1 | CC-BY 4.0 Open Access | rec: 25 Jul 2015, publ: 25 Jul 2015

## Figure 10 (on next page)

Effect of temperature on far-UV CD spectra of CaD mutants.

**Figure 10. A.** Far-UV CD spectra of W674A mutant of $CaD_{136}$ measured at different temperatures: 15°C (**1**); 40°C (**2**), 90°C (**3**) and 15°C after the cooling (**4**). All measurements were carried out at a protein concentration of 0.8 mg/ml, cell pathlength 0.1 mm. **B.** Effect of temperature on far-UV CD spectra of $CaD_{136}$ and its mutants: wild type (**1**), W674A (**2**), W707A (**3**), W737A (**4**) and W674A/W707A (**5**).



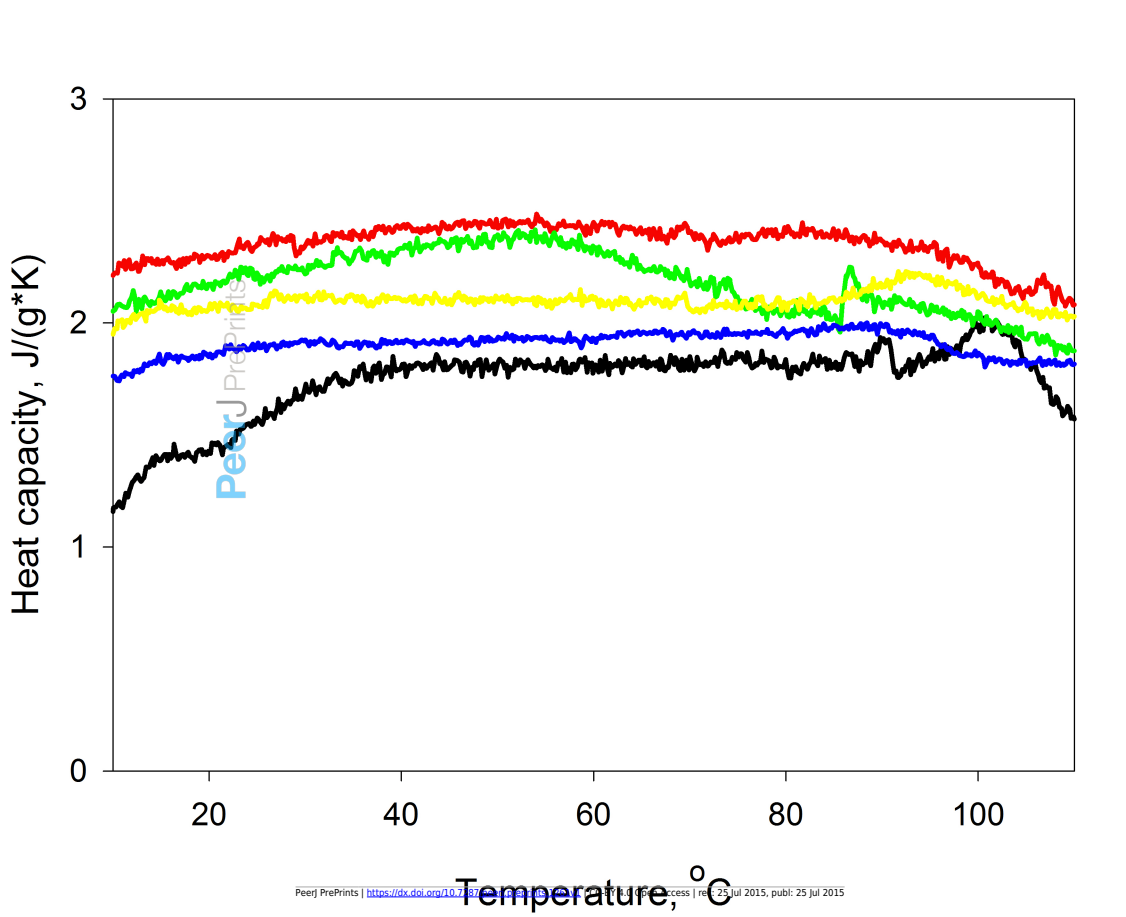



PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1261v1 | CC-BY 4.0 Open Access | rec: 25 Jul 2015, publ: 25 Jul 2015

## Figure 11 (on next page)

Calorimetric analysis of the thermal stability of CaD mutants.

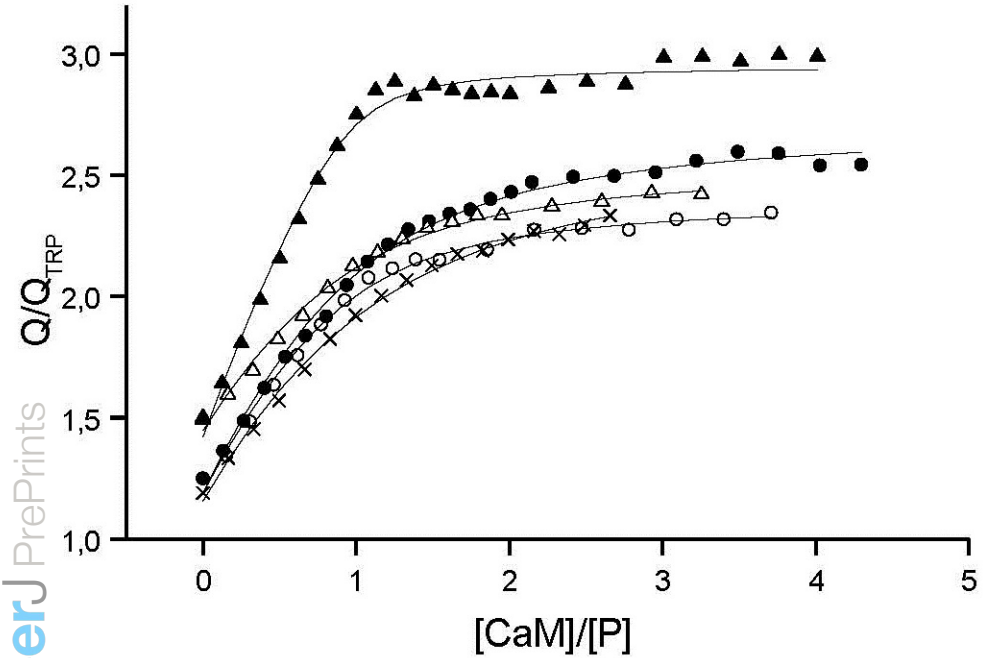
**Figure 11.** Calorimetric scans for wild type $CaD_{136}$ and its mutants in solution. Experiments were performed in 50 mM $H_3BO_3$ buffer, pH 8.0. Protein concentrations were 0.97 mg/ml, 1.38 mg/ml, 1.21 mg/ml, 1.56 mg/ml and 1.96 mg/ml for the wild type (black curve), W674A (red curve), W707A (green curve), W737A (yellow curve), and W674A/W707A (blue curve), respectively.

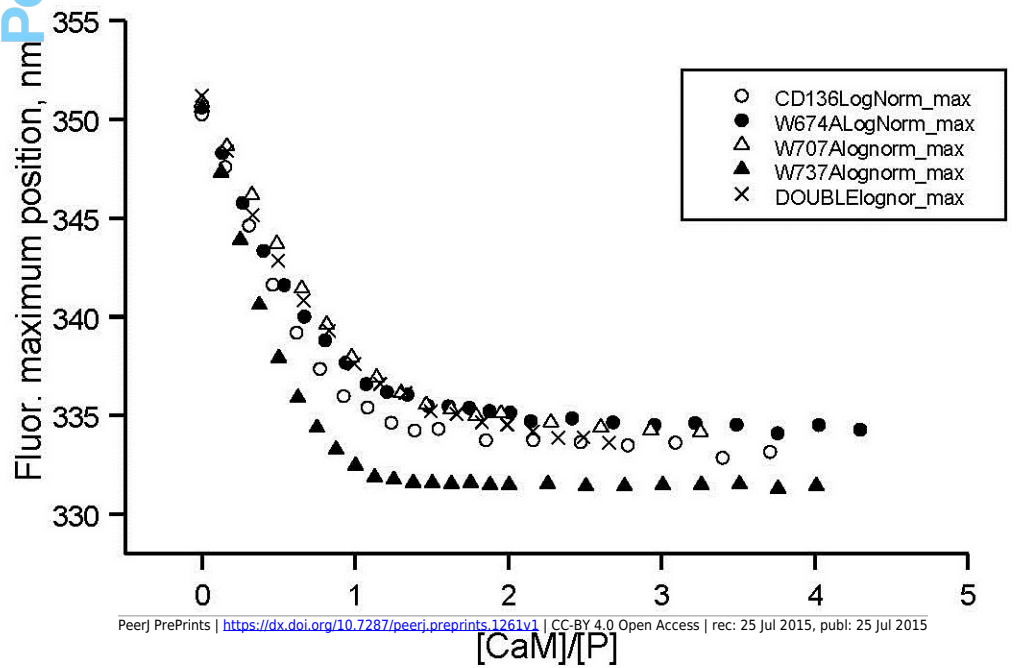


## Figure 12(on next page)

Spectrofluorimetric titration of the $CaD_{136}$ and its mutants by CaM.

**Figure 12.** Spectrofluorimetric titration of the $CaD_{136}$ and its mutants by CaM.



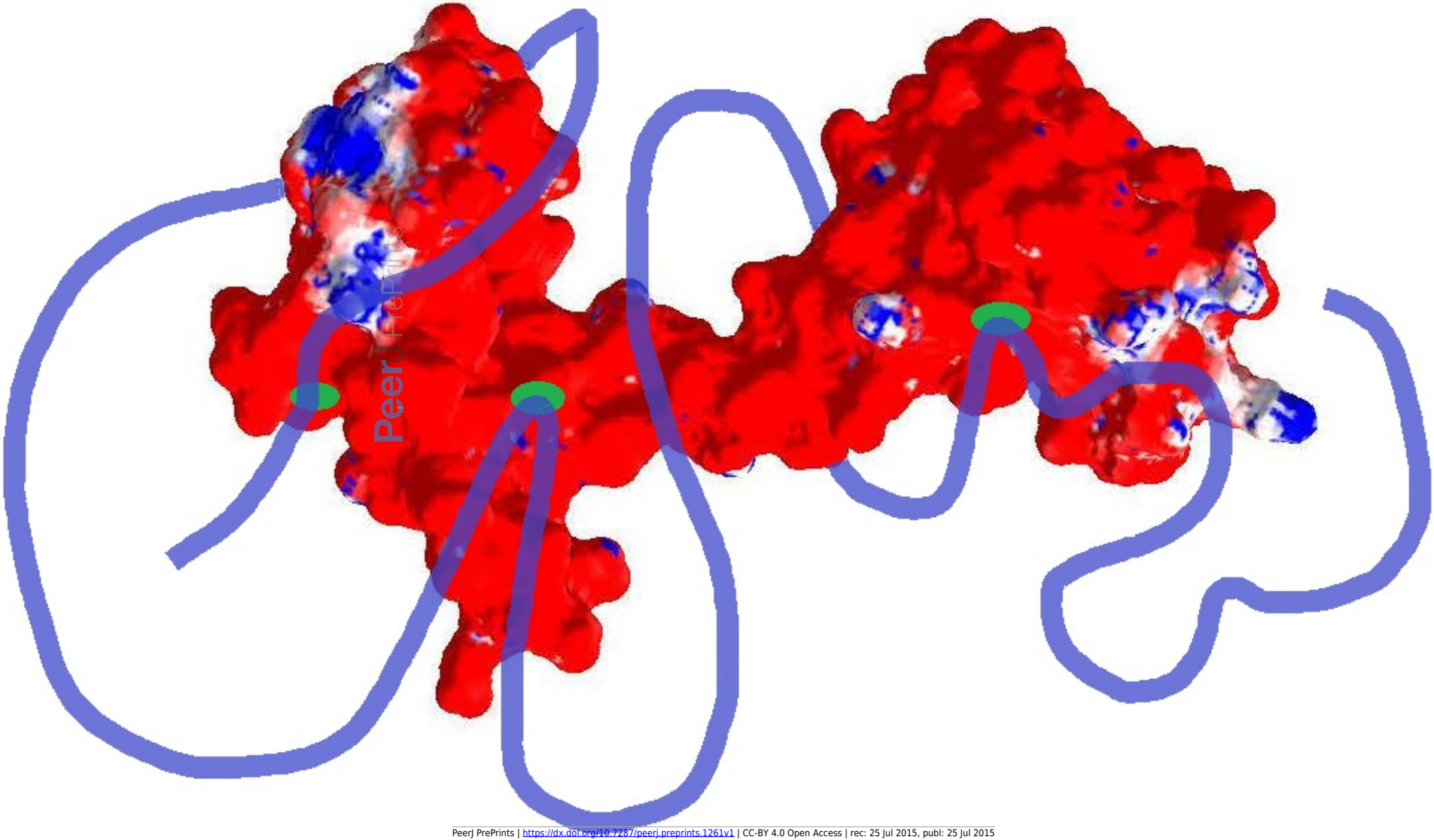



PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1261v1 | CC-BY 4.0 Open Access | rec: 25 Jul 2015, publ: 25 Jul 2015

## Figure 13 (on next page)

Model of CaD binding to CaM.

**Figure 13.** Schematic representation of the “buttons on a charge string” binding mode proposed in this study. Here, the $CaD_{136}$ is shown as a blue string containing three “buttons” (tryptophan-centric partially structured binding sites), whereas CaM is shown as mostly red surface. Note that positions of binding sites and length of the $CaD_{136}$ chain are arbitrary and used here only to illustrate an idea.





PeerJ PrePrints | https://dx.doi.org/10.7287/peerj.preprints.1261v1 | CC-BY 4.0 Open Access | rec: 25 Jul 2015, publ: 25 Jul 2015