Machine learning-based modulation of Ca²⁺-binding affinity in EF-hand proteins and comparative structural insights into site-specific cooperative binding

Mohit Mazumder^{1#}, Sanjeev Kumar^{1,2#}, Devbrat Kumar¹, Alok bhattacharya¹ and S. Gourinath^{1*}

¹School of Life Sciences, Jawaharlal Nehru University, New Delhi-110067

²Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, USA

Authors contributed equally

Abstract

Ca²⁺-binding proteins are present in almost all living organisms and different types display different levels of binding affinities for the cation. Here, we applied two new scoring schemes enabling the user to manipulate the binding affinities of such proteins. We specifically designed a unique EF-hand loop capable of binding calcium with high affinity by altering five residues of the loop based on the scoring scheme. We worked on the N-terminal domain of *Entamoeba histolytica* calcium-binding protein1 (NtEhCaBP1), and used site-directed mutagenesis to incorporate the designed loop sequence into the second EF hand motif of this protein. The binding isotherms calculated using ITC calorimetry showed a ~500-fold greater association constant (K_a) for the mutant. The crystal structure of the mutant was also determined, and displayed more compact Ca²⁺-coordination spheres in both of its EF loops than did the structure of the wildtype protein, consistent with the greater calcium-binding affinities of the mutant. The NtEhCaBP1 mutant was also shown to form a hexamer rather than just a trimer, and this hexamer formation was attributed to the position of the last helix of the mutant having been changed as a result of the strong calcium coordination. Further dynamic correlation analysis revealed that the mutation in the second EF loop changed the entire residue network of the monomer, resulting in a stronger coordination of Ca²⁺ even in other EF hand loop.

Keywords: Ca²⁺-binding affinity, EF-hand proteins, protein engineering, Isothermal calorimetry, X-ray crystallography, support vector machine, dynamic residue correlation.

Introduction

Ca^{2+} is a ubiquitous secondary messenger that controls various cellular processes like cell differentiation, muscle contraction, exocytosis, and apoptosis event [1, 2]. One of the well-studied Ca^{2+} binding proteins is Calmodulin (CaM), known to interact with many cellular proteins thereby performing various functions [3-5]. The CaM or CaM-like proteins has a conserved Ca^{2+} binding motifs referred as EF-hand motifs. Including CaM and other CaM like proteins, the number of EF-hand motifs varies across the different CaBPs [6-8]. The EF-hand containing proteins are abundantly present in the living organism and are classified more than 66 subfamilies and plays numerous roles in many crucial pathways [5].

The residues of EF-hand motifs are well conserved across the various species, and mostly by negatively charged amino acids. The coordination of amino acid residues of EF-hand motifs with Ca^{2+} occurred in a very conserved manner that forms a pentagonal bipyramidal geometry [5, 7, 9, 10]. In most of the Ca^{2+} binding proteins, including all CaMs, the residues in the Ca^{2+} binding loop at positions +X(1st residue of the loop), +Y(3rd residue of the loop), +Z(5th residue of the loop), -Y(7th residue of the loop), -Z (9th residue of the loop) and -X (12th residue of the loop) coordinate with Ca^{2+} directly, the 9th (-Z) position residue does not coordinate the Ca^{2+} directly, instead a water molecule involved in the coordination and 12th position residue contributes donates two electron and complete the coordination sphere [7, 9-12].

Binding of Ca^{2+} to CaM/CaM-like proteins, induces the conformational rearrangement which changes the energy landscape of these proteins, thereby interacts with many intracellular proteins [13]. The residue composition in the EF-hand binding loop is one of the major determinants, which dictate the Ca^{2+} binding affinity of all CaBPs; other factors such as the conformational cost upon Ca^{2+} binding, the EF- β -scaffold, cooperative binding, the physiological environment, also play crucial roles in the binding of the metal ion [5, 12, 14, 15]. These factors lead to a wide range of Ca^{2+} binding affinities, perhaps such diverse Ca^{2+} binding affinity exhibited by CaBPs may allow them to perform variety of functions.

Due to the importance of Ca^{2+} signaling, the Ca^{2+} binding proteins have been extensively studied; for example, various mutational studies have been reported on the Ca^{2+} -binding loops of CaM [16, 17]. Many site-specific binding affinity studies, using methods such as grafting the EF-hand loop

into different proteins such as CD2 proteins, have been conducted to understand the role EF-hand residues in Ca^{2+} binding [18-22].

Previously, we developed a method to identify and classify the Ca^{2+} -binding EF-hand loops by levels of their Ca^{2+} -binding affinities using support vector machine (SVM) based approaches operating through a classification-based scoring system [23]. The classifier uses a nonlinear SVM with a gaussian radial basis function kernel [24]. The program showed high accuracy for the prediction of Ca^{2+} binding motifs available from the literature but it only predicts the affinities qualitatively. Despite of huge implication of EF-hand motifs mediated Ca^{2+} signaling across the various species, there is no proper method has been developed yet which helps in manipulating Ca^{2+} binding affinity systematically. Although Ca^{2+} binding affinity can be manipulated by mutating critical residues of EF-hand motifs (1st, 3rd, 5th, 7th and 12th) of EF-hand motifs, but it often abolishes the Ca^{2+} binding affinity completely, which limits the study the function of CaBPs over a wide range of Ca^{2+} binding affinity.

To overcome this shortcoming in the field of Ca^{2+} signaling, we upgraded our previously available software in a way so that user can not only predict the Ca^{2+} binding affinity qualitatively rather one can manipulate the binding affinities *in-vitro* by mutating the EF-hand motif residues. For upgrading our current software, we extracted the margin distances for each prediction from our non-linear hyperplane which is the distance between the decision boundary and the support vectors and calculated the weighted position-specific scores (PSM) for each prediction [23]. With two additional parameters to further classify the EF-hand loops, we designed a unique EF-hand loop with predicted capability to bind Ca^{2+} with high affinity. The machine trained to identify Ca^{2+} binding sites & estimate affinity was further used to design and discriminate between two similar sites based on new scores.

To validate the binding abilities of the new designed site, we introduced the designed sequence in N-terminal domain of *Entamoeba histolytica* Ca^{2+} binding protein1, EF-hand loop 2 (Nt-EhCaBP1 EF-2 mutant) and we applied, biochemical, structural, and computational approach to validate our hypothesis. The binding affinity determined using ITC clearly shows Nt-EhCaBP1 EF-2 mutant forms of protein binds Ca^{2+} with higher affinity than Wild type-Nt-EhCaBP1 (Wt-Nt-EhCaBP1). Furthermore, we employed structural biology approach (X-ray crystallography) to understand the changes at the atomic level, that lead to changes in its Ca^{2+} binding affinity. Further, we used normal mode analysis (NMA) [25] to obtain insights leading into the coupled motions and

allosteric effects directly influencing the interactions in two Ca²⁺-binding sites of each protein. The analysis indicated a direct influence of the EF-hand loops on the central helix as a result of long and short-range interactions affecting the entire residue network and hence enhancing the allosteric interactions in high-affinity sites. Based on sequence and structural investigations followed by the network analysis, we derived the basis of allosteric cooperativity shown by different EF-hand Ca²⁺-binding loops.

A set of scripts with new scoring functions and a user-friendly webserver to predict, design and engineer EF-hand binding loops were developed. The webserver and the downloadable set of scripts are available at <http://sbl.jnu.ac.in/calb/index.html>.

Materials and Methods

Scoring system for each prediction

Previously, we developed a method to identify and classify Ca²⁺-binding EF-hand loops [23]. The classifier uses support vector machine (SVM) machine learning with C and Gamma parameters for specifically a nonlinear SVM with a Gaussian radial basis function (RBF) kernel [23]. To increase the efficiency of the classification, we extracted the margin distance between the decision boundary and the support vectors for each prediction.

In general, margin is referred to as the distance of the vector (x) which is the Euclidean distance of x from the separating hyperplane [26], is given by

$$\frac{w \cdot x + b}{\text{norm}(w)} ,$$

Where w = perpendicular to the hyperplane and b is, the parameter used to determine the offset of the hyperplane from the origin along the normal vector.

The distance to the origin was calculated as $\frac{|b|}{||w||}$. The margin (d (margin)) was calculated using the equation $d(\text{margin}) = m_{\text{positive}} + m_{\text{negative}}$. The output is a signed distance that is either negative/positive, depending on which side of the hyperplane the point x resides [27, 28].

PSSM based log likelihood scores (PSM_{LogL})

The conservation score based on PSSM (position specific scoring matrix) for each sequence submitted for classification was calculated by using the equation $G_{ij} = \log (S_{ij}/P_i)$. Where G_{ij} is the calculated PSSM score and B_{ij} is the frequency of amino acid residue i at position j in matrix S , and was calculated by using the equation

$$B_{ij} = q + \frac{bP_i}{n} + b \quad ,$$

where q denotes the observed counts of amino acid residue type i at position j , P_i is probability of amino acid residue type i , b is a pseudocount considered here as the square root of the total number of training sequences, and n is the number of training sequences. G_{ij} (PSSM score) represents the foreground model, which is the true homology and P_i represents the probability that a match occurs at random (background model) calculated using the BLOSUM62 substitution matrix [29]. In short, the PSSM-based scoring includes the relative frequencies obtained by counting the number of times each amino acid residue occurred at each position of the alignment, followed by normalization of the frequencies.

Design of the unique EF-loop site

We used SVM margin score (SVM_{Mar}) and the PSSM-based log likelihood score (PSM_{LogL}) values to design a unique Ca^{2+} -binding site, one does not present in any of the protein sequence databases [30, 31]. The design was developed in order to improve the predictions of the classifier as well as to understand the binding mechanism by deciphering the roles of Ca^{2+} -binding residues. The sequence design was carried out by substituting each of the 20 encoded amino acid residues in all possible 12 positions in the Ca^{2+} -binding loop; giving a unique score with every iteration. The sequence DKDGDGFIDFEE showed a high score and the proteome wide BlastP [30] showed the absence of the site in all known databases. Therefore, the 2nd EF-hand loop of NtEhCaBP1, i.e., DADGNGEIDQNE, was replaced with the DKDGDGFIDFEE sequence. This substitution was carried out using site-directed mutagenesis. Note that NtEhCaBP1 denotes the N-terminal domain of the EF-hand-containing Ca^{2+} -binding protein1, which has been well characterized in our laboratory [32-34]. The N-terminal domain has two Ca^{2+} -binding EF hand motifs. In order to construct the desired mutant, we incorporated 5-point mutations, namely A47K, N50D, E52F, Q55F and N56E, in the 2nd EF-loop of NtEhCaBP1.

We selected NtEhCaBP1 and the NtEhCaBP1 EF2 site as the model in this study since this set provided several advantages. 1) The sequence of the designed construct and that of the wild type were fairly similar so that relatively few single-site substitutions were required (Figure 1D). 2) We

used an optimized set of protocols and availability of biophysical and structural data on the N-terminal domain construct of the protein. 3) The NtEhCaBP1 protein construct with two known Ca²⁺-binding sites allowed us to investigate the effect of one binding site on the other and hence to potentially understand the mechanism of cooperative binding in this protein. 4) The designed mutant sequence was unique.

Cloning of the NtEhCaBP1 mutant

The gene fragments corresponding to the N-terminal domain of the EhCaBP1 protein were cloned by using an existing N-terminal clone of EhCaBP1 as a template in the bacterial expression vector pET28(b). The mutations were created (in NtEhCaBP1 EF-II) at positions 141, 150, 156, 165, and 168 by performing site-directed mutagenesis. The following primers were used for the mutation:

CaBP1 Mut K2 FP 5'-CAAATCTATTGATAAAGATGGAAATGG-3'

CaBP1 Mut K2 RP 5'-CCATTTCATCTTTATCAATAGATTTG-3'

CaBP1MutD5, F7 FP 5'-CTATTGATAAAGATGGAGATGGATTTATTGATCAAAATGAATTTGC-3'

CaBP1MutD5, F7 RP 5'-GCAAATTCATTTTGATCAATAAATCCATCTCCATCTTTATCAATAG-3'

CaBP1Mut F₁₀ FP 5'-ATTTATTGATTTTAATGAATTTGC-3'

CaBP1Mut F₁₀ RP 5'-GCAAATTCATTAATAAATCAATAAAT-3'

CaBP1Mut E₁₁ FP 5'-ATTTATTGATTTTGAAGAATTTGC-3'

CaBP1Mut E₁₁ RP 3'-GCAAATTCCTTCAAAATCAATAAAT-3'

CaBP1 FP -5'-CATGCCATGGCAATGGCTGAAGCACTTTTTTAAAG-3'

CaBP1 RP-5'-CGGCTCGAGGAGTGAAACTCAAGGAATTCTTC-3'

The mutations were confirmed by DNA sequencing.

Overexpression and purification of the NtEhCaBP1 EF-2 mutant

NtEhCaBP1 was expressed and purified as described in previously published literature [32]. The NtEhCaBP1 EF-2 mutant construct was transformed into *Escherichia coli* strain BL21 (DE3) for expression. Cells were grown in Luria-Bertani (LB) medium supplemented with 50 mg ml⁻¹ kanamycin at 37°C. The culture was induced with 0.8 mM IPTG when the OD reached 0.7 at A₆₀₀. It was then incubated at the same temperature for 3 h for further growth. Cells were harvested by subjecting by centrifugation at 7000 rpm for 10 min. The cell pellet was resuspended in suspension buffer (50 mM Tris pH 7.5, 2 mM EGTA). The cells were then lysed by freeze-thaw followed by sonication. A clear supernatant was obtained by carrying out centrifugation at 12000g for 30 min. The protein supernatant was passed through an ion-exchange chromatography using DEAE resin column, was pre-equilibrated with ten bed volumes of suspension buffer. The column was then

washed with 30-40 ml of wash buffer (50 mM Tris pH 7.5, 5 mM NaCl) to remove nonspecifically bound proteins. Finally, the protein was eluted with elution buffer (50 mM Tris pH 7.5, 5 mM CaCl₂). Further, an ion exchange purified protein was subjected to gel filtration chromatography using Superdex G-75 column pre-equilibrated in buffer containing, 50 mM Tris pH 7.5, 5 mM CaCl₂. The NtEhCaBP1-EF2 mutant protein peak was eluted at 65.93 ml (Figure 3.1). The molecular weight was calculated for observed peak using standard protein marker of different molecular weights, supplied by Sigma Aldrich. The purity of the protein was checked using 15% SDS-PAGE.

Isothermal titration calorimetry (ITC) to determine the Ca²⁺ binding affinity

Before performing ITC experiment both the wildtype (Wt-NtEhCaBP1) and mutant (NtEhCaBP1EF-2) proteins were prepared in Ca²⁺-free (apo form) form, as described previously [34] and the proteins were dialyzed against 10 mM of HEPES pH 7.4. ITC experiments were carried out using a Microcal ITC₂₀₀ instrument from GE Healthcare. Experiments were performed at 30°C in 10 mM HEPES buffer (pH 7.4) at 1 mM of CaCl₂ (titrant) concentration and 3.4 μM of the protein (analyte), either Wt-NtEhCaBP1 or the NtEhCaBP1EF-2 mutant were used. The volume of each titrant sample was 2 μl per injection and the fixed titrant pipette volume (40 μl) led us to perform 20 injections of titrant into a sample cell containing analyte (known concentration) as protein. The mixture was allowed to react for 2 min between injections and the heat release due to injection and dilution was obtained by titrating CaCl₂ into buffer containing protein. A raw thermogram was generated with rate of change of heat with respect to time by using Microcal iTC200 software. The data were fitted using the Origin software Microcal Analysis Launcher supplied by Microcal. The plot of heat change with respect to molar ratio of [ligand]/[protein] was derived from the raw thermogram by using Origin software Microcal Analysis Launcher. Iterations were performed until the chi square value did not decrease further and was stable. The data were fit by sequential binding mode which gave number of binding sites (N=2) for ligand (Ca²⁺), association constant (K_a), change in enthalpy (ΔH), and change in entropy (ΔS) values were obtained from the fitted data. Ligand to buffer titration was performed to check the background thermal noise of the buffer dilution. In order to decrease the thermal noise, a reference cell was filled with the same buffer used to prepare both the ligand and protein. The heat released and absorbed during titration was calculated by using the equation

$$q = V \Delta H \Delta [L_B],$$

where q = change in heat resulting from the change in the concentration of bound ligand,

$\Delta[L_B]$ = change in the concentration of bound ligand,

ΔH = change in enthalpy after ligand binding, and

V = reaction volume.

Crystallization of the NtEhCaBP1 EF-2 mutant

Prior to crystallization the SEC purified NtEhCaBP1 EF-2 mutant protein was concentrated to 15mg ml⁻¹ by using a 3kDa-cut-off centricon filter. Crystallization trials were performed by using hanging-drop vapour-diffusion method in 24-well plates. 2μl of protein sample was mixed with an equal volume of precipitant solution in hanging drops and equilibrated against 500 ml of reservoir solution (precipitant). Initially the same crystallization condition was tried in which Wt-NtEhCaBP1 was crystallized [32], but it did not yield crystals. Rather the NtEhCaBP1 EF-2 mutant was crystallized in conditions similar to the crystallization condition of full length EhCaBP1 [35]. The NtEhCaBP1 EF-2 mutant crystallize in the presence of MPD (58%- 63%), 50mM sodium acetate (pH 5.0-5.5), and 5mM CaCl₂.

X-ray diffraction, data collection, and processing and structure determination

NtEhCaBP1 EF-2 mutant crystals were soaked in cryoprotectant solution consisting of 65% MPD, 50 mM sodium acetate pH 5.3 and 5 mM CaCl₂. Single crystals were picked up in cryoloop and flash-cooled in liquid nitrogen. Data were collected at ESRF DBT-BM14, France. The crystals diffracted to a resolution of 1.9 Å. Diffraction data were processed and scaled using HKL2000 [36]. Crystal belonged to the space group P2₁2₁2₁, with unit cell parameters $a = 44.6$, $b = 101.3$, $c = 107.4$ Å. Matthews's coefficient, V_M , was 2.90 Å³Da⁻¹, suggested presence of six molecules in the asymmetric unit, with a solvent content of 57.5% [37]. Structure was determined by molecular replacement method using Phaser program [38] and the assembled trimer of wild-type structure of EhCaBP1 (PDB code 2NXQ) was used as a search model [35]. Overall, thirteen Ca²⁺ ions were identified in the electron density, two Ca²⁺ ions in each chain at the center of the EF-hand loop and one extra Ca²⁺ ion was observed which was not bound to an EF hand motif. All thirteen Ca²⁺ ions were included in the refinement. Structure was refined by REFMAC5 [39] and carrying out iterations of model building using the COOT graphics package [40]. For the final model, the R_{work} was 21.2 % and R_{free} was 24.9% and quality of structure using check using the program

PROCHECK [41] showed it to have good stereochemistry with 97.6% of residues lying in the most-favored regions of the Ramachandran plot. Structure factor and refined model of NtEhCaBP1 EF-2 mutant were deposited in the Protein Data Bank [42] with the accession code 5XOP. Data collection and final refinement statistics are shown in Table 1.

Table 1. Crystallographic parameters and data collection statistics for the NtEhCaBP1 EF-2 mutant.

X-ray source	ESRF BM14 Beamline
Wavelength (Å)	0.9788
Space Group	P2 ₁ 2 ₁ 2 ₁
Cell parameters (Å, °)	a = 44.6, b = 101.3, c = 107.4 $\alpha = 90, \beta = 90, \gamma = 90$
No. of molecules in an asymmetric unit	6
Resolution range (Å)	50-1.9 (1.93-1.9)
Total Reflections	301767 (15064)
Unique Reflections	39381(1966)
Completeness (%) ^{a,b}	99.9 (100)
Redundancy	7.7 (7.7)
Mean I/ σ	33.0 (1.9)
Rmerge (%) ^{b,c}	7.5 (88.8)
Refinement	
R _{work} (%)	21.2 (27.1)
R _{free} (%)	24.9 (33.6)
Mean B-factor	42.5
No. of atoms Protein/Ca/water/MPD	3128/13/174/1
Bond length (Å)	0.015
Bond angles (°)	1.71
Dihedral angle (°)	17.6
Ramachandran plot statistics	
Favoured region (%)	95.5
Additional allowed region (%)	4.2
Generously allowed regions (%)	0.3
Outliers	0.0
PDB code	5XOP

Structure and sequence analyses

We performed the sequence alignments using Clustal Omega and BioEdit [43, 44]. The structural alignment was performed using the Dali server [45]. Protein-protein interactions between the subunits were calculated using the *PDBsum* webserver [46] and Dimplot [47]. The Ca^{2+} coordination distances and angles were calculated using LIGPLOT [47] and the PLIP server [48] as well as using PyMol software [49]. Representations of the structure were prepared by using PyMol, Chimera [50] [45] and Photoshop software [51]. The individual residues involved in interactions of both the structures were analyzed by using the Arpeggio web server [52] and Hydrogen Bonds Computing Server (HBCS) [53]. The contact maps from the protein structures were calculated by using the Residue Interaction Network Generator (RING) web server [54].

Cross-correlation and dynamic network analysis

Network and dynamics cross-correlation map (DCCM) analyses were performed by using the Bio3D package [55]. Here, the proteins network systems were used to show insights into the structural relationships between three EF-hand-containing protein systems exhibiting different binding affinities. The dynamic networks were generated by using an ensemble of structures showing cross correlation between intramolecularly connected residues. We used normal mode analysis (NMA) [25] to obtain insights into the coupled motions and allosteric effects directly influencing the cooperativity of the two Ca^{2+} -binding sites in each protein. Furthermore, the cross-correlation matrix was calculated from the NMA, and a DCCM [55] was generated for three proteins using the Hessian matrix. C_{ij} values of the matrix close to 1 means the fluctuations of residues i and j are completely correlated (same period and same phase), a C_{ij} close to -1 suggests that the fluctuations of residues i and j are completely anti-correlated (same period and opposite phase), and a C_{ij} value equal to 0 indicates the fluctuations of i and j are not correlated. In the figure made, C_{ij} values are indicated in the right panel, and the cyan color signifies the positive correlation and pink shows the negative or anti correlation. Such a depiction allows for the identification of highly intramolecularly connected residues clustered as individual communities shown in the same color. The community clustering (clusters as communities in a graph), which deals with the graph partitioning problem, has been solved by the Girvan-Newman edge-betweenness approach which is implemented in Bio3D to generate the network [55]. The residues

belonging to the communities/clusters in the graph are also shown in different color codes along with position of each residue in the DCCM on the X axis and Y axis.

Results

Roles of EF-hand loop residues in Ca^{2+} binding

Mutations occurring in the sequences of many critical genes form the major driving mechanism leading to the evolution of complex organisms. These mutations have been seen in different protein families across all biological systems. But mutations are not prevalent in certain locations of proteins such as their structural cores, and the conserved nature of these residues are suggestive of their important functional roles [56]. Enhancing the ability of a protein to carry out a specific function requires an understanding of the functional domains and the roles of conserved residues. The EF-hand is a well-studied motif having many conserved residues, where the conservation is particularly strict at the Ca^{2+} -binding site. The well-studied high preference of certain amino acid residues at particular positions of the motif, such as aspartate at the beginning of the motif and glutamate at the end indicates that some ligands are indispensable for Ca^{2+} binding [17, 57]. These negatively charged residues placed at many positions in a short stretch of sequence makes an ideal site for the binding of a positively charged divalent ion such as Ca^{2+} or magnesium. In physiological conditions, both of these ions are present in the cell in high concentrations, yet the selectivity for the metal ion is highly specific in EF-hand proteins [58, 59]. Ca^{2+} -binding loops have evolved to reflect biological functions that require of the proteins a range of binding affinities [5]. In order to understand the role of individual residues in the Ca^{2+} -binding loop, we classified our data into two groups, namely high-affinity sites (HASs) and low-affinity sites (LASs) and aligned the sequences within each group to derive sequence motifs of the two groups (Figure 1). Figures 1A and 1B illustrate, for the two groups, the distributions of residues at each position of the EF-hand loop of the Ca^{2+} -binding site. The first position (X) of the loop is dominated by aspartic acid (D), known to be indispensable for Ca^{2+} binding due to its being involved in the critical hydrogen-bonding network and due to its providing an oxygen for coordinating the Ca^{2+} [60]. Some of the EF-hand Ca^{2+} -binding sites with a non-aspartate residue at the X position have shown low or even no affinity for Ca^{2+} . This observation has also been seen for numerous mutational studies of different EF-hand proteins [60]. Comparison of the high-affinity and low-affinity sequences (figure 1A, 1B) suggests that positively charged lysine in the second position

(non-interacting site) of the loop is present more often in HAS sequences than in LAS sequences. In this position of the LAS sequences showing the occurrence of a variety of amino acid residues such as alanine and isoleucine. In the third position (Y position), aspartic acid occurs most often followed by asparagine. The eighth position is mostly occupied by the hydrophobic residue isoleucine, valine and leucine. A close inspection of the sequences indicates a relatively high occurrence of serine at the 9th position of LAS sequences but not HAS sequences. In the twelfth position, glutamic acid is conserved in HAS sequences, not surprising since this residue provides a bidentate side-chain ligand for Ca²⁺ binding. In a few LAS sequences, aspartic acid is present in the twelfth position. Aspartic acid at the 12th position sometimes requires, because of the short side chain, one more water molecule to coordinate Ca²⁺, resulting in a lower affinity for the Ca²⁺ due to increased solvent entropy [60]. The analysis suggested that Ca²⁺-binding loop residues that do not directly coordinate the Ca²⁺ nevertheless have important roles for the binding of Ca²⁺, including stabilizing the loop, contributing to the extensive intricate hydrogen-bond network required for the proper folding of the loop, and properly positioning the negatively charged residues of the loop into the three-dimensional arrangement best suited for Ca²⁺ binding [60].

Scoring scheme to assist modeling of the EF-hand Ca²⁺ binding site

In an earlier study, we were able to classify EF-hand domains based on their affinities (high, low and none) for Ca²⁺ [61]. This classification was done on the basis of datasets trained in machine learning, specifically based on support vector machines (SVMs). The correlations derived from the use of various amino acid indexes showed high prediction accuracies when tested against many *EhCaBPs* [62]. In the present study, we extracted the margin distance from the decision boundary for each prediction, generated PSSM values, and incorporated the SVM_{Mar} and the PSM_{LogL} scores as added features to the previously developed method. Furthermore, we calculated the margin distances and conservation scores for all of the sequences of the three different previously published datasets consisting of 121 unique Ca²⁺-binding EF-hand loops. The details of the predictions for each site along with the new scores are shown in Appendix I. The scores indicated the possibility of classifying individual sites in an EF-hand protein on the basis of their distance from the decision boundary. The conservation score shows the probabilities of the residues occurring frequency.

Designing a high binding affinity EF-hand loop

So far, several mutational and design studies on EF-hand proteins have yielded distinctive impacts on their range of binding affinities, thus influencing the functions of those proteins [19, 23] [59]. In many of these studies, mutations were made to help assess the Ca^{2+} -binding capability of the protein. Substitutions in the first (X) and last (-X) positions of the binding site did yield desirable results, as these ligand positions are extremely important and require negatively charged chelating residues for proper function [57, 63]. In order to validate our prediction method, we designed a unique EF-hand loop, that is not presents in any protein sequence database (Figure 1C). The sequence design was executed by substituting every amino acid residue in each of the 12 positions of EF-hand motif. Each mutant was analyzed, and only those that increased the score were taken to the next iteration (see diagram for schematic representation). The designed binding loop with the amino acid residue sequence 1-DKDGDGFIDFEE-12 showed a high SVM_{Mar} score of 2.694 and PSM_{LogL} score of 6.46, “”, both comparable to those of the Ca^{2+} -binding sites with high affinity from the literature [23]. Sequence alignment performed with EF-hand containing proteins from different families (Figure 1C) suggested that the designed loop is most similar to the Ca^{2+} -binding pollen allergen [64], and similar to the structure of the *WtNtEhCaBP1* protein, which shows a head-to-tail arrangement with domain-swapped EF-hand pairing. We inserted the EF-hand loop sequence having the high predicted Ca^{2+} binding affinity into the well-characterized *WtNtEhCaBP1* Ca^{2+} -binding site to attempt to manipulate the Ca^{2+} binding affinity of this protein. We specifically mutated the sequence of the 2nd EF-hand motif of the Ca^{2+} -binding loop of *NtEhCaBP1*. The *NtEhCaBP1*-EF2 loop was selected because of the greater similarity between its sequence and that of the designed loop than between the sequences of the loops of other characterized Ca^{2+} -binding proteins of *E. histolytica* and that of the designed loop (Figure 1D). The crystal structure of the N-terminal construct of *NtEhCaBP1* has two Ca^{2+} -binding sites, while the full-length protein has four Ca^{2+} -binding EF-hand motifs [32, 35]. In a previous study, ITC experiments using full-length protein showed four binding sites with two sites displaying high Ca^{2+} -binding affinity and two sites with low binding affinity [65]. Calculations performed on the first Ca^{2+} -binding motif, that is the EF-1 (1st EF-loop-DVNGDGAVSYEE), yielded SVM_{Mar} and PSM_{LogL} scores of -1.045 and 4.976 and hence predicted EF-1 to have a relatively low binding affinity for Ca^{2+} . Calculations performed on the 2nd EF-loop, i.e., with the sequence DADGNGEIDQNE, yielded SVM_{Mar} and PSM_{LogL} scores of 1.001, and hence predicted the 2nd

EF-loop to have a Ca^{2+} -binding affinity lower than that of the first, i.e., EF-1, loop. To preserve the proper folding of the molecule the construct was built with two EF-hand loops. We were also interested in understanding how synthetic EF-hand loops influence cooperative binding in this system. In order to accomplish these aims and to understand the biophysical characteristics of the protein containing the engineered loop, we purified both WtNtEhCaBP1 and the NtEhCaBP1 mutant (EF2) and compared them by performing biophysical and computational experiments.

Figure 1. Sequence conservation and design of EF-hand loop sequence. A) Sequence logo of the residues forming high-affinity Ca^{2+} -binding EF-hand loops from the published dataset with experimentally known binding affinities. B) Sequence logo of the residues forming of low- Ca^{2+} affinity EF-hand loops from the published dataset with experimentally known binding affinities[62]. The alignment showed certain position of the EF-hand loop to be

relatively variable, but others to be quite conserved. C) The sequence alignment of the designed construct EF2 site with structurally closest homologs of EF-hand loops of NtEhCaBP1. D) The EF-hand loop sequences of the characterized and published constructs of EhCaBPs [31] [63][19] aligned with the designed peptide. This alignment showed the number of single-site substitutions (in pink) required to be performed in order to mutate the EF-hand loop to the designed construct.

NtEhCaBP1-EF2 mutant binds Ca^{2+} with higher affinity

To determine whether incorporating mutation in EF-2 loop of Nt-*Eh*CaBP1 (Nt-EhCaBP1-EF2 mutant) enhanced the Ca^{2+} -binding affinity, we performed the ITC experiments. Initially we performed the titration in presence of 1mM of CaCl_2 , but WtNt-EhCaBP1 did not shows the Ca^{2+} binding (data not shown) and curve did not achieve the saturation. In contrast, the binding of Ca^{2+} to Nt-*Eh*CaBP1-EF2 mutant is strong, and a well saturated curve obtained (Figure2B). To get better binding curve for WtNt-EhCaBP1, titration was performed with increasing concentration of Ca^{2+} and titration at 25mM of Ca^{2+} gave a better binding curve, with very low association constant than Nt-EhCaBP1-EF2 mutant form of protein (Table 2). The low association constant suggest that binding affinity is very weak in case of WtNt-EhCaBP1.

Figure 2 Isothermal titration calorimetric analysis of the binding of Ca^{2+} to the A) apo NtEhCaBP1 and B) apo NtEhCaBP1 EF-2 mutant. Plots are shown of kcal of heat absorbed/released per mole of injected CaCl_2 at 30°C. For

all titrations, the top panels represent the raw data (power versus time) and the bottom panels represent integrated binding isotherms. The solid line in each bottom panel represents the best nonlinear fit to the experimental data. The thermodynamic parameters obtained from these plots are summarized in Table 1.

The binding constant of Nt-EhCaBP1 EF-2 mutant is 45 and 77-fold higher than the WtNt-EhCaBP1. The WtNt-EhCaBP1 curve shows the exothermic process with a favorable change in enthalpy for site-1 ($\Delta H1$, -760.4) and unfavorable change in enthalpy ($\Delta H2$, 4857 kcal/mol) for site-2 (Table 1). The entropy at site-2 is very low (-7.38) which suggest the binding of Ca^{2+} at site-2 is not bringing any large structural or conformational changes.

The Nt-EhCaBP1 EF-2 mutant curve shows an exothermic process for site-1 with a favorable change in enthalpy ($\Delta H1$, -1866) whereas site-2 represent an endothermic process with unfavorable change in enthalpy ($\Delta H2$, 5699 kcal/mol) and comparatively higher entropy than the site-1 (Table 1). The unfavorable enthalpy and higher entropy indicate that Ca^{2+} binding at site 2 undergoes large conformational change as compared to site 1. Usually, the Ca^{2+} binding protein shows the cooperativity when their EF-hands bind with Ca^{2+} ion i.e., binding of metal ion at one site induced the binding at second site. As we observed that NtEhCaBP1 is unable to bind Ca^{2+} or has weak binding affinity but in case of Nt-EhCaBP1-EF2, mutation induced the binding of Ca^{2+} , therefore a μM range binding affinity obtained at both sites. This difference in binding affinity in both form of protein suggests that Nt-EhCaBP1 EF-2 mutant shows the phenomenon of cooperativity, which leads to obtain a better binding curve in comparison with Nt-EhCaBP1.

Protein	Site name	K_a (M^{-1})	ΔH (cal/mol)	ΔG (cal/mol)	ΔS (cal/mol/deg)
Nt-EhCaBP1	Site-1	$K1=1.05 \times 10^3 \pm 45$	-760.4 ± 21.1	4186 ± 21.1	11.3
	Site-2	$K2=77.3 \pm 5.8$	4857 ± 322	7094 ± 322	-7.38
NtEhCaBP1-EF2 mutant	Site-1	$K1= (4.80 \times 10^4 \pm 10^3)$	-1866 ± 120	-6504 ± 120	15.3
	Site-2	$K2= (1.44 \times 10^3 \pm 10^2)$	5699 ± 1900	-4396 ± 1900	33.3

Table 1. Summary of binding constants and thermodynamic parameters obtained from the Ca^{2+} -binding isotherms of ITC studies of NtEhCaBP1 and Nt-EhCaBP1 EF-2 mutant at 25°C.

Solution state oligomerization and three-dimensional structure of NtEhCaBP1 EF-2 mutant

In our previous reports it has been shown that either full length EhCaBP1 or its C-terminal deleted mutant (NtEhCaBP1) form of protein always forms a trimer in solution or in crystal form [32, 35]. In contrast to our previous observation, we found that while purifying NtEhCaBP1 EF-2 mutant protein differed in oligomeric state with respect to Nt-EhCaBP1. The size exclusion profile (SEC) of NtEhCaBP1 EF-2 mutant protein shows a peak correspond to molecular mass of about 42-44 kilodalton (kDa). As, the molecular weight of NtEhCaBP1 is 7.0 kDa, the SEC profile suggest the formation of hexamer in solution (Figure 3A).

Consistent with the solution state hexameric form, crystal structure also shows presence of six molecule in the asymmetric unit, and it is arranged in the form of two trimer, which leads to hexamer formation. The structure is well refined, and a representative electron density map of EF-1 and EF-2 are shown in (Figure 3B & 3C). Structural analysis suggested that each trimer in Nt-EhCaBP1 EF-2 mutant structure adopted similar trimeric arrangement as formed in case of Wt-NtEhCaBP1 structure [32] (Figure 3D).

Figure 3. A) Size exclusion chromatography (SEC) profile of the Nt-EhCaBP1-EF2 mutant) The SEC peak of Nt-EhCaBP1-EF2 mutant obtained from Superdex G-75 columns, shows the peak at elution volume of 64 mL, which

corresponding to molecular weight of around 39kDa. Considering the molecular weight of NtEhCaBP1-EF2 monomer is approximately 7kDa, the eluted peak at 63ml indicates the formation of hexamer in solution. B, C) Electron density maps and structural models of the EF-1 and EF-2 Ca^{2+} -binding loops of NtEhCaBP1 EF-2 mutant with bound Ca^{2+} . The images specifically show the portions of the 2Fo-Fc electron density maps, each at a 1.0 σ cutoff, corresponding to the residues of the N-terminal (EF1-orange) and C-terminal (EF2-yellow) Ca^{2+} -binding loops. All twelve amino acid residues (labeled) comprising each binding loop are represented in stick form inside each electron density map. D) The structure of the NtEhCaBP1-EF2 mutant forms hexamer, assembled via two trimers.

In the hexameric state, one trimer (interface-A) of the NtEhCaBP1-EF2 mutant interact with the other trimer (interface-B) (Figure 4D) where chain A was observed to interact with chain D (Figure S1) in the same manner chain B interact with chain F and chain C with chain E. The residues present on the binding interface of both the trimers forms multiple hydrogen bonds and non-bonded interactions. The oxygen atom from the C-terminal residues Ile-65, Ser-64 and Ile-65 are involved in hydrogen-bond interactions (Table 2) with nitrogen atom of glutamate-36 of the central helix (Figure 4E). This interaction brings the both the trimer in close proximity thereby leading to hexamer formation, in solution as well as in crystal form.

Trimer 1		Trimer 2		Distances
GLN-36 (NE2)	Chain A	Ile-65 (O)	Chain D	3.06
GLN-36 (NE2)	Chain B	Ser-64 (O)	Chain F	3.36
GLN-36 (NE2)	Chain C	Ile-65 (O)	Chain E	3.28

Table 2 The hydrogen bonds formed between the trimer1 (Chain: A, B, C) and trimer2 (Chain: D, E, F).

The interactions between the three assembled domains in each trimer of Nt-EhCaBP1-EF2 mutant structure were found to be similar as it was observed in case of Wt-NtEhCaBP1 structure.

There are some additional hydrogen bonding interactions were found in hexamer of NtEhCaBP1-EF2 mutant structure. The various interactions involved in hexameric state of Nt-EhCaBP1-EF2 structure are shown in supplementary figures S1 and S2 (Table S4 -S5).

An extra Ca^{2+} ion bound outside of EF-hand motif

Apart from the usual Ca^{2+} ions occupied at each EF-hand motifs, an extra Ca^{2+} ion was observed at the 2nd EF-hand motif of chain B. This extra ion occupied space outside of the regular Ca^{2+} -

binding coordination sphere. In order to understand the binding of two Ca^{2+} ions to one EF-hand motif, we calculated and plotted all the atomic interactions involving each of the two Ca^{2+} ions (Figure 4A). The first Ca^{2+} ion was observed to interact with X, Y, Z, -Y, -Z, and -X, including specifically with one of the side-chain oxygens (OD1) of each of the aspartate D-48 (Y) and D-50 (Z), to form a pentagonal bipyramidal geometry like conventional EF-hand loop. Whereas the second Ca^{2+} ion is bound outside of EF-hand motif and interacts with the side-chain oxygen atom (OD2) of D-48, D-50 and two water molecules. Most likely these two water molecules stabilize this extra Ca^{2+} ion (Figure 4A).

We also compared the EF2 loop of chain B with the EF2 loop of chain E (Figure 4B) including all the well-ordered water molecules within 5 Å of each loop. This analysis clearly indicated that the binding of the extra calcium to the EF2 loop of chain B was associated within the proximity of two water molecules to this calcium. In chain B the two water molecules interacted with the extra calcium at typical distances 2.30 Å. The occurrence of clusters of water or water-mediated hydrogen-bonding interactions have been seen in a greater percentage of non-EF-hand calcium-binding proteins (33.1%) than EF hand-containing proteins (13.1%). Our current comparison of the EF-2 loops of the different chains of the *NtEhCaBP1*-EF2 mutant suggested that binding of the additional calcium in chain B was not due to any additional structural changes but due to two close water-mediated interactions with the Ca^{2+} ion stabilizing the binding of Ca^{2+} outside of the EF-hand motif.

Figure 4 Two Ca^{2+} ions bound to the designed EF-2 of chain B of the NtEhCaBP1-EF2 mutant. A) The modified EF-2 loop of chain B of the NtEhCaBP1-mutant structure shown in a cartoon and lines representation and showing the locations of the two bound Ca^{2+} ions. The 2nd Ca^{2+} ion positioned outside of the Ca^{2+} -binding coordination sphere (colored in yellow). Both Ca^{2+} ions are separated with a distance of 5 Å, 2nd Ca^{2+} ion interact with oxygen atom of Asp48, Asp50 and two water molecules, both are located at a distance of 2.30 Å from Ca^{2+} ion.

Mutation induced bend in the third helix (H3) of NtEhCaBP1 EF-2 structure

To investigate whether the mutation induced any structural changes in the overall structure or in particular EF-hand motif, we superimposed the NtEhCaBP1 EF-2 mutant structure on Wt-NtEhCaBP1 by using PyMol. An alignment of all 66 C α atoms yielded an RMSD of 0.96 Å. Structural superimposition shows the difference in the orientation of the third α -helix (i.e., H3) (Figure 5A). Distance measurement between second helix and the C-terminal part of third helix indicates that NtEhCaBP1 EF-2 mutant structure acquire 76.3 Å whereas the wildtype structure shows the distance of 83.4 Å, suggesting the occurrence of a distinctive structural change due to this mutation of the Ca^{2+} -binding sites. We calculated this difference in the orientation of the third helix by measuring the angle between CD2 atom of residue Leu-36, CD1 atom of residue Ile-44, and the OE-1 atom of residue Gln-65 (Figures 5B & 5C). This measurement suggests a difference of about 7.1°, between the orientation of the third helix with respect to helix-2, in the mutant structure (Figure 5C).

Hydrophobic interactions leading to positive cooperativity

Hydrophobic interactions are one of the driving forces in EF hand proteins due to their role in the cooperative binding of the two EF hand motifs to each other [13, 17, 21]. The locations of hydrophobic residues in the NtEhCaBP1 structure and in the NtEhCaBP1 mutant structure

interacting are highlighted in green (Figure 5E and 5F), where the four-helices forming the hydrophobic core are tightly associated with each other. These interactions mostly involved phenylalanine and tyrosine residues, which appeared to play an important role in stabilizing the EF-hand loop by forming interactions between one EF hand loop and the other (Figure 5E and 5F). The interactions were aided by the formation of an antiparallel EF β -scaffold (Figure 5E), which helps the loop to close while folding without causing large structural change[64]. The stronger allosteric interactions between the odd-numbered and even-numbered EF loops were found to lead to a greater calcium binding affinity; in the NtEhCaBP1 mutant, these interactions appeared to be further enhanced due to the presence of two additional aromatic residues (E52F and Q55F). The observation of conserved hydrophobic residues at the 6th and 22nd positions of EF hand motifs suggested the communication between the two EF hands to be chiefly mediated by the noncovalent interactions.

Figure 5 Superimposition of the WtNtEhCaBP1 and NtEhCaBP1 mutant structures. A) The structure of WtNtEhCaBP1 is shown in violet and that of the NtEhCaBP1 mutant in pink. B) & C) Two additional views of this superposition, with these views showing highlighting the different angles between the C-terminal helix and central helix for the two proteins. The change of the orientation of the C-terminal helix is shown with a curved red arrow. D) E) & F) Detailed views of the EF-hand motifs showing conserved hydrophobic residues (F-61, F-58, F-55, Y-62, Y-19, F-25, F-6, Y-6, F-42, F-25) highlighted in green.

Trimer 1 (chains A, B, C) is shown in surface representation and trimer 2 (D, E, F) is shown in cartoon representation forming hexamer. The c-terminal residues are shown at the binding interface of protein-protein interactions. E) The interactions of the EF-1 and EF-2 motifs from two subunits of the NtEhCaBP1 mutant were found to mimic the pairing required for the proper functioning of the EF-hand proteins [53]. The hydrophobic residues are labeled and the residues in the EF-2 loop are shown in green. F) Similarly, the two chains from the trimer of NtEhCaBP1 is shown with residues forming the hydrophobic core.

Changes in overall charge distribution

To understand the differences between the charge distribution in the mutated loop and that in the wildtype loop, we calculated the overall surface charge distribution of the NtEhCaBP1 mutant and that of WtNtEhCaBP1 (Supplementary Figure S3) using ABPS plugin in PyMol. The analysis indicated a similar charge distribution in the calcium-binding loop of the mutant as in that of WtNtEhCaBP1, except for the replacement of wildtype negatively charged glutamate at position 52 with hydrophobic phenylalanine. We calculated the number of oxygen atoms available to the mutant and wild type EF-2 and found more oxygen atoms in the wild type than in the mutant protein (Figure S3). This result suggested that increasing the binding affinity for Ca^{2+} does not necessarily depend upon introducing more negatively charged atoms but in creating an optimum distribution of charges where the stabilizing hydrophobic interactions overcome the destabilizing electrostatic interactions caused by the negatively charged residues in close proximity [66].

Comparison of EF-hand-2 of Nt-EhCaBP1-EF2 mutant structure with Wt-NtEhCaBP1

To understand if enhanced Ca^{2+} binding observed in ITC data of Nt-EhCaBP1-EF2 mutant protein, contributed any changes in Ca^{2+} coordination sphere we compared the distances and angles of the bound Ca^{2+} with their respective interacting residues of the EF-2 loop of the mutant protein with the EF-2 of Wt-NtEhCaBP1(PDB code 2NXQ). The Ca^{2+} coordination distance (d1) acquired by various residues in Nt-EhCaBP1-EF2 mutant structure were comparatively small with respect to the Ca^{2+} coordinating distance (d2) acquired by respective residues of Wt-NtEhCaBP1 structure (Table 3). Compared to Wt-NtEhCaBP1 structure, each Ca^{2+} coordinating residues of Nt-EhCaBP1-EF2 mutant structure shows the difference of $\sim 0.1\text{\AA}$ - $0.5\text{\AA}$, suggesting that reduced distance in Ca^{2+} of Nt-EhCaBP1-EF2 structure attribute the high Ca^{2+} binding affinity.

Also, the oxygen-water- Ca^{2+} angles (θ_1), of Nt-EhCaBP1-EF2 mutant structure shows the difference, for example Asp-46, interacting from the top of the pyramidal geometry showed an

HOH-CA-OD1 angle of 172.6° compared to 158° for NtEhCaBP1-EF2 (Figure 6 A & 6B). The other coordinating contacts showed a variation of ~15° of angles (80° to 95°) amongst themselves (Table 3). These differences in bond lengths and angles between the mutant and wildtype structures may have caused the difference in ITC results, where the NtEhCaBP1-EF2 mutant protein sites showed higher binding affinities for Ca²⁺.

Residue	Position	Angle (Θ)	AA	Distance (d1)	Angle (Θ1)	AA	Distance (d2)	Angle (Θ2)
Binding Site (EF-2)			NtEhCaBP1-mutant			WtNtEhCaBP1		
46	X	OD1- HOH-CA	ASP	2.34	172.6	ASP	2.64	158
48	Y	OD1-HOH-CA	ASP	2.4	95.7	ASP	2.84	102.4
50	Z	OD1-HOH-CA	ASP	2.22	84.7	ASN	2.70	73.6
52	-Y	O-HOH-CA	PHE	2.27	97	GLU	2.56	105.3
57	-Z	OE1-HOH-CA	GLU	2.66	95.6	GLU	2.7	109.1
57		OE2-HOH-CA	GLU	2.5	79.5	GLU	2.4	115.2
water	-X	NA	HOH	2.32	NA	HOH	2.92	NA

Table 3 List of the Ca²⁺-binding ligands at the conserved positions and water molecules of the NtEhCaBP1-mutant EF-2 loop and WtNtEhCaBP1 EF2-loop as shown in the figure 5 (A and B) along with the distances (Å) between the interacting atoms/ residues and the angles (Θ) with respect to the water molecule. Specifically listed are residues of EF-2 of chain A interacting with Ca²⁺ through their coordinating side-chain oxygens.

The EF-1 loop of the NtEhCaBP1-EF2 mutant also shows reduced bond distance and bond angle

Although, we could not change any residues of EF-1 of NtEhCaBP1-EF2 mutant protein, but a comparative analysis of the Ca²⁺ coordinating distances of EF-1 of NtEhCaBP1-EF2 mutant structure also shows shrinkage in the Ca²⁺ coordinating sphere with respect to EF-1 of Wt-NtEhCaBP1 structure (Table 4) (Figure 6C and 6D). The OD1 atom of Asp-10, the first coordinating residue, was found to be positioned at the same site in both structures, but closer to the Ca²⁺ ion in the NtEhCaBP1-EF2 mutant than Wt-NtEhCaBP1 structure. Along with other residues which are involved in the Ca²⁺ coordination sphere, the 12th position Glu-21 ligand was also observed to be closer to the Ca²⁺ ion in the mutant than in WtNtEhCaBP1 (Table 4).

Residue	AA	Angle (Θ)	Position	Distance (d1)	Angle (Θ 1)	Distance (d2)	Angle (Θ 2)
	Binding Site (EF-I)			<i>Eh</i>CaBP1-mutant EF-I		<i>Eh</i>CaBP1 EF I	
10	ASP	OD1-HOH-CA	X	2.11	166.6	2.53	164.1
12	ASN	OD1-HOH-CA	Y	2.4	89.9	2.80	104.1
14	ASP	OD1-HOH-CA	Z	2.58	84.4	2.85	83.5
16	ALA	O-HOH-CA	-Y	2.25	96	2.43	89
21	GLU	OE1-HOH-CA	-Z	2.55	90.6	2.70	86.3
21	GLU	OE2-HOH-CA		2.46	93.5	2.70	107.6
Water	HOH	NA	-X	2.19	NA	2.80	NA

Table 4 List of the Ca^{2+} -binding ligands at the conserved positions and water molecules of the Nt*Eh*CaBP1-mutant EF-1 loop and WtNt*Eh*CaBP1 EF-1-loop as shown in the figure 5 (C and D) along with the distances ($\text{\AA}$) between the interacting atoms/ residues to the bound Ca^{2+} ion and the angles (Θ) with respect to the water molecule. Specifically listed are residues of EF-1 of chain A interacting with Ca^{2+} through their coordinating side-chain oxygens.

To overcome the problem of having small dataset, we included all six chains from two other reported structure of *Eh*CaBP1 (3LI6 and 2NXQ) and six chains of Nt*Eh*CaBP1-EF2 mutant structure (5XOP) for distance comparison of each Ca^{2+} ligating oxygen in EF-hand motifs (Figure 6E). The average Ca^{2+} coordination distance of Nt*Eh*CaBP1-EF2 mutant structure shows the shortest coordination distance for EF-2 than EF-1 and longest in EF1-of Wt*Eh*CaBP1 followed by EF-2 (Listings of the coordinating residues and distances for each of the individual chains from each structure are shown in Supplementary tables S7-S13).

Figure 6. The arrangements of residues in the EF-hand Ca²⁺-bound sites of NtEhCaBP1-mutant and WtNtEhCaBP1. A) The Ca²⁺-binding EF-2 site of the designed EhCaBP1-EF-2 mutant structure, showing the close contacts of the known ligands at the critical Ca²⁺-binding positions (X, Y, Z, -Y, -Z, -X) with the Ca²⁺ ion. The coordinating water molecule required for binding Ca²⁺ with a proper binding affinity was used to calculate various angles inside the cavity and to estimate the compactness of the site. B) The Ca²⁺-binding active site in the EF-2 loop of the WtEhCaBP1 protein. C) The EF1 loop of the NtEhCaBP1-EF2 mutant structure. D) The Ca²⁺-binding active site in the EF1 loop of the WtNtEhCaBP1 structure. E) The distributions of distances from the oxygen atoms of the

Ca²⁺-interacting residues to the Ca²⁺ ion for each of the two EF-hand sites, i.e., EF-1 and EF-2, of each of the two proteins, i.e., the WtNtEhCaBP1 (blue) and NtEhCaBP1-mutant (red) proteins. For each of the four cases, these distances were calculated by averaging the distances of all the chains of the crystal structure. EF-1-Wt stands for the 1st EF hand of the wildtype NtEhCaBP1, etc.

Cooperative binding enhances Ca²⁺ binding capability of the EF-1 loop of the mutant protein

EF-hand motifs mostly occur in pairs, and the Ca²⁺-binding mechanism and the resulting affinity also rely on how well these two motifs influence each other's presence by interacting with the residues at the interface of Ca²⁺-binding sites. The non-covalent interactions play very important role in the stabilization of a pair of EF-hand loops [61, 64]. The communication between the two EF-hands in our structure was observed to be mediated by the non-covalent interactions involving conserved hydrophobic residues. We used the Residue Interaction Network Generator (RING) to identify covalent and non-covalent bonds, including π - π stacking and π -cation interactions, between the EF-1 and EF-2 motifs in both the NtEhCaBP1 and NtEhCaBP1-EF2 mutant structures (Figure 7). The NtEhCaBP1 structure showed numerous such interactions, including both at the C-terminal helix where F-58 and Y-62 formed π - π stacking interactions and at the binding interface where I-53 formed a hydrogen bond with V-17 (Figure 7A). In addition to a hydrogen bond between I-53 and V-17 as seen in the wildtype protein, an additional van der Waals interaction was observed between F-42 and I-45 (Figure 7B). The non-bonded interactions in the mutant were further enhanced by the incorporation of two additional phenylalanine (E52F and Q55F) in site 2. The analysis suggested that the more extensive hydrophobic interactions in the mutant helped its EF1 loop to move closer by 0.2 Å towards EF2 loop (supplementary figure S6). The detailed structural analysis using protein network nodes and edges as interactions suggested that the incorporation of the ring structures (of the aromatic residue) allowed the loops to interact through pi-pi interactions mediated via F-52 and F-55 and several aromatic-amide π interactions (Supplementary figure S6 D&E) and also increased the polarizability due to the face-to-face and edge-to-face interactions of aromatic rings exhibiting strong van der Waals interactions in the NtEhCaBP1 mutant.

Insights from the high Ca²⁺ affinity of the calmodulin Ca²⁺-binding loop

In order to gain further insight into the different cooperative binding mechanisms exhibited by different pairs of EF-hand loops, we analyzed a high-Ca²⁺-affinity rat CaM (3SJQ) structure as a

third model for comparison., since this CaM was found to be the closest CaM structural homolog (supplementary table S5) of the *NtEhCaBP1* mutant. The CaM EF-1 (DKDGDGTITTKE) and EF-2 (DADGNGTIDFPE) shows SVM_{Mar} values of 0.93 and 1.87, respectively, and PSM_{LogL} scores of 5.58 and 6.39, respectively, hence both the sites were predicted a high affinity Ca²⁺ binder. The hydrophobic core formed by the interactions of two EF-hand motifs of CaM and *NtEhCaBP1* EF-2 mutant were superimposed to understand what makes a strong Ca²⁺ binder. Surprisingly, the alignment showed a low RMSD of 0.1-0.2 Å. In CaM structure, the two Ca²⁺ ions of the two EF-hand motifs were 11.7Å away from each other (Figure 7C), compared to 11.8 Å for the *NtEhCaBP1* mutant and 12Å for *WtNtEhCaBP1*. Comparison of the contact map of CaM (Figure 7D) with *NtEhCaBP1* EF-2 mutant shows the presence of more hydrophobic residues near the EF-1 motif of the high-affinity site (HAS) than low-affinity site (LAS). Analysis of the protein-protein interaction site in the *NtEhCaBP1* EF-2 mutant shows eight aromatic residues compared to seven in CaM, i.e., with this difference due to the presence of Phe-52 in the *NtEhCaBP1* EF-2 mutant but Thr-52 in CaM (Figure 7C). The close proximities of the two Ca²⁺-binding loops at the EF1-EF2 interface (Figure 7C) of CaM and *NtEhCaBP1* EF-2 mutant suggested a stronger communication. The formation of an EFβ-scaffold in CaM and in the *Nt-EhCaBP1* EF-2 mutant apparently allows the sites to interact with each other through changes in the hydrogen bond network and increases the stability which is a major factor for high intrinsic binding affinity for Ca²⁺ [61].

Figure 7 Structural basis of cooperativity of Ca^{2+} binding of WtNtEhCaBP1(LAS), the NtEhCaBP1 mutant and CaM (HAS). Contact maps of A) WtNtEhCaBP1 and B) the NtEhCaBP1 mutant where the residues at the interface between EF-1 and EF-2 are colored according to their physico-chemical properties: the hydrophobic amino acid residues are shown in green, positively charged residues in blue, negatively charged residues in red, polar residues in pink, and small residues such as alanine and glycine in yellow. The interactions are colored according to the type of bonding: hydrogen bonds are shown in cyan, van der Waals interactions in yellow, ionic interactions in blue, π - π stacking in red, and the metal-ion-mediated interactions in grey. The hydrophobic amino acid residues and the residues participating in π - π stacking are labelled. C) The two Ca^{2+} -binding sites in the CaM (pink) and NtEhCaBP1 mutant (red) proteins, showing the positions of aromatic residues (shown in green color: NtEhCaBP1 mutant) mediating hydrophobic and van der Waals interactions. The residues forming the Ca^{2+} -binding loop are labeled in both structures and the distance between the two- Ca^{2+} binding sites in CaM is also shown. D) The ionic interactions (blue), hydrogen bonds (cyan), and van der Waals interactions of the residues at the interface between EF-1 and EF-2.

Dynamically coupled networks and correlated motions in EF-hand proteins

One of the factors that has been not yet been well studied in Ca^{2+} -binding proteins is that how the motion of one domain of the protein influences that of the other and how the residues of these domains are networked together by long-range interactions so that they can manipulate each other's functions by undergoing small conformational changes. During the comparative analysis throughout this study, we noticed dissimilarities of many side chain interactions, with these dissimilarities due to different orientations of the side chains (Supplementary Figure S4). The three model systems (NtEhCaBP1, the NtEhCaBP1 EF-2 mutant, and CaM) comprising six Ca^{2+} -binding EF-hand motifs with different Ca^{2+} -binding affinities were further analyzed to understand the role of the all the residues in influencing each other's correlative movement. We calculated the dynamic cross-correlation (DCC) map and the dynamic protein residue network to understand the correlations amongst the residues in both a low- Ca^{2+} -affinity protein (NtEhCaBP1) and high- Ca^{2+} -affinity protein (NtEhCaBP1 mutant). Here, the protein structures were modeled as network systems by translating amino acid residues into nodes and both long- and short-range interactions into links (Figure 8). Comparing the full network of residues of NtEhCaBP1 (a monomer including EF1 and EF2), NtEhCaBP1 EF-2 mutant (a monomer including EF1 and EF2) and high Ca^{2+} affinity rat CaM (a monomer including EF2 and EF3). Interestingly this comparison revealed a pattern of communication leading to different correlated motions for the three proteins (Figures 8A, B and C). The community clustering map generated using the Girvan-Newman method showed the EF-hand motifs to be highly intramolecularly connected proteins with residues clustered as communities and numbered as clustered in the large networks. The NtEhCaBP1 with 65 residues, NtEhCaBP1 mutant with 66 residues, and N-terminal domain of CaM with 69 residues form the same set of nodes. NtEhCaBP1 forms 1527 edges, considerably more than the 1387 edges of the NtEhCaBP1 mutant and the 1397 edges of rat CaM. Both of the structures (i.e., NtEhCaBP1 mutant and CaM) with high- Ca^{2+} -affinity binding sites showed a more compact structure as the clustering of intramolecularly connected communities resulted in fewer community nodes and edges than in WtNtEhCaBP1 (supplementary table 11). The full network of NtEhCaBP1 formed the largest community3 (grey), comprising 23 residues (Figure 8A), whereas the high- Ca^{2+} -affinity NtEhCaBP1 mutant formed a massive community 2 comprising 48 residues, quite similar to the CaM residue network that showed the largest community 2 with 38 residues. The differences in the communication between the clustered communities and the differences in the number of edges

(connections) in the two very similar systems (NtEhCaBP1-mutant and NtEhCaBP1), i.e., differing by only 5 residues, was quite surprising. Six residues (Figure 8A) from the EF loops of NtEhCaBP1 were found to participate in the largest community (grey color) cluster compared to 13 residues in the NtEhCaBP1-mutant and 11 residues in CaM (community 2). The largest communities in each of the EF-hand motifs were determined to consist of residues mainly from the central helix. But the pattern varied in the high-affinity motifs, where the residues in the largest community were found to be distributed throughout the structure especially in the designed mutant (Figure 8B), all of whose C-terminal residues were observed to be part of the largest connected community. The high-Ca²⁺-affinity sites of CaM and NtEhCaBP1-mutant showed a high number of residues clustered together (Supplementary table S11). The determination of the number of nodes (residues) participating from individual domains shed light on the extent of the correlations amongst these 3 multiple EF-hand containing protein constructs. The N-terminal helix of NtEhCaBP1 showed residues connected to the largest community (grey), with the central helix having 13 residues and the C-terminal helix interconnected via 7 residues. The analysis of the high-affinity CaM and NtEhCaBP1-mutant proteins showed their N-terminal helices to have 13 and 12 residues connected to the largest community (red), their central helices to have 15 and 20 residues, and their C-terminal helices to have 9 and 15 residues, respectively. The comparison of these intra connected residues clearly suggested the extent of the direct correlation between the EF-loop site residues is higher in CaM than NtEhCaBP1, and highest in NtEhCaBP1-mutant.

The dynamic correlation motions of the residues in EF-hand proteins plotted as DCC maps (figure 8) showed strong positive and negative correlation. Both WtNtEhCaBP1 and NtEhCaBP1-mutant showed positive correlation in the EF1 loop, in contrast to that in the EF2 loop (Figure 8). The pattern of correlation shown by the EF-hand proteins suggested a direct correlation between the two Ca²⁺-binding sites as shown by the red arrow in the DCC map (Figure 8). The yellow arrow shows the negative correlation between the central helix and the EF loops (Figure 8). In summary, the dynamic correlation and the clustering analysis suggested that the coordination between two EF-hand motifs varies in high & low binding, in high binders it resulted in strong coordination of residues. All of the EF loops directly influence the central helix by long-and short-range interactions affecting the residue network thereby enhancing the allosteric interactions.

Figure 8 Comparison of the full networks of residues of three Ca^{2+} -binding EF-hand proteins. The left column of images shows all-residue networks of monomers of A) *NtEhCaBP1* B) *NtEhCaBP1* mutant and C) N-terminal domain of CaM, with the amino acid residues (nodes) as circles and the long- and short-range interactions (links) as lines. Each network is colored by highly correlated local substructures (communities) within which the network connections are dense. The residues in the EF-loop are highlighted to show the position of the loop. The network communication in each 2D network comprises different communities colored and is shown using the force directed layout. The middle column of images shows the DCCM maps calculated by using the NMA ensemble with the Hessian matrix; the C_{ij} value ranging from +1 to -1 are indicated by two colors, with cyan indicating positive correlation and pink indicating negative correlation amongst the residues. The leftmost column of images shows the community connections in each network. The rightmost column showing the largest communities numbered as per clustering are labeled in the image. The coarse grain network based on dynamically coupled communities shows the central helix residues forming the largest community connected with few residues from the N- and C-termini of *NtEhCaBP1*, in contrast to the *NtEhCaBP1*-mutant and CaM where the central helix is strongly connected to the residues from the EF loop site and the residues at the C-terminal end.

Discussion

In this study we upgraded our earlier designed software ‘Calbinder’ in a way so that along with identification of an EF-hand protein sequence user can also design and manipulate Ca^{2+} binding affinity of their protein of interest, as well. To confirm that our upgraded software helps in manipulating the Ca^{2+} binding affinity accurately or reaching to our expectation, we designed a high affinity EF-hand motif sequence and validated our hypothesis by exploiting biophysical and structural approach. As it was predicted, the newly designed EF-hand motifs of NtEhCaBP1-EF2 mutant protein yielded a higher Ca^{2+} binding affinity (ITC data) than the wild type protein. Further, with respect to Wt-NtEhCaBP1, the EF-2 of NtEhCaBP1-EF2 structure, shows a more compact Ca^{2+} coordination sphere, suggest the tight binding of Ca^{2+} . The compact Ca^{2+} coordination sphere induced structural bend in the helix-3, due to which helix-3 of one trimer interact with the helix-2 of another trimer leading to the hexamer formation in the solution as well as in the crystal state.

We anticipate that the incorporated mutation in NtEhCaBP1-EF-2 protein induced the cooperativity, i.e., binding of Ca^{2+} at one site induced the binding at second site. Due to the induced cooperative in Ca^{2+} binding, the NtEhCaBP1-EF-2 form of protein may binds Ca^{2+} with higher affinity.

As, the crystal structure of Wt-NtEhCaBP1 has Ca^{2+} bound at its both EF-hand motifs [28], therefore the observation of weak Ca^{2+} binding in ITC experiment of Nt-EhCaBP1 was a surprising observation. A justified statement for this concern is that the crystallization trials of Wt-NtEhCaBP1 was performed in the presence of high Ca^{2+} concentration (5mM), which allows Ca^{2+} ion to accommodate in the crystal lattice, therefore it observed in the crystal structure. Further, to check if higher concentration of Ca^{2+} may allow better binding for Wt-NtEhCaBP1 and the binding curve may achieve the saturation in ITC experiment, we stepwise increased the amount of Ca^{2+} till 25mM, however it did not help in achieving the saturation point similar or close to Nt-EhCaBP1 EF-2 mutant form of protein (data not shown). Suggesting that Wt-Nt-EhCaBP1 form of protein could not bind Ca^{2+} with high affinity even in presence high Ca^{2+} concentration.

Binding affinity of Ca^{2+} to EF-hand motifs containing CaBPs is widely governed by the residues present in EF-hand loop and the Ca^{2+} binding affinity can be manipulated usually by altering the 1st, 3rd, 5th, and 12th residues of the active site. Altering these critical Ca^{2+} binding residues of EF-hand motifs may drastically reduce the Ca^{2+} binding affinity or completely abolish the binding of Ca^{2+} . Such approach reduces the possibility study the function of these proteins over a wide range

of Ca^{2+} binding affinity. Our software addresses this issue and one can design/manipulate the EF-hand motif sequence to achieve the desired Ca^{2+} binding affinity, as per their experimental requirement. Most importantly with this software, manipulation of Ca^{2+} binding affinity is not limited only with the residues which are directly involved in Ca^{2+} coordination rather one can explore all the twelve residues of the EF-hand motif and manipulate the Ca^{2+} binding affinity by performing permutation combination of entire twelve residues present at EF-hand loop. Such flexibility may allow user to manipulate the Ca^{2+} binding affinity without affecting the conserve structural core (EF-hand motifs) of CaBPs. Although the experimental data presented in this manuscript is from only one protein, nevertheless it provides the strong confidence of machine learning based method of Ca^{2+} binding affinity manipulation.

With enough data, we expect it to be possible to factor the participation of cooperative binding into a predictive model that should provide deeper insight into the physiological Ca^{2+} -binding state. In summary, the non-bonded interactions at the interface of two EF loops and the ionic interactions at the Ca^{2+} -binding sites changed the binding characteristics of EF1 in the designed protein despite the residues of EF1 not having been altered. The allosteric communication from the coupling of structural changes at the Ca^{2+} -binding sites helped regulate the binding characteristic of the proteins. As an allosteric effector, the EF2 mutation was shown to modulate the behavior of the EF1 site. This form of intramolecular communication provided insight into the role of residues that participates in Ca^{2+} binding via dynamically coupled coordination with other residues in the structure. We expect the structure-function relationship obtained from the analysis of EF hand proteins to aid the building of highly accurate predictive models of Ca^{2+} binding in EF hand proteins. We also expect a better understanding of the various interactions carried out by all of the residues, i.e., not only those at the active sites but the entire EF hand motifs, to help in building machine-learned models capable of predicting the effect of cooperativity. Moreover, the understanding of accurate protein networks generated from high-resolution structural data can further enable the designing of similar proteins mimicking the EF-hand motif functions.

References

1. Bagur R, Hajnoczky G. Intracellular Ca(2+) Sensing: Its Role in Calcium Homeostasis and Signaling. *Mol Cell*. 2017;66(6):780-8. Epub 2017/06/18. doi: 10.1016/j.molcel.2017.05.028. PubMed PMID: 28622523; PubMed Central PMCID: PMC5657234.
2. Berridge MJ. Calcium signal transduction and cellular control mechanisms. *Biochim Biophys Acta*. 2004;1742(1-3):3-7. Epub 2004/12/14. doi: 10.1016/j.bbamcr.2004.08.012. PubMed PMID: 15590051.
3. Westerlund AM, Delemotte L. Effect of Ca²⁺ on the promiscuous target-protein binding of calmodulin. *PLoS Comput Biol*. 2018;14(4):e1006072. Epub 2018/04/04. doi: 10.1371/journal.pcbi.1006072. PubMed PMID: 29614072; PubMed Central PMCID: PMC5898786.
4. Maximciuc AA, Putkey JA, Shamoo Y, Mackenzie KR. Complex of calmodulin with a ryanodine receptor target reveals a novel, flexible binding mode. *Structure*. 2006;14(10):1547-56. Epub 2006/10/10. doi: 10.1016/j.str.2006.08.011. PubMed PMID: 17027503.
5. Halling DB, Liebeskind BJ, Hall AW, Aldrich RW. Conserved properties of individual Ca²⁺-binding sites in calmodulin. *Proc Natl Acad Sci U S A*. 2016;113(9):E1216-25. Epub 2016/02/18. doi: 10.1073/pnas.1600385113. PubMed PMID: 26884197; PubMed Central PMCID: PMC4780646.
6. Xie X, Dwyer MD, Swenson L, Parker MH, Botfield MC. Crystal structure of calcium-free human sorcin: a member of the penta-EF-hand protein family. *Protein Sci*. 2001;10(12):2419-25. Epub 2001/11/21. doi: 10.1110/ps.36701. PubMed PMID: 11714909; PubMed Central PMCID: PMC2374028.
7. Babu YS, Bugg CE, Cook WJ. Structure of calmodulin refined at 2.2 Å resolution. *J Mol Biol*. 1988;204(1):191-204. Epub 1988/11/05. doi: 10.1016/0022-2836(88)90608-0. PubMed PMID: 3145979.
8. Bitto E, Bingman CA, Bittova L, Frederick RO, Fox BG, Phillips GN, Jr. X-ray structure of *Danio rerio* secretagogin: A hexa-EF-hand calcium sensor. *Proteins*. 2009;76(2):477-83. Epub 2009/02/26. doi: 10.1002/prot.22362. PubMed PMID: 19241471; PubMed Central PMCID: PMC2742686.
9. Kawasaki H, Nakayama S, Kretsinger RH. Classification and evolution of EF-hand proteins. *Biometals*. 1998;11(4):277-95. Epub 1999/04/07. doi: 10.1023/a:1009282307967. PubMed PMID: 10191494.
10. Kumar S, Mazumder M, Gupta N, Chattopadhyay S, Gourinath S. Crystal structure of *Arabidopsis thaliana* calmodulin7 and insight into its mode of DNA binding. *FEBS Lett*. 2016;590(17):3029-39. Epub 2016/08/09. doi: 10.1002/1873-3468.12349. PubMed PMID: 27500568.
11. Kumar S, Mishra S, Gourinath S. Structural and functional diversity of *Entamoeba histolytica* calcium-binding proteins. *Biophys Rev*. 2020. Epub 2020/10/17. doi: 10.1007/s12551-020-00766-6. PubMed PMID: 33063237; PubMed Central PMCID: PMC7755952.
12. Rout AK, Patel S, Somlata, Shukla M, Saraswathi D, Bhattacharya A, et al. Functional manipulation of a calcium-binding protein from *Entamoeba histolytica* guided by paramagnetic NMR. *J Biol Chem*. 2013;288(32):23473-87. Epub 2013/06/21. doi: 10.1074/jbc.M112.411058. PubMed PMID: 23782698; PubMed Central PMCID: PMC395027.
13. Li W, Wang W, Takada S. Energy landscape views for interplays among folding, binding, and allostery of calmodulin domains. *Proc Natl Acad Sci U S A*. 2014;111(29):10550-5. Epub

2014/07/09. doi: 10.1073/pnas.1402768111. PubMed PMID: 25002491; PubMed Central PMCID: PMC4115553.

14. Gaertner TR, Kolodziej SJ, Wang D, Kobayashi R, Koomen JM, Stoops JK, et al. Comparative analyses of the three-dimensional structures and enzymatic properties of alpha, beta, gamma and delta isoforms of Ca²⁺-calmodulin-dependent protein kinase II. *J Biol Chem.* 2004;279(13):12484-94. Epub 2004/01/15. doi: 10.1074/jbc.M313597200. PubMed PMID: 14722083.

15. Hoeflich KP, Ikura M. Calmodulin in action: diversity in target recognition and activation mechanisms. *Cell.* 2002;108(6):739-42. Epub 2002/04/17. doi: 10.1016/s0092-8674(02)00682-7. PubMed PMID: 11955428.

16. Cates MS, Berry MB, Ho EL, Li Q, Potter JD, Phillips GN, Jr. Metal-ion affinity and specificity in EF-hand proteins: coordination geometry and domain plasticity in parvalbumin. *Structure.* 1999;7(10):1269-78. Epub 1999/11/05. doi: 10.1016/s0969-2126(00)80060-x. PubMed PMID: 10545326.

17. Piazza M, Taiakina V, Dieckmann T, Guillemette JG. Structural Consequences of Calmodulin EF Hand Mutations. *Biochemistry.* 2017;56(7):944-56. Epub 2017/01/26. doi: 10.1021/acs.biochem.6b01296. PubMed PMID: 28121131.

18. Drmota Prebil S, Slapsak U, Pavsic M, Ilc G, Puz V, de Almeida Ribeiro E, et al. Structure and calcium-binding studies of calmodulin-like domain of human non-muscle alpha-actinin-1. *Sci Rep.* 2016;6:27383. Epub 2016/06/09. doi: 10.1038/srep27383. PubMed PMID: 27272015; PubMed Central PMCID: PMC4895382.

19. Yang W, Lee HW, Hellinga H, Yang JJ. Structural analysis, identification, and design of calcium-binding sites in proteins. *Proteins.* 2002;47(3):344-56. Epub 2002/04/12. doi: 10.1002/prot.10093. PubMed PMID: 11948788.

20. Jones LM, Yang W, Maniccia AW, Harrison A, van der Merwe PA, Yang JJ. Rational design of a novel calcium-binding site adjacent to the ligand-binding site on CD2 increases its CD48 affinity. *Protein Sci.* 2008;17(3):439-49. Epub 2008/02/22. doi: 10.1110/ps.073328208. PubMed PMID: 18287277; PubMed Central PMCID: PMC2248323.

21. Zhou Y, Yang W, Kirberger M, Lee HW, Ayalasomayajula G, Yang JJ. Prediction of EF-hand calcium-binding proteins and analysis of bacterial EF-hand proteins. *Proteins.* 2006;65(3):643-55. Epub 2006/09/19. doi: 10.1002/prot.21139. PubMed PMID: 16981205.

22. Beccia MR, Sauge-Merle S, Lemaire D, Bremond N, Pardoux R, Blangy S, et al. Thermodynamics of Calcium binding to the Calmodulin N-terminal domain to evaluate site-specific affinity constants and cooperativity. *J Biol Inorg Chem.* 2015;20(5):905-19. Epub 2015/06/14. doi: 10.1007/s00775-015-1275-1. PubMed PMID: 26070361.

23. Mazumder M, Padhan N, Bhattacharya A, Gourinath S. Prediction and analysis of canonical EF hand loop and qualitative estimation of Ca(2)(+) binding affinity. *PLoS One.* 2014;9(4):e96202. Epub 2014/04/25. doi: 10.1371/journal.pone.0096202. PubMed PMID: 24760183; PubMed Central PMCID: PMC3997525.

24. Chang C LC. LIBSVM: A library for support vector machines. 2009;2(27):21-7.

25. Skjaerven L HS, Reuter N. Normal mode analysis for proteins. *J Mol Struct THEOCHEM.* 2009.

26. Hsu C CC, Lin C. A Practical Guide to Support Vector Classification. 2009;(1: 1-15.).

27. Lo SL, Cai CZ, Chen YZ, Chung MC. Effect of training datasets on support vector machine prediction of protein-protein interactions. *Proteomics.* 2005;5(4):876-84. Epub 2005/02/18. doi: 10.1002/pmic.200401118. PubMed PMID: 15717327.

28. Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, et al. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc Natl Acad Sci U S A*. 2000;97(1):262-7. Epub 2000/01/05. doi: 10.1073/pnas.97.1.262. PubMed PMID: 10618406; PubMed Central PMCID: PMCPMC26651.
29. Eddy SR. Where did the BLOSUM62 alignment score matrix come from? *Nat Biotechnol*. 2004;22(8):1035-6. Epub 2004/08/03. doi: 10.1038/nbt0804-1035. PubMed PMID: 15286655.
30. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-10. Epub 1990/10/05. doi: 10.1016/S0022-2836(05)80360-2. PubMed PMID: 2231712.
31. UniProt C. UniProt: a hub for protein information. *Nucleic Acids Res*. 2015;43(Database issue):D204-12. Epub 2014/10/29. doi: 10.1093/nar/gku989. PubMed PMID: 25348405; PubMed Central PMCID: PMCPMC4384041.
32. Kumar S, Ahmad E, Mansuri MS, Kumar S, Jain R, Khan RH, et al. Crystal structure and trimer-monomer transition of N-terminal domain of EhCaBP1 from *Entamoeba histolytica*. *Biophys J*. 2010;98(12):2933-42. Epub 2010/06/17. doi: 10.1016/j.bpj.2010.03.048. PubMed PMID: 20550906; PubMed Central PMCID: PMCPMC2884261.
33. Jain R, Kumar S, Gourinath S, Bhattacharya S, Bhattacharya A. N- and C-terminal domains of the calcium binding protein EhCaBP1 of the parasite *Entamoeba histolytica* display distinct functions. *PLoS One*. 2009;4(4):e5269. Epub 2009/04/23. doi: 10.1371/journal.pone.0005269. PubMed PMID: 19384409; PubMed Central PMCID: PMCPMC2668073.
34. Kumar S, Ahmad E, Kumar S, Khan RH, Gourinath S. Flexibility of EF-hand motifs: structural and thermodynamic studies of Calcium Binding Protein-1 from *Entamoeba histolytica* with Pb²⁺, Ba²⁺, and Sr²⁺. *BMC Biophys*. 2012;5:15. Epub 2012/08/22. doi: 10.1186/2046-1682-5-15. PubMed PMID: 22906057; PubMed Central PMCID: PMCPMC3483242.
35. Kumar S, Padhan N, Alam N, Gourinath S. Crystal structure of calcium binding protein-1 from *Entamoeba histolytica*: a novel arrangement of EF hand motifs. *Proteins*. 2007;68(4):990-8. Epub 2007/06/08. doi: 10.1002/prot.21455. PubMed PMID: 17554780.
36. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol*. 1997;276:307-26. Epub 1997/01/01. PubMed PMID: 27754618.
37. Matthews BW. Solvent content of protein crystals. *J Mol Biol*. 1968;33(2):491-7. Epub 1968/04/28. doi: 10.1016/0022-2836(68)90205-2. PubMed PMID: 5700707.
38. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. *J Appl Crystallogr*. 2007;40(Pt 4):658-74. Epub 2007/08/01. doi: 10.1107/S0021889807021206. PubMed PMID: 19461840; PubMed Central PMCID: PMCPMC2483472.
39. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*. 1997;53(Pt 3):240-55. Epub 1997/05/01. doi: 10.1107/S0907444996012255. PubMed PMID: 15299926.
40. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004;60(Pt 12 Pt 1):2126-32. Epub 2004/12/02. doi: 10.1107/S0907444904019158. PubMed PMID: 15572765.
41. Laskowski RA MM, Moss DS and Thornton JM. Procheck – a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993;26:283-91.
42. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235-42. Epub 1999/12/11. doi: 10.1093/nar/28.1.235. PubMed PMID: 10592235; PubMed Central PMCID: PMCPMC102472.

43. Sievers F, Higgins DG. Clustal omega. *Curr Protoc Bioinformatics*. 2014;48:3 13 1-6. Epub 2014/12/17. doi: 10.1002/0471250953.bi0313s48. PubMed PMID: 25501942.
44. TA H. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 1999.
45. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. *Nucleic Acids Res*. 2010;38(Web Server issue):W545-9. Epub 2010/05/12. doi: 10.1093/nar/gkq366. PubMed PMID: 20457744; PubMed Central PMCID: PMC2896194.
46. Laskowski RA. PDBsum new things. *Nucleic Acids Res*. 2009;37(Database issue):D355-9. Epub 2008/11/11. doi: 10.1093/nar/gkn860. PubMed PMID: 18996896; PubMed Central PMCID: PMC2686501.
47. Laskowski RA, Swindells MB. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J Chem Inf Model*. 2011;51(10):2778-86. Epub 2011/09/17. doi: 10.1021/ci200227u. PubMed PMID: 21919503.
48. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein-ligand interaction profiler. *Nucleic Acids Res*. 2015;43(W1):W443-7. Epub 2015/04/16. doi: 10.1093/nar/gkv315. PubMed PMID: 25873628; PubMed Central PMCID: PMC4489249.
49. Holec PV, Hackel BJ. PyMOL360: Multi-user gamepad control of molecular visualization software. *J Comput Chem*. 2016;37(30):2667-9. Epub 2016/09/21. doi: 10.1002/jcc.24489. PubMed PMID: 27645768.
50. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605-12. Epub 2004/07/21. doi: 10.1002/jcc.20084. PubMed PMID: 15264254.
51. McLean D. Adobe Photoshop and Illustrator techniques. *J Audiov Media Med*. 2001;24(3):132-5. Epub 2001/10/05. PubMed PMID: 11584604.
52. Jubb HC, Higuero AP, Ochoa-Montano B, Pitt WR, Ascher DB, Blundell TL. Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures. *J Mol Biol*. 2017;429(3):365-71. Epub 2016/12/15. doi: 10.1016/j.jmb.2016.12.004. PubMed PMID: 27964945; PubMed Central PMCID: PMC5282402.
53. Gurusaran M SP, Dinesh Kumar KS, Radha P, Thulaa Tharshan KPS, Sathesh SN, et al. Hydrogen Bonds Computing Server (HBCS): An online web server to compute hydrogen-bond interactions and their precision. *J Appl Crystallogr*. 2016.
54. Piovesan D, Minervini G, Tosatto SC. The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res*. 2016;44(W1):W367-74. Epub 2016/05/21. doi: 10.1093/nar/gkw315. PubMed PMID: 27198219; PubMed Central PMCID: PMC4987896.
55. Skjaerven L, Yao XQ, Scarabelli G, Grant BJ. Integrating protein structural dynamics and evolutionary analysis with Bio3D. *BMC Bioinformatics*. 2014;15:399. Epub 2014/12/11. doi: 10.1186/s12859-014-0399-6. PubMed PMID: 25491031; PubMed Central PMCID: PMC4279791.
56. Girvan M, Newman ME. Community structure in social and biological networks. *Proc Natl Acad Sci U S A*. 2002;99(12):7821-6. Epub 2002/06/13. doi: 10.1073/pnas.122653799. PubMed PMID: 12060727; PubMed Central PMCID: PMC122977.
57. Perica T, Kondo Y, Tiwari SP, McLaughlin SH, Kemplen KR, Zhang X, et al. Evolution of oligomeric state through allosteric pathways that mimic ligand binding. *Science*. 2014;346(6216):1254346. Epub 2014/12/20. doi: 10.1126/science.1254346. PubMed PMID: 25525255; PubMed Central PMCID: PMC4337988.

58. Waltersson Y, Linse S, Brodin P, Grundstrom T. Mutational effects on the cooperativity of Ca²⁺ binding in calmodulin. *Biochemistry*. 1993;32(31):7866-71. Epub 1993/08/10. doi: 10.1021/bi00082a005. PubMed PMID: 8347591.
59. Marsden BJ, Shaw GS, Sykes BD. Calcium binding proteins. Elucidating the contributions to calcium affinity from an analysis of species variants and peptide fragments. *Biochem Cell Biol*. 1990;68(3):587-601. Epub 1990/03/01. doi: 10.1139/o90-084. PubMed PMID: 2198059.
60. Kirberger M, Wang X, Deng H, Yang W, Chen G, Yang JJ. Statistical analysis of structural characteristics of protein Ca²⁺-binding sites. *J Biol Inorg Chem*. 2008;13(7):1169-81. Epub 2008/07/03. doi: 10.1007/s00775-008-0402-7. PubMed PMID: 18594878.
61. Gifford JL, Walsh MP, Vogel HJ. Structures and metal-ion-binding properties of the Ca²⁺-binding helix-loop-helix EF-hand motifs. *Biochem J*. 2007;405(2):199-221. Epub 2007/06/26. doi: 10.1042/BJ20070255. PubMed PMID: 17590154.
62. Godzik A, Sander C. Conservation of residue interactions in a family of Ca-binding proteins. *Protein Eng*. 1989;2(8):589-96. Epub 1989/08/01. doi: 10.1093/protein/2.8.589. PubMed PMID: 2813336.
63. Maniccia AW, Yang W, Johnson JA, Li S, Tjong H, Zhou HX, et al. Inverse tuning of metal binding affinity and protein stability by altering charged coordination residues in designed calcium binding proteins. *PMC Biophys*. 2009;2:11. Epub 2009/12/23. doi: 10.1186/1757-5036-2-11. PubMed PMID: 20025729; PubMed Central PMCID: PMC2816670.
64. Eijssink VG, Bjork A, Gaseidnes S, Sirevag R, Synstad B, van den Burg B, et al. Rational engineering of enzyme stability. *J Biotechnol*. 2004;113(1-3):105-20. Epub 2004/09/24. doi: 10.1016/j.jbiotec.2004.03.026. PubMed PMID: 15380651.
65. Gopal B, Swaminathan CP, Bhattacharya S, Bhattacharya A, Murthy MR, Surolia A. Thermodynamics of metal ion binding and denaturation of a calcium binding protein from *Entamoeba histolytica*. *Biochemistry*. 1997;36(36):10910-6. Epub 1997/09/09. doi: 10.1021/bi9702546. PubMed PMID: 9283081.
66. Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res*. 2018;46(W1):W363-W7. Epub 2018/06/04. doi: 10.1093/nar/gky473. PubMed PMID: 29860391; PubMed Central PMCID: PMC6031066.