Biochimica et Biophysica Acta 1824 (2012) 608–619

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Identification of the critical structural determinants of the EF-hand domain arrangements in calcium binding proteins

Ye-dan Feng [1], Jing Li [1], Wen-chang Zhou, Zhi-guang Jia, Qun Wei *

Department of Biochemistry and Molecular Biology, Beijing Normal University, Beijing Key Laboratory, Beijing 100875, PR China

ARTICLE INFO

Article history:
Received 29 July 2011
Received in revised form 2 December 2011
Accepted 4 January 2012
Available online 21 January 2012

Keywords:
EF-hand calcium binding protein
Calmodulin
Calcineurin B
Structural determinant
Molecular dynamics

ABSTRACT

EF-hand calcium binding proteins (CaBPs) share strong sequence homology, but exhibit great diversity in structure and function. Thus although calmodulin (CaM) and calcineurin B (CNB) both consist of four EF hands, their domain arrangements are quite distinct. CaM and the CaM-like proteins are characterized by an extended architecture, whereas CNB and the CNB-like proteins have a more compact form. In this study, we performed structural alignments and molecular dynamics (MD) simulations on 3 CaM-like proteins and 6 CNB-like proteins, and quantified their distinct structural and dynamical features in an effort to establish how their sequences specify their structures and dynamics. Alignments of the EF2–EF3 region of these proteins revealed that several residues (not restricted to the linker between the EF2 and EF3 motifs) differed between the two groups of proteins. A customized inverse folding approach followed by structural assessments and MD simulations established the critical role of these residues in determining the structure of the proteins. Identification of the critical determinants of the two different EF-hand domain arrangements and the distinct dynamical features relevant to their respective functions provides insight into the relationships between sequence, structure, dynamics and function among these EF-hand CaBPs.

1. Introduction

Calcium (Ca^{2+}) signals and calcium binding proteins (CaBPs) of the EF-hand superfamily are involved in the regulation of almost every process in life from birth to death [1,2]. The EF hand is the elementary structural motif in almost all the proteins in this superfamily. Although it is generally assumed that sequence similarity implies shared structure, function, and evolution, the CaBPs are highly homologous and yet are very diverse in their structures and in the targets with which they interact [3]. The CaBPs, such as calmodulin (CaM), calcineurin B (CNB), sarcoplasmic calcium binding protein (SCP) and S100B, which have significant sequence similarity, have different arrangements of their N- and C-terminal domains. Understanding this exception to the sequence-implies-structure paradigm, and addressing the underlying basis of structural diversity, could provide deeper insight into the relationships between sequence, structure, dynamics and function in these CaBPs.

CaM, the best studied and prototypical example of an EF-hand calcium sensor protein, is involved in the regulation of many important Ca^{2+}-dependent signaling pathways by moving between various cellular compartments and interacting with various targets [4]. CaM contains four EF hands: the first two combine to form a globular N-terminal domain, which is separated by a flexible linker from the homologous C-terminal domain containing the other two EF hands. Although CaM varies in the way it interacts with its targets, which is matched by the extensive and diverse list of targets that it regulates [5,6], the most classic conformation of CaM is the dumb-bell shape which exists when CaM is in a Ca^{2+}-saturated state and is not bound to any target [7–10]. A similar extended architecture can also be found in some other EF-hand proteins, such as troponin C (TnC) [11–13] and human calmodulin-like protein (hCLP) [14]. TnC, together with troponin I (TnI) and troponin T (TnT), make up the troponin (Tn) complex, which acts as a trigger in skeletal and cardiac muscle by switching on contraction at a critical Ca^{2+} concentration [15], while hCLP is found in mammary epithelial cells [16], and functions as a myosin-10 binding partner [17].

CNB has a more compact four EF-hand structure: the two globular domains align on the same side and are linked by a U-shaped linker [18,19]. CNB is the regulatory subunit of calcineurin, a Ca^{2+}/CaM-dependent serine/threonine protein phosphatase that plays a critical role in many cellular processes [20]. After the structure of CNB was established, three-dimensional structures of similar architecture were

Abbreviations: Ca^{2+}, calcium; CaBPs, calcium binding proteins; CaM, calmodulin; CNB, calcineurin B; SCP, sarcoplasmic calcium binding protein; TnC, troponin C; TnI, troponin I; TnT, troponin T; Tn, troponin; hCLP, human calmodulin-like protein; CIB1, calcium and integrin binding protein 1; CHPs, calcineurin B homologous proteins; Frq1, yeast frequenin; NCS, neuronal calcium sensor; GCAPs, guanylyl cyclase activating proteins; KChIPs, Kv channel interacting proteins; RECO, recoverin; CNA, calcineurin A; NHE1, Na^+/H^+ exchanger 1; Pik1, phosphatidylinositol 4 kinase; PDB, protein data bank; MD, molecular dynamics; RMSD, root mean square deviation; PME, particle mesh Ewald; ED, essential dynamics; PCA, principal component analysis

* Corresponding author. Tel./fax: +86 10 5880 7365.
E-mail address: weiq@bnu.edu.cn (Q. Wei).

[1] These authors contributed equally to this work.

1570-9639/$ – see front matter.
doi:10.1016/j.bbapap.2012.01.005

found in CIB1 [21,22], CHP1 [23], CHP2 [24], Frq1 (yeast frequenin) [25], NCS1 (neuronal calcium sensor 1, human frequenin) [26], neurocalcin-δ [27], recoverin [28], GCAP1 [29], GCAP2 [30], GCAP3 [31], KChIP1 [32] and KChIP4 [33]. CIB1 (calcium and integrin binding protein 1) was initially identified as a specific binding partner of platelet integrin αIIbβ3; it is also expressed in many other cell types and interacts with a number of other target proteins [34]. CHPs (calcineurin B homologous proteins) are involved in membrane trafficking; they bind multiple effectors and potentially modulate their functions [23], in particular inhibiting NHE1 (Na^+/H^+ exchanger 1) [35]. Also, frequenin, neurocalcin-δ, recoverin, GCAPs (guanylyl cyclase activating proteins) and KChIPs (Kv channel interacting proteins) constitute the 5 subgroups (A to E, respectively) of the NCS protein family; they regulate many cellular events in neurons and retinal photoreceptors, and have specific targets that do not overlap with those of CaM [36,37].

The EF-hand proteins mentioned above can be divided into two groups according to the arrangements of their N- and C-terminal domains: the CaM-like proteins and CNB-like proteins. It is interesting that some members of the opposing groups are more homologous at the sequence level than some proteins in the same group. A typical example is CNB, whose full primary sequence is more homologous to CaM (35% identity) than to recoverin (29% identity). Such proteins provide an opportunity to gain a further understanding of the relationships between sequence, structure, dynamics, and function in these EF-hand proteins. Does any factor other than strong sequence homology have a significant impact on protein folding and the arrangements of the EF-hand domains?

Pawlowski et al. have done comparative modeling of three EF-hand proteins with different arrangements of their N- and C-terminal domains, i.e., CaM, recoverin and SCP; they suggested that correlated mutations on the surface of two units, and the presence of additional fragments, are two factors that might stabilize the desired topologies [38]. Because of the limited numbers of three-dimensional structures available at the time, they could not compare the structures of a large number of different proteins to identify the determinants of the divergence in the EF-hand domain arrangements. Mouawad et al. analyzed the sequences of 17 CaBPs of known structure, and developed a novel procedure to predict the compactness of a CaBP of unknown structure. They suggested that the compactness of a CaBP is determined by the hydrophilicity of a few residues (average of 5) that link the two domains [39].

Here, we report the results of a study of the structural basis of the two EF-hand domain arrangements represented, respectively, by CaM and CNB, and the sequence–structure–dynamics–functional relationships among the CaBPs with four EF hands. We chose nine representative EF-hand proteins from the various subgroups (Table 1), namely 3 CaM-like proteins (CaM, TnC and hCLP) and 6 CNB-like proteins (CNB, recoverin, CIB1, NCS1, GCAP3 and KChIP1), all of whose structures were available in their Ca^{2+}-saturated states. We chose proteins whose functions had been well studied as this would be beneficial for later examination of the relationships between dynamics and function. Using structural alignments and molecular dynamics (MD) simulations on the 9 proteins, we quantified the respective features on structures and dynamics of the two groups of proteins. To address how sequence specifies structure and dynamics, we aligned partial (EF1, EF2–EF3, and EF4) and complete sequences of the 9 proteins, and tentatively identified the critical structural determinants in the EF2–EF3 region. Then we used a customized inverse folding approach by comparative modeling of CaM and CNB, to confirm the importance of these structural determinants. Finally, we surveyed the functions of the two groups of CaBPs as previously reported, especially their manner of interacting with target proteins. The functions of the proteins could be accurately inferred from the identified determinants of the structural and dynamical features of each of the domain arrangements. Our findings provide deeper insight into the sequence–structure–dynamics–functional relationships among the CaBPs.

2. Materials and methods

2.1. Preparation of structures, and structural alignments

Basic information on the 9 chosen proteins is given in Table 1. Their three-dimensional structures obtained from the PDB (protein data bank) [40] rendered with PyMOL [41] are shown in Fig. S1.

For the structural alignments and subsequent MD simulations, residues 133–137 of NCS1, 137–142 of CIB1, and 1–4 and 148 of CaM, which are all not visible in the respective crystal structures, were added with the Insight II Biopolymer Module (Accelrys, Inc., San Diego, CA). In the NMR structure 1JSA (recoverin), the first model in the PDB file was used.

The structural alignment matrix was generated using the Insight II Homology Module with the multiple sequence alignment method. The RMSDs (root mean square deviations) for the backbone between each pair of the 9 proteins were calculated.

2.2. MD simulations

MD simulations were performed using GROMACS [42,43] with the ROMOS96 force field [44]. Each protein was embedded in a box containing an SPC water model [45], with at least 12 Å between the protein and the edge of the box. To neutralize the system and reach physiological ionic strength of 100 mM, appropriate amounts of sodium and chloride ions were added at random positions. After minimization and equilibration for 100 ps under position restraints for the backbone of the protein, each MD production run was performed for 20 ns. In the MD simulations, the LINCS algorithm [46] was used to constrain the lengths of all bonds. The time step for simulations was 2 fs. The simulations were run under NPT conditions, using Berendsen's coupling algorithm to keep temperature and pressure constant ($P = 1$ bar; $T_p = 0.5$ ps; $T = 300$ K; $T_t = 0.1$ ps) [47]. For Van der Waals forces and short-range electrostatic interactions we used a cutoff of 12 Å. Long-range electrostatic forces were obtained by

Table 1
Basic information on the 9 proteins.

Protein	Source	PDB ID	UniProt ID	Length [a]	Ca^{2+}-binding	Reference
CaM	Human	1CLL	P62158	148/149	4	Chattopadhyaya et al. (Ref. [8])
TnC	Rabbit	1TN4	P02586	157/160	4	Houdusse et al. (Ref. [13])
hCLP	Human	1GGZ	P27482	144/149	4	Han et al. (Ref. [14])
CNB	Human	1TCO	P63098	169/170	4	Griffith et al. (Ref. [18])
RECO [b]	Bovine	1JSA	P21457	188/202	2	Ames et al. (Ref. [28])
CIB1	Human	1XO5	Q99828	180/191	2	Gentry et al. (Ref. [22])
NCS1	Human	1G8I	P62166	187/190	3	Bourne et al. (Ref. [26])
GCAP3	Human	2GGZ	O95843	165/209	3	Stephen et al. (Ref. [31])
KChIP1	Human	1S1E	Q9NZI2	181/227	2	Scannevin et al. (Ref. [32])

[a] The first number in each case is the number of residues in the three-dimensional structure in the PDB entry; the second number is the number of residues in the primary amino acid sequence in the UniProt entry.
[b] RECO: recoverin.