Biophysical Chemistry, 31 (1988) 133–137
Elsevier

BPC 01292

Conformational properties of Ca^{2+}-binding segments of proteins from the troponin C superfamily

Grzegorz Boguta [a] and Andrzej Bierzyński [b]

[a] *Department of Biophysics, Institute of Experimental Physics, University of Warsaw, Al. Zwirki i Wigury 93, 02-089 Warszawa, and* [b] *Department of Biophysics, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, ul. Rakowiecka 36, 02-532 Warszawa, Poland*

Received 5 September 1987
Accepted 1 February 1988

Ca^{2+}-binding protein; Protein structure; Protein folding

The troponin C superfamily consists of about 100 Ca^{2+}-binding proteins. Sequence variations observed in these proteins have been analyzed and lead to the following conclusions. (1) There are some strict rules defining the set of calcium ligands necessary for effective Ca^{2+} binding. (2) If they are fulfilled, the Ca^{2+} binding constant depends on tertiary interactions within a protein, as well as the free energy of secondary structures of its polypeptide chain. The former provide a constant contribution to the free energy of protein folding and the Ca^{2+}-binding process. (3) The observed variety in Ca^{2+}-binding constants of these proteins results from the various abilities of segments of these proteins to assume the correct secondary structure.

1. Introduction

The troponin C (TNC) superfamily consists of a large variety of regulatory proteins with high affinity for Ca^{2+}. Polypeptide chains of these proteins consist of 32-amino-acid, repetitive units with related sequences: in principle, each of the units can bind one Ca^{2+}. X-ray studies of a few representatives of the TNC superfamily [1–5] have shown a common structural pattern of all the units [6]. Two α-helices formed by residues 1–12 and 21–32 connected by a loop formed by central residues 13–20 are oriented almost perpendicularly to each other (fig. 1). The Ca^{2+}-coordinating ligands are provided by residues 13, 15, 17 and 19 situated within the loop, and residues 21 and 24 located at the N-terminus of the C-terminal α-helix. The ligands are arranged on verticals of the octahedron and, in the Cartesian coordinate framework, are denoted by X, Y, Z, $-Y$, $-X$ and $-Z$, respectively.

The first ligand (X) is always a β-carboxyl group of the aspartic acid * (table 1). The last one ($-Z$) is a side chain carboxyl group of either the glutamate or, less frequently, aspartate. All 'defective' units missing either Asp at position 13 or an acidic residue at position 24, are deprived of Ca^{2+}-binding activity. At position $-Y$, calcium is coordinated by a peptide oxygen. All the remaining ligands (Y, Z and $-X$) are, as a rule, oxygen atoms of side chains of Asp, Glu, Asn, Gln, Ser of Thr, but at least one of them must be a carboxyl group of either Asp or Glu. Sometimes, one of the positions 15, 17, or 19 is occupied by Gly. In these instances, a water molecule coordinates calcium.

There seem to be no strict rules concerning the occupancy of the other positions within the units,

Correspondence address: G. Boguta, Department of Biophysics, Institute of Experimental Physics, University of Warsaw, Al. Zwirki i Wigury 93, 02-089 Warszawa, Poland

* Calmodulin units appeared to be an exception to this rule until their sequences were corrected by Wada et al. [7].

Table 1

Amino acid frequencies at each position of Ca^{2+}-binding loops in proteins from the TNC superfamily

149 units have been analyzed [11a,13–18]. The method of loop identification will be described elsewhere (G. Boguta and A. Godzik, to be published).

Residue	Ligand:	*X*		*Y*		*Z*		*−Y*		*−X*			*−Z*
	Position:	13	14	15	16	17	18	19	20	21	22	23	24
Ala		0	17	0	7	1	0	3	0	0	12	19	0
Arg		0	7	0	7	0	1	6	0	0	3	0	0
Asn		0	3	36	13	22	1	2	0	8	2	6	0
Asp		149	0	104	2	88	4	5	0	54	1	41	14
Cys		0	1	0	0	0	0	2	3	0	0	0	0
Gln		0	16	0	4	2	0	9	0	2	4	5	0
Glu		0	14	7	2	1	0	12	0	22	13	48	135
Gly		0	1	0	89	5	141	0	0	19	8	1	0
His		0	0	0	4	0	1	1	0	0	0	0	0
Ile		0	11	0	0	0	0	4	103	0	3	0	0
Leu		0	2	0	0	0	0	3	12	1	10	0	0
Lys		0	43	0	21	0	0	20	0	0	10	16	0
Met		0	2	0	0	0	0	3	2	4	0	2	0
Phe		0	5	0	0	0	0	28	0	0	32	0	0
Pro		0	0	0	0	0	0	0	0	0	1	9	0
Ser		0	12	2	0	30	1	6	0	25	3	0	0
Thr		0	6	0	0	0	0	25	0	14	12	2	0
Trp		0	0	0	0	0	0	0	0	0	2	0	0
Tyr		0	1	0	0	0	0	18	0	0	13	0	0
Val		0	8	0	0	0	0	2	29	0	20	0	0

except for positions 18 and 20. The former is predominantly occupied by Gly and the latter by one of the following, closely similar, hydrophobic residues: Ile, Val, or Leu (table 1).

The Ca^{2+}-binding units, even when cleaved off and separated from the rest of the protein molecule, maintain their conformation and Ca^{2+}-binding ability to a significant extent. For example, CD measurements show 21% helix content in the apo form of the RSTNC unit III [8]. This increases to 39% upon Ca^{2+} binding [8] as compared with 55% in the intact metal-bound protein [9]. The binding constant, although considerably reduced, is still high: 2.6×10^5 M^{-1} [10].

As many as 149 Ca^{2+}-binding units with various sequences are known. For a number of these, the binding constants, ranging from 10^3 to 10^8 M^{-1} in intact proteins and from 10^2 to 10^5 M^{-1} in isolated units (for a review, see ref. 11a), have been determined. The proteins from the TNC superfamily therefore provide an excellent basis for the systematic analysis of the relationship between local sequence mutations and binding activity. Using the secondary structure prediction method of Garnier et al. [11b], we have shown that the affinity of a unit for calcium is correlated with the tendency of the units segments to adopt the correct secondary structure. From this we have elaborated [11a] a set of rules which makes it possible to evaluate, within an accuracy of one order of magnitude, the Ca^{2+}-binding constants of the isolated units as well as those in intact proteins.

In this work, we discuss some general conclusions that we have drawn from the results obtained with our method [11a]. These focus on the problem of the relevance of local conformational properties of polypeptide chains to protein folding and activity.

The conformational features of polypeptide chains are considered merely in terms of their tendency to adopt a particular type of structure.

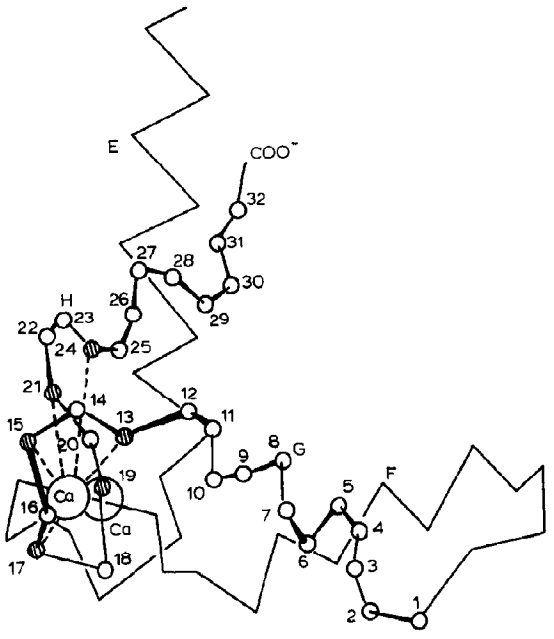

Fig. 1. Schematic representation of α-carbon positions in the second biunit of turkey skeletal muscle troponin C [3]. Residues in the C-terminal unit are numbered from 1 to 32, according to the nomenclature used in the text. Hatched circles denote Ca^{2+}-coordinating ligands. Helices E–H are labeled according to Kretsinger's notation [6].

No further attempts are made towards achieving a more precise determination of conformational free energies. This approach is the result of our choice of a rather primitive method of conformation analysis based on a procedure for structure prediction. Although very rough, it appears to give more trustworthy results, at least for a series of homologous peptides, than any approximate method of energy calculations. On the other hand, in the case of large peptides, such as those discussed here, precise calculations using, say, the Monte Carlo method are too time-consuming to be feasible, particularly if solvent effects cannot be neglected.

2. Isolated Ca^{2+}-binding units

The conformation of the Ca^{2+}-binding units in proteins, as revealed by X-ray crystallography [1–5], is as follows:

```
          5         10        15        20
H H H H H H H H H T T T T T T T C C C H H
          25        30
  H H H H H H H H H H
```

where T denotes the β-turn, C the random coil, and H the α-helical conformation. The affinity of an isolated unit for Ca^{2+} is strictly correlated with its propensity for this conformation [11a]. The highest binding constants, of the order of 10^5 M^{-1}, characterize those units for which the predicted secondary structure exactly matches the X-ray pattern. Any deviation from the X-ray pattern results in a drop in binding activity.

The most significant region in this respect is the loop segment (residues 13–20). If the predicted conformation of any n residues from this region deviates from the ideal conformation, the binding constant of the unit decreases by $n/2$ orders of magnitude. The exceptions to this rule are residues 17 and 18. For either one, the β-turn and random coil (but not helical) conformations are equally admissible.

The propensity of residues 21–32 for α-helix formation is also of great importance. Note that the correct orientation of the two ligands $-X$ and $-Z$ is fixed by the first turn of this helix. Because of the cooperative character of α-helix formation, the whole C-terminal segment tends to adopt the helical conformation. In the worst case, when no segment in this region is predicted as helical, the binding constants of the unit decreases by 4 orders of magnitude.

The importance of the N-terminal helix for Ca^{2+} binding is less clear. Unfortunately, it is a rather conserved region. Therefore, a large enough set of known sequence mutations is not available to determine the relationship of this structure to the affinity of the unit for Ca^{2+}. In intact proteins, a coordinated Ca^{2+} is situated almost on the axis of the N-terminal helix and close to its C-terminus. Most probably, an electrostatic interaction between the positive ion and the negative pole of the aggregated helix dipole contributes to the stability of the complex (cf. ref. 12). Experiments with fragments of the unit show that re-

moval of the N-terminal helix from the unit leads to a substantial drop in binding constant of about 2 orders of magnitude [10]. Nevertheless, it is not clear whether the orientation of the helices in isolated units is similar to that in the protein molecules. It is possible that they tend to become arranged in an anti-parallel fashion relative to each other, with their hydrophobic surfaces in close contact [19]. In this case, the N-terminal helix would enhance the binding ability of the unit through the stabilization of the α-helical conformation of its C-terminal segment.

Since the binding activity of non-defective units does not depend, to any noticeable degree, on the type of ligands coordinating Ca^{2+}, the net free energy of Ca^{2+} coordination is similar in all active units. The differences in binding constants of the unit arise, therefore, only from differences in free energies of conformational transitions from the apo to metal-bound forms.

3. Ca^{2+} binding by intact proteins

With the exception of parvalbumins and oncomodulin which will be not discussed here, all other known proteins from the TNC superfamily are composed of an even number of units arranged according to a common structural pattern. Beginning from the N-terminus, units I and II form a compact domain – the biunit [5]. In larger proteins, a second, similar domain is formed by units III and IV (see fig. 1). Both biunit domains bind calcium independently and are not in contact with one another [2–4]. Therefore, the biunits are treated as separate entities in the following considerations.

The biunit structure is maintained by the following tertiary interactions [20]:

(1) Residues 20 of both units are locked together by a hydrogen bond.

(2) Bulky hydrophobic side chains of residues 20 touch each other.

(3) the N-terminal helix of each unit is oriented almost antiparallel to, and in close contact with, the C-terminal helix of the other.

As a consequence, the affinity of both units for Ca^{2+} is enhanced considerably. Moreover, their conformational properties appear to be combined to result in a common set of characteristics: they bind Ca^{2+} with the same constant K_{Ca} [11a] which can be estimated according to the following expression:

$$K_{Ca} = f\sqrt{K'_{Ca}K''_{Ca}}$$

where K'_{Ca} and K''_{Ca} denote the binding constants of the isolated units. f, which can be treated as a binding-enhancement parameter, equals approx. 5000 for biunits of low and medium activity and drops when $\sqrt{K'_{Ca}K''_{Ca}} \geq 10^4$, reaching about 500 for the most active biunits ($K_{Ca} \geq 2 \times 10^8$ M^{-1}).

It is evident that the tertiary interactions between the units make a substantial contribution to the conformational stability of the metal-bound form of the biunits. Consequently, the free energy for the conformational transition from apo to metal-bound forms is lower in biunits by about 20–14 kJ/mol biunit, as estimated from the enhancement factor f.

In strongly binding units, even if they are isolated, the free energy for the transition is already low. This is the reason why in these units Ca^{2+} binding is not greatly enhanced by tertiary interactions. Note that, if the conformation of isolated units in the apo form were exactly the same as in the metal-bound form, tertiary interactions would have no effect whatsoever on Ca^{2+} binding.

It is interesting that even if one of the units cannot actually bind Ca^{2+} because some ligands are missing from it, eq. 1 can still be used, with a 'dummy' K_{Ca} value estimated from the primary structure of the non-binding unit. Evidently, the unit-unit interactions are the same, even if one of the units is defective and cannot bind Ca^{2+}. It is worth stressing here that in all known non-binding units, being components of biunit domains, positions 18 and 20, the most important positions for correct unit-unit interaction, are occupied predominantly by Gly and Ile, respectively (G. Boguta and A. Godzik, to be published).

4. Protein folding

The folding process has not been studied in proteins from the TNC superfamily. Nevertheless,

much can be deduced about the process from the conformational properties of protein fragments, and from conformational changes induced by Ca^{2+} binding in these proteins.

Even small fragments, containing only one helical segment, show a remarkable tendency to adopt the α-helical a conformation [10]. In protein fragments, the helices are almost certainly localized to the same regions of the polypeptide chain as in the native proteins, since the binding loops separating them comprise many strong helix breakers. Therefore, there is every reason to believe that the helical segments are the autonomous folding units, in the sense proposed by Shoemaker et al. [21]. When proteins fold, the helices form first and then recombine with each other into the tertiary structure of the apo form known from X-ray studies [3,4].

Conformational changes between the apo and metal-bound forms can be treated as the final step of the folding process, induced by Ca^{2+} binding. When calcium binds, the α-helix content increases [9,10] and the helical segments rearrange [19]. The Ca^{2+}-binding constants depend on the free energies for both processes. Our results show that the K_{Ca} values can be predicted reasonably well without taking into consideration the differences in tertiary interactions in various proteins. Therefore, in proteins from the TNC superfamily, tertiary interactions seem to be nonspecific, in the sense that they provide a constant contribution to the free energy of folding of these proteins, irrespective of sequence variations. Differences in the free energy of folding appear to arise solely from differences in the free energy of secondary structures of proteins.

References

1 R.H. Kretsinger and C.E. Nockolds, J. Biol. Chem. 248 (1973) 3313.

2 S.Y. Babu, J.S. Sack, T.J. Greenhough, C.E. Bugg, A.R. Means and W.J. Cook, Nature 315 (1985) 37.

3 O. Herzberg and M.N.G. James, Nature 313 (1985) 653.

4 M. Sundaralingam, R. Bergstrom, G. Strasburg, S.T. Rao, P. Roychowdhury, M. Greaser and B.C. Wang, Science 227 (1985) 945.

5 D.M.E. Szebenyi, S.K. Obendorf and K. Moffat, Nature 294 (1981) 327.

6 R.H. Kretsinger, CRC Crit. Rev. Biochem. 8 (1980) 119.

7 Y. Wada, K. Sobue, K. Morimoto, T. Matsuo, I. Katakuse, S. Kakiuchi, H. Matsuda and A. Hayashi, Biochim. Biophys. Acta 832 (1985) 383.

8 B. Nagy, J.D. Potter and J. Gergely, J. Biol. Chem. 253 (1978) 5971.

9 P.C. Leavis, S.S. Rosenfeld, J. Gergely, Z. Grabarek and W. Drabikowski, J. Biol. Chem. 253 (1978) 5452.

10 R.E. Reid, D.M. Clare and R.S. Hodges, J. Biol. Chem. 255 (1980) 3642.

11a G. Boguta, D. Stepkowski and A. Bierzynski, J. Theor. Biol. (1988) in the press.

11b J. Garnier, D.J. Osguthorpe and B. Robson, J. Mol. Biol. 120 (1978) 97.

12 P.T. van Duijnen and B.T. Thole, Biopolymers 21 (1982) 1749.

13 T. Takagi, K. Konishi and J.A. Cox, Biochemistry 25 (1986) 3585.

14 P.W. Wilson, M. Harding and D.E.M. Lawson, Nucleic Acids Res. 13 (1985) 8867.

15 T. Takagi, M. Nojiri, K. Konishi, K. Maruyama and Y. Nonomura, FEBS Lett. 201 (1986) 41.

16 T. Takagi, S. Kudoh and K. Konishi, Biochem. Biophys. Acta 874 (1986) 318.

17 J.H. Collins, R. Jakes, J. Kendrick-Jones, J. Leszyk, W. Barouch, J.L. Theibert, J. Spiegel and A.G. Szent-Gyorgyi, Biochemistry 25 (1986) 7651.

18 B. Watanabe, T. Maita, K. Konno and G. Matsuda, Biol. Chem. Hoppe-Seyler 367 (1986) 1025.

19 O. Herzberg, J. Moult and M.N.G. James, J. Biol. Chem. 261 (1986) 2638.

20 O. Herzberg and M.N.G. James, Biochemistry 24 (1985) 5298.

21 K.R. Shoemaker, R. Fairman, P.S. Kim, E.J. York, J.M. Stewart and R.L. Baldwin, Cold Spring Harbor Symp. (1987) submitted for publication.