Secondary-Structure Predictions of Calcium-Binding Proteins†

Patrick Argos*

ABSTRACT: The known tertiary structure of carp muscle parvalbumin is consistent with an "EF-hand" architecture (helix–loop–helix) for each calcium-ion binding site. Primary-sequence alignments have indicated four EF hands in rabbit skeletal muscle troponin C and in rabbit myosin alkali light chains. Five secondary-structure prediction methods, based on amino acid sequence only, have been fully computerized and used to calculate joint prediction histograms for several calcium-binding proteins. The joint histogram can suggest directly the extent and sequence of the helical- and loop-structural elements, as well as any secondary structural distortions or evolutionary developments. Since the histogram predicted well the length and sequence of secondary structural elements in carp muscle parvalbumin, it seemed reasonable to calculate the joint distribution for other proteins that might bind calcium through the EF-hand configuration. The histograms indicated the four EF-hand regions speculated for rabbit skeletal muscle troponin C but suggested only three such hands in bovine cardiac muscle troponin C with a distorted fourth hand. Considerable secondary structural distortion is postulated for the alkali light chains. Possible EF configurations consistent with the histogram results are speculated for *Escherichia coli* acyl-carrier protein and bovine prothrombin fragment 1, which have been shown to bind calcium. The secondary-structure-prediction algorithms appear to be a useful adjunct to sequence-alignment techniques, especially in cases where the primary sequence homology is weak or the evolutionary distance is large.

The tertiary structure of carp muscle calcium-binding parvalbumin has been determined to 1.9 Å resolution by x-ray crystallographic techniques (Kretsinger and Nockolds, 1973). Though its function is unknown, it is assumed to act as a mediator of Ca^{2+} concentration within muscle (Kretsinger and Nockolds, 1973). CBP[1] contains six regions of α helix, A–F, connected by loops or turns. Two Ca^{2+} are coordinated, respectively, in the loops joining helices C and D and helices E and F. Helix C, the CD turn, and helix D are related to the corresponding EF region by an intramolecular approximate twofold axis, as shown and discussed by Kretsinger and Nockolds (1973). Such regions are referred to as "EF hands". The helical lengths were observed as approximately 11 residues, while turns were about 9 residues long.

The coordination geometry of the calcium ions are represented by octahedra, which are also related by a twofold axis between the CD and EF hands. At five of the six octahedral vertices (designated sequentially X, Y, Z, −Y, −X, −Z from the amino terminus) oxygen-atom ligands are provided by appropriate amino acids, while the fourth ligand (−Y) is provided by a peptide oxygen.

Tufty and Kretsinger (1975) have proposed that EF-hand regions can be recognized in any given protein amino acid sequence if a minimum of 10 out of 16 critical structure-forming and liganding residues can be aligned to a 29-residue test sequence determined from CBP. Based on such primary-sequence comparisons, Tufty and Kretsinger and others (Collins, 1974; Weeds and McLachlan, 1974; Kretsinger and Barry, 1975) have proposed four EF hands in the calcium-binding component of rabbit skeletal muscle troponin C and in alkali-extractable light chains from rabbit skeletal myosin. However, the "10 of 16" test may be too simplistic to recognize EF hands in calcium-binding proteins distantly related to CBP (through deletions and insertions in the 29-residue segment) and, yet, containing the helix–loop–helix tertiary configuration.

Several methods have been proposed to predict secondary structural regions (helices, β sheets, and loops) based only on a knowledge of amino acid sequence and have been recently applied to adenylate kinase (Schulz et al., 1974) and T4 phage lysozyme (Matthews, 1975). The prediction algorithms assign hierarchies to amino acid patterns that give rise to such secondary structures in known proteins. A joint probability histogram is obtained by summing the number of methods predicting an amino acid to be in a certain secondary structural fold. By comparing the joint histogram predictions based on five methods with experimentally known protein structures, Argos et al. (1976) have shown that the joint probabilities predict well secondary structure in certain "favorable" cases, which include fish parvalbumins. Furthermore, the joint probability histogram can predict secondary structural regions at least as well as the most successful predicting method for a given protein sequence.

In the present work, the prediction histograms were first calculated for proteins suspected or known to bind calcium through the EF architecture. Though the prediction of the length and sequence of the helix–loop–helix secondary structure does not assure a tertiary EF configuration for calcium binding, it provides a sensitive test for those regions already predicted as calcium-binding sites by the Tufty–Kretsinger primary-sequence model.

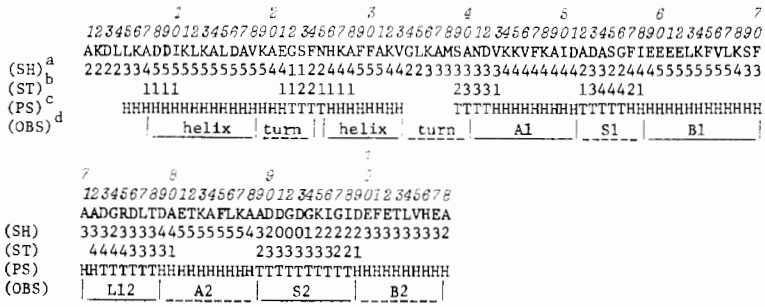
Many proteins contain helices and loops in sequence and yet do not bind calcium. However, if the predictive algorithms indicate the proper extent and sequence of secondary structural regions in proteins suspected of calcium binding, it is probable, or at least suggestive, that such regions form tertiary EF configurations. The predicted sites can then be chemically and crystallographically tested. It is in this spirit that predictions

† From the Department of Physics, Southern Illinois University, Edwardsville, Illinois 62026. *Received July 22, 1976.* This research was supported by a grant from the state of Illinois.

* Present address: Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907.

[1] Abbreviations used are: CBP, calcium-binding parvalbumin; TNC, troponin C; ALC-2, rabbit myosin alkali light chain 2; ACP, *Escherichia coli* acyl-carrier protein; PT1, vitamin K dependent fragment 1 portion of bovine prothrombin; Nbs_2, 5,5′-dithiobis(2-nitrobenzoic acid) extractable light chain; GLA, γ-carboxyglutamic acid.

TABLE I: The Predicted Secondary Structure of Pike Muscle Calcium Binding Protein.



[a] SH refers to the number of prediction methods (maximum five) predicting a residue as helical. For this and all subsequent tables, single letters are used for amino acids according to the following code: Ala, A; Arg, R; Asn, N; Asp, D; Cys, C; Gln, Q; Glu, E; Gly, G; His, H; Ile, I; Leu, L; Lys, K: Met, M; Phe, F; Pro, P; Ser, S; Thr, T; Trp, W; Tyr, Y; Val, V. [b] ST indicates turn predictions with four as a maximum score. [c] An H in the PS line indicates a predicted helical residue and T a residue predicted in a loop configuration. [d] The OBS line indicates observed helical structures in carp muscle calcium-binding protein as solid lines, while dashed lines refer to observed loop or turn regions. The EF-hand secondary structural elements are indicated by the A-S-B notation (helix–loop–helix) adopted in the text. The symbol L refers to the connecting loop between hand pairs.

of EF hands through joint histogram calculations are made for *E. coli* acyl-carrier protein and the vitamin K dependent fragment 1 portion of bovine prothrombin.

Primary-sequence comparisons generally demand close evolutionary relatedness, while secondary structural homology does not, as evidenced in recent nucleotide-binding domain comparisons in dehydrogenases (Rossmann et al., 1974). The joint distribution may thus point to EF hands that primary-sequence comparisons cannot detect. Furthermore, prediction methods can indicate within a few residues the starting and terminating points of secondary structural elements in polypeptide chains, rather than merely wide regions containing several secondary structure types as in EF hands. Certain secondary structural distortions or evolutionary developments in the classic EF configuration can also be recognized through structure predictions and related to different Ca^{2+}-binding affinities, or to a lack of such affinities.

The present paper discusses the application of the joint prediction histogram to fish parvalbumins, muscle troponin C's, rabbit alkali light chain 2, *E. coli* acyl-carrier protein, and the vitamin K dependent fragment 1 portion of bovine prothrombin. The predictions for CBP indicate the efficacy of the joint histogram in determining the proper length and sequence of helices and loops. The four EF hands speculated for skeletal muscle TNC and ALC-2 through primary-sequence alignment are also suggested by the prediction schemes, though considerable distortion is indicated in cardiac TNC and ALC-2. In ALC-2, the two regions having weak Ca^{2+} affinity are predicted with larger loop regions than those for high-affinity sites. The F helix of the third hand and E helix of the fourth are predicted as fused, indicating either a possible hand-to-hand configuration different from that in parvalbumin, or the nonexistence of a fourth EF fold. An EF hand is predicted for ACP, presently undetected by the sequence alignment tests. The joint histogram calculations for PT1 indicated helices and turns which might form eight EF hands for calcium binding.

Methods

Several prediction schemes have been developed to determine locations of secondary structure from amino acid sequence. There are presently twelve methods available for helical predictions, eight for sheets, and four for loop determinations. Matthews (1975) references the available methods. The techniques attempt to predict regions of secondary structure by trying to correlate the patterns of amino acids which give rise to such structures in proteins of known structure.

The methods used here for helical predictions are those of Nagano (1973, 1974), Burgess et al. (1974), Chou and Fasman (1974a,b), Kabat and Wu (1974), and Barry and Friedman (Schulz et al., 1974). Only the former four were used for sheet and turn predictions. The methods were fully automated (Argos et al., 1976) through computer programs but did not include an evaluation process made by some of the predictors themselves, who modify results obtained from a strict adherence to their rules due to overlapping regions of predicted helices, sheets, or turns. Despite elimination of this evaluation, it has been shown (Argos et al., 1976) that the five-method automated joint prediction histogram for helices in adenylate kinase is virtually equivalent to the ten-method histogram compiled from predictions by the authors themselves. Other methods were not used in the present work due to the extreme complexity in automating prediction rules (cf. method of Lim (1974a,b)) or to the unavailability of computer programs.

Argos et al. (1976) have demonstrated that the joint predictions are at least as good as the most successful individual prediction. However, by comparing the joint automated predictions with about 40 known protein structures, the predictions were shown to be good in only about half the cases; yet, the best predictions were for helical secondary structures.

The criteria used to indicate the accuracy of the predictions was the correlation coefficient, C (Matthews, 1975). The correlation, C, between helical prediction and observation is given by:

$$C = \frac{(h/N - \overline{H}_{obsd}\overline{H}_{pred})}{\{\overline{H}_{obsd}\overline{H}_{pred}(1 - \overline{H}_{obsd})(1 - \overline{H}_{pred})\}^{1/2}}$$

where N is the total number of residues in the protein, h is the number of residues predicted and observed to be helical, and $\overline{H}_{obsd}$ and $\overline{H}_{pred}$ are, respectively, the fraction of the protein observed and predicted to be helical. Similar definitions can be made for turn and sheet correlations. If $1.0 \geq C \geq 0.4$, the prediction is considered quite good or, at least, useful. A perfect prediction would give $C = 1.0$, while a random one yields $C =$

TABLE II: The Predicted Secondary Structures of Rabbit Skeletal Troponin C

```
                  2         3         4
        1234567890123456789012345678901234567890123456789
        LSEEMIAEFKAAFDMFDADGGGDISVKELGTVMRMLGQT
(SH)[a] 23445555445555332    1223444433333332 1
(ST)[b]                111344433311          2212
(PS)[c] HHHHHHHHHHHHHHHTTTTTTT HHHHHHHHHHHTTTT
(PAA)[d] |    A1    |  S1  |   B1   | L12 |

        5         6         7         8
        01234567890123456789012345678901234567890123456789
        PTKEELDAIIEEVDEDGSGTIDFEEFLVMMVRQMKEDAKG
(SH)    124444444444432  11225555555554555443323
(ST)    221          12233321              3443
(PS)    TTHHHHHHHHHHHHTTTTTT  HHHHHHHHHHHHHHTTTT
(PAA)   _|   A2   | S2 |   B2    | J23 |

                  1         1         1
        9         0         1         2
        012345678901234567890123456789012345
        KSEEELAECFRIFDRNADGYIDAEELAEIFRASGEH
(SH)    34445555442222121111234555555533311 22
(ST)    21       111244443321          1221
(PS)    THHHHHHHHH  TTTTTTTTHHHHHHHHHHTTTT
(PAA)   _|   A3   | S3 |   B3   |L34

        1   1         1         1
        2   3         4         5
        67890123456789012345678901234567890123
        VTDEEIESLMKDGDKNNDGRIDFDEFLKMMEGVQ
(SH)    34455544431        11134445555321
(ST)          133244443311221
(PS)    HHHHHHHHHH TTTTTTTTT  HHHHHHHHH
(PAA)   _|   A4   | S4 |   B4    |
```

[a] SH refers to the number of prediction algorithms (maximum five) indicating a residue as helical. The sequences of the four EF hands are aligned according to Tufty and Kretsinger (1975). [b] ST indicates loop predictions with a maximum score of four. [c] The line PS indicates residues as predicted helical (H) or in a turn configuration (T) from the joint histogram. [d] The line (PAA) is the prediction based on amino acid aligments (Weeds and McLachlan, 1974).

TABLE III: A Comparison of the Lengths of the Predicted Secondary Structural Regions for Rabbit Skeletal and Bovine Cardiac Muscle TNC from the Joint Prediction Algorithms and for Rabbit Skeletal Muscle TNC from Amino Acid Sequence Alignment.

(JP) Cardiac[a]	(JP) Skeletal[b]	(AA) Skeletal[c]	Structural[d] Element
4–28 (25)[e]	13–26 (14)	17–27 (11)	A1
31–36 (6)	28–34 (7)	29–38 (10)	S1
38–48 (11)	37–47 (11)	39–48 (10)	B1
49–54 (6)	48–53 (6)	49–53 (5)	L12
55–64 (10)	54–65 (12)	54–64 (11)	A2
65–74 (10)	66–71 (6)	65–73 (9)	S2
75–85 (11)	74–87 (14)	74–84 (11)	B2
86–92 (7)	88–92 (5)	85–93 (9)	J23
93–103 (11)	93–103 (11)	94–104 (11)	A3
104–112 (9)	104–111 (8)	105–113 (9)	S3
113–123 (11)	112–122 (11)	114–124 (11)	B3
124–127 (4)	123–126 (4)	125–129 (5)	L34
127–138 (12)	128–137 (10)	130–140 (11)	A4
139–150 (12)	139–147 (9)	141–149 (9)	S4
151–160 (10)	150–159 (10)	150–161 (12)	B4

[a] The extent of the secondary structural elements indicated for rabbit cardiac muscle troponin C by the joint prediction algorithms (JP). [b] The extent of the secondary structural elements indicated for bovine skeletal muscle troponin C by the joint prediction algorithms (JP). [c] The lengths of the secondary structural elements indicated for rabbit skeletal muscle troponin C by amino acid sequence alignment (AA). [d] The letters A, S, and B refer, respectively, to the first helix, Ca^{2+}-binding site loop, and second helix of an EF hand, while L and J indicate the loops connecting, respectively, hands or hand pairs. [e] The numbers in parentheses indicate the residue length of a secondary structural element. It must be noted that the amino acid numbering scheme follows that for cardiac troponin C (Van Eerd and Takahashi, 1975) and *not* the numerical scheme indicated in Table II for skeletal muscle troponin C.

0.0; total disagreement between observation and prediction would lead to $C = -1.0$.

The correlation coefficients for helical predictions determined from the joint automated histogram were, respectively, 0.745, 0.789, and 0.720 using the homologous hake (Capony et al., 1973), pike (Frankenne et al., 1973), and carp (Coffee and Bradshaw, 1973) muscle CBP amino acid sequences and the carp CBP known tertiary structure. The loop or turn correlation coefficients were, respectively, 0.788, 0.704, and 0.679. No sheet regions were predicted. Table I lists the pike primary sequence, the number of methods predicting a particular residue as in a helix or loop, the predicted structure, and the observed structure based on the carp crystallographic results (Kretsinger and Nockolds, 1973). It is evident that the EF-hand secondary structure is a "favorable" prediction case; it thus appears reasonable that the joint histogram should predict the appropriate secondary structural regions for calcium-binding sites suspected in other proteins.

The format used to obtain the best possible secondary-structure prediction from the automated histogram for the CBP's has been previously described (Argos et al., 1976). It should be noted that for all predictions to be discussed here no residue scored more than one or two predictions for β sheet, while most residues in calcium-binding proteins scored zero. The sheet predictions will, thus, not be discussed further.

Results and Discussion

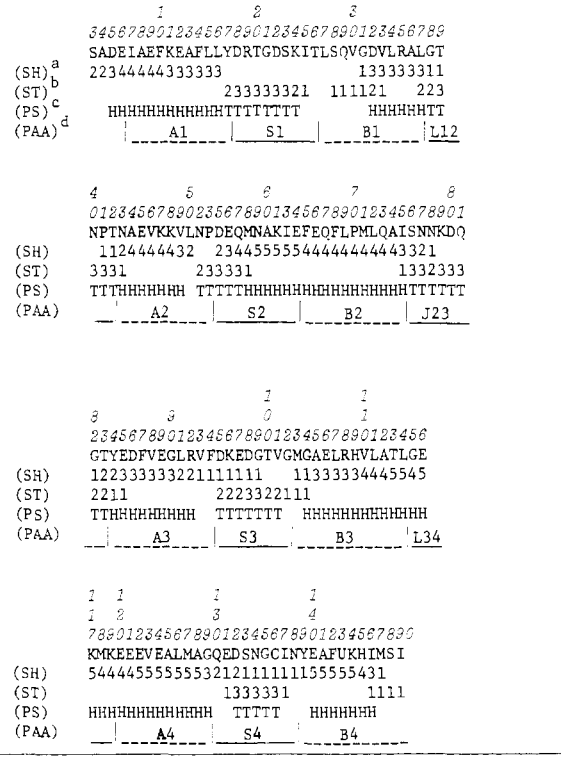
Troponin C Predictions. Troponin is a complex of three proteins (Greaser and Gergely, 1973) of which one component binds calcium (Hartshorne and Pyun, 1971) and is called troponin C. Troponin binds to actin and tropomyosin in the thin filaments of muscle. When calcium is released from the sarcoplasmic reticulum, it binds to TNC resulting in a tropomyosin shift relative to actin and thus allowing ATP hydolysis, the sliding of the thick filament relative to the thin filament, and muscular contraction (Weber and Murray, 1973). The amino acid sequences of rabbit skeletal (Collins et al., 1973) and bovine cardiac (Van Eerd and Takahashi, 1975) muscle TNC's have been determined and shown to be homologous (Van Eerd and Takahashi, 1975).

Collins (1974), Weeds and McLachlan (1974), and Tufty and Kretsinger (1975) have recognized through primary-sequence comparisons four EF hands in rabbit skeletal muscle TNC. This appears consistent with chemical data that TNC binds two Ca^{2+} with $pK_d \simeq 6.5$ and two with $pK_d \simeq 5.0$ (Potter and Gergely, 1974; Kretsinger, 1975). Kretsinger and Barry (1975) have predicted C_α coordinates for the tertiary structure of TNC based on the assumption of four EF hands arranged in two pairs with overall 222 symmetry.

A joint automated histogram was calculated for both rabbit skeletal and bovine cardiac muscle TNC. Table II shows the joint histogram and resulting predicted secondary structure for skeletal muscle TNC. The amino acids of each EF hand are aligned according to that proposed by Kretsinger and Barry. Table III compares the predicted helices and turns in both cardiac and skeletal TNC with those proposed by Weeds and McLachlan (1974) through primary-structure alignments.

For all four EF hands, the agreement between predictions made through primary sequence and the joint automated histogram is excellent, with the exception of the long A1 helix

TABLE IV: The Joint Automated Histogram Indicating the Predicted Secondary Structure of Rabbit Myosin Alkali Light Chain 2.



[a] SH refers to the number of prediction schemes indicating a residue helical. The sequences of the four EF-hand regions are aligned according to Kretsinger and Barry (1975). Amino acid insertions in the second EF hand (G(51), S(54), K(62)) have been omitted. [b] ST refers to the turn-prediction score. [c] The line PS shows the residues predicted helical (H) or in loops (T) using the rules of Argos et al. (1976) as applied to the joint histogram. [d] PAA refers to the prediction based on primary sequences, according to Weeds and McLachlan (1974), with dashed lines for helical regions and solid lines for loop configurations.

TABLE V: A Comparison of the Lengths of Secondary Structural Elements Observed in Carp Muscle CBP with Those Predicted by the Joint Histogram for Rabbit Skeletal Muscle Troponin C, Bovine Cardiac Muscle Troponin C, and Myosin Alkali Light Chain 2.

Structural Element	S-TNC[a]	C-TNC[b]	ALC-2[c]	CBP[d]
A1	14	25	12	11
S1	7	6	8	7
B1	11	11	6	12
L12	6	6	5	8
A2	12	10	8	10
S2	6	10	6	10
B2	14	11	18	9
J23	5	7	8	
A3	11	11	9	
S3	8	9	7	
B3	11	11	(25)[e]	
L34	4	4		
A4	10	12		
S4	9	12	5	
B4	10	10	7	

[a] Rabbit skeletal muscle troponin C. [b] Bovine cardiac muscle troponin. [c] Rabbit alkali light chain 2. [d] Carp muscle calcium-binding protein. [e] Helices B3 and A4 are predicted as fused in ALC-2.

of the amino-terminal EF hand in bovine cardiac TNC. Ebashi (1972) has indicated through chemical experiments that there are 2.4 calcium binding sites per cardiac TNC molecule. The amino acid sequences of the three C-terminal EF regions in cardiac TNC are very similar to corresponding regions in skeletal TNC, while the amino-terminal hand of cardiac TNC contains seven amino acid replacements and one additional residue (Van Eerd and Takahashi, 1975). Furthermore, there is a reduction in the number of suitable ligands for calcium binding in this region (Kretsinger and Nockolds, 1973). The long A1 helix predicted for cardiac, but not skeletal, TNC may be sufficient to distort or disguise the classic EF structure with the resulting loss of calcium-binding ability in this amino region.

These results clearly suggest that skeletal TNC and cardiac TNC contain, respectively, four and three units of the type A-S-B, where A is the first helix of an EF hand, B the second helix, and S the loop region providing the calcium-binding ligands. Weeds and McLachlan (1974) propose that this structure arose by duplication of the calcium-binding components of CBP to produce a tetrahedral molecule of the form: (A1-S1-B1-L12-A2-S2-B2)-J23-(A3-S3-B3-L34-A4-S4-B4), with the new joining piece J23. They found that the only change required to fit the skeletal TNC structure to CBP was to shorten the link region between helices B1 and A2, and B3 and A4. The secondary-structure predictions show these loop regions, L12 and L34, to contain fewer amino acids (6 and 4, respectively) than other turns in the skeletal and cardiac TNC's (Table III).

Weeds and McLachlan (1974), as well as Kretsinger and Barry (1975), suggest further that calcium sites S1 and S2 are the high-affinity binding locations in skeletal TNC, while S3 and S4 are the weaker ones, especially S4 with its many basic residues. The secondary-structure prediction results indicate that loops S1 and S2 have the least number of residues (a total of 13 in skeletal TNC and 16 in cardiac TNC), while S3 and S4 have a total of 17 and 21 residues predicted, respectively, in loop structures for skeletal and cardiac TNC's. The weaker affinity may result from the greater openness of the 3 and 4 binding sites. Furthermore, cardiac S4 has 12 predicted residues in the turn configuration, more than any other S site.

Alkali Light Chains. In some primitive species, the regulatory system of muscle contraction is associated with the light chains of myosin (Szent-Gyorgyi et al., 1972), rather than through troponin-mediated calcium sensitivity. The regulatory functions of the myosin light chains are not well known; yet, some light chains have been implicated in calcium binding and regulation in muscle. The ethylenediaminetetraacetic acid extractable light chain in thick-filament-controlled muscles (Szent-Gyorgyi et al., 1973) appears to be modulated by Ca^{2+}. The 5,5′-dithiobis(2-nitrobenzoic acid) extractable light chain does bind at least one Ca^{2+} (Weber et al., 1972; Collins, 1975), and may be phosphorylated in resting muscle (Collins, 1975).

Complete sequences have been determined for the alkali-extractable light chains from rabbit skeletal myosin (Weeds, 1967; Weeds and Frank, 1972); however, alkali light chains do not bind calcium at physiological concentrations of magnesium and when isolated from the heavy chain of myosin. Yet, from primary-sequence alignment with CBP, Tufty and Kretsinger (1975) and Weeds and McLachlan (1974) have suggested alkali light chains contain four EF hands.

The joint prediction histogram for the secondary structure of ALC-2 of rabbit myosin was calculated and is shown in

TABLE VI: The Predicted Secondary Structure of *E. coli* Acyl-Carrier Protein.

```
                  1         2         3         4         5         6         7
         1234567890123456789012345678901234567890123456789012345678901234567890123456
         STIEERVKKIIGENLGVKNEEVTDNASFVEDLGADSLDTVELVMALEEEFDTEIPDEEAEKITTVNAAIDYINGHQA
(SH)[a]  2233333332211111233444333333332211223344555555555443312455554444444433211
(ST)[b]             11          123223321          12111
(PJT)[c]          |____A1____|___S1___|____B1____|
```

[a] SH refers to the score (maximum five) for helical predictions. [b] ST refers to the number of prediction methods (maximum of four) indicating a residue, as in a turn configuration. [c] The EF-hand region predicted by the joint histogram.

Table IV. The predicted regions do not seem to follow well the classic EF-hand secondary structure in two respects. The A and B helices in CBP's and TNC's generally contain 10 to 14 residues and the S structures about 6 to 10 residues. As shown in Table V, the corresponding helices in ALC-2 contain as few as 6 residues (B1) and as many as 18 residues (B2). Furthermore, the B3-L34-A4 region is predicted as one long helix of 25 residues.

Weeds and McLachlan (1974) found it necessary to shorten the L12 and L34 links. The former shortening seems to agree with the ALC-2 histogram results which indicate 5 residues in L12; however, the L34 turn is predicted nonexistent with B3 and A4 fused. This fusion may break the typical twofold symmetry found between EF hands in CBP and lessen or eliminate calcium binding affinities in these latter two regions.

Given the erratic lengths of predicted helices in ALC-2, the structure may be sufficiently altered so as not to bind calcium, as presently observed for alkali light chains, or to bind fewer than four calciums, as appears to be the case for Nbs_2. Weeds and McLachlan (1974) speculate that lessened calcium affinity in ALC-2 is a result of distorted binding sites: an extra glycine residue following Leu-A2(8), a distorting proline residue at B2(6), an extra basic residue in front of Ile-S2(8), and bulky valine and methionine residues in S3. Thus, despite the primary structural similarity, ALC and TNC or CBP do not show as clear-cut a secondary structural similarity. Perhaps, ALC-2 lost an original function of Ca^{2+} binding at one or more sites in its evolution and has achieved a completely or partially different function (possibly connected with ATPase activity in myosin) and yet maintains a structural memory of its origins.

Acyl Carrier Protein. Acyl-carrier protein has been isolated from *Escherichia coli* and shown to function as a coenzyme in fatty acid biosynthesis (Sauer et al., 1964; Majerus, 1968). The primary sequence of ACP, which consists of 77 amino acids, is known (Vanaman et al., 1968a,b; Weeds and McLachlan, 1974). The protein binds two calciums per molecule at pH 6.6 (Schulz, 1972).

Application of the Tufty–Kretsinger procedures to ACP does not indicate any region fulfilling the "10 of 16" criterion. The joint prediction histogram shown in Table VI points to a polypeptide region between residues 21 and 50 that may well be an EF hand. Two helices and a loop are significantly predicted. If Glu-30 is considered an insertion and if an Ile has been deleted between Leu-37 and Asp-38, the Tufty–Kretsinger score becomes 9 of 16 key residues with 4 of 5 calcium ligands provided by Asp-31, Asp-35, Asp-38, and Glu-41.

There is a weak turn predicted between residues 52 and 56, which could point to an EF hand in the region 38 to 66. The Tufty–Kretsinger criteria gave an 8 of 16 key residue score with three calcium ligands. Since ACP binds two calciums and since two EF hands are predicted with an overlap region of residues 38 to 50, an "EFG fist" may exist between 21 and 66. An EFG fist is formed by two adjacent EF hands that overlap in the second helix of the first hand and the first helix of the second hand to form two calcium binding sites through a helix–loop–helix–loop–helix structure. Such a fist has been recently speculated for bovine intestinal calcium-binding protein (Moffat et al., 1975).

Schulz (1972) has suggested the importance of carboxyl groups for calcium binding in ACP. The chemical modification of the carboxyl groups of ACP (performed by reacting acetylated ACP with 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide and ethyl glycinate) indicates that the reactivity of at least one carboxyl group is decreased by the binding of Ca^{2+} to ACP and that the modification of this group results in extensive ACP inactivation. Reaction of ACP with radioactive ethyl glycinate yields a single peptide between residues 29 and 41 with largest radioactivity (Davie et al., 1969). Schulz suggests binding in the neighborhood of Ser-36. Since the joint histogram predicts strongly a loop region between amino acids 31 to 38, the EF-hand region from residues 21 to 50 seems a likely Ca^{2+}-binding site in ACP.

Prothrombin. Vitamin K is required for the biosynthesis of three of the four blood-coagulation proteins (called factors): prothrombin, factor IX (Christmas factor), and factor X_1 (Stuart factor) (Davie et al., 1969). The primary structure is known for the entire bovine prothrombin chain (Magnusson et al., 1975) consisting of 582 residues, for the light chain of bovine factor X_1 (Enfield et al., 1975) with 140 residues, and for amino-terminal portion of factor IX (Fujikawa et al., 1974) with only 14 known residues. These studies have demonstrated that the amino-terminal regions of the three blood-clotting proteins have considerable sequence homology and that the first 35 residues of factor X_1 light chain and prothrombin contain 10 γ-carboxyglutamic acid residues. This modified amino acid is necessary for calcium binding by these proteins (Nelsestuen et al., 1975).

The vitamin K dependent peptides obtained from prothrombin and factor X_1 are the amino-terminal tryptic peptides which contain most of the GLA residues and adsorb quantitatively to barium citrate. Vitamin K is essential to the γ carboxylation of glutamic amino acids (Bajaj et al., 1975). Prothrombin fragment 1 is the product of thrombin action on prothrombin and consists of the 274 amino-terminal prothrombin residues (Magnusson et al., 1975).

The GLA residues in the vitamin K dependent peptides are found primarily in pairs. A proposal based on model building involves a single calcium-binding site through coordination with the carboxyl groups of two adjacent GLA residues (Enfield et al., 1975). A calcium-binding site has been identified in prothrombin at positions 7 and 8 as γ-carboxyglutamic acid (Stenflo, 1974; Stenflo et al., 1974). Since three pairs of the GLA residues are homologous in the factor X_1 and pro-

TABLE VII: The Secondary-Structure Prediction Histogram for the Vitamin K Dependent Portion of Bovine Factor X_1 Light Chain.

```
                1         2         3         4
       1234567890123456789012345678901234567890123
       ANSFLEEVKQGNLERECLEEACSLEEAREVFEDAENTDNFWSKY
(SH)[a] 22244444321122333344444555444555555534433211
(ST)          12221
```

[a] SH and ST are as defined in previous tables. The three pairs of adjacent GLA residues proposed as calcium-binding sites are underlined. They were considered as glutamic acids for the prediction schemes.

thrombin sequences (Enfield et al., 1975) and since the maximum amount of calcium bound by the vitamin K dependent peptide of prothrombin is about 3 mol/mol (Nelsestuen et al., 1975), it would seem reasonable to infer three effective centers of calcium-ion chelation in the first 35 or so residues in prothrombin and factor X_1 light chain. However, prothrombin will bind about 10 mol of calcium/mol (Bajaj et al., 1975), leaving seven calcium sites in the fragment 1 portion of prothrombin between residues 35 and 274.

It has been shown that intact PT1 binds to phospholipid residues in the presence of calcium, while the vitamin K dependent peptide does not (Nelsestuen et al., 1975). This and other results led Nelsestuen et al. (1975) to the conclusion that the factors important for *tight* calcium binding by prothrombin in solution are also important for binding the protein to phospholipid. They state further that, since the presence of GLA residues do not provide particularly tight calcium-binding sites, secondary/tertiary protein structure is also required for binding prothrombin and factor X_1 to phospholipid. Their results do not eliminate the possibility that tight calcium binding can be provided by groups in fragment 1 other than by the carboxyl groups in GLA residues.

Since tight calcium binding is typical of the EF-hand structure, the joint prediction histogram was calculated for the first 274 residues of prothrombin (fragment 1) in an attempt to recognize seven EF calcium-binding sites supposedly not provided by the amino-terminal 35 residues, which contain three pairs of adjacent GLA acids. *It must be emphasized that the implications for calcium binding derived from the secondary-structure prediction methods are highly speculative but are given in the spirit of instigating further inquiry into alternative calcium-binding mechanisms in prothrombin.*

Table VII gives the secondary-structure prediction scores for the first 44 amino acids of factor X_1 light chain, which comprises the vitamin K dependent portion of the chain. The three pairs of adjacent GLA residues (6–7, 19–20, 25–26) are predicted in helical structures. A similar prediction was observed for the homologous region in prothrombin, except for a slight turn indication between Pro-22 and Ser-24. In factor X_1 Ala-21 is substituted for Pro-22, indicating its nonessentialness in calcium binding. Pro-22 probably causes a slight kink or distortion in the long second helix (residues 15–41 in factor X_1). The helix–loop–helix predicted for factor X_1 is reminiscent of the EF secondary-structure pattern. Perhaps, the vitamin K dependent portion of prothrombin and factor X_1 evolved from the classic EF structure found in other calcium-binding proteins.

The predicted secondary structure for bovine prothrombin fragment 1 is given in Table VIII. The seven EF units are indicated according to the symbolic scheme adopted for TNC. The functional significance of this region is not yet clear (Magnusson et al., 1976) and is proposed here to bind seven calcium ions. Some of the helical regions are predicted poorly, or short in length: A2, B4, A5, A7, and B7. However, if some of the residues predicted in the N- and C-terminal points of the rather long loop predictions are considered to be in distorted helical configurations, the above helical regions seem reasonable. This is especially obvious in the case of A7. The turn predictions throughout the sequence are substantial, possibly indicating distorted helices as have been observed, for example, in the insulin structure (Hodgkin, 1974). These deviations from

TABLE VIII: The Secondary Structure Prediction Histogram for Bovine Prothrombin Fragment 1.

```
        3    4         5         6         7         8         9
        5678901234567890123456789012345678901234567890123456789012345678
        LSATDAFWAKYTACESARNPREKLNECLEGNCAEGVGMNYRGNVSNTRSGIECQLWRSRYPHKP
(SH)[a] 333355555423332222113344443211   111111          233333321
(ST)[b]  11111111111     2222       1111111112333331112332 1112232121 22
(PJT)[c] |___A1____|  |S1|__B1__|L12|_A2_|___S2____|__B2__|_____

         1         1         1         1         1         1         1         1
        90        1         2         3         4         5         6         7
        90123456789012345678901234567890123456789012345678901234567890123456
        EINSTTHPGADLRENFCRNPDGSITGPWCYTTSPTLRREECSVPVCGQDRVTVEVIPRSGGSTTSOSPLLETCVPDRG
(SH)          122444444321              1233332211  1111112222          1122344442
(ST)  1133323443111  22233333233212211111         1322221    12323333331        12333
       L23__|__A3___|_____S3_____|__B3___|_L34__|__A4__|_S4__|__B4___|____

        1  1         1         2         2         2         2
        7  8         9         0         1         2         3
        789012345678901234567890123456789012345678901234567890123456789012345678
        REYRGRLAVTTSGSRCLAWSSEQAKALSKDNDFNPAVPLAENFCRNPDGDEEGAWCYVADNP
(SH)       111111    1111125555555555   1122334433221      233333111
(ST)  333322 111233321111111     122333332      1223333331111     1322
       L45|_A5__|_S5__|  |___B5____|_L56__|__A6___|_S6__|__B6____|___

        22        2         2         2
        34        5         6         7
        90123456789012345678901234567890123456789012345678901234
        GDFEYCNLNYCEEPVDGDLGDRLGEDPDPDAAIEGR
(SH)     111111222111                3333311
(ST)  23221    11   12333344443334333 1222
       L67|__A7___|____S7_____|_B7__|
```

[a] SH refers to the helical-prediction scores. [b] ST refers to the turn-prediction scores. [c] The helix–loop–helix configuration of a predicted EF hand is indicated by the symbols A-S-B for each of the seven hands. The interhand loops are designated by L.

TABLE IX: Results of the Application of the Tufty–Kretsinger Procedure to the Primary Sequence of Bovine Prothrombin Fragment 1.

Residue Region	Score[a]
4–32	7
33–61	5
64–92	5
106–134	8
144–172	6
175–203	5
211–239	6
245–273	4

[a] Score refers to the number of residues of a maximum of 16 satisfying the Tufty–Kretsinger criteria for a 29-residue segment.

the classic EF architecture might be expected given the phospholipid-binding properties of fragment 1 and the cooperative nature of calcium binding (Nelsestuen et al., 1975).

Residues 62–144 and 167–249 form two structures with mutual sequence homology (Magnusson et al., 1976) with 31 of 83 residues being identical. The prediction histogram shows structural homology in these regions as the first contains EF hands two and three while the second is made up of hands five and six.

Three spans are also predicted in a sheet configuration: 122–130 and 141–148 with prediction scores of three for each residue and 170–173 with scores of four. However, each of these regions has a significant and overlapping prediction for either turn or helix.

Application of the Tufty–Kretsinger test to the fragment 1 sequence gave an average background value of about 3 of 16 key residues. Table IX lists the test results for those regions proposed as EF hands which gave an average of 6 of 16 marker amino acids. It is interesting to note that the residue span 4–32 had a score of 7 of 16. Given the evolution and functional differences of fragment 1, it would not be expected to score as well as other strictly calcium-binding proteins. Though this proposal is highly speculative, it would appear to be more than merely fortuitous given that exactly seven EF-like secondary structural sequences were reasonably well predicted and that each region generally had a Tufty–Kretsinger score considerably larger than those for the other possible 29-residue segments in PT1.

Conclusions

In certain favorable cases, secondary-structure predictions methods are a useful adjunct to primary-sequence comparisons, especially in cases of ambiguous or tenuous alignment scores. Furthermore, prediction results can indicate variations in secondary structural elements with implications for function and evolutionary development. If the evolutionary distance is so large that the primary sequence homologies are not preserved, secondary-structural sequence required by biological function may still appear and be detected. However, as cautioned by Argos et al. (1976), prediction methods should not be used in a flippant manner but only where chemical evidence and "favorable" predictions suggest their application.

Kretsinger (1975) has recently listed many other calcium-binding proteins that may contain one or more EF hands. Among those noted as possible candidates are phosphodiesterase, phosphorylase b kinase, luciferin, aequorin, Nbs_2, ethylenediaminetetraacetic acid extractable light chain, and so forth. The application of secondary-structure prediction schemes to these primary sequences when they become available should prove useful.

Acknowledgments

The author is extremely grateful for helpful discussions with Drs. Robert Kretsinger of the University of Virginia, David Barry of the Washington University Medical School, and especially Thomas Vanaman of the Duke University Medical Center. Computer programs for prediction schemes were obtained from Drs. K. Nagano, D. Barry, and A. Burgess. John and James Schwarz helped write and assemble programs applying the methods of Chou and Fasman, and Kabat and Wu. The Southern Illinois University computer staff provided considerable technical assistance. Sharon Schaefer, Sharon Wilder, and Karen Babos helped prepare the manuscript.

References

Argos, P., Schwarz, J., and Schwarz, J. (1976), *Biochim. Biophys. Acta 439,* 261–273.

Bajaj, S. P., Butkowski, R. J., and Mann, K. G. (1975), *J. Biol. Chem. 250,* 2150–2156.

Burgess, A. W., Ponnuswamy, P. K., and Scheraga, H. A. (1974), *Isr. J. Chem. 12,* 483–494.

Capony, J. P., Ryden, L., Damaille, J., and Pechere, J. F. (1973), *Eur. J. Biochem. 32,* 97–104.

Chou, P. Y., and Fasman, G. D. (1974a), *Biochemistry 13,* 211–222.

Chou, P. Y., and Fasman, G. D. (1974b), *Biochemistry 13,* 222–245.

Coffee, C. J., and Bradshaw, R. A. (1973), *J. Biol. Chem. 248,* 3305–3312.

Collins, J. H. (1974), *Biochem. Biophys. Res. Commun. 58,* 301–308.

Collins, J. H. (1975), *Fed. Proc., Fed. Am. Soc. Exp. Biol. 34,* 539.

Collins, J. H., Potter, J. D., Horn, M. J., Wilshire, G., and Jackman, N. (1973), *FEBS Lett. 36,* 268–272.

Davie, E. W., Hougie, C., and Lundblad, R. L. (1969), in Recent Advances in Blood Coagulation, Poller, L., Ed., London, Churchill, pp 13–28.

Ebashi, S. (1972), *J. Biochem. (Tokyo) 72,* 787–790.

Enfield, D. L., Ericsson, L. H., Walsh, K. A., Neurath, H., and Titani, K. (1975), *Proc. Natl. Acad. Sci. U.S.A. 72,* 16–19.

Frankenne, F., Joassin, K., and Gerday, C. (1973), *FEBS Lett. 35,* 145–147.

Fujikawa, K., Coan, M. H., Enfield, D. L., Titani, K., Ericsson, L. H., and Davie, E. W. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 427–430.

Greaser, M. L., and Gergely, J. (1973), *J. Biol. Chem. 248,* 2125–2133.

Hartshorne, D. J., and Pyun, H. Y. (1971), *Biochim. Biophys. Acta 229,* 698–711.

Hodgkin, D. C. (1974), *Proc. R. Soc. London, Ser. B 186,* 191–215.

Kabat, E. A., and Wu, T. T. (1974), *Proc. Natl. Acad. Sci. U.S.A. 71,* 4217–4220.

Kretsinger, R. H. (1975), in Calcium Transport in Contraction and Secretion, Carafoli, E., Ed., Amsterdam North-Holland Publishing Co. pp. 469–478.

Kretsinger, R. H., and Barry, C. D. (1975), *Biochim. Biophys. Acta 405,* 40–52.

Kretsinger, R. H., and Nockolds, C. E. (1973), *J. Biol. Chem. 248,* 3313–3326.

Lim, V. I. (1974a), *J. Mol. Biol. 88,* 857–872.
Lim, V. I. (1974b), *J. Mol. Biol. 88,* 873–894.
Magnusson, S., Peterson, T. E., Sottrup-Jensen, L., and Claeys, H. (1975), in Proteases and Biological Control, Reich, E., Rifkin, D. B., Shaw, E., Eds., Cold Spring Harbor, New York, Cold Spring Harbor Laboratory, pp 123–150.
Magnusson, S., Sottrup-Jensen, L., Claeys, H., Peterson, T. E., and Zajdel, M. (1976), in Abstracts, The Eighth Miami Winter Symposia, Ribbons, D. W., Brew, K., Ahmad, F., Schultz, J., Eds., Miami, Florida, University of Miami, pp 30–34.
Majerus, P. W. (1968), *Science 159,* 428–430.
Matthews, B. W. (1975), *Biochim. Biophys. Acta 405,* 442–451.
Moffat, K., Fullmer, C. H., and Wasserman, R. H. (1975), *J. Mol. Biol. 97,* 661–664.
Nagano, K. (1973), *J. Mol. Biol. 75,* 401–421.
Nagano, K. (1974), *J. Mol. Biol. 84,* 337–372.
Nelsestuen, G. L., Broderius, M., Zytkovicz, T. H., and Howard, J. B. (1975), *Biochem. Biophys. Res. Commun. 65,* 233–240.
Potter, J. D., and Gergely, J. (1974), *Fed. Proc., Fed. Am. Soc. Exp. Biol. 33,* 1465.
Rossmann, M. G., Moras, D., and Olsen, K. W. (1974), *Nature (London) 250,* 194–199.
Sauer, F., Pugh, E. L., Wakil, S. J., Delaney, R., and Hill, R. L. (1964), *Proc. Natl. Acad. Sci. U.S.A. 52,* 1360–1366.
Schulz, G. E., Barry, C. D., Friedman, J., Chou, P. Y., Fasman, G. D., Finkelstein, A. V., Lim, V. I., Ptitsyn, O. B., Kabat, R. A., Wu, T. T., Levitt, M., Robson, B., and Nagano, K. (1974), *Nature (London) 250,* 140–142.
Schulz, H. (1972), *Biochem. Biophys. Res. Commun. 46,* 1446–1453.
Stenflo, J. (1974), *J. Biol. Chem. 249,* 5527–5535.
Stenflo, J., Fernlund, P., Egan, W., and Roepstarff, P. (1974), *Proc. Natl. Acad. Sci., U.S.A. 71,* 2730–2733.
Szent-Gyorgyi, A. G., Szentkiralyi, E. M., and Kendrick-Jones, J. (1972), *Cold Spring Harbor Symp. Quant. Biol. 37,* 47–53.
Szent-Gyorgyi, A. G., Szentkiralyi, E. M., and Kendrick-Jones, J. (1973), *J. Mol. Biol. 74,* 174–203.
Tufty, R. M., and Kretsinger, R. H. (1975), *Science 187,* 167–169.
Vanaman, T. C., Wakil, S. J., and Hill, R. L. (1968a), *J. Biol. Chem. 243,* 6409–6419.
Vanaman, T. C., Wakil, S. J., and Hill, R. L. (1968b), *J. Biol. Chem. 243,* 6420–6431.
Van Eerd, J., and Takahashi, K. (1975), *Biochem. Biophys. Res. Commun. 64,* 122–127.
Weber, A., and Murray, J. M. (1973), *Physiol. Rev. 53,* 612–673.
Weber, M. M., Gaffin, S. L., and Oplatka, A. (1972), *J. Mechanochem. Cell Motil. 1,* 91–96.
Weeds, A. G. (1967), *Biochem. J. 105,* 25C–27C.
Weeds, A. G., and Frank, G. (1972), *Cold Spring Harb. Symp. Quant. Biol. 37,* 9–16.
Weeds, A. G., and McLachlan, A. D. (1974), *Nature (London) 252,* 646–649.

Interactions of Platinum Complexes with the Essential and Nonessential Sulfhydryl Groups of Thymidylate Synthetase†

John L. Aull,* Allison C. Rice,‡ and Laura A. Tebbetts

ABSTRACT: Thymidylate synthetase (methylenetetrahydrofolate:deoxyuridylate *C*-methyltransferase) from *Lactobacillus casei* was progressively inactivated when incubated at 25 °C, pH 6.8, in the presence of *trans*-$Pt(NH_3)_2Cl_2$. The inhibition appeared to be irreversible, and the rate of activity loss was dependent on the inhibitor concentration. The corresponding cis isomer was incapable of inhibiting the enzyme under the same conditions. The presence of 2-mercaptoethanol protected the enzyme from inhibition, but did not reactivate enzyme preparations which had been inhibited prior to the addition of the thiol. The interactions of *cis*- and *trans*-$Pt(NH_3)_2Cl_2$ with the enzyme's sulfhydryl (-SH) groups were inferred from the results of spectrophotometric titrations of the enzyme with 5,5′-dithiobis(2-nitrobenzoic acid) and *p*-hydroxymercuribenzoate. The results suggested that the cis isomer reacted with an average of 1.3 of the enzyme's 4 -SH groups and that these were not essential for catalysis. The trans isomer reacted with a total of approximately 2.5 -SH groups, 1.2 of which are essential for catalysis. Neither the trans isomer nor a combination of both isomers was able to react with 1.2 of the 4 -SH groups. Further evidence that the Pt complexes are interacting with enzyme's -SH groups was obtained by reversibly blocking the -SH groups of thymidylate synthetase, and demonstrating the resistance of these preparations to inhibition by the trans Pt complex. Possible explanations for the preferential inhibition of thymidylate synthetase by only one of the two geometric isomers of $Pt(NH_3)_2Cl_2$ are considered.

Thymidylate synthetase catalyzes the reductive methylation of 2′-deoxyuridylate by (+)-5,10-methylenetetrahydrofolate to form 2′-deoxythymidylate and 7,8-dihydrofolate (Huennekens, 1968; Blakely, 1969) as shown in eq 1.

$$\text{dUMP} + (+)\text{-5,10-CH}_2\text{H}_4\text{-folate} \rightarrow \text{dTMP} + \text{dihydrofolate}^{1} \quad (1)$$

† From the Department of Chemistry, Auburn University, Auburn, Alabama 36830. *Received June 23, 1976.* Supported by a grant from Research Corporation and an Auburn University Grant-in-aid.

‡ Present address: Department of Microbiology, Vanderbilt University, Nashville Tennessee.

[1] Abbreviations used are: 5,10-CH_2H_4-folate, 5,10-methylenetetrahydrofolate; FdUMP, 5-fluoro-2′-deoxyuridylate; pHMB, *p*-hydroxymercuribenzoate; DTNB, 5,5′-dithiobis(2-nitrobenzoic acid); H_4-folate, 5,6,7,8-tetrahydrofolate; $NaDodSO_4$, sodium dodecyl sulfate.