Volume 274, number 1,2, 217-222 FEBS 09110 November 1990

Computer simulation of zinc finger motifs from cellular nucleic acid binding protein and their interaction with consensus DNA sequences

V. Kothekar

Department of Biophysics, All India Institute of Medical Sciences, New Delhi-110029, India

Received 16 July 1990; revised version received 17 August 1990

We report here a computer simulation of the three-dimensional structures of seven zinc finger motifs from cellular nucleic acid binding protein involved in negative feedback inhibition of cholesterol biosynthesis. The structures are optimised using steric constraints imposed by tetrahedral coordination of the zinc ion with Cys and His residues, by molecular mechanics technique. We have also optimised the structure of a finger-I with GpT sequence. The model for the interaction of seven fingered protein with single-stranded d(GTGCGGTG) from sterol regulatory element (SRE) is given on the basis of these results. We also propose a scheme for recognition of a multifingered regulatory protein with small single-stranded DNA fragments.

Zinc finger; CNBP protein; DNA binding; Computer modelling; Molecular mechanics

1. INTRODUCTION

Recently Tripathi et al. [1] isolated a 5-kDa positive clone cellular nucleic acid binding protein (CNBP) using oligonucleotide probes containing HMG CoA-reductase SRE. This protein could function in transcription regulation by sterols by selectively interacting with single-stranded DNA fragments d(GTGCGGTG) from human and Hamster HMG CoA reductase SRE, d(GTGGGGTG) from LDL receptor SRE, d(CTGGGGTG) from mouse apolipoprotein promoter, and d(GTGGCGCG) from fernesyl pyrophosphate reductase SRE. It did not bind to SRE like octanucleotide sequences d(GTGGGTAT), d(CATAGGTG), d(CATATCAT), which fail to confer sterol responsiveness [1]. CNBP protein is characterized by seven tandem repeats of 14 amino acids each with regular occurrence of Cys-Cys-His-Cys. The chemically similar amino acids such as Arg/Lys, Asp/Glu, Ser/Thr occupy identical positions. The protein has two, four and four residues in between Cys-Cys, Cys-His, and His-Cys. Although it belongs to the general class of 'zinc finger' proteins [2–8], it differs from the well-studied C2H2, C4 or C6 class. The positions of Cys/His have some resemblance to metalloproteins rubredoxin [9], aspartate transcarbamoylase [10] and azurin (Cu) [11]. It differs from TFIIIA in the number of intervening residues between different repeats/fingers which varies from 34 to 4. (TFIIIA has 9 continuous repeats of about 30 amino acids each [2].)

Correspondence address: V. Kothekar, Department of Biophysics, All India Institute of Medical Sciences, New Delhi-110029, India

2. MATERIALS AND METHODS

2.1. *Finger simulation*

Simulation of three-dimensional structure of the fingers with two flanking amino acids on both the sides (a sequence of 18 amino acids) was done with emphasis on steric constraints and energetics. The crude modelling was accomplished in three steps. (i) The fragments x-x-Cys-x-x-Cys, Cys-x-x-x-x-His and His-x-x-x-x-Cys-x-x (where x is Ala) were simulated on the basis of known structural data on similar fragments in metalloproteins [9–11]. (ii) They were constrained to give SG-SG, SG-NE2 and NE2-SG distance of 4.5 Å without any short contacts between the side chains or the backbone. (iii) The positively charged Zn^{2+} ion was introduced and the three fragments were arranged so as to yield tetrahedral coordination of SG, SG, NE2 and SG about it. The fragments were allowed to join smoothly by damped least square energy minimization (DLS) technique [12]. The refinement of the structures of seven fingers was achieved by Molecular Mechanics techniques using AMBER 3.0 [13] adopted for CYBER 180/930 system, using united atom force field parameters with distance dependent di-electric constant. The Zn atom was not considered explicitly but only as a doubly positive charge. The geometry around it was maintained by introducing distance and angle constraints. For this purpose a pseudomolecule was generated with Zn–S and Zn–NE2 distance 2.3 and 2.0 Å, respectively. Angle S–Zn–S or S–Zn–NE2 was kept at 109°. The zinc binding ligands from the pseudomolecule and the finger were overlapped. The tetrahedral constraint energy was calculated as the sum of squares of deviation in the cartesian coordinates of the zinc binding ligands in the pseudomolecule and finger. To get the energy equivalent this was multiplied by a weighting factor of 100. The structure of each finger was minimized in 3000–5000 cycles. The final structure had AMBER energy less than -267 kcal/mol. The structures were stabilized mainly by electrostatic and H-bonding interactions. The tetrahedral constraint variation energy was less than 12 kcal/mol, and the RMS deviation from ideal tetrahedron was 0.014. These values can be compared with those for the first finger from *Xenopus* protein XFIN/8/where the AMBER optimised energy for 25 amino acid fragment was -304 kcal/mol. The starting data in this case were taken from NMR spectroscopy. Almost all the backbone torsional angles (not shown here) lie in the allowed range. Those which are outside are mostly Gly. Only five torsional angles for non-glycine residues lie outside this range. They are in the bend regions.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

AMINO ACID SEQUENCE OF ZINC FINGER MOTIFS OF HUMAN CLONE CELLUAR NUCLEIC ACID BINDING PROTEIN (CNBP)

	1									10							
4	Asn	Glu	Cys	Phe	Lys	Cys	Gly	Arg	Ser	Gly	His	Trp	Ala	Arg	Glu	Cys	Pro	Thr
52	Asp	Ile	Cys	Tyr	Arg	Cys	Gly	Glu	Ser	Gly	His	Leu	Ala	Lys	Asp	Cys	Asp	Leu
72	Asp	Ala	Cys	Tyr	Asn	Cys	Gly	Arg	Gly	Gly	His	Ile	Ala	Lys	Asp	Cys	Lys	Glu
96	Gln	Cys	Cys	Tyr	Asn	Cys	Gly	Lys	Pro	Gly	His	Leu	Ala	Arg	Asp	Cys	Asp	His
117	Gln	Lys	Cys	Tyr	Ser	Cys	Gly	Glu	Phe	Gly	His	Ile	Gln	Lys	Asp	Cys	Thr	Lys
135	Val	Lys	Cys	Tyr	Arg	Cys	Gly	Glu	Thr	Gly	His	Val	Ala	Ile	Asn	Cys	Ser	Lys
156	Val	Asn	Cys	Tyr	Arg	Cys	Gly	Glu	Ser	Gly	His	Leu	Ala	Arg	Glu	Cys	Thr	Ile

Fig. 1. Amino acid sequence of CNBP in the seven repeats of 18 amino acids each. The sequence number of the starting amino acid of each repeat is given at the beginning. The sequence number of the residue from the beginning of the finger is given at the top of the first sequence. One can note the regular occurrence of Cys at position 3, 6 and 16, and His at 11. We also note Arg/Lys at position 14, Glu/Asp at 15, and Phe/Tyr at 4.

2.2. *DNA binding*

The octanucleotide sequence d(GTGCGGTG) was simulated in B form using coordinates from Arnott et al. [14] since there were no experimental data on this sequence. Theoretical conformational considerations and X-ray crystallographic data [15,16] indicate that small oligonucleotide sequences exist in B-conformation. There were two reasons for choosing single-stranded over double-stranded DNA: (i) known affinity of CNBP an estrogen receptor (hER) for single-stranded DNA [1,17]; (ii) topological considerations. Single-stranded DNA has more room for binding to a 177 amino acid protein. The interaction of finger-I with GpT was optimised using our IMF program [12].

3. RESULTS AND DISCUSSION

3.1. *Stereochemical aspect of zinc fingers*

The amino acid sequences of the seven zinc fingers of the CNBP protein were highly homologous (Fig. 1). As a result of this their three-dimensional structures showed several common stereochemical features. The most striking was the helical region extending between residues 1 and 6 (Fig. 2) in all the seven fingers. The helix was distorted and could be classified as in between α and 3_{10} on the basis of H-bonding scheme (H bonds were observed between $i,i+3$, and $i,i+4$ residues) and backbone torsional angles which were in between those for α and 3_{10} helix. In case of XFIN-31 [8] the helix extended from residues 12 to 24, (twice as long as CNBP) and was a mixture of α and 3_{10} helix. The structures in CNBP and XFIN-31 had some additional similarities, viz. orientation of polar and hydrophobic amino acid side chains. Vth position in all the seven fingers was occupied by the polar amino acids Lys, Arg, Asn, Ser while second residue was polar in the case of first and III–VII fingers. These projected away from the finger geometry similar to XFIN-31. Hydrophobic residues Leu, Ile, Val, Trp were buried inside. Trp-15 in CNBP-I stacked on His-14 (Fig. 2). As in the case of XFIN-31, CNBP fingers were characterized by two loose bends between residues 7–10 and 13–16, respectively, with no specific interchain H-bonds between them (Fig. 3). The 8th residue at the tip of BEND-I was basic in fingers I, III and IV and acidic in other cases. While the 9th residue could be both basic and hydrophobic. Contrary to this at the tip of BEND-II, the 14th position is occupied by basic amino acids Arg/Lys in all except finger VI while the 15th position is occupied by Glu/Asp. Positively and negatively charged side chains at the BEND-II came closer than those at the BEND-I because of the electrostatic interaction between them. As a result the structure is narrower towards BEND-II and wider towards BEND-I like a 'flower petal' with zinc ion in the middle of compact globular structure.

Because of the two bends, amino- and carboxy-terminals point in opposite directions and the whole structure looks like a left hand palm (Fig. 3) with carboxy-terminal pointing towards the thumb. While XFIN-31 resembled right hand palm. The difference was mainly because of the position of the helix. The bends were distorted and did not belong to any classical types. A typical view of the amino acids in the proximity of Zn ion is shown in Fig. 4. The average Zn–S and Zn–Ne2 distance was 2.4 and 2.02 Å and the angle X–Zn–X equal to 108.

3.2. *DNA binding*

The only way of interacting the seven-fingered, 177 amino acid CNBP protein with the octanucleotide CORE sequences from SRE is to follow a 'petalloid' arrangement where each finger binds to one DNA base.

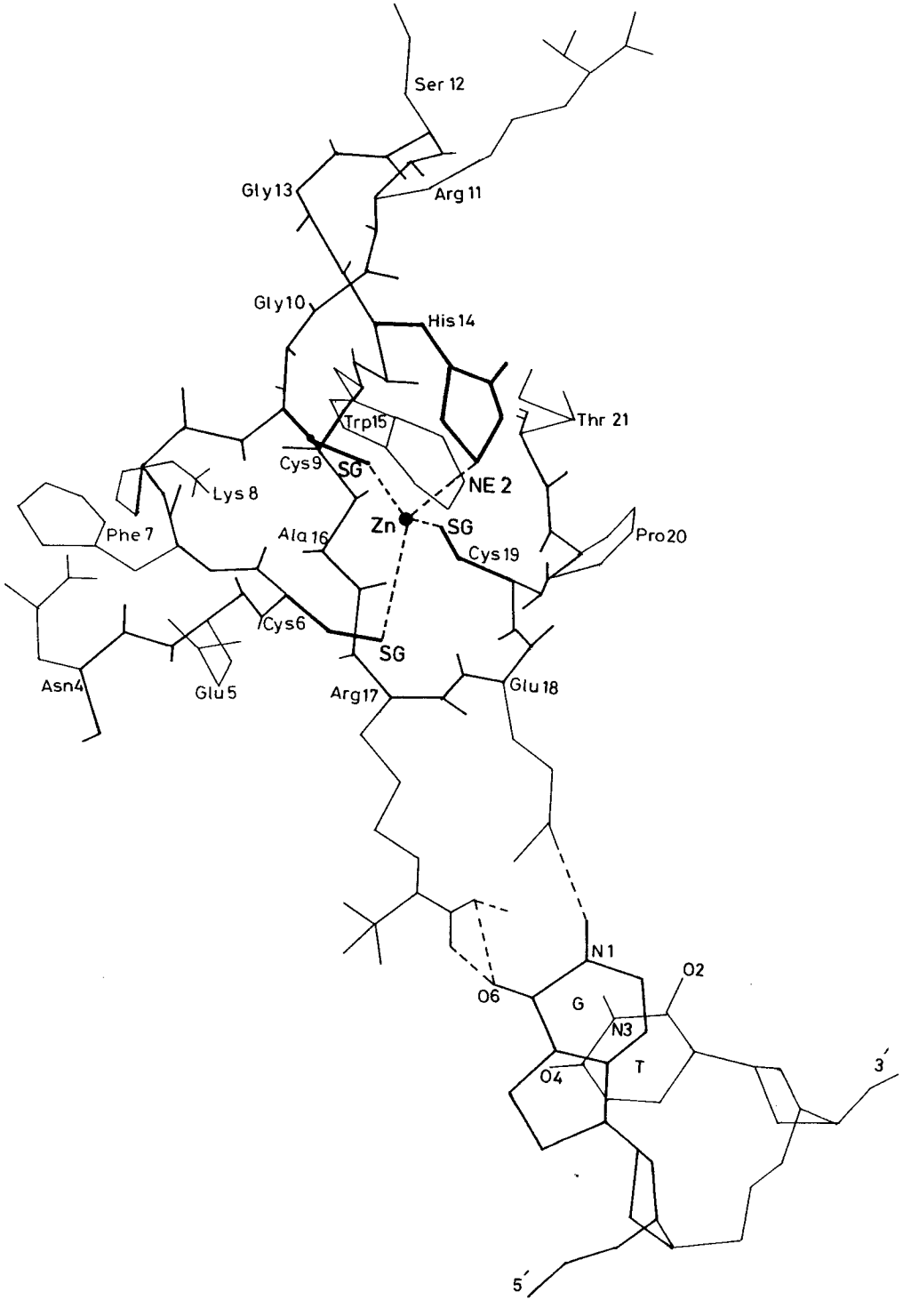

Fig. 2. Energy minimised model for the interaction of finger I with single-stranded GpT. H-bonds between residues 14 and 15 in the BEND-II with G are shown by dotted lines.

We attempted interacting both BEND-I and BEND-II of the protein with the DNA bases (models I and II, respectively). Although we could make few specific H-bonding contacts in model I, it did not yield a sterically feasible model since there was not enough room to accommodate the adjacent fingers. We discarded this model also because of the lesser number of H-bonding contacts it made with the DNA bases. The model II with Bend-II interacting with the DNA base (Fig. 2) seemed sterically feasible. Two sets of bifurcated H-bonds could be formed in this case between Arg with 06 of guanine and Glu with N2H2 of guanine. The rest of the

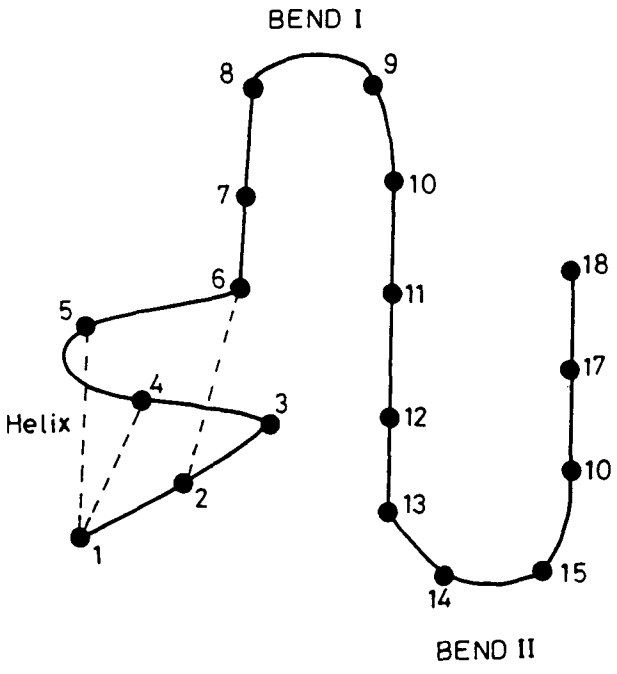

Fig. 3. Schematic representation of the secondary structural elements of CNBP fingers. The sequence number of the amino acid starts from the beginning of the finger. The H-bonds are depicted by dotted lines.

finger projects out with polar residues at BEND-I facing the exterior. The SRE core sequence d(GTGCGGTG) is rich in guanine and thymines. The inner binding sites of these nucleotides from the major groove side (06 of G, 04 of T, and N1H of G and N3H of T) are chemically identical. The only difference is that they are closer to the backbone in thymine. We have alternately spaced acceptors and donors (06/04,N1H/N3H) which can match with the alternately spaced donors and acceptors in the BEND-II in the model II. There is an extra base in the octanucleotide SRE sequence. This is essential for accommodating a large loop of 34 amino acids between finger I and II. This loop also has Arg and Ser in the 10th and 11th position which can make contact with the DNA base in a manner similar to the fingers. A schematic representation of the protein DNA

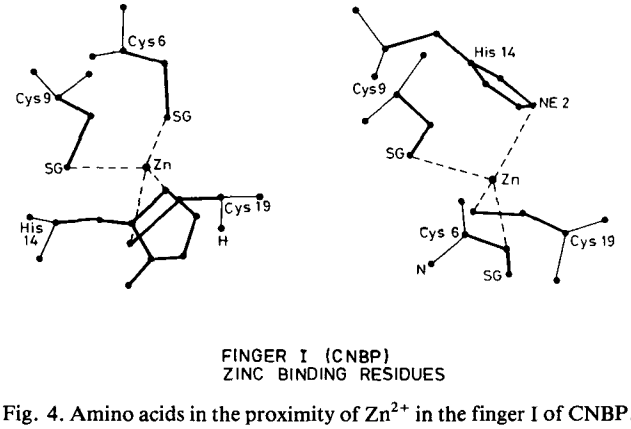

Fig. 4. Amino acids in the proximity of Zn^{2+} in the finger I of CNBP.

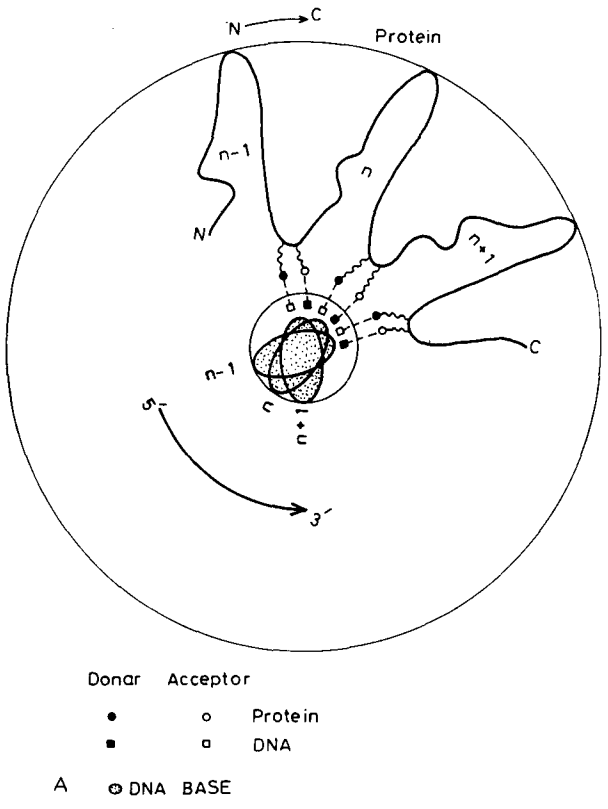

Fig. 5A.

binding scheme is shown in Fig. 5A. A complete model for seven-fingered CNBP protein with octanucleotide CORE sequence is presented in Fig. 5B. It is obtained by extrapolation of the model for finger I with GpT. Here the seven fingers along with a 34 amino acid loop contact the eight bases in octanucleotide SRE core sequence. It can be seen from Fig. 5A that the donor acceptor correspondence between protein requires that $i+1$ finger must interact with i-1 nucleotide base. In other words the direction of protein and DNA should be antiparallel, or protein N-C direction should be parallel to DNA 3′-5′ direction. Thus finger I contacts eighth base (G8), loop the 7th base (T7), finger II 6th (G6), finger III 5th (G5), and so on. There is no 'tangling' of the protein backbone.

A complete recognition scheme was developed on the basis of the proposed model. We give a score of + if one donor acceptor criterion is satisfied by purines and + + if both criteria are satisfied by the residues of position 14 and 15 in the BEND-II. If guanine is replaced by thymine we reduce the score to half for steric reasons. Zero score is given if neither criteria are satisfied. The loop is more flexible than the fingers. The donor-acceptor sequence at the tip of the loop is slightly different. Arg can make H-bond with thymine 04, Ser oxygen may contact N3H of thymine, while hydrogen may obstruct it. Hence we give a score of + for the loop. In the case of the VIth finger, the 14th residue at the BEND-II is Ile and not basic as in other cases. The 15th position is oc-

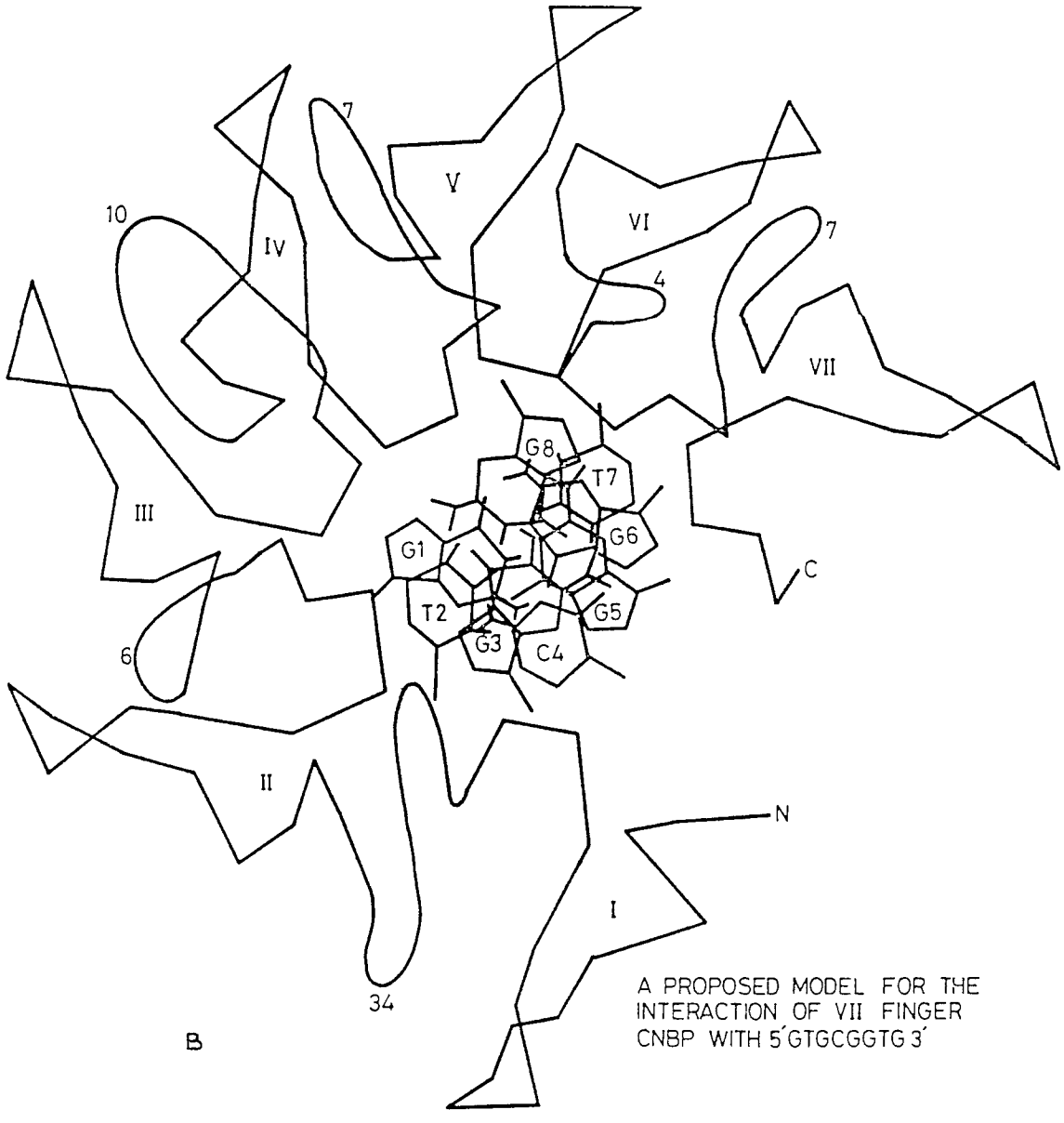

Fig. 5.(A) A schematic representation of the binding of three successive fingers to three successive DNA bases. The direction of growth of the protein is in the clockwise direction. The positively and negatively charged residues are placed alternately. The electron acceptors and donors from guanine and thymine are also placed alternately on a arc of a circle but increase in anticlockwise direction. As a result, the next finger tip can interact only with the preceding base. In other words, N-C direction of the protein runs parallel to 3'-5' direction of DNA. (B) A proposed model for the interaction of 7-fingered CNBP protein with octanucleotide CORE sequence 5'-GTGCGGTC- 3' obtained by extrapolation of the model for interaction of Finger I with GpT. The residues at the tip of the BEND-II (Arg/Cys, Asp/Gly) of finger I H-bond with 06, N1H of G8. Residues (Arg and Ser) from the loop H-bond with T7, terminal residues at the BEND-II of fingers II to VII H-bond with G6, G5, C4, G3, T2, G1. There is no 'tangling' of the protein backbone.

cupied by Asn with donor as well as acceptor atoms. These two can form a pair of hydrogen bonds with guanine or thymine. We give here a score of + since only one amino acid is involved. Table I gives the summary of the recognition scheme of CNBP by seven octanucleotide sequences GTGCGGTG (DNA1), GTGGGGTG (DNA2), CTGGGGTG (DNA3), GTGGCGCG (DNA4), GTGGGTAT (DNA5), CATAGGTG (DNA6) and CATATCAT (DNA7). The highest score of 14 is arrived at in the case of DNA2, an octanucleotide core sequence from LDL receptor promoter, followed by 13 for DNA1 in HMG Co-A reductase SRE. The lowest score is for sequences DNA6 and DNA7 which do not offer sterol responsiveness. Thus, the scheme seems to work well.

Acknowledgments: The author is thankful for the Department of Science and Technology of the Government of India for financial support and Director AIIMS for use of CYBER 180/930 system. We are also thankful to Prof. P.A. Kollman for sending us a copy of AMBER 3.0.

Table I

A scheme for recognition of DNA sequences by CNBP

CNBP	Fing. I	Loop 1	Fing. II	Fing. III	Fing. IV	Fing. V	Fing. VI	Fing. VII	score
	D A	D D	D A	D A	D A	D A	H D A	D A	
DNA 3′5′A	D	A D	A D	A D	A D	A D	A D	A D	
DNA1	G	T	G	G	C	G	T	G	
	+ +	+	+ +	+ +	+	+ +	+	+ +	13
DNA2	G	T	G	G	G	G	T	G	
	+ +	+	+ +	+ +	+ +	+ +	+	+ +	14
DNA3	G	T	G	G	G	G	T	C	
	+ +	+	+ +	+ +	+ +	+ +	+		12
DNA4	G	C	G	C	G	G	T	G	
	+ +		+ +		+ +	+ +	+	+ +	11
DNA5	T	A	T	G	G	G	T	G	
	+		+	+ +	+ +	+ +	+	+ +	11
DNA6	G	T	G	G	A	T	A	C	
	+ +	+	+ +	+ +		+			8
DNA7	T	A	C	T	A	T	A	C	
	+			+		+			3

D = Donor NH-Arg,Lys,N1H-G,N3H-T; A = Acceptor O-Glu,Asp.06-G,04-T.
DNA1 = d(GTGCGGTG); DNA2 = d(GTGGGGTG); DNA3 = d(CTGGGGTG); DNA4 = d(GTGGCGCG); DNA5 = d(GTGGGTAT); DNA7 = d(CATATCAT).
Fing. = Finger, H = Hydrophobic residue.

REFERENCES

[1] Tripathi, B.R., Taylor, A.K., Andalibi, A., Svenson, K.L. and Lusis, A.J. (1989) Science 245, 640–643.
[2] Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J. 4, 609–614.
[3] Rhodes, D. and Klug, A. (1986) Cell 46, 123–132.
[4] Berg, J.M. (1986) Science 232, 485–487.
[5] Berg, J.M. (1988) Proc. Natl. Acad. Sci. USA 85, 99–102.
[6] Evans, R.M. and Hollenberg, S.M. (1988) Cell 52, 1–3.
[7] Freedman, L.P., Luisi, B.F., Korszun, Z.R., Basavappa, R., Sigler, P.B. and Yamamoto, K.R. (1988) Nature 334, 543–546.
[8] Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A. and Wright, P.E. (1989) Science 245, 635–637.
[9] Watenpaugh, K.D., Sieker, L.C. and Jensen, L.H. (1980) J. Mol. Biol. 138, 615–633.
[10] Honzatko, R.B., Crawford, J.L., Monaco, H.L., Ladner, J.E., Ewards, B.F.P., Evans, D.R., Warren, S.G., Wiley, D.C., Ladner, R.C. and Lipscomb, W.N. (1982) J. Mol. Biol. 160, 219–263.
[11] Bernstein, F.C., Koetze, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) J. Mol. Biol. 112, 535–554.
[12] Kothekar, V. and Mrigank (1988) J. Phys. Educ. 5, 169–175.
[13] Singh, U.C., Weiner, P.K., Caldwell, J. and Kollman, P.A. (1986) AMBER (UCSF version 3.0).
[14] Arnott, S., Smith, P.J.C. and Chandrasekaran, R. (1976) in: CRC Handbook of Biochemistry and Molecular Biology, vol. 2 (Fasman G.D. ed.) pp. 411–422, CRC, Cleveland, OH.
[15] Pullman, B. and Saran, A. (1976) Prog. Nucleic Acids Res. Mol. Biol. 18, 216–322.
[16] Kalia, A. and Kothekar, V. (1986) Ind. J. Biochem. Biophys. 23, 174–178.
[17] Lannigan, D.A. and Notides, A.C. (1989) Proc. Natl. Acad. Sci. USA 86, 863–867.