ROLE OF PROTEIN BACKBONE IN SPECIFIC RECOGNITION OF NUCLEIC ACID BASE SEQUENCES: A HYPOTHESIS

RAMAKRISHNA V. HOSUR

Chemical Physics Group, Tata Institute of Fundamental Research, Homi Bhabha Road, Bombay 400 005, *India*

ABSTRACT

A novel H-bonding scheme of interaction between protein backbone and nucleic acid base pairs has been proposed. The importance of such an interaction in specific recognition of base sequences in double helical nucleic acids has been discussed. It is concluded that protein backbone can play very important role in specific recognition of base sequences.

1. INTRODUCTION

THE specific recognition of nucleic acid base sequences by proteins is a fundamental process at several steps of genetic expression. Several attempts have been made[1-12] to understand this key phenomenon. It has been suggested that proteins can interact with nucleic acids in four different ways[1], *viz.*, via electrostatic interaction, stacking, hydrogen bonding and hydrophobic interaction. Among these, hydrogen bonding interaction between amino acid residues and nucleic acid bases is, by far, the most efficient process in the specific recognition of nucleic acid bases, because of the directional nature of the H-bond. Several H-bonding schemes have been suggested and these are centered on specific interactions between amino acid side chains and bases in DNA. It has been suggested that at least two H-bonds are needed for specific recognition of a base pair[3]. In this communication we propose a new scheme of H-bonding interaction involving protein backbone and nucleic acid base pairs and discuss its role in specific recognition.

2. Scheme of Interaction

The proposed scheme of interaction with two H-bonds across a base pair is shown in Fig. 1. It is seen that two types of interactions are possible depending upon the direction of the protein backbone in the ring formed by H-bonds. If one chooses the C=O group of the backbone to bind to the $-NH_2$ group of adenine (cytosine) while going from the N-terminal to the C-terminal, an eight membered ring is formed. For the reverse direction of the backbone, a nine membered ring is formed. It is also clear from Fig. 1 that for A–U (T) and U (T)–A base pairs such an interaction with protein backbone can occur only in the major groove of the helix whereas G–C and C–G base pairs have binding sites in both the grooves. The directions of the various possible H-bonds for the different base pairs are schematically shown in Fig. 2. We have built molecular models with different homopolymeric peptides interacting through their backbone with the base pairs in a double helical nucleic acid. Fig. 3 shows as an example Poly (L-Ser) bound to two successive base pairs in the major groove of (A–U) double helix. Some qualitative conclusions may be drawn from these studies and these are as follows. The major and minor grooves of DNA double helix have very different characteristics. While both eight and nine membered rings can easliy be formed in the major groove, there are severe steric restrictions from the sugar ring at the pyrimidine sidh in the minor groove. We shall therefore henceforte

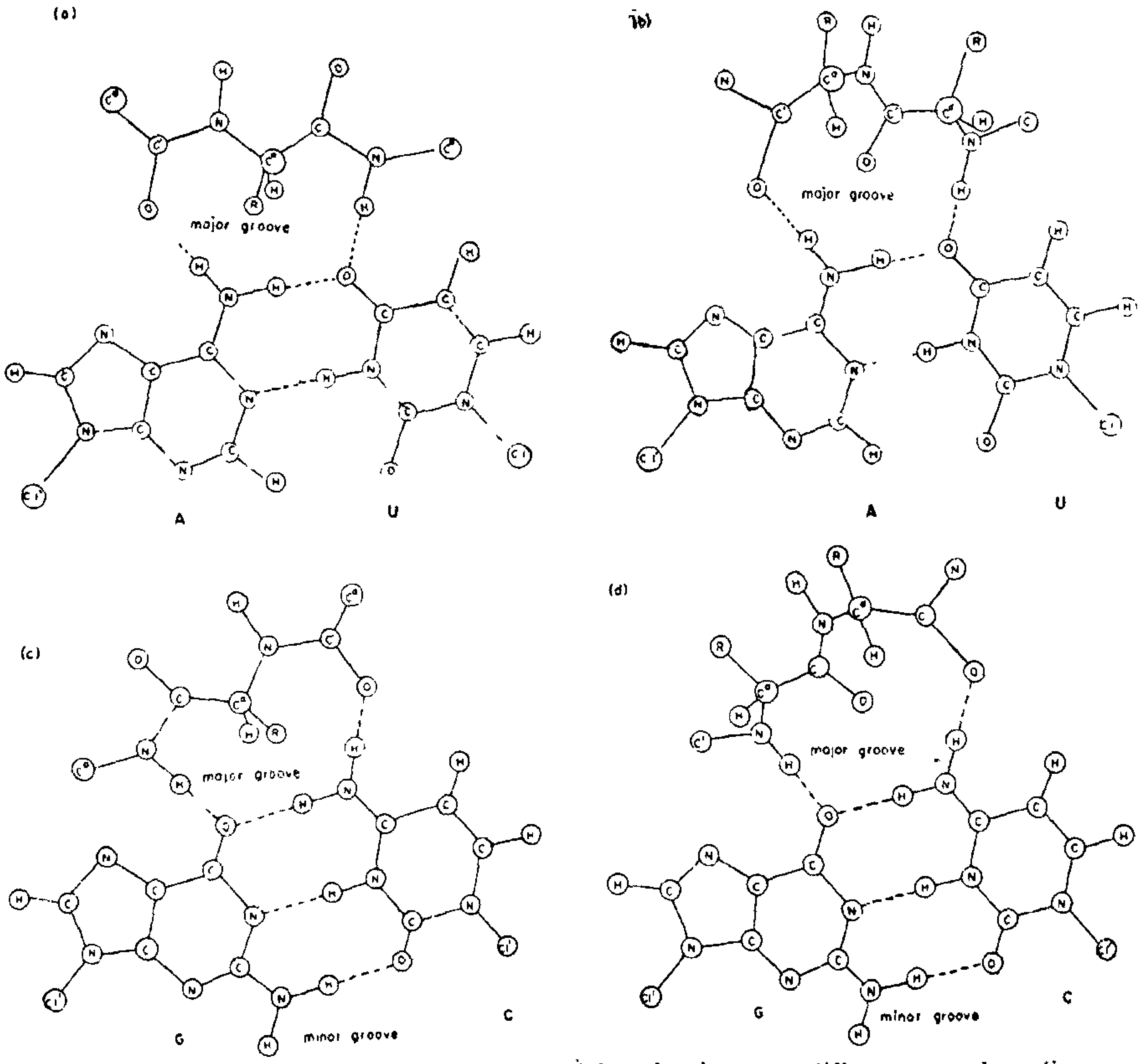

Fig. 1. Binding schemes of base pairs with proteins showing two different ways depending upon the direction of protein backbone : in (*a*) and (*c*) an eight membered ring is formed and in (*b*) and (*d*) nine membered ring is formed.

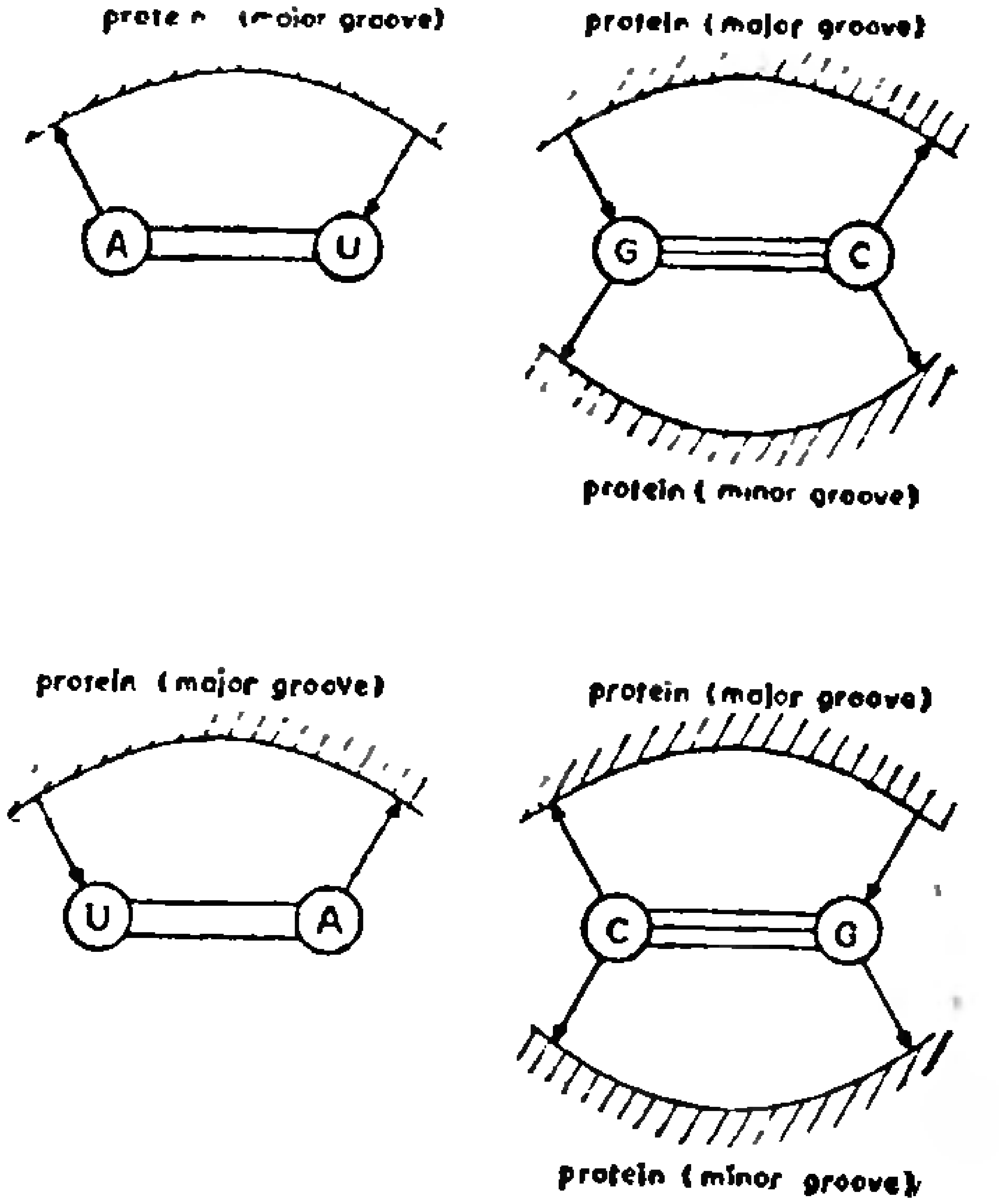

FIG. 2. Directions of the different H-bonds between different base pairs and protein backbone. The arrow indicates the vector (N → O).

restrict our attention only to the major groove. It is generally observed that the peptide backbone loops out considerably between two successive base pairs. The extent of looping depends upon the combination of eight and nine membered rings, between the two pairs. The conformations of the peptide backbone are different in the bound and unbound regions of the chain and are dependent on the nature of the side chain. In the nine mebmered ring the side chain protrudes out and thus poses no steric hindrances, while an eight membered ring is sterically disallowed if the side chain is bulkier than $-CH_2OH$ group (*i.e.*, Ser side chain). The conformations of the peptide backbone in the eight and nine membered rings are different. Table I lists the domains of allowed conformations. A domain covers a range of about 20° with respect to each dihedral angle in the main chain. Some of these fall within the allowed regions in the (ϕ, ψ) maps[13]. It is seen that, while only one set of (ϕ, ψ) is important for an eight membered ring, two sets are crucial for a nine membered ring. Further the nine membered ring has some conformational flexibility, whereas the eight membered ring is fairly rigid. In the nine membered ring, the sterically allowed conformations depend upon whether a purine is an acceptor and pyrimidine a donor or *vice versa*. For example, A–U (T) and C–G base pairs require different conformations of the protein backbone.

TABLE I

Allowed conformations of peptide backbone (polyserine) in the eight and nine membered rings

Conformation	Base pair	Size of ring	Dihedral angles* ψ_2	ϕ_2	ψ_1	ϕ_1
A	A–U	8			330	200
B	G–C	8			320	210
C	A–U	9	270	50	150	80
D	A–U	9	270	340	210	70
E	A–U	9	270	330	230	60
F	G–C	9	330	0	200	20
G	G–C	9	270	180	70	60

* ϕ, ψ angles are defined taking *trans* arrangement as the zero value[13]. For the nine membered ring, (ψ_2, ϕ_2) and (ψ_1, ϕ_1) are the angles at the acceptor and donor ends respectively of the peptide chain.

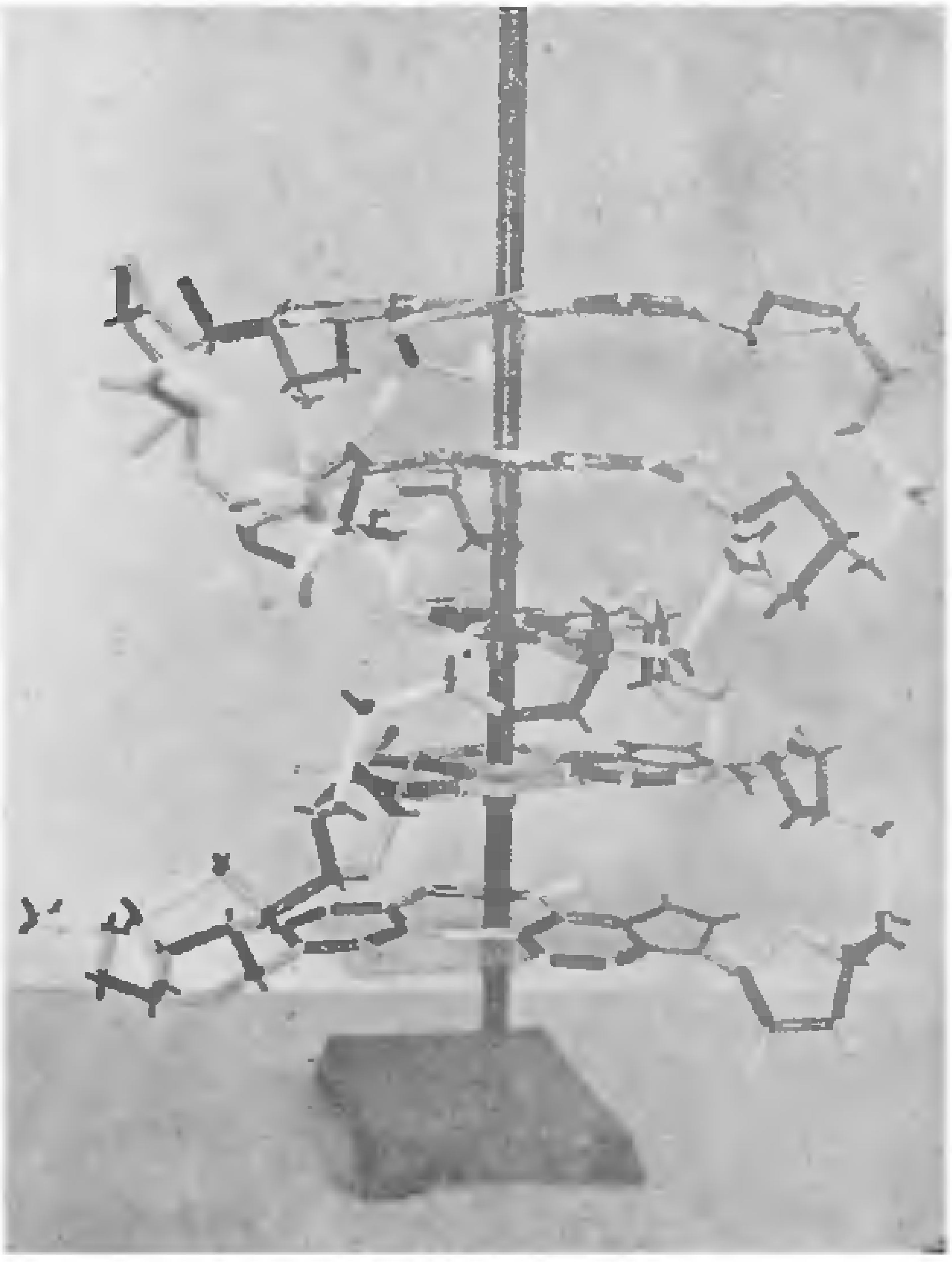

FIG. 3. Model of poly (A). poly (U) double helix (B-type), 5 base pairs in length, showing poly (L-ser) bound to two successive base pairs in the major groove. The peptide backbone loops out considerably between the two base pairs (top left). The upper base pair forms an eight membered ring, while the lower pair forms a nine membered ring. Polyserine has conformations C and A (Table I) respectively for eight and nine membered rings,

3. Specific Recognition of the Base Sequences by Protien Backbone

Specific recognition of base sequences by proteins involves (i) discrimination between A–U (T), U (T)–A, G–C and C–G base pairs and (ii) discrimination between different sequences in the nucleic acid double helix. The potential of protein backbone interaction in these two processes is discussed below.

A–U (T) is easily distinguishable from the U (T) pair due to the opposite directions of the H-bonds between the protein and the base pairs. The geometrical requirement on the part of the protein to bind so as to form eight or nine membered rings are different for the two pairs. Similarly A–U (T) pair is distinguishable from G–C pair and G–C is distinguishable from C–G for the same reasons. On the other hand, U (T)–A and G–C as well as A–U (T) and C–G pairs have the same directions of H-bonds (Fig. 2). However, in each of these two sets, the two base pairs have different sterically allowed conformations for the formation of nine membered rings, since one pair has purine as an acceptor while the other pair has a pyrimidine as an acceptor.

Discrimination between the different sequences requires that the protein backbone binds to every base pair in a given sequence. Depending on the combination of eight and nine membered rings in a sequence, the geometry of the binding protein will be different. This implies that a given sequence of bases can be recognized by more than one protein having different geometries at the binding site. Since the formation of eight membered ring is limited to a few amino acid residues (Gly, Ala, Ser) the number of such possibilities is very small. On the other hand, for a given protein with a definite geometry at the binding site, the specificity of binding could be very high. For example, let us consider the following sequences :

```
A — G — C — G              G — A — C — C
|   |   |   |              |   |   |   |
|   |   |   |     and      |   |   |   |
|   |   |   |              |   |   |   |
T — C — G — C              C — T — G — G
      I                          II
```

Let us assume that the geometry and sequence of a protein at the binding site are such that it can form only nine membered rings. If it is able to bind to sequence I, it is easy to see that it cannot bind to sequence II since the geometries for nine membered rings for A–T and G–C pairs are different. It is of course assumed that the protein does not undergo a conformational change. If it does, the specificity will be reduced to some extent.

In addition to the geometrical factors discussed above, the energetics of interaction between the different base pairs and the protein will also contribute to specificities.

Detailed experimental and theoretical investigations to check the above discussed ideas are in progress in our laboratory.

Acknowledgements

The author is thankful to Prof. G. Govil and Mr. N. Vasantkumar for useful discussions.

---

1. Helene, C., *FEBS Lett.*, 1977, **74**, 10.
2. —, *Nucleic Acids Res.*, 1975, **2**, 961.
3. Seeman, N. C., Rosenberg, J. M. and Rich, A., *Proc. Nat. Acad. Sci., U.S.A.*, 1976, **73**, 804.
4. Wehling, K., Arfmann, Hans-Adolf, Seipke, G. and Wagner, K. G., *Nucleic Acids Res.*, 1977, **4**, 513.
5. —, —, —, Standke, K. H. C. and Wagner, K. G., *Ibid.*, 1975, **2**, 799.
6. Seipke, G., Arfmann, Hans-Adolf and Wagner, K. G., *Biopolymers*, 1979, **18**, 855.
7. Mayer, R., Toulme, F., Monteay-Garestier, T. and Helene, C., *J. Biol. Chem.*, 1979, **254**, 75.
8. Brun, F., Toulme, J. J. and Helene, C., *Biochemistry*, 1975, **14**, 558.
9. Toulme, J. J., Charlier, M. and Helene, C., *Proc. Nat. Acad. Sci., USA*, 1974, **71**, 3185.
10. Govil, G. and Nanda, R. K., *Ind. J. Biochem. Biophys.*, 1978, **15**, 1.
11. Niu, Chien-Hua and Black, S., *J. Biol. Chem.*, 1979, **254**, 265.
12. Lancelot, G. and Helene, C., *Proc. Nat. Acad. Sci., USA*, 1977, **74**, 4872.
13. Pullman, B. and Pullman, A., *Advan. Protein Chem.*, 1974, **16**, 347.