PROTEIN–NUCLEIC ACID INTERACTIONS

Interaction of Tyrosine containing model oligopeptides with some deoxyoligonucleotides

A THESIS

Submitted in fulfilment of the requirements for the award of the degree of

DOCTOR OF PHILOSOPHY

By

ANITA AGRAWAL

247155
Acc. No. ---
Date 24-9-96

DEPARTMENT OF BIOSCIENCES AND BIOTECHNOLOGY
UNIVERSITY OF ROORKEE
ROORKEE - 247 667 (INDIA)

FEBRUARY, 1995

Gratis

DEDICATED TO MY PARENTS

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in this thesis entitled "PROTEIN-NUCLEIC ACID INTERACTIONS: Interaction of Tyrosine containing oligopeptide with some deoxyoligonucleotides" in fulfilment of the requirements for the award of the Degree of DOCTOR OF PHILOSOPHY, submitted in the Department of Biosciences and Biotechnology of the University is an authentic record of my work carried out during a period from June, 1988 to February, 1995 under the supervision of Dr.(Mrs.) Ritu Barthwal, Head and Reader, Department of Biosciences and Biotechnology, University of Roorkee, Roorkee.

The matter embodied in this thesis has not been submitted by me for the award of any other degree.

Date:17.2.95

A Agrawal
(ANITA AGRAWAL)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Ritu Barthwal 17.2.95
RITU BARTHWAL
Head & Reader, Deptt.
of Biosciences & Biotechnology

The candidate has passed the Viva-Voce examination held on Oct 16, 1995. The thesis is recommended for the award of Ph.D degree.

Ritu Barthwal
Signature of Guide
16.10.95

16-10-95
Signature of External Examiner

ABSTRACT

There has been a great interest in the precise determination of structure of DNA molecules of varying sequence and of their complexes with protein molecules. But these biological systems are very complex. So model systems have been extensively used. The obvious advantage is the possibility to investigate interactions of DNA-peptide system with controlled amino acid composition, sequence, chain lengths and predictable conformation which results in considerable simplification of the system as compared to the natural DNA-protein systems.

The present work is on a functional part of DNA binding loop of Gene V Protein comprising of residues, Lys^{24}-Pro-Tyr-Ser-Leu-Asn^{20}. We have studied interaction of this synthetic peptide with a single-stranded $d(A)_5$ and double-stranded DNA $(d\text{-}GACTCGTC)_2$ by NMR techniques. One dimensional NMR, 2D COSY and NOESY spectra are used to get changes in chemical shift, T_m, spin-spin couplings, interproton distances, intermolecular NOEs and hence the structural details. The results establish a role of Tyr, Lys, Pro, and Leu residues.

We also present results obtained by theoretical calculations using classical potential functions on the stacking of nucleic acid bases (A,T,G,C) and base-pairs (AT,CG) with aromatic amino acids. Intercalation of aromatic amino acids between model dinucleotide systems such as d-CG, d-CC, d-AT, d-AA have been studied. Unwinding and winding of DNA helix upon intercalation have been investigated. Conformational energy of complexes of amino acids with single-stranded DNA have also been computed. Overlap geometries in the optimised conformations have been obtained.

The trends in interaction energy suggest that Trp and His are involved in the formation of more stable complex than Tyr and Phe. Among dinucleotide model systems, homonucleotides show strong binding to aromatic amino acids. The molecular

mechanisms involved in these interactions show a specificity in protein-nucleic acid associations.

Abbreviations used:

NMR	Nuclear Magnetic Resonance
Tyr	Tyrosine
Phe	Phenylalanine
Trp	Tryptophan
His	Histidine
GVP	Gene V Protein used in thesis and also in literature.
G5P	Gene 5 protein used in literature.
NOE	Nuclear Overhauser effect
COSY	Correlation spectroscopy
NOESY	Nuclear Overhauser enhancement spectroscopy
τ_m	Mixing time
ssDNA	single-stranded DNA
dsDNA	double-stranded DNA
SS	single-stranded
RF	Replicative Form
t_1	Evolution period
t_2	Detection period
T_1	Spin-lattice relaxation time
T_2	Spin-spin relaxation time
T_m	Melting temperature
1D	One-dimensional
2D	Two-dimensional
FT	Fourier Transform
CD	Circular Dichroism
D_2O	Deuterium Oxide
DSS	sodium 2,2 Dimethyl-2 silapentane-5 sulphonate
HPLC	High Performance Liquid Chromatography
RPLC	Reverse Performance Liquid Chromatography
FPLC	Fast Performance Liquid Chromotography
EDTA	Ethylene Diamine Tetraacetic Acid
τ_c	Correlation time
3D	Three-dimensional

ACKNOWLEDGEMENTS

I wish to express my sincere thanks and gratitude to Dr. R.Barthwal, Head and Reader, Department of Biosciences and Biotechnology, University of Roorkee, Roorkee for providing her supervision, valuable criticism, expert guidance and constant encourgement throughout the completion of this work.

I wish to express my appreciation and thanks to my senior colleagues, especially, Anwer for his support and guidance. I enjoyed working with my labmates Uma and Nandana.

I am most grateful to all the members of FT-NMR National facility at TIFR, Bombay for the hospitality and friendly atmosphere extended to me during my stay at Bombay. My sincere thanks are due to Prof. G.Govil and Dr. R.V.Hosur for providing guidance during the course of research work.

I thank Prof. V.S.Chauhan, ICGEB, New Delhi for providing me all required research facilities to synthesize my samples.

Financial assistance rendered by Department of Science and Technology, Government of India, is gratefully acknowledged.

The blessing and moral support from my parents remained a constant source of encouragement. To them and to my brothers, Rajiv and Sanjiv, goes my deepest sense of gratitude and love. I don't find words to express my gratitude for my in-laws who have tolerated and provided love and support to bring this dissertation to fruition. Finally I wish to thank my husband, Sumeet Agrawal, for his constant encourgement at the last stage of my research work, without which it would have been impossible for me to complete this work.

A Agrawal

(ANITA AGRAWAL)

CONTENTS

CHAPTER I

INTRODUCTION

A General

Our present day knowledge of interactions between proteins and nucleic acids is rather limited. The rules for DNA-DNA interactions in terms of G-C and A-T base pairing have been known for almost three decades. However, the situation for protein-nucleic acid interactions has remained obscure.

DNA functions as the primary source of genetic information by interacting with proteins that copy it into a strand of RNA, and with other proteins that modulate the copying activity. In the `transcription' process a class of enzymes called the RNA polymerases construct a RNA copy of the DNA sequence. This RNA is then converted to protein by a process called `translation', the mechanism of which is not yet fully understood at the molecular level.

Another class of enzymes are the restriction endonucleases, which recognise specific base sequences in the DNA. These enzymes cleave foreign DNA molecules and prevent their transcription. Host DNA and other DNA molecules harboured by the host, are protected from cleavage by methylation at specific sites. The same sites are recognised by a restriction endonuclease that act only at the unmethylated positions. As yet not much is known about how the methylation works, but it is thought that it alters the interactions of DNA with protein that binds to it. This may be accompanied by subtle structural changes in the DNA molecule.

Besides this process there are many others in which protein molecules interact specifically with well defined sequences of nucleic acids. However, the specificity with which proteins interact with DNA varies. Proteins involved in DNA packing and repair are relatively less specific to base sequences as compared to those involved in gene control and expression.

In case of restriction enzymes, the specificity of binding is quite high. For example, for EcoRI which recognises the double stranded sequence d-(GAATTC), the ratio of free energy of binding to the cognate hexanucleotide to that at the noncognate site is about 10^5. In case of DNase, the specificity exists to a lower extent yet is significant. In fact, to understand these interactions at the molecular level, it is necessary to know their structures accurately.

Structure of several types of protein-templates recognised by DNA have been unravelled by X-ray crystallography. These include, restriction enzymes with homology to EcoRI, the helix-turn-helix motifs presented by repressor-operator systems, DNase I, DNA polymerase I, as well as structural proteins and their DNA complexes such as bacterial chromosomal protein, DNA binding proteins of Gene 5 and histone octamer.

Theoretical calculations have provided insight into the intermolecular interaction energies involved in protein-nucleic acid interactions. Solution structures of several protein binding segments of DNA are now available and as a consequence, the functional relationships in DNA expression as controlled by binding of regulatory proteins to DNA has begin to emerge.

B Structure of Nucleic Acids

The building blocks of nucleic acids are the nucleotide units. Each nucleotide unit is made up of:

1. Purine or Pyrimidine base: The purine bases found in nucleic acids are guanine and adenine and the pyrimidine bases in DNA are cytosine and thymine. The latter being replaced by uracil in RNA (Fig.1.1). In addition to these common bases found in nucleic acids there are a number of minor bases such as β-pseudouridine, inosine and 7-methyl guanosine, which are also found as components of nucleotides.

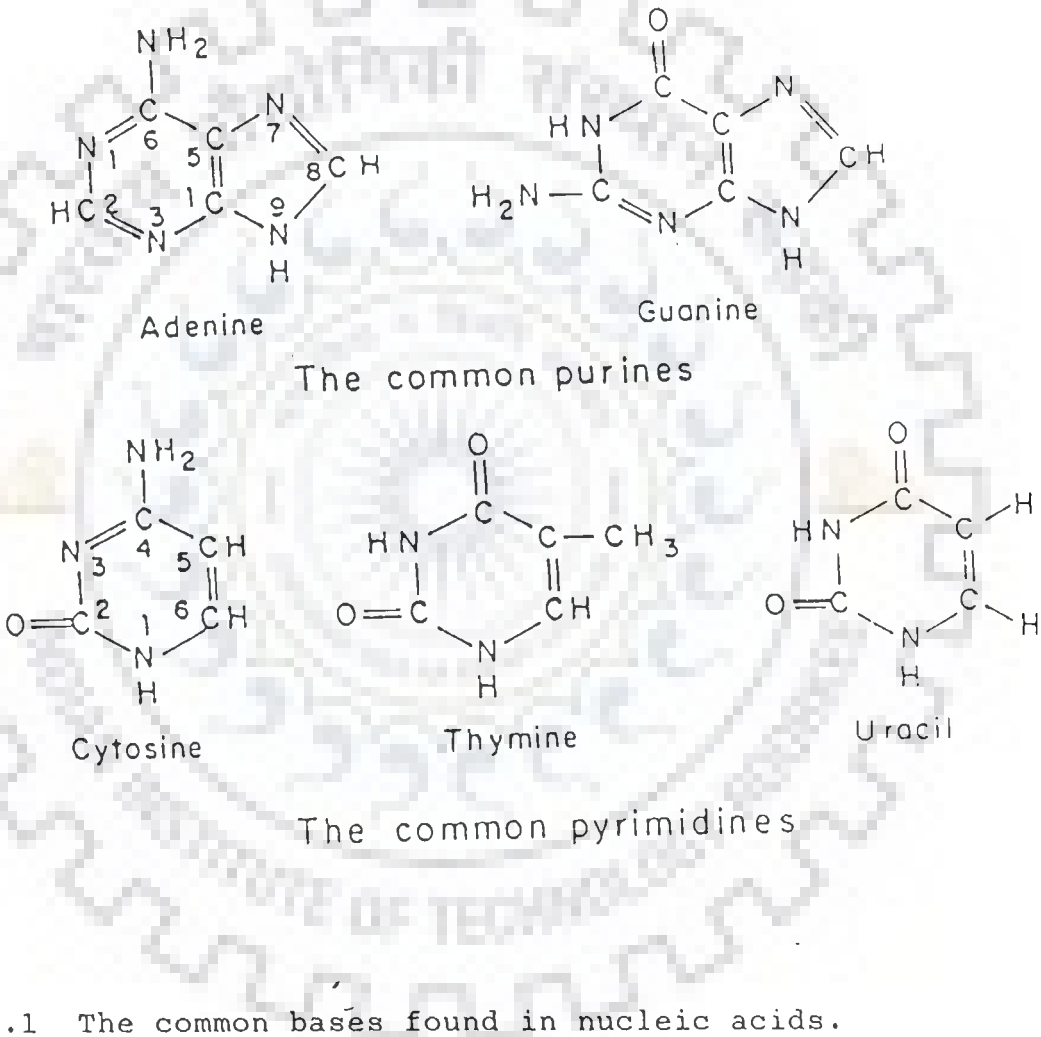
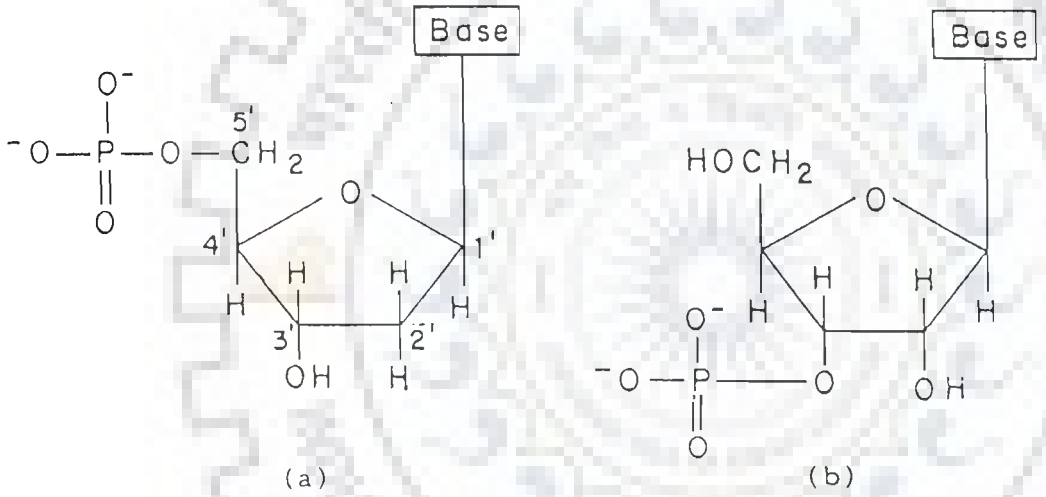


FIG. 1.1 The common bases found in nucleic acids.



G. 1.2 Five membered furanose ring (a) deoxyribose and (b) ribose found in DNA and RNA, respectively.

2. Sugar: A five membered furanose ring (Fig.1.2), ribose in RNA and deoxyribose in DNA, which have a well defined configuration, are known to exist. The nucleic acid bases are attached to the C1' position of the sugar moieties. The sugar base combination is known as nucleoside unit.

3. Phosphate Group: It is attached to the furanose ring, which bridges successive nucleotide units (Fig.1.3).

Polynucleotide chains are made up of many mononuclectide units, linked by a phosphoric acid bridge from the 5'end of one nucleotide to the 3' end of the other. Only a few nucleotides with 2'-5' phosphodiester bonds are known.

The primary structure of nucleic acids is determined by the sequence of bases along the nucleotide chain and the functions of nucleic acids namely replication, transcription and translation are governed by this sequence.

The secondary structure of nucleic acids i.e. three dimensional conformation, is governed by a number of torsion angles shown in Fig.1.3.

These torsion angles can be divided into three groups:
a. The glycosidic torsion angle (χ): This angle defines the orientation of the planar base with respect to the sugar ring. The sequence of atoms defining this angle is O1'-C1'-N9-C8 for the purines and O1'-C1'-N1-C6 for the pyrimidines. Alternate definitions have been used in the literature (129). In general, the anti conformation in nucleotides have a χ angle of approximately 30°, while the corresponding angle for a syn nucleotide is around 250°, the former being energetically more favourable.

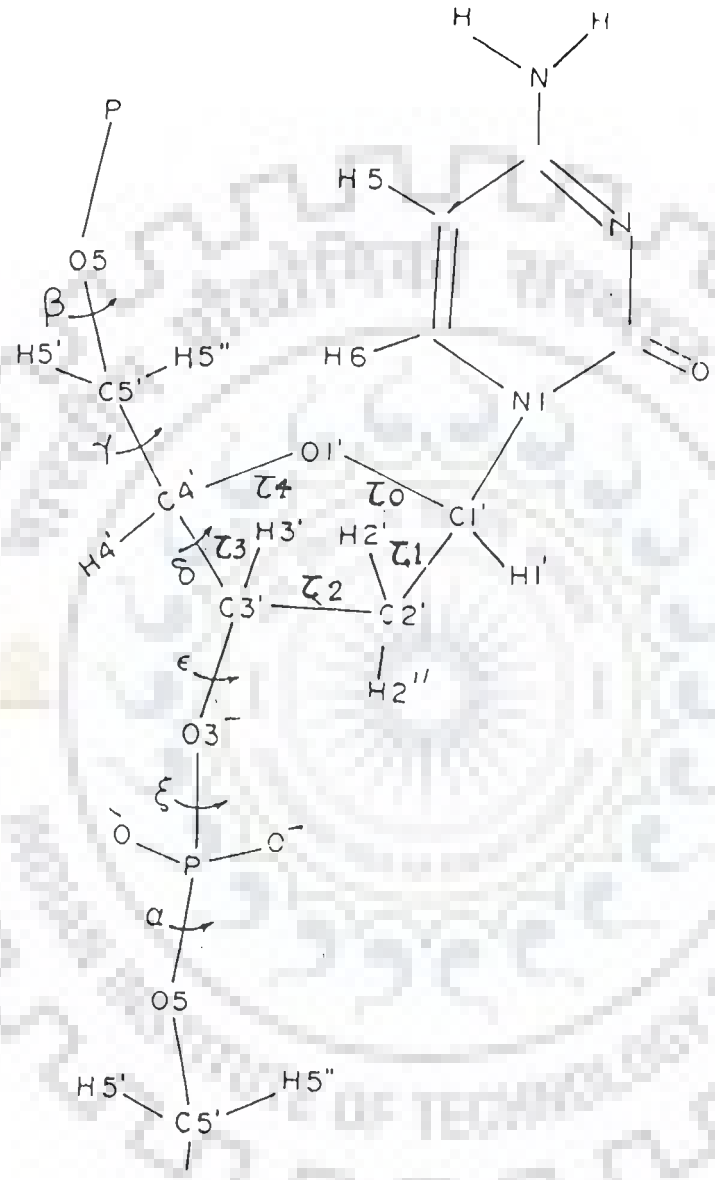


FIG. 1.3 A nucleotide unit showing the numbering of the atoms in the furanose ring and the backbone torsion angles.

b. The five torsion angles associated with the sugar ring : The sugar ring has an important position in the nucleotide unit since it forms part of the backbone unit and side chains. Its conformation is best described in terms of five endocyclic torsion angles about the bonds O1'-C1', C1'-C2', C2'-C3', C3'-C4' and C4'-O1' denoted by the symbols τ_0, τ_1, τ_2, τ_3 and τ_4, respectively. Since the sugar consists of five atoms only, two torsion angles are sufficient to define the three dimensional geometry of the ring, as the cyclic system introduces constraints due to closure of the ring.

The sugar ring is non planar and its conformation is described in terms of envelope (E) or twist (T) conformations. There are ten possible envelope conformations for the furanose ring and it has been observed that the most commonly occurring conformations of the sugar ring correspond closely to these forms. A convenient and more general description of the sugar ring is based on the concept of pseudorotation (7).

Each conformation of the furanose ring is unequivocally described by two pseudorotational parameters, P and δ. In this description the torsion angles are related by the equation:

$$\tau_j = \tau_m \cos[P + (j-2)\delta]$$

where

$$\tan P = \frac{(\tau_4 + \tau_2) - (\tau_3 + \tau_0)}{2\tau_3(\sin 36^\circ + \sin 72^\circ)}$$

and for a five membered ring τ_m (maximum amplitude) = 38°, δ = 144° and j = 0, 1, 2, 3 and 4. The conformations for P = 0 $\pm$ 18° are denoted by N and include the classical C3' endo and C2' exo conformations and those for P = $180^\circ \pm 18^\circ$ are denoted by S and include the C2'endo and C3' exo puckers of the ribose ring (66).

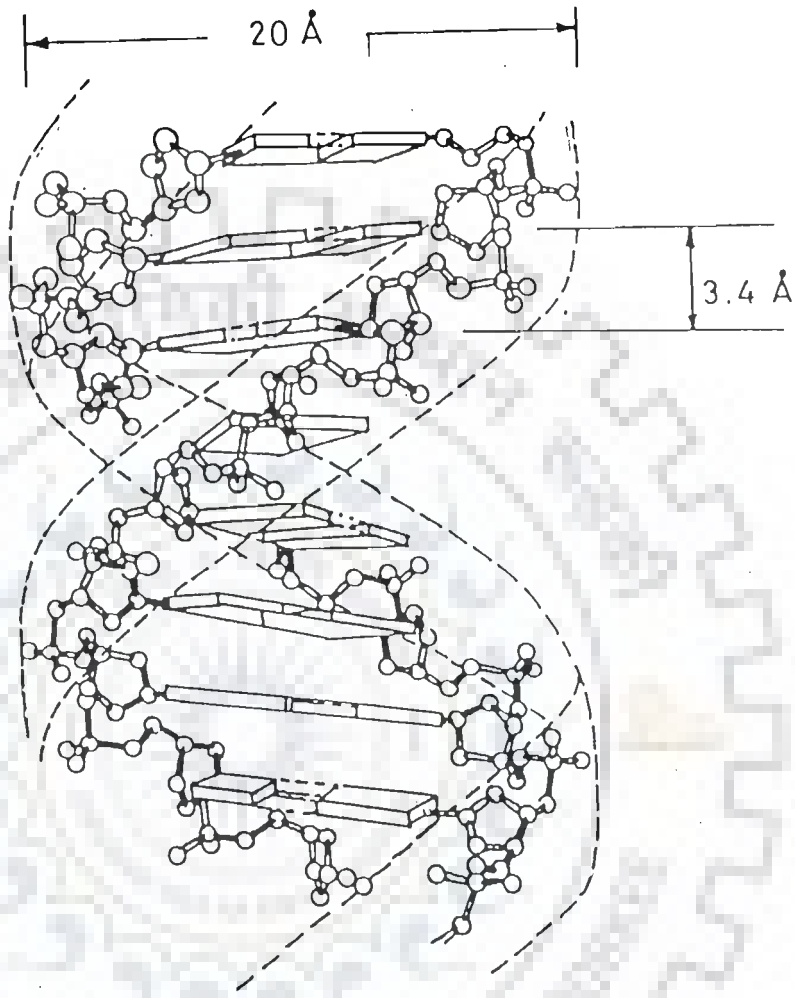


FIG. 1.4 Schematic representation of the DNA double helix. The sugar-phosphate backbones run at the periphery of the helix in antiparallel orientation. Base-pairs are stacked along the centre of the helix.

c. Six torsion angles along the sugar phosphate backbone: These are the most important parameters which determine the secondary structure of the polynucleotide chain. Under normal physiological conditions of temperature and pH, most double stranded nucleic acids exist as helices with well defined dihedral angles. Different types of helices have been observed depending upon conditions of humidity, ionic strength, etc. These differ in their helical parameters such as twist, pitch and number of units per turn. At elevated temperatures, the helical order gets lost and polymers go into what are called as random coils.

In naturally occurring nucleic acids, the helix consists of two strands (except in the case of tRNA) forming a double helix (Fig. 1.4). The two sugar phosphate backbones run along the outside of the helix in an antiparallel orientation and the bases are paired with each other along the centre of the helix. The outer envelope of the double helix is not smooth and can display one or two grooves of different width and depth. The DNA double helix is stabilized by base-base hydrogen bonds and stacking interactions. The hydrogen bonding arrangement in the double helix depends on specific interactions between the bases and is such that adenine forms a base pair with thymine and cytosine with guanine according to a scheme proposed by Watson and Crick (Fig.1.5).

In addition to this horizontal base-base hydrogen bonding in the solid state and in aqueous solution, the bases are also stacked vertically so that one base plane is at the van der Waal's distance of approximately 3.4 Å and parallel to the adjacent one. This vertical stacking is important for the stabilization of DNA helices and it has been found that stacking interactions between purines and pyrimidines follow the trend:

purine-purine > pyrimidine-purine > pyrimidine-pyrimidine.

The DNA double helix is capable of assuming many forms and reacts in many different ways. Basically it can exist as a right-handed

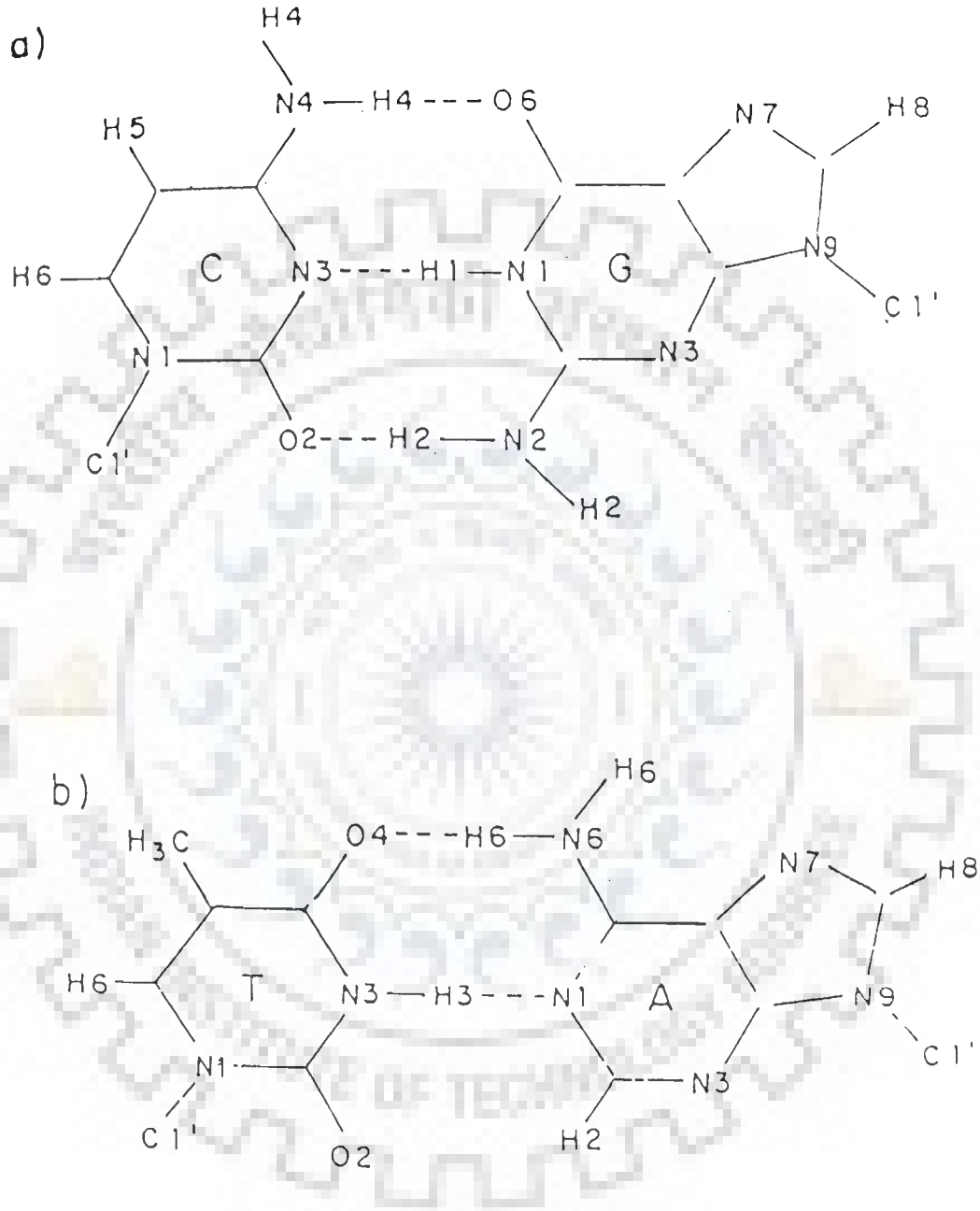


FIG. 1.5 Watson and Crick base-pairs showing the hydrogen bonding arrangement in double helical DNA.

or left-handed helix.

Within these two major groups, there exist a large number of families or subgroups of DNA structures. Different forms of DNA can be described in terms of conformation of sugar ring, the base orientation and backbone dihedral angles. DNA double helices are classified as A, B or Z type and the main features of these are given in Table 1.

C Structure of Protein Binding Domains of DNA

Recent studies have focussed attention on specific DNA sequences recognised by proteins. In all cases, the observations indicate the presence of a right-handed B-DNA structure. Significant sequence dependent alterations have been found. In many cases, the major alterations occur close to the sites specifically recognised by proteins.

The conformational variations result not only in sequence dependent orientation of the phosphate groups but also in relative positioning and stacking of the base-pairs, which are the important functional groups for protein binding. The conformational movements are somewhat correlated and a major structural variation can transmit its effect to several base-pairs in its neighbourhood.

Recently, structural studies on specific sequences of DNA have been reported. For example, the solution structure of d-GAATTC sequence recognised by EcoRI shows that the recognition segment acquires a B-DNA conformation in both phases with variations in sugar pucker ranging from C3' endo to C2' endo. Such studies indicate that within the B-DNA family of structures, the DNA molecule can show marked conformational fluctuations orienting the active group in a conformation, which makes it easier for the proteins to recognise specific segments. Further, since the B-DNA structure can readily adopt minor structural changes, it is likely that the proteins may cause further perturbation to the structure

TABLE 1: STRUCTURAL CHARACTERISTICS OF A, B AND Z-TYPE DOUBLE HELICES.

FAMILY TYPE	A-TYPE	B-TYPE	Z-TYPE
BASES PER TURN	11	10	12
ROTATION PER NUCLEOTIDE	30°to 32.7°	36°to 45°	-60°
BASE PAIR TILT	10°to 20.2°	-5.9°to 16.4°	-7°to 9°
HELIX PITCH (nm)	3.09 to 3.60	3.38	4.56
RISE PER RESIDUE	0.20	0.34	0.76/2
HELIX DIAMETER (nm)	2.30	1.93	1.81
HELIX SENSE	RIGHT-HANDED	RIGHT-HANDED	LEFT-HANDED
GLYCOSIDIC BOND ROTATION	ANTI	ANTI	ANTI (DEOXYCYTIDINE) SYN (DEOXYGUANOSINE)
SUGAR PUCKER	C3' endo	C2' endo	C2' endo (DEOXYCYTIDINE) C3' endo (DEOXYGUANOSINE)

for optimal binding.

D Structure of Proteins

The basic structural units of proteins are the amino acids. The proteins in all species are constructed from the same set of twenty amino acids, which can be grouped according to their side chains as:

a) Aliphatic : Glycine, Alanine, Valine, Leucine, Isoleucine and Proline.

b) Aromatic : Tyrosine, Phenylalanine and Tryptophan.

c) Basic : Lysine, Arginine and Histidine.

d) Acidic : Aspartic acid and Glutamic acid.

e) Amide : Asparagine and Glutamine.

f) Sulphur Side Chain : Cysteine and Methionine.

g) Hydroxyl Aliphatic : Serine and Threonine.

These amino acids are joined by peptide bonds to form a polypeptide chain which consists of a repeating sequence of peptide units $-NH-C^{\alpha}H(R)-C=O$. However the C-N bond of peptide linkage is shorter than most single bonds and has some double bond character. The four atoms comprising the peptide bond and the two α-carbon atoms lie in a single plane with the oxygen of the carbonyl group and the hydrogen of the -NH group trans to each other since this arrangement is energetically more favourable. A cis arrangement may occur with certain amino acid residues e.g. proline, where because of the spatial structure of the side chain, the cis and trans conformations have similar structural configurations and thus the difference in energy is small. The backbone of the peptide chain may thus be pictured as consisting of a series of relatively rigid planes, separated by substituted methylene groups.

The three dimensional structure of a single polypeptide chain is represented by specifying the backbone dihedral angles ω_i, ϕ_i and

ψ_i and the side chain rotational angles χ_1, χ_2, etc. According to the convention, torsion angles are measured with the cis conformation as 0° and the range of the backbone dihedral angles varies from -180° to $+180^\circ$. The χ angles are measured from 0° to 360°. Since the backbone structure can be defined by the two torsion angles, ϕ and ψ, a number of (ϕ, ψ) plots have been obtained using potential energy calculations, which depict the energetically allowed domains of these angles for a peptide residue (101,63). However these (ϕ, ψ) plots can be used only as a rough guide because the side chain interactions also play an important role in determining the conformation of the backbone.

The structure of proteins can be considered at different levels:
Primary Structure: It is the sequence of amino acids and the location of disulphide bridges, if any.

Secondary Structure: It is the folding of the polypeptide chains into different regular structures considering the constraints imposed by the peptide bonds and their specific dimensions. Some of the important secondary structures prevalent in proteins are:

a) Right-handed α-helix has 3.6 amino acid residues per turn of helix. This arrangement is particularly favourable because it is stabilized by intrachain hydrogen bonds between every fifth and first amino acid residue (Fig.1.6). In α-helix, $\phi = -58^\circ$ and $\psi = -46^\circ$ correspond to low energy regions. This form is often spontaneously assumed by peptide chains since it has the least free energy, provided there are no opposing interactions of the side groups or solvent. The nature and sequence of the side groups in the chain often inhibit the formation of an α-helix.

b) Left-handed α-helix is also stabilized by a hydrogen bonding pattern similar to that for a right-handed helix but is intrinsically less stable and therefore not so favourable.

c) β-structures are parallel or antiparallel polypeptide chains,

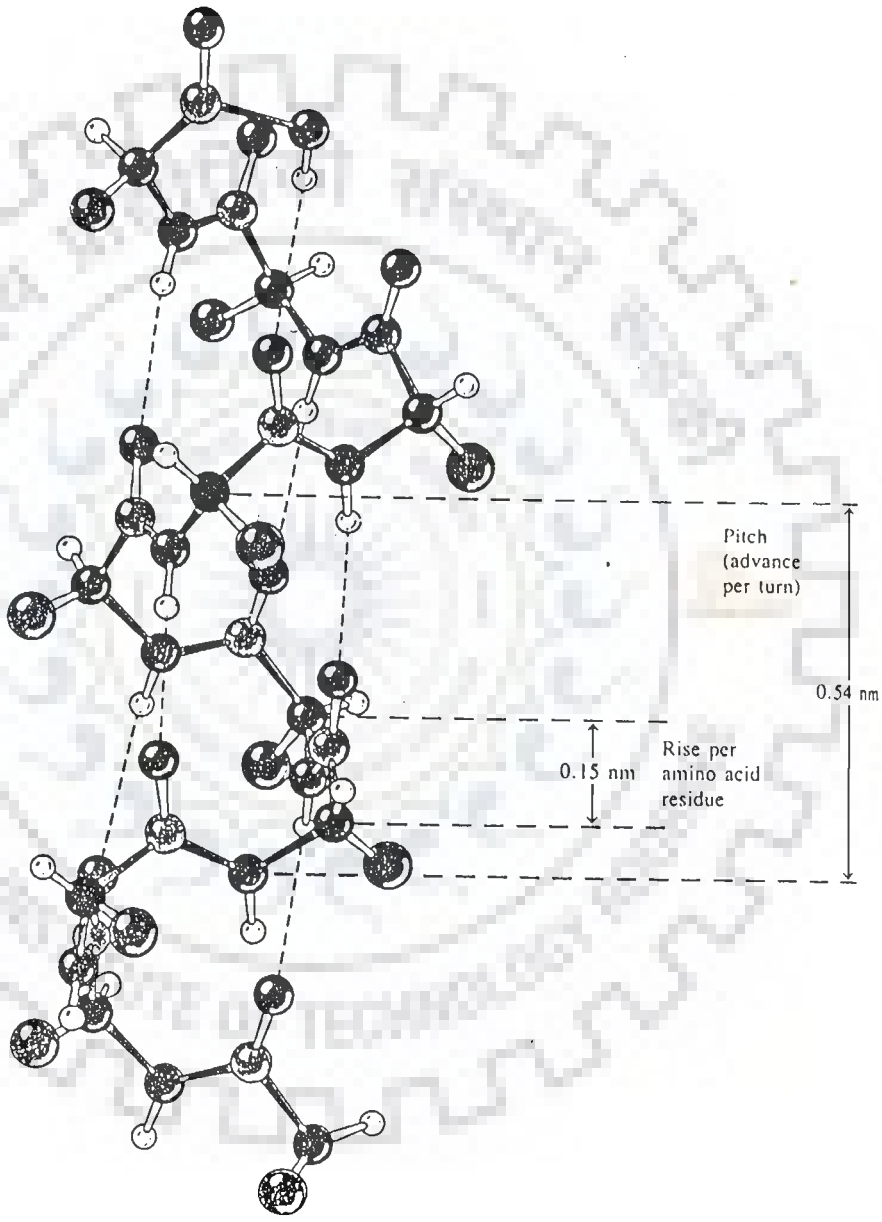


FIG. 1.6 Right handed α-helix. Equivalent positions in the helix recur every 0.54 nm. There are 3.6 amino acid residues per turn.

which represent the most extended chain conformations allowed by normal bond lengths and angles with a displacement of 3.47 Å per residue along helix axis. These polypeptide chains are cross linked by interchain hydrogen bonds between NH and CO groups (Fig.1.7) forming β-pleated sheets. If the adjacent chains are in the same direction, it is called parallel β-sheet, whereas if the chains are running in the opposite directions it is designated as antiparallel β-sheet. In the β-pleated sheets, the ϕ, ψ values are -119° and 113°, respectively for parallel chains and -139° and 135°, respectively for antiparallel chains.

d) β-turn is formed when a polypeptide chain folds back on itself forming a β-sheet. In this hairpin turn, the CO group of the n^{th} residue is hydrogen bonded to the NH group of the $(n + 3)^{rd}$ residue (Fig. 1.8). There are various β-loop structures, type I contains both the intervening peptide units in the L-configuration and the ϕ_2, ψ_2, ϕ_3, ψ_3 angles are -60°, -30°, -90° and 0°, respectively. Type II has a LD sequence and the values of the torsion angles are -60°, 120°, 80° and 0°, respectively. The third residue in type II β-turn can only be Gly due to steric constraints. Five membered proline ring has little conformational mobility and is ideally suited for second residue in reverse turn.

e) 3_{10} helix consists of 3 units per helix turn and has ϕ, ψ angles of -59° and -26°, respectively.

f) A unique type of single helical structure with no hydrogen bonding is polyproline helix. Polyproline I is a left-handed helix with 10 residues per turn and has all cis peptide bonds. Polyproline II is a left-handed helix with 3 residues per turn, displacement of 3.12 Å along helix axis and all trans peptide bonds. Polyproline helices are pure synthetic and are present in aqueous state. Closest to this secondary structure is collagen which imparts tensile strength to connective tissues such as skin, bone and tendon. Collagen dominated by Pro-Hyp-Gly are triple helices formed by twisting three polyproline II helices about each

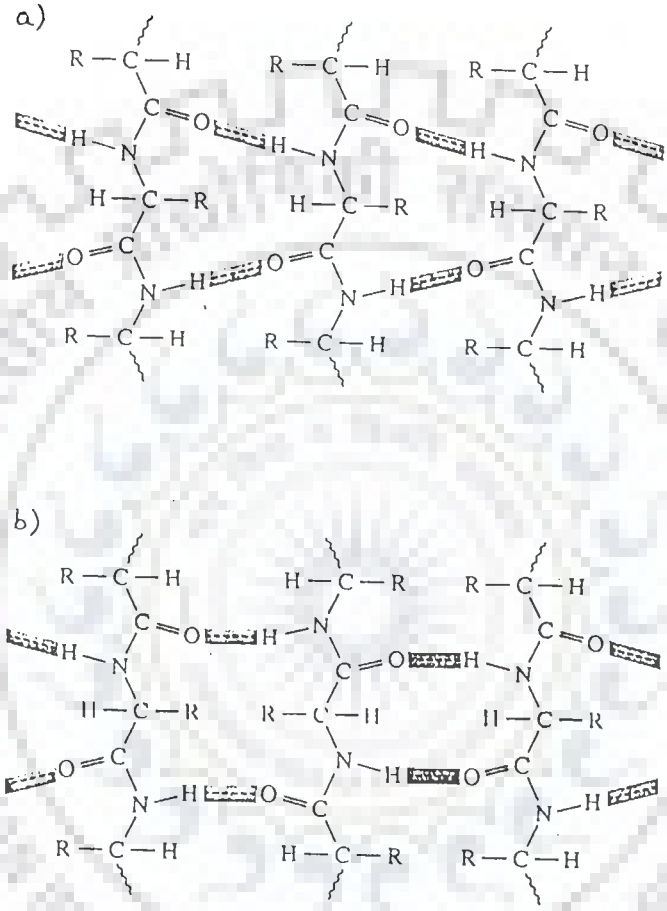


FIG. 1.7 β-pleated sheet. The hydrogen bonds are indicated by dotted lines, `R' represents the amino acid side chain.

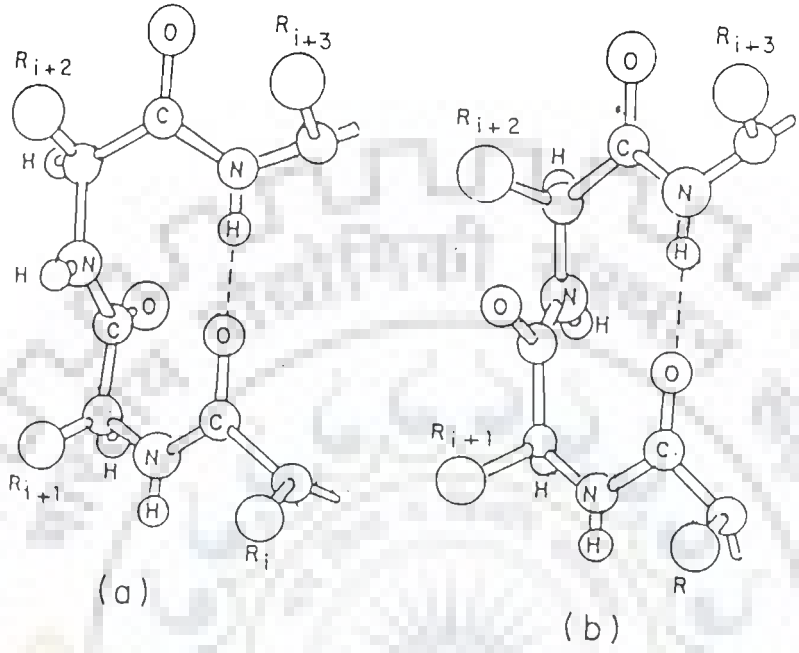


FIG. 1.8 β-turn of (a) type I and (b) type II.

other and stabilized by interchain hydrogen bonds. Unlike α-keratins, collagen does not stretch readily and it is distinctive in that one third of all its amino acid residues are glycine and a quarter or more proline or hydroxyproline. These regularly spaced residues force the chain to assume a peculiar kinked left-handed helix. Fibrous proteins have either α-helical structures as in the α-keratins (hair, wool) or β-pleated sheet structures as in the β-keratins which are formed when the α-keratins are subjected to moisture, heat and stretched, the transition being caused by the breakage of interchain hydrogen bonds which stabilize the tight α-helix.

Tertiary Structure: In globular proteins such as myoglobin, haemoglobin, lysozyme, ribonuclease, carboxypeptidase, etc. the polypeptide chains are not oriented along one direction but fold back upon themselves giving a compact structure. Globular proteins are highly convoluted structures with most of the hydrophobic residues on the inside and hydrophilic residues on the surface. They contain varying amounts of α-helix and β-structures and are usually cross-linked by interchain disulphide bridges. These covalent bonds stabilize the tertiary structure of globular proteins and limit their tendency to unfold completely. Segments of α-helix in proteins do not possess exactly the same dimensions and repeat distance and are not equally stable, since they have different amino acid contents and sequence. They contain local points of instability at which the helix is likely to bend in response to other forces, particularly the tendency of the entire chain to assume that conformation in which the hydrophobic R groups are maximally shielded from water. The characteristic tertiary structure of a polypeptide chain in an aqueous system is thus the resultant of intra or interchain hydrogen bonding and the tendency of the entire chain to bend at points of instability to assume the configuration which is most stable.

Quaternary Structure: Most globular proteins having a molecular weight greater than 50,000 daltons are oligomeric, consisting of

two or more separate chains. The characteristic manner in which the individual chains are arranged is called the quaternary structure and like the secondary and tertiary structure, it is determined by the primary amino acid sequence.

E Structure of Proteins interacting with Nucleic Acids

RNA-Binding Proteins

Like DNA, RNA too interacts with proteins in a great variety of functions. Few examples of RNA binding proteins are RNA polymerase, RNA ligase, reverse transcriptase, replicase, hydrolytic enzymes, exo- and endonucleases, etc.

Bovine pancreatic ribonuclease, RNase A and RNase S, bind specifically to pyrimidine and this specificity appears to reside in the formation of hydrogen bonds between the protein backbone or side chains with the pyrimidine bases. Several features of RNase S inhibitor-complexes are shared with complexes of dehydrogenase the coenzyme nicotinamide adenine pyrophosphoryl dinucleotide (NAD^+). The main difference between the two resides in the mode of binding. Unlike RNase S, the adenine moiety in NAD^+ appears to interact with the dehydrogenase via polar groups.

Glyceraldehyde-3-phosphate dehydrogenase as well as lactate dehydrogenase have been shown to bind to nucleic acids and to exhibit selectivity towards single-stranded structure. A study of oligonucleotide binding has shown that these dehydrogenases have a strong preference for dinucleotides of sequence CpA or UpA.

The study of the complex between tobacco mosaic virus (TMV) protein and RNA shows conformational changes on binding. The region of viral RNA has in its centre a sequence where every third residue is G. This feature might be important for the selectivity of protein-RNA interactions which initiate virus assembly.

DNA-Binding Proteins

Restriction enzymes

i) EcoRI: EcoRI endonuclease recognises double-stranded sequence d-(GAATTC) with high degree of specificity. It has a nonspecific binding with other DNA sequences which probably enhances the rate of formation of the specific complex by a facilitated diffusion along the DNA.

The two subunits of enzyme form a globular structure with the DNA embedded on one side. The major groove of DNA is in intimate contact with the protein while the minor groove is exposed to the solvent. The DNA lies in a distorted B double helical form. Evidently, these distortions are induced by protein binding. The disruptions appear to have structural consequences which propogate over long distances through twisting and bending. The resulting unwinding increases the separation of the DNA backbone across the major groove thereby facilitating access by the protein.

ii) DNase I : It is an endonuclease from bovine pancreas which degrades double-stranded DNA through hydrolysis of P-O3' bond. DNase I `footprinting' is regularly used to detect protected regions in protein-DNA complexes. The enzyme can distinguish hairpin loops and double helical segments of identical sequence.

The protein structure consists of eight α-helices and an exposed loop region between residues 70-74 i.e Arg-Asn-Ser-Tyr-Lys, has been implicated in binding to the minor groove of the DNA. This complex is stabilized by electrostatic interactions between the phosphate backbone of DNA and the positively charged amino acid residues Arg and Lys.

iii) DNA Polymerase I : DNA polymerase is found to be associated with error free replication. Polymerase I consists of two fragments : (a) a large C-terminal fragment and (b) a smaller fragment which retains the exonuclease activity. The larger fragment consists of two domains. One of the most striking

features of the larger domain is a deep crevice of appropriate size and shape for binding double-stranded B-DNA.

Repressor-Operator Systems

DNA binding proteins in this class have common features. The proteins form a dimer which makes contact with double helical operator DNA. The recognition segment of the proteins consists of two α-helices joined together by an arm. Both the helices are involved in the recognition process and the arm helps in the relative orientation of the two in the correct geometry for contact with DNA. This has given rise to the `helix-turn-helix' concept for protein-DNA interactions. Hydrogen bonding and electrostatic interactions are responsible for specificity in binding. Hydrogen bonding interactions occur mainly through the second helix of the helix-turn-helix motif. The DNA binding segments are of variable length 14-21 base-pairs and maintain a rough symmetry around the centre. Several base-pairs are conserved in different operators.

i) Cro Repressor : Cro is a basic protein (66 amino acids) which consists of three strands of antiparallel β-sheets and three α-helices. Model building studies indicate that the regions 15-38 and 54-66 are involved in binding. At least nine hydrogen bond contacts are involved in operator-repressor binding. In particular, the interaction of Glu 54, Lys 56 and Lys 63 with the nucleotide backbone has been suggested.

ii) Lambda Repressor : It has 236 amino acid residues and contains two domains. Cleavage with papain yields a 92 residue N-terminal fragment which binds to the operator region in phage DNA. A dimer of this fragment binds to the 17 base-pair operator region in the phage DNA. Lambda repressor consists of an extended arm and five α-helices. Residues in the N-terminal fragment seem to make major contacts. They can contact 3-4 base-pairs in the major groove. The orientation of the peptide dipole in the α-helix gives a partial positive charge at the N-terminus, which may allow

a favourable electrostatic interaction.

iii) CAP :- The catabolite gene activator protein (CAP), also called cyclic AMP protein regulates several catabolite-sensitive gene operons. The active form of CAP is a dimer of identical subunits. Each subunit consists of two distinct structural domains. The smaller domain presumably contains the DNA binding region. The complete structure is composed of the usual network of α and β structures.

iv) Trp Repressor :- This repressor consists of 107 amino acid residues and is folded predominantly into six α-helices which are connected by short turns. Binding of L-Trp to the aporepressor alters the conformation of the latter in particular regions leaving the rest of geometrical pattern unaltered. The Trp binding involves its indole ring which goes into a pocket formed by Arg 48, Gly 45 and Arg 54. The first two of these amino acid residues lie in the α-helix which is presumably involved in DNA binding.

Structure of DNA-protein systems

Complexes such as nucleosomes, chromatin and viral DNA-protein complexes have been studied.

i) Nucleosome core particle : It is a repeating unit of DNA in chromosome. It packs more than 90% of genome of eukaryotic cells. It consists of 146 base-pairs of DNA. The protein moiety is an octamer formed of pairs of the histone proteins. The DNA is wound around the histone core and has approximately 7.6 turns of double helix in one superhelical turn of nucleosome. The contacts between DNA and proteins are made on every turn of double helix.

ii) Structure of Gene V protein and bacteriophage M13/fd/f1 DNA complex

Structurally, the bacteriophage M13 consists of a circular, single-stranded genome encapsidated as a loop in a protein tube. The length of the ss DNA is about 6400 nucleotides. The wall of the protein tube is made up of 2700 copies of a protein, the gene

VIII product called as major coat protein. The end bears minor proteins which are specific to each end -thus the ends are differentiable from each other: morphologically, biochemically and functionally. The phage genome encodes 10 proteins (Fig.1.9) of which five are structural proteins, three are required for phage DNA synthesis, and two are required for virion assembly. In addition, there are two well conserved intergenic regions, which do not code for proteins but contain signal for the initiation of synthesis of both the (+) and the (-) strand of DNA, the initiation of synthesis of capsid formation, and the termination of RNA synthesis. The extended intergenic region lies between gene VIII and III and between gene II and IV.

The gene VIII protein encodes for the major coat protein and forms an integral part of the virion structure. Another gene III encoded protein is required for the termination of the virion, stabilization of the phage particle and for infectivity. The polypeptides encoded by the VI, VII and IX genes play an important role in making association with or integration into membranes.

Gene II encodes a specific nuclease topoisomerase which in the presence of Mg^{2+} either introduces a specific nick into the supercoiled replicative form (RF) or relaxes so as to leave a covalently closed circle without supercoils.

A major role is played by gene X encoded protein in single-strand DNA synthesis where probably it acts as an inhibitor of gene II function.

Gene V protein is a single-stranded DNA-binding protein and prevents synthesis of complementary strand by complexation with DNA. It is required for ss DNA synthesis.

Mechanism of DNA replication (27) in bacteriophage (Fig.1.10) involves three steps:

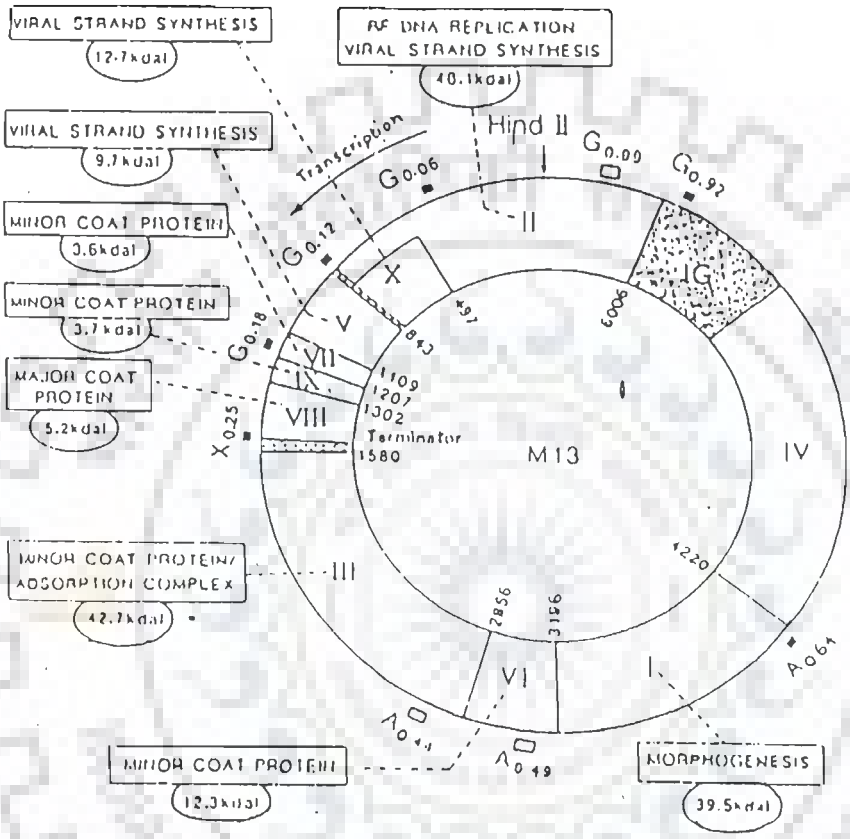


FIG. 1.9: Filamentous bacteriophage (fd/M13) genome organisation (27).

I) Synthesis of the complementary strand on the viral strand template which yields parental replicative form (RF) form DNA (ss to RF), following the infection. This stage is entirely carried out by the bacterial proteins. The various steps in this process are prepriming, priming, chain elongation, gap filling and ligation.

II) Replication of the parental RF DNA to yield the progeny RF DNA (RF to RF). RF DNA replication takes place by rolling circle mechanism and requires both viral and host proteins.

III) Asymmetric synthesis of the viral strand on the duplex RF template, which results in the accumulation and packaging of viral single-stranded DNA molecules. Both phage and host proteins are required in this final step (Fig.1.10).

STAGE I REPLICATION :

The phage genome shows little intrastrand base pairing and therefore remains largely in the single-stranded state. The DNA itself does not provide a suitable template for the replication reaction. It is first coated by the host single strand binding protein (SSB). SSB is a tetramer that binds cooperatively to single-stranded DNA, maintaining it in the extended state i.e. it stabilizes DNA which is in the single-stranded condition.

In the M13 phage the only part of the phage genome that is not bound to SSB is a sequence of 59 bases that forms a hairpin. This duplex region helps in RNA synthesis with the help of host RNA polymerase. The hairpin is disrupted by the RNA chain. SSB binds to the nontemplate half of the hairpin and the transcription is terminated. The RNA primer is extended by DNA polymerase III holoenzyme, which synthesizes a DNA complementary strand.

STAGE II REPLICATION

Amplification of the viral genome begins when the protein product of gene II introduces a nick at a specific site in the (+) strand

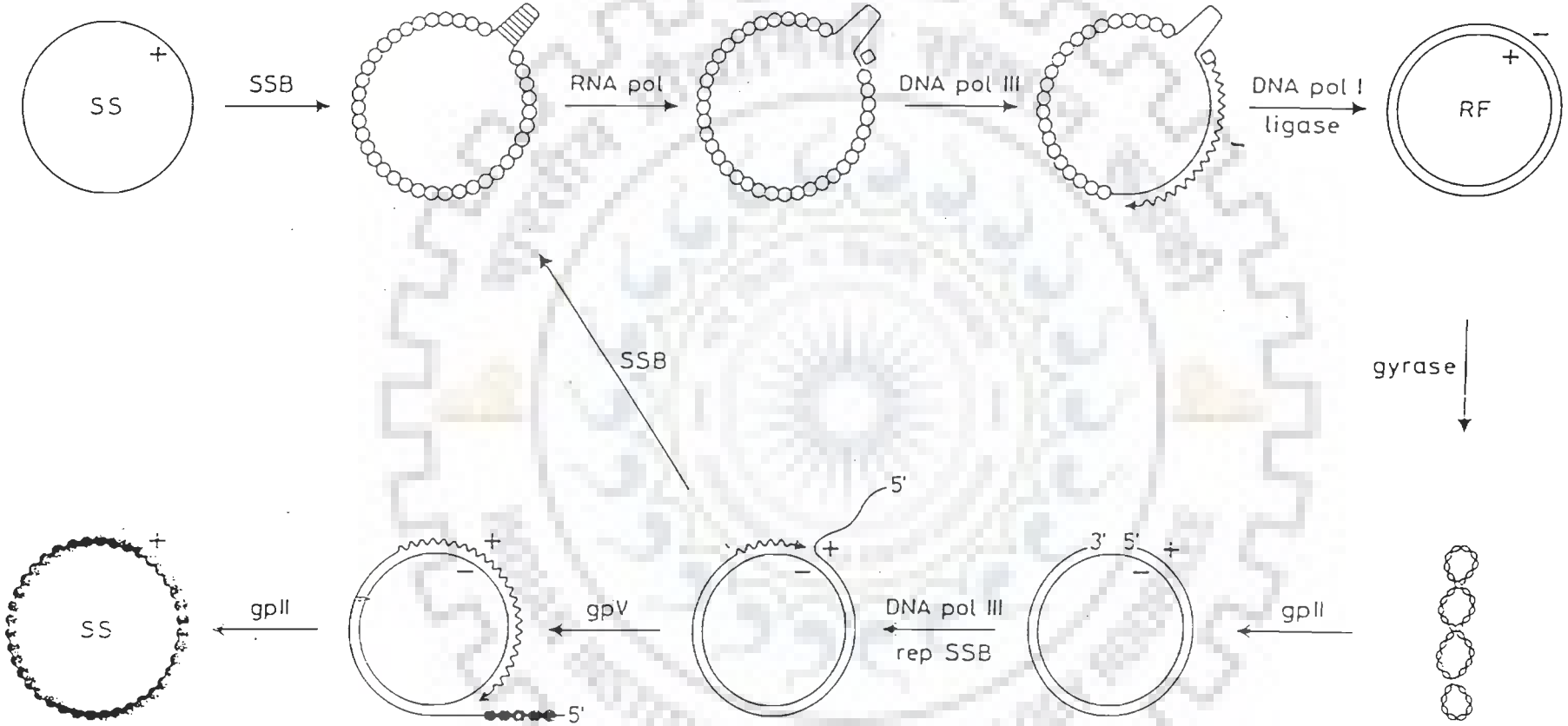


FIG. 1.10: Schematic representation of the two mechanisms of replication of bacteriophage M13. The upper part shows the discontinuous synthesis of the complementary strand (-) and the lower part shows the continuous synthesis of viral strand (+) (27).

of the parental RF DNA.

E. coli DNA polymerase I (host) then adds nucleotides to the free 3' hydroxyl terminus, progressively displacing the original (+) strand from the circular (-) strand template. After the replication fork has completed a full circle, the displaced (+) strand is cleared by the gene II product to generate a unit length viral genome that is then circularized. During first 15-20 minutes of infection, these progeny (+) strands are converted by cellular enzymes to closed circular RF DNA molecules, which serve as templates for further rounds of transcription and synthesis of additional (+) strands.

STAGE III REPLICATION :

When there are about 100-200 copies of RF DNA accumulated in the infected cell, there is enough single-stranded DNA binding protein (the product of gene V) to repress the translation of mRNA and bind to the newly synthesized (+) strands. The displaced viral strand is coated with approximately 1300 molecules of gene V protein and thus preventing further conversion to RF DNA. During morphogenesis the gene V protein is displaced from the viral DNA by capsid proteins at the cell surface. Removed gene V protein can then be recycled by associating with newly synthesized viral strands. Recycling of gene V protein is strongly coupled with morphogenesis, because inhibition of phage production immediately blocks the release of gene to protein from DNA-protein complex. Role of gene V protein is crucial as it maintains the ss form of DNA which is necessary for infection. Thus gene V protein has three primary roles in replication of bacteriophage M13:

1) It regulates the synthesis of the products of gene II and gene X.

2) It prevents further conversion of newly synthesized (+) strand to RF DNA and thus effectively limits the synthesis of phage protein at later stages, while allowing for their rapid expression

early in the infection cycle.

3) The gene V protein forms a stoichiometric complex with newly formed progeny single-stranded DNA viral strands. This intracellular complexation event is a necessary prepackaging step in the ultimate assembly of the viral particles.

STRUCTURE OF GENE V PROTEIN :

Structure of gene V protein was given by Brayer and Mcpherson (20). Structure of gene V protein consists of:

i) A major three stranded β-sheet originating from the initial two thirds of the sequence.

ii) A twisted β-ribbon that joins two monomers within the dimer.

iii) A broad connecting loop that substitutes for the most irregular portion of the element described previously as β-ribbon.

Monomer structure

Gene V protein is a T shaped molecule composed entirely of β-pleated structure. Starting from the N-terminus, the first ten residues are arranged in a short β-strand, followed by the large "DNA binding loop" residues 15-32, which is continued by the "complex loop" residues 33-40. After a short β-strand, residues 52-59, the polypeptide chain leaves the bar of the T and forms the stem with the large "dyad loop" represented by residues 61-82.

Dimer structure

In solution, in crystalline state and when binding to DNA, gene V protein exists as a dimer. The forces involved in dimer association are hydrophobic, and hence the dimer is stable under high ionic strength, extreme pH values and at elevated temperatures. The union between the monomers occur with the imposition of extended β-loops into the open channels that cross each individual monomer.

The principal role in aggregation are played by complex and the dyad β-loops. The dyad loops are aggregated and the T-bars are in

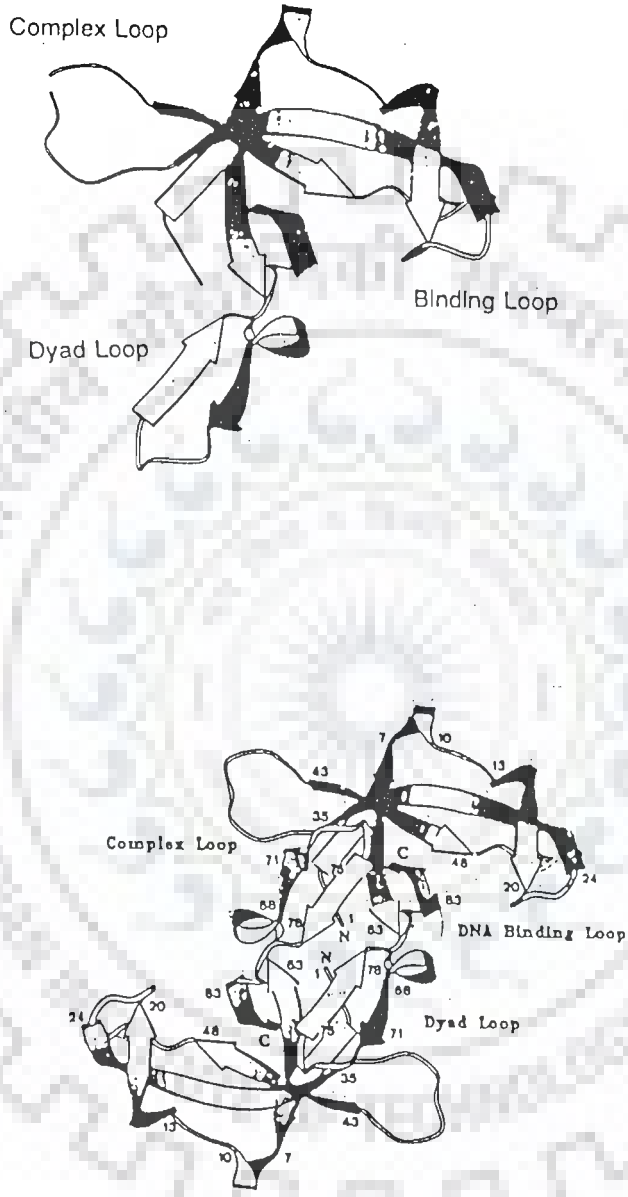


FIG. 1.11: Schematic representation of the structure of GVP-Y41H (a) monomer (53) (b) dimer (52).

antiparallel orientation.

Along the T bar, there is a DNA binding channel about 10 A wide and 35 A long formed by residues 15-32 and unperturbed by the dimer associations. This channel is long enough to accommodate about five nucleotides of DNA.

A major structural element is created in the gene V protein dimer by the quarternary interactions between symmetry related monomers. A compact and well delineated 6 stranded antiparallel β-cage is formed from the 2 extended dyad β-loops in association with N-terminal strands of both monomers. The intermolecular 2-fold axis is perpendicular to and bisects the barrel structure. The interactions that maintain the β-barrel are predominantly hydrophobic, which is contributed by 20 amino acids.

Recently based on the NMR results, a solution structure of gene V protein (Fig.1.11) have been proposed (51,52). It consists of:
1) Five stranded β-sheet comprising of residues 3-6, 30-35, 43-48, 59-63 and 83-85.
2) Two antiparallel β-loop comprising of residues 68-78 and 13-31.
3) 3_{10} helical turn 7-10 and 64-67.

They have predicted one more β-turn in the strand 13-31 but its position could not be ascertained (Fig.1.11).

Gene V protein and its binding to DNA are the subject of study for the present thesis work and have been discussed in later part of the thesis.

F. Elements of Protein-Nucleic Acid Interactions

i) **Electrostatic Interactions** : Most of the protein-nucleic acid complexes involve electrostatic interactions between positively charged lysyl or arginyl side chains and negatively charged phosphates. The energy of such interactions, particularly in regions of low water activity, are an order of magnitude higher

than those associated with the other nonbonded interactions. Such interactions are predominantly Coulombic, though not of simple monopole-monopole type. Because of the extensive charge delocalisation, significant contribution arises from monopoles and dipoles centered on atoms and bonds other than those directly associated with positive and negative charges in a classic sense. This imparts a high degree of structural dependence on the binding energy between the two moieties, for example, DNA in A-form interacts more strongly than the B-DNA. Further the interactions of single-stranded DNA are somewhat weaker.

Gresh and Pullman (64) have studied the interaction of ammonium and guanidium ions with phosphate anion by SCF abinitio and empirical procedures. Kumar and Govil (90) have made energy calculations for charged amino acids with a view to understand binding affinities and relative stabilities of their complexes with nucleic acids. They have shown that for a particular DNA structure, the binding energy for the amino acids follow the order Lys > His > Arg. Even at the simplest level of binding of the structural units of DNA and proteins, both sequence and three dimensional structural dependence of the interactions between the two molecules is evident. This can have several implications. Coulombic interactions may bring about conformational variation in the DNA geometry resulting in optimum interactions. Finally, different primary structures of proteins may recognise the same DNA segment with different affinities.

ii) Stacking Interactions: Aromatic amino acids Trp, Tyr, Phe and His (neutral) interact with DNA through stacking interactions (90,129). The main contribution to the total energy comes from dispersion term. In aqueous solutions, additional stabilization may result since the stacked conformers are favoured due to hydrophobic interactions. While stacking of aromatic residues can occur naturally in single-stranded regions of DNA, in double helical DNA, the base-pairs are stacked among themselves at distances of 2.8-3.4 Å. Thus the loop regions and kinks are

natural choices for stacking interactions. Further, the flexible DNA regions may enter into an induced fit type of interaction with protein. The G-C base-pairs show better stacking energies than A-T regions. On the other hand, for the same region, G-C rich DNA domains are more rigid than A-T rich domains. In loop region, purine bases may be preferred over pyrimidines which are poor stackers.

iii) Hydrogen bonding: It involves phosphate group, sugar and bases in nucleic acids and peptide bonds and hydrophilic amino acid side chains in proteins. Amino acid residues, Asp, Glu, Asn and Gln, can form a variety of hydrogen bonded structures with nucleic acid bases. Interactions involving single hydrogen bonds may be relatively rare in specific complexes both because the energy of interaction is low and because a single hydrogen bond may not be adequate to bring the specificity through the directional character of hydrogen bonding interactions (73). With two hydrogen bonds, the interaction energy becomes comparable to stacking interactions. The binding energy for hydrogen bonding involving charged amino acid residues (Arg, Glu and Asp) is considerably higher than that for uncharged residues. Even in double helical DNA structures, the bases can be contacted through hydrogen bonding both in major and in minor groove (73). However the ability to form more than one hydrogen bond requires that the donor and acceptor groups on DNA are placed in the correct orientation to be contacted by the complementary groups on the protein.

(iv) Hydrophobic Interactions : These interactions occur between nucleic acid bases and non polar amino acid side chains. Such interactions are usually weaker and have a non specific character but may contribute significantly to a binding motif formed by other specific interactions (73).

G Literature Review of Gene V Protein

Gene V of bacteriophage fd, M13 and f1 encodes a single-stranded

DNA binding protein (1). This protein is required for single-stranded DNA synthesis (8) and it acts by forming a complex with single-stranded DNA and preventing the synthesis of complementary strand.

The product of Gene V from the F specific phages M13, fd and f1 is identical; it is 87 amino acids in length (36,104). It forms a dimer in solution at physiological salt concentrations (28,117) and binds tightly and cooperatively to single-stranded but not double-stranded DNA (1). In the absence of Mg^{2+} and in moderate salt concentration, its affinity for single-stranded DNA is high enough to lower the T_m (melting temperature) as much as 40°C; it can therefore 'melt' DNA at room temperature. There is no indication that it will do so under conditions prevailing in the infected cell. Since the protein complexed any single-stranded DNA presented, it was described as binding to DNA nonspecifically. More detailed investigation has shown that in fact it does have some specificity; this can be masked by its high affinity and cooperativity. The protein is available in large quantities from filamentous phage-infected cell, and its properties have been studied extensively. Residues involved in binding and the binding mechanism have been studied by chemical modification, by DNA protection and by a variety of spectral methods including NMR and fluorescence depolarization.

^1H NMR (270 MHz) studies were carried out (35) on binding of Gene 5 Protein showing that a large number of aromatic protons undergo change in chemical shift on complex formation. Gene 5 protein contains five Tyr, three Phe, one His and no Trp residues. Nitration of the protein results in nitration of three of the five tyrosyl residues in the protein, Tyr 26, 41 and 56. Nitration of these residues is completely prevented by binding of a tetranucleotide. In addition, ^{19}F NMR (9) studies of gene V protein in which the tyrosyl residues were labelled with m-fluoro-tyrosine identify three surface tyrosyls which interact with oligodeoxy-nucleotides (9). Complexes of gene V protein with

tetra- and octadeoxynucleotides show that 12 of the 37 aromatic protons of the protein undergo upfield shifts upon nucleotide binding. In the complex with $d(T)_8$, the same resonance shifts upfield by 0.8 ppm. These are interpreted as ring current shifts induced by stacking of the phenyl rings of three of the five tyrosyl residues with the bases of the nucleotide. Chemical modification data combined with the base dependent upfield shifts observed for 35% of the aromatic protons of gene 5 protein on complex formation, were used to propose a model in which the bases of the nucleotide were pictured as intercalating with tyrosyl -26, 41 and 56 (35). The aromatic protons which shift on complex formation were assigned to the 2,6 and 3,5 protons of the tyrosyl residues. This study also shows that G5P has a greater affinity for adenine rich region than thymine rich region.

Unambiguous assignments of individual resonances in the aromatic region of the ^1H NMR spectrum of a protein are difficult, especially distinguishing resonances arising from the 2,6 protons of tyrosyl from those of phenylalanyl residues. Assignment can be made more precise by selectively deuterating individual groups of aromatic protons such as the 3,5 or the 2,6 tyrosyl protons. This required a reassignment of the aromatic resonances and indicated a more complex interaction of the aromatic residues with the nucleotides involving both tyrosyl and phenylalanyl residues. Combining the ^1H NMR data on the selectively deuterated nucleotide complexes with ^{19}F NMR data on complexes of the m-fluorotyrosyl labelled protein-oligodeoxynucleotide complexes allows a more detailed model of the interaction of the nucleotide with aromatic residues of the protein than what was possible earlier (35).

The aromatic residues of gene 5 protein in complexes with $d(A)_8$ and $d(T)_8$ have been determined by ^1H NMR of the protein in which the five tyrosyl residues have been selectively deuterated either in the 2,6 or in the 3,5 positions. Only the 3,5 protons of the three surface tyrosyls interact with the bases (9,34,35). The remainder of the aromatic protons undergoing base dependent

upfield ring current shifts on complex formation are phenylalanine protons, assigned to Phe-73 on the basis of model building. ^{19}F NMR of the complexes of the m-fluorotyrosyl labelled protein with $d(pT)_8$ and $d(pA)_8$ confirms the presence of ring current perturbations of nuclei at the 3,5 tyrosyl positions of the three surface tyrosyls. Differential expression of the ^{19}F and ^{1}H (35) nuclear Overhauser effect confirms the presence of two buried and three surface tyrosyl residues.

^{1}H and ^{31}P NMR have been employed (57,58) to study the unwinding of double helical self-complementary tetranucleotide d-CGCG in presence of gene 5 protein. It is shown that the protein is able to unwind the double helical fragment at 0°C. Binding of tetranucleotide causes change in the aromatic part of spectrum of the complex suggesting that aromatic residues, most likely tyrosines, take part in the protein-nucleic acid interactions. From the ^{31}P NMR spectra of the protein-nucleic acid complex, it follows that the pK value of the 5' terminal phosphate is lower than that for the free nucleic acid species. Same group has performed 360 MHz proton NMR studies on deuterated phenylalanine in gene 5 protein. This modified protein yields a simplification in spectrum in the aromatic region. The interaction of modified G5P with d(p-CGCG) and d(p-CGCGCG) shows changes in aromatic region of the NMR spectrum due to binding. It was found that at least one phenylalanine and one tyrosine are involved in the interaction with the oligonucleotides via stacking.

The structure of the gene 5 DNA unwinding protein from bacteriophage fd was determined by X-ray diffraction analysis (98). The protein monomer contains no α-helix. It consists of three secondary structural elements, a radically twisted three stranded antiparallel β-sheet and two distinct antiparallel β-loops which are joined by short segments of extended polypeptide chain. Two monomers are maintained as a dimer by the very close interaction of symmetry related ribbons about the molecular dyad axis. About six residues at the amino and carboxyl terminus are in

extended conformation.

They proposed a high resolution structure of G5P refined to a resolution of 2.3 Å (20). The molecular structure of this protein is found to be primarily β-sheet as proposed earlier. Two G5P monomers are tightly interlocked to form a compact dimer. The dimer is characterized by two symmetry related antiparallel clefts that traverse its surface in a plane essentially perpendicular to the dyad axis. Resident within these clefts are the side chains of those peptide residues that are implicated in nucleic acid complexation (20).

On the basis of difference fourier analysis, a model for the complex arising from the interaction of single-stranded DNA with the gene 5 DNA binding protein has been proposed (21). In this model, tyrosine 26, 34 and 41 are found to be involved in the binding of DNA through stacking of their respective aromatic groups and the DNA phosphate backbone is bound by a series of appropriately positioned lysyl and arginyl side chains. The DNA conformation is fully extended and each binding channel can accommodate up to five nucleotides. Only minor conformational changes in the native G5P structure are required to optimize the binding of DNA.

An excellent refined model of complex of gene 5 protein and DNA was presented by the same group (22). According to this model, phage DNA is wound to the outside of the helical protein ribbon that forms the core of intracellular complex at a density of five nucleotides per G5P monomer. Bound DNA strands are positioned in two contiguous binding channels, which are formed as a consequence of the interactions of complexed G5P dimers. These channels run just inside the outer extended β-loops, composed of residue 20-30, and are separated by approximately 3.2 nm. A number of sterically unacceptable contacts involving residues 38-42, prevent complexation of otherwise complementary dimer surfaces in the absence of nucleic acids. In the process of binding to DNA, these

residues change conformation thereby allowing self assembly of dimer units into a helical structure. These residues act as a two position stereochemical switch that allows or disallows complex formation in response to the absence or presence of DNA.

Further, they proposed a preliminary three dimensional model for the bacteriophage IKe DNA binding protein based on the known high resolution X-ray diffraction structure of a functionally related protein (G5P) from bacteriophage fd (23). Structural comparisons show residues conserved in the primary sequence of both proteins tend to cluster in two regions. The first is essential for the maintenance of dimer association. The second region exists where the two DNA binding channels cross the face of each dimer.

A 500 MHz ^{1}H proton NMR analysis was carried out (6) to investigate the role of lysine and arginine in binding of gene 5 protein to $d(A)_{25-30}$. The aliphatic region of the spectrum in particular 3.5-2.9 ppm region, where lysine ϵ and arginine δ protons are expected to resonate, was considered. Upon binding of $d(A)_8$ to the gene 5 protein, broadening and downfield shift of arginyl resonance was observed. The lysine ϵ proton resonance was found to broaden and shift somewhat to low field. The broadening effects, observed upon binding to $d(A)_8$, were more dramatic than that observed for oligo $d(A)_{25-30}$. The resonances were broadened beyond detection. This increase in linewidth was a result of immobilization of lysine and arginine residues upon complexation.

Selective decoupling and time resolved selective Overhauser experiments were performed on gene 5 protein-$d(A)_8$ complex (4). These experiments allowed a detailed interpretation of the aromatic part of the protein spectrum. The spectrum of the aromatic part of the complex was interpreted in a similar fashion. The ring protons of one phenylalanyl residue and two tyrosyl residues show rather large shifts upon complex formation. Direct evidence for the proximity of these aromatic rings and the DNA fragment in the complex was obtained by additional Overhauser

experiments. It turns out that the H3', H4' and H5' sugar protons of the oligonucleotide are situated near the ring protons of two or three of the aromatic residues, the resonances of which undergo large shifts upon complex formation.

The same group of workers (3) performed studies on the binding of gene 5 protein to oligodeoxyadenylic acids varying in length from two to sixteen nucleotides, by titrating the protein with the oligonucleotides. The mode of binding of the protein and the aromatic part of the spectra have been analysed by performing spectral simulations, starting from the assignments obtained from nuclear Overhauser enhancements. The proton spectra of the complexes of gene 5 protein with $d(A)_8$, $d(A)_{12}$, $d(A)_{16}$ differ from each other and from the spectra obtained from the complexes with longer oligonucleotides. However, binding of all oligonucleotides basically influences the same aromatic residues namely two tyrosines and one phenylalanine. In the protein-oligonucleotide complexes, one protein monomer covers three nucleotide residues, in contrast to the stoichiometry of 1:4 found for protein-polynucleotide complexes. It was found that the binding to the oligonucleotide is cooperative and ionic strength dependent but far less than that found for the binding to polynucleotides.

NMR studies were carried out to compare the gene 5 protein-$d(A)_8$ complex with gene 5 protein-$d(A)_{25-30}$ complex (2). Nuclear Overhauser experiments have been performed on both the systems to obtain structural information regarding the oligonucleotide-protein interactions. In these experiments also, oligonucleotides deuterated at the adenyl C8 positions have been used in order to distinguish between the adenyl H2 and H8 resonances. The results show that the nucleotides in the complexes are situated in such a way that the adenyl H8 and the sugar H1' protons are near the protein surface while the adenyl H2 proton is relatively far removed from all other protons, indicating that this side of the base is pointing away from the protein surface. It has been concluded that structurally different complexes can be obtained

for the $d(A)_{25-30}$ system. The complex with $d(A)_{25-30}$ undergoes a structural transition when going from excess oligonucleotide to excess gene 5 protein. This transition is identified with that caused by increasing length of oligoadenylate $d(A)_n$ (n is the number of residues) binding to protein as it increases from n=3 (oligonucleotide mode) to n=4 (polynucleotide mode). Further, the orientation of the oligonucleotides in the complex is such that adenyl H8 and sugar H1' protons are situated near the protein.

The complexes of gene 5 protein with dpA, $d(pA)_2$, $d(pA)_3$, $d(pA)_4$ and $d(pA)_8$ show that the oligonucleotides carrying a 5' phosphate form 1:1 complexes which involve the same specific binding site for the 5' phosphate dianion. Formation of the dpA mononucleotide complex induces small upfield shift in the resonances of one Phe and one Tyr residue. The magnitudes of these shifts increase as the length of the oligonucleotide increases (33). Phe protons shift by 0.05 ppm for dpA, 0.15 ppm for $d(pA)_2$, 0.25 ppm for $d(pA)_4$ and 0.6 ppm for $d(pA)_8$, while the associated shifts for the Tyr protons are 0.03, 0.14, 0.15 and 0.32 ppm, respectively. Thus the 5' base of the bound nucleotide must rotate closer to the pocket formed by Tyr and Phe or the stacking distance of the Tyr, Phe must decrease as the distal residues of an oligonucleotide are bound. This movement may be induced by conformational changes in the mobile part of the DNA binding domain which contains Tyr 26. As the nucleotide is lengthened, protons from two additional Tyr residues shift upfield. The ring current shifts in the protons of the phenylalanine and one of the tyrosine residues appears to be closely coupled and associated with these two residues at the 5' end of the DNA binding groove.

NMR studies (85,86) of $d(pA)_4$ and $d(pA)_8$ binding have shown that only two aromatic amino acids are involved in stacking interactions with nucleotide bases. Chemical shift changes that accompany binding of $d(pA)_4$, $d(A)_4$, $d(pT)_4$, $d(pA)_8$ combined with specific protein-nucleotide nuclear Overhauser effects (NOEs) suggest this. Chemical shift data also imply a role for Leu 28,

though this has not been confirmed by intermolecular NOEs. In the $d(pA)_4$-G5P complex an intermolecular NOE is observed between Tyr 26 and H1' of adenine 1, while Phe 73 has NOEs with the H2, H8 and H1' of adenine 2 and 3. In the $d(pA)_8$ complex the upfield ring current shifts of the Tyr 26 and Phe 73 resonances become greater than those caused by $d(A)_4$, implying that cooperative interactions result in tighter complex. A comparison of NMR data with X-ray crystallographic data suggests that Phe 73, Tyr 26 and Leu 28, acting in concert, form a dynamic clamp, that is the sole base binding element. This group also performed experiments on complexes between G5P and $d(pA)_{40-60}$, a system with the potential to form over one turn of supermolecular helix. This serves as an improved model system for ^1H NMR examination of the G5P-ss DNA interface under cooperative binding conditions. Selective deuteration of aromatic residues enables individual Tyr (3,5)H and (2,6)H resonances to be monitored in spectra of high molecular weight nucleoprotein assemblies. Analysis of complexation induced chemical shift changes and intermolecular NOEs indicates that Tyr 26 is the only tyrosine to interact directly with single-stranded DNA (86). It is proposed that Phe and Leu side chains stack on either side of a single base, while there is a possibility that Tyr 26 may form hydrogen bond with sugar phosphate backbone in addition or instead of stacking. Chemical exchange effects underscore the dynamic nature of binding at the pocket. A comparison of $d(pA)_{40-60}$ and oligodeoxyadenylates $(dA)_n$ induced chemical shift changes suggests that poly and oligonucleotide complexes have indistinguishable base-binding loci but appear to differ in their dimer-dimer interactions. Alternative polynucleotide binding stoichiometries are explicable in terms of a single base binding model. Chemical shift data are consistent with the proposal that the n=3 and n=4 modes differ basically in the extent to which the sugar phosphate backbone is stretched.

^{13}C NMR spectroscopy of ^{13}C enriched N, N- dimethyl lysyl residues of G5P shows that three of these residues were selectively perturbed by binding to the oligomer $d(pA)_7$. These lysines are

probably directly involved in the nucleic acid binding function of the protein. Negatively charged chelates of lanthanide ions were used to perturb the ^{13}C NMR resonances of labelled lysyl and amino terminal residues of G5P. ^{13}C resonances of Lys 24, Lys 46 and Lys 69 were maximally shifted by the chelate. However the data obtained indicates a substantial movement of the flexible DNA binding loops containing Lys 24 upon binding of the chelate (41).

Several group of researchers have carried out fluorescence depolarization studies on the binding properties of the protein (25,26). Some of the important conclusions from these studies are that the affinity of gene V protein for homopolymers differs from one homopolymer to another such that the binding to poly dA is 2 orders of magnitude higher than that to Poly dT. Further, the affinity for RNA, although lower than that for the corresponding deoxy polymers, is by no means negligible. The authors have also found that there are two different binding modes- one to oligonucleotides, with three nucleotides bound per monomer and one for DNA, with four nucleotides per monomer. The cooperativity of binding was found to be much lower for binding to oligonucleotides, in accord with the results obtained from NMR studies.

Fluorescence studies (39) of the binding of IKe gene V protein to various polynucleotides were carried out to understand the difference in characteristics of its binding from that of M13 gene V protein. It is found that protein-protein interaction surfaces cause the cooperativity in the binding of both proteins diverged during evolution.

The complex isolated from the infected cells contains one molecule of DNA and about 1300 copies of the gene V protein and it forms a filamentous structure which, at first glance, rather resembles the phage (1). It may differ substantially in morphology from that prepared *in vitro* by the addition of purified gene V protein to single-stranded DNA; particularly since gene V protein probably

binds to a short length of displaced single-stranded DNA as (+) strand synthesis occurs *in vivo*. It probably does so in the presence of other single-stranded binding proteins (E.Coli-SSB), and it would thus not be unreasonable for the *in vivo* complex to differ from the one produced *in vitro* by mixing the two purified components under conditions that may not closely resemble those existing in the infected cell.

The *in vivo* complex is 1.1 μm long, 20 nm wide and about 8 nm high and is a ribbon slightly longer than the virus. It is regular repeated structure and has a repeat of 16 nm, which has subsequently been interpreted as representing a helix. Another group of workers (92) crosslinked the protein to the DNA *in vivo*, thereby confirming the association *in situ*. The *in vitro* complex showed branching, which was denser, but it did not show the 16 nm repeats and had a knobbed structure (1). The *in vivo* complex dissociated when NaCl concentration was made higher than 0.1 M. Addition of DNA to the complex in the cell lysate occasionally caused a redistribution, so that some of the added DNA got coated with protein. When a small amount of radio-active DNA was added, it acquired protein from the complex, whereas if a sixfold excess of DNA was added, the complex was demonstrably depleted of protein.

The *in vivo* and *in vitro* complexes were directly compared (135) and there was a little difference between them. In this study, the complexes generated *in vitro* had helical repeats. Their complexes were shorter (0.9 μm) and narrower (10 μm) than those described earlier.

Gray *et al.* (61) extended and repeated these experiments using deuterated phage DNA, which allowed a better estimation of the radius of gyration of the shells occupied by the DNA and the protein. They proposed that the DNA is contained in a shell with a radius of gyration 1.18-2.18 nm while the protein is in a shell extending from 1.49 to 4.49 nm. Hence, the DNA is clearly in the

center of the complex (62). This study was carried out at a nucleotide/protein ratio of 3.

Based on X-ray results, a model for the gene V complex was put forward (21,22). They proposed that the protein dimer binds to DNA in two clefts located on the outside of the dimer pair. They also proposed that initial binding of the DNA causes a movement of Tyr 41 away from the position it occupies in the protein crystal structure. Tyr 41 has been implicated in DNA binding earlier (38). More recent work, however, suggests that Tyr 41 is not involved in DNA interactions (86).

The binding cleft can accommodate four nucleotides easily and a fifth one less easily. However, when the position of Tyr 41 is changed, a binding site is created if two adjacent dimers are bound to the DNA. X-ray researchers found that their model (21) is consistent with the description of the complex derived from electron microscopic images (135). They predicted the formation of a right-handed helix consisting of five nucleotides per monomer, 0.91 μm long with a diameter of 9.3 nm, having a helical pitch of 9 nm and 100 helical turns containing 6.4 dimers per turn, all in good agreement with many of the observed properties of the complex (1,9,28,38). Although, this model is functionally and aesthetically very pleasing, its predictions do not seem to be reconcilable with the conclusions of Gray et al. (61). The nucleotide/protein stoichiometry reported (2,6,25,26), which is four with DNA and three with oligonucleotides, does not agree with the predictions of the Brayer and McPherson model. A CD study gives a stoichiometry of three or four nucleotides per protein for the two binding modes. There is no experimental evidence for stoichiometry of five nucleotides per protein for binding of gene V protein to DNA. However at rather low salt concentrations, a stoichiometry of 3.1 has been reported (83). Recent CD results show that fd and IKe Gene 5 protein undergo minimal conformational changes upon binding to poly (rA) (126).

In addition to the NMR studies on complexes of Gene 5 protein, experiments have been carried out on other systems as well. Most of these studies have shown that aromatic residues of oligopeptides form stacked complexes with nucleic acid bases (42,56,72,95). Stacking was found to be favourable in single-stranded DNA (Poly A, UV-irradiated/denatured DNA) rather than double-stranded DNA.

It has been shown by photochemically induced nuclear dynamic polarisation (CINDP) studies that two out of four tyrosine residues in the N-terminal head piece of the lac repressor are protected in the complex with oligo d(AT) (24). The measurement of isolated headpiece shows that Tyr 7, 17, 47 and His 29 are no longer accessible to the photo sensitizer. These therefore exist in the contact region. The involvement of Tyr 7, 17 in the interaction of N-terminal region of the lac repressor with DNA has also been inferred from chemical modification experiments (8,46).

GVP functions as a master regulatory protein of the expression and replication of the M13 genome (152). With the aid of a binary plasmid *in vivo* test system it was demonstrated that the ss DNA binding protein encoded by gene V of bacteriophage M13 not only regulates the synthesis of its cognate DNA replication proteins at the level of translation, but also of the assembly proteins and the coat proteins encoded by gene I and III, respectively. Further GVP functions as a translational autoregulator of its own synthesis. Comparison of the mRNA levels of gene I and X in the presence and absence of wild type GVP indicated that gene V protein augments the physical stability of these mRNAs.

M13 GVP has the ability to translationally repress the expression of gene II while IKe GVP is unable to do so (151). Deletions in the major GVP target site on gene II mRNA shows that it does not deleteriously effect the production of viral ss DNA. Also the mechanism by which M13 GVP sequesters viral ss DNA differs from the mechanism by which this protein represses the translation of

M13 gene II mRNA.

Picture in solution of a β-loop in mutant Y41—>H of the single-stranded DNA binding protein encoded by gene V of the filamentous phage has been elucidated using 2D NMR techniques (137). They have demonstrated that an identical structural element (residues 13-31 form β-loop, 20-23 form β-turn) is present in wild type GVP and that this element is intimately involved in the binding of gene V protein to ss DNA. However the structure of the DNA binding wing deviates from that proposed for the same amino acid sequence on the basis of x-ray diffraction data.

Hutchinson et al. (81) proposed a model for GVP-DNA complex on the basis of the crystal structure. The model does not implicate Tyr 41 in ss DNA binding but suggests the involvement of Tyr 34 in ss DNA binding under cooperative binding conditions. For constructing the model they used known bond length constraints and postulated protein-nucleic acid interactions (determined from NMR and chemical modification studies). Physical properties of the complex and data from electron micrograph was also considered during model building.

With the help of series of single site mutant proteins of M13 GVP, sequence specific assignments have been reported (50). The solution properties of the mutants of the aromatic amino acid residues have been fully investigated. It has been shown that, for these proteins either none or local changes occur compared to the wild type molecule. Spin-labelled oligonucleotide binding studies of wild type and mutant gene V proteins indicate that Tyr 26 and Phe 73 are the only aromatic residues involved in binding to short stretches of ss DNA. The degree of aggregation of wild type GVP is dependent on both the total protein and salt concentration. The data obtained suggest the occurrence of specific protein-protein interactions between dimer GVP molecules in which the tyrosine residue at position 41 is involved. The so called solubility mutant Y41—>H of GVP has finally made it possible to study the

solution structure of GVP and its interaction with ss DNA by 2D NMR (49). The secondary structure elements present in the protein are deduced from a qualitative interpretation of the NOESY spectra and amide exchange data. The protein is entirely composed of antiparallel β-structure. It is shown that identical structural elements are present in wild type GVP. The N-terminal part of the protein, is in solution, part of a triple stranded β-sheet while in the crystal, it is an extended strand pointing away from the bulk of the protein dimer. One of the antiparallel β-sheets in the protein, which had been designated earlier as the complex loop has, in the solution structure, a different pair wise arrangement of the residues in its respective β-ladders. Residues 30 and 48 are opposite to one another in the solution structure. The amino acid sequences which make up the β-turn are 20-23 and 72-74 residues.

In the solution structure of the DNA binding wing, the β-sheet is shifted 4 amino acids with respect to that of crystal structure. For example, Tyr 26 which is present at the tip of DNA binding loop in crystal forms part of antiparallel β-sheet in solution. This shift in the actual positions of the residues in the β-loop is propogated into the rest of the molecule.

The complete assignment of the ^{1}H NMR of IKe GVP and the subsequent elucidation of its secondary structure has been proposed (138). The major part of this secondary structure is present as an antiparallel β-sheet, i.e. as two β-loops which partly combine into a triple-stranded β-sheet structure. It is shown that a high degree of homology exists with the secondary structure of the ss DNA binding protein encoded by gene V of the distantly related filamentous phage M13.

To study the structure function relationship of GVP, M13 gene V was inserted in a phagemid expression vector and a library of missense and nonsense mutants was constructed by random chemical mutagenesis (134). Phagemids encoding GVP with decreased

biological activities were selected and the nucleotide sequences of their gene V fragments were determined. Furthermore, the mutant proteins were characterized both with respect to their ability to inhibit the production of phagemid DNA transducing particles and their ability to repress the translation of a chimeric lac Z reporter gene whose expression is controlled by the promoter and translational initiation signals of M13 gene II. From the data obtained, it can be deduced that the mechanism by which GVP binds to ss DNA differs from the mechanism by which it binds to its target sequence in the gene II mRNA.

The binding characteristics of the single-stranded DNA binding protein encoded by gene V of bacteriophage M13, are affected by single site amino acid substitutions (133). The series of mutant proteins tested includes mutations in purported monomer-monomer interaction region as well as mutations in the DNA binding domain at positions which are thought to be functionally involved in monomer-monomer interaction or single-stranded DNA binding. The characteristics of the binding of the mutant proteins to homopolynucleotides poly (dA), poly (dU) and poly (dT), were studied by means of fluorescence titration experiments. The binding stoichiometry and fluorescence quenching of the mutant proteins are equal to, or lower than, the wild type gene V protein values. The binding affinities for poly (dA) decrease in the following order:
Y61H> wild type > F68L and F16H > Y41F and Y41H> F73L > R21C > Y34H > G18D/Y56H.
The conservation of binding affinity, for mutations in the ss DNA binding, as well as for mutations in the single-stranded domain, suggests that the binding to homopolynucleotides is largely non specific.

Recently, the DNA binding domain of the ss DNA binding protein GVP encoded by the bacteriophage M13 was studied by means of ^{1}H NMR, through use of a spin-labelled deoxytrinucleotide (51). Using 2D difference spectroscopy, a vast data reduction was accomplished

and the analysis of 2D spectra carried to probe the DNA binding domain and its surroundings.

The DNA binding domain is principally situated on two β-loops. The major loop of the two is the so called DNA binding loop (residues 16-28) of the protein where the residues which constitute one side of the β-ladder (in particular, residues Ser 20, Tyr 26 and Leu 28) are closest to the DNA spin-label. The other loop is part of the so called dyad domain of the protein (residues 68-78), and mainly its residues at the tip are affected by the spin-label (in particular, Phe 73). In addition, a part of the complex domain of the protein (residues 44-51) which runs contiguous to the DNA binding loop is in close vicinity to the DNA. In conclusion, the GVP predominantly utilizes two β-loops for complex formation with ss DNA which are divided over both functional units of the GVP dimer.

More recently, the solution structure of mutant Tyr 41—> His of the single-stranded DNA binding protein encoded by gene V of the filamentous bacteriophage M13 has been investigated by nuclear magnetic resonance spectroscopy (52). Two and three dimensional NMR experiments have been employed with a variety of NMR samples of gene V protein, some of which were uniformly enriched with ^{15}N or ^{13}C. A total of 20 structures were calculated for the M13 gene V protein mutant Tyr 41—>His based on approximately 1000 experimental restraints derived from the NMR data. The orientation of the exposed antiparallel β-loop (residues 16-28) with respect to the core could not be determined. The molecular architecture of each of the monomers include a five stranded β-barrel enclosing a hydrophobic core and two antiparallel β-loops (68-78, 13-31). The dimer structure is stabilized predominantly by hydrophobic residues primarily involving the symmetry related dyad domain (64-82) of the monomers. Residues which are close to bound ss DNA were identified previously from binding experiments (51) with spin-labelled oligonucleotides. The solution structure of gene V protein mutant is consistent with these binding data.

Using contact analysis and a series of restrained molecular dynamics simulations, a model of the complex between single-stranded DNA and the ss DNA binding protein encoded by GVP has been derived (53). Electron microscopy studies (59), indicating that the complex forms a flexible, left-handed helical coil with a diameter of 8 to 9 nm and an average pitch of 9 nm were taken into consideration. The contact analysis served to determine the helix parameters that permit the energetically most favourable packing of protein molecules. Then a protein super helix was built, into which two extended strands of DNA were modelled using restrained molecular dynamics. Specific constraints were included to ensure that the DNA would position itself into the binding groove of the protein. These constraints were based on NMR spin-label experiments (51) which offered a direct identification of the amino acids of the protein present in the DNA binding domain.

Various group of researchers have been looking into protein-nucleic acid interactions and a better picture of these interactions has started to emerge.

H SCOPE OF THE THESIS

Gaps in literature on Gene V Protein

As yet, attempts at direct visualization of the gene V protein-DNA interactions have not been successful. Brayer and McPherson (1984) (21) proposed a model based on X-ray diffraction studies. This model places a chain of five nucleotides, in fully extended conformation, along a DNA binding channel that extends from Tyr 26 to Tyr 41 in 5' to 3' direction. The phosphate backbone is bound by four basic residues, two from the DNA binding loop (comprising of residues 15-32) and two from main body of the protein while each nucleotide is stacked against an aromatic side chain. Though consistent with a reasonable portion of the available physicochemical data, the Brayer-McPherson model is at variance

247155

with several findings for GVP fd DNA and GVP homopolymer complexes.

Several NMR experiments have suggested role of lysines and arginines, Leu 28, Tyr 26, Tyr 41, etc. (2-4,85,86). Studies by use of spin-labelled $d(A)_3$ (51) and restrained molecular dynamics (52) have shown that even number residues in the sequence 24, 26, 28 of GVP fall on one side accessible to solvent and is closest to $d(A)_3$ in bound complex. Some other like Pro 25, Ser 27, Asn 29 are far from DNA surface in the bound complex. However the implication of specific residues by spin label NMR experiments and modelling differ with regard to proximity of several residues (53). The exact role of each and every residue which is in the binding domain or is close to surface of bound DNA complex is not established. The specific interaction -electrostatic, hydrophobic, hydrogen bonding, etc. is not clear (85,86,51,53). The specific change in conformation of DNA to which it binds is not clear.

Why Gene V Protein ?

Protein-nucleic acid interactions occur at all levels of DNA replication, and expression, regulation, cleavage of the specific sequence of DNA by restriction endonuclease, etc. The gene V protein is the smallest and simplest DNA unwinding protein. The gene V protein of bacteriophage fd (M13) plays a negative role in the production of viral single-stranded DNA. This protein blocks the function of a DNA polymerase which is capable of converting intracellular single strands to double-stranded replicative form molecules. Thus the availability of free gene V protein in the cell determines whether replicative form replication or single strand production occur. It also regulates the synthesis of the products of gene II and gene X. Further it forms stoichiometric complex with newly formed progeny single-stranded DNA viral strands. This intracellular complexation event is necessary prepackaging step in the ultimate assembly of the viral particle. It also maintains the single- stranded DNA in the single-stranded state by binding to them. Thus the role of GVP is very crucial as

it maintains DNA in single-stranded form which is necessary for infection. It is one such protein-nucleic acid system for which several studies by fluorescence, CD,NMR, etc. have been done. In particular x-ray crystallographic structure is available to enable us to understand its mode of interaction. The further elucidation of gene V protein-DNA interactions should be useful in the study of affinity exhibited by other DNA binding proteins such as bacteriophage λ, lac operon repressor protein, etc. The specific interaction at molecular level could be a prototype of protein-nucleic acid interactions, in general.

Why study model system ?

Model systems circumvent the problems inherent in natural systems. Due to overlap of resonance and lack of precise assignment, it is very difficult to extract useful information about complicated macromolecular complexes by NMR. A model system not only simplifies the problem but also gives exact information. With this in mind, we have chosen a hexapeptide from the binding domain of gene V protein so as to understand its binding properties.

Extensive research is being carried out on protein-nucleic acid complexes using 2D NMR, fluorescence, X-ray methods, etc. However very little is known about the intrinsic stabilities of the stacked complexes of aromatic amino acids with bases, base-pairs or between two base-pairs. Infact, the situation of protein-nucleic acid complexation is of several orders of magnitude more complicated than DNA-DNA interactions. Details about the energetics of such association are now known. Few theoretical studies have been reported on protein-nucleic acid complexes. Stacking energies of aromatic amino acids with bases and base-pairs have been reported (90,91). The major contribution comes from dispersion term and electrostatic term. It has been found that among pyrimidines, cytosine forms a stable complex whereas among purines guanine forms a stable complex. Overlap geometries indicate overlap of the aromatic ring and hence the formation of a stable complex. Much more information can be

obtained about the protein-nucleic acid associations by theoretical methods. Thus we have carried out potential energy calculations on various model systems.

Lys-Pro-Tyr-Ser-Leu-Asn as a model peptide

The X-ray structure of gene V protein is found to be in β conformation. It consists of three β-loops. One of these loops consisting of residues 15-32 is DNA binding loop. Various physicochemical methods have shown that it is this portion of gene V protein which is actually involved in binding. The peptide sequence Lys-Pro-Tyr-Ser-Leu-Asn is present in the DNA binding loop in the position 24-29. This sequence contains three residues Tyr 26, Lys 24 and Leu 28, whose role has been implicated in binding (85,86). Therefore this peptide was chosen as a model for the present studies.

$d(A)_5$ and $d-(GACTCGTC)_2$ as model oligonucleotides

It has been shown (52,53) that the DNA binding domain of Gene V protein is conserved/semi conserved in closely or distantly related bacteriophage systems. Therefore its binding to ss DNA is rather nonspecfic [133]. Further its affinity for ss DNA is so high that any such base sequence preference will be masked [1]. However few studies on homooligomers have shown that its binding to $d(A)_8$ is 2 orders of magnitude higher than that for $d(T)_8$ (25,26). Since several 1D NMR investigations have been carried out on GVP-oligoadenylate system (2,3,4,85,86) and X-ray crystallographic studies have been carried out to get the structure of $d(A)_5$-GVP complex (21,22), we have choose $d(A)_5$ as our model system.

The oligonucleotide $d-(GACTCGTC)_2$ was chosen as it has a mismatch in the central portion. As Gene V protein binds to DNA of any sequence, the sequence GACTCGTC was taken at random. It is expected that the mismatch in the central portion of octanucleotide will help us in determining a preference for single-stranded DNA or double-stranded DNA.It may be noted that

Garssen et al. (57,58) found that gene V protein binds to double-stranded DNA and showed that the protein has the unique property of opening DNA helix even at 0°C.

In this thesis an attempt has been made to understand the binding of gene V protein to single- and double-stranded DNA using model systems as they circumvent the difficulties inherent in natural systems. This kind of approach has particularly become more relevant with the recent advancements in spectroscopic techniques, particularly, two dimensional nuclear magnetic resonance. Theoretical studies on interaction of nucleic acid bases, base-pairs, two stacked base-pairs, and two stacked bases with aromatic amino acids have been carried out using classical potential functions. A brief introduction to the structure of nucleic acids, proteins and their binding domains alongwith literature survey have been given in this chapter. An introduction to the NMR technique used along with the theoretical methods employed have been outlined in Chapter II. 2D NMR technique has been used to understand the mode of binding of functional part of DNA binding loop of Gene V protein of bacteriophage fd. The results of these investigations on two model systems $d(A)_5$ and $d\text{-}(GACTCGTC)_2$ are given in Chapter III and IV, respectively. Chapter V-VII present the results of theoretical simulation studies. The final chapter presents conclusions on the theoretical and 2D NMR studies undertaken in this thesis.

CHAPTER II

MATERIALS AND METHODS

A Materials

Octanucleotide d-$(GACTCGTC)_2$ was synthesized on a 10 μmole scale on an Applied Biosystems DNA synthesizer (Model 381 A) using cyanoethyl phosphoramidites in Dr. R.V. Hosur's laboratory at TIFR, (Tata Institute of Fundamental Research) Bombay. The starting material is a solid support derivatized with a nucleoside which will become the 3'-hydroxyl end of the oligonucleotide. The nucleoside is bound to the solid controlled porous glass (CPG) support through a linker attached at the 3'-hydroxyl. The 5'-hydroxyl is blocked with a dimethoxytrityl (DMT) group. The steps of the DNA synthesis cycle are as follows:

i) The treatment of the derivatized solid support with acid removes the DMT group and thus frees the 5'-hydroxyl for the coupling reaction. An activated intermediate is created by simultaneously adding the phosphoramidite nucleoside monomer and tetrazole, a weak acid, to the reaction column. The intermediate is so reactive that addition is complete within 30 seconds.

ii) The next step, capping, terminates any chains which did not undergo addition. Capping is done with acetic anhydride and 1-methylimidazole. Capping minimizes the length of the impurities and thus facilitates their separation from the final product.

iii) During the last step, oxidation, the internucleotide linkage is converted from the phosphite to the more stable phosphotriester. Iodine is used as an oxidising agent and water as oxygen donor. This reaction is complete in less than 30 seconds.

After oxidation, the dimethoxytrityl group is removed with trichloroacetic acid, the cycle is repeated until chain elongation is complete. Treatment with concentrated ammonium hydroxide for one hour removes the β-cyanoethyl protecting groups and also

cleaves the oligonucleotide from the support. The benzoyl and isobutyryl base protecting groups are removed by heating at room temperature in ammonia for 8 to 15 hours. Purification was performed on Water's HPLC instrument (RPLC) with acetonitrile and triethyl ammonium acetate (0.01 M) buffer system. The synthesized sequence d-GACTCGTC was annealed by heating it in a water bath up to 70° C and then allowing it to cool until it attains room temperature. Thus a duplex of GACTCGTC i.e. d-$(GACTCGTC)_2$ was prepared. Pentanucleotide $d(A)_5$ was purchased from Centre for Cellular and Molecular Biology (CCMB), Hyderabad, which was purified by FPLC.

Hexapeptide, Lys-Pro-Tyr-Ser-Leu-Asn was synthesized on Model 431 A, Applied Biosystems peptide synthesizer in Prof. V.S.Chauhan's laboratory at International Centre for Genetic Engineering and Biotechnology (ICGEB), New Delhi. Coupling was carried out by the mixed anhydride method. The t-boc group was used to protect the amino termini and was removed with a solution of HCl in diethyl ether. Benzyl ester was used for C-terminal protection. The purification was done on HPLC using double distilled water (30%) and CH_3CN (70%) as buffer.

High Purity D_2O 99.96% was obtained from M/s Merck Sharpe & Dohme Ltd., Canada and M/s ICN Biomedicals Inc., Cambridge, USA.

B Sample Preparation

The oligonucleotide and peptide samples were dissolved in a known volume of phosphate buffer (0.01 M) having pH 7.2 and the concentrations were determined from the measurement of absorbance using Beckman DU-6 spectrophotometer. From this sample, solution of required concentration was prepared.

The values of extinction coefficient used were as follows:

Oligonucleotide/Peptide	Molar extinction coefficient
d-$(GACTCGTC)_2$	9.57×10^3 $M^{-1}cm^{-1}$ at 260 nm

$d(A)_5$	15×10^3 $M^{-1}cm^{-1}$ at 260 nm
Lys-Pro-Tyr-Ser-Leu-Asn	1300 $M^{-1}cm^{-1}$ at 280 nm

The samples were lyophilized and redissolved in D_2O twice. pH was adjusted before lyophilization by addition of small amounts of HCl or NaOH. Finally the sample was redissolved in 0.5 ml of phosphate buffer (0.01 M) in D_2O and transferred to a 5 mm NMR tube. Typically 1 μl of 0.1 M solution of sodium 2, 2 dimethyl-2 silapentane-5 sulphonate (DSS) was added to the sample as internal standard. Further, EDTA was added to the sample to chelate metal impurities. The concentration of EDTA was one tenth of total sample concentration.

8 mM $d(A)_5$ (mol.wt. 1644.5) sample in 0.5 ml D_2O was used for NMR studies whereas the concentration of $d-(GACTCGTC)_2$ (mol.wt. 2381) sample used was 10 mM. The concentration of hexapeptide Lys-Pro-Tyr-Ser-Leu-Asn (mol.wt. 774) was 5 mM.

Preparation of complex :
(a) 0.5 ml of 8 mM $d(A)_5$ sample and 0.5 ml of 5 mM hexapeptide samples were taken as the stock solutions for NMR studies.

A complex of $d(A)_5$ and hexapeptide was prepared by titration. 0.02 ml (V_2) of 8 mM (N_2) $d(A)_5$ was added in steps to 0.38 ml (V_3) of 5 mM (N_3) hexapeptide. The concentration of $d(A)_5$ (N_1) in total volume of 0.40 ml (V_1/V_4)was determined as follows:

$$N_1V_1 = N_2V_2$$
$$N_1 \times 0.40 = 8 \times 0.02$$
$$N_1 = 0.4 \text{ mM}$$

The peptide concentration (N_4) in this solution was determined as follows:

$$N_4V_4 = N_3V_3$$
$$N_4 \times 0.4 = 5 \times 0.38$$
$$N_4 = 4.75 \text{ mM}$$

Table 2.1: Various concentration ratios (N/P) for the complex formed between KPYSLN and $d(A)_5$

$d(A)_5$ conc. (mM)=N	Monomer conc. (mM)	Peptide conc. (mM)=P	Monomer/ Peptide	N/P
0.40	2.00	4.75	0.42	0.08
0.70	3.50	4.50	0.77	0.15
1.09	5.45	4.30	1.18	0.15
1.30	6.50	4.13	1.57	0.31
1.60	8.00	3.95	2.05	0.41
1.92	9.60	3.80	2.52	0.51

Table 2.2: Various concentration ratios (N/P) for the complex formed between KPYSLN and $d\text{-}(GACTCGTC)_2$

Nucleotide conc. (mM)=N	Peptide conc. (mM)=P	P/N
10.00	0.00	-
9.60	0.19	0.02
9.30	0.37	0.04
8.90	0.54	0.06
8.60	0.69	0.08

Likewise five other peptide-$d(A)_5$ complex of different nucleotide/peptide (N/P) or adenine monomer/peptide complex were prepared. The concentration of $d(A)_5$ (N), peptide (P), adenine monomer , adenine monomer/peptide and nucleotide/peptide (N/P) are shown in Table 2.1.

(b) 0.5 ml of 10 mM d-$(GACTCGTC)_2$ and 0.5 ml of 5 mM hexapeptide KPYSLN were taken as stock solutions for NMR studies.

Titration was performed to prepare the complex of d-$(GACTCGTC)_2$ and hexapeptide (KPYSLN). In this case, peptide was added to the nucleotide. This was done intentionally since the aim of the present study, is to monitor the change in conformation of d-$(GACTCGTC)_2$ (on adding peptide) which is a duplex and has a mismatch in the centre.

0.02 ml (V_2) of 5 mM KPYSLN (N_2) was added in steps to 0.5 ml (V_3) of 10 mM d-$(GACTCGTC)_2$ (N_3). The concentration of (N_1) hexapeptide in total volume of 0.52 (V_1/V_4) was determined as follows:

$$N_1V_1 = N_2V_2$$
$$N_1 \times 0.52 = 5 \times 0.02$$
$$N_1 = 0.19 \text{ mM}$$

Likewise, the nucleotide concentration (N_4) in this solution was determined as follows:

$$N_4V_4 = N_3V_3$$
$$N_4 \times 0.52 = 10 \times 0.5$$
$$N_4 = 9.6 \text{ mM}$$

Likewise three other hexapeptide - d-$(GACTCGTC)_2$ complexes of different P/N (Peptide/nucleotide) concentration ratios were prepared. The nucleotide concentration, peptide concentration and various concentration ratios are tabulated in Table 2.2.

CD measurements on hexapeptide, Lys-Pro-Tyr-Ser-Leu-Asn, in H_2O at 297 K were carried out on JASCO spectropolarimeter, located at Delhi University.

C NMR Methods

Particle spins can interact with a beam of electromagnetic radiations. If the beam has the same frequency as that of the precessing particle, it can interact coherently with the particle and energy can be exchanged. The phenomenon, is one of resonance. For nuclei, it is referred to as nuclear magnetic resonance. During recent years, nuclear magnetic resonance (NMR) has emerged as one of the most powerful tools to study the conformation of biological molecules in solution. Due to higher sensitivity, the sample requirement has decreased considerably and a quantity of few mg can now provide detailed information on biomolecular structure and interaction.

Basic Theory of Resonance Absorption

A nucleus of spin I, when placed in magnetic field B_o has 2I+1 equally spaced energy levels with the separation

$\Delta E = h \gamma B_o / 2\pi$

where γ is the magnetogyric ratio and B_o is the magnitude of the applied static magnetic field. The magnetogyric ratio is the proportionality constant which relates the observation frequency for a particular nucleus to the applied field. The observation frequency can be expressed in terms of the magnetogyric ratio and the applied field

$\omega_o = \gamma B_o$

where ω_o is the resonant frequency in radians/seconds. Therefore radiations of frequency ω_o will induce transitions between adjacent energy levels. Thus a nucleus of spin 1/2 has two allowed energy levels, whose state functions are labelled α and β, corresponding to the m_I values of +1/2 and -1/2, respectively, where m_I is the quantum number characterising the z component of I. The spin states for a system of spin 1/2 nuclei are represented as product function of α and β states. Thus there are 2^n basic product functions for a spin system of n spin 1/2 nuclei. Each of these states can be assigned a net m_I value which is the sum of the contributions of the individual spins.

For an ensemble containing a large number of identical nucleus of spin 1/2, all the moments precess at the same frequency and net macroscopic magnetisation is M which orients along Z axis. If a radiofrequency field B_1 is applied in the X-Y plane with a frequency $\omega_{rf} = \omega_o$ then energy will be absorbed. This results in tipping of M vector along Z axis.

NMR parameters

i) Chemical Shift: It is an important parameter which helps in identifying chemical groups present in a system. In a given molecule, nuclei of the same species are subjected to different local magnetic fields although the externally applied field is the same. This is represented by the equation:

$$B_{loc} = B_o (1 - \sigma)$$

where σ is called the screening constant and represents the effect of the electronic environment around the nucleus. The variations in nuclear shielding arise from many factors. Some of the important ones are variations in spatial shielding effects, hydrogen bonding, intramolecular fields due to magnetic anisotropy and the effects of unpaired electron spin. The unit used to express chemical shifts is parts per million (ppm) defined as

$$\delta = \frac{\nu_i - \nu_r}{\nu_r} \times 10^6$$

where ν_i refers to the resonance frequencies of the sample and ν_r refers to the corresponding value for the reference.

ii) Coupling constant : Spin-Spin coupling is the next important phenomenon which occurs between nuclei which are directly bonded such as ^{13}C-^{1}H or ^{31}P-^{1}H or when the number of intervening bonds are small (typically < 4). This coupling causes splitting of the lines. One of the most important parameters in conformational studies of all classes of molecules is the vicinal or three bond coupling constant (^3J) between nuclei A and D in the fragment

A-B-C-D. These couplings are related to the dihedral angle ϕ between plane A-B-C and B-C-D as shown below :

$$J = A \cos^2\phi + B \cos \phi + C$$

where A, B and C are constants which depend on the nature of four atoms in the fragment and the electronegativity of other substituents in the molecule. The values of these constants can be obtained from theoretical calculations or empirically (37). The above equation can be used to get the value of ϕ from coupling constant data.

iii) Relaxation time : The relaxation time is a measure of the time that the spin system takes to return to equilibrium after it has been excited. The spin-lattice relaxation time (T_1) is a measure of the rate with which the z component of magnetisation, M_z, approaches thermal equilibrium. It is defined as the time required for the difference between the excess spin population and its equilibrium value to be reduced by a factor `e' from the starting non-equilibrium value. It depends on several factors, e.g. inter nuclear dipole-dipole interaction, chemical shift anisotropy, scalar coupling, spin rotation, electron-nuclear relaxation, etc. The most important factor for a proton in a diamagnetic sample is the nuclear dipole-dipole interaction.

The spin-spin relaxation (T_2) also occurs through local magnetic field and is basically an exchange of energy between two nuclei of the same type but in opposite spin states. The spin-spin relaxation time (T_2) expresses the rate at which spins can exchange energy with each other. The relaxation time T_2 is reflected in the NMR line shape, the line width at half height being related to T_2 by the formula:

$$(\Delta\omega)_{\text{half height}} = \frac{1}{\pi T_2}$$

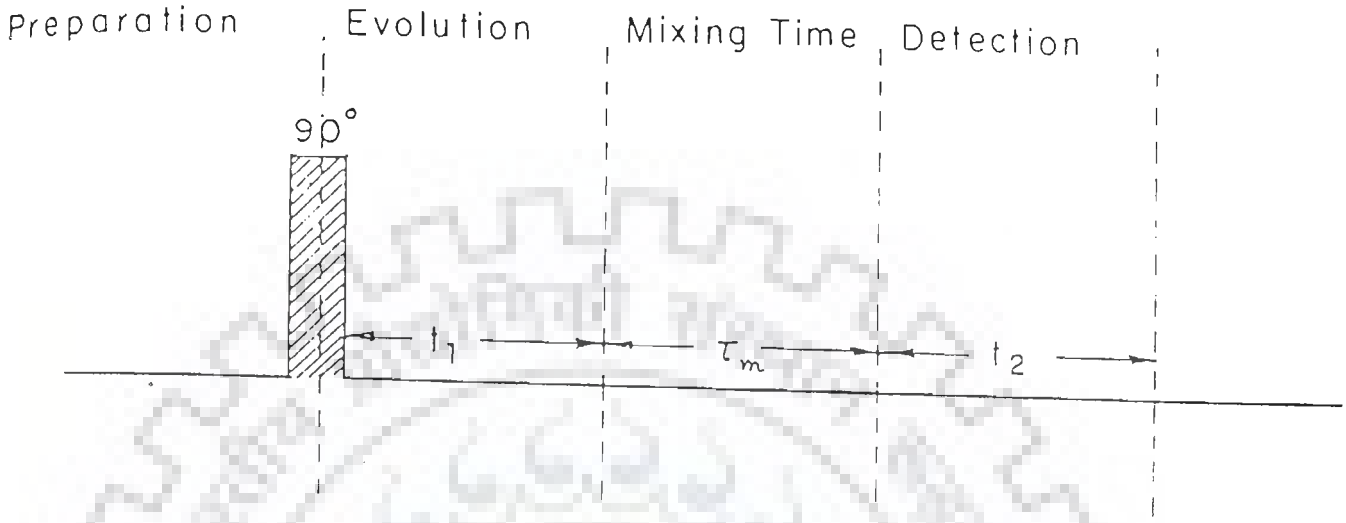


FIG. 2.1 Four different time segments of a 2D NMR experiment namely (i) preparation period (ii) evolution period (t_1) (iii)mixing period (τ_m) and (iv)detection period (t_2)(77).

In practice $$(\Delta\omega) = \frac{1}{\pi T_2^*}$$

where T_2^* is the effective relaxation time which has contributions from the true relaxation time and from field inhomogeneties, so that T_2^* is always less than T_2.

In order to carry out a NMR experiment one needs a static magnetic field B_o for the alignment of the nuclear spins, a radiofrequency field to stimulate absorption and a detector to record the resonance signals. In an FT-NMR spectrometer the sample is excited not by application of a continuous monochromatic radiofrequency field but by a short radiofrequency pulse which contains all the frequencies in the range of possible chemical shifts in a given nucleus. The magnetisation M_o of the sample in the direction of the field B_o is reduced and a transverse magnetisation builds up. The total magnetisation is thus tilted through an angle which depends on the energy of pulses.

The spectra of biological macromolecules and large organic molecules are often too complex and it may not be possible to obtain complete resonance assignments and unambiguous structural information. The advent of 2D NMR, which was first proposed by Jeener in 1971 (82) and later developed by Ernst, has helped a lot in solving such problems. 2D NMR is a simple extension of 1D FT-NMR technique.

2D NMR Techniques

In 2D NMR the frequency domain spectrum is a function of two independent frequency variables, ω_1 and ω_2, and is obtained after a two-dimensional fourier transformation of a time domain function $S(t_1, t_2)$. The two time variables are generated by suitable segmentation of the time axis of the FT-NMR experiment. In all two-dimensional experiments, four different time segments can be distinguished namely preparation period, evolution period (t_1), mixing period (τ_m) and detection period (t_2) (Fig.2.1).

(i) Preparation period - This consists of delay time or a sequence of pulses separated by fixed time intervals or saturation sequence, etc.

(ii) Evolution period (t_1)- The non-equilibrium state of the spin systems created during the preparation period evolves under the influence of a physical environment described by the hamiltonian H. The various spins are frequency labelled. During this period, further radiofrequency pulses may be applied according to the type of experiment being performed. The evolution period can be varied from one experiment to the next experiment.

(iii) Mixing time (τ_m) is a time constant period which consists of pulses and delays of fixed length.

(iv) Detection period (t_2) during which the signal is acquired as a function of t_1.

The basis for the COSY experiment whose pulse sequence is shown in Fig.2.2 is the classical Jeener sequence (82). After the preparation period, a 90° pulse constitutes brief mixing period whose effect is to mix the single quantum coherences into a whole range of orders of coherence. However, only the single quantum coherences will give rise to any measurable signal during the detection period. The mixing process interchanges orders of coherence, mixes coherences among the transitions associated with a given spin and exchanges coherences between spins having a mutual scalar coupling. Thus a magnetisation, initially associated with the A spin of an A-X spin system, may be transferred to spin X through the scalar coupling, J_{AX}. Therefore the A magnetisation in the X-Y plane will also depend upon the Larmor frequency ω_x and the 2D COSY will show signals with frequency coordinates (ω_A, ω_X) and (ω_X, ω_A) as well as at (ω_A, ω_x). The former are the characteristic cross peaks of the COSY spectrum and the latter one is the diagonal peak, which corresponds to the peak in a 1D

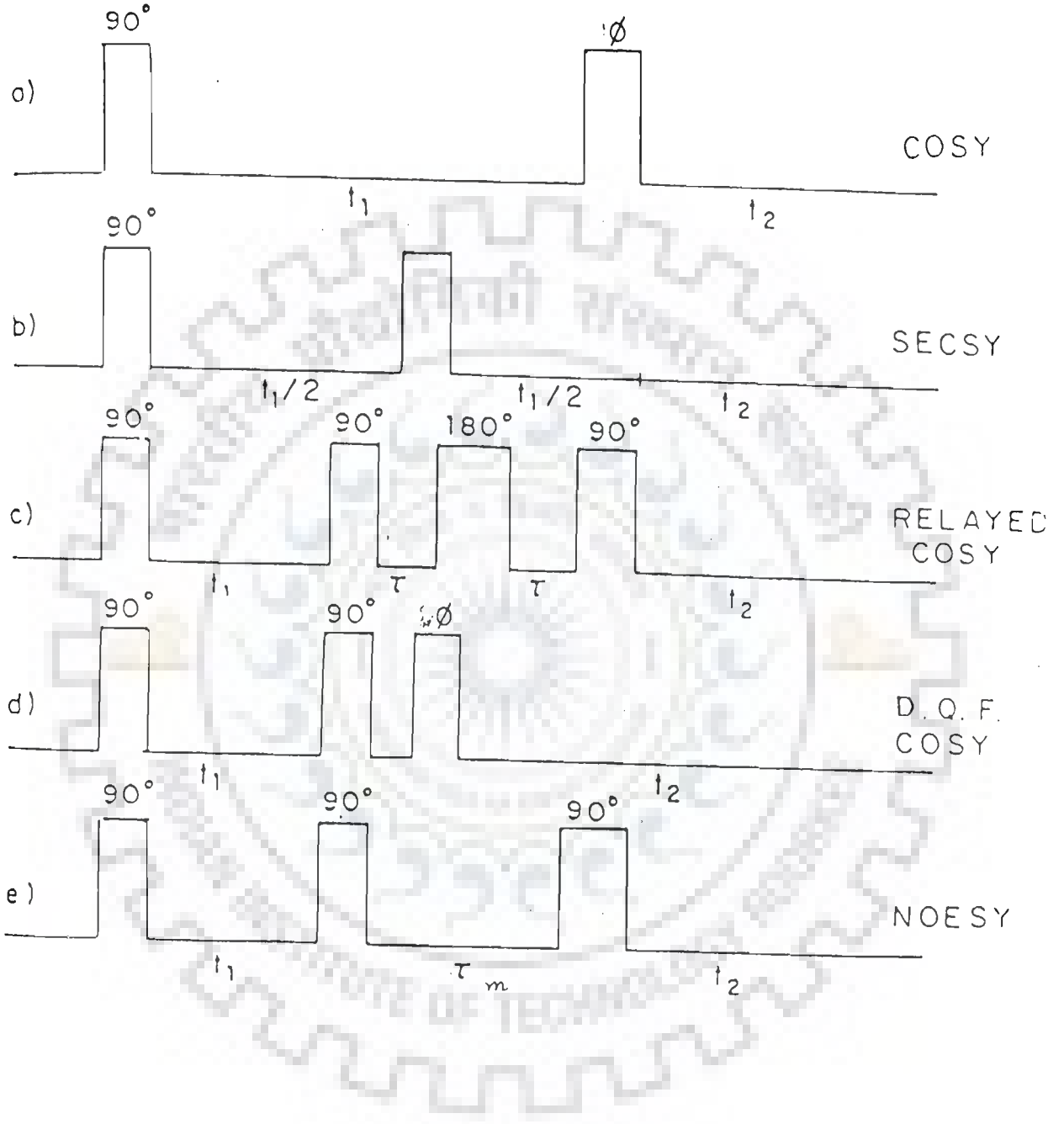


FIG. 2.2 Pulse schemes of various 2D NMR techniques (77).

spectrum. A single 2D spectrum reveals all the scalar couplings in a spin system much more efficiently and clearly than the type of information that could be provided by a series of homonuclear decoupling experiments.

COSY experiment can be carried out with special phase cycling and data processing to change the 2D line shape into pure 2D absorption mode, allowing the use of a phase sensitive display. There are two different methods in use, the first requires the results of two complete COSY experiments with different phase cycling to be added (132) and the second known as TPPI (Time Proportional Phase Incrementation) method uses a single experiment with phase cycling which changes with t_1 increment (19,84,96,120).

The advantage of phase-sensitive COSY is the resolution obtainable of the fine structure of the cross peaks. With pure 2D absorption, the line shapes overlap is much less and thus phase-sensitive COSY can be used for accurate measurement of chemical shift and coupling constants. However this is feasible for small molecules.

The correlation spectroscopy which reflects J-coupling correlation alone is not sufficient for obtaining a complete spectral assignment unless the approximate three-dimensional conformation of the molecule is known. In other words, information about the proximity of atoms in space should also be available. This is achieved in a two-dimensional nuclear Overhauser enhancement spectroscopy (NOESY), which reflects dipolar coupling correlations between protons and indicates which pairs of protons are nearby in space. In basic 2D experiment, the pulse sequence used for making correlation through dynamic NOE is 90°-t_1-90°-τ_m-90° acquisition.

As in COSY, the first 90° pulse produces transverse magnetisation that precesses in the X-Y plane during the evolution period. The 90° mixing pulse is followed by a mixing time τ_m which is of the order of spin-lattice relaxation time T_1. This time is necessary

to achieve magnetisation transfer due to dipolar interactions, unlike scalar interactions which are effective immediately. During the period τ_m the magnetisation can mix, either by direct chemical exchange or in context of the Overhauser experiment through cross relaxation. The first two pulses in effect label the magnetisation M_z of different spins with their chemical shift frequencies. Any exchange of magnetisation M_z during the time τ_m will lead to signals detected during t_2 which have modulation frequencies different from their precession frequencies during t_1 and hence on fourier transformation they will give rise to cross peaks in the 2D spectrum . The third 90° pulse transforms the resulting z magnetisation into transverse magnetisation, which is measured during the detection period. The NOESY experiment is used extensively in the study of biological molecules in solution to give information about protons nearby in space and hence on the 3D structure of the molecule.

Experimental Parameters

All proton NMR experiments were carried out at DST sponsored National FT-NMR Facility located at Tata Institute of Fundamental Research, Bombay and were recorded on a 500 MHz high resolution Bruker AM 500 FT-NMR spectrometer equipped with aspect 3000 computer. The typical parameters for 1D NMR experiments at different temperatures were pulse duration 11.6 μsec ; no. of data points 2-8 K, Spectral width 5000 Hz, no. of scans 60-100 and a resolution of 4.8 Hz/pt to 1.2 Hz/pt. Receiver gain value was optimized in every experiment to get the best signal to noise ratio in limits of good sensitivity. All 2D COSY and NOESY experiments were carried out at room temperature. NMR parameters were 1024-2048 data points along t_2 axis and 256-512 data points along t_1 axis, no of scans 60-200, pulse width 11.6 μsec and sweep width 5000 Hz, a resolution of 4.8 Hz/pt to 9.7 Hz/pt in ω_1 and ω_2 dimensions, relaxation delay of 1 sec and mixing time (τ_m) of 700 ms for $d(A)_5$ and 300 ms for $d\text{-}(GACTCGTC)_2$.

Use of 2D NMR in obtaining the solution structure of Nucleic Acids

The first step in determination of the structure is to obtain complete resonance assignments of the individual protons (48,55,70,78,80,93,101,120,125). Once the assignments have been made, the relative intensities of the various cross peaks in the 2D J-correlated and NOE correlated spectra can be used to derive information about the base orientation, sugar geometry, backbone structure and base-base stacking along the entire sequence of the molecule.

Spectral Assignment

From the NMR point of view, the protons in nucleic acid can be classified into two groups namely the exchangeable NH protons of the bases which resonate between 10-16 ppm and the non-exchangeable base and sugar protons which resonate between 7-9 ppm and 2.0-6.5 ppm, respectively with the exception of methyl protons which resonate between 0.5-2.0 ppm. In case of peptides, the different type of protons are found to resonate in specific regions. For e.g. α-CH's appear around 4-5 ppm. The region between 2-4 ppm contains the resonances belonging to β-CH, γ-CH and ε-CH of amino acids. The δ-CH's appear most upfield at about 1 ppm. The ring protons of aromatic amino acids resonate at 6.8-7.2 ppm.

The protons H1', H2', H2", H3', H4', H5', H5" in each sugar ring of nucleic acid, form a complex J-correlated network. The cross peaks in the 2D J-correlated spectra can be used to identify each spin system within the individual nucleotide units. The proton-proton coupling constants in the above pathway are sensitive to the sugar pucker of the deoxyribose ring with the values ranging between 0-10 Hz. Thus the intensities of the cross peaks in the COSY spectrum vary with the sugar pucker.

Each H1' is expected to show cross peak with the H2' and H2" protons. These cross peaks may have similar or dissimilar intensities or may even be absent depending on the sugar pucker. The H2' and H2" protons are coupled to the H3' proton. Again,

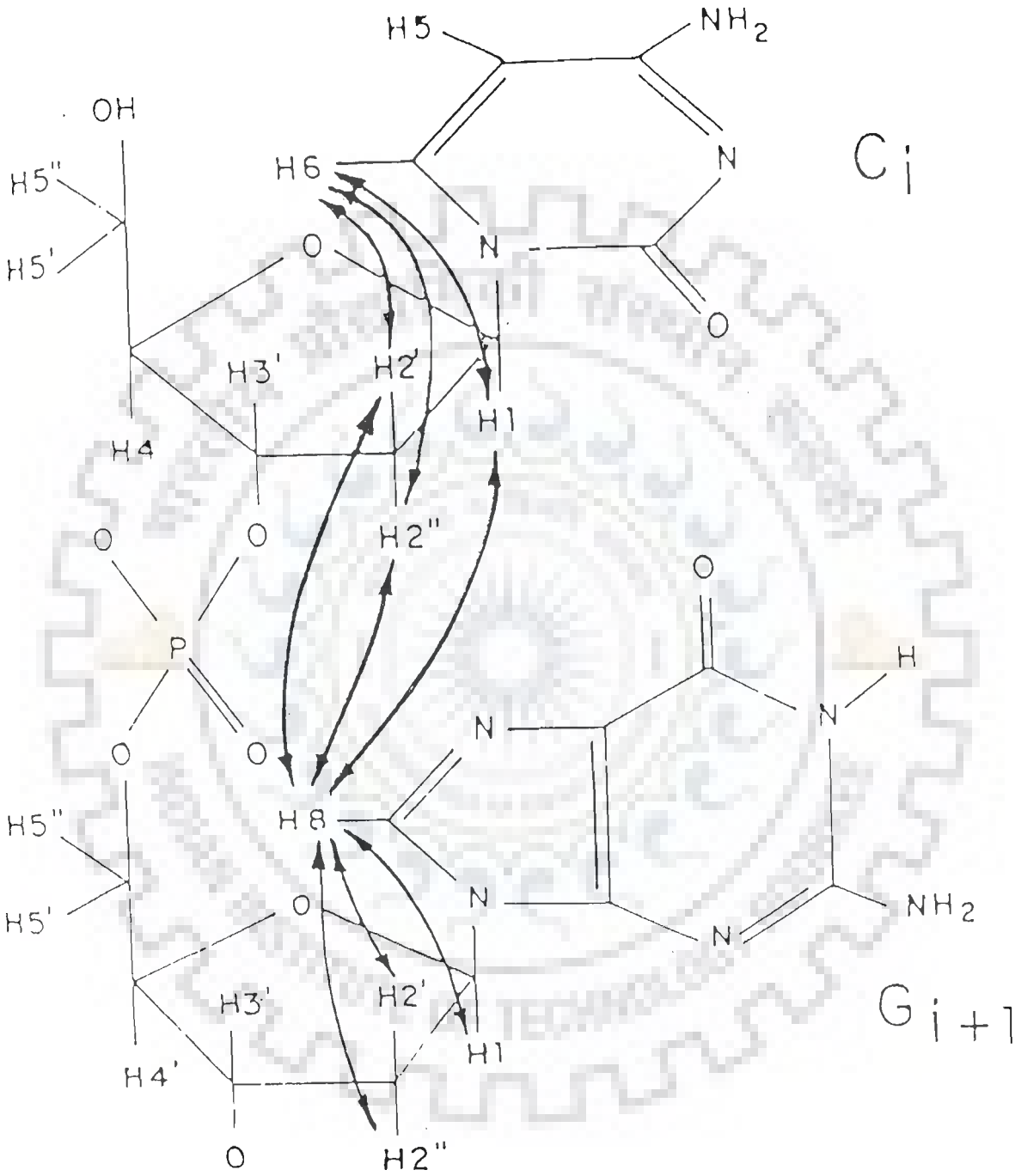


FIG. 2.3 Short interproton distances between adjacent nucleotide units in right-handed DNA which can be used for sequential assignment (80).

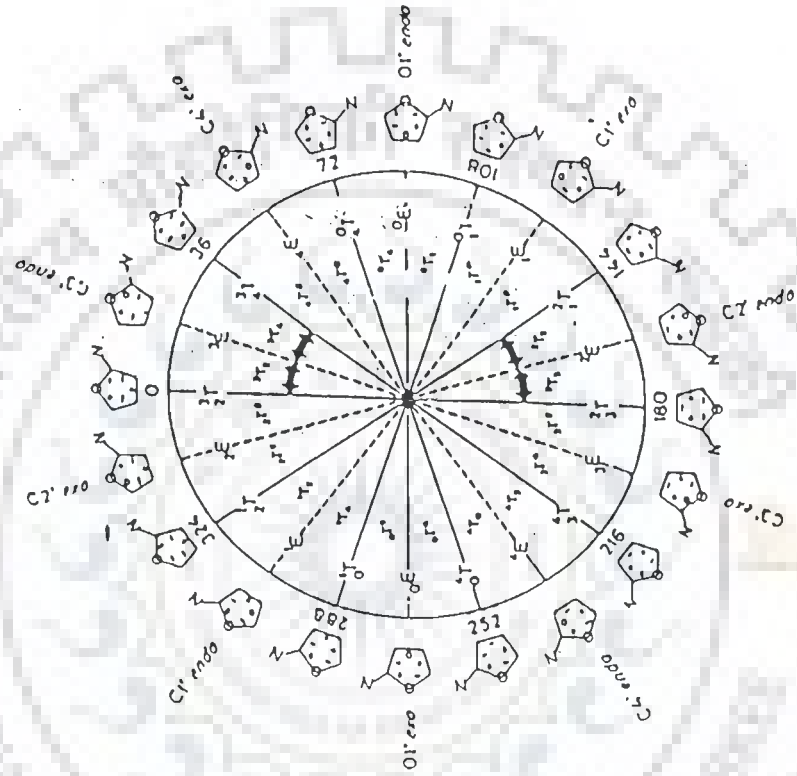


FIG. 2.4 Pseudorotation cycle of the furanose ring in nucleosides. Values of phase angles are given in mutiples of 36° (129).

depending on the sugar geometry, cross peaks may be seen from both the H2' and H2" protons to the H3' proton or from either H2' or H2" to the H3' proton. In turn H3' proton is coupled to the H4' proton, H4' to H5',H5" protons, etc. These variations in the intensities of the cross peaks in the COSY spectrum (magnitude mode) may be used with advantage. Alternatively phase-sensitive COSY spectra is used to estimate spin-spin coupling. This is described in later part of the thesis.

The sequential assignment is carried out using the NOESY spectrum which gives cross peaks corresponding to proton pairs which are close by in space. Short internucleotide distances between adjacent nucleotide units are used as shown in Fig.2.3 (80). In right-handed DNA there are three short internucleotide distances which can be used for the sequential assignment. They are:

(H6/H8)n -----> (H2', H2")n-1
(H6/H8)n -----> (H1')n-1
(H6/H8)n -----> (H8/H6)n+1

In the case of left-handed Z-DNA (119), where the repeating unit is a dinucleotide, the internucleotide pathway is-

Base (2_n - 1) - H5' (2_n - 1) - Base (2_n) - H1'(2_n) - H2' (2_n) and H2" (2_n) - base (2n+1)

Sugar Ring Conformation

The conformation of the deoxyribose ring can be determined by making use of the 3 bond coupling constants (^3J) between the various protons in the sugar ring (66,129). The five J-values of the H1'-H2', H1'-H2", H2'-H3', H2"-H3', H3'-H4' proton pairs are related to the relevant dihedral angle between them namely H-C-C-H according to the relation (77):

$$J = 10.2 \cos^2\phi - 0.8 \cos\phi$$

The dihedral angles ϕ can be calculated in terms of the pseudorotation angle `P'. Fig.2.4 shows the pseudorotation cycle of the furanose ring in nucleoside (129). In these calculations, a

value of 38° has been taken as τ_m. The effect of a larger τ_m is to increase the amplitude of the curves but the position of the maxima and minima remains unchanged. The values of the H1'-H2" and H2'-H3' coupling constant vary within a small range of 6-10 Hz. Thus they are relatively insensitive to the sugar geometry. The values of the H1'-H2', H2"-H3' and H3'-H4' coupling constants however show a much larger variation and can be used advantageously to determine the sugar geometry if the J values can be measured from the NMR spectra.

Under conditions of normal resolution, the intensities of the cross peaks are directly proportional to the magnitude of the coupling constants in a COSY spectra. This is due to the fact that the components of the cross peaks have an antiphase character and tend to cancel each other when the resolution is not high enough to resolve the J separation between them. For a given J value, the cancellation depends on two factors (a) T_2^* or the linewidth (b) the digital resolution along the ω_1 axis. In 2D spectroscopy, the second factor is more important since an attempt to improve the resolution places serious demands on the expensive instrument time. This dependence of the intensities on the J values can be used to obtain information regarding the sugar geometries in oligonucleotides.

From the above discussion, it is clear that depending upon the sugar geometry, certain cross peaks in the COSY spectrum will be more prominent than others and the peaks corresponding to low J values may even be absent. Fig.2.5 shows typical COSY spectra expected for various sugar conformations (80). The C3' endo sugar geometry gives rise to more intense H1'-H2" cross peaks ($J \simeq 10$ Hz) but less intense H1'-H2' cross peak ($J \simeq 0$ Hz). The H3'-H4' ($J \simeq 10$ Hz) cross peak will also be more intense. In case of C2' endo sugar geometry, the H1'-H2' cross peak would be more intense than the corresponding H1'-H2" cross peak and H3'-H4' cross peak would be less intense or absent. Thus even an approximate level of quantification of cross peak intensities leads to important

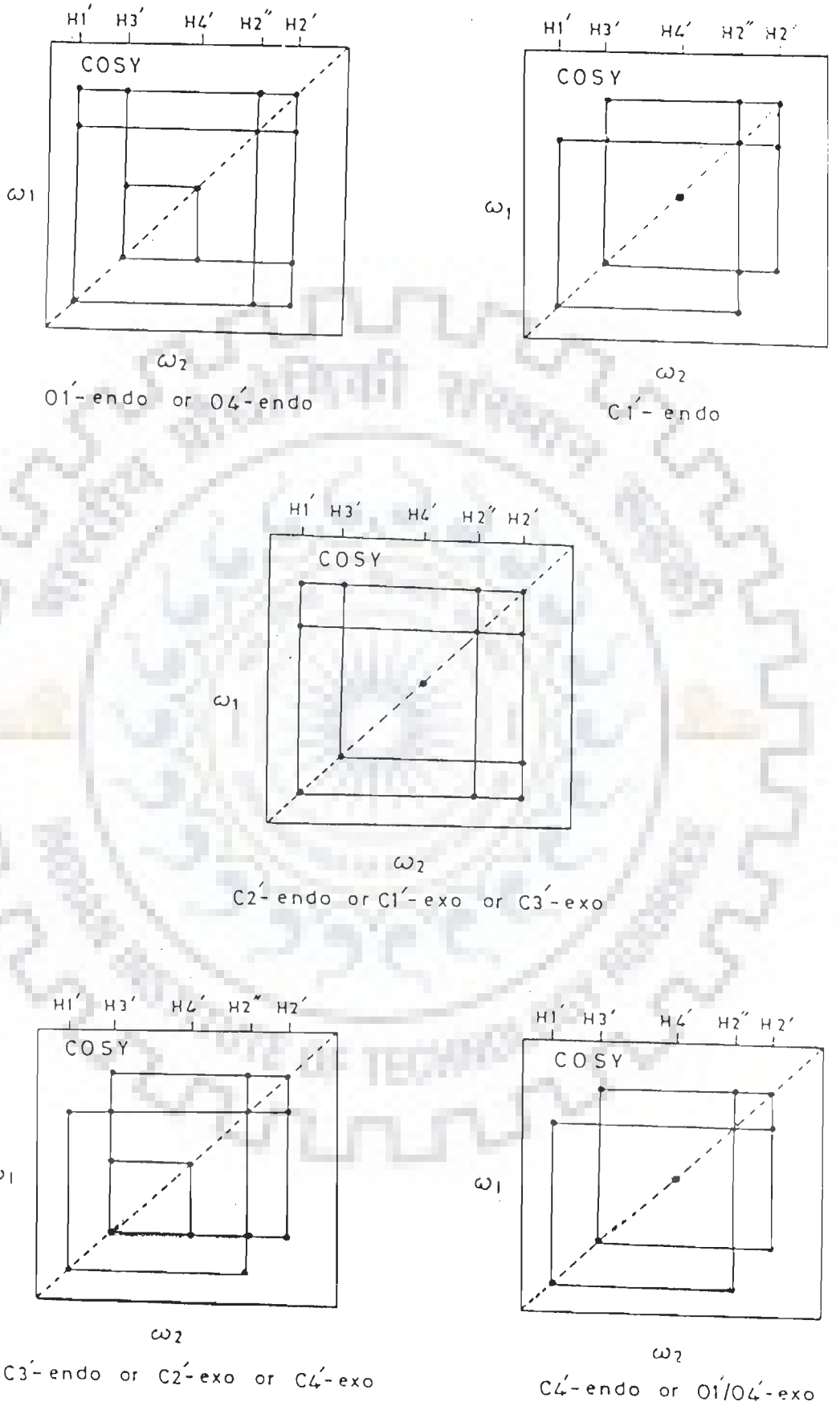


FIG. 2.5 Typical COSY spectra expected for various sugar geometries (80).

conclusions about the sugar geometry.

We have made use of the two strategies to determine the sugar conformation (a) the analysis of cross peak patterns of H1'-H2', H1'-H2", H2'-H3', H2"-H3', spin-spin coupled connectivities from phase-sensitive COSY spectra supported by J values measured as distance between + and - contours due to antiphase components (b) the estimate of intraresidue interproton distances from NOESY spectra. The details have been described in Chapter III (Section C).

Glycosidic bond rotation

Information regarding the glycosidic bond rotation (χ) can be obtained from NOESY spectrum (80,119). An important factor in this case is that in order to relate the intensity of the cross peaks to interproton distances, the experiment must be performed at low mixing time to avoid problems of magnetisation transfer due to spin diffusion. The interatomic distances between base (H8/H6) protons and those belonging to the sugar moiety (H1', H2', H2", H3', H4') depend on the glycosidic bond angle (χ) and the sugar geometry. Since the sugar geometry can be fixed independently from the COSY spectrum, information on glycosidic torsion angle can be obtained from NOESY spectrum using the following strategies.

(i) For a syn conformation, the intranucleotide NOE cross peaks between (H6/H8) and H1' protons are expected to be much stronger than those between the base and H2', H2" protons. In addition, the NOE cross peak between the (H6/H8)-H2" is expected to be stronger than the (H6/H8)-H2' cross peak. (ii) For an anti conformation, the base (H6/H8) protons will show much stronger cross peaks with the H2', H2" protons than with the H1' protons and the NOE peak between the base and H2' proton is expected to be stronger than that between the base and H2" protons. (iii) For the high anti conformation, the H2" protons will show strong NOEs to the base protons of the same nucleotide and the following nucleoside on 3' end.

The glycosidic bond angles have been estimated from the knowledge of the intranucleotide distances between the base protons and their corresponding sugar protons, the details of which have been discussed in Chapter III (Section C).

Estimation of interproton distances

The cross peaks in a NOESY spectrum arise from cross relaxation via the dipole-dipole interactions between protons. The NOESY spectrum can therefore be used to estimate ^1H-^1H distances. For large molecules which satisfy the condition $\omega\tau_c > 1$ (where τ_c is the effective correlation time, which accounts for the motional averaging process on the observed NOE and ω is the spectrometer frequency), cross relaxation is very efficient. If there are several protons in close proximity to each other then a quick diffusion of magnetisation occurs, leading to `spin diffusion'. The extent of diffusion depends on the length of the mixing time (τ_m) used in the NOESY experiment. For short τ_m (< 50 ms), the magnetisation transfer is restricted to a single step and under such conditions the intensity of a cross peak (I_{ij}) is proportional to a single cross relaxation rate ρ_{ij} (where i and j are the relaxing protons).

$$I_{ij} = \rho_{ij}\,\tau_m$$

The intensity of the cross peak in the above equation varies linearly with mixing time and therefore this condition is referred to as a `linear regime'. Under the conditions of an isotropic tumbling motion of a rigid molecule ρ_{ij} is linearly dependent on τ_c (for $\omega\tau_c \gg 1$) and inversely proportional to the sixth power of the interproton distance r_{ij}. The intensity I_{ij} can then be written in terms of these parameters as:

$$I_{ij} = \frac{\gamma^4 \hbar^2 \tau_c \tau_m}{10\, r_{ij}^6}$$

where γ is the proton magnetogyric ratio and $\hbar$ is Planck's constant divided by 2π.

It is thus clear that interproton distances (77) can be estimated by measuring the intensities of cross peaks in the linear regime. In order to determine the acceptable value of τ_m for this purpose, it is advisable to obtain NOE buildup curves as a function of τ_m for several cross peaks, since spin diffusion can be different for different protons. Estimations of correlation times, τ_c, can be obtained from T_2 and T_1 measurements, according to the equation:

$$\tau_c = 2\omega^{-1}(3T_2/T_1)^{-1/2}$$

which holds good for $\omega\tau_c \gg 1$. If protons i, j, k, l have similar τ_c values and if r_{ij} is a known distance, then the unknown distance r_{kl} in a single spectrum can be calculated using the following relationship:

$$\frac{I_{ij}}{I_{kl}} = \frac{r_{kl}^{\sigma}}{r_{ij}^{\sigma}}$$

In oligonucleotides, three reference distances can be utilised for this purpose, namely, (CH6-CH5), (H2'-H2"), and (TH6-TCH_3), where C and T refer to cytosine and thymine bases, respectively. These three vectors can have different effective correlation times and can be selected depending upon the type of cross peak being compared (77,121a).

D CD Methods

Circular dichroism results from the interaction of a given medium with linearly polarized light. The linearly polarized light is regarded as superposition of two equal left and right polarized components. Optically active media are characterised by different indices of refraction for the two components. Circular dichroism depends on wavelength. A plot of the rotation angle as a function of wavelength is designated as ORD spectrum, and that of the

difference in absorption as CD spectrum. Depending upon whether the refractive index or absorption coefficient for the left circularly polarized components have values higher or lower than those for the right circularly polarized component, ellipticity has a positive or negative sign. According to the sign of the CD band, positive and negative Cotton effects are distinguished.

CD could be used to determine the secondary structure of protein. It was assumed that oligopeptides exist only in a single conformation. For proteins, the main standards are the four forms namely, α-helix(H), β-sheet(β), β-turn and random coil(R), the spectra of which are shown in Fig.3.5 (76,68).

E Theoretical Methods

The charges on nucleic acid constituents and amino acids have been calculated by the Complete Neglect of Differential Overlap (CNDO) method (29). The interaction energy has been calculated as a sum of contributions from electrostatic, polarisation, dispersion and repulsion interactions using classical potential functions (CPF). Such an approach has been widely used in the study of protein structures, nucleic acid structures and drug-nucleic acid interactions (30,31,100). We have used perturbation method since the rigorous quantum mechanical method requires large computer resources. In this approximation, the total potential energy is assumed to be:

$$V = V_a + V_r + V_{el} + V_\theta$$

where, V_a is van der Waal's attraction energy

V_r is repulsion energy

V_{el} is the electrostatic energy and

V_θ is the torsional energy

The interactions due to stretching and bending of covalent bonds are neglected as the vibrational constants are considerably higher than the torsional barriers.

Van der Waal's attractive energy also known as London's dispersion

energy arises from the interaction between fluctuating transient dipoles on the two atoms concerned and is of the form $\sum_{i<j}(-A_{ij})/(R_{ij})^{6}$, where i and j are nonbonded atoms and A_{ij} depends on the polarisabilities of atoms i and j. An expression for A_{ij} has been derived as:

$$A_{ij} = \frac{\frac{3}{2}.\hbar.\sqrt{\mu} \;.\alpha_i \alpha_j}{\sqrt{(\alpha i/Ni)} + \sqrt{(\alpha j/Nj)}}$$

where α_i, α_j are the polarisabilities on atoms i and j and N_i and N_j are the effective number of polarisable electrons on atoms i and j.

The repulsion energy expression:

$$V_r = \sum_{i<j} \frac{B_{ij}}{R_{ij}^{12}}$$

is an empirical one where B_{ij} is calculated by imposing a condition that the minimum in total potential energy occurs when the separation between pairs of atoms is equal to the sum of Van der Waal's radii R_o for a given pair of atoms i and j.

The electrostatic energy expression:

$$V_{el} = \sum_{i<j} \frac{q_i q_j}{\in R_{ij}}$$

where q's are the partial charges on atoms arising due to differences in electronegativity of atoms participating in covalent bonds and calculated by standard quantum mechanical approaches. The effective dielectric constant $\in$ is taken as 4.0.

The torsional potential is characterised by the bond about which the rotation is considered and arises from the exchange interactions between the orbitals of the bonded atoms. It is taken as:

$V(\theta) = V1\ (1+ \cos 3\theta) + V2\ (1+\cos 2\theta)$

where V1 and V2 are the barrier heights and have been taken as 1.5

Kcal/mole.

The interaction energy between two molecules is calculated by treating the interaction as a perturbation to the isolated systems and partitioning it into physically meaningful energy terms. The details of method are reviewed elsewhere (29,31). The interaction energy is sum of electrostatic (E_{el}), polarization (E_{pol}), dispersion (E_{disp}) and repulsion energies (E_{rep}):

$$E_{int} = E_{el} + E_{pol} + E_{disp} + E_{rep}$$

The molecular charge distribution has been expressed as a set of atomic charges (q) and atomic dipoles (μ) as obtained by the CNDO method.

For electrostatic interaction:

$$E_{el} = E_{qq} + E_{q\mu} + E_{\mu\mu}$$

where, E_{qq} is monopole-monopole interaction

$E_{q\mu}$ is monopole-dipole interaction

$E_{\mu\mu}$ is dipole-dipole interaction

The explicit expressions for these are:

$$E_{qq} = \sum_{i}^{(1)(2)} \sum_{j} \frac{q_i q_j}{R_{ij}}$$

$$E_{q\mu} = \sum_{i} \sum_{j} q_i (\bar{\mu}_j \cdot \bar{R}_{ij}) \bar{} |R_{ij}|^{-3}$$

$$E_{\mu\mu} = \sum_{i} \sum_{j} (\bar{\mu}_i \cdot \bar{\mu}_j) |\bar{R}_{ij}|^{-3} - (\bar{\mu}_i \cdot \bar{R}_{ij})(\bar{\mu}_j \cdot \bar{R}_{ij}) |\bar{R}_{ij}|^{-5}$$

The polarization energy of a binary complex is given as

$$E_{pol} = E_{pol(2\rightarrow 1)} + E_{pol(1\rightarrow 2)}$$

where $E_{pol\ (1\rightarrow 2)}$ stands for energy due to polarization of molecule 2 by 1. Further

$$E_{pol(2\to1)} = -\frac{1}{2}\sum_{u}^{(1)} \cdot \underset{u}{\in}^{(2)} \cdot A_u \cdot \underset{u}{\in}^{(2)}$$

where $\in$ is electric field induced by molecule 2 at midpoint of bond μ of molecule 1 and A_u is the polarizability tensor of bond μ of molecule 1.

The dispersion and repulsion terms considered together in Kitagordoskii type formula are:

$$E_{disp} + E_{rep} = \sum_{i}^{(1)} \sum_{j}^{(2)} E(i,j)$$

where ,

$$E(i,j) = K_i K_j \left[\frac{-A}{Z_{ij}^{\sigma}} + B e^{-\gamma Z_{ij}}\right]$$

$$Z_{ij} = \frac{R_{ij}}{R_{ij}^{\circ}}$$

$$R_{ij}^{\circ} = \left[(2R_i^{\omega})(2R_j^{\omega})\right]^{-1/2}$$

where R_i^{ω} and R_j^{ω} are Van der Waal's radii of atoms i and j; A, B and γ are constants i.e. A = 0.214, B = 47000 and γ = 12.35; parameters K_i, K_j are taken from literature (29,31).

CHAPTER III

BINDING OF $d(A)_5$ TO FUNCTIONAL PART, Lys^{24}-Pro-Tyr-Ser-Leu-Asn^{29} OF DNA BINDING LOOP OF GENE V PROTEIN

We present here the results of NMR investigations on hexapeptide (Lys-Pro-Tyr-Ser-Leu-Asn), deoxy-penta-adenylate $d(A)_5$ and their complex. The spectral assignments of various protons in peptide and nucleic acid are discussed first. This is followed by the results on conformation of oligonucleotide in the unbound and bound form. The mixture of $d(A)_5$ and hexapeptide is made in oligonucleotide to hexapeptide concentration ratio varying from 0 to 0.51. The changes in chemical shifts of peptide and nucleotide protons in various complexes are then discussed. Finally the observed experimental results are analysed to arrive at a structure of the complex.

A) Spectral assignment in hexapeptide

The hexapeptide Lys-Pro-Tyr-Ser-Leu-Asn (KPYSLN) (Fig.3.1(a)) comprises of a part of the DNA binding loop, residue number 24 to 29 of Gene V protein of bacteriophage M13/fd/f1. The 500 MHz proton NMR spectra of 5 mM KPYSLN in D_2O is shown in Fig.3.1(b). The assignment of various resonance peaks is straightforward as none of the residues are repeated in the model hexapeptide. The chemical shift of the amino acid residues available in literature (150) serves as a guideline for the present work. The two doublets around 7 ppm are expected to be due to Tyr ring protons. The α-CH protons of each residue appears around 4.5 ppm and δCH_2, γCH_2, β-CH_2 protons (except Asn and Pro) appear in range 1.4-1.9 ppm. The exact position of all the other resonance peaks are however unambiguously assigned on the basis of 2D COSY spectra (Fig.3.2(a)) which gives all J-coupled connectivities. The expanded portions of some of the regions of phase-sensitive 2D COSY spectra are given in Fig.3.2(b-g) to highlight specific connectivities.

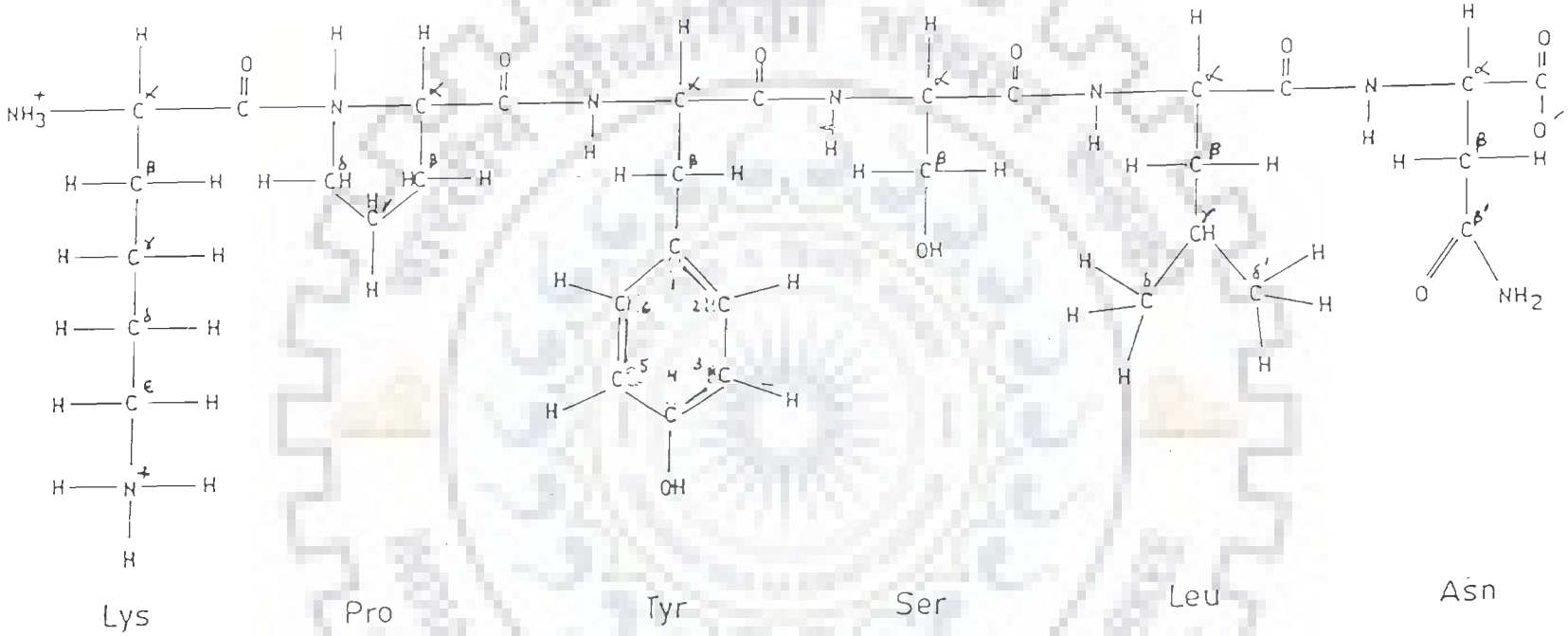


FIG. 3.1(a): Schematic representation of Lys-Pro-Tyr-Ser-Leu-Asn (KPYSLN).

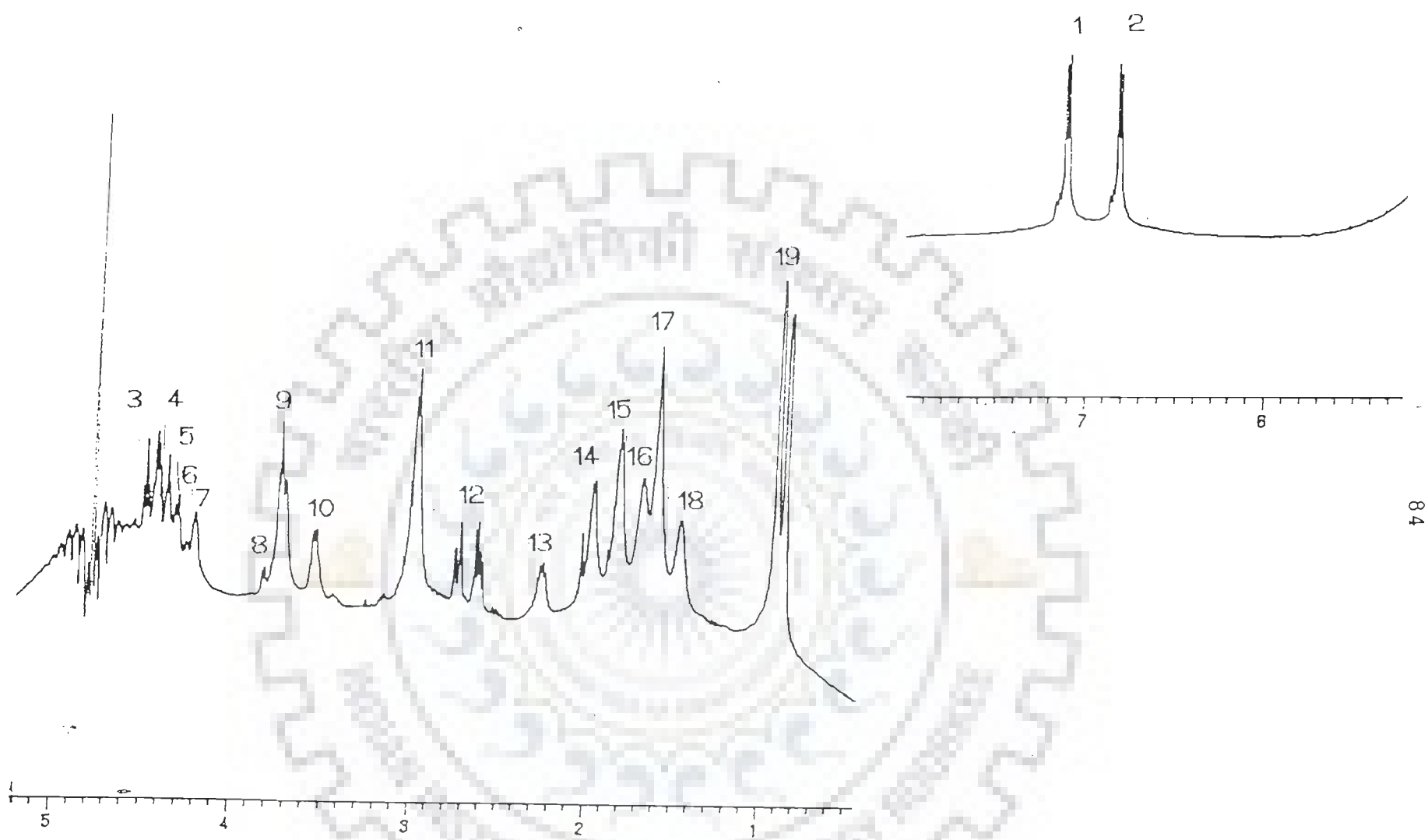


FIG. 3.1(b): 500 MHz proton NMR spectrum of 5 mM KPYSLN in D_2O solution (phosphate buffer 0.01 M) at 297 K. Inset shows Tyr(2,6)H and Tyr(3,5)H peaks. Various protons are labelled serially as (1) Tyr (2,6)H, (2) Tyr (3,5)H (3)Tyr α-CH (4)Pro α-CH, Lys α-CH (5)Asn α-CH (6)Ser α-CH (7)Leu α-CH (8)Ser β'-CH_2 (9)Pro δ'-CH_2, Serβ-CH_2 (10)Pro δ-CH_2 (11)Lys ε-CH_2, Tyr β-CH_2 (12)Asn β'-CH_2, Asn β-CH_2 (13)Pro β-CH_2 (14)Pro γ-CH_2 (15)Lys β-CH_2 (16)Lys δ-CH_2 (17)Leu β-CH_2, Leu γ-CH_2 (18)Lys γ-CH_2 (19)Leu δ-CH_3, Leu δ'-CH_3

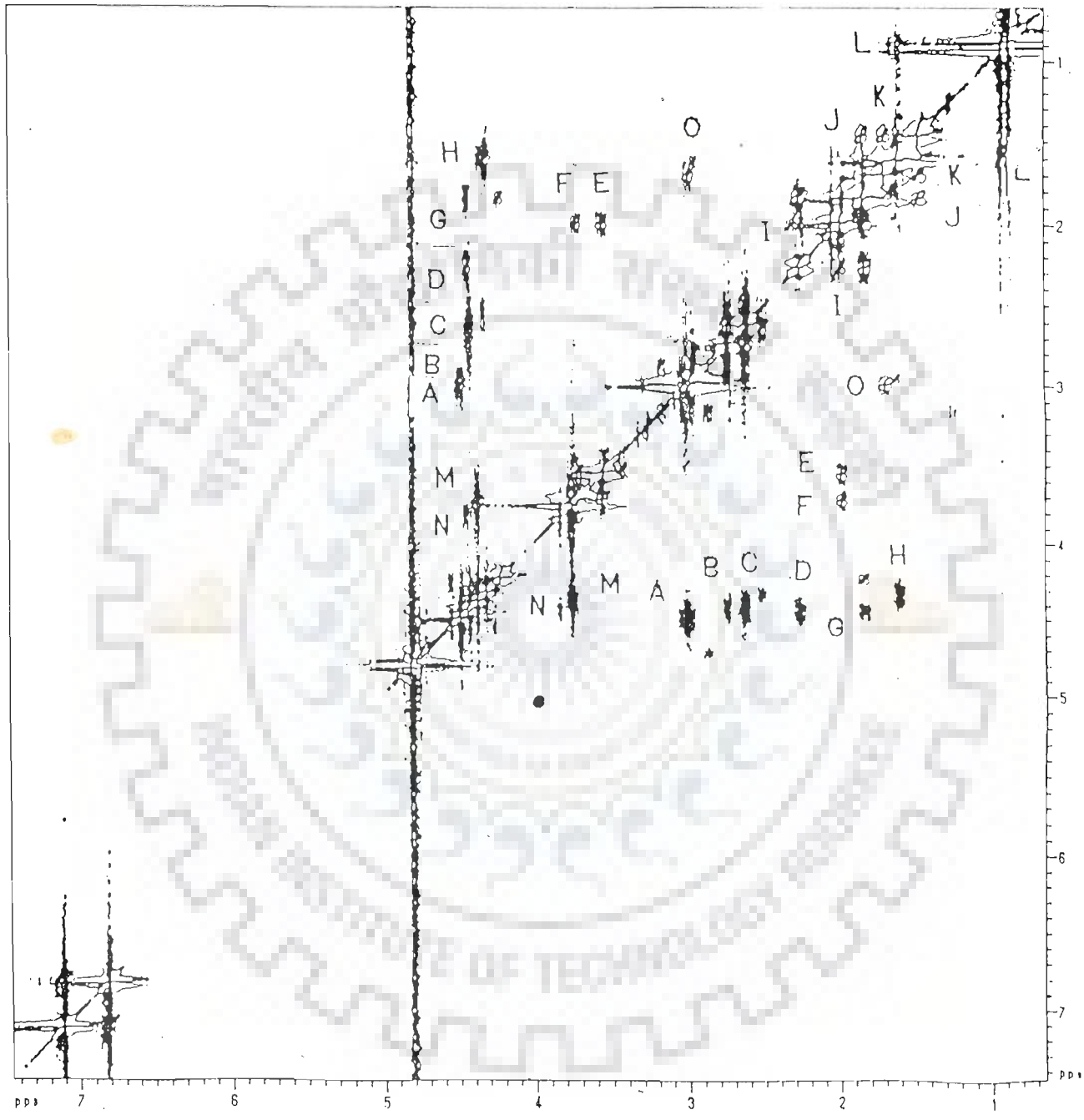


FIG. 3.2(a):Phase-sensitive COSY spectrum of 5 mM KPYSLN in phosphate buffer (0.01 M) at 297 K. Ref. DSS.

The Asn β, β'-CH_2 protons show strong geminal coupling with each other and are both coupled to Asn α-CH proton. The multiplets of Asn β, β'-CH_2 protons appearing at 2.63 and 2.75 ppm show strong J-connectivity with the multiplet at 4.43 ppm (Fig.3.2(b) and (d)) which therefore gets assigned to Asn α-CH proton. The methyl protons Leu δ-CH_3 and Leu δ'-CH_3 appear as sharp lines around 0.91 ppm. Leu δ-CH_3 are coupled to Leu γ-CH_2 which is coupled to Leu β-CH_2. The Leu β-CH_2 is further coupled to Leu α-CH proton as shown in Fig.3.2(f). Tracing these J-coupled connectivities in Fig.3.2(b-g), the α-CH at 4.32 ppm gets assigned to Leu α-CH. Ser β and Ser β'-CH_2 protons expected to resonate around 3.7 ppm give J-connectivity with one of the α-CH's Fig.3.2 (f, d and g), which therefore gets assigned to Ser α-CH proton. The Tyr β-CH_2 expected to resonate around 3 ppm gives J-connectivity with one of α-CH's which resonates at 4.50 ppm and gets assigned to Tyr α-CH proton (Fig.3.2b,d,e). In a similar way Pro δ-CH_2 is coupled to Pro γ-CH_2 which is coupled to Pro β-CH_2. The Pro β-CH_2 is coupled to Pro α-CH proton. These J-coupled connectivities assign the resonance at 4.45 ppm to Pro α-CH proton. The Lys ε-CH_2 appearing around 3 ppm are connected to Lys side chain methylene protons. Lys β-CH_2, γ-CH_2, δ-CH_2 appearing in the region 1.4-1.8 ppm show J-coupled connectivities (Fig.3.2(b-d)). The methylene peak at 1.84 ppm is coupled to the α-CH proton. Thus each and every proton of the hexapeptide gets assigned unambiguously. Table 3.1 gives a complete list of their chemical shift positions in the peptide spectrum.

B) Conformation of hexapeptide, Lys-Pro-Tyr-Ser-Leu-Asn

1) Fig.3.3(a) shows a typical CD spectrum of α-helix, β-sheet, β-turn and random coil conformations extracted from the spectra of proteins of known three-dimensional structure. The CD spectrum of type I and type II β-turn of $(Ala_2\text{-}Gly_2)_n$ and Pro-Gly-Leu, respectively, are shown in Fig.3.3(b) (76). In type I and type II β-turns, following characteristic features are observed:

FIG. 3.2(b): A portion of phase-sensitive COSY spectrum of 5 mM KPYSLN in phosphate buffer (0.01 M) at 297 K expanded to show specific connectivities : (A) Tyr α-CH - Tyr β-CH_2 (B) Asn α-CH - Asn β-CH_2 (C) Asn α-CH - Asn β'-CH_2 (D) Pro α-CH - Pro β-CH_2 (E) Pro γ-CH_2 - Pro δ-CH_2 (F) Pro γ-CH_2 - Pro δ'-CH_2 (G) Lys α-CH - Lys β-CH_2 (H) Leu α-CH - Leu β-CH_2

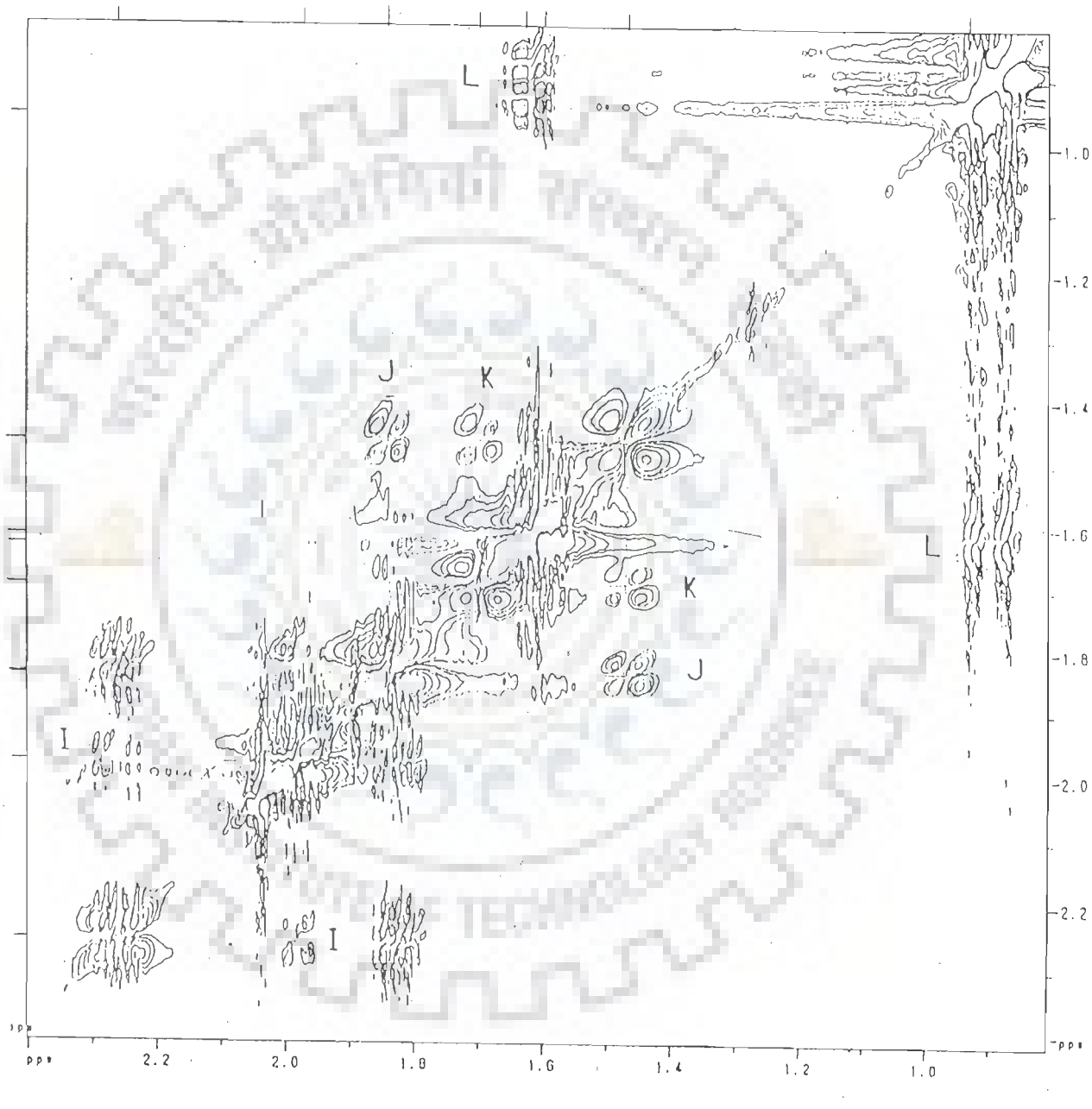


FIG. 3.2(c): A portion of phase-sensitive COSY spectrum of 5 mM KPYSLN in phosphate buffer (0.01 M) at 297 K expanded to show specific connectivities : (I) Pro γ-CH_2 - Pro β-CH_2 (J) Lys γ-CH_2- Lys β-CH_2 (K) Lys γ-CH_2 - Lys δ-CH_2 (L) Leu γ-CH_2 - Leu δ-CH_3

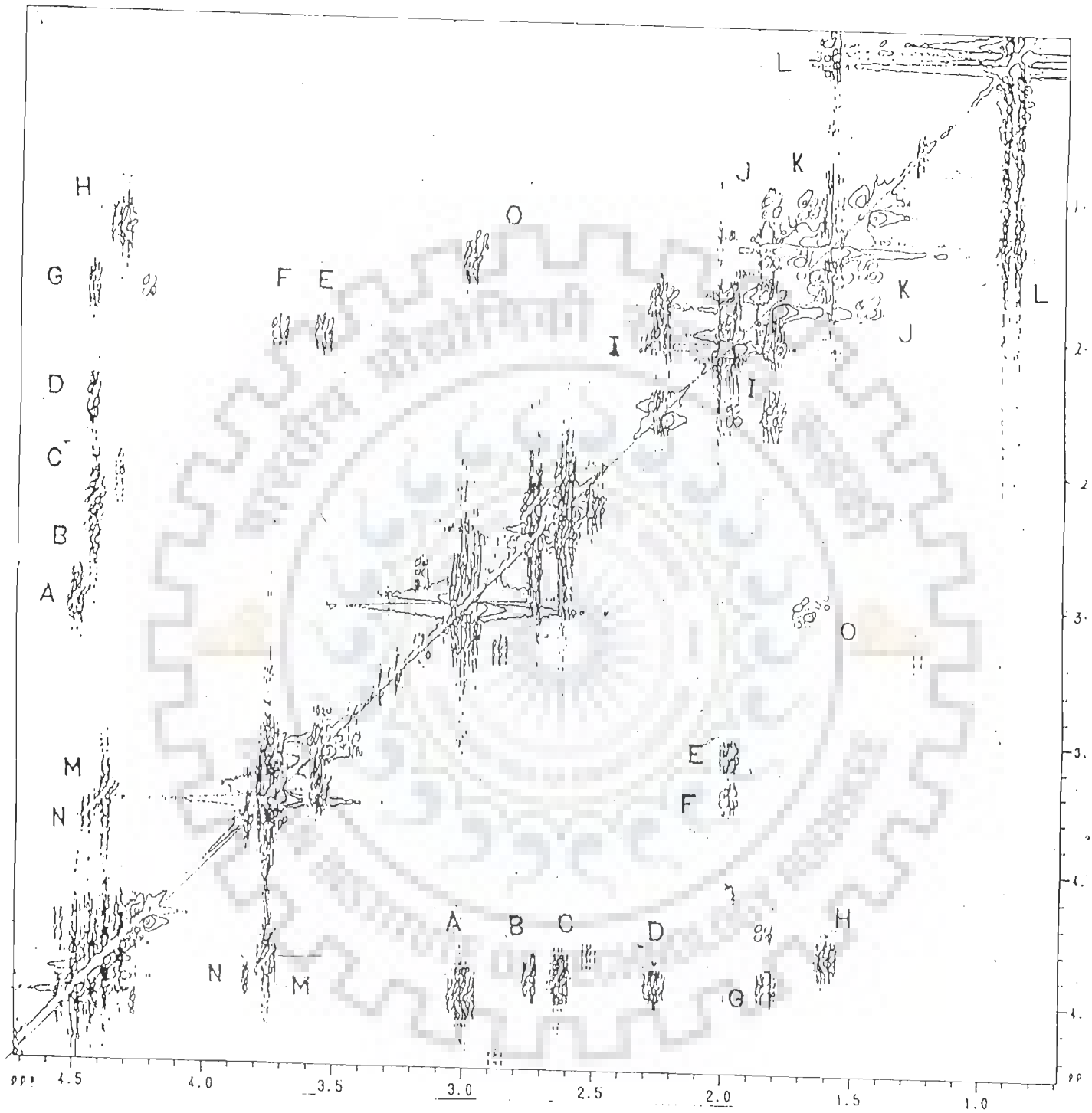


FIG. 3.2(d): A portion of phase-sensitive COSY spectrum of 5 mM KPYSLN in phosphate buffer (0.01 M) at 297 K expanded to show specific connectivities: (A) Tyr α-CH - Tyr β-CH_2 (B) Asn α-CH - Asn β-CH_2 (C) Asn α-CH - Asn β'-CH_2 (D) Pro α-CH - Pro β-CH_2 (E) Pro γ-CH_2 - Pro δ-CH_2 (F) Pro γ-CH_2 - Pro δ'-CH_2 (G) Lys α-CH - Lys β-CH_2 (H) Leu α-CH - Leu β-CH_2 (I) Pro γ-CH_2 - Pro β-CH_2 (J) Lys γ-CH_2 - Lys β-CH_2 (K) Lys γ-CH_2 - Lys δ-CH_2 (L) Leu γ-CH_2-Leu δ-CH_3 (M) Ser α-CH - Ser β-CH_2 (N) Ser α-CH - Ser β'-CH_2 (O) Lys δ-CH_2 - Lys ε-CH_2

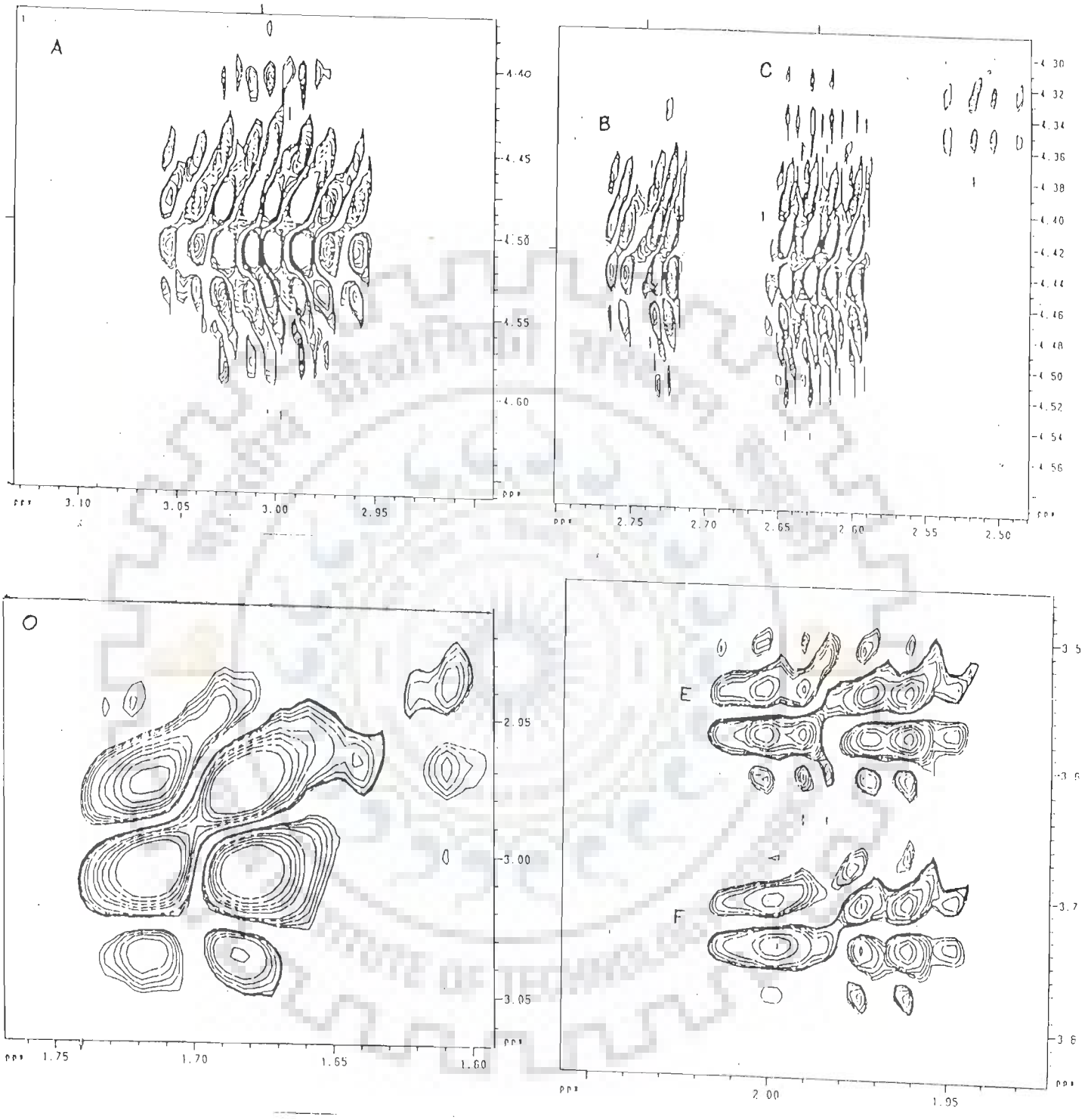


FIG. 3.2(e): A portion of phase-sensitive COSY spectrum of 5 mM KPYSLN in phosphate buffer (0.01 M) at 297 K expanded to show specific connectivities: (A) Tyr α-CH - Tyr β-CH_2 (B) Asn α-CH - Asn β-CH_2 (C) Asn α-CH - Asn β'-CH_2 (E) Pro γ-CH_2 - Pro δ-CH_2 (F) Pro γ-CH_2 - Pro δ'-CH_2 (O) Lys δ-CH_2 - Lys ε-CH_2

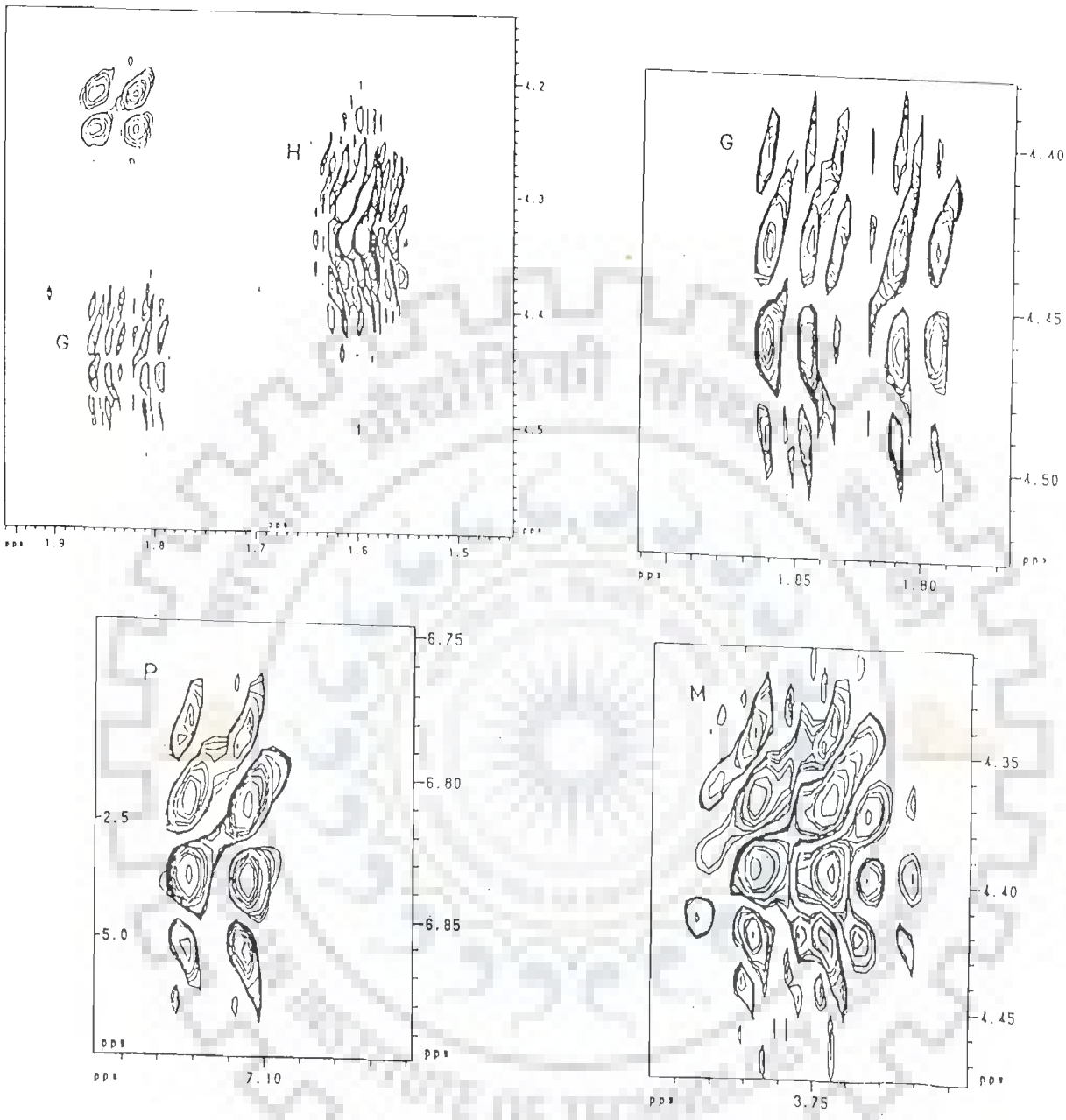


FIG. 3.2(f): A portion of phase-sensitive COSY spectrum of 5 mM KPYSLN in phosphate buffer (0.01 M) at 297 K expanded to show specific connectivities: (G) Lys α-CH - Lys β-CH_2 (H) Leu α-CH - Leu β-CH_2 (P) Tyr(2,6)H - Tyr(3,5)H (M) Ser α-CH - Ser β-CH_2

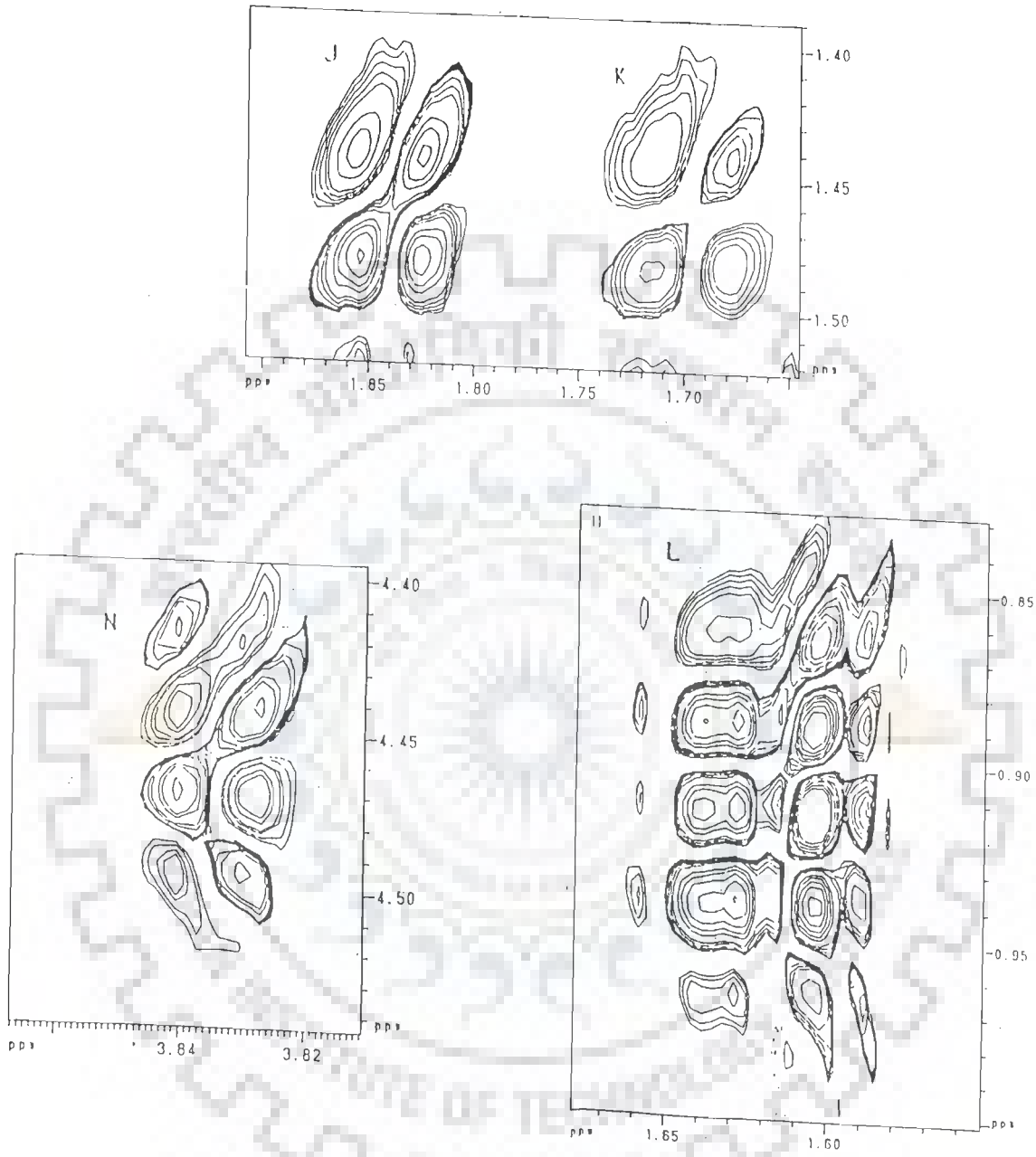


FIG. 3.2(g): A portion of phase-sensitive COSY spectrum of 5 mM KPYSLN in phosphate buffer (0.01 M) at 297 K expanded to show specific connectivities: (N) Ser α-CH - Ser β-CH_2 (L) Leu δ-CH_3 - Leu γ-CH_2 (J) Lys β-CH_2 - Lys γ-CH_2 (K) Lys δ-CH_2 - Lys γ-CH_2

Table 3.1: Chemical shift values δ (in ppm) of various protons of 5mM hexapeptide, KPYSLN in phosphate buffer (0.01 M) at 297 K.

	Lys	Pro	Tyr	Ser	Leu	Asn
α-CH	4.45	4.45	4.50	4.40	4.32	4.43
β-CH_2	1.83	2.26	3.06	3.75 (β'3.83)	1.61	2.63 (β'2.75)
γ-CH_2	1.47	1.96	-	-	1.62	-
δ-CH_2	1.70	3.56 (δ'3.72)	-	-	-	-
δ-CH_3	-	-	-	-	0.91	-
ε-CH_2	2.98	-	-	-	-	-
(3,5)H	-	-	6.87			
(2,6)H			7.15			

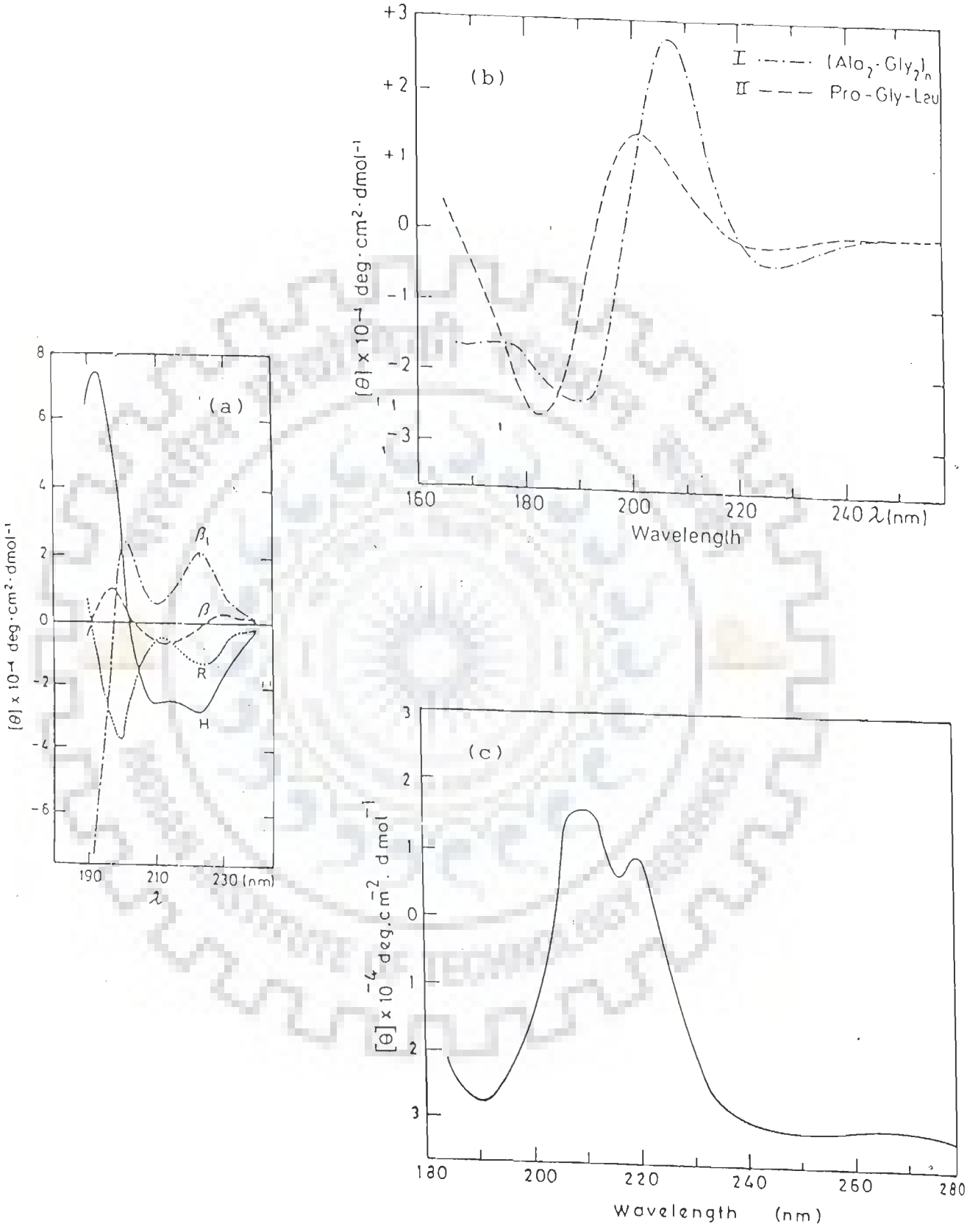


FIG. 3.3:(a) CD reference spectra for α-helix (H), β-form (β), β-turn (βt), and random coil (R) (76), (b) CD reference spectra of standard peptide; β-turn, type I and II and (c) CD spectrum of 0.14 mM KPYSLN in H_2O, at 297 K.

	Parameter determined	Type I	Type II
1.	Wavelength at which peak maxima occurs	208 nm	200 nm
2.	[Ø]x 10^{-4}deg cm^2dmol^{-1} at 200 nm	1.25	1.5
3.	[Ø]x 10^{-4}deg cm^2dmol^{-1} at 210 nm	2.75	0.7
4.	[Ø]x 10^{-4}deg cm^2dmol^{-1} at 240 nm	-0.3	-0.1

The data clearly indicates that in type I β-turn, the value of ellipticity at 210 nm is twice the ellipticity observed at 200 nm whereas at 240 nm, the value of ellipticity reduces to one-ninth of that observed at 210 nm. On the other hand, in type II β-turn, the relative value of ellipticity at 200 nm and 210 nm are just the reverse of that of type I β-turn.

We have taken CD spectra of 0.14 mM hexapeptide Lys-Pro-Tyr-Ser-Leu-Asn in H_2O at room temperature. Fig.3.3(c) shows the CD spectrum obtained. The following observations are made:

1.	Wavelength at which peak maxima occurs	=	210 nm
2.	[Ø]x 10^{-4}deg cm^2dmol^{-1} at 200 nm	=	1.39
3.	[Ø]x 10^{-4}deg cm^2dmol^{-1} at 210 nm	=	2.50
4.	[Ø]x 10^{-4}deg cm^2dmol^{-1} at 240 nm	=	0.24

It should be noted that the observed value of ellipticity at peak maxima is twice of that observed at 200 nm. However at 240 nm, the value reduces to one tenth of the value obtained at 210 nm. Since the peak is observed at 210 nm and the ratios of value of ellipticity at 200 nm and 210 nm observed in our case are same as that observed for type I β-turn, we conclude that our hexapeptide adopts a type I β-turn conformation.

2) Chou Fasman technique is used to predict the conformational preference for the various amino acids. The notion is that certain residues are more likely to exist as α-helices than others. The fractional occurrence of each residue in an α-helical section, fα,

is defined as $f\alpha = n\alpha/n$, where n is the total number of residues of a given kind and $n\alpha$ is the number of these residues that are in α-helical sections. Then $\langle f\alpha \rangle$ is the average value of $f\alpha$, that is, the sum of the $f\alpha$ values divided by the number of residues. The conformational parameter $P\alpha$ is defined as $f\alpha/\langle f\alpha \rangle$. $P\beta$ and Pt are defined in the same way.

$P\alpha$, $P\beta$ and Pt for various amino acids present in our hexapeptide are:

	$P\alpha$	$P\beta$	Pt
Lys	1.23	0.77	0.96
Pro	0.52	0.64	1.91
Tyr	0.72	1.25	1.05
Ser	0.82	0.95	1.33
Leu	1.30	1.02	0.59
Asn	0.9	0.76	1.23
Total	5.49	5.39	7.07
Average	0.91	0.89	1.01

It is interesting to note that Pt (conformational parameter for β-turn) is significantly greater than 1 for Pro, Ser, Tyr and Asn. On comparing the average values of $P\alpha$, $P\beta$ and Pt that is 0.91, 0.89 and 1.01, we observe $Pt > P\beta$, hence our hexapeptide is very likely to adopt β-turn conformation.

Although we have done this exercise using Chou Fasman technique, we expected the hexapeptide to be in β-turn conformation since it is a part of DNA binding loop of Gene V protein and is found to exist as β-turn in native state.

3) Theoretical technique- Conformational angles in hexapeptide were computed using program PCILO on CYBER at TIFR, to predict the secondary structure of hexapeptide theoretically. X-ray coordinates of various amino acids were obtained from literature (103b). The angles ϕ and ψ for each residue, hereafter labelled as

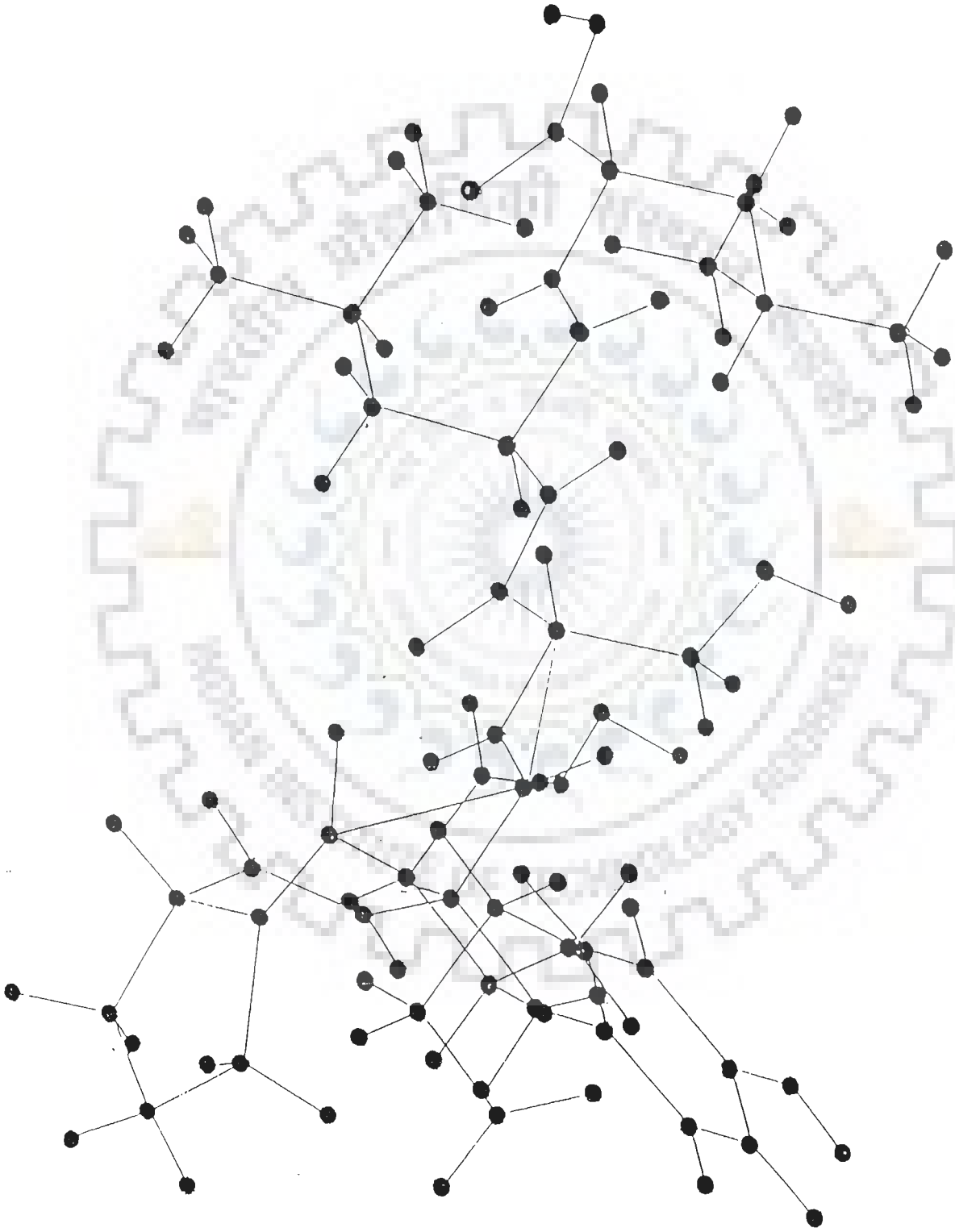


FIG. 3.4: Molecular graphics model of hexapeptide, KPYSLN.

ϕ_1,ψ_1; ϕ_2,ψ_2 and so on for the first, second, etc. peptide residue of hexapeptide, were varied independently one after the other, from 0° to 300°, in steps of 10° and energy was minimised. This procedure was first adopted for lysine and minimum energy was obtained at $\phi_1=-70^\circ$ and $\psi_1=60^\circ$. It is well known that proline has a rigid pyrrolidine ring. Its rotation angle ϕ is fixed at about -60° and allowed values for ψ are -55° and 145°. The relative ease with which it adopts the compact $\psi = -55^\circ$ conformation makes it a desirable candidate for accommodating turns and bends in a polypeptide chain. Therefore the angles for proline were varied in this range only and the values obtained were $\phi_2 = -59^\circ$ and $\psi_2 = -25^\circ$. For other residues, the minimum energy was obtained as in the case of lysine and a particular geometry of the hexapeptide was obtained. This conformation which corresponds to minimum energy, has $\phi_3 = -85^\circ$, $\psi_3 = 0^\circ$, $\phi_4 = -87^\circ$, $\psi_4 = 115^\circ$, $\phi_5 = -60^\circ$, $\psi_5 = 82^\circ$, $\phi_6 = -125^\circ$ and $\psi_6 = 88^\circ$ and the hexapeptide with these angles is shown in Fig.3.4.

C) Spectral assignment of deoxy-penta-adenylate, $d(A)_5$

The bases in $d(A)_5$, shown schematically in Fig.3.5(a) have been numbered from 5' end to 3' end as follows:

d-A_p	A_p	A_p	A_p	A
1	2	3	4	5

Fig.3.5(b) shows the one-dimensional NMR spectrum of 8 mM $d(A)_5$ in D_2O at 297 K. The expanded portion of resonances in the region 7.4-8.3 ppm is shown in Fig.3.5(c). The singlets of H2 and H8 resonances appear in the region of aromatic protons in the range 7.4-8.3 ppm. The five triplets of sugar H1' protons resonate in the range 5.8-6.4 ppm, sugar H2', H2" protons in the range 2.1-2.8 ppm, sugar H3' protons in the range 4.5-5.0 ppm, sugar H4' protons in the range 4.1-4.3 ppm and the sugar H5', H5" protons in the range 3.6-4.1 ppm. The specific assignments to various adenine residues are however made only by two-dimensional NMR experiments. The phase-sensitive (also expansions of magnitude mode spectra) 2D COSY and NOESY spectra are shown in Figs.3.6(a-i) and 3.7(a-h)

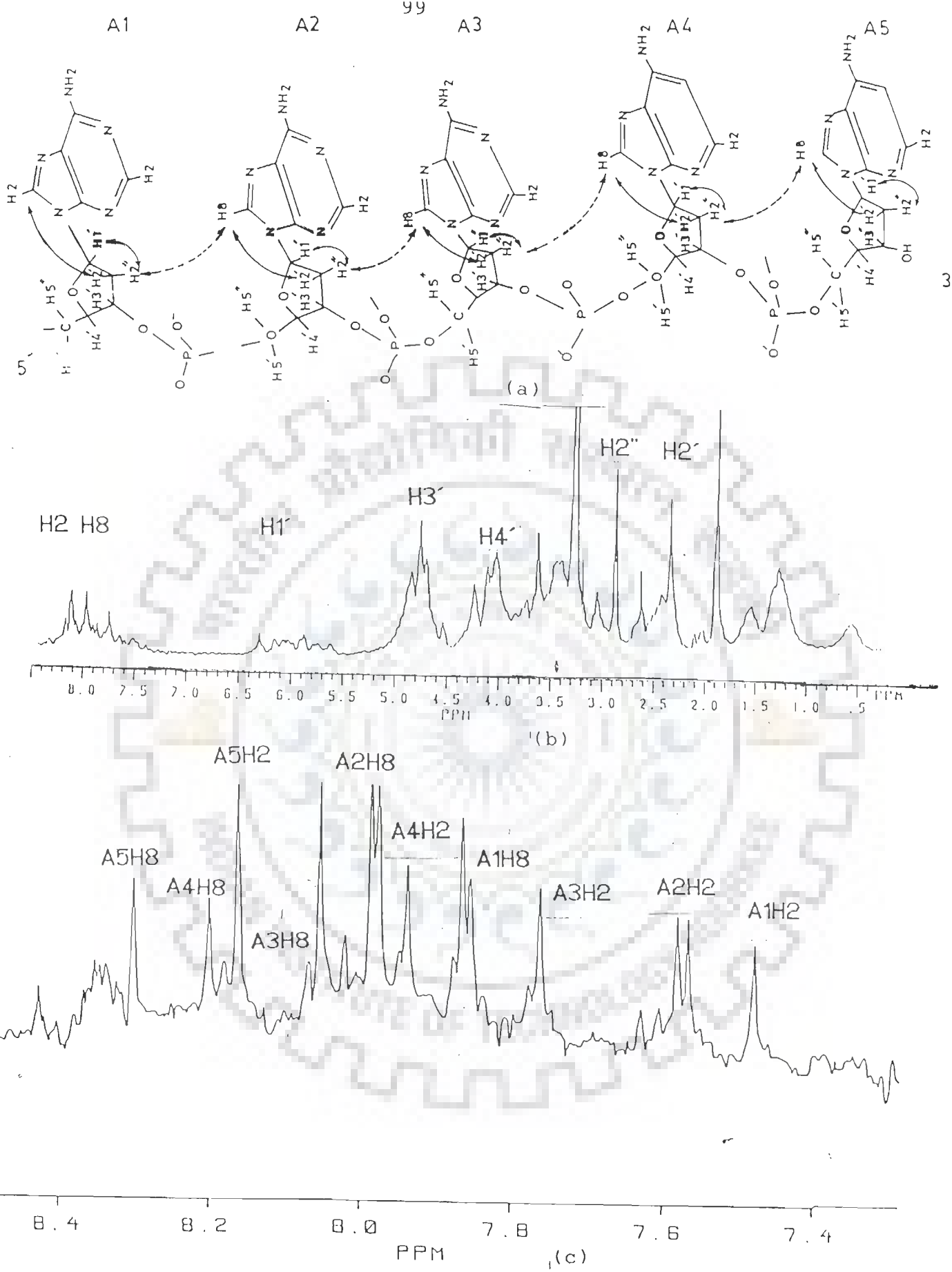


FIG. 3.5(a): Schematic representation of $d(A)_5$. Solid lines (————) show intramolecular connectivities while broken lines (----) show intermolecular connectivities (b) 500 MHz proton NMR spectrum of 8mM $d(A)_5$ in D_2O in phosphate buffer (0.01 M) at 297 K. (c) An expanded portion of (b) showing resonances in the region 7.4-8.3 ppm.

along with expansions of certain portions to highlight specific connectivities. The chemical shifts, δ (in ppm), for various protons are presented in Table 3.2.

The five sets of sugar protons are recognisable in the COSY spectrum. The H1' of each of the five sugars gives cross peak pattern with its corresponding H2' and H2" protons numbered serialy as 1 to 10 in Fig.3.6(b). The H2' and H2" protons show strong geminal couplings (numbered 11-15 in Fig.3.6(c)) with each other. These are further coupled to their corresponding H3' protons (numbered as 16-20 in Fig.3.6(d)) which are coupled to their corresponding H4' protons (Fig.3.6(e)). The H4' protons are coupled to H5', H5" protons (numbered as 26-30 in Fig.3.6(f)). The five sets of sugar protons identified in COSY spectra are assigned to specific adenine nucleotides of $d(A)_5$ by following the strategies available in literature (48,63,78,80,125), for sequential assignment in right-handed B-DNA with sugar in C3' endo/C2' endo/O1' endo conformation and glycosidic angle in anti domain.

The distances of base H8 proton from H1', H2' and H2" (150) protons are less than 4.0 Å so that the following internucleotide connectivities are expected:

$$(AH8)_n \quad \text{———} \quad (AH1')_{n-1}$$
$$(AH8)_n \quad \text{———} \quad (AH2")_{n-1}$$

Fig.3.7(b,c,f) show the intra residual NOE's of H1'-H2', H1'-H2" (cross peaks no. 1-10); H2'-H2" (cross peaks no. 11-15); H3'-H4' (cross peaks no. 21-25) and H4'-H5', H5" protons (cross peak no. 26-30). The internucleotide NOE connectivities are seen in NOESY spectra (Fig.3.7(g)) as cross peaks no. 31-43. Each H8 proton is thus expected to give two intranucleotide cross peaks with its corresponding H2', H2" proton and one internucleotide cross peak with H2" of nucleotide preceeding it except for the terminal A1 residue which does not have another residue preceeding to it in the sequence. Thus all H8's resonance are assigned sequentially as

Table 3.2: Chemical shift+ values of different protons of 8 mM $d(A)_5$ in D_2O at 297 K.

Proton	A1	A2	A3	A4	A5
H1'	5.94	5.81	6.13	6.23	6.37
H2'	2.15	2.37	2.45	2.51	2.51
H2"	2.41	2.51	2.63	2.76	2.75
H3'	4.75	4.90	4.95	4.58	4.58
H4'	4.17	4.27	4.31	4.15	4.15
H5'	3.68	4.05	4.07	3.73	3.73
H8	7.82	7.95	8.04	8.18	8.25
H2	7.45	7.55	7.74	7.92	8.14

+ Chemical shift values are in ppm with respect to sodium 2,2 Dimethyl-2-silapentane-5 suphonate (DSS).

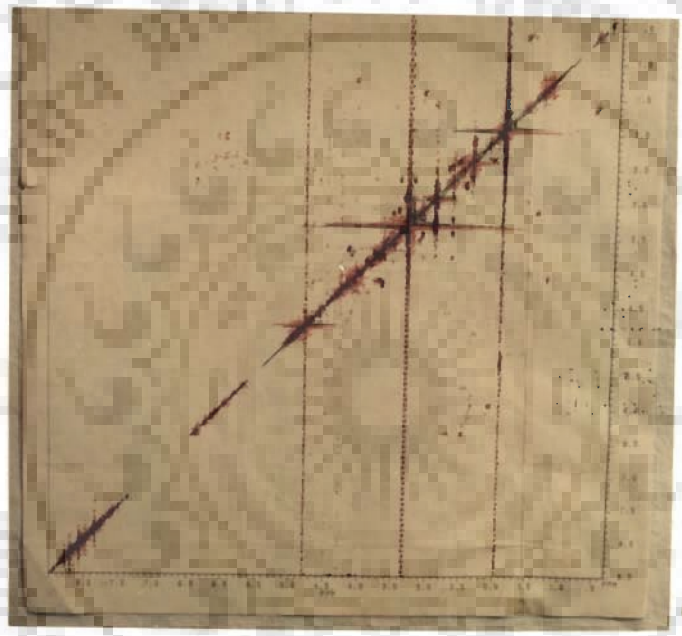


FIG. 3.6(a):Phase-sensitive COSY spectrum of 8 mM $d(A)_5$ in D_2O in phosphate buffer (0.01 M) at 297 K.

FIG.3.6(b):A portion of phase-sensitive COSY spectrum of 8 mM $d(A)_5$ in D_2O expanded to show specific connectivities: (1) A5H1'-A5H2' (2) A5H1'-A5H2" (3) A4H1'-A4H2' (4) A4H1'-A4H2" (5) A3H1'-A3H2' (6) A3H1'-A3H2" (7) A1H1'-A1H2' (8) A1H1'-A1H2" (9) A2H1'-A2H2' (10) A2H1'-A2H2"

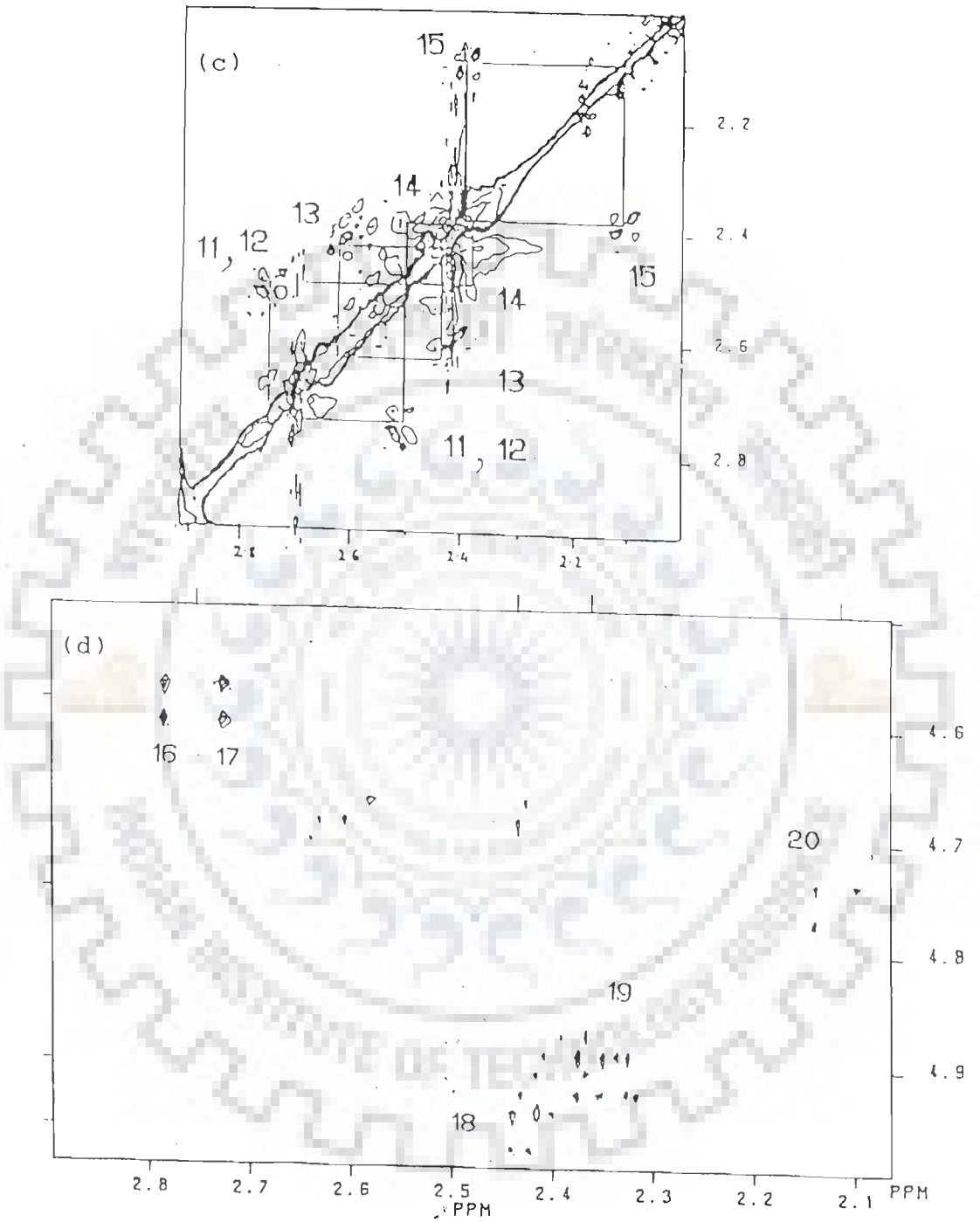


FIG. 3.6(c): A portion of phase-sensitive COSY spectrum of 8 mM $d(A)_5$ in D_2O expanded to show specific connectivities: (11) A5H2'-A5H2" (12) A4H2' -A4H2" (13) A3H2'-A3H2" (14) A2H2'-A2H2" (15) A1H2'- A1H2" (d): Expansion shows (16) A5H2"-A5H3' (17) A4H2" -A4H3' (18) A3H2'-A3H3' (19) A2H2'-A2H3' (20) A1H2'-A1H3'

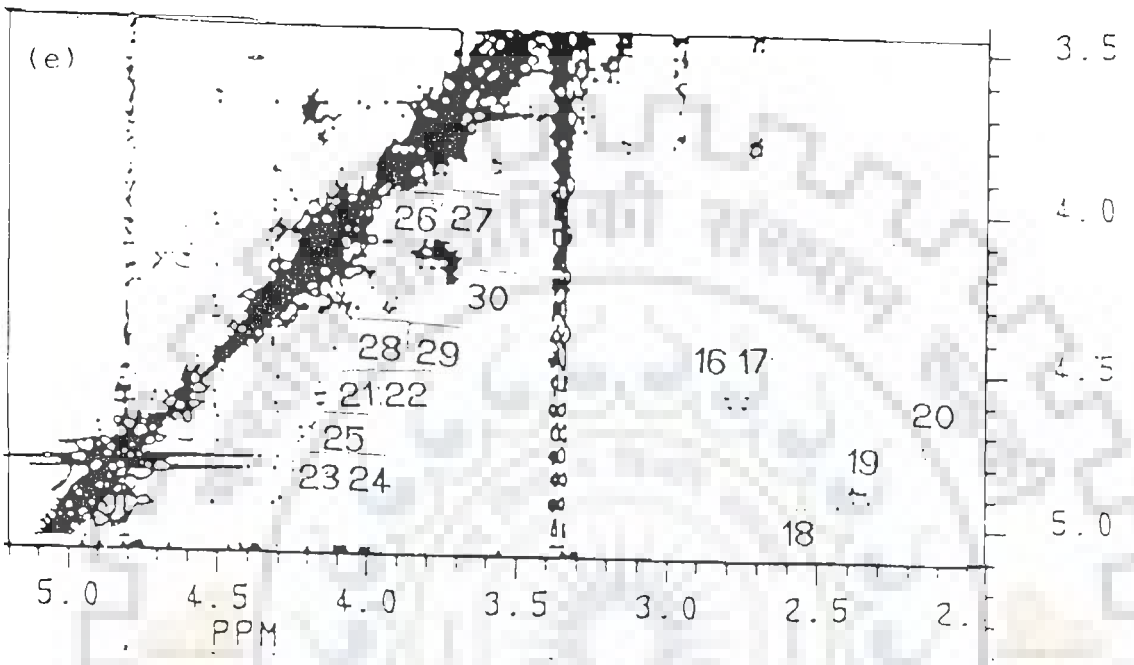


FIG. 3.6(e): A portion of phase-sensitive COSY spectrum of 8 mM $d(A)_5$ in D_2O expanded to show specific connectivities: (21) A5H3'-A5H4' (22) A4H3'-A4H4' (23) A3H3'-A3H4' (24) A2H3'-A2H4' (25) A1H3'-A1H4' (26) A5H4'-A5H5' (27) A4H4'-A4H5' (28) A3H4'-A3H5' (29) A2H4'-A2H5' (30) A1H4'-A1H5' (16) A5H2"-A5H3' (17) A4H2"-A4H3' (18) A3H2'-A3H3' (19) A2H2'-A2H3' (20) A1H2'-A1H3'

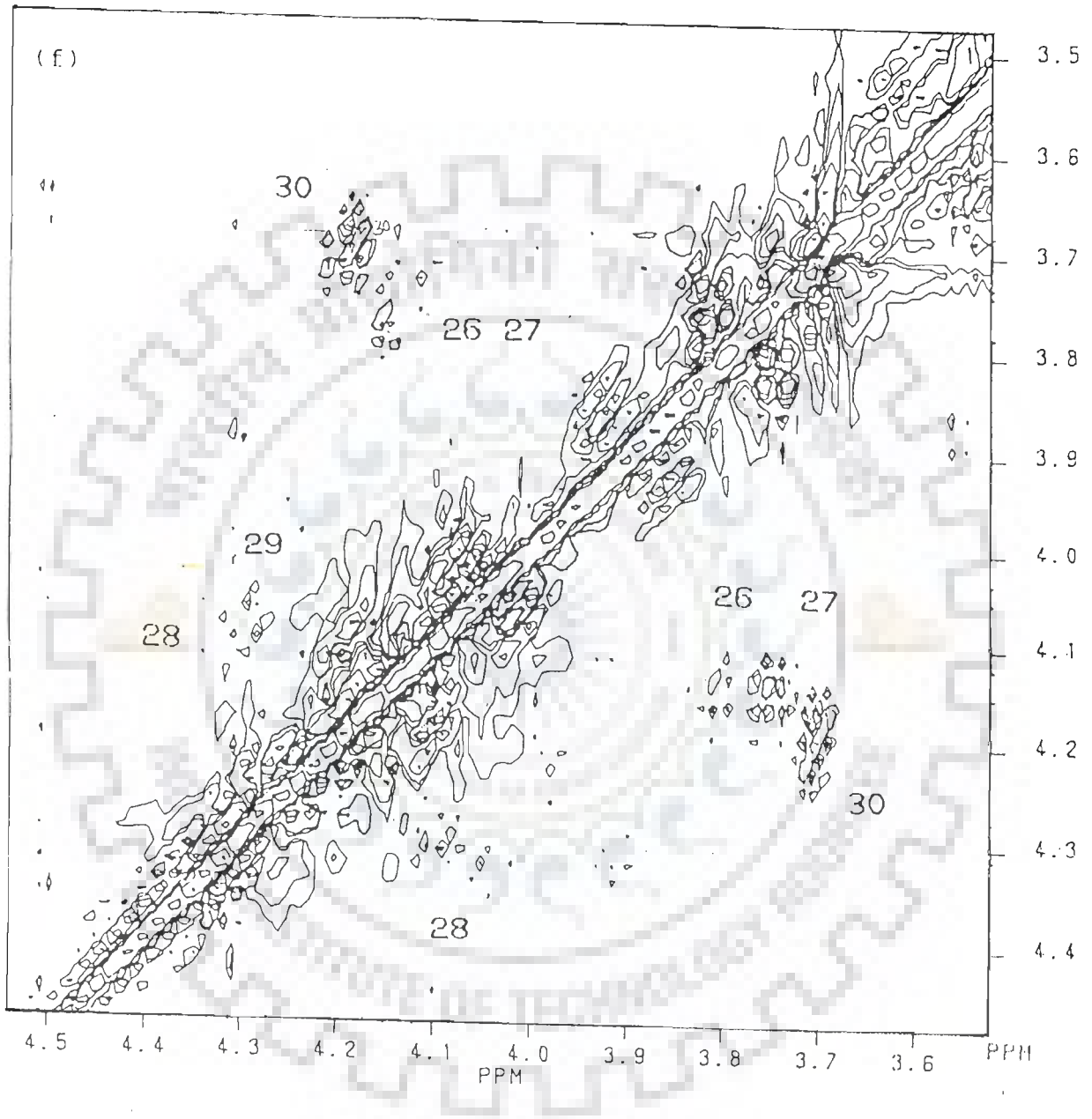


FIG. 3.6(f):A portion of phase-sensitive COSY spectrum of 8 mM $d(A)_5$ in D_2O expanded to show specific connectivities: (26) A5H4'-A5H5' (27) A4H4'-A4H5' (28) A3H4'-A3H5' (29) A2H4'-A2H5' (30) A1H4'-A1H5'

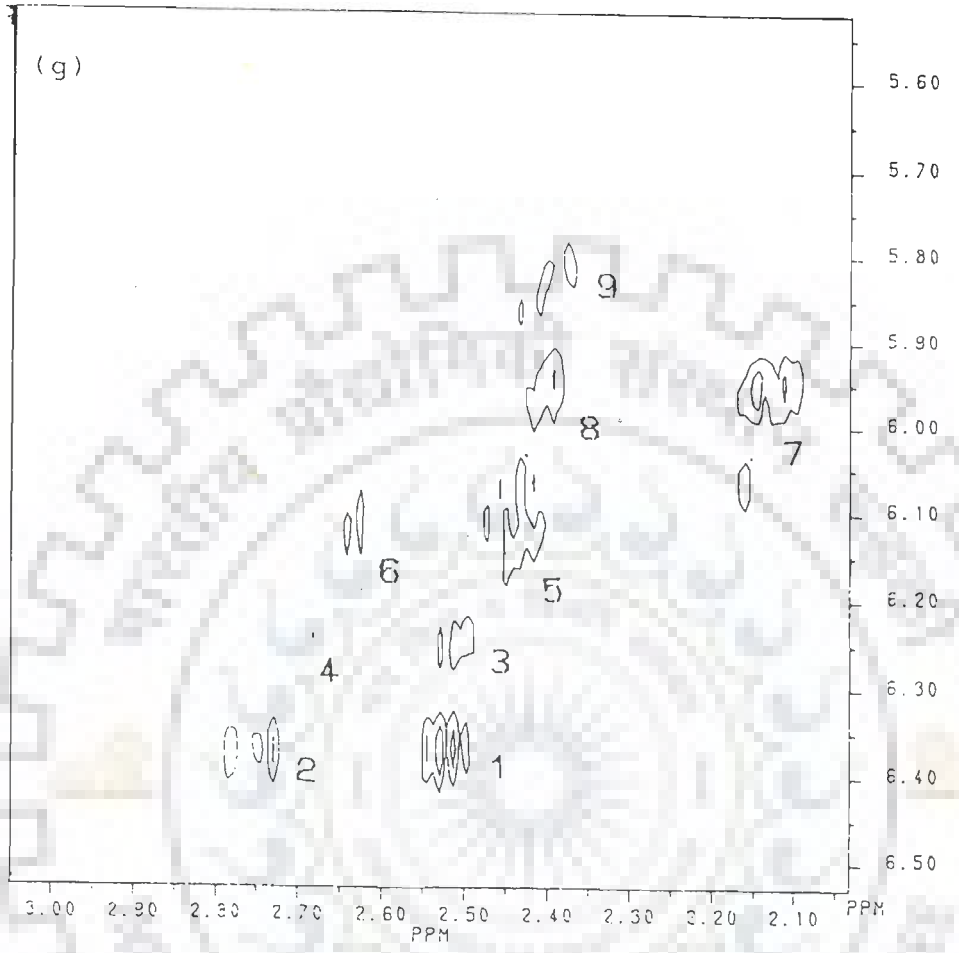


FIG. 3.6(g):A portion of magnitude mode COSY spectrum of 8 mM $d(A)_5$ in D_2O expanded to show specific connectivities: (1) A5H1'-A5H2' (2) A5H1'-A5H2" (3) A4H1'-A4H2' (4) A4H1'-A4H2" (5) A3H1'-A3H2' (6) A3H1'-A3H2" (7) A1H1'-A1H2' (8) A1H1'-A1H2" (9) A2H1'-A2H2'

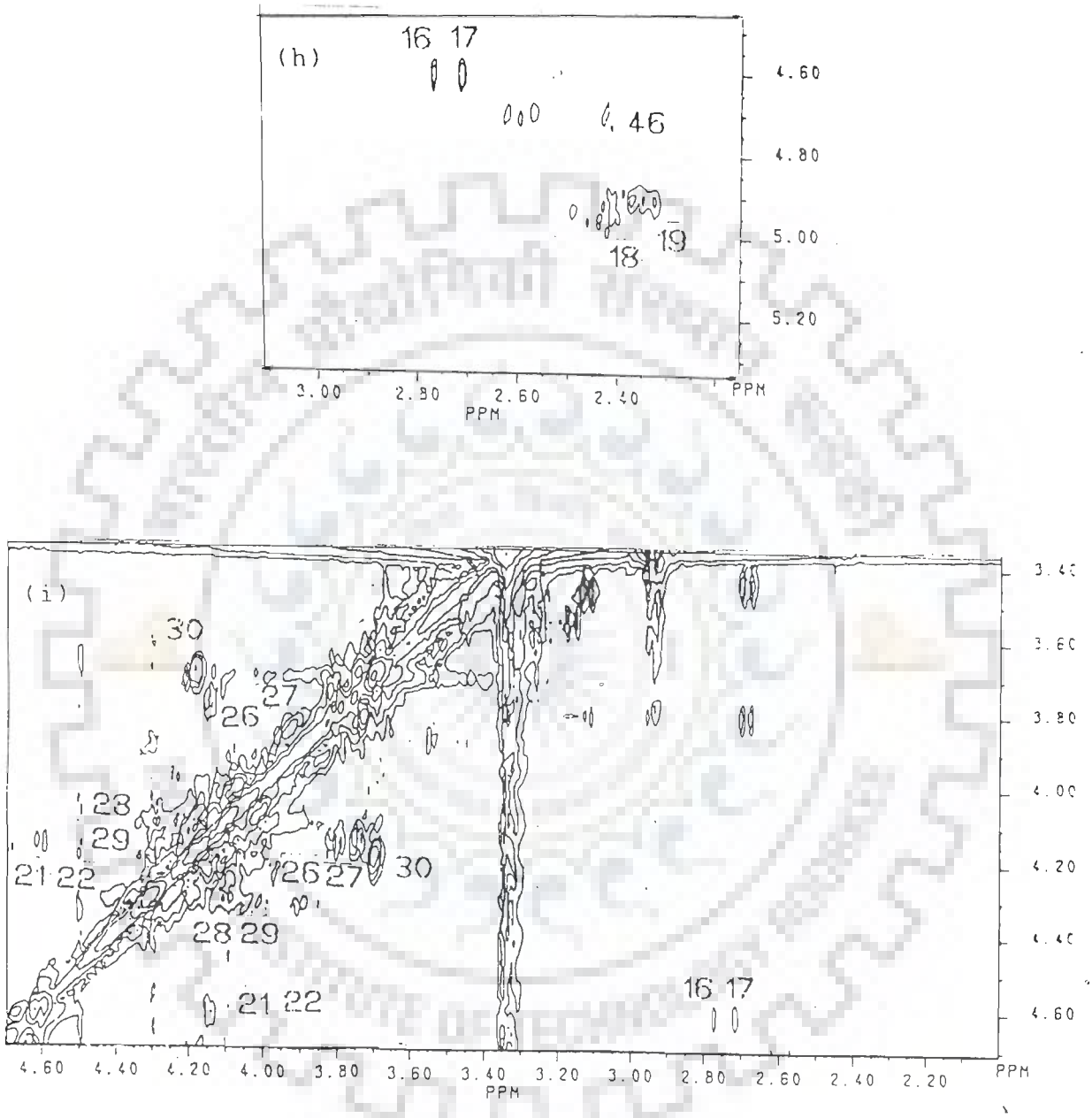


FIG. 3.6(h): A portion of magnitude mode COSY spectrum of 8 mM $d(A)_5$ in D_2O expanded to show specific connectivities: (16) A5H2"-A5H3' (17) A4H2"-A4H3' (18) A3H2'-A3H3' (19) A2H2'- A2H3' (46) A1H2"- A1H3' (i): Expansion shows: (16) A5H2"- A5H3' (17) A4H2"- A4H3' (26)A5H4'- A5H5' (27) A4H4'-A4H5' (28) A3H4'-A3H5' (29) A2H4'-A2H5'(30) A1H4' - A1H5' (21) A5H3'- A5H4' (22)A4H1' -A4H4'

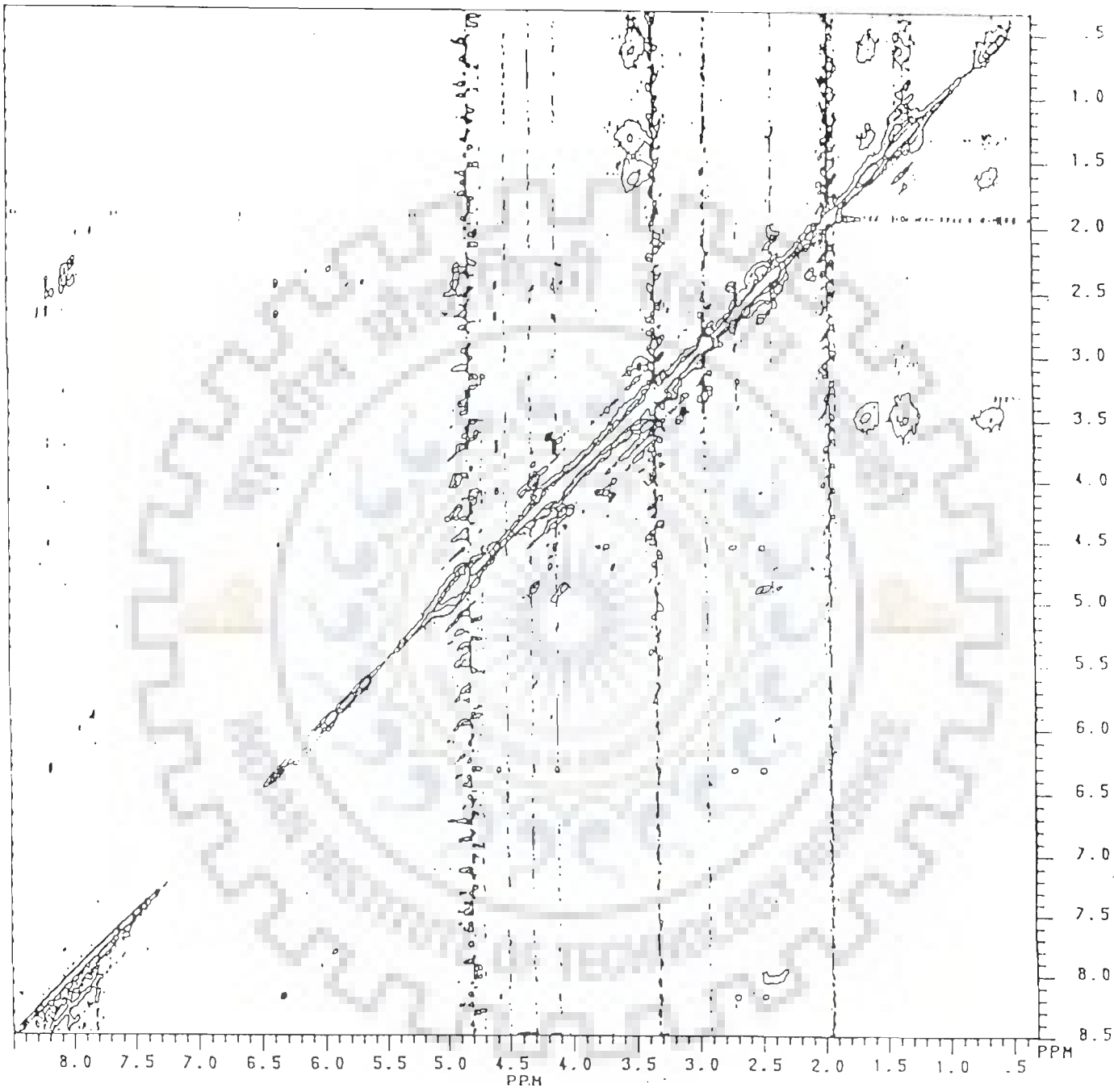


FIG. 3.7(a): Phase-sensitive NOESY spectrum of 8mM $d(A)_5$ recorded in D_2O solution at 297 K.

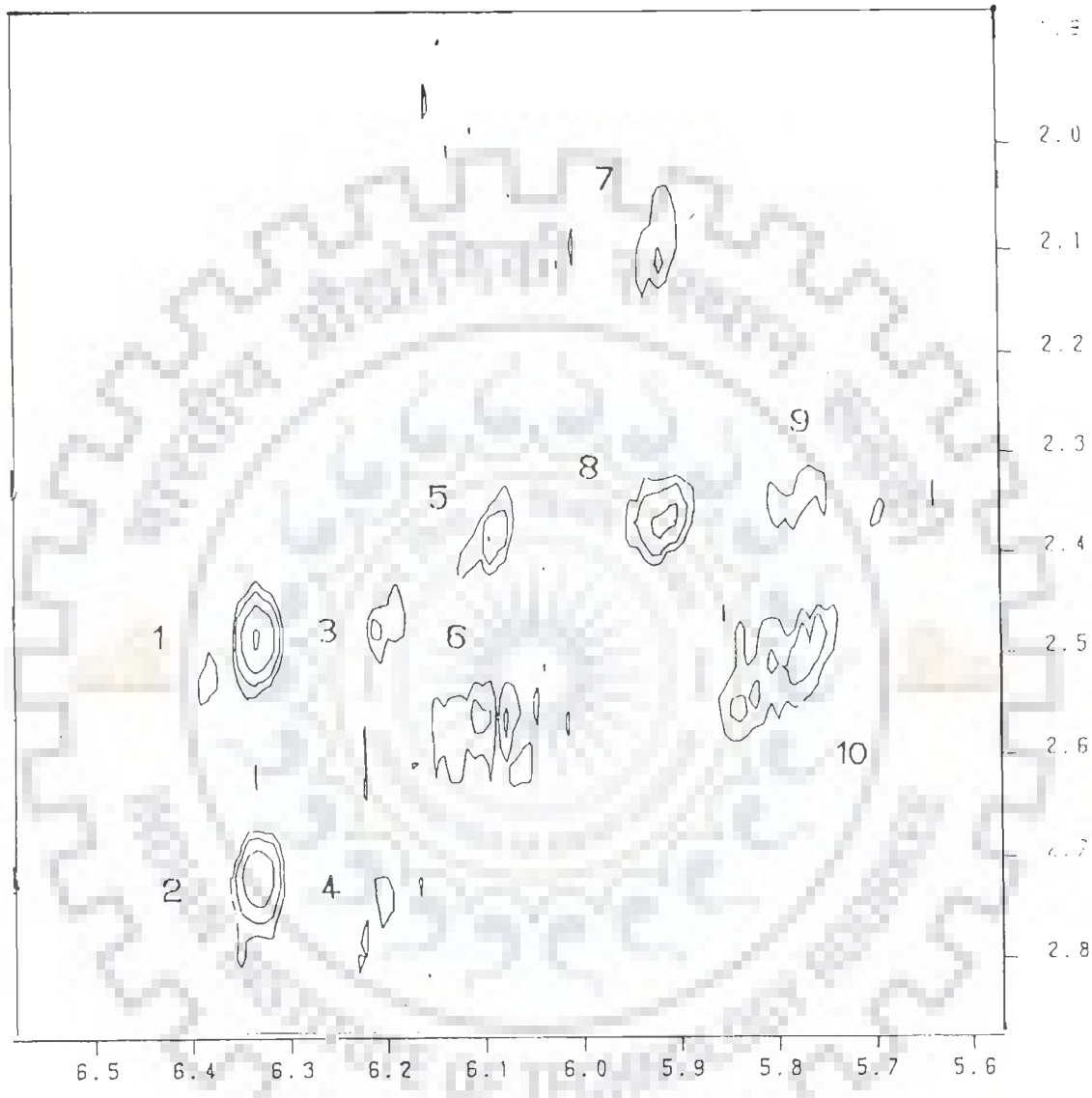


FIG. 3.7(b): A portion of phase-sensitive NOESY spectrum of 8 mM $d(A)_5$ in D_2O expanded to show intramolecular proton connectivities: (1) A5H1'-A5H2' (2) A5H1'-A5H2" (3) A4H1'-A4H2' (4) A4H1'-A4H2" (5) A3H1'-A3H2' (6) A3H1'-A3H2" (7) A1H1'-A1H2' (8) A1H1'-A1H2" (9) A2H1'-A2H2' (10) A2H1'-A2H2"

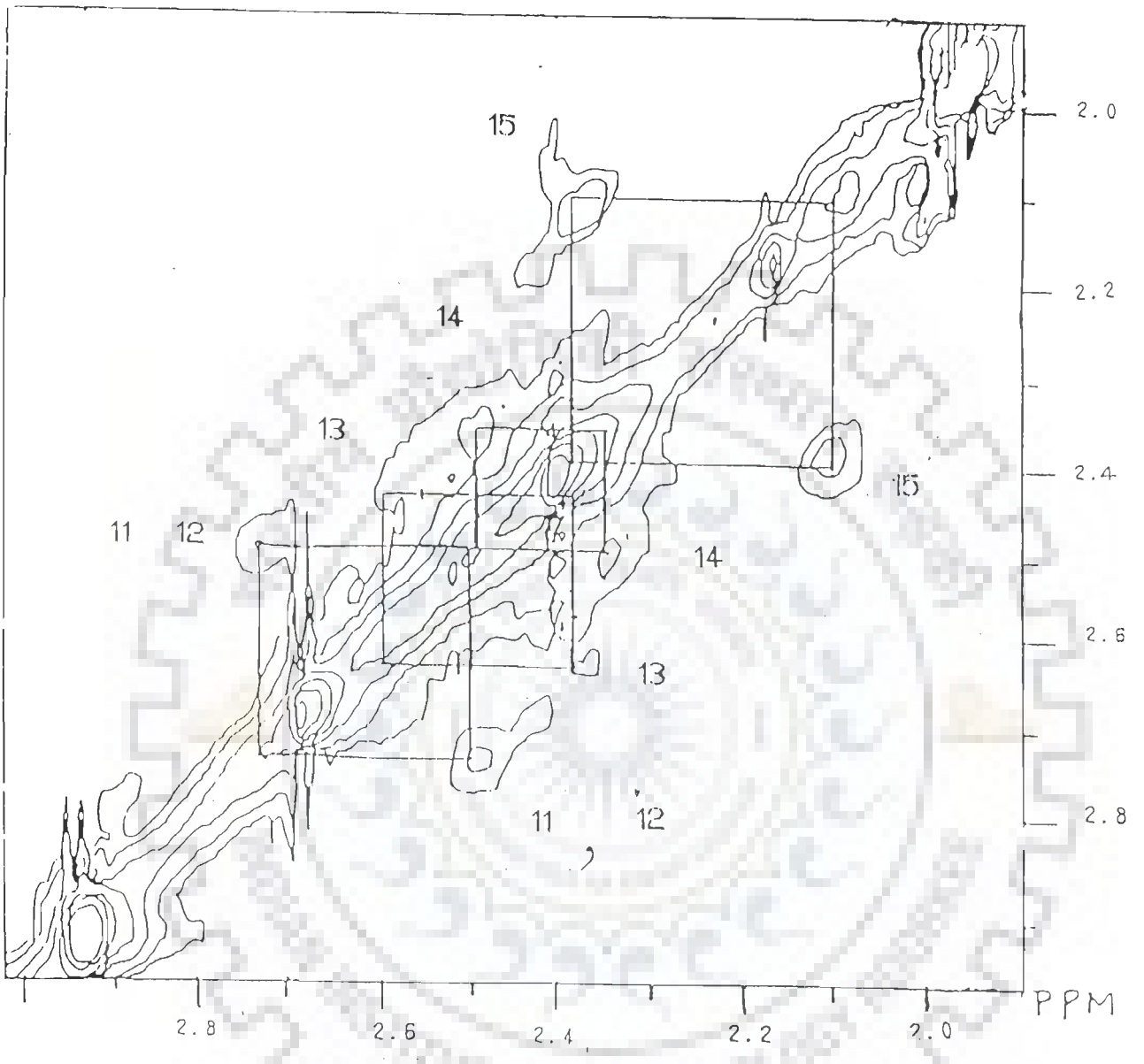


FIG. 3.7(c): A portion of phase-sensitive NOESY spectrum of 8 mM $d(A)_5$ in D_2O expanded to show intramolecular proton connectivities:(11) A5H2'-A5H2" (12) A4H2'-A4H2" (13) A3H2'-A3H2" (14) A2H2'-A2H2" (15) A1H2'-A1H2"

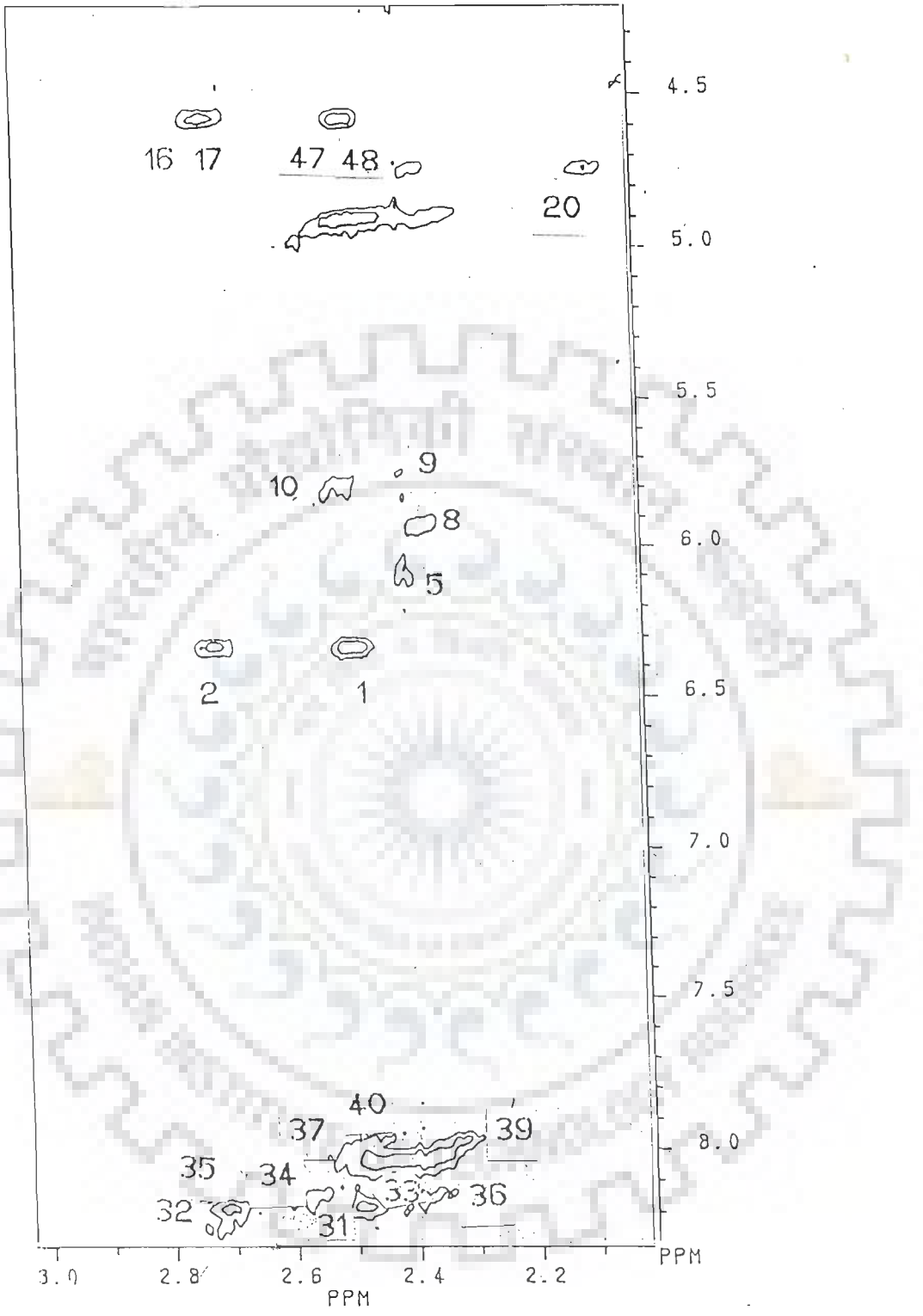


FIG. 3.7(d):A portion of phase-sensitive NOESY spectrum of 8 mM $d(A)_5$ in D_2O expanded to show intramolecular and intermolecular proton connectivities: (1) A5H1'-A5H2' (2) A5H1'-A5H2" (5) A3H1'-A3H2' (8) A1H1'-A1H2" (9) A2H1'-A2H2' (10) A2H1'-A2H2" (16) A5H2"-A5H3' (17) A4H2"-A4H3' (47) A5H2' - A5H3' (48) A4H2'-A4H3' (31) A5H8-A5H2' (32) A5H8-A5H2" (33) A4H8-A4H2' (34) A4H8-A3H2" (35) A3H8-A4H2" (36) A3H8-A3H2' (37) A3H8-A2H2" (39)A2H8 -A2H2" (40) A2H8-A1H2" (20) A1H2'-A1H3'

FIG. 3.7(e): A portion of phase-sensitive NOESY spectrum of 8 mM $d(A)_5$ in D_2O expanded to show intermolecular and intramolecular proton connectivities: (44) A5H8 - A4H1' (45) A3H8 - A2H1' (57) A4H8-A4H3'

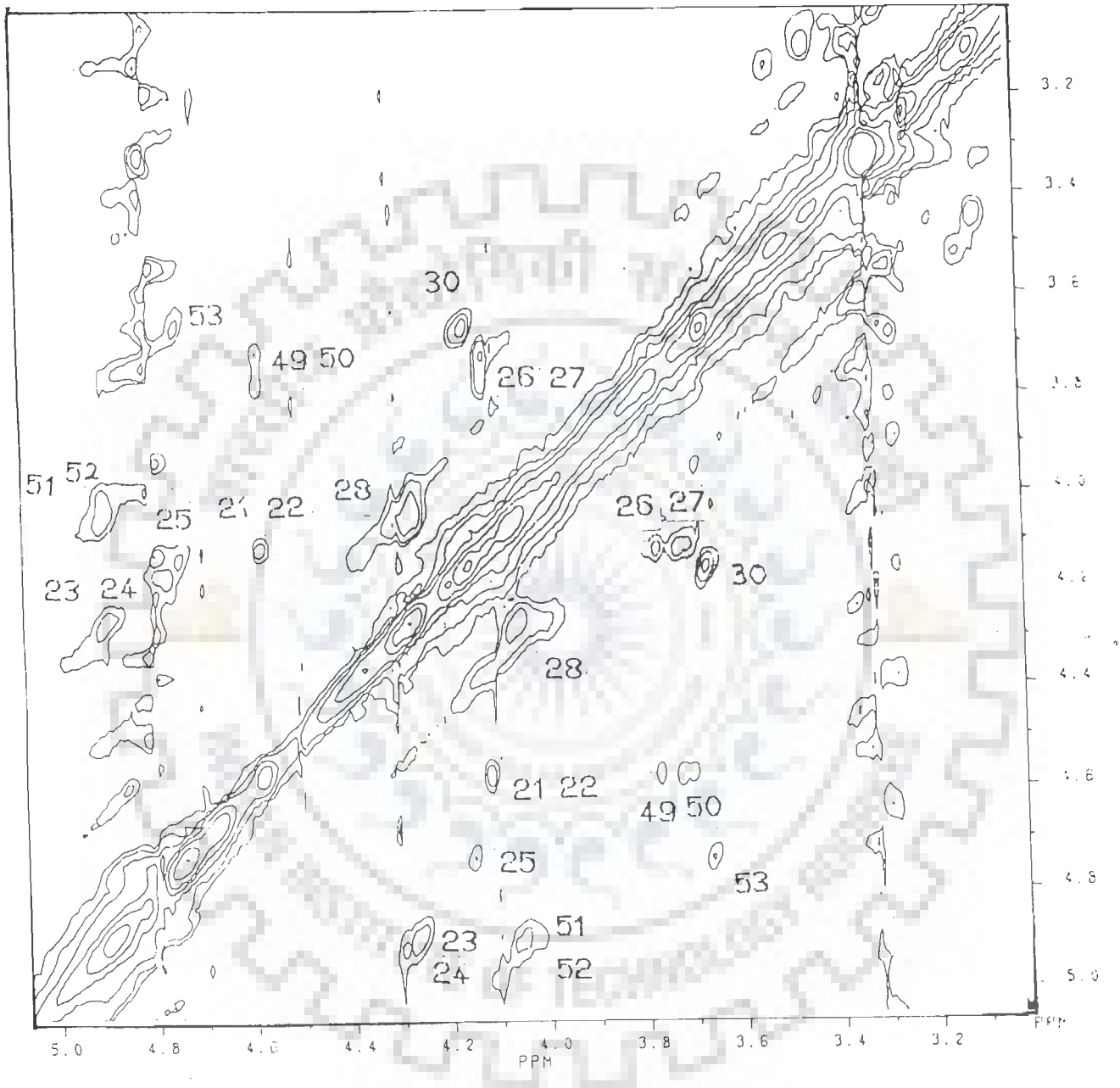


FIG.3.7(f):A portion of phase-sensitive NOESY spectrum of 8 mM $d(A)_5$ in D_2O expanded to show intramolecular proton connectivities:(21)A5H3'-A5H4' (22) A4H3'-A4H4' (23) A3H3'-A3H4' (24) A2H3'-A2H4' (25) A1H3'-A1H4' (26) A5H4'-A5H5' (27) A4H4'-A4H5' (28) A3H4'-A3H5' (29) A2H4'-A2H5' (30) A1H4' -A1H5' (49) A4H3'-A4H5' (50) A5H3'-A5H5' (53) A1H3'-A1H5'

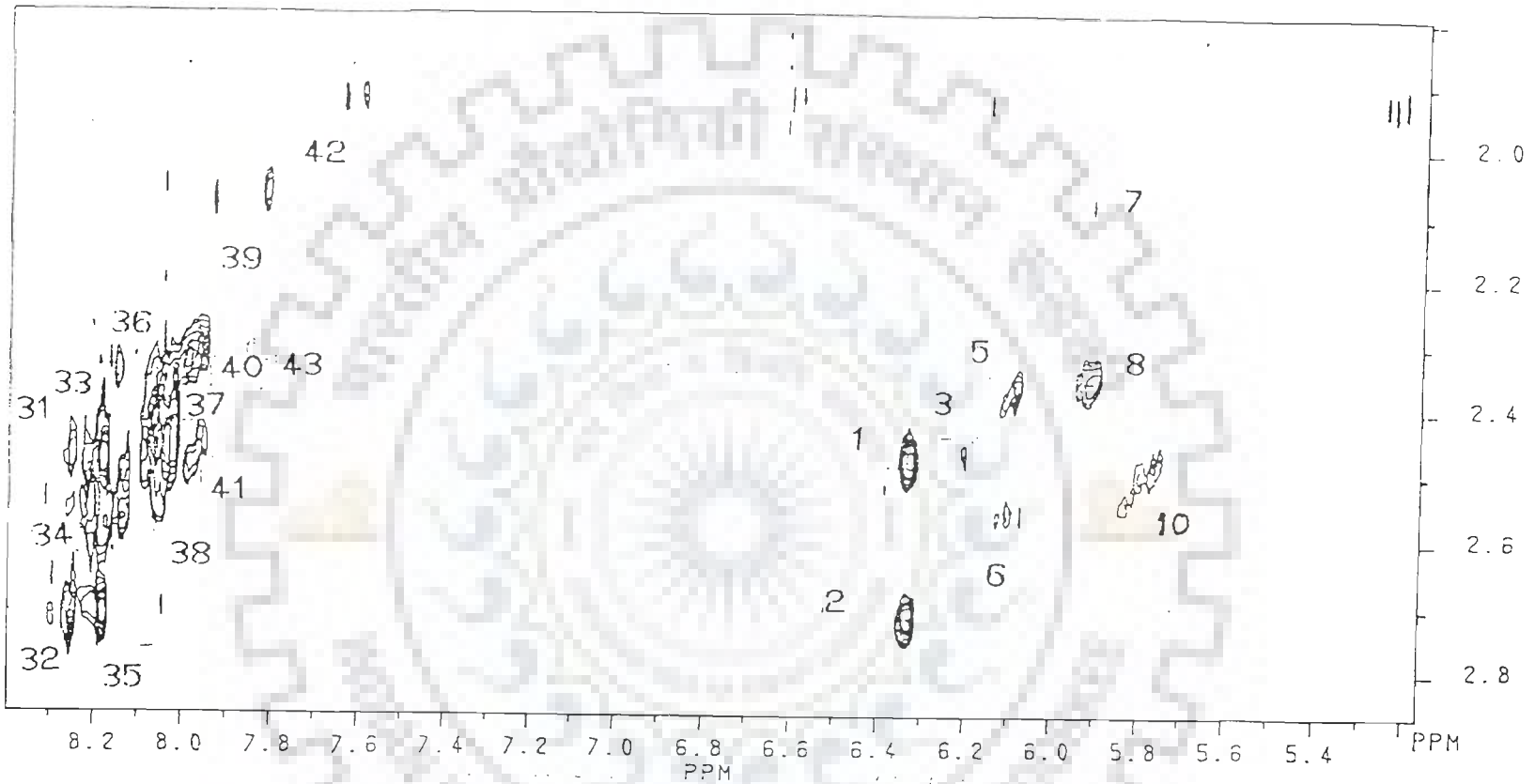


FIG. 3.7(g): A portion of phase-sensitive NOESY spectrum of 8 mM $d(A)_5$ in D_2O expanded to show intramolecular and intermolecular proton connectivities: (1) A5H1'-A5H2' (2) A5H1'-A5H2" (3) A4H1'- A4H2' (4) A4H1'- A4H2" (5) A3H1'-A3H2' (6) A3H1'-A3H2" (7) A1H1'-A1H2' (8) A1H1'-A1H2" (9) A2H1'-A2H2' (10) A2H1'-A2H2" (31) A5H8-A5H2' (32)A5H8-A5H2" (33) A4H8- A4H2' (34) A4H8 -A3H2" (35) A4H8-A4H2" (36) A3H8-A3H2' (37) A3H8-A2H2" (38)A3H8-A3H2" (39) A2H8-A2H2' (40) A2H8-A1H2" (41) A2H8-A2H2" (42) A1H8-A1H2' (43) A1H8-A1H2"

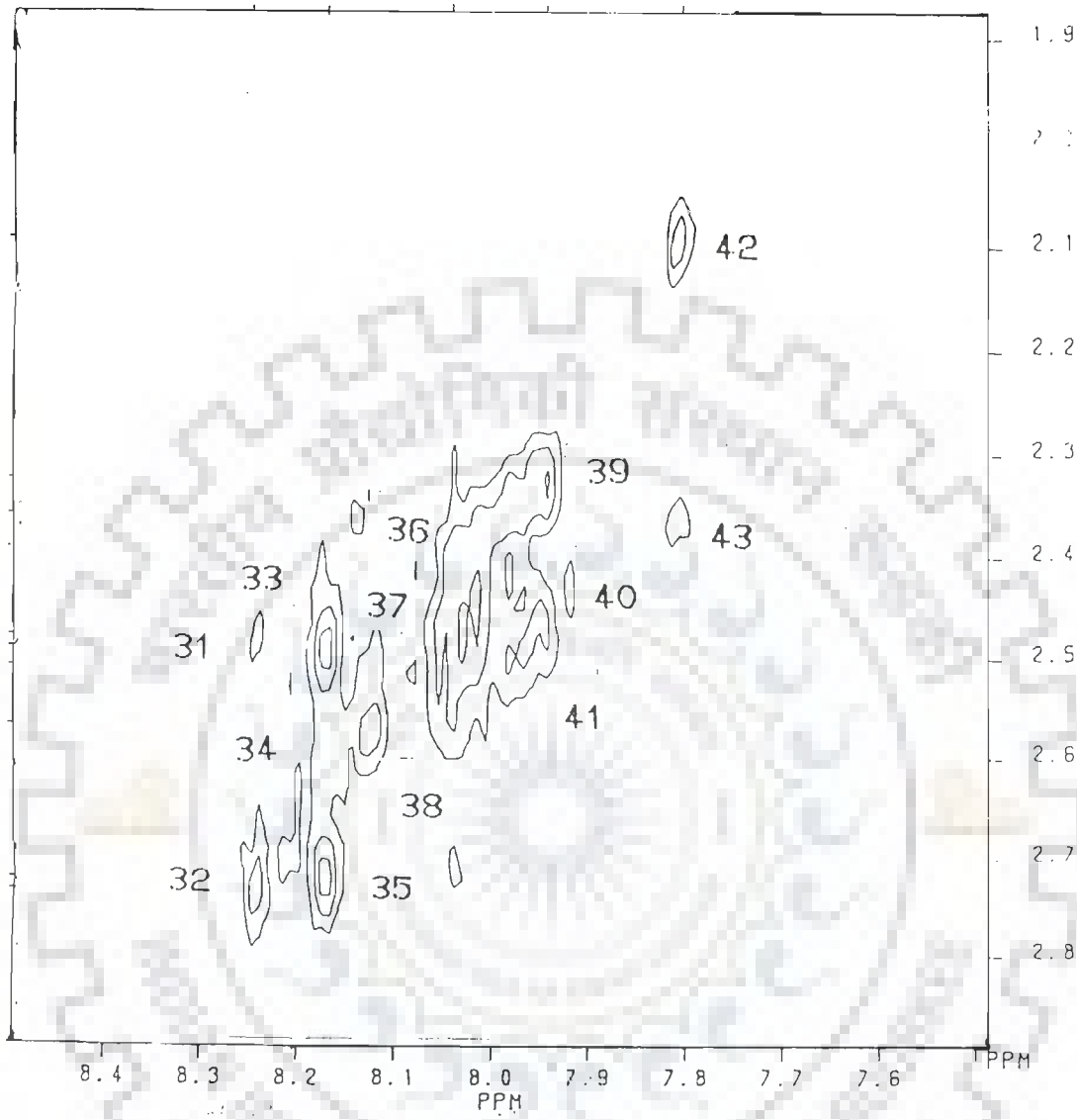


FIG. 3.7(h): A portion of phase-sensitive NOESY spectrum of 8 mM $d(A)_5$ in D_2O expanded to show intramolecular and intermolecular proton connectivities: (31) A5H8-A5H2' (32) A5H8-A5H2" (33) A4H8-A4H2' (35) A4H8-A4H2" (36) A3H8-A3H2' (37) A3H8-A2H2" (42) A1H8-A1H2' (43) A1H8-A1H2" (34) A4H8 -A3H2" (38)A3H8-A3H2" (39) A2H8-A2H2' (40) A2H8-A1H2" (41) A2H8-A2H2"

shown in Fig.3.7(h). On the basis of relative stacking of each nucleotide in the sequence, it has been shown that the chemical shift of AH2 protons follow the pattern of H8 resonances (85). Therefore the resonance at 7.45, 7.55, 7.74, 7.92, 8.14 ppm are assigned to A1H2, A2H2, A3H2, A4H2 and A5H2, respectively. The unambiguous assignment was thus made.

D) Conformation of $d(A)_5$ in unbound form

Conformation of deoxyribose sugar

The two-dimensional spectra have been used to obtain conformation of deoxyribose sugar and glycosidic bond rotation. We have estimated sugar geometry from the analysis of cross peak patterns of H1'-H2', H1'-H2", H2'-H3', H2"-H3', spin-spin coupled connectivities from phase-sensitive COSY spectra supported by J values measured as distance between + and - contours due to antiphase components (wherever possible).

The standard values of proton vicinal coupling constants for various pairs of sugar protons for different pseudorotation values are taken from literature (139,150). Using these we have generated the cross peak patterns expected for different sets of J-coupled protons in a phase-sensitive COSY spectrum and compared them with our experimentally observed patterns as described below.

In the simplest possible case of an AX spin (two spin) system, the phase-sensitive COSY spectra shows cross peaks between the protons that are J-coupled. These are in pure absorption phase while the diagonal peaks are in dispersive mode. The pure absorption mode cross peaks arising from transferred magnetisation shows well resolved multiplet fine structure along both the frequency dimensions. The coupling produces a square array of two positive (+) and two negative (-) peaks such that cross-sections along either frequency axis show an antiphase splitting (+ -) of magnitude J_{AX} shown schematically in Fig.3.8(a). A typical phase-sensitive spectra of AX spin system is shown in Fig.3.8(b)

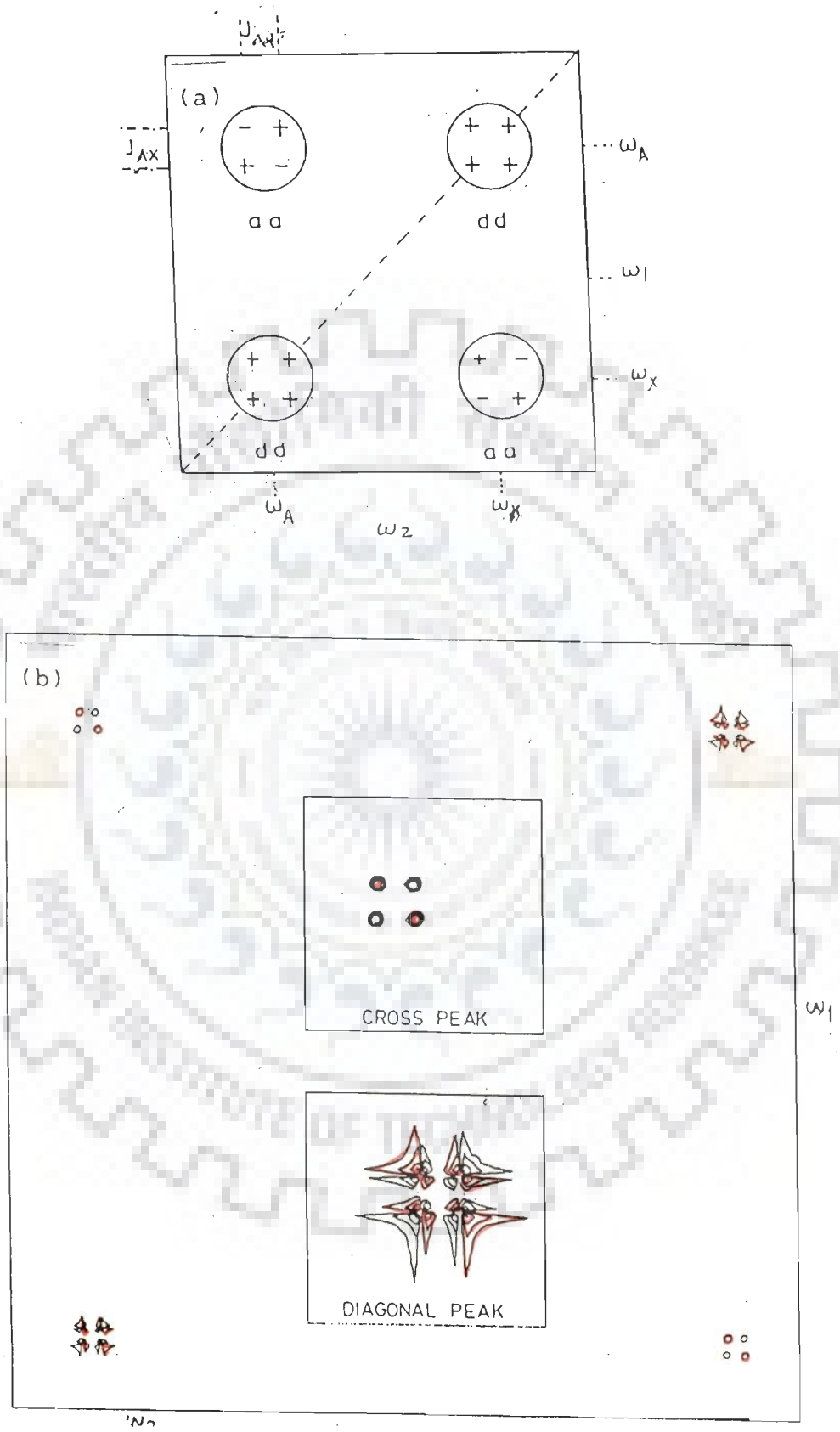


FIG. 3.8(a): Schematic COSY spectrum of a two spin system, AX. ω_A and ω_X are the resonance frequencies of spins A and X. J_{AX} is the active coupling constant. The cross peaks have antiphase character and absorptive character (aa) along both frequency axis. The diagonal peaks have in phase components and dispersive character (dd) along both axis. (b)A typical phase-sensitive spectra of AX spin system (43). Negative peaks are shaded with red colour.

wherein the positive contours are black and negative ones are red (43). Separation between black (+) and red (-) contours of cross peak shown in the inset is a measure of coupling constant J_{AX}. It may be noted that the diagonal peaks are dispersive (inset of Fig.3.8(b)) and overlap of wider dispersion lines leads to characteristic `angel' shapes.

In a three spin system, e.g. AMX, with non-vanishing spin-spin coupling constants J_{AM}, J_{AX} and J_{MX} (schematically shown in Fig.3.9(a)) the AX pattern is further split into identical arrays. The coupling between the protons connected by the cross peaks are active couplings while all others between these protons and any other proton are passive couplings. For example, the AX cross peak at $\omega_1 = \delta_X$ and $\omega_2 = \delta_A$ manifests along ω_2 the antiphase (+ -) splitting J_{AX} and inphase (+ +) splitting J_{AM}. Along ω_1 axis, it manifests antiphase splitting J_{AX} and inphase splitting J_{MX}. A typical phase-sensitive COSY (43) for an AMX system with positive and negative contours in red and black is shown in Fig.3.9(b). A "prototype AX pattern" is clearly evident in the expansion $\bar{B}$ as each of the four quarters of the multiplet (dividing parallel with ω_2 and ω_1). Therefore in the cross peak pattern $\bar{B}$, the smaller coupling is the active coupling J_{AX} (along both ω_1 and ω_2 axis) and the larger ones along ω_2 and ω_1 axis are passive couplings J_{AM} and J_{MX}, respectively. It is readily apparent from the Fig.3.9(a and b) that the appearance of the crosspeaks including the order in which positive and negative components are arranged depends critically on the relative size of the individual coupling constant.

The ^1H-^1H coupling constants of interest in oligonucleotides are J(H1'-H2'), J(H1'-H2"), J(H2'-H3'), J(H2"-H3'), J(H3'-H4') and J(H2'-H2"). Among these the five 3-bond coupling constants depend critically on sugar geometry (66,78) and thus their estimation from experimental spectra provides an effective tool for fixing the sugar geometries in oligonucleotides. The coupling constants

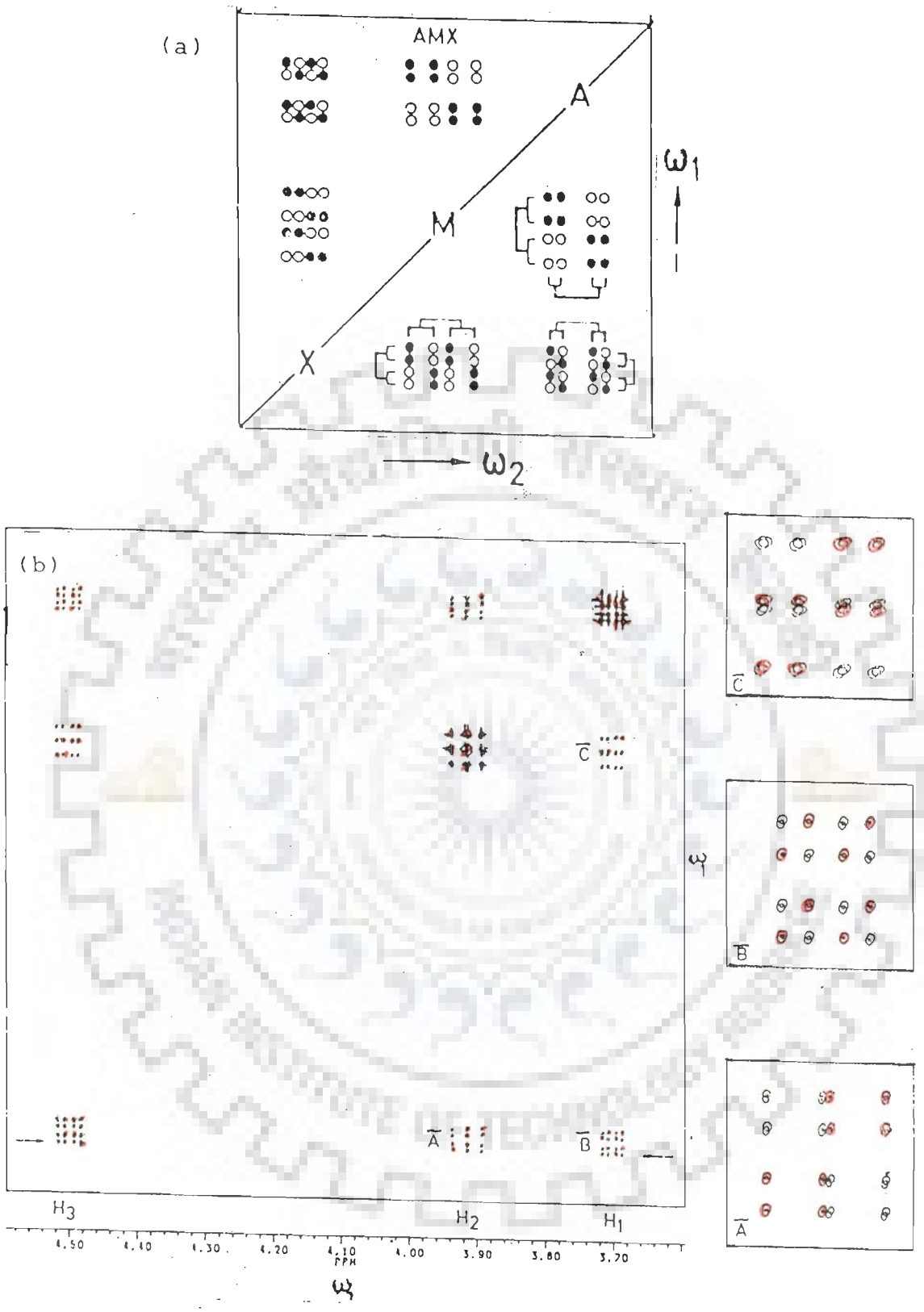


FIG. 3.9(a): Schematic COSY spectrum of a three spin system, AMX. The letters on the diagonal indicate the chemical shifts for the different group of protons. Filled circles represent positive signals, open circles represent negative signals and the area of the circles represents the peak intensity. In the lower right, active and passive couplings are indicated with thick and thin lines, respectively (150).(b) A typical phase-sensitive COSY spectrum for an AMX system (43). Negative peaks are shaded with red colour.

are calculated as a function of the sugar geometry using the Karplus type relationship (77)

$$J = 10.2 \cos^2 \theta - 0.8 \cos \theta$$

where θ is the relevant torsional angle. These values are given in Table 3.3. If the flip angle ϕ of the second pulse (please see Fig. 2.2(a) in Chapter II) is $\pi/2$ then the multiplet structure is simply calculated by combining the + and - along the ω_1 axis with those along the ω_2 axis. These are illustrated in Fig.3.10(a) for H1'-H2' cross peak in O1' endo sugar conformation. The fine structure along the ω_1 axis arises from the H1'-H2', H2'-H3' and H2'-H2" coupling constants while that along the ω_2 axis arises from H1'-H2' and H1'-H2" coupling constants. Along ω_2 and ω_1 axis, the (+ +) and (- -) pattern repeats itself as the active coupling J(H1'-H2')= 8.5 Hz is larger than its passive coupling i.e. J(H1'-H2")=6.7 Hz and J(H2'-H3')=8.4 Hz, respectively. On the other hand, in H1'-H2" cross peak pattern, the fine structure along the ω_2 axis arises from active H1'-H2" coupling ~ 6.7 Hz which is lesser than passive coupling i.e. J(H1'-H2') ~ 8.5 Hz. We have used this procedure to calculate the cross peak pattern for all J-coupled protons in different standard sugar geometries, the results of which are shown in Fig.3.10(a-c). The cross peak patterns generated as a result of calculation by procedure used by Hosur et al.(77) are shown in Fig.3.11 for H1'-H2' and H1'-H2" cross peaks for flip angle $\phi=\pi/2$ of the second pulse for three different sugar geometries. For the C3' endo geometry the H1'-H2" cross peak looks more elongated along the ω_1 axis and is compressed along the ω_2 axis. The coupling constants can be read directly from the scale along axis.

Fig.3.6(a) shows the phase-sensitive COSY spectrum of 8 mM $d(A)_5$ in D_2O solution at 297 K, portions of which showing specific connectivities are given in Fig.3.6(b-f). It is observed that the distance between outer peaks of H1' in the COSY spectrum Fig.3.6(b) along ω_2 axis which corresponds to Σ H1' = J(H1'-H2') + J(H1'-H2") lies in the range 14-15 Hz (Table 3.3). This rules out

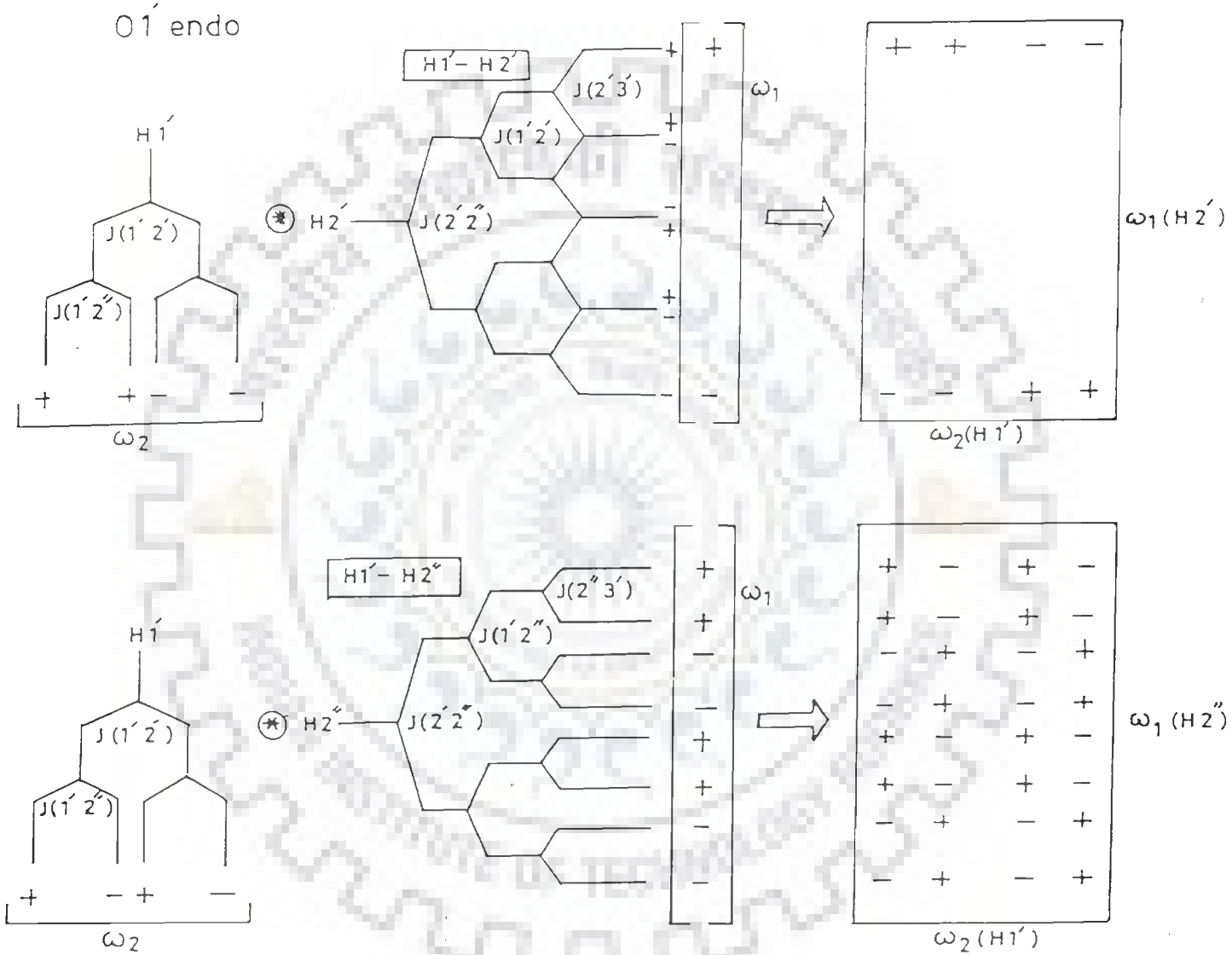
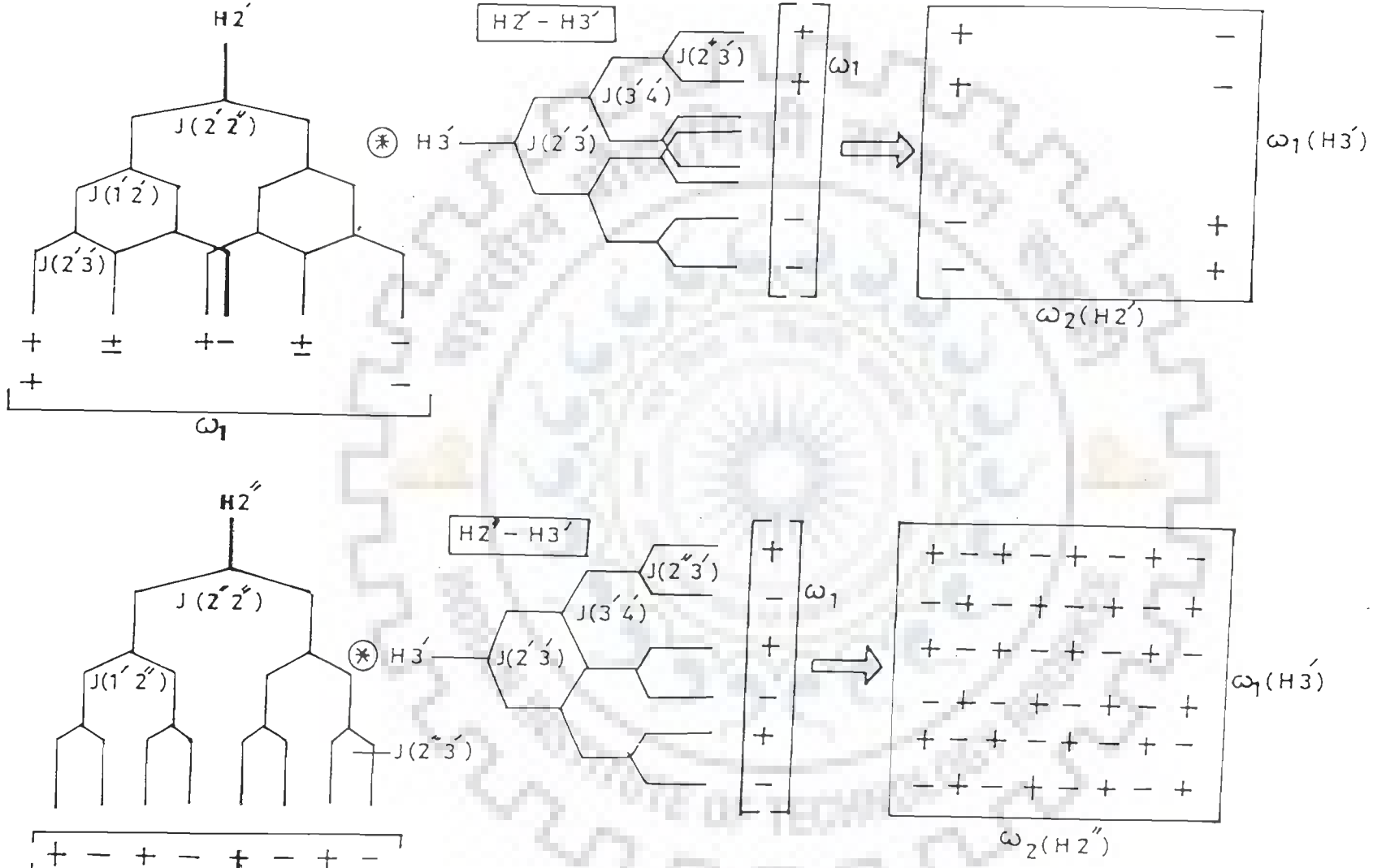
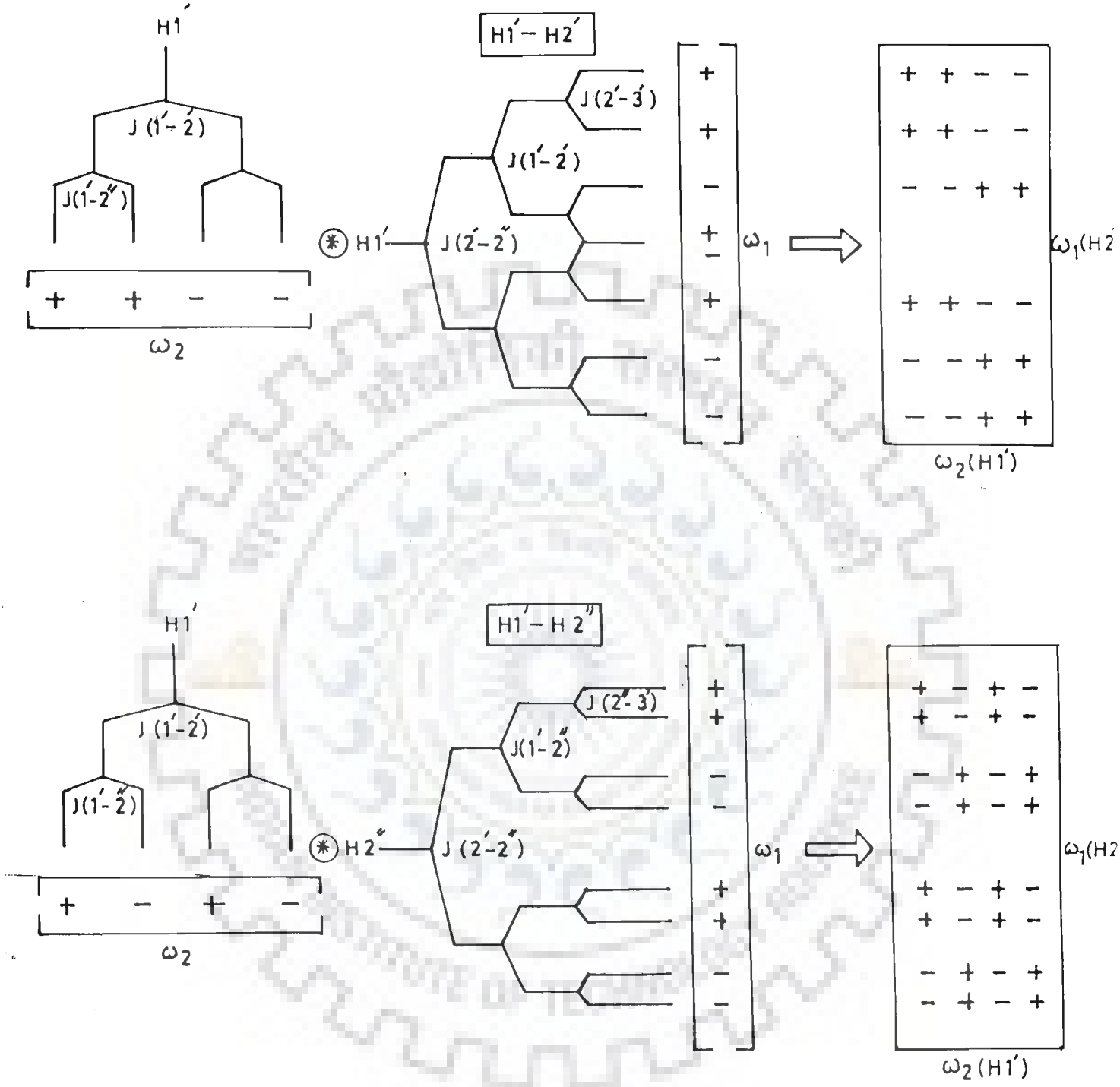


FIG. 3.10: Procedure for the calculation of crosspeak patterns in COSY spectra, illustrated for (a) O1' endo, (b) C2' endo and (c) C3'endo for (i) H1'-H2' (ii) H1'-H2" (iii) H2'-H3' and (iv) H2"-H3'

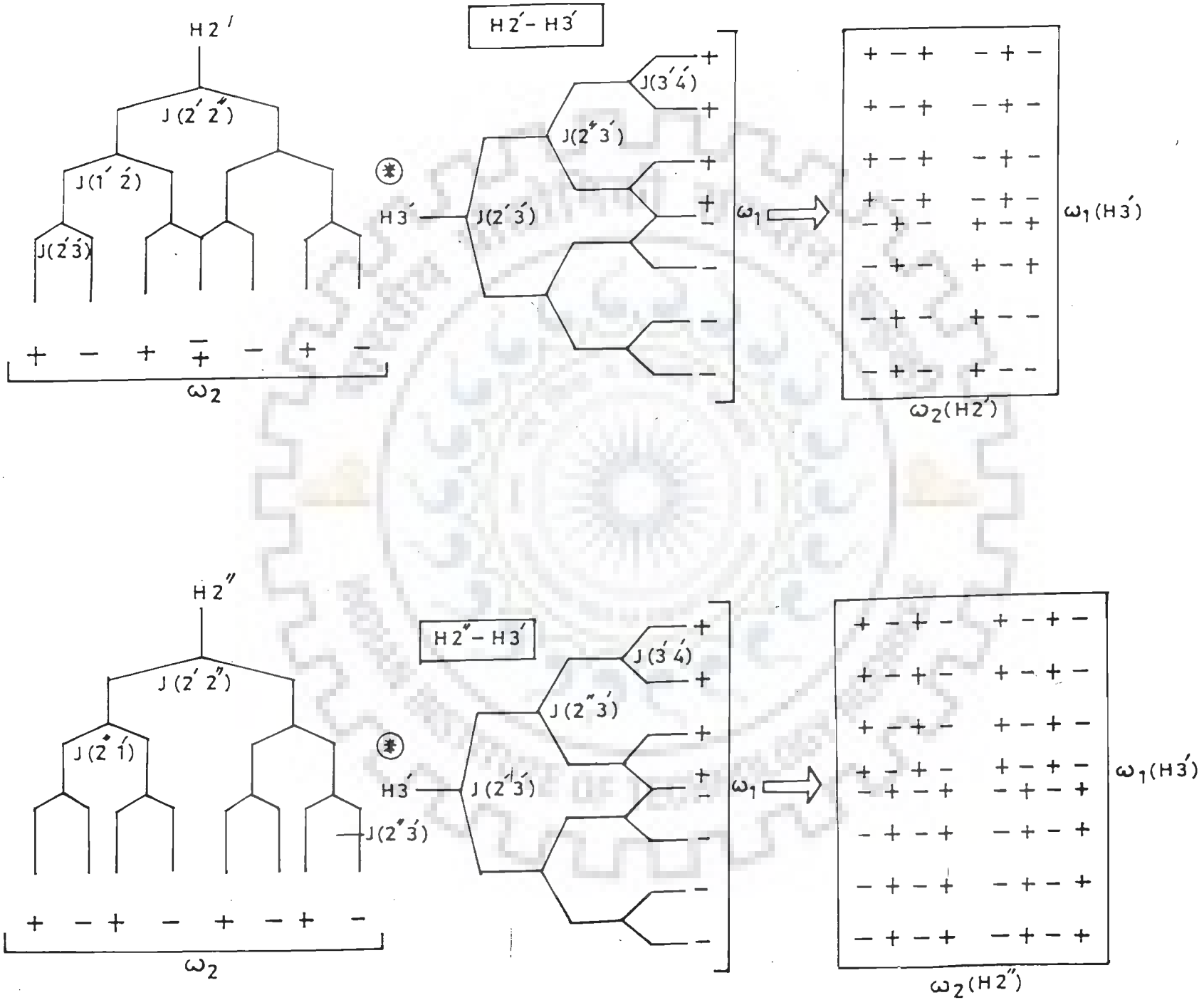
(a) 01′ endo


H2′
J(2′2″)
J(1′2′)
J(2′3′)
+ ± +− ± −
+ −
ω1
H2′ − H3′
J(2′3′)
J(3′4′)
H3′
J(2′3′)
+
+
ω1
−
−
+ −
+ −
ω1(H3′)
− +
− +
ω2(H2′)
H2″
J(2′2″)
J(1′2″)
H3′
J(2″3′)
+ − + − + − + −
ω2
H2″ − H3′
J(2″3′)
J(3′4′)
J(2′3′)
+
−
+
−
+
−
ω1
+ − + − + − + −
− + − + − + − +
+ − + − + − + −
− + − + − + − +
+ − + − + − + −
− + − + − + − +
ω1(H3′)
ω2(H2″)

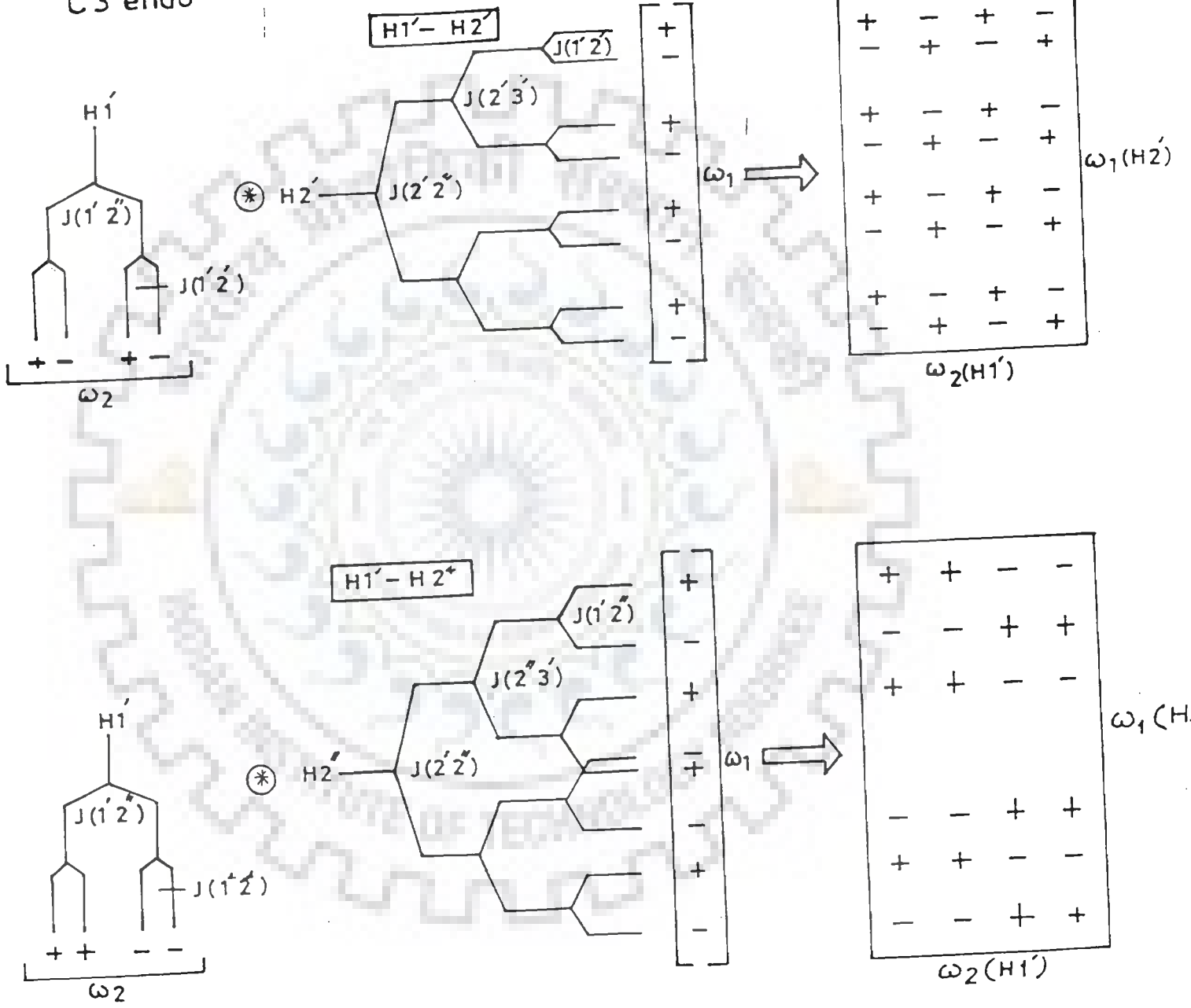
(b) C 2′ endo


H1′
J (1′-2′)
J(1′-2″)
+ + − −
ω2
H1′ − H2′
J(2′-3′)
J(1′-2′)
⊛ H1′
J(2′-2″)
+ + − −
ω1
ω1(H2
+ + − −
− − + +
+ + − −
− − + +
ω2(H1′)
H1′
J (1′-2′)
J(1′-2″)
+ − + −
ω2
H1′ − H2″
J(2′-3′)
J(1′-2″)
⊛ H2″
J (2′-2″)
ω1
ω1(H2
+ − + −
− + − +
+ − + −
− + − +
ω2(H1′)

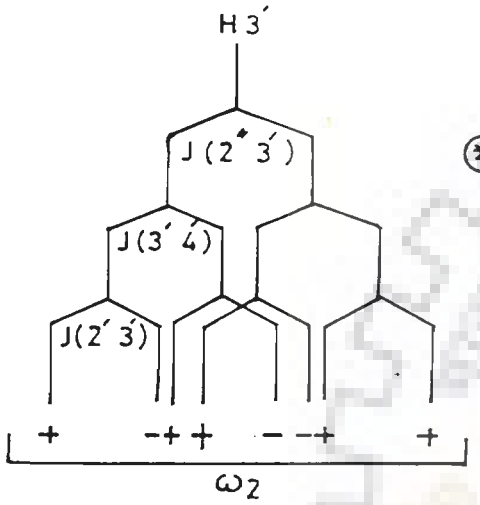
(b) C2′endo

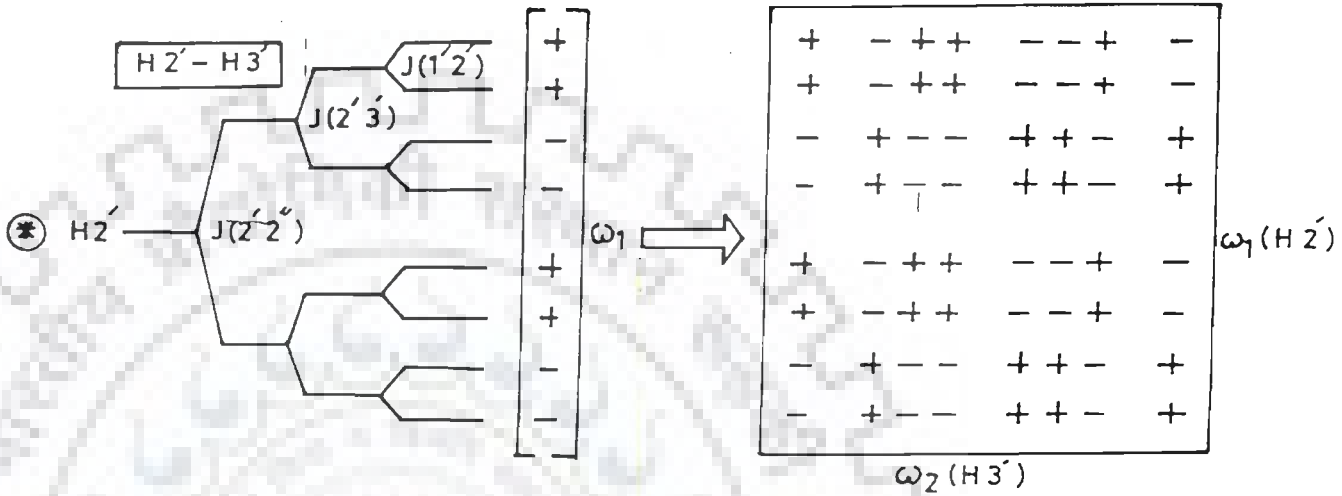

H2′
J(2′2″)
J(1′2′)
J(2′3′)
+ − + − + − + −
ω2
H2′ - H3′
H3′
J(2′3′)
J(3′4′)
J(2″3′)
ω1
ω1(H3′)
ω2(H2′)
H2″
J(2′2″)
J(2″1′)
J(2″3′)
H2″ - H3′
J(3′4′)
J(2″3′)
J(2′3′)
ω1
ω1(H3′)
ω2
ω2(H2″)

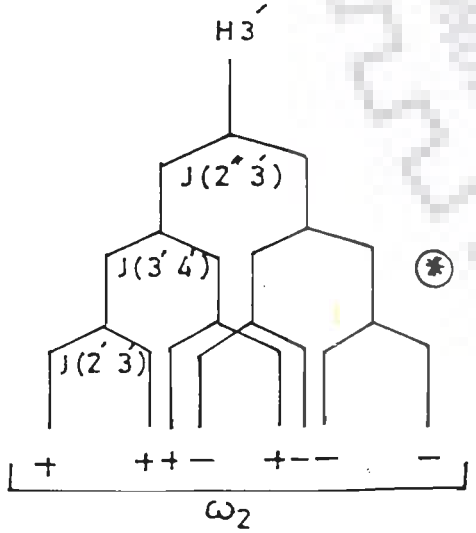
(c) C3´ endo

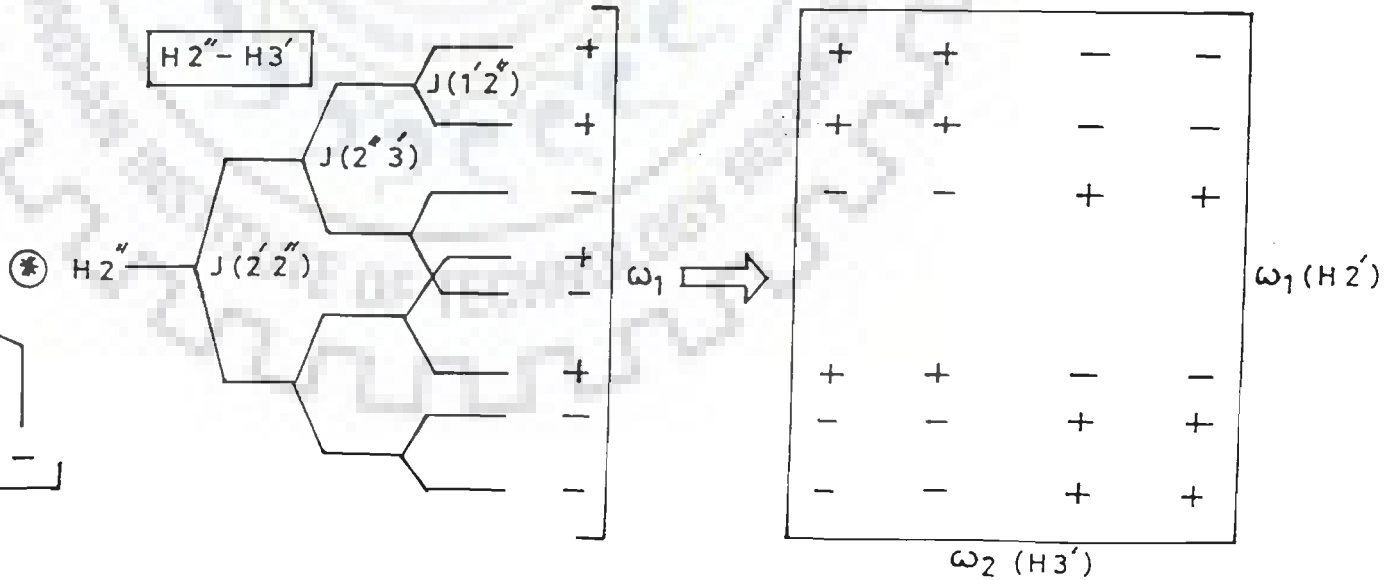

H1´
J(1´2´´)
J(1´2´)
+ − + −
ω2
H1´− H2´
J(1´2´)
J(2´3´)
H2´
J(2´2´´)
ω1
+ − + −
− + − +
+ − + −
− + − +
+ − + −
− + − +
+ − + −
− + − +
ω1(H2´)
ω2(H1´)
H1´
J(1´2´´)
J(1´2´)
+ + − −
ω2
H1´− H2´´
J(1´2´´)
J(2´´3´)
H2´
J(2´2´´)
ω1
+ + − −
− − + +
+ + − −
− − + +
+ + − −
− − + +
ω1(H2´´)
ω2(H1´)

(c) C3′endo


H3′
J(2′3′)
J(3′4′)
J(2′3′)
+ −+ + − −+ +
ω2


H2′ − H3′
J(1′2′)
J(2′3′)
H2′
J(2′2″)
ω1
ω1(H2′)
ω2(H3′)


H3′
J(2′3′)
J(3′4′)
J(2′3′)
+ ++ − +−− −
ω2


H2″ − H3′
J(1′2″)
J(2′3′)
H2″
J(2′2‴)
ω1
ω1(H2′)
ω2(H3′)

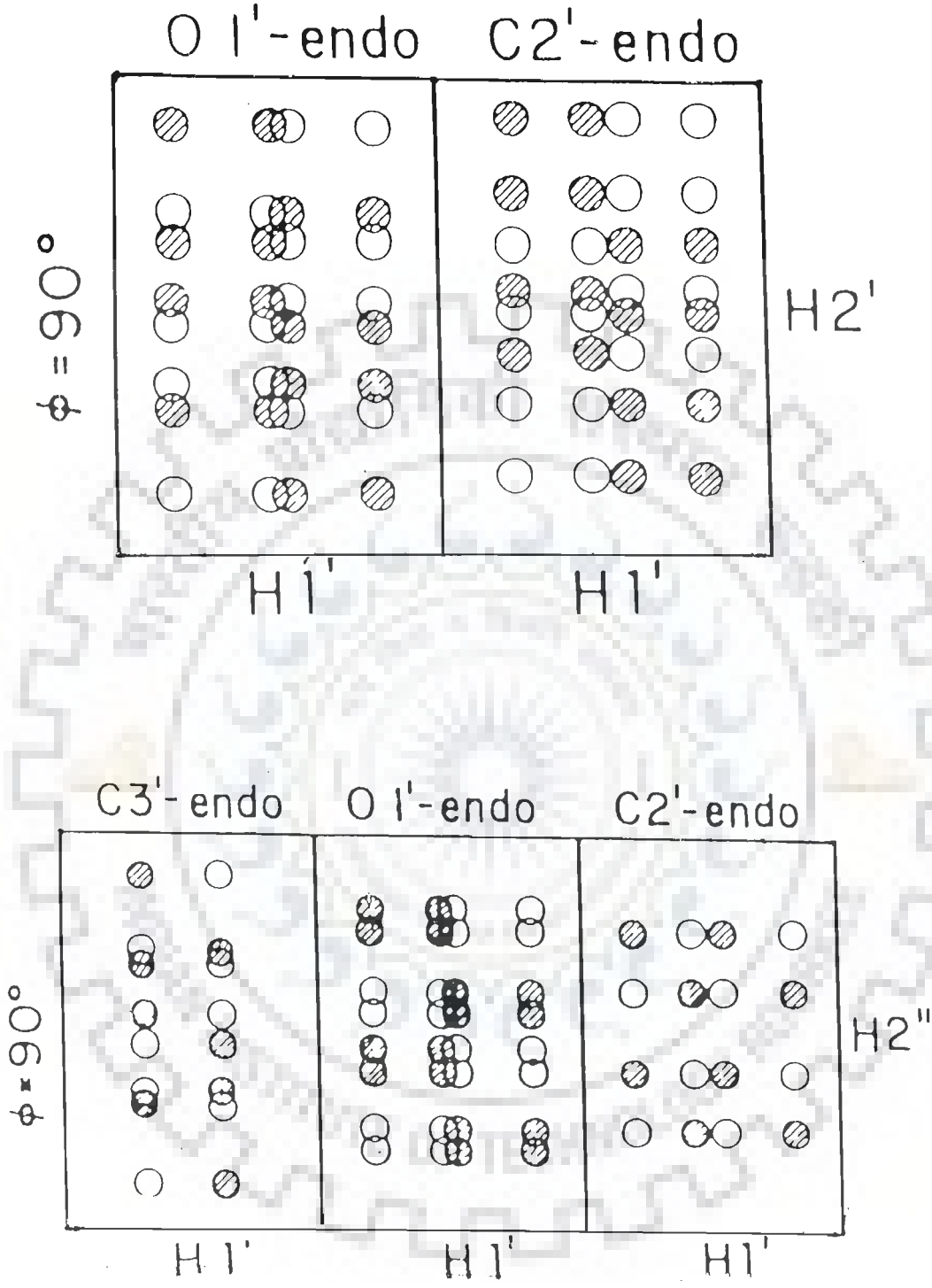


FIG. 3.11: Calculated H1'-H2' and H1'-H2" cross peak patterns for flip angle $\phi = \pi/2$ for C2'endo, O1'endo and C3'endo sugar geometries (77).

Table 3.3 : Standard ^{1}H -^{1}H coupling constants for various suga conformations and also for mixture of C3' endo and C2' endo sugar geometrie (Column 7 and 8).

S.No.	Connectivities	C3' endo	O1' endo	C1' exo	C2' endo	C3' exo	C2' exo	20:80	30:70
1.	H1'-H2'	1.4	8.5	10.5	10.2	8.5	0.7	8.4	7.5
2.	H1'-H2"	8.0	6.7	5.2	5.5	6.7	7.0	6.0	6.3
3.	H2'-H3'	6.8	8.4	6.9	5.1	5.3	6.8	5.4	5.6
4.	H2"-H3'	9.7	3.0	0.7	0.9	0.8	9.7	2.7	3.5
5.	H3'-H4'	8.6	6.8	3.2	0.8	0.5	6.8	2.4	3.2
6.	Σ1'	7.5	15.2	15.7	15.7	15.2	7.7	14.1	13.3
7.	Σ2'	20.0	31.9	32.4	30.3	28.8	22.5	28.2	27.2
8.	Σ2"	33.4	24.7	20.9	21.4	22.5	31.7	23.8	25.0

J(H2'-H2") is assumed as ~ 15 Hz.
Ref. Van Wijk et al.(139) for amplitude of pucker ϕ_m = 36°

TABLE 3.4: Observed Intensity of J-coupled Intrasugar connectivities in COSY spectra of unbound d(A)$_5$ (Fig.3.6(a-i)).

S.No.	Connectivities	A1	A2	A3	A4	A5
1.	H1'-H2'	ss	ss	ss	ss	ss
2.	H1'-H2"	s	s	s	s	s
3.	H2'-H3'	s	s	s	w*	w*
4.	H2"-H3'	w*	×	×	s	s
5.	H3'-H4'	×	×	×	s	s

ss- very strong; s- strong; w- weak; × cross peak not seen
* cross peak seen in magnitude mode COSY spectra only.

the presence of predominant C3' endo or C2' exo sugar geometry for which Σ H1' is expected to be ~ 7.5 Hz (Table 3.3). Further for all residues, the magnitude mode COSY spectra (Fig.3.6(g) clearly revealed that J(H1'-H2') was greater than J(H1'-H2") as more intense cross peaks were seen for H1'-H2' so that C3' endo sugar geometry does not exist.

The C1' exo and C2' endo are expected to give identical cross peak patterns which are different from that expected for O1' endo and C3' exo geometries. Considering the cross peak pattern marked as serial no.3 in Fig.3.6(b) which arises due to A4H1'-A4H2' coupling, along ω_2 axis, the (+ - + -) pattern corresponds to O1' endo/C3' exo since for C2' endo/C1' exo geometries a pattern [- - + +] is expected. This is confirmed further from cross peak pattern marked as serial no.4 arising due to A4H1'-A4H2" coupling which is expected to be [- - + +] and is observed as [- +] under low resolution with the central [- +] peaks merging as their separation is ~ 1.3 Hz. Along ω_1 axis, for H1'-H2" cross peak pattern for serial no. 2 as well as no. 4 the [- - + + - - + +] pattern expected show elongated [- -] and [+ +] peaks on merging due to low resolution.

Reverse is expected for C2' endo/C1' exo geometry for which along ω_2 axis the H1'-H2' cross peak will be [- - + +] while H1'-H2" cross peak will be [- + - +]. This pattern is observed in serial no. 7 and 8 indicating that A1 residue is predominantly C2' endo conformation. We have also obtained estimates of J(H1'-H2') and J(H1'-H2") values wherever possible. For pattern in serial no. 3 and 4, the J(H1'-H2') and J(H1'-H2") are about 7 Hz measured to an accuracy of ± 1.0 Hz. The patterns in serial no. 7 and 8 give J(H1'-H2') and J(H1'-H2") as 10 and 5 Hz, respectively. Proceeding in this way for each nucleotide residue of $d(A)_5$ it was observed that while A1, A2, A3 residues were predominantly C2' endo, the A4 and A5 residues were apparently different, being close to O1' endo geometry.

An important parameter which can be measured to a fairly good accuracy is the sum of coupling constants defined as follows:

$$\Sigma 1' = J(H1'\text{-}H2') + J(H1'\text{-}H2")$$

$$\Sigma 2' = J(H1'\text{-}H2') + J(H2'\text{-}H3') + J(H2'\text{-}H2")$$

$$\Sigma 2" = J(H1'\text{-}H2") + J(H2"\text{-}H3') + J(H2'\text{-}H2")$$

$$\Sigma 3' = J(H2'\text{-}H3') + J(H2"\text{-}H3') + J(H3'\text{-}H4') \quad \{1\}$$

If the chemical shift difference $\Delta\delta = \delta$ (H2')-δ (H2") well exceeds J(H2'-H2") ~ 15 Hz there is no overlap of signals and the Σ2' and Σ2" values measured as distances between outer peaks of H2' and H2" resonance patterns are measured from 1D NMR spectrum. Alternatively Σ1', Σ2' and Σ2" can be read directly from cross peak patterns of phase-sensitive COSY spectrum. The H3' overlap with residual water signal and is also phosphorus coupled so that Σ3' is not used. The importance of sum of J values has been discussed in detail by Van Wijk et al. (139) and are given for a range of sugar geometries of deoxyribose in nucleosides and nucleotides encountered along the pseudorotation itinerary (some values are given in Table 3.3). From Fig.3.6(b) it was found that Σ1' measured as distance between outermost peaks along ω_2 axis for A4 and A5 residues was shorter ~ (13.5-14.0 Hz) than that for other three residues (~15 Hz). Further the difference between Σ2' to Σ2", measured as distances between outermost peaks along ω_1 axis in Fig.3.6(c) for A4 and A5 residues was not ~9 Hz (30.3-21.4 ≅ 8-9 Hz) as expected for C2' endo sugar geometry (Table 3.3) but was ~ 4 Hz (± 1 Hz) only. Typically for A4, A5 residues, Σ2' and Σ2" were measured as 28 and 24 Hz, respectively. The same was also clear from the observed patterns of H2'-H2" cross peaks in phase-sensitive COSY spectra (Fig.3.6(c)). The pattern marked as serial no. 11/12 for A4/A5 residues is almost a square while that of serial no. 15 for A1 residue is a rectangle, with Σ2' being significantly larger than Σ2". Although A4, A5 residues appear to have apparently O1' endo sugar geometry, it was noted that difference between Σ2' and Σ2" was ~ 4 Hz and lesser than 7.2 Hz as expected for O1' endo geometry, the significance of which is

discussed later.

The expansion of phase-sensitive COSY spectra showing H2'-H3' and H2"-H3' connectivities (Fig.3.6(d)) shows that for A4, A5, residues the cross peak pattern for H2"-H3' existed while that for H2'-H3' was not seen. For A1, A2 and A3 residues the case was on the contrary. The H2'-H3' cross peak patterns was seen while those for H2"-H3' did not exist. The magnitude mode COSY of the same region (Fig.3.6(h)) showed a strong (H2"-H3') coupling and a very weak H2'-H3' coupling for A4, A5 residues. For A2, A3 residues, only H2'-H3' connectivity was seen. Further for A1 residue a weak H2"-H3' peak was seen (peak no. 46, Fig.3.6(h)). The results on relative intensities of observed J-coupled intrasugar connectivities are summarised in Table 3.4. From the standard values of J for various sugar geometries (Table 3.3) it is clear that an absence of H2"-H3' cross peak may be interpreted as the presence of C2' endo sugar for the adenine residues.

The H3'-H4' cross peak was clearly visible for A4, A5 residues in both phase-sensitive (Fig.3.6(e)) as well as magnitude mode COSY (Fig.3.6(i)) spectra. For A1 residue, a weak H3'-H4' cross peak is seen (Fig.3.6(e)) while for A2 and A3 residues no such cross peak pattern was found to exist. A very significant observation was that the intensity of H3'-H4' cross peak was as strong as that of H2"-H3' cross peak for A4, A5 residues. This is not expected if sugar geometry is O1' endo since in that case J (H2"-H3') ~ 3.0 Hz while J(H3'-H4') ~ 6.8 Hz (Table 3.3). Further the observed H2"-H3' cross peak was significantly more intense than its corresponding H2'-H3' cross peak (Fig.3.6(d)) for A4, A5 residues. This led us strongly to believe that existence of a single predominant sugar conformation cannot explain the observed results. Alternatively we must look for existence of a rapid interconversion between two sugar conformers i.e. a dynamic two-state model (17,102,139). This is commonly employed with DNA using one conformer from the S region and one from N region of the

pseudorotation cycle. In this model, the two conformation states are defined, a minor conformer (N) with pseudorotation phase angle, P_N, and pucker amplitude, ϕ_N, essentially same as those of C3' endo geometry and a dominant S conformer with a pucker amplitude range $\phi_S = 28^\circ\text{-}44^\circ$ and a pseudorotation phase angle P_S corresponding to approximately the C2' endo range of puckers but subject to some variation (139) in the range 180° to 162°. The resulting coupling constant, J (a-b), between two protons a and b is calculated as:

$$J(a - b) = \chi_S J_S + \chi_N J_N \qquad \{2\}$$

where χ_S and χ_N are the fractional populations of N and S conformers ($\chi_S + \chi_N = 1$), respectively. The approximate fraction of S-type conformer χ_S, is evaluated from the equation:

$$\chi = (\Sigma 1' - 9.4) / (15.7 - 9.4) \qquad \{3\}$$

which is then used to fit experimental values to J(a-b) calculated by using equation {2} by varying both P_S and ϕ_S (17,139). Since the resolution used in our experiments was not so good to yield these parameters with reliability, we did not make any attempt to make an exact quantitative analysis. However the observed $\Sigma 1' \sim 14$ Hz; $\Sigma 2' \sim 28$ Hz; $\Sigma 2'' \sim 24$ Hz ($\pm$ 1 Hz accuracy); and an equally intense (H2'-H3') and (H3'-H4') cross peak; fits into $\chi_S \simeq 0.7$ to 0.8 with $P_S = 162^\circ$ as major S-conformer and $\chi_N = 0.3$ to 0.2 with $P_N = 18^\circ$ as minor N-conformer. The J values expected in such cases using equation {2} are shown in Table 3.3 (column 7 and 8 for reference) and are found to fit into experimental data significantly better than presence of single O1' endo geometry (i.e. J and summation values in column 2).

The results on intensity of cross peaks (Table 3.4) were interpreted as follows:

a) Presence of minor N-conformer along with predominant S-conformer for A4, A5 residues.

b) Existence of S-conformer only for A2, A3 residues.

c) Evidence of minor N-conformer in A1 residue (not

unambiguously confirmed due to weak signals).

We have noted the relative intensities of various cross peaks (Table 3.6) in the NOESY spectra (Fig.3.7(a-h)) to get information on sugar conformation. Since spin diffusion is expected for the mixing time 700 ms and NOE buildup rates are not obtained, the relative intensities are used only to support inference made on the basis of J-couplings. No attempt is made to carry out a rigorous analysis of sugar conformation on the basis of interproton distances. The various intrasugar distances in B-DNA vary with the sugar geometry and are independent of glycosidic bond rotation, χ. These distances (150) and the observed relative intensities of corresponding NOE crosspeaks as seen in NOESY spectra (Fig.3.7(a-h)) are listed in Tables 3.5 and 3.6, respectively. All interproton distances less than 4 Å should be observed, however under the resolution used, the distances < 2.8-2.9 Å only are observable as NOE cross peaks. Considering the parameters in the B-DNA helix, it is found that the distances H1'-H2', H1'-H2", H2'-H3', H2"-H3' and H3'-H4' vary in a small range, that is 2.6-2.8, 2.1- 2.2, 2.2-2.3, 2.5-2.9 and 2.5-2.8 Å, respectively. The distance H1'-H2" being minimum, ~ 2.1-2.2 Å for all sugar geometries, gives the most intense cross peak (labelled as ss (very strong) in Table 3.6) of all the intra sugar cross peaks in NOESY spectra (Fig.3.7(g)). Further, since the distance H1'-H2" remains constant irrespective of the sugar pucker, the intensity of this peak is taken as a reference to obtain a comparatively rough estimate of other distances from the intensity of NOE cross peaks. The intensity of H1'-H2' cross peak is weaker in A1, A2 and A3 residues as compared to their corresponding H1'-H2" cross peaks since the distances involved are comparatively larger being ~ 2.6-2.8 Å irrespective of the sugar geometry. Distances H3'-H4' ~ 2.8 Å give weakly intense (designated as w in Table 3.6) peaks. Thus the observed intensities of cross peaks are in accord with the expected distances. Of the ten intrasugar interproton distances indicated in Table 3.5 (Fig.3.12), eight

TABLE 3.5: Standard intranucleotide distances (Å) between different hydrogen atoms in the deoxyribose ring in various sugar geometries (150).

S.No.	P Connectivities	C2' exo 342°	C3' endo 20°	O1' endo 90°	C1' exo 126°	C2' endo 162°
1.	H1'-H2'	2.6	2.6	2.8	2.8	2.8
2.	H1'-H2"	2.2	2.2	2.1	2.2	2.2
3.	H2'-H3'	2.3	2.3	2.2	2.3	2.3
4.	H2"-H3'	2.9	2.9	2.8	2.6	2.5
5.	H3'-H4'	2.8	2.8	2.8	2.6	2.5
6.	H1'-H4'	3.5	2.9	2.2	2.8	3.0
7.	H2"-H4'	2.5	2.3	3.0	3.7	3.8
8.	H2'-H4'	3.6	3.5	3.6	3.6	3.5
9.	H1'-H3'	3.7	3.6	3.7	3.6	3.6

Table 3.6: Intensities of various connectivities of intranucleotide deoxyribose ring proton as observed in phase-sensitive NOESY spectrum of unbound $d(A)_5$ (Fig.3.7(a-i)).

S.No.	Connectivities	A1	A2	A3	A4	A5
1.	H1'-H2'	s	s	s	ss	ss
2.	H1'-H2"	ss	ss	ss	ss	ss
3.	H2'-H3'	ss	s	s	ss	ss
4.	H2"-H3'	s	ss	x	ss	ss
5.	H3'-H4'	w	s	w	w	w
6.	H1'-H4'	x	x	x	w	w
7.	H2"-H4'	x	x	x	x	x
8.	H2'-H4'	x	x	x	x	x
9.	H1'-H3'	x	x	x	s	s
10.	H3'-H5'	s	s	s	s	s

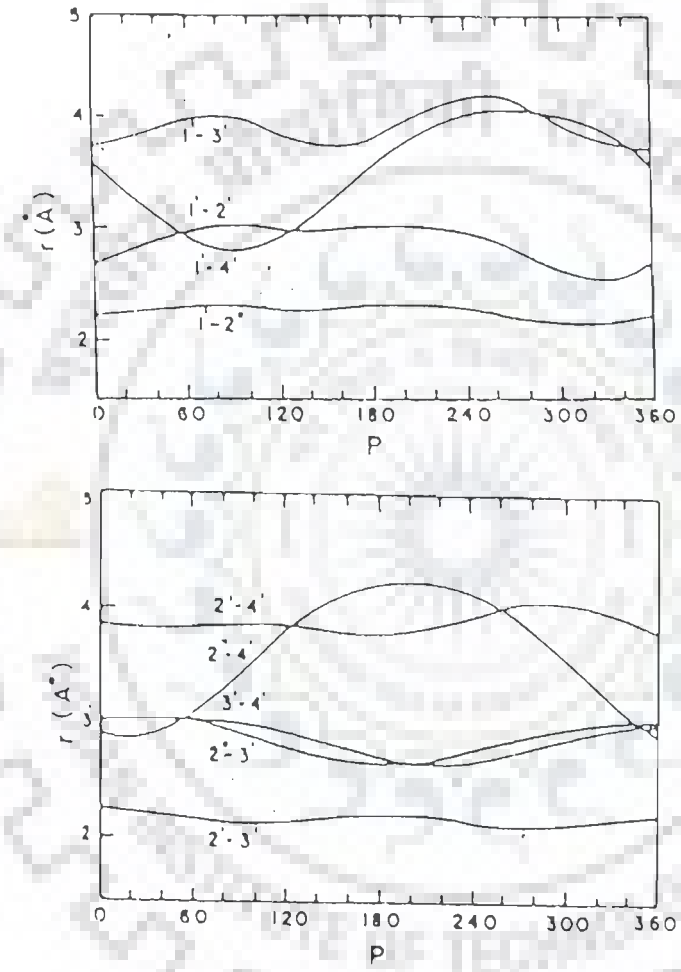


Fig 3.12: Calculated interproton distances in the d-ribose ring as a function of sugar geometry (150).

show a relatively small range of variation of distance, within ± 0.5 Å for different sugar conformations and two, that is, H1'-H4' and H2"-H4' show a wider variation of 2.2-3.5 Å and 2.3-3.8 Å, respectively. It is therefore, the H1'-H4' and H2"-H4' distances that can be used to provide conclusive evidence of the existence of a particular preferred sugar conformation (17,139). The H1'-H4' distance varies by more than 1 Å but the minimum distance (2.2 Å) lies in O1' endo (P=90°) region, thus, this distance cannot discriminate between N-(distance=2.9 Å) and S-(distance=3.0 Å) forms. We observe H1'-H4' cross peak in NOESY spectra only for A4 and A5 residues (Fig.3.7(a)) indicating possibility of O1' endo geometry. Absence of H1'-H4' NOE connectivity (Table 3.6) in A1, A2 and A3 residues demonstrates absence of O1' endo but presence of either C3' endo or C2' endo geometries. Only the H2"-H4' distance serves as a marker as it changes approximately 2.3-2.5 Å in the N region to about 3.6-4.0 Å in the S-conformational space (139). It was observed from the NOESY spectra (Fig.3.7(a)) that H4' peaks of A2 and A3 residues resonate at 4.27 and 4.31 ppm, respectively, did not give any cross peak with corresponding H2" peaks resonating at 2.51 and 2.63 ppm, respectively. Therefore one can safely conclude that A2 and A3 residues do not have N-conformer since the distances are not linear superpositions of the two populations present but are $\langle (rij)^{-6} \rangle$ means weighted by corresponding χ_S and χ_N and are therefore weighted in favour of the shorter of the two interproton distances (139). The result is thus in accord with that obtained from J values. Unfortunately due to low resolution (signal to noise ratio) no conclusive evidence was available for A1, A4 and A5 residues about the presence of N-conformer from NOESY.

Glycosidic bond rotation

In the glycosidic bond rotation, χ, defined as O1'-C1'-N9-C8, may be estimated from the knowledge of intranucleotide distances between the base protons AH8 and their corresponding sugar protons H1', H2', H2", H3' and H4'. Of these five distances, the AH8 to

H1' distance depends only upon the values of χ, while others depend upon both χ and pseudorotation angle P. Table 3.7 (also see Fig.3.13 (a)) gives values of these distances for the major S-conformer with P=162° for χ=-105°, the typical glycosidic bond rotation found in B DNA as well as for two other χ values of -90° and -60°. The base to sugar distance for the minor N-conformer of sugar with P=18° and χ=-105° the value associated with these sugar conformations, are also shown. The relative intensities of corresponding NOE cross peaks (Fig.3.7(a-h)) which have been used to obtain a rough estimate of these distances are shown in Table 3.8. Considering the AH8-AH1' distances for right-handed DNA, the distance varies in the small range 3.5-3.9 Å for χ in the range -90° to -150° (150). Under the resolution used in present experiment, no NOE cross peak is seen as expected. The H8-H2' cross peaks are observed to be fairly strong for all adenine residues (Table 3.8) of sugar pucker the distance is expected to be 2.2 Å. The H8-H2" cross peaks are observed to be weaker than their corresponding H8-H2' NOE cross peak but certainly do not reflect distance as large as ~ 3.6 Å expected for P=162° (i.e. S-conformer) and χ=-105°. Distance H8-H2" can be made possible only if χ is varied from -105° to higher values of say ~ -90° (150). A variation of χ in the opposite sense, that is, a decrease to say -130°, results in a further increase in distance to a value (150) of 4.0 Å contrary to the observed results. Thus we may infer that fairly intense H8-H2" NOE cross peaks leads to existence of higher χ values towards say -90° (±30° accuracy) termed generally as high anti. The base to H3' and H4' distance for S-conformer lies in the range 4.5-5.5 Å and is not expected to yield NOE cross peak. We observe a fairly intense NOE cross peak AH8-AH3' only for A4 residue which is possible only and only if N-conformer is present. The presence of minor N-conformer for A4 residue is in accord with our results based on COSY spectra discussed earlier.

Thus we conclude that $d(A)_5$ is present as a right-handed B-DNA with glycosidic bond angle as high anti and sugar in

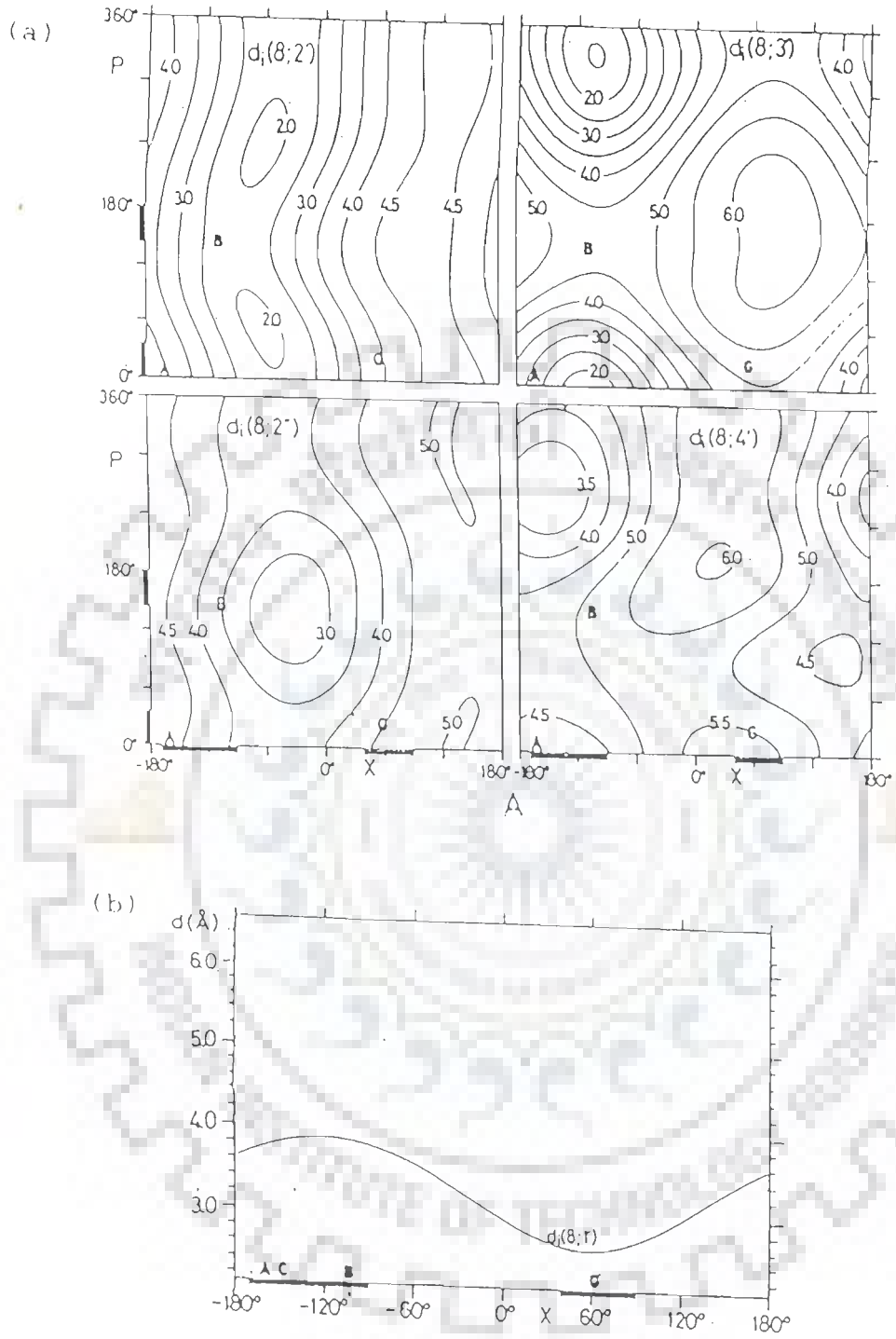


FIG. 3.13(a): P-χ plane with contour lines indicating the H8 to (i) H2'(ii) H2",(iii) H3' and (iv) H4' distances. The preferred regions C3' endo near 20° and C2' endo near 160° for P and syn near 60° and anti near 130° for χ are indicated with heavy lines on the left and the bottom. A, B, C and G identify the χ values for A DNA, B DNA and nucleotide G in the Z form of d-$(CGCGCG)_2$. (b): Intranucleotide distance between AH8 and H1' versus the torsional angle, χ. Preferred χ regions from -90° to -170° (anti) and 40° to 90° (syn) are indicated with heavy lines.

Table 3.7: Intranucleotide distances (Å) between base proton (AH8) and deoxyribose ring protons corresponding to different P and χ values (150).

S.No.	Connectivities	P = 18° χ = -105°	P = 90° χ = -105°	P = 162° χ = -105°	P = 162° χ = -90°	P = 162° χ = -60°
1.	AH8-H1'	3.7	3.8	3.8	3.8	3.7
2.	AH8-H2'	4.0	2.2	2.2	2.2	2.2
3.	AH8-H2"	4.5	4.1	3.6	3.2	2.8
4.	AH8-H3'	3.0	4.0	>4.5	>4.5	>4.5
5.	AH8-H4'	4.4	5.0	5.0	>5.0	>5.0

Table 3.8: Intensities of various connectivities between base proton (AH8) and deoxyribose ring protons as observed in phase-sensitive NOESY of unbound $d(A)_5$ (Fig.3.7(a-h)).

S.No.	Connectivities	A1	A2	A3	A4	A5
1.	AH8-H1'	x	x	x	x	x
2.	AH8-H2'	ss	s	s	s	s
3.	AH8-H2"	s	s	s	s	s
4.	AH8-H3'	x	x	x	s	x
5.	AH8-H4'	x	x	x	x	x

S-conformational state. A minor N-conformer is present for A4 and A5 residue.

D) Conformation of $d(A)_5$ bound to hexapeptide

We have made complex of $d(A)_5$ with hexapeptide KPYSLN at varying N/P ratios and recorded 1D NMR spectra at 297 K (discussed later). The phase-sensitive COSY (Fig.3.14(a-e)) and NOESY spectra (Fig.3.15(a-d)) are recorded for the mixture of 1.92 mM $d(A)_5$ and 3.80 mM KPYSLN (N/P = 0.51). The spectral assignment of specific nucleotide and peptide protons in the spectra of $d(A)_5$ bound to hexapeptide is made by using the corresponding 1D NMR and 2D NMR results of $d(A)_5$ alone as well as the strategies for spectral assignment detailed in section C. The chemical shift positions of various protons of $d(A)_5$ on interaction with hexapeptide are given in Table 3.9a. The chemical shift of hexapeptide protons in the mixture of 1.92 mM $d(A)_5$ and 3.80 mM KPYSLN are also shown in Table 3.9b. It may be noted that at N/P = 0.51 uncomplexed KPYSLN exists in solution if 1:1 complex is formed. However this does not complicate the analysis of conformation of $d(A)_5$ as most of the $d(A)_5$ in the mixture exists in bound state.

Conformation of deoxyribose sugar

We have followed the strategies used in determining the conformation of $d(A)_5$ alone for analysing the phase-sensitive COSY spectra of complex shown in Fig.3.14(a-e). The sum of couplings $\Sigma 1'$ measured as distance between + - contours of H1'-H2' and H1'-H2" cross peak patterns along ω_2 axis, was found to be ~ 14-15 Hz for all residues indicating the presence of S-conformer as major conformer of the deoxyribose sugar. The cross peak patterns H1'-H2' and H1'-H2" of A5 residue, i.e. marked as serial no. 1 and 2, respectively in Fig.3.14(b), are of the type [+ -] along ω_2 axis with identical separation between + and - contours. This is due to the fact that both [+ -+ -] and [+ + - -] patterns, expected for H1'-H2" and H1'-H2' cross peak, respectively for O1' endo sugar geometry look like [+ -] under the resolution used. On

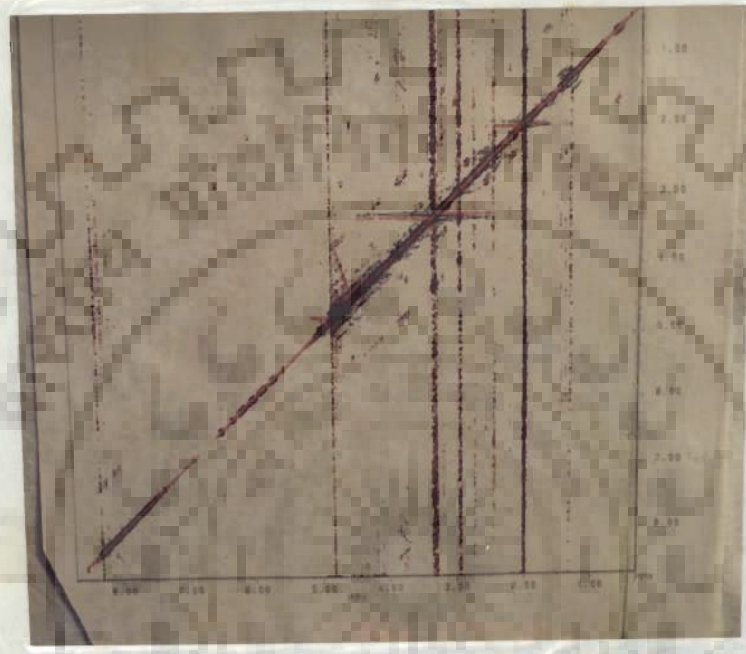


FIG. 3.14(a): Phase-sensitive COSY spectrum of a mixture of 1.92 mM $d(A)_5$ and 3.80 mM KPYSLN in D_2O solution in phosphate buffer (0.01M) at 297 K.

CENTRAL LIBRARY — IIT ROORKEE

FIG. 3.14(b): A portion of phase-sensitive COSY spectrum of a mixture of 1.92mM $d(A)_5$ and 3.80 mM KPYSLN in phosphate buffer (0.01 M) expanded to show specific connectivities: (1) A5H1'-A5H2' (2) A5H1'-A5H2" (3) A4H1' -A4H2'(4) A4H1'-A4H2" (6) A3H1'-A3H2" (7) A1H1'-A1H2' (8) A1H1'-A1H2"(9) A2H1'-A2H2'

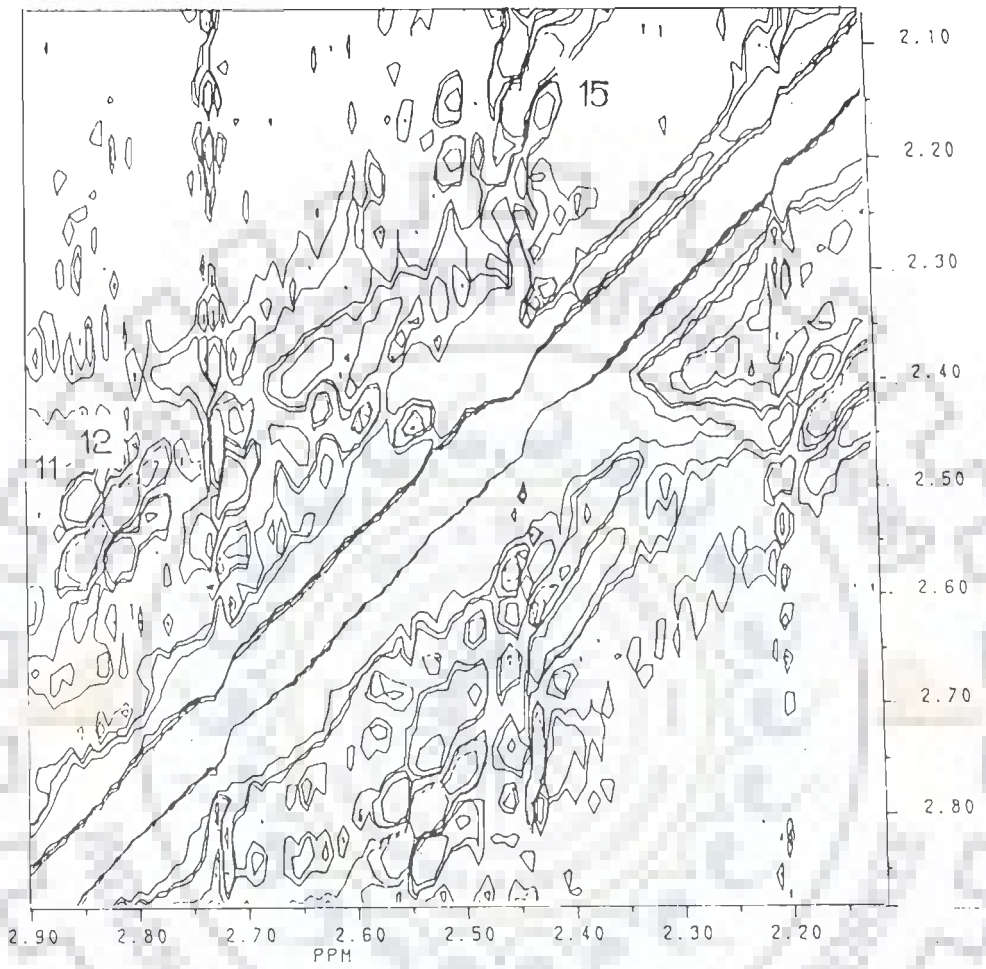


Fig. 3.14(c):A portion of phase-sensitive COSY spectrum of a mixture of 1.92mM $d(A)_5$ and 3.80 mM KPYSLN in phosphate buffer (0.01 M) expanded to show specific connectivities:(11) A5H2'-A5H2" (12) A4H2'-A4H2" (15)A1H2'-A1H2" .

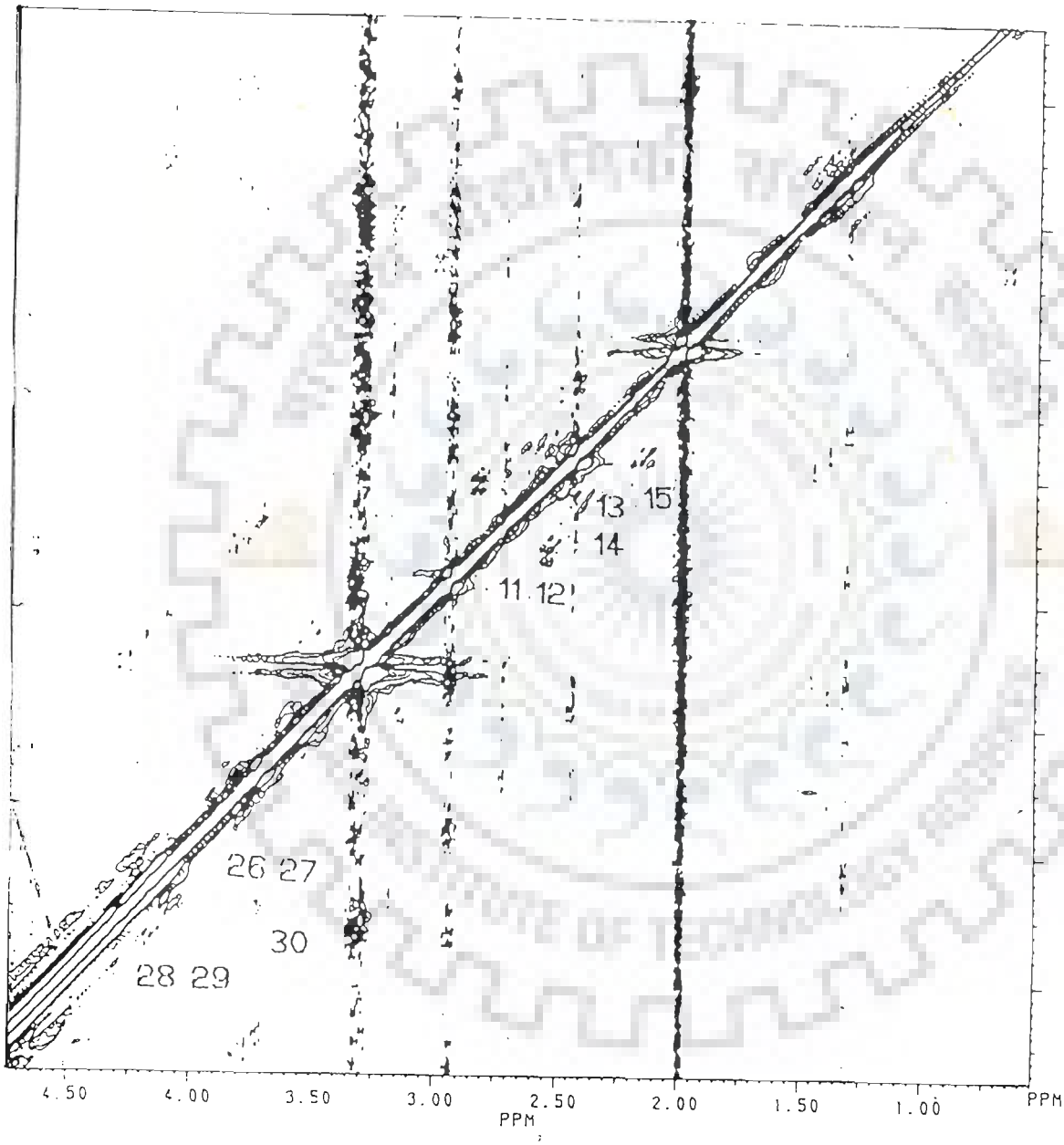


FIG. 3.14(d): A portion of phase-sensitive COSY spectrum of a mixture of 1.92 mM $d(A)_5$ and 3.80 mM KPYSLN in phosphate buffer (0.01 M) expanded to show specific connectivities: (11) A5H2'-A5H2" (12) A4H2'-A4H2" (13)A3H2' -A3H2" (14)A2H2'-A2H2" (15) A1H2'-A1H2" (26) A5H4'-A5H5' (27) A4H4'-A4H5' (28) A3H4'-A3H5' (29) A2H4'-A2H5' (30) A1H4'-A1H5'

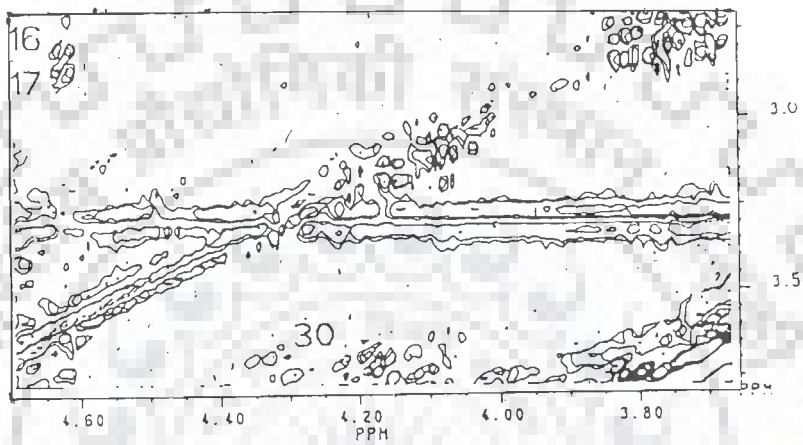


FIG. 3.14(e): A portion of phase-sensitive COSY spectrum of a mixture of 1.92 mM $d(A)_5$ and 3.80 mM KPYSLN in phosphate buffer (0.01 M) expanded to show specific connectivities: (16) A5H2"-A5H3' (17) A4H2"-A4H3' (30) A1H4' -A1H5'

the contrary the H1'-H2' cross peak pattern of A1 residue, marked as serial no. 7, shows + and - contours broadened along ω_2 axis as a pattern [+ + - -] is expected for C2' endo geometry. Thus by comparing with cross peak pattern expected for standard geometries shown in Fig.3.10, it is observed that there is no major change in sugar conformation of adenine residues of $d(A)_5$ on binding to hexapeptide. The patterns in serial no. 1-4 conform to O1' endo geometry while others are close to C2' endo geometry of deoxyribose. Further the cross peak pattern, marked as serial no. 11, 12, in Fig.3.14(c) due to H2'-H2" J-connectivity for A4 and A5 residues is almost square as compared to that for A1 residue, marked as serial no. 15, in Fig.3.14(c) which appears to be rectangular. The distances between [+ -] contours along ω_1 and ω_2 axis correspond to Σ2' and Σ2" values. The Σ2' and Σ2" differ by about 2-4 Hz for mixed sugar conformers, that is, 20-30% of N-conformer mixed with 80-70% of S-conformer. The difference is about 7 and 9 Hz for O1' endo and C2' endo geometries, respectively. Further fairly intense H2"-H3' cross peaks are seen for A4 and A5 residue in Fig.3.14(a and d). Thus all the features seen in COSY spectra of $d(A)_5$ in unbound form exist clearly in the penta-adenylate bound to hexapeptide. Thus the conformation of deoxyribose sugar in $d(A)_5$ does not change significantly on binding to hexapeptide.

Glycosidic bond rotation

Fig.3.15 (a-d) shows phase-sensitive NOESY spectra of $d(A)_5$ bound to hexapeptide in N/P ratio of 0.51 and Table 3.11 gives the relative intensities of some of the NOE cross peaks seen in these spectra. Comparing the results with those obtained for $d(A)_5$ alone (Fig.3.7(a-h) and Tables 3.6 & 3.8) we conclude that there is no major change in glycosidic bond rotation.

E) Changes in Chemical Shift of Peptide and Nucleotide protons on binding

Table 3.12 (a and b) shows change in chemical shift of adenine and

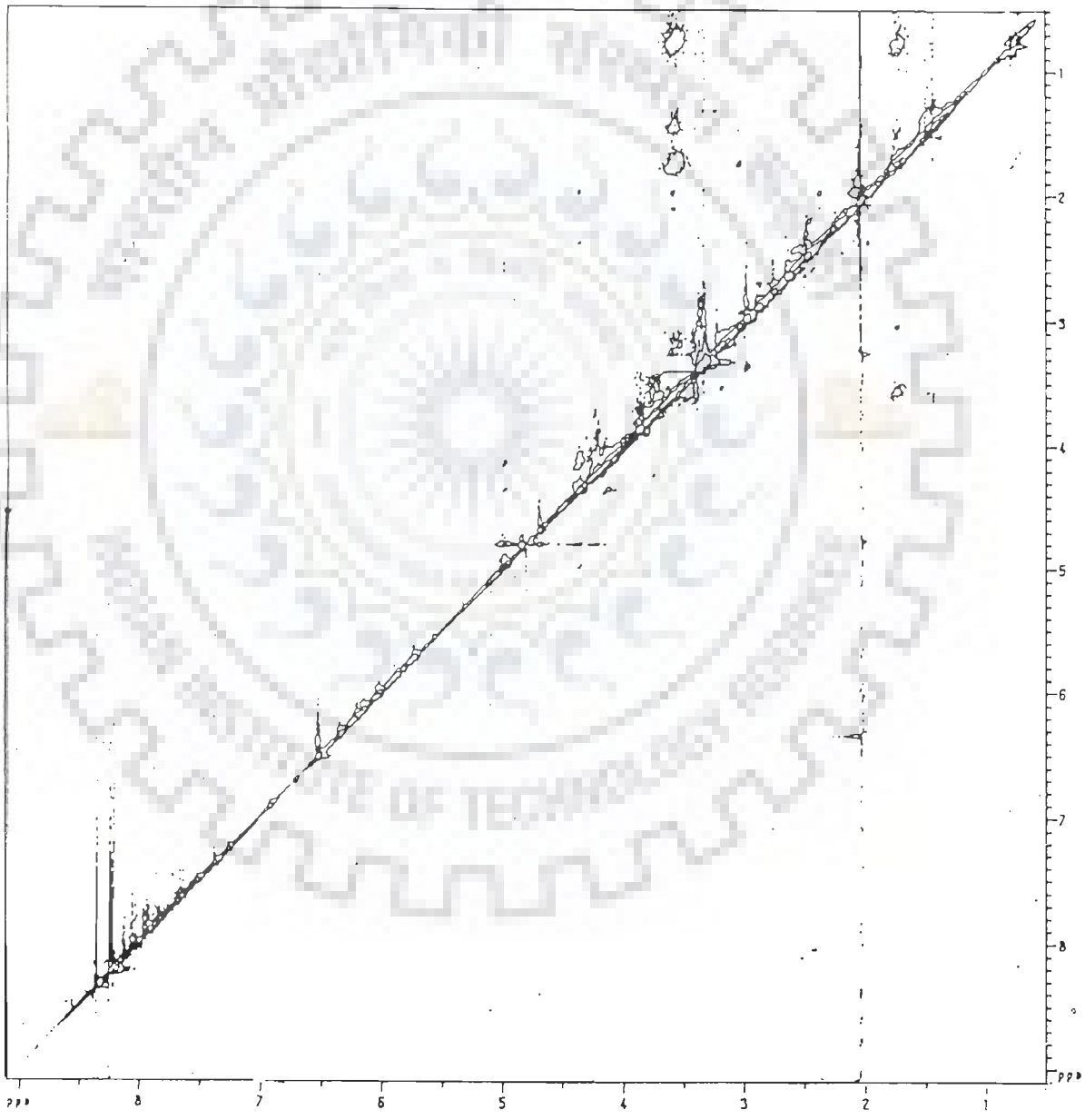


FIG. 3.15(a): Phase-sensitive NOESY spectrum of a mixture of 1.92 mM $d(A)_5$ and 3.80 mM KPYSLN in phosphate buffer (0.01 M) at 297 K.

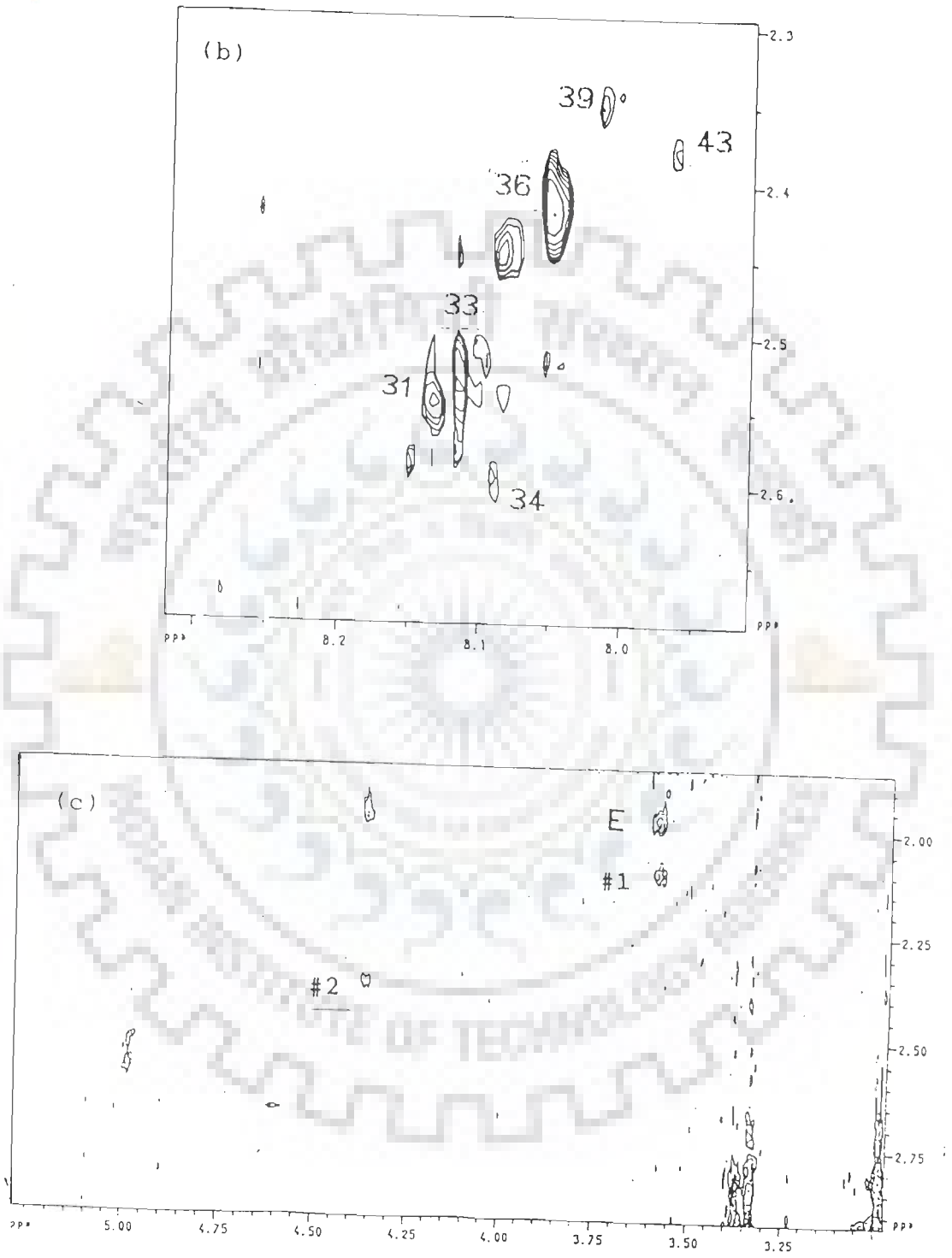


FIG. 3.15(b): Phase-sensitive NOESY spectrum of a mixture of 1.92 mM $d(A)_5$ and 3.80 mM KPYSLN at 297 K expanded to show specific connectivities: (31) A5H8-A5H2'(33) A4H8- A4H2' (34) A4H8-A3H2" (36) A3H8-A3H2' (37) A3H8-A2H2" (39) A2H8-A2H2" (43) A1H8- A1H2" (c): An expansion of Fig.3.15(a) to show specific connectivities:(E)Proδ-CH_2 - Pro γ-CH_2 (#1)Pro δ-CH_2 -A1H2' (#2) Leu α-CH - A2H2'

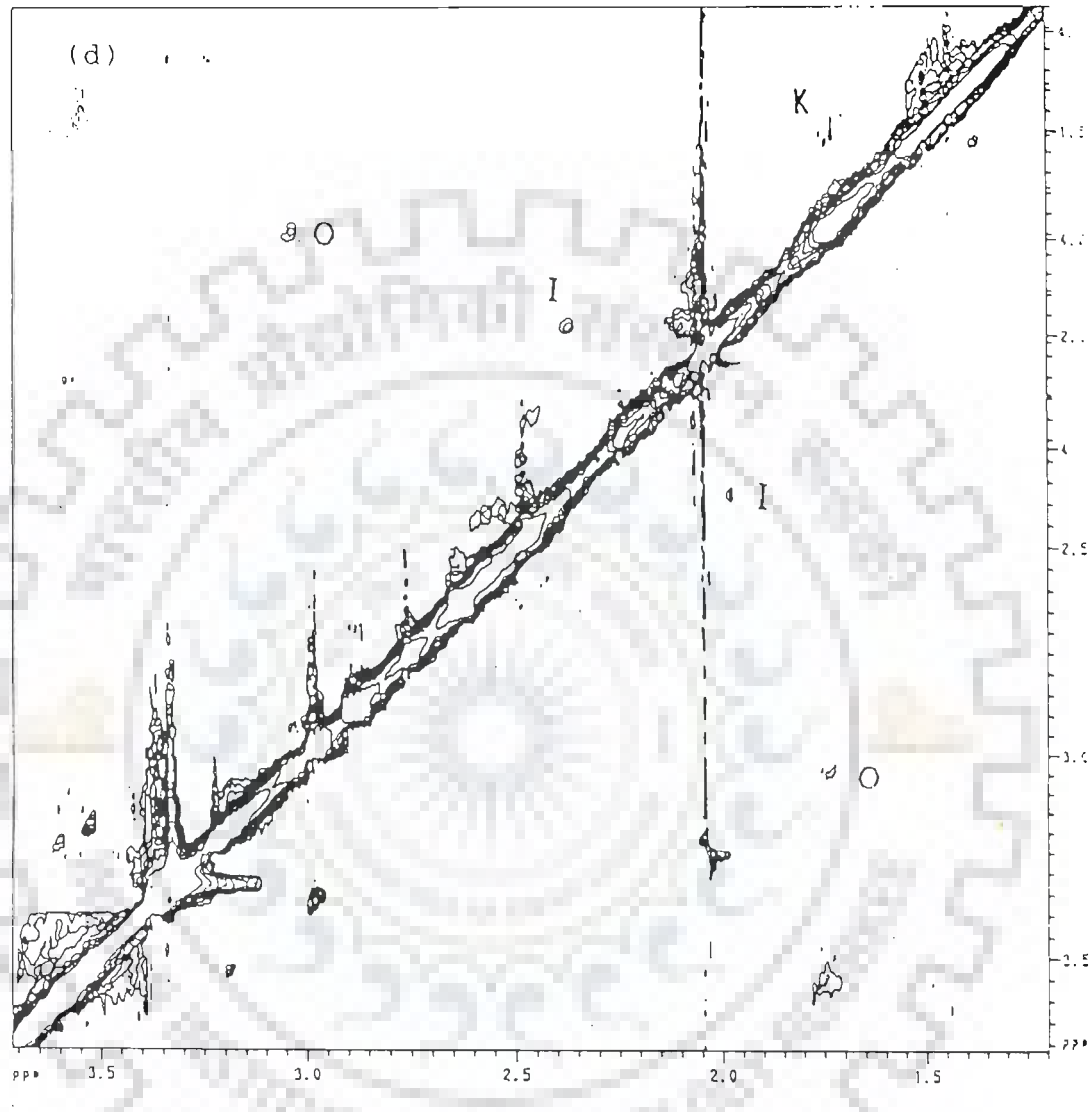


FIG. 3.15(d): An expansion of Fig.3.15(a) to show specific connectivities (K) Lys γ-CH - Lys δ-CH (I) Pro γ-CH - Pro β-CH (O) Lys δ-CH - Lys ε-CH

hexapeptide protons due to interaction at N/P = 0.51 as a difference of chemical shift observed in the spectra of mixture of 1.92 mM $d(A)_5$ and 3.80 mM KPYSLN, i.e. N/P = 0.51 (Table 3.9 a and b), and that in the spectra of the corresponding unbound forms (Table 3.1 and 3.2). To determine if the change in chemical shift varies gradually with increase in binding, we carried out 1D NMR experiments on a mixture of $d(A)_5$ and hexapeptide at six different concentration ratios, N/P in the range 0 to 0.51. These experiments are not aimed at determining the stoichiometry of the hexapeptide-$d(A)_5$ complex since it is fairly well established in literature (3, 21) that 4 or 5 nucleotide residues are covered/bound to each Gene V protein molecule on a single-stranded DNA. NMR spectra recorded at different N/P ratios are shown in Fig.3.16(a and b). Since all the proton resonances are assigned unambiguously, the chemical shift of every proton in each of the spectrum is ascertained. The chemical shifts as well as changes in chemical shifts are plotted as a function of N/P in Fig.3.17(a and b, c and d) and tabulated in Table 3.13 (a and b). A significant feature of the results on binding of $d(A)_5$ to hexapeptide is that only one set of NMR lines are observed in the bound complex for both $d(A)_5$ as well as hexapeptide. This is indicative of strong binding (17). In these experiments 5.0 mM hexapeptide is taken and $d(A)_5$ is added to it in steps (Table 2.1 Chapter II). For practically all adenine H2, H8, H1', etc. protons, it is observed (Fig.3.17(c)and Table 3.13b) that the change in chemical shift due to binding occurs at N/P ~ 0.08 or 0.15 and does not change any further on increasing N/P. This shows that as $d(A)_5$ concentration is increased, more of $d(A)_5$ exists completely in bound state in all the mixtures. However at N/P = 0.51, the hexapeptide is present in excess of $d(A)_5$ so that some of it is likely to be present in free state. The free and bound states may be exchanging fast on NMR time scale.

The results on change in chemical shift ($\Delta\delta$) of various nucleotide protons (Table 3.13 b and Fig.3.17((c)) due to interaction show

Table 3.9a: Chemical shift values of various protons of $d(A)_5$ in the bound form for N/P=0.51 in D_2O at 297 K.

Proton	A1	A2	A3	A4	A5
H1'	5.96	5.80	6.13	6.29	6.45
H2'	2.14	2.38	2.42	2.52	2.55
H2"	2.42	2.61	2.72	2.78	2.83
H3'	-	4.71	4.71	4.64	4.64
H4'	4.23	4.47	4.47	4.17	4.17
H5'	3.70	4.10	4.10	3.76	3.76
H8	7.96	8.03	8.07	8.12	8.13
H2	7.49	7.60	7.70	7.87	8.00

Table 3.9b: Chemical shift values δ (in ppm) of various protons of hexapeptide in a mixture of 1.92 mM $d(A)_5$ and 3.60 mM KPYSLN in phosphate buffer (0.01 M) at 297 K.

	Lys	Pro	Tyr	Ser	Leu	Asn
α-CH	4.48	-	-	-	4.35	-
β-CH_2	2.00	2.55	2.99	-	1.68	-
γ-CH_2	1.44	1.96	-	-	1.68	-
δ-CH_2	1.96	3.57	-	-	-	-
δ-CH_3	-	-	-	-	0.65	-
ε-CH_2	2.96	-	-	-	-	-
(3,5)H	-	-	6.85			
(2,6)H			7.12			

Table 3.10: Presence (marked as +) and absence (marked as -) of cross peak pattern as observed in phase-sensitive COSY spectra of intraresidue $d(A)_5$ bound to hexapeptide in N/P ratio of 0.51.

S.No.	Connectivities	A1	A2	A3	A4	A5
1.	H1'-H2'	+	+	+	+	+
2.	H1'-H2"	+	+	+	+	+
3.	H2'-H3'	-	-	-	-	-
4.	H2"-H3'	-	-	o	+	+
5.	H3'-H4'	-	-	-	o	o

o-overlap

Table 3.11: Intensities of some of the intraresidue NOE connectivities from the NOESY spectra of $d(A)_5$ bound to hexapeptide in N/P ratio of 0.51 (Fig.3.15(a-d)).

S.No.	Connectivities	A1	A2	A3	A4	A5
1.	H2'-H3'	o	o	ss	o	o
2.	H2"-H3'	x	x	x	x	x
3.	AH8-H1'	x	x	x	x	x
4.	AH8-H2'	o	s	s	s	s
5.	AH8-H2"	s	o	o	o	o

ss- very strong; s- strong; x- not seen; o- overlap

that the $\Delta\delta$ for base residues lies in the range -0.14 ppm (upfield) to + 0.14 ppm (downfield). The terminal adenines, in particular A1H8, A5H8 and A5H2, are affected the most on binding. Further the A1, A2 and A3 base protons show net downfield shift while A4 and A5 base protons show net upfield shift. However both H2 and H8 base protons of any one residue shift in the same direction, either upfield or downfield although to different extent (Table 3.13b and Fig.3.17(c)). This is expected, particularly if the binding involves stacking interaction. If a particular adenine residue of $d(A)_5$ experiences, net stacking, both H2 and H8 will be affected in same way but by different magnitudes due to their different locations on the aromatic base. The H1' atom of deoxyribose sugar being close to the aromatic ring of bases does experience upfield/downfield shift due to stacking/destacking. We observe a change in $\Delta\delta$ ~ -0.01 to +0.09 ppm in H1' resonance of different residues of $d(A)_5$. A striking observation was large change in H3' and H4' proton resonances of A2 and A3 residues of $d(A)_5$ (Table 3.12a) due to interaction. Although a straightforward interpretation of shifts in sugar resonances is problematic, a large change certainly reflects extent of involvement of these protons in complex formation.

Considering the change in chemical shift of various peptide protons due to interaction of hexapeptide KPYSLN with $d(A)_5$ (Table 3.12b, 3.13b and Fig.3.17(d)), it is observed that several protons show practically no significant change in their positions. The tyrosine ring protons, (3,5)H and (2,6)H, shift upfield by -0.05 ppm and -0.03 ppm, respectively. The upfield shift may be attributed to stacking with bases of $d(A)_5$ (85,86). The Tyr β-CH_2 experiences shift by -0.07 ppm as a result of interaction. The lysine β-CH_2 and δ-CH_2 experience large downfield shifts by 0.17 and 0.26 ppm, respectively (Figs.3.16(b) and 3.17(d)). On the other hand Lys ε-CH_2 does not shift significantly. The Pro β-CH_2 and Leu δ-CH_3 show remarkable change due to binding. The shift in Leu β-CH_2 and γ-CH_2 is to a lesser extent. The asparagine and

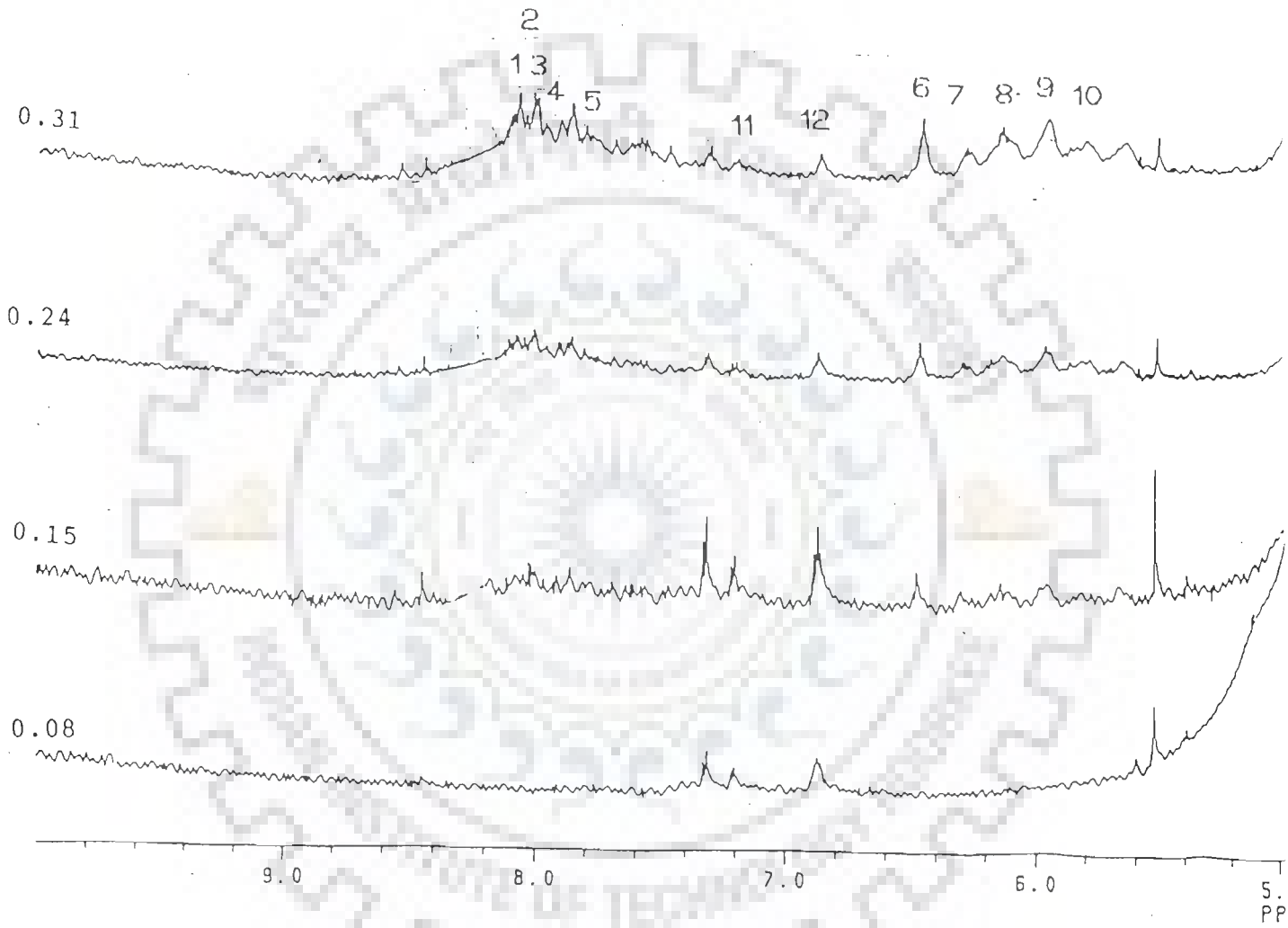
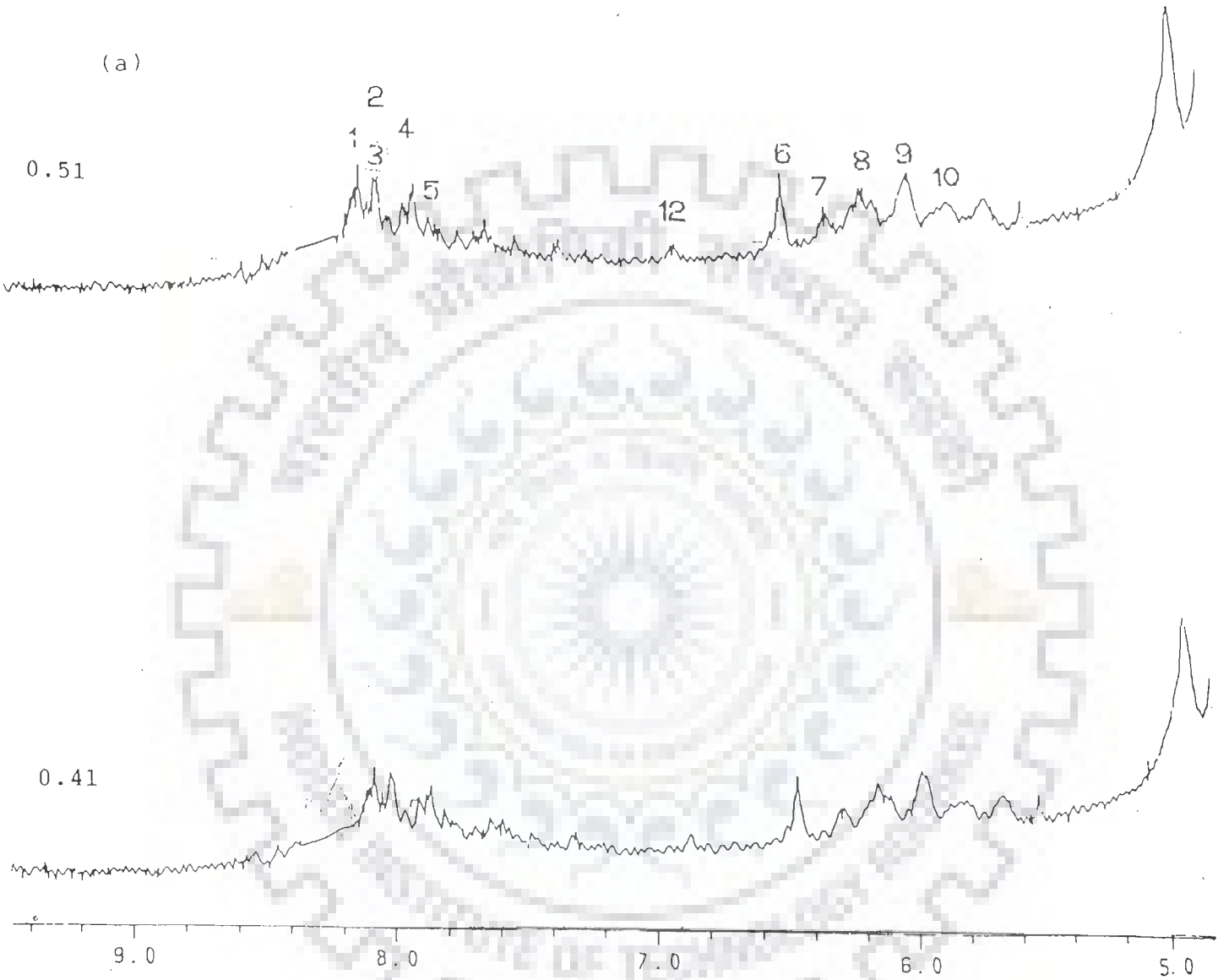


FIG. 3.16(a): Series of one dimensional spectrum recorded at varying concentration ratio (N/P). Peaks are labelled as follows:1-A5H8, 2-A4H8, 3-A3H8, 4-A2H8, 5-A1H8, 6-A5H1', 7-A4H1', 8-A3H1', 9-A1H1',10-A2H1', 11-Tyr(2,6)H, 12-Tyr(3,5)H.


(a)
0.51
0.41
1
2
3
4
5
6
7
8
9
10
12
9.0
8.0
7.0
6.0
5.0

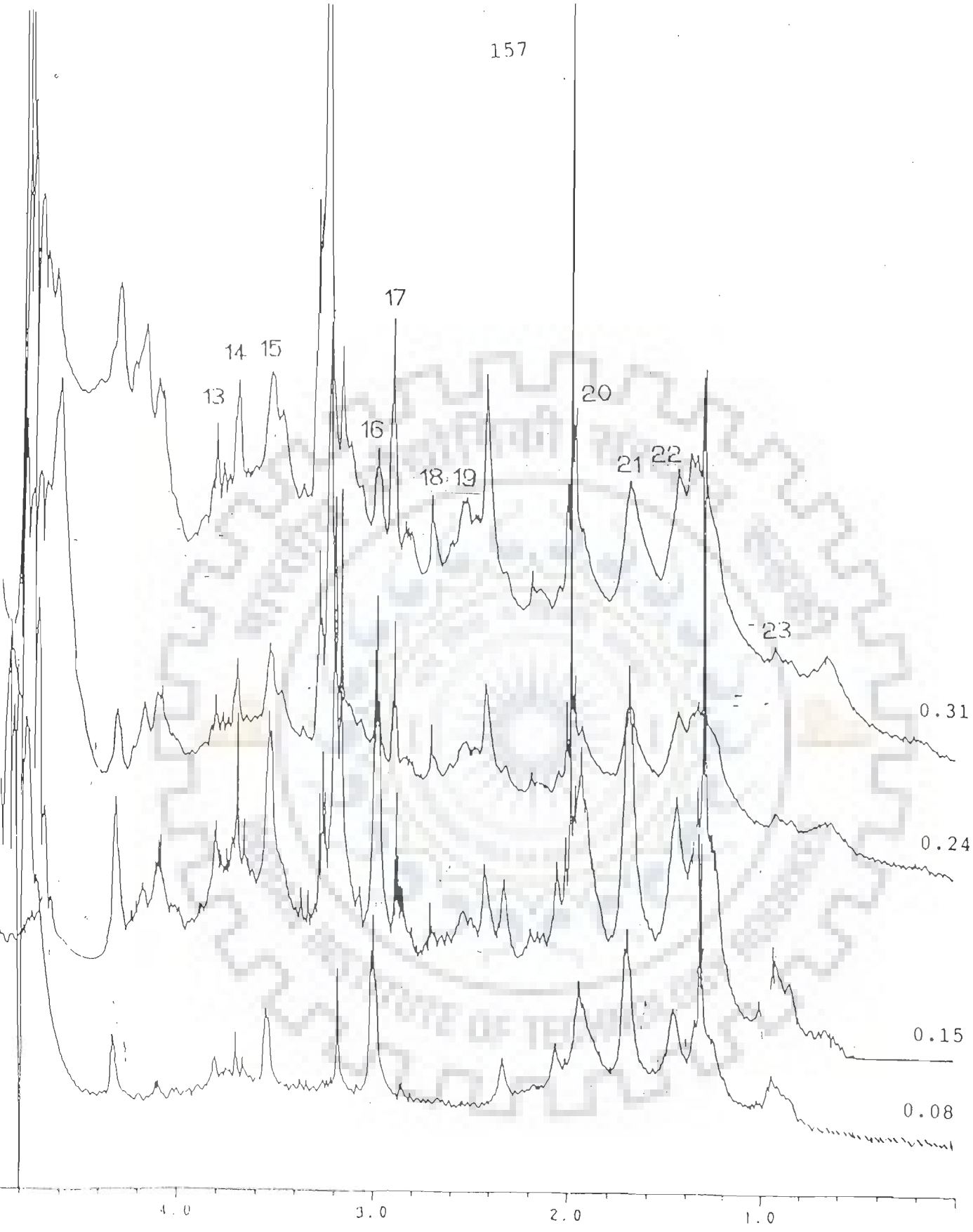
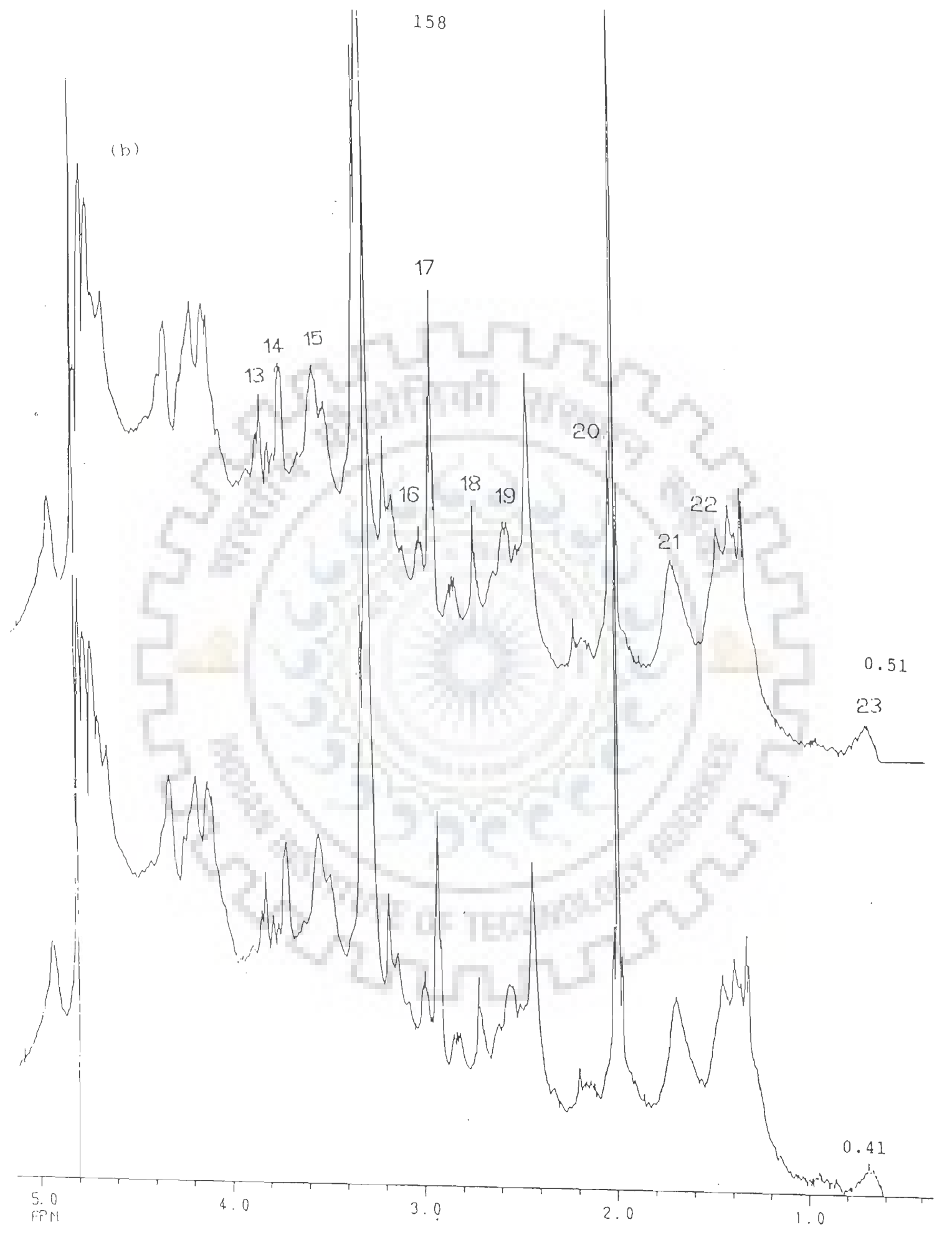


FIG. 3.16(b): Series of one dimensional spectrum recorded at varying concentration ratio (N/P). Peaks are labelled as follows: 13-Ser β'-CH_2, 14-Ser β-CH_2, 15-Pro δ-CH_2, 16-Tyr β-CH_2, 17-Lys ε-CH_2, 18-Asn β-CH_2, 19-Pro β-CH_2, 20 -Lys β-CH_2, Lys δ-CH_2, Pro γ-CH_2, 21-Leu β-CH_2, Leu γ-CH_2, 22-Lys γ-CH_2, 23-Leu δ-CH_3


158
(b)
13
14
15
16
17
18
19
20
21
22
23
0.51
0.41
5.0
PPM
4.0
3.0
2.0
1.0

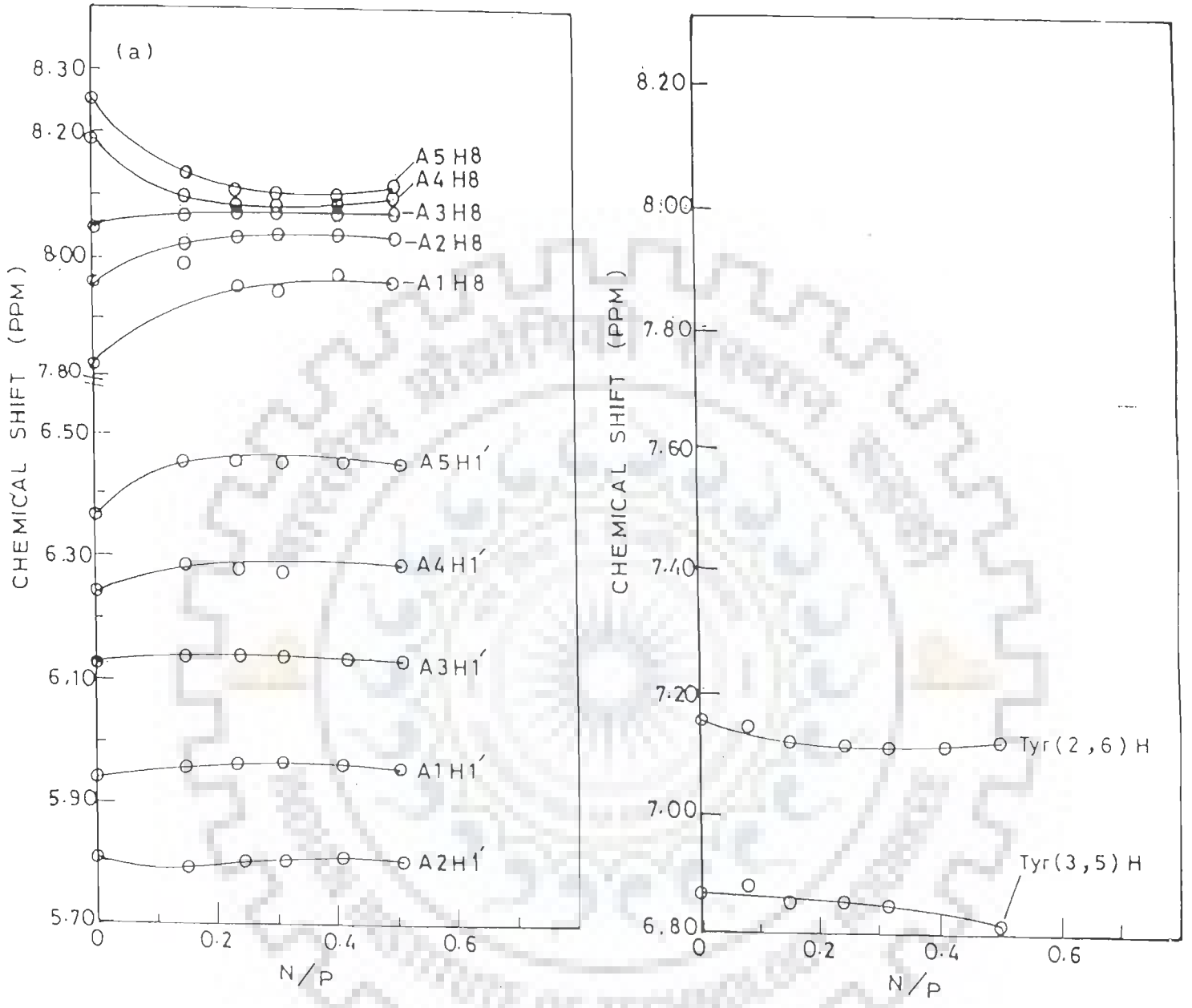
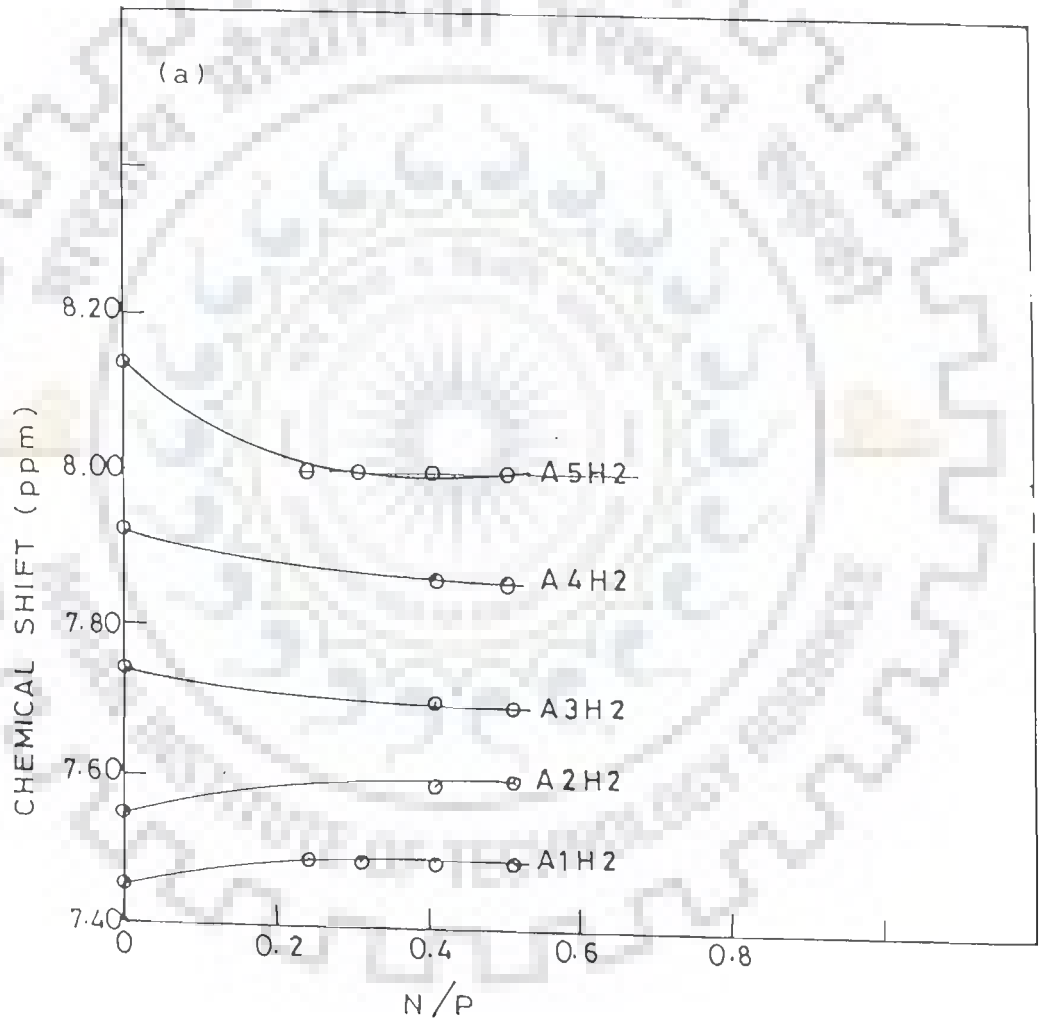
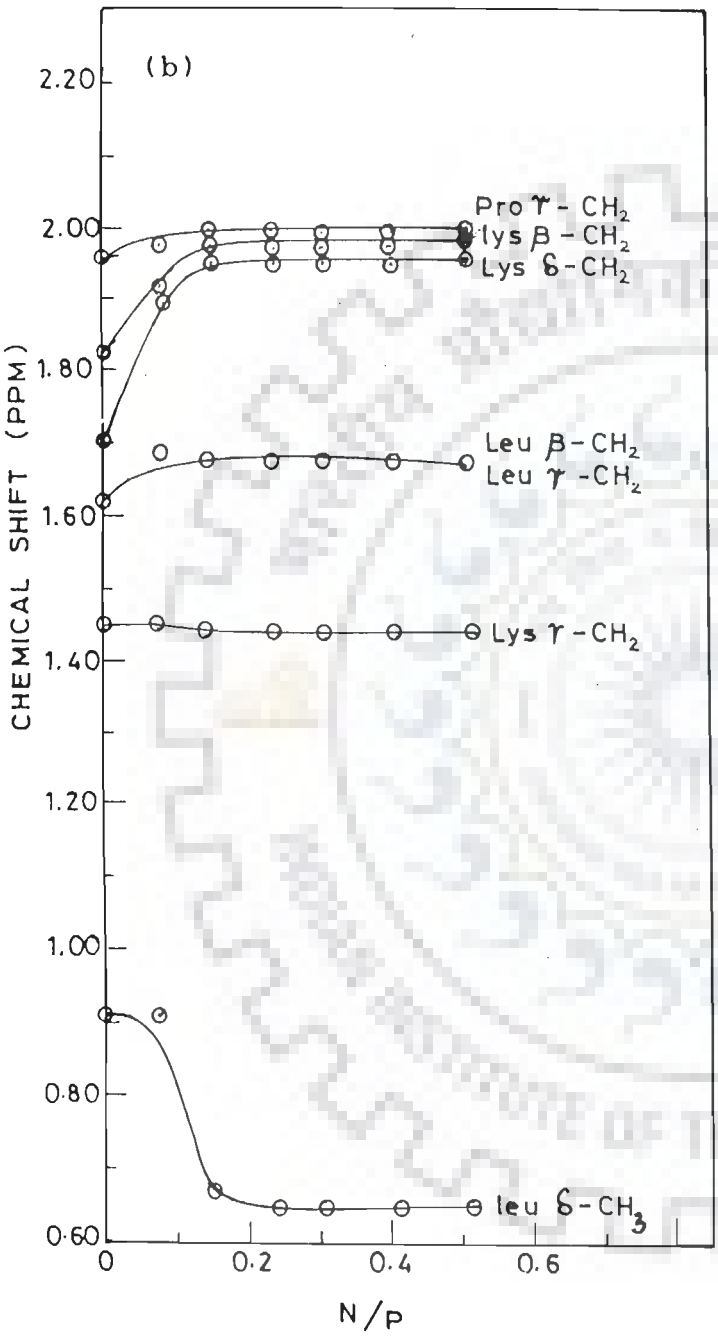
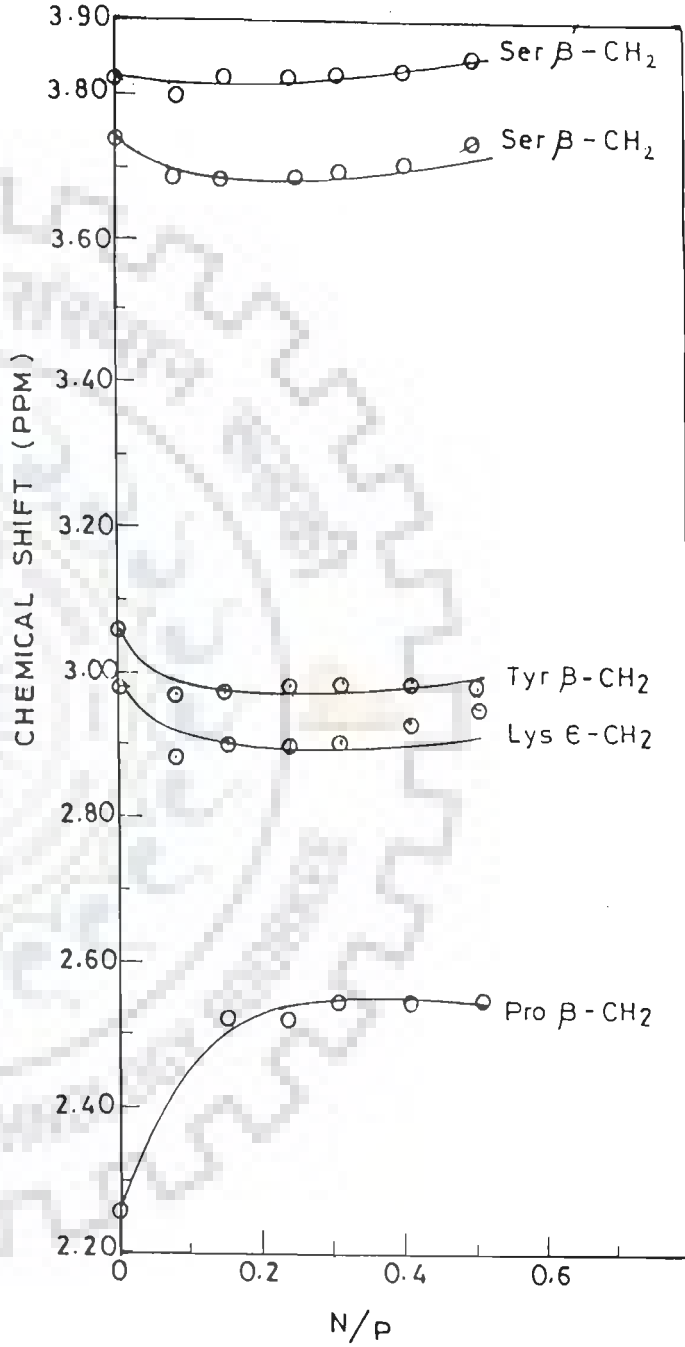


FIG. 3.17(a & b): Chemical shift of various peptide and nucleotide protons as a function of concentration ratio (N/P)


(a)
8.20
8.00
7.80
7.60
7.40
CHEMICAL SHIFT (ppm)
A5H2
A4H2
A3H2
A2H2
A1H2
0
0.2
0.4
0.6
0.8
N/P


(b)
CHEMICAL SHIFT (PPM)
2.20
2.00
1.80
1.60
1.40
1.20
1.00
0.80
0.60
0
0.2
0.4
0.6
N/P
Pro γ - CH2
lys β - CH2
Lys δ - CH2
Leu β - CH2
Leu γ - CH2
Lys γ - CH2
leu δ - CH3


CHEMICAL SHIFT (PPM)
3.90
3.80
3.60
3.40
3.20
3.00
2.80
2.60
2.40
2.20
0
0.2
0.4
0.6
N/P
Ser β - CH2
Ser β - CH2
Tyr β - CH2
Lys ε - CH2
Pro β - CH2

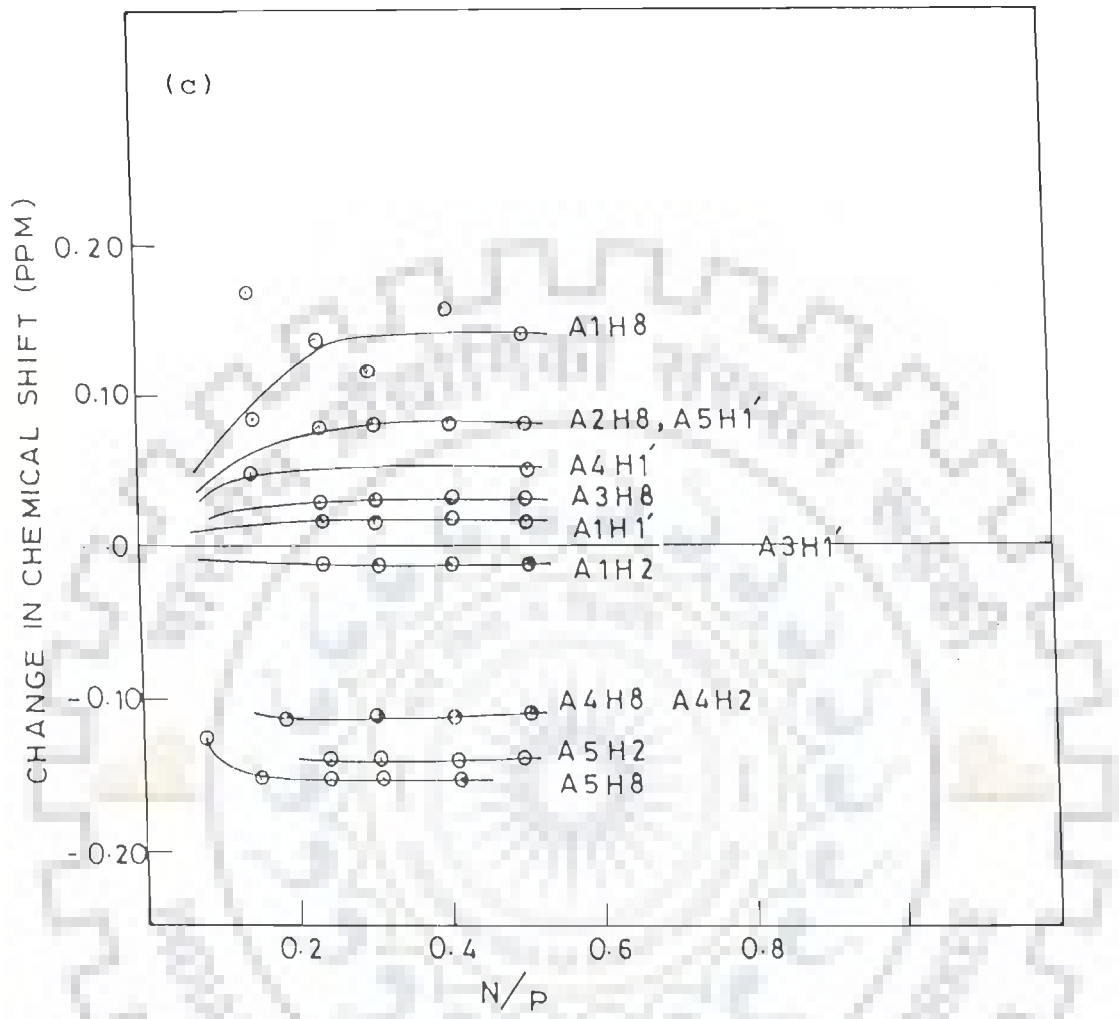


FIG. 3.17(c & d): Change in chemical shift of various peptide and nucleotide protons as a function of concentration ratio (N/P)

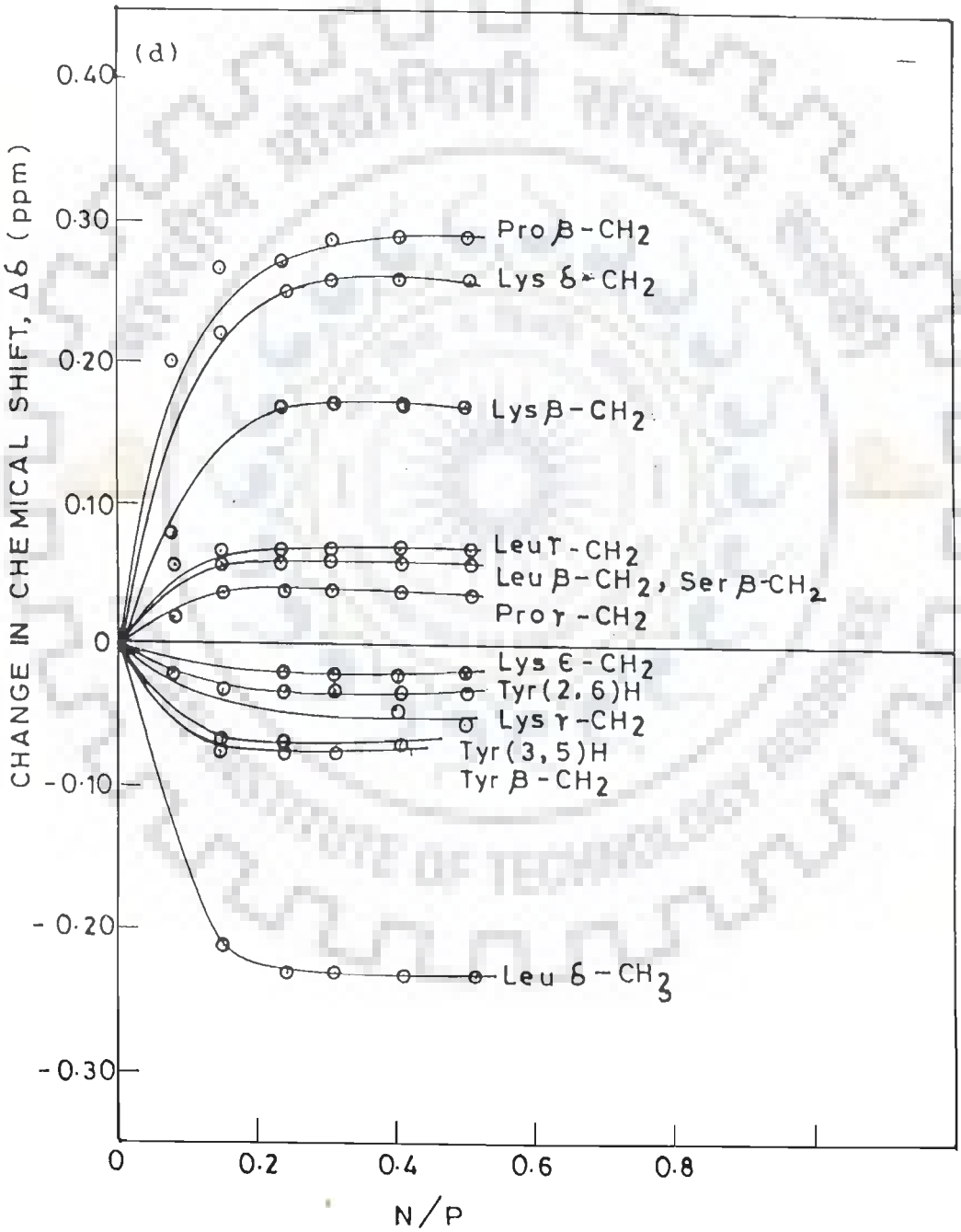

(d)
0.40
0·30
0·20
0·10
0
-0·10
- 0·20
-0·30
CHANGE IN CHEMICAL SHIFT, Δδ (ppm)
Pro β-CH2
Lys δ-CH2
Lys β-CH2
Leu γ-CH2
Leu β-CH2, Ser β-CH2
Pro γ-CH2
Lys ε-CH2
Tyr(2,6)H
Lys γ-CH2
Tyr(3,5)H
Tyr β-CH2
Leu δ-CH3
0
0.2
0.4
0.6
0.8
N/P

Table 3.12a: Change in chemical shift values ($\Delta\delta$) of various protons of $d(A)_5$ in the bound form N/P = 0.51in D_2O at 297 K. Upfield shifts are taken as negative.

Proton	A1	A2	A3	A4	A5
H1'	0.02	-0.01	0.00	0.06	0.08
H2'	-0.01	0.01	-0.03	0.01	0.04
H2"	0.01	0.10	0.09	0.02	0.08
H3'	-	-0.19	-0.24	0.06	0.06
H4'	0.06	0.20	0.16	0.02	0.02
H5'	0.02	0.05	0.03	0.03	0.03
H8	0.14	0.08	0.03	-0.06	-0.12
H2	0.04	0.05	0.04	-0.05	-0.14

Table 3.12b: Change in chemical shift values $\Delta\delta$ (in ppm) of various protons of hexapeptide in a mixture of 1.92 mM $d(A)_5$ and 3.60 mM KPYSLN in phosphate buffer (0.01 M) at 297 K. Upfield shifts are taken as negative.

	Lys	Pro	Tyr	Ser	Leu	Asn
α-CH	0.03	-	-	-	0.03	-
β-CH_2	0.17	0.29	-	-	0.07	-
γ-CH_2	-0.03	0.00	-	-	0.07	-
δ-CH_2	0.26	0.01	-	-	-	-
δ-CH_3	-	-	-	-	-0.26	-
ε-CH_2	0.02	-	-	-	-	-
(3,5)H	-	-	-0.02			
(2,6)H			-0.03			

Table 3.13(a): Chemical shift, δ (in ppm) of peptide and nucleotide protons in a mixture of $d(A)_5$ to KPYSLN at varying concentration ratios (N/P).

N/P	A1H8	A2H8	A3H8	A4H8	A5H8
0*	7.82	7.95	8.04	8.18	8.25
0.08	-	-	-	-	-
0.15	7.99	8.03	8.07	-	8.10
0.24	7.96	8.03	8.07	8.08	8.10
0.31	7.94	8.03	8.07	8.08	8.10
0.41	7.98	8.03	8.07	8.08	-
0.51	7.96	8.03	8.07	8.12	8.13

N/P	A1H2	A2H2	A3H2	A4H2	A5H2
0	7.45	7.55	7.74	7.92	8.14
0.08	-	-	-	-	-
0.15	-	-	-	7.82	-
0.24	7.49	-	-	7.82	8.00
0.31	7.49	-	-	7.82	8.00
0.41	7.49	7.59	7.70	7.87	8.00
0.51	7.49	7.60	7.70	7.87	8.00

N/P	A1H1'	A2H1'	A3H1'	A4H1'	A5H1'
0	5.94	5.81	6.13	6.24	6.37
0.08	-	-	-	-	-
0.15	5.96	5.79	6.13	6.29	6.46
0.24	5.96	5.80	6.13	6.27	6.45
0.31	5.96	5.80	6.13	6.27	6.45
0.41	5.96	5.80	6.13	-	6.45
0.51	5.96	5.80	6.13	6.29	6.45

N/P	Leu δ-CH_3	Lysγ-CH_2	Leuγ-CH_2	Leuβ-CH_2	Lysδ-CH_2
0*	0.91	1.47	1.62	1.61	1.70
0.08	0.91	1.45	1.69	1.68	1.90
0.15	0.67	1.44	1.68	1.68	1.96
0.24	0.65	1.44	1.68	1.68	1.96
0.31	0.65	1.44	1.68	1.68	1.96
0.41	0.65	1.44	1.68	1.68	1.96
0.51	0.65	1.44	1.68	1.68	1.96

N/P	Lysβ-CH_2	Proβ-CH_2	Proγ-CH_2	Lysε-CH_2	Proδ-CH_2
0*	1.83	2.26	1.96	2.98	3.56
0.08	1.92	-	1.98	2.98	3.56
0.15	1.98	2.53	2.00	2.99	3.53
0.24	1.98	2.53	2.00	2.99	3.53
0.31	1.98	2.55	2.00	2.99	3.53
0.41	1.98	2.55	2.00	2.99	3.53
0.51	2.00	2.55	2.00	2.99	3.53

N/P	Serβ-CH_2	Serβ'-CH_2	Tyrβ-CH_2	Tyr(3,5)H	Tyr(2,6)H
0*	3.75	3.83	3.06	6.87	7.15
0.08	3.69	3.80	2.98	6.88	7.12
0.15	3.69	3.83	2.99	6.85	7.12
0.24	3.69	3.83	2.99	6.85	7.12
0.31	3.70	3.83	2.98	-	-
0.41	3.71	3.83	2.99	6.82	-
0.51	3.74	3.85	2.99	-	-

* 5 mM peptide alone

Table 3.13 (b): Change in chemical shift, $\Delta\delta$ (in ppm) of peptide protons in a mixture of $d(A)_5$ and KPYSLN at varying N/P concentration ratios. Upfield shifts are taken as negative.

N/P	A1H8	A2H8	A3H8	A4H8	A5H8
0.08	-	-	-	-	-
0.15	0.17	0.08	0.03	-	-0.15
0.24	0.14	0.08	0.03	-0.10	-0.15
0.31	0.12	0.08	0.03	-0.10	-0.15
0.41	0.16	0.08	0.03	-0.10	-0.15
0.51	0.14	0.08	0.03	-0.06	-0.12

N/P	A1H2	A2H2	A3H2	A4H2	A5H2
0.08	-	-	-	-	-
0.15	-	-	-	-0.10	-
0.24	0.04	-	-	-0.10	-0.14
0.31	0.04	-	-	-0.10	-0.14
0.41	0.04	0.04	0.04	-0.05	-0.14
0.51	0.04	0.05	0.04	-0.05	-0.14

N/P	A1H1'	A2H1'	A3H1'	A4H1'	A5H1'
0.08	-	-	-	-	-
0.15	0.02	-0.02	0.00	0.05	0.09
0.24	0.02	-0.01	0.00	0.03	0.08
0.31	0.02	-0.01	0.00	0.03	0.08
0.41	0.02	-0.01	0.00	-	0.08
0.51	0.02	-0.01	0.00	0.05	0.08

N/P	Leu δ-CH_3	Lysγ-CH_2	Leuγ-CH_2	Leuβ-CH_2	Lysδ-CH_2
0.08	0	-0.02	0.08	0.06	0.20
0.15	-0.22	-0.03	0.07	0.06	0.22
0.24	-0.24	-0.03	0.07	0.06	0.25
0.31	-0.24	-0.03	0.07	0.06	0.26
0.41	-0.24	-0.03	0.07	0.06	0.26
0.51	-0.24	-0.03	0.07	0.06	0.26

N/P	Lysβ-CH_2	Proβ-CH_2	Proγ-CH_2	Lysε-CH_2	Proδ-CH_2
0.08	0.09	-	0.02	0.00	0.03
0.15	0.15	0.27	0.04	0.00	0.03
0.24	0.15	0.27	0.04	0.01	0.03
0.31	0.15	0.29	0.04	0.01	0.03
0.41	0.15	0.29	0.04	0.01	0.03
0.51	0.17	0.29	0.04	0.01	0.03

N/P	Serβ-CH_2	Serβ'-CH_2	Tyrβ-CH_2	Tyr(3,5)H	Tyr(2,6)H
0.08	0.06	0.03	-0.08	0.00	0.00
0.15	0.06	0.00	-0.07	0.01	-0.03
0.24	0.06	0.00	-0.07	-0.02	-0.03
0.31	0.05	0.00	-0.08	-0.02	-0.03
0.41	0.04	0.00	-0.07	-	-
0.51	0.01	0.02	-0.07	-0.05	-

serine protons seem to be least/not affected on binding.

Complexation of hexapeptide with $d(A)_5$ causes more line broadening than might be expected from a consideration of relative rotational correlation times based on molecular weight (85). A careful look at the 1D NMR spectra of a mixture of $d(A)_5$ with KPYSLN (Fig.3.16(a and b)) reveals that Leu δ-CH_3 broadens remarkably. The sharp doublet resonating at 0.91 ppm changes to a broad signal at ~ 0.65 ppm (peak no. 23 in Fig.3.16(b)). Similarly Tyr (3,5)H and (2,6)H protons resonating at 6.87 and 7.15 ppm, respectively (Fig.3.1b) become increasingly broad as N/P increases (peak no. 11 and 12 in Fig.3.16(a)) and merge with baseline at N/P ~ 0.51. Further, Tyr β-CH_2, Leu γ-CH_2 and Leu β-CH_2(peak nos. 16 and 21 in Fig.3.16(b)) also get significantly broadened on binding. The line broadening in Pro δ-CH_2 (peak no. 15 in Fig.3.16(b)) due to binding is to lesser extent. On the other hand, several other proton resonance of hexapeptide, particularly those of Asn, Ser residues remain as very sharp resonances in the complex. The implications of the changes in chemical shift and the line broadening are discussed in the next section.

F) Structure of $d(A)_5$ - KPYSLN complex

The observed experimental results show specific changes in peptide and $d(A)_5$ protons on complexation indicating therby that a well defined structural complex is formed. We observe following intermolecular NOEs in the NOESY spectra of the complex (Fig.3.15(c)):

#1 Pro δ-CH_2 - A1H2'

#2 Leu α-CH - A2H2'

Under the resolution used, NOE is not expected to be seen for distances > 3.5 Å so that the above referred intermolecular NOE connectivities pertain to a distance in the range ~ 2.5-3.5 Å which is used in deriving a structure of complex by Dreiding stereo model.

For a long time, it has been believed that Tyr 26 residue of Gene V protein binds to ssDNA through stacking interactions and considerable evidence has been shown to this effect in literature (2-5,51,53,85,86,88) on the basis of X-ray crystallography, fluorescence and NMR investigations. Considering the hypothesis that tyrosine ring of hexapeptide interacts with $d(A)_5$ through stacking interactions, two of the five adenine bases of $d(A)_5$ get completely unstacked and aromatic ring of tyrosine intercalates between them. This has been shown by Brayer and McPherson (21) in the X-ray crystal structure of complex of Gene V protein with $d(A)_5$. In this case, three adenine bases remain unaffected and hence stacked as in the uncomplexed form. The adenine base protons of two adenine residues within which the Tyr ring intercalates are expected to show change in chemical shift as a result of (a) destacking of adjacent adenine bases (induces downfield shift) and (b) stacking of adenine ring with intercalating ring (induces upfield shift). The net result may be upfield or downfield shift depending upon the size and position of intercalating ring. The atomic diamagnetic susceptibility anisotropy due to ring currents (118) have been utilised to check this hypothesis. Fig.3.18(a) shows the overlap geometry when Tyr ring is intercalated between two adenine rings moved apart to a distance of 6.8 Å. If tyrosine ring stacks completely, it may be placed in a typical position occupied earlier by adenine. It is well known that relative stacking among different bases decreases in the order: purine-purine > purine-pyrimidine > pyrimidine-pyrimidine. Since Tyr ring may be considered like a pyrimidine base, the stacking of adenine-adenine bases in the uncomplexed $d(A)_5$ is expected to be more effective than that of adenine-Tyr ring as in $d(A)_5$-KPYSLN complex. Therefore the stacking is effectively reduced. This is expected to result in a small downfield shift due to net stacking. Such net downfield shift in base protons ($\sim$ 0.02 ppm) have been observed on interaction of Tyr containing tripeptides with Poly (A) as well as double stranded d-GCGC, d-CpG and d-CCGG (13-16). On the other hand, the drugs/dyes intercalating between base-pairs

often consist of 3 or more conjugated aromatic rings and cause more stacking with a purine than that observed for purine-purine stacks. In such a case net upfield shift is observed in base H6/H8 protons (17). We observe net downfield shift in H2 and H8 protons of A1, A2, A3 residues and net upfield shift in A4, A5 residues of $d(A)_5$. This cannot be explained solely on the basis of stacking interaction. Further H1' proton experiences ring current in a similar way as H8 resonance being close to aromatic ring of the base. We observe upfield shift in H8 resonance of A4 and A5 residues whereas the corresponding H1' resonance shift downfield (Table 3.13b). This also cannot be explained solely on the basis of stacking interaction. King and Coleman (85) observe changes in H2, H8 and H1' protons on complexation of Gene V protein to $d(A)_4$, $d(pA)_4$ and $d(pA)_8$, etc. that are similar to our results but do not offer a clear interpretation of the same.

It may be noted that we do not observe any significant change in conformation of $d(A)_5$ on complexation. Earlier studies on intercalation of actinomycin, proflavin, ethidium bromide, etc. in the double-stranded DNA have shown (113) that the sugar conformation at the point of intercalation changes to C3' endo 3'—>5' C2' endo. However later it has been shown that change in sugar conformation as C3' endo 3'—>5' C2' endo is not an essential feature of intercalation (153-156). In fact theoretical modelling as well as X-ray crystallographic studies on intercalation of large rings of daunomycin and adriamycin in d-CGATCG, d-TGTACA, etc. (153-155), the deoxyribose remains in major S-conformational state with glycosidic bond rotation in anti or high anti conformational state. It is shown that by changing the ε backbone angle in DNA (defined as C4'-C3'-O3'-P) from near trans value of -169° to a near gauche value of -104° in the C5 residue, it is possible to separate the neighbouring C5-G6 bases to 6.8 Å distance in daunomycin- d-CGTACG complex (153).

However the more crucial test for the hypothesis of Tyr ring

intercalating between two adenine bases, as suggested in X-ray crystallographic studies by Brayer and Mcpherson (21) is the change in chemical shift of ring protons of tyrosine. From Fig.3.18 (a), it is clear that Tyr ring protons are expected to show upfield shifts ~ 0.2 to 0.4 ppm depending upon exact position of ring due to magnetic anisotropy of neighbouring adenine bases. We do not observe such large upfield shifts in Tyr(3,5)H and Tyr(2,6)H on complexation of hexapeptide with $d(A)_5$. Upfield shift ~ 0.15 to 0.30 ppm have been observed by us on binding of Lys-Tyr-Lys to d-GCGC, d-CGCG, d-CCGG, d-CpG, etc. (13-15). The corresponding upfield shifts in aromatic protons of drugs/dyes on intercalation are observed to be ~ 0.2 to 0.5 ppm (17,114) perhaps due to greater extent of stacking in these cases. Thus, from the chemical shift change in adenine and Tyr ring protons, it is very unlikely that Tyr ring intercalates between adenine rings of $d(A)_5$.

Considering the second hypothesis in which Tyr may partially stack by coming close to adenine bases of $d(A)_5$, as schematically shown in Fig.3.18(b), we find that such a geometry will result in small changes in Tyr ring protons, with practically no change in conformation of $d(A)_5$ as observed. However it still cannot explain the increased extent of stacking in A4, A5 residues on complexation resulting in net large upfield shift ~ -0.14 ppm in H2 and H8 resonances as well as net downfield shift in H1' protons of the same A4, A5 residues (Table 3.13b). This clearly shows that interaction other than the stacking interactions may be involved in complex formation. This may also explain a close proximity of Leu α-CH with A2H2' and Pro δ-CH_2 with A1H2'. Such a proximity of peptide residues may also result in large change in chemical shift of A2H3', A2H4', A1H2", A3H3', A3H4' and A3H2" protons as observed (Table 3.12(a)). Such a possibility has recently been suggested by King and Coleman (86) and is discussed later.

It may be noted that conformational analysis of d-ApA using

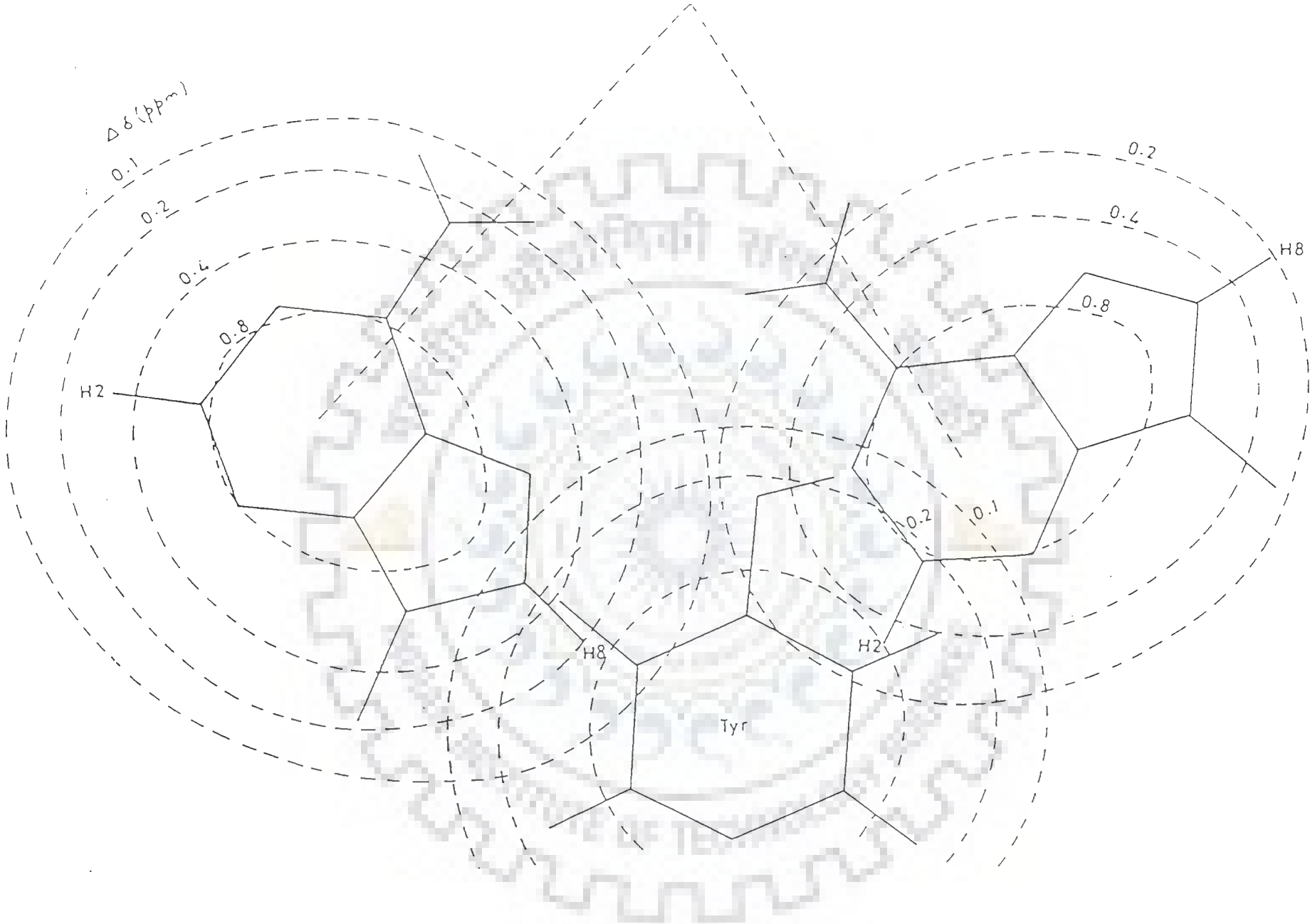


FIG. 3.18 (a): Proposed model of overlap geometry, when Tyr ring is fully stacked between two adenine bases.

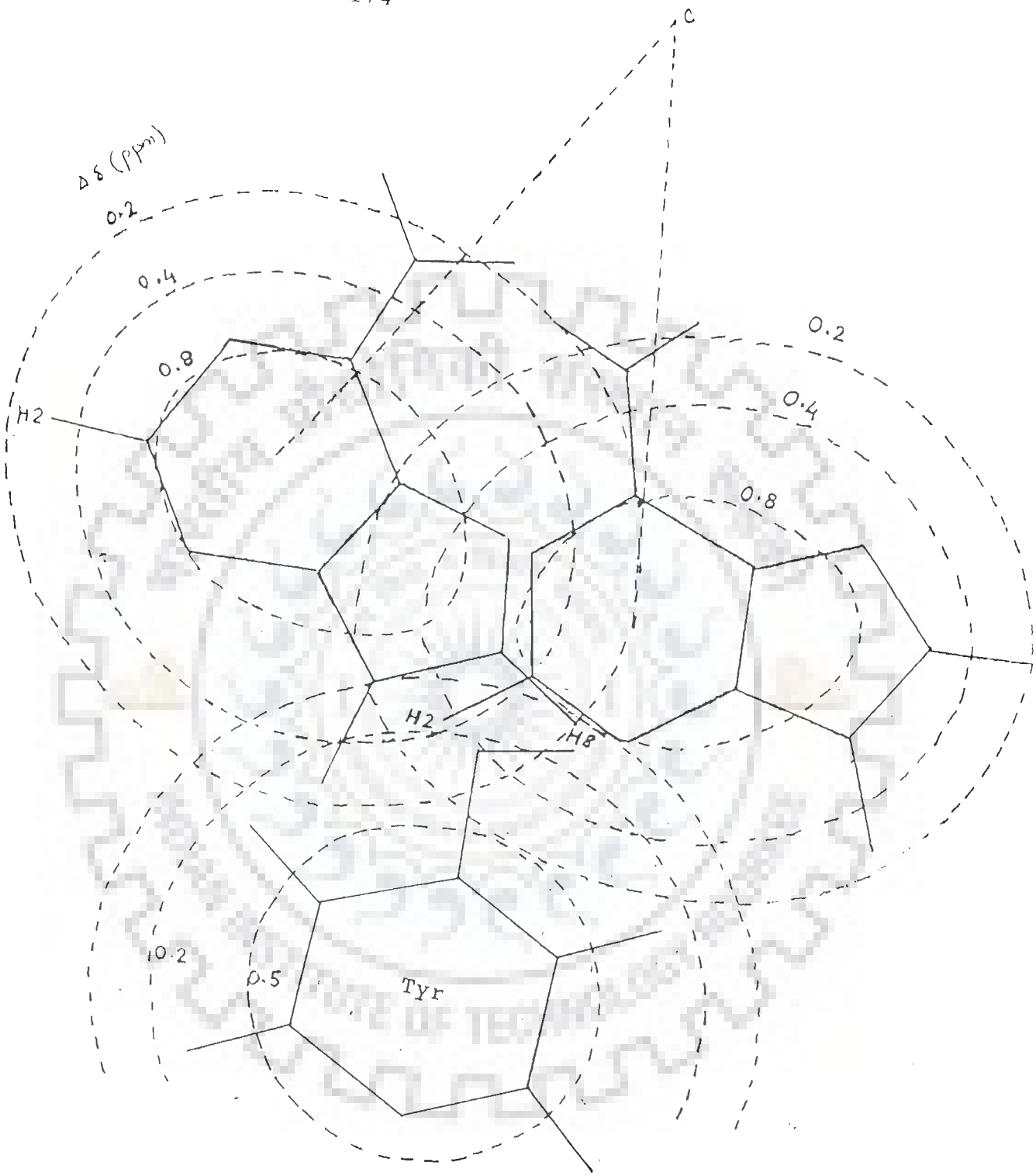


FIG. 3.18 (b): Proposed model of overlap geometry, when Tyr ring is in proximity of adenine bases.

J-couplings in NMR spectra (108) as well as potential energy minimisation studies on d-ApA (108b) and d-ApApA (108c) have shown that single-stranded dApA or d-ApApA adopt a conformation very close to that of one strand of B-DNA (85). We therefore constructed a Dreiding stereo model of single-stranded $d(A)_5$ using these backbone torsional angle. We positioned the hexapeptide close to $d(A)_5$ in an extended conformation. The side chains of alternate residues face one side of the hexapeptide shown schematically in Fig.3.19. The side chains of Leu and Pro residues are close to the sugar-phosphate backbone of $d(A)_5$ so as to satisfy the observed two intermolecular NOEs. The Lys-NH_3^+ group attached to ε-CH_2 atom of the side chain can come close to the phosphate group of $d(A)_5$. The alternate residues Lys, Tyr and Leu comprising of hydrophobic -CH_2/-CH_3 groups and aromatic ring are close to the hydrophobic region of bases of $d(A)_5$ with Tyr-OH group pointing away from base. Ser and Asn are towards the sugar-phosphate backbone. In the Drieding stereo model so formed (Fig.3.20) it is observed that it is possible for Leu side chain and Tyr aromatic ring to be positioned in the hydrophobic core of $d(A)_5$ as though the hydrophobic interactions are the way of anchoring the hexapeptide to $d(A)_5$. This could explain the dramatic line broadening observed in the case of Leu δ-CH_3 and Tyr ring proton resonances (Fig.3.16(a and b)) as though they are completely immobilised on binding to $d(A)_5$.

It has earlier been suggested that Tyr 26 of Gene V protein stacks with bases of ssDNA (85). It was observed that Tyr (2,6)H and Tyr (3,5)H experience upfield shift of -0.11 to -0.18 ppm and -0.13 to -0.23 ppm, respectively on interaction with $d(pA)_4$, $d(A)_4$ and $d(pA)_8$ (85). An intermolecular cross peak corresponding to NOE connectivity AH1'-Tyr 26 (3,5)H was also seen (85). Our results are in sharp contrast to earlier NMR results (2,4,85,86). We neither get large upfield shift nor any NOE connectivity involving Tyr protons. On the other hand we get remarkable broadening of Tyr ring protons and Tyr β-CH_2 indicating immobilisation of this

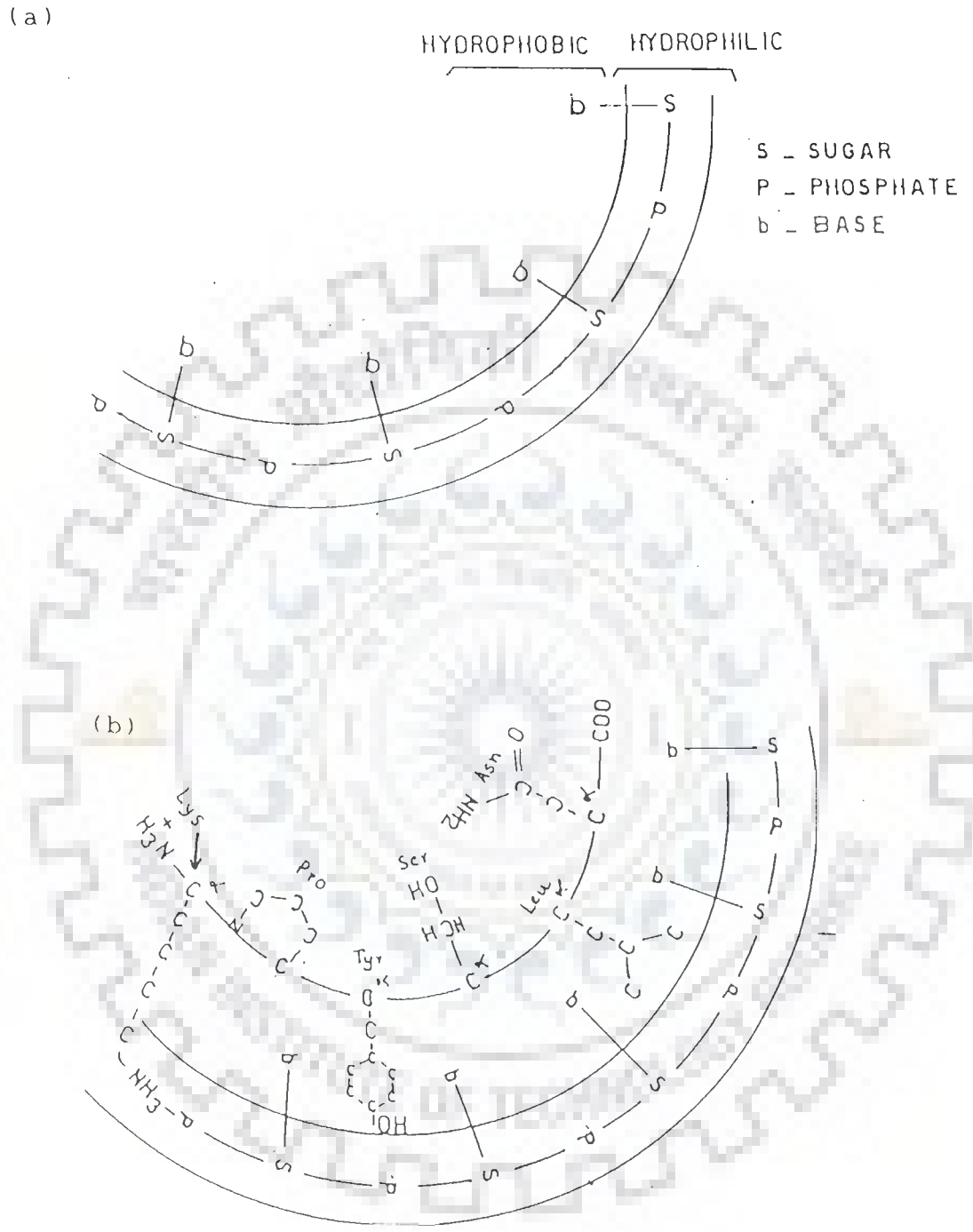


Fig. 3.19:(a) Schematic representation of the arrangement of sugar-phosphate backbone with bases stacked towards helix axis (b) Schematic representation of the arrangement of KPYSLN stretched on sugar phosphate backbone of DNA.

residue on interaction.

It is observed in the model that Tyr ring can be positioned close to the hydrophobic environment of bases without destacking the bases. This may explain the large shift in Tyr β-CH_2 resonance alongwith broadening of Tyr β-CH_2, Tyr (3,5)H and Tyr (2,6)H protons. This is in accord with results published by the research group of Prof. C.W. Hilbers (49-53,86). To facilitate the discussion we briefly describe the structure of Gene V protein here. Folkers et al. (51) explored the DNA binding domain of Gene V protein encoded by bacteriophage M13 with the aid of spin-labelled oligonucleotides by proton NMR. The nucleoprotein structure is modelled using restrained molecular dynamics, the schematic representation of which by Molscript programme (52) is given in Fig.1.11(a and b, chapter I). Fig.3.21(a) shows schematically the topology of monomer of Gene V protein determined by Skinner et al. (157) by X-ray diffraction method. The Gene V protein monomer consists predominantly of a distorted five stranded β-barrel from which two β-hairpins and a large loop protrude (using the nomenclature of Brayer and McPherson (22)). The first hairpin comprising residues 15 to 32 is the DNA binding loop. The amino acid sequence 16-28 is conserved best when closely and distantly related ssDNA binding proteins are compared (138(b)). The large loop is composed of residues 33 to 49 and is termed as complex loop. It is assumed that this region is involved in dimer-dimer interaction i.e. responsible for the cooperative character of binding of Gene V protein to ssDNA. Mutating Tyr 41, which is at the tip of this loop, has a pronounced effect on the solubility of the protein, its tendency to aggregate and cooperativity of binding (50). The second hairpin, the so called, dyad loop runs from residue 61 to 82 and is nearest to the two fold axis of the dimer molecule. This β-ladder is in close contact with the corresponding region of the second monomer and together they constitute the hydrophobic core of the molecule.

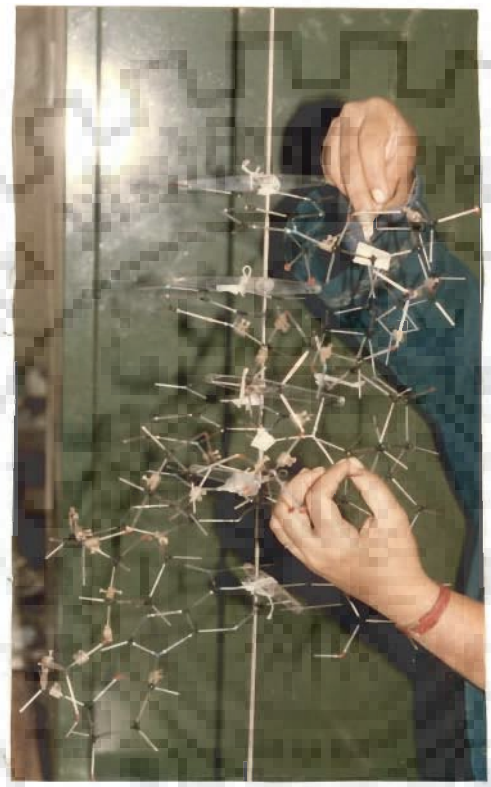


FIG. 3.20: Dreiding stereo model of $d(A)_5$ - hexapeptide complex.

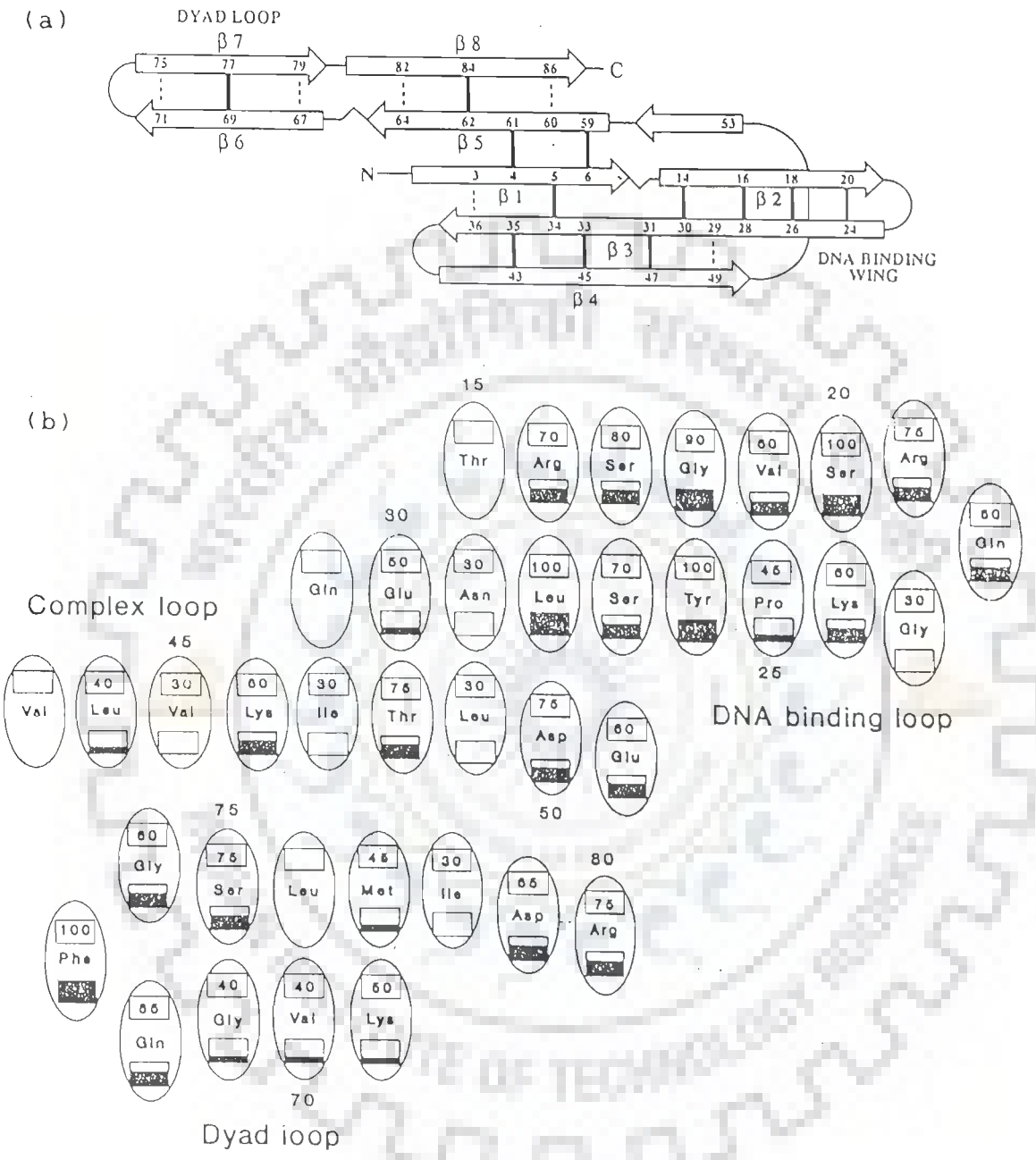


FIG. 3.21(a): Schematic representation of topology of monomer of Gene V Protein (157).

FIG. 3.21(b): The percentage intensity decrease in NOE crosspeaks of GVP-Y41H on binding to spin-labelled $d(A)_3$ (51).

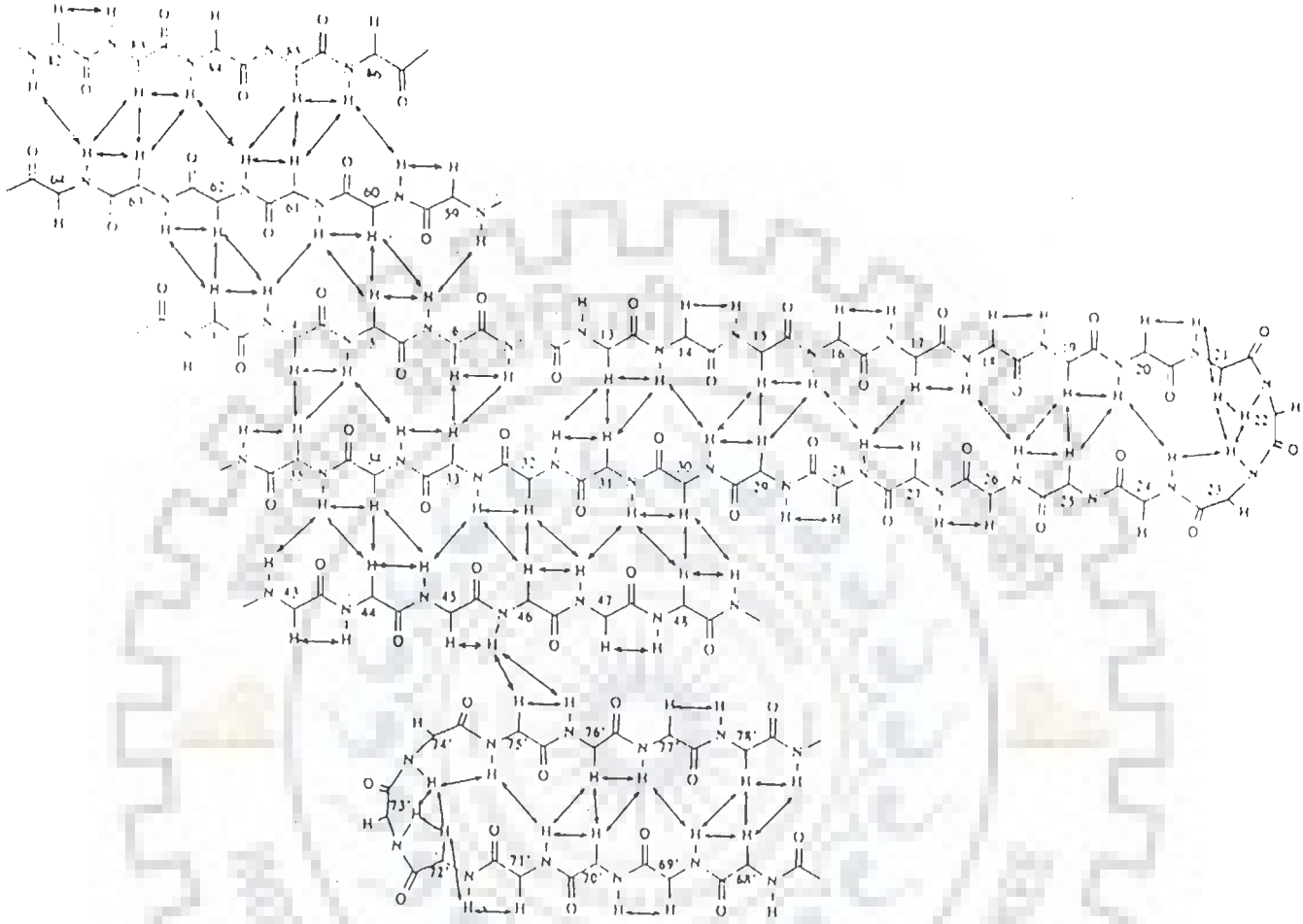


FIG. 3.22: Schematic representation of the complete secondary structure of M13 GVP Y41H (51).

A schematic representation of the complete secondary structure of mutant GVP-Y41H of bacteriophage M13 having same binding domain as the corresponding wild type derived from 2D NMR data by Folkers et al. (51), is shown in Fig.3.22 , .. It may be noted that in the DNA binding domain, β-ladder comprises two sides: one side is formed by the side chains of the residues Ser 17, Val 19, Pro 25 and Ser 27, the other side is formed by the side chains of residues Arg 16, Glu 18, Ser 20, Lys 24, Tyr 26 and Leu 28 (Fig. 3.22 . The percentage intensity decrease in NOE cross peaks of GVP- Y41H on binding to spin-labelled $d(A)_8$ (shown schematically in Fig. 3.21b) is used to estimate the distance of different residues from spin-labelled $d(A)_8$. It is found that the distance between the spin-label and the residues which are apart of the surface constituting residues Arg 16, Gly 18, Ser 20, Lys 24, Tyr 26 and Leu 28 must on the average be shorter than that between the spin-label and the opposite surface. The DNA is positioned on one of the sides of both strands of the DNA binding loop and is not running along one of the strands of that β-loop as the X-ray model suggests (51). It is observed that Ser 20, Tyr 26, Leu 28 and Phe 73 are closest to the DNA spin-label. Skinner et al. (157) have also suggested that Tyr 26, Leu 28 and Phe 73 are expected to be closest to ss DNA which wraps around the outside of protein.

Binding of oligonucleotides to Gene V protein results in broadening of resonances of amino acids which are in close vicinity to the spin-labelled $d(A)_8$ (52). In particular it has been shown that this applies to residues Lys 24, Tyr 26 and Leu 28 (and some other residues). It has also been shown that Tyr 26 is separated from bases by only 2.0 to 4.0 Å (53). But the NMR data shows (85) that complete unstacking of adenine bases does not occur to enable Tyr ring to intercalate as suggested by Brayer and McPherson (21,22). In fact there is a possibility that Tyr 26 may hydrogen bond to the sugar-phosphate backbone in addition to, or instead of stacking (86). Our results support the same to some extent but the mode of binding appears to be hydrophobic

interaction being in the vicinity of bases resulting in complete immobilisation of Tyr ring and Tyr β-CH_2 protons.

Considering the role of Leu 28 in binding of gene V protein to ssDNA, it has been shown that Leu δ-CH_3 protons shift upfield by -0.13 ppm (85) to -0.50 ppm (86) on binding to oligoadenylates and polyadenylates, respectively. The Leu δ-CH_3 protons show NOE connectivity with adenine H2/H8/H1' protons (86). It is proposed that Leu side chain stack on either side of a single base (86). The Leu residue is broadened (52) on binding to spin-labelled $d(A)_3$ and is close to DNA being at a distance ~ 7.0 to 9.0 Å (53). We observe an unusually large broadening of Leu δ-CH_3 (peak no. 23 in Fig.3.16(b)) protons followed by a significant upfield shift of ~0.26 ppm. The Leu β-CH_2 and Leu γ-CH_2 are also shifted upfield and broadened though to a lesser extent than Leu δ-CH_3 protons. Thus our results are in accord with those available in literature. The surface accessibility of Leu residue is observed to be high (52) indicating that their interaction with the hydrophobic nucleotide bases would be energetically favourable.

The long side chains of lysines are known to lend the ε-CH_2 group a considerable amount of motional freedom allowing the hydrophilic NH_3^+ group to interact with solvent. Therefore most of the Lys ε-CH_2 protons are expected to resonate at the position characteristic of a free lysyl residue. Lysines that are involved in intermolecular interaction with other residues appear at a different resonance position as was observed for Lys-41 in the bovine pancreatic trypsin inhibitor which is interacting with a tyrosine (24b).

In a NMR study of role of lysines in the binding of Gene V protein to oligoadenylates (6), it has been shown that out of the two Lys ε-CH_2 resonating at 3.065 ppm and 3.052 ppm, one at 3.065 ppm shifts downfield and broadens beyond detection. The other Lys ε-CH_2 at 3.052 remains a sharp resonance and hence mobile. We do

not observe any significant shift (Table 3.14) or broadening in Lys ε-CH_2 protons. The peak (no. 17 in Fig.3.16) remains narrow and retains its motional freedom in the complex and is to a considerable extent still free to move. It may interact with solvent but is certainly not immobilised. However Lys δ-CH_2 and β-CH_2 experience large downfield shift on complexation. For comparison, typical changes in chemical shift of Lys protons on interaction of tripeptide Lys-Tyr-Lys with different oligo- and polynucleotide systems (14-16) are shown in Table 3.14. It may be noted that the observed shift of 0.17-0.26 ppm in Lys β-CH_2 and Lys δ-CH_2, respectively are significantly higher than those observed in literature (13-16) which have generally been attributed to electrostatic interactions. Thus apparently it seems that lysine residue has a greater role to play in hexapeptide-$d(A)_5$ interaction than mere electrostatic interactions.

It is noteworthy that Gene V protein X-ray structure (21,22) reveals that Tyr 26, Leu 28 and Phe 73 are located in the region of electropositive cluster formed by the side chains of Arg 16, Arg 21, Lys 24 and Lys 46. Series of fluorescence binding studies have shown that protein-protein interactions and cooperativity of the binding is largely ionic-strength independent. On the other hand protein-DNA interactions are salt-dependent. Therefore the changes in spectral parameters of Arg δ-CH_2 and Lys ε-CH_2 protons are attributed to electrostatic interaction (13-16). Further only one or two lysyl residues (positions of lysine residues not exactly identified) show broadening of Lys ε-CH_2 protons (13-16). The spin-label (51) and chemical modification studies (41b) show that Lys 24, Lys 46 and Lys 69 are involved in binding to ssDNA. As such a direct evidence of Lys 24 binding through electrostatic interactions does not exist in literature. With the aid of spin-labelled oligonucleotides $d(A)_3$ (51), the studies on the distance dependence of the spin-label induced relaxation in Gene V protein- $d(A)_3$ complex show that the distance between the

Table 3.14: Chemical shift of lysine protons at 285 K upon complex formation, on binding of Lys-Tyr-Lys to different oligo- and polynucleotide systems as reported in literature (13-16).

System	β-CH_2	γ-CH_2	δ-CH_3	ε-CH_2	α-CH
d-CpG	0.00	0.00	0.00	0.00	0.00
d-GpC	0.02	0.01	0.01	0.00	0.01
d-CCGG	0.01	0.02	0.01	0.01	0.01
d-GCGC	0.02	0.04	0.02	0.03	0.03
d-CGCG	0.03	0.01	0.03	0.03	0.02
Poly(A)	0.07	0.25	0.10	0.01	0.02
Poly(C)	0.07	0.28	0.11	0.01	0.02
$d(A)_5$*	0.17	-0.03	0.26	0.02	-

* Present investigation.

spin-label and the Lys 24 is fairly short. The relative intensity difference in NOESY spectra is about 60 % as compared to 100 % of Tyr 26 and Leu 28 residues (Fig.3.21b). The molecular dynamics simulations show that the distance between Lys 24 residue and phosphorus atom on DNA is ~ 7-9 Å (53). The Lys ε-NH_2 is also seen to have a hydrogen bond with oxygen atom of Arg 21 residue (53). Since we observe very sharp resonances of Lys ε-CH_2 (no broadening) and a significant shift of Lys β-CH_2 and Lys δ-CH_2 resonances, it is suggested that Lys side chain may be present in the hydrophhobic region of $d(A)_5$ but Lys ε-CH_2 is exposed to solvent.

Considering the role of proline residue in binding of hexapeptide to $d(A)_5$, we observe a large downfield shift of Pro β-CH_2 $\cong$ 0.29 ppm, broadening of Pro δ-CH_2 resonance and intermolecular NOE connectivity Pro δ-CH_2-A1H2'. Pro residue is therefore certainly in proximity of $d(A)_5$ in the complex and it is more likely near the hydrophobic region encompassed by the aromatic bases. We have for the first time shown that Pro residue has a definite role in these interactions. So far not a single study by NMR or other methods have shown the possibility of any role of Pro 25 in Gene V protein-ssDNA interaction. However it is important to note that a secondary structure investigation of Gene V protein encoded by the filamentous bacteriophage IKe (138b) by 2D NMR techniques has shown that there is homology between M13 and IKe Gene V protein. Their secondary structure as well as spatial folding is same so that their ssDNA binding properties can be passed on to the protein- ssDNA interaction model (131). It was observed (138a) that the amino acid sequences which make up the turns formed by residues 21-24 and 39-42 in the IKe Gene V protein end with a proline residue. The sequential NOE contacts indicate that these proline residues adopt a trans conformation. Further these Pro residues and their trans conformations are conserved in M13 Gene V protein (49). This indicates that these residues probably have a function in stabilising the structures of the turns in these

regions. It has already been shown that Pro 25 residue of Gene V protein is at a distance of 9 to 12 Å from DNA (53) and can participate in binding to DNA. Our experimental results corroborate these predictions.

As yet attempts at direct visualisation of Gene V protein-DNA complex in solution, has not been very succesful, though lot of research work have been carried out, on this system. Resonance overlap and lack of resonance assignment being the major problem. With these problems, researchers faced difficulty in interpreting the NMR data in terms of specific nucleotide-peptide interactions. To avoid these problems, we carried out the present investigation, on functional part of Gene V protein rather than full Gene V protein.

A comparison of present investigations with the X-ray structure and earlier solution studies suggests that the hexapeptide model simplifies the problem of studying the interaction of Gene V protein with DNA. Each resonance is clearly resolved and the specific role of Lys, Pro, Leu and Tyr residue is established. The Lys-NH_3^+ and Tyr-OH group may be exposed to solvent while rest of hydrophobic chemical groups of Lys and Tyr as well as Leu and Pro residues participate in hydrophobic interaction with the nucleotide bases (52). We do not see any direct evidence of participation of Ser and Asn residue in complex formation although it will be energetically favourable if their side chain could interact with hydrophilic sugar-phosphate backbone. In particular, there is evidence in literature (53) to show that Ser 27 is within 6.5 to 9.0 Å of DNA in the complex.

CHAPTER IV

INTERACTION OF d-(GACTCGTC)$_2$ WITH HEXAPEPTIDE, Lys-Pro-Tyr-Ser-Leu-Asn

The binding of DNA binding loop of Gene V Protein (GVP) comprising of residues, Lys^{24}-Pro-Tyr-Ser-Leu-Asn^{29}, with a double-stranded DNA segment, d-(GACTCGTC)$_2$ have been investigated using 500 MHz proton NMR. The following spectra of octamer system have been investigated: (a) One-dimensional NMR spectra of d-(GACTCGTC)$_2$ as a function of temperature in the range 275-335 K, (b) 1D NMR spectra of the complex of d-(GACTCGTC)$_2$ with hexapeptide at a concentration ratio = 0.08 as a function of temperature, (c) One dimensional NMR spectra of complex of octanucleotide with hexapeptide (KPYSLN) at P/N ratio varying in the range 0 to 0.08 at room temperature and (d) Two - dimensional NMR spectra of a mixture of d-(GACTCGTC)$_2$ and KPYSLN at P/N = 0.08 at room temperature: phase-sensitive COSY and NOESY. Conformation of peptide has already been discussed in previous chapter (Chapter III). In the following we describe spectral assignment of nucleic acid followed by conformational analysis in unbound and bound form, the changes in chemical shift due to binding and hence the structural implications.

A) Spectral Assignment in deoxyoctanucleotide d-(GACTCGTC)$_2$

The bases of d-(GACTCGTC)$_2$ shown schematically in Fig.4.1 have been numbered as follows:

			1	2	3	4	5	6	7	8		
5'	d	-	G	A	C	T	C	G	T	C	3'	
		3'	C	T	G	C	T	C	A	G	-	d 5'
			8	7	6	5	4	3	2	1		

The octamer has a mismatch in the centre involving residue T4 C5 which is discussed later (Section D).

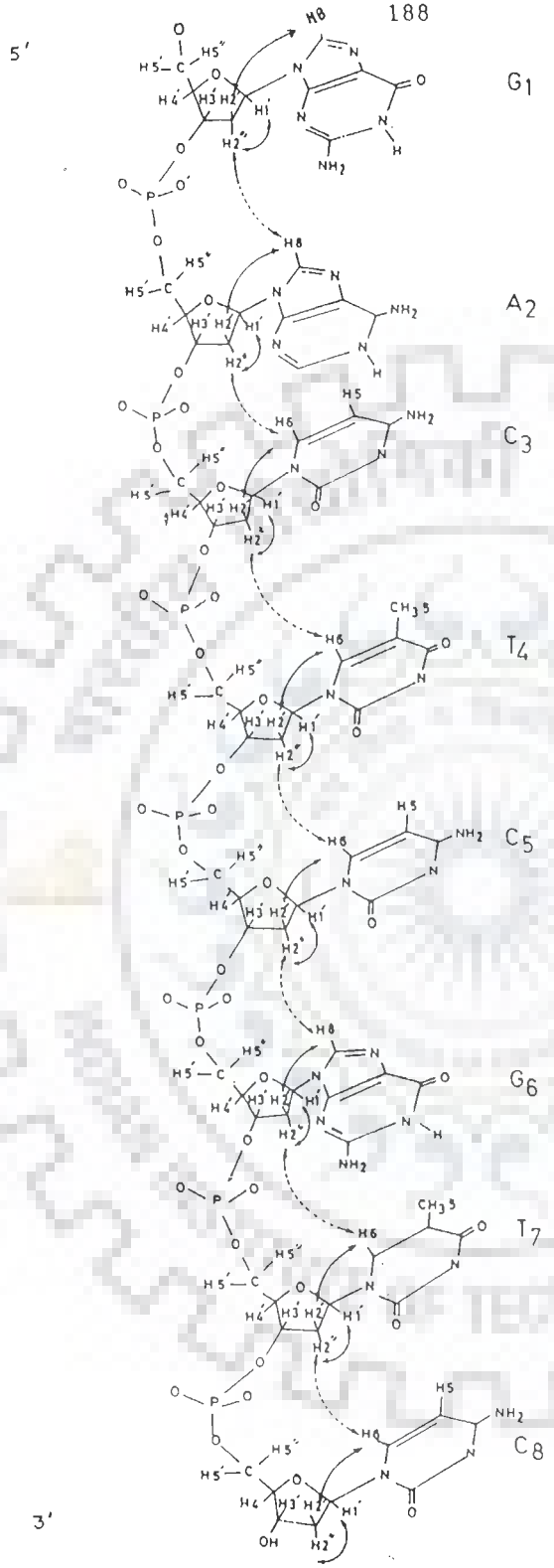


FIG. 4.1: Schematic representation of octamer d-(GACTCGTC)$_2$. The Intra -residue and Inter-residue NOE connectivities are shown as ⟨ ——— ⟩ and ⟨ ----- ⟩, respectively.

Fig.4.2 shows the one-dimensional NMR spectrum of the octamer in D_2O at 297 K at pH = 7.2. The region between 7.2 and 8.3 ppm shows resonances due to the aromatic protons, that is, the AH8, AH2, GH8 appearing as singlets, CH6 resonances as doublets (peaks marked as e,f,g in Fig.4.2(a)) and TH6 as quartets (peak h, quartet is seen in expansion not shown here). The region between 5.4 and 6.4 ppm consists of three doublets of CH5 (peaks q,r,s) and eight triplets of sugar H1' protons of 8 residues of the octamer (peaks i to n, p, o) some of which overlap with each other. The resonances between 4.0 and 5.1 ppm are due to sugar H3', H4', H5' and H5" protons. The sugar H2' and H2" protons resonate between 1.75 and 3.1 ppm. The doublet of methyl group at fifth position of thymine appears around ~ 1.5 ppm. The assignment of a base/H1'/H2'/H2" etc. proton to a specific residue of octamer is made by using results of two-dimensional NMR experiments. The strategies used for assignment are those available in literature for right-handed DNA structures having primarily sugar in S-conformational state (48,63,78,125, also discussed in Section C of Chapter III). The chemical shift of all protons thus ascertained from 1D and 2D NMR experiments are given in Table 4.1a.

The steps involved in spectral assignment from results of 2D NMR experiments are as follows:
Fig.4.3(a) and 4.4(a) show the phase-sensitive COSY and NOESY spectra of 10 mM d-$(GACTCGTC)_2$, respectively at 297 K while the Figs. 4.3(b-i) and 4.4(b-g) are the expansions of certain portions of the corresponding spectra to highlight specific connectivities. The cross peaks in COSY and NOESY spectra are numbered serialy as per scheme given in Table 4.2.

The three doublets of cytosine H6 protons are J-coupled to their corresponding H5 proton resonances and being close to each other in space, show NOE cross peaks (cross peak no. 25 to 27 in Fig.4.4(c)). Therefore, the CH6 protons resonating at 7.75, 7.56 and 7.50 ppm have the corresponding CH5 protons resonating at 5.92, 5.56 and 5.49 ppm, respectively. The geminal coupling

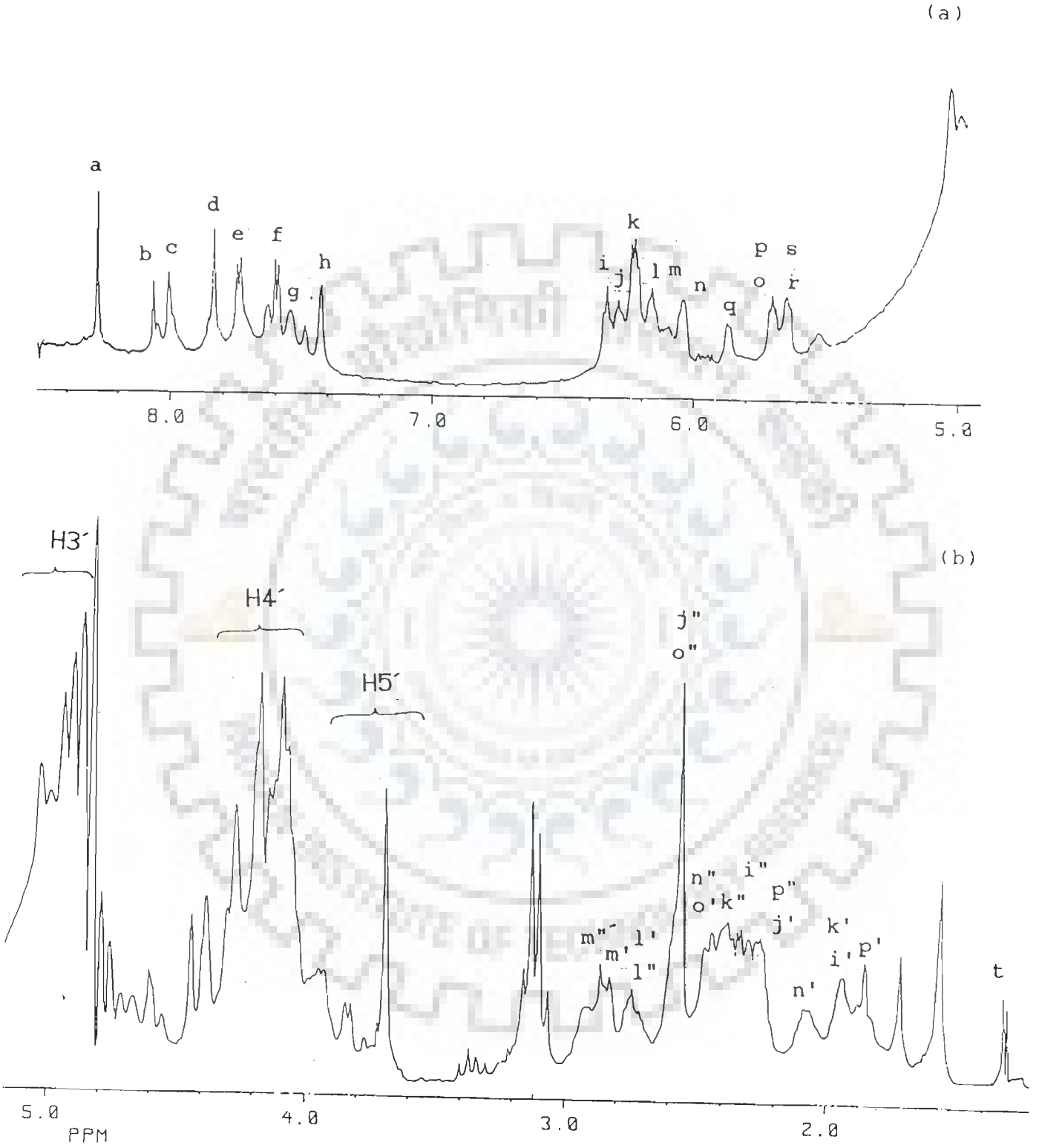


FIG. 4.2: 1D NMR spectra of d-$(GACTCGTC)_2$ at 297 K in (a) 5.0 to 8.3 ppm region and (b) 1.2 to 5.0 ppm region. Peaks are labelled as follows:a-A2H8, b-A2H2, c-G1H8, d-G6H8, e-C5H6, f-C8H6, g-C3H6, h-T4H6, T7H6, i-C5H1', j-G6H1', k-T7H1', l-G1H1', m-A2H1', n-C3H1', o-T4H1', p-C8H1' i'-C5H2', j'-G6H2', k'-T7H2', l'-G1H2', m'-A2H2', n'-C3H2', o'-T4H2', p'-C8H2', i"-C5H2", j"-G6H2", k"-T7H2", l"-G1H2", m"-A2H2", n"-C3H2", o"-T4H2",p"-C8H2", q-C5H5, r-C8H5, s-C3H5, t-$T4CH_a$

Table 4.1a: Chemical shift (in ppm) of various protons of 10 mM d-$(GACTCGTC)_2$ in unbound form at 297 K.

	G1	A2	C3	T4	C5	G6	T7	C8
H8/H6	7.99	8.27	7.50	7.41	7.75	7.84	7.41	7.56
H1'	6.19	6.20	6.04	5.66	6.36	6.32	6.26	5.64
H2'	2.90	2.75	2.04	2.38	2.02	2.24	2.06	1.72
H2"	3.02	2.82	2.42	2.58	2.30	2.59	2.28	2.20
H3'	-	5.06	-	4.60	4.60	4.92	4.60	-
H4'	-	-	-	-	4.14	-	4.14	-
H5'/H5"	-	-	-	-	3.71	-	3.71	-
CH5	-	-	5.49	-	5.92	-	-	5.56
TCH_3	-	-	-	1.50	-	-	1.50	-
H2	-	8.08	-	-	-	-	-	-

Table 4.1 b: Chemical shift of four sets of H3', H4' and H5'/H5" protons of octamer at 297 K from Fig.4.4(g).

	I	II	III	IV
H3'	4.60	4.82	4.66	4.46
H4'	4.42	4.17	4.32	4.20
H5'/H5"	4.16	3.71	4.08	3.92

Table 4.2: Numbering scheme of cross peaks observed in COSY and NOESY spectrum of d-$(GACTCGTC)_2$ (Figs. 4.3, 4.4, 4.6 and 4.7).

1.	G1H1' - G1H2'	31.	T4H2' - T4H3'
2.	G1H1' - G1H2"	32.	C5H2' - C5H3'
3.	A2H1' - A2H2'	33.	G6H2' - G6H3'
4.	A2H1' - A2H2"	34.	T7H2' - T7H3'
5.	C3H1' - C3H2'	35.	C8H2' - C8H3'
6.	C3H1' - C3H2"	36.	G1H3' - G1H4'
7.	T4H1' - T4H2'	37.	A2H3' - A2H4'
8.	T4H1' - T4H2"	38.	C3H3' - C3H4'
9.	C5H1' - C5H2'	39.	T4H3' - T4H4'
10.	C5H1' - C5H2"	40.	C5H3' - C5H4'
11.	G6H1' - G6H2'	41.	G6H3' - G6H4'
12.	G6H1' - G6H2"	42.	T7H3' - T7H4'
13.	T7H1' - T7H2'	43.	C8H3' - C8H4'
14.	T7H1' - T7H2"	44.	G1H8 - G1H2'
15.	C8H1' - C8H2'	45.	G1H8 - G1H2"
16.	C8H1' - C8H2"	46.	A2H8 - A2H2'
17.	G1H2' - G1H2"	47.	A2H8 - A2H2"
18.	A2H2' - A2H2"	48.	C3H6 - C3H2'
19.	C3H2' - C3H2"	49.	C3H6 - C3H2"
20.	T4H2' - T4H2"	50.	T4H6 - T4H2'
21.	C5H2' - C5H2"	51.	T4H6 - T4H2"
22.	G6H2' - G6H2"	52.	C5H6 - C5H2'
23.	T7H2' - T7H2"	53.	C5H6 - C5H2"
24.	C8H2' - C8H2"	54.	G6H8 - G6H2'
25.	C3H6 - C3H5	55.	G6H8 - G6H2"
26.	C5H6 - C5H5	56.	T7H6 - T7H2'
27.	C8H6 - C8H5	57.	T7H6 - T7H2"
28.	G1H2' - G1H3'	58.	C8H6 - C8H2'
29.	A2H2' - A2H3'	59.	C8H6 - C8H2'
30.	C3H2' - C3H3'	60.	C5H2" - C5H3'

between H2' and H2" protons give rise to intense COSY cross peaks in Fig. 4.3(a) in the range 1.75-3.10 ppm for all the eight residues. The expansion of some of H2'-H2" cross peak regions are shown in Fig.4.3(c and d). Since they are close in space the corresponding NOE connectivities are easily seen in Fig.4.4(b) and (g). These 8 pairs of H2'-H2" protons give COSY cross peak patterns with their corresponding H1' protons. Fig.4.3(b) shows five such sets of patterns of H1'-H2' and H1'-H2" (cross peak numbered 3-4, 5-6, 7-8, 11-12 and 13-14) as well as H1'-H2' pattern of one residue (cross peak no. 1) and H1'-H2" patterns of 2 residues (cross peak no. 10 and 16).

Depending upon the pseudorotation angle adopted by the deoxyribose sugar, the H2'/H2" protons are coupled to their corresponding H3' proton. Some of these patterns are easily identified and shown in Fig.4.3 (e, f and g). The H3' proton is not coupled to H4' in the standard B-DNA having sugar pucker in S-conformational state, that is, close to C2' endo geometry. However one set of this coupling is observed as shown in Fig.4.3(h) for which further H4'-H5'/H5" J-coupling is shown in Fig.4.3(i) (Table 4.1b).

The assignment of a specific base and sugar proton to a specific nucleotide residue has been done using following 3 facts:

i) Within a nucleotide residue the base proton H6/H8 is closer to its corresponding H2' proton (distance 2.0-2.3 Å) than its corresponding H2" proton (distance ~ 3.3 Å) (150). Therefore base to H2' NOE cross peak is expected to be much more intense than base to H2" NOE cross peak (often not seen).

ii) The reverse is true for sugar H1' proton. Within a nucleotide residue H1' proton is closer to its corresponding H2" proton (distance ~ 2.1-2.2 Å) than its corresponding H2' proton (distance 2.6-2.8 Å) (150). Therefore H1'-H2" NOE cross peak is expected to be more intense than H1'-H2' NOE cross peak (often seen as weak peaks).

iii) If n is the position of a nucleotide residue in 5'-3' direction of octamer the base $(CH5/TH6/GH8/AH8)_n$ is closer to the sugar $(H2")_{n-1}$ proton of the preceeding residue and is referred to as sequential connectivity. The base proton is also close to $(H1')_{n-1}$ though this distance is greater than that for $(H2")_{n-1}$ proton. Fig.4.4(c), (d), (e) and (f) are expansions of NOESY spectra showing base-H2'/H2" and H1'-H2'/H2" connectivities. The following sequential NOE connectivities are seen in Fig.4.4(c), (d), (e) and (f):

NOE connectivity	marked as
C3H6 - A2H1'	s1
G6H8 - C5H1'	s2
A2H8 - G1H2"	s3
C3H6 - A2H2"	s4
C5H6 - T4H2"	s5
G6H8 - C5H2"	s6
T7H6 - G6H2"	s7
C8H6 - T7H2"	s8

These lead to unambiguous assignment of all base protons and sugar H1', H2', H2" protons. Now using the COSY cross peak patterns in Fig.4.3(e), (f), (g) and (h), the H3' protons resonating at 5.06, 4.92 and 4.60 ppm get assigned to A2, G6 and T4/C5/T7 residues, respectively. From Fig.4.3(h) and (i), the H4', H5'/H5" protons resonating at 4.14 and 3.71 ppm get assigned to C5/T7 residues. From the NOESY spectra shown in Fig.4.4(g), four sets of H3', H4' and H5'/H5" protons are identified as shown in Table 4.1b. One of these i.e. set I is assigned to T4 residue as its H3' proton resonates at 4.60 ppm as ascertained from COSY pattern in Fig.4.3(f). Other 3 sets could not be assigned to specific residues due to noise in the region where H2'/H2"-H3' NOE cross peaks are expected to be seen.

B) Conformation of octamer in unbound state

Conformation of deoxyribose sugar

The strategy used for determining conformation of deoxyribose sugar is already discussed in Chapter III (section C). It is observed that the distance between outer peaks of H1' triplet in

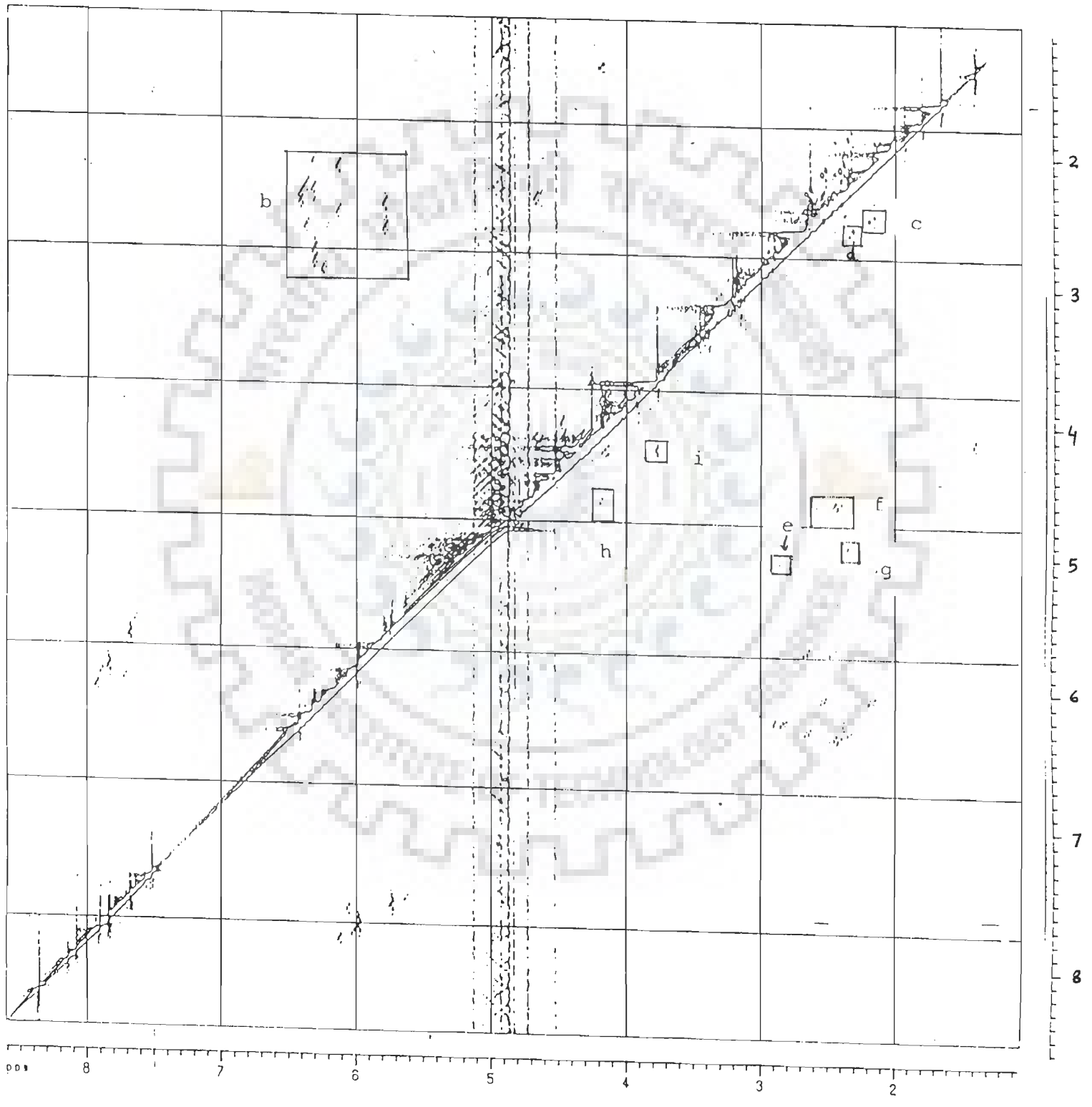


FIG. 4.3(a): Phase-sensitive COSY of 10 mM d-$(GACTCGTC)_2$ at 297 K expansions of certain regions to highlight specific connectivities are given in Fig.4.3(b) to 4.3(i).

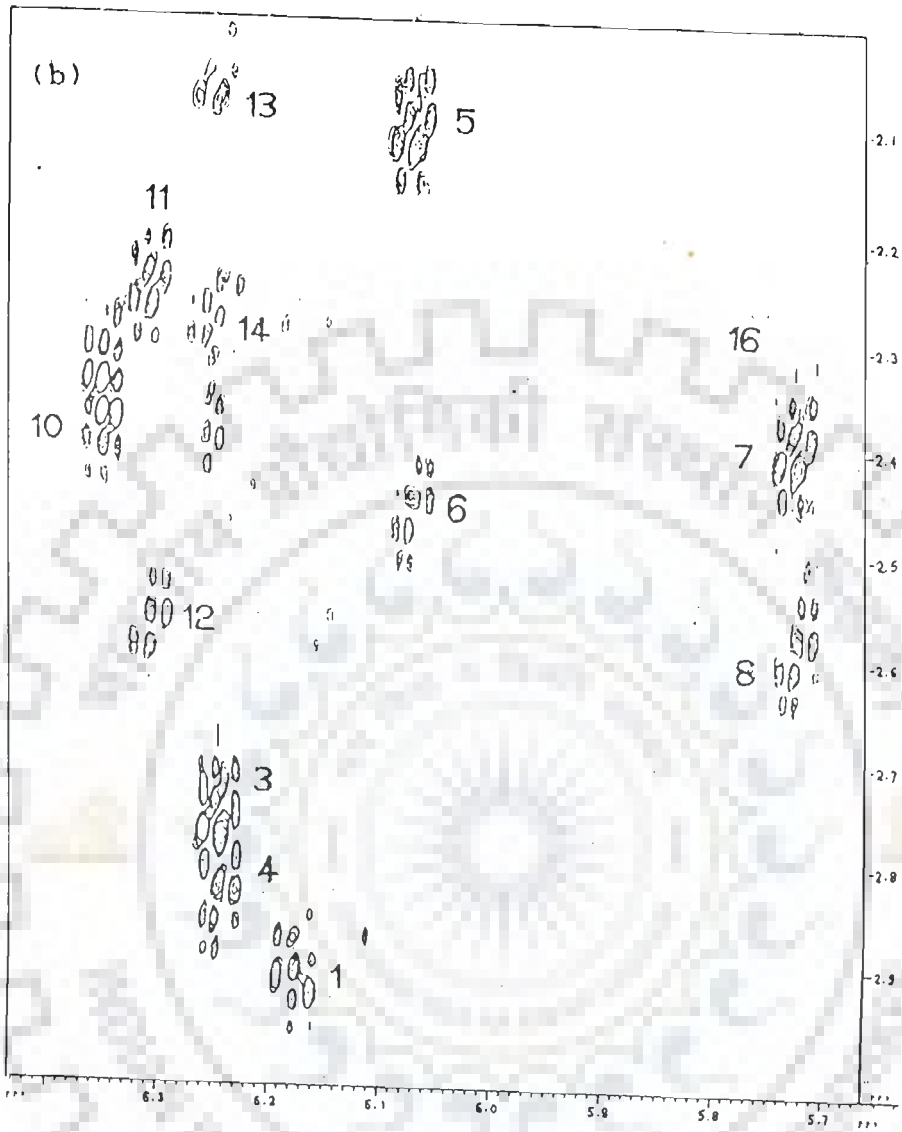


FIG. 4.3(b): H1'-H2' and H1'-H2" COSY cross peak patterns of A2 (no. 3, 4) C3(no. 13,14) , T4 (no. 7,8), G6 (no.11,12) and T7 (no. 13, 14) residues of octamer. H1'-H2' cross peak of G1 (no. 1) and H1'-H2" cross peaks of C5 (no 10) and C8 (no. 16) are also shown.

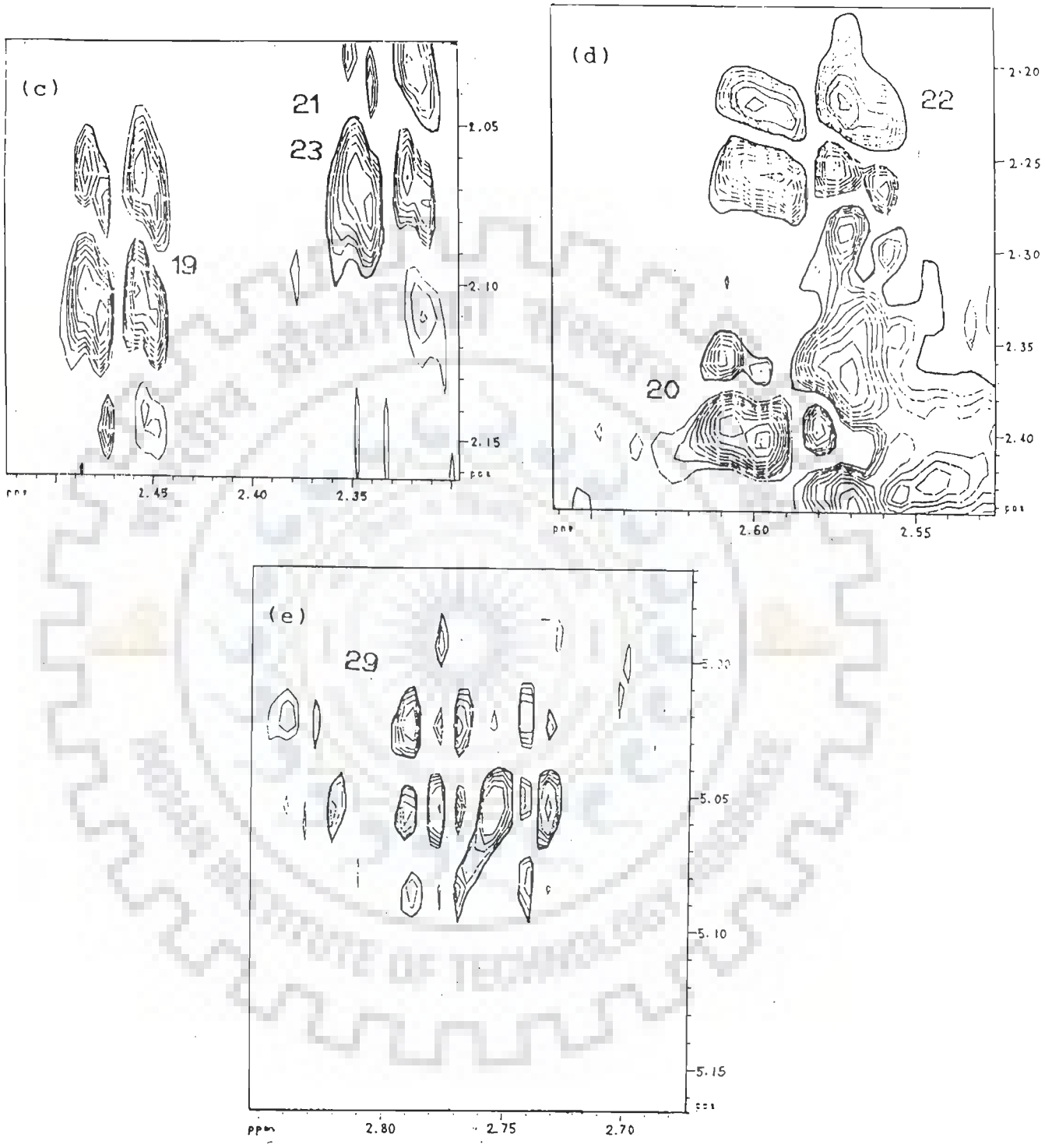


Fig 4.3(c): H2'-H2" COSY cross peak pattern of C3 (no. 19), C5 (no. 21) T4 (no. 20), G6 (no. 22) and T7 (no. 23) residues of octamer. H2'-H3' COSY cross peak pattern of A2 (no. 29) residue of octamer is also shown.

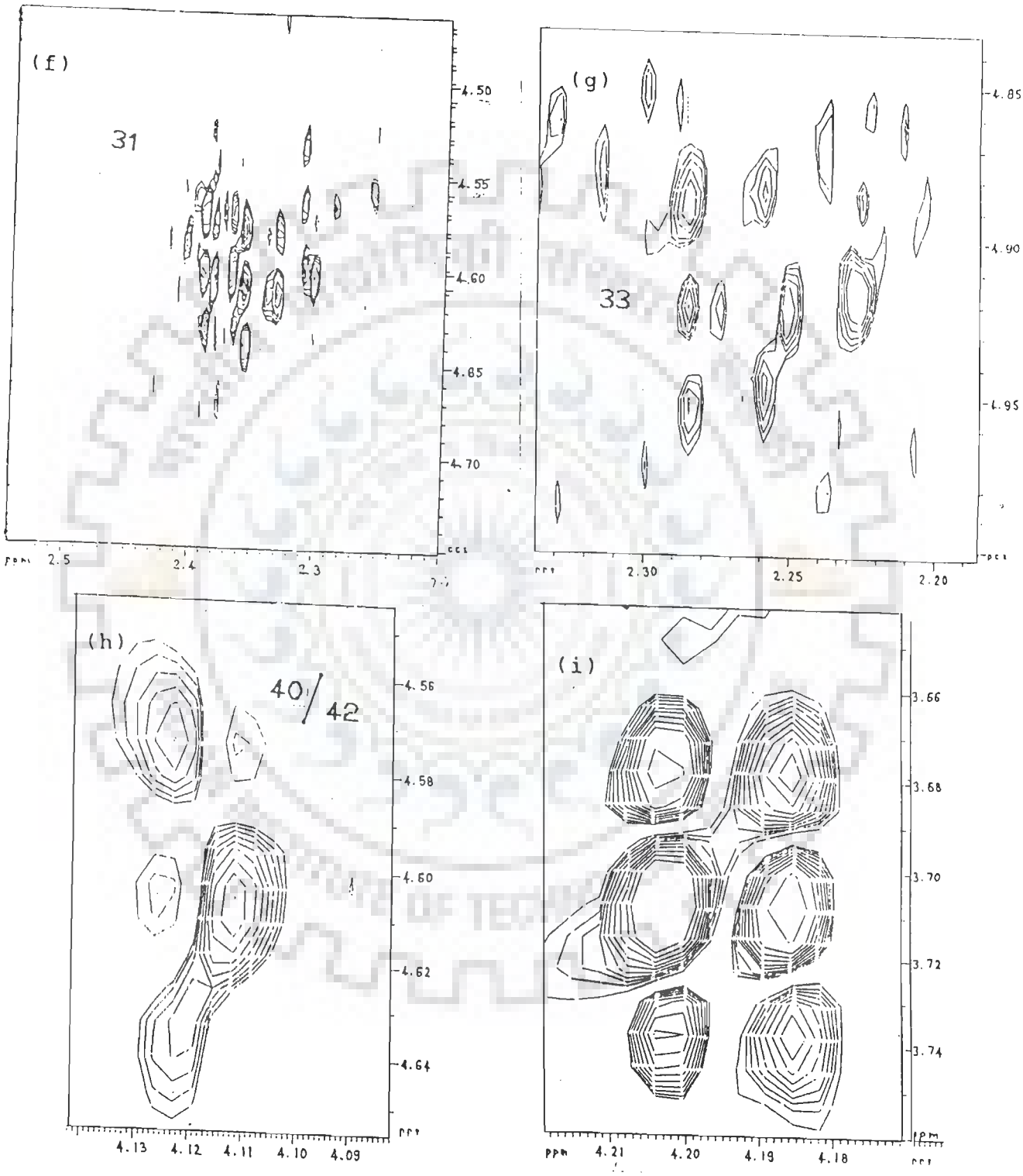


FIG. 4.3(f): H2'-H3' COSY cross peak pattern of T4 (no. 31) residue of octamer. Also seen here is a less intense cross peak pattern of H2"-H3' COSY cross peak of C5 and T7 residues (g):H2'-H3' COSY cross peak pattern of G6 (no. 33) (h): H3'-H4' COSY cross peak pattern of C5/T7 (no. 40/42) residues (i): H4'-H5' COSY cross peak pattern of C5/T7 residue of octamer.

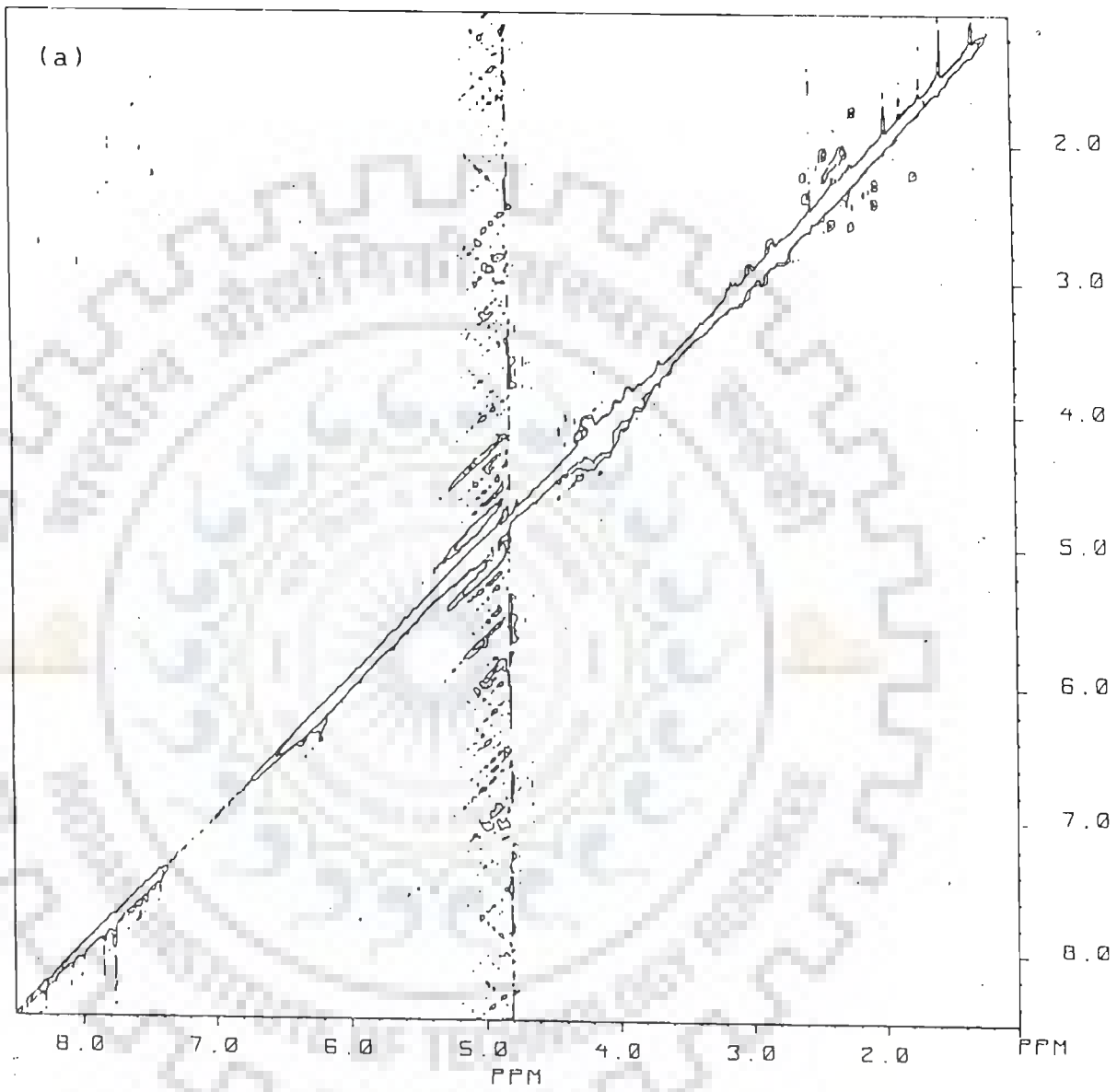


FIG. 4.4(a): Phase-sensitive NOESY of d-$(GACTCGTC)_2$ at 297 K. Expansion of certain regions to highlight specific connectivities are given in Fig.4.4(b to Fig.4.4(g).

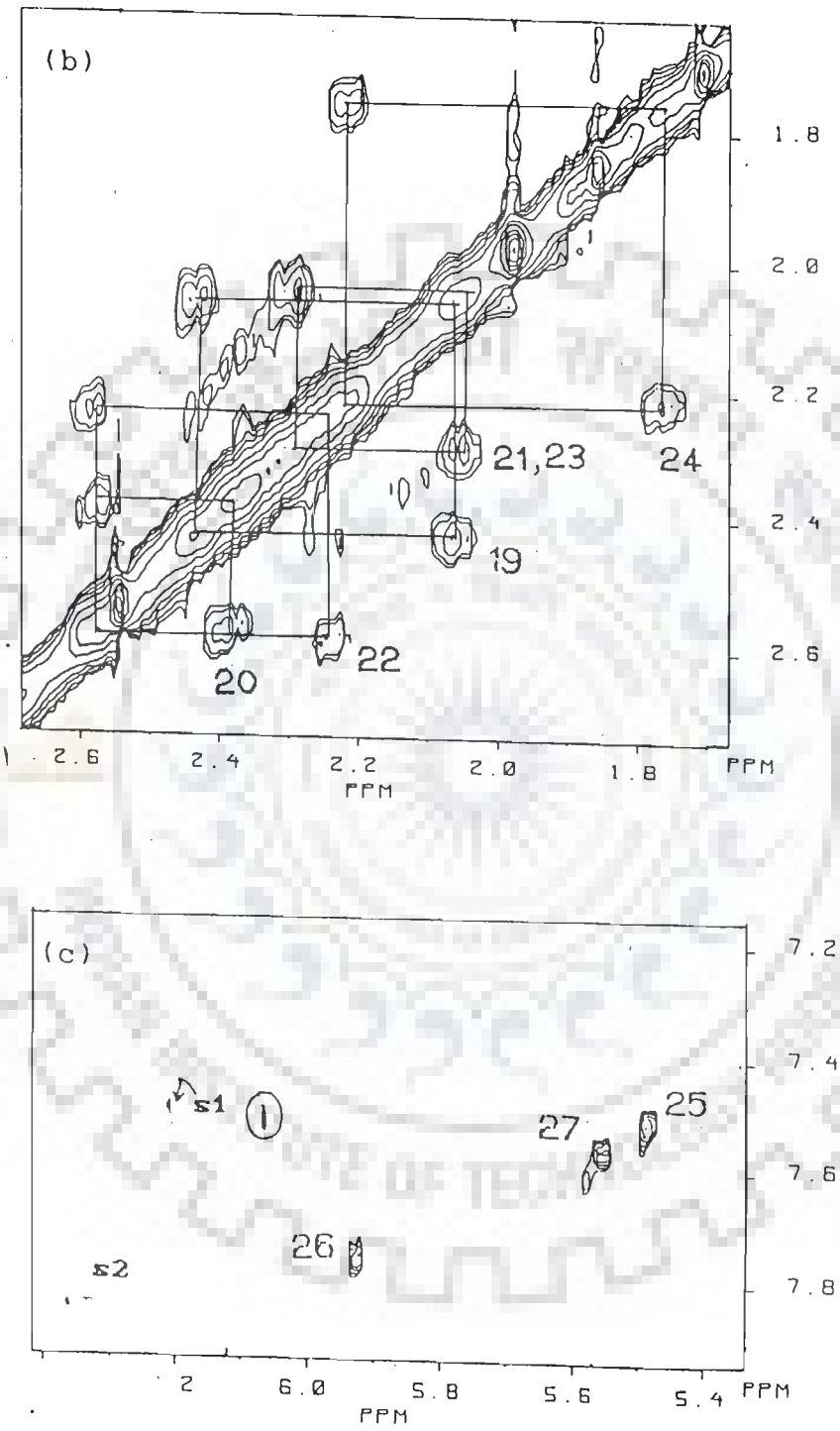


FIG. 4.4(b): NOESY cross peak of H2'-H2" protons of C3 (no.19), T4 (no.20), C5 (no.21), G6 (no.22), T7 (no.23) and C8 (no.24) residues of octamer. (c): NOE cross peak of cytosine H6 - H5 protons of C3 (no. 25), C5 (no. 26) and C8 (no.27) residues of octamer. Also seen here are intra residue weakly inten-se NOE connectivity of C3H6-C3H1' protons (encircled). The sequential NOE connectivities C3H6-A2H1' (marked as s1) and G6H8-C5H1' (marked as s2) are also seen.

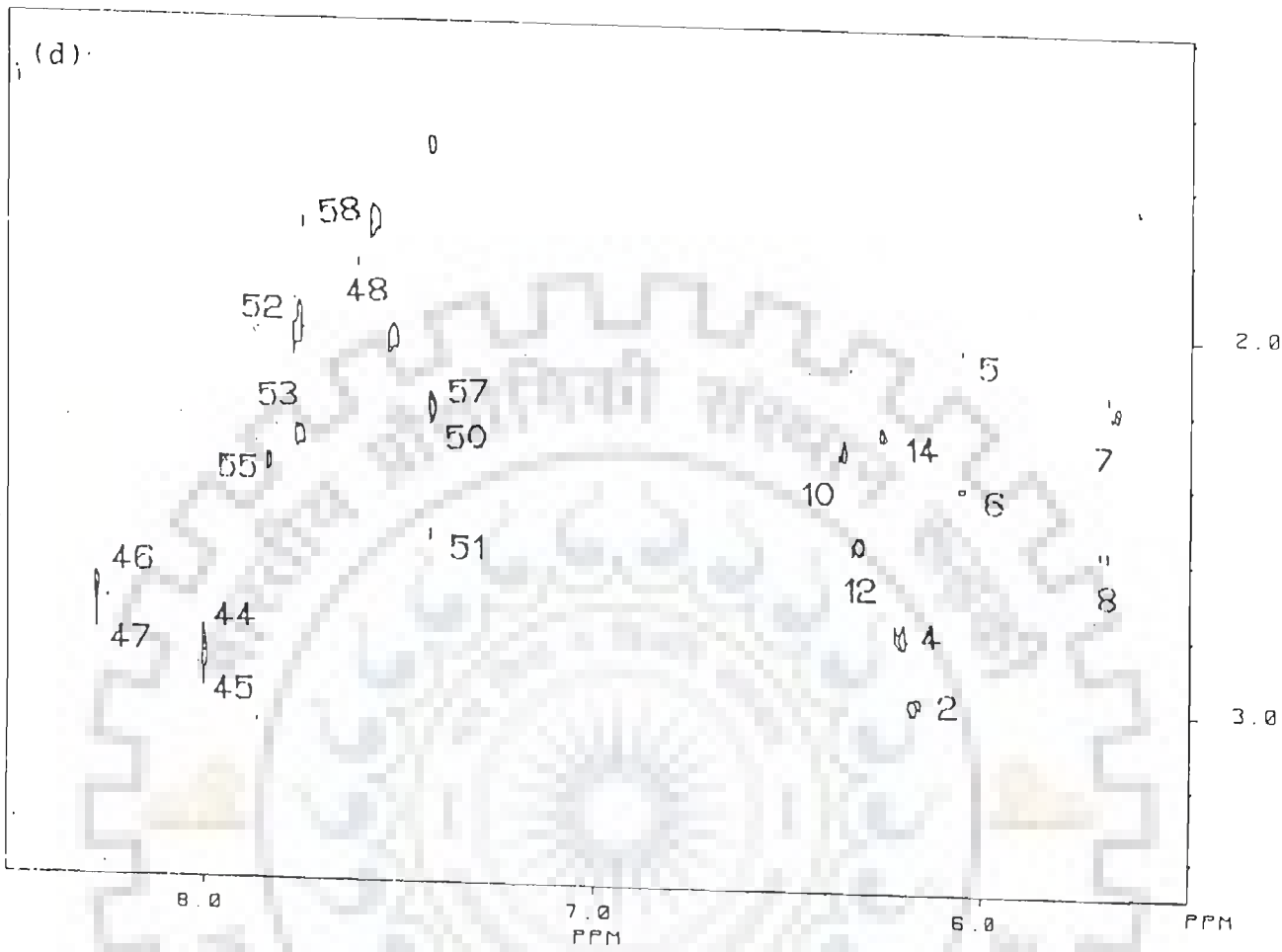


FIG. 4.4(d): NOE cross peak of intra residue base to H2'/H2" protons and H1' to H2'/H2" protons and sequential cross peak of base $(H6/H8)_n$ to $(H2")_{n-1}$ protons. The sequential peaks are marked on s3 to s8 and others are numbered according to scheme given in Table 4.2. Also seen here is the NOE cross peak of CH_3 protons of thymine residues resonating at ~ 1.50 ppm with their corresponding H6 protons.

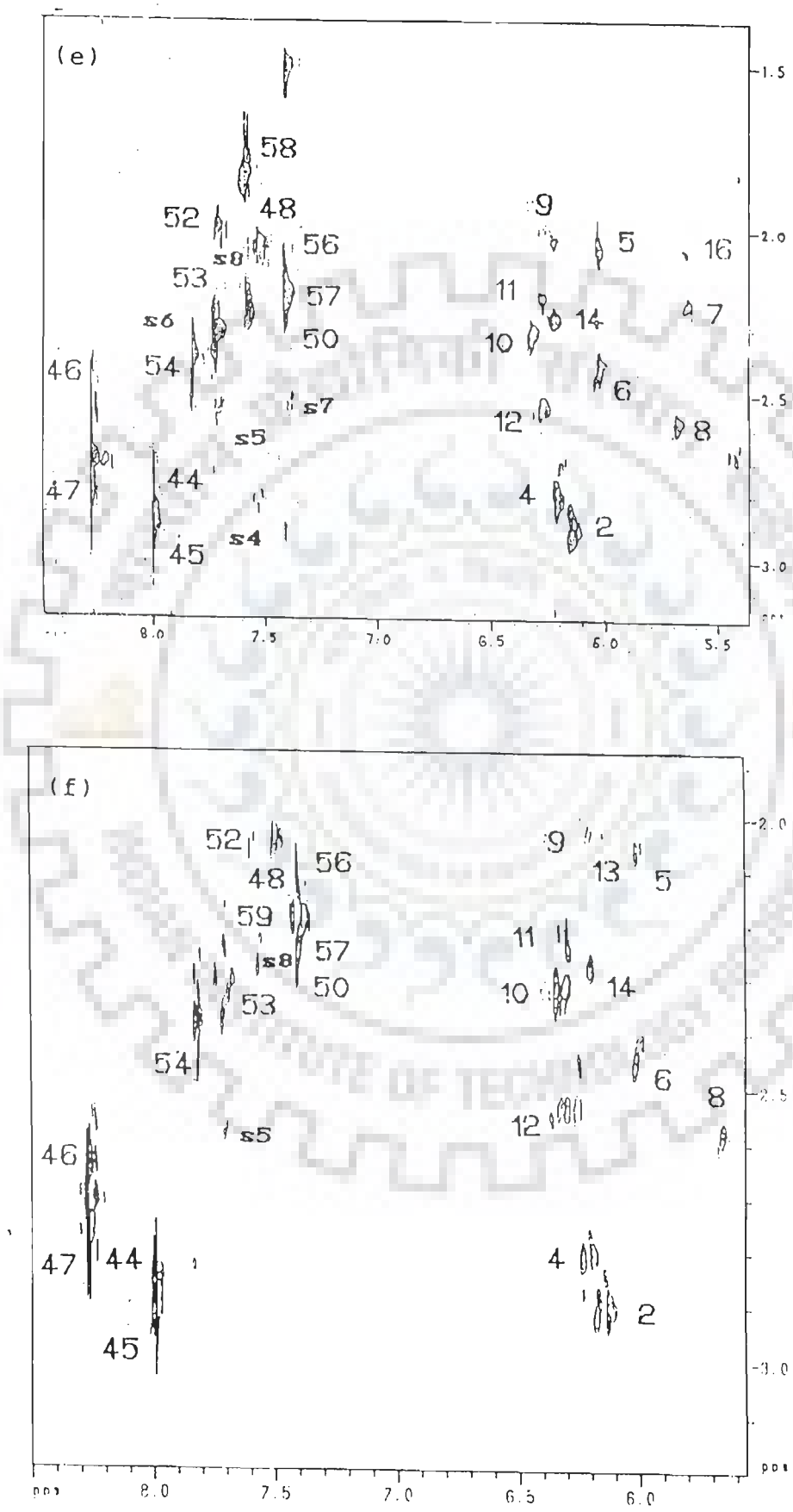


FIG. 4.4(e) and 4.4(f): Same as Fig.4.4 (d) but at lower sensitivity (greater noise) to show some of the weakly intense and sequential NOE cross peaks.

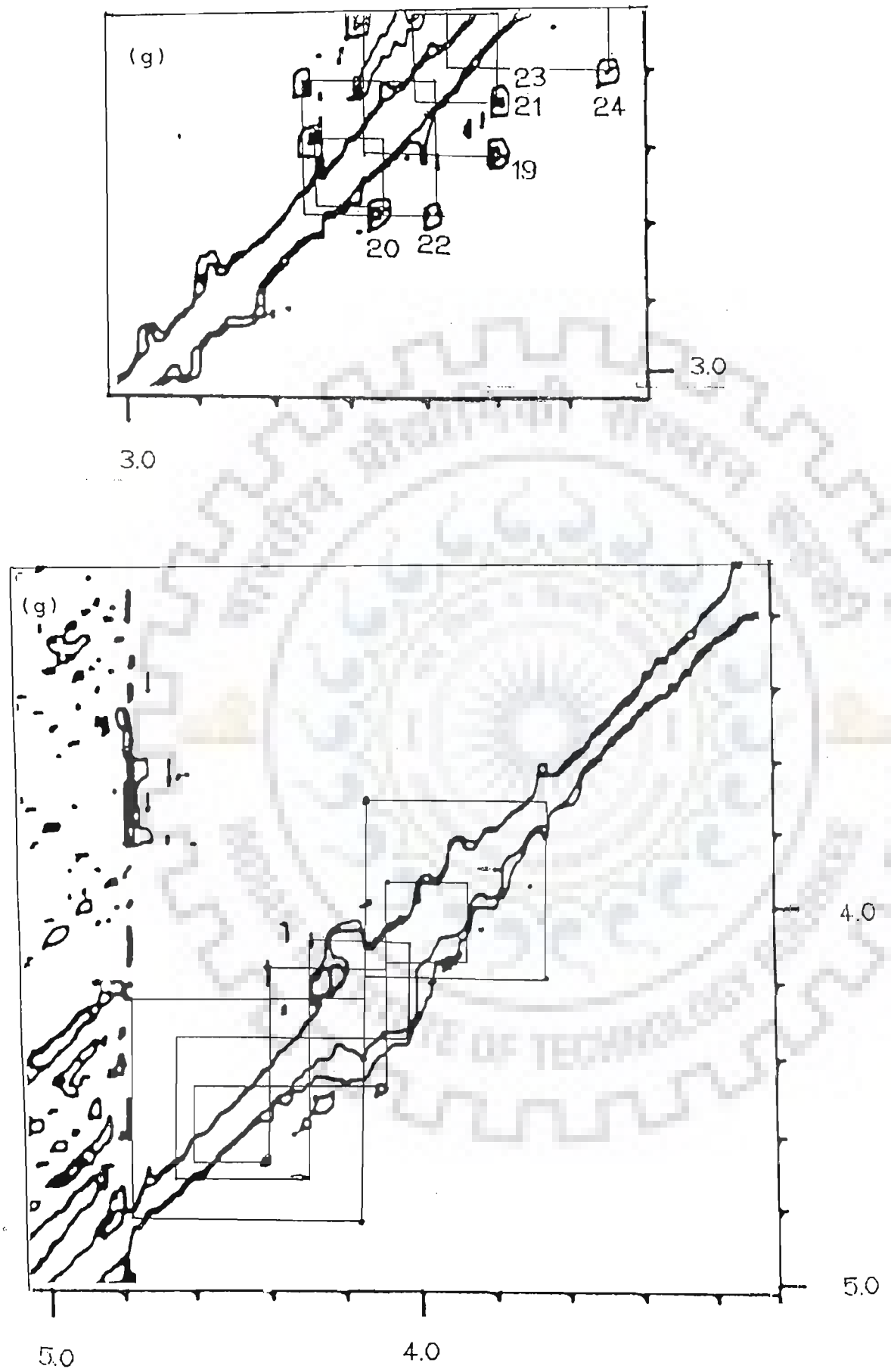


FIG. 4.4(g): NOE cross peaks of H2'-H2" proton of all residues of octamer, H3'-H4' and H4'-H5'/H5" NOE cross peaks of atleast 4 sets of residues are also seen.

the 1D NMR spectrum (Fig.4.2(a)) as well as distance between outer contours along ω_2 axis in COSY spectrum (Fig.4.3(b)) (say cross peak pattern no. 1, 5, 6, 7, 8, 10, 12, etc.) lie in the range 14.2 to 15.5 Hz. This distance corresponds to $\Sigma 1'$ = J(H1'-H2') + J(H1'-H2") (Refer to Table 3.3. Chapter III) and hence rules out the presence of a predominant C3' endo or C2' exo sugar geometries for which $\Sigma 1'$ is expected to be ~ 7.5 Hz (Table 3.3). Fig.4.5 shows the calculated H1'-H2', H1'-H2" cross peak patterns for flip angle of $\pi/4$ for C3' endo, C2' endo and O1' endo sugar geometries. It has been shown (77) that it offers certain advantages over that with $\phi = \pi/2$. In the latter, the components on the central two lines of multiplet crosspeak pattern are close and because of the opposite phases, tend to cancel their intensities. As a result the H1'-H2' and H1'-H2" coupling constants cannot be measured directly along the same axis. In contrast, the situation with $\phi = \pi/4$ is different. Both H1'-H2' and H1'-H2" coupling constants can be directly measured along the same axis in cross peak (Fig.4.5). We have used a flip angle of $\pi/4$ in phase-sensitive COSY spectrum. Comparing the cross peak pattern H1'-H2" of C3 and T4 residues (pattern marked as no. 6 and 8 in Fig.4.3(b)) with that in Fig.4.5, it is easily inferred that sugar geometry is predominantly C2'endo. Further the cross peak pattern H1'-H2' of C3 and T4 residues (no. 5 and 7 in Fig.4.3(b)) compare well with the corresponding patterns expected for C2' endo geometry (Fig.4.5). A close look at expansions of all cross peak patterns of Fig.4.3(b) revealed that G1, A2, C3, T4, G6 and C8 (not seen in Fig.4.3(b) but seen in other expansions available) residues have C2' endo geometry. Comparison of the H1'-H2" cross peak pattern of C5 and T7 residues (no. 10 and 14 in Fig.4.3(b)) with the corresponding patterns for other residues (for example pattern no. 6 and 8 in Fig.4.3(b)) reveals distinct difference between the two sets. It is not possible to infer straightaway if C3' endo or O1' endo are predominant sugar forms as the observed pattern no. 10 and 14 do not exactly match with corresponding ones in Fig.4.5. More realistically, it appears to be a mixture of these so that one may infer that sugar geometry in C5 and T7 residues has other

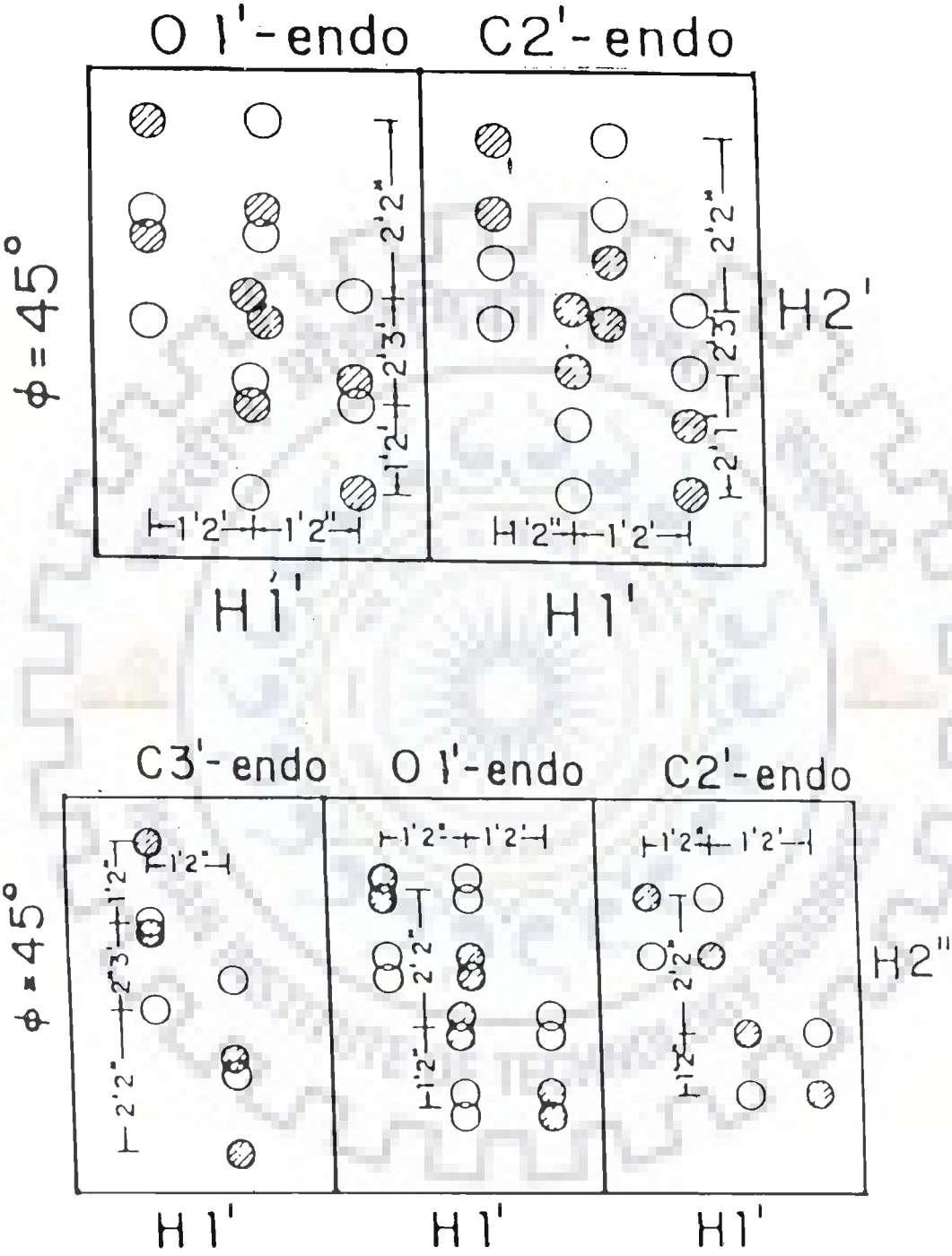


FIG. 4.5: Calculated H1'-H2' and H1'-H2" cross peak patterns for flip angle $\phi = \pi/4$ from C3' endo C2' endo and O1' endo sugar geometries (77).

puckers existing besides C2' endo. This is in accord with other results obtained from COSY spectra and is discussed later.

An estimate of coupling constant was made from the blow ups of the observed H1'-H2' and H1'-H2" cross peak patterns of Fig.4.3(b). By comparing them with calculated patterns shown in Fig.4.5, the peak separations revealed that for G1, A2, C3, T4, G6, C8, residues the corresponding J values are H1'-H2' ~ 5 ± 0.5 Hz, H1'-H2" ~ 10 ± 0.5 Hz, H2'-H2" ~ 15 ± 0.5 Hz, Σ1' = 14.2 to 15.5 Hz, Σ2' = 28.8 to 29.4 Hz, Σ2" = 24 to 25 Hz. Comparing these with the standard values given in Table 3.3 (Chapter III), it is confirmed that in all these residues C2' endo sugar is the predominant sugar geometry. It is possible that 10-20 % of N-conformer is present (column 7 Table 3.3) the exact quantitative estimate of which is not attempted due to low resolution of spectra.

Typical cross peak patterns expected for H2'-H3' J-coupled protons in C2' endo sugar geometry are observed for A2, T4 and G6 residues shown in Fig.4.3(e, f and g), respectively. This enables us to estimate J(H2'-H3'), J(H1'-H2'), J(H2'-H2") and Σ2' from separation of peaks using Fig.3.10 described earlier in Chapter III. The observed coupling constants for these residues are J(H2'-H3') ~ 4.4 ± 0.5 Hz, J(H1'-H2') ~ 10.7 ± 0.5 Hz, J(H2'-H2") ~ 14.3 ± 0.5 Hz and Σ2' = 29.4 ± 0.5 Hz, which are in accord with the standard values for C2' endo geometry (Table 3.3).

It is observed that weak cross peak patterns of H2"-H3' J-coupled connectivity exist for C5 and T7 residue (Fig.4.3(f))(Table 4.3). Comparing this pattern with that for J(H2"-H3') pattern expected for C3' endo geometry (Fig.3.10), a very rough estimate of J(H2"-H3') ~ 9-10 Hz and J(H1'-H2") ~ 7.8 Hz was made. This H3' resonance gives a COSY cross peak pattern with H4' seen clearly in Fig.4.3(a) and expanded in Fig.4.3(h). The H4' is further J-coupled to its corresponding H5', shown as expansion in Fig.4.3(i) and hence gets assigned to C5/T7 residues from the results of COSY spectra.

The existence of a clear H3'-H4' J-coupled cross peak indicates the presence of C3' endo/O1' endo sugar geometry in C5/T7 residues besides a predominant S-conformer. This is in accord with the fact that H1'-H2' and H1'-H2" cross peak pattern for these residues (no. 10, 13 and 14 in Fig.4.3(b)) are significantly different from the corresponding patterns for other residues (say no. 5-6, 7-8 in Fig.4.3(b)). In fact H1'-H2' cross peak pattern is not observed for C5 residue while that for T7 residue is weak in intensity. Thus we conclude that C5 sugar has more N-conformer than that of T7 while deoxyribose sugar of all other nucleotide residues are primarily in S-conformational state.

As described in Chapter III (section C), the relative intensities of intrasugar NOE connectivities may be used to ascertain sugar conformation since the intrasugar distances vary with psuedorotation and are independent of glycosidic bond rotation (Table 3.5 Chapter III). Table 4.4 gives the relative intensities of cross peaks seen in Fig.4.4(d-g). The distance H1'-H2" being minimum and about 2.1-2.2 Å for all sugar geometries (Table 3.5) gives the most intense cross peak. The H1'-H2' distance of 2.6-2.8 Å yields relatively weakly intense NOE cross peaks for all residues. The intensities of H2'-H3' and H2"-H3' NOE cross peak could not be ascertained due to noise in NOESY spectra in the corresponding regions. The H3'-H4' and H4'-H5'/H5" NOE connectivities are seen for several residues. But the H3' resonance of C5/T7 residue at 4.60 ppm does not give NOE connectivity with its corresponding H4' proton resonating at 4.14 ppm (Table 4.4) (Fig.4.4(g)) which further shows NOE cross peak with its corresponding H5'/H5" protons resonating at 3.71 ppm. Therefore the H3'-H4' peaks which are J-coupled for C5/T7 residues do not give NOE cross peak. Since H2"-H4', H1'-H4' are not seen for any residue due to noise in that region, sugar geometry could not be ascertained from relative intensities of intrasugar NOE connectivities.

Table 4.3: Presence (marked as +) or absence (marked as -) of J-coupled intrasugar connectivities in phase-sensitive COSY spectra (Fig.4.3 (a-i)) of octamer in unbound form.

	G1	A2	C3	T4	C5	G6	T7	C8
H1'-H2'	+	+	+	+	+	-	+	+
H1'-H2"	+	+	+	+	+	+	+	+
H2'-H3'	-	+	-	+	-	+	-	-
H2"-H3'	-	-	-	-	+	-	+	-
H3'-H4'	-	-	-	-	+	-	+	-

Table 4.4: Intensity[#] of cross peaks of intrasugar NOE connectivities in octamer in unbound state from NOESY spectra (Fig.4.4(d-g)).

	G1	A2	C3	T4	C5	G6	T7	C8
H1'-H2'	w	w	w	w	w	w	w	x
H1'-H2"	ss	ss	ss	ss	ss	ss	ss	ss
H3'-H4'	ss	o	ss	ss	x	o	x	ss

ss - very strong; s - strong; w - weak; x - not seen; o - overlap

Table 4.5 : Intensity[#] of cross peaks of base H6/H8 to sugar NOE connectivities in octamer in unbound state from Fig.4.4 (d-g).

	G1	A2	C3	T4	C5	G6	T7	C8
Base - H1'	x	x	w	x	x	x	x	x
Base - H2'	ss	ss	ss	ss	ss	ss	ss	ss
Base - H2"	w	w	x	x	s	x	w	s

ss - very strong; s - strong; w - weak; x - not seen

Glycosidic bond rotation

The glycosidic bond rotation χ = O1'-C1'-N9-C8 (purines) or O1'-C1'-N1-C6 (pyrimidines) may be estimated from the knowledge of intranucleotide distances between base CH6/TH6/GH8/AH8 proton and their corresponding sugar protons (Table 3.7, Chapter III). The observed relative intensities of corresponding cross peaks in octamer, as seen in Fig.4.4 (d-g), are given in Table 4.5. The base to H2' distance being the minimum and about 2.2 Å gives very strong intensity of cross peak for all residues. On the other hand, base to H2" NOE connectivities are relatively much weaker as expected for P = 162° (major S-conformer of deoxyribose) and χ =-105°, for C1, A2, C3, T4, G6, T7 residues. For C5 and C8 residues, the base to H2" NOE connectivity was relatively strong and certainly does not reflect distance as large as ~ 3.6 Å. Lesser distance is possible only if χ is varied from -105° to higher values of say χ = -90° (150). Since a variation of χ in the opposite sense, that is, a decrease to say χ = 130° results in a further increase in distance to a value of 4.0 Å contrary to the observed results. Thus we infer that glycosidic bond rotation is in anti conformational state for all residues except C5 and C8 for which it is likely to be high anti.

C) Conformation of octamer d-$(GACTCGTC)_2$ bound to hexapeptide

Fig.4.6(a-f) and Fig. 4.7(a-d) show the phase-sensitive COSY and NOESY spectra of 8.60 mM octamer d-$(GACTCGTC)_2$ bound to 0.69 mM hexapeptide KPYSLN in D_2O at 297 K. Table 4.6 gives the chemical shift values of various protons of octamer in complexed form at 297 K ascertained from these 2D COSY and NOESY spectra as well as 1 D NMR spectra (Fig. 4.10, shown later while discussing the changes in chemical shift). The resonance peaks of various protons of octamer as well as that of peptide in all NMR spectra are easily recognisable by comparing the spectra with the corresponding spectra in unbound form (Fig.4.2 and Fig.3.1(b)). The cross peaks in COSY and NOESY spectra are numbered as in corresponding spectra of unbound form (Table 4.2). The strategies used in spectral assignment are same as that described in section

Table 4.6a: Chemical shift values (in ppm) of various protons of d-$(GACTCGTC)_2$ in the mixture of 8.60 mM d-$(GACTCGTC)_2$ and 0.69 mM KPYSLN at 297 K.

	G1	A2	C3	T4	C5	G6	T7	C8
H8/H6	-	8.26	7.50	7.40	7.70	7.71	7.40	7.54
H1'	6.18	6.23	6.10	5.73	6.32	6.32	6.25	5.75
H2'	2.87	2.76	2.08	2.38	2.00	2.24	2.08	1.82
H2"	2.98	2.81	2.42	2.58	2.35	2.59	2.27	2.20
H3'	-	-	-	4.72	4.56	-	-	-
H4'	-	-	-	4.22	4.20	-	-	-
H5'/H5"	-	-	-	-	3.68	-	-	-
CH5	-	-	5.55	-	5.91	-	-	5.61
TCH_3	-	-	-	1.50	-	-	1.50	-
H2	-	8.14	-	-	-	-	-	-

B for unbound octamer.

In order to determine the conformation of deoxyribose sugar, the distance between the outer peaks of H1' proton along ω_2 axis is measured in Fig.4.6(b). For the various cross peak patterns e.g. 5,6,7,8,13,14,11,12, this distance is found to lie in the range 14.2 to 15.1 Hz. Since it corresponds to $\Sigma 1'$ = J(H1'-H2') + J(H1'-H2"), the presence of a predominant C3' endo or C2' exo sugar geometry is ruled out. A flip angle, ϕ, of $\pi/2$ is used in getting COSY spectra. A close look at blow ups of the cross peak patterns of Fig.4.6(b) shows that H1'-H2' patterns of all residues are similar (no. 1,3,5,7,11,13,15). The corresponding pattern for C5 residue (no.9) is not observed. Comparing the various H1'-H2" patterns with each other it is noted that patterns for C3, T4, G6, T7, C8 (no. 6,8,12,14,16) are exactly same but are significantly different from the corresponding pattern for C5 residue (no. 10). H1'-H2" pattern of A2, G1 residue (no. 2,4) are also same as that for most of the residues (no. 6, 8, 12, etc.) although they are merged in Fig.4.6 (b) with their corresponding H1'-H2' patterns. Therefore the sugar conformation of C5 residue is apparently different from that of all other residues.

The cross peak patterns of Fig.4.6(b) are compared with the expected pattern shown in Fig.3.10 particularly with that for C2' endo geometry, the major S-conformer. For the H1'-H2' pattern (Fig. 3.10) the pattern [- - + +] expected along ω_2 axis is seen as [- +] with wide positive and negative peak due to merging of two [- -] and two [+ +] peaks as one each [-] and [+] peak. This is clear in all patterns e.g. no. 11, 13, 5, 7 etc. The H1'-H2' pattern along ω_1 axis (Fig.3.10) expected as [+ + - + - -] looks like [+ -] in the cross peak pattern no. 11,13,5,7, etc. due to merging of two [+ +] and two [- -] peaks on left hand and right hand corners. The active coupling H1'-H2' is easily read along ω_2 axis as ~ 9.9 ± 0.5 Hz (varies between 9.4 and 10.4 Hz in various cross peak patterns (no. 8,12,6, etc.) and the H2"-H3' coupling is buried within the width of peaks.

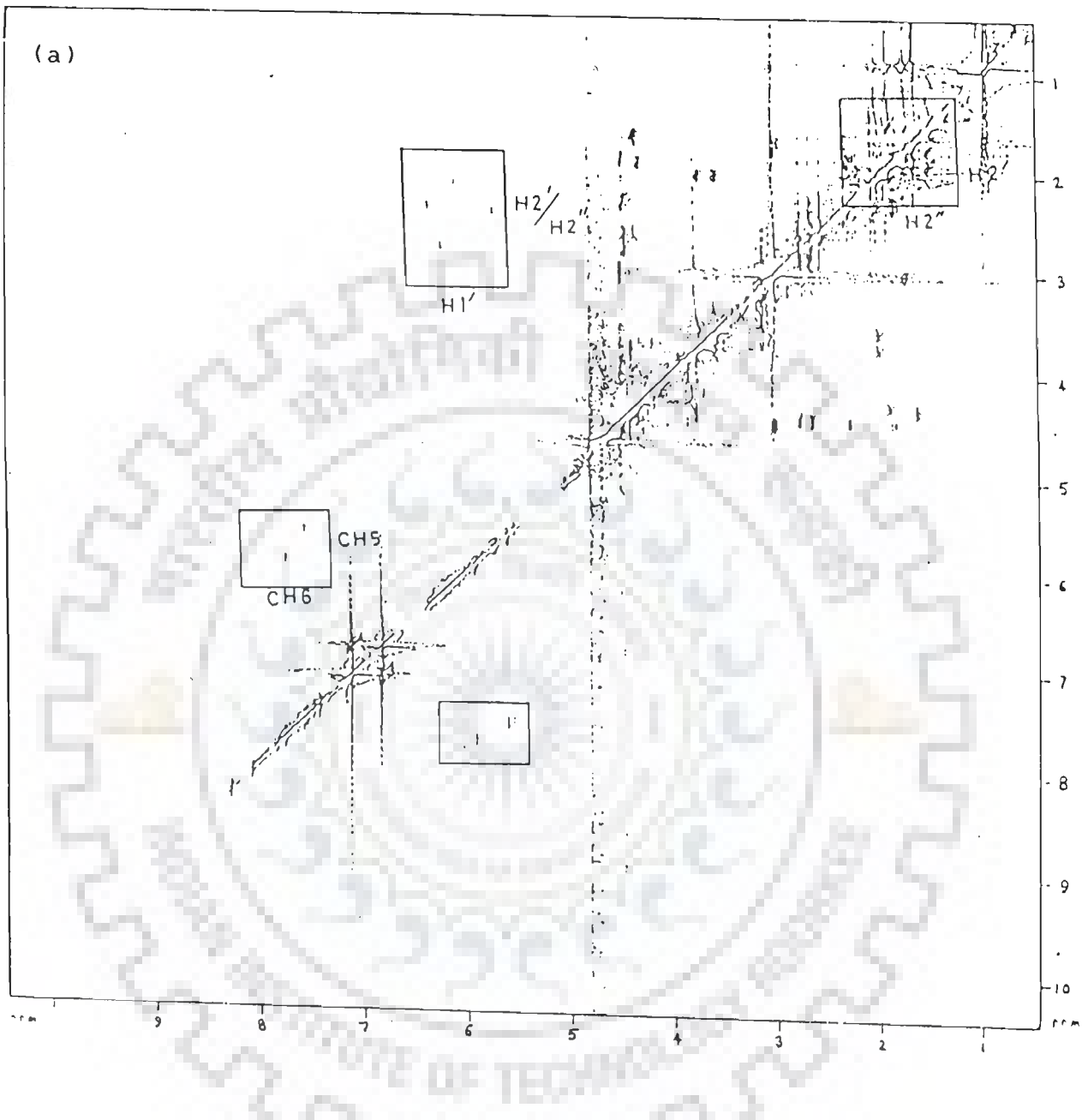


Fig. 4.6(a): Phase-sensitive COSY spectra of a mixture of 8.60 mM d-(GACTCGTC)$_2$ and 0.69 mM hexapeptide KPYSLN (P/N = 0.08) in D_2O at 297 K.

(b)

Fig. 4.6(b):An expansion of Fig.4.6(a) to highlight specific connectivities no. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 and 16 (Table 4.2).

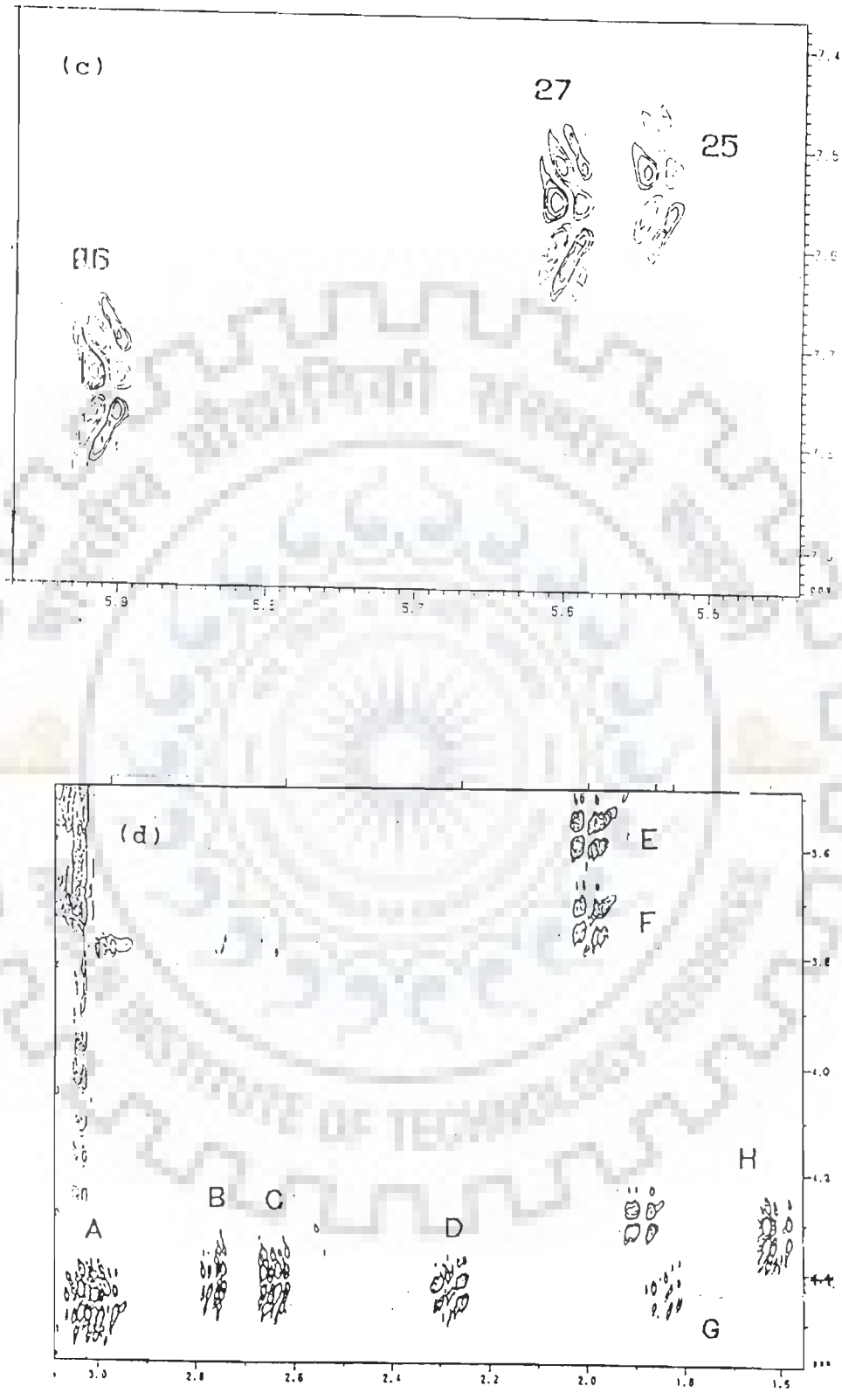
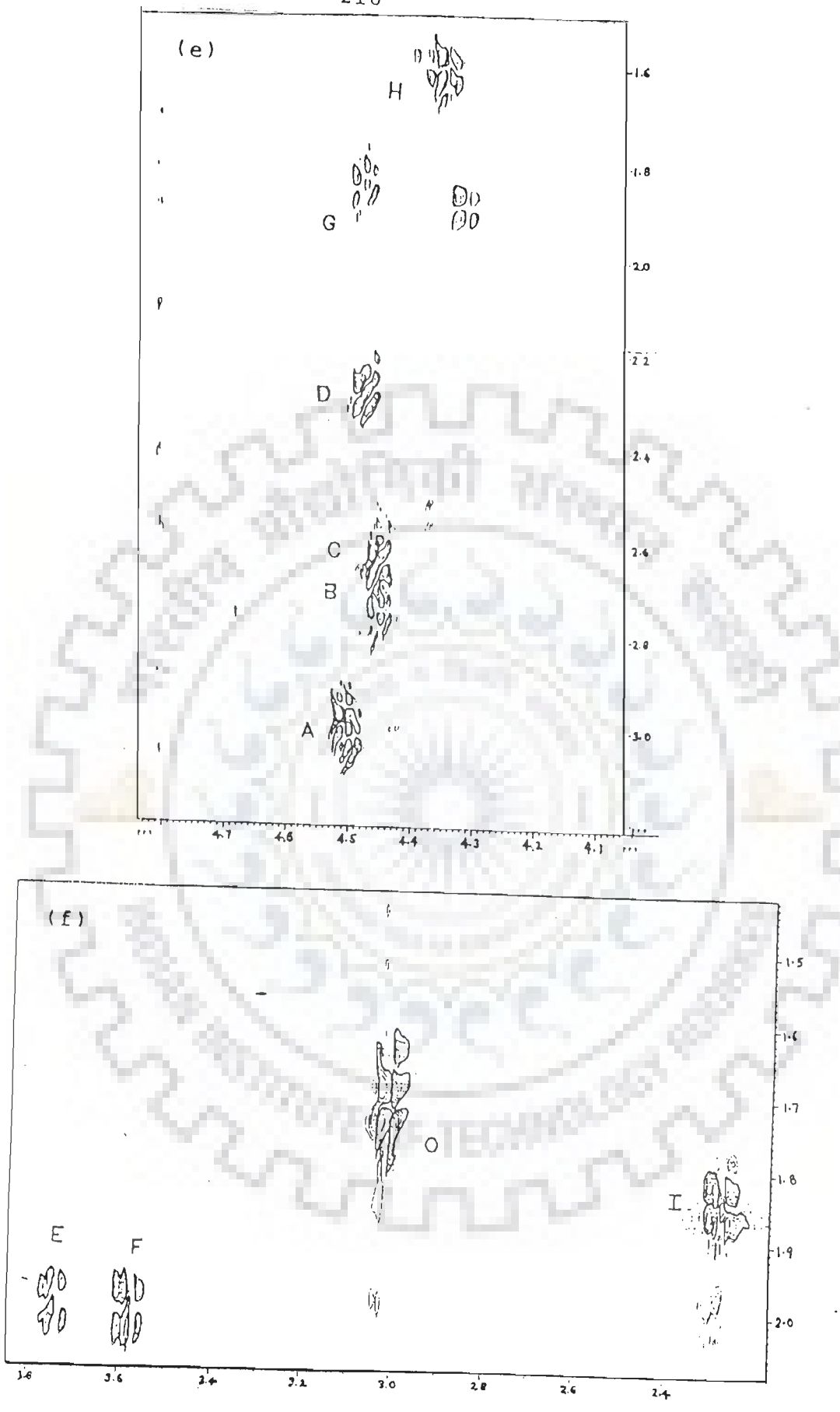


Fig. 4.6(c):An expansion of Fig.4.6(a) to highlight specific connectivities no. 25, 26 and 27 (Table 4.2) (d): Expansion shows specific connectivities no.(A)Tyr α-CH - Tyrβ-CH_2 (B)Asn α-CH - Asn β'-CH_2 (C) Asn α-CH - Asn β-CH_2 (D) Pro α-CH - Pro β-CH_2 (E) Pro γ-CH_2 - Pro δ-CH_2 (F)Pro γ-CH_2 - Pro δ'-CH_2 (G) Lys α-CH - Lys β-CH_2 (H) Leu α-CH - Leu β-CH_2.



g. 4.6(e): An expansion of Fig.4.6(a) to highlight specific connectivities
.(A) Tyr α-CH - Tyr β-CH_2 (B)Asn α-CH - Asn β'-CH_2 (C)Asn α-CH - Asn β-CH_2
) Pro α-CH - Pro β-CH_2 (G) Lys α-CH - Lys β-CH_2 (H) Leu α-CH - Leu β-CH_2
): Expansion shows specific connectivities no. (O) Lys δ-CH_2 - Lys ε-CH_2
)Pro γ-CH_2 - Pro δ-CH_2 (F) Pro γ-CH_2 - Pro δ'-CH_2

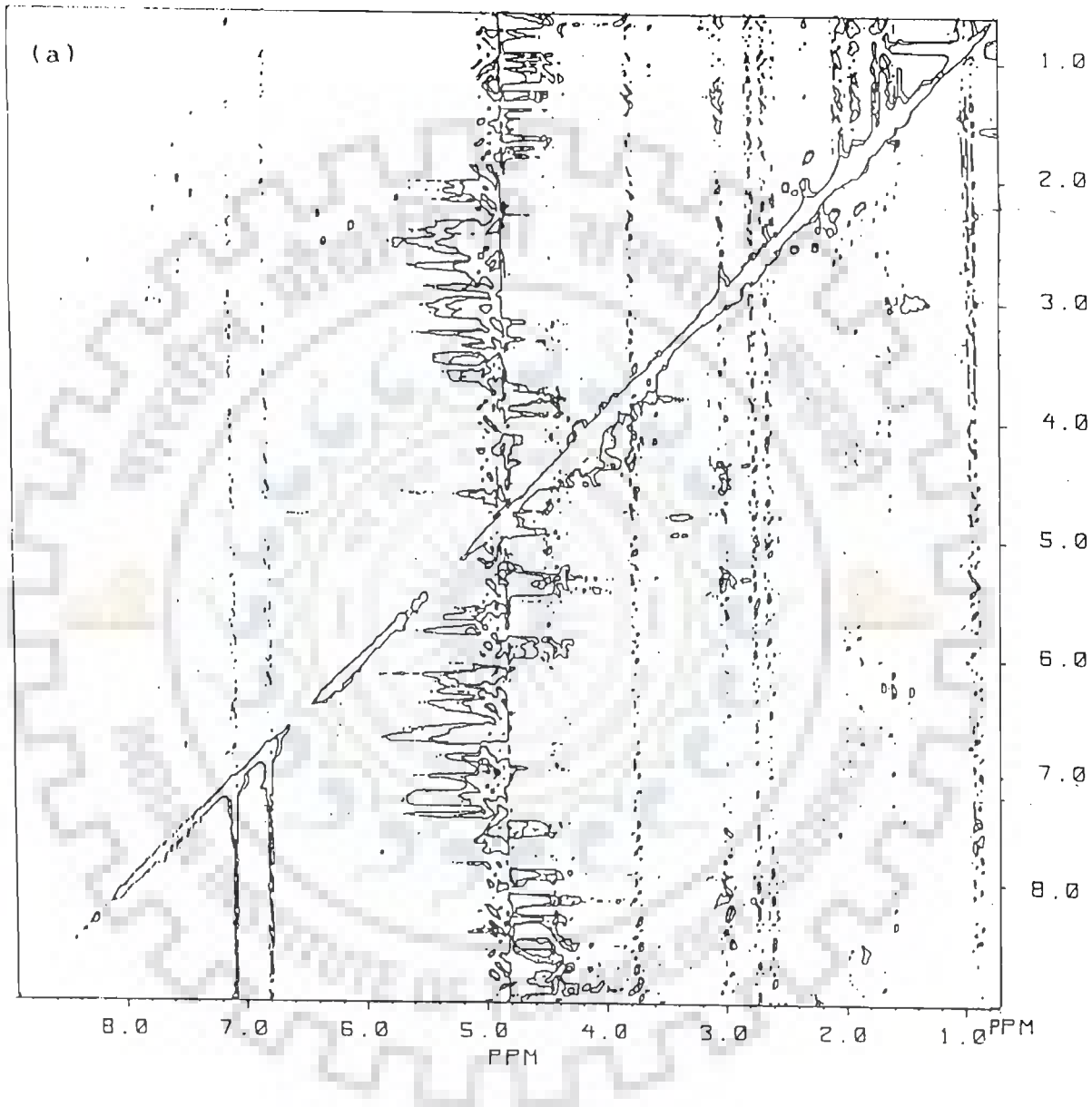


Fig. 4.7(a):Phase - sensitive NOESY spectra of a mixture of 8.60 mM d-$(GACTCGTC)_2$ and 0.69 mM hexapeptide KPYSLN (P/N= 0.08) in D_2O at 297 K.

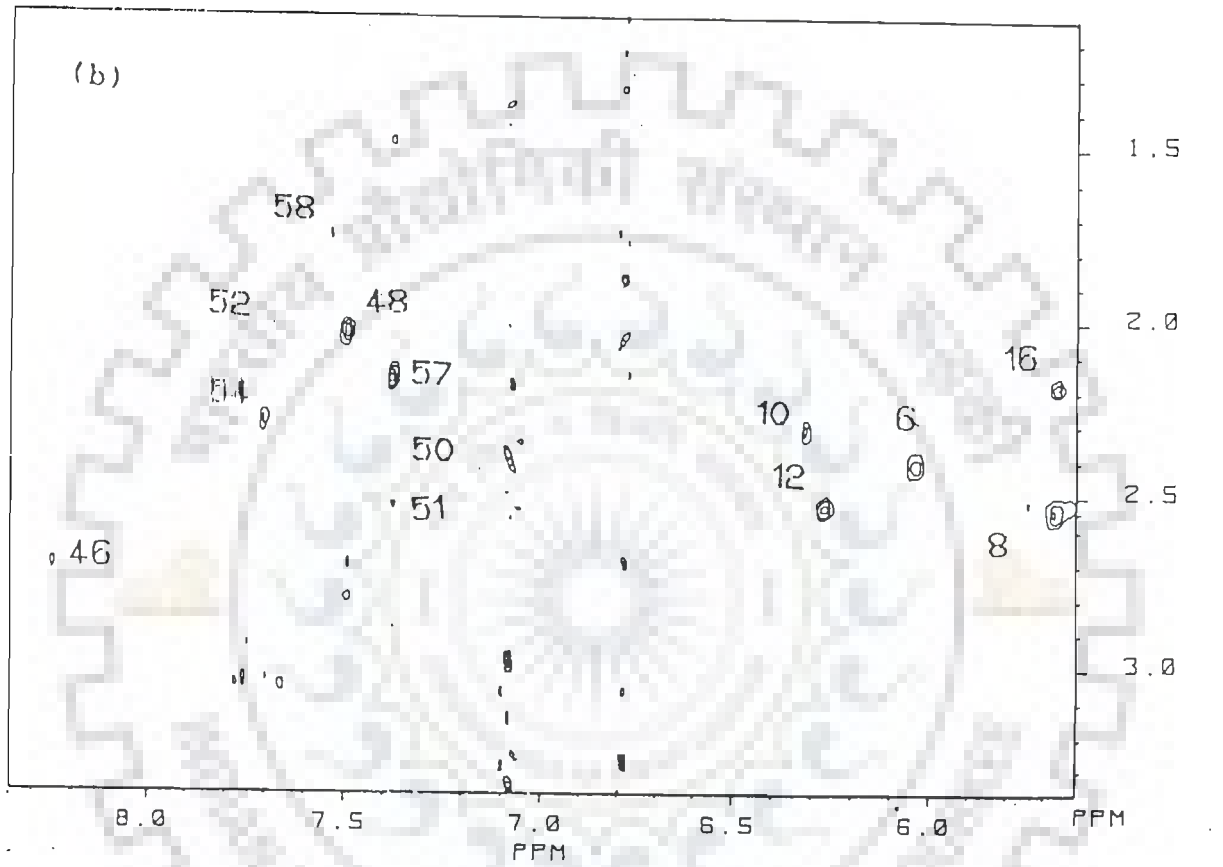


Fig. 4.7(b): An expansion of Fig.4.7(a) to highlight specific connectivities no. 6, 8, 12, 10, 16, 46, 48, 51, 52, 54, 57, 58 (Table 4.2).

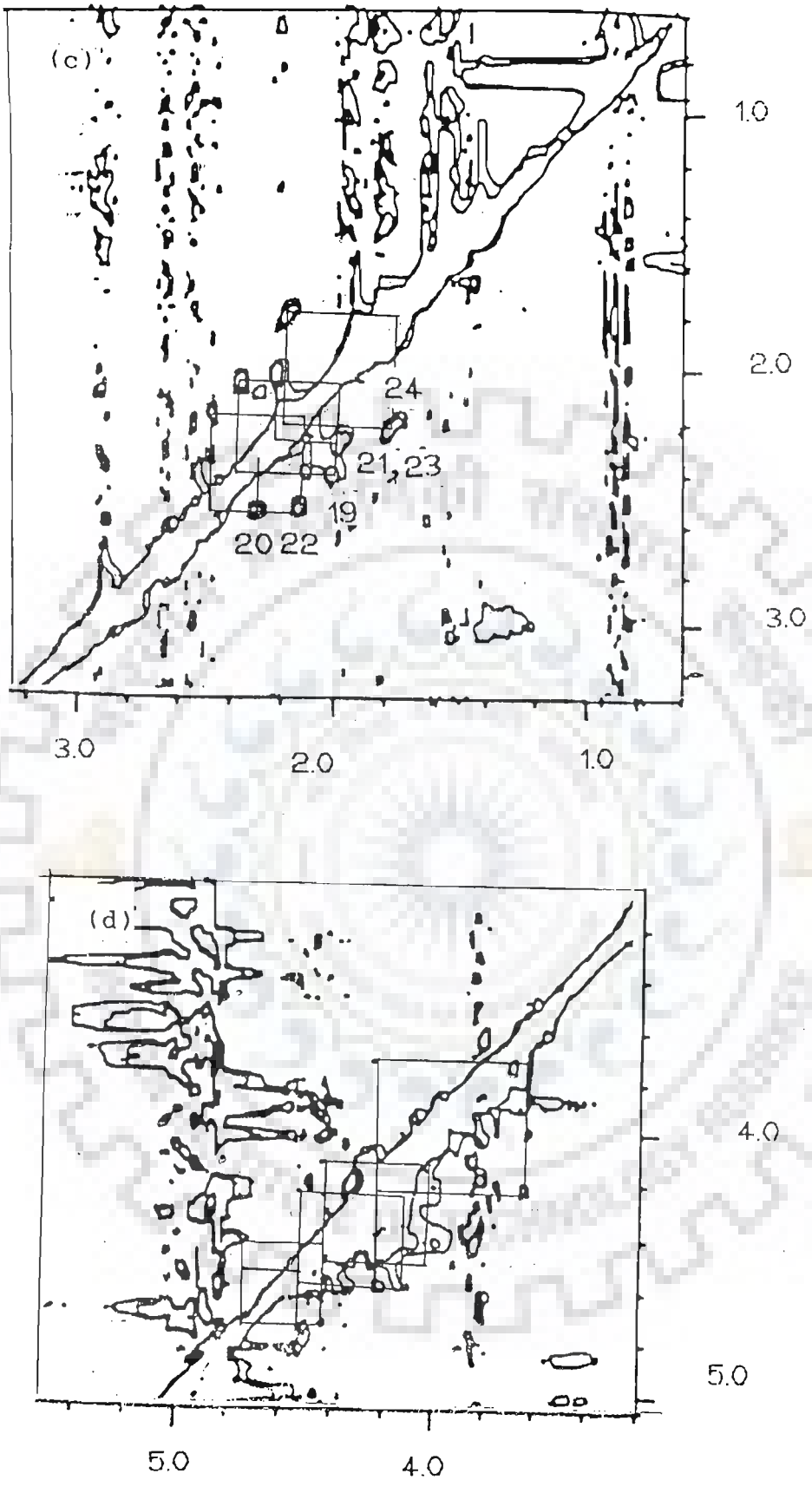


Fig. 4.7(c): An expansion of Fig.4.6(a) to highlight specific connectivities of H2'-H2" protons of all residues peak no. 19, 20, 22, 23,24 (d): Expansion shows H4'- H5'/H5" NOE cross peak pattern of at least three residues of octamer.

The pattern along ω_1 axis therefore, expected as [+ + - - + + - -] looks like [+ - + -] whereas that along ω_2 axis expected as [- + - +] looks like [- +] with the central two merging together due to width of peaks. The average values of J(H1'-H2'), J(H1'-H2"), J(H2'-H2") and Σ1' read from the expansions of crosspeaks are 9.9, 5.6, 15.0 and 14.7 Hz, respectively to an accuracy of ± 0.5 Hz.

The analysis of H1'-H2" cross peak pattern (no. 10) of C5 residue reveals that the observed pattern is more close to that expected for O1' endo geometry than that for C3' endo geometry (Fig.3.10). The active coupling is ~ 6.6 while Σ1' ~ 15 Hz to an accuracy of ± 0.5 Hz. Comparing these values with the spin-spin coupling for standard sugar geometries (Table 3.3), we can infer that either O1' endo or a mixture of C3' endo and C2' endo (say for example in ratio 20:80) geometries exist for C5 residue.

The relative intensities of intra residue NOE connectivities in octamer in complexed form as seen in Fig.4.7 are given in Table 4.7. Intense peaks are seen for H1'-H2" NOE connectivities for which distances lie in the range 2.1-2.2 Å irrespective of sugar geometry. The NOE connectivity H1'-H2' is not observed, so that distance > 2.8 Å do not yield NOE cross peak. Base H6/H8 to H2' NOE connectivities are found to be very intense while the base to H2" NOE connectivities are not observed at all. Using the standard values of base-H2'/H2" distances for various sugar geometries and glycosidic bond rotation (Table 3.5 and 3.7), we conclude that deoxyribose sugar is present in a major S-conformational state and glycosidic bond rotation is in anti conformation for all residues.

D) Changes in chemical shift and structure of complex

Figs. 4.8 and 4.9 show the proton NMR spectra at different temperatures and the temperature profile, respectively of various oligonucleotide protons in the temperature range 275-335 K (Table 4.8). The change in chemical shift with temperature is ascribed to opening of the helix (14) into single strands. The various base

Table 4.6b: Change in chemical shift values (in ppm) of various protons of octamer in a mixture of 8.60 mM d-$(GACTCGTC)_2$ and 0.69 mM KPYSLN at 297 K. Upfield shifts are taken as negative.

	G1	A2	C3	T4	C5	G6	T7	C8
H8/H6	-	-0.01	0.00	-0.01	0.05	-0.13	0.01	-0.02
H1'	-0.01	0.03	0.06	0.07	-0.04	0.00	-0.01	0.11
H2'	-0.03	-0.01	0.04	0.00	-0.02	0.00	0.02	0.10
H2"	-0.04	-0.01	0.00	0.00	0.05	0.00	-0.01	0.00
H3'	-	-	-	0.12	-0.04	-	-	-
H4'	-	-	-	-	0.06	-	-	-
H5'/H5"	-	-	-	-	-0.03	-	-	-
CH5	-	-	0.06	-	-0.01	-	-	0.05
TCH_3	-	-	-	0.00	-	-	0.00	-
H2	-	0.06	-	-	-	-	-	-

Table 4.6c: Chemical shift values of various peptide protons in a mixture of 8.60 mM d-$(GACTCGTC)_2$ and 0.69 mM KPYSLN in D_2O at 297 K.

	Lys	Pro	Tyr	Ser	Leu	Asn
α–CH	4.45	4.45	4.50	4.40	4.32	4.45
β-CH_2	1.86	2.28	3.00	3.77	1.62	2.64
						(β'2.77)
γ-CH_2	1.43	1.96	-	-	1.62	-
δ-CH_2	1.68	3.56	-	-	-	-
		(δ'3.74)				
δ-CH_3	-	-	-	-	0.94	-
ε-CH_2	3.02	-	-	-	-	-
(3,5)H			6.87			
(2,6)H			7.15			

Table 4.6d: Change in chemical shift values (Δδ) of various peptide protons in a mixture of 8.60 mM d-$(GACTCGTC)_2$ and 0.69 mM KPYSLN in D_2O at 297 K. Upfield shifts are taken as negative.

	Lys	Pro	Tyr	Ser	Leu	Asn
α–CH	0.00	0.00	0.00	0.00	0.00	0.02
β-CH_2	0.03	0.02	0.06	0.02	0.01	0.01
						(β'0.02)
γ-CH_2	0.04	0.00	-	-	0.00	-
δ-CH_2	0.02	0.00	-	-	-	-
		(δ'0.02)				
δ-CH_3	-	-	-	-	0.03	-
ε-CH_2	0.04	-	-	-	-	-
(3,5)H			0.00			
(2,6)H			0.00			

Table 4.7: Intensities of some of the intraresidue NOE connectivities as observed in NOESY spectrum (Fig.4.7(a-d)) of a mixture of 8.6 mM d-$(GACTCGTC)_2$ and 0.69 mM KPYSLN in D_2O at 297 K.

	G1	A2	C3	T4	C5	G6	T7	C8
H1'-H2'	x	x	x	x	x	x	x	x
H1'-H2"	x	x	ss	ss	ss	ss	x	ss
H8/H6-H2'	x	s	ss	x	x	ss	ss	s
H8/H6-H2"	x	x	x	x	x	x	x	x

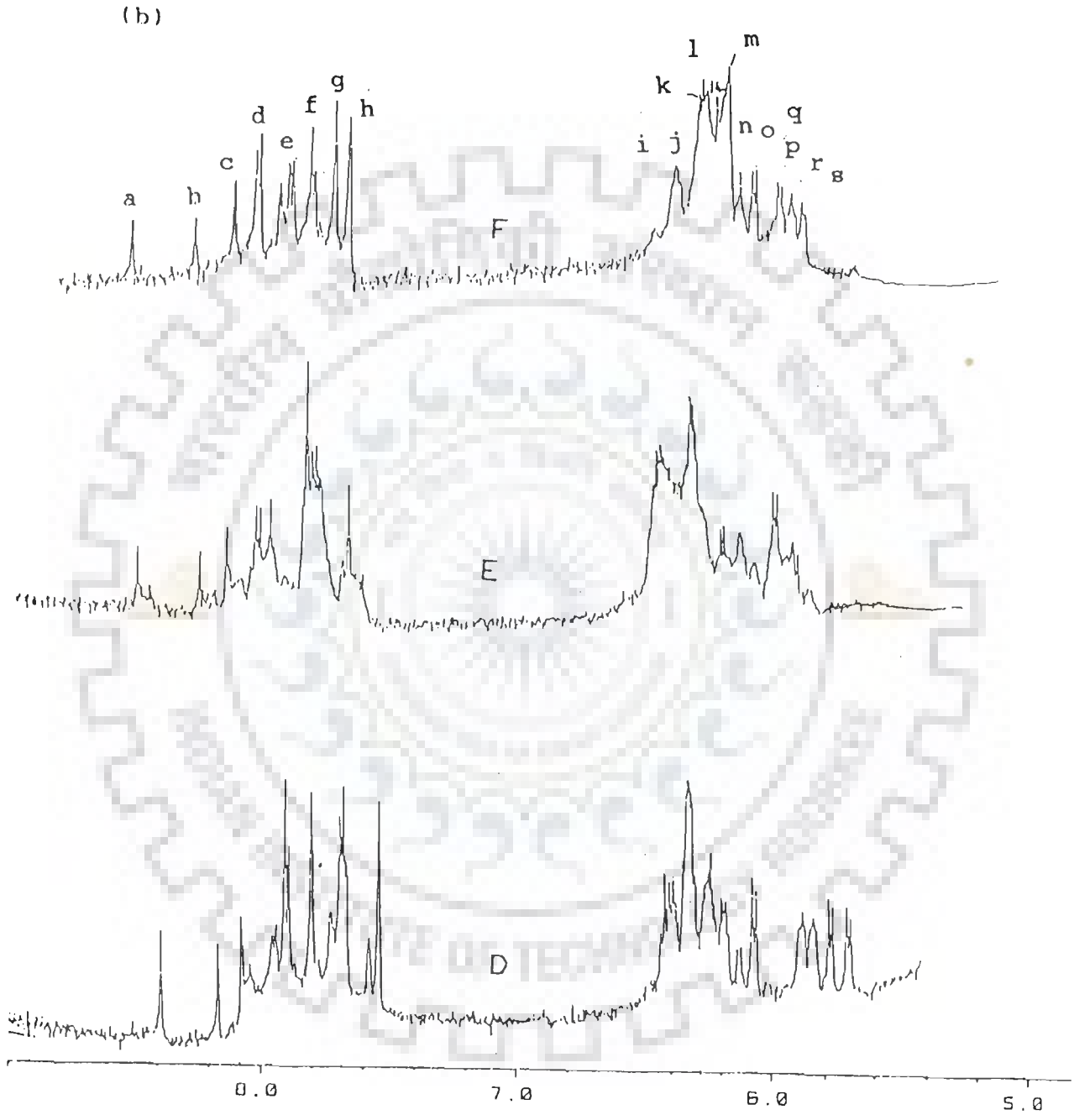
ss- very strong; s-strong; x-not seen

Table 4.8: Chemical shift values (in ppm) of various protons of 10 mM d-$(GACTCGTC)_2$ in D_2O (pH 7.2) at different temperatures.

Temp.(K)	G1H8	A2H8	A2H2	C3H6	T4H6	C5H6	G6H8	T7H6	C8H6
275	7.82	8.05	7.86	7.32	7.18	7.61	7.68	7.18	7.48
285	7.91	8.16	7.98	7.43	7.29	7.69	7.75	7.29	7.60
295	7.99	8.27	8.08	7.50	7.41	7.75	7.84	7.41	7.56
305	8.07	8.38	8.17	7.68	7.54	7.89	7.80	7.54	7.72
315	8.14	8.49	8.25	7.80	7.67	8.02	7.97	7.67	7.84
335	8.26	8.63	8.40	7.97	7.84	8.17	8.09	7.84	8.06

Temp.(K)	G1H1'	A2H1'	C3H1'	T4H1'	C5H1'	G6H1'	T7H1'	C8H1'
275	-	5.97	5.77	5.41	6.07	6.17	6.12	5.29
285	-	6.09	5.91	5.53	6.18	6.26	6.23	5.44
295	6.19	6.20	6.04	5.66	6.36	6.32	6.26	5.64
305	6.25	6.26	5.88	5.84	6.38	6.43	6.39	5.82
315	6.35	6.44	6.15	6.08	6.46	6.48	-	6.08
335	6.51	6.55	6.41	6.27	6.57	6.63	-	6.24

g. 4.8(a & b):500 MHz proton NMR spectrum of d-(GACTCGTC)$_2$ 10 mM in D_2O at fferent temperatures (A)275 K (B)285 K (C)295 K (D)305 K (E)315 K (F)335 Peaks are labelled as follows: a-A2H8, b-A2H2, c-G1H8, d-G6H8, e-C5H6, -C8H6, g-C3H6, h-T4H6, T7H6, i-C5H1', j-G6H1', k-T7H1', l-G1H1', m-A2H1', -C3H1', o-T4H1', p-C8H1' q-C5H5, r-C8H5, s-C3H5.


(b)
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
F
E
D
8.0
7.0
6.0
5.0

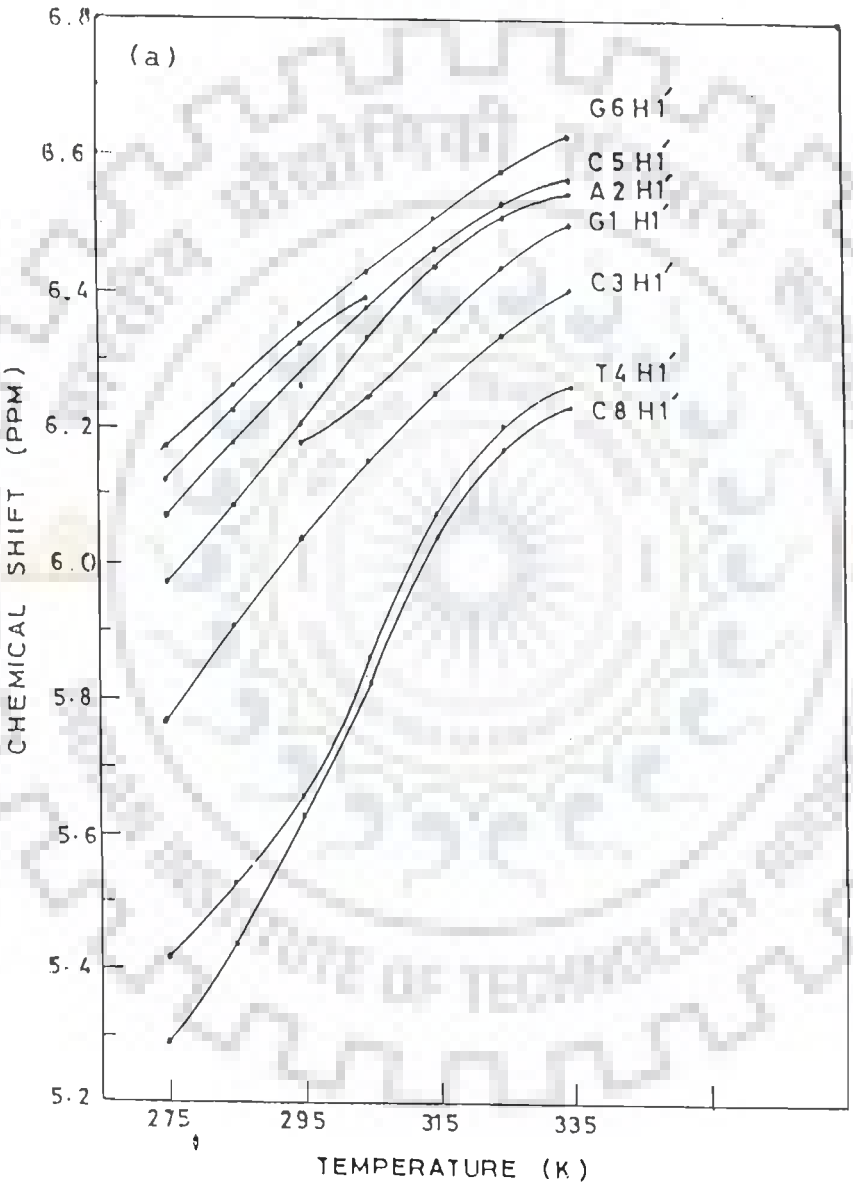


Fig. 4.9(a): Chemical shift of H1' sugar protons of d-$(GACTCGTC)_2$ as a function of temperature (275 -335 K).

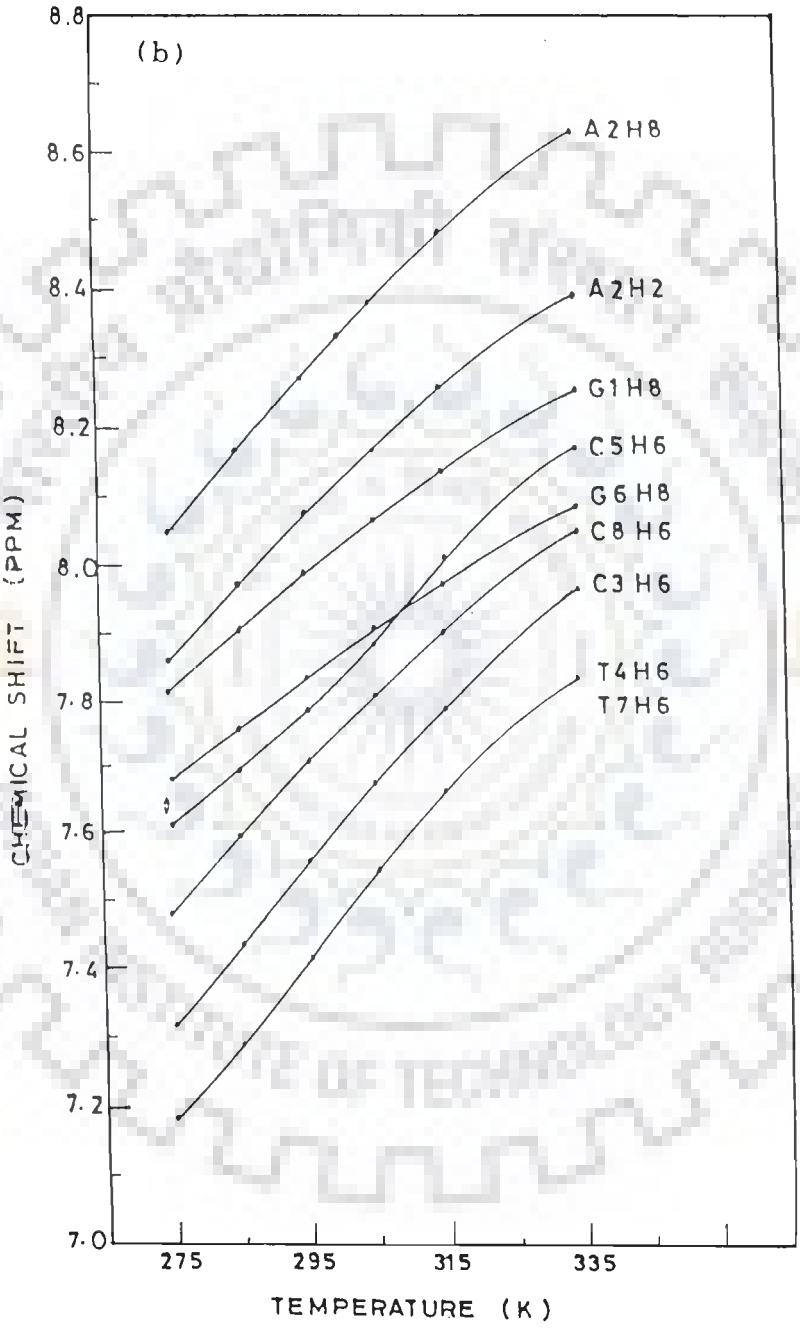


Fig. 4.9(b): Chemical shift of various base protons of d-$(GACTCGTC)_2$ as a function of temperature (275 -335 K).

protons shift downfield with temperature. The total change in chemical shift with temperature , that is, $\Delta\delta = \delta$ (335 K)-δ (275 K) lies in the range 0.41-0.66 ppm.

The H1' protons being close to the aromatic rings of the bases too experience the effect of magnetic anisotropy due to ring currents and shift downfield with temperature by 0.46-0.95 ppm. The melting temperature, Tm, is measured from the temperature profiles of various protons as the temperature at which half the total change in chemical shift i.e. $\Delta\delta = \delta$(335 K)-δ(275 K) occurs. The values of Tm, measured to an accuracy of ± 1K for various base protons are as follows:

Base Protons	G1H8	A2H8	C3H6	T4H6	C5H6	G6H8	T7H6	C8H6
Tm	302 K	301 K	302 K	307 K	303 K	301 K	304 K	301 K

A long range NOE between C3H6-C5H6 protons is seen. In the normal double-stranded DNA, existence of this NOE cannot be explained anyway. The T4 and C5 of 8 mer do not have a matching base-pair in duplex. Due to this, there is a bulge in the centre of d-$(GACTCGTC)_2$ which bring C3H6 close to C5H6 in the structure.

GAC TC GTC
CTG CT CAG

Presence of such long range NOE's have been observed in d-GAATTCCCGAATTC (79).

GAATT CC CGAATTC
CTTAA——GCTTAAG
CC

NOE cross peaks C6H6-G9H6, C6H6-C8H6 and C7H5-TCH_3 indicated above were seen.

Figs.4.10 and 4.11 shows a part of the proton NMR spectra in the range 5.4-9.0 ppm of a mixture of d-$(GACTCGTC)_2$ and KPYSLN (P/N=0.08) at different temperatures (275-335 K). Fig.4.11 shows the plots of chemical shift positions of various protons of

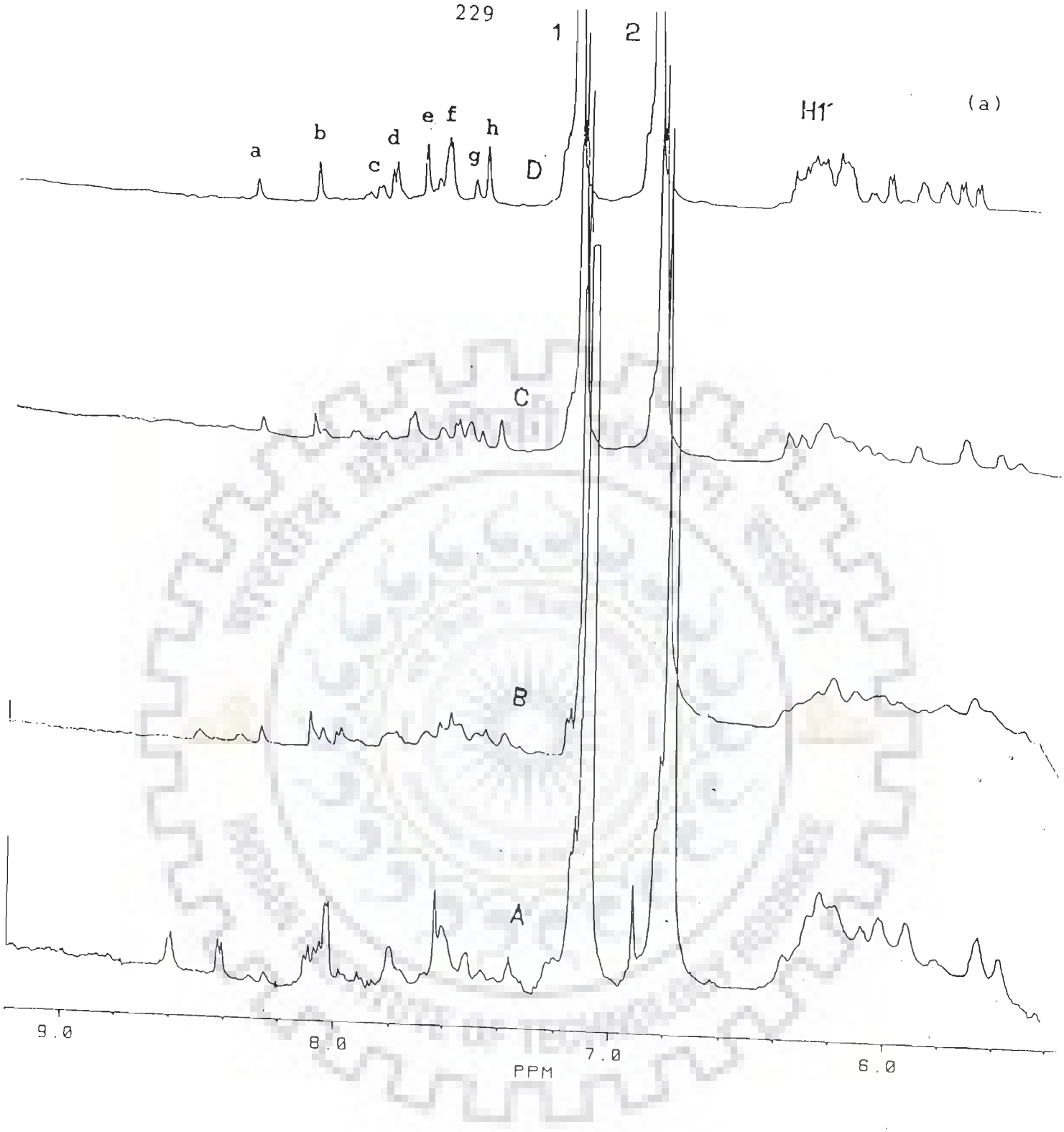
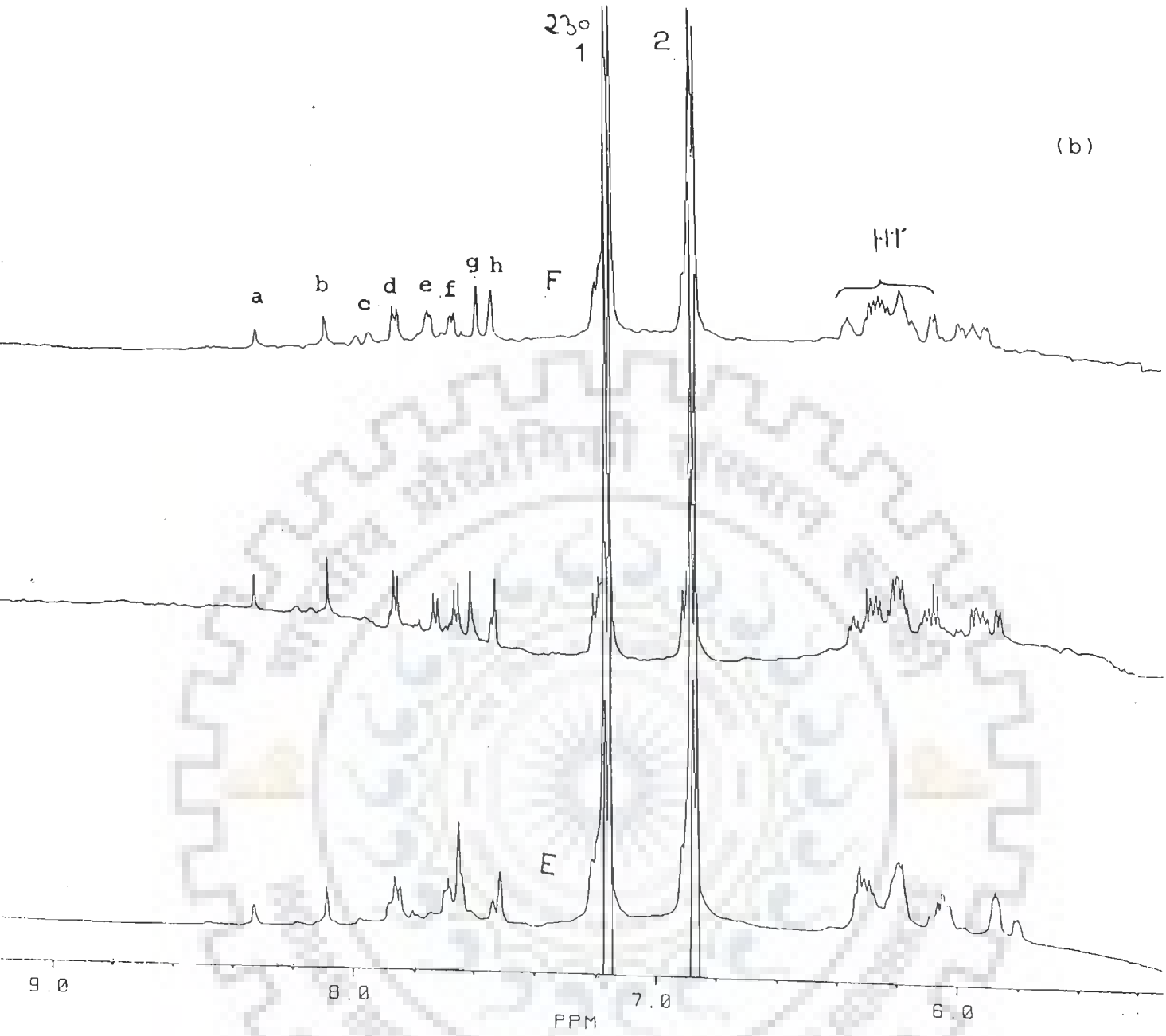


Fig 4.10(a & b):500 MHz proton NMR spectrum of a mixture of d-(GACTCGTC)$_2$ 8.60 mM and 0.69 mM KPYSLN in D_2O at different temperatures (A)275 K (B) 285 K (C)295 K (D)305 K (E)315 K (F)335 K. Peaks are labelled as follows :a -A2H8, b-A2H2, c-G1H8, d-G6H8, e-C5H6, f-C8H6, g-C3H6, h-T4H6, T7H6.


230
1
2
(b)
III
a
b
c
d
e
f
g
h
F
E
9.0
8.0
7.0
6.0
PPM

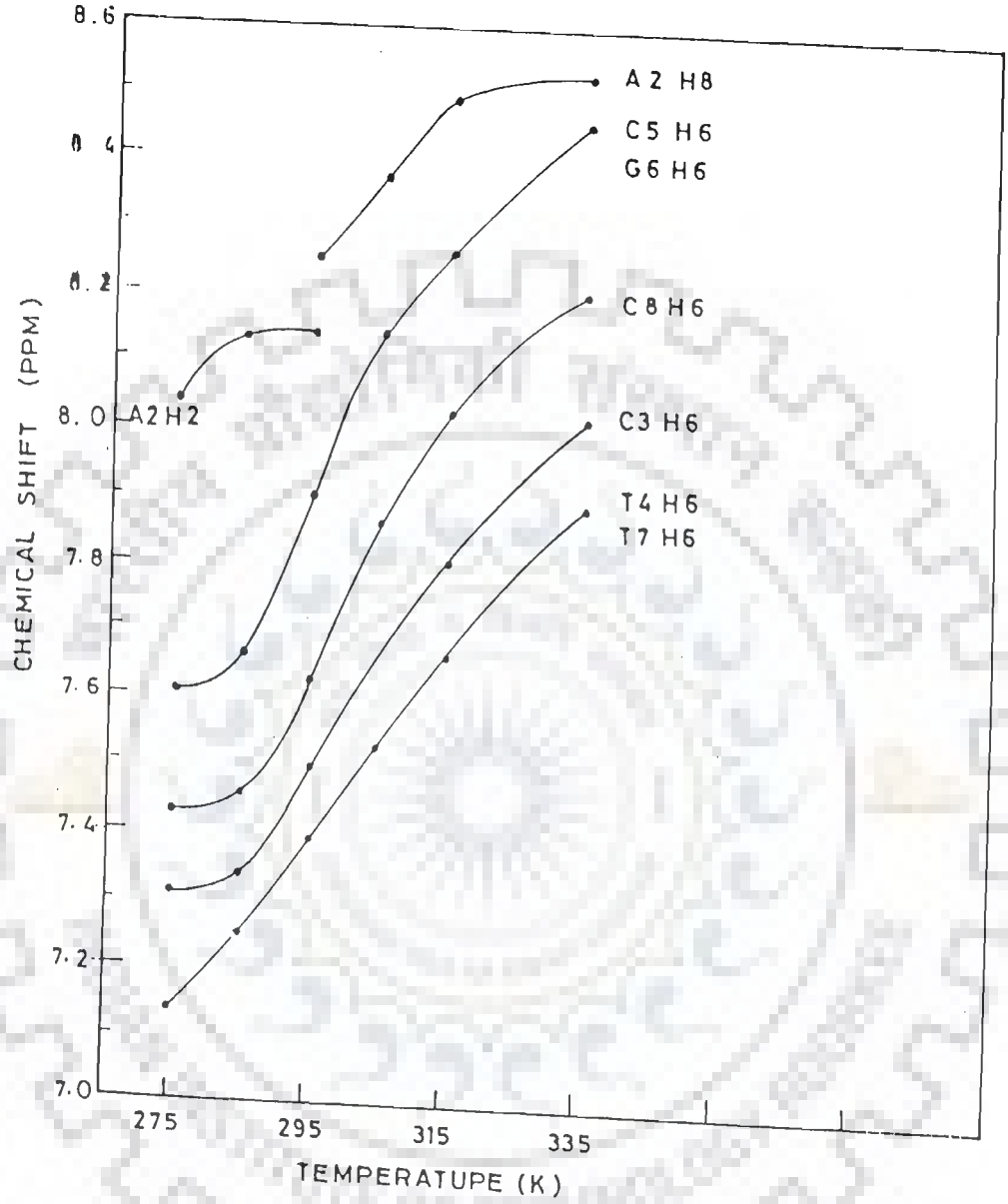


Fig. 4.11 :Chemical shift of base protons of d-$(GACTCGTC)_2$ in a mixture of 8.60 mM d-$(GACTCGTC)_2$ and 0.69 mM KPYSLN at P/N = 0.08 as a function of temperature (275 - 335 K).

octamer as a function of temperature (Table 4.9). It is found that most of the base protons and H1' protons shift downfield with temperature as in the case for unbound octamer. The T_m of various protons is as follows:

Base Protons	A2H8	C3H6	T4H6	C5H6	G6H8	T7H6	C8H6
Tm	310 K	305 K	305 K	300 K	300 K	305 K	303 K

There is no significant change in T_m or $\Delta\delta = \delta(335\ K) - \delta(275\ K)$ on interaction of octamer with hexapeptide.

Fig.4.11 shows the 1D NMR spectra of mixtures of octamer and KPYSLN at four different concentrations ratios P/N= 0.02, 0.04, 0.06 and 0.08. Tables 4.10a and 4.10b show the chemical shift and change in chemical shift of several protons on increasing P/N ratio from 0.02 to 0.08, respectively. For reference, the chemical shift of unbound octamer and KPYSLN are also shown against the P/N value of 0 in Table 4.10a. It is observed that the change in several nucleotide protons is insignificant except for shift of 0.06, -0.05 and -0.13 ppm in A2H2, C5H6 and G6H8 protons, respectively. Even at P/N=0.08, most of the octamer is likely to be present in free form as its concentration is nearly ten times greater than that of peptide. Therefore change in nucleotide protons with P/N is not expected. From Table 4.6b, it may be noted some of the sugar protons of C3, T4, C5 residues are also shifted in the range -0.04 to 0.12 ppm. From Tables 4.6d and 4.10a and b it is seen that there is no significant change in chemical shift of any of the peptide protons. It may be noted that G1H8 broadens on complexation and is not observable. Also G6H8 is broadened to a significant extent. Among the peptide peaks Tyr ring proton, Proδ-CH_2, Leuβ-CH_2, Leuγ-CH_2 are somewhat broadened. Further, Leuδ-CH_3 are very sharp peak in contrast to that observed on complexation of KPYSLN with $d(A)_5$.

Comparing all these results with that observed in case of binding of KPYSLN with $d(A)_5$, we find that the interaction with octamer is

Table 4.9: Chemical shift values of various protons of d-$(GACTCGTC)_2$ in the mixture of 8.60 mM d-$(GACTCGTC)_2$ and 0.69 mM KPYSLN at different temperatures.

Temp.(K)	G1H8	A2H8	A2H2	C3H6	T4H6	C5H6	G6H8	T7H6	C8H6
275	-	-	8.04	7.31	7.15	7.60	7.60	7.15	7.43
285	-	-	8.14	7.33	7.25	7.66	7.66	7.25	7.45
295	8.06	8.26	8.14	7.50	7.40	7.70	7.71	7.40	7.54
305	-	8.37	-	7.66	7.52	8.15	8.15	7.52	7.86
315	-	8.50	-	7.80	7.66	8.25	8.25	7.66	8.02
335	-	8.63	-	8.03	7.90	8.46	8.46	7.90	8.21

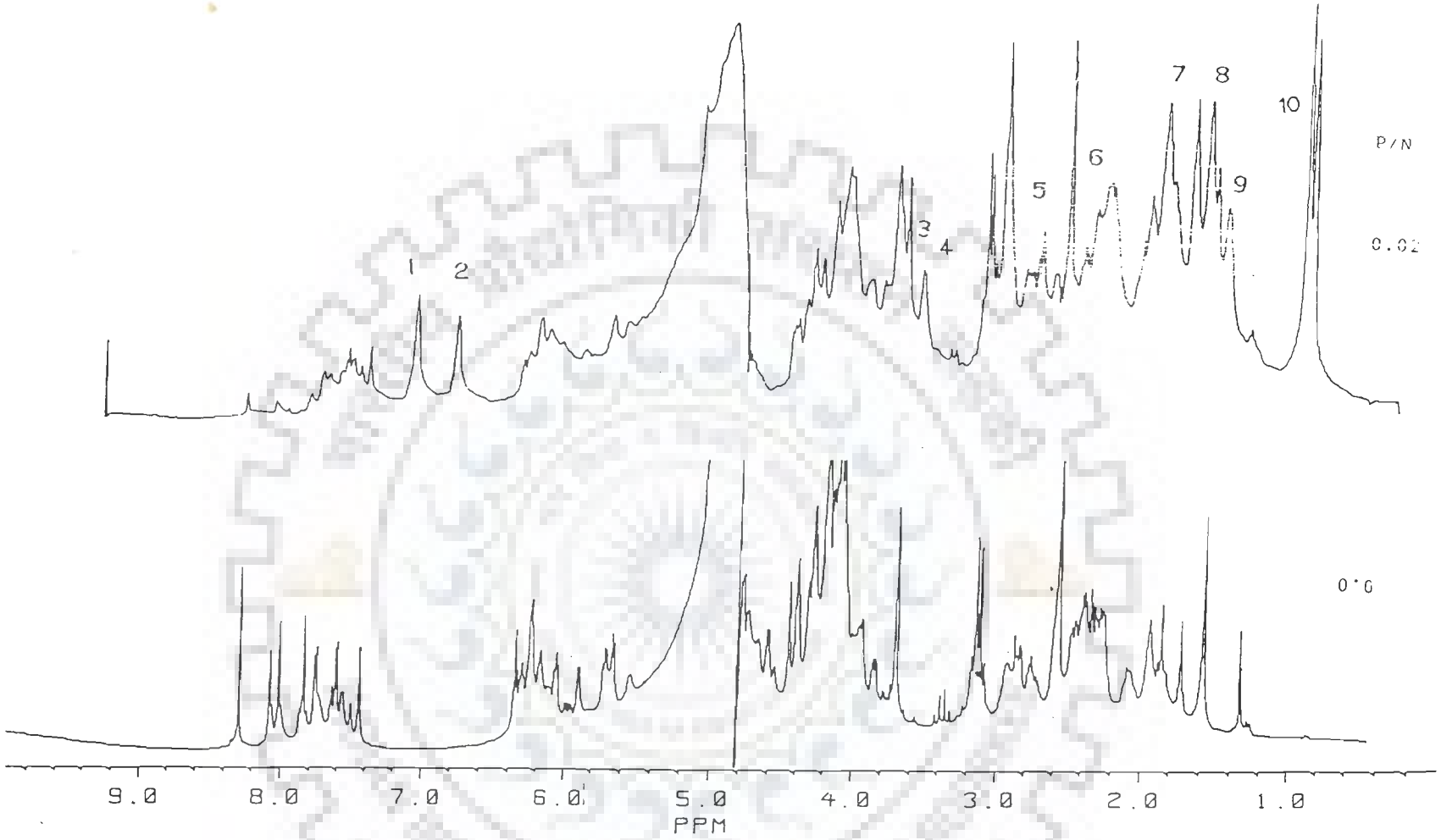
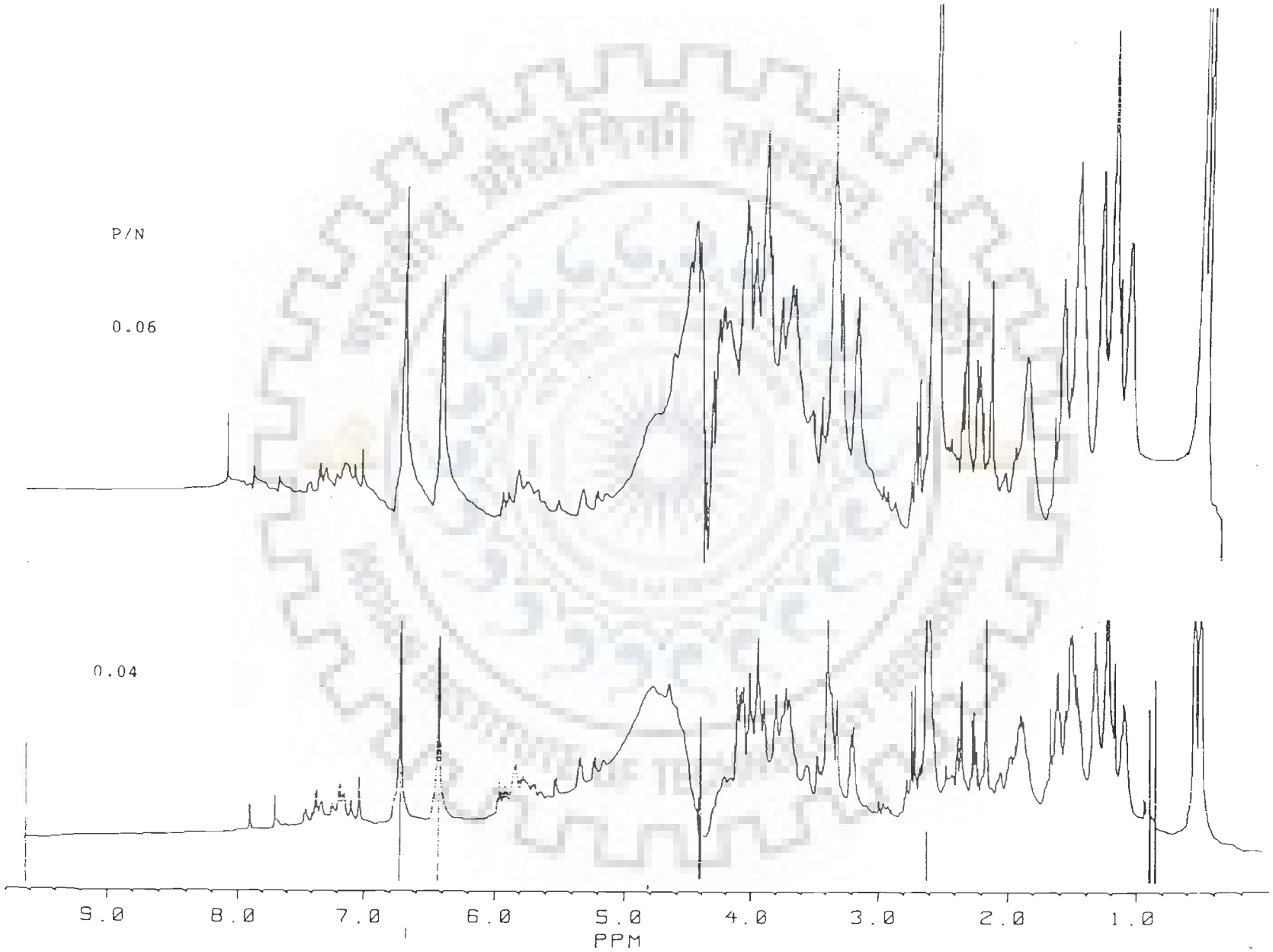


Fig.4.12: Series of one dimensional spectrum recorded at varying concentration ratio of peptide:oligonucleotide (P/N). Peaks are labelled as (1) Tyr (2,6)H, (2) Tyr (3,5)H (3)Ser β-CH_2 (4)Pro δ-CH_2 (5)Asn β-CH_2 (6) Pro β-CH_2 (7)Pro γ-CH_2, Lys β-CH_2, Lys δ-CH_2 (8)Leu β-CH_2, Leu γ-CH_2 (9) Lys γ-CH_2 (10) Leu δ-CH_3


P/N
0.06
0.04
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
PPM

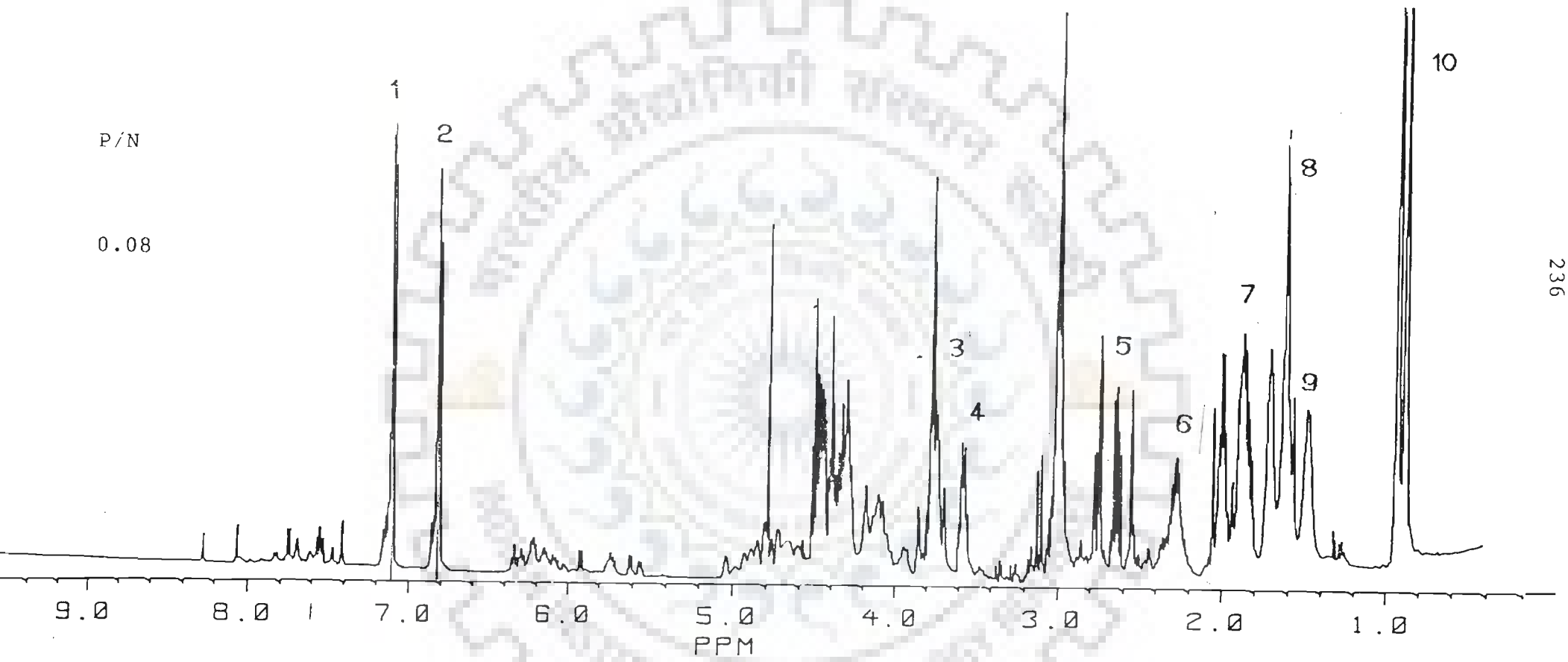

P/N
0.08
1
2
3
4
5
6
7
8
9
10
9.0
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
PPM

Table 4.10a: Chemical shift values of various nucleotide and peptide protons in the mixture of d-$(GACTCGTC)_2$ and KPYSLN at different concentration ratios.

P/N	A2H8	A2H2	C3H6	T4H6	C5H6	T7H6	C8H6
0*	8.27	8.06	7.50	7.41	7.75	7.41	7.56
0.02	8.27	8.06	7.53	7.38	7.71	7.38	7.57
0.04	8.27	8.06	7.51	7.40	7.70	7.40	7.55
0.06	8.26	8.06	7.50	7.40	7.70	7.40	7.55
0.08	8.26	8.14	7.50	7.40	7.70	7.40	7.54

P/N	Leuδ-CH_3	Lysγ-CH_2	Leuβ-CH_2	Lysδ-CH_2	Asnβ'-CH_2
0*	0.91	1.47	1.61	1.70	2.75
0.02	0.94	1.43	1.62	1.68	2.77
0.04	0.94	1.43	1.62	1.68	2.77
0.06	0.94	1.43	1.62	1.68	2.77
0.08	0.94	1.43	1.62	1.68	2.77

P/N	Lysε-CH_2	Serβ-CH_2	Proδ-CH_2
0*	2.98	3.75	3.56
0.02	3.01	3.77	3.58
0.04	3.01	3.77	3.58
0.06	3.02	3.77	3.56
0.08	3.77	3.77	3.56

* 10 mM d-GACTCGTC alone.

Table 4.10b: Changes in chemical shift values of various nucleotide and peptide protons in the mixture of d-$(GACTCGTC)_2$ and KPYSLN at different concentration ratios.

P/N	A2H8	A2H2	C3H6	T4H6	C5H6	T7H6	C8H6
0.02	0.00	-0.02	0.03	-0.03	-0.04	-0.03	0.01
0.04	0.00	-0.02	0.01	-0.01	-0.05	-0.01	-0.01
0.06	-0.01	0.00	0.00	-0.01	-0.05	-0.01	-0.02
0.08	-0.01	0.06	0.00	-0.01	-0.05	-0.01	-0.02

P/N	Leuδ-CH_3	Lysγ-CH_2	Leuβ-CH_2	Lysδ-CH_2	Asnβ'-CH_2
0.02	0.03	-0.04	0.01	-0.02	0.02
0.04	0.03	-0.04	0.01	-0.02	0.02
0.06	0.03	-0.04	0.01	-0.02	0.02
0.08	0.03	-0.04	0.01	-0.02	0.02

P/N	Lysε-CH_2	Serβ-CH_2	Proδ-CH_2
0.02	0.03	0.02	0.02
0.04	0.03	0.02	0.02
0.06	0.04	0.02	0.00
0.08	0.04	0.02	0.00

rather weak. Relatively small changes seen in G1, C3, T4, C5 and G6 residues may be due to a nonspecific binding to KPYSLN wherein some of the hydrophobic groups- Tyr ring, Proδ-CH_2, Leuβ-CH_2, Leuγ-CH_2 come close to these bases.

CHAPTER V

POTENTIAL ENERGY CALCULATIONS ON THE STACKING OF AROMATIC AMINO ACIDS WITH BASES AND BASE PAIRS

Stacking energies have been computed using second order perturbation method. The stacking of aromatic amino acids Trp, Tyr, Phe and His with nucleic acid bases A, T, G, C and base-pairs AT and GC have been looked into. The structure of nucleic acid bases and base-pairs have been taken from the results of single crystal X-ray diffraction studies (141). The geometries of amino acids have been taken from the published data (45,54). The charge distributions of these molecules have been calculated by the Complete Neglect of Differential Overlap (CNDO) method (112). The charges on various atoms of aromatic amino acids, bases and base-pairs have been shown in Figs.5.1 and 5.2 (a & b). The formulae used for the calculation of interaction energy are given in chapter II.

A Minimum energy configuration

In order to find the minimum energy and hence the most stable geometry of two stacked molecules, the relative orientation of molecules is represented in Fig.5.3. In the orthogonal coordinate system XYZ, the base moiety/base-pair is held rigid such that the planar base ring is in the X-Y plane with Z axis passing approximately through the centre of the base moiety. Initially aromatic amino acid is oriented parallel to the X, Y plane at a distance of 3.0 Å and then is optimised by varying separation of molecule along Z axis by an interval of 0.2 Å. The configurational space is scanned by rotating the aromatic amino acid about Z axis (angle 0°) at an interval of 10° and sliding it along X and Y axis for each orientation by a displacement of 0.1 Å. Further refinement of rotation of aromatic amino acids about Z-axis at an interval of 1° is done and interaction energy of each step is calculated.

B Stacking with bases and base-pairs

Table 5.1 summarises the calculated stacking energy values for

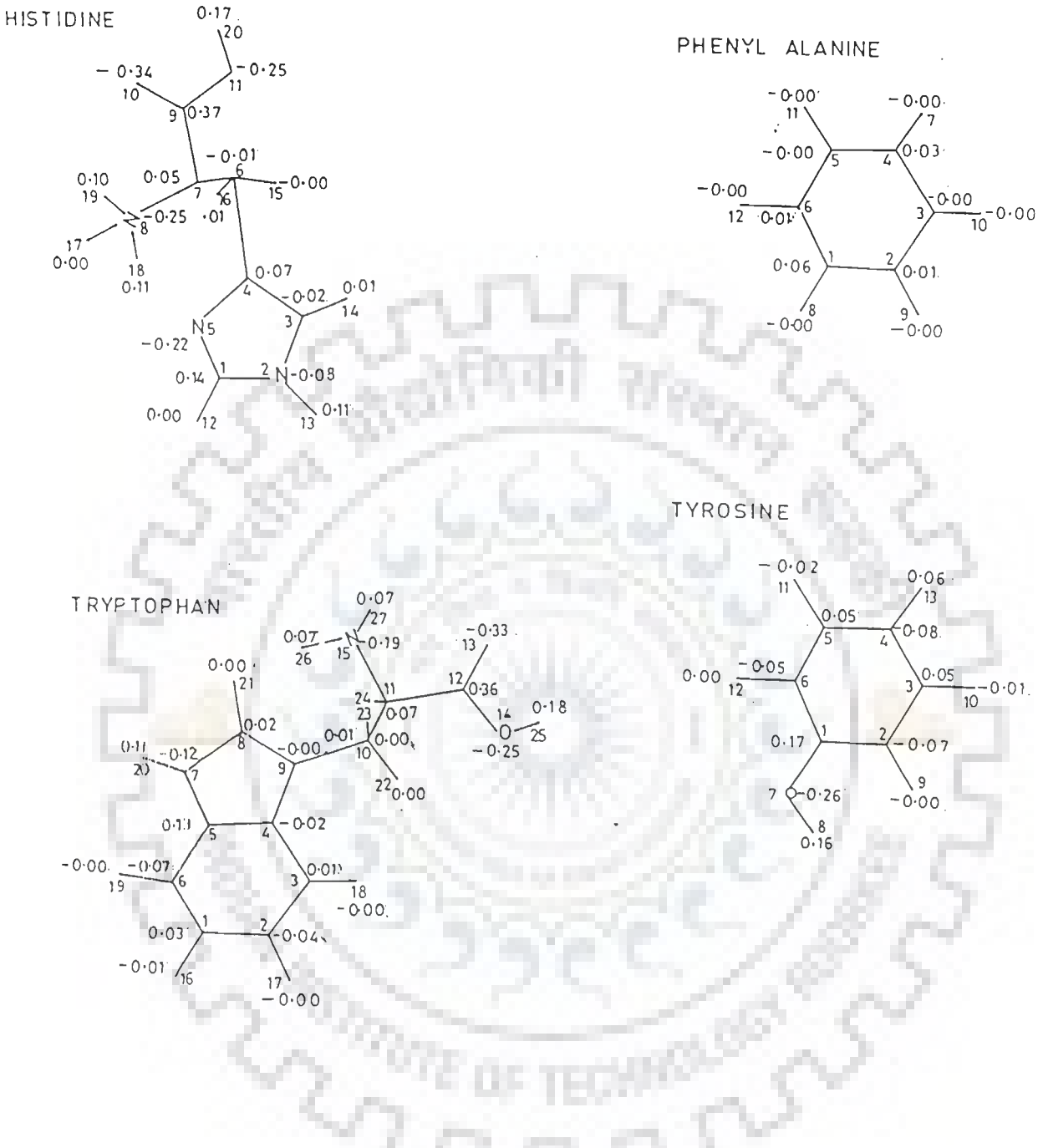


FIG. 5.1 Charge distribution on aromatic amino acids.

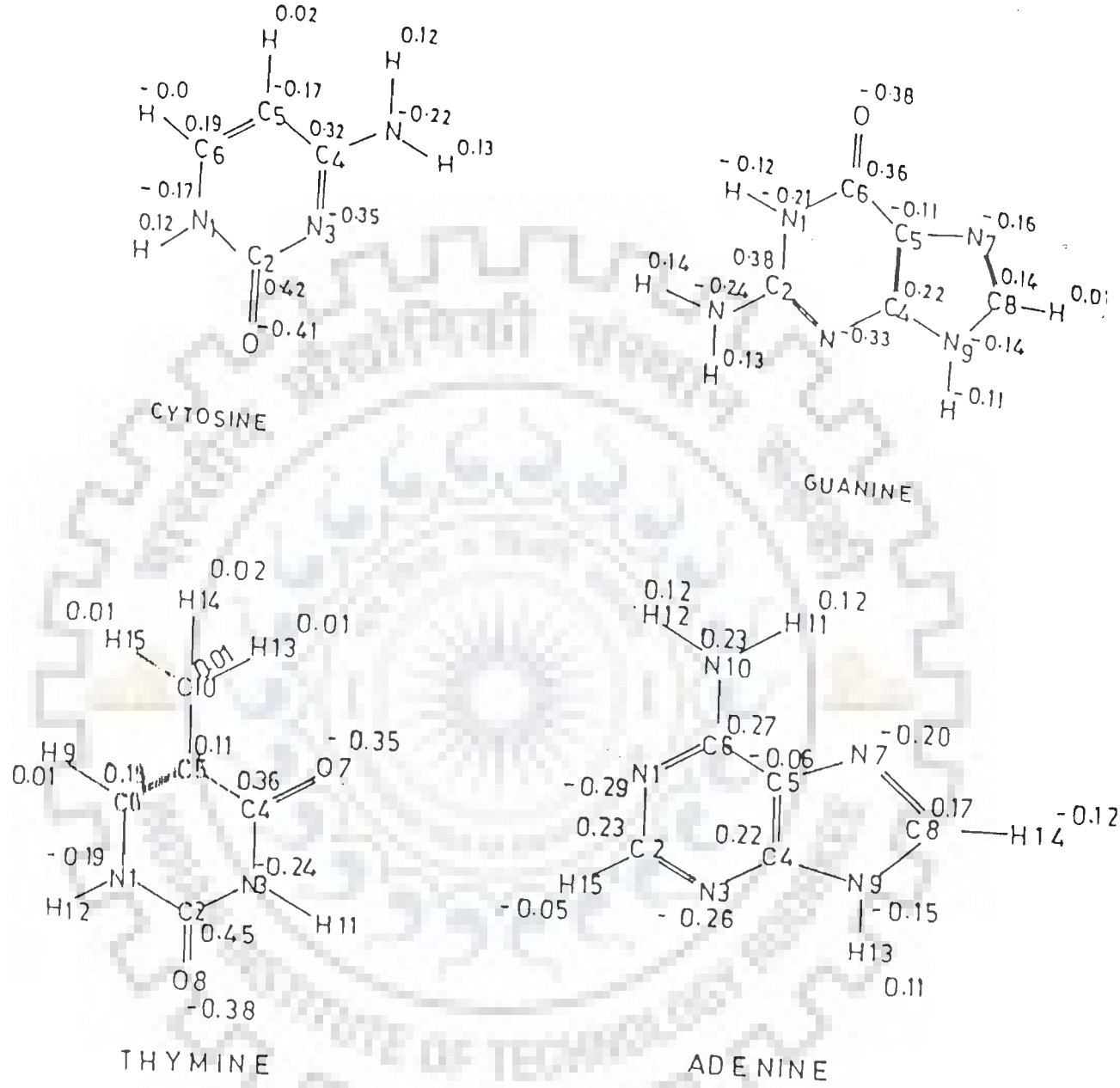


FIG. 5.2(a): Charge distribution on nucleic acid bases.

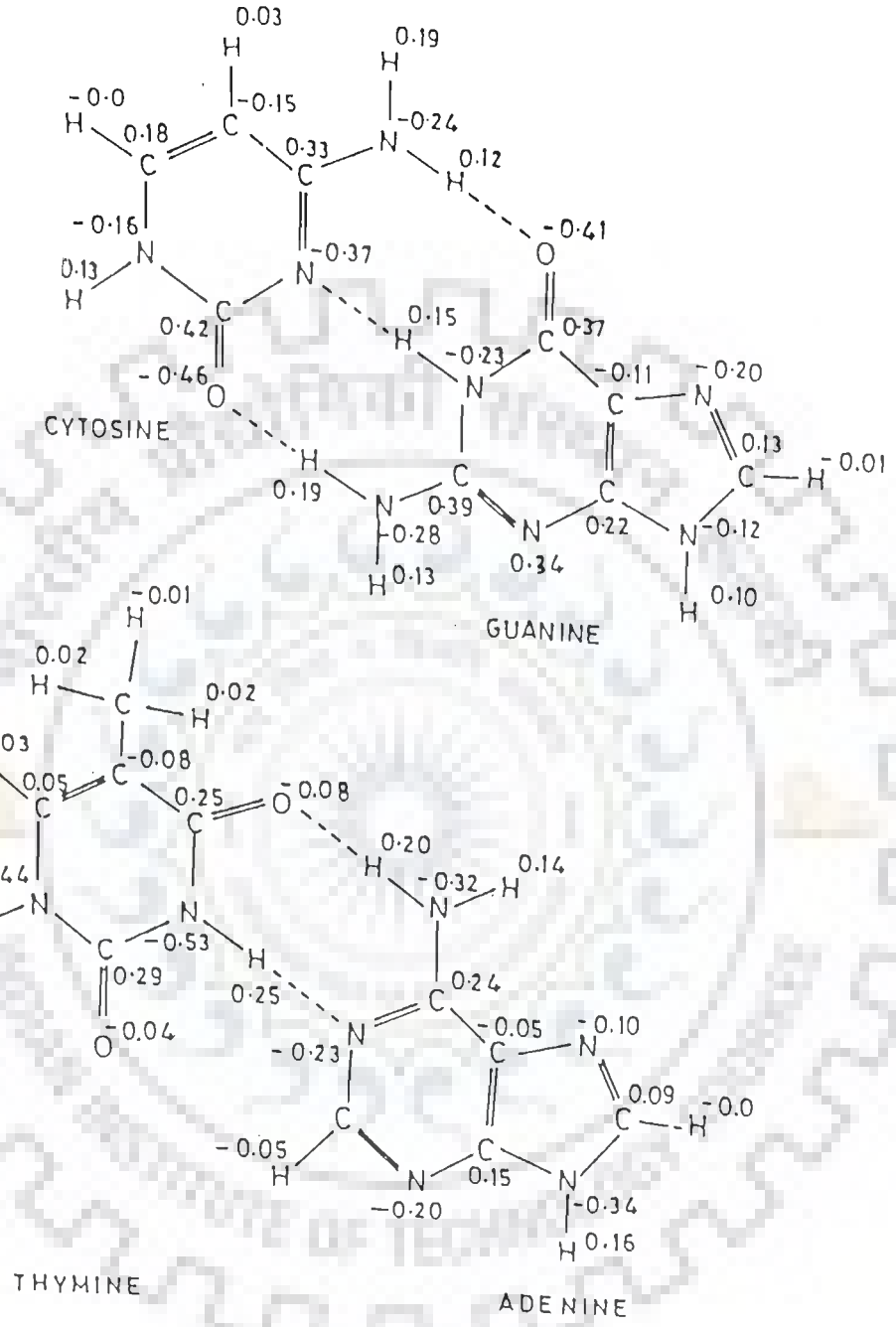


FIG. 5.2(b): Charge distribution on nucleic acid base-pairs.

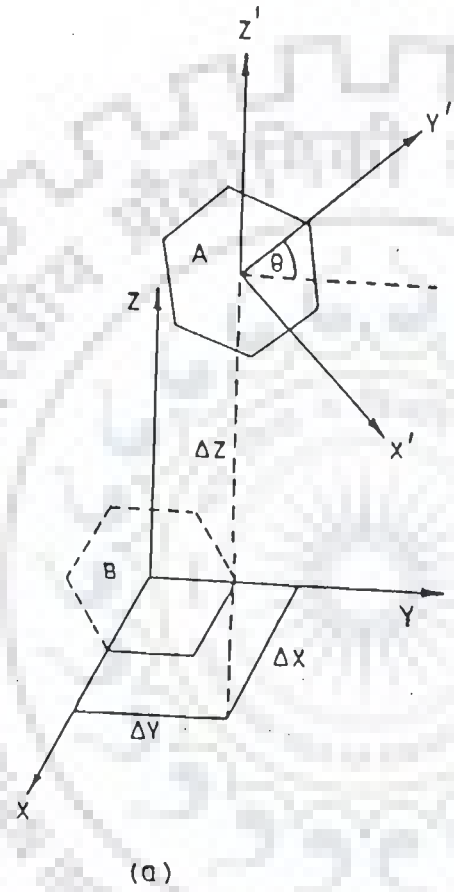


FIG. 5.3 Schematic representation of the relative orientation base/base-pairs with respect to aromatic amino acids.

TABLE 5.1 : Summary of stacking energies (Kcal/mole).

	Trp	Tyr	Phe	His
A	-9.2	-7.7	-6.3	-8.6
T	-12.1	-8.4	-7.3	-11.2
G	-10.8	-10.1	-7.8	-13.2
C	-11.5	-9.5	-6.3	-12.5
AT	-12.4	-9.7	-7.7	-12.0
CG	-16.1	-10.8	-8.4	-12.4

complexes of aromatic amino acids Trp, Tyr, Phe and His with nucleic acid bases (A, G, C, T) and base-pairs (AT, CG). Table 5.2 gives the partitioning of total interaction energies into various components for these complexes. It is observed that the maximum contribution to stacking interactions comes from the dispersion term. The results are in agreement with those reported in the literature (90). The energy values associated with the complexes of bases show variation which is of the order of 1-1.5 Kcal/mole. Comparison of stacking energies of the complexes of aromatic amino acids with bases reveal that His and Trp are involved in more efficient stacking than Tyr and Phe. Generally, His complexes appear to have minimum energy. It may be noted that stacked complexes of His with G and C are associated with stronger binding energies than that with A and T. The electrostatic term seems to be responsible for such behaviour. Complexes of Tyr with bases are more stable than the corresponding complexes of Phe. The relative order of binding energies of the complexes of aromatic amino acid with bases, generally, follows the pattern His > Trp > Tyr > Phe.

The interaction energies of base-pairs AT and CG with aromatic amino acids decrease in the order Trp > His > Tyr > Phe (Table 5.1 and Table 5.2). The difference in stacking energies has been observed to be 2 Kcal/mole between His and Tyr complexes. Same is true for Tyr and Phe complexes. Among base-pairs, CG complexes are always associated with lower energies. Overlap geometries of aromatic amino acids with bases are shown in Fig.5.4. These geometries correspond to the minimum energies (Table 5.1). Fig.5.4 shows that among bases, the overlapping decreases in the order His > Trp > Tyr > Phe. It may be noted that the overlapping is more for His complexes. Overlap geometries of base-pairs with aromatic amino acids are shown in Fig.5.5. It is found that the overlap geometries of the stacks involving base-pairs are different from those of component bases. With CG base-pair, the overlap of aromatic ring with G is relatively greater than C except for Trp. While for AT base-pair, the overlap of amino acid is more with A than T.

TABLE 5.2 : Partitioning of stacking energy (Kcal/mole) into various components. D is the optimised distance between the base and aromatic amino acid in Å

	D	E_{el}	E_{pol}	E_{disp}	E_{rep}	E_{tot}
A-Trp	3.4	-1.7	-0.7	-11.7	4.9	-9.2
A-Tyr	3.3	-2.5	-0.6	-9.6	4.9	-7.7
A-Phe	3.4	-1.2	-0.5	-8.8	4.2	-6.3
A-His	3.2	-3.5	-1.0	-9.8	5.7	-8.6
T-Trp	3.3	-3.6	-1.7	-13.1	6.3	-12.1
T-Tyr	3.2	-3.1	-1.0	-10.0	5.7	-8.4
T-Phe	3.3	-1.7	-1.0	-9.5	5.0	-7.3
T-His	3.0	-6.3	-1.9	-9.6	6.6	-11.2
G-Trp	3.4	-1.6	-1.4	-15.2	7.5	-10.8
G-Tyr	3.4	-3.6	-1.6	-10.2	5.4	-10.1
G-Phe	3.4	-1.8	-1.2	-9.0	4.2	-7.8
G-His	3.3	-6.8	-1.4	-11.1	6.0	-13.2
C-Trp	3.2	-2.8	-2.9	-14.7	8.9	-11.5
C-Tyr	3.2	-4.3	-1.3	-9.0	5.1	-9.5
C-Phe	3.3	-0.7	-1.6	-8.3	4.3	-6.3
C-His	3.0	-7.7	-2.1	-9.1	6.3	-12.5
AT-Trp	3.4	-2.7	-1.1	-15.7	7.1	-12.4
AT-Tyr	3.2	-2.8	-0.6	-12.2	5.9	-9.7
AT-Phe	3.3	-0.9	-0.4	-10.5	4.2	-7.7
AT-His	3.2	-3.4	-1.1	-14.0	6.5	-12.0
CG-Trp	3.2	-4.7	-2.3	-17.1	8.0	-16.1
CG-Tyr	3.2	-3.1	-0.6	-12.9	5.7	-10.8
CG-Phe	3.2	-0.9	-0.8	-12.0	5.2	-8.4
CG-His	3.2	-4.6	-1.2	-13.5	6.5	-12.9

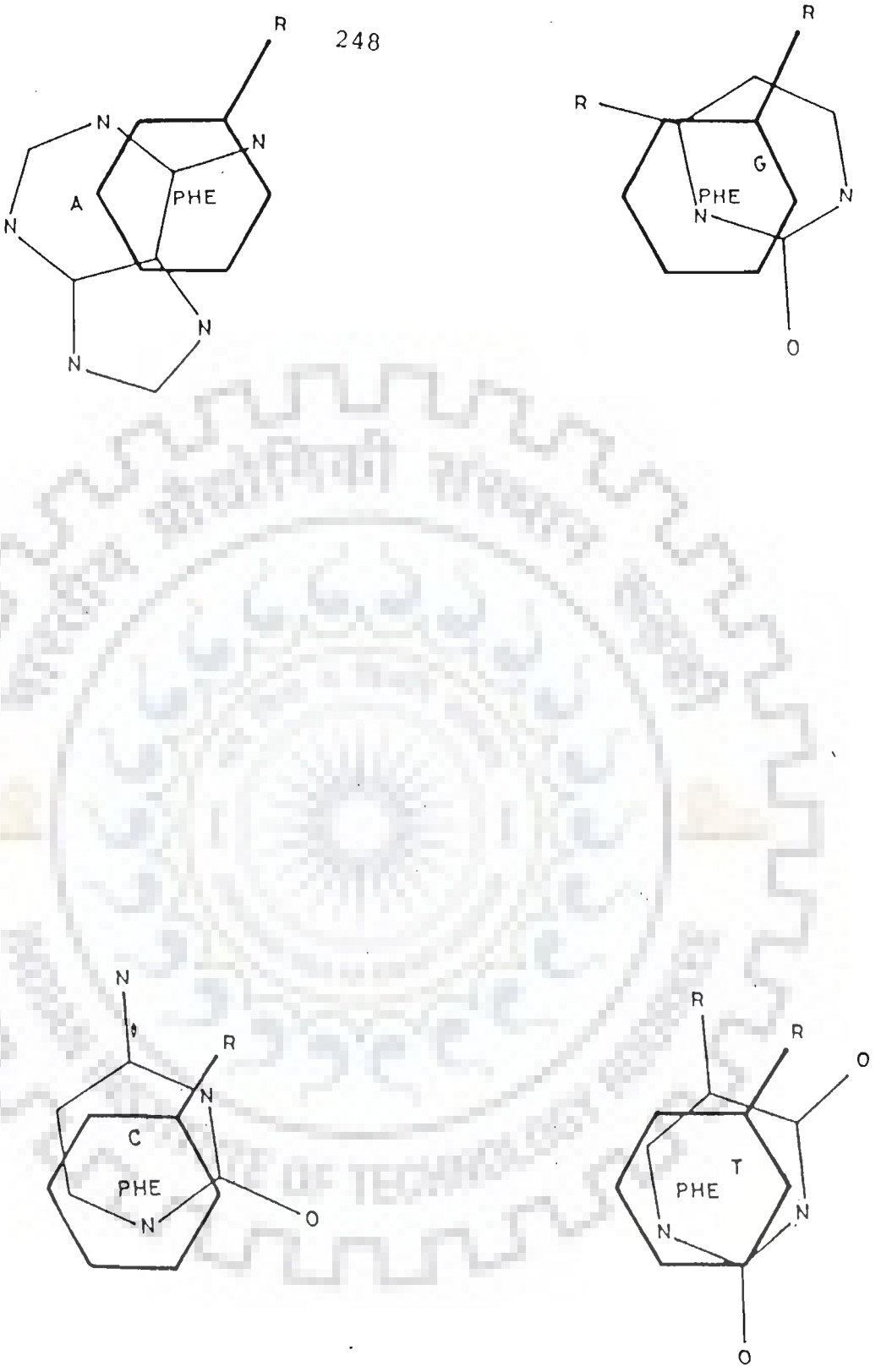
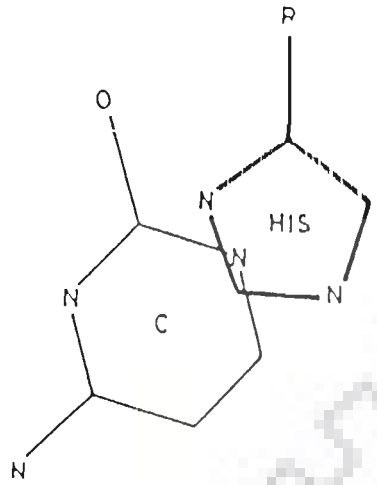
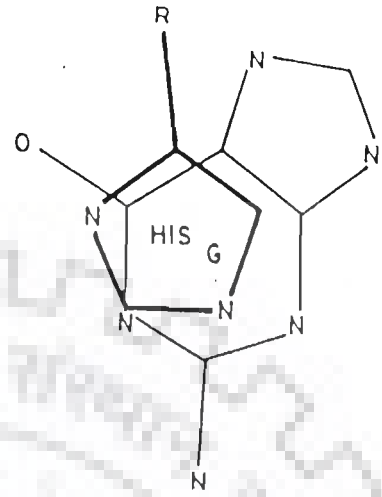
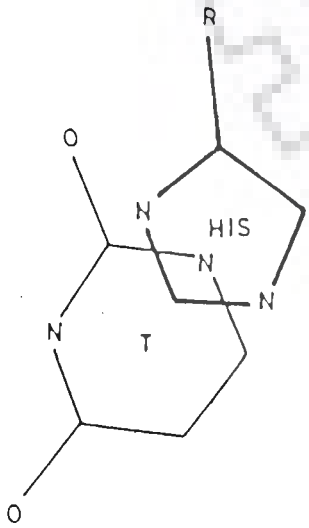
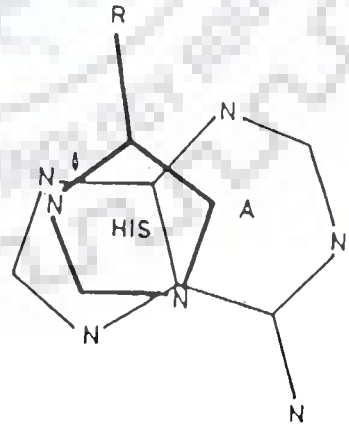


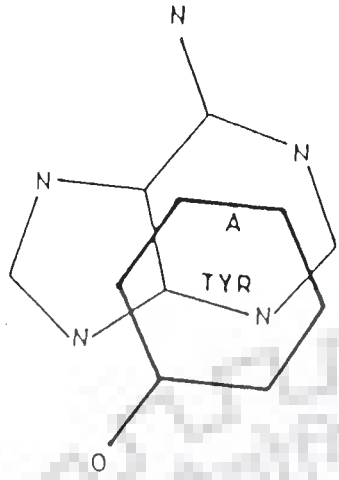
FIG. 5.4 Overlap geometries of A, T, G and C with Trp, Tyr, Phe and His.

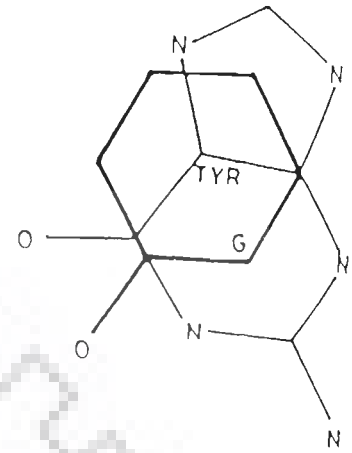

R
O
N
HIS
N
N
C
N
N

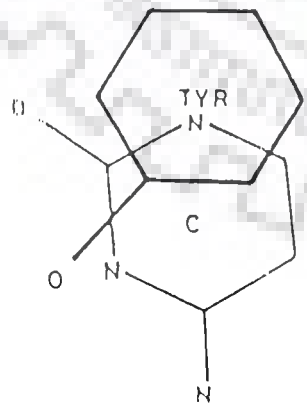

R
N
N
O
N
HIS
G
N
N
N
N

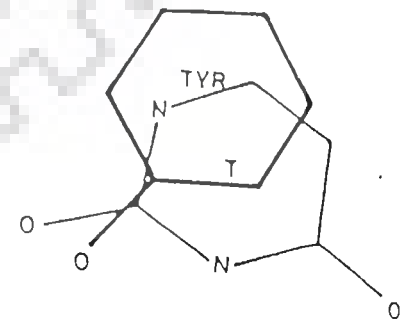

R
O
N
HIS
N
N
N
T
O

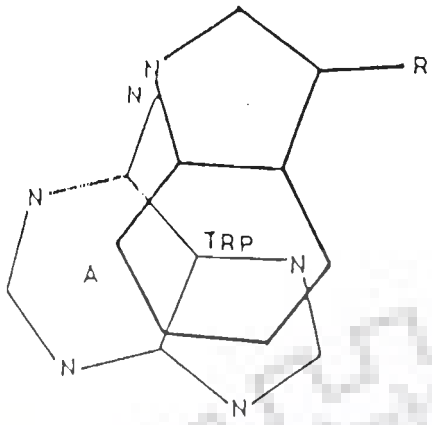

R
N
N
N
HIS
A
N
N
N

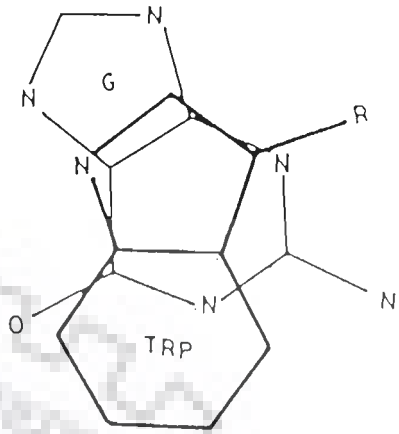

N
N
N
A
TYR
N
N
O

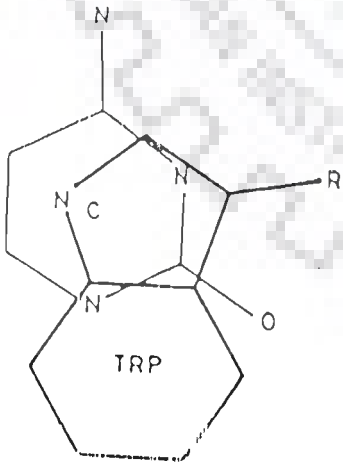

N
N
TYR
O
G
N
O
N
N

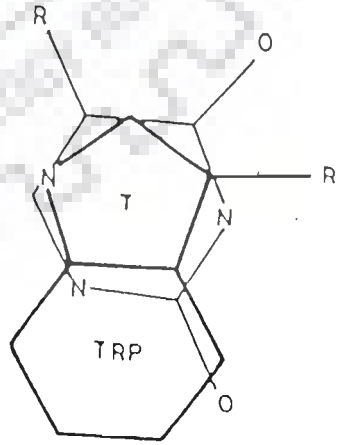

TYR
N
O
C
O
N
N


TYR
N
T
O
O
N
O









Contd....

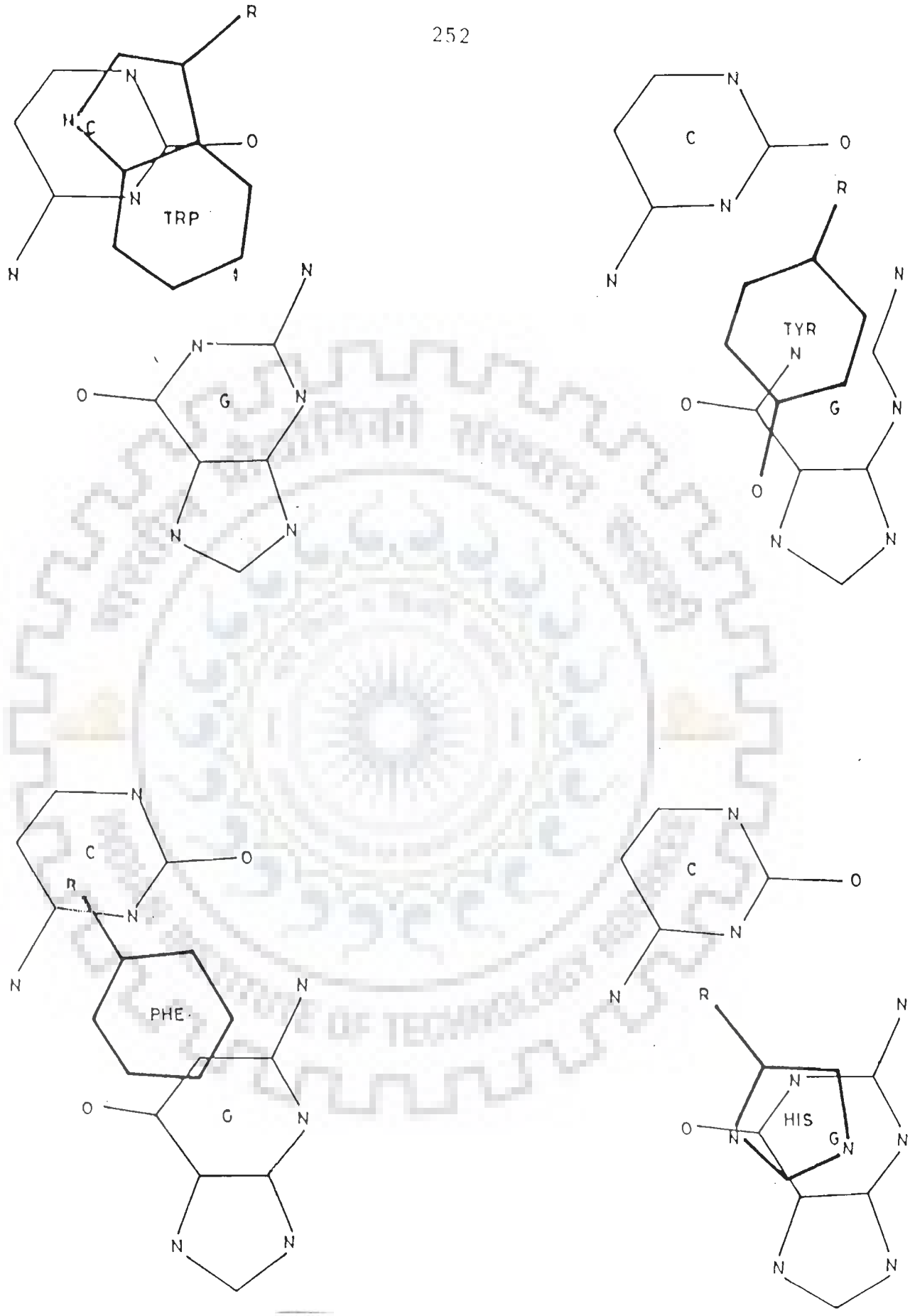
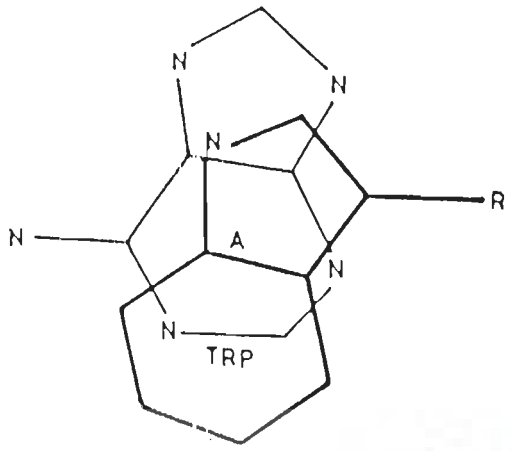
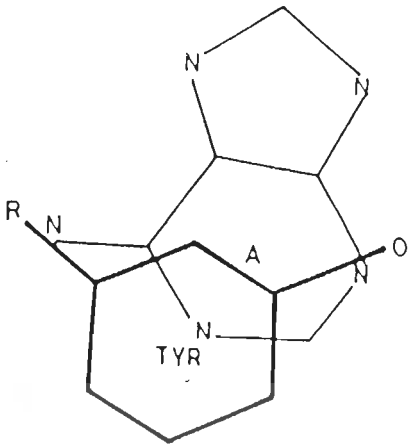
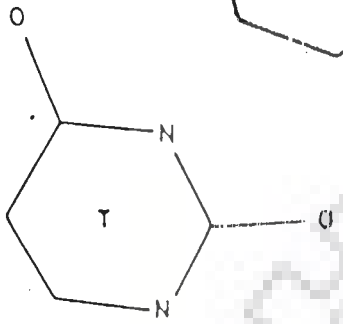
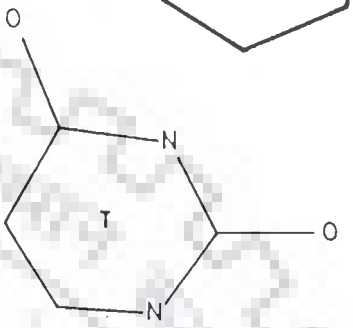


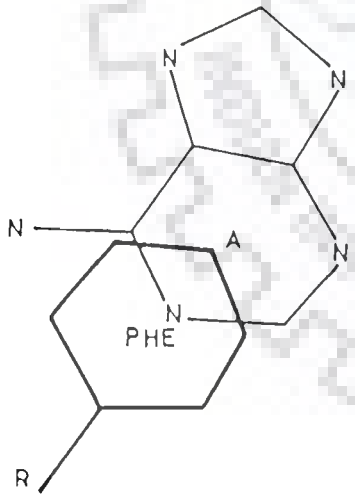
FIG. 5.5 Overlap geometries of AT and GC base-pairs with Trp, Tyr, Phe and His.

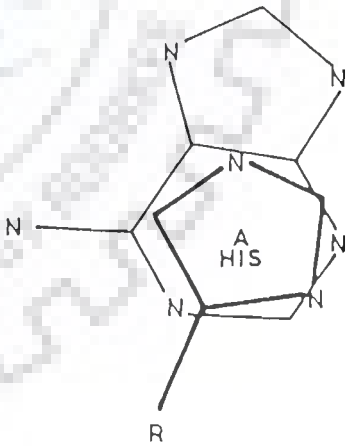

N
N
N
R
N
A
N
N
TRP

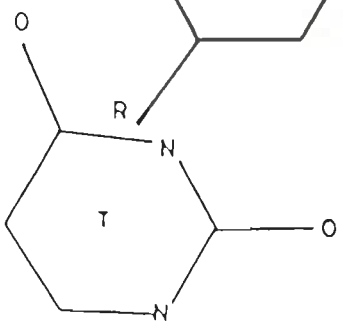

N
N
R
N
A
O
N
N
TYR

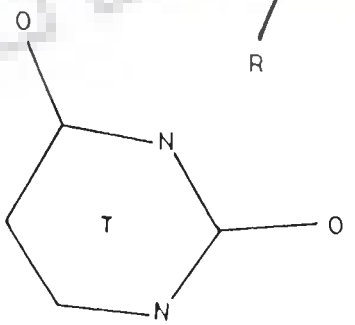

O
N
T
O
N


O
N
T
O
N


N
N
N
A
N
N
PHE
R


N
N
N
A
HIS
N
N
N
R


O
N
T
O
N


O
N
T
O
N

The experimental results available in the literature support our findings. It has been shown that solubility enhancement of guanosine in presence of aromatic amino acid decreases in the order Trp > Tyr > Phe (10). The binding behaviour of 5'GMP with polymer containing aromatic amino acid has been studied which suggests that the contribution of the aromatic residues to the binding affinity of nucleotide decreases in the order Trp > Tyr > Phe (11).

The stacking energy is the sum of electrostatic, dispersion and repulsion energies. Contributions of these energy terms to the total energy has been given for the optimized conformations in Table 5.2. The predominant contributions to the total interaction energy comes from dispersion energy. This energy is greater than the total energy in all the complexes. The electrostatic and polarisation energies are smaller and lesser than the repulsion energy.

These studies on the stacking of nucleic acid bases and base-pairs have important implications. Among base-pairs, CG base-pair forms more stable complex than AT base-pair. Among aromatic amino acids, tryptophan and histidine form minimum energy configurations which have energy lower than the corresponding minimum energy configurations for tyrosine and phenylalanine. Thus specificity in protein-nucleic acid associations can be explained on the basis of these results .

CHAPTER VI

THEORETICAL STUDIES ON THE INTERCALATION OF AROMATIC AMINO ACIDS BETWEEN BASE-PAIRS OF DINUCLEOTIDE MODEL SYSTEMS

The present chapter deals with the interaction of Trp, Tyr, Phe and His when placed between two CG, CC, AT and AA base-pairs. Interaction energies have been calculated using the formulae described elsewhere (Chapter II).

A Minimum energy configuration

The search for minimum energy configuration was made by varying the relative orientation of amino acids with respect to base-pair, that is, angle θ and space coordinates X and Y without considering the phosphate backbone.

The following algorithm was used. The distance between two base-pairs (Z) was increased from 6.0 to 6.8 Å in steps of 0.2 Å. For the optimum distance, the configurational space was scanned by varying θ by 10° and sliding it along X and Y-axis for each orientation by a displacement of 0.1 Å and calculating interaction energy of each step. For the energy minimum thus obtained, global search for unwinding angle, $\Delta\alpha$, is carried out in a range of $+180^{\circ}$ to -180° in steps of 20°. Local refinement of unwinding angle value is done in steps of 1°.

B Interaction energies on intercalation of aromatic amino acids between two base-pairs

Fig.6.1 shows the schematic representation of aromatic ring of amino acid, X, (X = Trp, Tyr, Phe and His) positioned exactly midway between the two base-pairs having a relative orientation, α, of 36°. Minimum energy of -28.8 Kcal/mole is observed for Trp for a distance of 6.6 Å between the base-pairs, CG and CG. For Tyr, Phe and His the minimum potential energy values are found to be -22.0, -19.0 and -29.2 Kcal/mole, corresponding to a distance of 6.4 Å, 6.4 Å and 6.2 Å, respectively (Table 6.1). Such a result

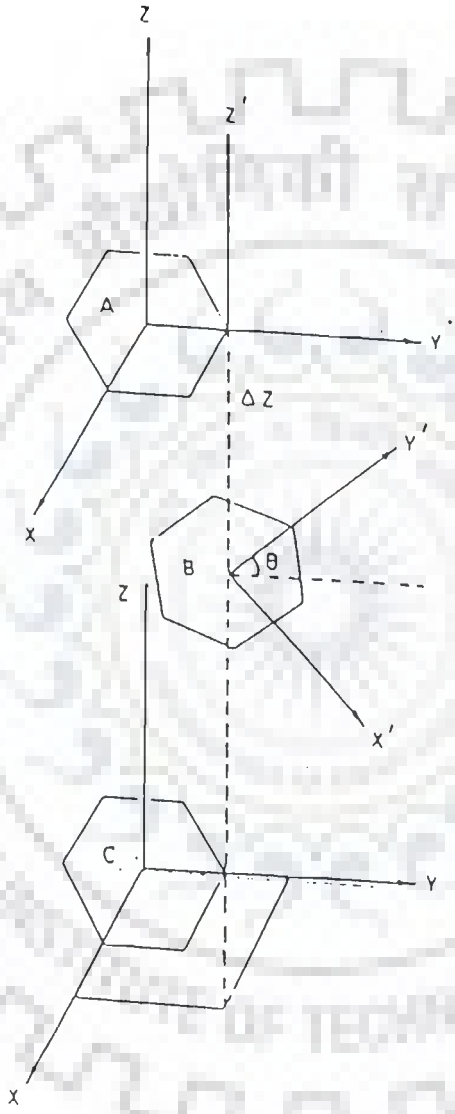


FIG. 6.1 Schematic representation of the relative orientation of two base-pairs with aromatic amino acid intercalated inbetween.

TABLE 6.1 : Summary of stacking energies (Kcal/mole)

	Trp	Tyr	Phe	His
CG / CG	-29.0	-22.0	-19.0	-29.2
AT / TA	-20.7	-18.9	-14.8	-20.6
AT / AT	-22.5	-19.9	-15.5	-20.1
CG / GC	-21.9	-18.3	-13.9	-16.7

TABLE 6.2 : Summary of unwinding angles (degree)

	Trp	Tyr	Phe	His
CG / CG	-22	-41	-35	-43
AT / TA	-30	36	34	28
AT / AT	-24	-40	-33	-43
CG / GC	-6	36	-34	7

is consistent with expectation, that in order to accommodate an aromatic amino acid moiety, the nucleic acid base-pairs must reorganize by opening out from their usual stacking distance of 2.9-3.4 Å to 6.2-6.8 Å in order to penetrate an intercalation site (110).

Fig.6.2 shows the effect of translating the amino acid with respect to the two base-pairs held rigidly at a distance of optimum Z (as obtained in Fig. 6.1(a)) with relative orientation among themselves to be 36°(α). For each amino acid, it is seen that the interaction energy increases rapidly as the aromatic ring of amino acid moves away from the centre of nucleic acid base-pairs in either X or Y direction. This is understandable as the energy is expected to be minimum in the orientation in which the overlap of aromatic amino acid is maximum with base-pairs.

Another parameter involved in intercalation is the change in angular disposition of the bases on either side (unwinding angle). Fig.6.3 shows the results of a global search for the best unwinding angle for Trp, Tyr, Phe and His intercalated between different sets of base-pairs. For example for Tyr intercalated between CG and CG base-pairs, minimum energy configurations are observed for three sets of unwinding angles i.e. -41°, 80° and -140°. Of the three values, the minima at -140°and +80° are sterically forbidden in presence of phosphate backbone. Further a large change in the overlap angles may cause long range disorders in double-stranded nucleic acids. Thus the configuration corresponding to an angle of -41° with a total potential energy of -22 Kcal/mole is the optimised minima. In a similar way for Trp, Phe and His, the values of best unwinding angle Δα, are found to be -22°,-35° and -43°, respectively.

It is found that the angles vary in range -43° to 36°(Table 6.2). It may be emphasised here that the results obtained are the first attempt to study these types of interactions with aromatic amino acids on the basis of classical potential functions. Reports on

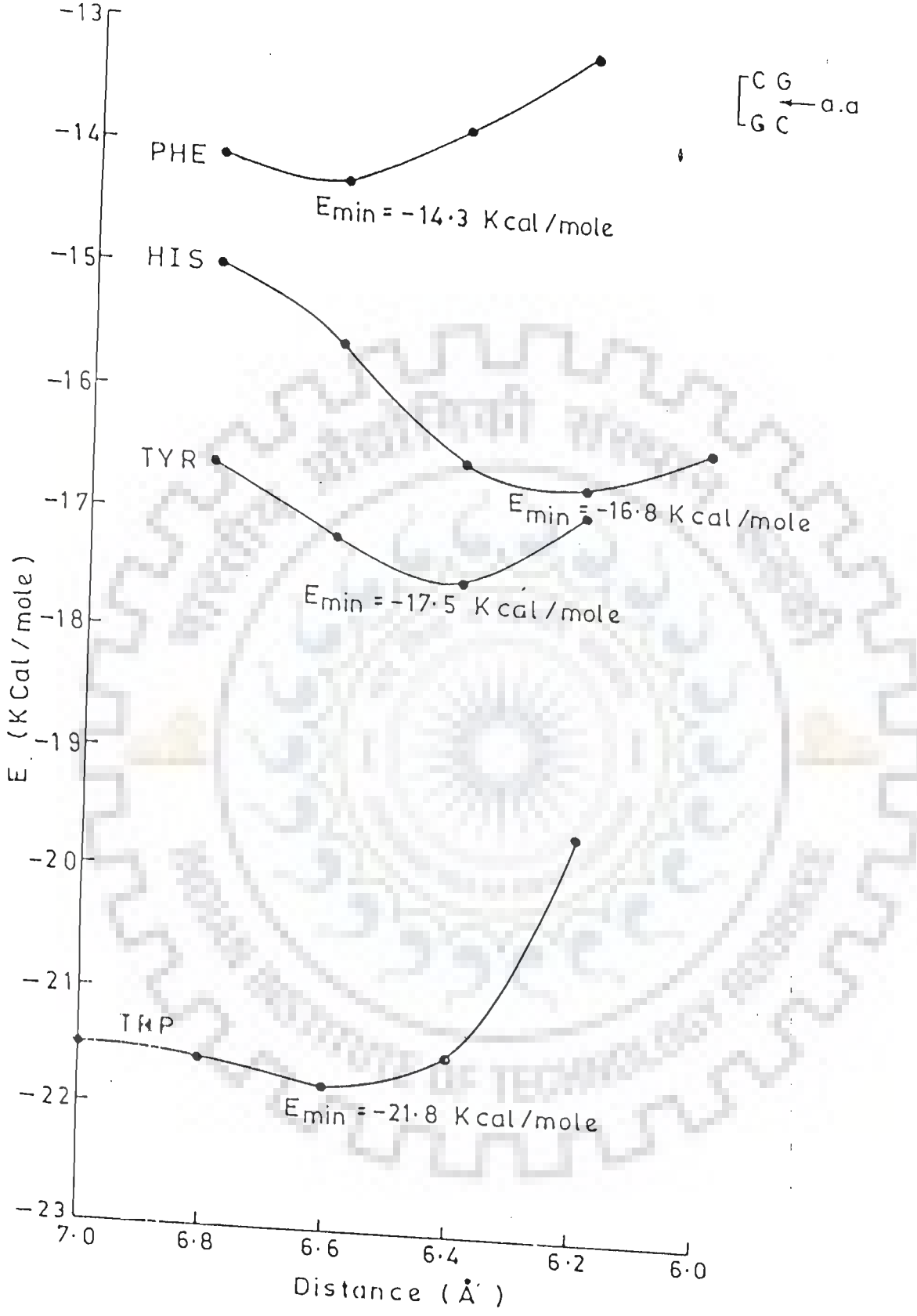
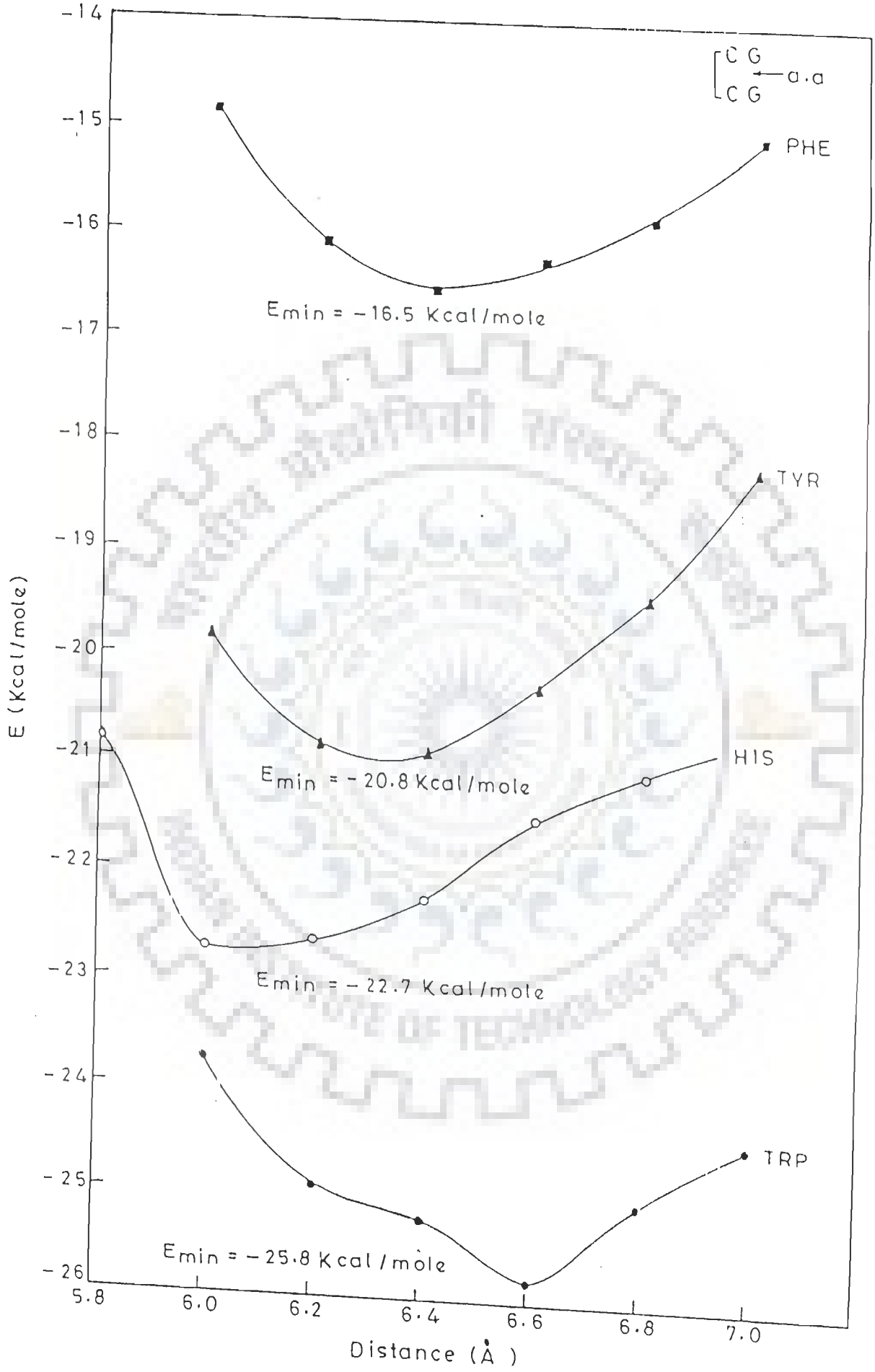
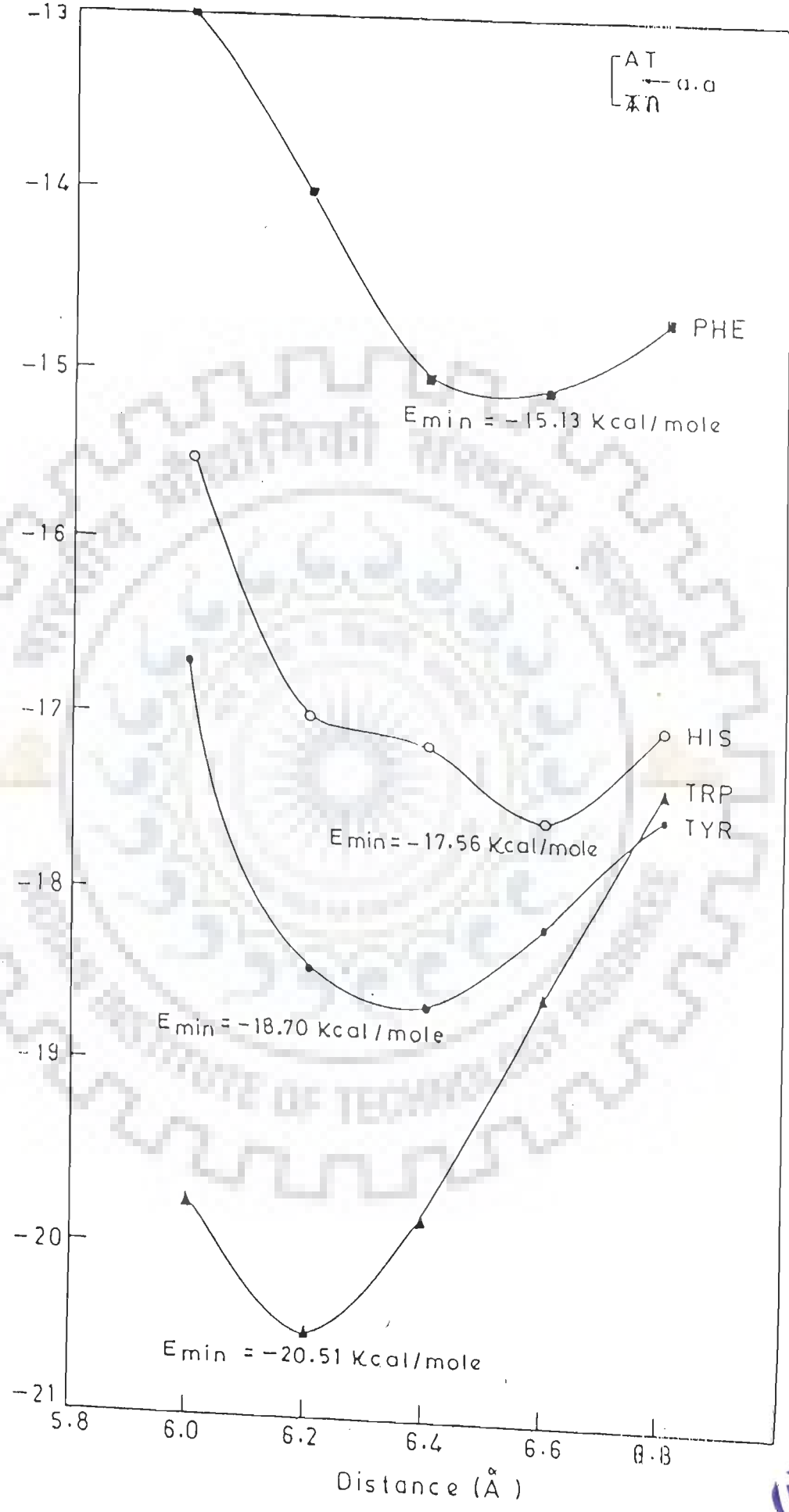
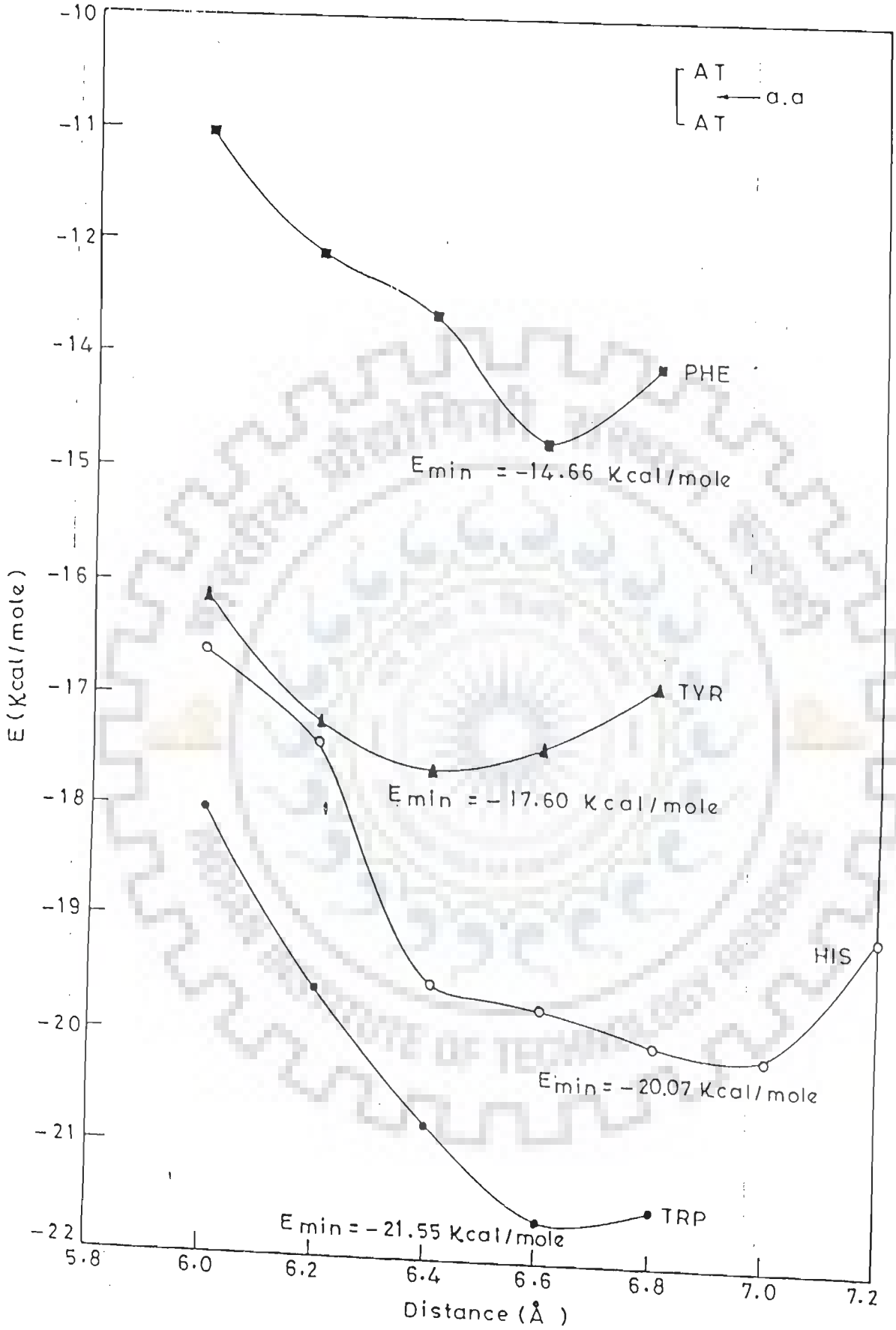


FIG. 6.2 Interaction energy as a function of distance between two base-pairs (a) CG & GC (b) CG & CG (c) AT & TA and (d) AT & AT due to intercalation of aromatic amino acids Trp, Tyr, Phe and His.


C G
C G
a.a
PHE
Emin = −16.5 Kcal/mole
TYR
Emin = −20.8 Kcal/mole
HIS
Emin = −22.7 Kcal/mole
TRP
Emin = −25.8 Kcal/mole
E (Kcal/mole)
−14
−15
−16
−17
−18
−19
−20
−21
−22
−23
−24
−25
−26
5.8
6.0
6.2
6.4
6.6
6.8
7.0
Distance (Å)


AT
a.a
PHE
Emin = −15.13 Kcal/mole
HIS
TRP
TYR
Emin = −17.56 Kcal/mole
Emin = −18.70 Kcal/mole
Emin = −20.51 Kcal/mole
E (Kcal/mole)
−13
−14
−15
−16
−17
−18
−19
−20
−21
5.8
6.0
6.2
6.4
6.6
8.8
Distance (Å)

Contd....


AT
a.a
AT
PHE
Emin = -14.66 Kcal/mole
TYR
Emin = -17.60 Kcal/mole
HIS
Emin = -20.07 Kcal/mole
TRP
Emin = -21.55 Kcal/mole
E (Kcal/mole)
-10
-11
-12
-14
-15
-16
-17
-18
-19
-20
-21
-22
5.8
6.0
6.2
6.4
6.6
6.8
7.0
7.2
Distance (Å)

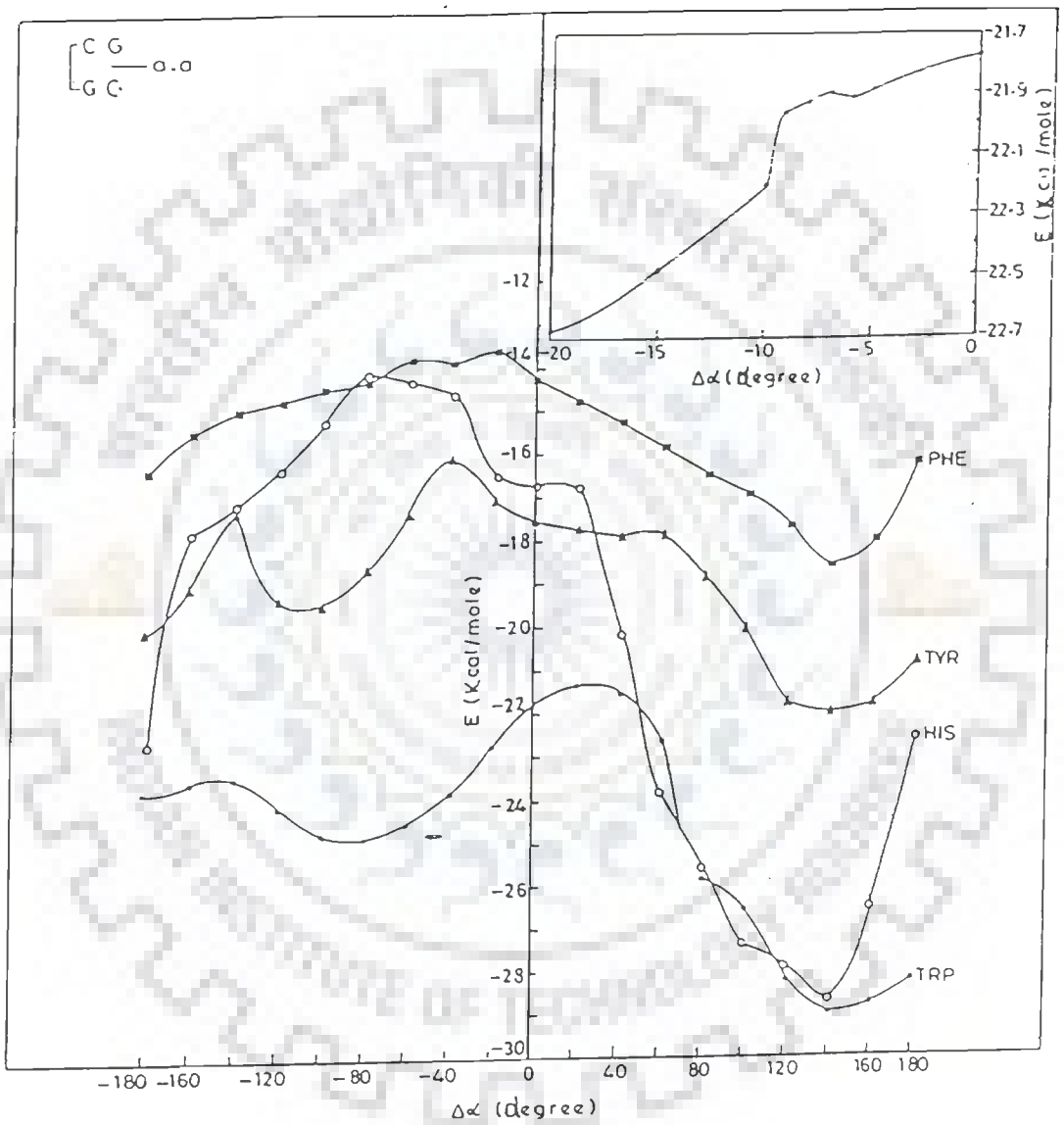
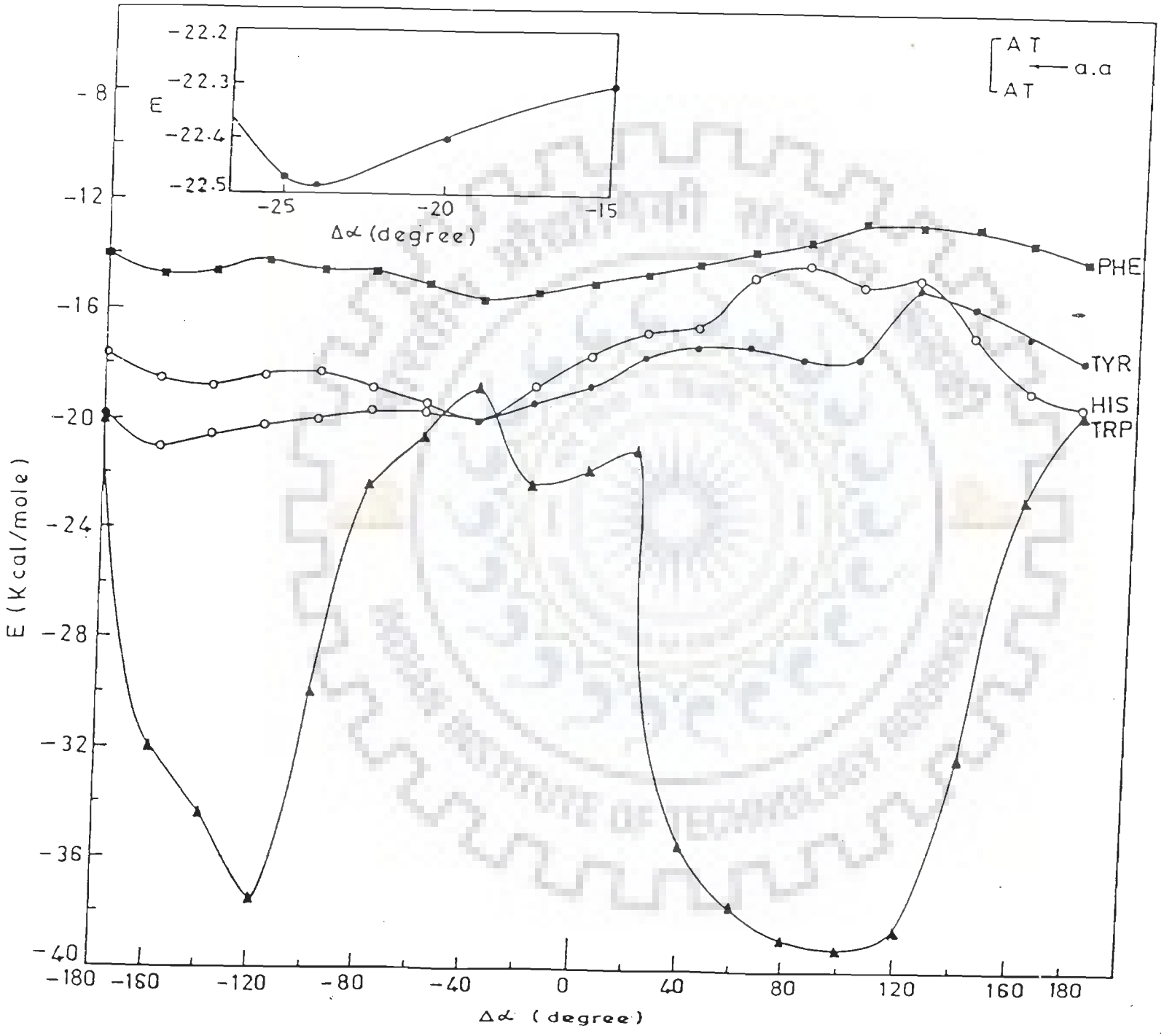
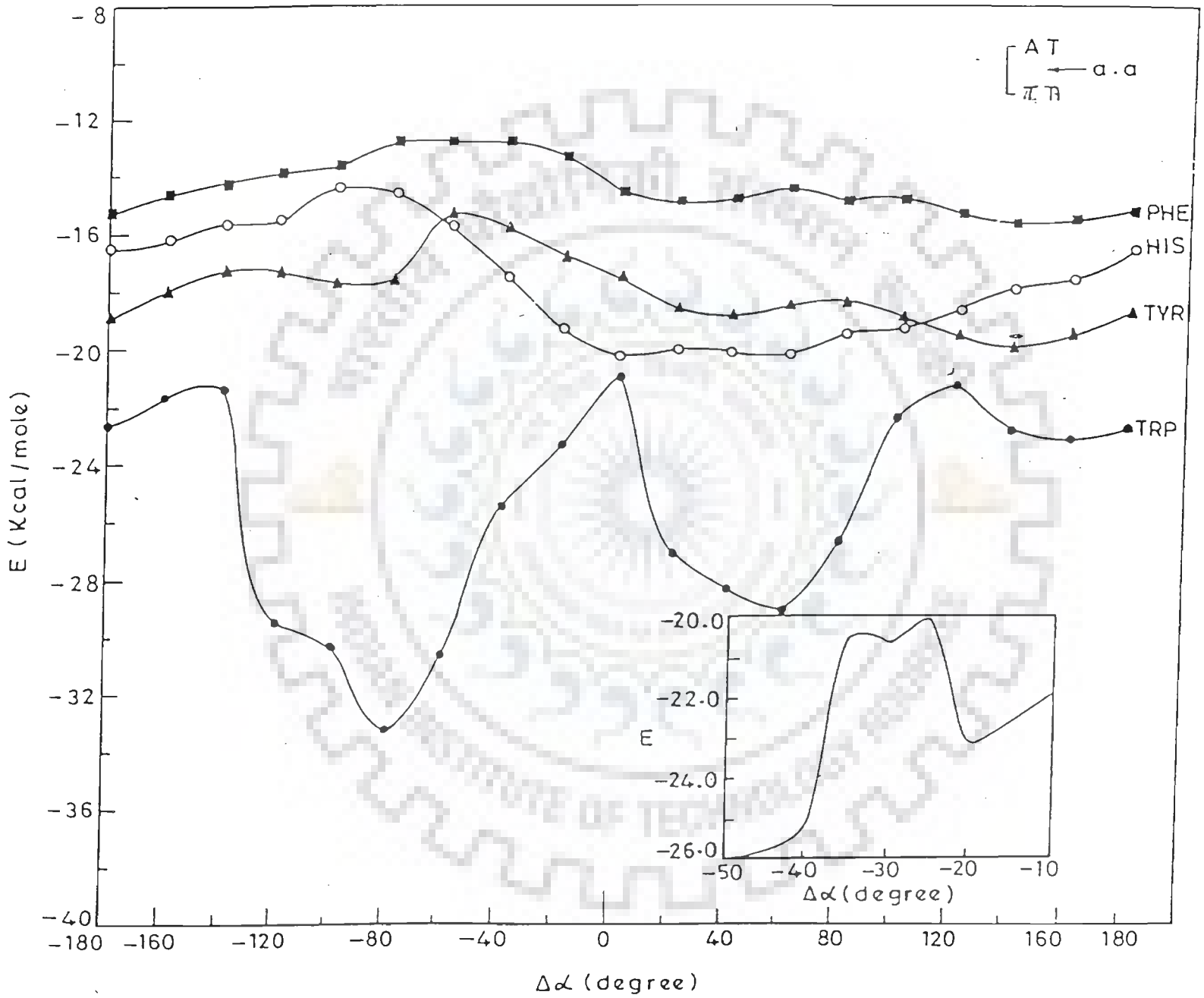
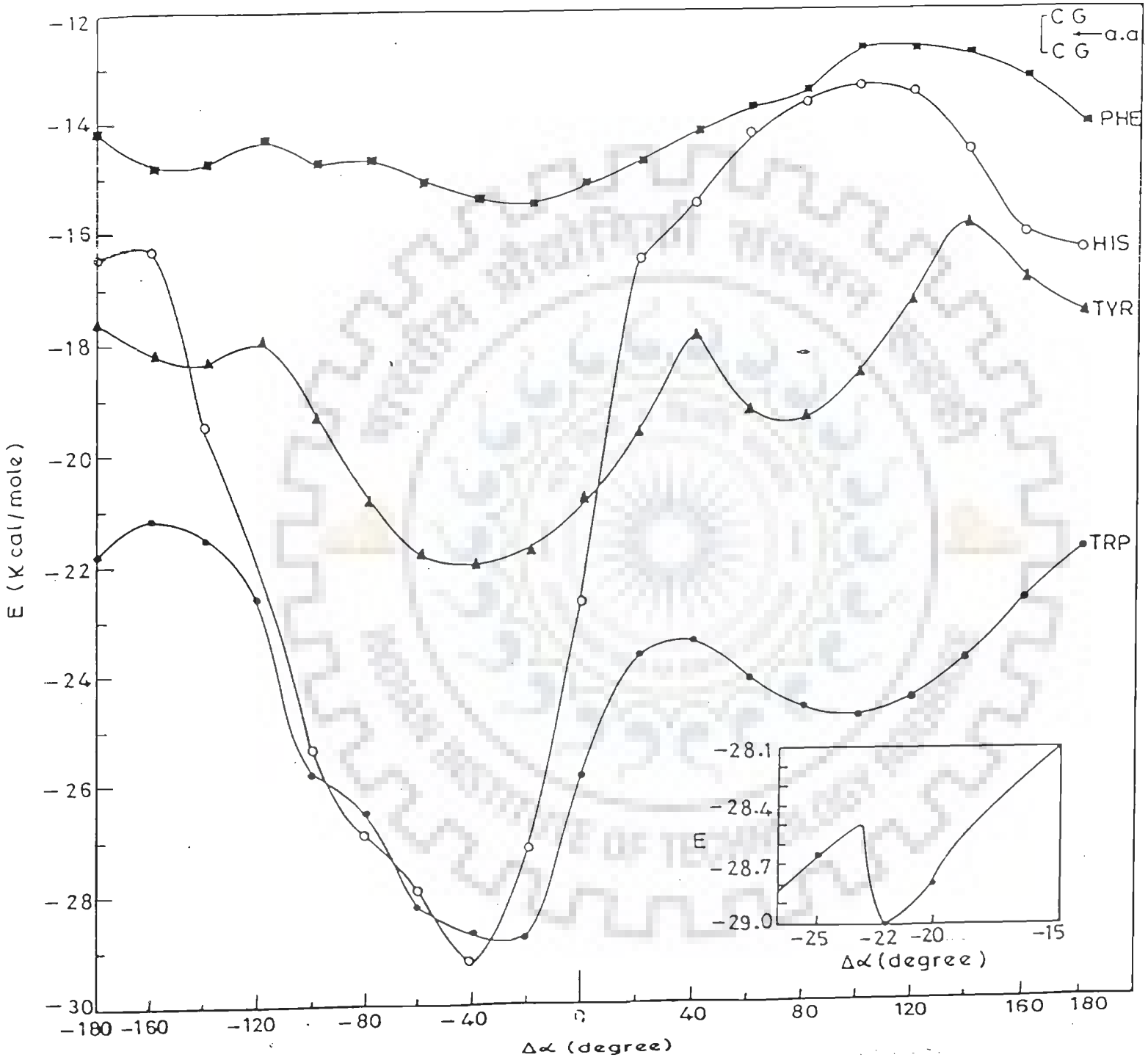


FIG. 6.3 Global search for best unwinding angle for aromatic amino acid intercalated inbetween base-pairs shown.


AT
a.a
AT
E
Δα (degree)
PHE
TYR
HIS
TRP
E (Kcal/mole)
Δα (degree)


A T
a.a
π T
PHE
HIS
TYR
TRP
E (Kcal/mole)
Δα (degree)
-8
-12
-16
-20
-24
-28
-32
-36
-40
-180
-160
-120
-80
-40
0
40
80
120
160
180
-20.0
-22.0
-24.0
-26.0
E
-50
-40
-30
-20
-10
Δα (degree)

Contd....


C G
C G
a.a
PHE
HIS
TYR
TRP
E (K cal/mole)
Δα (degree)
-12
-14
-16
-18
-20
-22
-24
-26
-28
-30
-180 -160 -120 -80 -40 0 40 80 120 160 180
-28.1
-28.4
-28.7
-29.0
E
-25 -22 -20 -15
Δα (degree)

interaction of antibiotic drugs and aromatic dyes with oligonucleotides by X-ray and theoretical studies exist in the literature (30,90,100,105). Unwinding of DNA helix in the range 20° to 30° have been observed on intercalation of drugs by X-ray techniques (130) and conformation of intercalation sites thus generated has been obtained.

The total potential energy corresponding to the best unwinding angle for different sets of base-pairs are given in Table 6.3. The results show that the total interaction energy on intercalation decreases in the order Trp > Tyr > Phe for all base-pair systems. For His, the stacking energy is practically same as that for Trp.

This is consistent with the existing experimental results. It was found that the effect of different aromatic amines on the CD spectrum of Poly (A) follows the order Trp > Tyr > His (44). Recent studies on the interaction of oligopeptides Lys-Tyr-Lys, Lys-Trp-Gly-Lys OtBu and Lys-Phe-Lys with deoxynucleotides d-CpG and d-GpC by NMR show that binding decreases in the order Trp > Tyr > Phe (13,14,15).

It is also clear from our results that intercalation between two identical base-pairs is energetically more favourable than that for corresponding alternative base sequence, that is:

$$\{^{CG}_{CG} > \{^{CG}_{GC} \quad \text{and} \quad \{^{AT}_{AT} > \{^{AT}_{TA}$$

Further for amino acids, Tyr and Phe, the stability of intercalated complex in different base-pair systems decreases in the following order:

$$\{^{CG}_{CG} > \{^{AT}_{AT} > \{^{AT}_{TA} > \{^{CG}_{GC}$$

For amino acids Trp and His, the results are only marginally different.

The stacking energy is sum of electrostatic, dispersion and repulsion energies. Contributions of these terms to the total energy have been given for optimised conformations in Table 6.3.

TABLE 6.3 : Partitioning of stacking energies (Kcal/mole) into various components.

		E_{el}	E_{pol}	E_{disp}	E_{rep}	E_{total}
CG/CG	Trp	-6.3	-4.9	-30.1	12.3	-28.8
	Tyr	-6.1	-1.5	-25.8	11.4	-22.0
	Phe	-1.7	-6.0	-24.6	13.2	-19.0
	His	-13.3	-4.7	-23.8	12.5	-29.2
AT/TA	Trp	-1.1	-3.5	-32.3	16.2	-20.7
	Tyr	-4.2	-1.9	-23.9	11.1	-18.9
	Phe	-1.0	-2.3	-21.3	9.8	-14.8
	His	-7.7	-1.6	-20.6	9.3	-20.6
AT/AT	Trp	-3.4	-3.9	-34.9	19.7	-22.5
	Tyr	-5.5	-1.7	-24.4	11.8	-19.9
	Phe	-1.7	-1.1	-21.0	8.3	-15.5
	His	-8.4	-1.7	-18.2	8.1	-20.2
CG/GC	Trp	-3.3	-2.5	-28.1	12.1	-21.9
	Tyr	-3.7	-0.9	-25.5	11.8	-18.3
	Phe	-0.4	-0.8	-22.3	9.6	-13.9
	His	-2.8	-1.7	-23.5	11.2	-16.7

The predominant contribution to the total interaction energy comes from dispersion term. This term is greater than the total energy in all complexes. The electrostatic and polarisation terms are smaller and lesser than the repulsion term.

Fig.6.4 shows the overlap geometries of amino acids with different base-pair systems corresponding to optimised energy values. These structures indicate the diverse nature of the stacking geometries. It can be seen that Trp and His intercalate with some base-pairs only partially, in agreement with results obtained by Gabbay *et al.*, 1976 (56). The stacking patterns support the selective bookmark hypothesis which states that the nucleotide sequences can be recognized by amino acid side chains acting like bookmarks.

Unfortunately crystallographic data on such protein-nucleic acid complexes are not available. It may be seen that the trend in extent of overlap for different amino acids does not necessarily follow the trend in values of minimum energies. Such a variation may result from marginal differences in corresponding stacking energies. In fact there may be many other configurations which are associated with binding energies within 0.5 Kcal/mole of the optimised structure. Such conformational freedom of intercalated complexes has important implications, since a variety of active site geometries, either inherent or induced, can facilitate the binding of protein with nucleic acids.

An important question that arises at this stage is the preference of aromatic amino acids for intercalation in nucleic acid duplex as against stacking at the ends of DNA or RNA duplexes. For CG base-pair, the binding energy for Trp due to stacking interaction is calculated to be -16.1 Kcal/mole. For intercalation the following terms are to be considered:

(i) Destacking of nucleic acid base from their usual distances in DNA/RNA of 3-3.4 Å leading to a destablisation of 14 Kcal/mole (109).

(ii) Changes in sugar-phosphate angles leading to a small but

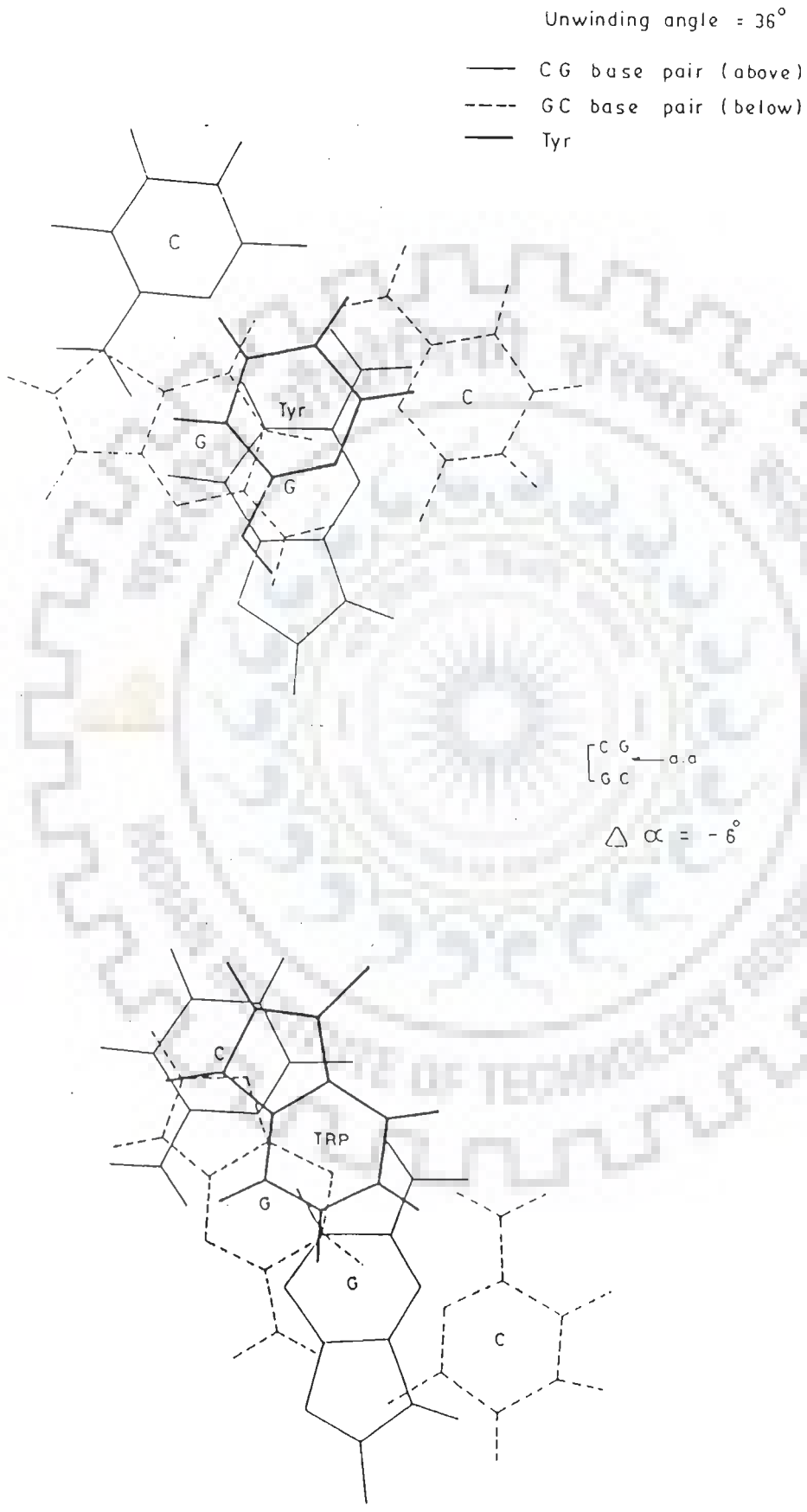
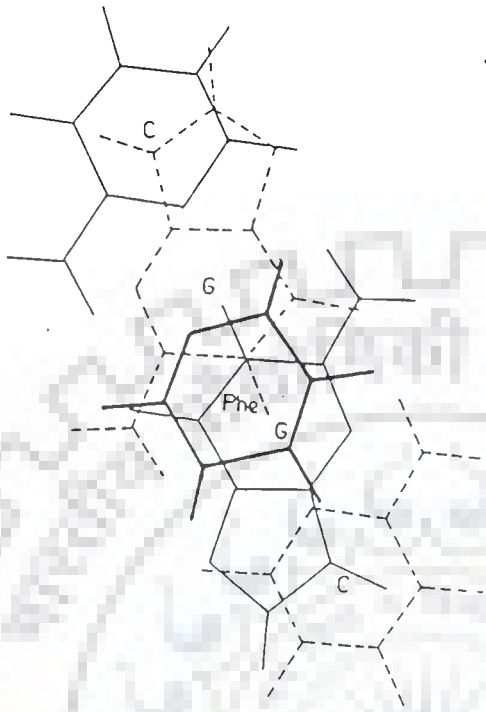


FIG. 6.4 Overlap geometries of aromatic amino acids intercalated in model dinucleotide duplex of d-CG, d-CC, d-AT and d-AA.

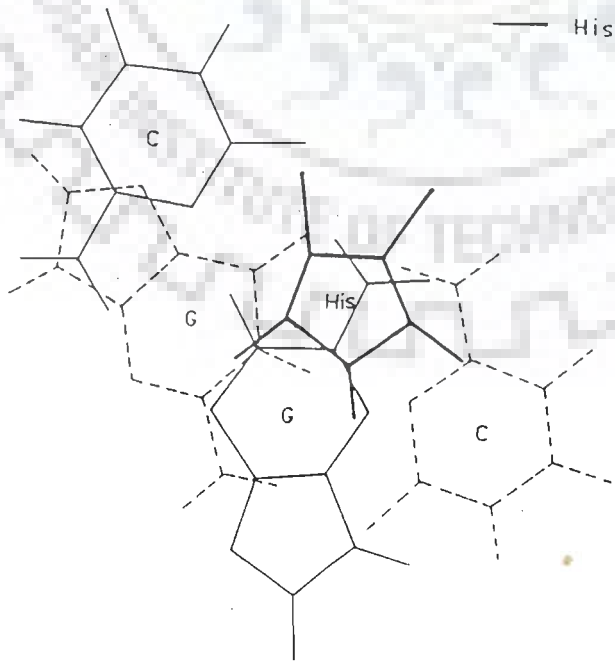
Unwinding angle = -34°

—— C G base pair (above)
---- G C base pair (below)
—— Phe



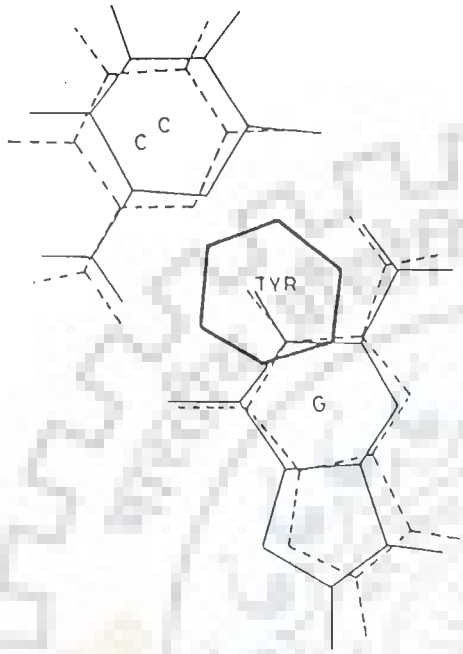
Unwinding angle = 7°

—— C G base pair (above)
---- G C base pair (below)
—— His



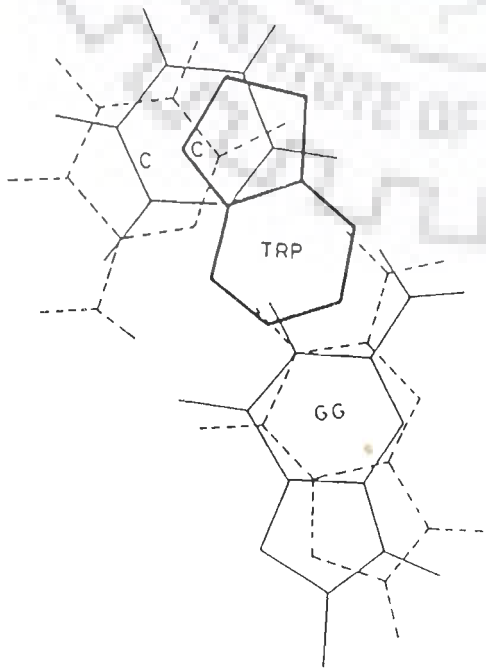
C G
C G ← a.a

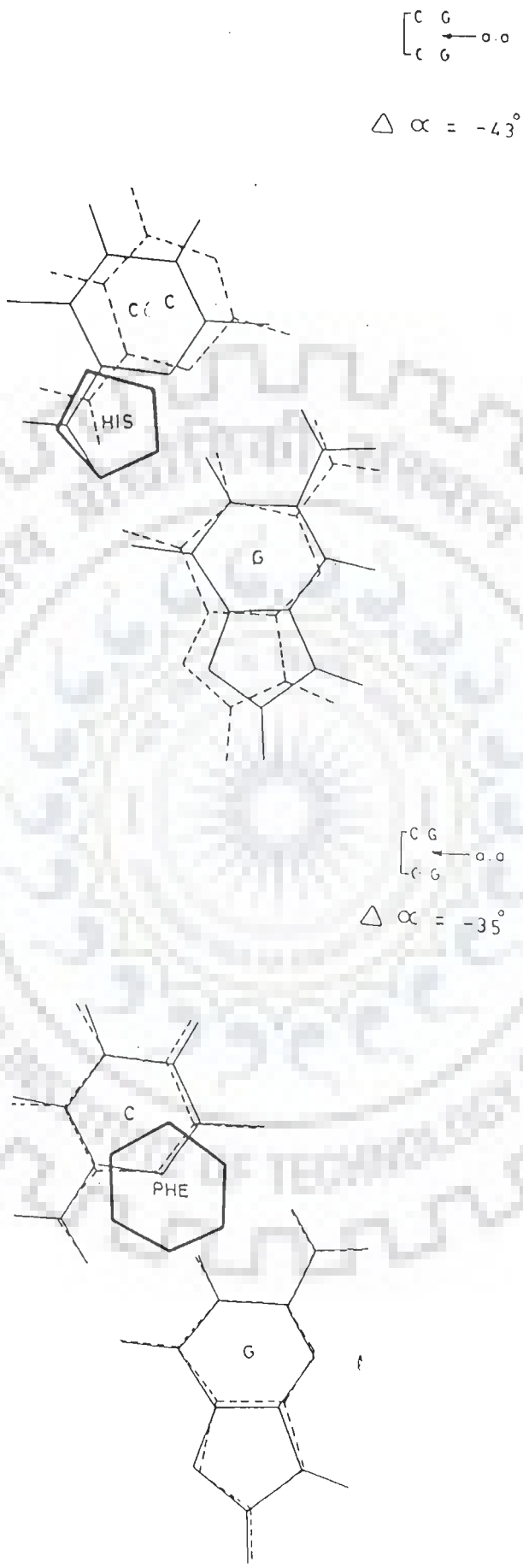
$\Delta \alpha = -41^{\circ}$

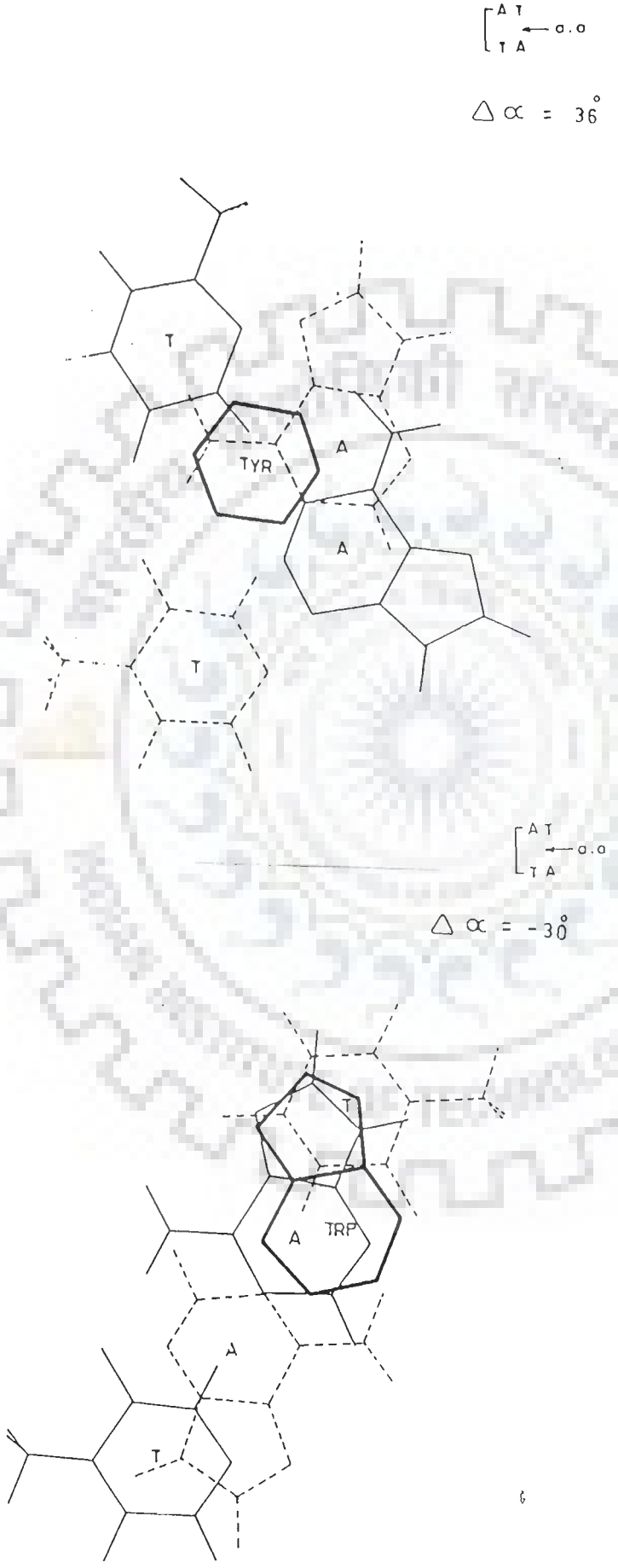


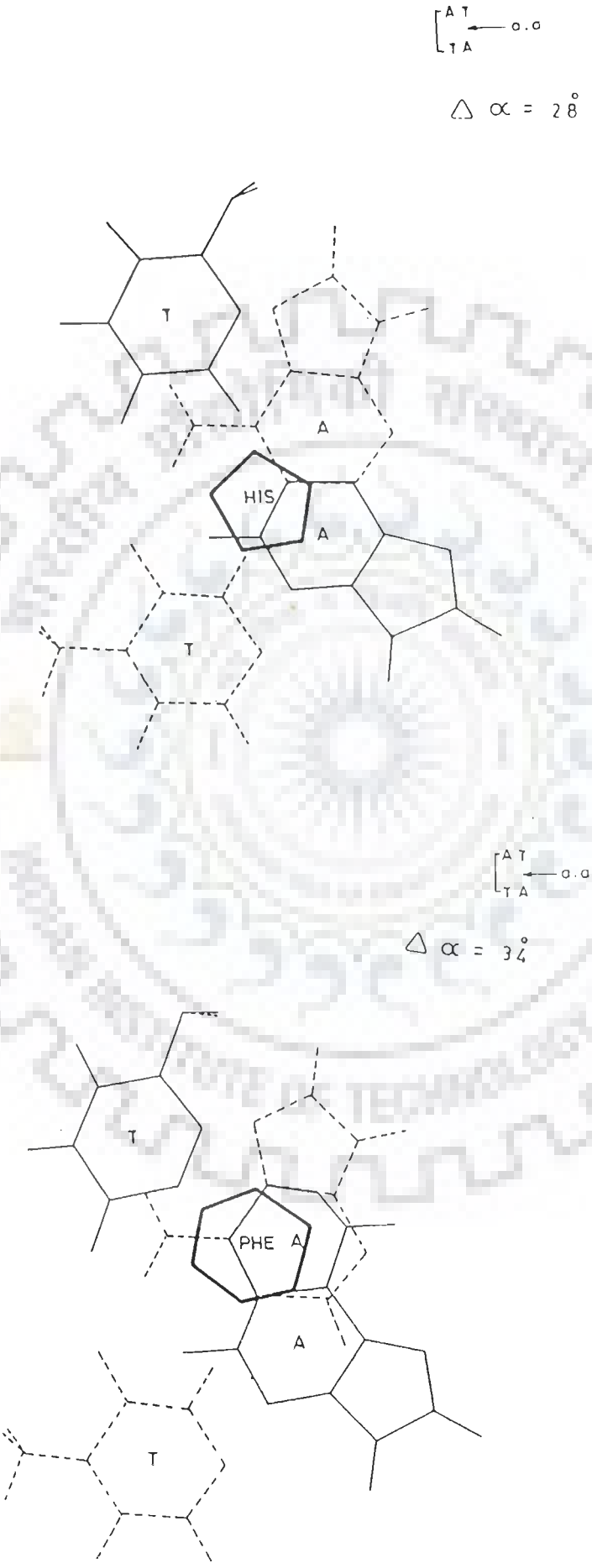
C G
C G ← a.a

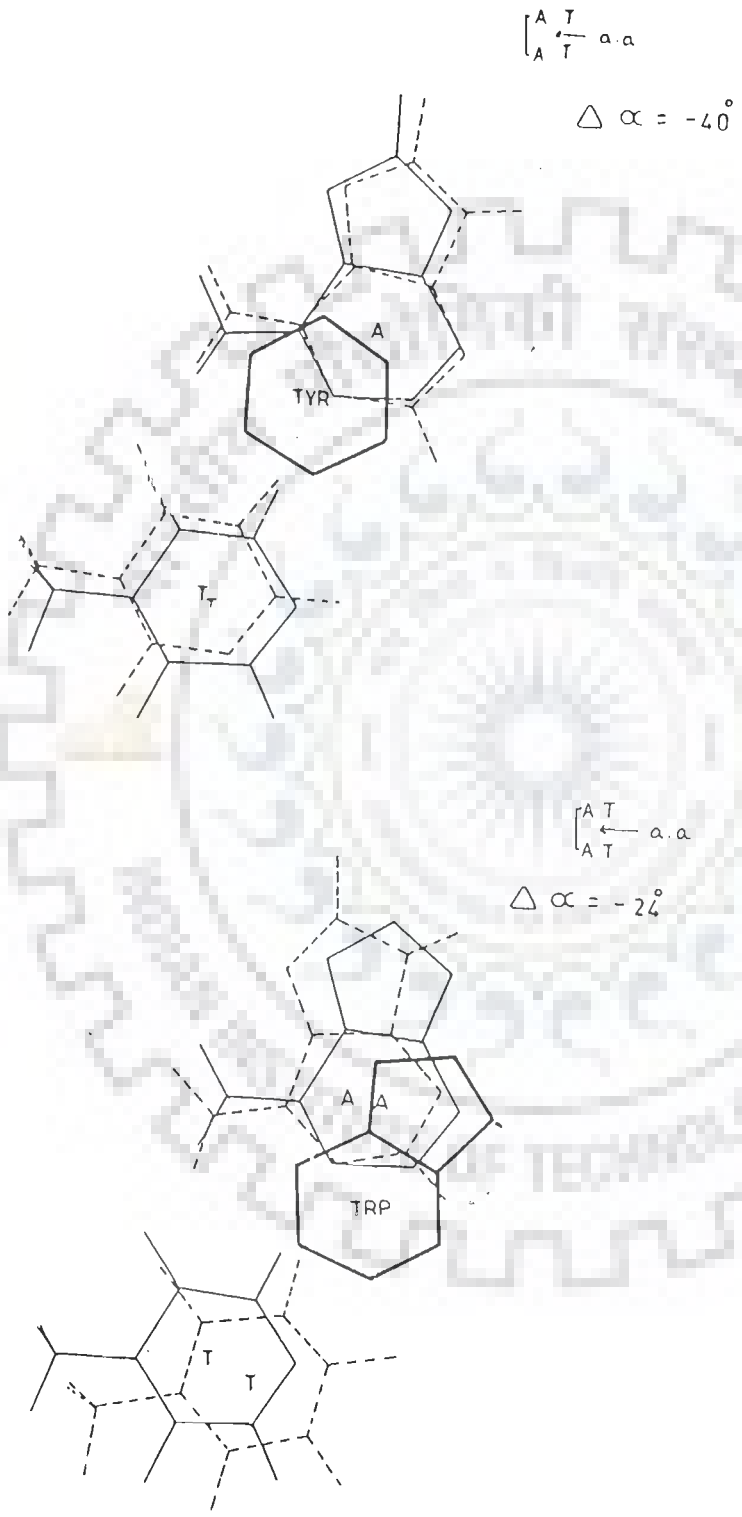
$\Delta \alpha = -22^{\circ}$

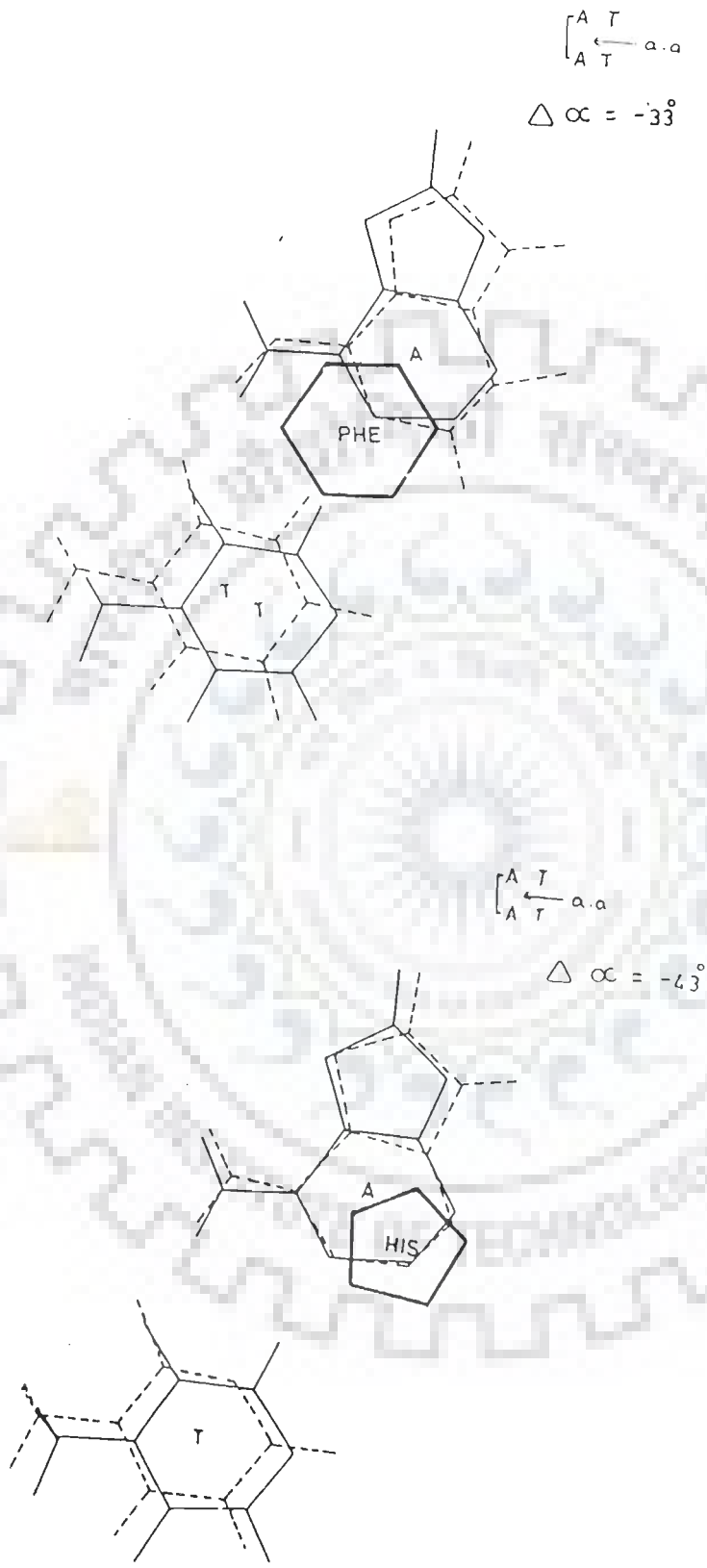



C G
C G
o.o
Δ α = -43°
C C
HIS
G
C G
C G
o.o
Δ α = -35°
C
PHE
G


A T
T A
o.o
Δ α = 36°
T
TYR
A
A
T
A T
T A
o.o
Δ α = -30°
T
A
TRP
A
T


A T
T A
o.a
Δ α = 28°
T
A
HIS
A
T
A T
T A
o.a
Δ α = 34°
T
PHE
A
A
T


A T
A T
a.a
△ α = -40°
A
TYR
T
T
A T
A T
a.a
△ α = -24°
A
A
TRP
T
T


A T
A T a.a
△ ∝ = -33°
A
PHE
T
T
A T
A T a.a
△ ∝ = -43°
A
HIS
T

finite destablisation energy.
(iii) Stablisation due to stacking interactions of Trp with, say for example, CG and CG base-pairs at its two ends (-21.9 Kcal/mole).

The sum of (i) and (iii) is about -8 Kcal/mole which is lower bound for interaction energies for intercalation. Comparing it with the stacking energies at either ends of nucleic acids (-16.1 Kcal/mole), it is obvious that Trp will show preference for end stacking as compared to intercalation in the middle of DNA/RNA duplex. It is clear from energy values in Table 5.2 that all amino acids would prefer end stacking. The same has been observed on binding of tripeptide Lys-Tyr-Lys to deoxytetranucleotides d-GpCpGpC by proton 2D-NMR (14).

The present calculations yield changes in unwinding angle on intercalation in the range -43° (unwinding) to +36° (winding). This apparently large range of variation of unwinding angle may only be taken as a rough estimate of range due to simplification in our DNA such as neglect of phosphate backbone. But the fact that these results are in similiar range as those reported for drugs, is encouraging. A significant result is the winding of DNA helix on intercalation, that is, $\Delta\alpha=36^{\circ}$ in case of Tyr. The winding of DNA helix on intercalation has also been reported in literature (111,123).

Through simple geometrical representations and model building studies (111), it has been shown that winding of DNA helix upto an angle of +16° is possible on intercalation when stretching of phosphate backbone occurs above and below the intercalation sites. A model in which the purine pyrimidine base-pairs are at their normal hydrogen bonded configuration at their normal displacement from the helix axis, perpendicular to the helix axis and the backbone continues distorsion free to the nucleotide above and below the separated pair at the normal internucleotide spacing was proposed. Paoletti and Le Pecq (111) proposed a modified version

of above intercalation model which permits stretching of backbone above and below the intercalation site. This backbone stretching permits a positive change of torsion of DNA helix that is winding of 4°. A change of 2 Å in the diameter of helix would permit to reach winding angle as large as 16°. This model permits either an unwinding of the helix included between 15° and 0° or a winding of the helix upto 16°. It was observed by viscometric measurements that unwinding of helix takes place upon intercalation of various drugs in DNA from PM2 bacteriophage. However by resonance energy transfer experiments, it was found that ethidium bromide intercalation leads to winding of DNA helix.

Table 6.4 tabulates the total energy which is required to open up the helix and then intercalate the moiety between two base-pairs, that is,

$$E_{total} = E_{CA} + E_{stack}$$

where E_{stack} is the total energy due to stacking interaction calculated by potential energy calculations in present study and E_{CA} is the energy required for opening the base-pairs of double helix. The value of E_{CA} has been taken from literature (110). It has been found that E_{CA} is always positive whereas E_{stack} is negative. From the results obtained, it can be inferred that the interaction energy for copolymer system follows the order:

$$\left\{\begin{matrix}CG\\CG\end{matrix}\right. > \left\{\begin{matrix}GC\\GC\end{matrix}\right. > \left\{\begin{matrix}TA\\TA\end{matrix}\right. > \left\{\begin{matrix}AT\\AT\end{matrix}\right.$$

for all aromatic amino acids. As is clearly seen that the combination $\begin{matrix}Py & Pu\\Py & Pu\end{matrix}$ form more stable complexes than $\begin{matrix}Pu & Py\\Pu & Py\end{matrix}$. It is noted that poly AT and poly GC forms more stable complexes. The opening of helix followed by intercalation is the most preferred mode of binding when aromatic amino acids bind to DNA.

From Table 6.4, it is seen that E_{stack} is not related to $\Delta\alpha$, unwinding angle. We expect E_{CA} only to be related to α as unwinding is opening of helix to accommodate aromatic rings. So for greater value of E_{CA}, more will be the opening and hence large unwinding of the helix are expected. E_{stack} is only destacking of

TABLE 6.4 : Total interaction energy (E_{total}) calculated using E_{CA} (ref) and E_{stack} (present calculations).

		Trp			Tyr			Phe			His		
	E_{CA}	Estack	Etotal	α	Estack	Etotal	α	Estack	Etotal	α	Estack	Etotal	α
AT TA	3.5	-20.7	-17.2	..	-22.0	-18.5	..	-19.0	-15.5	..	-29.2	-23.7	..
TA AT	16.7	-20.7	-4.0	30	-22.0	-5.3	36	-19.0	-2.3	34	-29.2	-12.5	28
AT AT	12.2	-22.5	-10.3	-24	-18.9	-6.7	-40	-14.8	-2.6	-33	-20.6	-8.4	-43
TA TA	12.0	-22.5	-10.5	..	-18.9	-6.9	..	-14.8	-2.8	..	-20.6	-8.6	..
CG GC	14.9	-21.9	-7.0	-6	-19.9	-5.0	36	-15.5	-0.6	-34	-20.1	-5.2	7
GC CG	3.9	-21.9	-18.0	..	-19.9	-16.0	..	-15.5	-11.6	..	-20.1	-16.2	..
CG CG	6.0	-28.8	-22.8	-22	-18.3	-12.3	-41	-13.9	-7.9	-35	-16.7	-10.7	-43
GC GC	6.5	-28.8	-22.3	..	-18.3	-11.8	..	-13.9	-7.4	..	-16.7	-10.2	..

base-pairs followed by stacking of aromatic ring of amino acids residues. This part of interaction involves only base-pairs and not the backbone of DNA. Variation of E_{stack} with different base-pair sequences is not directly related to variation in $\Delta\alpha$. Our aim to carry out the present study was to reflect the significance of stacking interactions, their variation with different aromatic rings of amino acids and their relation with $\Delta\alpha$ and stacking configurations. These orders of variation with different base-pairs do not necessarily reflect total energies of interaction in complete system as E_{CA} varies in an altogether different way for different sequences as it involves conformational adjustment of nucleic acid. Nevertheless the present study established the significant role and gives a further insight of stacking interactions.

In conclusion, we note that unlike drugs, intercalation of aromatic amino acids in between DNA/RNA may be partial and give rise to several configurations having energy differences of only about 0.5 Kcal/mole. Unwinding or winding of helix on intercalation occurs to different extent for different amino acids depending upon the base-pairs involved. The observed large extent of flexibility may be due to comparatively small size of aromatic rings of these amino acids with respect to base-pairs. However, different configurations in each of these intercalated complexes exhibit selectivity of interactions and has important implications in recognition of nucleic acids by proteins.

We carried out theoretical calculations to understand the stacking of aromatic acids, in particular, tyrosine. We observe that the theoretical results support our experimental findings.

CHAPTER VII

THEORETICAL ENERGY CALCULATIONS ON THE STACKING OF AROMATIC AMINO ACIDS BETWEEN TWO BASES.

A Minimum energy configuration

In order to calculate the minimum energy configuration corresponding to the most stable complex, the relative orientation of aromatic amino acid with respect to the two bases was exactly done in a manner described elsewhere (chapter VI).

B Intercalation of aromatic amino acids between two bases

Fig.7.1(a) shows the interaction energy of aromatic amino acid X (X = Trp, Tyr, Phe and His) positioned exactly midway between two bases (A, T, G, C) having a relative orientation, α, of 36° as a function of distance between two bases. Minimum interaction energy of -17.9 Kcal/mole is observed for Trp for a distance of 6.6 Å between two T bases. However for Tyr, Phe and His, the minimum potential energy values are found to be -13.6, -12.0 and -14.5 Kcal/mole corresponding to a distance of 6.4, 6.6 and 6.4 Å, respectively. This result is consistent with the findings that single-stranded DNA adopts extended forms on binding to proteins (129).

Figs.7.1(b) and 7.1(c) show the effect of translating the aromatic amino acid with respect to the two bases in X-Y plane at an optimised Z distance. It may be noted that the interaction energy decreases as the aromatic ring of amino acid is displaced from centre in either X or Y direction. This is understandable as the energy is expected to be minimum when there is more overlap. However this is not found to be always true. Fig.7.1(d) shows the effect of rotation of amino acid ($\Delta\theta$) after optimising translation along X and Y direction. Minima are observed at an angle of 40° and 220° in case of Trp when placed between two G bases. In most of the cases, two values of $\Delta\theta$ are 180° apart corresponding to maximum overlap geometry.

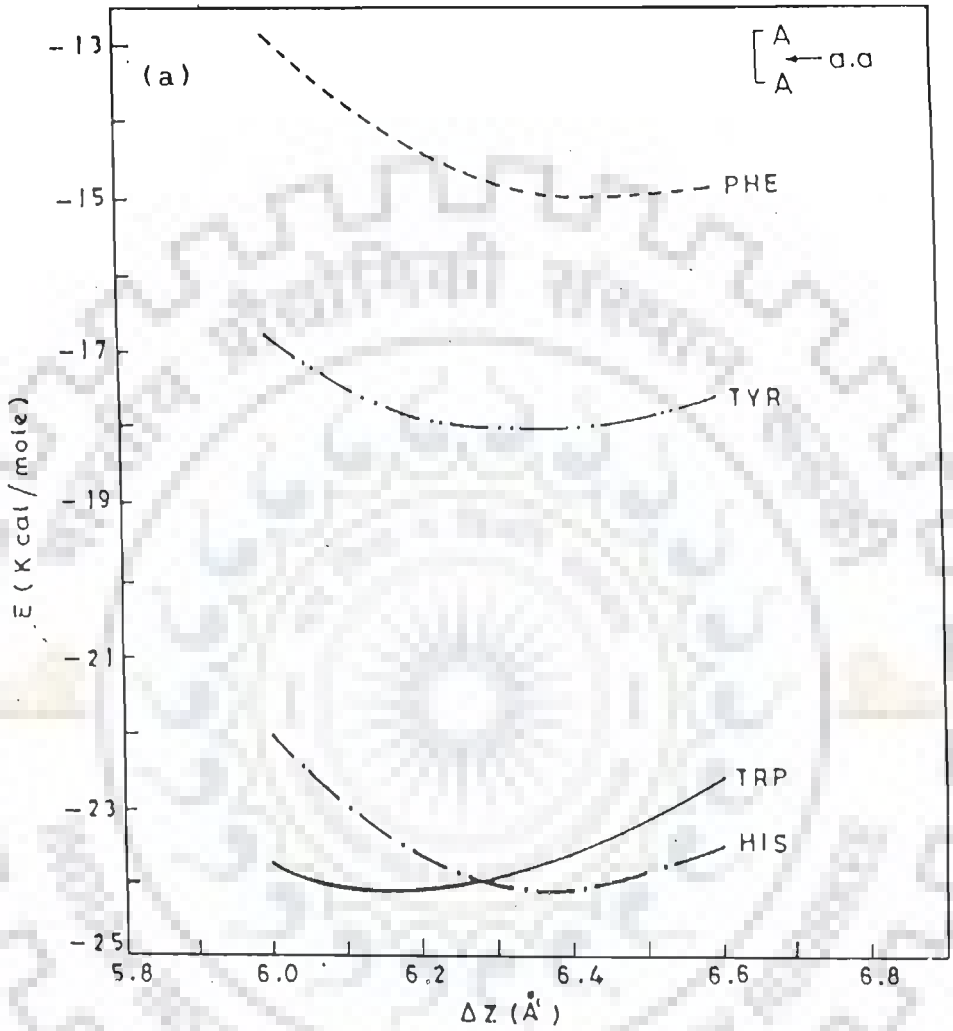
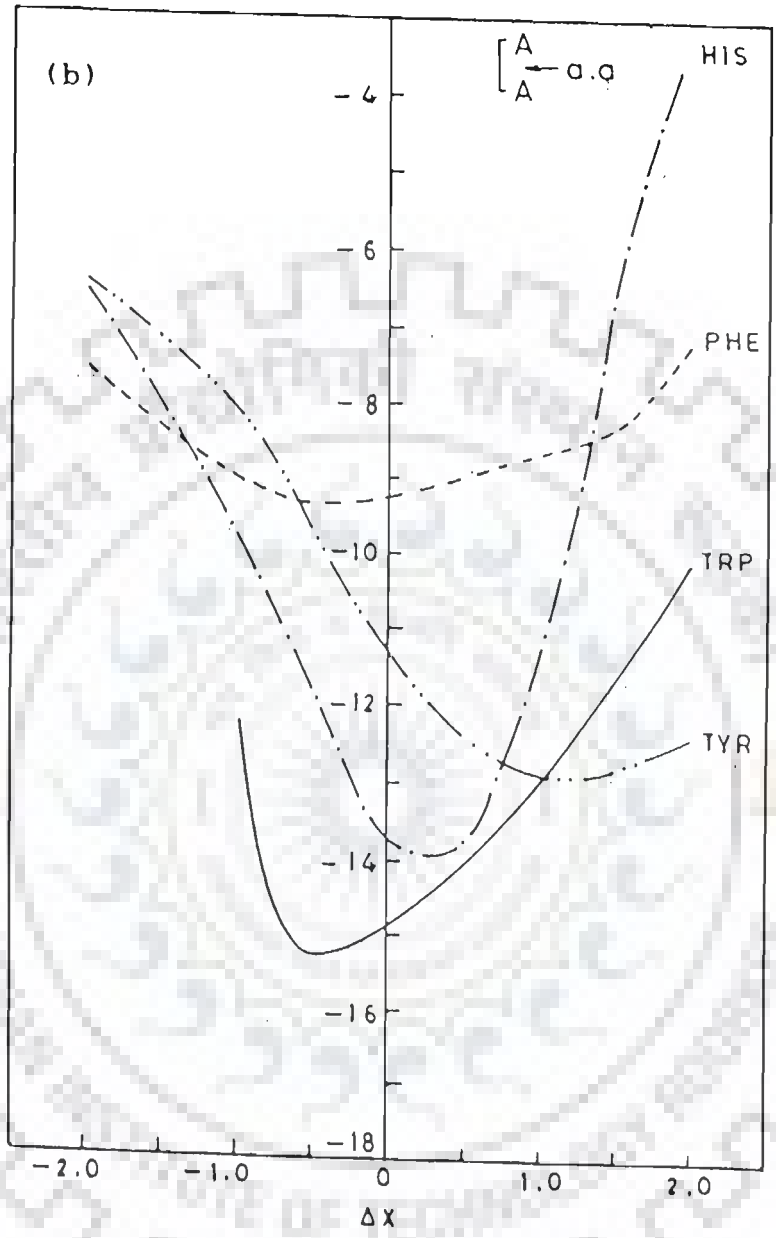
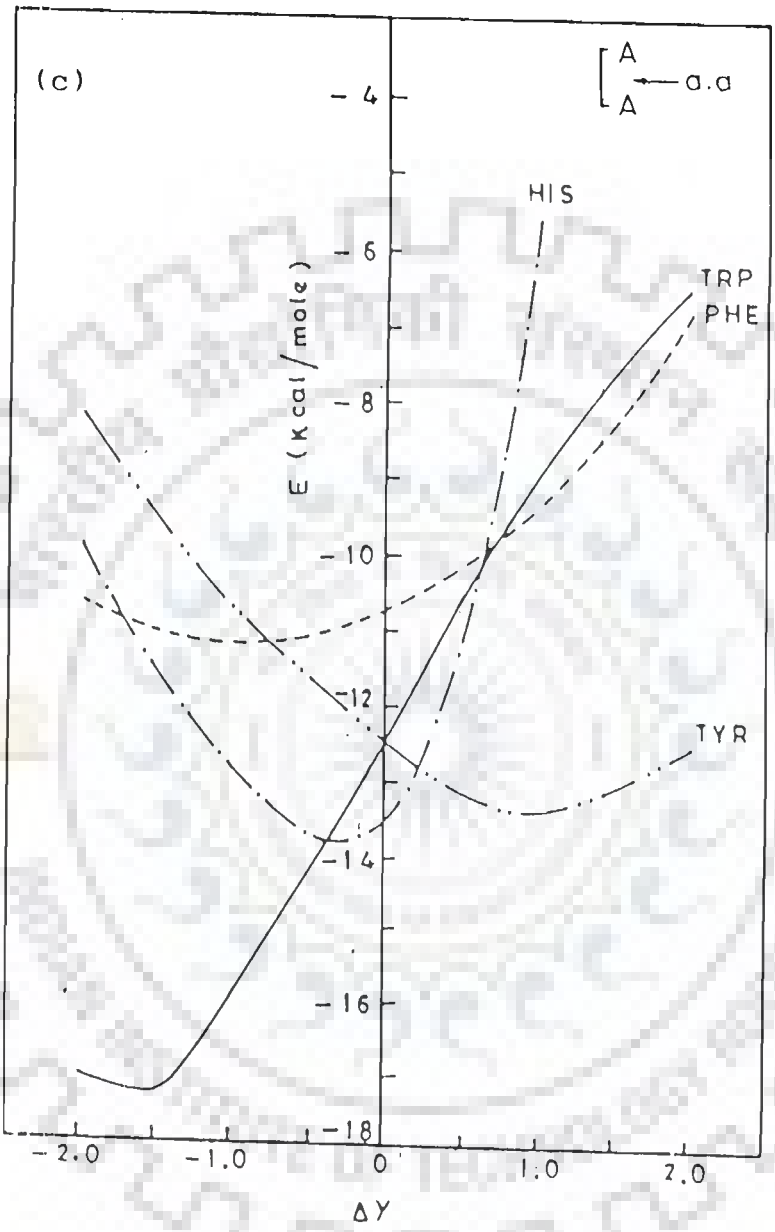
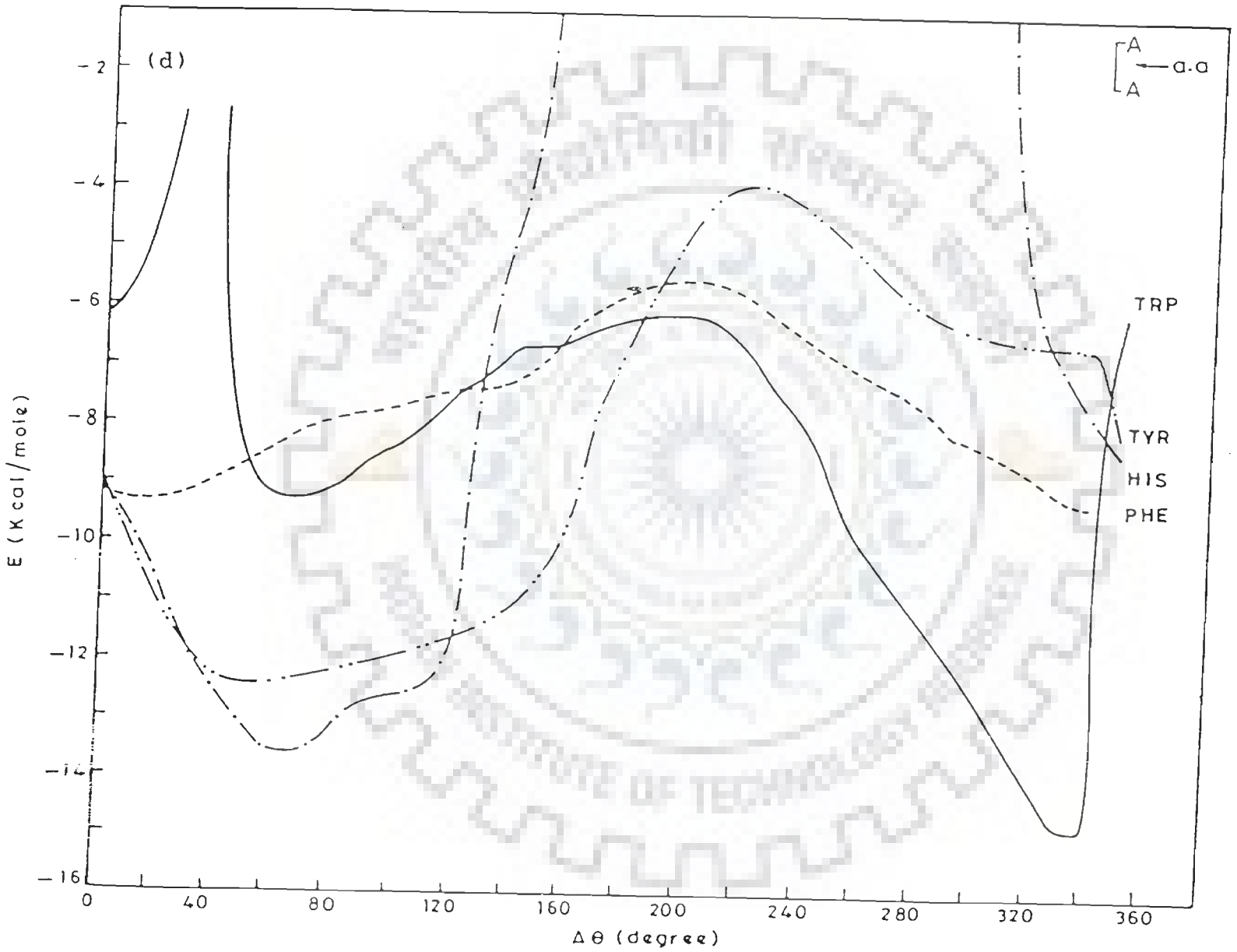


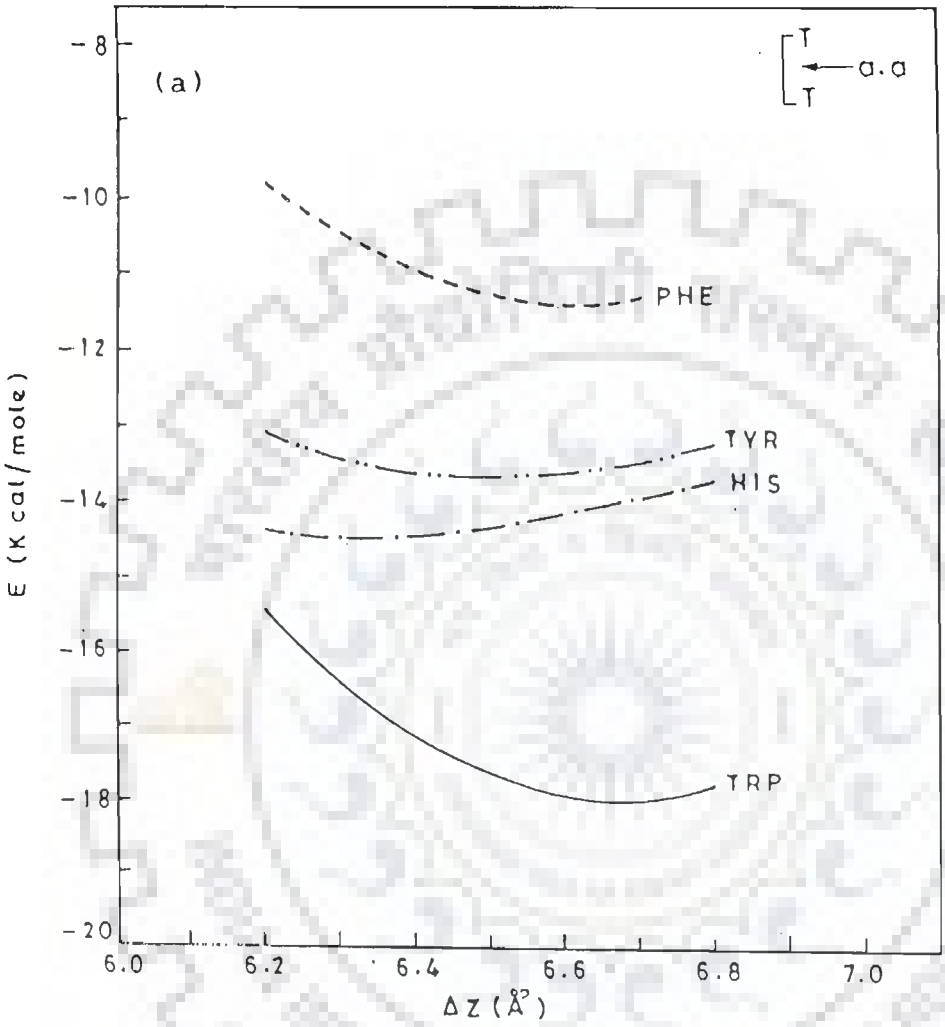
FIG. 7.1 Interaction energy as a function of :
a) ΔZ b) X
c) Y d) Δθ
when aromatic amino acid is placed midway between two bases. A, T, G and C.


(b)
[A A ← a.a
HIS
PHE
TRP
TYR
-4
-6
-8
-10
-12
-14
-16
-18
-2.0
-1.0
0
1.0
2.0
ΔX

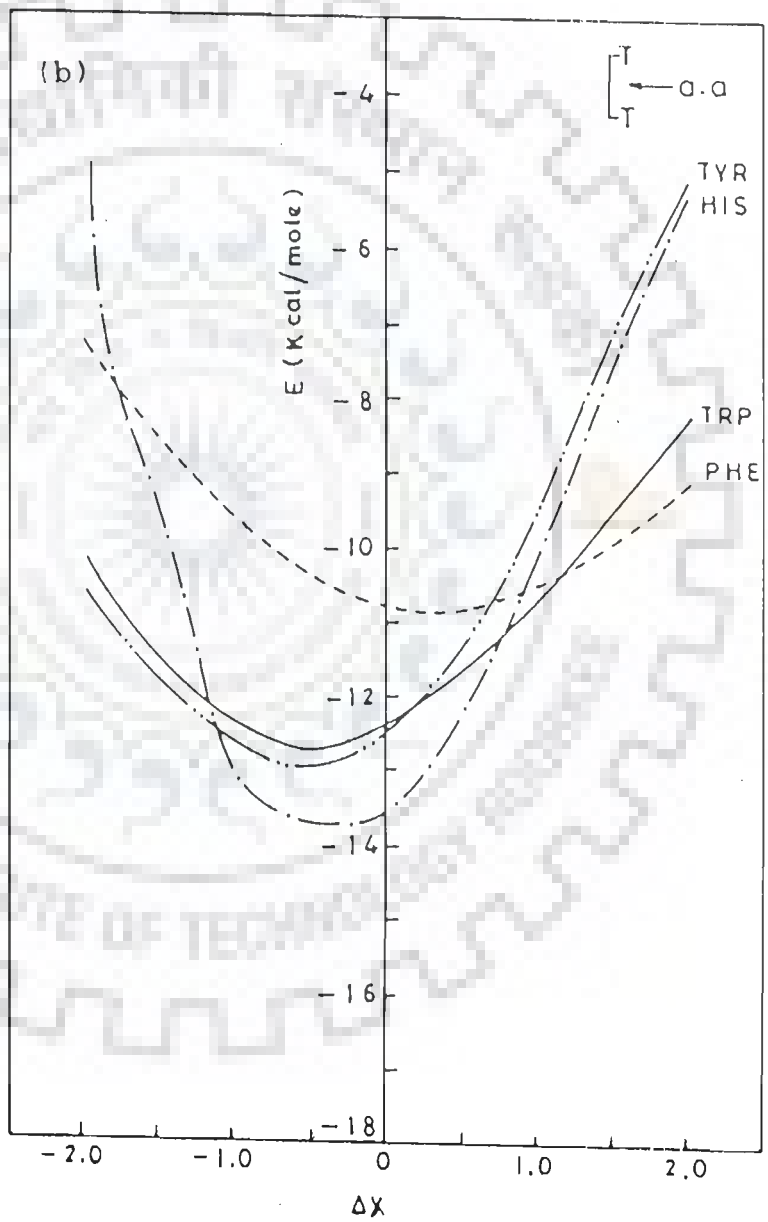

(c)
A
A
a.a
HIS
TRP
PHE
TYR
E (K cal / mole)
-4
-6
-8
-10
-12
-14
-16
-18
-2.0
-1.0
0
1.0
2.0
ΔY

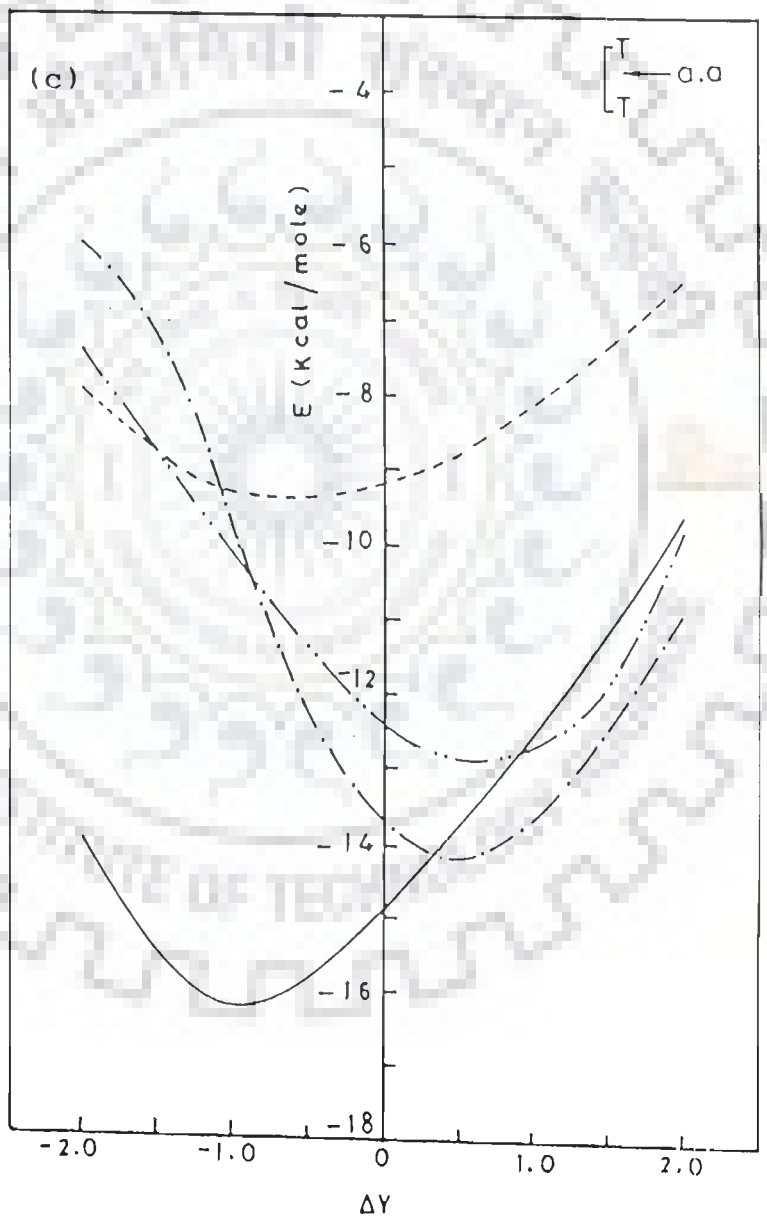
Contd...


(d)
A
A
a.a
TRP
TYR
HIS
PHE
E (K cal/mole)
-2
-4
-6
-8
-10
-12
-14
-16
0
40
80
120
160
200
240
280
320
360
Δθ (degree)

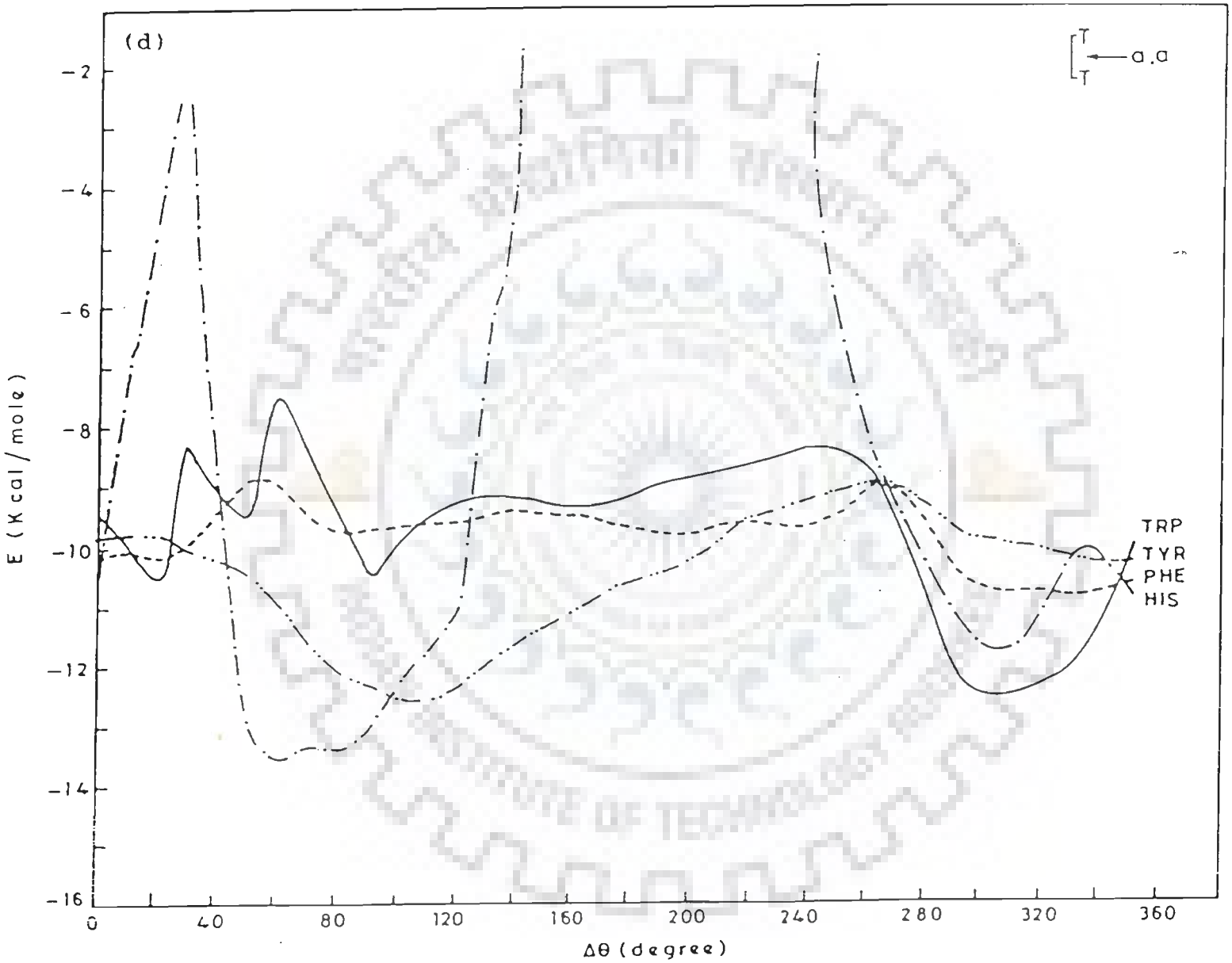

(a)
T
a.a
T
PHE
TYR
HIS
TRP
E (K cal/mole)
-8
-10
-12
-14
-16
-18
-20
6.0
6.2
6.4
6.6
6.8
7.0
ΔZ (Å)

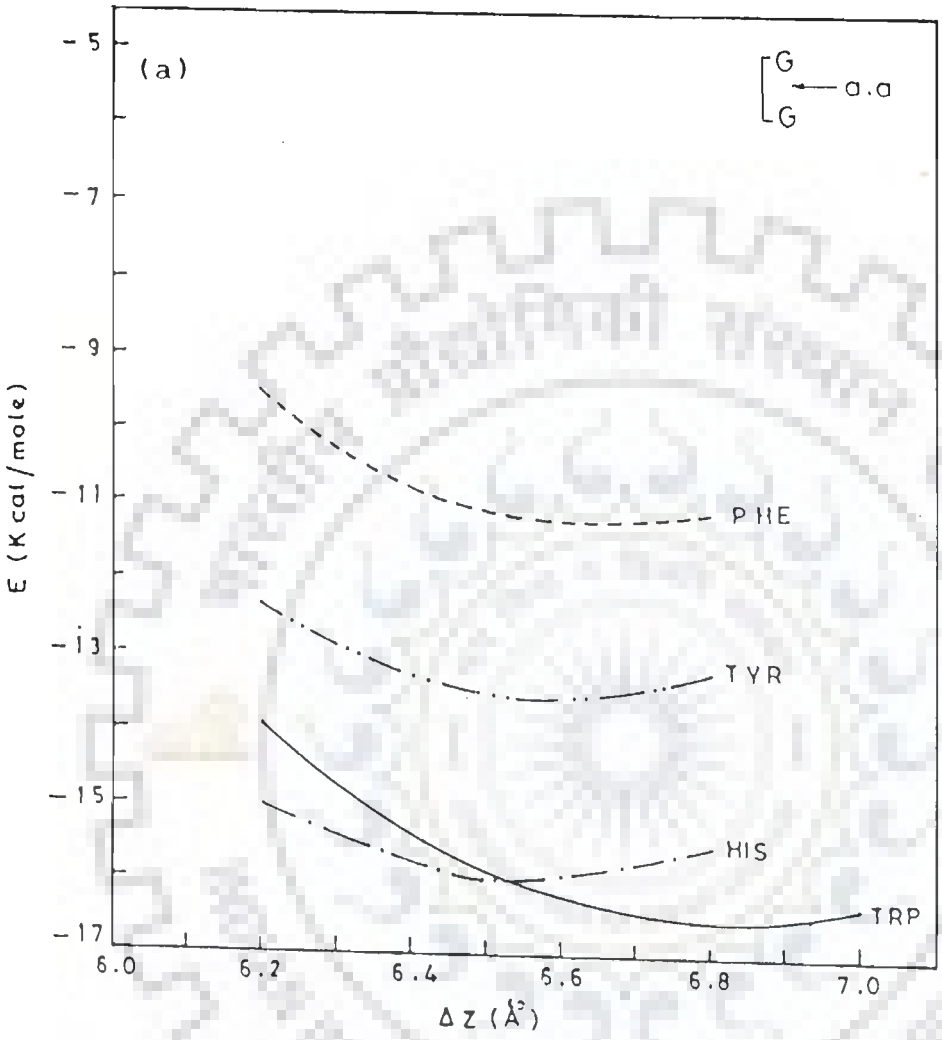
Contd....


(b)
T
a.a
T
TYR
HIS
TRP
PHE
E (K cal/mole)
-4
-6
-8
-10
-12
-14
-16
-18
-2.0
-1.0
0
1.0
2.0
ΔX

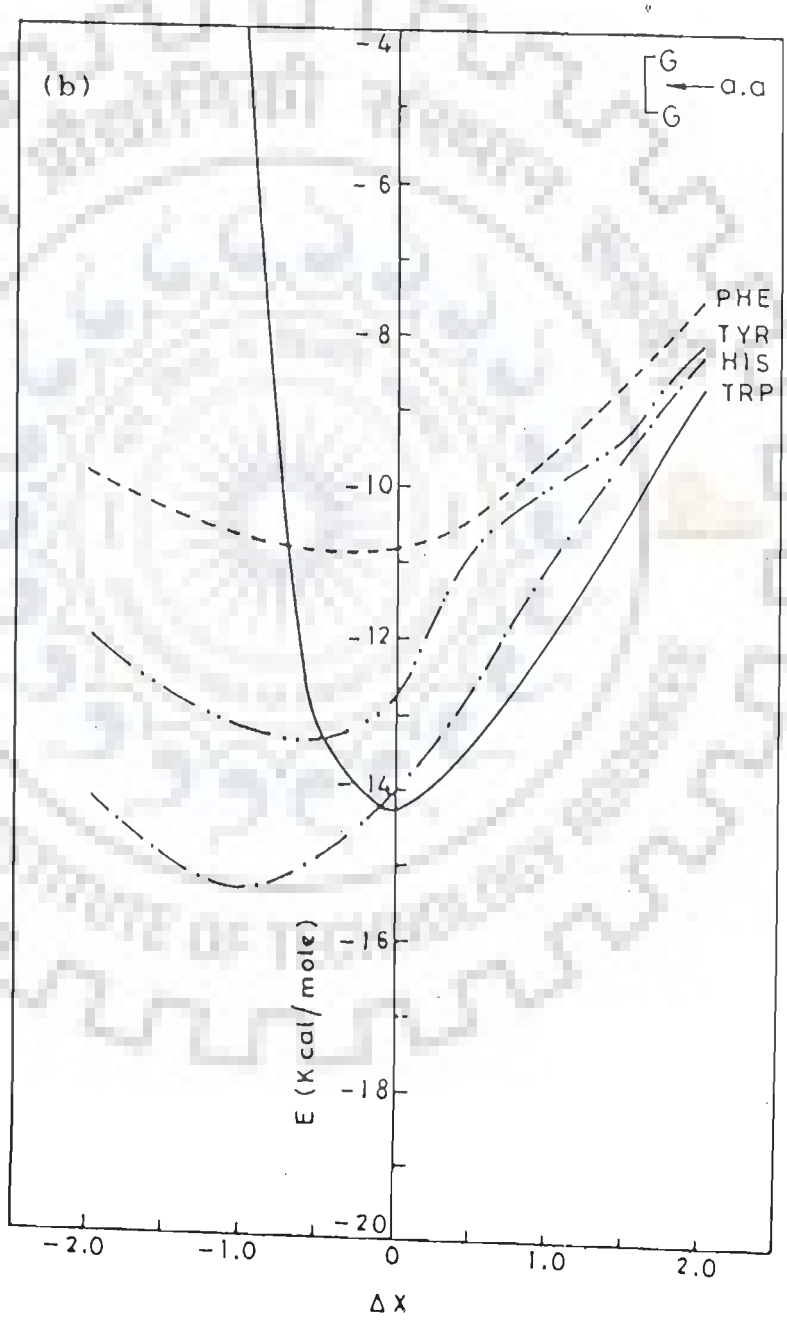


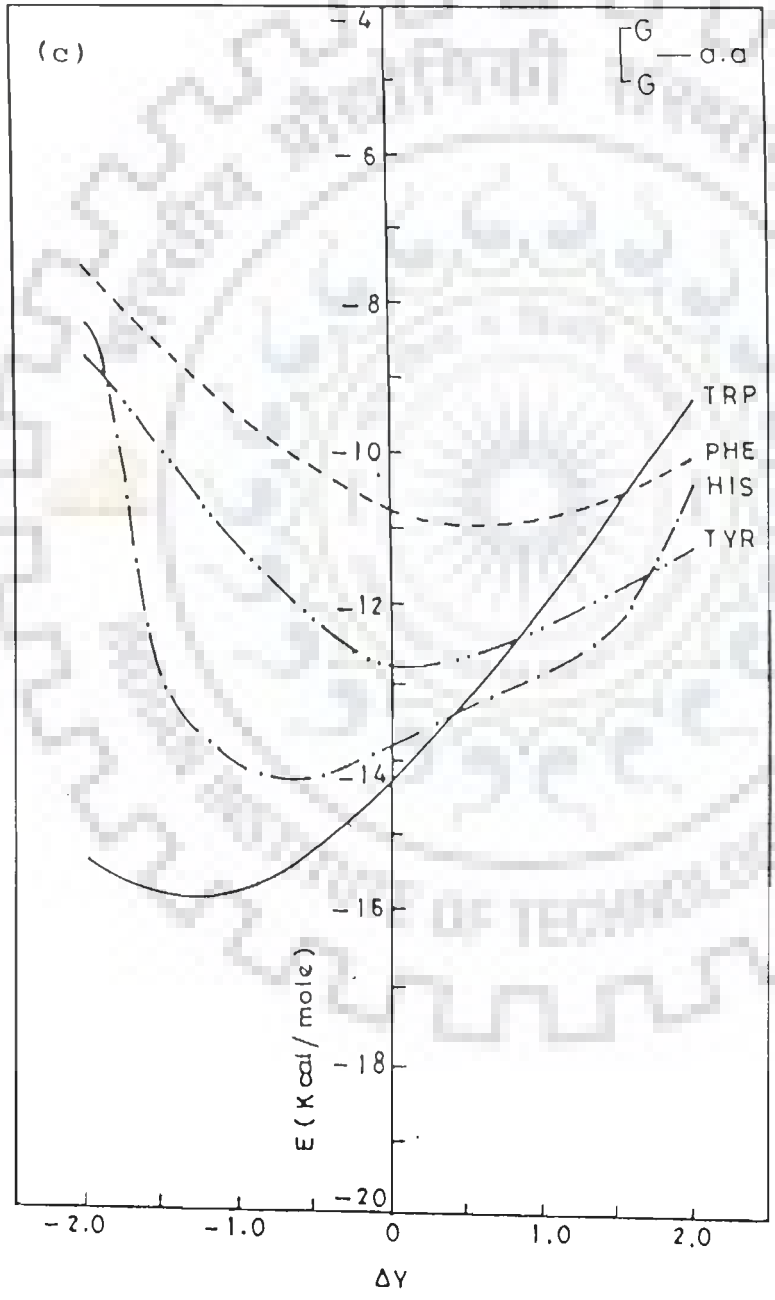
Contd....


(d)
T
T
a.a
-2
-4
-6
-8
-10
-12
-14
-16
E (K cal / mole)
0
40
80
120
160
200
240
280
320
360
Δθ (degree)
TRP
TYR
PHE
HIS

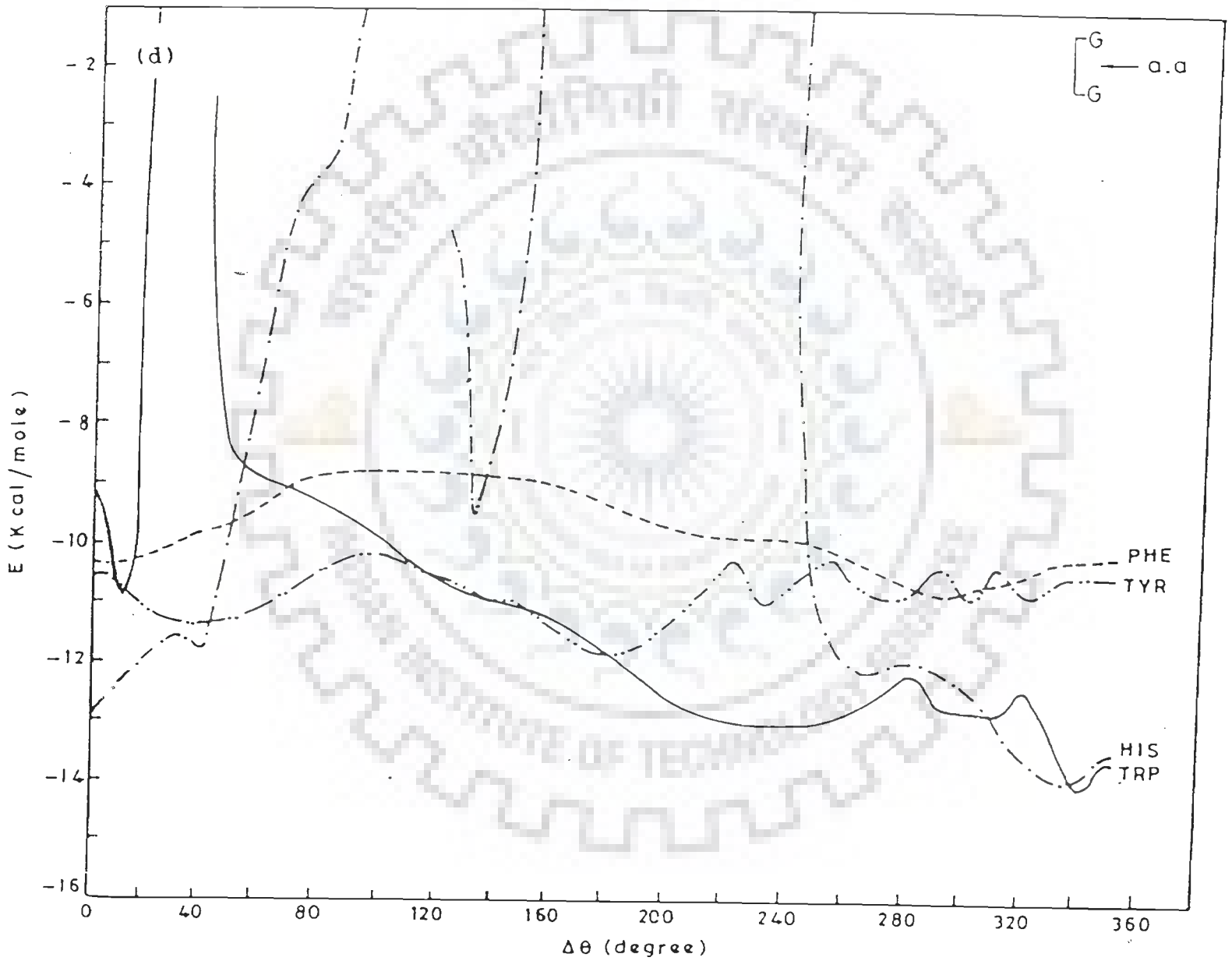

(a)
G
G
a.a
PHE
TYR
HIS
TRP
E (K cal/mole)
-5
-7
-9
-11
-13
-15
-17
6.0
6.2
6.4
6.6
6.8
7.0
ΔZ (Å)

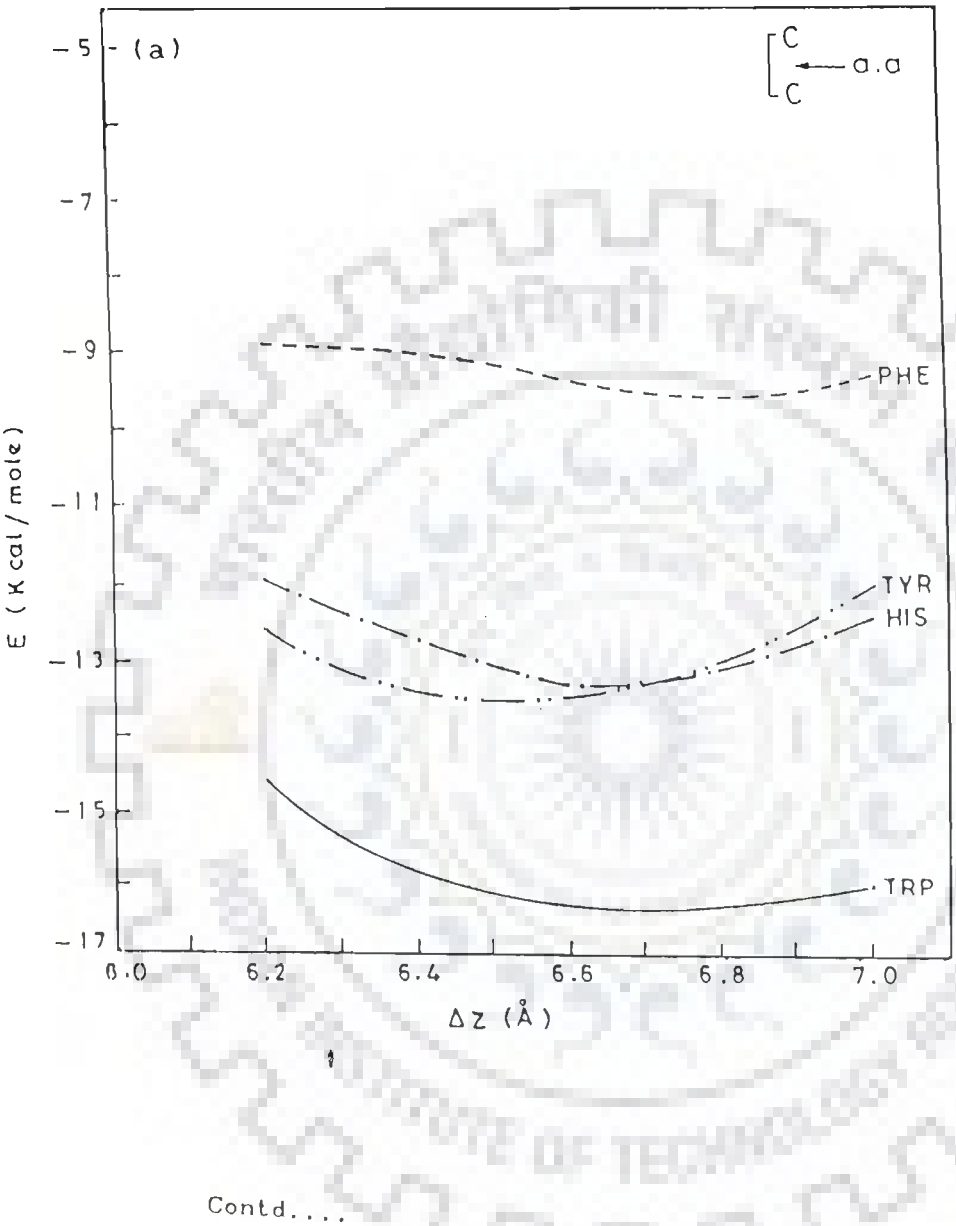
Contd..


(b)
G
G
a.a
PHE
TYR
HIS
TRP
-4
-6
-8
-10
-12
-14
-16
-18
-20
E (K cal/mole)
-2.0
-1.0
0
1.0
2.0
Δ X

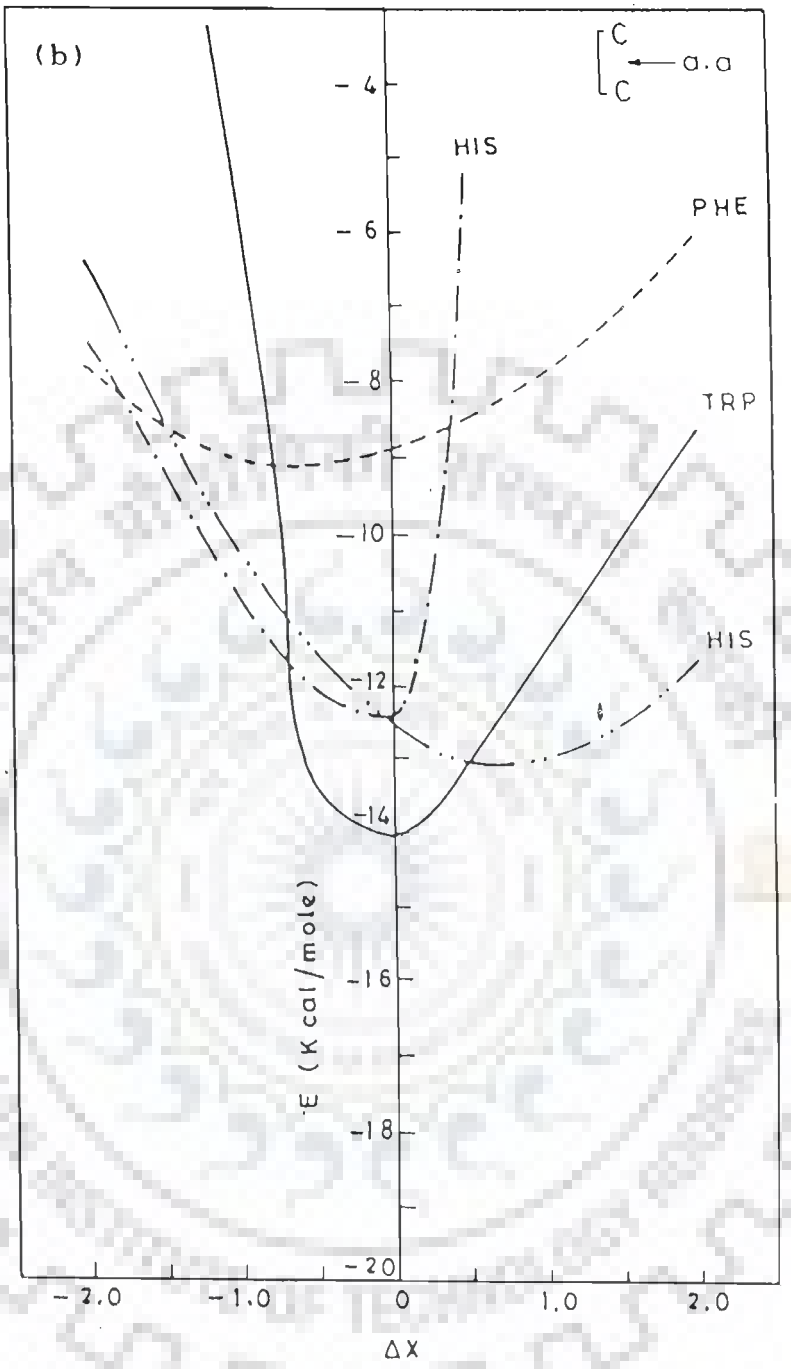

(c)
G
G
a.a
TRP
PHE
HIS
TYR
E (Kcal/mole)
ΔY
-4
-6
-8
-10
-12
-14
-16
-18
-20
-2.0
-1.0
0
1.0
2.0

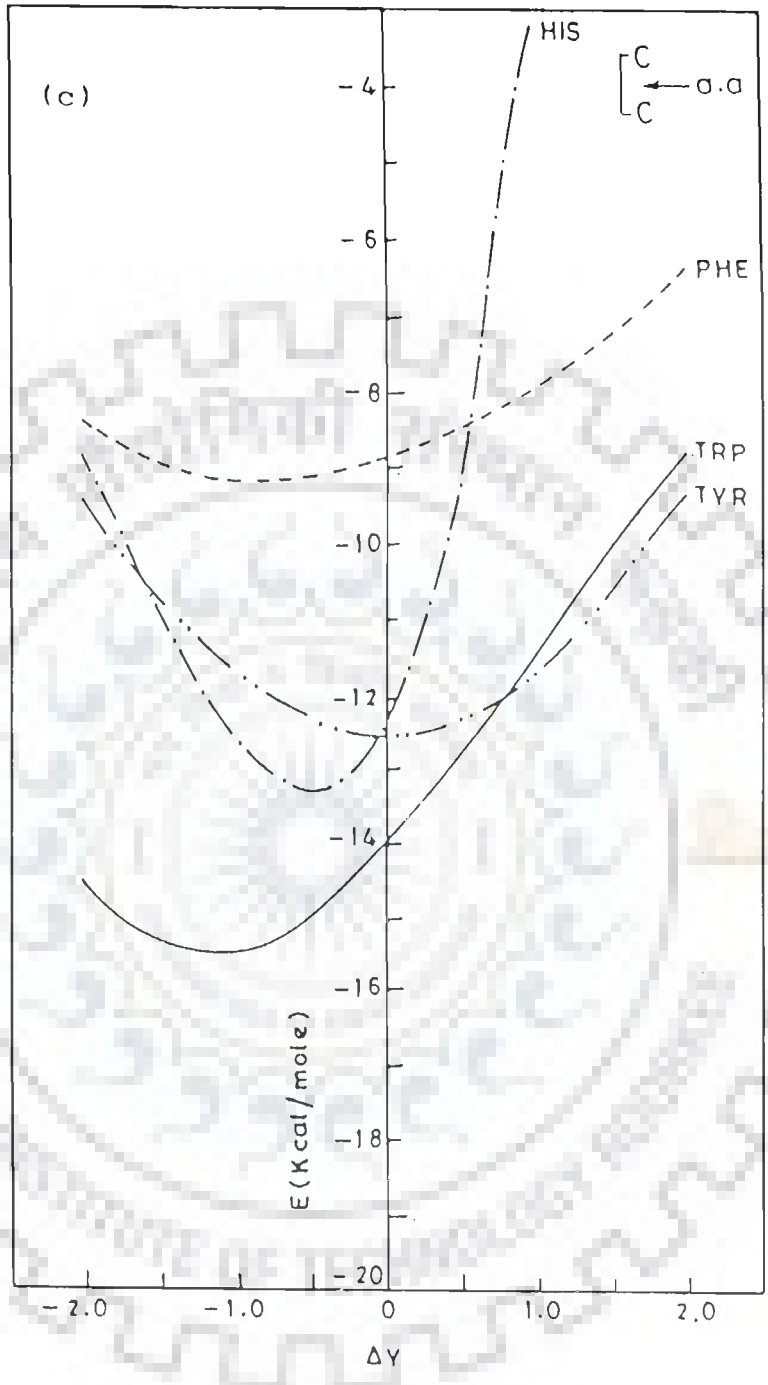
Contd....


(d)
G
a.a
G
-2
-4
-6
-8
-10
-12
-14
-16
E (K cal/mole)
0
40
80
120
160
200
240
280
320
360
Δθ (degree)
PHE
TYR
HIS
TRP


(a)
C
C
a.a
PHE
TYR
HIS
TRP
E (K cal / mole)
ΔZ (Å)
-5
-7
-9
-11
-13
-15
-17
6.0
6.2
6.4
6.6
6.8
7.0

Contd....


(b)
C
C
a.a
HIS
PHE
TRP
HIS
-4
-6
-8
-10
-12
-14
-16
-18
-20
E (K cal/mole)
-2.0
-1.0
0
1.0
2.0
ΔX


(c)
HIS
C
C
a.a
PHE
TRP
TYR
-4
-6
-8
-10
-12
-14
-16
-18
-20
E (K cal/mole)
-2.0
-1.0
0
1.0
2.0
ΔY

Contd....

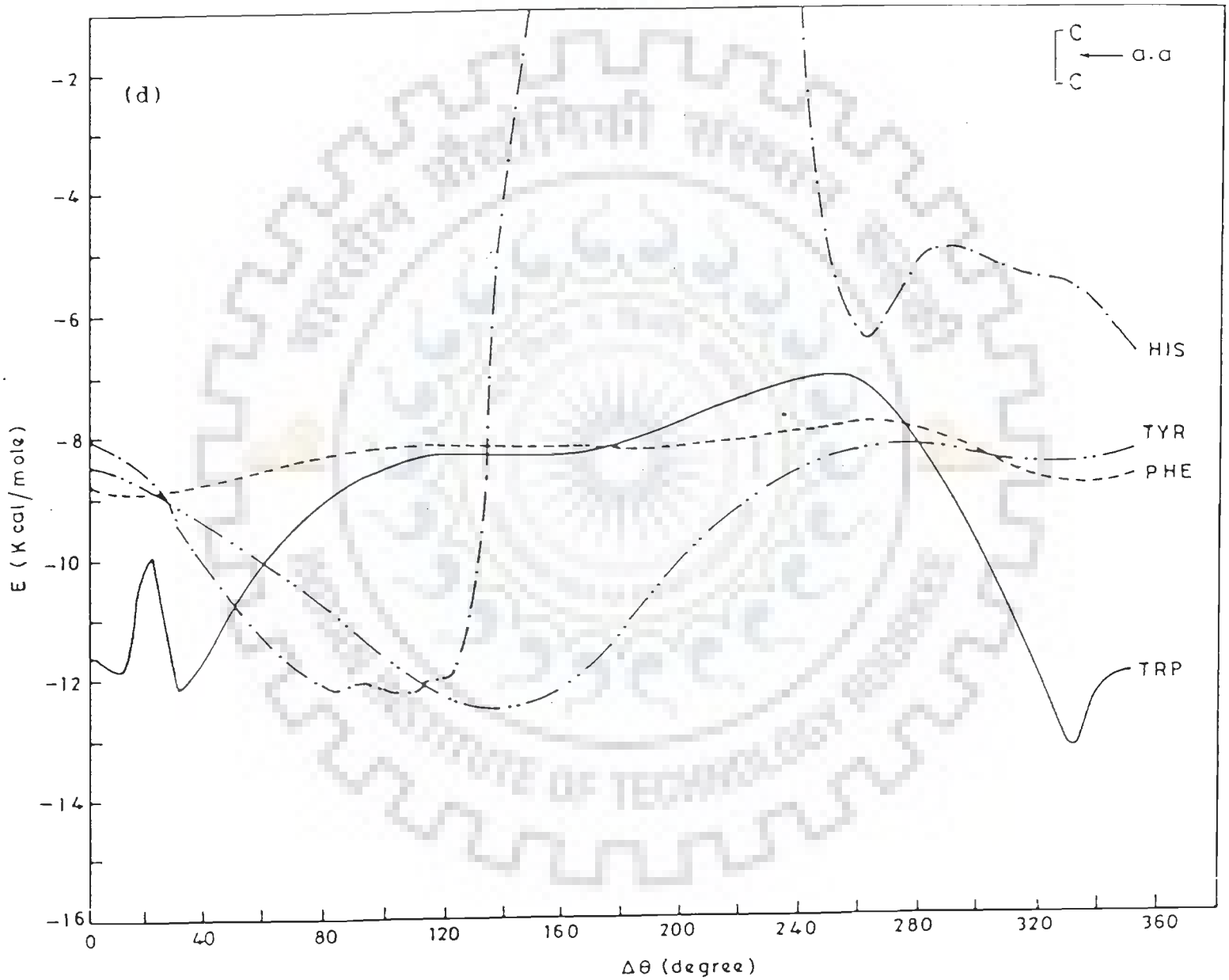

(d)
C
a.a
C
HIS
TYR
PHE
TRP
E (K cal/mole)
-2
-4
-6
-8
-10
-12
-14
-16
0
40
80
120
160
200
240
280
320
360
Δθ (degree)

Fig.7.2 shows the overlap geometries of aromatic amino acids placed between two bases corresponding to the minimum interaction energies. The extent of overlapping is more with the lower base in all the cases however His in general shows less overlap with the two bases. Thus the trend in extent of overlap does not follow the trend in values of minimum energies. Such a variation may result from the marginal differences in corresponding stacking energies. In fact there may be many other configurations which are associated with binding energies within ± 0.5 Kcal/mole of that of the optimised structure. Such conformational freedom is very significant from the biological point of view.

Table 7.1 tabulates the optimised stacking energy values for complex of aromatic amino acid Trp, Tyr, Phe and His when placed midway between two bases. Table 7.2 gives the partitioning of total stacking energies into various components for these complexes. It is observed that maximum contribution to stacking energies comes from dispersion term. Comparison of stacking energies of the complexes of aromatic amino acids between two bases shows variation which is of the order of 1-7 Kcal/mole. Complexes of Trp are involved in more efficient stacking than complexes of other aromatic amino acids. The relative interaction energy, in general decreases in the order:

$$\text{amino acid} - \begin{matrix} A \\ A \end{matrix} > \begin{matrix} T \\ T \end{matrix} > \begin{matrix} C \\ C \end{matrix} > \begin{matrix} G \\ G \end{matrix}$$

Among aromatic amino acids the stability of the complexes follows the trend:

Trp > His > Tyr > Phe.

It is found that the double-stranded DNA (Table 6.1 and Table 7.1) form stabler complex than single-stranded DNA. These results are consistent with those reported in literature (65). Further, the complex of aromatic amino acid with two bases is more stable than that with a base. This suggests that more energy is required to

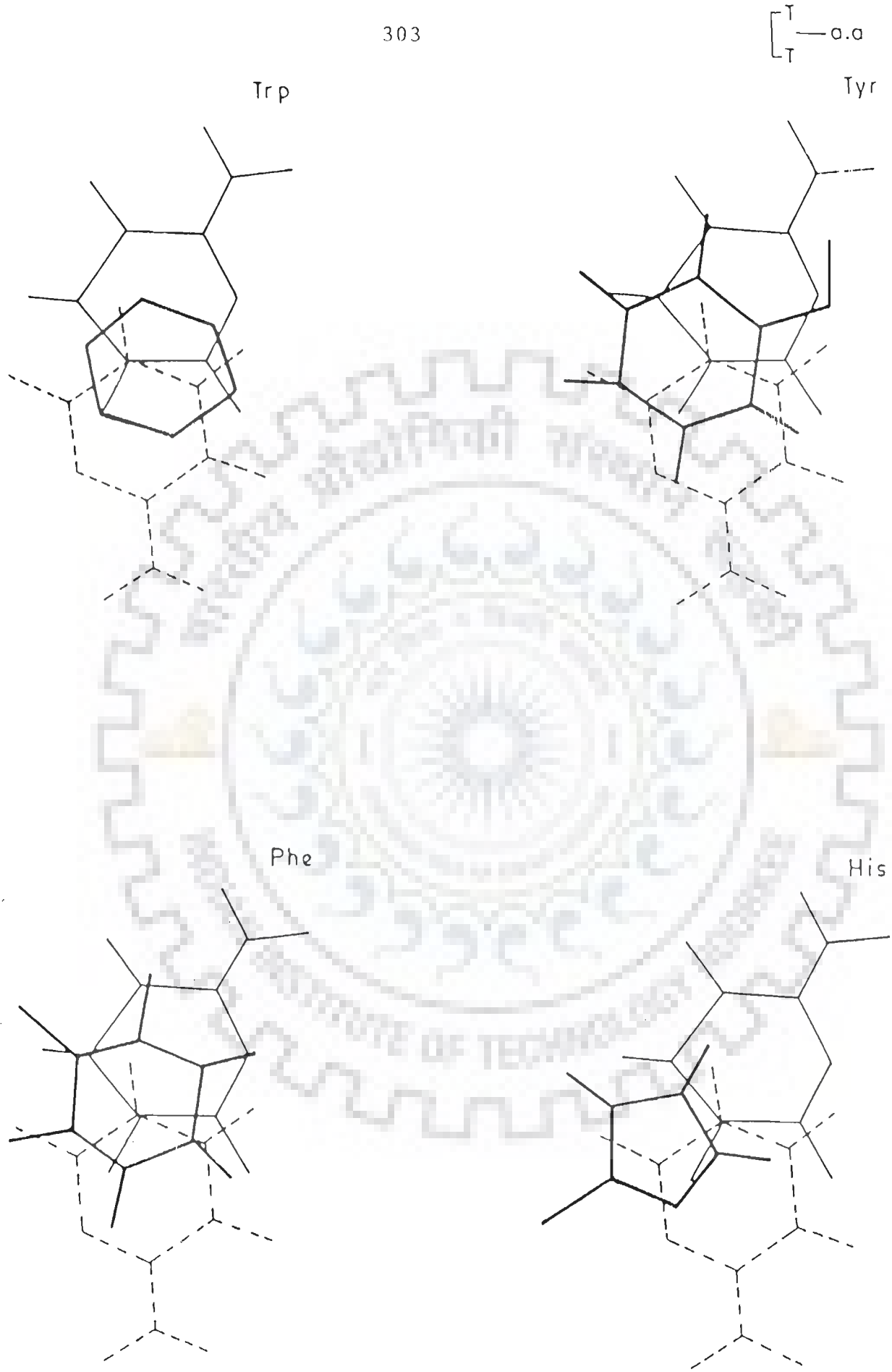
TABLE 7.1 : Summary of stacking energies (Kcal/mole)

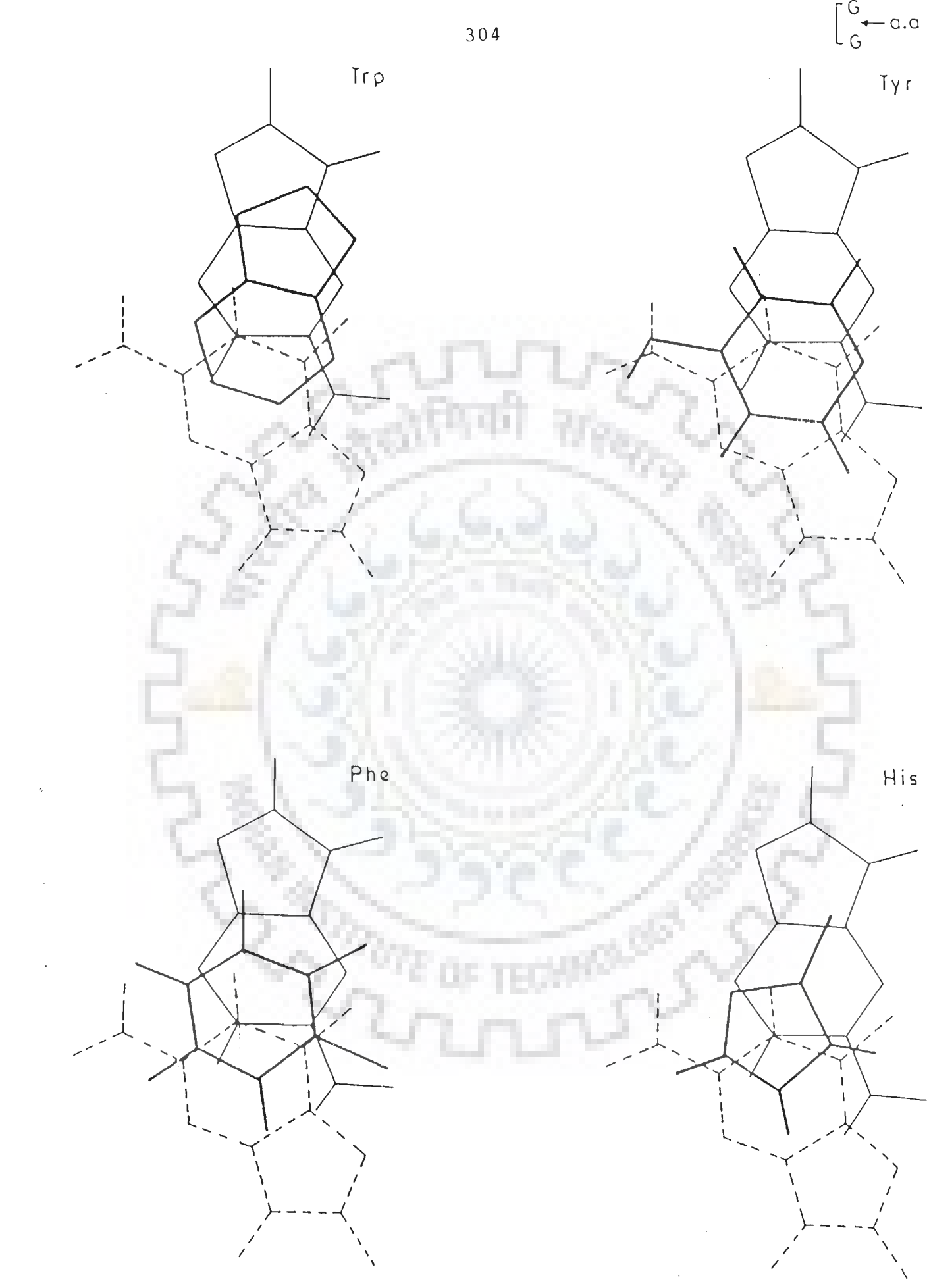
	Trp	Tyr	Phe	His
AA	-16.67	-13.17	-9.58	-14.87
TT	-17.89	-13.63	-12.00	-14.54
GG	-16.49	-13.46	-11.18	-15.90
CC	-16.30	-13.36	-9.55	-13.28

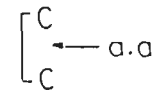
TABLE 7.2 : Partitioning of stacking energies (Kcal/mole) into various components.

		D	E_{el}	E_{pol}	E_{disp}	E_{rep}	E_{total}
AA	Trp	6.2	-4.5	-3.1	-28.0	18.9	-16.7
	Tyr	6.4	-4.1	-1.6	-17.5	10.0	-13.2
	Phe	6.4	-0.5	-2.5	-18.2	11.7	-9.6
	His	6.4	-6.1	-1.4	-16.7	9.3	-14.9
TT	Trp	6.6	-3.6	-2.0	-25.6	13.3	-17.9
	Tyr	6.4	-3.9	-1.3	-18.9	10.5	-13.6
	Phe	6.6	-1.9	-1.7	-16.5	8.1	-12.0
	His	6.4	-4.8	-1.9	-15.2	7.3	-14.5
GG	Trp	6.8	-1.7	-1.9	-23.3	10.4	-16.5
	Tyr	6.6	-2.4	-1.5	-17.9	8.4	-13.5
	Phe	6.6	-0.7	-1.4	-17.9	8.8	-11.2
	His	6.6	-3.2	-2.2	-19.3	8.8	-15.9
CC	Trp	6.8	-3.2	-2.3	-19.8	8.9	-16.3
	Tyr	6.6	-4.0	-1.4	-15.1	7.2	-13.4
	Phe	6.8	-0.9	-1.1	-13.2	5.6	-9.6
	His	6.6	-4.0	-1.9	-14.7	7.2	-13.3

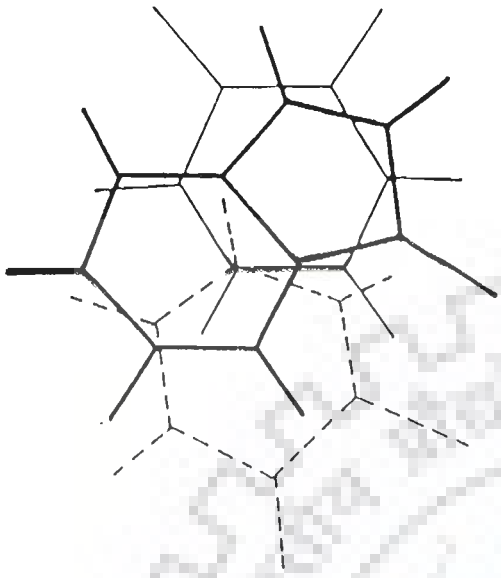
FIG. 7.2 Overlap geometries of complexes of aromatic amino acid placed between two bases.


T
T
a.a
Trp
Tyr
Phe
His

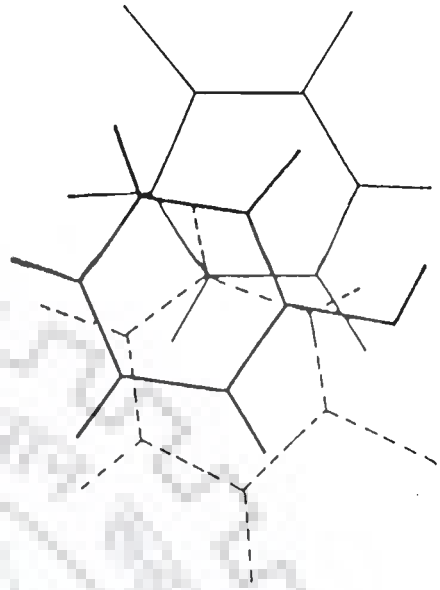

G
G
a.a
Trp
Tyr
Phe
His



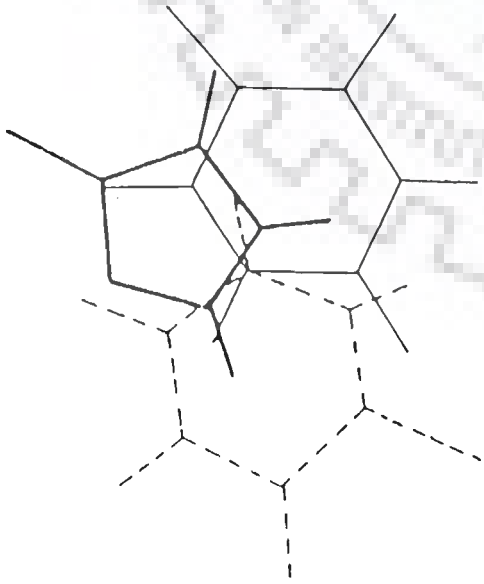
Trp



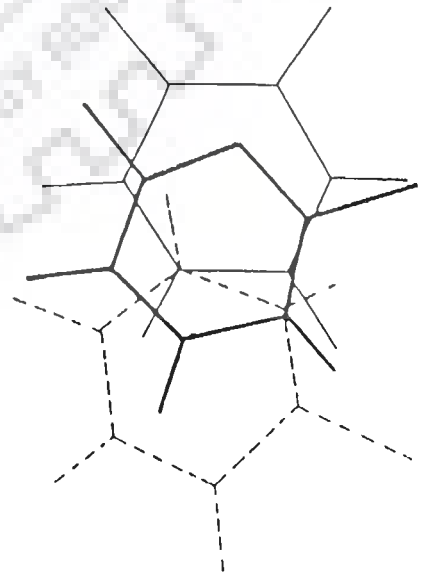
Tyr



His



Phe



Contd....

bring the base closer to aromatic amino acid. This is found to be true in all the cases.

The present study has been undertaken to understand the specificity of binding of aromatic amino acids to nucleic acid constituents. This has important implications. A special class of proteins called `DNA binding proteins' binds to DNA specifically. One such protein is Gene V protein of bacteriophage fd. This protein has the unique property of identifying strandedness of DNA. Thus energetically speaking, it is easy to compare the minimum energies of the complexes and determine the most suitable conformation. The present results indicate preference to double-stranded DNA than single-stranded DNA which is consistent with that reported in literature (65). This can be explained by the unwinding property of these DNA binding proteins. It may be possible that this class of proteins binds to double-stranded DNA, opens it up and finally binds to single-stranded DNA.

CHAPTER VIII

CONCLUSIONS

The detailed 1D and 2D NMR analysis based on spin-spin coupling and interproton distances show that $d(A)_5$ is present as a right handed B-DNA with glycosidic bond angle as high anti and sugar in S-conformational state. A minor N-conformer is present in deoxyribose sugar of A4 and A5 residue. On interaction with hexapeptide Lys-Pro-Tyr-Ser-Leu-Asn, it does not undergo any major change in conformation. The Leu δ-CH_3, Tyr(3,5)H, Tyr(2,6)H and Tyr β-CH_2 are broadened remarkably on interaction of $d(A)_5$ with KPYSLN. The Pro δ-CH_2, Lys β-CH_2 and Lys γ-CH_2 are also broadened significantly. Large changes in chemical shift ~ 0.17 to 0.29 ppm are observed in Lys β-CH_2, Lys δ-CH_2, Pro β-CH_2 and Leu δ-CH_3 proton due to binding. The base protons of $d(A)_5$ shift downfield/upfield upto 0.14 ppm while some of the sugar protons of A2, A3 residue shift upto 0.24 ppm. NOE connectivities Pro δ-CH_2 - A1H2' and Leu α-CH - A2H2' reflect proximity of Pro and Leu residues to $d(A)_5$ in the complex. A direct evidence of participation of Ser and Asn residue is not observed. The $d(A)_5$-KPYSLN complex appears to be stabilised by hydrophobic interactions. Complete unstacking of adenine bases to intercalate Tyr aromatic ring seems to be highly unlikely. The interaction of KPYSLN with octamer d-$(GACTCGTC)_2$ is relatively very weak and nonspecific. The present study on model systems has thus shown the role of each peptide residue of KPYSLN.

The energy calculations on stacking of aromatic amino acids Trp, Tyr, Phe and His with nucleic bases (A,T,G,C), base-pairs (AT,CG) show that G and C form stabler complexes with aromatic amino acids than A and T. CG forms a more stable complex than AT base-pair. Further among the model dinucleotide systems, d-CC and d-AA form stabler complexes with aromatic amino acids than d-CG and d-AT. In the energetically optimized configuration, the constituent aromatic moiety overlaps only partially. Unwinding/winding of the DNA helix have been observed. The optimised unwinding angle for all systems varies within a range similar to that reported for drugs. The energies for intercalation of aromatic amino acids between two bases have been computed. Among bases, Trp forms more

stable complexes. These calculations were performed to understand the relative stacking of aromatic acids with nucleic acid constituents and the results suggest that stacking contribute to specific recognition.

REFERENCES

1. Alberts, B. and Frey, L. (1972) Isolation and characterisation of gene V protein of filamentous bacterial viruses. J. Mol. Biol. 68, 139.

2. Alma, N.C.M., Harmsen, B.J.M., van Boom, J.H., van der Marel, G. and Hilbers, C.W. (1983) A 500 MHz proton NMR study of the structure and structural alterations of G5P - oligo (deoxyadenylic acid) complexes. Biochemistry 22, 2014.

3. Alma, N.C.M., Harmsen, B.J.M., van Boom, J.H., van der Marel, G. and Hilbers, C.W. (1982) ^{1}H NMR studies of the binding of bacteriophage M13 encoded gene 5 protein to oligodeoxyadenylic acids of varying length. Eur. J. Biochem. 122, 319.

4. Alma, N.C.M., Harmsen, B.J.M., Hull, W.E., van Boom, J.H., Van der Marel, G. and Hilbers, C.W. (1981) Double resonance experiments at 500 MHz on G5P and its complex with octariboadenylic acid. Biochemistry 20, 4419.

5. Alma, N.C.M., Harmsen, B.J.M., DeJong, E.A.M., Van der ven, J. and Hilbers, C.W. (1983) Fluorescence studies of the complex formation between the gene 5 protein of bacteriophage M13 and polynucleotides. J. Mol. Biol. 163, 47.

6. Alma, N.C.M., Harmsen, B.J.M., Hilbers, C., van der Marel, G.and van Boom, J.H. (1981) A 500 MHz ^{1}H NMR study of the role of lysines and arginines in the binding of the gene 5 protein to oligoadenylic acids. FEBS Letts. 135, 15.

7. Altona, C. and Sundaralingam, M. (1973) Conformational analysis of the sugar ring in nucleosides and nucleotides improved method of interpretation of proton magnetic

resonance coupling constants. J. Amer. Chem. Soc. USA 95, 2333.

8. Alexander, M.E., Burgum, A.A., Noall, R.A., Shaw, M.D. and Matthews, K.S. (1977) Modification of tyrosine residues of the lactose repressor protein. Biochem. Biophys. Acta. 493, 367.

9. Anderson, E., Nakashima, Y. and Coleman, J. E. (1975a) Chemical modifications of functional residues of a gene 5 DNA-binding protein. Biochemistry 14, 907. (b) Anderson, E., Nakashima, Y. and Koningsberg, W. (1975b) Photo-induced cross-linkage of gene 5 protein and bacteriophage fd DNA, Nucleic Acids Res. 2, 361.

10. Arcaya, G., Pantoja, M.E., Pieber, M., Romero, C. and Toha, J.C. (1971) *Z. Naturforsch Tiel.* B26, 1026. Reference cited in N.V.Kumar's Thesis , TIFR Bombay.

11. Arfmann, H.A., Labitzke, R., Lawaczeck, R. and Wagner, K.G. (1974) Aromatic amino acid-lysine copolymers. Conformation and specificity of nucleotide interaction. Biochimie 56, 53.

12. Aue, W.P., Bartholdi, E. and Ernst, R.R. (1976) Two dimensional Spectroscopy: Application to nuclear magnetic resonance. J. Chem. Phys. 64, 2229.

13. Barthwal, R., Kukreti, S. and Mujeeb, A. (1992) A 500 MHz proton NMR study on stacking interactions: Binding of d-CpG to tripeptides Lys-Tyr-Lys and Lys-Phe-Lys. Ind. J. Biochem. Biophys., 29, 394.

14. Barthwal, R., Kukreti, S., Mujeeb, A., Gupta, A. and Govil, G. (1991) A 500 MHz proton NMR study : Binding of tetranucleotide d-GpCpGpC to tripeptide Lys-Tyr-Lys. J. Mol. Recog. 4, 45.

15. Barthwal, R., Agarwal, A., Kukreti, S. and Mujeeb, A. (1987) Interaction of tryptophan containing oligopeptide with d-CpGpCpG by proton NMR. Physiol. Phys. Chem. Med. NMR. 19, 125.

16. Barthwal, R., Lancelot, G., Agarwal, A., Mujeeb, A. and Kukreti, S. (1988) Proton magnetic resonance studies of the binding of oligopeptides containing tryptophan to polyribonucleotides Poly(A), Poly(U) and Poly(C) Physiol. Phys. Chem. Med. NMR. 20, 145.

17. Barthwal, R., Mujeeb, A. and Govil, G. (1994) Interaction of Daunomycin with Deoxydinucleotide d-CpG by Two Dimensional Proton Magnetic Resonance Techniques. Archives of Biochemistry and Biophysics 313, 189.

18. Bobst, E.V, Perrino, F.W., Meyer, R.R. and Bobst, A.M. (1991) An EPR study to determine the relative nucleic acid binding affinity of single-stranded DNA binding protein from Escherichia coli. Biochimica et Biophysica Acta. 1078, 199.

19. Bodenhausen, G., Freeman, R. and Turner, D.L. (1977) Supression of artifacts in two dimensional spectroscopy. J. Magn. Reson. 27, 511.

20. Brayer, G.D. and McPherson, A. (1983) Refined structure of the gene 5 DNA-binding protein from bacteriophage fd. J. Mol. Biol. 169, 565.

21. Brayer, G.D. and McPherson, A. (1984a) Cooperative interactions of the gene 5 protein. J. Biomol. Struct. Dyn. 2, 495. Brayer, G.D. and McPherson, A. (1984b) Mechanism of DNA binding of the gene 5 protein of bacteriophage fd. Biochemistry 23, 340.

22. Brayer, G.D. and McPherson, A. (1985a) A model for intracellular complexation between gene 5 protein and bacteriophage fd DNA. Eur. J. Biochem. 150, 287. Brayer, G.D. and McPherson, A. (1985b) Topological comparison of 2 helix destablising proteins-ribonuclease A and the gene 5 DNA binding protein. J. Biomol. Struct. Dyn. 3, 173.

23. Brayer, G.D. (1987) A preliminary structure for the DNA binding protein from bacteriophage IKe. J. Biomol. Struct. Dyn. 45, 859.

24. Buck, F., Ruterjans, H., Kaptein, R. and Beyreuther, K. (1980) Photochemically induced dynamic nuclear polarisation investigation of complex formation of the NH_2-terminal DNA binding domain of lac-repressor with poly d(AT). Proc. Natl. Sci. U.S.A. 77, 5145. (b) Brown, L.R., Demarco, A., Wagner, G. and Wuthrich, K. (1976) Eur. J. Biochem. 62, 103.

25. Bulsink, H., Harmsen, B.J.M. and Hilbers, C.W. (1981)Specificity of the binding of bacteriophage M13 encoded gene 5 protein to DNA and RNA studied by means of fluorescence titrations. J. Biomol. Struct. Dyn. 3, 227.

26. Bulsink, H., Vanresandt, R.W.W., Harmsen, B.J.M. and Hilbers, C.W. (1986) Different DNA binding modes and cooperativities for bacteriophage M13 gene 5 protein revealed by means of fluorescence depolarisation studies. Eur. J. Biochem. 157, 329.

27. Model, P. and Russel, M. In `The Bacteriophages' Vol II, Plenum press, New York, pp 375, (1988).

28. Cavalieri, S.J., Neet, K.E. and Goldthwait, D.A. (1976) Gene V protein of bacteriophage fd: A dimer which interacts cooperatively with DNA. J. Mol. Biol. 102, 697.

29. Caillet, J. and Claverie, P. (1975) Theoretical evaluation of the intermolecular interaction energy of a crystal: Application to the analysis of crystal geometry. Acta Cryst. A31 , 448.

30. Chen, K. X., Gresh, N. and Pullman, B. (1987) A theoretical exploration of conformational aspects of ethidium bromide intercalation into a $d(CpG)_2$ minihelix. Biopolymers 26, 831

31. Claverie, P. In 'Intermolecular Interactions : From Diatomics to Biopolymers' Pullman, B. Ed. John Wiley, New York, pp 69 (1978).

32. Coleman, J.E. and Oakley, J.L. (1980) Physico Chemical studies of the structure and function of the DNA binding (helix destablising) proteins. Crit. Rev. Biochem. 7, 247.

33. O'Connor, T.P. and Coleman, J.E. (1983) Proton NMR (500 MHz) of mono-, di-, tri-, and tetradeoxynucleotide complexes of gene 5 protein. Biochemistry 22, 3375.

34. Coleman, J. E., Anderson, R.A., Radcliffe, R.G. and Armitage, I.M. (1976) Structure of gene 5 protein-oligodeoxynucleotide complexes as determined by ^{1}H, ^{19}F and ^{31}P nuclear magnetic resonance. Biochemistry 15, 25.

35. Coleman, J.E. and Armitage, I.M. (1978) Tyrosyl base phenylalanyl intercalation in G5P-DNA complexes. Proton magnetic resonance of selectively deuterated G5P. Biochemistry 17, 5038.

36. Cuypers, T., van der Ouderaa, F.J. and de Jong, W.W. (1974) The amino acid sequence of gene V protein of bacteriophage M13. Biochem. Biophys. Res. Commun. 59, 557.

37. Davies , D.B. (1978) Jerus. Sympos. Quant. Chem. Biochem., 11 , 71.

38. Day, L.A. (1972) Circular dichroism and ultraviolet absorption of a deoxyribonucleic acid binding protein of filamentous bacteriophage. Biochemistry 12, 5330.

39. De Jong, E.A.M., Harmsen, B.J.M., Konings, R.N.H. and Hilbers, C.W. (1987) Binding of Gene V protein to polynucleotides. Fluorescence binding experiments of IKe Gene V protein and mutual cooperativity of IKe and M13 Gene V proteins. Biochemistry 26, 2039.

40. De Jong, E.A.M, Harmsen, B.J.M., Konings, R.N.H., Prinse, C.W.J.M. and Hilbers, C.W. (1987) ^{1}H NMR studies of the gene 5 protein encoded single-stranded DNA binding protein of the filamentous bacteriophage IKe. Eur. J. Biochem. 167, 563.

41. Dick, L.R., Geraldes, C.F.G.C., Gerry, D.A., Gray, C.W. and Gray, D.M. (1989) ^{13}C NMR of methylated lysines of fd Gene 5 Protein: Evidence for a conformational change involving lysine 24 upon binding of a negatively charged lanthanide chelate. Biochemistry, 28 7896. (b)Dick, L.R., Sherry, A.D., Nenkirk, M.M. and Gray, D.M. (1988) J. Biol. Chem. 263, 18864.

42. Dimicoli, J.L. and Helene, C. (1974a) Interactions of aromatic residues of proteins with nucleic acids I.Proton magnetic resonance studies of the binding of tryptamine and tryptophan containing peptides to poly(adenylic acid) and deoxyribonucleic acid. Biochemistry 13, 714. Dimicoli, J.L. and Helene, C. (1974b) Interactions of aromatic residues of proteins with nucleic acids II. Proton magnetic resonance studies of the binding of tyramine and tyrosine- containing peptides to poly(adenylic acid) and deoxyribonucleic acid.

Biochemistry 13, 724.

43. Derome, A.E. In `Modern NMR Techniques for Chemistry Research' Volume 6 Pergamon Press, Oxford U.K.(1987).

44. Durand, M., Borazan, H.N., Maurizot, J.C. and Helene, C. (1975) Interaction of aromatic residues of protein with nucleic acids. Circular dichroism studies of the binding of oligopeptides to polyadenylic acid. Biochemistry 14, 563.

45. Eddington, P. and Harding, M.M. (1974) The crystal structure of DL-Histidine. Acta Cryst. 30, 204

46. Fanning, T.G. (1975) Iodination of E.Coli lac repressor. Effect of tyrosine modification on repressor activity. Biochemistry 14, 2512.

47. Feigon, J., Wang, A.H.J., van der Marel, G.A., van Boom, J.H. and Rich, A. (1984) A one and two dimensional NMR study of the B to Z transition of $(m^5dC\text{-}dG)_3$ in the methanolic solution. Nucleic Acids Res. 12, 1243.

48. Feigon, J., Wright, J.M., Leupin, W., Denny, W.A. and Kearns, D.R. (1982) Use of two dimensional NMR in the study of a double-stranded decamer. J. Amer. Chem. Soc. USA 104, 5540.

49. Folker, P.J.M., van Duynhoven, J.P.M., Jonker, A.J., Harmsen, B.J.M., Konings, R.N.H. and Hilbers, C.W. (1991) Sequence specific ^1H NMR assignment and secondary structure of the Tyr 41-->His mutant of the single-stranded DNA binding protein, Gene V protein, encoded by the filamentous bacteriophage M13. Eur. J. Biochem. 202, 349.

50. Folkers, P.J.M., Stassen, A.P.M., van Duynhoven, J.P.M., Harmsen, B.J.M., Konings, R.N.H. and Hilbers, C.W. (1991) Characterization of wild type and mutant M13 gene V proteins

by means of ^{1}H-NMR. Eur. J. Biochem. 200, 139.

51. Folkers, P.J.M, van Duynhoven, J.P.M., van Lieshout, H.T.M., Harmsen, B.J.M, van Boom, J.H., Tesser, G.I., Konings, R.N.H. and Hilbers, C.W. (1993) Exploring the DNA binding domain of Gene V Protein encoded by bacteriophage M13 with the aid of spin-labeled oligonucleotides in combination with ^{1}H-NMR. Biochemistry 32, 9407.

52 Folkers, P.J.M., Nilges, M., Folmer, R.H.A., Konings, R.N.H. and Hilbers, C.W. (1994) The solution structure of the Tyr 41—>His mutant of the single-stranded DNA binding protein encoded by Gene V of the filamentous bacteriophage M13. J. Mol. Biol. 236, 229.

53. Folmer, R.H.A., Nilges, M., Folkers, P.J.M., Konings, R.N.H. and Hilbers, C.W. (1994) A model of the complex between single-stranded DNA and the single-stranded DNA binding protein encoded by Gene V of filamentous bacteriophage M13. J. Mol. Biol. 240, 341.

54. Frey, M.N., Koetzle, T.F., Lehmann, M.S. and Hamilton, W.C. (1973) Precision neutron diffraction structure determination of protein and nucleic acid components. A comparison between the crystal and molecular structure of L-tyrosine and L-tyrosine hydrochloride. J. Chem. Phys. 58, 2547.

55. Frechet, D., Cheng, D.M., Kan, L.S. and Tso, P.O.P. (1983) Nuclear Overhauser effect as a tool for the complete assignment of non-exchangeable proton resonances in short deoxyribonucleic acid helices. Biochemistry 22, 5194.

56. Gabbay, E.J., Sanford, K., Baxter, C.S. and Kapicak, L. 1973) Specific interaction of peptide with nucleic acids. Evidence for a "selective bookmark" recognition hypothesis. Biochemistry 12, 4021.

57. Garssen, G.J., Hilbers, C.W., Schoenmakers, J.G.G. and van Boom, J.H. (1977) Studies on the DNA unwinding: Proton and phosphorus NMR studies of G5P from bacteriophage M13, interacting with d-(CGCG). Eur. J. Biochem. 81, 453.

58. Garssen, G.J., Tesser, G.I., Schoenmakers, J.G.G. and Hilbers, C.W. (1980) NMR studies of the interaction of G5P of bacteriophage M13 with oligonucleotides. Biochem. Biophys. Acta. 607, 361.

59. Gray, C.W., Kneale, G.G., Leonard, K.R., Siegrist, H. and Marvin, D.A. (1982) A nucleoprotein complex in bacteria infected with Pf1 filamentous virus: identification and electron microscopic analysis. Virology 116, 40.

60. Gray, C.W. (1989) Three dimensional structure of complexes of single-stranded DNA binding proteins with DNA. J. Mol. Biol. 208, 57.

61. Gray, D.M., Gray, C.W. and Carlson, R.D. (1982) Neutron scatterring data on reconstituted complexes of fd deoxyribonucleic acid and gene 5 protein show that the deoxyribonucleic acid is near the center. Biochemistry 21, 2702.

62. Gray, C.W., Page, G.A. and Gray, D.M. (1984) Complex of fd gene V protein and double-stranded RNA. J. Mol. Biol. 75, 553.

63. Gronenborn, A. M. and Clore, G.M. (1985) Investigation of the solution structure of short nucleic acid fragments by means of nuclear Overhauser enhancement measurements. Progress in NMR spectroscopy 17, 1.

64. Gresh, N. and Pullman, B. (1979) A theoretical study of the

interaction of ammonium and guanidium ions with the phosphodiester linkage. Theor. Chim. Acta. 52, 67.

65. Govil, G., Hosur, R.V. and Miles, H.T. (1988) Protein-DNA interactions. Proc. Indian Natl. Sci. Acad. 54, 667.

66. Govil, G. and Hosur, R.V. In `Conformation of Biological Molecules: New results from NMR', Springer Verlag, Heidelberg (1982).

67. Govil, G., Khetrapal, C.L. and Saran, A. In `Magnetic resonance in Biology and Medicine' Tata McGraw Hill (1985).

68. Guschlbauer, W. In `Nucleic Acid Structure' Springer-Verlag New York (1975).

69. Hanstock, C.C. and Lown, J.W. (1984) Two dimensional double quantum spectroscopy for the proton NMR assignment of oligonucleotides. J. Magn. Reson. 58, 167.

70. Hare, D.H., Wemmer, D.E., Chou, S.H., Drobny, G. and Reid, B.R. (1983) Assignment of the non-exchangeable proton resonances of d-$(CGCGAATTCGCG)_2$ using two dimensional methods. J. Mol. Biol. 171, 319.

71. Hasnoot, C.A.G., van der Ven, J. M. and Hilbers, C.W. (1984) Sequential assignments for the ^{1}H and ^{31}P atoms in the backbone of oligonucleotide by two dimensional NMR. J. Amer. Chem. Soc. USA 105, 1652.

72. Helene, C. and Dimicoli, J.L. (1972) Interaction of oligopeptides containing aromatic amino acids with nucleic acids. Fluorescence and proton magnetic resonance studies. FEBS Letts. 26, 6.

73. Helene, C. and Lancelot, G. (1982) Interaction between

functional groups in Protein-nucleic acid associations. Prog. Biophys. Mol. Biol. 39, 1.

74. Hilbers, C.W., Garssen, G.J., Kaptein, R., Schoenmakers, J.G.G. and van Boom, J.H. (1978) Studies of G5P-nucleic acid interactions: NMR in Molecular Biology. Ed. Pullman Proceedings of XI Jerusalem symposium on Quantum chemistry and Biochemistry held on April 3-7, 1978.

75. Hopfinger, A.J. In `Conformational properties of macromolecules', New York Academic Press (1973).

76. Hoppe, W., Lohman, W., Markl, H. and Zeigler, H. In `Biophysics' Springer Verlag, Berlin Heidelberg, (1983).

77. Hosur, R.V., Govil, G. and Miles, H.T. (1988) Application of two dimensional NMR spectroscopy in the determination of solution conformation of nucleic acids. Magn. Reson. Chem. 26, 927.

78. Hosur, R.V., Ravikumar, M., Chary, K.V.R., Sheth, A., Govil, G., Tan-Zu-Kun and Miles, H.T. (1986) Solution structure of d-GAATTCGAATTC by 2D NMR: a new approach to determination of sugar geometries in DNA segments. FEBS Letts. 205, 71.

79. Hosur, R.V., Sheth, A., Chary, K.V.R, Ravikumar, M. and Govil, G. (1986) A novel loop structure observed in d-GAATTCCCGAATTC by 2D NMR. Biochem. Biophys. Res. Comm. 139, 1224.

80. Hosur, R.V., Ravikumar, M., Roy, K.B., Tan-Zu-Kun, Miles, H.T. and Govil, G. In `Magnetic Resonance in Biology and Medicine' (Eds. G.Govil, C.L.Khetrapal and A.Saran) Tata McGraw Hill, New Delhi, pp-305 (1985).

81. Hutchinson, D.L., Barnett, B.L. and Bobst, A.M. (1990) Gene 5

Protein-DNA complex: Modeling Binding Interactions. J. Biol. Struc. Dyn. 1, 1.

82. Jeener, J. (1971) Ampere International summer school Basko. Polje Yogaslavia (unpublished data).

83. Kansy, J.W., Clack, B.A. and Gray, D.M. (1986) The binding of fd gene 5 protein to polydeoxynucleotides: Evidence from CD measurements for two binding modes. J. Supramol. Struct. 3, 1079.

84. Keeler, J. and Neuhaus, D. (1985) Comparison and evaluation of methods for two dimensional NMR spectra with absorbtion mode lineshape. J. Magn. Reson. 63, 454.

85. King, G.C. and Coleman, J.E. (1987) Two dimensional ^{1}H NMR of Gene 5 Protein indicates that only two aromatic rings interact significantly with oligodeoxynucleotide bases. Biochemistry 26, 2929.

86. King, G.C. and Coleman, J.E. (1988) The Gene 5 Protein-$d(pA)_{40-60}$ complex: ^{1}H NMR supports a localized base-binding model. Biochemistry 27, 6947.

87. Kim, Y.T., Tabor, S., Churchich, J.E. and Richardson, C.C. (1992) Interactions of Gene 2.5 Protein and DNA Polymerase of Bacteriophage T7. J. Biol. Chem. 267, 15032.

88. Kneale, G.G. and Plyte, S.E. (1991) Structural parameters of the Pf1 Gene 5 Protein-DNA complex in solution by Neutron scattering. J. Mol. Biol. 221, 755. (b) Plyte, S.E. and Kneale, G.G. (1993) Identification of a compact DNA-binding Domain in the Gene 5 Protein of Pf1 bacteriophage. Biochemistry 32, 3623.

89. Kumar , N.V. Ph.D Thesis, University of Bombay, 1983.

90. Kumar, N.V. and Govil, G. (1984) Theoretical studies on protein nucleic acid interactions III. Stacking of aromatic amino acids with bases and base-pairs of nucleic acids. Biopolymers 23, 2009.

91. Kowalczykowski, S.C., Bear, D.G. and Von Hippel, P.H. (1981) Single-stranded DNA binding proteins in `The Enzyme' 14 (Ed. P.D.Boyer) 373, Academic press, New York.

92. Lica, L. and Ray, D.S. (1977) Replication of bacteriophage M13 XII. *In vivo* cross-linking of a phage-specific DNA binding protein to the single-stranded DNA bacteriophage M13 by ultraviolet irradiation. J. Mol. Biol. 115, 45.

93. Lown, J.W., Hanstock, C.C., Bleackley, R.C., Imbach, J.L., Rayner, B. and Vasseur, J.J. (1984) Synthesis, complete ^1H assignment and conformations of the self complementary hexadeoxynucleotide d(CpGpApTpCpG)$_2$ and its fragment by high field NMR. Nucleic Acids Res. 12, 2519.

94. Maurizot, J.C., Durand, M., Dimicoli, J.L. and Helene, C. (1973) Stud. Biophys. 40, 91. Reference cited in N.V.Kumar's Thesis TIFR, Bombay.

95. Mayer, R., Toulme, F., Montenay-Garestier, T. and Helene, C. (1978) The role of tyrosine in the association of proteins and nucleic acids. J. Biol. Chem., 254, 75.

96. Marion, D. and Wuthrich, K. (1983) Application of phase-sensitive two dimensional correlated spectroscopy (COSY) for measurement of ^1H-^1H spin coupling constants in proteins. Biochem. Biophys. Res. Commun. 113, 967.

97. Majumdar, A. and Hosur, R.V. (1992) Simulation of 2D spectra for determination of solution conformations of nucleic acids.

Progress in NMR spectroscopy 23, 109.

98. McPherson, A., Jurnak, F.A., Wang, A.H.J., Molineux, I. and Rich, A. (1979) Structure at 2.3 A resolution of the gene V protein of bacteriophage fd : A DNA unwinding protein. J. Mol. Biol. 134, 379.

99. McPherson, A., Wang, A.H.J., Jurnak, F.A., Molineux, F.K. and Rich, A. (1980) X-ray diffraction studies on crystalline complexes of G5P DNA unwinding protein with deoxyoligonucleotides. J. Biol. Chem. 255, 3174.

100. Miller, K.J. In `Biomolecular Stereodynamics' II (Ed. Sarma, R.H.) Adenine press, New York, (1981).

101. Munt, N.A. and Kearns, D.R. (1984) Poly (dA-dT) has a right handed B conformation in solution: A two dimensional NMR study. Biochemistry 23, 791.

102. Mujeeb, A., Kerwin, S.M., Egan, W., Kenyon, G.L. and James, T.L. (1992) A potential gene target in HIV-1: Rationale, Selection of a conserved sequence, and determination of NMR distance and torsion angle constraints. Biochemistry 31, 9325.

103. Momany, F.A., McGuire, R.F., Burgess, A.W. and Scheraga, H.A. (1975) Energy parameter in polypeptide VII Geometric parameters, partial atomic charges, non-bonded interactions, hydrogen bonded interactions and intrinsic torsional potential for naturally occurring amino acids. J. Phys. Chem. 79, 2361. (b) Scheraga, H.A. (1968) Calculations of conformations of polypeptides. Adv. Phys. Org. Chem. 6, 103.

104. Nakashima, Y., Dunker, A.K., Marvin, D.A. and Koningsberg, W. (1974) The amino acid sequence of a DNA binding protein, the gene V product of fd filamentous bacteriophage. FEBS Letts.

40, 290.

105. Nilges, M., Clore, G.M., Gronenborn, M., Brunger, A.T., Karplus, M. and Nilsson, L. (1987) Refinement of the solution structure of the DNA hexamer 5'-d(GCATCG)$_2$. Combined use of nuclear magnetic resonance and restrained molecular dynamics. Biochemistry 26, 3318.

106. Nuss, M.E., Marsh, F.J. and Kollman, P.A. (1979) Theoretical studies of drug-dinucleotide interactions. Empirical energy function calculations on the interaction of Ethidium, 9-Aminoacridine and Proflavin cations with the base-paired dinucleotides GpC and CpG. J. Amer. Chem. Soc. USA 101, 825.

107. Olsthoorn, C.S.M., Haasnoot, C.A.G. and Altona, C. (1980) Circular dichroism studies of 6-N-methylated adenylyladenosine and adenylyluridine and their parent compounds. Thermodynamics of stacking . Eur. J. Biochem. 105, 85.

108. Olsthoorn, C.S.M., Bostelaar, L. J., van Boom, J. H. and Altona, C. (1980) Conformational characterstics of the trinucleoside diphosphate dApApA and its constituents from NMR and CD studies. Eur. J. Biochem. 112, 95. (b) Hingerty, B. and Broyde, S. (1978) Nucleic Acids Res. 5, 3249. (c) Thiyagarajan, P. and Ponnuswamy, P.K. (1979) Biopolymers 18, 789.

109. Ornstein, R.L., Rein, R., Breen, D.L. and MacElroy, R.D. (1978) An optimized potential function for the calculation of the nucleic acid interaction energies: Base stackings. Biopolymers 17, 2341.

110. Ornstein, R.L. and Rein, R. (1979) Energetics and structural aspects of ethidium cation intercalation into DNA minihelices. Biopolymers 18, 2821.

111. Paoletti, J. and Le Pecq, J.B. (1971) The change of torsion of the DNA helix caused by intercalation. Biochemie 53, 969.

112. Pople, J.A. and Segal, G.A. (1965) Approximate self consistent molecular orbital theory. I Invariant procedures. J. Chem. Phys. 43, 5136.

113. Patel, D.J. and Shapiro, L. (1985) Biochemie 67, 887.

114. Patel, D.J. (1979) Helix-coil transition of dG-dC-dG-dC self complementary duplex and complex formation with daunomycin in solution. Biopolymers 18, 553.

115. Pardi, A., Walker, R., Rapoport, H., Wider, G. and Wuthrich, K. (1983) Sequential assignment for the ^{1}H and ^{31}P atoms in the backbone of oligonucleotide by two dimensional nuclear magnetic resonance. J. Amer. Chem. Soc. USA 105, 1652.

116. Porschke, D. and Jung, M. (1982) Stability decrease of RNA double helices by phenylalanine, tyrosine and tryptophan amide. Analysis in terms of site binding and in relation to melting proteins. Nucleic Acids Res. 10, 6163.

117. Pretorius, H.T., Klein, M. and Day, L.A. (1975) Gene V protein of fd bacteriophage : Dimer formation and the role of Tyrosyl groups in DNA binding. J. Biol. Chem. 250, 9262.

118. Pullman, B. In 'Quantum Mechanics of molecular conformations' (Ed. B. Pullman) John Wiley and Sons, London (1976).

119. Ravikumar, M., Hosur, R.V., Roy, K.B., Miles, H.T. and Govil, G. (1985) Resonance assignment of the 500 MHz proton NMR spectrum of self complementary dodecanucleotide d-GGATCCGGATCC: Altered conformation at BamHI cleavage sites. Biochemistry 24, 7703.

120. Redfield, A.G., Kunj, S. and Ralph, E.K. (1975) Quadrature fourier NMR detection, simple multiplex for dual detection and discussion. J. Magn. Reson. 19, 116.

121. Reid, B.R., Baukes, K., Flynn, P. and Nerdal, W. (1989) NMR distance measurements in DNA duplexes: sugar and bases have the same correlation times. Biochemistry 28, 10001.

122. Rizo, J., Dhingra, M. and Gierasch, L.M. (1990) Peptide model for reverse turns. The role of Asparagine in the i position of a β turn. In `Molecular Conformation and Biological Interactions' by P. Balaram and S. Ramaseshan, Indian Academy of Science, Bangalore.

123. Saucier, J.M., Festy, B. and LePecq, J.B. (1971) The change of torsion of the DNA helix caused by intercalation: Measurement of the relative change of the torsion induced by various intercalating drugs. Biochemie 53, 973.

124. Sanyal, N.K., Roychoudhury, M. and Ojha, R.P. (1984) Molecular basis of drug action of some antibiotics. J. Theor. Biol. 110, 505.

125. Scheek, R.M., Russo, N., Boelens, R., Kaptein, R. and van Boom, J.H. (1983) Sequential resonance assignments in DNA ^{1}H NMR spectra by two dimensional NOE spectroscopy. J. Amer. Chem. Soc. USA 105, 2914.

126. Sang, B.C. and Gray, D.M. (1987) fd Gene V protein binds to double stranded polydeoxyribonucleotides poly(dA.dT) and poly [d(A-T).d(A.T)]. Biochemistry 26, 7210.

127. Sang, B.C. and Gray, D.M. (1989) CD measurements show that fd and IKe Gene V protein undergo minimal conformational changes upon binding to poly(rA). Biochemistry 28, 9502.

128. Salstrom, J.S and Pratt, D. (1971) Role of coliphage M13 in single-stranded DNA production. J. Mol. Biol. 61, 489.

129. Saenger, W. In `Principles of Nucleic Acid Structure' Springer Verlag, New York Inc (1984).

130. Sarma, M.H., Gupta, G. and Sarma, R.H. (1987) DNA structure in which an Ade-Cyt mismatch forms an integral part of the double helix. Biochemistry 26, 7707.

131. Sobell, H.M. and Jain, S.C. (1972) Stereochemistry of actinomycin binding to DNA. II. Detailed molecular model of actinomycin-DNA complex and its implications. J. Mol. Biol. 68, 21.

132. States, D.J., Haberkorn, R.A. and Ruben, D.J. (1982) A two dimensional nuclear Overhauser experiment with pure absorption phase in four quadrants. J. Magn. Reson. 48, 286.

133. Stassen, A.P.M., Harmsen, B.J.M., Schoenmakers, J.G.G., Hilbers, C.W. and Konings, R.N.H. (1992) Fluorescence studies of the binding of bacteriophage M13 gene V mutant proteins to polynucleotides. Eur. J. Biochem. 206, 605.

134. Stassen, A.P.M., Zaman, G.J.R., van Deursen, J.M.A, Schoenmakers, J.G.G. and Konings, R.N.H. (1992) Selection and characterization of randomly produced mutants of gene V protein of bacteriophage M13. Eur. J. Biochem. 204,1003.

135. Torbet, J., Gray, D.M., Gray, C.W., Marvin, D.A. and Siegrist, H. (1981) Structure of fd DNA-gene 5 protein complex in solution: A neutron small angle scattering study. J. Mol. Biol. 146, 305.

136. Ts'o, P.O.P (Ed.) In `Basic principles in Nucleic Acid

Chemistry', Vols. 1 and 2. Academic press, New York (1974).

137. van Duynhoven, J.P.M., Folkers, P.J.M., Stassen, A.P.M., Harmsen, B.J.M., Konings, R.N.H. and Hilbers, C.W. (1990) Structure of the DNA binding wing of the gene V encoded single-stranded filamentous bacteriophage M13. FEBS 261, 1.

138. van Duynhoven, J.P.M., Folkers, P.J.M., Prinse, C.W.J.M., Harmsen, B.J.M., Konings, R.N.H. and Hilbers, C.W (1992) Assignment of the ^{1}H NMR spectrum and secondary structure elucidation of the single-stranded DNA binding protein encoded by the filamentous bacteriophage IKe. Biochemistry 31, 1254. (b) van Duynhoven, J.P.M., Nooren, I.M.A., Swinkel, D.W., Folkers, P.J.M., Prinse, C.W.J.M., Harmsen, B.J.M., Konings, R.N.H., Tesser, G.L., Hilbers, C.W. (1993) Exploration of the ssDNA binding domains of gene V protein encoded by the filamentous bacteriophages IKe and M13 by means of a spin-labelled oligonucleotide and lanthanide-DOTP complexes. Eur. J. Biochem. 216, 507.

139. van Wijk, J., Huckriede, B.D., Ippel, J.H. and Altona, C. Furanose sugar conformations in DNA from NMR coupling constants. Methods in enzymology (Eds. Lilley, D.M.J. and Dahlberg, J.E.) Academic Press, New York 211, 286 (1992).

140. van dee Ven, F.J.M. and Hilbers, C.W. (1988) Nucleic acids and nuclear magnetic resonance Eur. J. Biochem. 178, 1.

141. Voet, D. and Rich, A. (1979) The crystal structure of purines, pyrimidines and their intermolecular complexes. Prog. Nucleic Acids Res. Mol. Biol. 10, 183.

142. Wagner, G., Kumar, A and Wuthrich, K. (1981) Systematic application of two dimensional ^{1}H nuclear magnetic resonance technique for studies of proteins. II Combined use of

correlated spectroscopy and nuclear Overhauser spectroscopy for sequential assignment of backbone resonances and elucidation of polypeptide secondary structures. Eur. JBiochem. 114, 375.

143. Wang, J. C. (1974) The degree of unwinding of the DNA helix by ethidium. I Titration of twisted PM2 bacteriophage in alkaline cesium chloride density gradients. J. Mol. Biol. 89, 783.

144. Waring, M. J. (1970) Variation of supercoils in circular closed DNA by binding of antibiotics and drugs: evidence of molecular models involving intercalation. J. Mol. Biol. 54, 247.

145. Wemmer, D.R. and Reid, B.R. (1986) Ann. Rev. Phys. Chem. 36, 105.

146. Westerlink, H.P., van der Marel, G.A., van Boom, J.H. and Hasnoot, C.A.G. (1984) Conformational analysis of r(CGCGCG) in aqueous solution: A type double helical conformation studied by two dimensional NOESY. Nucleic Acids Res. 12, 4323.

147. Wijmenga, S.S., Mooren, M.M.W. and Hilbers, C.W. In `NMR in macromolecules' (Ed. Roberts G.C.K.) pp. 217, IRL Press, Oxford (1993).

148. Woodson, S.A. and Crothers, D.M. (1988) Preferential location of bulged guanosine internal to a GC tract by NMR. Biochemistry 27, 436.

149. Wuthrich, K. In `NMR in Biological research: Peptides and Proteins', North Holland Amsterdam (1976).

150. Wuthrich, K. In `NMR of Proteins and Nucleic acids' John

Wiley, New York (1986).

151. Zaman, G.J.R., Kaan, A.M., Schoenmakers, J.G.G. and Konings, R.N.H. (1992) Gene V Protein mediated translational regulation of the synthesis of gene II protein of the filamentous bacteriophage M13: a dispensable function of the filamentous phage genome. J. Bacteriol. 174, 595.

152. Zaman, G., Smetsers, A., Kaan, A., Schoenmakers, J. and Konings, R. (1991) Regulation of expression of the genome of bacteriophage M13. Gene V protein regulated translation of the mRNAs encoded by gene I, III, V and X. Biochimica Biophysica Acta. 1089, 183.

153. Wang, A.H.J., Ughetto, G., Quigley, G.J. and Rich, A. (1987) Interactions between an anthracycline antibiotic and DNA: Molecular structure of daunomycin complexed to d-(CpGpTpApCpG) at 1.2 Å resolution. Biochemistry 26, 1152.

154. Nunn, C.M., Meervelt, L.V., Zhang, S., Moore, M.H. and Kennard, O. (1991) Drug DNA interactions: The crystal structures of d(TGTACA) and d(TGATCA) complexed with daunomycin. J. Mol. Biol. 222, 167.

155. Frederick, C.A., Williams, L.D., Ughetto, G., van der Marel, G.A., van Boom, J.H., Rich, A. and Wang, A.H.J. (1990) Structural comparison of anticancer drug-DNA complexes: Adriamycin and Daunomycin. Biochemistry 29, 2538.

156. Prabhakaran, M. and Harvey, S.C.(1988) Molecular dynamics of structural transition and interactions in DNA. Biopolymers 27 1239.

157. Skinner, M.M., Zhang, Hong, Leschnitzer, D.H.,Guan, Y., Bellawy, H., Sweet, R.M., Gray, C.W., Konings, N.H., Wang, A.H.J. and Terwilliger, T.C. (1994) Structure of the gene V

protein of bacteriophage f1 determined by multi wavelength X-ray diffraction on the selenomethionyl protein. Proc. Natl. Acad. Sci. 19, 2071.