COMMITTEE I
Unifying Principles in the Sciences

DRAFT - 8/15/86
For Conference Distribution Only

SYMMETRY IN BIOMOLECULES: ITS PHYSIOLOGICAL MEANING

by

Guido Pincheira V.
Professor of Genetics
Faculty of Sciences
University of Chile
Santiago, CHILE

The Fifteenth International Conference on the Unity of the Sciences
Washington, D.C. November 27-30, 1986

INTRODUCTION

The fundamental aspiration of biologists to achieve a unified understanding of living matter is basically confronted with the problem of deciding what structural and dynamical details of biomolecules are truly primary.

Although the discover of the chemical nature of the genetic material, followed by partial explanations of its coding functions, has been a very important advance, it still is a very challenging task to relate levels of macroscopic description of biological phenomena with the microscopic physical and chemical details of the work of biomolecules.

The large majority of biomolecules are made of variable numbers of a few chemical groups, such as OH, NH_2, CH_3, etc., being really surprising the high specificity with which these molecules tend to aggregate and organize themselves. The binding specificity of these chemical groups is, of course, related to energy levels that atoms and molecules have associated with their structure.

The study of the structural characteristics of living matter has demonstrated the special importance of the notion of symmetry to explain energy states associated with the interactions and transitions that take place between molecules in living processes.

Symmetry is a very general phenomenon in nature. It exists in many different ways at the different levels of organization of matter and may be defined as the "harmony or balance in the proportions of parts to the whole" (2). As such, symmetry is associated with the correspondence of elements on

opposite sides of a line, a point or a plane which may be called an axis, a center or a plane of symmetry, respectively.

In Biology, symmetry is exhibited in many different ways by the different levels of organization of living matter; within the molecular structure, in the organization of molecules to form cellular structures, in the distribution of the different organs of an organism, etc. As such, this characteristic is useful to classify organisms in different groups:

a) The bilateral symmetric type (mammals, birds, fishes), in which the body exhibits a single plane of symmetry. The right half of the human body is related to the left half as if the right were obtained by reflection of the left across a mirror plane, which can be imagined dividing right from left. The human hands are another example of symmetry in the human body.
b) The radially symmetric type, with an infinite number of planes and rotational axes of summetry, is exhibited by sea urchins, jelly fishes, etc.
c) The serially symmetric type, in which elements are repeated at regular intervals, is found in earthworms.
d) The asymmetric type, where no symmetry exists at the organismic level, is exemplified by the *Paramecium*.

This very rich variability of biological symmetry may be considered in a relative simpler and unifying scheme by studying it in spatial arrangement of atoms in biomolecules and its significance evaluated in physiological processes in which these biomolecules participate.

A molecule contains a symmetry element if by carrying out certain geometric operations on its original configuration, the molecule can be transformed into another configuration that is superimposable on the

original, although its orientation may be changed.

The existence of different elements of molecular symmetry (planes, centers, axes), is important to consider the possibility of different types of symmetry operations (rotations, reflections, inversions, etc.) involved in the behavior of organic molecules.

A symmetry operation means to move a structure in such a way that, after the operation, it is possible to get points that coincide in position with those that the object had in its original position. As such one refers to Planes of symmetry, Axes of symmetry, and Center of symmetry.

The majority of small molecules have one or several planes of symmetry. Others, such as linear molecules, have an infinite number of planes of symmetry, because any plane passing through the molecular axis will be a plane of symmetry.

The water molecule, so important in life, has a two fold axis of symmetry and two planes of symmetry (Fig. 1). One of them is coincident with the plane of the molecule. The other includes an Oxigen (O) atom and is perpendicular to the molecular plane. The effect of reflection through this second plane maintains fixed the O atom, but exchanges the H atoms.

The molecule of ammonia (NH_3) has a three fold axis of symmetry and three vertical planes of symmetry (Fig. 2). The NH_3 molecule is an example of a molecule with a pyramidal form, with all planes of symmetry including the N atom and one of the three H atoms.

The molecules that are not able to superimpose upon its mirror image are called dysymmetric. They may have a certain degree of symmetry. In fact, the concept of symmetry implies qualitative and quantitative aspects.

In Biology the qualitative aspect predominates as may be illustrated by a tree, whose branches may differ to a certain extent in size and distribution on one side as compared to the other, but at the moment of considering the tree in two halves, it is possible to see that it is a symmetrical, equilibrated and harmonius structure, but not absolutely perfect.

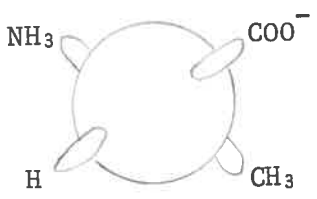
In qualitative symmetry it is possible to have equilibrium of asymmetric components. Then, we can see that qualitative symmetry also includes the concept of asymmetry, which, as we shall see, is very much related to living processes.

Asymmetry

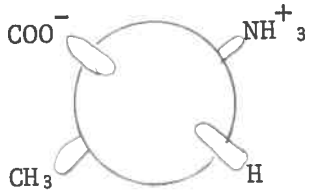
In a molecule, asymmetry may arise from the occurrence of an atom asymmetrically located, as may be the case of a Carbon (C) atom having four different substituents on its four valences. In such condition, the C atom becomes an asymmetric center, originating two different distributions of substituents.

As a consequence of these different substitutions, we have molecules with the same number and kinds of atoms, same molecular weights and formula, but differing in their 3-dimension configuration. They are called stereoisomers. Examples of this situation are the monosaccharides glucose, galactose, mannose; molecules that differ only in the orientation of their hydroxyl groups.

Two isomers that are mirror images of each other have the same chemistry and therefore are given the same name; distinguished by the prefix D (dextrogyre) or L (levogyre). Ex.:



L-alanine



D-alanine

Looking along the H-C bond, the methyl, amino and carboxyl groups appear in a clockwise direction in L-alanine, and in a counterclockwise direction in D-alanine.

Some of these isomers are called optical isomers, because they have optical activity detected by the ability of the compound to rotate polarized light when exposed to it in a crystal form. This optical activity will be absent if the molecule exhibits any plane of symmetry.

Molecular structure may be studied by X-Ray diffraction techniques that try to locate all the atoms in molecules regularly arranged in a three dimentional matrix of a crystal.

Single crystal analysis allows the determination of structural details of a molecule, revealing aspects of symmetry that may be the clue to explain biological processes in physîco-chemical terms.

The overall picture gives insight into molecular self-organization and intermolecular interactions, such as hydrogen bonding; stacking of heterocyles, and dipolar, van der Waals and electrostatic forces that are the essence of most biological phenomena.

BIOLOGICAL MACROMOLECULES

Some of the main biological macromolecules, such as proteins, nucleic acids, polysaccharides, are built of asymmetric molecules (α-aminoacids, nitrogen bases, sugars). The polymerization of these monomers generally gives rise to macromolecules with a linear skeleton in which symmetric elements and symmetry operations may emerge in association with energy changes taking place in living phenomena.

In the case of aminoacids, all of them, except glycine, have an asymmetric α-carbon, hence they are called α-aminoacids. α-aminoacids found in nature are predominantly of L configuration. Sugars, such as glucose or ribose, have a D-configuration.

A mixture of D and L forms of molecules is called a racemic mixture, and is optically inactive.

This situation generates questions about how in the course of chemical evolution, molecules optically active could have been segregated abiotically from a racemic mixture, biologically inert and, later, how in

biological evolution of matter, it was possible the formation of so many different living forms with the same type of basic molecular constituents. In this respect, a study of the different steps in the evolution of living matter must not be dissociated from the study of evolution of optical activity.

AMINOACIDS

They may exist in D and L forms. Although the L-form is the most predominant in nature, the existence of D-aminoacids in peptides present in bacterial membranes, antibiotics (gramicidin), malignant cells, etc., originated many speculations about the mechanisms that could have determined such predominance.

D and L aminoacids have identical covalent bonds, but their properties to link themselves to other molecules, such as enzymes, are very different. In this respect, the large majority of known enzymes are specific to L-aminoacids, and they do not recognize D-aminoacids. It has been suggested that the predominance of L-aminoacids in the majority of living forms could have occurred by change, as a result of the action of symmetrical forces (chemical and physical) participating in a series of reactions. Among the mechanisms facilitating this selective process, it is possible to imagine the intervention of enzymes favored in its action by characteristics of molecular symmetry in the enzyme-substrate recognition process.

PROTEINS

Aminoacids form polymers with a linear skeleton, in which a specific group (-CO-NH-) is repeated along the molecule. A chain of aminoacids joined in this manner is a polypeptide.

An helical structure is the natural conformation for polymers of this type, because it places each monomer in identical orientation within the molecule, with the possibility to form secondary linkages (hydrogen bonds) between groups of the same polypeptide. These linkages determine the winding of the polypeptide skeleton to form an helical structure.

The α-helix is the most important regular configuration in the polypeptidic chain. Nevertheless, the presence of irregular lateral groups of aminoacids connected to the regular central skeleton determines that the protein structure is, usually, irregular. Practically there are no proteins with a simple helical structure, rather, proteins have helical regions and non-helical regions. One reason is that one of the aminoacids, Proline, has no amino group, and when present, H bonding is interrupted. Another reason is the formation of -S-S linkages between Cysteine residues, producing distorsions of the helical structure. A third factor that determines irregularities of proteins is the diverse chemical nature of aminoacids lateral groups, which will tend to form the most favorable secondary interactions from the energy point of view.

Some proteins may have no helical structure at all. Proteins such as silk, and specially those rich in glycine and alanine, exhibit a secondary laminar structure, called β-sheet, formed by fully extended polypeptidic

chains, held together by H bonds between groups of the different chains. A protein β-sheet may exhibit a parallel or an antiparallel disposition. The last possibility happens when an extended polypeptide folds back and forth upon itself, satisfying the requirements of dyad symmetry.

In addition, the possibility of proteins to fold up gives rise to tertiary structures and increases the number of possible shapes in proteins, even in those of small size. Among all these possibilities, in biological conditions the protein adopts only one of these conformations, usually the most energetically convenient.

Although it is possible to figure or to obtain beautiful pictures of three dimensional structures of proteins, illustrated by the positions of all atoms, we still do not clearly understand how proteins fold up.

Besides, it still is a general belief that these structures have a high degree of rigidity, instead of highly dynamic structures, involved in cooperative phenomena between similar or different polypeptide units in which symmetry may emerge and play an important physiological role.

NUCLEOTIDES

A nucleotide is made of three major molecular components: a nitrogen base (adenine, guanine, cytosine, thymine, uracil) connected by a β-glycosidic linkage with a sugar (ribose or deoxyribose) in a cyclic furanoside form to produce a nucleoside. If the 3'OH or 5'OH groups of the sugar of a nucleoside is phosphorilated, we have a nucleotide. Nucleotides may be polymerized to form the nucleic acids.

Nucleosides, besides of being constituents of the genetic material, may have several physiological roles in living organisms. Let us consider the case of adenosine (4):

a) As di- or triphosphate adenosine (ADP or ATP) is a source of energy for many enzymatic processes.

b) As 3', 5'-cyclic phosphate adenosine (cAMP) is a messenger molecule, participating in the activity of peptide hormones to regulate genetic activity.

c) As a constituent of puromycin, adenosine inhibits the biosynthesis of proteins.

d) As an arabino or 8-azaderivative, adenosine is an antibiotic.

A high conformational flexibility is possible in a nucleotide and to understand its very diverse biological function it is essential to know its structural characteristics specially its electronic and three-dimensional structure.

NUCLEIC ACIDS

DNA

a) Structure

The Watson-Crick model of DNA, known as the double helix, propose two helical chains coiled round the some axis. These chains (excluding the bases) are related by a dyad perpendicular to the helix axis, thus the two chains are antiparallel. The structural studies of natural DNA generally means to deal with a heterogenous population of molecules, being difficult to obtain them in the form of single crystals, suitable for X-Ray analysis. Besides,

X-Ray diffraction data with a resolution of 3 Å defy any direct determination of the structure of such a large molecule. Nevertheless the possibility to synthesize short oligonucleotides has made possible to produce single crystals, able to diffract polarized light with a high resolution and allowing experimental measurements to observe fine details of DNA conformation. Synthetic oligonucleotides of a defined sequence may be crystallized alone or forming complexes with metal ions, drugs or proteins bound with them. X-Ray crystallography can yield precise structure for molecules of weight up to 2000 daltons. This correspond to hexanucleotides (350 daltons = average MW/nucleotide). A few DNA oligomers have been atomically resolved, and they exhibit a perfectly regular structure with asymmetric units consisting of one nucleotide (3). In addition to the high conformational flexibility of one nucleotide, it has been possible to detect the same characteristic at the junction of two neighbor nucleotides in a polymer.

Symmetry elements are also found in these synthetic polynucleotides. They are related to the glycosil bonds linking the nitrogen bases with sugars and to the attached sugar phosphate backbones. At least, two cyclic hydrogen bonds must form to produce a stable nitrogen base pair, which may display two kinds of two-fold symmetry, depending on the orientations of the -glycosyl C -N bonds. Two-fold rotation axes can be arranged either perpendicular to or within base-pair planes. This has an influence upon the orientation of the attached sugar-phosphate groups or backbones of DNA (4).

The crystal structure of a B-DNA dodecamer has been studied at high resolution. Nevertheless, it is too short and irregular to precisely estimate the pitch, diameter and offset of a large DNA molecule, made of many

thousands of nucleotides.

In the purine-pyrimidine pairs of the Watson-Crick model, the operation of the dyad is restricted to the glycosyl-C_1 -N linkage and not applicable to the bases. It is therefore called a pseudo-dyad.

The symmetry of the various DNA double helices is represented by indexed numbers (N_m).

N = the number of nucleotides after which exact repeats along the helix axis is achieved.

m = indicates the number of helical turns of the molecule.

For native B-DNA, on 92% relative humidity, helix symmetry $N_m = 10_1$.

For A-DNA, on 75% relative humidity, $N_m = 11_1$.

b) Organization

Specially important for certain aspects of DNA physiology is the existence of inverted sequences of nucleotides, called palindromes, in which a sequence of nucleotides read in the 5' → 3' direction of one DNA strand is identical to a sequence of nucleotides read from the other end of the palindrome on the other strand in the 5' → 3' direction. Ex:

5' C A G C T G 3'
3' G T C G A C 5'
axis of symmetry

A palindrome is a DNA region of dyad symmetry, in which the axis of symmetry identifies the central point about which the sequence of nucleotides is the same on either side.

As we shall see later on, palindromes play a very important rol in the process of recognition of DNA by enzymes called nucleases

tRNAs (transfer RNA)

Perhaps, the best and most thoroughly studied nucleic acids are transfer RNAs (tRNAs). These molecules, with a molecular weight around 26.000 daltons, can be crystallized from aqueous medium by addition of alcohol or salts. These tRNA crystals contain 30-80% solvent and may be considered a concentrated solution. Nevertheless, again, due to the limited resolution of X-Ray data, atoms cannot be resolved in an electron density map of tRNA, specially in non-helical regions of the molecule. As such, the same molecule under the same crystallographic conditions may be subject of different interpretations by different research groups.

tRNAs are polynucleotide chains made of 75 - 90 units. Their nucleotide composition include the standard nucleotides and other special nucleotides, such as dihydrouridine, ribothymidine, etc. The primary sequences of nucleotides of tRNA give rise to secondary structures that ressemble a cloverleaf form, with four double-helical stems and three loop regions (see Fig. 3).

Each one of the four stem regions of this cloverleaf contains 4-7 base pairs organized in double helices. They are called the acceptor (it accepts the aminoacid that will be transported to the site of protein synthesis), the anticodon, the D (for dihydrouridine) and the T (or ribothymidine) stems.

The subsequent folding of the cloverleaf into a tertiary structure confers an L shape to the tRNA molecule (Fig. 4). The L shape is composed of two nearly perpendicular RNA helices, each one having a length of 70 Å and a thickness of 20 Å. The two ends of the L configurations are: the acceptor

CCA triplet that will bind the aminoacid, and the anticodon loop that will recognize the messenger RNA codon in the process of protein synthesis.

The two arms of the L tRNA tertiary configuration may be separated by a pseudo-dyad axis of symmetry that bisects the L angle and relates the anticodon/D helix to the acceptor/T helix.

This aspect of tRNA symmetry seems to be very important for the interaction with the aminoacyl-tRNA synthetases, the enzymes carrying the aminoacids to the acceptor loop. In this respect, the substrate recognition may be facilitated and, perhaps, it may give rise to an enzyme-tRNA complex with a two fold symmetrical structure.

PROTEIN-NUCLEIC ACID INTERACTIONS

The interaction between nucleic acids, specially DNA, and proteins is one of the central processes in all living cells. The interactions occur at all levels of nucleic acid physiology (replication, transcription, repair, etc.). Unfortunately, actual knowledge about these processes is still very limited.

Some proteins interact with single stranded nucleic acids: others with double stranded nucleic acids. A third type of protein interacts with globular structures of nucleic acids.

It is also possible to distinguish interactions based in the protein specific recognition of particular sequences of nucleotides from interactions based in a general recognition that does not discriminate special sequences of nucleotides. In the last case, the protein must primarily recognize structural details of the sugar-phosphate backbone of the nucleic acid. It is

possible that a certain complementation between the backbone of the nucleic acid and the protein exists, facilitating the possible interactions. One type of compatibility is the distance of 7 Å between adjacent phosphate groups of the double strand DNA backbone, when measured on the outer perimeter of the phosphate groups, and an equivalent distance between two positively charged groups of arginine or lysine in a protein α-helix, when separated by three other aminoacids.

A similar dimensional correspondence exists between double stranded DNA and protein antiparallel β-sheet or fully extended polypeptides.

This intermolecular compatibility affects elements of rotational symmetry of these molecules. DNA, in A or B form, has two types of pseudo two-fold axes per base pair, one on the plane of each base pair and one in between two adjacent base pairs. A protein antiparallel β-sheet also exhibits the two types of pseudo two-fold axes (5).

So, we have structural elements in double stranded nucleic acids and proteins exhibiting similar dimensions and symmetries.

These compatibilities facilitate the formation of hydrogen bonding between the backbone of a peptide and the backbone of the nucleic acid and several models have been proposed to explain a general type of double stranded DNA-protein recognition.

The outer DNA double helix is not cylindrically smooth, but displays two grooves of different width and depth (Fig. 5). They are called the major and the minor grooves.

Portions of each base pair as constituent of a DNA molecule are exposed in two separate grooves. Each of the four possible base pair

arrangements (A-T; T-A; G-C or C-G) may be recognized by the different arrangements of the atoms that protrude, allowing a unique hydrogen bonding pattern for each of four base pairs.

When one examines the contact region of a major groove of DNA with the α-helix of a protein, it is possible to find structural alignments in which certain C atoms of the α-helix are roughly coplanar with a nitrogen base pair plane. This situation may facilitate the formation of H bonds according a sort of "recognition code" between aminoacids and nucleotides.

In addition, steric recognition seems to play a very important role. In this respect, the most important group would be the CH_3 group of thymidine, because it has the largest accesible area of any group on the major groove.

A similar situation takes place in the interaction of proteins with the minor groove of B-DNA.

Examples of DNA-protein interaction.

1) The action of nucleases, enzymes that cut nucleic acids.

Some endonucleases recognize and cleave DNA in regions characterized by the presence of symmetrical sequences of nucleotides, particularly tetra or hexanucleotides. As indicated previously, these sequences are called palindromes and exhibit a two-fold axis of symmetry and two-fold rotational symmetry. For example, the enzyme Eco Ri recognizes the following DNA palindromic sequence: 5' -GAATTC-3'

3' -CTTAAG-5'

Owing to the two-fold symmetry of the palindrome, the respective nucleases may at least be dimeric in structure. When these palindromes bind proteins

that contain identical subunits, a better interaction may take place on the basis of coincidence of the axes of symmetry of the protein and symmetry elements of nucleic acid structure. This type of symmetric interaction may be expected to add a favorable entropic contribution to the total free energy involved in the process of DNA-protein recognition and cooperation. In this respect, it is important to mention that it is not necessary that the bound DNA sequences by exactly symmetric. The reason is that the atomic interactions involved in the protein-nucleic acid interaction provide a margin of flexibility. For example, the N atom in position 7 is common to both purines in the large groove and the O atom in position 2 is common to both pyrimidines in the small groove. If those atoms are the only contact for a given base pair, then approximately symmetric sequences may be sufficient for the recognition of action of the nuclease. Nevertheless, the fact that the rates of activity for some endonucleases in sites in different locations of DNA vary so much, suggests that sequences around the recognition sites are also involved in interactions with the enzyme.

2) DNA recognition by regulatory proteins. Regulatory proteins may start or shut off gene expresion.

Recently two DNA binding proteins, cro repressor of bacteriophage lambda (λ) and *Escherichia coli* catabolite gene activator protein (CAP) have been studied intensively (1).

a) Interaction of B-DNA with α-helices and β-pleated sheets of cro repressor.

The cro protein is the product of cro gene of bacteriophage lambda. It is

a polypeptide, 66 aminoacids-long, that recognizes at least 6 near-palindromic sequences in double stranded DNA of phage lambda, binds to them and prevents expression of several genes. The structure of cro peptide consists of three α-helices and three antiparallel β-sheets.

In a cryatal, cro protein is a tetramer with near 222 point groups symmetry. When forming complex with DNA, cro protein probably binds as a dimer or a tetramer. The dyad axis of cro as a dimer coincides with the intrinsic dyad symmetry of the near-palindromic region of lambda B-DNA to which cro repressor binds facilitating the formation of hdyrogen bonds between certain aminoacids and the nucleotides

b) Interaction of B-DNA with catabolite gene activator protein (CAP).

Catabolite gene activator protein (CAP) activates several genes in a very specific way. To do that, it requires,first, to be activated by cyclic adenosine monophosphate (cAMP), which acts as an allosteric effector.

When the intracellular levels of cAMP increase, CAP activates the expression of several genes in bacteria, such as those coding for enzymes for the catabolism of lactose, arabinose, maltose and other sugars.

The activation of gene expression in bacteria by CAP must include the binding of cyclic AMP to CAP and the interaction of this complex with segments, called promotors, of the target genes.

The cAMP-CAP complex makes its major contacts in two succesive major grooves of DNA, but the entire area of interaction includes the minor groove situated between the major grooves and one minor groove at each end of the DNA region (3).

The structure of the catabolite gene activator protein (CAP).

CAP is composed of two identical subunits, 210 aminoacids each (1). The two subunits are related by a dyad axis of symmetry (see Fig. 6).

CAP has two domains: a carboxyl-terminal, or DNA binding domain and a larger amino-terminal domain, responsible for: the interaction with cAMP; the interaction between the subunits to form a dimer and, probably the regulation of the binding with DNA.

This two domains structure was clearly shown by the 2.9 Å resolution crystal structure of CAP complexed with cAMP (1).

The carboxyl-terminal domain of each CAP subunit consist of three α-helices, designated as D, E and F, connected by short β-sheet structures. In each subunit, the F α-helices are thought to provide the major interaction sites with the target DNA.

DNA structure for binding with CAP.

The segment of DNA for binding CAP has an extension of 25 nucleotide pairs. In this stretch of DNA we have:

A nucleotide sequence 5' TGTGA 3', which is critical for CAP binding,

A block of 6 nucleotides that shows little sequence preferences,

A second sequence that contains an inverted repeat of the TGTGA motif

In some cases, the sequence of 6 nucleotides generates a two-fold symmetry in the DNA sequences. This symmetry is functional in the case of the gene lac, as has been proved by mutations that prevent the binding CAP (1).

Experimental results indicate that, when CAP is bound at the lac gene, the symmetrically arranged sequences of nucleotides must recognize approximately the same protein configuration. In other cases, the described symmetry is not observed and it may be replaced by a symmetrically arranged second version of 5' TGTGA 3', or another type of sequence.

The affinity of CAP for DNA appears to be higher when these sequences of nucleotides are symmetrically arranged than when they are not symmetrical. The two peptides in the CAP dimer are in near-parallel arrangement and related by a pseudo-dyad axis running between them. An adequate interaction for CAP dimer with DNA may depend on the coincidence of the dyad axis of the protein with the dyad axis of the DNA palindromic sequence.

The binding of cAMP-CAP complex to a fragment containing a lac promoter is favored by 4 kcal/mole over the binding to fragments that do not contain a binding site. The binding of CAP to the gene mal T is less stable by roughly 0.6 kcal/mole and 1.4 kcal/mole when bound to the gal site in comparison with the binding to the lac site.

This hierarchy of CAP binding affinities to different DNA regions may explain why different cAMP sensitive points are activated by different concentrations of cyclic AMP *in vivo*.

EVOLUTIONARY ASPECTS

X-Ray diffraction comparative studies of the molecular structures of cro , cI repressor of phage lambda, and CAP of the bacterium *Escherichia coli* reveal a striking structural homology in the relative position and orientation of two consecutive α-helices (E and F) in the carboxil terminal domain of

CAP and two α-helices (II and III) in the amino terminal part of cro protein. The path of 24 α-Carbon atoms in the structural unit that contains these two helices in CAP can be superimposed on the path of the α-Carbons of the two homologous helices in cro. A similar structural homology is found between 2 helices of cro and two similar helices of lambda cI repressor.

Thus, a conserved helix-turn-helix motif, essential for interactions with DNA, is found in each of the three proteins.

Cro, cI repressor and CAP interact with DNA in a basically similar manner. Each of these proteins use symmetrically related subunits to interact with two-fold related nucleotide sequences in the regions that operate the corresponding genes. These DNA operator sequences have approximate two-fold symmetries and the complexes that they may form with the regulatory proteins may also be symmetric.

Although the basic design of DNA interaction unit appears to be conserved, most of these different proteins recognize different sites in DNA. This situation suggests that the specificity in DNA recognition must also be provided by the aminoacid side chains in the motif, and possibly, by other adjacent residues in the proteins.

X-Ray diffraction data of cro and cI proteins allow the possibility to suggest models for the interaction with DNA. Lambda cI repressor recognizes the same sites that cro proteins binds, but with a difference in affinity.

In the case of cro protein, a model proposes that one α-helix interacts with a major groove in DNA, and one β-sheet contacts a minor groove.

Since the active form of cro-repressor is a dimer, it must interact with DNA symmetrical sequences in two succesive major grooves along one face of the DNA double helix. Adjacent α-helices of cro-repressor make contact the DNA backbone and may help to orient the recognition process. Nevertheless these models and suggestions need to be tested with additional crystallographic, biochemical and genetic studies.

Conclusions

Biomolecules display a wide variety of symmetries at the different levels of their organization; an aspect that is projected in the process of morphogenesis of unicellular and multicellular structures of living organisms.

These symmetrical characteristics of biomolecules may be interpreted in terms of the relationships between their structure, forces and energy fields displayed in basic living phenomena. As such, we may consider the existence of a very dynamic symmetry; alternating its existence with the possibility of broken symmetry or asymmetry in the process of biosynthesis of the different biomolecules. In this respect symmetries may be related to the better performance or organization of molecules to fullfil requirements for more efficient living processes.

At the genetic level, it seems evident that aspects of interactions between nucleic acids and proteins are related to the matching of their symmetry properties.

Coincidence of axes of symmetry of proteins and symmetry elements of the nucleic acid may add a favorable entropic contribution to the total free

energy involved in the process of recognition and interaction between these two basic biomolecules. This situation may be reflected in an easier formation of hydrogen bonds between biomolecules in the processes of replication, transcription and regulation of the gene activity. This is an area of research that is expanding very rapidly and where physics and chemistry are expected to contribute decisively to the understanding of biological phenomena.

References

1) Crombrugghe, Benoit de; Stephen Busby & Henri Buc. 1984. Cyclic AMP receptor protein: Role in transcription activation. Science 224: 831 - 837.

2) Ferraro, J.R. and J.S. Ziomek. Introductory group theory and its application to molecular structure. Plenum Press. New York. 1969.

3) Mizobuchi, Kiyoshi; Itaru Watanabe and James D. Watson (eds.). Nucleic Acid Research Future Development. Academic Press 1983.

4) Saenger, Wolfram. Principles of nucleic acid structure. Springer - Verlag. New York, Berlin, Heidelberg, Tokyo. 1984.

5) Sung-Hou Kim. Structural modes for DNA-Protein recognition. in Nucleic Acid Research Future Development. Ed. Mizobuchi, K. et al. Academic Press. 1983.

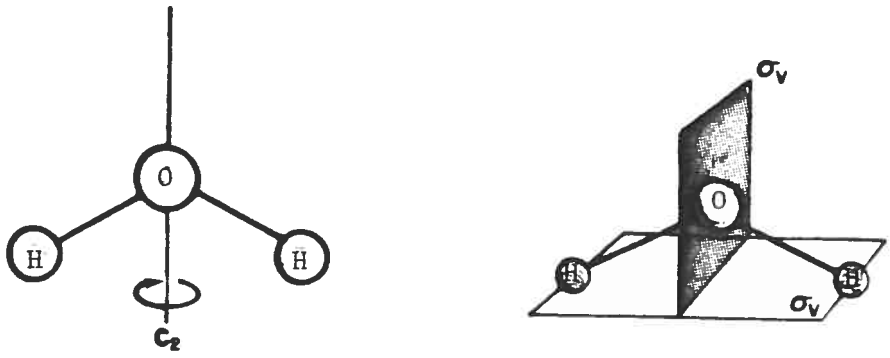


FIG. 1. Symmetry elements in the H_2O molecule.

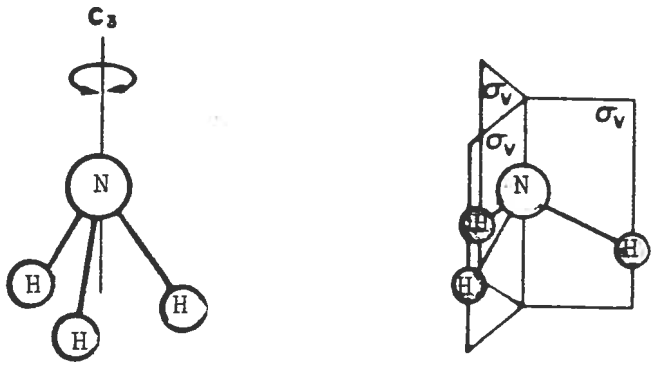


FIG. 2. Symmetry elements in the NH_3 molecule.

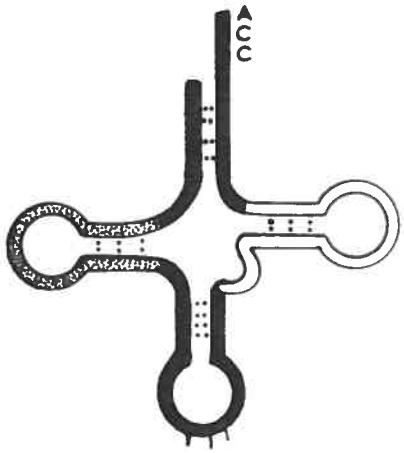


FIG. 3. Scheme of tRNA secondary structure.

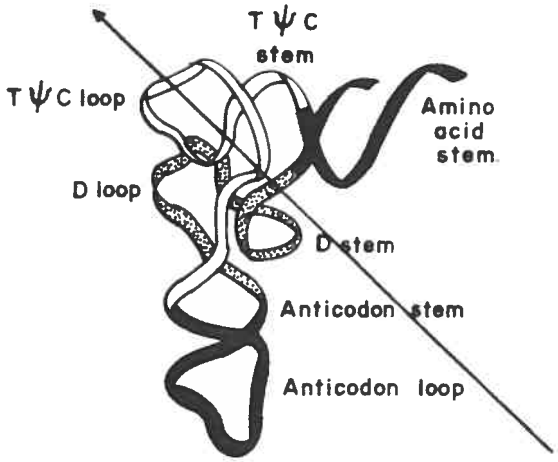


FIG. 4. Scheme of tRNA tertiary structure. The arrow represents a pseudo-dyad axis of symmetry.

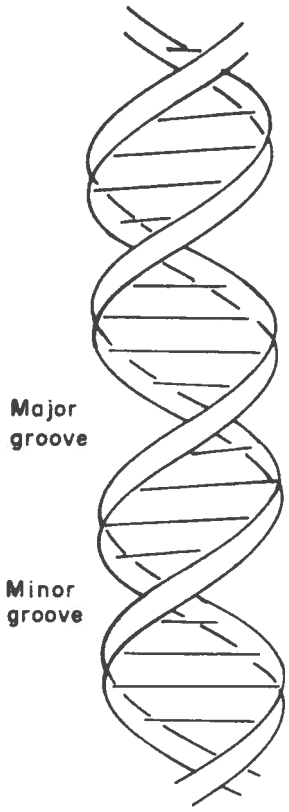


FIG. 5. Scheme of B-DNA
Double helix indicating the positions
of major and minor grooves.

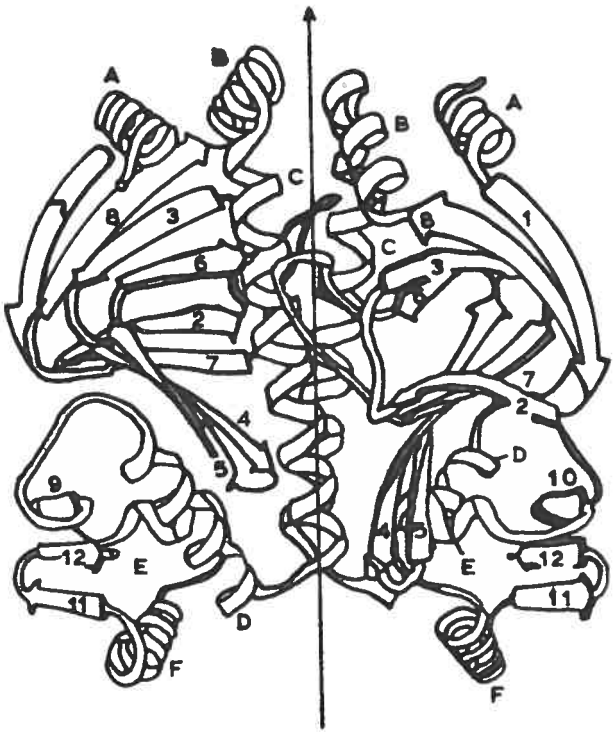


FIG. 6. Scheme of CAP as a dimer.
The drawing shows the existence of α-helices (designated by letters A-F) and β-sheets (designated by numbers1-12). The arrow represents a dyad axis of symmetry relating the two identical subunits of the dimer.