CHEMICAL EVOLUTION: AN APPROACH INVOLVING MOLECULAR INTERACTIONS

M. VIJAYAN and C. G. SURESH
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India.

ABSTRACT

The processes by which the first self-replicating systems arose through different stages from a milieu of racemic mixtures of biomolecules, obtained through prebiotic organic synthesis, constitute the least understood link in the evolutionary history of the biosphere. An approach primarily involving molecular interactions observed in crystals, towards understanding these processes is discussed. Amino acids and peptides have a strong intrinsic propensity to arrange themselves in such a way as to bring the main chain amino and carboxylate groups into periodic hydrogen bonded proximity. The geometrical condition for the non-enzymatic condensation is satisfied in these arrangements. It appears that chiral separation could be achieved in favourable cases through molecular interactions. The primitive molecular assemblies which gave rise to early self-replicating systems could have been similar in nature to the self-assembly systems, generated by non-covalent interactions, which occur at the lowest levels in the organisational hierarchy of present day organisms. Thus, the study of biologically significant non-covalent interactions, including specific ones, is important for understanding not only the present biosystems but also the primitive biomolecular assemblies. In addition to their probable relevance to prebiotic phenomena, the investigations discussed here have led to results which are of intrinsic interest in relation to biomolecular aggregation and interactions.

INTRODUCTION

CHEMICAL evolution and biological evolution constitute two successive phases in the history of the biosphere. Chemical evolution in turn is believed to have consisted of broadly three overlapping stages, namely, the origin of the monomers of biopolymers and other small molecules, the condensation of monomers into meaningful sequence-specific polymers, and the organisation of biopolymers and small biomolecules into self-replicating systems, perhaps the primitive forms of the cell. Chiral selection must also have taken place during this process. The celebrated Oparin-Haldane hypothesis, which has been supported by several laboratory experiments under simulated prebiotic conditions, provides a plausible description of the first stage of chemical evolution, namely, the abiotic synthesis and the accumulation of monomers of biopolymers and other molecules in primordial oceans. Subsequently, several meteorites (carbonaceous chondrites) were found to contain amino acids, nucleotide bases and other organic materials, in proportions comparable to those found in laboratory experiments under simulated prebiotic conditions. Thus, it would appear that abiotic organic synthesis was not exclusively a terrestrial phenomenon. Furthermore, spectroscopic and other evidences for the presence of abiotic organic molecules in interstellar space points to the universality of such a synthesis.

As at present, no universally accepted explanation for the subsequent stages of chemical evolution exists. Indeed, the processes by which the first self-replicating systems arose through different stages from a milieu of racemic mixtures of biomolecules, including monomers of biopolymers, constitute the least understood link in the evolutionary history of the biosphere. The present article is primarily concerned with an
approach involving molecular aggregation, developed in our laboratory, towards understanding these processes. This approach essentially arose out of studies in this laboratory on the possible geometrical features of biomolecular interactions, through the preparation and x-ray analysis of crystalline complexes involving amino acids and peptides, among themselves as well as with other biomolecules\(^7\).\(^8\).

**MONOMERS TO POLYMERS: PROBABLE ROLE OF MOLECULAR ALIGNMENT**

In the present-day organisms, proteins are synthesised employing a complex machinery using the information encoded in nucleic acid molecules. Nucleic acid molecules are synthesised using information contained in other nucleic acid molecules with the help of biological catalysts, the enzymes, which are proteins. Thus nucleic acids are necessary to synthesise proteins and proteins are required to synthesise nucleic acids. Which came first? This is the well-known chicken and egg problem in evolution. It would appear that primitive proteins or polypeptides on the one hand and nucleic acids on the other, could have originated independently. A primitive simple machinery for nucleic acid directed protein synthesis could have evolved at a later stage through the interactions of those two types of molecules. The probable nature of this primitive machinery is currently a topic of intense research\(^9\)–\(^11\). However, the problem still remains as to how the two types of fundamentally important biopolymers originated independently in the first place.

In attempts to simulate the possible independent (of nucleic acids) non-enzymatic synthesis of polypeptides in prebiotic conditions, polymerisation of amino acids has been carried out with the aid of condensing agents which might have existed in the prebiotic milieu\(^12\). Polymerisation could also take place on clay particles when monomers are adsorbed on them\(^13\). It has been demonstrated that polypeptides could be formed by heating amino acids in comparatively dry conditions\(^14\). The thermal polypeptides so formed have been shown to have non-random sequences of amino acids, pointing to the possibility of "self-ordering" of amino acids during condensation without the aid of nucleic acids\(^15\).\(^16\).

The experiments outlined above indicate that amino acids could independently polymerise, possibly with non-random sequences, non-enzymatically. However, the mechanisms of such polymerisation and the conditions in these experiments that favoured it, have not been clearly elucidated. *A priori* it would appear that a condition necessary for the non-enzymatic condensation of monomers is the proximity and the favourable juxtaposition of the reacting groups. It is in this context that some of the observations made by us on the aggregation of amino acids and peptides become interesting in relation to chemical evolution.

*Head-to-tail sequences of amino acids*

An interesting common feature observed in most of the crystalline complexes, x-ray analysed in this laboratory\(^7\).\(^8\) is the aggregation of amino acid molecules in ‘head-to-tail’ sequences of the type,

\[
\text{---NH}_3^+\text{CHR-COO}^--\text{---NH}_3^+\text{CHR-COO}^--\text{---NH}_3^+\text{CHR-COO}^--
\]

in which the \(\alpha\)-amino and the \(\alpha\)-carboxylic groups are brought into periodic hydrogen bonded proximity in a peptide-like arrangement. It was realised that the condition, referred earlier, necessary for the non-enzymatic condensation of monomers, is satisfied in these sequences. Hence, it was suggested that such sequences could well have been made use of prebiotic polymerisation\(^17\). While pursuing the hypothesis, it was important to enquire if head-to-tail sequences were indeed an intrinsic feature of amino acid aggregation. It was also interest to investigate the geometrical features of these sequences, which could conceivably affect condensation reactions, and the influence of chemical nature of the side chains in their systematic study of the occurrence and geometrical features of head-to-tail sequences.
the crystal structures of naturally occurring common amino acids, and their racemates and complexes was therefore carried out\textsuperscript{18}.

The above study showed that head-to-tail sequences are indeed an almost universal feature of amino acid aggregation in the solid state. These sequences largely belong to two broad categories in terms of the geometrical arrangement of amino acid molecules in them. As illustrated schematically in figure 1, the sequences in the first category consist of straight chains of molecules related mostly by the shortest cell translation in the crystals while those in the second category form hydrogen-bonded two-fold helices centred around crystallographic $2_1$ screw axes. A detailed examination of the geometrical features of the hydrogen bonds and the orientation of the amino acid molecules, lead to further subdivisions in the two main categories. The shortest cell translation in most crystals containing amino acids has a value in the neighbourhood of 5.3 Å and corresponds to the periodicity of a straight head-to-tail sequence or, less frequently, that of a helical sequence or both. In fact, the crystal structures of amino acids and their complexes can be classified in terms of the occurrence and the geometrical disposition of different types of head-to-tail sequences in them. Such a classification appears to have some chemical significance. For example, two types of straight sequences coexist, as shown in figure 2, in the crystals of most of the hydrophobic L-amino acids whereas the crystals of hydrophilic L-amino acids tend to be made up of a straight sequence and a helical sequence (figure 3). Thus the aggregation of, and hence approach between, amino-acids is affected by the nature of the side chain. This perhaps provides a handle for approaching sequence selection.

The near ubiquitous occurrence of head-to-tail sequences of amino acids cannot be explained exclusively on the basis of electrostatic

---

**Figure 1.** Schematic illustration of (a) straight and (b) two-fold helical head-to-tail sequences. Broken lines indicate hydrogen bonds in this and the subsequent figures.

**Figure 2.** Mutual disposition of two straight sequences observed in the crystals of most hydrophobic L-amino acids.
interactions between the z-amino and the z-carboxylate groups. First, the formation of infinitely long sequences containing alternating positive and negative charges is only one among the several possible geometrical patterns capable of ensuring microscopic charge compensation in condensed systems containing an equal number of negative and positive charges. Secondly, and perhaps more importantly, unlike what one would normally expect, the most basic group does not necessarily interact with the most acidic group in crystals involving amino acids. This particular aspect was specifically explored through the x-ray analysis of the crystals of arginine acetate\textsuperscript{19} and lysine acetate\textsuperscript{20}. The side-chain guanidyl group in arginine and the side chain amino group in lysine are considerably more basic than the corresponding z-amino groups. Likewise, in both the cases, the z-carboxylate group is considerably more acidic than the acetate group. Therefore, one would normally expect the z-carboxylate group to interact with the side chain guanidyl or amino group of a neighbouring molecule and the acetate ion with the z-amino group, on the basis of simple electrostatic considerations. In the crystal structures, however, the acetate ion interacts primarily with the side chain guanidyl or amino group, leaving the z-amino and the z-carboxylate groups to take part in head-to-tail sequences. Thus electrostatic effects appear to be strongly modulated by other factors when amide molecules aggregate. The modulating effect resulting from the characteristic geometry of amino acid molecules and the charge distribution in them are presumably such as to promote head-to-tail sequences.

**Head-to-tail sequences and other patterns of peptide aggregation**

Having established head-to-tail sequences to be the prominent feature of amino acid aggregation, the obvious next step was to investigate the aggregation of peptides using the available structural data. Such an investigation showed that, despite the increased molecular flexibility of peptides in comparison to amino acids and the presence of additional hydrogen bonding groups in them, head-to-tail sequences remain the important feature of peptide aggregation in solid state\textsuperscript{21}. In fact, such sequences are further stabilised in peptides by hydrogen-bonding sequences involving the peptide groups. It turned out that aggregation of peptides in their crystal structures follow a few well-defined patterns which could be predicted with a reasonable degree of success, at least in the case of dipeptides, in terms of a few basic elements, such as sequences mentioned above, involving only main chain atoms. For example, a predicted idealized pattern (for L-alanyl-L-alanine) and the observed pattern in the crystal structure of serglycine are shown in figure 4. Thus peptide aggregation appears to be controlled primarily by interactions involving main chain atoms. Polar groups in side-chains and water molecules when present, disrupt some hydrogen bonding involving main chain atoms, but the basic terms of aggregation are most often left undisturbed. In particular, head-to-tail sequences are retained in most cases despite the presence of such disturbing factors.
Each of the crystal structures used in the above study contains only one type of peptide molecules. It was of interest to investigate the aggregation patterns in situations involving more than one type of peptide molecules. Therefore, the ongoing programme on crystalline complexes in this laboratory\textsuperscript{7,8} was extended to include those involving peptides as well. As part of this programme, the preparation and the x-ray analysis of a heavily hydrated 1:1 complex between L-histidyl-L-serine and glycyl-L-glutamic acid has been carried out\textsuperscript{22}. The crystal structure of this dipeptide-dipeptide complex, the first of its kind to be analysed, is shown in figure 5. As in the case of most of the amino acid complexes analysed in this laboratory\textsuperscript{8,23-27}, the unlike molecules aggregate into separate alternating layers in this complex also. The adjacent layers are held together by interactions involving side chains and an intricate net work of water molecules. (This network, incidentally, provides a good model, at atomic resolution, for the water structure in protein crystals.) Despite the fact that the two layers are made up of different types of molecules, the aggregation patterns in them are remarkably similar. This common pattern, illustrated in figure 6, is in turn very similar to one of the idealized patterns derived for L-alanyl-L-alanine using a few basic structural

Figure 4. (a) A predicted idealized crystalline pattern for L-alanyl-L-alanine. (b) A similar pattern observed in the crystal structure of L-serylglucose.
elements (vide figure 4). Thus the basic patterns of peptide aggregation, with head-to-tail sequences as their central feature, are apparently retained even in complex situations, such as that found in the dipeptide-dipeptide complex, involving disrupting elements like polar side chains and water molecules.

CHIRALITY AND MOLECULAR INTERACTIONS

Chiral selection is among the outstanding problems in evolution. The abiotic organic synthesis of biomolecules would have obviously resulted in a racemic mixture of different molecules. Living organisms, however, contain proteins made up exclusively of L-amino acids and nucleic acids with D-sugars only. Thus, chiral selection must have taken place fairly early in the evolutionary history of the biosphere. Several theories involving, among other things, interactions which do not conserve parity, earth’s magnetic field and polarised light, have been advanced to explain chiral selection. None of these has received universal acceptance. The possibility of chiral selection by accident has also been seriously considered. It is possible that systems having proteins made up of amino acids of the same chirality, are more efficient than others. If, in addition, L-amino acids have a closer affinity to nucleic acids with D-sugars.
indeed appears to be the case\textsuperscript{10}, there could have been an equal number of systems with L-amino acids and D-sugars on the one hand, and with D-amino acids and L-sugars on the other. It could well have been an accident that systems with L-amino acids and D-sugars got selected in living organisms. However, even chiral selection by accident presupposes chiral separation. Chiral separation of biomolecules is usually difficult. For example, it has been demonstrated that DL-amino acids are in general more stable than the corresponding L and D compounds\textsuperscript{30,31}. Although optical resolution has been achieved using techniques such as seeding\textsuperscript{32} and occlusion\textsuperscript{33}, no indication of the possibility of chiral separation of amino acids without the aid of external agencies has so far been demonstrated. It is in this context that the preliminary results obtained by us during co-crystallization experiments involving histidine and aspartic acid become interesting.

Crystallization experiments were carried out using aqueous solutions with acetone or ethanol as precipitants. The experiments involving L-histidine and L-aspartic acid with acetone as the precipitant yielded crystals of a 1:1 complex between the two amino acids. The crystal structure of this complex has already been reported\textsuperscript{25}. Crystals with an identical diffraction pattern were obtained when D-histidine and D-aspartic acid instead of the corresponding L-amino acids, were used in the experiments. Similar experiments with ethanol as the precipitant yielded another crystal form of 1:1 LL and DD complexes\textsuperscript{34}. However, the components crystallized separately when L-histidine and D-aspartic acid or D-histidine and L-aspartic acid were used in the experiments. Attempts to obtain crystals from experiments involving DL-histidine and DL-aspartic acid have not yet been successful. Thus, to summarise, L-histidine and L-aspartic acid as well as D-histidine and D-aspartic acid form crystalline complexes whereas L-histidine and D-aspartic acid as well as D-histidine and L-aspartic acid do not, under identical conditions. These results perhaps constitute the first clear demonstration of the possibility of stronger interactions between different amino acids of the same chirality than between different amino acids of opposite chirality. They appear to suggest that the possibility of chiral separation occurring through intermolecular interactions merits serious consideration.

The choice of histidine for the experiments outlined above appears to have been ideal, in retrospect. All the crystalline complexes between amino acids prepared and x-ray analysed so far have involved a basic amino acid and an acidic amino acid. Electrostatic interactions appear to be necessary for the formation of such solid state complexes. Histidine, however, is less basic than the other two common basic amino acids, namely, arginine and lysine, and electrostatic interactions involving it is not apparently strong enough to override chiral effects. Arginine and lysine, on the other hand, are highly basic and electrostatic effects, override the effects, if any, arising from chiral specificity, in their interactions with acidic amino acids. Therefore, complexation takes place irrespective of the chirality of the molecules involved. Indeed, crystalline complexes of L-arginine with D-aspartic acid and D-glutamic acid have already been prepared and x-ray analysed\textsuperscript{34}. The patterns of molecular aggregation in these structures are strikingly different from those found in the corresponding complexes involving the L-isomers only\textsuperscript{24,26}.

**PRIMITIVE MULTI-MOLECULAR SYSTEMS AND SELF-ASSEMBLY**

The current understanding of the processes by which biopolymers and other small molecules could have come together to form the first self-replicating systems, is extremely hazy. Functional assemblies in the present-day organisms themselves might conceivably provide some clue for approaching this problem. For example, the information available in DNA for synthesising a protein is confined to that about its amino acid sequence. The polypeptide, once synthesised on the ribosomes, folds into a characteristic three-dimensional form through a self-assembly process mediated by non-covalent interactions. The components of a multi-subunit protein can be taken apart; they most often spontaneously
reassemble into the original functional form when put together again in appropriate conditions. The same is true of organelles like ribosomes and most viruses. However, more complex organelles like mitochondria or cells or multicellular organisms do not self-assemble. Therefore, in the hierarchy of biological organisation, self-assembly operates at the lower levels. It appears reasonable to expect the early self-replicating systems to be akin to these lower, and simpler, levels of organisational hierarchy. Therefore, the study of self-assembly processes perhaps provides a fruitful approach towards understanding early biological systems.

Self-assembly, almost invariably, is mediated through comparatively weak non-covalent interactions. A detailed study of the biologically significant non-covalent interactions is therefore important in understanding not only the present-day biological systems, but also the primitive self-assembly systems. Among such interactions, hydrophobic and van der Waals’ interactions are largely responsible for the gross features of biological structure and phenomena while their finer features are often determined by hydrogen bonds, ionic interactions and, to a limited extent, charge-transfer interactions. It is with the latter type of interactions that a part of the X-ray structural work in this laboratory has been concerned. The results of this work, which have been described elsewhere, amply demonstrate the high directionality of these interactions, even at the monomer level. The attractive possibility of the occurrence of specific interactions, implying recognition, and interaction patterns involving two or more hydrogen bonds, has also been explored. The structural and functional information that could be generated using these interactions is substantial and such information could well have been of considerable importance in the emergence of the first primitive functional biomolecular assemblies.

CONCLUSIONS AND PROSPECTS

The studies outlined earlier conclusively demonstrate the propensity of amino acids and short peptides to arrange themselves in such a way as to bring the main chain amino and carboxylate groups into hydrogen-bonded proximity. The geometrical pre-requisite for the occurrence of non-catalytic condensation is satisfied in these head-to-tail arrangements which are retained in the solid state even in complex situations involving the presence of other interacting groups. Admittedly, prominent effects are often over emphasized in crystals on account of the rigorous requirements of translational periodicity and rotational symmetry, and a higher degree of flexibility is likely to exist in the non-crystalline state. However, the main features of aggregation observed in the crystalline state, the only state in which precise studies can be carried out at present on the geometrical aspects of large systems, are generally found to occur even when the rigorous translational and rotational symmetry requirements are absent. Thus the near ubiquitous occurrence of head-to-tail arrangements of amino acids and peptides in crystals, despite variations in their symmetry and composition, indicates that these arrangements are an intrinsic feature of amino acid and peptide aggregation. The investigations that could be, and indeed are being, carried out taking advantage of this feature include an exploration, using theoretical methods, of possible self ordering of amino acid residues in peptides.

As shown earlier, crystallization experiments involving histidine and aspartic acid of L and D forms suggest that chiral separation could be achieved in favourable cases through molecular interactions. The crystal structures of only two complexes containing L and D amino acids, are available at present. Although the aggregation of molecules in these complexes is strikingly different from that in the corresponding LL complexes, more complexes involving L and D amino acids need to be analysed before the possible relevance of aggregation patterns in them to chiral separation or selection, could be explored.

It is reasonable to expect primitive molecular assemblies which gave rise to the early self-replicating systems, to be similar in nature to the self-assembly systems which occur at the lowest levels in the organisational hierarchy of the present day organisms. The structure and action
of these self-assembly systems critically depend upon, as the primitive molecular assemblies must have done, the presence of different types of non-covalent interactions. The geometrical details of some of these interactions, including specific ones, have already been elucidated. However, the number of patterns and specificities that could be generated through non-covalent interactions involving relevant molecules are likely to be very large, and the geometrical and other details of a substantial part of them need to be elucidated before one can begin to understand the self-assembly processes which might have led to the formation of early self-replicating systems.

In studies on chemical evolution, one is trying to elucidate events which have taken place billions of years ago, of which few traces exist. Hence, all ideas on chemical evolution, including our own, have to be necessarily tentative. Yet it appears almost certain that non-covalent molecular interactions must have played a decisive role in the stage of chemical evolution starting from the abiotic condensation of monomers through the formation of primitive self-replicating systems. In any case, as it often happens, the quest for understanding prebiotic phenomena has led in the present case to several results which are of intrinsic interest in relation to biomolecular aggregation and interactions.

ACKNOWLEDGEMENTS

We thank the SERC, Department of Science and Technology, Government of India, for financial support.

24 April 1985

ANNOUNCEMENT

THE STATE OF THE ENVIRONMENT 1985

Air pollution, water pollution, noise, toxic chemicals and hazardous waste risks need further curbing and water, land, forest and wildlife resources need better management according to a new survey on the environment by the OECD. The survey will be submitted to Environment Ministers at their meeting at OECD headquarters in Paris on June 18th–20th, 1985.

The report lauds progress made in the past 15 years in cleaning up rivers, city air, municipal wastes and protecting beauty spots. Among the remaining black spots, the survey notes:

- air pollution from the yearly emission of 55 million tonnes of sulphur oxides, 37 million tonnes of nitrogen oxides, and 38 million tonnes of hydrocarbons in OECD countries is still a health risk and affects ecosystems;
- millions of households are still without access to waste water treatment facilities;
- intense traffic noise is the lot of 110 million people daily in OECD countries;
- hazardous wastes moved within and between countries (more than 100 000 border crossings a year) and existing toxic chemicals as a health and environment hazard, are growing problems;
- other potential problems, on a global scale, arise from the possible impacts of chlorofluorocarbons on the ozone layer and of increases in concentrations of CO₂ in the atmosphere on climate.

The extent of environmental progress and the problems remaining vary on OECD countries—for example, more than 65 per cent of the population in the USA, France, Germany, Sweden and the UK, Kingdom and less than 35 per cent in Japan, Belgium, Greece, Italy, Portugal and Spain are served by water treatment. And high road traffic noise levels (over 65dBA a day) affect about 19 million people in North America, 38 million in Pacific countries, at million in Europe.

The report tracts new environmental port, including the transfer of pollutants between land and water, 'new' pollutants such as a number of organic compounds, the degradation of natural sources (ground water, soil, tropical forests, wildlife resources), and the risks of energy-related or industry-related environmental accidents.

The report takes stock of the latest press imposed by agricultural practices on the environment including pollution of surface water and ground by nitrogenous fertilizers and pesticides, and degradation relating to changing agricultral practices.

And, resulting partly from the impact of the terial demands of the OECD economies, these issues that affect the developing world. They are important too for the OECD countries for they relate to the substantiability of the resource base of economic development and to North-South relations.

The OECD’s ‘State of the Environment’ report, published together with a companion volume, is the kind, of internationally comparable environmental statistics.

Journalists may obtain copies of the reports from the OECD Press Division, 2 rue Andre-Pascal, 75006 Paris, Cedex 16, France.