Proc. Indian Acad. Sci. (Chem. Sci.), Vol. 93, No. 4, May 1984, pp. 687–701.

Evolution of the genetic code: Chemical studies on the genesis of coded α-amino acids

D RANGANATHAN, S RANGANATHAN* and S BAMEZAI
Department of Chemistry, Indian Institute of Technology, Kanpur 208 016, India

Abstract. The choice of the twenty coded L-α-amino acids from innumerable such possibilities has been correlated to selection based on structural versatility and the ability to accommodate new entrants as a function of time. This, in turn, favours an evolutionary profile for the genetic code, an idea that finds much support on the basis of a close examination of the pattern of degeneracy that is present. The focus of the experiments reported in the present paper is to delineate pathways that would lead to a centrally controlled synthesis of α-amino acids and to the selective replacement of one amino acid with another in intact peptides. These endeavours would also be appropriate in the context of the current demonstration that the replacement of a single amino acid residue with another could bring about profound biological changes.

N-trityl serine methyl ester-O-mesylate and N-benzoyl serine methyl ester-O-mesylate have been demonstrated as alanyl transfer synthons. The latter synthon is quite effective and it has been already transformed to N-benzoyl phenylalanine methyl ester, N-benzoyl leucine methyl ester and N-benzoyl tryptophan methyl ester via reaction with appropriate organo cuprate reagents. In principle, as many as 14 of the twenty coded amino acids could be derived in a similar manner from this synthon.

N-benzoyl phenyl alanine methyl ester (BzN-Phe-OMe), BzN-Phe-Phe-OMe and BzN-Phe-Phe-Phe-OMe have been transformed to, respectively, Bz-Asp-OMe, Bz-Asp-Asp-OMe and BzN-Asp-Asp-Asp-OMe in good yields via controlled oxidation with RuO_4 in aqueous acetonitrile. Benzoyl aspartic acid cyclohexyl amide-β-methyl ester has been transformed to NBz-Leu-OMe via Grignard addition and selective dehydroxylation. Thus, the methodology for the Phe → Leu change, of importance with respect to evolution of code, has been delineated.

Whilst the structural relationship between phenyl alanine and leucine is obvious, that in the case of the pair histidine-glutamine is not. Glutamine precursors have been obtained in a single step *via* oxidation of Bz-His-OMe with RuO_4. The mechanism of formation of the products has been delineated by model studies. A formamidomethyl transfer synthon, of potential use in the Gln → His change, has been prepared, and a cycloaddition strategy for the transformation is described.

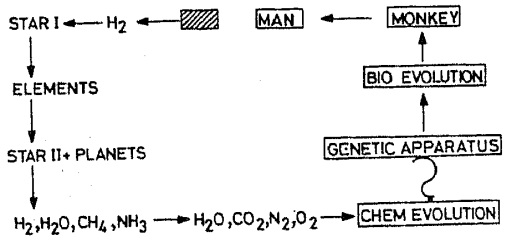
Keywords. Coded α-amino acids; centrally controlled synthesis; side-chain interconversions.

The realisation that Nature obeys the laws of chemistry has led to endeavours relating to the understanding, on a molecular basis, the origin of life itself. A system can be said to be alive if it can sustain and replicate with the assistance of its inbred memory and mechanisms. In chart I, a rational interpretation leading to the evolution of life in our planet is presented.

Whilst processes including the chemical evolution on the one hand and those related to the Darwinian evolution on the other, can be comprehended on a rational basis, the transformation of molecules of diverse complexity to a self-replicating proto-cell is an area that remains to be understood. The overall metamorphosis associated with the organization of molecules to a proto-cell is a consequence of the operation of

* To whom all correspondence should be addressed.

CHART I



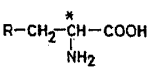
innumerable molecular changes and it is currently generally accepted (Kuhn 1972; Kuhn *et al* 1981) that the best way to understand the formation of a self-replicating proto-cell is to cogently explain the most crucial, if not all, the events that must be related to the formation of life on a molecular basis and in terms of the accepted pathways for molecular transformations.

The present paper endeavours to synthesize, using theory, inference and experiments, an approach to the understanding of the genesis of the coded amino acids, which, in turn, linked through the genetic code, could lead to an important contribution pertaining to the genesis of the information system, which is a prerequisite for life. This is a domain that has hitherto been rarely dwelled upon. Indeed, even basic questions such as the logic of selection and sustenance over millennia, of 20 (L) α-amino acids, from a pool of several hundred thousand possibilities of which about 400 are found in nature and the non-randomness of the code await rational explanation. There is no doubt that the choice of these 20 amino acids must be a result of evolution over long periods, which would imply that the genetic code must have inherent in it, facets that would be a pointer to the evolution of the genetic code itself. Consequently, there is a crucial relationship between the structures of the coded amino acids, that vary only as a function of side chain, and the 64 triplets that constitutes the genetic code.

Time has obliterated direct evidence relating to the evolution of the code, if such were the case. Amongst the lines that may be pursued to secure indirect evidence pertaining to this, could be the understanding of logic behind the amino acid changes that are associated wth essential protein families such as cytochromes, haemoglobins, or hormones such as insulin, oxytocin, vasopressin, as a function of evolution across a large time span. Another possible approach to this problem could be the synthesis of peptides, which might be an ancient precursor to those that exist today based on the probability that amino acid introdunction was a progressive feature and accommodated by lifting the degeneracy of the genetic code. Both these approaches take cognizance of the choice of the 20 L-α-amino acids in our information system, from amongst the many possibilities available. From a totally synthetic point of view, an advantageous approach to such a choice would be to generate these from as few precursors or synthons as possible. Indeed, in nature also, more than half the complement of the α-amino acids arise from side chain transformation of serine and aspartic acid. Besides, the 20 α-amino acids could be arranged into two major structural types as illustrated in chart II and chart III.

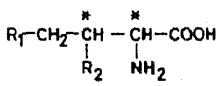
Ten of the 20 coded amino acids could be considered as derivatives of alanine, that is, they could, in principle, be generated from the common grouping, namely, the alanyl synthon and therefore structurally represented as R-ala. Both from biosynthetic and synthetic perception, they could arise by further changes of terminal groups that are

CHART II



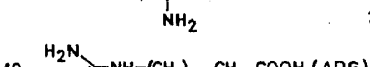
No.	R	NAME	NATURE	No.	R	NAME	NATURE
1	-H	ALA	H	5	-OH	SER	P
2	$-CH(CH_3)_2$	LEU	H	6	-SH	CYS	P
3		PHE	H	7		TYR	P
4		TRP	H	8		HIS	P
				9	$-CONH_2$	ASN	P
				10	-COOH	ASP	A

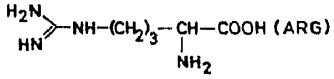
CHART III



No.	R_1	R_2	NAME	NATURE	No.	R_1	R_2	NAME	NATURE
11.	$-SCH_3$	-H	MET	P	14.	-H	$-CH_3$	VAL	H
12.	$-CONH_2$	-H	GLN	P	15.	-H	$-CH_2CH_3$	ILE	H
13.	-COOH	-H	GLU	A	16.	-H	-OH	THR	P

17. H_2N-CH_2-COOH (GLY)

18. 

19. 

20. 

present in serine (R=–OH), cysteine (R=–SH), aspartic acid (R=–COOH) and asparagine (R=–$CONH_2$). In the present work, aspects of this have been successfully simulated synthetically. As mentioned previously, biosynthetically also, most coded amino acids arise from precursors cited above. In chart III are shown a set of six amino acids, three of which, namely, methionine, glutamine and glutamic acid could be considered as derived from the homoalanyl synthon and the remaining three as β-substituted homoalanines. Thus, sixteen of the twenty coded amino acids can be correlated to common precursors. Of the remaining four (chart III), glycine is a unique one, being the simplest and the only one having no chiral centre. The basic amino acids lysine and arginine have their functional groups well separated from the peptide backbone to minimize rupture of the protein backbone by intramolecular protonation. The remaining coded amino acid, namely, proline is the only one that defies classification.

All rationalizations, at the molecular level, relating to the formation of the proto cell from precursors, are in agreement with the development of the information system or the genetic code as a result of increasingly sophisticated mechanisms involving nucleic acid-nucleic acid interactions on one hand and nucleic acid–protein on the other. This would imply a relatively late origin for the triplet code. Consequently, the experimental establishment for an evolutionary nature of the code will provide support for theory, viewed from diverse perspectives. It is almost certain that the code initially referred to fewer than the 20 amino acids, for which it stands today. Two questions logically arise, namely, the nature of the primitive code and the descritption of the primitive amino acids. With reference to the code, another logical criteria has to be annexed, that is, any change of its profile should not disrupt evolutionary processes already in vogue (Crick 1968). A simple solution to this would be an assignment of total degeneracy to the code with reference to the last letter. In sum, it would imply a doublet code. This is reasonable, since, even in the present day translation machinary, 7 out of the 20 amino acids have a "de-facto" doublet code. According to this rationalization, in its early history, all the codes were doublets and, as a function of evolution, late amino acids that imparted advantages were incorporated by lifting of the degeneracy of the code. With reference to the nature of the amino acids that were coded from the beginning, the doublet code could have recognised perhaps as few as eight α-amino acids and these too in terms of their ability to influence the environment and not as function of exact

molecular structure (Woese 1967). Parenthetically, the 20 α-amino acids currently in vogue could be classified into four categories, namely, hydrophobic, polar, largely basic and largely acidic. A striking feature of the code is that the triplets related to each four categories are bunched together, an aspect that would be in harmony with an initial selection that was not necessarily based on exact molecular structure. The implication of this is that, enzyme activity which plays a central role, had a modest or even fortuitous origin and that the enzyme systems are connected with imparting the needed tilt in favour of assertive organisations. The genetic code is presented in table 1.

The genetic code should hold key facets that are relevant to the very origins of life. This is best reflected in the fact that although degeneracy is anticipated since 64 triplets stand for 20 α-amino acids, this degeneracy is not uniform. Chart IV illustrates this aspect of the genetic code.

We have, in chart IV, re-arranged the genetic code into five groups. Type A refer to seven amino-acids, which have a "de-facto" doublet code even today, in the sense that the third letter is totally degenerate. Of great significance are the ten amino acids presented as five pairs in category B, wherein each pair has a common doublet as well as the consistent use of, as the third letter, pyrimidines (U, C) to denote one partner and

Table 1. Genetic Code for amino acids.

First position (5′ end)	Second position U	C	A	G	Third position (3′ end)
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Term	Term	A
	Leu	Ser	Term	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

CHART IV

A	B	D
LEU (CU)	PHE-LEU (UU/U,C-A,G)	ILE (AU/U,C,A)
VAL (GU)	SER-ARG (AG/U,C-A,G)	MET (AUG)
SER (UC)	ASN-LYS (AA/U,C-A,G)	E
PRO (CC)	HIS-GLN (CA/U,C-A,G)	CYS (UG/U,C)
THR (AC)	ASP-GLU (GA/U,C-A,G)	TERM (UGA)
ALA (GC)	C	TRP (UGG)
GLY (GG)	TYR-TERM (UA/U,C-A,G)	

purines (A, G) the other. This is unlikely to be a matter of coincidence. We are of the opinion, that these pairs are historically related and that the degeneracy pertaining to the third letter was partially lifted to accommodate a more recent and in many cases a structurally similar amino acid. For example, it could be envisaged that in certain situations, leucine residues in early proteins, were, as a function of evolution, replaced by phenylalanine. Most pairs in category-B, could be historically so related. This notion is amenable to experimental scrutiny, provided suitable methods can be developed to replace one of the partners with the other in intact proteins. Such an approach would, in addition, lead to strategies for the restructuring of the protein *via* side chain metamorphosis, thus reducing synthetic arduours of classical procedures. In type C, we have tyrosine-termination codon pair much as in the situation in catagory B. In this case, it is possible that the degeneracy pertaining to termination codon was lifted partially to accommodate the late arrival tyrosine. Perhaps, a converse of this is illustrated in type D, where isoleucine degeneracy is partially lifted to accommodate the initiator codon methionine. Again, a similar situation can be observed in type E, where the cysteine has taken up the pyrimidine pair and the purines shared between a terminal codon and tryptophan, perhaps the most complex of the coded α-amino acids. Tryptophan, is most definitely a candidate for late inclusion and this could explain the unique triplet code assigned to it.

Common functional proteins and hormones are present in living systems that are separated by a very large time gap. Amino acid analysis of these compounds (Eck *et al* 1966) has demonstrated replacement of some amino acid residues with others. Barring few exceptions, it has so far been not possible to design a common logical pattern for such replacements, with reference to the needs of the particular living species. However, it would be too unrealistic to conclude that these replacements are a result of only random mutations.

A key feature of the evolutionary model presented here for the genetic code is the notion that incorporation of late amino acids is accommodated by lifting of the degeneracy that is present. As pointed out earlier, since, by and large, amino acids of the same type are bunched together in the code, this incorporation would amount to the adaption of the α-amino acids of the same characteristics. That this alteration is of profound significance is best illustrated with the relatively simple molecules that constitute the neurohypophyseal hormones which are produced in the pituitary and which are present in species separated in the evolutionary scale, by as much as three hundred million years. In each of these species two hormones are present which have a common structural pattern characterised by nine amino acid residues with a di-sulphide bridge connecting the amino acids in position 1 and 6 (chart V).

Isotocin, mesotocin and oxytocin exert control in reproduction, and vasotocin, vasopressin in the regulation of salt and water balance. It may be noted, that the transformation of vasotocin which is present in amphibians to vasopressin that is present in land animals, requires only the substitution of phenyl-alanine for isoleucine in position 3 (chart V). Although, chemically this is merely an interchange between two hydrophobic groups, it precipitates great biological changes which are required to maintain a vastly different ion water retention pattern required in land animals (Acher 1966; Woese 1967). Thus, whereas vasotocin promotes water secretion, vasopressin conserves retention of water. It must be admitted, that this is one of the rare cases where the replacement of a single amino acid residue can be well correlated with evolution. However it is likely that in several cases, there may be present, valid reasons for the

CHART V

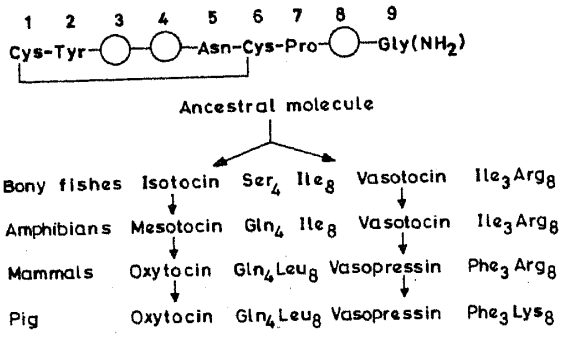
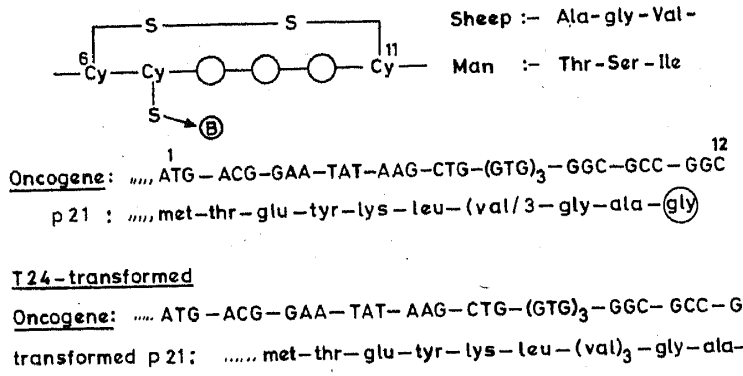


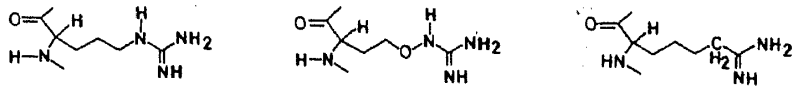
CHART VI



choice of particular amino acid. Less direct proof for the relationship between pairs of amino acids cited in chart IV(B) is also available from comparison of the amino acid sequences of cytochromes of several species separated by a long time span in evolution which show that in several cases leucine has been replaced by phenylalanine, lysine with asparagine and glutamic acid with aspartic acid (Woese 1967). Parenthetically, there appears to be a special relationship between lysine and asparagine. The protein sequence of haemoglobin from patients suffering from hereditary disorders demonstrates that in several cases lysine and asparagine are interchanged. The replacement of as little as one amino acid residue in a protein containing as many as hundred amino acids could have profound biological implications. This has been most dramatically illustrated from studies on the mechanism of the transformation of a harmless oncogene in humans to one that is malignant. Painstaking experimentation pertaining to this transformation by two independent groups has shown that protein p 21 arising from the oncogene on being made infectious by T 24 causes a single mutation in the appropriate genome, transforming the triplet GGC to GTC, which, on translation, would involve the replacement of a glycine residue with valine! Thus, the transformed malignant protein p 21 differs from normal one by one amino acid! Further experiments have shown that even this apparently minor change causes marked differences in the electrophoretic behaviour of the two proteins (chart VI) (Tabin *et al* 1982; Reddy *et al* 1982).

Currently, there is a great deal of interest in the ~ 400 plant non-protein amino acids that have been isolated with respect to the relationship of their structures with

CHART VII



ARGININE CANAVANINE INDOSPICENE

biological, biochemical and toxicological properties. Many toxic situations arise, leading to severe biological malfunctioning, as a result of incorporation of such amino acids in the place of the coded ones by cells that are unable to distinguish the structural difference. Such differences can indeed be very minor and this aspect can best be demonstrated by comparison of the coded amino acid arginine with the toxic non-coded amino acids, canavanine and indospicine (chart VII).

The first book on this subject, which has recently appeared, gives an account of the profound changes that may take place by even a minor error in the choice of appropriate amino acids (Rosenthal 1982).

The long range objectives of our work in this area, the preliminary results of which are presented here, are,

(i) to develop practical methods, using the tools available to the organic chemist, for the preparation of as many coded α-amino acids as possible from a common precursor.

(ii) the selective transformation of one amino acid residue to another in intact peptides with focus in the first phase on the transformation of α-amino acid pairs cited under category B (chart IV). It could be readily seen that the experiments planned under (i) would be also related to the objectives under (ii).

It is hoped that these objectives would be of relevance not only with respect to the problem at hand but also could provide an alternate method for protein synthesis as well as to capabilities for chemically effecting unit replacements. A useful outcome of such an accomplishment would be the ability to replace, as can be done enzymatically presently, the five amino acid residues from sheep insulin with the corresponding unit of human insulin as illustrated in chart VI.

The preparation of an alanyl transfer synthon, a precursor to chiral alanine derived α-amino acids:

The strategy that was adopted for the preparation of an alanyl transfer synthon is shown in chart VIII.

It was anticipated that the readily available α-amino acid, serine can be transformed to an alanyl unit possessing a desired leaving group, which, in turn, could be replaced by an appropriate "R" moiety. The alanyl synthon thus conceived could belong to the several types as shown in chart VIII.

N-trityl serine methyl ester-O-mesylate was prepared from L-serine. In preliminary experiments, when this synthon was reacted with nucleophiles having an alkali metal counter ion, the major product isolated was the dehydro-amino acid, which, in principle, could serve as an achiral alanyl synthon. Interestingly, it was found that the reaction of mesylate with NaI/Zn in DME gave in excellent yields, the chiral cyclic amino acid ester (chart IX). Subsequently, it was possible to make quantities of this N-protected aziridine ester directly from serine methyl ester hydrochloride in a one flask reaction (Smrt *et al* 1957) (chart X).

CHART VIII

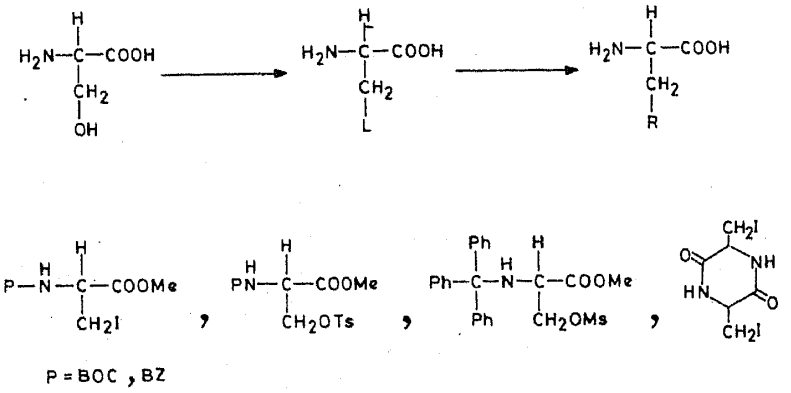


CHART IX

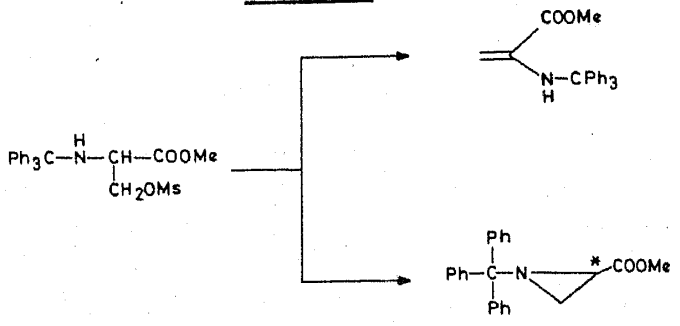
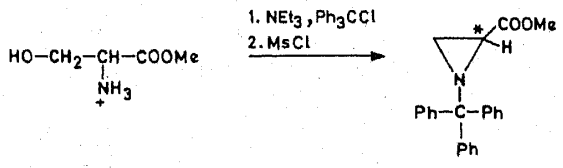


CHART X



The ready availability of the chiral N-trityl aziridine, suggested it to be an attractive synthon for the generation of the alanyl residue *via* nucleophilic attack at the unsubstituted methylene carbon. Accordingly, the aziridinium ion (chart XI) was prepared at $-70°$ by addition of exactly one equivalent of ethereal HCl. In the event, every attempt to open the aziridinium system (Deyrup, 1982) led to the isolation of only the chloride opened products, thus indicating the propensity for the internal return rather than external nucleophilic displacement (chart XI).

In control experiments (chart XI), the aziridine system was treated with one equivalent of ethereal HCl at $-70°$ and allowed to warm up to room temperature. When chloroform was used as a solvent, the product was the expected N-trityl-β-chloroalanine methyl ester. In sharp contrast in DME the β alanine product was formed. Finally, the aziridine, when treated with excess of ethereal HCl, underwent deprotection and ring opening giving rise to β-chloro alanine methyl ester HCl (chart XI).

These events led to the original concept of making the N-protected serine methyl

CHART XI

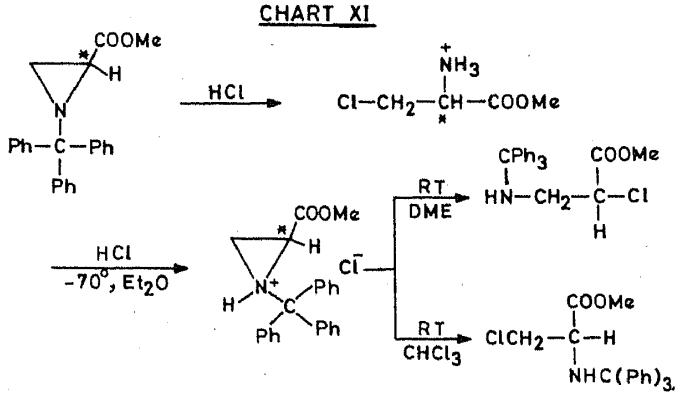
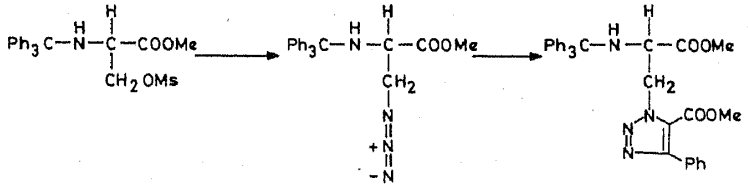


CHART XII

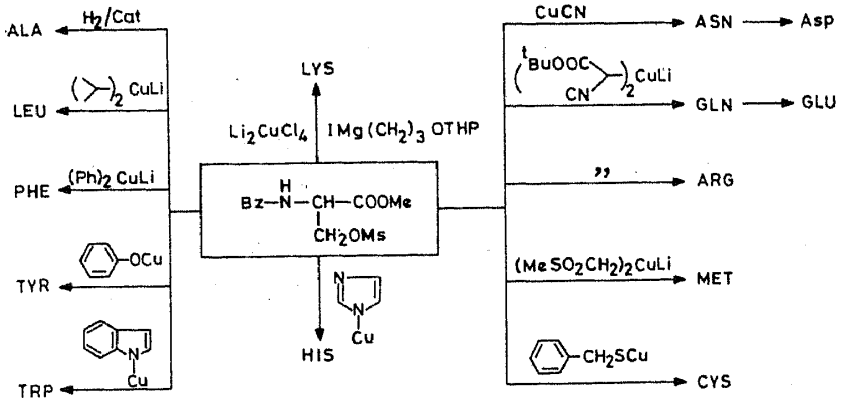


esters having an appropriate leaving group as the alanyl synthon. In preliminary experiments, the feasibility of this was established by reaction of N-trityl-Ser-OMe mesylate with azide ion followed by cycloaddition with methyl phenyl propiolate leading to a novel triazolyl alanine system (chart XII).

In subsequent experiments, it was found that a portion of the N-tritylated substrates underwent deprotection under conditions of the reactions thereby leading to reduced yields. Other protecting groups which were tried without success included the N-Benzyl as well as the N-t-butyl sulphenyl. Similarly, substrates with the nitrogen function protected *via* a carbonyl linkage were also not suitable for this reaction. The best synthon so far studied that could transfer the chiral alanine residue, is the N-Benzoyl serine methyl ester, having the leaving group either as tosylate or mesylate. Nucleophiles, having alkali metal counter ions were not suitable, since they brought about a great deal of elimination. We report, that the alanyl transfer synthon N-Benzoyl serine methyl ester-O-mesylate, could be effectively used in the synthesis of coded α-amino acids *via* reaction with appropriate organo-copper reagents. Thus, reaction of the alanyl synthon with freshly prepared diphenyl lithium cuprate at $-70°$ gave N-benzoyl phenyl-alanine methyl ester in good yield (chart XIII).

The structural assignment for the product is based on spectral comparison with an authentic sample. Similarly, treatment of the alanyl synthon with *in situ* generated di-isopropyl copper lithium and indolyl copper led to the formation of, respectively N-benzoyl leucine methyl ester and N-benzoyl tryptophan methyl ester. Thus, the usefulness of N-benzoyl serine methyl ester-O-mesylate as an alanyl transfer synthon has been clearly established. The capabilities of this synthon are illustrated in chart XIII, where it could be further used for the preparation of as many as fourteen of the twenty amino acids! We plan to explore further the full potential of this synthon and try to translate such changes in intact peptides possessing serine residues. It may be pointed

CHART XIII



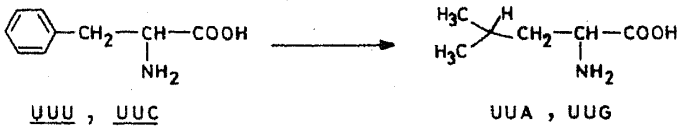
out that methodologies developed in this direction would enable the transformation of, at will, a serine residue in peptides to nearly most of the coded amino acids, thus affording a powerful tool in the synthesis of peptides.

Selective α-amino acid → α-amino acid metamorphosis in intact peptides:

In this section we describe our endeavours pertaining to the transformation of amino acids to the corresponding partner with which it shares a doublet (chart IV) (vide supra). The most appealing amongst these, is the transformation of phenyl alanine to leucine, which would, in sum, amount to the metamorphosis of a phenyl residue to an isopropyl one. Molecular models show that the structural changes associated with the replacement of a phenyl residue in a peptide with leucine would be marginal. Therefore it would be of great interest to study the properties of representative peptides, where a phenylalanine residue has been replaced by leucine, with reference to the possibility that the latter could have been a precursor to the former. In the case of small peptides that serve as hormones, the receptor site adaptability is of consequence and therefore a study of such modified hormones would enable, in addition, an assessment of the receptor site changes, as a function of evolution. From a synthetic point of view, there are only few methods available for the transformation of an aromatic residue with an aliphatic successor. All of these take advantage of the limited nucleophilic activity of the aromatic residue in terms of its reaction with oxidizing reagents. The task at hand then involved the transformation of the benzene ring of phenylalanine residue to an isopropyl one, not only with respect to the individual amino acid but also in intact peptides (chart XIV).

Of the several procedures that were attempted to oxidize the phenyl ring, the most promising turned out to be ruthenium tetroxide in aqueous acetonitrile (Sharpless *et al* 1981). Thus, it was possible to transform N-benzoyl phenylalanine methyl ester to N-benzoyl aspartic acid- α-methyl ester in excellent yields and with complete chiral retention. Parenthetically, as stated earlier, the side chain residue of aspartic acid and serine are good precursors for amino acid replacement in peptides. Therefore, the transformation of the phenyl residue in phenylalanine to the carboxyl function is of additional importance. We have subsequently demonstrated convincingly that this oxidation is feasible in intact peptides by transformation of both Bz-Phe-Phe-OMe and Bz-Phe-Phe-Phe-OMe to, respectively, Bz-Asp-Asp-OMe and Bz-Asp-Asp-Asp-

CHART XIV



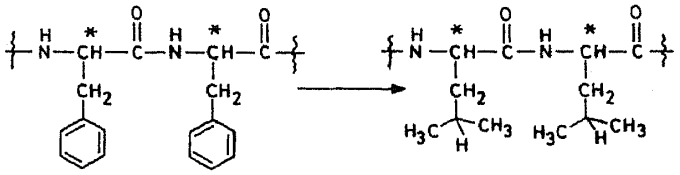
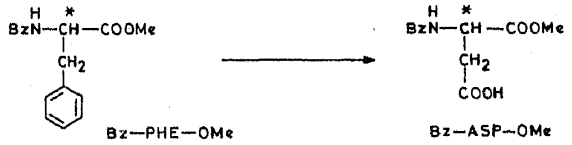
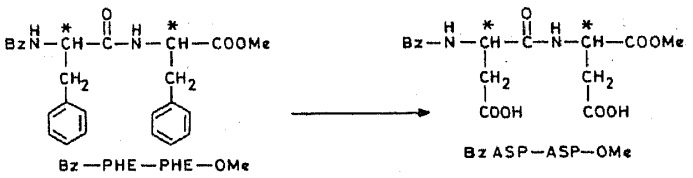


CHART XV





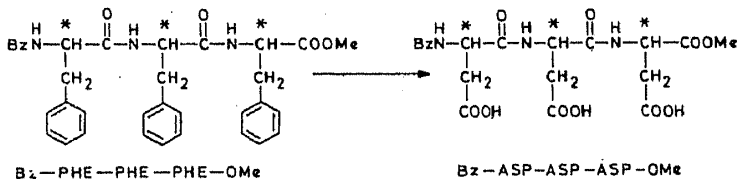
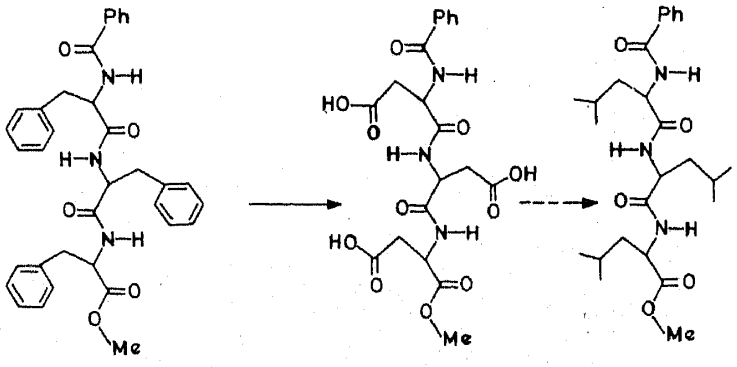


CHART XVI



OMe in good yields and chiral purity (chart XV). In chart XVI the Bz-Phe-Phe-Phe-OMe → Bz-Asp-Asp-Asp-OMe change is shown with focus on the peptide backbone.

The second stage in the Phe-Phe-Phe → Leu-Leu-Leu change (chart XVI) envisages the transformation of the carboxyl function to an isopropyl grouping. This aspect has

been demonstrated with the transformation of N-benzoyl aspartic acid cyclohexylamide-β-methyl ester, wherein the side chain, namely, $-CH_2COOCH_3$ function, is in an environment similar to that in peptides, to N-benzoyl-leucine-cyclohexylamide, *via* sequence, methyl Grignard addition to the tertiary alcohol and dehydroxylation with trimethyl silyl chloride and Zn/acetic acid (chart XVII).

Thus, it is now possible to transform a phenylalanine side chain in peptides to a leucine residue, provided there are no interfering residues. Work is in progress relating to the Phe → Leu change, by the procedures described here, in selected, biologically active peptides and to assess the potency of such modified proteins.

Whilst the structural similarity between phenylalanine and leucine is obvious, that between histidine and glutamine—another pair of amino acids that belong to category B chart IV—is latent. A common feature can however be discerned in the sense that the polar terminal nitrogen is situated in both these amino acids from the peptide backbone at the same distance. Additionally, it could be conceived that histidine may arise by reaction of glutamine, in a regio selective manner, with formamide or its equivalent. Amongst all the 20 coded α-amino acids, histidine plays the most crucial role in the sense that its presence has been identified at the active sites of enzymes more than any other amino acid. In view of this, it was considered to charter practical routes for the transformation of histidine (Code-CAU, CAC) to glutamine (Code-CAA, CAG) on the one hand and the reverse operation on the other. At the outset, it was considered expedient to transplant our experience pertaining to the oxidation of the aromatic ring in phenylalanine to aspartic acid, to the imidazole ring present in histidine. In the event, reaction of histidine with ruthenium tetroxide in acetonitrile was quite complex and the highly functionalized, crystalline carboxylic acid obtained in low yields from these experiments has been assigned, on mechanistic grounds (*vide infra*), a pyrrole structure (chart XIX). Reaction of N-benzoyl histidine methyl ester gave encouraging results. Reaction of this protected histidine with ruthenium tetroxide at pH 6 in aqueous acetonitrile gave two crystalline compounds, the first of which has been identified as N-

CHART XVII

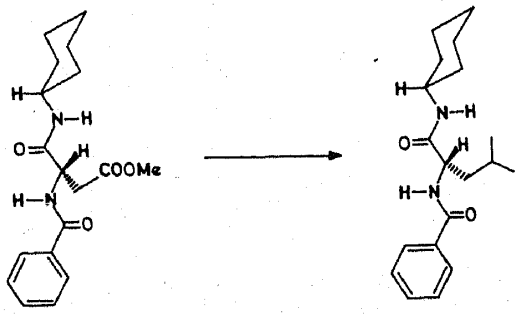
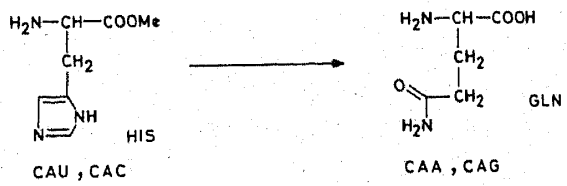


CHART XVIII



benzoyl aspartic acid- α-methyl ester by comparison with an authentic sample (chart XIX). The second crystalline product was neutral and could be obtained consistently in about 30% yields. Based on spectral and analytical data and studies with model system, this compound is assigned the imidazolidine dione structure, shown in chart XIX.

This structure is of interest not only because it is very closely related to glutamine, to which it can be transformed, in principle, by rupture of the ring -CH-NH bond followed by hydrolysis, but also because it represents an intermediate stage in the transformation of glutamine to histidine, *via* incorporation of elements of formamide. It may be noted that the acidic product obtained could, by classical methods, be homologated readily to N-benzoyl-glutamine methyl ester. In order to understand the course of reaction of ruthenium tetroxide on histidine, the reaction was studied on the model system, tetrahydrobenzimidazole, prepared from α-chloro-cyclohexanone. This reaction gave as the sole product the interesting bicyclic system shown in chart XX.

We rationalized the formation of this by primary interaction of the reagent with the 2-position of the imidazole (Wasserman *et al* 1981), leading to a novel electrophilic and unstable 2,5-diazacyclopentadienone system which readily accepts elements of water giving rise to observed product. A similar mechanistic sequence would readily explain the products arising from the reaction of N-benzoyl-histidine methyl ester with ruthenium tetroxide as shown in chart XXI.

A similar series of events, as shown in chart XXII, would explain the formation of the highly polar carboxylic acid obtained when unsubstituted histidine was reacted with ruthenium tetroxide. It may be noted from chart XXII, that in this case, the key

CHART XIX

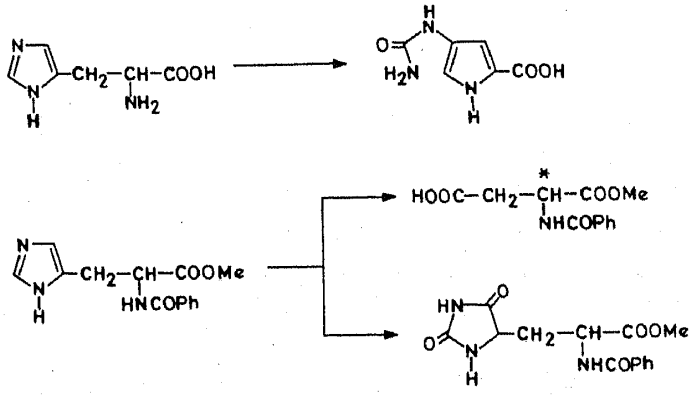


CHART XX

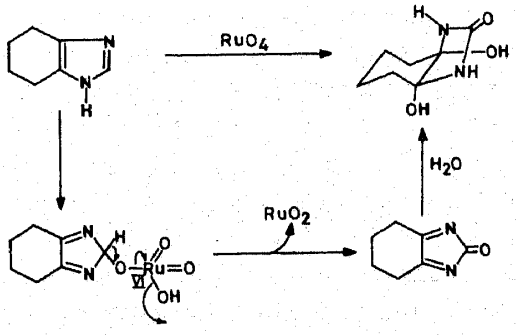


CHART XXI

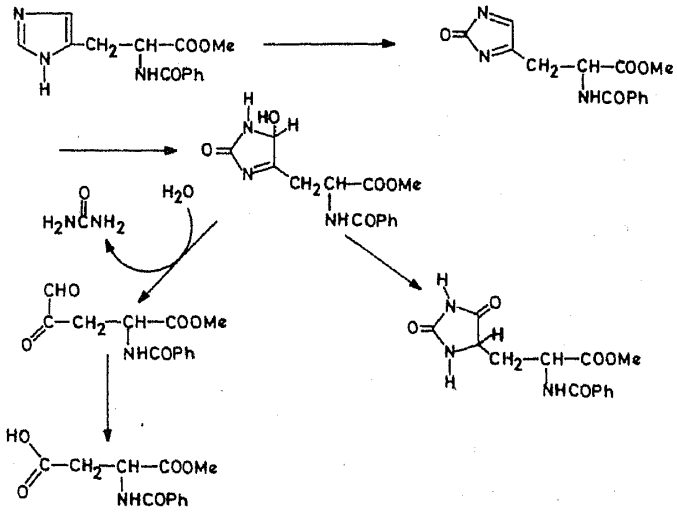
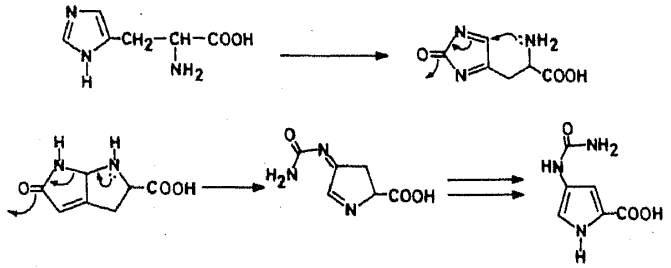


CHART XXII



electrophilic diazacyclopentadienone undergoes intramolecular Micheal type addition with the free amino group and the resulting bicyclic system readily reorganises to the aromatic pyrrrole nucleus, the entire experience amounting to the liberation of the imidazole nitrogens from ring system, to be replaced by the pyrrole nucleus involving the cyclization of the α-amino acid residue. Pathways are currently under study for an effective method for the transformation of the imidazolidine dione (chart XIX) to glutamine.

The perception that an amide system could lead to imidazole *via* incorporation of elements of formamide led us to explore the possibility of the preparation of an appropriate synthon from formamide that could react with either an amide function or its equivalent leading to imidazole. Reaction of formamide with formalin and dimethylamine followed by methylation of the resulting Mannich product, gave, in good yields, the crystalline formamidomethyltrimethyl ammonium iodide, which has been demonstrated to be an excellent forimidomethyl transfer synthon, on reaction with appropriate nucleophiles. Such formamido compounds have been transformed with triphenyl phosphine-CCl_4 to isocyanides and their conjugate bases have been shown to form imidazoles *via* cycloaddition with substituted nitrile functions (chart XXIII) (Van Heusen *et al* 1976).

Finally, asparagine has been converted to N-benzyl oxycarbonyl β-cyanoalanine methyl ester and endeavours are under way to transform this to N-benzyloxycarbonyl

CHART XXIII

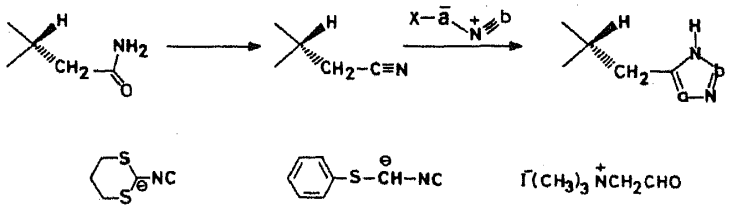
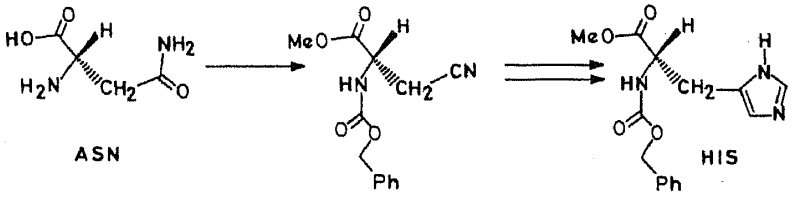


CHART XXIV



histidine methyl ester *via* cycloaddition with synthons described in chart XXIII (chart XXIV).

The results from investigations reported here, provide sufficient incentive to proceed further to reach the objectives outlined in the earlier part of the paper. Further, they are likely to lead to the addition of novel facets to synthetic methodologies.

Acknowledgements

We are most grateful to Mr Kamesh Dixit and Mr Wahe Guru Pal Singh for noteworthy experimental help. We thank Dr Nityanand and Dr Popli of CDRI, Lucknow, for mass spectral and analytical facilities. Financial assistance from DST, New Delhi, is gratefully acknowledged.

References

Acher R 1966 *Angew. Chem. Int.* **5** 798
Crick F H C 1968 *J. Mol. Biol.* **38** 367
Deyrup J E 1982 *Aziridines* (New York: Wiley Interscience)
Eck R V and Dayhoff M O 1966 *Atlas of protein sequence and structure* (Georgetown National Biomedical Research Foundation)
Kuhn H 1972 *Angew. Chem. Int.* **11** 798
Kuhn H and Waser J 1981 *Angew. Chem. Int.* **20** 500
Reddy E P, Reynold R K, Santos E and Barbacid M, 1982 *Nature* (*London*) **300** 149
Rosenthal G A 1982 *Plant nonprotein amino and imino acids* (New York: Academic Press)
Sharpless K B, Carlsen P H J, Katsuki T and Martin V S 1981 *J. Org. Chem.* **46** 3936
Smrt J, Beranek J, Sicher J and Sorm F 1957 *Chem. Listy* **51** 112 (1957 *Chem. Abstr.* **51** 14683)
Tabin C J, Bradley S M, Bargmann C I, Weinberg R A, Papageorge A G, Scolnick E M, Dhar R, Lowy D R and Chang E H, 1982 *Nature* (*London*) **300** 143
Van Heusen A M and Schut J 1976 *Tetrahedron Lett.* 285
Wasserman H H, Wolff M S, Stiller K, Saito I and Pickett J E 1981 *Tetrahedron* **37** *Suppl.* 1 191.
Woese C R 1967 *The genetic code* (New York: Harper and Row)