Protein evolution: The deciphering of latent facets – correlation of synthesis profiles of ribosomally directed proteins and enzyme directed peptides[§]

S RANGANATHAN*,†, D KUNDU** and N TAMILARASU

Biomolecular Research Unit, Regional Research Laboratory, Trivandrum 695 019, India

**Present address: Discovery Laboratory, Organic Chemistry Division III, Indian Institute of Chemical Technology, Uppal Road, Hyderabad 500 007, India*

***Department of Chemistry, Scottish Church College, Calcutta 700 006, India*

†*Corresponding author (Fax, 91-40-7173757; Email, ranganathan@iict.ap.nic.in).*

Key facets pertaining to the evolution of proteins have been probed, using as springboard, the relevant data bases constructed from (i) 60 ribosomally directed proteins, whose 3D structures are known and having 10,000 residues and (ii) from 73 enzyme directed peptides, comprising of 524 residues. The preference profiles, both in terms of the choice of neighbours and the placement of the peptide bonds, have been delineated with respect to each of the 20 coded amino acids. By and large, the preference profile from both the sets are similar, thus giving importance to the nature of the side chains of the coded amino acids. The predictive power of the preference profile has been tested with good results, thus demonstrating the evidence of common preference pathways for peptide formation during evolution. The ribosomally directed protein synthesis, controlled by the genome, proceeds by the addition of single residues at a time. On the other hand, the enzyme directed peptide synthesis largely operates in a more energy conscious block mode, where each constituent of a large enzyme ensemble is engaged in precisely assembling the modules and transfering them to the adjacent one, thus realizing a sequence specific peptide synthesis. Of significance is the fact that, in spite of such divergence in assembly, predictions for neighbour preferences in ribosomally directed protein synthesis work well when applied to enzyme directed peptide synthesis. The findings here are significant since they provide (i) a clear picture of directed peptide synthesis in the absence of direct genomic control, (ii) evidence for the preferred formation of peptide bonds using protein templates, (iii) they also provide evidence for the presence of protein like structures, with catalytic activity, prior to the freezing of the genetic code arising from dominance of the information system and (iv) a logical approach to the evolution of a hierarchical pattern.

1. Introduction

Across the living domain, the functional system is uniformly comprised of 20 α-amino acids (AA), which reflect high selectivity and versatility pertaining to the critical role that they play to regulate life processes. Key to the understanding of protein evolution would be an explanation for the specific manner in which proteins are aggrandized and the controls are exercised by the side chains of AA. We have evidence to show that the selectivity observed in protein structures is rooted in neighbour preferences that can manifest in two ways, namely, gross preferences and those which relate to the placement of the neighbour at the amino end (left) or the carboxyl end (right), with focus on the central residue. By tandem computer analysis of 60 functional proteins having 10,000 residues, whose 3D structures have been established by high resolution X-ray, both the preferences

[§]Presented at the National Symposium on Evolution of Life.

Keywords. Protein evolution: enzyme directed peptide synthesis model

have been delineated for each of the 20 coded AA. Thus, for the first time, an element of preference predictability has been introduced about a targeted AA, progressing either to the left or right. In addition, the knowledge of the 3D structure of the proteins enables the analysis of secondary structural elements like α-helices and β-sheets, where the expected variation in the neighbour preference was clearly seen. The predicted neighbour preference has been confirmed by experiments involving peptidation of diverse AA in water using water soluble carbodiimides. A step has been taken to answer the poser by Bernal (1967), as to the nature of peptides that are likely to be formed when the 20 coded AA are allowed to form peptide bonds in water (Kundu 1994). An area that has received scant attention is the synthesis of peptides in a totally sequence specific manner, in the cytoplasm, by the use of enzyme ensembles, in the place of polysome templates. In the context of protein evolution, the significance here lies in the fact that the ribosomal system operates in step-wise and the enzyme directed system in a modular fashion. From point of energy vantage, a modular approach for the construction of biopolymers is superior. However, the dominance of DNA necessitated the operation of a step-wise mode, in the ribosomal, genome directed protein synthesis. From an evolutionary view point, the enzyme mediated peptide synthesis being genome independent, could take the path of a modular economic mode, as is seen. A correlation of ribosomally directed proteins and enzyme directed peptides, in terms of neighbour preferences of relevant amino acids with respect to peptide bond formation will be significant for the following reasons:

(i) A clearer picture of protein synthesis in the absence of genomic control. (ii) Provide evidence for the formation of peptide using protein templates. (iii) Make feasible the presence of protein like structures with catalytic activity prior to the genetic code, denoting the supremacy of the information system. (iv) Provide a logical approach to the evolution of proteins on a hierarchical pattern.

The very accurately working, but rather complex amino acid sequencing machinery, on the ribosome templates represents the final step in the genetic information transfer. As such, it transmits from generation to generation information about the synthesis of specific proteins. The protocols governing the ribosomally directed protein synthesis orchestrate a rather unusual chemical creation, composed as it is of various sequences of 20 different amino acids which have remained unchanged throughout the evolutionary process beginning with bacteria. Protein structure has the strange and useful property, depending on AA sequence, of giving rise to an inexhaustible variety of catalysts. An event so complicated, so sophisticated and so precise could not have been born, so to say, in one fell swoop. The search for the "missing link" has led to the delineation of alternate pathways in peptide bond formation which take place largely in the cytoplasm of prokaryotes and mediated by enzyme conglomerates. Although the enzymatic formation of essential peptides such as glutathione and patetheine was already known in the preribosmal era, the elucidation of the biosyntheses of more complex peptides followed the unraveling of the genetic code in sixties (Lipman 1971, 1973; Kleinkauf and von Dohren 1982a, b; von Dohren and Kleinkauf 1988). A prediction of poly or multi-enzymatic pathways to peptides had been proposed as early as 1954 (Lipman 1954), which has been now verified for various types of peptides. The ribosomally directed protein synthesis being universal in nature, is evolved to pick up only the 20 coded amino acids all having L configuration. On the other hand, the enzyme ensemble that direct peptide synthesis permit wide variation to incorporate not only L and D configurations but also a host of other substrates as well. It is not surprising therefore, that whereas mRNA coding for 200 or more residues are not uncommon, the largest peptide known to be formed by an enzyme system is still the 20 residue alamethicin (Kleinkauf and von Dohren 1990, 1996; von Dohren *et al* 1997). In enzyme mediated peptide synthesis substructures are organized into multi enzymes and the protein template structure functions either as a set of interacting multi enzymes or a single unit. So far, no other differently organized enzymes system forming identical products have been characterized (Froyshov 1975; Kleinkauf and von Dohren 1982a, b; Kurahashi *et al* 1982; Ghosh *et al* 1986; Keller 1987;

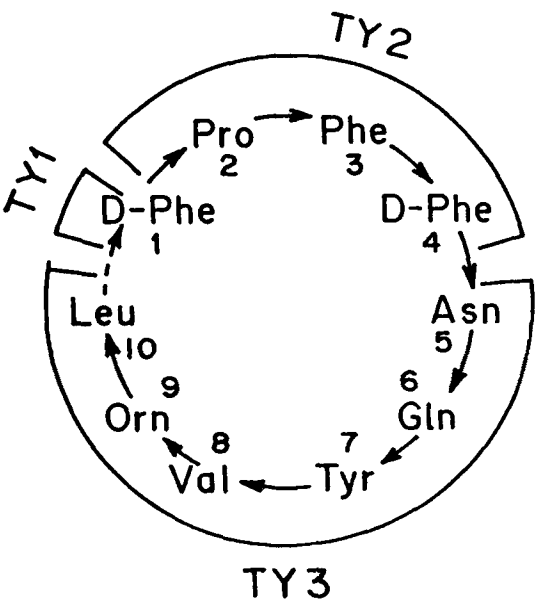
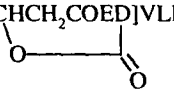
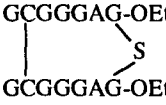
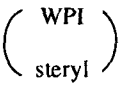
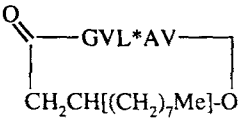
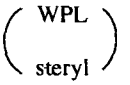
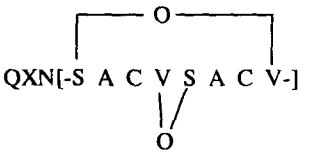
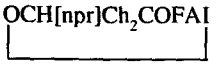


Figure 1. Enzyme assembly of tyrocidine.

Table 1. Enzyme mediated peptides used in the construction of the database.

No.	Name	Coded AA	D-AA	Others	ε	Sequence
1	Antamanide	10	–	–	10	[-VPPAFFPPFF-]
2	Anaphylatoxin	08	–	–	08	ASHLGLAR
3	Bacitracin-F	09	–	01	10	LFI[-KOrnIFHD-]D(NH_2)
4	Bacillomycin L	04	03	01	08	[-βaaDY*D*SES*T-]
5	Bacitracin A	07	02	03	12	NH_2[CLE*LE*I[-KOrnIF*HD^{dl}-]D^{dl}COOH (S)
6	Cycloamanide-I	07	–	–	07	[-PSFFFPI-]
7	Cycloamanide-II	08	–	–	08	[-PMLGFLVL-]
8	Cycloamanide-III	08	–	–	08	[-PMLGFLPL-]
9	Bottromycin A_2	03	–	–	03	[Subs.-GPV-]
10	BE-4	04	–	01	05	[ALDA-Subs.ISOQ]
11	*Bacillus subtilis* C-756 metabolite	07	–	–	07	R(CH_2)$_6$[CHCH$_2$COELLVDLL-]
12	Cyclolinopeptide	09	–	–	09	[-LIILVPPFF-]
13	Desthiomalformin	02	03	–	05	NH_2-IA*A*VL*OH
14	Deoxybouvardin	06	–	–	06	[-YAAYAY-]
15	Didemnin-A	05	–	–	05	L[-TSubs.LPW-]
16	Evolidine	07	–	–	07	[-SFLPVNL-]
17	Enkephalin	05	–	–	05	YSGFSOH (-Si-O-Si-)
18	Echinocandin-D	05	–	–	05	[p-hydroxy phenyl TPTSubs.PT-]
19	Fungisporin	04	04	–	08	[-F*FV*VF*FV*V-]
20	Gramicidin-A	09	06	–	15	HCO-VGAL*AV*VV*WL*WL*WL*W-NH-CH$_2$CH$_2$OH
21	Gramicidin-SA	06	02	02	10	[-VOrnLF*PVOrnLF*P-]
22	Gramicidin-J_1	03	02	02	7	[-VOrnF*L*FPOrn-]
23	Gramicidin-J_2	02	02	02	06	[-VOrnL*F*POrn-]
24	Griselimycin	06	–	04	10	MeCOMeVP[-MeTLPLMeVPMeL*G-]
25	Gratisin	10	–	02	12	[-LFPYVOrnLFPYVOrn-]
26	Leupeptin	03	–	–	03	MeCOLLR
27	Longicatenamycin	02	02	02	06	LI*K*β-hydroxy-5-chloroW*G
28	Lophyrotomin	04	04	–	08	PhA*F*VID*DE*QOH
29	Malformin A	02	03	–	05	C*VC*L*I
30	Malformin C	01	04	–	05	C*C*VLL
31	Mulndocandin	04	–	–	04	[-PSubs$_1$TPSubs$_2$S-]
32	Mycobacillin	07	06	–	13	[-AD*PD*E*YDYSD*LE*D*-]
33	Mycosubtilin	04	04	–	08	[-NQPY*N*N*SD-βamines-]
34	Norsufactin	07	–	–	07	Me(CH_2)$_{10}$-[CHCH$_2$COELLVDLL-]
35	Polymyxin E	07	01	–	08	Pelargonyl KTKK(NHNH$_2$)TKL*K
36	Polymyxin M	07	01	–	08	Pelargonyl KTKK(NHNH$_2$)TKT*K
37	Peptidoglycan	02	02	01	05	ArE*KA*A*
38	Phalloin	07	–	–	07	[-AWLAACP-]
39	Retroantamanide	10	–	–	10	[-FFPPFFAPPV-]
40	Retrogramicidin S	06	02	02	10	[-F*LOrnVPF*LOrnVP-]
41	Tyrocidin A	07	02	01	10	[-VOrnLF*PFF*NQY-]
42	Tyrocidin B	07	02	01	10	[-VOrnLF*PYF*NQY-]
43	Tyrocidin C	07	02	01	10	[-VOrnLF*PYW*NQY-]
44	Tyrocidin E	07	02	01	10	[-VOrnLF*PFF*NQF-]
45	Tuberactinomycin	08	02	–	10	[-HECF*GHECF*G-]
46	Terlipressin	12	–	–	12	GGGCYQNCPKGNH$_2$
47	TL-119	06	–	01	07	AcFLF[-TVA$\alpha\beta$-hydroxyT-]

Table 1. (*Contd.*)

No.	Name	Coded AA	D-AA	Others	ε	Sequence
48	Rhozonin A	06	–	02	08	[-IVVVβ-furylALAβ-furylA-]
49	Rhozonin B	05	–	02	07	[-VVVβ-furylALAβ-furylA-]
50	Viscunamide	02	03	–	05	[-L*IL*IL*-]
51	HC-Toxin	01	02	–	03	[AoeP*AA*-]
52	Viscosin	05	–	–	05	LGSVT
53	Gramicidin S	06	02	02	10	[-VOrnVF*PVOrnVF*P-]
54	Actinomycin-C_2	04	02	04	10	–[-TV*PSarMeV-]–[-TaI*PSarMeV-]
55	Aniso-Actinomycin D	04	02	04	10	–[-TV*PSarMeV-]–[-TaV*PSarMeV-]
56	Actinomycin X	03	02	05	10	–[-TV*PSarMeV-]–[-TV*SarSarMeV-]
57	Esperin	04	01	–	05	$CH_3(CH_2)_9$–[CHCH$_2$COED]VLL* 
58	Geodiamolides	03	–	–	03	A—Y—A—
59	Glumamycin	07	–	02	09	–D 3Me DDGDG ery αβAibu VPNH
60	Lanthionine	14	–	–	14	GCGGGAG-OEt  GCGGGAG-OEt
61	Mycoplanecin A	09	–	–	09	EtCOCOVP[-TLP aminoseplanyl VPLG-]
62	Norphalloin	07	–	–	07	AWVAT 
63	Nummularia	03	–	–	03	
64	Protodestruxin	04	–	01	05	[-O-CH(CH$_2$iPr)COPIVAβ-Ala-]
65	Isarin	04	01	–	05	
66	Sativanine E	03	–	–	03	
67	Triostin A	08	–	–	08	
68	Vernamycin B	04	–	01	05	–[-TAPF-]
69	Dorcidin	05	–	01	06	[-TAPFDPhG-]
70	Ostreogrycin B	04	–	01	05	–[-TAPF-]
71	K-13	03	–	–	03	Ac[YYY]
72	Beavellide	03	–	–	03	OCH[npr]Ch$_2$COFAI 
73	Dextruxin	03	–	02	05	[-NMeAβAlaOCH[ipr]PIV-]

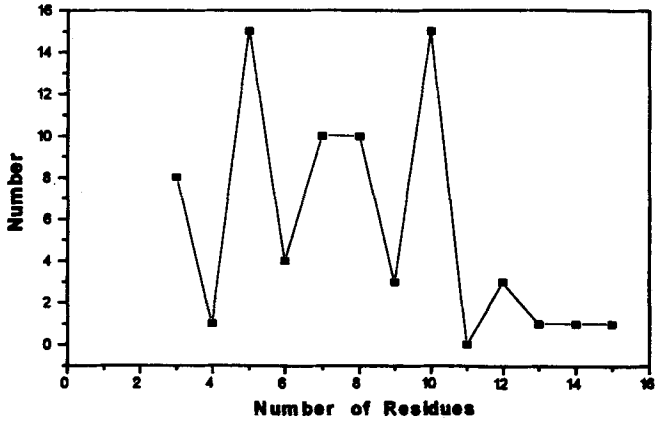


Figure 2. Amino acid residue number distribution in enzyme mediated peptides from the database.

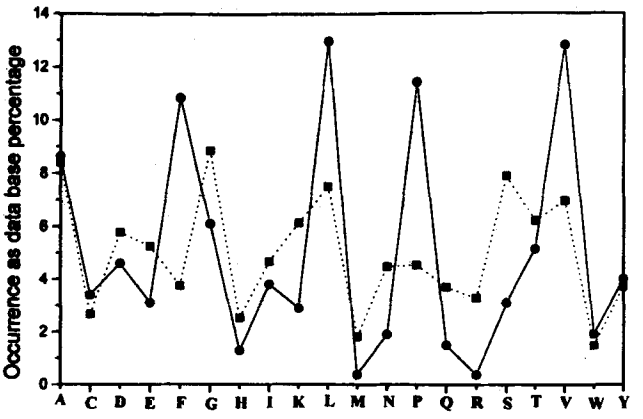


Figure 3. Amino acid percentage occurrence profiles of ribosomally directed proteins (- - - - -) and enzyme directed peptides (——).

Table 2. Percentage occurrence profile.

	Enzyme mediated peptides		Ribosomally directed proteins	
Amino acid	Total	Per cent	Per cent	Total
Alanine	44	8·40	8·65	815
Cysteine	18	3·40	2·67	252
Aspartic acid	24	4·60	5·78	545
Glutamic acid	16	3·10	5·25	495
Phenylalanine	57	10·87	3·75	354
Glycine	32	6·10	8·86	835
Histidine	7	1·30	2·53	239
Iso leucine	20	3·80	4·66	439
Lysine	15	2·90	6·14	579
Leucine	68	12·98	7·52	709
Methionine	2	0·38	1·82	172
Asparagine	10	1·90	4·47	421
Proline	60	11·45	4·53	427
Glutamine	8	1·50	3·70	349
Arginine	2	0·38	3·28	309
Serine	16	3·10	7·91	745
Threonine	27	5·15	6·22	586
Valine	67	12·80	6·96	656
Tryptophan	10	1·90	1·47	139
Tyrosine	21	4	3·71	350
	524	100	100	9,416

Vater 1990). Typical enzyme directed peptide synthesis can be illustrated with tyrocidine (figure 1).

During the simultaneous production of at least two different enzyme directed peptides (tyrocidine-gramicidin A, surfactin-iturin) no mixed products were detected, thus reinforcing the notion of individual collection of modules for each of the peptides, which do not intermix (von Dohren and Marahiel 1990; von Dohren 1995). It is tempting to speculate that the enzyme ensemble also evolved and the initiation of the peptide synthesis and the need to have the assembly of all components to form products, also evolved from precursors that have ability to recognize specific AA in a predetermined arrangement. Enzyme directed peptide synthesis also takes place in eukaryotic organisms. The best illustration here would be that of cyclosporin, produced by the eukaryotic *Beauvenia nivea*. A single multifunctional enzyme orchestrates all the events. It seems that, while prokaryotic synthetase systems comprise of module ensembles, necessitating trans peptidyl transfer, eukaryotic peptide synthetases are fully integrated to giant enzyme modules. Thus, cyclosporin synthetase represents the largest known enzyme (690 kDa) and catalyses a total of 40 reactions (Lawen and Zocher 1990).

2. Methods

The generation of an extensive database from 60 proteins, whose 3D structures are known, having 10,000 residues, enabled the demonstration of neighbour preferences both in terms of choice and in the formation of peptide bonds. The leads here were confirmed by experiments (Kundu 1994). Thus, in this first phase, the focus was to establish the presence of peptide non random sequences, in the early stages of protein evolution. A correlation between existing peptide modules and functional proteins was secured in the second phase, on the premise that the early imprints of protein evolution should manifest in both the systems. A similar database analysis of 135 zinc finger modules [30 residue peptide in a $(\text{thiolate})_2$ $(\text{histidine})_2$ Zn environment which effect DNA triplet recognition] showed good match with that from functional proteins (vide supra). In addition, this provided a link between DNA and proteins and made feasible the presence of sequence specific DNA-peptide composites, possible harbingers of DNA and proteins respectively. Sequence

specific peptides that are constructed in the cytoplasm and not under direct genetic control and their correlation with data secured from the earlier two phases form the focus of endeavours reported here. A correlation here would be significant, since the protocols for peptide assembly are divergent. The reasonable assumption that in both the ribosomally directed protein synthesis and the enzyme directed peptide synthesis, the peptidation follows similar mechanistic pathways, would dictate that a preference profile for neighbours should be similar in the latter as well. Endeavours to secure database pertaining to sequences in peptides of non ribosomal origin did not succeed, consequently the construction of such a database had to be performed manually. In table 1 is presented 73 non ribosomal peptides which have been used in the construction of database pertaining to the present analysis.

3. Results and discussion

In figure 2 is presented the amino acid residue distribution in the enzyme directed peptide database. The distribution is uneven, but brings out the major presence of 5 and 10 residue constructs.

Unlike proteins of ribosomal origin, a rational analysis here is complicated by the intervention of non coded AA and other structural units within the peptide frame work. This aspect has clearly been shown in table 1. Necessarily, therefore, the sequence analysis had to be restricted to AA that are present in the code complement including their D-analogues. The inclusion of D-analogues seems justified since in the analysis of neighbour preferences for peptidation, both antipodes would proceed through identical transition states that differ only in the spatial distribution of ligands. The AA distribution of the coded AA or their D-analogues in the basic set of 524 residues, from enzyme directed peptide database is presented in table 2, along with the percentage presence of each in the set. Similar data from 10,000 residues from 60 ribosomally directed proteins is also presented in table 2. A comparison of percentage occurrence of ribosomally directed proteins and enzyme directed peptides, presented in figure 3, is striking. The excessive presence of phenylalanine, leucine, proline and valine and the paucity of lysine, arginine, asparagine, glutamine and serine should make the profile of enzyme directed peptides as hydrophobic. The excessive presence of proline is related to the antibiotic action of these class of peptides.

Neighbour residues analysis of the enzyme directed peptide set was performed using the program presented in figure 4. The results presented in table 3 are compared with a neighbour non preference profile.

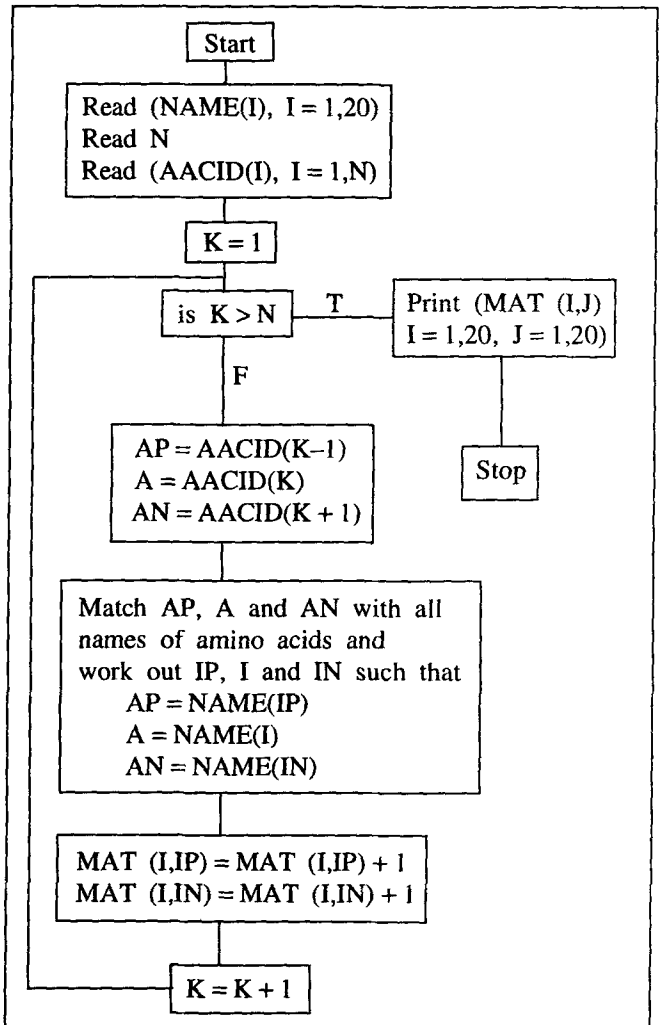
The non preference value (NPV) was computed as follows:

Total number of residues in the set = 524
let 'a' be the number of amino acid A in the set

$\therefore$ fraction of 'a' in the set = $a/524$

$\therefore$ if 'b' were to be distributed as neighbours, without preference, over the set, A will receive, $[a/524] \times [b]$

since each non terminal residue can have 2 neighbour sets [i.e., AB, BA] the non preference neighbour distribution [NPV] for A with respect to B will be,



NAME: Array of dimension 20 which contains names of all amino acids (20).
N = Total number of amino acids present in proteins.
AACID = Array of dimension N, stores distribution of N amino acids in proteins.
MAT = Matrix of dimension 20 by 20. Value of MAT(I,J) represents number of occurrence when NAME(J) found to be NAME(I) of neighbour.

Figure 4. Neighbour residue analysis.

Table 3. Relative neighbour preferences observed in a total of 524 pairs of examples (top right) and those expected on the basis of non preference for neighbour (bottom left).

	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	
	10	3	3	0	4	3	0	2	1	8	0	0	7	0	1	3	4	6	2	6	A
A	6	2	0	2	2	3	0	1	0	2	0	1	3	0	0	0	1	5	0	1	C
C	2	1	6	4	1	3	2	1	0	4	0	0	2	0	0	3	0	3	0	4	D
D	3	2	2	0	0	0	2	2	0	5	0	0	0	1	0	2	0	0	0	1	E
E	2	1	1	1	22	5	2	2	0	19	0	3	25	2	0	3	1	10	0	2	F
F	8	3	4	2	10	6	2	0	1	7	0	0	1	0	0	2	0	2	0	0	G
G	2	1	3	2	6	3	0	0	0	1	0	0	0	0	0	1	0	0	0	0	H
H	1	0	1	0	2	1	0	2	3	8	0	0	5	0	0	0	0	4	0	0	I
I	2	1	2	1	3	1	1	2	4	2	0	0	1	0	0	0	10	0	0	0	K
K	3	1	1	1	2	1	0	1	1	16	2	1	11	0	1	1	2	9	7	0	L
L	9	4	5	3	13	6	2	4	3	14	0	0	2	0	0	0	0	0	0	0	M
M	0	0	0	0	0	0	0	0	0	1	0	2	0	6	0	1	0	1	1	1	N
N	2	1	1	1	2	1	0	1	1	2	0	0	10	1	0	2	5	17	3	5	P
P	8	3	4	4	11	5	2	4	2	13	0	2	12	0	0	0	0	0	0	3	Q
Q	1	1	1	0	2	1	0	1	0	1	0	0	2	0	0	0	0	0	0	0	R
R	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	2	0	2	S
S	2	1	1	1	3	1	0	1	1	3	0	1	3	0	0	1	0	6	0	0	T
T	3	1	1	2	4	2	1	0	1	5	0	1	3	1	0	1	1	14	2	5	V
V	8	3	4	3	11	5	2	3	2	11	1	2	12	1	1	2	4	11	0	1	W
W	2	1	1	1	2	1	0	1	1	2	0	0	2	0	0	1	1	2	0	6	Y
Y	3	1	2	1	4	2	1	2	1	4	0	1	4	1	0	1	1	3	1	2	
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	

NON PREFERENCE VALUE

COMPUTED VALUE

$$[2a/524] \times [b]. \quad (1)$$

In the case of proteins, since the terminal residues would be quite small, the above equation will be valid. However, in the case of enzyme directed peptides, both terminal residues as well as those sandwiched between non coded amino acids, have to be taken into account in the calculation of NPV, since in these cases, the neighbour will be cited only once. Therefore, in the case of enzyme directed peptides, equation (1) has to be modified as follows; of the total 'a' residues of A in the set, if a_x are non terminal, equation (1) becomes:

contribution from non terminal residues
$= [2a_x/524] \times [b]$;

contribution from terminal and sandwiched residues
$= [[a - a_x]/524] \times b$;

$\therefore$ NPV for A with respect to neighbour B
$= [[a + a_x]/524] \times b.$ (2)

The NPV presented in table 3, have been derived using equation (2), where the non terminal residues were identified manually. It could be readily seen from table 3 that the choice of neighbours is selective as was observed for proteins (Kundu 1994). Significant deviations from NPVare presented in table 4.

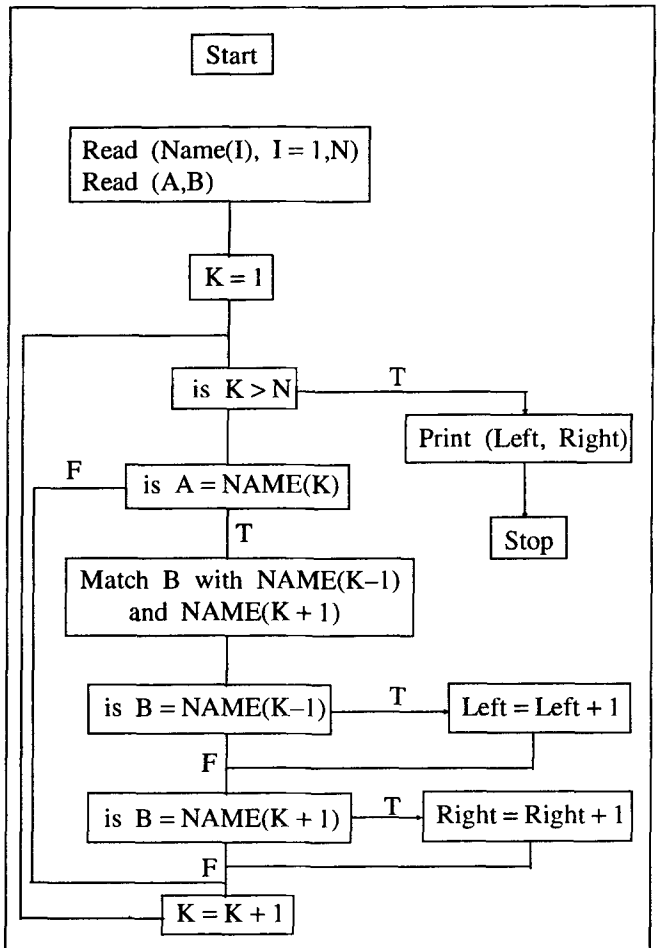
Table 4. Significant deviations from NPV.

	Positive deviation			Negative deviation	
Neighbour	NPV	Observed	Neighbour	NPV	Observed
DD	2	6	AF	8	4
FF	10	22	AV	8	6
FP	11	25	DF	4	1
KK	1	4	DP	4	2
KT	1	10	EP	4	0
LW	2	7	EV	3	0
NQ	0	6	FT	4	1
			GP	5	1
			GV	5	2
			LY	4	0

Particularly notable in the positive preference is FP and NQ. Table 4 highlights preference for clustering of

like residues (DD, FF, KK, LW, NQ). The negative preference bring out selectivity as shown by Asp, Glu and Pro. No significant correlation could be seen on comparison of the preference profiles seen here with that from the protein database. This is reasonable, since even amongst proteins, secondary structural elements like α-helix and β-sheet had their own preference profiles.

Using a second program (figure 5) the neighbour preference was refined to left-right preferences (table 5). This would show that in the construction of enzyme directed peptides, both neighbour preference as well as that for placement of a particular residue either to the left or right are seen.



NAME: Array of dimension n containing the protein sequence.
N: Total number of amino acids in the protein sequence.
A,B: Amino acids B is the residue to be tested for presence on the left or right of A.
Left = Number of times B comes to the left of A.
Right = Number of times B comes to the right of A.

Figure 5. 'Left-right' preference analysis.

The test for common features in ribosomally directed and enzyme directed synthesis can be applied at two key points. As stated earlier, in enzyme directed synthesis, each constituent of the enzyme ensemble orchestrates the construction of the specific module, aligns it for forward peptidation and transfer. The module sequence actually so constructed can be compared with one predicted on the basis of neighbour preference from the relevant database (table 3). Additionally, the direction of peptidation can also be predicted (table 5). We illustrate the above with successful application to the biosynthesis of tyrocidine (figure 6).

This cyclic peptide having 10 residues is constructed by a three enzyme ensemble. The synthesis is initiated with peptide bond formation involving D-Phe, harboured by the light enzyme, by transfer to an intermediate enzyme that constructs and holds the module Pro-Phe-D-Phe. The tetra peptide, D-Phe-Pro-Phe-D-Phe, thus formed is then transferred from the intermediate enzyme to heavy enzyme involving the D-Phe-Asn peptide bond. The most complex system involved in tyrocidine biosynthesis, namely, the heavy enzyme, assembles the hexa peptide Asn-Gln-Tyr-Val-Orn-Leu and accepts the tetra peptide module from the light enzyme to provide the open deca-peptide. The biosynthetic cycle is then completed with peptide bond formation involving Leu (residue 10) and D-Phe (residue 1). The total biosynthetic cycle can be analysed in terms of the construction of the module Pro-Phe-D-Phe by the intermediate enzyme and module Asn-Gln-Tyr-Val-Orn-Leu by the heavy enzyme and peptide bond formation at 3 sites, namely, D-Phe → Pro, D-Phe → Asn and Leu → D-Phe. As could

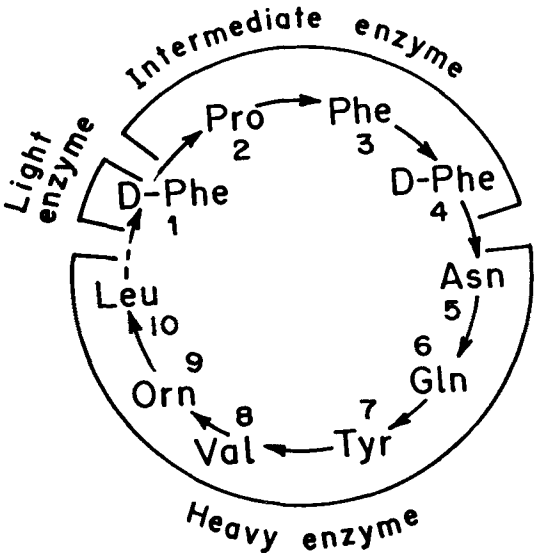


Figure 6. Module alignment in tyrocidine biosynthesis.

Table 5. Master analysis for neighbour (left-right) preference (enzyme mediated peptides, 482 residues).

NEIGHBOUR	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	05/05	03/00	01/02	00/00	02/02	01/02	00/00	01/01	00/01	02/06	00/00	00/00	04/03	00/00	01/00	01/02	01/03	03/03	02/00	03/03
C	00/03	01/01	00/00	00/02	02/00	01/02	00/00	00/01	00/00	02/00	00/00	00/01	03/00	00/00	00/00	00/00	00/01	04/01	00/00	01/00
D	02/01	00/00	03/03	02/02	00/01	02/01	00/02	00/01	00/00	03/01	00/00	00/00	01/01	00/00	00/00	01/02	00/00	01/02	00/00	02/02
E	00/00	02/00	02/02	00/00	00/00	00/00	00/02	02/00	00/00	02/03	00/00	00/00	00/00	01/00	00/00	01/01	00/00	00/00	00/00	01/00
F	02/02	00/02	01/00	0/00	11/11	02/03	02/00	00/02	00/00	08/11	00/00	03/00	15/10	01/01	00/00	01/02	01/00	05/05	00/00	00/02
G	02/01	02/01	01/02	00/00	03/02	03/03	02/00	00/00	00/01	02/05	00/00	00/00	01/00	00/00	00/00	01/01	00/00	01/01	00/00	00/00
H	00/00	00/00	02/00	02/00	00/02	00/02	00/00	00/00	00/00	01/00	00/00	00/00	00/00	00/00	00/00	00/01	00/00	00/00	00/00	00/00
I	01/01	01/00	01/00	00/02	02/00	00/00	00/00	01/01	03/00	03/05	00/00	00/00	01/04	00/00	00/00	00/00	00/00	03/01	00/00	00/00
K	01/00	00/00	00/00	00/00	00/00	01/00	00/00	03/00	02/02	01/01	00/00	00/00	00/01	00/00	00/00	00/00	05/05	00/00	00/00	00/00
L	06/02	00/02	01/03	03/02	11/08	05/02	00/01	05/03	01/01	08/08	00/02	00/01	07/04	00/00	01/00	01/00	01/01	04/05	03/04	00/00
M	00/00	00/00	00/00	00/00	00/00	00/00	00/00	00/00	00/00	02/00	00/00	00/00	00/02	00/00	00/00	00/00	00/00	00/00	00/00	00/00
N	00/00	01/00	00/00	00/00	00/03	00/00	00/00	00/00	00/00	01/00	00/00	01/01	00/00	05/01	00/00	01/00	00/00	00/01	00/01	00/01
P	03/04	00/03	01/01	00/00	10/15	00/01	00/00	04/01	01/00	04/07	02/00	00/00	05/05	00/01	00/00	01/01	03/02	07/10	01/02	05/00
Q	00/00	00/00	00/00	00/01	01/01	00/00	00/00	00/00	00/00	00/00	00/00	01/05	01/00	00/00	00/00	00/00	00/00	00/00	00/00	03/00
R	00/01	00/00	00/00	00/00	00/00	00/00	00/00	00/00	00/00	00/01	00/00	00/00	00/00	00/00	00/00	00/00	00/00	00/00	00/00	00/00
S	02/01	00/00	02/01	01/01	02/01	01/01	01/00	00/00	00/00	00/01	00/00	00/01	01/01	00/00	00/00	00/00	01/00	01/01	00/00	00/02
T	03/01	01/00	00/00	00/00	00/01	00/00	00/00	00/00	05/05	01/01	00/00	00/00	02/03	00/00	00/00	00/01	00/00	05/01	00/00	00/00
V	03/03	01/04	02/01	00/00	05/05	01/01	00/00	01/03	00/00	05/04	00/00	01/00	10/07	00/00	00/00	01/01	01/05	07/07	01/01	00/05
W	00/02	00/00	00/00	00/00	00/00	00/00	00/00	00/00	00/00	04/03	00/00	01/00	02/01	00/00	00/00	00/00	00/00	01/01	00/00	00/01
Y	03/03	00/01	02/02	00/01	02/00	00/00	00/00	00/00	00/00	00/00	00/00	01/00	00/05	00/03	00/00	02/00	00/00	05/00	01/00	03/03

CENTRAL RESIDUE

READ OUT FORMAT: CENTRAL RESIDUE X CAN HAVE NEIGHBOUR
PROFILE YX [LEFT PREFERENCE] OR XY [RIGHT PREFERENCE]
EXAMPLE: FOR P, FP : PF :: 15 : 10

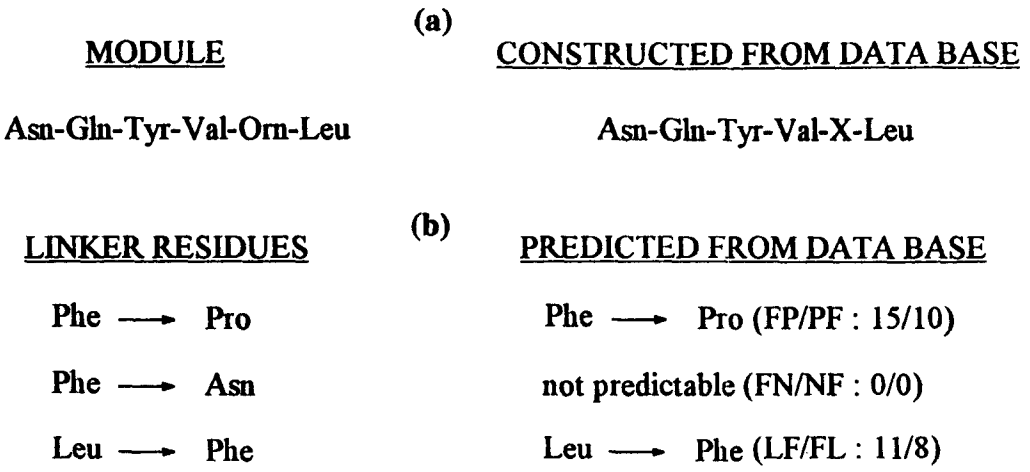


Figure 7. Tyrocidine biosynthesis: Database correlation with (**a**) hexa peptide from heavy enzyme module and that predicted from database neighbour preference correlation. (**b**) The direction of intermodule peptidation based on left right preferences derived from the database.

be seen from figure 7 the actual module, Asn-Gln-Tyr-Val-Orn-Leu matches perfectly from that constructed from the neighbour preference database (table 3). In addition the direction of peptide bond formation is also as predicted (table 5). These findings support the notion that preferences in the choice of neighbours as well as their placement with respect to peptide bond formation, must have played a key role in the evolution of the enzyme ensemble associated with the tyrocidine biosynthesis. Most significantly, similar analysis was largely applicable to several enzyme directed peptides, having open and closed frames.

The analysis of peptide synthesis by enzyme directed mode, can be correlated to two significant events, namely, the evolution of individual synthetase molecules and their subsequent assembly, appropriate for trans peptidation and transfer. The evolution of such modules can be considered as arising from neighbour selective amalgamation of sub modules that have ability to hold single residues. Thus the assembly of peptides in a specific sequence within each module can be understood. The modules could be either reversibly assembled to the synthetase complex as is seen in prokaryotes or covalently linked to a large module as with eukaryotes in such a fashion as to conform to the desired preference for peptidation, as derived in the present work. DNA sequence analysis of gene clusters of synthetases that direct peptidation have revealed extensive homology amongst prokaryotic systems (Kleinkauf and von Dohren 1996) which would imply the conservation of structures suitable to the task at hand, namely, the acceptance of specific AA, oriented peptidation to blocks and transfer to the neighbour, if needed. A protein system that can recognize individual AA would naturally correspond to early stages of evolution leading to an assembly line mode for the generation of sequence specific peptides. These could be the harbingers of functional and information system, by favourable interactions attended with increasing structural complexity and system optimization. Our findings that in functional proteins, zinc finger peptides and enzyme directed peptides, the same traits-neighbour preference, left-right preference prevail, is suggestive of a hierarchical pattern in protein evolution. In sum, the concept of preference profile and the direction of peptidation are features that are common in all peptidations. The fact that the two pathways aggrandize amino acid residues by different mechanisms yet following the same basic principle, do highlight the key role played by the organic chemistry of the side chains of α-amino acids. Thus a deeper understanding here would have ramifications in diverse domains of peptides, such as the understanding of their structures, protein folding and protein design.

Acknowledgements

We thank Dr Darshan Ranganathan and Dr G P Singh for their advice. Financial assistance from the Indian National Science Academy, New Delhi is gratefully acknowledged.

References

Bernal J D 1967 *The origin of life* (London: Weidenfeld and Nicolson)

Froyshov O 1975 Enzyme bound intermediates in the biosynthesis of Bacitracin; *Eur. J. Biochem.* **59** 201–206

Ghosh S K, Majumdar S, Mukopadhyay N K and Bose S K 1986 Role of ATP and enzymebound nascent peptides in the control of elongation for mycobacillin synthesis; *Biochem. J.* **240** 265–268

Keller U 1987 Actinomycis synthetases. Multifunctional enzymes responsible for the synthesis of the peptide chains of Actinomycin; *J. Biol. Chem.* **262** 5852–5856

Kleinkauf H and von Dohren H 1982a *Trends in antibiotic research* (Tokyo: Antibiotics Association)

Kleinkauf H and von Dohren H 1982b *Peptide antibiotics-biosynthesis and functions* (Berlin: de Gruyter)

Kleinkauf H and von Dohren H 1990 Nonribosomal biosynthesis of peptide antibiotics; *Eur. J. Biochem.* **192** 1–15

Kleinkauf H and von Dohren H 1996 A nonribosomal system of peptide synthesis; *Eur. J. Biochem.* **236** 335–351

Kundu D 1994 *Protein evolution: The deciphering of latent facets,* Ph D thesis, Indian Institute of Technology, Kanpur

Kurahashi K, Komura S, Akashi K and Nishio C 1982 Biosynthesis of peptide antibiotics; in *Peptide antibiotics-biosynthesis and functions* (eds) H Klenkauf and H von Dorfen (Berlin: de Gruyter) pp 275–288

Lawen A and Zocher R 1990 Cyclosporin synthetase. The most complex peptide synthesizing multi enzyme polypeptide so far described; *J. Biol. Chem.* **265** 11355–11360

Lipman F 1971 Attempts to map a process evolution of peptide biosynthesis. Synthesis of peptide antibiotics from thiol linked amino acids parallels fatty acid synthesis; *Science* **173** 875–884

Lipman F 1954 *The mechanism of enzyme action* (Baltimore: The Johns Hopkins University Press)

Lipman F 1973 Nonribosomal polypeptide synthesis on polyenzyme templates; *Acc. Chem. Res.* **6** 361–367

Vater J 1990 *Biochemistry of peptide antibiotics* (Berlin: de Gruyter)

von Dohren H 1995 Peptide and peptide derived antibiotics; in *Genetics and biochemistry of antibiotic production* (eds) L C Vining and C Stuttard (Boston: Butterworth) pp 129–172

von Dohren H, Keller U, Vater J and Zocher R 1997 Multifunctional peptide synthesis; *Chem. Rev.* **97** 2675–2705

von Dohren H and Kleinkauf H 1988 Biosynthesis of peptide antibiotics; in *The roots of modern biochemistry* (eds) H von Dohren and R Jaenicke (Berlin: de Gruyter) pp 355–368

von Dohren H and Marahiel M A 1990 Specific biosynthesis of peptides in a mixed medium; *Proc. Gen. Ind. Microbiol. Meet,* vol. 2 (Strasbourg) pp 691–702

MS received 28 April 1997; accepted 16 November 1998

Corresponding editors: RAGHAVENDRA GADAGKAR and AMITABH JOSHI