Pure & Appl. Chem., Vol. 62, No. 7, pp. 1437–1440, 1990.
Printed in Great Britain.

Chemical approaches to the re-structuring of proteins

S. Ranganathan*, D. Ranganathan, S. Bamezai, W.P. Singh, D. Bhattacharyya, G.P. Singh, R. Rathi, S. Saini, N. Jayaraman and B.K. Patel

Department of Chemistry, Indian Institute of Technology, Kanpur 208 016, India

Abstract Methodologies for protein re-structuring have been developed enabling site specific side chain modification to either coded or non-coded residues. Procedures for the modification of the protein profile by attachment of either pendants or structural motifs have been delineated. Selectivity has been demonstrated on the basis of modulation of the $-CH_2-$ spacer, which interposes the functional group and the peptide backbone.

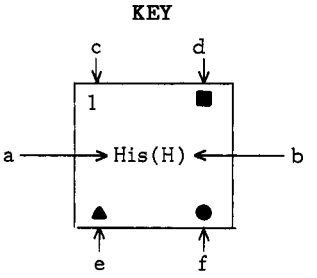
Re-structuring of proteins -protein engineering- can be achieved either through the information system (genetic engineering) or directly by chemical means. The latter methodology is of recent origin and has demonstrated that even in complex enzymes, site specific changes could be accomplished. Whilst much progress has to be made to reach the present versatility of genetic engineering, the chemical approach holds sufficient promise with respect to alteration of peptides and hormones and for the introduction of non-coded amino acid side chains (ref.1). The problems pertaining to admixture of protein by-products inherent in the genetic engineering methodology and superior economical prospects are added incentives to the chemical approach. The power and potential of the chemical methodology finds illustration in recent examples with papain, chymotrypsin and subtilisin (ref.2).

The genetic code, incorporating information relating to the profile of the side chains of amino acids that are of relevance to our endeavours, is presented in Table 1. In view of the fact that an infinite number of α-amino acid structures are possible and that over seven hundred such compounds exist in Nature (ref.3), the choice of 20 amino acids in the code complement is of structural significance. Further, in spite of the immense diversity of the living systems, the fidelity of the code is maintained. We have scrutinized the code from diverse perspectives to provide useful leads.

Table 1 shows that the amino group of lysine- the only coded amino acid that carries a basic residue- is placed farthest from the peptide backbone. A logical explanation

Table 1. The genetic code

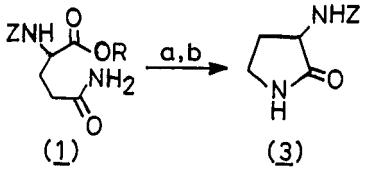
	U	C	A	G	
U	1 Phe(F) ■ 1 Leu(L) ▲	1 Ser(S) ■	1 Tyr(Y) ■ ● Stop	1 Cys(C) ■ ▲ ● Stop 1 Trp(W) ■	U C A G
C	1 Leu(L) ▲	3 Pro(P) ■	1 His(H) ■ ▲ ● 2 Gln(Q)	3 Arg(R) ■ ●	U C A G
A	0 Ile(I) 2 Met(M) ■	0 Thr(T) ■	1 Asn(N) ■ 4 Lys(K) ■ ●	1 Ser(S) ■ 3 Arg(R) ■ ●	U C A G
G	0 Val(V)	1 Ala(A)	1 Asp(D) ■ ● 2 Glu(E) ●	0 Gly(G)	U C A G



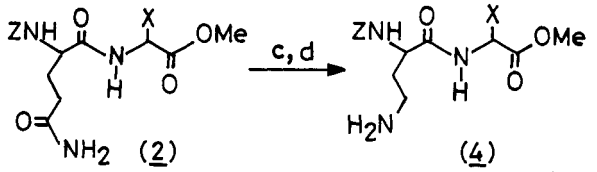
a. 3 letter abbreviation
b. 1 letter symbol
c. Spacer methylenes number
d. Oxidizable side chain
e. participant in templates
f. Capability to form trans-annular bridges.

would be that any lesser spacing would either make activation and consequent loading on the tRNA impossible as a result of lactam formation, or that the proteins arising from incorporation of such units be destabilized because of intramolecular cyclization. Our finding (ref.4) of the spontaneous cyclization of the chain shortened analog of lysine, generated from N-protected glutamine esters (Scheme 1), provided proof to the effect that such analogs can not be supported on tRNA. Of relevance is the demonstration that such units generated from glutamine precursors, in a peptide environment, are quite stable, thus making it possible to transform the neutral Gln side chain to the basic γ-aminoethyl unit in proteins. Another valuable finding of this study is the unreactivity of Asn peptides, under conditions where the Gln analogs are totally transformed ((2)→(4), Scheme 1).

Scheme 1



a. $PhI(TFA)_2$, CH_3CN-H_2O-Py, rt, 3h
b. $NaHCO_3(aq)-CH_2Cl_2$, rt, 3h

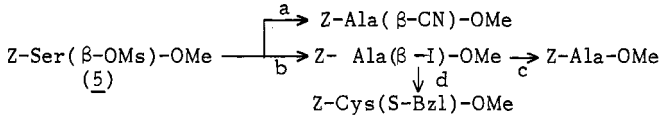


(2a): X = H (2b): = $-CH_2Ph$ (2c): = $-CH_2{}^{i}Pr$
c. $PhI(TFA)_2$, $DMF-H_2O-Py$, rt, 3h
d. $NaHCO_3(aq)$

Our recent studies have shown that the above reactivity difference between Gln and Asn is also found in the side chain methyl esters of Glu and Asp, which also differ by a single methylene spacer (Table 1).

Coded amino acids that have one or more spacer methylenes can be represented as R-Ala (excepting proline!). This family has 15 members or 75% of the code complement. Thus, Ala transfer from chiron from Ser appeared attractive for the generation of coded amino acid side chains in a peptide environment. The feasibility of this approach has been demonstrated with the chemical simulation of the transformation of Ser to Ala, Asp and Cys, that takes place in bacteria (Scheme 2).

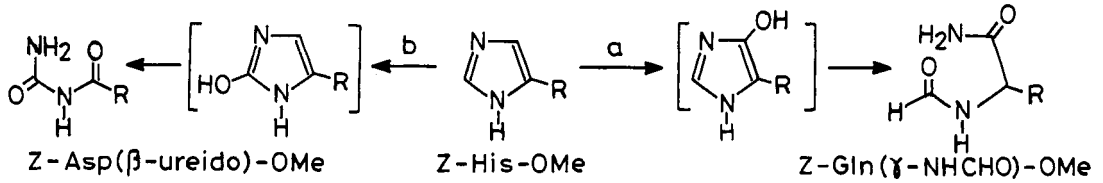
Scheme 2



a. NaCN, 18 crown 6
b. NaI, acetone
c. $(n-Bu)_3SnH$
d. Bzl-SH, NaH, DMSO

Side chain alteration along the metabolic pathway would be a good lead to useful proteins having non-coded amino acid residues. Here, His would emerge as the first choice since its oxidative degradation follows a most unusual pathway, involving hydroxylation of the 4 position followed by hydrolytic ring rupture. We have been able to precisely simulate this path using Z-His-OMe and the novel reagent 4-^{t}butyliodoxybenzene (ref.5). Interestingly, oxidation with Ru VIII generated the novel β-aspartoyl urea side chain, as a result of 2-hydroxylation and cleavage (Scheme 3).

Scheme 3

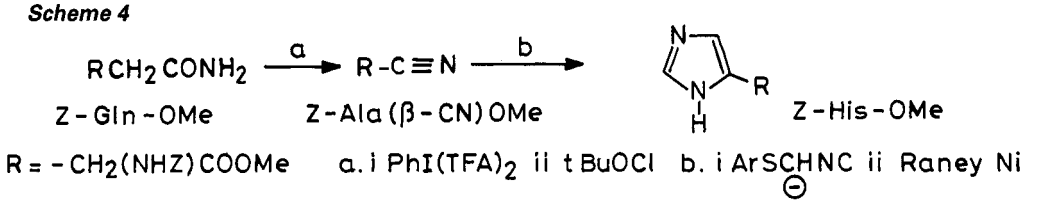


R = $-CH_2CH(NHZ)COOMe$ a: IO_2, $PhCl, \Delta, 7h$ b: Ru^{VIII}, $MeCN-H_2O-CCl_4$, rt

Although not that obvious, the ynamine equivalent of Gln on acceptance of formamide and cyclization can give His. Our endeavours to define Gln and Asn as His equivalents, additionally attractive because of relevance to protein synthesis, are outlined in Scheme 4. The final step has eluded success, although the same has been demonstrated with several non-amino acid ligands (ref.6).

A novel approach to protein engineering, that has been explored in some detail, has as its basis, the oxidative transformation of coded amino acid side chains, since, as Table 1 would show, 11 of the 20 coded amino acid side chains are susceptible to oxidation.

Scheme 4



The strategy here called for a common reagent that can effect, with ease, the oxidation of the recalcitrant Phe as well as the most susceptible Cys or Met and then to bringabout selectivity by control of the experimental protocol. Ru(VIII) generated in situ and recycled using the auxiliary oxidizing agent, sodium metaperiodate, in a two phase system, was used throughout. As a result of extensive experimentation involving 13 of the individual N,C protected amino acids concerned (Gln and Asn in addition to those indicated in Table 1) a selection of dipeptides and some polypeptides, it was possible to arrive at a preference profile shown in Table 2. We can thus control the oxidation of coded amino acid side chains from a profile of general reactivity to one specifically targeted to Trp.

Table 2: Selectivity in Ru(VIII) oxidations

IO_4^-(mmol)	t(h)	pH	Ser	Thr	Phe	Tyr	Trp	His	Lys	Arg	Pro
											poor
15	60	5.5	✓	✓	✓	✓	✓	✓	✓	✓	
8	8	5.5	✓	✓	X	✓	✓	✓	✓	✓	X
8	0.5	5.5	✓	✓	X	X	✓	X	poor	poor	X
8	0.5	3.0	X	X	X	X	✓	X	X	X	X

A number of noteworthy findings were made during the course of the above extensive study, prominent amongst which are: the complete oxidative replacement of the indole ring with a COOH unit (ref.7); the stability of usual N-protecting groups, Ac, Bz, Boc, Z as well as the side chains of Gln and Asn; the non interference of an already oxidized residue with respect to the oxidation of a neighbour (Bz-F-F-F-OMe ⟶ Bz-D-D-D-OMe); total chiral retention; the increased susceptibility of the terminal pro towards oxidation; feasibility to generate non-coded amino acid chains by reaction control (Trp ⟶ N^{ω}-For-Kyn; His ⟶ β-aspartoyl urea/ γ-formamido glutamine).

The methodology is currently in the second stage of development. The oxidation of a number of peptides and proteins is planned and the outcome would be assessed from vantage of yields, purity, specificity, secondary structure retention and changes in reaction profile. Results thus far obtained in this phase, presented in Table 3, are encouraging.

Table 3: Selective Ru(VIII) oxidation of peptides

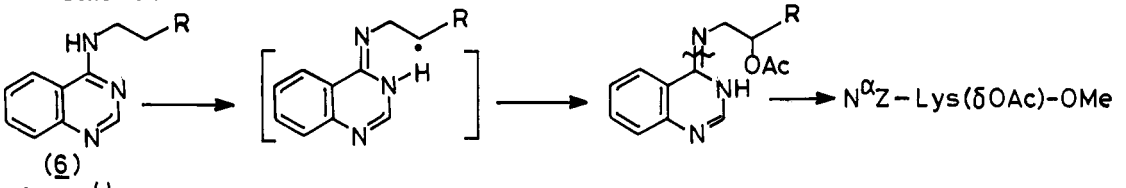
Original sequence	Altered sequence
Z-(LPL<u>W</u>)-OMe (tetrapeptide)	Z-(LPL<u>D</u>)-OMe
Boc-(LVL<u>F</u>LPLAALG)-OBzl (c.lysozyme signal segment)	Boc-(LVL<u>D</u>LPLAALG)-OBzl
H-(GIGAVLKVLTTGLPALIS<u>W</u>IKRKRQQ)-NH_2 (melittin (bee venom protein)	H-(GIGAVLKVLTTGLPALIS<u>D</u>IKRKRQQ)-NH_2
OHC-(VGAL*AV*VV*WL*WL*WL*<u>W</u>)-X gramicidin A(ion channel protein); *=D configuration; X=-$NHCH_2CH_2OAc$	OHC-(VGAL*AV*VV*WL*WL*WL*<u>D</u>)-X

The transformed melittin exhibits CD close to that of the parent, thus indicating retention of the secondary structure. Preliminary studies reveal that the ion conductance

properties of the gramicidin, closely associated with its transport profile, are changed in the transformed product (ref.8).

The understanding and chemical simulation of non trivial post translational protein modifications would provide valuable leads to protein engineering. Most unusual amongst such changes is the hydroxylation of the unreactive δ-methylene of lysine in collagen causing age manifestation. We envisaged the attachment of a 4-quinazoline pendant to the ω-position of N,C-protected lysine, which is in a peptide environment, thus enabling the functionalization of the unreactive position via a six membered transition state, followed by detachment of the pendant from the functionalized substrate (Scheme 5). Compound (6) has been prepared. Whilst halogen exchange to (6) did not succeed, $PhI(OAc)_2$ gave acetoxylated products.

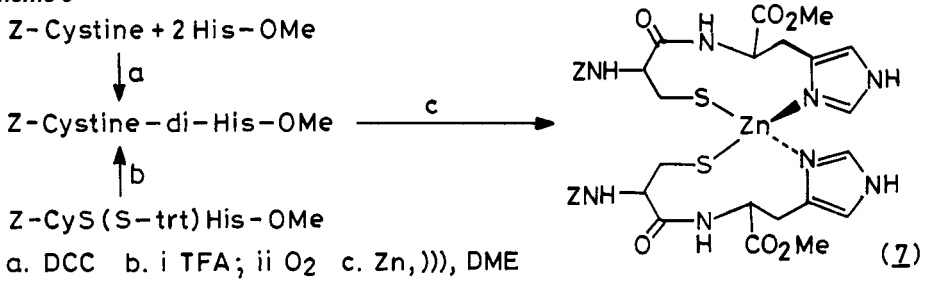
Scheme 5



$N^{\alpha}Z-N^{\omega}$ Quin Lys OMe ($R=-(CH_2)_2CH(NHZ)CO_2Me$)

We here report the synthesis of the novel zinc template (7). The significance of such motifs in DNA recognition is a very recent discovery (ref.9). It is hoped that the attachment of (7) to proteins would give rise to new recognition systems. The preparation of (7) is illustrated in Scheme 6.

Scheme 6



In conclusion, our recent endeavours are based on the present recognition of the protein architecture as an ensemble of structural units (ref.10). As a positive illustration we have shown that the -S-S- bridge could be generated from N,C-protected Ser precursors via activation followed by treatment with the newly introduced tetrathiomolybdate reagent (ref.11). We have also found that such bridges could be snapped with silver carbonate supported on celite to two dehydroalanine units. The generation of salt bridges, from neutral residues, using a combination of the degradative ($PhI(TFA)_2$) and oxidative(RuVIII) methodologies presented here, is also under study.

Acknowledgements Financial assistance from DST, CSIR and IEL is gratefully acknowledged.

REFERENCES

1. R.E. Offord, Protein Engineering, 1, 151 (1987).
2. E.T. Kaiser, Angew.Chem.Int.Ed.Engl., 27, 913(1988); J.B. West, J. Scholten, N.J. Stolowich, J.L. Hogg, A.I. Scott, C.H. Wong, J.Am.Chem.Sco., 110, 3079(1988); Z.P. Wu, B. Hilvert, J.Am.Chem.Soc., 111, 4513 (1989).
3. G.A. Rosenthal, Plant Nonprotein Amino and Imino Acids, Academic Press (1982).
4. S. Ranganathan, D. Ranganathan, W.P. Singh, Tetrahedron Lett., 29, 3111 (1988).
5. S. Ranganathan, D. Ranganathan, S.K. Singh, D. Bhattacharyya, J.Chem.Soc., Chem. Commun., 1887 (1987).
6. S. Ranganathan, W.P. Singh, Tetrahedron Lett., 29, 1435 (1988).
7. S. Ranganathan, D. Ranganathan and D. Bhattacharyya, J.Chem.Soc.,Chem.Commun., 1085 (1987).
8. R. Nagaraj, Centre for Cellular and Molecular Biology, Hyderabad, India, Personal Communication.
9. L.M. Green, J.M. Berg, Proc.Natl.Acad.Sci.USA, 86, 4047 (1989).
10. J. King, Chem.Eng.News, 67, 32 (April 10, 1989).
11. P. Dhar, S. Chandrasekaran, J.Org.Chem., 54, 2998 (1989).