Protein Engineering vol 4 no 7 pp 709–710, 1991

Commentary

Going beyond the code

Protein engineering is all about obtaining proteins not immediately available in Nature. Recombinant methods are, rightly, the most often used for that purpose and normally should be considered as the first option. Recombinant methods (when all goes well) are rapid and the simplicity with which they can be transferred to a new protein system is unrivalled by any other approach. They are, in short, conveniently adapted to the needs of the 'kit generation' of biological scientists and are all the better for that.

On the other hand, methods calling for direct intervention at the protein level have more to offer than mere historical interest: when they work, they work well, and in the most favourable cases with a convenience that it is hard to imagine being bettered in any other way, recombinant approaches included. It is not a coincidence, for example, that much of the human insulin used by the World's diabetics is obtained not by recombinant means, but semisynthetically, by chemical transformation of porcine insulin. Moreover, the advantages of the protein–chemical approach are not limited to being appropriate for bulk production. To remain with insulin, an earlier Commentary (Offord, 1987) discussed the elegant combination of recombinant and semisynthetic methods used at the laboratory scale by Markussen's group (Markussen *et al.*, 1989a,b) to produce a whole range of new insulin molecules. Finally, when mentioning non-recombinant approaches, we should not forget total synthesis as a promising means of producing analogues of small and even medium-sized proteins (e.g. Woo *et al.*, 1989).

Chemical methods are particularly worthy of consideration when substitutions are required that involve either non-coded residues or non-peptide links in the main chain, or where some site-specific post-translational modification, e.g. C-terminal amidation, defeats the expression systems normally used. This general topic, has been reviewed by Offord (1990), and by Rose *et al.* (1991a). However, it was long predicted (e.g. Offord, 1980) that the recombinant approach would also be able to contribute in this area and, as is well known, this has indeed happened in the last couple of years. Some recent papers (Bain *et al.*, 1991; Mendel *et al.*, 1991, Abrahamsén *et al.*, 1991; Rose *et al.*, 1991b) permit us to evaluate the current state of the field, both as regards recombinant and semisynthetic methods.

First, two papers on the recombinant approach, one (that of Bain *et al.*) primarily methodological and one (that of Mendel *et al.*) directed to exploiting the methodology to study a protein structure–function relationship.

The brief review of the origins of the recombinant approach to working outside the code, given in the introduction to the paper by Bain *et al.*, although written entirely from a molecular geneticist's, rather than a protein chemist's standpoint, is well worth reading in itself. But the main interest in the paper is the extension of the authors' previous methodological contribution in this area. Briefly, Bain *et al.* prepared semisynthetic nonsense-suppressor tRNA, chemically acylated with L-3-[^{125}I]iodotyrosine, *N*-methyl-L-phenylalanine, D-phenylalanine or L-phenyllactic acid. The parent tRNA sequence was tRNA^{Gly3}, but having, among other modifications, its anticodon replaced by the appropriate suppressor triplet. Bain *et al.* used principally the amber suppressor but also did some work with tRNAs having the ochre and opal anticodon. The object, of course, was to see if these chemically acylated tRNAs would suppress the effect of the corresponding 'stop' codon, i.e. instead of bringing peptide chain elongation to an end, insert their aminoacyl residue in the chain, and permit synthesis to continue until one of the other possible 'stop' signals was encountered. To study this possibility in a completely controlled way, the authors constructed an mRNA which, in a cell-free system, would give an octapeptide if the target 'stop' codon were allowed its normal function, and a hexadecapeptide (with the chemically introduced residue in the ninth position) if the 'stop' signal was suppressed.

Careful study of the behaviour of all these components in reticulocyte lysates gave firm evidence that things had worked as planned. Using the iodotyrosine derivative with the amber suppressor, as much as 65% of the translated sequence appeared to be the wanted hexadecapeptide. The *N*-methyl amino acid was incorporated with about the same efficiency, whereas the D-amino acid was not incorporated at all. The incorporation efficiency of phenyllactate (thus, with an ester bond between residues 9 and 10 of the peptide chain) was 46%, a truly remarkable result.

The other leading group in this field (Mendel *et al.*, 1991) have also further developed their early work. These authors have used their version of the same general approach as that just described to construct a mutant of T4 lysozyme in which Asp20 should be specifically esterified by *o*-nitrobenzoic acid. The reason for doing so is that Asp20 forms part of the catalytic site, in that it is thought to stabilize a carbocation intermediate by electrostatic interaction. Even leaving the question of its bulk aside, the nitrobenzoic ester would clearly not be able to do this, and the mutant should be inactive. But, since *o*-nitrobenzoic esters are photolabile, irradiation of the mutant enzyme should restore activity. A suppressor tRNA was prepared, chemically acylated with β(*o*-nitrobenzoyl)-Asp, and protein biosynthesis carried out in a cell-free system, directed by a plasmid encoding either the enzyme with the appropriate changes to the codon corresponding to position 20 or the wild-type sequence. Everything worked as hoped for: an inactive mutant enzyme was produced, identified with suitable controls by radioautography of gel electrophoretograms, at several tens of μg/ml, and shown to be photoactivatable. This finding constitutes useful support to the original hypothesis concerning the enzyme's mechanism, and although the evidence does not yet meet the standard required in protein-chemical exercises of this type (direct analytical demonstration that the substitution is the one intended, and that it really is at the intended site, and only there), there is good inferential evidence that all is well.

In both of these papers the quantity of product is small, and in Bain *et al.* (1991) identification and characterization is quite indirect, relying on the detection of relatively modest quantities of radioactivity. In Mendel *et al.* (1991), detection depends on radioautography and enzyme assay. That more material is not available is due to the current need to work in cell-free rather than *in vivo* systems, and perhaps also to the great complexity (greater than in most modern protein semisyntheses) of the chemical work required to make the acylated suppressor tRNAs. Neither of these two papers seeks to conceal these difficulties, and in the words of Bain *et al.* 'we are currently expanding our research to address these two issues'. There is quite a way to

go yet, but given the recent history of molecular genetics, one can hardly be other than optimistic.

Meanwhile protein-based methodology continues to make progress. Most current work is based on the stratagem, originally made popular in the field by the work of Laskowski (1978), that proteases can be made to work in reverse (i.e. as ligases) even with large protein substrates, while retaining much of the specificity of action. Thus, to take but one of the many enzymes now used for this purpose, trypsin normally cuts on the C-terminal side of lysyl and arginyl residues, but can be made to ligate peptides with C-terminal Lys or Arg to a wide variety of other compounds, including amino-acid derivatives and peptides. None of the complicated side-chain protection schemes of conventional peptide synthesis are needed, since the preference of proteins for the substituents of the α-carbon will normally mean that there is no danger of reactions involving side chains.

Space does not permit us to develop the theory of this method here [see Laskowski (1978) and Kullmann (1987) for a complete discussion]. It is perhaps sufficient to say that there is a kinetic approach to the problem and a thermodynamic one. The kinetic approach exploits the fact that, under appropriate conditions, ligation products are transiently formed on the way to hydrolysis, and can be trapped. The thermodynamic approach relies both on mass action to push the reaction in the direction of synthesis, and on changes to the equilibrium constant itself, brought about by altering the dielectric constant of the medium by the addition of a water-miscible, uncharged organic solvent (e.g. glycerol, various diols, dimethylsuphoxide, etc). Since the standard free energy of reaction does not greatly favour hydrolysis, if at all, these quite simple manoeuvres often permit rapid and clean ligations.

Of two recent papers, one (Abrahmsén *et al.*, 1991) is, formally speaking, related to the kinetic approach, while the other (Rose *et al.*, 1991b) deals with the thermodynamic approach. Abrahmsén *et al.*, recognizing the potential of recombinant techniques to tailor proteases to function as better ligases, have engineered an analogue of subtilisin which ligates efficiently in water. They present to the enzyme the substrate that donates its α-COOH to the new peptide bond not as a free protein or peptide, but as the α-ester. In the mutant enzyme, to improve esterolysis while reducing amidase activity, the catalytic serine has been replaced by cysteine, something that, as Abrahmsén *et al.* point out, has been done previously by other workers. However, in the present case, crowding caused by the greater size of the -SH group relative to the OH-group was alleviated by mutating a second site, one helix turn away from the catalytic residue (proline in the wild type, alanine in the mutant). As hoped, the mutant proved to be a quite efficient esterase but a very inefficient amidase. Thus, ester substrates readily form the customary acyl-enzyme intermediates, whereas there would be much less tendency to cleave other peptide bonds. As in most cases of ligation by proteases, whether kinetic or thermodynamic, the new peptide bond is formed by attack on the acyl-enzyme intermediate, not by water (which is what is required for hydrolysis) but by the amino group of the second component destined to complete the peptide bond. In the present case, not only did the mutant enzyme perform to specification in trial ligations, but the authors even carried out further mutations at the specificity site to give three new variants, each having different preferences in the P_1 site.

It would have been interesting to see what degree of ligation would have been obtained, and under what conditions, using a non-esterified substrate. The principle of microscopic reversibility shows that such substrates, too, must give the acyl intermediate. Of course, while in most thermodynamically driven ligations the free α-COOH group actually renders the process more rapid and efficient than one involving the prior removal or a substituent of the α-COOH, such a result would scarcely be expected in the present case.

This is a very detailed paper, including the 1.5Å X-ray structure of the principal mutant, and deserves to be read in detail.

It would not be fair to give too extensive a commentary on a paper coming from this Laboratory, but the reader might like to glance at Rose *et al.* (1991b) to see some of the things of which the thermodynamic approach is now capable. This work, directed at finding ways of using ligation for the site-specific modification of antibodies and other target-directed proteins, shows, in a simple protein system, how readily non-coded functional groups (aldehydes and hydrazides) can be introduced at the C-terminus of a peptide chain. Chosen for their potential value in site-specific conjugation to other molecules, these groups have a good reactivity in mild aqueous conditions, while nonetheless having a profile of chemical response quite unlike that of any amino acid side chain.

References

Abrahmsén,L., Tom,J., Burnier,J., Butcher,K.A., Kossiakoff,A and Wells,J A (1991) *Biochemistry*, **30**, 4151–4159.

Bam,J.D., Diala,S.E., Glabe,C G., Wacker,D.A , Lyttle,M H , Dix,T.A and Chamberlin,A.R. (1991) *Biochemistry*, **30**, 5411–5421

Kullmann,W. (1987) *Enzymatic Peptide Synthesis* CRC Press, Boca Raton.

Laskowsky,M ,Jr (1978). In Offord,R E. and Di Bello,C. (eds). *Semisynthetic Peptides and Proteins*. Academic Press, London, pp 255–262.

Markussen,J., Hougaard,P., Ribel,U , Sorensen,A R. and Sorense,E. (1987a) *Protein Engng*, **1**, 205–214

Markussen,J , Diers,I., Engesgaard,A., Hansen,M T , Hougaard,P., Langkjaer,L., Norris,K , Ribel, U., Sorensen,A.R and Sorensen,E. (1987b) *Protein Engng*, **1**, 215–233

Mendel,D., Ellman,J.A. and Schultz,P.G. (1991) *J Am. Chem. Soc* , **113**, 2758–2760.

Offord,R.E. (1980) *Semisynthetic Proteins*. John Wiley Chichester and New York

Offord,R E (1987) *Protein Engng*, **1**, 271

Offord,R.E. (1990) In Ilook,J.B. and Poste,G. (eds). *Protein Design and the Development of New Therapeutics and Vaccines*. Plenum Press, New York, pp. 253–282.

Rose,K., Fisch,I , Vilaseca,L A., Meunier,A., Werlen,R., Pochon,S., Jones,R.M.L., Dufour,B., Rossito,I., Regameyn,P.-O. and Offord,R E (1991a) In Epston,R. (ed.), *Innovation and Perspectives in Solid Phase Synthesis and Related Technologies* SPCC, Birmingham, in press.

Rose,K., Vilaseca,L A , Werlen,R., Meunier,A , Fisch,I., Jones,R.M.L. and Offord,R.E. (1991b) *Bioconjugate Chem* , **2**, 154–159.

Wood,D D , Clark-Lewis,I , Chait,B T. and Kent,S B (1989) *Protein Engng*, **3**, 29–37.

Robin Offord
Département de Biochimie
Centre Médical Universitaire
Université de Genève
9 avenue de Champel
1211 Genève 4, Switzerland