

Translational Modification of Proteins

This has been listed as a lecture on site-specific mutagenesis; but a broader title is appropriate, covering a number of ways that proteins can be modified through the normal biosynthetic mechanism. These can be broadly described as mutagenesis, i.e. changing the DNA sequence of the gene to change or remove a residue in the final protein; suppressor mutagenesis, in which even an unnatural amino acid of any structure can be incorporated into the protein, but so far only in *in vitro* synthesis, which is limited in amount; and analog replacement, which involves fooling the tRNA synthetase and incorporating an analog of the natural amino acid, though generally one very similar to it.

'Ordinary' mutation was a reason to study protein structure and function before it was a method - consider for instance sickle cell hemoglobin, which is liable to polymerization and distortion of red blood cells, particularly when the blood pH drops a little. Hemoglobin was in fact the first protein whose mutations were widely studied, since a single individual can readily donate enough hemoglobin for almost any study, and there is a broad screening mechanism, all the general practitioners in the world examining anemic patients. Hundreds of mutations of human hemoglobin have been described and characterized - it is I think the only protein to have its own journal!

We should think about what mutation can do to a protein. Firstly and worstly, the protein may not fold up correctly; no functional protein will be formed, and cellular mechanisms that degrade and remove denatured proteins will remove it. Monomers of the protein may fold correctly, but fail to assemble to active oligomers. Even assembled oligomers may be inactive; in these two cases we may observe what is called "cross reactive material", i.e. protein which reacts with antibody to the native protein but is inactive. The protein may be **temperature-sensitive**, active but denatured at a lower temperature than the wild type; this may be interesting in defining what holds it together, and it has long been a useful way of finding mutations of an essential gene, because you can still grow the organism at a lower temperature. On the other hand, there is much interest in finding mutants which are *more* temperature-stable, usable at higher temperatures; the detergent industry and the cellulose-to-glucose people want this. Mutations making the protein more stable in organic solvents have also been sought. Finally and most interestingly there are mutants which have normal stability, but reduced, or even enhanced, activity; this is an approach to understanding the chemical basis of the activity of the protein. We shall see in a later lecture a very complete example of the use of mutant forms of an enzyme to define how it catalyzes a reaction.

We can also consider kinds of mutation. There are 'nonsense' mutations, which disrupt a large part of the protein; frame-shifts which cause a completely erroneous amino acid sequence beyond the mutation - these are not interesting to us, though they are to Bruce Ames's mutagenesis assay - and mutations to the termination codons UAG, UGA and UAA, which terminate the protein at this point. It has long been known that there are 'suppressor' mutations which suppress this effect, producing tRNAs which are charged with a natural amino acid but have an anticodon pairing to a termination codon, so that they can introduce the amino acid they bear at that codon and allow the protein to continue. Of course they can't be completely efficient, else proteins would never stop at that codon, but production of *some* of the protein is generally enough, as long as it is still functional with the introduced amino acid at that point. As we shall see, they very often are.

Some protein functions can be investigated efficiently by **deletion mutagenesis**, removal of a portion of the polypeptide chain. For instance, alanyl- tRNA synthetase is natively 875 amino acids long; it could be shown using deletion mutants that essentially only the first 385 amino acids were needed for alanine adenylate formation (at half the specific activity of the native protein); the first 461 amino acids were needed for transfer of alanine to tRNA; for assembly of the native tetrameric form of the protein residues 699 to 808 were needed (the monomer is active, but it is the tetramer which

regulates expression of its own gene by binding to DNA.) Deletions can be achieved by cutting the cloned gene with restriction enzymes at conveniently located sites (but there may not be conveniently located sites), or by the same methods used for point mutation, depending on a synthetic oligonucleotide binding to the cloned gene in a single-stranded vector. Deletions can be used somewhat blindly, like chemical modification, to try to locate important functions in the polypeptide chain, since removing, say, 20 amino acids at a time from a 200-amino acid protein would require only 20 mutants; one could scan the entire protein for important parts, and narrow down the focus when one had found an important stretch (without which the protein is inactive). However, a deletion in the middle of a folded domain - a portion of a protein which folds into the native structure on its own - is likely to prevent proper folding, unless carefully located so that it only shortens a loop; removal of a piece which stretches across the protein is likely to disrupt folding quite completely, resulting in a protein which is inactive, not properly folded, and probably rapidly removed by proteolysis. Deletion mutagenesis is most appropriate at the N- and C-termini of a protein, removing more and more until one has a protein which is inactive even though largely native in folding (if you are lucky). It is particularly appropriate for determining how much of an N-terminal leader sequence is necessary for proper direction of a protein into or through membranes.

More usually, we are interested in **replacing** one or more amino acid residues with other amino acids, by molecular biological methods. This is a very powerful way of investigating the roles of individual amino acids in function - not only catalytic activity, but stability, subunit interaction, relationship of domains in the protein, delivery to specific compartments in the cell. In favorable cases, when the mutated gene can be substituted for the wild type in the original organism, and the protein is not required for survival under all conditions - for instance, proteins of the photosynthetic apparatus in cyanobacteria, which can be grown heterotrophically - one can look at the effects of the mutation on the protein *in situ*, in its normal environment in the cell, without having to reconstruct a protein complex *in vivo*, as would be the case with a chemical modification.

One can here use **random** mutagenesis, generating many mutants without looking at any position specifically. This can be done either with chemical mutagens or UV light acting on the host organism - the traditional method - or with PCR (polymerase chain reaction) under conditions where it makes lots of mistakes, ligating the product into a vector and putting it back into a host organism. The advantage of PCR mutagenesis is that you are only mutating the gene you are making a copy of, and only getting changes of one base and one amino acid, not deletions, cross-linking, etc. One must then have a good **screen**, way of looking for interesting mutants - at a temperature where the original enzyme isn't active, or in presence of an inhibitor. One hopes to isolate an improved version of the protein and then do another round of mutation.

Site-directed mutagenesis replaces one amino acid **at one specific position** in the polypeptide chain, a position of your choice, or sometimes several amino acids in a row, with other natural amino acids, without changing the length of the polypeptide chain. One can change size without changing polarity, or polarity without changing size, as in changing aspartate to asparagine. One can substitute a smaller amino acid for a larger, and be sure that the effect is due to the chemical change, not to increased bulk of a chemically modified residue. One can usually be sure that the effect results from the specific change made, not from modification at other, more distant residues. One can attack any amino acid, chemically reactive or unreactive, buried or exposed; one can make a wide variety of changes. In a seminar at Waksman, I heard about substitutions in staphylococcal nuclease, which included a lysine for a buried valine; the lysine went into the inside of the protein too, even though that meant that the lysine remained uncharged down to the lowest pH where the protein was at all stable, $pK_a \leq 5.8$; from this it could be calculated that the dielectric constant inside the protein was 12 or below, vs. 81 in H_2O . *But* one is normally limited to the natural amino acids as replacements.

Tryptophan, which fluoresces and is therefore a natural 'reporter group', is a particularly useful natural amino acid to insert, where there is room for it; the fluorescence is sensitive to the group's surroundings and often changes when the protein changes conformation. A general approach (Atkinson et al., Biochem. Soc. Transactions **15**:991-3 [1987]) is to first convert the natural tryptophans to tyrosines, to remove 'background' fluorescence - this converts to a smaller amino acid of similar polarity, so it is generally acceptable; then to reinsert single tryptophans at specific positions, either original positions or others which are expected to move in a conformational change. Changes in the fluorescence (intensity, polarization, energy transfer from or to a bound group such as NADH) can then be monitored during the fluorescence change. [For instance, a *Bacillus stearothermophilus* lactate dehydrogenase was quite functional with all its trp converted to tyr. Substitution of lys¹⁰⁶ by trp made it less heat-stable, due to loss of a surface ion pair, but put a reporter in the 'coenzyme loop' which in X-ray crystallography structures (of dogfish and pig LDHs) moves 13 Å to bury the nicotinamide ring of the coenzyme in nonpolar amino acids. The mutant trp¹⁰⁶ shows a 15% increase in fluorescence on rapid mixing with NAD⁺ and oxalate (to form an E·coenzyme·inhibitor complex) due to reduced solvent quenching. The rate of movement could be monitored in a stopped-flow spectrofluorimeter, $k = 2.7 \text{ s}^{-1}$ at -16°, 250 s⁻¹ at 25°, the same rate as that of a single turnover of the enzyme (mixing E·NADH with pyruvate, 2.1 s⁻¹ at -16°).]

However, site-specific mutagenesis requires as an essential condition that the gene - or at least the cDNA - for the protein in question has been cloned and can be produced in *E. coli* or other host in quantities sufficient for purification and study of the protein. This is now routine, indeed proteins can be produced in usable quantities much more easily by such a procedure than by purification from their natural source. The only problem is if the protein undergoes significant post-translational modification such as glycosylation, hydroxylation, cross-linking, etc., these are not likely to occur in *E. coli*, though they may if the protein is produced in yeast, insect cells, or Chinese hamster ovary cells, all now used for production of recombinant proteins. Having the gene or cDNA cloned of course means having the DNA sequence and thus that of the protein.

I shall not talk about *how* you do site-directed mutagenesis, since this is a molecular biology procedure, described in Fersht, and you can buy kits to do it.

Best use can be made of site-specific mutagenesis only when the three-dimensional structure of the protein, and its complexes with other molecules such as substrates and products, is known in considerable detail by X-ray crystallography, or possibly in small proteins by solution nmr. It is such knowledge that tells you what residues to modify - because they are at the catalytic site, or other site one wants to investigate structure-function relationships at. One usually cannot try all possible mutations of all positions in a protein, though one can by other means (such as chemical mutagenesis) introduce random point mutations, and given a screening mechanism select those with the most interesting effects. Thus a contrast can be drawn to chemical modification, which though crude compared to site-specific mutagenesis, and liable to modify more than one residue, does have the ability to identify important residues in the absence of detailed structural information, because it is easier to carry out a chemical modification experiment and see whether the protein's function is affected than to construct mutants with each of a specific residue modified and determine whether function is affected. Again, the primary sequence of the protein is needed to make this information most useful. Site-specific mutagenesis can be used to change amino acids identified as important by chemical modification or other chemical or physical means, when the 3-D structure is still lacking. The exam paper I will be giving out is an example.

[Example: the manganese-stabilizing protein of the photosynthetic center of the cyanobacterium *Synechocystis*, which Dr. Zilinskas' lab was studying. There is evidence that the Mn ligands are carboxyl groups, not surprising, and one could systematically change each aspartate or glutamate to the corresponding amide and look for decreased stability of Mn binding and perhaps decreased function in

photosynthesis. However, one would like to narrow down the possibilities a bit more. One might in principle do this by chemically modifying carboxyl residues of the protein in isolation, showing that the modified protein cannot bind to the photosynthetic complex, and then carry out the modification reaction when the protein is bound to the photosynthetic complex and the carboxyls interacting with manganese are buried and unreactive. One would then seek to identify which carboxyls are modified in the free protein and not in the complex; these, hopefully no more than three or four, would then be the targets of specific site-directed mutagenesis. In this case the host organism is easily transformed by native DNA of an appropriate vector, and the wild-type gene replaced by specific recombination, so that a clone with the mutant gene can be isolated and grown heterotrophically even if photosynthesis is disrupted.]

Another approach is called 'alanine scanning mutagenesis'. While I never find papers on this when I want to, I believe the idea is replacing blocks of five or so amino acids of the normal sequence with alanine, and looking for effects on activity. Since alanine is smaller than any other amino acid except glycine, this cannot prevent the folding of the protein (unless you replace a critical glycine at a bend), but will remove any chemically significant side chain interactions. Thus you can look for criticality of all amino acids in a protein with a finite number of experiments, although the modified protein may be seriously less stable than the wild type if several important interactions were removed.

When the 3-D structure of the protein and (in the case of an enzyme) its complexes with substrates are known, one can modify specific residues in the active site which are believed to interact with the substrate and see how important they are, what their role is. We shall see the most thorough example of this in the case of tyrosyl-tRNA synthetase. One also hopes that modifications will be discovered which will make an enzyme **more** active, by higher V_{\max} or lower K_m , or more specific or stable; for instance, mutants of lysozyme with an added cysteine, which forms an additional disulfide bond, are more stable than the wild-type.

I mentioned what I called "suppressor mutagenesis". The idea takes advantage of suppressor tRNAs which have anticodons complementary to the terminator codons UAG, UGA and UAA, but usually UAG, the so-called "amber" codon. Whatever amino acid the tRNA is charged with is then incorporated into the protein at that point. One use of this has been made by Jeffrey Miller at UCLA, studying the *lac* repressor of *E. coli*. He has obtained 5 natural suppressor tRNAs and made nine more - all one has to do is mutate the anticodon triplet of the gene for a natural tRNA. He (and a number of people in his lab) inserted the UAG codon at each position in the *lac* repressor gene from position 2 to 329 - deletions have shown that residues beyond 330 are necessary only to assemble the native tetramer, but the dimer form is active. They combined these 328 mutants with insertions of 13 possible amino acids by the appropriate suppressor tRNAs, generating over 4000 mutants. I haven't read the paper with the methods, so I don't know how they screened all these, but at least it was easier than constructing 4000 separate mutations. They found two regions in the protein, roughly 5 through 60 and 239 through 293, where mutations were not well tolerated; but outside those regions most positions would accept almost any amino acid. Ninety-three of the 328 sites (28%) would accept anything; another 51 would accept all but one tried - usually proline was the one rejected. Another 48 would accept conservative substitutions, one small a.a. for another or any hydrophobic amino acid. Stretches of five to 14 amino acids where anything was accepted are characterized as "spacer" regions between key features, and would accept substitutions of runs of alanine for them - but not deletion, the length remained important.

This approach was invented a very long time ago. The element selenium has been shown to be present as selenocysteine in a small number of enzymes of eukaryotes, prokaryotes and archæa; the best known cases are a selenopolypeptide of *E. coli* formate dehydrogenase and mouse glutathione peroxidase; others are a component of glycine reductase in *Clostridium sticklandii*, where Thressa Stadtman originally discovered this substitution, mammalian thioredoxin reductase, and the enzyme which converts tetraiodothyronine to triiodothyronine, i.e. the hormone thyroxine. The genes contain an in-frame TGA (opal) termination codon at the position where selenocysteine is found in the protein.

Genetic studies of *E. coli* mutants deficient in selenium metabolism and unable to convert formate to CO₂ identified a number of genes, one of which, *selC*, has as product a special tRNA, called tRNA^{Sec}, with a UCA anticodon matching the opal codon. This tRNA is charged with serine by seryl-tRNA synthetase. The product of the *selA* gene catalyzes elimination of water from the serine, giving an enzyme-bound aminoacryloyl-tRNA, to which selenium is transferred from selenophosphate, the actual selenium donor, to form selenocysteinyl-tRNA. The *selD* product forms selenophosphate, from HSe⁻ or a selenium carrier protein, and ATP. Selenophosphate is also the selenium donor for 5-methylamino-methyl-2-selenouridine, which is found in tRNA^{Lys} and tRNA^{Glu} of *E. coli*.

The actual incorporation into protein requires several factors in addition to the charged tRNA and the UGA codon: a stem-loop structure in the RNA, called SECIS, which has five invariant residues, including two non-Watson-Crick base pairs A-G; a binding protein called SBP2, and the specific chain elongation factor EF-Sec, product of the *SelB* gene, similar to the elongation factor Ef-Tu but specific for incorporation of selenocysteine from selenocysteinyl-tRNA^{Sec}. Like Ef-Tu, it binds GTP and better GDP, and it binds the charged tRNA^{Sec}, which Ef-Tu doesn't, and competes with the release factor 2 which terminates peptide formation at the UGA codon. The SECIS loop is immediately downstream of the UGA codon in bacteria, but in the 3'-untranslated region in eukaryotes and archæa. SBP2 binds to the SECIS element, EF-Sec binds to SBP2, and the charged tRNA^{Sec} binds to EF-Sec; in bacteria the SBP2 and EF-Sec functions are combined in a single protein SELB. This puts the charged tRNA in position to be handed over to the ribosomal A site and the selenocysteine to be incorporated. Reference: D.H. Hatfield & V.M. Gladyshev, *Mol. Cell. Biol.* **22**:3565-3576 (2002).

Mammals have a similar opal suppressor tRNA^{Sec} which can be charged with serine. The serine OH can be phosphorylated and apparently phosphoserine can thus be incorporated into proteins. The same tRNA carries selenocysteine, but as of this paper proof was lacking that phosphoseryl-tRNA is an intermediate in formation of selenocysteinyl-tRNA.

A further advance has been made by Peter Schultz and co-workers at Berkeley. He uses a cloned gene, in the published case for β -lactamase, with a nonsense codon TAG replacing a natural codon, in his case for Phe-66. They had a large paper in *Science*, vol. **244**: 182-187, April 1989. Yeast tRNA^{Phe} has its anticodon chemically replaced with CUA, making it an amber suppressor tRNA; this process produces tRNA lacking the last two nucleotides, pCpA, where the amino acid is normally put on. Instead, in this case an *o*-nitrophenylsulfenyl-amino acid is chemically attached to the 2' and 3' ribose hydroxyls of the dinucleotide (with the NH₂ of the cytosine blocked to prevent adding there). This charged dinucleotide is then deblocked and ligated to the pCpA-less tRNA_{CUA}, yielding a charged tRNA which will insert the amino acid of choice, **not necessarily a 'normal' amino acid**, at the TAG codon. This is used in *in vitro* protein synthesis. With tRNA_{CUA} acylated with [³H]phe they produced 5.5-7.5 μ g/ml active β -lactamase, which they were able to purify to homogeneity. They also charged the tRNA with D-phenylalanine, *p*-nitrophenylalanine, homophenylalanine (one more CH₂ between the α -carbon and the ring), *p*-fluorophenylalanine, 3-amino-2-benzylpropionic acid (an analog of Phe with the positions of the NH₂ group and the benzene ring reversed) and 2-hydroxy-3-phenylpropionic acid. D-Phe, 3-amino-2-benzylpropionic acid and 2-hydroxyphenylpropionic acid predictably yielded no synthesized enzyme (monitored by [³⁵S]-met incorporation). *p*-Nitrophenylalanine substitution gave an enzyme with the same K_m and *k*_{cat} about half that of the wild-type enzyme; *p*-fluorophenylalanine gave an enzyme with the same K_m and a slightly **higher** *k*_{cat} than the wild type; homophenylalanine gave an enzyme with a slightly higher K_m and a *k*_{cat} about one-sixth that of the wild type. A tyrosine-containing mutant, prepared by standard site-directed mutagenesis, had a slightly lower K_m and a *k*_{cat} half that of the wild type. The *p*-nitrophenylalanine and homophenylalanine mutants were too unstable to purify, as was an alanine-containing mutant.

This technique thus makes it possible to introduce unnatural amino acids at a specific site, although the amount of protein that can be produced is very small and the mutant protein so produced may not be stable. The authors state that "sufficient protein can be purified to characterize the catalytic constants and specificity of the mutants, to carry out limited mechanistic and mapping studies, and to probe protein structure with techniques such as ESR and fluorescence spectroscopy." They hope to be able to make milligram amounts. There is great interest in being able to do this *in vivo* and thus make a lot more protein, but that will require mutating a tRNA synthetase as well, and also ensuring that it does *not* charge the tRNA with a normal amino acid. [One step that has been used by another group, who express a protein in egg cells of the frog *Xenopus* into which they inject mRNA with the termination codon and the suppressor tRNA, is to use a natural suppressor tRNA^{Gln} from the unicellular alga *Tetrahymena*, which recognizes the UAG codon but isn't charged by *Xenopus* tRNA synthetases.]

Another way this has been done is by using four base codons, replacing rarely used normal codons (Hohsaka et al., *Biochemistry* **40**:11060-11064 (2001). I don't know how they charge their tRNA, but they add an extra base to the cloned gene to create a four base codon (preferably removing any similar 3 base codons elsewhere in the gene), and a corresponding base to the tRNA. Thus if the ribosome incorporates the original amino acid, the message is thereafter is out of frame and soon runs into a termination codon; only if the analog amino acid is incorporated is the full length protein synthesized. They say this is more efficient than the suppressor system, and allows incorporation of unnatural amino acids at several points in the same protein.

Amino acid analogs can in principle replace natural amino acids in proteins, *if* they can fool the activating enzymes, which is where all specificity is found; if you can get the amino acid on the tRNA you can incorporate it into protein. But these enzymes have evolved ways of being very specific, beyond what you can expect for the difference in binding between isoleucine and valine, for instance. Analogs which can be incorporated generally are very similar in structure to natural amino acids. You hope that the analog-containing protein will still be active but have interestingly altered properties. However, most substitutions either give fully active proteins (especially in the case of hydrophobic substitutions such as *p*-fluorophenylalanine, 3-fluorotyrosine, 5-fluorotryptophan, 7-azatryptophan, δ -trifluoroleucine) or inactive, unassembled proteins, especially with azetidine-carboxylic acid, the four-membered-ring analog of proline, and 1,2,4-triazolealanine, a histidine analog. Other substitutions include norleucine, ethionine, selenomethionine and even telluromethionine for methionine - the last is useful for X-ray crystallography.

Analog incorporation is best carried out into a specific inducible protein in a microorganism which cannot make the amino acid in question, and even lacks salvage pathways for resynthesis of the amino acid (such as tryptophanase for tryptophan). The organism is grown under non-inducing conditions with a limiting amount of the normal amino acid, to a level perhaps two generations short of what other constituents of the medium would allow in presence of an excess of that amino acid; this is because most analogs allow two generations or less growth, because *some* protein is likely to be seriously affected, as are control mechanisms. When growth stops due to exhaustion of the natural amino acid, the analog is added, together with the inducer of the protein desired, which may be cloned behind the *lac* promoter. Growth then continues for two generations or even not at all, with essentially all of the desired protein being produced under conditions where only the analog can be incorporated.

One use of this technique is incorporation of labeled amino acids as reporter groups, especially in nuclear magnetic resonance. In principle one can use nmr to look at the surroundings of any atom with an odd number of nucleons - ¹H, ¹³C, ¹⁹F, ³⁵Cl, ³¹P. Looking at all the protons, or even all the C atoms by natural abundance ¹³C nmr, is just too many peaks, at least for all but the smallest proteins - with one exception: the proton at the 2 position of the imidazole ring of histidine, between the N atoms, is well resolved from all other resonances, and proteins with as many as 11 his have been examined and specific resonances separated out. These can be looked at at different pH and the individual pK_as

determined, and correlated with the structural surroundings when the 3D structure is determined. Nmr studies have suggested that the imidazole anion, which in free histidine in solution has a pK_a above 13, may occur in proteins with a pK_a as low as 8 and play an important role in enzyme mechanisms.

A better way to look at other amino acids is to incorporate a synthetic amino acid, enriched in ^{13}C at a specific position, up to 90%, by the means described above. This has the advantage that an amino acid natural except for a ^{13}C nucleus does not affect growth and biological regulation. Unfortunately, the ^{13}C nuclear resonance is generally sensitive only to the nature of the atoms to which it is immediately bonded, not to the general surroundings, unless these include a paramagnetic atom such as Fe^{+++} or Mn^{++} , with which interesting results have been achieved. In principle one could grow a microorganism on a fully deuterated medium (not too difficult, deuterated rats have been grown) with addition of an amino acid specifically labeled with ^1H at one position, deuterated at other positions - this is the expensive part. In practice, ^{19}F -fluorotyrosine, which is not quite a natural amino acid, has been most useful.

A different and interesting approach is summarized in a paper by Proudfoot *et al.*, *J. Biol. Chem.* **264**:8764-8770 [1989]. In many cases two pieces of a protein will associate non-covalently in the native conformation, even though one natural peptide bond is not present; the classic case is ribonuclease S, with the peptide bond between residues 21 and 22 cut by subtilisin. The fragments can be separated, and the 1-21 peptide, called the S-peptide, has no structure on its own, but re-forms a helix on associating with the remainder of the RNase, called the S-protein; this process can be followed by binding of the inhibitor 2'-CMP, with a change in absorbance at 254 nm on binding. This process can be used to assemble ribonuclease S, which is active, with a synthetic or semisynthetic S-peptide, which could incorporate unnatural amino acids, such as a pyrazole-alanine, with $pK_a \approx 2.5$, at position 12 instead of a histidine ($pK_a \approx 7$).

Cytochrome *c* is another protein of which fragments assemble to form complexes of near-native conformation and activity. Notably, the two cyanogen bromide fragments 1-65 and 66-104 not only complex, but resynthesize the peptide bond between them. Cyanogen bromide cleavage leaves the original methionine residue as homoserine lactone, which is a mild activation of the C-terminal carboxyl group of the 1-65 peptide. When the two fragments are held together in a complex, the carboxyl and amino ends are held together so well that peptide bond formation occurs spontaneously. This reaction has been used to couple natural 1-65 with synthetic versions of 66-104 containing analogs, and natural 66-104 from other species.

Proudfoot *et al.* generalized this process considerably, though still working with cytochrome *c*. They cleave this protein, acetimidylated so that cleavage occurs only at arg-38, with trypsin. They then use reverse proteolysis with trypsin or other serine protease in 80-90% butanediol to add an amino acid ester, preferably a 2,4-dichlorophenyl ester, as residue 39. Meanwhile the amino terminal residue of the 39-104 fragment is removed by Edman degradation. This 40-104 fragment is then combined with the 1-39 dichlorophenyl ester in phosphate buffer pH 7; the residue 40 NH_2 displaces the 2-,4-dichlorophenol to reform a peptide bond, and the complete protein is separated from the fragments by Sephadex G-50 gel filtration. Yields are typically about 50%. A further change can be made by adding N^ϵ -Boc-lys-*t*-butyl ester as residue 39, removing the protecting groups, and adding an amino acid dichlorophenyl ester as residue 40. This could then be combined with a 41-104 fragment generated from 39-104 by two cycles of Edman degradation.

This approach could be used in any case where a proteolytic cleavage can be made at a position where the products will combine tightly right up to the point of cleavage - for instance, the 65-66 break occurs in the middle of a perfect amphipathic helix at the protein surface. One could use it to insert different amino acids at positions right after the break, or to recombine a synthetic peptide with a natural rest of the protein. However, they admit that it doesn't always work - it didn't with CuZn superoxide dismutase or α -lactalbumin.