

## Chemical Modification of Proteins

By "chemical modification" I mean formation or cleavage of covalent bonds, generally with the side chains, though cleavage of peptide bonds and modification of the  $\alpha$ -amino terminal can formally be included. These are generally not reversible by dilution, gel filtration or dialysis, though a few - Schiff base formation with aldehydes, some reactions of arginine - are thus reversible. Some modifications are reversed by changing the conditions, for instance lowering the pH, which can be very useful, but most are not. Some modifications are hydrolyzed in 6N HCl, as you do before amino acid analysis; some are stable even to that. Noncovalent interactions also have a place - competitive inhibitors which bind at active sites; low dielectric solvents change the UV absorbance of groups exposed to solvent, fluorescence-quenching molecules quench the fluorescence of exposed tryptophan residues. Even  $H^+$  can be looked at as a reversible modifier; one can judge the degree of exposure of peptide backbone NH by the rate of exchange of tritium on and off.

Chemical modification can have two types of specificity - chemical, i.e. for a particular amino acid side chain, and steric, i.e. fitting a particular site on the protein. This lecture is devoted to the former, the next to the latter. But remember that even a simple reagent may bind in some specific way to a particular protein and even modify in an unexpected way - for instance, iodoacetate, normally thought of as reacting with sulfhydryl groups, or possibly histidine or lysine, reacts with a carboxyl in ribonuclease T<sub>1</sub>.

Another way of categorizing protein reagents is whether the protein is expected to maintain activity under the conditions used for the reaction. Reagents used under extreme conditions are good for some things, such as sequence analysis, but for most purposes we want to retain activity, and use mild conditions such as pH 7 or 8, room temperature, little or no organic solvent (some is often needed to dissolve the reagent).

Still another way of categorizing reactions is whether they seek to find out something about the protein, investigate its structure or activity, or seek to *use* the protein, modify it to change its properties, attach it to a solid support, etc.

I want to concentrate on purposes and strategies of chemical modification, leaving particular reactions to the end, or to the long list I am giving you. I generally don't have time to discuss particular reactions in this lecture, but may be able to at the end of the next lecture. You will note that there are hardly any references after about 1975; there are enough reagents that there hasn't been much need to discover more.

Purposes:

1) Analysis: How many of a certain amino acid are present in a protein? Of course this is usually done by amino acid analysis after hydrolysis, but tryptophan is destroyed in acid hydrolysis and is often determined by chemical reactions. More interestingly, how many are exposed to solvent, to reagents in solution? One can determine total residues by carrying out the reaction under denaturing conditions, the exposed residues by carrying the reaction out under non-denaturing conditions. Non-covalent methods such as solvent perturbation of the absorption spectrum or fluorescence quenching, can be used here. There is a technique called isotope competition, which judges the relative reactivity - as a result of both chemical and steric effects - of groups on the surface of a protein: one reacts the protein with a small amount of a radioactively labeled reagent, and groups on the surface *compete* for it, the most reactive getting the most of it; then one floods with non-radioactive reagent, so that all reactive groups on the protein are fully modified in a chemically identical way, which simplifies analysis. The protein is then taken apart by enzymic or chemical proteolytic reactions, the peptides isolated and identified, and their specific radioactivity determined to determine which residue in the sequence is most reactive, which may correlate with participation in biological activity.

A related topic is labeling of membrane proteins, using reagents which do not penetrate the membrane to determine what parts of the protein are exposed on the outside of the cell or organelle. See Rosenberg, pp. 25-32.

2) As an aid in sequence analysis. These reactions do not require mild conditions, since the protein is being taken apart anyway, and can be done on proteins bound to membranes after electrophoresis, or even in the gel. These include enzymatic and chemical cleavage of peptide bonds; see Rosenberg pp. 183-195 on cleavage of proteins at specific peptide bonds. But also, SH groups are generally converted to stable carboxymethylcysteine derivatives, so that you don't get mixtures of oxidized and reduced cysteine during peptide purification. Or they are oxidized with performic acid, H<sub>2</sub>O<sub>2</sub> in HCOOH, to cysteic acid, the best form for amino acid analysis. If the protein has disulfide bonds, the procedure to determine what cysteine is linked to what other one is to cut it up into peptides with the disulfide bonds intact, then oxidize them and analyze the individual peptides. One method is to carry out then layer chromatography or electrophoresis, oxidize the disulfides after movement in one direction, then chromatograph or electrophorese again under the same conditions. Peptides not containing disulfides will all be on a diagonal line, peptides resulting from disulfide cleavage will be off the diagonal. They can be eluted and sequenced.

3) To understand mechanism of action. What groups are at the site responsible for biological activity? How many? Which ones in the sequence? For enzymes we aim to characterize the chemical mechanism, which requires identifying the participating groups, eventually to characterize the conformation of the active site, and beyond that to understand changes in conformation during catalysis. Much of this is best done by X-ray crystallography, but a complex with an inhibitor bound at the active site is invaluable in determining where that site is. Conformation may be evaluatable through the use of chemically similar but sterically different reagents, those hindered in approaching the active site being less reactive. Similarly, the distance between groups on the enzyme's surface is often measured by seeing whether bifunctional reagents, with two reactive groups, will react with both.

The basic approach is of course to react the protein with various reagents and see which ones affect the activity. The chief problem is in distinguishing between residues which are *essential* for activity (catalysis or whatever), those not essential but sterically important - their modification blocks access to the active site or causes or prevents a critical conformational change; and finally the general denaturation of the protein due to widespread modification. The last problem is ideally avoided by the use of *reversible* reagents, which can be removed with recovery of activity, proving that the enzyme retained its proper conformation and the inactivation was specific. One can also observe the protein by physical methods such as circular dichroism which distinguish between native and random conformations. Essential and merely blocking modifications can be distinguished by active site titration, which I have described earlier and will mention again later. Another good way is *subtractive* modification, making the side chain *smaller* rather than larger, so that no steric blocking can have been introduced; but there are few chemical ways of doing this. Site directed mutagenesis can easily do this, mutating a larger amino acid to alanine.

It is worthwhile to make definitions: the *catalytic* site is the residues actually involved in catalysis, interacting chemically with the substrate, or possibly activating other residues which interact with it. The *substrate binding* site is generally larger, including many other residues which bind normal substrates; a smaller substrate may still be acted on after a reaction in the substrate binding site. However, binding interactions are often important in adjusting the catalytic site for catalysis. Then there are *effector* sites, where other molecules bind and affect the biological activity without themselves being substrates or competitive inhibitors. These sites can be investigated by chemical modification in the same way as the catalytic and binding sites.

A technique often used to demonstrate that chemical modification *is* at the active site is *substrate protection*: the reaction is carried out in the presence, and separately in the absence, of a high concentration of substrate or good competitive inhibitor. This should prevent a loss of activity seen in

absence of the substrate and decrease the number of amino acid residues modified. Some caution must be exercised; in Koshland's early work with phosphoglucomutase some six residues had their reactivity increased or decreased by the binding of one substrate molecule, more than it could cover without a conformational change - and an *increase* in reactivity in presence of substrate or inhibitor is very likely to indicate a conformational change, exposing a previously buried residue.

If the results are not complicated, this technique can be used to identify a particular residue at an active site. Modification in absence of substrate or inhibitor results in modification of, say, six cysteines; modification in presence of substrate modifies five cysteines, one is protected. You carry out the modification reaction with cold (non-radioactive) reagent in *presence* of substrate, which protects one residue. Then you remove excess reagent and substrate by dialysis or gel filtration, and repeat the modification, this time with radioactive reagent, which now labels only the residue previously protected by the substrate. Then you cut up the protein with proteases, separate peptides, identify the peptide which is radioactively labeled, sequence it, and thus determine exactly which residue in the total enzyme is at the active site. Before the ready determination of protein sequences through their DNA sequences, such identification of an "active site peptide" often preceded determination of the complete amino acid sequence of the protein.

[Here I might give an example. When I went to Brookhaven National Laboratory as a post-doc, over thirty years ago, it had just been shown that trypsin hydrolyzed ethyl *p*-guanidinobenzoate, with the guanidinobenzoyl residue remaining attached to the active site serine, hydrolyzing off very slowly. There was a difference spectrum - guanidinobenzoyl-trypsin had more absorbance in the 260-275 nm region than was seen with trypsin and ethyl guanidinobenzoate in the same light path in a tandem cell. This suggested that the benzene ring of the guanidinobenzoate was binding in a hydrophobic region, perhaps up against a tyrosine, whose absorbance thereby would be increased because its surroundings would be more non-polar. My mission was to try to identify a tyrosine which would react with a modifying reagent, tetranitromethane, in the free enzyme, but not in guanidinobenzoyltrypsin, being substrate-protected. I was not very successful, getting at best a difference between 4.9 tyrosines nitrated in the free enzyme and 4.3 in the guanidinobenzoyl-enzyme, but this should at least illustrate a discrete problem and experimental approach.]

Chemical modification can also identify hyper-reactive residues, i.e. those which are *more* reactive than others of their kind, due to their chemical surroundings, as for instance being better nucleophiles because their proton, on a cysteine or histidine, is attracted off by a nearby group. Of course many residues are *less* reactive than the normal free amino acid, because they are more or less buried in the protein, but surprisingly often a single residue per subunit will be so much more reactive than others as to be the only one to react under suitable mild conditions. This is often thought to indicate a residue involved in catalysis, though there is no guarantee of this. The classic example is the active site serine in proteases and esterases, which reacts with all sorts of reagents which modify neither other serine residues (serine is normally quite unreactive, corresponding to its very high pK<sub>a</sub>, about 13) nor other reactive residues.

**4) Physical modification** - change in the physical characteristics of the protein. For instance, destabilization of complexes, including solubilization of hydrophobic proteins normally found in membranes, especially by succinylation, which converts a positively charged lysine to a negatively charged group and thus give the whole protein a much more negative charge. Reaction of SH groups not needed for activity can also lead to depolymerization, as in hemoglobin and aspartate transcarbamylase. Conversely, bifunctional reagents can be used to tie complexes together so that they can be copurified. Another example is modification of proteins with polyethylene glycol to make them non-immunogenic, the invention of our Drs. Davis and Van Es upon which a whole company, Enzon, was founded.

**5) Cross-linking and attachment to supports**: much of cross-linking I have just mentioned, but there is also insolubilization of protein by intermolecular cross-linking, so that for instance insoluble but

active trypsin can be removed from a digestion by centrifugation. Attachment of proteins to solid supports was first carried out by Ephraim Katchalski in 1954 to mimic membrane-bound enzymes - the properties in terms of pH optimum and even  $K_m$  are often change due to surface and diffusional effects. More often proteins are immobilized to make enzyme reactors, or for affinity chromatography, including immunoaffinity. I mentioned some of the methods in that context. The whole industry of immobilized enzymes is based here. Example: cold beer often shows a haze due to precipitated protein; the public is believed to like their beer crystal clear, and the protein can be dissolved by papain digestion. But the public also doesn't like to read any ingredients on the label other than barley malt, perhaps other grains, hops and spring water. Solution: run it through a column of immobilized papain; the protein is broken down, but the papain doesn't end up in the beer. You'd be surprised how much of food science is driven by keeping the list of ingredients down.

6) Hapten attachment to carrier proteins, to elicit antibodies to the hapten, which you wouldn't get if it were the free small molecules.

7) Attachment of reporter groups sensitive to their environment - chromophores whose absorbance is pH-sensitive, for instance *p*-nitrophenols whose absorbance changes with the pH in their immediate vicinity; fluorescence probes, spin labeled groups for electron paramagnetic resonance, heavy atoms for X-ray crystallography. Iodinated compounds are often used here, but other heavy atoms which can be incorporated into organic compounds could be used.

In most cases one wishes to determine the degree of modification, how many groups have been introduced per mole of protein. Ideally the modification is stable to acid hydrolysis and you can quantitate the modified amino acid by amino acid analysis, which also gives the moles of protein by the same analysis, based on the amount of some amino acid present as only a few residues per mole of protein. Otherwise, modification may generally be determined by incorporation of a radioactively labeled reagent, or by difference, i.e. carrying out some chemical determination of *un*modified residues - for instance, modification of lysines may be measured as decrease in the number able to react with trinitrobenzenesulfonic acid, or modification of cysteine as decrease in the number able to react with 5,5'-dithiobis(2-nitrobenzoate), DTNB for short (draw). Some modifications can be measured directly by spectrophotometry, as for instance 3-nitrotyrosine at 428 nm. But you also have to be able to measure the amount of protein you have, in order to get the ratio moles reagent incorporated/mole protein. And if for instance you are modifying tyrosines you have no faith in the Lowry method, because how will the nitrotyrosines react? A version of the biuret reaction is good here. Amino acid analysis is best, if you have an analyzer available, because it gives amount of modified amino acid and amount of protein on the same sample, and because it should be unequivocal in identifying the modified amino acids. Many reagents, especially alkylators such as iodoacetate, can react with more than one kind of amino acid. Determination by analysis is also important as you break up the protein into peptides to determine which amino acid in the sequence has been modified.

Then one correlates degree and rate of modification with loss of activity. You plot number of residues per mole protein vs. % loss of activity, and extrapolate to zero activity; this usually indicates the number of residues which must be modified to abolish activity. The actual plot usually doesn't quite hit zero, as other residues are modified at higher reagent concentration or time of reaction. You get the different points of this plot usually by carrying out the modification over a period of time, removing aliquots at different times, stopping the reaction, and determining the amount of modification. Of course if you find that complete inactivation depends on reacting with two equivalents of reagent, this doesn't necessarily mean that two residues of the modified amino acid are important for activity; probably two residues react at about the same rate, at least one of which is essential for activity. If there are two essential residues, modification of *either* inactivating, the plot of % activity remaining vs. residues modified should be curved, actually showing a quadratic dependence, though the data won't be good enough to define this clearly.

The problem of determining what modified residue was critical for activity was exemplified in the study of Koshland, Strumeyer and Ray, Brookhaven Symposium in Biology, 1962, in which chymotrypsin was being photo-oxidized - a dye which on absorption of light is promoted to a triplet state interacts with molecular oxygen, normally in a triplet state, and converts it to a more reactive singlet state, which oxidizes cysteine, methionine, histidine and tryptophan in proteins. In this case both histidine and methionine were being oxidized, with loss of enzyme activity; which was significant?

The rate of loss of activity could be determined - in general one plots  $\ln \% \text{ active}$ , or  $\ln [E]/[E]_0$  where  $[E]_0$  is the original activity, vs time, and the slope of this log plot is  $-k_{\text{obs}}$ , the rate of loss of activity under the conditions used. If the fullest possible modification of the enzyme does not completely inactivate it, this plot will not be a straight line going down indefinitely, but will level off at a constant non-zero activity, which might be expressed  $[E]_{\infty}$ , activity at infinite time of modification. Then one plots  $\ln ([E]-[E]_{\infty})/[E]_0$  vs  $t$  to get  $k_{\text{obs}}$ .  $k_{\text{obs}}$  can then be plotted vs some other variable - for instance, in what I call "ordinary" chemical modification, in which reaction depends on random encounters between protein and reagent, a plot of  $k_{\text{obs}}$  vs reagent concentration will be a straight line going up indefinitely, while in active-site-directed reaction, as we shall see,  $k_{\text{obs}}$  levels off as reagent concentration increases, just like  $v$  with increasing  $[S]$  in an enzyme-catalyzed reaction.

To return to the chymotrypsin case: the rate of loss of histidine and methionine could be determined similarly, by plotting  $\ln (\text{his remaining/original his content})$  vs time, and both rates were slower than the rate of activity. At this point they devised an active site titrant, cinnamoylimidazole, the first active site titrant used - it reacted with the active site to give a cinnamoyl-enzyme and free imidazole, with a loss of specific absorbance. They could then similarly determine the rate of loss of active site, and show that this rate was the same as the rate of loss of histidine. This was the first direct evidence that there was an essential histidine in the catalytic site of chymotrypsin, the histidine now well known as part of the 'catalytic triad'. In the discussion after this paper, Elliott Shaw stood up and gave evidence for a catalytic histidine based on its modification by an active-site-directed reagent, the first report of such. The methionine turned out to be met<sup>192</sup>, very near the critical ser<sup>195</sup> in the sequence, but not catalytic - its oxidation merely increased the  $K_m$  for the substrate enough so that in the standard assay the enzyme appeared to lose activity.

Let me say a few things about the chemistry of modification reactions. Most are with the *basic* form of an amino acid, reacting as a nucleophile. The cysteine sulfhydryl, reacting as S<sup>-</sup>, is the most reactive, but also, in order of decreasing reactivity, the uncharged forms of histidine and lysine, the anions of tyrosine and the carboxylic acids, the uncharged sulfur of methionine, and, rarely, the indole nitrogen of tryptophan. The rate of reaction of any of these which are protonatable will therefore decrease as the group is protonated; indeed, the pH dependence of the rate of chemical modification is a very good way of determining the pK<sub>a</sub> of the group in the free enzyme, as the pH dependence of V<sub>max</sub> shows the pK<sub>a</sub> of the group in the enzyme-substrate complex, while the pH dependence of V<sub>max</sub>/K<sub>m</sub> generally shows the pK<sub>a</sub> in the free enzyme, but may be influenced by other binding effects (and by the pK<sub>a</sub>s, if any, of the substrate). Among protonatable nucleophiles, the higher the pK<sub>a</sub>, in general, the better the nucleophile. Thus specificity, for rather non-specific reagents such as many alkylating reagents, depends on pH; iodoacetamide may react with methionine at pH 5, histidine at pH 7, and lysine at pH 9; cysteine, however, is always most reactive, and will react first with such reagents. It is sometimes necessary to block cysteine with a *reversible* modification in order to investigate the effect of modifying some other amino acid. Arginine is protonated under all normal circumstances, and does not then react as a nucleophile; however, there is a special category of reagents for arginine, generally with two adjacent carbonyl groups, as butane-2,3-dione, phenylglyoxal; some of these give a modification which is stable only in the presence of borate, which makes them conveniently reversible by removing excess borate.

**Acylation** gives amides with lysine, thioesters with cysteine, acylimidazoles with histidine, esters with tyrosine and rarely serine and threonine. The modifications of cysteine, histidine and tyrosine are not very stable, and are removed quantitatively by neutral hydroxylamine, leaving the acylated lysines. Anhydrides such as maleic anhydride, with the other carboxyl group in the acyl-lysine nearby, come off again at more acid pH by reformation of the anhydride, which can be useful - the pH at which they come off depends on the exact structure of the acyl group, tetrahydrophthalate, with the carboxyls held close to each other not only by a double bond between them but also by the ring structure, comes off at pH 6.

**Amidination and guanidination** are reactions only of lysines; they maintain a positive charge on the group, whereas acylation neutralizes it. Sometimes having the positive charge is necessary for stability of the protein. Positive charge is also maintained when the amino group is merely methylated, by Schiff base formation with formaldehyde and borohydride reduction. Amidination and guanidination tend to require rather high pH, depending on the reagent; they are reactions of unprotonated lysine.

Also shown on the list are some reagents with some specificity for histidine and tyrosine. Cyanuric fluoride could be called either acylating or arylating.

**Alkylation** is the widest class of reactions, the most used. They depend on attack of a nucleophilic group of the protein on an activated carbon atom, usually next to a carbonyl group - even in an acid or amide - or to an aromatic nucleus. Usually the nucleophile displaces a halogen atom from the activated carbon, though in some cases it adds to a double bond, or a three-membered ring as in ethyleneimine and ethylene oxide. Iodine is the most easily displaced, so iodoacetate and iodoacetamide are general alkylating agents, usually of cysteine but occasionally of histidine, lysine, methionine and carboxylic acids. Bromine and chlorine are successively less easily displaced - which means that they generally react only when the reagent binds non-covalently to the enzyme - and fluorine is displaceable only from sulfonyl fluorides, like the well known inactivator of serine proteases, phenylmethane sulfonyl fluoride. The products of these reactions are generally acid-stable, and some give products which are well characterized in amino acid analysis, such as carboxymethylcysteine.

**Arylation** is similar, but the nucleophile reacts directly with the aromatic nucleus; here fluorine is the *most* easily displaced, in the familiar 1-fluoro-2,4-dinitrobenzene (Sanger's reagent).

**Reduction of disulfides** is usually by excess small molecule sulfhydryl compound, such as 2-mercaptoethanol, or 1,4-dithiothreitol, whose oxidized form is a 6-membered ring, which pulls the reaction in that direction and allows a much lower concentration to be used - 1 mM dithiothreitol is as good as 50 mM mercaptoethanol and much less smelly. A few other reagents have been used; NaBH<sub>4</sub> selectively reduced specific SS bonds in the reference given.

**Disulfide interchange** is used either to measure free SH, with Ellmann's reagent, 5,5-dithiobis(2-nitrobenzoate), the released 5-thio-2-nitrobenzoate having a high absorbance,  $\epsilon_{412} = 13,600$ , or to put a removable blocking agent on free SH groups before modifying other groups with a reagent which otherwise would react with the SH groups. Tetrathionate and methyl methanethiosulfonate are used for this purpose.

Sulfhydryl groups can be oxidized to several states - the first state, a sulfinic acid, -SOH, is not stable in small molecules, which dismute to a sulfenic acid -SO<sub>2</sub>H and a sulfhydryl, but can occur when well isolated on a protein. The usual oxidation is all the way to cysteic acid, which is easier to measure by amino acid analysis than cysteine, which after acid hydrolysis is a mixture of cysteine, cystine and *mesocystine*, cystine with one  $\alpha$  carbon inverted in configuration to give a compound separable on the amino acid analyzer. It is easier to oxidize all the cysteine to a single product. Cysteine of course also reacts with **mercury compounds** - the reaction with *p*-chloromercuribenzoate was the old way of measuring free SH groups, by a change in A<sub>255</sub>. Mercury compounds are removable by a large excess of mercaptoethanol, or even better by passage through an immobilized SH column; similarly, proteins may be adsorbed on an immobilized mercurial column and eluted with mercaptoethanol, or the fluorescent dansyl group may be attached through a mercury.

There are various ways of **oxidizing** tryptophan and methionine - though they also oxidize cysteine. You probably know about cleavage of the peptide bond of methionine with cyanogen bromide; 3-bromo-2(2-nitrophenylsulfenyl) skatole does the same thing at tryptophan. Oxidative reactions are otherwise rather hard to control, and typically do not yield a single identifiable product. The oxidative cleavages of the peptide chain are useful to molecular biologists: you can determine the N-terminal sequence of each peptide produced and be sure it was preceded by a methionine or tryptophan, which decreases the degeneracy of the oligonucleotide probe designed from the sequence - you want as few codon choices as possible in the oligonucleotide sequence, and Met and Trp each have only one codon.

There are a few strong **reducing agents** - Raney nickel, which reduces cysteine to alanine, and diborane and alkylboranes, which can reduce carboxyl groups to CH<sub>2</sub>OH. These then *decrease* the size of the side chain. Unfortunately the reaction tends to be rather incomplete, and they are not often used.

Carboxyl groups are usually modified by **ester** or **amide** formation. Some reagents form the ester directly, such as *p*-bromophenacyl bromide reacting with active-site carboxyls in pepsin, but more usually the carboxyls are activated with a water-soluble carbodiimide, or rarely Woodward's K reagent, and then react with a nucleophile such as aminomethanesulfonic acid, which can be measured on the amino acid analyzer. The first chemical modification of a protein was Frankel-Conrat's esterification of the carboxyls of ribonuclease by refluxing in methanolic HCl! RNase is a *very* stable protein.