THE PROTEIN CHEMISTRY OF ENZYMES

R. N. PERHAM

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, England, UK

1. Introduction

Enzymes are proteins, things of beauty and a joy forever. The first part of that statement no longer brooks dispute, the second, I confess, is an opinion. When Kühne coined the term 'enzyme' (literally 'in yeast') in 1876, the controversy between Pasteur and Liebig was as its height, Liebig professing that fermentation was brought about by the action of what we, following Kühne, would call enzymes, whereas Pasteur clung to the view that fermentation was inseparable from living cells. This residue of belief in vital forces was laid to rest when Buchner (1897) demonstrated that a cell-free yeast extract could catalyse alcoholic fermentation, that is, an 'organized ferment' could act outside the living cell. Thus there was no difference in principle from the 'unorganized ferments', such as gastric juice, secreted by cells.

For reasons of experimental accessibility, extracellular enzymes, in particular those of digestion, have figured prominently in the history of the subject. As early as 1783, Spallanzani had been able to show that gastric juice would digest meat in vitro and Schwann (1836) later termed the active substance pepsin. Kühne himself appears to have given trypsin its present name, though its existence in the intestine had been suspected since the early 19th century, and he demonstrated that freshly-secreted pancreatic juice was proteolytically inactive, an early indication of the existence of zymogens.

Efforts to purify enzymes met with little success until the 1920s. Many of the early workers had thought enzymes to be protein but the evidence was indirect. For example, enzyme inactivation and protein denaturation were observed to have comparable and uniquely high temperature coefficients. Then, in 1926, Sumner crystallized urease from jack bean meal and announced it to be a simple protein [1]. On the other hand, the distinguished chemist Willstätter and his colleagues purified a number of enzymes but, probably because of the diluteness of the solutions, failed to identify them as proteins [2]. Willstätter argued in terms of an 'active prosthetic group' and a 'colloidal carrier' but Sumner's claim that an enzyme could be a simple protein was soon substantiated by a classic series of experiments on the isolation of crystalline proteolytic enzymes by Northrop and his colleagues [3], beginning with pepsin in 1930. By that time, sufficient criteria of purity were becoming available, such as ultracentrifugation, boundary electrophoresis and tests of constant solubility, for Northrop's work to be put beyond challenge.

Perhaps it was a legacy of the dispute over 'organized' and 'unorganized ferments' that work on the purification of intra-cellular enzymes did not begin in earnest until the later 1930s. Certainly the problems involved are vastly more complex but the fascinating contemporary work on the dissection of the reactions of glycolysis and the tricarboxylic acid cycle inevitably focussed attention on the enzymes responsible for those reactions and, by the late 1940s, many of them were available in pure form and in sufficient quantity for studies of their chemical structure of be undertaken. Since that time, the number of enzymes brought to purity, even crystallized, has grown to massive proportions. In the purification of enzymes, as in the other areas of protein chemistry discussed later, it was improvement in techniques, for example, the better understanding of factors affecting protein solubility, the introduction of ion-exchange chromatography on suitable supports, the advent of gel filtration, that led to this impressive sudden growth of knowledge [4]. More recently, affinity chromatography, which relies on the special ability of an enzyme in a mixture of

North-Holland Publishing Company – Amsterdam

proteins to bind non-covalently to an immobilized small molecule that closely resembles its substrate or cofactor, has placed another powerful tool in the hand of the chemist [5]. The distinction between the wanted enzyme and all other proteins in the mixture is thus put to the best of uses in its purification.

The recognition that enzymes are specific proteins meant of course that the methods of protein chemistry could be brought to bear in the analysis of their structure and activity. As a corollary, much of our knowledge of the chemistry of proteins stems from experiments with enzymes. Our modern views of the chemistry of proteins and enzymes are therefore to a large extent inseparable and we must consider them together.

The primary structure of enzymes

The first amino acid to be isolated from a protein was leucine, by Proust, in 1819 and, with the growth of technical facility in organic chemistry, most of the amino acids commonly occurring in proteins had been characterized by the beginning of the 20th century. It is, however, salutary to recall that threonine remained unidentified until 1936 [6]. As early as 1872, Ritthausen had suggested that the amino acid composition might be characteristic of a protein and, indeed, by 1915 wide differences of amino acid composition between proteins were commonly accepted. Remarkably accurate results could be obtained, for example, by admirably painstaking gravimetric analysis of amino acids selectively precipitated from a hydrolysate of a protein [7]. The introduction of newer methods, such as enzymic or microbiological assay, for the analysis of certain amino acids led to distinct improvements [7], the determination of the amino acid composition of β-lactoglobulin by Brand and his colleagues testifying to the precision that could be attained [8].

However, the basis for the dramatic advances in amino acid analysis, which can be attributed principally to Moore and Stein [9], was laid in the early 1940s in the development of liquid–liquid chromatography by Martin and Synge. After valiant efforts with starch columns that permitted the amino acid composition of a protein to be determined in about ten days, Moore and Stein turned to ion-exchange chromatography on sulphonated polystyrene resins. By 1958 the procedure had been automated [10] and the advances since that time have been largely concerned with shortening the time taken for an analysis (to little over an hour) and reducing the amount of protein required (to about 10 μg) [11]. The only drawback is the high cost of the sophisticated apparatus needed and manual methods have a part still to play on occasion [12].

Not surprisingly, the improvements in amino acid analysis outstripped our knowledge of the arrangement of amino acids in proteins. When Fox in 1945 reviewed the study of the terminal residues of proteins [13], only one position of one residue in one protein was known with certainty. That was the presence of phenylalanine as an N-terminal residue in insulin, which had been deduced by Jensen and Evens ten years before [14] from their isolation of the phenylhydantoin of that amino acid from a hydrolysate of insulin treated with phenylisocyanate. A widely discussed hypothesis at the time was due to Bergmann and Niemann [15] who, on the basis of amino acid analyses of various proteins, had proposed that the total number of amino acids in a protein molecule could be expressed by the formula $2^m \times 3^n$, where m and n are integers but not zero. Neuberger and also Pirie [7] were swift to point out that the analytical data were imprecise enough to account for the apparent regularities but these ideas of periodic chemical structures in proteins had their contemporary physical counterparts. The concept of repeating sequences of amino acids arranged in systematic patterns of closed structures was developed by Wrinch in her 'cyclol theory' which, however, was shortly dismissed by Pauling and Niemann [16]. Svedberg, too, drew attention to regularities in molecular weights determined with the aid of the newly developed ultracentrifuge and, in proposing a common plan of protein structure built up by aggregation of units, essayed a parallel with the concept of isotopes unravelled not long before [17].

The determination in 1951 of the amino acid sequence of the β-chain of the hormone, insulin, by Sanger and Tuppy saw the beginning of the end of all such speculations. Sanger observed, in reviewing the work in 1952 [18], that the fact that a unique sequence could be deduced was evidence that a protein

was really a single chemical substance, adding, perceptively, that this in turn implied 'an absolute specificity for the mechanisms responsible for protein synthesis and this should be taken into account when considering such mechanisms.' The mechanisms were of course unknown at that time and the structure of DNA was unsolved. The impact of this aspect of Sanger's work at a critical time in the development of our modern views of protein biosynthesis and molecular genetics is sometimes overlooked.

Other workers were quick to apply Sanger's methods to the study of enzyme proteins. By the early 1950s, the terminal residues of many enzymes had been determined and fragmentary sequence information was becoming available. Sanger had pioneered the use of fluorodinitrobenzene to identify N-terminal residues and relied heavily on liquid chromatography, principally on paper supports following the work of Consden, Gordon, Martin and Synge [19], and paper electrophoresis to separate the peptide products of partial hydrolysis of a protein. It should be remarked that the possibility of using proteolytic enzymes to bring about specific degradation of protein molecules without sequence rearrangement had been put in doubt by work in Bergmann's laboratory some years before. The method therefore had to be validated by Sanger and his colleagues before reliance could be placed on the results [18]. In 1950, Edman [20] advanced an alternative strategy for sequence determination, namely, stepwise degradation of a peptide chain from the N-terminus with phenylisothiocyanate. Moore and Stein, who, as we have seen, were bringing about striking advances in amino acid analysis, subsequently adapted this strategy by choosing to analyse the peptide after each removal of the N-terminal residue, introducing the subtractive Edman degradation that became the basis of their attack on the amino acid sequence of the enzyme ribonuclease and which was widely copied. The structure of insulin (51 residues) was completed in 1955 [21] and the primary structure of the first two enzymes to be analysed, ribonuclease [22] and egg-white lysozyme [23] followed in 1963. Both of these enzymes, simple proteins, contain about 120 residues.

The wish to improve the sensitivity of amino acid sequence analysis saw the introduction, in 1963, of the dansyl-Edman degradation [24] in which the fresh N-terminal residue exposed at each step of an Edman degradation of a peptide is identified as the fluorescent dansyl derivative. This technique, cheap yet elegant and effective, has since played a prominent part in the determination of the amino acid sequence of many enzymes [25]. Thus, the 200-residue barrier was broken in 1964 with chymotrypsinogen [26], the 300-residue mark was passed in 1967 with glyceraldehyde 3-phosphate dehydrogenase [27] and the massive 500-residue chain of glutamate dehydrogenase was completed in 1972 [28]. The main change of emphasis in the past 5 years has been the growing importance of automated methods based on the Edman degradation. The liquid-phase sequencer, first described by Edman and Begg in 1967 [29], is now widely used to identify the first 40 or so residues of a polypeptide chain [30] and a comparable instrument in which the peptide under study is first covalently attached to a solid support is proving very successful for work with peptides of up to, say 30 residues [31]. Mass spectrometry, too, has come to play an increasingly important part and may indeed emerge as the method of choice in many instances [32].

Most of the early work on amino acid sequence determination was concerned with the relatively small extra-cellular enzymes but interest shifted naturally to the intra-cellular enzymes as they became available in pure form; glyceraldehyde 3-phosphate dehydrogenase [27] was in fact the first intra-cellular enzyme for which a complete sequence was established. The extra-cellular enzymes tend to be small but rich in disulphide bridges, the intra-cellular enzymes are mostly larger, often possessing quaternary structure. Many primary structures are now listed in a valuable compendium published from time to time [33]. The awesome industry of the past 20 years, recorded there in the form of complete or partial amino acid sequences for many enzymes, has given no indication that enzymes are other than proteins with unique amino acid sequences assembled using the conventional peptide bond first described by Hofmeister and Fischer in 1902 (neglecting, of course, the occasional presence of prosthetic groups, essential metal ions and so on). Nor has it permitted any resurrection of the concept of regularities in the sequences [34,35]. Each and every enzyme therefore presents its own particular challenge to the chemist.

The active site

From the earliest work on the general disparity in size of enzyme and substrate, the concept of specificity and the kinetic evidence for the existence of enzyme-sbustrate complexes, it became apparent that the conversion of substrate to product must take place at a very restricted part of an enzyme molecule, which soon became known as the 'active site' or 'active centre' [4]. Clearly, however, experiments to identify the amino acid side-chains that might be taking part in the catalysis had to await the isolation of pure enzymes and the development of chemical techniques to deal with the problems such work threw up. The thoughtful pioneering experiments of Herriott on the newly purified pepsin deserve mention here. In 1935 he showed that acetylation of the phenolic hydroxyl groups of tyrosine residues caused loss of activity and, in 1937, that iodination of tyrosine residue also inhibited the enzyme [3,36].

With intra-cellular enzymes, however, only crude extracts were available at first. Lundsgaard, in 1930, had drawn attention to the important effect of halogenoacetic acids as inhibitors of glycolysis and fermentation and Quastel, Rapkine and Dickens, independently in 1933 [37], were able to show that a likely cause was alkylation of the thiol groups of an enzyme or coenzyme such as glutathione. Rapkine later (1938) correctly identified the susceptible enzyme as that we know today as glyceraldehyde 3-phosphate dehydrogenase [38]. In due course, the purification of the enzyme in various laboratories enabled Racker and others to implicate a thiol-ester intermediate in the catalytic action of the enzyme, an important step forward in our knowledge of the mechanism of group transfer reactions [39], even though the true chemical identity of the thiol group concerned was not revealed as a unique cysteine residue in the enzyme until the sequence work of Harris and Park a further decade later [40]. These latter experiments, by chemical analysis of the thiol-ester intermediate, demonstrated unequivocally that inhibition by treatment with iodoacetate is due to alkylation of the cysteine residue that normally carries the substrate in thiol-ester linkage. Such a demonstration was of particular interest because it had long been recognized that modification of a residue other than one taking part directly in the catalytic action of an enzyme might inhibit by interfering sterically with the approach of the substrate to the active site [41]. To be sure, sophisticated thinking about the interpretation of the effect of chemical modification of an enzyme was much in evidence even earlier in a typically elegant paper from Hopkins and colleagues [42], in which they showed, among other things, that substrate and competitive inhibitors were able to protect succinic dehydrogenase from inhibition by oxidized glutathione. Testing the ascription of functional groups to the active sites of enzymes by such protection experiments thereafter became a standard practice.

The early successes in understanding the reactions of thiol groups with various compounds led to a flood of experiments on inhibition of enzymes with such reagents: enzymes were classified as dependent or not on thiol groups by this criterion [43]. But the dangers we now appreciate of assuming the identity of chemical reaction between an inhibitor and different enzymes only became manifest when the nature of such reactions could be pursued with pure enzymes and more sophisticated protein chemical techniques. Thus, iodoacetic acid was found to inhibit pancreatic, ribonuclease not by reaction with thiol groups – there is none in the enzyme – but by alkylation of histidine residues at the active site [44]. Most remarkably of all, inhibition of ribonuclease T_1 by iodoacetic acid was found to be due to specific alkylation of the γ-carboxyl group of glutamic acid-58 in the enzyme polypeptide chain [45]. In 1945, in a good review of the knowledge of protein denaturation at that time [46], Anson had occasion to observe that 'in general, no case has as yet been discovered in which an enzyme increases the reactivity of any group in native protein'. As we have seen from just one example, evidence in favour of unusual reactivitiy of functional groups in enzymes became available in due course [62], but to dispel any notion that catalytically essential groups are always hyper-reactive, the depressed reactivity of the carboxyl group of glutamic acid-35 in lysozyme was later demonstrated [47]. Numerous other examples of unusual functional group reactivity can be cited [62].

I have dwelt at some length on the reactions of thiol groups in enzymes since historically this represents an important strand in the development of our knowledge

of enzyme chemistry. Another strand that has proved of major significance concerns the proteolytic enzymes and esterases and their inhibition by organophosphorus compounds, first detected as an adjunct of work during the Second World War on chemical warfare agents [48]. Such compounds were observed to react irreversibly and specifically on a 1:1 molar basis only with the active, native form of the susceptible enzyme and not with the denatured polypeptide chain: the integrity of the active site was therefore seen to be critical. It was soon shown that in all such enzymes the phosphoryl group becomes attached to the hydroxyl group of a serine residue [49], not commonly regarded as a reactive side-chain in a protein, and sequence work on many esterases and proteases quickly followed [50]. Remarkable sequence homology around the modified serine residue in the various enzymes was uncovered, a point taken up later in this article; and the significance of the particular serine residue in chymotrypsin (and, by analogy, in the other enzymes) became apparent when, in 1958, it was demonstrated that this self-same residue carries the acyl group of the substrate (an acyl-enzyme was already thought to be a catalytic intermediate) in ester linkage with the hydroxyl group [51]. (This study was an ideal model for the work on the thiol-ester intermediate for glyceraldehyde 3-phosphate dehydrogenase referred to above.) The joint work of protein chemists and crystallographers in providing a complete molecular structure for chymotrypsin was later satisfyingly to reveal that the unique reactivity of the serine residue is caused by its participation in a 'charge-relay system' with the imidazole side-chain of a histidine residue and the β-carboxyl of a buried aspartic acid residue [52]. Other 'serine proteases' were subsequently ascribed comparable structures at the active site.

In retrospect, the work with organophosphorus inhibitors assumes an extra significance. Much of the chemical work to probe the active sites of enzymes did, and still does, depend on the use of group specific reagents, selectivity of reaction being conferred by the (unforseeable) hyper- or hypo-reactivity of functionally important side chains or, perhaps, by substrate protection. The organophosphorus inhibitors were restricted to a class of enzymes and were manifestly not group (serine!) specific reagents. The possibility of utilizing the specific binding properties of a substrate analogue to carry an otherwise unspecific chemical 'warhead' into the active site of an enzyme was realized independently by several workers in 1961 and the development of this approach has been well reviewed recently by one of its principal begetters [53]. Suffice it to say here, of chymotrypsin, that particular histidine and methionine residues were soon allocated positions in the active site of that protease. Moreover, although the idea that an active site consists of a constellation of amino acid side-chains brought into some efficacious proximity by the specific folding of the polypeptide chain had won general acceptance by the middle 1950s [4], the chemical evidence in favour of it was mostly indirect. The fact that different active-site-directed inhibitors could react with different amino acid residues contributed by widely separated regions of the primary structure of the enzyme was to add conviction to the proposition. The method has recently been given a new gloss by the proposal to produce the 'warhead' on the inhibitor in situ by photogeneration [88].

Study of the chemistry of the active site of certain enzymes was greatly aided by the recognition that they contained covalently-bound prosthetic groups. The inception of this approach may be found in the classic work of Tuppy and Maleus [54] on the amino acid sequence in the haem-binding site of cytochrome *c* but comparable systems were under investigation elsewhere. Thus, for example, phosphorylation of a serine residue was found to accompany conversion of phosphorylase *b* into phosphorylase *a* and to be part of the mechanism of phosphoglucomutase (reviewed by Sanger [50]). The special virtue of this approach lies in the fact that these chemical modifications are catalysed by the enzymes themselves and their significance cannot therefore be doubted. Similarly, we have already seen how acyl-enzyme intermediates of the normal catalytic reaction came to be identified for chymotrypsin and glyceraldehyde 3-phosphate dehydrogenase. In 1962 this concept was carried one stage further when it was demonstrated that the enzyme–substrate complex (an imine) of aldolase and dihydroxyacetone phosphate could be 'frozen' by reduction with sodium borohydride, thereby permitting chemical identification of the lysine residue that contributed the amino group [55], this work following the comparable use of borohydride to characterize the (Schiff-base) binding-site of pyridoxal phosphate in phosphorylase [56].

The existence of unusual protein structural features in the active site of an enzyme has also been recognized as a point of attack on the chemistry of the active site. For example, as summarized elsewhere [57], protein disulphide bridges that become alternately oxidized and reduced in the course of the reactions catalysed by several flavoprotein dehydrogenases have been characterized and, in each case, shown to be a small intra-chain disulphide loop. Finally, to stand an earlier principle on its head, it has recently been observed that in certain enzymes a functionally important group becomes reactive towards group-specific reagents only in the *presence* of the substrate. Thus, a single tyrosine residue in aspartate aminotransferase became susceptible to nitration only in the presence of the substrate pair, aspartate and 2-oxoglutarate [57]. Efforts to arrest enzymes 'in flight' may yet have something to tell us about the events, including protein conformational changes, that accompany catalysis, to complement the growing picture, detailed yet static, we have been building up of the chemistry of functional groups in active sites.

In the chapter on the chemistry of the active site in the second edition of their excellent book [4], Dixon and Webb could point to little in the way of detailed knowledge of the functional groups in enzymes. The identification of catalytically important groups from pK values obtained by studies of the pH-dependence of the reaction catalysed had come to be regarded with proper skepticism as the perturbations induced by proteins became apparent (though such investigations had, rightly as it turned out, thrust histidine to the fore as a residue likely to be found in many active sites); and degradation studies that purported to show that large tracts of primary structure could be digested away by means of exopeptidases without loss of catalytic activity were shortly to be discredited. The 1960s changed all that, thanks largely to the growing sophistication of protein chemical techniques, not least among which must rank the widespread use of radioisotopically-labelled reagents and inhibitors.

Topography and three-dimensional structure

The methods of protein chemistry cannot of course allow the determination of the three-dimensional structure of a protein in atomic detail, yet chemical experiments have played an important part in the development of our present ideas. The observation that denatured but not native egg albumin gives the nitroprusside test for free thiol groups was made early in this century and Mirsky and Pauling in 1936, by way of explanation, proposed that functional groups in the interior of a native, coiled protein become accessible when the molecule is opened up by denaturation. However, as late as 1945, Anson [46], though highly dubious, could not wholly discount alternative explanations, such as that of Linderstrom-Lang and Jacobsen, who had ingeniously suggested that unreactive thiol groups are blocked as thiazolidines which are broken open when the protein is denatured. Anson much preferred the idea that denaturation causes changes of reactivity of functional groups in proteins by means of unfolding the native structure and rupture of hydrogen bonds. As we now know, he came commendably close to the mark. What greatly puzzled him, understandably, was the evidence from chemical modification experiments on enzymes and other proteins that amino- and carboxyl-groups were on the outside of native protein molecules whereas other side-chains, such as those of cysteine and tyrosine residues were, on this interpretation, in the interior. Why was there this broad difference in accessibility of different classes of functional groups?

The answer did not emerge until the three-dimentional structures of proteins and enzymes were revealed by X-ray diffraction techniques and other experiments began to point to the underlying mechanisms of protein folding. By 1960, the primary structure of an enzyme was demonstrated to contain all the information needed by the protein spontaneously to adopt its native conformation (even its quaternary structure) from the denatured state [59]. X-Ray crystallographic analysis of protein structure soon revealed the tendency of hydrophobic groups to be associated with the interior of protein molecules whereas charged side-chains are almost invariably found on the outside, although the concept of a protein 'surface' has since been seen to be more complicated than the use of this simple term might imply [60]. However, the fact that so many studies of the chemical reactivities of functional groups in enzymes have subsequently been rationalized in terms of the three-

dimensional structures of those enzymes determined by X-ray crystallography (e.g. [61,76]) leads to the conclusion that the structures in the crystal and in solution must closely approximate one another, a conclusion comforting to crystallographers and protein chemists alike [62]. Chemical modification of functional groups as a predictive probe of protein topography therefore carries weight despite the complexities of interpretation that may be necessary [62,25]. Although a discussion of the pathways of protein folding is beyond the scope of the present article, it should be emphasized that chemical modification and other elegant techniques of protein chemistry, such as structural complementation of enzyme fragments, have played a major part in allowing the recent rapid advance of knowledge in this area too (well reviewed by Anfinsen [59]).

In the past fifteen years there has been a growing interest in examining the spatial arrangement of amino acid residues by chemical cross-linking experiments [63]. Some of the predictions of early experiments were at variance with the later results of X-ray crystallography and it took one or two undoubted successes to restore confidence in the method. A good example would be the demonstration that the catalytically essential cysteine and histidine residues in the active sites of several plant proteases must lie within 0.5 nm of each other since they could be cross-linked with dibromoacetone [64], a result confirmed by X-ray crystallography. Similarly, Anfinsen and his colleagues showed how a little clever chemistry could lead to the specific cross-linking of residues in the active site of staphylococcal nuclease [65]. More recently still, treatment of enzymes with general cross-linking reagents not restricted to the active site, for example bifunctional imidoesters, has given valuable information about subunit contacts and quaternary structure [66,67]. With complex enzymes, indeed, the combination of protein chemical techniques with physical methods such as electron microscopy and arguments from symmetry must serve in the absence of detailed X-ray crystallographic analysis [68,69].

Chemical modification of functional groups in enzymes has another use whose importance has also grown immeasurably in the past decade, namely, the introduction of probe molecules. Changes in the spectral properties of aromatic side-chains had long been used as a test of conformational change in enzymes (e.g. Shugar's classic work on ribonuclease [70]). In 1964, Burr and Koshland suggested that chromophoric groups bound to enzymes would be useful reporters of changes in their micro-enviroment during catalysis [71]. Many interesting experiments (e.g. [72,73]) have since been carried out on numerous enzymes using a variety of spectral probes (chromophores, fluorophores, spin-labels and n.m.r.) introduced at chemically defined amino acid side-chains, not necessarily in the active site. Reversible chemical modification to produce hybrid forms of oligomeric enzymes promises to be a welcome addition to the techniques for studying subunit interactions [86,87]. However, the relevance to the present article of such experiments lies not so much in the value of the results they have produced but in the fact that they would not have been conceived of without the development of our general knowledge of the protein chemistry of enzymes outlined above.

The Evolution of Enzymes

No attempt to analyse the development of present knowledge of the protein chemistry of enzymes would be complete without some brief reference to evolution. With the sequence work on enzymes in its early stages of development and Ingram's success in establishing the molecular lesion in haemoglobin-S still very fresh [74], Anfinsen observed in 1959 [75] that analysis of structural differences between the 'same' enzyme from different species might be used to determine whether particular residues are essential for activity. Retention of a residue cannot prove that it is essential but a change clearly signifies that it is not, an approach that has been usefully applied in numerous studies and which forms a valuable complement to chemical modification. The latter technique, ex hypothesi, can only deal with residues that can be chemically modified; Nature, through evolution, can in principle vary any amino acid in a protein.

It will be recalled that striking sequence homology around the functional serine residue in the active sites of various proteases and esterases became apparent early on [50]. The culmination of many years work in several laboratories was the remarkable conclusion that the pancreatic 'serine' proteases of differing

specificities have arisen by divergent evolution from a common ancestral form whereas the serine protease, subtilisin, of *B. subtilis* has converged to the same charge-relay mechanism at the active site from a completely different ancestral sequence [76]. An equally astonishing observation was the close homology between sequences of the milk protein, α-lactalbumin, and egg-white lysozyme, again indicative of gene duplication and divergent evolution [77]. The concept of enzyme families arising in this way was strengthened by the demonstrations that the sequences determined for different members of a family could be fitted without undue difficulty of the three-dimensional model of one of them deduced from X-ray crystallography [78,79], which held out hopes that one three-dimensional structure analysis might suffice for all members of a given family. These hopes have largely been verified by more recent work [80].

The study of amino acid sequences has also had an important impact in another area of enzyme structure and evolution. For several aminoacyl-tRNA synthetases it has recently been shown that the polypeptide chain comprises two repeating sequences that presumably have arisen by gene duplication and fusion [81]. Chemical evidence suggests that gene fusion probably explains the existence of the unusual 'double-headed' enzyme, aspartokinase-I: homoserine dehydrogenase-I of *E. coli* [82]. Promising attempts are now being made to study mechanisms of enzyme evolution in several systems [83,84] and arguments about adaption and neutral mutation are still eminently unsettled. The whole subject has been well reviewed at length elsewhere [85].

Concluding remarks

I have tried to trace the emergence of our current view of the protein chemistry of enzymes by means of examples that illuminate the stepping stones to our present knowledge. The path is not always uniquely defined, even with hindsight: the prodigious growth of interest in this area in the past 25 years has meant that other examples could no doubt have been chosen to illustrate some of the points. I can best relieve myself of this charge, at least in respect of recent work, by drawing attention to two comprehensive reviews of the topic [88,89]. Accounts of advances in the subject since 1969 are also published year by year elsewhere [90].

The first enzyme was crystallized, albeit impure, and shown to be a protein in 1926. In 1950, proof that a given protein had a unique amino acid sequence was still awaited. By 1970 the complete molecular structures of several enzymes were established and plausible reaction mechanisms could be discussed. As Synge so rightly observed in 1943 [91], advances in technique were the prerequisite for advances in our knowledge of protein chemistry. In the issue of *Brighter Biochemistry* dated May 1931 (the light-hearted house journal of the Department of Biochemistry in Cambridge from 1924–1931), a spoof examination paper for the undergraduates of twenty-five years later was offered by J.B.S. Haldane, himself a notable contributor to the early studies of enzymes [92]. The first question bears repetition here:

1. Write down the structural formula of human type C oxyhaemoglobin and briefly summarize the evidence on which it is based. (Structural formulae should be written stereoscopically. A stereoscope is provided.)

Although even today's undergraduates would not be called upon to enact the feat of memory playfully forecast by Haldane, prescient as he was, he might have been surprised at the dramatic change in our knowledge of enzyme chemistry over the past two decades.

I began this article by applying to enzymes some famous words from Keats' Endymion. It would be niggardly indeed not to take pleasure in the elegance of the enzyme, in concept and as we envisage it in action a century after Kühne named it. May I therefore conclude by quoting at slightly greater length?

A thing of beauty is a joy forever:
Its loveliness increases; it will never
Pass into nothingness

References

[1] Sumner, J. B. (1926) J. Biol. Chem. 69, 435–441.
[2] Willstätter, R. (1928) Untersuchungen über Enzyme, Vol 2. Springer, Springer.

[3] Northrop, J. H., Kunitz, M. and Herriott, R. M. (1948) Crystalline Enzymes, 2nd Edn, Columbia University Press.
[4] Dixon, M. and Webb, E. C. (1964) Enzymes, Longmans, London.
[5] Guilford, H. (1975) Ann. Rep. Prog. Chem. B, 71, 56–74.
[6] Cohn, E. J. and Edsall, J. T. (1943) Proteins, Amino Acids and Peptides Reinhold Publishing Co., New York.
[7] Chibnall, A. C. (1942) Proc. Roy. Soc. B 131, 136–160.
[8] Brand, E. L., Saidel, J., Goldwater, W. H., Kassell, B. and Ryan, F. J. (1945) J. Amer. Chem. Soc. 67, 1524–1533.
[9] Moore, S. (1972) in: Chemistry and Biology of Peptides: Proceedings of the Third Peptide Symposium (J. Meienhofer, ed.) Ann Arbor Science Publications Inc.
[10] Spackman, D. H., Stein, W. H. and Moore, S. (1958) Analyt. Chem. 30, 1190–1206.
[11] Benson, J. R. (1975) in: Instrumentation in Amino Acid Sequence Analysis (R. N. Perham, ed.) pp 1–39, Academic Press, London.
[12] Brown, J. P. and Perham, R. N. (1973) Eur. J. Biochem. 39, 69–73.
[13] Fox, S. W. (1945) Adv. Prot. Chem. 2, 156–177.
[14] Jansen, H. and Evans, E. A. (1935) J. Biol. Chem. 108, 1–9.
[15] Bergmann, M. and Niemann, C. (1938) J. Biol. Chem. 122, 577–596.
[16] Pauling, L. and Niemann, C. (1939) J. Amer. Chem. Soc. 61, 1860–1867.
[17] Svedberg, T. (1937) Nature 139, 1051–1062.
[18] Sanger, F. (1952) Adv. Prot. Chem. 7, 1–67.
[19] Consden, R., Gordon, A. H., Martin, A. J. P. and Synge, R. L. M. (1947) Biochem. J. 41, 596–602.
[20] Edman, P. (1950) Acta Chem. Scand. 4, 277–293.
[21] Sanger, F. (1959) Science, 129, 1340–1344.
[22] Smyth, D. G., Stein, W. H. and Moore, S. (1963) J. Biol. Chem. 238, 227–234.
[23] Canfield, R. E. (1963) J. Biol. Chem. 238, 2698–2707.
[24] Gray, W. R. and Hartley, B. S. (1963) Biochem. J. 89, 379–380.
[25] Hartley, B. S. (1970) Biochem. J. 119, 805–822.
[26] Hartley, B. S. (1964) Nature, 201, 1284–1287.
[27] Davidson, B. E., Sajgo, M., Noller, H. F. and Harris, J. I. (1967) Nature, 216, 1181–1185.
[28] Moon, K., Piskiewicz, D. and Smith, E. L. (1972) Proc. Nat. Acad. Sci. USA, 69, 1380–1383.
[29] Edman, P. and Begg, G. (1967) Eur. J. Biochem. 1, 80–91.
[30] Waterfield, M. D. and Bridgen, J. (1975) in: Instrumentation in Amino Acid Sequence Analysis (R. N. Perham, ed.) pp. 41–71, Academic Press, London.
[31] Laursen, R. A., Bonner, A. G. and Horn, M. J. (1975) in: Instrumentation in Amino Acid Sequence Analysis (R. N. Perham, ed.) pp. 73–110, Academic Press, London.
[32] Morris, H. R. and Dell, A. (1975) in: Instrumentation in Amino Acid Sequence Analysis (R. N. Perham, ed) pp. 147–191, Academic Press, London.
[33] Atlas of Protein Sequence and Structure (1972) Vol 5 (M. O. Dayhoff, ed.) National Biomedical Research Foundation, Maryland.
[34] Williams, J., Clegg, J. B. and Mutch, M. O. (1961) J. Mol. Biol. 3, 532–540.
[35] McLachlan, A. D. (1972) J. Mol. Biol. 64, 417–437.
[36] Herriott, R. M. (1947) Adv. Prot. Chem. 3, 169–225.
[37] Dickens, F. (1933) Biochem. J. 27, 1141–1151.
[38] Rapkine, L. (1938) Biochem. J. 32, 1729–1739.
[39] The Mechanism of Enzyme Action (1954) (W. D. McElroy and B. Glass eds.) The Johns Hopkins Press, Baltimore.
[40] Harris, I., Meriwether, B. P. and Park, J. H. (1963) Nature, 198, 154–157.
[41] Singer, T. P. (1948) J. Biol. Chem. 174, 11–21.
[42] Hopkins, F. G., Morgan, E. J. and Lutwak-Mann, C. (1938) Biochem. J. 32, 1829–1848.
[43] Barron, E. S. G. and Singer, T. P. (1943) Science, 97, 356–358.
[44] Crestfield, A. M., Stein, W. H. and Moore, S. (1963) J. Biol. Chem. 238, 2413–2419.
[45] Takahashi, K., Stein, W. H. and Moore, S. (1967) J. Biol. Chem. 242, 4682–4690.
[46] Anson, M. L. (1945) Adv. Prot. Chem. 2, 361–386.
[47] Lin, T. Y. and Koshland, D. E., Jr. (1969) J. Biol. Chem. 244, 505–508.
[48] The Biochemical Reactions of Chemical Warfare Agents: Biochemical Society Symposia No.2 (1948) (R. T. Williams, ed.) Cambridge University Press.
[49] Schaffer, N. K., May, S. C. and Summerson, W. H. (1954) J. Biol. Chem. 206, 201–207.
[50] Sanger, F. (1963) Proc. Chem. Soc. 76–83.
[51] Oosterbaan, R. A. and van Adrichem, M. E. (1958) Biochem. Biophys. Acta 27, 423–425.
[52] Blow, D. M., Birktoft, J. J. and Hartley, B. S. (1969) Nature, 221, 337–340.
[53] Shaw, E. (1970) Physiol. Revs. 50, 244–296.
[54] Tuppy, H. and Paléus, S. (1955) Acta Chem. Scand. 9, 353–364.
[55] Morse, D. and Horecker, B. L. (1968) Adv. Enzymol. 31, 125–181.
[56] Fischer, E. H. (1964) in: Structure and Activity of Enzymes: FEBS Symposia No.1(T. W. Goodwin, J. I. Harris and B. S. Hartley, eds.) pp. 111–120, Academic Press, London.
[57] Jones, E. T. and Williams, C. H. Jr. (1975) J. Biol. Chem. 250, 3779–3784.
[58] Christen, P. and Riordan, J. F. (1970) Biochemistry 9, 3025–3034.
[59] Anfinsen, C. B. (1973) Science, 181, 223–230.
[60] Lee, B. and Richards, F. M. (1971) J. Mol. Biol. 55, 379–400.
[61] Oppenheimer, H. L., Labouesse, B. and Hess, G. P. (1966) J. Biol. Chem. 241, 2720–2730.
[62] Cohen, L. A. (1970) in: The Enzymes (P. D. Boyer, ed.) vol.1, pp. 147–211. Academic Press, New York.
[63] Wold, F. (1967) Methods in Enzymology 11, 617–640.

[64] Husain, S. S. and Lowe, G. (1970) Biochem. J. 117, 333–346.
[65] Cuatrecasas, P., Fuchs, S. and Anfinsen, C. B. (1969) J. Biol. Chem. 244, 406–412.
[66] Davies, G. E. and Stark, G. R. (1970) Proc. Nat. Acad. Sci. USA, 66, 651–656.
[67] Hucho, F., Müllner, H. and Sund, H. (1975) Eur. J. Biochem. 59, 74–87.
[68] Cohlberg, J. A., Pigiet, V. P. Jr. and Schachman, H. K. (1972) Biochemistry, 11, 3396–3411.
[69] Perham, R. N. (1975) Phil Trans. Roy Soc. B, 272, 123–136.
[70] Shugar, D. (1952) Biochem. J. 52, 142–148.
[71] Burr, M. E. and Koshland, D. E. Jr. (1964) Proc. Nat. Acad. Sci. USA, 52, 1017–1024.
[72] Gutfreund, H. and McMurray, C. H. (1970) in: Chemical Reactivity and Biological Role of Functional Groups in Enzymes: Biochemical Society Symposia No. 31, (R.M.S. Smellie, ed.) pp. 39–47. Academic Press, London.
[73] Campbell, I. D., Dwek, R. A., Price, N. C. and Radda, G. K. (1972) Eur. J. Biochem. 30, 339–347.
[74] Ingram, V. M. (1957) Nature, 180, 326–328.
[75] Anfinsen, C. B. (1959) J. Cell. Comp. Physiol. 54 (Supplement 1), 215–220.
[76] A Discussion on the Structure and Functions of Proteolytic Enzymes (organized by D. C. Phillips, D. M. Blow, B. S. Hartley, and G. Lowe) Phil. Trans. Roy, Soc. B, 257, 63–266.
[77] Brew, K., Castellino, F., Vanaman, T. C. and Hill, R. L. (1970) J. Biol. Chem. 245, 4570–4582.
[78] Browne, W. J. North, A. C. T., Phillips, D. C., Brew, K, Vanaman, T. C. and Hill, R. L. (1969) J. Mol. Biol. 42, 65–86.
[79] McLachlan, A. D. and Shotton, D. M. (1971) Nature New Biology, 229, 202–205.
[80] Delbaere, L. T. J., Hutcheon, W. L. B., James, M. N. G. and Thiessen, W. E. (1975) Nature, 257, 758–763.
[81] Koch, G. L. E., Boulanger, Y. and Hartley, B. S. (1974) Nature, 249, 316–320.
[82] Véron, M., Falcoz-Kelly, F. and Cohen, G. N. (1972) Eur. J. Biochem. 28, 520–527.
[83] Betz, J. L., Brown, P. R., Smyth, M. J. and Clarke, P. H. (1974) Nature, 247, 261–264.
[84] Rigby, P. W. J., Burleigh, B. D. Jr. and Hartley, B. S. (1974) Nature, 251, 200–204.
[85] Williams, J. (1974) in: MTP International Review of Science: Biochemistry Series One. Vol.1 (H. Gutfreund, ed.) pp. 1–56 Butterworths, London.
[86] Gibbons, I. and Perham, R. N. (1974) Biochem. J. 139, 331–342.
[87] Gibbons, I., Yang, Y. R. and Schachman, H. K. (1974) Proc. Nat. Acad. Sci. USA, 71, 4452–4456.
[88] Knowles, J. R. (1974) in: MTP International Review of Science: Biochemistry Series One. Vol.1 (H. Gutfreund, ed.) pp. 149–184, Butterworths, London.
[89] Thomas, J. O. (1974) in: Companion to Biochemistry (A. T. Bull, J. R. Lagnado, J. O. Thomas and K. F. Tipton, eds.) pp. 87–138 Longmans, London.
[90] Specialist Periodical Reports: Amino Acids, Peptides and Proteins (1969) The Chemical Society, London.
[91] Synge, R. L. M. (1943) Chem. Revs. 32, 135–172.
[92] Haldane, J. B. S. (1930) Enzymes, Longmans, London.