THE CHEMISTRY OF PROTEOLYTIC ENZYMES

B. KEIL

Czechoslovak Academy of Sciences, Prague, Czechoslovakia

The study of the relationship between the genetic evolution, the structure and the function of enzymes is given a variety of names, such as molecular biology, molecular genetics, and evolutional biochemistry. In our Laboratory we started from organic chemistry, and we try to use the methods of chemistry as much as possible in this field of biologically active macromolecules.

About ten years ago, the main task of the chemist working with an enzyme was to determine the amino acid sequence of its polypeptide chain and to identify, by means of chemical modifications, the groups responsible for the biological activity. Then came the most valuable intervention of x-ray crystallographers who were able to trace the spatial arrangement of hundreds of amino acids forming the whole macromolecule. The answer to the question how the actual conformation of the macromolecule influences the activity of an enzyme has now become urgent. Today the chemist is no longer faced with the problem of following merely a reaction mechanism involving one or two functional groups, but has to consider a concerted effect of many groups in environments of hydrophobic and hydrophilic interactions. Such groupings fixed in a very rigid conformation form clefts and holes on the surface of a macromolecule. Recently an eminent biochemist said that we have to deal nowadays with a special branch of chemistry, the chemistry of holes. A part of this review can thus be considered as a contribution to this field.

It is about twenty years ago that Sanger's pioneer work established the complete amino acid sequence of insulin. This new information brought about the collapse of many old theories on the structure of proteins. The sequences of amino acids in the two chains of insulin showed no regularities or clear-cut pattern. Moreover, it was impossible to explain the hormonal activity from the primary structure. It became clear that if there existed any regularities or patterns in the structure of proteins, they would hardly be detectable without knowledge of the exact chemical architecture of many proteins.

There were two lines of chemical approach to the question whether there exist structural similarities between related proteins or not: to study the amino acid sequences either of functionally similar proteins from different organisms, or of functionally different proteins from the same cell.

Cytochrome C and later the haemoglobins from different organisms have attracted much interest. Studies on these substances have revealed the rules governing the differentiation of protein structures in the course of biological evolution of species.

Fourteen years ago, in our Laboratory in Prague, we started to follow the second line of approach, that is, we looked for related proteins formed in the same cell. These efforts were aimed at experimental verification of the

general theory of similarities in protein structures, formulated in 1954 by Šorm[1], who also participated in most of the work reported here.

Chymotrypsinogen and trypsinogen from beef pancreas were good subjects for such a study. They represent the major fraction of all enzymogens formed by the pancreas cell (*Table 1*). Their chemical composition and physicochemical properties differ considerably. *Table 2* shows our analyses[2, 3] of the amino acid composition of the two zymogens. The differences between the corresponding pairs of figures are considerable. The values for nine out of twenty amino acids differ by more than 30 per cent (e.g. in the case of threonine and tryptophan).

Table 1. Relative content of some enzymes in bovine pancreatic juice

Protein	%
Ribonuclease A	3
Chymotrypsinogen A	14
Chymotrypsinogen B	4
Trypsinogen	23
Amylase	5·5
Procarboxypeptidase A	29
Procarboxypeptidase B	6
Lipase	1·2

Table 2. Amino acid composition of trypsinogen (TG), chymotrypsinogen (CHTG) and pancreatic trypsin inhibitor (TI).

	TG	CHTG	TI		TG	CHTG	TI
Arg	2	4	6	Gln	12	10	1
Met	2	2	1	Cys	12	10	6
Glu	2	5	2	Ala	14	22	6
His	3	2	—	Leu	14	19	2
Phe	3	6	4	Ile	15	10	2
Trp	4	8	—	Lys	15	14	4
Pro	8	9	4	Asn	18	14	3
Asp	8	9	2	Val	18	23	1
Thr	10	23	3	Gly	25	23	6
Tyr	10	4	4	Ser	34	28	1
				Total	229	245	58

Our first comparison of cystine-containing peptide fragments derived from the two zymogens[4–6] was very encouraging—the fragments were similar. However, they represented only very small parts of the two molecules and therefore we decided at that time to determine the whole primary structures of the two proteins (*Table 3*). The determination of the full amino acid sequence of chymotrypsinogen A in our Laboratory[7] was largely due to Meloun, Vaněček, Kostka, and Kluh. The sequence was simultaneously elucidated by Hartley and coworkers[8] in Cambridge, and large parts of it were also determined by Desnuelle and coworkers[9] in Marseille. *Table 3* also shows the amino acid sequence of the second protein

Table 3. Comparison of the primary structure of chymotrypsinogen and trypsinogen

20

Cys.Gly.Val.Pro.Ala.Ile.Gln.Pro.Val.Leu.Ser. Gly.Leu.Ser. Arg.Ile.Val.Asn.Gly.Glu.Glu.Ala.Val.Pro.Gly. Ser. Trp.Pro.Trp.Gln.
Val.Asp.Asp.Asp.Asp.Lys. Ile.Val.Gly.Gly.Tyr.Thr.Cys.Gly.Ala.Asn.Thr.Val.Pro. Tyr.Gln.

40 60

.Val.Ser.Leu.Gln.Asp.Lys.Thr.Gly.Phe.His.Phe.Cys.Gly.Gly.Ser.Leu.Ile.Asn.Glu.Asn.Trp.Val.Val.Thr.Ala.Ala.His.Cys.Gly.Val.
.Val.Ser.Leu.— .Asn.—. Ser. Gly.Tyr.His.Phe.Cys.Gly.Gly.Ser.Leu.Ile.Asn.Ser. Gln.Trp.Val.Val.Ser. Ala. Ala.His.Cys.Tyr.Lys.

80

.Thr.Thr.Ser.Asp.Val.Val.Val. Ala.Gly.Glu.Phe.Asp.Gln.Gly.Ser. Ser. Ser. Glu.Lys. Ile. Gln.Lys.Leu.Lys.Ile. Ala.Lys.Val.Phe.Lys.
.Ser. Gly.Ile. Gln.Val.Arg.Leu.Gly.Gln.Asp.Asn. Ile. Asn.Val.Val.Glu.Gly.Asn.Gln.Gln.Phe.Ile. Ser. Ala.Ser.Lys.Ser. Ile. Val.His.

100 120

.Asn.Ser.Lys.Tyr.Asn.Ser.Leu.Thr.Ile. Asn.Asn.Asp.Ile.Thr. Leu.Leu.Lys.Leu.Ser.Thr.Ala.Ala.Ser.Phe.Ser. Gln.Thr.Val.Ser.Ala.
.Pro.Ser.—.Tyr.Asn.Ser.Asn.Thr.Leu.Asn.Asn.Asp.Ile.Met.Leu.Ile. Lys.Leu.Lys.Ser. Ala.Ala.Ser.Leu.Asn.Ser. Arg.Val.Ala.Ser.

140

.Val.Cys.Leu.Pro.Ser. Ala.Ser.Asp.Asp.Phe.Ala.Ala.Gly.Thr.Thr.Cys.Val. Thr.Thr.Gly.Trp.Gly.Leu.Thr.Arg.Tyr.Thr.Asn.Ala.Asn.
.Ile. Ser. Leu.Pro.Thr.Ser.Cys.Ala. —. —.Ser. Ala.Gly.Thr.Gln.Cys.Leu.Ile. Ser. Gly.Trp.Gly.Asn.Thr.Lys. Ser. Ser. Gly.Thr.Ser.

160

.Thr.Pro.Asp.Arg.Leu.Gln.Gln.Ala.Ser. Leu.Pro.Leu.Leu.Ser.Asn.Thr.Asn.Cys.Lys.Lys. —.Tyr.Trp.Gly.Thr.Lys.Ile.Lys. Asp.Ala.
.Tyr.Pro.Asp.Val.Leu.Lys. Cys.Leu.Lys.Ala. Pro.Ile. Leu.Ser.Asn.Ser. Ser. Cys.Lys. Ser. Ala.Tyr.Pro.Gly. —.Gln.Ile.Thr.Ser. Asn.

180 200

.Met.Ile. Cys.Ala.Gly.Ala. Ser. —. —.Gly.Val.Ser. Ser.Cys.Met.Gly.Asp.Ser.Gly.Gly.Pro.Leu.Val.Cys. —.Lys.Lys.Asn.Gly.Ala.
.Met.Phe.Cys.Ala.Gly.Tyr.Leu.Glu.Gly.Gly.Lys. Asp.Ser.Cys.Gln. Gly.Asp.Ser.Gly.Gly.Pro.Val.Val.Cys.Ser.Gly.Lys. —. —. —.

220

.Trp.Thr.Leu.Val.Gly.Ile.Val.—.Ser.Trp.Gly.Ser. Ser.Thr.Cys. — .Ser. Thr.Ser. Thr.Pro.Gly.Val.Tyr.Ala. Arg.Val.Thr.Ala. Leu.
. —. — .Leu.Gln.Gly.Ile.Val.—.Ser.Trp.Gly.Ser.Gly. —.Cys.Ala.Gln.Lys. Asn.Lys. Pro.Gly.Val.Tyr.Thr.Lys. Val.Cys.Asn.Tyr

240

.Val.Asn.Trp.Val.Gln.Gln.Thr.Leu.Ala.Ala.Asn
.Val.Ser. Trp.Ile. Lys. Gln.Thr.Ile. Ala.Ser. Asn

studied, namely trypsinogen. This formula was established by Mikeš, Tomášek and Holeyšovský[10] in our Laboratory, and it was corroborated independently by Neurath[11] in Seattle.

There were three points of interest with respect to the primary structure of the two zymogens. Firstly, the mechanism of activation to the active enzymes, chymotrypsin and trypsin. Secondly, the chemical nature of the active groupings responsible for the enzymatic activity. Thirdly, the general similarities in the structures of the two enzymes compared.

The changes in the molecules of the zymogens were studied mostly in the laboratories of Neurath and of Desnuelle[12–14]. *Figure 1* shows the comparison of the two activation schemes. In both cases the first change which gives rise to the active enzyme is the hydrolysis of the peptide bond adjacent to a basic amino acid residue which liberates a terminal hydrophobic isoleucyl–valyl sequence. We will consider this change later when dealing with three dimensional structures.

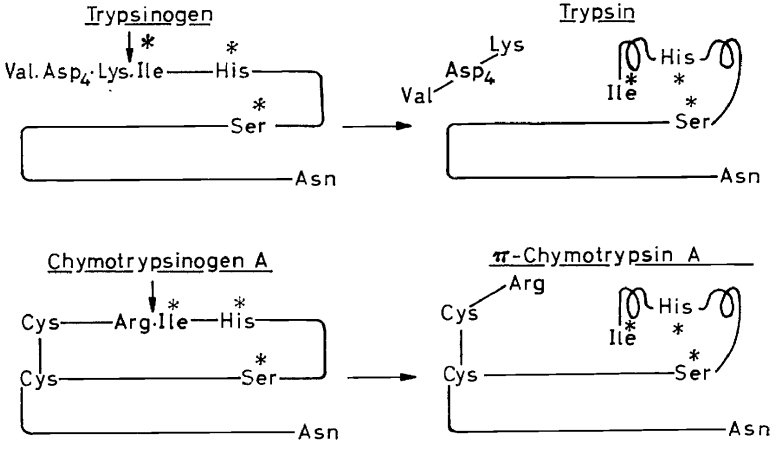

Figure 1. Activation of chymotrypsinogen and trypsinogen

Now, what is the chemical composition of the active part of the molecule, responsible for the enzymatic activity? By selective substitution reactions it has been established that two residues are of prime importance, namely a serine and a histidine residue. Without going into experimental details, attention should be directed to the fact that there again exists a far-reaching similarity in the amino acid sequences, especially between those parts of the two enzymes which are of crucial importance for their activity (*Table 4*). Later it was found that this structural similarity of the active sites is not confined merely to trypsin and chymotrypsin. *Table 5* shows a list of proteinases and esterases of very different origin which show analogies in the structures of their active centres. From this information we can also learn that there is a characteristic deviation from the very similar common picture in the case of bacterial proteolytic enzymes, where the active serine is still present but its environment is quite different.

The similarities in trypsinogen and chymotrypsinogen are not restricted to the immediate vicinity of their active sites. *Table 3* shows how far the

similarities go along the whole polypeptide chain. The identical sequences are underlined. Out of 229 amino acid residues in trypsinogen 101 residues occupy positions analogous to those in chymotrypsinogen, that is about 44 per cent. The similarities would be, of course, very much greater if replacements of amino acids due to one-step mutation in a triplet were considered. Then the analogy rate would rise to 61 per cent. These far-reaching analogies in the primary structure of the two polypeptide chains obviously mirror the development of the two enzymes from a common ancestor protease.

Table 4. Similarities between trypsin (T) and chymotrypsin (CH) connected with their function

(T)	..Tyr.Cys.*His.Ala.Ala.Ser..
(CH)	..Gly.Cys.His.Ala.Ala.Thr..
(T)	..Ser.Cys.Gly.Gly.Asp.*Ser.Gly.Gly.Pro.Val..
(CH)	..Ser.Cys.Met.Gly.Asp.Ser.Gly.Gly.Pro.Leu..

Table 5. Active centres of some hydrolases

	↓
Trypsin	.Ser.Cys.Gln.Gly.Asp.Ser.Gly.Gly.Pro.Val.Val.Cys.
Chymotrypsin	.Ser.Cys.Met.Gly.Asp.Ser.Gly.Gly.Pro.Leu.Val.Cys.
Elastase	.Gly.Cys.Gln.Gly.Asp.Ser.Gly.Gly.Pro.Leu.His.Cys.
Thrombin	.Asp.Ser.Gly.
Ali-esterase	.Gly.Glu.Ser.Ala.Gly.Gly.
Pseudocholin esterase	.Phe.Gly.Glu.Ser.Ala.Gly.
Acetylcholin esterase	.Glu.Ser.Ala.
Yeast peptidase	.Glu.Ser.Val.
Subtilisin	.Asn.Gly.Thr.Ser.Met.Ala.
Aspergillopeptidase	.Gly.Thr.Ser.Met.Ala.
	↑

From studies on other groups of proteins it became evident that the three dimensional arrangement of the polypeptide chains is determined by its amino acid sequence. In the case of trypsin and chymotrypsin we know now about the far reaching similarity in their sequences. How far can we extend the assumption that not only the primary structure, but also the three dimensional arrangement of the two enzymes will be analogous?

One fact stands out against this hypothesis. There is a long sequence of about 26 residues in the middle of the chain which is entirely different in the two enzymes. This can, of course, change completely the relative positions of very similar segments of the molecule. On the other hand, there are three indications in favour of a similar three dimensional pattern.

Firstly, the sequences around the active serine and histidine, forming two independent loops, are similar. Therefore it can be assumed that a parallelism in activity would also require a parallelism in conformation of these loops.

Secondly, we know the relative positions of the disulphide bridges. In chymotrypsinogen there are five disulphide bridges, in trypsinogen six. In *Figure 2* their relative positions are indicated by brackets. Four out of five bridges occupy analogous positions. This indicates clearly that not only the active sites but also large segments including disulphide bridges, and consequently their three dimensional arrangements, are similar[15].

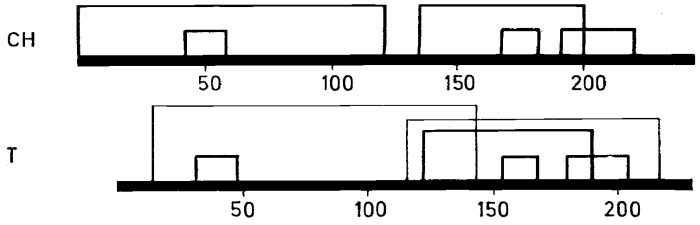

Figure 2. Disulphide bridges in chymotrypsin and trypsin

The third indication in favour of a similarity in spatial arrangement is the relative location of certain so called "invariant" amino acids in the polypeptide backbone. Some amino acid residues, like glycine or proline, were found in analogous proteins to occupy "invariant", identical, positions. The explanation for this phenomenon can be the vital importance of these residues for the overall shape of the polypeptide chain, and therefore for the biological function. When we consider our two enzymes, then we find that out of 23 glycines in chymotrypsinogen 17 are in positions identical with those in trypsinogen. The numbers for prolines are 6 out of 9, and as regards four tryptophans in trypsinogen all are in exactly identical positions along the polypeptide chain.

All this was, of course, only a very modest exploration of an unknown spatial model. Without direct knowledge of the three-dimensional arrangement of at least one of the two enzymes there was little hope of further progress.

The great break came from the laboratory of Blow in Cambridge when they reported the three-dimensional structure of tosyl-α-chymotrypsin[16]. They showed the general orientation of the polypeptide chain, the relative positions of the disulphide bridges and of three residues, namely histidine No. 57, tyrosine No. 171, and serine No. 195.

When we interpreted in our Laboratory the general shape of the model given as a schematic drawing by the Cambridge group, and placed all amino acid residues of the known primary structure along the chain, we obtained a first rough approximation of the relative position of all residues. We also built a similar model of trypsin, in which we maintained the same orientation of the polypeptide backbone as in chymotrypsin. Immediately it became evident that the two disulphides which are found in trypsin, but not in chymotrypsin, fit beautifully into the model, because their half-cystine components are located in adjacent loops and can be joined without difficulty. In the schematic drawing (*Figure 3*) the arrows mark the position of these two bridges in trypsin[17]. The idea of conformational homology between chymotrypsin and trypsin was formulated also by Neurath[18] and by Blow[19]. No wonder that groups who have worked on the same projects for a couple of years have logically arrived at the same conclusions.

This fact, together with other observations on the two models, which will be mentioned later, indicate that analogous structures in two proteins can be maintained in a given conformation by different types of bonds.

Once the orientation of the polypeptide backbone had been given, and the primary sequence determined, we tried to build a more precise model of

Figure 3. Hypothetical three-dimensional schemes of chymotrypsin (left) and trypsin (right). Hydrophobic residues are in black, the positions of the two additional disulphide bridges in trypsin are marked by arrows.

chymotrypsin including the amino acid residues (*Figure 4*; given in the section printed on art paper), using a kit similar to that employed by Phillips[20] for lysozyme, which can be combined with space-filling hydrogen and oxygen atoms.

Again this can be considered only as an approximation. When we constructed this model we had available only the facts which had already been published. I recently had the opportunity of visiting Dr. Blow in his laboratory in Cambridge and comparing our ideas with his new unpublished data on x-rays measurements of chymotrypsin. We have now to revise our model in the light of these new facts, but, generally speaking, these changes are not in contradiction with the results obtained by chemical means, and they do not affect the hypothesis of the structural analogy between chymotrypsin and trypsin.

The second structure given in *Figure 4* represents an attempt to create a model of trypsin. It is, of course, a tentative approach of temporary value. It was based uniquely on the analogy between trypsin and chymotrypsin, on the known amino acid sequence. and on general considerations concerning the orientation of hydrophilic and hydrophobic residues. However, before x-ray data on trypsin become available, it can serve as a working hypothesis as will be shown later. On the other hand, we await with great interest the results of x-ray studies on chymotrypsin using a higher resolution. A comparison will show how far we are from reality, and whether one is justified at all in reconstructing a protein model knowing merely the orientation of the polypeptide backbone and the amino acid sequence.

One of the questions which may be asked is what is the molecular volume? In the case of another protein whose three dimensional structure is well known, myoglobin, the interior of the molecule is so tightly space-filled that only about 3 molecules of water can be embedded. According to data obtained by x-ray measurements of chymotrypsin[16] the length of the coordinates is 45, 38, and 35 Ångströms. When we take into account the sum of the partial volumes of all amino acid residues, then the overall density is less than the value for a fully space-filled model. In other words, it seems that chymotrypsin has a sort of Emmenthaler cheese molecule with holes or channels on its surface.

It is premature to discuss in detail the orientation of the hydrophobic and hydrophilic side chains in those sections where the polypeptide chain forms the surface of the protein, but after the construction of our models we have found that hydrophobic residues are situated predominantly in the interior (*Figure 3*). The nonpolar residues of different chains can thus form a sort of oil drop with a stabilizing effect.

The distribution of charges within the model is given in *Figure 5*. A close inspection of the model shows that practically all charged groups are on the surface. This does not mean of course, that they are all entirely free; they can engage in ionic interactions. For a check on the validity of the model proposed for trypsin it was very important to know whether the location of charged amino acid residues in trypsin in those parts which differ from the sequences in chymotrypsin complies with the general distribution of charged residues. There are 15 such replacements and we found all the charged groups on the surface of the trypsin model, a fact which favours its reliability.

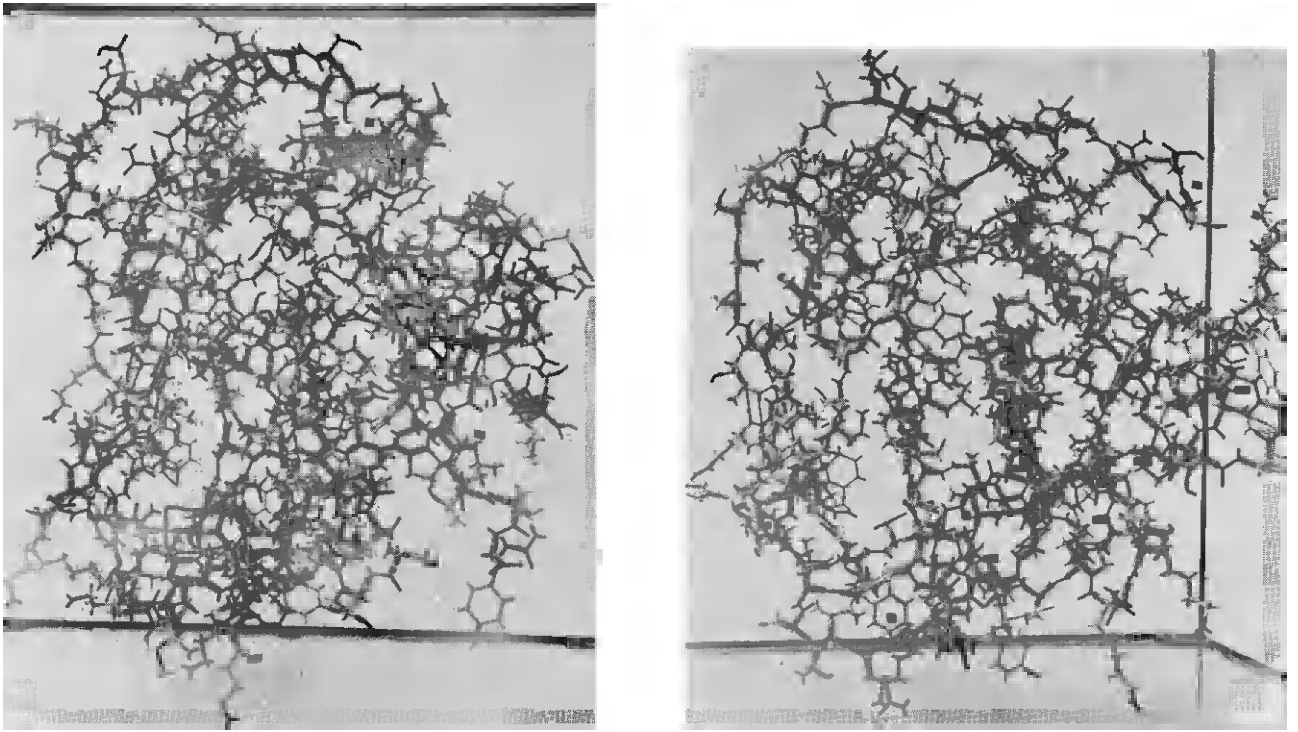

Figure 4. Hypothetical three-dimensional models of chymotrypsin (left) and trypsin (right).

Figure 11. Hypothetical three-dimensional model of the pancreatic trypsine inhibitor

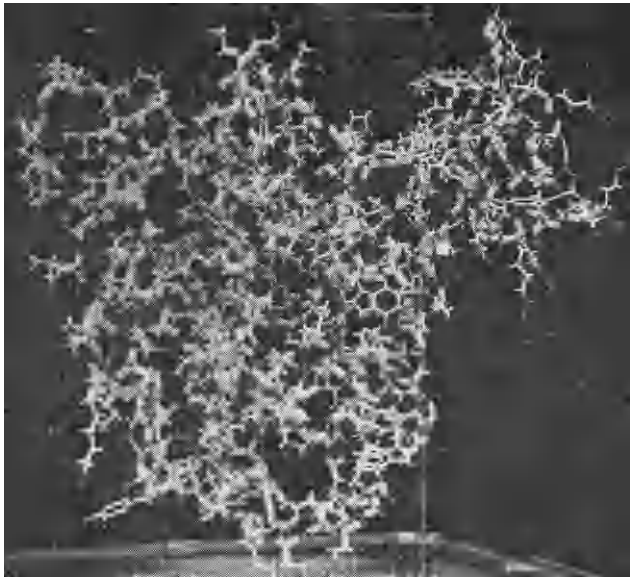

Figure 13. Hypothetical three dimensional model of the trypsin–trypsin inhibitor complex.

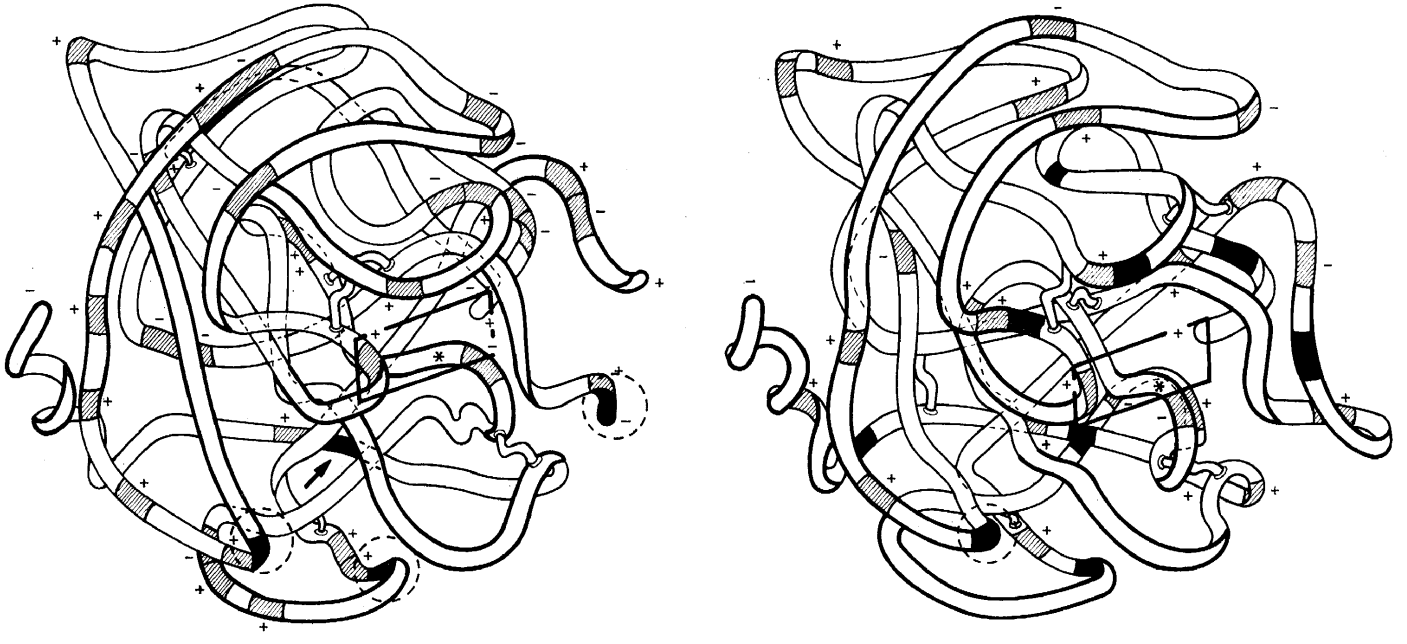

Figure 5. The distribution of charged groups in the hypothetical scheme of chymotrypsin (left) and trypsin (right). Tyrosine residues accessible to nitration in chymotrypsin are marked by circle, buried tyrosine by arrow.

There is a great difference between chymotrypsin and trypsin in the content of tyrosine residues (*Figure 5*). In chymotrypsin there are 4 tyrosines whereas in trypsin there are 10. Together with Drs. Shlyapnikov and Meloun, we have in our Laboratory checked the distribution of the tyrosine residues in chymotrypsinogen and chymotrypsin by nitration using tetranitromethane[21].

Knowing the distribution of tyrosine residues, we expected that tyrosine No. 228 in chymotrypsin would not react because it is buried in the interior, whereas the other three are located on the surface. After the isolation and analyses of the nitrated and non-nitrated peptides, Shlyapnikov and Meloun were able to demonstrate that tyrosine residue No. 228 does not react with tetranitromethane, according to our expectation. However, the reactivity of the remaining three residues was different in chymotrypsinogen and in chymotrypsin. All three tyrosines are accessible to the reagent in chymotrypsinogen, but in chymotrypsin tyrosine No. 94 does not react. During the activation process there must occur a conformational change which involves the loop containing this tyrosine.

Parallel experiments with trypsinogen and trypsin were carried out in our Laboratory by Holeyšovský[22]. Only a limited number of tyrosine residues reacted with tetranitromethane, out of ten only four in the case of trypsinogen, and only three in the case of trypsin. Although the experimental work has not been completed yet, we conclude that here again we meet with preferential substitution of tyrosine No. 82 in the bottom loop of trypsinogen, and that the majority of tyrosine residues are buried.

Now, how can we explain the mechanism of activation of trypsinogen to trypsin? According to the scheme shown in *Figure 1*, an *N*-terminal hexapeptide will be split off, containing four aspartic acid residues and one lysine. With respect to the general topology of the trypsin molecule it can be proposed that this strongly polar hexapeptide is held in position by ionic interactions. As soon as the lysine bond is split, the new hydrophobic isoleucyl–valyl terminal liberated swings to the interior of the molecule and its free amino group No. 7 enters into a new interaction with the aspartic acid residue No. 182 adjacent to the active serine.

And so we come to the structural basis of the proteolytic activity. I will mention only two schemes proposed for chymotrypsin. One of them was proposed by Hartley[23]; in this scheme two histidines and one serine are involved. The second scheme proposed by Bender[24], invokes only one histidine and the active serine (*Figure 6*). This second scheme seems more plausible, because according to the x-ray measurements[16] the second histidine is too far away for a direct interaction.

The characteristic grouping of amino acid residues involved in the activity was detected directly in the x-ray patterns by Blow and his coworkers[16] and it is shown in *Figure 7*. The crucial amino acids of the active centre—histidine No. 57, serine No. 195, aspartic acid No. 194, and isoleucine No. 16—lie in this region. This pattern shows clearly the very specific role played by the aspartic acid residue. The negatively charged carboxyl group forms an ion pair with the positively charged amino terminal group of isoleucine No. 16. This ionic interaction presumably stabilizes the active configuration of the enzyme.

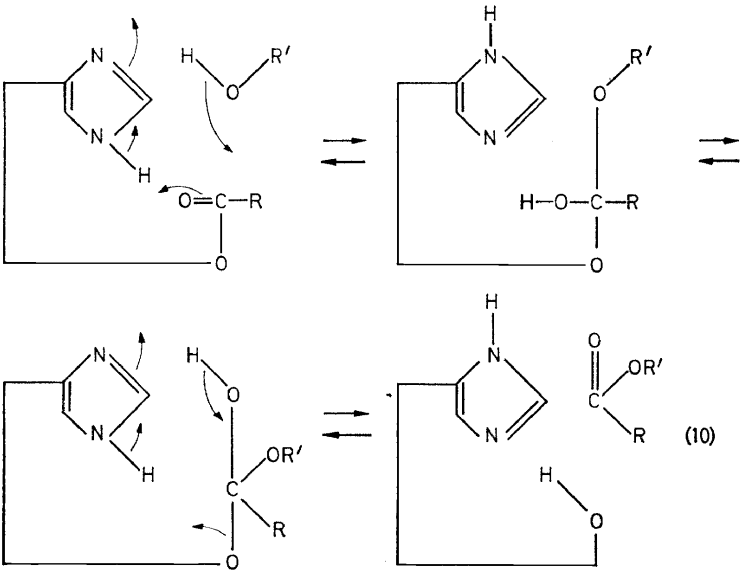

Figure 6. Schematic representation of chymotrypsin action[24].

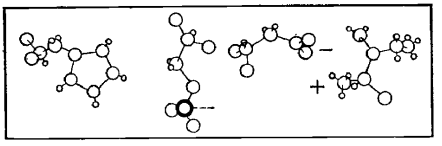

Figure 7. Active site of α-chymotrypsin (after Matthews *et al.*[16]).

It was quite tempting to try to build a model of the active site section of our three dimensional model of chymotrypsin, with space-filling hydrogen and oxygen atoms. On this model, a cleft with the active histidine and serine was clearly visible on one side, the interaction of aspartic acid with the amino end of isoleucine on the other side. As regards the loop containing both active serine and aspartic acid, it can probably swing from the left to right, the "left" configuration existing in the zymogen. During the activation the attraction of aspartic acid by the *N*-terminal isoleucine brings about the shift to the "right" configuration, thus establishing the correct conformation of histidine and serine residues necessary for their concerted action. The cleft in our model corresponds approximately to the dimensions 11 × 9·5 × 6·5Å of the active site, which were found by Erlanger[25] in his study of specific substrates with different lengths of side chains.

Here we have the "machinery" required for the catalytic action; but where is the moiety responsible for the specific binding of the substrate, the binding site? One of the prerequisites for a binding site in chymotrypsin would be its hydrophobicity, the second one is its position close to the active site. Only a few residues in the model can comply with these conditions.

We suspect that the role is filled by one of the two phenylalanines No. 39 or 41 which look in the space for a substrate like an antenna. But we have to wait for the exact numbers which should be obtained from direct x-ray measurements of the enzyme–substrate complex. Chemistry can hardly offer more information in this respect.

It is only a guess that in the case of trypsin the active site could have a similar form. Instead of phenylalanine residue No. 39 in chymotrypsin here we have tyrosine residue No. 28. Trypsin binds specifically substrates with a strongly positive charge, like arginine or lysine residues. It would be more comprehensible to find in close vicinity to the active site a strongly negative group of glutamic or aspartic acid.

In the case of low molecular weight synthetic substrates trypsin cleaves preferentially the bond adjacent to arginine. A study was recently undertaken in our laboratory by Keilová[26] in an effort to compare the relative rates of cleavage of arginine- and lysine-vasopressin by trypsin. These compounds have all the features of polypeptide substrates, and are identical except for the side chain containing the only basic amino acid. *Figure 8* shows that arginine–vasopressin is preferentially cleaved throughout the whole pH-range. Although for the binding of the substrate an interaction of the arginine side chain with tyrosine phenolic group may be proposed, we are not too happy with this explanation. It cannot explain the remarkable specificity of trypsin and many other data. We must wait for direct experimental evidence on this point.

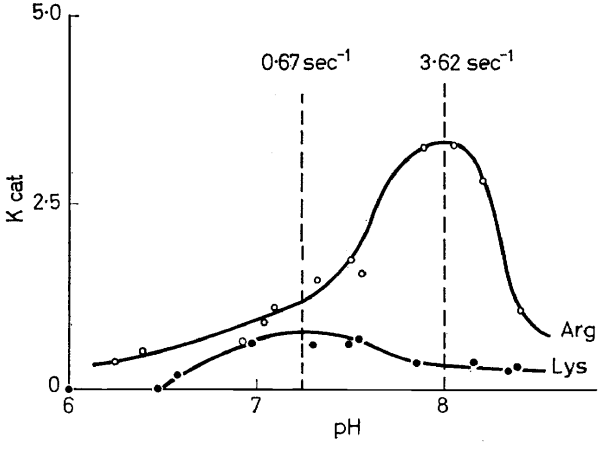

Figure 8. Dependence of catalytic constants (K_{cat} sec^{-1}) on pH of oxidized Arg-vasopressin (Arg) and Lys-vasopressin (Lys) substrates cleaved by trypsin.

Before starting with the second part of my lecture, I would like to summarize that for the proper function of a proteolytic enzyme several configurational factors are indispensable—the active site, the binding site, and a system of supports or springs which fixes the optimal conformation. The sequence studies show that five members of the trypsin family, namely bovine and pig trypsin, bovine chymotrypsin A and B and pig elastase, contain identical amino acid residues in segments around the active sites, which, together

with identical disulphide bridges, give to this machinery an adequate rigidity (*Figure 9*).

Now I am coming to the second part of the trypsin story, to that of its interaction with the pancreatic trypsin inhibitor. Since the time that the systematic crystallization of pancreas proteins was carried out by Kunitz[27] 18 years ago, it has been well known that trypsin forms a crystalline, very firm complex with a polypeptide inhibitor from pancreas. This compound seemed to us to be a very promising model for a study of the specific interactions of two high-molecular weight polypeptide structures. To furnish the first necessary prerequisite it was decided to elucidate the amino acid sequence of the inhibitor.

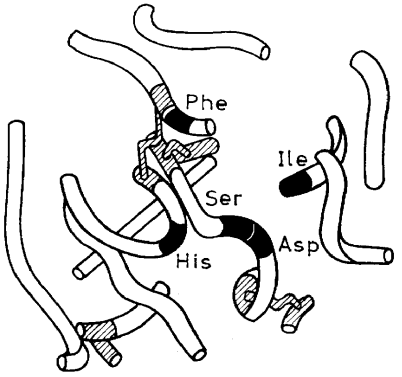

Figure 9. Identical fragments and disulphide bridges in the direct environment of the active site of bovine and pig trypsin, bovine chymotrypsin A and B, and bovine elastase.

The molecular weight of the inhibitor is 6500. It contains no histidine and no tryptophan (*Table 2*). The experimental work on the determination of the complete amino acid sequence (*Figure 10*) was done by Dlouhá and Pospíšilová in our Laboratory[28]. This sequence was also determined independently in the laboratory of Laskowski by Kassel and coworkers[29], and in the laboratory of Acher[30].

At first glance the inhibitor appears to be a very rigid structure because its single chain of 58 residues is cross-linked by three disulphide bridges. For a better understanding of its inhibitory action on trypsin a three dimensional model is highly desirable. We could not wait for exact information from x-ray crystallographers, and the only approach left was therefore the one via chemical and physicochemical topography. This task was fulfilled largely by Meloun in our laboratory.

There are four tyrosines in the molecule (see *Figure 10*). The inhibitor was therefore nitrated by tetranitromethane and the substituted and free tyrosines were identified by analyzing peptides isolated from a partial hydrolysate[31]. It was found that the tyrosine residue nitrated first is the one in position 10, and that next comes tyrosine No. 21. The two remaining tyrosine residues No. 23 and No. 35 are resistant to nitration. The substitution of tyrosine No. 10 results in a change in the Cotton effect which can be accounted for by an interaction of this tyrosine with another aromatic residue, presumably that of phenylalanine No. 33 in the native inhibitor. Additional

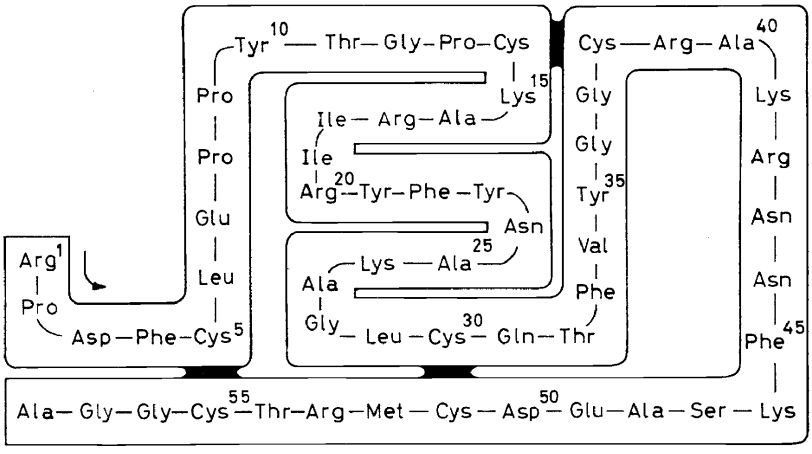

Figure 10. Primary structure of the pancreatic trypsin inhibitor.

substitution of tyrosine No. 21 leading to a dinitro-derivative does not affect the Cotton effect. The residual Cotton effect of the molecule can be ascribed to an interaction of the two "buried" tyrosines No. 23 and No. 35.

The next reaction used by Meloun was the glycination by glycine methyl ester in the presence of a carbodiimide derivative, a reaction described by Hoare and Koshland[32]. The experiments have not been completed, but the results clearly show that the residues of aspartic acid No. 3 and glutamic acid No. 7 are substituted readily, the carboxyl of the C-terminal alanine only at a limited rate, whereas one of the two residues at positions No. 49 and No. 50 is resistant[33]. It is assumed that these residues may be linked by hydrogen bonds to the two "resistant" tyrosines.

It has been found independently in the laboratory of Laskowski[34] and in our laboratory[35] that the disulphide bridge No. 14–No. 38 can be preferentially reduced. When the intact inhibitor is subjected to O.R.D. measurement it shows a marked Cotton effect which disappears only after a total reduction of all disulphide bridges. There is no detectable change in the Cotton effect after the preferential reduction of one disulphide bond, which means that no dramatic change has occurred in conformation of the polypeptide chain.

All these results, substitution experiments together with the known amino acid sequence, stimulated Meloun's efforts to build a hypothetical three-dimensional model of the inhibitor (*Figure 11*; given in the section printed on art paper). Because of the three disulphide bridges there is not an unlimited freedom in the arrangement of the polypeptide backbone. In the proposed model this arrangement is consistent with the results of studies on the reactivity of individual residues, and most of the polar groups are oriented towards the surface of the molecule.

Now, having the three-dimensional models of both trypsin and trypsin inhibitor, we can try to throw light on the problem of their interaction.

The first information came from studies based on chemical modification. Chauvet and Acher[36] have shown that lysine No. 15 (*Figure 10*) plays an

important role in the interaction of the inhibitor with trypsin. The reduction of the disulphide bond No. 14–No. 38 did not result in the loss of activity[34,35,37]. Meloun in our laboratory has shown that after the conversion of the half-cystine residue No. 14 into a neutral derivative by reaction with acrylonitrile, the product remains active[33]. On the other hand, conversion of the same residue into the carboxymethyl derivative leads to an inactive product[35]. This finding can be explained by postulating that there is an interaction between the negatively charged carboxyl of carboxymethyl-cysteine No. 14 and the positively charged lysine No. 15, which as a result of this interaction can no longer interact with a negatively charged group in trypsin.

The carboxyl groups of aspartic acid No. 3 and glutamic acid No. 7 can be left out of consideration, because the glycinated product is active. The same holds true for the nitro-derivatives. The inhibitor with both tyrosines nitrated retains its full inhibitory effect. This evidence was corroborated in a study on the nitration of the complex of the inhibitor with the enzyme. Even in this complex both tyrosines, No. 10 and No. 21 were readily nitrated and, after subsequent dissociation of the complex, the modified inhibitor was able to reassociate with the enzyme.

Additional information on the site of interaction of the inhibitor with trypsin was furnished by studies on the trypsin moiety. Since the trypsin–trypsin inhibitor complex is stable in alkaline media, Dlouhá in our Laboratory tried to split it proteolytically under conditions which did not bring about its dissociation[38]. By analyzing the combined tryptic–chymotryptic digest of the complex she was able to identify a fragment of trypsin, which remains bound to the inhibitor and represents the contact site of trypsin (*Figure 12*). It contains all the residues, shown in *Figure 9*, which were found to be vitally important for the enzymatic activity with the exception of isoleucine No. 16.

Obviously, the regions of the molecule which are not digested at all when

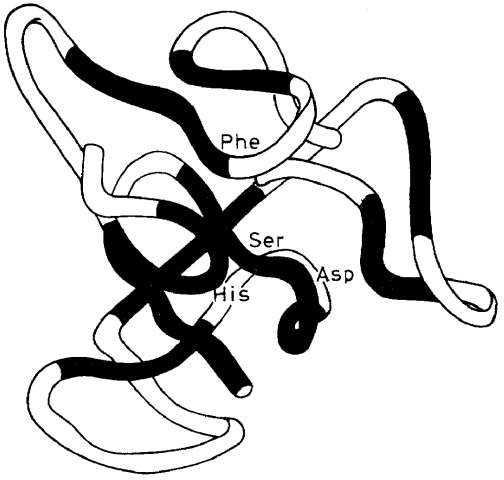

Figure 12. A fragment of trypsin molecule bound by the pancreatic trypsin inhibitor. In black are areas identical with the drawing on *Figure 9.*

trypsine is bound in the complex are those which are sterically protected by the inhibitor or buried in the interior. The complex of the inhibitor with this fragment which represents about one half of the trypsin molecule, can be dissociated, and after the peptide is set free the inhibitor becomes fully active. On the other hand, the dissociation and the release of the fragment from the inhibitor moiety is irreversible. This indicates, that trypsin can interact with the inhibitor, only if its three-dimensional structure is intact. The bonds which hold together the complex, however, are so stable that they are not broken even if the complex is subjected to profound proteolytical cleavage. Hydrogen bonds can be excluded as possible main forces because the complex is stable even in high concentrations of salts.

The result of amino acid and end-group analyses of the inhibitor isolated from the proteolyzed complex indicate that the inhibitor is not split in the process of complex formation. Therefore the mechanism of inhibition is obviously different from the mechanism proposed for the interaction of soy bean inhibitor with trypsin by Fírkenstadt and Laskowski[39]. The resistance of the pancreatic inhibitor to proteolytic enzymes is due to the rigid three-dimensional structure of its molecule since the inhibitor remains intact even when it is bound in the complex and its inhibitory effect is therefore neutralized.

When we take into account all the information at hand, what will be the probable shape of the two interacting molecules? In *Figure 13* (given in the section printed on art paper) is presented our idea as to how it works. We suppose that the inhibitor does not interact directly with the active histidine and serine, but with three negatively charged groups, probably aspartic acid residues (No. 139, 178 and 182?); and that three basic residues of the inhibitor, namely lysine No. 15, arginine No. 20, and arginine No. 1 are their partners in these ionic interactions.

The main effect of the inhibitor would be due therefore, to the competition of one of its basic amino acids with the amino group of trypsin isoleucine No. 7 for the aspartic acid residue No. 182. As a result of this shift the conformation of the active site is deteriorated.

There remain many known experimental facts which should be discussed in this connection, such as the lack of reactivity of the complex with diisopropylfluorophosphate, or the interaction of the inhibitor with trypsinogen, and many other questions which are being examined or which should be examined experimentally in the near future. Moreover we hope that the time will soon come when the x-ray crystallographers will show us how far we are from reality.

References

[1] F. Šorm. *Coll. Czech. Chem. Commun.* **19**, 1003 (1954).
[2] Z. Zmrhal. *Coll. Czech. Chem. Commun.* **27**, 2934 (1962).
[3] Z. Zmrhal. *Coll. Czech. Chem. Commun.* **28**, 2356 (1963).
[4] B. Keil and F. Šorm. *Coll. Czech. Chem. Commun.* **19**, 1018 (1954).
[5] B. Keil and F. Šorm. *Coll. Czech. Chem. Commun.* **24**, 1558 (1959).
[6] B. Keil and F. Šorm. *Coll. Czech. Chem. Commun.* **24**, 1573 (1959).
[7] B. Meloun, I. Kluh, V. Kostka, L. Morávek, Z. Prusík, J. Vaněček, B. Keil, and F. Šorm. *Biochim. Biophys. Acta* **130**, 543 (1966).
[8] B. S. Hartley, R. Brown, D. L. Kauffman, and L. B. Smillie. *Nature* **207**, 1157 (1965).

[9] S. Maroux, M. Rovery, and P. Desnuelle. *Feder. Europ. Biochem. Soc., Symposium No. 1*, p. 63, Goodwin, Harris, Hartley Ed., London 1964.
[10] O. Mikeš, V. Tomášek, V. Holeyšovský, and F. Šorm, *Coll. Czech. Chem. Commun.* **32**, 655 (1967).
[11] K. A. Walsh, and H. Neurath. *Proc. Natl. Acad. Sci. US* **52**, 84 (1964).
[12] P. Desnuelle, in *The Enzymes*, Vol. 4 Ed. P. Boyer, H. Lardy, and K. Myrbäck, Academic Press, New York 1960.
[13] E. Davie and H. Neurath. *J. Biol. Chem.* **212**, 515 (1955).
[14] M. Charles, M. Rovery, A. Guidoni, and P. Desnuelle. *Biochim. Biophys. Acta* **69**, 115 (1963).
[15] B. Keil and F. Šorm. *Feder. Europ. Biochem. Soc., Symposium No. 1*, p. 37, Goodwin, Harris, Hartley Ed., London 1964.
[16] B. W. Matthews, P. B. Sigler, R. Henderson, and D. M. Blow. *Nature* **214**, 652 (1967).
[17] B. Keil, V. Dlouhá, V. Holeyšovský, and F. Šorm. *Coll. Czech. Chem. Commun.* **33**, 2321 (1968).
[18] H. Neurath, K. A. Walsh, and W. P. Winter. *Science* **158**, 1638 (1967).
[19] P. B. Sigler, D. M. Blow, B. W. Matthews, and R. Henderson, in press.
[20] D. C. Phillips. *Scientific American* **215**, 78 (1966).
[21] S. V. Shlyapnikov, B. Meloun, B. Keil, and F. Šorm. *Coll. Czech. Chem. Commun.* **33**, 2306 (1968).
[22] V. Holeyšovský, J. Augustyniak, and B. Keil. FEBS Ann. Meeting Abstracts (Prague 1968).
[23] B. S. Hartley. *Feder. Europ. Biochem. Soc., Symposium No. 1*, p. 47, Goodwin Harris, Hartley Ed. London 1964.
[24] M. L. Bender and F. J. Kézdy. *Ann. Rev. Biochem.* **34**, 49 (1965).
[25] B. F. Erlanger. *Proc. Natl. Acad. Sci. US* **58**, 703 (1967).
[26] H. Keilová, V. Pliška, B. Keil, and F. Šorm. to be published.
[27] M. Kunitz. *J. Gen. Physiol.* **33**, 349 (1950).
[28] V. Dlouhá, D. Pospíšilová, B. Meloun, and F. Šorm. *Coll. Czech. Chem. Commun.* **30**, 1311 (1965).
[29] B. Kassel, M. Radicevic, M. J. Ansfield, and M. Laskowski, Sr. *Biochem. Biophys. Res. Comm.* **30**, 1311 (1965).
[30] J. Chauvet, G. Nouvel, and R. Acher. *Biochim. Biophys. Acta* **92**, 200 (1964).
[31] B. Meloun, I. Frič, and F. Šorm. *European J. Biochem.* **4**, 112 (1968).
[32] D. G. Hoare and D. E. Koshland, Jr. *J. Biol. Chem.* **242**, 2447 (1967).
[33] B. Meloun, in press.
[34] L. F. Kress and M. Laskowski, Sr. *J. Biol. Chem.* **242**, 4925 (1967).
[35] B. Meloun, I. Frič, and F. Šorm. *Coll. Czech. Chem. Commun.* **33**, 2313 (1968).
[36] J. Chauvet and R. Acher. *J. Biol. Chem.* **242**, 4274 (1967).
[37] D. Pospíšilová, B. Meloun, I. Frič, and F. Šorm. *Coll. Czech. Chem. Commun.* **32**, 4108 (1967).
[38] V. Dlouhá, B. Keil, and F. Šorm. *Biochem. Biophys. Res. Commun.* **31**, 66 (1968).
[39] W. R. Firkenstadt and M. Laskowski, Jr. *J. Biol. Chem.* **240**, PC962 (1965).