Some Comments on the Crystallography and the Evolution of the Catalytic Structure in Some Serine Hydrolases[*],[#]

Eleanor Dodson[a] *and Guy Dodson*[a,b,**]

[a]*Department of Chemistry, University of York, York YO10 5DD, UK*

[b]*National Institute for Medical Research, London NW7 1AA, UK*

X-ray analysis has provided a 3-dimensional structural basis through which the complex processes of biochemistry and biology can be investigated and understood. The crystallographic studies on the serine hydrolases have provided accurate structures for many of the enzyme families and have shown that although the families evolved independently, their active atoms in their catalytic structures have equivalent stereochemistry. There has also been remarkable variation in the amino acid composition of the catalytic structures. Analysis has shown that their essential catalytic chemistry is achieved through the interactions of carboxyl or carbonyl oxygens, a base and a potential nucleophile. Finally the evolutionary changes in some enzyme families, such as the Ntn-hydrolases, are so extensive that their evolutionary relationships can only be detected by comparisons of their 3-dimensional structures. Thus X-ray analysis gives the biologist chemical, functional and evolutionary insight into protein molecules.

Key words: serine proteases, serine hydrolases, X-ray crystallography, catalytic triad, molecular evolution, nucleophilic attack, amidase, esterase, Ntn-hydrolases.

INTRODUCTION

It is a pleasure to contribute a paper to this volume of Croatica Chemica Acta dedicated to Boris Kamenar, who as a crystallographer and a chemist has always been interested in the ability of X-ray analysis to determine structures, including protein structures, and the chemical mechanisms that can be deduced from them.

The remarkable power of single crystal X-ray analysis is its ability to reveal individual atomic positions. It was this that made X-ray crystallography so fundamentally important in chemistry. It provided first an accurate structural framework to fit the theories of the chemical bond, secondly it established the huge 3-dimensional library of chemical molecules, and thirdly it removed the labour of synthesis to prove chemical constitution.[1] The impact of X-ray analysis on biological science however is probably even more important. The technique has established the principles of macro-molecular structure and has defined the active centres and surfaces in a huge range of proteins and their complexes. Such is the power and vitality of the crystallographic method that in some 30 years, really an incredibly short time, the 3-dimensional structure of some 8000 biological molecules, many of key importance, have now been determined. At present the rate is about 15 structures a week and this is increasing.[2]

X-ray analysis has defined not only the 3 dimensional atomic structure of proteins but also the stereochemistry of the amino acid groupings, and where present, the chemical cofactors responsible for their function. Often it has been possible to deduce the chemical mechanisms of the protein molecule's reactions – bringing a new and fundamental insight to the biochemistry. But biological molecules have another importance, they are subject to evolutionary selection and change. The determination of their 3-dimensional structures has revealed the existence of both divergent and convergent evolutionary relationships. The former are often very ancient indeed and undetectable from the amino acid sequences alone. The analysis of these relationships gives X-ray analysis an additional impact on biological science.

In this review some enzymes from one large class, the serine hydrolases, has been chosen to illustrate variation in 3-dimensional structure and to identify examples of divergent and convergent evolution. At the same time comparisons of the enzymes' active sites will illustrate how the requirements of their reaction mechanisms have imposed common chemical and structural features on the catalytic machinery on these very different protein molecules.

EVOLUTIONARY RELATIONSHIPS IN THE SERINE HYDROLASES

The serine hydrolases are a very large and diverse class of enzymes which act on a wide variety of substrates. They employ a common mechanism of nucleophilic attack through an activated seryl hydroxyl oxygen.

In the late 1960s the crystal structures of two serine hydrolases, alpha-chymotrypsin and subtilisin were determined.[3,4,5] The enzymes belong to a large family within the serine hydrolases, the serine proteases. Both enzymes carry out the same chemistry of nucleophilic attack by the seryl oxygen on the carbonyl carbon in a peptide bond. The two enzymes have completely independent evolutionary histories and completely different amino acid sequences and 3 dimensional structures. The X-ray analyses reveal however that their catalytic structures both consisted of the same three amino acids: Asp-His-Ser.[4] Comparison of these two local structures showed moreover that their active H-bonded atoms were in essentially identical positions. This was a most important demonstration of locally convergent evolution in which the stereochemical requirements for catalysis were imposed on the chemical and structural variations driven by the evolutionary process. Figure 1 illustrates the overlap of the catalytic triad in trypsin (from *Streptomyces griseus*) with that in subtilisin (from *Bacillus subtilis*) and shows the similar stereochemistry of their active atoms.

Many other crystal structures of chymotrypsin related and subtilisin related enzymes have since been determined. It is apparent that within these two families there are very considerable changes in the sequences and some distinct changes in chain structure, particularly in segments connecting
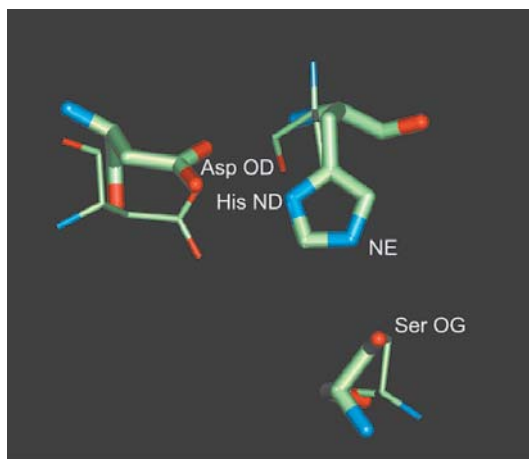

Figure 1. The overlay of the catalytic triad from the serine protease trypsin [pdb listing 1sgt] from the organism *Streptomyces griseus* on that from subtilisin, derived from *Bacillus subtilis* [pdb listing 1svn]. The atoms are coloured according to the usual conventions. The trypsin catalytic residues are drawn with thick bonds, those from subtilisin with thin bonds. The aspartic acid is to the left, the histidine is central while the serine lies below. Note the close overlap of the Asp OD, His ND and NE, and Ser OG atoms.

strands and helices. The stereochemistry at the catalytic structures in the two families however does not change. Moreover the positioning of the catalytic amino acids on the protein polypeptide chain of the chymotrypsin-related and of the subtilisin-related serine protease families as a rule does not vary.[4,6,7]

The same catalytic structure of Asp-His-Ser has since been identified in a number of other enzyme families. One such family is that of the alpha/beta hydrolases.[8] The enzymes in this family have a central sheet of sequential parallel strands, linked by loops and helices which vary enormously between the different species. The active serine however always sits on a specific strand which is connected by a tight loop to a helix. Figure 2A shows the architecture, represented schematically, of the lipase from the yeast *Mucor miehei* typical of the members of this family. These particular enzymes hydrolyse the ester bond in acyl-glycerides by the same mechanism of nucleophilic attack as seen in the serine proteases. Inspection of the catalytic machinery revealed the presence again of the Asp-His-Ser triad, and further it was shown that the active atoms overlapped closely onto their equivalents in the serine proteases,[9] as illustrated in Figure 3. Comparison of enzymes in the alpha/beta hydrolase family shows that variation in the position of the catalytic residues on the polypeptide backbone can occurr and that the aspartic acid can change to glutamic acid, or even be substituted by a carbonyl oxygen.[10] This relaxation of the stereochemistry at the catalytic triad in these lipases is in contrast to the tightly defined stereochemistry in the serine proteases. It probably reflects the more labile character of the ester bond compared to that of the amide bond.

Two recent determinations of esterase structures provide striking examples of molecular evolution. The first is a methyl esterase from *Salmonella typhimirium* from where the arrangement of the beta-strands and some of the helices are very reminiscent of the alpha/beta hydrolases.[8,11] A comparison between this esterase and the lipase from *Mucor miehei* is shown in Figure 2A and B. Since the connectivity linking the strands is importantly different, the enzymes have in all probability evolved independently. Most intriguingly however the catalytic serine in both enzymes is sitting on a loop between a central strand and a helix. With this arrangement both enzymes can exploit favourable access to a substrate and its presence as a common feature in the two enzymes could therefore just as well be an example of convergent evolution. It is also intriguing that the two subsequent strands and the subsequent helix have the same organisation. This might suggest a very remote common evolutionary origin for the two enzymes. But the small size and simple organisation of the motif makes convergent evolution a feasible mechanism for this structure in the two molecules as well.
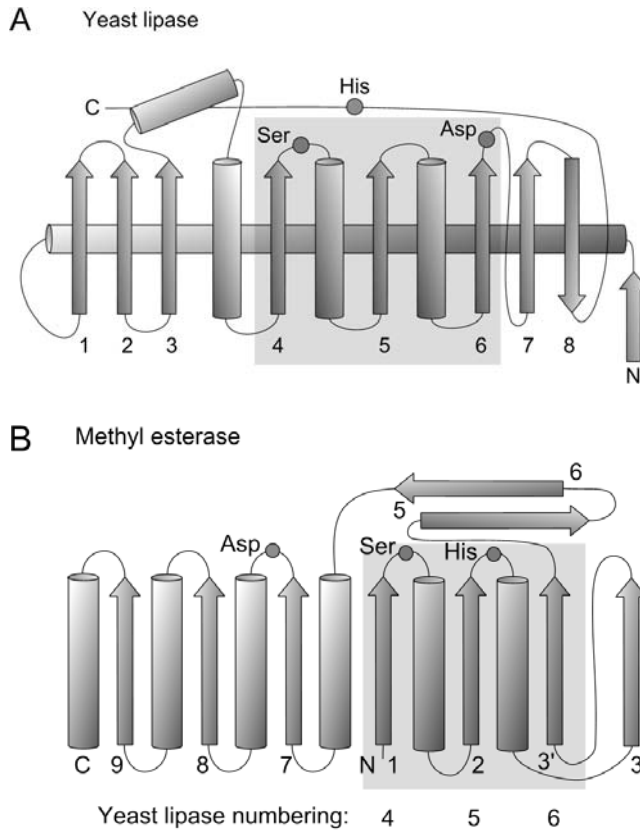
Figure 2. The organisation of main chain secondary structure with helices represented as cylinders and β-strands represented as arrows in two serine hydrolases. This simplified schematic emphasises the connectivity along the elements of secondary structure, but loses some aspects of the 3-dimensional organisation. The catalytic residues – Asp, His, Ser – are labelled and represented as circles.
A, the lipase from *Mucor miehei*.
B, a methyl esterase from *Salmonella typhimurium*.
Both enzymes contain 7 central parallel β-strands, though their connectivity is different. It can be seen that in the lipase, the β-strands are arranged in the same order as the sequence, in the methyl esterase this is not the case. The individual β-strands nonetheless share a very similar structural architecture imposed by the H-bonding interactions between the polypeptide chains. There is also a difference in the ordering of the sequence of the catalytic residues. In the lipase the order from the sequence N-terminus is Ser → Asp → His; in the methyl esterase, Ser → His → Asp. The three strands in the shaded box however match in connectivity and structure. It may be that this small motif was incorporated in both enzymes, or that the structure was arrived at independently. On the basis of the connectivity and the sequence order of the catalytic residues, it appears that the similarity in secondary structure positioning is the result of a convergent process.

A further factor in favour of convergent evolution is that in the methyl esterase the histidine and aspartic acid are recruited from quite different regions of the sequence and structure compared to the alpha/beta hydrolases.
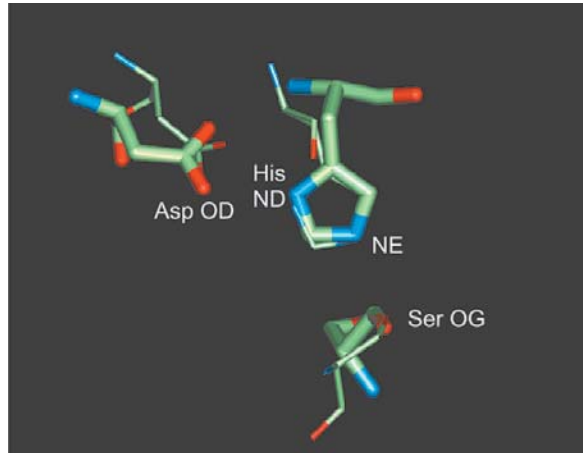

Figure 3. The overlap of the catalytic residues in the trypsin from *Streptomyces griseus* (thick bonds) and the lipase from *Mucor miehei* (thin bonds). The aspartic acid is to the left, the histidine is central while the serine lies below; the angle of view is slightly different to that in Figure 1. Note again the overlap of the active atoms.

The second example of evolutionary change at the molecular level comes from an acetyl hydrolase that functions in neural cells.[12] In this enzyme the polypeptide chain has a 3-dimensional fold essentially the same as the proteins, called G-proteins, that hydrolyse GTP. The small G-proteins are key molecules in the cell signalling pathways. Their particular structure is widespread and is generally associated with guanosine di- or tri-phosphate binding. The secondary structure organisation of the GTPase molecules and the esterase are illustrated schematically in Figure 4A and B. The amino acid sequences of the two molecules show no detectable similarity. This suggests that the presence of the GTPase fold in the main chain imposes few restrictions on the amino acids during its evolution to different functions. In Figure 5 the overlap of the catalytic triad from the serine protease trypsin and from the brain acetylase is shown. Again the convergent evolution acting locally on the catalytic residues can be seen in the close overlap of the active atoms, though this is not as close as the overlaps seen in the serine proteases.
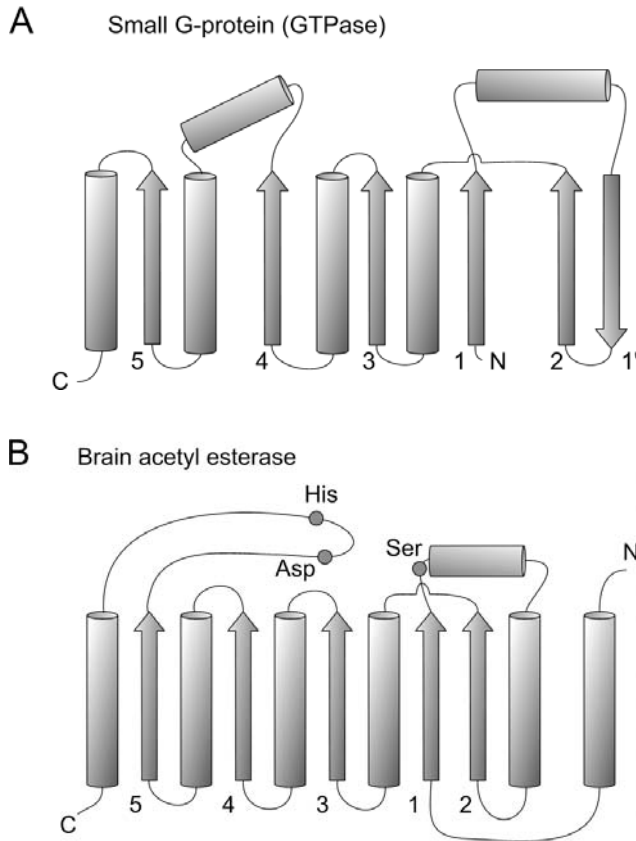
Figure 4. Schematic representations of a small G-protein, rac–1 (pdb listing 1mih), A; and of a brain acetyl esterase (pdb listing 1wab), B. The connectivity of the β-strands structure in these two proteins is the same except that the G-protein has an extra strand (1'). There are also helices which can be seen to connect the strands in a generally identical fashion.
The active residues in the two proteins are different and are performing very different functions: in the G-protein they stabilise the bound guanosine di- or tri-phosphate and in the esterase they form a catalytic site. They come from different points of the sequence but those involved in the GTPase activity are not shown here.

VARIATIONS IN THE CATALYTIC STRUCTURE IN SERINE HYDROLASES

Variations at the Acid/H-bond Acceptor

The comparison of the catalytic triads Asp-His-Ser in a selection of these enzymes has revealed a number of variations that inform us about the
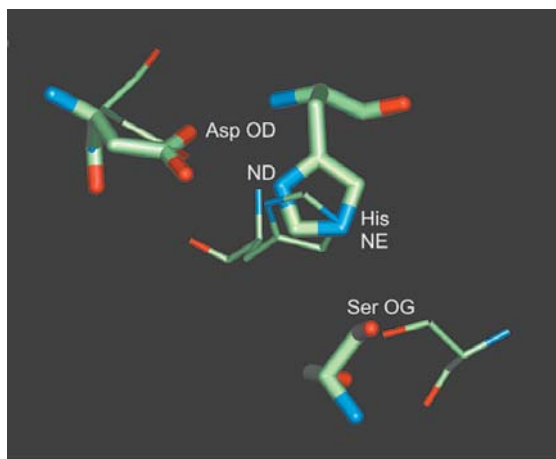
Figure 5. The overlap of the catalytic residues in the trypsin from *Streptomyces griseus* (thick bonds) and in the brain acetyl hydrolase. The view is similar to that in Figure 1. The aspartic acid is to the left, the histidine is central while the serine lies below. In this comparison the match of the active atoms is poorer than those in Figures 1 and 3.

chemistry of the catalytic reaction. The three residues have the character of acid, acting as a H-bond acceptor, base and nucleophile. Survey of the structural data base shows that this combination of chemical properties is preserved in the serine hydrolases but that there are variations.[13] The appearance of oxygen functions other than the aspartic acid carboxylate group in certain esterases has already been noted, see above. Another variation, aspartic acid to histidine, was recently reported in the structure of a new family of viral proteases.[14,15,16,17] In the crystal structure of this dimeric enzyme we can see that the acid function served by aspartic acid has been replaced by histidine and that the carboxylate oxygen function is substituted by the imidazole epsilon-nitrogen. Thus the catalytic triad is not Asp-His-Ser but His-His-Ser. For a chemist there is no surprise in seeing an imidazole functioning as a H-bond acceptor once it is realised that the special environment in the protein is reducing the pK of this histidine side chain.

Variations at the Base

The histidine which acts as a base in the classic triad is replaced by lysine in the enzyme asparaginase and in the serine beta-lactamase.[18,19] This substitution has two consequences. The first of these is steric, the contacts to the epsilon-amino group will be tetrahedral while those to histidine will be planar. The second is that lysine has a higher pK than histidine, 10.5 *cf*. 6.8 which will affect the local chemistry. To act as a base the lysine must be

deprotonated. It is evident that the local displacement of water molecules and the burial of the active site by the substrate, accompanied by structural changes, favour the proton's movement and the structural and chemical events in the reaction pathway. The presence of an amino group acting as base in the catalytic residues also occurs in the newly identified Ntn-hydrolase family of enzymes. The nucleophilic residue in these enzymes is at the N-terminus; it is the N-terminal amino group that operates as the base. The conformation assumed by this amino acid creates a H-bond inter-action between the alpha-amino group and the side chain. These interac-tions that create catalytic potency in the Ntn-hydrolases are more equiva-lent to those seen in asparaginase (seen in Figure 6) or the beta-lactamase, though they exist within a single amino acid, and may also recruit a water molecule into the reaction.[20] The catalytic site in the Ntn-hydrolase, Penicil-lin G amidase or acetylase, is illustrated in Figure 6.
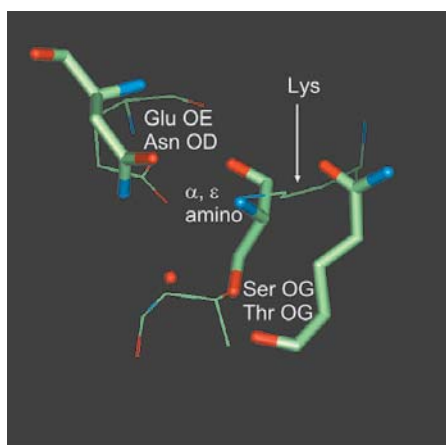

Figure 6. The overlay of the catalytic groups of asparaginase and that from penicillin G acylase both from *Escherichia coli*. Note that in the penicillin acylase, illustrated with thick bonds, the nucleophile containing side chain and the base belong to the N-terminal residue. In asparaginase which is drawn with thin bonds, the aspartic acid, the lysine, acting as the base, is central and the threonine (equivalent to serine) lies below. These match on to a carbonyl oxygen from a glutamine side chain, the alpha-amino function and the serine residue respectively from the penicillin G amidase.

Variations at the Nucleophile-containing Residue

Although the families of enzymes under discussion are referred to gen-erally as the serine hydrolases, the nucleophile containing residue in some cases has been found to change to cysteine or threonine. There is for exam-

ple one chymotrypsin-like serine protease from the hepatitis A virus, in which the enzyme's nucleophilic residue has changed to a cysteine.[21,22] More remarkably, the Ntn-hydrolase family has been found to contain members with serine, cysteine and threonine as the catalytic nucleophile.[23] For example the Ntn-hydrolase penicillin V amidase with a catalytic N-terminal cysteine removes the sidegroup from the beta-lactam by cleaving the linking amide bond.[24] The enzyme performs an identical amide hydrolysis to that of penicillin G acylase, which curiously is also a Ntn-hydrolase[23] – it however has a cysteine as its catalytic residue. Replacement of the serine by cysteine in the viral protease and in the amidase will introduce a stronger nucleophile and there may be need, as yet unidentified, for this in these enzymes.

The proteasome assembly also contains proteases which belong to the Ntn-hydrolase family.[25] This enzyme complex degrades peptides and proteins in the cell. Its catalytic residue is a threonine which in terms of the nucleophilic reaction is the same as serine. The presence in threonine of the gamma methyl will however introduce steric constraints which in serine proteases would interfere with the conformation and the movement of the side chain in the catalytic process. One other enzyme that has threonine as its catalytic residue is asparaginase, referred to above. It acts as an amidase to convert the side chain amide to a carboxylic acid by the same mechanism.[26] An overlay of the catalytic structures in the Ntn-hydrolase and in asparaginase is shown in Figure 6. Note that the match of active atoms is as close as in other comparisons.

CONCLUSION

This survey is unavoidably selective but it is evident from the 3 dimensional structures that there is an evolution of the catalytic residues in serine hydrolases which can inform the chemist about the constraints acting on the catalysis and which also can give the biologist clues about the physiological conditions the organism concerned lives in. There is no doubt that the progress in stuctural analysis is going to lead to many more surveys of functional sites across the rapidly growing number of known structures.[27] These are likely to reveal the presence of functions unsuspected by the biochemist. Very recently for example the crystal structure of an alpha-amylase extracted from an Antarctic organism was determined.[28] The authors discovered that on the enzyme's surface there was a triad Glu-His-Ser whose active atoms matched those of the classic triad in a serine protease, or more exactly the identical triad in the lipase from the yeast *Geotrichum candidum*.[8] This finding has given the biologists questions – does the enzyme have a hydrolytic function, and if so what is it there for. It seems to many of us that there will be many other discoveries of this nature in the

future. Thus crystallography continues to bridge disciplines, to challenge the ideas of chemists and biologists and more importantly is bringing both research communities together.

REFERENCES

1. J. Dunitz, in: G. Dodson, J. P. Glusker and D. Sayre, (Eds.), *Structural studies on molecules of biological interest,* Oxford University Press, 1981, pp. 47–59.
2. W. Hendrickson and K. Wuthrich, (Eds.), *Macromolecular Structures*, Current Biology Ltd., London, 1991–1997.
3. D. M. Blow, *Trends Biochem. Sci*. **22** (1996) 405–408.
4. J. Kraut, *Annu. Rev. Biochem*. **46** (1977) 331–358.
5. J. Drenth, W. Hol, J. N. Jansonius, and R. Koekoek, *Cold Spring Harbor Symp. Quant. Biol*. **36** (1971) 107–116.
6. C. A. Orengo, A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells, and J. M. Thornton, *Structure* **5** (1994) 1093–1108
7. A. C. Wallace, R. A. Laskowski, and J. Thornton, *Protein Sci.* **5** (1996) 1001–1013.
8. D. L. Ollis, E. Cheah, M. Cygler, B. Dijkstra, F. Frolow, S. M. Franken, M. Harel, S. J. Remington, I. Silman, J. Schrag, J. L. Sussman, J. H. G. Verschueren, and A. Goldman, *Protein Eng*. **5** (1992) 197–211.
9. L. Brady, A. M. Brzozowski, Z. S. Derewenda, E. Dodson, G. Dodson, S. Tolley, J. P. Turkenburg, L. Christiansen, B. Huge Jensen, L. Norskov, L. Thim, and U. Menge, *Nature* **343** (1990) 767–770.
10. D. Lawson, A. M. Brzozowski, and G. G. Dodson, *Curr. Biol*. **2** (1992) 473–475.
11. A. H. West, E. Martinez-Hackert, and A. M. Stock, *J. Mol. Biol*. **250** (1995) 276–290.
12. Y. S. Ho, L. Swenson, U. Derewenda, L. Serre, Y. Y. Wei, Z. Dauter, M. Hattori, T. Adachi, J. Aoki, H. Arai, K. Inoue, and Z. S. Derewenda, *Nature* **385** (1997) 89–93.
13. M. Paetzel and R. E. Dalbey, *Trends Biochem. Sci*. **22** (1997) 29–31.
14. P. Chen, H. Tsuge, R. J. Almassy, C. L. Gribskov, S. Katoh, D. L. Vanderpool, S. A. Margosiak, C. Pinko, D. A. Matthews, and C. C. Kan, *Cell* **86** (1996) 835–843.
15. L. Tong, C. G. Qian, M. J. Massariol, P. R. Bonneau, M. G. Cordingley, and L. Lagace, *Nature* **383** (1996) 272–275.
16. X. Y.. Qiu, J. S. Culp, A. G. DiLella, B. Hellmig, S. S. Hoog, C. A. Janson, W. W. Smith, and S. S. AbdelMeguid, *Nature* **383** (1996) 275–279.
17. H. S.. Shieh, R. G. Kurumbail, Stevens, A. M. Stegeman, E. J. Sturman, J. Y. Pak, A. J. Wittwer, M. O. Palmier, R. C. Wiegand, B. C. Holwerda, and W. C. Stallings, *Nature* **383** (1996) 279–282.
18. G. J. Palm, J. Lubkowski, C. Derst, S. Schleper, K. H. Rohm, and A. Wlodawer, *FEBS Letts*. **390** (1996) 211–216.

19. N. C. J. Strynadka, R. Martin, S. E. Jensen, M. Gold, and J. B. Jones, *Nat. Struct. Biol.* **3** (1996) 688–695.

20. H. J. Duggleby, S. P. Tolley, C. P. Hill, E. J. Dodson, G. G. Dodson, and P. C. E. Moody, *Nature* (1995) **373** 264–268.

21. D. A. Matthews, W. W. Smith, R. A. Ferre, B. Condon, G. Budahazi, W. Sisson, J. E., Villafranca, C. A. Janson, H. E. Mcelroy, C. L. Gribskov, and S. Worland *Cell* **77** (1994) 761–771.

22. E. M. Bergmann, S. C. Mosimann, M. M. Chernaia, B. A. Malcolm, and M. N. G. James, *J. Virol.* **71** (1997) 2436–2448.

23. J. A. Brannigan, G. Dodson, H. J. Duggleby, P. C. E. Moody, J. L. Smith, D. R. Tomchick, and A. G. Murzin, *Nature* **378** (1995) 416–419.

24. C. G. Suresh, A. V. Pundle, H. SivaRaman, K. N. Rao, J. A. Brannigan, C. E. McVey, C. S. Verma, Z. Dauter, E. J. Dodson, and G. G. Dodson. *Nat. Struct. Biol.* (1998), accepted for publication.

25. J. Lowe, D. Stock, R. Jap, P. Zwickl, W. Baumeister, and R. Huber, *Science* **268** (1995) 533–539.

26. A. L. Swain, M. Jaskolski, D. Housset, J. K. M. Rao, and A. Wlodawer, *Proc. Natl. Acad. Sci. (U.S.A.)* **90** (1993) 1474–1478.

27. G. Dodson and A. Wlodawer, *TIBS* **23** (1998) 347–352.

28. N. Aghajari, G. Feller, C. Gerday, and R. Haser, *Protein Sci.* **7** (1998) 564–576.

SAŽETAK

Kristalografska tumačenja i procjena katalitičkog ustroja nekih serin-hidrolaza

Eleanor Dodson i Guy Dodson

Rentgenska strukturna analiza pruža trodimenzijsku strukturnu osnovicu putem koje se složeni biokemijski i biološki procesi mogu istraživati i razumjeti. Kristalo-grafsko izučavanje serin-hidrolaza pribavilo je točnu građu mnogih skupina enzima i pokazalo da, premda se skupine razvijaju neovisno, aktivni atomi u njihovim kata-litičkim strukturama imaju ekvivalentnu stereokemiju. Uočila se također i značajna promjena u aminokiselinskom sastavu katalitičkih struktura. Analizom je pokazano da se njihova bitna katalitička kemija odvija kroz interakcije karboksilnih ili karbo-nilnih kisikovih atoma, osnovnog i možebitnog nukleofila. Konačno, razvojne promje-ne u nekim skupinama enzima kao što su Ntn-hidrolaze tako su opsežne da se nji-hove razvojne srodnosti mogu otkriti tek usporedbom njihove trodimenzijske građe. Rentgenska strukturna analiza nudi biološkom kemičaru uvid ne samo u funkcijske već i razvojne značajke molekula proteina.