Protein Science (1996), 5:1001-1013. Cambridge University Press. Printed in the USA.
Copyright © 1996 The Protein Society

Derivation of 3D coordinate templates for searching structural databases: Application to Ser-His-Asp catalytic triads in the serine proteinases and lipases

ANDREW C. WALLACE, ROMAN A. LASKOWSKI, AND JANET M. THORNTON
Biomolecular Structure and Modelling Unit, Department of Biochemistry and Molecular Biology, University College, Gower Street, London WC1E 6BT, England

(RECEIVED December 14, 1995; ACCEPTED March 13, 1996)

Abstract

It is well established that sequence templates (e.g., PROSITE) and databases are powerful tools for identifying biological function and tertiary structure for an unknown protein sequence. Here we describe a method for automatically deriving 3D templates from the protein structures deposited in the Brookhaven Protein Data Bank. As an example, we describe a template derived for the Ser-His-Asp catalytic triad found in the serine proteases and triacylglycerol lipases. We find that the resultant template provides a highly selective tool for automatically differentiating between catalytic and noncatalytic Ser-His-Asp associations. When applied to nonproteolytic proteins, the template picks out two "non-esterase" catalytic triads that may be of biological relevance. This suggests that the development of databases of 3D templates, such as those that currently exist for protein sequence templates, will help identify the functions of new protein structures as they are determined and pinpoint their functionally important regions.

Keywords: catalytic triad; cyclophilin; 3D structure; lipases; serine proteinases

The use of protein sequence motifs and templates as a tool for the identification of biological function and prediction of tertiary structure is already well established (see reviews by Taylor, 1988; Hodgman, 1989; Taylor & Jones, 1991). These templates are in essence a 1D protein sequence signature that is identified by the analysis of information in known protein structures and in data from sequence alignments and pattern-matching techniques. The information is summarized in databases such as PROSITE (Bairoch & Bucher, 1994) and PRINTS (Attwood et al., 1994) which, along with automatic sequence alignment algorithms, enables swift assessment of an unknown protein sequence.

There has been detailed analysis of the 3D topologies of metal-binding sites, both in proteins and in small molecules (for reviews see Glusker, 1991; Jernigan et al., 1994). However, there is not a database of 3D templates of functionally important units in proteins, analogous to the sequence templates of PROSITE. As the number of known protein structures increases, so the need for a 3D equivalent of PROSITE grows with it—especially for identifying likely functions of proteins whose biological role is unknown and, equally usefully, for locating the functional regions and residues involved. A 3D template can provide a quantitative description of the relative dispositions of, for example, the key residues in an enzyme active site based solely on the coordinates. The template can then be used to scan a database of known protein structures to identify putative catalytic centers.

Here we demonstrate the power of such search templates and the novel information they can provide. We take as our example the Ser-His-Asp catalytic triad of the serine proteases and lipases, which is one of the best known and most intensively studied of all functional mechanisms. An added advantage of using this triad is that there are many examples in the protein structures deposited in the Brookhaven Protein Data Bank (PDB) (Bernstein et al., 1977). The triads are present in several protein families, and the template provides a means of quantifying the differences in the conformational geometry across the different structural and functional families.

In the Ser-His-Asp catalytic triad, the three residues, which occur far apart in the amino acid sequence of the enzyme, come together in a specific conformation in the active site to perform the hydrolytic cleavage of the appropriate bond in the substrate. This triad was first identified in the serine proteinases (Blow et al., 1969; Wright et al., 1969), which cleave peptides at the amide bond. This is an ubiquitous group of proteolytic enzymes responsible for a range of physiological responses, such as the onset of blood clotting (Mann, 1987) and digestion (Blow,

Reprint requests to: Janet M. Thornton, Biomolecular Structure and Modelling Unit, Department of Biochemistry and Molecular Biology, University College, Gower Street, London WC1E 6BT, England; e-mail thornton@bsm.bioc.ucl.ac.uk.

1976). They also play a major role in the tissue destruction associated with arthritis, pancreatitis, and pulmonary emphysema. Each enzyme is highly specific for its own peptide substrate and this specificity is governed by the substrate residue that fits into the P′ subsite, or specificity pocket, immediately adjacent to the scissile bond. Perona and Craik (1995) have recently published a comprehensive review of the structural basis of substrate specificity in serine proteinases.

Barth et al. (1993) have performed an extensive steric comparison of the active site residues in the serine proteases. They analyzed the differences in the relative conformations of the Ser-His-Asp residues by performing RMS fits of all against all occurrences and, on the basis of the differences and similarities, were able to classify the serine proteases according the chymotrypsin and subtilisin families. As a result of their analyses, they were able to identify an additional serine, which is present, and highly conserved, in the active site of the serine proteases, and suggested it to be part of a "catalytic tetrad." They also found several examples of Ser-His-Asp triads in nonproteolytic proteins and in approximately similar conformations to that of the serine proteases (Barth et al., 1994).

Artymiuk et al. (1994) have used a graph-theoretic approach for the identification of 3D patterns of amino acid side chains in protein structures. For example, they constructed a search template from the side-chain atoms of the Ser 195-His 57-Asp 102 catalytic triad of chymotrypsin and, depending on the allowed interatomic distance tolerances, different numbers of catalytic triads were identified from their data set. They also identified an unusual triad from the pro-enzymes chymotrypsinogen and trypsinogen, which does not exist in the active form.

A different structural comparison of the serine proteases, using a less specific technique, has been performed by Fischer et al. (1994). Their method, derived from geometric hashing methods used in computer vision research, treats all Cα atoms in a protein as points in space and compares proteins purely on the geometric relationships between these points. It can detect recurring substructural 3D motifs, and was able to identify the structural similarities of the active sites of the trypsin-like and subtilisin-like serine proteases based solely on the similarities of the Cα geometries of their constituent residues.

Apart from the serine proteases, the Ser-His-Asp catalytic triad also occurs in the triacylglycerol lipases, which are responsible for hydrolyzing triglycerides into diglycerides and, subsequently, monoglycerides and free fatty acids. For example, pancreatic lipase hydrolyzes water-insoluble triacylglycerols in the intestinal lumen and thereby plays an important role in dietary fat absorption. Lipases are stable in both aqueous and organic media and this makes them suitable as catalysts for a number of synthetic processes that would otherwise require harsh conditions to proceed. Like the serine proteinases, the catalytic mechanism is effected by way of a catalytic serine (Blow, 1990; Brady et al., 1990). The catalytic site is buried beneath a short stretch of helix, known as the "lid." A number of crystallographic studies have confirmed the hypothesis that the lid is displaced during activation (Brzozowski et al., 1991; Derewenda et al., 1992), being rolled back as a rigid body into a hydrophilic trench filled previously by water molecules, exposing the active site.

In this study, we used a data set of serine proteases and lipases to automatically compute a highly specific 3D template for the Ser-His-Asp catalytic triad in these proteins, thereby defining its conformation. Our method differs substantially from previous methods (e.g., Barth et al., 1993; Artymiuk et al., 1994; Fischer et al., 1994) in that a simple template, specific to the Ser-His-Asp catalytic triad, is derived. This template allows such triads to be identified uniquely in other structures and quantifies the differences in the triads from related proteins. The application of this procedure to other systems is discussed.

We used the derived template to search for similar triads in other proteins, including non-enzymes, to see how often they occur outside the serine proteinases and lipases. We found two proteins with the correct triad conformation and these are discussed. Our study demonstrates that 3D templates derived from the PDB can provide interesting suggestions concerning the function of a novel protein once its structure has been determined.

Results

The data sets

Two data sets were used, both extracted from the January 1995 release of the PDB. The first comprised the serine proteinases and lipases used in deriving the Ser-His-Asp 3D templates. The serine proteinases, lipases, and related enzymes were identified on the basis of their Enzyme Classification (E.C.) number (Bielka et al., 1992), which characterizes an enzyme's function in terms of the reaction it catalyzes, the substrate on which it operates, and any associated co-factors. To ensure the most complete data set, the sequence of each structure in the PDB was cross-referenced against the SWISS-PROT database (Bairoch & Boeckmann, 1994) and the relevant E.C. numbers identified.

The resultant data set consisted of 192 serine proteinases, 4 serine-type carboxypeptidases, and 9 triacylglycerol lipases. Some of the enzymes have more than one chain, and therefore more than one catalytic triad. Table 1 lists the data set in terms of all the individual chains: 205 serine proteinases, 7 serine-type carboxypeptidases, and 13 lipases. The chains are grouped into four main fold groups, numbered 1–4, according to their overall structure. This structural classification was achieved using the program SSAP (Orengo et al., 1993), which computes a similarity score (SSAP score) between two proteins; the higher the score, which ranges from 0 to 100, the more similar the overall structures. We used an SSAP score of >80, which is the minimum score generally used to identify homologues, to group together enzymes having similar overall folds.

As can be seen in Table 1, the serine proteinases come in two distinct fold groups: Group 1 members have a β-sandwich fold, characterized by trypsin, whereas Group 2 members have an alternating α-β fold, characterized by subtilisin. Group 1 is further subdivided into subgroups 1a, 1b, and 1c on the basis of sequence similarity: chains in different subgroups have a lower than 30% pairwise sequence identity. These chains, being in the same fold group, still have very similar overall structures and are almost certainly derived from a common ancestor. Group 3 contains the serine-type carboxypeptidases and Group 4 the triacylglycerol lipases. The SSAP scores between members of Groups 2, 3, and 4 are <70, reflecting the significant structural differences between them. However, the overall folds of the serine-type carboxypeptidases of Group 3 and the triacylglycerol lipases of Group 4 have some similarity in that both are members of a superfamily of α/β hydrolases, as defined in SCOP (Murzin et al., 1995) and Ollis et al. (1992). Within each

Table 1. *Data set of enzyme chains containing the Ser-His-Asp catalytic triad*

FOLD GROUP 1: Serine proteases
β-Sandwich-trypsin-like fold

Subgroup 1a
Chymotrypsin E.C.3.4.21.1
1acb E 1cgi E 1cgj E 2cga A 2cga B 1chg 1cho E 2cha 4cha A 4cha B 5cha A 5cha B
6cha A 6cha B 1gcd 1gct A 2gch 2gct A 3gch 3gct A 4gch 5gch 6gch 7gch
8gch 1gha E 1ghb E 1gmc A 1gmd A 1gmd B 1gmh 2gmt

Trypsin E.C.3.4.21.4
1bit 1bra 1brb E 1brc E 1gbt 1mct A 1ntp 1ppc E 1ppe E 1pph E 2ptc E 2ptn
3ptb 3ptn 4ptp 1sgt 1smf E 1tab E 2tbs 1tgb 1tgc 1tgn 1tgs Z 1tgt
2tga 2tgd 2tgp Z 2tgt 1tld 1tng 1tnh 1tni 1tnj 1tnk 1tnl 1tpa E
1tpo 1tpp 3tpi Z 4tpi Z

Thrombin E.C.3.4.21.5
1abi H 1abj H 1bbr H 1bbr K 1bbr N 1dwb H 1dwc H 1dwd H 1dwe H 1etr H 1ets H 1ett H
1fph H 1hag E 1hah H 1hai H 2hat H 1hgt H 2hgt H 1hlt H 1hlt K 2hnt E 2hpp H 2hpq H
1hrt H 4htc H 1hut H 1ihs H 1iht H 1nrn H 1nrn R 1nro H 1nro R 1nrp H 1nrp R 1nrq H
1nrr H 1nrs H 1ppb H 1thr H 1ths H 1tmb H 1tmt H 1tmu H

Tissue kallikrein E.C.3.4.21.35
2kai A 2kai B 2pka B 2pka Y 1ton

Pancreatic elastase E.C.3.4.21.36
1ela A 1elb A 1elc A 1esa 1esb 1est 2est E 3est 4est E 5est E 6est 7est E
8est E 9est 1inc 1jim

Leukocyte elastase E.C.3.4.21.37
1hne E 1ppf E 1ppg E

Subgroup 1b
α-Lytic protease E.C.3.4.21.12
2alp 1lpr A 2lpr A 3lpr A 4lpr A 5lpr A 6lpr A 7lpr A 8lpr A 9lpr A 1p01 A 1p02 A
1p03 A 1p04 A 1p05 A 1p06 A 1p08 A 1p09 A 2p07 1p10 A 1p11 E 1p12 E

Streptogrisin A E.C.3.4.21.80
1sgc 2sga 3sga E 4sga E 5sga E

Streptogrisin B E.C.3.4.21.81
3sgb E 4sgb E

Subgroup 1c
1.3 Lysyl endopeptidase E.C.3.4.21.50
1arb 1arc

FOLD GROUP 2: Serine proteases
Alternating α/β-subtilisin-like fold

Subtilisin E.C.3.4.21.62
1cse E 1mee A 1s01 1s02 1sbc 1sbn E 1sbt 2sbt 1sca 1scb 1scd 1scn E
1sel A 1sel B 2sec E 1sib E 2sic E 3sic E 5sic E 2sni E 1st2 1st3 2st1 1sub
1suc 1sud 1vmo A 1vmo B

Endopeptidase K E.C.3.4.21.64
1pek E 2pkc 2prk 3prk E 1ptk

Thermitase E.C.3.4.21.66
1tec E 2tec E 3tec E 1thm

FOLD GROUP 3: Serine-type carboxypeptidase
α/β

Serine type carboxypeptidase E.C.3.4.16.5
3sc2 A 3sc2 B 3sc2 A 1whs A 1whs B 1whs A 1wht A 1wht B 1wht A 1ysc

FOLD GROUP 4: Triacylglycerol lipase
α/β

Triacylglycerol lipase E.C.3.1.1.3
1crl 1hpl A 1hpl B 1tah B 1tah A 1tah C 1tah D 1tgl 3tgl 4tgl 5tgl 1thg
1trh

of the fold groups in Table 1, the enzyme chains are further grouped according to their E.C. number, reflecting their different functional roles.

The second of our two data sets was a representative set of protein structures in the PDB. The 3D templates derived from the enzyme data set were applied to this second data set to see if any Ser-His-Asp triplets, in the catalytic conformation, are present in any other proteins. The second data set was a set of unique protein chains, including homologues, but excluding identical or trivially different chains such as single-residue mutants. It was compiled by extracting protein chains from the PDB such that no two had a sequence identity greater than 95%. The resultant 639 protein chains are listed in Table 2.

Derivation of the 3D templates

The first step in generating the 3D coordinate templates was to automatically extract all occurrences of interacting Ser, His, and Asp residues, catalytic as well as noncatalytic, and irrespective of relative conformation. The interacting triplets were located in the enzyme data set using a program called DISTRIB (R.A.L.). Residues were considered to be interacting if at least one interatomic contact was less than the sum of the van der Waals radii of the contacting atoms plus 1 Å. Each extracted triplet was transformed onto a common reference frame defined by the planar ring of the His: the His was placed in the x-y plane with its C^{γ} at the origin, its C^{β} on the negative y-axis, and its $N^{\delta 1}$ atom with positive x and y values.

The second step was to automatically filter from these Ser-His-Asp triplets only those that were catalytic triads with well-conserved conformations. These would form the basis for calculating the final 3D templates. The alternative would have been to locate all the catalytic triads manually and use these for deriving the templates. However, there are two principal advantages of the automatic procedure. First, it is more generally applicable to other systems, including those where the triads of interest are not so well-known beforehand. Second, the automatic procedure selects only those cases that are well conserved. The degree of this structural conservation is defined by the RMS distance cut-off; essentially, this separates those interactions that are defined as "catalytic" from those that are "noncatalytic." Therefore, we can quantify the distortion of those triads which, although known to be catalytic, have a conformation in the crystal structure that has somehow been perturbed—such as by the binding of an inhibitor. By eliminating these perturbed conformations, one gets a more accurate and reliable final template.

Thus, the aim of the filtering process was to identify those triplets where the Asp and Ser residues were in approximately the correct positions relative to the His. The procedure we used was an iterative one, illustrated in Figure 1, and involved taking one of the known catalytic triads as a starting point, or seed. The seed triad was Ser 195-His 57-Asp 102 from α-lytic proteinase 1lpr (Bone et al., 1991), although any other unperturbed triad would have done just as well. Initially, we calculated the RMS distance of all the Ser and Asp side-chain atoms relative to the seed triad; however, after detailed inspection, it became clear that the relative positions of the two functional oxygens Asp $O^{\delta 2}$ and Ser O^{γ}, governed the conformation of the catalytic triad. Therefore, these two atoms were taken as reference points and this, in essence, constituted an initial trial 3D template. Each Ser-His-Asp triplet found by DISTRIB was then

Table 2. *Nonhomologous data set of enzyme and non-enzyme proteins, where no two proteins have a sequence identity greater than 95%*

1191 1351 155c 1aaf 1aaj 1aak 1aap A 1aat 1ab2 1aba
1abk 1abm A 1abt A 1aca 1ace 1aco 1acp 1acx 1adn 1adr
1ads 1aec 1aep 1afc A 1ahd P 1ain 1aiz A 1ak3 A 1ake A 1ala
1alb 1alc 1ald 1alk A 1aoz A 1apa 1apm E 1apo 1aps 1arb
1arp 1arq A 1atn A 1atx 1ave A 1avh A 1ayh 1baf H 1bal 1bbh A
1bbi 1bbo 1bbp A 1bbt 1 1bds 1bgc 1bge A 1bgh 1bha 1bia
1bllE 1bmv 1 1bod 1bov A 1brn L 1bsr A 1btc 1bus 1bw3 1c2r A
1c5a 1caa 1cau A 1cb1 1cbn 1cc5 1ccd 1ccr 1cd8 1cdb
1cde 1cdg 1cdt A 1cew I 1cgi I 1cgt 1chb D 1chr A 1cid 1cll
1cmb A 1cob A 1col A 1cor 1coy 1cpb 1cpc A 1cpt 1crl 1csc
1cse E 1ctf 1cth A 1cvo 1cy3 1d66 A 1dbb H 1dfb H 1dhr 1drf
1dri 1dtk 1dtx 1dxi A 1eaf 1eca 1ede 1egf 1ego 1end
1etr H 1ezm 1f3g 1fas 1fba A 1fc2 C 1fcb A 1fdh G 1fdl H 1fdx
1fgv H 1fha 1fia A 1fkb 1flv 1fnr 1frr A 1fus 1fvc A 1fvd A
1fxa A 1fxd 1fxi A 1gal 1gat A 1gb1 1gca 1gct A 1gd1 O 1gdh A
1gf1 1gf2 1ggb H 1ghl A 1gky 1gla G 1glt 1glu A 1gly 1gmf A
1gmp A 1gof 1gox 1gp1 A 1gpb 1gpr 1gps 1gpt 1gsr A 1gss A
1guh A 1hbg 1hbq 1hcc 1hdd C 1hds A 1hdz A 1hem 1hev 1hfh
1hge A 1hhl 1hil A 1hip 1hiv A 1hle A 1hmy 1hna 1hne E 1hoe
1hra 1hrh A 1hsa A 1hsb A 1hsp 1hst A 1huw 1hyp 1i1b 1ifc
1igf H 1ind H 1ipd 1isu A 1ith A 1jhl H 1kdu 1kst 1lab 1lcc A
1lct 1ldn A 1lec 1len A 1lfb 1lfi 1lga A 1lh1 1lis 1lla
1llc 1lld A 1lmb 3 1lpe 1lpf A 1ltb C 1lte 1lts A 1lvl 1lya A
1lz1 1maj 1mam H 1mat 1mba 1mbd 1mbs 1mcp H 1mct A 1mda H
1mdc 1mee A 1mfa H 1mfa L 1min A 1mio A 1mpp 1mup 1myg A 1myp A
1myt 1nar 1nbt A 1nbv H 1nca H 1ndk 1nea 1nip A 1noa 1nor
1npc 1npx 1nrc A 1nrd 1nsc A 1ntx 1nxb 1ofv 1oma 1omf
1onc 1opa A 1osa 1ova A 1ovb 1paf A 1pal 1paz 1pba 1pbx A
1pca 1pda 1pdc 1pdg A 1pfk A 1pgd 1pgx 1pha 1phh 1pho
1pi2 1pii 1pk4 1pkp 1pkr 1plc 1pnj 1poa 1poc 1pod
1poh 1pox A 1pp2 L 1ppa 1ppb H 1ppl E 1ppn 1ppo 1prc C 1ptf
1pya A 1pyp 1r09 1 1r69 1rai A 1rbp 1rcb 1rdg 1rds 1rec
1rei A 1rfb A 1rhd 1rhg A 1rib A 1ril 1rip 1rne 1rop A 1rro
1rtc 1rtp 1 1rve A 1s01 1sbp 1sdy A 1sgt 1sh1 1sha A 1shf A
1shg 1shp 1sim 1siv A 1slt A 1smr A 1sos A 1spa 1srd A 1sry A
1st3 1stf I 1stp 1sub 1tab I 1tbs 1ten 1tet H 1tfd 1tfi
1tgl 1tgs I 1thb A 1thg 1thm 1tie 1tim A 1tlk 1tme 1 1tml
1tnc 1tnf A 1ton 1top 1tpk A 1tpl A 1tpm 1trb 1tre A 1trm A
1tta A 1ttf 1ubq 1ula 1utg 1vaa A 1vab B 1vil 1vna 1vsg A
1wsy A 1xim A 1xis 1xla A 1yat 1ycc 1yea 1yeb 1ymb 1ypc 1
1ypi A 1ysa C 1zaa C 256b A 2aaa 2aai B 2abx A 2ach A 2act 2alp
2apr 2atc A 2bat 2bb2 2bbk H 2bjl 1 2bop A 2bpa 1 2cab 2cas
2cba 2ccx 2ccy A 2cdv 2cmd 2cna 2cpl 2cro 2ctc 2cts
2ctv A 2ctx 2cyp 2dnj A 2ech 2er7 E 2fb4 H 2fbj H 2fcr 2fx2
2fxb 2gbp 2gcr 2gst A 2hhm A 2hhr A 2hip A 2hmb 2hmq A 2hpd A
2hpr 2ig2 H 2igg 2ihl 2imn 2ldx 2lhb 2ltn A 2mad H 2mcg 1
2mcm 2mev 1 2mhb A 2mhr 2mip A 2mml 2mnr 2msb A 2mta C 2nck L
2nn9 2ohx A 2ovo 2pcb B 2pf1 2pia 2pka A 2pkc 2plt 2plv 1
2pmg A 2pna 2pol A 2reb 2rhe 2rn2 2rsp A 2sas 2sga 2sic I
2sn3 2sns 2snv 2stv 2tbv A 2tgf 2tgi 2tmd A 2tmn E 2tmv P
2tpr A 2trx A 2ts1 2tsc A 2uce 2wrp R 2yhx 2yhx 351c 3adk
3b5c 3bcl 3blm 3c2c 3cd4 3chy 3cla 3cms 3dfr 3eca A
3est 3fxc 3gap A 3grs 3hfm H 3il8 A 3ink C 3lad A 3ldh 3mds A
3mon A 3ovo 3p2p A 3pfk 3pgk 3pgm 3psg 3rp2 A 3rub L 3sc2 A
3sdh A 3sdp A 3sgb E 3trx 3xia 4azu A 4bp2 4cpv 4dfr A 4enl
4fab H 4fgf 4fxn 4gcr 4gpd 1 4hvp A 4icb 4mdh A 4mt2 4ptp
4rcr H 4sbv A 4sgb 1 4tms 5cyt R 5fbp A 5fd1 5ldh 5p21 5pal
5pti 5rub A 5tim A 6ins E 6ldh 6rxn 6taa 7aat A 7api A 7cat A
7fab H 7icd 7pcy 7rsa 8abp 8dfr 8fab A 8i1b 8rub L 8rxn A
9ldt A 9pcy 9rnt 9wga A

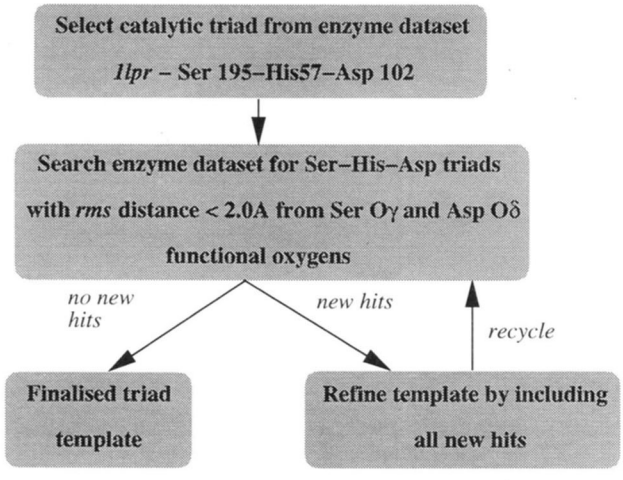

Fig. 1. Flow diagram showing the main steps involved in the calculation of the 3D template.

compared against this template. The RMS distance of its closest Asp carboxyl oxygen and its Ser O^{γ} to the reference atoms was calculated. An RMS cut-off distance of 2.0 Å was used, all triplets with a lower distance being retained.

Once all matching triplets had been extracted, a new template was calculated using their coordinates. However, to avoid biasing the results toward the largest of the fold groups (Group 1 in Table 1), a separate mean template was calculated for each of the four fold groups and these means were averaged to give an overall mean template. As before, the template involved only the coordinates of the two functional oxygens in the triad. The resultant 3D template was then taken as the new starting template and the entire above procedure was repeated. At each cycle, a few more triads would be pulled in, and the 3D template further refined, until no more new hits were obtained. A similar iterative procedure was used by Barth et al. (1993) to extract Ser-His-Asp triads from the serine proteases, although their method differed in the fitting of the triads and the definition of matches.

We will refer to the final 3D template that resulted from the above procedure as the "functional template," because it gives only the coordinates of the two functional Asp and Ser oxygens. A second template, referred to as the "side-chain template," was also computed. This used the coordinates of all the Asp and Ser side-chain atoms, namely Ser C^{α}, Ser C^{β}, Ser O^{γ}, Asp C^{α}, Asp C^{β}, Asp $O^{\delta 1}$, and Asp $O^{\delta 2}$. Again, a separate template was computed for each of the four fold groups, as well as an overall one. The calculation was slightly complicated by the Asp having two carboxyl oxygens, either one of which, $O^{\delta 1}$ or $O^{\delta 2}$, might be the functional one, their names being defined in the PDB files according to the IUPAC conventions and defined by the appropriate torsion angle. Therefore, we took whichever of the oxygens was hydrogen bonded to the His imidazole ring as the functional oxygen, and the other as the nonfunctional one. In computing the overall RMS distance, the distances between corresponding functional and nonfunctional oxygens were used. In some cases, this definition gives an artificially increased RMS distance. For example, consider the case where the two functional oxygens of equivalent Asp's coincide, but their nonfunctional oxygens are on opposite sides of this oxygen. Our RMS distance will be larger than the value a standard RMS distance comparison would give but, as a result, has the advantage of allowing triads with unusual side-chain conformations to be identified more easily.

The Ser-His-Asp templates

The overall "functional template," derived by the iterative procedure described above, is shown in Table 3. It comprises the 3D coordinates of the reference His side chain plus the two functional oxygens on the Ser and Asp residues. As will be seen shortly, the positions of these two functional oxygens are well conserved across all four fold groups. Furthermore, as will also be shown, they are specific to the catalytic triads; that is, the oxygen positions in noncatalytic associations of Ser, His, and Asp residues differ sufficiently to make the template particularly selective for the catalytic associations of these three residues.

The side-chain template involves all the Ser and Asp side-chain atoms and, because these are not conserved across the fold groups, no meaningful overall template can be derived for the side chains as a whole. Indeed, as will be seen, the Ser and Asp side chains adopt quite different conformations in the different structural groups while managing to conserve the positions of their important functional oxygens. This variability among the serine protease and lipase families has been observed before (see, for example, Barth et al., 1993, 1994), but is of interest here in that it is automatically picked up as a by-product of template generation and can be quantified easily for comparative purposes. The differences can be seen in Figure 2, which shows a representative Ser-His-Asp triad from each of the four fold groups, and one can see that there are quite marked differences between them, as will be discussed later.

Functional templates

Let us consider the functional template first, this being the one extracted and refined from a starting seed catalytic triad. Figure 3 illustrates just how good the resultant template is at discriminating between catalytic and noncatalytic associations of Ser, His, and Asp residues. The black bars in the histogram correspond to the catalytic triads with an RMS distance of <2.0 Å from the functional template, whereas the white bars correspond

Table 3. *Coordinates of the 3D functional template for the Ser–His–Asp catalytic triad*

			Coordinates		
Residue	Number	Atom	*x*	*y*	*z*
Ser	195	O^{γ}	−1.15	4.87	−0.07
Asp	102	$O^{\delta 2}$	3.68	0.06	0.06
His	57	$C^{\delta 2}$	−1.09	0.80	0.00
His	57	C^{γ}	0.00	0.00	0.00
His	57	$N^{\delta 1}$	1.11	0.82	0.00
His	57	C^{β}	0.07	−1.50	0.01
His	57	$C^{\epsilon 1}$	0.70	2.09	0.00
His	57	$N^{\epsilon 2}$	−0.66	2.09	0.00
Ser	214[a]	O^{γ}	5.01	2.26	1.71
Ser	125[a]	O^{γ}	2.28	5.71	−2.81

[a] Mean coordinates of the O^{γ} atom in the noncatalytic Ser 214 of fold. Group 1 enzymes and the Ser 125 of fold Group 2 enzymes.

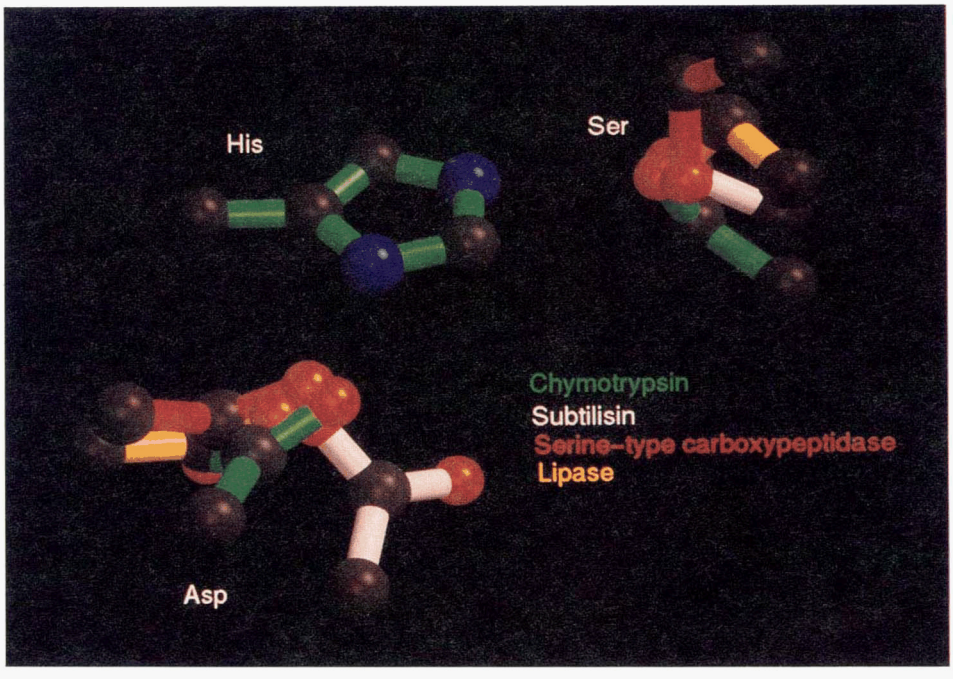

Fig. 2. Conformations of representative catalytic triads from each of the four fold groups: Group 1 in green, chymotrypsin 1cho (Fujinaga et al., 1987); Group 2 in white, subtilsin 2sic (Takeuchi et al., 1991); Group 3 in red, serine-type carboxypeptidase 3sc2 (Liao et al., 1992); and Group 4 in yellow, lipase 1tah (Noble et al., 1993), showing the different conformations adopted by the Ser and Asp side chains. Triads have all been superimposed on their histidine residue. The diagram was produced using Raster3D (Bacon & Anderson, 1988; Merritt & Murphy, 1994).

to those too far from the 2.0-Å cutoff to be picked up. Examination of the latter structures reveals that, in all cases, the conformation of the Ser-His-Asp triad has been perturbed by the presence of an inhibitor, so displacing the residues from their uncomplexed positions.

What is more important in Figure 3, however, is that none of the noncatalytic associations (grey and hatched bars) lie within the 2.0-Å cutoff; all are outside it, and most a long way outside. In other words, the automatic filtering procedure is capable of discriminating between catalytic and noncatalytic associations of the Ser, His, and Asp residues. This in turn suggests that the conformation of the two functional oxygens in the catalytic triad is in some way special; the oxygens are not found at these positions when the triplet is a noncatalytic one. Indeed, it can be seen from Figure 3 that a cutoff of 2.0 Å may even be too generous. The black bars have a strong peak at around 0.4–1.0 Å, and the majority lie below 1.4 Å. The majority of structures in the data set have inhibitors bound to the active site; examination of the structures with RMS distances between 1.4 Å and 2.0 Å reveals that the active sites are distorted by these inhibitors and that those above the 2.0-Å cut-off have covalently bound inhibitors. Figure 3 has other interesting features that are discussed below.

Conservation of the functional oxygens

The strong conservation of the two functional oxygens is illustrated in Figure 4, which shows the functional templates for each of the four fold groups. The tighter of the two clusters corresponds to the functional Asp carboxyl oxygen, which is in a position where it can hydrogen bond to the $N^{\delta 1}$ of the His. The other cluster corresponds to the Ser O^{γ} positions, which are within hydrogen bonding distance of the His $N^{\epsilon 2}$.

The differences in these positions across the four fold groups are quantified in Table 4. This shows the RMS distances between the functional templates of each group, together with their distances from the overall functional template. Among the groups, the smallest difference is 0.34 Å, which is between the templates of Groups 1 and 2 (both serine proteases), and the largest is 1.04 Å, between the trypsin-like serine proteases of Group 1 and the serine-like carboxypeptidases of Group 3.

The differences between each of the four group templates and the overall template ranges from 0.17 Å for Group 2 to 0.65 Å for Group 3. In other words, the overall template lies closer to that of the Group 2 enzymes than to the other group templates. The template for the largest group of enzymes, the trypsin-like serine proteases of Group 1, lies 0.47 Å away, and this accounts for the large peak at around 0.5 Å in Figure 3; this peak corresponds to the highly populated Group 1 enzymes.

Column 4 in Table 4 shows the measure of the variability of the functional oxygen positions within each fold group. The separations are quite small, ranging from 0.45 Å to 0.65 Å. They tend to be less than the differences between the functional groups, suggesting that the differences are genuine and arise from the differences in the overall folds of the proteins.

Side-chain templates

As mentioned above, the side-chain templates are not well conserved across the four fold groups because the Asp and Ser side chains tend to have markedly different orientations while

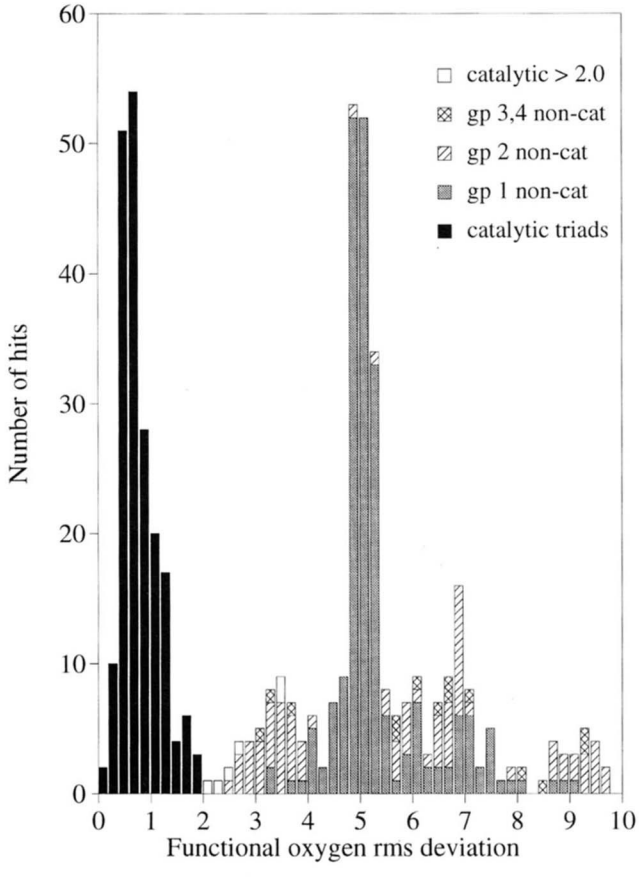

Fig. 3. Histogram of the RMS distance of the Ser O^{γ} and Asp carboxyl oxygen atom from the overall mean consensus template position for all Ser, His, and Asp associations in the enzyme data set. This histogram shows how the majority of catalytic triads, in black, are within an RMS distance of 2.0 Å from the consensus template, and indeed can be separated from the noncatalytic interactions using 2.0 Å as a cut-off. In the remaining catalytic triads (white bars), lying beyond the 2.0-Å cut-off, the conformation of the triad is perturbed by the binding of an inhibitor. In addition, the triads at an RMS distance of 2–5.5 Å in the histogram represent structurally conserved noncatalytic triads that play a role in hydrogen bonding to and orientating the catalytic triads (see text).

still managing to present their functional oxygens to the His in strongly conserved positions. Table 5 quantifies the RMS differences between these templates, together with their distances from the overall side-chain template. The most similar templates are those of Groups 3 and 4, both with an RMS distance of 0.87 Å. The similarities in their Asp and Ser conformations can be seen in Figure 2. In terms of sequence homology, the enzymes in these two groups are not particularly similar, sharing a sequence identity of 11%. The side-chain templates of Groups 1 and 2 are also reasonably similar (RMS 1.33 Å) and the sequence identity, at 16%, is again low.

The most marked differences are between Groups 2 and 3 (RMS 2.76 Å), Groups 2 and 4 (RMS 2.30 Å), and Groups 1 and 3 (RMS 2.23 Å). These differences can be seen in Figure 2, where, first, the Ser side chain of Groups 1 and 2 originates from below the plane of the His, whereas that of Groups 3 and 4 comes from above. The side-chain templates of Groups 3 and 4 have the greatest similarity, indeed they both have similar α/β folds. There are also differences in the conformations of the Asp side chains for each of the four groups, most noticeably in the subtilisin-like serine proteases (Group 2), where a different oxygen on the Asp is the functional one.

Within the Group 1 serine proteases, there are three subgroups (see Table 1) where the sequence identity between members of different subgroups are <30%. Despite these differences, the side-chain templates for the catalytic triads are very similar, with the template of subgroup 1a having an RMS distance of 0.35 Å from that of subgroup 1b α-lytic proteinase, and 0.53 Å from subgroup 1c lysyl endopeptidase.

The "catalytic tetrad"

Returning to Figure 3, which shows the distribution of the RMS distances of all Ser-His-Asp triplets from the overall functional template, one sees that there are actually very many noncatalytic Ser-His-Asp associations (grey and hatched bars) in this data set. What is more, those that come from the Group 1 enzymes (grey bars) show a marked peak at around 5.0 Å, suggesting a conserved conformation of a different, noncatalytic Ser-His-Asp triad. It turns out that this triad is strongly associated with the catalytic triad; in fact, it involves the catalytic His 57 and Asp 102 residues, together with a different Ser than the catalytic Ser 195, namely Ser 214. There is a near 1:1 ratio of the triads at the peaks of 0.5 Å and 5.0 Å because Ser 214 is found in a structurally conserved position in all Group 1 active sites. Indeed, these four residues, the catalytic triad plus Ser 214, have been termed the "catalytic tetrad" (Barth et al., 1993).

To check the extent of this structural conservation, we derived a consensus template for the Ser 214-His 57-Asp 102 triplet in

Table 4. *RMS distances between overall "functional templates"*

	No. of chains[a]	No. of catalytic triads[b]	Mean RMS distance of group[c]	Overall template	Fold group 1	Fold group 2	Fold group 3	Fold group 4
Overall template:	225	195	0.77	–	0.47	0.17	0.65	0.39
Fold group 1	170	152	0.62	0.47	–	0.34	1.04	0.71
Fold group 2	35	29	0.58	0.17	0.34	–	0.80	0.42
Fold group 3	7	4	0.65	0.85	1.04	0.80	–	0.87
Fold group 4	13	10	0.45	0.39	0.71	0.42	0.87	–

[a] Total number of chains for each fold group in the enzyme data set.
[b] Number of catalytic triads identified using overall functional template.
[c] Mean RMS deviation from the group's mean template of each of the fold group's members.

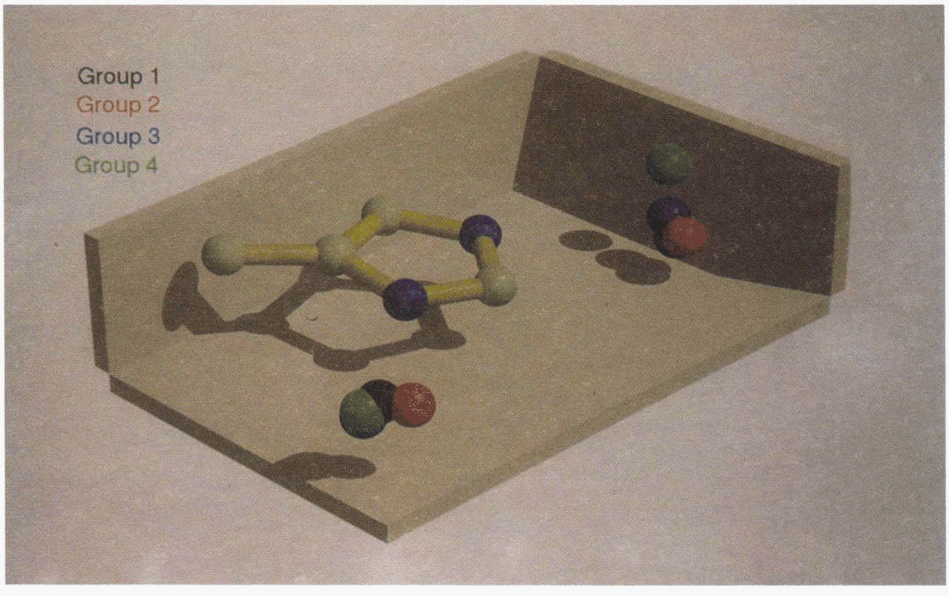

Fig. 4. Box-plot showing the mean positions of the Ser O^{γ} and the Asp carboxyl oxygen atom for each of the four fold groups (black, yellow, blue, and green, respectively). These atoms all converge at favorable hydrogen bonding positions relative to the nitrogens of the His ring. The tighter of the two clusters at the bottom of the picture corresponds to the Asp oxygen positions, and the other cluster represents the Ser oxygen.

the Group 1 enzymes. The method was the same as before, but using the Ser 214-His 57-Asp 102 triad of 1lpr as the starting seed. We found that these triplets, like the catalytic triads, all have an RMS distance of less than 2.0 Å from their overall mean. In other words, the position of the noncatalytic Ser 214 is just as tightly defined as the catalytic Ser 195. The relative coordinates of its functional oxygen are given in Table 3. It appears that the Ser 214, together with the peptide backbone of the catalytic His 57, forms a network of hydrogen bonds that enable the Asp 102 to be presented in the optimal position for interaction with the His 57 imidazole ring. Indeed, it has been suggested that the Ser 214 residue performs an electrostatic stabilization role. Mutation of this residue causes decreases in free energy of catalysis, in agreement with electrostatic calculations (McGrath et al., 1992). In another mutation experiment, Corey et al. (1992) investigated the effect of swapping this noncatalytic serine with the functional Asp. They formed the double mutation of D102S and S214D into the gene coding for rat anionic trypsin, expressed this in *Escherichia coli*, and then solved the X-ray structure. The Asp in this mutant was in a totally different position to that in the catalytic triad, though it still formed a hydrogen bond to the His imidazole ring. There was catalytic activity, but the k_{cat} was reduced 100-fold, indicating that, although a charged Asp in the vicinity of the His is sufficient for low catalytic activity, the position and hydrogen bond interactions of the Ser 214 residue are important for efficient catalysis. In addition, Noble et al. (1993) expressed the mutant D263A in the catalytic triad of the lipase of *Pseudomonas glumae* and found it still exhibited catalytic activity. They postulate that a neighboring Glu residue may adopt the role of the Asp. This suggests that the Asp residue in this particular lipase is both dispensable and in an unusual chemical environment.

Figure 3 has another peak at RMS distance 2–4.5 Å (hatched bars). This, as in the Group 1 case above, corresponds to a second Ser residue associated with the catalytic His and Asp, this time in the Group 2 enzymes (the subtilisin-like serine proteases).

Table 5. *RMS distances between overall "side-chain templates"*

	No. of chains[a]	No. of catalytic triads[b]	Mean RMS distance of group[c]	Overall template	Fold group 1	Fold group 2	Fold group 3	Fold group 4
Overall template:	225	195	1.32	–	1.49	3.27	2.39	1.59
Fold group 1	170	152	0.67	1.49	–	1.33	2.23	1.74
Fold group 2	35	29	0.70	3.27	1.33	–	2.76	2.30
Fold group 3	7	4	0.83	2.39	2.23	2.76	–	0.87
Fold group 4	13	10	1.06	1.59	1.74	2.30	0.87	–

[a] Total number of chains for each fold group in the enzyme data set.
[b] Number of catalytic triads identified using overall functional template.
[c] Mean RMS deviation from the group's mean template of each of the fold group's members.

The triplet consists of Ser 125-His 64-Asp 32. Again, we derived a consensus template for this triplet, this time using the Ser 125-His 64-Asp 32 triplet from 2sic as the starting seed. We found that the position of the functional oxygen of the Ser 125 is also strongly conserved: there were 29 Ser 215-His 64-Asp 32 triplets identified from a total of 35 chains with all RMS distances well below 1.00 Å. The coordinates of the functional oxygen are given in Table 3.

The serine-type carboxypeptidases of Group 3 do not have a conserved noncatalytic serine. However, it has been suggested that Asn 176 plays a similar role (Liao et al., 1992), but, unlike the two serines, the Asn carboxylate is out of the plane of the imidazole ring, which calls into question the significance of this residue.

There is a further peak in Figure 3 at 6–7 Å RMS distance, which corresponds to noncatalytic triplets where the Ser O^{γ} hydrogen bonds to His $N^{\delta 1}$, whereas the Asp carboxyl O hydrogen bonds to His $N^{\epsilon 2}$. This is the opposite hydrogen bonding conformation to that of a catalytic triad and is catalytically inactive because the Ser lies close to the histidine backbone and so would cause steric hindrance to a substrate.

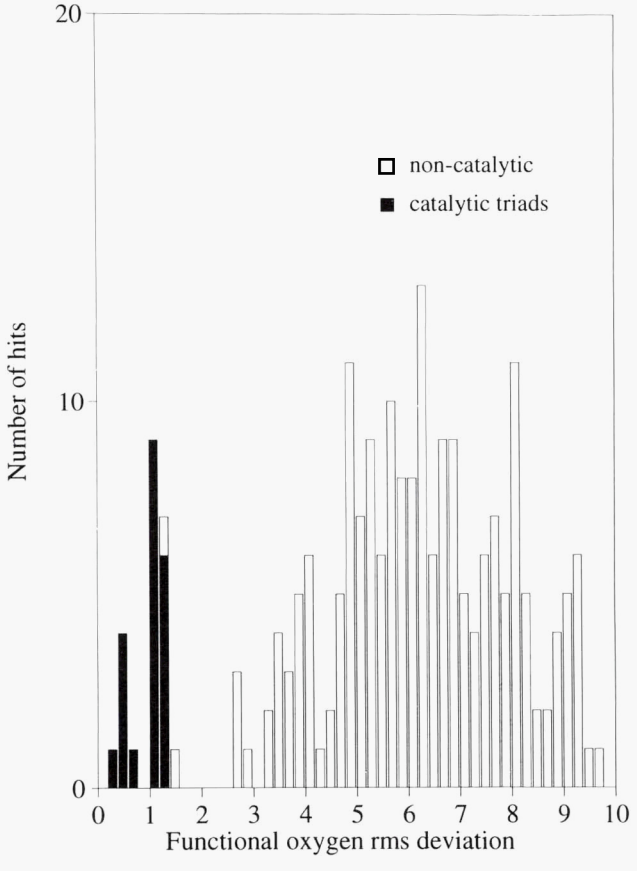

Fig. 5. Histogram showing the RMS deviation from the "functional" consensus template of all Ser-His-Asp interactions extracted from a data set of nonhomologous proteins. The serine proteinases are shown in black and these are clearly separated from the other, noncatalytic associations. There are, however, two proteins that are not serine proteinases, cyclophilin and immunoglobulin, shown in grey, that appear to have a Ser-His-Asp triad in the catalytic conformation.

Search for Ser-His-Asp triads in other PDB entries

We have established that the overall functional template, along with an RMS distance cut-off of 2.0 Å, is able to identify all Ser-His-Asp catalytic triads with the exclusion of all noncatalytic associations. We next applied the template to search a representative data set of 639 protein chains taken from the PDB (Table 2). This data set contains some of the structures present in the enzyme data set, the ones omitted being those excluded on the basis of having a higher than 95% sequence identity. The results of the search are shown in the histogram in Figure 5. As before, the black bars correspond to the catalytic triads in the serine proteases and lipases present in the data set. These all lie well within 2.0 Å of the template. Most of the other triplets (grey bars), which come from the other structures, have an RMS distance of > 2.0 Å. However, there are two proteins that are neither serine proteinases nor lipases, but that have an RMS deviation below the 2.0-Å cut-off. These are the Ser 99-His 92-Asp 123 triad of cyclophilin A, 2cpl (Ke, 1992), which has an RMS of 1.38 Å, and Ser 191-His 225-Asp 222 of chain H of immunoglobulin G1 2ig2, with an RMS of 1.57 Å (Marquart et al., 1980).

Cyclophilin A

The first of these, cyclophilin A, is a binding protein for the immunosuppressive drug cyclosporin A and is also an enzyme with peptidyl-prolyl *cis-trans* isomerase activity. Figure 6 shows the location of the identified Ser 99-His 92-Asp 123 triad (white bonds) in the protein. Although the protein's enzymatic mechanism is not yet fully known, various residues have been identified as possibly important for catalysis. One of these is His 126, which is shown with black bonds in Figure 6, very close to the

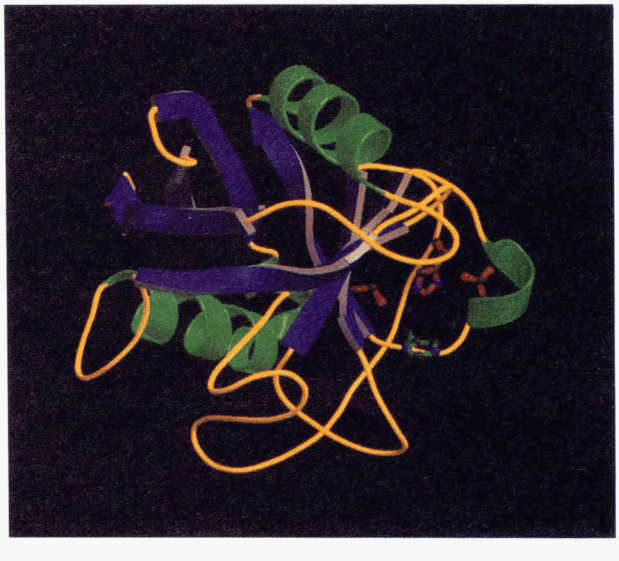

Fig. 6. MOLSCRIPT diagram (Kraulis, 1991) of cyclophilin A showing the Ser 99-His 92-Asp 123 triad, in white bonds, picked up by the Ser-His-Asp functional template. This triad may enable cyclophilin A to exhibit protease activity. Also shown is His 126 (black bonds), which is thought to be involved in the peptidyl-prolyl *cis-trans* isomerase activity.

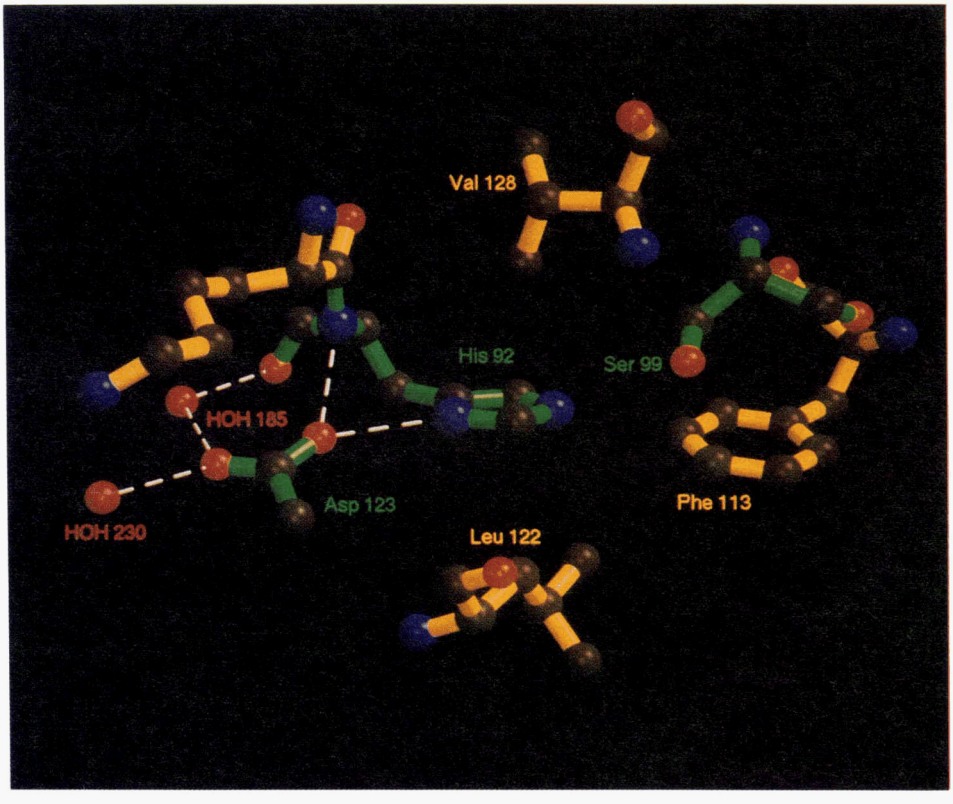

Fig. 7. Ser-His-Asp triplet from cyclophilin A, which adopts a catalytic triad conformation yet is not known to be catalytically active. The "catalytic" Ser O^{γ} has the side chain of Phe 113 lying directly below it, making the binding of a substrate sterically unfavorable.

triad, lending weight to the possibility of this protein having proteinase activity. Figure 7 shows a close-up of the cyclophilin A "catalytic" triad. The Ser O^{γ} and Asp carboxyl oxygen appear to be in optimal hydrogen bonding positions. The Asp carboxyl oxygens form a network of stabilizing hydrogen bonds with the surrounding residues, making the Asp very similar in nature to a real catalytic aspartate.

However, the triad is surrounded by three hydrophobic residues: Val 128 and Leu 122 lying above and below the His 92 ring, and Phe 113 directly below the catalytic Ser 99 O^{γ}. It appears that the steric hindrance caused by these residues would inhibit any substrate binding. There is the possibility that the binding of a substrate could cause a conformational change in the enzyme, forcing the Phe 113 to move out of the way of the Ser and enable catalysis. To test this possibility, the accessibility and reactivity of the Ser to substrate could be tested using diisopropylphosphofluoridate, which should form an irreversible covalent adduct with the Ser, as it does in all serine proteinases and lipases (Hayashi et al., 1973). To date, no proteinase-like activity has been reported for cyclophilin A, although, because enzymes are highly specific for substrate, the protein may not have been sufficiently assayed for proteinase activity.

The overall structure of cyclophilin A (Fig. 7) is a β-barrel comprising eight antiparallel β-strands with two α-helices sitting on the top and bottom of the barrel. Because this topology is unlike any of the fold groups in Table 1, the protein is unlikely to be evolutionarily related to the serine proteases or lipases.

Immunoglobulin G1

The second non-enzyme with an apparent catalytic triad is immunoglobulin G1, 2ig2 (Marquart et al., 1980). The triad is Ser 191-His 225-Asp 222, which has an RMS deviation of 1.57 Å from the overall functional template. Figure 3 indicates that the majority of catalytic triads lie within 1.4 Å of the mean template, and those between 1.4 and 2.0 Å have bound inhibitors. This suggests that this triad may not have the stringent conformation needed to be catalytically active. Indeed, its location in the structure (Fig. 8) is at the C-terminus of the molecule, near the hinge region, and would probably be buried if the X-ray structure of the whole Ig molecule were available. In addition, it is far away from the hapten binding site, which is situated at the N-terminus.

A close-up of the triad is shown in Figure 9, where it can be seen that it is surrounded by mostly hydrophobic residues (green bonds). It also looks rather different from the triads in the serine proteases and lipases (cf. Fig. 2) because the Asp side chain approaches the His from a completely different orientation. This is reflected in a high RMS distance of 4.20 Å between this triad and the overall side-chain template.

The triad lies on the surface of the heavy chain of the immunoglobulin molecule with the Ser O^{γ} pointing out, toward the surface. There is also a second Ser residue in the vicinity, Ser 192 (Fig. 9), which might be compared with the Ser 214 or Ser 125 residues of the Group 1 and Group 2 enzymes, respectively (Ta-

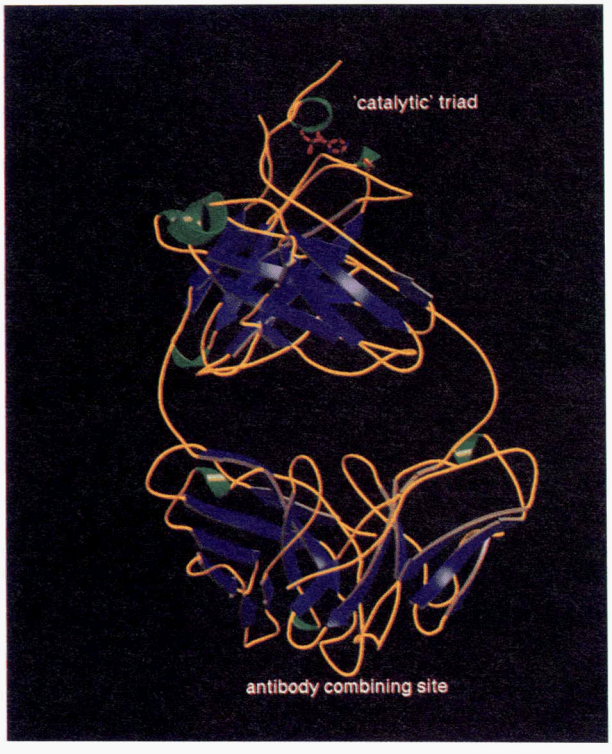

Fig. 8. MOLSCRIPT diagram (Kraulis, 1991) of the intact immunoglobulin fragment with the position of the "catalytic" triad shown. The triad lies at the C-terminus of the molecule, whereas the hapten binding site is at the N-terminus.

ble 1). However, its location, although next to the "catalytic" Ser 191, is quite different.

Discussion

We have derived a simple 3D template for the well-known and much-studied Ser-His-Asp triads of the serine proteases and lipases. The template reveals a number of new and interesting aspects of these triads. The first is just how strongly the positions of the functional oxygens are conserved across all the different fold groups, with the majority having an RMS distance of under 1.4 Å. What is more, the general, noncatalytic associations of Ser, His, and Asp residues tend not to make use of these special oxygen locations. Taken together, this strong conservation of position and its uniqueness to the catalytic systems, enables the use of the template as an automatic filtering process for the catalytic triads. It also provides a means of measuring how unusual any given triad conformation is, particularly where the binding of an inhibitor disrupts it.

The strong conservation of the functional oxygen positions is achieved in different ways in the four different fold groups. The orientations of the Asp and Ser residues vary markedly from one group to the next, although they are well conserved within each group.

When the template was used to search through a representative data set of structures in the PDB, apart from identifying correctly all the known Ser-His-Asp catalytic triads, it also located two very interesting matches: one in cyclophilin A, which is known to have some enzymatic activity, although not protease activity, and in immunoglobulin G1, where any enzyme activity at the located triad is most unlikely. A similar search performed by Barth et al. (1994) revealed several other matches. However, the triads were fitted on the C^{α} and C^{β} of the Asp and Ser as well as the $N^{\delta 1}$ and $N^{\epsilon 2}$ of the His, explaining the discrepancies in the matches. Perhaps the most interesting result of our search through the PDB is that we found only these two noncatalytic matches. One might expect a large enough data set of proteins to contain some fortuitous Ser-His-Asp associations that satisfy the template, yet do not have any catalytic function. The fact that only the immunoglobulin G1 appears to be the only fortuitous "hit"—and then with an RMS outside the usual catalytic triad cut-off of 1.4 Å—demonstrates the special nature of the conformation of the oxygens in the catalytic triad.

In this paper, we have only dealt with the Ser-His-Asp catalytic triad. However, there are many similar systems to which the approach could be applied. For example, not all catalytic triads employ an Asp as the acid group, but instead use a glutamate. This is true in two esterases: *Geotrichum candidum* lipase (Schrag et al., 1991) and *Torpedo californica* acetylcholinesterase (Sussman et al., 1991). Both of these proteins have Ser-His-Glu catalytic triads. Comparison with our overall functional template shows that the RMS distances from the template of the Glu $O^{\epsilon 1}$ and the Ser O^{γ} in these triads are 0.70 Å and 0.72 Å, respectively. These values lie well within the RMS cut-off that defines a catalytic triad, supporting the view that the Glu and Asp have the same role in their respective triads.

In addition, Wei et al. (1995) have identified a novel catalytic triad in *Streptomyces scabies* that employs the main-chain carbonyl of a Trp residue to hydrogen bond to the His $N^{\delta 1}$ in place of the Asp carboxyl oxygen. Here the catalytic triad is Ser 144-His 283-Trp 280. Again, applying our template shows that the RMS distance from our template of the two functional oxygens—the Ser O^{γ} and the main-chain Trp carbonyl oxygen—is 1.92 Å. This is larger than the 1.4 Å seen for "native" enzymes, but is still below the 2.0-Å cut-off defined above. Examination of this structure shows that the Ser O^{γ} is distorted away from the mean template position by 2.64 Å because the structure contains the covalent inhibitor bis-*p*-nitrophenylmethylphosphonate. We have already noted that the Ser-His-Asp catalytic triad is sometimes distorted by binding of unusual or covalent inhibitors.

Of course, there are other enzyme systems present in the PDB that employ a His residue as part of their catalytic machinery; for example, papain (Cys, His, Asn) and malate dehydrogenase (His, Asp). An extension of this work would be to use our template triad to investigate the structural similarities in catalytic centers of such enzymes. This work is currently in progress.

An additional application of the method would be for cases where there is no prior knowledge of functionally important residue associations. The filtering procedure described here requires only a single triad as a starting seed to act as an initial functional template. At each cycle of the iterative procedure, similar triads, which are sufficiently close in conformation to the starting template, are brought in and used in refining the template for the next cycle. The final result is a functional template plus all the residue associations that match it. The whole procedure merely needs a initial seed conformation to start it off. Such a seed can be generated automatically with no prior knowledge of its functional importance. Thus, for example, for an association of residues X-Y-Z, the first such association encountered

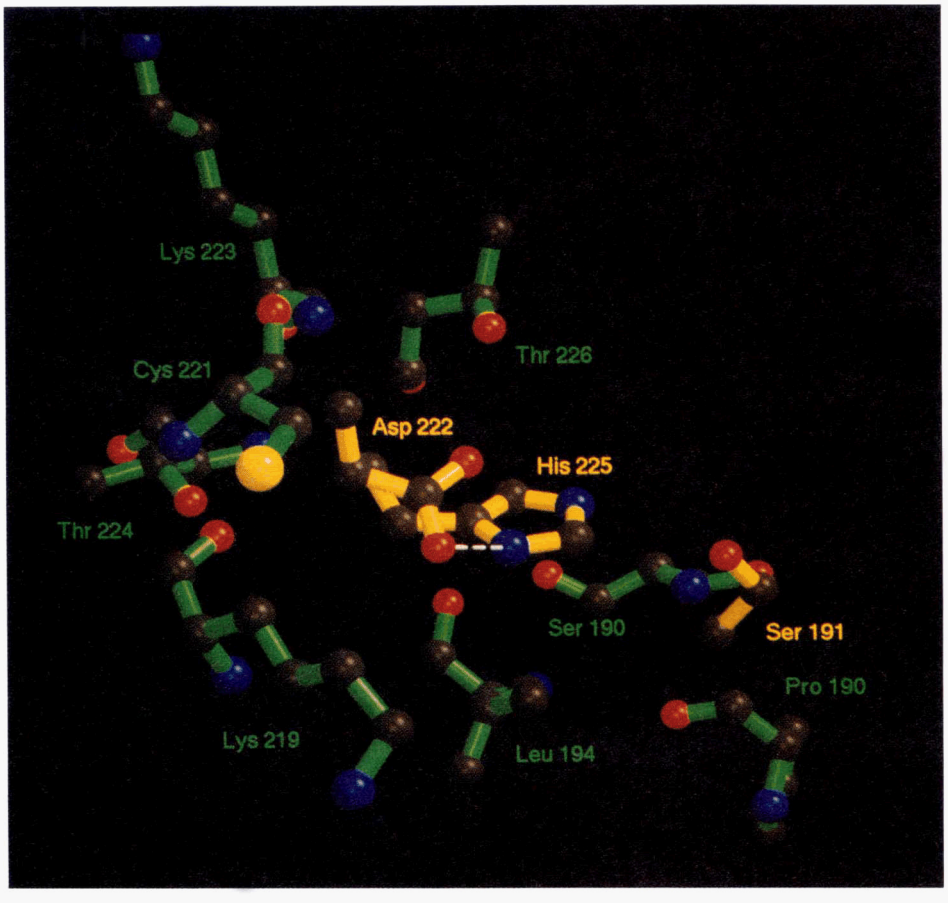

Fig. 9. Ser 191-His 225-Asp 222 triad (yellow bonds) found in the immunoglobulin molecule G1 (2ig2, Marquart et al., 1980).

in the data set of structures could be taken as the starting seed. If sufficient matches were found during the subsequent cycles, the template could be stored as a common triad motif that might have either functional or structural significance. With this approach, it should be possible to build a database of common 3D residue motifs, much like the functional motifs used in the PROSITE database (Bairoch & Bucher, 1994).

Acknowledgments

We thank Drs. Y. Wei, J.L. Schottel, U. Derewenda, L. Swenson, S. Patkar, and Z.S. Derewenda for kindly providing their *Streptomyces scabies* esterase coordinates. We also thank Neera Borkakoti, Laurence Pearl, Keith Brocklehurst, Christine Orengo, and Alex Michie for their help and useful suggestions. A.C.W. is supported by a BBSRC CASE studentship, sponsored by Roche Products, Ltd., and R.A.L. is supported by a grant from Parke Davis, Inc.

References

Artymiuk PJ, Poirrette AR, Grindley HM, Rice DW, Willett P. 1994. Graph-theoretic approach to the identification of three-dimensional amino acid side-chains in protein structures. *J Mol Biol 243*:327–344.

Attwood TK, Beck ME, Bleasby AJ, Parry-Smith DJ. 1994. PRINTS—A database of protein motif fingerprints. *Nucleic Acids Res 22*:3590–3596.

Bacon DJ, Anderson WF. 1988. A fast algorithm for rendering space-filling molecular pictures. *J Mol Graph 6*:219–220.

Bairoch A, Boeckmann B. 1994. The SWISS-PROT protein sequence data bank: Current status. *Nucleic Acids Res 22*:3578–3580.

Bairoch A, Bucher P. 1994. PROSITE: Recent developments. *Nucleic Acids Res 22*:3583–3589.

Barth A, Frost K, Wahab M, Brandt W, Schlader HD, Franke R. 1994. Classification of serine proteases derived from steric comparisons of their active site geometry, part II: Ser, His, Asp arrangements in proteolytic and non-proteolytic proteins. *Drug Design and Discovery 12*:89–111.

Barth A, Wahab M, Brandt W, Frost K. 1993. Classification of serine proteases derived from steric comparisons of their active sites. *Drug Design and Discovery 10*:297–317.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol 112*:535–542.

Bielka H, Dixon HBF, Karlson P, Liebecq C, Sharon N, Van Lenten EJ, Velick SF, Vliegenthart JFG, Webb EC. 1992. *E.C. Enzyme Nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the nomenclature and classification of enzymes*. Nomenclature Committee of the International Union of Biochemistry. London: Academic Press.

Blow DM. 1976. Structure and mechanism of chymotrypsin. *Acc Chem Res 9*:145–152.

Blow DM. 1990. More of the catalytic triad. *Nature 221*:337–340.

Blow DM, Birktoft JJ, Hartley BS. 1969. Role of a buried acid group in the mechanism of action of chymotrypsin. *Nature 221*:337–340.

Bone R, Fujishige A, Kettner CA, Agard DA. 1991. Structural basis for broad specificity in α-lytic protease mutants. *Biochemistry 30*:10388–10398.

Brady L, Brzozowski AM, Derewenda ZS, Dodson E, Dodson G, Tolley S, Turkenburg JP, Christianson L, Huge Jensen B, Norskov L, Thim L, Menge U. 1990. A serine protease triad forms the catalytic centre of triacylglycerol lipase. *Nature 343*:767–770.

Brzozowski AM, Derewenda U, Derewenda ZS, Dodson GG, Lawson DM, Turkenburg JP, Bjorkling F, Huge Jensen B, Patkar SA, Thim L. 1991. A model for interfacial activation in lipases from the structure of a fungal lipase-inhibitor complex. *Nature 351*:491–497.

Corey DR, McGrath ME, Vasquez JR, Fletterick RJ, Craik CS. 1992. An alternative geometry for the catalytic triad of serine proteases. *J Am Chem Soc 114*:4905–4907.

Derewenda U, Brzozowski AM, Lawson DM, Derewenda ZS. 1992. Catalysis at the interface: The anatomy of a conformational change in a triglyceride lipase. *Biochemistry 31*:1532–1541.

Fischer D, Wolfson H, Lin SL, Nussinov R. 1994. Three-dimensional, sequence order-independent comparison of a serine protease against the crystallographic database reveals active site similarities: Potential implications to evolution and to protein folding. *Protein Sci 3*:769–778.

Fujinaga M, Sielecki AR, Read RJ, Ardelt W, Laskowski M, James MNG. 1987. Crystal and molecular structures of the complex of α-chymotrypsin with its inhibitor turkey ovomucoid domain at 18 Å resolution. *J Mol Biol 195*:397–418.

Glusker JP. 1991. Structural aspects of metal liganding to functional groups in proteins. *Adv Protein Chem 42*:1–76.

Hayashi R, Moore S, Stein WH. 1973. Serine at the active center of yeast carboxypeptidase II. *J Biol Chem 248*:8366–8369.

Hodgman TC. 1989. The elucidation of protein function by sequence motif analysis *CABIOS 5*:1–13.

Jernigan R, Raghunathan G, Bahar I. 1994. Characterization of interactions and metal-ion binding-sites in proteins. *Curr Opin Struct Biol 4*:256–263.

Ke H. 1992. Similarities and differences between human cyclophilin A and other β-barrel structures. *J Mol Biol 228*:539–550.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Liao DI, Breddam K, Sweet R, Bullock T, Remington SJ. 1992. Refined atomic model of wheat serine carboxypeptidase II at 22 Å resolution. *Biochemistry 31*:9796–9812.

Mann KG. 1987. The assembly of blood clotting complexes on membranes. *Trends Biochem Sci 12*:229–233.

Marquart M, Deisenhofer J, Huber R, Palm W. 1980. Crystallographic refinement and atomic models of the intact immunoglobulin molecule kol and its antigen-binding fragment at 30 Å and 19 Å resolution. *J Mol Biol 141*:369–391.

McGrath ME, Vasquez JR, Craik CS, Yahg AS, Honig B, Fletterick RJ. 1992. Perturbing the polar environment of Asp 102 in trypsin: Consequences of replacing the conserved Ser 214. *Biochemistry 31*:3059–3064.

Merritt EA, Murphy MEP. 1994. Raster3D version 2.0: A program for photorealistic molecular graphics. *Acta Crystallogr D 50*:869–873.

Murzin AG, Brenner SE, Hubbard T, Chothia C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol 247*:536–540.

Noble MEM, Cleasby A, Johnson LN, Egmond MR, Frenken LGJ. 1993. The crystal structure of triacylglycerol lipase from *Pseudomonas glumae* reveals a partially redundant catalytic aspartate. *FEBS Lett 331*:123–128.

Ollis D, Cheah E, Cygler M, Dijkstra B, Frolow F, Franken SM, Harel M, Remington SJ, Silman I, Schrag J, Sussman JL, Verschueren KHG, Goldman A. 1992. The α/β hydrolase fold. *Protein Eng 5*:197–211.

Orengo CA, Flores TP, Taylor WR, Thornton JM. 1993. Identification and classification of protein fold families. *Protein Eng 6*:485–500.

Perona J, Craik C. 1995. Structural basis of substrate specificity in the serine proteases. *Protein Sci 4*:337–360.

Schrag JD, Yungi Li, Shan Wu, Miroslaw C. 1991. Ser-His-Glu triad forms the catalytic site of the lipase from *Geotrichum candidum*. *Nature 351*:761–764.

Sussman JL, Harel M, Frolow F, Oefner C, Goldman A, Toker L, Silman I. 1991. Atomic structure of acetylcholinesterase from *Torpedo californica*: A prototypic acetylcholine-binding protein. *Science 253*:872–879.

Takeuchi Y, Satow Y, Nakamura KT, Mitsui Y. 1991. The refined crystal structure of the complex of subtilisin and *Streptomyces subtilisin* inhibitor at 18 Å resolution. *J Mol Biol 221*:309–325.

Taylor WR. 1988. Pattern matching in protein sequence comparison and structure prediction. *Protein Eng 2*:77–86.

Taylor WR, Jones DT. 1991. Templates, consensus patterns and motifs. *Curr Opin Struct Biol 1*:327–323.

Wei Y, Schottel JL, Derewenda U, Swenson L, Patkar S, Derewenda ZS. 1995. A novel variant of the catalytic triad in the *Streptomyces scabies* esterase. *Nature Struct Biol 2*:218–223.

Wright CS, Alden RA, Kraut J. 1969. Structure of subtilsin bpn at 25 Å resolution. *Nature 221*:235–242.