Protein Engineering vol.7 no.1 pp.57–64, 1994

Evolutionary divergence and conservation of trypsin

Wojciech R.Rypniewski[1], Anastassis Perrakis, Constantin E.Vorgias and Keith S.Wilson

European Molecular Biology Laboratory (EMBL), c/o DESY, Notkestrasse 85, 22603 Hamburg, Germany

[1]To whom correspondence should be addressed

The trypsin sequences currently available in the data banks have been collected and aligned using first the amino acid sequence homology and, subsequently, the superposed crystal structures of trypsins from the cow, the bacterium *Streptomyces griseus* and the fungus *Fusarium oxysporum*. The phylogenetic tree constructed according to this multiple alignment is consistent with a continuous evolutionary divergence of trypsin from a common ancestor of both prokaryotes and eukaryotes. Comparison of crystal structures reveals a strict conservation of secondary structure. Similarly, in the alignment of all the sequences, insertions and deletions occur only in regions corresponding to loops between the secondary structure elements in the known crystal structures. The conserved residues cluster around the active site. Almost all conserved residues can be associated with one of the basic functional features of the protein: zymogen activation, catalysis and substrate specificity. In contrast, the residues of the hydrophobic core of the protein and the calcium ion binding sites are generally not conserved. The conserved features of trypsin and the nature of the conservation are discussed in detail.

Key words: alignment/evolution/structure/trypsin

Introduction

Trypsin (EC 3.4.21.4) is a member of a large and diverse family of serine proteinases characterized by a common catalytic mechanism involving three essential residues: serine, histidine and aspartate, known as the catalytic triad. Trypsin specifically cleaves the peptide bond on the carboxyl side of lysine or arginine residues [for reviews, see Steitz and Shulman (1982) and Kraut (1977)]. In the pancreas, trypsin functions not only as a digestive enzyme but is also responsible for activating all the pancreatic enzymes, including itself, by cleaving a short propeptide from the N-terminus of the inactive zymogens. Bovine trypsin was among the first proteolytic enzymes isolated and analysed (Northrop *et al.*, 1948). It has since been identified in a wide variety of organisms. Of particular significance was the discovery of a trypsin in a bacterium *Streptomyces griseus* (Olafson *et al.*, 1975). Its close homology to mammalian trypsins, comparable to the homology between trypsin and related enzymes within the same organism, was puzzling and led Hartley (1970, 1979) to hypothesize that a gene transfer had taken place from a mammal to a bacterium. Hewett-Emmett *et al.* (1981) disputed that by constructing an evolutionary tree with the bacterial enzyme at the root. An examination of the crystal structure led Read and James (1988) to suggest that the need to stabilize the buried charge of Asp189, required for tryptic specificity, could impose stricter structural requirements and account for the apparently greater conservation of sequence in trypsin than in related enzymes. Another bacterial trypsin, from *Streptomyces erythraeus* (Miyamoto *et al.*, 1979; Yamane *et al.*, 1991) was investigated and found to have similar properties to that from *S.griseus*. Recently a trypsin was found in the mould *Fusarium oxysporum* and its crystal structure solved in our laboratory (Rypniewski *et al.*, 1993). A number of other trypsin genes have recently been sequenced.

In this paper the trypsin sequences currently available in the data banks are collected and aligned using both amino acid sequence homology and superposed, available crystal structures. The evolution of trypsin is examined by constructing a phylogenetic tree. The sequences are then compared and the significance of the conserved features discussed.

Materials and methods

Trypsin sequences were extracted from the four major data banks using the computing facilities of EMBL Heidelberg utilizing the GCG package. The sequences were identified by searching for the pattern 'trypsin' among the entries of the Swissprot-PIR protein sequence as well as the EMBL and GenBank nucleotide sequence data banks. The selected entries were manually evaluated. Only complete sequences were selected. Nucleotide sequences were translated into protein. The sequence of trypsin from *F.oxysporum* (Rypniewski *et al.*, 1993) was appended. Initial alignment was carried out using the program CLUSTALV (Higgins *et al.*, 1992).

The atomic coordinates of the crystal structures of trypsin from *Bos taurus* (cow) (Chambers and Stroud, 1979) and *S.griseus* (Read and James, 1988), obtained from the Brookhaven Protein Data Bank (Bernstein *et al.*, 1977; Abola *et al.*, 1987) and *F.oxysporum* were superposed by minimizing the least-squares differences in position between the main chain atoms of those residues which are identical in all three proteins (Figure 1). The alignment of the superposed structures was obtained by examining them on an Evans and Sutherland ESV graphics station running FRODO. The secondary structure elements were determined from the crystal structures using the program DSSP (Kabsch and Sander, 1983).

The structure alignment was used to manually modify the initial, automatic alignment of all the sequences. The three sequences were made to match as in the crystal structures while maintaining their alignment with the other sequences. In a few cases, in regions where there was no sequence homology or crystal structure to use as a guide, the alignment was ambiguous. In such cases care was taken not to introduce breaks or insertions where secondary structure elements could be expected from the crystal structures. This procedure is justified by the remarkable conservation of the secondary structure evident from the crystallographic analysis.

A pairwise percent divergence matrix was calculated taking into account identities in all positions where there were no gaps in each pair. A phylogenetic tree was constructed from the percent

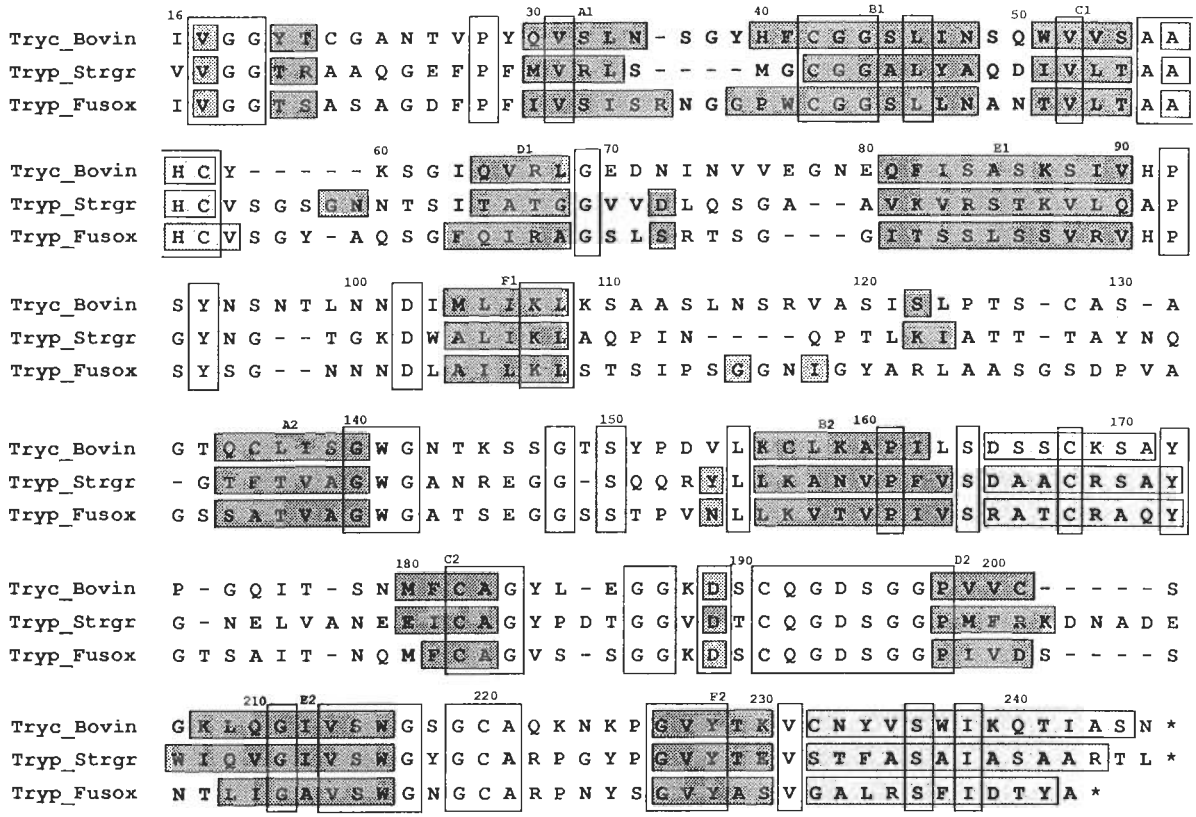

Fig. 1. Alignment of sequences of trypsins from *B.taurus* (cow) (Tryc_Bovin), *S.griseus* (Tryp_Strgr) and *F.oxysporum* (Tryp_Fusox) based on the superposed crystal structures. The structures were superposed by minimizing the least-square differences in position between the main chain atoms of those residues that are identical in all three proteins (indicated by boxes). Secondary structure elements are shaded: β-strands and isolated β-bridges as defined in Kabsch and Sander (1983) (dark shading) and helices (light shading). The β-strands labelled A1-F1 and A2-F2 form the N-terminal and the C-terminal domains respectively. The residues are numbered according to the convention used in bovine trypsin.

divergence matrix using the Neighbour Joining method of Saitu and Nei (1987) supplied with CLUSTALV. No correction was applied for multiple substitutions. The phylogenetic tree was drawn with the DRAWTREE program from the PHYLIP package (Felsenstein, 1993).

Results and discussion

Sequence alignment

The automatic sequence alignment correctly matched the conserved residues known to be essential for catalysis and specificity, with the exception of Asp102 which lies in a poorly conserved region. Varying the gap penalty parameters or grouping sequences and matching their profiles did not remove the mismatches and they had to be corrected by hand. More seriously, the automatic alignment of the sequences, for which the crystal structures were known, introduced gaps and mismatches in several β-strands and contained several mismatches in the loop regions. As the mismatches occur in regions of low sequence similarity it is not surprising that the program failed to match the sequences correctly in the absence of information of 3-D structures. However, this clearly demonstrates the problems of homology modelling of structures based on sequences which are not very closely related.

The alignment of the three crystal structures, shown in Figure 1, made it possible to check and amend the alignment of the other sequences with a reasonable degree of confidence. This was helped by the fact that the three proteins came from distantly related organisms: a bacterium, a fungus and a mammal. For example, bovine trypsin matches the other mammalian trypsins unambiguously and to a lesser extent, the other vertebrates. The homology of trypsin from *F.oxysporum* is approximately equally distant from the bacterial enzymes and from the mammals and shares some features with both. The sequences from insects and, to a lesser degree, from fish presented the most ambiguity as they contain regions of little sequence homology to other enzymes.

Two sequences were removed from the initial data set. An entry in the PIR database, described as a trypsin from papaya, was found to be identical to bovine cationic trypsin. On checking the original reference it became clear that the sequence was placed in the database by mistake. One of the sequences from *Aedes aegypti* (AATRYP), tentatively identified by the authors (Kalhok *et al.*, 1992) as a trypsin, has a serine in position 189. This residue is in the specificity pocket and has an important role in defining the substrate specificity. Serine is consistent with chymotrypsin, not trypsin which requires an aspartate in this position. Although this sequence was removed from the final data set it was included as an 'outlier' in the calculation of the phylogenetic tree. Table I lists the trypsin sequences used in the alignment. The final sequence alignment, shown in Figure 2, is based both on sequence homology and on knowledge of the available crystal structures.

Phylogenetic tree

The phylogenetic tree, shown in Figure 3, is consistent with a continuous evolution of trypsin from a single ancestral gene. This resolves the question of the origin of trypsin in prokaryotes. The gene duplication and subsequent divergence of cationic and anionic forms of the enzyme is reflected in the consistent

Table I. The 31 trypsins used in this study

No.	Organism	Proposed name	Reference
1	*Streptomyces griseus*	Tryp_Strgr	Olfason *et al.* (1975)
2	*Saccharopolyspora erythraea*	Tryp_Sacer	Miyamoto *et al.* (1979)
3	*Fusarium oxysporum*	Tryp_Fusox	Rypniewski *et al.* (1993)
4	*Choristoneura fumiferana*	Tryp_Chofu	Wang *et al.* (1992)
5	*Aedes aegypti*	Tryp_Aed5G	Kalhok *et al.* (1992)
6	*Aedes aegypti*	Tryp_Aed3A	Kalhok *et al.* (1992)
7	*Drosophila melanogaster*	Trya_Drome	Davis *et al.* (1985)
8	*Drosophila melanogaster*	Tryb_Drome	Magoulas and Hickey (1992)
9	*Drosophila melanogaster*	Trye_Drome	Magoulas and Hickey (1992)
10	*Astacus fluviatilis*	Tryp_Astfl	Titani *et al.* (1983)
11	*Squalus acanthias*	Tryp_Squac	Hermondson *et al.* (1973)
12	*Pleuronectes platessa*	Tryp_Plepl	Leaver and George (1992)
13	*Salmon salar*	Try1_Salsa	Male *et al.* (1992)
14	*Salmon salar*	Try2_Salsa	Male *et al.* (1992)
15	*Salmon salar*	Try3_Salsa	Male *et al.* (1992)
16	*Xenopus laevis*	Tryp_Xenla	Shi and Brown (1990)
17	*Sus scrofa*	Tryp_Pig	Hermondson *et al.* (1973)
18	*Bos taurus*	Trya_Bovin	LeHuerou *et al.* (1990)
19	*Bos taurus*	Tryc_Bovin	Mikes *et al.* (1966)
20	*Canis familiaris*	Try1_Canfa	Pinsky *et al.* (1985)
21	*Canis familiaris*	Try2_Canfa	Pinsky *et al.* (1985)
22	*Mus musculus*	Tryp_Mouse	Stevenson *et al.* (1986)
23	*Rattus norvegicus*	Try1_Rat	MacDonald *et al.* (1982)
24	*Rattus norvegicus*	Try2_Rat	MacDonald *et al.* (1982)
25	*Rattus norvegicus*	Try3_Rat	Fletcher *et al.* (1987)
26	*Rattus norvegicus*	Try4_Rat	Luetcke *et al.* (1989)
27	*Rattus norvegicus*	Try5a_Rat	Kang *et al.* (1992)
28	*Rattus norvegicus*	Try5b_Rat	Kang *et al.* (1992)
29	*Homo sapiens*	Try1_Human	Emi *et al.* (1986)
30	*Homo sapiens*	Try2_Human	Emi *et al.* (1986)
31	*Homo sapiens*	Try3_Human	Tani *et al.* (1990)

The proposed names are based on the Swissprot nomenclature system. Where the name was given by Swissprot it was not changed, otherwise it was assigned by us.

clustering of the mammalian sequences in two groups. This supports the hypothesis that a single gene duplication event occurred prior to the divergence of mammals. The 'outlier' sequence Chym_Aed (Kalhok *et al.*, 1992), which we had identified as a likely chymotrypsin, does not cluster with the other insect proteins and occupies the longest branch on the tree.

Using the sequence alignment based only on sequence homology, without incorporating the information from the crystal structure alignment, gives a very similar tree. Presumably, in the current data set enough diverse species are represented to derive their relationship robustly. However, in our earlier attempt, with only 16 sequences, a consistent branching could only be obtained when the 3-D structural information was used to amend the initial sequence alignment.

Conservation of secondary structure

The conservation of secondary structures is evident in the superposition of the crystal structures (Figure 1), even in regions where sequence similarity is low. Insertions and deletions occur only in loops between the secondary structure elements. Similarly, the final alignment in all sequences (Figure 2) has consistent insertions and deletions, all lying in segments corresponding to loop regions in the crystal structures.

Conserved residues

The conserved residues are listed in Table II. A few substitutions have been allowed in defining the conserved residues. Many of these are conservative substitutions, preserving the polarity or size of the side chain. The conserved residues cluster in the region of the active site (Figure 4). They can be classified according to their likely role in the basic functions of trypsin: zymogen activation, catalysis, specificity and structural stability. Although this classification is useful for the purpose of discussion, it is inevitably imperfect, not only because the function of a particular residue is, with few exceptions, difficult to determine, but also because some residues may be important in more than one way.

Zymogen activation. After cleaving off the propeptide the new N-terminus is buried in the C-terminal domain forming an ion pair with Asp194 and making several hydrogen bonds. The point of burial of the N-terminus is located a considerable distance from the N-terminal domain. The distance is bridged by a segment of 13 residues ending at conserved Pro28. Gly18 and Gly19 define a sharp turn into the body of the protein allowing the burial of the N-terminus which plays a key role in the activation mechanism (Huber and Bode, 1978). Asp194 is conserved in all the sequences listed in this study and its importance is indicated by the fact that it is an internal, charged residue. Its functional importance must outweigh the cost of burying the charge. Adjacent to Asp194 is the highly conserved segment Gly140–Trp141–Gly142 and three internal water molecules observed in the crystal structures. The waters participate in hydrogen bonds linking this segment with Asp194 and the catalytic Ser195 while the side chain of Trp141 stacks against another conserved residue Leu155. This cluster of apparently invariant residues (Figure 5), probably represents the essential features of the means of maintaining the active conformation for

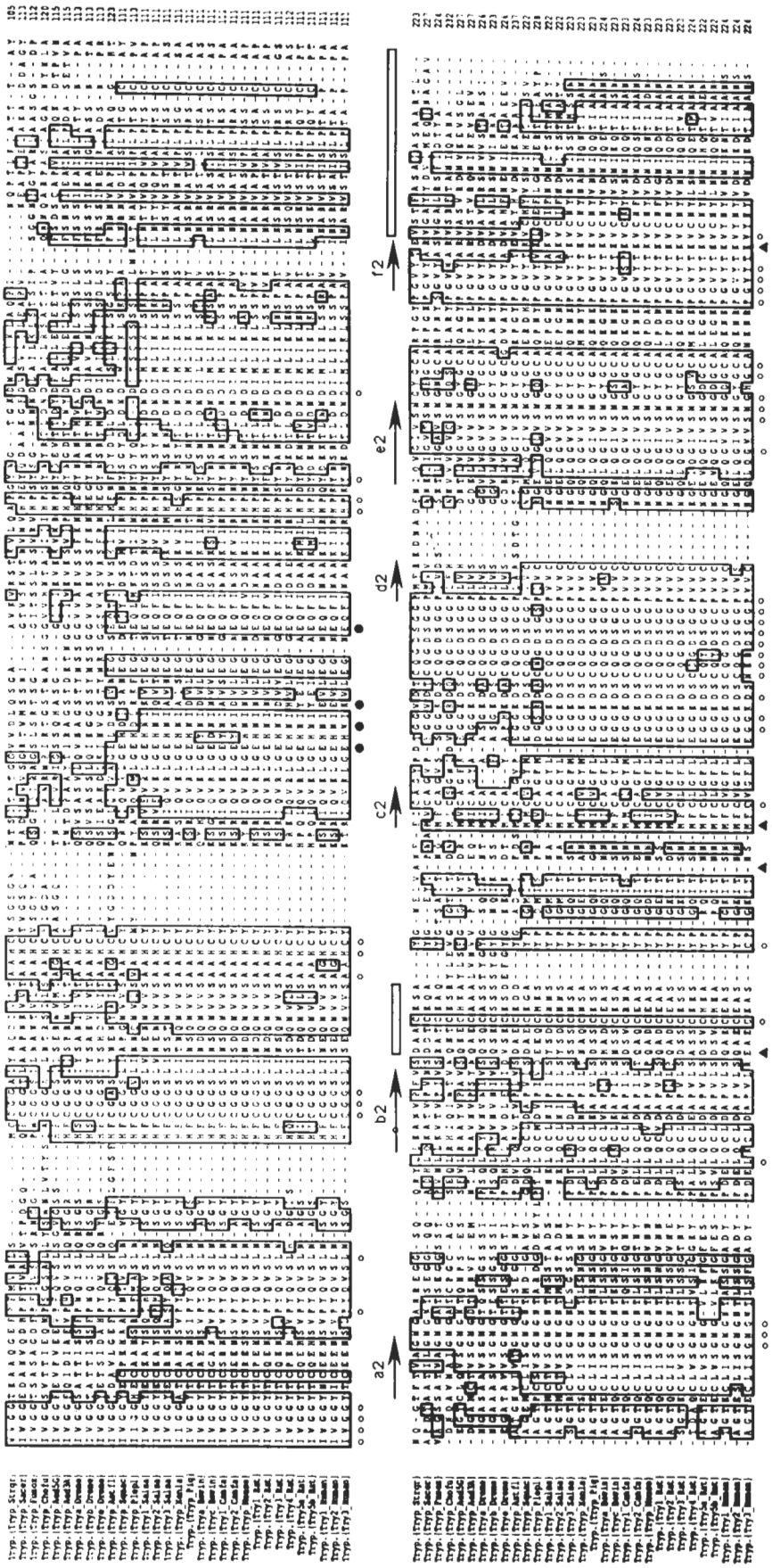

Fig. 2. The sequences of the 31 trypsins aligned as described in Materials and methods and named as in Table I. The secondary structure elements, as observed in the crystal structures, are indicated along the top as arrows (β-strands) and bars (helices). The highly conserved residues, listed in Table II, are indicated underneath by small circles. Residues participating in binding calcium in the known crystal structures, are indicated by filled circles (Tryc__Bovin) and triangles (Tryp__Strgr). Regions with at least 50% homology are boxed.

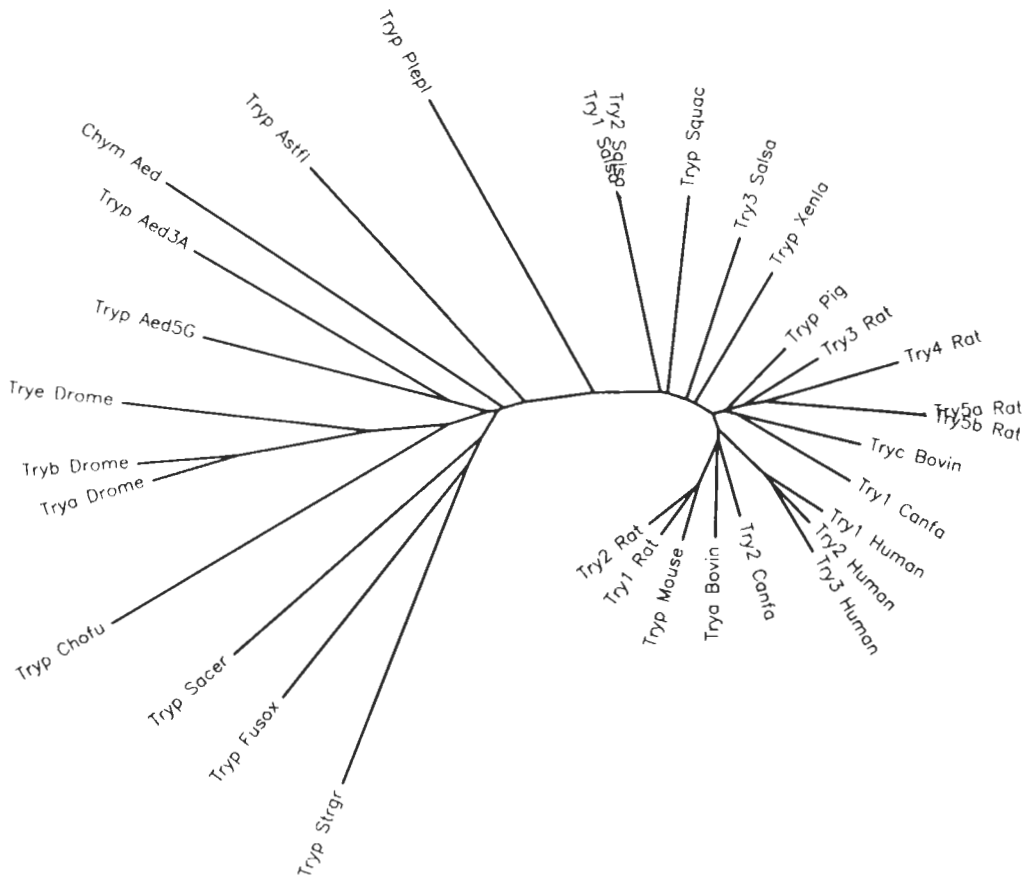

Fig. 3. The unrooted phylogenetic tree of trypsin inferred from the amino acid sequence alignment shown in Figure 2. The tree was constructed as described in Materials and methods.

catalysis while countering the destabilizing effect of the buried charge of Asp194.

Catalysis. In addition to the catalytic triad His57, Asp102 and Ser195, a number of neighbouring residues are strongly conserved. Gly43 forms a hydrogen bond with the carbonyl oxygen of Ser195. The adjacent Gly44 is strongly conserved. Ala55 lies directly underneath the catalytic triad and its methyl group is in close contact with His57, Asp102 and Cys42. In Tryp__Plepl this residue is replaced by a valine and in Tryp__Aed3A by an arginine. It is hard to see how a larger side chain, particularly an arginine, can be accommodated without disrupting the alignment of the catalytic residues or causing a major rearrangement of the neighbouring structure. Residues His91, Pro92 and Tyr94 lie on the exposed side of the loop containing the catalytic Asp102 and seem to play a role in maintaining its position while the top of the loop shows substantial variability in the crystal structures. The carbonyl oxygen of His91 is hydrogen bonded to the amide nitrogen of Tyr94. The side chain of Tyr94 is in contact with the catalytic residues Asp102 and His57. The substitution of His91 by Ala in Tryp__Strgr is accompanied by a change of Trp/Phe to Ala at position 238. A compensating substitution of Ile/Leu by Trp at position 103 fills the vacant space. Gly193 forms a part of the loop holding Ser195 and its amide group stabilizes the reaction intermediate by hydrogen bonding the carbonyl oxygen of the substrate (Krieger *et al.*, 1974). Gly196 and Gly197, adjacent to the catalytic Ser195 are also strongly conserved. In one case only, in Tryp__Plepl, Gly197 is replaced by an alanine. The methyl group could be accommodated by displacing the conserved internal water seen in the crystal structures. The OH group of Ser214 makes a hydrogen bond to OD2 of Asp102 thus stabilizing its interaction with the catalytic histidine.

Specificity. Asp189 lies at the bottom of the specificity pocket and is required for the tryptic specificity for arginine and lysine. The conserved Gly186 and Gly187 facilitate its optimal positioning. Most of the residues 214–220, which form a loop defining the rim of the specificity pocket, are strongly conserved. Residue 217 is not conserved but its side chain points outwards from the specificity pocket. The loop is completed by the disulphide bridge between Cys220 and Cys191. In the crystal structures Gly216 and Gly219 have dihedral angles that would be sterically unfavourable for other residues. Gly226 is in contact with Asp189 in such a way that any substitution at this position would severely disrupt the orientation of the aspartate. The OH group of Tyr172 forms a hydrogen bond with Pro225, which is at the bottom of the specificity pocket, while its ring interacts with Trp215 and Val227. Tyr172 may play an indirect role in stabilizing the geometry of the specificity pocket, although it is substituted in several of the sequences. The crystal structure shows that the substitution of Pro225 by serine in Tryp__Fusox has not changed the main chain conformation. This stability is probably maintained by hydrogen bonds from N and OG to OE1 of Gln171 which in this protein replaces the usual Ser/Ala. Gln192 is located at the entrance of the binding pocket. In the various crystal structures of trypsin the side chain is only weakly ordered and in the zymogen the residue is completely disordered.

Table II. The conserved residues of trypsins used in this study

Residue	Comments	Exceptions
Ile16	Terminal amino group forms ion pair with Asp194, essential for zymogen activation	Tryp__Strgr (Val)
Val17	β-Sheet interaction stabilizing N-terminal ion pair	Tryp__Plepl, Tryp__Xenla (Ile)
Gly18	N-terminal bend	
Gly19	As above	
Pro28	End of N-terminal arm	Try1__Salsa (Thr), Try2__Salsa (Ser)
Leu33	Hydrophobic core	Tryp__Astfl (Phe), Tryp__Fusox (Ile)
Cys42	Disulphide bridge with Cys58	
Gly43	Contact with Ser195 loop. Internal residue	
Gly44	As above	Tryp__Astfl (Ala)
Ala55	Close contact with His57, Asp84 and Cys42	Tryp__Aed3A (Arg), Tryp__Plepl (Val)
His57	Catalytic triad residue	
Cys58	Disulphide bridge with Cys42	
His91	Residues 91−94 outer side of Asp102 loop	Tryp__Strgr (Ala)
Pro92	As above	all Try*x*__Drome, Tryp__Astfl (Glu), Tryp__Xenla (Ser)
Tyr94	As above. Contact with catalytic Asp102 and His57	Tryp__Astfl, Tryp__Pig, Try1__Rat (Phe)
Asp102	Catalytic triad residue	
Gly140	In hydrophobic core residues 139 and 141 hydrogen bonded to internal	
Trp141	waters, Trp stacks against Leu155	
Gly142		
Leu155	Hydrophobic core, stacks against Trp141	
Cys168	Disulphide bridge with Cys182	
Tyr172	Interacts with Trp215 and Val227. OH hydrogen bonded to Pro225 close to specificity pocket	Tryp__Chofu, Tryp__Aed3A (Val), Tryp__Aed5G (Phe), Try3__Human (Cys)
Met180	Hydrophobic core	Tryp__Strgr (Glu)
Cys182	Disulphide bridge with Cys168	
Gly186	Turn into specificity pocket	Trya,b__Drome (Ser), Trye__Drome (Pro)
Gly187	As above	Trye__Drome (His)
Asp189	Determines specificity	
Cys191	Disulphide bridge with Cys220, close to specificity pocket	Try3__Human (Trp)
Gln192	Closes active site	Tryp__Plepl (Asn), Try4__Rat (Asp),
Asp194	Buried charge. Ion pair with terminal amino group	
Ser195	Catalytic triad residue	
Gly196	Catalytic Ser195 loop	
Gly197	As above. In Tryp__Plepl side chain could replace conserved water	Tryp__Plepl (Ser)
Pro198	Internal residue	
Gly211	As above	
Ser214	OH hydrogen bonded to OD2 of Asp102	
Trp215	Hydrophobic core	
Gly216	Part of specificity pocket	
Gly219	As above	
Cys220	Disulphide bridge with Cys191	
Pro225	Mutation in Tryp__Fusox has little effect on local structure	Tryp__Fusox (Ser)
Gly226	Contact to Asp189	
Val227	Interacts with Trp215. Near specificity pocket	
Tyr228	As above. Hydrogen bonded through conserved water to OD1 of Asp189	Tryp__Chofu (Asp, Try1__Dog (Ser)
Val231	Hydrophobic core; start of C-terminal helix	Tryp__Plepl (Leu)

Krieger *et al.* (1974) suggested that the side chain of Gln192 plays a role in providing a polar environment for the polar side chains of the substrates. In three of the sequences Gln192 is replaced by other polar residues (see Table II).

Conserved residues of the hydrophobic core. Most residues of the hydrophobic core of the protein are not conserved. Several of those that are conserved are located in the vicinity of the active site and play a role in maintaining its conformation (see preceding sections and Table II). The conserved residues that appear not to be essential for enzymatic activity include Leu33, Met180, Pro198, Gly211 and Val231. In Tryp__Strgr Met180 is replaced by a glutamate. This does not affect the structure of the neighbouring residues, Trp215 and Val227, which take part in defining the specificity pocket. Val231 marks the end of the β-strand, F2 and the beginning of the C-terminal helix.

Disulphide bridges. Bovine trypsin has six disulphide bridges. Of these only three are conserved in the trypsins from *S.griseus* and *F.oxysporum*. These bridges are clearly important for structural stability. The disulphide bond between Cys191 and Cys220 also plays a role in maintaining the structure of the binding site by completing the loop of residues 214−220 which defines the rim of the specificity pocket. In Try3__Human Cys191 is replaced by a tryptophan. This implies that this disulphide bridge is not essential although it may affect the enzymatic activity.

Calcium binding sites. A number of serine proteinases have been found to require calcium for thermal stability and resistance to degradation [for reviews, see Kretsinger (1976) and Martin (1984)]. It has been suggested (Kretsinger, 1976) that the requirement for calcium ensures that the enzymes are not active in the cytoplasm where the calcium concentration is low. The calcium binding residues in the known structures are indicated in Figure 2. The site observed in bovine trypsin is conserved in vertebrates but not in lower animals. No Ca^{2+} binding sites

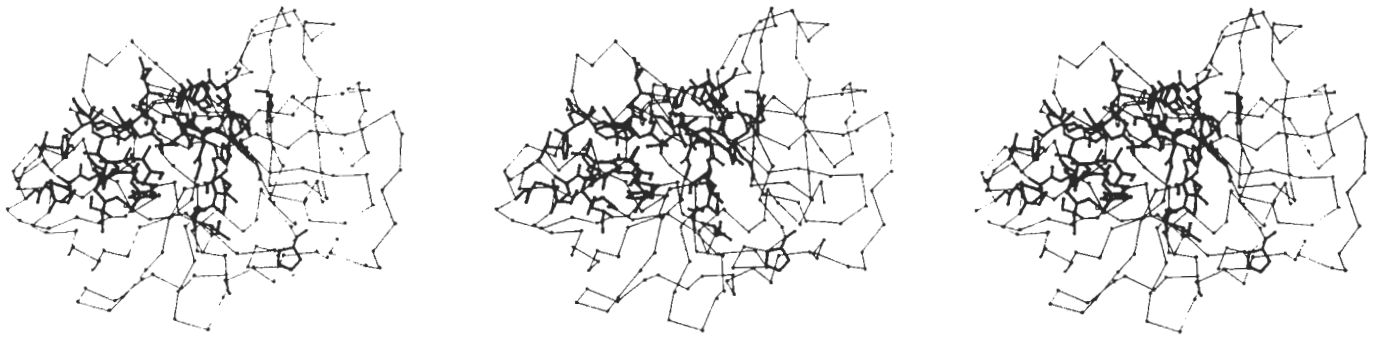

Fig. 4. The conserved residues, listed in Table II, drawn as in the structure of *F.oxysporum* trypsin (Tryp__Fusox). For the other residues only α-carbons are shown. The position of the active site is indicated by the inhibitor, DIP (diisopropyl phosphoryl), drawn with open bonds.

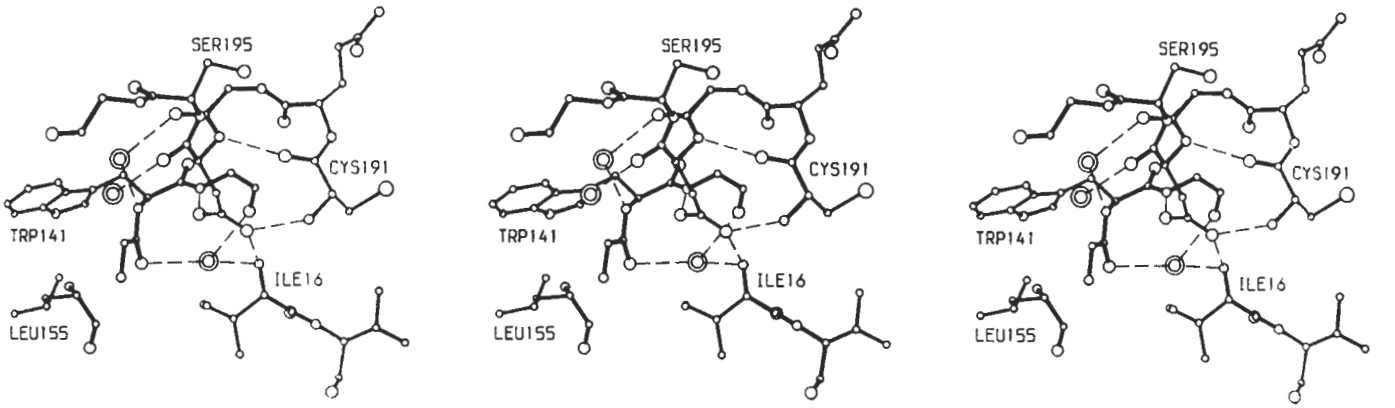

Fig. 5. The cluster of conserved residues and water molecules around Asp194 as seen in the structure of *F.oxysporum* trypsin. The water molecules are marked by two concentric circles. Main chain bonds are drawn with thicker lines. Hydrogen bonds and ion pairs are drawn with dashed lines.

have been found in the trypsins from *S.erythraeus* (Yamane *et al.*, 1991) and *F.oxysporum* (Rypniewski *et al.*, 1993). The trypsin from *S.griseus* (Read and James, 1988) has a calcium binding site located in a different position to that in the bovine trypsin. A possible candidate for having a Ca^{2+} binding site similar to that in *S.griseus* trypsin is Tryp__Astfl where residues Asp165 and Glu230, whose side chains participate in binding the calcium ion, are present.

Concluding remarks

The hydrophobic core interactions are generally not specific. This means that the internal residues should be replaced at an approximately constant rate determined by random mutations subject only to the constraint that complementary side chain substitutions are required to preserve the integrity of the protein interior. Consequently, the conservation of the internal, hydrophobic residues could be used as a measure of evolutionary distance between related proteins. In contrast, the residues necessary for the specific function of the protein are subject to absolute constraints. Comparing the rates of substitution of the internal, hydrophobic residues and other conserved residues gives a measure of the strength of these constraints. Given a sufficient number of sequences and structures of a protein from distantly related species one might expect that the residues subject to non-specific constraints have not been conserved. The conserved features should define the parts of the molecule that are essential for activity and structural integrity, even if the function of some of these residues is not fully understood.

Ancient proteins like trypsin are well suited for such studies. The list of conserved residues in Table II is an attempt at defining the most invariant and, therefore, most probably, the essential features. In the case of trypsin, specific constraints appear to act on ~30 residues or more than 10% of the protein. As more sequences become available it is possible to define the features essential for enzyme function more accurately. One reservation to this approach is that features such as the dipoles of the peptide bonds, postulated by Read and James (1988) to play a major role in stabilizing the buried, charged residues, do not depend directly on the nature of the side chains, but on the positioning of the main chain atoms. Sequence comparisons are not sufficient to investigate the role of such peptide bond dipoles. This emphasizes the need for detailed 3-D structural information to complement the sequence data.

The sequences and the alignment presented in this paper can be obtained from the authors by sending a request to the following Email address: Wojtek@EMBL-Hamburg.de.

Acknowledgement

We would like to express our appreciation and many thanks to C.Ouzounis for helping us in several computer problems.

References

Abola,E.E., Bernstein,F.C., Bryant,S.H., Koetzle,T.F. and Weng,J. (1987) In Allen,F.H., Bergerhoff,G. and Sievers,R. (eds), *Crystallographic Database—Information Content, Software Systems, Scientific Applications.* Data Commission of the International Union of Crystallography, Bonn/Cambridge/Chester, pp. 107–132.

Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F.,Jr, Brice,M.D., Rodgers,J.R., Kennard,O., Shomanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.

Chambers,J.L. and Stroud,R.M. (1979) *Acta Crystallogr.*, **35B**, 1861–1879.
Craik,C.S., Choo,Q.-L., Swift,G.H., Quinto,C., MacDonald,R.J. and Rutter,W.J. (1984) *J. Biol. Chem.*, **259**, 14255–14264.
Davis,C.A., Riddell,D.C., Higgins,M.J., Holden,J.J.A. and White,B.N. (1985) *Nucleic Acids Res.*, **13**, 6605–6619.
Emi,M., Nakamura,Y., Ogawa,M., Yamamoto,T., Nishide,T., Mori,T. and Matsubara,K. (1986) *Gene*, **41**, 305–310.
Felsenstein,J. (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Department of Genetics, University of Washington.
Fletcher,T.S., Alhadeff,M., Craik,C.S. and Largman,C. (1987) *Biochemistry*, **26**, 3081–3086.
Hartley,B.S. (1970) *Phil. Trans. R. Soc. Lond.*, **B257**, 77–87.
Hartley,B.S. (1979) *Proc. R. Soc. Lond.*, **B205**, 443–452.
Hermondson,M.A., Ericsson,L.H., Neurath,H. and Walsh,K.A. (1973) *Biochemistry*, **12**, 3146–3153.
Hewett-Emmett,D., Czelusniak,J. and Goodman,M. (1981) *Ann. NY Acad. Sci.*, **370**, 511–527.
Higgins,G.D., Bleasby,J.A. and Fuchs,R. (1992) *CABIOS*, **8**, 189–191.
Huber,R. and Bode,W. (1978) *Acc. Chem. Res.*, **11**, 114–122.
Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.
Kalhok,S., Tabak,L.M., Prosser,D.E., Downe,A.E.R. and White,B.N. (1992) EMBL AC X64362, X64363.
Kang,J., Wiegand,U. and Mueller-Hill,B. (1992) *Gene*, **110**, 181–187.
Kraut,J. (1977) *Annu. Rev. Biochem.*, **46**, 331–358.
Krieger,M., Kay,L.M. and Stroud,R.M. (1974) *J. Mol. Biol.*, **83**, 209–230.
Kretsinger,R.H. (1976) *Annu. Rev. Biochem.*, **45**, 239–266.
Leaver,M.J. and George,S.G. (1992) EMBL AC X56744.
LeHuerou,I., Wicker,C., Guilloteau,P. and Puigserver,A. (1990) *Eur. J. Biochem.*, **193**, 767–773.
Luetcke,R., Rausch,U., Vasiloundes,P., Scheele,G.A. and Kern,H.F. (1989) *Nucleic Acids Res.*, **17**, 6736.
MacDonald,R.J., Stary,S.J. and Swift,G.H. (1982) *J. Biol. Chem.*, **257**, 9724–9732.
Magoulas,C. and Hickey,D. (1992) EMBL AC M96372.
Male,R., Lorens,J.B., Smals,A.O., Jensen,M.F. and Torrissen,K.P. (1992) EMBL AC X70073, X70074, X70075.
Martin,R.B. (1984) In Siegel,H. (ed.), *Metal Ions in Biological Systems*, Vol. 17. Marcel Dekker, New York, pp. 1–49.
Mikes,O., Holeysovsky,V., Tomasek,V. and Sorm,F. (1966) *Biochem. Biophys. Res. Commun.*, **24**, 346–352.
Miyamoto,K., Matsuo,H. and Narita,K. (1979) In *Dai 30 Kai Tanpakushitsu Kouzou Touronkai Kouen Youshishuu* (in Japanese), pp. 77–80.
Northrop,J.H., Kunitz,M. and Herriott,R. (1948) In *Crystalline Enzymes*, 2nd edn, Ch. 6. Columbia Press, New York, pp. 125–167.
Olafson,R.W., Jurasek,L., Carpenter,M.R. and Smillie,L.B. (1975) *Biochemistry*, **14**, 1168–1177.
Pinsky,S.D., LaForge,K.S. and Scheele,G. (1985) *Mol. Cell. Biol.*, **5**, 2669–2676.
Read,R.J. and James,M.N.G. (1988) *J. Mol. Biol.*, **200**, 523–551.
Rypniewski,W.R., Hastrup,S., Betzel,Ch., Dauter,M., Dauter,Z., Papendorf,G., Branner,S. and Wilson,K.S. (1993) *Protein Engng*, **6**, 341–348.
Saitu,N. and Nei,M. (1987) *Mol. Biol. Evol.*, **4**, 406–425.
Shi,Y.-B. and Brown,D.D. (1990) *Gene Devel.*, **4**, 1107–1113.
Stevenson,B.J., Habenbuechle,O. and Wellauer,P.K. (1986) *Nucleic Acids Res.*, **14**, 8307–8330.
Steitz,T.A. and Shulman,R.G. (1982) *Annu. Rev. Biochem. Biophys.*, **11**, 419–444.
Tani,T., Kawashima,K., Mita,K. and Takiguchi,Y. (1990) *Nucleic Acid Res.*, **18**, 1631.
Titani,K., Sasagawa,T., Woodbury,R.G., Ericsson,L.H., Dorsam,H., Kraemer,M., Neurath,M. and Zwilling,R. (1983) *Biochemistry*, **22**, 1459–1465.
Wang,S., Magoulas,C. and Hickey,D.A. (1992) SwissProt AC L04749.
Yamane,T., Kobuke,M., Tsutsui,H., Toida,T., Suzuki,A., Ashida,T., Kawata,Y. and Sakiyama,F. (1991) *J. Biochem.*, **110**, 945–950.

Received May 5, 1993; revised July 21, 1993; accepted July 30, 1993