Extremophiles (1999) 3:205–219

ORIGINAL PAPER

Hanna-Kirsti Schrøder Leiros · Nils Peder Willassen
Arne O. Smalås

Residue determinants and sequence analysis of cold-adapted trypsins

Received: January 14, 1999 / Accepted: March 31, 1999

Abstract The digestive enzyme trypsin is among the most extensively studied proteins, and its structure has been reported from a large number of organisms. This article focuses on the trypsins from vertebrates adapted to life at low temperatures. Cold-adapted organisms seem to have compensated for the reduced reaction rates at low temperatures by evolving more active and less temperature-stable enzymes. We have analyzed 27 trypsin sequences from a variety of organisms to find unique attributes for the cold-adapted trypsins, comparing trypsins from salmon, Antarctic fish, cod, and pufferfish to other vertebrate trypsins. Both the "cold" and the "warm" active trypsins have about 50 amino acids that are unique and conserved within each class. The main unique features of the cold-adapted trypsins attributable to low-temperature adaptation seem to be (1) reduced hydrophobicity and packing density of the core, mainly because of a lower (Ile + Leu)/(Ile + Leu + Val) ratio, (2) reduced stability of the C-terminal, (3) lack of one warm trypsin conserved proline residue and one proline tyrosine stacking, (4) difference in charge and flexibility of loops extending the binding pocket, and (5) different conformation of the "autolysis" loop that is likely to be involved in substrate binding.

Key words Cold adaptation · Psychrophilicity · Residue determinants · Sequence comparison · Trypsin

Communicated by K. Horikoshi

H.-K. Schrøder Leiros · A.O. Smalås (✉)
Protein Crystallography Group, Department of Chemistry, Faculty of Science, University of Tromsø, N-9037 Tromsø, Norway
Tel. +47 77 64 40 70; Fax +47 77 64 47 65
e-mail: arne.smalas@chem.uit.no

N.P. Willassen
Institute of Medical Biology, Faculty of Medicine, University of Tromsø, N-9037 Tromsø, Norway

Introduction

In recent years, increasing attention has been drawn to the cold adaptation of enzymes, but the molecular basis of the phenomenon is still poorly understood, probably because of the limited number of amino acid sequences and structural data available. Examples of X-ray structures for cold-adapted enzymes reported so far are Atlantic salmon trypsin and elastase (Berglund et al. 1995a,b; Smalås et al. 1994), cod pepsin (Karlsen et al. 1998), α-amylase (Aghajari et al. 1998), and triosephosphate isomerase (Alvarez et al. 1998), while homology models have been determined on the basis of gene sequences for Antarctic fish trypsin and elastase (Aittaleb et al. 1997; Genicot et al. 1996), Atlantic cod trypsin and chymotrypsin (Gudmundsdóttir et al. 1993; Leth-Larsen et al. 1996), bacterial subtilisin (Davail et al. 1994), bacterial lipase (Arpigny et al. 1997), bacterial citrate synthase (Gerike et al. 1997), bacterial isopropylmalate dehydrogenase (Wallon et al. 1997), and bacterial isocitrate dehydrogenase (Ishii et al. 1993).

In this study the digestive enzyme trypsin is used as a model system for exploring the molecular basis of physical properties such as pH stability, temperature stability, stability in the presence of denaturants, primary and secondary binding specificities of substrates and inhibitors, and catalytic efficiency. Homologous sets of trypsins from different species show variations in these properties in spite of high sequence identity and three-dimensional structural similarity. The differences in physical properties must therefore be the result of small and subtle differences of the molecules. Even though the catalytic mechanism of serine proteinases in general and trypsin in particular is well characterized and documented, much less is known about the molecular basis of the above-mentioned properties. Because of its frequent citation in the literature, trypsin is well suited for the present study in searching for the molecular basis of enzyme cold activity. The EMBL and SWISSPROT databases include about 70 trypsin sequences, and many of the trypsins are characterized in terms of their kinetics and stability. A number of three-dimensional X-ray structures have been

determined, from vertebrates such as the pig (Huang et al. 1993), rat (Earnest et al. 1991), cow (Marquart et al. 1983), and human (Gaboriaud et al. 1996), from the bacteria *Streptomyces griseus* (Read and James 1988), and from the fungi *Fusarium oxysporum* (Rypniewski et al. 1995). In addition to the mentioned X-ray structures of the psychrophilic Atlantic salmon trypsin (Berglund et al. 1995a; Smalås et al. 1994), the structure of a cationic variant, also from Atlantic salmon, has been solved (Schrøder et al. 1998), but this enzyme is not classified as a cold-adapted trypsin (Outzen et al. 1996). Trypsin sequences and structures have been compared to look for evolutionary relationships (Roach et al. 1997; Rypniewski et al. 1994), and comparisons have been carried out to look for structure–function relationships for cold-adapted trypsins (Genicot et al. 1996; Heimstad et al. 1995; Smalås et al. 1994).

There seems to be a general difference between trypsins originating from cold-adapted fish species and higher vertebrate trypsins in terms of activity and stability. Trypsins from Atlantic cod (Ásgeirsson et al. 1989), Greenland cod (Simpson and Haard 1984a,b), Atlantic salmon (Outzen et al. 1996), and Antarctic fish (Genicot et al. 1988), for example, are all shown to be more catalytically efficient at all temperatures studied, but also in general much less stable at high temperatures, compared to mammalian counterparts. The enhanced catalytic efficiency (k_{cat}/K_m) is mostly the result of a lower K_m, indicating a better substrate affinity, which results in a 4- to 35-fold-higher efficiency for the cold-adapted fish trypsins. The fish and other trypsins of marine origin are also found to be unstable at low pH, e.g., trypsin from capelin (Hjelmeland and Raa 1982), Atlantic cod (Ásgeirsson et al. 1989), Greenland cod (Simpson and Haard 1984b), trout and Antarctic fish (Genicot et al. 1988), Atlantic salmon (Outzen et al. 1996), carp (Cohen et al. 1981a,b), mackerel (Kalác 1976), starfish (Kozlovskaya and Elyakova 1974; Winter and Neurath 1970), anchovy (Martínez et al. 1988), krill (Osnes and Mohr 1985), catfish (Yoshinaka et al. 1983a), and chum salmon (Uchida et al. 1984). In contrast, the cationic bovine trypsin (BT-C) is most stable at pH 3 (Simpson and Haard 1984b). The nonpsychrophilic cationic salmon trypsin (CST) is also stable at low pH, whereas the cold-adapted anionic counterpart is rapidly denaturated below pH 5 (Outzen et al. 1996). Whether the reduced stability at acidic pH for the psychrophilic trypsins is a consequence of cold adaptation, however, is still not known, but is likely attributable to structural characteristics.

Reviews have recently been presented on cold-adapted enzymes in general (Arpigny et al. 1994; Feller and Gerday 1997; Feller et al. 1996, 1997; Gerday et al. 1997; Marshall 1997) and on serine proteinases in particular (Kristjansson et al. 1997), but only a limited number of sequences were included. Pairwise comparison has been carried out for Antarctic fish and bovine trypsin (Genicot et al. 1988, 1996), Atlantic salmon and bovine trypsin (Heimstad 1996; Heimstad et al. 1995; Outzen et al. 1996; Schrøder et al. 1998; Smalås et al. 1994), Atlantic cod and bovine trypsin (Ásgeirsson et al. 1989), Greenland cod and bovine trypsin (Simpson and Haard 1984a,b), and mackerel and bovine trypsin (Kalác 1976). These comparisons are in some cases based on amino acid composition and in some cases on sequence identities, but probably because of the small number of sequences included have not been able to give general "rules" for the molecular basis of cold adaptation of enzymes and therefore of the psychrophilic nature of the organisms. In this study we are summarizing features at the level of amino acid sequence that are common and unique for the cold-active trypsins, and in this we try to rationalize the physical observations of the high catalytic efficiency and instability at low pH and high temperatures.

Materials and methods

Selection of amino acid sequences

A total of 74 trypsin sequences were located from a search in the EMBL and SWISSPROT protein data banks (February 1998). This number was reduced to 27 by considering only vertebrate trypsins and excluding isoforms with 99% sequence identity (Table 1). The trypsins originate from diverse species, ranging from the primitive sea lamprey to man, and have been subjected to highly different levels of characterization in the literature. For some, the amino acid sequence is almost the only information available, whereas others are thoroughly characterized in terms of activity, stability, and three-dimensional structure. The aim of the present study is to correlate residue determinants to physical behavior, but for the sake of comparison we have chosen to include as many eukaryote trypsins as possible, also those for which the enzyme is poorly characterized. A large selection may allow a better discrimination between features common to the cold-adapted enzymes and those that solely originate from close evolutionary relations. Among the higher vertebrate trypsins, the enzymes from pig, cattle, rat, and man are all well characterized (Bartunik et al. 1989; Craik et al. 1984; Gaboriaud et al. 1996; Kimland et al. 1989; Voytek and Gjessing 1971), whereas the others are much less extensively described in the literature. Trypsins from salmon, Antarctic fish, and cod (AST-II, AST, AFT, CT-I, CT-X) are well characterized in terms of activity and stability (Ásgeirsson et al. 1989; Genicot et al. 1988; Gudmundsdóttir et al. 1993; Male et al. 1995; Outzen et al. 1996; Raae and Walther 1989; Simpson and Haard 1984a,b) and all are, with the exception of cationic salmon trypsin (CST), found to be cold active as described in the Introduction. Pufferfish trypsin (PFT) is less well studied, but is included in the group referred to as cold-adapted, "cold fish," or "cold" trypsins. The cationic salmon trypsin, which shows no cold activity, is grouped with trypsins from sea lamprey and dogfish, and referred to as "other fish trypsins." The 17 other sequences from pig, cow, rat, dog, man, mouse, chicken, and frog are referred to as higher vertebrate trypsins.

Table 1. Origin, abbreviation, and references for the 27 trypsin sequences used in this study

No.	Data bank abbreviation	Short abbreviation	Organism	Species name	Reference
Higher vertebrate trypsins					
1	Tryp_Pig	PT	Pig	*Sus scrofa*	Hermodson et al. 1973
2	Try1_Bovin	BT-C	Bovine	*Bos taurus*	Mikes et al. 1966
3	Try1_Rat	RT-1	Rat	*Rattus norvegicus*	MacDonald et al. 1982
4	Try2_Rat	RT-2	Rat	*Rattus norvegicus*	MacDonald et al. 1982
5	Try3_Rat	RT-3	Rat	*Rattus norvegicus*	Fletcher et al. 1987
6	Try4_Rat	RT-4	Rat	*Rattus norvegicus*	Lütcke et al. 1989
7	Trya_Rat	RT-5A	Rat	*Rattus norvegicus*	Kang et al. 1992
8	Try1_Canfa	DT-I	Dog	*Canis familiaris*	Pinsky et al. 1985
9	Try2_Canfa	DT-II	Dog	*Canis familiaris*	Pinsky et al. 1985
10	Try1_Human	HT-I	Human	*Homo sapiens*	Emi et al. 1986
11	Try2_Human	HT-II	Human	*Homo sapiens*	Emi et al. 1986
12	Try3_Human	HT-III	Human	*Homo sapiens*	Tani et al. 1990
13	Try2_Bovin	BT-A	Bovine	*Bos taurus*	Le Huerou et al. 1990
14	Tryp_Mouse	MT	Mouse	*Mus musculus*	Stevenson et al. 1986
15	Try1_Chick	CHT-I	Chicken	*Gallus gallus*	Wang et al. 1995
16	Try2_Chick	CHT-II	Chicken	*Gallus gallus*	Wang et al. 1995
17	Try1_Xenla	FT	Frog	*Xenopus laevis*	Shi and Brown 1990
Other fish					
18	Try3_Salsa	CST	Atlantic salmon	*Salmo salar*	Male et al. 1995
19	AF011352	LT-A	Sea lamprey	*Petromyzon marinus*	Roach et al. 1997
20	AF011900	LT-B	Sea lamprey	*Petromyzon marinus*	Roach et al. 1997
21	Tryp_Squac	DFT	Dogfish	*Squalus acanthias*	Titani et al. 1975
Cold-adapted fish					
22	Try2_Salsa	AST-II	Atlantic salmon	*Salmo salar*	Male et al. 1995
23	Try1_Salsa	AST	Atlantic salmon	*Salmo salar*	Male et al. 1995
24	Tryp_Parma	AFT	Antarctic fish	*Paranotothenia magellanica*	Genicot et al. 1988
25	Try1_Gadmo	CT-I	Atlantic cod	*Gadus morhua*	Gudmundsdóttir et al. 1993
26	Tryx_Gadmo	CT-X	Atlantic cod	*Gadus morhua*	Gudmundsdóttir et al. 1993
27	TR25747	PFT	Pufferfish	*Fugu rubripes*	Roach et al. 1997

Sequence analysis

The 27 sequences were aligned by the Hein method (DNASTAR Inc. 1992) and the hydrophobicity indices were calculated using the scale of Kyte and Doolittle (1982) with a window size of 5 residues. Definition of internal and external residues was taken from our earlier work (Schrøder et al. 1998) in which a criterion of water-accessible surface $\leq 10 Å^2$ was set for internal residues. The mean hydrophobic indices for core, surface, and all residues were calculated by averaging the total values for the respective trypsins. Criteria for considering residues as possible determinants for higher vertebrate or cold-adapted trypsins were conservation in at least 9 of the 17 higher vertebrate trypsins and in 4 of the 6 cold fish trypsin sequences. The description of the trypsin regions, i.e., the detailed division into loops and regular secondary structure elements, is based on the existing X-ray structures (PDB entries 3PTB, 1TLD, 2TBS, 1BIT, 1A0J) and the Kabsch and Sander (1983) definition. Because the mean displacement among the trypsin structures already reported is small (Smalås et al. 1994), the definitions are likely to be valid for all the included trypsins with unknown three-dimensional structures. The structural elements are named according to their structural or functional task if possible (Fig. 1). Loops named as the Nβ1–Nβ2 loop, for example, indicate a location between the Nβ1 and Nβ2 β-strand (Fig. 1). The sequence numbering system is adopted from chymotrypsinogen A (Hartley and Kauffman 1966).

Results

Sequence alignment and identity

The sequence comparison includes 27 trypsin sequences (Table 1) divided into three groups: the higher vertebrate trypsins, the cold-adapted fish trypsins (AST-II, AST, AFT, CT-I, CT-X, PFT), and the other fish trypsins (CST, LT-A, LT-B, DFT). For simplicity, the enzymes that do not show features of cold adaptation are sometimes referred to as "warm" trypsins. All sequences studied are believed to have the typical trypsin fold, based on the conservation of the 3 catalytic residues Asp102, His57, and Ser195, Asp189 at the bottom of the specificity pocket, and Asp194, which forms an internal salt bridge to the N-terminal amino group of Ile16. Gly216 and Gly226 at the entrance of the specificity pocket, and Ser214, which is known to interact with Asp102, are also present for all the selected sequences (Fig. 1). Tyr172, which is suggested to be a trypsin determinant (Hedstrom et al. 1994b), is also present in all the trypsins except for HT-III. The total number of residues in the 27 selected eukaryote trypsins are 222 (8 sequences), 223 (11

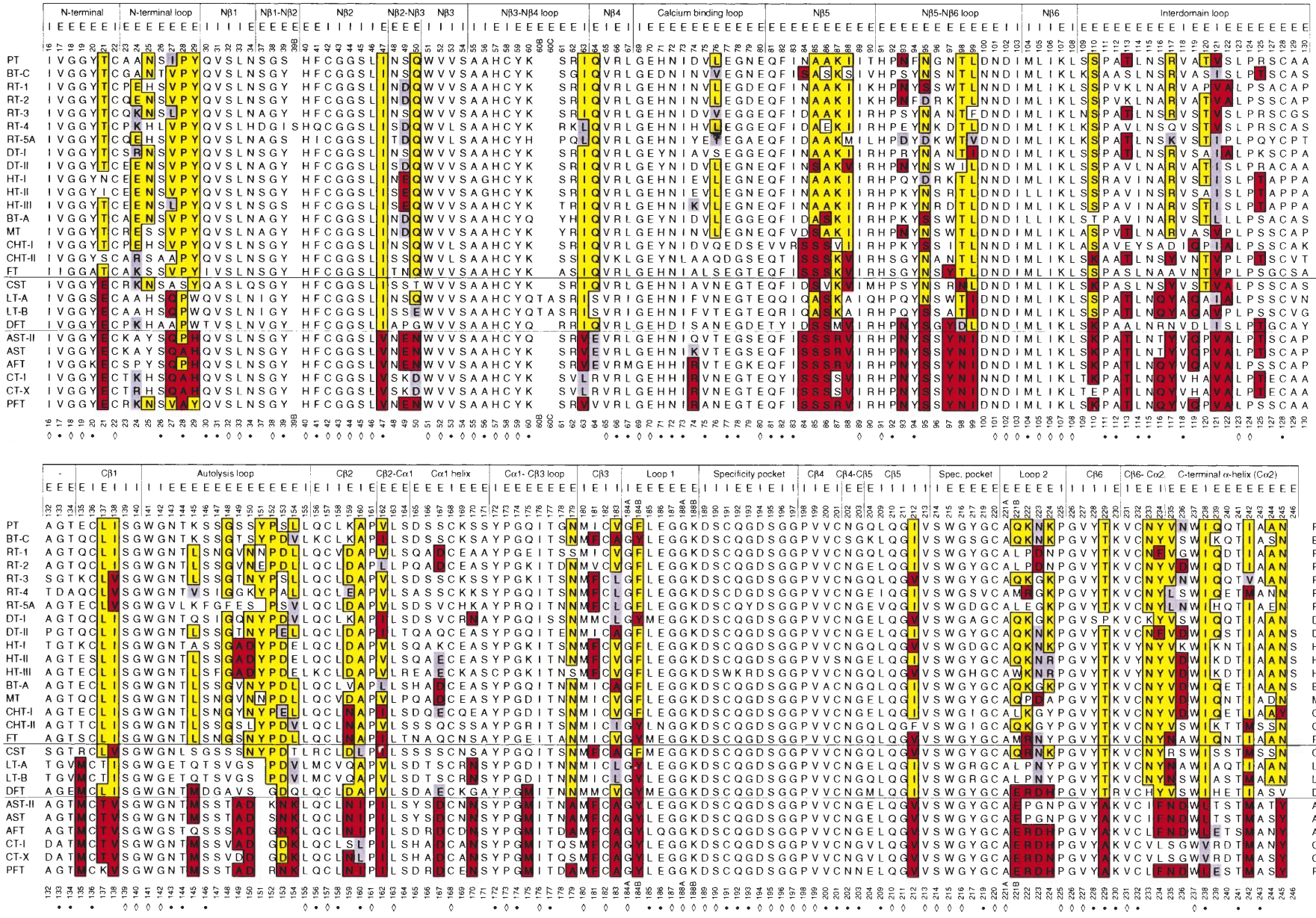

Fig. 1. Amino acid sequence alignment of the 27 vertebrate trypsins named according to Table 1. Strictly conserved residues are indicated with (✧), while those found in more than 75% of the sequences (≥20) are marked with a *dot*. Residue determinants of the cold fish trypsin sequences (in ≥4) or the higher vertebrate trypsins (in ≥9) are color coded in *red* or *yellow*, respectively. Residues in *lilac* denote homology with the remaining residues, in terms of polarity, charge, and size

	PT	BT-C	RT-1	RT-2	RT-3	RT-4	RT-5A	DT-1	DT-II	HT-I	HT-II	HT-III	BT-A	MT	CHT-I	CHT-II	FT	CST	LT-A	LT-B	DFT	AST-II	AST	AFT	CT-I	CT-X	PFT	
PT		82	78	81	86	74	71	79	82	78	79	77	79	78	73	72	74	73	72	71	67	66	66	63	62	63	66	PT
BT-C			73	74	79	71	68	79	76	75	74	71	74	73	72	74	70	73	69	68	65	65	65	61	62	62	65	BT-C
RT-1				90	75	69	68	74	81	78	77	75	80	87	77	67	74	68	69	69	67	67	67	64	65	65	67	RT-1
RT-2					77	69	69	75	82	80	79	76	80	87	75	66	73	68	68	67	65	67	67	65	63	64	68	RT-2
RT-3						77	73	80	79	78	76	75	79	77	71	72	72	76	69	69	64	66	66	64	63	63	66	RT-3
RT-4							72	70	71	68	68	68	71	70	66	65	66	68	63	64	62	60	60	58	60	60	60	RT-4
RT-5A								70	69	67	68	67	69	69	65	63	64	63	65	63	61	59	59	56	55	56	58	RT-5A
DT-I									76	75	75	71	75	73	72	72	72	71	71	69	64	66	66	65	62	63	66	DT-I
DT-II										80	82	78	84	83	74	69	75	72	67	66	65	67	67	64	64	65	68	DT-II
HT-I											89	85	78	75	70	68	70	68	66	66	63	63	63	60	62	61	64	HT-I
HT-II												87	80	78	70	67	70	68	67	66	65	65	65	62	62	63	65	HT-II
HT-III													76	74	67	65	70	65	63	63	63	61	62	58	60	60	61	HT-III
BT-A														81	75	68	73	69	68	68	66	65	65	62	61	62	65	BT-A
MT															75	67	75	68	69	69	67	66	66	64	62	63	67	MT
CHT-I																70	70	63	66	65	64	65	65	62	60	60	65	CHT-I
CHT-II																	74	71	64	67	64	64	64	60	61	61	63	CHT-II
FT																		72	67	67	66	65	65	63	63	64	67	FT
CST																			65	67	65	68	69	65	67	67	70	CST
LT-A																				93	64	67	66	66	64	64	65	LT-A
LT-B																					65	69	68	67	65	66	66	LT-B
DFT																						65	65	63	66	65	66	DFT
AST-II																							97	87	78	79	87	AST-II
AST																								87	79	80	88	AST
AFT																									79	79	88	AFT
CT-I																										96	80	CT-I
CT-X																											80	CT-X
PFT																												PFT

Fig. 2. Percent sequence identity among the 27 trypsin amino acid sequences. Sequence similarity: *red*, ≤64%; *green*, 65%–67%; *black*, 68%–79%; *blue*, ≥80%. The comparisons between cold fish, other fish, and higher vertebrate sequences are *framed*

sequences) or 224 (8 sequences). Deletions and insertions among them are confined to three loops and the C-terminal helix. The Nβ1–Nβ2 loop of RT-4 is 1 residue longer than for the other trypsins (residue 39B), and the Nβ3–Nβ4 loop of LT-A and LT-B includes 2 extra residues (60B, 60C), while there is a 1-residue deletion (Tyr151) in the "autolysis loop" of the six cold-adapted fish trypsins, LT-A, LT-B, DFT, and RT-5A. At the C-terminal the three human trypsins (HT-I, II, III), DT-II, and BT-A have an additional serine residue, Ser246. Based on the position of the cysteine residues (Fig. 1), six disulfide bridges seem to be conserved (22–157, 42–58, 128–232, 136–201, 168–182, 191–220) among all the trypsins with the exception of the three from human sources. Judged from the crystal structure of HT-I (PDB entry 1TRN) (Gaboriaud et al. 1996), lack of the 128–232 disulfide bridge seems to be compensated for by stacking interactions between Pro128 and Tyr232, which is possible for the HT-II and HT-III as well (Fig. 1).

Of the maximum 224 residues, 73 (32.6%) are strictly conserved among the 27 sequences in this study, and 136 residues (60.7%) are conserved in more than 20 sequences (Fig. 1). The homology in terms of percentage sequence identity, shown in Fig. 2, groups the trypsins in relatively distinct classes. The mammalian trypsins, with the exceptions of RT-4 and RT-5A, match each other with approximately 65%–90% identity, while frog trypsin (FT) and the two chicken trypsins (CHT-I, -II) are about equally distant from the mammalian and fish trypsins. The fish trypsins in relation to the higher vertebrate counterparts possess identities in the range 55%–76%. Among the fish trypsins, the six cold enzymes clearly form a uniform group with high sequence identity (78%–97%), while other fish trypsins (CST, LT-A, LT-B, and DFT) are equally distant from all other trypsins, with 61%–76% identity compared to the higher vertebrates and 63%–70% compared to the cold fish trypsins.

Phylogenetic relations

The phylogenetic tree (Fig. 3) is consistent with continuous evolution of trypsin starting from a single ancestral gene. The four branches succeeding the second diverging nodes group the trypsins into mammalian, "other" higher vertebrates, and anionic fish trypsins, while the cationic salmon trypsin (CST) singularly forms the fourth branch. The six cold fish trypsins separated at a relatively late stage in evolution, while CST diverged from the anionic fish trypsins at an early stage. Atlantic salmon seems to have kept an "early form" of trypsin (the CST), while anionic forms (AST, AST-II) have evolved.

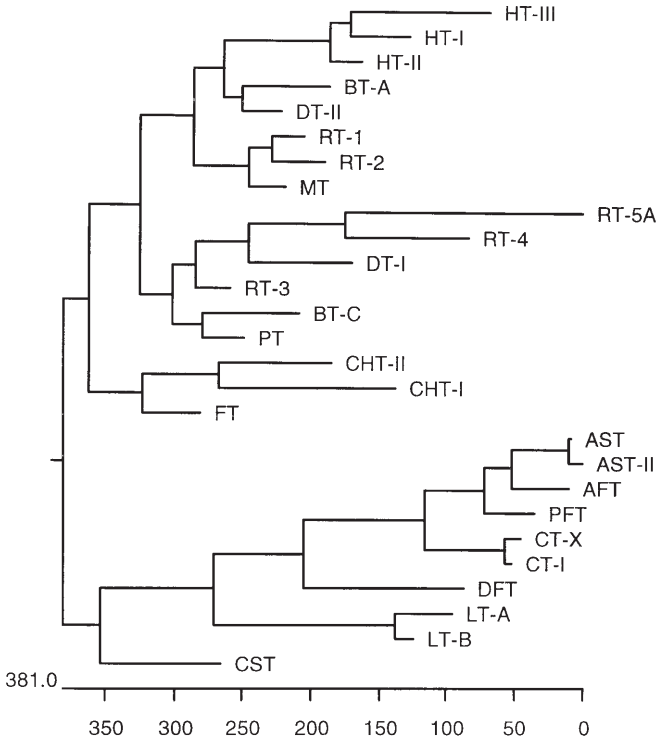

Fig. 3. Phylogenetic tree for the 27 trypsin sequences. The cladogram was constructed using the Jotun–Hein method (DNASTAR 1992), and branch distances represent the length of sequence divergence

The amino acid sequences of the cationic (CST) compared to the anionic salmon trypsin (AST), and RT-1 compared to RT-5A, are much more different than any of the other multiple forms within a single species. This, along with the unique difference in activity and stability between CST and AST, makes CST an interesting "outlier" among the fish trypsins.

Amino acid composition

In Table 2, the average amino acid composition of each of the three groups of trypsins is shown: the higher vertebrate, the cold fish, and the other fish trypsins. The most pronounced feature of the compositions is the high number of charged (RKHDE) and polar residues (QNCSTYRKHDEG) and the low number of hydrophobic residues (AILFWVPM) in the cold fish trypsins compared to the other vertebrate trypsins. All the cold-adapted fish trypsin sequences are anionic, reflected through a higher number of Asp and Glu and a lower number of Lys, but actually including more Arg than for the remaining trypsins.

The cold-adapted fish trypsins contain fewer Gly, Pro, and Gln, whereas there are more His, Ser, and Tyr than in the remaining trypsins. The total number of hydrophobic residues is somewhat lower for the cold-adapted fish trypsins, but even more pronounced is the lower number of isoleucines and leucines and the higher number of valines. On average there are about 5 more isoleucines in the warm trypsins compared to the cold fish trypsins, resulting in a low (Ile + Leu)/(Ile + Leu + Val) ratio for the six cold fish trypsins compared to the other vertebrate trypsins (Table 2). The other fish trypsins resemble the higher vertebrate trypsins in terms of high Ile content and high (Ile + Leu)/(Ile + Leu + Val) ratio. A high number of the flexible methionine residues, as in the cold-adapted trypsins, is also found in the other fish trypsins. For the anionic cold fish trypsins most of the additional methionines are confined to the C-terminal domain, whereas they are more evenly distributed for the remaining fish trypsins. The number of serine residues varies from 16 in RT-5A to 39 in CST, and is on average significantly higher for the cationic than for the anionic forms. The fish trypsins are however not different from the other anionic trypsins in this respect. The mean number of histidines is notably higher for the anionic trypsins (5.7) than for the cationic isoforms (3.8), and a high histidine content is particularly evident for the cold fish trypsins (average, 6.8).

Hydrophobicity

The mean hydrophobicity indices calculated for each of the 27 trypsin sequences are shown in Fig. 4. Classification of residues as external and internal are based on water-accessible surfaces as found for trypsins with known three-dimensional structure (Schrøder et al. 1998). The internal residues of the cold fish trypsins are, with the exceptions of the two cod trypsins (CT-I, X), significantly less hydrophobic than the average of the remaining trypsins. The other fish trypsins are in this sense similar to the higher vertebrate trypsins. For surface residues there is a more pronounced difference with increased hydrophilicity of the cold-adapted fish trypsins, as they are all among the 11 trypsins with lowest mean hydrophobicity index. CST is in this sense more similar to the cold fish trypsins, whereas LT-A, LT-B, and DFT resemble the higher vertebrate trypsins. More hydrophilic surfaces compared to the mammalian counterparts are also found in the case of cold-adapted bacterial subtilisin (Davail et al. 1994) and cold-adapted bacterial α-amylase (Feller et al. 1992). Enhanced overall hydrophilicity, as observed for the cold fish trypsins compared to the warm trypsins, is also found for cod elastase compared to the porcine equivalent (Gilberg and Øverbø 1990). The three human trypsins (HT-I, -II, -III) and RT-3 seem to be the exceptions among the higher vertebrate trypsins, as their mean hydrophobicity indices for external and all residues are low and in the same range as for the cold fish trypsins. CST, however, seems to have retained the stability and activity features of the warm trypsins even with surface and overall hydrophilicity in the same range as the cold fish enzymes.

Charge and arginine residues

Even though all the low-temperature-adapted trypsins are anionic, there does not seem to be any direct connection between phylogenetic evolution and the isoelectric nature

Table 2. Mean amino acid composition (with SD inparentheses) of all sequences (27 sequences), higher vertebrates (17 sequences), the cold fish (6 sequences) and other fish (4 sequences) trypsin sequences

	All	Higher vertebrates	Cold fish	Other fish
pI (calculated)	6,21 (1,3)	6,33 (1,6)	6,01 (0,4)	5,99 (1,5)
Net charge	−3,2 (5,3)	−2,5 (6,0)	−4,5 (1,5)	−4,3 (6,2)
Classes				
Charged	35,2 (6,7)	34,5 (7,0)	38,7 (4,8)	32,8 (7,1)
Acidic	16,7 (4,4)	16,4 (5,0)	18,2 (2,0)	15,8 (4,9)
Basic	13,4 (3,3)	13,8 (3,6)	13,7 (2,2)	11,5 (3,4)
Polar	139,5 (4,3)	138,4 (4,4)	142,0 (1,1)	140,3 (6,1)
Hydrophobic	83,5 (4,5)	84,9 (4,4)	80,0 (1,1)	83,0 (6,1)
Aromatic	18,3 (1,6)	18,0 (1,6)	19,5 (1,4)	17,8 (1,0)
Ala	14,8 (2,1)	14,9 (2,3)	14,3 (1,2)	15,0 (2,7)
Cys	11,7 (0,9)	11,5 (1,1)	12,0 (0,0)	12,0 (0,0)
Asp	8,9 (2,5)	8,6 (2,5)	9,5 (2,1)	8,8 (3,5)
Glu	7,8 (2,7)	7,7 (3,2)	8,7 (1,0)	7,0 (1,6)
Phe	3,3 (1,1)	3,5 (1,1)	3,5 (0,8)	2,3 (1,3)
Gly	23,8 (1,4)	24,0 (1,2)	**22,7** (0,8)	24,5 (2,4)
His	5,1 (1,9)	4,4 (1,4)	**6,8** (2,2)	5,5 (1,7)
Ile	14,0 (2,8)	15,2 (1,8)	**10,0** (0,9)	14,8 (3,0)
Lys	8,4 (2,7)	9,4 (2,7)	7,5 (0,8)	5,3 (1,9)
Leu	15,6 (2,3)	16,8 (2,0)	**13,8** (1,0)	13,3 (1,5)
Met	3,7 (2,0)	2,4 (0,9)	**6,0** (0,0)	5,8 (2,2)
Asn	14,8 (2,8)	15,3 (2,8)	14,3 (3,0)	13,3 (2,8)
Pro	9,9 (1,6)	10,3 (1,6)	**8,8** (1,7)	9,8 (1,3)
Gln	10,3 (2,4)	10,9 (1,8)	8,2 (1,2)	11,0 (4,4)
Arg	5,1 (1,9)	4,4 (1,7)	**6,2** (1,9)	6,3 (1,7)
Ser	23,9 (5,4)	23,1 (5,4)	**24,7** (1,0)	26,3 (9,3)
Thr	9,2 (2,0)	8,9 (1,7)	9,5 (2,1)	10,0 (2,9)
Val	17,9 (1,8)	17,5 (1,6)	**19,5** (2,0)	17,3 (1,0)
Trp	4,4 (0,6)	4,4 (0,6)	4,0 (0,0)	5,0 (0,8)
Tyr	10,6 (1,7)	10,1 (1,8)	**12,0** (0,9)	10,5 (1,3)
SUM	223,0 (0,8)	223,3 (0,6)	222,0 (0,0)	223,3 (1,0)
Arg/(Arg + Lys)	0,38 (0,12)	0,32 (0,09)	**0,44** (0,08)	0,55 (0,05)
(Ile + Leu)/(Ile + Leu + Val)	0,62 (0,05)	0,65 (0,03)	**0,55** (0,02)	0,62 (0,01)

Residues included in the various class definitions are net charge (KR-DE), charged (RKHDE), acidic (DE), basic (KR), polar (QNCSTYRKHDEG), hydrophobic (AILFWVPM), and aromatic (FWY). The ratios Arg/(Arg + Lys) and (Ile + Leu)/(Ile + Leu + Val) are also shown. Numbers in **bold** indicate interesting deviations for the cold fish trypsins

of the proteins. The anionic trypsins contain at average 2–3 more aspartic and glutamic acids and 3–4 fewer lysine residues. The number of arginines, however, is not much different between the cationic and the anionic trypsins, and is in fact higher for the anionic cold fish trypsins (average, 6.2) compared to all the cationic trypsins (average, 5.4) and the remaining anionic forms (average, 4.4). Arginine is believed to stabilize the protein structure by its ability to form salt bridges and a large number of hydrogen bonds (Mrabet et al. 1992). For other cold-adapted enzymes, a low Arg/(Arg + Lys) ratio has been observed, e.g., in Antarctic fish elastase (Aittaleb et al. 1997) and in α-amylase (Feller et al. 1994), in which it is thought to destabilize the structure (Arpigny et al. 1994; Feller et al. 1996). The average Arg/(Arg + Lys) ratio, however, is higher for the cold fish trypsins (see Table 2), and thus in the case of trypsin cannot explain the observed high-temperature instability.

Apart from the arginine content, there are no obvious differences in overall numbers of charged residues between the anionic fish trypsins and the other anionic forms. Still, based on calculated isoelectric points (pI) (Fig. 5), the cold-adapted fish trypsins cluster in a relatively narrow range around pI 6, while most of the remaining anionic forms have lower pI values. In general, the charge of the N-terminal domain is very similar for all trypsins, regardless of overall anionic or cationic character. The main charge differences between the cationic and anionic trypsins thus arise from differences of the C-terminal domain. The difference in net charge between the N- and C-terminal domains for some of the anionic forms is as high as 10 point charges, resulting in highly polar protein molecules. The charge differences between the two molecular halves are on average similar for both cold-adapted fish trypsins and the other anionic trypsins, with a few exceptions. However, at the level of amino acid sequence, some notable differences are evident for the charged residues among the anionic forms (Table 3).

Uniquely conserved residues in higher vertebrate and cold fish trypsins

Tables 3 and 4 give a survey of distinct amino acid conservations among the cold-adapted fish trypsins and other higher vertebrate trypsins, extracted from the alignment in Fig. 1. There are 56 residues that are unique for the cold fish

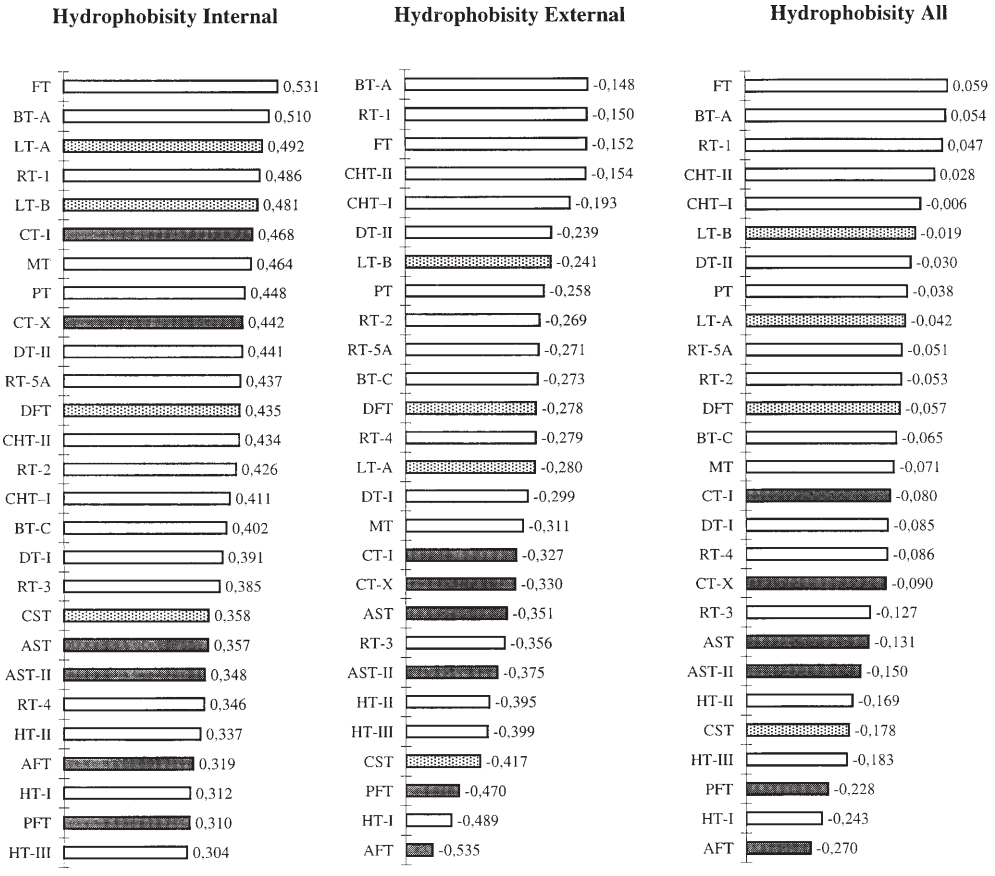

Fig. 4. Histogram of the mean hydrophobicity indices for internal, external, and all residues for each of the 27 trypsins included in the study. The bars representing the cold fish and other fish trypsins are marked in *gray* and with *dots*, respectively

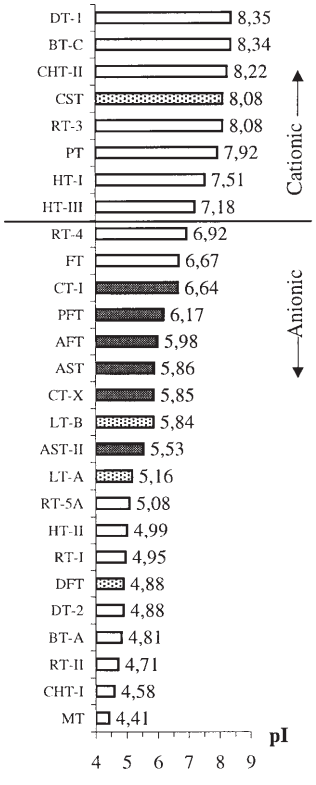

Fig. 5. The calculated isoelectric points (pI) of the respective amino acid sequences, sorted in increasing order. *Gray bars*, numbers for cold fish trypsins; *dotted bars*, other fish trypsins

trypsins (conserved in 4 or more of the 6 sequences), and 49 residues that are uniquely conserved among the higher vertebrate trypsins (found in at least 9 of the 17 sequences).

Hydrophobicity and core residues

It is generally believed that isoleucine, through its Cβ-branching and size, packs and stabilizes the protein structure more efficiently than the other aliphatic residues (Britton et al. 1995), and also through increased burial of hydrophobic surface area (Malakauskas and Mayo 1998). For some thermophilic enzymes, increased isoleucine content has been observed (Britton et al. 1995; Haney et al. 1997; Tanner et al. 1996; Yip et al. 1995), and enhanced stability when increasing the number of isoleucines by mutation (Malakauskas and Mayo 1998) is also reported. In the case of the cold-adapted fish trypsins the low number of isoleucines (average, 10.0) compared to the higher vertebrate trypsins (average, 15.2), may be one factor explaining the reduced stability of the former. Seven isoleucines are unique to the warm trypsins (47, 63, 88, 138, 212, 238, 242), while 3 are found in the cold fish trypsins only (99, 160, 162) (see Fig. 1 and Table 3). Four of these in the warm trypsins and 1 in the cold fish trypsins are internal, indicating less optimized core packing of the latter trypsins. The four fish trypsins that do not possess "cold activity" resemble the higher vertebrates in this respect. Four other internal residues (121, 183, 229, 234; see Table 3) are replaced by

Table 3. Unique amino acid determinants for cold fish or higher vertebrate trypsins

No.	Higher vertebrates	Cold fish
Isoleucines in higher vertebrates:		
47	**Ile**	**Val**
63	Ile	Val
88	Ile	**Val**
138	Ile	**Val**
212	Ile	**Val**
238	**Ile**	Leu
242	Ile	**Met**
Isoleucines in cold fish trypsins only:		
99	Leu	**Ile**
160	**Ala**	Ile
162	Val	**Ile**
Internal residue differences:		
121	*(Val/Ile)	**Val**
183	Val	**Ala**
229	Thr	**Ala**
234	Tyr	Phe
Hydrophobic external residues in higher vertebrates:		
27	Val	Gln
76	Leu	*
85	Ala	**Ser**
86	Ala	**Ser**
137	**Leu**	Thr
154	Leu	**Lys**
235	Val	Asn
Methionines in cold fish trypsins:		
135	*	**Met**
145	Leu	Met
175	*	**Met**
242	*Ile*	***Met***
Charged residues in cold fish trypsins:		
21	Thr	**Glu**
29	**Tyr**	His
49	* (Asp/Glu)	Glu
74	*	Arg
87	Lys	Arg
110	Ser	Lys
150	Asn	**Asp**
154	*Leu*	***Lys***
167	*	**Asp**
221B	Gln	**Glu**
222	Lys	Arg
223	*	Asp
224	Lys	His
236	*	Asp
Charged residues in higher vertebrates:		
24	Glu	* (Lys/Arg)
117	Arg	**Tyr**
153	Asp	Asn
159	Asp	Asn

Residues are shown in **bold** when conserved in all sequences of either of the two classes; *italic* indicates that the residue has been listed earlier in the table; and * means any amino acid

smaller amino acids in the cold fish trypsins, compared to the higher vertebrate counterparts, in which the substitutions at position 121 and 183 result in reduced hydrophobicity of the cold trypsins. At an additional seven positions, the higher vertebrate trypsins include hydrophobic residues, whereas the cold fish trypsins have polar ones and for residue 229 and 234 the opposite is the case.

Methionine residues

A high methionine content is not uniquely found for the cold fish trypsins, but is in fact common for all trypsins of marine origin (see Table 2). Two methionines, Met104 and Met180, are conserved in virtually all the trypsins, while the four additional methionines in the cold-adapted trypsins, Met135, Met145 (not AFT), Met175, and Met242, are well conserved within this class (see Fig. 1, Table 3). A number of other digestive proteinases of marine origin have also been reported with a high methionine content. Examples are carp trypsin (Cohen et al. 1981a), catfish trypsin (Yoshinaka et al. 1983b), anchovy trypsin (Martínez et al. 1988), krill trypsin (Osnes and Mohr 1985), anionic chum salmon trypsin (Uchida et al. 1984), tunicate trypsin (Roach et al. 1997), cod elastase (Gudmundsdóttir et al. 1996), salmon elastase (Berglund et al. 1995b), Antarctic fish elastase (Aittaleb et al. 1997), cod chymotrypsin (Gudmundsdóttir et al. 1994), tuna fish pepsin (Tanji et al. 1996), and cod pepsin (Karlsen et al. 1998).

Conserved charged residues

The cold fish trypsins are all distinguishable by a high number of charged residues, and their charge distribution is also partly conserved. Fourteen charged residues (DEHKR) are invariant in the cold fish sequences, as listed in Table 3. Among all the higher vertebrate trypsins only 4 charged residues are conserved, probably because of the mixture of cationic and anionic forms. His29 and His224 are both unique for the cold fish trypsins, and because acid-induced denaturation of many proteins is found to be caused by protonation of histidines (Barrick et al. 1994), these and the overall high number of histidine residues in the six cold trypsins (Tables 2, 3) may be a main reason for their reduced stability at low pH.

Substitutions in loop regions

A large portion of the trypsin surface area is formed by loops of various lengths (see Fig. 1), and in common with all classes of homologous proteins, the degree of amino acid conservation is lowest for the loop regions. As reflected by the increased surface hydrophilicity of the cold fish trypsins, the amino acid substitutions between warm and cold trypsins for loops are generally toward more hydrophilic residues. However, because a particular type of side chain is less critical for loop regions, the conservation within the two groups of trypsins is rather of properties than of identities of the residues. Nevertheless (as shown in Table 4), some amino acids in loop regions are conserved to a high degree within both trypsin classes, but are also distinctly different between them.

Table 4. Possible residue determinants in the higher vertebrate or cold fish trypsins, located in distinct structural regions of the molecule

No.	Higher vertebrates	Cold fish	Comments
Connecting loops and C-terminal α-helix			
N-terminal and N-terminal loop			
21	*Thr*	***Glu***	The loop in the higher vertebrate trypsins could be stabilized by the non-pH-dependent Pro28-Tyr29 motif.
24	*Glu*	**(Lys/Arg)*	
25	Asn	*	
27	*Val*	*Gln*	
28	**Pro**	Ala	
29	***Tyr***	*His*	
Interdomain loop			
110	*Ser*	*Lys*	The loop connects the two β-barrels in the trypsin structure and is located along the molecule. Arg117 is found to make a salt bridge with Glu77 in a molecular dynamic simulation of BT-C (Brandsdal, Heimstad, Sylte, and Smalås, unpublished), an interaction that is not possible in the cold fish trypsins.
113	*	**Thr**	
116	*	Gln	
117	*Arg*	***Tyr***	
119	*	Gln	
120	Thr	*	
121	**(Val/Ile)*	***Val***	
122	*	Ala	
125	*	Thr	
Cβ6-Cα2-loop and the C-terminal α-helix			
233	Asn	*	Asp236 in the cold-active enzymes could stabilize the positively charged helical macro dipole. Asn245 in higher vertebrate and other fish trypsins is more stabilizing in terms of helical propensity than the Tyr245 in the cold fish trypsins. One of the cold fish conserved methionines, Met242, is also located in the helix.
234	*Tyr*	*Phe*	
235	*Val*	*Asn*	
236	*	*Asp*	
238	***Ile***	*Leu*	
239	Gln	*	
242	*Ile*	***Met***	
244	Ala	*	
245	Asn	**Tyr**	
Other loops and residues			
Nβ5			
84	*	**Ser**	The three serine residues (84,85,86) are all found in the cold fish sequences. The strand includes Lys87, which in the higher vertebrate trypsins and CST forms an ion pair to the carboxyl group of the C-terminal. This interaction is not observed in the crystal structures of anionic salmon trypsin (Smalås et al. 1994; Berglund et al. 1995a; Helland et al. 1998).
85	*Ala*	***Ser***	
86	*Ala*	***Ser***	
87	Lys	Arg	
88	*Ile*	***Val***	
Nβ5-Nβ6 loop			
93	*	**Asn**	A number of residues are conserved in all the cold fish sequences. The loop includes Asp102, which is a member of the catalytic triad.
95	Asn	**Ser**	
97	*	**Tyr**	
98	Thr	**Asn**	
99	*Leu*	***Ile***	
Autolysis loop			
145	*Leu*	*Met*	The autolysis loop is known to fold very differently in AST compared to BT-C, RT-II, and CST (Smalås et al. 1994; Schrøder et al. 1998). All the cold fish trypsins show high sequence homology in this loop with Met145, deletion of Tyr151, and no Pro152.
148	Gly	*	
149	*	Ala	
150	*Asn*	***Asp***	
151	Tyr	**Missing**	
152	Pro	*	
153	*Asp*	*Asn*	
154	*Leu*	***Lys***	
Loop 2			
221B	*Gln*	***Glu***	This loop is found to be important for the specificity when converting chymotrypsin to trypsin (Hedstrom et al. 1992, 1994a, 1994b). The higher vertebrate trypsins have charge +2 (222, 223) and the cold fish trypsins have −1 or 0 (221B, 222, 223, (224)).
222	*Lys*	*Arg*	
223	*	*Asp*	
224	*Lys*	*His*	
Other residue differences			
50	Gln	Asn	Gln/Asn50 is in an environment close to Lys107.
64	Gln	*	
170	*	**Asn**	
179	Asn	Ala	
181	*	**Phe**	
184B	Phe	**Tyr**	

Residues are shown in **bold** when conserved in all sequences; *italic* indicates residue has been listed earlier in table; *, any amino acid

All the higher vertebrate trypsins have Pro28 and Tyr29 in the N-terminal loop; the six cold trypsins have other residues in these positions (mostly Ala and His), but the most recent X-ray structure of AST does have Pro28 (Helland et al. 1998). The proline residue, succeeded by a large bulky residue, will probably reduce the flexibility of this loop, and could thus contribute significantly to increased overall stability of the warm trypsins as the cold trypsins lack Tyr29.

Residue 145, in the autolysis loop (residues 141–155), is known to be a major autolysis site in bovine trypsin (Fehlhammer and Bode 1975). Crystal structure analysis of anionic salmon and bovine trypsins (Smalås et al. 1994) showed that the largest structural difference between the two enzymes was confined to this loop. This so-called autolysis loop takes totally different directions in the two trypsins. From the sequence comparison, it is likely that the conformation of this loop is similar to that of bovine trypsin for all the higher vertebrate trypsins, while the cold trypsins probably are similar to anionic salmon trypsin in this respect. With the exception of CST, all fish trypsins have a one-residue deletion at position 151, a residue that is present as a tyrosine in the majority of the warm trypsins. The proline at position 152 in the higher vertebrate trypsins, as well as in CST, LT-A and LT-B, probably determines the conformation of the loop, while all cold trypsins have other residues at this position, e.g., glycine for three trypsins. Pro152 and Pro28 are also missing in the psychrophilic cod chymotrypsin whereas the prolines are conserved in the mammalian chymotrypsins (Gudmundsdóttir et al. 1994). The impact of the differences in amino acid sequence and conformation of the autolysis loop is not clear, but comparison of the crystal structures of anionic salmon and bovine trypsins complexed with the bovine pancreatic trypsin inhibitor (BPTI) showed that the loop is involved in inhibitor binding in different manners for the two enzymes (Helland et al. 1998). This result could imply that the loop is important for substrate binding and catalysis. It should also be mentioned that a basic residue is present at residue 145 only for three trypsins (BT-C, PT, and RT-5A), and is thus not an autolysis site for the majority of the trypsins.

Loop 2 (residues 221A–225) is distinguished with positively charged residues in most of the higher vertebrate trypsins, while both positively and negatively charged residues are present in the cold fish trypsins. The loop, which is an extension of the specificity pocket, has been found to be crucial for trypsin activity by Hedstrom et al. (1992, 1994a,b). It is therefore possible that the presumed differences in electrostatic potential of this region also impose differences in catalytic efficiency among trypsins.

β-Sheets and α-helices

Amino acids of β-strands are to a high degree conserved among all trypsins, indicating very similar folds. Of the 72 β-strand forming residues (Fig. 1), 75.0% are identical in 20 or more of the 27 sequences, and 44.4% are identical in all sequences. Differences in amino acid sequences of the two helices are, however, more evident and distinguishable between the cold fish and the warm trypsins. N-capping residues of the C-terminal helix (Cα2) slightly favor the cold fish trypsins in terms of stability, as the cold trypsin-conserved Asn235 and Asp236 could stabilize the positively charged helical macro-dipole (Fersht and Serrano 1993; Muñoz and Serrano 1995). At the C-cap of the helix, Asn245 is conserved in most of the warm trypsins, while the cold trypsin conserved Tyr245 is known to be a poor C-capping residue.

Discussion

From this study, many residue determinants have been found to be common and unique for the cold-adapted trypsins, but the main challenge has been to discriminate between common features that can be identified as cold-adaptation determinants and those that arise purely from close evolutionary relations. Trypsin is regarded as a highly suitable model system because the enzyme has been well studied in a diversity of species, and many amino acid sequences and crystal structures are available. The amounts of data reported for cold-adapted fish trypsins and mammalian trypsins are generous, but still more data on trypsins from ectothermic species that do not possess low-temperature activity are needed. The cationic salmon trypsin, which is well characterized and resembles the mammalian equivalents in three-dimensional structure, activity, and stability (Outzen et al. 1996; Schrøder et al. 1998), was therefore considered as an important member to study in this context. To be able to isolate single effects that might be responsible for the cold-adaptation features, we have only considered trypsins that were likely to possess the same or highly similar three-dimensional folds. This issue was also the main reason for omitting the prokaryote trypsins from this comparison. The present comparison and discussion are therefore based on the assumption that the classification of internal and external residues, secondary structure elements, etc. of trypsins with known three-dimensional structures also are valid for those where only the sequences are known.

Although the amino acid sequence isology is high among all the trypsins, there are notable differences in terms of both overall amino acid composition and specific residues at distinct positions. Both the cold and warm trypsins have about 50 amino acids that are unique and conserved within each class. The most notable overall difference between the warm and cold trypsins is the increased hydrophilicity of the latter. This change seems to be achieved by both a reduced hydrophobicity of internal residues and increased hydrophilicity of surface-exposed groups (see Fig. 4). The impact of increased hydrophilicity of groups accessible to solvent is not fully understood, but Davail et al. (1994) suggested that the resulting increased interactions with the surrounding water reduce the compactness of the external protein shell and thereby reduce the stability. However, increased hydrophilicity is also found for several high-temperature-stable

enzymes (Haney et al. 1997), and is suggested to be a factor for increased thermal stability (Jaenicke 1991).

The number of both charged and other polar amino acids is significantly higher in the cold trypsins. A large number of arginines and a high Arg/(Arg + Lys) ratio, as found in all the marine trypsins, are thought to stabilize a protein structure through providing a strong ability to form hydrogen bonds (Mrabet et al. 1992). Arginines are, however, typically external and thus are mostly involved in interactions with water, and unless involved in salt bridges their stabilizing effect is questionable. A thermophilic triosephosphate isomerase (TIM) was found to have a high Arg/(Arg + Lys) ratio, but the structural comparison with mesophilic TIM structures did not correlate this with increased number of hydrogen bonds and higher stability of the thermophilic enzyme (Delboni et al. 1995). One should also keep in mind that, at low temperature, polar interactions such as salt bridges, hydrogen bonds, and aromatic interactions must be more important than the hydrophobic interactions because electrostatic interactions are formed exothermically while hydrophobic interactions are formed endothermically (Jaenicke 1990). The generally higher number of charged residues, predominantly acidic groups, of the cold trypsins might be responsible for their reduced stability, but one cannot rule out that this is associated with adaptation to high salt concentrations for these extracellular fish enzymes.

The reduced hydrophobicity of internal groups is accomplished by an apparently systematic substitution to smaller and less hydrophobic residues in the cold-active trypsins. In particular, the reduced number of isoleucines and correspondingly increased number of valines (expressed by a low (Ile + Leu)/(Ile + Leu + Val) ratio) in the core is notable. The less densely packed core results in reduced van der Waals interactions and increased rotational and vibrational freedom of internal groups, and could be a main factor for the low temperature stability. Furthermore, the reduced hydrophobicity of buried residues also reduces the stability by the decreased positive free energy of hydration compared to the warm active enzymes.

The number of proline and glycine residues is not dramatically different between the warm and cold trypsins. The additional Pro–Tyr interaction in the N-terminal loop and the proline in the autolysis loop of the warm trypsins undoubtedly reduce the flexibility of these two loops and could also in turn be a factor that rigidifies and increases the overall stability. The different conformation of the autolysis loop between bovine (BT-C) and anionic salmon trypsin (AST) (Smalås et al. 1994) is, based on the amino acid sequence, probably a general difference between the cold- and warm-active trypsins. The impact of the different loop conformations is not fully understood, but the autolysis loop is found to be involved in inhibitor binding in different manners for BT-C and AST, as seen from the respective complexes with bovine pancreatic trypsin inhibitor, and could therefore affect the activity (Helland et al. 1998).

The high number of methionines found for many cold-adapted proteins is thought to decrease the stability (Aittaleb et al. 1997; Gudmundsdóttir et al. 1996; Leth-Larsen et al. 1996), a theory supported by the observed replacement of methionines to leucine or isoleucine in the thermophilic adenylate kinases (Haney et al. 1997). The mechanism by which methionines destabilize the protein is not known, but mutation studies on T4 lysozyme revealed a less stable structure when more methionines were introduced (Gassner et al. 1996), and lower refolding yield was observed when an extra methionine residue was mutated at the N-terminus of hen egg-white lysozyme (Mine et al. 1997). The six cold fish trypsins all contain six methionines, whereas the higher vertebrate trypsins on average have about two, for example. A high number of methionines does, however, seem not to be unique for the cold trypsins but common to all trypsins with marine origin, and is therefore probably not a main factor responsible for the reduced stability of the cold-adapted trypsins.

Cationic salmon trypsin (CST), which shows similar activity and stability profiles to the mammalian trypsins, shares structural features with both the higher vertebrate and the cold fish trypsins. CST resembles the warm-active trypsins with a similar packing of the core and number of charged residues and has a similar orientation of the autolysis loop, while it resembles the cold trypsins in the number of methionine residues and hydrophilicity. By elimination, it therefore seems likely that the main cold-adaptation determinants of trypsins are the reduced hydrophobicity and packing of the core (mainly, fewer isoleucines), the lack of the Pro28–Tyr29 stacking, no Pro152, and the deletion of Tyr151 in the autolysis loop. The apparently reduced stability of a few loops and of the C-terminal helix might also be of importance. Mutational studies have shown that the total stability of proteins is enhanced additively (Eijsink et al. 1995; Serrano et al. 1993; Watanabe et al. 1994; Zhang et al. 1995); thus, the reduced stability of the cold trypsins is indeed likely explained by the specific amino acid substitutions pointed out in this study. The increased catalytic efficiency of the cold enzymes is more difficult to explain from the specific substitutions, but the anionic character and the autolysis loop conformation could be important. However, amino acid replacements causing reduced stability are, in many cases, also likely to increase the flexibility of the molecule. Whether increased dynamic behavior is the only explanation for the increased activity can probably only be resolved by extensive mutational studies in combination with more three-dimensional structural data.

To focus on the residue determinants of the cold-adapted trypsins and include all the available vertebrate trypsin sequences, we deliberately omitted a detailed structural discussion. In a subsequent paper we intend to study further the structural impact of the suggested determinants for a selection of both cold- and warm-active trypsin structures.

Acknowledgments Professor Lars Kr. Hansen is thanked for critical reading of the manuscript. The Norwegian Research Council (AOS, HKSL) provided financial support.

References

Aghajari N, Feller G, Gerday C, Haser R (1998) Crystal structures of the psychrophilic α-amylase from *Alteromonas haloplanctis* in its native form and complexed with an inhibitor. Protein Sci 7:564–572

Aittaleb M, Hubner R, Lamotte-Brasseur J, Gerday C (1997) Cold adaptation parameters derived from cDNA sequencing and molecular modelling of elastase from Antarctic fish *Notothenia neglecta*. Protein Eng 10:475–477

Alvarez M, Zeelen JP, Mainfroid V, Rentier-Delrue F, Martial JA, Wyns L, Wierenga RK, Maes D (1998) Triose-phosphate isomerase (TIM) of the psychrophilic bacterium *Vibrio marinus*. Kinetic and structural properties. J Biol Chem 273:2199–2206

Arpigny JL, Feller G, Davail S, Genicot S, Narinx E, Zekhnini Z, Gerday C (1994) Molecular adaptations of enzymes from thermophilic and psychrophilic organisms. In: Gilles R (ed) Advances in comparative and environmental physiology, vol 20. Springer, Berlin Heidelberg New York, pp 269–295

Arpigny JL, Lamotte J, Gerday C (1997) Molecular adaptation to cold of an Antarctic bacterial lipase. J Mol Catal B 3:29–35

Ásgeirsson B, Fox JW, Bjarnason JB (1989) Purification and characterization of trypsin from the poikilotherm *Gadus morhua*. Eur J Biochem 180:85–94

Barrick D, Hughson FM, Baldwin RL (1994) Molecular mechanisms of acid denaturation. The role of histidine residues in the partial unfolding of apomyoglobin. J Mol Biol 237:588–601

Bartunik HD, Summers LJ, Bartsch HH (1989) Crystal structure of bovine β-trypsin at 1.5 Å resolution in a crystal form with low molecular packing density. Active site geometry, ion pairs and solvent structure. J Mol Biol 210:813–828

Berglund GI, Smalås AO, Hordvik A, Willassen NP (1995a) Structure of anionic salmon trypsin in a second crystal form. Acta Crystallogr D 51:725–730

Berglund GI, Willassen NP, Hordvik A, Smalås AO (1995b) Structure of native pancreatic elastase from North Atlantic salmon at 1.61 Å resolution. Acta Crystallogr D 51:925–937

Britton KL, Baker PJ, Borges KM, Engel PC, Pasquo A, Rice DW, Robb FT, Scandurra R, Stillman TJ, Yip KS (1995) Insights into thermal stability from a comparison of the glutamate dehydrogenases from *Pyrococcus furiosus* and *Thermococcus litoralis*. Eur J Biochem 229:688–695

Cohen T, Gertler A, Birk Y (1981a) Pancreatic proteolytic enzymes from carp (*Cyprinus carpio*)-I. Purification and physical properties of trypsin, chymotrypsin, elastase and carboxypeptidase B. Comp Biochem Physiol B 69:639–646

Cohen T, Gertler A, Birk Y (1981b) Pancreatic proteolytic enzymes from carp (*Cyprinus carpio*)-II. Kinetic properties and inhibition studies of trypsin, chymotrypsin, and elastase. Comp Biochem Physiol B 69:647–653

Craik CS, Choo QL, Swift GH, Quinto C, MacDonald RJ, Rutter WJ (1984) Structure of two related rat pancreatic trypsin genes. J Biol Chem 259:14255–14264

Davail S, Feller G, Narinx E, Gerday C (1994) Cold adaptation of proteins. Purification, characterization, and sequence of the heat-labile subtilisin from the Antarctic psychrophile *Bacillus* TA41. J Biol Chem 269:17448–17453

Delboni LF, Mande SC, Rentier-Delrue F, Mainfroid V, Turley S, Vellieux FM, Martial JA, Hol WG (1995) Crystal structure of recombinant triosephosphate isomerase from *Bacillus stearothermophilus*. An analysis of potential thermostability factors in six isomerases with known three-dimensional structures points to the importance of hydrophobic interactions. Protein Sci 4:2594–2604

Earnest T, Fauman E, Craik CS, Stroud R (1991) 1.59 Å structure of trypsin at 120 K: comparison of low temperature and room temperature structures. Proteins 10:171–187

Eijsink VG, Veltman OR, Aukema W, Vriend G, Venema G (1995) Structural determinants of the stability of thermolysin-like proteinases. Nat Struct Biol 2:374–379

Emi M, Nakamura Y, Ogawa M, Yamamoto T, Nishide T, Mori T, Matsubara K (1986) Cloning, characterization and nucleotide sequences of two cDNAs encoding human pancreatic trypsinogens. Gene 41:305–310

Fehlhammer H, Bode W (1975) The refined crystal structure of bovine β-trypsin at 1.8 Å resolution. I. Crystallization, data collection and application of Patterson search techniques. J Mol Biol 98:683–692

Feller G, Arpigny JL, Narnix E, Garday C (1997) Molecular adaptations of enzymes from psychrophilic organisms. Comp Biochem Physiol A 118:495–499

Feller G, Gerday C (1997) Psychrophilic enzymes: molecular basis of cold adaptation. Cell Mol Life Sci 53:830–841

Feller G, Lonhienne T, Deroanne C, Libioulle C, VanBeeumen J, Gerday C (1992) Purification, characterization, and nucleotide sequence of the thermolabile α-amylase from the Antarctic psychrotroph *Alteromonas haloplanctis* A23. J Biol Chem 267:5217–5221

Feller G, Narinx E, Arpigny JL, Aittaleb M, Baise E, Genicot S, Gerday C (1996) Enzymes from psychrophilic organisms. FEMS Microbiol Rev 18:189–202

Feller G, Payan F, Theys F, Qian M, Haser R, Gerday C (1994) Stability and structural analysis of α-amylase from the Antarctic psychrophile *Alteromonas haloplanctis* A23. Eur J Biochem 222: 441–447

Fersht AR, Serrano L (1993) Principles of protein stability derived from protein engineering experiments. Curr Opin Struct Biol 3:75–83

Fletcher TS, Alhadeff M, Craik CS, Largman C (1987) Isolation and characterization of a cDNA encoding rat cationic trypsinogen. Biochemistry 26:3081–3086

Gaboriaud C, Serre L, Guy-Crotte O, Forest E, Fontecilla-Camps JC (1996) Crystal structure of human trypsin 1: unexpected phosphorylation of Tyr151. J Mol Biol 259:995–1010

Gassner NC, Baase WA, Matthews BW (1996) A test of the "jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. Proc Natl Acad Sci USA 93:12155–12158

Genicot S, Feller G, Gerday C (1988) Trypsin from antarctic fish (*Papanotothenia magellanica forster*) as compared with trout (*Salmo gairdneri*) trypsin. Comp Biochem Physiol B 90:601–609

Genicot S, Rentier-Delrue F, Edwards D, VanBeeumen J, Gerday C (1996) Trypsin and trypsinogen from an Antarctic fish: molecular basis of cold adaptation. Biochim Biophys Acta 1298:45–57

Gerday C, Aittaleb M, Arpigny JL, Baise E, Chessa JP, Garsoux G, Petrescu I, Feller G (1997) Psychrophilic enzymes: a thermodynamic challenge. Biochim Biophys Acta 1342:119–131

Gerike U, Danson MJ, Russell NJ, Hough DW (1997) Sequencing and expression of the gene encoding a cold-active citrate synthase from an Antarctic bacterium, strain DS2-3R. Eur J Biochem 248:49–57

Gilberg A, Øverbø K (1990) Purification and characterization of pancreatic elastase from Atlantic cod (*Gadus morhua*). Comp Biochem Physiol B 97:775–782

Gudmundsdóttir A, Gudmundsdóttir E, Óskarsson S, Bjarnason JB, Eakin AK, Craik CS (1993) Isolation and characterization of cDNAs from Atlantic cod encoding two different forms of trypsinogen. Eur J Biochem 217:1091–1097

Gudmundsdóttir A, Óskarsson S, Eakin AE, Craik CS, Bjarnason JB (1994) Atlantic cod cDNA encoding a psychrophilic chymotrypsinogen. Biochim Biophys Acta 1219:211–214

Gudmundsdóttir E, Spilliaert R, Yang Q, Craik CS, Bjarnason JB, Gudmundsdóttir A (1996) Isolation and characterization of two cDNAs from Atlantic cod encoding two distinct psychrophilic elastases. Comp Biochem Physiol B 113:795–801

Haney P, Konisky J, Koretke KK, Luthey-Schulten Z, Wolynes PG (1997) Structural basis for thermostability and identification of potential active site residues for adenylate kinases from the archaeal genus *Methanococcus*. Proteins 28:117–130

Hartley BS, Kauffman DL (1966) Corrections to the amino acid sequence of bovine chymotrypsinogen A. Biochem J 101:229–231

Hedstrom L, Farr-Jones S, Kettner CA, Rutter WJ (1994a) Converting trypsin to chymotrypsin: ground-state binding does not determine substrate specificity. Biochemistry 33:8764–8769

Hedstrom L, Perona JJ, Rutter WJ (1994b) Converting trypsin to chymotrypsin: residue 172 is a substrate specificity determinant. Biochemistry 33:8757–8763

Hedstrom L, Szilagy L, Rutter WJ (1992) Converting trypsin to chymotrypsin: the role of surface loops. Science 255:1249–1253

Heimstad ES (1996) Structural and dynamical comparison of two homologous enzymes. Exploring the protein dynamics of salmon and bovine trypsin with molecular dynamics simulations. Doctoral thesis,

Institute of Mathematical and Physical Sciences, Department of Chemistry, University of Tromsø, Norway
Heimstad ES, Hansen LK, Smalås AO (1995) Comparative molecular dynamics simulation studies of salmon and bovine trypsins in aqueous solution. Protein Eng 8:379–388
Helland R, Leiros I, Willassen NP, Berglund GI, Smalås AO (1998) The crystal structure of anionic salmon trypsin in complex with bovine pancreatic trypsin inhibitor. Eur J Biochem 256:317–324
Hermodson MA, Ericsson LH, Neurath H, Walsh KA (1973) Determination of the amino acid sequence of porcine trypsin by sequenator analysis. Biochemistry 12:3146–3153
Hjelmeland K, Raa J (1982) Characteristics of two trypsin type isozymes isolated from the arctic fish capelin (*Mallotus villosus*). Comp Biochem Physiol B 71:557–562
Huang Q, Liu S, Tang Y (1993) Refined 1.6 Å resolution crystal structure of the complex formed between porcine β-trypsin and MCTI-A, a trypsin inhibitor of the squash family. Detailed comparison with bovine β-trypsin and its complex. J Mol Biol 229:1022–1036
Ishii A, Suzuki M, Sahara T, Takada Y, Sasaki S, Fukunaga N (1993) Genes encoding two isocitrate dehydrogenase isozymes of a psychrophilic bacterium, *Vibrio* sp. strain ABE-1. J Bacteriol 175:6873–6880
Jaenicke R (1990) Protein structure and function at low temperatures. Philos Trans R Soc Lond B Biol Sci 326:535–551
Jaenicke R (1991) Protein stability and molecular adaptation to extreme conditions. Eur J Biochem 202:715–728
Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637
Kalác J (1976) Comparative study of mackerel (*Scomber scombrus*) anionic trypsin and bovine trypsin. Biologia (Bratisl) 31:973–982
Kang J, Wiegand U, Müller-Hill B (1992) Identification of cDNAs encoding two novel rat pancreatic serine proteases. Gene 110:181–187
Karlsen S, Hough E, Olsen RL (1998) The crystal structure and proposed amino acid sequence of a pepsin from Atlantic cod (*Gadus morhua*). Acta Crystallogr D 54:32–46
Kimland M, Russick C, Marks WH, Borgstrom A (1989) Immunoreactive anionic and cationic trypsin in human serum. Clin Chim Acta 184:31–46
Kozlovskaya EP, Elyakova LA (1974) Purification and properties of trypsin-like enzymes from the starfish *Lysastrosoma anthosticta*. Biochim Biophys Acta 371:63–70
Kristjansson MM, Ásgeirsson B, Bjarnason JB (1997) Serine proteinases from cold-adapted organisms. Adv Exp Med Biol 415:27–46
Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157:105–132
Le Huerou I, Wicker C, Guilloteau P, Toullec R, Puigserver A (1990) Isolation and nucleotide sequence of cDNA clone for bovine pancreatic anionic trypsinogen. Structural identity within the trypsin family. Eur J Biochem 193:767–773
Leth-Larsen R, Ásgeirsson B, Thorolfsson M, Norregaard-Madsen M, Hojrup P (1996) Structure of chymotrypsin variant B from Atlantic cod, *Gadus morhua*. Biochim Biophys Acta 1297:49–56
Lütcke H, Rausch U, Vasiloudes P, Scheele GA, Kern HF (1989) A fourth trypsinogen (P23) in the rat pancreas induced by CCK. Nucleic Acids Res 17:6736–6736
MacDonald RJ, Stary SJ, Swift GH (1982) Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide sequences of the cloned cDNAs. J Biol Chem 257:9724–9732
Malakauskas SM, Mayo SL (1998) Design, structure and stability of a hyperthermophilic protein variant. Nat Struct Biol 5:470–475
Male R, Lorens JB, Smalås AO, Torrissen KR (1995) Molecular cloning and characterization of anionic and cationic variants of trypsin from Atlantic salmon. Eur J Biochem 232:677–685
Marquart M, Walter J, Deisenhofer J, Bode W, Huber R (1983) The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. Acta Crystallogr B 39:480–490
Marshall CJ (1997) Cold-adapted enzymes. Trends Biotechnol 15:359–364
Martínez A, Olsen RL, Serra JL (1988) Purification and characterization of two trypsin-like enzymes from the digestive tract of anchovy *Engraulis encrasicholus*. Comp Biochem Physiol B 91:677–684
Mikes O, Holeysovsky V, Tomasek V, Sorm F (1966) Covalent structure of bovine trypsinogen. The position of the remaining amides. Biochem Biophys Res Commun 24:346–352
Mine S, Ueda T, Hashimoto Y, Imoto T (1997) Improvement of the refolding yield and solubility of hen egg-white lysozyme by altering the Met residue attached to its N-terminus to Ser. Protein Eng 10:1333–1338
Mrabet NT, van den Broeck A, van den Brande I, Stanssens P, Laroche Y, Lambeir AM, Matthijssens G, Jenkins J, Chiadmi M, van Tilbeurgh H, Rey F, Janini J, Quax WJ, Lasters I, de Maeyer M, Wodak SJ (1992) Arginine residues as stabilizing elements in proteins. Biochemistry 31:2239–2253
Muñoz V, Serrano L (1995) Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. J Mol Biol 245:275–296
Osnes KK, Mohr V (1985) On the purification and characterization of three anionic, serine-type peptide hydrolases from Antarctic krill *Euphausia superba*. Comp Biochem Physiol B 82:607–619
Outzen H, Berglund GI, Smalås AO, Willassen NP (1996) Temperature and pH sensitivity of trypsins from Atlantic salmon (*Salmo salar*) in comparison with bovine and porcine trypsin. Comp Biochem Physiol B 115:33–45
Pinsky SD, LaForge KS, Scheele G (1985) Differential regulation of trypsinogen mRNA translation: full-length mRNA sequences encoding two oppositely charged trypsinogen isoenzymes in the dog pancreas. Mol Cell Biol 5:2669–2676
Raae AJ, Walther BT (1989) Purification and characterization of chymotrypsin, trypsin and elastase like proteinases from cod (*Gadus morhua* L.). Comp Biochem Physiol B 93:317–324
Read RJ, James MN (1988) Refined crystal structure of *Streptomyces griseus* trypsin at 1.7 Å resolution. J Mol Biol 200:523–551
Roach JC, Wang K, Gan L, Hood L (1997) The molecular evolution of the vertebrate trypsinogens. J Mol Evol 45:640–652
Rypniewski WR, Dambmann C, von der Osten C, Dauter M, Wilson KS (1995) Structure of inhibited trypsin from *Fusarium oxysporum* at 1.55 Å. Acta Crystallogr D 51:73–84
Rypniewski WR, Perrakis A, Vorgias CE, Wilson KS (1994) Evolutionary divergence and conservation of trypsin. Protein Eng 7:57–64
Schrøder H-K, Willassen NP, Smalås AO (1998) Structure of a non-psychrophilic trypsin from a cold-adapted fish species. Acta Crystallogr D 54:780–798
Serrano L, Day AG, Fersht AR (1993) Step-wise mutation of barnase to binase. A procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. J Mol Biol 233:305–312
Shi YB, Brown DD (1990) Developmental and thyroid hormone-dependent regulation of pancreatic genes in *Xenopus laevis*. Genes Dev 4:1107–1113
Simpson BK, Haard NF (1984a) Purification and characterization of trypsin from the Greenland cod (*Gadus ogac*). 1. Kinetic and thermodynamic characteristics. Can J Biochem Cell Biol 62:894–900
Simpson BK, Haard NF (1984b) Trypsin from Greenland cod, *Gadus ogac*. Isolation and comparative properties. Comp Biochem Physiol B 79:613–622
Smalås AO, Heimstad ES, Hordvik A, Willassen NP, Male R (1994) Cold adaption of enzymes: structural comparison between salmon and bovine trypsins. Proteins 20:149–166
Stevenson BJ, Hagenbuchle O, Wellauer PK (1986) Sequence organisation and transcriptional regulation of the mouse elastase II and trypsin genes. Nucleic Acids Res 14:8307–8330
Tani T, Kawashima I, Mita K, Takiguchi Y (1990) Nucleotide sequence of the human pancreatic trypsinogen III cDNA. Nucleic Acids Res 18:1631–1631
Tanji M, Yakabe E, Kageyama T, Takahashi K (1996) The primary structure of the major pepsinogen from the gastric mucosa of tuna stomach. J Biochem (Tokyo) 120:647–656
Tanner JJ, Hecht RM, Krause KL (1996) Determinants of enzyme thermostability observed in the molecular structure of *Thermus aquaticus* D-glyceraldehyde-3-phosphate dehydrogenase at 2.5 Å resolution. Biochemistry 35:2597–2609
Titani K, Ericsson LH, Neurath H, Walsh KA (1975) Amino acid sequence of dogfish trypsin. Biochemistry 14:1358–1366

Uchida N, Tsukayama K, Nishide E (1984) Properties of two main anionic trypsins from the pyloric caeca of chum salmon. Bull Jpn Soc Sci Fish 50:313–321

Voytek P, Gjessing EC (1971) Studies of an anionic trypsinogen and its active enzyme from porcine pancreas. J Biol Chem 246:508–516

Wallon G, Lovett ST, Magyar C, Svingor A, Szilagyi A, Zavodszky P, Ringe D, Petsko GA (1997) Sequence and homology model of 3-isopropylmalate dehydrogenase from the psychrotrophic bacterium *Vibrio* sp. I5 suggest reasons for thermal instability. Protein Eng 10:665–672

Wang K, Gan L, Lee I, Hood L (1995) Isolation and characterization of the chicken trypsinogen gene family. Biochem J 307:471–479

Watanabe K, Masuda T, Ohashi H, Mihara H, Suzuki Y (1994) Multiple proline substitutions cumulatively thermostabilize *Bacillus cereus* ATCC7064 oligo-1,6-glucosidase. Irrefragable proof supporting the proline rule. Eur J Biochem 226:277–283

Winter WP, Neurath H (1970) Purification and properties of a trypsin-like enzyme from the starfish *Evasterias trochelii*. Biochemistry 9: 4673–4679

Yip KS, Stillman TJ, Britton KL, Artymiuk PJ, Baker PJ, Sedelnikova SE, Engel PC, Pasquo A, Chiaraluce R, Consalvi V (1995) The structure of *Pyrococcus furiosus* glutamate dehydrogenase reveals a key role for ion-pair networks in maintaining enzyme stability at extreme temperatures. Structure (Lond) 3:1147–1158

Yoshinaka R, Sato M, Suzuki T, Ikeda S (1983a) Enzymatic characterization of anionic trypsin of the catfish (*Parasilurus asotus*). Comp Biochem Physiol B 77:1–6

Yoshinaka R, Suzuki T, Sato M, Ikeda S (1983b) Purification and some properties of anionic trypsin from the catfish pancreas. Bull Jpn Soc Sci Fish 49:207–212

Zhang XJ, Baase WA, Shoichet BK, Wilson KP, Matthews BW (1995) Enhancement of protein stability by the combination of point mutations in T4 lysozyme is additive. Protein Eng 8:1017–1022