brought to
provided by Helsingin yliopisto
citation and similar papers at core.ac.uk

High Molecular Weight Cysteine Proteinase Inhibitors in Atlantic Salmon and Other Fish Species

ANNE OLONEN

Institute of Biotechnology
Protein Chemistry Laboratory and
Department of Biological and Environmental Sciences
Division of Biochemistry
University of Helsinki

Dissertationes Biocentri Viikki Universitatis Helsingiensis

18/2004

HIGH MOLECULAR WEIGHT CYSTEINE PROTEINASE INHIBITORS IN ATLANTIC SALMON AND OTHER FISH SPECIES

Anne Olonen

Institute of Biotechnology

Protein Chemistry Laboratory and

Department of Biological and Environmental Sciences

Division of Biochemistry

Faculty of Biosciences

University of Helsinki

Academic dissertation

To be presented for public criticism, with the permission of the Faculty of Biosciences, University of Helsinki, in the auditorium 2402 (Telkänpönttö) at Viikki Biocenter (Viikinkaari 1, Helsinki), on August 13th, 2004, at 12 o'clock.

Helsinki 2004

Supervised by
Docent Nisse Kalkkinen
Institute of Biotechnology
University of Helsinki

Reviewed by
Docent Jaana Vesterinen
Institute of Biomedicine
University of Helsinki

Professor Heikki Rauvala
Neuroscience Center
University of Helsinki

Opponent
Professor Jukka Finne
Department of Medical Biochemistry and
Molecular Biology
University of Turku

ISBN 952-10-1923-9 (print)
ISBN 952-10-1924-7 (ethesis, pdf)
ISSN 1239-9469

Press:: N-Paino, Lahti 2004

Hei, peikot!
Tulkaas nyt jo alas;
juusto nousee kuuseen.

"Vaikka nousis kalakukko,
olet hiljaa, Pränttiukko,
kunnes sulta kysytään! —
Niin, ja sitten — mutta Kala,
mihinkäs me jäimmekään?"

Lauri Viita

CONTENTS

ABBREVIATIONS i
LIST OF ORIGINAL PUBLICATIONS ii
INTRODUCTION iii
1. REVIEW OF THE LITERATURE 1
1.1. Cysteine proteinases 1
1.1.1. Regulation of cysteine proteinases 2
1.1.2. Cysteine proteinases in fish 2
1.2. Cysteine proteinase inhibitors 3
1.2.1. The cystatin superfamily 3
1.2.1.1. Cystatin family 1. Stefins 5
1.2.1.2. Cystatin family 2. Cystatins 5
1.2.1.3. Cystatin family 3. Kininogens 5
1.2.2. Other types of inhibitors 6
1.2.3. Cysteine proteinase inhibitors in fish 7
1.3. Glycosylation of proteins 7
1.3.1. Protein glycosylation in general 7
1.3.2. Glycosylation of fish proteins 9
1.4. Fish species 9
2. AIMS OF THE STUDY 12
3. MATERIALS AND METHODS 13
3.1. Fish materials 13
3.2. Purification and characterisation of the inhibitors 13
3.2.1. Chromatographic methods 13
3.2.2. Inhibition assay 14
3.2.3. SDS/PAGE and N-terminal sequencing 14
3.2.4. Enzymatic digestions and peptide separation 15
3.2.5. Mass spectrometry of proteins and peptides 15
3.2.6. Capillary isoelectric focusing 16
3.3. Glycosylation analysis 16
3.3.1. Preparation of glycans 16
3.3.2. Permethylation 17
3.3.3. Removal of O-acetyl groups 17
3.3.4. Periodate oxidation 17
3.3.5. Gas phase β-elimination 17
3.3.6. Exoglycosidase digestions 18
3.3.7. Mass spectrometry of glycans and glycopeptides 18

3.4. Molecular cloning and DNA sequencing of the inhibitors 18
3.4.1. Isolation of RNA and DNA 18
3.4.2. Primers 18
3.4.3. Polymerase chain reaction (PCR) 19
3.4.4. Rapid amplification of cDNA ends (RACE) 19
3.4.5. DNA sequencing of the salarin gene 20

4. RESULTS 21
4.1. Purification of cysteine proteinase inhibitors from fish skin 21
4.2. Characterization of the inhibitors 22
4.2.1. Chemical and physical properties 22
4.2.2. Protein sequencing 23
4.2.3. Functional properties 25
4.3. Analysis of glycan structures 26
4.4. Cysteine proteinase inhibitors in other fish species 28
4.5. Molecular cloning and DNA sequencing of salarin 29

5. DISCUSSION 31
5.1. Cysteine proteinase inhibitors from fish skin 31
5.2. Salarin 32
5.3. Glycosylation of cysteine proteinase inhibitors 32

6. CONCLUSIONS 35

7. ACKNOWLEDGEMENTS 36

8. REFERENCES 38

ABBREVIATIONS

BANA	N_α-benzoyl-DL-arginine-2-naphtylamide
CAPS	3-(cyclohexylamino)propane-1-sulphonic acid
cDNA	complementary deoxyribonucleic acid
ConA	Concanavalin A
Da	dalton, molecular mass unit (1 Da = 1/12 ^{12}C mol [g])
Gal	D-galactose
GlcNAc	N-acetylglucosamine
HPLC	high performance liquid chromatography
KDN	ketodeoxynononic acid
MALDI TOF MS	matrix assisted laser desorption ionization time of flight mass spectrometry
NDV	Newcastle Disease Virus
NeuNAc	N-acetylneuraminic acid
PAGE	polyacrylamide gel electrophoresis
pI	isoelectric point
PCR	polymerase chain reaction
PVDF	polyvinylidene difluoride
RP	reversed phase
RACE	rapid amplification of cDNA ends
SDS	sodium dodecyl sulfate

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following articles which are referred to in the text by their Roman numerals (I-IV):

I Ylönen, A., Rinne, A., Herttuainen, J., Bøgwald, J., Järvinen, M., Kalkkinen, N. (1999).
Atlantic salmon (*Salmo salar* L.) skin contains a novel kininogen and another cysteine proteinase inhibitor. *Eur. J. Biochem.* 266, 1066-1072.

II Ylönen, A., Kalkkinen, N., Saarinen, J., Bøgwald, J., Helin, J. (2001)
Glycosylation analysis of two cysteine proteinase inhibitors from Atlantic salmon skin: di-O-acetylated sialic acids are the major sialic acid species on N-glycans. *Glycobiology* 11, 523-531[a].

III Ylönen, A., Helin, J., Bøgwald, J., Jaakola, A., Rinne, A., Kalkkinen, N. (2002)
Purification and characterization of novel kininogens from spotted wolffish and Atlantic cod. *Eur. J. Biochem*. 269, 2639-2646.

IV Olonen, A., Kalkkinen, N., Paulin, L. (2003)
A new type of cysteine proteinase inhibitor – the salarin gene from Atlantic salmon (*Salmo salar* L.) and Arctic charr (*Salvelinus alpinus*). *Biochimie* 85, 677-681.

[a] Reprinted from *Glycobiology* by permission of Oxford University Press and Society for Glycobiology.

INTRODUCTION

Proteolytic enzymes are extensively distributed in all types of organisms; 2% of all gene products are peptidases [1]. They catalyze the cleavage of the peptide bond. Depending on the cleavage site they can be divided into proteases which remove terminal residues and proteinases which act on peptide bonds in the inner polypeptide chain. Proteolytic enzymes can be assigned to four distinct groups on the basis of their catalytic types: serine, cysteine, aspartic and metallo proteinases [2]. The names of the enzyme groups refer to the components of their active site.

Cysteine proteinases are a large group including lysosomal cathepsins. Physiologically cysteine proteinases have an important role in protein metabolism and turnover. In addition, processing and activation of many prohormones, proenzymes and peptides is conducted by cysteine proteinases.

The action of cysteine proteinases can be regulated and inhibited by endogenous, natural inhibitors as well as by compartmentalization and surrounding conditions e.g. pH or chelating agents. If this well-balanced control system is disturbed it can lead to serious damage. Many pathogens have their own cysteine proteinases to invade hosts. Disturbed expression of cysteine proteinases and their natural inhibitors is associated with some pathological conditions like arthritis, neurological disorders and cancer.

The interactions between cysteine proteinases and their inhibitors have been a target of intensive studying for the last two decades. The specific inhibitors of cysteine proteinases are needed in preventing unwanted, potentially destructive proteolysis; they can be utilized in therapy and in research [3]. A deeper understanding of these interactions may give us the means to control the proteolytic activities. New structural data also adds information about the evolution of both cysteine proteinases and their protein inhibitors as well as about the species studied.

The endogenous inhibitors of cysteine proteinases, cystatins, are reported to act as defensive agents against bacteria, viruses and pests. One of the first cystatins was discovered from the rat skin [4] and soon after that from human skin [5]. As fishes live in pathogen-rich environments, it is possible that their skin too contains this non-specific form of defence mechanism.

The present study characterizes high molecular weight cysteine proteinase inhibitors isolated from the skins of some salmonid and non-salmonid fish species.

1. REVIEW OF THE LITERATURE

1.1. Cysteine proteinases

The cysteine proteinase superfamily consists of about 20 families which are grouped into six clans (CA, CB, …, CF) [1]. All cysteine proteinases have a similar catalytic mechanism where cysteine and histidine residues are located in the active centre [6]. Their classification is based on the order of the catalytic residues (His/Cys or Cys/His) and other structural features, but it is likely to change as more structural data becomes available [6]. The most studied family is papain-like cysteine proteinases grouped into clan CA with catalytic residues Cys-25 and His-159 (papain numbering) [1]. In this clan there are enzymes from all organisms; animals (cathepsins, calpain), plants (papain, ficin), protozoa (cruzipain), bacteria (streptopain) and viruses (V-cath proteinase) [1]. Comprehensive information about different types of proteinases can be found in the MEROPS database [7,8].

Cysteine proteinases of the papain family are either secreted or lysosomal enzymes. In plants they are located in vacuoles and in secreted fluids like the latex of papaya. They are synthesized as inactive zymogens with signal peptides and with an N-terminal (usually glycosylated) propeptide region [6]. The proregion is needed for proper folding, stability and transport of the enzyme [9]. It also inhibits proteolytic activity until the enzyme is needed. Thus, cleavage of the propeptide is essential for activation of the enzyme [6, 10].

The most widely characterized enzymes in papain-like cysteine proteinase family are cathepsins [reviewed in 1, 11, 12]. On the basis of their propeptides papain-like cysteine proteinases can be divided into two subfamilies. Cathepsin L -like enzymes have a 100 amino acid long propeptide with the interspersed amino acid motif ERFNIN [10, 13]. This motif is not found in cathepsin B -like enzymes which have shorter propeptides (60 amino acids) (Figure 1). [9, 10].

In mammals the major function of cysteine proteinases is lysosomal protein degradation and specific protein processing [10], but they also have a role in cancer [14],

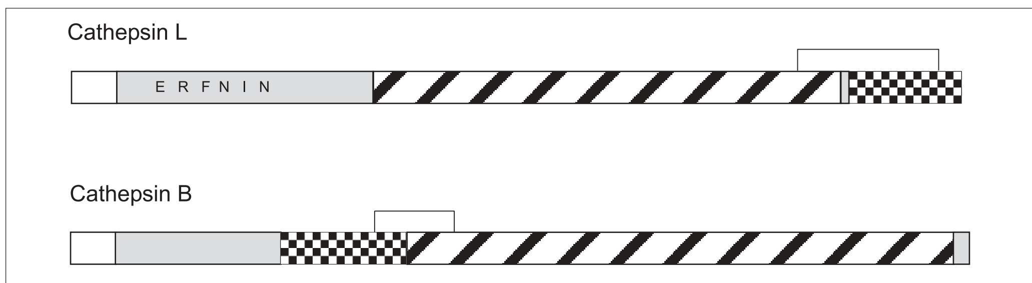

Figure 1. A picture of the structural properties of two types of cathepsins belonging to the papain family. Human cathepsin L (P07711) has a 96 amino acid long propeptide region (gray box) with the conserved interspersed motif ERFNIN. The mature cathepsin L consists of a heavy chain (hatched box) and a light chain (squared box) connected with a disulfide bridge. Human cathepsin B (P07858) has a 62 amino acid long propeptide region. The mature cathepsin B consists likewise of heavy chain and light chain connected with a disulfide bridge. Both cathepsins have a short signal sequence (white box) in their N-terminus.

antigen processing [15], arthritis [16], Alzheimer's disease, muscular disorders and gingivitis [17]. Several cysteine proteinases of pathogen origin have also been characterized in great detail: histolysin and amoebapain of *Entamoeba histolytica*, aid the amebiasis infection and invasion by proteolytic actions [18], cruzipain, a cysteine proteinase of *Trypanosoma cruzi* causing Chagas disease [19] and falcipain , a malarial haemoglobinase of *Plasmodium falciparum* [1]. The regulation of destructive proteolysis could prevent pathogen invasion and severe tissue damage in disease.

1.1.1. Regulation of cysteine proteinases

The biological activity of cysteine proteinases is regulated by inhibitors, pH and zymogen activation [20]. Non-specific inhibitors of peptidases in general are ±-macroglobulins, which bind and inhibit irreversibly most proteinases; peptide aldehydes (leupeptin and antipain) acting as reversible transition-state analogues and peptide chloromethyl ketones which inactivate proteinases by binding the inhibitor and alkylating the active site [21]. Peptide diazomethanes and peptide epoxides alkylate active-site cysteine inhibiting particularly cysteine proteinases but they can not discriminate between different cysteine proteinases [21]. New synthetic cysteine proteinase inhibitors are under research [22]. Natural endogenous inhibitors include proregions of proteinases and inhibitory molecules with protein structure such as cystatins (see below).

All papain-like cysteine proteinases are synthesized as inactive proenzymes, zymogens. The proregion is needed for proper folding and transport of newly synthesized enzyme [10]. Cathepsins are transported from the Golgi apparatus to the lysosomes where the pH is acidic [10]. This acidic environment has two functions. It triggers the proteolytic cleavage of propeptide yielding mature cathepsin [23] and provides the optimum pH for these enzymes to function [24]. A neutral pH as well too acidic pH rapidly inactivates cathepsins [25]. Further, one mechanism of regulation is that cathepsins are stable only in a certain pH range which differs for every cathepsin. Throughout the maturation the pH declines in lysosomes. Therefore different cathepsins are activated at different maturation stages of lysosomes [26]. In mature lysosomes pH can reach values as low as 3.8 which is sufficiently low to cause irreversible denaturation of cathepsins B, S and L [11].

1.1.2. Cysteine proteinases in fish

Elevated levels of cathepsins B, D, H and L have been reported in the white muscle of chum salmon (*Oncorhynchus keta*) in spawning migration [27]. Cathepsin L [28] and cathepsin B [29] have been purified and characterized from the chum salmon. One kind of cysteine proteinase, miltpain, is found in the milt of both chum salmon [30] and Pacific cod [31]. The determined N-terminal sequences (17 amino acid residues) for miltpains are not similar to each other and neither do they show any significant homology to any other proteins in the databases. However, both miltpains share other properties like substrate specificity towards paired basic residues. These results suggest that miltpain is

a new type of cysteine proteinase [31]. Preliminary studies have also been made on *Mujil auratus* cathepsin B [32, 33] and arrowtooth flounder (*Atheresthes stomias*) cathepsin L [34, 35].

In the GenBank there is a gene sequence (accession number AF281331) from rainbow trout encoding a protein which was submitted as a cathepsin. The gene is found by a method where a pool of clones were sequenced (suppression subtractive hybridization) and then analysed with the BLASTX program. The most likely hit for this gene was cathepsin which was given as identification for the new sequence and the result was not verified [36], as discussed in more detail in the discussion section.

Salmons cease feeding in fresh water before spawning [37], hence the cathepsins are probably needed at this stage to degrade proteins as a source of energy. Autolysis of mature chum salmon muscle has been shown to be caused by cathepsins [38]. Similar observations have also been made for hake (*Merluccius hubbsi*) where autolysis in muscle takes place even after frozen storage [39]. Both chum salmon and hake are alimentary fish, so their muscle autolysis deteriorates the quality of their fillets.

Epidermal proteinases may play a role in the natural immunity of fish. This non-specific defence mechanism has been studied in different eel species [40] and in Japanese flounder [41]. In these species cathepsins B and L have been suggested to act as antibacterial agents. Our collaborator group purified cathepsins B, L, D and H from Atlantic salmon (manuscript in preparation). The results of their immunohistochemical localization studies show that cathepsins B and L are localized mainly in lysosomal-like vesicles of the kidney and liver and also in potential antigen presenting cells in immunologically important organs [42].

1.2. Cysteine proteinase inhibitors

1.2.1. The cystatin superfamily

The first inhibitor of papain was reported from chicken egg white in 1968 [43]. When this protein was further characterized it was named cystatin [44], being the first member of the cystatin superfamily. Other cysteine proteinase inhibitors were reported independently in several laboratories during the 70-80's [45, 46, 47, 48, 49] thus forming the basis for the cystatin superfamily. The nomenclature and classification of cysteine proteinase inhibitors was decided at the First International Symposium on Cysteine Proteinases and their Inhibitors in 1985 [50]. The cystatin superfamily is divided into three structurally related families: stefins, cystatins and kininogens. It has been proposed that one of their functions is to defend the host against pathogens and parasites [51].

The evolutionary relationships of cystatins were studied on the basis of sequence homology. One model suggests that there has been one archetypal unit which has evolved into four families according to sequence similarity and biological function [52]. Another study proposes that cystatins have evolved from two units: A and B [53]. In this model stefins are made only from unit A thus forming the archetype of the superfamily. Cystatins have units A and B with disulphide bonds whereas kininogens have arisen from gene

triplication for the formula $(AB)_3$. Both models, however, leave a missing link between cystatins and kininogens, consisting of two AB units [53]. A schematic drawing of cystatin structures is presented in Figure 2. According to all the models a distinct feature of cystatin is the amino acid sequence QVVAG.

Also, on a structural basis fetuins, histidine-rich glycoprotein and other cystatin-related proteins, with no inhibitory activity, could be included in the cystatin superfamily [54]. Cystatins of plant origin (phytocystatins) differ slightly from cystatins and stefins by having a consensus sequence LARFADEHN which is absent from other cystatins [55]. Therefore the phytocystatins could be classified as a distinct family in the cystatin superfamily.

Also, completely new types of cysteine proteinase inhibitors have emerged. Our recent comprehension [56] is further expanding the cystatin superfamily into three different groups: The first group are *sensu stricto* cystatins, structurally related proteins inhibiting cysteine proteinases. The second group has a structural similarity to conventional cystatins but no inhibitory activity, like fetuins. The third group are cysteine proteinase inhibitors which share no structural homology with group 1, like chagasin [56]. Table 1 presents this classification with some examples. The protease database MEROPS [8] has its own classification system for all types of protease inhibitors.

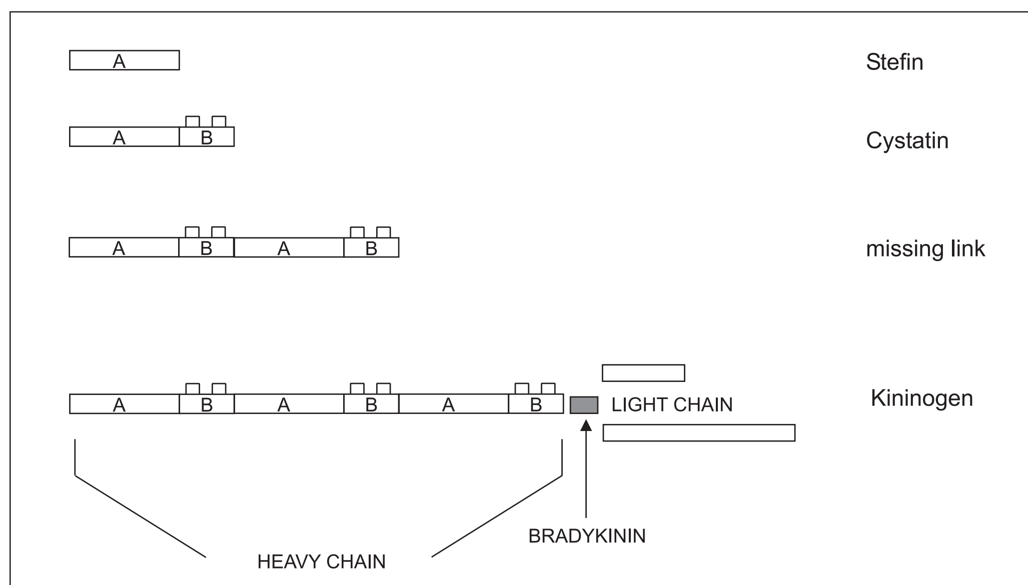

Figure 2. A schematic picture of protein structure in cystatin superfamily. Unit A is archetypal cystatin and contains conserved QVVAG sequence. Unit B has two disulfide bonds. The heavy chain of kininogen consists of three cystatin domains $(AB)_3$ and the light chain of kininogen can be shorter or longer thus determining LMW or HMW kininogen.

Table 1. Classification of cystatins and other cysteine proteinase inhibitors and examples of them in human and fish.

Group I	**Traditional cystatin superfamily**		
	- 1. Stefins	Stefin A	
	- 2. Cystatins	Cystatin C	chum salmon cystatin C
	- 3. Kininogens	HMW Kininogen	Salmon kininogen
	- 4. Phytocystatins	-	-
Group II	**Structurally cystatin-like molecules with no inhibitory activity**		
	Histidine rich glycoprotein (HRG)	human HRG	
	Fetuins	human fetuin	zebrafish fetuin
Group III	**Cysteine proteinase inhibitors with no structural homology to group I**		
	Thyropins	equistatin	chum salmon inhibitor
	Cathepsin L proregion-like	-	salarin

1.2.1.1. Cystatin family 1. Stefins

Stefins are single-chain proteins with neither disulphide bonds nor carbohydrates (Fig 1). Their molecular weight is about 11 kDa. Stefins are localized mainly intracellularly but have been found also in extracellular fluids [26]. The nomenclature of stefins can be confusing because they are also called cystatins e.g. cystatin A means stefin A [51, 57]. The cystatins of human [5], rat [46] and bovine [58,59] origin have been reported. This group is considered to be similar to archetypal cystatins [52, 53].

1.2.1.2. Cystatin family 2. Cystatins

Cystatins contain two characteristic disulfide bonds near their C-terminus (Fig 1). Their molecular weight is a bit higher than that of stefins, about 13 kDa [56]. Usually they occur in secretions like chicken egg-white and other fluids including blood plasma [51]. Cystatins are non-glycosylated with the exception of the rat cystatin C [56]. Some cystatins are also phosphorylated, but this phosphorylation does not effect their inhibitory activity [60, 61].

1.2.1.3. Cystatin family 3. Kininogens

Kininogens consist of an N-terminal heavy chain combined with a bradykinin sequence to a variable length light chain (Fig 1). The heavy chain has three cystatin-like domains

and the length of the light chain determines whether the kininogen is a high molecular weight kininogen (HMW kininogen, 120 kDa) or a low molecular weight kininogen (LMW kininogen, 68 kDa) [56]. Bradykinin is a conserved vasoactive peptide controlling the action of smooth muscle cells. Three types of kininogens are known: HMW, LMW and T-kininogen (only in rat), most of them being sialyl-glycosylated [51]. Kininogens have many different functions in the blood coagulation cascade, muscle contraction, acute phase response and inhibition of cysteine proteinase [62]. Some of these functions are mediated by bradykinin. Bradykinin can be released by the action of kallikrein leaving the N-terminal heavy chain and the C-terminal light chain connected by a single disulfide bridge [62].

The second domain of kininogens is known also to inhibit calpain, although other cystatins are not effective against this enzyme [63].

1.2.2. Other types of cysteine proteinase inhibitors

The first larger group of other cystatins are thyropins (**thyro**globulin type-1 **p**roteinase **in**hibitors) which are inhibitors of cysteine proteinases containing at least one thyroglobulin type-1 domain [64]. Equistatin from sea anemone inhibits cysteine proteinases and is structurally similar to thyroglobulin type-1 [65]. It also inhibits cathepsin D which is an aspartic proteinase [66] whereas traditional cystatins do not inhibit other classes of proteinases. Saxiphilin is an inhibitor of papain-like cysteine proteinases isolated from bullfrog. It has two thyroglobulin type-1 domains [67].

One thyropin has also been isolated from the eggs of chum salmon (see below) [68].

Clitocybin (P82314) is a new type of cysteine proteinase inhibitor from a mushroom. Clitocybin appears to be related to fungal lectins and the team that found it suggested a new family of cysteine proteinase inhibitors called mycocypins [69]. Chagasin (Q966X9) is a cysteine proteinase inhibitor found in *Trypanozoma cruzi* inhibiting both cruzipain and papain, but it has no homology with cystatins or with other published sequences in the databanks [70].

Proregions of cysteine proteinases are known to selectively inhibit their cognate enzymes. There are studies on mammalian proteinases as cathepsin B [71], cathepsin L [72] and cathepsins K, L and S [73] which are selectively inhibited by their respective recombinant or synthetic propeptides. Also trypanosomal cysteine proteinases congopain and cruzipain are inhibited by their propeptides but not by cathepsin propeptides [74].

A novel inhibitor protein from the silkmoth *Bombyx mori* is homologous to propeptide regions of cysteine proteinases [75]. Other propeptide-like inhibitors are CTLA-2 [76] and Drosophila CTLA-2-like protein [77]. These findings suggest that propeptide regions inhibiting cysteine proteinases have evolved as autonomous elements. As shown in our studies (I, IV) also a cysteine proteinase inhibitor from the skin of Atlantic salmon has structural similarity with propeptide regions of cathepsins.

1.2.3. Cysteine proteinase inhibitors in fish

A cysteine proteinase inhibitor was described from a fish for the first time in 1988 [78]. The 50 kDa inhibitor was isolated from the muscle of carp and it was shown to inhibit both papain and calpain [78]. No structural data or other evidence of the family of this carp inhibitor was presented, but it may have been kininogen, based on the molecular weight and the fact that kininogens inhibit calpains unlike the other cystatins [63].

Both the proteinases and their inhibitors have been intensively studied in chum salmon (*Oncorhynchus keta*). A thyroglobulin-like cysteine proteinase inhibitor has been found in the eggs of chum salmon [68, 79]. Also, cystatin is found in the muscle of chum salmon [80]. A cystatin homologous to mammalian cystatin C was first isolated and sequenced from the pituitary gland of chum salmon [81] and later cloned and expressed [82]. Cystatin C is reported also to exist in rainbow trout [83, 84]. Cystatin has also been found from carp [85], fugu [86] and zebrafish [87]. These inhibitors are low molecular weight cystatins.

A preliminary report of high molecular weight cysteine proteinase inhibitors in Atlantic salmon (*Salmo salar* L.) appeared in 1998 [88]. As these results were partly erroneous the study was continued leading to publications I, II, III and IV.

There have been no other reports describing fish kininogens prior to our publications. However, earlier fish bradykinin findings imply that kininogens exist in fishes too. Bradykinin had been found from the plasma of steelhead trout [89], bowfin [90] and Atlantic cod [91]. Also, the sarcopterygian lungfish has bradykinin [92].These results were obtained by incubating fish plasma with kallikrein or trypsin and the peptides were sequenced and compared with mammalian bradykinin.

In addition to the carp inhibitor [78] the only high molecular weight cysteine proteinase inhibitor from fish has been thyropin from chum salmon [79]. No reports of proteinase proregion -like cysteine proteinase inhibitors in fish have been observed.

1.3. Glycosylation of proteins

1.3.1. Protein glycosylation in general

Glycosylation is perhaps the most important post-translational modification of proteins. It is needed for proper folding and stabilization of proteins [93]. Some glycan structures are crucial either for the function or the proper location of proteins.

The N-glycosylation of proteins is an essential part in the quality control of polypeptide folding in endoplasmic reticulum [94]. Targeting of lysosomal enzymes is mediated by the specific glycan structure, phosphomannosyl marker, which is recognized by mannose-6-phosphate receptor in P-type lectins [95]. Glycan structures are also involved in many important recognition events, for example adhesion of leukocytes to endothelial cells by L-selectin [96]. Important as it is the glycosylation of proteins is not crucial on all occasions. There are enzymes which function normally after deglycosylation [97] and

some experiments show that recombinant proteins without their glycan part function properly [98]. However, in vivo their transport would be impaired.

Vertebrates have two main types of oligosaccharides; N-linked and O-linked. N-linked glycans are linked to the asparagine residue of a polypeptide chain in the consensus sequence Asn-X-Thr/Ser. O-linked glycans are linked to serine or threonine residues in the polypeptide chain. Figure 3 summarizes the most common structures of N- and O-glycans. [99].

Glycans can be modified in several ways to yield an extended complexity. Many of the explanations for complex glycan structures are not yet understood [100].

The terminal part of many glycans is sialic acid. The elementary units of sialic acids are N-acetylneuraminic acid (NeuNAc), N-glycolylneuraminic acid (Neu5Gc) and 2-

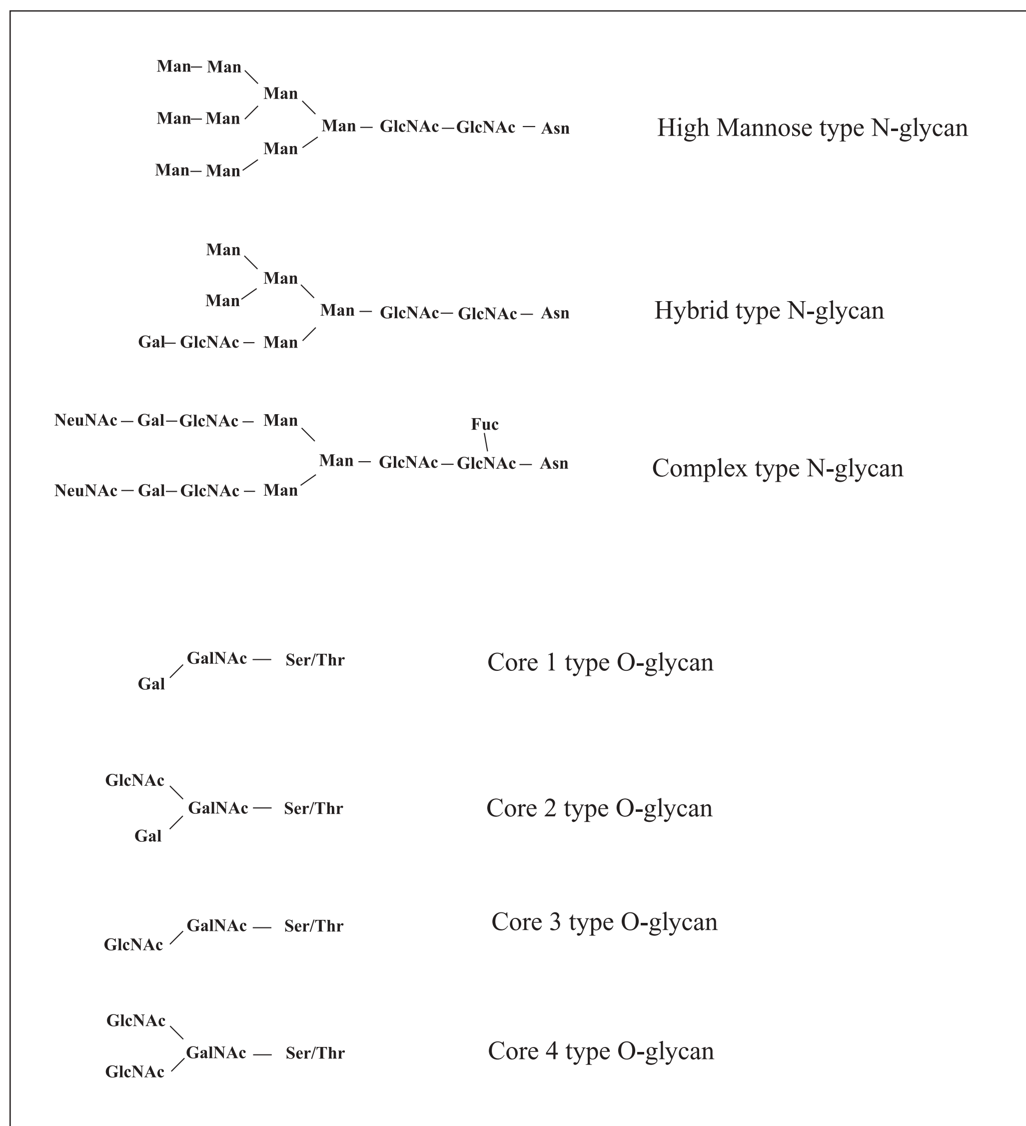

Figure 3. Examples of common structures of vertebrate N– and O-linked glycans.

keto-3-deoxy-nonulosonic acid (KDN) [101]. Sialic acids can be modified in many ways and over 40 modifications are known [102]. One of the most common modification is O-acetylation [102]. O-acetyl-groups can be attached to positions 7, 8, 9 and 4 of the sialic acid [101]. The function of O-acetylation is not completely understood but at least it has a protective function, as many sialic acid hydrolyzing enzymes (sialidases) are inhibited by O-acetylation [102]. It also inhibits the binding of certain viruses, but on the other hand O-acetylated sialic acids are also receptors for certain other viruses [101].

1.3.2. Glycosylation in fish proteins

The best characterised fish glycans originate from fish egg glycoproteins. Rainbow trout (*Oncorhynchus mykiss*) eggs were the second source where disialosyl groups were described on a glycoprotein [103], this structure had previously been shown to exist in rat brain [104, 105]. Several salmonid species were then reported to have even tetrasialylated O-glycans on their egg glycoproteins [106]. These proteins were named polysialoglycoproteins (PSGP) and even 50% of their weight consists of sialic acids [106]. O-acetylated sialic acid residues of PSGPs have been reported as well [107]. Fish eggs also have small glycophosphoproteins with sialylated N-glycans. Furthermore these glycans are found in the free form and they seem to be species specific [108]. Hyosophorin from the unfertilized eggs of dace (*Tribolodon hakonensis*) contain highly sulphated multiantennary N-glycans [109]. Rainbow trout ovarian fluid has tetra-antennary α2,3-disialylated poly-N-lactosamine chains with Lex and I antigens as terminal parts [110].

Antifreeze glycoproteins (AGFP) in polar fish have also been widely studied [111]. The most common glycan structure in AFGPs is Gal^{2}1,3GalNac, but the essential components for antifreeze activity have not been determined [112].

Atlantic salmon antithrombin has only three carbohydrate chains [113] whereas human antithrombin contains four carbohydrate chains. Interestingly, albumin from brown trout (*Salmo trutta*) is reported to be glycosylated but the albumin from Chinook salmon (*Oncorhynchus tshawytscha*) is not. As albumin in mammals is not glycosylated it is possible that the ancestral albumin was glycosylated [114]. Very few detailed structural studies on fish glycans are available.

1.4. Fish species

More than half of living vertebrate species are fishes. The systematic classification of fishes is largely based on anatomical investigations. Based on anatomical criteria the Atlantic salmon is classified as a rather primitive fish, Atlantic cod is more advanced and the spotted wolf fish is the most advanced of the three species. [115].

The Atlantic salmon (*Salmo salar* L.) is popular both as gamefish and in aquaculture. It is native in both the west and east Atlantic. It is anadromous species, which means that is lives in fresh water for the first 2-3 years before it migrates to sea. Smolts migrate to sea for 1-2 years before they return to fresh water to spawn. Before spawning they stop

eating. One difference between the Atlantic and Pacific salmon is that Pacific salmon dies soon after spawning. The subfamily *Salmoninae* contains seven genera including *Salmo, Salvelinus* and *Oncorhynchus*. [115, 37].

Atlantic cod (*Gadus morhua*) is a popular gamefish as well as important market fish. It can grow to nearly 2 m long and is omnivorous. Cod lives at bottom of the sea and its scales are covered with epidermis. Recently, the farming of cod has been under research and development. [116, 117].

The spotted wolf fish (*Anarchichas minor*) inhabits the North Atlantic and Barents Sea. It is found mainly in cold and deep waters. Sharp-teethed wolf fishes are popular gamefish and food. The farming of the wolf fish is under research and development. It is very resistant to most diseases and it has a leatherlike skin which can be used for clothes. [115, 117].

The nomenclature of fish species mentioned in this study can be found in Table 2.

Table 2. Fish names and classification according to Ref. 37.

Class	***Osteichthyes***	**Luukalat**
Order	***Anguilliformes***	**Ankeriaskalat**
Japanese eel	*Anguilla japonica*	Japanin ankerias
Order	***Salmoniformes***	**Lohikalat**
Atlantic salmon	*Salmo salar*	lohi
Arctic charr	*Salvelinus alpinus*	nieriä
Chum salmon	*Oncorhynchus keta*	koiralohi
Rainbow trout	*Oncorhynchus mykiss*	kirjolohi
Sockeye salmon	*Oncorhynchus nerka*	punalohi
Order	***Gadiformes***	**Turskakalat**
Atlantic cod	*Gadus morhua*	turska
Pacific cod	*Gadus macrocephalus*	Tyynenmeren turska
Hake	*Merluccius hubbsi* Marini	kummeliturska
Order	***Cypriniformes***	**Karppikalat**
Carp	*Cyprinus carpio*	karppi
Dace	*Tribolodon hakonensis*	seipi
Zebrafish	*Danio rerio*	seeprakala
Order	***Perciformes***	**Ahvenkalat**
Spotted wolf fish	*Anarhichas minor*	täplikäs merikissa
	Mujil auratus	mullo
Order	***Pleuronectiformes***	**Kampelakalat**
Turbot	*Scophthalmus maximus*	piikkikampela
Order	***Tetraodontiformes***	**Jäykkäleukakalat**
Fugu	*Fugu rubribes*	pallokala

2. AIMS OF THE STUDY

This study was started as a part of EU-project CT97-3508 "Fish cysteine proteinase inhibitors and infectious diseases". The objective of this study was to purify high molecular weight cysteine proteinase inhibitors from the skin of Atlantic salmon, Atlantic cod and spotted wolf fish for structural studies. It was suggested that the inhibitors might have a role in defence mechanisms of fish against pathogens. Therefore the purified inhibitors were biologically tested by other partners of the project. During the project the specific aims were formulated as:

- development of a purification method generally applicable for these inhibitor proteins
- structural and functional characterization of the proteins
- glycan analysis of the proteins
- cloning and sequencing of the gene of the previously unknown protein salarin.

3. MATERIALS AND METHODS

3.1. Fish materials

Fishes were obtained from Aquaculture station of Norwegian College of Fishery Science at Kårvika, Tromsø, Norway, except for the rainbow trout which was purchased from fish market in Finland.

Skin extracts from Atlantic salmon (*Salmo salar* L.), Arctic charr (*Salvelinus alpinus L.*), Atlantic cod (*Gadus morhua* L.), spotted wolf fish (*Anarhichas minor* Olafsen), rainbow trout (*Oncorhynchus mykiss* Walbaum) and turbot (*Scophthalmus maximus*) were used as the source for protein extraction. The skin (1 kg) was homogenized with Waring-Blender type homogenizer in 1 l of 10 mM Tris/HCl pH 7.4, 10 mM EDTA, 0.25 M sucrose, 0.1 mM phenylmethanesulfonyl fluoride, 5 mM benzamidine and 15 mM sodium azide. The homogenate was centrifuged at 6000 g for 30 min at 4 °C and the supernatant was collected. The supernatant was ultracentrifuged at 100 000 g for 2 h at 4 °C and the clear extract was collected underneath the floating fat. Skin extracts were kept in -20 °C until used.

Liver and skin from Atlantic salmon (*Salmo salar* L.), liver from Arctic charr and liver from spotted wolf fish (*Anarhichas minor* Olafsen) were used for RNA and DNA isolation. Organs were collected from freshly killed animals, frozen with liquid nitrogen and kept in -70 ºC until used.

3.2. Purification and characterization of the inhibitors

3.2.1. Chromatographic methods

A four-step purification scheme was developed for the purification of cysteine proteinase inhibitors from the fish skin extracts. Affinity chromatography of the inhibitors was performed by carboxymethyl-papain-Sepharose 4B (Pharmacia Biotech, Uppsala, Sweden) [118]. The affinity chromatography column was equilibrated with 20 mM sodium phosphate, pH 7,5 at a flow rate of 0.5 ml/min and the clear skin extract was applied to the column. The column was then washed with 2 column volumes of equilibration buffer and the non-specifically bound proteins were removed by washing with 1 column volume of the 1 M NaCl in equilibration buffer. Then the column was washed again with one column volume of the equilibration buffer and the bound proteins were finally eluted with 20 mM sodium phosphate, pH 11.8, 0.1 M NaCl, monitoring at 280 nm.

The inhibitory activity of the eluted fractions was determined by inhibition assay (see below) and the fractions expressing the inhibitory activity were pooled and concentrated with pressure microfiltration (Omegacell 10 K, Pall Filtron, Northborough, MA, USA). The concentrate (500 µl) was subjected to gel filtration on a Superdex 200 HR10/30 column (Pharmacia Biotech) equilibrated with 50 mM Tris/HCl pH 8.0, 300 mM NaCl at a flow rate of 0.5 ml/min. The gel filtration was monitored at 280 nm and the eluted fractions with inhibitory activity were pooled and diluted with water (1:1)

before anion-exchange chromatography. Anion-exchange chromatography was performed on a Mono Q HR 5/5 column (Pharmacia Biotech). The column was equilibrated with 50 mM Tris/HCl pH 8.0 and elution was performed with a linear salt gradient (0-0.5 M NaCl in 40 min) at a flow rate of 1.0 ml/min in the equilibration buffer and monitored at 280 nm.

Reversed-phase chromatography was performed at a flow rate of 1.0 ml/min on 0.46 x 4 cm, TSK TMS 250 (C1, TosoHaas Corporation, Japan) column equilibrated with 0.1 % trifluoroacetid acid. Elution was performed with a linear gradient of acetonitrile (15-60 % in 60 min) in 0.1 % trifluoroacetic acid and monitored at 214 nm. Analytical reversed-phase chromatography for purity and quantity estimations of proteins was performed as above but on a 0.21 x 10 cm TSK TMS 250 a column using a flow rate of 0.2 ml/min.

3.2.2. Inhibition assay

The inhibitory activity of chromatographic fractions was determined by inhibition assay with papain as an enzyme and BANA (N_{α}-benzoyl-DL-arginine-2-naphtylamide, Fluka Chemie AG, Buchs, Switzerland) as a substrate. 25 µl of papain solution (0.3 mg/ml, Sigma, St Louis, MO, USA) was incubated with 25 µl of chromatographic fractions (as the inhibitor) in 25 µl of assay buffer (0.2 M Tris/HCl pH 7.6, 8 mM EDTA, 4mM dithiotreitol) in a 96 well microtiter plate (Falcon, Becton-Dickinson) at room temperature for 10 min. Then 25 µl of BANA solution (5mM in 20 % methanol) was added and incubation continued for 10 min at 37 °C. The reaction was stopped by adding 100 µl of dimethylamino-benzaldehyde- reagent (500 mg 4-(dimethylamino)-benzaldehyde (Merck)) dissolved in 50 ml methanol and 50 ml 1 M Na-acetate pH 1.4). After 30 min at room temperature the developed colour was measured at 450 nm with a microwell reader (Multiskan® MCC/340, Labsystems, Espoo, Finland).

The inhibitory activity was defined as follows. The activity of non-inhibited papain (incubated without any inhibitors) was taken as 100 % activity and the activity of partially inhibited papain (incubated with the inhibitory fractions) was expressed relative to 100 % activity calculated comparing the absorbances of their reaction mixtures. Thus when the absorbance of the reaction mixture with non-inhibited papain was 0,899 and absorbance of the reaction mixture with papain and the inhibitor was 0,237, the inhibitory activity was calculated from the formula (0.899-0.237) / 0.899 * 100 = 73,6 so giving the inhibitory activity as 73.6 % meaning that 73.6 % of papain activity is inhibited.

All assays were made in duplicates and an established proteinase inhibitor leupeptin in serial dilutions was used as a control. This method is adapted from [46]. The inhibitory activity of the purified fish inhibitors was also tested by the same method using ficin, bromelain and trypsin as the proteinase.

3.2.3. SDS/PAGE and N-terminal sequencing

The purified proteins were separated by SDS/PAGE in 12 % gel and stained with Coomassie Brilliant Blue or with silver stain [119].

Protein and peptide sequencing was performed using a Procise 494 A sequencer.

For N-terminal sequence analysis the proteins were electroblotted from a SDS/PAGE gel on to a PVDF (polyvinylidene difluoride membrane ProBlott™, Perkin-Elmer Applied Biosystems Division) in 10 mM CAPS [3-(cyclohexylamino)propane-1-sulphonic acid (pH 11)/10 % (v/v) methanol] as described [120]. After staining with Coomassie Brilliant Blue (0.1 % in 1 % acetic acid/40 % methanol) the protein bands were cut out and loaded on the sequencer. Alternatively, purified proteins were directly adsorbed to the polyvinylidene difluoride membrane of a ProSorb™ (Perkin-Elmer, Applied Biosystems Division) device and loaded to the sequencer.

Purified peptides were sequenced on a glass fiber disc pretreated with 1.5 mg BioBrene™ (Perkin Elmer, Applied Biosystems Division, Foster City, CA, USA).

3.2.4. Enzymatic digestions and peptide separation

For in-gel digestion the proteins were separated by SDS/PAGE and stained with Coomassie Brilliant Blue (0.1 % in 0.5 % acetic acid/30 % methanol) and destained (30 % methanol). The proteins in the excised bands were alkylated with iodoacetamide and digested with trypsin (Sequencing Grade Modified Trypsin, Promega Corporation, Madison, WI, USA) or endoproteinase LysC (Wako Chemicals GmbH, Germany) as described [121]. For in-solution digestion the purified proteins were alkylated with 4-vinylpyridine and digested with trypsin or endoproteinase GluC (Sequencing Grade Endoproteinase GluC, Promega). For endoproteinase GluC digestion a 30 µg sample of alkylated protein was dissolved in 50 µl of 50 mM ammonium acetate, pH 4.3 and digested with 0.5 µg of the enzyme by incubation at 37 °C overnight [III]. Peptides were separated by narrow bore reversed-phase chromatography on a 0.1 x 15 cm Vydac C8 column (300 Å, 5 µm, LC-Packings, The Netherlands) using a SMART™ System (Pharmacia Biotech, Uppsala, Sweden). Elution was performed using linear gradients of acetonitrile (0-40% in 120 min, 40-100 % in 40 min) in 0.1% trifluoroacetic acid. The flow rate was 50 µl/min, chromatography was monitored at 214 nm and the peptides were collected automatically.

3.2.5. Mass spectrometry of proteins and peptides

Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI TOF MS) was performed in the positive-ion mode with a BIFLEX™ mass spectrometer (Bruker Daltonik, Bremen, Germany), equipped a 337 nm nitrogen laser. Protein samples were analysed in sinapic acid (Fluka) as the matrix as described [122]. Peptide samples were analysed by mixing 1 µl of sample with 1 µl of α-cyano-4-hydroxy-cinnamic acid matrix (saturated solution in 30 % acetonitrile, 0.1 % aqueous trifluoroacetic acid) and dried with a stream of air. External calibration was performed with angiotensin II (Sigma, A9525) and ACTH (adrenocorticotropic hormone) fragment 18-39 (Sigma, A0673).

3.2.6. Capillary isoelectric focusing

Capillary isoelectric focusing (IEF) was performed using a BioFocus™ 3000 Capillary Electrophoresis System (BioRad Laboratories, Hercules, CA, USA). Focusing was performed in Bio-Lyte Ampholyte (pH 3-10, BioRad), in an eCAP™ neutral capillary (50 µm x 50 cm, Beckman). BioMark™ CIEF markers (pI values 5.3, 6.4, 7.4, 8.4 and 10.4, BioRad) together with trypsin inhibitor (pI 4.6, Sigma T-1021) were used for internal calibration. UV detection was at 280 nm.

3.3. Glycosylation analysis

3.3.1. Preparation of glycans

N-glycans of the inhibitors were released by enzymatic hydrolysis with N-glycosidase F. A dry protein sample (approximately 30 µg) was dissolved in 100 µl of 20 mM sodium phosphate, pH 7.2, 0.1% SDS, 1% β-octylglycoside and 2 U of N-glycosidase F (Roche Biochemicals, Switzerland) was added. The reaction mixture was incubated at 37 °C for 48 h and then centrifuged in a ProSpin™ sample preparation cartridge (PerkinElmer Applied Biosystems Division). The de-N-glycosylated protein was adsorbed to polyvinylidene difluoride membrane and released N-glycans could be collected from the flow-trough and further purified from proteins and detergents by passing through a BondElut C18 extraction cartridge (Varian SPP, Harbor City, CA, USA). The eluted N-glycan pool was dried and desalted by gel filtration on a Superdex ®Peptide column in 50 mM NH_4HCO_3, monitoring at 205 nm and the glycan pool was dried in a vacuum centrifuge. The amount of N-glycans was estimated by their UV absorbance in gel filtration chromatography compared with external N-acetylglucosamine and N-acetylneuraminic acid standards [II].

O-glycans were isolated by subjecting the polyvinylidene difluoride membrane containing the de-N-glycosylated protein to reductive β-elimination. The membrane was incubated in 150 µl of 1 M $NaBH_4$ in 0.1 M NaOH for 48 h at 37 °C. The reaction was terminated by adding 5 µl acetic acid to the mixture. Borate was removed by repeated addition of 1% acetic acid in methanol and evaporation. The O-glycan pool was desalted by gel filtration as described above [123].

The major glycoproteins of the salmon skin extract were isolated by affinity chromatography on a 5 x 45 mm column of ConA sepharose (Amersham Pharmacia Biotech). Samples of skin extract were first run on a Bio-Gel P-6 DG column (10 x 100 mm) in 20 mM Tris/HCl, 250 mM NaCl, 1 mM CaCl2, 1 mM MgCl2 (ConA buffer) to remove saccharose present in the extract. Then aliquots of 1-3 ml were applied to the ConA column. Nonbinding proteins were removed by washing with five column volumes of the ConA buffer, and the N-glycosylated proteins were then eluted with 5 ml of 0.5 M α-methylmannoside in ConA buffer. Pooled glycoprotein fractions were concentrated

with pressure microfiltration (Omegacell 10 K) and separated on a C1 reversed phase chromatography column. This chromatography yielded a major peak carrying two glycoproteins, 80 kDa and 55 kDa, when analysed on SDS/PAGE. Approximately 40 µg of total protein was vacuum-dried and subjected to N-glycosidase F treatment and purified as above.

3.3.2. Permethylation

Dry samples were permethylated by dissolving them in 70 µl NaOH/DMSO suspension. After 1 h incubation at room temperature, 35 µl methyliodide was added and the mixture was incubated for 1 h in the dark and vigorously vortexed occasionally [124]. The permethylated glycans were purified by adsorption to POROS R3 material (Perseptive Biosystems, Framingham, MA). About 1 µl of POROS material was packed into an Eppendorf gel-loader tip and washed with water, then the permethylated glycans were applied to the tip in 20% aqueous methanol. The column was washed with 20 µl of 10% aqueous methanol and eluted with 5 µl of 50% aqueous methanol.

3.3.3. Removal of O-acetyl groups

O-acetyl groups were removed by mild acidic hydrolysis (saponification) of samples. Dry samples were dissolved in 0.1 M NaOH and incubated at 8 °C for 80 min. The reaction was terminated by adding one-half volume of 1 M acetic acid. An aliquot of the reaction mixture was drop-dialyzed for 30 min against water on VSWP 02500 membranes (Millipore, Bedford, MA) prior to MALDI-TOF MS analysis.

3.3.4. Periodate oxidation

Saponified N-glycans were incubated in 65 µl of 5 mM $NaIO_4$ on ice in the dark for 15 min, and the excess of periodate was destroyed by addition of 10 µl of 0.3 M ethylene glycol [125]. After 30 min of incubation, 50 µl of 1 M $NaBH_4$ in 0.2 M Na_2CO_3 was added and the mixture was incubated at room temperature for 4 hours longer. The reaction was terminated by adding 2 M acetic acid, and prior to MALDI-TOF MS analysis the N-glycans were desalted by gel filtration.

3.3.5. Gas-phase β-elimination

Glycopeptide sample was dried in a glass tube, and the tube was placed in a larger glass tube containing 300 µl of 40% aqueous methylamine [126]. The tube was evacuated until cool to touch and the sample was allowed to react with methylamine gas for 2.5 h at 55 °C. The deglycosylated peptide was dissolved in 50% aqueous acetonitrile and an aliquot was analysed in MALDI-TOF MS.

3.3.6. Exoglycosidase digestions

Digestions with NDV (Newcastle disease virus) neuraminidase and *S. pneumoniae* β1,4-galactosidase (Oxford Glycosciences, Abingdon, UK) were carried out in 10 µl of 50 mM sodium acetate, pH 5.5, at 37 °C. Digestion with jack bean β-galactosidase (Glyko, Novato, CA, USA) was carried out in 10 µl of 100 mM sodium citrate, pH 4.5, at 37 °C. Aliquots of 1 µl were removed after 16 h of digestion, drop-dialyzed against water, and analyzed by MALDI-TOF MS.

Pure oligosaccharide standards were used as controls for the glycosidase reactions: Neu5Acα2,3Galβ1,4(Fucα1,3)GlcNacβOMe for NDV neuraminidase, and Galβ1,4GlcNAcβ1,6[Galβ1,3GlcNAcβ1,3]Galβ1,4Glc for β1,4-galactosidase.

3.3.7. Mass spectrometry of glycans and glycopeptides

MALDI TOF MS was perfomed on a Biflex™ time-of-flight instrument (Bruker Daltonik, Bremen, Germany) equipped with a nitrogen laser operating at 337 nm.

Underivatized N-glycans were analysed in either positive ion delayed extraction reflector mode or negative ion delayed extraction linear mode, using 2,4,6-trihydroxyacetophenone (Fluka Chemie AG, Buchs, Switzerland) (3 mg/ml in acetonitrile/ 20 mM aqueous diammonium citrate, 1:1, by volume) as the matrix as described [127]. Permethylated samples were analysed in the positive ion delayed extraction reflector mode using 10 mg/ml DHB (Aldrich, Steinheim, Germany) in H_2O as the matrix.

Peptide samples were analysed using α-cyano-4-hydroxy-cinnamic acid matrix.

3.4. Molecular cloning and sequencing of the inhibitors

3.4.1. Isolation of RNA and DNA

RNA was isolated with QuickPrep™ mRNA purification kit (Amersham Pharmacia Biotech, Uppsala, Sweden) according to the manufacturer's instructions or by an established method [128]. The isolated RNA was reverse transcribed with a Ready-to-go T-primed First-Strand kit (Amersham Pharmacia Biotech) and the obtained cDNA was used as a template for polymerase chain reactions. Genomic DNA was isolated with QIAamp DNA MiniKit (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

3.4.2. Primers

All primers were synthesized in DNA synthesis laboratory of Institute of Biotechnology, University of Helsinki. In the beginning of the study several degenerate primers designed on the basis of obtained peptide sequences were used before one combination gave a positive result. Then the primers were designed on the basis of the DNA sequence. All of the used primers are listed in Table 3.

Table 3. Sequences and location of primers used in PCR of salarin from Atlantic salmon and acrtic charr.

Name	Sequence (5' → 3')	Direction	Position in AJ550228
N-term	GTNCAYAARGARTTYGARATGAC	+	
P14r	TGRTTCATNCCCATNGTRAA	-	
SAL1f	CTAAAGGACTTTTGCCCATGC	+	1545-1564
SAL2f	GCTTCAAGATGGGGAGGAAG	+	3247-3266
SAL3f	GTCTACAATCACAGGGAGAAG	+	1-21
SAL4f	AATGGATTGGAGAGTTACAC	+	1254-1273
SAL5f	GAAAGACCTATCCGTCCAC	+	3400-3418
SAL1r	CATGGTGTAACTCTCCAATC	-	1277-1258
SAL2r	GAACCTCCTCAGCAGTCAAG	-	3706-3687*
SAL3r	AAAATAGTTAATTCCTCATAG	-	4086-4066
SAL4r	TCCATGTTTCAAACTCCTTGTGC	-	3296-3274
GSPr1	TCCACCTCTGCTTCCTCCCCTCCATC	-	2697-2672
iSALr1	CAAGCTTTCACCTGGATTCA	- (intron)	619-599
iSALr2	AGGCCATATTGATATTGATACACT	- (intron)	308-285
PolyTN	AAGCAGTGGTATCAACGCAGAGTAC(T)$_{30}$(AGC)		
UPM	CTAATACGACTCACTATAGGGCAAGCAGTGGTAACAACGCAGAGT		

Y= C,T
R= A,G
N= A,C,G,T
*= mismatch in the 3' end

3.4.3. Polymerase chain reaction (PCR)

PCR was performed with 20 pmol of two primers, 1 µl of cDNA or genomic DNA, 8 µl of 1.25 mM dNTP mix, 5 µl buffer and water was added to 50 µl. Enzymes, DyNAzyme II DNA Polymerase (Finnzymes, Espoo, Finland) or Taq DNa polymerase (MBI Fermentas, Vilnius, Lithuania) 1 U/50 µl were added either before or after the initial denaturation step.

Amplification conditions were; initial denaturation 5 min at 95 °C followed by 30 cycles 1 min at 95 °C, 1 min at annealing temperature (see below) and 1 min at 72 °C and a final extension at 72 °C for 5 min.

Optimal annealing temperature varied between 51-70 °C depending on the primer pair used. The genomic PCR was done similarly except for a longer elongation time of 90 s and a longer final extension time of 10 min. Obtained PCR products were separated on 1.5% agarose gels, stained with ethidium bromide and visualised under UV light.

3.4.4. Rapid Amplification of cDNA Ends (RACE)

A SMART RACE cDNA amplification kit (Clontech Laboratories, Palo Alto, CA, USA) utilizes the principle of RACE to obtain full-length cDNA copies of mRNA transcripts [129]. This kit was used to determine the 5'-end of the salarin gene. A gene specific

primer (GSPr1, Table 3) was designed on the basis of the sequence already obtained. It was used together with the UPM primer from the RACE kit. Selected PCR products were ligated into pCR®2.1-TOPO vector and cloned with TOPO-TA cloning kit (Invitrogen). The colonies were screened using a direct colony PCR and universal forward (UP) and reverse (RP) primers [130]. Obtained products were analysed in agarose gel and those with inserts were used for sequencing.

3.4.5. DNA sequencing of the salarin gene

At least three colonies from the cloned PCR products were sequenced or sequencing was done directly on the obtained PCR products when applicable. Sequencing was done using BigDye terminator chemistry and the obtained reactions were analyzed on an ABI 3100 16 capillary sequencer (Applied Biosystems, CA, USA). The obtained sequences were edited and assembled using the Staden Package programs Trev and Gap [131] on a SUN workstation. Homology search from the Genebank sequences was performed by the BLAST program [132]. The obtained salarin sequences were submitted to the EMBL gene bank with the accession numbers AJ550228 for Atlantic salmon and AJ550229 for arctic charr.

4. RESULTS

4.1. Purification of cysteine proteinase inhibitors from fish skin

All cysteine proteinase inhibitors were purified following a similar four-step scheme: papain affinity chromatography, gel filtration chromatography, anion-exchange chromatography and finally reversed phase chromatography. This method was developed for purification of cysteine proteinase inhibitors from the skin of Atlantic salmon in study (I). An overview of this purification is presented in Figure 4.

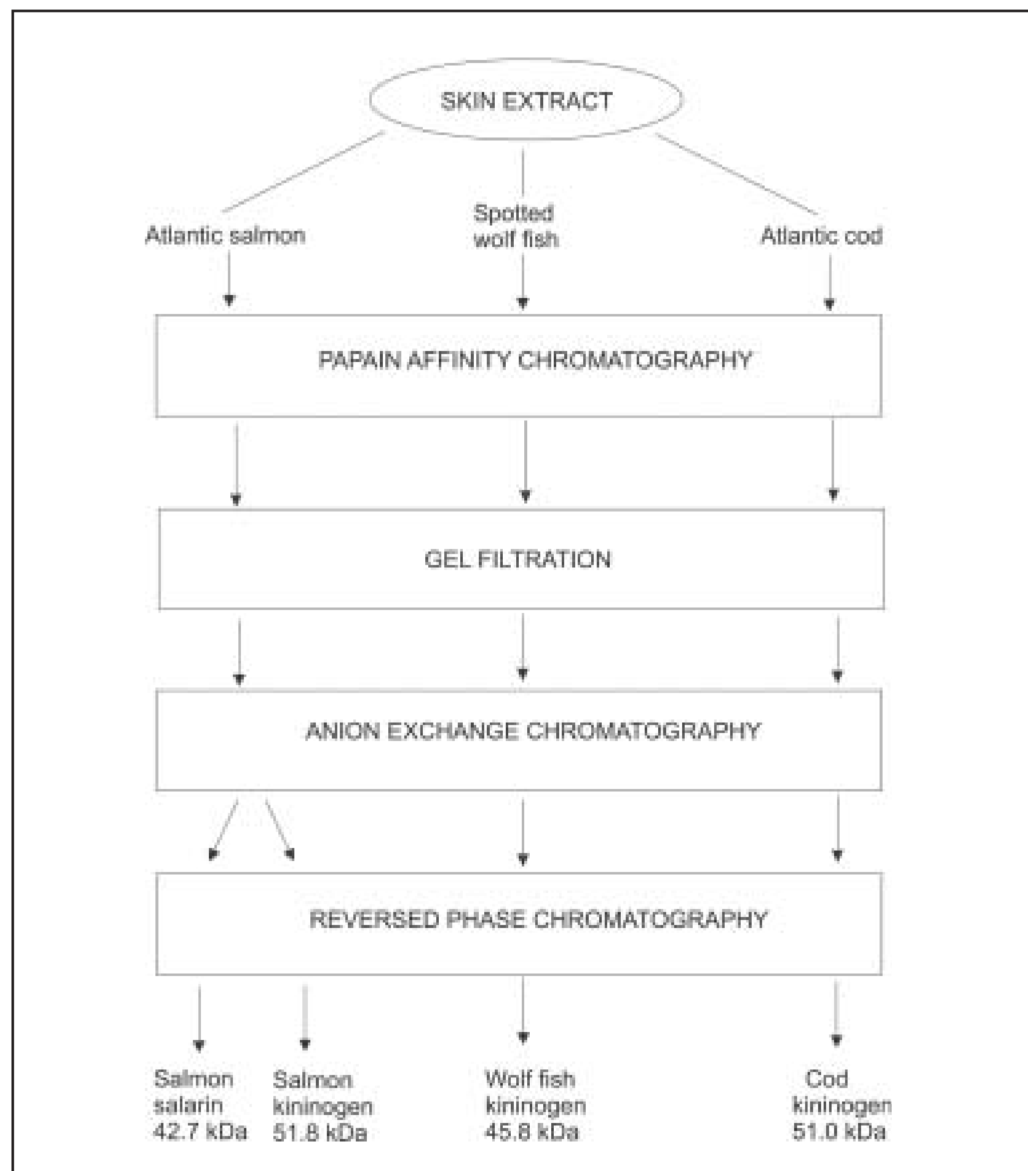

Figure 4. An overview of the purification method developed for the high molecular weight cysteine proteinase inhibitors from the skin of fish.

The first step in purification of cysteine proteinase inhibitors from the fish skin extract was affinity chromatography on immobilized papain. The proteins bound to papain-Sepharose were eluted and the fractions showing inhibitory activities over 70% were pooled, concentrated and further purified by gel filtration. The fractions from gel filtration exhibiting inhibitory activity were subjected to anion-exchange chromatography. With salmon skin extract the inhibitory activity was separated into two main fractions by anion-exchange chromatography. In further fractionation by reversed-phase chromatography salmon skin contained two cysteine proteinase inhibitors (I and II), which in SDS-PAGE migrated at positions relative to apparent molecular weights of 43 kDa and 76 kDa.

In contrast to the salmon skin only one fraction with inhibitory activity was obtained from cod skin extract on anion exchange chromatography. This fraction was further purified by reversed phase chromatography after which the main protein migrated in SDS-PAGE relative to apparent molecular weight of 78 kDa. Similarly treated wolf fish skin extract also gave one fraction with inhibitory activity in anion exchange chromatography, and after purification the major protein in this fraction showed an apparent molecular weight of 67 kDa in SDS-PAGE. These four proteins with cysteine proteinase inhibitory activity were further characterized.

4.2. Characterization of the inhibitors

4.2.1. Chemical and physical properties

The exact molecular masses of the purified inhibitors were determined by MALDI TOF mass spectrometry where the salmon inhibitor I gave a molecular mass of 42.7 kDa and salmon inhibitor II, a molecular mass of 51.8 kDa. In similar analysis cod inhibitor gave a molecular mass of 51.0 kDa and wolf fish inhibitor a molecular mass of 45.8 kDa. From the differences between SDS/PAGE and mass spectrometry we suspected that these proteins were glycosylated, this was later confirmed (see below). The isoelectric points of the purified inhibitors were determined by capillary isoelectric focusing. Salmon inhibitor I had a pI 5.1 and salmon inhibitor II has pI's 4.0, 4.2. and 4.6. Also, cod and wolf fish inhibitors had several isoelectric points (pI's 3.6, 3.9 and 4.4 for cod inhibitor and pI's 4.1, 4.3, 4.35 and 4.4 for wolf fish inhibitor). The obtained molecular masses, isoelectric points as well as other properties of the inhibitors are summarized in Table 4.

Table 4. A summary of the properties of the purified fish inhibitors (n.d. = not determined)

Fish	Inhibitor	Molecular weight (kDa) by SDS/PAGE	and MALDI-TOF MS	pI	specific inhibition
Atlantic salmon	salarin	43	42.7	5.1	2900 U/mg
Atlantic salmon	kininogen	76	51.8	4.0, 4.2, 4.6	2250 U/mg
Atlantic cod	kininogen	78	51.0	3.6, 3.9, 4.4	940 U/mg
spotted wolf fish	kininogen	67	45.8	4.1, 4.3, 4.35, 4.4	6200 U/mg
arctic charr	salarin	43	43.2	n.d.	n.d.
arctic charr	kininogen	75	n.d.	n.d.	n.d.
rainbow trout	salarin	43	43.6	n.d.	n.d.
rainbow trout	kininogen	78	51.2	n.d.	n.d.
turbot	kininogen	78	n.d.	n.d.	n.d.

4.2.2. Protein sequencing

All the purified inhibitors were subjected to N-terminal sequencing as well as enzymatic digestions followed by peptide separation and N-terminal sequencing of the peptides. The kininogens were also compared by peptide mass fingerprinting. All obtained sequences are presented in Table 5.

To investigate the possible homology of these inhibitors to other known proteins the obtained peptide sequences were compared with proteins in databanks. It was shown that salmon inhibitor I had a low similarity with some cysteine proteinases but it had no close match with any previously known protein so we named it salarin. Salmon inhibitor II had peptide sequences matching with kininogens, so we named it salmon kininogen. Cod and wolf fish inhibitors had sequences matching with kininogens too so we named them cod kininogen and wolf fish kininogen, respectively. The peptide sequences of the three kininogens are aligned with human kininogen in Fig 5.

```
1 KNH_HUMAN   MKLITILFLC SRLLLSLTQE SQSEEIDCND KDLFKAVDAA LKKYNSQNQS
2 SAL_KIN                        ELE PEQVLAFCDD KDVEAAVFLA LVKYN  LPY
3 COD KIN                        RHE VPQANLECDE GAMDLK            IST
4 WF KIN                         XLV QPGVLIFCDD PS            YNSMSDS

1 KNH_HUMAN   NNQFVLYRIT EATKTVGSDT FYSFKYEIKE GDCPVQSGKT WQDCEYKDAA
2 SAL_KIN     GNQLALYQIL ESSK                           V WRDCDYLPTG
3 COD KIN     GNMVALYQIL SASK                D SDCPAGGAVT WTD
4 WF KIN      THLFTLHFVX Y  SENGSDS VYSLQFTSR  SDCPAGSNKP WTECDYLSYE

1 KNH_HUMAN   KAATGECTAT VGKRSSTKFS VATQTCQITP AEGPVVTAQY DCLGCVHPIS
2 SAL_KIN     NK                    VLAVFCDP VEAPVVAERT TCLGCPREID
3 COD KIN
4 WF KIN      RR

1 KNH_HUMAN   TQSPDLEPIL RHGIQYFNNN TQHSSLFMLN EVKRAQRQVV AGLNFRITYS
2 SAL_KIN     VEG    DPL TYSITRFNAD SDSSHHFILN
3 COD KIN                                            QVV AGLR
4 WF KIN                        FNER LSTGHK

1 KNH_HUMAN   IVQTNCSKEN FLFLTPDCKS LWNGDTGECT DNAYIDIQLR IASFSQNCDI
2 SAL_KIN                       XFVP LTVLEPGQEK
3 COD KIN
4 WF KIN                                                          QVY

1 KNH_HUMAN   YPGKDFVQPP TKICVGCPRD IPTNSPELEE TLTHTITKLN AENNATFYFK
2 SAL_KIN
3 COD KIN                 ICMGCPVE LDLESEELKV PVAVSISK
4 WF KIN      CLLDDVIIPE KAPCLGCPME VDENSEDLKF PLSVSISK

1 KNH_HUMAN   IDNVKKARVQ VVAGKKYFID FVARETTCSK ESNEELTESC ETKKLGQSLD
2 SAL_KIN                                          ELNDEC HPDPDVELAH
3 COD KIN
4 WF KIN

1 KNH_HUMAN   CNAEVYVVPW EKKIYPTVNC QPLGMISLMK RPPGFSPFR
2 SAL_KIN     CNSTVDVAPW R                     RPPGWSPLR
2 COD KIN                                    R RPPGWSPLR
3 WF KIN                           A GALPTMFTRR RPPGWSPLR
```

Figure 5. Comparison of amino acid sequences of kininogens. The determined peptide sequences from salmon kininogen (`2 SAL KIN`), cod kininogen (`3 COD KIN`) and wolffish kininogen (`4 WF KIN`) are aligned with the heavy chain of human kininogen (`1 KNH_HUMAN`) sequence.

Table 5a. N-terminal and peptide sequences determined from salarin. X = not identified amino acid.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
N-terminal	K	P	L	P	E	D	S	E	A	G	V	H	K	E	F	E	M	T	K	V	K	Y	G	K	X	Y	P				
1	L	M	V	V	F	P	X	R	D	G	G	E	E	A	E	V	D	K	E	F	E	T	W	K							
2	G	L	L	P	M	P	R	P	E	E	E	E	V	D	K	E	F	E	M	W	K										
3	T	T	A	E	V	T	G	R	R	L	Q	F	G	E	E	A	E	V	H	K											
4	T	T	A	E	V	T	G	R	R	L	Q	D	G	E	E	A	E	V	H	K											
5	T	Y	P	S	T	V	E	E	A	K																					
6	S	Y	P	S	T	E	E	E	A	K																					
7	E	M	W	L	A	T	R	K																							
8	E	I	W	L	A	T	R	K																							
9	M	V	M	E	H	N	K	R																							
10	V	M	E	H	N	T	R	A	G	N	G	L	E	S	Y	T	M	A	V	N	H	L	A	D	T						
11	A	G	N	G	L	E	S	Y	T	M	A	V	N	H	L	A	D	L	T	T	E	E	V	P	K						
12	A	E	X	G	S	E	S	F	T	M	G	I	N	Y	F	S	D	M	T	Q	E	E	I	P	K						
13	A	E	N	G	L	E	S	F	T	M	G	V	N	H	F	A	D	L	T	A	E	E	V	P	R						
14	R	A	E	T	G	S	E	S	F	T	M	G	M	N	H	L	S	D	K												

Table 5b. N-terminal and peptide sequences determined from salmon kininogen. X = not identified amino acid.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
N-terminal	no result																														
1	V	P	R	P	C	K																									
2	A	T	V	H	M	S	E	T	D	K																					
3	V	W	R	D	C	D	Y	L	P	T	G	N	K																		
4	L	P	Y	G	N	Q	L	A	L	Y	Q	I	L	E	S	S	K														
5	E	L	N	D	E	C	H	P	D	P	D	V	E	L	A	H	C	N	S	T	V	D	V	A	P	W	R				
6	E	L	N	D	E	C	H	P	D	P	D	V	E	L	A	H	C	X	S	T	V	D	V	A	P	W	R				
7	D	P	L	T	Y	S	I	T	R	F	N	A	D	S	D	S	S	H	H	F	I	L	N								
8	E	L	E	P	E	Q	V	L	A	F	C	D	D	K	D	V	E	A	A	V	D	L	A	L	V	K					
9	V	L	A	V	F	C	D	P	V	E	A	P	V	V	A	E	R	T	T	C	L	G	C	P	R	E	I	D	V	E	G
10	C	D	P	V	E	A	P	V	V	A	E	R	T	T	C	L	G	C	P	R	E	I	D	V	V						
11	A	Q	X	D	S	C	T	Q	Y	F	V	E	F	N	S	R															
12	X	F	V	P	L	T	V	L	E	P	G	Q	E	K																	
13	H	E	T	A	E	A	N	V	E	C	A	P	G	P	L	D	N	F	D	V	F	R									
14	C	A	P	G	P	L	D	N	F	P	V	F	R	R	R	P	P														
15	R	R	P	P	G	W	S	P	L	R																					

Table 5c. N-terminal and peptide sequences determined from cod kininogen.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
N-terminal	no result																														
1	A	A	V	T	S	F	N	E	K																						
2	E	F	S	P	P	A	P	P	S	R																					
3	H	E	V	P	Q	A	N	L	E	C	D	E	G	A	M	D	L	K													
4	R	R	P	P	G	W	S	P	L	R																					
5	I	C	M	G	C	P	V	E	L	D	L	E	S	E	E	L	K	V	P	V	A	V	S	I	S	K					
6	I	S	T	G	N	M	V	A	L	Y	Q	I	L	S	A	S	K														
7	Q	V	V	A	G	L	R																								
8	D	S	D	C	P	A	G	G	A	V	T	W	T	D																	
9	Y	S	L	Y	F	D	M	R	K																						

Table 5d. N-terminal and peptide sequences determined from wolf fish kininogen. X = not identified amino acid.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
N-terminal	no result																														
1	X	L	V	Q	P	G	V	L	I	F	C	D	D	P	S																
2	K	A	P	C	L	G	C	P	M	E	V	D	E	N	S	E	D	L	K												
3	Q	V	Y	C	L	L	D	D	V	I	I	P	E	K																	
4	R	K	K	P	I	S	C	N	A	T	V	Y	M	T	E	T	E	A	D	T	K										
5	A	G	A	L	P	T	M	F	T	R	R	R	P	P	G	W	S	P	L	R											
6	S	E	X	G	S	D	S	V	Y	S	L	Q	F	T	S	R															
7	Q	V	Y	C	L	L	D	D	V	I	I	P	E	K																	
8	D	L	N	D	L	C	V	P	D	D	Q	N	A	G	Y	A	N	C	X	S	T	V	N	V	A						
9	S	D	C	P	A	G	S	N	K	P	W	T	E	C	D	Y	L	S	Y	E	R	R									
10	Y	N	S	M	S	D	S	T	H	L	F	T	L	H	F	V	X	Y													
11	F	N	E	R	L	S	T	G	H	K																					

4.2.3. Functional properties

The inhibitory activity of purified inhibitors was determined against papain, ficin, bromelain and trypsin by inhibition assay. All inhibitors were found to inhibit papain and ficin (Fig 6.) but had no effect on trypsin (serine proteinase) and very weak effect on bromelain (inhibition of 0-20%). The specific activities against papain were calculated from the inhibition curves and are presented in Table 4.

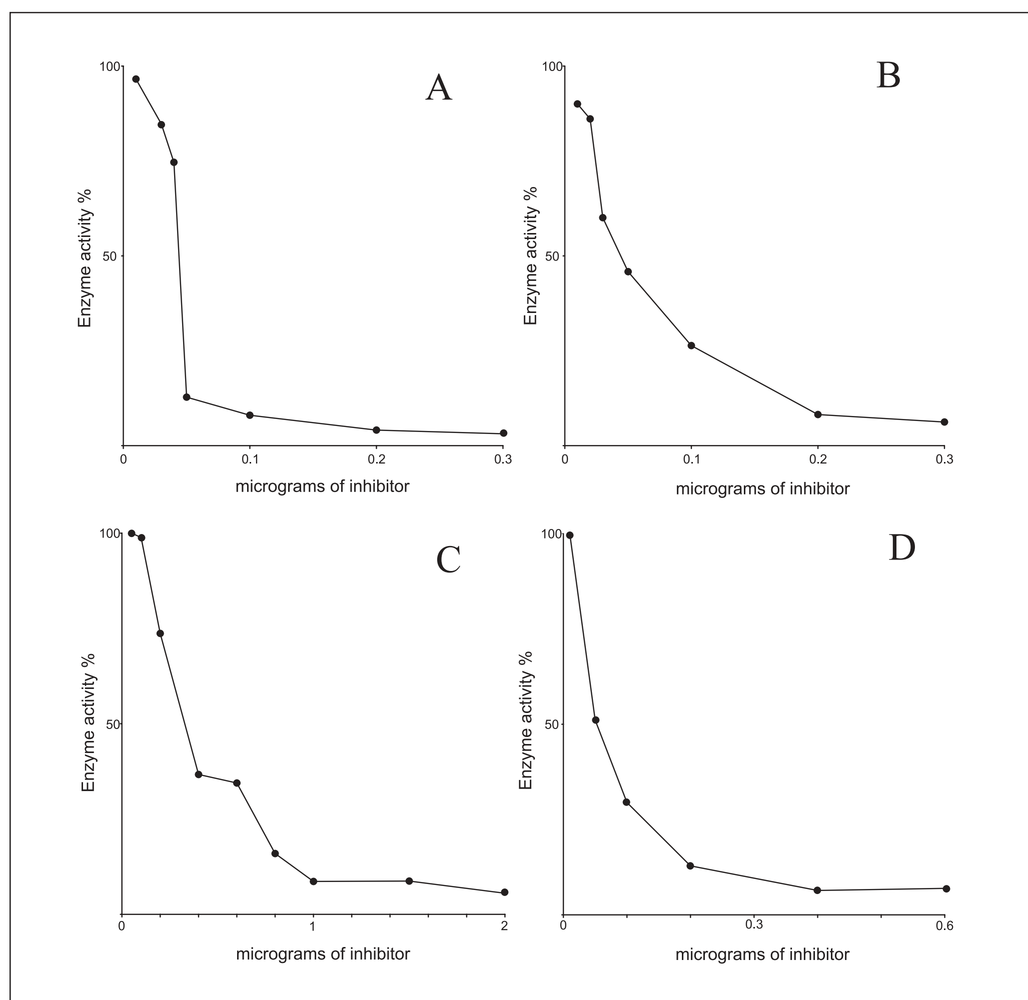

Figure 6. The inhibitory activities of purified cysteine proteinase inhibitors from fish skin. Inhibition of 0.25 µg of papain by different amounts of A) salarin, B) salmon kininogen, C) cod kininogen and D) wolf fish kininogen.

4.3. Analysis of glycan structures

Glycosylation of inhibitors was suspected from the difference between molecular weight determined by SDS/PAGE and by MALDI TOF mass spectrometry (Table 4) and it was confirmed by digestion with N-glycosidase F. To release N-glycans for more detailed characterization all inhibitors were treated with N-glycosidase F.

Approximately 1 nmol of N-glycans were released from 0.6 nmol of salmon kininogen, thus implying two N-glycosylation sites. The MALDI TOF spectrum of salmon kininogen N-glycans showed three clusters of ions with intervals of 42 Da (Fig 7A). This implied

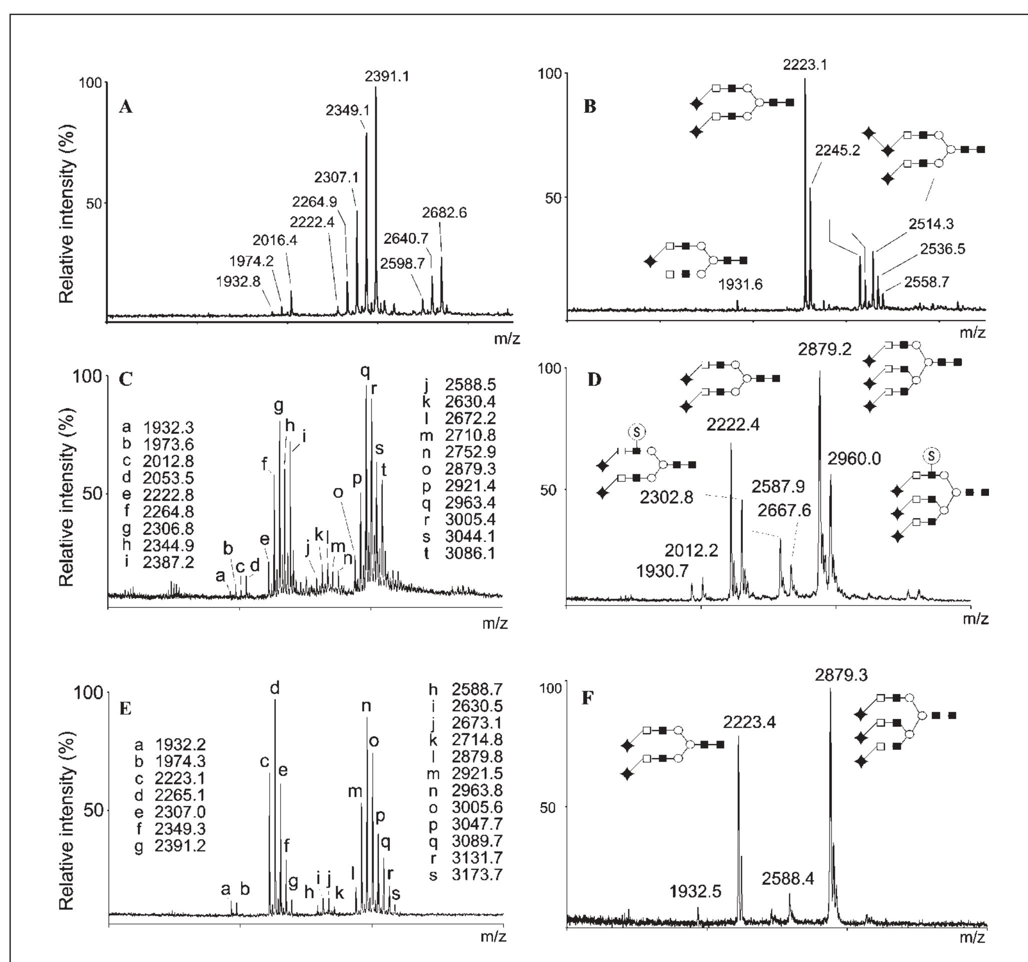

Figure 7. MALDI-TOF spectra of N-glycans isolated from inhibitors. A) Intact glycans from salmon kininogen. B) Salmon kininogen N-glycans after saponification. C) Intact glycans from cod kininogen. D) Cod kininogen N-glycans after saponification. E) Intact glycans from wolf fish kininogen. F) Wolf fish kininogen glycans after saponification. The proposed glycan structures are shown: ■, N-acetylglucosamine, ○, mannose, □, galactose, ◆, N-acetylneuraminic acid, circled S, sulphate. All spectra were recorded in linear negative ion mode, signals are $[M-H]^-$ ions and average mass values are shown.

variable degree of acetylation in the glycan structures, most probably O-acetylation of sialic acids [133].When relative intensities of the signals are used to the estimate molar ratio of the components [134], about 60% of sialic acids seem to be di-O-acetylated.

The tentative N-glycan structure was verified by removal of O-acetylation by short alkaline hydrolysis (saponification) [135] and subsequent digestions of saponified glycans with Newcastle disease virus (NDV) neuraminidase (specific only to α 2,3- and α 2,8-linked sialic acids) and with *S. pneumoniae* β1,4-galactosidase (specific only to β1,4-galactose). After saponification the MALDI TOF MS spectrum showed one major signal (m/z 2223.1 $[M-H]^-$) appropriate for disialylated biantennary N-glycan. The glycan structure is drawn in Fig 7B. Trisialylated N-glycan is also observed in this spectrum (m/z 2514.3 $[M-H]^-$). When permethylated trisialylated N-glycans were analyzed with ESI MS they were shown to have a disialosyl (α2,8) group.

A sample of salarin (30 µg, approx. 0.7 nmol) was treated by the same methods as above, yielding 0.5 nmol of N-glycans which indicated there was only one N-glycosylation site. When the above mentioned structural studies were performed the results were practically identical with salmon kininogen N-glycans.

A MALDI TOF MS spectrum of isolated Atlantic cod N-glycans revealed two major clusters with adducts of + 42 and +80 Da, so in addition to probable acetylation observed in salmon inhibitors there was another adduct, possibly phosphorylation or sulfation (Fig 7C). Alkaline phosphatase treatment did not remove the +80 Da adduct so it was presumed to be a sulphate (SO_3) group.

After saponification and subsequent digestions with NDV sialidase, *S. pneumoniae* β1,4-galactosidase (known to be unable to hydrolyze galactose from Galβ1-4GlcNAc-R type structures where either Gal or GlcNAc is sulphated [136]) and with jack bean β-galactosidase (inhibited by sulphate on Gal unit only) the cod N-glycan structures were verified as bi- and triantennary N-glycans terminated by highly O-acetylated α2,3-sialic acids. About 1/3 of the glycans carried sulphate at GlcNAc. The structures are presented at Fig 7D.

The isolated N-glycans of the wolf fish kininogen exhibited a similar complicated pattern in MALDI-TOF MS (Fig. 7E), with two major clusters of ions. Adducts of + 42 Da are again evident in the spectrum, implying extensive O-acetylation of sialic acids but no adducts of +80 Da are observed. Again the subsequent saponification and digestions with NDV sialidase and *S. pneumoniae* β1,4-galactosidase confirmed the N-glycan structures of wolf fish kininogen as α 2,3-sialylated bi- and triantennary N-glycans, with extensive sialic acid O-acetylation (Fig 7F).

Two glycopeptides were isolated from both salmon kininogen and salarin and their amino acid sequences were determined. N-terminal sequencing of the salmon kininogen glycopeptides showed that both peptides had N-glycosylation sites (Appendix B. Table 1b.). One glycopeptide of salarin was stable against N-glycosidase F treatment thus implying O-glycosylation which was confirmed by N-terminal sequencing (Appendix B Table 1a)

O-glycosidically bound glycans in salmon kininogen and salarin were liberated by alkaline reductive β-elimination. The isolated, permethylated O-glycans from both

inhibitors yielded a MALDI TOF spectrum suggesting a composition $(Neu5Ac)_2(Hex)_1(HexNAc\text{-}ol)_1$, a disialylated core type-1 O-glycan (Figure 8). With fragmentation analysis by MS/MS it was shown that two sialic acids do not form a diNeu5Ac unit as in N-glycans.

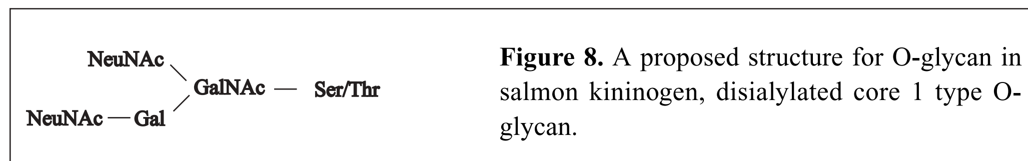

Figure 8. A proposed structure for O-glycan in salmon kininogen, disialylated core 1 type O-glycan.

O-glycosidic glycans were also recovered from cod and wolffish kininogen. The permethylated O-glycan pool was analyzed by MALDI-TOF MS, and the spectrum revealed the same major $[M+Na]^+$ signal at m/z 1256.63 matching with O-glycan structure in salmon kininogen and salarin.

Two major glycoproteins from salmon skin were isolated and their N-glycan structures were isolated by a similar method and analyzed by MALDI TOF MS. Their glycan spectrum was very similar to those obtained from salmon kininogen and salarin, indicating that extensive O-acetylation of sialic acids is also found in other glycoproteins than inhibitors.

4.4. Cysteine proteinase inhibitors in other fish species

Since salarin was found only in Atlantic salmon but not in cod or wolf fish the question about its existence in other fish species arose. Thus the skin extracts of a few other fish species were investigated using the purification method developed for the salmon inhibitors. The skin extract of Arctic charr was treated by the same method as above. In anion exchange chromatography the inhibitory activity was separated into two main fractions which were then further purified by reversed phase chromatography. When they were analysed by SDS/PAGE they migrated to positions according to molecular weights of 43 kDa and 70 kDa. When the purified proteins were N-terminally sequenced the N-terminus of 43 kDa protein was identical with salarin and the N-terminus of 70 kDa protein was blocked. The proteins were then enzymatically digested and a peptide mass fingerprint was obtained thus confirming them to be charr salarin and charr kininogen.

Rainbow trout skin extract was treated similarly and it yielded two inhibitors with molecular masses of 43.5 kDa and 51.2 kDa. The N-terminus of the 43 kDa protein was identical to that of salarin. In SDS/PAGE and in MALDI TOF mass spectrometry the salarin also showed a smaller fragment of 34.8 kDa which had an identical N-terminus.

When the skin extract of turbot was analysed with a similar method it produced only one inhibitor with an apparent molecular weight of 70 kDa. It was presumed to be a kininogen, but no further experiments were performed to verify this assumption.

These results are summarized in Table 4.

4.5. Molecular cloning and sequencing of the salarin cDNA

Salarin cDNA sequence was obtained from salmon liver by RT-PCR with degenerate primers which were designed on the basis of the peptide sequences. Several degenerate primer pairs were tried before two degenerate primers, N-term and P14r (Table 3) corresponding to the N-terminus and peptide P14 of salarin (Table 5a), yielded a successful product. After this first salarin DNA stretch was sequenced it was possible to design gene specific primers for PCR and thus obtain the whole salarin coding sequence of 1026 bp. The predicted amino acid sequence of salarin has a putative 19 amino acid signal sequence and a 323 amino acid mature protein. The protein sequence consists of four almost identical domains differing by only a few amino acids (Fig. 9). The previously obtained peptide sequences matched almost precisely with the predicted amino acid sequence.

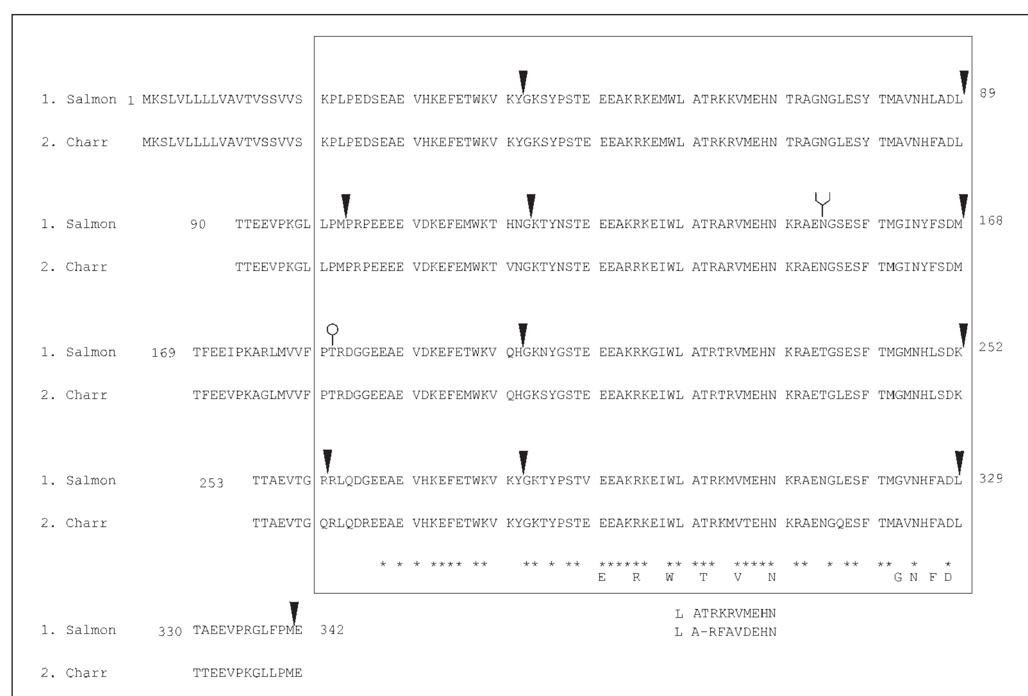

Figure 9. The translated amino acid sequences of salarin from Atlantic salmon (line 1) and arctic charr (line 2). Repeated amino acid sequences are outlined in square box. The N-glycosylation site is marked with Y and the O-glycosylation site is marked with ♀. Black triangles show intron sites. Stars denote identical amino acids in all four domains of salmon salarin. Underneath the stars is depicted the cathepsin L –proregion motifs ERFNIN and GNDF with modifications. Also phytocystatin motif LARFAVDEHN can also be found in the salarin sequence.

PCR using cDNA made from salmon skin as template did not result in any products. This suggests that salarin is not transcribed in skin but transported there.

With the charr liver cDNA as a template for PCR the salmon salarin primers resulted in similar products as in salmon. The charr gene has an identical signal sequence with the salmon gene and the amino acid sequence has the same four domains and only 19 amino acids differ from the salmon protein (Fig. 9).

We also isolated RNA from the liver of spotted wolf fish and made cDNA from it. PCR amplifications with the primer pairs SAL 1f +SAL 2r and SAL 3f +SAL1r were negative. This confirmed our previous negative result for the obtained protein data from wolf fish.

Genomic DNA was isolated from Atlantic salmon liver. To evaluate possible introns in the gene different primer combinations were tried in PCR to amplify the genomic sequence. The products were sequenced and combined thus yielding a complete salmon salarin gene. The salarin gene altogether contained 12 introns. All the intron borders followed the rule GT/AG except for the last intron (#12) which was GT/CA. It is interesting that intron 1 is located only 15 bp from the ATG start codon and contains two GT repeats and the last exon (13) contains only one amino acid from the coding gene. Allelic variation in the GT repeat was detected as part of the clones contained variable length GT. Furthermore, two different sequences were obtained differing by five bases in intron 11 indicating allelic variation in intron length.

A picture of the salarin gene structure is presented in Fig. 10.

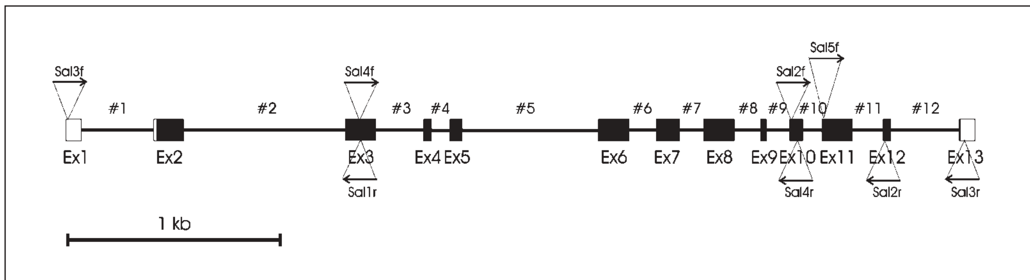

Figure 10. A schematic picture of the salarin gene. Primers are shown with arrows numbered according to Table 3. Boxes indicate exons, white boxes UTRs and black boxes coding areas. Introns are marked with # and a running number.

5. DISCUSSION

5.1. Cysteine proteinase inhibitors from fish skin

In this study four high molecular weight cysteine proteinase inhibitors from three different fish species were purified and characterized. Further, five similar inhibitors were shown to exist in three other fish species. Three of the purified inhibitors were kininogens which have not been isolated from any fish species prior to these studies. Kininogens were found from all six fish species studied. The fourth of the purified inhibitors was named salarin, this was found only in salmonid species.

All three kininogens were N-terminally blocked and a smaller fragment with open N-terminus copurified with the full size protein. The smaller fragment might be a differentially processed form of kininogen. Most of the peptide sequences obtained from fish kininogens matched with the human kininogen heavy chain which is conserved among different kininogens (Fig. 5a). Those peptides which did not match with the kininogen heavy chain may originate from the light chain or represent a more variable region in the heavy chain. Bradykinin peptide was found from each fish kininogen thus fortifying their identification. The fish bradykinin sequence differed from that of human one by two amino acids (Fig 5B).

All kininogens inhibited cysteine proteinases like papain and ficin but their specific inhibitory activities against papain varied from 940 U/mg of cod kininogen to 6200 U/mg of wolf fish kininogen, where U is defined as µg of papain inhibited per mg of inhibitor (Table 3). Interestingly, salmon cathepsin B was not at all inhibited by kininogens at test levels of 76 nM salmon kininogen and 300 nM wolf fish kininogen. On the other hand, cathepsin L from cod, salmon and wolf fish were inhibited by salmon kininogen and wolf fish kininogen (Dr. Ekkehard Weber, personal communication). The inhibition of cathepsin B by cystatins is reported to be weaker than that of papain or cathepsins H and L [51].

The antibacterial effects of kininogens and salarin were studied by comparing the inhibition zones around filter discs immersed into inhibitor solution and then placed onto the bacteria plate [42]. The antiviral effects were studied by fish cell cultures grown with inhibitors and then inoculated with viruses [42]. It has been shown that low molecular weight cystatins are able to inhibit the growth of bacteria [137]. In contrast to previous reports, the inhibitors had no effect on viral or bacterial growth in this study. However, it is possible that the inhibitors studied here are too large to act as antibiotics in these assays. Kininogens and salarin did not inhibit a virus-induced apoptosis in fish cell lines either [138]. Again, one reason for the negative results may be the size of kininogens and salarin as 11 kDa acid cysteine proteinase inhibitor (ACPI) is reported to inhibit rhabdovirus induced apoptosis in fish cell line [139]. It is also possible that the protective mechanism of high molecular weight inhibitors is mediated inhibiting the penetrative action of bacterial proteinases rather than inhibiting the bacterial growth on the whole.

5.2. Salarin

Salarin from Atlantic salmon is a 42.7 kDa glycoprotein which has four similar domains. A 30 kDa fragment with identical N-terminus copurified with the 42.7 kDa protein. Salarins theoretical mass calculated from the amino acid sequence is 37.5 kDa thus leaving 5.2 kDa for glycans which appears reasonable. When compared with other known cysteine proteinase inhibitors salarin has very little structural similarity. However, phytocystatins have a LARFAVDEHN motif [55] which is partially present in salarin (Fig 9.). As this motif is not found in other cystatins, it can not be essential to their inhibitory activity.

Proregions of cathepsin L –like proteinases have an interspersed ERFNIN –motif which is found in salarin too. Also, a short conserved sequence GNFD from cathepsin proregion [11] is present in the salarin sequence (Fig. 9). One function of the cathepsin proregion is to inhibit the proteolytic activity of cathepsin before entering the correct subcellular compartment [9]. Autonomous proregion -like inhibitors are also known [77].

A database search with salarin sequence results in one close match, AF281331, a 115 amino acid cathepsin from rainbow trout (*O. mykiss*). This sequence was identified by the closest hit from a BLAST search and no other tests were performed [36]. It is possible that the similarity was observed on the proregion of cathepsins, thus misguiding the authors to think that they had sequenced a cathepsin when they actually had sequenced a cathepsin L proregion-like inhibitor.

This idea is also supported by the fact that salarin effectively inhibited cathepsin L from salmon, cod and wolf fish but it had only little effect on cathepsin H and no effect on cathepsin B of salmon (Weber, personal communication). As cathepsin proregions are known to inhibit most effectively their cognate enzymes it is very likely that salarin is a cathepsin L proregion -like cysteine proteinase inhibitor.

The appearance of salarin in our studies was limited only to salmonid species, this was unexpected. It may be that the size of salarin has decreased during evolution, although then it should have been found also from wolf fish cDNA. However, the existence of a salmonid specific cysteine proteinase inhibitor appears quite unlikely.

Equally surprising was the number of introns in the salarin gene. The sequencing of the complete genome of pufferfish (*Fugu rubribes*), the first fish genome, was recently reported [140]. Compared to the human genome it is only one-ninth of size, but contains the same amount of genes. The pufferfish genome has shorter introns and a smaller amount of repetitive DNA [141]. Comparison of the salarin sequence against zebrafish genome resources (NCBI) identifies cathepsin L as a highly similar sequence.

5.3. Glycosylation of cysteine proteinase inhibitors

We isolated and characterized the glycan structures from the purified cysteine proteinase inhibitors. The salmon kininogen and salarin were shown to have biantennary N-glycans with mono-, di- or tri-O-acetylated α2,3-linked sialic acids. Trisialylated N-glycan carries an α2,8-disialosyl group with up to four O-acetyl groups. The O-glycans from salmon

kininogen and salarin were mostly disialylated type-1 core glycan with no indication of a disialosyl group (Fig. 8).

Wolf fish kininogen carries an α2,3-sialylated biantennary and triantennary N-glycans and disialylated core type-1 O-glycans. Cod kininogen similarly carries O-acetylated α2,3-sialylated biantennary and triantennary N-glycans but about 1/3 of them are sulphated at their antennal N-acetylglucosamine units. Also the cod kininogen O-glycan had a type-1 core with two sialic acids.

We also studied the major N-glycoproteins, transferrin and an unidentified 55 kDa glycoprotein, from the skin of Atlantic salmon. Their glycosylation was very similar to that of salmon kininogen and salarin with a high degree of O-acetylation.

O-acetylation of sialic acids has been described before in fish. The polysialoglycoprotein from the unfertilized eggs of the kokanee salmon (*Oncorhynchus nerka adonis*) contains N-glycolylneuraminic acid (Neu5Gc) with 4-, 7- and 9-O-acetylation and 9-O-acetylated ketodeoxynononic acid (KDN) [107].

These glycan structures, sialylated biantennary and triantennary N-glycans, are common vertebrate glycans. The high amount of O-acetylation in sialic acids was however unexpected. Practically all terminal sialic acid residues were O-acetylated. To the best of our knowledge, these represent by far the most extensively O-acetylated N-glycans characterized from any species to date. One reason for the scarce reports of O-acetylation might be that they are easily destroyed by standard glycan isolation procedures [142].

O-acetylated polysialic acid chains are resistant to bacterial sialidases specific for hydrolysing this kind of α2,3-linkages [107]. The same phenomenon is also observed in O-acetylated sialoglycoproteins [135]. Albeit, the biological function of O-acetylation is not fully understood, it has been proposed that is serves as a protective agent, as O-acetylation inhibits the binding of several pathogens like influenza A and B viruses [143] or malarial parasites to murine erythrocytes [144]. O-acetylation in the colon appears in response to microbial colonization [145] and the intestinal level of O-acetylation varies with the amount of microbes [146]. In addition, O-acetylated sialic acids are detected in neural gangliosides where they have a role in the development of certain organs [147]. One function of O-acetylation is to provide a primary barrier for host cells and tissues against attacking pathogens. Therefore, it is quite natural that this kind of protective modification is observed in unfertilized eggs of salmon which are usually laid in the bottom of rivers whose waters are full of micro-organisms. On the other hand, the skin of fish is also a barrier against pathogens so in this context our finding of extensive O-acetylation could have been expected.

In salmon kininogen we demonstrated a Neu5Acα2,8-Neu5Ac disialosyl group from N-glycans.

An α2,8-sialyltransferase has been described from the eggs of rainbow trout (*Oncorhynchus mykiss*) [148]. Later also an α2,8-disialylated N-glycan was reported from rainbow trout ovarian fluid [110]. Disialosyl groups were first described from the rat brain in 1977 [104, 105].

Cod kininogen carries a sulphate group in its antennal N-acetylglucosamine. The most well-known sulphated glycans are chondroitin sulphate and heparan sulphate [99] but sulphated glycans have been described on glycoproteins like HIV gp120 [149], respiratory mucins [150], thyroglobulin [151] and hen egg albumin [152]. Sulphated glycoproteins have also been detected from fish; the eggs of dace have highly sulphated multiantennary N-glycans [109]. The biological significance of sulphation is not fully understood but sulphation modifies glycoconjugates to a more acidic and viscose direction [150], increase structural diversity for recognition and signalling purposes. As other kininogens were not sulphated this explanation however is not convincing for the sulphation of cod kininogen. Sulphation is also known to increase protease resistance, water-holding capacity and high charge density and it is often associated with sialylation and O-acetylation [153]. Cod live at the bottom of the sea [37] and therefore it is possible that sulphation in cod skin kininogen is just adding protection for the fish living in this hazardous environment.

We detected only Neu5Ac as sialic acid species in fish skin glycoproteins and no evidence of Neu5Gc or KDN was seen. This is slightly surprising as several fish eggs are reported to have both NeuGc and KDN [107], with or without O-acetylation [154]. This might be a sign of the different function of different sialic acids or perhaps targeting of glycoproteins is mediated by different types of sialic acids.

6. CONCLUSIONS

Four cysteine proteinase inhibitors with interesting characteristics were found. Kininogens were found in all fish species studied. Salarin is a new type of cysteine proteinase inhibitor and it was cloned. As salarin is a new type of inhibitor it will be interesting to see whether salarin is found in other fish and animal species. Salarin belongs to a new family of cysteine proteinase inhibitors, proregion-like cysteine proteinase inhibitors, but further studies are needed to characterize its specificity in detail. It also remains to be seen whether salarin is a salmonid specific inhibitor.

In this study, the analyzed glycans were isolated from the fish skin. Structural analysis of the glycans isolated from inner organs of the studied fish is needed to see whether the unusually high O-acetylation found is a characteristic of fish skin.

Although the original suggestion that high molecular weight cysteine proteinase inhibitors from the skin of fish could be defensive agents against viruses, bacteria and pathogens, could not be confirmed, it is possible that in tests more suitable for large molecules the effect of these inhibitors could be seen. As these inhibitors are transported to skin and they have glycan structures of a protective type it is very likely that they act as defensive agents in some yet undefined way.

It has been fascinating to realize that proteinases and their inhibitors are on opposite sides but they still work together to keep up nature's balance.

7. ACKNOWLEDGEMENTS

The experimental work of this study was carried out in the Protein Chemistry Laboratory at the Institute of Biotechnology, University of Helsinki during 1998-2000. I thank Professor Mart Saarma, Head of the Institute of Biotechnology, for providing excellent research facilities and scientific atmosphere.

I want to express my deepest gratitude to my supervisor, Docent Nisse Kalkkinen, for sharing his knowledge of protein chemistry and for guiding me towards this goal. It has been a privilege to work in his well-equipped laboratory where nearly everything is possible.

Professor Ari Rinne is thanked for coordinating the EU-project CT97-3508 “Fish cysteine proteinase inhibitors and infectious diseases”. All the project partners in Tromsö, Halle, Oulu and Turku are thanked for unforgettable moments. Especially I want to thank Doctor Jarl Bøgwald, Docent Mikko Järvinen, Doctor Ekkehard Weber and Dagmar Günther for smooth co-operation and positive atmosphere they brought to the project.

This manuscript was carefully reviewed by Professor Heikki Rauvala and Docent Jaana Vesterinen. I warmly thank them both for their time and improving suggestions.

Docent Jari Helin is warmly thanked for introducing me to the world of glycobiology and mass spectrometry. I truly appreciate his help and patience and will always remember our cascading discussions as well as his humour which has carried me over the bad days.

My colleagues at the protein lab are thanked for their help at practical things as well as sharing the good and bad moments of research and life. Thank you Anu, Annika, Elina, Gunilla, Jari, Jukka, Jussi, Keijo, Leena, Marko, Marjo, Minna, Mira, Tero, Tuula and Ulla. Very special thanks are due to two mass men, Jari and Jussi, for their invaluable help in mass spectroscopy and to Ulla who originally got me involved with this story.

Lars Paulin, Head of the DNA laboratory, is thanked for his help, enthusiasm and bright ideas. I warmly thank Noora, Päivi, Ritu and Sini for both their professional help in DNA work and their friendship. I also want to thank all members of the DNA lab for the joyful moments when we still shared the lab and coffee breaks.

During the writing process I had an opportunity to work at well-organized laboratory of Carbion Inc.. I want to thank all Carbion people for regarding me as a member of a family. Special thanks are due to Ritva and Annari for their friendship.

Cami, Eve, Ragna and Tiina are thanked for supportive and refreshing discussions on science and non-science.

I am thankful to Else, Repe, Pauli, Petri, Tinde and Tommi likewise all the personnel at Viikki Biocenter for making everyday life easier.

I have been blessed with a wide variety of dear friends whose help and support is gratefully acknowledged. I sincerely thank my old friends Sari and Maarit for their lasting friendship and Minna for her thoughtful care. Countless thanks are due to my mother, my father, my mother-in-law and other relatives for all their help and support.

My dearest ones, Pasi and Otso, have shown me that everything is possible. I deeply thank Pasi for his love and respect and fulfilment of my dreams and little Otso for intensely embracing me with his charm.

anneirmeli

May 2004, Hollola

8. REFERENCES

[1] Barrett, A. J. (1998) Cysteine peptidases and their clans. In Barrett, A. J., Rawlings, N. D., and Woesner, J. F (eds.) Handbook of proteolytic enzymes, Academic Press Inc. 1998.

[2] Barrett, A. J. (1994) Classification of peptidases. Methods in Enzymology 244, 1-15.

[3] Hernandez, A. A. and Roush, W. R. (2002) Recent advances in the synthesis, design and selection of cysteine protease inhibitors. Current Opinion in Chemical Biology 6, 459-465.

[4] Järvinen, M. and Hopsu-Havu, V. K. (1975) α-N-benzoylarginine-2-naphthylamide hydrolase (cathepsin B1 ?) from rat skin. II. Purification of the enzyme and demonstration of two inhibitors in the skin. Acta Chemica Scandinavica B 29, 772-780.

[5] Järvinen, M. (1978) Purification and some characteristics of the human epidermal SH-protease inhibitor. Journal of Investigative Dermatology 71, 114-118.

[6] Rawlings, N. D. and Barrett, A. J. (1994) Families of cysteine peptidases. Methods in Enzymology 244, 461-486.

[7] Rawlings, N. D., O'Brien, E. and Barrett, A.J. (2002) MEROPS: The protease database. Nucleic Acids Research 30, 343-346.

[8] MEROPS database, Internet address http://www.merops.ac.uk

[9] Tao, K., Stearns, N. A., Dong, J., Wu, Q. L., and Sahagian, G. G. (1994) The proregion of cathepsin L is required for proper folding, stability, and ER exit. Archives of Biochemistry and Biophysics 311, 19-27.

[10] Cygler, M. and Mort, J. S. (1997) Proregion structure of members of the papain superfamily. Mode of inhibition of enzymatic activity. Biochimie 79, 645-652.

[11] Turk, B., Turk, D. and Turk, V. (2000) Lysosomal cysteine proteases: more than scavengers. Biochim.Biophys.Acta 1477, 98-111.

[12] Turk, V., Turk, B. and Turk, D. (2001) Lysosomal cysteine proteases: facts and opportunities. EMBO Journal 20, 4629-4633.

[13] Karrer, K. M., Peiffer, S. L. and DiTomas, M. E. (1993) Two distinct gene subfamilies within the family of cysteine protease genes. Proceedings of the National Academy of Sciences 90, 3063-3067.

[14] Kos, J. and Lah, T. T. (1998) Cysteine proteinases and their endogenous inhibitors: Target proteins for prognosis, diagnosis and therapy in cancer. Oncology Reports 5, 1349-1361.

[15] Turk, V., Turk, B., Guncar, G., Turk, D. and Kos, J. (2002) Lysosomal cathepsins: structure, role in antigen processing and presentation, and cancer. Advances in Enzyme Regulation 42, 285-303.

[16] Kaneko, M., Tomita, T., Nakase, T., Ohsawa, Y., Seki, H., Takeuchi, E., Takano, H., Shi, K., Takahi, K., Kominami, E., Uchiyama, Y., Yoshikawa, H. and Ochi, T. (2001) Expression of proteinases and inflammatory cytokines in subchondral bone regions in the destructive joint of rheumatoid arthritis. Rheumatology 40, 247-255.

[17] Turk, B., Turk, V. and Turk, D. (1997) Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. Biological Chemistry 378, 141-150.

[18] Que, X and Reed, S. L. (2000) Cysteine proteinases and the pathogenesis of amebiasis. Clinical Microbiology Reviews 12, 196-206.

[19] Stoka, V., Nycander, M., Lenarcic, B., Labriola, C., Cazzulo, J. J., Björk, I. and Turk, V. (1995) Inhibition of cruzipain, the major cysteine proteinase of the protozoan parasite, *Trypanosoma cruzi*, by proteinase inhibitors of the cystatin superfamily. FEBS Letters 370, 101-104.

[20] Turk, V., Podobnik, M, Turk, B. and Turk, D. (1997) Regulation of cysteine proteinase activity by protein inhibitors, zymogen activation and pH. In Hopsu-Havu, V. K., Järvinen, M., and Kirschke, H. (eds) Proteolysis in cell functions. pp. 128-136. Amsterdam, IOS Press.

[21] Salvesen, G. and Nagase, H. (1989) Inhibition of proteolytic enzymes. In Bond, J. S. and Beynon, R. J. (eds) Proteolytic enzymes: A practical approach. Practical approach series vol. 5, pp. 83-104. Oxford, IRL Press.

[22] Otto, H-H. and Schirmeister, T. (1997) Cysteine proteases and their inhibitors. Chemical Reviews 97, 133-171.

[23] Jerala, R., Zerovnik, E., Kidric, J. and Turk, V. (1998) pH-induced conformational transitions of the propeptide of human cathepsin L. A role for a molten globule state in zymogen activation. Journal of Biological Chemistry 273, 11498-11502.

[24] Kirschke, H., Barrett, A. J. and Rawlings, N. D. (1995) Proteinases 1: Lysosomal cysteine proteinases. Protein Profile 2, 1587-1643.

[25] Turk, B., Dolenc, I., Lenarcic, B., Krizaj, I., Turk, V., Bieth, J.G. and Björk, I. (1999) Acidic pH as a physiological regulator of human cathepsin L activity. European Journal of Biochemistry 259, 926-932.

[26] Turk, V. and Bode, W. (1991) The cystatins: protein inhibitors of cysteine proteinases. FEBS Letters 285, 213-219.

[27] Yamashita, M. and Konagaya, S. (1990) High activities of cathepsins B, D, H, and L in the white muscle of chum salmon in spawning migration. Comparative Biochemistry & Physiology 95B, 149-152.

[28] Yamashita, M. and Konagaya, S. (1990) Purification and characterization of cathepsin L from the white muscle of chum salmon, *Oncorhynchus keta*. Comparative Biochemistry & Physiology 96B, 247-252.

[29] Yamashita, M. and Konagaya, S. (1990) Purification and characterization of cathepsin B from the white muscle of chum salmon, *Oncorhynchus keta*. Comparative Biochemistry & Physiology 96B, 733-737.

[30] Kawabata, C. and Ichishima, E. (1997) Miltpain, new cysteine proteinase from the milt of chum salmon, *Oncorhyncus keta*. Comparative Biochemistry & Physiology 117B, 445-452.

[31] Kawabata, C., Doi, Y., and Ichishima, E. (2000) Miltpain, a cysteine proteinase, from milt of Pacific cod (*Gadus microcephalus*): purification and characterization. Comparative Biochemistry & Physiology 125B, 533-542.

[32] Bonete, M. J., Manjon, A., Llorca, F. and Iborra, J. L. (1984) Acid proteinase activity in fish I. Comparative study of extraction of cathepsins B and D from *Mujil auratus*. Comparative Biochemistry & Physiology 78B, 203-206.

[33] Bonete, M. J., Manjon, A., Llorca, F., and Iborra, J. L. (1984) Acid proteinase activity in fish II. Purification and characterization of cathepsins B and D from *Mujil auratus* muscle. Comparative Biochemistry & Physiology 78B, 207-213.

[34] Visessanguan, W., Menino, A. R., Kim, S. M. and An, H. (2001) Cathepsin L: a predominant heat-activated proteinase in arrowtooth flounder muscle. Journal of Agricultural & Food Chemistry 49, 2633-2640.

[35] Visessanguan, W., Benjakul, S. and An, H. J. (2003) Purification and characterization of cathepsin L in arrowtooth flounder (*Atheresthes stomias*) muscle. Comparative Biochemistry & Physiology 134B, 477-487.

[36] Bayne, C. J, Gerwick, L., Fujiki, K., Nakao, M. and Yano, T. (2001) Immune-relevant (including acute phase) genes identified in the livers of rainbow trout, *Oncorhynchus mykiss*, by means of suppression subtractive hybridization. Developmental & Comparative Immunology 25, 205-217.

[37] Koli, Lauri. (1998) Suomen kalat.

[38] Yamashita, M. and Konagaya, S. (1992) Differentiation and localization of catheptic proteinases responsible for extensive autolysis of mature chum salmon muscle (*Oncorhynchus keta*). Comparative Biochemistry & Physiology 103B, 999-1003.

[39] Perez-Borla, O., Roura, S.I., Montecchia, C.L., Roldan, H. and Crupkin, M. (2002) Proteolytic activity of muscle in pre- and post-spawning hake (*Merluccius hubbsi Marini*) after frozen storage. Lebensmittel-Wissenschaft Und-Technologie-Food Science and Technology 35, 325-330.

[40] Aranishi, F. (1999) Possible role for cathepsins B and L in bacteriolysis by Japanese eel skin. Fish & Shellfish Immunology 8, 61-64.

[41] Aranishi, F. and Mano, N. (2000) Antibacterial cathepsins in different types of ambicoloured Japanese flounder skin. Fish & Shellfish Immunology 10, 87-89.

[42] Tähtinen, V., Weber, E., Gunther, D., Ylönen, A., Kalkkinen, N., Olsen, R., Järvinen, M., Söderström, K. O., Rinne, A., Björklund, H. and Bogwald, J. (2002) Immunolocalization of cysteine proteinases (cathepsins) and cysteine proteinase inhibitors (salarin and salmon kininogen) in Atlantic salmon, *Salmo salar*. Cell &Tissue Research 310, 213-222.

[43] Fossum, K. and Whitaker, J.R. (1968) Ficin and papain inhibitor from chicken egg white. Archives of Biochemistry & Biophysics 125, 367-375.

[44] Barrett, A. J. (1981) Cystatin, the egg white inhibitor of cysteine proteinases. Methods in Enzymology 80, 771-777.

[45] Fräki, J. E. (1976) Human skin proteases. Separation and characterization of two acid proteases resembling cathepsin B1 and cathepsin D and of an inhibitor of cathepsin B1. Archives of Dermatological Research. 255, 217-220.

[46] Järvinen, M. (1976) Purification and properties of two protease inhibitors from rat skin inhibiting papain and other SH-proteases. Acta Chemica Scandinavica B30, 933-940.

[47] Sen, L.C. and Whitaker, J.R. (1973) Some properties of a ficin-papain inhibitor from avian egg white. Archives of Biochemistry and Biophysics 158, 623-632.

[48] Sasaki, M., Minakata, K., Yamamoto, H., Niwa, M., Kato, T. and Ito, N. (1977) A new serum component which specifically inhibits thiol proteinases. Biochemical and Biophysical Research Communications 76, 917-924.

[49] Lenney, J.F., Tolan, J.R., Sugai, W.J. and Lee, A.G. (1979) Thermostable endogenous inhibitors of cathepsins B and H. European Journal of Biochemistry 101, 153-161.

[50] Barrett, A. J., Fritz, H., Muller-Esterl, W., Grubb, A., Isemura, S., Järvinen, M., Katunuma, N., Machleidt, W, Sasaki, M., Turk, Vito, and Turk, V. Nomenclature and classification of the proteins homologous with the cysteine proteinase inhibitor chicken cystatin. In Turk, V. (ed.) Cysteine proteinases and their inhibitors, Proceedings of the international symposium. pp. 1-2. Walter de Gryuter.

[51] Barrett, A. J., Rawlings, N. D., Davies, M. E, Machleidt, W, Salvesen, G, and Turk, V. (1986) Cysteine proteinase inhibitors of the cystatin superfamily. In Barrett, A. J. and Salvesen, G. S.(eds) Proteinase inhibitors. pp. 515-596. Elsevier Science Publisher.

[52] Rawlings, N.D. and Barrett, A.J. (1990) Evolution of proteins of the cystatin superfamily. Journal of Molecular Evolution 30, 60-71.

[53] Muller-Esterl, W., Fritz, H., Kellerman, J., Lottspeich, F., and Machleidt, W. (1985) Genealogy of mammalian cysteine proteinase inhibitors; Common evolutionary origin of stefins, cystatins and kininogens. FEBS Letters 191, 221-226.

[54] Brown, W.M. and Dziegielewska, K.M. (1997) Friends and relations of the cystatin superfamily - new members and their evolution. Protein Science 6, 5-12.

[55] Margis, R., Reis, E.M. and Villeret, V. (1998) Structural and phylogenetic relationships among plant and animal cystatins. Archives of Biochemistry and Biophysics 359, 24-30.

[56] Vray, B., Hartmann, S., and Hoebeke, J. (2002) Immunomodulatory properties of cystatins. Cellular & Molecular Life Sciences 59, 1503-1512.

[57] Barrett, A.J. (1987) The cystatins: a new class of peptidase inhibitors. Trends in Biochemical Sciences 12, 193-196.

[58] Turk, B., Ritonja, A., Bjork, I., Stoka, V., Dolenc, I. and Turk, V. (1995) Identification of bovine stefin A, a novel protein inhibitor of cysteine proteinases. FEBS Letters 360, 101-105.

[59] Turk, B., Krizaj, I., Kralj, B., Dolenc, I., Popovic, T., Bieth, J.G., and Turk, Vito. (1992) Bovine stefin C, a new member of the stefin family. Journal of Biological Chemistry 268, 7323-7329.

[60] Gerhartz, B., Engh, R.A., Mentele, R., Eckerskorn, C., Torquato, R., Wittmann, J., Kolb, H.J., Machleidt, W., Fritz, H. and Auerswald, E.A. (1997) Quail cystatin: isolation and characterisation of a new member of the cystatin family and its hypothetical interaction with cathepsin B. FEBS Letters 412, 551-558.

[61] Laber, B., Krieglstein, K., Henschen, A., Kos, J., Turk, V., Huber, R. and Bode, W. (1989) The cysteine proteinase inhibitor chicken cystatin is a phosphoprotein. FEBS Letters 248, 162-168.

[62] Muller-Esterl, W., Iwanaga, S. and Nakanishi, S. (1986) Kininogens revisited. Trends in Biochemical Sciences 11, 336-339.

[63] Salvesen, G, Parkes, C., Abrahamson, M., Grubb, A. and Barrett, A.J. (1986) Human low molecular weight kininogen contains three copies of a cystatin sequence that are divergent in structure and in inhibitory activity for cysteine proteinases. Biochemical Journal 234, 429-434.

[64] Lenarcic, B. and Bevec, T. (1998) Thyropins - new structurally related proteinase inhibitors. Biological Chemistry 379, 105-111.

[65] Lenarcic, B., Ritonja, A., Strukelj, B., Turk, B. and Turk, V. (1997) Equistatin, a new inhibitor of cysteine proteinases from *Actinia equina*, is structurally related to thyroglobulin type-1 domain. Journal of Biological Chemistry 272, 13899-13903.

[66] Lenarcic, B. and Turk, V. (1999) Thyroglobulin type-1 domains in equistatin inhibit both papain-like cysteine proteinases and cathepsin D. Journal of BiologicalChemistry 274, 563-566.

[67] Lenarcic, B., Krishnan, G., Borukhovich, R., Ruck, B., Turk, V. and Moczydlowski, E. (2000) Saxiphilin, a saxitoxin-binding protein with two thyroglobulin type 1 domains, is an inhibitor of papain-like cysteine proteinases. Journal of Biological Chemistry 275, 15572-15577.

[68] Yamashita, M. and Konagaya, S. (1996) A novel cysteine protease inhibitor of the egg of chum salmon, containing a cysteine-rich thyroglobulin-like motif. Journal of Biological Chemistry 271, 1282-1284.

[69] Brzin, J., Rogelj, B., Popovic, T., Strukelj, B. and Ritonja, A. (2000) Clitocypin, a new type of cysteine proteinase inhibitor from fruit bodies of mushroom *Clitocybe nebularis*. Journal of Biological Chemistry 275, 20104-20109.

[70] Monteiro, A.C.S., Abrahamson, M., Lima, A.P.C.A., Vannier-Santos, M.A. and Scharfstein, J. (2001) Identification, characterization and localization of chagasin, a tight-binding cysteine protease inhibitor in *Trypanosoma cruzi*. Journal of Cell Science 114, 3933-3942.

[71] Fox, T., de Miguel, E., Mort, J. S. and Storer, A.C. (1992) Potent slow-binding inhibition of cathepsin B by its propeptide. Biochemistry 31, 12571-12576.

[72] Carmona, E., Dufour, E., Plouffe, C., Takebe, S., Mason, P., Mort, J.S. and Menard, R. (1996) Potency and selectivity of the cathepsin L propeptide as an inhibitor of cysteine proteases. Biochemistry 35, 8149-8157.

[73] Guay, J., Falgueyret, J.P., Ducret, A., Percival, M.D., and Mancini, J.A. (2000) Potency and selectivity of inhibition of cathepsin K, L and S by their respective propeptides. European Journal of Biochemistry 267, 6311-6318.

[74] Lalmanach, G., Lecaille, F., Chagas, J.R., Authie, E., Scharfstein, J., Juliano, M.A., and Gauthier, F. (1998) Inhibition of trypanosomal cysteine proteinases by their propeptides. Journal of Biological Chemistry 273, 25112-25116.

[75] Yamamoto, Y., Watabe, S., Kageyama, T. and Takahashi, S.Y. (1999) A novel inhibitor protein for *Bombyx* cysteine proteinase is homologous to propeptide regions of cysteine proteinases. FEBS Letters 448, 257-260.

[76] Delaria, K., Fiorentino, L., Wallace, L., Tamburini, P., Brownell, E. and Muller, D. (1994) Inhibition of cathepsin L-like cysteine proteases by cytotoxic T- lymphocyte antigen-2 β. Journal of Biological Chemistry 269, 25172-25177.

[77] Yamamoto, Y., Kurata, M., Watabe, S., Murakami, R., and Takahashi, S.Y. (2002) Novel cysteine proteinase inhibitors homologous to the proregions of cysteine proteinases. Current Protein and Peptide Science 3, 231-238.

[78] Toyohara, H., Makinodan, Y., and Ikeda, S. (1988) Detection of a cysteine protease inhibitor in carp muscle. Nippon Suisan Gakkaishi 54, 157.

[79] Yamashita, M. and Konagaya, S. (1991) Cysteine protease inhibitor in egg of chum salmon. Journal of Biochemistry 110, 762-766.

[80] Yamashita, M. and Konagaya, S. (1991) A comparison of cystatin activity in the various tissues of chum salmon *Oncorhynchus keta* between feeding and spawning migrations. Comparative Biochemistry & Physiology 100A, 749-751.

[81] Koide, Y. and Noso, T. (1994) The complete amino acid sequence of pituitary cystatin from chum salmon. Biosci.Biotechnol.Biochem. 58, 164-169.

[82] Yamashita, M. and Konagaya, S. (1996) Molecular cloning and gene expression of chum salmon cystatin. Journal of Biochemistry 120, 483-487.

[83] Li, F, An, H., Seymour, T.A., Bradford, C.S., Morrissey, M.T., Bailey, G.S., Helmrich, A. and Barnes, D.W. (1998) Molecular cloning, sequence analysis and expression distribution of rainbow trout (*Oncorhynchus mykiss*) cystatin C. Comparative Biochemistry & Physiology 121B, 135-143.

[84] Li, F, An, H., Seymour, T.A., and Barnes, D.W. (2000) Rainbow trout (*Oncorchyncus mykiss*) cystatin C: expression in *Escherichia coli* and properties of the recombinant protease inhibitor. Comparative Biochemistry & Physiology 121B, 493-502.

[85] Tsai, Y.J., Chang, G.D., Huang, C.J., Chang, Y.S. and Huang, F.L. (1996) Purification and molecular cloning of carp ovarian cystatin. Comparative Biochemistry & Physiology 113B, 573-580.

[86] Macrae, A.D. and Brenner, S. (1995) Analysis of the dopamine receptor family in the compact genome of the puffer fish *Fugu rubripes*. Genomics 25, 436-446.

[87] Gong, Z., Yan, T., Liao, J., Lee, S.E., He, J. and Hew, C.L. (1997) Rapid identification and isolation of zebrafish cDNA clones. Gene 201, 87-98.

[88] Synnes, M. (1998) Purification and characterization of two cysteine proteinase inhibitors from the skin of Atlantic salmon (*Salmo salar* L.). Comparative Biochemistry & Physiology 121B, 257-264.

[89] Conlon, J.M. and Olson, K.R. (1993) Purification of a vasoactive peptide related to lysyl-bradykinin from trout plasma. FEBS Letters 334, 75-78.

[90] Conlon, J.M., Platzack, B., Marra, L.E., Youson, J.H., and Olson, K.R. (1995) Isolation and biological activity of [Trp5]bradykinin from the plasma of the phylogenetically ancient fish, the bowfin and the longnosed gar. Peptides 16, 485-489.

[91] Platzack, B. and Conlon, J.M. (1997) Purification, structural characterization, and cardiovascular activity of cod bradykinins. American Journal of Physiology 272, R710-R717.

[92] Li, Z., Secor, S.M., Lance, V.A., Masini, M.A., Vallarino, M. and Conlon, J.M. (1998) Characterization of bradykinin-related peptides generated in the plasma of six sarcopterygian species (African lungfish, amphiuma, coachwhip, bullsnake, gila monster, and Gray's monitor). General and Comparative Endocrinology 112, 108-114.

[93] Varki, A. (1993) Biological roles of oligosaccharides: all of the theories are correct. Glycobiology 3, 97-130.

[94] Ellgaard, L. and Helenius, A. (2003) Quality control in the endoplasmic reticulum. Nature Reviews Molecular Cell Biology 4, 181-191.

[95] Dahms, N.M., Lobel, P. and Kornfeld, S. (1989) Mannose-6-phosphate receptors and lysosomal enzyme targeting. Journal of Biological Chemistry 264, 12115-12118.

[96] McEver, R.P., Moore, K.L. and Cummings, R.D. (1995) Leukocyte trafficking mediated by selectin-carbohydrate interactions. Journal of Biological Chemistry 270, 11025-11028.

[97] Smith, S.M., Kane, S.E., Mason, R.W. and Gottesman, M.M. (1989) Glycosylation of procathepsin L does not account for species molecular-mass differences and is not required for proteolytic activity. Biochemical Journal 262, 931-938.

[98] Kelker, H.C., Yip, Y.K., Anderson, P. and Vilcek, J. (1983) Effects of glycosidase treatment on the physicochemical properties and biological activity of human interferon-gamma. Journal of Biological Chemistry 258, 8010-8013.

[99] Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Hart, G, and Marth, J. (1999) Essentials of glycobiology. Cold Spring Harbor Laboratory Press, NY.

[100] Gagneux, P. and Varki, A. (1999) Evolutionary considerations in relating oligosaccharide diversity to biological function. Glycobiology 9, 747-755.

[101] Traving, C. and Schauer, R. (1998) Structure, function and metabolism of sialic acids. Cellular and Molecular Life Sciences 54, 1330-1349.

[102] Klein, A. and Roussel, P. (1998) O-acetylation of sialic acids. Biochimie 80, 49-57.

[103] Inoue, S. and Iwasaki, M. (1978) Isolation of a novel glycoprotein from the eggs of rainbow trout: occurrence of disialosyl groups on all carbohydrate chains. Biochemical and Biophysical Research Communications 83, 1018-1023.

[104] Finne, J., Krusius, T. and Rauvala, H. (1977) Occurrence of disialosyl groups in glycoproteins. Biochemical and Biophysical Research Communications 74, 405-410.

[105] Finne, J., Krusius, T., Rauvala, H. and Hemminki, K. (1977) The disialosyl group of glycoproteins. European Journal of Biochemistry 77, 319-323.

[106] Iwasaki, M. and Inoue, S. (1985) Structures of the carbohydrate units of polysialoglycoproteins isolated from the eggs of four species of salmonid fishes. Glycoconjugate Journal 2, 209-228.

[107] Iwasaki, M., Inoue, S. and Troy, F.A. (1990) A new sialic acid analogue, 9-O-acetyl-deaminated neuraminic acid, and alpha -2,8-linked O-acetylated poly(N-glycolylneuraminyl) chains in a novel polysialoglycoprotein from salmon eggs. Journal of Biological Chemistry 265, 2596-2602.

[108] Iwasaki, M., Seko, A., Kitajima, K., Inoue, Y. and Inoue, S. (1992) Fish egg glycophosphoproteins have species-specific N-linked glycan units previously found in a storage pool of free glycan chains. Journal of Biological Chemistry 267, 24287-24296.

[109] Taguchi, T., Iwasaki, M., Muto, Y., Kitajima, K., Inoue, S., Khoo, K.H., Morris, H.R., Dell, A. and Inoue, Y. (1996) Occurrence and structural analysis of highly sulfated multiantennary N- linked glycan chains derived from a fertilization-associated carbohydrate-rich glycoprotein in unfertilized eggs of *Tribolodon hakonensis*. European Journal of Biochemistry 238, 357-367.

[110] Funakoshi, Y., Taguchi, T., Sato, C., Kitajima, K., Inoue, S., Morris, H. R., Dell, A., and Inoue, Y. (1997) Occurrence of terminal alpha 2->8-linked disialylated poly-N-acetyllactosamine chains with Le(X) and I antigenic glycotopes in tetraantennary arms of an N-linked glycoprotein isolated from rainbow trout ovarian fluid. Glycobiology 7, 195-205.

[111] Davies, P.L. and Hew, C.L. (1990) Biochemistry of fish antifreeze proteins. FASEB Journal 4, 2460-2468.

[112] Harding, M. M., Anderberg, P.I. and Haymet, A.D.J. (2003) ‘Antifreeze’ glycoproteins from polar fish. European Journal of Biochemistry 270, 1381-1392.

[113] Andersen, O., Flengsrud, R., Norberg, K. and Salte, R. (2000) Salmon antithrombin has only three carbohydrate side chains, and shows functional similarities to human beta-antithrombin. European Journal of Biochemistry 267, 1651-1657.

[114] Metcalf, V.J., Brennan, S.O., Chambers, G.K. and George, P.M. (1998) The albumin of the brown trout (*Salmo trutta*) is a glycoprotein. Biochimica et Biophysica Acta 1386, 90-96.

[115] Nelson, J.S. (1994) Fishes of the world. 3rd. edition. New York, John Wiley & Sons, Inc.

[116] Internet address “www.FishBase.org”

[117] A broschure of Norwegian College of Fishery Science, University of Tromsø. 1998.

[118] Anastasi, A., Brown, M.A., Kembhavi, A.A., Nicklin, M.J., Sayers, C.A., Sunter, D.C. and Barrett, A.J. (1983) Cystatin, a protein inhibitor of cysteine proteinases. Improved purification from egg white, characterization, and detection in chicken serum. Biochemical Journal 211, 129-138.

[119] O'Connell, K.L. and Stults, J T. (1997) Identification of mouse liver proteins on two-dimensional electrophoresis gels by matrix-assisted laser desorption/ionization mass spectrometry of in situ enzymatic digests. Electrophoresis 18, 349-359.

[120] Matsudaira, P. (1987) Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. Journal of Biological Chemistry 262, 10035-10038.

[121] Kinnunen, T., Kaksonen, M., Saarinen, J., Kalkkinen, N., Peng, H.B. and Rauvala, H. (1998) Cortactin-Src kinase signaling pathway is involved in N-syndecan-dependent neurite outgrowth. Journal of Biological Chemistry 273, 10702-10708.

[122] Helin, J., Caldentey, J., Kalkkinen, N. and Bamford, D. (1999) Analysis of the multimeric state of proteins by matrix assisted laser desorption/ionization mass spectrometry after cross-linking with glutaraldehyde. Rapid Communications in Mass Spectrometry 13, 185-190.

[123] Nyman, T.A., Kalkkinen, N., Tölö, H. and Helin, J. (1998) Structural characterisation of N-linked and O-linked oligosaccharides derived from interferon-α2b and interferon-α14c produced by Sendai-virus-induced human peripheral blood leukocytes. European Journal of Biochemistry 253, 485-493.

[124] Ciucanu, I and Kerek, F. (1984) A simple rapid method for the permethylation of carbohydrates. Carbohydrate Research 131, 209-217.

[125] Van Lenten, L. and Ashwell, G. (1971) Studies on the chemical and enzymatic modification of glycoproteins. A general method for the tritiation of sialic acid-containing glycoproteins. Journal of Biological Chemistry 246, 1889-1894.

[126] Mirgorodskaya, E. and Roepstorff, P. (1998) Mass spectrometric determination of O-glycosylation sites using beta-elimination and partial acid hydrolysis. Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics, Orlando, FL, 215.

[127] Saarinen, J., Welgus, H.G., Flizar, C.A., Kalkkinen, N. and Helin, J. (1999) N-Glycan structures of matrix metalloproteinase-1 derived from human fibroblasts and from HT-1080 fibrosarcoma cells. European Journal of Biochemistry 259, 829-840.

[128] Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate- phenol-chloroform extraction. Analytical Biochemistry 162, 156-159.

[129] Frohman, M.A. (1990) RACE: Rapid amplification of cDNA ends. In PCR protocols: A guide to methods and applications. Pp. 28-38. Academic Press, Inc.

[130] Saris, P.E.J., Paulin, L.G. and Uhle´n, M. (1990) Direct amplification of DNA from colonies of *Bacillus subtilis* and *Escherichia coli* by the polymerase chain reaction. Journal of Microbiological Methods 11, 121-126.

[131] Staden, R., Beal, K.F. and Bonfield, J.K. (1998) The Staden package, Methods in Molecular Biology 132, 115-130.

[132] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25, 3389-3402.

[133] Schauer, R. (1991) Biosynthesis and function of N- and O-substituted sialic acids. Glycobiology 1, 449-452.

[134] Papac, D.I., Wong, A. and Jones, A.J. (1996) Analysis of acidic oligosaccharides and glycopeptides by matrix- assisted laser desorption/ionization time-of-flight mass spectrometry. Analytical Chemistry 68, 3215-3223.

[135] Varki, A. and Diaz, S. (1984) The release and purification of sialic acids from glycoconjugates: methods to minimize the loss and migration of O-acetyl groups. Analytical Biochemistry 137, 236-247.

[136] Hemmerich, S., Leffler, H. and Rosen, S.D. (1995) Structure of the O-Glycans in GlyCAM-1, an endothelial-derived ligand for L-selectin. Journal of Biological Chemistry 270, 12035-12047.

[137] Björck, L., Åkesson, P., Bohus, M., Trojnar, J., Abrahamson, M., Olafsson, I. and Grubb, A. (1989) Bacterial growth blocked by a synthetic peptide based on the structure of a human proteinase inhibitor. Nature 337, 385-386.

[138] Johansson, T. R. (2001) Fish pathogenic rhabdoviruses: molecular characterization, phylogeny and inhibition of virus-induced apoptosis by cysteine proteinase inhibitors. A dissertation of Institute of Parasitology, Department of Biology, Åbo Akademi University.

[139] Björklund, H.V., Johansson, T.R. and Rinne, A. (1997) Rhabdovirus-induced apoptosis in a fish cell line is inhibited by a human endogenous acid cysteine proteinase inhibitor. Journal of Virology 71, 5658-5662.

[140] Aparacio, S and Chapman, J. (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297, 1301-1310.

[141] Hedges, S.B and Kumar, S. (2002) Vertebrate genomes compared. Science 297, 1283-1285.

[142] Butor, C., Diaz, S. and Varki, A. (1993) High level O-acetylation of sialic acids on N-linked oligosaccharides of rat liver membranes. Differential subcellular distribution of 7- and 9-O-acetyl groups and of enzymes involved in their regulation. Journal of Biological Chemistry 268, 10197-10206.

[143] Higa H.H., Rogers G.N., and Paulson J.C. (1985) Influenza virus hemagglutinins differentiate between receptor determinants bearing N-acetyl-, N-glycolyl-, and N,O-diacetylneuraminic acids. Virology 144, 279-282.

[144] Klotz F.W., Orlandi P.A., Reuter G., Cohen S.J., Haynes J.D., Schauer R., Howard R.J., Palese P. and Miller L.H. (1992) Binding of *Plasmodium falciparum* 175-kilodalton erythrocyte binding antigen and invasion of murine erythrocytes requires N-acetylneuraminic acid but not its O-acetylated form. Molecular & Biochemical Parasitology. 51, 49-54.

[145] Muchmore, E.A., Varki, N.M., Fukuda, M. and Varki, A. (1987) Developmental regulation of sialic acid modifications in rat and human colon. The FASEB Journal 1, 229-235.

[146] Robbe, C., Capon, C., Maes, E., Rousset, M., Zweibaum, A., Zanetta, J.P. and Michalski, J.C. (2003) Evidence of regio-specific glycosylation in human intestinal mucins: Presence of an acidic gradient along the intestinal tract. Journal of Biological Chemistry , 2529200.

[147] Varki, A. (1992) Diversity in the sialic acids. Glycobiology 2, 25-40.

[148] Kitazume, S., Kitajima, K., Inoue, S., Inoue, Y. and Troy, F.A. (1994) Developmental expression of trout egg polysialoglycoproteins and the prerequisite α2,6-, and α2,8-sialyl and α2,8- polysialyltransferase activities required for their synthesis during oogenesis. Journal of Biological Chemistry 269,10330-10340.

[149] Shilatifard, A., Merkle, R.K., Helland, D.E., Welles, J.L., Haseltine, W.A. and Cummings, R.D. (1993) Complex-type N-linked oligosaccharides of gp120 from human immunodeficiency virus type 1 contain sulfated N-acetylglucosamine. The Journal of Virology 67, 943-952.

[150] Lo-Guidice, J.M., Wieruszeski, J.M., Lemoine, J., Verbert, A., Roussel, P. and Lamblin, G. (1994) Sialylation and sulfation of the carbohydrate chains in respiratory mucins from a patient with cystic fibrosis. Journal of Biological Chemistry 269, 18794-18813.

[151] de Waard, P., Koorevaar, A., Kamerling, J.P. and Vliegenthart, J.F. (1991) Structure determination by 1H NMR spectroscopy of (sulfated) sialylated N-linked carbohydrate chains released from porcine thyroglobulin by peptide-N4-(N-acetyl-beta-glucosaminyl)asparagine amidase-F. Journal of Biological Chemistry 266, 4237-4243.

[152] Yamashita, K., Ueda, I., and Kobata, A. (1983) Sulfated asparagine-linked sugar chains of hen egg albumin. Journal of Biological Chemistry 258, 14144-14147.

[153] Wu A.M. Csako G. and Herp A. (1994) Structure, biosynthesis, and function of salivary mucins. Molecular & Cellular Biochemistry 137, 39-55.

[154] Angata, T., Nakata, D., Matsuda, T., Kitajima, K. and Troy, F.A. (1999) Biosynthesis of KDN (2-Keto-3-deoxy-D-glycero-D-galacto-nononic acid). Identification and characterization of a kdn-9-phosphate synthetase activity from trout testis. Journal of Biological Chemistry 274, 22949-22956.