A New Look into Protein C Inhibitor

Posttranslational Modifications and their Functions

WEI SUN
Dissertation presented at Uppsala University to be publicly examined in C10:301, BMC, Husarg. 3, Uppsala, Friday, November 5, 2010 at 09:15 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English.

Abstract

The influences of posttranslational modifications on the functions of the versatile serpin protein C inhibitor (PCI) were studied. PCI is a serine protease inhibitor that is expressed in many tissues and secreted to various fluids in human, including blood plasma, seminal plasma, and urine. PCI in blood can act both as an anticoagulant and a procoagulant and is believed to play a role in pathogen defence. PCI in reproductive tissues is believed to regulate human reproduction at several steps, including the fertilization process. Due to the broad protease specificity and the contradictory activities, the physiological role of PCI is elusive. In this work the inhibitor was purified from blood and seminal plasma by immunoaffinity chromatography. Blood-derived PCI was found to be highly heterogeneous, due to variations in posttranslational modifications. The occupancy and structures of N- and O-glycans attached to blood plasma PCI and N-glycans of seminal plasma PCI were determined by mass spectrometry. An O-glycosylation site at Thr 20 was identified in PCI derived from blood. N-glycan structures of PCI isolated from blood and seminal plasma differed markedly, demonstrating that they are expressed in a tissue-specific manner. Proteolytic processing also appeared to be tissue-specific, since N-terminally cleaved PCI was found in PCI isolated both from blood and seminal plasma, but the length of the lacking segment differed. The effects of the N-linked glycans and the N-terminus of PCI on protease inhibition were determined using enzymatic measurements with chromogenic substrates. The N-glycans and the N-terminus had different effects on the inhibition of thrombin, factor Xa and prostate specific antigen, demonstrating that posttranslational modifications of PCI affect its functional specificity. These findings enhance the understanding of the regulation of the various functions of PCI and may potentially be used for the production of specialized PCI variants for medical purposes.

Keywords: Protein C inhibitor, posttranslational modifications, N-glycan, O-glycan, mass spectrometry, factor Xa, thrombin, prostate specific antigen

Wei Sun, Department of Medical Biochemistry and Microbiology, Box 582, Uppsala University, SE-75123 Uppsala, Sweden.

© Wei Sun 2010

ISSN 1651-6206
urn:nbn:se:uu:diva-131126 (http://urn.kb.se/resolve?urn:nbn:se:uu:diva-131126)
To my parents

献给我的父母
This thesis is based on the following papers, which are referred to in the text by their Roman numerals.


* Contributed equally to the work


Reprints were made with permission from the respective publishers.
## Contents

Introduction .................................................................................................................. 9

Posttranslational modifications .................................................................................. 9
  Proteolytic cleavage ................................................................................................. 9
  Glycosylation ........................................................................................................ 10
  γ-Carboxylation .................................................................................................. 14

Hemostasis .............................................................................................................. 14

Serine proteases ....................................................................................................... 17
  Thrombin ............................................................................................................. 17
  Factor Xa .......................................................................................................... 19
  Activated protein C (APC) ............................................................................... 21
  Prostate specific antigen (PSA) ................................................................. 22

Serine protease inhibitors (Serpins) ....................................................................... 23
  Protein C inhibitor (PCI) .................................................................................. 26

Present investigation ............................................................................................... 30
  Aims ................................................................................................................ 30
  Results and discussion ....................................................................................... 31
  Conclusions ...................................................................................................... 35

Future perspectives ................................................................................................. 38

Acknowledgements ................................................................................................. 41

References ............................................................................................................... 43
Abbreviations

APC  activated protein C
AT  antithrombin
Fuc  fucose
FXa  factor Xa
GAGs  glycosaminoglycans
Gal  galactose
GalNAc  N-acetylgalactosamine
Gla  γ-carboxyglutamic acid
GlcNAc  N-acetylglucosamine
Glu  glucose
HMWK  high-molecular-weight kininogen
$k_2$  second-order rate constant
Man  mannose
N-terminus  amino-terminus
NeuAc  N-acetylneuraminic acid
NeuGc  N-glycolyneuraminic acid
PAR  protease activated receptor
PCI  protein C inhibitor
PSA  prostate specific antigen
RCL  reactive center loop
TF  tissue factor
TM  thrombomodulin
t-PA  tissue-type plasminogen activator
u-PA  urokinase-type plasminogen activator
Introduction

Posttranslational modifications

The collection of proteins produced by an organism is termed the proteome. While the proteome is encoded by the genome in each organism, and therefore a one-to-one correspondence of proteome to genome might be expected, proteomes are typically of much higher complexity than genomes. For example, estimation of the human genome put the number of human genes at about 30 000. The human proteome is thought to have a $>10$–$100$-fold higher complexity with $>300$ 000–$3$ 000 000 distinct protein forms [1]. The two major mechanisms for proteome expansion are posttranslational modifications and at the transcriptional level, by mRNA splicing.

As the name implies, posttranslational modifications occur after RNA has been translated into proteins. The immature proteins are then subjected to a battery of specific enzyme-catalyzed modifications on the side chains or backbones. About 5% of the genomes of higher eukaryotes can be dedicated to enzymes that carry out posttranslational modifications of the proteomes [1]. The posttranslational modifications of proteins play a crucial role in generating heterogeneity in proteins and help in utilizing identical proteins for different cellular functions in different cell types. Some posttranslational modifications involved in this thesis are reviewed below.

Proteolytic cleavage

Generally, proteolytic cleavage is the process of cleaving the peptide bonds between amino acids in proteins, and is the most common type of posttranslational modification. Probably all mature proteins have been so modified, at least by the proteolytic removal of their leading Met (or fMet) residues shortly after emerging from the ribosome. Some enzymes are synthesized as inactive precursors, so-called zymogens. These precursors must be activated, for instance by limited proteolysis, to become mature and functional. Secretory proteins have their signal peptide removed by specific signal peptidases before release to the extracellular environment or specific organ.

The proteolytic cleavage site is denoted P1-P1', representing the two residues between which cleavage occurs. Incrementing the numbering in the
N-terminal direction of the cleaved peptide bond, the subsequent residues can be denoted P2, P3, P4, etc. On the carboxyl side of the cleavage site, numbering is likewise incremented (P1', P2', P3', etc.).

Glycosylation
Other important posttranslational modifications include covalent modifications of proteins, i.e. specific chemical derivatizations, both at the functional groups of their side chains and at their terminal amino and carboxyl groups. Over 150 different types of side chain modifications, involving all side chains except Ala, Gly, Ile, Met, and Val, are known. These include glycosylation, acetylation, hydroxylation, phosphorylation, carboxylation, methylation, and sulfation.

Among all the covalent modifications of proteins, glycosylation is the most complex. Many eukaryotic proteins are glycoproteins; that is, they are covalently associated with carbohydrates. Carbohydrates constitute 1–90% of the weight of glycoproteins; occur in all forms of life; and have functions that span the entire spectrum of protein activities, including those of enzymes, transport proteins, receptors, hormones, and structural proteins [2].

The carbohydrate moieties have several important biological roles, but in many cases their functions remain enigmatic. Unlike the synthesis of polypeptides that follow a rigid genetic template, the carbohydrate chains are generated by a series of enzymes and, thus, have variable compositions, a phenomenon known as micro-heterogeneity.

The two main types of protein glycosylation are N-glycosylation and O-glycosylation.

N-glycosylation
N-linked glycans in glycoproteins are attached to the amide nitrogen of asparagine side chains. In animals, the sugar linked to an asparagine residue is almost always N-acetylglucosamine (GlcNAc) and the linkage is always β conformation (Figure 1).

Glycosylated asparagine residues are found almost exclusively in the sequence Asn-X-Ser or Asn-X-Thr, where X can be any amino acid except proline. Rare cases of glycosylation on Asn-X-Cys have also been described.
The biosynthetic pathway of N-linked glycosylation are divided into three stages: (1) formation of the precursor oligosaccharide, \( \text{Glc}_3\text{Man}_9\text{GlcNAc}_2 \), linked to the lipid dolichol (for structures and names of the monosaccharide units, see Figure 2), on the endoplasmic reticulum (ER) membrane; (2) transfer of the oligosaccharide from dolichol to the polypeptide in the ER lumen by oligosaccharyl-transferase; and (3) processing of the oligosaccharide in the ER and Golgi apparatus by a few or all of the following enzymes: glucosidases, mannosidases, \( \text{GlcNAc}^- \), galactosyl-, sialyl-, and fucosyl-transferases [3].

Figure 1. Attachment of an N-linked glycan to an asparagine residue.

Figure 2. Structures, names and symbols of monosaccharides referred to in this thesis. Adapted from Prof. Anne Dell’s picture.
Depending on the manner and degree of the processing, N-glycans can be
further classified into three types: Glycans terminated in unsubstituted
mannose residues are denoted as high-mannose type oligosaccharides;
hybrid structures are defined as those with both substituted (GlcNAc
linkage) and unsubstituted mannose residues; complex N-glycans refer to
those in which mannose residues are substituted with GlcNAc moieties
(Figure 3). When total N-glycans are analyzed, most vertebrate extracellular
N-glycans are found to be of the complex subtype.

It is worth noting that although glycosylation is often referred to as a
posttranslational modification, glycan transfer usually occurs for N-
glycosylation as the immature polypeptide emerges in the lumen of the
endoplasmic reticulum and, thus, is actually a cotranslational event.

![Figure 3. Some examples of N-glycans (upper panel) and special antennae
(lower panel) mentioned in this thesis. The conservative structure is shaded
in pink. Common antennae of mammalian N-glycans, such as GlcNAcGal,
are not listed. Adapted from Prof. Anne Dell’s picture.](image)

A great number of studies in naturally occurring and experimentally
induced mutations in N-glycan biosynthesis have shown that deficiencies of
enzymes for N-glycan biosynthesis are lethal or pathogenic for vertebrates
[4-8].
O-glycosylation

O-type glycosylation refers to the covalent addition of oligosaccharides to the hydroxyl group of serine or threonine residues on the proteins. The biosynthesis of O-glycan is simpler compared to N-linked glycosylation, since a lipid-linked oligosaccharide precursor for transfer to protein is not required. All the monosaccharides are added one at a time in a stepwise series of reactions, starting by the addition of GalNAc to serine or threonine residues. Moreover, in contrast to N-glycosylation, a consensus sequence for GalNAc attachment to polypeptides has not been found, because numerous transferases can attach GalNAc to serine and threonine residues. These enzymes have different specificity for different sequences of amino acids around the glycosylation target; thus, examining the entire glycosylation profile does not reflect the specificity of any one enzyme. Also, unlike N-glycosylation, adding core GalNAc residues to serine and threonine occurs posttranslationally only in the Golgi apparatus [9].

O-glycans are less branched than most N-glycans and are commonly bi-antennary structures. Further addition of one or two sugars to the initiating GalNAc gives a range of core structures where cores 1, 2, 3 and 4 are the most common (Figure 4). Antennae and terminal sialylation are built onto the cores in a similar manner as for complex type N-glycans.

![Figure 4. Most common O-glycan core structure types 1, 2, 3, and 4 of glycoproteins. Examples of antennae that include Lewis types of oligosaccharides are shown in Figure 3. Drawn by Prof. Anne Dell.](image)

O-glycosylation can result in the formation of mucin-type molecules. Mucin is defined as a cell surface or secreted protein with a large number of clustered O-glycans. The major function of many mucins is retaining water at surfaces that are not protected by moisture-impermeable layers, such as
the skin. Typical O-linked glycans on the mucins are sialylated, as the strong negative charge of sialic acids increases the capacity of binding water. In contrast to each mucin O-glycan, which is small and consists of only a few monosaccharides, another class of heavily O-glycosylated proteins, termed proteoglycans, contains O-glycans of as many as up to ~200 sugars, i.e. glycosaminoglycans (GAGs). GAGs are linear polysaccharides consisting of repeating disaccharide units and are highly negatively charged. Proteoglycans also bind water, but unlike mucins, they provide structure rather than lubrication.

O-glycans also perform roles in maintaining the stability of glycoproteins [10,11]. O-linked oligosaccharides are crucial elements in many glycoprotein-protein interactions, as has been shown by the loss of recognition function after deglycosylation. Typical examples are the selectin-ligands and the zona pellucida glycoproteins [12-16]. Moreover, intracellular proteins can be modified by β-O-linked GlcNAc (O-GlcNAc) at serines and threonines that would otherwise be phosphorylated. O-GlcNAc has been found to play roles in cell signaling [17].

γ-Carboxylation

γ-Carboxylation is the acidic modification of the glutamate residues of proteins to form glutamates with an extra carboxyl group, denoted γ-carboxyglutamate. This modification is introduced by the vitamin K-dependent enzyme, γ-glutamyl carboxylase. γ-Carboxyglutamic acid (Gla) is a unique amino acid that binds to calcium. The biosynthesis of protein γ-carboxylation requires vitamin K as a cofactor; therefore, the proteins that contain the Gla or Gla domain are denoted as “vitamin K-dependent proteins.” In vertebrates most vitamin K-dependent proteins are coagulation factors that all have an N-terminal Gla domain that is about 45 amino acid residues long and contains from 9 to 12 Gla residues [18,19]. Blood clotting factors that contain a Gla domain include prothrombin, FVII, FIX and FX, which are the clotting factors of the traditional coagulation system. Additionally, protein C and protein S of the protein C anticoagulation system [19-21] and protein Z, a regulator of blood coagulation, have Gla domains [21]. The Gla domain is required for the interactions of these proteins with phospholipids in cellular membranes. More about the Gla domain in relation to the function of factor Xa and APC is discussed in a later section.

Hemostasis

Blood circulation in the human body is essential for transportation of oxygen, nutrients, minerals, and metabolic products, etc. Humans, therefore,
have evolved an efficient, complex series of hemodynamic, cellular, and biochemical mechanisms in response to vessel injury to prevent excessive bleeding. Upon vessel injury, activated platelets adhere to macromolecules in the subendothelial tissues and then aggregate to form the primary hemostatic plug [22]. Platelets are activated by several substances that bind to the receptors on their surface. One of the receptors, PAR, is activated by thrombin. Platelet activation leads to a change of platelets’ shape and release of several compounds involved in hemostasis and, thus, facilitates their aggregation and adhesion to the vessel wall. This primary plug is friable and needs to be subsequently stabilized by fibrin produced by the secondary hemostatic reactions. The secondary hemostatic action is a series of steps in which plasma zymogens of serine proteases (see later chapter) are transformed into active enzymes (Figure 5). These enzymes act to convert their procofactor substrates to cofactors. The sequential nature of the reactions, in which the resulting product serves as the enzyme of the next reaction, amplifies the overall velocity of the reaction. The final event is the formation of thrombin, which converts the soluble fibrinogen into fibrin to form an insoluble polymer that stabilizes the clot [23].

The coagulation factors have two pathways of activation: the extrinsic and the intrinsic (Figure 5). Plasma coagulation factors are identified by roman numerals, and activated factors are designated by “a”.

The extrinsic (or tissue factor) pathway is considered to play the most important role for initiating coagulation in vivo [24]. It is initiated when blood comes in contact with tissue factor (TF, factor III), which is exposed on cellular membranes of subendothelial tissue at a vessel injury. The circulating plasma protein factor VII then forms a complex with it. The enzymatically active form of this clotting factor, factor VIIa, is present in the blood, and the rate of formation is amplified when the coagulation system is active. This factor VIIa-TF complex, in the presence of Ca$^{2+}$ and phospholipids, converts factor IX to factor IXa and factor X to factor Xa [25]. Once factor Xa is generated, the remaining part of the cascade is the same as the intrinsic pathway.

Exposing blood to collagen in a damaged vascular wall, or exposing the blood to a nonendothelial "surface," will initiate the intrinsic (or contact factor) pathway. Factor XII is converted to factor XIIa. Activated factor XII is a protease, which activates factor XI to factor XIa. This reaction requires the presence of high-molecular-weight kininogen (HMWK) and prekallikrein. Activated factor XI can further activate factor IX to factor IXa [26]. Factor IXa then converts factor X to factor Xa. The activation of factor X is greatly accelerated by factor VIIIa. Factor Xa then functions as a protease that converts the inactive prothrombin to its active form, thrombin. This step requires the presence of factor Va [27]. Thrombin then converts soluble fibrinogen to insoluble fibrin, which then polymerizes to form a stable clot. Once thrombin is generated, it can also catalyze the conversion
of factors V [28] and VIII [29] to their activated forms through a positive feedback mechanism and, thus, convert more prothrombin to thrombin. In this manner, thrombin is able to accelerate the entire cascade, resulting in the formation of large amounts of fibrin. It is important to note that when the cascade is activated, the amount of product formed in the individual reactions increases logarithmically as one moves down the cascade.

Figure 5. Blood coagulation cascade. Small “a” indicates the active form of the enzyme. Factors in fuchsia are co-factors. Drawn by Dr. Sophia Schedin Weiss.

The formation of a blood clot and its inhibition are essential, as the latter prevents excessive thrombin formation and thrombosis. Three systems of coagulation inhibitors have been identified so far: (1) Antithrombin, a plasma protein that inhibits thrombin and factors IXa and Xa [30-32]. (2) Protein C and its cofactor, protein S. Protein C is activated by the thrombomodulin-thrombin complex and then forms a complex with protein S. The active protein C-protein S complex specifically degrades factors Va and VIIIa [31,33,34]. (3) Tissue factor pathway inhibitor (TFPI), a protein that circulates bound to lipoproteins and specifically inhibits tissue factor [35].

Once hemostasis is restored and the tissue is repaired, the clot or thrombus must be removed from the vessel, which is achieved by fibrinolysis. A proenzyme, plasminogen, can be converted to the active enzyme plasmin by different plasminogen activators, i.e. a t-PA or u-PA (tissue- or urokinase-type plasminogen activator) [36]. Plasmin is a potent proteolytic enzyme with a broad spectrum of activity, including mediating the solubilization of fibrin.
Serine proteases

Almost one-third of all proteases can be classified as serine proteases, named after the presence of a unique reactive serine residue. Serine proteases have a common catalytic triad, usually consisting of three amino acids: serine, aspartate, and histidine, in their active site [37,38]. Novel catalytic triads have also been discovered recently, including Ser-His-Glu, Ser-Lys-His, His-Ser-His and N-terminal Ser [39]. When the linear sequence of amino acids folds into its tertiary structure, the triad is arranged in a structure that results in the side chain of the serine residue becoming negatively charged through the loss of the hydrogen of the hydroxyl group to histidine. This nucleophile can then attack the carbonyl group of the peptide bond that is the target for the cleavage.

Chymotrypsin-like proteases are the most abundant serine proteases in nature, with over 240 recognized [40], and they are of extremely widespread occurrence and diverse function. Representative mammalian chymotrypsin-like proteases include chymotrypsin, trypsin, and pancreatic elastase that work as digestive proteases; factors VIIa, IXa, Xa, XIa, thrombin, and protein C that regulate blood coagulation; urokinase, t-PA, plasmin and kallikrein that regulate fibrinolysis; and prostate specific antigen (PSA) and acrosin that play important roles in reproduction [41].

Thrombin

Thrombin is a serine protease that circulates in the plasma at 2-5 nM [42] and plays a key role in blood coagulation. The cleavage of prothrombin (plasma concentration 2 μM) by factor Xa on the bond 327-328 results in the formation of active thrombin. Thrombin exerts either procoagulant or anticoagulant functions in hemostasis [43]. The procoagulant role involves the conversion of soluble fibrinogen into insoluble fibrin clots, platelet activation through cleavage of protease-activated receptors (PAR), and stimulation of its own formation through cleavage activation of the essential cofactors V and VIII. Thrombin also stabilizes fibrin clots by activating factor XIII, which can cross-link fibrin polymers; and TAFI (thrombin activatable fibrinolysis inhibitor), which inhibits the proteolytic degradation of clots. Thrombin plays an important role as an anticoagulant through protein C activation, but only when it is bound to thrombomodulin, a membrane receptor present on the surface of endothelial cells [44]. Activated protein C downregulates thrombin formation through cleavage inactivation of factors Va and VIIIa (Figure 5) [44,45].

An effective modulator of the opposing functions of thrombin is Na⁺, which triggers the transition of the enzyme from the anticoagulant (slow) form to the procoagulant (fast) form [46] (Figure 6). Na⁺ binding is required
for optimal cleavage of fibrinogen [47] and activation of Factors V [48], VIII [49] and XI [50], but is dispensable for cleavage of protein C [51]. Na⁺ binding also promotes the prothrombotic and signaling functions of the enzyme by enhancing cleavage of PAR1, PAR2 and PAR4 [45,46]. The Na⁺ binding site of thrombin is nestled between the 220- and 186- loops that contribute to the primary specificity of serine proteases [41,52,53]. This effect is exquisitely allosteric because the Na⁺ binding site is located away from residues of the catalytic triad or involved directly in substrate recognition. Crystal structures of thrombin in the presence or absence of Na⁺ have been solved by different groups and are deposited in the Protein Data Bank (PDB), with codes 1SG8, 1SG1 [54,55] and 1JOU_AB, 1JOU_EF [56]; however, the exact mechanism of the allosteric change has not been agreed upon. It is noteworthy that under physiological conditions the two forms occur almost equally because the $K_d$ for Na⁺ binding is 110 mM at 37 °C and, thus, the physiologic [Na⁺] (140mM) is not sufficient for saturation [57].

![Schematic of the multiple roles of thrombin. In blood, the protease is present in a Na⁺-free slow form (~40%) and a Na⁺-bound fast form (~60%). The fast form is responsible for the cleavage of fibrinogen to fibrin and activation of factors V, VIII and XI, PAR1, PAR3 and PAR4. The slow form activates anticoagulant protein C with assistance from the cofactor thrombomodulin. The picture is from [58].](image)

Thrombin is composed of two polypeptide chains, A and B, which are covalently linked through a disulfide bond. The A chain has no documented functional role, while the B chain hosts the entrance to the active site and all functional epitopes of the enzyme (Figure 7). It has an active site triad (in chymotrypsin numbering, His57-Asp102-Ser195) that is a characteristic of all plasma-clotting enzymes. Thrombin features a trypsin-like specificity and cuts preferably at Arg residues of substrate. Unlike trypsin, thrombin selectively cleaves at specific Arg sites by, in addition to the active site,
using ancillary interactions from “exosites” [59]. Thrombin has extra surface structures that influence the interactions with macromolecular substrates and thus make it a more discriminating protease. One of these subsites is called the anion-binding exosite-I (or exosite I). Exosite I of thrombin has been implicated in fibrinogen and thrombomodulin binding [45,60,61]. Another surface structure of thrombin is the anion-binding exosite-II (or exosite II). Exosite II is responsible primarily for glycosaminoglycan binding to thrombin [62-66]. Heparin glycosaminoglycans are potent anticoagulants, as they accelerate the inhibition of thrombin by antithrombin in several orders of magnitude. Details are discussed in a later chapter.

Prothrombin has three N-glycosylation sites at Asn 78, Asn 100, and Asn 373. The cleavage of the prothrombin by factor Xa leaves off only one potential N-glycan on its heavy chain. Neither the occupancy nor the function of N-glycans on thrombin has been reported.

![Crystal structure of native thrombin drawn from PDB structure 1SG8. Chain A (olive) runs in the back of chain B (cyan). Residues that form the catalytic triad (in chymotrypsin numbering, His57-Asp102-Ser195) are in blue, hot pink, and red, respectively. The sodium ion is in orange.](image)

**Factor Xa**

Factor X is secreted into the plasma as the zymogen form of the serine protease (plasma concentration 0.13 μM) and is converted to the active form, factor Xa (plasma concentration 10 nM [67]), by the factor VIIa-TF complex in the extrinsic pathway or by the factor IXa-factor VIIIa complex in the intrinsic pathway [68]. In both complexes, the activation peptide of factor X is released by limited proteolytic processing to form mature factor Xa. Factor Xa plays a vital role in blood coagulation by converting prothrombin to thrombin. In the presence of Ca\(^{2+}\) ions, factor Xa forms the
prothrombinase complex with factor Va on the phospholipid membrane of the activated platelets.

Factor Xa consists of a light and heavy chain linked by a single disulfide bond. The light chain contains the N-terminal Gla domain and two epidermal growth factor (EGF)-like domains. The Gla domain consists of 11 $\gamma$ carboxyglutamic acid residues and enables binding to the negatively charged phospholipid membrane in the presence of $\text{Ca}^{2+}$. In factor Xa, the Gla domain is in random disordered conformation in the absence of $\text{Ca}^{2+}$ but folds properly to its native conformation in the presence of $\text{Ca}^{2+}$ [69]. $\text{Ca}^{2+}$ is thus necessary for the interaction of heparin with factor Xa [70,71]. The Gla domain is followed by a short stack of hydrophobic residues, and two epidermal growth factor (EGF)-like domains: the EGF-like 1 and the EGF-like 2. The roles of EGF domains are not fully understood but are believed to be important for protein-protein interactions [72]. The heavy chain contains a trypsin-like serine protease domain similar to those of other blood coagulation enzymes such as factor VIIa, factor IXa, and protein C [68]. Like thrombin, three residues (in chymotrypsin numbering, His57-Asp102-Ser195) form a catalytic triad at the active site (Figure 8). Factor X is converted to factor Xa by removal of the activation peptide from the N-terminus of the heavy chain, and this activation peptide contains both the two N-glycans and the two O-glycans on factor X [73], so that the resulting factor Xa carries no N- or O-glycans.

Figure 8. Crystal structure of Gla domain-free factor Xa drawn from PDB code 1XKB. The heavy chain is in cyan. EGF-like domains 1 and 2 are in yellow and orange, respectively. Residues that form the catalytic triad (in chymotrypsin numbering, His57-Asp102-Ser195) are in blue, red, and salmon, respectively. A calcium ion each is bound to the first EGF domain and the catalytic heavy chain. Crystal structure of the Gla domain is not available.
Activated protein C (APC)

The protein C pathway is an important anticoagulant system that regulates blood coagulation in vivo. The key component of this pathway is the activated protein C (APC). It is a vitamin K-dependent serine protease structurally similar to factor Xa, but functionally opposite. It functions as a potent physiological anticoagulant by degrading cofactors Va and VIIIa [44], which serve as membrane-bound cofactors to factors Xa and IXa, respectively (Figure 5).

In addition to its anticoagulatory effect, APC has anti-inflammatory and fibrinolytic properties and has been shown to inactivate type 1 plasminogen activator inhibitor (PAI-1), thereby promoting fibrinolysis [74,75]. Inflammatory cytokines both activate the coagulation cascade and inhibit fibrinolysis [76]. Disseminated intravascular coagulation, one of the most feared complications of sepsis, is caused by dysregulated coagulation. Reduced levels of protein C in patients with sepsis have been correlated with increased risk of death [77-79]. Administration of activated protein C has been found to benefit patients with severe sepsis.

In vivo, the proteolytic activity of APC is regulated by two serpins (serine protease inhibitors), namely α1-antitrypsin and protein C inhibitor (PCI) [80,81] (See the next chapter for an introduction to serpins and PCI). Additionally, it was recently reported that APC activity is inhibited by zinc ions [82].

Protein C circulates in plasma at a concentration of 70 nM. After activation on endothelial cells by thrombin bound to thrombomodulin (TM), the concentration of APC is 40 pM [83]. The activation of protein C depends on the presence of divalent cations, calcium being the most effective [84]. APC circulates in plasma as a light and heavy chain held together by a single disulfide bond [85]. Like factor Xa, APC has an N-terminal light chain consisting of a Gla domain that contains nine γ-carboxylated glutamic acid residues and two EGF-like domains (Figure 9).

APC has four potential N-glycosylation sites: Asn 97 on the light chain and Asn 248, Asn 313 and Asn 329 on the heavy chain. The fourth site does not have the typical Asn-X-Ser/Thr sequence; a cysteine replaces the usual Ser or Thr residue. About 30% of plasma protein C, denoted β Protein C, is not glycosylated at the fourth site [86]. There is also a γ-form of APC that is glycosylated only at Asn 97 and Asn 313 [87].
Prostate specific antigen (PSA)

Prostate specific antigen (PSA), one of the most abundant serine proteases in seminal fluid secreted by the human prostate epithelium [88,89], is a member of the kallikrein family of serine proteases where it is known as human glandular kallikrein-3 (hK-3). The name PSA derives from the highly restricted expression in prostate epithelial cells in men. It has been proven that PSA is also present at very low concentrations in male and female serum; PSA in females is related to female androgen excess [90]. PSA is a serine protease with biochemical properties similar to the protease involved in blood clotting. PSA degrades the seminal vesicle proteins seminogelin I and II, resulting in liquefaction of the seminal coagulum [91]. PSA is also connected to the prostate and prostate carcinoma and has become the most important tumor marker for diagnosing and monitoring patients with prostate cancer [92,93]. An increase of PSA in serum can be caused by prostate carcinoma and by nonmalignant prostatic alterations [94,95].
PSA is a 33-kD single chain glycoprotein and occurs at concentrations of 15-60 \( \mu \text{M} \) in seminal fluid [96-98]. PSA has one potential N-glycosylation site at Asn 45. Ser 69, Thr 70, and Ser 71 are possible O-glycosylation sites [99]. PSA has been demonstrated to exhibit proteolytic activity similar to that of chymotrypsin and trypsin, and the His-Asp-Ser charge-relay catalytic triad is also similar to that of chymotrypsin-like serine proteases.

![Figure 10. Crystal structure of prostate specific antigen drawn from PDB code 2ZCH. The residues that form the catalytic triad (in chymotrypsin's number, His57-Asp102-Ser195) are labeled in blue, pink, and green, respectively.](image)

Serine protease inhibitors (Serpins)

The serpin (serine protease inhibitor) superfamily is a group of single-chain proteins containing a conserved domain structure consisting of 370-390 amino acid residues [100]. Typical structural features include seven to nine \( \alpha \)-helices (A-I); three \( \beta \)-sheets (A, B, C); and a surface-exposed reactive center loop (RCL) (Figure 12). Serpins were first found as a set of proteins that are able to inhibit serine proteases and are extremely widely distributed among eukaryotes. So far, 36 serpins have been identified in human and are found to play important roles in highly regulated physiological processing, such as inflammation, blood coagulation and fibrinolysis [101]. Of the 36 known human serpins, at least seven are non-inhibitory, meaning that they are not known to inhibit any proteases. The non-inhibitory serpins include hormone-binding serpins such as corticosteroid-binding globulin and thyroxine-binding globulin [102,103].

In the native conformation, the reactive center loop lies on the surface of the serpin and is thus exposed to the solvent. The reactive bond that interacts
with the active site of target proteases is located on the loop. The length of this loop shows only small variation in inhibitory serpins, particularly on the N-terminal side of the reactive bond that is usually 17 residues, but 16 residues for a few serpins. On the C-terminal side, the length is less constrained and has been found to be from five to nine residues long [102].

Inhibitory serpins, including protein C inhibitor (PCI) and antithrombin, share a unique suicide mechanism for protease inhibition, outlined in Figure 11. First, the protease recognizes and binds to the RCL of the serpin to form an initial non-covalent Michaelis complex. The initial encounter complex, with forward rate constant of $k_1$ and back rate constant of $k_{-1}$, is a reversible formation. Then the reactive bond of the serpin, denoted P1-P1’, is cleaved to form a covalent acyl intermediate complex with rate constant $k_2$. After cleavage, one pathway is the release of serine protease from the cleaved serpin, which occurs with overall rate constant $k_3$. In another pathway, the amino terminal portion of the RCL inserts into $\beta$-sheet A to form an additional strand, while dragging the serine protease along the surface of the serpin to the other pole. This occurs with the rate constant $k_4$. Instead of being released from the inhibitor, the protease remains attached to the serpin, and this complex is rapidly cleared from the circulation [102]. During the translocation, there is still chance for the protease to finish the substrate reaction and escape, with a rate constant $k_5$ (Figure 11). The serpin-bound protease may be released with rate constant $k_3$. It has been estimated that $k_3$ is at least five to seven orders of magnitude smaller than $k_5$ [102], and in a predominant serpin inhibitory reaction, $k_3<<k_4$. A prevailing simplified model of investigating a second-order reaction, for instance the serpin inhibition of a serine protease, uses a pseudo first-order condition. As in the protease-inhibitor association, if the initial inhibitor concentration is much higher than the initial protease concentration, the inhibitor concentration remains effectively constant during the course of the reaction, and only the protease concentration changes with respect to time.

Some inhibitory serpins found in vertebrate blood plasma are activated by binding to glycosaminoglycans (GAGs) or other linear negatively charged molecules. These serpins are termed heparin-binding serpins and include the blood coagulation and fibrinolysis regulators antithrombin, heparin cofactor II, PAI-1, PCI, and protease nexin 1. The resulting enhancement of protease inhibition can be up to several thousand-fold [104,105]. There are two principal mechanisms behind this enhancement: (1) a bridging effect in which the linear GAG binds both to the protease and the serpin and thus brings them together in an appropriate orientation for effective interaction of the serpin RCL with the protease active site, and (2) a conformational change-based mechanism in which GAG binding to the serpin results in a change in the serpin’s conformation, rendering the inhibitor more reactive toward the protease. Antithrombin and heparin cofactor II employ both, while PCI, PAI-1, and protease nexin 1 use only the bridging mechanism
Antithrombin, heparin cofactor II, PAI-1, and protease nexin 1 utilize their D-helices for heparin binding. Antithrombin additionally uses basic residues in the adjacent A-helix and in a region immediately N-terminal to the D-helix that forms a new short helix upon heparin binding [106]. PCI is the outlier, as it is thought to utilize helix H and a basic N-terminal extension [107]. Binding of heparin to helix H of PCI has been confirmed by mutagenesis studies [108,109]. The involvement of N-terminal residues in heparin binding was suggested by demonstrating that an anti-PCI antibody that specifically binds the N-terminus blocks heparin binding [107].

A characteristic of such a bridging effect is the bell-shaped dependence of the rate of acceleration on the heparin concentration [110]. This results from binding of both serpin and protease to the same GAG chain at low GAG concentrations, but binding of each protein to a separate chain at very high GAG concentrations [102].

![Figure 11. Schematic of the serpin mechanism of protease inhibition. A serine protease (blue) recognizes and binds to the RCL of the serpin (white) to form a Michaelis complex with a forward rate constant $k_1$ and back rate constant $k^{-1}_1$. The reactive bond P1-P1' is cleaved by the serine protease, forming an acyl intermediate with overall rate constant of $k_2$. The amino-terminal portion of the RCL inserts into $\beta$-sheet A (red arrows symbolize $\beta$-sheet A) to form an additional strand, dragging the serine protease along the surface of the serpin to the other pole, with rate constant $k_4$. The protease-serpin complex is rapidly cleared from circulation. During translocation, there is still a chance for protease to finish the substrate reaction, with a rate constant $k_3$. $k_5$ reflects the effectiveness of serpin trapping the protease. Adapted from Dr. Sophia Schedin Weiss's picture.](image)
Protein C inhibitor (PCI)

PCI, also known as plasminogen activator inhibitor 3, is a glycoprotein made up of 387 amino acids with a molecular weight of 57-kD. It is produced by several cell types and found in many tissues. In human, PCI is synthesized in liver and various other tissues such as kidney, testis, and seminal vesicle [111-113], and thus exists in blood, skin, urine, seminal plasma and other secretions. The concentration of PCI in blood, seminal plasma and urine is about 100 nM [114], 3-4 μM [113,115] and 4-10 nM [116], respectively.

PCI in hemostasis

The name PCI derives from the ability to inhibit the anticoagulant activated protein C in the circulation [117], which was the first function described for PCI. Later it was found to also inhibit a variety of other serine proteases, including blood coagulation factors [118,119], fibrinolytic enzymes [120], tissue kallikreins [121], and the seminal proteases [122,123]. In blood coagulation, PCI acts as an anticoagulant by directly inhibiting thrombin, factor Xa, factor XIa and plasma kallikrein [118,119], and as a procoagulant by inhibiting activated protein C [117,118,124] and thrombin bound to thrombomodulin [125]. Blood coagulation is a complicated process and involves a balance between procoagulant, anticoagulant, and fibrinolytic mechanisms; in each system a role for PCI has been suggested.

The lack of documentation of patients with abundancy, deficiency, or specific mutation of PCI makes it difficult to determine the exact physiologic function of PCI in plasma. However, high or low PCI levels have been described in association with certain diseases. For instance, a previous study showed that plasma PCI levels are elevated in a group of survivors of myocardial infarction [126]. The Leiden Thrombophilia Study, a case-control study of more than 450 patients with a first deep vein thrombosis, indicates that high levels of PCI may constitute a mild risk factor for venous thrombosis [126]. Activation of the protein C pathway is a sensitive marker for the activation of the coagulation system, and the APC generated in plasma forms complexes with PCI, which is present in excess over APC. Therefore, the presence of APC-PCI complexes in plasma could be a marker for the detection of intravascular thrombin formation. APC-PCI complex levels are significantly elevated in different groups of patients with atherosclerotic diseases [127]. Complexes of PCI and APC have also been found in plasma samples from patients with disseminated intravascular coagulation [115].

PCI in fertilization

PCI has been identified both in the human male and female reproductive tracts. The concentration of PCI in follicular fluid is similar to that in
plasma, while the concentration of PCI in seminal plasma is 40-fold higher [113]. Immunohistochemical staining has shown that PCI is present almost throughout the male reproductive system including testis, epididymis, prostate, seminal vesicle [113], and acrosomal regions of human spermatozoa [128]. Among all these regions, seminal vesicle has the most intense stain. Moreover, PCI antigen is absent in the seminal plasma from the patient with dysfunctional seminal vesicle [113]. No PCI immunoreactivity was seen in the intraluminal secretions of the prostate. Seminal plasma of vasectomized patients appears to have normal PCI concentration, suggesting that the epididymal gland is not a major source of PCI. These findings imply that the predominant origin of seminal plasma PCI is the seminal vesicle, which are a pair of simple tubular glands behind and below the urinary bladder in males. The excretory duct of seminal vesicle gland opens into the vas deferens as it enters the prostate gland. Espana et al. have shown that PCI in the seminal vesicle secretion is 80-95% functionally active and the activity is not decreased within several hours [129]. When PCI was determined several hours after ejaculation, its activity was less than 3% [130]. Given that most seminal enzymes have a prostatic origin, it is likely that the seminal PCI is inactivated by prostatic enzyme(s) after ejaculation [130]. Seminal PCI has been found in complexes with PSA [129,131,132]; t-PA; and u-PA [116] in the semen. Although the function of PCI in seminal plasma is not completely known, published evidence shows that PCI plays a significant role in male fertility. In contrast to human, PCI is exclusively expressed in the reproductive tract in mouse and rat [133,134]. PCI knockout mice appear to be healthy, but males of this genotype are infertile due to abnormal spermatogenesis as the Sertoli cell barrier is destructed [134]. PCI is also capable of inhibiting acrosin, a serine protease stored in its zymogen form in the acrosome of spermatozoa. Different groups have reported the localization of PCI in the acrosomal membranes of sperm heads, indicating a scavenger role of PCI for prematurely activated acrosin, to protect intact surrounding cells and seminal plasma proteins from possible proteolytic damage [122,135]. In a clinical investigation, the inhibitory activities of PCI towards u-PA and t-PA were absent in two infertile patients, suggesting that the formation of PCI complexes with u-PA and t-PA plays a role in human fertilization [136]. Given that the physiological role of PSA is the degradation of the major proteins of seminal coagula, Sg-I and Sg-II, PCI also appears to be involved in regulating semen liquefaction [137]. Moreover, PCI has been found to inhibit the binding and penetration of human sperm to zona-free hamster oocytes [128,138]. The inhibitor thus appears to be necessary for several steps in fertilization.

Other functions of PCI
Recent studies suggest that PCI has antimicrobial activity. PCI was found to be able to penetrate cells, and this cellular internalization necessarily
requires phosphatidylethanolamin (PE) [139]. Treating *E. Coli* or *Streptococcus pyogenes* bacteria with PCI triggers membrane disruption followed by the efflux of bacterial cytosolic contents and bacterial killing [140]. The antimicrobial activity of PCI is thought to depend on a region between helix H and D of the serpin, which is close to its heparin-binding site [139,140].

**Glycosylation of PCI**

Human plasma PCI has three predicted *N*-glycosylation sites at Asn 230, Asn 243 and Asn 319 [141] (Figure 12). Two potential O-glycosylation sites at Thr 20 and Ser 39 were proposed by Suzuki [141]. Neither the composition nor the structures of *N*- and O-glycans of native PCI has been studied. Interestingly, molecular cloning and sequence analysis studies show that monkey PCI has three potential *N*-glycosylation sites as does human, whereas mouse and rat PCI have only one potential *N*-glycosylation site at Asn 230. Monkey has the same potential O-glycosylation sites as human, while mouse and rat lack the residue corresponding to Thr 20 in human [133,142,143].

**Cofactor enhanced PCI inhibition**

In addition to heparin, thrombomodulin has been found to be a potent cofactor that accelerates inhibition of thrombin by PCI. Thrombomodulin is

---

Figure 12. Crystal structure of human native PCI drawn from PDB code 2OL2 in a standard orientation with the hot pink β-sheet A facing the back. The three *N*-glycosylation sites (Asn230, Asn243, Asn319) are marked in blue. The RCL is in green, with the reactive site Arg354-Ser355 labeled in red. Unlike antithrombin and some other serpins, PCI utilizes helix H (cyan) instead of helix D (grey) for heparin binding. The N-terminus (not crystallized), in connection with helix A (yellow), is also thought to contribute to heparin binding of PCI.
an integral membrane protein expressed on the surface of endothelial cells. It consists of a lectin-like N-terminal region, followed by six contiguous EGF domains, an O-glycosylated Ser/Thr-rich domain, a transmembrane domain, and a short cytoplasmic tail [144]. The EGF domain 456 is responsible for the cofactor activity of thrombomodulin. It has been reported that thrombomodulin enhances the PCI inhibition of thrombin by two orders of magnitude [125]. Thrombomodulin and heparin bind to different exosites on thrombin, whereas they bind to the same or similar sites of PCI [145,146].
Present investigation

Compared to other serpins, little is known about the fundamental function of human PCI, an unusually ubiquitous and multifunctional protease inhibitor, due to the lack of documentation of patients with PCI defects. Moreover, proper animal models are lacking because PCI is not normally expressed in mouse blood. Many functional studies have been done using recombinant PCI, resulting in a lack of knowledge about the functions of posttranslational modifications of PCI. The aim of this study was to gain a deeper understanding on the driving force for PCI towards various functions and the roles that the posttranslational modifications play in PCI inhibition of proteases.

Aims

- To characterize PCI derived from the two most abundant resources in human, i.e. blood plasma and seminal plasma.
- To investigate the existence and occupancy of posttranslational modifications such as N- and O-glycosylation and proteolytical processing on PCI.
- To investigate the influence of posttranslational modifications of PCI on inhibiting various proteases including thrombin, factor Xa, and PSA, in the presence or absence of cofactors.
Results and discussion

Paper I

N-glycans and the N-terminus of protein C inhibitor affect the cofactor-enhanced rates of thrombin inhibition

PCI was purified from a pool of human blood plasma samples by three consecutive chromatographic steps, two of which employed immobilized monoclonal antibodies against PCI. SDS-PAGE revealed at least six PCI bands. Mass spectrometric analysis confirmed the purity of these six bands and verified that the heterogeneity of the protein was caused by differences in N-glycan structures, N-glycosylation occupancy, and the length of the N-terminus. Approximately 3% of total PCI lacked the N-glycan attached at Asn 243 and 18% was cleaved at the N-terminus, lacking the six most N-terminal amino acid residues, HRHHPR. A previous study on recombinant N-glycosylated PCI mutants showed that Asn 243 is required for optimal rates of thrombin and plasma kallikrein inhibition [147]. In other studies, the positively charged N-terminal region (also called the A+ -helix region) has been proposed to contribute to heparin binding to PCI [107,148]. Our results thus suggest that the presence of a natural PCI form lacking the Asn 243-linked glycan and the N-terminally cleaved form are functionally significant.

Structural analysis of the N- and O-glycans by mass spectrometry revealed the presence of bi-, tri-, and tetra-antennary types of N-glycan structures and core-1 type of O-glycan structures in human blood-derived PCI. The large distribution of various N-glycan structures verified that these contribute to the size heterogeneity of PCI. In contrast, the differences in molecular weight for the variable O-glycan structures are insufficient for causing size heterogeneity in the SDS-PAGE system used.

A method was developed to isolate N-terminally cleaved PCI from the full-length form by heparin chromatography. We were therefore able to do a number of kinetic measurements on thrombin inhibition by N-terminally-cleaved and N-glycan-free PCI in the presence or absence of cofactors. The second-order rate constants of thrombin inhibition by native full-length, native Δ6-N-cleaved, deglycosylated full-length, and deglycosylated Δ6-N-cleaved PCI were found to be similar, leading to the conclusion that neither the Δ6-N-region nor the N-glycans of PCI have any major effects on thrombin inhibition in the absence of cofactors. In the presence of varying concentrations of heparin, the dependence of the pseudo first-order rate constant on the heparin concentration followed bell-shaped curves for inhibition of thrombin by all PCI variants, suggesting that the enhancing effect of heparin is due to a bridging effect. The Δ6-N-region enhanced the maximal heparin-catalyzed second-order rate constant three-fold, whereas the presence of N-linked glycans reduced the maximal second-order rate
constants approximately two-fold. In contrast, the thrombomodulin-enhanced rate of thrombin inhibition was two-fold enhanced by the Δ6-N-region, but was unaffected by the removal of N-glycans. The N-terminal region, also called A+ε-helix region, has been proposed to contribute to heparin binding in some studies. Positively charged amino acids in this region have been implicated in heparin binding [107]. A synthetic peptide corresponding to the N-terminal region was found to compete with PCI for heparin binding in one study [148], but was found not to compete in another [149]. According to a study where recombinant PCI expressed in E.coli was used, the N-terminal region did not contribute to heparin affinity [108,150]. In contrast, another study using PCI expressed by baby hamster kidney cells found that the N-terminal region decreased the affinity for heparin [148].

Since our results demonstrated that PCI also carries a limited repertoire of short O-glycans, the factors that can differ between the expression systems are several and include the occupation and structures of N- and O-linked glycans and the length of N-terminus. Such factors presumably contribute to the opposite results of previous studies.

Thrombin inhibition by PCI variants was also determined at increased temperature or in a condition where Na⁺ was replaced by K⁺. Increased temperature elevated the uncatalyzed and thrombomodulin-enhanced rates of thrombin inhibition for all variants and enhanced the heparin-catalyzed inhibitory rates for all variants, except the one lacking both the N-glycans and the N-terminus. Moreover, the full-length variant reached the highest heparin-enhanced rate in the presence of K⁺, whereas the Δ6-N-cleaved form reached the highest heparin-enhanced rate in the presence of Na⁺. These findings may reflect different regulatory mechanisms for these two variants, suggesting that the cofactors and posttranslational modifications and the solution conditions contribute to bringing specificity to this serpin.

**Paper II**

*Further insight into the roles of the glycans attached to human blood protein C inhibitor*

This study was undertaken to investigate whether the heterogeneity of PCI exists in single individuals and to identify the O-glycan attachment site and its potential heterogeneity. We found that the micro-heterogeneity was present in single individuals and conserved between donors of Asian and European extraction, further supporting the argument that the micro-heterogeneity of PCI has some functional importance.

PCI has two predicted O-glycosylation sites on Thr 20 and Ser 39 [124]. We have, for the first time, provided experimental evidence that PCI is O-glycosylated on Thr 20 with the O-glycan structure composition NeuAcGalGalNAc. The N-terminal region of PCI, including Thr 20, has been reported to be highly flexible [150] and has never been visualized in
any of the PCI crystal structures. Therefore, we modeled a part of the N-terminal region including Thr 20. Two alternatives for the secondary structure in this region seemed possible, i.e. a coil sticking out from the surface or an α-helix, but the latter was excluded because the A-helix of PCI has been proposed to be shortened by two turns compared to other serpins and thus begin at Asp 28 [151]. One function of the O-glycan at Thr 20 may thus be to affect the secondary structure by preventing the formation of α-helical structure in this region, which in turn presumably affects the interactions of PCI with several ligands. Ligands that are believed to interact with PCI in proximity to the N-terminal region include retinoic acid and phosphatidylethanolamine (PE). Docking studies have placed the binding site for these lipids in a cavity at the N-terminal side of helix A, which is created by the shortening of this helix [139,151]. Furthermore, several previous studies have proposed that the N-terminal sequence of PCI, in addition to helix H, contributes to heparin binding [110,148]. Finally, the thrombomodulin-enhanced rate of thrombin inhibition by PCI is affected by the N-terminal sequence of the inhibitor, as we showed in paper I. The O-glycan at Thr 20 may thus affect several PCI-ligand interactions with biological implications.

Paper III

Heparin enhances the inhibition of factor Xa by protein C inhibitor in the presence, but not the absence, of Ca^{2+}

Previous studies have shown that heparin has no marked effect on the rate of factor Xa inhibition by PCI in the absence of Ca^{2+} [149]. The aim of the present study was to investigate the effects of the combination of Ca^{2+} and heparin on the rates of factor Xa inhibition by two PCI variants.

Two forms of PCI from human blood plasma, i.e. full-length and Δ6-N-cleaved, were used, because we showed in paper I that the N-terminus contributes to heparin binding. The effects of these PCI variants on the inhibition of factor Xa were studied in the absence or presence of physiological concentration of Ca^{2+}. In the absence of Ca^{2+} and heparin, factor Xa inhibition by the full-length form was four-fold more efficient than by the Δ6-N-cleaved form. In the absence of Ca^{2+}, heparin did not affect the rate of PCI inhibition of factor Xa. In the presence of Ca^{2+}, however, the dependence of the rate of factor Xa inhibition by PCI on heparin concentration followed bell-shaped curves, for both forms of PCI. Moreover, both the peak height of the bell-shaped curve and the heparin concentration at which this peak was observed differed for the two forms of PCI, suggesting that the N-terminal sequence contributes to the affinity for heparin and to the inhibitory mechanism for PCI.

Ca^{2+} has previously been found to affect the heparin-enhanced rate of antithrombin inhibition of factor Xa by promoting the formation of the
antithrombin-heparin-factor Xa ternary complex [71,152]. Unlike PCI, heparin enhances antithrombin inhibition of factor Xa by both allosteric and bridging mechanisms. The bridging effect is present only to a small extent in the absence of Ca$^{2+}$ but increases in the presence of Ca$^{2+}$ [71,153]. The Ca$^{2+}$-induced heparin enhancing effect on antithrombin inhibition of factor Xa [70] is highly similar to the 14-fold Ca$^{2+}$-induced heparin enhancement observed here for the rate of factor Xa inhibition by full-length PCI. Thus, Ca$^{2+}$ presumably affects the heparin-enhanced PCI inhibition of factor Xa by affecting the conformation of factor Xa so that it exposes its exosite for heparin, making the bridging mechanism more efficient, as for antithrombin.

**Paper IV**

*N-glycans of human seminal plasma protein C inhibitor: tissue specific expression and function*

PCI exists in many tissues and secretions in human, among which seminal plasma has the highest PCI concentration. Although the exact function of PCI in semen is not clear, both clinical and preclinical studies have reported the importance of PCI in relation to fertilization [128,134,136,138]. In this study, PCI was purified from seminal plasma that originated from several donors, using basically the same procedure as for blood PCI purification. The N-glycan profile of seminal plasma PCI was determined by mass spectrometry. The most abundant N-glycan structure has a composition of Fuc$_3$Hex$_5$HexNAc$_4$, consistent with a core fucosylated bi-antennary glycan with terminal Lewis$^a$. Recent studies have shown that the sperm-oocyte binding and penetration is inhibited by PCI [135,138] and that sperm binding to egg through one of the glycoproteins ZP-3, which forms the zona pellucida, is inhibited by glycans that contain Lewis$^a$ in a competitive manner [154,155]. Together, these findings indicate that the Lewis$^a$ on PCI may be of biological significance and involved in the receptor binding of PCI.

Moreover, mass spectrometric analysis showed that PCI purified from seminal plasma is totally inactivated, forms a complex with PSA to a large extent, and is cleaved at both the N- and C-terminal sides. The N-glycan structures differed as compared to those of both blood-derived and urinary PCI, providing evidence that the N-glycans of PCI are expressed in a tissue-specific manner (see Figure 13 for comparison of major N-glycan structures of PCI derived from the three different tissues) Second-order rate constants for PSA inhibition by PCI were $4.3\pm0.2$ and $4.1\pm0.5$ M$^{-1}$s$^{-1}$ for the natural full-length PCI and a form lacking six amino acids at the N-terminus, respectively. Moreover, second-order rate constants were $4.7\pm0.1$ and $34\pm0.3$ M$^{-1}$s$^{-1}$ for the PNGase F-treated full-length and Δ6-N-terminal-cleaved PCI, respectively, demonstrating that the rate of PSA inhibition was not affected by removing N-glycans or the six-residue N-terminal peptide.
alone, whereas the combination of overall removal of N-glycans and the N-terminal peptide resulted in a seven- to eight-fold increase in the rate of PSA inhibition by PCI. This finding suggests that these posttranslational modifications together are responsible for the slow PSA-PCI reaction velocity. A reasonable speculation is that the combined removal of one or more of the N-linked glycans and the N-terminal peptide gives rise to a new or improved exosite interaction for the protease.

![Figure 13. Tissue-specific expression of N-glycans on PCI. Major N-glycans of PCI derived from blood plasma, seminal plasma and urine are framed in light red, blue and yellow, respectively. N-glycan structure of urinary PCI is drawn according to [156]. * represents the most abundant glycan. Glycan marked with “S” has a special lacdiNAc (GalNAcβ1-4GlcNAc) structure (constitutes about 12% of the total N-glycans in urinary PCI [156]). This structure has not been observed in blood or seminal plasma PCI N-glycans.](image)

**Conclusions**

In this study we have, for the first time, provided a full explanation for the marked size heterogeneity of blood-derived PCI in human and shown that it exists in single individuals of different ethnicities. N-glycans of blood PCI were composed of bi-, tri-, and tetra-antennary complex structure. Fucose residues were identified both on the core GlcNAc and as parts of sialyl-Lewis* epitopes. PCI was N-glycosylated at all three potential N-glycosylation sites, Asn 230, Asn 243, and Asn 319, but a small fraction of PCI lacked the N-glycan at Asn 243. O-glycans of PCI were more than 95% composed of either mono- or di-sialylated core type 1 structures. We
provided evidence that one of the predicted O-glycosylation sites, Thr 20, is occupied with the O-glycan structure NeuAcGalGalNAc. A PCI variant that lacks six amino acids attached at the N-terminus was identified.

The N-glycans of PCI did not contribute to the rate of thrombin inhibition in the absence of cofactors or presence of thrombomodulin, but affected the heparin-enhanced rate of thrombin inhibition. N-glycans had no effect on PSA inhibition by PCI.

The N-terminus of PCI did not affect the rate of thrombin inhibition without cofactors, but increased both the heparin- and thrombomoduline-enhanced rate of thrombin inhibition. In contrast, the N-terminus of PCI increased the rate of inhibition of factor Xa both in the absence and presence of heparin. PSA inhibition was not affected by N-glycans or N-terminus solely, but was markedly enhanced by the combined removal of the N-glycans and the N-terminal peptide. See Figure 14 for a summary of the conclusion.

Interestingly, the N-glycan structures of PCI were expressed in a tissue-specific manner. The N-glycans of seminal plasma PCI were mainly core-fucosylated bi-antennary Lewis^x and/or Lewis^y-capped structures completely devoid of sialic acids. In a previous study, urinary PCI glycans were reported to contain lacdiNAc (GalNAcβ1-4GlcNAc) antennae, a sequence that was
observed in either blood or seminal plasma PCI. These findings suggest that the N-glycosylation of PCI displays a highly tissue-specific expression (Figure 13).
These findings provide evidence that the posttranslational modifications of PCI are important for its inhibitory specificity towards different proteases.
Future perspectives

Although found decades ago, the exact function of PCI has not been fully understood, due to its multifunctional ability and the complex regulatory effect on it. By investigating the properties of posttranslational modifications on PCI derived from its original biological sources in human, we were able to demonstrate some functional importance of posttranslational modifications for PCI functions. Further experiments producing PCI variants expressing designed posttranslational modifications or modified glycan structures or N-terminus, e.g. sialyl-Lewis epitopes, in appropriate eukaryotic cell systems will be helpful for a deeper understanding. Sialyl-Lewis epitopes were found in a small portion of the glycans of PCI derived from blood, and the low concentration made purifying this form difficult. In PCI derived from seminal plasma, Lewis\(^X\) and Lewis\(^Y\) were identified. Recombinant PCI expressing these epitopes would help understand the effects of various Lewis-containing epitopes on PCI cell binding and interactions.

Crystal structures of PCI in various complexes have been made for the purpose of understanding the binding feature of this serpin to target proteases. However, none of the crystal structures have defined the structure of the N-terminus. We have, for the first time, confirmed the occupancy of O-glycan on the N-terminus and proposed that the location may be of great importance, as it is close to several ligand-binding sites for the inhibitor. Identifying the structure of O-linked glycans allows their enzymatic removal for functional studies. However, no enzyme is comparable to PNGase F for removing intact O-linked sugars. Monosaccharides may be sequentially hydrolyzed by a series of hexoglycosidases. We have attempted to use a combination of neuraminidase, \(\beta(1-4)\)galactosidase, \(\beta\)-N-acetylhexosaminidase and O-glycanase; however, conditions need to be optimized to achieve complete removal of O-glycans in future work. Moreover, mutation of the Thr 20 site in a recombinant PCI expressed in an appropriate system would help address the exact function of O-glycan attached to this site.

PCI derives its name from the ability to inhibit activated protein C, which was the first function described for the inhibitor. PCI is a remarkably slow inhibitor of APC. The reaction is accelerated 20-200-fold in the presence of heparin and other sulfated polysaccharides, but still slow compared to that of other serpins; for instance, the inhibition of thrombin by antithrombin. We
therefore tried repeatedly to study the effects of the PCI variants on the inhibition of APC, because the inhibition of APC by PCI under pseudo first-order conditions displays a multi-phasic behavior (not included in the papers). In other words, the inhibition seems to be composed of a fast and slow inhibition phase. This phenomenon, however, has never been reported before for PCI inhibition of APC. This multi-phasic inhibitory behavior of PCI toward APC observed throughout our measurements is presumably caused by heterogeneous APC. Such heterogeneity could, for instance, be explained by conformational variants and/or glycosylation variants of APC in the preparations. Determining the cause of the multi-phasic behavior would improve kinetic determinations, and may lead to important medical implications, since the fast phase is much faster than published values for the overall phase.

In this study, we conducted PCI inhibition of thrombin, factor Xa, PSA, and APC; however, PCI has inhibitory specificity towards many other proteases. It would be intriguing to investigate the effect of different PCI variants on the inhibition of other proteases such as urokinase and kallikreins and seminal plasma proteases such as acrosin in the presence and absence of cofactors. Such experiments will be important to further understand the protease-specific regulatory function of posttranslational modifications of PCI.

The different profiles of glycosylation of PCI in semen, blood, and urine are highly interesting. This tissue-specific expression of PCI presumably has different functions in different fluids. Unfortunately, the purified seminal plasma PCI isolated was cleaved and partially in complex with seminal proteases. Attempts to purify active seminal PCI should thus be made to investigate the functional differences of blood and seminal plasma PCI.

Glycans such as bi-antennary structures can take a considerably big proportion of the protein, both in weight and volume. For PCI, the actual and calculated (based on the primary structure) molecular weights are 57- and 43-kD, respectively. Since glycans are located on the protein surfaces, the relative contribution of glycans to the surface properties is probably much higher than their relative contribution to the molecular weight. Unfortunately, current techniques do not allow the crystallization of these glycans. Modeling or crystallizing of the protein with glycans of native structure would greatly improve the understanding on the importance of these attached sugars.
摘要

这篇论文研究了蛋白C抑制物 (protein C inhibitor, 简称PCI) 的翻译后修饰对其的多功能性的影响。PCI是一种丝氨酸蛋白酶抑制物 (serine protein inhibitor,简称serpin)。PCI在人体很多器官中表达，并被分泌到血液、精液、尿液等各种体液中。在血液中，PCI既有凝血作用又有抗凝作用，并且还被广泛认为起到病原防御的作用。在生殖系统中，PCI还在很多步骤上调控生殖过程，其中包括受精。PCI对酶的广泛的抑制作用，加上其自身既有凝血作用又有抗凝作用的矛盾特性，使我们至今不完全了解其在生理过程中的确切地位。在该项课题中，我们运用亲和层析技术从血液及精液中提纯了PCI，我们发现血液中的PCI具有高度的异质性。这种异质性是由蛋白的不同翻译后修饰引起的。我们用质谱技术研究了血液PCI的N-和O-糖基化的位点和结构以及精液PCI的N-糖基化位点及结构，并确定了血液PCI在Thr-20位点具有O-糖基化。研究结果显示血液及精液中的PCI糖基化结构不同，证明了PCI在人体中的表达呈组织特异性。不同来源的PCI的氨基末端还被不同程度的水解。在这个基础上，我们用测量显色底物以计算酶活的方法研究了N-糖基和氨基末端在PCI抑制蛋白酶的过程中作用。结果显示PCI的N-糖基和氨基末端在不同程度上影响PCI对凝血酶 (thrombin), 凝血因子Xa (factor Xa) 和前列腺特异性抗原 (prostate specific antigen) 的抑制作用，说明翻译后修饰对PCI的特异性有影响。这些发现增进了我们对PCI的多功能性及其调控机制的理解，以及为可能的在医学上应用不同种类的PCI提供了依据。

关键词：蛋白C抑制物，翻译后修饰，N-糖基，O-糖基，质谱，凝血因子Xa, 凝血酶，前列腺特异性抗原
Acknowledgements

I would like to express my sincere gratitude to the following people:

My supervisor Sophia Schedin Weiss, for introducing me to the world of science with excellent guidance, great patience, support and encouragement. I feel so lucky for being your student.

All the collaborators and friends in Imperial college London, Lund University and Uppsala University: Prof. Anne Dell, for accepting me to learn MS in your lab. I learned so much in those two weeks. Margareta Kjellberg, for sharing your great knowledge about PCI and everything, this work would never have gone this far without your help. Prof. Johan Stenflo, for kindly letting me purify PCI in your lab. Aris and PohChoo, for your friendship and patient answers to all my questions. Åke Engström, for being very helpful and efficient in the collaboration. Julius Hreinsson, for collecting the experimental sample, and being very enthusiastic on our project.

My examiner Erik Fries, for the nice talk and advice on my project. Prof. Ulf Lindahl, Jin-ping Li, Dorothe Spillmann, Lena Kjellén, Birgitta Tomkinson, Per Jemth, Pia Ek, Liya Wang and my supervisor’s supervisor Prof. Ingmar Björk, for your help, giving advices and being my role models of great scientists.

Previous and present roommates: Sandra and Susanne for your friendship, nice chat, and company. Anna, for being like sunshine, you always know how to make me laugh.

Numerous colleagues and friends in B9:4 and D9:4 corridor, and our very helpful administrative staff, for the memorable time we spent, helping me in different ways, no one is forgotten, nothing is forgotten. I am going to miss every one of you when I leave IMBIM.

My Chinese friends in Uppsala, especially the badminton pals, you know who you are, ¯\_(ツ)_/¯ instead of thanking you, I want to thank the destiny for bringing you to me. I don’t get to say this very often, but I love you all!
Zhanchun and Lijun, for your love, concern, and support.

Last but not least, I am very thankful to Sweden for bringing to me my other half and the love of my life, Hao.
References


[77] F. Fourrier, C. Chopin, J. Goudemand, S. Hendrycx, C. Caron, A. Rime, A. Marey, and P. Lestavel, Septic shock, multiple organ failure, and disseminated intravascular coagulation. Compared patterns of


[145] L. Yang, C. Manithody, T.D. Walston, S.T. Cooper, and A.R. Rezaie, Thrombomodulin enhances the reactivity of thrombin with protein C inhibitor by providing both a binding site for the serpin and


Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 601

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title “Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine”.)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-131126