Structural and Functional Studies on Proteinaceous Metallocarboxypeptidase Inhibitors

Joan López Arolas

Structural and Functional Studies on Proteinaceous Metallocarboxypeptidase Inhibitors

Structural and Functional Studies on Proteinaceous Metallocarboxypeptidase Inhibitors

Doctoral thesis presented by Joan López Arolas for the degree of PhD in Biochemistry from the University Autonomous of Barcelona

Institute of Biotechnology and Biomedicine, Protein Engineering laboratory. Thesis supervised by Prof. Francesc X. Avilés Puigvert and Dr. Salvador Ventura Zamora

Joan López Arolas

Francesc X. Avilés Puigvert

Salvador Ventura Zamora

Barcelona, March 2005

PREFACE

The present thesis consists of six independent research works that are situated in the field of metallocarboxypeptidase inhibitors: their folding, stability, structure and function are studied. For a better comprehension of the thesis, the different works have been grouped into two sections (I and II).

The first work comprises the isolation and cDNA cloning of a new carboxypeptidase inhibitor from ticks, named TCI. The recombinant form of this protein is extensively characterized in terms of stability and function, and its possible biological activity is discussed. This work was done in part in Munich, in the Department of Clinical Biochemistry of the Ludwig-Maximilians University under the supervision of Prof. Christian Sommerhoff. Drs. Julia Lorenzo, Ana Rovira and Joaquim Castellà participated in the realization of the project in Barcelona, under the direction of Prof. Francesc X. Avilés.

The second work presents the crystal structure of TCI in complex with either bovine carboxypeptidase A or human carboxypeptidase B. The structure of TCI is characterized in detail as well as its mechanism of inhibition toward metallocarboxypeptidases. Applications of the information derived from this study are discussed. This work was done in Munich between the Department of Clinical Biochemistry and the Department of Structure Research, Max-Planck Institute for Biochemistry, under the supervision of Profs. Christian Sommerhoff, Robert Huber and Tad Holak. Grzegorz Popowicz and Dr. Julia Lorenzo participated in this work, which was also supervised by Prof. Francesc X. Avilés from Barcelona.

In the third part, the role of each residue of the secondary binding site of potato carboxypeptidase inhibitor (PCI) in the folding, structure and function of this protein is determined. The collected data clarify the folding determinants of the primary and secondary binding sites of PCI and their contribution to the inhibition of carboxypeptidases. This work was done with the participation of Drs. Julia Lorenzo, Ana Rovira, Josep Vendrell and Salvador Ventura, and under the direction of Prof. Francesc X. Avilés.

The fourth work elucidates the oxidative folding and reductive unfolding pathways of leech carboxypeptidase inhibitor (LCI). The intermediates formed along the folding process are characterized by structural and kinetic analysis for a better understanding of the determinants underlying the reaction. This project was done with the participation of Drs. Sílvia Bronsoms and Julia Lorenzo, and with the collaboration of Prof. Jui-Yoa Chang from the Research Center for Protein Chemistry in the University of Texas, Houston. Prof. Francesc X. Avilés and Dr. Salvador Ventura supervised this study.

In the fifth work, the III-A intermediate, the major rate-limiting step in the oxidative folding of LCI, is directly purified from the folding reaction and structurally characterized by NMR

spectroscopy. The results obtained in this study are of great interest in the field of disulfide-rich proteins and folding intermediates. Theoretical approaches based on topological constrains are also used here to predict the formation of folding intermediates. This work was done in part in Munich, in the Department of Structure Research of the Max-Planck Institute for Biochemistry, under the supervision of Prof. Tad Holak. Loyola D'Silva and Grzegorz Popowicz participated in the realization of this work in Munich. Prof. Francesc X. Avilés and Dr. Salvador Ventura directed the project in Barcelona.

The last work deals with the other major folding intermediate of LCI: III-B intermediate. In this work, an analog of this intermediate is constructed and extensively analyzed. The derived data allows a better understanding of the folding process, conformational stability and functionality of LCI. This work was also done in part in the Max-Planck Institute of Munich under the supervision of Prof. Tad Holak and with the participation of Grzegorz Popowicz. Dr. Sílvia Bronsoms participated in the project in Barcelona, where Prof. Francesc X. Avilés and Dr. Salvador Ventura carried out the supervision.

The six research works of this thesis are connected by a general introduction, a section of objectives, and by a general discussion and conclusions. In the first part of the thesis, the general introduction describes proteases, especially metallocarboxypeptidases: classification, physiological role, and main characteristics of their inhibitors. The rest of introduction is dedicated to general aspects of folding; in particular to folding of disulfide-rich proteins such as metallocarboxypeptidase inhibitors. The next chapters present each of the six works that constitute the thesis, including each one a summary, introduction, experimental procedures, results, discussion and references. The last part of the thesis, containing the general discussion, conclusions and perspectives, is followed by a list of bibliographic references.

The six research works presented in this thesis are the basis of the following scientific papers:

1. Arolas JL, Lorenzo J, Rovira A, Castellà J, Aviles FX & Sommerhoff CP. "A carboxypeptidase inhibitor from the tick *Rhipicephalus bursa*. Isolation, cDNA cloning, recombinant expression, and characterization". *J Biol Chem* 2005; 280: 3441-3448

2. Arolas JL, Popowicz GM, Lorenzo J, Sommerhoff CP, Huber R, Aviles FX & Holak TA. "The three-dimensional structures of tick carboxypeptidase inhibitor in complex with A/B carboxypeptidases reveal a novel double-headed binding mode". *Submitted for publication*

3. Arolas JL, Lorenzo J, Rovira A, Vendrell J, Aviles FX & Ventura S. "Secondary binding site of the potato carboxypeptidase inhibitor. Contribution to its structure, folding, and biological properties". *Biochemistry* 2004; 43: 7973-7982

4. Arolas JL, Bronsoms S, Lorenzo J, Aviles FX, Chang JY & Ventura S. "Role of kinetic intermediates in the folding of leech carboxypeptidase inhibitor". *J Biol Chem* 2004; 279: 37261-37270

5. Arolas JL, D'Silva L, Popowicz GM, Aviles FX, Holak TA & Ventura S. "NMR structural characterization and computational predictions of the major intermediate in the oxidative folding of leech carboxypeptidase inhibitor". *Submitted for publication*

6. Arolas JL, Bronsoms S, Popowicz GM, Aviles FX, Holak TA & Ventura S. "Characterization of an analog of a major intermediate in the oxidative folding of leech carboxypeptidase inhibitor". *In manuscript*

ABBREVIATIONS

αLA	α-lactalbumin
BPTI	bovine pancreatic trypsin inhibitor
CD	circular dichroism
CP	carboxypeptidase
Cys/Cys-Cys	cysteine/cystine
D/H	deuterium to proton
DTT	dithiothreitol
EGF	epidermal growth factor
GdnHCl	guanidine hydrochloride
GSH/GSSG	reduced/oxidized glutathione
IGF-1	insulin-like growth factor
K_i	inhibition constant
LCI	leech carboxypeptidase inhibitor
MALDI-TOF MS	matrix-assisted laser desorption/ionization - time of flight mass spectrometry
N	native form
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser enhancement
NOESY	NOE spectroscopy
ORF	open reading frame
PCI	potato carboxypeptidase inhibitor
PCP	pro-carboxypeptidase
PCR	polymerase chain reaction
R	reduced/denatured form
RACE	rapid amplification of cDNA ends
Rmsd	root mean square deviation
RNase A	ribonuclease A
RP-HPLC	reversed-phase high performance liquid chromatography
TAFI(a)	(activated) thrombin-activatable fibrinolysis inhibitor
TAP	tick anticoagulant peptide
TCI	tick carboxypeptidase inhibitor
TFA	trifluoroacetic acid
TM	thrombomodulin
tPA	tissue-type plasminogen activator
wt	wild-type

INDEX

GENERAL INTRODUCTION

PEPTIDASES AND INHIBITORS

Peptidases: a view of classification

Proteolytic enzymes, also known as peptidases or proteases[*], catalyze the hydrolytic cleavage of peptide bonds in other proteins. Some peptidases are involved in intra- and extra-cellular protein digestion, but most of them perform much more specialized tasks such as activation of zymogens, processing of hormones and neuropeptides from precursors, translocation through membranes, trimming of assembling proteins or activation of receptors (Neurath, 1984). Thus, peptidases are essential for a great variety of biological actions in living organisms, ranging from bacteria to mammals, acting in alimentary processes, regulation of blood coagulation/fibrinolysis, tissue remodeling, apoptosis, inflammation, defense mechanisms, fertilization, and many others (Hook et al., 1994; Peters, 1994; Kalafatis et al., 1997; Thornberry & Lazebnik, 1998; Creagh et al., 2003; Kuzuya & Iguchi, 2003). Besides the physiological necessity, many of these peptidases are potentially hazardous to their natural environment therefore their proteolytic activity must be carefully controlled by the cell/organism. This control is normally achieved by regulated expression/secretion and/or activation of the pro-peptidases, by specific degradation of the mature enzymes, and by blockage of their proteolytic activity through inhibition (Bode & Huber, 1992). As indicated by the accumulation of high levels of inhibitors in various body fluids, this latter regulation mode through inhibition plays a key role in many cells, tissues and organisms.

The accelerating rate of research on peptidases is being rewarded by a high rate of discovery. At the present time about 700 distinct peptidases, including over 200 that are expressed in mammals, are recognized and new ones are being discovered almost daily. This fact makes difficult the task of classification and nomenclature of these enzymes, which is currently managed by a combination of two partially overlapped systems: the EC system, recommended by the Nomenclature Committee of IUBMB, and the MEROPS system of peptidase clans and families (Barrett, 1999). Both can be found on the WWW (http://www.chem.qmw.ac.uk/iubmb/enzyme/EC34 and http://merops.sanger.ac.uk/, respectively). In both systems peptidases are organized in a three-layer classification by (i) *catalytic type*, (ii) *molecular structure*, and (iii) *individual peptidases*. In the EC system,

[*] The EC system favours the term "protease" over "peptidase", whereas MEROPS favours the latter to stress that more frequently proteolytic enzymes act on peptides rather than on proteins.

catalytic type is used to subdivide the peptidases in 14 sub-subclasses (Table 1), of which metallocarboxypeptidases (a major issue in this thesis) constitutes one sub-subclass.

Table 1. Classification of peptidases according to the EC recommendations

Sub-subclass	Kind of peptidase
3.4.11	Aminopeptidases
3.4.13	Dipeptidases
3.4.14	Dipeptidyl-peptidases and tripeptidyl-peptidases
3.4.15	Peptidyl-dipeptidases
3.4.16	Serine-type carboxypeptidases
3.4.17	Metallocarboxypeptidases
3.4.18	Cysteine-type carboxypeptidases
3.4.19	Omega peptidases
3.4.21	Serine endopeptidases
3.4.22	Cysteine endopeptidases
3.4.23	Aspartic endopeptidases
3.4.24	Metalloendopeptidases
3.4.25	Threonine endopeptidases
3.4.99	Endopeptidases of unknown catalytic mechanism

While the EC recommendations may not have much to offer in the higher levels of classification of peptidases, they certainly come into their own at the lowest level, in which individual peptidases are recognized. A name is recommended for each enzyme, and other names that may be encountered in the literature are also listed. Perhaps more importantly, each peptidase is assigned a unique EC number that serves as an unmistakable reference to this enzyme. In contrast to EC system, MEROPS has most to contribute in the upper two levels. The similarities in primary structure tend to reflect shared evolutionary origins, and a wealth of biological meaning can be extracted from this. Thus, the classification of peptidases into families is at the heart of this system. Also, it is well established that similarities in protein fold persist in evolution much longer than do close similarities in amino acid sequence, and accordingly, the folds can reveal distant relationships that cannot be seen clearly in the primary structures. Such distantly related groups are termed clans by the MEROPS system (Rawlings et al., 2002; Rawlings et al., 2004a).

Peptidase inhibitors

Peptidase inhibitors are important tools of nature for regulating the proteolytic activity of their target proteases, and for blocking them in emergency cases (Bode & Huber, 1992). All naturally occurring inhibitors directed toward endogenous peptidases, i.e. against proteolytic enzymes from the same organism, seem to be proteins. Some microorganisms, however, produce and secrete small non-proteinaceous inhibitors, which impair the proteolytic activity of host peptidases (Bode & Huber, 2000). The number of peptidase inhibitors isolated and identified so far is extremely large (Rawlings et al., 2004b). These inhibitors can have very different polypeptide scaffolds and can be grouped into a vast number of protein families, with some of them existing as bis- or multi-headed tandem proteins. In most cases, all members of a specific inhibitor (super)family are directed against the target peptidases of the same mechanistic class. Only a very few protein inhibitors exhibit "dual" activity simultaneously exerted toward peptidases from different classes; e.g. the inhibitor from sea anemone, equistatin, which inhibits papain-like cysteine enzymes and the aspartic protease cathepsin D via different domains (Lenarcic & Turk, 1999). For a few inhibitors, functions other than protease blockage have also been found such as growth factor activities, receptor clearance signaling or involvement in carcinogenesis (Clawson, 1996).

The majority of protein inhibitors known and characterized are those directed toward serine proteases (Otlewski et al., 1999; Krowarsch et al., 2003), enzymes with important biological roles in e.g. blood coagulation. These inhibitors can be grouped into at least 16 different families, based on sequence and topological similarity, and mechanism of binding. Among them, BPTI (Kunitz) family, Kazal family, serpins and hirudin have been extensively studied. Within the last few years a large number of protein inhibitors of cysteine proteases have also been discovered and characterized. The best-known family of cystein proteases is the papain family (Berti & Storer, 1995). Their inhibitors, called cystatins, are grouped into stefins, cystatins and kininogens (Turk & Bode, 1991). Recently an increasing interest has been focused on caspases and calpains, two families of cysteine proteases. The former family participate in apoptosis, and thus in the development of degenerative and autoimmune diseases like multiple sclerosis, diabetes or cancer. The latter family is involved in many pathological processes, including Alzheimer's disease, inflammation or muscular dystrophy (Thornberry & Lazebnik, 1998). Great attention has been also paid to matrix metalloproteases (Hooper, 1994). These enzymes with

endoproteolytic activity, together with their inhibitors (TIMPs), have a central role in the turnover of the extracellular matrix and are involved in many diseases such as arthritis, cancer or artherosclerosis (Nagase, 1996). The inhibitors of metallocarboxypeptidases, another family of metalloproteases, are thoroughly described in next sections. Finally, only two protein inhibitors of aspartic proteases, pepsin and equistatin, are known to date.

In principle, very different mechanisms of inhibition can be imagined, such as blocking the enzyme's active site in a substrate-like manner, docking adjacent to the active/substrate binding site, or allosterically impairing the proteolytic activity/substrate interaction of the enzyme via binding to quite distantly located enzyme exosites (Bode et al., 1999). Variations of these concepts have been observed in proteinase-protein inhibitor complex structures analyzed by crystallographic techniques. The substrate-like binding is observed in a large number of structures of serine/metallocarboxypeptidase-inhibitor complexes, while the interaction of cystatins with papain-like cysteine proteases represents a non-substrate binding mode. Protease-inhibitor complexes have been for long time, together with the antibody-antigen complexes, the principal protagonists in the studies of protein-protein interactions.

Metallocarboxypeptidases: an overview

Carboxypeptidases (CPs) catalyze the hydrolysis of peptide bonds at the C-terminus of peptides and proteins as a step in the degradation of substrate molecules or during the maturation of others (Vendrell et al., 2004). As for every type of protease, the physiological effect of their hydrolytic action is varied and site- and organism-dependent. The carboxypeptidase action may be carried out by three kinds of enzymes with different catalytic mechanisms: using an active site with serine, cysteine or zinc. The latter group is referred to as metallocarboxypeptidases and possesses a tightly bound Zn^{2+} atom, which is directly involved in catalysis (Hooper, 1996). Normally the classification of carboxypeptidases is carried out following their catalytic type, their evolutionary relationship and the type of reaction that catalyze. The data presented in Table 2 shows the group of mammalian metallocarboxypeptidases, which are included in the family M14 (Skidgel, 1996). Two subfamilies can be defined based upon sequence homology and overall structure: carboxypeptidase A and carboxypeptidase H (also named carboxypeptidase E) (Vendrell & Aviles, 1999; Vendrell et al., 2004). The degree of amino acid sequence identity is greater than 40% among members of the same subfamily and

about 20% between subfamilies. The metallocarboxypeptidases could also be subdivided upon the basis of their differential involvement in physiological processes (Vendrell & Aviles, 1999). Thus, pancreatic carboxypeptidases A (also named A1), A2 and B typically function only as digestive enzymes, whereas the rest of mammalian carboxypeptidases, including carboxypeptidase A from mast-cells (also named A3) and carboxypeptidase U, exert their action in different highly selective processing reactions, mainly in non-digestive tissues and fluids, and have been called *regulatory carboxypeptidases*.

Table 2. Mammalian metallocarboxypeptidases[a]

Subfamily	Enzyme	E.C. number	Substrate specificity	Tissue localization
CPA	Carboxypeptidase A(1)	3.4.17.1	Hydrophobic	Exocrine pancreas
	Carboxypeptidase A2	3.4.17.15	Hydrophobic	Exocrine pancreas
	Carboxypeptidase B	3.4.17.2	Basic	Exocrine pancreas
	Mast-cell carboxypeptidase A(3)	3.4.17.-	Hydrophobic	Mast-cells
	Carboxypeptidase U(R)	3.4.17.20	Basic	Plasma, liver
CPH	Carboxypeptidase H(E)	3.4.17.10	Basic	Neuroendocrine
	Carboxypeptidase M	3.4.17.12	Basic	Broad
	Carboxypeptidase N	3.4.17.3	Basic	Plasma, liver
	Carboxypeptidase D	3.4.17.22	Basic	Broad
	Carboxypeptidase Z	3.4.17.-	Basic	Placenta, embryonic tissues

[a] Other mammalian metallocarboxypeptidases such as CPA4, CPA5, CPA6, CPO, CPX1 and CPX2 have been recently discovered and are not included in this table.

All metallocarboxypeptidases from family M14, including the digestive and regulatory enzymes, may in general also be classified upon the basis of their substrate specificities (Vendrell & Aviles, 1999). Carboxypeptidase A-like enzymes have a preference for hydrophobic C-terminal residues, and B-like enzymes exert their action on C-terminal Lys or Arg residues. Remarkably, most regulatory carboxypeptidases are of the B type therefore are also called “basic carboxypeptidases”. The pancreas of different species not only contains carboxypeptidases types A and B, but also a carboxypeptidase A-like enzyme named carboxypeptidase A2, which shows a preference for C-terminal bulky

aromatic residues such as Trp (Garcia-Saez et al., 1997). The only exception in specificity among regulatory carboxypeptidases is that of mast-cell carboxypeptidase A, which shares a higher degree of homology with carboxypeptidase B than with carboxypeptidase A but its specificity is toward hydrophobic amino acids.

Digestive metallocarboxypeptidases

Pancreatic carboxypeptidases are approximately 34-36-kDa proteins found in soluble and non-glycosylated form. They are synthesized as inactive precursors containing a 94-96 residues activation segment and are stable as zymogens during storage in the pancreatic granules until secretion occurs. Trypsin-promoted limited proteolysis generates active enzymes with 305-309 residues (Vendrell et al., 1991), which are optimally active in the neutral pH range and contribute to the degradation of dietary proteins in the intestine. Allelomorphic forms have been described for pancreatic carboxypeptidases A and B. The degree of sequence homology between the pancreatic enzymes (A, A2 and B forms) varies between 42-64% in the mature enzymes (Aloy et al., 1998). However, the sequence homology is substantially lower in the pro-segment regions. Bacterial carboxypeptidases share a comparable degree of homology between them but their sequence similarity with the digestive enzymes is very low.

The existence of methods for the isolation of digestive carboxypeptidases from both natural sources and recombinant systems (Reverter et al., 1998a; Ventura et al., 1999) has allowed the determination of several 3D structures of carboxypeptidases. The first available structures were those of procarboxypeptidase A and B from porcine pancreas and A from bovine pancreas (Rees et al., 1983; Coll et al., 1991; Guasch et al., 1992). The three enzymes share similar conformational folds and conserve a similar architecture at the active site, with identical catalytic residues and the expected variations in residues at the main specificity-subsites for substrate side chain binding. The core of the tertiary fold is an eight-strand β-sheet over which eight α-helices pack on both sides to form a globular molecule (Fig. 1). The active site is located between the β-sheet and two helices, and residues located in turns or loops protruding from the secondary structure elements hold the Zn^{2+} atom. In the pro-enzyme structures, the globular region of the pro-segment covers the active site and establishes specific interactions with residues important for substrate recognition. The conformations of the pro-segment moieties of porcine

procarboxypeptidase A and B are very similar despite the low degree of sequential identity (32%).

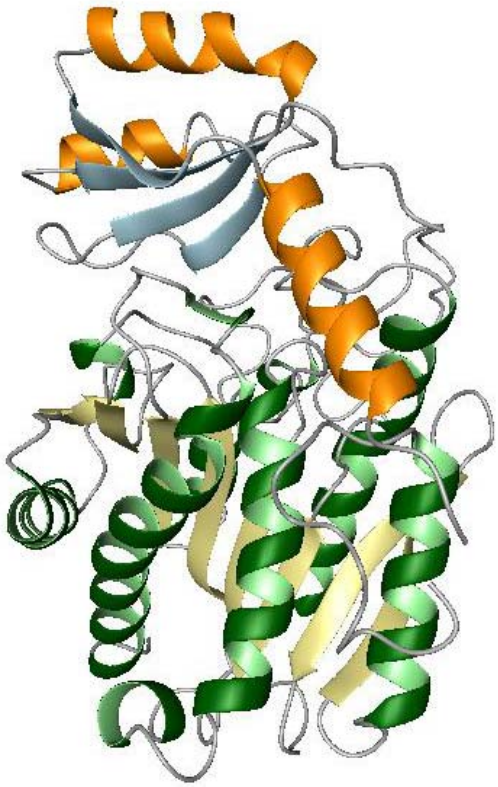


Fig. 1. Three-dimensional structure of bovine pancreatic procarboxypeptidase A. The helixs and β-strands of the pro-segment are colored orange and light blue, respectively.

Bovine pancreatic carboxypeptidase A was among the first proteases to be discovered and characterized, and constitutes the best-known example of metallocarboxypeptidases (Vallee & Neurath, 1955). Its refined structure is used as reference for the description of the catalytic mechanism of carboxypeptidases and as an example of catalytic mechanism in many textbooks. However, its catalytic mechanism has been in controversy for almost 50 years and is yet not clear. From crystallographic studies it is known that the zinc binding residues are His69, Glu72 and His196 (numbering of carboxypeptidase A). A great number of amino acids have been defined as important for substrate binding and catalysis and can be classified into several subsites: Asn144, Arg145, Tyr248 in S1'; Arg127 and Glu270 in S1; Arg71, Ser197, Tyr198 and Ser199 in S2; Phe279 in S3; and Glu122, Arg124 and Lys128 in S4 (Aviles et al., 1993). The terminal carboxylate group of the scissile peptide bond becomes positioned near Glu270, Arg127 and zinc. All of those residues are conserved among digestive carboxypeptidases and differences are restricted to amino acids defining specificity. In carboxypeptidase B, an

Asp residue is at position 255 instead of Ile in A and A2. This residue is located at the specificity pocket and allows the binding of basic residues in the B forms (Barbosa-Pereira et al., 2002). A further replacement of Ile243 by Gly in the B forms contributes to the creation of a polar environment. In A2, Met and Ala replaces Leu203 and Thr268 from A, respectively, rendering a binding pocket more capable of accommodating a bulky substrate side chain.

Pancreatic metalloprocarboxypeptidases are activated by limited proteolysis in the duodenum, where they arrive after being secreted from the granules of the acinar pancreatic cells. The release of the pro-segments, and thus activation, is mainly promoted by trypsin in the case of the procarboxypeptidase B and by trypsin complemented with other pancreatic serine proteases, such as elastase and chymotrypsin, in the case of procarboxypeptidase A. In general, the activation course of these pro-enzymes is dependent on the capability of the scissed fragments of the pro-segment to bind and inhibit the active enzyme. This seems to be strong in the A forms and weak or null in the A2 and B forms. Unfolding-refolding studies on the pro-segment of pancreatic metalloprocarboxypeptidases indicate that the globular part of this region acts as an independent folding unit *in vitro* (Villegas et al., 1998). Its presence greatly facilitates the heterologous expression of the pro-enzymes therefore the pro-segment probably behaves as a co-chaperone in the folding of pancreatic carboxypeptidases *in vivo*. Pancreatic metalloprocarboxypeptidases A and A2 occur in different species either as monomers and/or as binary or ternary complexes with zymogens of serine proteases, such as chymotrypsinogen C, chymotrypsinogen B or proproteinase E. Although the biological role of these digestive complexes is unknown, it has been proposed that they could protect the subunits against inactivation in the acidic conditions of the upper duodenal tract, modulate or coordinate their transformation from zymogens to active forms, or enhance the complementary action of the different components once activated (Gomis-Rüth et al., 1995).

Regulatory metallocarboxypeptidases

Following the criteria based in the physiological function, mast-cell carboxypeptidase A and carboxypeptidase U, two enzymes that belong to the carboxypeptidase A subfamily based upon structural homology, are commented in this section. Knowledge about regulatory carboxypeptidases has increased significantly over

the last two decades as a consequence of the research upon biologically active peptides and the recognition of the involvement of carboxypeptidases in their processing. The first regulatory carboxypeptidase to be described was carboxypeptidase N, a plasma enzyme found to inactivate bradykinin (Matthews et al., 2004). Carboxypeptidases M, H, U and mast-cell carboxypeptidase A were described in the 1980s. Carboxypeptidases D and Z are the newest members of this group of enzymes (Vendrell & Aviles, 1999).

Most regulatory carboxypeptidases appear as single polypeptide chain with one exception: carboxypeptidase N is a dimmer of heterodimers, with each heterodimer containing one catalytic subunit and one non-catalytic subunit (Levin et al., 1982). It is also the only case where the catalytic chain is not glycosylated, rendering it highly unstable in the isolated state and thus needing the presence of the glycosylated accompanying subunit to stabilize it and circulate in the blood. Unlike A/B carboxypeptidases, the H enzymes do not appear to be produced as inactive precursors that require proteolysis to produce the active form (Gomis-Rüth et al., 1999). Instead, these enzymes rely on their substrate specificity and sub-cellular compartmentalization to prevent inappropriate cleavages that would otherwise damage the cell (Reznik & Fricker, 2001). In contrast, mast-cell carboxypeptidase A and carboxypeptidase U, the two regulatory carboxypeptidases from the A subfamily, are synthesized in the form of zymogens.

The enzymes within the carboxypeptidase H subfamily contain a 380-400 residues domain, which is preceded and followed by N- and C-terminal sequences (Fig. 2). The mature enzymes have a higher molecular mass than the digestive forms, either because are glycosylated, have extensions at the C-, and/or N-terminus, or both. All metallocarboxypeptidases are synthesized with a signal peptide but only the pancreatic-like enzymes possess a pro-peptide capable of maintaining the enzymes in the zymogen inactive state until its removal. Regardless of the classification of the M14 family into two groups, they share the same zinc-binding motif. They also show residue conservation at most of the catalytically important residues. Thus, His69, Glu72 and His196 (carboxypeptidase A numbering system) are conserved, as are also Asn144, Arg145 and Tyr248, involved in substrate binding, and Glu270, essential for catalysis. A difference is observed in the catalytic residue Arg127. There are other differences in residues involved in substrate binding: Arg71, Ser197 and Tyr198, and in residues related to specificity: Leu203, Ile243, Ile255 or Thr268. This is expected in the latter case given the variability already observed among pancreatic-like enzymes.

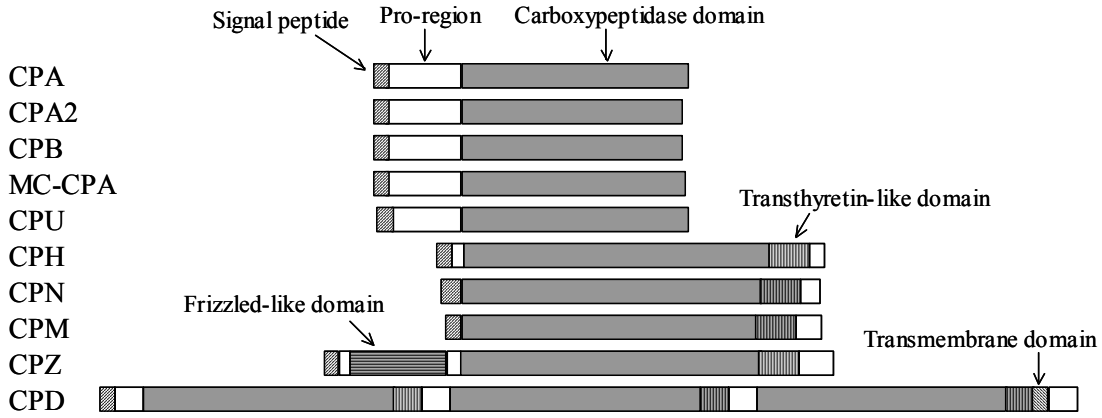


Fig. 2. Comparative sizes of digestive and regulatory carboxypeptidases.

Functional roles of metallocarboxypeptidases

The primary function of digestive carboxypeptidases A and B is to break down peptides in the gut, following the action of chymotrypsin and trypsin on ingested proteins. Similar to the digestive carboxypeptidases, mast-cell carboxypeptidase also cleaves peptides with C-terminal aliphatic/aromatic residues with the highest efficiency and functions in the destruction of proteins and peptides carried out by mast cells presumably following the action of chymase (Reynolds et al., 1989). This carboxypeptidase is found in the secretory granules of these cells forming macromolecular complexes with proteoglycans. Its precise role is still to be defined, although it may obviously be related to pathological conditions in which mast-cells have been implicated, such as allergic response, inflammation and others. The rest of regulatory carboxypeptidases cleave C-terminal basic residues, with different affinities for Arg or Lys depending on the enzyme. They exert their function on peptides related to a considerable number of relevant physiological processes, both normal and pathological, as shown in the following paragraphs.

Carboxypeptidase H, which is also known as carboxypeptidase E or enkephalin convertase, was discovered during searches for a carboxypeptidase that would remove C-terminal Lys and Arg residues from neuroendocrine peptide precursors (Fricker & Snyder, 1983). Most neuroendocrine peptides are initially produced as small proteins that require cleavage at basic amino acid-containing sites by a pro-hormone convertase. Following the action of this convertase, the carboxypeptidase step is then usually required to generate the bioactive peptide. It is involved then in the biosynthesis of numerous peptide hormones

and neurotransmitters such as Met- and Leu-enkephalin, insulin, vasopressin, oxytocin and many others. Carboxypeptidase H is a membrane-bound enzyme and is broadly distributed in the neuroendocrine system (Fricker, 1988). Within cells, this enzyme is further localized in the regulated secretory pathway where it exists in a soluble and membrane-associated form. Carboxypeptidases M and N have also been proposed to function in the processing of peptide hormones, but in contrast to carboxypeptidase H, these other carboxypeptidases function extracellularly. Carboxypeptidase M was first purified and characterized from human placental microvilli (Tan et al., 1989). It is a widely distributed membrane-bound ectoenzyme which is found at high levels in placenta and lungs and also in a variety of tissues and cells including blood vessels, brain and peripheral nerves (Reverter et al., 2004). It was considered to be a marker of differentiation from peripheral blood monocytes to macrophages *in vitro*. This enzyme releases Arg faster than Lys. Its soluble form has been detected in amniotic fluid, seminal plasma and urine. Carboxypeptidase M could participate in the control of peptide hormone activity at the cell surface, in the modulation of receptor specificity through the generation of C-terminal shortened new agonists or antagonists and in extracellular protein processing or degradation. In addition, it has been implicated in the metabolism of growth factors.

Carboxypeptidase N is produced by the liver and circulates in plasma as a protein complex at high concentrations. The glycosylated subunit protects the catalytic from degradation and removal from the blood by glomerular filtration. This carboxypeptidase cleaves an ample variety of peptide and protein substrates and inactivates or alters the specificity of potent vasoactive peptides such as kinins and anaphylatoxins C3a, C4a and C5a (Campbell et al., 2001; Campbell et al., 2002). It is believed that carboxypeptidase N has a general protective function in blood. The precise functional roles of carboxypeptidases D and Z are still to be defined (Reznik & Fricker, 2001). It is believed that the D form may have a broad role in the processing of proteins that transit in the regulatory pathway (Song & Fricker, 1995). Carboxypeptidase Z is thought to perform an extracellular function (Song & Fricker, 1997; Novikova et al., 2000). This enzyme is unique among members of the metallocarboxypeptidase family by the presence of a 120-residue cysteine-rich domain that has 20-35% amino acid sequence identity to *Drosophila* and mammalian Frizzled proteins (Sidyelyeva & Fricker, 2002). These proteins are receptors for members of the Wingless/Wnt family, which are important signaling molecules in early development. Thus carboxypeptidase Z more likely functions as an extracellular Wnt-binding protein.

Carboxypeptidase U is a plasma enzyme that is produced in the liver (Eaton et al., 1991; Bajzar et al., 1995). This enzyme has also been named carboxypeptidase R, plasma carboxypeptidase B, and thrombin-activatable fibrinolysis inhibitor (TAFI). It circulates in plasma primarily in its inactive precursor form and becomes unstable upon removal of the glycosylated pro-segment. Since it has affinity for basic C-terminal residues, carboxypeptidase U may release Lys and Arg residues at the binding sites for plasminogen on the surface of cells and fibrin and thus modulate plasminogen activation and the rate of fibrinolysis.

Inhibitors of metallocarboxypeptidases

Metallocarboxypeptidases are inhibited by several chelating agents such as 1,10-phenanthroline. Benzyl succinic acid inhibits A-type enzymes, and compounds like guanidinoethylmercaptosuccinic acid and 2-mercaptomethyl-3-guanidinoethylthiopropa noic acid inhibit B-type enzymes, with inhibition constants (K_i) in the order of micromolar to nanomolar (Vendrell & Aviles, 1999). Other synthetic inhibitors of carboxypeptidase A have been used in order to investigate its catalytic mechanism. In contrast to the numerous natural inhibitors of endoproteinases, only few protein inhibitors of metallocarboxypeptidases have been found (Vendrell et al., 2000). Four exogenous protein inhibitors from *Solanacea* plants and the parasites *Ascaris suum* and *Hirudo medicinalis* and one mammalian endogenous carboxypeptidase inhibitor have been reported.

Potato carboxypeptidase inhibitor

The potato carboxypeptidase inhibitor (PCI or CPI) is the most extensively studied among the natural inhibitors (Hass et al., 1975). It was discovered in 1966 as an impurity in crystals of potato chymotrypsin inhibitor I by its ability to inhibit carboxypeptidase B (Ryan, 1966). Five different variants of PCI were found because of isomorphic and processing differences (Hass & Derr, 1979a). They present very little sequential differences and identical inhibitory activities toward carboxypeptidases A and B, with sizes ranging from 37 to 39 residues (Mw 4.1-4.3 kDa). All of these PCI forms competitively inhibit metallocarboxypeptidases with a K_i in the nanomolar range (Hass & Ryan, 1981a). Potato tuber contains the highest levels of PCI, although this inhibitor can also be detected in leaves, sprouts and upper stems (Hass & Derr, 1979b). The cloning of its cDNA showed that PCI is synthesized in potato with a 29-residue N-terminal signal peptide, a 27-residue

N-terminal pro-region, a 39-residue mature region, and a 7-residue C-terminal extension (Villanueva et al., 1998). Northern blot analysis demonstrated that the gene of PCI is transcriptionally activated by wounding, which can be induced by abscisic acid and jasmonic acid. By immunocytochemistry and electron microscopy it was also shown that this inhibitor accumulates inside the vacuoles in the cells. The role of the C-terminal extension of PCI is possibly, apart of the vacuolar sorting function, the modulation of the PCI activity. The function of the N-terminal pro-region in PCI was thought to be important during its folding process. However, a recent study indicated that PCI does not depend on this pro-sequence for folding neither *in vitro* nor *in vivo* in *E. coli* (Bronsoms et al., 2003).

The three-dimensional structure of PCI was determined in solution by NMR spectroscopy (Clore et al., 1987; Gonzalez et al., 2003) and by crystallography in complex with bovine pancreatic carboxypeptidase A (Rees & Lipscomb, 1982). The 39 amino acids of PCI are organized in a 27-residue globular core, a 7-residue N-terminal tail and a 5-residue C-terminal tail (Fig. 3). The core, stabilized by three disulfide bridges (Cys8-Cys24, Cys12-Cys27, and Cys18-Cys34) forms a scaffold called T-knot. Other proteins also share this scaffold (Lin & Nussinov, 1995); proteins with different functions such as squash serine protease inhibitors, toxins from snake venoms, or some growth factors. Few regular secondary structure elements are found in PCI: only a short 3/10 helix (residues 14 to 19) and a very small β-sheet made by two short antiparallel strands (residues 26-28 and 33-35). The uncomplexed PCI in solution and the complexed form in crystal present very similar structures, although the residues of the tails are disordered in solution.

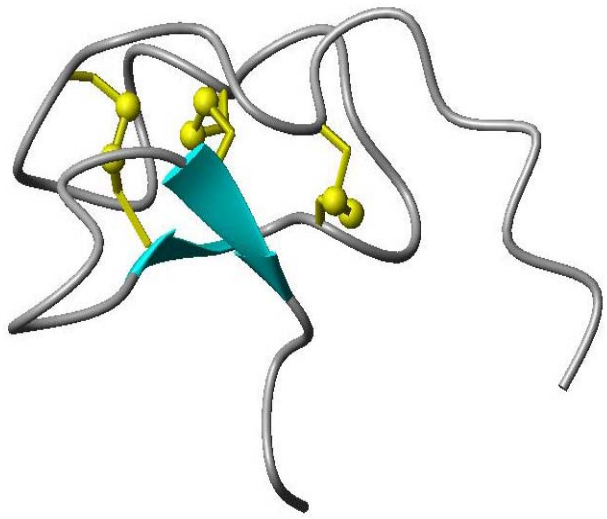


Fig. 3. Three-dimensional structure of PCI. The disulfide bridges are indicated in yellow.

In order to ensure homogeneity in the preparations of PCI, a synthetic gene using the *E. coli* codon usage was constructed (Molina et al., 1992). Several methods for its over-expression using fed-batch aerobic culture in a fermentor were developed (Marino-Buslje et al., 1994). Large quantities (about 200 mg/L) of recombinant PCI could be easily purified from the supernatant of these cultures. This allowed additional studies on this inhibitor (e.g. folding and unfolding), which are commented in next sections.

Tomato carboxypeptidase inhibitor

Tomatoes contain a metallollocarboxypeptidase inhibitor (MCPI) with an amino acid sequence highly similar to that of PCI (Hass & Hermodson, 1981b). Of the 37 residues of these inhibitors, 26 are identical, and 7 of the 11 replacements could have arisen by single base changes. Moreover, all six residues of half-cystine, which occur as three disulfide bridges, are aligned without the need to introduce gaps in either sequence. The K_i-values of both inhibitors are virtually indistinguishable for carboxypeptidases A and B (in the nanomolar range).

The carboxypeptidase inhibitor from Ascaris suum

Ascaris suum are intestinal parasites that produce inhibitors toward pepsins of the host stomach, and inhibitors of chymotrypsin, elastase and trypsin of the host intestine. The metallocarboxypeptidase inhibitor found in this parasite is a 65- or 66-amino acid residues polypeptide (Mw 7.5-7.6 kDa) (Homandberg et al., 1989). Its sequence bears no obvious homology with either the chymotrypsin/elastase or trypsin inhibitors from *Ascaris*. However, some homology was observed with a trypsin inhibitor from pumpkin, and with factors VIII and X. In addition, it displays some similarity with the inhibitors from *Solanacea* at a short 9-residue internal stretch and at the C-terminal 5 residues.

Leech carboxypeptidase inhibitor

Several protein inhibitors of serine protease have been isolated from the medical leech *Hirudo medicinalis*. The leech carboxypeptidase inhibitor (LCI) is the first metallocarboxypeptidase inhibitor found in leeches and does not show similarity in sequence with any other protein, except at its C-terminus tail, which is very similar to those of the *Solanacea* carboxypeptidase inhibitors (Reverter et al., 1998b). LCI is a cysteine-rich polypeptide of 66 residues (Mw ~7.4 kDa) that behaves as a tight binding and

competitive inhibitor of metallocarboxypeptidases with a K_i in the nanomolar range. The LCI cDNA codes for an 81-amino acid sequence containing a 15-residue hydrophobic signal peptide, which precedes the N-terminal Ser of the mature protein. Heterologous expression in *E. coli* allowed further structural studies on this molecule, which was produced with an extra glycine as N-terminal residue (67 residues). CD and 1D NMR spectroscopy indicated that recombinant LCI is a compactly folded globular protein, very stable to a wide range of pH and denaturing conditions. Its three-dimensional structure was determined in solution by NMR and in complex with human carboxypeptidase A2 by X-ray crystallography (Reverter et al., 2000). LCI folds in a compact domain consisting of a five-stranded antiparallel β-sheet and a short α-helix (Fig. 4). The β-sheet involves residues Glu 7–Tyr 12 (β1), Gln 16–Arg 23 (β2), Glu 33–His 37 (β3), Val 51–Tyr 53 (β4) and Gly 56–Ile 63 (β5) in a β3-β1-β2-β5-β4 topology, and the short α-helix is located between residues Pro 41 and Gly 46. The α-helix packs onto the most compact part of the β-structure, interacting with the end and the beginning of the β1 and β2 strands, respectively. Nearly 45% of the LCI residues belong to regular secondary structure elements, a high percentage when compared to other proteins of similar size.

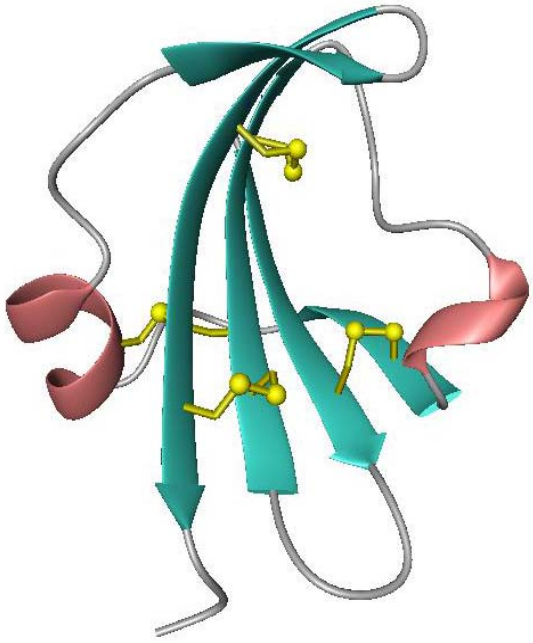


Fig. 4. Three-dimensional structure of LCI. The disulfide bridges are indicated in yellow.

The structures of free LCI in solution and that of LCI in complex with the carboxypeptidase are very similar. However, the C-terminus tail of LCI becomes rigid in the crystal upon binding with the protease. This inhibitor is stabilized by the presence of

four disulfide bridges formed between cysteines 11-34, 18-62, 19-43, and 22-58, all of them located within secondary structure elements.

The endogenous carboxypeptidase inhibitor

This endogenous inhibitor is known as tissue carboxypeptidase inhibitor or latexin and is the largest protein metallocarboxypeptidase inhibitor reported to date. It was purified from rat brain after detection of inhibition of added bovine pancreatic carboxypeptidase A to this tissue. The cloned full-lenght cDNA encodes a protein of 223 amino acid residues (Mw ~26 kDa) (Normant et al., 1995). Its sequence does not display significant homology with the rest of described carboxypeptidase inhibitors. However, it shows a limited degree of similarity with the activation segment of porcine carboxypeptidase B. Its recombinant form produced in *E. coli* appears as a hardly reversible, noncompetitive, and potent inhibitor of carboxypeptidases A, A2, B and mast-cell carboxypeptidase. Initially this inhibitor was found widely in the rat (e.g. in brain, lung, or digestive tract) and its apparently cytosolic localization pointed to a rather general functional role in the control of cytosolic protein degradation. Further studies in rat have shown the expression of latexin in a cell type-specific manner in both central and peripheral nervous systems. Latexin is also constitutively expressed at high basal levels in mouse macrophages and can be upregulated by stimulation of the cells with a growth factor or pro-inflammatory stimulus. Upregulation occurs in parallel with several other types of protease inhibitors and target proteases. Very recently, its structure has been determined by X-ray crystallography (Aagaard et al., 2005). Although there is no evidence of internal symmetry from the polypeptide sequence, its crystal structure reveals two structurally related domains linked by a connecting helix (Fig. 5). Each domain comprises an extended α-helix followed by a strongly twisted four-stranded antiparallel β-sheet of simple up-and-down connectivity that embraces the helix establishing hydrophobic contacts. Latexin shows no structural relationship with the other carboxypeptidase inhibitors. However, the individual domains share the topology of the cystatin/monellin family of proteins, suggesting an evolutionary and possibly functional relationship.

A homologue of latexin has been identified in humans (Liu et al., 2000). Its cDNA contains an open reading frame encoding 222 amino acids. The comparison between the deduced amino acid sequences of rat and human latexins reveals high sequence identity (85%). The expression levels of this human protein are high in heart, prostate, ovary,

kidney, pancreas, and colon, and moderate or low in other tissues like brain. The recombinant form of this inhibitor, expressed in *E. coli*, inhibits A/B carboxypeptidases with affinities in the nanomolar range and displays a very similar structure to that of latexin from rat (Pallarès et al., 2005).

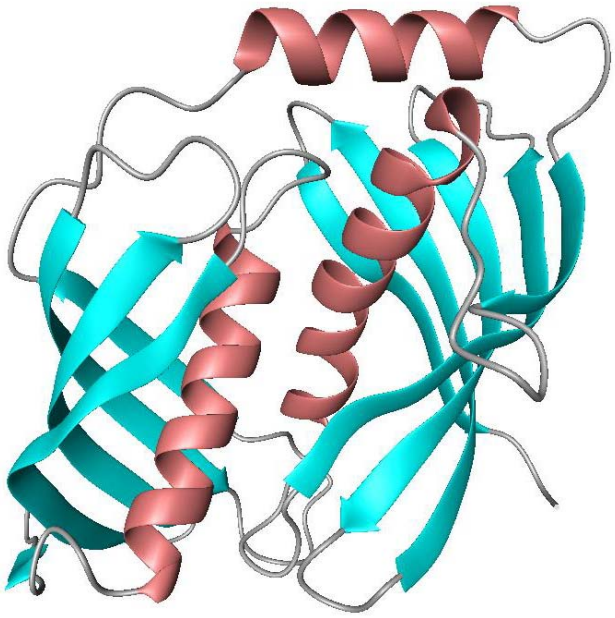


Fig. 5. Three-dimensional structure of latexin.

Mechanism of inhibition of metallocarboxypeptidases by their protein inhibitors

The mechanism of inhibition of PCI and LCI relies on the interaction of their C-terminus tail with the active site groove of the carboxypeptidase in a way that mimics substrate binding (Vendrell et al., 2000). As observed from the crystal structures of PCI and LCI in complex with bovine carboxypeptidase A and human carboxypeptidase A2, respectively (Rees & Lipscomb, 1982; Reverter et al., 2000), the C-tail of the inhibitor is inserted into the active site of the enzyme and forms the primary contact region (Fig. 6); the C-terminal residue (Gly39 in PCI and Glu66 in LCI) is split off but remains bound in the S1' subsite of the carboxypeptidase, buried by the rest of the inhibitory moiety. The new C-terminal residue (Val38/65) is coordinated with the active site Zn^{2+} and hydrogen-bonded to the phenol oxygen of Tyr248 in the enzyme. Several side chain atoms of Val38 are at van der Waals distance of carbonyl oxygen atoms of enzyme residues Ser197 and Tyr198 and of side chain atoms of Ile247, Tyr248, Tyr198 and Phe279. In the LCI complex, Val65 also forms hydrogen bonds with the carboxylate group of Glu270 and with

the guanidinium group of Arg127. Moreover, the carbonyl oxygen of Tyr37/64 is hydrogen-bonded to the guanidinium group of Arg71 of the enzyme. A few additional secondary contacts confer stability to the complex. This secondary region involves residues 15, 22, and 27-30 of PCI, residues from the β1-β2 loop, β3-α1 loop and α1 of LCI, and mainly residues 246-248 of the enzymes.

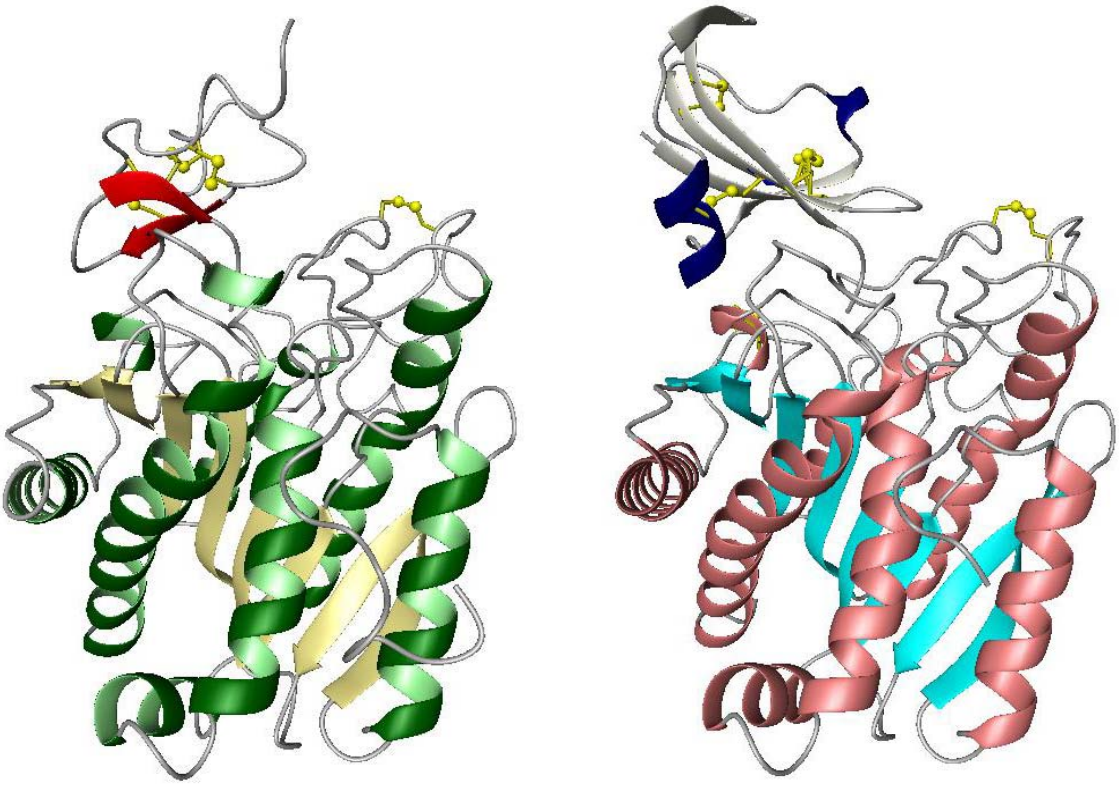


Fig. 6. Three-dimensional structures of PCI-CPA (*left*) and LCI-CPA2 (*right*) complexes. The inhibitors PCI and LCI are colored red and blue, respectively.

Thus, sites S1, S2 and S3 of the carboxypeptidase are involved in contacts with the inhibitor. The numerous interactions with residues essential for substrate binding and catalysis probably account for the highly efficient inhibition. In addition, there is an excellent fit between the interfaces of the globular region of the inhibitors and the carboxypeptidases, which excludes any further penetration and promotes the formation of a tight complex.

The structures of PCI and LCI in these complexes are completely different, except at the C-tail, where the backbone fold is almost identical in both proteins. Given that the sequence of the C-terminus residues is the only region conserved between both inhibitors,

it could be considered a clear example of convergent evolution dictated by the target metallocarboxypeptidases (Reverter et al., 1998b). Site-directed mutagenesis studies were used to analyze the structure-function relationships in PCI. A series of PCI mutants were produced by truncation or replacement of residues in the primary contact region of PCI (the C-tail) toward bovine carboxypeptidase A (Molina et al., 1994; Marino-Buslje et al., 2000). In summary, these studies confirmed the key role of Val38 in the binding and stability of the PCI-carboxypeptidase complex through van der Waal interactions with the active site of the enzyme. In addition, it was shown that Pro36 plays an important role in reducing the mobility of the PCI C-tail.

The C-terminus of the endogenous inhibitor found in mammals does not seem to be a suitable substrate for metallocarboxypeptidases. This protein interacts with carboxypeptidases through the interface of its two subdomains (Fig. 7) (Pallarès et al., 2005).

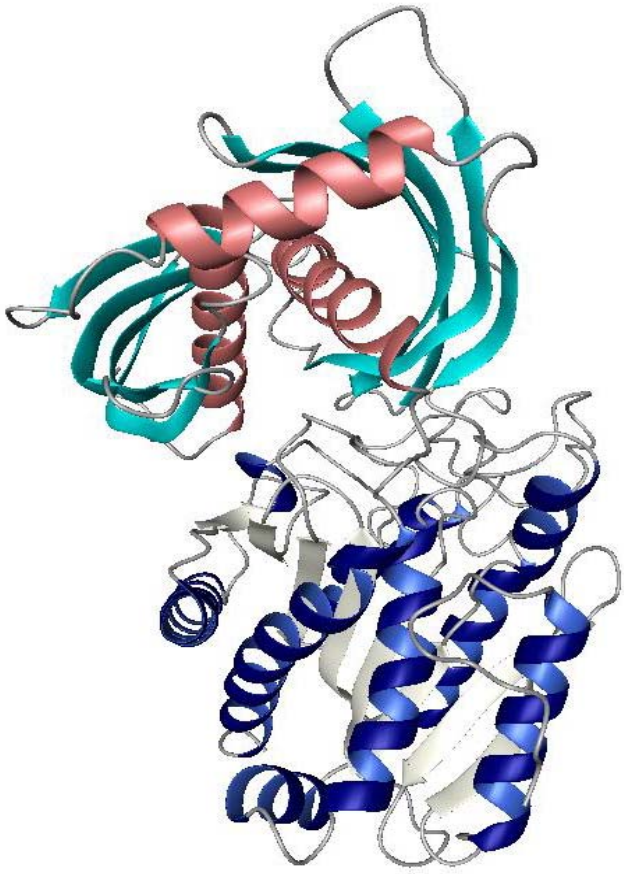


Fig. 7. Three-dimensional structure of the latexin-CPA4 complex. The helixs and β-strands of the inhibitor are colored light red and light blue, respectively.

In the latexin-human CPA4 complex, the inhibitor sits on top of the funnel rim, mainly clamping a loop encompassing residues Asp273-Pro282 of the protease moiety through its lower barrel surface at the subdomain interface. The complex occludes a large contact surface but makes rather few contacts, despite a nanomolar K_i.

Biotechnological and biomedical applications of metallocarboxypeptidase inhibitors

Among several defense proteins, protease inhibitors appear to be an important group in plants (Casaretto & Corcuera, 1995). Protease inhibitors are often found in storage organs and are known to be inducible by injuries such as insect damage. Since digestive carboxypeptidases are significant factors in the overall process of protein digestion, it can be thought that the natural function of potato and tomato inhibitors is probably related to plant defense mechanisms against insects. Metallocarboxypeptidase inhibitors could be then used in plants for the development of defense strategies against pests. The use of protease inhibitors, with a highly specificity and minimal undesirable impact on the environment, could replace conventional toxic pesticides (Ryan, 1989). However, this is an emerging field and much work is yet to be done.

It is known that epidermal growth factor (EGF) and its receptor (EGFR) are involved in many aspects of the development of carcionomas such as tumor cell growth, vascularization, and invasiveness. EGFR is over-expressed in many epithelial tumors, being a potential target for anti-tumor therapy. PCI seems to behave as an EGF antagonist due to its structure in T-knot, very similar to that of this growth factor (Mas et al., 1998). Thus, it competes with EGF for binding to EGFR and inhibits EGFR activation and cell proliferation. Indeed, PCI suppresses the growth of several human pancreatic adenocarcinoma cell lines, both *in vitro* and in nude mice (Blanco-Aparicio et al., 1998). It is interesting to note that the occurrence of similar topologies among unrelated proteins is an attractive theme in structural biology. The T-knot scaffold, a disulfide-reinforced structural motif shared by toxins, protease inhibitors from plants, and EGF-like proteins, represents a compact structural unit suitable for the design of minimized proteins (Lin & Nussinov, 1995).

TAFIa is generated from TAFI by thrombin produced during coagulation, in complex with thrombomodulin, present both on the endothelial cell surface and in plasma (Bajzar et al., 1998). TAFIa attenuates fibrinolysis by catalyzing the removal of C-terminal Lys and Arg residues present in plasmin-degraded fibrin and fibrin degradation products (Wang et

al., 1998). Because C-terminal Lys residues facilitate tissue plasminogen activator (tPA)-mediated plasminogen activation, TAFIa likely exerts its anti-fibrinolytic effect by down regulating tPA-mediated plasminogen activation. Consequently, the inhibition of TAFIa may potentiate the thrombolytic therapy by preventing the TAFIa-mediated attenuation of plasminogen activation (Bouma & Meijers, 2003). Although there is no known inhibitor of TAFIa present in plasma, metallocarboxypeptidase inhibitors are shown to inhibit TAFIa. Several studies demonstrate that commercial PCI and the small organic inhibitor GEMSA act as pro-fibrinolytic agents in the lysis of clots *in vitro* through inhibition of TAFIa (Klement et al., 1999; Mutch et al., 2003; Schneider & Nesheim, 2003; Walker et al., 2003). PCI also reverses prolongation of clot lysis *in vivo* (Nagashima et al., 2000). This effect was achieved on tPA-induced clot lysis using commercial PCI in a rabbit jugular vein thrombolysis model.

The fibrinolytic activity of recombinant LCI has also been studied *in vitro* (Salamanca, 2003a). This molecule also accelerates the tPA-induced lysis of human plasma clots being approximately 10-times more potent than PCI. All of these studies are of great interest because of the high levels of TAFIa recently found in patients with an increased risk for venous thrombosis and with recurrent thromboembolisms (Silveira et al., 2000; van Tilburg et al., 2000; Eichinger et al., 2004). The physiological function of TAFIa is not restricted to fibrinolysis; this enzyme also seems to have a broad anti-inflammatory role (Myles et al., 2003; Bajzar et al., 2004). Mast-cell carboxypeptidase A is also related with inflammation as commented in previous sections (Springman et al., 1995). The possible use of metallocarboxypeptidase inhibitors in inflammatory diseases such as arthritis is yet to be tested (So et al., 2003).

PROTEIN FOLDING

Protein folding problem: historical perspective

The challenge of this century in protein chemistry will be to understand and characterize how proteins perform all the biological functions needed for life. Understanding these processes at the molecular level requires knowledge about folding and three-dimensional structure of proteins, fields of fundamental importance in biological science (Honig, 1999). Because of the key role played by protein folding and unfolding within the cell, it is inevitable that mistakes in folding give rise to the malfunctioning of biological processes and hence to disease (Dobson, 2004). A large number of diseases are already associated with misfolding and more are being added each year. Understanding of protein folding and misfolding is therefore crucial for the rational development of therapeutic strategies directed against these diseases. In addition, progress in protein folding is of great interest since unfolded and partially folded states are known to be important in events such as translocation across membranes, the trafficking of proteins to particular locations within the cell, and in targeting for destruction those proteins that have served their function and are no longer needed.

From Anfinsen to pathways

The main discoveries in the field of protein folding started 40 years ago. The famous experiments of Christian B. Anfinsen and co-workers beginning in the early 1960s suggested that the three-dimensional structure of proteins is codified by the amino acid sequence, information that was called the second genetic code (Anfinsen, 1973). However, later emerged that in many cases additional factors, known as chaperones, might be required to assist protein folding (Hendrick & Hartl, 1995). In these initial studies, Anfinsen showed that proteins can fold reversibly, implying that the native structures of some small globular proteins such as ribonuclease A (RNase A) are thermodynamically stable, and therefore are conformations at the global minima of their accessible free energies (Anfinsen et al., 1961). In the late 1960s Cyrus Levinthal made the argument, later called "Levinthal paradox" (Levinthal, 1968; Zwanzig et al., 1992), that there are too many possible conformations and insufficient time for proteins to find the "needle" (the native structure) in the "haystack" (conformational space) by random searching. Levinthal concluded that proteins must fold by specific "folding pathways".

Levinthal framed the puzzle as if the two goals, achieving the global minimum and doing it quickly, were mutually exclusive. These two mutually exclusive options came to be called thermodynamic control and kinetic control. The former control meant that a protein reaches its native structure (global minimum in energy) through a pathway-independent folding. It takes a long time because it requires an extensive search. Kinetic control meant that folding happens quickly because it is pathway-dependent and therefore the protein may reach only local optima. Levinthal's argument led to a search for folding pathways over many years.

The modern era of folding experiments began soon after Levinthal's thought with the pioneering works of Ikai and Tanford (Ikai & Tanford, 1971), and Tsong, Baldwin and Elson (Tsong et al., 1971) in 1971 that initiated the search for folding intermediates. Immediately another question arose: Are folding intermediates on-pathway or off-pathway? The implication was that only on-pathway intermediates could teach us the folding code whereas off-pathway intermediates were uninteresting dead ends. These studies, known as *classical view* of folding, are based on simple phenomenological models and use decays of optical properties that monitor changes in the protein structure after a jump to folding or unfolding conditions. When a single exponential decay is observed in both folding and unfolding directions, kinetics is called "two-state" because only comprises a native state (N) and a denatured state (D). When multiple exponentials are observed, more complex models are required with additional intermediate conformations (Fig. 8). Models are chosen based on which one gives the best fit to the experimental data. The classical experiments generally probe only the average behavior of the protein and are not able to resolve much at atomic detail.

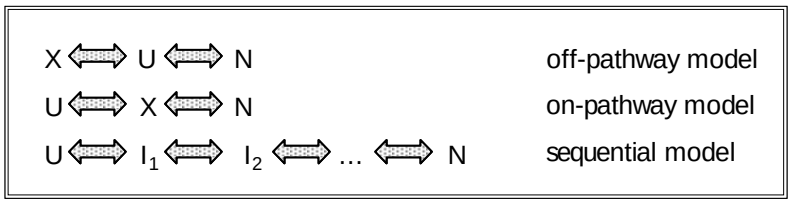


Fig. 8. Three of the most important classical models in protein folding. U represents the fully unfolded denatured state, N the native state, and X or I represent intermediate states that have properties between those of U and N.

In the 1970s and 1980s the search for mechanisms in protein folding led to a multitude of proposals (Daggett & Fersht, 2003). The first, called *nucleation-growth*, proposed that tertiary structure propagates rapidly from an initial nucleus of local secondary structure (Wetlaufer, 1973). However, nucleation dropped from favor as it predicts the absence of folding intermediates, and the field of protein folding in these years was dominated by the study of folding intermediates. Two alternative models prevailed (Daggett & Fersht, 2003); the first was the *framework model*, described by Kim and Baldwin, and the related diffusion-diffusion model of Karplus, in which secondary structure is proposed to fold first, followed by docking of the pre-formed secondary structural units to yield the native, folded protein (Kim & Baldwin, 1982; Kim & Baldwin, 1990; Karplus & Weaver, 1976). The second was the *hydrophobic collapse model*, in which hydrophobic collapse drives compaction of the protein so that folding can take place in a confined volume, therefore narrowing the conformational search to the native state (Tanford, 1962; Baldwin, 1989). The framework model gained support from studies mostly carried out in the 1980s on small, relatively stable, helical peptides. Previously, it was assumed that secondary structural segments were not stable enough to form in the absence of tertiary contacts. Support for the hydrophobic collapse model came from early studies showing that the hydrophobic driving force provided by the expulsion of water from the burial of non-polar surfaces is substantial. However, the idea that the conformational search is facilitated within a nonspecific hydrophobic globule presents a problem because an excess of interactions will hinder reorganization of both the polypeptide chain and side chains. The discussions that followed incorporated the molten globule intermediate, proposing that the secondary structure is formed in the process of collapsing. Another problem with the simple views of protein folding provided by the hydrophobic collapse and framework models that has become evident in the past decade is that the denatured state is rarely a random, unstructured coil in which side chain interactions are fleeting and all amino acids behave independently. Instead, proteins generally adopt residual structure, which can be in the form of fluctuating secondary structure and dynamic side chain interactions.

These three above-mentioned mechanisms were replaced in the early 1990s by the *nucleation-condensation model*, which unites features of both the hydrophobic collapse and framework mechanisms (Fersht, 1995; Itzhaki et al., 1995; Fersht, 1997). Nucleation-condensation invokes the formation of long range and other native hydrophobic

interactions in the transition state to stabilize the otherwise weak secondary structure. In this way, secondary and tertiary structure is formed in parallel as proteins undergo a general collapse.

From pathways to funnels

The *new view* of protein folding (Baldwin, 1994; Baldwin, 1995; Baldwin, 1999), which derives from advances in both, experiments and theory, gives faster and more detailed structural information, down to the atomic level. The new experimental techniques include high-resolution hydrogen exchange, mass spectrometry, mutational studies, and fast-triggered methods that explore very early events in folding. The theoretical advances comprise statistical mechanics models with highly simplified representations of chain geometries and interactions, analyzed by analytical methods and computer simulations. These models recognize that macroscopic states are really distributions or *ensembles* of individual chain conformations; many of the same molecular conformations may contribute to different macroscopic states, i.e. can overlap, therefore the new view replaces the pathway concept of sequential events with the *funnel* concept of parallel events (Dill & Chan, 1997). This new models also use the concept of *energy landscapes*. An energy landscape is the free energy of each conformation as a function of the degrees of freedom, such as the dihedral bond angles along the peptide backbone. The vertical axis of a funnel represents the internal free energy of a given chain configuration, while the many lateral axes represent the conformational coordinates (Fig. 9). The high dimensionality of this representation reflects the many degrees of freedom of a protein chain. Each conformation is represented by a point on the multidimensional energy surface and those that are similar geometrically are close to one another on the energy landscape. Energy landscapes can have many kinds of features; hills correspond to high-energy conformations and valleys are configurations that are more favorable than others nearby.

The kinetic process of folding or unfolding a protein can be likened to rolling a ball on this energy surface. Upon initiation of folding conditions, the protein tends to change its conformation in ways that decrease its energy, but it is also constantly buffeted into different conformation by Brownian motion. In this analogy, each individual protein molecule corresponds to a ball rolling on the energy landscape following some particular trajectory, winding through the hills and valleys while being randomly redirected by Brownian forces; uphill steps also happen but less often.

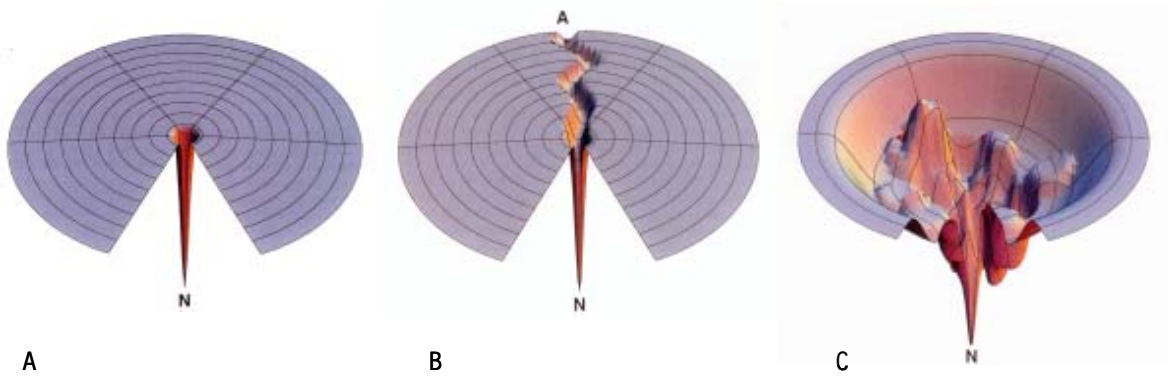


Fig. 9. Different energy landscapes to illustrate protein folding. A. The Levinthal "golf-course" landscape. The chain searches for the native conformation (N) randomly. B. The "pathway" solution to the random search problem of A. A pathway is assumed to lead from a denatured conformation (A) to N, so conformational searching is more directed and folding is faster than for random searching. C. A rugged energy landscape with kinetic traps, energy barriers, and some narrow throughway paths to N. Folding can be multi-state.

The new view of folding recognizes that the fundamental problem with Levinthal's solution is the concept of "pathway" itself (Dill & Chan, 1997). The Levinthal's paradox is little more than a misconception about how any physical, chemical or biological system that is governed by thermodynamics can reach its stable states in measurable times. The two goals of reaching a global energy minimum and doing so quickly are not mutually exclusive. The paradox is an artifact of framing the folding problem in terms of the landscape of Fig. 9A. A pathway can lead from a specific point A to a specific point N, as in Fig. 9B, but folding a protein does not involve starting from one specific conformation (A). The denatured state of a protein is not a single point on the landscape; it is all the points on the landscape, except for N. A pathway is too limited an idea to explain the flow from everywhere else, the denatured ensemble, to one point N. The new view recognizes that the solution to Levinthal's paradox is funnels, not tunnels. While pathways lead us to think in terms of the sequential assembly of specific structures, funnels lead us to think in terms of parallel microscopic multi-pathway processes and ensembles. Folding funnels may not be generally smooth and simple and Fig. 9C indicates a possible scenario with broad shapes, as in mountains, involving hills, valleys, ridges, channels, moguls, plains, and so on. Whereas the pathway of a classical model is a one-dimensional route through configuration space, a funnel describes the progressive reduction in dimensionality of the

accessible conformational space, beginning from the many degrees of freedom available to denatured chains, ultimately down to the nearly complete lack of freedom of the native chain. Better knowledge of these ideas will come when new experiments become available for measuring not just average of structural observables, but also correlations among their fluctuations.

Topology constraints, folding pathways and computer models

The strong correlation found between the folding rates of proteins with two-state folding mechanisms and the complexity of their topology (Plaxco et al., 1998) suggests that the folding of small proteins may be simpler than originally thought (Baker, 2000). Moreover, it was found experimentally that proteins sharing similar topologies have similar transition-state ensembles (Martinez & Serrano, 1999). On the basis of the energy landscape theory and considering only native interactions (Go-like models), several groups have proposed elegant methods to predict both folding rates (Munoz & Eaton, 1999) and the structure of the transition-state ensemble (Alm & Baker, 1999; Munoz & Eaton, 1999; Plaxco et al., 2000). As the use of simple potentials is sufficient to capture the major characteristics of the folding process, it has been suggested that the complex and detailed network of interactions stabilizing the folded protein may not be as important as protein topology for the selection of folding pathways. The 'topological frustrations' encountered by the polypeptide chain during folding are likely to overwhelm the effect of the 'energetic frustrations' due to specific interactions. A good qualitative agreement between predictions and experiments has been found not only for two-state folding proteins, but also for larger proteins exhibiting transient on-pathway intermediates (Plaxco et al., 2000). Thus, computer algorithms that can predict the mechanisms underlying the folding process have promising applications. One of them is the design of folding pathways and optimization of the folding process, which may be used as a strategy to reduce misfolding and formation of amyloid fibrils.

Disulfide bonds and protein folding

The most widely adopted approach to study the pathway of protein folding is to unfold the protein in the presence of strong denaturant or by extreme pH and temperature. Following the removal of denaturant, pH jump, or temperature jump, denatured proteins usually refold spontaneously to form the native structure. The pathway of refolding is then

commonly monitored using the restoration of physicochemical signals of fluorescence, circular dichroism, infrared, UV, and NMR coupled with amide-proton exchange. However, it does not permit, in most cases, isolation of folding intermediates because of their short half-life. To date, much of our knowledge about protein folding has been acquired by the study of disulfide-containing proteins (Wedemeyer et al., 2000). For these proteins, unfolding and refolding are generally followed by reduction and oxidation of the native disulfides. Since breaking and formation of disulfide bonds can be chemically trapped and characterized, the disulfide unfolding and folding pathways can thus be constructed on the basis of the heterogeneity and structures of the trapped intermediates (Creighton, 1990).

Disulfide bonds between cysteine residues are a key component of many proteins (Creighton et al., 1995). In extracellular proteins, the regeneration of their native disulfides is a critical step in their maturation; during protein synthesis, such proteins are threaded into the endoplasmic reticulum where these bonds may be formed. Disulfide bonds are relatively rare in intracellular proteins because the cytosol is a strongly reducing environment. Disulfide formation *in vivo* is catalyzed by specialized enzymes belonging to a thioredoxin superfamily, such as protein disulfide isomerase (PDI) in eukaryotes. PDI is localized in the endoplasmic reticulum where catalyzes the formation and rearrangement of disulfides during the maturation of most secretory and membrane-bound proteins. Recent studies have analyzed the effect of PDI on the folding mechanism of proteins such as RNase A (Shin & Scheraga, 2000). Disulfide bonds can stabilize significantly the native structure of a protein both by lowering the entropy of the unfolded state and by forming stabilizing interactions in the native state. However, the native pairing of disulfide bonds is the end result of an often complicated process involving covalent reactions such as oxidation, reduction, and disulfide reshuffling (when a protein thiol group attacks a disulfide bond of the same protein). Stable tertiary structure also affects the formation of disulfide bonds (Creighton et al., 1996). It can inhibit the contacts between thiols and disulfide bonds by controlling the accessibility, reactivity, and proximity of these reactive groups. For example, the burial of the disulfide bonds in stable tertiary structure inhibits their reduction and reshuffling, while the sequestering of thiol groups can inhibit their oxidation, i.e. the formation of disulfide bonds. Thus, the rate-determining step in protein folding is often the formation of a disulfide intermediate with stable tertiary structure; disulfide species with buried thiols and disulfide bonds are often called metastable (Creighton & Goldenberg, 1984).

A well-established method for analyzing the mechanism of folding in proteins with disulfide bonds is *oxidative folding*. This method, pioneered by Creighton (Creighton, 1986), describes the composite process by which a reduced, unfolded protein recovers both its native disulfide bonds (disulfide bond regeneration) and its native structure (conformational folding); proteins are fully reduced and denatured, and allowed to refold after removal of the reducing and denaturing reagents. Elucidation of the pathway of oxidative folding requires the trapping, isolation and characterization of intermediates accumulated along the process of folding (Narayan et al., 2000). The study of concentration, heterogeneity, and structure of disulfide isomers permits the construction of a detailed folding pathway. Oxidative folding also allows the analysis of the influence of redox agents on the folding pathway and the efficiency of recovery of the native protein.

Mainly two techniques are used for oxidative folding: pulsed-label NMR, which permits trapping and identification of amide groups that are engaged in the structured elements. This technique can in principle be applied to all types of proteins but the trapped intermediates are not amenable to chromatographic purification. An established alternative is to quench the disulfide pairing (*disulfide quenching*) during the course of folding by acid or alkylating reagents and to follow the mechanism of formation of native disulfides (Wu et al., 1998) (Fig. 10). This technique is limited to the analysis of disulfide-containing proteins but a unique advantage is that species of trapped intermediates can be further purified for characterization, normally by using reversed-phase high performance liquid chromatography (RP-HPLC). The acid trapping, performed at pH 3-4, represents a very simple and effective method for quenching the reshuffling of disulfide bonds. The reversibility of this technique allows further stop/go experiments of the trapped and isolated disulfide isomers.

Application of the oxidative folding and acid trapping method has allowed the analysis of folding pathways of several disulfide proteins such as bovine pancreatic trypsin inhibitor (BPTI), RNase A, hirudin, EGF, tick anticoagulant peptide (TAP), PCI and LCI, among others. Recently, a new method has been used in the study of disulfide folding pathways. It allows oxidizing the thiols of proteins with disulfides faster than the rate at which thiol-disulfide exchange reactions can take place. It uses the compound trans-$[Pt(en)_2Cl_2]^{2+}$ and elucidates the "native tendency" of the reduced polypeptide chain to form the native set of disulfide bonds (Narayan et al., 2003). This native tendency is defined as the propensity of the polypeptide chain to promote the direct formation of the full native set of disulfide bonds over non-native ones, and it is represented by the ratio of

the formed native protein to the fully oxidized non-native species during the course of oxidation.

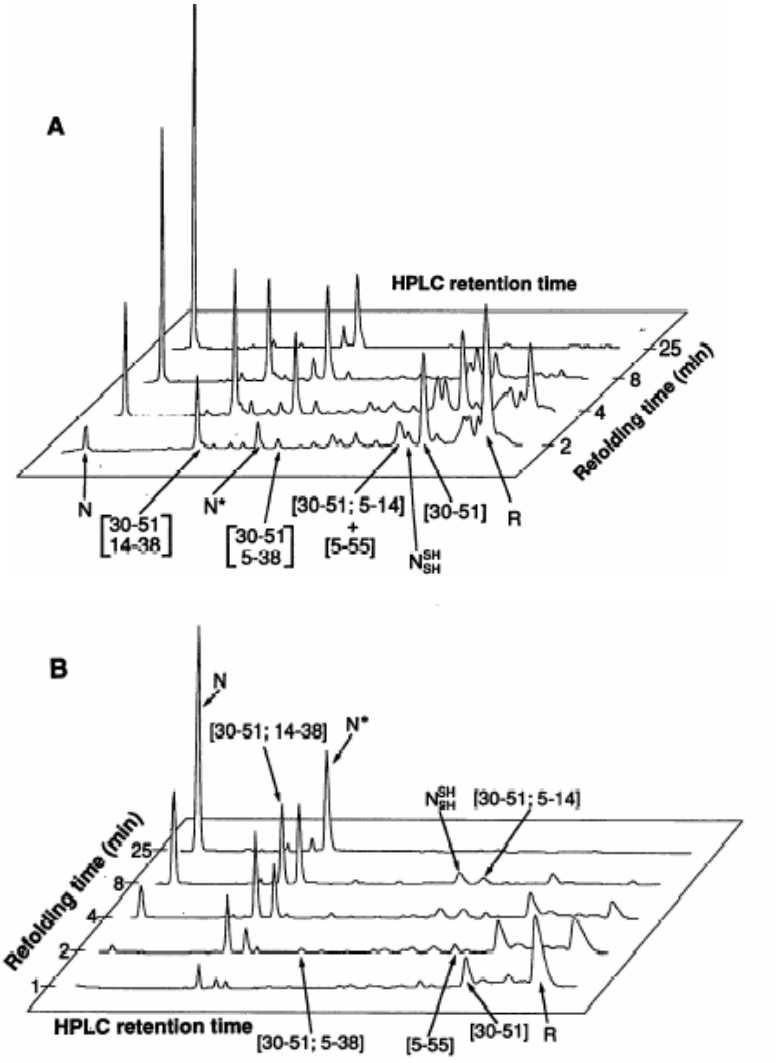


Fig. 10. HPLC chromatograms obtained after various times of BPTI folding. The folding reaction was quenched with either iodoacetate (A) or formic acid (B).

Disulfide bonds and protein unfolding

Knowledge about stability and unfolding behavior of proteins constitutes a basis for the development of variants of these molecules with enhanced activity and/or stability. The conformational stability of proteins has been regularly studied with denaturation curves, where changes in parameters (e.g. fluorescence, UV, circular dichroism) that can be correlated with the three-dimensional structure of the protein are measured as a function of denaturant concentration in the protein solution. These approaches are not able to determine the degree of unfolding of a protein in a given denatured state, or the presence

and concentration of intermediates. However, this information can be obtained from the study of unfolding in disulfide-containing proteins.

Unfolding of these proteins can be achieved conventionally by either reduction of disulfide bonds in the absence of denaturant (*reductive unfolding*) or by denaturation (e.g. urea, GdnHCl) in the absence of reductant (disulfide-intact unfolding). In the latter case, the unfolded protein retains intact native disulfides. A recent methodology (called *disulfide scrambling*) allows determination of stability toward denaturants on the basis that the presence of trace amounts of a thiol initiator during unfolding by denaturants generates a mixture of scrambled species mostly consisting of non-native disulfides that still maintain the native number of disulfide bonds (*scrambled* isomers are fully oxidized species that contain at least two non-native disulfide bonds) (Chang, 1997). The composition of this mixture depends on the type of denaturant and its concentration. The trapping (e.g. acid trapping) of such scrambled species and its analysis by chromatographic methods (e.g. RP-HPLC) permit an evaluation of the extent of protein denaturation by measuring the amount of native protein converted into scrambled form and the degree of unfolding through the relative composition of the different scrambled species. This method has been applied to characterize the unfolding pathway and denaturation curves of several disulfide-containing proteins such as hirudin, TAP, RNase A, PCI and LCI (Bulychev & Chang, 1999; Chang, 1999). These studies demonstrate the potential of this approach for comparing the stability of disulfide-rich proteins and characterizing the extent of unfolding of the denatured proteins.

Folding of small disulfide-containing proteins

Several 3- and 4-disulfide proteins have been studied by the technique of oxidative folding. The results have shown that even among three-disulfide proteins, the folding mechanisms vary substantially. These differences are illustrated by the extent of heterogeneity of folding intermediates, the predominance of intermediates containing native disulfide bonds, and the accumulation of scrambled isomers as intermediates. BPTI and hirudin represent two important cases of such divergent folding mechanisms. The mechanism of *BPTI folding* has been a subject of intensive investigation and is one of the most characterized models (Goldenberg, 1992). Its folding pathway was originally determined by Creighton (Creighton, 1978), who identified seven predominant intermediates; two 1-disulfide species and five 2-disulfide species, with 75% of the

disulfide bonds being native. In the revised model proposed by Weissman and Kim (Weissman & Kim, 1991), five well-populated intermediates, two 1-disulfide and three 2-disulfide species, were described, and all of them were shown to contain only native disulfide bonds. Despite this inconsistency, the model of BPTI folding maintains three principal conclusions: (a) the folding intermediates are only 1- and 2-disulfide species; (b) specific non-covalent interactions that stabilize the native BPTI play a crucial role in guiding the folding early on and hence dictating the formation of limited numbers of well-populated intermediates that admit mainly native disulfides (funneling the conformations toward the native state to prevent the accumulation of nonproductive intermediates); and (c) the rate-limiting step of BPTI folding is the intramolecular rearrangements of 2-disulfide intermediates to generate the intermediate processor that rapidly form the third and final native disulfide.

The *folding pathway of hirudin* (Chatrenet & Chang, 1993; Chang, 1994a) and other single-domain proteins, *PCI and TAP* (Chang et al., 1994b; Chang, 1996; Chang & Ballatore, 2000a), all containing three disulfides as well as a similar size to that of BPTI, can be dissected in two stages: an initial stage of nonspecific disulfide formation (packing) that leads to the formation of scrambled 3-disulfide species, followed by disulfide reshuffling (consolidation) of the heterogeneous scrambled population which leads to the native structure. Their folding mechanism differs from that of BPTI in two crucial aspects. Their folding intermediates, including 1- and 2-disulfide species, are far more heterogeneous than those described for BPTI. The most noticeable difference, however, is the presence of scrambled 3-disulfide isomers as folding intermediates, which has not been observed in BPTI. Surprisingly, TAP and BPTI are two proteins that belong to the Kunitz-type inhibitor family and share close structural homology in terms of disulfide pattern and three-dimensional structure (Antuch et al., 1994). The main differences in their folding pathways cannot be more intriguing.

Another protein that has been studied is *EGF* (Chang et al., 1995; Chang et al., 2001). Its folding mechanism displays both similarity and dissimilarity to that of BPTI and hirudin. Thus, it proceeds through 1-disulfide intermediates and accumulates rapidly as a single stable 2-disulfide intermediate (EGF-II), which represents up to more than 85% of the total protein along the process. Among the five 1-disulfide intermediates that have been characterized, only one is native. Extensive accumulation of EGF-II indicates that it

accounts for the major kinetic trap of EGF folding. This intermediate contains two of the three native disulfide bonds of EGF. However, formation of the third native disulfide is slow and does not occur directly. EGF-II reaches the native structure via rearrangements through 3-disulfide scrambled isomers (Fig. 11).

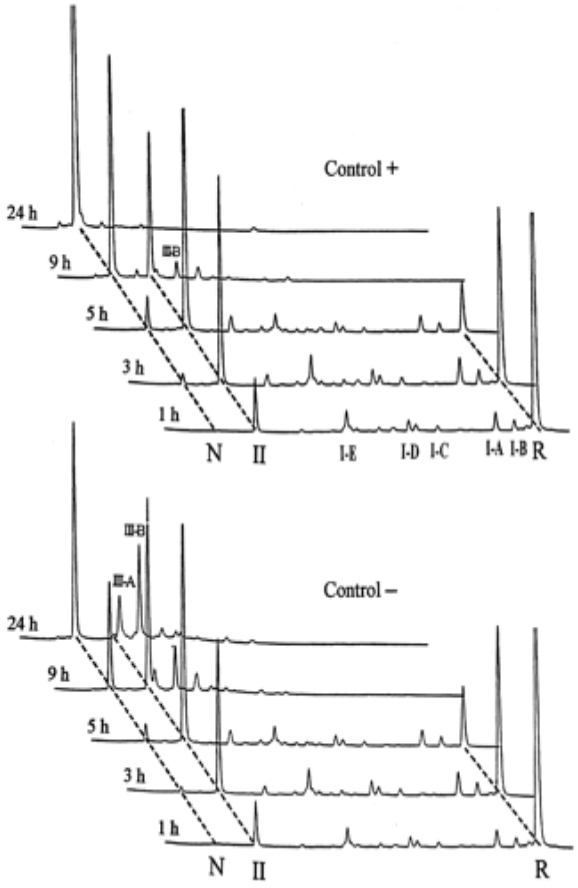


Fig. 11. Intermediates of oxidative folding of EGF trapped by acidification. Folding was carried out with (Control +) or without (Control –) 2-mercaptoetha nol (0.25 mM). Folding intermediates were trapped with trifluoroacetic acid and analyzed by RP-HPLC. *N* and *R* stand for the native and the fully reduced EGF. *II*, *I-A*, *I-B*, *I-C*, *I-D*, and *I-E* indicate the predominant 2-disulfide intermediate and five identified 1-disulfide intermediates. *III-A* and *III-B* are two major species of 3-disulfide scrambled isomers.

These studies indicate that scrambled proteins play an essential role along the pathway of productive folding. This proposal, however, contradicts conventional wisdom, which considers scrambled species as abortive structures of "off-pathway" folding.

Distinction between the mechanism of BPTI and hirudin bears implications of fundamental importance because they represent two very different models of protein folding. The BPTI folding is in line with the "framework model" that stresses the importance of local interactions in reducing the conformational search and guiding the efficient protein folding through hierarchic condensation of native-like elements. The hirudin model, on the other hand, is more consistent with the "collapse model". This depicts that protein folding undergoes an initial stage of rapid hydrophobic collapse

followed by slower annealing in which specific interactions refine the structure rather than dominate the folding code.

The underlying cause of such a diversity of folding pathways seems to be associated with the fashion of how local structure elements of disulfide proteins are being stabilized and can be predicted from the mechanism of reductive unfolding of these proteins (Chang, 1997; Chang et al., 2000b). Those with their native disulfide bonds reduced in a collective and simultaneous manner (*all-or-none* mechanism) typically exhibit both a high degree of heterogeneity of folding intermediates and the accumulation of scrambled isomers along the folding pathway (e.g. hirudin, PCI and TAP) (Fig. 12). A *sequential* reduction of native disulfide bonds is indicative of the existence of stable structural domains. In these cases (e.g. BPTI and EGF) limited types of intermediates adopting native-like structures tend to dominate the folding pathway (Fig. 12).

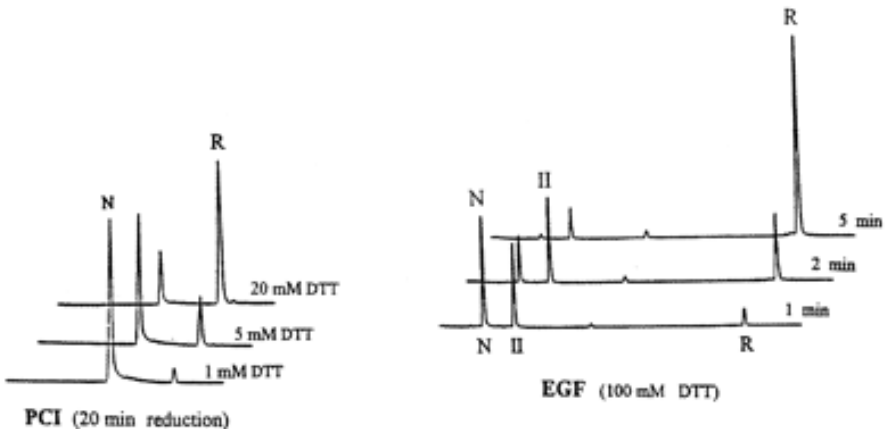


Fig. 12. Mechanisms of reductive unfolding of PCI and EGF. The native protein of PCI and EGF was treated with the indicated concentration of dithiothreitol (DTT). Time course intermediates were trapped with acid and analyzed by RP-HPLC. *N*, *II,* and *R* indicate the elution positions of the 3-disulfide native species, the 2-disulfide species, and the fully reduced species, respectively.

Analysis of the folding pathway of 4-disulfide proteins presents yet another level of technical challenge of isolation and characterization of intermediates due to the exponential increase in the number of disulfide isomers. One of the most extensively investigated models is bovine RNase A. This protein contains four disulfide bonds formed within very different secondary structure elements. While the disulfide bonds 40-95 and 65-72 occur in relatively flexible loop segments, bonds 26-84 and 58-110 join an α-helix to a β-sheet and stabilize the surrounding local hydrophobic core. Scheraga and co-workers

showed that *the folding pathway of RNase A* proceeds through a pre-equilibrium involving ensembles of unstructured species, followed by rate-determining steps from two intermediates, which then proceed to the native protein (Scheraga et al., 1984; Scheraga et al., 2001). The unstructured species comprises a heterogeneous population of intermediates with one, two, three and four (scrambled) disulfide bonds. However, only two intermediates adopting three native disulfide bonds (des[40-95] and des[65-72]; i.e. intermediates missing the disulfide bonds between Cys40-Cys95 and Cys65-Cys72, respectively) convert rapidly to the native form (Rothwarf et al., 1998). These two species are productive intermediates, whereas other intermediates such as des[26-84] and des[58-110] are metastable dead-end species that preferentially reshuffle (Welker et al., 1999). The key factor in determining the kinetic fate of these des species is the relative accessibility of both their thiol groups and disulfide bonds. Productive intermediates tend to be *disulfide-secure*, meaning that their structural fluctuations preferentially expose their thiol groups, while keeping their disulfide bonds buried (Welker et al., 2001). In contrast, dead-end species tend to be *disulfide-insecure*, in that their structural fluctuations expose their disulfide bonds in concert with their thiol groups.

Another 4-disulfide protein that has been characterized in detail is *α-lactalbumin* (αLA) (Chang, 2002a; Chang & Li, 2002b). This protein comprises an α-helical domain and a β-sheet calcium binding domain (Rao & Brew, 1989). In the absence of calcium, oxidative folding of αLA proceeds through highly heterogeneous species of 1-, 2-, 3-, and 4-disulfide (scrambled) intermediates to reach the native structure. In addition, its reduction follows a nearly all-or-none mechanism. The presence of calcium serves to generate a stable β-sheet domain, and then the folding intermediates of αLA comprise two predominant isomers (αLA-IIA and αLA-IIIA) adopting exclusively native disulfide bonds, including the two disulfide bonds located within the β-sheet calcium binding domain (Ewbank & Creighton, 1993a; Ewbank & Creighton, 1993b). Besides scrambled isomers are not observed in the folding reaction performed in presence of calcium. The native disulfide bonds of calcium-bound αLA are reduced in a sequential manner, with selective and sequential reduction of the two native disulfide bonds located within the α-helical domain. Consequently, analysis of the folding pathway(s) of αLA both in the absence and presence of calcium provides valuable information for corroborating the underlying diversity of folding pathways observed among many small disulfide-containing proteins (Chang, 2004).

Potato carboxypeptidase inhibitor

PCI is a 39-residue protein stabilized by three disulfides. Its folding pathway was elucidated by structural analysis and stop/go folding experiments of both acid and iodoacetate trapped intermediates (Chang et al., 1994b). The results revealed that as in the case of hirudin and TAP, folding of PCI proceeds through an initial stage of packing, followed by the consolidation of scrambled isomers into the native form. The process of nonspecific disulfide formation (packing) involves a sequential flow of fully reduced PCI through 1- and 2-disulfide intermediates and leads to the formation of scrambled 3-disulfide species. All three classes of intermediates are highly heterogeneous. Disulfide reshuffling occurs at the final stage where scrambled forms reorganizes and consolidates to attain the native conformation (Fig. 13).

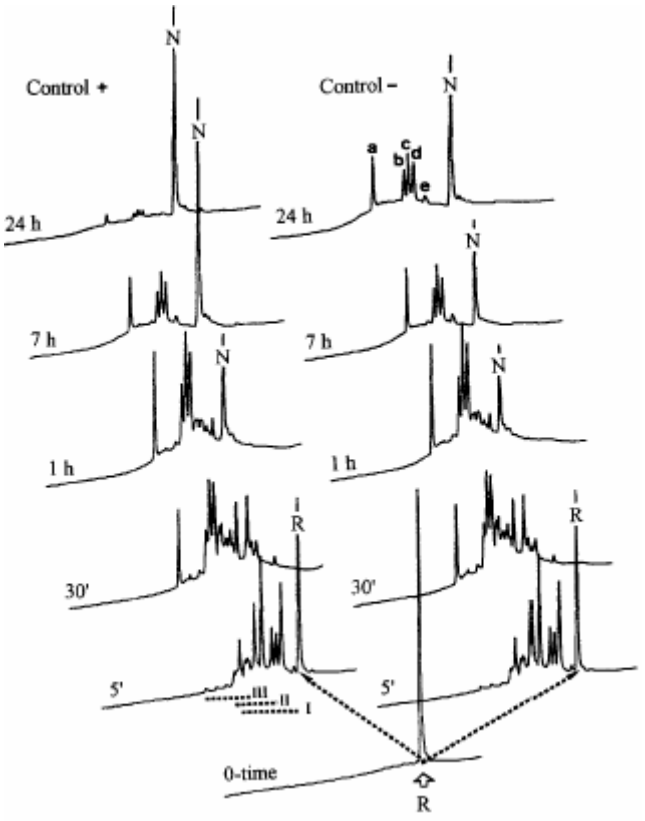


Fig. 13. RP-HPLC analysis of the acid trapped folding intermediates of PCI. Reduced/denatured PCI was allowed to refold in the presence (Control +) or absence (Control –) of 0.25 mM 2-mercaptoethanol. Folding intermediates were trapped with acid and analyzed by RP-HPLC. *N* and *R* indicate the elution positions of the native and fully reduced species of PCI, respectively. Scrambled PCI are eluted within five major fractions, *a-e*.

The efficiencies of disulfide formation and disulfide reshuffling can be selectively regulated by redox agents. Disulfide formation is promoted by cystine or oxidized glutathione, whereas disulfide reshuffling requires free thiols such as cysteine, reduced glutathione or β-mercaptoethanol. Consolidation of scrambled species to form the native

PCI represents the major rate-limiting step and when the folding reaction is carried out in the presence of for example cystine (2 mM), more than 98% of the intermediates accumulates as scrambled species after 1 min of refolding. In contrast, in presence of 0.25 mM β-mercaptoethanol, more than 90% of native PCI is recovered within 24 h. Denaturant mainly disrupts the final stage of PCI folding and exerts no apparent influence on the early stage of packing. Under conditions in which the non-covalent interactions are abrogated by denaturants scrambled structures accumulate as the end products. It suggests that non-covalent specific interactions do not actively participate in guiding PCI folding during the early phase of folding (Venhudova et al., 2001).

The unfolding and denaturation curves of PCI were investigated using the technique of disulfide scrambling (Chang et al., 2000c). In the presence of denaturant and thiol initiator, the native PCI denatures by shuffling its native disulfide bonds and converts to form a mixture of scrambled species that consist of 9 out of a possible 14 isomers. This study reveals the existence of structurally defined unfolding intermediates and a progressive expansion of the polypeptide chain. In addition, this work shows that the PCI sequence is unable to fold quantitatively into a single native structure. Thus, under physiological conditions the scrambled isomers of PCI (about 5 %) are in equilibrium with native PCI.

Leech carboxypeptidase inhibitor

LCI is a 67-amino acid protein containing four disulfide bonds that stabilize its structure. Its folding pathway was elucidated using structural and kinetic analysis of the folding intermediates trapped by acid quenching (Salamanca et al., 2003b). Reduced and denatured LCI refolds through a sequential flow of 1- and 2-disulfide intermediates and reaches a rate-limiting step in which a mixture of three major 3-disulfide species and a heterogeneous population of 4-disulfide scrambled isomers coexist and act as kinetic traps (Fig. 14). A high degree of heterogeneity of intermediates is observed during the early stages of folding and the HPLC patterns remain indistinguishable regardless of the presence of β-mercaptoethanol (Fig. 14).

Two 3-disulfide species contain only native disulfide pairings, while the other one contains one native and two non-native disulfide bonds. The addition of oxidized glutathione increases the rate of disulfide formation by approximately 12-fold. In the absence of any thiol reagent, scrambled intermediates become almost all trapped and are

unable to reshuffle efficiently their non-native disulfide bonds and convert to the native structure during the folding. Thus, in the absence of β-mercaptoethanol only approximately 40% of the protein reaches the native structure after 48 h of folding, while the remaining 60% is trapped as intermediates. Among them, 35% corresponds to III-A and III-B and 25% to scrambled forms. In the presence of reducing agent the recovery of native LCI is much higher, with more than 90% of protein in native state after 48 h of reaction.

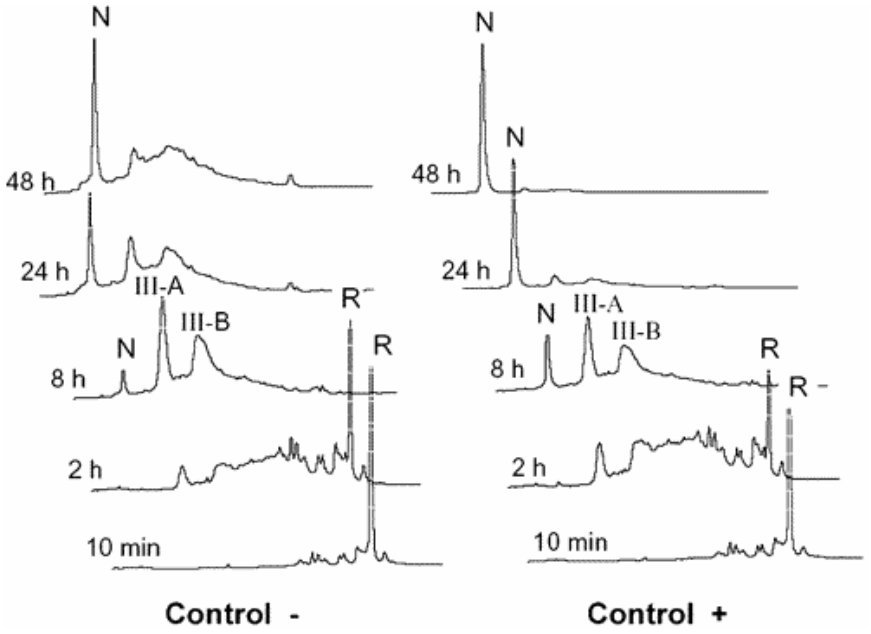


Fig. 14. RP-HPLC analysis of the folding intermediates of LCI trapped by acidification. Folding was carried out in the presence (Control +) or absence (Control –) of 0.25 mM 2-mercaptoethanol. Folding intermediates were trapped with acid and analyzed by RP-HPLC. *N* and *R* indicate the elution positions of the native and fully reduced species of LCI, respectively. *III-A* and *III-B* are two major fractions of three-disulfide intermediates identified along the pathway of oxidative folding of LCI.

The unfolding and denaturation curves of LCI were also elucidated using the technique of disulfide scrambling (Salamanca et al., 2002). In the presence of thiol initiator and denaturant, the native LCI denatures by shuffling its native disulfide bonds and transforms into a mixture of scrambled species; 9 of 104 possible isomers can be distinguished. The rate constants of unfolding of LCI are low when compared with those of other disulfide-containing proteins. The denaturation curves of LCI, determined by the fraction of native protein converted to scrambled isomers under increasing concentrations of denaturant, show that LCI is a very stable molecule; high concentrations of denaturants

such as GdnHCl or GdnSCN are required to affect LCI (Table 3), a fact that reflect the high contribution of secondary structure to the stability of the LCI fold.

Table 3. Concentration of different denaturants required to achieving 50% of denaturation for several disulfide-containing proteins. These data were determined by the method of disulfide scrambling between the native and scrambled isomers

Protein	GdnSCN	GdnHCl	Urea
		M	
Hirudin	2	5	>8
TAP	1	4.2	4
PCI	0.7	1.45	>8
LCI	2.4	3.6	>8
EGF	2.1	3.9	>8
BPTI	3.25	7.5	>8
RNase A	0.75	2.25	5.75
αLA (calcium-depleted)	0.4	1.1	3.4
αLA (calcium saturated)	-	2.15	4.75

The present thesis can be subdivided in two sections:

I. An initial section that deals with a new carboxypeptidase inhibitor from ticks (TCI): its functional and structural characterization and its study as a protein with potential biomedical applicability (works 1 and 2).

II. Another section in which two carboxypeptidase inhibitors, PCI and LCI, are used as protein models to study the determinants that account for the folding, stability and functionality of disulfide-rich proteins (works 3-6).

OBJECTIVES

The following objectives were proposed for each section:

I.
- To find, isolate and clone a new carboxypeptidase inhibitor from ticks.
- To obtain it in a pure state by recombinant expression and further chromatographic purification.
- To characterize in detail this inhibitor in terms of homology, stability and function and test its potentiality as a pro-fibrinolytic compound.
- To determine its structure in complex with carboxypeptidases and elucidate its mechanism of inhibition toward the target enzymes.

II.
- To apply site-directed mutagenesis to evaluate the importance of the secondary binding site of PCI for its binding to carboxypeptidases.
- To characterize mutations located at the PCI globular core to provide insights into the folding, unfolding and stability of this protein.
- To study in depth the oxidative folding and reductive unfolding pathways of LCI by analysis of the acid-trapped folding intermediates.
- To determine the structures of the two major folding intermediates in the oxidative folding of LCI (III-A and III-B) and the reasons for their accumulation.
- To merge experimental and theoretical approaches to postulate general rules that govern the folding of disulfide-containing proteins.

Section I. Work 1

A Carboxypeptidase Inhibitor from the Tick *Rhipicephalus bursa*

ISOLATION, cDNA CLONING, RECOMBINANT EXPRESSION, AND CHARACTERIZATION

Summary

A novel proteinaceous metallocarboxypeptidase inhibitor, named TCI (tick carboxypeptidase inhibitor), was isolated from the ixodid tick *Rhipicephalus bursa* and N-terminally sequenced. The complete cDNA encoding this protein was cloned from tick mRNA by RT-PCR and rapid amplification of cDNA ends (RACE) techniques. The full-length TCI cDNA contains an open reading frame coding for a precursor protein of 97 amino acid residues that consists of a predicted signal peptide of 22 residues and of mature TCI, a 75-residue cysteine-rich protein (12 Cys). The deduced amino acid sequence shows no homology to other known proteins; the C-terminus, however, resembles those of other protein metallocarboxypeptidase inhibitors, suggesting a common mechanism of inhibition. Recombinant TCI expressed in *Escherichia coli* is fully functional and inhibits carboxypeptidases of the A/B subfamily with equilibrium dissociation constants in the nanomolar range. Structural analyses by circular dichroism and nuclear magnetic resonance indicate that TCI is a protein strongly constrained by disulfide bonds, unusually stable over a wide pH range and highly resistant to denaturing conditions. As a tight binding inhibitor of plasma carboxypeptidase B, also known as TAFIa, recombinant TCI stimulates fibrinolysis *in vitro* and thus may have potential for applications to prevent or treat thrombotic disorders.

Keywords: metallocarboxypeptidase inhibitor; blood sucking parasites; cDNA cloning; recombinant expression; pro-fibrinolytic compounds.

Introduction

The biological actions of many proteases are tightly controlled by specific interactions with proteinaceous inhibitors (1, 2). In contrast to the wide variety of structurally and functionally diverse inhibitors of endoproteinases (3, 4), however, only few specific inhibitors that bind to metallocarboxypeptidases have been identified. The

inhibitors described so far are from *Solanacea* (tomato and potato; 38-39 residues) (5, 6), the intestinal parasite *Ascaris suum* (65 residues) (7), the medical leech *Hirudo medicinalis* (66 residues) (8), and from rat and human tissues (223 residues) (9, 10). These carboxypeptidase inhibitors differ in their mechanism of action: While the inhibitor from mammalian tissues apparently binds to carboxypeptidases through a small region that shows sequence similarity to the inhibitory loop present in their propeptides (9, 11, 12), all others known rely on the interaction of their C-terminal tail with the active site groove in a way that mimics substrate binding (13).

In recent years interest in carboxypeptidase inhibitors has been fostered by the identification of non-digestive carboxypeptidases in extrapancreatic tissues and fluids that possess diverse physiological functions (13, 14). In particular, two carboxypeptidases have been found in plasma, i.e. carboxypeptidase N (CPN) (15) and plasma CPB, also known as thrombin-activatable fibrinolysis inhibitor (TAFI) (16, 17). CPN is a constitutively active enzyme that cleaves numerous endogenous vasoactive peptides and proteins such as kinins and anaphylatoxins (18). TAFI circulates in plasma as a zymogen; the activated form TAFIa down-regulates fibrinolysis presumably by removing C-terminal lysine residues from fibrin that is already partially digested by plasmin (19). Removal of these residues, which act as ligands for the lysine-binding sites of plasminogen and tissue-type plasminogen activator (tPA), results in a reduced plasminogen activation by tPA and consequently in a prolongation of clot lysis time (20). This mechanism provides a rationale for inhibiting TAFIa in order to enhance clot lysis. Indeed, experiments performed both *in vitro* and *in vivo* have shown that the potato carboxypeptidase inhibitor (PCI) significantly enhances clot lysis induced by tPA via the inhibition of TAFIa (21, 22). Similar effects have recently been demonstrated for the leech carboxypeptidase inhibitor (LCI) (Salamanca et al., manuscript in preparation). Taken together, these data suggest a potential application of carboxypeptidase inhibitors as pro-fibrinolytic agents or lead compounds for the prevention or treatment of thrombotic disorders.

Blood-sucking animals such as leeches, insects and ticks are known to contain complex mixtures of compounds that interfere with their host's haemostatic system, allowing fast and successful feeding of blood (23, 24). Several protein inhibitors of serine proteases have been isolated from leeches and ticks that include anticoagulants such as thrombin and factor Xa inhibitors (25, 26) as well as others whose biological functions are not yet fully understood (27, 28). In the present study, we have detected, isolated, cloned, expressed and characterized a novel carboxypeptidase inhibitor from the ixodid tick

Rhipicephalus bursa (see photographs of this tick species in "Supplementary Information"). This protein, named TCI (for tick carboxypeptidase inhibitor), is the first carboxypeptidase inhibitor found in ticks. Most of TCI does not show sequence similarity with other known proteins; its C-terminus, however, resembles those of the inhibitors from *Solanacea*, *Ascaris* and *Hirudo*, suggesting a common mechanism of action. TCI inhibits different carboxypeptidases of the A/B subfamily with nanomolar affinities. Moreover, the generated recombinant form of the inhibitor (rTCI) accelerates fibrinolysis approximately 10 times more effectively than rPCI and thus may have potential for applications in coagulation disorders.

Experimental Procedures

Materials — *R. bursa* ticks were collected from cows and sheep in Catalonia, immediately frozen in liquid nitrogen, and stored at -80ºC. Restriction enzymes, ligases and DNA polymerases were obtained from Roche Applied Science and Ecogen. Roche Applied Science synthesized all oligonucleotides. Dithiothreitol, guanidine hydrochloride, reduced glutathione and oxidized glutathione were purchased from Sigma. Recombinant PCI was produced and purified as described previously (29). Rabbit lung thrombomodulin (TM) was obtained from American Diagnostica, human thrombin from Sigma, and recombinant tPA (Actilyse) from Boehringer Ingelheim. Bovine CPA and porcine CPB were purchased from Sigma, and human CPN from Elastin Product Co. Recombinant human CPA1, CPA2 and CPB were produced and purified following described procedures (30, 31). Human and bovine TAFI were a generous gift from Prof. J. J. Enghild (University of Aarhus, Denmark). The chromogenic substrates N-(4-methoxyphenylazoformyl)-Phe-OH and N-(4-methoxyphenylazoformyl)-Arg-OH were obtained from Bachem and N-(3-[2-Furyl]Acryloyl)-Ala-Lys from Sigma.

Purification of Natural TCI — Whole ticks were homogenized in Tris-acetate buffer (20 mM, pH 7.5) using a Politron, and the insoluble residue was removed by centrifugation at 20,000 x g for 30 min. The supernatant was loaded onto a Sep-Pak C_{18} Cartridge (Waters), which was subsequently eluted using different concentrations of acetonitrile (0-50%). After evaporation of acetonitrile fractions were analyzed by CPA inhibition tests, and inhibitor-containing fractions were lyophilized.

For further enrichment of TCI an affinity purification step using a CPA-Sepharose column was required. The column was prepared by coupling bovine CPA to CBr-activated

Sepharose 4B according to the guidelines of the manufacturer (Amersham Biosciences). The column was equilibrated with Tris-HCl buffer (50 mM, pH 7.5) containing 100 mM NaCl. The lyophilized TCI was reconstituted in this buffer and applied onto the CPA-Sepharose column, which was subsequently washed with water and phosphate buffer (50 mM, pH 10.0). The proteins retained on the column were eluted with phosphate buffer (50 mM, pH 12.0) and subjected to reversed-phase high performance liquid chromatography (RP-HPLC) on a Vydac C_{18} column using a 10-60% linear acetonitrile gradient in 0.1% trifluoroacetic acid (TFA) at a flow rate of 0.75 ml/min over 60 min. The purity of the isolated inhibitor was verified by matrix-assisted laser/desorption ionization - time of flight mass spectrometry (MALDI-TOF MS) and N-terminal amino acid sequence analysis.

Molecular Cloning of the Full-Length TCI cDNA — Total RNA was isolated from frozen ticks using the Nucleospin kit, and poly(A)$^+$ RNA was purified using the Nucleotrap kit (both Macherey-Nagel) according to the manufacturer's instructions. Synthesis of the first-strand of *R. Bursa* cDNA was performed using the adaptor oligonucleotide R_0R_1-dT (R_0, 5'-CCGGAATTCACTGCAG-3'; R_1 5'-GGTACCCAATACGACTCACTATAGGGC -3') and AMV Reverse Transcriptase (Roche Applied Science).

For the cloning of the TCI cDNA three degenerated oligonucleotides were designed based on the N-terminal sequence of the purified TCI: N1, 5'-AAYGARTGYGTIWSIAARGGITTYGG-3' with Y = C or T, R = A or G, I = inosin and W = A or T (corresponding to the N-terminal 9 amino acids of the isolated TCI protein); N2, 5'-AARGGITTYGGITGYYTICCICA-3' (amino acids 6-13); N3, 5'-CCICARWSIGAYTGYCCICARGARGC-3' with S = G or C (amino acids 12-20). The 3'-end fragment of the TCI cDNA was obtained using a 3'-rapid amplification of cDNA ends protocol. For the first round of PCR tick cDNA was amplified using oligonucleotide N1 and adaptor oligonucleotide R_0. PCR was conducted with 30 cycles each at 94ºC for 1 min, annealing at 55ºC for 1 min, and extension at 72ºC for 2 min. PCR mixtures were diluted 20-fold and re-amplified using the N2 and R_1 nested oligonucleotides. Subsequently, PCR mixtures were diluted again and re-amplified using the N3 and R_1 oligonucleotides. PCR products were separated by electrophoresis on 2% agarose gels. The prominent PCR product was recovered from the gel and cloned into the pGEM-T Easy vector (Promega) to generate the clone pGEM-TCI-N3/R_1.

Subsequently, the 5'-end fragment of TCI-cDNA was obtained by 5'-RACE using the Smart RACE cDNA amplification kit (BD Biosciences Clontech). Based on the sequence of the 3' cDNA end two nested oligonucleotides were designed: Gene-specific primer 1

(GSP1), 5'-TTGTAACTCGGTGCGATATCCTCG-3' (corresponding to nucleotides 444-421 of the complete cDNA; see Fig. 1), and gene-specific primer 2, (GSP2) 5'-TGATCCACGACGACATCGTGCGCC-3' (corresponding to nucleotides 420-397). After second strand cDNA synthesis and ligation of the adaptor oligonucleotide "semi-nested" PCR was performed with the adaptor primer mix and the gene-specific primers GSP1 and GSP2 in a first and second round of PCR, respectively. PCR products were analyzed and cloned as described above.

Construction of the TCI Expression Vectors — PCR was used to append the 5' bases coding for the N-terminal amino acids of mature TCI to the partial TCI-cDNA obtained by 3'-RACE. Two oligonucleotides were designed using *E. coli* codon usage: A, 5'-AACGAATGCGTGAGCAAAGGCTTTGGCTGCCTGCCGCAGTCGGATTGTCCGCA-3' and B, 5'-CAAGCTTGGTCATTAGTGCAGCCAGACGCA-3'. PCR was performed with these oligonucleotides and the plasmid pGEM-TCI-N3/R_1 as template using 30 cycles each with 94ºC for 1 min, annealing at 50ºC for 2 min, and extension at 72ºC for 2 min.

For the cloning of this semisynthetic TCI cDNA into the expression vector pBAT-4 (32) three oligonucleotides were designed using *E. coli* codon usage: C, 5'-CGGAATTCCATGAACGAATGCGTGAGCAAA-3'; D, 5'-TGGCACTGGCTGGTTT CGCTACCGTAGCGCAGGCCAACGAATGCGTGAGCAAAGG-3'; and E, 5'-CCGG AATTCATGAAAAAGACAGCTATTGCGATTGCAGTGGCACTGGCTGGTTTCGCT-3'. The product obtained by two successive PCRs (conditions see above) using the oligonucleotides D/B and E/B, respectively, to introduce the OmpA signal sequence was cleaved with Eco RI and Hind III and ligated with the pBAT-4 vector to generate pBAT-4-OmpA-TCI. Similarly, the product obtained using oligonucleotides C and B was ligated using the same restriction sites with pBAT-4 to generate pBAT-4-TCI. The pBAT-4-OmpA-TCI vector was transformed into *E. coli* strain BL21(DE3) whereas pBAT-4-TCI was used to transform *E. coli* strains BL21(DE3) and B834.

Heterologous Expression of TCI — For the production of recombinant TCI (rTCI) an overnight culture of *E.coli* BL21(DE3)/pBAT-4-OmpA-TCI grown in Luria-Bertani medium was used to inoculate 0.5 liter of M9CAS medium containing 0.5% glycerol. At an OD_{550} of 1 the culture was induced with 1 mM IPTG. After 24 h of induction the culture was centrifuged and the supernatant was applied onto a Sep-Pak C_{18} Cartridge. The column was washed with water followed by 10% acetonitrile, and the bound proteins were eluted with 30% 2-propanol and then concentrated in a Roto-Vapor to remove the organic solvent. The rTCI was further purified by cation exchange chromatography (Fractogel

EMD SO_3^-, Merck) using MES buffer (50 mM, pH 5.5) containing 1 M NaCl as elution buffer. Inhibitor-containing fractions were kept at 20ºC for 24 h with 1 mM reduced/0.5 mM oxidized glutathione at pH 8.5 to promote the formation of disulfide bonds. Finally, the protein was purified by HPLC on a Vydac C_4 column using a linear gradient from 10-40% acetonitrile in 0.1% TFA at a flow rate of 0.75 ml/min for 40 min.

For the intracellular expression of rTCI BL21(DE3)/pBAT-4-TCI or B834/pBAT-4-TCI were grown in Luria-Bertani medium and used to inoculate 0.5 liter of the same medium. At an OD_{550} of 0.8-1.0 IPTG was added to a final concentration of 1 mM. After 8 h of induction the culture was centrifuged and cells were sonicated in Tris-HCl buffer (20 mM, pH 8.5) containing 0.5 mM EDTA. After centrifugation the pellet was washed with Tris-HCl buffer containing 0.5 mM EDTA and 2% Triton X-100, dissolved in Tris buffer containing 6 M guanidine hydrochloride and 30 mM dithiothreitol, and kept for 6 h at 20°C. Subsequently, the sample was dialyzed against Tris buffer containing 1 mM reduced/0.5 mM oxidized glutathione for 24 h before measurements of CP inhibitory activity.

Molecular Mass Determination, Carboxymethylation, and Amino Acid Sequencing — The molecular masses of natural and of recombinant TCI were determined by MALDI-TOF MS on a Bruker Ultraflex spectrometer. Samples were prepared by mixing equal volumes of the protein solution and matrix solution (sinapic acid in 30% acetonitrile with 0.1% TFA). To determine the number of cysteines the inhibitor (0.5 μg) was dissolved in ammonium bicarbonate buffer (0.2 M, pH 8.5) and further denatured and reduced by incubation with 5 M guanidine hydrochloride and 30 mM dithiothreitol for 2 h at 50°C. Carboxymethylation of cysteines was achieved by addition of 0.5 M iodoacetamide for 1 h at room temperature in the dark. The solution was diluted ten times with water and analyzed using the same spectrometer.

N-terminal amino acid sequence analysis was carried out by automatic Edman degradation using a Beckman LF3000 Protein Sequencer.

Circular Dichroism and NMR Spectroscopy — Samples for circular dichroism (CD) spectroscopy were prepared by dissolving the proteins to a final concentration of 0.5 mg/ml in 0.1% aqueous TFA (pH 2.0) or sodium phosphate buffer (20 mM, pH 6.5 or 8.0). CD analyses were carried out in a Jasco J-715 spectrometer at 25ºC or 90ºC using a cell with 2 mm path length. Reduction and/or denaturation were achieved by treating rTCI with 10 mM dithiothreitol for 24 h and/or dissolving the protein in 6 M guanidine

hydrochloride. Protein samples for ^{1}H and 2D NMR experiments were prepared by dissolving the proteins in phosphate buffer (20 mM, pH 6.5) to a final concentration of 1 mM. NMR spectra were acquired on a Bruker AMX 500 MHz spectrometer at different temperatures using both normal and deuterated water.

CP Inhibitory Activity — The inhibitory activity of rTCI was assayed by measuring the inhibition of the hydrolysis of the chromogenic substrate N-(4-methoxyphenylazoformyl)-Phe-OH by carboxypeptidases type A and N-(4-methoxyphenylazoformyl)-Arg-OH by carboxypeptidases type B at 350 nm. TAFI was activated to generate TAFIa immediately before use by incubation (2.2 μM) with TM (200 nM) and thrombin (160 nM) in Tris-HCl buffer (20 mM, pH 7.4) containing 150 mM NaCl at 23ºC for 40 min. All assays were performed in Tris-HCl buffer (50 mM, pH 7.5) containing 100 mM NaCl and 100 μM substrate. Inhibition constants (K_i) for the complexes of recombinant TCI and PCI with the different carboxypeptidases were determined using presteady-state kinetics as described for tight binding inhibitors (33). The concentration of rTCI used in these assays was determined based on the A_{280} and a calculated extinction coefficient of $E_{0.1\%}$=0.97.

The inhibitory activity of rTCI toward CPN was tested by measuring its effect on the hydrolysis of the chromogenic substrate N-(3-[2-Furyl]Acryloyl)-Ala-Lys. The assay was performed in Tris-HCl buffer (50 mM, pH 7.5) containing 100 mM NaCl and 0.35 mM of the substrate at 37ºC and 340 nm.

In Vitro Clot Lysis Assays — Clot lysis assays were performed in 96-well microtiter plates using control human plasma (Dade Behring) diluted 1:3 with Hepes buffer (20 mM, pH 7.4) containing 0.15 M NaCl. Clotting and subsequent fibrinolysis was initiated by simultaneous addition of thrombin (5 nM final concentration), $CaCl_2$ (10 mM) and tPA (1.2 or 2.4 nM) in the absence or presence of TM (10 nM). The effects of rPCI and rTCI on clot lysis time were assessed by addition of various concentrations (0-2000 nM) into the assay mixture. In all experiments microtiter plates were sealed with clear tape to avoid evaporation and the turbidity of the clots at 405 nm was monitored at 37ºC using a Biolumin 960 plate reader (Molecular Dynamics). The time required for 50% clot lysis was determined graphically as the midpoint between the maximum absorbance after clot formation and the minimum after complete clot lysis.

Results

Purification, Protein Sequence Analysis, and Mass Determination of TCI — TCI was isolated from *Rhipicephalus bursa* ticks using a combination of reversed-phase and affinity chromatography. A crude homogenate was prepared from 200 adult specimens, clarified by centrifugation, and injected into a reversed-phase cartridge (Sep-Pak C_{18}). Fractions were eluted with increasing acetonitrile concentrations, evaporated and tested for inhibition of bovine CPA. The majority of the inhibitory activity was recovered in the fraction eluted at 20% acetonitrile. The material was subsequently loaded at intermediate pH onto an affinity column prepared by immobilization of bovine CPA on a Sepharose 4B resin. The inhibitor was strongly bound and could only be eluted using an extreme-alkaline pH (phosphate buffer, pH 12); the recovery of the active inhibitor at such an extreme pH indicated a strong affinity toward CPA and a high stability of the protein. In a final RP-HPLC step a virtually homogeneous TCI preparation was obtained in a peak with a retention time of ~31 min (10-60% linear acetonitrile gradient over 60 min). Using this procedure, approximately 2 μg TCI were isolated from 200 ticks.

The purified TCI was analyzed by automated Edman degradation and MALDI-TOF MS. The sequence NH_2-NECVSKGFGCLPQSDCPQEARLSYGGCSTV was obtained for the N-terminal 30 residues. Mass spectrometry revealed a molecular mass of 7798 Da for the native inhibitor. After reduction and carboxymethylation MS showed the incorporation of 12 carboxymethyl groups and no incorporation when the previous reduction was absent, suggesting that TCI contains 12 cysteine residues that are involved in disulfide bond formation.

Cloning and Sequence Analysis of TCI cDNA — A combination of 3' and 5' RACE was used to clone the cDNA of TCI. Three degenerated oligonucleotides (N1, N2 and N3) were designed based on the N-terminal peptide sequence derived from the isolated TCI protein. First strand cDNA synthesized from *R. bursa* poly(A)$^+$ RNA was used as template for three successive rounds of PCR with these oligonucleotides in combination with the two 3'-RACE adaptor oligonucleotides as primers. A major PCR product of about 340 base pairs was generated, purified and cloned. Sequence analysis of the plasmid showed the presence of a 3'-cDNA fragment (nucleotides 166-494 of the complete cDNA; see Fig. 1). This partial cDNA sequence was used to design two specific oligonucleotides for 5'-RACE. A fragment of ~ 450 base pairs was amplified by nested PCR and cloned. Seven

clones were sequenced that all contain fragments representing nucleotides 1-420 of the TCI cDNA.

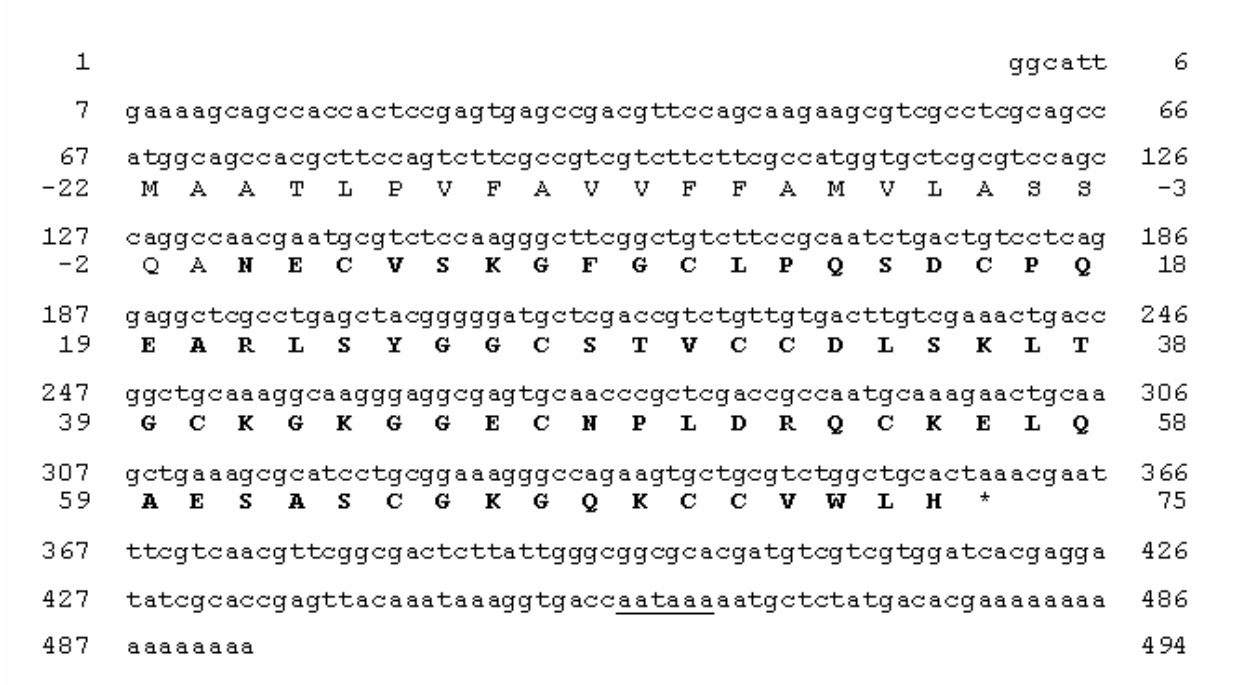


Fig. 1. Nucleotide and deduced amino acid sequence of TCI. The cDNA sequence was assembled from two overlapping cDNA clones (nucleotides 1-420 and 166-494, respectively). The amino acid sequence of mature TCI is printed in bold; amino acid residues 1-30 were also obtained by sequencing of the protein isolated from ticks. The consensus polyadenylation signal is underlined.

As shown in Figure 1 the full-length TCI cDNA consists of 494 nucleotides and contains a 66-nt 5'-untranslated sequence, an open reading frame (ORF) of 291 nucleotides, and a 137-nt 3'-untranslated region. A canonical polyadenylation signal AATAAA is found 16 nucleotides upstream the poly(A)$^+$ tail. The deduced protein consists of 97 amino acid residues. A 22-residue hydrophobic signal peptide, predicted by the program SignalP (34), precedes the N-terminal Asn residue found in the TCI protein isolated from ticks. With 7935 Da the calculated molecular mass of the deduced mature protein (75 amino acid residues) is larger than that of natural TCI determined by MS (7798 Da). The observed difference of 137 Da corresponds to the mass of the C-terminal His^{75} residue that most likely is removed by proteolytic processing during the CPA-affinity chromatography. Computer search analyses did not show significant homology with any deposited amino acid or nucleotide sequence.

Heterologous Expression of rTCI — PCR was used to generate the 5'-bases that are missing in the partial cDNA obtained by 3'-RACE and code for the 20 N-terminal amino

acid residues of mature TCI. The resulting semi-synthetic cDNA was ligated into the expression vector pBAT-4 (see "Experimental Procedures" for details) that was previously utilized for the recombinant production of PCI and LCI (29). The resulting expression plasmid pBAT-4-OmpA-TCI was used to transform *E. coli* strain BL21(DE3). After induction of expression in shaker flasks recombinant TCI was predominantly detected in the culture medium, suggesting that the OmpA signal peptide correctly directed the immature protein through the plasma membrane into the bacterial periplasmic space with concomitant removal of the signal peptide and subsequent release of mature rTCI through the outer membrane into the culture medium. Approximately 2 mg of rTCI per liter of supernatant were produced of which 40-50% was recovered by purification using reversed-phase and cation exchange chromatography (see "Experimental Procedures"). In this recombinant preparation several rTCI species were detected by RP-HPLC; these species most likely reflect different folding intermediates because one predominant species was obtained after treatment with reduced/oxidized glutathione (pH 8.5, 24 h), which showed a chromatographic retention time similar to that of natural TCI isolated from ticks. N-terminal amino acid sequence analysis and MALDI-TOF MS (7935 Da) verified the identity and correct processing of rTCI. In addition, MS analysis after treatment with the alkylating reagent vinylpyridine showed no incorporation of vinylpyridine groups suggesting that all cysteine residues of this rTCI preparation are involved in disulfide bond formation.

During optimization of rTCI production also intracellular expression systems were examined. Unfortunately, neither soluble intracellular TCI nor inclusion bodies were detectable after recombinant expression using the systems BL21(DE3)/pBAT-4-TCI or B834/ pBAT-4-TCI.

Conformational Properties of rTCI — The structural features of recombinant TCI were analyzed using CD and NMR spectroscopy. rTCI displays a peculiar CD spectrum with a well-defined minimum of ellipticity at 204 nm (Fig. 2A). This band may be related to the presence of a high percentage of residues in β–structures and loops (35). CD measurements at different pH values show that rTCI essentially has the same structure at pH 2.0, 6.5 and 8.0; the obtained spectra are virtually identical. The CD spectrum of TCI is similar to those of PCI and LCI (8, 36), but does not possess a maximum of ellipticity at 228 nm; the presence of this 228-nm band was attributed to a characteristic environment of the Tyr residue located at the C-terminus of both PCI and LCI.

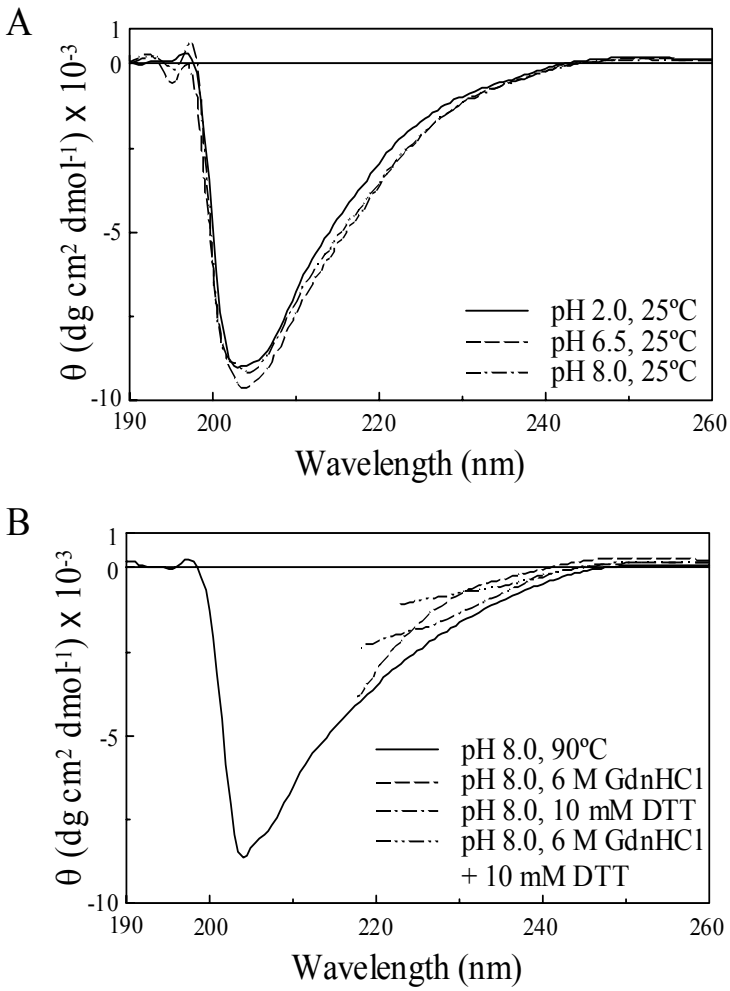


Fig. 2. Circular dichroism analysis of rTCI. A. The CD spectrum of native rTCI was recorded at 25ºC in 0.1% aqueous TFA (pH 2.0) and in sodium phosphate buffer (20 mM, pH 6.5 or 8.0) at a protein concentration of 0.5 mg/ml. B. CD analysis of rTCI was carried out at 90ºC in sodium phosphate buffer (20 mM, pH 8.0) and after incubation at 25ºC for 24 h in buffer containing 6 M guanidine hydrochloride, 10 mM dithiothreitol or both reagents.

CD spectroscopy measurements have shown that the degree of denaturation of PCI and LCI correlates with the disappearance of their characteristic CD ellipticity bands (8, 36). The 204-nm band of TCI maintains most of its intensity when the temperature is raised from 25 to 90ºC, and the native spectrum is completely recovered after lowering the temperature back to 25ºC. Even treatment with 6 M guanidine hydrochloride does not significantly affect the characteristic spectrum of TCI (Fig. 2B). However, the reducing agent dithiothreitol (10 mM) strongly decreases the intensity of the 204-nm band, and this effect is even stronger in the presence of denaturants, causing an almost complete disappearance of such band.

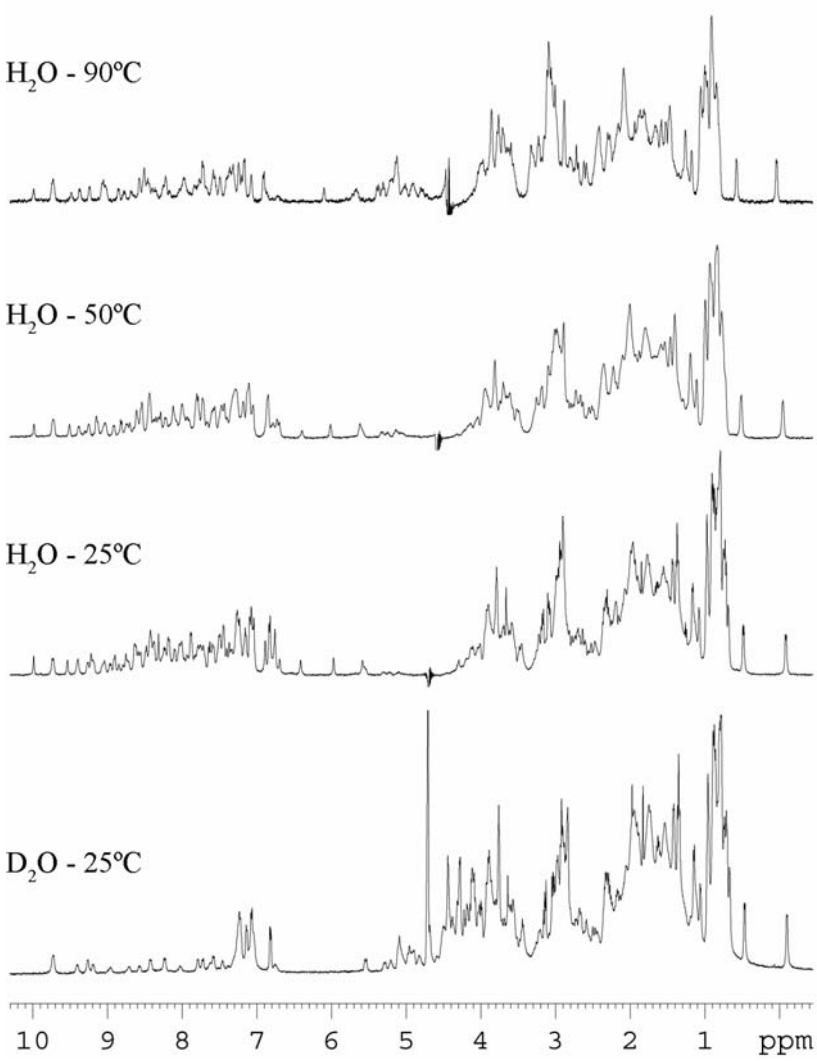


Fig. 3. ^{1}H-NMR analysis of rTCI in H_2O and D_2O. Monodimensional spectra of rTCI at 1 mM concentration were recorded in a 500 MHz NMR Bruker spectrometer. Spectra were obtained at 25, 50, and 90ºC in phosphate-buffered H_2O and at 25ºC in phosphate-buffered D_2O (each 20 mM, pH 6.5).

^{1}H-NMR spectra of rTCI showed a wide signal dispersion of resonances at both the low (amide and aromatic region) and high fields (methyl region) with a significant number of potential methyl protons in the 0-1 ppm region, and rich in exchangeable resonances in the NH region (Fig. 3). A high resistance of NH protons to the exchange with the solvent was detected after raising the temperature from 25 to 90ºC in H_2O. In contrast, most of these protons were already exchanged at 25ºC in D_2O (Fig. 3). Interestingly, at high temperatures the shape/dispersion of the spectra in the 0-1 ppm region was very similar to that at 25ºC in both H_2O and D_2O. The 2D-NMR analysis of rTCI also showed a significant resistance of NOEs contacts against temperature. Thus, as shown in Fig. 4, the

fingerprint region is similar at 25ºC and 75ºC in 10% D_2O. Taken together, CD and NMR analyses indicate a high conformational thermostability and resistance against denaturants of rTCI, which requires the presence of reducing agent to unfold.

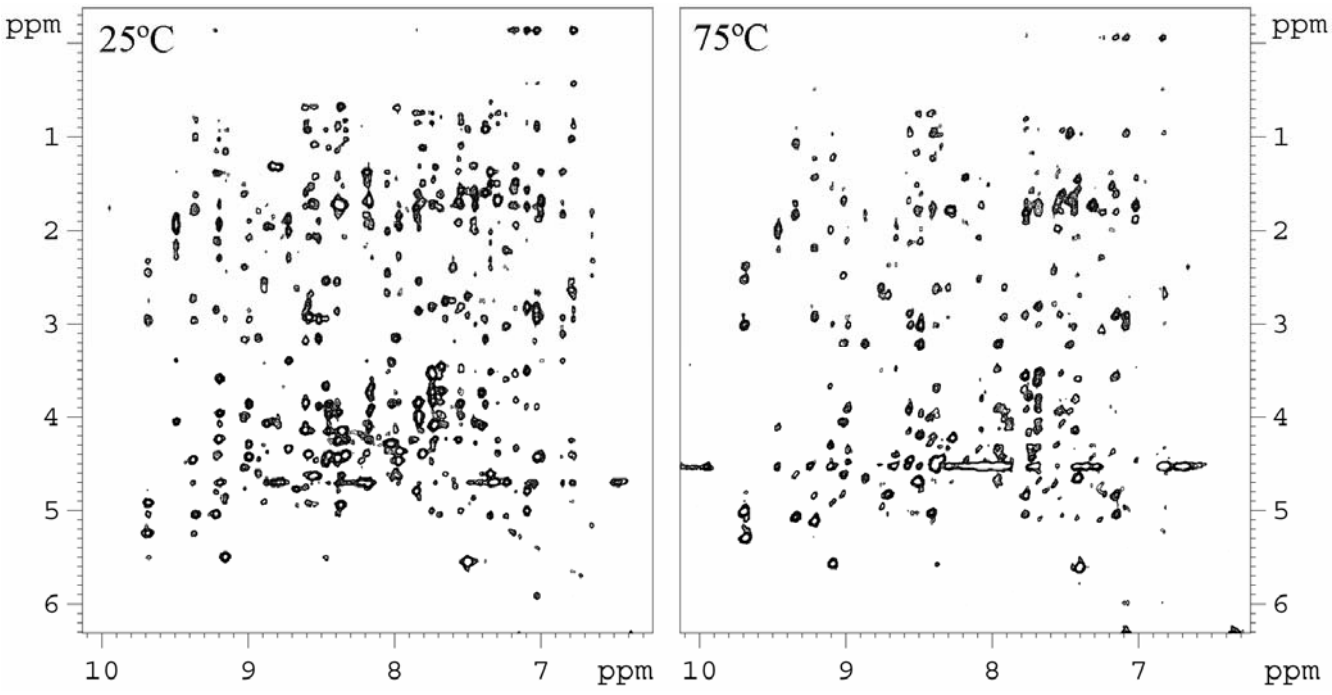


Fig. 4. Fingerprint region of the NOESY spectra of rTCI. 2D-NMR spectra were obtained both at 25 and 75ºC at a protein concentration of 1 mM in 10% D_2O. The spectra are unsymmetrized.

Inhibitory Activity and Selectivity of TCI — Equilibrium dissociation constants for the complexes of rTCI with different metallocarboxypeptidases were determined using a presteady-state approach. It is deduced that rTCI is a tight binding, competitive inhibitor of different carboxypeptidases of the A/B subfamily with K_i values in the nanomolar range (see Table 1). Compared to rPCI, the affinity of rTCI toward porcine CPB and human TAFIa is ~5-fold higher whereas bovine CPA and human CPB are inhibited with similar K_i-values. rTCI is also ~2-fold better inhibitor toward CPA2 and bovine TAFIa than the potato inhibitor. Both rTCI and rPCI have no effect on the enzymatic activity of the plasma "regulatory carboxypeptidase" CPN even at a much higher concentration than in the assays with the other carboxypeptidases (100 μM vs. ~10 nM). Titration experiments suggest that both rTCI and rPCI inhibit bovine CPA and TAFIa with a 1:1 stoichiometry and that the recombinant proteins are fully functional (specific activity > 90% of the theoretical values).

Table 1. Dissociation constants K_i for the complexes of rTCI and rPCI with different metallocarboxypeptidases

	rTCI	rPCI
Carboxypeptidase	K_i [nM]	
Bovine CPA	1.1 ± 0.3	1.6 ± 0.2
Human CPA1	1.2 ± 0.4	1.6 ± 0.3
Human CPA2	3.6 ± 0.5	8.8 ± 0.7
Porcine CPB	1.6 ± 0.3	7.2 ± 0.6
Human CPB	1.3 ± 0.2	1.8 ± 0.3
Bovine TAFIa	1.3 ± 0.3	2.5 ± 0.4
Human TAFIa	1.2 ± 0.4	5.3 ± 0.5
Human CPN	ϕ	ϕ

ϕ, no inhibition at 100 μM concentration of inhibitor

It is worth mentioning that the K_i-values determined using rTCI and TCI isolated from ticks are virtually identical (data not shown) verifying that the recombinant protein is correctly folded. In addition, these results demonstrate that the inhibitory activity of TCI is not significantly affected by the presence of the C-terminal His^{75} residue that, supposedly, is removed by proteolytic processing during the isolation of TCI from ticks. Similarly, the removal of the C-terminal residue of PCI and LCI does not significantly affect their affinity toward different carboxypeptidases (8, 37).

Effect of rTCI on the tPA-induced Lysis of Clots — Clot formation and lysis assays were performed to compare the effectivity of rTCI and rPCI to inhibit TAFIa during fibrinolysis. Clots lysis was induced by addition of 1.2 nM tPA in the absence and presence of either rPCI (Fig. 5, panel A) or rTCI (Fig. 5, panel B). In this assay thrombomodulin (TM) enhances the activation of TAFI by thrombin (38), thus prolonging clot lysis time (see curves with open symbols in Fig. 5). High concentrations of either rPCI or rTCI restore clot lysis time to that observed in the absence of TM (see closed triangles) because their pro-fibrinolytic effect is due to the inhibition of TAFIa. However, depending on the concentration used PCI either prolongs (circles) or shortens (triangles) lysis, which has been explained by its effect on both the activity and the stability of TAFIa (39). rTCI has a similar biphasic effect but the prolongation of fibrinolysis is much less pronounced than that caused by rPCI.

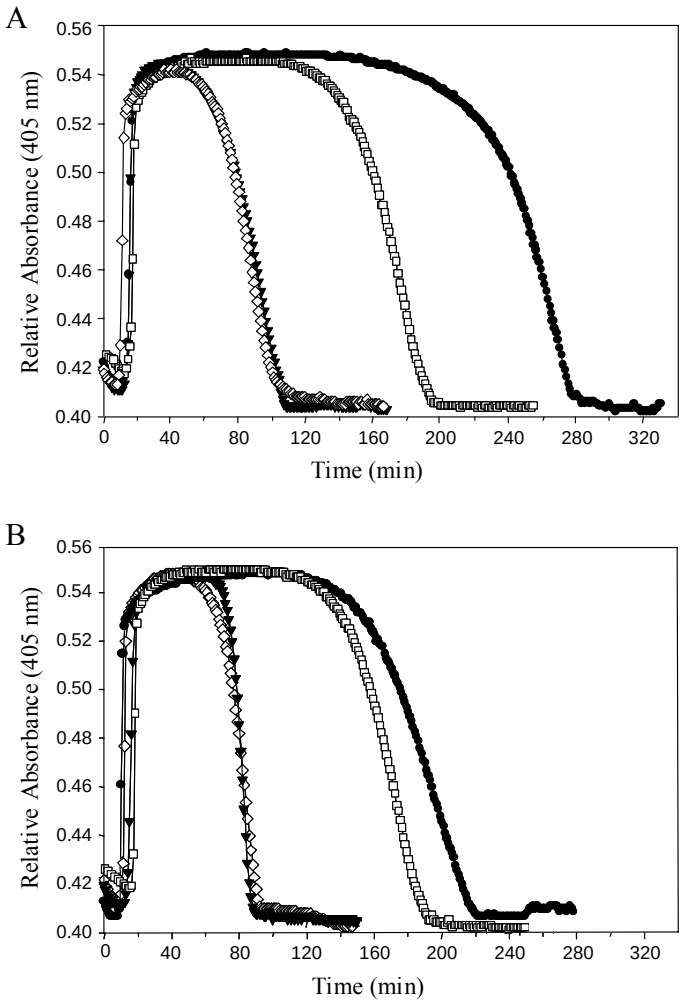


Fig. 5. Effects of rTCI and rPCI on clot formation and lysis. Coagulation and subsequent fibrinolysis were induced by concomitant addition of thrombin (5 nM), thrombomodulin (10 nM), calcium chloride (10 mM), and tPA (1.2 nM) to diluted human plasma. The extend of clotting was followed by measuring the change of turbidity over time and compared to that observed in the absence (◇) and presence of thrombomodulin (□). A. rPCI significantly delays fibrinolysis at low concentrations (50 nM, ●) and accelerates clot lysis at higher concentrations (1000 nM, ▼). B. Compared to rPCI, low concentrations of rTCI (2 nM, ●) retard fibrinolysis to a much smaller extend whereas higher concentrations (100 nM, ▼) accelerates lysis comparable to rPCI.

In order to compare the effects of rTCI and rPCI on fibrinolysis, clot lysis times were calculated as the time required for 50% clot lysis. Figure 6 shows the effects of rPCI (panel A) and rTCI (panel B) on clot lysis induced by either 1.2 or 2.4 nM tPA. Depending on the concentration used, both rTCI and rPCI prolong or shorten lysis time. The anti-fibrinolytic potential of rPCI is substantial: At a concentration of 20 nM it prolongs clot lysis induced by 1.2 nM tPA ~1.7-fold (clot lysis time 170 ± 7 and 285 ± 13 min in the absence and

presence of rPCI, respectively; mean ± SD, n=4). rPCI concentrations ≥ 200 nM are required to accelerate fibrinolysis, and a 2.0-2.5-fold acceleration is achieved at concentrations ≥ 1000 nM (clot lysis time 170 ± 7 and 80 ± 6 min in the absence and presence of 1500 nM rPCI and 1.2 nM tPA, respectively). These results are in agreement with previous studies performed using commercial PCI (39, 40). Compared to rPCI, the effect of rTCI differs in three aspects: 1) rTCI has only a minor pro-fibrinolytic effect at low concentrations; 2) approximately 10-fold lower concentrations (≥ 15 nM) are sufficient to accelerate fibrinolysis; and 3) the maximal pro-fibinolytic effect is more pronounced (2.8-fold acceleration at 500 nM).

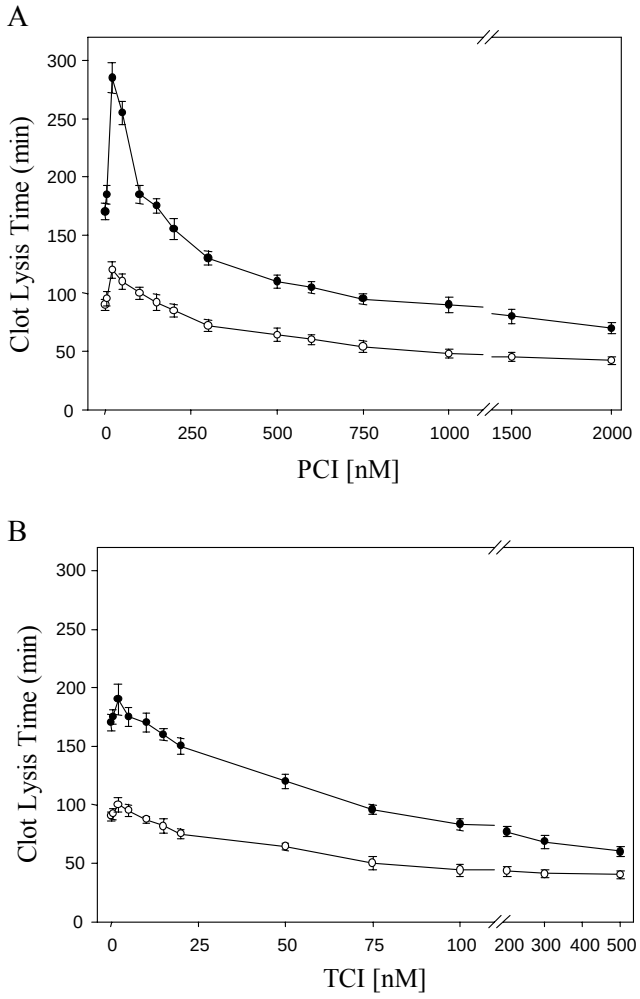


Fig. 6. Effects of rTCI and rPCI on clot lysis time. The effects of rPCI (A) and rTCI (B) on fibrinolysis induced by either 1.2 nM (●) or 2.4 (○) nM tPA were quantified as described in "Experimental Procedures". Clot lysis time was defined as the time required to reduce clotting by 50%. Data are presented as mean ± SD of 4 independent experiments.

Discussion

In the present study we have isolated, cloned, expressed and characterized a novel metallocarboxypeptidase inhibitor, named TCI, from the ixodid tick *Rhipicephalus bursa*. TCI is a 75-residue protein rich in glycines (10 residues) and cysteines (12 residues), the latter involved in the formation of 6 disulfide bridges. The cDNA of TCI was obtained by a combination of 3' and 5' RACE approaches; it contains an open reading frame coding for a protein of 97 residues that consists of a hydrophobic signal sequence (22 residues) preceding mature TCI. A signal peptide cleavage site is predicted between Ala^{-1} and Asn^{+1} (see Fig. 1), verifying the N-terminal asparagine residue obtained by the N-terminal amino acid sequencing of TCI isolated from ticks. The mature protein deduced from the cDNA contains an additional C-terminal His^{75} residue that was not detected by MALDI-TOF MS of the isolated protein and most likely is proteolytically removed during affinity purification from the tick extract (see below). The cDNA and amino acid sequences of TCI do not show significant similarity to those of known proteins.

For recombinant periplasmatic expression in *E. coli* a semi-synthetic gene coding for a fusion protein of the OmpA signal sequence and mature TCI was constructed by PCR based on the 3' RACE cDNA clone. The expression yield was lower than those observed for rPCI and rLCI using the same expression system: 2 mg vs. 5 and 11 mg per liter culture, respectively (29; Arolas et al., unpublished data), most likely reflecting the higher number of disulfide bridges of TCI (6 vs. 3 and 4 in PCI and LCI, respectively). Indeed, during purification several rTCI species were detected by RP-HPLC that appear to be folding intermediates because only one predominating species accumulated after treatment with reduced/oxidized glutathione. The elution profile and the inhibitory properties of this recombinant protein are very similar to those of the naturally occurring inhibitor, confirming that rTCI is correctly folded and functionally equivalent to the protein isolated from ticks. In addition, titration experiments indicate that rTCI displays a fully inhibitory capability. Attempts to improve rTCI yields e.g. by using an intracellular *E. coli* expression system failed so far, neither soluble rTCI not inclusion bodies were detectable in the cytoplasm.

CD and NMR studies demonstrate that rTCI retains a well-folded conformation over a wide range of pH and temperatures and is unusual resistant to denaturing conditions. The CD spectra of this small disulfide-rich molecule suggest a high content in residues forming β-structures and/or loops, in agreement with its richness in glycine and cysteine residues.

However, the anomalous dichroic bands that are similar to those reported for PCI and LCI hinder conformational predictions (8, 36). TCI seems to possess few secondary structures that could be involved in the formation of an internal core, because the protons in the NH-NMR region exchange with the solvent (D_2O) on a very short time scale at low temperatures. Thus, the observed maintenance of a proper conformation under extreme pH and temperature conditions appears to depend primarily on the establishment of disulfide bonds that constrain its structure.

TCI is a tight binding inhibitor that inhibits different metallocarboxypeptidases with equilibrium dissociation constants in the nanomolar range. Several reasons provide support for the hypothesis that the C-terminus of TCI is the primary binding site for binding to these carboxypeptidases. First, the C-terminal residue of TCI as well as those of PCI and LCI (His, Gly and Glu, respectively) is removed after binding to carboxypeptidases. Despite the absence of such C-terminal residue, the activity of these inhibitors is unaltered. Secondly, the C-tail of TCI is similar to those of other carboxypeptidase inhibitors in amino acid composition and certain sequence features (Fig. 7). This similarity observed in otherwise completely different molecules that are derived from evolutionary highly distant species, i.e. blood-sucking animals and *Solanacea* plants, is likely to result from convergent evolution dictated by the three-dimensional structures of the target metallocarboxypeptidases.

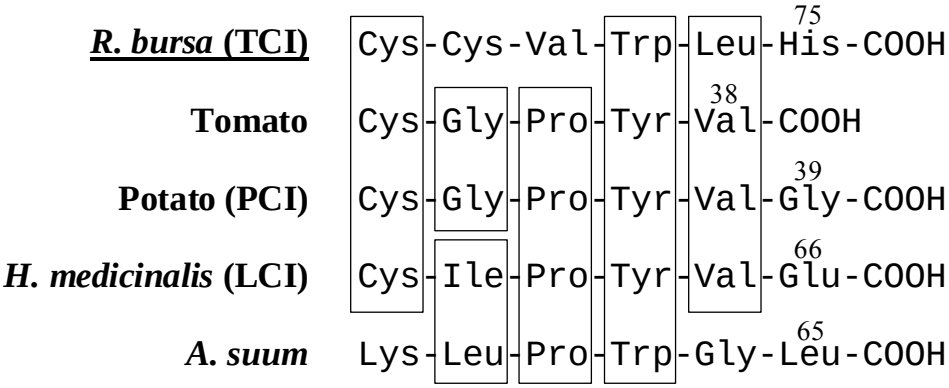


Fig. 7. Comparison of the C-terminal amino acid sequences of TCI and other carboxypeptidase inhibitors. Similar residues are boxed. The sequences of the inhibitors are taken from references 5-8 and from this work.

X-ray crystallographic studies have shown that only two residues of PCI and LCI are located within the active site of the inhibited carboxypeptidase, i.e. the penultimate Tyr and the newly generated C-terminal Val residue, the latter playing a key role in the inhibition because its carboxylate group is coordinated to the catalytic zinc of the enzyme (37, 41). Similarly, mutagenesis studies using PCI have indicated that these Tyr and Val are important residues in the interaction with the carboxypeptidase while the preceding Pro mainly contributes to maintain the rigidity of the tail (42, 43). Tyr and Val can be substituted by another aromatic residue and Leu, respectively, without loss of inhibitory capability. Thus, after removal of the His^{75} residue, the C-terminal Trp^{73} and Leu^{74} of des-His^{75}-TCI could mimic the Tyr and Val present in PCI and LCI, resulting in a similar conformational arrangement in the complex with a carboxypeptidase (see Fig. 8).

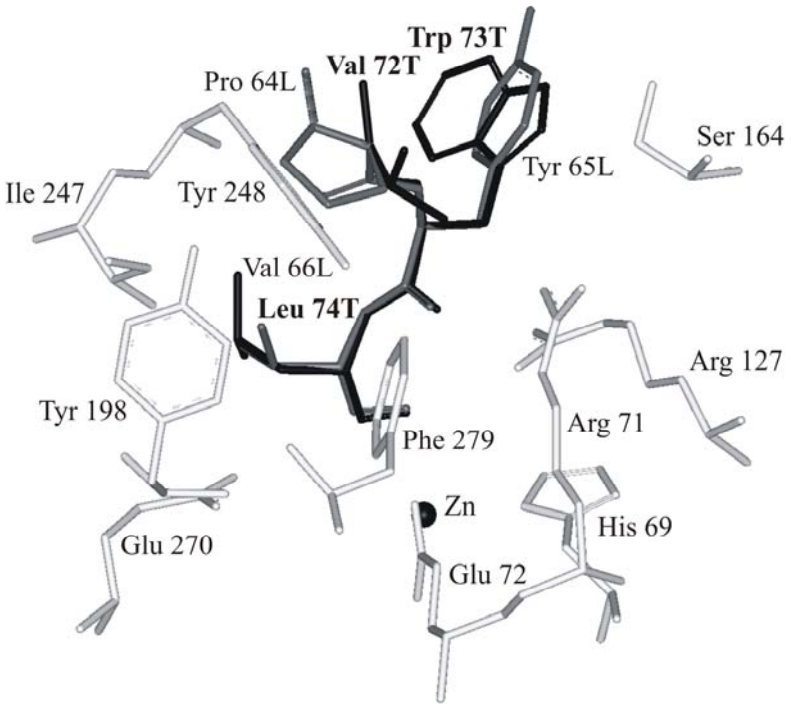


Fig. 8. Model of the interaction of the C-terminus of TCI with CPA. The position of the C-terminal amino acids of des-His^{75}-TCI within the active site groove of CPA was modeled based on the X-ray structure of the LCI-CPA primary binding site (41) using Swiss-Pdb Viewer and optimized by 2000 cycles of the steepest descent algorithm using the GROMOS package. CPA residues are shown in light gray and the active site Zn in black. LCI C-tail residues (Pro 63L, Tyr 64L, Val 65L) are indicated in dark gray and TCI residues (Val 72T, Trp 73T and Leu 74T) in black.

Finally, TCI binds and inhibits carboxypeptidases of the A/B subfamily without affecting CPN, a carboxypeptidase that belongs to the "regulatory" subfamily (also known

as subfamily N/H). This is consistent with the behavior of other protein carboxypeptidase inhibitors (44). The crystal structure of the duck CPD domain II, another regulatory carboxypeptidase subfamily member, reveals a long loop that shapes the border of the funnel into the active site and probably hinders the binding of these inhibitors through their C-terminus (45).

It has been well established that the inhibition of TAFIa by PCI affects tPA-induced clot lysis both *in vitro* and *in vivo* (21, 22, 46). More recently, PCI and the synthetic inhibitor 2-guanidinoethylmercaptosuccinic acid (GEMSA) have been shown to either accelerate or inhibit fibrinolysis depending on the concentration used, which has been explained by a dual effect on the activity and the stability of TAFIa (39, 40). Motivated by the higher affinity of rTCI compared to (r)PCI toward TAFIa we have also compared their fibrinolytic activity. The results (see Fig. 5 and 6) show that both inhibitors have a biphasic effect on tPA-mediated clot lysis of human plasma. However, rTCI is more potent that rPCI, i.e. both approximately 10-fold lower concentrations are required to accelerating fibrinolysis and the pro-fibrinolytic effect at high concentrations is more pronounced. These differences may in part originate from the ~5-fold higher affinity of TCI compared to PCI toward TAFIa (see Table 1). More importantly, compared to PCI rTCI has only a minor pro-fibrinolytic effect at low concentrations. This may result from a higher stability of the rTCI/TAFIa complex and reflect the larger size of TCI (nearly twice of that of PCI) and additional or more efficient secondary contacts with the enzyme. Based on these results rTCI has therapeutic potential as an adjuvant to augment tPA-based strategies in thrombosis. In particular, the minor pro-fibrinolytic “adverse” effect at low concentrations might be important at the beginning and the end of a therapeutic regime e.g. for the prevention or treatment of thrombotic disorders (47-49).

Considering the few carboxypeptidase inhibitors identified so far, it is interesting to note that TCI together with LCI, an inhibitor previously isolated from the medical leech (8), is the second inhibitor of metallocarboxypeptidases identified in haematophagous ectoparasites. These species are well known to contain a variety of serine proteinase inhibitors (25, 26), most prominently among them the thrombin and factor Xa inhibitors that are potent anticoagulants. It appears likely that TCI and LCI contribute to the maintenance of blood liquidity during feeding and inside the animal by stimulation of fibrinolysis via inhibition of TAFIa. Similarly, vampire bats rely on the stimulation of fibrinolysis rather than the inhibition of coagulation during feeding by secretion of plasminogen activators in their saliva (50). In addition, TCI and LCI may modulate

inflammation and host defense mechanisms e.g. by hindering the inactivation of kinins by TAFIa (51-53), thus increasing blood flow at the site of the bite, or by affecting the activity of CPA released from mast cell that are activated during parasitic infections (54). Such mechanisms to modulate inflammation and host defense may be particularly relevant for leeches and hard ticks (Family Ixodidae) that feed for several days or even weeks with its mouthpart embedded in their vertebrate hosts.

Acknowledgments — We thank M. Rodríguez and S. Streicher for technical assistance and helpful comments, and S. Prieto for his help in the collection of ticks. We gratefully acknowledge the generous gift of bovine and human TAFI from J. J. Enghild and Z. Valnickova.

The nucleotide and protein sequences reported in this work have been submitted to GenBank™ with accession number AY794405.

References

1. Bode W & Huber R. Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem* 1992; 204: 433-451
2. Bode W & Huber R. Structural basis of the endoproteinase-protein inhibitor interaction. *Biochim Biophys Acta* 2000*;* 1477: 241-252
3. Turk B, Turk V & Turk D. Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. *Biol Chem* 1997*;* 378: 141-150
4. Rawlings ND, Tolle DP & Barrett AJ. Evolutionary families of peptidase inhibitors. *Biochem J* 2004; 378: 705-716
5. Hass GM, Nau H, Biemann K, Grahn DT, Ericsson LH & Neurath H. The amino acid sequence of a carboxypeptidase inhibitor from potatoes. *Biochemistry* 1975; 14: 1334-1342
6. Hass GM & Hermodson MA. Amino acid sequence of a carboxypeptidase inhibitor from tomato fruit. *Biochemistry* 1981; 20: 2256-2260
7. Homandberg GA, Litwiller RD & Peanasky RJ. Carboxypeptidase inhibitors from *Ascaris suum*: the primary structure. *Arch Biochem Biophys*1989*;* 270: 153-161
8. Reverter D, Vendrell J, Canals F, Horstmann J, Aviles FX, Fritz H & Sommerhoff CP. A carboxypeptidase inhibitor from the medical leech *Hirudo medicinalis*. Isolation, sequence analysis, cDNA cloning, recombinant expression, and characterization. *J Biol Chem* 1998; 273: 32927-32933

9. Normant E, Martres MP, Schwartz JC & Gros C. Purification, cDNA cloning, functional expression, and characterization of a 26-kDa endogenous mammalian carboxypeptidase inhibitor. *Proc Natl Acad Sci USA* 1995*;* 92: 12225-12229
10. Liu Q, Yu L, Gao J, Fu Q, Zhang J, Zhang P, Chen J & Zhao S. Cloning, tissue expression pattern and genomic organization of latexin, a human homologue of rat carboxypeptidase A inhibitor. *Mol Biol Rep* 2000; 27: 241-246
11. Guasch A, Coll M, Aviles FX & Huber R. Three-dimensional structure of porcine pancreatic procarboxypeptidase A. A comparison of the A and B zymogens and their determinants for inhibition and activation. *J Mol Biol* 1992; 224: 141-157
12. Garcia-Saez I, Reverter D, Vendrell J, Aviles FX & Coll M. The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen. *EMBO J* 1997; 16: 6906-6913
13. Vendrell J, Querol E & Aviles FX. Metallocarboxypeptidases and their protein inhibitors. Structure, function and biomedical properties. *Biochim Biophys Acta* 2000; 1477: 284-298
14. Reznik SE & Fricker LD. Carboxypeptidases from A to Z: implications in embryonic development and Wnt binding. *Cell Mol Life Sci* 2001; 58: 1790-1804
15. Levin Y, Skidgel RA & Erdos EG. Isolation and characterization of the subunits of human plasma carboxypeptidase N (kininase i). *Proc Natl Acad Sci USA.* 1982; 79: 4618-4622
16. Eaton DL, Mallory BE, Tsai SP, Henzel W & Drayna D. Isolation, molecular cloning, and partial characterization of a novel carboxypeptidase B from human plasma. *J Biol Chem* 1991; 266: 21833-21838
17. Bajzar L, Manuel R & Nesheim ME. Purification and characterization of TAFI, a thrombin-activable fibrinolysis inhibitor. *J Biol Chem* 1995; 270: 14477-14484
18. Matthews KW, Mueller-Ortiz SL & Wetsel RA. Carboxypeptidase N: a pleiotropic regulator of inflammation. *Mol Immunol* 2004; 40: 785-793
19. Wang W, Boffa MB, Bajzar L, Walker JB & Nesheim ME. A study of the mechanism of inhibition of fibrinolysis by activated thrombin-activable fibrinolysis inhibitor. *J Biol Chem* 1998; 273: 27176-27181
20. Bouma BN & Meijers JC. Thrombin-activatable fibrinolysis inhibitor (TAFI, plasma procarboxypeptidase B, procarboxypeptidase R, procarboxypeptidase U). *J Thromb Haemost* 2003; 1: 1566-1574
21. Klement P, Liao P & Bajzar L. A novel approach to arterial thrombolysis. *Blood* 1999; 94: 2735-2743
22. Nagashima M, Werner M, Wang M, Zhao L, Light DR, Pagila R, Morser J & Verhallen P. An inhibitor of activated thrombin-activatable fibrinolysis inhibitor potentiates tissue-type plasminogen activator-induced thrombolysis in a rabbit jugular vein thrombolysis model. *Thromb Res* 2000; 98: 333-342

23. Bowman AS, Coons LB, Needham GR & Sauer JR. Tick saliva: recent advances and implications for vector competence. *Med Vet Entomol* 1997; 11: 277-285
24. Basanova AV, Baskova IP & Zavalova LL. Vascular-platelet and plasma hemostasis regulators from bloodsucking animals. *Biochemistry (Mosc)* 2002; 67: 143-150
25. Salzet M. Anticoagulants and inhibitors of platelet aggregation derived from leeches. *FEBS Lett* 2001; 492: 187-192
26. Mans BJ & Neitz AW. Adaptation of ticks to a blood-feeding environment: evolution from a functional perspective. *Insect Biochem Mol Biol* 2004; 34: 1-17
27. Söllner C, Mentele R, Eckerskorn C, Fritz H & Sommerhoff CP. Isolation and characterization of hirustasin, an antistasin-type serine-proteinase inhibitor from the medical leech *Hirudo medicinalis*. *Eur J Biochem* 1994; 219: 937-943
28. Sommerhoff CP, Söllner C, Mentele R, Piechottka GP, Auerswald EA & Fritz H. A Kazal-type inhibitor of human mast cell tryptase: isolation from the medical leech *Hirudo medicinalis*, characterization, and sequence analysis. *Biol Chem Hoppe-Seyler* 1994; 375: 685-694
29. Arolas JL, Lorenzo J, Rovira A, Vendrell J, Aviles FX & Ventura S. Secondary binding site of the potato carboxypeptidase inhibitor. Contribution to its structure, folding, and biological properties. *Biochemistry* 2004; 43: 7973-7982
30. Reverter D, Ventura S, Villegas V, Vendrell J & Aviles FX. Overexpression of human procarboxypeptidase A2 in Pichia pastoris and detailed characterization of its activation pathway. *J Biol Chem* 1998; 273: 3535-3541
31. Ventura S, Villegas V, Sterner J, Larson J, Vendrell J, Hershberger CL & Aviles FX. Mapping the pro-region of carboxypeptidase B by protein engineering. Cloning, overexpression, and mutagenesis of the porcine proenzyme. *J Biol Chem* 1999; 28: 19925-19933
32. Peranen J, Rikkonen M, Hyvonen M & Kaariainen L. T7 vectors with modified T7lac promoter for expression of proteins in *Escherichia coli*. *Anal Biochem* 1996; 236: 371-373
33. Morrison JF. The slow-binding and slow, tight-binding inhibition of enzyme-catalysed reactions. *Trends Biochem Sci* 1982; 7: 102-105
34. Nielsen H, Engelbrecht J, Brunak S & von Heijne G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 1997; 10: 1-6
35. Chen YH, Yang YT & Martinez HH. Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* 1972; 13: 3350-3359
36. Venhudova G, Canals F, Querol E & Aviles FX. Mutations in the N- and C-terminal tails of potato carboxypeptidase inhibitor influence its oxidative refolding process at the reshuffling stage. *J Biol Chem* 2001; 276: 11683-11690
37. Rees DC & Lipscomb WN. Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 A resolution. *J Mol Biol* 1982; 160: 475-498

38. Bajzar L, Nesheim M, Morser J & Tracy PB. Both cellular and soluble forms of thrombomodulin inhibit fibrinolysis by potentiating the activation of thrombin-activable fibrinolysis inhibitor. *J Biol Chem* 1998; 273: 2792-2798
39. Walker JB, Hughes B, James I, Haddock P, Kluft C & Bajzar L. Stabilization versus inhibition of TAFIa by competitive inhibitors in vitro. *J Biol Chem* 2003; 278: 8913-8921
40. Schneider M & Nesheim M. Reversible inhibitors of TAFIa can both promote and inhibit fibrinolysis. *J Thromb Haemost* 2003; 1: 147-154
41. Reverter D, Fernandez-Catalan C, Baumgartner R, Pfander R, Huber R, Bode W, Vendrell J, Holak TA & Aviles FX. *Nat Struct Biol* 2000; 7: 322-328
42. Molina MA, Marino C, Oliva B, Aviles FX & Querol E. C-tail valine is a key residue for stabilization of complex between potato inhibitor and carboxypeptidase A. *J Biol Chem* 1994; 269: 21467-21472
43. Marino-Buslje C, Venhudova G, Molina MA, Oliva B, Jorba X, Canals F, Aviles FX & Querol E. Contribution of C-tail residues of potato carboxypeptidase inhibitor to the binding to carboxypeptidase A A mutagenesis analysis. *Eur J Biochem* 2000; 267: 1502-1509
44. Mao S, Colussi D, Bailey CM, Bosserman M, Burlein C, Gardell SJ & Carroll SS. Electrochemiluminescence assay for basic carboxypeptidases: inhibition of basic carboxypeptidases and activation of thrombin-activatable fibrinolysis inhibitor. *Anal Biochem* 2003; 319: 159-170
45. Aloy P, Companys V, Vendrell J, Aviles FX, Fricker LD, Coll M & Gomis-Ruth FX. The crystal structure of the inhibitor-complexed carboxypeptidase D domain II and the modeling of regulatory carboxypeptidases. *J Biol Chem* 2001; 276: 16177-16184
46. Mutch NJ, Moore NR, Wang E & Booth NA. Thrombus lysis by uPA, scuPA and tPA is regulated by plasma TAFI. *J Thromb Haemost* 2003; 1: 2000-2007
47. Silveira A, Schatteman K, Goossens F, Moor E, Scharpe S, Stromqvist M, Hendriks D & Hamsten A. Plasma procarboxypeptidase U in men with symptomatic coronary artery disease. *Thromb Haemost* 2000; 84: 364-368
48. van Tilburg NH, Rosendaal FR & Bertina RM. Thrombin activatable fibrinolysis inhibitor and the risk for deep vein thrombosis. *Blood* 2000; 95: 2855-2859
49. Eichinger S, Schönauer V, Weltermann A, Minar E, Bialonczyk C, Hirschi M, Schneider B, Quehenberger P & Kyrle PA. Thrombin-activatable fibrinolysis inhibitor and the risk for recurrent venous thromboembolism. *Blood* 2004; 103: 3773-3776
50. Renatus M, Stubbs MT, Huber R, Bringmann P, Donner P, Schleuning WD & Bode W. Catalytic domain structure of vampire bat plasminogen activator: a molecular paradigm for proteolysis without activation cleavage. *Biochemistry* 1997; 36: 13483-13493

51. Campbell W, Okada N & Okada H. Carboxypeptidase R is an inactivator of complement-derived inflammatory peptides and an inhibitor of fibrinolysis. *Immunol Rev* 2001; 180: 162-167
52. Campbell WD, Lazoura E, Okada N & Okada H. Inactivation of C3a and C5a octapeptides by carboxypeptidase R and carboxypeptidase N. *Microbiol Immunol* 2002; 46: 131-134
53. Myles T, Nishimura T, Yun TH, Nagashima M, Morser J, Patterson AJ, Pearl RG & Leung LLK. Thrombin activatable fibrinolysis inhibitor, a potential regulator of vascular inflammation. *J Biol Chem* 2003; 278: 51059-51067
54. Springman EB, Dikov MM & Serafin WE. Mast cell procarboxypeptidase A. Molecular modeling and biochemical characterization of its processing within secretory granules. J Biol Chem 1995; 270: 1300-1307

Supplementary information

The photographs show *Rhipicephalus bursa* ticks, from which TCI has been isolated. These haemoparasites are commonly found in ruminants in Europe, Northern Africa, and some places of Asia (Paleartic Region).

Section I. Work 2

The Three-Dimensional Structures of Tick Carboxypeptidase Inhibitor in Complex with A/B Carboxypeptidases Reveal a Novel Double-Headed Binding Mode

Summary

The tick carboxypeptidase inhibitor (TCI) is a proteinaceous inhibitor of metallocarboxypeptidases present in the blood-sucking tick *Rhipicephalus bursa*. The three-dimensional crystal structures of recombinant TCI bound to bovine carboxypeptidase A and to human carboxypeptidase B have been determined and refined at 1.7 and 2.0 Å resolution, respectively. TCI consists of two domains that are structurally similar despite the low degree of sequence homology. The domains, each consisting of a short α-helix followed by a small twisted antiparallel β-sheet, show high structural homology to proteins of the β-defensin-fold family. TCI anchors to the surface of mammalian carboxypeptidases in a double-headed manner not previously seen for carboxypeptidase inhibitors: the last three carboxy-terminal amino acids interact with the active site of the enzyme in a way that mimics substrate binding, and the N-terminal domain binds to an exosite distinct from the active site groove. The structures of these complexes should prove valuable in the applications of TCI as a thrombolytic drug and as a basis for the design of novel bivalent carboxypeptidase inhibitors.

Keywords: metallocarboxypeptidase; carboxypeptidase inhibitor; inhibitor-enzyme complex; crystal structure; pro-fibrinolytic drug.

Introduction

Metallocarboxypeptidases (CPs) comprise a family of zinc-containing exopeptidases that catalyze the cleavage of C-terminal peptide bonds in proteins and peptides. These enzymes can be subdivided structurally into the pancreatic/digestive or CPA/B subfamily and the regulatory or CPN/E subfamily (1). The members of the former subfamily are synthesized as inactive zymogens called procarboxypeptidases (PCPs) and are subsequently activated by limited proteolysis (2). Traditionally, these carboxypeptidases have been classified into A and B forms based on their substrate specificity: CPA has a

preference for aliphatic and aromatic C-terminal residues, and CPB for basic C-terminal residues (3). The members of the CPN/E subfamily, commonly referred to as "regulatory" carboxypeptidases, do not appear to be produced as inactive precursors, and are located in subcellular compartments to prevent inappropriate cleavages that would otherwise damage the cell (4, 5).

Normally, the biological actions of proteases are controlled by specific interactions with proteinaceous inhibitors (6). So far, however, only a few protein inhibitors have been identified for metallocarboxypeptidases (1). Among them, the potato and leech carboxypeptidase inhibitors (PCI and LCI) have been structurally studied in detail by NMR (7-9) and by X-ray crystallography in their complexes with bovine CPA and human CPA2 (8, 10). Both proteins inhibit carboxypeptidases via a substrate-like interaction of their C-terminal tail with the active site groove of the enzyme (11, 12). Numerous protease inhibitors have been described in hematophagous organisms, most notably those directed toward their host's blood coagulation factors such as thrombin and factor Xa (13, 14). The structures of several of these inhibitors in complex with different target proteases have been solved by X-ray crystallography. These include hirudin, ornithodorin, triabin, and tick anticoagulant peptide, among others (15-18). A common feature of these non-homologous inhibitors is their small size and their stabilization by several disulfide bridges.

Tick carboxypeptidase inhibitor (TCI) is a novel metallocarboxypeptidase inhibitor recently identified by our group in the ixodid tick *Rhipicephalus bursa* (19). TCI is a 75-amino acid protein stabilized by 6 disulfide bridges that inhibits various members of the A/B carboxypeptidase subfamily with nanomolar affinity, including plasma CPB (also known as thrombin-activatable fibrinolysis inhibitor, TAFI). Its amino acid sequence shows no homology with other known proteins except at its C-terminus, which resembles that of PCI and LCI, suggesting a similar mechanism of inhibition. Most likely, ticks secrete TCI during feeding and digestion to maintain the liquid state of blood because TCI strongly inhibits TAFI, a well-known attenuator of fibrinolysis (20, 21). Indeed, we have recently shown that recombinant TCI stimulates fibrinolysis *in vitro* and thus may have potential for applications to prevent or treat thrombotic disorders (19). In addition, TCI could also modulate host defense mechanisms during the parasitic infection by affecting mast-cell CPA and/or TAFI, both of which are thought to be involved in inflammatory processes (22, 23).

Here we report the X-ray structures of TCI in complex with bovine CPA (bCPA) and human CPB (hCPB) at 1.7 and 2.0 Å resolutions, respectively. The TCI-hCPB structure is

the first crystallographic study of a complex between a protein inhibitor and a carboxypeptidase type B. The protease-inhibitor interactions in both complexes are analyzed in detail and compared to those of the PCI-bCPA and LCI-hCPA2 complexes. Surprisingly, TCI shows a novel binding mode to carboxypeptidases, thereby providing information about the evolution of protein metallocarboxypeptidase inhibitors. Because a experimental structure of TAFI has not yet been obtained and given the high degree of sequence identity between this enzyme and hCPB (24), the TCI-hCPB structure may help to understand the properties of TAFI and may foster the design of drugs that modulate its biological activity.

Experimental Procedures

Protein Expression and Purification — Expression and purification of recombinant TCI were carried out as described previously (19). Briefly, TCI was produced in the *Escherichia coli* strain BL21(DE3) using the pBAT-4-OmpA plasmid for periplasmic expression, and was purified from the culture medium using a Sep-Pak C_{18} cartridge followed by cation-exchange chromatography and reversed-phase high performance liquid chromatography on a Vydac C_4 column. Recombinant human pancreatic PCPB and the derived active CPB were produced as published elsewhere in detail (25, 26). Bovine pancreatic CPA was purchased from Sigma and used without further purification. The identity and purity of the proteins were verified by mass spectrometry, SDS-PAGE, and N-terminal sequence analysis.

Complex Formation — The TCI-bCPA and TCI-hCPB complexes were prepared by incubating equimolar quantities of inhibitor and enzyme in 10 mM tris-HCl (pH 7.5), 50 mM NaCl for 2 hours at 20ºC. The complete inhibition of the carboxypeptidases was monitored by continuous photometric assays with specific chromogenic substrates (19). The complexes were purified by gel-filtration chromatography (Superdex 75 HiLoad 26/60), and finally concentrated to about 10-12 mg/ml.

Crystallization and Diffraction Data Collection — For the TCI-bCPA complex, crystals suitable for data collection were obtained after a few days at 20ºC using the hanging-drop vapor-diffusion method (see photographs of the crystals for both complexes in "Supplementary Information"). Drops were prepared by mixing equal volumes of protein solution and reservoir buffer containing 0.2 M sodium cacodylate (pH 6.5), 0.2 M zinc acetate dihydrate, 7% (w/v) PEG 8000, and 10% (w/v) dried dioxan. For the TCI-

hCPB complex, single crystals of suitable size were grown within a few weeks at 20ºC using the sitting-drop vapor-diffusion method. Drops were made by mixing equal volumes of one of the assayed reservoir solutions (0.1 M bis-tris (pH 5.5), 0.2 M lithium sulfate monohydrate, 25% (w/v) PEG 3350) and the protein solution. Both TCI-bCPA and TCI-hCPB crystals contained one complex per asymmetric unit. 1.7 Å (TCI-bCPA) and 2.0 Å (TCI-hCPB) native data sets were obtained from plunge-frozen crystals at 100 K. Crystals were measured using an in-house rotating anode X-ray source. Rotation images were collected on Mar345 image plate. Data were indexed, integrated, and scaled using XDS software (27). The data statistics and crystals' space groups are summarized in Table 1.

Structure Determination and Refinement — Both structures were determined by molecular replacement. The known crystallographic structures of bovine CPA (pdb code 1M4L) and human PCPB (pdb code 1KWM) were used as probes (24, 28). The N-terminal pro-segment of hPCPB (95 residues long) was removed from the search model. Rotation and translation searches were performed using Molrep (29). For the TCI-bCPA complex, rotation search in Patterson space yielded a peak of height 21.2 σ over the highest noise peak of 4.8 σ. Translation search gave peaks of 20.9 σ over the noise height of 4.9 σ. Corresponding values for CPB were 12.33 σ over 6.36 σ, and 51,75 σ over 28,04 σ. The R-factors of the initial models without the inhibitor were 43% for bCPA and 46% for hCPB. At this stage, free atom model improvement was introduced using Arp/wArp (30) for both structures. As the next step, iterative model building was performed for CPA using Arp/wArp (31). The model thus obtained (360 residues) was completed and revised manually using Xfit software (32). Similarly, the structure of CPB was built using Arp/wArp, but the initial phases calculated from the improved model were corrected by density modification in the DM program. The resulting model of 356 residues was also completed using Xfit. For both structures, Arp/wArp was used to add solvent atoms (33). Structures were finally refined with Refmac5 (29). Refinement statistics are shown in Table 1. Final electron density maps were of very high quality. The TCI-bCPA model was built from residues Ser3 to Leu305, and has no interpretable density for residues Ser134-Ser135 and side chains of residues Lys177, Lys190, Lys216, Gln221, Lys239, Arg276, Gln13i, and Gln58i (residues of TCI are suffixed with an i). The TCI-hCPB model has all residues present from Thr5 to Tyr309 but the following amino acids have no interpretable electron density in the side chain area: Arg92, Glu93, Lys122, Lys237, Arg 276, Lys298, Lys6i, Gln13i, Gln18i, and Glu19i. These side chains were removed from the models. The

Ramachandran plot calculated for CPA model shows only one residue (Ser199) in a "disallowed" region of the ϕ, ψ plane. It agrees with the previously reported structures of native CPA. The TCI-bCPA contains four additional surface-bound zinc atoms probably as an artifact caused by the use of zinc in the crystallization buffer.

Table 1. Data collection and refinement statistics

	TCI-CPA	TCI-CPB
A. *Data collection*		
Space group	$P2_1$	$P4_32_12$
Cell constants (Å)	a=47.07 c=57.89 b=68.60 β=104.72	a=b=74.20 b=163.55
Resolution range (Å)	20-1.7	20-2.0
Wavelength (Å)	1.542	1.542
Observed reflections	215232	339432
Unique reflections	36248	29538
Whole range		
Completeness (%)	98.9	99.9
R_{merge}	6.7	4.6
$I/\sigma(I)$	12.9	28.5
Last shell		
Resolution range (Å)	1.7-1.8	2.0-2.1
Completeness (%)	72.0	49.3
R_{merge}	16.6	18.4
$I/\sigma(I)$	5.7	8.9
B. *Refinement*		
No. of reflections	39235	27853
Resolution (Å)	20-1.7	20-2.0
R-factor (%)	15.4	16.1
R_{free} (%)	18.5	21.5
Average B (Å^2)	15.2	19.38
r.m.s.d. bond length (Å)	0.007	0.010
r.m.s.d. angles (°)	1.112	1.273
C. *Content of asymmetric unit*		
No. of protein complexes	1	1
No. of protein residues/atoms	375/2908	380/2963
No. of solvent atoms	379	505
Other atoms or molecules (Zn)	5	1

Results and Discussion

Crystal Structure of TCI — The structures of TCI are similar in its complexes with bCPA and hCPB, with a root mean square (r.m.s.) deviation of 0.51 Å for the backbone chain atoms and 0.87 Å for both backbone and side chains. The structure consists of two clearly separated regions named here as the "N-terminal" (residues Asn1-Lys36) and "C-terminal" (residues Gly39-Leu74) domains that are linked by residues Leu37 and Thr38 (Fig. 1). Although both domains show only a low degree of sequence identity (27%; see alignment in Fig. 1S of "Supplementary Information") they are structurally very similar, displaying a r.m.s. deviation of 1.26 Å for main chain atoms (Fig. 1), and thus could have arisen by gene duplication. Similarly, several thrombin inhibitors from blood-sucking organisms such as rhodniin, ornithodorin and savignin are double-headed inhibitors (16, 34, 35).

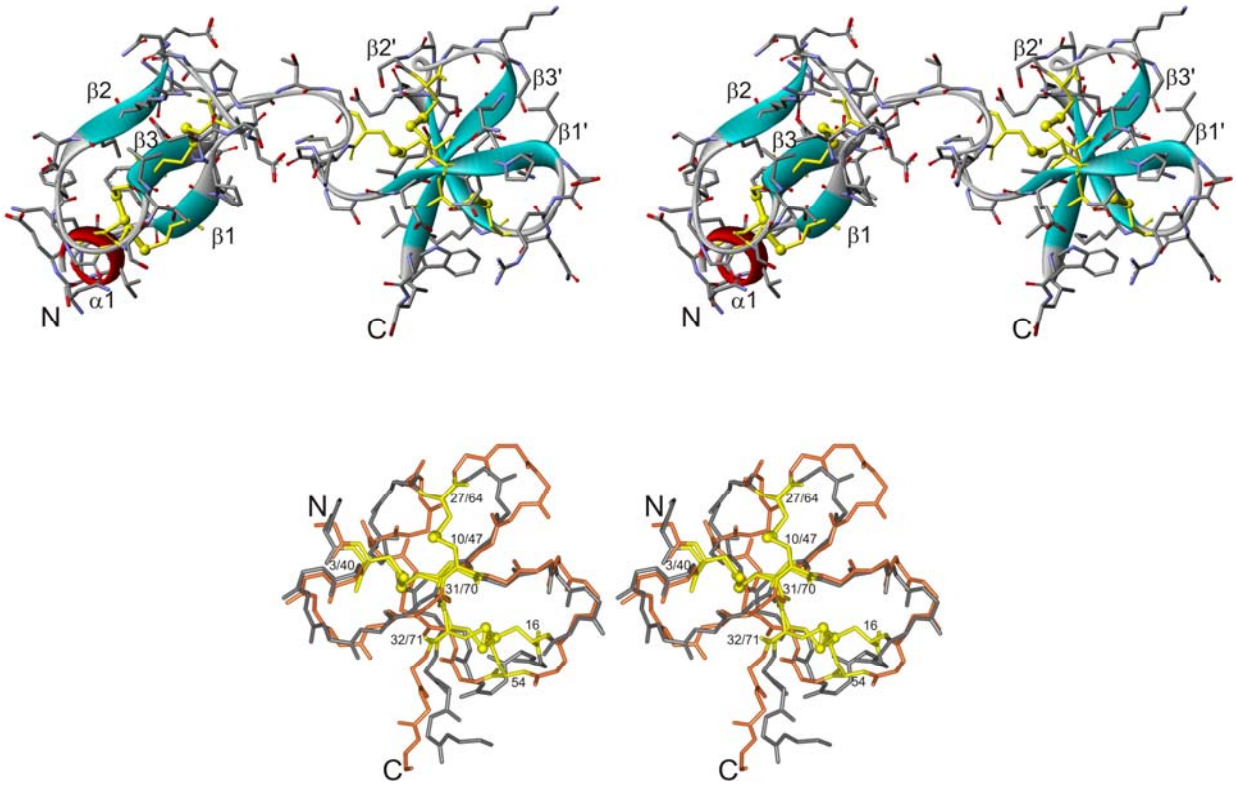


Fig. 1. Three-dimensional structure of TCI. *Upper panel.* Stereo view of TCI shown in a ribbon representation. The helix (α1) and β-strands (β1-β3 and β1'-β3') are colored red and light blue, respectively. The six disulfide bridges are shown in yellow. N and C indicate the N- and C-terminus of TCI. *Lower panel.* Superposition of the backbone chains of the N-terminal (gray) and C-terminal (orange) domain of TCI shown in a stereo view. The disulfide pairing of both domains is shown in yellow.

The N-terminal domain of TCI consists of a short α-helix located between residues Glu2 and Ser5 and an antiparallel triple-stranded β–sheet involving residues Gly9-Leu11 (β1), Ala20-Leu22 (β2) and Val30-Asp33 (β3) with α-β1-β2-β3 topology (Fig. 1). In the C-terminal domain, residues Glu46-Leu50 (β1'), Lys55-Leu57 (β2') and Gly 67-Val72 (β3') also form a triple-stranded antiparallel β–sheet, which is connected to the N-terminal domain by a loop that is structurally similar to a small α-helix (residues Gly39-Gly44). The disulfide pairing of TCI was determined during the three-dimensional structure calculations. Disulfide bridges are formed between Cys3-Cys31, Cys10-Cys27, Cys16-Cys32, Cys40-Cys70, Cys47-Cys64, and Cys54-Cys71 (Fig. 1). The large number of disulfides probably contributes to the high stability of the protein against temperature and denaturing agents observed in previous biochemical studies (19). Also, these S-S bridges are probably responsible for the high percentage of residues in loops and for the absence of a defined hydrophobic core. However, approximately 38% of the TCI residues belong to regular secondary structure elements, a high content when compared to other proteins of similar size. A computer search employing DALI server (36) showed that the structures of both domains and their disulfide pattern are similar to those of the β-defensin-fold family (37). The fold of these proteins generally consists of a short helix or turn followed by a small twisted antiparallel β-sheet stabilized by six cysteine residues that are paired in a 1-5, 2-4, 3-6 fashion and maintain a compact structure. It is remarkable that the superposition of each TCI domain with several members of this family (e.g. the venom toxin crotamine and human β-defensin-2) displays a r.m.s. deviation of only 1.3-1.5 Å for backbone atoms (Fig. 2).

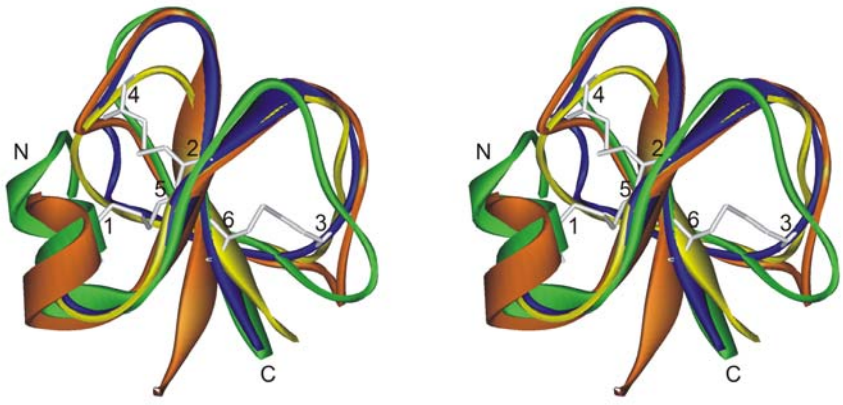


Fig. 2. Comparison of the TCI domains with members of the β-defensin-fold family. The domains of TCI (N-terminal in yellow and C-terminal in dark blue), crotamine (pdb code 1H5O; in dark orange) and β-defensin-2 (pdb code 1FD3, chain A; in green) are superimposed and are shown as

ribbons. The N- and C-terminal ends are labeled. Cysteine residues, which are paired in a 1-5, 2-4, 3-6 fashion in all molecules, are indicated in light gray.

For both potato (PCI) and leech (LCI) carboxypeptidase inhibitors, the overall structure in solution is very similar to that bound to carboxypeptidases (8, 9). NMR studies indicated however that the N- and C-terminal tails of both inhibitors are highly flexible but become more rigid upon binding to a carboxypeptidase (8). Unlike PCI and LCI, the N- and C-terminal ends of TCI appear more rigid due to their involvement and proximity to secondary structure elements and disulfides. The N-terminus is located within an α-helix, while the C-terminal tail protrudes from the β3-strand of the second domain that is greatly stabilized by both secondary structure interactions and disulfide bridges (Fig. 1). It is worth mentioning that some flexibility is to be expected between the two domains of TCI because their relative position is slightly different when bound to bCPA and hCPB and no inter-domain interactions are observed in the TCI structure. Thus, the domains probably can adopt different orientations when binding to different carboxypeptidases. In addition, unlike in the cases of PCI and LCI that are compact one-domain proteins, conformational changes could also take place upon binding to the enzyme. A better understanding of these changes would require the determination of the structure of TCI in its free state; unfortunately, attempts to crystallize it alone have been unsuccessful so far.

Crystal Structures of Bovine CPA and Human CPB — The structure of bovine CPA in its complex with TCI is almost identical to that of the unliganded enzyme described previously (r.m.s. deviation 0.43 Å for the backbone atoms) (28). Bovine CPA, one of the most thoroughly studied carboxypeptidases, consists of a central mixed eight-stranded twisted β-sheet that is surrounded by eight α-helices forming a globular α/β motif (Fig. 3). Similar to porcine CPA and human CPA1 (3, 38) bovine CPA contains a single disulfide bond between Cys138 and Cys161. In contrast to the unbounded form, in the complex the catalytic water of the active site Zn^{2+} is substituted by the C-terminus of TCI. The largest differences between the free and TCI-bound structures of CPA are seen in residues 246 to 249 (shifted ~ 1.1 Å), a region directly involved in binding TCI. Among these residues, the movement of the Tyr248 side chain from the native "up" conformation (in isolated CPA) to the "down" position (in the complex) is noteworthy. This conformational change has also been reported in the complexes of other carboxypeptidases with protein inhibitors (8). Additional differences are found in the side chains of residues Arg127, Arg145, Thr164 and Glu270, residues important for substrate binding and catalysis (1).

The structure of human CPB in its complex with TCI is very similar to that of the pro-enzyme form (24), with a r.m.s.d. between active enzymes of 0.48 Å for main chain atoms. The structure of the catalytic domain reported here shows the characteristic α/β hydrolase topology of carboxypeptidases described above for bCPA (see Fig. 2S of "Supplementary Information"). However, hCPB contains two additional disulfide bridges (Cys66-Cys79 and Cys152-Cys166). As seen in the TCI-bCPA complex, the catalytic water of the active site is substituted by the C-terminus of TCI. The largest differences between the native proform and the TCI-bound active hCPB are again found in the region involved in binding of TCI, i.e residues 245 to 250 that are shifted by ~ 1.45 Å. In PCPB the Tyr248 presents a side chain flip and is in the "down" conformation when compared to the porcine counterpart or bCPA. Upon binding to TCI, the Tyr248 side chain is in a "down" conformation, with a localization even deeper inside the carboxypeptidase. The side chains of residues Arg127, Arg145, Cys152 and Tyr165 are also hidden due to the binding of TCI. The side chain of Glu163 is moved to allow the "new" location of Arg127, while the Ile247 and Glu270 side chains are in a different conformer when compared to that in the free enzyme.

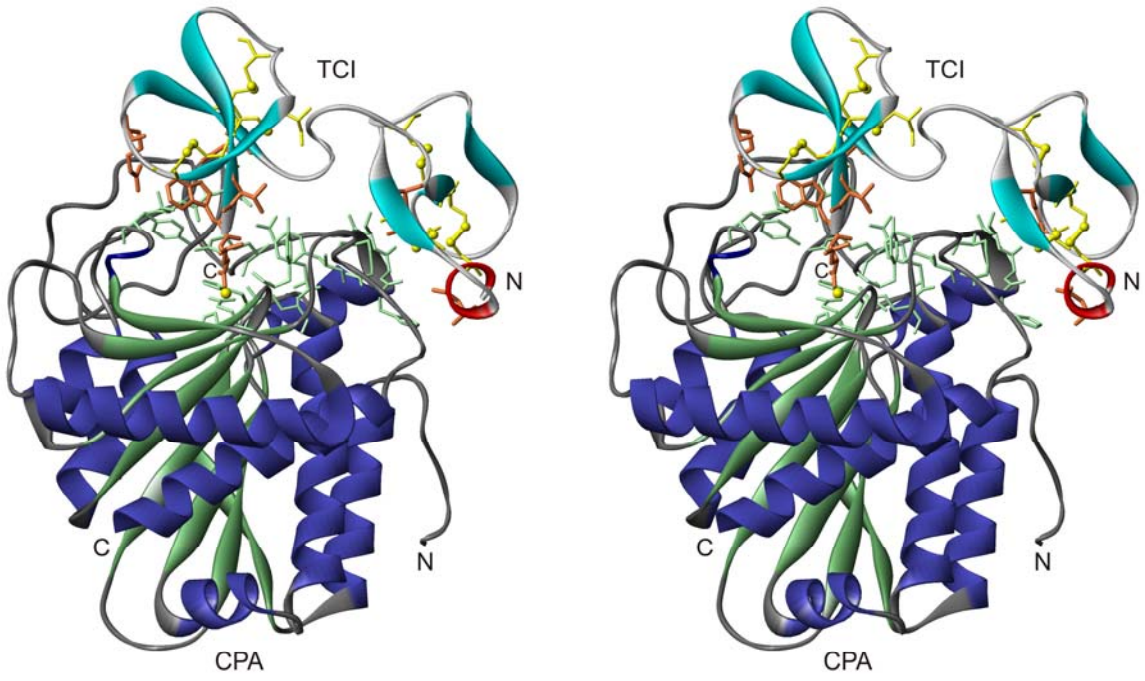


Fig. 3. Structure of the TCI-CPA complex shown in a stereo ribbon representation. The helix and β-strands of TCI are shown in red and light blue, respectively, and the disulfide bridges are indicated in yellow. The helices and β-strands of CPA are shown in dark blue and dark green, respectively, and the catalytic zinc atom is represented by a yellow sphere. The side chains of the

residues involved in the interaction are colored in orange (TCI residues) and light green (CPA residues). The N- and C-termini of TCI and CPA are labeled.

Binding Interactions between TCI and the Carboxypeptidases — TCI interacts extensively with both carboxypeptidases, with a contact area of 1229 and 1476 $Å^2$ in its complex with bCPA and hCPB, respectively. The interaction of the C-terminus of TCI with the active site groove of the carboxypeptidase constitutes the "primary binding region" (Fig. 3). The binding subsites S1, S2 and S3, which are conserved between bCPA and hCPB, can be located based on their interaction with the C-terminal residues Leu74, Trp73, and Val72 of the inhibitor, respectively (Fig. 4). After leaving the active site groove, the polypeptide chain of TCI first turns away from CPA/CPB but later approaches the enzymes' surface again to form a separate contact region (see Fig. 3 and below).

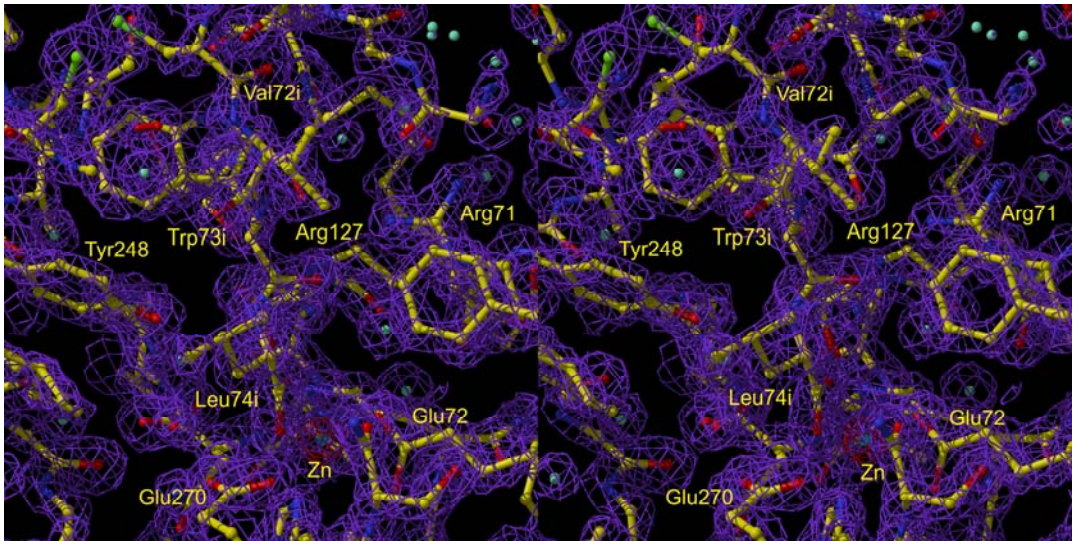


Fig. 4. Stereo plot of the final ($2F_{obs}$-F_{calc}) electron density map (dark violet) of the C-terminal tail of TCI bound within the active site groove of CPA. Residues of TCI are suffixed with an i. The zinc atom is represented by a red sphere.

Unlike those of PCI and LCI, the C-terminal residue of TCI (His75i; in this section residues of TCI are suffixed with an i), which is cleaved off by the enzyme, is not trapped in the S1' subsite of CPA/CPB (formed by Asn144, Arg145 and Tyr248). The new C-terminal Leu74i residue coordinates with the active site Zn through one carboxylate C-terminal oxygen atom (2.1 Å) in the TCI-CPA complex and two carboxylate C-terminal oxygen atoms (2.51 Å for O, and 2.23 Å for OT) in TCI-CPB. Thus, in these complexes the coordination shell of Zn consists of five and six protein atoms, respectively, i.e. four

from CPA or CPB (Oε1 and Oε2 of Glu72, and the Nδ1 atoms of His69 and His196), and one or two from TCI (O and OT of Leu74i). In comparison, in the LCI-CPA2 and PCI-CPA complexes two and one oxygen atoms from the inhibitor are coordinated to zinc, respectively. In the TCI-CPA/CPB complexes, the catalytic water, which in the free enzyme attacks the C-terminal peptide bond during the enzymatic reaction, is displaced by TCI. Water molecules are bound to the guanidinium group of Arg145, and to Asn144, Tyr248 and Glu270.

The S1 subsite of CPA/CPB (Arg127 and Glu270) is shielded by Leu74i, which interacts with the enzyme through several hydrogen bonds, i.e. two with the side chain of Glu270 (2.83/2.86 Å and 2.96/2.98 Å for TCI-CPA/CPB) and one with the guanidinium group of Arg127 (2.83/3.0 Å) (Fig. 4). These important interactions are also present in the LCI-CPA2 complex. Leu74i forms an additional hydrogen bond with the side chain of Glu72 (3.20/3.19 Å), which formally does not belong to the S1 subsite but is conserved in both carboxypeptidases. The main chain of Leu74i is also hydrogen bonded with the side chain of Tyr248 (2.92/2.79 Å).

The major interaction in the S2 subsite (formed by Arg71, Ser197, Tyr198 and Ser 199) is the hydrogen bond between the carbonyl oxygen of Trp73i and the guanidinium group of Arg71 (3.04/3.07 Å) (Fig. 4). The side chain of Trp73i is also in contact with Thr164 (conserved in both carboxypeptidases) and forms an internal hydrogen bond with the side chain of Glu46i. Equivalent internal hydrogen bonds are found in the PCI/LCI-CPA complexes (8). Another remarkable intramolecular hydrogen bond is established between Val72i and Lys55i. There are no hydrogen bonds between Val72i and the S3 subsite (Phe279), although the Val72i residue is in contact with the phenyl ring of Tyr198 and Phe279. In both complexes several polar residues of TCI (Lys41i, Glu46i, Arg52i, Glu53i and Lys55i) surround the TCI C-terminus, thereby covering the remainder of the CPA/CPB active site.

The only C-terminal residue of TCI that is likely to be charged is the carboxylate group of Leu74i (after cleavage). The active site of the carboxypeptidase contains several charged groups that are buried in both complexes, i.e. the guanidinium groups of Arg71, Arg124, Arg127 and Arg145 and the carboxylate group of Glu270. These charged residues could be important for the proper approach of substrates to the active site, for their proper positioning in it, and for their cleavage (2).

Similar to PCI and LCI, residues from TCI establish “secondary contacts” with regions close to the carboxypeptidase active site groove (Fig. 5). Thus, e.g. side chains of

residues Glu46i and Arg52i of the C-terminal domain form hydrogen bonds with Glu163 and Thr164 of CPA, respectively. However, unlike other inhibitors, TCI binds to carboxypeptidases in a double-headed fashion: The N-terminal domain of TCI establishes interactions with the enzyme surface away from those made by its C-terminal domain or those made by PCI and LCI (Fig. 5 and 6). These additional "secondary contacts" of the N-terminal domain define an "exosite" on the enzyme surface not identified previously that is located around the S4 subsite of the carboxypeptidase (formed by Gln122/Lys122, Arg124 and Lys128). The interaction of TCI with this exosite comprises several hydrogen bonds between Asn1i and His120/Trp120 (3.26/3.06 Å for TCI-CPA/CPB), Cys10i and Arg124 (2.8/2.81 Å), Ser28i and Ser121/Thr121 (3.06/2.95 Å), and Thr29i and Gln122/Lys122 (2.71/2.58 Å).

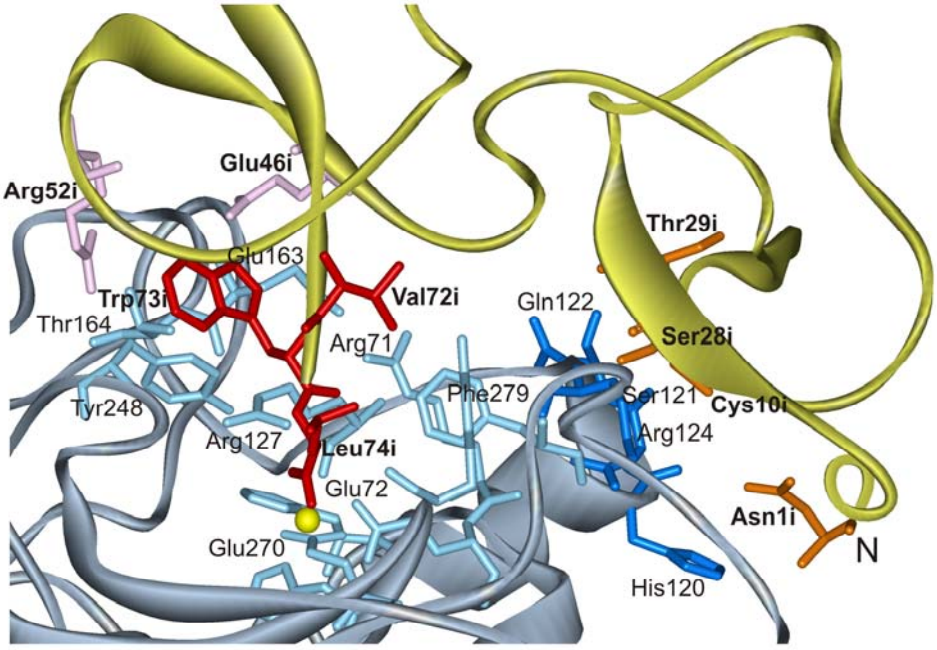


Fig. 5. Close-up view of the double-headed binding of TCI to CPA. TCI (yellow) and CPA (gray-blue) are shown in a ribbon representation. The N-terminus of TCI is labeled and the zinc atom of the CPA active site is represented by a yellow sphere. The residues involved in binding are labeled (TCI residues with a suffixed i), and their side chains are shown as stick models. Residues constituting the "primary binding site", i.e. those of the TCI C-tail and the CPA active site, are colored red and light blue, respectively. Residues forming "secondary contacts" are colored in light purple (TCI C-terminal domain) and light blue (CPA), and those that define the "exosite" on the CPA surface in orange (N-terminal domain of TCI) and dark blue (CPA).

Mechanism of Inhibition of Carboxypeptidases by TCI —Many of the features of the inhibition of carboxypeptidases by TCI characterized previously in enzymatic studies of TCI can be explained based on the three-dimensional structures reported here and comparisons with related inhibitors. The product generated after complex formation, TCI-desHis^{75}, has the same inhibitory activity as native TCI (19), indicating that both forms bind similarly to the enzyme. After the initial trimming, TCI-desHis^{75} remains intact. This is not surprising given the excellent fit of the surfaces of TCI and the carboxypeptidase that hinders any further penetration (see surface representation of the complex in Fig. 3S of "Supplementary Information"). The highly efficient inhibition is probably due to the abundant interactions established between the inhibitor and the enzyme, which affect or shield many residues of the carboxypeptidase that are essential for substrate binding and catalysis. The coordination of the enzyme's catalytic Zn with the carboxylate group of Leu74 also contributes to the stabilization of the complex. Taken together, numerous interactions account for the formation of a tight and stable complex.

The interaction interfaces in both complexes are nearly indistinguishable, demonstrating that TCI binds and inhibits forms A and B of carboxypeptidases by the same mechanisms. However, the position of the two TCI domains is slightly different in the two complexes suggesting that the inhibitor might adjust its structure, i.e. the position of its two domains relative to each other, for proper interaction with different carboxypeptidases. Curiously, such a behavior is reminiscent of a strategy recently introduced in drug design, called "tethering" (39), in which synthetic inhibitors are optimized by combining a small fragment that binds to the enzyme's active site with an adaptable exosite ligand.

Overall, the number of enzyme-inhibitor contacts seen in the complexes solved in this work is larger than those seen in the PCI-bCPA and LCI-hCPA2 complexes (Fig. 6). However, many similarities in binding mode are observed among all of them. The C-terminal tail also constitutes the primary binding site of PCI and LCI, docking into the active site of the enzyme and leading to a stopper-like inhibition (2, 12). The C-terminal residue of PCI and LCI (Gly39 and Glu66) also is cleaved off by the carboxypeptidase, but unlike in TCI-CP complexes, this last residue remains bound in the S1' subsite of the enzyme's active site (8, 10). The number of interactions between this cleaved-off residue and the active site of the carboxypeptidase might determine its presence (PCI, LCI) or absence (TCI) after the crystallization process. In addition, in TCI-CPA/CPB Trp73 may

also prevent its presence in the active site due to the larger size of the Trp side chain compared to the Tyr37 in PCI and Tyr65 in LCI.

Significantly, the overall folds of TCI, PCI and LCI in the complexes are completely different except at their C-tail, where a similar sequence accounts for an almost identical backbone fold (Fig. 6). This is the only motif conserved among the carboxypeptidase inhibitors isolated from evolutionarily highly distant organisms, i.e. blood-sucking animals and *Solanacea* plants, representing a good example of convergent evolution dictated by the three-dimensional structure of the target enzymes.

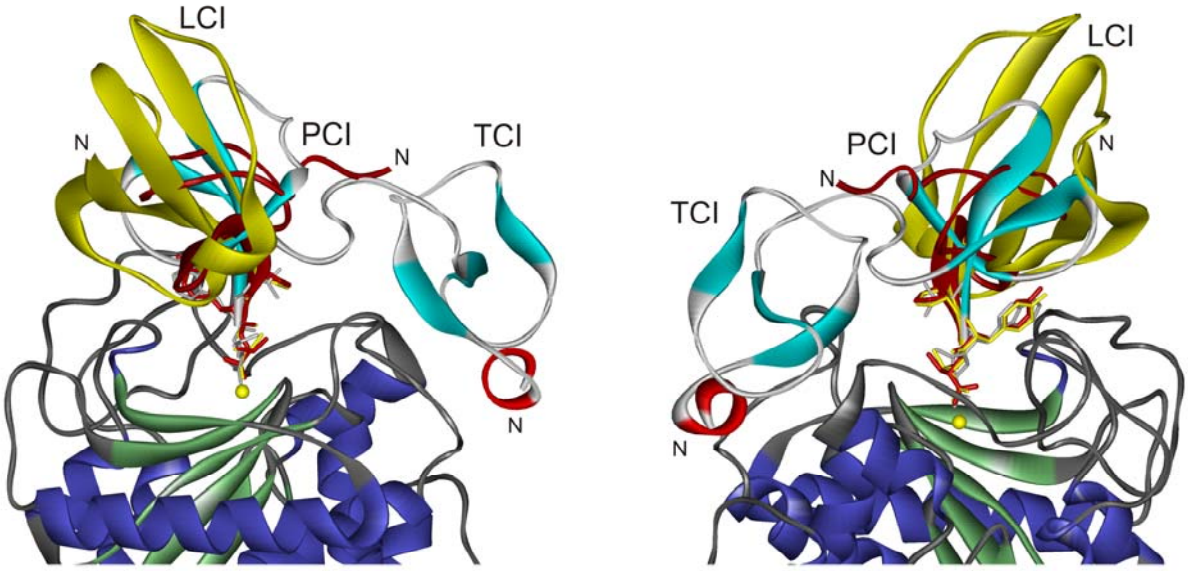


Fig. 6. Comparison of the binding of TCI, PCI and LCI to CPA. The CPA moieties of the crystal structures were superimposed; only the CPA moiety of the TCI-CPA complex is shown with helices and β-strands colored in dark blue and dark green, respectively. The catalytic zinc atom is represented by a yellow sphere. The inhibitors TCI (gray, light blue and red), PCI (dark red) and LCI (yellow) are represented by ribbons; their binding is shown in two different orientations (rotated by ~180º). For all inhibitors the side chains of the C-terminal residues are shown as stick model and the N-terminus is labeled.

Biomedical Implications — Besides their classical role as digestive enzymes, metallocarboxypeptidases also participate in selective regulatory processes like blood coagulation/fibrinolysis, inflammation, local anaphylaxis and prohormone/neuropeptide processing (5, 12). Specific information about the mechanisms of control of these enzymes as well as their complexed structures with inhibitors are essential for a rational drug design for biotechnological and biomedical applications.

TAFI has attracted considerable medical interest in recent years because it constitutes a risk factor for thrombosis and coronary artery disease (40, 41) and it is probably involved in inflammatory processes (23, 42). Since attempts to crystallize TAFI have been unsuccessful so far, the present three-dimensional structure of human CPB, the carboxypeptidase most closely related to TAFI (48% sequence identity; see alignment in Fig. 4S of "Supplementary Information"), may help to better understand its biological action. Activated TAFI down-regulates blood fibrinolysis by removal of carboxy-terminal lysine and arginine residues from partially degraded fibrin (21). Consequently, its inhibition by PCI and TCI results in an enhanced tissue plasminogen activator-induced clot lysis (19, 43, 44). TCI shows ~5-fold higher affinity toward TAFI and accelerates fibrinolysis ~10 times more effectively than PCI (19). These data, together with the structural information derived from this work, will be useful to modulate the activity of TAFI by using more specific and potent drugs, e.g. bivalent inhibitors.

Finally, it is worth mentioning that the domains of TCI are structurally related to the β-defensin-fold family (37). This fold is found in a wide range of small disulfide-rich proteins from e.g. venoms of snakes (45) and sea anemones. Proteins presenting this fold have numerous pharmacological activities including ion-channel inhibition and analgesic and myonecrotic actions. Several anti-microbial defensins from vertebrates also belong to this family (46). Because *Rhipicephalus* ticks feed for several days or even weeks with their mouthpart embedded in their vertebrate hosts, it appears possible that TCI, primarily a pro-fibrinolytic compound, could also participate in the modulation of inflammation and host defenses.

Acknowledgments — We thank Prof. W. Bode for helpful discussions and critical revision of this work. We are grateful to Drs. S. Segura and J. Vendrell for kindly providing the clone of PCPB and the TAFI model and for overall help.

The coordinates of the structures reported in this work have been deposited with the Protein Data Bank (accession code YXXX for the TCI-bCPA complex and YXXX for the TCI-hCPB complex).

References

1. Vendrell J, Querol E & Aviles FX. Metallocarboxypeptidases and their protein inhibitors. Structure, function and biomedical properties. *Biochim Biophys Acta* 2000; 1477: 284-298
2. Aviles FX, Vendrell J, Guasch A, Coll M & Huber R. Advances in metalloprocarboxypeptidases. Emerging details on the inhibition mechanism and on the activation process. *Eur J Biochem* 1993; 211: 381-389
3. Guasch A, Coll M, Aviles FX & Huber R. Three-dimensional structure of porcine pancreatic procarboxypeptidase A. A comparison of the A and B zymogens and their determinants for inhibition and activation. *J Mol Biol* 1992; 224: 141-157
4. Gomis-Rüth FX, Companys V, Qian Y, Fricker LD, Vendrell J, Aviles FX & Coll M. Crystal structure of avian carboxypeptidase D domain II: a prototype for the regulatory metallocarboxypeptidase subfamily. *EMBO J* 1999; 18: 5817-5826
5. Reznik SE & Fricker LD. Carboxypeptidases from A to Z: implications in embryonic development and Wnt binding. *Cell Mol Life Sci* 2001; 58: 1790-1804
6. Bode W & Huber R. Structural basis of the endoproteinase-protein inhibitor interaction. *Biochim Biophys Acta* 2000; 1477: 241-252
7. Clore GM, Gronenborn AM, Nilges M & Ryan CA. Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. *Biochemistry* 1987; 26: 8012-8023
8. Reverter D, Fernandez-Catalan C, Baumgartner R, Pfander R, Huber R, Bode W, Vendrell J, Holak TA & Aviles FX. Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2. *Nat Struct Biol* 2000; 7: 322-328
9. Gonzalez C, Neira JL, Ventura S, Bronsoms S, Rico M & Aviles, FX. Structure and dynamics of the potato carboxypeptidase inhibitor by 1H and 15N NMR. *Proteins* 2003; 50: 410-422
10. Rees DC & Lipscomb WN. Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. *J Mol Biol* 1982; 160: 475-498
11. Bode W & Huber R. Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem* 1992; 204: 433-451
12. Vendrell J, Aviles FX & Fricker LD. Metallocarboxypeptidases. In Handbook of Metalloproteins. Messerschmidt A, Bode W, Cygler M. Wiley & Sons 2004; pp. 167-192
13. Salzet M. Anticoagulants and inhibitors of platelet aggregation derived from leeches. *FEBS Lett* 2001; 492: 187-192
14. Mans BJ & Neitz AWH. Adaptation of ticks to a blood-feeding environment: evolution from a functional perspective. *Insect Biochem Mol Biol* 2004; 34: 1-17

15. Rydel TJ, Ravichandran KG, Tulinsky A, Bode W, Huber R, Roitsch C & Fenton JW II. The structure of a complex of recombinant hirudin and human alpha-thrombin. *Science* 1990; 249: 277-280
16. van de Locht A, Stubbs MT, Bode W, Friedrich T, Bollschweiler C, Höffken W & Huber R. The ornithodorin-thrombin crystal structure a key to the TAP enigma? *EMBO J* 1996; 15: 6011-6017
17. Fuentes-Prior P, Noeske-Jungblut C, Donner P, Schleuning WD, Huber R & Bode W. Structure of the thrombin complex with triabin, a lipocalin-like exosite-binding inhibitor derived from a triatome bug. *Proc Natl Acad Sci USA* 1997; 94: 11845-11850
18. Wei A, Alexander RS, Duke J, Ross H, Rosenfeld SA & Chang CH. Unexpected binding mode of tick anticoagulant peptide complexed to bovine factor Xa. *J Mol Biol* 1998; 283: 147-154
19. Arolas JL, Lorenzo J, Rovira A, Castella J, Aviles FX & Sommerhoff CP. A carboxypeptidase inhibitor from the tick *Rhipicephalus bursa*. Isolation, cDNA cloning, recombinant expression, and characterization. *J Biol Chem* 2005; 280: 3441-3448
20. Wang W, Boffa MB, Bajzar L, Walker JB & Nesheim ME. A study of the mechanism of inhibition of fibrinolysis by activated thrombin-activable fibrinolysis inhibitor. *J Biol Chem* 1998; 273: 27176-27181
21. Bouma BN & Meijers JC. Thrombin-activatable fibrinolysis inhibitor (TAFI, plasma procarboxypeptidase B, procarboxypeptidase R, procarboxypeptidase U). *J Thromb Haemost* 2003; 1: 1566-1574
22. Springman EB, Dikov MM & Serafin WE. Mast cell procarboxypeptidase A. Molecular modeling and biochemical characterization of its processing within secretory granules. *J Biol Chem* 1995; 270:1300-1307
23. Bajzar L, Jain N, Wang P & Walker JB. Thrombin activatable fibrinolysis inhibitor: not just an inhibitor of fibrinolysis. *Crit Care Med* 2004; 32: S320-324
24. Barbosa-Pereira PJ, Segura-Martín S, Oliva B, Ferrer-Orta C, Aviles FX, Coll M, Gomis-Rüth FX & Vendrell J. Human procarboxypeptidase B: three-dimensional structure and implications for thrombin-activatable fibrinolysis inhibitor (TAFI). *J Mol Biol* 2002; 321: 537-547
25. Reverter D, Ventura S, Villegas V, Vendrell J & Aviles FX. Overexpression of human procarboxypeptidase A2 in *Pichia pastoris* and detailed characterization of its activation pathway. *J Biol Chem* 1998; 273: 3535-3541
26. Ventura S, Villegas V, Sterner J, Larson J, Vendrell J, Hershberger CL & Aviles FX. Mapping the pro-region of carboxypeptidase B by protein engineering. Cloning, overexpression, and mutagenesis of the porcine proenzyme. *J Biol Chem* 1999; 28: 19925-19933

27. Kabsch W. Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* 1993; 26: 795-800
28. Kilshtain-Vardi A, Glick M, Greenblatt HM, Goldblum A & Shoham G. Refined structure of bovine carboxypeptidase A at 1.25 Å resolution. *Acta Crystallogr D* 2003; 59: 323-333
29. Collaborative Computational Project Number 4. The CCP4 Suite: programs for protein crystallography. *Acta Crystallogr D* 1994; 50: 760-763
30. Perrakis A, Harkiolaki M, Wilson KS & Lamzin VS. ARP/wARP and molecular replacement. *Acta Crystallogr D* 2001; 57: 1445-1450
31. Perrakis A, Morris R & Lamzin VS. Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 1999; 6: 458-463
32. McRee DE. XtalView/Xfit - A versatile program for manipulating atomic coordinates and electron density. *J Struc Biol* 1999; 125: 156-165
33. Lamzin VS & Wilson KS. Automated refinement of protein models. *Acta Crystallogr D* 1993; 49: 129-149
34. van de Locht A, Lamba D, Bauer M, Huber R, Friedrich T, Kroger B, Hoffken W & Bode W. Two heads are better than one: crystal structure of the insect derived double domain Kazal inhibitor rhodniin in complex with thrombin. *EMBO J* 1995; 14: 5149-5157
35. Mans BJ, Louw AI & Neitz AW. Amino acid sequence and structure modeling of savignin, a thrombin inhibitor from the tick, *Ornithodoros savignyi*. *Insect Biochem Mol Biol* 2002; 32: 821-828
36. Holm L & Sander C. Protein structure comparison by alignment of distance matrices. *J Mol Biol* 1993; 233: 123-138
37. Torres AM & Kuchel, PW. The β-defensin-fold family of polypeptides. *Toxicon* 2004; 44: 581-588
38. Aloy P, Catasus L, Villegas V, Reverter D, Vendrell J & Aviles FX. Comparative analysis of the sequences and three-dimensional models of human procarboxypeptidases A1, A2 and B. *Biol Chem* 1998; 379: 149-155
39. Erlanson DA, Wells JA & Braisted AC. Tethering: fragment-based drug discovery. *Annu Rev Biophys Biomol Struct* 2004; 33: 199-223
40. Franco RF, Fagundes MG, Meijers JC, Reitsma PH, Lourenco D, Morelli V, Maffei FH, Ferrari IC, Piccinato CE, Silva WA Jr & Zago MA. Identification of polymorphisms in the 5'-untranslated region of the tafi gene: relationship with plasma TAFI levels and risk of venous thrombosis. *Haematologica* 2001; 86: 510-517
41. Eichinger S, Schonauer V, Weltermann A, Minar E, Bialonczyk C, Hirschl M, Schneider B, Quehenberger P & Kyrle PA. Thrombin-activatable fibrinolysis inhibitor and the risk for recurrent venous thromboembolism. *Blood* 2004; 103: 3773-3776

42. So AK, Varisco PA, Kemkes-Matthes B, Herkenne-Morard C, Chobaz-Péclat V, Gerster JC & Busso N. Arthritis is linked to local and systemic activation of coagulation and fibrinolysis pathways. *J Thromb Haemost* 2003; 1: 2510-2515
43. Nagashima M, Werner M, Wang M, Zhao L, Light DR, Pagila R, Morser J & Verhallen P. An inhibitor of activated thrombin-activatable fibrinolysis inhibitor potentiates tissue-type plasminogen activator-induced thrombolysis in a rabbit jugular vein thrombolysis model. *Thromb Res* 2000; 98: 333-342
44. Walker JB, Hughes B, James I, Haddock P, Kluft C & Bajzar L. Stabilization versus inhibition of TAFIa by competitive inhibitors *in vitro*. *J Biol Chem* 2003; 278: 8913-8921
45. Nicastro G, Franzoni L, Chiara C, Mancin AC, Giglio JR & Spisni A. Solution structure of crotamine, a Na^+ channel affecting toxin from *Crotalus durissus terrificus* venom. *Eur J Biochem* 2003; 270: 1969-1979
46. Hoover DM, Rajashankar KR, Blumenthal R, Puri A, Oppenheim JJ, Chertov O & Lubkowski J. The structure of human beta-defensin-2 shows evidence of higher order oligomerization. *J Biol Chem* 2000; 275: 32911-32918

Supplementary Information

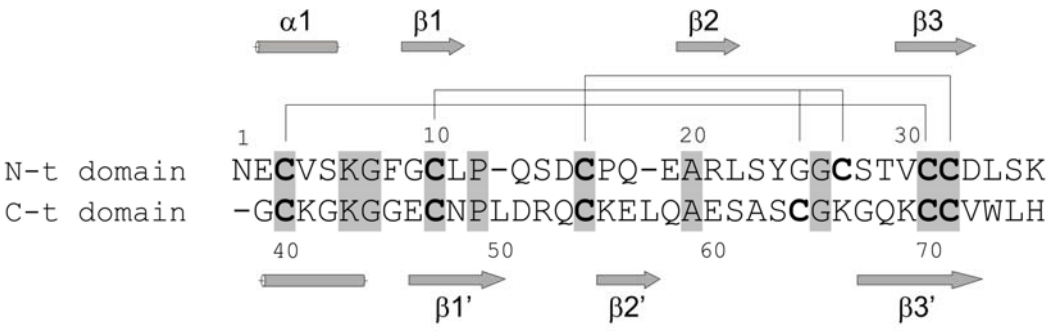


Fig. 1S. Amino acid sequence alignment of the N- and C-terminal domains of TCI. Identical residues are shaded and cysteine residues are printed in bold. The disulfide pairings of both domains are shown. The helices and β-strands are indicated by cylinders and strands, respectively.

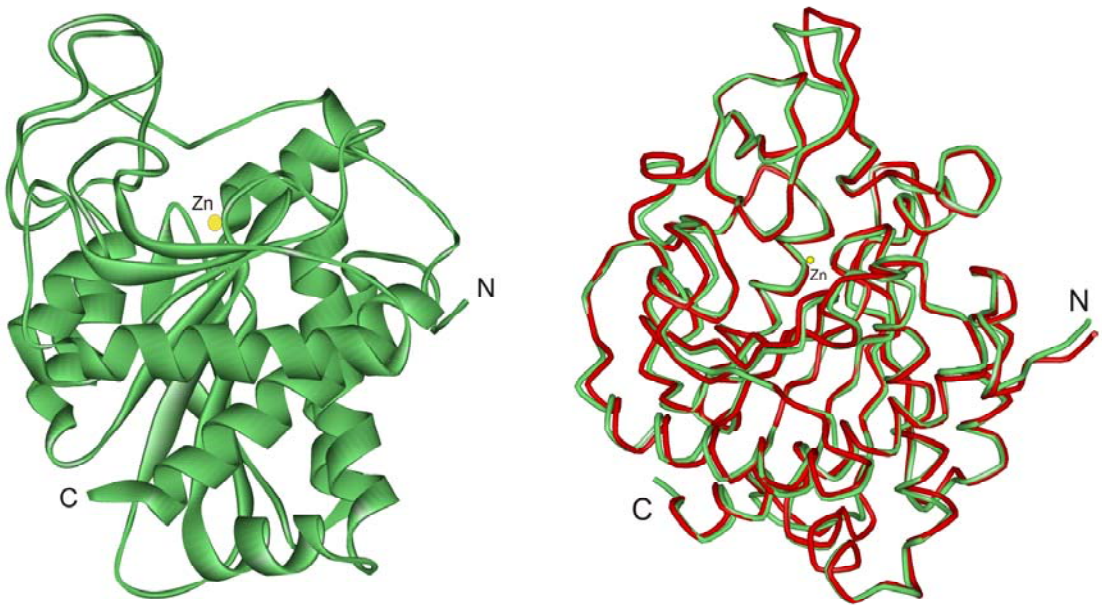


Fig. 2S. Three-dimensional structure of human CPB. *Left panel*. The structure of CPB is shown as a ribbon plot in green. The N- and C-terminal tails are labeled and the catalytic zinc atom is represented by a yellow sphere. *Right panel.* Superposition of human CPB and human TAFIa. The backbone traces of human CPB (green) and of a model of human TAFIa (red), which was constructed previously by our group (Barbosa Pereira et al, 2002), were superimposed showing a high structural similarity (r.m.s.d. 0.77 Å).

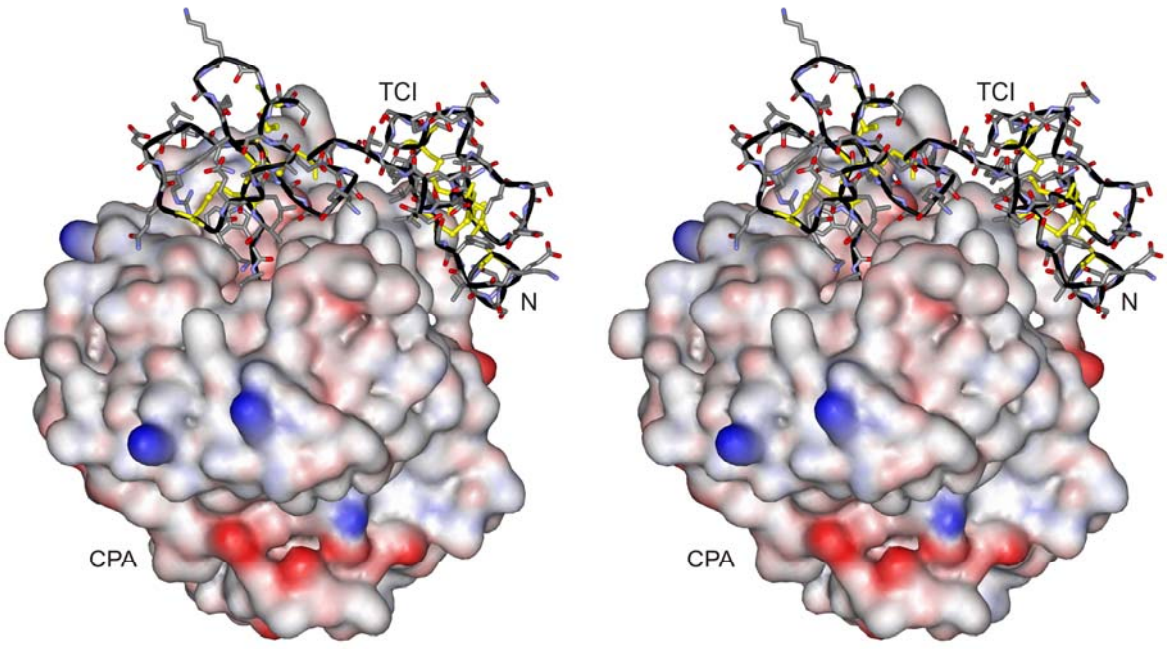


Fig. 3S. Interaction of TCI with CPA. CPA is shown in a solid surface representation; positive surface charges are colored in blue and negative charges in red. TCI is shown in a grey stick model and the disulfide bridges are highlighted in yellow. The N-terminus of TCI is labeled.

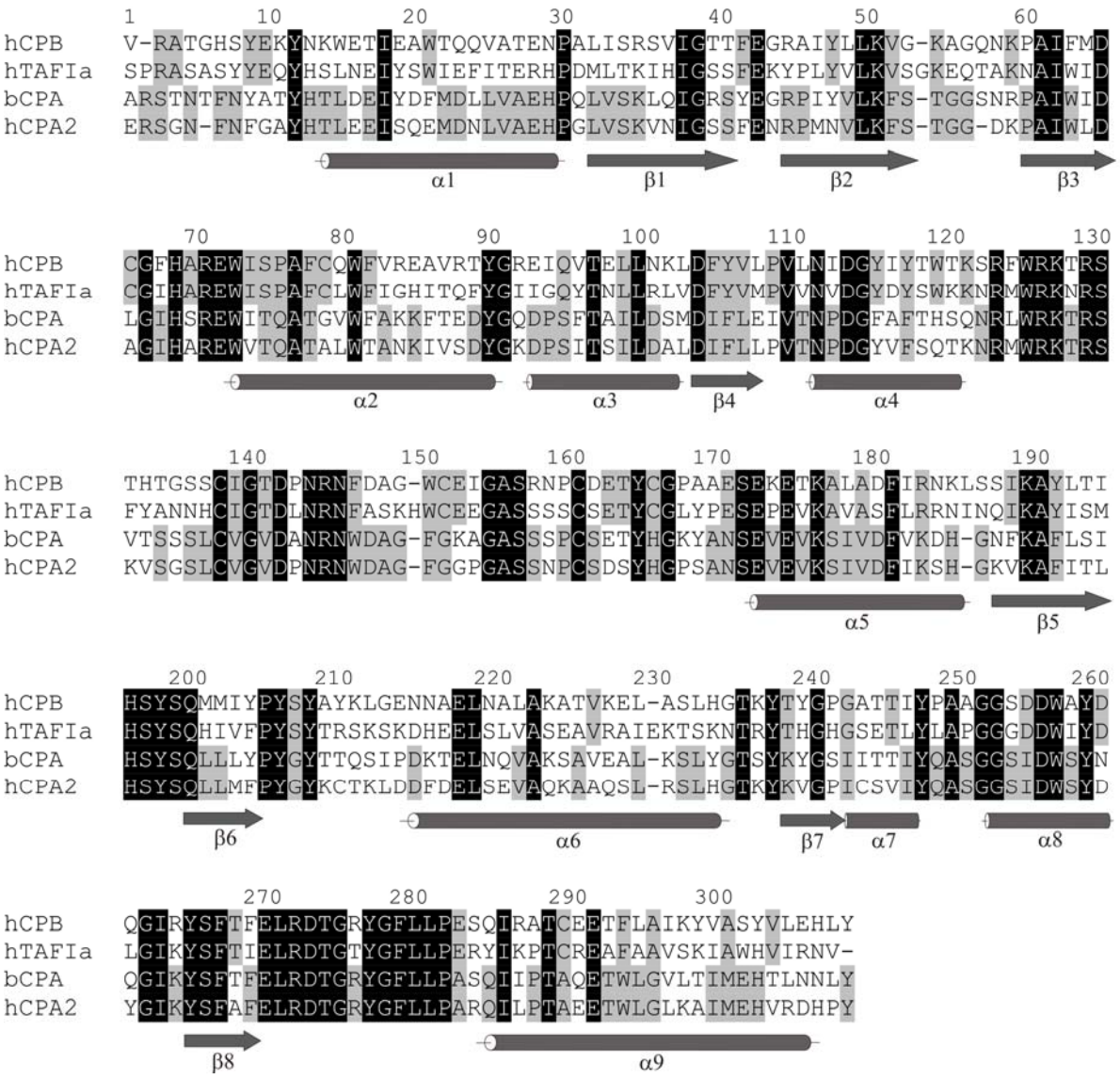
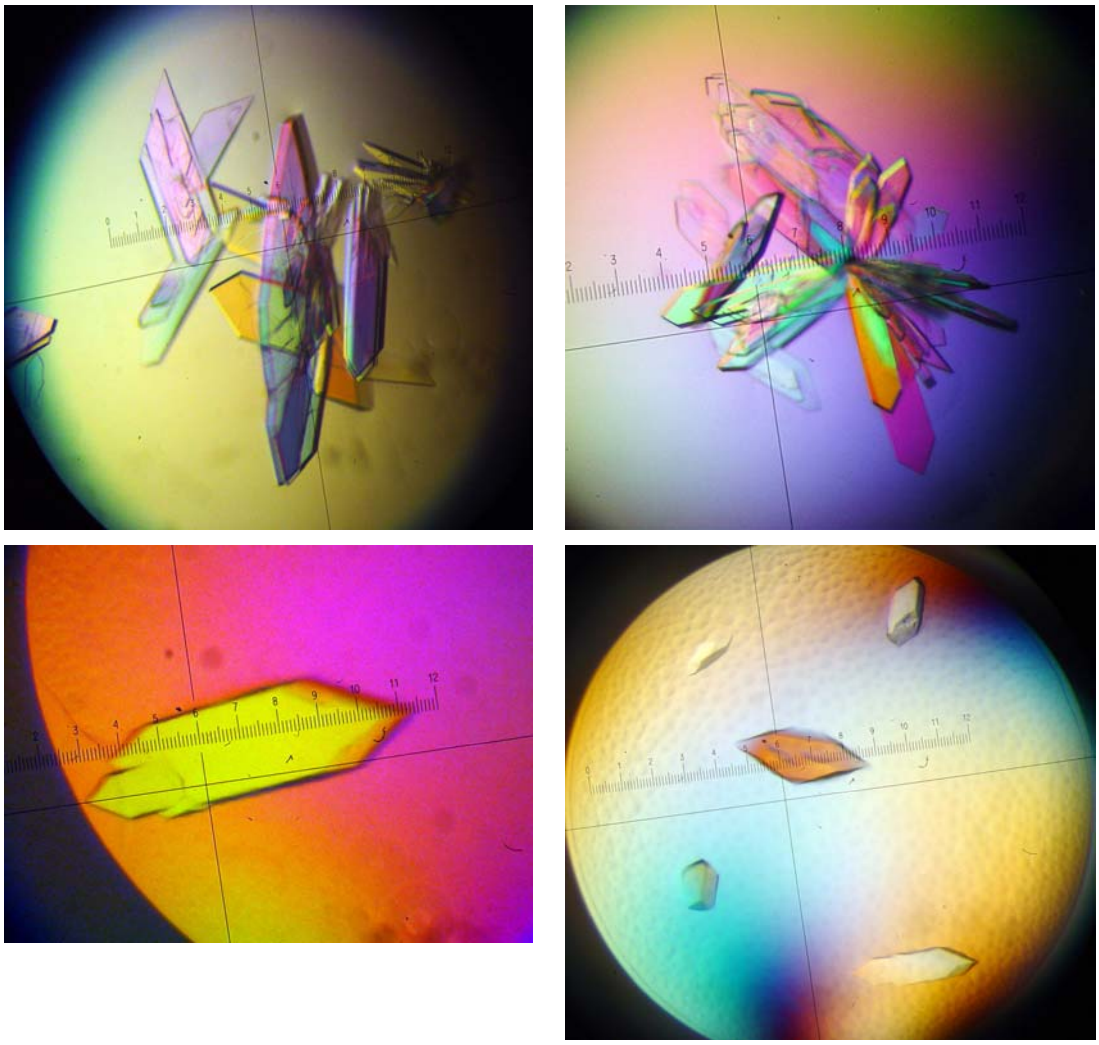


Fig. 4S. Amino acid sequence alignment of human pancreatic carboxypeptidase B (hCPB), human TAFIa (hTAFIa), bovine pancreatic carboxypeptidase A (bCPA), and human pancreatic carboxypeptidase A2 (hCPA2). Residues that are identical in all sequences are printed as white letters on black background. Residues identical in hCPB and hTAFIa or bCPA and hCPA2 are shaded. The helices and β-strands are indicated by cylinders and strands, respectively.

The photographs show the crystals obtained for TCI-bCPA (*top*) and TCI-hCPB (*bottom*) complexes.



Section II. Work 3

The Secondary Binding Site of Potato Carboxypeptidase Inhibitor. Contribution to Its Structure, Folding and Biological Properties

Summary

The contribution of each residue of the potato carboxypeptidase inhibitor (PCI) secondary binding site to the overall properties of this protein has been examined using alanine-scanning mutagenesis. Structural and enzymatic studies, performed on a series of PCI mutants, demonstrate that the proper positioning of the primary site for efficiently binding and inhibition of carboxypeptidase A is significantly dependent on such a secondary contact region. The aromatic residues in this region play a key role in the stabilization of the PCI-enzyme complex, whereas polar residues contribute little to this task. A comparative study of the oxidative folding of these PCI mutants has been carried out using the disulfide quenching approach. The data, together with the structural characterization of some of these mutants, clearly indicate that noncovalent forces drive the refolding of this small disulfide-rich protein at the reshuffling stage, the rate-limiting step of the process. Moreover, it reveals that by introducing new noncovalent intramolecular contacts in PCI, we may create more stable variants, which also show improved folding efficiency. Taken together, the collected results clarify the folding determinants of the primary and secondary binding sites of PCI and their contribution to the inhibition of the carboxypeptidase, providing clues about PCI evolution and knowledge for its biotechnological redesign.

Keywords: potato carboxypeptidase inhibitor; carboxypeptidase A; enzyme-inhibitor complexes; alanine-scanning mutagenesis; protein engineering; protein folding.

Introduction

In the past several years, much effort has been devoted to understanding the interactions that drive the formation of protein-protein complexes (1-3). The associations between proteases and their specific inhibitors are very adequate as models for such studies given the large number of different species, structures, and stabilities that are encompassed (4). The detailed characterization of the energetic and structural basis that lay behind

protease-inhibitor interactions is also of great biotechnological relevance because of their involvement in fundamental biological processes, such as insect attack-vegetal defenses strategies, parasite invasion, blood coagulation/fibrinolysis, inflammation, local anaphylaxis, fertilization, hormone/neuropeptide processing, carcinogenesis, and invasiveness, among others (5-8).

The carboxypeptidase inhibitor (PCI) is a small, 39-amino acid, globular protein naturally occurring in potatoes that can competitively inhibit several carboxypeptidases, forming complexes with an inhibition constant (K_i) in the nanomolar range (5, 9). Its structure has been studied in detail in solution by NMR (10, 11), and by X-ray crystallography in complex with bovine carboxypeptidase A (CPA) (12). This small protein is folded into a 27-residue globular core that lacks regular secondary structures, except for a short five-residue helix and a very small β-sheet, and is stabilized by three disulfide bridges. PCI shares a peculiar disulfide-stabilized loop scaffold, known as the cystine knot or T-knot, with other plant protease inhibitors and with animal peptidic growth factors, such as the epidermal growth factor (EGF) (13). From its hydrophobic central core protrudes a five-residue C-tail (residues 35-39), which is inserted into the active site of the carboxypeptidase and shapes the primary binding site with the enzyme. A short stretch of core residues located around Trp28 (residues 15, 22, 23, and 28-30) constitutes the secondary binding site (12). The contribution of the primary binding site of PCI both to the formation of its complex with CPA and to its thermodynamic stability has been previously evaluated in detail using a site-directed mutagenesis approach (14, 15). These studies suggested that, in addition to the C-terminal tail of PCI, the secondary binding site also plays an important role in PCI function.

The folding process of small disulfide-rich proteins is strongly driven and constrained by the proper disulfide bond formation. The folding reaction of these proteins can be studied using the disulfide acid-trapping approach, in which the folding intermediates that form during their oxidative refolding are trapped at low pH, purified by reversed-phase high performance liquid chromatography (RP-HPLC), and analyzed (16, 17). On the other hand, their unfolding pathways and conformational stabilities can be assayed by disulfide scrambling, based on the observation that the presence of trace amounts of a thiol initiator during unfolding by denaturants generates a mixture of native and disulfide-scrambled species (18, 19).

In PCI, as in hirudin, EGF, or tick anticoagulant peptide (TAP), the folding of the fully reduced and denatured protein to the native form proceeds through an initial stage of

nonspecific disulfide formation (packing), followed by a rate-limiting step of disulfide reshuffling of partially packed intermediates to acquire the native structure (consolidation) (20-27). These small disulfide-containing proteins seem also to share a common unfolding pathway that encompasses a gradual expansion and relaxation of the polypeptide conformation toward a more linear structure (18, 19, 28). In addition, PCI is an example of protein sequences that are unable to fold quantitatively into single native structures, since scrambled isomers exist in equilibrium with the native fold under physiological conditions (28).

In the work presented here, we have used alanine-scanning mutagenesis (29, 30) to gain insights into the role of secondary binding site residues in the folding, structure, and function of PCI. The analysis of the inhibitory, conformational, thermodynamic, and folding characteristics of the different protein mutants has allowed us to dissect the individual contribution of each residue of the secondary binding site in PCI properties. Potential biotechnological applications of these results are discussed.

Experimental Procedures

Materials — All chemicals used in DNA manipulation procedures and the oligonucleotides for site-directed mutagenesis were obtained from Roche Applied Science. Cysteine (Cys), cystine (Cys-Cys), urea, guanidine hydrochloride (GdnHCl), 2-mercaptoethanol, and dithiothreitol, with purities of >99%, were purchased from Sigma. The chromogenic substrate N-(4-methoxyphenylazoformyl)-L-phenylalanine was obtained from Bachem. Bovine CPA was from Sigma.

Molecular Cloning and Site-Directed Mutagenesis — The synthetic gene for PCI (31) was cloned into the pBAT4 plasmid, fused in frame to the OmpA signal sequence (Bronsoms et al., manuscript in preparation). Site-directed mutagenesis was performed on this construction using the QuickChange site-directed mutagenesis kit from Stratagene according to the procedure recommended by the manufacturer. F23A and N29A mutants were used as templates in the PCR to generate PCI forms F23A/W28A and N29A/S30A, respectively. All constructs were verified by DNA sequencing.

Protein Expression and Purification — Recombinant PCI forms were obtained by heterologous expression in *Escherichia coli* strain BL21(DE3). Proteins were purified from the culture medium using a Sep Pak C_{18} cartridge (Waters), followed by anion exchange chromatography on a TSK-DEAE 5PW column (Tosohaas) and by gel filtration

chromatography on a Superdex Peptide column (Amersham Biosciences). Molecular masses were confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). The concentration of the purified solutions of recombinant PCIs was determined from the A_{280} of the final solution (PCI extinction coefficient: $E_{0.1\%}$=3.0). For mutants lacking a Trp residue, the extinction coefficient ($E_{0.1\%}$) was estimated to be 1.7 from the sequence.

Oxidative Refolding — The native protein (0.2 mg/mL) was dissolved in Tris-HCl buffer (0.5 M, pH 8.4) containing 5 M GdnHCl and 30 mM dithiothreitol and kept at 23ºC for 2 h. The reduced and denatured sample was then passed through a PD-10 gel filtration column (Amersham Biosciences) previously equilibrated with Tris-HCl buffer (0.1 M, pH 8.4). The protein was eluted in 1.2 mL and split in three parts that were immediately diluted in the same Tris-HCl buffer to a final protein concentration of 60 μg/mL, in the absence (control -) and presence of Cys (1 mM) or Cys and Cys-Cys (4 and 2 mM, respectively). In some cases, refolding experiments were carried out at a concentration of 0.5 mg/mL in the Tris-HCl buffer, both in the absence (control -) and in the presence (control +) of 0.25 mM 2-mercaptoethanol. Folding intermediates were trapped in a time course manner at selected times by mixing aliquots from the different solutions with an equal volume of 2% trifluoroacetic acid (TFA) in water and analyzed by RP-HPLC on a Protein C_4 column (Vydac, 4.6 mm x 150 mm). Using water containing 0.1% TFA as solvent A and acetonitrile containing 0.1% TFA as solvent B, a linear gradient from 20 to 40% B over the course of 30 min was applied, at a flow rate of 0.75 mL/min.

Denaturation and Unfolding in the Presence of a Denaturant — The native protein (0.5 mg/mL) was dissolved in Tris-HCl buffer (0.1 M, pH 8.4) containing 0.25 mM 2-mercaptoethanol and selected concentrations of denaturants (urea and GdnHCl). The reaction was allowed to reach equilibrium and was typically performed at 23ºC for 20 h. The reactions were then quenched with 2% TFA and analyzed by HPLC using the same conditions described in “Oxidative Refolding”. To follow the time course of unfolding, the native protein (0.5 mg/mL) was dissolved in the same buffer containing 6, 7, or 8 M GdnHCl. At given incubation times, aliquots were removed, the reactions were quenched with 2% TFA, and the mixtures were analyzed by HPLC.

Reductive Unfolding — The native protein (0.5 mg/mL) was dissolved in Tris-HCl buffer (0.1 M, pH 8.4) containing 5 mM dithiothreitol. Reduction was carried out at 23ºC. To monitor the kinetics of unfolding, aliquots of the sample were removed at various time

intervals, the reactions were quenched with 2% TFA, and the mixtures were analyzed by HPLC as described in detailed in "Oxidative Refolding".

Circular Dichroism and NMR Spectroscopy — Samples for circular dichroism (CD) spectroscopy were prepared by dissolving the protein to a final concentration of 0.2 mg/mL in 0.1% aqueous TFA (pH 2.0) or sodium phosphate buffer (20 mM, pH 7.0). CD spectra were collected on a Jasco J-715 spectrometer at 25ºC using a cell with a path length of 2 mm. Protein samples for NMR experiments were prepared by dissolving the protein in H_2O and D_2O (9:1 ratio by volume) at a concentration of ~2 mM at pH 4.0. NMR spectra were acquired on a Bruker AMX 500-MHz spectrometer at 35ºC.

Deuterium to Proton Exchange Followed by MALDI-TOF MS — The native protein (0.2 mg/mL) was dissolved in deuterated Tris-DCl buffer (0.5 M, pD 8.4) containing 5 M GdnHCl and 30 mM dithiothreitol. The sample was acidified to pH 3.0 and passed through a PD-10 column, equilibrated with glycine buffer (20 mM, pH 3.0). The reduced and denatured protein was collected and lyophilized. The sample was then refolded in the deuterated Tris-DCl buffer containing Cys and Cys-Cys (4 and 2 mM, respectively). The native deuterated protein was diluted 1:9 with ammonium citrate (50 mM, pH 5.0) to start the hydrogen exchange. Aliquots were taken at different times and analyzed by MALDI-TOF MS on a Bruker Ultraflex spectrometer until an exchange plateau was reached. Samples were prepared by mixing equal volumes of the protein solution and matrix solution (sinapic acid in aqueous 30% acetonitrile with 0.1% TFA). Six samples were analyzed twice at each exchange time. The average of the mass values corresponding to the centroid of the peaks was calculated for each exchange time and compared to an external unlabeled control, whose mass was determined by duplicate measurements.

CPA Inhibitory Assays — Inhibition constants of the different PCI forms were calculated ($K_i=K_{off}/K_{on}$) by pre-steady-state analysis (32). Kinetic association (K_{on}) and dissociation (K_{off}) constants were determined by a continuous photometric assay with the chromogenic substrate N-(4-methoxyphenylazoformyl)-L-phenylalanine. The assay was performed in Tris-HCl buffer (50 mM, pH 7.5) containing 0.1 M NaCl, with a substrate concentration of 100 μM and a bovine CPA concentration of 2 nM. Several independent K_i determinations were made for each PCI form and corrected for substrate competition. The dissociation free energy of the PCI-CPA complexes was calculated according to the formula $\Delta G_d^o = -RT \ln K_i$.

Results

Role of the PCI Secondary Binding Site in CPA Inhibition — In the PCI-CPA complex, the C-terminal amino acid tail of PCI docks into the active site of the enzyme, leading to a stopper-like inhibition mechanism (5, 12). The carboxy-terminal residue of PCI (Gly39) is cleaved off by CPA during the first stages of the binding, and remains bound in the S1' subsite of CPA. The carboxylate group of the previous residue (Val38) coordinates with the Zn^{2+} of the active site. The rest of the polypeptide chain of PCI remains away from the CPA surface until residues 27-30. These residues pass along the CPA molecule forming a secondary contact region which also includes the side chains of His15, Trp22, and Phe23 (12), and is shown in Fig.1.

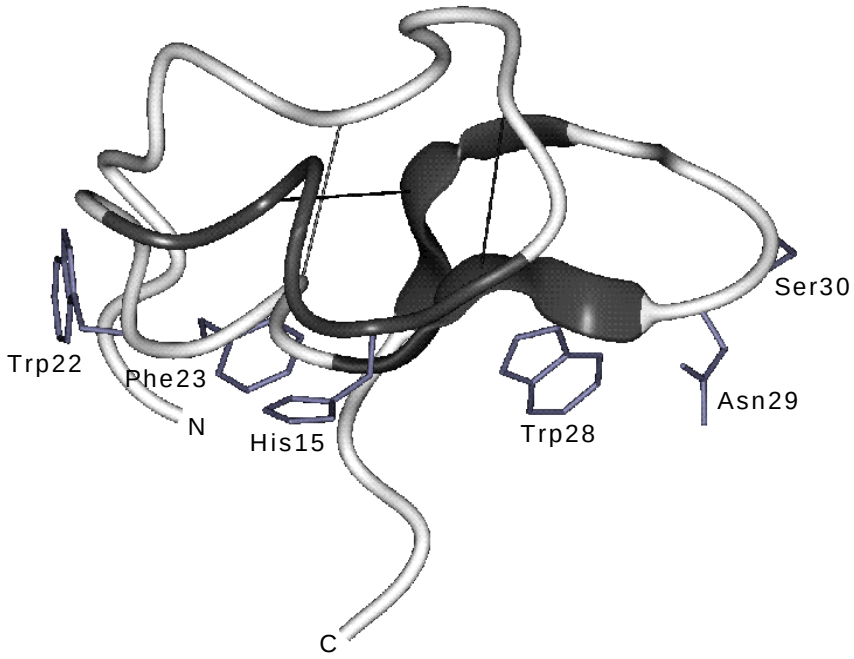


Fig. 1. Schematic view of the native three-dimensional structure of PCI. The secondary binding site residues are depicted in the structure. *N* and *C* denote the locations of N- and C-terminal regions, respectively, of PCI. The Protein Data Bank entry for the structure of PCI is 1H20.

To analyze in a systematic and unbiased way the role and energetic contributions of these residues in the complex, every side chain in this interface was substituted with that of Ala. Moreover, F23A/W28A and N29A/S30A double mutants were produced to asses the additive or cooperative nature of side chains contributions. All mutant proteins were expressed in *E. coli* as secreted soluble proteins and purified to homogeneity, and their identities were confirmed by MS. The purity was greater than 98% in all cases. The RP-

HPLC chromatographic retention displayed by the mutants correlated well with the introduced changes in protein polarity due to amino replacements.

The characterization of the interaction of wild-type (WT) PCI and PCI mutants with CPA in terms of their K_i and the free energy of dissociation of the complexes (ΔG_d^o) are shown in Table 1. Such results allow us to classify residues of the secondary binding site into two groups according to their binding to CPA and side chain characteristics.

Table 1. Association (K_{on}), dissociation (K_{off}) and inhibition (K_i) kinetic constants of wild-type and mutant PCIs for CPA and Gibbs free energy of dissociation (ΔG_d^o) of the respective PCI-CPA complexes

PCI form	K_{on} (M^{-1} s^{-1})	K_{off} (s^{-1})	K_i (nM)	ΔG_d^o of complex (kcal/mol)
WT	2.1×10^6	3.4×10^{-3}	1.6 ± 0.2	12.0
H15A	2.1×10^6	4.8×10^{-3}	2.3 ± 0.3	11.8
W22A	5.2×10^5	4.1×10^{-3}	7.9 ± 0.7	11.1
F23A	1.3×10^5	2.4×10^{-3}	18.5 ± 1.9	10.6
W28A	2.5×10^5	3.3×10^{-3}	13.2 ± 1.5	10.8
N29A	2.6×10^6	4.0×10^{-3}	1.5 ± 0.2	12.0
S30A	3.4×10^6	4.6×10^{-3}	1.4 ± 0.2	12.1
F23A/W28A	8.5×10^3	3.7×10^{-3}	440 ± 37	8.7
N29A/S30A	1.8×10^6	3.4×10^{-3}	1.9 ± 0.4	11.9
N29G	1.1×10^6	3.9×10^{-3}	3.5 ± 0.4	11.5

- *Role of the Secondary Binding Site Polar Residues in CPA Inhibition* — In the PCI-CPA complex, residues Asn29 and Ser30 of PCI are hydrogen bonded to the main chain of residues Ile247 and Thr246 of CPA, respectively. The change of either Asn29 or Ser30 to Ala did not significantly affect the K_i value of the binding, with K_{on} and K_{off} values of the mutant proteins being very similar to those of the WT form. Also, the double Ala mutant (N29A/S30A) retains an inhibitory activity identical to that of WT PCI, clearly indicating that the polar side chains of these residues are not involved in the stabilization of the PCI-CPA complex (Table 1). In the PCI-CPA crystal structure, a hydrogen bond is established between the imidazole group of His 15 and the CPA Tyr248 side chain ring. This contact, again, does not seem to be relevant for complex formation or stabilization since the His to Ala substitution results just in a very moderate increase in the K_i value (Table 1).

- *Role of the Secondary Binding Site Aromatic Residues in CPA Inhibition* — The aromatic residues of the secondary binding site of PCI (Trp22, Phe23, and Trp28) appear to play an important role in its inhibitory capability. All these aromatic nonpolar side chains are solvent-exposed in the uncomplexed molecule, and their mutation to Ala results in PCI variants with seriously decreased inhibitory activity (Table 1). Thus, mutant W22A exhibited a nearly 5-fold increase in the K_i compared to that of WT PCI. As deduced from the crystal structure of the WT complex, this mutant would lose interactions with side chains of CPA residues Phe277 and Arg276. Substitution of Phe23 with Ala caused the largest effect on the inhibitory activity of PCI, with an approximately 12-fold increase in the K_i, indicating that the side chain of Phe23 contributes with 1.4 kcal/mol to the binding energy of the complex. From such a significant contribution, and from the crystal structure, it can be deduced that this mutant would be unable to establish an important interaction with the Phe279 aromatic ring of CPA. In addition, Phe23 of PCI accounts for 21% of the surface area buried in the secondary binding region, and its mutation to Ala diminishes the size of the PCI-CPA contact surface. The third mutant, W28A, exhibited a K_i about 8-fold higher than that of WT PCI, an effect that could be attributed to the loss of the interactions with residues Tyr198, Ser199, and Leu202 of CPA.

Finally, the double substitution F23A/W28A caused a dramatic decrease in inhibitory activity toward CPA, with a K_i that is about 275 times greater than that of WT PCI. From the measured rate constants, it can be seen that such an increase in K_i is mainly due to its slow association with CPA (K_{on} = 8.5 x 10^3). It is worth mentioning that the double mutation causes a 3.3 kcal/mol decrease in the binding energy, which is greater than the 2.6 kcal/mol value expected from a simple additive effect, indicating that a coordinated action of both aromatic side chains takes place in the stabilization of the PCI-CPA complex.

Structural Reasons behind the Role of Phe23 and Trp28 in Complex Stabilization — The conformation of the singly and doubly Ala-substituted mutants was analyzed by CD and NMR spectroscopy in an effort to understand the enthalpic contribution of both Phe23 and Trp28 residues to CPA binding. Although PCI has a small content of secondary structures, far-UV CD spectroscopy was previously reported to be a helpful tool in monitoring the conformational state of PCI variants (21). PCI displays a peculiar CD spectrum with a strong minimum of ellipticity at 204 nm and a maximum at 228 nm, the latter being due to residue Tyr37 (Fig. 2) (21). Interestingly, the scrambled forms of WT PCI, which possess all the residues of the native form but display different conformations,

lack this maximum that could thus be attributed to a native-like environment around Tyr37 (Aviles et al., unpublished data). The CD spectrum of the F23A mutant is very similar to that of the WT protein, while the positive band at 228 nm is absent both in the W28A and F23A/W28A mutants (Fig. 2). It suggests that a native-like conformation prevails in the first case and that Tyr37 is not properly oriented in the last two mutants.

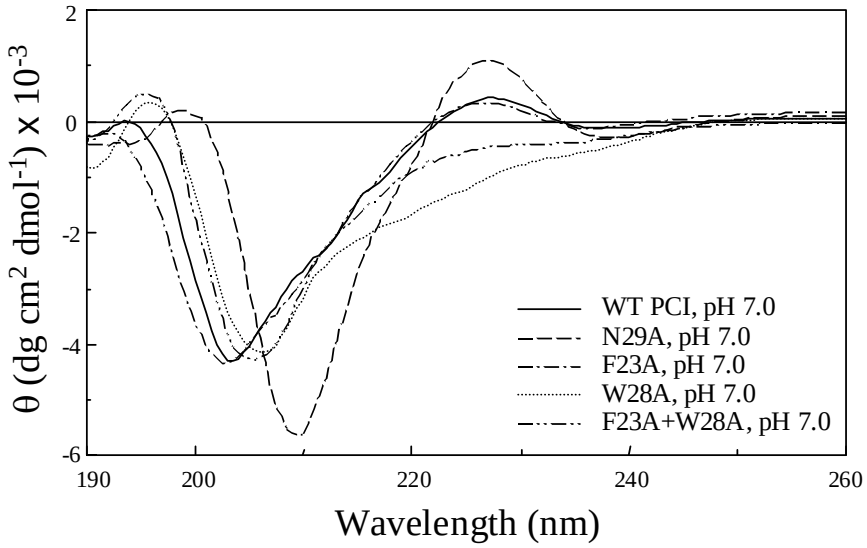


Fig. 2. Circular dichroism spectra of PCI forms. CD analyses of native and mutant PCI forms were carried out in 0.1% aqueous trifluoroacetic acid (pH 2.0) and in sodium phosphate buffer (20 mM, pH 7.0) with a protein concentration of 0.2 mg/mL. Spectra were very similar for each PCI form under both pH conditions.

We have recently shown by NMR that the conformation of PCI in solution does not differ in general terms from that adopted in the crystal when complexed to CPA (11). In the present work, two-dimensional (2D) NMR studies of selected PCI mutant forms were carried out at pH 4.0 and 35ºC. In accord with CD data, both the fingerprint and NH-NH regions of the F23A mutant are very similar to those in the NMR spectra of the WT form (Fig. 3A and 3B), indicating that no important changes in the globular structure of PCI are introduced by this mutation. Thereby, its increased K_i toward CPA can be mainly assigned to the truncation of the Phe23 side chain, which renders this mutant unable to establish proper interactions with CPA.

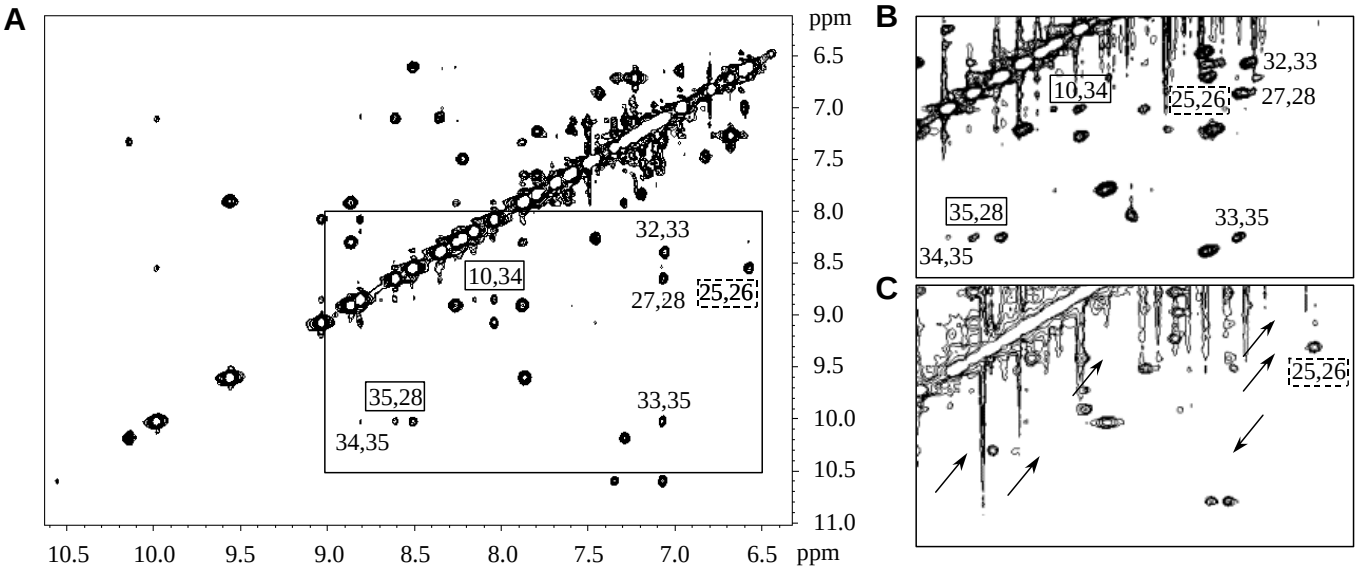


Fig. 3. NH-NH regions of the NOESY spectra of WT (A), F23A (B), and W28A (C) PCI forms. NMR spectra were obtained at 35ºC and a protein concentration of 2 mM in 10% D_2O. Selected cross-peaks between amide protons are shown. The arrows denote WT cross-peaks absent in the mutant. The boxes indicate long-range NOEs. The spectra are unsymmetrized.

In contrast, when compared to WT PCI, the low signal dispersion in the one-dimensional (1D) NMR spectra, the scarce spatial contacts in the fingerprint region and some weaker cross-peaks in the NH-NH region suggest that the mutant W28A has a high degree of flexibility and probably a less compact globular nucleus. In particular, contacts C27-W28, K10-C34, W28-G35, R32-T33, T33-G35, and C34-G35 are lost in this mutant (Fig. 3C). Previous molecular dynamics studies have suggested that the contacts of Trp28 with Gly35 could play an important role in the maintenance of the C-tail of the free inhibitor in an appropriate orientation for binding CPA (33). Thus, the increase in K_i observed for the Ala-substituted mutant can be attributed not only to loss of PCI-CPA contacts but also to an increase in flexibility (entropy) at the C-tail of this mutant. The 1D NMR spectrum of the F23/W28 mutant, although it shows reduced signal dispersion compared to that of WT PCI, still corresponds to a protein with a compact three-dimensional structure. However, the low recombinant expression level of this mutant prevented the acquisition of 2D NMR spectra.

Oxidative Folding of PCI Forms — Oxidative refolding studies of PCI mutant forms were carried out in the absence (control -) and presence of redox agents. Folding intermediates were trapped by acidification at different times during the refolding process

and analyzed by RP-HPLC. The assays were always performed in parallel with a control refolding experiment of WT PCI under the same conditions. The typical results observed for the refolding of WT PCI are shown in Fig. 4 for selected times. The first stage of nonspecific packing of WT PCI to reach scrambled three-disulfide species is very quick, being the consolidation of scrambled species the major rate-limiting step. As previously reported, the disulfide reshuffling of the scrambled intermediates was very slow in the buffer without added thiols (control -), leading to a final trapped mixture with less than 10% of the native species at the end of the folding process. In contrast, in the presence of supplementing thiols (Cys or Cys and Cys-Cys), more than 90% of native PCI was recovered within 3 h of folding. In this reaction, Cys promotes the conversion of scrambled PCIs to the native form without accelerating the flow between fully reduced PCI and scrambled species or altering the apparent composition of one- and two-disulfide intermediates, whereas Cys-Cys drastically enhances the flow between reduced PCI and scrambled species (20).

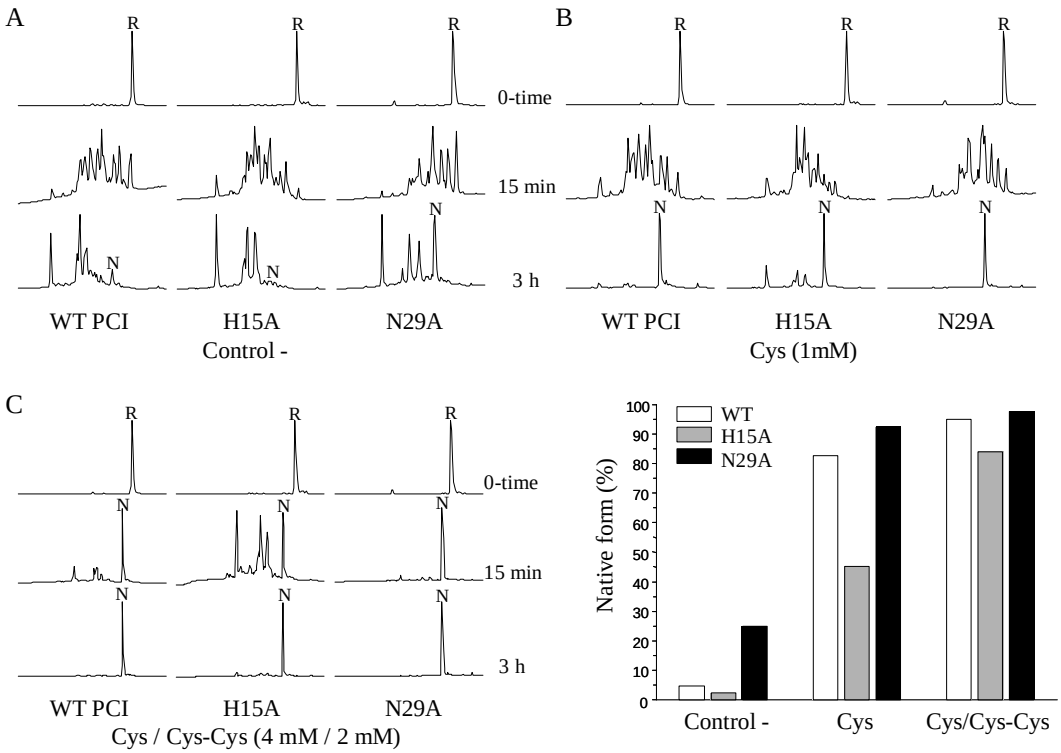


Fig. 4. Oxidative folding of PCI forms in the presence of redox agents. The figure shows the oxidative refolding process of WT, H15A, and N29A PCI forms. Reduced and denatured PCIs were allowed to refold in Tris-HCl buffer (0.1 M, pH 8.4) in the absence (A, control -) and presence of Cys (B, 1 mM) or Cys and Cys-Cys (C, 4 and 2 mM, respectively). Acid-trapped

intermediates were analyzed at the noted times by HPLC as described in detail in "Experimental Procedures". *N* and *R* denote the elution positions of native and fully reduced species, respectively. The graphics show the percentages of native species for each PCI form obtained under different conditions (A-C) after 3 h of refolding, calculated from the peak areas in the corresponding HPLC chromatograms.

The oxidative refolding of all Ala mutants was characterized in detail to test the role of the secondary binding site residues in the folding pathway of PCI. No differences were detected for any mutant, as compared with WT PCI, in the initial stage of nonspecific disulfide formation, neither in folding speed nor in intermediates heterogeneity. This is illustrated for some mutants in Fig. 4. It has been argued that the second step (consolidation) in the refolding process of PCI is extremely slow, when compared to those of other small disulfide-rich proteins such as hirudin (20, 22), because it requires the exposure of hydrophobic sites previously hidden in the scrambled species, increasing the enthalpic cost of the folding reaction (20). To test this hypothesis, we analyzed the folding process of W22A, F23A, and W28A PCI mutants. The folding speed and efficiency of these less hydrophobic variants were similar to that of the WT PCI (Fig. 5). However, the double mutant F23A/W28A exhibited reduced folding efficiency with respect to the WT or singly mutated forms. This behavior does not seem to correspond to an additive effect in loss of hydrophobicity but, as the above-mentioned structural studies suggest, to a somehow altered native conformation.

With regard to the role of polar residues, no significant differences in the folding process were observed when Ser30 was substituted with Ala, while important changes were detected for the other polar residues in the secondary binding site (Fig. 5). Thus, the H15A mutant exhibited slower folding kinetics than the WT form, together with lower percentages of the native form at the end of the folding reaction (~50% of WT in control -). However, the most important changes in the folding process were those observed in mutant N29A. Whereas its initial stage of folding was similar to that of WT PCI, as shown in Fig. 4, the reshuffling process exhibited a high efficiency in the absence of added thiols, with percentages of the native form of about 6-fold higher than those of WT PCI at the end of the process (see Fig. 5). The reshuffling of scrambled species into the native form is considerably accelerated in this mutant: 2 times faster than in the WT form, both without supplementing thiols and with 2-mercaptoethanol, when the refolding is performed at a high protein concentration (0.5 mg/mL; see "Experimental Procedures").

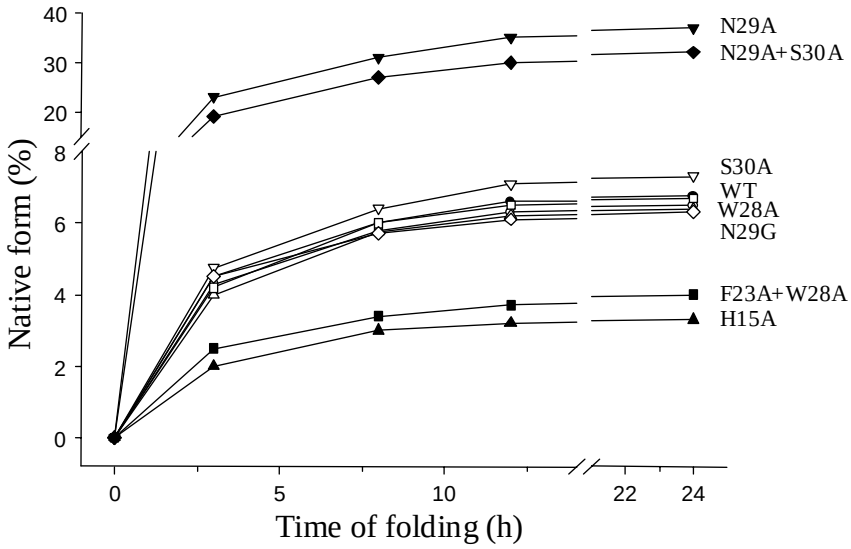


Fig. 5. Folding kinetics of the different PCI forms. Refolding experiments with WT and PCI mutants were performed in the absence of redox agents (control -). Folding intermediates were trapped in a time course manner by acidification, and were then analyzed by HPLC. The symbols show the percentages of native species for the different PCI forms along the folding process, calculated from the peak areas in the corresponding HPLC chromatograms: WT (●), H15A (▲), W22A (○), F23A (△), W28A (□), F23A+W28A (■), N29A (τ), N29G (◇), S30A (▽), and N29A+S30A (◆).

It is worth mentioning that Asn29 is located in a nonregular β-turn linking the two short antiparallel β-strands of PCI, between residues 26-28 and 33-35 (Fig. 1). The analysis of the PCI-CPA crystal structure shows that the three residues in the turn following Asn29 (Ser30, Ala31, and Arg32) have phi and psi angles that are out of the allowed regions in the Ramachandran plot. To test whether the N29A mutant is improving its folding efficiency just because a reduction of the side chain size will give rise to either less entropic cost, less strain, or more flexibility in this turn region, the folding process of a new mutant, in which Asn29 was substituted with Gly, was analyzed. Surprisingly, N29G folded with the same speed and efficiency as the wild-type (Fig. 5), indicating that the efficient folding of the N29A variant can be attributed to the methyl group on the Ala29 side chain and very unlikely to an improved turn flexibility or reduced strain. The double mutant N29A/S30A exhibited the same accelerated kinetics and high folding efficiency as the N29A mutant. It clearly indicates a position-dependent effect of the Ala substitutions in this turn.

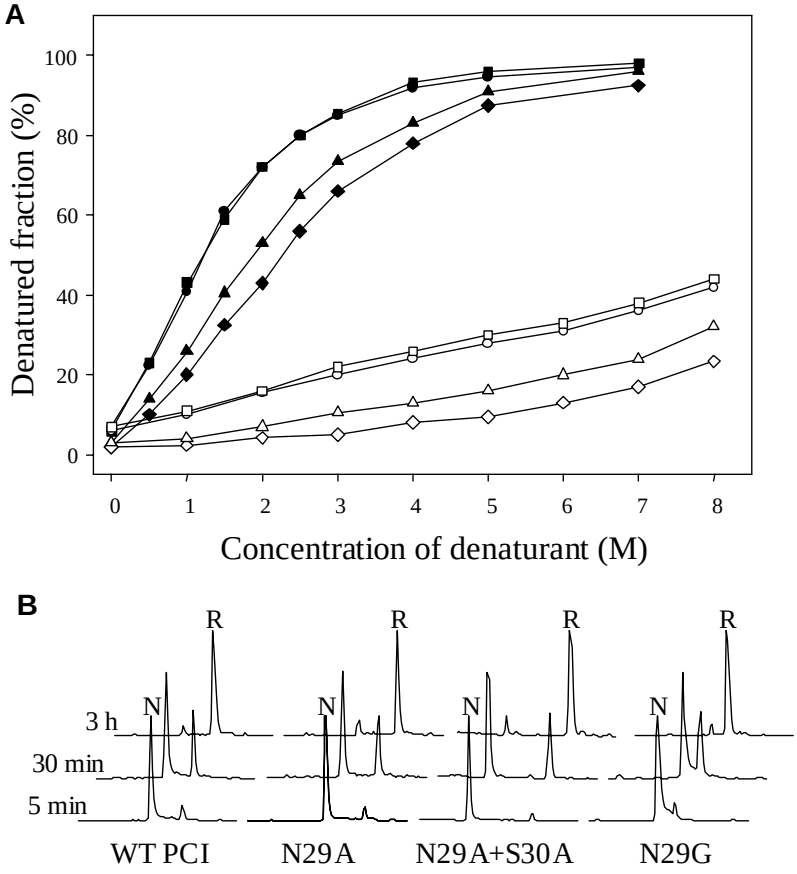


Fig. 6. Denaturation curves and reductive unfolding of PCI forms. (A) The denatured fraction is the percentage of each PCI form that is converted to scrambled isomers. Data for WT, N29G, N29A, and N29A+S30A are depicted with the following symbols: ●, ■, ◆, and ▲, respectively. The denaturants used are GdnHCl (filled symbols) and urea (empty symbols). Denaturation was carried out at 23ºC, for 20 h, in Tris-HCl buffer (0.1 M, pH 8.4) containing 2-mercaptoethanol (0.25 mM) and the indicated concentration of denaturant. (B) Native WT PCI, and N29A, N29A/S30A, and N29G mutants were treated with 5 mM dithiothreitol in Tris-HCl buffer (0.1 M, pH 8.4); the reactions were quenched with 2% aqueous trifluoroacetic acid, and the mixtures were analyzed by HPLC as described in "Experimental Procedures". The peaks corresponding to native (N) and reduced (R) species are indicated.

Conformational Stability of WT, N29A, N29G and N29A+S30A PCI — To study whether the improved folding efficiency of the N29A mutants can be related to a stabilization of the native state of the protein, we derived the denaturation curves of the WT and mutated forms. The extent of unfolding, and hence the equilibrium constant between scrambled species and the native proteins, was clearly dependent upon the strength of the denaturant (Fig. 6A). On the basis of the concentration that is required to

achieve the same extent of PCI denaturation, GdnHCl was about 5-fold more potent than urea, which was unable to fully denature any of the native forms, in agreement with previous reports (28, 34, 35). PCI, while being highly stable in front of GndHCl, unfolds in a cooperative-like manner that suggests the progressive loss of noncovalent interactions during the unfolding event. The denaturation curves, calculated from the fraction percent of the PCI form converted to the scrambled isomers in the presence of the denaturant and thiol initiator, are shown in Fig. 6A. The concentrations of GdnHCl required to achieve 50% denaturation were 2.4 and 1.9 M, for N29A and N29A/S30A, respectively, and 1.2 M for the WT form, indicating that the relative stability of these mutants was significantly higher than that of the wild-type. The N29G mutant exhibited the same stability as WT PCI, in front of both urea and GdnHCl. The unfolding rate constants for WT and N29A PCI at 6, 7, and 8 M GdnHCl were found to be very similar (data not shown).

The conformational stability of WT and N29A PCI was also analyzed by deuterium to proton (D/H) exchange experiments followed by MALDI-TOF MS. The kinetics of interconversion were similar for both proteins, but the extent of hydrogen exchange was fairly different. WT PCI only retains five deuterons at the end of the reaction because of its low secondary structure content, while the N29A mutant retains two additional protected deuterons. This 40% increase in the level of protected deuterons is in agreement with the increase in stability deduced from the denaturation curves. It is worth to mentioning that the N29A mutant showed a recombinant expression yield approximately 2-fold higher than that of WT PCI, another indication of higher stability.

Structural Reasons behind the Improved Conformational Stability and Folding Efficiency of N29A Form — The 228 nm spectral maximum related to the N-terminal Tyr 37 environment in the CD spectrum of the N29A mutant is almost 2-fold more intense than in WT PCI (Fig. 2). The minimum in ellipticity is located at 209 nm, instead of the 204 nm band typical for the WT form, and is also more intense. These features, and the sensitivity of such bands to denaturants (data not shown), suggest that this mutant is a more compact folded and structured protein. The analyses of the 2D NOESY NMR spectra show an increase in the number and intensity of NOEs contacts for the N29A form (see Fig. 1S of "Supplementary Information"). Most of the new or stronger NOEs observed in the NH-NH region involve residues Ala26 to Gly35 (A26-G35, Q25-G35, C27-W28, W28-G35, A31-R32, R32-T33, C34-G35, and P36-T37) (Fig. 7). Although Ala 29 does not seem to be directly involved in any of the new contacts, they are indicative of stronger noncovalent interactions, and a more stable β-hairpin region.

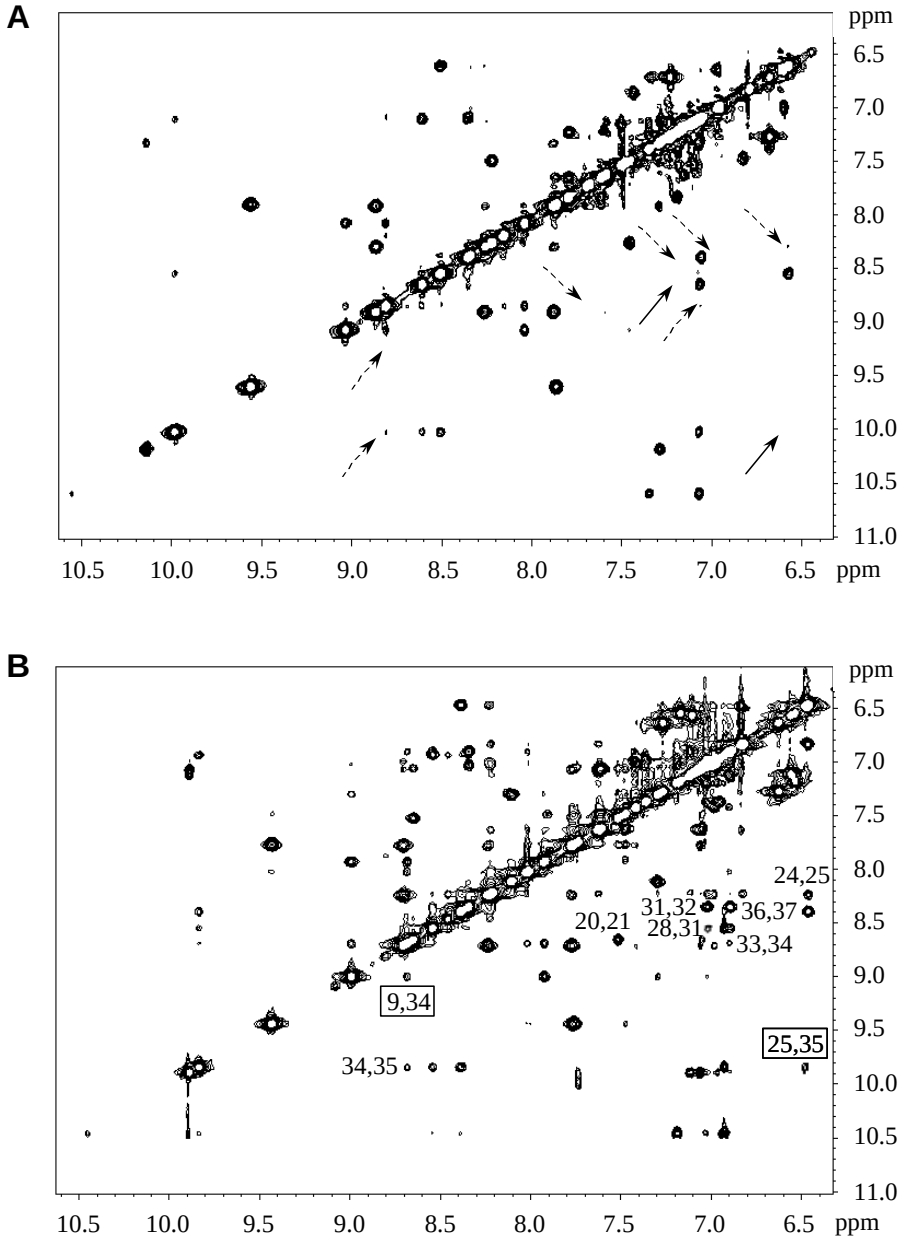


Fig. 7. NH-NH regions of the NOESY spectra of WT (A) and N29A (B) PCI forms. The experimental conditions were the same as those described in the legend of Figure 3. Selected cross-peaks between amide protons are shown in the N29A spectrum. The cross-peaks absent in the WT form (versus the mutant) are denoted with arrows. The boxes indicate long-range NOEs. The spectra are unsymmetrized.

Surprisingly, the change of Asn29 to Ala not only results in local changes but also has a long-range effect, since new contacts far away from the mutation point were observed, mainly located in the short 3_{10}-helix generated by two consecutive type III reverse turns between residues 14 and 19. The new NOEs are: C12-K13, C12-T33, K13-

D17, D16-D17, and G20-A21. Interestingly enough, this 3_{10}-helix constitutes, together with the above-mentioned short β-sheet, the only secondary structure elements in PCI.

The Mechanism of Reductive Unfolding of WT, N29A, N29G and N29A+S30A PCI Forms — In the absence of denaturant, a high concentration of the reducing agent is needed to break the three PCI native disulfides (28). Reduction of the three native disulfide bonds of this protein follows the "all-or-none" mechanism observed in numerous small single-domain proteins such as hirudin and TAP (18, 36). In this mechanism, reduction of the first disulfide bond is a rate-limiting step, after which the remaining cystines are quickly reduced, there is little accumulation of partially reduced species, and protein structure is lost.

Reduction of WT, N29A, N29G, and N29A/S30A PCI at a fixed concentration of dithiothreitol (5 mM) was followed over time to test whether the additional noncovalent forces present in the N29A variant may influence the reductive unfolding reaction of PCI. Under this condition, reduction of the three native disulfide bonds of these proteins followed the described "all-or-none" mechanism (Fig. 6B), suggesting that their native disulfide bonds are stabilized in a concerted and interdependent manner (18). Kinetic analysis revealed that the reduction rates of these proteins were identical (data not shown).

Relationship between Conformational Stability and Structural Uniqueness in PCI — Our group previously reported that when native PCI is incubated in a physiological buffer (pH 7.4-8.4) containing trace amounts of a thiol agent, a minute fraction of the protein spontaneously shuffles its native disulfide bonds and isomerizes to form scrambled species (28). To study whether there is any relationship between conformational stability and the ratio of non-native ensembles under physiological conditions, the equilibrium constant for some PCI variants with altered stability was determined in the presence of 0.25 mM 2-mercaptoethanol to "jump start" such spontaneous conversion. The reaction reached a state of equilibrium after 24 h, and the percentages of scrambled species for the different variants were as follows: 6% for WT (K'_{eq} = 0.06, ΔG_d^o = 1.7 kcal/mol), 7% for N29G (K'_{eq} = 0.07, ΔG_d^o = 1.6 kcal/mol), 2% for N29A (K'_{eq} = 0.02, ΔG_d^o = 2.3 kcal/mol), 3% for N29A/S30A (K'_{eq} = 0.03, ΔG_d^o = 2.1 kcal/mol) and 14% for W28A (K'_{eq} = 0.14, ΔG_d^o = 1.2 kcal/mol).

Discussion

Role of the PCI Secondary Binding Site in CPA Inhibition — Although the structure of the PCI-CPA complex is relatively well known, the knowledge of the basis of the complex formation and of its thermodynamic stability is still scarce. To gain insight into the latter issues, our group previously analyzed the contribution of C-tail residues of PCI to the binding to CPA (14, 15). Such analysis dissected the role played by each residue in the PCI primary binding site and suggested a significant involvement of the secondary binding site residues in the interaction. Our study shows that the polar residues in the secondary binding region of PCI (His15, Asn29, and Ser30) are not likely to be involved in the stabilization of the PCI-CPA complex, whereas aromatic residues (Trp22, Phe23, and Trp28) appear to play an important role in it since about 30% of the free energy of binding can be attributed to them.

A common feature of proteins with the cysteine knot fold is the presence of large hydrophobic patches on the protein surface, which are stabilized by the covalent disulfide scaffold when function requires those large hydrophobic surfaces to be exposed in small volumes (13). This is also the case for PCI, where the surface aromatic patch formed by residues 22, 23, and 28 is an important factor in its inhibitory activity. Our studies allow us to dissect the individual contribution of each side chain in this patch. The structural results in this work show that the F23A mutant keeps a structure very similar to that of WT PCI, whereas W28A loses some of the contacts between the secondary contact region and the C-tail present in the WT form. Therefore, although both residues are basic for PCI function, as deduced from the weak inhibitory capability of their Ala mutants, they clearly play different roles in this interaction. The importance of Phe23 in the binding of PCI to CPA can probably be attributed to the interactions established between its aromatic side-chain and the carboxypeptidase surface. In contrast, the importance of Trp28 seems to rely on the maintenance of the C-terminal tail in a proper orientation, providing the conformational rigidity needed for the binding to the enzyme. These observations may facilitate the rational improvement of the PCI inhibitory properties toward carboxypeptidases.

Structural Reasons behind the low Folding Efficiency of PCI — The folding mechanism of several small disulfide-rich proteins, such as PCI, hirudin, and TAP, can be clearly dissected in two distinct stages (20, 22, 26). An initial stage of nonspecific disulfide pairing leads to the formation of scrambled species as the crucial folding intermediates. The final stage of the folding process involves the reshuffling of the scrambled disulfides

into the native conformation. For these proteins, the folding forces from noncovalent bonds alone are apparently insufficient to drive the unfolded protein across the energy barrier, and a reduction of the conformational search by disulfide bond formations is needed. It is only at the final stage of folding where the attainment of the native structure is probably driven by noncovalent interactions (18, 20). According to this theory, the second stage of folding in PCI, which contains fewer noncovalent bonds than hirudin or TAP, is extremely inefficient compared to these topologically related proteins.

Two alternative explanations have been proposed for the very slow consolidation process of PCI (20). First, its disulfide structure is different from that in hirudin and TAP, also with three disulfide bonds: the two disulfides linking sequences Cys8-Asn-Lys-Pro-Cys12 and Cys24-Glu-Ala-Cys27 form a ring to which the third disulfide (Cys18-Cys34) is threaded. This tight conformation has been suggested to impose structural constraints and prevent efficient formation of the native structure. However, our finding that some PCI mutants that display the same disulfide pattern exhibit striking improved folding efficiencies runs against this hypothesis. Second, in PCI, unlike in hirudin and TAP, the scrambled forms are less hydrophobic than the properly folded protein, suggesting that an exposition of previously hidden hydrophobic sites occurs during the consolidation process with the consequent enthalpic cost. From our studies, however, it is clear that the unfavorable change in environment suffered by the hydrophobic side chains of PCI during the consolidation process does not increase the energy barrier of the folding reaction, since the truncation of side chains of W22, F23, or W28 with an Ala mutation does not change either the folding speed or the efficiency, ruling out this hypothesis.

In this work, we have found that the N29A variant exhibits a high refolding efficiency and an accelerated reshuffling process in comparison to WT PCI. In the PCI-CPA crystal structure, Asn29 is located in a nonregular reverse turn that lacks the characteristic 4-1 hydrogen bond and links the two short antiparallel β-strands of PCI (12). In this crystal structure, the ϕ and ψ dihedral angles for the second and third residues of this turn are located in disallowed areas of the Ramachandran plot. Steric clashes make protein folding energetically expensive in this zone and may constitute yet another reason for PCI low folding efficiency. The reduction in side chain size in the N29A mutant may relax the strain in this area by promoting a more canonical conformation or an increased flexibility. Mutant N29G, expected to perform equal or better than N29A, folded in a manner similar to that of the WT form, showing that the improved characteristics of N29A variants are not related to better turn properties but rather restricted to the effect of the

Ala29 side chain. These results agree with the refined structure of free PCI in solution reported by our group "during the development of the present work" (11). In the NMR structure, this turn is more regular and relaxed than that in the complex of PCI with CPA, indicating a clear conformational shift of this turn upon enzyme binding. In solution, this type III β-turn exhibits the characteristic CO(*i*)-NH(*i*+3) backbone hydrogen bonds, and in addition, all turn residues of free PCI are distributed in favorable regions of the Ramachandran plot.

Thus, according to our results, the higher folding efficiency of the N29A mutant can be attributed unequivocally to the effect of the methyl group of the Ala side chain. CD and NMR studies show that N29A is more compact and structured than the WT, probably as a result of both, new short- and long-range effects. Locally, the short β-sheet including this turn appears to be stabilized by the presence of more contacts between the two β-strands. Interestingly, some of the new long-range noncovalent contacts are observed between this region and the 3_{10}-helix of PCI, thus connecting the only structured regions of this small molecule. In PCI, it is the difficult consolidation of the scrambled isomers into the native form that makes its folding process extremely inefficient as compared to those of other similar proteins. In the case of N29A, the mutation introduces new noncovalent interactions that drive more efficiently this stage of reshuffling, leading to a higher efficiency of the folding reaction. All these points indicate that the low efficiency of folding exhibited by PCI is probably due to the scarce noncovalent contacts displayed by this molecule.

Conformational Stability and Structural Uniqueness in PCI — PCI is a very stable fold probably designed to resist hard external conditions and still be functional (35). The conformational stability of N29A mutants is significantly higher than that of WT PCI, as deduced from their denaturation curves, and correlates with a higher refolding efficiency. The introduction of new long- and short-range interactions in N29A decreases the free energy of the native state of PCI and renders it less susceptible to denaturation. The deuterium to proton exchange method monitored by MS (37), also applied to characterize the conformational stability of this mutant relative to the WT form, shows an increase in the amount of deuterons retained by the N29A mutant. It agrees with the data from its denaturation curves and NMR spectra, establishing a relationship between protection to D/H exchange and stability and/or compactness. While both new short- and long-range

interactions would contribute to the increased protein stability, long-range contacts would lead to increased compactness.

PCI is unable to fold quantitatively into a single structure under physiological conditions, following a behavior of "one-sequence multiple structures" involving the formation of native and scrambled species (28). Our data show that these non-native low-energy packing alternatives (scrambled species) indeed exist in PCI in a state of equilibrium with the native structure. We observed that N29A-containing variants occur with small amounts of scrambled species in equilibrium than with WT PCI, whereas in less stable variants, such as W28A, the equilibrium is more displaced toward the non-native conformations, thus confirming that there is a correlation between the conformational stability of the native ensemble and its structural uniqueness. For a polypeptide chain to fold into a unique conformation, a free energy gap of several kilocalories per mole between the correctly folded protein and alternative folds is required. As we show, increasing this gap by lowering the free energy of the native ensemble will also increase its structural uniqueness. Our observations are in total agreement with those made in proteins causing conformational diseases where stabilization of the native structure has been used as a tool to restrict conformational fluctuations that may drive aggregation of prone ensembles (38).

Biological and Biotechnological Implications — It is interesting to note that the primary site of binding of PCI to CPA is located in a region, the C-tail, that is much less important for its conformation and folding than the secondary binding site. Although both sites are conformationally and functionally inter-related, as we have shown in this work, it is surprising that the secondary binding site resides in a region so critical for the folding of this protein, a potential constraint for functional evolution. The comparison of PCI with other proteic inhibitors of metallocarboxypeptidases shows that the C-tail structure is much more conserved than that of the secondary binding site (5). The latter is located in a region of the protein that changes both in composition and topology from one inhibitor to another. Such structural characteristics are probably important factors in the evolution of these proteins, which would be the consequence of the convergent evolution dictated by the target molecules, the carboxypeptidases (39). The features of the primary and secondary binding sites of PCI described above should be taken into account when used as a lead compound for drug design. Actually, PCI has been proposed as a potential fibrinolytic agent for treatment and/or prevention of thrombotic diseases (40-42) because of its ability to inhibit plasma CPB (TAFI) (43), and great interest in the design of minimized inhibitors appeared (44).

Acknowledgments — We are grateful to Drs. S. Bronsoms, G. Venhudova and F. Canals for technical assistance and helpful discussions. We also thank O. Conchillo for computational support.

References

1. Otlewski J & Apostoluk W. Structural and energetic aspects of protein-protein recognition. *Acta Biochim Pol* 1997; 44: 367-387
2. Edwards AM, Kus B, Jansen R, Greenbaum D, Greenbatt J & Gerstein M. Bridging structural biology and genomics: assessing protein interaction data with known complexes. *Trends Genet* 2002; 18: 529-536
3. Nooren IM & Thornton JM. Diversity of protein-protein interactions. *EMBO J* 2003; 22: 3486-3492
4. Bode W & Huber R. Structural basis of the endoproteinase-protein inhibitor interaction. *Biochim Biophys Acta* 2000; 1477: 241-252
5. Vendrell J, Querol E & Aviles FX. Metallocarboxypeptidases and their protein inhibitors. Structure, function and biomedical properties. *Biochim Biophys Acta* 2000; 1477: 284-298
6. Coughlin SR. Thrombin signalling and protease-activated receptors. *Nature* 2000; 14: 258-264
7. Turk B, Stoka V, Rozman-Pungercar J, Cirman T, Droga-Mazovec G, Oreic K & Turk V. Apoptotic pathways: involvement of lysosomal proteases. *Biol Chem* 2002; 383: 1035-1044
8. Stamenkovic I. Extracellular matrix remodelling: the role of matrix metalloproteinases. *J Pathol* 2003; 200: 448-464
9. Hass GM & Ryan CA. Carboxypeptidase inhibitor from potatoes. *Methods Enzymol* 1981; 80: 778-791
10. Clore GM, Gronenborn AM, Nilges M & Ryan CA. Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. *Biochemistry* 1987; 26: 8012-8023
11. Gonzalez C, Neira JL, Ventura S, Bronsoms S, Rico M & Aviles FX. Structure and dynamics of the potato carboxypeptidase inhibitor by ^{1}H and ^{15}N NMR. *Proteins* 2003; 50: 410-422
12. Rees DC & Lipscomb WN. Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. *J Mol Biol* 1982; 160: 475-498
13. Lin SL & Nussinov R. A disulphide-reinforced structural scaffold shared by small proteins with diverse functions. *Nat Struct Biol* 1995; 2: 835-837
14. Molina MA, Marino C, Oliva B, Aviles FX & Querol E. C-tail valine is a key residue for stabilization of complex between potato inhibitor and carboxypeptidase A. *J Biol Chem* 1994; 269: 21467-21472

15. Marino-Buslje C, Venhudova G, Molina MA, Oliva B, Jorba X, Canals F, Aviles FX & Querol E. Contribution of C-tail residues of potato carboxypeptidase inhibitor to the binding to carboxypeptidase A. A mutagenesis analysis. *Eur J Biochem* 2000; 267: 1502-1509
16. Creighton TE. Disulfide bonds as probes of protein folding pathways. *Methods Enzymol* 1986; 131: 83-106
17. Creighton TE. Protein folding. *Biochem J* 1990; 270: 1-16
18. Chang JY. A two-stage mechanism for the reductive unfolding of disulfide-containing proteins. *J Biol Chem* 1997; 272: 69-75
19. Chang JY. Denatured states of tick anticoagulant peptide. Compositional analysis of unfolded scrambled isomers. *J Biol Chem* 1999; 274: 123-128
20. Chang JY, Canals F, Schindler P, Querol E & Aviles FX. The disulfide folding pathway of potato carboxypeptidase inhibitor. *J Biol Chem* 1994; 269: 22087-22094
21. Venhudova G, Canals F, Querol E & Aviles FX. Mutations in the N- and C-terminal tails of potato carboxypeptidase inhibitor influence its oxidative refolding process at the reshuffling stage. *J Biol Chem* 2001; 276: 11683-11690
22. Chatrenet B & Chang JY. The disulfide folding pathway of hirudin elucidated by stop/go folding experiments. *J Biol Chem* 1993; 268: 20988-20996
23. Chang JY. Controlling the speed of hirudin folding. *Biochem J* 1994; 300: 643-650
24. Wu J, Yang Y & Watson JT. Trapping of intermediates during the refolding of recombinant human epidermal growth factor (hEGF) by cyanylation, and subsequent structural elucidation by mass spectrometry. *Protein Sci* 1998; 7: 1017-1028
25. Chang JY, Li L & Lai PH. A major kinetic trap for the oxidative folding of human epidermal growth factor. *J Biol Chem* 2001; 276: 4845-4852
26. Chang JY. The disulfide folding pathway of tick anticoagulant peptide (TAP), a Kunitz-type inhibitor structurally homologous to BPTI. *Biochemistry* 1996; 35: 11702-11709
27. Chang JY & Ballatore A. Structure and heterogeneity of the one- and two-disulfide folding intermediates of tick anticoagulant peptide. *J Protein Chem* 2000; 19: 299-310
28. Chang JY, Li L, Canals F & Aviles FX. The unfolding pathway and conformational stability of potato carboxypeptidase inhibitor. *J Biol Chem* 2000; 275: 14205-14211
29. Cunningham BC & Wells JA. High-resolution epitope mapping of hGH-receptor interactions by alanine-scanning mutagenesis. *Science* 1989; 244: 1081-1085
30. Ashkenazi A, Presta LG, Marsters SA, Camerato TR, Rosenthal KA, Fendly BM & Capon D J. Mapping the CD4 binding site for human immunodeficiency virus by alanine-scanning mutagenesis. *Proc Natl Acad Sci USA* 1990; 87: 7150-7154
31. Molina MA, Aviles FX & Querol E. Expression of a synthetic gene encoding potato carboxypeptidase inhibitor using a bacterial secretion vector. *Gene* 1992; 116: 129-138

32. Morrison JF. The slow-binding and slow, tight-binding inhibition of enzyne-catalysed reactions. *Trends Biochem Sci* 1982; 7: 102-105
33. Oliva B, Wästlund M, Nilsson O, Cardenas R, Querol E, Aviles FX & Tapia O. Stability and fluctuations of the potato carboxypeptidase A protein inhibitor fold: a molecular dynamics study. *Biochem Biophys Res Commun* 1991; 176: 616-621
34. Bulychev A & Chang JY. Unfolding of hirudin characterized by the composition of denatured scrambled isomers. *J Protein Chem* 1999; 18: 771-778
35. Salamanca S, Villegas V, Vendrell J, Li L, Aviles FX & Chang JY. The unfolding pathway of leech carboxypeptidase inhibitor. *J Biol Chem* 2002; 277: 17538-17543
36. Chang JY, Li L & Bulychev A. The underlying mechanism for the diversity of disulfide folding pathways. *J Biol Chem* 2000; 275: 8287-8289
37. Villanueva J, Canals F, Villegas V, Querol E & Aviles FX. Hydrogen exchange monitored by MALDI-TOF mass spectrometry for rapid characterization of the stability and conformation of proteins. *FEBS Lett* 2000; 472: 27-33
38. Chiti F, Taddei N, Bucciantini M, White P, Ramponi G & Dobson CM. Mutational analysis of the propensity for amyloid formation by a globular protein. *EMBO J* 2000; 19: 1441-1449
39. Reverter D, Vendrell J, Canals F, Horstmann J, Aviles FX, Fritz H & Sommerhoff CP. A carboxypeptidase inhibitor from the medical leech *Hirudo medicinalis*. Isolation, sequence analysis, cDNA cloning, recombinant expression, and characterization. *J Biol Chem* 1998; 273: 32927-32933
40. Nagashima M, Werner M, Wang M, Zhao L, Light DR, Pagila R, Morser J & Verhallen P. An inhibitor of activated thrombin-activatable fibrinolysis inhibitor potentiates tissue-type plasminogen activator-induced thrombolysis in a rabbit jugular vein thrombolysis model. *Thromb Res* 2000; 98: 333-342
41. Schneider M & Nesheim M. Reversible inhibitors of TAFIa can both promote and inhibit fibrinolysis. *J Thromb Haemost* 2003; 1: 147-154
42. Walker JB, Hughes B, James I, Haddock P, Kluft C & Bajzar L. Stabilization versus inhibition of TAFIa by competitive inhibitors in vitro. *J Biol Chem* 2003; 278: 8913-8921
43. Bouma BN & Meijers JCM. Thrombin-activatable fibrinolysis inhibitor (TAFI, plasma procarboxypeptidase B, procarboxypeptidase R, procarboxypeptidase U). *J Thromb Haemost* 2003; 1: 1566-1574
44. Barrow JC, Nantermet PG, Stauffer SR, Ngo PL, Steinbeiser MA, Mao SS, Carroll SS, Bailey C, Colussi D, Bosserman M, Burlein C, Cook JJ, Sitko G, Tiller PR, Miller-Stein CM, Rose M, McMasters DR, Vacca JP & Selnick HG. Synthesis and evaluation of imidazole acetic acid inhibitors of activated thrombin-activatable fibrinolysis inhibitor as novel antithrombotics. *J Med Chem* 2003; 46: 5294-5297

Supplementary Information

Fig. 1S. Fingerprint regions of the NOESY spectra of WT (A) and N29A (B) PCI forms. The experimental conditions were the same as those described in the legend of Figure 3. Cross-peaks are indicated for residues A26 to Y37. The spectra are unsymmetrized.