COMPARISON OF SOLUTION STRUCTURES AND STABILITIES OF NATIVE,
PARTIALLY UNFOLDED AND PARTIALLY REFOLDED PEPSIN

A Thesis

Presented to

The Faculty of Graduate Studies

of

The University of Guelph

by

DEREK ROY DEE

In partial fulfilment of requirements

for the degree of

Master of Science

August, 2006

© Derek Roy Dee, 2006

Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 978-0-494-23191-3

Our file Notre référence

ISBN: 978-0-494-23191-3

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

ABSTRACT

COMPARISON OF SOLUTION STRUCTURES AND STABILITIES OF NATIVE, PARTIALLY UNFOLDED AND PARTIALLY REFOLDED PEPSIN

**Derek Roy Dee
University of Guelph, August 2006**

**Advisor:
Professor Rickey Y. Yada**

This thesis reports on the characterization of the stability and low-resolution tertiary and secondary structure of native, alkaline unfolded and acid refolded porcine pepsin. Through a combination of small-angle neutron scattering (SANS), circular dichroism and differential scanning calorimetry, acid refolded pepsin was shown to have secondary and tertiary structure intermediate between the alkaline denatured and native forms, but was found to be thermodynamically stabilized relative to the native state. It was also observed that the acid refolded state of pepsin was dependent on the protein concentration during refolding, as a concentrated-refolded pepsin (> 10 mg/ml) sample had native-like secondary and tertiary structure. It is proposed that pepsin in its native conformation is thermodynamically metastable and that this is the reason for its irreversible denaturation. Furthermore, the domain architecture and the disruption of the active site cleft in the denatured states could be discerned by modelling of the SANS data.

Acknowledgements

I would like to give thanks to those who have enabled me to come this far in my education and in life. First, thank you to Dr. Rickey Y. Yada for taking a chance on a Foodie from out West and taking me into the lab. Also, thank you for your flexibility as a supervisor and even acting as a witness at my wedding.

Thank you to my advisory committee members, Dr. Leonid Brown and Dr. John Dutcher, and to my examination committee members Dr. Mira de Orduña and Dr. Dalglish, for your time and thoughts during the preparation and examination of this thesis.

Acknowledgements must go to my fellow lab mates Andy, Brian, Charity, Huogen, Reena and Miguel for showing me the ropes, pointing out my mistakes and lending me a key to the attic when necessary. A special thanks goes to Miguel for all the “limon” shrimp and tequila.

I would also like to thank our collaborators at Chalk River Laboratories, Jeremy Pencer, Mu-Ping Nieh and John Katsaras, and Susan Krueger at the NCNR at NIST, for all the discussions and help with the SANS experiments. A special thanks to JP for his attention to detail and many questions, which surely helped make this a better thesis.

Much gratitude goes to my family and friends for supporting me throughout my many university years. Your encouragement has not gone unnoticed or unfelt. Finally, I would like to thank my wife Vindya for the many things she does and puts up with.

Table of Contents

1. Introduction.....	1
2. Literature Review.....	3
2.1 Protein Folding and Thermodynamics.....	3
2.2 Protein Folding Models.....	5
2.3 Kinetic vs. Thermodynamic Control of the Folding Landscape.....	8
2.4 Pepsin-like Aspartic Peptidases.....	9
2.5 Pepsin structure.....	10
2.6 Studies on the Irreversible Denaturation of Pepsin.....	11
2.7 SANS and the Study of Protein Folding.....	15
3. Materials and Methods.....	18
3.1 Materials.....	18
3.2 Sample Preparation.....	18
3.3 Kinetic Analysis.....	19
3.4 Small-Angle Neutron Scattering	20
3.4.1 SANS Data Collection.....	20
3.4.2 SANS Data Analysis.....	20
3.5 Densitometry.....	23
3.6 Circular Dichroism Spectroscopy.....	23
3.7 Differential Scanning Calorimetry.....	24
3.8 Intrinsic Tryptophan Fluorescence.....	27
4. Results.....	27

4.1 Pepsin Behaviour in H ₂ O and D ₂ O.....	27
4.2 Small-Angle Neutron Scattering.....	31
4.2.1 SANS Data Reduction.....	31
4.2.2 Comparison of SANS Results to the Crystal Structure.....	32
4.2.3 Pepsin Radii of Gyration via Guinier Plots.....	33
4.2.4 Kratky Plots.....	40
4.2.5 The Pair Distance Distribution Function, $P(R)$	40
4.2.6 Low Resolution Structure via Simulated Annealing:.....	41
GASBOR	
4.2.7 Low Resolution Structure via Scattering Envelopes: SASHA...41	
4.3 Densitometry.....	47
4.4 Circular Dichroism Analysis.....	47
4.4 Intrinsic Tryptophan Fluorescence.....	51
4.5 Differential Scanning Calorimetry.....	55
4.6 Refolded Pepsin Activity.....	68
5. Discussion.....	68
5.1 Unfolding.....	68
5.2 Refolding.....	70
5.3 Folding Cooperativity.....	72
5.4 Molten Globule vs. Partly Folded States.....	73
5.5 Proposed Role of the Prosegment in Pepsin Folding.....	74
6. Conclusions.....	75
7. References.....	78

List of Tables

Title	Page
Table 1. Techniques for studying protein folding and structure.	16
Table 2. Kinetic constants determined for pepsin in H₂O and D₂O.	30
Table 3. Radii of gyration (<i>R_g</i>) obtained from Guinier approximation of the scattering data.	38
Table 4. Secondary structure estimated from the CD spectra for Np (pH 5.3), Ip (pH 8.0), Rp (pH 5) and CRp (pH 5) in H₂O and D₂O.	50
Table 5. Fluorescence spectral parameters for pepsin in H₂O and D₂O buffer.	54
Table 6. Parameters derived from the DSC scans of the Np, Ip, Rp and CRp states of pepsin in both H₂O and D₂O buffers, for either one or two transitions where observed.	66
Table 7. Cooperativity of unfolding of Np, Rp and CRp indicated by $\Delta H^{\text{cal}} / \Delta H^{\text{vh}}$ ratios.	67
Table 8. Relative stability of Np, CRp and Rp compared using apparent free energy values ΔG_{app} (25 °C).	67

List of Figures

Title	Page
Figure 1. Crystal structure of pepsin (PDB code: 4PEP).	12
Figure 2. Structure of pepsinogen, the zymogen form of pepsin.	13
Figure 3. Kinetic activity of pepsin in H₂O and D₂O.	29
Figure 4. Reduction of raw SANS data to produce background and instrument corrected, absolute-scaled data.	34
Figure 5. Averaging and subtraction of incoherent background from SANS data.	35
Figure 6. Comparison of experimental and theoretical SANS data.	36
Figure 7. Comparison of SANS data from pepsin in D₂O and H₂O.	37
Figure 8. Radii of gyration R_g determined by Guinier approximation in the low-angle region.	39
Figure 9. Kratky plots, showing q^2I vs. q for Np, Ip, Rp and CRp in D₂O.	43
Figure 10. Pair distance distribution functions, $P(R)$, for pepsin in D₂O.	44
Figure 11. Ab initio modelling of the SANS data.	45
Figure 12. Domain structure indicated by low-resolution models.	46
Figure 13. Comparison of circular dichroism spectra from pepsin in four conformational states.	49
Figure 14. Ribbon representation of the porcine pepsin A structure (PDB: 4PEP) highlighting the position of the five tryptophan residues.	52
Figure 15. Fluorescence spectra for pepsin (5.8 μM) in three different states.	53
Figure 16. Two apparent transitions observed upon heating pepsin.	59
Figure 17. Four conformations of pepsin as observed by DSC.	60
Figure 18. Irreversibility of Np unfolding by DSC.	61

List of Figures

Title	Page
Figure 19. Comparison of thermally and pH refolded Rp by CD.	62
Figure 20. Non two-state approximation of DSC transitions.	63
Figure 21. Two-state approximation of DSC transitions.	64
Figure 22. Melting of native pepsin occurred as, at least, a three-state transition.	65
Figure 23. Summary of findings on the irreversible denaturation of pepsin.	77

List of Abbreviations

ANS - 1-anilino-8-naphthalene sulphonic acid

APs – Aspartic Peptidases

CD – Circular Dichroism

CRp – Concentrated Refolded pepsin

DSC – Differential Scanning Calorimetry

Ip – Alkaline denatured pepsin

Np – Native pepsin

MG – Molten Globule

NMR – Nuclear Magnetic Resonance

Rp – Refolded pepsin

Rp_{melt} – Thermally Refolded pepsin

SAS – Small-Angle Scattering

SANS – Small-Angle Neutron Scattering

SAXS – Small-Angle X-ray Scattering

Up – Unfolded pepsin

1. Introduction

Entire genomes can be sequenced and the primary sequences of proteins can be selectively altered. Yet, the fundamental link between the primary sequence of a protein and the functional, folded conformation remains elusive, but is critical to understand to make full use of the genetic information. For example, although the Human Genome Project was successful in mapping most of the human genome (1, 2), without a complete understanding of this link the structure and function of the translational products of most of the ~ 25,000 human genes remain unknown. Relationships between gene and gene-products have and will continue to be established empirically, however, a truly *ab initio* approach is required.

An understanding of protein folding from first principles would allow (i) the accurate prediction of the structure and possible functions of a gene product, and (ii) the precise engineering of proteins for specific applications. The diversity and impact of being able to accurately predict the structure and possible functions of a gene product cannot be overstated. Progress in all areas of the biosciences (including health, agriculture, environment) would benefit greatly, and new bio-industries could develop. For example, the impact on the food cycle, that is, how food is grown (genetic engineering), processed (processing aids, entirely new products and processes), and consumed by humans (nutrigenomics, diet-health relationships) would be substantial.

In establishing the groundwork to understand the links between sequence, structure and function, a general folding mechanism has been sought. It is believed that, generally, all the information required for a polypeptide chain to fold into the correct

native conformation is contained within the amino acid sequence. As a result, studies of the protein folding problem have been dominated by approaches in which the native structure is perturbed, and the unfolding/refolding processes characterized. From the knowledge generated by these studies, it is apparent that a general folding mechanism must allow for unfolded, intermediate and native states, and both thermodynamic and kinetic contributions to stability.

Protein folding studies have also tended to focus on single domain proteins (3-5) that appear to conform to a two-state folding mechanism, and thus are generally able to be reversibly unfolded. However, an interesting opportunity is presented by examining the unfolding and refolding of proteins that are inherently unable to refold to their native conformation – that is, they are unable to refold not due to extraneous circumstances (such as aggregation), but possibly due to a lack of folding competence within the primary sequence. Proteins that do not appear to obey Anfinsen's rule (6) and are unable to refold to the native structure are useful models for the study of folding intermediates. Many members of the peptidase family of enzymes fit this latter description (7-10).

The following thesis seeks to examine the application of small-angle neutron scattering and related analysis methods to the study of protein folding. In doing so, complementary solution structure techniques were used to characterize various folded states of pepsin, an aspartic peptidase believed to be irreversibly denatured.

2. Literature Review

2.1 Protein Folding and Thermodynamics

The underlying principle widely believed to be responsible for how polypeptides are able to fold into specific, complex protein structures is described by the *thermodynamic hypothesis* (6), which states that, under a certain set of defined environmental conditions, the native conformation of a polypeptide chain will be “the one in which the Gibbs free energy of the whole system is lowest.” The thermodynamics of protein folding are generalized in the following equation:

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

Thus, the folding of a protein is a compromise between (i) the heat released upon formation of bonds (i.e., hydrogen bonds and van Der Waals contacts), the change in enthalpy (ΔH) and (ii) the loss of conformational entropy upon folding of the polypeptide (ΔS). In general, the native, folded state of a protein is only marginally stable compared to the unfolded state, with ΔG_{fold} values typically on the order of -21 to -42 kJ/mol (11). This indicates that the addition or loss of interactions can substantially affect the balance between folded/unfolded states.

Interactions with the solvent also need to be considered, as water molecules are believed to form ordered arrays (12) around exposed non-polar residues (lower relative entropy), such that upon burying of non-polar groups (i.e. folding), the water molecules

are free to assume other configurations (13). It has also been shown that the translational entropy of water molecules increases in proportion to the decrease in excluded volume upon protein folding, and that this factor may compete with the protein conformational entropy gain of unfolding (14). Thus, there may be an increase in solvent entropy upon folding.

A potentially useful parameter in describing protein-solvent interactions is the partial heat capacity, C_p and its change upon protein unfolding, ΔC_p . The unfolding of proteins is most often observed to occur with an increase in C_p . The transfer of non-polar groups into water was shown to increase the heat capacity, while exposure of polar groups was shown to have the opposite effect, decreasing the heat capacity (15, 16). Thus, the positive ΔC_p upon protein unfolding is believed to be a result of exposure of non-polar groups, and as such, is taken to indicate the importance of hydrophobic interactions in protein folding (17). The C_p for the native folded state is proposed to consist of three factors (18), the most significant being (i) a covalent/primary structure factor (85%), followed by (ii) a hydration contribution (15 %) and finally (iii) a non-covalent, secondary/tertiary structure factor (< 3 %). The increase in C_p upon unfolding is largely due to the increase in the hydration factor, which may account for ~ 40 % of C_p in the unfolded state (18). Values for the change in heat capacity upon hydration of non-polar residues were estimated to range from 2.24 (5 °C) to 1.66 (125 °C) J K⁻¹ mol⁻¹ Å⁻² (19).

While hydrophobic interactions and related solvent entropy effects are important to understand the stability of a protein, enthalpic factors likely dominate the stability of the folded state (19). However, the observation that some multisubunit proteins dissociate

at low temperature (cold denaturation) (20, 21) supports that hydrophobic interactions play a dominant role in protein stability in some cases.

The application of thermodynamics to protein folding requires that (i) the denatured state approximates the unfolded state, (ii) the transition is reversible (so that equilibrium criteria are met) and (iii) the transition is cooperative (two-state only) (19).

2.2 Protein Folding Models

The simplest model of protein folding is the two-state model, in which folding occurs as a reversible, cooperative transition between the only two possible states: native and unfolded. As mentioned above, a strict application of thermodynamic analysis to protein folding requires that the transition be two-state. However, while thermodynamic analysis is commonly employed, no protein actually folds according to a two-state mechanism, and protein folding can only be approximated as such. Essentially, rather than defining two discrete conformational states (native and unfolded), two non-overlapping regions of conformational space can be ascribed to the native and unfolded states. As most (22, 23) current methods are unable to resolve the conformation of a single molecule, data collected from native and denatured states actually reports on these two conformational regions anyway, and can usually be adequately fit as a simple two-state transition [see for example, (24)]. In terms of defining a protein folding mechanism, the two-state approximation is an oversimplification and more complicated models are required.

The importance of hydrophobic interactions in protein stability led to the hypothesis that these interactions may actually drive protein folding (13). Specifically, the protein initially becomes compact due to non-polar interactions, which allows the search for the native state to occur in a smaller conformational space, greatly speeding up the folding reaction (4).

A folding model that may reconcile both thermodynamic and kinetic factors was proposed, in which a protein folds through a series of kinetic intermediates (at least two), each successive intermediate being more native like. Native-like secondary structure elements are formed in the first intermediate. The second intermediate represents the formation of the tertiary structure fold (relative orientations of secondary structure elements), but lacking native tertiary contacts and rigidity. The formation of secondary structure first, followed by the folding pattern and then the final, rigid native structure is often referred to as the framework model (4, 5, 17).

It was from the second intermediate, or folding pattern level of structure, that the concept of a molten globule (MG) state originated (17). Evidence for the existence of the MG comes from the observation that many denatured proteins are not completely unfolded [see the following for reviews: (17, 25, 26)]. Criteria for the MG include a native-like compactness, native-like secondary structure, absence of stabilizing tertiary structure contacts, more hydrophobic surface area exposure, reduced cooperativity and enhanced dynamic fluctuations (17, 19).

A potential shortcoming of the framework model is that the proteins shown to fold according to this model are predominantly α -helical (27-29), as these structures have

local contacts to provide stability. The model becomes less suitable for β -strand structures, which require tertiary contacts (4).

A nucleation-condensation growth model has emerged, which is a combination of the framework and hydrophobic collapse models (4). The reasoning behind the model is that, as secondary structure is not very stable in isolation (30, 31), hydrophobic collapse and the formation of tertiary contacts stabilize the secondary structure elements.

Denatured and intermediate states are characterized by nucleation-growth sites, which are secondary structure elements stabilized by tertiary contacts. The native structure propagates outwards from these sites. Thus, the framework model becomes an extension of this model, if the proteins have unusually stable secondary structural elements [such as α -helices (32)]. Alternatively, the folding of a protein consisting of more hydrophobic content and less stable secondary structure elements could be driven mostly by the nucleation/condensation factor. In this way, the hydrophobic collapse and framework mechanisms are proposed to be opposite extremes of the same basic model of folding (4). Regardless of the exact energy landscape traversed during folding, or what number and type of intermediates are significantly populated, the nucleation-condensation folding process occurs as a search for the free energy minimum – thus for the native structure. If native-like interactions are strongest, the search is not random and becomes considerably hastened as it is guided by the local energy landscape, determined by the primary sequence (33).

2.3 Kinetic vs. Thermodynamic Control of the Folding Landscape

That many proteins can be reversibly folded has given wide support to the thermodynamic hypothesis and the folding mechanisms described above generally hold, as their basis, that the native conformation is at the global energy minimum. However, there are reasons to believe that protein folding is not solely under thermodynamic control and these have been reviewed (34). For example, in light of Levinthal's paradox, only a limited amount of the total available conformational free energy space could possibly be sampled by a protein in the relatively short time it takes to fold. Otherwise, it would take a protein far too long to randomly search the entire conformational space to find the global lowest energy state. As a result, the final folded state reached could be a local rather than the global energy minimum (34). According to Baker and Agard (1994), the limited conformational space occupied during folding could correspond to the kinetic pathway, resulting in a local or global (although this is less likely) energy minimum (34). However, the finding that proteins are often able to fold to the native state under widely varying conditions (e.g., in vivo and in vitro) may indicate that they do not fold according to a defined pathway, but rather to an energy minimum (17). Under varying conditions, kinetically driven folding would be expected to result in different final folded states (34).

Examples of kinetically stabilized proteins include many zymogen-derived enzymes. Proteases such as α -lytic protease (35, 36), subtilisin (37) and *S. griseus* protease B (38) all require the presence (covalent or non-covalent) of a prosegment in order to fold to the native state. The folding mechanism may be that the prosegment somehow lowers the free energy barrier to the native state, catalyzing the folding reaction

(34). For example, on its own, the α -lytic protease denatured state is stable for long periods, but rapidly (rate of folding increased by 1.7×10^7) folds to the native state upon addition of its prosegment. It was calculated that the prosegment lowered the folding energy barrier to the native state of α -lytic protease by 23 kcal/mol (from 27 to 4 kcal/mol) (36), and catalyzed the folding reaction with a second order rate constant of $k_{cat}/K_m = 1300 \text{ M}^{-1} \text{ s}^{-1}$ (35). Other examples of proteins found to be kinetically stabilized include non-zymogen derived proteins, such as plasminogen activator inhibitor PAI-1, *Influenza* hemagglutinin and luciferase [reviewed in (34)]. In addition, a recent study has suggested that the native state of a mouse prion protein is thermodynamically metastable. After unfolding the prion protein using urea, dilution resulted in the immediate formation of the native, predominantly α -helical state. However, after incubation for five weeks (at room temperature with 10 or 30 μM of protein in a pH 3.6 buffer, 0.2 M NaCl) a non-native, β -sheet state formed that was found to be thermodynamically more stable (39).

2.4 Pepsin-like Aspartic Peptidases

Pepsin has historically (40) been examined as an archetype of the aspartic peptidases (APs), or more specifically, the A1 family of APs. This family includes pepsin-like enzymes originating from mammalian, fungal, protozoan, plant and viral sources (41). APs are a subclass of a broader group of enzymes known as peptidases, and are named as such because they contain two aspartic acid residues in the catalytic centre. APs catalyze the cleavage of peptide bonds – a function so fundamental to virtually all organisms that AP enzymes are at the centre of many areas of study. For example, APs

are of importance to industrial, physiological and pathological processes. APs can be used in industrial processes/products such as in detergents, meat tenderizing (42), production of cheese (43, 44) and other dairy products (45, 46). In the body, APs are involved in digestion, controlling cell death, regulating blood pressure (47), and are central to pathologies such as Alzheimer's disease (48), malaria (49) and HIV (50).

Most peptidases known to date are naturally produced as zymogens, containing a prosegment domain that is removed from the main enzyme body upon activation. Although the nature of the prosegment function is not entirely understood, it is evident that many peptidases require the presence of the prosegment in order to fold to the native structure (51).

2.5 Pepsin structure

The crystal structure (4PEP) of pepsin (52) is shown in Figure 1. Native pepsin has a bilobal conformation with two catalytic Asp residues (Asp 32 and Asp 215) on either side of the active site cleft, formed between the N- and C-terminal domains. The structure is predominantly β -strand (~ 50 %) and random coil (~ 38 %) with limited α -helical segments (~ 12 %). The two domains are connected via a six-stranded β -sheet plate, which may be viewed on its own as a third domain (41). Regardless of whether a two- or three-domain classification is used, the N- and C-lobes are similar in structure such that it is hypothesized that pepsin and other APs originated by gene duplication (53)

Pepsin is naturally synthesized by the parietal cells lining the stomach and folded at neutral pH as a zymogen containing a 44 residue prosegment. As shown in Figure 2,

the prosegment lies between the N- and C-terminal domains and sterically blocks the active site cleft in pepsinogen (54). Through a variety of mechanisms (55), pepsinogen is capable of autoactivation at acidic pH to yield the mature enzyme, pepsin, of approximately 34.6 kDa (56). As a gastric enzyme, pepsin displays optimum activity in the pH range from 1—4 (57), but has minimal activity above pH 5.

Different mechanisms of AP catalyzed peptide hydrolysis have been proposed (58, 59), which are generally quite similar but differ in the exact nature of H-bonding and proton transfer from the catalytic residues. Some proposals place a shared proton between Asp32 and Asp215 (59), while others argue that the molecular geometry observed in high-resolution structures suggests that all H-bonds are mediated through a molecule of H₂O, not directly between the Asps (58, 60). In any case, the oxygen of the bound H₂O acts as a nucleophile, attacking the carbonyl carbon of the substrate peptide bond. Recent data (x-ray and neutron diffraction and NMR) support (60) that in the resulting transition state complex formed, Asp215 is protonated while Asp32 is deprotonated. The nitrogen of the leaving group of the scissile peptide bond, NH-R, is then protonated either by Asp215 or the solvent, resulting in hydrolysis (58).

2.6 Studies on the Irreversible Denaturation of Pepsin

Pepsin is believed to be incapable of refolding to the native state upon denaturation. In contrast, pepsinogen, the inactive, zymogen form of pepsin, can be reversibly unfolded by means of urea or alkaline pH (61, 62). The irreversibility of pepsin unfolding has been the subject of a number of studies in the last few decades, but remains

Figure 1. Crystal structure of pepsin (PDB code: 4PEP). Ribbon representation illustrating the bilobal structure of the N- (left side) and C-domains (right side), which form the active site cleft. Asp32 and Asp215 side chains are coloured green, while the backbone is coloured according to secondary structure (yellow – strand, red – helix, black – other).

prosegment	mature pepsin
L1p – L44p	I1 – A326

Figure 2. Structure of pepsinogen, the zymogen form of pepsin. (Top) Mature pepsin consists of 326 residues and is derived from the zymogen form after cleavage of a 44 residue prosegment. (Bottom) The crystal structure of the inactive, zymogen form of pepsin – pepsinogen (3PSG). The N-terminal, 44 residue prosegment (blue) is bound across the active site cleft. The side chains of Lys36p and Tyr37p are also shown H-bonding (dashed lines) with the catalytic Asp32 and Asp215 residues. Orientation and colour are the same as in Figure 1.

unsolved. Research by Amad and McPhie indicated that neutral pH denatured pepsin (in a covalently inhibited form) fully retained its disulfide linkages and furthermore, that a large amount of ordered structure persisted as further unfolding transitions could be brought about by addition of Gdn-HCl (63) and also by urea (64). A calorimetric study by Privalov et al. suggested that the N- and C-terminal domains are independent folding units, with the C-terminal domain being much more stable and capable of reversible unfolding (65). It was later shown that only the C-terminal domain of neutral-pH denatured pepsin (termed Ip) could form an active complex when combined with its recombinant partner domain (66). By introducing stabilizing mutations into the N-terminal domain, Tanaka and Yada (67) found that the rate of inactivation at pH 7.0 could be substantially reduced relative to the wild-type inactivation; furthermore, by incorporating a disulphide bridge between the N-terminus and the bottom β -sheet plate (G2C/L167C), complete alkaline denaturation of pepsin could be prevented even after many hours (67). Thus, prevention of the dissociation of the N-terminal lobe was related to the prevention of denaturation. More recent structural investigations, notably results from ^1H -NMR, suggest that while the N-terminal domain contains relatively little organized structure, the C-terminal domain retains a compact and rigid, yet non-native, organization upon neutral-pH denaturation (68, 69).

It has been established that neutral-pH denatured pepsin (Ip) is partially unfolded (63, 64, 68, 69), with the N-terminal domain undergoing more extensive loss of native structure compared to the C-terminal domain (68, 69). The complete loss of activity and partial loss of secondary structure upon alkaline denaturation of pepsin has been fairly

well documented (63, 64, 68-71), but there have been comparatively few studies of tertiary structure of partially unfolded pepsin (68) as well as renatured conformations.

There has only been one small-angle scattering study characterizing the tertiary structure of pepsin in its native (Np) and Ip states (68). Previous small-angle x-ray scattering results (68) showed that the tertiary structure of pepsin changed on denaturation, becoming extended, but no measurements were made to determine whether or not the tertiary structure of renatured pepsin more closely resembles the denatured or native form.

2.7 SANS and the Study of Protein Folding

Numerous techniques can be used to examine protein folding, each with its own unique advantages and information acquired (Table 1). However, while many of the techniques indirectly report on the tertiary structure, only the small-angle scattering methods (SANS and SAXS) are able to directly measure the tertiary/quaternary protein structure of unfolded states. SAS is a method whereby the rotationally averaged, low-resolution ($r_{\max} \sim 5\text{--}10 \text{ \AA}$) structure of macromolecules in solution can be measured (75, 76). The low-resolution, spherically averaged nature of small-angle scattering data makes SAS a potentially ideal method for examining the higher orders of structure of various folded states, from compact and native to completely unfolded proteins. Using SAXS to study protein structure generally requires the use of synchrotron radiation, which can pose degradation problems for the sample, but the high flux can offer short measuring times and good counting statistics (77, 78). SANS offers the advantage that changing the

Table 1. Techniques for studying protein folding and structure^a.

Technique	Information obtained
Fluorescence	
<i>Intrinsic</i>	Environment of Trp and/or Tyr residues
<i>ANS binding</i>	Hydrophobic surface area exposure
<i>Fluorescence Resonance Energy Transfer (FRET)</i>	Distances between labels on a protein
Site-Directed Spin Labelling (SDSL) with electron paramagnetic resonance (EPR) Spectroscopy ^b	Environment, mobility and distance between labels
Circular dichroism	
<i>Far-UV</i>	Secondary structure
<i>Near-UV</i>	Environment of aromatic groups – assessment of tertiary structure
FTIR	Secondary structure
UV absorbance	Environment of aromatic groups
Real-time NMR	Residue environment
Hydrogen-exchange MS	Protection of H atoms – population of folding states
Atomic Force Microscopy (AFM)	Force required to cause unfolding
Small-angle scattering (SAXS and SANS)	Protein size, shape and protein-protein interactions
Differential Scanning Calorimetry ^b	Protein stability, enthalpy and heat capacity changes of unfolding/folding
Protein Engineering	Effects of residues on folding and stability
Enzyme kinetic activity	Active site structure

^a Table adapted from (5, 72).

^b Information on SDSL-EPR and DSC can be found in (73) and (74), respectively. References for the other techniques can be found in (5, 72).

composition of $^1\text{H}/^2\text{H}$ in the sample can dramatically alter the contrast between the analyte and the solution (79, 80).

With SANS, neutrons which are scattered elastically ($\Delta \text{energy} = 0$) and coherently ($\Delta \text{phase difference} = 0$) from a dilute, isotropic solution of (ideally non-interacting) protein molecules are detected on a two-dimensional area detector as a function of scattering angle, 2θ (79). Converting this data to a one-dimensional plot of scattered intensity as a function of scattering angle, information on the size and shape, and thus tertiary or quaternary protein structure is obtained.

While there are relatively fewer reports, SANS has been used to study the structure of denatured states. The scattering data from proteins under strong denaturing conditions, obtained using chemical denaturants, has been interpreted in terms of size analysis and polymer theory (24, 81, 82). For example, by modelling the scattering data from phosphoglycerate kinase using atomic coordinates, fixed into different extended chain-of-spheres models, changes in the scattered intensity profile were related with changes in the backbone torsion angles via Ramachandran plots (83). The chemically unfolded form of neocarzinostatin, a small globular protein, was shown to differ from the heat unfolded form as the scattering curve from the chemically denatured state was more closely approximated by an excluded volume chain while the thermally denatured state was better described as an ideal chain (84). The folding of very large proteins can be measured by SANS, as was done in the study of fibronectin ($M_w = 530 \text{ kDa}$), a multidomain, extracellular matrix protein. Under various denaturing and potentially renaturing conditions, fibronectin could accurately be modelled using a chain of 56

spheres ($r = 25 \text{ \AA}$) as the conformation alternated between globular, swollen-coil and open-chain states (85).

The present thesis examined the use of SANS and complementary techniques (CD, DSC, fluorescence, densitometry) to the study of protein folding. The model folding system was chosen to be the irreversible denaturation of pepsin. Under different conditions of pH and protein concentration, the structure of pepsin under native (pH 5.3), denaturing (pH 8.0) and potentially renaturing (taken from pH 8 to pH 5) conditions was examined.

3. Materials and Methods

3.1 Materials

High purity porcine pepsin A (EC 3.4.23.1) was purchased from Sigma (St. Louis, MO) and used without further purification. The peptide substrate KPAEFF(NO₂)AL was synthesized at the Core Facility for Protein/DNA Chemistry located at Queens University (Kingston, ON). Deuterium oxide (D₂O, 99.9%) was purchased from ACP Chemicals (Montreal, QC). Other chemicals were of reagent grade.

3.2 Sample Preparation

Stock solutions of pepsin samples of known concentration [determined by mass and/or A_{280} , where $E_{280} = 38,600 \text{ M}^{-1} \text{ cm}^{-1}$ (68, 86)] were prepared and subsequently

diluted in known volumes of the appropriate buffers. Pepsin was examined under three different conditions of pH using the following buffers: (a) 20 mM sodium acetate/acetic acid, pH 5.3, (b) 20 mM Tris/HCl, pH 8.0, and (c) 1:1 mixture of 10 mM Tris/HCl, pH 8.0 and 40 mM sodium acetate/acetic acid, pH 4.7, with a final pH of 5.

In the preparation of pepsin in buffer (c), a quantity of pepsin was first dissolved in one part 10 mM Tris/HCl, pH 8.0, incubated at 21°C for at least ten minutes, then further diluted with one part 40 mM, sodium acetate/acetic acid, pH 4.7. Refolded samples for CD, DSC and fluorescence experiments were prepared from alkaline unfolded (pH 8.0) pepsin samples ranging in concentration from 0.4 – 0.8 mg/ml (CD and fluorescence), from 0.8 – 1.6 mg/ml (DSC), and 5 mg/ml (SANS). These samples are collectively referred to as refolded pepsin, or Rp. For SANS, CD and DSC experiments, refolded samples were also prepared from a 20 mg/ml alkaline unfolded pepsin solution. These samples are referred to as concentrated refolded pepsin, or CRp.

3.3 Kinetic analysis

Kinetic measurements were taken using the octapeptide substrate KPAEFF(NO₂)AL (87) in which initial reaction rates ($-\Delta A/\text{min}$) were measured as the linear decrease in absorbance at 300 nm using a DU640 spectrophotometer (Beckman Instruments, Fullerton, CA). At least ten substrate concentrations were employed, in the range 0.002 – 0.16 mM octapeptide, with 10 nM pepsin. Initial reaction velocity data were plotted versus substrate concentration and fitted according to the Michaelis-Menten model using least-squares non-linear regression analysis. Measurements were repeated at

least in triplicate for each substrate concentration and constants are reported as the mean $\pm$ SD as determined from multiple fits to the Michaelis-Menten model.

3.4 Small-Angle Neutron Scattering

3.4.1 SANS Data Collection

SANS measurements were performed using the 30 m NG3 instrument located at the National Institute of Standards and Technology (NIST, Gaithersburg, MD). Three sample-to-detector distances, 13 m, 5 m, and 1.5 m were used with 5.5 Å ($\Delta\lambda/\lambda = 0.12$) neutrons to obtain a total range in scattering vector $0.00637 < q < 0.4272 \text{ Å}^{-1}$, where the scattering vector $q = 4\pi\sin(\theta/2)/\lambda$.

For SANS measurements, pepsin samples were prepared in both H₂O and D₂O buffers and diluted to a final concentration of 10 mg/ml (H₂O and D₂O), 5 mg/ml or 2.5 mg/ml (D₂O only). Protein solutions were then injected into cylindrical cells with 1 and 2 mm path lengths for H₂O and D₂O solutions, respectively. SANS measurements, typically lasting ~ 3 h per sample, were performed immediately after dilution of the protein.

3.4.2 SANS Data Analysis

Data reduction consisted of correction for instrument response, empty cell scattering, incoherent scattering background, and scaling to absolute intensity based on

the incident beam flux. The data reduction procedures were carried out using the software Igor Pro 4.05, provided by NIST. Mean and standard deviation values for the scattering intensity were determined by averaging the scattered intensity within rings of distance r from the beam centre on the two-dimensional detector.

Radii of gyration R_g were obtained from the SANS data using the Guinier approximation (88), in which the solution was assumed to be ideal (no solute interactions) and monodisperse (all particles are identical). In the small angle limit, for values of q such that $R_g^2 q^2 < 1$ (approaching $q = 0$), the intensity was approximated as:

$$I(q) \approx I(0)\exp(-R_g^2 q^2/3) \quad (2)$$

A plot of $\ln[I(q)]$ versus q^2 gave R_g from the slope of the linear regression, and was performed using Igor Pro 4.05.

At large scattering vectors ($q \rightarrow \infty$), the limit of the scattering intensity $I(q)$ varies as q^{-2} for a Gaussian chain, while for a globular object $I(q)$ varies as q^{-4} (from Porod's law for the asymptotic behaviour of scattering intensity), such that the scattering curves of globular (bell shape) or unfolded proteins (plateau) can usually be differentiated. To this end, the SANS data were plotted as $q^2 I(q)$ versus q , in what is commonly known as a Kratky plot (79, 89).

SANS data were further analyzed using several programs freely available from the Biological Small Angle Scattering Group (European Molecular Biology Hamburg Outstation, Hasylab). In particular, SANS data were compared to crystal structure predictions using CRYSON (90), pair-distance distribution functions $P(r)$ were calculated

using GNOM (91), low resolution dummy residue models were produced using GASBOR (92), and low resolution model scattering envelopes were calculated using SASHA (93).

The program CRYSON (90) calculates a SANS profile from the crystal structure of a protein. The program can also be used to fit small-angle scattering data by combining the calculated SANS profile with a varying thickness or density of a bound water layer and level of incoherent background. The reduced and normalized scattering data were transformed from reciprocal space ($I(q)$), to real space using GNOM (91). This program calculates a pair distance distribution function, $P(R)$ that corresponds to the distribution of distances between different scattering points, R_{ij} . Thus, $P(R)$ is sensitive to shape and $R_{\max}$ gives an estimate of the maximum dimension of the molecule.

Starting from the experimental $P(R)$ and the number of residues in the sequence (326 for pepsin), the program GASBOR (92) calculates a chain-like distribution of dummy residues ($d = 5 \text{ \AA}$) to approximate the scattering data. Utilizing simulated annealing, a conformation retaining a chain-like distribution of dummy residues and yielding the best fit to the scattering data is obtained. The GASBOR fitting routine was performed five times for each sample data set and the resulting models were averaged using DAMAVER (94) to produce the final model.

As an alternative to the simulated annealing procedures, SASHA (93) was used to calculate low resolution, model scattering envelopes directly from the $I(q)$ data. This program calculates angular envelope functions using spherical harmonic approximations to a real-space representation of the protein scattering length density.

3.5 Densitometry

The densities of pepsin solutions and buffers were measured via the mechanical oscillator technique (95, 96) with the use of two Anton Paar DMA 602 external cells using a DMA 60 measuring unit in the phase lock loop mode (A. Paar, K. G., Graz, Austria). The temperature of the external cells was controlled by circulating thermostated fluid from a Neslab RTE-4 constant temperature bath. Instrument calibration and the determination of the densities of the various solutions were performed as described in (95, 96). The partial specific volume (v) of pepsin at the various pH conditions was determined via:

$$v = [1 - (\rho_p - \rho_b) / c] / \rho_b \quad (3)$$

where v is the specific volume of the protein, ρ_p is the density of the protein suspension, ρ_b is the density of the buffer and c is the mass concentration of protein in solution. Measurements on each solution were performed at least in triplicate and the results presented correspond to the mean and standard deviations of each condition.

3.6 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectra were recorded using a Jasco J-600 spectropolarimeter (Japan Spectroscopic, Tokyo, Japan). Data were recorded in the far-UV range from 250 – 190 nm with a 1 nm bandwidth, 100 nm/min scan rate, 0.25 s

response time and $4 \times$ accumulation. In each case, 0.2 mg/ml of pepsin solution was analyzed in a cuvette with a 0.1 cm pathlength and a minimum of three measurements were made per sample. Prior to each pepsin sample run, the corresponding buffer alone was run and was subtracted from the subsequent pepsin spectra before further analysis. All solutions were filtered and degassed before measurement. Ellipticity values recorded as a function of wavelength (λ) were converted into mean residue ellipticity $[\theta]_{\text{mrw},\lambda}$ using:

$$[\theta]_{\text{mrw},\lambda} = MRW \times \theta_{\lambda} / 10 \times d \times c, \quad (4)$$

where MRW is the mean residue weight [107 g/mol for pepsin (70)], θ_{λ} is the measured ellipticity at a particular wavelength (degrees), d is the pathlength (0.1 cm) and c is the protein concentration (g/cm^3). Secondary structure proportions were estimated using three different algorithms [SELCON3 (97), CONTINLL (98) and CDSSTR (99)], each with two different protein reference sets (set 4 and 7, where set 7 contains data from four denatured proteins) via the DICROWEB server (100, 101). Results from estimates that gave a normalized root mean square deviation from the experimental data of less than 0.25 were accepted (102) and averaged to give the final estimate of secondary structure.

3.7 Differential Scanning Calorimetry

Calorimetry experiments were carried out with a MicroCal VP-DSC (MicroCal, Northampton, MA) using a scan rate of $1.5^{\circ}\text{C}/\text{min}$ over a temperature range from 10 –

110°C, 16 s filtering period (time delay), and passive feed-back response. The final concentration of pepsin samples used was ~ 0.4 mg/ml, with the corresponding buffer used as the reference. Data reduction and analysis was performed using MicroCal Origin v7.0 (OriginLab, Northampton, MA). Curve-fitting with Origin was carried out as follows, starting from the total molar heat capacity of the system, C_p , consisting of independent transitions A , B , etc...

$$C_p = C_{pN} + \left[\frac{K_A \Delta C_{pA}}{1 + K_A} + \frac{K_A \Delta H_A^* \Delta H_A}{(1 + K_A)^2 RT^2} \right] + \dots \quad (5)$$

For a two-state transition, the van't Hoff enthalpy ΔH^* is assumed to be equal to the calorimetric ΔH , and if the heat capacity of the native state, C_{pN} , is assumed to be linear with temperature ($C_{pN} = B_0 + B_1 T$), then Equation (5) becomes:

$$C_p(T) = B_0 + B_1 T + \left[\frac{K_A(T) \Delta C_{pA}}{1 + K_A(T)} + \frac{K_A(T) \Delta H_A(T)^2}{(1 + K_A(T))^2 RT^2} \right] + \dots \quad (6)$$

Making the following substitutions,

$$\Delta H_A(T) = \Delta H_{mA} + \Delta C_{pA}(T - T_{mA}) \quad (7)$$

and

$$K_A(T) = \exp \left\{ \frac{-\Delta H_{m,A}}{RT} \left(1 - \frac{T}{T_{m,A}} \right) - \frac{\Delta C_{p,A}}{RT} \left(T - T_{m,A} - T \ln \frac{T}{T_{m,A}} \right) \right\} \quad (8),$$

the Origin software calculates $C_p(T)$ with the user input of the number of transitions and values (or initial guesses) for $T_{m,A}$, $T_{m,B}$, etc. Estimations of the fitting parameters, B_0 , B_1 , $\Delta C_{p,A}$, $\Delta H_{m,A}$ and $T_{m,A}$ are made until no improvement in the fit of $C_p(T)$ to the experimental curve, $C_p(T)_{\text{exp}}$ is obtained.

In fitting a non-two state transition, $\Delta C_{p,A}$ is assumed to be zero and a baseline must first be subtracted, then Equation (5) can be expressed as:

$$C_p(T) = \frac{K_A(T) \Delta H_{m,A}^* \Delta H_{m,A}}{(1 + K_A(T))^2 RT^2} + \dots \quad (9)$$

where the equilibrium constant $K_A(T)$ is given by :

$$K_A(T) = \exp \left\{ \frac{-\Delta H_{m,A}^*}{RT} \left(1 - \frac{T}{T_{m,A}} \right) \right\} \quad (10)$$

In a similar fashion as described above for fitting a two-state transition, Origin was used to estimate values for $T_{m,A}$, $\Delta H_{m,A}$ and $\Delta H_{m,A}^*$.

Estimates for an apparent Gibbs free energy of unfolding (ΔG_{unf}) were calculated using:

$$\frac{\Delta G}{T} = \Delta H(T_M) \left(\frac{1}{T} - \frac{1}{T_M} \right) + \Delta C_p \left(\ln \frac{T_M}{T} - \frac{T_M}{T} + 1 \right) \quad (11)$$

where values for $\Delta H(T_m)$, ΔC_p and T_m were obtained by fitting the data as a two-state transition.

3.8 Intrinsic Tryptophan Fluorescence

Fluorescence spectra were recorded using a Shimadzu RF-540 spectrofluorophotometer (Shimadzu Corporation, Kyoto, Japan) with a 1-cm quartz cell at room temperature and the following settings: excitation at 295 nm (the optimal excitation wavelength for measurements of Trp fluorescence), emission scan from 305 – 450 nm, and excitation and emission slit widths at 5 nm. All samples consisted of 0.20 mg/ml pepsin (5.8 μ M) and were measured at least in triplicate.

4. Results

4.1 Pepsin Behaviour in H₂O and D₂O.

Due to the strong incoherent scattering nature of hydrogen in H₂O, SANS data were collected from pepsin solutions made up with D₂O to improve the signal-to-noise ratio and contrast between the protein and solvent. As some proteins are observed to

display anomalous behaviour in pure D₂O (103-105), it was necessary to show that pepsin retained native activity, and thus native conformation, in a deuterated solution.

To this end, enzyme kinetic data were collected from pepsin in H₂O and D₂O buffered solutions (pH 5.3). The activities followed Michaelis-Menten kinetics for both conditions (Figure 3), and comparison of the kinetic parameters indicated that pepsin retained native activity in deuterated solvent (results shown in Table 2).

Kinetic data were collected at initial substrate concentrations ([S₀]) in the range 0.002 – 0.16 mM octapeptide, which resulted in the [S₀] range covering approximately 0.04—3.4 K_m (aqueous solution) and 0.09—7.3 K_m (deuterated solution). As the [S₀] range used should be sufficient to avoid bias in the derived parameters (106), the estimates for V_{max} and K_m could be considered to be accurate.

While the V_{max} (and thus k_{cat}) and K_m values for pepsin in D₂O solution were significantly lower ($p \leq 0.05$) than for pepsin in H₂O, the specificity constants (k_{cat} / K_m) were not significantly different ($p > 0.05$) for the two conditions. This result suggested that pepsin retained its native conformation in D₂O solution. The solvent effect on V_{max} is not entirely unexpected and kinetic isotope effects are commonly studied (107, 108). During the course of substrate-peptide hydrolysis, O-¹H and conversely O-²H bonds are broken – with the activation energy E_a of the later chemical reaction being considerably greater which could be expected to result in a decrease in the reaction rate [refer to (109, 110) for an analysis of isotope effects on enzyme kinetics].

Figure 3. Kinetic activity of pepsin in H₂O and D₂O. Initial reaction rate versus substrate concentration for cleavage of a synthetic substrate by pepsin, in H₂O (upper data) and D₂O (lower data), fitted with a Michaelis-Menten model (bold curve, H₂O; thin curve, D₂O). Error bars represent the standard deviation of at least three measurements at each substrate concentration.

Table 2. Kinetic constants determined for pepsin in H₂O and D₂O. Samples were made up in a 20 mM acetic buffer at pH 5.3, consisting of either H₂O or D₂O as the solvent.

Solvent	$V_{\max}$ ($\mu\text{M}\cdot\text{min}^{-1}$)	K_m (μM)	k_{cat} (s^{-1})	k_{cat} / K_m ($\text{s}^{-1}\cdot\mu\text{M}^{-1}$)
H ₂ O	44 ± 9	47 ± 14	73 ± 16	1.6 ± 0.6
D ₂ O	31 ± 4	22 ± 6	51 ± 6	2.3 ± 0.7

Calorimetric measurements showed that the heat capacity profiles of native pepsin in the two solvents were comparable, although the melting transition in D₂O ($67.61 \pm 0.19^\circ\text{C}$) was significantly ($p \leq 0.05$) greater than that in H₂O ($62.16 \pm 0.29^\circ\text{C}$). This implied that D₂O had a stabilizing effect on pepsin, as has been observed for other proteins (24, 111). For pepsin under unfolded, dilute and concentrated refolded conditions, the heat capacity profiles did not change substantially with isotopic conditions. While the absolute intrinsic fluorescence of the Trp residues of pepsin was higher in D₂O, compared to H₂O, the spectral changes that accompanied the structural changes of pepsin were similar in the two solvents. The secondary structure and changes in structure revealed by CD appeared to be independent of isotopic composition of the medium. SANS results indicated, indirectly by comparison with the crystal structure, and directly by comparison of the scattering from pepsin in both solvents, no difference between pepsin conformation in H₂O and D₂O.

4.2 Small-Angle Neutron Scattering

4.2.1 SANS Data Reduction

The steps carried out in data reduction are indicated in Figure 4, using a 5 mg/ml Np sample as an example. The scattering from the empty sample cell and from the background (taken with the beam blocked) were first subtracted from the sample data. Secondly, the data were corrected for detector efficiency using the isotropic scattering of a standard (vanadium), which accounted for non-uniformities in the multiwire detector.

Thirdly, the data were scaled to absolute intensity using a scaling factor calculated from the total neutron beam flux incident on the sample area. Finally, data around the detector edges and the beamstop were ignored by masking, and the remaining data circularly averaged to obtain 1 D data sets.

Figure 5 shows the 1 D scattering curve for 5 mg/ml Np, in which data sets collected at 5 m and 1.5 m have been combined. The final step was to subtract an incoherent background, largely due to incoherent scattering from the solvent. In doing so, either the scattering from buffer only, or a line fitted to the highest- q , linear portion of the sample scattering curve, could be subtracted to obtain the final data set.

4.2.2 Comparison of SANS Results to the Crystal Structure

The program CRYSON (90) was used to calculate a theoretical $I(q)$ profile from the crystal structure of pepsin (5PEP), and to compare the calculated and experimental curves. In fitting the theoretical SANS profile to those obtained experimentally, the incoherent background and thickness of bound water layer in the model function were fixed to be zero.

In Figure 6, the predicted scattering curve is shown plotted against the experimental results for Np, Ip, Rp and CRp. The experimentally measured SANS curve for Np clearly matched that predicted from the crystal structure. When pepsin was in its denatured form at pH 8, the measured SANS curves followed the predicted SANS profile at low q , but diverged above $q \approx 0.1 \text{ \AA}^{-1}$, demonstrating that denaturation of pepsin led to a substantial change in the tertiary structure. When the denatured protein was diluted

from pH 8 back to pH 5 under dilute protein concentration ($c < 10$ mg/ml), comparison between the experimental and predicted scattering curves showed that pepsin did not recover its native tertiary structure. In comparison, pepsin refolded at higher concentration ($c > 10$ mg/ml) was characterized by a native-like tertiary structure.

From the SANS curves in Figure 6 it was also evident that pepsin in D₂O showed enhanced forward scattering, or an upswing in the intensity at low q , indicative of either protein aggregates or protein-protein interactions (112, 113). Figure 7 shows plots, at each pH condition, of the scattering curves obtained in H₂O superimposed over those taken in D₂O. Comparison between the scattering curves for each pH condition showed that, for $q \geq 0.03 \text{ \AA}^{-1}$, the scattering curves from pepsin in D₂O were similar to those in H₂O. This demonstrated that, in this q -range, the influence of isotope (i.e., as it effects aggregation or the structure factor) on the protein conformation (form factor) was negligible. Rp was also examined in pure H₂O, but this data is not shown because the coherent signal was too low with respect to the incoherent background to resolve.

4.2.3 Pepsin Radii of Gyration via Guinier Plots

Figure 8 shows Guinier plots and the corresponding fits to the scattering data from Np, Ip, Rp and CRp, in D₂O. The R_g obtained from the fits are given in Table 3. Denaturation of pepsin at pH 8 led to an increase in R_g , while dilution of alkaline denatured pepsin back to pH 5 returned R_g to its original value. Notably, R_g for Ip was found to be greater at 5 mg/ml protein concentration than at 10 mg/ml. It is likely that R_g

Figure 4. Reduction of raw SANS data to produce background and instrument corrected, absolute-scaled data.

Figure 5. Averaging and subtraction of incoherent background from SANS data.

Reduced SANS data were circularly averaged and then data sets collected at different instrument configurations (5m s-d and 1.5m s-d) were combined. These combined sets (hollow square/top curve) were corrected for incoherent background scattering (solid line) to produce the final data set (solid square/bottom curve).

Figure 6. Comparison of experimental and theoretical SANS data. SANS curves for Np (pH 5.3), Ip (pH 8.0), Rp (pH 8→5) and CRp (pH 8→5) in D₂O buffer. Both the experimental (*square*) and theoretical (calculated from the crystal structure, 5PEP) (*solid red curve*) data are shown. Curves are offset along the y-axis for clarity. Data points and error bars correspond to mean and standard deviation from circular averaging.

Figure 7. Comparison of SANS data from pepsin in D₂O and H₂O. SANS curves for pepsin in D₂O (black symbols) overlaid with those in H₂O (grey symbols) at pH 5.3 (top), 8.0 (middle) and 8→5 (bottom). Data points and error bars correspond to mean and standard deviation from circular averaging.

determined at 10 mg/ml was underestimated due to the influence of the protein structure factor. Due to the higher protein charge at pH 8 than at pH 5 [~ -38 and -17 , respectively (114)], it was speculated that the protein structure factor for Ip would be more pronounced, and thus more sensitive to protein concentration than that for Np, Rp or CRp. It is proposed here that this was why a concentration effect on R_g was observed for Ip, but not for the Np and refolded states.

Table 3. Radii of gyration (R_g) obtained^a from Guinier approximation of the scattering data.

	c (mg/ml)	R_g (Å)	Fitting range	norm. χ^2
Np (pH 5.3)	10	19.5 ± 0.2	$0.99 < qR_g < 1.3$	1.2
	5	19.8 ± 0.2	$0.67 < qR_g < 1.1$	2.0
Ip (pH 8.0)	10	22.6 ± 0.3	$0.68 < qR_g < 1.1$	1.3
	5	29.3 ± 0.7	$0.81 < qR_g < 1.3$	1.2
Rp (pH 8 $\rightarrow$ 5)	2.5	19.9 ± 0.3	$0.55 < qR_g < 1.3$	1.3
CRp (pH 8 $\rightarrow$ 5)	10	19.4 ± 0.1	$0.70 < qR_g < 1.3$	0.93
	5	20.3 ± 0.2	$0.37 < qR_g < 1.0$	0.62

^aValues $\pm$ SD determined from single linear regression fits of SANS data.

Figure 8. Radii of gyration R_g determined by Guinier approximation in the low-angle region. Guinier fits to the SANS data for Np, Ip, Rp and CRp states, in D_2O . Data points and error bars correspond to mean and standard deviation from circular averaging.

4.2.4 Kratky Plots

A qualitative method for assessing protein structure, utilizing the asymptotic behaviour of the scattering function at high q , is possible by replotting the data as $q^2 I$ vs. q , commonly known as a Kratky plot (89). For globular proteins, the scattering decays as q^{-4} , while for proteins which are completely unfolded i.e., random coil, the scattering decays as q^{-2} . Kratky plots will show a plateau region at high q in the case of random coil structure, or a peak at finite q followed by a monotonic decrease in the case of globular structure (89). Examination of Kratky plots (Figure 9) of the SANS data for pepsin in its native, partially unfolded and dilute and concentrated refolded states indicated that the protein was globular in all four states.

4.2.4 The Pair Distance Distribution Function, $P(R)$

The pair-distance-distribution function, or $P(R)$, is obtained from a model independent transformation (115) of small-angle scattering data and essentially describes the scattering length density weighted correlation between mass elements within a protein (116). The $P(R)$ curves evaluated from the SANS data are shown in Figure 10. Again, the data in D_2O and H_2O (data not shown) were not substantially different; however, analysis of $P(R)$ curves obtained for pepsin in H_2O was not useful due to the poor resolution and additional noise in the case of H_2O . The $P(R)$ for Ip and Rp were found to differ from that for Np and extended out to a $D_{\max}$ of $\sim 100 \text{ \AA}$ and $\sim 72 \text{ \AA}$, respectively, while that for

CRp was essentially the same as the native form ($D_{\text{max}} \sim 60 \text{ \AA}$), consistent with the observations above.

4.2.5 Low Resolution Structure Via Simulated Annealing: GASBOR

Using the $P(R)$ profiles obtained by GNOM, it was possible, via simulated annealing fits, to obtain a low resolution dummy residue model representation of pepsin. Since a single structural solution to the SANS data obtained by this method need not be unique, the most generally accepted approach is to obtain a set of solutions from which an average is obtained (92). Since the SANS data obtained for pepsin in H_2O had a narrower range of usable q data, but showed no considerable difference from that obtained in D_2O , only the D_2O data were fit. In Figure 11 space filling representations of structures for Np, Ip, Rp and CRp are shown.

The overall shape of Np determined by GASBOR was found to be qualitatively similar to the crystal structure, with the bilobal and active-site cleft features partially resolved. However, the structure of Ip was found to be quite different, i.e., elongated, and nearly cylindrical. Again, consistent with the results presented above, it was shown that CRp had a tertiary structure that closely matched that of native pepsin, while that for Rp was intermediate between the Ip and Np forms.

4.2.6 Low Resolution Structure Via Scattering Envelopes: SASHA

While GASBOR utilizes dummy residues and simulated annealing to fit the $P(R)$ function determined by GNOM, SASHA fits the experimentally determined SANS profiles directly using spherical harmonics approximation. Thus, SASHA represents a complementary analysis method. In Figure 11, visual representations of the scattering envelopes determined by SASHA fits are shown alongside the results obtained using GASBOR, with the structures produced from both methods found to be in close agreement. Again, Np showed a compact, bilobal shape, Ip showed an extended, rod-like conformation and Rp appeared to recover some tertiary structure while CRp had a native-like conformation. The detailed, or small length-scale, features in the SASHA envelope functions, especially those of the Ip and Rp models, were considered with caution, for the following reasons. The generation of aberrant features by SASHA can occur due to the limited number of fitting parameters used in the spherical harmonic approximation. By increasing the number of parameters a smoother shape may be obtained, however, the experimental data set is usually too limited such that visibly different shapes can fit the experimental data equally well (76, 117). Also, unlike the many possible outcomes possible from GASBOR, which are averaged to obtain the most probable model, only a single solution is obtained from SASHA (93). However, the general shapes of the SASHA models were considered to be reliable because of their agreement with the GASBOR models.

With the models generated using both SASHA and GASBOR, the domain architecture, the relative positions of the N- and C-lobes, and thus the active site cleft formed between them, could be discerned (Figure 12). Although the models agreed well with the atomic structure, assignment of the N- and C-domains for the Np models could

Figure 9. Kratky plots, showing $q^2 I$ vs. q for Np, Ip, Rp and CRp in D_2O . The curves are normalized with respect to their maximum values. Data points and error bars correspond to mean and standard deviation from circular averaging.

Figure 10. Pair distance distribution functions, $P(R)$, for pepsin in D_2O . The Np and CRp curves overlap. Error bars correspond to the standard deviation estimated by GNOM through propagation of the experimental error from circular averaging.

Figure 11. Ab initio modelling of the SANS data. Low resolution (i.e., $r > 10 \text{ \AA}$) models of four states of pepsin, based on scattering data, calculated from (left side) a model using spherical harmonics approximation of the protein scattering density – SASHA, and (right side) a dummy residue model based on simulated annealing and fits of the $P(R)$ curve – GASBOR. Models are viewed from the side.

Figure 12. Domain structure indicated by low-resolution models. The active site cleft formed between the N- and C-lobes of Np could be identified in the models generated using both SASHA (left) and GASBOR (right). For comparison, the crystal structure of pepsin (5PEP; wireframe) overlays the low-resolution structures. From top to bottom, the views presented are from the side, top and end.

not be established conclusively, as they appeared similar at the low resolution. However, it was speculated that the most extended lobe of the Ip model corresponded to the N-domain. In the Rp state, it appeared that both domains were of similar conformation, being compact but non-native.

4.3 Densitometry

Densitometry measurements were performed to determine whether the conformational changes of pepsin were accompanied by substantial changes to its apparent partial specific volume, v_p . No significant difference ($p > 0.05$) was found among the Np, Ip and CRp states, which had partial specific volumes of 0.784 ± 0.040 , 0.782 ± 0.004 and $0.784 \pm 0.010 \text{ cm}^3/\text{g}$, respectively.

4.4 Circular Dichroism Analysis

In order to determine possible conformational changes on smaller length scales (i.e., changes to secondary structure), circular dichroism spectra were collected from pepsin under native, unfolded and two refolded conditions.

Examination of the CD spectra (Figure 13) clearly showed a shift in the minimum of the mean residue ellipticity (MRE) from $\sim 210 \text{ nm}$ to $\sim 204 \text{ nm}$ on going from pH 5.3 (native) to pH 8 (unfolded). The shift in MRE was qualitatively consistent with a loss of β -sheet structure in favour of random coil structure, as was observed in a previous study (70). The spectrum for Rp differed slightly from that of the Ip state, suggesting a small

portion of organized structure was reformed, whereas that for CRp matched closely with the native spectra suggesting a greater recovery of secondary structure. By estimating secondary structure proportions (Table 4) it was found that, while the α -helical and β -turn content remained more or less unchanged upon unfolding and refolding, there was a significant ($p \leq 0.05$) loss of strand and concurrent increase in unordered structure upon alkaline denaturation. Interestingly, either a partial or nearly complete reversal of this process was observed upon dilute or concentrated refolding, respectively.

Figure 13. Comparison of circular dichroism spectra from pepsin in four conformational states: (1) Np, pH 5.3 (*bold, solid line*) (2) Ip, pH 8.0 (*thin, lowest solid line*) (3) Rp, pH 5 (*dashed line*) and (4) CRp, pH 5 (*dotted line*). Protein concentration was 0.2 mg/ml in H₂O buffer.

Table 4. Secondary structure estimated from the CD spectra for Np (pH 5.3), Ip (pH 8.0), Rp (pH 5) and CRp (pH 5) in H₂O and D₂O. Values^a are given as the mean $\pm$ SD of six (or five where noted) separate estimates.

	Helix (%)	Strand (%)	Turn (%)	Random (%)
Np, H ₂ O	8.2 $\pm$ 2.9 _A	38.6 $\pm$ 4.9 _A	22.4 $\pm$ 1.4 _A	28.3 $\pm$ 3.6 _A
Ip, H ₂ O ^b	9.7 $\pm$ 2.3 _A	23.8 $\pm$ 3.8 _B	21.0 $\pm$ 4.5 _A	46.1 $\pm$ 10.4 _B
Rp, H ₂ O	9.9 $\pm$ 2.5 _A	30.8 $\pm$ 2.8 _C	21.4 $\pm$ 1.3 _A	37.4 $\pm$ 5.2 _B
CRp, H ₂ O	12.4 $\pm$ 7.1 _A	38.1 $\pm$ 0.6 _A	20.7 $\pm$ 3.9 _A	28.2 $\pm$ 3.6 _A
Np, D ₂ O	12.2 $\pm$ 6.3 _A	38.5 $\pm$ 0.4 _A	20.8 $\pm$ 3.6 _A	28.2 $\pm$ 3.3 _A
Ip, D ₂ O	6.6 $\pm$ 2.3 _A	28.6 $\pm$ 4.4 _{B, C, E}	21.1 $\pm$ 3.7 _A	43.2 $\pm$ 9.1 _B
Rp, D ₂ O ^c	9.2 $\pm$ 2.7 _A	32.5 $\pm$ 2.2 _{C, D, E}	21.9 $\pm$ 0.6 _A	35.5 $\pm$ 5.3 _{B, C}
CRp, D ₂ O ^c	8.9 $\pm$ 2.0 _A	33.8 $\pm$ 1.8 _D	22.3 $\pm$ 0.7 _A	35.0 $\pm$ 2.5 _C

^aMeans in each column sharing the same letter are not significantly different, $p > 0.05$.

^bFive estimates: SELCON3 – reference set 4 results omitted.

^cFive estimates: SELCON3 – reference set 7 results omitted.

4.4 Intrinsic Tryptophan Fluorescence

Pepsin contains five tryptophan residues, three of which are distributed evenly along the bottom β -sheet plate (Trp141, 181 and 190), while the other two Trp residues (Trp39 and Trp299) reside in a more central region of each lobe (Figure 14).

Analysis of the pepsin structure (PDB: 4PEP) using DeepView (118) indicated that residues Trp141 and Trp181 were quite solvent exposed and would be expected to contribute the least to the fluorescence signal. Trp299, 190 and 39 appeared to be relatively solvent inaccessible, but were found to be within 3-4 Å of carboxyl groups, which could lead to some quenching (119). However, Trp39, Trp190 and Trp299 could be expected to contribute the greatest to the measured fluorescence signal.

Trp-fluorescence was used to monitor the conformational changes of pepsin upon unfolding and refolding. The transfer of Trp from a polar to non-polar environment results in a blue shift of the fluorescence emission. Conversely, exposure to aqueous solution of previously “buried” Trp residues will be reflected by a red shift in the emission spectrum [for further reading, refer to (119)]. The fluorescence spectra for pepsin in three states (Np, Ip and Rp) are given in Figure 15. A comparison of the maximum fluorescent intensities (F) and their corresponding wavelengths (λ_F) (Table 5), showed that for Ip F decreased to ~ 72 % of the value for Np and was red-shifted by ~ 3 nm. Refolding of the denatured pepsin gave a spectrum characteristically very similar to that of the native pepsin, in that F was increased and λ_F was blue-shifted relative to the Ip state.

Figure 14. Ribbon representation of the porcine pepsin A structure (PDB: 4PEP) highlighting the position of the five tryptophan residues. The five Trp residues are shown in Vdw radii coloured according to solvent accessibility (green – most, blue – least accessible) and with the solvent accessible surface shown (dot surface). Ribbon colouration is according to the following domain structure: blue (N-terminal lobe) grey (bottom β -sheet plate) and red (C-terminal lobe).

Figure 15. Fluorescence spectra for pepsin ($5.8 \mu\text{M}$) in three different states: (1) Np, pH 5.3 (*solid, uppermost curve*) (2) Ip, pH 8.0 (*solid, lowest curve*) and (3) Rp, pH 5 (*dotted curve*). Representative scans are shown. Measurements were performed at least in triplicate for each sample.

Table 5. Fluorescence spectral parameters^a for pepsin in H₂O and D₂O buffer.

Condition (pH)	λ_F (nm)	F	F_{relative} (%)
Np, H ₂ O	348.0 $\pm$ 0.8 _{A, D}	67.8 $\pm$ 2.6 _{A, E}	100
Ip, H ₂ O	352.3 $\pm$ 0.2 _B	49.1 $\pm$ 3.2 _B	72
Rp, H ₂ O	349.0 $\pm$ 0.7 _A	65.1 $\pm$ 1.3 _A	96
Np, D ₂ O	348.0 $\pm$ 0.3 _A	76.9 $\pm$ 5.5 _C	100
Ip, D ₂ O	351.2 $\pm$ 0.5 _C	57.4 $\pm$ 8.1 _D	75
Rp, D ₂ O	346.6 $\pm$ 0.4 _D	68.3 $\pm$ 0.5 _{C, E}	89

^aMeans in each column sharing the same letter are not significantly different, $p > 0.05$.

The λ_F values suggested (120) that those Trp residues that contributed substantially to the spectra for Np and Rp were partially exposed, whereas those for Ip were completely exposed to the solvent. As free Trp in solution can yield a maximum fluorescent intensity at 353 nm (120, 121), the data presented in this research suggested that Trp39, Trp190 and Trp299 were more exposed (and/or quenched) in Np than one would predict from the crystal structure using DeepView (118), perhaps due to dynamic fluctuations.

4.5 Differential Scanning Calorimetry

As shown below (Table 6), DSC measurements were used to assess the conformational stability of the various states of pepsin. For all but the Ip samples, the excess heat capacity scans gave rise to two apparent melting events (termed here peak one, P1 and peak two, P2). P2 was assumed to correspond to the main melting or unfolding transition of both pepsin domains, while the origin of P1 was unclear. Although two separate melting transitions of pepsin have previously been observed and associated with the independent unfolding of the N- and C-terminal domains (65), that possibility was ruled out for these measurements. In contrast to results reported by Privalov et al. (65), where the N- and C-domain melting transitions were of comparable enthalpy and the two transitions converged to a single transition for pepsin samples below pH 6, in the present study the calorimetric enthalpy (ΔH^{cal}) of P1 was nearly 1/10th the magnitude of that for P2 (Table 6). Furthermore, in some heat capacity profiles, P1 appeared to more closely resemble a “step” in the heat capacity profile than an actual

peak (Figure 16), and so it was not clear whether P1 corresponded to an unfolding transition. Nevertheless, for comparative analysis P1 was treated as an enthalpic transition.

Figure 17 shows the heat capacity profiles measured from heating scans of pepsin under the various conditions (in H₂O). It was found that Np gave a fairly sharp, high enthalpy transition relative to Ip, which showed a much broader transition with lower enthalpy and a concomitantly lower transition temperature (T_m). Measurements by DSC revealed unexpected differences in the behaviour of Rp and CRp. In the case of the concentrated-refolded protein the transition was narrow, with a width and enthalpy comparable to that of the native state. Importantly, the T_m of CRp was found to be significantly ($p \leq 0.05$) higher than that of the native state, indicating a conformational difference between the native and concentrated refolded states of pepsin. In the dilute-refolded case the heat capacity profile was quite low and broad, although there was also a significant ($p \leq 0.05$) shift of T_m to a higher value than that observed for Np.

Figure 18 shows the first scan of Np along with the second, third and fourth re-scans of the same sample. Although the Np transition was found to be irreversible, with a second scan showing a transition with only $\sim 20\%$ ΔH^{cal} of the original transition, the second scan gave an excess heat capacity profile similar to that for Rp. Successive heating scans of the melted Np sample revealed that this transition was highly reversible, giving 95 % and 85% of the initial ΔH^{cal} on the second and third scans, respectively. In addition, Np samples taken from the DSC after the first heating/cooling cycle were analyzed by CD and yielded spectra that overlaid well with that of Rp (Figure 19).

The ratio of the calorimetric enthalpy (ΔH^{cal}) to the van't Hoff enthalpy (ΔH^{vh}) can be taken as a measure of the cooperativity of the protein unfolding (74). A $\Delta H^{\text{cal}}/\Delta H^{\text{vh}}$ ratio of one indicates a completely cooperative and two-state unfolding process, while any deviation from unity indicates either a reduction in cooperativity (i.e., cooperative units smaller than the whole protein) or non two-state behaviour of the protein (74). $\Delta H^{\text{cal}}/\Delta H^{\text{vh}}$ ratios (Table 7) were estimated by fitting the Np, CRp and Rp endotherms (in H₂O) with a single non two-state transition (Figure 20). The larger $\Delta H^{\text{cal}}/\Delta H^{\text{vh}}$ for Np compared to that for CRp suggested that unfolding of CRp was slightly more cooperative, whereas Rp unfolding was substantially more cooperative than Np. Overall, the data suggested that unlike Rp, the unfolding of both CRp and Np proceeded through a non two-state mechanism.

Although the Np and CRp transitions were only semi-cooperative and non-reversible, their approximation as a two-state transition (7) allowed for the comparison of the relative thermodynamic stability of Np and CRp with that of Rp. Fitting the endotherms from Np, CRp and Rp (in H₂O) with a single two-state transition (Figure 21), the resulting values of T_m , $\Delta H(T_m)$ and heat capacity change at the transition midpoint, $\Delta C_p(T_m)$, were used to calculate apparent Gibbs free energy values, ΔG_{app} (Table 8). Previous studies of Gdm-HCl denaturation of pepsin monitored by activity and absorbance (300 nm) yielded an apparent free energy value of $\Delta G_{\text{app}} = 5.42 \pm 0.36$ kcal/mol (25 °C) (122), which is in close agreement with the value for Np determined in the present study. ΔG_{app} (25 °C) for CRp was roughly twice that of Np while that for Rp was slightly higher, demonstrating that both CRp and Rp were significantly ($p \leq 0.05$) more stable than Np.

The stabilization of Rp relative to Np appeared to be due to both a higher T_m and that the change in apparent heat capacity ΔC_p upon melting of Rp ($\Delta C_p = 447 \pm 226$ cal/mol/°C) was found to be considerably less than that for Np ($\Delta C_p = 5170 \pm 198$ cal/mol/°C), as can be seen in Figure 22A. The formation of Rp upon unfolding of Np is a possible explanation for the broadening of the Np excess heat capacity profile. This was supported by fitting the Np data with a single, non two-state transition and then adding this curve to the Rp excess heat capacity function, which yielded a curve similar to the Np heat capacity profile (Figure 22B).

Figure 16. Two apparent transitions observed upon heating pepsin. Heat capacity profiles are shown for Np, Rp and CRp after solvent subtraction. P1 appeared as a step in the heat capacity (solid curve), a small transition (symbol+line) or somewhat indistinguishable between a step/peak (dashed curve). Representative scans are shown. Measurements were performed at least in triplicate for each sample.

Figure 17. Four conformations of pepsin as observed by DSC. Endotherms shown are for Np, Ip, Rp and CRp. For example, “refolded, 0.8-0.4 mg/ml” indicates that 0.8 mg/ml of alkaline unfolded pepsin was refolded (thus diluted) to 0.4 mg/ml. Representative scans are shown. Measurements were performed at least in triplicate for each sample.

Figure 18. Irreversibility of Np unfolding by DSC. Upon cooling after the first scan of Np (solid curve), the second, third and fourth scans (broken curves) were highly reversible. Measurements were performed at least in triplicate for each sample.

Figure 19. Comparison of thermally and pH refolded Rp by CD. CD spectra for an Np sample taken from the DSC after the first heating/cooling cycle, termed Rp_{melt}, (solid curve) was identical to that of Rp (dashed curve). Measurements were performed at least in triplicate for each sample.

Figure 20. Non two-state approximation of DSC transitions. The excess heat capacity profiles for Np, CRp and Rp (thick/black curves) were fit as a non two-state transition excluding a heat capacity term (thin/red curves). A baseline (cubic connect) was subtracted from the data before fitting. Fittings were performed at least in triplicate; representative fits are shown.

Figure 21. Two-state approximation of DSC transitions. The excess heat capacity profiles for Np, CRp and Rp (thick/black curves) were fit as a two-state transition including a heat capacity term (thin/red curves) and progress baseline (broken curve). Fittings were performed at least in triplicate; representative fits are shown.

Figure 22. Melting of native pepsin occurred as, at least, a three-state transition. (A) DSC data, after buffer subtraction: upon melting Np (solid curve), the second scan (broken curve) was identical to that for Rp. (B) The function resulting from a single, non two-state fit of the Np data was added to the excess heat capacity profile for Rp,melt, resulting in a profile similar to the Np curve. Scans were collected at least in triplicate.

Table 6. Parameters^{a, b, c} derived from the DSC scans of the Np, Ip, Rp and CRp states of pepsin in both H₂O and D₂O buffers, for either one or two transitions where observed.

Sample	T_{m1} (°C)	ΔH_1^{cal} (kcal/mole)	T_{m2} (°C)	ΔH_2^{cal} (kcal/mole)
Np, H ₂ O	33.63 ± 1.53 _A	10.55 ± 2.25 _A	62.16 ± 0.29 _A	176.72 ± 2.32 _A
Ip, H ₂ O			56.47 ± 0.48 _B	42.83 ± 1.33 _B
Rp, H ₂ O	35.04 ± 1.83 _{A, B}	6.70 ± 2.14 _B	67.00 ± 2.05 _{C, D}	44.67 ± 4.27 _B
CRp, H ₂ O	35.82 ± 1.51 _{A, B}	26.66 ± 6.41 _C	65.11 ± 0.13 _C	152.83 ± 3.05 _C
Np, D ₂ O	35.17 ± 0.54 _{A, B}	11.98 ± 2.20 _A	67.61 ± 0.19 _D	157.83 ± 15.96 _{A, C}
Ip, D ₂ O			61.95 ± 0.34 _A	42.05 ± 1.98 _B
Rp, D ₂ O	37.21 ± 1.98 _B	9.99 ± 6.01 _{A, B}	63.68 ± 3.09 _{A, C}	20.94 ± 4.19 _H
CRp, D ₂ O	35.97 ± 0.42 _{A, B}	12.14 ± 1.97 _A	71.03 ± 0.22 _E	161.39 ± 11.39 _C

^aThese parameters were obtained by integrating the area between the baseline (created as a cubic curve) and the experimental curve. Obtaining data in this way assumes $\Delta C_p = 0$ for the transitions.

^bValues were determined as the mean ± SD from at least three separate DSC scans.

^cMeans in each column sharing the same letter are not significantly different, $p > 0.05$.

Table 7. Cooperativity of unfolding of Np, Rp and CRp indicated by $\Delta H^{\text{cal}} / \Delta H^{\text{vh}}$ ratios.

	Np,H ₂ O	CRp,H ₂ O	Rp,H ₂ O
$\Delta H^{\text{cal}} / \Delta H^{\text{vh}}$	1.36 ± 0.08	1.17 ± 0.04	1.02 ± 0.07

^aValues are the mean ± SD of results from multiple DSC scans, n > 3, not of a single fit.

Table 8. Relative stability of Np, CRp and Rp compared using apparent free energy values ΔG_{app} (25 °C)^a.

	Np,H ₂ O	CRp,H ₂ O	Rp,H ₂ O
ΔG_{app} (25 °C) (kcal/mol)	4.85 ± 0.58	11.19 ± 0.953	5.75 ± 0.41

^aValues are the mean ± SD of results from multiple DSC scans, n > 3, not of a single fit.

4.6 Refolded Pepsin Activity

Activity measurements were made on refolded pepsin samples as described above. While Rp displayed no activity, CRp gave an initial rate of $-\Delta A/\text{min} = 0.012 \pm 0.003 \text{ min}^{-1}$. For comparison, Np had a significantly greater ($p \leq 0.05$) initial rate of $-\Delta A/\text{min} = 0.031 \pm 0.005 \text{ min}^{-1}$ under the same conditions. Thus, the activity of CRp was found to be about 39 % of that of the native form.

5. Discussion

5.1 Unfolding

The present study confirmed and supports previous (68) research regarding the Ip state of pepsin. Similar to results reported by Konno et al. (68), an increase in R_g was found upon alkaline unfolding (from $\sim 20 \text{ \AA}$ to $\sim 30 \text{ \AA}$). Although enzymatically inactive at pH 8, pepsin still retained a high degree of secondary structure, exhibited a globular rather than random coil conformation, and underwent a heat induced, enthalpic unfolding transition. In particular, ab initio modelling of the SANS data illustrated the nature of the Ip conformation as being relatively compact yet distinctly elongated relative to Np, Rp and CRp.

The Ip state was partially unfolded, as CD measurements indicated a loss of β -strand and an increase in random structure, in agreement with previous studies (64, 68, 70). Also, DSC measurements showed that the melting transition of Ip was much broader,

decreased in T_m and had a reduced enthalpy compared to that of Np. However, the heat capacity profile demonstrated that Ip still retained residual structure, which was denatured on heating.

SANS measurements showed a retention of tertiary structure via Guinier plots (Figure 8) and low resolution structural reconstruction (Figure 11), which showed that the Ip state was still comparatively compact, and by Kratky plots (Figure 9) which indicated that, although Ip differed from Np, it still retained globular features. Densitometry measurements also showed that there was no difference in the apparent partial specific volume of Ip and Np, though it is not clear a priori whether a protein should always show a volume expansion upon denaturation, as various other proteins have been reported to decrease or increase in volume (95, 123) or to show no change (95).

In an earlier functional study (66), in which pepsin activity was recovered by the addition of recombinant N-terminal domain of pepsinogen to denatured pepsin (Ip), it was postulated that the C-terminal domain retained its native structure in the Ip state. From recent structural investigations (68, 69), it was shown that the C-terminal lobe of Ip had non-native secondary structure and further, that the N-terminal lobe did retain some degree of non-random structure. From the SANS data in the present study it appeared that the N-terminal lobe of pepsin was not fully unfolded, as Kratky plots indicated a relatively compact form for the Ip state with no apparent contribution to the scattering profile from a random coil conformation. Furthermore, from the models produced by SASHA, and in particular GASBOR (Figure 11), it appeared that both N- and C-domains may have undergone unfolding as the two domains separated. The results of Lin et al. (66) and Privalov et al. (65), rather than suggesting a native C-domain in the Ip state, may

support the notion that the C-domain retains the ability to refold and form a functional domain, whereas the N-domain is incapable of refolding.

5.2 Refolding

Relative to that of the Ip state, information available in the literature on the structure of renatured pepsin is limited and somewhat conflicting. Some studies (63, 71) have reported that alkaline denatured and subsequently acid renatured pepsin generated a CD spectrum very similar to the native spectrum, while it has recently been reported that renatured pepsin did not give a native-like CD spectrum (68). Measurements by CD (Figure 13) and DSC (Figure 17) showed that the refolding of pepsin was concentration dependent. Two regimes of pepsin renaturation were carried out, one with dilute Ip (< 10 mg/ml) and the other with concentrated Ip (> 10 mg/ml), yielding two distinct refolded states of pepsin, i.e., Rp and CRp. Both the secondary and tertiary structures of Rp were intermediate between the Np and Ip forms and using DSC showed a reversible, low enthalpy transition that was approximately two-state. In contrast, CRp exhibited secondary and tertiary structures similar to Np, but with a greater conformational stability. Although the concentrated refolding of pepsin from the Ip state led to nearly complete recovery of both its secondary and tertiary structures, it was found that CRp had considerably lower activity for the synthetic substrate than Np, suggesting that CRp was misfolded with respect to the native state.

To our knowledge, this is the first report of the tertiary structure of pepsin under renaturing conditions. The partially unfolded conformation of the Ip state has been

characterized as remaining globular yet extended. Combined with the findings that both domains contain non-native structure (68, 69), it is hypothesized that the two domains unfold while separating, as supported by the models in Figure 11. Re-acidifying the alkaline denatured state was shown (Figure 11) to result in a partial reversal of this process. In the Rp state, it appeared that both domains became more compact. Furthermore, the fluorescence spectrum of Rp was similar to that of Np (Table 5), which also indicated that Rp was more compact than Ip.

Comparison of the DSC heat capacity profiles of Rp, CRp and Np (Figure 17; Table 6) revealed that both Rp and CRp had higher T_m values, and as the T_m can be taken as a reliable measure of stability (124), this result suggested that the refolded states had a greater conformational stability than Np. Calculation of an apparent Gibbs free energy change of unfolding (ΔG_{app}) further supported that Rp and CRp were thermodynamically more stable than Np. However, the transitions for Np and CRp were both completely irreversible, in that after the first melting and cooling cycle of Np or CRp samples, successive heating scans produced excess heat capacity profiles similar to those of Rp. Furthermore, Np samples that were heated and cooled by DSC gave CD spectra identical to that of Rp. Thus, it was observed that upon melting Np and cooling the sample Rp was formed, indicating that the Rp state could be formed both by thermal and pH induced denaturation/renaturation of pepsin. Thus, it is hypothesized that Np at pH 5.3 represents a metastable state, whereas Rp represents a truly thermodynamically stable state.

The findings reported in the present study and elsewhere (63, 68, 71) have shown that, for equivalent final conditions, different refolded states of pepsin could be observed depending on the initial conditions (e.g., buffer, pH, protein concentration, etc.)

before/during refolding. From a thermodynamic perspective, the starting conditions should not determine the final folded state, only the final conditions (34); thus, the present results are suggestive of a strong kinetic influence on the folding mechanism of pepsin.

Comparison of our initial calorimetry data for Np, especially with that of Rp but also with that of CRp, suggests that non-native conformations of pepsin exist which are thermodynamically more stable than the native structure. This would imply that the Np state is kinetically stabilized, similar to α -lytic protease (35, 36, 125), *S. griseus* protease B (38, 125) and possibly subtilisin (37). Therefore, the present findings favour the hypothesis that the reason why denaturation of pepsin is an irreversible, or poorly reversible (126) process, is that the native conformation is energetically disfavoured.

5.3 Folding Cooperativity

The DSC data supported that Rp was formed subsequent to the unfolding of Np, and that the broadening of the Np excess heat capacity profile towards higher temperatures was due to the formation of Rp, as shown in Figure 22. Therefore, the excess heat capacity profile of Np could be summarized as constituting (at least) a three state transition, in which an initial, irreversible step (the loss of the Np conformation) was followed by a second, reversible transition, which was approximately two-state:

The cooperativity of the unfolding of the N- and C-domains in the Np $\rightarrow$ Up transition was estimated using a single non two-state fit of the entire profile, and was found to be semi-cooperative as indicated by the $\Delta H^{\text{cal}}/\Delta H^{\text{vh}}$ ratio (Table 7). In comparison, the calorimetry data collected on pepsin (under different buffer conditions) by Privalov et al (65) indicated that the N- and C-domains could unfold in a completely non-cooperative manner, as two well separated transitions of comparable enthalpy were observed. Thus, the cooperativity of the initial, irreversible transition presented above may be quite sensitive to solvent conditions. Experimental conditions can play a large role in how a protein is observed to unfold, such that extrapolation to how a protein folds (mechanism) should be done with caution. For example, it was recently shown that the all α -helical domains R16 and R17 of chicken α -spectrin, connected in tandem, only display cooperative unfolding at 25 °C in the absence of salt; at higher temperature, or in the presence of salts, the unfolding is not cooperative (127).

5.4 Molten Globule vs. Partly Folded States

Viscosity measurements made by McPhie (63) indicated that alkaline denatured pepsin was quite compact (intrinsic viscosity, 7.8 mg/ml) relative to the completely unfolded state (intrinsic viscosity, 31.5 mg/ml) and was more comparable with the native state (intrinsic viscosity, 3.2 mg/ml). The SANS data presented above are in agreement with these findings, as Ip was found to be elongated relative to Np yet remained globular. Furthermore, the binding of Ip to GroEL was reported to occur without substantial contributions from hydrophobic interactions (86) and Ip was found to not interact with

the ANS fluorescent probe (68). Although this latter finding could be due to charge repulsions, as both ANS and Ip (114) are highly negatively charged, these studies (68, 86) support that Ip does not have a considerably larger hydrophobic surface area exposed. Combined with the significant loss of secondary structure (Table 4), it is likely that Ip is best described as a partially unfolded state and not a molten globule intermediate.

The results show that the secondary structure of Rp is more closely native-like than the Ip state, and the SANS and fluorescence data suggest a native-like compactness of Rp. Interestingly, by DSC it was found that Rp had a relatively small ΔC_p , indicating (16-18) a larger area of hydrophobic surface exposure in Rp. Together these findings suggest that Rp could suitably be called a molten globule state; however, the findings that Rp was thermodynamically more stable and cooperative than the other states examined strongly excludes (19) this classification. Thus, it did not seem that a framework folding model (4, 17), consisting of molten globule states, correctly described the folding of pepsin.

5.5 Proposed Role of the Prosegment in Pepsin Folding

As mentioned above in Section 5.1, the findings of Lin et al. (66) have often (128) been interpreted to suggest that the C-domain remains in a native conformation in the Ip state, and that the source of the irreversibility in the folding of pepsin is due solely to the irreversible denaturation of the N-domain. However, in the previous study (66) the folding competence of the N- and C-domains were compared under uneven conditions. That is, when native N-lobe was added to Ip (sample regained activity at pH 2), the

prosegment was present; but, when native C-lobe was added to Ip (sample had no activity at pH 2), the prosegment was not present. As the addition of the prosegment to other denatured peptidases has been shown to result in recovery of native activity in the past (36, 38, 129), it is quite plausible that the research by Lin et al. (66) has actually shown that Ip can, in the presence of the prosegment, be refolded to an active form. In this light, the findings presented in this thesis on pepsin denaturation, combined with previous results on other zymogen-derived peptidases (125), indicate that the role of the pro-region in these proteins is not (only) to stabilize or inhibit the active site, but to catalyze the folding to kinetically trapped native states.

6. Conclusions

The use of SANS and complementary solution structure techniques provided a thorough characterization of the conformations of various folded pepsin states. In particular, the combination of SAS and DSC proved to be a powerful means to probe the elements of structure, cooperativity and stability in the folding of a protein.

The main findings presented in this thesis are summarized in Figure 23. The alkaline denatured state, Ip, was partially unfolded and marginally stable but retained a globular structure. It is hypothesized that both the N- and C-domains were partially unfolded. The acid renatured state, Rp, was partially refolded with both domains becoming more compact, yet was thermodynamically stable. It was concluded that because native pepsin is thermodynamically metastable, denaturation and subsequent renaturation results in the formation of Rp. Np is likely a kinetically stabilized state that

is only practically realizable when folded in the presence of the prosegment, which catalyzes folding to the active conformation.

Future studies on pepsin folding should include the examination of exactly how the prosegment functions as a folding catalyst. One possible method to do this would be to identify key residues of the prosegment responsible for lowering the folding activation energy. As pepsinogen can be reversibly folded, the effect of systematic mutations to the prosegment on the stability of the folding transition state could be measured and interpreted thermodynamically and kinetically.

Further studies on the structures of the Np, Rp and CRp states should help to understand what differentiates a kinetically stabilized fold from a thermodynamically stabilized fold. One hypothesis to test would be whether a kinetically stabilized state is dynamically constricted relative to the thermodynamically stable state. It has been hypothesized that extracellular bacterial peptidases may have developed kinetic stability to enhance resistance to proteolysis, and that this resistance is due to the native states being more rigid (38, 125, 130). Complementary methods such as neutron spectroscopy and NMR can be used, respectively, to measure faster side chain and slower conformational fluctuations of proteins. Using these methods to investigate the dynamic fluctuations of the Rp and Np states would allow a unique opportunity to address these questions.

Figure 23. Summary of findings on the irreversible denaturation of pepsin. Both the N- and C-domains appear to partially unfold (Ip state) and partially refold (Rp state), resulting in a non-native but thermodynamically stable conformation (Rp).

7. References

1. The International Human Genome Mapping Consortium. (2001) A physical map of the human genome. *Nature* 409, 934-941.
2. The Celera Genomics Sequencing Team. (2001) The sequence of the human genome. *Science* 291, 1304-1351.
3. Jackson, S. E. (1998) How do small single-domain proteins fold? *Fold. Des.* 3, R81-R91.
4. Daggett, V., and Fersht, A., R. (2003) Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* 28, 18-25.
5. Radford, S. (2000) Protein folding: progress made and promise ahead. *Trends Biochem. Sci.* 25, 611-618.
6. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science* 181, 223-230.
7. Brown, E. D., and Yada, R. Y. (1991) A kinetic and equilibrium study of the denaturation of aspartic proteinases from the fungi, *Endothia parasitica* and *Mucor miehei*. *Biochim. Biophys. Acta.* 1076, 406-415.
8. Song, J., Xu, P., Xiang, H., Su, Z., Storer, A. C., and Ni, F. (2000) The active-site residue cys-29 is responsible for the neutral-pH inactivation and the refolding barrier of human cathepsin B. *FEBS Lett.* 475, 157-162.
9. Beldarrain, A., Acosta, N., Montesinos, R., Mata, M., and Cremata, J. (2000) Characterization of *Mucor pusillus* rennin expressed in *Pichia pastoris*: enzymic, spectroscopic and calorimetric studies. *Biotechnol. Appl. Biochem.* 31, 77-84.
10. Solís-Mendiola, S., Gutiérrez-González, L. H., Arroyo-Reyna, A., Padilla-Zúñiga, J., Rojo-Domínguez, A., and Hernández-Arana, A. (1998) pH dependence of the activation parameters for chymopapain unfolding: influence of ion pairs on the kinetic stability of proteins. *Biochim. Biophys. Acta.* 1388, 363-372.
11. Petsko, G. A., and Ringe, D. (2004) *Protein Structure and Function*. New Science Press Ltd, London.
12. Scharnagl, C., Reif, M., and Friedrich, J. (2005) Stability of proteins: temperature, pressure and the role of the solvent. *Biochim. Biophys. Acta.* 1749, 187-213.
13. Dill, K. A. (1990) Dominant forces in protein folding. *Biochemistry* 29, 7133-7155.
14. Harano, Y., and Kinoshita, M. (2005) Translational-entropy gain of solvent upon protein folding. *Biophys. J.* 89, 2701-2710.

15. Privalov, P. L., and Gill, S. J. (1988) Stability of protein structure and hydrophobic interaction. *Adv. Protein Chem.* 39, 191-234.
16. Loladze, V. V., Ermolenko, D. N., and Makhatadze, G. I. (2001) Heat capacity changes upon burial of polar and nonpolar groups in proteins. *Protein Sci.* 10, 1343-1352.
17. Ptitsyn, O. B. (1995) Molten clobule and protein folding. *Adv. Protein Chem.* 47, 83-229.
18. Gómez, J., Hilser, V. J., Xie, D., and Freire, E. (1995) The heat capacity of proteins. *Proteins Struct. Funct. Genet.* 22, 404-412.
19. Makhatadze, G., and Privalov, P. (1995) Energetics of protein structure. *Adv. Protein Chem.* 47, 307-425.
20. Kunugi, S., and Tanaka, N. (2002) Cold denaturation of proteins under high pressure. *Biochim. Biophys. Acta.* 1595, 329-344.
21. Franks, F. (1995) Protein destabilization at low temperatures. *Adv. Protein Chem.* 46, 105-139.
22. Weiss, S. (2000) Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat. Struct. Biol.* 7, 724-729.
23. Schuler, B., Lipman, E. A., and Eaton, W. A. (2002) Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 419, 743-747.
24. Russo, D., Durand, D., Calmettes, P., and Desmadril, M. (2001) Characterization of the denatured states distribution of neocarzinostatin by small-angle neutron scattering and differential scanning calorimetry. *Biochemistry* 40, 3958-3966.
25. Ptitsyn, O. B. (1995) Structures of folding intermediates. *Curr. Opin. Struct. Biol.* 5, 74-78.
26. Dobson, C. M. (1992) Unfolded proteins, compact states and molten globules. *Curr. Opin. Struct. Biol.* 2, 6-12.
27. Myers, J. K., and Oas, T. G. (2001) Preorganized secondary structure as an important determinant of fast protein folding. *Nat. Struct. Biol.* 8, 552-558.
28. Peng, Z., and Kim, P. S. (1994) A protein dissection study of molten globule. *Biochemistry* 33, 2131-2141.
29. Redfield, C., Smith, R. A. G., and Dobson, C. M. (1994) Structural characterization of a highly ordered 'molten globule' at low pH. *Nat. Struct. Biol.* 1, 23-29.
30. Epand, R. M., and Scheraga, H. A. (1968) The influence of long-range interactions on the structure of myoglobin. *Biochemistry* 7, 2864-2872.

31. de Prat Gay, Gonzalo, Ruiz-Sanz, J., Neira, J. L., Itzhaki, L. S., and Fersht, A. R. (1995) Folding of a nascent polypeptide chain in vitro: cooperative formation of structure in a protein module. *Proc. Natl. Acad. Sci. U. S. A.* 92, 3683-3686.
32. Ptitsyn, O. B. (1992) Secondary structure formation and stability. *Curr. Opin. Struct. Biol.* 2, 13-20.
33. Dobson, C. M. (2003) Protein folding and misfolding. *Nature* 426, 884-890.
34. Baker, D., and Agard, D. A. (1994) Kinetics versus thermodynamics in protein folding. *Biochemistry* 33, 7505-7509.
35. Sohl, J. L., Jaswal, S. S., and Agard, D. A. (1998) Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature* 395, 817-819.
36. Baker, D., Sohl, J. L., and Agard, D. A. (1992) A protein-folding reaction under kinetic control. *Nature* 356, 263-265.
37. Eder, J., Rheinacker, M., and Fersht, A. R. (1993) Folding of subtilisin BPN': characterization of a folding intermediate. *Biochemistry* 32, 18-26.
38. Truhlar, S. M. E., Cunningham, E. L., and Agard, D. A. (2004) The folding landscape of *Streptomyces griseus* protease B reveals the energetic costs and benefits associated with evolving kinetic stability. *Protein Sci.* 13, 381-390.
39. Baskakov, I. V., Legname, G., Prusiner, S. B., and Cohen, F. E. (2001) Folding of prion protein to its native alpha-helical conformation is under kinetic control. *J. Biol. Chem.* 276, 19687-19690.
40. Fruton, J. S. (2002) A history of pepsin and related enzymes. *Quart. Rev. Biol.* 77, 127-147.
41. Dunn, B. M. (2002) Structure and mechanism of the pepsin-like family of aspartic peptidases. *Chem. Rev.* 102, 4431-4458.
42. Lipatov, N. N., Boreskov, V. G., Krokha, N. G., Mitaseva, L. F., and Stefanov, A. V. (1989) Biotechnological modification of meat trimmings with a high level of connective tissue. *Izv. Vyssh. Uchebn. Zaved. , Pishch. Tekhnol.* 5, 12-15.
43. Barbano, D. M., and Rasmussen, R. R. (1992) Cheese yield performance of fermentation-produced chymosin and other milk coagulants. *J. Dairy Sci.* 75, 1-12.
44. Ohtani, H. (1998) Coagulated milk and enzymes. *New Food Ind.* 30, 65-79.
45. Champagne, E. T., Hinojosa, O., and Clemetson, C. A. B. (1990) Production of ascorbate free radicals in infant formulas and other media. *J. Food Sci.* 55, 1133-1136.

46. Tomita, M., Wakabayashi, H., Yamauchi, K., Teraguchi, S., and Hayasawa, H. (2002) Bovine lactoferrin and lactoferricin derived from milk: production and applications. *Biochem. Cell Biol.* 80, 148.
47. Cooper, J. B., and Myles, D. A. (2000) A preliminary neutron laue diffraction study of the aspartic proteinase endothiapepsin. *Acta Crystallogr. Sect. D* 56 (Pt 2), 246-248.
48. Hong, L., Koelsch, G., Lin, X., Wu, S., Terzyan, S., Ghosh, A. K., Zhang, X. C., and Tang, J. (2000) Structure of the protease domain of memapsin 2 (beta-secretase) complexed with inhibitor. *Science* 290, 150-153.
49. Westling, J., Cipullo, P., Hung, -H. S., Saft, H., Dame, J. B., and Dunn, B. M. (1999) Active site specificity of plasmepsin II. *Pro. Sci.* 8, 2001-2009.
50. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B. K., Bladwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., and Kent, S. (1989) Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* 245, 616-621.
51. Baker, D., Shiau, A. K., and Agard, D. A. (1993) The role of pro regions in protein folding. *Curr. Opin. Cell Biol.* 5, 966-970.
52. Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S., and James, M. N. G. (1990) Molecular and crystal structures of monoclinic porcine pepsin refined at 1.8 Å resolution. *J. Mol. Biol.* 214, 143-170.
53. Tang, J., James, M. N. G., Hsu, I. N., Jenkins, J. A., and Blundell, T. L. (1978) Structural evidence for gene duplication in the evolution of the acid proteases. *Nature* 271, 618-621.
54. Hartsuck, J. A., Koelsch, G., and Remington, S. J. (1992) The high-resolution crystal structure of porcine pepsinogen. *Proteins* 13, 1-25.
55. Richter, C., Tanaka, T., and Yada, R. Y. (1998) Mechanism of activation of the gastric aspartic proteinases: pepsinogen, progastricsin and prochymosin. *Biochem J.* 335, 481-490.
56. Lin, X., Wong, R. N. S., and Tang, J. (1989) Synthesis, purification, and active site mutagenesis of recombinant porcine pepsinogen. *J. Biol. Chem.* 264, 4482-4489.
57. Cornish-Bowden, A. J., and Knowles, J. R. (1969) The pH-dependence of pepsin-catalysed reactions. *Biochem. J.* 113, 353-362.
58. Veerapandian, B., Cooper, J. B., Sali, A., Blundell, T. L., Rosati, R. L., Dominy, B. W., Damon, D. B., and Hoover, D. J. (1992) Direct observation by X-ray-analysis of the tetrahedral intermediate of aspartic proteinases. *Protein Sci.* 1, 322-328.

59. Northrop, D. B. (2001) Follow the protons: a low-barrier hydrogen bond unifies the mechanisms of the aspartic proteases. *Acc. Chem. Res.* 34, 790-797.
60. Coates, L., Erskine, P. T., Mall, S., Gill, R., Wood, S. P., Myles, D. A. A., and Cooper, J. B. (2006) X-ray, neutron and NMR studies of the catalytic mechanism of aspartic proteinases. *Eur Biophys J. Epub ahead of print.*
61. Ahmad, F., and McPhie, P. (1978) Thermodynamics of the denaturation of pepsinogen by urea. *Biochemistry* 17, 241-246.
62. Frattali, V., Steiner, R. F., and Edelhoch, H. (1965) Native and unfolded states of pepsinogen. I. the molecular conformation in water and in urea. *J. Biol. Chem.* 240, 112-121.
63. Ahmad, F., and McPhie, P. (1978) The denaturation of covalently inhibited swine pepsin. *Int. J. Pept. Protein Res.* 12, 155-163.
64. McPhie, P. (1989) A reversible unfolding reaction of swine pepsin; implications for pepsinogen's folding mechanism. *Biochem. Biophys. Res. Commun.* 158, 115-119.
65. Privalov, P. L., Mateo, P. L., Khechinashvili, N. N., Stepanov, V. M., and Revina, L. P. (1981) Comparative thermodynamic study of pepsinogen and pepsin structure. *J. Mol. Biol.* 152, 445-464.
66. Lin, X. L., Loy, J. A., Sussman, F., and Tang, J. (1993) Conformational instability of the N- and C-terminal lobes of porcine pepsin in neutral and alkaline solutions. *Protein Sci.* 2, 1383-1390.
67. Tanaka, T., and Yada, R. Y. (2001) N-terminal portion acts as an initiator of the inactivation of pepsin at neutral pH. *Protein Eng.* 14, 669-674.
68. Konno, T., Kamatari, Y. O., Tanaka, N., Kamikubo, H., Dobson, C. M., and Nagayama, K. (2000) A partially unfolded structure of the alkaline-denatured state of pepsin and its implication for stability of the zymogen-derived protein. *Biochemistry* 39, 4182-4190.
69. Kamatari, Y. O., Dobson, C. M., and Konno, T. (2003) Structural dissection of alkaline-denatured pepsin. *Protein Sci.* 12, 717-724.
70. Yada, R. Y., and Nakai, S. (1986) Secondary structure of some aspartyl proteinases. *J. Food Biochem.* 10, 155-183.
71. Favilla, R., Parisoli, A., and Mazzini, A. (1997) Alkaline denaturation and partial refolding of pepsin investigated with DAPI as an extrinsic probe. *Biophys. Chem.* 67, 75-83.
72. Brockwell, D. J. (2000) Protein folding mechanisms: new methods and emerging ideas. *Curr. Opin. Struct. Biol.* 10, 16-25.

73. Wayne L. Hubbell, Wayne L., Cafiso, D. S., and Altenbach, C. (2000) Identifying conformational changes with site-directed spin labeling *Nat. Struc. Biol.* 7, 735-739.
74. Privalov, P. L., and Potekhin, S. A. (1986) Scanning microcalorimetry in studying temperature-induced changes in proteins. *Methods Enzymol.* 131, 4-51.
75. Wall, M. E., Gallagher, S. C., and Trewheella, J. (2000) Large-scale shape changes in proteins and macromolecular complexes. *Annu. Rev. Phys. Chem.* 51, 355-380.
76. Koch, Michel H. J., Vachette, P., and Svergun, D. I. (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys.* 36, 147-227.
77. Glatter, O., and Kratky, O. (1982) *Small Angle X-Ray Scattering*. Academic Press, New York.
78. Vachette, P. and Svergun, D. (2000) Small-Angle X-Ray Scattering by Solutions of Biological Macromolecules, in *Structure and Dynamics of Biomolecules: Neutron and Synchrotron Radiation for Condensed Matter Studies* (Franchon E., Geissler E., Hodeau J-L., Regnard J-R., Timmins P.A., Eds.) pp 199-237, Oxford University Press Inc., New York.
79. Higgins, J. S., and Benoit, H. C. (1994) *Polymers and Neutron Scattering*. 8th ed., Oxford University Press, New York.
80. Zaccai, G. (2000) Small-Angle Neutron Scattering, in *Structure and Dynamics of Biomolecules: Neutron and Synchrotron Radiation for Condensed Matter Studies* (Franchon E., Geissler E., Hodeau J-L., Regnard J-R., Timmins P.A., Eds.) pp 238-250, Oxford University Press, New York.
81. Petrescu, A. J., Receveur, V., Calmettes, P., Durand, D., Desmadril, M., Roux, B., and Smith, J. C. (1997) Small-angle neutron scattering by a strongly denatured protein: analysis using random polymer theory. *Biophys. J.* 72, 335-342.
82. Petrescu, A. J., Receveur, V., Calmettes, P., Durand, D., and Smith, J. C. (1998) Excluded volume in the configurational distribution of a strongly-denatured protein. *Protein Sci.* 7, 1396-1403.
83. Petrescu, A. J., Calmettes, P., Durand, D., Receveur, V., and Smith, J. C. (2000) Change in backbone torsion angle distribution on protein folding. *Protein Sci.* 9, 1129-1136.
84. Russo, D., Durand, D., Desmadril, M., and Calmettes, P. (2000) Study of thermally and chemically unfolded conformations of a small beta-protein by means of small-angle neutron scattering. *Phys. B.* 276-278, 520-521.
85. Lairez, D., Pauthe, E., and Pelta, J. (2003) Refolding of a high molecular weight protein: salt effect on collapse. *Biophys. J.* 84, 3904-3916.

86. Aoki, K., Taguchi, H., Shindo, Y., Yoshida, M., Ogasahara, K., Yutani, K., and Tanaka, N. (1997) Calorimetric observation of a GroEL-protein binding reaction with little contribution of hydrophobic interaction. *J. Biol. Chem.* 272, 32158-32162.
87. Dunn, B. M., Jimenez, M., Parten, B. F., Valler, M. J., Rolph, C. E., and Kay, J. (1986) A systematic series of synthetic chromophoric substrates for aspartic proteinases. *Biochem. J.* 237, 899-906.
88. Guinier, A., and Fournet, G. (1955) *Small-Angle Scattering of X-Rays*. John Wiley and Sons, New York.
89. Glatter, O. (1982) Interpretation, in *Small-angle X-ray scattering* (O. Glatter, and O. Kratky, Eds.) pp 167-196, Academic Press, Toronto.
90. Svergun, D., Richard, S., Koch, M., Sayers, Z., Kuprin, S., and Zaccai, G. (1998) Protein hydration in solution: experimental observation by X-ray and neutron scattering. *Proc. Natl. Acad. Sci. U. S. A.* 95, 2267-2272.
91. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* 25, 495-503.
92. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. J. (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J.* 80, 2946-2953.
93. Svergun, D. I., Volkov, V. V., Kozin, M. B., and Stuhmann, H. B. (1996) New developments in direct shape determination from small-angle scattering. 2. uniqueness. *Acta. Crystallogr. Sect. A.* 52, 419-426.
94. Volkova, V. V., and Svergun, D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Cryst.* 36, 860-864.
95. Kratky, O., Leopold, H., and Stabinger, H. (1973) The determination of the partial specific volume of proteins by the mechanical oscillator technique. *Methods Enzymol.* 27, 98-110.
96. Epand, R. M., and Epand, R. F. (1980) Studies of thermotropic phospholipid phase transitions using scanning densitometry. *Chem. Phys. Lipids.* 27, 139-150.
97. Sreerema, N., and Woody, R. W. (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* 209, 32-44.
98. Van Stokkum, I. H. M., Spoelder, H. J. W., Bloemendal, M., Van Grondelle, R., and Groen, F. C. A. (1990) Estimation of protein secondary structure and error analysis from CD spectra. *Anal. Biochem.* 191, 110-118.
99. Manavalan, P., and Johnson, W. C., Jr. (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem.* 167, 76-85.

100. Whitmore, L., and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* 32, W668-W673.
101. Lobley, A., Whitmore, L., and Wallace, B. A. (2002) DICHROWEB: An interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* 18, 211-212.
102. Kelly, S. M., Jess, T. J., and Price, N. C. (2005) How to study proteins by circular dichroism. *Biochim. Biophys. Acta.* 1751, 119-139.
103. Liu, X. Q., and Sano, Y. (1998) Effect of Na⁺ and K⁺ ions on the initial crystallization process of lysozyme in the presence of D₂O and H₂O. *J. Protein Chem.* 17, 479-484.
104. Verheul, M., Roefs, S. P. F. M., and de Kruif, Kees G. (1998) Aggregation of β -lactoglobulin and influence of D₂O. *FEBS Lett.* 421, 273-276.
105. Chakrabarti, G., Kim, S., Gupta, M. L., Barton, J. S., and Himes, R. H. (1999) Stabilization of tubulin by deuterium oxide. *Biochemistry* 38, 3067-3072.
106. Parkin, K. L. (2003) Putting Kinetic Principles into Practice, in *Enzyme Kinetics A Modern Approach* (A. G. Marangoni, Ed.) pp 174-192, John Wiley & Sons, Inc., Hoboken, New Jersey.
107. Wang, L., Tharp, S., Selzer, T., Benkovic, S. J., and Kohen, A. (2006) Effects of a distal mutation on active site chemistry. *Biochemistry* 45, 1383(10)-1393.
108. Papajak, E., Kwiecień, R. A., Rudziński, J., Sicińska, D., Kamiński, R., Szatkowski, Ł., Kurihara, T., Esaki, N., and Paneth, P. (2006) Mechanism of the reaction catalyzed by DL-2-haloacid dehalogenase as determined from kinetic isotope effects. *Biochemistry* 45, 6012-6017.
109. Cook, P. F. (1991) *Enzyme Mechanism from Isotope Effects*. CRC Press, Ann Arbor, MI.
110. Leskovac, V. (2003) *Comprehensive Enzyme Kinetics*. Kluwer Academic / Plenum Publishers, New York.
111. Bonnete, F., Madern, D., and Zaccai, G. (1994) Stability against denaturation mechanisms in halophilic malate dehydrogenase adapt to solvent conditions. *J. Mol. Biol.* 244, 436-447.
112. Pederson, J. S., Hansen, S., and Bauer, R. (1994) The aggregation behavior of zinc-free insulin studied by small-angle neutron scattering. *Eur. Biophys. J.* 22, 379-389.
113. Tuinier, R., and Brûlet, A. (2003) On the long-range attraction between proteins due to nonadsorbing polysaccharide. *Biomacromolecules* 4, 28-31.

114. Horn, D., and Heuck, C. (1983) Charge determination of proteins with polyelectrolyte titration. *J. Biol. Chem.* 258, 1665-1670.
115. Glatter, O. (1977) A new method for the evaluation of small-angle scattering data. *J. Appl. Crystallogr.* 10, 415-421.
116. Glatter, O. (1979) The interpretation of real-space information from small-angle scattering experiments. *J. Appl. Crystallogr.* 12, 166-175.
117. Svergun, D. I., and Koch, M. H. J. (2003) Small-angle scattering studies of biological macromolecules in solution. *Rep. Prog. Phys.* 66, 1735-1782.
118. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18, 2714-2723.
119. Lakowicz, J. R. (2000) *Topics in Fluorescence Spectroscopy: Protein Fluorescence*. Kluwer Academic / Plenum Publishers, New York.
120. Burstein, E. A., Vedenkina, N. S., and Ivkova, M. N. (1973) Fluorescence and the location of tryptophan residues in protein molecules. *Photochem. Photobiol.* 18, 263-279.
121. Lala, A. K., and Kaul, P. (1992) Increased exposure of hydrophobic surface in molten globule state of alpha-lactalbumin: fluorescence and hydrophobic photolabeling studies. *J. Biol. Chem.* 267, 19914-19918.
122. Yoshimasu, M. A., Tanaka, T., Ahn, J. -K., and Yada, R. Y. (2004) Effect of N-linked glycosylation on the aspartic proteinase porcine pepsin expressed from *Pichia pastoris*. *Glycobiology* 14, 417-429.
123. Kanjilal, S., Taulier, N., Le Huerou, J. -Y., Gindre, M., Urbach, W., and Waks, M. (2003) Ultrasonic studies of alcohol-induced transconformation in β -lactoglobulin: the intermediate state. *Biophys. J.* 85, 3928-3934.
124. O'Brian, R., and Haq, I. (2004) Applications of Biocalorimetry: Binding, Stability and Enzyme Kinetics, in *Biocalorimetry II: Applications of Calorimetry in the Biological Sciences* (J. E. Ladbury, and M. L. Doyle, Eds.) pp 1-34, John Wiley & Sons, Ltd., West Sussex.
125. Jaswal, S. S., Truhlar, S. M., Dill, K. A., and Agard, D. A. (2005) Comprehensive analysis of protein folding activation thermodynamics reveals a universal behavior violated by kinetically stable proteases. *J. Mol. Biol.* 347, 355-366.
126. Kurimoto, E., Harada, T., Akiyama, A., Sakai, T., and Kato, K. (2001) In vitro refolding of porcine pepsin immobilized on agarose beads. *J. Biochem.* 130, 295-297.
127. Batey, S., Randles, L. G., Steward, A., and Clarke, J. (2005) Cooperative folding in a multi-domain protein. *J. Mol. Biol.* 349, 1045-1059.

128. Campos, L. A., and Sancho, J. (2003) The active site of pepsin is formed in the intermediate conformation dominant at mildly acidic pH. *FEBS Lett.* 538, 89-95.
129. Zhu, X., Ohta, Y., Jordan, F., and Inouye, M. (1989) Pro-sequence of subtilisin can guide the refolding of denatured subtilisin in an intermolecular process. *Nature* 339, 483-484.
130. Jaswal, S. S., Sohl, J. L., Davis, J. H., and Agard, D. A. (2002) Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability. *Nature* 415, 343-346.