Serpins Flex Their Muscle

II. STRUCTURAL INSIGHTS INTO TARGET PEPTIDASE RECOGNITION, POLYMERIZATION, AND TRANSPORT FUNCTIONS*

Published, JBC Papers in Press, May 24, 2010, DOI 10.1074/jbc.R110.141408

James C. Whisstock‡, Gary A. Silverman‡, Phillip I. Bird‡, Stephen P. Bottomley‡, Dion Kaiser‡, Cliff J. Luke‡, Stephen C. Pak‡, Jean-Marc Reichhart§, and James A. Huntington¶

From the ‡Department of Biochemistry and Molecular Biology and ARC Centre of Excellence in Structural and Functional Microbial Genomics, Monash University, Clayton, Victoria 3800, Australia, the ¶Departments of Pediatrics and Cell Biology and Physiology, Children’s Hospital of Pittsburgh and Magee-Womens Hospital, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15201, the §Université de Strasbourg, CNRS UPR 9022, Institut de Biologie Moléculaire et Cellulaire, 67084 Strasbourg Cedex, France, and the ¶Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, United Kingdom

Inhibitory serpins are metastable proteins that undergo a substantial conformational rearrangement to covalently trap target peptidases. The serpin reactive center loop contributes a majority of the interactions that serpins make during the initial binding to target peptidases. However, structural studies on serpin-peptidase complexes reveal a broader set of contacts on the scaffold of inhibitory serpins that have substantial influence on guiding peptidase recognition. Structural and biophysical studies also reveal how aberrant serpin folding can lead to the formation of domain-swapped serpin multimers rather than the monomeric metastable state. Serpin domain swapping may therefore underlie the polymerization events characteristic of the serpinopathies. Finally, recent structural studies reveal how the serpin fold has been adapted for non-inhibitory functions such as hormone binding.

Although amino acid sequence similarity varies from 17 to 95%, key conserved residues facilitate the folding of inhibitory serpins into a metastable conformation typically comprising three β-sheets, eight to nine α-helices, and a solvent-exposed reactive center loop (RCL) (Fig. 1A) (1). The conformational “stressed-to-relaxed” transition that classical inhibitory serpins undergo upon interaction with target peptidases (Figs. 1A and 2A) has been commented on in previous reviews and will not be discussed extensively here (2–5). Suffice it to say that structural snapshots are available for native forms with the RCL fully expelled or partially (e.g. P14) inserted into β-sheet A and for inactive forms with the RCL partially (e.g. P12) or fully (e.g. RCL-cleaved serpins, the final peptidase complex, and the intact but latent conformer) inserted into β-sheet A (6).

Together, these data provide a comprehensive picture of the range of conformational states that the serpin scaffold adopts, as well as the structural details of the conformational rearrangement that occurs upon RCL cleavage by a target peptidase (Figs. 1A and 2A). Indeed, during peptidase inhibition, it is this latter event that triggers RCL insertion, which stabilizes the acyl-enzyme complex by transposing the enzyme and deforming its active site before deacylation occurs (4).

Serpins are broadly distributed throughout all major branches of life; hence, their irreversible inhibitory mechanism must provide a distinct advantage over standard mechanism inhibitors when regulating proteolytic circuits. These concepts are discussed in the accompanying minireview (7). The main purpose of this minireview is to highlight new structural information regarding serpin-peptidase complex formation, hormone binding, and pathologic polymerization.

Serpin Exosite Interactions Enhance Target Peptidase Recognition

In 2001, the crystal structure of the Michaelis complex formed between the Manduca sexta P1 Lys serpin 1B with rat S195A trypsin was published (8). As expected, the RCL is bound in a substrate-like fashion by the peptidase, poised for attack of the P1–P1′ bond. The core interactions involve residues from P4 to P3′, with no contacts between the body of the serpin and the peptidase (i.e. no exosite contacts). This structure is consistent with the notion that the RCL is flexible and positioned away from the body of the serpin as an isolated peptide loop. Additional evidence suggesting that exosite contacts are not extensively involved in serpin-peptidase recognition came from changes in serpin specificity by mutations within the RCL (principally P1) and an NMR study of the Michaelis complex between α1-antitrypsin (α1-AT) Pittsburgh and trypsin showing that the each molecule rotated as if in isolation. After 2001, however, several new crystal structures of serpin-peptidase Michaelis complexes were solved: two non-physiological pairings with trypsin, four pairings with thrombin, and one pairing each with factors Xa (Fxa) and IXa. These structures show that extensive exosite interfaces are a common feature involved in the recognition of serpins by target peptidases and involve residues outside P4–P3′ in addition to the RCL (Fig. 1, B–D) (3).

* This work was supported by National Institutes of Health Grants DK079806 and DK081422 (to G. A. S.) and HL68629 (to J. A. H.); National Health and Medical Research Council Australia Program Grant 490900 (to J. I. B., and G. A. S.); National Institutes of Health Grant DK81422 (to G. A. S.) and HL68629 (to J. A. H.); National Health and Medical Research Council Australia Principal Research Fellow. Additional support was provided by the Monash University Department of Haematology and the Australian Research Council Federation Fellowship to J.-M. R. This minireview will be reprinted in the 2010 Minireview Compendium, which will be available in January, 2011.

1 Australian Research Council Federation Fellow and Honorary National Health and Medical Research Council Australia Principal Research Fellow. To whom correspondence may be addressed. E-mail: james.whisstock@monash.edu.au.

2 To whom correspondence may be addressed. E-mail: gsilverman@upmc.edu.

3 The abbreviations used are: RCL, reactive center loop; α1, AT, α1-antitrypsin; Fxa, factor Xa; s, strand; PCL, protein C inhibitor; HCII, heparin cofactor II; PAI-1, plasminogen activator inhibitor 1; PEDF, pigment epithelium-derived factor; CBG, cortisol-binding globulin; TBG, thyroxine-binding globulin.

4 RCL numbering is in accordance with the convention of Schechter and Berger (Schechter, I., and Berger, A. (1967) Biochem. Biophys. Res. Commun. 27, 157–162).
The RCL of most serpins extends from the P15 residue at the top of strand (s) 5A to the P3’ residue at the beginning of s1C. There is a high degree of flexibility on the non-prime side, consistent with the need for rapid incorporation in physiological target such as heparin cofactor II (HCII/SERPIND1) and α1 AT have short P’ sides (P2’ and P3’ of α1 AT are conformationally restrictive prolines) (12). The hallmark of multispecific (not promiscuous) serpins thus appears to be the extension of the P’ region.

By calculating total interaction surface areas, with and without the RCL, we can assess the relative importance of RCL and exosite contacts in determining serpin specificity (Table 1). The surface area buried in non-physiological complexes (e.g. the two involving trypsin) is typically <1000 Å², 90% of which involves the RCL. By contrast, the physiologically relevant pairings all bury >1000 Å² and rely to varying degrees on exosite contacts. Indeed, there appears to be a tradeoff between the quality of the RCL sequence and the dependence on exosite contacts. This analysis is most interesting with respect to thrombin recognition by various serpins. For example, the disfavored P1 Leu of HCII and the P2 Gly of antithrombin necessitate large exosite contacts of over 1000 and 500 Å², respectively, whereas the favorable P2 Pro and P1 Arg sequence of PCI requires an exosite contact of only 150 Å² for efficient recognition by thrombin.

A second important finding from these structural data is the demonstration that exosites on the serpin scaffold play a crucial role in facilitating initial serpin-peptidase interactions (Table 1). Interestingly, different peptidases appear to rest in different ways on the top of the serpin scaffold (even where the serpin component is the same). In several complexes, the peptidase does far over the “front” of the serpin scaffold and predominantly forms contacts with a conserved single turn helix that precedes s4C as well as surrounding residues (Fig. 1B) (10). Conversely, in other complexes, the peptidases dock “on the back foot” (e.g. the PCI-thrombin-heparin complex) by forming interactions with residues on β-sheet B/s1C, the N-terminal end of s2C, and the C-terminal portion of s3C (9).

Finally, it is interesting to note that three human serpins use protein sequences outside the serpin scaffold as key exosites. HCII utilizes an N-terminal extension to bind to exosite I of thrombin (12), and similarly, α2-antiplasmin contains an extensive C-terminal extension that functions to bind the Kringle domains of plasmin (13). The x-ray crystal structure of α2-antiplasmin reveals that the C terminus is positioned appropriately near the RCL to bind to the peptidase (Fig. 1C). Finally, the crystal structure of the protein Z-dependent inhibitor (SERPINA10) of FXa reveals that the serpin recruits protein Z via a binding site centered on sheet C (14). Whereas the full ternary complex (SERPINA10-protein Z-FXa) remains to be determined, modeling studies suggest that the epidermal growth factor domain of protein Z is placed to assist in recruiting FXa (Fig. 1D).

Taken together, these results show that the RCL sequence itself provides insufficient information for determining actual peptidase target(s), as some targets poorly recognize certain RCL sequences in the absence of regulatory cofactors. Thus, the identification of specificity-determining exosites is of therapeutic interest because these sites represent potentially important new targets for development of molecules that block or enhance serpin-peptidase interactions.

**Domain-swapped Model of Serpin Polymerization**

The use of serpins to modulate diverse molecular pathways predicts that serpin mutations will yield a wide range of disease phenotypes. Indeed loss-of-function mutations associated with α1 AT, antithrombin, and C1 esterase inhibitor (SERPING1)
result in emphysema, thrombosis, and angioedema, respectively. Historically, these loss-of-function mutations underpin the clinical deficiencies that are classified as either type 1 (absent or decreased circulating levels below a critical threshold of a functionally normal protein) or type 2 (normal circulating levels of a dysfunctional protein). A spectrum of genetic mutations (missense, nonsense, indels) leads to these classes of deficiencies (15). However, several of the missense mutations also induce toxic gain-of-function phenotypes by encoding full-length molecules that are prone to misfolding and/or polymer formation. For example, the Z mutation (E342K) of α1AT leads to the accumulation of misfolded and polymerized protein within the endoplasmic reticulum of hepatocytes (16). The marked decrease in circulating levels of α1AT (the major antipeptidase in extracellular fluids) predisposes to emphysema, a loss-of-function phenotype (17). In contrast, the accumulation of misfolded or aggregated α1AT in hepatocytes probably leads to overloading of proteostasis control systems, resulting in cellular injury and the development of cirrhosis, a toxic gain-of-function phenotype. Collectively, toxic gain-of-function phenotypes observed with destabilizing mutations have been termed the serpinopathies (18). Because full-length serpins associated with the serpinopathies retain some functional inhibitory activity, a better understanding of the mechanism of serpin polymer formation could lead to therapeutic strategies designed to block serpin accumulation and to enhance secretion and thereby treat both disease phenotypes.

The early observation that inhibitory serpins accept their RCL as an additional strand in β-sheet A led to the suggestion that RCL insertion in trans may represent the physiological basis for serpin polymerization. Although alternative models have been suggested, the “loop-sheet A” model of serpin polymerization has been generally accepted until recently (Fig. 2B). Specifically, it was suggested that mutations might destabilize β-sheet A of the native serpin and enhance the ability of this region to accept another serpin RCL in trans. The black dotted arrow indicates a gap in β-sheet A that accommodates s5A to form the s4A (RCL) exposed native form in A or the s5A and s4A domain-swapped structure in C.

FIGURE 2. Serpin mechanism, folding, and misfolding. A, active serpins fold into a metastable state. Following the initial interaction with a target peptidase and RCL cleavage, the serpin undergoes a radical conformational change (RCL/s4A incorporation into β-sheet A) that culminates in peptidase inhibition via distortion of the catalytic residues. B, loop-sheet model of serpin polymerization. The RCL of one molecule is inserted into open β-sheet A of another. C, domain-swapped model of serpin polymerization. The model is based on the structure of a domain-swapped antithrombin dimer, where s5A and s4A (RCL) of a donor molecule insert into β-sheet A of a recipient. D, normal serpins may fold through a polymerogenic intermediate that is stabilized by certain mutations. The black dotted arrow indicates a gap in β-sheet A that accommodates s5A to form the s4A (RCL) exposed native form in A or the s5A and s4A domain-swapped structure in C.
Merogenic serpin variant and the wild-type counterpart reveal only modest differences in thermal stability or inhibitory activity (19). Finally, it has proven difficult to build physiochemically reasonable loop-sheet A polymer models that contain both an RCL linkage and completed β-sheet A hydrogen bonding. (In the illustrative model shown in Fig. 2B, the top of β-sheet A is open.)

Recently, the x-ray crystal structure of a domain-swapped antithrombin dimer suggested a new model for serpin polymerization (20). Strikingly, this structure reveals that both s5A and the RCL are incorporated into β-sheet A of another serpin molecule (Fig. 2C). Although the crystal structure was that of a self-terminating antithrombin dimer, and thus not able to propagate further, it was easy to open this model to form long chain polymers. Limited proteolysis data, together with disulfide trapping experiments, support the formation of such polymers in vitro in response to chemical denaturants and heat.

The domain-swapped serpin dimer has a number of implications for the mechanism of serpin polymerization. In particular, this model suggests that merogenic serpin variants, which generally cluster on and around s5A and s6A, interfere with the final stages of β-sheet A assembly and permit the domain-swapping event (Fig. 2, C and D). Thus, the merogenic M* intermediate may resemble a serpin with a substantially incomplete or disordered β-sheet A (Fig. 2D). A key question is whether similar domain-swapped serpin polymers of polymeric variants form in the endoplasmic reticulum in vivo.

Although the role of most prokaryote serpins is not understood, some of these molecules (serpins from Clostridium thermocellum) localize to the cellulosome, a multiprotein extracellular complex that digests material such as cellulose (23). These serpins may protect the cellulosome against unwanted peptidase activity.

Considering the conformational lability of most serpins in eukaryotes, it was surprising to find a large number of inhibitory serpins encoded in extremophilic organisms. In addition to C. thermocellum serpin, exemplars include thermopin (from Thermobifida fusca; 55 °C), tengpin (from Thermoanaerobacter tengcongensis; 75 °C), and aeropin (from Pyrobaculum aerophilum; 100 °C). Biophysical studies reveal that both the native and cleaved states of these latter molecules have substantially elevated stability and that these serpins still function as metastable inhibitors in essentially the same way as their mesophilic counterparts. Interestingly, however, one of these molecules, thermodin, employs a novel strategy to fold at elevated temperatures. Structural studies reveal that thermodin possesses an extreme C-terminal sequence that folds across the front of the molecule and interacts with the top of sheet A (Fig. 3A) (21, 24). Biophysical studies show that the C-terminal sequence plays no detectable role in influencing the stability of the native fold but instead appears to be important for stabilizing a folding intermediate. This concept is intriguing in light of the recent discovery of how native serpins may fold and domain swap (20). As suggested above, the final stage of serpin folding is most likely the assembly of s5A into β-sheet A (Fig. 2D). Thus, the C-terminal sequence of thermodin could provide an extra set of interactions that assist in efficient recruitment of s5A to the folding serpin. An alternative explanation is that the C-ter-

### TABLE 1

| PDB code | Complex | Serpin exosite | RCL contacts | Buried surface area (Å²) at interface (% contributed by exosite) |
|----------|---------|---------------|--------------|---------------------------------------------------------------|
| 2GD4 (3.3 Å) and 1SR5 (3.1 Å) | Antithrombin-S195A factor Xa-pentasaccharide complex | Interactions between thrombin “140 loop” and helical turn 231–234 (preceding s4C), s4C, and s3C; also minor interaction between thrombin “30 loop” and C-terminal end of s1B. | P4–P6’ | 1257 (29.6% exosite) and 1031 (12.9% exosite) |
| 1OPH (2.3 Å) | α1-AT (Pittsburgh) in complex with S195A trypsin | Few exosite interactions: one contact between Asn207 (trypsin) and Asp280 (antitrypsin, on loop between helix H and s2C). | P3–P2’ | 788 (10.3% exosite) |
| 1JMO (2.2 Å) | Crystal structure of heparin cofactor II-S195A thrombin complex | Interactions between thrombin 140 loop and helical turn 285–289 (preceding s4C), s4C, and s3C; also minor interaction between thrombin 30 loop and C-terminal end of s1B and minor interaction between thrombin “60 loop” and s2C; also major interactions between thrombin and flexible N-terminal “tail” of HCII (this region is sandwiched between the serpin and the peptidase). | P4–P3’ | 1834 (59.3% exosite) |
| 1K9O (2.75 Å) | Insect serpin (Ala serpin) in complex with trypsin (non-physiological paring) | Interaction between trypsin 140 loop and helical turn 189–193 (preceding s4C). | P4–P4’ | 999 (12.9% exosite) |
| 3B9F (1.6 Å) | PCI-thrombin-heparin complex | Thrombin 60 loop and N-terminal end of s3C. | P4–P4’ | 1183 (12.9% exosite) |
| 1TB6 (2.5 Å) (also 1SR5 at lower resolution) | Antithrombin-thrombin-heparin ternary complex | Multiple interactions with thrombin and helical turn 231–234 (preceding s4C); also interactions with s1C, s2C, s3C, and s4C. | P4–P5’ | 1500 (32.7% exosite) |
FIGURE 3. Recent serpin structures. A, structure of native thermopin (24). The serpin is in green, with the RCL in magenta (part of the RCL is not visible in electron density; the dotted line shows connectivity). The C-terminal sequence is in addition to the serpin domain and folds across the front of β-sheet A. B, structure of native tengpin (26). The serpin is in green, with the RCL in magenta and helix E/s1A in dark blue. The N-terminal sequence (orange coil) is crucial for maintaining the native metastable state. C, structure of PAI-1 in complex with the somatomedin B domain of vitronectin (orange). The serpin is in green, with the RCL in magenta and helix E/s1A in dark blue. D, structure of native TBG in complex with thyroxine (28). The serpin is in green, the RCL is in magenta, and thyroxine is in cyan spheres.

A seminal sequence contributes to folding by making the intermediate less likely to undergo a domain-swapping event. Bioinformatic studies on aeropin also reveal the presence of a C-terminal extension; however, the role of this region remains to be investigated (25).

Another interesting insight from the study of bacterial serpins is the discovery that tengpin contains an N-terminal sequence that makes extensive contacts with the first β-strand of sheet A (s1A) and helix E (Fig. 3B) (26). Deletion of this region results in a protein that initially folds to the native state but rapidly undergoes conformational change to adopt the latent conformation. The N-terminal region thus functions to stabilize the native state. Interesting parallels can be drawn between this situation and that of mammalian plasminogen activator inhibitor 1 (PAI-1/SERPINE1), which, in the absence of the cofactor vitronectin, rapidly switches from the native to the latent conformation. The x-ray crystal structure of PAI-1 in complex with vitronectin reveals that the cofactor similarly interacts with s1A and helix E (Fig. 3C) (27). Although the folds of the N terminus of tengpin and vitronectin are different, these data nonetheless point toward a common means of stabilizing the metastable serpin fold.

Structural Studies on Non-inhibitory Serpins:
Hormone-binding Serpins, Pigment Epithelium-derived Factor, and Maspin

Of the 36 known human serpins, six have evolved to perform homeostatic non-inhibitory functions: maspin (SERPINB5), pigment epithelium-derived factor (PEDF/SERPINF1), HSP47 (SERPINH1), cortisol-binding globulin (CBG/SERPINA6), thyroxine-binding globulin (TBG/SERPINA7), and angiotensinogen. Structural studies on the hormone-binding serpins CBG and TBG have provided seminal insights into how these molecules interact with and transport the hormones cortisol and thyroxine, respectively. For both serpins, the native conformation has a higher affinity (CBG, ~10-fold; and TBG, ~3-fold) for ligand than the cleaved conformation. Until recently, the precise molecular mechanism of ligand binding and release was obscure, although it was postulated that hormone release may be mediated via RCL cleavage by promiscuous proteolytic activity at the site of delivery. Recent structures of both serpins reveal the unexpected finding that cortisol and thyroxine bind into a pocket formed by helix H, s3B, s4B, s5B, and the loop between s4B and s5B (Fig. 3D) (28, 29). The structures also reveal how RCL cleavage results in substantial mobility in a loop at the top of helix D, which forms extensive contacts with residues in β-sheet B that form the ligand-binding site. This finding suggests that the enhanced flexibility of this region results in a hormone-binding region less able to adopt the conformation required for high affinity binding.

Unfortunately, determination of the structures of PEDF and maspin has not provided insights into their mechanisms of action, as their partners and molecular functions remain somewhat enigmatic. Compelling evidence from cell and animal models implicate PEDF (an extracellular glycoprotein) in the maintenance of tissue homeostasis via cell-surface receptors and interaction with extracellular matrix components (30). For example, it inhibits angiogenesis by suppressing the migration and proliferation of endothelial cells and by promoting apoptosis. Although maspin is also implicated in tissue development and homeostasis, as evidenced by early developmental failure of null mice and its loss in carcinomas (31, 32), it is a nucleocytoplasmic protein (33), which may bind transcription factors (34). The only structure of the human non-inhibitory serpins that remains to be determined is HSP47, and this is perhaps the most tantalizing, as it may shed light on how this serpin performs its critical intracellular chaperone function in collagen biosynthesis (35).

Conclusion

Taken together, these studies underscore the versatility of the large complex metastable serpin scaffold in regulating peptidase and non-proteolytic functions. In addition to the specificity afforded by the primary amino acid sequence of the RCL, inhibitory serpins achieve fine-tuning of specificity via the use of specific exosites present both inside and outside the serpin domain. Accordingly, a single serpin is able to effectively modulate the function of more than one target peptidase and, furthermore, ensure tight physiological specificity even when confronted by a group of peptidases whose cleavage specificity is similar. The cost of complexity is, however, that serpins are unusually vulnerable to mutations that impact their folding pathway. As a consequence, and in common with many other proteins that cause conformational disease (36), serpins are able to domain swap. These events most likely underlie the physiological basis for serpin polymerization.

Acknowledgments—We are indebted to colleagues in the serpin community for knowledge and data and whose collective insights underpin the concepts presented here.

REFERENCES
1. Irving, J. A., Pike, R. N., Lesk, A. M., and Whisstock, J. C. (2000) Genome Res. 10, 1845–1864
2. Gettins, P. G., and Olson, S. T. (2009) J. Biol. Chem. 284, 20441–20445
3. Huntington, J. A. (2006) Trends Biochem. Sci. 31, 427–435
4. Huntington, J. A., Read, R. J., and Carrell, R. W. (2000) Nature 407, 923–926
MINIREVIEW: Serpin Structural Insights