Serpins fold into a native metastable state and utilize a complex conformational change to inhibit target proteases. An undesirable result of this conformational flexibility is that most inhibitory serpins are heat sensitive, forming inactive polymers at elevated temperatures. However, the prokaryote serpin, thermopin, from *Thermobifida fusca* is able to function in a heated environment. We have determined the 1.8 Å x-ray crystal structure of thermopin in the native, inhibitory conformation. A structural comparison with the previously determined 1.5 Å structure of cleaved thermopin provides detailed insight into the complex mechanism of conformational change in serpins. Flexibility in the shutter region and electrostatic interactions at the top of the A-sheet (the breach) involving the C-terminal tail, a unique structural feature of thermopin, are postulated to be important for controlling inhibitory activity and triggering conformational change, respectively, in the native state. Here we have discussed the structural basis of how this serpin reconciles the thermodynamic instability necessary for function with the stability required to withstand elevated temperatures.

Serpins (peptidase inhibitor family I4) are the largest superfamily of protease inhibitors with over 800 members identified to date (1–4). A recent authoritative classification of known peptidase inhibitors reveals that serpins are the only inhibitor family present in all three superkingdoms of life (Eukarya, Bacteria, and Archaea) as well as certain viruses (1). The majority of serpins are found in multicellular eukaryotes and inhibit serine or (more rarely) cysteine proteases. Inhibitory serpins are unusual molecules that fold into a native metastable state and utilize a complex conformational change to achieve protease inhibition (4–6). Upon docking with a target protease, a flexible exposed region termed the reactive center loop (RCL)1 is cleaved and the region N-terminal to the site of proteolysis (termed P1-P15)2 inserts into the center of the large A-β-sheet, forming an additional β-strand and causing a large conformational change (termed the Stressed (S) to Relaxed (R) transition) throughout the molecule (8–10). Following cleavage, and throughout the S to R transition, the protease remains covalently attached to the serpin via an acyl bond between the side chain of the active site serine (or cysteine) and the carbonyl oxygen of the P1 residue in the RCL. The x-ray crystal structure of antitrypsin in complex with trypsin revealed that the enzyme is translocated 75 Å to the base of the inhibitor where it is trapped in a distorted, inactive conformation (11–14). The events following successful protease inhibition are irreversible, and serpins are thus termed suicide inhibitors. The size and conformational flexibility of serpins allow for an exquisite degree of functional control, and many serpins require specific co-factors to effectively inhibit target enzymes. For example, the thrombin/factor Xa inhibitor antithrombin circulates in a relatively inactive conformation until interaction with the co-factor heparin induces a conformational switch to an activated state (15–17).

An undesirable consequence of this conformational flexibility is that serpins are susceptible to inappropriate conformational change, whereby the RCL of one molecule inserts into the A-β-sheet of another, forming inactive long chain loop-sheet polymers (18, 19). Biophysical studies on antitrypsin reveal that this molecule passes through a polymerogenic intermediate during folding or unfolding (20, 21). The resulting serpin polymers can accumulate in the endoplasmic reticulum of cells primarily responsible for protein synthesis, eventually result-
ing in cell death and organ damage. Numerous mutations in human serpins have been identified that promote polymerization, misfolding, and disease (the “serpinopathies”) (22). For example, polymerization of antitrypsin and neuroserpin results in liver cirrhosis and dementia, respectively (22–24). Polymerization of native inhibitory serpins can also be induced by mild heating (19, 25, 26).

The available genomic data reveal that serpins are sparsely distributed in prokaryotes (27). Surprisingly, predicted serpin genes have recently been discovered in the genomes of three extremophiles, the hyperextremophile archaea Pyrococcus abyssi (growth temperature 100 °C), the anaerobic bacteria Thermotoga maritima (optimum growth temperature 80 °C), and the moderate thermophilic bacteria Thermobifida fusca (optimum growth temperature 55 °C). Biochemical, structural, and bioinformatic studies suggest that these molecules function as bona fide protease inhibitors and retain the ability to undergo conformational change (27, 28). In a previous study it was shown that the serpin thermopin from T. fusca was able to inhibit bovine chymotrypsin and, in comparison to human antitrypsin, possessed enhanced stability at elevated temperatures (28). The 1.5 Å X-ray crystal structure of the cleaved form of thermopin was determined and revealed that the molecule adopted the typical serpin fold, albeit with several significant variations. In particular, thermopin lacks the G-helix and possesses a C-terminal region (or “tail”) that interacts with a cluster of conserved residues at the top of the A β-sheet. Based upon biophysical and mutagenesis data, it was suggested that C-terminal tail is required for efficient folding at elevated temperatures (28). Structural analysis revealed that the thermostability of cleaved thermopin is achieved by improved hydrogen bonding and salt bridging, consistent with the results of studies on a range of thermostable proteins (29–34).

The native, metastable state of a serpin from a thermophilic organism is intriguing because it represents an evolutionary trade-off between stability and function: how does thermopin maintain a metastable conformation at elevated temperatures while retaining the ability to inactivate target proteases? To answer this question we have determined the 1.8 Å resolution X-ray crystal structure of thermopin in the native conformation. We contrasted and compared the high resolution structures of native and cleaved thermopin and discussed possible structural mechanisms underpinning this paradox. Furthermore, the structures have allowed the most detailed investigation to date of the mechanism of conformational change in serpins. In this regard, the availability of high resolution data is of great importance, previous comparisons between serpin conformers having been limited by the relatively low resolution of one of the structures (8–10). Finally, the role of the C-terminal tail in modulating the stability and function of native thermopin was discussed.

**EXPERIMENTAL PROCEDURES**

**Cloning**—A clone containing the full open reading frame of the T. fusca serpin gene was obtained from the Joint Genome Institute, United States Department of Energy. The gene was initially cloned between the BamHI and SacI sites of the pQE30 (Qiagen) multicloning site as described previously (28). A pQE30 variant (pQE30HIS) was created that lacked the nucleotide sequence encoding the RGS epitope and BamHI restriction endonuclease sites of the vector sequence. The oligonucleotide linker was constructed by annealing two 5′-phosphorylated primers, 5′-aattctcaagagaggagaggattcagtgg and 5′-gatccagctgatatatttctcttcattg. A BamHI/SacI fragment containing the serpin gene was then excised from the initial pQE30-T.fusca construct and ligated with BamHI/SacI-digested pQE30HIS to create the pQE30aHIS-T.fusca construct used in this study.

**Mutagenesis**—The QuikChange site-directed mutagenesis kit (Stratagene) utilizing a PCR-based strategy was used to introduce the C41S mutation in thermopin. The following oligonucleotide and its antisense counterpart was used (mutation is underlined): 5′-catcgctgtggcgtcctccggcctccgc-3′. Positive clones were confirmed by DNA sequencing and transformed into SG13009(pREP4) Escherichia coli cells (Qiagen) for expression.

**Expression and Purification**—Six liters of 2YT broth (100 µg ml⁻¹ ampicillin, 30 µg ml⁻¹ kanamycin) was inoculated with an overnight culture of SG13009 cells harboring pQE30HIS-T.fusca. Expression was induced with isopropyl-1-thio-D-galactopyranoside (0.5 mM) at A<sub>600</sub> of ~0.7. Cells were harvested after 5 h and the pellet resuspended in 50 mM Hepes, 50 mM NaCl, 1 mM EDTA before flash freezing (N<sub>2</sub> liquid). Lysozyme (0.5 mg ml⁻¹) was added to the thawed pellet and allowed to incubate for 1 h at 4 °C. The sample was sonicated and heated to 55 °C for 20 min. Centrifugation at 18,000 rpm removed all insoluble material. The supernatant was loaded onto a pre-equilibrated (50 mM Hepes, 50 mM NaCl, 1 mM EDTA, pH 7.5) 5-mL HiTrap SP-Sepharose FF column (Amersham Biosciences), washed to baseline, and eluted using a 60-ml linear gradient into 50 mM Hepes, 1 mM NaCl, 1 mM EDTA, pH 7.8. Peak fractions (SDS-PAGE) were combined, and 4 mM NaCl was added until the final salt concentration was ~2 M. The high salt protein solution was loaded onto a pre-equilibrated (50 mM Hepes, 2 mM NaCl, 1 mM EDTA, pH 7.8) 5-mL HiTrap phenyl-Sepharose HP column (Amersham Biosciences), washed to baseline, and eluted using a 60-ml linear gradient into 50 mM Hepes, 1 mM EDTA, pH 7.8. Fractions corresponding to pure native-state thermopin were combined, buffer-exchanged into 50 mM Hepes, 5 mM EDTA, pH 7.8, and concentrated (Ultrafree-5 centrifugal filter unit; Amicon) to a final concentration of ~12 mg ml⁻¹ (determined spectrophotometrically using the following relationship: A<sub>280</sub> = 1 mg/ml). Small aliquots were flash frozen and stored at -80 °C for future use.

**Spectroscopic Methods**—Circular dichroism experiments were performed on a Jasco 810 spectropolarimeter (Jasco, Tokyo) at 20 °C. θ<sub>222</sub> measurements were made with the signal averaged over 15 s. The protein concentration used was 0.2 mg/ml with a 0.1-cm path length. Thermal denaturation was performed at a heating rate of 1 °C/min, at a protein concentration of 0.05 mg/ml.

**Chemical Denaturation**—Stock solutions of guanidine-HCl (GdnHCl) in 20 mM NaPO₄, pH 8.0, were prepared and filtered through 0.22-µm membranes before use. The GdnHCl concentration was determined by refractive index measurements as described previously (35). Equilibrium unfolding curves were obtained by incubating the protein at various GdnHCl concentrations for 2 h at 20 °C and plotting the signal change at 222 nm (far UV range) as a function of denaturant concentration. No differences were observed in experiments where spectroscopic measurements were taken after equilibration for greater lengths of time or within a concentration range between 0.01 and 0.2 mg/ml.

**Crystalization**—Crystals of native thermopin were obtained by the hanging drop vapor diffusion method (36). Optimization of crystallization conditions led to a reservoir mixture containing 0.1 M HEPES, 0.1 M sodium thioglycolate, 40% polyethylene glycol 4K, 0.2 M ammonium sulfate, 0.1 M sodium cacodylate, pH 6.6. Crystals were grown by mixing equal volumes of 12 mg ml⁻¹ protein solution (2 µl) with the reservoir solution (2 µl), with the prior addition of 0.1 M l-cysteine (0.4 µl) to the reservoir. Large rod-shaped crystals grew overnight, and SDS-PAGE analysis confirmed the crystallized material was in the native state (data not shown). The crystals were flash frozen in liquid nitrogen using the reservoir solution as a cryoprotectant.

**X-ray Data Collection, Structure Determination, and Refinement**—The crystals diffracted to 1.8 Å resolution and belong to space group P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, with unit cell dimensions of a = 43.53 Å, b = 81.17 Å, c = 163.63 Å, consistent with one monomer/asymmetric unit. The data were merged and processed with the HKL suite (37). Subsequent crystallographic and structural analysis was performed using the CCP4i interface (38) to the CCP4 suite (39), unless stated otherwise. 5% of the data set was flagged for calculation of the free R factor (R<sub>f</sub>free) with neither a σ nor a low resolution cutoff applied to the data (40). A summary of statistics is provided in Table I.

The structure was solved using the molecular replacement method (41) with the program AMORE (41). An initial search model was built using regions of the cleaved thermopin structure (1MTP) predicted to remain relatively unchanged in the native structure (residues 5–78, 168–309, and 333–367). The molecule packed well within the unit cell, and together with the unbiased features in the initial electron density maps, the correctness of the molecular replacement solution was confirmed.

For structure refinement a hybrid model was created, based on the
sequence alignment and structures of cleaved thermopin (1MTP) and native α1-antitrypsin (1QLP) (28). Residues 5–78, 168–309, and 333–367 of cleaved thermopin were combined with residues 168–309 of a polyalanine α1-antitrypsin model. This was then superimposed with the molecular replacement solution to give a starting model. Maximum likelihood refinement using REFMAC (42) produced a 10% drop in the \( R_{	ext{free}} \) and \( R_{	ext{cryst}} \). The progress of refinement was monitored using the \( R_{	ext{free}} \) value. Iterative cycles of maximum likelihood refinement using REFMAC and model building in O (43) were used to further improve the model until there was no further drop in the \( R_{	ext{free}} \) value. A bulk solvent correction (Babinet model with mask) was used within REFMAC. Water molecules were added to the model using ARP/wARP when \( R_{	ext{cryst}} \) reached 30%. Solvent molecules were retained only if they had acceptable hydrogen bonding geometry contacts of 2.5–3.5 Å with existing solvent and were in good 2Fo − Fc and Fc − Fo electron density. At this stage, residues that could be modeled in two alternate conformations were included in the refinement. Disordered atoms in these residues were given occupancies of 0.5. The dimethylarsinoyl moiety was built into the model in the latter stages of refinement, using coordinates obtained from Ligand Depot (44).

The final model contains residues 2–326, 333–367, 1 dimethylarsinoyl moiety, 5 sulfate groups, and 457 water molecules. It has an \( R_{	ext{free}} \) of 18.7% and an \( R_{	ext{cryst}} \) of 30%. Solvent molecules were retained only if they had acceptable hydrogen bonding geometry contacts of 2.5–3.5 Å with existing solvent and were in good 2Fo − Fc and Fc − Fo electron density. At this stage, residues that could be modeled in two alternate conformations were included in the refinement. Disordered atoms in these residues were given occupancies of 0.5. The dimethylarsinoyl moiety was built into the model in the latter stages of refinement, using coordinates obtained from Ligand Depot (44).

The final model contains residues 2–326, 333–367, 1 dimethylarsinoyl moiety, 5 sulfate groups, and 457 water molecules. It has an \( R_{	ext{free}} \) of 18.7% and an \( R_{	ext{cryst}} \) of 22.2% for all reflections between 64 and 1.76 Å. All residues are in the most favored and allowed regions of the Ramachandran plot. A summary of statistics is provided in Table I. The coordinates have been deposited in the Protein Data Bank (accession code 1SNQ).

**Kinetics**—The stoichiometry of inhibition (SI) of bovine chymotrypsin by thermopin, ThermopinC41S, and arsenate-modified thermopin at 37 °C was performed as described (27).

**RESULTS**

**Structure of Native Thermopin**—The 1.8 Å structure of native thermopin is one of the highest resolution structures of a native serpin determined to date (Fig. 1A and Table I). The quality of the electron density of the final model is excellent throughout, and the model has excellent stereochemistry. The final model contains residues 2–326 and 333–367, 1 dimethylarsinoyl moiety, 5 sulfate groups, and 457 water molecules. No electron density was observed in the region of the RCL (residues 327–332), this portion of the molecule being disordered.

**Structural Analysis**—Hydrogen bonds (not including water-mediated bonds) were calculated using the WHATIF optimal hydrogen bonding network server (45). Calculation of the number of salt bridges, the network server (44). Calculation of the number of salt bridges, the

**TABLE I**

Data collection and refinement statistics

| Values in parentheses refer to the highest resolution shell. |
|---------------------------------------------------------------|
| **Data collection**                                           |
| Temperature                                                  | 100 K                           |
| X-ray source                                                 | Station 14BM-C, BioCARS, APS    |
| Detector                                                    | Quantum 4                        |
| Space group                                                 | P2,2,2                           |
| Cell dimensions (Å): a, b, c                                 | 45.35, 81.17, 106.63             |
| Resolution (Å)                                              | 1.76                             |
| Total number of observations                                 | 126, 990                         |
| Number of unique observations                                | 39, 513                          |
| Multiplicity                                                 | 3.2                              |
| Data completeness (%)                                        | 98.8 (98.6)                      |
| \( R_{	ext{free}} \) (%)                                    | 14.6 (2.8)                       |
| **Structure refinement**                                     |
| Nonhydrogen atoms                                           |
| Protein                                                     | 2739                             |
| Solvent                                                     | 457                              |
| Sulfates                                                    | 25                               |
| Dimethylarsinoyl dimethylarsinoyl                           | 3                                |
| \( R_{	ext{cryst}} \) (%)                                   | 22.2                             |
| \( R_{	ext{free}} \) (%)                                    | 18.6                             |
| R.m.s. deviation from ideality                              |
| Bond lengths (Å)                                            | 0.005                            |
| Bond angles (°)                                              | 1.032                            |
| Improper (°)                                                 | 1.03                            |
| Dihedrals (°)                                                | 28.1                             |
| Ramachandran plot                                           |
| Most favoured                                               | 92.4                             |
| Additional allowed regions (%)                               | 7.3                              |
| B factors (Å)                                               |
| Mean main chain                                             | 15.2                             |
| Mean side chain                                             | 17.2                             |
| Mean water molecule                                         | 32.6                             |
| Mean sulfate molecule                                       | 47.7                             |
| Mean dimethylarsinoyl moiety                                | 19.4                             |
| R.m.s. deviation bonded Bs                                  | 1.19                             |

Footnotes:

\( a \) Agreement between intensities of repeated measurements of the same reflections and can be defined as \( \frac{\sum (I_{h,i} - \langle I_{h,i} \rangle)^2}{\sum I_{h,i}} \), where \( I_{h,i} \) are individual values and \( \langle I_{h,i} \rangle \) is the mean value of the intensity of reflection \( h \).

\( b \) The free \( R \) factor was calculated with the 5% of data omitted from the refinement.
Native-state Serpin Structure

1A and 2A). The A β-sheet contains five β-strands (Fig. 1A). The RCL is fully expelled from the A β-sheet, and only the first two residues of the RCL can be seen protruding from the molecule at the top of strand s5A. Analysis by SDS-PAGE confirmed that the protein crystals contained the serpin in the native state (data not shown). The six-stranded B β-sheet comprises the majority of the hydrophobic core of the molecule, and the four-stranded C β-sheet includes strand s1C located at the C-terminal end of the RCL (Fig. 1A). In comparison with other structurally characterized serpins, thermopin contains an extended C-terminal sequence that interacts with the top of the A sheet; and on the opposite end of the molecule on the exposed surface of the C sheet (Fig. 1B).

Throughout the course of the work, the structure of native thermopin was compared with its cleaved counterpart as well as the mesophilic eukaryote serpin antitrypsin. Despite low sequence similarity (~25%), native thermopin and native antitrypsin superpose with an r.m.s. deviation of 1.0 Å over 217 Ca atoms (the respective cleaved forms superpose to 1.27 Å/atom over 289 Ca atoms). Antitrypsin is one of only two other inhibitory serpin pairs for which both native (at 2.0 Å resolution) and cleaved (at 3.0 Å resolution) conformations have been structurally characterized.

Structural Comparison with Cleaved Thermopin—Native and cleaved thermopin are both more compact than the corresponding native and cleaved states of the mesophilic serpin antitrypsin (Table II). Consistent with the thermostable nature of thermopin (28), the absence of the RCL in native thermopin precludes precise calculation of the change in accessible surface area that occurs upon cleavage. The secondary structure content of native and cleaved thermopin is very similar. In comparison with native antitrypsin, native thermopin contains an extra 14 salt bridges and a similar number of hydrogen bonds (Table II). Conformational change within native thermopin to the R conformation clearly results in a dramatic increase in both the number of salt bridges (an additional 28 salt bridges are formed in the cleaved state) and hydrogen bonds (Table II and Ref. 28). Comparison of native and cleaved antitrypsin reveals that the S to R transition does not result in such a dramatic increase in electrostatic interactions in this serpin; however, it must be noted that whereas the structure of native antitrypsin is of comparable resolution to native thermopin, the structure of cleaved antitrypsin has only been determined to 3.0 Å. Average residue packing calculations reveal that the cleaved state of thermopin is better packed than the native state, and both are better packed than either conformation of antitrypsin (Table II). Again, packing comparisons must be used with caution because a correlation between packing density and resolution has been observed (higher resolution structures appear better packed) (50).

A structural comparison of cleaved and native thermopin reveals that thermopin can be divided into three major fragments that shift upon the molecule undergoing the S to R transition (Table III and Figs. 1B and 2B). The largest fragment (1) comprises the majority of the molecule and can be considered as the scaffold upon which the other two fragments move. Fig. 2B shows a superposition of native and cleaved thermopin on fragment 1A. Broadly, in order to accommodate the RCL, strands s1A-s3A move as a rigid unit (fragment 2) as do the F- and E-helices (fragment 3). Two regions of plastic deformation connect the mobile fragments; the top of the D-helix links fragments 1 and 2 and crumples as a result of RCL insertion, and the loop between h1 and s5A deforms in response to movement of the E-helix (Table III and Fig. 1B). The N- and C-terminal regions are also flexible.

We have used the native structure of thermopin to investi-
gate how this molecule can create two alternative interfaces (i.e. the A-sheet with and without the inserted RCL) against which the F-helix and its associated loop (residues 115–155) can pack. Relative to the mean plane of the sheet, the F-helix is angled so that the top of the helix makes the closest contact (Fig. 2B). Upon cleavage, the A-sheet of the native serpin undergoes two distinct subtitle “twists” to accommodate the inserting RCL. The backbones of strands s5A/s6A rotate around a point at the top of the sheet so as to maintain the approximate position of three residues on s5A that contact the F-helix (Ile-300, Gln-302, and Arg-304). In native thermopin Pro-264 distorts strand s6A and disrupts the hydrogen bond network between s5A and s6A. In the cleaved form no such distortion is apparent and proper β-sheet hydrogen bonding is maintained. On the opposite side of the inserting RCL strands s1A-s3A rotate around a point at the bottom of the sheet, again ensuring that residues contacting the F-helix maintain approximately the same positions. The S to R transition is thus accompanied by the A-sheet adopting a “pigeon-toed” conformation, the backbone atoms moving apart to create room for the RCL while many of the side-chain atoms maintain contact with the F-helix. Where original contacts are not approximately maintained, the side-chain atoms of RCL residues often occupy similar positions and replace residues that have moved, particularly at the top of the F-helix. For example, after cleavage, Ala-315 on the RCL replaces the position of Ala-161 on s3A. In shifting across, Ala-161 occupies the position of the Cβ atom of Arg-86 on strand s2A. At the very top of the F-helix Ala-313 on the RCL replaces the position of the Cβ atom of Trp-163 on strand s3A. After shifting across, the Cβ atom of Trp-163 occupies the position of Ala-84 on strand s2A. These interactions are shown in detail in Fig. 2B. Packing between the F-helix and the bottom of the A β-sheet is looser. The position of Ile-157 is approximately maintained, because it is close to the “pivot” point, and the side chain of Met-321 on the inserted RCL fills space unoccupied by any residue in the native structure.

Two regions have been shown to be important for controlling conformational change and inhibitory activity of serpins (9, 10). First, the breach is located at the top of the A β-sheet and is the initial point of insertion of the RCL. Second, the shutter is located in the middle of the molecule, centered on the top of the B-helix, and is postulated to control sheet opening and RCL insertion. Numerous mutations that cause polymerization clusters in the shutter region (51). A detailed analysis of the breach and shutter of native/cleaved thermopin is presented below.

**The Breach**—Thermopin contains a unique C-terminal extension or tail that adopts an extended conformation, packing against the face of the top of the A β-sheet (the breach). Although the tail is required for correct folding and forms interactions with highly conserved amino acids (in particular Glu-309 and Arg-258) at the top of strands s5A and s6A in both cleaved and native forms, biophysical studies reveal that it does not contribute to the stability of the cleaved or native state (28). Comparison between the native and cleaved conformations reveals that the tail adopts a similar conformation in the native state, but there are some clear differences in its interaction with the rest of the protein (Fig. 3). There are five hydrogen bonds between residues in the tail and the breach in the native structure, compared with only three in the cleaved structure. In the cleaved structure the backbone amide group of Ala-367 forms a hydrogen bond with the side chain of Glu-309 at the top of s5A. This interaction is not present in the native structure, because the tail bends back on itself so that the terminal carboxyl group forms a salt bridge with the side chain of Arg-306 (strand s5A, Fig. 3A). Asp-308 also forms a salt bridge with Lys-165, completing an ionic “bridge” across s3A and s5A. Upon undergoing the S to R transition, the side chain of Glu-313 of the RCL (P13 Glu in standard nomenclature (7) forms a salt bridge with Arg-306 (Fig. 3B, s5A). In addition, Lys-165 switches salt bridge partners to interact with Asp-82 at the top of s2A. Throughout, the conformation of the side chain of Glu-309 is preserved. It is interesting to note the unusually close (4.5 Å) proximity between the positively charged amino groups of Arg-306 and Lys-165. It is possible that this interaction provides the necessary “strain” at the top of the A sheet that facilitates opening of strands s5A and s6A upon strand insertion in the S to R transition.

**The Shutter**—During model building, analysis of the shutter region revealed the presence of a significant peak (20 σ) in the F–F electron density, adjacent to Cys-41. The extra density was attributed to a dimethylarsinoyl molecule covalently attached to the thiol group of Cys-41 (Fig. 4A), consistent with the presence of 100 mM cacodylate (dimethylarsenic acid) in the crystallization buffer. This moiety could be easily built into the electron density and included in refinement and was further supported by calculation of a difference Fourier map (using anomalous dispersion differences for structure factor amplitudes and phases corresponding to the refined atomic model.
between the tail and the surrounding residues are represented as dashed green lines shown. A-sheet strands are labeled.

B) thermopin. cleaved (B) thermopin. The RCL is colored magenta and represented as a dashed line for the native structure. Hydrogen bonds are drawn as dashed green lines. Water molecules that mediate hydrogen bonds between the tail and the surrounding residues are represented as red spheres. For clarity, some elements of secondary structure are not shown. A-sheet strands are labeled.

lacking the metal atom. The resulting map showed a single outstanding peak (15 σ) at the metal center. Further refinement resulted in no significant peaks in the $F_o - F_c$ electron density in this region, with the dimethylarsinoyl moiety adopting a well ordered conformation (mean B-factor of side-chain atoms = 19.4 Å²). The observation of an arsenic-derivatized cysteine has never been reported for serpin structures, but there are more than 10 examples in the literature of such modifications in protein crystals grown in the presence of cacodylate (for examples, see Refs. 52–54). The shutter region in native thermopin is considerably less well packed than in the cleaved structure. Consistent with this finding, there is also more ordered solvent in this region of the native structure (four solvent molecules compared with one in the cleaved structure). Crystals of native thermopin could not be obtained in the absence of cacodylate buffer.

Analysis of the electron density around Cys-41 in the cleaved thermopin structure revealed no such modification was present, consistent with the absence of cacodylate buffer. Indeed, in cleaved thermopin the cavity occupied by the metal atom in the native form is filled, primarily by Asn-160 (strand s3A) and Thr-87 (strand s2A), both of which move across in response to RCL insertion (Fig. 4). The derivatization of Cys-41 in the native form is intriguing because the shutter region is important for controlling serpin stability and conformational change.

We therefore generated the conservative substitution C41S to investigate the function of Cys-41 and the effect of this variant on thermostability, folding, and inhibitory activity. ThermopinC41S and Arsenate-modified Thermopin Possess Impaired Inhibitory Activity—We have previously shown that thermopin is an inhibitor of bovine chymotrypsin (SI of 8.0 ± 0.4 and an apparent second order rate constant ($k_{app}$) of 8.4 ± 0.4 × 10⁴ M⁻¹ s⁻¹ (28). The mutant ThermopinC41S and arsenate-modified thermopin were assayed against bovine chymotrypsin and found to have an SI of 29 ± 0.62 and 16 ± 0.7, respectively (Fig. 5). Thus, arsenate modification at Cys-41 significantly impairs inhibitory activity and the conservative substitution C41S effectively abolishes the inhibitory activity of the molecule. ThermopinC41S was found to have an identical heat denaturation profile to wild type material (Fig. 5B, $T_m = 67 °C$). In addition, the equilibrium unfolding of the mutant was investigated and found to be indistinguishable from wild type material (Fig. 5B).

DISCUSSION

We have determined the x-ray crystal structure of the bacterial serpin thermopin in the native conformation. Thermopin adopts the typical native serpin fold, possessing a 5-stranded A β-sheet with the RCL fully expelled from the top of the A β-sheet. The RCL of thermopin is five residues longer than that of human antitrypsin; despite the high quality of the data, much of the RCL is disordered, indicating that this region is extremely flexible. The length and flexibility of this region is a possible explanation for our observation that the non-cognate protease bovine chymotrypsin is able to cleave in multiple places with the RCL (28). Using our previously determined structure of cleaved thermopin, the high resolution data have allowed a detailed analysis of the conformational rearrangements that the serpin undergoes during the S to R transition. Our current understanding of this complex mechanism is based upon medium resolution crystal structures of antitrypsin (9) and is thus improved upon significantly by the high resolution structure presented here and previously (28). Comparison with cleaved thermopin revealed that the majority of the molecule acts as a scaffold upon which two smaller fragments (strands s1A-s3A and the E-F-helix) shift upon loop insertion. A careful examination of the structures revealed that the A β-sheet undergoes subtle distortion in order to maintain a similar interface under the F-helix and its associated loop (linking to s3A).

The data revealed the presence of a network of salt bridges in the breach, centered on Lys-165 and bridging across s3A and s5A. During the transition to the cleaved conformation, the salt

FIG. 3. Interactions between the C-terminal tail (yellow bonds) and the top of the A-sheet in the breach region of native (A) and cleaved (B) thermopin. The RCL is colored magenta and represented as a dashed line for the native structure. Hydrogen bonds are drawn as dashed green lines. Water molecules that mediate hydrogen bonds between the tail and the surrounding residues are represented as red spheres. For clarity, some elements of secondary structure are not shown. A-sheet strands are labeled.

FIG. 4. Interactions in the shutter region of native (A) and cleaved (B) thermopin. The dimethylarsinoyl moiety attached to Cys-41 is drawn as ball-and-stick with the arsenic atom represented as a gray sphere. Hydrogen bonds are drawn as dashed magenta lines. Water molecules are represented as red spheres. For clarity, some elements of secondary structure are not shown.
bridging pattern in this region undergoes significant rearrangement. In particular, whereas in the cleaved conformation the top of the A-sheet is stabilized by a salt bridge between Lys-165 and Arg-82, the close proximity of Lys-165 and Arg-306 in the native structure (Fig. 3A) may provide electrostatic strain in the breach that allows for triggering of conformational change. The presence of discreetly dispersed residues in a channel from the base of the A-sheet, along strand s2A via the shutter region, finishing at the top of the A-sheet (Fig. 1B) is consistent with the plasticity in these regions that is necessary for conformational change and inhibitory function.

Unexpectedly, native thermopin contains a dimethylarsinoyl moiety covalently attached to Cys-41, a residue on the B-helix. The derivatization of the cysteine has most likely occurred as a result of the presence of cacodylate buffer in the crystallization conditions; thus this interaction is unlikely to occur in vivo. However, the ability to derivatize Cys-41 is consistent with the presence of a deep solvent-accessible cavity centered on the shutter region and bounded by the D-helix, s2A, and the top of the E-helix. The shutter is crucial for controlling conformational change; numerous mutations in this region in human serpins have been shown to result in polymerization and disease (for a review see Ref. 51). Furthermore, the presence of an analogous cavity in native antitrypsin has been identified as a possible target for the rational design of compounds that prevent serpin polymerization (55). These authors demonstrated that cavity-filling mutations in strand s2A result in enhanced stability and resistance to heat-induced polymerization. Although these variants were still able to inhibit bovine chymotrypsin, the efficiency of the interaction was impaired with respect to wild type (55). Interestingly, non-conservative mutations at the analogous residue to Cys-41 in the human plasma serpin antithrombin (Thr-85) to either methionine (antithrombin_wibble) or lysine (antithrombin_wobble) resulted in massive instability, polymerization, and spontaneous transition to the latent conformation, highlighting the importance of this position in other serpins (56).

Kinetic analysis of arsenate-modified thermopin revealed that the presence of the dimethylarsinoyl moiety significantly impaired inhibitory activity, this material possessing an SI of 16 against bovine chymotrypsin compared with an SI of 8 for wild type. We therefore further investigated the role of Cys-41 in thermopin and generated the conservative mutation ThermopinC41S. Our data revealed the exquisite sensitivity of the serpin molecule to variation in this region: ThermopinC41S, a conservative mutation, is essentially indistinguishable from wild type in regards to thermal stability and unfolding profile, but this variant is an extremely poor inhibitor with a SI of 29 against bovine chymotrypsin compared with an SI of 8 for wild type. The presence of a covalently attached metal atom in the shutter region provides structural support for the rational design of anti-polymerogenic compounds targeting this cavity (55). However, significant challenges remain in modulating serpin behavior because the presence of the dimethylarsinoyl moiety and conservative substitutions such as ThermopinC41S result in significantly impaired inhibitory activity against target proteases.

Structural analysis and comparison with our previously determined 1.5 Å structure of thermopin in the cleaved state allowed us to propose a mechanism for how a serpin from a thermophilic organism reconciles the thermodynamic instability necessary for function with the stability required to withstand elevated temperatures. First, the native state of thermopin is relatively flexible and loosely packed. Overpacking in the native state of antitrypsin, particularly in the shutter region, has been postulated to cause local strain that is utilized to regulate the S to R transition (57–60). In contrast, there is no evidence of destabilization induced by molecular overpacking and strain in native thermopin. Second, and in contrast to native and cleaved antitrypsin, the transition to the cleaved conformation results in a dramatic increase in the number of salt bridges. Taken together, we have suggested that the absolute stabilities of the native and cleaved conformations of thermopin are increased, providing the necessary stability at elevated temperatures while maintaining the relative difference in stability between native and cleaved states that is critical for rapid conformational change and inhibitory function.

There is evidence to suggest that in many cases the evolution of proteins has occurred to optimize function at the expense of stability (61–64); for example, there are exquisite stereochemical requirements for catalysis, whereas protein stability requirements usually are less onerous (61–64). A serpin from a thermophilic organism represents an interesting twist to this hypothesis, as the evolutionary selection pressure is complicated by the presence of two structurally and energetically unique folded states (65). Serpins appear to represent an example of how sophisticated functionality can be achieved by exploiting this window of stability. At least in the case of thermopin, native-state flexibility combined with improved overall electrostatic interactions in the cleaved state achieves a fine balance between sufficient thermostability to prevent polymerization at high temperatures and surplus stability that would impair function.

Acknowledgment—We thank the staff at BioCARS and the Australian Synchrotron Research Program for assistance.

REFERENCES

1. Rawlings, N. D., Tolle, D. P., and Barrett, A. J. (2004) Biochem. J. **378**, 705–716
The High Resolution Crystal Structure of a Native Thermostable Serpin Reveals the Complex Mechanism Underpinning the Stressed to Relaxed Transition
Kate F. Fulton, Ashley M. Buckle, Lisa D. Cabrita, James A. Irving, Rebecca E. Butcher, Ian Smith, Shane Reeve, Arthur M. Lesk, Stephen P. Bottomley, Jamie Rossjohn and James C. Whisstock

J. Biol. Chem. 2005, 280:8435-8442.
doi: 10.1074/jbc.M410206200 originally published online December 7, 2004

Access the most updated version of this article at doi: 10.1074/jbc.M410206200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 64 references, 13 of which can be accessed free at http://www.jbc.org/content/280/9/8435.full.html#ref-list-1