Clan CA, family C1 cysteine peptidases (CPs) are important virulence factors and drug targets in parasites that cause neglected diseases. Natural CP inhibitors of the I42 family, known as ICP, occur in some protozoa and bacterial pathogens but are absent from metazoa. They are active against both parasite and mammalian CPs, despite having no sequence similarity with other classes of CP inhibitor. Recent data suggest that *Leishmania mexicana* ICP plays an important role in host-parasite interactions. We have now solved the structure of ICP from *L. mexicana* by NMR and shown that it adopts a type of immunoglobulin-like fold not previously reported in lower eukaryotes or bacteria. The structure places three loops containing highly conserved residues at one end of the molecule, one loop being highly mobile. Interaction studies with CPs confirm the importance of these loops for the interaction between ICP and CPs and suggest the mechanism of inhibition. Structure-guided mutagenesis of ICP has revealed that residues in the mobile loop are critical for CP inhibition. Data-driven docking models support the importance of the loops in the ICP-CP interaction. This study provides structural evidence for the convergent evolution from an immunoglobulin fold of CP inhibitors with a cystatin-like mechanism.

Characterization of the structure and mechanisms of action of natural inhibitors of cysteine peptidases (CPs)\(^1\) has provided important insights into the functional roles of the inhibitors themselves and also those of the target CPs. CP inhibitors occur widely in nature, and several distinct groups with unrelated primary structures are recognized (1). Members of the cystatin superfamily (clan IH, family 125 (2)) have been extensively characterized at the molecular level (3–5). A second group of natural CP inhibitors, assigned to clan IX, family 131, are similar to cystatin in inhibiting CPs of clan CA, family CI, such as papain, but are distinguished by the presence of a thyroglobulin type I domain (6). The recently discovered chagasin family of CP inhibitors, designated ICP for inhibitor of cysteine peptidases (clan I-1, family I42), also inhibit papain-like CPs and yet have no significant sequence identity with the other groups of CP inhibitors (7). This raises the intriguing question: have the different groups of CP inhibitors evolved related tertiary structures to interact with the same target CPs in a similar fashion, as predicted from *in silico* analysis by Rigden et al. (8), or does ICP act in a different way?

ICP family members appear to occur in species from a very limited phylogenetic range (some parasitic protozoa, bacteria (including *Pseudomonas aeruginosa*), and Archaea, but no metazoa), suggesting that the genes have been acquired by horizontal gene transfer or retained for some special function. Although there is good evidence that the *in vivo* target for ICPs are clan CA, family C1 CPs, it remains uncertain whether the main function of the inhibitors is to modulate the activity of enzymes of the parasite itself (as is suggested for protozoan parasite *Trypanosoma cruzi* (9)) or the host (as suggested for the related parasite *Leishmania* (10)). Interestingly, no clan CA, family C1 CPs appear to be present in the *P. aeruginosa* genome, which provides further support for the suggestion that one role of ICPs in pathogens might be to regulate host CP activity and so facilitate infection.

This has been investigated with *Leishmania* by gene targeting to create parasite lines that either lack or overexpress the ICP gene (10). ICP null mutants grow normally axenically in *vitro* and are as infective to *Macrophages in vitro* as wild type parasites. However, they have reduced infectivity to mice. Lines that overexpress ICP also show markedly reduced virulence *in vivo*. Thus, ICP may be important after it is released in the interaction of the parasite with its mammalian host or as a mechanism of preventing damage from host CPs taken up by the parasite through endocytosis (*e.g.* from the parasitophorous vacuole, which contains host lysosomal enzymes).

The ICP proteins are small (≈13 kDa) and very divergent, with *L. mexicana* ICP having only 31% identity with ICP of *T. cruzi* and 24% identity with ICP of *P. aeruginosa* (11). Nevertheless, there are highly conserved motifs that suggest important functional regions. This has facilitated the identification of predicted ICPs from genome data and recombinant ICPs have been produced from the *L. mexicana, T. brucei, Entamoeba histolytica,* and *P. aeruginosa* genes and confirmed to have potent inhibitory activity toward CPs, notably cathepsin L homologues (7, 11, 12).

To date, the structural basis of the inhibitory activity of ICP is unknown. Previous threading studies have suggested that the binding site of *T. cruzi* ICP may be located on the loops between β-strands in a fold that resembles immunoglobulin light chain variable domains (8, 13). Another study drew parallels between the sequence conservation in predicted loops of the ICP family and the peptidase-binding regions of the cystatin family (12). We have now determined the structure of *L. mexicana* ICP in solution by NMR spectroscopy, confirmed residues key for its inhibitory activity using site-directed mutagenesis, and investigated how the key residues may bind to the model clan CA, family C1 peptidase papain, and an important *L. mexicana* CP, known as CPB (14).
NMR Structure of Leishmania mexicana ICP

EXPERIMENTAL PROCEDURES

Protein Production—Recombinant L. mexicana ICP was expressed from a pET28 (Novagen)-derived plasmid in Escherichia coli BL21 (DE3) cells as described previously (11). 15N,13C-labeled protein was produced by growing the cells in M9 medium using 15NH4Cl and [13C]glucose (Spectra Stable Isotopes) as the sole nitrogen and carbon sources. The fusion protein was purified by nickel chelate chromatography and digested with thrombin (Novagen). The cleaved histidine tag and thrombin were removed by nickel chelate and benzamidine-Sepharose (Sigma) affinity chromatography. The protein comprising the complete native sequence (Q868H13; CAD68975) with the addition of three residues (GSH) at the N terminus (designated ICP-(2–113)) was buffer-exchanged into 25 mM sodium phosphate, pH 4.5, 50 mM NaCl, 0.001% NaN3 by extensive diafiltration using a 5,000 MWCO centrifugal concentrator (Vivascience) and concentrated to ~1 mM. D2O was added to a final concentration of 10% (v/v).

NMR samples of L. mexicana ICP-(2–113) underwent protonation over 2–3 days under NMR sample conditions to produce an N-terminally truncated protein starting at residue serine 6 (ICP-(6–113)) as confirmed by mass spectrometry, which then remained stable. No difference in Kd for L. mexicana CPB could be detected between ICP-(2–113) and ICP-(6–113).

Interaction studies were carried out using papain from Papaya latex (Sigma) and L. mexicana CPB2.8Δ CTE, produced as described previously (14). In each case, pepstatin was mixed with an excess of 15N-labeled ICP in NMR sample buffer, and the complex was isolated by gel filtration on a Superdex 75 HR10/30 column (APBiotech) and then concentrated using a 10,000 MWCO centrifugal concentrator.

NMR Spectroscopy and Data Analysis—Resonance assignments were determined using standard triple resonance NMR techniques and have been deposited as described (15). Distance restraints for structure calculation were derived from three-dimensional 15N and 13C HSQC-NOESY spectra recorded with 100 ms mixing times recorded at an 800 MHz Bruker Avance spectrometer. Slowly exchanging amide protons were identified by redissolving a lyophilized sample in D2O and recording a series of 15N HSQC spectra. Spectra were processed with AZARA4 and analyzed using CCPN analysis (16).

Structure Calculation—Assigned, partially assigned, and ambiguous NOESY cross-peaks were used to generate distance constraints within CCPN analysis that were directly exported to CNS/XPLOR format and used as input for structure calculations using CNS v1.17 (17) using a modified version of the PARALLHHDG 5.3 forcefield (18) with IUPAC-recommended nomenclature (19). Structures were generated from random atomic coordinates following the scheme of the rand, dga, and refine scripts from the XPLOR 3.1 manual (20) reimplemented in CNS and modified to incorporate floating chirality at prochiral centers (21) using a metropolis acceptance criterion. The tools provided by the ARIA (22) module within CNS were used to identify consistently violated distance restraints for checking, to reduce the ambiguity of ambiguous distance restraints based on the ensemble of structures calculated, to remove duplicate restraints, and to recalibrate the distance-intensity mapping. Distance restraints representing hydrogen bonds were incorporated for slowly exchanging amides where corroborating NOE signals existed and an acceptor atom could be unambiguously identified.

Atomic coordinates have been deposited at the PDB (PDB code 2C34), and structural statistics are summarized in Table 1.

Table 1

| NMR structural statistics for L. mexicana ICP |
|---------------------------------------------|
| NMR distance constraints                     |
| Distance constraints                          |
| Total NOE                                     | 2370 |
| Ambiguous                                     | 812  |
| Unambiguous                                   | 1558 |
| Intra-residue                                 | 869  |
| Inter-residue                                 | 302  |
| Sequential (|i−j|=1)                              | 83   |
| Medium-range (|i−j|<4)                           | 304  |
| Long-range (|i−j|>5)                            | 22   |

| Structure Statistics                         |
|---------------------------------------------|
| Violations (mean and S.D.)                  |
| Distance constraints (Å)                     | 0.052 ± 0.002 |
| Distance constraint violations > 0.5 Å      | 1.3 ± 0.98    |

| Deviations from idealized geometry          |
|---------------------------------------------|
| Bond lengths (Å)                            | 2.54e-3 ± 1.00e-4 |
| Bond angles (°)                             | 0.582 ± 1.60e-2 |
| Impropers (°)                                | 0.354 ± 1.71e-2 |

| Ramachandran plot (% residues in)           |
|---------------------------------------------|
| Most favoured regions                       | 52.5  |
| Additionally allowed regions                | 44.4  |
| Generously allowed                          | 2.7   |
| Disallowed regions                          | 0.5   |

| Root mean square deviation to the unbiased mean structure (%) | 0.854 ± 0.189 |
|---------------------------------------------------------------|
| Backbone                                                      | 0.586 ± 0.181 |

15N Relaxation Measurements—Protein dynamics were probed through 15N relaxation measurements of the backbone amide groups. 15N T1, T2, and 1H,13N-heteronuclear NOE experiments were recorded at 600.13 MHz (1H), and the data were analyzed according to the Lipari-Szabo model-free formalism using the programs curvfit and modelfree (23).

Docking—Data-driven molecular docking was carried out using HADDOCK v1.3 (24) using the default parameters and using chemical shift perturbation data as input. In the HADDOCK terminology, the “active” ICP residues are those whose backbone or side-chain amide chemical shifts are significantly perturbed in the complex with papain and which are surface-exposed. The “passive” ICP residues are their immediate, surface-exposed neighbors. In the absence of chemical shift perturbation data for the peptidease, and given the superficial similarity between the ICP and stefin B structures, the “active” papain residues were defined as those residues in close contact with stefin B in the crystal structure of the stefin B/papain complex. The passive papain residues are their immediate, surface-exposed neighbors. “Semiflexible” regions are defined to be two residues either side of the active and passive residues in the primary sequence of both proteins. The ICP-D-E loop was made “fully flexible” to reflect its high mobility in unbound ICP.

Mutations—Mutations of L. mexicana ICP were incorporated into the expression vector using the QuikChange site-directed mutagenesis kit (Stratagene) and the following pairs of complementary primers (mutated sites in lowercase): NT303, CGCGGACCATGAGGcaTATGAGGACCCGC and NT301, CGGGTGCTCCACATGgcTCCAGTTGGCGG to generate pBP199 (encoding Y34A); NT282, CATCCTGACCTCAGGacGagGGAGaTGGTGGCATCTAC and NT283, GTAGATGCGACCATCTCCGgtcAGGATCGAGGATG to generate pBP193 (encoding M64D, V68D, V70D); NT298, CCTCAGCTCCTATGGTgcAGTtcTccATCTACGTGGTGCTCG and NT299, CGAGACACACAGTAGATggAgAgAAcCggCA-CCTAGAGACGGTGGG to generate pBP194, (encoding G69P, G71P, G72P); NT284, CTGTCTACACGgcCCTCCGcCGAGGCGC.
ATCAAGC and NT285, GCTTGATGCCCTCGgcGGGGgcCGTG- to generate pBP197, (encoding R94A, F96A); NT280, GAAGGGCAACCCGgcTGGATACATGTGGACG and NT281, CGTCCACATGTATCCAccGccCGGGTGGCCCTTC to generate pBP199 (encoding T31,T32G). The mutated plasmids were verified by nucleotide sequencing.

FIGURE 1. The solution structure of L. mexicana ICP-(6–113). Conserved regions of the B-C, F-G, and D-E loops are colored blue, green, and magenta, respectively. Residues of the central hydrophobic core are colored orange and of the secondary hydrophobic cluster yellow. A, a stereoview of the ensemble of 38 structures superposed on C$_\alpha$ atoms of residues 14–60 and 73–112. The heavy atoms of the residues that make up the two hydrophobic clusters are shown. B, orthogonal views of a schematic of the structure. C, structure-based sequence alignment of ICPs confirmed as inhibitors numbered according to L. mexicana ICP sequence. The locations of the β-strands are indicated by arrows.

CGTCCACATGTATCCAccGccCGGGTGGCCCTTC to generate pBP199 (encoding T31,T32G). The mutated plasmids were verified by nucleotide sequencing.
**RESULTS**

**ICP Adopts an Immunoglobulin Fold**—In solution, ICP adopts an immunoglobulin-like (Ig) fold with seven β-strands (Fig. 1). One β-sheet is formed by anti-parallel strands B, E, and D, whereas the other is formed by anti-parallel strands G, F, and C with strand A parallel to strand G. A search for structural homologues using the DALI server (27) identified the N-terminal Ig domain from α-dystroglycan (PDB:1u2c (28)) as the closest match with a Z-score of 5.1 and an RMSD of 3.2 Å over 82 residues. The SCOP data base (29) classifies this as a cadherin-like superfAMILY fold, which also closely resembles the I-set immunoglobulin-like fold (30, 31).

The β-sheets enclose a well packed core of principally hydrophobic residues that correspond to the alternating pattern of conserved residues along the strands (see Fig. 2). The C-D loop of ICP folds back on strand C with a number of hydrophobic side-chains making contacts that are not part of the main core of the molecule. These interactions involving residues Met-35, Thr-37, Val-39, Val-42, Val-91, Thr-93, Pro-95, and Ile-99 serve to pin the F-G loop back to the outer face of one β-sheet. The prediction of the ICP family structure based on threading (13) correctly identified the Ig-like fold, but the details of the predicted strands and loops were out of register after the C-D loop, possibly because of the misidentification of the hydrophobic residues in the C-D loop as representative of an extended C strand.

The three highly conserved groups of residues previously noted (8, 11, 12) are all located in loops at one end of the molecule (Fig. 1B). The GNPTTGY motif lies in the B-C loop, the GXGG motif lies in the highly mobile D-E loop and the RPW/F motif in the F-G loop. The co-location of all three motifs, and the finding that none of the residues appear to have roles that are key to the integrity of the fold of the protein, suggests that they together form the CP-binding site. Residues 30–32 in the B-C loop form a turn of 3₁₀-helix projecting the side chains of Asn-29, Thr-31, and Thr-32 toward the solvent. Gly-28 and Gly-33 allow the loop to be accommodated between the B and C strands by adopting conformations with positive ϕ dihedral angles. The conserved Tyr-34, which caps the main hydrophobic core of the protein, interacts with residues in the F-G loop, and contributes to the solvent-exposed surface at this end of the protein. Pinned back to the GFC β-sheet, the 11-amino acid F-G loop (residues 93–103) forms a broad saddle with residues 96–101 running roughly perpendicular to the direction of the β-strands. The side chain of conserved Arg-94 interacts with Tyr-34 and, along with Tyr-96, projects out of this end of the protein accompanied by Glu-97, Lys-100, and Glu-102. The D-E loop (residues 61–72) projects from this end of the protein on the other side of the B-C loop and is poorly defined by the experimental data.

**The D-E Loop Is Flexible**—15N relaxation measurements were used to probe the backbone dynamics of ICP-(6–113). 15N NOE, 13C T₁, and 15N T₂ were measured at 60.8 MHz (15N), the derived order parameters and correlation times are shown in Fig. 2. The majority of the residues display uniform relaxation rates close to the average values typical of a...
compactly folded monomeric protein of this size, and their behavior could be modeled using either a single order parameter ($S^2$) or $S^2$ and an internal correlation time. The exceptions are the N terminus up to residue 13 and residues 61–72 in the D-E loop, which display depressed NOE values and lengthened $T_2$ values. These residues were best modeled with two order parameters representing internal motion on distinct timescales and an internal correlation time of the order of a nanosecond. In addition, residue Asn-29, which could not be well fitted by any model, has a significantly shortened $T_2$ value indicative of millisecond timescale chemical exchange, which may reflect backbone flexibility, or given the proximity of Tyr-92 may simply be because of the combination of local motion and the ring current shift from this aromatic side chain.

The Conserved Motifs in the Loops Are Involved in CP Binding—Chemical shift perturbation mapping was used to identify the CP-binding site on ICP. Complexes of 15N-labeled ICP with \textit{L. mexicana} CPB and with \textit{P. latex} papain were purified, and their 15N HSQC spectra were compared with free ICP. A similar subset of backbone H$_N$ and side chain cross-peaks are perturbed from their free ICP chemical shifts in both complexes, and the most distinctive are shown in Fig. 3 for the papain complex. It is clear that all the most perturbed residues fall in the ranges 28–37 (covering the B-C loop), 59–74 (covering the D-E loop), and 94–104 (covering the F-G loop). Chemical shifts are sensitive to changes in chemical environment caused either by proximity to a ligand or by propagated structural changes. The co-location of all the significant changes in the vicinity of the three conserved loops strongly suggests that this is because of their involvement in binding to target CPs.

Residues in the D-E Loop Are Critical for CP Binding—Based on our \textit{L. mexicana} ICP structure and on sequence similarity between ICP family members, we made a limited number of mutant ICPs to test their effect on the inhibitory activity of ICP. In the highly conserved B-C loop, we mutated the pair of threonines to glycine (T31G,T32G) and mutated the conserved tyrosine at the base of the loop to alanine (Y34A). In the flexible D-E loop, we made two triple mutants, one to replace the semiconserved hydrophobic residues with charged residues (M67D,V68D,V70D) and the other to replace the glycines by more conformationally restricted residues (G69P,G71P,G72P). In the F-G loop, we made a double mutant of the conserved arginine and aromatic residues (R94A,F96A).

The D-E loop mutants lacked inhibitory activity against \textit{L. mexicana} CPB. To ensure that these mutants had retained their three-dimensional fold we analyzed them using CD spectroscopy. Wild type ICP and mutant M67D,V68D,V70D possess superimposable CD spectra in both the near and far UV regions indicating very similar, if not identical, secondary and tertiary structures. Mutant G69P,G71P,G72P gave rise to a slightly different far UV spectrum, while retaining similar spectral features in the near UV region, and CDSSTR analysis placed the regular secondary structure estimates within a few percent of the wild type ICP. Of the other mutants, the B-C loop mutations resulted in somewhat lower $K_s$, whereas the $K_s$ of the F-G loop mutant was ~3-fold higher (see Table 2).

A Model for the Interaction between ICP and CPs—The identification of the three conserved loops as the CP-binding site prompted a comparison of the ICP structure with other CP inhibitors. A slight, but suggestive, structural similarity was detected with the cystatin, stefin B (see below), where the CP-binding site is formed by a short central loop flanked by a flexible N-terminal peptide and a longer flattened loop that incorporates a proline (see Fig. 4). We speculated that ICP might bind target CPs in a similar fashion with the flexible D-E loop binding in place of the N-terminal peptide of stefin B. To produce a model for the interaction between ICP and a CP, we used the chemical shift perturbation data for ICP in complex with papain, together with information from

![Figure 3](https://example.com/figure3.png) 

**TABLE 2**

| Loop          | $K_s$     |
|---------------|-----------|
| \textit{L. mexicana} ICP | 132.8 ± 21.1 |
| T31G,T32G     | 53.0 ± 8.3  |
| Y34A          | 46.4 ± 6.5  |
| M67D,V68D,V70D| ND, no inhibition observed |
| G69P,G71P,G72P| ND, no inhibition observed |
| R94A,F96A     | 428.5 ± 86.5 |

![Table 2](https://example.com/table2.png)
the crystal structure of the complex between papain and stefin B, to drive a docking simulation using HADDOCK (24). The input data dictated that ICP should bind into the active site of papain but did not, a priori, dictate that ICP should bind the peptidase in a stefin-like orientation. Of the 200 lowest energy structures from the HADDOCK run, all adopted a stefin-like mode of binding (Fig. 4) with residues 67–70 from the D-E loop filling the “unprimed” sites of the active site cleft, whereas the B-C loop approaches the active site cysteine-histidine pair, and side chains from residues in the F-G loop fill the “primed” end of the active site cleft. The preferred orientation for the interaction additionally brings the C-D loop into contact with the short helix (residues 139–143) on the top of the R domain of papain. The feasibility of this contact is born out by the observation that the chemical shifts of residues 43 and 46 in the C-D loop can clearly be seen to change on complex formation (Fig. 3).

DISCUSSION

This study has shown that ICP has an Ig fold that acts as a scaffold for three interstrand loops carrying the most highly conserved residues in the chagasin family of ICP proteins. These three loops are located adjacent to one another at one end of the protein with the most highly conserved B-C loop lying between the structured F-G loop and the flexible D-E loop. Together, the three loops form a ridge that is of the correct dimensions to be complementary to the active sites of clan CA,
family C1 CPs. Chemical shift perturbation studies identify these loops as forming the binding interface with the two representative CPs, papain and *L. mexicana* CPB.

Comparison of the ICP structure with CP inhibitors of other families reveals no similarity at the level of the overall fold. However, there are suggestive similarities between the CP-interacting regions of ICP and the cystatin family of inhibitors. The ICP F-G loop (residues 93–103) bears a structural resemblance to the second binding loop of the tripartite wedge of stefins A and B (4, 32) and superposing the structures on these features brings the B-C loop and C-terminal end of ICP strand B into the vicinity of the first interacting loop and strand B of the stefins (Fig. 4A). Superposition of both these structural features places the flexible D-E loop of ICP in the vicinity of the stefin N-terminal trunk (Fig. 4A). This suggests that the three conserved ICP loops may form a similar tripartite wedge to that of the cystatin family, both adapted to fit the active site cleft of clan CA, family C1 CPs. The implications of this similarity would be that ICP recognizes target CPs with the flexible D-E loop adapting to the recognition sites for the substrate residues N-terminal to the cleavage site (thought to be a main determinant of specificity for this group of CPs) and the F-G loop binding to the distal end of the substrate recognition groove. The importance of these interactions for the binding of ICP is demonstrated by our data for the ICP mutants (Table 2). In particular, the changes to the D-E loop abrogated inhibitory activity. It should be highly informative to express this “dead” mutant in *L. mexicana* so that the role of ICP in the parasite and its interaction with its host can be studied. Additional support for the importance of the F-G loop is provided by the observation that ICP inhibits cathepsin B considerably less well than cathepsin L (7). This is also a feature of stefins (33) and can be explained by the steric hindrance afforded by the occluding loop, which is characteristic of cathepsin B-like CPs and is partially responsible for their substrate specificity.

To investigate the feasibility of this model, we carried out data-driven docking between ICP and papain with the assumption that ICP would contact the same region of the active site as the cystatins because the flexibility of the D-E loop makes the results of an any less biased approach such as rigid body docking difficult to evaluate. The HADDOCK modeling surprisingly revealed just one clearly favored orientation for the interaction in which the fold of ICP places the highly conserved NPTT sequence motif in the B-C loop close to catalytic residues of the target CP. The ICP B-C loop is longer and bulkier than the first interacting loop of the cystatin family and, given its disposition relative to the loops on either side, may interact intimately with the CP active site residues. Thus it is surprising that the two B-C loop mutations had no deleterious effect on inhibitory activity despite the high level of sequence conservation between bacterial and protozoan ICPs in this region. However, the changes introduced into the mutants both decreased the bulk of the B-C loop. It is known that CP active sites are expanded when in complex with stefins (4, 5), and so, taking into account the bulkiness of the ICP B-C loop, wild type ICP may be larger than is optimal for the tightest interaction with this peptidase active site. The most likely explanation for this is that the loop is optimized for inhibition of an as yet untested CP other than those investigated, such as a host CP that has a wider active site. This suggestion is consistent with previous findings derived by generating mutants of *Leishmania* that lack or overexpress ICP (10).

Proteins with Ig folds are rare in species other than higher eukaryotes and the viruses that infect them. Hitherto they seemed to be entirely absent from protozoa. ICP is the first protein with a cadherin-like Ig domain to be discovered in a non-metazoa. This raises the possibility that ICPs were acquired by the pathogens from their hosts, although for this to be the case the event would have had to have happened on multiple occasions as gene transfer between the parasites seems unlikely. A more likely explanation is that a crucial role in the host-pathogen interaction has ensured retention of an otherwise unimportant gene by the pathogens.

Most cadherin-like Ig domains in cell-adhesion molecules mediate protein-protein interactions. However, for the cadherin domains themselves, these principally involve the faces of the β-sheets (34). This differs from the binding mode of ICP, which is more akin to antigen recognition by Ig variable domains. The finding that ICP has an inhibitory mechanism potentially rather similar to the cystatins, despite the unrelated primary structures and folds of the proteins, clearly demonstrates the likelihood that there has been convergent evolution. This apparent similarity also suggests that this type of protein–protein interaction is efficient for the regulatory functions of these inhibitors and, moreover, perhaps that such a tripartite binding mechanism is required.

The discovery of the binding mode of ICP may also be informative for the design of chemical inhibitors of CPs that could have therapeutic value, both against pathogens and also in the treatment of CP-related mammalian disorders. Perhaps a compound that mimics the binding of the D-E loop and also has an active site-directed nucleophile would show good activity against the CPs. Indeed the peptidyl vinyl sulphonate compound currently in development for use against *T. cruzi* infections (35, 36) presumably is acting in just this way. The discovery of the physiological targets of each ICP should allow design of inhibitors that have optimal specificity for these target CPs and therefore potentially have value in therapies directed against them.

Acknowledgments—NMR data were acquired at the University of Edinburgh Biomolecular NMR Unit. We thank Sharon Kelly and Tommy Jess for assistance with the CD spectroscopy and analysis.

REFERENCES

1. Rawlings, N. D., Tolle, D. P., and Barrett, A. J. (2004) *Biochem. J.* 376, 705–716
2. Rawlings, N. D., Tolle, D. P., and Barrett, A. J. (2004) *Nucleic Acids Res.* 32, 160–164
3. Rode, W., Engui, R., Musil, D., Thiele, U., Huber, R., Karashkov, A., Brzin, J., Kos, J., and Turk, V. (1988) *EMBO J.* 7, 2593–2599
4. Stubbs, M. T., Laber, B., Rode, W., Huber, R., Jeral, R., Lenarcic, B., and Turk, V. (1990) *EMBO J.* 9, 1939–1947
5. Jenko, S., Dolenc, I., Guncar, G., Dobesker, A., Podobnik, M., and Turk, D. (2003) *J. Mol. Biol.* 326, 875–885
6. Guncar, G., Pungencevic, G., Klemenc, I., Turk, V., and Turk, D. (1999) *EMBO J.* 18, 793–803
7. Monteiro, A. C. S., Abrahamson, M., Lima, A. P. C. A., Vannier-Santos, M. A., and Scharfstein, J. (2001) *J. Cell Sci.* 114, 3933–3942
8. Rigden, D. J., Mosolov, V. V., and Galperin, M. M. (2002) *Protein Sci.* 11, 1971–1977
9. Santos, C. C., Santos, A. C., Fernandez, A., Barreiro, E., and de Lima, A. (2005) *J. Mol. Biol.* 347, 383–388
10. Besteiro, S., Coombs, G. H., and Mottram, J. C. (2004) *FEBS Lett.* 54, 1224–1236
11. Sanderson, S. J., Westrop, G. D., Scharfstein, J., Mottram, J. C., and Coombs, G. H. (2003) *FEBS Lett.* 542, 12–16
12. Riekenberg, S., Wites, B., Saric, M., Bruchhaus, I., and Scholze, H. (2005) *FEBS Lett.* 579, 1573–1578
13. Rigden, D. J., Mottram, J. C. S., and De Sá, M. F. G. (2001) *FEBS Lett.* 504, 41–44
14. Sanderson, S. J., Pollock, K. G., Hillery, J. D., Mello, A. L., Hilaire, P. A., Mijano, M. A., Juliano, L., Mottram, J. C., and Coombs, G. H. (2000) *Biochem. J.* 347, 383–388
15. Smith, B. O., Westrop, G. D., Mottram, J. C., and Coombs, G. H. (2005) *J. Biomol. NMR,* in press
16. Breimer, W. R., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, P., Ulrich, E. L., Markley, J. L., Ionides, J., and Lue, E. D. (2005) *Proteins Struct. Funct. Bioinf.* 59, 687–696
17. Bünzli, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszelewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D. Biol. Crystallogr.* 54, 905–921
18. Linge, J. P., Williams, M. A., Pronk, C. A. E. M., Bonvin, A. M. J. J., and Nilges, M. (2003) *Proteins Struct. Funct. Genet.* 50, 496–506
19. Markley, J. L., Bax, A., Arata, Y., Hilbers, C. W., Kappetein, R., Sykes, B. D., Wright, P. E.,
NMR Structure of Leishmania mexicana ICP