

The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites

Brett S. Chevalier¹, Raymond J. Monnat, Jr.² and Barry L. Stoddard¹

¹Fred Hutchinson Cancer Research Center and the Graduate Program in Molecular and Cell Biology, University of Washington 1100 Fairview Ave. N. A3-023 Seattle, Washington 98109 USA. ²University of Washington, Department of Pathology, Box 357705 Seattle, Washington 98195 USA.

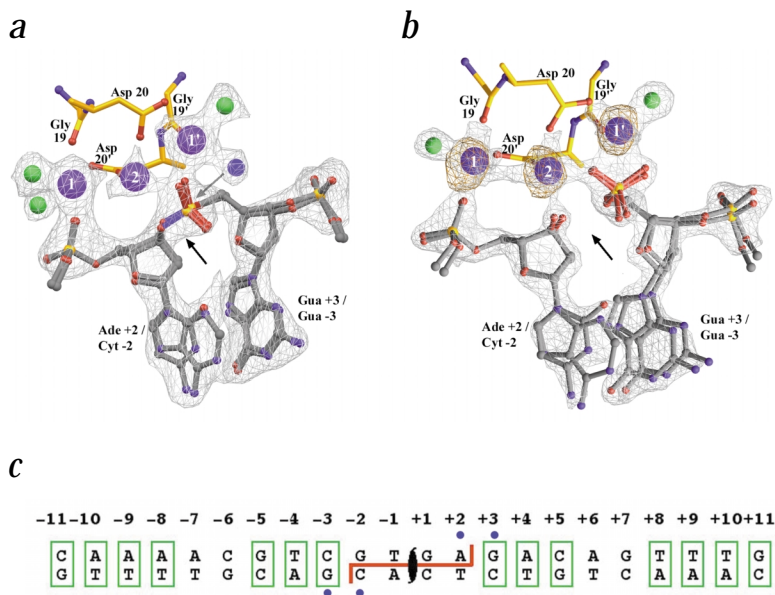
Homing endonucleases, like restriction enzymes, cleave double-stranded DNA at specific target sites. The cleavage mechanism(s) utilized by LAGLIDADG endonucleases have been difficult to elucidate; their active sites are divergent, and only one low resolution cocrystal structure has been determined. Here we report two high resolution structures of the dimeric I-CreI homing endonuclease bound to DNA: a substrate complex with calcium and a product complex with magnesium. The bound metals in both complexes are verified by manganese anomalous difference maps. The active sites are positioned close together to facilitate cleavage across the DNA minor groove; each contains one metal ion bound between a conserved aspartate (Asp 20) and a single scissile phosphate. A third metal ion bridges the two active sites. This divalent cation is bound between aspartate residues from the active site of each subunit and is in simultaneous contact with the scissile phosphates of both DNA strands. A metal-bound water molecule acts as the nucleophile and is part of an extensive network of ordered water molecules that are positioned by enzyme side chains. These structures illustrate a unique variant of a two-metal endonuclease mechanism is employed by the highly divergent LAGLIDADG enzyme family.

ophile and is part of an extensive network of ordered water molecules that are positioned by enzyme side chains. These structures illustrate a unique variant of a two-metal endonuclease mechanism is employed by the highly divergent LAGLIDADG enzyme family.

The homing endonucleases are a diverse family of proteins encoded by open reading frames in genetically mobile introns or inteins^{1–4}. They have been identified in unicellular eukaryotes, *Archaea* and eubacteria^{1–4}. These proteins cleave long DNA target sites (14–40 bp) in homologous alleles that lack the intron or intein. The cleavage event initiates the transfer of the mobile sequence to these sites by a targeted transposition mechanism termed ‘homing’. At least four homing endonuclease families have been identified on the basis of conserved sequence motifs that provide residues critical for enzyme folding and catalysis^{1,2}. The LAGLIDADG family is the largest of these families with over 200 known members, each containing one or two copies of a motif that resembles the consensus ‘LAGLIDADG’ sequence^{2,5}.

Enzymes that contain a single copy of the LAGLIDADG motif — for example, I-CreI and I-CeuI — are homodimers of 15–20 kDa per subunit and recognize pseudo palindromic homing sites^{6–9}. The LAGLIDADG motif, located near the N-terminus of each monomer, forms an α -helix and packs against its counterpart to form a dimer interface¹⁰. The last, strictly conserved acidic residue of each motif is found in the active site of each subunit. Enzymes with two copies of the LAGLIDADG motif — for example, I-DmoI and the endonuclease component of the PI-SceI intein — are roughly twice as large as an I-CreI subunit and act as monomers. These members use the LAGLIDADG motif as an intramolecular domain interface^{11–13}. The monomeric LAGLIDADG enzymes usually cleave less symmetric DNA target sites than I-CreI^{14–19}. Both the homodimeric and monomeric LAGLIDADG endonucleases generate 4-nucleotide, 3'-extended cohesive ends. Crystal structures have been determined for the intron-encoded I-CreI²⁰ and I-DmoI¹¹

Fig. 1 Electron density at the cleavage site. **a**, The substrate complex. **b**, The cleaved product complex. The electron density corresponds to omit $F_o - F_c$ electron density maps contoured over all residues and nucleotides near the scissile phosphates. Divalent cations are purple, the nucleophilic water in the substrate complex in (a) is blue, and water molecules are green. All atoms modeled in this figure were omitted from the phase calculation for maps. The structure of the substrate complex was determined in the presence of calcium; the scissile phosphodiester bond is intact (black arrow). The structure of the cleaved product complex was determined in the presence of magnesium; the scissile phosphodiester bond is fully cleaved and the 5' phosphate is rotated away from the adjoining ribose sugar. The red density corresponds to the strongest features of an anomalous difference map calculated from data collected from crystals grown in the presence of manganese (contoured at 8 σ). The central shared metal ion (# 2) bridges the enzyme active sites and contacts both scissile phosphates and both LAGLIDADG Asp 20 residues (Asp 20 and 20'). The two flanking metal ions (1 and 1') are each bound by individual active sites and contact a conserved aspartate residue and scissile phosphate groups. The density at base 2 of each half-site corresponds to an equal mixture of adenine and cytosine because the pseudo symmetric DNA is bound in both orientations relative to the I-CreI homodimer (see text). Both refined orientations are shown. **c**, The sequence of the native homing site is shown for reference: bases conserved between homing half-sites are boxed, and blue dots beneath indicate the nucleotide bases shown in the panels. Figs 1–3 made using program RIBBONS³¹.



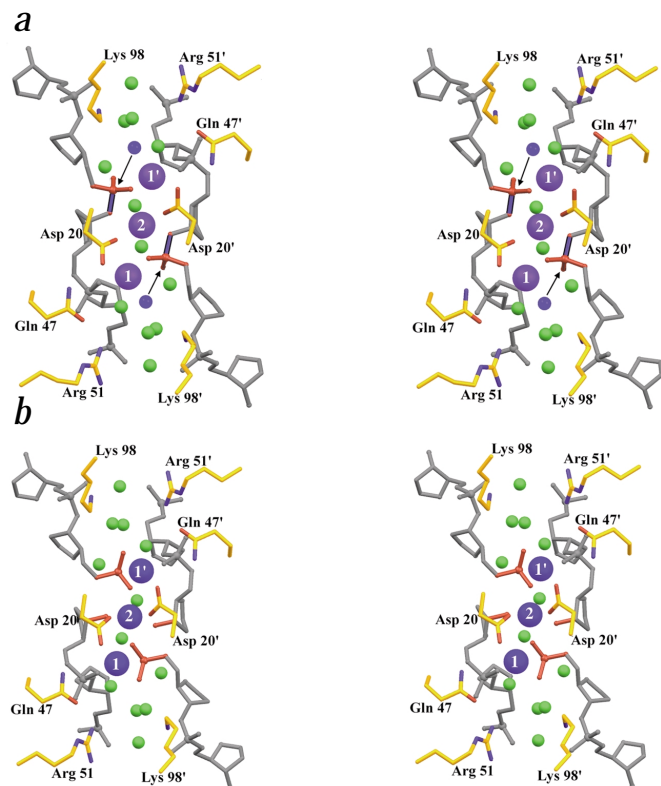


Fig. 2 The DNA–protein interface. **a**, The substrate complex. **b**, The cleaved product complex. Stereo view is shown looking directly into the minor groove at the cleavage site. Only phosphate groups and DNA sugars of the DNA are shown for clarity. The scissile phosphodiester bond is shown in (a) as a blue P–O bond segment; the direction of nucleophilic water attack on the corresponding red phosphate group is shown by a black arrow. The coloring is the same as in Fig. 1.

(see Methods). We therefore modeled and refined the I-CreI structures with a mixture of these two DNA orientations. The cleavage state of the DNA, the conformation of the DNA backbone, the position of the cleaved DNA ends and the conformation of protein side chains in the protein–DNA interface are isomorphous between the two DNA orientations and between the separate DNA strands.

Positions of bound metals and nucleophilic waters

The two active sites are positioned close to one another with a spacing that matches the separation between the scissile phosphates (Fig. 2). The aspartate residues in the C-terminal ends of the LAGLIDADG helices (Asp 20) are separated by and very close to the two-fold dimer symmetry axis. A total of three divalent cations are bound by the enzyme–DNA complex. The binding site of each metal has been verified by anomalous difference maps using crystals grown in the presence of manganese (Fig. 1b). Each individual active site binds one divalent cation that is coordinated by an octahedral arrangement of six ligands. In the uncleaved complex (Figs 2a, 3a, 4a), calcium is bound by one oxygen atom from the conserved aspartate residue (Asp 20), the main chain carboxyl oxygen of Gly 19, a nonbridging oxygen from the scissile phosphate (between nucleotide bases +2 and +3), a second nonbridging oxygen from a phosphate group on the opposite DNA strand (between bases -1 and -2) and two water molecules. One of these water molecules (number 24 in Fig. 4a) is located 3 Å from the scissile phosphate and is well-positioned for in-line

endonucleases and for the intein-associated PI-*SceI*¹³ and PI-*PfiI*¹² endonucleases. A crystal structure of a bound complex with DNA has only been reported for I-CreI¹⁰ to 3 Å resolution. Despite low sequence homology outside the LAGLIDADG motif, the topologies of these enzymes are quite similar. However, with the exception of the C-terminal acidic residue of each LAGLIDADG motif (underlined and in bold), their active site residues are divergent. This has made it difficult to assign roles for individual active site residues on the basis of structural conservation alone. The study reported here directly illustrates the structural mechanism of DNA cleavage for the I-CreI homing endonuclease.

Overall structure of the complex

The native homing-site DNA is bound in a relatively unperturbed B-form conformation and exhibits a curvature around the enzyme that reduces the distance between minor groove scissile phosphates to ~8 Å. In the presence of calcium, the DNA is uncleaved (Fig. 1a), whereas in the presence of magnesium or manganese the DNA is completely cleaved between bases 2 and 3 (Fig. 1b). The only significant conformational change upon cleavage is a movement of 1.5 Å by the liberated 5' phosphate group away from the 3' leaving group. The pseudo symmetric homing-site DNA construct is present in both orientations relative to the enzyme homodimer in the crystals

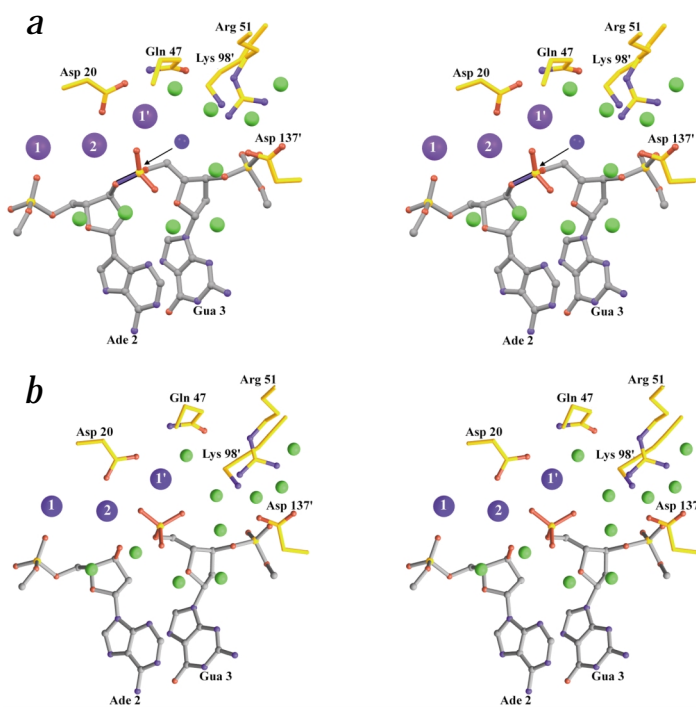


Fig. 3 Stereo side views of the DNA–protein interface. **a**, The substrate complex. **b**, The cleaved product complex. Only one DNA strand and one active site are shown for clarity. The residue and metal numbers and labels are the same as in Figs 1 and 2. Note the network of water molecules (green spheres) surrounding the scissile phosphate, the nucleophilic water and the 3' oxygen leaving group, and the movement of the cleaved 5' phosphate in the product complex in (b).



letters

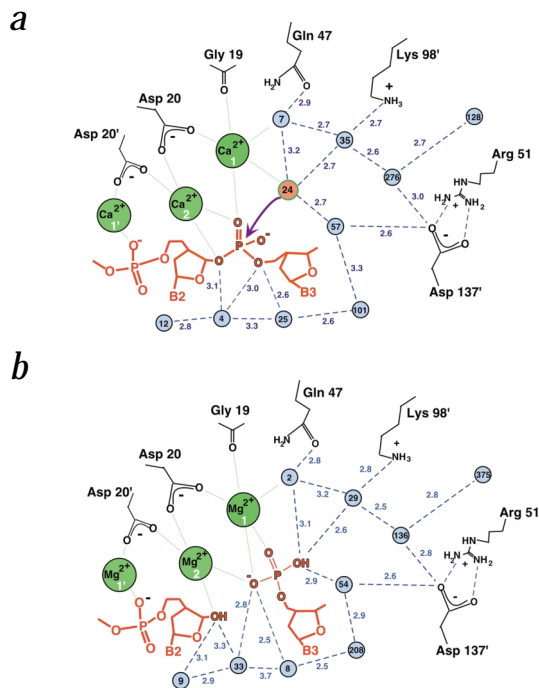


Fig. 4 Schematic of the active site interactions and solvent network. **a**, The substrate complex. **b**, The cleaved product complex. The residue numbers and labels are the same as in Figs 1–3. Water molecules are numbered consistently with the corresponding PDB files and are shown in blue with the exception of the proposed nucleophilic water molecule (orange 24) in the top panel. Bond distances are given in Å. All direct contacts between the bound metal ions and protein side chains and water molecules and the scissile phosphate are shown by thin green bonds; the contacts to phosphate atoms on the opposite DNA strand are omitted for clarity. The metal ions display octahedral coordination, with ligand bond distances ranging from 2.0 to 2.1 Å for magnesium and 2.4 to 2.5 Å for calcium.

attack and hydrolysis of this phosphate (angle from the water nucleophile through the scissile phosphate to the 3' oxygen leaving group). In the cleaved product complex with magnesium (Figs 2b, 3b, 4b), the ligands coordinating the metal are identical with the exception that the nucleophilic water is now a covalently bound oxygen atom on the free 5' phosphate group; this oxygen atom is still directly bound to the metal ion.

A third metal ion (metal ion 2 in Figs 1–3) is bound at the interface of the enzyme subunits and is octahedrally coordinated by three identical ligands from each active site. In the uncleaved complex, the metal ion contacts an oxygen atom from each of the two conserved aspartate residues of the LAGLIDADG motif, the 3' bridging oxygen of the scissile phosphate on each DNA strand and a non-bridging oxygen from the same scissile phosphates. In the cleaved complex, this coordination is maintained, with the exception that the 3' oxygen is now a hydroxyl group on each cleaved DNA strand.

The metal ions in each individual active site (metal ions 1 and 1') appear to stabilize the phosphoanion transition state during cleavage and to position an activated water molecule for nucleophilic attack. The metal ion bound between the active sites (metal ion 2) simultaneously contacts both scissile phosphates and appears to stabilize the 3' oxygen leaving group on both DNA strands. Therefore, strand scission occurs by a two-metal hydrolytic mechanism, but with an essential bound metal ion shared between the two active sites. To the best of our knowledge, this is the first example of a shared metal ion among active sites for a nuclease or for any enzyme carrying out metal-dependent phosphoryl transfer. This finding should have implications for the kinetic mech-

anism of double strand cleavage. The hydrolysis of individual phosphates may be more concerted than for other endonuclease families because it is possible that all metal sites must be occupied in the complex before cleavage of either strand can occur. This result may also explain why it has been difficult to create enzymes with 'nicking' activity, in which only one strand of DNA is cut, by mutating individual active site aspartate residues in monomeric LAGLIDADG endonucleases²¹.

Active site residues and an extensive catalytic solvent network

In addition to the utilization of a shared metal ion by both active sites, the positions and roles of solvent molecules and additional catalytic side chains in I-CreI are unusual. Apart from the conserved aspartic acids, three residues in the I-CreI active site have been identified as particularly important for catalysis: Lys 98, Arg 51 and Gln 47 (ref. 22). The counterparts of these residues are

Table 1 Data and refinement statistics

Data set	Product Complex (Mg ²⁺)	Substrate Complex (Ca ²⁺)	Anomalous (Mn ²⁺)
Source	5.0.2 (ALS)	Rotating Anode	Rotating Anode
Resolution limit (Å)	1.80	2.05	2.60
Wavelength (Å)	1.10	1.54	1.54
Space group	P21	P21	P21
Unit cell parameters			
a,b,c (Å)	43.0, 67.9, 88.3	43.2, 67.8, 88.3	43.1, 67.8, 88.3
angle (°)	91.6	91.6	91.6
Measured reflections	153,455	76,042	163,556
Unique reflections	44,949	30,080	15,211
R _{merge} ¹	3.8 (18.5)	6.7 (36.0)	7.4 (31.0)
Completeness ¹ (%)	95.4 (75.3)	94.1 (81.7)	97.9 (95.1)
Refinement			
R-factor (%)	20.4	20.7	
R _{free} (%) ²	24.8	26.0	
Resolution (Å)	50–1.80	50–2.05	
Number of atoms	5,294	4,868	
Number of waters	857	435	
R.m.s. deviations			
Bond length (Å)	0.005	0.005	
Bond angles (°)	1.15	1.16	
Mean B value (Å ²)			
Overall	30.1	37.6	
Protein	31.2	39.5	
DNA	25.4	32.7	
Solvent	47.9	48.5	
Cations	15.6 (Mg ²⁺)	29.9 (Ca ²⁺)	

¹The numbers in parentheses are statistics from the highest resolution shell.

²R_{free} was calculated with 5% of the data.



also important in the LAGLIDADG homing endonuclease *I-CeuI*⁷. These residues primarily interact with a network of solvent molecules that surround the nucleophilic water molecule (Fig. 4) and extend around the scissile phosphate to the 3' oxygen leaving group. This network includes a water molecule (number 4 in Fig. 4a) that is positioned near the 3' leaving group. This water molecule is not directly coordinated to a metal ion and, therefore, is not likely to be an ideal proton donor. However, because the 3' oxyanion leaving group of the scissile phosphate directly interacts with a metal ion, departure of this oxygen and its eventual protonation is still favorable. *I-CreI* appears to utilize a mechanism for DNA strand cleavage in which scissile phosphates contact two divalent cations while being extensively hydrated by several water molecules. This hydration shell is in turn structured and polarized by interactions with several basic side chains.

In most type II restriction endonucleases, individual residues directly contact the scissile phosphate oxygens, bind metal ions and help activate the water nucleophile through direct polar contacts²³. These residues tend to be conserved among large numbers of restriction enzymes and are usually part of the (PD...^{(D/E)X^{K/R}}) signature for these enzymes, where ^{D/E} and ^{K/R} represent either of two residues at a single position, X is any residue at a single position and '...' represents a variable number of residues after the signature of these enzymes. Additional residues in type II endonucleases that are peripheral to the site of cleavage and often interact with solvent molecules near the DNA substrate are also catalytically important and result in diminished catalytic efficiency when mutated. These residues tend to be less strictly conserved across the entire enzyme superfamily. The majority of important active site residues in the *I-CreI* endonuclease and, presumably, in the LAGLIDADG family appear to resemble these latter type II residues in their catalytic role and in their degree of structural conservation. Apart from the C-terminal acidic residues in the LAGLIDADG motif, none of the active site residues that play a role in *I-CreI* catalysis are strictly conserved throughout the LAGLIDADG family^{1,5}. For example, Lys 98 and 98' in *I-CreI* have counterparts in the monomeric *PI-SceI* intein–endonuclease as Lys 301 and Lys 403 (ref. 24). In contrast, a Lys 98 equivalent is present in only one domain of *I-DmoI* (as Lys 120) (ref. 11) or of *PI-PfuI* (as Lys 322) (ref. 12). Similarly, *I-CreI* residues Gln 47 and Arg 51 do not have consistent counterparts in these other LAGLIDADG homing endonucleases.

The catalytic mechanism we have proposed for *I-CreI* can be generalized to other members of the LAGLIDADG family and may explain some of the divergence of active site residues. The use of basic side chains to position and polarize a network of waters surrounding the scissile phosphate groups might allow a broader range of substitutions than if these residues were used for direct, specific contacts to either the substrate or the transition state during catalysis. This active site structure and catalytic mechanism may have been further diversified by the independent and separate fusion of ancestral homodimeric LAGLIDADG endonucleases¹.

Methods

Crystallization. The *I-CreI* endonuclease was overexpressed and purified as previously described²⁵ with the exception that the induction and subsequent cell growth was conducted at 15 °C to facilitate efficient overexpression. The DNA was purchased from Oligos Etc. and consisted of two strands of sequence: 5'-GCAAAACGTCGTGAGACAGTTTCG-3' and its complement 5'-CGAAACTGTCTACGACGTTTTGC-3'. The construct forms a 24-bp

blunt-ended pseudo palindromic duplex that differs at 4 positions between the two half-sites (Fig. 1). Crystals were grown using a 2.7:1 molar ratio solution of DNA:protein by hanging drop vapor diffusion against a reservoir containing 20 mM NaCl, 100 mM MES pH 6.3–6. and PEG 400 (v/v) 20–35%. The final concentration of *I-CreI* in the DNA:protein complex solution was 3.5 mg ml⁻¹. The crystallization drops also contained 10 mM of divalent cations that either inhibited (CaCl₂) or promoted (MgCl₂) cleavage, or permitted cleavage while allowing verification of the metal binding sites by using anomalous difference Fourier maps (MnCl₂). For all three separate experiments, the crystal belong to space group P2₁, with unit cell dimensions a = 43 Å, b = 68 Å and c = 88 Å (Table 1). Although the DNA constructs are identical, and the crystallization conditions similar to that used in previous crystallographic studies²⁰, the resulting crystal form represent a new space group that diffracts to 2 Å resolution or higher.

Data collection. Crystals were transferred sequentially to aliquots of the crystallization reservoir with the concentration of PEG 400 increased to (v/v) 30 to 35%. Crystals were suspended in a fiber loop, frozen in liquid nitrogen and maintained at 100 K during data collection. Data corresponding to crystals grown in the presence of calcium or manganese were collected on an in-house Rigaku RAXIS IV area detector mounted on an RU200 rotating anode X-ray generator equipped with mirror focusing optics (Molecular Structure Corporation). Data from crystals grown in the presence of magnesium were collected at the Advanced Light Source (beamline 5.0.2). Data were reduced using the DENZO/SCALEPACK crystallographic data reduction package²⁶ (Table 1).

Structure refinement. The structures of the complexes were solved by molecular replacement with the program EPMR²⁷, using the low resolution model of the enzyme–DNA complex²⁰ as a search probe. Subsequent maps were of excellent quality. The initial models were refined using CNS²⁸ with 5% of the reflections set aside for R_{free}²⁹. For subsequent rebuilding, omit maps were calculated. The pseudo symmetric DNA is present in both possible orientations relative to the homodimeric endonuclease; this is readily apparent when omit maps are examined at each of the four basepairs in each half-site that violate palindromic symmetry within the DNA (positions ± 1, 2, 6 and 7). Refinement of the complexes with the DNA in single orientations rather than in both orientations causes an increase in R_{free} from ~25% to ~27%. This result is confirmed by experiments in which an iododeoxyuracil base is incorporated into the end of one DNA strand. Difference Fourier maps of that complex also indicate a 50/50 mixture of two DNA orientations (B.L.S., unpublished data). The final refined model consists of residues 2–153 for each enzyme subunit, 12 base pairs of DNA for each DNA half-site (modeled and refined as a mixture of two DNA orientations), 3 metal ions per complex and 435 or 857 water molecules for the substrate and product complexes, respectively. For the crystals grown in the presence of manganese, anomalous difference Fourier maps unambiguously confirm the locations of bound metal ions. Geometric analysis of the structure using PROCHECK³⁰ indicates that there are no residues with generously allowed or unfavorable backbone dihedral angles and that 89% of residues are in the core region of the Ramachandran plot.

Coordinates. Atomic coordinates of the structures have been deposited in the Protein Data Bank (accession codes 1G9Y for the substrate complex and 1G9Z for the product complexes).

Acknowledgments

We acknowledge B. Shen, E. Galburt, C. Spiegel and R. Strong for advice during the crystallographic analysis, P. Rupert, P.-W. Li and A. Ferre-D'Amare for assistance collecting data, and the staff of ALS beamline 5.0.2 for technical assistance. This work was supported by the NIH for B.L.S. and R.J.M. and by the National Cancer Institute for B.C.

Correspondence should be addressed to B.L.S. email: bstoddar@fred.hhcr.org

Received 7 December, 2000; accepted 1 February, 2001.



letters

1. Jurica, M.S. & Stoddard, B.L. *Cell. Mol. Life Sci.* **55**, 1304–1326 (1999).
2. Belfort, M. & Roberts, R.J. *Nucleic Acids Res.* **25**, 3379–3388 (1997).
3. Belfort, M. & Perlman, P.S. *J Biol. Chem.* **270**, 30237–30240 (1995).
4. Lambowitz, A.M. & Belfort, M. *Annu. Rev. Biochem.* **62**, 587–622 (1993).
5. Dalgaard, J.Z. *et al. Nucleic Acids Res.* **25**, 4626–4638 (1997).
6. Durrenberger, F., Thompson, A.J., Herrin, D.L. & Rochaix, J.D. *Nucleic Acids Res.* **24**, 3323–3331 (1996).
7. Turmel, M., Otis, C., Cote, V. & Lemieux, C. *Nucleic Acids Res.* **25**, 2610–2619 (1997).
8. Argast, G.M., Stephens, K.M., Emond, M.J. & Monnat, R.J., Jr. *J. Mol. Biol.* **280**, 345–353 (1998).
9. Marshall, P. & Lemieux, C. *Nucleic Acids Res.* **20**, 6401–6407 (1992).
10. Heath, P.J., Stephens, K.M., Monnat, R.J., Jr. & Stoddard, B.L. *Nature Struct. Biol.* **4**, 468–476 (1997).
11. Silva, G.H., Dalgaard, J.Z., Belfort, M. & Van Roey, P. *J. Mol. Biol.* **286**, 1123–1136 (1999).
12. Ichihyanagi, K., Ishino, Y., Ariyoshi, M., Komori, K. & Morikawa, K. *J. Mol. Biol.* **300**, 889–901 (2000).
13. Duan, X., Gimble, F.S. & Quijcho, F.A. *Cell* **89**, 555–564 (1997).
14. Komori, K., Ichihyanagi, K., Morikawa, K. & Ishino, Y. *Nucleic Acids Res.* **27**, 4175–4182 (1999).
15. Hu, D., Crist, M., Duan, X. & Gimble, F.S. *Biochemistry* **38**, 12621–12628 (1999).
16. Pingoud, V. *et al. Biol. Chem.* **274**, 10235–10243 (1999).
17. Aagaard, C., Awayez, M.J. & Garrett, R.A. *Nucleic Acids Res.* **25**, 1523–1530 (1997).
18. Wende, W., Grindl, W., Christ, F., Pingoud, A. & Pingoud, V. *Nucleic Acids Res.* **24**, 4123–4132 (1996).
19. Lykke-Andersen, J., Garrett, R.A. & Kjems, J. *Nucleic Acids Res.* **24**, 3982–3989 (1996).
20. Jurica, M.S., Monnat, R.J., Jr. & Stoddard, B.L. *Mol. Cell* **2**, 469–476 (1998).
21. Gimble, F.S. & Stephens, B. W. *J. Biol. Chem.* **270**, 5849–5856 (1995).
22. Seligman, L.M., Stephens, K.M., Savage, J.H. & Monnat, R.J., Jr. *Genetics* **147**, 1653–1664 (1997).
23. Pingoud, A. & Jeltsch, A. *Eur. J. Biochem.* **246**, 1–22 (1997).
24. Gimble, F.S., Duan, X., Hu, D. & Quijcho, F.A. *J. Biol. Chem.* **273**, 30524–30529 (1998).
25. Stephens, K.M., Monnat, R.J., Jr., Heath, P.J. & Stoddard, B.L. *Proteins* **28**, 137–139 (1997).
26. Otwinowski, Z. & Minor, W. *Methods Enzymol.* **276**, 307–326 (1997).
27. Kissinger, C.R. & Gehlhaar, D.K. *EMPR: A program for crystallographic molecular replacement by evolutionary search.* (Agouron Pharmaceuticals, La Jolla, CA; 1997).
28. Brunger, A. *et al. Acta Crystallogr. D* **54**, 905–921 (1998).
29. Brunger, A. *Acta Crystallogr. D* **47**, 24–36 (1993).
30. Laskowski, R.J., MacArthur, M.W., Moss, D.S. & Thornton, J.M. *J. Appl. Crystallogr.* **26**, 383–290 (1993).
31. Carson, M. *Method Enzymol.* **277**, 493–505 (1997).

Human cystatin C, an amyloidogenic protein, dimerizes through three-dimensional domain swapping

Robert Janowski¹, Maciej Kozak^{1,2}, Elzbieta Jankowska³, Zbigniew Grzonka³, Anders Grubb⁴, Magnus Abrahamson⁴ and Mariusz Jaskolski^{1,5}

¹Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Grunwaldzka 6, 60-780 Poznan, Poland. ²Present address: Department of Macromolecular Physics, Faculty of Physics, A. Mickiewicz University, Umultowska 85, 61-614 Poznan, Poland. ³Department of Organic Chemistry, University of Gdansk, Sobieskiego 18, 80-952 Gdansk, Poland. ⁴Department of Clinical Chemistry, University of Lund, Sweden. ⁵Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Noskowskiego 12/14, 61-704 Poznan, Poland.

The crystal structure of human cystatin C, a protein with amyloidogenic properties and a potent inhibitor of cysteine proteases, reveals how the protein refolds to produce very tight two-fold symmetric dimers while retaining the secondary structure of the monomeric form. The dimerization occurs through three-dimensional domain swapping, a mechanism for forming oligomeric proteins. The reconstituted monomer-like domains are similar to chicken cystatin except for one inhibitory loop that unfolds to form the 'open interface' of the dimer. The structure explains the tendency of human cystatin C to dimerize and suggests a mechanism for its aggregation in the brain arteries of elderly people with amyloid angiopathy. A more severe 'conformational disease' is associated with the L68Q mutant of human cystatin C, which causes massive amyloidosis, cerebral hemorrhage and death in young adults. The structure of the three-dimensional domain-swapped dimers shows how the L68Q mutation destabilizes the monomers and makes the partially unfolded intermediate less unstable. Higher aggregates may arise through the three-dimensional domain-swapping mecha-

nism occurring in an open-ended fashion in which partially unfolded molecules are linked into infinite chains.

Cystatins are single-chain proteins that reversibly inhibit cysteine proteases belonging to the papain (C1) and legumain (C13) families^{1,2}. Three types of cystatins are present in higher animals: type 1, without signal peptides (cystatins A and B); the secretory type 2 cystatins (C, D, E, F, S, SN, SA) and the multidomain type 3 cystatins (high and low molecular weight kininogens)¹. Human cystatin C (HCC) is composed of 120 amino acids¹ and contains, as do other type 2 cystatins, four Cys residues forming two characteristic disulfides (Fig. 1a). Wild type HCC is a high-affinity inhibitor of human C1 family enzymes — for example, cathepsins B, H, K, L, and S. In pathological processes, it forms part of the amyloid deposits in brain arteries of young adults, which leads to fatal cerebral hemorrhage³.

Crystallographic and NMR studies of three cysteine protease inhibitors, chicken cystatin^{4–6}, cystatin B in complex with papain⁷ and cystatin A⁸, have revealed similar overall structure, with three regions implicated in interactions with the enzyme (Fig. 1b). These regions include the N-terminal segment and two hairpin loops, L1 and L2, that are aligned in a wedge-like fashion. The crystal structure of chicken cystatin⁴ (Protein Data Bank (PDB) accession code 1CEW) has defined the general fold of monomeric inhibitors belonging to the cystatin family. Its canonical features include a long α 1 helix running across a large, five-stranded antiparallel β -sheet. The connectivity within the β -sheet is: (N)- β 1-(α 1)- β 2-L1- β 3-(AS)- β 4-L2- β 5-(C), where AS is a broad 'appending structure' that is unrelated to the compact core of the molecule and positioned on the opposite ('back side') end of the β -sheet relative to the N-terminus and loops L1 and L2. The development of effective cysteine protease peptide inhibitors for the treatment of *inter alia* tissue-degenerative diseases (like osteoporosis) and bacterial and viral infections¹ would be greatly facilitated by a three-dimensional structure of HCC. Similarly, such a structure is necessary to understand the ability of HCC to inhibit mammalian legumain⁹, which is important for antigen processing¹⁰. The nature of this inhibitory site is unknown, but it does not overlap with the binding site for papain-like proteases and seems to involve residue Asn 39 (ref. 11). Finally, a three-dimensional model of HCC is necessary for the elucidation of the pathophysiological background of the cerebral hemorrhage produced by this protein, particularly its L68Q variant.