

- teria, M. E. Michel-Beyerle, Ed., *Springer Series in Biophysics Part 6* (Springer, New York, 1990), p. 293.
34. M. Bixon, J. Jortner, M. E. Michel-Beyerle, A. Ogrodnik, *Biochim. Biophys. Acta* **977**, 273 (1989).
35. E. C. Kellogg, S. Kolaczowski, M. R. Wasielewski, D. M. Tiede, *Photosynthesis Res.* **22**, 47 (1989).
36. T. O. Yeates, H. Komiya, D. C. Rees, J. P. Allen, G. Feher, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 6438 (1987); M. Losche, G. Feher, M. Y. Okamura, *ibid.*, p. 7537.
37. M. A. Stephan, K. Lao, S. G. Boxer, *Science* **264**, 810 (1994); D. J. Lockhart and S. G. Boxer, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 107 (1988); V. Palaniappan and D. Bocian, *J. Am. Chem. Soc.* **117**, 3647 (1995).
38. M. E. Michel-Beyerle *et al.*, *Biochim. Biophys. Acta* **932**, 52 (1988); P. O. J. Scherer and S. F. Fischer, *Chem. Phys. Lett.* **131**, 115 (1989); R. Friesner and Y. Won, *Biochem. Biophys. Acta* **977**, 99 (1989).
39. M. A. Thompson and M. C. Zerner, *J. Am. Chem. Soc.* **113**, 8210 (1991); M. Marchi, J. N. Gehlen, D. Chandler, M. Newton, *ibid.* **115**, 4178 (1993). These calculations and those in (26) consistently place the states on the M-side at higher ΔG than the corresponding states on

the L side, and place $P^*BChl_{100}^-$ above P^* in keeping with the native photochemistry.

40. Supported by grant MCB-9405248 from the National Science Foundation and by an Individual National Research Service Award to B.A.H. from the National Institutes of Health. We thank J. Babcock, D. Bocian, and S. Boxer for helpful comments on an early version of the manuscript and D. Daly for assistance with the RC preparations.

5 April 1995; accepted 12 July 1995

Crystal Structure of a Conserved Protease That Binds DNA: The Bleomycin Hydrolase, Gal6

Leemor Joshua-Tor,* H. Eric Xu,† Stephen Albert Johnston, Douglas C. Rees

Bleomycin hydrolase is a cysteine protease that hydrolyzes the anticancer drug bleomycin. The homolog in yeast, Gal6, has recently been identified and found to bind DNA and to act as a repressor in the Gal4 regulatory system. The crystal structure of Gal6 at 2.2 Å resolution reveals a hexameric structure with a prominent central channel. The papain-like active sites are situated within the central channel, in a manner resembling the organization of active sites in the proteasome. The Gal6 channel is lined with 60 lysine residues from the six subunits, suggesting a role in DNA binding. The carboxyl-terminal arm of Gal6 extends into the active site cleft and may serve a regulatory function. Rather than each residing in distinct, separable domains, the protease and DNA-binding activities appear structurally intertwined in the hexamer, implying a coupling of these two activities.

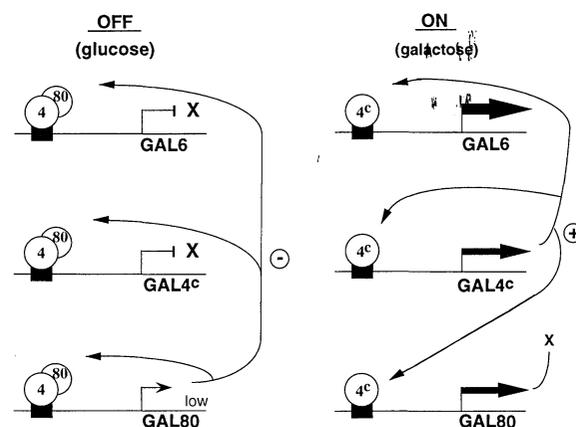
Bleomycin is a small glycometallopeptide produced by *Streptomyces verticillus* that is used as an anticancer drug because of its ability to catalytically cleave double-stranded DNA (1). The endogenous enzyme, bleomycin hydrolase (BH), is a cysteine protease that detoxifies bleomycin by hydrolysis of an amide group (2). The amounts of this enzyme in the cell limit the effectiveness of the drug, because cancer cells with increased expression of BH are resistant to drug treatment, and because tissues that naturally contain small amounts of BH (skin and lung) are especially sensitive to toxic side effects of bleomycin (3).

BH is present in mammals, birds, and reptiles in all tissues tested (2). Homologs of BH which also have peptidase activities have been identified in yeast (4, 5) and bacteria (6). The yeast BH has 35 percent amino acid identity with the bacterial forms (PepC) and 43 percent identity with the available partial cDNA sequence of the mammalian BH (276 of ~450 amino acids), implying an evolutionarily conserved role from bacteria to mammals for this peptidase activity. As in mammalian cancer cells, over-expression of the yeast BH in yeast confers increased resistance to bleomycin (4, 5).

The yeast form of BH was unexpectedly purified as a DNA-binding protein that binds to upstream activating sequences (UAS_G) of the GAL system (4). The 454-residue protein preferentially binds single-stranded DNA ($K_d \sim 10$ nM) over double-stranded ($K_d \sim 1$ μ M). Further, the amounts of BH protein in yeast BH are regulated by the Gal4 regulatory protein, prompting the designation of the gene encoding yeast BH as GAL6 (4). Gal4 is a positive regulatory protein that has a central role in the control of gene expression in the galactose metabolism system of yeast. Gal4 and the GAL genes it controls serve as a model system for studying eukaryotic gene regulation (7). Gal4 regulates GAL6 expression by binding a site in the promoter of GAL6. Gal6 itself apparently negatively regulates GAL gene expression and when GAL6 is deleted, the levels of the GAL gene RNA become three to five times higher (8). Thus, Gal6 appears to be part of a negative feedback regulatory system in yeast, adding another dimension to the cellular roles of BH.

The BHs apparently have cellular functions that have been conserved from bacteria to humans. Presumably these functions involve both proteolysis and DNA binding. We now describe the three-dimensional structure of Gal6. Unlike many bifunctional proteins, the DNA-binding and protease activities of Gal6 are intimately intertwined at both the tertiary and quaternary structural levels with an organization strikingly analogous to that of the 20S proteasome (9). This feature may be key to understanding the conserved cellular functions of this

Fig. 1. A general system for the overproduction of proteins in yeast as applied to Gal6. The target gene (*GAL6*) is under control of a Gal4 responsive promoter (*GAL1*) and is on a multicopy plasmid. Gal4 is a positive regulatory protein that is repressed by Gal80 protein, which is also under Gal4 transcriptional control. A mutant form of Gal4, Gal4^c, is also under Gal4 regulation. In the absence of the inducer, galactose, there is no expression of the target gene or of the *GAL4^c* gene. On induction, the *GAL6* gene is expressed. Its expression is maximized by overexpression of the Gal4^c protein, which is not subject to repression by Gal80 protein and saturates the Gal6 promoter sites. In this system, a modification of earlier ones (39, 40), the target protein is produced in a eukaryotic host; expression of the target protein is completely repressed until induction, allowing production of toxic proteins; the negative feedback regulation of the Gal80 protein (which is required to keep the system off until induction) is bypassed after induction by overproduction of the mutant Gal4^c protein. When applied to Gal6, protein production saturates 4 to 6 hours after induction, with Gal6 comprising 5 to 10 percent of the total cellular protein (the most abundant protein band in total cell extracts). The yeast strain stops growing, presumably because of overproduction of the protein. Gal6 protein was purified after induction by conventional chromatography (4) to more than 98 percent purity.



unusual family of proteins.

The Gal6 protein was initially purified from wild-type yeast (4). However, for the crystallographic studies, the protein was purified using a yeast overproduction system that produces more than 100 mg of purified protein from 10 liters of culture and has several attractive features (Fig. 1). Details of the crystallization and structure determination of Gal6 at 2.2 Å resolution are provided in Table 1. The six identical subunits of the Gal6 hexamer are arranged in a ring with 32-point group symmetry that creates a prominent channel of ~22 Å diameter along the threefold axis (Fig. 2A). The monomers have a papain-like polypeptide fold as the core, with additional structural and functional modules inserted into loop regions (Fig. 3). As a result, the Gal6 monomer has a three-domain structure composed of a catalytic (protease) domain, an oligomerization or hook domain, and a helical domain (Fig. 3A). Elements of α -helical and β -sheet secondary structure are designated by letters and numbers, respectively, increasing consecutively from the amino to the carboxyl terminus (Fig. 3B). Prime symbols are used in reference to a symmetry-related monomer.

Subunit structure. The catalytic domain forms the core of the Gal6 subunit structure, with an overall fold similar to that of members of the papain family, including the plant enzymes papain (10), papaya proteinase Ω (11), actinidin (12), and calotropin (13), and human cathepsin B (14). The root mean square (rms) deviation of the backbone of the residues forming the core (126 residues with 25 percent identity) is 1.9 Å between Gal6 and papain, which is comparable to the expected deviation based on an empirical relation described by Chothia and Lesk (15). The structural integrity of the fold is maintained even though Gal6 has no disulfides, in contrast to the three or more disulfides present in members of the monomeric papain family.

The catalytic domain of Gal6 consists of two subdomains; these are designated the L and R subdomains by analogy to papain. The active site consists of residues from both subdomains. The L subdomain consists of the three helices E, H, and K, and part of a fourth helix F. The active site cysteine (Cys⁷³) is positioned at the amino terminus of helix E, facing the R subdomain. The first two turns of helix F super-

impose on the second helix in the plant proteases and cathepsin B. The remaining four turns of helix F extend into the helical domain. Polypeptide insertions occur in the L subdomain at positions in the papain fold where there are significant differences between previously determined structures of members of this family, such as the location of the occluding loop in cathepsin B (14). The main feature of the R subdomain is two highly twisted β -pleated sheets, which fold into an open barrel-like structure with an extension. Helix L closes the barrel on one side, as in the other proteases; however, unlike the other proteases, in Gal6 a short β -hairpin (strands 8 and 9), which is part of a six-stranded sheet, closes the other side of the barrel.

The carboxyl-terminal region of the Gal6 polypeptide chain also contributes to

the catalytic domain, and may have a regulatory function. From β -strand 13, which is part of the papain fold, the polypeptide chain extends outward to the surface of the molecule, forming helix O, and then attaches back at the edge of the second β sheet with strand 14. The polypeptide chain then heads directly toward the active site of the catalytic domain, occupying much of the active site cleft and blocking the approach to Cys⁷³ in the active site (Fig. 4). Although the DNA-derived protein sequence, which was used in the overexpression system, should end with a lysine residue at the carboxyl terminus, we do not see any evidence for this lysine in the electron density maps. Carboxyl-terminal sequencing and mass spectrometry confirmed that this lysine is indeed absent in the mature protein. Apparently the carboxyl-terminal lysine is removed posttranslationally. The

Table 1. Crystallographic data and statistics. Data sets from the hexagonal native and derivative crystals of Gal6 (28) were collected by cryocrystallographic techniques at the Stanford Synchrotron Radiation Laboratory (SSRL). The data were collected on a 300-mm MAR Research imaging plate detector system and integrated and scaled with the use of the HKL package (29). Data from the orthorhombic crystal were collected on an R-Axis IIC imaging plate system mounted on an RU200 rotating anode x-ray generator equipped with a Molecular Structure Corporation (MSC) cooling system. These data were integrated with the MSC software and scaled by means of the CCP4 package (30). Each data set was collected from a single crystal. Single isomorphous replacement combined with anomalous differences (SIRAS) phases from the native and derivative data sets of the hexagonal form were calculated with the use of PHASES (31), initially with two mercury sites found by interpretation of difference Patterson maps. The phasing power was 2.45 for the isomorphous differences and 1.25 for the anomalous differences, with a mean figure of merit of 0.54 at 2.2 Å resolution. Solvent flattening was initially done with PHASES and continued with iterative cycles of solvent flattening and histogram matching with SQUASH (32). A polyaniline chain containing approximately 75 percent of the residues could be fitted into this map with the program O (33). The electron density maps were improved by averaging (34) between the hexagonal and orthorhombic crystal forms. The coordinate transformations were established with Crowther's fast rotation function (35) and a locally written "brute force" translation search. The molecular envelope was defined via MAMA (36) and manually expanded at places where the electron density appeared to extend beyond the envelope. Positional refinement of the model against the derivative data set was performed with X-PLOR (37) and calculated phases were combined with the SIRAS phases with SIGMAA (30), at 2.5 Å resolution. Electron density maps calculated with these phases were used in an iterative process to complete the molecular model. More than 90 percent of the (ϕ , ψ) angles of the model are in the most favored regions and all residues are in allowed regions of the Ramachandran plot (38).

Crystal Source (wavelength, Å)	Native-P6 ₃ 22 SSRL (1.08)	EMTS-P6 ₃ 22 SSRL (1.08)	Native-C222 ₁ RU200 (1.54)
Resolution (Å)	2.20	2.20	3.12
Reflections			
Observed	183105	202481	68854
Unique	29482	31451	31094
Redundancy	6.4	6.2	2.2
Completeness (percent)	94.6	99.8	59.9
Last shell	92.1	99.8	42.6
R_{merge}^*	0.073	0.074	0.124
<i>Refinement statistics</i>			
Resolution range (Å)	10.0 to 2.2	Crystallographic R factor†	0.209
Number of reflections	30018	Free R ‡ factor	0.274
σ cutoff	None	rms deviations from ideality	
Number of atoms:		Bonds (Å)	0.008
Protein (nonhydrogen)	3584	Angles (°)	1.38
Hg	4	Dihedrals (°)	21.70
Water	239	Impropers (°)	1.18
Sulfates	3		
Glycerol	1		

L. Joshua-Tor and D. C. Rees are in the Division of Chemistry and Chemical Engineering, 147-75CH, California Institute of Technology, Pasadena, CA 91125, USA. H. E. Xu and S. A. Johnston are in the Departments of Medicine and Biochemistry, University of Texas Southwestern Medical Center, Dallas TX 75235-8573, USA.

*To whom correspondence should be addressed.

†Present address: Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

* $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$, where I is the observed intensity and $\langle I \rangle$ is the average intensity obtained from symmetry related and multiple measurements. † R factor = $\sum |F_o| - k|F_c| / \sum |F_o|$ summed over 90 percent of the reflections. ‡Free $R = \sum |F_o| - k|F_c| / \sum |F_o|$ summed over 10 percent of randomly selected reflections not included in the refinement.

bacterial forms of Gal6 show high sequence divergence at this position, suggesting that the carboxyl terminus may be removed from the ones that extend beyond Ala⁴⁵³ (6).

The hook domain consists of two modules attached to the core, but separated by 290 amino acids (Fig. 3). The first of these modules, which makes up most of the hook domain, is located at the amino terminus (residues 2 to 54) and consists of four α helices (A to D). The first two α helices (A and B) are connected by a short turn to form an L-shaped structure. The α helices C and D are also perpendicularly oriented to each other. These four α helices extend out from the rest of the subunit and resemble a hook. The second module of the hook domain, termed the loop-helix module, is closer to the carboxyl terminus (residues 342 to 365) and consists of a long loop that

extends from the catalytic domain, followed by helix M. The amino-terminal module wraps around helices I' and J' of the helical domain of a second monomer, and the loop-helix module binds the helical domain from the other side, thereby clamping the two subunits together to form a dimer (Fig. 2B). A pair of B helices also form a parallel coiled coil between two different monomers that are diagonally related. This coiled-coil interaction, which is formed between a top monomer from one dimer with the bottom monomer of an adjacent dimer, provides a structural contribution to hexamer formation.

The helical domain is formed by two pairs of antiparallel α helices that extend from the catalytic domain. The first pair of helices is composed of the last four turns of helix F and the entire helix G, and the second pair con-

sists of helices I and J. Helices F and G are positioned against the loop that connects β strands 5 and 6. A potentially important interaction between the helical and catalytic domains is formed by the stacking of the Arg¹²⁹ side chain in helix G with the side chain of Trp⁴⁴⁶ in the carboxyl-terminal arm. This region of the carboxyl-terminal arm is located after β strand 14, just before the arm enters the active site cleft. These residues are stacked on their dimer counterparts Arg^{129'} and Trp^{446'} forming a four-residue stack. Possible implications of this interaction are discussed below.

The hexamer. The toroidal Gal6 hexamer has overall dimensions ~ 125 by 115 by 85 Å (Fig. 2). Extensive intersubunit contacts stabilize the hexamer; the accessible surface area of an isolated monomer is 21800 Å², whereas that of the hexamer is 95400 Å², or 15900 Å² per monomer. Consequently, about 6000 Å² or 27 percent of the solvent-accessible surface of each monomer is buried on oligomerization. The hexameric arrangement is maintained in both crystal forms and in solution, as demonstrated by sedimentation experiments (16). Most of the buried surface area of each subunit (~ 4000 Å² or 19 percent) is associated with dimer contacts where the hook domain of one monomer clamps the helical domain of a second monomer. Because of the extensive dimer interaction, the hexamer may be considered as a trimer of dimers. The dimers are arranged around the threefold axis to form the hexamer through the coiled-coil interactions involving B helices from different subunits described above. When the top and bottom trimers are separated, the hook domains protrude from the surfaces of the trimers. The β hairpin modules are positioned at the two rims of the channel, with the turn between the two strands pointing into the channel. The two protease active sites within each dimer are separated by about 32 Å and are positioned nearly on top of each other, when viewed down the threefold axis of the hexamer. Active sites between threefold related subunits in the same trimer are separated by about 45 Å.

Among the striking features of the hexamer is the channel traversing its center. The channel may be divided into three regions; two symmetry-related outer regions that are approximately 22 Å in diameter and 25 Å in length, that open out into a third central region that is approximately 45 Å in diameter and 30 Å in length. The surface of the channel is lined with 60 lysine residues, 10 from each monomer. Three sets of three lysines, situated on the turn between strands 2 and 3 of the β hairpin (residues 242, 244, and 245), surround the outer rims of the channel. As a

Fig. 2. The three-dimensional structure of the Gal6 hexamer. The six subunits are shown in different colors. This figure, as well as Figs. 3A, 4, and 6, A and C, were prepared with the program MOLSCRIPT (41) and rendered with RASTER3D (42). **(A)** A top view along the threefold axis. The central channel is a prominent feature. The dimeric interactions are between top and bottom subunits (specifically, between the purple and blue subunits, the red and pink subunits, and the green and yellow subunits). Trimer interactions occur between diagonally related subunits (the blue and red, the pink and green, and the yellow and purple subunits). A portion of the helical domain extends from the equator of the hexamer. The turn between strands 2 and 3 of the β hairpin is pointing toward the channel at the rim, three on either side; **(B)** A side view perpendicular to the top view. The dimer interactions are apparent in this view, with the oligomerization domain of the top-front subunit (purple) wrapping around the helical domain of the bottom-front subunit (blue) and vice versa. The coiled-coil of the trimer interaction is visible at the back of the hexamer in this view (pink and green subunits). The helical domains are at the equator of the hexamer. Helices I and J of each subunit extend from the surface of the hexamer like pot handles.

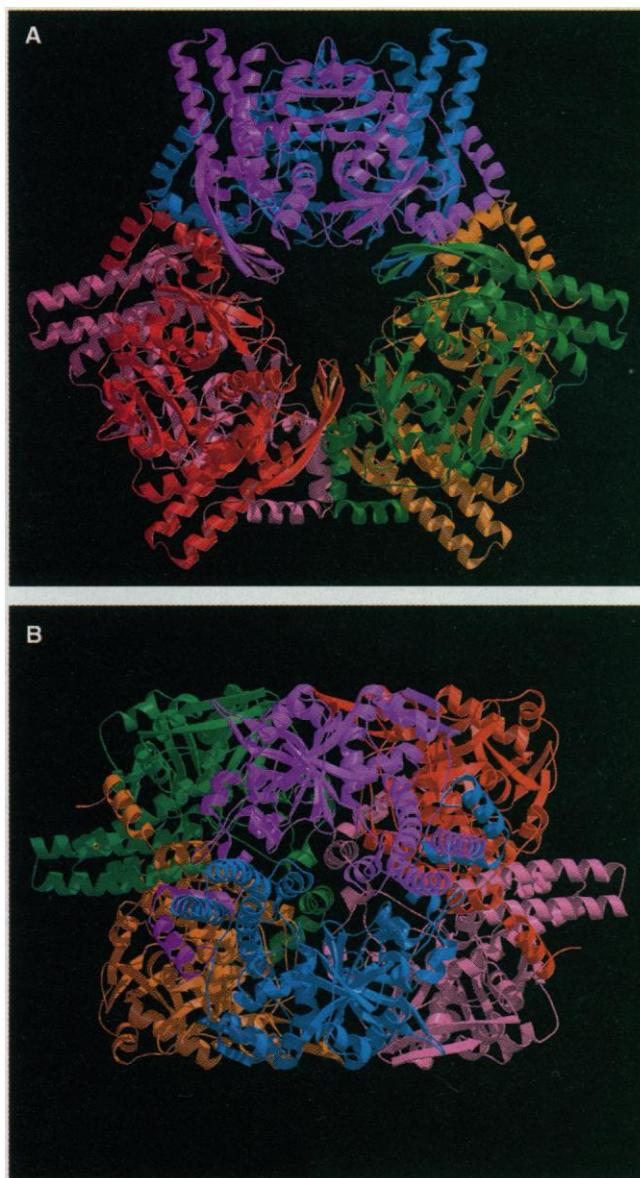


Fig. 3. The Gal6 subunit structure. **(A)** A stereo pair with the oligomerization domain shown in purple, the catalytic domain in blue, the helical domain in yellow, and the carboxyl-terminal arm shown in a dark pink. **(B)** Sequence alignment of Gal6 and its homologs: *Lactococcus lactis* pepC (lcPepC) (43) and rabbit bleomycin hydrolase (rBH) for which only a partial cDNA is available (2). Identical residues are highlighted in blue. A structure-based sequence alignment of papain is shown on the bottom row. The structure of papain was superimposed on the core of Gal6, and the structural sequence alignment of 126 residues was done by visual inspection. The secondary structure assignments of Gal6 are shown above the sequence: α helices are shown as zig-zag elements and β strands as arrows and color-coded according to the domains shown in (A).

consequence of these lysines, and the ones at positions 31, 42, 68, 93, 175, 333, 334, 396, and 400, a high positive electrostatic potential exists inside the channel, which is indicated by the blue-colored surface in Fig. 5. Since the overall charge of the Gal6 hexamer is neutral at pH 7, the positively charged channel reflects an asymmetric charge distribution over the protein surface, and not an overall positive charge.

The prominent positive electrostatic potential inside the channel suggests that it represents the region of Gal6 involved in DNA binding. A double-stranded piece of DNA can be modeled into the channel of Gal6, although the fit is rather snug. Since the hexamer seems stable, we assume that Gal6 does not assemble around DNA, but rather binds to it as a hexamer. Therefore, although the channel may contain the site for DNA binding, Gal6 may undergo conformational changes upon DNA binding that would better allow accommodation of a DNA helix or hairpin in the channel. These conformational changes could also regulate accessibility of the protease active site. Alternatively, DNA may bind to the outer surface of Gal6 through other basic residues near the surface of each monomer, or DNA may bind through the α helices projecting from the surface of the hexamer. Initial mutagenesis experiments, however, demonstrate that, at least, lysines 242, 244, and 245 of the β hairpin loop of the channel are required for single-stranded DNA binding (17). Loops are also used for DNA binding in p53 (18) and in NF- κ B (19).

The active site. The active site triad of Gal6, formed by the side chains of Cys⁷³, His³⁶⁹, and Asn³⁹² resembles the active site triad of all other structurally characterized cysteine proteases. As in papain, the side chain of a glutamine residue, Gln⁶⁷, is positioned to stabilize the oxyanion hole in Gal6. A major difference in the active site cleft of Gal6, relative to papain, is the projection of the carboxyl terminus of Gal6

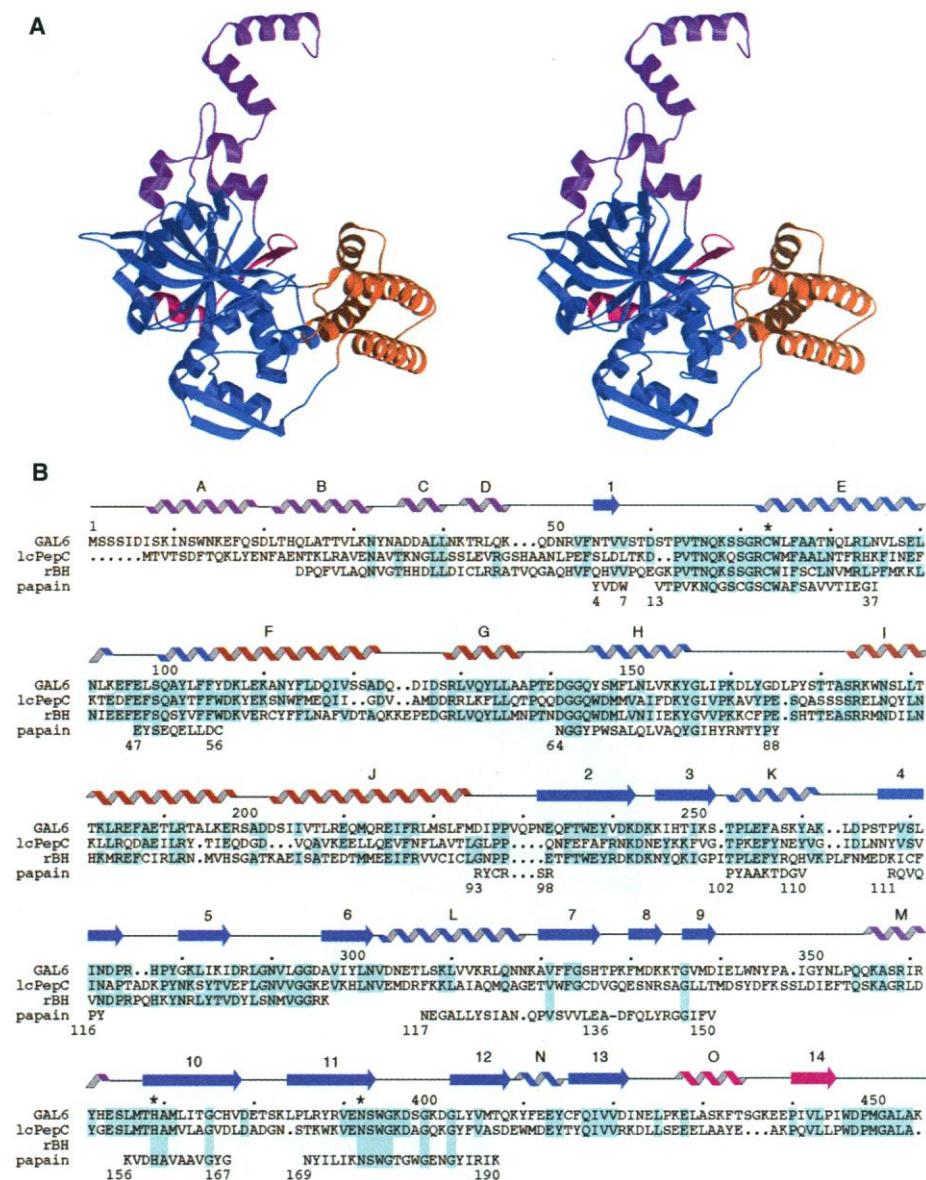


Fig. 4. The active site of Gal6 showing the carboxyl-terminal arm in the active site cleft. Gal6, shown in blue with the conserved carboxyl-terminal and active site residues in a stick representation, starting with Trp⁴⁴⁶, is superimposed on a complex of papain with the cysteine protease inhibitor leupeptin. The papain active site residues and leupeptin are shown in orange.

into its active site. The replacement of Asp¹⁵⁸ of papain, which has been associated with the absence of carboxypeptidase activity in papain because of electrostatic repul-

sion with a free carboxyl terminus (20), by Thr³⁶⁸ in Gal6, may facilitate the binding of the carboxyl terminus of Gal6 in the active site. Remarkably, the structure of

Fig. 5. GRASP (44) representation of the electrostatic potential at the surface of Gal6. The coloring scheme depicts potentials $< -10 k_B T$ in red and $> 10 k_B T$, in blue, and where k_B is the Boltzmann constant and T is the absolute temperature. For this calculation, the dielectric constants were 80 for the solvent and 2 for the protein, and the ionic strength was 0. The left panel shows the view down the threefold axis. On the right is a slice approximately along the threefold axis showing the positive interior of the channel.

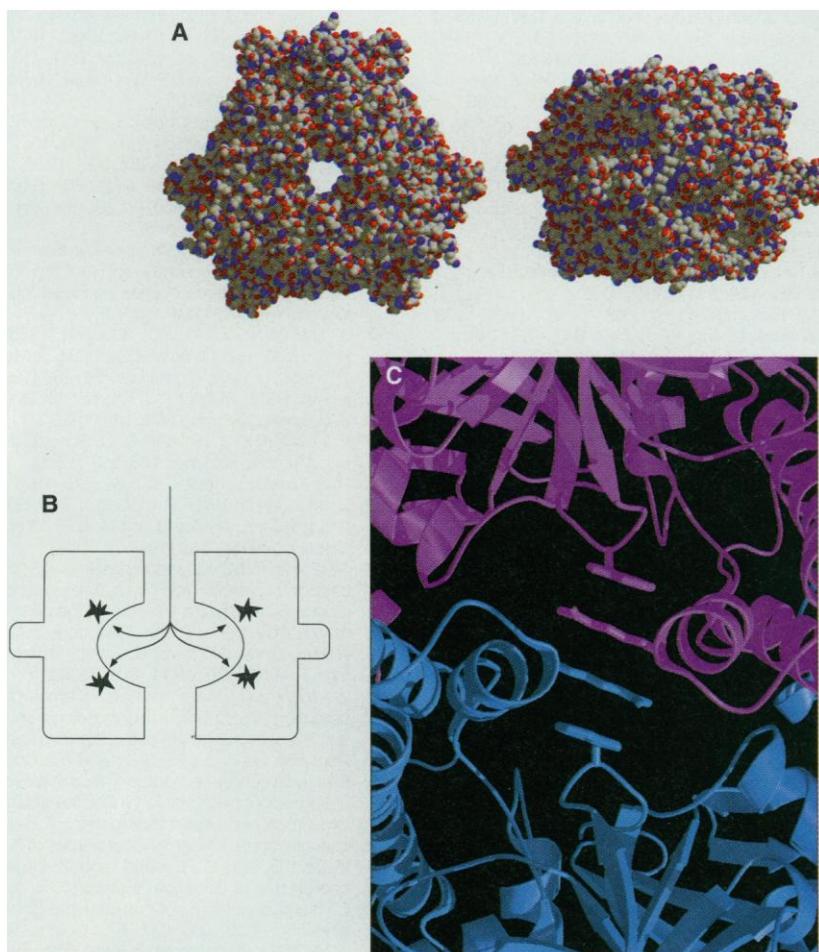
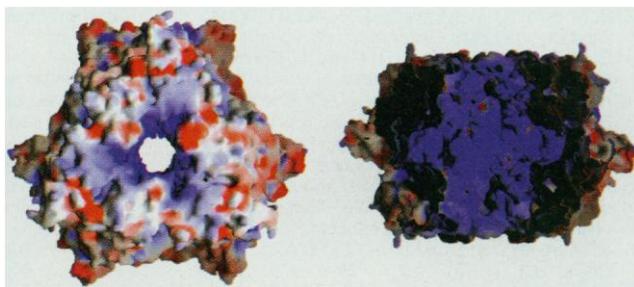


Fig. 6. (A) A space-filling representation of the Gal6 hexamer. The only opening to the interior of the protein is through the channel. The left panel shows a view along the threefold axis, while the right panel provides a side view. (B) A schematic showing the "inner approach" to the active site. A slice through the middle of the protein along the threefold axis with a side view. Four out of the six active sites are shown and are represented as distorted stars. (C) The putative "outer approach" site showing the stacking interaction between the two carboxyl-terminal arm Trp⁴⁴⁶ and the two Arg¹²⁹ side chains at the dimer interface.

the carboxyl terminus resembles the structures of inhibitors complexed with papain (for example, leupeptin, E-64, chloromethyl ketones, and stefin B) (21). The peptide aldehyde leupeptin (Ac-Leu-Leu-arginal) forms a reversible, covalent transition state analog with papain. In the superposition between Gal6 and this complex, the carboxyl terminus of Gal6 closely

matches the position of the inhibitor in that the carbonyl group of Ala⁴⁵¹ is located where the acetyl group of leupeptin is placed; Leu⁴⁵² coincides with the first leucine of leupeptin at subsite P3, and Ala⁴⁵³ overlaps with the middle leucine of leupeptin at subsite P2 (Fig. 4). There is an arginyl group in leupeptin at P1, but no equivalent in Gal6 as the carboxyl-termi-

nal Lys⁴⁵⁴ was found to be cleaved from the mature protein.

The approach to the active site cleft. One of the puzzling aspects about Gal6 is how substrates reach the protease active site. In addition to the presence of the carboxyl-terminal residues in the active site cleft, the site is also essentially inaccessible from the outer surface of the hexamer. The question of how a substrate can reach the active site is critical. The six active site clefts open out toward the wider, central region of the channel, and they can be accessed from that central region of the channel interior through a path labeled the "inner approach" (Fig. 6). This is very reminiscent of the arrangement of the active sites within the channel of the proteasome (9). As the opening of the channel is ~ 20 Å, a putative protein substrate must be either of limited size or partially unfolded to reach the active site. Should larger substrates be cleaved by Gal6, either the channel must be enlarged, or alternative access routes to the active sites must be created. The latter scenario forms the basis of a second, speculative approach to the active site. The active site clefts would be accessible from the outer surface of the hexamer if the helical domains of two monomers in a dimer could be spread apart. A potential pathway ("the outer approach") could be generated by subunit motions that separate the four stacked side chains at the dimer interface. These residues, Trp⁴⁴⁶ from the carboxyl-terminal arm, Arg¹²⁹ from the helical domain, along with their counterparts in the dimer-related subunit (Fig. 6C), are conserved in the Gal6 homologs. In addition, Arg¹²⁹ hydrogen bonds to the carbonyl oxygen of Asp¹²⁷ of the dimer partner. Separation of the two dimeric helical domains through repositioning of Trp⁴⁴⁶ (shown in stick representation also in Fig. 4) may also help dislocate the carboxyl termini from the active site clefts, with the end result of allowing access to the active site from the outer side of the Gal6 hexamer. This scenario implies that the enzymatic activity of Gal6 could be regulated by interactions with other macromolecules that mediate changes in subunit positioning in the hexamer.

The combination of the overall architecture of the hexamer and the presence of the carboxyl-terminal arm in the active site cleft limiting the access to the active site cleft may explain why the peptidase activity observed so far in vitro has been limited to small peptides (5). If the carboxyl-terminal arm is not displaced, there may be room for only one residue between the carboxyl terminus of Gal6 and the catalytic sulfhydryl group. In addition, the carboxyl terminus may form ionic interactions with an amino terminus of a substrate, thus locking it in a proper orientation for cleavage, which may account for

the aminopeptidase activities of the PepC homologs of Gal6. The spatially restricted active site may also help select the amide bond in the β -amino alanine group of bleomycin for cleavage during hydrolysis of the drug. PepC from *Lactococcus lactis* has been crystallized (22), and a comparison of the structures of these two proteins may be instructive.

Many of the characteristics of Gal6 are provided by the subunit arrangement in this hexameric molecule. Rather than having distinct DNA-binding and protease domains, these two activities appear intimately connected by the overall architecture of the oligomer. The overall toroidal structure of the Gal6 hexamer is reminiscent of the structures of other DNA-binding proteins such as the β subunit of DNA polymerase III (23), PCNA (24), and topoisomerase I (25), where the hole or channel is thought to be the site for nucleic acid binding. Topoisomerase I is proposed to open and close around the DNA (25), whereas an auxiliary protein is required to assemble the stable PCNA trimer around DNA (26). Ringlike structures are not restricted to DNA binding proteins; the recent crystal structure of the 20S proteasome from *Thermoplasma acidophilum* (9) is instructive, because there are many parallels between the proteasome structure and the structure of Gal6. Both proteins are ringlike oligomers that contain several protease active sites situated in a central channel protected and isolated from the outside. Further, both proteins have a larger central cavity and two side cavities of smaller diameter, with the active sites lining the central cavity. The activity of the 20S proteasome particle is regulated by interactions with other molecules, leading to activation or inhibition of different proteolytic activities (27). In parallel, it is likely that Gal6 protease activity is affected by DNA-binding and by interactions that modulate the effects of the carboxyl-terminus of Gal6 in the active site cleft. The structure of Gal6 suggests that the protease and DNA-binding activities of Gal6 may be linked by the quaternary structure arrangement of the subunits, which could be modulated by binding of other molecules. Definition of the structural relationship between these activities should facilitate a better understanding of the basic, conserved role of Gal6 in its cellular functions from bacteria to humans.

REFERENCES AND NOTES

1. J. Stubbe and J. W. Kozarich, *Chem. Rev.* **87**, 1107 (1987); S. A. Kane and S. M. Hecht, *Prog. Nucleic Acids Res. Mol. Biol.* **49**, 313 (1994).

2. S. M. Sebt, J. E. Mignano, J. P. Jani, S. Srimatkandada, J. S. Lazo, *Biochemistry* **28**, 6544 (1989).
3. S. M. Sebt, J. P. Jani, J. S. Mistry, E. Gorelik, J. Lazo, *Cancer Res.* **51**, 227 (1991).
4. H. E. Xu and S. A. Johnston, *J. Biol. Chem.* **269**, 21177 (1994). In this reference Gal6 was predicted to be a tetramer based on gel filtration and native gel electrophoresis. However, analytical centrifugation results show that it is a hexamer. Other estimates of the multimerization state of Gal6 homologs have been tetramer, pentamer, or hexamer.
5. N. G. Kambouris, D. J. Burke, C. E. Creutz, *J. Biol. Chem.* **267**, 21570 (1992); C. Enekel and D. H. Wolf, *ibid.* **268**, 7036 (1993); U. Magdolen, G. Müller, V. Magdolen, W. Bandlow, *Biochim. Biophys. Acta* **1171**, 299 (1993).
6. M.-P. Chapot-Chartier, M. Nardi, M.-C. Chopin, A. Chopin, J.-C. Gripon, *Appl. Environ. Microbiol.* **59**, 330 (1993); Y. Wohrlab and W. Bockelmann, *Int. Dairy J.* **3**, 685 (1993); L. Fernández, T. Bhowmik, J. L. Steele, *Appl. Environ. Microbiol.* **60**, 333 (1994); M.-P. Chapot-Chartier, F. Rul, M. Nardi, J.-C. Gripon, *Eur. J. Biochem.* **224**, 497 (1994); E. Vesanto, P. Varmanen, J. L. Steele, A. Palva, *ibid.*, p. 991.
7. M. Johnston, M. Carlson, in *The Molecular Biology of the Yeast *Saccharomyces cerevisiae**, J. R. Broach and E. W. Jones, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1992), pp. 193–281.
8. H. E. Xu, W. Zheng, S. A. Johnston, unpublished results.
9. J. Löwe *et al.*, *Science* **268**, 533 (1995).
10. I. G. Kamphuis, K. H. Kalk, B. A. Swarte, J. Drenth, *J. Mol. Biol.* **179**, 233 (1984).
11. R. W. Pickersgill, P. Rizkallah, G. W. Harris, P. W. Goodenough, *Acta Crystallogr.* **B47**, 766 (1991).
12. E. N. Baker, *J. Mol. Biol.* **141**, 441 (1980).
13. U. Heinemann, G. P. Pal, R. Hilgenfeld, W. Saenger, *ibid.* **161**, 591 (1982).
14. D. Musil *et al.*, *EMBO J.* **10**, 2321 (1991).
15. C. Chothia and A. M. Lesk, *EMBO J.* **5**, 823 (1986).
16. A. Wu, L. Joshua-Tor, S. A. Johnston, D. C. Rees, unpublished results.
17. W. Zheng and S. A. Johnston, unpublished results.
18. Y. Cho, S. Gorina, P. D. Jeffrey, N. P. Pavletich, *Science* **265**, 346 (1994).
19. G. Ghosh, G. V. Duyne, S. Ghosh, P. B. Sigler, *Nature* **373**, 303 (1995); C. W. Müller, F. A. Rey, M. Sodeoka, G. L. Verdine, S. C. Harrison, *ibid.* **373**, 311 (1995).
20. G. Lowe and Y. Yuthavong, *Biochem. J.* **124**, 107 (1971).
21. E. Schröder, C. Phillips, E. Garman, K. Harlos, C. Crawford, *FEBS Lett.* **315**, 38 (1993); K. I. Vrughese *et al.*, *Biochemistry* **28**, 1330 (1989); J. Drenth, K. H. Kalk, H. M. Swen, *ibid.* **15**, 3731 (1976); M. T. Stubbs *et al.*, *EMBO J.* **9**, 1939 (1990).
22. M. Y. Mistou *et al.*, *J. Mol. Biol.* **237**, 160 (1994).
23. X. P. Kong, R. Onrust, M. O'Donnell, J. Kuriyan, *Cell* **69**, 425 (1992).
24. T. S. R. Krishna, X. P. Kong, S. Gary, P. M. Burgers, J. Kuriyan, *Cell* **79**, 1233 (1994).
25. A. Mondragón, C. D. Lima, J. C. Wang, *Nature* **367**, 138 (1994).
26. T. Tsurimoto and B. Stillman, *J. Biol. Chem.* **266**, 1950 (1991).
27. M. Rechsteiner, L. Hoffman, W. Dubiel, *ibid.* **268**, 6065 (1993).
28. Gal6 was crystallized by the hanging drop vapor diffusion method; mixing equal volumes of protein (36 mg per milliliter of solution of Gal6 in 25 mM Tris buffer, pH 8.5, and 10 percent glycerol) with equal volumes of a reservoir solution containing 21 percent polyethylene glycol (PEG) 4K, 0.3 M $(\text{NH}_4)_2\text{SO}_4$, 2 mM dithiothreitol (DTT), and 100 mM Tris buffer, pH 8.5, and equilibrating against the reservoir solution at 22°C. Thicker (70 to 100 μm) rodlike hexagonal crystals were obtained after recrystallization by the addition of water to the hanging drop in a 2:1 volume ratio. The hexagonal crystals are of space group $P6_322$, with cell dimensions $a = b = 151.15 \text{ \AA}$, $c = 89.88 \text{ \AA}$ and one monomer in the asymmetric unit. The crystals diffract beyond 2.2 Å resolution when a synchrotron radiation source is used. They are extremely radiation sensitive, and the use of cryocrystallographic techniques was essential for the structure determination. Derivative crystals were obtained by cocrystallization with 1 mM sodium ethylmercurithiosalicylate (EMTS), without the use of the recrystallization step. Crystals of a second, orthorhombic form, were grown under similar conditions to the hexagonal form, except that neither $(\text{NH}_4)_2\text{SO}_4$ nor the recrystallization procedure was used. Orthorhombic crystals grow in space group $C222_1$, with cell dimensions $a = 85.80 \text{ \AA}$, $b = 206.22 \text{ \AA}$, $c = 170.24 \text{ \AA}$ and three monomers in the asymmetric unit. These crystals diffracted rather poorly, to $\sim 3.1 \text{ \AA}$ resolution, but proved useful in the structure determination.
29. Z. Otwinowski, in *Data Collection and Processing*, L. Sawyer, N. Isaacs, S. Burley, Eds., (Science and Engineering Research Council, Warrington, UK, 1993), pp. 56–62.
30. S. Bailey, *Acta Crystallogr.* **D50**, 760 (1994).
31. W. Furey and S. Swaminathan, *American Crystallographic Association Mtg. Abstr. Ser.* **2**, 18, 73 (1990).
32. K. Y. J. Zhang, *Acta Crystallogr.* **D49**, 213 (1993).
33. T. A. Jones, J. Y. Zou, S. W. Cowan, M. Kjeldgaard, *ibid.* **A47**, 110 (1991).
34. M. G. Rossman, Ed., *Molecular Replacement Method* (Gordon and Breach, New York, 1972); G. Brice, *Acta Crystallogr.* **A32**, 832 (1976).
35. R. A. Crowther, in *The Molecular Replacement Method*, M. G. Rossman, Ed. (Gordon Breach, New York, 1972), pp. 173–178.
36. G. J. Kleywegt and T. A. Jones, in *From First Map to Final Model*, S. Bailey, R. Hubbard, D. Waller, Eds., (Science and Engineering Research Council, Warrington, UK, 1994), pp. 59–66.
37. A. T. Brünger, *J. Mol. Biol.* **203**, 803 (1988).
38. R. A. Laskowski, R. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Crystallogr.* **26**, 283 (1993).
39. K. King, H. G. Dohlman, J. Thorner, M. G. Caron, R. J. Lefkowitz, *Science* **250**, 121 (1990).
40. L. M. Mylin, K. J. Hofman, L. D. Schultz, J. E. Hopper, *Methods Enzymol.* **185**, 297 (1990).
41. P. J. Kraulis, *J. Appl. Cryst.* **24**, 946 (1991).
42. D. J. Bacon and W. F. Anderson, *J. Mol. Graphics* **6**, 219 (1988); E. A. Merritt and M. E. P. Murphy, *Acta Crystallogr.* **D50**, 869 (1994).
43. Two other PepC sequences are known: from *Lactobacillus helveticus* and from *Streptococcus thermophilus*, which are 47 and 70 percent identical, respectively, to PepC from *Lactococcus lactis*.
44. A. Nicholls, K. A. Sharp, B. Honig, *Proteins; Structure, Function, Genet.* **11**, 281 (1991).
45. We thank M. Stowell for help in synchrotron data collection; W. Zheng for discussions and unpublished results; members of the Stanford Synchrotron Radiation Laboratory (SSRL) and the Cornell High Energy Synchrotron Source (CHESS) for assistance during data collection; A. Wu for the sedimentation experiments and mass spectrometry; K. Swiderek for sequencing; R. Huber for coordinates of human cathepsin B, and W. Saenger and R. Hilgenfeld for coordinates of calotropin. Supported in part by the Caltech Consortium in Chemistry and Chemical Engineering (L.J. and D.C.R.), grants from NIH (GM40700) and Human Frontier Science Program (S.A.J.), and by a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund for Medical Research (L.J.). The diffraction facility was supported in part by USPHS grant GM45162 and the Beckman Institute (D.C.R.), and the Howard Hughes Medical Institute (to P. J. Bjorkman). The rotation camera facility at the Stanford Synchrotron Radiation Laboratory is supported by the Department of Energy, Office of Basic Energy Sciences and the National Institutes of Health Biomedical Research Technology Program, Division of Research Resources. X-PLOR calculations were done on the CRAY C90 at the San Diego Supercomputer Center, supported by the NSF. Coordinates have been deposited in the Brookhaven Data Bank (IGCB) and are available by e-mail (leemor@citray.caltech.edu).

3 March 1995; accepted 7 July 1995