Protein ⑤ Science

High resolution refinement of beta-galactosidase in a new crystal form reveals multiple metal-binding sites and provides a structural basis for alpha-complementation [In Process Citation]

DH Juers, RH Jacobson, D Wigley, XJ Zhang, RE Huber, DE Tronrud and BW Matthews

References	Article cited in: **http://www.proteinscience.org/cgi/content/abstract/9/9/1685#otherarticles**
Email alerting service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

Notes	

To subscribe to *Protein Science* go to:
http://www.proteinscience.org/subscriptions/

High resolution refinement of *β*-galactosidase in a new crystal form reveals multiple metal-binding sites and provides a structural basis for *α*-complementation

DOUGLAS H. JUERS,[1] RAYMOND H. JACOBSON,[1,3] DALE WIGLEY,[2] XUE-JUN ZHANG,[1,4]
REUBEN E. HUBER,[1,5] DALE E. TRONRUD,[1] AND BRIAN W. MATTHEWS[1]

[1]Institute of Molecular Biology, Howard Hughes Medical Institute and Department of Physics, 1229 University of Oregon,
 Eugene, Oregon 97403-1229
[2]Sir William Dunn School of Pathology, University of Oxford, South Parks Road, Oxford OX1 3RE, United Kingdom

Abstract

The unrefined fold of *Escherichia coli* *β*-galactosidase based on a monoclinic crystal form with four independent tetramers has been reported previously. Here, we describe a new, orthorhombic form with one tetramer per asymmetric unit that has permitted refinement of the structure at 1.7 Å resolution. This high-resolution analysis has confirmed the original description of the structure and revealed new details. An essential magnesium ion, identified at the active site in the monoclinic crystals, is also seen in the orthorhombic form. Additional putative magnesium binding sites are also seen. Sodium ions are also known to affect catalysis, and five putative binding sites have been identified, one close to the active site. In a crevice on the protein surface, five linked five-membered solvent rings form a partial clathrate-like structure. Some other unusual aspects of the structure include seven apparent cis-peptide bonds, four of which are proline, and several internal salt-bridge networks. Deep solvent-filled channels and tunnels extend across the surface of the molecule and pass through the center of the tetramer. Because of these departures from a compact globular shape, the molecule is not well characterized by prior empirical relationships between the mass and surface area of proteins. The 50 or so residues at the amino terminus have a largely extended conformation and mostly lie across the surface of the protein. At the same time, however, segment 13–21 contributes to a subunit interface, and residues 29–33 pass through a "tunnel" formed by a domain interface. Taken together, the overall arrangement provides a structural basis for the phenomenon of *α*-complementation.

Keywords: *α*-complementation; clathrate; glycosidase; operon model

β-Galactosidase (EC 3.2.1.23) from *Escherichia coli* hydrolyses lactose and other *β*-galactosides into monosaccharides. The enzyme is the gene product of the lacZ operon and, as such, has a unique place in the history of molecular biology. Purification of *β*-galactosidase (Wallenfels & Weil, 1972) was first carried out in the 1950s by Cohn and Monod. Later, the regulation and production of *β*-galactosidase by *E. coli* led Jacob and Monod (1961) to their classic model of the operon. Today, *β*-galactosidase is widely used as an indicator molecule in a variety of different assays (Fowler & Zabin, 1983).

The functional form of *β*-galactosidase is a tetramer of four identical subunits (Appel et al., 1965), each consisting of 1,023 amino acid residues (Fowler & Zabin, 1978; Kalnins et al., 1983). The tetramer (M_r = 465,412 Da) contains four catalytic sites that show no cooperativity or allosteric effectors. Taken together, these two observations make one ask (1) why is the protein so big, and (2) why does it need to be a tetramer?

The enzyme has three activities that ultimately result in the complete breakdown of the disaccharide lactose into galactose plus glucose. First, *β*-galactosidase cleaves lactose into galactose plus glucose. Second, the enzyme acts as a transglycosylase, converting lactose into allolactose. Third, it hydrolyzes allolactose into galactose plus glucose. Glu537 is thought to be the catalytic nucleophile, forming a covalent bond with the substrate (Gebler et al., 1992).

In binding substrates, the enzyme is highly specific for the galactose moiety, but is very promiscuous with regard to the remain-

der. This has permitted the development of a series of substrates such as X-Gal (5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside) that incorporate a chromophore that allows the activity of the enzyme to be immediately recognized by a distinct change in color.

The divalent cation Mg^{2+} (which can be substituted by Mn^{2+}) is required for maximal activity, although the exact role that it plays in catalysis has been controversial (Wallenfels & Weil, 1972; Sinnott & Withers, 1978; Sinnott, 1990; Richard et al., 1996). Likewise, the exact role played by Na^+ in facilitating the activity of the enzyme has been unclear. A first step in addressing these issues is to identify the number and locations of the various metal-binding sites.

In addition to the ability of β-galactosidase to cleave chromogenic substrates, it has other attributes that have contributed to its usefulness as a tool in molecular biology. For example, other polypeptide sequences can be fused up to residue 23 of the enzyme without affecting hydrolysis (Fowler & Zabin, 1983). Also certain deletions in the protein that are inactive can be complemented *in trans* by adding appropriate short polypeptides, producing active enzyme (Ullmann et al., 1965). This phenomenon is know as α-complementation.

Using a monoclinic crystal form that included four independent tetramers (Jacobson & Matthews, 1992), it was possible to determine the three-dimensional structure of β-galactosidase (Jacobson et al., 1994). The refinement of this structure to 2.5 Å resolution is briefly reported here. We also describe the crystallization and structure determination of the enzyme in a new crystal form with one tetramer per asymmetric unit. This new structure, which has been refined to 1.7 Å resolution, confirms the fold as determined previously and provides new insights regarding metal binding sites and the phenomenon of α-complementation.

Results

Structure determination and refinement, space group $P2_1$

The initial determination of the structure of β-galactosidase (Jacobson et al., 1994) was in space group $P2_1$ (Fig. 1). As described in Materials and methods, this structure has been refined to 2.5 Å resolution both with and without the imposition of noncrystallographic symmetry (Tables 1, 2). The model includes residues 3–1,023 for each chain with ~93 solvent molecules per monomer. In general, the electron density (Fig. 2A) is of high quality throughout the structure. However, limited regions exhibit weak density. These include the two amino-terminal residues omitted from each chain, a largely solvent-exposed loop including residues 578–583 that appears to be quite mobile, and an extended region of chain between residues 727 and 733 that appears to display multiple conformations. The Luzatti plot (not shown) for the constrained model suggests a root-mean-square deviation coordinate error of <0.35 Å. The Ramachandran plot (also not shown) has 83% of the residues in the most favored regions.

The four tetramers of β-galactosidase that make up the asymmetric unit of these crystals each have 222 point symmetry and are arranged with a high degree of pseudo-symmetry (Fig. 1). The organization of the tetramers may be described as a pair of "octamers" related roughly by a translation of (0.5, 0.0, 0.5). Within each of the two "octamers," the tetramers are related by an approximate twofold rotation about an axis nearly parallel to the crystallographic 2_1 screw axis. This chain-like arrangement may be related to multimeric forms of β-galactosidase (Appel et al., 1965) including the "fiber-like" structures seen by electron microscopy (Karlsson et al., 1964).

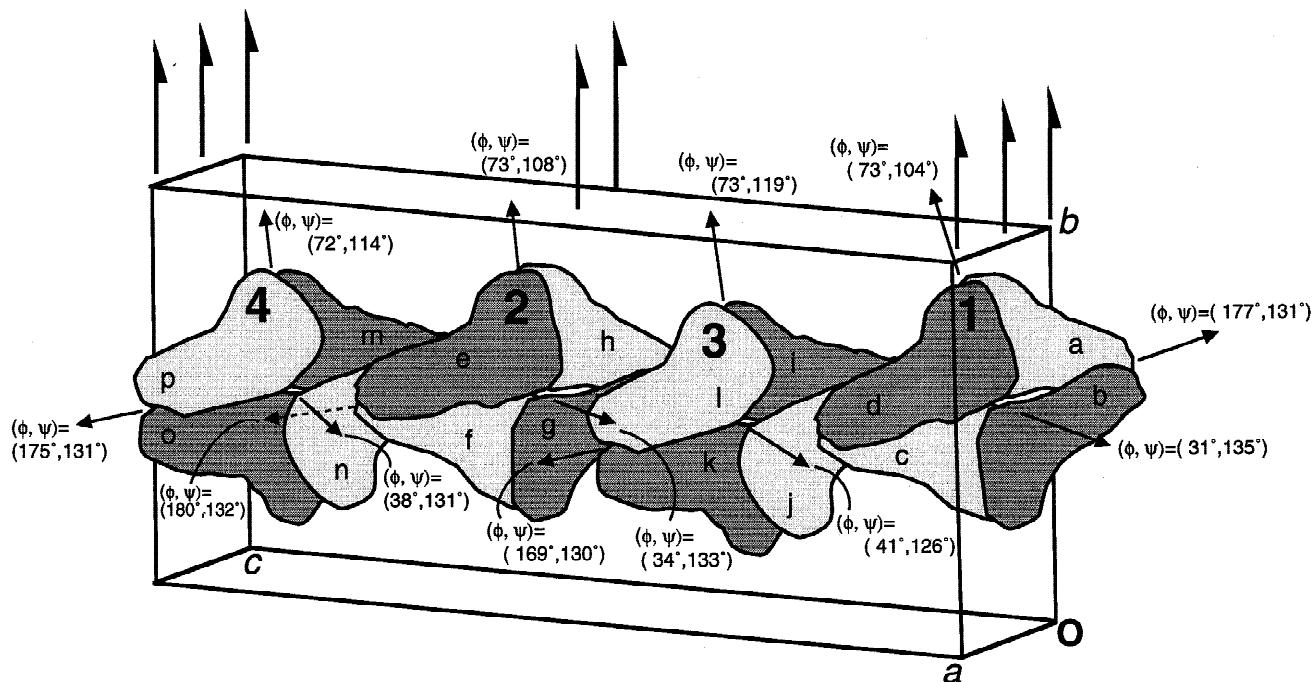**Fig. 1.** The arrangement of the four tetramers of β-galactosidase in the asymmetric unit of the $P2_1$ crystal form. Tetramers 1 and 3 are related to tetramers 2 and 4 by an approximate translation of (0.5, 0, 0.5). Tetramers 1 and 3 (and 2 and 4) are related to each other by a twofold axis approximately parallel to *b*. The various noncrystallographic twofold axes within and between tetramers are shown. Their orientations are specified in polar angles: ψ is the angle of declination from *b* and ϕ is the azimuthal rotation about *b* with *a** corresponding to $\phi = 0°$.

Table 1. *X-ray data collection* [a]

Space group	P2$_1$	P2$_1$2$_1$2$_1$	P2$_1$2$_1$2$_1$
Temperature	RT	RT	~95K
Mode of collection	Photon factory	Daresbury	ALS
Measured reflections	1,321,660	299,596	2,201,152
Unique reflections	559,917	116,158	543,188
R_{merge} (%)	7.0	9.6 (32.1)	6.0 (34.6)
Completeness (%)	73	88.3 (71.0)	98.6 (97.0)
Resolution limit (Å)	2.5	25.0–2.8	30.0–1.7
Cell dimensions			
a (Å)	107.9	153.4	149.6
b (Å)	207.5	173.4	168.4
c (Å)	509.9	204.4	200.7
$β$ (°)	94.7	—	—

[a] The data in space group P2$_1$ are from Jacobson et al. (1994). Numbers in parentheses correspond to the outermost shell of data.

Structure determination and refinement, space group P2$_1$2$_1$2$_1$

Crystallization and determination of the structure of β-galactosidase at 2.8 Å resolution in a new (orthorhombic) crystal form is described in Materials and methods and in Table 1. The use of flash freezing permitted data collection and refinement to high resolution (Table 2). The coordinate error is ~0.15 Å as judged by a Luzatti plot. This model serves as the basis for the detailed description of the structure. It includes residues 13–1,023 for each of the four chains, several Mg^{++} and Na$^+$ ions, 4,424 water molecules, and 112 dimethylsulfoxide (DMSO) molecules. The electron density (Fig. 2B) is of high quality throughout except that

residues 684–690 and 730–735 have very weak density and are likely in multiple conformations. They are both in loop regions that are quite solvent exposed. The first 12 residues at the N-terminal end of the molecule appear to be disordered, perhaps in part because the sequence of the first eight of these is not the same as wild-type (see Materials and methods). This was also verified by sequencing the thrombin digested protein (data not shown). This change is expected to have little effect on the overall structure. Also the identity of the first 23 amino acids has little effect on hydrolytic activity (Fowler & Zabin, 1983). It was verified with a capillary electrophoresis assay (Zeleny et al., 1997) that the modified enzyme has the expected transglycosylase activity (data not shown). Figure 3 shows the Ramachandran plot for the tetramer.

The high resolution refinement in the new crystal form confirms the structure as originally described (Jacobson et al., 1994). Including the constrained refinements, there are 26 models for a β-galactosidase monomer, which differ by either space group, temperature, or refinement protocol (constrained vs. unconstrained). The agreement of C^α positions in the 325 possible monomer–monomer comparisons varies from 0.2 to 0.7 Å. The best agreements (0.2–0.3 Å) are between two monomers in the 1.7 Å model, between the constrained models and their own descendants, and between the two constrained models. The monomers having poorest agreement (0.4–0.7 Å) are from the nonconstrained models in different crystal forms and at different temperatures.

Structure of the monomer

Most of the 1,023 residues that form the β-galactosidase monomer form five well-defined structural domains (Jacobson et al., 1994;

Table 2. *X-ray refinement* [a]

Temperature	Space group P2$_1$		Space group P2$_1$2$_1$2$_1$		
	RT	RT	RT	RT	~95K
Mode of refinement	Constrained (16 monomers)	Unconstrained	Constrained (4 monomers)	Unconstrained	Unconstrained
Resolution (Å)	93–2.5	8.0–2.5	25.0–2.8	25.0–2.8	15.0–1.7
Protein atoms	131,712 (16 × 8,232)	131,168	32,952 (4 × 8,238)	32,952	32,500
Solvent atoms	6,992 (16 × 437)	1,486	1,472 (4 × 368)	853	4,908
$Δ_{bond}$ (Å)	0.018	0.016	0.015	0.016	0.018
$Δ_{angle}$ (deg)	2.7	2.6	2.6	2.8	2.8
$Δ_B$ (Å^2)	5.5		5.2	6.2	7.5
$⟨B⟩_{main}$ (Å^2)	31.2	29.5	31.3	29.2	16.8
$⟨B⟩_{side}$ (Å^2)	37.7	35.5	37.0	34.7	22.2
$⟨B⟩_{solvent}$ (Å^2)	48.0	28.0	47.9	33.9	31.5
R-factor (%)	19.9	17.4	16.7	13.7	15.7
R-free (%)	20.7	—	19.8	27.9	21.1
Ksol	0.98	0.8	0.80	0.77	0.66
Bsol (Å^2)	625	200	356	367	126
B11 (Å^2)	−3.2	—	−3.7	−3.1	−1.6
B22 (Å^2)	5.7	—	3.4	4.1	1.7
B33 (Å^2)	−2.5	—	0.2	−1.0	−0.1
B13 (Å^2)	148.1	—	0	0	0
PDB code	1F49	1BGL/1BGM	1F4A	1F4H	1DPO

[a] $Δ_{bond}$, $Δ_{angle}$, and $Δ_B$ give the average deviations of the bond lengths, bond angles, and B-factors from expected values. $⟨B⟩_{main}$, $⟨B⟩_{side}$, and $⟨B⟩_{solvent}$ give the average thermal factors of the main chain, side chain, and solvent atoms. K_{sol} and B_{sol} are the parameters specifying the bulk solvent model used by TNT (Tronrud, 1997). B11, B22, B33, and B13 are the parameters defining the overall anisotropic scaling of the data (Stout & Jensen, 1989). No special precautions were taken in selecting the reflections used to calculate R-free so the quoted values may be artificially low due to the presence of noncrystallographic symmetry.
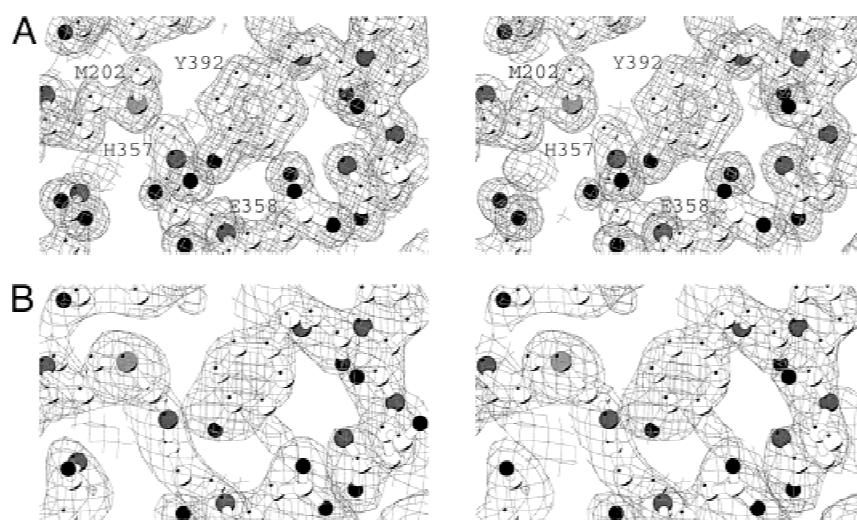
Fig. 2. A: Stereoview of representative electron density in space group P2₁2₁2₁. Coefficients are $2F_o - F_c$ where F_o and F_c are the observed and calculated structure amplitudes. The map is calculated at 1.7 Å resolution and contoured at 1σ. The refined model is superimposed. **B:** Same view as above for the refined structure in space group P2₁. The resolution is 2.5 Å. Figure prepared with MOLSCRIPT (Kraulis, 1991).
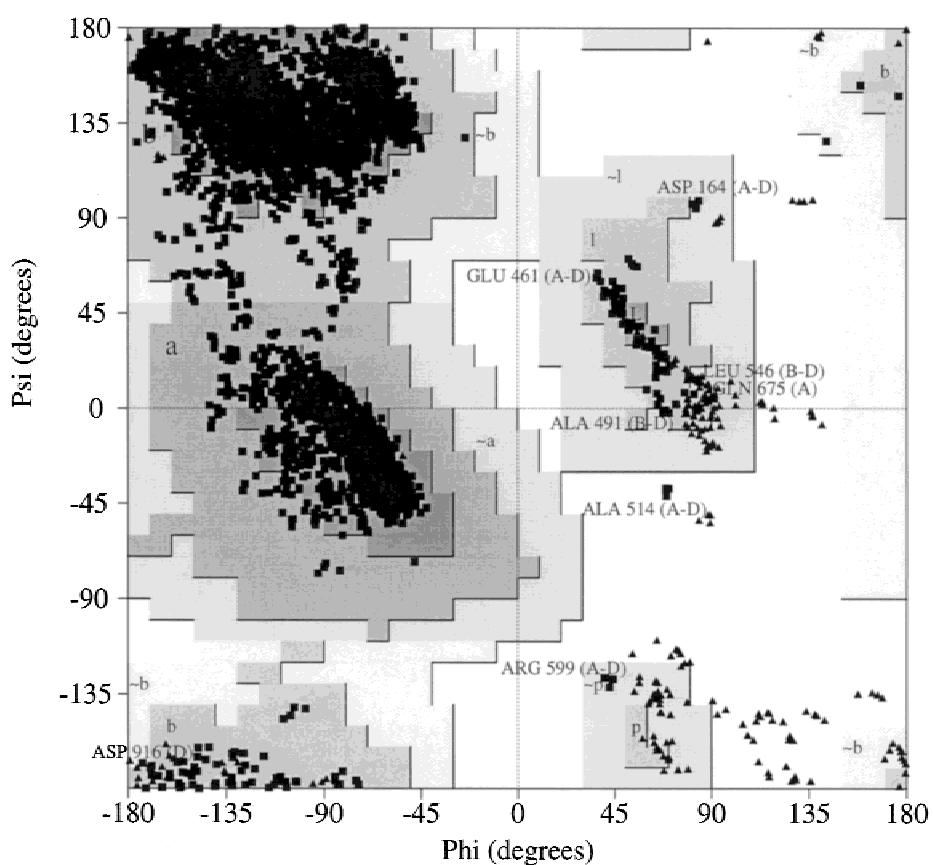

Fig. 3. Ramachandran diagram for the nonconstrained model of the β-galactosidase tetramer in space group P2₁2₁2₁. Glycine residues are indicated by triangles and nonglycines by squares. According to the criteria of Laskowski et al. (1993) 87.7% of the residues are in most favored regions, 11.6% in additionally allowed regions, 0.6% in generously allowed regions, and 0.1% in disallowed regions. Residues that are putative outliers are labeled; the letters A–D identify the four monomers. Glu461 and Arg599 are active site residues. Ala514 is at the apex of a tight turn that packs against the donated loop in the activating interface and Ala491 is in the (missing) helix 5 of the $(\alpha/\beta)_8$ barrel (Fig. 6C). There are no distinctive structural elements associated with residues D164, L546, or D916.

Juers et al., 1999) (Fig. 4). These include one jelly-roll type barrel (Domain 1, ~170 residues) two fibronectin type III-like barrels (Domains 2, 4, ~110 residues each), a large 19-stranded β-sandwich that exhibits a unique topology (Domain 5, ~300 residues), and the central TIM barrel (Domain 3, ~300 residues). Approximately the first 50 residues of the polypeptide chain are in a rather extended conformation and are not obviously categorized as being associated with any of the five well-defined domains. This portion of the chain makes contacts with the first, second, and third domains from the same chain.

As with all other known enzymes that contain a TIM barrel, the active site of β-galactosidase is located at the C-terminal end of the central core of this domain. For β-galactosidase the active site forms a deep pit that intrudes well into the central core of the TIM barrel. The active sites also includes portions of loops from the first, second, and fifth domains of the monomers.

Structure of the tetramer

The tetramer has 222-point symmetry (Fig. 5). It is roughly ellipsoidal, with dimensions $175 \times 135 \times 90$ Å along the twofold axes. There is a continuous system of channels running along the surface (Fig. 5B) and within the tetramer (Fig. 5C,D). These channels appear to be accessible to bulk solvent and vary in width from 5–20 Å. The four active sites are located at the bottom of such surface channels (Fig. 5B).

Subunit interfaces and formation of the active site

There are two principal subunit interfaces (Fig. 6). The "long" interface buries about 4,000 Å^2 (48% polar atoms) and the "activating" interface buries 4,600 Å^2 (45% polar). There is also a third interface in the tetramer which is much smaller, burying 230 Å^2 (75% polar).

The long interface (Fig. 6A) is formed from Domains 3, 4, and 5 and has two separate regions. The first region is a Domain 3–Domain 3 contact and accounts for about 40% of the interface.

This region includes two buried arginine residues (Arg561 from each domain). Each guanidinium group is surrounded by four backbone carbonyls (one from the neighboring subunit) and two solvent molecules, which bridge the interface. The second region of contact includes parts of Domains 4 and 5.

The activating interface (Fig. 6B) is more contiguous than the long interface and involves mostly Domain 2, Domain 3, and the complementation peptide. It is "S" shaped, in contrast with the long interface, which is fairly planar. About half is formed by two equivalent interactions between Domain 3 and a loop that includes residues 272–288 (Fig. 6C). This loop, which is donated by Domain 2 of one subunit, extends across the interface and completes the active site within Domain 3 of the neighboring subunit. The remaining half of the activating interface is formed by interactions involving Domain 3 and the complementation peptide. Interactions involving the complementation peptide are more polar (~65% vs. ~45%) than the bulk of the interfacial regions. On the other hand, the Domain 3–Domain 3 interaction, which forms a four-helix bundle, is considerably less polar (~30% vs. ~45%).

Solvent structure

In the orthorhombic model, there are ~4,900 solvent atoms, including water, DMSO, Na$^+$, and Mg^{++}. This may seem excessive but is equivalent to about one per residue. About 70% of the solvent atoms are in equivalent locations in the four monomers.

About 80% of the presumed water molecules are within 3.5 Å of polar protein atoms suggesting that they are "first shell" (i.e., interacting directly with the protein). Fifty-five percent of the hydrogen bonds made by protein are to ordered water, and 21% of these are to "buried" water molecules.

The overwhelming majority of ordered solvent molecules interact with other ordered solvent to form rings, clusters, and networks. Ring sizes vary between 4 and 6, with 5 being most common. There are several linked rings, and in the vicinity of Arg721, there are five linked five-membered water rings forming part of a clathrate-like structure. The outside of this "clathrate" forms polar
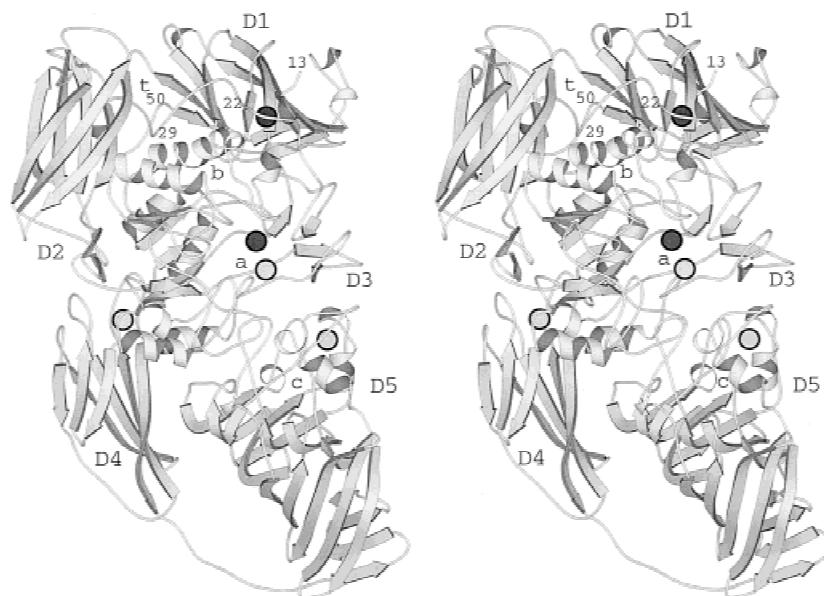

Fig. 4. Stereo drawing illustrating the fold of the monomer. Three of the solvent molecules identified as Na$^+$ ions are shown as light-gray spheres. The magnesium ion identified at the active site (labeled *a*) is drawn as a dark sphere as is another tightly-bound ion, also presumed to be Mg^{2+} (see text). Additional putative Na$^+$ and Mg^{2+} binding sites are not shown. D1–D5 identify the five domains. The amino-terminus, as seen in the electron density map, starts at residue 13, which is labeled. This also corresponds to the α-complementation peptide that extends roughly to residue 50. The tunnel through which this peptide is threaded is labeled *t*. The pair of helices that form half of the four-helix bundle at the activating interface are labeled *b* while the polar core of domain D5 is labeled *c*. Figure prepared with MOLSCRIPT (Kraulis, 1991).
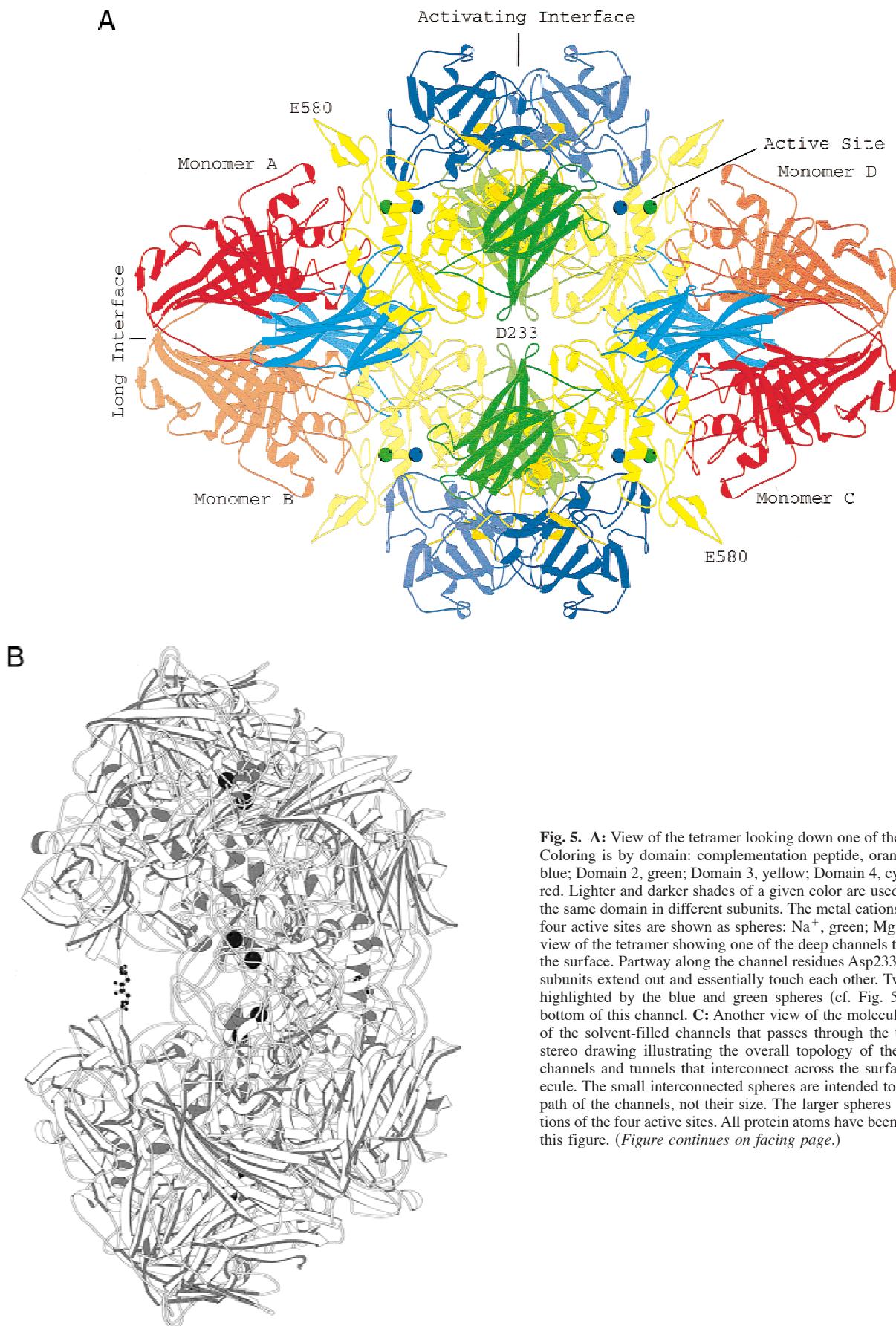
Fig. 5. **A:** View of the tetramer looking down one of the twofold axes. Coloring is by domain: complementation peptide, orange; Domain 1, blue; Domain 2, green; Domain 3, yellow; Domain 4, cyan; Domain 5, red. Lighter and darker shades of a given color are used to distinguish the same domain in different subunits. The metal cations in each of the four active sites are shown as spheres: Na^+, green; Mg^{++}, blue. **B:** A view of the tetramer showing one of the deep channels that runs across the surface. Partway along the channel residues Asp233 from different subunits extend out and essentially touch each other. Two active sites, highlighted by the blue and green spheres (cf. Fig. 5A), are at the bottom of this channel. **C:** Another view of the molecule showing one of the solvent-filled channels that passes through the tetramer. **D:** A stereo drawing illustrating the overall topology of the solvent-filled channels and tunnels that interconnect across the surface of the molecule. The small interconnected spheres are intended to show only the path of the channels, not their size. The larger spheres show the locations of the four active sites. All protein atoms have been removed from this figure. (*Figure continues on facing page.*)

contacts to solvent and to Arg721 (Fig. 7). Inside the partial clathrate is electron density of uncertain origin. It is poorly modeled by one or more water molecules and might, for example, be a disordered molecule of dimethylsulfoxide (DMSO), or some adduct involving His878.

There are 112 presumed DMSO molecules in the model, reflecting the high concentration (30% v/v) used for low-temperature data collection. They bind in 34 distinct sites—most are in pockets and crevices and none are in interior cavities. Although the crevices are topologically on the surface of the protein, some are quite deep. The DMSO molecules generally bind with the sulfonyl groups making polar contacts.

Cis-peptide bonds

There are seven apparent cis-peptide bonds in each monomer, at Asp164, His391, and Trp568, as well as four proline residues (87,

112, 422, and 902). All of these residues occur within domains at turns and other breaks in secondary structure. His391 and Trp568 are at the active site while Asp165 is about 15 Å away forming a tight turn. Pro112 and Pro422 participate in the activating interface, Pro112 interacts with the complementation peptide and Pro422 with the donated loop. Pro87 and 902 are in relatively solvent-exposed regions of Domains 1 and 5.

Polar "core" of Domain 5

Each domain has a well-defined hydrophobic core. Within some domains, however, there are also substantial polar networks and salt bridges. This is especially true of Domain 3, the TIM barrel, and Domain 5, the β-supersandwich. Domain 3 has a salt-bridge network through the outer core (Arg356–Asp375–Arg611) connecting strand 1 to a helix and a loop. This domain also includes a four-residue charged network in the active site that includes
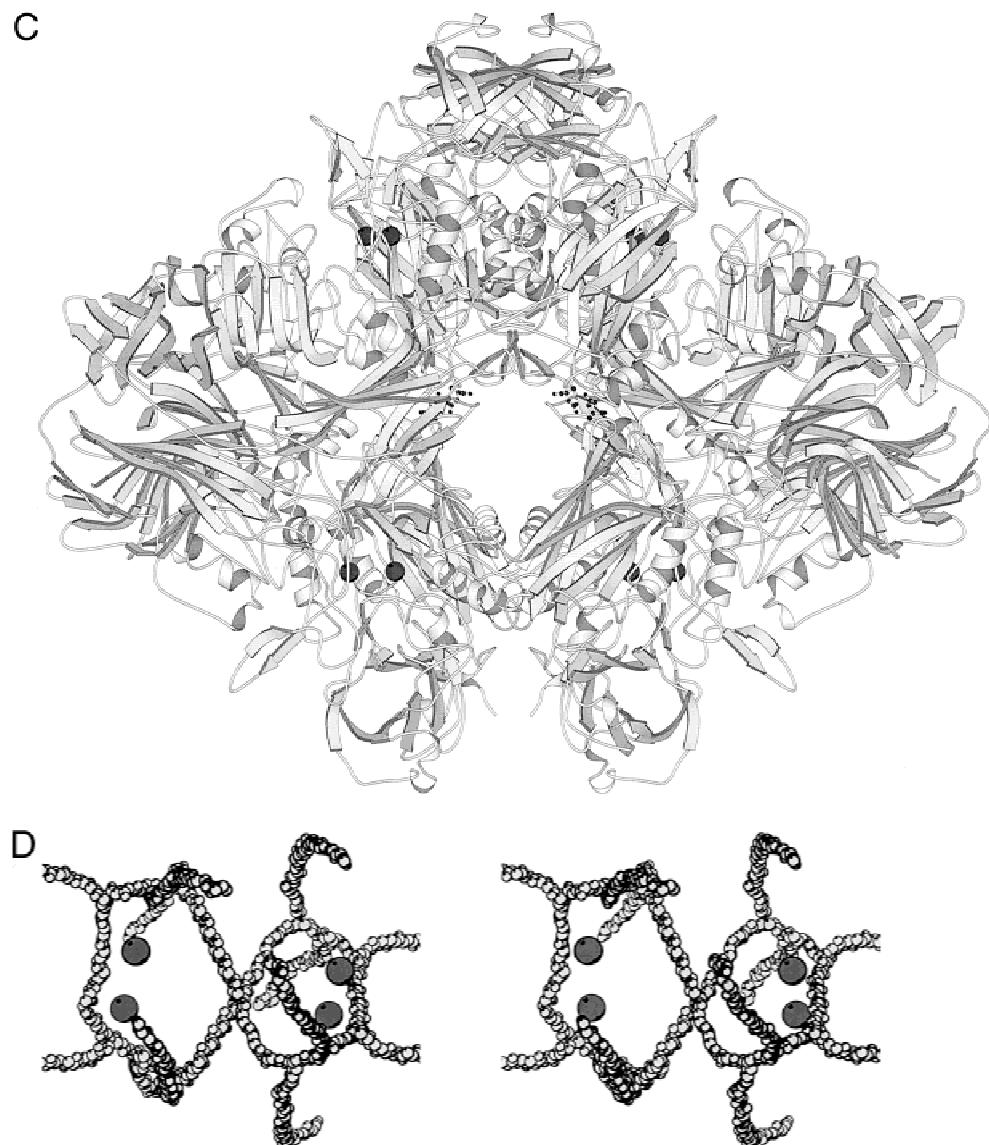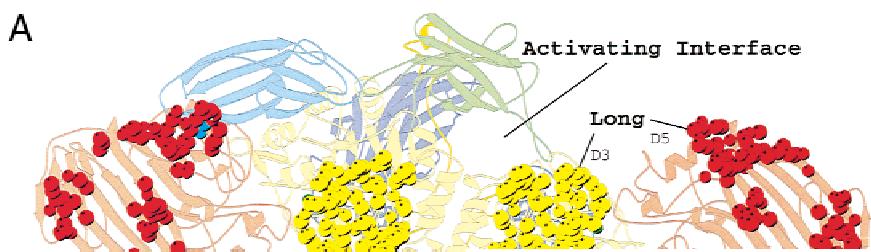

Fig. 5. *Continued.*

A

Activating Interface

Long

D3   D5

His391, Asp412, Arg388, and Glu537. More striking is a buried charged network within Domain 5. In this case, it appears that the domain has folded back on itself so that "outside" has become "inside." The network involves six residues (Arg786, Asp792, Arg881, Glu934, Asp987, and His990) whose side chains are completely buried within the domain, occupying about 650 Å (Fig. 8). Most of these residues are conserved in enzymes homologous to β-galactosidase (e.g., see Fig. 3 of Jacobson et al., 1994).
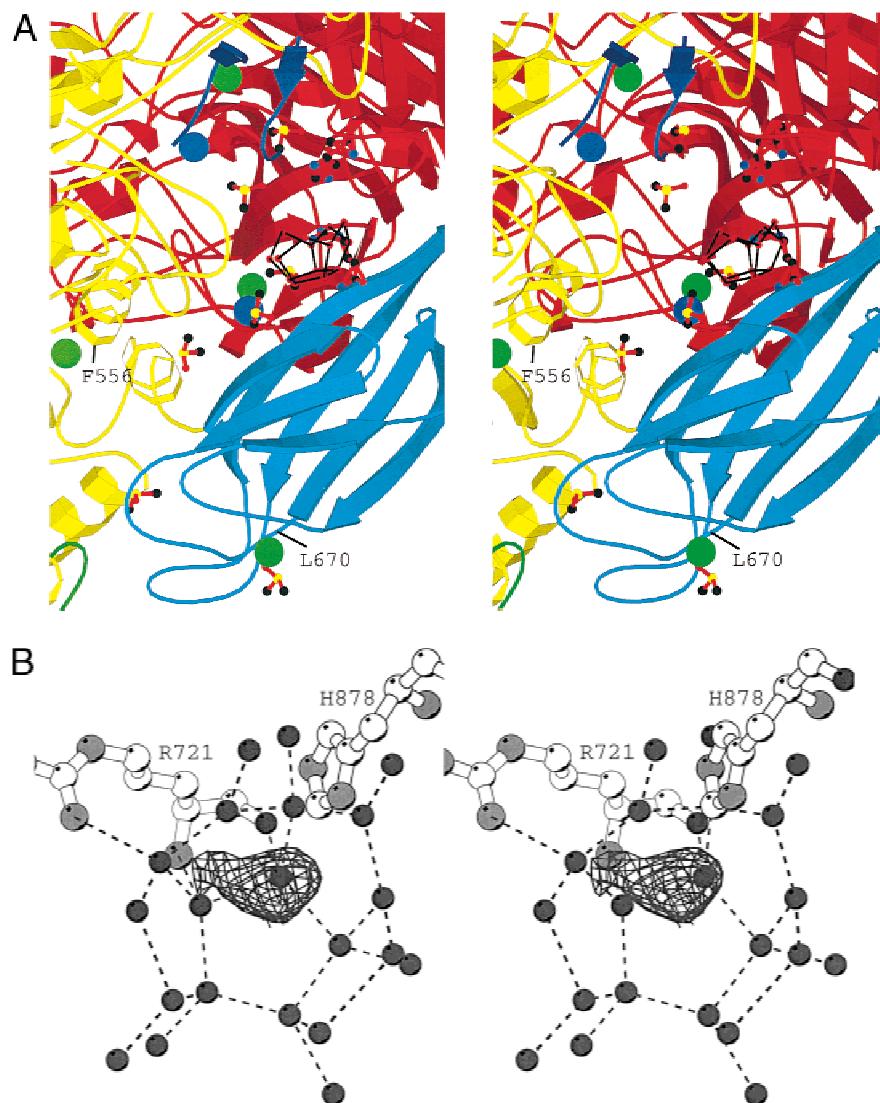
Fig. 7. A: Stereo drawing showing the general location of the partial clathrate-like structure in a surface crevice at the long interface. Coloring is as follows: Domain 4 (monomer A), cyan; Domain 5 (monomer B), red; Domain 3 (monomer B), yellow. The blue sphere identifies a Mg^{++} ion. The green spheres indicate either Na$^+$ or K$^+$ binding sites. The site near L670 has been identified as Na$^+$ and that near the clathrate as K$^+$ while the remaining two are uncertain. The active site is in the upper left, marked by its Mg^{++} and Na$^+$/K$^+$ sites. Also shown are several DMSO molecules. **B:** Close-up stereoview of the partial clathrate-like structure. Water molecules and protein oxygen atoms are shown as black spheres, nitrogen atoms light gray, and carbon atoms as open circles. Presumed hydrogen bonds are shown as broken lines. The electron density is seen in a "residual" map with coefficients $(F_o - F_c)$ where the structure factors F_c and phases correspond to the final refined model. The map is contoured at $+3.0\sigma$ where σ is the root-mean-square density throughout the unit cell. Figure prepared with BOBSCRIPT (Esnouf, 1999).

Discussion

Because β-galactosidase is a relatively large protein, and also because oligomerization is related to activity, we briefly describe the interactions that occur at different levels of association, i.e., between domains, monomers, and tetramers.

Overall, the tetramer has about 135,000 Å^2 of accessible surface area and buries about 18,000 Å^2 at the subunit interfaces. Although there are no structures reported of other tetramers of comparable size, a study including smaller tetramers (up to 230 kDa) suggested that the mass M in Daltons, and the accessible surface area A_s in Å^2 for oligomeric proteins follow the relationship $A_s = 5.3 \, M^{0.76}$ (Miller

et al., 1987). β-Galactosidase, however, shows a 27% discrepancy (135,000 Å^2 vs. 106,000 Å^2). The same authors propose another relationship for monomeric proteins, $A_s = 6.3 \, M^{0.73}$. They suggest that this should be applicable to subunits of oligomeric proteins that involve only a small fraction of their surface area in subunit–subunit contacts. Even though the β-galactosidase subunits can be considered to be in this category, they show a 21% discrepancy from the monomer power law. When considered separately, however, the five individual domains follow the monomer power law better, with discrepancies of 0, 12, 6, 9, and 4%, respectively.

The discrepancies with the scaling relationships suggest that β-galactosidase exposes more surface area to solvent than ex-
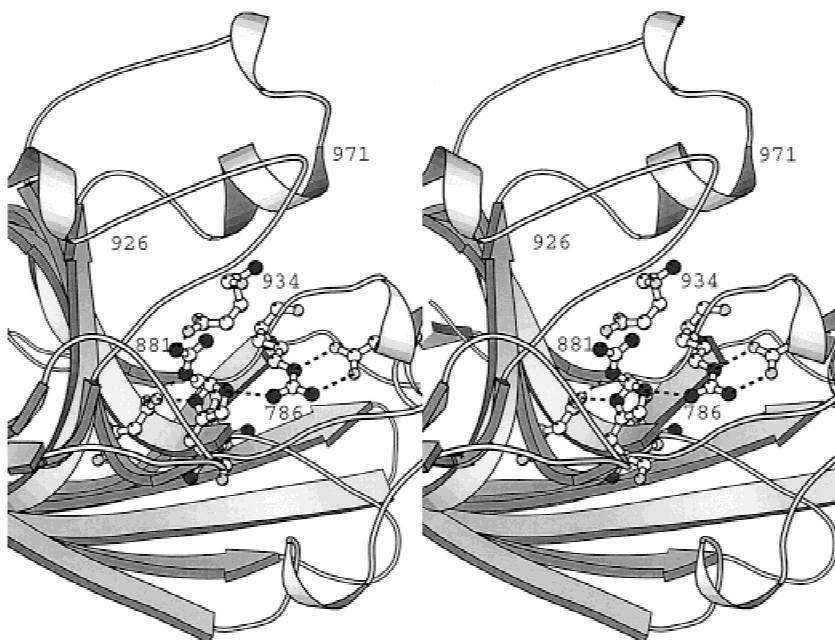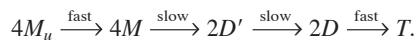
Fig. 8. Stereoview of the polar network within Domain 5 that includes Asp792, Glu934, Asp987, Arg786, Arg881, and His990. This network also interacts with other polar protein atoms and two buried water molecules. Figure 4 shows the location of this network within the monomer. Figure prepared with MOLSCRIPT (Kraulis, 1991).

pected in comparison with other oligomeric proteins. Because the domains follow the monomer scaling law better than the monomer, it suggests that the discrepancy is due to the domain associations rather than the domains themselves. Indeed, as shown in Figures 5B, 5C, and 5D there are deep solvent-filled channels that extend across the surface of the molecule as well as channels that pass through the middle of the tetramer.

There is a decrease in buried hydrophobic area (from 61 to 40% of overall surface area) as one moves up the hierarchy from interactions within domains, to those between domains and ultimately to those between tetramers. At the same time, the frequency of hydrogen bonds decreases. In other words, the higher in the hierarchy the fewer hydrogen bonds per unit area of buried polar interface. Generally, the interfaces at the crystal contacts are quite different in character from the other interfaces. They are significantly more polar yet have fewer hydrogen bonds per unit area. In addition, the area buried by bridging waters is greater at the crystal contacts than the other interfaces. This illustrates the nonspecific nature of these contacts.

It has recently been suggested that β-galactosidase folds according to the following mechanism (Nichtl et al., 1998):

$$4M_u \xrightarrow{\text{fast}} 4M \xrightarrow{\text{slow}} 2D' \xrightarrow{\text{slow}} 2D \xrightarrow{\text{fast}} T.$$

According to this scheme, the unfolded monomer chains (M_u) first give folded monomers (M). There is then a slow bimolecular event to form dimers (D). These then undergo a slow first order event to form dimers (D) competent for fast association to tetramers (T). Overall this is consistent with the hierarchical stability and specificity described above. In particular, a much larger surface area is buried between domains than between monomers, suggesting that the monomers could fold independently prior to forming dimers or tetramers. The buried surface of the former is also more hydrophobic in character suggesting greater stability. Which dimer in-

terface is formed in the $4M \rightarrow 2D$ step and which must wait for the $2D \rightarrow 2D$ event has not been established. The long interface is relatively flat and unstructured. In contrast, the activating interface is S-shaped and its formation involves interdigitation of the donated loop and proper positioning of the complementation peptide. This suggests that $2D$ is a dimer formed by association of two monomers at the long interface. The rate-limiting step in the $2D \rightarrow 2D$ association could correspond to the proper positioning of the complementation peptide or the donated loop. This would be consistent with the kinetics of α-complementation, which show a first order event of similar rate following binding of the complementation peptide (Zabin, 1982; Nichtl et al., 1998)

α-Complementation and the role of the amino-terminus

β-Galactosidase is widely used because of its easy colorimetric assay and because hybrids with other polypeptides can be made extending up to at least residue 26 and still result in active enzyme (Müller-Hill & Kania, 1974; Ullmann, 1992). Also, deletions of residues 23–31 or 11–41 result in inactive dimers (called α-acceptors) that can be complemented by certain peptides (α-donors) to reconstitute the active tetramer (Ullmann et al., 1967; Ullmann, 1992). Two common α-donors encompass residues 3–41 or 3–92. This phenomenon of α-complementation is the basis for the common blue/white screening used in cloning and other procedures.

Figure 9 is a sketch illustrating in a highly simplified fashion the parts of the β-galactosidase structure that appear to be important for hybrids and for α-complementation. The 50 or so residues at the amino terminus have an irregular, largely extended conformation and mostly lie across the surface of the protein. At the same time, however, residues 13 and 15 contribute to the activating interface while segment 29–33 passes through a "tunnel" formed by a domain–domain interface. Residues 22–31 are located fairly close to the activating interface but most of the contacts made by these residues are with Domain 1 and parts of the four-helix bun-
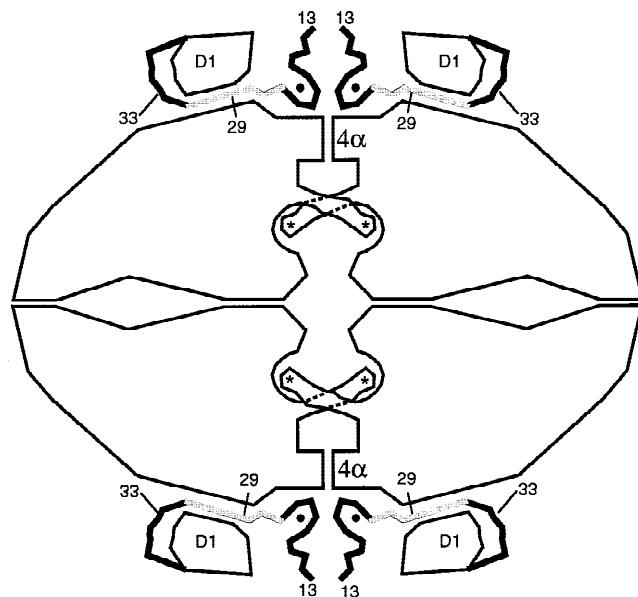
Fig. 9. Sketch summarizing key features of the β-galactosidase tetramer. At the amino terminus, residues 1–12 are not seen in the electron density map due to presumed disorder. Residues 13–50 (shown as thick lines) pass through a tunnel between the first domain (labeled D1) and the rest of the protein (see also Fig. 4). The region shaded gray (residues 23–31) is deleted in one of the α-donors (see text). A magnesium ion (shown as a small solid circle) bridges between the complementation peptide and the rest of the protein. The four active sites are labeled with asterisks. The activation interface runs vertically through the middle of the figure. An important part of this interface is a bundle of four α-helices in the region labeled 4α. When the activation interface is formed the four equivalent loops that include residues 272–288 extends across the interface to complete the active sites within the four recipient subunits.

dle, both of which are within the same subunit (Figs. 4, 6C, 9). There is a presumed magnesium ion that is coordinated by Asp15, Asn18, Val21, Gln163, and Asp193. This ion therefore bridges between the complementation peptide and the rest of the protein (Fig. 6C).

Studies of hybrids by Fowler and Zabin (1983) showed that variants of β-galactosidase with the first 26 residues replaced were prone to dissociation to dimers, whereas substitution of the first 23 residues did not show this behavior. Since deletion of residues 23–31 results in inactive dimers, residues 27–31 are presumably critical for tetramer formation. This is consistent with the structure illustrated in Figure 9. The lesser importance of residues 1–23 is consistent with the observations that part of this region (residues 1–12) are disordered while the remainder (residues 13–23) contribute only weakly to interactions across the activating interface (mostly by reciprocal salt bridges between Arg13 and Asp15). Residues 22–31 help stabilize the four-helix bundle that is a major part of the interface. The particular importance of residues 27–31 is explained by the observation that residues 29–33 pass through the tunnel, stabilizing the junction of domains 1–3 (Fig. 4). It may also be noted that mutant M15 β-galactosidase, which is missing residues 11–41, is an inactive dimer, and is much more labile to proteases than the native enzyme, particularly with regard to the Arg431–Trp432 peptide bond (Edwards et al., 1988). This can be rationalized in terms of protection of the peptide bond both with intrasubunit interactions (involving the complementation peptide) and intersubunit interactions (involving the activating interface).

As illustrated in Figure 9, the formation of each active site requires that each half of the activating interface be present. Dissociation of the β-galactosidase tetramer into dimers removes the Glu281 loop from the remainder of the active site. Thus, dissociation of the tetramer to dimers is synonymous with deactivation.

Typically, α-complementation has employed the α-acceptors M15 or M112, which have deletions of residues 11–41 and 23–31, respectively, and the α-donors 3–41 and 3–92 (3–92 is usually called CNBr$_2$). Both acceptors can be complemented by either donor. Complemented β-galactosidase has catalytic activity essentially identical with the native enzyme but is more heat and urea labile. The region of overlap between the α-donors and α-acceptors includes the segment of the β-galactosidase structure in which the polypeptide chain is threaded through the "tunnel" (Figs. 4, 9). This helps to rationalize some of the nuances of the complementation reaction. The α-acceptors, which result in the substitution of residues 29–31 with nonnative amino acids, would make it less favorable for this segment to occupy the tunnel region. The donors, in contrast, include the appropriate amino acid sequence to occupy the tunnel and to substitute the interactions present in the wild-type protein. The long α-donor, including residues 18–92, presumably occupies the tunnel and displaces from the acceptor not only residues 29–33 but also residues extending to 60–90 within domain D1 (Fig. 9), making them susceptible to proteases and available for binding by antibodies (Zabin, 1982). The shorter donor, spanning residues 3–41, presumably also occupies the tunnel, but does not displace residues in the vicinity of 60–90.

Although the N-terminal 23 residues appear to be relatively unimportant for tetramer formation in hybrids, they can have an effect on α-complementation. For example, the mutation E17Y in the donor 3–92 reduces α-complementation and also decreases the stability of complemented enzyme. This residue makes no intersubunit contacts, but does participate with Arg14 and the backbone amide of Val114 in a small intrasubunit polar network. Likewise, Trp16, which is fairly well conserved in homologous enzymes, does not participate in subunit contacts but is largely buried within its own subunit. Also, deletion of residues 3–17 of the 3–92 donor eliminates complementation activity, suggesting that some of these residues are critical. In particular, Asp15, Asn18, and Val21 presumably contribute to the binding of the complementation peptide via their coordination of the Mg^{2+} ion that bridges to the rest of the protein (Figs. 6C, 9). This is also consistent with the observation that Mg^{++} stabilizes the complemented protein (Gallagher & Huber, 1999).

In a donor-acceptor, complex binding energy is required to offset the entropic cost of keeping the two polypeptide chains in contact. This is not necessary either for the native enzyme or for a covalently-linked chimera. Thus, the supplemental interactions provided by residues 3–17 may be dispensable in the latter two cases but required for α-complementation. We assume that the complementation peptide binds within the tunnel shown in Figure 9. This unusual arrangement may confer two advantages. First, the intimate mode of association presumably enhances the thermodynamic binding affinity. Second, the enclosure of the peptide within the tunnel presumably increases the activation energy for its removal, thus providing kinetic stability as well.

Metal binding sites

Both Mg^{2+} and Na$^+$ are required for maximal activity of β-galactosidase (Wallenfels & Weil, 1972). Putative sodium-

binding sites were identified by collecting X-ray data for crystals soaked in both potassium and rubidium. This analysis identified five such sites. It also suggested there is at least one site that binds potassium and rubidium, but not sodium. The electron density map for the sodium ion that binds in the vicinity of the active site is shown in Figure 10. Because of its close proximity to the active site (Fig. 4) it is highly likely that its removal would perturb this region and reduce activity. The other presumed sodium ions bind on the surface liganded by backbone carbonyls, water molecules, and in one case, a DMSO oxygen.

A presumed magnesium ion at the active site was identified, both in the monoclinic and orthorhombic structures using X-ray data for crystals soaked with solutions containing EDTA (data not shown). Various inhibitors and substrate analogs bind close to this ion (not shown), consistent with it being required for catalysis. The identities of ions at several other presumed metal sites are less clear, although each has been modeled as magnesium based on its octahedral geometry and the nature of its ligands. Most have five or six water ligands, and three occur at crystal contacts mediating intermolecular interactions with no direct contacts to protein. An apparent ion-binding site in Domain 1 (Fig. 4) was not affected by the presence of potassium, rubidium, or EDTA, suggesting that the putative ligand at this site, which has also been modeled as a magnesium, is bound both tightly and specifically.

Consistency with protease-sensitivity and insertion mutagenesis

Limited exposure of the enzyme to chymotrypsin results in cleavage of the Trp585–Ser586 and Phe601–Cys602 peptide bonds, both being protected by Na^+ or K^+ (Edwards et al., 1988). This is consistent with the former bond being solvent-exposed and mobile and the latter bond being close to the Na^+ binding site (Fig. 10).

Likewise, elastase cleaves the Ala732–Ala733 peptide bond (Edwards et al., 1988) which is also highly mobile in both crystal structures.

The observed structure of β-galactosidase is also consistent with insertion mutants constructed by Breul et al. (1991). Those insertions that have little effect on activity are located in solvent-exposed loops while those that reduce activity tend to occur in the more rigid parts of the protein.

Materials and methods

Expression and purification

Purification of β-galactosidase for the initial structure determination in space group P2$_1$ was as described (Jacobson & Matthews, 1992).

Protein used for crystallizing the P2$_1$2$_1$2$_1$ crystal form was initially prepared by growing *E. coli* strains BL21(DE3) and B834(DE3) and purifying the endogenous β-galactosidase. Cells were resuspended in 25 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM DTT, 10% glycerol (TEGD), and lysed. A 40% ammonium sulfate precipitation (4°C) was performed on the lysate and the pellet redissolved in TEGD. The protein was applied to a Q-Sepharose HP column (Pharmacia, Uppsala, Sweden), equilibrated in TEGD and eluted with a 0–0.5 M NaCl gradient. The β-galactosidase was concentrated and run on a Superdex 200 sizing column (Pharmacia), equilibrated to TEGD + 150 mM NaCl. The purity was slightly improved with a Mono Q HR 5/5 column (Pharmacia), again with TEGD and NaCl. The room temperature data were collected using crystals from this protein.

Subsequently, higher levels of β-galactosidase were obtained by overexpressing the protein using Induction Control B for the pET system from Novagen (Madison, Wisconsin). This is the pET 15b
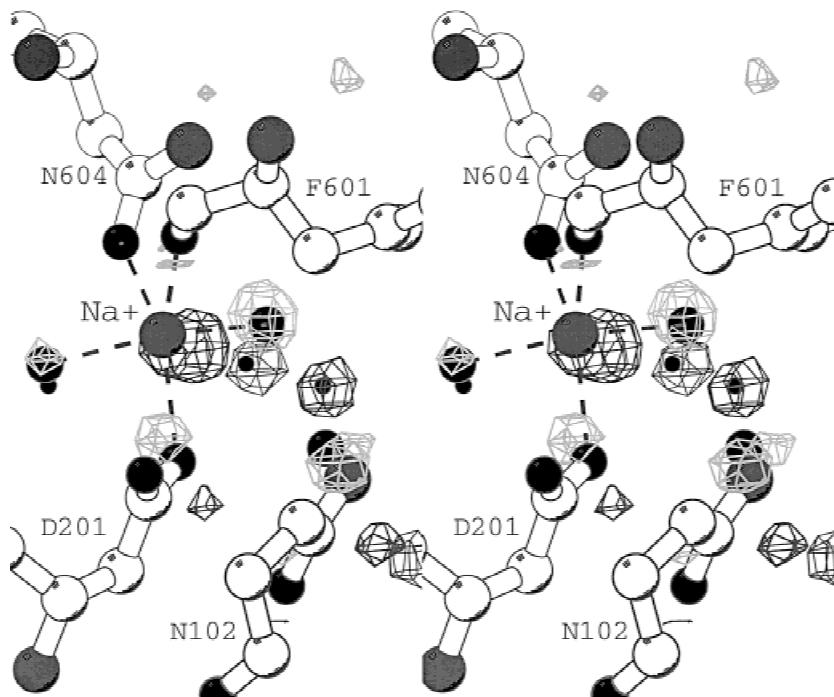

Fig. 10. Stereoview of sodium binding site in the active site. The large sphere shows the sodium ion and the smaller spheres show the locations of refined water molecules in the potassium soak. Electron density is calculated from a map with coefficients $F_o(K^+) - F_o$, where the $F_o(K^+)$ are the structure amplitudes for the crystals soaked in K^+ and F_o are the structure amplitudes of the native (i.e., Na^+-containing) crystals. The map is contoured at $\pm 6\sigma$ (black = +, gray = −). The density suggests the sodium has been replaced by potassium and in response the cation ligands have slightly expanded. The solvent structure has also slightly reorganized. Figure prepared with MOLSCRIPT (Kraulis, 1991).

plasmid with a lacZ insert and includes an N-terminal six-histidine tag. It also has the N-terminal sequence Gly-Ser-His-Met-Leu-Glu-Asp-Pro rather than the wild-type sequence of Thr-Met-Ile-Thr-Asp-Ser-Leu-Ala. This protein was used for the low-temperature data collection. Cells of *E. coli* strain BL21 (DE3) with this plasmid were grown either in shaker flasks or a fermenter at 37 °C and induced with 1 mM IPTG for 3 h. The cells were spun down and resuspended in 20 mM Tris, pH 7.9, 500 mM NaCl, 5 mM imidazole, and 2 mM β-mercaptoethanol. After sonication for 5 min, the cell lysate was loaded on a nickel column (Qiagen, Hilden, Germany). Usually sonication was repeated for better yield. The column was washed with the loading buffer and the protein was eluted with a 5–200 mM imidazole gradient (500 mL total volume). After dialyzing vs. 2 × 4 L 25 mM Tris, pH 7.9, 125 mM NaCl, 2.5 mM CaCl$_2$, and 2 mM β-mercaptoethanol, thrombin (Pharmacia) was added to cleave the histidine tag. This was allowed to incubate for 2–3 days at room temperature, and the cleavage was monitored via native polyacrylamide gels (Phastsystem, Amersham Pharmacia Biotech, Piscataway, New Jersey) or anion exchange chromatography (Biocad Perceptive Biosystems, Framingham, Massachusetts). The digested protein was further purified with anion exchange chromatography. The best results were obtained with a PI column (Perceptive Biosystems, Framingham, Massachusetts) at pH 7.0 with a 0–1 M NaCl gradient in a Bis-Tris/Tris buffer. Pooled fractions from the anion exchange step were quite pure, but typically contained higher order oligomers as judged from native polyacrylamide gel electrophoresis (Phastsystem). Therefore, the protein was concentrated by ammonium sulfate precipitation and run on a sizing column (Sephacryl S-200 at 0.1 mL/min) after resuspending to ~20 mg/mL in 100 mM Bis-Tris, pH 6.5, 200 mM MgCl$_2$, 1 mM DTT, and 5 mM NaCl. Fractions from the sizing run corresponding to the pure tetramer were concentrated to ~10 mg/mL for crystallization with centriprep concentrators.

Crystals

Monoclinic crystals of *E. coli* β-galactosidase (Table 1) were obtained as previously described (Jacobson & Matthews, 1992). A number of other crystal forms were identified, some by Dale Wigley in York and some in Eugene. Of these, the one most promising crystallized as pyramids. The best crystals were obtained by seeding and using a mother liquor of 10% PEG 8K, 100 mM Bis-Tris, pH 6.5, 200 mM MgCl$_2$, 100 mM NaCl, and 10 mM DTT. Seed solutions were created by diluting a drop of initial, small crystals into 0.1–10 mL mother liquor. Drops were then set up using 5 μL of protein solution and 5 μL of seed solution. Pyramids (and occasionally plates) appeared in 1–3 days, and growth appeared to be complete in 2–3 weeks. The largest crystals were approximately 0.8 × 0.7 × 0.6 mm. Macroseeding was also successful and produced some of the largest crystals. Temperature was an important factor, 15° usually giving the best yield of large crystals. Room temperature usually produced fewer crystals, while 4 °C often produced poorly formed ones.

Although the crystals diffracted well, they decayed significantly after a few hours in the beam, suggesting cryocrystallography would be required for high resolution data collection. For freezing, several solutions were tried. Glycerol, MPD, and ethylene glycol gave poor diffraction, while PEG 200, PEG 400, PEG 550, glucose, and sucrose were more promising. Dimethylsulfoxide was the best cryosolvent (70% mother liquor, 30% DMSO). Crystals were equilibrated by adding 25 μL aliquots of DMSO to crystals sitting in 0.7 mL mother liquor over the course of 6 h or more. Crystals could then be flash frozen in a cold stream with only small effects on the mosaicity and diffraction.

Model building and refinement, space group P2$_1$

An initial model of one monomer was built into the 3.5 Å resolution electron density map (Jacobson et al., 1994) with fragments from a library of well-refined protein structures using the automated routines in O (Jones et al., 1991). This model was then used to generate the 16 copies present in the P2$_1$ cell. Assuming an overall average Wilson *B*-value of 27.5 Å^2 the initial *R*-factor was 38% for data at 4 Å resolution.

All refinement was done with the TNT package of programs (Tronrud et al., 1987; Tronrud, 1996, 1997). Constrained NCS refinement was carried out on the atomic positions using data from 8.0 to 3.5 Å, 3.0 Å, and finally 2.5 Å resolution. At this point, the averaged maps were used to locate the missing regions of the model and refinement continued. About 100 water molecules and two magnesium ions were located in the averaged density and built into the prototype molecule. At this point, releasing the NCS constraints and refining both positions and *B*-factors resulted in the model described in Table 2.

From a practical standpoint, it was desirable to also have a model with 16-fold constrained NCS, so refinement was continued with constrained NCS using all data to 2.5 Å. Several rounds of positional refinement, model building, and solvent addition resulted in an *R*-factor of 27.3%. Subsequently, coordinates and *B*-factors were refined simultaneously with the *B*-factors restrained to the TNT *B*-correlation library (Tronrud, 1996). Eight more rounds of building and refinement, including further refinement of the NCS transformations resulted in an *R*-factor of 21.1%.

Anisotropic scaling was then included, which lowered the *R*-factor from 21.1 to 20.4%. Final model building and refinement resulted in a model which included one subunit (residues 3–1023), 437 solvent molecules, two Mg^{++}, five side chains modeled with two conformations, and three cysteines derivatized by β-mercaptoethanol. The entire model was constrained by the noncrystallographic symmetry.

It might be noted that initial attempts at *B*-factor refinement using all data to 2.5 Å resolution were unsuccessful in that the *B*-factors of many interior atoms decreased to near zero. Also the scaling profile of F_o to F_c was not well fit by the solvent model employed by TNT (Fig. 11A). A series of tests (not shown) suggested that the poor fit was caused by the lack of ordered solvent molecules in the model. To some extent the problem could be circumvented by including only the data between 8.0 and 2.5 Å resolution. A better procedure, however, was to include ordered solvent molecules prior to any *B* refinement and, as well, to include anisotropic scaling of F_o to F_c. The anisotropic scaling helped but the *B* refinement was still unstable without the addition of ordered solvent. The scaling profile for the final model is shown in Figure 11B.

Data collection, space group P2$_1$2$_1$2$_1$

Using nonfrozen crystals, data were initially collected at Daresbury to 2.8 Å resolution (Table 1). Data for structure refinement at high resolution were collected on a frozen crystal using beam line 5.02 at the Advanced Light Source with 30 s exposures, 0.5° oscillations and a wavelength of 1.0 Å. Reflections were visible to
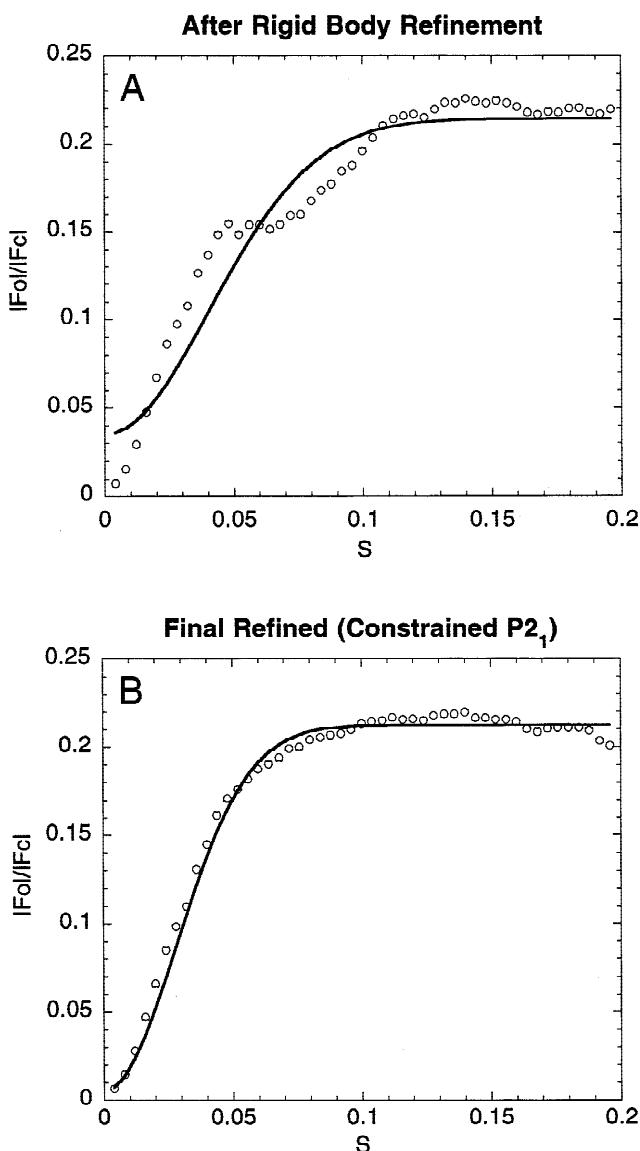
After Rigid Body Refinement

Final Refined (Constrained P2₁)

Fig. 11. Behavior of the scaling profile used to account for the scattering of bulk solvent. The model for bulk solvent used in TNT is based on Babinet's principle, which states that, at low resolution, the scattering of the bulk solvent is the inverse of that from the protein. This leads to a scaling function for F_o, which is defined by the functional form $(1/K)\exp(-B\sin^2\theta/\lambda^2)[1 - K_{sol}\exp(-B_{sol}\sin^2\theta/\lambda^2)]$ where K, B, K_{sol}, and B_{sol} are adjustable parameters that define the solvent scaling. In the figures the circles show the values of $\Sigma F_o/\Sigma F_c$ where F_o and F_c are the observed and calculated structure factor amplitudes calculated in increasing ranges of $\sin\theta/\lambda$. The solid line shows the scaling profile as derived from the best fit to the data of the above equation. **A:** Scaling after rigid-body refinement with the B-factor of all atoms at 27.5 Å². The poor fit is presumably due to the lack of ordered solvent in the model. **B:** Scaling for the final refined model including ordered solvent and with anisotropic scaling.

1.5 Å, and data were processed to 1.7 Å resolution with Mosflm/Scala (Kabsch, 1988; Leslie, 1990; Evans, 1993).

Structure determination and refinement, space group $P2_12_12_1$

The native Patterson function had a large peak (40% of the origin) at (0.0, 0.5, 0.48), suggesting that a local twofold axis was parallel

to a crystallographic 2_1 screw axis. The self-rotation function also suggested two pairs of perpendicular twofold axes in the xz plane rotated ~25° (+ or −) about the y axis.

Molecular replacement was based on the 2.8 Å resolution data set (Table 1) and the averaged β-galactosidase structure (see above) as a search model. Using the MRCHK suite of molecular replacement programs and GLRF (Tong & Rossmann, 1990; Zhang & Matthews, 1994), the rotation search gave peaks that aligned the tetramer so that its 222 axes coincided with diad axes observed in the self rotation function, including one parallel to the y axis.

The systematic absences alone did not clearly differentiate between the two possible space groups $P2_12_12_1$ and $P2_12_12$. Translation searches showed strong peaks for various combinations of monomers in both space groups. After rigid-body and positional refinement, the model in space group $P2_12_12_1$ had an R-factor of 20%, while that in space group $P2_12_12$ had an R-factor of 30%, suggesting that the former was the correct solution.

Following further rigid-body refinement at the tetramer, monomer, domain, and secondary structure levels, the model was averaged and subsequent refinement was done with constrained noncrystallographic symmetry. Several rounds of model inspection, solvent addition, and minimization resulted in a model with an overall R-factor of 16.8% at 2.8 Å resolution (Table 2). When the symmetry constraints were released this model refined to an R-factor of 13.6%.

After the higher resolution data were collected (Table 1), another model was built, again starting from the averaged P2₁ model with thermal factors set to the Wilson B of 17 Å². After rigid-body refinement each monomer was refined independently using all data to 1.7 Å resolution. During model building, one chain was inspected and adjusted, solvent molecules were added, and the overall structure re-refined. This procedure was then repeated for each subunit in turn. Many adjustments were necessary, mostly repositioning side chains. By the time that each chain had been rebuilt once, ~2,500 solvent molecules had been added, including water molecules, dimethylsulfoxide molecules, Mg^{++} ions, and Na^+ ions. At this point, the Automated Refinement Procedure (ARP) was implemented, adding about 2,500 more solvent molecules (Lamzin & Wilson, 1993). Approximately 1,000 of those added by ARP were subsequently removed by hand because the electron density and solvent-protein contacts were unconvincing. Several more cycles of model building resulted in a model with an overall R-factor of 15.7% at 1.7 Å resolution (Table 2). Several solvent molecules have been set with occupancies of 0.5 or 0.25. These molecules drifted out of density during refinement when their occupancies were 1.0. Halving the occupancy usually eliminated the drift. If it did not, the occupancy was halved again.

Analysis and calculations

EDPDB (Zhang & Matthews, 1995) was used for coordinate manipulations, solvent accessible surface area calculations, and to generate crystal contacts. MSRoll (Connolly, 1993) was used for volume and molecular surface calculations. All surface area calculations used a probe of radius 1.4 Å. Whatif (Vriend, 1990) was used to determine hydrogen bonds.

References

Appel SH, Alpers DH, Tomkins GM. 1965. Multiple molecular forms of β-galactosidase. *J Mol Biol 11*:12–22.

Breul A, Kuchinke W, von Wilcken-Bergmann B, Müller-Hill B. 1991. Linker mutagenesis in the *lacZ* gene of *Escherichia coli* yields variants of active β-galactosidase. *Eur J Biochem 195*:191–194.

Connolly ML. 1993. The molecular surface package. *J Mol Graphics 11*:139–141.

Edwards LA, Tian MR, Huber RE, Fowler AV. 1988. The use of limited proteolysis to probe interdomain and active site regions of β-galactosidase (*E. coli*). *J Biol Chem 263*:1848–1854.

Esnouf RM. 1999. Further additions to Molscript version 1.4, including reading and contouring electron density maps. *Acta Crystallogr D5*:938–940.

Evans PR. 1993. In: Sawer L, Isaacs N, Bailey S, eds. *Data reduction, proceedings of a CCP4 study weekend on data collection and processing.* Daresbury, UK: Daresbury Laboratory. pp 114–122.

Fowler A, Zabin I. 1978. Amino acid sequence of β-galactosidase. *J Biol Chem 253*:5521–5525.

Fowler A, Zabin I. 1983. Purification, structure, and properties of hybrid β-galactosidase proteins. *J Biol Chem 258*:14354–14358.

Gallagher CN, Huber RE. 1999. Stabilities of uncomplemented and complemented M15 β-galactosidase (*Escherichia coli*) and the relationship to α-complementation. *Biochem Cell Biol 77*:109–118.

Gebler JC, Aebersold R, Withers SG. 1992. Glu 537, not Glu 461, is the nucleophile in the active site of (*lac*Z) β-galactosidase from *Escherichia coli*. *J Biol Chem 267*:11126–11130.

Jacob F, Monod J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol 3*:318–356.

Jacobson RH, Matthews BW. 1992. Crystallization of β-galactosidase from *Escherichia coli*. *J Mol Biol 233*:1177–1182.

Jacobson RH, Zhang X-J, DuBose RF, Matthews BW. 1994. Three-dimensional structure of β-galactosidase from *E. coli*. *Nature 369*:761–766.

Jones TA, Zou JY, Cowan SW, Kjelgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A47*:110–119.

Juers DH, Huber RE, Matthews BW. 1999. Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between β-galactosidase and other glycohydrolases. *Protein Sci 8*:122–136.

Kabsch W. 1988. Evaluation of single X-ray diffraction data from a position sensitive detector. *J Appl Crystallogr 21*:916–924.

Kalnins A, Otto K, Ruther U, Müller-Hill B. 1983. Sequence of the lacZ gene of *Escherichia coli*. *EMBO J 2*:593–597.

Karlsson U, Koorajian S, Zabin I, Sjöstrand FS, Miller A. 1964. High resolution electron microscopy on highly purified β-galactosidase from *Escherichia coli*. *J Ultra Res 10*:457–469.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Lamzin VS, Wilson KS. 1993. Automated refinement of protein models. *Acta Crystallogr D49*:129–147.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Crystallogr 26*:283–291.

Leslie AGW. 1990. Molecular data processing. In: Moras D, Podjarny AD, Thierry JC, eds. *Crystallographic computing*. New York: Oxford University Press.

Miller S, Lesk AM, Janin J, Chothia C. 1987. The accessible surface area and stability of oligomeric proteins. *Nature 328*:834–836.

Müller-Hill B, Kania J. 1974. Lac repressor can be fused to β-galactosidase. *Nature 249*:561–563.

Nichtl A, Buchner J, Jaenicke R, Rudolph R, Scheibel T. 1998. Folding and association of β-galactosidase. *J Mol Biol 282*:1083–1091.

Richard JP, Huber RE, Lin S, Heo C, Amyes TL. 1996. Structure-reactivity relationships for β-galactosidase (*Escherichia coli*, lacZ). 3. Evidence that Glu-461 participates in Brönsted acid-base catalysis of beta-D-galactopyranosyl group transfer. *Biochemistry 35*:12377–12386.

Sinnott ML. 1990. Catalytic mechanisms of enzymic glycosyl transfer. *Chem Rev 90*:1171–1202.

Sinnott ML, Withers SG. 1978. The necessity of magnesium cation for acid assistance aglycore departure in catalysis by *Escherichia coli* (lacZ) betagalactosidase. *Biochem J 175*:539–546.

Stout GH, Jensen LH. 1989. *X-ray structure determination, a practical guide*, 2nd ed., Appendix B. New York: John Wiley & Sons.

Tong L, Rossmann MG. 1990. The locked rotation function. *Acta Crystallogr A46*:783–792.

Tronrud DE. 1996. Knowledge-based B-factor restraints for the refinement of proteins. *J Appl Crystallogr 29*:100–104.

Tronrud DE. 1997. TNT refinement package. *Methods Enzymol 277*:306–319.

Tronrud DE, Ten Eyck LF, Matthews BW. 1987. An efficient general-purpose least-squares refinement program for macromolecular structures. *Acta Crystallogr A43*:489–503.

Ullmann A. 1992. Complementation in β-galactosidase: From protein structure to genetic engineering. *Bioessays 14*:201–205.

Ullmann A, Jacob F, Monod J. 1967. Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the β-galactosidase structural gene of *Escherichia coli*. *J Mol Biol 24*:339–343.

Ullmann A, Perrin D, Jacob F, Monod J. 1965. Identification par complémentation in vitro et purification d'un segment peptidique de la β-galactosidase d'*Escherichia coli*. *J Mol Biol 12*:918–923.

Vriend G. 1990. A molecular modeling and drug design program. *J Mol Graphics 8*:52–56.

Wallenfels K, Weil R. 1972. β-Galactosidase. In: Boyer P, ed. *The enzymes*, 3rd ed., Vol. 7. London: Academic Press. pp 617–663.

Zabin I. 1982. β-Galactosidase α-complementation. *Mol Cell Biochem 49*:87–96.

Zeleny R, Altmann F, Praznik W. 1997. A capillary electrophoretic study on the specificity of β-galactosidases from *Aspergillus oryzae*, *Escherichia coli*, *Streptococcus pneumoniae*, and *Canavalia ensiformis*. *Analyt Biochem 246*:96–101.

Zhang X-J, Matthews BW. 1994. Enhancement of the method of molecular replacement by incorporation of known structural information. *Acta Crystallogr D50*:675–686.

Zhang X-J, Matthews BW. 1995. EDPDB: A multi-functional tool for protein structure analysis. *J Appl Crystallogr 28*:624–630.