A STRUCTURAL VIEW OF

BETA-GALACTOSIDASE

IN ACTION

by

DOUGLAS H. JUERS

A DISSERTATION

Presented to the Department of Physics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2000

"A Structural View of Beta-Galactosidase in Action", a dissertation prepared by Douglas H. Juers in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Physics. This dissertation has been approved and accepted by:

_____

Dr. Brian W. Matthews, Chair of the Examining Committee

May 26 '00

Date

Committee in charge: Dr. Brian W. Matthews, Chair
                         Dr. Bruce Branchaud
                         Dr. Stephen Gregory
                         Dr. S. James Remington
                         Dr. John Toner

Accepted by:

_____

Dean of the Graduate School

An Abstract of the Dissertation of

Douglas H. Juers   for the degree of   Doctor of Philosophy

in the Department of Physics   to be taken   June 2000

Title: A STRUCTURAL VIEW OF BETA-GALACTOSIDASE IN ACTION

Approved:_____ _____ _____
                   Dr. Brian W. Matthews

An atomic-level description of the presumed catalytic action of β-galactosidase is described. This large enzyme, from *E. coli*, carries out two reactions which allow the bacterium to live on the disaccharide lactose. First, it breaks down lactose to the two monosaccharides galactose and glucose. Second, it converts lactose into another disaccharide, allolactose, which is the inducer for the lac operon, and thus is the signal to the bacterium to produce more β-galactosidase. The work is based on high resolution x-ray crystallography and enzyme kinetics. A crystal form of β-galactosidase was isolated that permits data collection up to 1.5 Å resolution. Using this crystal form, the structures of several ligands bound to the enzyme were determined. These ligands were chosen to

mimic various points in the reaction: binding of substrate, covalent intermediates, transition states, and products. Together these complexes suggest a reaction coordinate for β-galactosidase which clarifies and enhances previous ideas about the reaction mechanism. The reaction includes a conformational change triggered by the progression of the substrate towards the transition state. Additional investigation suggests that this conformational change is involved in determining whether the enzyme carries out its hydrolysis or isomerization reaction. Considerations of the structure in the context of other related enzymes suggest an evolutionary path for β-galactosidase. It is suggested that a progenitor enzyme which catalyzed the hydrolysis of long polysaccharide substrates recruited additional domains which permit β-galactosidase to act on smaller substrates and produce the inducer, allolactose.

This dissertation includes both my previously published and co-authored material.

CURRICULUM VITA

NAME OF AUTHOR: Douglas H. Juers

PLACE OF BIRTH: Portland, Maine

DATE OF BIRTH: May 24, 1965

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

> University of Oregon
> Cornell University

DEGREES AWARDED:

> Doctor of Philosophy in Physics, 2000, University of Oregon
> Bachelor of Arts in Physics, 1987, Cornell University

AREAS OF SPECIAL INTEREST:

> Enzyme Structure and Mechanism
> The Application of Physics to Problems in Biology

PROFESSIONAL EXPERIENCE:

> Research Assistant, Department of Physics and Institute of Molecular Biology, University of Oregon, Eugene, 1993-2000

> Teaching Assistant, Department of Physics, University of Oregon, Eugene,1992-1993

> High School Teacher, Cycling & Snowboard Coach, and Dorm Parent, Carrabassett Valley Academy, Carrabassett Valley, Maine,1988-1992

> Crew Coach and Substitute Teacher, Cornell University and Ithaca City Schools, Ithaca, NY, 1987-1988

GRANTS:

PUBLICATIONS:

Holland, D.R., Hausrath, A.C., Juers, D. & Matthews, B.W. Structural analysis of zinc substitutions in the active site of thermolysin. *Protein Science* **4**, 1955-1965 (1995).

Juers, D.H., Huber, R.E. & Matthews, B.W. Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between β-galactosidase and other glycohydrolases. *Protein Science* **8**, 122-136 (1999).

ACKNOWLEDGMENTS

TABLE OF CONTENTS

APPENDIX

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Enzymes as Tiny Little Machines

Every living organism is full of enzymes. These are tiny protein molecules (~50-500 Å) which are the workhorses of life. They are responsible for building and maintaining infrastructure (membranes, filaments, tubules, bones, vessels, tendons, skin...), carrying oxygen and carbon dioxide, digesting food, producing movement, interpreting input signals (light, smell, sound), and generating responses. Anything a living organism does ultimately has an enzyme catalyzing a chemical reaction at its root. The purpose of the research presented here is to understand in atomic detail how one of these tiny machines works.

Enzyme Adaptation and the Operon Model for Gene Expression

Each organism has its own genes, which code for thousands of enzymes and other proteins. At any given moment, many of the possible enzymes aren't actually made, but are held in reserve to deal with situations as they arise. Organisms have evolved ways of responding to environmental changes (temperature, pressure, light, salinity, food) which

involve manufacturing enzymes specific for the new environmental condition and destroying enzymes specific for the old one. In this way energy is conserved by keeping on hand only the tools necessary for the job. The requirement for this system is the ability to sense the change in the environment and respond appropriately.

One environmental condition that organisms respond to is a change in food type. Early studies showed that yeast can grow on many different sugars, including glucose and galactose. [1, 2] Further, if yeast are grown on a mixture of glucose and another sugar, the glucose is hydrolyzed first, followed by the other sugar. Similar studies were later done in bacterial systems.

The adjustment to a new food source was thought to be controlled by a phenomenon called enzyme adaptation, for which two explanations were proposed. The "instructive theory" suggested there is a preexisting enzyme, which in the presence of the new sugar adapts to break it down. In the alternative, the "selective theory", the appearance of the new sugar triggers new production of an enzyme specific for its break down. In both cases the new sugar acts as an inducer, resulting in an enzyme which can break it down.

The lac System from *E. coli*

Several bacterial systems were used to study induction in enzyme adaptation[2]. However, the lac system in *Escherichia coli* proved to be the most fruitful. In this system, the new sugar is the disaccharide lactose. The appearance of lactose induces an active enzyme, β-galactosidase, which can hydrolyze lactose into the two

monosaccharides glucose and galactose.

The lac system had several advantages. First, it occurs in *E. coli*, so genetic crosses could be performed. Second, neither β–galactosidase activity nor its induction was specific for only lactose. Other galactosides could be hydrolyzed and/or act as inducers, which allowed for tests to separate hydrolysis from induction. And third, in 1950, a chromogenic substrate became available (ortho-nitrophenyl-β-D-galactoside), which made kinetic assays much easier than in other systems[3].

With these advantages, Jacques Monod and colleagues at the Institute Pasteur carried out activity and induction tests on a variety of small molecules. They showed that inducers are not necessarily substrates and substrates are not necessarily inducers. This suggested the instructive theory, which predicted all substrates and substrate analogues should be inducers, was incorrect.[4] Subsequently, using radioactively labeled nutrients, it was shown that the induced enzyme was synthesized from amino acids *de novo* only after the appearance of the inducer[5]. Additional studies suggested the existence of another protein, called the lac permease, which acted as a pump to move the inducer into the cell. Further investigations were focused on using genetic analysis to understand the mode of induction by distinguishing between the ideas of positive and negative control. Induction with positive control worked by initiating production of new enzyme, whereas with negative control, induction worked by stopping repression of enzyme production. The genetic experiments showed that production of β-galactosidase was governed by negative control.[6] Ultimately, this resulted in the operon model for gene expression.[7]

In the operon model, two types of genetic elements govern the synthesis of proteins. The structural elements code for proteins, such as β-galactosidase, whose

functions are the goal of induction. The other genetics elements, the regulator and

operator genes, control the expression of the structural elements through intermediate

cytoplasmic molecules, repressors and inducers (Figure 1).

In the lac system, there are three structural elements, the products of the Z,Y and

A genes. The Z gene codes for β-galactosidase, which will hydrolyze lactose to galactose

and glucose, thus permitting the bacterium to live on lactose[8]. The Y gene
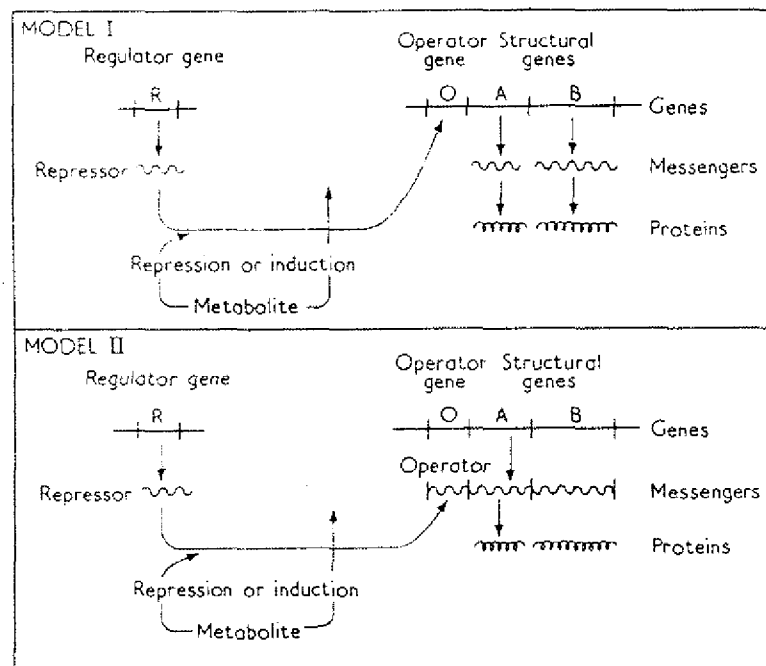
FIG. 6. Models of the regulation of protein synthesis.

Figure 1. The regulation of protein synthesis as originally presented by Jacob & Monod. The figure is from their review in the Journal of Molecular Biology.[7] Genes (DNA) are transcribed into messengers (RNA), which are translated into proteins. The repressor affects this process by binding to the operator, and binding is modulated by metabolites. In this figure, the repressor is shown as RNA, although it was later shown to be protein. The principal mode of regulation is via Model I. Translational regulation (Model II) also exists, but had been studied in much less detail.

codes for lac permease, a membrane bound transport protein responsible for transporting the lactose into the cell[7]. The A gene codes for thiogalactoside transacetylase, which appears to be involved in detoxification[9, 10]. It is also now know that the R gene also codes for a protein (the repressor). All the genes appear sequentially on the *E. coli* genome forming the lac operon.

The repressor protein binds to the DNA, preventing expression of the structural genes. When the inducer molecule binds to the repressor, the repressor releases from the DNA, allowing expression of the structural genes. Curiously, the inducer for the lac operon is not lactose itself, but another disaccharide, allolactose, which is made by β-galactosidase (Figure 2). Thus β-galactosidase makes is own inducer.

There are two apparent incongruities in this system. First is a chicken and egg paradox involving β-galactosidase, which makes the inducer, and the inducer, which triggers the synthesis of β-galactosidase. This paradox is resolved with the observation that gene expression is "leaky". The repressor has a certain probability to release from the DNA even in the absence of inducer[11, 12]. Thus there will always be a few molecules of lac permease and β-galactosidase in the cell to import lactose and create inducer (the basal level is about 10 molecules of each per cell)[13]. Second, there must be a way of processing the inducer to shut down the system once the lactose is depleted. This occurs via hydrolysis of allolactose to galactose and glucose by β-galactosidase[14]. So eventually β-galactosidase converts all of the lactose to galactose and glucose, but there is a transient build up of the inducer, allolactose.

Function of β-galactosidase
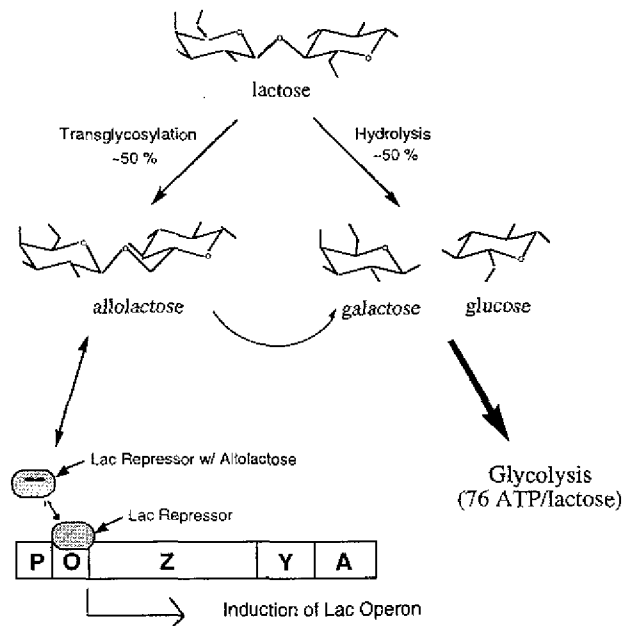
Figure 2. Schematic showing the function of β-galactosidase in *E. coli* metabolism. The enzyme hydrolyzes lactose, produces allolactose, and hydrolyzes allolactose.

It should be noted that even under high concentrations of lactose, the lactose system is repressed in the presence of glucose. This is due to two other effects. Inducer exclusion involves the inhibition of lac permease by glucose, while catabolite repression is another genetic regulatory system that responds to glucose levels.

The purpose of this dissertation is to explain in atomic detail how the enzyme β-galactosidase carries out its two reactions, hydrolysis of lactose to galactose and glucose, and conversion of lactose to allolactose.

<u>Biochemistry of β-galactosidase</u>

General Features of Hydrolysis

As mentioned above, β–galactosidase will cleave a variety of small disaccharide sized substrates. It is highly specific for galactosides, but will also cleave fucosides and arabinosides about 10^{-3} and 10^{-2} as efficiently (Figure 3). In addition to acting on O-galactosides, it will hydrolyze galactosides with N, S, and F linkages. The specificity is much less stringent at the aglycon position (corresponding to glucose in the lactose molecule), and kinetics on ~50 galactosides with different aglycons have been determined. Many offer easy chromogenic assays and the most common substrates for kinetic analysis are ortho-nitrophenyl-β-D-galactoside (onpg), para-nitrophenyl-β-D-galactoside (pnpg), and glucosyl-1,4-β-D-galactoside (lactose).

Hydrolysis by β-galactosidase involves retention of stereochemistry, suggesting a two-step reaction (Figure 4) [15]. The first step (galactosylation) involves bond cleavage and formation of a galactosyl-enzyme intermediate. The second step (degalactosylation) involves the addition of an acceptor molecule (usually water) to the intermediate releasing a product, which has retained the same stereochemistry as the starting state.

Two classes of experiments demonstrate the existence of two steps. First are methanol competition experiments done with various galactosides as substrates[16]. Here, methanol substitutes for water to form methyl-β-D-galactopyranoside (MG) as a product rather than galactose. For all aglycons, the ratio of galactose/MG produced is constant for a given concentration of methanol, suggesting the aglycon is released prior to reaction

with water or methanol. Second, under certain conditions with certain substrates, a burst

of aglycon is observed in the pre-steady state, indicating a fast first step, followed by a

slower rate limiting step[17].
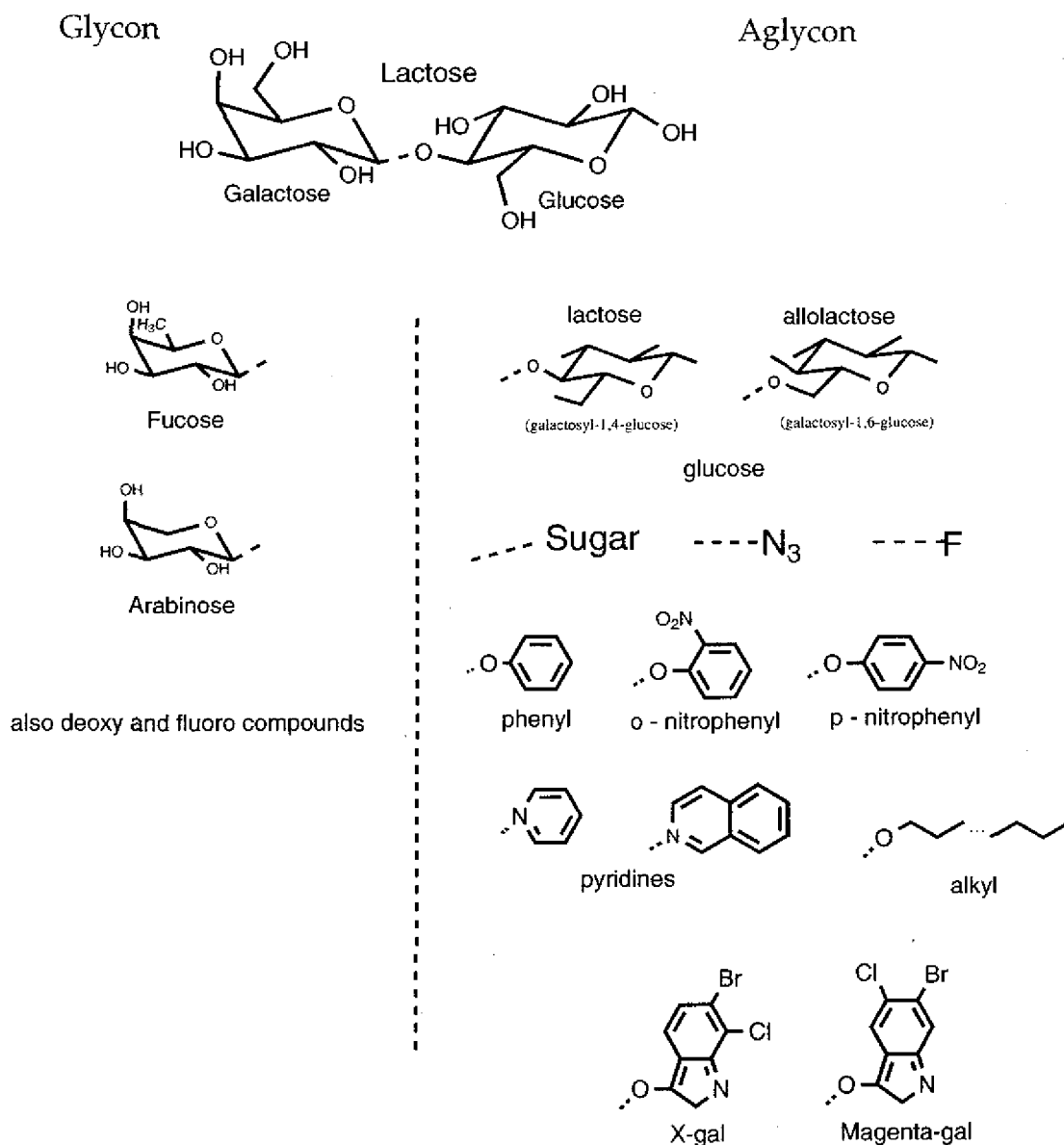
Glycon   Aglycon

Figure 3. Substrate specificity of β-galactosidase. Only a small fraction of possible
aglycons is shown. The specificity for the glycon is much more stringent. The bond
cleaved by β-galactosidase is shown as a dashed line.

a. A two step reaction retains stereochemisty

Step 1

enzyme intermediate with
inverted stereochemsitry

Step 2

product with same stereochemistry
as starting state
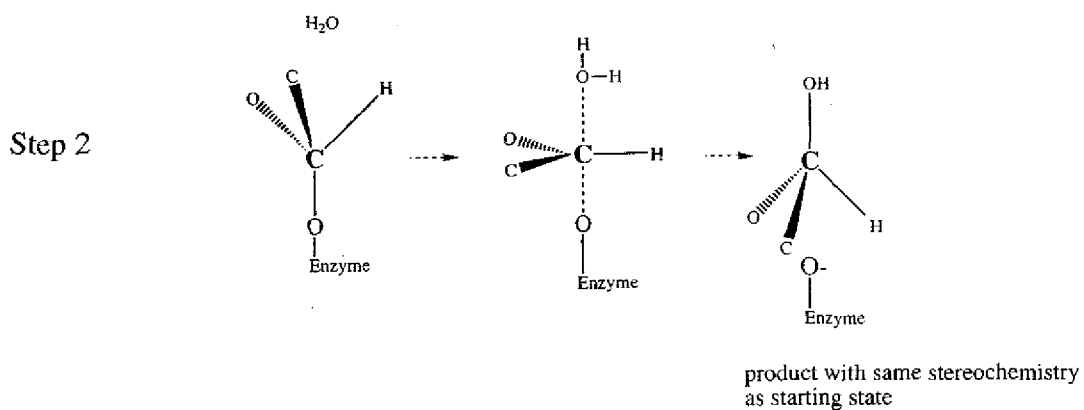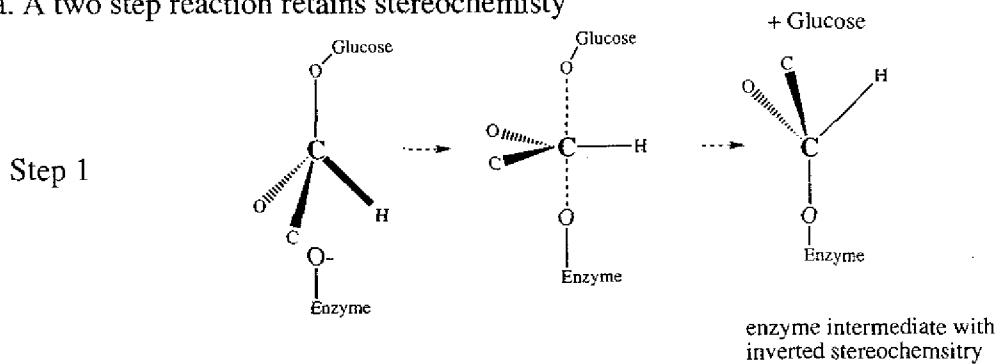
b. A single step reaction inverts stereochemisty

Step 1

product with
inverted stereochemsitry
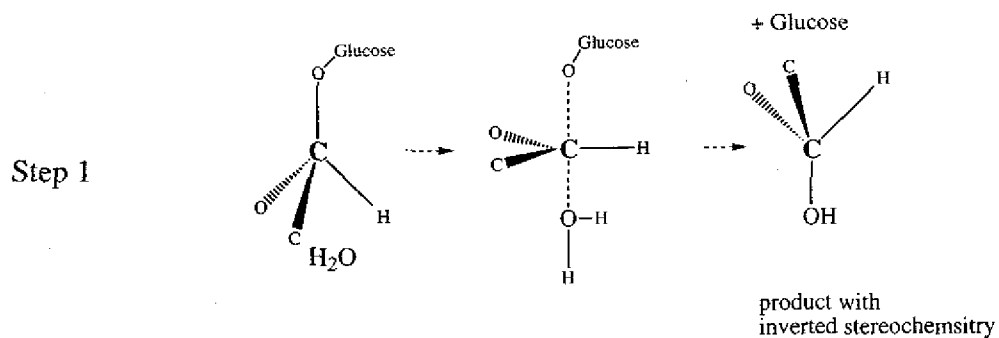
Figure 4. Retention vs inversion with nucleophilic substitutions. In both cases, water is the nucleophile substituting for glucose.

Overview of Reactions and Inhibition

A general overview of β-galactosidase catalyzed reactions is shown in Figure 5.

The principal reactions involving lactose (Figure 5(a)) are hydrolysis (k_2) to glucose and

Figure 5. Reactions catalyzed by β-galactosidase. All of the rate constants (k) and equilibrium constants (K) shown have been measured or approximated with inhibition studies(See Table 9, Chapter 3) . E + X means enzyme and X are separate in solution. E•X and E-X signify non-covalent and covalent complexes respectively.

a. Lactose is substrate. Lac=lactose; Gal=galactose; Glu=glucose; Allo=allolactose.
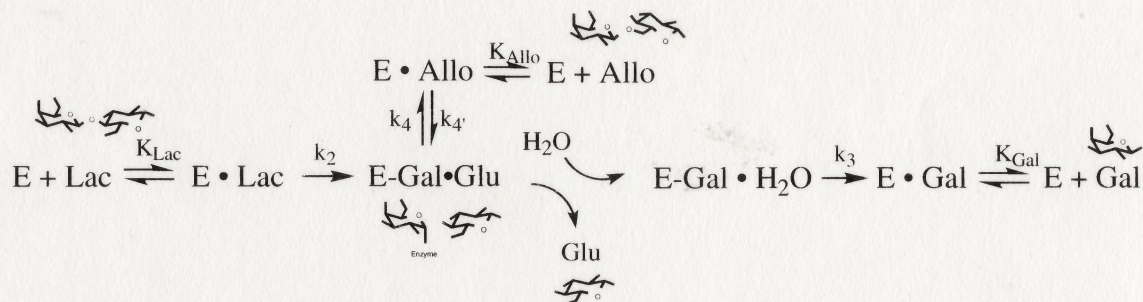
b. Generalized substrate with effector present. S = substrate, shown here as a galactoside; I = intermediate; A = effector; P1-3 = products. The red (left) path shows competitive inhibition by A. The blue (upper) path shows uncompetitive inhibition by A, although it includes the possibility of completing a reaction to produce product P3. For true uncompetitive inhibition k4 must be zero.
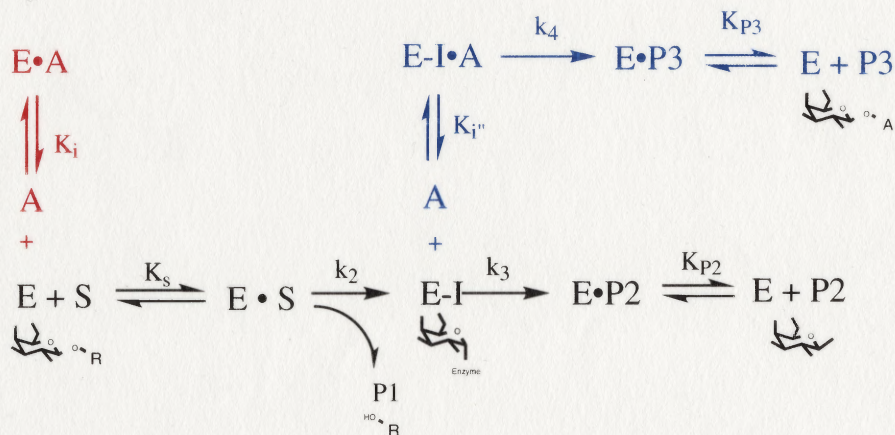
galactose and transglycosylation to allolactose. Dissociation of glucose after hydrolysis is shown as irreversible because experiments with radioactively labeled glucose have shown, at least at low concentrations of lactose (0.3 mM) and glucose (3 mM), that the formation of allolactose does not occur through binding of exogenous glucose to theintermediate but through reattachment of the just-cleaved glucose[18]. Also, the action of β-galactosidase on allolactose produces primarily glucose and galactose. Lactose is not produced at detectable levels, although other transglycosylation products are formed at high allolactose concentrations (> 50 mM)[14]. Thus the path from E-Gal•Glu to E•Lac is shown as irreversible, while interconversion between E-Gal•Glu and E•Allo is allowed.

The ratio of allolactose production/hydrolysis is about 1/1 in native enzyme[19]. This ratio should be important in the response of *E. coli* to lactose – small values should give poor induction, while large values should give excessive induction. Factors known to affect the ratio are pH and the presence of Mg^{++} ions (see below). Two of the questions addressed by the research presented here are first, what determines whether the enzyme simply hydrolyzes lactose or does the transglycosylation to produce allolactose and second, why is allolactose the preferred transglycosylation product?

Figure 5(b) shows the action of β-galactosidase on a generalized substrate with an exogenous effector present. The principal reaction, in black, is governed by two rate constants, k_2 for galactosylation and k_3 for degalactosylation, and a dissociation constant K_s for the binding of substrate to enzyme. It produces two products, P1 and P2, which are galactose and glucose if lactose is the substrate. Alternate pathways are shown in color. Red shows the binding of the effector A to free enzyme in competitive inhibition. Blue shows the same effector binding to the intermediate in uncompetitive inhibition. In many

cases, the intermediate bound effector can also act as an acceptor to release a third

product P3.

The inhibition by many small molecules has been measured. The best competitive inhibitors have nanomolar dissociation constants, whereas K_s for the most common substrates is in the range 0.1-1.0 mM. Uncompetitive inhibitors (those which bind more tightly to the intermediate than the free enzyme) usually bind at about 10-100 mM to the intermediate and 100-1000 mM to free enzyme. Non-competitive inhibitors (those which bind equally well to the intermediate and free enzyme) typically bind at about 100-500 mM. Another question addressed by the research presented here is what is the structural basis for the various types of inhibition?

Details of the Transition State

There is some disagreement regarding the details of the transition state for β–galactosidase hydrolysis. The basic argument goes as follows[20]. The enzyme will efficiently cleave galactosyl pyridinium salts, for which galactosylation (step 1) is rate limiting [21]. Because they have an aromatic nitrogen (which has no free lone pair) at the O4 position, the pyridinium salts are precluded from undergoing acid catalysis. They show significant α–deuterium isotope effects and reasonably linear Bronsted plots of slope ~ 1, suggesting that bond cleavage is rate limiting and that it occurs without acid catalysis in an SN1 process. These parameters depend little on the presence of Mg^{++}.

For aryl-O-galactosides, which are able undergo acid catalysis[22], the catalytic parameters in the absence of Mg^{++} (including α–deuterium isotope effects and Bronsted

plots) are similar to the pyridinium salts. This suggests that for all substrates, without Mg^{++} there is little acid assistance with galactosylation.

With Mg^{++} present, the aryl galactosides show more complex behavior. The Bronsted plots are more scattered, and there is no α–deuterium isotope effect for slow substrates[23]. This casts doubt on whether bond cleavage is rate limiting for the aryl galactosides in the presence of Mg^{++}. However, it was further shown that for a slow substrate (para-nitrophenyl-β-D-galactopyranoside, or pnpg) there is a primary isotope effect on both k_{cat} and k_{cat}/K_m[24]. Additionally, for hydrolysis of pnpg, is there is a solvent isotope effect on k_{cat} and this is due to a single proton. However, there is no solvent isotope effect on k_{cat}/K_m, meaning that steps up through the first irreversible step are insensitive to this proton transfer (or there are offsetting changes to k_{cat} and K_m)[25]. Together these data have been interpreted in two ways.

One possible interpretation is that because of the primary isotope effect, C-O bond cleavage is rate limiting for aryl galactosides in the presence of Mg^{++}. The absence of an α–deuterium isotope effect suggests this process is either SN2 or involves a substrate-assisted element to maintain the hybridization at the anomeric center. The scatter in the Bronsted plot is due to structural variations in the substrates, which apparently matters only if Mg^{++} is present. Acid catalysis would be facilitated by an enzyme group donating a proton to the O4 of the leaving group in concert with C-O bond cleavage in the rate determining process, accounting for the solvent isotope effect. This model, however, does not account for the lack of solvent isotope effect on k_{cat}/K_m for pnpg hydrolysis.

The other possibility is that the Mg^{++} acts as a direct electrophilic catalyst, forming a complex with O4 of the leaving group in concert with the formation of a galactosyl enzyme intermediate[20]. The rate determining process would be release of the Mg-O-Ar complex via donation of a proton. This would account for both the primary and solvent isotope effects. And because the galactosyl moiety is not involved in the rate determining step, there would be no α–deuterium isotope effect.

The basic difference between the two interpretations is the mode of acid catalysis. In the first interpretation, acid catalysis is accomplished by donating a proton (Bronsted catalysis) from an enzyme group to the leaving group oxygen. In the second interpretation, acid catalysis is accomplished when the Mg^{++} accepts a lone pair (Lewis catalysis) from the leaving group oxygen. One of the objectives of the research presented here is to distinguish between these two interpretations by determining the location of the leaving group oxygen relative to the Mg^{++} at various points along the reaction path.

The above discussion is based on experiments done on two classes of substrates – the pyridinium salts and the aryl galactosides. Bronsted parameters have also been determined for a third class of substrates, the alkyl galactosides. These substrates have more basic leaving groups and are therefore expected to better mimic the natural substrate lactose. With Mg^{++}, β_{lg} for the alkyl galactosides is –0.49, suggesting at least some acid assistance for C-O bond cleavage[26].

Details concerning degalactosylation are less controversial, but there are still ambiguities. There is a significant α–deuterium isotope effect for this step, but no solvent isotope effect. Both of these observations are consistent with an SN1 type

process. However, the rate of degalactosylation depends on the acceptor concentration, suggesting involvement of the acceptor in the transition state, which is inconsistent with an SN1 process. To explain this paradox, Sinnott has proposed a reversible ion pair between the galactosyl intermediate and the enzyme nucleophile[17]. Nucleophilic attack of the intermediate then occurs on the ion pair rather than the covalently bound galactoside. This accounts for both the α–deuterium isotope effect and the dependence of the rate on the acceptor concentration.

Interactions between Substrate and Enzyme

Interactions between substrate and enzyme are the source of the catalytic power of an enzyme, so the details of these interactions are expected to help in understanding how the enzyme works. The goal of the research presented here is to delineate these interactions at different points along the reaction coordinate.

Due to the two step nature of the reaction, there is expected to be a nucleophilic enzyme group that stabilizes the enzyme bound intermediate. Early pH studies and analogy with esterases suggested that this group might be a histidine side chain [27]. Later studies on other glycosidases, including lysozyme, implicated a pair of carboxylates, one as nucleophile and one as an acid to donate a proton to the leaving group [28]. With β–galactosidase, labeling experiments were used to identify these important catalytic groups.

A galactosyl carbonium ion was used to label Met 502, locating it to the active site[29]. This residue was rejected as a catalytic acid, but the possibility of Tyr 503 acting

in an acid role was noted. This is consistent with the pH profile, which shows a downturn in k_{cat} at pH ~9, and mutations made at this site supported the idea of Tyr 503 acting as an acid/base catalyst[30, 31]. Subsequently, Glu 461 was labeled with an epoxide and it was initially proposed as the nucleophile in the reaction[32]. Enzymes altered at this site had substantially decreased activity, consistent with its proposed role as a nucleophile[33]. However, dinitrophenyl-2-F-galactoside, a specific, mechanism based inactivator, was used to label Glu 537, suggesting that it, instead of Glu 461 was the nucleophile[34]. These authors suggested that Glu 461 was the acid/base and discussed the Glu 461 mutations in this light. Subsequently, mutations at Glu 537 were made, and found to have larger decreases in activity than those at Glu 461 (10^3-10^5 decrease, vs 10^2-10^3 decrease for Glu 461)[35]. Thus, prior to any structural information from x-ray crystallography, there was good evidence that Glu 537 acts as the active site nucleophile, with either Glu 461 or the active site Mg^{++} acting as a catalytic acid/base.

The initial structure determination of β–galactosidase located the above residues in a deep pocket on the surface of the enzyme – the presumed active site[36]. Subsequently, other site-directed mutagenesis studies have implicated Histidines 540, 391, and 357 as interacting with the substrate during catalysis[37-39]. These will be discussed in more detail later.

A final curious case is the observation of a mutation resulting in an increased rate of lactose hydrolysis – G794D. This mutation was identified in a temperature-sensitive screen as a β–galactosidase variant with lower thermal stability than native enzyme[39]. Subsequently, it was isolated and kinetics showed it to have a much larger galactosylation rate than native enzyme[40]. Mutations resulting in increased activity are uncommon, and

one of the goals of the research presented here is to understand the role of residue 794

in the catalytic mechanism.

Energetics

Figure 6 shows the nonenzymatic hydrolysis of the three common substrates of

β–galactosidase: ortho-nitrophenyl-β-D-galactopyranoside (onpg), para-nitrophenyl-β-D-

galactopyranoside (pnpg), and glucosyl-1,4-β-D-galactopyranoside (lactose). The rate

constants for these degradations are k_{non} = 2.2 x 10^{-9}, 3.2 x 10^{-9}, and < 10^{-10} s^{-1}

respectively. The corresponding rate constants for the enzyme catalyzed breakdown

from the Michaelis complex are k_{cat} = 1200, 90, and 60 s^{-1}. This results in a rate

enhancement of k_{cat}/k_{non} = 5 x 10^{11}, 3 x 10^{10}, and > 6 x 10^{11} for hydrolysis of onpg. pnpg,

and lactose. Several studies have been done suggesting quantitatively how much this

rate enhancement is facilitated by various galactosyl substituents and enzyme groups[37, 38,

41]. To the extent that binding interactions control the energetics, any reaction coordinate

proposed from crystallography should make sense in the context of these studies.

Parameters Affecting Catalysis

pH

Studies on the pH dependence of β–galactosidase catalyzed hydrolysis show a

relatively broad maximum at a pH of 7.0 [42]. For onpg hydrolysis, there is one

Nonenymatic Hydrolysis (Room Temp)
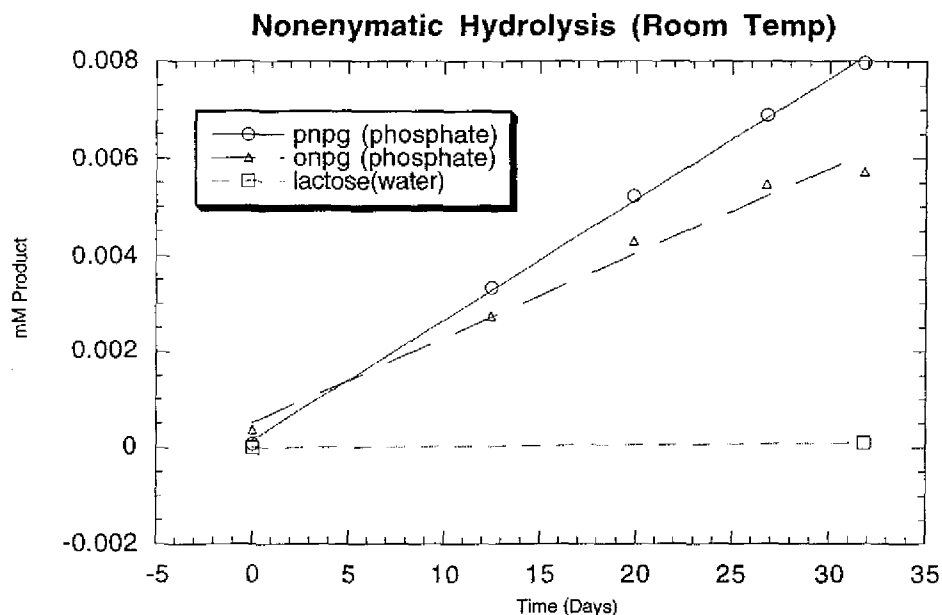
Figure 6. Non-Enzymatic hydrolysis of onpg, pnpg, and lactose using 1 mM solutions at room temp. Onpg and pnpg hydrolysis were measured in 50 mM Na_2HPO_4, 1 mM $MgCl_2$ using absorption spectroscopy to monitor the signal from nitrophenol at 420 nM. The single data point for lactose hydrolysis is an estimation based on the fact that less than 0.2% of a solution of 50 mM lactose (in water) was hydrolyzed over 1 year. Lactose and its products were measured with a capillary electrophoresis assay. The corresponding rate constants are: onpg: 2.2×10^{-9} s^{-1}, pnpg: $3.2 \times 10^{-9}s^{-1}$, lactose $3.6 \times 10^{-11}s^{-1}$.

protonated group which ionizes at alkaline pH (about 8.5) and at least one unprotonated group which ionizes at acidic pH (below 6). In the absence of Mg^{++}, the group ionizing at pH 8.5 shifts to a pKa of about 6.5.

Metal Ions

Both magnesium and alkali ions are required for maximal activity, but the enzyme is active in the absence of either ion. Manganese can substitute for magnesium with

almost no observable difference in the kinetics while calcium binds but does not

activate[43]. The alkali metals have more complex behavior. The β-galactosidase system

was one of the first enzymes observed to have a metal activation that depended on

substrate. Sodium ions activate hydrolysis of onpg greater than lactose and pnpg, while

potassium ions activate hydrolysis of lactose and pnpg greater than onpg[8]. Additionally,

some synergy has been observed with mixtures of alkali ions in which activity with

sodium and cesium is greater than the sum of the individual activities[44]. Another goal of

the research presented here is to understand the metal ion activation.

Temperature

The temperature dependence of β-galactosidase-catalyzed hydrolysis of several

substrates shows Arrhenius behavior for k_{cat} over the range 0 to 30° C[45]. Above 30° C the

activity deviates from Arrhenius. The activity in the range from −80° C to 10° C to has

also been measured for onpg hydrolysis with DMSO as a cryosolvent[46]. In this case the

Arrhenius plot shows a slow decrease in slope with increasing temperature. However,

the mechanism for β-galactosidase catalyzed hydrolysis is complex, involving two

separate steps, so the Arrhenius behavior is likely to be nonlinear.

Pressure

There have been at least two studies involving the effects of pressure on β-

galactosidase activity[47-49]. In one study, the pressure dependence on activity was

compared with alkali metal activation. It was found that increasing pressure (up to

1500 atm) decreased K^+ activation with all substrates (onpg, pnpg and lactose). In

contrast, pressure increase Na^+ activation with pnpg and lactose but had a substrate

dependent effect with onpg. A more detailed study on onpg hydrolysis with Na^+ showed

that increasing pressure decreased k_{cat} mostly by decreasing k_3. Also, glucose was a better

inhibitor with pressure and lactose was a worse inhibitor while the inhibition of galactose

was unaffected by pressure.

Molecular Biology of β–Galactosidase

Beta-galactosidase has several properties that have made it an invaluable tool for

molecular and cell biologists. First, as mentioned above, various chromogenic substrates

exist for the enzyme. In particular, X-gal is a galactoside that turns from colorless to blue

upon cleavage. Second, other polypeptide sequences can be fused at sites up to residue 23

of β–galactosidase without affecting hydrolysis[50]. And third, certain deletions in the

β–galactosidase protein which are inactive can be complemented in trans by adding an

appropriate separate short polypeptide, producing active enzyme[51]. This phenomenon is

known as either α or ω complementation.

Fusion proteins with β–galactosidase have been used extensively as reporters to

monitor gene expression. In these experiments, the lac Z gene (which codes for

β–galactosidase) is fused to the promoter of the gene of interest. When this gene is

turned on, β–galactosidase is produced and can be monitored with one of the

chromogenic substrates for β–galactosidase.

Alpha–complementation has been used extensively in cloning procedures.

Here, a strain of *E. coli* that bears a deletion in the lac Z gene is used. The cloning vector contains the gene sequence coding for the β–galactosidase complementation polypeptide. If this vector is transferred into the *E. coli* an active complemented β–galactosidase is produced and colonies grown on X-gal will turn blue. The cloning vector is constructed such that insertion of the gene of interest disrupts the β–galactosidase complementation peptide, so complemented β–galactosidase is no longer produced. Thus colonies that do not turn blue suggest that the gene of interest has been successfully incorporated into the cloning vector.

Part of the research presented here is to understand how complementation works in β–galactosidase.

Summary

Beta–galactosidase has a rich history but until recently there has been no high resolution structural information about the enzyme or its interactions with substrate. The main goal of the research presented here is to gain a deeper understanding of how β–galactosidase works by defining in detail the interactions between the enzyme and its substrate during its catalytic cycle and interpreting these interactions in light of previous genetic, biochemical and biophysical experiments. A secondary goal, which is an outgrowth of the structural studies presented here, is to understand more fully what happens when a protein crystal is flash frozen.

This dissertation contains co-authored material in Chapters 2 and 4. I am grateful to my co-authors for their insight and effort in preparing the manuscripts. The introduction to each of these chapters contains more details about the co-authorship.

<u>References</u>

1.   Dienert, F. Sur La Fermentation Du Galactose. *Annales de l'Institute Pasteur* **14**, 139-189 (1900).

2.   Müller-Hill, B. *The lac Operon*, (Walter de Gruyter, New York, 1996).

3.   Lederberg, J. The beta-D-galactosidase of Escherichia coli, strain K-12. **60**, 381-392 (1950).

4.   Monod, J., Germaine, C.-B. & Cohn, M. Sur La Biosynthese del la ß-Galactosidase (Lactase) chez *Escherichia coli*. La Specificite de l'Induction. *Biochimica et Biophysica Acta* **7**, 585-598 (1951).

5.   Hogness, D.S., Cohn, M. & Monod, J. Studies on the induced synthesis of beta-galactosidase in Escherichia coli: the kinetics and mechanism of sulfur incorporation. *Biochim Biophys Acta* **16**, 99 (1955).

6.   Pardee, A.B., Jacob, F. & Monod, J. The Genetic Control and Cytoplasmic Expression of "Inducibility" in the Synthesis of ß-galatosidase by *E. coli*. *JMB* **1**, 165-178 (1959).

7.   Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *JMB* **3**, 318-356 (1961).

8.   Cohn, M. & Monod, J. Purification et proprietes de la ß-galactosidase (lactase) d'Escherichia coli. *Biochimica et Biophysica Acta* **7**, 153-174 (1951).

9.   Musso, R. & Zabin, I. Substrate Specificity and Kinetic Studies on Thiogalactoside Transacetylase. *Biochemistry* **12**, 552-557 (1973).

10.   Andrews, K. & Lin, E. Thiogalactoside Transacetylase of the Lactose Operon as an Enzyme for Detoxification. *Journal of Bacteriology* **128**, 510-512 (1976).

11.   Berg, O.G., Winter, R.B. & von Hippel, P.H. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids 1. Models and Theory. *Biochemistry* **20**, 6929-6948 (1981).

12. Wong, P., Gladney, S. & Keasling, J.D. Mathematical Model of the lac Operon: Inducer Exlusion, Catabolite Repression, and Diauxic Growth on Glucose and Lactose. *Biotechnology Progress* **13**, 132-143 (1997).

13. Beckwith, J.R. & Zipser, D. The lactose operon. (Cold Spring Harbor, New York, 1970).

14. Huber, R.E., Wallenfels, K. & Kurz, G. The action of beta-galactosidase (Escherichia coli) on allolactose. *Can J Biochem* **53**, 1035-1038 (1975).

15. Koshland, D.E. Stereochemstry and the Mechanism of Enzymatic Reactions. *Biological Reviews* , 416-436 (1953).

16. Stokes, T.M. & Wilson, I.B. A common intermediate in the hydrolysis of - galactosides by -galactosidase from Escherichia coli. *Biochemistry* **11**, 1061-1064 (1972).

17. Sinnott, M.L. Ions, ion-pairs and catalysis by the lacZ beta-galactosidase of Escherichia coli. *FEBS Lett* **94**, 1-9 (1978).

18. Jobe, A. & Bourgeois, S. *lac* Repressor-Operator Interaction VI. The Natural Inducer of the *lac* Operon. *JMB* **69**, 397-408 (1972).

19. Huber, R.E., Kurz, G. & Wallenfels, K. A quantitation of the factors which affect the hydrolase and transgalactosylase activities of beta-galactosidase (E. coli) on lactose. *Biochemistry* **15**, 1994-2001 (1976).

20. Sinnott, M.L. Catalytic Mechanisms of Enzymic Glycosyl Transfer. *Chemical Reviews* **90**, 1171-1202 (1990).

21. Sinnott, M.L. & Withers, S.G. The beta-galactosidase-catalysed hydrolyses of beta-d-galactopyranosyl pyridium salts. Rate-limiting generation of an enzyme-bound galactopyranosyl cation in a process dependent only on aglycone acidity. *Biochem J* **143**, 751-762 (1974).

22. Sinnott, M.L. & Withers, S.G. The necessity of magnesium cation for acid assistance aglycone departure in catalysis by Escherichia coli (lacZ) beta-galactosidase. *Biochem J* **175**, 539-546 (1978).

23. Sinnott, M.L. & Souchard, I.J. The mechanism of action of beta-galactosidase. Effect of aglycone nature and -deuterium substitution on the hydrolysis of aryl galactosides. *Biochem J* **133**, 89-98 (1973).

24. Rosenberg, S. & Kirsch, J.F. Oxygen-18 Leaving Group Kinetic Isotope Effects on the Hydrolysis of Nitrophenyl Glycosides. 1. ß-Galactosidase-Catalyzed Hydrolysis. *Biochemistry* **20**(1981).

25.   Selwood, T. & Sinnott, M.L. A solvent-isotope-effect study of proton transfer during catalysis by Escherichia coli (lacZ) beta-galactosidase. *Biochem J* **268**, 317-323 (1990).

26.   Richard, J.P., Westerfield, J.G., Lin, S. & Beard, J. Structure-Reactivity Relationships for ß-Galactosidase (*Escherichia coli*, lac Z). 1. Reactions of the Galactosyl-Enzyme Intermediate with Alcohols and Azide Ion. *Biochemistry* **34**, 11713-11724 (1995).

27.   Wallenfels, K. & Molhotra, O.P. Galactosidases. *Advances in Carbohydrate Chemistry* **16**, 239-298 (1961).

28.   Parsons, S.M. & Raftery, M.A. Ionization behavior of the cleft carboxyls in lysozyme-substrate complexes. *Biochemistry* **11**, 1633-1638 (1972).

29.   Sinnott, M.L. & Smith, P.L. Affinity Labelling with a Deaminatively Generated Carbonium Ion. *Biochemical Journal* **175**, 525-538 (1978).

30.   Ring, M., Bader, D.E. & Huber, R.E. Site-directed mutagenesis of beta-galactosidase (E. coli) reveals that tyr-503 is essential for activity. *Biochem Biophys Res Commun* **152**, 1050-1055 (1988).

31.   Ring, M. & Huber, R.E. Multiple replacements establish the importance of tyrosine-503 in beta-galactosidase (Escherichia coli). *Arch Biochem Biophys* **283**, 342-350 (1990).

32.   Herrchen, M. & Legler, G. Identification of an essential carboxylate group at the active site of lacZ beta-galactosidase from Escherichia coli. *Eur J Biochem* **138**, 527-531 (1984).

33.   Cupples, C.G., Miller, J.H. & Huber, R.E. Determination of the roles of Glu-461 in beta-galactosidase (Escherichia coli) using site-specific mutagenesis. *J Biol Chem* **265**, 5512-5518 (1990).

34.   Gebler, J.C., Aebersold, R. & Withers, S.G. Glu-537, not Glu-461, is the nucleophile in the active site of (lac Z) beta-galactosidase from Escherichia coli. *J Biol Chem* **267**, 11126-11130 (1992).

35.   Yuan, J., Martinez-Bilbao, M. & Huber, R.E. Substitutions for Glu-537 of beta-galactosidase from Escherichia coli cause large decreases in catalytic activity. *Biochem J* **299**, 527-531 (1994).

36.   Jacobson, R.H., Zhang, X.-J., DuBose, R.F. & Matthews, B.W. Three-dimensional structure of ß-galactosidase from *E. coli*. *Nature* **369**, 761-766 (1994).

37.   Roth, N.J. & Huber, R.E. The beta-galactosidase (Escherichia coli) reaction is partly facilitated by interactions of His-540 with the C6 hydroxyl of galactose. *J Biol Chem* **271**, 14296-14301 (1996).

38.   Roth, N.J., Rob, B. & Huber, R.E. His-357 of beta-galactosidase (Escherichia coli) interacts with the C3 hydroxyl in the transition state and helps to mediate catalysis. *Biochemistry* **37**, 10099-10107 (1998).

39.   Huber, R. Personal Communication. (2000).

40.   Martinez-Bilbao, M., Holdsworth, R.E., Edwards, L.A. & Huber, R.E. A highly reactive beta-galactosidase (Escherichia coli) resulting from a substitution of an aspartic acid for Gly-794. *J Biol Chem* **266**, 4979-4986 (1991).

41.   McCarter, J.D., Adam, M.J. & Withers, S.G. Binding energy and catalysis. Fluorinated and deoxygenated glycosides as mechanistic probes of Escherichia coli (lacZ) beta-galactosidase. *Biochem J* **286**, 721-727 (1992).

42.   Tenu, J.P., Viratelle, O.M., Garnier, J. & Yon, J. pH dependence of the activity of beta-galactosidase from Escherichia coli. *Eur J Biochem* **20**, 363-370 (1971).

43.   Huber, R.E., Parfett, C., Woulfe-Flanagan, H. & Thompson, D.J. Interaction of divalent cations with beta-galactosidase (Escherichia coli). *Biochemistry* **18**, 4090-4095 (1979).

44.   Neville, M.C. & Ling, G.N. Synergistic Activation of ß-Galactosidase by Na+ and Cs+. *Archived of Biochemistry and Biophysics* **118**, 596-610 (1967).

45.   Kuby, S.A. & Lardy, H.A. Purification and Kinetics of ß-D-Galactosidase from *Escherichia coli. Biochemistry* **75**, 890-896 (1953).

46.   More, N., Daniel, R.M. & Petach, H.H. The effect of low temperature on enzyme activity. *Biochemical Journal* **305**, 17-20 (1995).

47.   Degraeve, P. & Lemay, P. High pressure-induced modulation of the activity and stability of Escherichia coli (lacZ) β-galactosidase: Potential applications. *Enyzme and Microbial Technology* **20**, 550-557 (1997).

48.   Degraeve, P., Delorme, P. & Lemay, P. Pressure-induced inactivation of E. coli beta-galactosidase: influence of pH and temperature. *Biochim Biophys Acta* **1292**, 61-68 (1996).

49.   Becker, V.E. & Evans, H.J. The influence of monovalent cations and hydrostatic pressure on beta-galactosidase activity. *Biochimica et Biophysica Acta* **191**, 95-104 (1969).

50.   Fowler, A. & Zabin, I. Purification, structure, and properties of hybrid ß-galactosidase proteins. *J. Biol. Chemistry* **258**, 14354-14358 (1983).

51.   Ullmann, A., Perrin, D., Jacob, F. & Monod, J. Identification par Complementation in vitro et Purification d'un Segment Peptidique de la ß-Galactosidase d'*Escherichia coli*. *JMB* **12**, 918-923 (1965).

CHAPTER II

THE HIGH RESOLUTION STRUCTURE OF BETA-GALACTOSIDASE

Introduction

The three dimensional structure of β-galactosidase was originally determined in a monoclinic crystal form which includes 16 independent chains[1]. This structure has been refined to 2.5Å resolution. Because of the large unit cell for these crystals (108 Å x 210 Å x 510 Å, β=95°) data collection is problematic and cannot be performed in house to high resolution. Therefore for further mechanistic studies a different crystal form was desirable. Such a form was isolated by Dr. Dale Wigley at Oxford and reproduced in Eugene with the help of Dr. Gene Huber from the University of Calgary while he was on sabbatical leave. This orthorhombic crystal form has smaller cell dimensions (154 Å x 172 Å x 204 Å), permitting data collection in house up to 2.5 Å resolution. Additionally with these crystals, synchrotron radiation has produced diffraction up to 1.1 Å resolution and data sets up to 1.5 Å resolution. This chapter describes the structure determination and refinement in the orthorhombic form to 1.7 Å resolution.

This chapter includes co-authored material. Brian Matthews provided much help in writing the manuscript. Ray Jacobson, who originally determined the structure of β-galactosidase in the monoclinic crystal form, was very helpful as he transferred the

project to me, teaching me the finer points of dealing with such a large molecule and unit cell. Additionally, Ray did the initial refinement in the monoclinic form, and I continued where he left off. Cai Zhang and Dale Tronrud were very helpful with discussions concerning molecular replacement and refinement. Finally, I am indebted to Dale Wigley for sharing his crystallization conditions and Gene Huber for helping me reproduce these orthorhombic crystals here in Eugene.

Results

Refinement, Space Group $P2_1$

The initial determination of the structure of β–galactosidase in space group $P2_1$ was described by Jacobson et al. [1]. This structure was initially refined using TNT [2] to an R-factor of 17.4% without the imposition of non-crystallographic symmetry constraints or restraints (see Methods and Tables 1 and 2). This model includes residues 3-1023 for each chain with approximately 93 solvent molecules per monomer. In general the electron density (Figure 7(b)) 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 which appears to be quite mobile, and an extended region of chain between residues 727 and 733 that appears to display multiple conformations. Examination of a Luzatti plot (not shown) suggests an r.m.s. coordinate error of less than 0.35Å. The model was also refined with imposition of 16-fold constrained non-crystallographic symmetry. In this

case the R-factor was 19.9% (Table 2). Although R_{free} is quite low (20.7%) the high

degree of NCS biases the randomly chosen test set towards the working set [3](Kleywegt &

Brünger, 1996). The Ramachandran plot (not shown) has 83% of the residues in the most

favored regions, and acceptable overall stereochemistry.

Table 1. X-ray data collection – native data. The data in space group $P2_1$ are from
Jacobson et al. (1994)[1]. Numbers in parentheses correspond to the outermost shell of
data.

Space group	$P2_1$	$P2_12_12_1$	$P2_12_12_1$
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	--	--

Table 2. X-ray refinement of native structures. 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 mainchain, sidechain and solvent atoms. K_{sol} and B_{sol} are the parameters specifying the bulk solvent model used by TNT [2]. B11, B22, B33 and B13 are the parameters defining the overall anisotropic scaling of the data [4]. Cons. = constrained and Uncons. = unconstrained.

	Space group P2₁		Space group P2₁2₁2₁		
Temperature	RT	RT	RT	RT	~95K
Mode of refinement	Cons. (16 mon.)	Uncons.	Cons. (4 mon.)	Uncons.	Uncons.
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 x 8232)	131,168	33,312 (4 x 8328)	32,952	32,500
Solvent atoms	6992 (16 x 437)	1486	1472 (4 x 368)	856	4908
Δ_{bond} (Å)	0.018	0.016	0.016	0.016	0.018
Δ_{angle} (°)	2.7	2.6	2.6	2.8	2.8
_B (Å2)	5.5		4.7	6.2	7.5
$<B>$main (Å2)	31.2	29.5	30.2	28.3	16.8
$<B>$side (Å2)	38.7	35.5	36.6	33.7	22.2
$<B>$solvent (Å2)	48.0	28.0	46.6	32.6	31.5
R-factor (%)	19.9	17.4	16.8	13.6	15.7
R-free (%)	20.7	--	19.7	28.0	21.1
Ksol	0.98	0.8	0.84	0.81	0.66
Bsol (Å2)	625	200	465	475	126
B11 (Å2)	-3.2	--	-3.8	-3.1	-1.6
B22 (Å2)	5.7	--	3.6	4.1	1.7
B33 (Å2)	-2.5	--	0.1	-1.0	-0.1
B13 (Å2)	148.1	--	0	0	0
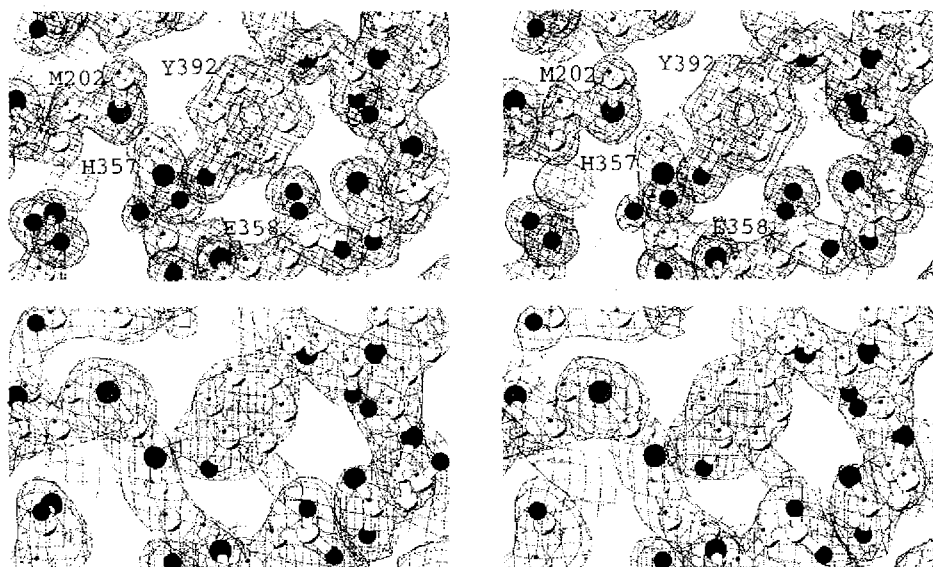
Figure 7. Electron density: P2$_1$ vs P2$_1$2$_1$2$_1$ (a-top) Stereo view of representative electron density in space group P2$_1$2$_1$2$_1$. Coefficients are 2F$_o$-F$_c$ where F$_o$ and F$_c$ are the observed and calculated structure amplitudes. The map is calculated at 1.7Å resolution and contoured at 1σ. The refined model is superimposed. (b-bottom) Same view as above for the refined structure in space group P2$_1$. The resolution is 2.5Å. Figure prepared with MOLSCRIPT (Kraulis, 1991).

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 the new (orthorhombic) crystal form is described in Methods and in Table 1. The use of flash freezing permitted data collection and refinement to high resolution (1.7 Å). This model serves as the basis for the detailed description of the structures. It includes residues 13-1023 for each of the four chains, several Mg^{++} and Na$^+$ ions, 4424 water molecules and 112 dimethylsulfoxide (DMSO) molecules. The electron density (Figure 7(a)) 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 wildtype (see 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[5]. It was verified with a capillary electrophoresis assay [6] that the modified enzyme has the expected transglycosylase activity (data not shown). The R-factor is 15.7% and R-free is 21.1% (Table 2), with a coordinate error of about 0.15Å as judged by a Luzatti plot. Figure 8 shows the Ramachandran plot for the tetramer.

The high resolution refinement in the new crystal form confirms the structure as originally described.[1] Including the constrained refinements, there are then 26 models for a β-galactosidase monomer, which differ by either space group, temperature, or refinement protocol (constrained *versus* unconstrained) (Table 3). The agreement of C^{α} positions in the 325 possible monomer-monomer comparisons varies from 0.2Å to 0.7Å. The best agreements (0.2-0.3Å) are between two monomers in the 1.7Å model, between the constrained models and their own descendants, and between the two constrained models. In the range from 0.3 to 0.4Å are comparisons between monomers in the same crystal form at the same temperature, and between a constrained model in one crystal form and the monomers from the other crystal form at room temperature. The monomers having poorest agreement (0.4-0.7Å) are from the non-constrained models in different crystal forms and at different temperatures. Taken together, these data suggest that the freezing process has a larger effect on the enzyme conformation than packing differences between crystal forms.
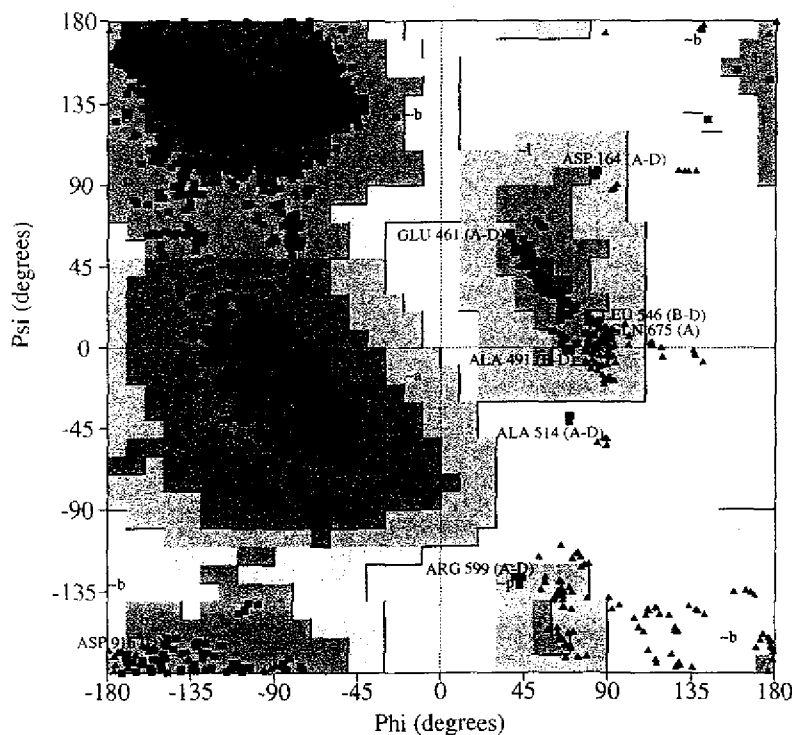
Figure 8. Ramachandran diagram for the non-constrained model of the β-galactosidase tetramer in space group $P2_12_12_1$. Glycine residues are indicated by triangles and non-glycines by squares. According to the criteria of Laskowski et al. (1993)[7] 87.7% of the residues are in most favored regions, 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. Glu-461 and Arg-599 are active site residues. Ala-514 is at the apex of a tight turn which packs against the donated loop in the activating interface and Ala-491 is in the (missing) helix 5 of the $(\alpha/\beta)_8$ barrel (Figure 11(c)). There are no distinctive structural elements associated with residues D164, L546 or D916.

Table 3. Monomer-monomer C_α rmsd (Å). Each value is an average over all monomer-monomer comparisons within each set. $P2_1$ = 16 monomers. $P2_12_12_1(LT)$ = 4 low temperature monomers. $P2_12_12_1(RT)$ = 4 room temperature monomers. $P2_1(Con)$ = $P2_1$ model constrained by non-crystallographic symmetry(NCS). $P2_12_12_1(Con)$ = $P2_12_12_1$ model constrained by NCS.

	$P2_1$	$P2_12_12_1(LT)$	$P2_12_12_1(RT)$	$P2_1(Con)$	$P2_12_12_1(Con)$
$P2_1$	0.32	0.55	0.41	0.27	0.34
$P2_12_12_1(LT)$		0.32	0.47	0.55	0.43
$P2_12_12_1(RT)$			0.34	0.41	0.24
$P2_1(Con)$				0.0	0.26
$P2_12_12_1(Con)$					0.0

Structure of the Monomer

Most of the 1023 residues that form the β-galactosidase monomer form five well-defined structural domains [1,8] (Figure 9). 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 and also interacts with the equivalent amino-terminal portion of the chain from a neighboring monomer. Both the inter- and intra-chain contacts appear to be important in forming one of the dimer interfaces that make up the tetramer.

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 site also includes portions of loops from the first, second and fifth domains of the monomer.
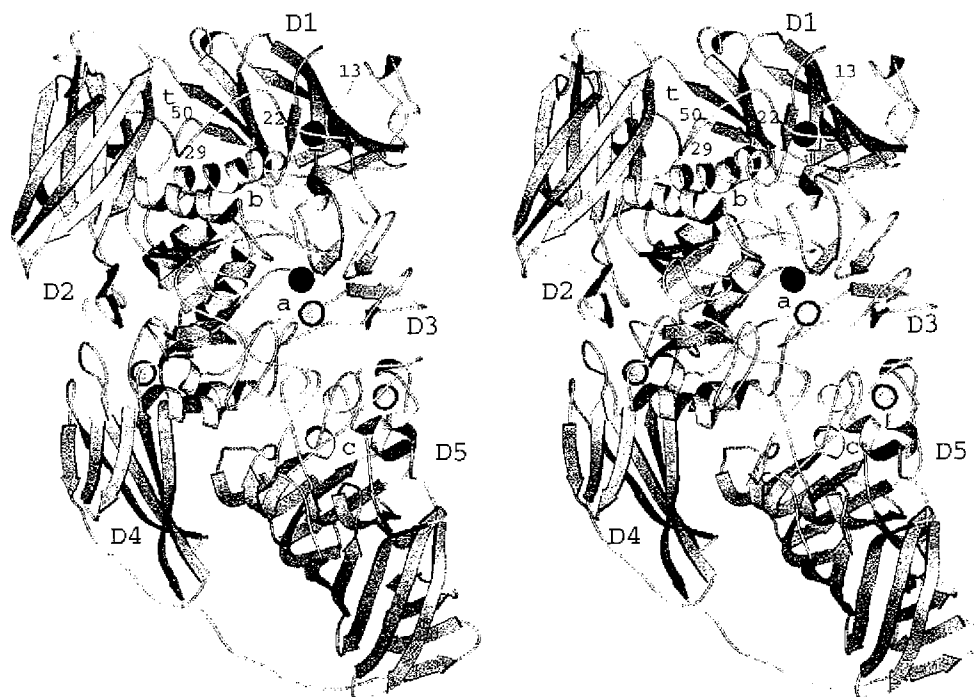
Figure 9. Stereo drawing illustrating the fold of the monomer. Three of the solvent molecules identified as Na$^+$ ions are shown as light-gray spheres. The magnesium ion identified at the active site (labeled *a*) is drawn as a dark sphere as is another tightly-bound ion, also presumed to be Mg^{++} (see text). Additional putative Na$^+$ and Mg^{++} 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 which extends roughly to residue 50. The hole through which this peptide is threaded is labeled *t*. The pair of helices which 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).

Structure of the Tetramer

The tetramer has 222-point symmetry (Figure 10). It is roughly ellipsoidal, with

dimensions 175 x 135 x 90 Å along the two-fold axes. There is a continuous system of

channels running along the surface (Figure 10(b)) and within the tetramer (Figures 10(c),

10(d)). These channels appear to be accessible to bulk solvent and vary in width from 5-

20Å. The active sites are located at the bottom of such surface channels (Figure 10(b)).

Figure 10. Several views of the tetramer. (a) (Following page). View of the tetramer looking down one of the two-fold 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.

Activating Interface

E580

Monomer A

Active Site
Monomer D

Long Interface

D233

Monomer B

Monomer C

E580

38

39
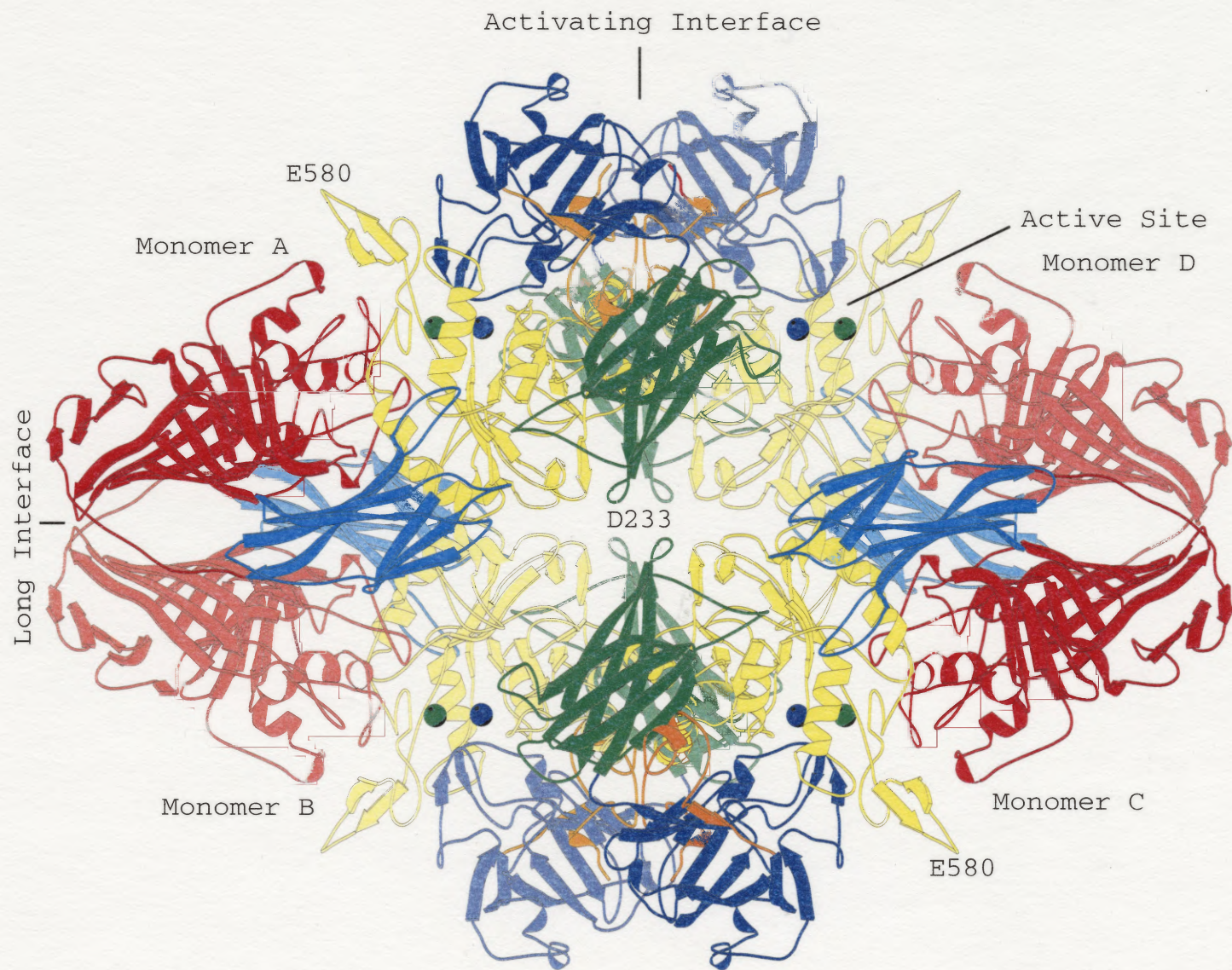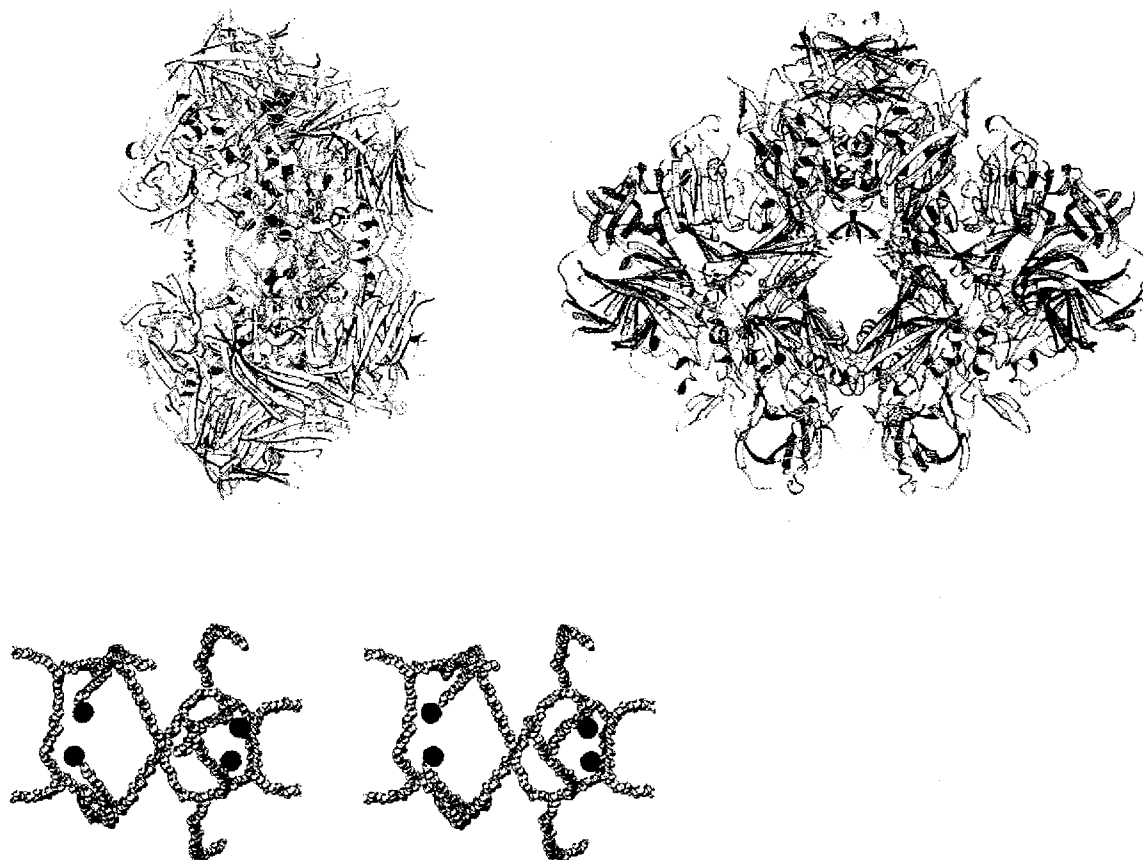
Figure 10 (b-upper left). A view of the tetramer showing one of the deep channels that runs across the surface. Part of the way along the channel residues Asp-233 from different subunits extend out and essentially touch each other. Two active sites, highlighted by the blue and green spheres (cf. Figure 10(a)), are at the bottom of this channel.

Figure 10 (c-upper right). Another view of the molecule showing one of the solvent-filled channels that passes through the tetramer.

Figure 10 (d-bottom). 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.

Subunit Interfaces and Formation of the Active Site

There are two principal subunit interfaces (Figure 11). The "long" interface buries about 4000Å^2 (48% polar atoms) and the "activating" interface buries 4600Å^2 (45% polar). There is also a third interface in the tetramer which is much smaller, burying 230Å^2 (75% polar).

The long interface (Figure 11(a)) 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 (Figure 11(b)) 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(Figure 11(c)). 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 is formed by interactions involving Domain 3 and the complementation peptide. Interactions involving the complementation peptide are more polar (~65% *versus* ~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% *versus* ~45%).

Figure 11. The dimer interfaces. (Following page) (a-top). View of the "long" interface. The β-galactosidase tetramer is viewed perpendicular to the long interface with the two subunits closer to the viewer removed so as to expose the interface. The remaining two subunits shown in the figure are A and D. Atoms that form part of the long interface are shown as spheres and shown in somewhat brighter coloring. They cluster into two patches and are mainly from Domains 3 (yellow) and 5 (red). The color coding is as in Figure 10. The junction between the two subunits shown here constitutes part of the activating interface which is shown "face-on" in the following panel. (b-bottom). "Face-on" view of the "activating interface" (cf. Figure 11(a)). In this case the two subunits that are included are A and B. The atoms that contribute to the activating interface (shown as brightly-colored spheres) are mainly from Domains 2 (green) and 3 (yellow), as well as the complementation peptide (orange).

Activating Interface

Long

D5

D3

D3

D5

Long

Complementation Peptide

Activating

D2

D3

Long Interface

D3

D2

Activating

Complementation Peptide
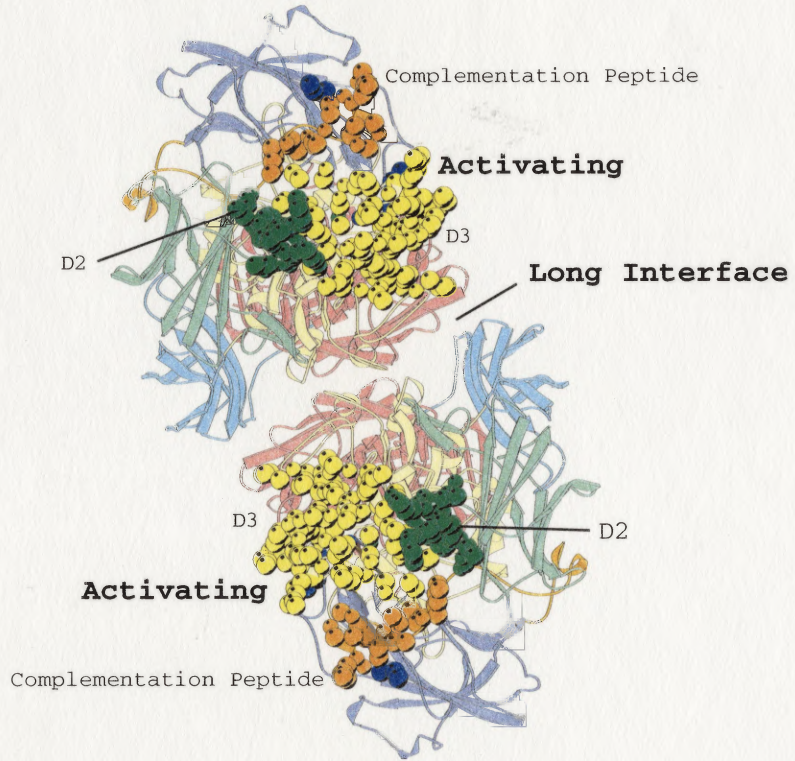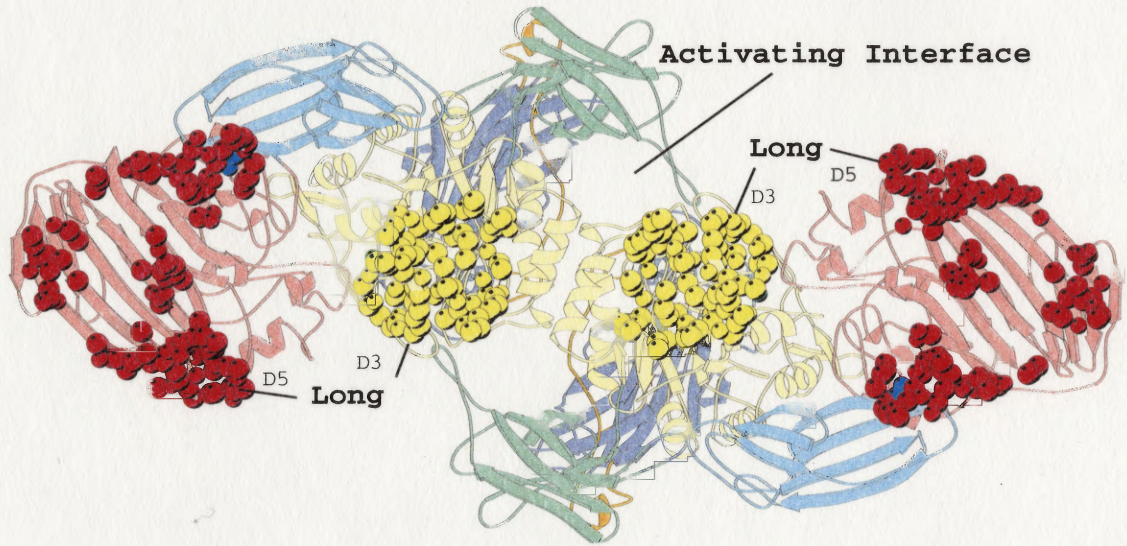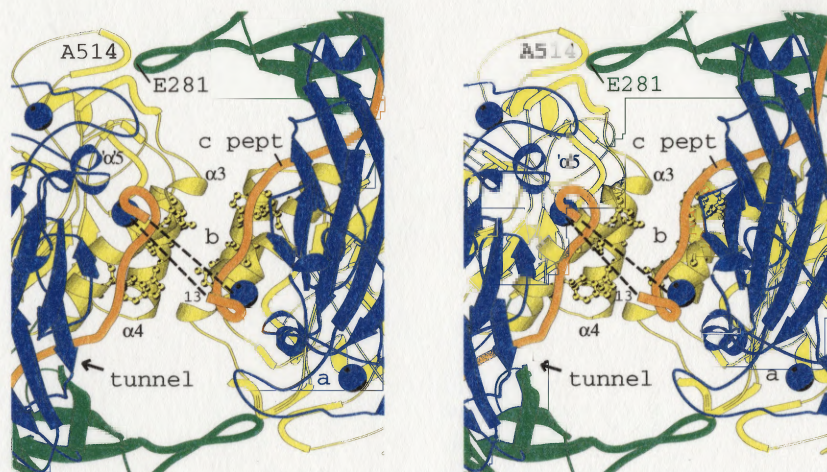
Figure 11(c). Stereo view of the activating interface in the vicinity of the four-helix bundle (yellow) and the complementation peptide (orange). The direction of view is essentially the same as in Figure 11(a). The four-helix bundle is formed by helices 3 and 4 of the $(\alpha/\beta)_8$ barrels of the subunits that form the interface. Also the loop that includes Glu281 (in green) extends across the interface and occupies a position labeled 'α'5 which corresponds to the "missing" helix 5 of the $(\alpha/\beta)_8$ barrel. The complementation peptide, labeled "c pept", starts at residue Arg-13 (near the letter b) and extends through a tunnel made by Domains 1-3. Close to the beginning of the complementation peptide there are two salt bridges, shown as dashed lines, that connect one subunit with the other. The two-fold axis of symmetry passes between the two salt bridges and relates one with the other. The two subunits have been separated slightly for clarity. Figure prepared with MOLSCRIPT.

Solvent Structure

In the orthorhombic model there are approximately 4900 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 3.5Å of polar protein atoms (Table 4) suggesting that they are "first shell" (i.e. interacting directly with the protein). 10% of the water molecules lie beneath the surface of the protein, and all of these make polar contacts with protein and/or solvent. (Each subunit has two buried

water molecules that make polar contacts only to other water molecules - they are in a buried water cluster.) The buried water molecules satisfy about 1300 hydrogen bonds for the protein (Vriend, 1990). The remaining 4000 water molecules make about 4800 hydrogen bonds to protein, while there are about 4700 protein-protein hydrogen bonds. Thus 55% 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 Arg-721, there are five linked five-membered water rings forming part of a clathrate-like structure. The outside of this "clathrate" forms polar contacts to solvent and to Arg-721 (Figure 12). 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 perhaps some adduct involving His-878.

There are 112 presumed DMSO molecules in the model (Table 4), 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 carbonyl groups making polar contacts.

Table 4. The environments of the solvent and other atoms associated with the tetramer of β-galactosidase in the $P2_12_12_1$ crystal form. Water molecules are segregated into different categories depending on whether they are at the solvent-exposed surface of the protein ("shell"), are buried, or form bridging interactions between subunits or domains or at crystal contacts ("crystal"). The "shell" water molecules are further segregated into first shell, second shell, etc. For the DMSO molecules the contacts are segregated into those made by the oxygen atom and those made by carbon plus sulfur. The "polar contacts" are the average number of either polar protein atoms or water molecules within 3.5Å of each type of solvent atom. First shell molecules are within 3.5Å of polar protein atoms. Second shell molecules are within 3.5Å of first shell molecules. Buried molecules are those whose center is greater than 1.4Å from the molecular surface of the tetramer (protein only) as calculated with MSRoll (1.4Å probe). Bridging waters are less than 3.5Å from two polar atoms that are greater than 4.0Å apart.

| | Count | Average B-factor (Å^2) | Polar contacts | |
			Protein	Water
All solvent	4908	32	1.4	1.6
Water	4424	31	1.5	1.6
Shell				
First	3429	28	1.9	1.5
Second	925	38	0.0	2.4
Third	118	42	0.0	1.8
Rest	52	45	0.0	0.5
Buried	398	16	3.4	0.9
Bridging				
Crystal	58	29	2.7	1.6
Subunit	129	25	3.0	1.4
Domain	352	19	3.0	1.4
Na^+	20	24	3.0	2.5
Mg^{++}	16	23	2.8	2.8
DMSO	112	40		
O	112	37	1.1	1.3
S,C	336	41	0.6	0.6
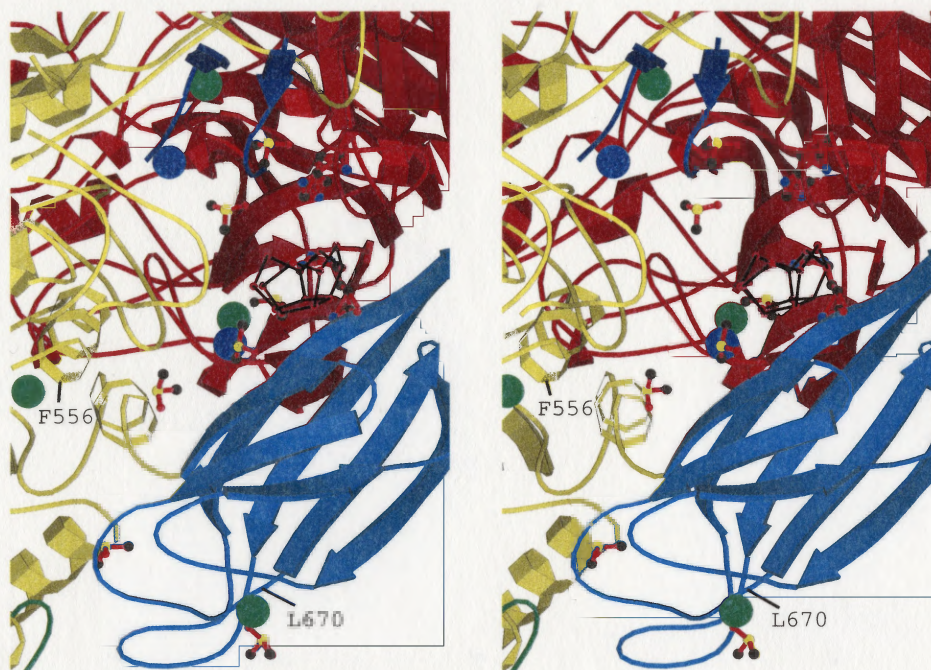
Figure 12(a). Stereo drawing showing the general location of the partial clathrate-like structure in a surface crevice at the long interface. Coloring is as follows: Domain 4 (monomer A), cyan; Domain 5 (monomer B), red; Domain 3 (monomer B), yellow. Putative cation binding sites are shown as colored balls (Mg^{++}, blue; Na^+, green; K^+, magenta). The active site is in the upper left, marked by its Mg^{++} and Na^+/K^+ sites. Also shown are several DMSO molecules.

Cavities

Figure 13 illustrates some of the characteristics of the cavities within the tetramer as well as those volumes that are inaccessible to bulk solvent but contain one or more water molecules. The most striking result is that the volumes that are most hydrophilic in character are occupied by solvent whereas those that are largely non-polar are empty (Hubbard et al., 1994). The volume of the largest such empty cavity, which is in Domain 5, is $113Å^2$. A partial 4.0 Å data set collected in the presence of iodoethane suggested that iodoethane binds in this cavity.
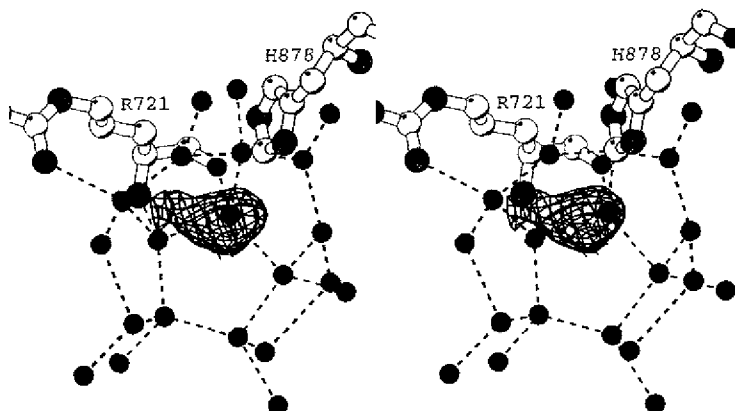
Figure 12(b). Close-up stereo view of the partial clathrate-like structure. Water molecules and protein oxygen atoms are shown as black spheres, nitrogen atoms light gray and carbon atoms as open circles. Presumed hydrogen bonds are shown as broken lines. The electron density is seen in a "residual" 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 +3.0σ where σ is the root-mean-square density throughout the unit cell. Figure prepared with BOBSCRIPT (Esnouf, 1999).

Cis-Peptide Bonds

There are seven apparent cis-peptide bonds in each monomer, at Asp-164, His-391 and Trp-568, as well at 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 Trp-568 are at the active site while Asp165 is about 15Å away forming a tight turn. Pro112 and Pro422 participate in the activating interface, Pro-112 interacts with the complementation peptide and Pro-422 with the donated loop. Pro-87 and 902 are in relatively solvent-exposed regions of Domains 1 and 5.

Cavity Volume vs Cavity Surface Hydrophilicity
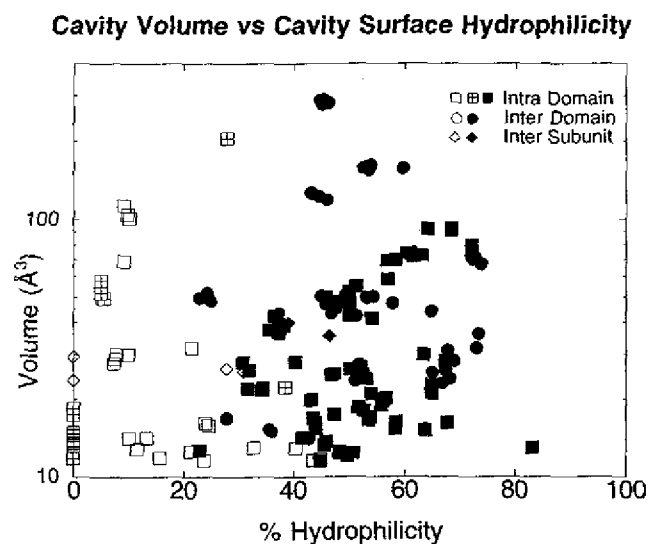
Figure 13. Plot summarizing the nature of both empty cavities and solvent-binding sites in the tetramer that do not have access to bulk solvent. Solid symbols indicate sites occupied by solvent, half-filled symbols indicate partially occupied sites and open symbols indicate cavities that appear to be empty. Squares indicate sites within domains; circles indicate sites between domains and diamonds indicate sites between monomers.

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 His-391, Asp-412, Arg-388 and Glu-537. 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 (Arg-786, Asp-792, Arg-881, Glu-934, Asp-987 and His-990) whose sidechains

are completely buried within the domain, occupying about 650 Å^2 (Figures 9, 14).

Most of these residues are conserved in enzymes homologous to β-galactosidase (e.g. see

Figure 3 of Jacobson et al., 1994).
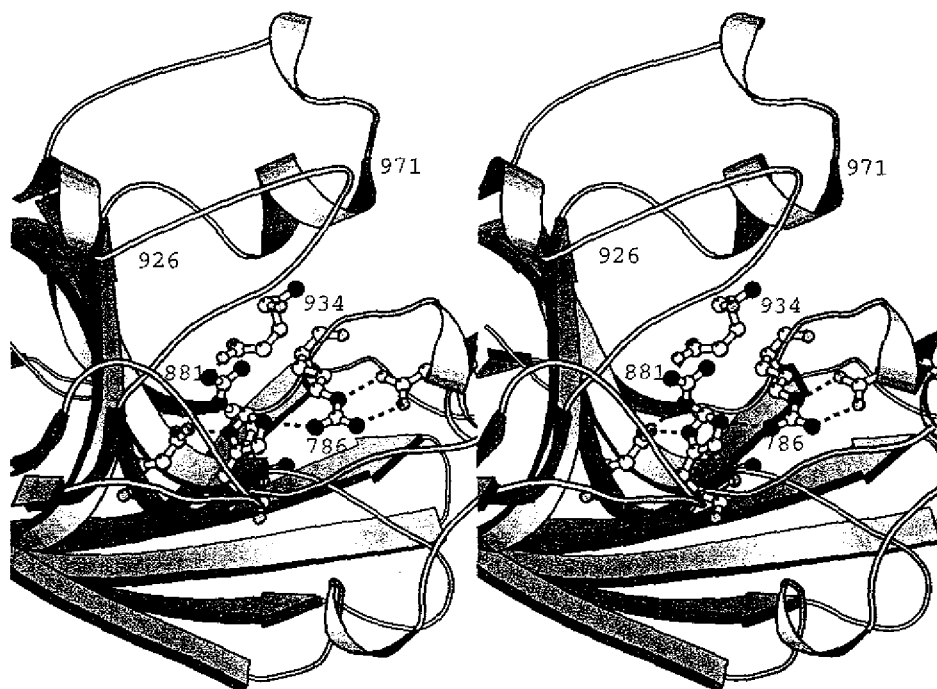
Figure 14. Stereo view of the polar network in the core of Domain 5 that includes Asp-792, Glu-934, Asp-987, Arg-786, Arg-881 and His-990. This network also interacts with other polar protein atoms and two buried water molecules. Figure 9 shows the location of this network within the monomer. Figure prepared with MOLSCRIPT (Kraulis, 1991).

Discussion

The structure of β-galactosidase from *E. coli* was originally determined to a

resolution of 2.5Å using the monoclinic crystal form with four tetramers in the

asymmetric unit [1]. The orthorhombic crystals described here have one tetramer per

asymmetric unit and higher crystallographic symmetry, allow for easier data collection. More important, these crystals diffract to higher resolution, which is helpful for detailed mechanistic studies.

The crystallization conditions for both forms make use of PEG 8000 as the precipitant at about the same concentration. The monoclinic crystals require a high concentration of β-mercaptoethanol (BME) (70-140 mM) and at least three cysteines appear to be derivatized with BME, one near a crystal contact. While the orthorhombic crystals do not require BME, they need 200 mM $MgCl_2$. Attempts at chelating the Mg^{++} with EDTA or exchanging it with Mn^{++} produced extremely mosaic or cracked crystals. This is consistent with octahedral centers (modeled as Mg^{++}-solvent clusters) at three crystal contacts.

Because the β-galactosidase molecule is large it provides an opportunity to evaluate the characteristics of the interactions that occur at different levels of association, i.e. between domains, monomers and tetramers.

Overall, the tetramer has about $135,000\text{\AA}^2$ of accessible surface area and buries about $18,000\text{\AA}^2$ at the subunit interfaces (Tables 5,6). 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 $\text{\AA}^2$ for oligomeric proteins follow the relationship $A_s = 5.3\ M^{0.76}$ [9]. β-galactosidase, however, shows a 27% discrepancy ($135,000\text{\AA}^2$ versus $106,000\text{\AA}^2$). In a similar study, the same authors propose a scaling law for monomeric proteins, $A_s = 6.3\ M^{0.73}$ [10]. They suggest that this should be applicable to subunits of oligomeric proteins which 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 (Table 5), 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 expected 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 10(b)-10(d) 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.

Table 6 shows that there is a decrease in buried hydrophobic 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.

Table 5. Characteristics of the surface area of the β-galactosidase tetramer and isolated monomers and domains. The five domains are designated D1-D5. If the solvent-accessible surface area of, for example, the complementation peptide is calculated in the context of a single monomer, then 52% of the surface area is buried. If the calculation is made in the context of the tetramer, 61% (i.e. an additional 9%) is buried.

| | Solvent-accessible surface area | | | | | |
	Total ($\mathring{A}^2$)	Polar (%)	Nonpolar (%)	Buried by monomer (%)	Buried by tetramer (%)	Volume ($\mathring{A}^3$)
Tetramer	134815	48	52	--	--	549312
Monomer	38109	48	52	--	11	136381
Complementation peptide (13-50)	4160	48	52	52	61	4542
D1 (51-219)	8449	43	57	35	35	22504
D2 (220-332)	6983	53	47	24	34	14230
D3 (333-626)	13581	47	53	38	54	39106
D4 (627-726)	6268	46	54	19	25	12868
D5 (735-1023)	13023	49	51	14	21	38451

Table 6. Characteristics of the surface area that forms the interfaces in the β-galactosidase crystals in space group $P2_12_12_1$ and between individual monomers and domains. The surface area buried in crystal interfaces was determined by calculating the surface area of an isolated tetramer and subtracting the surface area accessible within the crystal. The intradomain surface was determined by calculating the solvent-accessible surface area of the polypeptide chain in an extended conformation and subtracting the combined surface areas of the complementation peptide plus the five domains.

| | Nature of interface | | | | |
	Total area ($\mathring{A}^2$)	Polar surface (%)	Nonpolar surface (%)	H-bonds	H-bond density (H bonds/100$\mathring{A}^2$)
Crystal interfaces	6315	61	39	17	0.44
Subunit interfaces	17481	47	53	84	1.02
Domain interfaces	62551	44	56	370	1.34
Intradomain interface	444241	40	60	3636	2.05

It has recently been suggested that β-galactosidase folds according to the following mechanism [11]:

$$\begin{array}{ccccccccc} & fast & & slow & & slow & & fast & \\ 4M_u & \rightarrow & 4M & \rightarrow & 2D' & \rightarrow & 2D & \rightarrow & T \end{array}$$
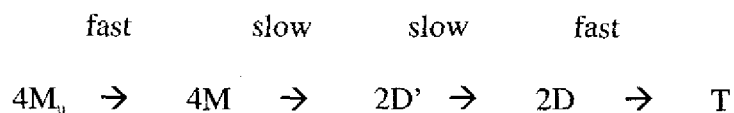
According to this scheme the unfolded monomer chains (M_u) first give folded monomers (M). There is then a slow bimolecular event to form dimers (D'). These then undergo a slow first order event to form dimers (D) competent for fast association to tetramers (T). Overall this is consistent with the hierarchical 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 interface 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[11, 12].

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[13, 14]. Also, deletions of residues 23-31 or 11-41 result in inactive dimers (called α–acceptors) which can be complemented by certain peptides (α-donors) to reconstitute the active tetramer [14, 15]. 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 15 is a sketch illustrating in highly simplified fashion the parts of β–galactosidase 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 3 domain junction. 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 bundle, both of which are within the same subunit (Figures 9, 11c, 15). There is a presumed magnesium ion that is coordinated by Asp15, Asn18, Val21, Gln163 and Asp 193. This ion therefore bridges between the complementation peptide and the rest of the protein (Figure 11(c)).
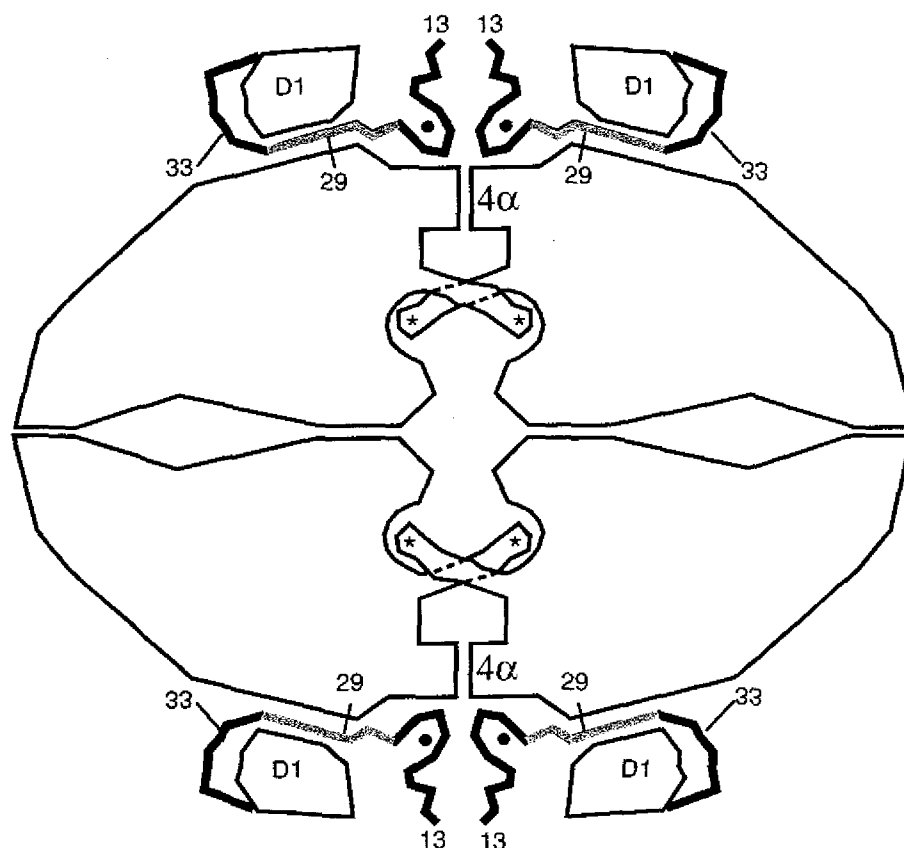
Figure 15. Sketch summarizing key features of the β–galactosidase tetramer and α–complementation. At the amino terminus, residues 1-12 are not seen in the electron density map due to presumed disorder. Residues 13-50 (shown as thick lines) pass through a tunnel between the first domain (labeled D1) and the rest of the protein (see also Figure 9). 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 activating 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 activating 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.

Studies of hybrids by Fowler and Zabin[5] 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 the deletion of residues 23-31 results in inactive dimers, residues are presumably critical for tetramer

formation. This is consistent with the structure illustrated in Figure 15. The lesser importance of residues 1-23 is consistent with the observation that parts of this region (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 Arg-13 and Asp-15). Residues 22-31 help stabilize the four-helix bundle which 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 (Figure 9). 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 Arg-431-Trp-432 peptide bond [16]. This can be rationalized in terms of protection of the peptide bond with both intrasubunit interactions (involving the complementation peptide) and intersubunit interactions (involving the activating interface).

As illustrated in Figure 15, 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 tetramers to dimers is synonymous with inactivation.

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 CNBr₂). 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 is threaded through the "tunnel" (Figures 9,15). 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 non-native amino acids would make is 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 wildtype 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 (Figure 15), making them susceptible to proteases and available for binding antibodies [12]. 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 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 Arg-14 and the backbone amide of Val-114 in a small intrasubunit polar network. Likewise, Trp-16, 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, Asp-15, Asn-18, and Val-21 presumably contribute to the binding of the complementation peptide via their coordination of the Mg^{++} ion that bridges to the rest of the protein (Figures 11(c), 15). This is also consistent with the observation that Mg^{++} stabilizes the complemented

protein[17].

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 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 in the tunnel as shown in Figure 15. This unusual arrangement may confer two advantages. First the intimate association presumably enhances 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^{++} and Na^+ are required for maximal activity of β–galactosidase[18]. Putative sodium ions 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 which 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 16. Because of its close proximity to the active site (Figure 9) it is highly likely that its removal would perturb this region and reduce activity. The role of this ion in the catalytic mechanism is discussed in Chapter 3. The other 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). The role of this ion is also discussed in Chapter 3. The

identities of the 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
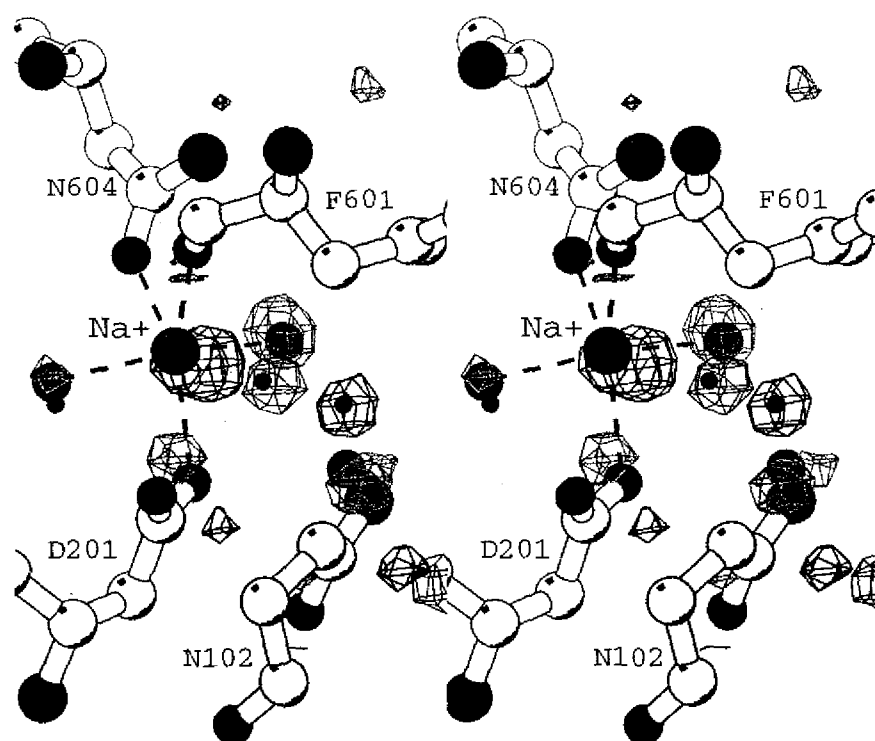
Figure 16. Stereo view of sodium binding site in the active site. The large sphere shows the sodium ion and the smaller spheres show the locations of refined water molecules in the potassium soak. Electron density is calculated from a map with coefficients $F_o(K^+)$ - F_o where the $F_o(K^+)$ are the 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 $+ 6\sigma$ (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).

contacts mediating intermolecular interactions with no direct contacts to protein. An

apparent ion-binding site in Domain 1 (Figure 9) was not affected by the presence of

potassium, rubidium, or EDTA, suggesting that the putative ligand at this site, which has

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+[16].

This is consistent with the former bond being solvent-exposed and mobile and the latter

bond being close to the Na+ binding site (Figure 16). Likewise, elastase cleaves the

Ala732-Ala733 peptide bond [16] 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.[19]. 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 P21 was as described[20].

Protein used for crystallizing the P212121 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 sulphate 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), 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. This is the pET

15b 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 wildtype

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

hours. 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 minutes,

the cell lysate was loaded on a nickel column (Qiagen). 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

versus 2x4 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) or anion exchange

chromatography (Biocad). The digested protein was further purified with anion exchange

chromatography. The best results were obtained with a PI column (Perceptive

Biosystems) 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 [20]. A number of other crystal forms were identified, some 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 x 0.7 x 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 hours or more. Crystals could then be flash frozen in a cold stream with only small effects on the mosacity and diffraction.

Model Building and Refinement, Space Group $P2_1$

An initial model of one monomer was built 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 27.5Å^2 the initial R-factor was 38% for data at 4Å resolution.

All refinement was done with the TNT package of programs [2, 21, 22] After rigid-body refinement, positional refinement was initiated using constrained NCS (non-

crystallographic symmetry). For this procedure the prototype model is expanded using the NCS operators derived from rigid-body refinement. Gradients (and curvatures in later cycles) are then calculated for all atoms using diffraction data and stereochemical restraints. After the gradients are combined using the chain rule, again with the NCS operators, shifts are calculated and applied to the prototype molecule.

This constrained NCS refinement was carried out on the atomic positions using data from 8.0 to 3.5Å, 3.0Å, and finally 2.5Å 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 have a model with 16-fold constrained NCS. Rebuilding such a model would require inspection of fewer residues (1000 *versus* 16,000) and would also provide a much more favorable ratio of observations to parameters. So refinement was continued with constrained NCS using all data to 2.5Å. Additionally, a random set of 1600 reflections was removed for R_{free} calculations, which were used to help determine the best refinement protocol (see below).

Prior to B-factor refinement, several rounds of positional refinement, model building and solvent addition were carried out and 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 [21]. 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 sidechains modeled

with two conformations, and three cysteines derivatized by β-mercaptoethanol. The

entire model was constrained by the non-crystallographic symmetry.

It might be noted that initial attempts at B-factor refinement using all data to 2.5Å

resolution were unsuccessful in that the B-factors of many interior atoms decreased to
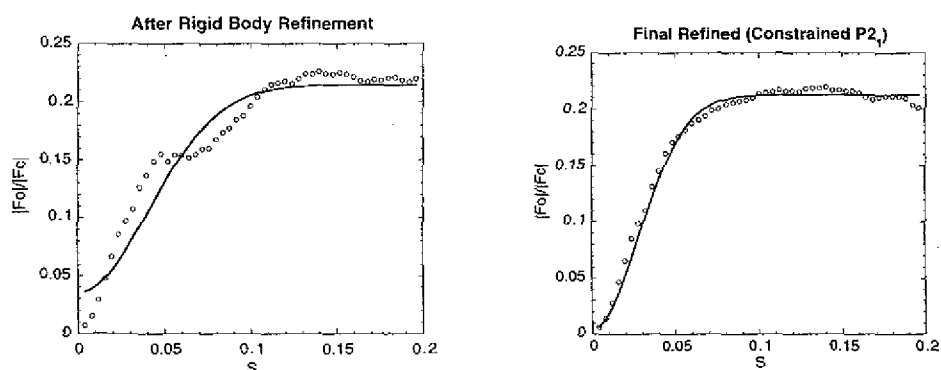
Figure 17. Behavior of the scaling profile used to account for the scattering of bulk
solvent. The model for bulk solvent used in TNT is based on Babinet's principle which
states that, at low resolution, the scattering of the bulk solvent is the inverse of that from
the protein. This leads to a scaling function for F_o which is defined by the functional

form:
$$\frac{1}{K}e^{-B(\frac{\sin\theta}{\lambda})^2}\,[1-K_{sol}e^{-B_{sol}(\frac{\sin\theta}{\lambda})^2}]$$

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 set at
27.5Å². 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.

near zero. Also the scaling profile of F_o to F_c was not well fit by the solvent model

employed by TNT (Figure 17(a)). 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 8.0Å and 2.5Å

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 17(b).

Data Collection, Space Group $P2_12_12_1$

Using non-frozen crystals, data were initially collected at Daresbury to 2.8Å

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 second

exposures, 0.5° oscillations and a wavelength of 1.0Å. Reflections were visible to 1.5Å,

and data were processed to 1.7Å resolution with Mosflm/Scala [23-25].

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 two-fold axis was parallel to a crystallographic 2_1

screw axis. The self-rotation function also suggested two pairs of perpendicular two-fold

axes in the xz plane rotated approximately 25° (+ or -) about the y axis.

Molecular replacement was based on the 2.8Å 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 [26, 27], the rotation search gave peaks which 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 non-crystallographic symmetry. Several rounds of model inspection, solvent addition, and minimization resulted in a model with an overall R-factor of 16.8% at 2.8Å 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 $P2_1$ model with thermal factors set to the Wilson B of $17Å^2$. After rigid-body refinement each monomer was refined independently using all data to 1.7Å 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 sidechains. By the time that each chain had been rebuilt once,

approximately 2500 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 2500 more solvent molecules [28]. Approximately 1000 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 1.7Å 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

Calculations on the refined coordinates were carried out using a variety of programs. EDPDB [29] was used for coordinate manipulations, solvent accessible surface area calculations, and to generate crystal contacts. MSRoll [30] was used for volume and molecular surface calculations. Cavity calculations were also performed with MSRoll and were supplemented with INSIDE_MSP (M. Quillin, unpublished) to determine the locations of water molecules relative to cavities. All surface area calculations used a probe of radius 1.4Å. Whatif [31] was used to determine hydrogen bonds. Calculations to determine bridging waters and contact atoms, to create extended models of the domains, and to rename water molecules were carried out with unpublished programs by the authors.

References

1.   Jacobson, R.H., Zhang, X.-J., DuBose, R.F. & Matthews, B.W. Three-dimensional structure of ß-galactosidase from *E. coli*. *Nature* **369**, 761-766 (1994).

2.   Tronrud, D. TNT refinement package. in *Methods in Enzymology* Vol. 277 306-319 , 1997).

3.   Kleywegt, G. & Brunger, A. Checking your imagination: applications of the free R value. *Structure* **4**, 897-904 (1996).

4.   Stout, G. & Jensen, L. *X-ray Structure Determination, A Practical Guide, 2nd Edition*, (John Wiley & Sons, New York, 1989).

5.   Fowler, A. & Zabin, I. Purification, structure, and properties of hybrid ß-galactosidase proteins. *J. Biol. Chemistry* **258**, 14354-14358 (1983).

6.   Zeleny, R., Altmann, F. & Praznik, W. A capillary electrophoretic study on the specificity of beta-galactosidases from Aspergillus oryzae, Escherichia coli, Streptococcus pneumoniae, and Canavalia ensiformis (jack bean). *Anal Biochem* **246**, 96-101 (1997).

7.   Laskowski, R., MacArthur, M., Moss, D. & Thornton, J. PROCHECK: A program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **26**, 283-291 (1993).

8.   Juers, D.H., Huber, R.E. & Matthews, B.W. Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between ß-galactosidase and other glycohydrolases. *Protein Science* **8**, 122-136 (1999).

9.   Miller, S., Lesk, A.M., Janin, J. & Chothia, C. The accessible surface area and stability of oligomeric proteins. *Nature* **328**, 834-836 (1987).

10.   Miller, S., Janin, J., Lesk, A.M. & Chothia, C. Interior and surface of monomeric proteins. *J Mol Biol* **196**, 641-656 (1987).

11.   Nichtl, A., Buchner, J., Jaenicke, R., Rudolph, R. & Scheibel, T. Folding and association of ß-galactosidase. *JMB* **282**, 1083-1091 (1998).

12. Zabin, I. ß-Galactosidase alpha complementation. *Molecular and Cellular Biochemistry* **49**, 87-96 (1982).

13. Muller-Hill, B. & Kania, J. Lac repressor can be fused to beta-galactosidase. *Nature* **249**, 561-563 (1974).

14. Ullmann, A. Complementation in β-Galactosidase: From Protein Structure to Genetic Engineering. *BioEssays* **14**, 201-205 (1992).

15. Ullman, A., Jacob, F. & Monod, J. Characterization by *in vitro* Complementation of a Peptide corresponding to an Operator-proximal Segment of the ß-Galactosidase Structural Gene of *Escherichia coli. JMB* **24**, 339-343 (1967).

16. Edwards, L.A., Tian, M.R., Huber, R.E. & Fowler, A.V. The use of limited proteolysis to probe interdomain and active site regions of beta-galactosidase (Escherichia coli). *J Biol Chem* **263**, 1848-1854 (1988).

17. Gallagher, C.N. & Huber, R.E. Stabilities of uncomplemented and complemented M15 beta-galactosidase (Escherichia coli) and the relationship to alpha-complementation. *Biochem Cell Biol* **77**, 109-118 (1999).

18. Wallenfels, K. & Weil, R. Beta-Galactosidase. in *The Enzymes (VII)* (ed. Boyer, P.) , 1972).

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

20. Jacobson, R. & Matthews, B. Crystallization of β-Galactosidase from *Escherichia coli. Journal of Molecular Biology* **223**, 1177-1182 (1992).

21. Tronrud, D. Knowledge-based B-factor restraints for the refinement of proteins. *Journal of Applied Crystallography* **29**, 100-104 (1996).

22. Tronrud, D. Unpublished Results. .

23. Kabsch, W. Evaluation of single X-ray diffraction data from a position sensitive detector. *Journal of Applied Crystallography* **21**, 916-924 (1988).

24. Leslie, A. in *Crystallographic Computing* (Oxford University Press, 1990).

25. Evans, P. Data Reduction. in *Proceedings of a CCP4 Study Weekend on Data Collection and Processing* 114-122 , 1993).

26.   Zhang, X.-J. & Matthews, B. Enhancement of the method of molecular replacement by incorporation of known structural information. *Acta Crystallographica* **D50**, 675-686 (1994).

27.   Tong, L. & Rossmann, M. The locked rotation function. *Acta Crystallographica* **A46**, 783-792 (1990).

28.   Lamzin, V. & Wilson, K. Automated refinement of protein models. *Acta Crystallographica* **D49**, 129-147 (1993).

29.   Zhang, X.-J. & Matthews, B. EDPDB: A multi-functional tool for protein structure analysis. *Journal of Applied Crystallography* **28**, 624-630 (1995).

30.   Connolly, M. The molecular surface package. *Journal of Molecular Graphics* **11**, 139-141 (1993).

31.   Vriend, G. A molecular modeling and drug design program. *Journal of Molecular Graphics* **8**, 52-56 (1990).

CHAPTER III

STUDIES ON THE REACTION MECHANISM OF BETA-GALACTOSIDASE

Diffusion of the Substrate to the Active Site

Before an enzyme can carry out its activity, the substrate must find the active site.

This occurs via diffusion, which is characterized by a second order rate constant

describing the frequency of collisions between the substrate and enzyme. This is

generally between 10^9 and 10^{11} s^{-1} M^{-11} For β–galactosidase, k_{cat}/K_m for hydrolysis is

$\sim 10^5$ s^{-1} M^{-1} for lactose and 10^7 for onpg, suggesting there are typically ~ 100-10,000

collisions between the enzyme and the active site before a reaction takes place.

Imaging Reaction Coordinate Complexes with X-Ray Crystallography

Once the substrate finds the active site, β–galactosidase catalyzes the hydrolysis

or transglycosylation in about 15 ms. In the ideal experiment, a single enzyme molecule

would be imaged at atomic resolution performing the catalysis in real time. This would

be repeated many times to build up statistics on the reaction pathway.

Although some spectroscopic experiments have been performed with single

enzyme molecules,[2, 3] these are typically done with longer wavelengths than the X-rays

required for atomic resolution. Because X-rays interact relatively weakly with proteins, many molecules in an ordered lattice are necessary to amplify the signal/noise. A typical β–galactosidase crystal has ~10^{13} molecules.

For the amplification effect to be useful, all the active sites must be synchronized in their catalysis. Methods have been devised to allow certain enzyme reactions to be synchronously initiated throughout a crystal with nanosecond laser pulses. However, once the reactions have begun, there will be a distribution of trajectories. Unless this distribution is very narrow, the resulting electron density maps are typically very difficult to interpret. This is particularly true if there are significant motions of the substrate and enzyme during the reaction. In addition to the difficulties in synchronization, data collection within submillisecond time scales is also problematic, although some experiments have been performed[4].

The most straightforward method of recording atomic resolution images relevant to an enzyme catalyzed reaction is to use altered substrates, enzymes, or buffer conditions that slow down or stop the reaction. In this way the active sites pause in the same conformation long enough either to flash freeze the crystal (minutes) or to collect the diffraction data at room temperature(minutes-hours).

Regardless of the exact experimental design, the substrate of interest is soaked into the crystal, diffusing throughout it, occupying all of the active sites. Critical to the success of these methods is first the observation that protein crystals are intimately dependent on solvent. Generally 25-75 % of the volume of the crystal is occupied by disordered solvent, occurring in channels which run throughout the crystal. This allows the substrate to diffuse throughout the crystal and occupy all of the active sites. Second, the enzyme

must be able to catalyze its reaction in the crystal. Sometimes lattice packing interactions either prevent access to the active site or preclude motions required for catalysis or the buffer conditions needed for crystal stability are incompatible with activity. However, in both crystal forms of β–galactosidase used in this study the enzyme is active in the crystal This is shown especially clearly by the use of the substrate X-gal, which turns blue within the crystal (Figure 18).
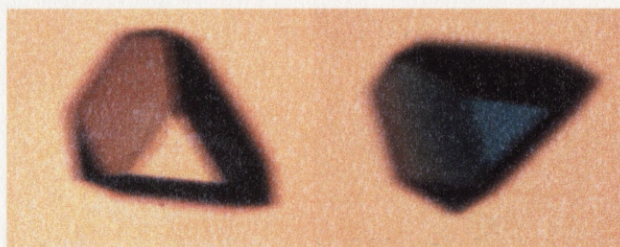
Figure 18. Crystal of β-galactosidase ($P2_12_12_1$) in the absence (left) and in the presence (about 2 hours) (right) of the substrate X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside). X-gal turns blue upon hydrolysis, showing the enzyme can perform hydrolysis in the crystal. A capillary electrophoresis assay also showed that in solution the enzyme can produce allolactose in the crystallization buffer (not shown).

Once all of the active sites are occupied by substrate, the diffraction data can be collected. An example of a diffraction image for β–galactosidase is shown in Figure 19.

1.5 Å
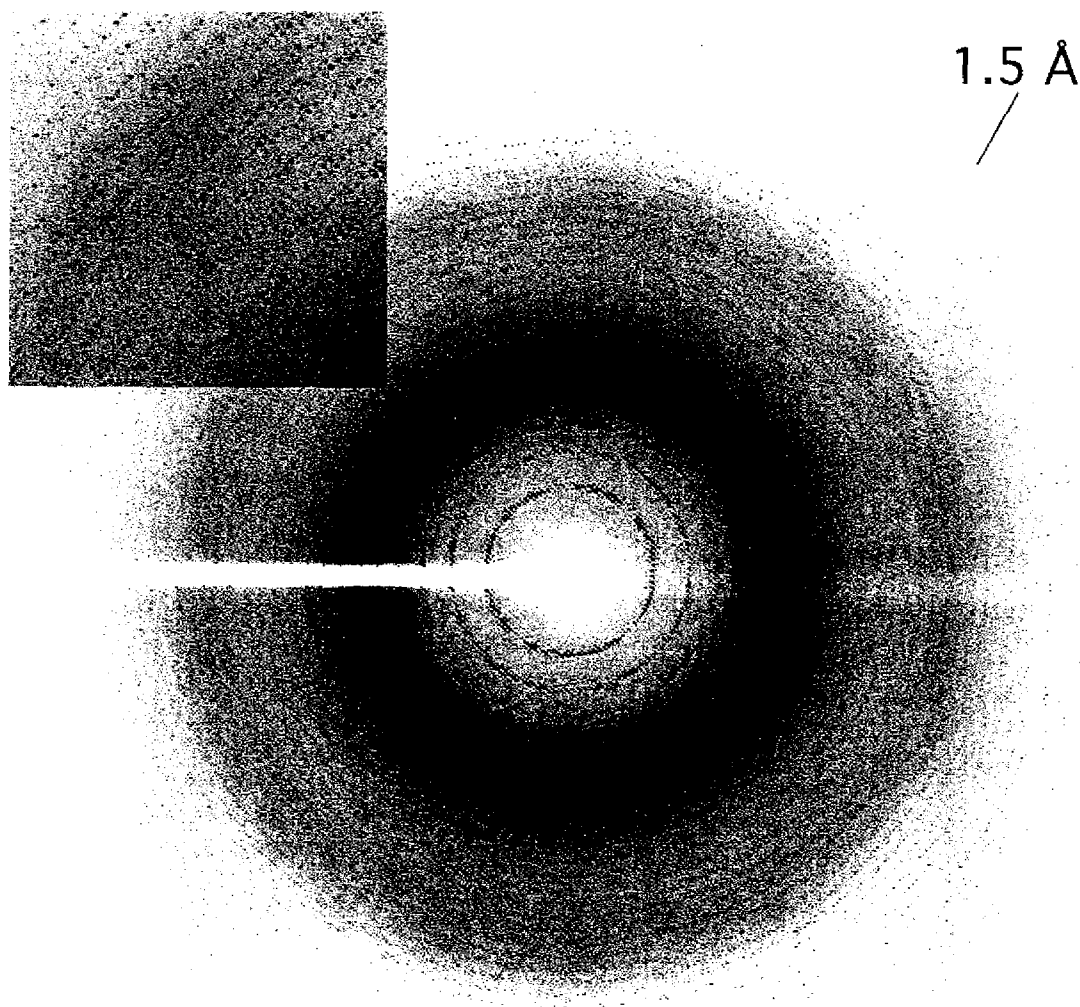
Figure 19. A diffraction image from an orthorhombic crystal of β-galactosidase. The crystal (the E537Q variant with allolactose) was exposed for 15 seconds at the Stanford Synchrotron Radiation Laboratory Beam Line 9-2 with the beam roughly parallel to the b* axis. The detector edge is at 1.5 Å resolution. There are about 30000 Bragg spots on the image and about 30 levels of reciprocal space can be discerned as the concentric rings. Longer exposures (60 seconds) of the same crystal produced spots up to 1.1 Å resolution. The row of spots in the blowup in the upper left corner shows that the long edge (c*) runs diagonal to the upper right and the short edge (a*) runs diagonal to the upper left.

The intensity of each Bragg spot can be used as a coefficient in a Fourier series to give an image of the electron density:

$$\rho(\vec{r}) = \sum_{k} F(\vec{k}) e^{i\phi(\vec{k})} e^{2\pi i \vec{r} \cdot \vec{k}}$$

where ρ is the electron density at a point $\vec{r}$, the sum is over reciprocal lattice vectors $\vec{k}$, $F(\vec{k})$ is the modulus of the intensity and $\phi(\vec{k})$ is the phase. The amplitude, F, can be measured for each Bragg reflection, but the phase, ϕ, associated with it is lost in the diffraction experiment and it must be estimated by other means. In the cases presented here, the structure of the unliganded, native enzyme is used to calculate an estimate, $\phi_c(\vec{k})$, for each phase:

$$\tilde{F}_c(\vec{k}) = F_c(\vec{k}) e^{i\phi_c(\vec{k})} = \int d\vec{r} \rho^n(\vec{r}) e^{2\pi i \vec{r} \cdot \vec{k}}$$

where ρ is the electron density derived from the coordinates.

To determine the location of a bound ligand, typically two diffraction experiments are performed, one without ligand ("native") and another with ligand. Difference electron density can then be calculated:

$$\rho^l(\vec{r}) - \rho^n(\vec{r}) = \sum_{k} \left[F_o^l(\vec{k}) - F_o^n(\vec{k}) \right] e^{i\phi_c^n(\vec{k})} e^{2\pi i \vec{r} \cdot \vec{k}}$$

Here the superscript refers to liganded (l) or native (n) and the subscript refers to

observed (o) or calculated (c). Such electron density is called "F_o-F_o" density because

both sets of coefficients come from observed data. Two examples are shown in Figure 20.
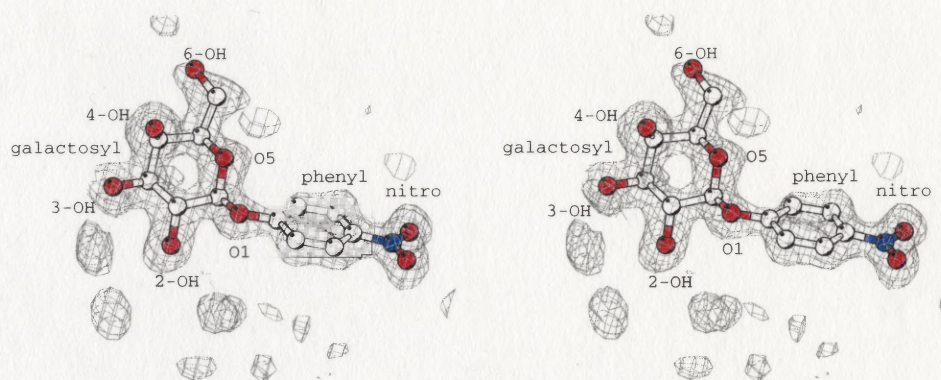
Figure 20(a). Stereo view of 1.7 Å F_o-F_o electron density for para-nitrophenyl-β-D-galactoside (pnpg) bound to the E537Q variant. Only positive density is shown, contoured at 4 sigma (0.24 electrons/Å^3). The binding mode of the ligand is very clear from the density and a ball-and stick model for the refined coordinates is also shown. Oxygens are red, carbons are white and nitrogens are blue.
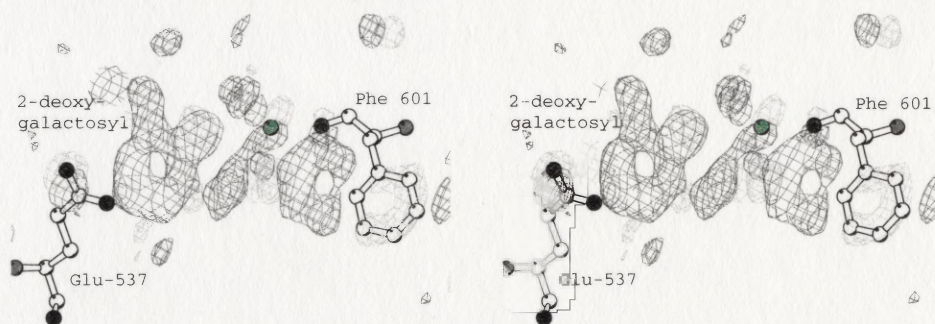
Figure 20(b). Stereo view of 1.7 Å F_o-F_o electron density for the 2-deoxy-galactosyl intermediate (with native enzyme). The map is contoured at +/- 6 sigma (0.36 electrons/Å^3) with black positive and gray negative and native protein coordinates are shown. This map is more complex than Fig 20(a) because of changes in the enzyme. Density for the 2-deoxy-galactosyl moiety can be seen next to the nucleophile, Glu-537, with a covalent bond between the two. Phe-601 clearly swings to the left. More subtle features can also be discerned, including shifts of Glu-537 and the sodium ion.

The crystal may be thought of as a convolution between a single molecule and a lattice:

$$\rho(\vec{r}) = m(\vec{r}) \otimes \sum_n \delta(\vec{r} - \vec{x}_n)$$

where $m(\vec{r})$ is the electron density for a single protein molecule, $\rho(\vec{r})$ is the electron density for the whole crystal, and the sum is over real space lattice vectors $\vec{x}_n$. The diffraction experiment will give information about the Fourier transform of the crystal. But the Fourier transform of a convolution is the product of the Fourier transforms:

$$FT(\rho(\vec{r})) = M(\vec{k}') \cdot \sum_n \delta(\vec{k}' - \vec{k}_n)$$

where $\vec{k}'$ is an arbitrary vector in reciprocal space, $M(\vec{k}')$ is the Fourier transform of the single molecule (the molecular transform), and the Fourier transform of the real space lattice gives another lattice in reciprocal space, which is defined by the vectors $\vec{k}_n$. Therefore, this second lattice can be thought of as a mask placed over the molecular transform, defining a sampling of the molecular transform which is recorded by the diffraction experiment.

F_o-F_o electron density is only useful if the lattice parameters for both crystals are nearly identical. Otherwise, the molecular transform is sampled differently, and differences may be due to inherent variations in the molecular transform rather than changes in the molecular transforms due to ligand binding. Several factors can cause variations in lattice parameters, including ligand binding and freezing. Freezing is

particularly problematic for β–galactosidase crystal variation. In these cases, the

difference electron density calculation must use calculated amplitudes for the native

conformation rather than the observed ones:

$$\rho^l(\vec{r}) - \rho^n(\vec{r}) = \sum_{\vec{k}} \left[F_o^l(\vec{k}) - F_c^n(\vec{k}) \right] e^{i\phi_c^n(\vec{k})} e^{2\pi i \vec{r} \bullet \vec{k}}$$

Such electron density is called "F_o-F_c", the liganded amplitudes coming from

observed data and the native amplitudes being calculated from a model of the native

enzyme which has been correctly positioned in the new crystal. This electron density is

sometimes less clear than F_o-F_o density because of difficulties in calculating the

amplitudes at low resolution. However, the binding mode for ligands can usually be

discerned from F_o-F_c density as well. An example is shown in Figure 21.
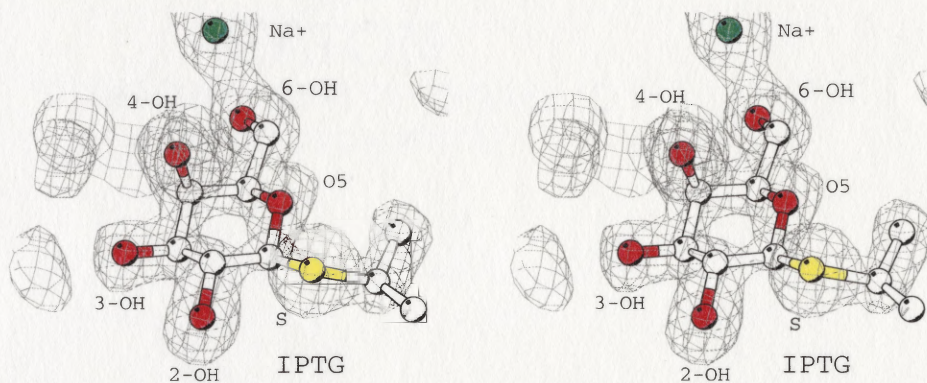
Figure 21. Stereo view of 1.75 Å F_o-F_c electron density for isopropyl-thio-β-D-galactoside (iptg) bound to native enzyme. In this case, because the crystal lattice was slightly repacked relative to native enzyme, an F_o-F_o map could not be calculated. After correctly placing the protein in the unit cell an F_o-F_c map could be calculated, resulting in the map shown, which clearly shows the binding mode for iptg. Only positive contours at 4 sigma are shown.

Assuming the binding mode of the ligand can be discerned from either F_o-F_o or F_o-F_c electron density, a model for the ligand is built on a graphics workstation. Also, the enzyme model is inspected and necessary changes to it are made. The coordinates for this model are then refined using a least squares procedure which tries to minimize the difference between the observed and calculated amplitudes by shifting the atoms around while keeping the whole model consistent with expected geometric parameters such as bond lengths, bond angles, and contact distances. The shifted coordinates are used to calculate another F_o-F_c map, which is used to further rebuild the model. After several cycles of refinement and model building, the end result is a model for the enzyme with bound ligand. This model then serves as a starting point for analyzing enzyme ligand interactions to understand how the enzyme carries out its activity.

Table 7 lists data collection and refinement statistics. The complexes are organized conceptually by experiment. The discussion which follows is organized along the listing of data complexes. Figure 22 shows some of the ligands whose binding modes to β-galactosidase were determined.

Substrates & Substrate Analogs
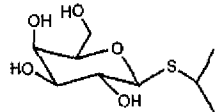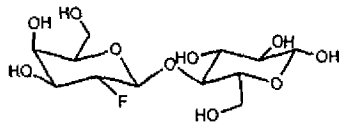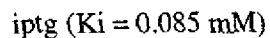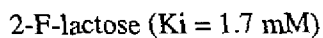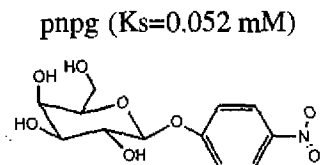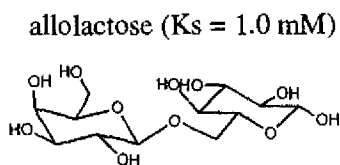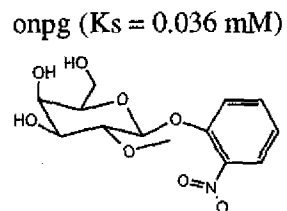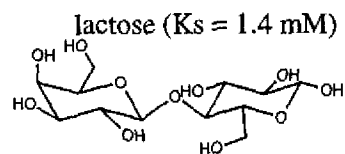
Transition State Analogs

galactonolactone (Ki = 0.5 mM)

lactose (Ks = 1.4 mM)

onpg (Ks = 0.036 mM)
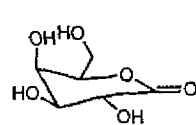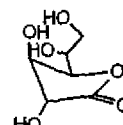
1,5 lactone (trace)   1,4 lactone (73 %)   galactonic acid (27%)

allolactose (Ks = 1.0 mM)

pnpg (Ks=0.052 mM)

L-ribose (Ki = 0.21 mM)   galactal (Ki = 0.016 mM)

2-F-lactose (Ki = 1.7 mM)
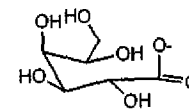
iptg (Ki = 0.085 mM)
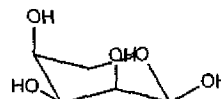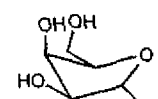
L-ribopyranose(~80 %)   L-ribofuranose (~20 %)
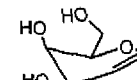
Galactotetrazole (Ki = 0.001 mM)
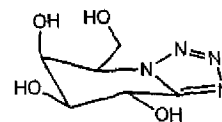
N-4-Bromobenzyl-1,5–deoxy-β-D-galacto-pyranosylamine (BBG)

(Ki = ~5 nM)
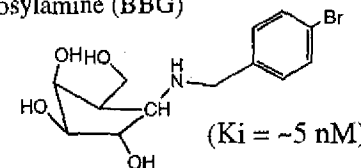
Figure 22. Schematics of some of the ligands whose binding mode was determined with crystallography. Left - substrate and substrate analogs. Right - transition state analogs.

Table 7(a) Data collection statistics for β-galactosidase complexes. Ligand = ligand; *=~1 mM Bis-Tris, otherwise 100 mM Bis-Tris; 2-d-g-E = 2-deoxy-galactosyl-enzyme; 2-F-g-E = 2-F-galactosyl-enzyme. Conc. = concentration of the ligand in the crystal. Res. = maximum resolution. R-Merge = average agreement between symmetry related reflections $(\Sigma_m(I_n-<I>)/<I>$, where the mean is over symmetry related reflections). All = all data, High = data in the high resolution shell. $<I>/<Sig(I)>$ = <signal>/<noise>. Red. = #total reflections/#unique reflections.

				R-Merge (%)		Completness (%)		$<I>/<Sig(I)>$		Reflections	
Identification	Ligand	Conc. (mM)	Res (Å)	All	High	All	High	All	High	Unique	Red
Reaction Coordinate											
Early											
Unliganded											
Native	none	-	1.70	6.0	34.6	99	91	17.1	3.0	523322	4.2
Real Substrates/E537Q											
lactose	lactose	100	1.80	6.6	18.8	97	82	12.5	4.1	450429	2.6
pnpg2	pnpg	150	1.55	4.2	27.6	98	97	10	4	715583	2.7
onpg	onpg	100	2.00	5.4	21.5	87	47	10.2	2.1	297446	2.7
galactal	galactal	~100	1.70	5.0	20.7	95	87	11.1	4.7	521708	1.8
Substrate Analogs											
gal-S-R											
Iptg	iptg	7	1.75	7.0	40.9	90	88	13	2.0	397254	3.8
Iptglu	ipt-glucose	40	2.70	8.3	27.9	92	94	11	3.3	128048	3.9
2-F-gal											
2-F-lac	2-F-lactose	89	2.70	9.0		66				408368	1.8
Middle											
Intermediates											
Hydrolysis											
2-d-gal	2-d-g-E	~50	1.75	6.1	45.5	94	88	14	2.2	471435	3.3
2-F-gal	2-F-g-E	~7	2.60	7.1		70				477611	1.9
Gtl-lobt	2-d-g-E*	10	2.10	8.1	27.0	90	80	7.5	2.3	265076	2.7

Table 7(a) continued

Identification	Ligand	Conc. (mM)	Res (Å)	R-Merge (%) All	High	Completness (%) All	High	<I>/<Sig(I)> All	High	Reflections Unique	Red
Ttransglycosylation											
2dG/Glc-1obt	2-d-g-E/glucose*	10/500	1.90	4.0	24.1	93	86	11	3.1	397875	2.4
2dG/Glc2-1obt	2-d-g-E/glucose*	50/500	2.30	8.3	27.6	96	95	10.4	2.6	219154	2.4
2FG/Glc	2-F-g-E/glucose	7/1000	2.10	6.0	34.2	91	61	10.3	2.8	275362	1.0
Transition State Analogs											
L-ribose	L-ribose	0.21	1.50	5.3	30.7	97	80	16	3.1	773703	3.9
Lactone	galactonolactone	100	1.80	6.8	41.5	91	87	12.7	2.0	420164	3.5
Tetrazole	galactotetrazole	0.1	2.10	8.8	37.9	86	87	12.4	2.7	288130	4.0
JG142	JG142	0.01	1.40	4.7	35.4	89	43	9.6	2	881391	3.2
Late											
Hydrolysis											
Galactose	galactose	400	1.50	7.4	22.6	90	87	9.1	2.5	750588	4.2
Glucose	glucose	500	2.80	9.4	30.8	99	98	9.3	3.4	131804	3.2
Transglycosylation											
E537Q/allo	allolactose	~30	1.50	5.4	24.6	98	85	12.4	3.2	778084	2.6
Lactose											
Lactose	lactose	130	1.70	4.8	26.7	93	75	11.6	4	739824	3.5
Na+ & Mg++ Function											
Potassium	K+	100	1.50	6.4	38.9	86	52	10.7	1.8	688912	2.6
Rubidium	Rb+	100	1.60	5.0	24.1	98	99	10.1	4.7	643587	3.4
K+/2dG	K+/2-d-g-E	100/10	1.85	8.2	31.3	94	95	6.9	2.1	407391	2.8
EDTA	edta	50	2.50	6.2	20.1	93	84	12.3	3.8	166292	2.6
Loop swing (room temp)											
L-rib-rt	L-ribose	21	2.70	7.7	29.6	98	95	9	2.2	145888	2.8
L-rib/dmso-rt	L-ribose/DMSO	21/30%	3.00	8.2	30.4	98	97	8.3	2.8	105972	2.5
Lactone-rt	galactonolactone	500	3.00	12.4	44.4	75	62	7.3	1.7	81588	2.1
Lactone/dmso-rt	lactone/DMSO	~100/30%	3.50	10.9	23.2	88	76	7.6	2.9	60013	1.9
2dG-hibt-rt	2-d-g-E	200	3.10	10.8	33.8	98	98	6.3	2.5	96563	2.5
2dG-hibt/dmso-rt	2-d-g-E/DMSO	200	3.00	9.9	33.9	96	94	7.3	2.1	104005	2.4

Table 7(a) continued

Identification	Ligand	Conc. (mM)	Res (Å)	R-Merge (%)		Completness (%)		<I>/<Sig(I)>		Reflections	
				All	High	All	High	All	High	Unique	Red
2dG-lobt-rt	2-d-g-E*	200	3.00	8.6	29.2	98	96	8.6	3.1	106546	2.6.
2dG-lobt/dmso-rt	2-d-g-E/DMSO*	200	3.00	9.9	33.5	99	96	7	2.1	106607	2.4
Tetrazole-rt	galactotetrazole	1	3.00	8.5	33.4	90	89	9.6	2.6	97866	2.5
Galactose-rt	galactose	40	2.90	9.1	32.6	98	99	8	3.6	118495	3.7

Glucose search (rt)

Identification	Ligand	Conc. (mM)	Res (Å)	All	High	All	High	All	High	Unique	Red
L-rib/glc-rt	L-ribose/glucose	21/900	2.80	8.7	34.8	99	99	8.9	3.3	132722	3.9
2dG/Glc-lobt-rt	2-d-g-E/glucose*	200/500	2.90	7.8	32.5	97	93	9.7	2.6	116775	2.4

Variants

E537Q (see also above)

Identification	Ligand	Conc. (mM)	Res (Å)	All	High	All	High	All	High	Unique	Red
E537Q-L-ribose	L-ribose	150	1.70	7.0	33.7	96	94	10.2	2.7	527809	2.5

F601A

Identification	Ligand	Conc. (mM)	Res (Å)	All	High	All	High	All	High	Unique	Red
F601A	none	-	1.55	5.8	17.5	91	67	9.0	5.2	671736	3.6
F601A-iptg	iptg	~10	1.70	5.7	23.5	90	87	8.1	3.6	492738	2.2
F601A-lactone	lactone	~100	1.80	6.8	34.9	86	79	6.7	1.8	392368	2.2
F601A-rt	none	-	3.20	11.6	35.7	98	99	6.9	3.4	88197	2.9
F601A-lrib-rt	l-ribose	21	3.50	10.0	25.3	98	100	11	7.1	67111	4.3

G794A

Identification	Ligand	Conc. (mM)	Res (Å)	All	High	All	High	All	High	Unique	Red
G794A	none	-	1.60	7.5	36.0	96	92	6.4	2.7	632326	3.5
G794A-iptg	iptg	~10	1.60	5.6	27.3	95	90	7.6	3	623810	3.2
G794A-rt	none	-	3.00	9.5	37.1	94	90	8.4	1.8	104632	2.1
G794A-gg-rt	gtl/glucose	100/250	3.10	9.2	31.6	96	91	7.3	2.2	94653	2.1
G794A-glc-rt	glucose	500	3.00	8.7	31.2	99	99	9.6	4.4	107677	3.9
G794A-ltn-rt	lactone	50	3.50	10.6	23.1	99	99	7.7	4.5	68323	2.9

DMSO Titration

Identification	Ligand	Conc. (mM)	Res (Å)	All	High	All	High	All	High	Unique	Red
DMSO0	none	-	3.00	9.2	29.4	95	99	8.9	2.3	103758	2.2
DMSO1	dmso	1%	3.00	8.2	28.6	99	99	12.7	3	107609	2.8
DMSO10	dmso	10%	3.30	10.7	27.6	93	95	11.8	4.3	76404	2.9
DMSO30	dmso	30%	3.20	11.5	29.0	95	96	7.7	2.5	84174	2.3

Table 7(b). Refinement statistics. All refinement was done with the TNT package (see methods). R-Fac = $\Sigma|F_o-F_c|/F_o$. R-Free is an R-factor with a small subset of data has not used in refinement. RMS Deviations are RMS deviations in bond lengths (Leng), bond angles (Ang) and B-factors (B-fac) from expected values for these parameters from a small molecule data base as defined in TNT. Units are Å, degrees and Å², respectively. Mean refined atomic B-factors are given for subsets of atoms: protein, solvent , and the two parts of the ligand (glycon and aglycon) if present. The scaling parameters define the model at low resolution (see discussion in Chapter 2). Blank entries were not refined.

Identification	R-Fac	R-Free	RMS Deviations			Number of atoms		<B-factor> (Å2)				Scaling				
			Leng	Ang	B-fac	Protein	Solvent	Prot	Solv	Glyc	Aglyc	K_{sol}	B_{sol}	B_{11}	B_{22}	B_{33}
Reaction Coordinate																
Early																
Unliganded																
Native	15.7	21.1	0.018	2.9	7.5	32500	4908	19.5	31.5	-	-	0.66	128.0	-1.6	1.7	-0.1
Real Substrates/E537Q																
lactose	15.4	21.8	0.018	3.0	8.0	32512	4945	16.8	27.7	8.5	23	0.69	183.0	-0.9	2.0	-1.1
pnpg2	18.0	22.9	0.019	2.9	7.4	32522	4942	19.2	32.1	12.1	19.5	0.83	217.0	-0.9	1.8	-0.9
onpg	21.3	32.2	0.017	3.0	5.5	32512	3421	34.8	42	30.8	61.9	0.87	226.0	-7.6	16.7	9.1
galactal	16.1	21.3	0.019	3.0	7.7	32512	4629	18.7	31.1	20.2	-	0.86	222.0	-1.6	1.7	-0.2
Substrate Analogs																
gal-S-R																
Iptg	16.8	24.5	0.015	2.8	5.5	32500	3956	25.1	34.3	18.6	22.3	0.74	130.0	-2.1	2.5	-0.4
Iptglu																
2-F-gal																
2-F-lac	21.4	22.5	0.019	2.7	4.3	16 x 8220	16 x 281	23*	28.9	35.1	52.6	1.00	693.0	0.0	0.0	0.0
Middle																
Intermediates																
Hydrolysis																
2-d-gal	15.8	22.2	0.016	2.7	8.4	32506	4823	22.1	33.9	13.2	27.3	0.70	115.0	-2.3	4.0	-1.7
2-F-gal	21.2	22.4	0.021	2.9	4.3	16 x 8220	16 x 293	23.8*	28.7	18.2	-	1.00	863.0	0.0	0.0	0.0
Gtl-lobt																

Table 7(b) continued

Identification	R-Fac	R-Free	RMS Deviations			Number of atoms		<B-factor> (Å2)				Scaling				
			Leng	Ang	B-fac	Protein	Solvent	Prot	Solv	Glyc	Aglyc	K_{sol}	B_{sol}	B_{11}	B_{22}	B_{33}
Transglycosylation																
2dG/Glc-lobt	24.0	24.3	0.014	2.3	4.1	31016	4513	32.4	48.5	?	?	0.88	262.0	-5.0	12.7	-7.7
2dG/Glc2-lobt	16.4	27.8	0.013	2.6	4.0	32500	4879	35.3	52.2	30.6	94	0.75	140.0	-0.9	7.8	-6.9
2FG/Glc	16.6	26.7	0.014	2.8	5.2	32500	4326	30.2	59.6	24.8	?	0.90	261.0	0.4	5.8	-6.3
Transition State Analogs																
L-ribose	17.4	22.0	0.018	2.9	3.7	32180	4811	19.5	33.2	14.6	-	0.87	219.0	-1.0	0.5	0.4
Lactone	16.5	23.8	0.014	2.7	7.1	32260	4664	25.2	38.9	20.5	-	0.74	143.0	-1.8	3.4	-1.6
Tetrazole	16.1	26.9	0.017	2.8	6.0	32500	3319	33	39.7	28.4	-	0.76	141.0	-6.2	10.7	-4.5
JG142	17.2	21.2	0.020	2.9	7.6	32643	5039	17.1	28.8	6.6	10.4	0.84	219.0	-1.6	2.2	-0.6
Late																
Hydrolysis																
Galactose	17.6	21.9	0.020	3.0	7.7	32508	4657	19.4	31.3	18.9	-	0.89	204.0	0.5	0.6	-1.1
Glucose																
Transglycosylation																
E537Q/allo						32500	4644	16	28	16	83.3					
Lactose																
Lactose	19.3	24.0	0.019	2.5	3.6	32500	4623	18.4	32.2	28.6	?	0.87	209.1	-1.7	2.3	-0.6
Na+ & Mg++ Function																
Potassium	16.2	21.4	0.019	3.0	7.9	32500	4751	18.6	29.3	-	-	0.85	218.0	-0.8	1.0	-0.2
Rubidium	18.2	23.6	0.019	3.0	8.6	32500	4399	22.1	32.3	-	-	0.85	202.0	-2.6	3.2	-0.6
K+/2dG	18.4	25.8	0.019	3.0	4.8	32500	4551	20.2	30	10.9	-	0.85	219.0	-1.6	2.9	-1.3
EDTA	28.8	34.7	0.017	2.6	6.6							0.84	103.0	-2.0	8.4	-6.3
Loop swing (room temp)																
L-rib-rt																
L-rib/dmso-rt																
Lactone-rt																
Lactone/dmso-rt																
2dG-hibt-rt	15.2	25.0	0.012	2.6	4.6	32628	748	36.6	40.2	48.6	?	0.75	140.0	-6.8	6.5	0.2
2dG-hibt/dmso-rt	15.1	28.8	0.012	2.8	5.9	32952	752	34	37.1	47.8	63.6	0.75	140.0	0.1	4.1	-4.2

Table 7(b) continued

Identification	R-Fac	R-Free	Leng	Ang	B-fac	Protein	Solvent	Prot	Solv	Glyc	Aglyc	K_{sol}	B_{sol}	B_{11}	B_{22}	B_{33}
			RMS Deviations			Number of atoms		<B-factor> (Å2)					Scaling			
2dG-lobt-rt	16.2	26.2	0.017	3.0	5.4	32596	744	35.8	39.8	38.1	-	0.80	190.0	0.5	-0.4	-0.1
2dG-lobt/dmso-rt	16.6	26.8	0.013	2.8	4.8	32756	748	35.6	39.3	49.3	-	0.75	140.0	0.7	2.0	-2.7
Tetrazole-rt																
Galactose-rt																
Glucose search (rt)																
L-rib/glc-rt																
2dG/Glc-lobt-rt	16.0	24.5	0.012	2.6	4.4	32628	756	40	43.4	35.4	?	0.75	140.0	-4.4	6.5	-2.1
Mutants																
E537Q (see also above)																
E537Q-L-ribose	17.7	26.1	0.018	2.9	9.5	32180	4448	21.8	32.8	17.2	-	0.86	212.4	-2.0	2.8	-0.8
F601A																
F601A																
F601A-iptg																
F601A-lactone																
F601A-rt																
F601A-lrib-rt																
G794A																
G794A	19.2	25.2	0.035	3.6	6.3	32410	4332	21.8	35.7	-	-	0.74	154.7	-0.4	1.6	-1.1
G794A-iptg																
G794A-rt																
G794A-gg-rt																
G794A-glc-rt																
G794A-lrn-rt																
DMSO Titration																
DMSO0																
DMSO1																
DMSO10																
DMSO30																

Table 7(c) Binding results from map inspections. Maps (F_o-F_o, otherwise F_o-F_c-marked with an asterisk) were inspected for ligand binding and the 794-804 loop conformation. Bound = positive feature (at least 4 sigma) for ligand binding. Not Bound = no obvious feature which corresponds to ligand binding. Open = no features to suggest the loop changes from its native conformation. Closed = both positive and negative features to suggest the loop moves from its native conformation to the closed conformation (e.g. Figure 34. Dest Open = negative features on the loop to suggest the native conformation is destabilized, but no obvious positive features to suggest where it goes. Mixed = different loop conformations can be discerned for different monomers. ICB = Novagen induction control B (see Chapter 2 methods). RT = room temperature (~294 K). SG = Space Group (4=P2$_1$, 19=P2$_1$2$_1$2$_1$). Both P2$_1$ data sets had β=95.0°.

Identification	Variant	Temp (K)	SG	a	b	c	glycon	Difference Map Result aglycon	794-804 loop
Reaction Coordinate									
Early									
Unliganded									
Native	ICB	~93	19	149.6	168.4	200.7	-	-	
Real Substrates/E537Q									
lactose	ICB/E537Q	~93	19	149.6	168.6	200.9	Bound	Bound	Open
pnpg2	ICB/E537Q	~93	19	149.7	168.6	201.1	Bound	Bound	Open
onpg	ICB/E537Q	~93	19	151.4	166.8	201.8	Bound*	Bound*	Open*
galactal	ICB/E537Q	~93	19	149.4	168.1	200.6	Bound*	-	Open*
Substrate Analogs									
gal-S-R									
Iptg	ICB	~93	19	151.8	161.2	202.9	Bound*	Bound*	Open*
Iptglu	ICB	~93	19	151.7	167.0	201.3	Not bound*	Not Bound*	Open*
2-F-gal									
2-F-lac	none	298	4	107.6	207.3	510.3	Bound	Bound	Open
Middle									
Intermediates									
Hydrolysis									
2-d-gal	ICB	~93	19	149.6	168.2	200.7	Bound	Bound(bis-tris)	Open
2-F-gal	none	298	4	107.5	207.2	510.2	Bound	-	Closed
Gtl-lobt	ICB	101	19	149.5	169.0	200.8	Bound	Not Bound	

Table 7(c) continued

Difference Map Result

Identification	Variant	Temp (K)	SG	a	b	c	glycon	aglycon	794-804 loop
Transglycosylation									
2dG/Glc-lobt	ICB	~93	19	151.7	167.5	202.0	Bound*	Bound *	Open*
2dG/Glc2-lobt	ICB	~93	19	151.8	168.0	202.8	Bound*	Bound *	Mixed/disordered*
2FG/Glc	ICB	~93	19	151.5	167.7	201.6	Bound*	Bound*	
Transition State Analogs									
L-ribose	ICB	~93	19	149.3	168.2	200.5	Bound	-	Dest Open
Lactone	ICB	~93	19	149.7	168.0	201.0	Bound	-	Dest Open
Tetrazole	ICB	~93	19	149.7	166.8	200.9	Bound	-	Open
JG142	ICB	~93	19	149.6	168.8	200.6	Bound	Bound	Mixed
Late									
Hydrolysis									
Galactose	ICB	~93	19	149.6	166.5	200.6	Bound*	-	Open*
Glucose	ICB	RT	19	153.8	171.8	204.8	Not Bound	-	Open
Transglycosylation									
E537Q/allo	ICB/E537Q	~93	19	149.4	168.7	200.9	Bound	Bound	Open
Lactose									
Lactose	ICB	~93	19	149.5	168.4	200.5	Mixed	Mixed	Open
Na+ & Mg++ Function									
Potassium	ICB	~93	19	149.3	168.4	200.4	Bound	-	Open
Rubidium	ICB	~93	19	149.5	168.2	200.5	Bound	-	Open
K+/2dG	ICB	~93	19	149.2	168.0	200.3	Bound	-	Open
EDTA	ICB	~93	19	151.6	166.6	202.2	Gone*	-	Open*
Loop swing (room temp)									
L-rib-rt	ICB	RT	19	153.9	171.6	204.4	Bound	-	Closed
L-rib/dmso-rt	ICB	RT	19	153.1	171.0	203.9	Bound	-	Closed
Lactone-rt	ICB	RT	19	153.8	171.5	204.5	Bound	-	Closed
Lactone/dmso-rt	ICB	RT	19	153.5	170.9	204.8	Bound	-	Closed
2dG-hibt-rt	ICB	RT	19	151.9	171.5	204.4	Bound	Bound (bis-tris)	Closed
2dG-hibt/dmso-rt	ICB	RT	19	151.9	171.2	204.2	Bound	Bound (bis-tris)	Dest Open (slight)

Table 7(c) continued

Identification	Variant	Temp (K)	SG	a	b	c	glycon	Difference Map Result aglycon	794-804 loop
2dG-lobt-rt	ICB	RT	19	153.9	171.7	204.4	Bound	Not Bound	Closed
2dG-lobt/dmso-rt	ICB	RT	19	153.5	171.0	204.0	Bound	Not Bound	Dest Open & closed ı
Tetrazole-rt	ICB	RT	19	153.8	171.5	204.5	Bound	-	Closed
Galactose-rt	ICB	RT	19	153.8	171.5	204.5	Bound	-	Open
Glucose search (rt)									
L-rib/glc-rt	ICB	RT	19	153.9	171.8	204.5	Bound	Not Bound	Closed
2dG/Glc-lobt-rt	ICB	RT	19	153.9	171.9	204.7	Bound	Bound (gluc)	Closed
Variants									
E537Q (see also above)							Bound	-	Open
E537Q-L-ribose	ICB/E537Q	~93	19	149.6	168.1	200.9			
F601A									
F601A	ICB/F601A	~93	19	149.7	168.6	200.7	-	-	Open
F601A-iptg	ICB/F601A	~93	19	149.6	168.0	200.2	Bound	Bound	Open
F601A-lactone	ICB/F601A	~93	19	149.4	166.7	200.5	Not bound	-	Open
F601A-rt	ICB/F601A	RT	19	153.7	171.9	204.4	-	-	Dest Open
F601A-lrib-rt	ICB/F601A	RT	19	153.5	171.8	204.6	Bound	-	Dest Open
G794A									
G794A	ICB/G794A	RT	19	149.7	169.6	201.0	-	-	Dest Open
G794A-iptg	ICB/G794A	~93	19	149.5	168.3	200.4	Bound	Bound	Open
G794A-rt	ICB/G794A	RT	19	153.9	171.6	204.4	-	-	Closed
G794A-gg-rt	ICB/G794A	RT	19	153.9	171.9	204.6	Bound	Not Bound	Closed
G794A-glc-rt	ICB/G794A	RT	19	153.9	171.8	204.8	Not bound	-	Closed
G794A-lm-rt	ICB/G794A	RT	19	153.8	171.7	204.7	Bound	-	Closed
DMSO Titration									
DMSO0	ICB	RT	19	153.9	171.4	204.5			Open
DMSO1	ICB	RT	19	153.9	171.4	204.5			Open
DMSO10	ICB	RT	19	153.9	171.2	204.3			Open
DMSO30	ICB	RT	19	153.5	171.0	204.1			Open

<u>Complexes Along the Reaction Coordinate of β-Galactosidase</u>

Early Points in the Reaction

Early points on the β–galactosidase reaction path were mimicked using two methods. First, natural substrates were bound to the altered enzyme, E537Q. In this variant, residue Glu-537 is replaced by a glutamine, resulting in an enzyme that is catalytically incompetent ($<10^{-4}$ x native activity – data not shown and [5]). Second, two classes of non-hydrolyzable analogs of β–galactosides– thiogalactosides and 2-F-lactose– were bound to native enzyme. In sum, atoms were changed at three positions to prevent substrates from being hydrolyzed: O → N on Glu-537, O4 → S on the substrate and O2 → F on the substrate. Attempts to image the natural substrate, lactose, with native enzyme were unsuccessful, but resulted in an image of galactose bound to the enzyme with some residual density at the aglycon position.

All of the "early reaction" ligands bind nearly identically. A representative example from this group, lactose bound to the E537Q variant, is shown in Figure 23. The ligand sits on the face of Trp-999 and the galactosyl hydroxyls 2,3 and 4 make specific contacts to enzyme and bound water, while the 6 hydroxyl contacts enzyme and a Na^+ ion. A summary of these interaction is shown in Table 8. The aglycon makes almost no specific contacts in any of the cases above with the important exception of His-418 (see below). The mean B–factor for aglycon atoms is about double that for the glycon atoms – reflecting the fact that there are very few specific contacts made by the aglycon. There is no obvious distortion in any of the ligands – the sugars are in a standard chair

conformation - and those for which there is a structure in the Cambridge Small

Molecule Database (CSD) bind in a conformation similar to that structure.   This

verification was made while determining geometry restraints for refinement.
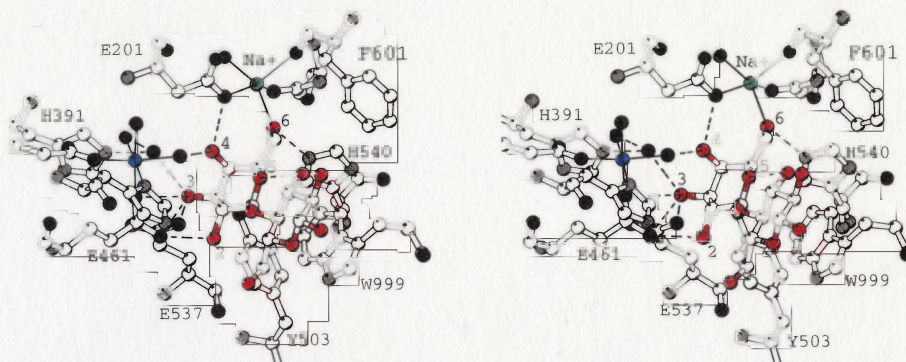
Figure 23. Stereo view showing the binding of lactose to the E537Q variant. The model shown is from the final refined coordinates to 1.5 Å. Lactose is shown in white (carbon) and red (oxygen), while the protein is shown in white (carbon), black (oxygen) and gray (nitrogen). The lactose molecule lies flat on Trp-999 with the galactosyl hydroxyls (numbered) making interactions to the protein and solvent. The 6-OH is directly liganded to a bound Na^+ ion (green), which has square pyramidal geometry. The 4-OH and 3-OH contact protein and Mg^{++}(blue) coordinated solvent.   The 2-OH has one polar contact to Glu-461. The glucose moiety, which lies further out of the active site, make no polar contacts with protein, although there is a hydrogen bond between the glucose and the galactosyl ring oxygen. The bound Mg^{++}(blue) shows octahedral geometry, with three solvent ligands and three protein ligands (Glu-416, His-418 and Glu-461).

A comparison of the binding modes of onpg and pnpg (Figure 24) shows that the

onpg nitro group contacts His-418 while the pnpg nitro group does not. This suggests an

explanation for the long observed elevated rate of ortho substituted aryl glycosides

relative to para substituted ones[6]. Galactosylation for onpg is about 15x faster than for

pnpg (Table 9). Since this probably involves delocalization of an electron on the pi

electron cloud of the leaving group, an interaction between His-418 and the nitro group

might help stabilize the developing negative charge. Since pnpg appears unable to

make this interaction with His-418, its rate for galactosylation is lower than for onpg.
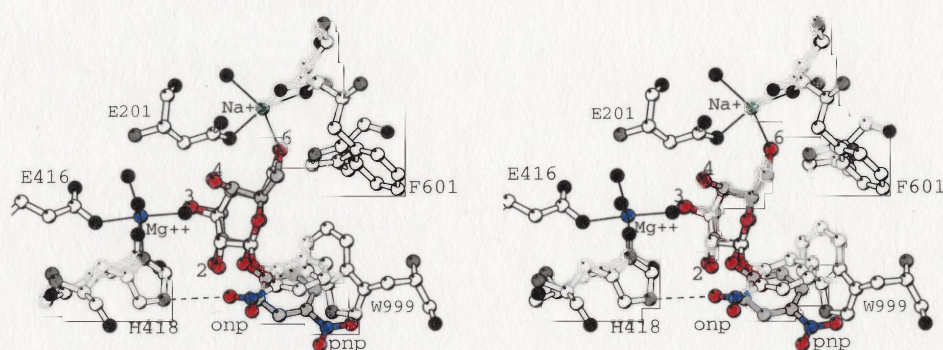
Figure 24. Stereo view comparing the binding of onpg (2.0 Å) and pnpg (1.5 Å) to
E537Q. Both ligands bind similarly to lactose. The interaction between the onpg nitro
group and His-418 is not possible with pnpg. This interaction could help bond cleavage
by slightly stabilizing the developing negative charge on the leaving group. Another onpg
molecule (not shown) was weakly bound such that its nitrophenyl ring stacked with the
one shown.

In these "early reaction" complexes, C1 is about 6 Å from the nucleophile Glu-

537, while O4 is also about 6 Å from the potential acids, Mg^{++} and Glu-461. Therefore,

the substrate would need to reposition relative to the enzyme before nucleophilic or acid

catalysis can take place.

Table 8. Distances between polar enzyme groups and the galactosyl substituents. Also shown are distances between the active site Mg^{++} and Na$^+$ ions and the enzyme. Distances were determined based on Monomer A for each complex. Distances greater than 3.2 Å are not shown. If a substituent is absent in a ligand, an "a" is entered in the table.

| | | Early Complexes | | | | | | Native | Intermediates | | | T - State Analogs | | Products | |
| | | | | E537Q Complexes | | | | | | | | | | | |
		native (Na+)	native (K+)	lactose	onpg	pnpg	galactal	iptg	2-F-gal	2-deoxy-gal (Na+)	2-deoxy-gal(K+)	lactone	tetrazole	galactose	E537Q/allolactose
2-OH(F)	N460 (od1)						a		3.2	a	a				
	N460 (nd2)						a		3.0	a	a				
	E461 (oe1)						a			a	a	3.1	3.0	3.0	
	E461 (oe2)			2.6	2.9	2.6	a	2.5		a	a	3.2			2.5
	E537 (oe1)						a		2.8	a	a	2.6	2.7	2.7	
	E537 (oe2)						a			a	a	3.1	3.1	3.1	
	HOH			3.2		3.1	a			a	a				
3-OH	H391								2.7	2.7	2.6	2.8	2.5	2.8	
	H357-HOH								2.7	2.7	2.6	2.7	3.0	2.8	
	E461 (oe1/oe2)			3.2											3.2
	E461 (oe1/oe2)			3.0		3.1	2.9	2.7							3.1
	E(Q)537 (1)			3.2	2.9	3.0									3.1
	Mg-HOH (3)								2.6	2.7	2.6	2.7	2.8	2.8	
	H391-HOH			2.8	2.6	2.8	2.7	3.0							2.8
4-OH	D201			2.7	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.6	2.5	2.6
	Mg-HOH (1)			2.5	2.7	2.6	2.4	2.8	2.8	2.6	2.6	2.7	2.4	2.6	2.6
	H391-HOH				2.9			3.0							
	Glyco - O														
6-OH	D201			3.2	3.2	3.2									
	H540			2.8	2.8	2.8	2.8	2.6	2.8	2.8	2.8	2.8	2.6	2.8	2.8
	N604			2.9	3.0	2.9	2.9	3.1	3.1	2.9	3.1	2.9	2.9	2.9	2.9
	Na+			2.3	2.4	2.3	2.3	2.4	2.5	2.5	2.6	2.5	2.6	2.4	2.3
	W568											3.1	3.0		

| | | Early Complexes | | | | | Native | Intermediates | | | T- State Analogs | | Products | |
| | | E537Q Complexes | | | | | | | | | | | | |
		native (Na+)	native (K+)	lactose	onpg	pnpg	galactal	iptg	2-F-gal	2-deoxy-gal (Na+)	2-deoxy-gal(K+)	lactone	tetrazole	galactose	E537Q/al-lolactose
O1	E461 (oe2)						a		a	a	a	2.5		2.5	
	M502 (sd)						a		a	a	a	3.0		3.2	
O5	N102			3.1	3.2			3.1							
	E537 (oe2)								2.6	2.4	2.4	3.1			
	Y503 (oh)									2.9	3.0	3.1			
C1	E537											2.9		3.0	
Mg++	E416 (oe1)	2.1	2.1	2.0	2.2	2.0	2.1	2.0	2.1	2.1	2.1	2.1	2.1	2.1	2.0
	E461 (oe1)	2.2	2.2	2.2	2.3	2.1	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.0	2.0
	H418 (ndx)	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.1	1.9	2.1	2.2
	HOH (1)	2.2	2.1	2.2	2.1	2.1	2.1	2.1	2.2	2.2	2.1	2.1	2.2	2.0	2.1
	HOH (2)	2.1	2.1	2.2	2.3	2.1	2.1	2.2	2.2	2.2	2.2	2.1	2.2	2.1	2.2
	HOH (3)	2.1	2.1	2.1	2.3	2.1	2.2	2.0	2.2	2.2	2.2	2.2	2.1	2.2	2.1
Na+	F601 (o)	2.4	2.6	2.4	2.4	2.4	2.4	2.3	2.4	2.5	2.7	2.5	2.3	2.3	2.4
	N604 (od2)	2.4	2.6	2.4	2.4	2.4	2.3	2.3	2.5	2.5	2.7	2.4	2.4	2.4	2.4
	D201 (oe1)	2.4	2.6	2.3	2.3	2.3	2.4	2.4	2.3	2.3	2.6	2.4	2.5	2.4	2.3
	HOH	2.3	2.4	2.4	2.4	2.2	2.2	2.4	2.3	2.3	2.6	2.4	2.3	2.3	2.3
	HOH/6-OH	2.3	2.8	2.3	2.4	2.3	2.3	2.4	2.5	2.5	2.6	2.5	2.6	2.4	2.3

Table 9. Representative kinetic parameters for β-galactosidase. Parameters for lactose, pnpg and onpg hydrolysis are from phosphate buffer [7]. Parameters for lactose transglycosylation and allolactose hydrolysis are from tris buffer [8]. Inhibition for substrate analogs is from TES buffer[9]. Most of the transition state analogs were measured in TES buffer[10], with the exception of the tetrazole[11] and BBG[12]. The galactal kinetics were done in imidazole buffer[13] and the 2-F-dnp-gal kinetics in phosphate[14].

	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m $(s^{-1}\,M^{-1})$	K_s (mM)	k_2 (s^{-1})	k_3 (s^{-1})
lactose (hyd)	60	1.4	4.3×10^4	1.3	65	900
lactose (tg)	61	1.4	4.4×10^4			
allolactose	92	1.0	9.2×10^4			
pnpg	90	0.04	2.2×10^6	0.036	100	900
onpg	550	0.14	3.9×10^6	0.052	1580	900
galactal			270^a		54^b	4.6×10^{-3}
2-F-dnp-gal			13^c	0.78^d	0.01^e	$4.2 \times 10^{-6 f}$

[a] 2nd order rate for inactivation [b] assumes $K_s \sim 200$ mm. Galactal inhibition has been measured up to 10 mM w/o saturation. [c] 0.01/(0.78/1000) [d] K_i [e] rate constant for inactivation [f] rate constant for reactivation

Inhibitor	K_i (mM)	K_i"(mM)	k_4 (s^{-1})
Substrate Analogs			
β-D-glucose	>300	17	380
β-D-glucose (2-F-gal)		460	7.6×10^{-5}
β-D-galactose	40		
2-deoxy-β-D-galactose	160		
phenyl-thio-β-D-galactoside	0.19		
o-nitro-phenyl-thio-β-D-galactoside	0.3		
isopropyl-thio–β-D-galactoside	0.085		
lactose	1		
Transition State Analogs			
galactonolactone	0.5		
L-ribose	0.21		
D-galactal	16×10^{-3}		
galactotetrazole	1×10^{-3}		
"BBG"	$\sim 1 \times 10^{-6}$		

Intermediate Points in the Reaction

True Catalytically Competent Intermediates

True intermediates along the reaction path were isolated by using two different

substrates for which the rate of step 2 is much smaller than step 1, allowing the

intermediate population in the crystal to build up long enough to collect diffraction data.

Dinitrophenyl-2-F-galactoside was prepared by John McCarter, Lloyd MacKenzie and

Steve Withers at the University of British Columbia, Vancouver, BC. The dinitrophenyl

group increases k_2 because the larger pi electron cloud makes it easier to delocalize the

negative charge on the leaving group. The 2OH → F change is thought to have two

effects. First, the interactions with enzyme should be different due to the replacement of

the OH hydrogen by a F lone pair, eliminating the possibility of H-bond donation by the

2-OH (see below). Second, because the F is more electronegative it destabilizes any

positive oxocarbonium ion which develops in the transition state. The other substrate

used to image a covalent intermediate is the slow-binding inhibitor, galactal (Figure 22).

This is thought to form a 2-deoxy-galactosyl-enzyme intermediate with a half-life of 10

minutes[13]. A 40 minute soak time was enough for the galactal to diffuse throughout the

crystal and form the 2-deoxy-galactosyl intermediate, which was trapped with flash

freezing. With higher concentrations of the galactal, room-temperature data collection

could also be performed, resulting in lower resolution images of the intermediate.

Figure 25 compares the native/2-deoxy-galactosyl intermediate with the

E537Q/lactose complex. There is a covalent bond between the ligand and Glu-537 and

the galactosyl ring has moved deeper into the active site, rotating ~90° to rest on Trp-568. The 2- and 3- substituents have moved to new binding locations, which were (roughly) occupied by water molecules in the unliganded enzyme. Changes in the binding of the 4- and 6-OHs are more subtle, although both are deeper in the active site. The 6-OH is still liganded to the Na$^+$, although the O-Na$^+$ distance has increased by 0.1-0.2 Å. Phe-601 has swung into a different rotamer conformation to pack against the hydroxymethyl of the galactosyl moiety. There are additional changes to the enzyme linked to the Phe 601 swing which will be discussed later.

Contacting both Glu-461 and the galactosyl ring oxygen is a water molecule (or other acceptor molecule – see below). Presumably this is the water molecule that will bond to C1 to form the product. In these complexes, this water molecule does not have
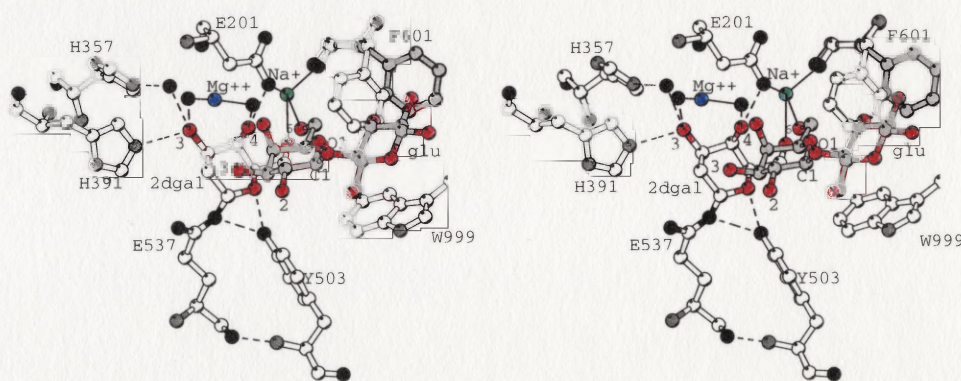
Figure 25. Stereo view comparing the 2-deoxy-galactosyl intermediate (1.7 Å, white carbons) to the lactose complex (gray carbons) In the intermediate, the galactosyl moiety has rotated about 90 degrees and moved deeper into the active site stacking on Trp-568 (not shown in the figure, but in the parallel to the plane of the paper beneath 2dgal). This substantially changes the environments of the 2 and 3 substituents. The 4 and 6-OHs change less, although the 6-OH moves about 0.1-0.2 Å farther from the Na$^+$ ion (Table 8). Phe-601 swings toward the Na$^+$ ion filling in for the departing hydroxymethyl group. Additional changes in enzyme structure are discussed later in the chapter.

direct access to C1, suggesting the galactosyl enzyme bond is broken prior to attack.

by the water molecule, consistent with an SN1 mechanism for degalactosylation

Significantly, the Mg^{++} ion does not interact directly with the substrate, while its

four of its ligands, including Glu-461, do interact with the substrate. This suggests first,

that the active site acid is Glu-461, not Mg^{++}. Second, the Mg^{++} may play a structural role

in properly positioning Glu-461, His-418 and water molecules, or an electronic role in

tuning the pKas of its protein ligands, Glu-461, Glu-416 and His-418.

These complexes represent intermediates for hydrolysis. Three attempts were

made to determine the structure of the intermediate for transglycosylation by soaking

either the 2-F-galactosyl or 2-deoxy-intermediate in high concentrations of glucose (500-

1000 mM). A lower resolution (3.0 Å) F_o-F_o map between the 2-deoxy-galactosyl

intermediate with and without glucose shows a 8 σ feature in the active site on Trp-999

(Figure 26). F_o-F_o maps could not be calculated for frozen data sets due to different

lattice changes caused by freezing (the high concentrations of glucose appear to alter the

freezing properties). In these cases, F_o-F_c maps also suggest glucose binding adjacent to

Trp-999, interacting with Glu-461, Asn-102 and perhaps His-418. However, a

convincing model could not be refined. Additionally, in some of the active sites this

density appears more linear than planar. Overall, this suggests the glucose molecule is

bound to the intermediate in multiple conformations, and since none of the conformations

are highly occupied, none of the binding modes can be discerned.
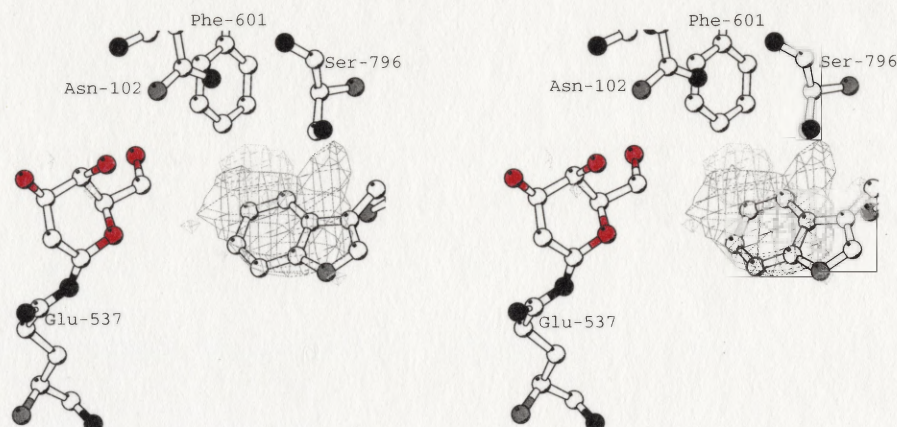
Figure 26. Stereo view of four-fold averaged 3.0 Å F_o-F_o electron density map contoured at +/- 4 σ (0.08 electrons/$Å^3$) for glucose binding to the intermediate. The data sets were collected at room-temperature on the 2-deoxy-galactosyl intermediate in the presence and absence of 500 mM glucose. The feature on Trp 999, near Ser-796 and Asn-102 suggests a general binding location for glucose, but a model could not be refined, suggesting multiple binding modes.

The suggestion that glucose has multiple binding modes is supported by biochemical experiments in the literature[15] The second order rate constant for glucose addition to the intermediate is k_{Glc}=1.2 x 10^4 M^{-1} s^{-1}, significantly below the diffusion limit. Also, the first order rate constant for glucose addition to the intermediate after binding is smaller than for the reaction with water[16] . This suggests that glucose has other populated binding mode(s) which do not lead to allolactose production.

Transition State Analogs

Whether bond cleavage going to or from the intermediate is an SN1 or SN2 process, the anomeric center should be planar near the transition state(Fig 4, Chapter 1). In the case of an SN1 process, there will also be a positive charge on the galactosyl ring (an oxocarbonium ion), whereas for an SN2 process there will be no formal charge. An

ideal transition state analog, then, would be planar at C1. To mimic the TS for the

SN1 cleavage the analog should also be positively charged.

Three ligands were used as transition state analogs – galactonolactone,

galactotetrazole (kindly provided by Thomas Heightman and Andrea Vasella, ETH,

Zurich) and L-ribose. The lactone and tetrazole are both planar at C1. Although they

both lack a formal charge at pH 7.0, they both have a dipole moment with the positive

end at C1. L-ribose is not expected to be planar, but generally follows the inhibition due

to other transition state analogs. All three bind in about the same mode as the

intermediates. Glu-461 contacts the atoms corresponding to the glycosidic oxygen in the

lactone and the tetrazole (there is no such atom with the L-ribose – see below), further

supporting its role as the active site acid/base. The lactone binds in the 1-5 form,

although this is by far the least populated of the three forms for the lactones ([17] and

Figure 22), illustrating the specificity of the enzyme.

Differences Among the Intermediate Complexes and the Reaction Coordinate in the Vicinity of the Transition State

Although all of the intermediate complexes bind similarly, there are some subtle

differences which offer insight into the reaction coordinate near the transition state.

The chief difference between the two covalent intermediates (2-F and 2-deoxy) is the

location of Glu-537 (Figure 27). This residue is slightly rotated in the 2-deoxy-galactosyl

complex, presumably due to the covalent bond between the sugar and Glu-537. The

rotation is increased in the 2-F-galactosyl intermediate, probably due to repulsion

between the fluorine and side chain carbonyl of Glu-537 . The ligand stays the same

suggesting its binding orientation is determined by the specific interactions with

hydroxyls 3,4 and 6.
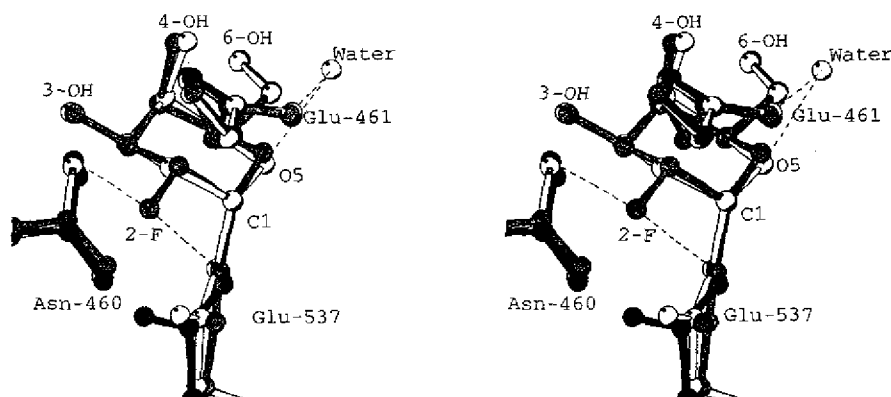
Figure 27. Stereo view comparing the 2-deoxy-galactosyl intermediate(1.7 Å, gray), the 2-F-galactosyl intermediate(2.1 Å, white) and unliganded enzyme (1.7 Å, black). The galactosyl ring sits nearly identically, suggesting its binding mode is dominated by the covalent bond to Glu-537 and noncovalent interactions made by hydroxyls 3,4 and 6. The major difference between the three structures is the rotation of Glu537. This residue rotates 25 degrees (χ3) in the 2-deoxy-intermediate, presumably from forming the covalent bond. The rotation increases by 15 degrees in the 2-F-intermediate, probably from repulsion of the Glu-537 oxygen by the fluorine.

The lactone and tetrazole bind similarly and rotated relative to the intermediates,

slightly shifting the 2 substituent by ~0.7-0.8 Å and the other substituents by ~0.2-0.5 Å

(Figure 28). This shift is probably due to steric repulsion between Glu-537 and the C1

mimic. Since the 2 substituent is closest to C1, it changes the most. Since the true

transition state would involve a partial bond between Glu-537 and the C1 mimic, it is

likely that the lactone and the tetrazole are not ideal transition state mimics, but represent

a position on the reaction coordinate between the transition state and the products (or

reactants).
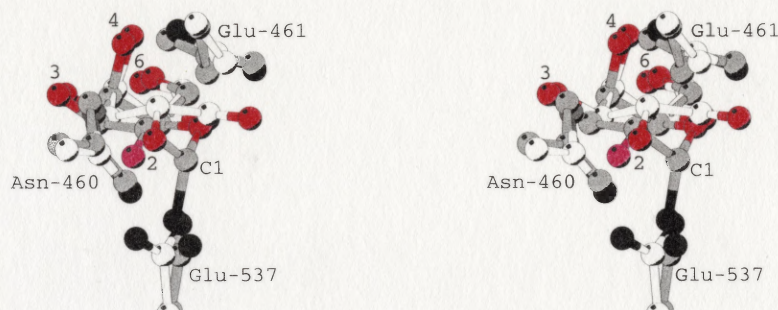
Figure 28. Stereo view comparing the planar transition state analog galactonolactone (1.7 Å, white carbons) to the 2-F-galactosyl intermediate (1.7 Å, gray carbons). The ring flattening of the lactone is accomplished primarily by moving the ring oxygen (or nitrogen in the case of the tetrazole, which is not shown) without affecting the relative geometry of substituents 2-6. The lactone rotates and shifts, probably due to repulsion between C1 and Glu-537 . The 2 substituent, closest to C1, moves the most.

The Binding of L-ribose is Atypical

The binding of the third transition state analogue, L-ribose is a curious case. Inhibitor binding studies suggested that L-ribose should bind in its furanose form, because, first, the temperature dependence of its inhibition closely follows the temperature dependence of the furanose population[17]. Second, in the furanose form, the hydroxyls are oriented similarly to hydroxyls 2,3 and 4 of galactose. In the pyranose form, however, L-ribose clashes at the 2-OH – galactopyranose is equatorial at this position whereas L-ribopyranose is axial. Despite these considerations, the structure of L-ribose bound to β–galactosidase shows that it binds in the pyranose form, but in a non standard orientation (Figure 29). It is in a normal chair conformation, but rotated 60 degrees and puckered inversely from galactopyranose. The effect is that the "3","4" and

"6"-OHs and ring oxygen make nearly identical interactions to the transition state

analogs and intermediates, but the "2"-OH moves slightly.
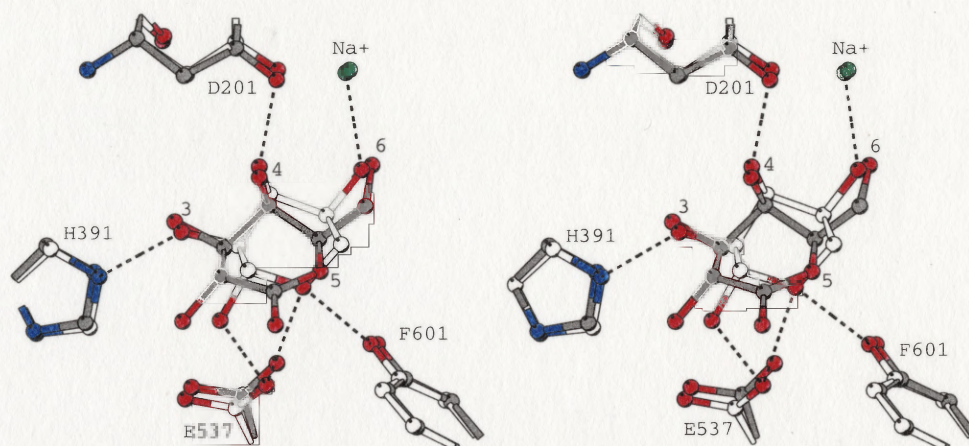
Figure 29. Stereo view comparing the binding of L-ribose (1.5 Å, white carbons) to the lactone (1.7Å, gray carbons). Somewhat surprisingly, L-ribose binds its pyranose form with the ring oxygen and the C1 equivalent swapped. However, because only it has a hydroxyl instead of a hydroxymethyl, it can fulfill almost identically the interactions made by the 3,4 and 6-OHs and the ring oxygen. The positioning of the 2-OH shifts the most.

In both the lactone and L-ribose structures, Phe601 is in the intermediate

conformation, whereas with the tetrazole, the native conformation is favored. This is

probably due to some extra density sitting next to the tetrazole in the region of Phe-601 –

perhaps a DMSO molecule. A room temperature structure in the absence of DMSO

shows that Phe-601 swings closed and the additional enzyme changes also occur.

Product State Complexes

Structures of product states for hydrolysis were determined by collecting data on

crystals with high concentrations of galactose and glucose. The galactose complex shows

that galactose binds similarly to the lactone and tetrazole with the 2-, 3- and 4- hydroxyls

in similar positions (Figure 30). The main difference is that the 6-OH has shifted along

the Na^+ sphere back to the early conformation, displacing Phe-601, which has opened to

the native conformation. This is probably due to steric repulsion between Glu-537 and

C1 created by puckering the galactose ring.
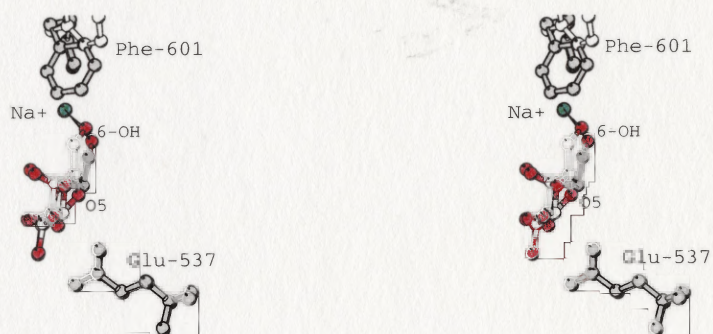
Figure 30. Stereo view comparing the product galactose (1.5 Å, white carbons) to the transition state analog galactonolactone (1.7 Å). The puckering of the galactose results in a shift of the hydroxymethyl group (including the 6-OH), while leaving the 2, 3 and 4-OHs unmoved. Although this shift is very slight, it is enough to move Phe-601 back into the native conformation, which resets the enzyme conformational change.

One attempt was made to determine the structure of glucose bound to the active

site by soaking a native crystal in 500 mM glucose. F_o-F_o maps showed no obvious

electron density greater than 3 sigma, suggesting that either the glucose was not bound

specifically or was competed out by something else in the buffer. This is not surprising,

because the dissociation constant for glucose from the active site is about 400 mM

(Table 9). The crystals were sensitive to higher concentrations of glucose, inhibiting data

collection under these conditions.

A complex between allolactose and the E537Q variant was used as a mimic for

the product complex for transglycosylation. Native enzyme could not be used because

allolactose is a substrate of β–galactosidase. In this complex, the galactosyl moiety binds

as in the early complexes. Beyond the galactosyl moiety a linkage that involves 2 atoms

to branching density can be seen, consistent with the allolactose 1-6 bond. (Figure 31).

Part of the glucose moiety is not visible, suggesting disorder. Three glucose

conformations can be modeled that are consistent with the branching density and not anti

to the 1-6 bond. Out of all of these conformations there are only 3 contacts to protein <

3.5 Å, and all are > 3.1 Å. These observations suggest that although a clear binding

mode cannot be seen for the glucose moiety, there is little reason to expect one. In the

case of lactose, onpg and pnpg, the galactose and aglycon are connected by a single atom,

which severely limits the conformational freedom of the aglycon in comparison to

allolactose. Furthermore, in lactose there is an interaction between the galactose ring

oxygen and O2 of the glucose which further restrains the lactose conformation.
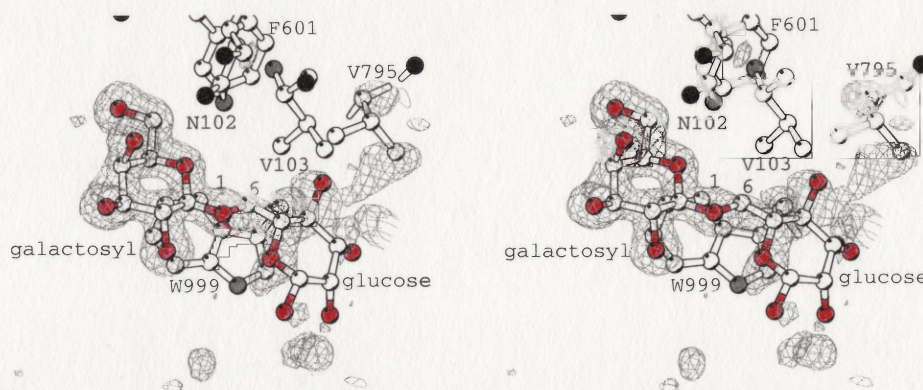
Figure 31. Stereo view showing the binding of allolactose to the E537Q variant. The electron density is positive 1.5 Å F_o-F_c density (3 sigma, or 2.13 electrons/Å^3) with F_c coming from the refined coordinates without the allolactose. The density clearly shows the binding of the galactosyl moiety and a two atom link to branching density, which is consistent with the 1-6 linkage in allolactose. The branching suggests three conformations for the glucose molecule, all with reasonable torsion angles with the 1-6 bond (trans, gauche + and gauche -). Only one of these conformations is shown. None of these conformations have distances to protein atoms less than 3.1 Å.

An Enzyme Conformational Switch

Induction of the Switch by Ligand Binding

As mentioned above, the swing that Phe-601 goes through between the early complexes and intermediate is linked to further changes in enzyme conformation (Figure 32). Arg-599 packs with Phe-601 and is ordered in the early complexes, but becomes disordered in the intermediate. More strikingly, the 794-804 loop swings toward the active site with residue shifts up to 9 Å.

One effect of the intermediate, or closed, conformation is to restrict access to the intermediate and the galactosyl group probably can not be released from the active site in this conformation. However, the change does not act as a lid, sequestering the active site
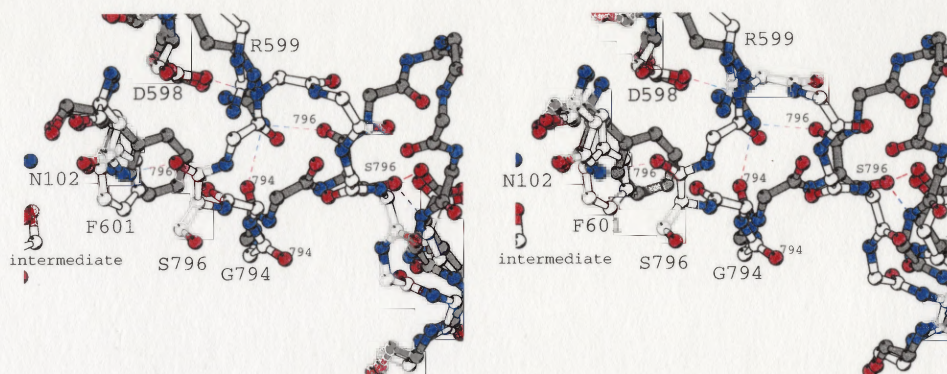
Figure 32. Stereo view showing the conformational switch. The native, or open conformation is shown with gray carbons and the intermediate, or closed conformation is shown with white carbons. When the substrate moves to the intermediate, Phe-601 can swing to the left. This destabilizes Arg-599 and its hydrogen bonds to the backbone carbonyls of Gly-794 and Ser-796. The 794-804 loop swings to the left toward the active site making two new polar interactions with its backbone to Asn-102 and Asp-598. The Ser-796 side chain, which was involved in an intraloop hydrogen bond in the open conformation (on the right) moves about 9 Å and its C_α and C_β pack on Phe-601.

from bulk solvent as has been observed with other enzymes[18].

This conformational change might be triggered, for instance, by changes in the charge states of Glu-461 and Glu-537 in going to the intermediate. However, this possibility seems unlikely since the conformational change is observed with the transition state analogs L-ribose, galactonolactone and galactotetrazole, and, most importantly, with L-ribose bound to the inactive variant E537Q. Thus the conformational change appears to be a function of binding interactions only. Inspection of the structure suggests that the conformational change is triggered by a subtle change in the coordination geometry of the active site Na^+ ion in going from the early complexes to the intermediates.

In the native, unliganded structure, the Na^+ ion has two water ligands and three protein ligands. In the early complexes, one of the water ligands is replaced by the galactosyl 6-OH in about the same place (Figure 23, pg. 92). This water or 6-OH

prevents Phe-601 from swinging into the new rotamer well. In moving to the intermediate, the 6-OH remains liganded to the Na^+, but moves about 20 degrees on the Na^+ sphere and 0.1-0.2 Å further away from the Na^+, opening some room for Phe-601 to swing in (Figure 25, pg. 98). This appears to destabilize Arg-599, breaking its contacts to two backbone carbonyls of the 794-796 loop. The destabilized loop finds two new contacts in the closed conformation to Asp-598 and Asn-102 (Figure 32). Furthermore, a new hydrophobic interaction is created between Phe-601 and Ser-796. The change in galactosyl position required to trigger the conformational change is very small, as the galactose complex is open, while the lactone complex is closed (Figure 30, pg. 105).

Induction of the Switch by Amino Acid Substitution

Figure 33 shows a Ramachandran plot for the 794-804 loop in the two conformations. In the native conformation, Gly-794 is in a region that would be high energy for a non glycine residue, but moves to a lower energy conformation (for non-glycine) in the closed conformation. Therefore, it seemed reasonable that a mutation at this position would force the enzyme into its intermediate conformation without the intermediate present. This position is of further interest because it has been shown to be functionally important (see below). The variant G794A was constructed and the structure determined. Figure 34 shows that, as predicted, the G794A enzyme is in the closed conformation without ligand present. Further structural studies show iptg binding to G794A forces the enzyme into its open conformation.

Phi-Psi Difference Plot of Open vs Closed Loop
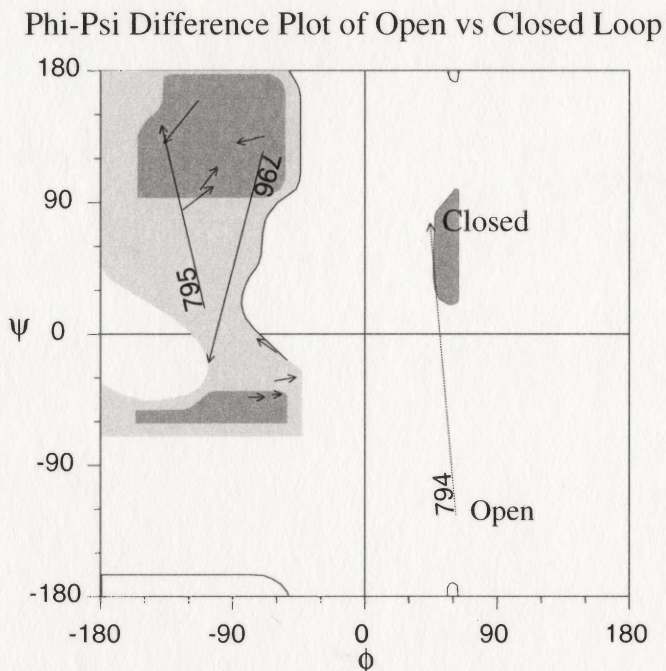
Figure 33. Phi-psi plot for the 794-804 loop in the open and closed conformation. Gray portions show lower energy conformations for non-glycine residues. Glycine 794 moves to a low energy region in the closed conformation. Altering Gly-794 to Ala forces the loop into the closed conformation (see text). Figure prepared with dphipsi (M.Quillin, unpublished program.)
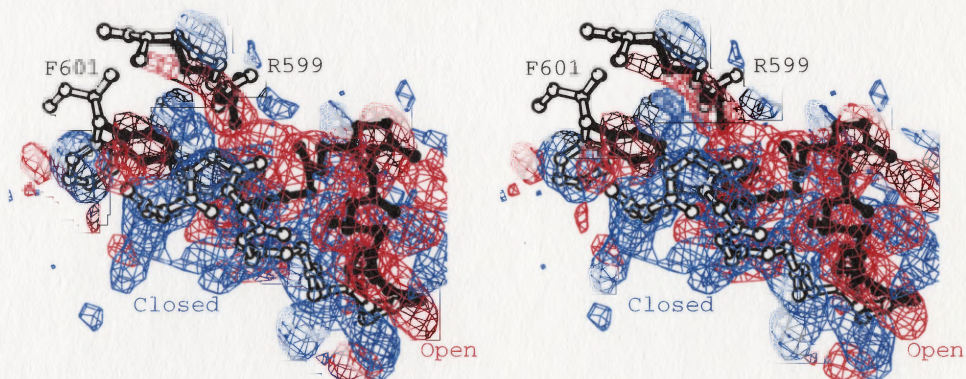
Figure 34. F_o-F_o electron density map between G794A and native enzyme contoured at +4/-5 σ. Native coordinates are shown in black and G794A coordinates are shown in white. The density clearly shows the loop swings to the left into a new conformation as a result of the mutation. Phe-601 also swings to the left and Arg-599 becomes disordered. Position 794 is at one end of the loop and not visible in the figure.

Because Phe-601 appears to play an important role in the conformational change by affecting the stability of Arg-599 and drawing Ser-796 in, the variant Phe-601-Ala was constructed. This appears to have the general affect of destabilizing both conformations, although iptg and L-ribose bind in the same modes as in native and G794A (Table 7(c)).

Besides the obvious changes described above there are much more subtle changes involving the overall enzyme conformation. Figure 35 shows the correlation between changes in domain 3 interatomic distances relative to native protein induced by the various ligands or mutations. It is defined as:

$$\left\langle \delta d^1 \bullet \delta d^2 \right\rangle = \frac{1}{4} \sum_{a=1}^{4} \frac{\sum_{i,j} \left(\delta d_{i,j}^{1^a} - \overline{\delta d^{1^a}} \right) \cdot \left(\delta d_{i,j}^{2^a} - \overline{\delta d^{2^a}} \right)}{N^{1^a} \cdot N^{2^a}}$$

where, $\delta d^1_{i,j}$ refers to the change in distance between the ith and jth atoms in going from native enzyme to structure 1. The superscript over the 1 refers to which monomer is being considered and an average is taken over the four monomers in the asymmetric unit. This gives a measure of how similarly a perturbation (ligand binding or mutation) affects the structure of the enzyme. Although the δds are typically less than 0.01 Å and the maximum δds are about 0.5 Å there is still an interpretable correlation in whether the enzyme is in its open or closed form. Of particular interest are E537Q/L-ribose and G794A, which fit with the closed complexes and G794A/iptg which fits better with the open complexes. Thus the gross features of the conformational change described above propagate to a small overall change in the C_α trace for domain 3.
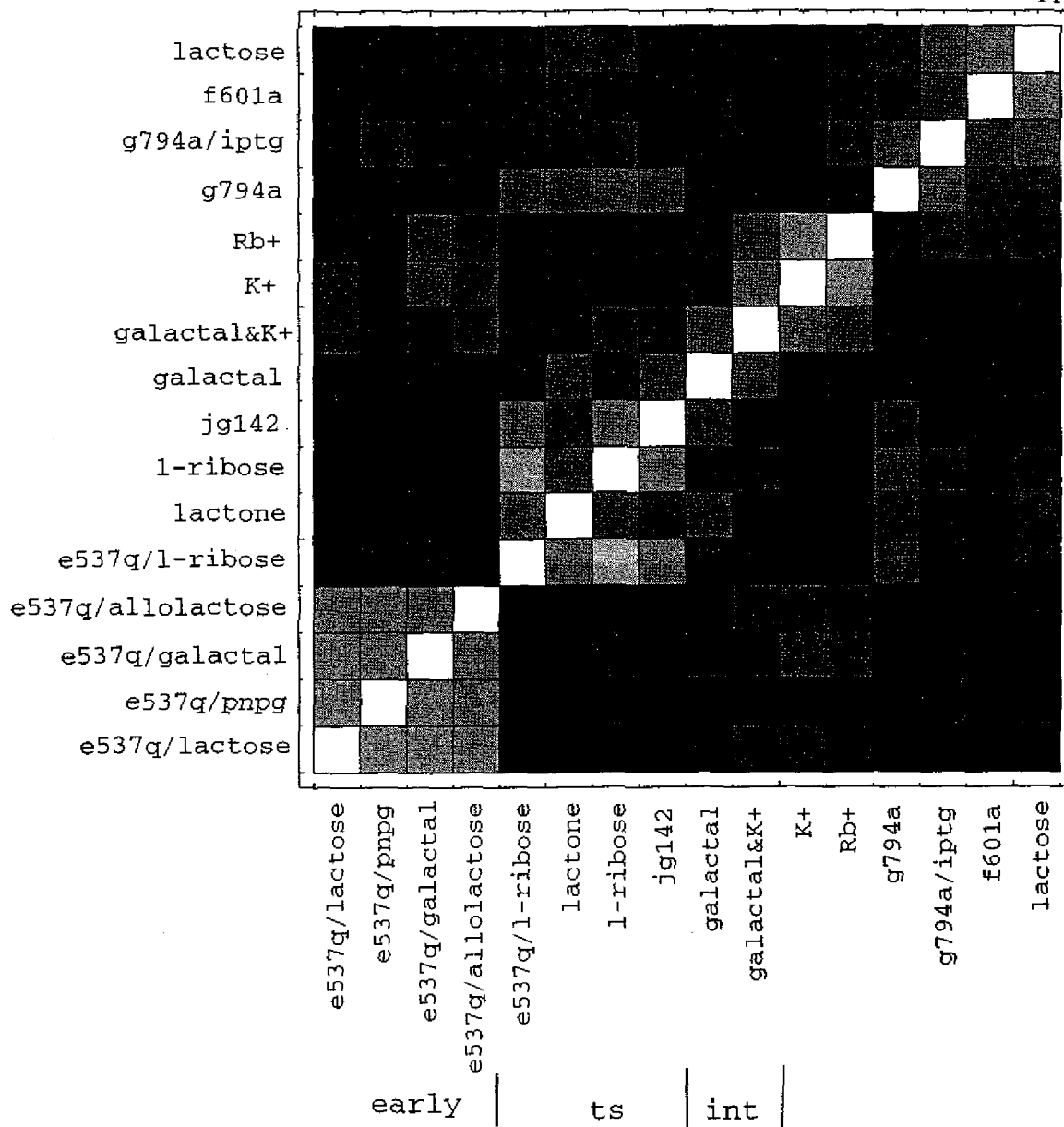
Figure 35. Correlations between delta-difference plots for Cα atoms of domain 3 (see the text for the formulation). Light squares mean the two ligands and/or mutants have a similar effect on native enzyme, whereas dark squares mean the effects are different. All the "early" complexes have a similar effect, as do the transition state analogs. The K⁺ and Rb⁺ complexes have an "early" effect, whereas G794A has a transition state effect. G794A with iptg, however, as a weak "early" effect. L-ribose has the same effect on the E537Q variant as it does on native, suggesting the changes due to ligand binding over shadow changes due to the mutation.

Energetics of the Switch

Inhibition studies can be used to estimate the free energy required for the conformational change:

$$\Delta G(open \leftrightarrow closed) \cong \Delta\Delta G_{binding} = RT\ln(\frac{Ki^{G794A}}{Ki^{Native}})$$

This assumes that inhibitor binding results in the same conformation for both enzymes. That is, iptg binds to the open conformation in both native and G794A, and lactone or L-ribose bind to the closed conformation in both cases. These studies, done in collaboration with Shamina Hakda and Gene Huber at the University of Calgary, CA, show that iptg binds 5x worse to G794A than native, because it must use some of its binding energy to open the loop. On other hand, the lactone and L-ribose bind 4x and 14x worse to native, because they must use their binding energy to close the loop. This suggests that in the absence of ligand, the free energy difference between the two enzyme conformations is about 1-1.5 kcal/mol, or 2-3 kT at 300 K. This is relatively small, suggesting the conformational change would be fairly accessible to thermal fluctuations.

Effects of Freezing, Solvent and Crystal Packing on the Switch

Some of the low temperature data sets show ambiguous density for part of the 794-804 loop. In these cases, a room temperature data set was also collected (usually to ~3 Å resolution), sometimes also collected in the presence of 30% DMSO. Fourfold

averaged F_o-F_o maps were calculated to determine the state of the conformational change. The results are listed in Table 7(c).

The open conformation, which is seen in all of the early complexes, is unaffected by solvent, freezing, or crystal packing (meaning that all four monomers are about the same). The closed conformation, however, is more variable. The transition state analogs L-ribose and galactonolactone clearly show the closed conformation at room temperature with or without DMSO. At low temperature, the electron density is more ambiguous, with residues ~794-797 appearing closed but 798-801 probably in multiple conformations. The loop in the tetrazole complex is closed at room temperature (without DMSO) but open at low temperature, possibly due to DMSO binding near both the tetrazole and Phe-601.

The behavior of the intermediates is more complex, depending on DMSO, bis-tris and temperature. Also, different monomers show different degrees of disorder for Phe-601, Arg-599 and the 794-804 loop. DMSO appears to destabilize the closed conformation by binding weakly on Phe-601, disrupting the Phe-601 – Ser-796 interaction. This binding could not be confirmed with F_o-F_o maps because high concentrations of DMSO change the cell dimensions. However, F_o-F_c density for many of the frozen data sets show fairly strong features on Phe-601, suggestive of DMSO binding. Kinetics with DMSO shows weak mixed inhibition in the 1 M range, which is ~1/3 the DMSO concentration used for data collection.

In the case of the F601A mutation, in the frozen complexes, which are in the presence of DMSO, there is sometimes density in place of the Phe-601 side chain, which

is, nevertheless, disconnected from the main chain. In these cases it appears that something is filling in for the F601 side chain, perhaps DMSO.

Most strikingly, bis-tris binds to the 2-deoxy-galactosyl intermediate, making interactions with Asn-102, His-418 and the attacking water molecule (Figure 36). There is little density for such a molecule in the native structure, suggesting the bis-tris binds better to the intermediate. This is supported by inhibition studies, which suggests that bis-tris acts an uncompetitive inhibitor, binding 5-10 x better to the intermediate than the ground state (Table 10). These studies also suggest that bis-tris can act as an intermolecular acceptor to form a bis-tris-galactoside at similar efficiency to the formation of allolactose when glucose is the acceptor. The 2-deoxygalactosyl-enzyme-bis-tris complex points to Asn-102 and His-418 playing a role in orienting the intermolecular acceptors.
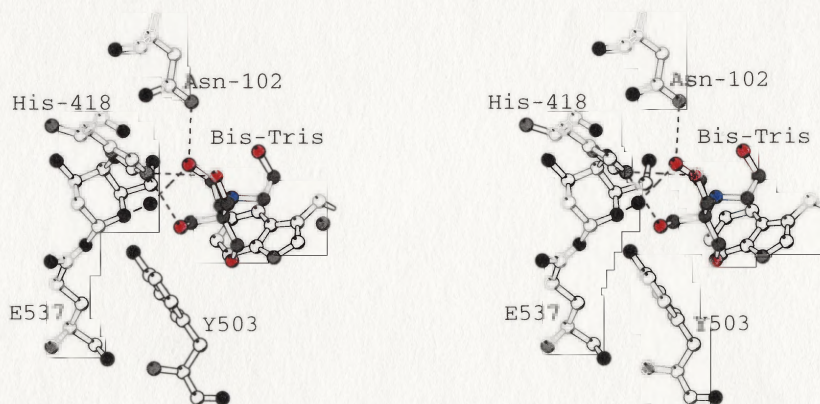
Figure 36. Stereo view of the the binding of the uncompetitive inhibitor bis-tris to the 2-deoxy-galactosyl intermediate. Despite its tremendous flexibility, the buffer molecule binds clearly in the active site, making polar interactions with Asn-102, the water molecule which will attack the intermediate, possibly the Trp-999 ring nitrogen and a bidentate interaction with His-418.

Table 10. Kinetics of β-galactosidase variants. See methods for experimental details. Ki is the competitive inhibition constant, Ki" is the uncompetitive inhibition constant and k_4 is the rate constant for the acceptor reaction between the galactosyl-enzyme intermediate and the given ligand. Results with an asterisk were provided by S. Hakda from the laboratory of G. Huber, University of Calgary.

	k_{cat} (onpg) (s-1)	K_m (onpg) (mM)	k_{cat}/K_m (onpg) (s^{-1}M^{-1})	k_{cat} (pnpg) (s^{-1})	K_m (pnpg) (mM)	k_{cat}/K_m (pnpg) (s^{-1}M^{-1})
Native	530	0.101	5.2×10^6	53	0.044	1.2×10^6
G794A				76	0.241	3.2×10^5
F601A				97	0.234	$4.1 \times 10_5$

	Ki (mM)	Ki^u (mM)	k_4 (s^{-1})
L-ribose			
Native	0.28*		
G794A	0.02*		
F601A	1		
lactone			
Native	0.13*		
G794A	0.035*		
iptg			
Native	0.11*		
G794A	0.59*		
Glucose			
Native	230*	21	450
G794A	43*	1*	22*
Bis-Tris			
Native	102	17	420
DMSO			
Native	900 (300)		

Transition State Stabilization

Kinetic studies done on G794A and other substitutions at position 794 (Asp, Asn and Glu) show that k_3 decreases, while k_2 slightly increases for lactose and pnpg hydrolysis [7, 19, 20]. These studies also show that the 794 variants bind transition state analog inhibitors better and substrate analog inhibitors worse than native enzyme. With the structural data this suggests that in the closed conformation the enzyme is better prepared to recognize the first transition state than in the open conformation and it is more reluctant to release the intermediate.

Allolactose Production

Although the position 794 variants are a slightly improved for hydrolysis of slow substrates, Figure 37 shows the variant G794A has diminished ability to produce allolactose relative to native enzyme. Surprisingly, acceptor studies with G794A suggest that glucose binds to the intermediate better than with native enzyme (Table 10). Taken together, the kinetics and structural data suggest that if the enzyme is in the closed conformation, hydrolysis is favored over transglycosylation via nonproductive binding of glucose. Allolactose production must occur primarily from a conformation other than the closed one.
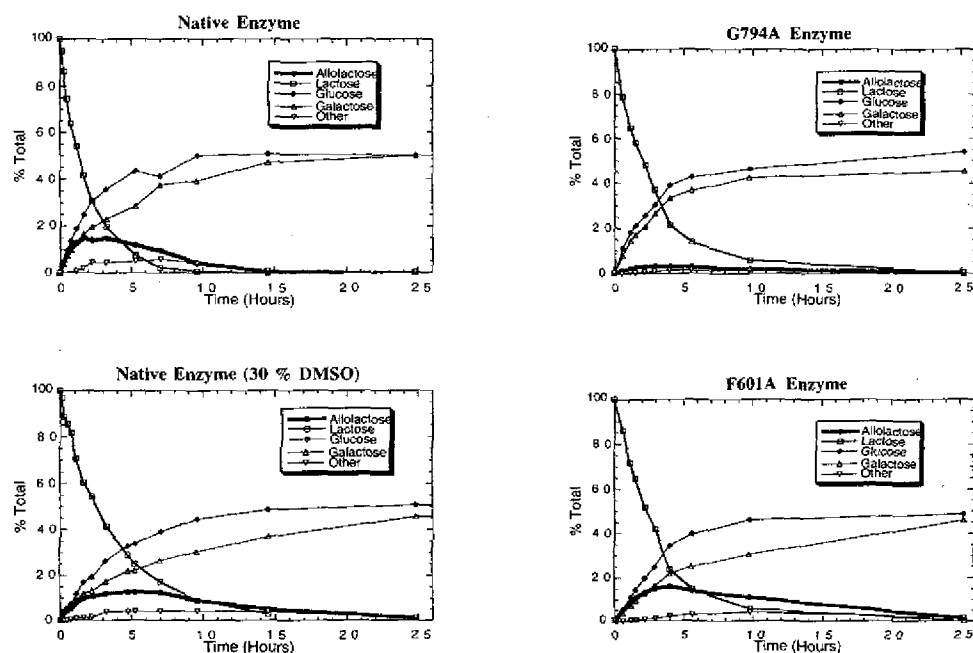
Figure 37. Activity assays for the action of β-galactosidase variants on lactose. The assay is a capillary electrophoresis based assay[21] and was done at room temperature with 50 mM Na_2HPO_4, 1mM $MgCl_2$ pH 7.0. Native enzyme both with and without 30% DMSO and F601A all produce significant amounts of allolactose, whereas G794A produces about 10-fold less.

Figure 38 shows a sequence alignment from various β-galactosidases in the regions of the conformational change. Most of the residues involved in the conformational switch (Phe-601, Asn-102, Asp-598 and Gly-794) are well conserved, except for the proteins coded by the ebg gene from E. coli and the K. lactis, a yeast gene. The region preceding Gly-794 is also well conserved except in ebg and yeast, although the loop itself is poorly conserved. As this region preceding Gly-794 underlies Trp-999 and Phe-601, it is probably important in shaping the active site pocket. Other key active site residues: Asp-201, His-357, Arg-388, Arg-391, Glu-416, His-418, Glu-461, Glu-537, Glu-540 are completely conserved in all sequences, suggesting that the basic mechanism for hydrolysis is the same.

```
S.the  FGSPQYVNT 103    598 DYEFCGNGIV 607   785 TRAYTDNDRG AGYPFEMAGW 804
C.ace  YDKCQYINT            DYNFSGNGLI           WRATTDNDRG NRHEFRCSQW
L.del  FGQPQYVNV            DYEFCGNGLV           WRALTDNDRG AGYGYDLARW
L.lac  YGQIQYINT            DYAFSGNGLM           WRATTDNDHG SGFSVKSAQW
E.col  YDAPIYTNV            DRQFCMNGLV           TRAPLDNDIG VSEATRIDPN
E.clo  FDTPIYTNC            DRQFCLNGLV           SRAPLDNDIG VSEATKIDPN
K.pne  YDAPIYTNC            DRQFCMNGLV           IRAPLDNDIG VSEVERIDPN
T.mar  YGKPIYTNC            DGNFCINGVV           WRVPTDNDIG NRMPQRLA..
ebg    HGKLQYTDE            NYNFCLDGLI           ...HRQEYEG LWQPNHLQ..
Athro  YGRPIYTNV            DGNFVMDGMI           WRAPTDNDKG QGFGAYGPED
K.lac  YGKPIYTNV            DGVFIMDGLC           WRPPTNNDEP RDFKN.....
```

Figure 38. Multiple sequence alignment in the regions of the conformational switch. All sequences are bacterial, except for K.lac, which is a sequence from yeast. S.the=*Streptococcus thermophilus*, C.ace=*Clostridium acetobutilicum*, L.del=*Lactobacillus delbrueckii*, L.lac=*Leuconostoc lactis*, E.col=*Escherichia coli*, E.clo=*Enterobacter cloacae*, K.pne=*Klebsiella pneumonia*, T.mar=*Thermotoga maritima*, ebg=E. coli ebg, Athro=*Athrobacter*, K.lac=*Kluyveromyces lactis*.

The ebg gene from *E. coli* encodes another β-galactosidase. It has a similar operon structure to the lacZ gene with its own repressor, and was identified in strains without lacZ. The ebg β-galactosidase cannot support growth on lactose. Selection experiments showed that the single point mutants Asp-102 (Asn and Trp-999 (Cys (lac Z numbering) allow ebg to support growth with iptg present[22]. These enzymes can hydrolyze lactose but cannot produce allolactose. The double mutant Asp-102 → Asn & Trp-999 → Cys can produce allolactose and supports growth on lactose without iptg present[23]. This suggests further that Asn-102 is important for allolactose production.

Since the sequence of 794-804 is poorly conserved except for 794, the identity of its sidechains are presumably unimportant. However, because the glycine is conserved, whether it is open or closed is probably important and this influence may exerted through Asn 102 positioning. The only sequence besides ebg without Gly-794 is the yeast sequence, which has a proline and a shorter loop, although the region preceding Gly-794 is well conserved. The rest of the sequences are from bacteria. Since gene organization is

different in eukaryotes, the yeast gene is probably regulated differently. For example, with yeast galactose is a gratuitous inducer[24], whereas with *E. coli*, galactose is a very weak inducer[25].

Summary – The Role of the Conformational Switch in Catalysis

The preceding sections have described a conformational switch in β-galactosidase which can be induced by substitutions at Gly-794. The 794 substitutions have the effect of increasing hydrolysis at the expense of allolactose production. This occurs by forcing the enzyme into its closed conformation, which favors non-productive binding of glucose Thus, Gly-794 appears to be conserved because any other residue would likely limit allolactose production and thus inhibit the ability of the bacterium express the genes of the lac operon.

Because ligand binding, in particular the binding of catalytic intermediates, also triggers this conformational switch, it may also play a role in the catalytic mechanism of native enzyme. In particular, if the enzyme is in its closed conformation, it is likely that allolactose will not be produced, and hydrolysis will be favored. This suggests that if the enzyme is in its open conformation, allolactose production would be favored over hydrolysis.

Whether the loop conformation affects catalysis with native enzyme depends on the time scale for loop closure relative to the second step. If the second step happens too fast, the loop may not have time to close. With the 2-F- and 2-deoxy galactosyl-intermediates, the closed conformation will very likely be visited, since the lifetime of

these complexes are hours-minutes. The lifetime of the normal galactosyl-enzyme, however is about 1 ms (k_3=900 s^{-1} - Table 9). This would suggest that the loop must fluctuate on a microsecond time scale in order for its effects to be felt.

Fluctuations in protein structure occur over many orders of magnitude in time. Bond vibration are picosecond events[1], while the rotation of methyl group occurs in nanoseconds[26]. The binding of hydrophobic ligands to a cavity in T4 lysozyme takes microseconds-milliseconds [27]. Other substrate induced conformational changes have been observed on 100 μs - 0.1 sec time scales[1], while flipping of buried aromatic rings can take seconds[28].

One of the best studied systems similar to the loop swing studied here is in triose phosphate isomerase. This enzyme includes a 10 residue loop which closes upon substrate binding to sequester the reaction from bulk solvent[18]. NMR experiments on this loop suggested it jumps between its two conformations at a rate of 3 x 10^4 s^{-1}, or a lifetime of 33 μs[29]. While this has no direct application to the 794-804 loop in β-galactosidase, it is suggestive that the loop swing might happen faster than 1 ms. Additionally, the 794-804 loop has relatively high B-factors, is relatively solvent exposed, and parts of it are disordered in several structures, suggesting it can fluctuate relatively easily.

Although both the F601A substitution and DMSO binding both appear to perturb the loop structure, neither has an appreciable effect on the relative rates of hydrolysis and allolactose production (Figure 37). This suggests that either the loop conformation is unimportant for this ratio, or that in these cases, although the loop conformation is

perturbed, it can still explore the conformations necessary for the enzyme to catalyze

hydrolysis and transglycosylation at about the same rate.

The Mechanism of Action for β–Galactosidase

Overall Description

Based on the work presented here and the extensive biochemical studies in the

literature, previously described mechanisms for hydrolysis and transglycosylation can be

clarified and enhanced. Figure 39 shows the presumed mechanism in light of the

observations presented here.

The substrate binds by stacking on Trp-999 (not shown in Figure 39 – see Figures

23-25, pp. 92-98) with several specific interactions made by the galactosyl hydroxyls and

few specific interactions made by the aglycon. Should the substrate bind exactly as in the

"early complexes", catalysis would be unable to proceed, because a water molecule

would be trapped deeper in the active site where the galactosyl group is presumed to

occupy to form the intermediate (Figure 39(a)). In the event that the substrate binds

without this water molecule present there will be unsatisfied interactions unless the

substrate moves in so its hydroxyls occupy this water binding site and another water

binding site which is unoccupied in the early complexes (Figure 39(b)). These

interactions made by the 2,3,4 and 6 hydroxyls draw the substrate deep into the active

site, positioning C1 near the nucleophile, Glu-537, the glycosidic oxygen near Glu-461

and the sugar ring on a different tryptophan, Trp-568. Whether the substrate binds

shallow or deep (Figure 39(a) or (b)), the active site will be unavailable for another substrate molecule, and saturation kinetics will result.

For the fastest substrates, which have degalactosylation rate limiting, bond cleavage could occur via an SN1 process. For onpg, this could be assisted by the interaction between His-418 and a nitro group on the leaving group (not shown in Figure 39, see Figure 24, pg. 93). Slower substrates such as lactose, which have more basic leaving groups or do not have the correct geometry to utilize His-418 need to have the acid assistance of Glu-461 and must move deeper into the active site before bond cleavage can occur.

The active site has been preorganized for this task[30], and once the substrate is in the correct position, Glu-461 can donate a proton to the glycosidic oxygen in concert with formation of the intermediate with Glu-537 (Figure 39(c)). This is likely predominantly an SN2 process and results in the formation of the galactosyl-enzyme intermediate (Figure 39(d)). However, the possibility of a substrate assisted reaction stabilized by an intramolecular interaction between the 2-OH and C1 cannot be excluded.

As the substrate moves through the transition state towards the intermediate the 6-OH moves to a new position on the sodium ion, opening room for Phe-601 to swing into its new rotamer well which destabilizes Arg-599 and favors the closed loop conformation (Figure 39(c) & (d)). The closed loop introduces a polar interaction between the backbone carbonyl of residue 796 and the Asn-102 side chain, probably favoring a single Asn-102 rotamer.

As with the unsatisfied binding interactions drawing the substrate deeper into the active site, the conformational change is another example of potential positive

interactions which can be realized only by moving further along in the reaction.

Interactions between the enzyme and the 2-OH probably fall in the same category. As

the bond is broken, the 2-OH-Glu537 H-bond strength will probably increase due to the

developing negative charge on Glu-537, or possibly optimizing the geometry of this H-

bond (see below).

Figure 39 (following 2 pages) A proposed mechanism for β-galactosidase based on the previous biochemistry and the work presented here. (a) The "early complexes" identified crystallographically. Solvents sites are shown as circles – filled (occupied) or open (empty). One of the filled sites must be empty for the reaction to proceed. (b) Another proposed binding mode on the way to the transition state. (c) The transition state for step 1 (d) The covalent intermediate with the enzyme in its "closed" conformation. The open conformation should also be possible. (e) The intermediate. (f) The transition state for step 2. (g) The product galactose, bound in the active site. (h) The enzyme in its unbound state. Ordered water molecules occupy much of the binding site.

e

HOH

+ HOR

f

g

+ Galactose

h

126
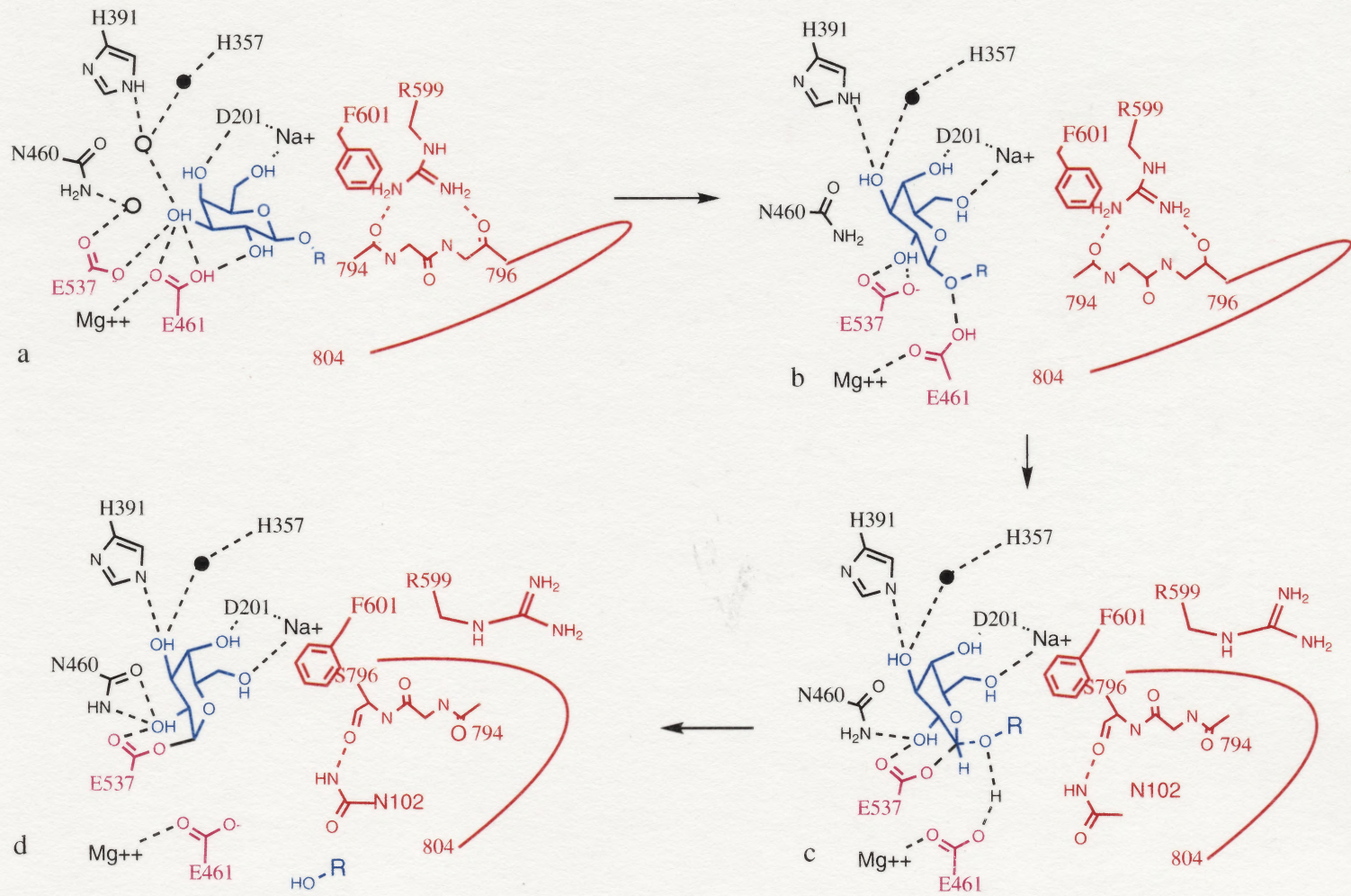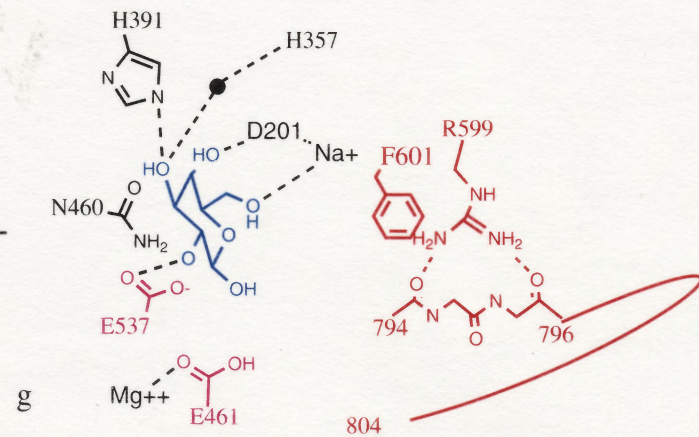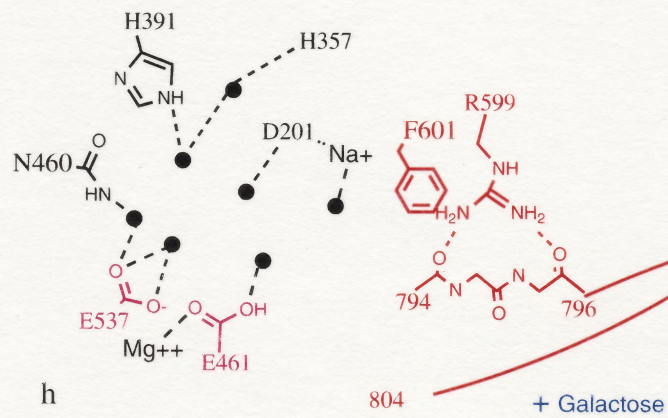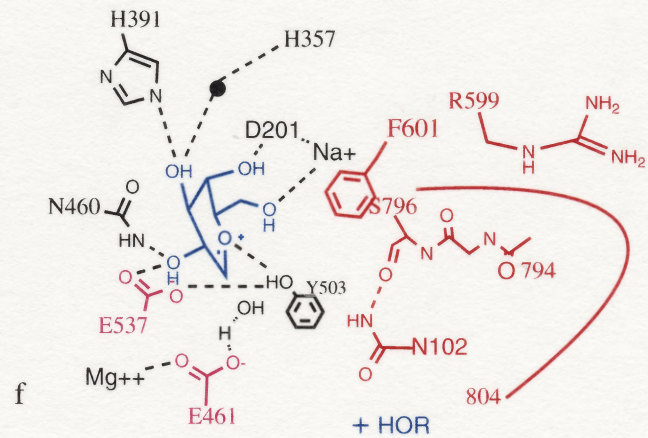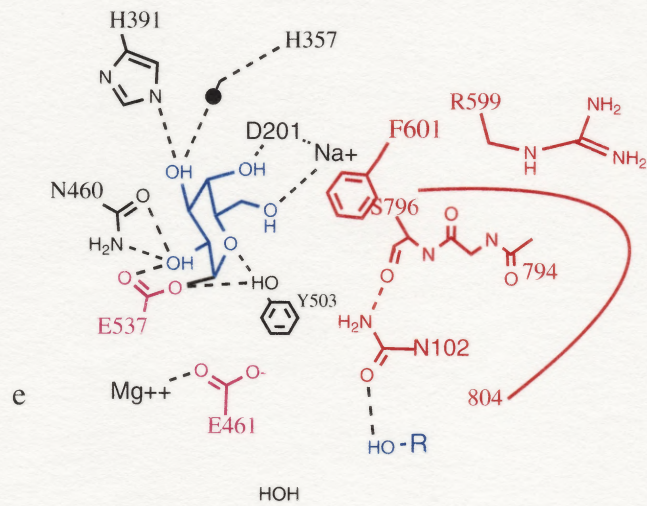
The second half of the reaction (Figure 39(d)-(f)) occurs via an SN1 process,

in which the galactosyl group is transferred from the nucleophile Glu-537 to an acceptor

molecule. During this process, a planar oxocarbonium ion develops, which is stabilized

by an interaction between Glu-537, Tyr-503 and the galactosyl ring oxygen. The

transition state is also stabilized by interactions between the 2-OH and Glu-537 and Asn-

460. Optimal geometry between the 2-OH and Glu-537 may be realized in the transition

state when the hydrogen bond between the two is coplanar with the Glu-537 carboxylate

group. Glu-461 facilitates attack of the oxocarbonium ion by the acceptor molecule by

abstracting a proton, which also regenerates the protonated Glu-461 (Figures 39(f) & (g)).

Whether the acceptor molecule is water, or in the case of allolactose production,

glucose, is determined by binding interactions between the acceptors and residues in the

active site – possibly Asn-102 and His-418. Glucose binds in multiple modes to the

intermediate, not all of which are productive for forming allolactose. The conformation

of the 794-804 loop has an impact on glucose binding, and if it is in its closed form, as in

Figure 39(e)& (f), the outcome will likely be hydrolysis, perhaps because the Asn-102

amino group is unavailable to properly bind the glucose for allolactose production.

In the case of hydrolysis, when the bond is cleaved the leaving group diffuses

away and a water molecule comes in, hydrogen bonding to Glu-461 and the ring oxygen

of the galactosyl moiety. The intermediate is released in an SN1 process, and is trapped

by the water molecule, which has been activated by the abstraction of a proton by Glu-

461. The move toward the transition state lengthens the O-C1 bond forcing the sugar to

rotate. As the pucker in the galactose product forms from the planar transition state, the

sugar rotates farther, shifting the 6-OH along the Na⁺ towards Phe-601. This causes

Phe-601 to open, allowing the galactose to diffuse out of the active site.

Other Considerations about the Mechanism

What about the Michaelis Complex ?

The discussion thus far has avoided explicitly defining any of the inhibitor

complexes as a model for the Michaelis complex. This a postulated enzyme-substrate

complex which rapidly associates and dissociates, but eventually leads directly to the

transition state. With the results presented here, the observed early complexes require

repositioning of the substrate before the transition state can be reached (at least for slow

substrates such as lactose). This repositioning can be modeled by overlaying lactose on

the galactose complex and adjusting the glucose molecule (Figure 39b). In this way,

lactose can be modeled without steric clashes at the glucose. Although the internal

glucose-galactose hydrogen bond must be broken in this process, there is potential for a

new hydrogen bond between the glucose and Asn-102.

A binding mode like the repositioned lactose is very likely on the reaction path,

however it was not seen in any of the early complexes. For 2-F-lactose, this could be

because the 2-substituent probably has to donate a hydrogen bond to Glu-537 or Glu-461,

which is impossible with the fluorine. For the thio-galactosides this could be because the

Glu461-S interaction is not as strong as a Glu461-O interaction. However, for the E537Q

variants it is less clear why the repositioned conformation is not observed, because all the

required polar interactions appear to be possible. This depends, however, on the protonation state of Glu-461 which could be altered by the Glu-537 alteration.

The simplest model to account for these results is that the Michaelis complex consists of two binding modes in equilibrium with each other. The saturation kinetics will not be affected by the relative population of these two modes, although pre-steady state kinetics might be able to detect them. In one extreme, the early mode observed by crystallography would be extremely high energy. In this case the rapid equilibrium would be with the repositioned mode directly and the early mode would never be visited. In the other extreme, the repositioned mode would be high energy and visited only on the way to the transition state. In the middle extreme, the free energy between the two modes would be zero, and they would be equally populated, although the reaction would only proceed from the repositioned mode.

Although none of the evidence presented here can distinguish between the two modes, it is worth pointing out the mutagenesis studies on His-357[31]. This residue does not interact directly with substrate. Instead, it ligands a water molecule which forms an H-bond to the 3-OH in intermediate complexes, but is quite far from the galactosyl in the early complexes (Figure 25, pg. 98 & Figure 39). The mutagenesis studies showed that with His-357 variants, the rate for step 1 decreased much more than for step 2. If the intermediate and both transition states are affected about the same, then step 2 would be unaffected because the barrier from intermediate to the second transition state would be unchanged. Since the intermediates and the transition state analogs have similar structures, it is reasonable to expect them to be affected the same by the His-357 variants. In this vein, for step 1 to decrease, the ground state must be less affected than the

transition state. This is more consistent with a Michaelis complex like the early mode than the repositioned mode, which would make about the same interactions to the His-357 water molecule as the intermediates.

The Role of Mg^{++} and Na^+.

The complexes presented here strongly suggest that the active site acid is Glu-461 and not the Mg^{++}. Recent biochemical work has also supported Glu-461 as a Bronsted acid rather than Mg^{++} as a Lewis acid[32]. The obvious role for the Mg^{++} is then structural, as it ligands Glu-461 as well as two water molecules which provide interactions for the 3 and 4-OHs (Figures 23 and 25, pg. 92). Whether it has a further role is unclear. It could be involved in tuning the pKa of Glu-461 and it has been suggested that the Mg^{++}-Glu-461 interaction distance might decrease on going to the intermediate, decreasing the pKa of Glu-461, allowing it to be deprotonated in the intermediate and setting it up to abstract a proton from the acceptor molecule[32]. However, in the structures presented here no decrease in the Mg^{++}-Glu-461 distance in going to the intermediate is observed (Table 8).

Another possible role of the Mg^{++} would be to provide correct geometry between Glu-461 and His-418 to bind glucose correctly for allolactose production. Removal of Mg^{++} decreases the ratio of transglycosylation/hydrolysis, highlighting its importance for producing allolactose. Recent studies by the Huber lab have also shown that His-418 is important for glucose binding[20]. However, such binding is apparently too transient or poorly occupied to produce a good model via X-ray crystallography.

The role of the Na^+ is to directly ligand the galactosyl 6-OH (or a water in unliganded enzyme). This distance increases from 2.2-2.4 Å in the early complexes to 2.4-2.5 Å in the intermediates and ts analogs and back to 2.3-2.4 Å in the products(Table 8). Although this change is small, it is consistent across all four monomers in the tetramer (data not shown). Assuming that increasing this distance is unfavorable, then replacing the Na^+ with K^+, which is larger, should make it easier to get to the intermediate and more difficult to leave it. This would speed up the first step and slow down the second step, which is consistent with the kinetics. Both lactose and pnpg, which are limited by the first step, prefer K^+ and onpg, which is limited by the second step prefers Na^+[33].

Energetics of Transition State Stabilization

Several experiments have been done to suggest limiting values for transition state stabilization provided by interactions between the various galactosyl substituents and enzyme groups. McCarter and Withers measured kinetics for a series of deoxy and fluoro analogs of 2,4 dinitrophenyl galactoside[14]. These substitutions affect both electrostatics and binding interactions. The electron withdrawing fluorine group will destabilize any positive carbocation that develops in the transition state, while the electropositive hydrogen in the deoxy compounds will stabilize the carbocation. Therefore, the deoxy variants represent a lower limit for the transition state stabilization resulting from interactions between the enzyme and the altered substituent. Their results suggested that interactions between the enzyme and the 3, 4 and 6 hydroxyls provide at

least 4 kcal/mol of transition state stabilization for galactosylation, whereas the 2-OH

provides about 8 kcal/mol. The special role of the 2-substituent correlates well with the

structures presented here, which suggest the 2-substituent changes its interactions the

most between the various complexes. In this model, the stabilization provided by the 2-

OH is due to interactions with Asn-460 and Glu-537 and the role of the other hydroxyls is

to properly orient the galactosyl moiety.

The Role of Strain and the Stereoelectronic Effect

Enzyme catalyzed hydrolysis of glycosidic bonds is often thought to involve

distortion of the sugar ring next to the cleaved bond prior to the transition state. This

distortion, which has been observed crystallographically in a number of systems, is

thought to be required to allow the substrate to bind to the active site. If the distortion

has the correct geometry for the transition state, then some of the work to reach the

transition state will have been done in the initial binding step.

Since many glycosidases include an SN1 bond cleavage, if binding interactions

distort the sugar from a relaxed chair to a high energy half chair (which is planar at C1)

catalysis will be facilitated. Furthermore, for retaining β–glycosidases such as

β–galactosidase, stereoelectronic considerations suggest that an inverted boat

conformation for the sugar ring would facilitate catalysis by providing orbital overlap

between the ring oxygen and the anomeric center to help form the oxocarbonium cation.

This idea has been supported with the observation of several inverted boat inhibitor

complexes in various glycosidases[34].

As was discussed above, the first step for β–galactosidase probably does not involve an oxocarbonium ion, so distortion to an inverted boat for optimal stereoelectronic would not be advantageous. However, distortion to a half chair could facilitate an SN2 reaction since the anomeric center should be planar at the transition state.

Nevertheless, the data presented here show no direct evidence for galactosyl distortion in the mechanism of β–galactosidase. In all of the early complexes, the galactosyl ring binds in a normal, full chair conformation, as does the product galactose. The repositioned lactose also offers no suggestion that galactosyl distortion would be required for binding in this hypothetical mode.

Assuming that distortion is the rule, then one possible reason for the lack distortion β–galactosidase catalysis is that the enzyme acts on a small substrate with strong specificity for only one sugar. Often the distortion is observed in enzymes that act on long substrates. The distortion is paid for with substrate binding energy comes the undistorted saccharide units. In the case of β–galactosidase, there are only two saccharides, and the glucose has very few specific contacts, so glucose binding might not be able to pay for distortion in the galactose ring

The Asymmetry in the Transition States

As presented, and in accord with the biochemistry, the two steps of β–galactosidase catalysis appear to have slightly different mechanisms with different

transition states. Step 1 appears to be predominantly SN2 (except perhaps for very fast substrates), whereas step 2 appears to be predominantly SN1.

One reason for this asymmetry may be because the geometry at the anomeric center differs. Step 1 is a substitution on a β-linked substrate, while step 2 is a substitution on an α-linked substrate. Stereoelectronic considerations suggest that in the absence of distortion an α-linked substrate should form an oxocarbonium ion more easily than a β-linked substrate. So cleavage of the galactosyl-enzyme intermediate can proceed via an SN1 process, whereas the initial substrate cleavage must proceed via SN2. Additionally, the leaving groups are different for the two reactions – glucose for step 1 and protein for step 2. In this case, the protein appears to be a better leaving group than the glucose molecule.

Charge States of Key Residues

A poorly understood but critical aspect of catalysis by β–galactosidase and other glycosidases is how the electrostatic environment changes during the reaction. Usually there is a residue, such as Glu-461 which acts as both an acid and a base during the reaction, which means it must be alternatively protonated and unprotonated at different points in the reaction.

Crystallographic studies can give little direct information about the detailed electrostatic environment in the active site. The locations of protons must usually be inferred from putative hydrogen bonding patterns, and these are often ambiguous. With

smaller proteins, however, NMR can be used to learn about the charge states key

residues in the active site.

NMR studies on an enzyme similar to β–galactosidase, a xylanase from *Bacillus*

circulans, suggest that the pKa of the active site acid cycles during the reaction.[35]. The

explanation suggests that the nucleophile (Glu-537 in β–galactosidase), which is

negatively charged in the unliganded enzyme, elevates the pKa of the nearby acid to 6.7

from the unperturbed 4.1, making it protonated at that enzyme's pH optimum of 5.7. In

the intermediate, the nucleophile forms a covalent bond and has a smaller effect on the

acid, lowering its pKa to 4.2, making it deprotonated and available to abstract a proton

from the acceptor molecule. This sort of mechanism has been also discussed for

β–galactosidase, although NMR experiments are not currently possible because of its

large size[36].

Glu-537, the β–galactosidase nucleophile, will have a large effect on the active

site electrostatics because it should be negatively charge in the free enzyme. It does,

however, interact with the OH of Tyr 503 and with Arg 388. Substitutions at 503 have

been studied and suggest that k_3 decreases more than k_2 if the substitution does not have

transferable protons (Phe, Ser and Ile) but decreases about the same if the substitution

does have transferable protons (His and Cys). This suggests a role for Tyr-503 in

helping to stabilize the negative charge which develops on the nucleophile as the

galactosyl-enzyme intermediate is hydrolyzed. Arg-388 may play a similar role, from the

other side of the Glu-537 carboxylate group. Alternatively, since Glu-537, Tyr-503 and

the ring oxygen form a tripartite interaction in some of the intermediate complexes Glu-

537 and Tyr-503 could mutually stabilize the developing oxocarbonium ion in the

transition state for degalactosylation. This would be expected to be less important for

k_2, if galactosylation occurs via an SN2 process.

Thus, in this model, the electrostatic environment of β–galactosidase is dependent

synergistically on a number of residues. The pKa of Glu-461 is elevated by the negative

charge on Glu-537. But this charge is also stabilized by Tyr-503. Formation of the

covalent bond in the intermediate decreases the pKa of Glu-461 via shielding of the

nucleophile by the intermediate and perhaps other factors. Breaking the galactosyl-

enzyme intermediate bond is facilitated by hydrogen bond donation from Tyr-503 to Glu-

537.

Considerations of the Mechanism in the Context of the Biology

It is often noted that an enzyme has evolved to perfection when it is diffusion

controlled. That is, the rate limiting step is diffusion, so that every time a substrate binds

to the active site a reaction happens. As noted in the introduction, β–galactosidase is not

diffusion controlled and there are probably ~1000 binding events between lactose and the

active site before a reaction takes place. Although it appears, therefore, that

β–galactosidase has not evolved to perfection, further consideration suggests that

diffusion control per se may not be advantageous for the bacterium.

Beta-galactosidase has three roles in the biology of *E. coli*. First, is to hydrolyze

lactose, second is to produce the inducer, allolactose and third is to hydrolyze allolactose.

Therefore, β–galactosidase must not only produce allolactose, but also hydrolyze it. If

the enzyme were diffusion controlled for hydrolysis, then allolactose might never make it

out of the active site, because it would be hydrolyzed as soon as it was produced.

Since the active site is in such a deep pocket on the surface, apparently making it more

difficult to bind appropriately, this might make it more likely for the allolactose to escape

once produced.

The progression of the galactosyl group to deep within the active site pocket in

the intermediate offers an explanation for why allolactose is produced and not lactose or

other transglycosylation products. Glucose apparently is exploring many binding modes

while the enzyme is in the intermediate state. Putting aside specific binding interactions

between the glucose and the enzyme, the most likely substituent to react with the

intermediate is the 6-OH, since it has an extra atom connecting it to the sugar ring.

Methods

Crystallography

The general method for determining the structure of bound ligands and altered

enzymes was described in the first section of the chapter. Most of the ligands were

purchased from Sigma. The 2-F- derivatives (2-F-lactose and the dinitrophenyl-2-F-

galatoside used to generate the 2-F-galactosyl-enzyme intermediate) were synthesized by

John McCarter, Lloyd McKenzie and Steve Withers at the University of British

Columbia, Vancouver, CA. The galactotetrazole was synthesized by Thomas Heightman

in the laboratory of Andrea Vasella at the Swiss Federal Institute of Technology, Zurich,

Switzerland. The "BBG" inhibitor was synthesized by Jörg Greul and Volker Jäger at the

University of Stüttgart. Allolactose was synthesized enzymatically using β-galactosidase after Huber et al (1975)[37]. To 50 mL of 50 mM onpg and 170 mM glucose in 50 mM Na_2HPO_4, 1mM $MgCl_2$, 1 mg β-galactosidase was added at room temperature. This was allowed to stir for about 45 minutes and the reaction was halted by heating to 80° C. The nitrophenol was extracted (twice) with 50 mL chloroform by stirring at room temperature for 30 minutes. The resulting aqueous solution was filtered (0.2 μm) and lyophilized. This was then run on a gel filtration column (Biorad P3) in water and the fractions were identified using circular dichroism and capillary electrophoresis assay (CE-see below). The allolactose peak was identified as allolactose with mass spectometry (the mass was identical with lactose) and using thin layer chromatography (the allolactose ran slower than lactose [38]). As judged by the CE, the allolactose was 98% pure.

All ligands were soaked into the crystal, usually stepwise, to the concentration shown in Table 7. Typically this was 100 x K_s where K_s is the dissociation constant for the ligand. For the time sensitive ligands, dnp-2-F-galactoside, galactal and lactose, soaking times of 30, 30 and 10 minutes were used, without stepwise equilibration. Generally, if the final ligand concentration was more than 100-200 mM, stepwise equilibration aided in minimizing the mosaicity and maximizing the diffraction. Stepwise equilibration was not necessary for lower ligand concentrations. Metal soaks were done similarly. Based on the soaking protocol, the K^+ soak had < 16 nM NaCl. The $P2_12_12_1$ EDTA soak had < 25 μM Mg^{++} and 50 mM EDTA.

Data collection was carried out in house and at synchrotron sources. The $P2_1$ data were collected at the Photon Factory, Beam Line 6A2, in Tsukuba, Japan. $P2_12_12_1$ data

were collected at the Stanford Synchrotron Radiation Laboratory Beam Lines 1-5, 7-1, 9-1 and 9-2 and at the Advanced Light Source Beam Line 5.0.2. Data were processed either with Mosflm/Scala or Denzo/Scalepack, generally to edge of the detector. Due to the high spot density, it was often necessary to decrease the size of the integration box in order to increase the completeness (Figure 19, pg. 75). In one case (the 2-deoxy-galactosyl intermediate with K^+) the data quality were improved by using the program PrOW (Profile fitting for Overlapped or Weak data)[39], which can deconvolute overlapped reflections.

Refinements were done generally as outlined in Chapter 2. The starting structure was native protein with all ordered solvent. After rigid body refinement at the tetramer, monomer, domain and secondary structure level, individual atom parameters (x,y,z and B) were refined. Each residue was then inspected by overlaying the four chains onto each other and stepping through the sequence, examining all four versions of each residue in succession. The ligand was then built into F_o-F_c and F_o-F_o electron density and refinement continued. Geometry restraints for the ligands were usually derived from the Cambridge Small Molecule Database. Cycles of model building and water molecule addition and deletion using ARP were repeated until fewer than 50 water molecules were added. The DMSO molecules and solvent ions were inspected and adjusted, flipped chiral centers were fixed and after a few more rounds of refinement, the model was judged to be finished. If the maximum resolution was lower than 2.4 Å, non-crystallographic symmetry restraints were used, and if the resolution was lower than 2.9 Å, non-crystallographic symmetry constraints were used. Prior to refinement, a test set

was set aside for R-free calculations (this was matched to a template R-free set, so that all complexes have the same test set).

Kinetics

Enzyme kinetics were carried out using onpg, pnpg and lactose as substrates. The onpg and pnpg assays were done with absorption spectroscopy by monitoring the signal from nitrophenol (the first product of the reaction) at 420 nm over time. Assays were done at 25° C (or sometimes room temperature: 21-22° C) in 50 mM Na_2HPO_4, 1 mM $MgCl_2$. pH 7. Initial velocities were obtained with a linear fit of a 10 minute scan. Low enzyme concentrations were used (usually 50 pM) to ensure linearity. Usually 6 different substrate concentrations were used and the kinetic parameters were determined by directly fitting the Michaelis-Menten equation to the data (see Appendix 1). Competitive inhibition constants were determined similarly using the parameters obtained in the absence of inhibitor. Uncompetitive inhibition is more complex, because of the possibility of the inhibitor acting as an acceptor, and this analysis is described in Appendix 1. G794A kinetics with onpg and pnpg were performed by Shamina Hakda in the laboratory of Gene Huber at the University of Calgary.

With lactose as the substrate, a reaction profile over time was measured using a capillary electrophoresis(CE) assay[21]. Briefly, a solution of 0.375 M lactose in 50 mM Na_2HPO_4, 1 mM $MgCl_2$ and 250 nM β-galactosidase was allowed to react. Samples were taken out (30 μL) at various times and the reaction was stopped by adding 60 μL 10% acetate in methanol and putting to 60°C. After about 24 hours of time points, the samples

were derivatized by adding 30 μL derivatizing solution (1% w/v of NaBH$_3$CN in

ABEE stock solution, which is 10 % w/v ethyl 4-aminobenzoate and 10% acetate in

methanol) and heating at 70°C for ~30 min. This adds a charged aromatic ring to the

reducing end of the sugar, allowing it to migrate in an electric field and be monitored

with absorption spectroscopy at 306 nm. The derivatized sugar was extracted with

300/300 μL chloroform/water and centrifuged for 5 minutes. This was then run on the

CE using a 23 mm capillary at 50-100 μA in a lithium borate buffer (pH 10.5).

Galactose, glucose and lactose standards were used to identify the peaks, with the

remaining major peak allolactose (identified as described above). Integration of these

peaks gave relative amounts of these sugars, producing the plots in Figure 37, pg. 118.

Protein Purification and Mutagenesis

Protein was purified as in Chapter 2. The E537Q, F601A and G794A variants

were constructed using the QuikChange mutagenesis kit from Stratagene with assistance

from Leslie Gay. The resulting plasmid was sequenced across the mutation site to verify

that the desired change had occurred. Because the complete gene was not sequence, there

is the possibility of accidental mutations.

References

1.   Fersht, A. *Enzyme Structure and Mechanism*, (W.H. Freeman & Company, New York, 1985).

2.   Xue, Q. & Yeung, E. Differences in the chemical reactivity of individual molecules of an enzyme. *Nature* **373**, 681-683 (1995).

3.   Lu, H.P., Xun, L. & Xie, X.S. Single-Molecule Enzymatic Dynamics. *Science* **282**, 1877-1882 (1998).

4.   Srajer, V. *et al.* Photolysis of the Carbon Monoxide Complex of Myoglobin: Nanosecond Time-Resolved Crystallography. *Science* **274**, 1726-1729 (1996).

5.   Yuan, J., Martinez-Bilbao, M. & Huber, R.E. Substitutions for Glu-537 of beta-galactosidase from Escherichia coli cause large decreases in catalytic activity. *Biochem J* **299**, 527-531 (1994).

6.   Sinnott, M.L. & Withers, S.G. The necessity of magnesium cation for acid assistance aglycone departure in catalysis by Escherichia coli (lacZ) beta-galactosidase. *Biochem J* **175**, 539-546 (1978).

7.   Martinez-Bilbao, M. & Huber, R.E. Substitutions for Gly-794 show that binding interactions are important determinants of the catalytic action of beta-galactosidase (Escherichia coli). *Biochem Cell Biol* **72**, 313-319 (1994).

8.   Huber, R.E., Parfett, C., Woulfe-Flanagan, H. & Thompson, D.J. Interaction of divalent cations with beta-galactosidase (Escherichia coli). *Biochemistry* **18**, 4090-4095 (1979).

9.   Deschavanne, P.J., Viratelle, O.M. & Yon, J.M. Conformational adaptability of the active site of beta-galactosidase. Interaction of the enzyme with some substrate analogous effectors. *J Biol Chem* **253**, 833-837 (1978).

10.   Hakda, S. Master's, University of Calgary (1997).

11.   Heightman, T. & Vasella, A. Personal Communication. , 1994).

12.   Jaeger, V. Personal Communication. , 1998).

13.   Wentworth, D.F. & Wolfenden, R. Slow binding of D-galactal, a "reversible" inhibitor of bacterial beta-galactosidase. *Biochemistry* **13**, 4715-4720 (1974).

14.   McCarter, J.D., Adam, M.J. & Withers, S.G. Binding energy and catalysis. Fluorinated and deoxygenated glycosides as mechanistic probes of Escherichia coli (lacZ) beta-galactosidase. *Biochem J* **286**, 721-727 (1992).

15.   Richard, J.P., Westerfield, J.G. & Lin, S. Structure-Reactivity Relationships for ß-Galactosidase (*Escherichia coli*, lac Z). 2. Reactions of the Galactosyl-Enzyme Intermediate with Alcohols and Azide Ion. *Biochemistry* **34**, 11703-11712 (1995).

16.   Huber, R.E., Gaunt, M.T. & Hurlburt, K.L. Binding and reactivity at the "glucose" site of galactosyl-beta-galactosidase (Escherichia coli). *Arch Biochem Biophys* **234**, 151-160 (1984).

17.   Huber, R.E. & Brockbank, R.L. Strong inhibitory effect of furanoses and sugar lactones on beta-galactosidase Escherichia coli. *Biochemistry* **26**, 1526-1531 (1987).

18.   Sampson, N. & Knowles, J. Segmental Movement: Definition of the Structural Requirements for Loop Closure in Catalysis by Triose Phosphate Isomerase. *Biochemistry* **31**, 8482-8487 (1992).

19.   Martinez-Bilbao, M., Holdsworth, R.E., Edwards, L.A. & Huber, R.E. A highly reactive beta-galactosidase (Escherichia coli) resulting from a substitution of an aspartic acid for Gly-794. *J Biol Chem* **266**, 4979-4986 (1991).

20.   Huber, R. Personal Communication. , 2000).

21.   Zeleny, R., Altmann, F. & Praznik, W. A capillary electrophoretic study on the specificity of beta-galactosidases from Aspergillus oryzae, Escherichia coli, Streptococcus pneumoniae and Canavalia ensiformis (jack bean). *Anal Biochem* **246**, 96-101 (1997).

22.   Hall, B.G., Betts, P.W. & Wootton, J.C. DNA Sequence Analysis of Artificially Evolved *ebg* Enzyme and *ebg* Repressor Genes. *Genetics* **123**, 635-648 (1989).

23.   Rolseth, S.J., Fried, V.A. & Hall, B.G. A Mutant Ebg Enzyme That Converts Lactose into an Inducer of the lac Operon. *Journal of Bacteriology* **142**, 1036-1039 (1980).

24.   Hsieh, H.B. & Da Silva, N.A. Development of a LAC4 promoter-based gratuitous induction system in Kluyveromyces lactis. *Biotechnol Bioeng* **67**, 408-416 (2000).

25.   Müller-Hill, B., Rickenberg, H.V. & Wallenfels, K. Specificity of the Induction of the Enzymes of the Lac Operon in *Escherichia coli*. *JMB* **10**, 303-318 (1964).

26.   Nicholson, L.K. *et al.* Dynamics of methyl groups in proteins as studied by proton-detected 13C NMR spectroscopy. Application to the leucine residues of staphylococcal nuclease. *Biochemistry* **31**, 5253-5263 (1992).

27.   Feher, V.A., Baldwin, E.P. & Dahlquist, F.W. Access of ligands to cavities within the core of a protein is rapid. *Nat Struct Biol* **3**, 516-521 (1996).

28.   Wagner, G., DeMarco, A. & Wuthrich, K. Dynamics of the aromatic amino acid residues in the globular conformation of the basic pancreatic trypsin inhibitor (BPTI). I. 1H NMR studies. *Biophys Struct Mech* **2**, 139-158 (1976).

29.   Williams, J. & McDermott, A. Dynamics of the flexible loop of triosphosphate isomerase: the loop motion is not ligand gated. *Biochemistry* **34**, 8309-8319 (1995).

30.   Cannon, W.R., Singleton, S.F. & Benkovic, S.J. A perspective on biological catalysis. *Nat Struct Biol* **3**, 821-833 (1996).

31.   Roth, N.J., Rob, B. & Huber, R.E. His-357 of beta-galactosidase (Escherichia coli) interacts with the C3 hydroxyl in the transition state and helps to mediate catalysis. *Biochemistry* **37**, 10099-10107 (1998).

32.   Richard, J.P., Huber, R.E., Lin, S., Heo, C. & Amyes, T.L. Structure-Reactivity Relationships for ß-Galactosidase (*Escherichia coli*, lac Z). 3. Evidence that Glu-461 Participates in Bronsted Acid-Base Catalysis of ß-D-Galactopyranosyl Group Transfer. *Biochemistry* **35**, 12377-12386 (1996).

33.   Cohn, M. & Monod, J. Purification et proprietes de la ß-galactosidase (lactase) d'Escherichia coli. *Biochimica et Biophysica Acta* **7**, 153-174 (1951).

34.   Rose, G. & White, A. Mechanism of catalysis by retaining β-glycosyl hydrolases. *Current Opinion in Structural Biology* **7**, 645-651 (1997).

35.   McIntosh, L.P. *et al.* The pKa of the general acid/base carboxyl group of a glycosidase cycles during catalysis: a 13C-NMR study of bacillus circulans xylanase. *Biochemistry* **35**, 9958-9966 (1996).

36.   Richard, J.P. The enhancement of enzymatic rate accelerations by Bronsted acid-base catalysis. *Biochemistry* **37**, 4305-4309 (1998).

37.   Huber, R.E., Wallenfels, K. & Kurz, G. The action of beta-galactosidase (Escherichia coli) on allolactose. *Can J Biochem* **53**, 1035-1038 (1975).

38.   Burstein, C., Cohn, M., Kepes, A. & Monod, J. Role du Lactose et de ses Produits Metaboliques dans l'Inuction de l'Operon Lactose ches *Escherichia coli. Biochimicha et Biophysica Acta* **95**, 634-639 (1965).

39.   Bourgeois, D. New processing tools for weak and/or spatially overalpped macromolecular diffraction patterns. *Acta Crystallographica* **D55**, 1733-1741 (1999).

CHAPTER IV

THE EVOLUTION OF BETA-GALACTOSIDASE

Introduction

As has been discussed briefly in Chapter 2, β-galactosidase has a modular

structure. This chapter describes comparisons between the different modules composing

β-galactosidase and other proteins to understand how β-galactosidase evolved. This

chapter contains co-authored material and was published in 1999[1]. My role was to do the

computational work involved. The manuscript was written together with Gene Huber

and my advisor, Brian Matthews.

Summary

Beta-galactosidase (lac Z) from *E. coli* is a 464 kDa homotetramer. Each subunit

consists of five domains, the third being an α/β barrel that contains most of the active site

residues. A comparison is made between each of the domains and a large set of proteins

representative of all structures from the protein data bank. Many structures include an

α/β barrel. Those that are most similar to the α/β barrel of *E. coli* β-galactosidase have

similar catalytic residues and belong to the so-called "4/7 superfamily" of glycosyl

hydrolases. The structure comparison suggests that β-amylase should also be included in this family. Of three structure comparison methods tested, the "ProSup" procedure of Zu-Kang and Sippl and the "Superimpose" procedure of Diederichs were slightly superior in discriminating the members of this superfamily, although all procedures were very powerful in identifying related protein structures.

Domains 1, 2 and 4 of *E. coli* β-galactosidase have topologies related to "jelly-roll barrels" and "immunoglobulin constant" domains. This fold also occurs in the cellulose binding domains (CBDs) of a number of glycosyl hydrolases. The fold of Domain 1 of *E. coli* β-galactosidase is closely related to some CBDs, and the domain contributes to substrate binding, but in a manner unrelated to cellulose binding by the CBDs. This is typical of Domains 1, 2, 4 and 5 which appear to have been recruited to play roles in β-galactosidase that are unrelated to the functions that such domains provide in other contexts. It is proposed that β-galactosidase arose from a prototypical single domain α/β barrel with an extended active site cleft. The subsequent incorporation of elements from other domains could then have reduced the size of the active site from a cleft to a pocket to better hydrolyze the disaccharide lactose and, at the same time, to facilitate the production of inducer, allolactose.

Background

β-galactosidase (lac Z) from *E. coli* is a 464 kDa homotetramer. Each subunit (1023 residues) is composed of five domains [2]. The overall structure is built around the third domain (Figure 40) which is a so-called "$(\alpha/\beta)_8$", "$(\beta/\alpha)_8$" or "TIM barrel" domain

(hereafter " α/β barrel"). This domain contains most of the active site residues. The

other domains consist primarily of β-sheet with Domain 1 having a jelly- roll barrel fold

and Domains 2 and 4 having immunoglobulin folds. Domain 5 is a 288-residue β-

sandwich with one structural homolog amongst known protein folds.
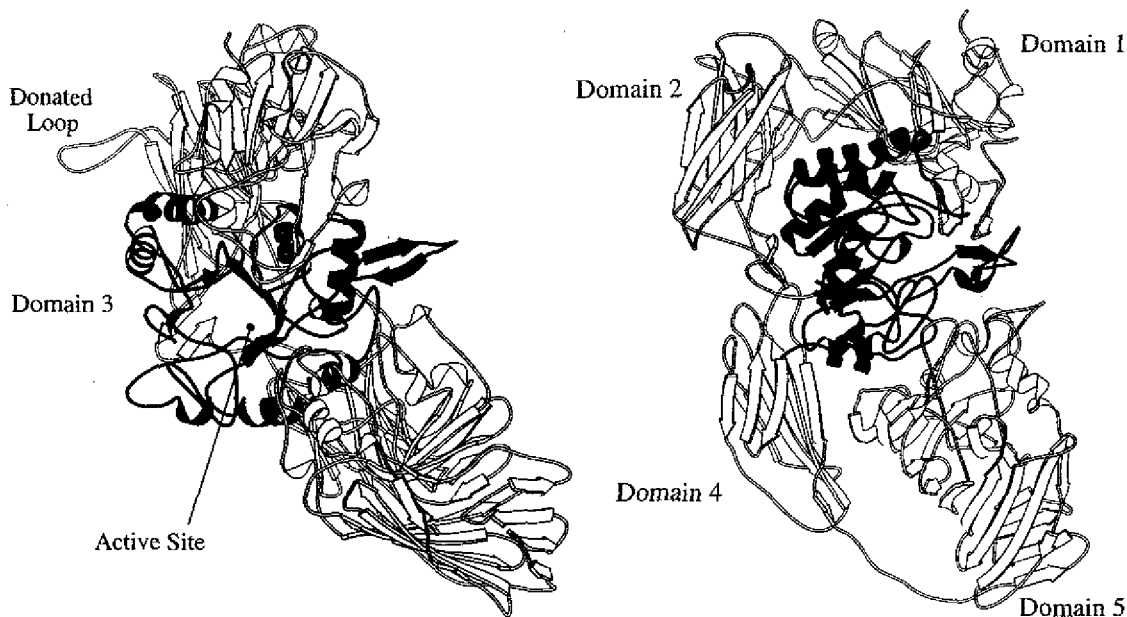
Figure 40. Orthogonal views of a single β-galactosidase monomer. (a) Looking straight
down the α/β barrel axis of Domain 3 (black). The cleft in the barrel runs up-down with
Domains 1 and 2 above, and 4 and 5 below-right. The "donated" loop extends across the
activating interface to complete the neighboring active site. (b) View from the left of
Figure 40(a). Except for the donated loop, Domains 2 and 4 are far from the active site,
which is on the right. The cleft in the barrel runs up-down on the right side with Domain
1 filling in the top and Domain 5 filling in the bottom.

Many organisms have enzymes with sequences related to that of *E. coli* β-

galactosidase. The closest relatives are summarized in Figure 41. Most of these enzymes

are β-galactosidases, but one branch contains mammalian and prokaryotic β-

glucuronidases. Some other enzymes, including β-galactosidase from *Bacillus circulans*

and some eukaryotic β-mannosidases, have weaker sequence correspondence (less than

25% identity) and are not included in the figure.
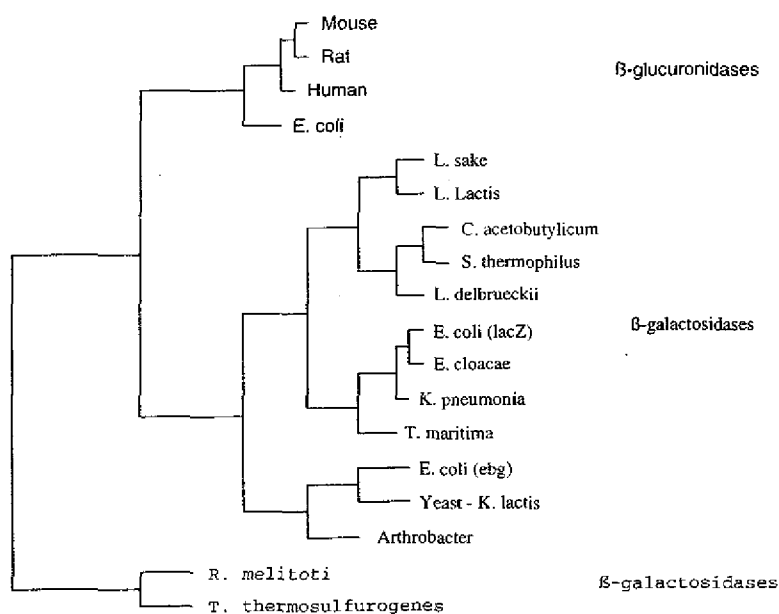
Figure 41. An evolutionary sequence tree for β-galactosidase and its closest relatives. The group of four enzymes at the top of the tree are β-glucuronidases; all others are β-galactosidases. The tree was created using the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) [3] and includes homologs with sequences at least 25% identical with Domain 3 (i.e. the α/β barrel) of *E. coli* β-galactosidase. The distances were calculated using the Jukes-Cantor method based on a multiple sequence alignment of regions of each protein corresponding to *E. coli* β-galactosidase Domain 3 [4]. All steps were done with the Wisconsin package [5]. Sequences used for the alignment came from a BLAST search with the amino acid sequence of *E. coli* β-galactosidase Domain 3.

Based on sequence similarities, the glycosyl hydrolases have been classified into

more than 50 families [6, 7]. All of the enzymes shown in Figure 40 are members of Family

2. Other considerations such as similarities in structure and catalytic mechanism have

been used to link various families into "superfamilies" or "clans"[8-10]. These

classifications have been based in large part on similarities of catalytic residues within

the active site domain. There is, however, more information potentially available from

consideration of the overall structure including the other four domains.

The objectives of the present study are: (1) to examine the domain modularity

and scaffolding of β-galactosidase (*E. coli*); (2) to use structural comparisons to

investigate the superfamily relationships; (3) to try to understand how changes in the

structure during evolution might explain differences in catalytic activity; and (4) to use

the abundance of structural data that are available for the glycohydrolases to test different

methods of structure comparison.

Results

Structure Comparisons

The overall backbone structure of each domain from *E. coli* β-galactosidase was

compared to a representative set of protein structures from the Protein Data Bank. Such a

list is published by the European Molecular Biology Laboratory under the name "PDB

Select"[11]. We were interested in detecting relatively distant relationships, not necessarily

apparent from sequence comparison and therefore chose a cutoff such that sequences

included were not more than 25% identical. This results in a sample, as of October 1997,

of 680 structures, including *E. coli* β-galactosidase. (Those proteins with sequences that

are more than 25% identical to *E. coli* β-galactosidase are identified in Figure 40.)

Comparisons of the five domains of β-galactosidase with the 680 representative

structures, carried out with ProSup [12](see Methods) are shown in Figures 42(a)-41(e).

These comparisons are characterized by the number of structurally equivalent residues

and the root-mean-square deviation (rmsd) of their Cα positions. It is convenient to first

discuss Domain 3, the α/β barrel.

Figure 42 (following page). Graphs for each domain of *E. coli* β-galactosidase showing the results of a structural search with ProSup for similar folds in a representative set of 680 known protein structures having less than 25% sequence identity with each other and with *E. coli* β-galactosidase. Program parameters were gap penalty s = 10, cutoff distance c = 3.5 , maximum seed p = 10, search atom = C_α. The plots show the percentage of Cα atoms in the β-galactosidase domain judged to be a structurally equivalent and the root-mean-square discrepancy between them. In each graph, triangles represent proteins identified in the Brookhaven Data Bank as having "jelly-rolls", squares represent proteins identified as immunoglobulins, circles indicate "α/β barrels" and the crosses indicate proteins that were not designated as belonging to any of the above categories. Structures that either score high or have similar function are identified including the Data Bank identification. (a) Domain 1 (residues 49-218). (b) Domain 2 (residues 219-334). (c) Domain 3 (residues 335-624). Solid symbols are α/β barrels with catalytic residues similar to β-galactosidase. Their Brookhaven ID codes and family names are identified. (d) Domain 4 (residues 625-725). (e) Domain 5 (residues 726-1023

Domain 1

Discrepancy of C' atoms (Å)

Residues structurally equivalent (%)

CBD IIa (1exg)
chitobiase (F 20, 1qba)
glucanase (F16, 2ayh)
CBD IIi (1nif)
polio virus coat (1pvc)
neuraminidase (F33, 1eut)
galactose oxidase
delta endotoxin (1dlc)
CBD Iv (1ulo)
PNGase (1pgs)
β gal domains 2&4

Domain 2

Discrepancy of C^α atoms (Å)

Residues structurally equivalent (%)

chitinase (F18, 1ctn)
β gal domain 4
neuraminidase (F33, 1eut)
EPO receptor (1ebp)
P53 (1tsr)
fibronectin III (1fnf)
growth hormone receptor (3hhr)
purple acid phosphatase (4kbp)
β gal domain 1
chitobiase (F20, 1qba)

Domain 3

Discrepancy of C^α atoms (Å)

Residues structurally equivalent (%)

1xyz
1ecb
1byb
2myr
1goi
1gow
1ece

▲	Family 1
◆	Family 5
▼	Family 10
■	Family 14
●	Family 17
○	Non 4/7 glycosidase α/β
○	Non glycosidase α/β
□	Immunoglobulin
△	Jelly roll
·	Other

Domain 4

Discrepancy of C^α atoms (Å)

Residues structurally equivalent (%)

chitobiase (F20, 1qba)
chitinase (F18, 1ctn)
rhamgalactan (1rmg)
cell adhesion protein (1neu)
β gal domain 2
hemagglutinin (1hgf)
neuraminidase (F33, 1eut)
β gal domain 1
P53 (1tsr)

Domain 5

Discrepancy of C^α atoms (Å)

Residues structurally equivalent (%)

copper amine oxidase (1oac)

Domain 3

The most distinctive graph (Figure 42(c)) is for Domain 3. The comparisons fall
into two quite distinct groups. On the right (shown as circles and solid symbols and
comprising 7% of the sample) are all the proteins that contain α/β barrels. The eight
structures to the extreme right of the figure (indicated by solid symbols) are all
glycosidases and have α-carbons that are structurally equivalent to more than 50% of the
E. coli β-galactosidase α/β barrel. Also the root-mean-square discrepancies are
generally somewhat less than those of the other α/β barrel proteins (open circles).
Together with β-galactosidase, these nine structures represent six glycosyl hydrolase
families which are functionally similar in that they hydrolyze 1,4 glycosidic bonds using
similar catalytic machinery. Jenkins et al. (1995) grouped five of these six families into
the "4/7 superfamily" because they have a catalytic acid/base on strand 4 of the α/β barrel
and a catalytic nucleophile on strand 7. (They were also identified by Henrissat et al.
(1995) and called "Clan GH-A".) Other glycosidases that have α/β barrels but are not in
the 4/7 superfamily have poorer structural similarity, although the similarity is generally
better than with the rest of the α/β barrels. Because we did not include proteins with
greater than 25% sequence identity, some members of the 4/7 superfamily do not appear
in Figure 42(c). These are, however, included in Figure 43 which is a phenogram
showing the result of comparing all α/β barrels with each other. Note that this
phenogram is based on structural correspondence (see Methods). In general it groups the

different proteins into the same subfamilies that have previously been proposed on the

basis of sequence relationships. The comparisons were repeated using different cutoff

parameters and the results were qualitatively similar (not shown).

Comparisons of the backbone of the α/β barrel of *E. coli* β-galactosidase with

other structures using the alignment algorithms DALI [13] and Superimpose[14] (see

Methods) are shown in Figures 43(a) and 43(b). Qualitatively, the results are similar to

those obtained using ProSup (Figure 42(c)). Both DALI and Superimpose differentiate

between α/β barrels and other structures although there are differences in the extent to

which the groups are separated.

We now return to the other domains of β-galactosidase.

Domain 1

Domain 1 is a so-called "jelly-roll barrel" [15]. The graph summarizing the

comparison of this domain with the 680 representative structures from the databank

(Figure 41(a)) shows that most of the structures with the greatest number of

Figure 43 (following page). Structure based tree showing the relationships between 93 α/β barrels based on their structural correspondence. The identification includes the data bank access code followed by the name of the protein. The figure was made by combining the structural agreement provided by ProSup with the Growtree algorithm of the GCG package (see Methods). All members of the 4/7superfamily (bold type), including β-amylase, appear on the same branch. The number in parentheses is the family number of Henrissat [6, 16]. Representatives of additional families identified by Henrissat are shown in Helvetica type.

Phenogram of 93 α/β Barrels

Based on Structure

- 1ald-aldolase
- 1fba-fbp-aldolase
- 1btm-tim
- 1hti-tim
- 1tph-tim
- 7tim-tim
- 5tim-tim
- 1tmh-tim
- 1tre-tim
- 1dhp-dhdps
- 1nal-n-acetylneuraminate-lyase
- 1dos-aldolase-class-ii
- 1igs-igps
- 1pii-igps
- 1nsj-prai
- 1pii-prai
- 1pkm-pyruvate-kinase
- 1pkn-pyruvate-kinase
- 1pky-pyruvate-kinase
- 1ttq-tryptophan-synthase
- 1pud-tgt
- 1sft-alanine-racemase
- 1bur-rubisco
- 1rbl-rubisco
- 3rub-rubisco
- 5rub-rubisco
- 1muc-cic
- 2chr-cmci
- 2mnr-mandelate-racemase
- 1pdz-enolase
- 4enl-enolase
- 1dor-dhodh
- 1gox-glycolate-oxidase
- 1ltd-flavocytochrome-b2
- 1oyc-old-yellow-enzyme
- 2tmd-trimethylamine-dehydrogenase
- **1bhg-beta-glucuronidase** *(2)*
- **d3pr-beta-galactosidase** *(2)*
- **1cbg-beta-glucosidase** *(1)*
- **2myr-myrosinase** *(1)*
- **1pbg-6-p-beta-gal** *(1)*
- **1gow-beta-glycosidase** *(1)*
- **1ceo-cellulase-celc** *(5)*
- **1edg-endoglucanase-a** *(5)*
- **1ece-endocellulase-el** *(5)*
- **1ghr-beta-glucanase** *(17)*
- **1ghs-beta-glucanase** *(17)*
- **1byb-beta-amylase** *(14)*
- **1cfx-xylanase-a** *(10)*
- **1xyz-xylanase** *(10)*
- **2exo-cellulase/xylanase** *(10)*
- 1amy-alpha-amylase *(13)*
- 2amg-alpha-amylase *(13)*
- 1bpl-alpha-amylase *(13)*
- 1pig-alpha-amylase *(13)*
- 1smd-amylase *(13)*
- 1cdg-cgtase *(13)*
- 1pam-cgtase *(13)*
- 1ciu-cgtase *(13)*
- 1cyg-cgtase *(13)*
- 2aaa-acid-alpha-amylase *(13)*
- 6taa-alpha *(13)*
- 1cnv-concanavalin-b *(18)*
- 2hvm-hevamine *(18)*
- 1ctn-chitinase-a *(18)*
- 1edt-endo-h *(16)*
- 2ebn-endoglycosidase-f1 *(18)*
- 1nar-narbonin
- 1qba-chitobiase *(20)*
- 1dxi-d-xylose-isomerase
- 2gyi-xylose-isomerase
- 2xis-xylose-isomerase
- 6xia-d-xylose
- 1xim-d-xylose-isomerase
- 4xia-d-xylose-isomerase
- 1tml-beta-glucanase *(6)*
- 1dik-ppdk
- 1ucw-transaldolase
- 1fkx-adenosine-deaminase
- 1psc-phosphotriesterase
- 1pta-phosphotriesterase
- 2kau-urease
- 1frb-fr-1-protein
- 2acq-aldose-reductase
- 2acs-aldose-reductase
- 2alr-aldehyde-reductase
- 1lwi-3-alpha-hydroxysteroid
- 1fvp-flavoprotein-390
- 1nfp-luxf
- 1luc-bacterial-luciferase
- 1gym-piplc
- 1qas-plc-delta-1
- 1req-methylmalonyl-coa-mutase

Other Glycosidases 4/7 Glycosidases

Figure 44. Comparison of two structural comparison programs. "Superimpose" and "DALI" were used to search the structural data base for folds similar to the α/β barrel from *E. coli* β-galactosidase. The set of structures searched with Superimpose was the same "PDB Select" list searched with ProSup (Figure 41(c)). The set searched with DALI is defined by the Dali server, and includes a set of unique structures similar to the "PDB Select" list. Solid symbols represent α/β barrels with similar catalytic machinery to *E.coli* β-galactosidase. These are identified by their Data Bank access code. Open circles represent other α/β barrels. Crosses indicate structures in the Data Bank not identified as α/β barrels. (a) Comparisons using "DALI". (b) Comparisons using "Superimpose".

structurally-equivalent residues have also been identified as "jelly-rolls". There is, however, overlap amongst the "jelly-rolls", the "immunoglobulins", and other proteins. "Jelly-rolls" and "immunoglobulins" have similar topology and can be viewed as variants of the same fold. The higher-scoring proteins in Figure 42(a) that are neither jelly-rolls nor immunoglobulins have the jelly-roll topology but have one or more insertions. Some of the jelly-rolls are found in glycosidases. These include CBDs [17-20], glucanase [21], neuraminidase [22] and chitobiase [23].

Domains 2 and 4

Domains 2 and 4 of β-galactosidase have identical topology, both having the fibronectin III fold. This is a close variant of the immunoglobulin constant domain [24]. Among the 680 proteins in the data base, those with most Cα atoms in common with Domain 2 of β-galactosidase are mostly identified as having "immunoglobulin" domains (Figure 42(b)). As with Domain 1, however, there is some overlap between the "immunoglobulin" and "jelly-roll" folds. Among all the matches seen in Figure 42(b), one of the best is between Domain 2 and Domain 4 of β-galactosidase. The inverse comparison is seen in Figure 42(d). Again, there are other glycosidases (e.g. a glycosyltransferase [25], chitinase [26], neuraminidase [22], chitobiase [23]) that have structures similar to Domains 2 and 4 of E. coli β-galactosidase.

Domain 5

The graph summarizing structural comparisons with Domain 5 of β-galactosidase is quite striking (Figure 42(e)). There is only one structure with significant similarity, namely the large β-sandwich domain of copper amine oxidase. This is also a multidomain oligomeric (dimeric) enzyme, and parts of the dimer interfaces of both β–galactosidase and copper amine oxidase are formed by roughly the same region of the -sandwich. The respective domains have the same overall fold and differ mainly in the length of the loops between the strands. With β-galactosidase, some residues from one end of the sandwich are contributed to form part of the active site. In the case of copper

amine oxidase, the active site residues are also on loops between the β-strands.

Topologically some of the active site residues of the two enzymes overlap, while others

do not.

Discussion

Domain Modularity and Scaffolding

One of the most striking aspects of the β-galactosidase structure is that it is

constructed from domains that, in most cases, serve very different functions in other

proteins. In addition, the parts of the domains used by β-galactosidase are different from

the parts that are functional in the other proteins. Of the five domains, only Domain 3

(the catalytic domain) is uniformly found in closely-related glycosyl hydrolases. The fact

that the other four domains are used in different roles in β-galactosidase suggests that

they have been appropriated as modules that provide stable scaffolds upon which binding

residues can be presented to render specificity to β-galactosidase.

Proteins similar to Domain 1 of β–galactosidase include viral coat proteins and

domains from toxins. Some of the other related structures (Figure 42(a)) are cellulose

binding domains (CBDs) of glycosyl hydrolases. Cellulose binding domains are usually

separated from their catalytic domains by flexible linkers. The CBDs occur in many

glycosyl hydrolases and are themselves classified into nine families (CBD:I-CBD:IX)

based on sequence similarities. The structures of representatives of CBD:I, CBD:II,

CBD:III and CBD:IV are known and β–galactosidase Domain 1 has appreciable

structural similarity to all but CBD:I. These relationships are illustrated in Figure 45.

As shown in Figure 45(a), Domain 1 of β–galactosidase has long loops that extend into

the active site region of the α/β barrel and help define the shape of the substrate binding

region. CBD:II (Figure 45(b)) has somewhat shorter loops. The loops are even shorter in

CBD:IV (Figure 45(d)). In the case of CBD:III (Figure 45(c))the loops make contact

with the α_6/β_6 catalytic domain [20]. Also the polypeptide chain extends to make a
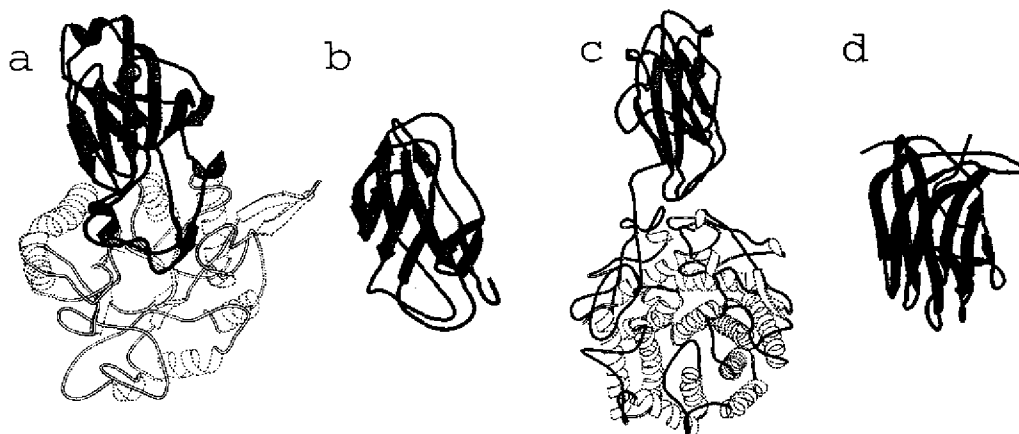
Figure 45. Comparison of Domain 1 of β–galactosidase with cellulose binding domains (CBDs) from the three structurally related families CBD:II, CBD:III and CBD:IV. (a) Domain 1 of β–galactosidase (shaded gray) with loops penetrating into the active site region within the α/β barrel (broken lines). (b) CBD:II domain from a bacterial xylanase (Xu et al., 1995) (1exh). (c)CBD:III domain (gray) with loops contacting the catalytic domain (broken lines) of a bacterial cellulase [20] (1tf4). (d) CBD:IV domain of a bacterial glucanase [18] (1ulp)

covalent connection between the two domains. The binding mechanisms that have been

proposed for the different cellulose binding domains [18, 19] are quite different from that

used by β–galactosidase and involve different parts of the respective structures. This is

also the case for the Family 16 bacterial endoglucanase which is one of the higher-

scoring Domain 1 analogs [18, 21]. There are other cases of this sort.

The structures with the highest similarity to Domain 2 of β–galactosidase are all fibronectin type III (fnIII) folds and immunoglobulin folds. They occur in a variety of proteins, often participating in cell surface interactions. The highest scorers include fnIII itself and a domain from the glycosyl hydrolase Family 33 bacterial neuraminidase [24] [22]. Structures with slightly poorer agreement include domains from glycosyl hydrolase Families 9 (a bacterial endoglucanase), 13 (a bacterial cyclodextrin glycosyltransferase), 18 (a bacterial chitinase) and 20 (a bacterial chitobiase) [25, 27] [23, 26].

As will be discussed, the structures most similar to Domain 3 of β–galactosidase represent the glycosyl hydrolase Families 1, 2, 5, 10, 14 and 17.

Domain 4 is like Domain 2. Hood et al. (1978) noted a similarity in sequence between residues 1-379 and 398-781 of β–galactosidase[28]. On this basis they proposed that these segments evolved via gene duplication and fusion. Their sequence alignment is, in part, remarkably close to the structural correspondence that is observed between Domains 2 and 4 (Figures 42(b), 42(d)). It therefore seems likely that these two domains were produced via gene duplication. However, the larger-scale duplication proposed by Hood et al. (1978) is not supported by any further structural correspondence.

The only closely matching structural homolog of Domain 5 is one of the domains of copper amine oxidase. Again, it is a protein with a function completely different from that of β–galactosidase.

These findings indicate that Domains 1, 2, 4 and 5 of β–galactosidase are, in essence, independent folding modules that serve to supplement or to modify the central role that is played by Domain 3. As shown in Figure 45(a), two loops of Domain 1 of β–galactosidase completely fill one end of the cleft in Domain 3. Inhibitor complexes

(Chapter 3) have shown that residues on these two loops contact the non-reducing end of the galactosyl moiety (residue Asp-201 in particular) and thus appear to be partially responsible for the specificity of β–galactosidase for β–D-galactosides. Domain 2 contributes residues to the active site of a neighboring subunit via a loop which reaches across a subunit interface. These residues appear to make no direct substrate contacts although some are located within 7Å of catalytic residues. Besides completing the α/β barrel, this loop also helps to convert the cleft to a pocket. The removal of this loop should substantially perturb the active site and it is therefore reasonable to assume that at least a dimer is required for activity (See also Figure 15, pg. 55 and discussion in Chapter 2 concerning α-complementation). Domain 2 also appears to function as a linker between Domains 1 and 3. Domain 4 is largely absent in some β–galactosidases and in those in which it is present the homology is poor. In the E. coli enzyme it has the least extensive interdomain contacts, contributes no residues to the active site and may indeed simply function as a linker between Domains 3 and 5. Domain 5 fills the end of the cleft opposite Domain 1 and also provides the important residue Trp-999, as well as the 794-804 loops. The role of Domain 5 is uncertain. Indeed, since the glucuronidases and some of the β–galactosidases require only Domains 1-3 to cleave disaccharide-sized substrates, it can be asked why the β–galactosidase of E. coli and some related organisms require two additional domains. The answer may lie in the ability of the enzyme not only to hydrolyze but also to transglycosylate. While β–glucuronidase can facilitate transglycosylation if the acceptor concentration is sufficiently high, this is apparently not part of its physiological role [29]. The same effect is found with β–galactosidase from T. thermosulfurogenes which has much of Domain 5 deleted [30]. The β–galactosidase from

E. coli, on the other hand, is tuned so that the kcat values for hydrolysis and

transglycosylation are about the same [31]. Domain 5 appears to contribute those residues

at the active site that bind and orient the glucose moiety of lactose. Thus its main role

may be to contribute to the dual activity of β-galactosidase, especially since Trp-999

stacks with glucose and the 794-804 loop is important for glucose binding.

Membership of the 4-7 Superfamily

Henrissat and Bairoch [6, 7] provided an overall classification of 57 glycosyl

hydrolase families. Based on topological similarity, Jenkins et al.(1995) identified five of

these as having a possible catalytic nucleophile on strand 7 and a possible acid/base on

strand 4. They identified these enzymes as the "4-7 Superfamily"[9]. Based on a structural

alignment of the α/β barrels (not shown) interfamily pairs show only 7-10% sequence

identity. There are only two residues (corresponding to Glu-461 and Glu-537 of

β-galactosidase) that are strictly conserved for all five families. Figure 46 shows the

structural coincidence of these active-site sidechains in the five family members plus

β-amylase. Also shown in the figure are an active site Trp/Phe that coincides, as well as

a cis peptide bond that occurs in four of the six families. Glu-537 or its equivalent has

been shown for many of these enzymes to be the active site nucleophile which attacks

the anomeric center [32-36]. Glu-461 or its equivalent appears to act as an acid/base in
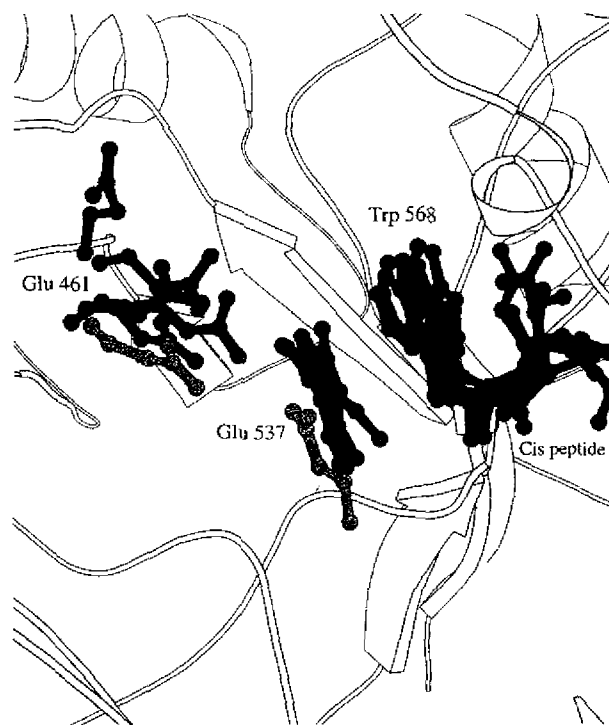
Figure 46. Close-up superposed view of selected sidechains in the active site of the β–galactosidase and the other 4/7 enzymes. The superpositions are based on the α-carbon correspondence determined by ProSup and one enzyme per family is shown (Family 1 - L. lactis 6-p-galactosidase (1pbg), Family 2 - *E. coli* β–galactosidase (1bgl), Family 5 - C. thermocellum β–glucanase (1cec), Family10 - C. thermocellum xylanase (1xyz), Family 14 - soybean β–amylase (1byb), Family 17 -barley β–glucanase (1ghr)). The β–galactosidase backbone is shown with sidechains for the presumed nucleophile (Glu-537), the acid/base (Glu-461), a tryptophan (Trp-568) and the conserved cis peptide bond. Dark gray side-chains correspond to the enzymes that cleave β–linked substrates with retention of configuration. The light gray sidechains correspond to β–amylase. The active site cleft runs roughly from lower left to upper right.

protonating the leaving group and activating a water to attack the enzyme-bound intermediate [37-39].

A phenogram based on the structures of 93 α/β barrels (Figure 43) suggests that Families 1, 2, 5, 10, 14 and 17 should be included in the 4-7 superfamily. Among this grouping is β–amylase (Family 14) which is included among the structures shown in Figure 46. β–Amylase is somewhat atypical in that it lacks the equivalent of Trp-568 and the neighboring cis peptide bond and also lacks a distinctive loop between strand 8

and helix 8. One of the major functional differences between β–amylase and the

other members of the 4/7 superfamily is that it is an "inverting" rather than a "retaining"

enzyme. For this reason the anomeric center is presumably attacked by a water molecule

rather than an acid residue [40] (Figure 47). The structural consequence (Figure 46) is a
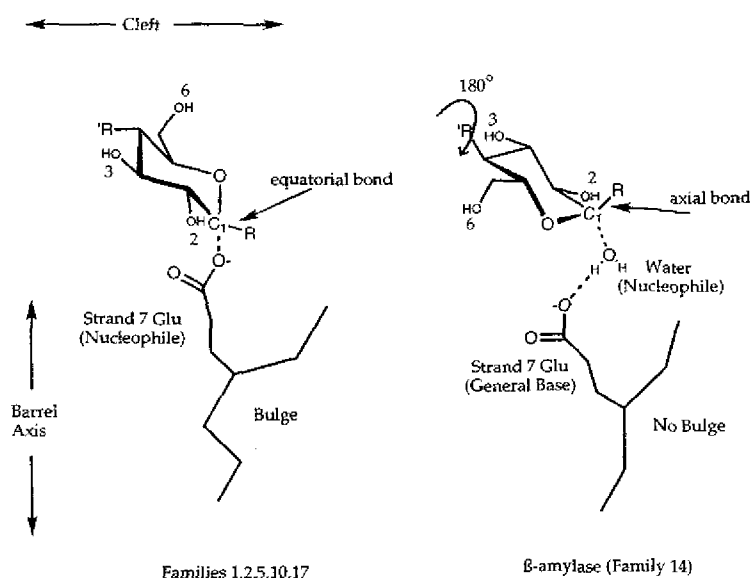
Figure 47. Schematic drawing comparing substrate binding in β–amylase (Family 14) with substrate binding in Families 1, 2, 5, 10 and 17 . There are two main differences. First, in Families 1, 2, 5, 10 and 17,which are retaining enzymes, the nucleophile is located on a bulge in strand 7 and directly attacks the C1 carbon of the substrate. β–Amylase lacks the bulge, so that the residue corresponding to the nucleophile is "retracted", leaving room for a water molecule to attack and produce an inverted substrate. Second, in Families 1, 2, 5, 10 and 17, the cleaved bond is equatorial whereas it is axial in β–amylase. However, the β–amylase substrate is rotated approximately 180° relative to that for Families 1, 2, 5, 10 and 17, giving the nucleophile similar access to the C1 carbon. This rotation puts the sugar substituents and the faces of the sugar in different locations relative to the cleft, resulting in different enzyme-substrate contacts.

slight withdrawal of the group that normally acts as a nucleophile into the barrel relative

to the "retaining" enzymes, allowing room for this water molecule. Also, although the

β–amylase substrate has an axial glycosidic bond, C1 is similarly presented to the

nucleophile because the substrate is bound with the opposite side of the sugar facing into

the cleft [40] (Figure 47). As a consequence, the sugar substituents all interact with different parts of the protein relative to the other 4/7 enzymes. For example, most of the 4/7 enzymes except β–amylase have an asparagine preceding the acid/base and with β–galactosidase this residue appears to bind the 2-hydroxyl of the cleaved sugar. With β–amylase, which lacks this asparagine, the 2-hydroxyl is on the opposite side of the cleft.

Similar arguments apply to additional active-site residues that are largely conserved among the other 4/7 enzymes. Despite these structural differences, β–amylase still retains the main characteristics of the 4/7 family. In particular the residue that corresponds to the nucleophile on strand 7 and the acid/base on strand 4 are the key elements of the 4/7 family. Taken together, the structural and functional relationships justify the inclusion of β–amylase in the 4/7 superfamily. The similarity of Family 14 to the above grouping was also discussed by Sakon et al. (1996) who considered possible relationships between Families 1, 2, 5, 6, 10, 13, 14, 17 and 18.

Henrissat and coworkers [8, 10] used amino acid sequence comparison and hydrophobic cluster analysis to include in the 4/7 superfamily additional subfamilies for which three-dimensional structures are as yet unknown. This technique was also used to study a subset of the 4/7 superfamily of clinical importance, the lysosomal hydrolases [41]. The overall characteristics of all of the putative 4/7superfamily members are given in Table 11. There are 12 families included, the first six of which have representatives with known three-dimensional structures. These six families are also the ones identified by our study as having α/β barrels like that of β–galactosidase. The smallest members of these six families are the Family 17 glucanases with 300-450 residues. Families 5 and

10 also have some single α/β barrel enzymes of 300-400 residues, but other members have additional domains. The enzymes of these three families appear to be involved in the degradation of either cell walls or cellulose and occur primarily in bacteria and fungi (Families 5 and 10) and plants (Family 17). Families 26 and 53 appear to have members that fit in with this grouping, having barrels of 300-400 residues, occurring in bacteria and fungi, and being involved in cell wall degradation.

The largest single-domain members are the Family 1 enzymes (~500 residues). This family carries out a variety of functions, although generally not cell wall degradation. Families 30, 35, and 39 include enzymes of similar overall length and all four families include mammalian lysosomal hydrolases. Within this group the extra length of the polypeptide chain tends to be correlated with a change in the active site from a cleft into a pocket.

Table 11. Some characteristics of putative members of the 4/7 superfamily of glycosyl hydrolases. This table summarizes the main characteristics of putative members of the 4/7 superfamily of glycosyl hydrolases. The family members are from Jenkins et al. (1995), Henrissat et al. (1995), Henrissat and Davies (1997) and the present work. The sources are from Bacteria (Eubacteria, B; Archaebacteria, B*), Fungi (F), Plants (P) and Animals (A). "Overall length" is the number of amino acids in the complete polypeptide chain. "Barrel length" is the number of residues in the a/b barrel, in cases where the structure is known. Such structures are identified by their Brookhaven Protein Data Bank access code. The substrate of each enzyme is identified as "small" (i.e. a disaccharide or equivalent), "long (i.e. oligosaccharide) or "small-large" (i.e. both). The orientation of the cleaved glycosidic bond (e = equatorial, a = axial) is also given. In cases where the three-dimensional structure is known the number of domains in the protein is indicated. Where the likely number of domains can be inferred from sequence homologies this is indicated inparentheses. Uncertain cases are indicated by question marks. References corresponding to each PDB coordinate file are as follows: 1cbg, [36]Barrett et al.(1995); 1gow, [42]; 1pbg, [43]; 2myr, [44]; 1bgl, [2]; 1bhg, [45];1cec, [46]; 1edg, [47]; 1ece, [48]; 1clx, [49]; 1xas, [50]; 1xyz, [46]; 2exo, [51]; 1byb, [40]; 1ghr and 1ghs, [52].

Family	Enzyme	Enzyme code	B	F	P	A	Overall length	Number of domains	Barrel length	Type of substrate	Active site	PDB code(s)
1	β–glucosidase	3.2.1.21	B	-	P	-	425-566	1	491	e-small	Pocket	1cbg
	β-galactosidase	3.2.1.23	B*	-	-	-	489-491	1	489	e-small-large	Cleft	1gow
	6-p-β-galactosidase	3.2.1.85	B	-	-	-	468-474	1	468	e-small	Pocket	1pbg
	thio-β-glucosidase	3.2.3.1	-	-	P	-	541-548	1	499	e-small	Pocket	2myr
	6-p-β-glucosidase	3.2.1.86	B	-	-	-	465-479	(1)	No structure	e-small		
	Lactase	3.2.1.108	-	-	-	A	1926-1927	(4)	No structure	e-small		
	Glycosylceramidase	3.2.1.62	-	-	-	A	1926-1927	(4)	No structure	e-long		
2	β-galactosidase	3.2.1.23	B	F	-	-	897-1037	4-5	277	e-small	Pocket	1bgl
	β-glucuronidase	3.2.1.31	B	-	-	A	603-651	3	290	e-long	Pocket	1bhg
5	Endo-1,4, β-glucanase	3.2.1.4	B	F	-	-	312-1039	1-?	332,380,362	e-long	Cleft	1cec,1edg1ece
	Exo-1,3-β-glucanase	3.2.1.58	-	F	-	-	421-570	?	No structure			
	Endo-1,4-β-mannosidase	3.2.1.78	B	-	-	-	363	?	No structure			
10	Endo-1,4-β–xylanase	3.2.1.8	B	F	-	-	312-1157	1-?	345,299,320	e-long	Cleft	1clx,1xas,1xyz
	Exo-1,4-β-cellobiosidase	3.2.1.91	B	-	-	-	484-1087	?	315	e-long	Cleft	2exo
	Endo-1,4-β-glucanase	3.2.1.4	B	F	-	-	387-1039	?	No structure			
14	β-amylase	3.2.1.2	B	-	P	-	488-575	1-?	491	a-long	Pocket	1byb
17	Endo-1,3-1,4-β–glucanase	3.2.1.73	-	-	P	-	312-370	1-?	306	e-long	Cleft	1ghr
	Endo-1,3-β–glucanase	3.2.1.39	-	-	P	-	305-478	1-?	306	e-long	Cleft	1ghs
	Exo-1,3-β–glucanase	3.2.1.58	-	F	-	-	308-313	1-?	No structure			

Table 11(continued)

Family	Enzyme	Enzyme code	Sources B	Sources F	Sources P	Sources A	Overall length	Number of domains	Barrel length	Type of substrate	Active site	PDB code(s
26	Endo-1,4-β–glucanase	3.2.1.4	B	-	-	-	900	?	No structure	e-long		
	Endo-1,4-β–mannosidase	3.2.1.78	B	F	-	-	360-398	?	No structure	e-long		
30	β–glucocerebrosidase	3.2.1.45	-	-	-	A	515-536	?	No structure	e-long		
35	β–galactosidase	3.2.1.23	B	F	P	A	536-1006	?	No structure	e-long		
39	Exo-1,4-β–xylosidase	3.2.1.37	B	-	-	-	488-500	?	No structure	e-long		
	α-L-iduronidase	3.2.1.76	-	-	-	A	634-655	?	No structure	e-long		
42	β–galactosidase	3.2.1.23	B	-	-	-	672	?	No structure	e-small		
53	Endo-1,4-β–galactanase	3.2.1.89	B	F	-	-	350-376	?	No structure	e-long		

Evolution of *E. coli* β–Galactosidase and Other Members of the 4/7 Superfamily

The coincidence of structural and functional similarity seen in Figure 42(c)

suggests possible evolutionary routes for the origin of β–galactosidase. Four of the

families represented by the solid symbols (Families 5, 10, 14 and 17) cleave a long-chain

polysaccharide substrate which binds in an active site cleft or tunnel in the α/β barrel. In

E. coli β–galactosidase the active site has a pocket-like shape. If, however, Domains 1,

2, 4 and 5 are stripped away, Domain 3 has an open, extended cleft (Figure 48(a)). Thus,

the extra domains increase the overall size of the enzyme and, at the same time, change

Figure 48. Overlay of 4/7 enzymes looking down the active site cleft. (a-left) The α/β barrels from the three families identified by ProSup that have a pronounced cleft (*C. thermocellum*, β-xylanase (1xyz), *C. thermocellum* β–glucanase (1cec), barley β–glucanase (1ghr)). Also superimposed is the structure of the α/β barrel of E. coli β–galactosidase (1bgl) (Domain 3) with Domains 1, 2, 4 and 5 removed. (b-right) Same view of representatives of the other two families (*L. lactis* 6-p-β–galactosidase (1pbg), and soybean β–amylase (1byb)). There is no cleft because of the longer loops.

the active site from a cleft to a pocket, which better accommodates the smaller substrate

(the disaccharide lactose). Let us assume that the prototypical enzyme had a structure

similar to that of the Family 17 barley endoglucanases (Figures 48(a), 49). The

enzymes in this family are, in a sense, the simplest members of the 4/7 superfamily

(Table 11). They have the shortest polypeptide chains, and function as monomeric

enzymes having only a single domain which is like Domain 3 of E .$coli$ β–galactosidase.

One imagines that in such cases the polysaccharide substrate would bind in an open cleft

running along the barrel with a number of the saccharide units making specific contacts

with the protein and providing binding energy to be used for catalysis. The activity of

such a prototype could be modified by altering the shape of the binding cleft. This might

happen in different ways. One possibility would be to extend the loops at the end of the

α/β barrel, partly filling the active site cleft and making it more appropriate for binding

smaller substrates. This is illustrated in Figure 48(b). Such an enzyme might, at least to

some extent, cleave a saccharide from the non-reducing end of an oligosaccharide, or

hydrolyze a smaller substrates such as a disaccharide. Another possibility would be to

add complete domains (CBDs) on one side or the other of the active site cleft. The sorts

of ways in which this might occur are illustrated in Figures 45(a) and 45(c). Yet another

possibility would be to utilize a combination of loop extension and the addition of new

domains. Of the alternative paths from the oligosaccharide to disaccharide hydrolase, one

would imagine that lengthening loops would be more economical than adding domains.

This is clearly the way that the Family 1 disaccharide hydrolases converted a cleft into a

pocket (Figure 48(b)). The evolutionary selection, however, is presumably for function in

preference to efficiency and shows that the extant state may be very path dependent.

Other enzymes appear to have added domains in order to hydrolyze smaller substrates.

These include the Family 20 chitobiases, which include structural homologs to
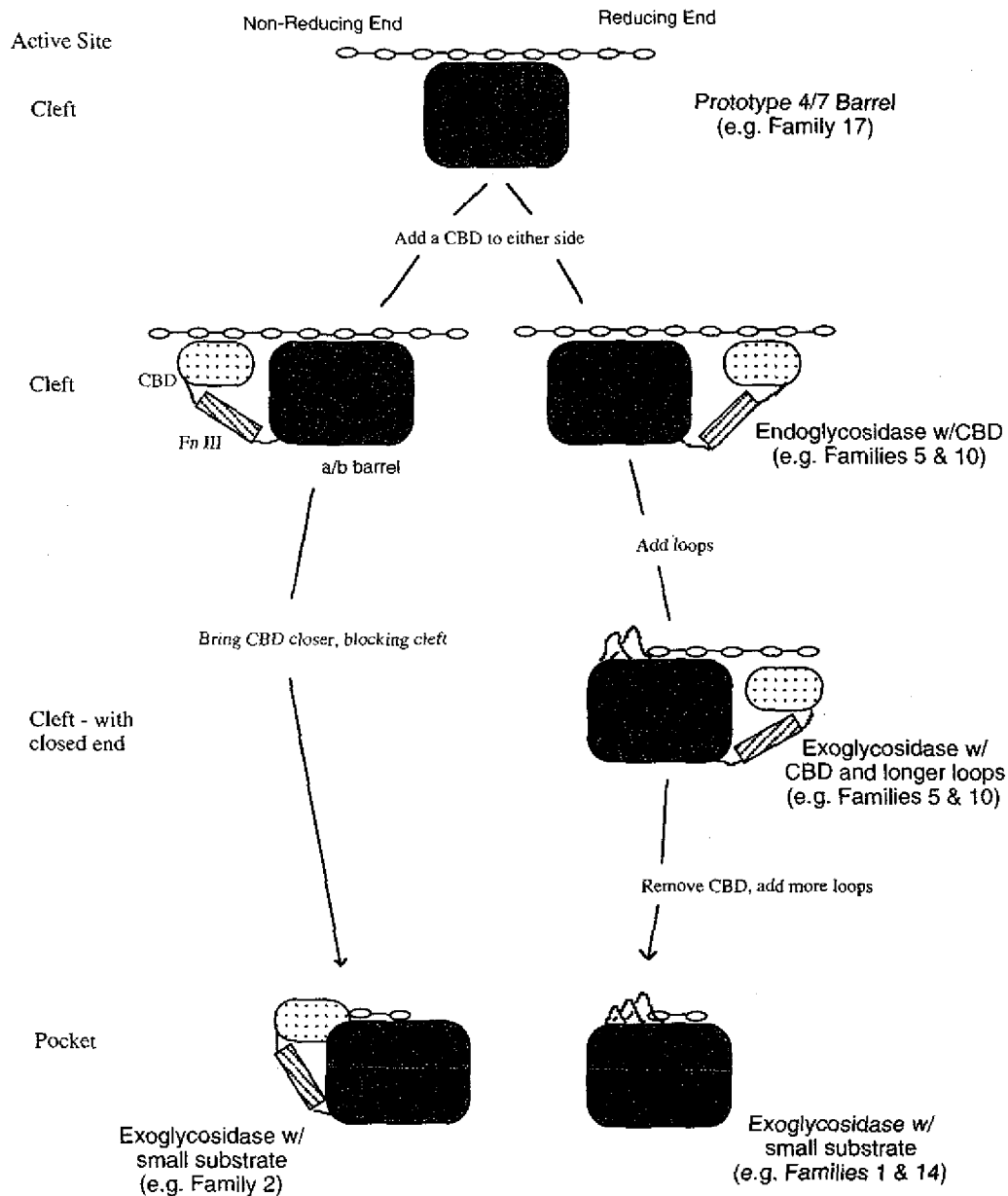
Figure 49. Model for the evolution of β–galactosidase and other 4/7 enzymes. A presumed prototypical α/β barrel enzyme (shaded) with a long, groove-like active site cleft and an extended polysaccharide substrate is shown at the top. The addition of a cellulose binding domain (CBD)to either terminus of the α/β barrel could give rise to both endoglycosidases and exoglycosidases (e.g. Families 5 and 10). The generation of an enzyme that hydrolyses small substrates might have occurred by either of the routes shown utilizing loops and/or a CBD to turn the active site from a cleft into a pocket.

Domains 1 and 2 as well as an additional domain that decorates a central α/β barrel

catalytic domain. Another example is the Family 33 bacterial neuraminidase. It has a

structure that is strikingly similar to domains 1-3 of β–galactosidase. Here domains very

similar to Domains 1 and 2 are built around a barrel structure which in this case is a six-

fold β-propeller rather than an α/β barrel [22]. This apparently is also how Family 2 has

evolved but in the case of β–galactosidase from E. coli and some closely related

enzymes, two extra domains have been added.

The β–glucuronidases (Figure 43) are about 600 residues in length and are in the

same family as β–galactosidase. Their structures appear to correspond to the first three

domains of the β–galactosidases with which they have about 25% overall sequence

identity. This structural relationship was confirmed recently with the determination of

the three-dimensional structure of human β–glucuronidase [45]. As discussed previously,

β–galactosidase from E. coli and some related β–galactosidases have two extra domains.

Domain 4 was presumably added by gene duplication of Domain 2 and seems to serve as

a "linker" to Domain 5 which was probably added and adapted specifically to bind

glucose in such a way that the enzyme has the capability to produce allolactose. Based

on this scenario, Domains 1-3 of β–galactosidase might be capable of forming a stable

folded structure. This may be the basis for the phenomenon of complementation, in

which a deletion of the C-terminal third of the lac Z gene (the acceptor) can be

complemented by a deletion of the N-terminal two-thirds of the lac Z gene (the donor)[53,

54]. Studies with antibodies suggested that each of these elements could fold

independently [55]. The β–galactosidase structure suggests that the acceptor comprises

Domains 1-3 while the donor consists of Domain 5 or Domains 4 and 5. Also consistent

with the above explanation, the subunits for *L. sake* and *L. lactis* β–galactosidase

(Figure 41) actually comprise two separate polypeptides roughly corresponding,

respectively, to Domains 1-3 and Domain 5 of the E. coli enzyme. Most of the

equivalents of Domain 4 are missing, suggesting that this region may not be critical for

function. These two enzymes together with the three to which they are most closely

related all have a deletion in the "donated loop" that reaches across the activating

interface to complete the active site (Figure 15, pg. 55). This suggests that they do not

need to be tetrameric. Indeed it has been shown that *L. delbrueckii* β–galactosidase is

active as a dimer (Adams et al., 1994). The enzymes from *R. melitoti* and *T.*

thermosulfurogenes (Figure 41) are smaller than the *E. coli*-type β–galactosidases (about

700 residues) and have lower sequence identity (about 20%). These two enzymes have

large deletions of Domain 4 and at least one of them (*T. thermosulfurogenes*) is a dimer

[30]. They also have long deletions within Domain 5 and do not produce allolactose except

in the presence of high concentrations of glucose. The deletions of regions of Domain 5

may also relate to their dimeric structures[30].

Structure Comparisons

In many respects the results from the three comparison programs ProSup, DALI

and Superimpose are comparable. All identify structural domains in other proteins

related to those in E. coli β–galactosidase. For this reason we have compared the results

of all three methods only for Domain 3 (Figures 42(c), 44(a), 44(b)). Each technique

consistently groups Families 1, 5 and 17 separately from all other structures in the similarity plots and also the α/β barrels separately from the non-α/β barrels.

There are also some differences. ProSup and Superimpose both differentiate the five TIM barrels that are glycosidases and appear to have the same catalytic machinery. Members of this group are indicated by the solid symbols in Figure 42(c). In the ProSup comparison, these structures have lower discrepancy between C_ atoms and/or more structurally equivalent residues than all the other α/β barrels. For DALI there is some overlap between the different barrel structures (Figures 44(a), 44(b)). There is not a consensus concerning the relative similarity of the various α/β barrels to β–galactosidase. For instance, ProSup and Superimpose group β–amylase with the other members of the 4/7 family. For DALI this distinction is much less clear-cut (Figure 44(a)).

In Figure 42(c), the non-α/β barrels with the most equivalent atoms correspond to typically 3/8 of the barrel. That is, the span of the structural segments being compared includes about three strands plus three helices. In contrast, the most poorly scoring α/β barrels (left-most open circles) correspond to more than half of a complete barrel. The presence of the gap suggests that the α/β barrel is an "all-or-nothing" structure. In other words, one does not have a gradual succession of structures with 3, 4, 5,...-strands eventually leading to the complete 8-stranded barrel.

The α/β barrels have pseudo eight-fold symmetry yet the apparent nucleophile is always found on strand 7. Given the symmetry of the α/β barrel, if these enzymes evolved convergently there would be no obvious reason to favor the location of

the nucleophile on one strand over another. Thus these different structures appear to have arisen by divergent evolution.

Following the preceding discussion there appear to be at least two factors that may have contributed to the large size of β–galactosidase. First, it may have evolved from a much smaller enzyme which cleaved long polysaccharides. In the process of modifying the active site to have shape complementarity towards the smaller substrate lactose, additional domains may have been added. Second, the addition of these domains may also have been associated with the development of a second activity (transglycosylation).

Methods

Structure Alignments

Three methods were used to search for structures similar to the domains of *E. coli* β–galactosidase. These are described below.

Superposition of Spatially Equivalent Backbone Segments

The method "ProSup" of Zu-Kang and Sippl (1996) is based on the superposition of backbone segments that are spatially superimposable[12]. Initial sets of equivalent segments, called seeds, are found and the alignment given by each seed is then optimized. To generate the seeds, every possible pair of fragments of length 5 between the two

structures is superimposed. Each fragment is then extended by adding residues onto either end if the Cα-Cα distance between these residues is less than a specified cutoff distance. This cycle is repeated until convergence (i.e. the equivalent fragments do not change). The result is a set of initial seeds. The alignments given by the seeds with the most equivalent Cα pairs are then refined using a dynamic programming algorithm. The resulting refined alignments are compared and redundant alignments are discarded. Although ProSup provides multiple alignments for each parities comparison, we only used the one with the most equivalent Cα pairs.

In the case of the α/β barrels, the structural correspondence as measured by ProSup was used to generate a structure-based phenogram. For every pair of α/β barrels the following ratio was determined: Rms Cα discrepancy/percentage of residues structurally equivalent in the smaller barrel. This was then input to the UPGMA algorithm of the GCG package [5] to obtain the phenogram shown in Figure 43 (page 154).

Superposition Using Distance Matrices

The program "DALI", introduced by Holm and Sander (1993), uses Cα-Cα distance matrices to find the optimal superposition of two structures. In principle the method allows for comparison of structures that have similar elements of secondary structure, but connected in a different sequence, although this is not normally done. The strategy is to break up the distance plot into overlapping 6-mer squares. Pairs of similar subdistance plots define equivalenced residues. Using a Monte Carlo algorithm, a chain of equivalenced segments is created to give the overall alignment. The similarity is

measured by a similarity score which is related to the distances between equivalent

pairs of residues.

Generalized Superposition

The method of Diederichs (1995) ("Superimpose") was also tested as a general

way to search for similarities in C_α positions. It can be applied with or without regard to

topological equivalency. The algorithm uses a six-dimensional search strategy to find the

alignment between two structures that maximizes a similarity function. The molecule is

rotated through discrete angular increments. For each angle a translation search is

performed to maximize the number of equivalent C_α atoms. This involves finding the

most common intermolecular C_α–C_α vector. The algorithm is general, allows for

different topologies, and the user can employ a number of different "filters". The "filter"

used in this study is the "Topological Similarity Score" which favors alignments with

many equivalent atoms and high sequence correlation between pairs.

Sequence Searches and Alignments

Amino acid sequence searches were done with BLAST at the National Center for

Biotechnology Information [56]. Non-redundant tblastn and blastp data bases were

searched. Amino acid sequence alignments were performed with the Wisconsin program

package [5]. Default parameters were used (swgappcp.cmpmatrix, gap creation = 3.0, gap

extension = 0.1) with pileup (multiple sequence alignment) and Bestfit (pairwise sequence alignment).

Structure Databases and Fold Assignments

The comparisons using ProSup and Superimpose were made using the PDB Select 10/97 release with a "25% identity cutoff" [11]. The Family II CBD from C. fimi (PDB code 1EXH) was also included because it is the only CBD II structure known. Although it has greater than 25% sequence identity with three structures in the PDB Select list, it scores much higher in structural comparisons with Domain 1 of β-galactosidase. The bacterial sialidase (code 1EUT) was also included because of its similar domain organization to β-galactosidase. Comparisons with DALI were done by submitting coordinates to the DALI server.

Structures were assigned to one or more of the categories α/β barrel, jelly-roll, immunoglobulin and "other" based on the identifications provided with the coordinates as well as a visual inspection of the structures. In cases where the assignment was uncertain a check was made with the SCOP data base [57]. While most of the structures with high similarity to any of the domains fall clearly in one of the categories, those with poorer similarity are more difficult to classify. This is especially true of the immunoglobulins and the jelly-rolls, which can be viewed as variants of each other. These classifications are therefore meant to serve as overall identifications rather than precise descriptors.

References

1.	Juers, D.H., Huber, R.E. & Matthews, B.W. Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between ß-galactosidase and other glycohydrolases. *Protein Science* **8**, 122-136 (1999).

2.	Jacobson, R.H., Zhang, X.-J., DuBose, R.F. & Matthews, B.W. Three-dimensional structure of ß-galactosidase from *E. coli*. *Nature* **369**, 761-766 (1994).

3.	Sneath & Sokol. *Numerical Taxonomy*, (Freeman, San Fransisco, 1973).

4.	Jukes, T. & Cantor, C. Evolution of protein molecules. in *Mammalian protein metabolism* (ed. Munro, H.) 21-132 (Academic Press, New York, 1969).

5.	Genetics_Computer_Group. *Program Manual for the Wisconsin Package, Version 8, 575*, , Madison, 1994).

6.	Henrissat, B. & Bairoch, A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **293**, 781-788 (1993).

7.	Henrissat, B. & Bairoch, A. Updating the sequence-based classification of glycosyl hydrolases [letter]. *Biochem J* **316**, 695-696 (1996).

8.	Henrissat, B. *et al.* Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc Natl Acad Sci U S A* **92**, 7090-7094 (1995).

9.	Jenkins, J., Lo-Leggio, L., Harris, G. & Pickersgill, R. Beta-glucosidase, beta-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold beta/alpha architecture and with two conserved glutamates near the carboxy-terminal ends of beta-strands four and seven. *Febs Lett* **362**, 281-285 (1995).

10.	Henrissat, B. & Davies, G. Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* **7**, 637-644 (1997).

11.	Hobohm, U. & Sander, C. Enlarged representative set of protein structures. *Protein Sci* **3**, 522-524 (1994).

12.   Zu-Kang, F. & Sippl, M.J. Optimum superimposition of protein structures: ambiguities and implications. *Folding & Design* **1**, 123-132 (1996).

13.   Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233**, 123-138 (1993).

14.   Diederichs, K. Structural superposition of proteins with unknown alignment and detection of topological similarity using a six-dimensional search algorithm. *Proteins* **23**, 187-195 (1995).

15.   Branden, C. & Tooze, J. *Introduction to Protein Structure*, (Garland, New York, 1991).

16.   Bairoch, A. http://www.expasy.ch/cgi-bin/lists?glycosid.txt Website describing various characteristics of the glycosyl hydrolase families. , 1997).

17.   Xu, G.Y. *et al.* Solution structure of a cellulose-binding domain from Cellulomonas fimi by nuclear magnetic resonance spectroscopy. *Biochemistry* **34**, 6993-7009 (1995).

18.   Johnson, P.E., Joshi, M.D., Tomme, P., Kilburn, D.G. & McIntosh, L.P. Structure of the N-terminal cellulose-binding domain of Cellulomonas fimi CenC determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **35**, 14381-14394 (1996).

19.   Tormo, J. *et al.* Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *Embo J* **15**, 5739-5751 (1996).

20.   Sakon, J., Irwin, D., Wilson, D.B. & Karplus, P.A. Structure and mechanism of endo/exocellulase E4 from Thermomonospora fusca. *Nat Struct Biol* **4**, 810-818 (1997).

21.   Hahn, M., Keitel, T. & Heinemann, U. Crystal and molecular structure at 0.16-nm resolution of the hybrid Bacillus endo-1,3-1,4-beta-D-glucan 4-glucanohydrolase H(A16-M). *Eur J Biochem* **232**, 849-858 (1995).

22.   Gaskell, A., Crennell, S. & Taylor, G. The three domains of a bacterial sialidase: a beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure* **3**, 1197-1205 (1995).

23.   Tews, I. *et al.* Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. *Nat Struct Biol* **3**, 638-648 (1996).

24.   Leahy, D.J., Hendrickson, W.A., Aukhil, I. & Erickson, H.P. Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science* **258**, 987-991 (1992).

25.   Lawson, C.L. *et al*. Nucleotide sequence and X-ray structure of cyclodextrin glycosyltransferase from Bacillus circulans strain 251 in a maltose-dependent crystal form. *J Mol Biol* **236**, 590-600 (1994).

26.   Perrakis, A. *et al*. Crystal structure of a bacterial chitinase at 2.3 A resolution. *Structure* **2**, 1169-1180 (1994).

27.   Juy, M. *et al*. Three dimensional structure of a thermostable bacterial cellulase. *Nature* **357**, 89-91 (1992).

28.   Hood, J.M., Fowler, A.V. & Zabin, I. On the evolution of ß-galactosidase. *PNAS* **75**, 113-116 (1978).

29.   Levvy, G. & Conchie, J. Glucuronidase and the hydrolysis of glucuronides. in *Glucuronic Acid, Free and Combined* (ed. Dutton, G.) 301-364 (Academic Press, New York, 1966).

30.   Huber, R.E., Roth, N.J. & Bahl, H. Quaternary structure, Mg2+ interactions, and some kinetic properties of the beta-galactosidase from Thermoanaerobacterium thermosulfurigenes EM1. *J Protein Chem* **15**, 621-629 (1996).

31.   Huber, R.E., Gupta, M.N. & Khare, S.K. The active site and mechanism of the beta-galactosidase from Escherichia coli. *Int J Biochem* **26**, 309-318 (1994).

32.   Gebler, J.C., Aebersold, R. & Withers, S.G. Glu-537, not Glu-461, is the nucleophile in the active site of (lac Z) beta-galactosidase from Escherichia coli. *J Biol Chem* **267**, 11126-11130 (1992).

33.   Chen, L., Fincher, G.B. & Hoj, P.B. Evolution of polysaccharide hydrolase substrate specificity. Catalytic amino acids are conserved in barley 1,3-1,4- and 1,3-beta-glucanases. *J Biol Chem* **268**, 13318-13326 (1993).

34.   McCarter, J.D. & Withers, S.G. Mechanisms of enzymatic glycoside hydrolysis. *Curr Opin Struct Biol* **4**, 885-892 (1994).

35.   Wang, Q. *et al*. Glu280 is the nucleophile in the active site of Clostridium thermocellum CelC, a family A endo-beta-1,4-glucanase. *J Biol Chem* **268**, 14096-14102 (1993).

36.   Barrett, T., Suresh, C.G., Tolley, S.P., Dodson, E.J. & Hughes, M.A. The crystal structure of a cyanogenic beta-glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure* **3**, 951-960 (1995).

37. Navas, J. & B'eguin, P. Site-directed mutagenesis of conserved residues of Clostridium thermocellum endoglucanase CelC. *Biochem Biophys Res Commun* **189**, 807-812 (1992).

38. Keresztessy, Z., Kiss, L. & Hughes, M.A. Investigation of the active site of the cyanogenic beta-D-glucosidase (linamarase) from Manihot esculenta Crantz (cassava). II. Identification of Glu-198 as an active site carboxylate group with acid catalytic function. *Arch Biochem Biophys* **315**, 323-330 (1994).

39. MacLeod, A.M., Lindhorst, T., Withers, S.G. & Warren, R.A. The acid/base catalyst in the exoglucanase/xylanase from Cellulomonas fimi is glutamic acid 127: evidence from detailed kinetic studies of mutants. *Biochemistry* **33**, 6371-6376 (1994).

40. Mikami, B., Degano, M., Hehre, E.J. & Sacchettini, J.C. Crystal structures of soybean beta-amylase reacted with beta-maltose and maltal: active site components and their apparent roles in catalysis. *Biochemistry* **33**, 7779-7787 (1994).

41. Durand, P. *et al.* Active-site motifs of lysosomal acid hydrolases: invariant features of clan GH-A glycosyl hydrolases deduced from hydrophobic cluster analysis. *Glycobiology* **7**, 277-284 (1997).

42. Aguilar, C.F. *et al.* Crystal structure of the beta-glycosidase from the hyperthermophilic archeon Sulfolobus solfataricus: resilience as a key factor in thermostability. *J Mol Biol* **271**, 789-802 (1997).

43. Wiesmann, C., Hengstenberg, W. & Schulz, G.E. Crystal structures and mechanism of 6-phospho-beta-galactosidase from Lactococcus lactis. *J Mol Biol* **269**, 851-860 (1997).

44. Burmeister, W.P. *et al.* The crystal structures of Sinapis alba myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase. *Structure* **5**, 663-675 (1997).

45. Jain, S. *et al.* Structure of human beta-glucuronidase reveals candidate lysosomal targeting and active-site motifs. *Nat Struct Biol* **3**, 375-381 (1996).

46. Dominguez, R. *et al.* A common protein fold and similar active site in two distinct families of beta-glycanases. *Nat Struct Biol* **2**, 569-576 (1995).

47. Ducros, V. *et al.* Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. *Structure* **3**, 939-949 (1995).

48. Sakon, J., Adney, W.S., Himmel, M.E., Thomas, S.R. & Karplus, P.A. Crystal structure of thermostable family 5 endocellulase E1 from Acidothermus cellulolyticus in complex with cellotetraose. *Biochemistry* **35**, 10648-10660 (1996).

49.   Harris, G.W. *et al.* Structure of the catalytic core of the family F xylanase from Pseudomonas fluorescens and identification of the xylopentaose-binding sites. *Structure* **2**, 1107-1116 (1994).

50.   Derewenda, U. *et al.* Crystal structure, at 2.6-A resolution, of the Streptomyces lividans xylanase A, a member of the F family of beta-1,4-D-glycanases. *J Biol Chem* **269**, 20811-20814 (1994).

51.   White, A., Withers, S.G., Gilkes, N.R. & Rose, D.R. Crystal structure of the catalytic domain of the beta-1,4-glycanase cex from Cellulomonas fimi. *Biochemistry* **33**, 12546-12552 (1994).

52.   Varghese, J.N. *et al.* Three-dimensional structures of two plant beta-glucan endohydrolases with distinct substrate specificities. *Proc Natl Acad Sci U S A* **91**, 2785-2789 (1994).

53.   Ullmann, A., Perrin, D., Jacob, F. & Monod, J. Identification par Complementation in vitro et Purification d'un Segment Peptidique de la ß-Galactosidase d'*Escherichia coli. JMB* **12**, 918-923 (1965).

54.   Goldberg, M.E. Tertiary Structure of *Escherichia coli* ß-D-galactosidase. *JMB* **46**, 441-446 (1969).

55.   Celada, F., Ullmann, A. & Monod, J. An Immunological Study of Complementary Fragments of ß-Galactosidase. *Biochemistry* **13**, 5543-5547 (1974).

56.   Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410 (1990).

57.   Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247**, 536-540 (1995).

CHAPTER V

THE EFFECTS OF FREEZING ON PROTEIN CRYSTALS

Introduction

Many protein crystals are subject to significant damage during X-ray data

collection. Figure 50 shows the decay of a β-galactosidase crystal at room temperature

and one flash frozen to ~100 K. Damage occurs hundreds of times faster at room

temperature (Table 12). At a synchrotron, where the beam intensity is up to 1000x

greater, damage occurs about 2000x faster (Table 12) and data collection at room

temperatures is practically impossible, because of the short life of the crystals. In the

case of β-galactosidase, a single data set at 1.7 Å would likely require 50 to 100 crystals,

whereas at low temperature a complete data set to 1.5 Å resolution can be collected on a

single crystal. The exact mechanism of the decay is unknown, but is thought to involve

free radicals generated by the interaction of the incident x-ray radiation with the solvent

or protein.

The search for suitable conditions under which to freeze crystals is often

problematic and screening is usually necessary to find a suitable cryoprotectant and a

freezing protocol. The cryoprotectant is often necessary in order to prevent ice crystals

from forming, which can destroy the protein lattice. Once conditions are found it is

usually possible to collect data to higher resolution than possible without freezing.
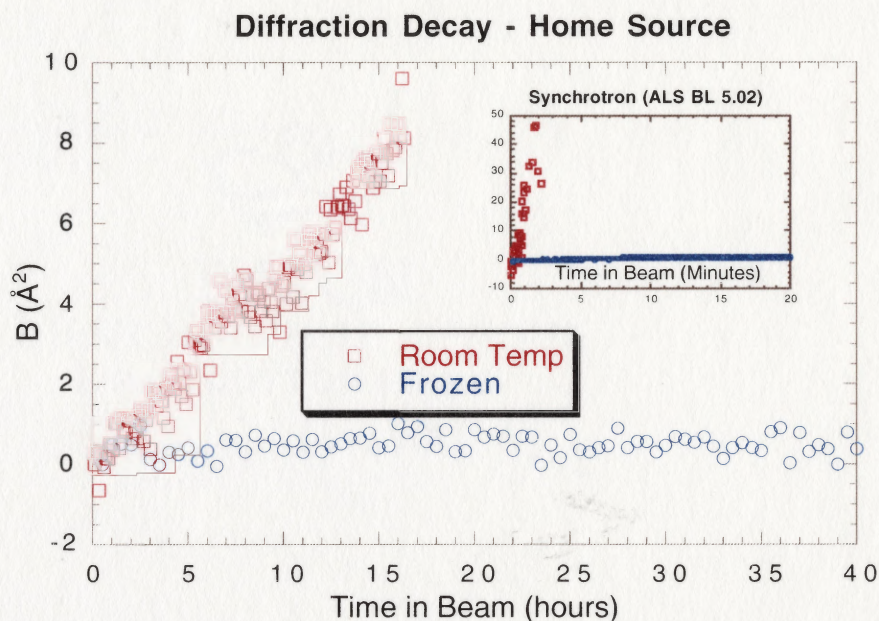
Diffraction Decay - Home Source

Figure 50. Plots of B vs Time in Beam. Each spot represents one exposure during data collection. B is derived from a scaling algorithm used to scale each exposure to the first one. It describes the falloff of diffraction intensity with increasing scattering angle: where theta is the scattering angle and lambda is the wavelength. Over time, the

$$I \sim e^{-B(\sin\theta/\lambda)^2}$$

radiation destroys the lattice, causing the diffraction intensity to falloff faster with theta, increasing B. The crystals used were: Room-Temp/Home(Raxis4): E537Q/X-gal; Low-Temp/Home(Raxis2):Native/Galactal; Room-Temp/Synchrotron: E537Q; Low-Temp/Synchrotron: E537Q/Galactal

Table 12. Crystal decay rates under various conditions measured with the rate of change of B, based on linear fits of the data in Figure 50.

	Room Temp	Low Temp	RT/LT
Synchrotron (ALS BL 5.0.2)	1380 Å^2/hour	3.6 Å^2/hour	380
Home source	0.50 Å^2/hour	0.003 Å^2/hour	200
Synchrotron/Home source	3000	1200	

Although freezing of crystals is an important element of macromolecular
crystallography it has not been studied in much detail. There have been a few reports of
the surprising possibility of recovering "lost" diffraction by using "annealing" protocols,
and this technique is gaining in popularity[1-3]. However, there has been little discussion
of more basic issues about what happens to crystals when they freeze. The intent of this
chapter is to describe some observations and experiments designed to address some basic
issues concerning freezing β-galactosidase crystals.

Results

1. Freezing Decreases the Unit Cell Volume

Figure 51 shows a histogram of unit cell volumes for different crystals of $P2_12_12_1$
β-galactosidase. On average, there is a 7 % decrease upon freezing. Both sets of crystals
(frozen and non-frozen) include a variety of protein ligand complexes and three different
mutations, as well as data sets processed with two different data processing programs.
Both groups also include data sets collected in house and at synchrotrons, although most
of the frozen data sets are from synchrotrons, while all but one of the room temperature
data sets were measured house. Additionally, the room temperature data sets are of lower
resolution than the frozen data sets. These differences may have contributed to the
spread in the two distributions, but it is highly unlikely that there is a systematic error
grouping the crystals in a way to lead to the 7 % decrease in cell volume upon freezing.
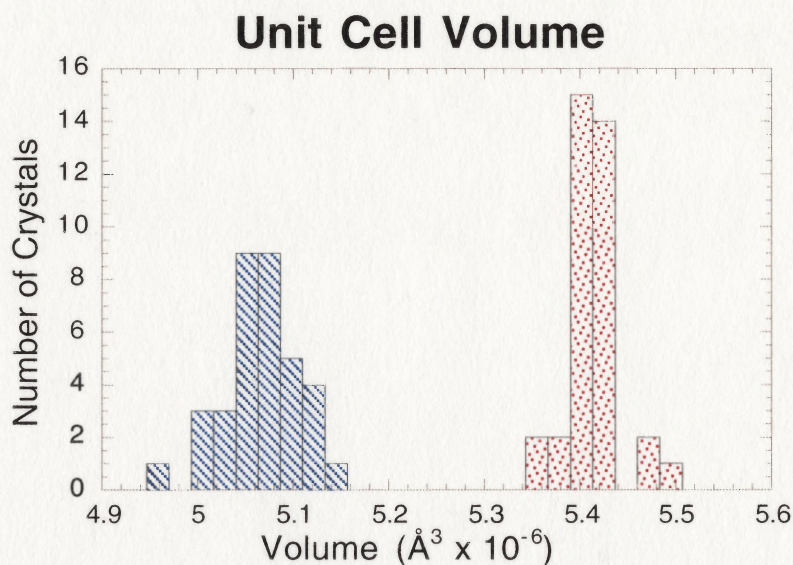
Figure 51. Histogram of orthorhombic β-galactosidase unit cell volumes. Blue cross hatching shows crystals at low temperature (~100 K) and red dots shows crystals at room temperature (~295 K). There is a ~7% decrease in volume with freezing.

2. Freezing Repacks the Crystal Lattice

As calculated with molecular surface programs, the area buried at crystal contacts increases by about 80 % upon freezing[4, 5]. The area buried at intersubunit contacts within a tetramer also increases, but not as much. This suggests that there is repacking of the crystal lattice upon freezing and slight repacking of the monomers within the tetramer. Examination of the coordinates shows that there are shifts of the tetramers in the crystal lattice relative to each other upon freezing. This is shown schematically in Figure 52.
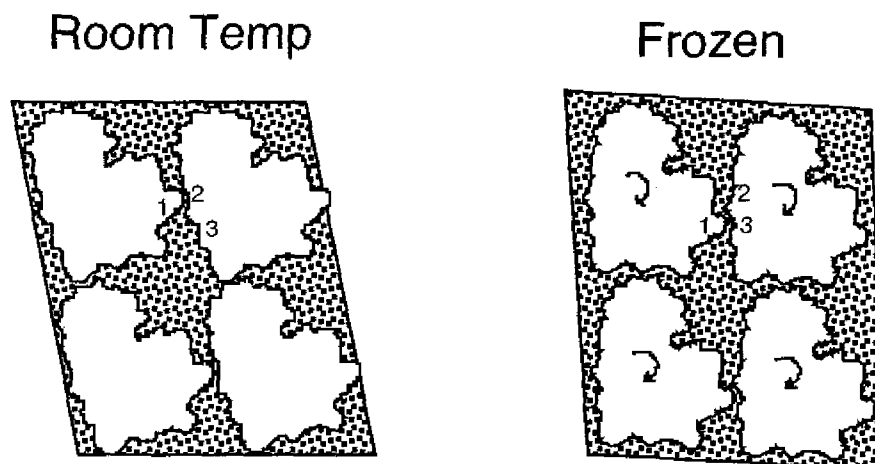
Room Temp

Frozen

Figure 52. Schematic illustrating changes in crystal packing upon freezing. The cell parameters change, and the molecules reorient. The 1-2 crystal contact becomes the 1-3 crystal contact. In $P2_12_12_1$ β-galactosidase, the reorientation includes rotations up to 3 degrees and shifts at the crystal contacts of several angstroms.

3. Freezing Does Not Change the Protein Volume

As calculated with molecular surface programs, the volume of the protein actually increases by ~ 1% upon freezing[4, 5]. This is probably near the noise level for the calculation. Since the volume of the protein doesn't change appreciably, the change in unit cell volume observed on freezing cannot be due to thermal contraction of the protein itself.

4. The Effects of Freezing are Reversible

A natural question, then, is whether the observed changes in unit cell volume are reversible. To test this, a crystal was mounted at room temperature in the cryo buffer using the standard method in a glass capillary and several diffraction images were taken

to determine the unit cell volume. The buffer was 10% PEG8000, 100 mM bis-tris

pH 6.5, 200 mM $MgCl_2$, 100 mM NaCl, 10 mM DTT, and 30% DMSO as the

cryoprotectant. The crystal was then expelled from the capillary, flash frozen, and several

more images were taken. The crystal was then "flash melted" by bringing a large drop of

crystal buffer up to the crystal and immersing the crystal directly in the drop. After

remounting in a glass capillary, more images were taken at room temperature. This was

repeated as many times as possible. Up to three cycles of freeze-melt could be attained

before the crystal broke apart. Figure 53(a) shows that the change in cell volume is

reversible. The volume decreases upon freezing and increases upon melting.

Furthermore, the mosaicity also changes reversibly, increasing with freezing and

decreasing with melting.

A second experiment was done to explore the reversibility in smaller temperature

jumps. Figure 53(b) shows that the cell change in the low temperature range is

reversible.

For the reversibility experiments, temperatures above ~170 K could not be

measured, because when the crystals were warmed to this temperature they stopped

diffracting. The disappearance of the protein diffraction was accompanied by the

appearance of a powder diffraction pattern which corresponded to the expected powder

diffraction pattern for hexagonal ice (space group $P6_3/mmc$, a=4.49 Å, c=8.33 Å). This

behavior could be reproduced with buffer only (no crystal), and also with water and

DMSO only. It appears then that flash freezing the buffer produces a glassy material.

Warming this material through about 170 K allows the water to crystallize

(devitrification). This phenomenon of devitrification has also been observed in

DMSO/water mixtures using differential thermal analysis[6]. The formation of these

ice crystals destroys the protein crystal lattice.
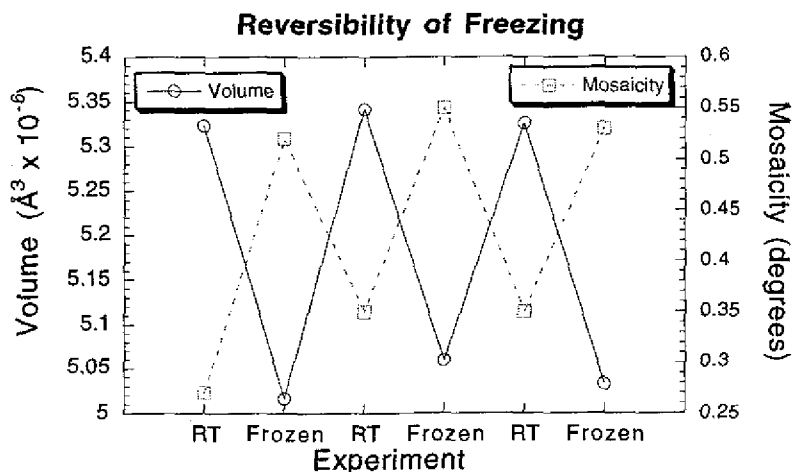
Figure 53. Reversible nature of the effects of freezing. (a) The crystal can be repeatedly frozen and melted cycling the volume and mosaicity between room temperature and low temperature values. There are, however, some indications of hysteresis in the mosaicity.
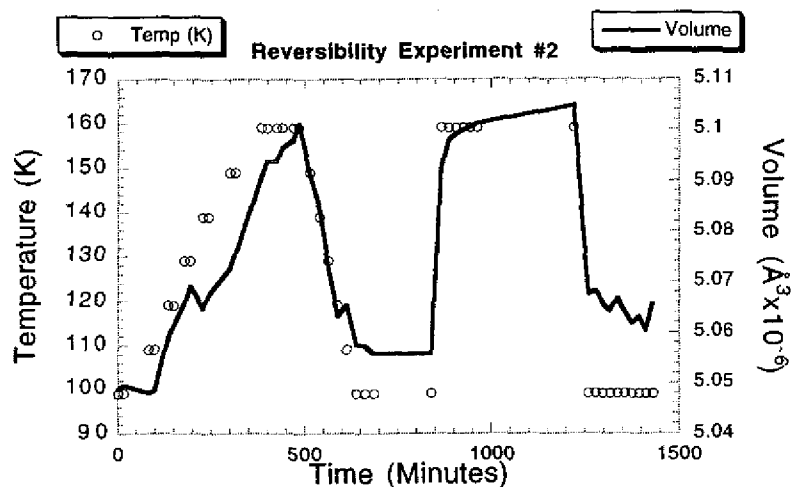
(b). Reversibility tests at low temperature. The crystal was first flash frozen to 100 K, then warmed to 160 K in 10 degree steps, cooled back to 100 K in 10 degree steps, warmed to 160 K in a 60 degree jump, and cooled back to 100 K in a 60 degree jump. Changes in cell volume can be detected even with 10 degree temperature jumps.

Discussion

Although the low temperature crystal packing is well ordered, diffracting to high resolution, it is apparently never produced with the initial crystallizations at 15° C. The reason for this could be a path dependence on the crystallization process. That is, the low temperature packing might be a stable configuration at room temperature, but only accessible through freezing (and then thawing). However, the reversible nature of the cell parameter change argues against this idea (Figure 53). It suggests, instead, that each temperature has a different optimal packing configuration which is governed by thermodynamics at each temperature.

The protein volume (and therefore density) does not change appreciably on freezing. Changes in density of bulk solvent could have an effect on the unit cell volume. However, the density of the vitrified bulk solvent after flash freezing has not been measured. The density of vitrified water has been measured with cryo electron microscopy and found to be ~0.93 g/cm[37], suggesting that water slightly expands upon vitrification. Assuming this bears on the behavior of the bulk solvent, it is then apparent that inherent changes in the density of neither the protein nor the bulk solvent can account for the decrease in unit cell volume, and in fact they may be expected to increase the cell volume.

The change in unit cell volume is probably be due, then, to the temperature dependence of the protein-protein interactions and protein-solvent interactions which define the crystal lattice. For example, it will be easier to order long amino acid side

chains and form a salt bridge at lower temperature. Hydrophobic interactions at 100

K are probably also quite different from those at room temperature.[8]

Some flexibility in the proteins and the lattice itself will be required to allow the

molecules to explore different packing arrangements during the freezing process. If the

lattice as formed at room temperature is very strong and rigid, then repacking will be less

likely upon freezing. If the lattice is very weak then it will be easier for the molecules to

reorient. In either case, the repacking is probably a cooperative process.

References

1.   Yeh, J. & Hol, W. A flash-annealing technique to improve diffraction limits and lower mosaicity in crystals of glycerol kinase. *Acta Crystallographica* **D54**, 479-480 (1998).

2.   Harp, J., Timm, D. & Bunick, G. Macromolecular Crystal Annealing: Overcoming Increased Mosaicity Associated with Cryocrystallography. *Acta Crystallographica* **D54**, 622-628 (1998).

3.   Harp, J., Hanson, B., Timm, D. & Bunick, G. Macromolecular crystal annealing: evaluation of techniques and variables. *Acta Crystallographica* **D55**, 1329-1334 (1999).

4.   Connolly, M. The molecular surface package. *Journal of Molecular Graphics* **11**, 139-141 (1993).

5.   Zhang, X.-J. & Matthews, B. EDPDB: A multi-functional tool for protein structure analysis. *Journal of Applied Crystallography* **28**, 624-630 (1995).

6.   Rasmussen, D. & MacKenzie, A. Phase Diagram for the System Water-Dimethylsulfoxide. *Nature* **220**, 1315-1317 (1968).

7.   Dubochet, J. *et al.* Cryo-electron miscroscopy of vitrified specimens. *Quarterly Review of Biophysics* **21**, 129-228 (1988).

8.   Privalov, P. & Gill, S. The hydrophobic effect: a reappraisal. *Pure and Applied Chemistry* **61**, 1097-1104 (1989).

CHAPTER VI

SUMMARY AND CONCLUSIONS

The structure of β-galactosidase was originally determined in a monoclinic crystal

form to 2.5 Å resolution.. This dissertation describes the crystallization and structure

determination of β-galactosidase in an orthorhombic crystal form to 1.7 Å resolution.

Analysis of this structure resulted in new insights concerning Mg^{++} and Na^+ binding, α-

complementation and the hierarchical construction of β-galactosidase(Chapter 2).

The mechanism of β-galactosidase includes two alternate pathways – one for

hydrolysis and one to produce allolactose, the inducer for the lac operon. This

mechanism was studied by determining the binding modes of several ligands to β-

galactosidase and β-galactosidase variants to high resolution (up to 1.5 Å). These

structures, which include two trapped covalent intermediates and products, as well as

structures designed to mimic early points in the reaction and transition states, suggest a

putative reaction pathway for β-galactosidase. The pathway involves the progression of

the substrate to a deep pocket where interactions between the galactosyl hydroxyls and

the enzyme position the substrate for a concerted nucleophilic attack of the galactosyl

group and proton donation to the leaving group. Detailed consideration of this reaction

pathway in the context of previous biochemical studies offers new insights and questions

about the mechanism for β-galactosidase (Chapter 3).

A conformational change in the enzyme is observed which appears to be triggered by the progression of the substrate toward the intermediate. This conformational change could also be triggered via amino acid substitution and this β-galactosidase variant appears to favor a structure better suited for binding the transition state for hydrolysis. However, the variant is deficient for allolactose production, suggesting the residues involved in the conformational change are important for determining which reaction pathway the enzyme follows. Possible roles for this conformational change in the action of native enzyme are discussed (Chapter 3).

The evolution of β-galactosidase was investigated by considering structures related to its 5 domains. A model is proposed involving a progenitor catalytic domain which could hydrolyze a long polysaccharide substrate by binding it in a cleft. The addition of other domains transformed the active site from a cleft into a pocket, drastically increasing the size of the enzyme, but allowing it to hydrolyze a much smaller substrate(Chapter 4).

Finally, a brief investigation of the effect of freezing on the orthorhombic β-galactosidase crystals is described. It is shown that the unit cell volume decreases by 7% upon freezing. This effect is reversible, suggesting that each temperature has a slightly different optimal packing determined by the nature of the intermolecular interactions at that temperature (Chapter 5).

APPENDIX

DERIVATION OF EXPRESSIONS FOR KINETICS

Typically, an enzyme catalyzed reaction follows saturation kinetics:

$$v_o = \frac{E_o \cdot k_{cat} \cdot [S]}{K_m + [S]} \quad (1)$$

where v_o is the initial rate of the reaction, E_o is the enzyme concentration and S is the substrate concentration. Experimentally, v_o is measured for different S at fixed E_o. This results in data that can be fit to the above equation using the two parameters: k_{cat}, which is the turnover number of the enzyme (reactions/second) and K_m, which is the substrate concentration at 1/2 the maximum rate.

A minimal model for an enzyme catalyzed reaction with two steps, as with β-galactosidase is:

$$E + S \underset{k_{-1}}{\overset{k_1}{\Longleftrightarrow}} E \cdot S \overset{k_2}{\rightarrow} E - I + P1 \overset{k_3}{\rightarrow} E + P2$$
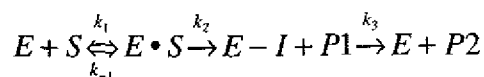
where E = enzyme, S = substrate, E•S = enzyme-substrate complex, E-I = enzyme-intermediate, P1 = product 1, P2 = product 2, k_1 = rate constant for association of E + S, k_{-1} = rate constant for dissociation of E•S, k_2 = rate constant for step 1, k_3 = rate constant for step 2.

The initial velocity for any model of an enzyme reaction is given by:

$$v_o = E_o \cdot k_{net} \quad (2)$$

where k_{net} is the net rate constant for the model. In this case, k_{net} is given by:

$$\frac{1}{k_{net}} = \frac{1}{k_1'} + \frac{1}{k_2} + \frac{1}{k_3} \quad (3)$$

where:

$$k_1' = S k_1 \frac{k_2}{k_2 + k_{-1}} \quad (4)$$

is a net rate constant for the association of enzyme and substrate. It is the rate of association times the probability that the association will continue down the reaction pathway.

Combining the equations 2-4, and rearranging to look lik equation 1 equation,

gives:
$$v_o = \frac{S\dfrac{k_2 k_3}{k_2 + k_3}}{\dfrac{(k_{-1} + k_2)k_3}{k_1(k_2 + k_3)} + S} \quad (5)$$
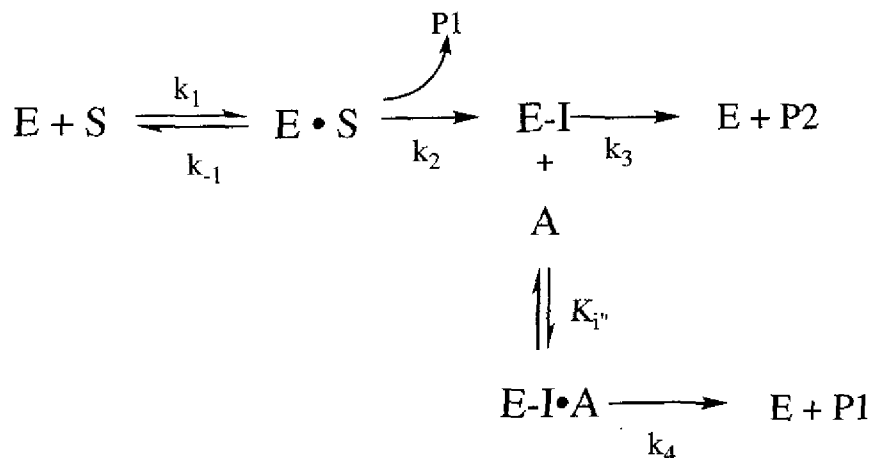
and therefore:
$$k_{cat} = \frac{k_2 k_3}{k_2 + k_3} \quad \text{and} \quad K_m = \frac{(k_{-1} + k_2)k_3}{k_1(k_2 + k_3)} \quad (6).$$

If substrate binding is a rapid equilibrium ($k_{-1} \gg k_2$), then:

$$K_m = K_s \frac{k_3}{k_2 + k_3} \quad (7)$$

where $K_s = k_{-1}/k_1$ is a dissociation constant for the enzyme-substrate complex. Thus if the rate of step 1 is comparable to the rate of step 2, an appreciable fraction of the enzyme is in the enzyme-intermediate complex, $K_m < K_s$. This is the case with onpg for β-galactosidase.

A model for β-galactosidase with an intermolecular acceptor, A, in rapid equilibrium with the intermediate is:

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E \bullet S \xrightarrow[k_2]{\quad P1 \quad} E\text{-}I \xrightarrow{k_3} E + P2$$

$$+$$

$$A$$

$$\Big\updownarrow K_{i''}$$

$$E\text{-}I \bullet A \xrightarrow{k_4} E + P1$$

With a net rate constant for given by:

$$\frac{1}{k_{net}} = \frac{1}{k_1} + \frac{1}{k_2} + \frac{1}{k_3'} \quad (8)$$

$$k_3' = k_3 \cdot \frac{1}{1 + \alpha} + k_4 \cdot \frac{\alpha}{1 + \alpha} \quad (9)$$
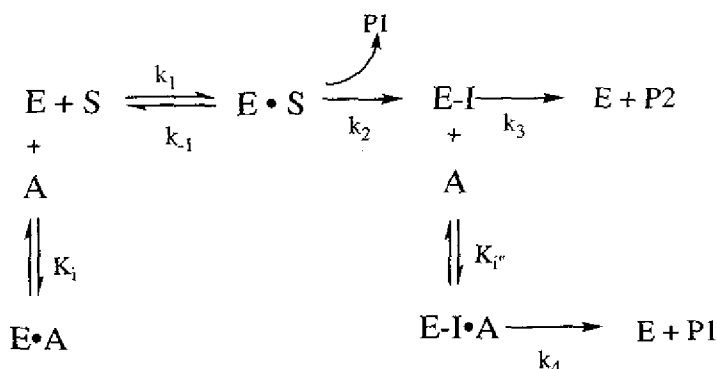
where $\alpha = \dfrac{A}{K_i''}$. Here, k_3 and k_4 are multiplied by the fraction of the intermediate which is following the k_3 or k_4 path.

Following the algebra through yields:

$$k_{cat} = \frac{k_2(k_3 + k_4\alpha)}{k_2 + k_3 + (k_2 + k_3)\alpha} \quad (10)$$

$$K_m = K_s \frac{(k_3 + k_4\alpha)}{k_2 + k_3 + (k_2 + k_3)\alpha} \quad (11)$$

Adding in the possibility that the acceptor A can also act as a competitive inhibitor:

decreases the amount of free enzyme to $E/(1+A/K_i)$, but doesn't directly affect the other complexes. Therefore, the net rate constant k_1' will decrease to $k_1'/(1+A/K_i)$. Following this through yields:

$$K_m = K_s(1 + \frac{A}{K_i})\frac{(k_3 + k_4\alpha)}{k_2 + k_3 + (k_2 + k_3)\alpha} \quad (12)$$

kcat is unaffected because a competitive inhibitor can always be overcome by increasing the substrate concentration. These expressions for k_{cat} and K_m (eqns 10-12) can be directly fit to data in A.

Alternatively, Deschavanne et al (1978) describe the use of these equations for kinetic analysis of various effectors. Determining k_{cat} and K_m for a series of concentrations of A allows one to determine K_i, K_i'' and k_4 by using:

$$\frac{K_m}{k_{cat}} = \frac{K_s}{k_2}(1 + \frac{I}{K_i}) \quad (13)$$

$$k_{cat} = \frac{(k_{cat} - k_{cat}^0)}{I} \frac{k_2 + k_3}{k_2 + k_4} K_i'' + \frac{k_2 k_4}{k_2 + k_4} \quad (14)$$

BIBLIOGRAPHY

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403-10 (1990).

Andrews, K., & Lin, E. Thiogalactoside Transacetylase of the Lactose Operon as an Enzyme for Detoxification. *Journal of Bacteriology* **128**, 510-512 (1976).

Bairoch, A. http://www.expasy.ch/cgi-bin/lists?glycosid.txt
Website describing various characteristics of the glycosyl hydrolase families. , 1997).

Barrett, T., Suresh, C. G., Tolley, S. P., Dodson, E. J., & Hughes, M. A. The crystal structure of a cyanogenic beta-glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure* **3**, 951-60 (1995).

Becker, V. E., & Evans, H. J. The influence of monovalent cations and hydrostatic pressure on beta-galactosidase activity. *Biochimica et Biophysica Acta* **191**, 95-104 (1969).

Beckwith, J. R., & Zipser, D. The lactose operon. (Cold Spring Harbor, New York, 1970).

Berg, O. G., Winter, R. B., & von Hippel, P. H. Diffusion-Driven Mechanisms of Protein Translocation on Nucleic Acids 1. Models and Theory. *Biochemistry* **20**, 6929-6948 (1981).

Bourgeois, D. New processing tools for weak and/or spatially overalpped macromolecular diffraction patterns. *Acta Crystallographica* **D55**, 1733-1741 (1999).

Branden, C., & Tooze, J. *Introduction to Protein Structure*, (Garland, New York, 1991).

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

Burstein, C., Cohn, M., Kepes, A., & Monod, J. Role du Lactose et de ses Produits Metaboliques dans l'Inuction de l'Operon Lactose ches *Escherichia coli*. *Biochimicha et Biophysica Acta* **95**, 634-639 (1965).

Cannon, W. R., Singleton, S. F., & Benkovic, S. J. A perspective on biological catalysis. *Nat Struct Biol* **3**, 821-33 (1996).

Celada, F., Ullmann, A., & Monod, J. An Immunological Study of Complementary Fragments of β–Galactosidase. *Biochemistry* **13**, 5543-5547 (1974).

Chen, L., Fincher, G. B., & Hoj, P. B. Evolution of polysaccharide hydrolase substrate specificity. Catalytic amino acids are conserved in barley 1,3-1,4- and 1,3-beta-glucanases. *J Biol Chem* **268**, 13318-26 (1993).

Cohn, M., & Monod, J. Purification et proprietes de la β–galactosidase (lactase) d'Escherichia coli. *Biochimica et Biophysica Acta* **7**, 153-174 (1951).

Connolly, M. The molecular surface package. *Journal of Molecular Graphics* **11**, 139-141 (1993).

Cupples, C. G., Miller, J. H., & Huber, R. E. Determination of the roles of Glu-461 in beta-galactosidase (Escherichia coli) using site-specific mutagenesis. *J Biol Chem* **265**, 5512-8 (1990).

Degraeve, P., Delorme, P., & Lemay, P. Pressure-induced inactivation of *E. coli* beta-galactosidase: influence of pH and temperature. *Biochim Biophys Acta* **1292**, 61-8 (1996).

Degraeve, P., & Lemay, P. High pressure-induced modulation of the activity and stability of Escherichia coli (lacZ) β-galactosidase: Potential applications. *Enyzme and Microbial Technology* **20**, 550-557 (1997).

Deschavanne, P. J., Viratelle, O. M., & Yon, J. M. Conformational adaptability of the active site of beta-galactosidase. Interaction of the enzyme with some substrate analogous effectors. *J Biol Chem* **253**, 833-7 (1978).

Diederichs, K. Structural superposition of proteins with unknown alignment and detection of topological similarity using a six-dimensional search algorithm. *Proteins* **23**, 187-95 (1995).

Dienert, F. Sur La Fermentation Du Galactose. *Annales de l'Institute Pasteur* **14**, 139-189 (1900).

Dubochet, J., Adrian, M., Chang, J., Homo, J.-C., Lepault, J., McDowell, A., & Schultz, P. Cryo-electron miscroscopy of vitrified specimens. *Quarterly Review of Biophysics* **21**, 129-228 (1988).

Durand, P., Lehn, P., Callebaut, I., Fabrega, S., Henrissat, B., & Mornon, J. P. Active-site motifs of lysosomal acid hydrolases: invariant features of clan GH-A glycosyl hydrolases deduced from hydrophobic cluster analysis. *Glycobiology* **7**, 277-84 (1997).

Edwards, L. A., Tian, M. R., Huber, R. E., & Fowler, A. V. The use of limited proteolysis to probe interdomain and active site regions of beta-galactosidase (Escherichia coli). *J Biol Chem* **263**, 1848-54 (1988).

Evans, P. (1993) in *Proceedings of a CCP4 Study Weekend on Data Collection and Processing* pp 114-122.

Feher, V. A., Baldwin, E. P., & Dahlquist, F. W. Access of ligands to cavities within the core of a protein is rapid. *Nat Struct Biol* **3**, 516-21 (1996).

Fersht, A. *Enzyme Structure and Mechanism*, (W.H. Freeman & Company, New York, 1985).

Fowler, A., & Zabin, I. Purification, structure, and properties of hybrid β-galactosidase proteins. *J. Biol. Chemistry* **258**, 14354-14358 (1983).

Gallagher, C. N., & Huber, R. E. Stabilities of uncomplemented and complemented M15 beta-galactosidase (Escherichia coli) and the relationship to alpha-complementation. *Biochem Cell Biol* **77**, 109-18 (1999).

Gaskell, A., Crennell, S., & Taylor, G. The three domains of a bacterial sialidase: a beta-propeller, an immunoglobulin module and a galactose-binding jelly-roll. *Structure* **3**, 1197-205 (1995).

Gebler, J. C., Aebersold, R., & Withers, S. G. Glu-537, not Glu-461, is the nucleophile in the active site of (lac Z) beta-galactosidase from Escherichia coli. *J Biol Chem* **267**, 11126-30 (1992).

Genetics_Computer_Group. *Program Manual for the Wisconsin Package, Version 8, 575*, , Madison, 1994).

Goldberg, M. E. Tertiary Structure of *Escherichia coli* β–D-galactosidase. *JMB* **46**, 441-446 (1969).

Hahn, M., Keitel, T., & Heinemann, U. Crystal and molecular structure at 0.16-nm resolution of the hybrid Bacillus endo-1,3-1,4-beta-D-glucan 4-glucanohydrolase H(A16-M). *Eur J Biochem* **232**, 849-58 (1995).

Hakda, S. The Properties of β–Galactosidases from Escherichia coli with Substitutions for Glycine 794 and Typtophan 999. *Master's Thesis in Biological Sciences* (University of Calgary, Calgary, 1997).

Hall, B. G., Betts, P. W., & Wootton, J. C. DNA Sequence Analysis of Artificially Evolved *ebg* Enzyme and *ebg* Repressor Genes. *Genetics* **123**, 635-648 (1989).

Harp, J., Hanson, B., Timm, D., & Bunick, G. Macromolecular crystal annealing: evaluation of techniques and variables. *Acta Crystallographica* **D55**, 1329-1334 (1999).

Harp, J., Timm, D., & Bunick, G. Macromolecular Crystal Annealing: Overcoming Increased Mosaicity Associated with Cryocrystallography. *Acta Crystallographica* **D54**, 622-628 (1998).

Heightman, T., & Vasella, A. (Personal Communication., 1994).

Henrissat, B., & Bairoch, A. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* **293**, 781-8 (1993).

Henrissat, B., & Bairoch, A. Updating the sequence-based classification of glycosyl hydrolases [letter]. *Biochem J* **316**, 695-6 (1996).

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., & Davies, G. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc Natl Acad Sci U S A* **92**, 7090-4 (1995).

Henrissat, B., & Davies, G. Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* **7**, 637-44 (1997).

Herrchen, M., & Legler, G. Identification of an essential carboxylate group at the active site of lacZ beta-galactosidase from Escherichia coli. *Eur J Biochem* **138**, 527-31 (1984).

Hobohm, U., & Sander, C. Enlarged representative set of protein structures. *Protein Sci* **3**, 522-4 (1994).

Hogness, D. S., Cohn, M., & Monod, J. Studies on the induced synthesis of beta-galactosidase in Escherichia coli: the kinetics and mechanism of sulfur incorporation. *Biochim Biophys Acta* **16**, 99 (1955).

Holm, L., & Sander, C. Protein structure comparison by alignment of distance matrices. *J Mol Biol* **233**, 123-38 (1993).

Hood, J. M., Fowler, A. V., & Zabin, I. On the evolution of β–galactosidase. *PNAS* **75**, 113-116 (1978).

Hsieh, H. B., & Da Silva, N. A. Development of a LAC4 promoter-based gratuitous induction system in Kluyveromyces lactis. *Biotechnol Bioeng* **67**, 408-16 (2000).

Huber, R. (Personal Communication., 2000).

Huber, R. E., & Brockbank, R. L. Strong inhibitory effect of furanoses and sugar lactones on beta-galactosidase Escherichia coli. *Biochemistry* **26**, 1526-31 (1987).

Huber, R. E., Gaunt, M. T., & Hurlburt, K. L. Binding and reactivity at the "glucose" site of galactosyl-beta-galactosidase (Escherichia coli). *Arch Biochem Biophys* **234**, 151-60 (1984).

Huber, R. E., Gupta, M. N., & Khare, S. K. The active site and mechanism of the beta-galactosidase from Escherichia coli. *Int J Biochem* **26**, 309-18 (1994).

Huber, R. E., Kurz, G., & Wallenfels, K. A quantitation of the factors which affect the hydrolase and transgalactosylase activities of beta-galactosidase (*E. coli*) on lactose. *Biochemistry* **15**, 1994-2001 (1976).

Huber, R. E., Parfett, C., Woulfe-Flanagan, H., & Thompson, D. J. Interaction of divalent cations with beta-galactosidase (Escherichia coli). *Biochemistry* **18**, 4090-5 (1979).

Huber, R. E., Roth, N. J., & Bahl, H. Quaternary structure, Mg2+ interactions, and some kinetic properties of the beta-galactosidase from Thermoanaerobacterium thermosulfurigenes EM1. *J Protein Chem* **15**, 621-9 (1996).

Huber, R. E., Wallenfels, K., & Kurz, G. The action of beta-galactosidase (Escherichia coli) on allolactose. *Can J Biochem* **53**, 1035-8 (1975).

Jacob, F., & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *JMB* **3**, 318-356 (1961).

Jacobson, R., & Matthews, B. Crystallization of β-Galactosidase from *Escherichia coli*. *Journal of Molecular Biology* **223**, 1177-1182 (1992).

Jacobson, R. H., Zhang, X.-J., DuBose, R. F., & Matthews, B. W. Three-dimensional structure of β-galactosidase from *E. coli*. *Nature* **369**, 761-766 (1994).

Jaeger, V. (Personal Communication., 1998).

Jain, S., Drendel, W. B., Chen, Z. W., Mathews, F. S., Sly, W. S., & Grubb, J. H. Structure of human beta-glucuronidase reveals candidate lysosomal targeting and active-site motifs. *Nat Struct Biol* **3**, 375-81 (1996).

Jenkins, J., Lo-Leggio, L., Harris, G., & Pickersgill, R. Beta-glucosidase, beta-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold beta/alpha architecture and with two conserved glutamates near the carboxy-terminal ends of beta-strands four and seven. *FEBS Lett* **362**, 281-5 (1995).

Jobe, A., & Bourgeois, S. *lac* Repressor-Operator Interaction VI. The Natural Inducer of the *lac* Operon. *JMB* **69**, 397-408 (1972).

Johnson, P. E., Joshi, M. D., Tomme, P., Kilburn, D. G., & McIntosh, L. P. Structure of the N-terminal cellulose-binding domain of Cellulomonas fimi CenC determined by nuclear magnetic resonance spectroscopy. *Biochemistry* **35**, 14381-94 (1996).

Juers, D. H., Huber, R. E., & Matthews, B. W. Structural comparisons of TIM barrel proteins suggest functional and evolutionary relationships between β–galactosidase and other glycohydrolases. *Protein Science* **8**, 122-136 (1999).

Jukes, T., & Cantor, C. (1969) in *Mammalian protein metabolism* (Munro, H., Ed.) pp 21-132, Academic Press, New York.

Juy, M., Amit, A. G., Alzari, P. M., Poljak, R. J., Claeyssens, M., Beguin, P., & Aubert, J.-P. Three dimensional structure of a thermostable bacterial cellulase. *Nature* **357**, 89-91 (1992).

Kabsch, W. Evaluation of single X-ray diffraction data from a position sensitive detector. *Journal of Applied Crystallography* **21**, 916-924 (1988).

Keresztessy, Z., Kiss, L., & Hughes, M. A. Investigation of the active site of the cyanogenic beta-D-glucosidase (linamarase) from Manihot esculenta Crantz (cassava). II. Identification of Glu-198 as an active site carboxylate group with acid catalytic function. *Arch Biochem Biophys* **315**, 323-30 (1994).

Koshland, D. E. Stereochemstry and the Mechanism of Enzymatic Reactions. *Biological Reviews*, 416-436 (1953).

Kuby, S. A., & Lardy, H. A. Purification and Kinetics of β–D-Galactosidase from *Escherichia coli. Biochemistry?* **75**, 890=896 (1953).

Lamzin, V., & Wilson, K. Automated refinement of protein models. *Acta Crystallographica* **D49**, 129-147 (1993).

Lawson, C. L., van Montfort, R., Strokopytov, B., Rozeboom, H. J., Kalk, K. H., de Vries, G. E., Penninga, D., Dijkhuizen, L., & Dijkstra, B. W. Nucleotide sequence and X-ray structure of cyclodextrin glycosyltransferase from Bacillus circulans strain 251 in a maltose-dependent crystal form. *J Mol Biol* **236**, 590-600 (1994).

Leahy, D. J., Hendrickson, W. A., Aukhil, I., & Erickson, H. P. Structure of a fibronectin type III domain from tenascin phased by MAD analysis of the selenomethionyl protein. *Science* **258**, 987-91 (1992).

Lederberg, J. The beta-D-galactosidase of Escherichia coli, strain K-12. **60**, 381-392 (1950).

Leslie, A. (1990) in *Crystallographic Computing*, Oxford University Press.

Levvy, G., & Conchie, J. (1966) in *Glucuronic Acid, Free and Combined* (Dutton, G., Ed.) pp 301-364, Academic Press, New York.

Lu, H. P., Xun, L., & Xie, X. S. Single-Molecule Enzymatic Dynamics. *Science* **282**, 1877-1882 (1998).

MacLeod, A. M., Lindhorst, T., Withers, S. G., & Warren, R. A. The acid/base catalyst in the exoglucanase/xylanase from Cellulomonas fimi is glutamic acid 127: evidence from detailed kinetic studies of mutants. *Biochemistry* **33**, 6371-6 (1994).

Martinez-Bilbao, M., Holdsworth, R. E., Edwards, L. A., & Huber, R. E. A highly reactive beta-galactosidase (Escherichia coli) resulting from a substitution of an aspartic acid for Gly-794. *J Biol Chem* **266**, 4979-86 (1991).

Martinez-Bilbao, M., & Huber, R. E. Substitutions for Gly-794 show that binding interactions are important determinants of the catalytic action of beta-galactosidase (Escherichia coli). *Biochem Cell Biol* **72**, 313-9 (1994).

McCarter, J. D., Adam, M. J., & Withers, S. G. Binding energy and catalysis. Fluorinated and deoxygenated glycosides as mechanistic probes of Escherichia coli (lacZ) beta-galactosidase. *Biochem J* **286**, 721-7 (1992).

McCarter, J. D., & Withers, S. G. Mechanisms of enzymatic glycoside hydrolysis. *Curr Opin Struct Biol* **4**, 885-92 (1994).

McIntosh, L. P., Hand, G., Johnson, P. E., Joshi, M. D., Korner, M., Plesniak, L. A., Ziser, L., Wakarchuk, W. W., & Withers, S. G. The pKa of the general acid/base carboxyl group of a glycosidase cycles during catalysis: a 13C-NMR study of bacillus circulans xylanase. *Biochemistry* **35**, 9958-66 (1996).

Mikami, B., Degano, M., Hehre, E. J., & Sacchettini, J. C. Crystal structures of soybean beta-amylase reacted with beta-maltose and maltal: active site components and their apparent roles in catalysis. *Biochemistry* **33**, 7779-87 (1994).

Miller, S., Janin, J., Lesk, A. M., & Chothia, C. Interior and surface of monomeric proteins. *J Mol Biol* **196**, 641-56 (1987a).

Miller, S., Lesk, A. M., Janin, J., & Chothia, C. The accessible surface area and stability of oligomeric proteins. *Nature* **328**, 834-6 (1987b).

Monod, J., Germaine, C.-B., & Cohn, M. Sur La Biosynthese del la β-Galactosidase (Lactase) chez *Escherichia coli*. La Specificite de l'Induction. *Biochimica et Biophysica Acta* **7**, 585-598 (1951).

More, N., Daniel, R. M., & Petach, H. H. The effect of low temperature on enzyme activity. *Biochemical Journal* **305**, 17-20 (1995).

Müller-Hill, B. *The lac Operon*, (Walter de Gruyter, New York, 1996).

Muller-Hill, B., & Kania, J. Lac repressor can be fused to beta-galactosidase. *Nature* **249**, 561-3 (1974).

Müller-Hill, B., Rickenberg, H. V., & Wallenfels, K. Specificity of the Induction of the Enzymes of the Lac Operon in *Escherichia coli. JMB* **10**, 303-318 (1964).

Murzin, A. G., Brenner, S. E., Hubbard, T., & Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247**, 536-40 (1995).

Musso, R., & Zabin, I. Substrate Specificity and Kinetic Studies on Thiogalactoside Transacetylase. *Biochemistry* **12**, 552-557 (1973).

Navas, J., & B'eguin, P. Site-directed mutagenesis of conserved residues of Clostridium thermocellum endoglucanase CelC. *Biochem Biophys Res Commun* **189**, 807-12 (1992).

Neville, M. C., & Ling, G. N. Synergistic Activation of β-Galactosidase by Na+ and Cs+. *Archived of Biochemistry and Biophysics* **118**, 596-610 (1967).

Nicholson, L. K., Kay, L. E., Baldisseri, D. M., Arango, J., Young, P. E., Bax, A., & Torchia, D. A. Dynamics of methyl groups in proteins as studied by proton-detected 13C NMR spectroscopy. Application to the leucine residues of staphylococcal nuclease. *Biochemistry* **31**, 5253-63 (1992).

Nichtl, A., Buchner, J., Jaenicke, R., Rudolph, R., & Scheibel, T. Folding and association of β-galactosidase. *JMB* **282**, 1083-1091 (1998).

Pardee, A. B., Jacob, F., & Monod, J. The Genetic Control and Cytoplasmic Expression of "Inducibility" in the Synthesis of ß-galatosidase by *E. coli. JMB* **1**, 165-178 (1959).

Parsons, S. M., & Raftery, M. A. Ionization behavior of the cleft carboxyls in lysozyme-substrate complexes. *Biochemistry* **11**, 1633-8 (1972).

Perrakis, A., Tews, I., Dauter, Z., Oppenheim, A. B., Chet, I., Wilson, K. S., & Vorgias, C. E. Crystal structure of a bacterial chitinase at 2.3 A resolution. *Structure* **2**, 1169-80 (1994).

Privalov, P., & Gill, S. The hydrophobic effect: a reappraisal. *Pure and Applied Chemistry* **61**, 1097-1104 (1989).

Rasmussen, D. H. & MacKenzie, A. P. Phase Diagram for the System Water-Dimethylsulfoxide. *Nature* **220**, 1315-1317 (1968).

Richard, J. P. The enhancement of enzymatic rate accelerations by Bronsted acid-base catalysis. *Biochemistry* **37**, 4305-9 (1998).

Richard, J. P., Huber, R. E., Lin, S., Heo, C., & Amyes, T. L. Structure-Reactivity Relationships for β-Galactosidase (*Escherichia coli*, lac Z). 3. Evidence that Glu-461 Participates in Bronsted Acid-Base Catalysis of β-D-Galactopyranosyl Group Transfer. *Biochemistry* **35**, 12377-12386 (1996).

Richard, J. P., Westerfield, J. G., & Lin, S. Structure-Reactivity Relationships for β–Galactosidase (*Escherichia coli*, lac Z). 2. Reactions of the Galactosyl-Enzyme Intermediate with Alcohols and Azide Ion. *Biochemistry* **34**, 11703-11712 (1995a).

Richard, J. P., Westerfield, J. G., Lin, S., & Beard, J. Structure-Reactivity Relationships for β–Galactosidase (*Escherichia coli*, lac Z). 1. Reactions of the Galactosyl-Enzyme Intermediate with Alcohols and Azide Ion. *Biochemistry* **34**, 11713-11724 (1995b).

Ring, M., Bader, D. E., & Huber, R. E. Site-directed mutagenesis of beta-galactosidase (*E. coli*) reveals that tyr-503 is essential for activity. *Biochem Biophys Res Commun* **152**, 1050-5 (1988).

Ring, M., & Huber, R. E. Multiple replacements establish the importance of tyrosine-503 in beta-galactosidase (Escherichia coli). *Arch Biochem Biophys* **283**, 342-50 (1990).

Rose, G., & White, A. Mechanism of catalysis by retaining b-glycosyl hydrolases. *Current Opinion in Structural Biology* **7**, 645-651 (1997).

Rosenberg, S., & Kirsch, J. F. Oxygen-18 Leaving Group Kinetic Isotope Effects on the Hydrolysis of Nitrophenyl Glycosides. 1. β–Galactosidase-Catalyzed Hydrolysis. *Biochemistry* **20**(1981).

Roth, N. J., & Huber, R. E. The beta-galactosidase (Escherichia coli) reaction is partly facilitated by interactions of His-540 with the C6 hydroxyl of galactose. *J Biol Chem* **271**, 14296-301 (1996).

Roth, N. J., Rob, B., & Huber, R. E. His-357 of beta-galactosidase (Escherichia coli) interacts with the C3 hydroxyl in the transition state and helps to mediate catalysis. *Biochemistry* **37**, 10099-107 (1998).

Sakon, J., Irwin, D., Wilson, D. B., & Karplus, P. A. Structure and mechanism of endo/exocellulase E4 from Thermomonospora fusca. *Nat Struct Biol* **4**, 810-8 (1997).

Sampson, N., & Knowles, J. Segmental Movement: Definition of the Structural Requirements for Loop Closure in Catalysis by Triose Phosphate Isomerase. *Biochemistry* **31**, 8482-8487 (1992).

Selwood, T., & Sinnott, M. L. A solvent-isotope-effect study of proton transfer during catalysis by Escherichia coli (lacZ) beta-galactosidase. *Biochem J* **268**, 317-23 (1990).

Sinnott, M. L. Ions, ion-pairs and catalysis by the lacZ beta-galactosidase of Escherichia coli. *Febs Lett* **94**, 1-9 (1978).

Sinnott, M. L. Catalytic Mechanisms of Enzymic Glycosyl Transfer. *Chemical Reviews* **90**, 1171-1202 (1990).

Sinnott, M. L., & Smith, P. L. Affinity Labelling with a Deaminatively Generated Carbonium Ion. *Biochemical Journal* **175**, 525-538 (1978).

Sinnott, M. L., & Souchard, I. J. The mechanism of action of beta-galactosidase. Effect of aglycone nature and -deuterium substitution on the hydrolysis of aryl galactosides. *Biochem J* **133**, 89-98 (1973).

Sinnott, M. L., & Withers, S. G. The beta-galactosidase-catalysed hydrolyses of beta-d-galactopyranosyl pyridium salts. Rate-limiting generation of an enzyme-bound galactopyranosyl cation in a process dependent only on aglycone acidity. *Biochem J* **143**, 751-62 (1974).

Sinnott, M. L., & Withers, S. G. The necessity of magnesium cation for acid assistance aglycone departure in catalysis by Escherichia coli (lacZ) beta-galactosidase. *Biochem J* **175**, 539-46 (1978).

Sneath, & Sokol. *Numerical Taxonomy*, (Freeman, San Fransisco, 1973).

Srajer, V., Teng, T.-y., Ursby, T., Pradervand, C., Ren, Z., Adachi, S.-i., Schildkamp, W., Bourgeois, D., Wulff, M., & Moffatt, K. Photolysis of the Carbon Monoxide Complex of Myoglobin: Nanosecond Time-Resolved Crystallography. *Science* **274**, 1726-1729 (1996).

Stokes, T. M., & Wilson, I. B. A common intermediate in the hydrolysis of β–galactosides by β-galactosidase from Escherichia coli. *Biochemistry* **11**, 1061-4 (1972).

Stout, G., & Jensen, L. *X-ray Structure Determination, A Practical Guide, 2nd Edition*, (John Wiley & Sons, New York, 1989).

Tenu, J. P., Viratelle, O. M., Garnier, J., & Yon, J. pH dependence of the activity of beta-galactosidase from Escherichia coli. *Eur J Biochem* **20**, 363-70 (1971).

Tews, I., Perrakis, A., Oppenheim, A., Dauter, Z., Wilson, K. S., & Vorgias, C. E. Bacterial chitobiase structure provides insight into catalytic mechanism and the basis of Tay-Sachs disease. *Nat Struct Biol* **3**, 638-48 (1996).

Tong, L., & Rossmann, M. The locked rotation function. *Acta Crystallographica* **A46**, 783-792 (1990).

Tormo, J., Lamed, R., Chirino, A. J., Morag, E., Bayer, E. A., Shoham, Y., & Steitz, T. A. Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *Embo J* **15**, 5739-51 (1996).

Tronrud, D. Unpublished Results. .

Tronrud, D. Knowledge-based B-factor restraints for the refinement of proteins. *Journal of Applied Crystallography* **29**, 100-104 (1996).

Tronrud, D. (1997) in *Methods in Enzymology* pp 306-319.
Ullmann, A. Complementation in β–Galactosidase: From Protein Structure to Genetic Engineering. *BioEssays* **14**, 201-205 (1992).

Ullmann, A., Jacob, F., & Monod, J. Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of Escherichia coli. *J Mol Biol* **24**, 339-43 (1967).

Ullmann, A., Perrin, D., Jacob, F., & Monod, J. Identification par Complementation in vitro et Purification d'un Segment Peptidique de la β–Galactosidase d'*Escherichia coli*. *JMB* **12**, 918-923 (1965).

Vriend, G. A molecular modeling and drug design program. *Journal of Molecular Graphics* **8**, 52-56 (1990).

Wagner, G., DeMarco, A., & Wuthrich, K. Dynamics of the aromatic amino acid residues in the globular conformation of the basic pancreatic trypsin inhibitor (BPTI). I. 1H NMR studies. *Biophys Struct Mech* **2**, 139-58 (1976).

Wallenfels, K., & Molhotra, O. P. Galactosidases. *Advances in Carbohydrate Chemistry* **16**, 239-298 (1961).

Wallenfels, K., & Weil, R. (1972) in *The Enzymes (VII)* (Boyer, P., Ed.).
Wang, Q., Tull, D., Meinke, A., Gilkes, N. R., Warren, R. A., Aebersold, R., & Withers, S. G. Glu280 is the nucleophile in the active site of Clostridium thermocellum CelC, a family A endo-beta-1,4-glucanase. *J Biol Chem* **268**, 14096-102 (1993).

Wentworth, D. F., & Wolfenden, R. Slow binding of D-galactal, a "reversible" inhibitor of bacterial beta-galactosidase. *Biochemistry* **13**, 4715-20 (1974).

Williams, J., & McDermott, A. Dynamics of the flexible loop of triosphosphate isomerase: the loop motion is not ligand gated. *Biochemistry* **34**, 8309-8319 (1995).

Wong, P., Gladney, S., & Keasling, J. D. Mathematical Model of the lac Operon: Inducer Exlusion, Catabolite Repression, and Diauxic Growth on Glucose and Lactose. *Biotechnology Progress* **13**, 132-143 (1997).

Xu, G. Y., Ong, E., Gilkes, N. R., Kilburn, D. G., Muhandiram, D. R., Harris-Brandts, M., Carver, J. P., Kay, L. E., & Harvey, T. S. Solution structure of a cellulose-binding domain from Cellulomonas fimi by nuclear magnetic resonance spectroscopy. *Biochemistry* **34**, 6993-7009 (1995).

Xue, Q., & Yeung, E. Differences in the chemical reactivity of individual molecules of an enzyme. *Nature* **373**, 681-683 (1995).

Yeh, J., & Hol, W. A flash-annealing technique to improve diffraction limits and lower mosaicity in crystals of glycerol kinase. *Acta Crystallographica* **D54**, 479-480 (1998).

Yuan, J., Martinez-Bilbao, M., & Huber, R. E. Substitutions for Glu-537 of beta-galactosidase from Escherichia coli cause large decreases in catalytic activity. *Biochem J* **299**, 527-31 (1994).

Zabin, I. β–Galactosidase alpha complementation. *Molecular and Cellular Biochemistry* **49**, 87-96 (1982).

Zeleny, R., Altmann, F., & Praznik, W. A capillary electrophoretic study on the specificity of beta-galactosidases from Aspergillus oryzae, Escherichia coli, Streptococcus pneumoniae, and Canavalia ensiformis (jack bean). *Anal Biochem* **246**, 96-101 (1997).

Zhang, X.-J., & Matthews, B. Enhancement of the method of molecular replacement by incorporation of known structural information. *Acta Crystallographica* **D50**, 675-686 (1994).

Zhang, X.-J., & Matthews, B. EDPDB: A multi-functional tool for protein structure analysis. *Journal of Applied Crystallography* **28**, 624-630 (1995).

Zu-Kang, F., & Sippl, M. J. Optimum superimposition of protein structures: ambiguities and implications. *Folding & Design* **1**, 123-132 (1996).