Protein Engineering, Design & Selection vol. 23 no. 2 pp. 103–113, 2010
Published online December 11, 2009 doi:10.1093/protein/gzp076

Analysis of galactosemia-linked mutations of GALT enzyme using a computational biology approach

A. Facchiano and A. Marabotti[1]

Institute of Food Science, CNR, Via Roma, 64, 83100 Avellino, Italy

[1]To whom correspondence should be addressed.
E-mail: amarabotti@isa.cnr.it

Received April 20, 2009; revised September 1, 2009; accepted November 9, 2009

Edited by Luis Serrano

We describe the prediction of the structural and functional effects of mutations on the enzyme galactose-1-phosphate uridyltransferase related to the genetic disease galactosemia, using a fully computational approach. One hundred and seven single-point mutants were simulated starting from the structural model of the enzyme obtained by homology modeling methods. Several bioinformatics programs were then applied to each resulting mutant protein to analyze the effect of the mutations. The mutations have a direct effect on the active site, or on the dimer assembly and stability, or on the monomer stability. We describe how mutations may exert their effect at a molecular level by altering H-bonds, salt bridges, secondary structure or surface features. The alteration of protein stability, at level of monomer and/or dimer, is the main effect observed. We found an agreement between our results and the functional experimental data available in literature for some mutants. The data and analyses for all the mutants are fully available in the web-accessible database hosted at http://bioinformatica.isa.cnr.it/GALT.

Keywords: database/galactose metabolism/galactose-1-phosphate uridyltransferase/single point mutants/structure prediction

Introduction

The study of protein mutations is of high interest to unravel the complex relationships between structure, function and dynamics in a protein, and also for biotechnological and medical applications. In particular, the study of inherited pathologies would greatly benefit of the ability to interpret the effects caused by genetic mutations not only in terms of sequence alteration, but also in terms of their impact on protein structure–function relationships. However, for the last decades, the attention of people working on genetic diseases has been focused mainly at determining the type and frequency of mutations present at genetic level, or their clinical outcome on the patients, often discarding what happens to the protein side. This has raised up the amount of information available about the variability of genes, leading to a corresponding increase in the number of databases and websites devoted to cataloging disease-associated mutations or polymorphisms (Wishart, 2006). Although these information are necessary for the knowledge and diagnosis of the disease, they are not sufficient to understand what really happens in the cell, if the proteins coded by the genes are not taken into account.

More recently, the interest for the analysis at protein level has increased, and few sophisticated online resources such as MutDB (Dantzer *et al*., 2005) have been developed to visualize the location of mutations directly on protein structures, when available, or to identify the likely underlying molecular effects of mutations on the basis of scores predicted by multiple sequence alignments (Ng and Henikoff, 2001). Obviously, the availability of the structures of the mutants and analyses performed on them would give more complete and direct information on the effects of mutation on protein structure and function. Yet, the study of the molecular consequences of protein mutations is very hard when a protein can be found in several different mutant forms. This is especially true for many genetic diseases that are typically caused by dozens or even hundreds of different point mutations at gene level causing dysfunctions at protein level (Beaudet *et al*., 2006). In these cases, it could be very difficult to obtain a suitable amount of protein in laboratory for so many mutants, in a reasonable time and with a low cost. Therefore, an experimental approach might be not applicable.

The simulation of protein structures is a well-established issue with reliable results in many cases, as proved by the last CASP experiment (Moult *et al*., 2009), and thanks to continuous improvements in hardware and software, the results are available in a reduced amount of time. Several programs are also available to analyze the structures of the proteins, in order to understand the structural or functional effects caused by mutations. It is evident that computational approaches suffer for the limits of methodology and are not as reliable as experimental studies. However, despite their potential errors, they can suggest interpretations of the molecular phenomena that keep place in the protein or help researchers to direct their effort towards a specific study. Therefore, this procedure could be successfully used to face the problem of understanding the effects of hundreds of mutants on a protein structure, as it has been made in some published examples (see for example Martin *et al*., 2002; Garza-Garcia *et al*., 2007; Rovida *et al*., 2007).

In this paper, we present the analysis performed on mutations affecting galactose-1-phosphate uridyltransferase (GALT) (E.C. 2.7.7.12). This enzyme is a member of branch III of the histidine triad superfamily and it shows a specific nucleoside monophosphate transferase activity (Brenner, 2002). GALT is a homodimer and the structures of the protein from *Escherichia coli* (Wedekind *et al*., 1995, 1996; Thoden *et al*., 1997) reveal for each monomer a single domain architecture composed of 6 α-helices and a β-sheet formed by 13 antiparallel and 1 parallel strand. Two active sites are present in the enzyme, each one formed by amino acid residues contributed by both subunits (Thoden *et al*., 1997). The catalytic mechanism of this enzyme has been extensively studied in the past (Frey, 1996; Geeganage and

© The Author 2009. Published by Oxford University Press. All rights reserved.
For Permissions, please e-mail: journals.permissions@oxfordjournals.org

Downloaded from http://peds.oxfordjournals.org/ at CharitÃ© -Med. Bibliothek on September 3, 2013

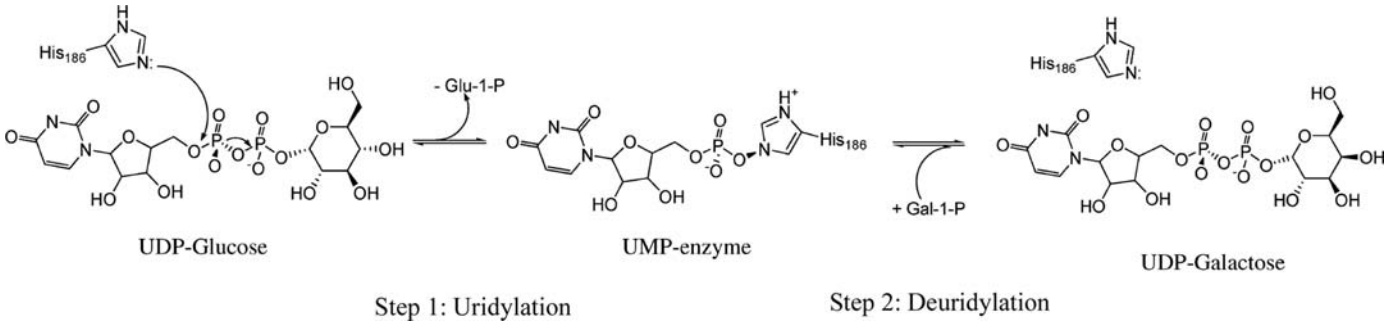

Fig. 1. Scheme of the two steps of the reaction catalyzed by GALT.

Frey, 1998) and is described by a ping-pong kinetics (Fig. 1). In the first step, the histidine of the active site attacks the R-phosphorus of UDP-glucose, displacing glucose-1-phosphate and forming a covalent intermediate (UMP-enzyme). In the second step, this intermediate reacts with galactose-1-phosphate that displaces the histidine to produce UDP-galactose.

The genetic disorder called 'classical galactosemia' (OMIM: #230400) is linked to the impairment of this enzyme caused by more than 180 sequence variations in the GALT gene, of which about 150 are missense mutations (Elsas and Lai, 1998; Tyfield *et al*., 1999; Tyfield, 2000; Fridovich-Keil and Walter, 2006). Symptoms can raise from mild to severe and include gastrointestinal complaints, hepatomegaly, cataracts, mental retardation and ovarian failure in females. Dietary restrictions can produce a rapid clinical improvement, but not necessarily ensure normal physical growth and cognitive development (Tyfield, 2000; Fridovich-Keil and Walter, 2006). Classical galactosemia is also characterized by a high allelic heterogeneity, with a typical distribution of mutations among several populations and ethnic groups (Tyfield, 2000). Taken together, the high level of allele heterogeneity, clinical variations among people carrying the same mutation(s) and metabolic factors have impaired to date to establish a strict and reciprocal relationship between mutations and illness, and this represents an important issue of the 'galactosemia enigma' (Segal, 1998). Some biochemical characterizations of the most frequent GALT mutants were made in the past (Elsevier and Fridovich-Keil, 1996; Elsevier *et al*., 1996; Lai *et al*., 1998, 1999; Crews *et al*., 2000; Henderson *et al*., 2000; Lai and Elsas, 2001; Riehman *et al*., 2001; Christacos and Fridovich-Keil, 2002; see also Fridovich-Keil and Walter, 2006 and references therein), but the high number of possible mutations and especially the absence of structural information on the human enzyme impaired the complete characterization of the effect of all mutations at protein level.

Considering that human GALT shares more than 40% sequence identity with the enzyme from *E.coli*, previously in our laboratory we created by homology modeling methods a model of the structure of wild-type enzyme and of its most frequent mutation, Q188R (Marabotti and Facchiano, 2005). In that work, the analysis of these structures allowed us to identify the position and the interactions of the mutant residue, and to investigate its structural and functional features, such as amino acid conservation, inclusion in secondary structures, solvent accessibility, interactions with the substrate and between the two monomers of the enzyme. This allowed us to hypothesize how Q188R mutation affects protein activity (hypotheses in agreement with experimental data) and to suggest a molecular explanation for the partial dominant negative effect caused by this mutation on heterodimeric association of GALT subunits (Elsevier and Fridovich-Keil, 1996).

During the past years, we decided to expand our case set and to simulate the effect of known missense mutations on GALT structure and activity, by creating the mutants and by analyzing the impact of each mutation on structure, function and quaternary assembly of GALT enzyme. In this paper, we present and discuss the results of our simulations, selecting some examples for each class of effects. The full results of our works are stored in the database freely accessible to all people via a web-based interface that we developed for this scope (http://bioinformatica.isa.cnr.it/GALT/) (d'Acierno *et al*., 2009), with the aim of sharing these data with the largest number of people in an interactive and up-to-date way, and also to allow people to interact providing us new mutations and/or suggesting new analyses.

Materials and methods

The 3D model of human GALT enzyme (Marabotti and Facchiano, 2005) was used as the starting point to simulate the effect of mutations. It was deposited in the PDB database (Berman *et al*., 2007) (PDB code: 1R3A) and thus it is freely available to scientific community. Information about known gene mutations has been retrieved from past literature (see Elsas and Lai, 1998; Tyfield *et al*., 1999, for review) and in the public database of GALT mutations at genetic level (GALTdb) developed by Calderon *et al*. (Calderon *et al*., 2007), available at: http://arup.utah.edu/database/galactosemia/GALT_welcome.php. Only missense mutations with published references were selected for our project.

We modeled the structures of GALT mutants using a Python script which is based on the program MODELLER v. 9.2 (Sali and Blundell, 1993), freely downloadable from the Modeller website: http://salilab.org/modeller/wiki/Mutate_model. This script implements a fully automated procedure to model mutations in protein structures (Feyfant *et al*., 2007). For mutations affecting residues with at least one atom within 5 Å from the cavity of the substrate, we created also the complex between the enzyme and the UDP-galactose, by merging the ligand into the active site, and then we applied a mild minimization using 500 steps of Steepest Descent method with a gradient limit of 0.1 kcal/mol·Å, in order to relax the possible steric clashes, in analogy with the

Downloaded from http://peds.oxfordjournals.org/ at CharitÃ© -Med. Bibliothek on September 3, 2013

procedure adopted for the wild-type GALT complexes (Marabotti and Facchiano, 2005).

An analysis of the impact of mutation on the protein structure has been made first by visual inspection of the enzyme using the Insight II package (Version 2000.1, Accelrys, Inc.; 2000). Protein structure analysis software DSSP (Kabsch and Sander, 1983), NACCESS (Hubbard *et al*., 1991) and HBPLUS (McDonald and Thornton, 1994) have been used to detect variation on secondary structures, relative solvent accessibility and H-bond patterns, respectively. Mutations were predicted to produce effects on the secondary structure to which the mutant residue belongs when a different code is identified by DSSP for the wild-type and mutated residue. Solvent accessibility was classified into states accordingly to a commonly used definition (Rost and Sander, 1994): buried, intermediate, exposed. Accessibility was considered significantly modified when two conditions were simultaneously verified: the variation of the state and a difference of more than 10% between the sums of the relative accessibilities of the two subunits in the dimeric wild-type and mutant protein. Differences in the H-bond pattern in which the residue is possibly involved are detected by comparing the results of HBPLUS with those of the wild-type residue. Identification of salt bridges was made according to Kumar and Nussinov (Kumar and Nussinov, 1999), and comparison was made with the native residue. To investigate possible interchain effects, we adopted three criteria at decreasing priority: (i) the presence of H-bonds or salt bridge with residues belonging to the opposite subunit; (ii) the decrease in accessibility of the residue after dimer assembly; (iii) the distance between the residue of interest and residues belonging to the opposite subunit lower than the threshold value of 5 Å.

Each mutant structure was then submitted to two different web servers to predict the mutation-induced change of protein stability with respect to the wild-type enzyme: PoPMuSiC (Gilis and Rooman, 2000) and DMUTANT (Zhou and Zhou, 2002). Since the two servers use different criteria to evaluate the impact of mutations on stability (the former estimates the changes in stability using linear combinations of database-derived potentials, whose coefficients depend on the solvent accessibility of the mutated residues; the latter is based on DFIRE all-atom potential, obtained using a new distance-scaled, finite ideal-gas reference state), we decided to consider reliable results only for whom both predictors reached a consensus, and the mutant protein was classified as 'less stable', 'unchanged' or 'more stable' with respect to the wild-type protein. When a consensus between the server is not reached, the effect of mutation on protein stability is classified as 'not determined'.

Finally, in addition to structural analyses, we also performed an investigation about the conservation of each residue in the whole GALT family. Mutations affecting conserved residues could in fact result in more negative effects on the enzyme structure and activity since it is supposed that residues conserved among the entire family play a major role in proteins belonging to that family. We used the Scorecons server at http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/valdar/scorecons_server.pl (Valdar, 2002). Given a multiple sequence alignment file, Scorecons calculates the degree of amino acid variability in each column of the alignment and returns this information to the user. We applied a mutation data matrix score (Valdar and Thornton, 2001) to an alignment based on 41 GALT sequences from different organisms: 3 from vertebrates (human, rat and mouse), 2 from invertebrates (fruitfly and *C.elegans*), 1 from plants (*A.thaliana*), 2 from yeasts (*S.cerevisiae* and *K.lactis*), and the others from several prokaryotic systems. For each residue, we obtained a score representing its conservation in the GALT family (Supplementary data available at *PEDS* online, Fig. S1).

Results and discussion

3D model of wild-type GALT enzyme

The 3D model of GALT enzyme was created using MODELLER program, as described in Marabotti and Facchiano (Marabotti and Facchiano, 2005). Briefly, the structures of GALT enzyme from *E.coli* complexed with UDP-galactose and UDP-glucose (Thoden *et al*., 1997) were used as templates to model the monomeric structure of human GALT enzyme. Two monomers were then assembled in the native dimeric form using the templates as reference, and the complex was submitted to a mild minimization in order to relieve steric clashes without distorting too much the structure. The evaluation of stereochemical and energetic properties of our model, as well as the assessment made by ADIT validation server at the time of deposition in PDB database, confirmed its overall good quality. The reliability of models obtained by comparative modeling procedures is dependent essentially on two factors: the extent of structural divergence between the target and the template and the quality of the sequence alignment between the two protein sequences. An analysis made by Chakravarty *et al*. on the accuracy of comparative models of proteins showed that the error in models based on approx. 40% template-target sequence identity is of the same order of magnitude as the differences between structure-derived properties measured from NMR and X-ray structures of the same protein (Chakravarty *et al*., 2005). Considering that in our case, the sequence identity between template and target is higher than 40%, that the two sequences can be unequivocally aligned, and that our model was created using multiple templates and was further refined, we can be confident that our model may be comparable in accuracy to structures determined by NMR spectroscopy.

3D model of mutants of GALT enzyme

Several programs have been developed to predict side-chain conformation in a protein and to introduce single-point mutations in a structure, and most of them are freely available on the web (for a review, see Marabotti, 2008). In general, these programs are developed to repack the side chains in a protein, or to add side chains to the structure of a backbone. On the contrary, the procedure included in Mutate_model script implemented in MODELLER 9v2 that we used to create the models of the 107 mutants of GALT enzyme is especially developed to introduce point mutations in a protein structure (Feyfant *et al*., 2007). It acts with a combination of two cycles of conjugate gradient minimization and molecular dynamics with simulated annealing, and it is possible to set an 'environment' formed by atoms at a distance of $\sim$5 Å. In the first cycle, only those non-bonded atom pairs that contain the set of atoms selected for

Downloaded from http://peds.oxfordjournals.org/ at CharitÃ© -Med. Bibliothek on September 3, 2013

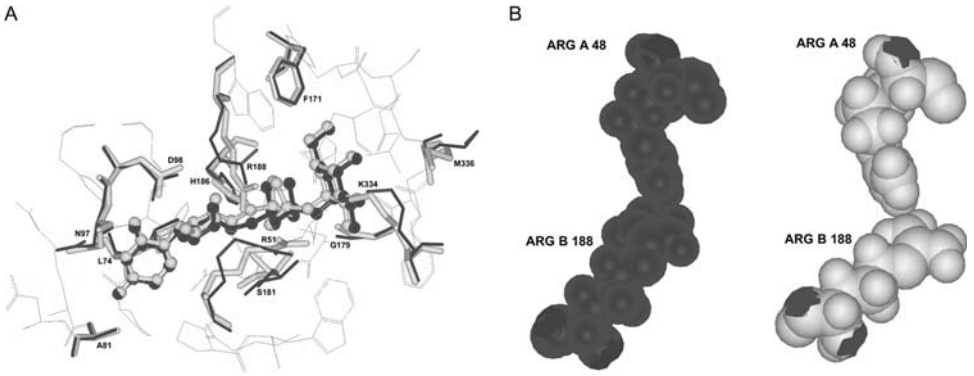

Fig. 2. (**A**) Comparison between the active sites of the model of GALT Q188R mutant simulated in our previous work (Marabotti and Facchiano, 2005) (light gray) and with the current procedure (dark gray). UDP-galactose is shown in ball-and-stick mode, residues of the protein involved in other galactosemia-related mutations are in stick mode and labeled, the other residues in the active site are represented with thin lines. (**B**) Close-up of the steric clash between Arg 188 and Arg 48 of the opposite subunit in our previous model (light gray) and in the current model (dark gray).

optimization are taken into account (i.e. the side chain does not 'feel' its environment). In the second cycle, the atom pairs that contain up to one environment atom are also included in the energy function (i.e. the side chain does 'feel' its environment). Moreover, the scoring function takes advantage of homology-derived dihedral angle restraints. An 'in-house' comparison of the results of some very popular programs showed that this procedure, applied to the simulation of single-point mutants of proteins, outperforms other programs (Marabotti and Facchiano, manuscript in preparation). For these reasons, we decided to apply this program to simulate our single-point mutants.

In our previous work (Marabotti and Facchiano, 2005), we used a different and more time-consuming procedure to create the model of mutant Q188R, since the one applied in this work was not available. In order to check for the reproducibility of the results obtained with these two different procedures, we decided to re-simulate this mutation with the current procedure and to compare the results. The overall RMSD between the mutants obtained with the two simulations is 0.56 Å and generally the positions of the residues within 5 Å from the substrate and involved in other galactosemia-linked mutations are conserved, with the exception of Ser 181 of the same subunit, and Arg 51, Lys 334 and Met 336 of the opposite subunit (Fig. 2A). The guanidinium group of the side chain of Arg 188 simulated with the new procedure is about 1.7 Å far from the guanidinium group of Arg 188 simulated previously, but also in the newly simulated mutant there is the contact with Arg 48 of the opposite side chain (Fig. 2B), discussed in our previous work and considered responsible of the partial dominant negative effect. Moreover, in both models of Q188R mutant, the H-bond pattern of the wild-type residue is similarly altered. This allows us to confirm that the procedure we adopted in this work produces results in agreement with the ones published previously.

Data set creation and overview of the results

The data set analyzed is formed by 107 different GALT missense mutations modeled with the procedure described in Materials and Methods section. A summary of the mutations analyzed and of their main effects is reported in Supplementary data available at *PEDS* online, Table S1. The data and analyses for all the mutants are fully available in our website http://bioinformatica.isa.cnr.it/GALT/ (d'Acierno *et al*., 2009), together with the models we created, that are downloadable in PDB format.

In the following paragraphs, we will describe in more details the structural effects of mutations: in particular, we will explore if they affect the active site of the enzyme, or intersubunit communications, or protein stability, in order to find, whenever possible, a molecular explanation for the impairment of the activity of GALT enzyme.

Mutations with predicted impact on active site

To identify the active site of GALT enzyme, we selected all the residues with at least one atom within 5 Å of distance from the substrate. Residues included in this selection and affected by mutations in galactosemia-linked GALT enzymes are Leu 74, Ala 81, Asn 97, Asp 98, Phe 171, Gly 179, Ser 181 and Gln 188 of the same subunit, Arg 51, Lys 334 and Met 336 of the other subunit. In addition to analyses obtained with the software DSSP, NACCESS and HBPLUS, we made a visual inspection of their position and interactions, taking as a reference the wild-type enzyme bound to the same ligand.

The most studied mutation belonging to this group is Q188R (Elsas and Lai, 1998; Lai *et al*., 1999; Tyfield *et al*., 1999; Fridovich-Keil and Walter, 2006). As already discussed in our previous work (Marabotti and Facchiano, 2005) and as confirmed here, this mutation can affect not only the catalytic activity, but also intersubunit relationships. Gln 188 is a fully conserved residues in the whole GALT family (Supplementary data available at *PEDS* online, Figure S1), with the role of stabilizing the catalytic intermediate and probably of favoring the nucleophilic attack of galactose-1-phosphate (Lai *et al*., 1999). Its replacement with Arg, which is unable to interact with the phosphoric moiety of ligands in the same way as Gln 188, as proved by its different hydrogen bonds network, causes the loss of activity of the enzyme. Moreover, the contact between Arg 188 and the side chain of Arg 48 belonging to the other subunit (Fig. 2B) allows to give a molecular explanation for the partial dominant negative effect exerted by the Q188R mutation on heterodimer activity.

Downloaded from http://peds.oxfordjournals.org/ at CharitÃ© -Med. Bibliothek on September 3, 2013

Arg 51 forms a network of interactions with Arg 48, Arg 328 and Asp 348, and is also able to create H-bonds with the oxygen atoms belonging to the phosphate moiety of the substrate. Its replacement with Leu disrupts the network of contacts with the other charged residues, and the shorter side chain of this hydrophobic residue is placed near the ribose moiety, causing a steric hindrance to the substrate. A similar effect is caused by the replacement of Arg 51 with Gln.

The methyl group of Ala 81 is located at ~5 Å of distance from the centroid of the uridyl moiety of the substrate. Its replacement with Thr increases the steric hindrance in this part of the enzyme, and this would likely have a negative impact on substrate binding. Moreover, this residue is placed at the boundary between the two subunits of the enzyme and may affect intersubunit relationships.

Asn 97 forms H-bonds with the oxygen in positions O2′ and O3′ of the ribose moiety of the substrate, and with the oxygen O2 of the uridyl moiety, and it is also H-bonded to Cys 126 with both its backbone nitrogen and its side chain. The visual inspection of the structure shows that Asn 97 and Gln 188 might form an H-bond (not recognized by HBPLUS because the angle H-A-AA is slightly less than the lower threshold limit fixed by the program to find H-bonds automatically). Mutation N97S causes the loss of the H-bonds with the ribose, and the alteration of the H-bond pattern of the residue, with the formation of a new H-bond with the backbone nitrogen of Phe 99. Moreover, there are no longer conditions to form the putative H-bond with Gln 188, which may be important to keep this last residue in the correct place and/or to promote the catalytic action. This mutation is also reported as resulting in a cryptic splice acceptor site in exon 3, resulting in shorter protein (Calderon *et al*., 2007).

Asp 98 interacts with nitrogen N3 of the uridyl moiety of the substrate. Its replacement with Asn, although this is generally considered a conservative substitution, disrupts this interaction, since the side chain of Asn is pointing away from the substrate. A possible explanation for this fact may be that Asp 98 is kept in the correct position by an ion pair with Arg 80. The loss of the negative charge impairs the formation of this interaction, and then the position of the mutant side chain is changed.

Phe 171 is a conserved residue in the 16 sequences more similar to human GALT sequence, whereas in the others a Gly is present. In the wild-type enzyme, Phe 171 is H-bonded with its backbone nitrogen to the backbone oxygen of Gln 188, and the phenyl side chain lies near Asn 173. This last residue is involved in H-bond with the oxygen atom O2A in the phosphate moiety of the substrate, *via* its amidic group. In mutant F171S, there is a different pattern of H-bonds and probably this is the cause of the impairment of the activity of this mutant. Moreover, this mutation replaces a hydrophobic residue in the core of the protein with a polar one, with possible destabilizing effects. Mutation F171S was simulated in a previous work, together with unnatural mutations F171L and F171Y (Crews *et al*., 2000). In that work, the effect of this mutation was explained by a hypothetical displacement of Gln 188 from its wild-type position, with the creation of a new H-bond between the amide group of this residue and the oxygen of Ser 171. In our model, the distance between the amidic group of Gln 188 and the oxygen of Ser 171 is more than 7.5 Å, and it appears unlikely that this mutation induces the formation of such H-bond. These different results can be ascribed to the use of different modeling procedures which evidently generate models with some differences.

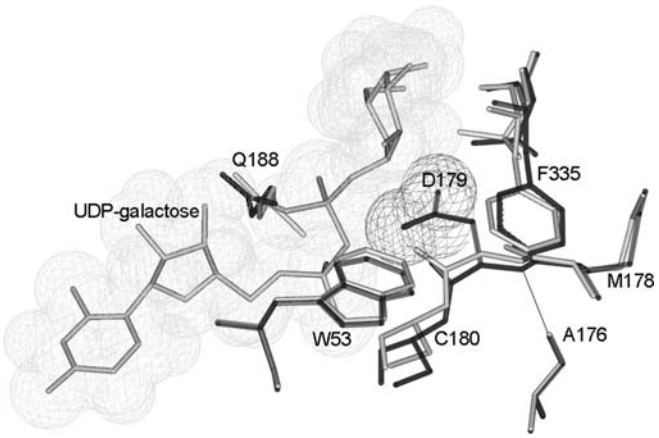

Fig. 3. Comparison of the structure of wild-type GALT (light gray) and mutant G179D (dark gray). The van der Waals surface of UDP-galactose and of the carboxilic moiety of D179 is represented as a sphere with lines. Residues near the mutation and cited in the text are shown in stick mode and labeled.

Gly 179 is a strictly conserved residue in all GALT enzymes (Supplementary data available at *PEDS* online, Figure S1). It is placed near the galactose or glucose moiety of substrate and also at the subunit interfaces, near two bulky residues (Trp 53 and Phe 335) of the opposite subunit. This particular structural location explains why there is the need of a very small residue in this position. Its replacement with Asp introduces a bigger and charged residue in this zone, with an obvious negative impact on substrate binding and also on intersubunit interactions. The side chain of Asp 179 causes a steric clash with the sugar moiety of the substrate, and also determines the displacement of several residues (Asn 173, Met 178, Cys 180 and Gln 188 of the same subunit, Phe 335 and Val 337 of the other subunit). Moreover, the conformation of the backbone of this segment is altered, and a new H-bond is formed with Ala 176 (Fig. 3).

Lys 334 is a residue able to form H-bonds either with the substrate (atom O1 of the galactose/glucose moiety) and with Gln 346, another residue belonging to the active site interacting with substrate, and also a salt bridge with Glu 340. The new side chain of the mutant K334R is accommodated far from the substrate and the H-bond pattern is perturbed, with the loss of the interaction with both residues.

Finally, for three residues of the active site, Leu 74, Ser 181 and Met 336, it is difficult to interpret how their replacement, respectively, with Pro, Ala and Leu could perturb the active site of the enzyme. Ser 181 interacts with the phosphate oxygen of the substrate *via* its backbone nitrogen. However, its replacement with Ala seems not to have a major impact on substrate binding and/or catalysis. Leu 74 and Met 336 are not interacting with the substrate although they are near the molecule. However, Met 336 is also at the interface of the two subunits of the enzyme, and its replacement could affect the correct assembly of the oligomeric protein.

Other mutations can affect indirectly the active site when they involve residues that keep contacts with active site

Downloaded from http://peds.oxfordjournals.org/ at CharitÃ© -Med. Bibliothek on September 3, 2013

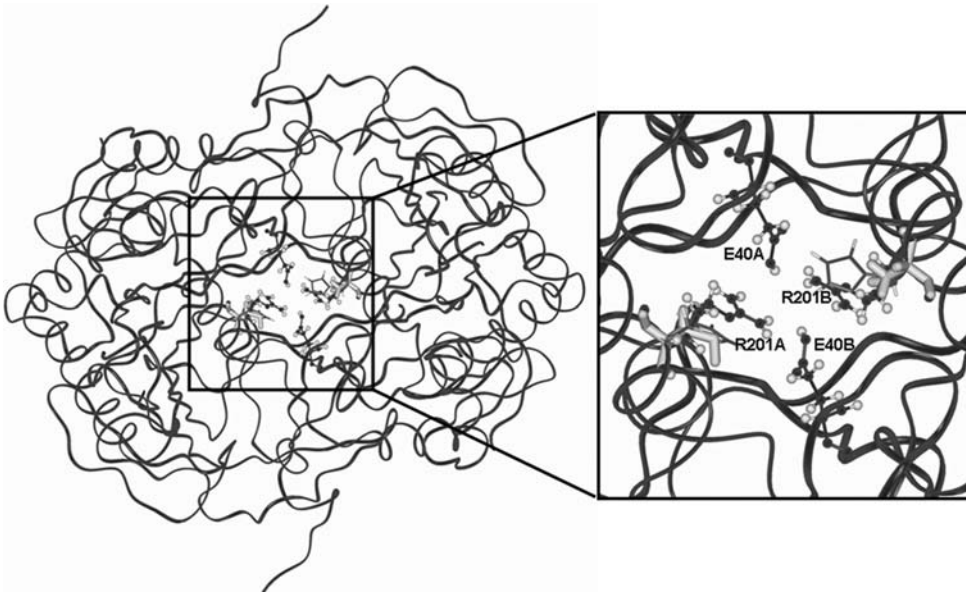

Fig. 4. Close-up of the network of salt bridges involving residues Arg 201-Glu 40 in the middle of GALT enzyme, in wild type and mutants R201H and R201C. Residues Arg 201 and Glu 40 are represented in ball-and-stick mode and dark gray, residue His 201 is represented in thin stick mode and medium gray, residue Cys 201 is represented in thick stick mode and light gray.

residues. Mutation W154G affects one of the most conserved residues in the protein (Supplementary data available at *PEDS* online, Figure S1). Trp 154 is inserted in helix H3, and is forming, by means of its indolic nitrogen, an H-bond with the sulfur moiety of Cys 187, placed in strand S6 of the central β-sheet, between His 186 (the catalytic residue) and Gln 188. Therefore, the loss of this interaction affects a crucial place for enzymatic activity. Moreover, the replacement of Trp 154, in the middle of a helix, with a helix-disfavoring residue (Gly) may also perturb the secondary structure in which this residue is inserted. Another mutation with a predictable effect on active site is R328H. Arg 328 makes H-bonds with Arg 51, and its replacement with His, a residue with a shorter side chain, disrupts this interaction.

Mutations with a predicted effect on interface relationships

Forty-eight galactosemia-related mutations involve residues <5 Å far from amino acids belonging to the other subunit. Among those, 30 mutations affect directly the interchain surface, as deduced by the differences in the exposure of the residue in the monomer towards the dimer, or by their involvement in interchain H-bonds or salt bridges (Supplementary data available at *PEDS* online, Table S1). The perturbation of these residues can have dramatic effects on intersubunit relationships, reducing GALT activity, and could also impair the correct *in vivo* dimerization of the protein. Some of the mutations belonging to this group (R51L, R51Q, L74P, A81T, N97S, F171S, G179D, S181A, Q188R, M336L and R328H) are also involved in active site perturbation, since the active site is formed by residues belonging to both subunits (Thoden *et al*., 1997), and were discussed in the previous paragraph.

Probably, the most critical effect in this group is represented by mutations affecting residue Arg 201. Arg 201 is involved in a salt bridge with Glu 40 of the opposite subunit. The two ion pairs Arg 201-Glu 40 from both subunits are placed exactly in the centre of the protein (Fig. 4), and they participate in the formation of cross-interactions that bring together the two monomers. The replacement of Arg 201 with Cys or His causes the disruption of this central unique network of interactions, and the two subunits of the protein are deprived of a very important intersubunit interaction.

Another interesting case is mutation G55C. The introduction of a Cys residue in this position may determine the formation of a new disulfide bridge. In fact, Cys 180, belonging to the other subunit, is placed near Gly 55, and the manual exploration of the conformational space of the two Cys residues in the mutant showed that the two sulfur moieties can reach a minimal distance of 2.74 Å, compatible with the formation of a disulfide bridge (Fig. 5). The formation of a double intersubunit disulfide bridge in the structure of this protein could obviously have a deep impact on the whole protein. Moreover, Cys 180 is placed near the active site, and therefore the enzymatic activity may be perturbed by the 'stiffening' of this part of protein.

Other residues involved in intersubunit H-bonds networks that are perturbed by galactosemia-related mutations are Arg 33, Ser 45, Gln 118, Leu 342 and Gln 344. On the contrary, there is no direct evidence of a perturbation of intersubunit H-bonds in mutants I32N and I198M/T that bind residues of the other subunit *via* their main-chain atoms. Other considerations, such as the variation of physicochemical features or of the steric hindrance of the residues, may explain the perturbation at the subunit interface.

Mutations that can perturb protein stability and structural organization

Apart from perturbations on the active site of the enzyme, or on intersubunit relationships, other mutations can have different effects on the structure of the enzyme, causing instability or incorrect folding.

We analyzed with two web servers, PoPMuSiC (Gilis and Rooman, 2000) and DMUTANT (Zhou and Zhou, 2002), the

Downloaded from http://peds.oxfordjournals.org/ at CharitÃ© -Med. Bibliothek on September 3, 2013

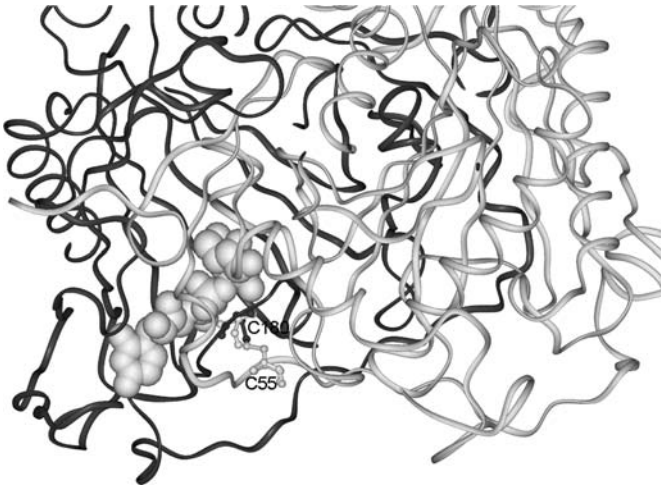

Fig. 5. Visualization of the position of disulfide bridge hypothetically formed in the G55C mutant, between C55 of subunit A (light gray) and C180 of subunit B (dark gray). The two residues are shown in stick mode. UDP-galactose is shown in CPK mode.

possible effect of mutations on GALT stability. We chose these servers since they are quite popular and both of them use as input data not only the sequence, but also the coordinates of the protein, therefore taking into account the information derived from the 3D structure.

Sixty-four mutations inserted in GALT are predicted by both servers to affect the stability of this protein. In 56 cases, the predicted alteration of protein stability is associated to the identification of another effect on the protein's structure by the analyses made on secondary structures, H-bond and salt bridge networks, and in 9 cases the variation of stability is apparently not directly associated with these structural effects. In some of these last cases (V151A, Y251S, I278N, L282V and P324S), the wild-type residue is deeply buried into the protein, and probably its replacement, especially with residues carrying polar moieties, may alter the hydrophobic interaction stabilizing the protein, or at least create a cavity, being in all cases the new side chain smaller than the original one. In the other cases, the residue is at the surface of the protein and no major changes are evident in side-chain conformations in the neighbors of each mutation. Probably long-range interactions, not valuable by our analyses, concur to induce destabilization.

With the aid of DSSP, we analyzed whether the mutations introduced were able to induce variations in the secondary structure to which they belong. The current protocol for introducing mutations allows only a limited rearrangement of backbone conformation; therefore, we cannot exclude that this kind of effects may be larger than those seen here. Further studies applying a more complex protocol to simulate more extensively the secondary structure destabilization and/or unfolding are planned in the future. Despite these limitations, mutations F117S, G179D, P183T and P185H are predicted to cause perturbation on the secondary structure to which they belong. Two of these mutations involve Pro residues that are generally critical for backbone conformation. In particular, mutation P183T is predicted to induce the gain of secondary structure (from a random coil to a bend conformation) in the strict proximity of the active site. Probably, the presence of a disordered region near the catalytic residues allows a higher flexibility that favors the catalysis, and the loss of this flexibility is responsible for the decrease of the activity. On the contrary, the other three mutations are predicted to disrupt the secondary structure in which they are inserted. In particular, we already discussed the case of G179D that involves a 100% conserved residue that is inserted in a zone with high steric impairment. The introduction of a bulkier and charged side chain can have a deep impact also on secondary structure of this segment.

In GALT, there are also mutations from other residues to Pro. Four of them (L139P, R204P, L217P and L226P) involve residues placed in helices. Although our models did not highlight alterations of secondary structures in these cases, we cannot exclude that the insertion of a known helix-disrupting residue in the centre of a helix may exert a negative effect on this structure.

Several mutations are predicted to introduce alterations in the network of interactions that the wild-type residue possesses with its neighbors. This may result both in short- and long-range effects with different consequences on protein structure and activity. Often, these networks of interactions seem to play a major role in connecting different secondary structure elements in the enzyme, thus keeping them in reciprocal contact and contributing to increase protein stability. Some examples are briefly discussed below.

K285N is the second most common mutation of GALT enzyme, and is prevalent in some countries of north central Europe, where it is found on 25–40% of mutant chromosomes, whereas it is rare in patients whose ancestries are non-European (Tyfield *et al*., 1999). Lys 285 is a residue enclosed in a very long helix (H8), which interacts *via* an H-bond with the backbone of Val 233 that belongs to the central β-sheet of the protein (strand S7). Our analyses suggest that mutation K285N disrupts this interaction because the side chain of Asn is shorter than that of Lys (Fig. 6). The loss of this interaction may cause instability to the protein structure.

Other similar examples are: Arg 33 is in the first strand (S1) of the protein forming an isolated small β-sheet, interacting with helix H10 *via* a H-bond with Glu 352. The replacement of Arg 33 with His causes the loss of this interaction. Lys 127 is placed in strand S4 belonging to the central β-sheet of GALT monomer and forms a salt bridge with Glu 160, that is the C-terminal residue of helix H3, one of the two long α-helices flanking the β-sheet. The charge inversion caused by replacement of Lys with Glu may induce repulsion and destabilize the favorable interactions between these two secondary structures. A similar effect may be induced by mutations affecting residue Arg 148, which is inserted in the middle of the same helix and forms a network of ion pairs with Asp 273, inserted in the other long helix (H8) flanking the central β-sheet. The replacement of Arg 148 with either Trp, or Gln, or Gly, disrupts this network of interactions and weakens the interactions between these two helices. Analogously, the replacement of residue Arg 272 with Gly in helix H8 disrupts the ion pair with Asp 152 in helix H3; the replacement of Arg 259 with Trp disrupts the ion pair with residue Glu 271, and the interaction between strand S9 of the central β-sheet and helix H8; the replacement of His 319 with Gln disrupts the contact between helix H5 and strand S11 *via* the ion pair between Glu 202 and His 319.

Downloaded from http://peds.oxfordjournals.org/ at CharitÃ© -Med. Bibliothek on September 3, 2013

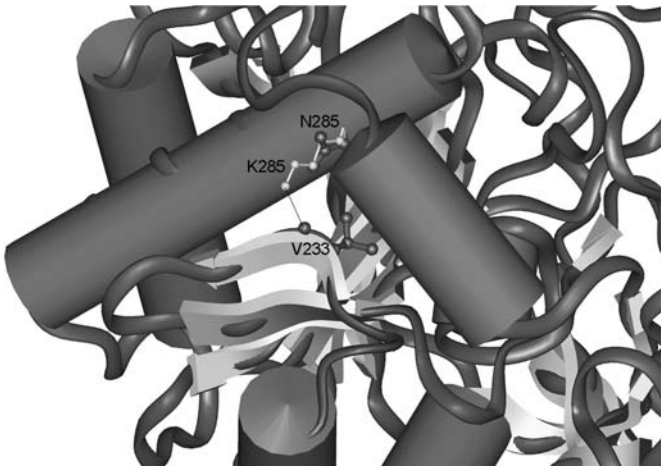

Fig. 6. Close-up of the interactions between residue 285 and residue 233 in wild-type and mutant K285N GALT enzyme. Residue K285 is represented in ball-and-stick mode and colored in light gray; residue N285 is colored in dark gray. The backbone of the protein is represented as a dark gray ribbon. α-Helices are represented as cylinders, and β-strands as light gray ribbons. The line represents the H-bond between K285 and V233, that is lost in K285N mutant.

Finally, we analyzed also if the introduction of a mutation can affect the solvent accessibility of the residue. In 16 cases, the mutation causes a significant variation in the solvent accessibility of the side chain of the new residue with respect to the former one. Often, this can be simply related to changes in the size of the side chain: for example, big residues exposed to solvent replaced by smaller ones show a decreased solvent accessibility simply because of their reduced dimensions, and *vice versa*. However, sometimes the change in size alone is not sufficient to justify these differences: for example, Leu 195 and Leu 217 in wild-type protein are buried residue, but their Pro mutants accommodate the ring in such a way that the side chain becomes partially exposed to solvent. In mutants R51L and R51Q, the two new side chains are not only shorter, but also differently oriented with respect to Arg. The side chain of Phe 194 is pointing towards the interior of the protein, whereas the side chain of Leu in F194L is pointing towards the exterior. The side chain of Ser 135 is small and buried in the wild-type protein, whereas the bigger side chain of Trp in the mutant S135W exposes a higher surface to the solvent; therefore, a large hydrophobic portion is present at the surface of the mutant, whereas in this portion of the wild-type protein there is a polar moiety. A variation in the solvent accessibility of these residues may cause anomalous interactions with other proteins, or instability.

Mutations with no predicted effects on the structure and function of the GALT enzyme

In 10 cases, our analyses on the structure of GALT mutant did not allow us to detect any perturbation able to justify the negative outcome on protein structure and function. This is the case for mutations R67C, D113N, H132Y, S135L, T138M, S143L, V150L, R258C, L289F and A330V.

Mutations R67C and R258C involve two charged surface residues that are replaced by Cys. Maybe this residue at the protein surface could promote the formation of intermolecular disulfide bridges with other molecules. In the other cases, however, the mutation involves residues not apparently involved in any known activity. A possible explanation could be that some of these mutations were identified only in combined heterozygous patients; therefore, a simple polymorphism with no negative effects on the protein activity was erroneously associated to galactosemia because of its presence in patients whose galactosemia is caused by other mutations. Another possible explanation is that these mutations can be associated to *in vivo* phenomena, such as the anomalous compartmentalization of the protein in subcellular organelles, or the premature degradation of the enzyme.

Comparisons between predictions and biochemical studies

Among the large quantity of papers describing GALT mutations, only a limited percentage of them report biochemical studies useful to test our predictions. For example, several studies report the identification of new mutations of GALT gene/protein, but often they are carried on heterozygous patients showing other mutations in addition to the newly identified ones and then it is not possible to isolate the effect of each single mutation on the final activity of the enzyme. Other studies report conflicting results when different expression models are used, or when different methodologies to test GALT activity in galactosemic patients are used. We selected few studies describing the biochemical characterization of isolated single-point mutants of GALT enzymes, expressed as homodimers or as wild-type heterodimers, and we compared our predictions to those data, assuming that the detection of an alteration on protein structure or stability could be linked to an impairment of its activity. In Table I, we report the list of mutants used as reference, together with the citation of the references for their experimental characterization, and the predicted molecular effects stored for each mutant in our database. We also made a 'reverse' test. N314D is now classified as a polymorphism, since it does not cause GALT impairment (Calderon *et al.*, 2007). We simulated also this mutation, though we did not include it in the data set, and in fact we did not predict any negative effect of this mutation on protein structure (data not shown).

Mutant Q188R is certainly the most studied one, with a complete molecular characterization of its effects. When we simulated this mutation the first time (Marabotti and Facchiano, 2005), we found that our predictions were in agreement with the experimental results about lack of activity of the homodimeric enzyme and partial dominant negative effect on heterodimer activity. In this work, we repeated the simulations with a different protocol, as we discussed in more details previously, and also in this case we found an agreement with experimental results. In particular, the involvement of this residue in intersubunit relationships is able to explain also the partial dominant negative effect responsible for an activity less of 50% in the heterodimeric wild-type/mutant enzyme.

Mutant F171S was expressed in yeast, purified and experimentally characterized (Crews *et al.*, 2000; Christacos and Fridovich-Keil, 2002). Experimental results clearly show that the homodimeric mutant enzyme is not active, and that the heterodimeric wild-type/mutant enzyme has about 15% activity, with a partial dominant negative effect. Our predictions state that the replacement of Phe 171 with Ser deeply affects the active site of the enzyme and the intersubunit

Downloaded from http://peds.oxfordjournals.org/ at CharitÃ© -Med. Bibliothek on September 3, 2013

Table I. Comparison between our predictions and mutants characterized experimentally

Mutation	References	% Activity (wt = 100%)[a]	Predicted alterations on protein structure[b]
R67C	Riehman *et al*., 2001	2.3	None
L139P	Riehman *et al*., 2001	1.9	Monomer stability
V151A	Fridovich-Keil *et al*., 1995; Riehman *et al*., 2001	3.1–4.6	Monomer stability
F171S	Crews *et al*., 2000; Riehman *et al*., 2001; Christacos and Fridovich-Keil, 2002	0	Substrate binding, intersubunit interactions, monomer stability
P183T	Riehman *et al*., 2001	45.2	Intersubunit interactions, monomer stability
Q188R	Elsevier *et al*., 1996; Elsevier and Fridovich-Keil, 1996; Lai *et al*., 1999; Riehman *et al*., 2001; Christacos and Fridovich-Keil, 2002	0	Substrate binding, intersubunit interactions, monomer stability
R201H	Riehman *et al*., 2001; Calderon *et al*., 2007	62.8	Intersubunit interactions, monomer stability
R231H	Ashino *et al*., 1995; Riehman *et al*., 2001	0–15.0	Intersubunit interactions, monomer stability
R259W	Riehman *et al*., 2001	0	Monomer stability
K285N	Riehman *et al*., 2001	0	Monomer stability
E291K	Riehman *et al*., 2001	62.8	Monomer stability
N314D[c]	Reichardt and Woo, 1991; Lai *et al*., 1998; Riehman *et al*., 2001; Christacos and Fridovich-Keil, 2002	100–102.5	None
Y323D	Riehman *et al*., 2001	9.6	Intersubunit interactions, stability
R333W	Elsevier *et al*., 1996; Riehman *et al*., 2001; Christacos and Fridovich-Keil, 2002	0	Intersubunit interactions
T350A	Riehman *et al*., 2001	9.9	Monomer stability

[a]When more than a value was present in the literature, the interval between the lower-upper value is reported in the table.
[b]as reported in the corresponding 'Summary of the effects of the mutation on the structural features of protein' available online at http://bioinformatica.isa.cnr.it/GALT.
[c]Not inserted in GALTprot database because classified as a polymorphism.

relationships. Therefore, our predictions are perfectly in agreement with experimental results.

In the work by Riehman *et al*., 2001, several human GALT mutants were expressed in yeast, and the GALT activity was assayed on soluble whole cell lysate. In that work, the mutants were not purified, and therefore there is not a complete biochemical characterization of the mutants, but the authors report only the different percentages of residual activity with respect to wild-type protein. Anyway, our predictions detected a perturbation of protein structure of all these mutants but R67C, which is reported to have about 2.3% of wild-type activity. As we already discussed in the previous paragraph, our model failed to identify for this mutant any perturbation of protein structure and stability at the level of the single molecule, although we hypothesize the formation of an interchain disulfide bridge as a consequence of this mutation.

The interpretation of another biochemically characterized mutation is more difficult from a structural point of view. Arg 333 is mapped at the interface between two GALT subunits both in the X-ray structure of *E.coli* enzyme (Wedekind *et al*., 1995, 1996; Thoden *et al*., 1997) and in our models. Therefore, a mutation involving this residue would be expected to affect dimer interface, though our analyses show that Arg 333 does not interact with any other residue to form e.g. intersubunit H-bonds and/or salt bridges. Mutant R333W is reported to have null activity in homodimeric state, but about 50% activity in heterodimeric wild-type state (Elsevier *et al*., 1996; Elsevier and Fridovich-Keil, 1996; Riehman *et al*., 2001; Christacos and Fridovich-Keil, 2002), suggesting that the replacement of this residue does not affect intersubunit relationships. Arg 333 is included in a potential phosphorylation site as predicted by ScanProsite analysis (De Castro *et al*., 2006), and, although we did not find reports in the literature about the presence of posttranslational modifications in GALT, it could be possible that the replacement of this residue with another one is the cause of inactivation of a single subunit due to the lack of this posttranslational modification in humans.

Finally, we were not able to predict any structural effect for a mutation observed especially in African and Afro-American people (Tyfield *et al*., 1999): S135L. However, also the experimental characterization of this mutant produced different, and in some cases controversial, results. For example, this mutation was firstly classified as a simple polymorphism when the mutant was overexpressed in COS cells (Reichardt *et al*., 1992). On the contrary, the mutant protein expressed in yeast showed not more than 3% activity with respect to wild type (Riehman *et al*., 2001). Patients homozygotes for this mutation show no GALT activity in their erythrocytes or lymphoblasts, but a residual GALT activity is present in their leukocytes, and they show near normal total body (^{13}C)galactose oxidation to $^{13}CO_2$ in breath (Lai and Elsas, 2001). The kinetic parameters of the isolated enzyme show a markedly decreased V_{max}, but an increased K_M for the two substrates (Lai and Elsas, 2001). Riehman *et al*. mapped this mutation onto the structure of *E.coli* enzyme (Wedekind *et al*., 1996), and stated that His 115 (the residue equivalent to Ser 135 in this bacterial enzyme) is 'near the active state' (Riehman *et al*., 2001). Yet, the distance between His 115 and the nearest atom of the ligand is more than 11 Å. Similarly, in our model of human GALT, Ser 135 is ~11 Å far from the active site, so the interaction of this residue with the substrate seems unlikely. Several hypotheses on the effects of this mutation were made (impairment of the binding of an ion necessary for catalysis, participation in catalysis, destabilization of the enzyme) but none has been proved (Lai and Elsas, 2001). Probably, the fact that also our predictions failed to identify a direct effect of this mutation on GALT structure and

Downloaded from http://peds.oxfordjournals.org/ at CharitÃ© -Med. Bibliothek on September 3, 2013

stability is a further confirmation that this mutation might have a very complex effect, perhaps involving other elements that interact with this protein. This mutation should be studied with different approaches than those used up to now, including proteomics approaches able to highlight global phenomena in the cells.

Conclusions

Despite decades of study, the mechanism of pathophysiology in galactosemia is still unclear, and, although early diagnosis and dietary intervention has caused a clear improvement of the quality of life of galactosemic patients, there is still a long and winding road to reach a complete success. The analysis of structural features of GALT enzyme, presented in this paper, has been made to predict the effect of mutations at protein level, with methodologies that allow to obtain results quickly with respect to experimental procedures. This allows to fill a gap in the knowledge of this enzyme, since the three-dimensional structures of GALT and of its mutants are not yet available. This approach should not be considered as the 'magic wand' able to solve all the problems related to the diagnosis and cure of classical galactosemia, at least in the (near) future. However, these information, especially if combined with other obtained with different approaches, such as multidisciplinary and 'omics' studies, could be useful to find a correlation between the kind of mutation and the expected evolution of the disease during the years, or the severity of the symptoms. Moreover, the results from this study could be used as a guide for planning future experiments to clarify or confirm deductions obtained from this computational analysis.

Other considerations can be made from our results. For an enzyme, it is usually expected that the mutations with higher effect on the activity should involve directly the substrate binding. This common hypothesis is only partially confirmed from the comparison of our analysis and the experimental data on biological activity available for 15 mutants (see Table I). Only two totally inactive mutants are affected at the level of substrate binding, in both cases together with intersubunit interactions and monomer stability, while other three mutations, which totally inactivate the enzyme, modify only intersubunit interactions or monomer stability. For other five mutants, having activity in the range 0–15%, intersubunit interactions or monomer stability is affected by the mutations, as for other three mutants with activity in the range 45–63%. As a summary of data reported in Supplementary data available at *PEDS* online, Table S1 and concerning the whole set of 107 mutations, we note that only 12 mutations affect residues directly involved in the substrate binding, 47 affect intersubunit interactions and 91 affect monomer stability. It could be concluded that alteration of protein stability, at the level of monomer and/or dimer, is a major effect observed, which is able to modulate the GALT activity.

Possible further investigations could be made on mutants, with simulations that could allow to study directly the impact of the mutation on protein dynamics and folding. This could clarify the effects of those mutations that apparently do not perturb the protein structure and activity from a static point of view, or to find novel or unseen effects for those with a predicted impact on the enzyme.

This approach opens also different future perspectives; first of all, the possibility of characterizing individual GALT structural features. The vast majority of galactosemic patients are compound heterozygous, i.e. they show different combinations of mutations in their GALT gene. Virtually, we can recreate each possible combination of mutations found in galactosemic patients, and therefore, we could provide a personalized 'molecular portrait' of GALT enzyme for each patient, in a much shorter time than experimental studies. Finally, a further application of our work can be considered by viewing it as an example of methodology for approaching the study of pathologies in which mutations affect the structure and function of key proteins. Many rare diseases are due to point mutations, so this computational methodology can be applied to increase the knowledge about molecular bases of these pathologies.

Acknowledgements

The authors wish to acknowledge Dr Andrew C. R. Martin for fruitful discussions during the first planning of this work, and Dr Antonio d'Acierno for continuous database updating.

Supplementary data

Supplementary data are available at *PEDS* online.

Funding

This work has been developed in the frame of the CNR-Bioinformatics project and was partially supported by RNBIO Network and by 'Programma Italia-USA Farmacogenomica Oncologica' (grant number 527/A/3A/5).

References

Ashino,J., Okano,Y., Suyama,I., Yamazaki,T., Yoshino,M., Furuyama,J., Lin,H.C., Reichardt,J.K. and Isshiki,G. (1995) *Hum. Mutat.*, **6**, 36–43.

Beaudet,A.L., Scriver,C.R., Valle,D. and Sly,W.S. (2006) In Valle,D., Beaudet,A.L., Vogelstein,B., Kinzler,K.W., Antonarakis,S.E. and Ballabio,A. (eds), *The Online Molecular and Metabolic Basis of Inherited Disease—OMMBID.* part 1, chapter 1. John Wiley & Sons, New York.

Berman,H., Henrick,K., Nakamura,H. and Markley,J.L. (2007) *Nucleic Acids Res.*, **35**, D301–D303.

Brenner,C. (2002) *Biochemistry*, **41**, 9003–9014.

Calderon,F.R., Phansalkar,A.R., Crockett,D.K., Miller,M. and Mao,R. (2007) *Hum. Mutat.*, **28**, 939–943.

Chakravarty,S., Wang,L. and Sanchez,R. (2005) *Nucleic Acids Res.*, **33**, 244–259.

Christacos,N.C. and Fridovich-Keil,J.L. (2002) *Mol. Genet. Metab.*, **76**, 319–326.

Crews,C., Wilkinson,K.D., Wells,L., Perkins,C. and Fridovich-Keil,J.L. (2000) *J. Biol. Chem.*, **275**, 22847–22853.

d'Acierno,A., Facchiano,A. and Marabotti,A. (2009) *Genomics Proteomics Bioinformatics*, **7**, 71–76.

Dantzer,J., Moad,C., Heiland,R. and Mooney,S. (2005) *Nucleic Acids Res.*, **33**, W311–W314.

De Castro,E., Sigrist,C.J.A., Gattiker,A., Bulliard,V., Langendijk-Genevaux,P.S., Gasteiger,E., Bairoch,A. and Hulo,N. (2006) *Nucleic Acids Res.*, **34**, W362–W365.

Elsas,L.J., 2nd and Lai,K. (1998) *Genet. Med.*, **1**, 40–48.

Elsevier,J.P. and Fridovich-Keil,J.L. (1996) *J. Biol. Chem.*, **271**, 32002–32007.

Elsevier,J.P., Wells,L., Quimby,B.B. and Fridovich-Keil,J.L. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 7166–7171.

Feyfant,E., Sali,A. and Fiser,A. (2007) *Protein Sci.*, **16**, 2030–2041.

Frey,P.A. (1996) *FASEB J.*, **10**, 461–470.

Fridovich-Keil,J.L. and Walter,J.H. (2006) In Valle,D., Beaudet,A.L., Vogelstein,B., Kinzler,K.W., Antonarakis,S.E. and Ballabio,A. (eds), *The*

Downloaded from http://peds.oxfordjournals.org/ at CharitÃ© -Med. Bibliothek on September 3, 2013

Online Molecular and Metabolic Basis of Inherited Disease—OMMBID. part 7, chapter 72. John Wiley & Sons, New York.
Fridovich-Keil,J.L., Langley,S.D., Mazur,L.A., Lennon,J.C., Dembure,P.P. and Elsas,L.J., 2nd (1995) *Am. J. Hum. Genet.*, **56**, 640–646.
Garza-Garcia,A., Patel,D.S., Gems,D. and Driscoll,P.C. (2007) *Hum. Mutat.*, **28**, 660–668.
Geeganage,S. and Frey,P.A. (1998) *Biochemistry*, **37**, 14500–14507.
Gilis,D. and Rooman,M. (2000) *Protein Eng.*, **13**, 849–856.
Henderson,J.M., Wells,L. and Fridovich-Keil,J.L. (2000) *J. Biol. Chem.*, **275**, 30088–30091.
Hubbard,S.J., Campbell,S.F. and Thornton,J.M. (1991) *J. Mol. Biol.*, **220**, 507–530.
Kabsch,W. and Sander,C. (1983) *Biopolymers*, **22**, 2577–2637.
Kumar,S. and Nussinov,R. (1999) *J. Mol. Biol.*, **293**, 1241–1255.
Lai,K. and Elsas,L.J. (2001) *Mol. Genet. Metab.*, **74**, 264–272.
Lai,K., Langley,S.D., Dembure,P.P., Hjelm,L.N. and Elsas,L.J., 2nd (1998) *Hum. Mutat.*, **11**, 28–38.
Lai,K., Willis,A.C. and Elsas,L.J. (1999) *J. Biol. Chem.*, **274**, 6559–6566.
Marabotti,A. (2008) *Curr. Chem. Biol.*, **2**, 200–214.
Marabotti,A. and Facchiano,A.M. (2005) *J. Med. Chem.*, **48**, 773–779.
Martin,A.C., Facchiano,A.M., Cuff,A.L., Hernandez-Boussard,T., Olivier,M., Hainaut,P. and Thornton,J.M. (2002) *Hum. Mutat.*, **19**, 149–164.
McDonald,I.K. and Thornton,J.M. (1994) *J. Mol. Biol.*, **238**, 777–793.
Moult,J., Fidelis,K., Kryshtafovych,A., Rost,B. and Tramontano,A. (2009) *Proteins*, **77**(Suppl. 9), 1–4.
Ng,P.C. and Henikoff,S. (2001) *Genome Res.*, **11**, 863–874.
Reichardt,J.K. and Woo,S.L. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 2633–2637.
Reichardt,J.K., Levy,H.L. and Woo,S.L. (1992) *Biochemistry*, **31**, 5430–5433.
Riehman,K., Crews,C. and Fridovich-Keil,J.L. (2001) *J. Biol. Chem.*, **276**, 10634–10640.
Rost,B. and Sander,C. (1994) *Proteins*, **20**, 216–226.
Rovida,E., Merati,G., D'Ursi,P., Zanardelli,S., Marino,F., Fontana,G., Castaman,G. and Faioni,E.M. (2007) *Hum. Mutat.*, **28**, 345–355.
Sali,A. and Blundell,T.L. (1993) *J. Mol. Biol.*, **234**, 779–815.
Segal,S. (1998) *J. Inherit. Metab. Dis.*, **21**, 455–471.
Thoden,J.B., Ruzicka,F.J., Frey,P.A., Rayment,I. and Holden,H.M. (1997) *Biochemistry*, **36**, 1212–1222.
Tyfield,L. (2000) *Eur. J. Pediatr.*, **159**(Suppl. 3), S204–S207.
Tyfield,L., *et al*. (1999) *Hum. Mutat.*, **13**, 417–430.
Valdar,W.S. (2002) *Proteins*, **48**, 227–241.
Valdar,W.S.J. and Thornton,J.M. (2001) *Proteins*, **42**, 108–124.
Wedekind,J.E., Frey,P.A. and Rayment,I. (1995) *Biochemistry*, **34**, 11049–11061.
Wedekind,J.E., Frey,P.A. and Rayment,I. (1996) *Biochemistry*, **35**, 11560–11569.
Wishart,D.S. (2006) In Valle,D.D., Beaudet,A.L., Vogelstein,B., Kinzler,K.W., Antonarakis,S.E. and Ballabio,A. (eds), *The Online Molecular and Metabolic Basis of Inherited Disease—OMMBID*. part 2, chapter 3.1. John Wiley & Sons, New York.
Zhou,H. and Zhou,Y. (2002) *Protein Sci.*, **11**, 2714–2726.

Downloaded from http://peds.oxfordjournals.org/ at CharitÃ© -Med. Bibliothek on September 3, 2013