J Mol Model (2011) 17:2639–2649
DOI 10.1007/s00894-011-0958-y

ORIGINAL PAPER

The effect of a Pro^{28}Thr point mutation on the local structure and stability of human galactokinase enzyme—a theoretical study

Balázs Jójárt · Milán Szőri · Róbert Izsák · István Marsi · Aranka László · Imre G. Csizmadia · Béla Viskolcz

Received: 14 September 2010 / Accepted: 4 January 2011 / Published online: 25 January 2011

Abstract Galactokinase is responsible for the phosphorylation of α-D-galactose, which is an important step in the metabolism of the latter. Malfunctioning of galactokinase due to a single point mutation causes cataracts and, in serious cases, blindness. This paper reports a study of the Pro^{28}Thr point mutation using a variety of theories including molecular dynamics (MD), MM-PBSA/GBSA calculations and AIM analysis. Altered H-bonding networks were detected based on geometric and electron density criteria that resulted in local unfolding of the β-sheet secondary structure. Another consequence was the decrease in stability (5–7 kcal mol^{-1}) around this region, as confirmed by ΔG_{bind} calculations for the extracted part of the whole system. Local unfolding was verified by several other MD simulations performed with different duration, initial velocities and force field. Based on the results, we propose a possible mechanism for the unfolding caused by the Pro^{28}Thr point mutation.

Keywords Human galactokinase · Molecular dynamics · AIM · MMPBSA/GBSA · Local unfolding · β-sheet stability

Abbreviations

MD	Molecular dynamics
WT	Wild-type form of galactokinase
MT	Point-mutated form of galactokinase
GALK	Galactokinase
WTA	Chain A of wild-type galactokinase
WTB	Chain B of wild-type galactokinase
MTA	Chain A of point-mutated galactokinase
MTB	Chain B of point-mutated galactokinase
AIM	Atoms in molecules theory
a.u.	Atomic unit
bcp	Bond critical points
PM	Point mutation
ΔG_{bind}	Binding free energy
MTA2, MTA3, MTA4, MTA5	Shorter (6 ns long) MD simulation of the MTA using different initial velocities
n.p.	Not possible
$\Delta\Delta G_{bind}$	Relative binding free energy

Introduction

The prediction of the structure and stability of proteins is a fundamental goal in the molecular sciences (molecular biology, protein engineering, computer-aided drug design etc.). A fairly tractable solution to this problem is to predict

Electronic supplementary material The online version of this article (doi:10.1007/s00894-011-0958-y) contains supplementary material, which is available to authorized users.

B. Jójárt (✉) · M. Szőri · R. Izsák · I. Marsi · I. G. Csizmadia · B. Viskolcz
Department of Chemical Informatics, University of Szeged,
Boldogasszony sgt. 6,
Szeged 6725, Hungary
e-mail: jojartb@jgypk.u-szeged.hu

A. László
Department of Pediatrics, Albert Szent-Györgyi Medical Center,
University of Szeged,
Korányi fasor 14-15,
Szeged 6725, Hungary

I. G. Csizmadia
Department of Chemistry, University of Toronto,
80 St. George Street,
Toronto, ON M5S 3H6, Canada

changes in structure and stability induced by point mutations [1]. An approach to achieve this aim is to use statistical potentials derived from geometric and environmental propensities and correlations of residues in X-ray crystal structures. These kinds of functions are not that accurate in predicting protein stability changes, but they may be used to suggest stabilizing mutations in the absence of detailed structural information [2]. In a more holistic structural description, molecular dynamics (MD) calculations using all atom force fields can also address this issue [3]. For instance, free energy simulations have been used to provide accurate predictions of properties of point mutants involving relative stabilities and structural properties [4–6]. To understand the effect of local structural changes in a protein caused by a point mutation, quantum chemical calculations on tripeptides have also proved a successful approach [7]. The effect of a point mutation can be amplified via such local structural changes, causing the protein to lose its enzyme functions partially or totally. This can give rise to serious problems in the metabolism of important biomolecules. Here, we present the results of a study on the effect of a point mutation on α-D-galactose metabolism.

The evolutionarily conserved pathway for α-D-galactose metabolism involves three enzymes: galactose-1-phosphate uridyl transferase, galactokinase (GALK), and uridine diphosphate galactose-4-epimerase [8]. Among these enzymes, GALK catalyzes the conversion of α-D-galactose to galactose-1-phosphate. This enzyme has attracted significant attention among researchers because of its important metabolic role, and the fact that functional defects in the human enzyme can result in galactosemia [9]. The decreased activity or absence of fully functional GALK results in the buildup of unmetabolized galactose in cells and in blood plasma. This is a particular problem in the lens cells of the eye where, in humans, the enzyme aldose reductase converts galactose to its corresponding sugar alcohol, galactitol (dulcitol). Although galactose can be transported across the cell membrane, galactitol cannot, and consequently it accumulates in lens cells. This leads to the osmotic uptake of water, swelling of cells, and eventual apoptosis and lysis [10]. Furthermore, this genetic disease is responsible for an increased risk of early onset cataracts and, in more severe cases, results in kidney, liver and/or brain damage [11].

A number of different mutations in the GALK gene, including both deletions and point mutations (PMs), that give rise to galactosemia have been described. Mutations can be divided into two groups based on their biochemical consequences (for a detailed description see [12]): (1) mutations causing proteins to become insoluble ($Pro^{28}Thr$, $Val^{32}Met$, $Gly^{36}Arg$), and (2) mutations resulting in altered kinetic parameters ($His^{44}Tyr$, $Arg^{68}Cys$, $Gly^{346}Ser$) [13]. Although Pro has only a medium level mutability when compared to other essential amino acids [14], the effect of the $Pro^{28}Thr$ point mutation can be dramatic as seen in the case of galactokinase. The $Pro^{28}Thr$ PM was detected particularly in several Romani families many years ago [15, 16] and the phenotype of this mutation is manifested in decreased or undetectable enzyme activity and the protein becoming insoluble in *Escherichia coli*. Kalaydjieva et al. [16] performed secondary structure–prediction analysis around this region of the enzyme. They concluded that this mutation causes loop shifting and a β-turn alteration. Following determination of the 3D structure by X-ray crystallography, Thoden and coworkers explained that the mutation of Pro^{28} to Thr^{28} will result in improper folding around this region, rather than any significant conformational changes [12]. Their reasoning was based on the fact that an –OH group is introduced into a hydrophobic pocket.

However, no quantitative description of the mechanism and structural rearrangement around this PM has yet been established. Therefore, in the present study we have made an attempt to quantify the effect of the $Pro^{28}Thr$ PM on the local structure and stability alteration of human GALK.

Methods

The effect of mutation on the stability of GALK

In silico characterization of the stabilization/destabilization effects of GALK PMs was studied via the l-Mutant2.0 webserver [17]. The method is a support vector machine-based technique that is able to predict changes in stability due to PM in a protein structure or sequence. The correlation coefficient between the experimental and calculated $\Delta\Delta G$ is predicted to be 0.71 with a standard error of 1.30 kcal mol^{-1}. The $\Delta\Delta G$ values were predicted for the following mutations at T=310 K and pH=7.4 for both A and B chains: (1) mutations resulting insoluble protein in *E. coli*: $Pro^{28}Thr$; $Val^{32}Met$; $Gly^{36}Arg$; $Thr^{228}Met$; $Ala^{384}Pro$; (2) mutations causing altered kinetic parameters: $His^{44}Tyr$, $Arg^{68}Cys$, $Arg^{239}Gln$, $Gly^{346}Ser$ and $Gly^{349}Ser$; and (3) mutation where the kinetic properties are the same as in WT: $Ala^{198}Val$.

Molecular dynamics simulation

The crystal structure of the human GALK enzyme (PDB ID: 1WUU [12]) was downloaded from the PDB database [18, 19] (Fig. 1).

After removing the substrates and water molecules, missing residues and side chains were added to the structure using the Sequence Editor module of the Molecular Operating Environment (version 2007.09) program package

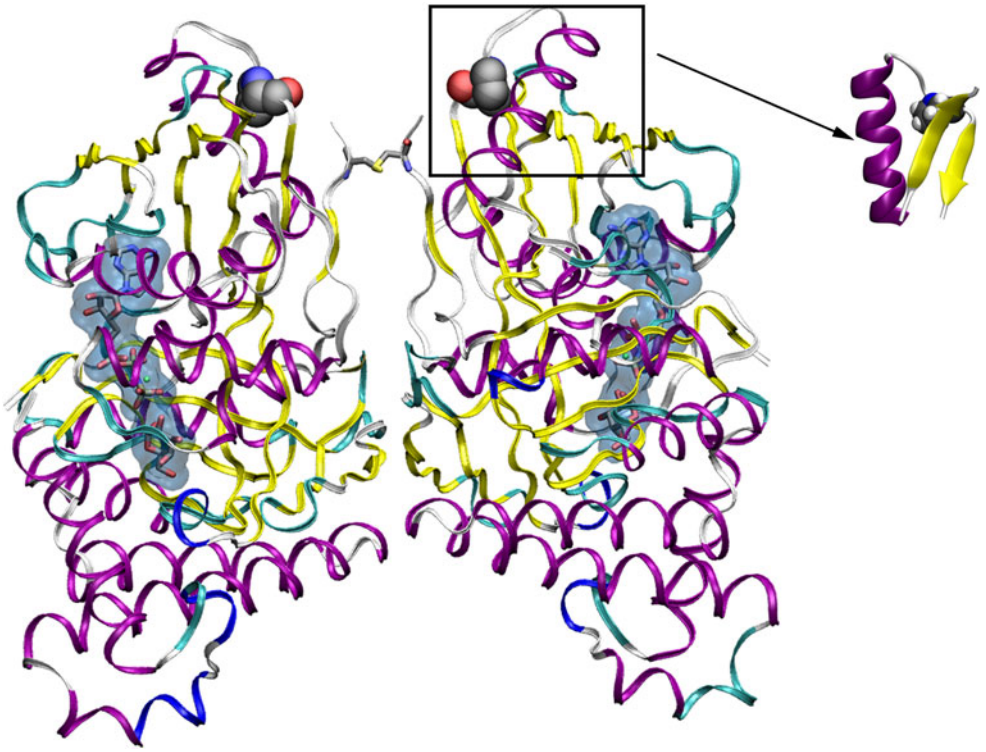


Fig. 1 Three-dimensional (3D) structure of the human galactokinase (GALK) enzyme (PDB ID: 1WUU), showing local secondary structure elements around the $Pro^{28}Thr$ point mutation (PM)

[20], while the Pro^{28} amino acid was modified manually to Thr^{28} and selenomethionines were treated as methionines. PDB2PQR software was employed in the optimization of possible H-bond interactions [21]. (Further discussion on the crystallographic water molecules may be found in the Electronic supplementary material.)

Initial structures for the MD simulation were prepared using the Desmond GUI in Maestro. Structures were solvated with TIP3P water molecules [22] (18,421), 10 Na^+ ions were added as counter ions, and 51 Na^+/Cl^- ions were also added to the system in order to ensure a 0.15 M salt concentration. The protein was modeled using CHARMM27 [23] parameters with CMAP correction terms [24], and the ions were described with the parameters of Beglov et al. [25].

All MD simulations were performed with the Desmond v20108 program package [26] applying the following protocols for the wild-type (WT) and mutated (MT) GALK: (1) minimization with the steepest-descent and L-BFGS methods was performed (maximum step number: 2,000, gradient tolerance: 1.0 kcal mol^{-1} $Å^{-1}$, 10 kcal mol^{-1} $Å^{-2}$ positional restraint was applied to the solute); (2) minimization, as described previously, without restraints; (3) 10 kcal mol^{-1} $Å^{-2}$ positional restraint was again applied to the solute, which was removed in the first 240 ps of our simulations, where the temperature was increased from 250 K to 310 K linearly using the DESMOND annealing plug-in; (3) free MD simulation was performed at 310 K. The calculations were performed in the NPT ensemble (T= 310 K, p=1 atm) with isotropic pressure scaling using the Berendsen thermo- and barostat methods (τ_p=1.0 ps, τ_T= 0.1 ps) [27]. The cut-off value of the near non-bonded interactions was set to 10 Å, and the far electrostatic interactions were calculated by the particle mesh Ewald (PME) method [28] with 100×84×84 Fourier mesh points. The time step was set to 2 fs, bonded and short non-bonded interactions were calculated in every step, while long range electrostatics was calculated at every third step.

The calculations were 24 ns long at the first step, thereafter the calculation was continued for another 48 ns. Additionally, five independent trajectories with different initial velocities were generated, each of 6 ns duration [29].

The force field dependence was also tested by applying the OPLSAA-2005 force field parameter set [30] for the protein and ions, in which case the duration of the simulations was 12 ns long.

Atoms in molecules calculations

For the 24 ns long WT and MT simulations in the last 21.504 ns, average structures were calculated for the region of interest: Xxx^{28}-Glu^{29}-Leu^{30}-Ala^{31}-Val^{32} and Leu^{63}-Val^{64}-Gly^{65} (Xxx≡Pro or Thr for WT and MT, respectively (Fig. 2).

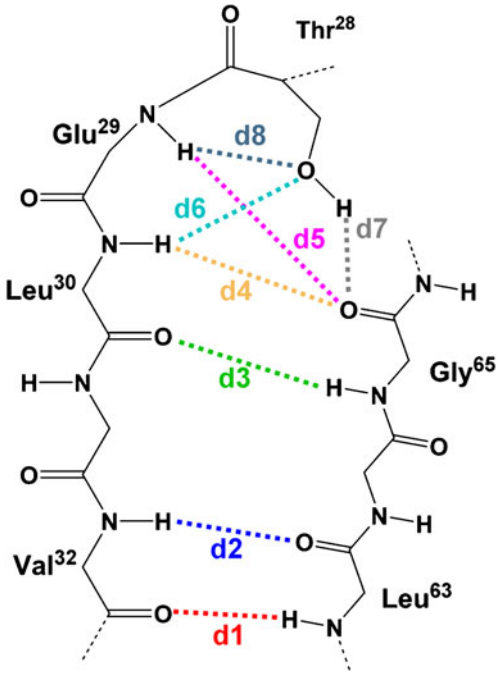


Fig. 2 Possible stabilizing H-bonds near the Pro[28]Thr mutation. The d1, ..., d5 interactions exist only in the wild type (WT), while the d6, d7 and d8 H-bonds are possible only in the mutant (MT) form

Using these structures, the electron density was calculated at the B3LYP/6-31 G(d) level of theory and then used to generate the electron density topography of the model system according to Bader's atoms in molecules (AIM) theory [31, 32]. AIM is the mathematical analysis of molecular electron density, where nuclei act as point attractors immersed in a negative charge cloud—the electron density $\rho(r)$. The existence of a bond is indicated by the presence of a bond critical point (bcp) and a corresponding bond path. The bcp is a 2nd order saddle point of the electron density between two nuclei, where the Hessian of the electron density has one positive eigenvalue along the bond axis (minimum) and two negative eigenvalues perpendicular to the bond axis (maximum). A bond is a path of maximum electron density from one nucleus to another. To avoid false interpretation arising from the numerical inaccuracy in the AIM analysis, only those bcps where the electron density is at least 0.001 a.u. are discussed. As the formation of any chemical bond/interaction results in corresponding characteristic changes in the spatial distribution of electrons, the identification of critical points in the electron density provides a general tool for the description of these phenomena. Here, we focused on the hydrogen bond network around the location of the point mutation. Although AIM analysis does not consider explicitly internuclear repulsion, which is another significant factor influencing the stability of chemical species, this is of little importance with hydrogen bonds due to the relatively large internuclear separation. Therefore, the numerical results of such analysis can be used in a comparative manner to obtain quantitative and qualitative information on the strength of the hydrogen bonds discussed, i.e., increased electron densities at the bcps represent stronger H-bonds.

MM-PBSA/GBSA calculations

The stability of the two strands of the β-sheet near the PM (Fig. 1) was characterized by calculating the binding free energy (ΔG_{bind}) between them using the MM-PBSA/GBSA method [33, 34] (in the following the standard nomenclature of receptor–ligand interactions is used). The (Pro/Thr)[28]-Glu[29]-Leu[30]-Ala[31]-Val[32] [denoted as R (receptor) in the following equations and in Fig. 3] and Leu[63]-Val[64]-Gly[65] [denoted as L (ligand) in the following equations and in Fig. 3] part of the enzyme structure were extracted from the trajectory and the calculations were performed with the mmpbsa module of Amber9 [35].

The ΔG_{bind} was calculated by applying the thermodynamic cycle presented in Fig. 3 and using the following equation:

$$\Delta G_{bind,s} = \Delta G_{bind,g} + \Delta G_{solv,C} - \left(\Delta G_{solv,L} + \Delta G_{solv,R}\right),$$

where

$\Delta G_{bind,s}$	binding free energy in solvent
$\Delta G_{bind,g}$	binding free energy in gas phase
$\Delta G_{solv,R}$	solvation free energy of the receptor
$\Delta G_{solv,L}$	solvation free energy of the ligand
$\Delta G_{solv,C}$	solvation free energy of the complex.

The solvation free energy was calculated as the sum of the electrostatic (ΔG_{polar}) and non-polar contributions ($\Delta G_{non\text{-}polar}$). The ΔG_{polar} term was calculated with the linearized Poisson Boltzmann (MM-PBSA, with the following parameters, ε_{int}=1, ε_{ext}=80, with a grid spacing of 0.5 Å) and generalized Born (MM-GBSA) equations. $\Delta G_{non\text{-}polar}$ was approximated with the solvent accessible surface area (SASA) using the following equation: $\Delta G_{non\text{-}}$

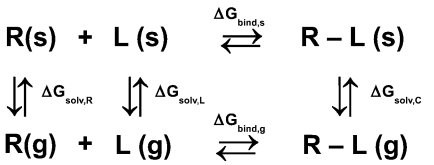


Fig. 3 Thermodynamic cycle used in the MM-PBSA/GBSA calculation. *R* corresponds to (Pro/Thr)[28]-Glu[29]-Leu[30]-Ala[31]-Val[32] and *L* to Leu[63]-Val[64]-Gly[65]. $\Delta G_{bind,s}$ Binding free energy in solvent, $\Delta G_{bind,g}$ binding free energy in gas phase, $\Delta G_{solv,R}$ solvation free energy of the receptor, $\Delta G_{solv,L}$ solvation free energy of the ligand, $\Delta G_{solv,C}$ solvation free energy of the complex. $\Delta G_{bind,s} = \Delta G_{bind,g} + \Delta G_{solv,C} - \left(\Delta G_{solv,L} + \Delta G_{solv,R}\right)$

$_{polar}$=λ×SASA (λ=0.0072 kcalÅ^{-2}). $\Delta G_{bind,g}$ was calculated as $\Delta G_{bind,g}=\Delta E_{MM} - T\Delta S$, where ΔE_{MM} is the average interaction energy between the receptor and the ligand. The entropy contribution was also calculated using the quasi-harmonic approximations [33, 36].

The calculations applied here are not classical ΔG_{bind} calculations in the sense that R [(Pro/Thr)^{28}-Glu^{29}-Leu^{30}-Ala^{31}-Val^{32})] and L (Leu^{63}-Val^{64}-Gly^{65}) are in the same part of the protein. Nevertheless, the Gibbs free energy calculations are straightforward for these segments of the structure, and the ΔG_{bind} obtained gives information about the local stability alterations due to the point mutation.

Analysis of MD trajectories

The 3D structure preparation and analysis were performed with Visual Molecular Dynamics software [37]. The secondary structure elements were calculated using the dssp algorithm [38] as implemented in the Simulaid program package [39].

Table 1 Computed protein stability changes in human galactokinase (GALK) enzyme using the l-Mutant2.0 method (pH=7.4; T=310 K). ΔΔG values are in kcal mol^{-1}, negative values of protein stability changes correspond to destabilizing mutations

Mutation	$\Delta\Delta G_{chainA}$	$\Delta\Delta G_{chainB}$
Pro^{28}Thr[a]	−0.99	−0.97
Val^{32}Met[a]	−3.05	−2.91
Gly^{36}Arg[a]	−1.46	−1.39
Thr^{288}Met[a]	−0.22	−0.14
Ala^{384}Pro[a]	−1.36	−1.32
His^{44}Tyr[b]	−0.79	−0.82
Arg^{68}Cys[b]	−1.21	−1.33
Arg^{239}Gln[b]	−1.62	−1.86
Gly^{346}Ser[b]	0.21	0.07
Gly^{349}Ser[b]	0.13	0.09
Ala^{198}Val[c]	0.61	0.69

[a] Mutations resulting in insoluble protein expressed in *Escherichia coli*

[b] Mutations altering the kinetic parameters of the enzyme

[c] Changes neither the parameters nor the solubility of GALK

Results and discussion

Description of enzyme structure, and the environment of the point mutation

The GALK enzyme is a homodimer; chains A and B are connected by a disulfide bridge formed between the neighboring Cys^{391} residues. The 28th position, where the PM takes place, is located far from the binding site (the shortest distance is about 15–16 Å) in a random coil structure between an α-helix and one strand of an anti-parallel β-sheet (Fig. 1).

In silico stability investigation of the point mutated form of GALK

Using the l-Mutant-2.0 method, stability changes in GALK due to PM were investigated (Table 1). Considering the standard error of this technique (1.30 kcal mol^{-1}) the data obtained do not differ significantly.

The most positive values (0.61 and 0.69 kcal mol^{-1}) were obtained for Ala^{198}Val, which resulted in similar kinetic parameters compared to WT.

The most negative values (−3.05 and −2.91 kcal mol^{-1}) were detected for Val^{32}Met. Val^{32} is located near the covalently linked Cys^{391} residue at the C-terminus. We speculate that the PM causes significant rearrangement of the structure around the linkage, which results in an insoluble protein.

The Pro^{28}Thr PM, which was studied in more detail in our study, has a ΔΔG value of about −0.90 kcal mol^{-1}. This mutated enzyme becomes insoluble after purification despite the fact that detectable enzyme activity was measured in patients with this mutation. These two facts are inconsistent, because an insoluble protein cannot function as an enzyme. Nevertheless, the ΔΔG value calculated does not show a high destabilizing effect, and a PM with similar values (His^{44}Tyr) is soluble. Therefore, we hypothesized that, during protheosyntesis, the folded 3D structure evolves, but due to the PM at position 28, a fast rearrangement takes place that results in an insoluble protein.

Local structural alteration of the enzyme

The secondary structure elements of the α-helical part of the protein were stable during the whole simulation (see Electronic supplementary material). The anti-parallel β-sheet-forming residues near the PM are: Val^{32}×××Leu^{63}, Leu^{30}×××Gly^{65}, with one additional H-bond formed between Glu^{29}×××Gly^{65}. After the mutation, some possible additional H-bonding sites appear (d5–d8), due to the incorporation of an –OH group in the structure (Fig. 2).

These possible H-bonds were characterized by the average distance (with standard deviation) between the appropriate heavy atoms as well as corresponding electron density values for average structures (Table 2). The distributions of the distances are also depicted in Fig. 4. The first 520 frames (2.496 ns) were considered as equilibration, and the distances were calculated from the remaining part of the simulation.

Far from the mutation (d1 and d2), there is no significant difference between the values obtained, all of which

Table 2 Measured geometric parameters (d1, ..., d8; see Fig. 2) and calculated electron densities between the possible H-bonding atoms near the mutation. *WTA* chain A of the wild type enzyme form; *WTB* chain B of the wild type enzyme form, *MTA* chain A of the mutated type enzyme form, *MTB* chain B of the mutated type enzyme form

	Geometric parameters (Å)				Electron density (a.u.)			
	WTA	WTB	MTA	MTB	WTA	WTB	MTA	MTB
d1	2.99±0.16	2.95±0.15	2.97±0.15	2.97±0.17	0.025	0.021	0.028	0.016
d2	2.94±0.13	3.00±0.15	2.95±0.14	2.99±0.15	0.035	0.026	0.039	0.023
d3	2.85±0.12	2.88±0.13	3.31±0.25	2.91±0.14	0.053	0.031	0.036	0.032
d4	2.95±0.15	2.93±0.15	4.50±0.24	2.98±0.16	0.051	0.044	0.000	0.041
d5	3.06±0.22	2.97±0.19	5.48±0.21	3.04±0.22	0.050	0.054	0.000	0.050
d6	n. p.[a]	n. p.	3.22±0.21	5.94±0.51	n. p.	n. p.	0.022	0.000
d7	n. p.	n. p.	2.75±0.15	5.08±0.55	n. p.	n. p.	0.041	0.000
d8	n. p.	n. p.	2.98±0.12	4.45±0.34	n. p.	n. p.	0.024	0.000

[a] Not possible

indicate strong H-bonds. For d3, the value obtained is somewhat high in MTA, but it remains an acceptable H-bond. d4 and d5 are the most interesting and important. The values obtained for MTA are significantly higher (4.50± 0.24, 5.48±0.21 Å) than the other, WTA, WTB and MTB cases. The original strong H-bond interactions were disrupted in the former case, and remain stable in the other cases during the time investigated. In line with the vanishing of these interactions, new H-bonding interactions are formed (d6, d7 and d8) between the side-chain of Thr^{28} and Leu^{30} or Gly^{65} in the MTA case that were not detected for MTB.

The asymmetric behavior of H-bond networks between MTA and MTB can be caused by the structural difference in the local environment of the PM, although the enzyme is a homodimer.

As opposed to geometric criteria, AIM analysis enables us to determine the existence of hydrogen bonds in a more explicit manner. The possible H-bonding interactions are depicted with green spheres at their bcps in Fig. 5.

In these cases, the electron densities are between 0.016 and 0.054 a.u. (0.035±0.012 a.u., one order of magnitude lower compared to that of the polar –N–H bond, where this value is 0.24 a.u.). It is important to note that whenever the possibility of an H-bond is ruled out by the large distance of the participating atoms, electron density values also become zero. There is a greater variation in density in cases where H-bonding does occur (mostly within a distance of about 3 Å), but they all still indicate significant interactions The blue spheres correspond to the –C=O×××H-–C– and –C–H×××H-–C– interactions, where the electron densities varied from 0.010 to 0.030 a.u. (0.025±0.009 a.u.). Besides the alteration of H-bond distances and the variation of electron densities, another consequence of the $Pro^{28}Thr$ mutation is clearly depicted in Fig. 5. Due to the insertion of the side chain of Thr^{28} between the two strands of the anti-parallel β-sheet in the case of the MTA, the parallel character of this secondary structure element vanishes.

Based on our previous discussion of the investigated β-sheet, we obtained altered H-bonding networks. Therefore it is very important to discover whether there is any alteration in stability as well behind the structural shift.

The results of MM-PBSA/GBSA calculations are depicted in Fig. 6.

The calculated relative ΔG_{bind} ($\Delta\Delta G_{bind}$) values [$\Delta\Delta G_{bind}=\Delta G_{bind}(WTA) - \Delta G_{bind}(MTA)$ etc.] are presented in Table 3.

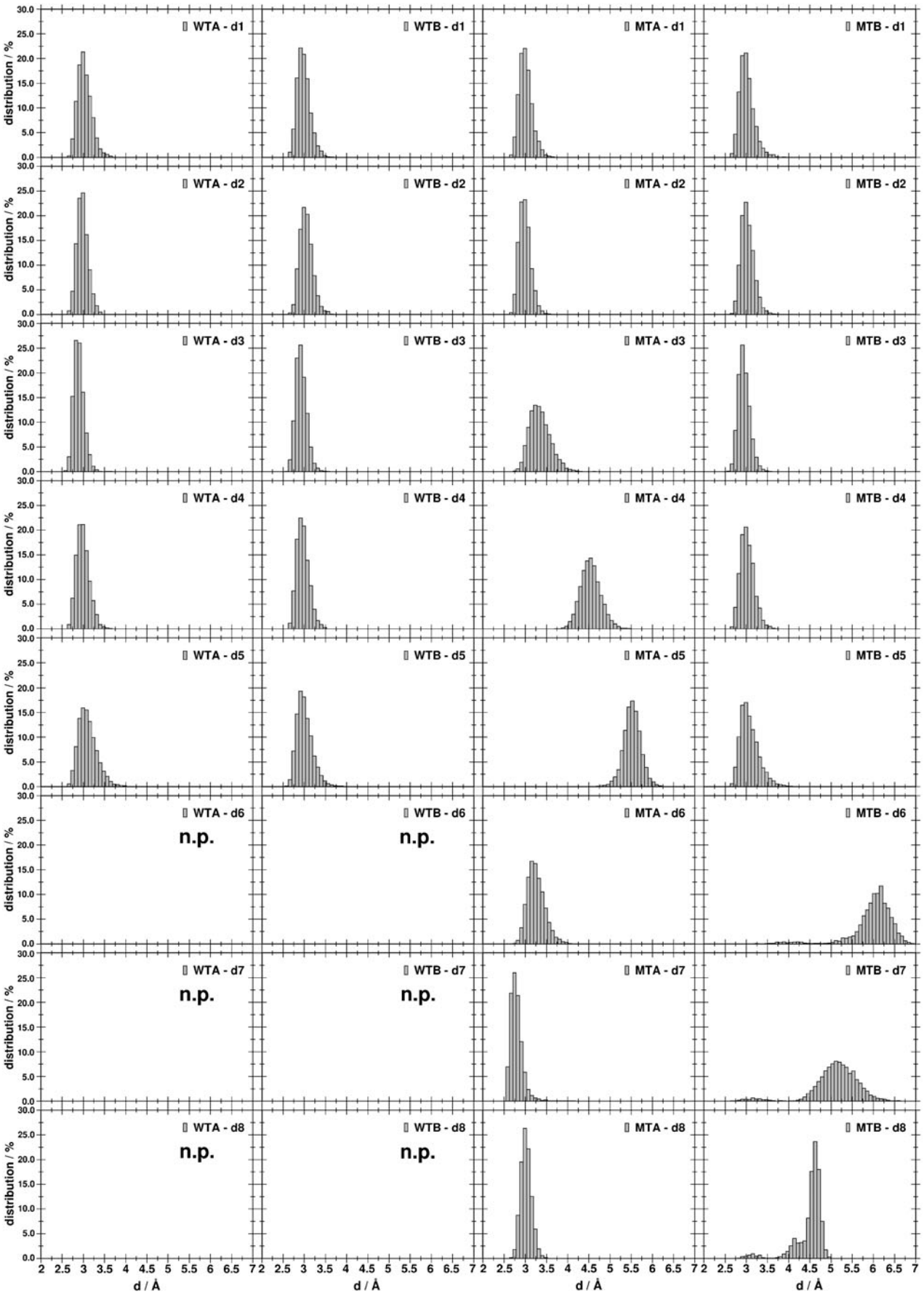
Tobias and co-workers described that the ΔG_{bind} of a model β-sheet (alanine dipeptide) is −5.5 kcal mol^{-1} [40], which is more stable compared to the ΔG_{bind} of formamide (−0.38 kcal mol^{-1}) [41].

The $\Delta\Delta G_{bind}$ values obtained indicate that, for MTB, WTA and WTB are more stable compared to the MTA system. The decreased stability (about 4–7 kcal mol^{-1}) is comparable with the alanine dipeptide model system [40], where two H-bonds vanished. In the case of MTA, one interaction point (d3) is weakened, d4 and d5 vanished completely, and a new one, d7, was formed [only inter-strand interactions were considered, the intra-strand interaction (d6) has no effect on the inter-strand stability].

Thoden et al. [12] suggested that the effect of $Pro^{28}Thr$ mutation manifested in improper folding, rather than in large conformational changes in the enzyme structure. The vanished H-bonds and the weakened interaction between the strands suggested that the proper folding of this part of the enzyme does not occur, and these findings are in good agreement with the statement of these authors. Nevertheless, no large conformational change was detected during the simulations, as confirmed by calculating the radius of gyration (R_{gyr}). This parameter provides information about large conformational changes during the course of the simulation (Fig. 7).

It can be seen from Fig. 7 that there is no significant alteration in the value of R_{gyr} compared to an unfolding

Fig. 4 Distribution (d_{min}=2 Å, d_{max}=7 Å, δd=0.2 Å) of the geometric parameters (d1, ..., d8; see Fig. 2) near the mutation. *WTA* Chain A of the wild type enzyme form; *WTB* chain B of the wild type enzyme form, *MTA* chain A of the mutated type enzyme form, *MTB* chain B of the mutated type enzyme form, *n.p.* not possible ▶


WTA - d1
WTB - d1
MTA - d1
MTB - d1
WTA - d2
WTB - d2
MTA - d2
MTB - d2
WTA - d3
WTB - d3
MTA - d3
MTB - d3
WTA - d4
WTB - d4
MTA - d4
MTB - d4
WTA - d5
WTB - d5
MTA - d5
MTB - d5
WTA - d6
WTB - d6
MTA - d6
MTB - d6
WTA - d7
WTB - d7
MTA - d7
MTB - d7
WTA - d8
WTB - d8
MTA - d8
MTB - d8
n.p.
distribution / %
d / Å

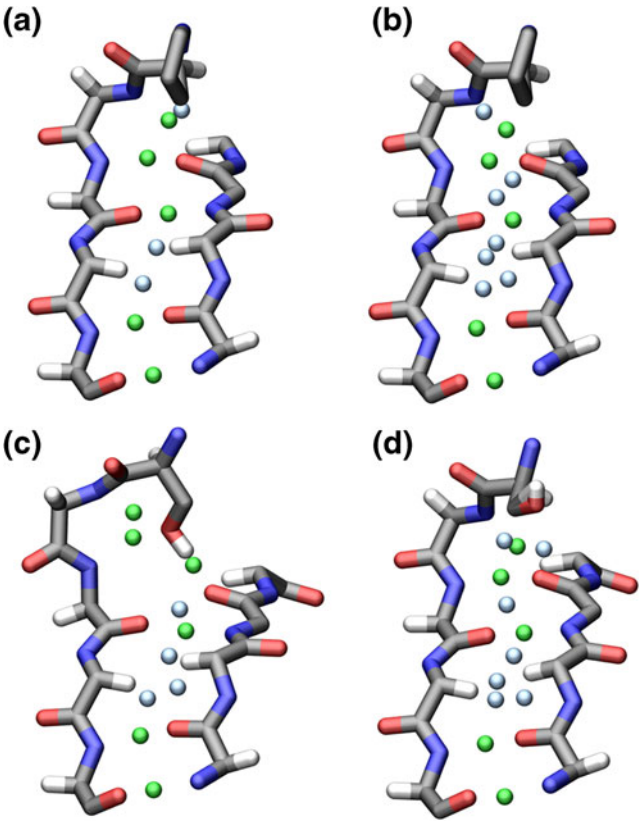


Fig. 5 The location of the bond critical points (bcps): *green points* are located in the path of H-bonding interactions, whereas *blue points* correspond to other interactions. **a** WTA, **b** WTB, **c** MTA, **d** MTB

process at high temperature [42], indicating that no large conformational shift took place. R_{gyr} was calculated in the extended (72 ns) trajectory as well, where the same behavior was detected (see Electronic supplementary material).

Based on the above results, we can conclude that there is a stability decrease in parallel with the β-sheet unfolding around the PM region in the case of MTA. Due to the loss of stability, the protein has become more sensitive to the alteration of the environment (pH, salt concentration, temperature), and any minor alteration in these parameters may cause defolding/unfolding of the whole protein structure, which may result in an insoluble protein.

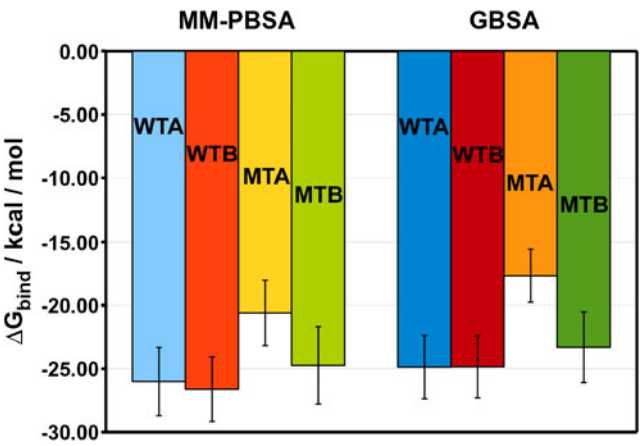


Fig. 6 Calculated ΔG_{bind} values with standard deviations for the model systems using either the MM-PBSA or the MM-GBSA methods

Table 3 Calculated $\Delta\Delta G_{bind}$ values (kcal mol^{-1}) obtained from MM-PBSA/GBSA calculations

	$\Delta\Delta G_{bind,\ MM\text{-}PBSA}$	$\Delta\Delta G_{bind,\ MM\text{-}GBSA}$
WTA	−5.41	−7.21
WTB	−6.03	−7.18
MTB	−4.14	−5.64

Further confirmation of the results

In 1998 Karplus and co-workers [29] described that the conformational space of a protein can be explored in more detail if the simulations are conducted with different initial velocities. To confirm that the insertion of the Thr^{28} side chain between the β-sheet strands during the calculation is not an artifact we conducted further 6 ns long MD simulations for WT and MT using different random numbers for generating the initial velocities. The average distances and standard deviations of the investigated interactions were calculated, and the values obtained for MTA are presented in Table 4.

The remaining tables for WTA, WTB and MTB are available in the Electronic supplementary material. Basically, the same results were obtained as with the 24 ns long MD simulations: (1) the d1 and d2 interactions, which are far from the mutation, are stable; (2) d3, d4 and d5 are

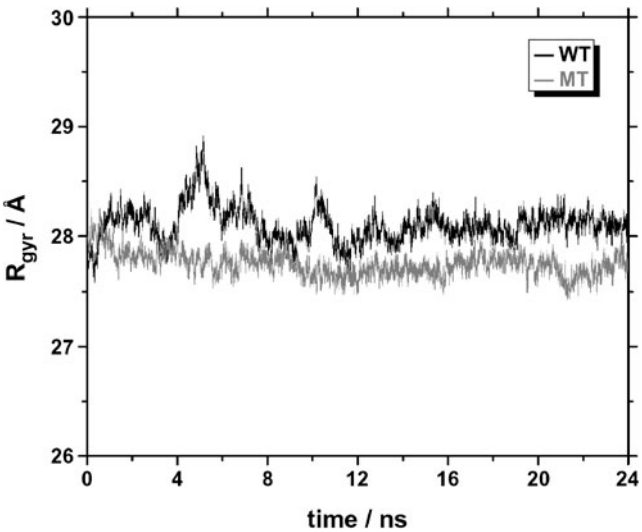


Fig. 7 Time evolution of the radius of gyration (R_{gyr}) in the course of the simulation

Table 4 Measured heavy atom distances between the possible H-bonding atoms near the mutation. Values were calculated in the last 3.504 ns. For definition of d1, ..., d8, see Fig. 2. MTA2, ..., MTA5: shorter (6 ns long) MD simulation of the MTA using different initial velocities

	MTA2	MTA3	MTA4	MTA5
d1	2.96±0.15	2.85±0.12	2.98±0.16	2.96±0.15
d2	2.95±0.13	2.89±0.13	2.94±0.13	2.94±0.13
d3	3.35±0.26	3.24±0.39	3.41±0.26	3.38±0.27
d4	4.48±0.22	4.03±0.82	4.66±0.26	4.53±0.26
d5	5.44±0.20	4.64±1.14	5.56±0.20	5.47±0.20
d6	3.25±0.22	3.41±0.38	3.31±0.24	3.21±0.21
d7	2.75±0.15	2.87±0.24	2.79±0.17	2.74±0.14
d8	2.96±0.12	3.00±0.13	3.01±0.13	2.95±0.12

weakened during the simulation; and (3) interactions were formed between the side chain –OH group and the appropriate backbone heavy atoms (d6, d7 and d8).

The initial 24 ns long MD simulation was extended to 72 ns and the above mentioned geometric parameters were calculated. These parameters do not change compared to the previous MD simulations.

Besides altering the initial conditions, the force field dependence of the results was also investigated. MD simulations of 12 ns duration were performed by applying the OPLSAA-2005 force field parameter set. The results for the interactions are available as Electronic supplementary material. The characteristics of average values and standard deviations are very similar to the parameters obtained with the CHARMM27 force field. We obtained somewhat higher values compared to previous simulations, especially for the interactions that vanished (d4 and d5) or formed (d6–d8) (see Electronic supplementary material). These higher values can be explained by the higher water influx to this site of the enzyme, which may indicate the important role of water molecules.

Possible mechanism of degradation of the β-sheet

The insertion of the side chain in the MTA cases developed quickly (already during the preequilibration stage), with the exception of MTA3, where the interactions responsible for the stability of the β-sheet (d3, d4 and d5) were stable during only the first ~4 ns, and vanished thereafter.

Together with the disappearance of these interactions, the increased fluctuation of d6 and d7 decreased, and remained stable in the last 2 ns (see Electronic supplementary material). The point at which the interactions change characteristically permitted us to study the mechanism of disruption of this part of the β-sheet. Until 3.960 ns, d6 and d7 form and disrupt several times (Fig. 8a).

At this point, one water molecule appears near the d4 interaction (Fig. 8b) and forms H-bonds with Leu^{30}@N and Gly^{65}@O with the following parameters (the angle formed between –D–H×××A is given in parenthesis): 2.96 Å (165.65°) and 2.78 Å (140.51°), respectively (Fig. 8c). In parallel with the water-mediated H-bond, a side chain -OH×××Leu^{30}@N /Gly^{65}@O interaction also forms, and remains stable for the last part of the simulation (Fig. 8d). It should be noted here that, before water attacks at 3.960 ns, there was no water influx in the 3.5 Å proximity of the Leu^{30}@N atom.

Based on the proposed mechanism, we can conclude that disruption of the β-sheet is caused by the collective effect of the Thr^{28} -OH group and the water molecule. Without Thr^{28} there is no destabilization effect on the d4 interaction, and without the water molecule, stable Leu^{30}@N-Thr^{28}@OH cannot be formed.

Conclusions

In this study we applied various theoretical methods including MD, MM-PBSA/GBSA and AIM analysis to the WT and MT form of the human galactokinase enzyme. Our conclusions may be summarized as follows:

(1) In the β-sheet of MTA close to the PM, significant alterations in the geometric structure of the H-bonding network were observed. Two backbone–backbone H-bonds disappear and three new backbone–side chain

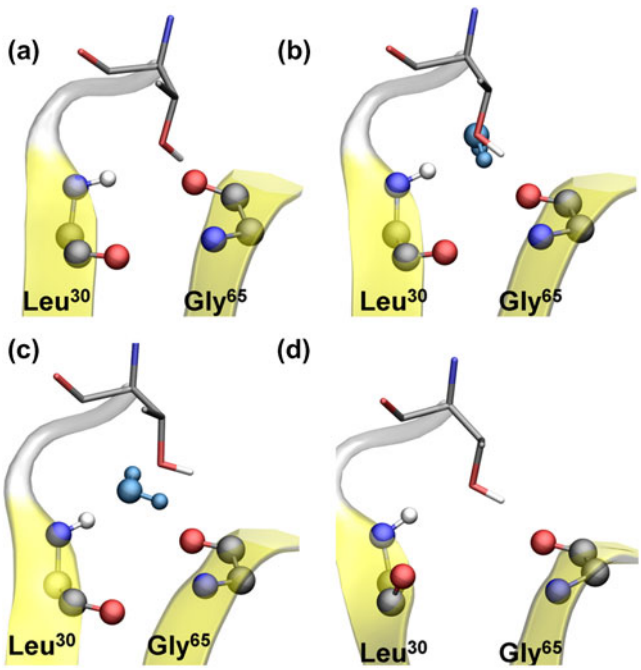


Fig. 8 Representative snapshots before (**a**), during (**b**, **c**) and after (**d**) the water attack at the Leu^{30}@N×××Gly^{65}@O interaction

interactions form, resulting in local unfolding of the secondary structure. Evaluating the R_{gyr} parameter allows us to conclude that there is no significant overall conformational change. These observations are in good agreement with the hypothesis of Thoden et al. [12].

(2) To support the conclusions in (1), AIM analysis was performed to improve the description of H-bonding interactions. The same disrupting and forming H-bond interactions were observed for all species examined.

(3) Other than structural investigations, we also examined the energetic consequences of local unfolding. ΔG_{bind} calculations indicated a ~5–7 kcal mol^{-1} decrease in stability corresponding to the unfolding.

(4) For further confirmation, some additional calculations were performed. These include: 5×6 ns MD applying different initial velocities, an extended 72 ns MD, and finally checking our results with the OPLS-AA force field. All these yield very similar results to what have been discussed previously. As similar results were obtained with different types of calculations, it may be concluded that our calculations are reliable.

(5) Based on our results, we propose a possible mechanism of local unfolding due to the $Pro^{28}Thr$ point mutation involving insertion of the –OH group of Thr^{28} between the β-sheets and the attack of a water molecule on the H-bonding networks responsible for its stability.

Acknowledgments This work was supported by "Társadalmi Megújulás Operatív Program" (TÁMOP-4.2.1/B-09/1/KONV-2010-0005). The authors thank M. Labádi for technical support at the High Performance Computing Centre of the University of Szeged. The help of Methos L. Müller in the preparation of the graphics is acknowledged.

References

1. Bordner AJ, Abagyan RA (2004) Large-scale prediction of protein geometry and stability changes for arbitrary single point mutations. Proteins Struct Func Bioinf 57:400–413
2. Gilis D, Rooman M (1999) Prediction of stability changes upon single-site mutations using database-derived potentials. Theor Chem Acc 101:46–50
3. Carlsson P, Koehler KF, Nilsson L (2005) Glucocorticoid receptor point mutation V571M facilitates coactivator and ligand binding by structural rearrangement and stabilization. Mol Endocrinol 19:1960–1977
4. Sneddon SF, Tobias DJ (1992) The role of packing interactions in stabilizing folded proteins. Biochemistry 31:2842–2846
5. Dang LX, Merz KM, Kollman PA (1989) Free-energy calculations on protein stability: Thr-1573Val-157 mutation of T4 lysozyme. J Am Chem Soc 111:8505–8508
6. Piana S, Laio A, Marinelli F, Van Troys M, Bourry D, Ampe Ch, Martins JC (2008) Predicting the effect of a point mutation on a protein fold: the villin and advillin headpieces and their Pro62Ala mutants. J Mol Biol 375:460–470
7. Sahai MA, Viskolcz B, Pai EF, Csizmadia IG (2007) Quantifying the intrinsic effects of two point mutation models of proline proline diamino acid diamide: a first-principle computational study. J Phys Chem B 111:11592–11602
8. Gitzelmann R, Hansen RG (1974) Galactose biogenesis and disposal in galactosemics. Biochim Biophys Acta 372:374–378
9. Holden HM, Thoden JB, Timson DJ, Reece RJ (2004) Galactokinase: structure, function and role in type II galactosemia. Cell Mol Life Sci 61:2471–2484
10. Kinoshita JH, Dikmak E, Satoh K, Merola L (1962) Osmotic changes caused by accumulation of dulcitol in lenses of rats fed With galactose. Nature 194:1085–1087
11. Tsakiris S, Schulpis KH, Marinou K, Behrakis P (2004) Protective effect of L-cysteine and glutathione on the modulated suckling rat brain Na+, K+ATPase and Mg2+ATPase activities induced by the in vitro galactosaemia. Pharmacol Res 49:475–479
12. Thoden JB, Timson DJ, Reece RJ, Holden HM (2005) Molecular structure of human galactokinase—implications for type II galactosemia. J Biol Chem 280:9662–9670
13. Timson DJ, Reece RJ (2003) Functional analysis of disease-causing mutations in human galactokinase. Eur J Biochem 270:1767–1774
14. Nowicka A, Mackiewicz P, Dudkiewicz M, Mackiewicz D, Kowalczuk M, Cebrat S, Dudek MR (2003) In: Sloot PMA et al (eds) Correlation between mutation pressure, selection pressure, and occurrence of amino acids. ICCS, LNCS 2658:650–657
15. Kalaydjieva LV, Perez-Lezaun A, Angelicheva D, Onengut S, Dye D, Bosshard NU, Jordanova A, Savov A, Yanakiev P, Kremensky I, Radeva B, Hallmayer J, Markov A, Nedkova V, Tournev I, Aneva L, Gitzelmann R (1999) A founder mutation in the GK1 gene is responsible for galactokinase deficiency in Roma (Gypsies). Am J Hum Genet 65:1299–1307
16. Hunter M, Heyer E, Austerlitz F, Angelicheva D, Nedkova V, Briones P, Gata A, De Pablo R, Laszlo A, Bosshard L, Gitzelmann R, Tordai A, Kalmar L, Szalai C, Balogh I, Lupu C, Corches A, Popa G, Perez-Lezaun A, Kalaydjieva LV (2002) The P28 mutation in the GALK1 gene accounts for galactokinase deficiency in Roma (Gypsy) patients across Europe. Pediatr Res 51:602–606
17. Capriotti E, Fariselli P, Casadio R (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic Acids Res 33:W306–W310
18. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242
19. Berman H, Henrick K, Nakamura H (2003) Announcing the Worldwide Protein Data Bank. Nat Struct Biol 10:980
20. Molecular Operating Environment (2007) C.C.G. Inc, Montreal, Quebec, Canada
21. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA (2004) Pdb2Pqr: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. Nucleic Acids Res 32: W665–W667
22. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of Simple Potential Functions for Simulating Liquid Water. J Chem Phys 79:926–935
23. MacKerell A, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616
24. Mackerell AD, Feig M, Brooks CL (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein

conformational distributions in molecular dynamics simulations. J Comput Chem 25:1400–1415
25. Beglov D, Roux B (1994) Finite representation of an infinite bulk system—solvent boundary potential for computer-simulations. J Chem Phys 100:9050–9063
26. Bowers KJ, Chow E, Xu H, Dror RO, Eastwood MP, Gregersen BA, Klepeis JL, Kolossvary I, Moraes MA, Sacerdoti FD, Salmon JK, Y, Shaw DE (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, SC'06
27. Berendsen HJC, Postma JPM, Van Gunsteren WF, Dinola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81:3684–3690
28. Darden T, York D, Pedersen L (1993) Particle mesh Ewald—an N×Log(N) method for Ewald sums in large systems. J Chem Phys 98:10089–10092
29. Caves LSD, Evanseck JD, Karplus M (1998) Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. Protein Sci 7:649–666
30. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field forpProteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 105:6474–6487
31. Bader RFW (1985) Atoms in molecules. Acc Chem Res 18:9–15
32. Bader RFW (1991) A quantum-theory of molecular-structure and its applications. Chem Rev 91:893–928
33. Srinivasan J, Cheatham TE, Cieplak P, Kollman PA, Case DA (1998) Continuum solvent studies of the stability of DNA, RNA and phosphoramidate-DNA helices. J Am Chem Soc 120:9401–9409
34. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE III (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 33:889–897
35. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Pearlman DA, Crowley M, Walker RC, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong KF, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews DH, Schafmeister C, Ross WS, Kollman PA (2006) AMBER 9. University of California, San Francisco
36. Brooks BR, Janežič D, Karplus M (1995) Harmonic analysis of large system. I. Methodology. J Comput Chem 16:1522–1542
37. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Graph 14:33–38
38. Kabsch W, Sander C (1983) Dictionary of protein secondary structure—pattern-recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637
39. Mezei M (2010) Simulaid: a simulation facilitator and analysis program. J Comput Chem 31:2658–2668
40. Tobias DJ, Sneddon SF, Brooks CL (1992) Stability of a model beta-sheet in water. J Mol Biol 227:1244–1252
41. Sneddon SF, Tobias DJ, Brooks CL (1989) Thermodynamics of amide hydrogen-bond formation in polar and apolar solvents. J Mol Biol 209:817–820
42. Seshasayee A (2005) High-temperature unfolding of a Trp-cage mini-protein: a molecular dynamics simulation study. Theor Biol Med Model 2:7–11