Insight into the mechanism of galactokinase: role of a critical glutamate residue and helix/coil transitions

Margaret McAuley[1], Meilan Huang[2] and David J. Timson[1,3*]

[1] School of Biological Sciences Queen's University Belfast, Medical Biology Building, 97 Lisburn Road, Belfast, BT9 7BL. UK.

[2]School of Chemistry and Chemical Engineering, Queen's University Belfast, David Keir Building, Stranmillis Road, Belfast, BT9 5AG. UK.

[3] School of Pharmacy and Biomolecular Sciences, University of Brighton, Huxley Building, Lewes Road, Brighton, BN2 4GJ. UK.

* Author to whom correspondence should be addressed.

School of Pharmacy and Biomolecular Sciences, University of Brighton, Huxley Building, Lewes Road, Brighton, BN2 4GJ. UK.

Telephone +44(0)1273641623

Fax +44(0)1273642090

Email d.timson@brighton.ac.uk

Abstract

Galactokinase, the enzyme which catalyses the first committed step in the Leloir pathway, has attracted interest due to its potential as a biocatalyst and as a possible drug target in the treatment of type I galactosemia. The mechanism of the enzyme is not fully elucidated. Molecular dynamics (MD) simulations of galactokinase with the active site residues Arg-37 and Asp-186 altered predicted that two regions (residues 174-179 and 231-240) had different dynamics as a consequence. Interestingly, the same two regions were also affected by alterations in Arg-105, Glu-174 and Arg-228. These three residues were identified as important in catalysis in previous computational studies on human galactokinase. Alteration of Arg-105 to methionine resulted in a modest reduction in activity with little change in stability. When Arg-228 was changed to methionine, the enzyme's interaction with both ATP and galactose was affected. This variant was significantly less stable than the wild-type protein. Changing Glu-174 to glutamine (but not to aspartate) resulted in no detectable activity and a less stable enzyme. Overall, these combined *in silico* and *in vitro* studies demonstrate the importance of a negative charge at position 174 and highlight the critical role of the dynamics in to key regions of the protein. We postulate that these regions may be critical for mediating the enzyme's structure and function.

Keywords: active site; molecular dynamics; protein flexibility; biocatalysis; galactosemia; GALK1

Introduction

Galactokinase (EC 2.7.1.6) catalyses the ATP-dependent site and stereospecific phosphorylation of the hexose monosaccharide galactose [1, 2]. This reaction is critical since it forms the first committed step of the Leloir pathway of galactose metabolism [3]. This pathway is required for the conversion of galactose into the glycolytic intermediate glucose 6-phosphate since hexokinase (EC 2.7.1.1; the first enzyme of the glycolytic pathway) has negligible activity with galactose [4]. The Leloir pathway has particular significance in young mammals since the main sugar present in milk is lactose, a disaccharide of glucose and galactose. Reduced activity of any of the enzymes in the Leloir pathway can result in the inherited metabolic disease galactosemia [5]. This disease has a wide spectrum of manifestations ranging from alterations in blood chemistry with only mild effects for the patients to death in childhood [6]. Mutations in the gene encoding galactokinase (*GALK1*) can result in type II galactosemia in humans and other mammals [7, 8]. This is considered to be the mildest form of galactosemia with early onset cataracts being the most important manifestation. These cataracts can be managed by diet or surgery and there appear to be few long-term complications for the patients [9]. In more severe forms of the disease (types I and III galactosemia) more significant pathology is thought to be linked to the build up of excess galactose 1-phosphate, the product of the reaction catalysed by galactokinase [10, 11]. Therefore, considerable efforts have been made to identify specific inhibitors of galactokinase in order to block the reaction it catalyses [12-17]. In effect, galactokinase inhibition would convert the more severe forms of galactosemia into the more manageable type II [9].

Galactokinase has also attracted some interest as a potential biocatalyst [18]. The site-specific modification of monosaccharides is chemically challenging. However, galactokinase targets C_1-OH for phosphorylation [19]. The high specificity of the human and yeast enzymes towards α-D-galactose (and some closely related sugars) limits their application in the synthesis of a wider range of sugar 1-phosphates [19-21]. Consequently, a number of studies have focussed on identifying galactokinases with broader specificity or broadening the specificity of the enzyme [22-29]. Many bacterial galactokinases have broader specificity than the human or yeast enzymes [30]. For example, the *Escherichia coli* enzyme catalyses the phosphorylation of a range of galactose derivatives, but is inactive with D-glucose [26]. Its activity with fluorinated galactose derivatives enabled its use in the enzymatic synthesis of the *O*-fluoroglucoside of *N*-methylanthranilate and fluorinated Thomsen–Friedenreich (T) antigens [31, 32]. The promiscuity of the *E. coli* enzyme has been expanded by protein engineering methods to include a wider range of D- and L-sugars [22, 23, 25]. Galactokinase from *Bifidobacterium infantis* is active with D-galacturonic acid and can also use a variety of

phosphate donors instead of ATP including dATP, GTP, dGTP, ITP and dTTP [33]. This versatility has enabled the use of the enzyme in the synthesis of Galacto-*N*-biose derivatives [34]. *Streptococcus pneumoniae* TIGR4 galactokinase is active with D-glucose, some L-monosaccharides and, unusually, with *N*- acetyl-D-galactosamine [35, 36]. This enzyme has been applied in the synthesis of complex oligosaccharides (e.g. globotriose) and UDP-sugars [37-39].

The catalytic mechanism of galactokinase is generally accepted to involve the initial removal of a proton from the C_1-OH and it is believed that this is done by an aspartate residue acting as a base in the active site [2, 40]. This generates a highly nucleophilic, negatively charged species which attacks the oxygen atom bridging the β- and γ-phosphates of ATP. The γ-phosphate is thus transferred to the galactose molecule and ADP is released. The initial state of the enzyme is regenerated by loss of the proton to water. This mechanism is not universally accepted [41]. Galactokinase is a member of the GHMP kinase family of enzymes (named from some of the enzymes originally assigned to the family: galactokinase, homoserine kinase, mevalonate kinase and phosphomevalonate kinase) [42, 43]. While the majority of these enzymes have an aspartate (or glutamate) residue in a structurally equivalent position to the putative active site base in galactokinase, at least one member does not. In homoserine kinase (EC 2.7.1.39) the equivalent residue is an asparagine, a residue which lacks the capacity to act as a base in this manner [44, 45]. It has been proposed, based on crystal structures and enzyme kinetic analysis, that homoserine kinase catalyzes the reaction partly by stabilizing the transition state [44]. It is possible that, in this enzyme, negatively charged oxygens on the γ-phosphate of ATP abstract the proton from homoserine facilitating a direct transfer of the phosphate group [2]. Furthermore, in order to act as a base capable of removing a proton from the monosaccharide, the aspartate in galactokinase would need a pK_a value considerably higher than that of the free amino acid (4.8 [46]) or the pK_a of the C_1-OH would need to be substantially reduced. Such alterations are possible in the interior of proteins. However, previous computational chemistry work suggests that the pK_a of this active site aspartate (Asp-186) in human galactokinase is 5.3-6.3 [47]. While this is displaced from the free solution value, it may not be sufficient to facilitate transfer of the proton. Nevertheless, a number of studies have demonstrated that this aspartate is critical for the function of galactokinase from a variety of species [16, 40, 47, 48]. In addition, an adjacent arginine residue (Arg-37 in human galactokinase) is also believed to be important and may play a role in modifying the pK_a of the aspartate or stabilising the negative charge that develops on the sugar as it loses the proton [40]. Experiments in which active site residues are altered using site-directed mutagenesis are always compounded by the possibility that they may also cause alterations to the protein's structure or stability. In the case of Asp-186 in human galactokinase, alteration to either asparagine or alanine decreased the protein's resistance to denaturation by urea [47]. Therefore, it

is difficult to deduce that the loss of enzyme activity resulting from these alterations is caused entirely by the loss of chemical functionality at the active site.

In an attempt to resolve some of these mechanistic issues, we previously conducted a detailed quantum mechanics/molecular mechanics (QM-MM) study of galactokinase [49]. This predicted that Asp-186 does not, directly, participate in the reaction and that there is direct transfer of the phosphate group from ATP to galactose. Its role may be to polarise and weaken the oxygen-hydrogen bond in C_1-OH facilitating the direct attack by ATP [47, 49]. This study also identified a number of other residues which may play a key role in the enzyme's mechanism. Arg-105 and Glu-174 were predicted to cooperate in the formation of a hydrogen bonding network which restricts the mobility of ATP in the active site. Arg-228 may stabilise the transient negative charge which develops on the oxygen bridging the β- and γ-phosphates of ATP during bind breakage [49].

A greater understanding of this enzyme's mechanism is important for the development of selective inhibitors of galactokinase and the fully exploiting its potential as a biocatalyst. Catalytically important residues could be sterically hindered or modified by drug-like molecules. The dynamic behaviour of enzymes is critical to their functions, including specificity and catalysis [50, 51]. Knowing the structural and dynamic requirements for catalysis may enable further engineering of the enzyme's specificity. Therefore, we conducted a combined molecular dynamics (MD) and enzymological study to further elucidate the roles of Arg-105, Glu-174 and Arg-228.

Materials and Methods

Molecular dynamics simulations

Protein Preparation and Molecular dynamics simulations were carried out as described by Huang et al, 2013 [49]. Chain A of the GALK crystal structure (1WUU) was altered using Biovia Discovery studio (Dassault Systèmes), Ser230 and Leu231 were added, selenomethionine residues were varied back to native methionine, AMP.PNP was altered to ATP and Mg^{2+} and the two coordinating water molecules were added to the active site. Variants were produced using the mutate feature in Biovia Discovery studio. Partial and RESP charges of galactose, Mg^{2+} and ATP were determined as previously described using the Gaussian 09 package [52] and antechamber encoded in Amber 10 [53] respectively. All MD simulations were carried out using the Amber 10 package and the Amber Parm 99 forcefields [54]. Prior to simulation the protein, ATP and galactose were soaked in a TIP3P water box using tleap encoded in Amber 10. The dimensions of the box for all proteins were 71.071 Å X 80.431 Å X 87.517

Å and the minimum distance to the boundary of the protein was set to 8 Å. Sodium ions were added to neutralise the system WT. E174D, R37K and R228K required seven ions, E174Q, D186A and D186N required 6 ions, both R105M and R228M required eight ions and R37E required nine. Minimisation and simulation steps were carried out as described by Huang et al [49]. The system was subjected to 1250 steps of first steepest descent minimisation then conjugate gradient minimisation, the latter was carried out with a 0.5 kcal mol^{-1} $Å^{-2}$ [55]. The system was then heated from 0K to 300K for 50ps and with a collision frequency of 5.0 ps^{-1}, using the Langevin dynamics method [56]. Equilibration was then carried out using an NVT ensemble, a periodic boundary was applied and the system maintained at 300K for 50 ps. Production simulation was then carried out for 7 ns with a time step of 1 fs, reference pressure of 1 atm and at 300 K. A cut-off distance for Van der Waals interactions was set at 10 Å and these long range interactions were calculated using the particle mesh Ewald method [57]. Hydrogen convalent bonds were constrained via the SHAKE method [58].

Analysis of simulations

Simulations were monitored using perl encoded in Amber and resulting graphs produced using GraphPad Prism version 5.03 for Windows (GraphPad Software, San Diego California USA) (Supplementary Figure S1). Cluster analysis, secondary structure analysis and RMSF calculations were performed using the ptraj command in Amber 10 and structures visualised using Biovia Discovery Studio. VMD software [59] was used to visualise trajectories, the timeline plugin was used to visualise secondary structure and Normal mode wizard was used to carry out principal component analysis (PCA) both were calculated using 1 ns of equilibrated trajectory, determined by monitoring RMSD over time (Supplementary Figure S2) [60].

Expression and purification of human galactokinase

Recombinant human galactokinase was expressed in, and purified from, *Escherichia coli* HMS174(DE3) as previously described except that cobalt agarose resin (His-Select, Sigma, Poole, UK) was used in place of nickel agarose [47, 61]. Site-directed mutations were carried out by the QuikChange method [62] and the mutated DNA sequences were verified (GATC Biotech, London, UK). Variant proteins were expressed and purified using the same protocol as used for the wild-type protein. Purified proteins were stored in buffer A (50 mM Hepes-OH, pH 7.5, 150 mM NaCl, 10% (v/v) glycerol, 1 mM DTT), frozen at -80 °C in aliquots of 20-100 µl until required.

Steady-state kinetic analysis of galactokinase

Galactokinase activity was measured by coupling the reaction to those catalysed by pyruvate kinase (EC 2.7.1.4) and lactate dehydrogenase (EC 1.1.1.27) [63]. Rates were measured (in triplicate) at 37 °C in 96 well plates by monitoring the decline in absorbance at 340 nm in a Thermo Scientific Multiskan spectrum platereader. The total reaction volume was 160 µl and each reaction contained 50 mM Hepes-OH, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.4 mM phospho*enol*pyruvate, 1mM NADH, 10 %(v/v) glycerol. Reactions were initiated by the addition of enzyme (125-690 nM depending on the variant), monitored for 30 min and the linear portion of the reaction extracted by visual examination. Rates were converted to molar units using the extinction coefficient for NADH (6220 l mol^{-1} cm^{-1} [64]) and a standard curve to correct for the reduced pathlength in the 96 well plate.

Typically reactions were arrayed in a eight by ten grid in which ATP concentration was varied along one axis and galactose along the other [20, 61]. This enabled the extraction of ten datasets with varying ATP concentration and constant galactose concentrations along with eight datasets with varying galactose concentration and constant ATP concentrations. For each of these 18 datasets, rates were plotted against the variable substrate concentration and these data fitted to the Michaelis-Menten equation (1) using non-linear curve fitting as implemented in GraphPad Prism 6.0 (GraphPad Software, CA, USA) [65, 66]:

$$v = \frac{V_{max,app}[\mathrm{S}]}{K_{m,app}+[\mathrm{S}]} \qquad (1)$$

where v is the initial rate, $V_{max,app}$ is the apparent maximal rate, $K_{m,app}$ is the apparent Michaelis constant and [S] is the concentration of the variable substrate. Note that these are apparent kinetic parameters since the experiment was carried out in sub-saturating concentrations of substrates. Since human galactokinase has an ordered ternary complex mechanism [61], the steady state constants were estimated by plotting the $V_{max,app}$ values obtained for the various constant concentrations of ATP against these ATP concentrations and fitting these data to equation (2) [67, 68]:

$$V_{max,app} = \frac{V_{max}[\mathrm{ATP}]}{K_{m,ATP}+[\mathrm{ATP}]} \qquad (2)$$

where $K_{m,ATP}$ is the Michaelis constant for ATP. A similar process was used to determine $K_{m,gal}$, the Michaelis constant for galactose.

Analytical methods

Differential scanning fluorimetry (DSF) was used to estimate the thermal unfolding temperature (or "melting" temperature, T_m) essentially as previously described [69]. Galactokinase (6.8 µM) was mixed with Sypro Orange (Sigma; 5 ×; manufacturer's concentration definition) in a total volume of 20 µl. Fluorescence was monitored as a function of temperature (25 °C to 90 °C 1°C per step, 5 seconds each step) using a RotorGeneQ qPCR machine (Qiagen). Data were analysed using RotorGeneQ software.

Native gel electrophoresis was used to show the resistance of the protein to denaturation by the chaotropic compound urea. A discrete band represents folded protein which becomes less discrete and more smeared as the urea concentration increases. The protocol was based on that of Megarity et al [47]. Protein (2.55 µM) was incubated with increasing concentrations of Urea (0-2 M) in 10 mM HEPES-OH, pH 7.0 and in the presence of native loading buffer (125 mM Tris-HCl, pH8.8, 20%(v/v) glycerol, 1%(w/v) DTT and 0.0002%(w/v) bromophenol blue) for 30 min at 37 °C. Samples were then immediately loaded onto 15% native polyacrylamide gels (378 mM Tris-HCl, pH 8.8) and electrophoresed for 3 h at 20 mA. Gels were then stained using 0.25% (w/v) Coomassie Blue R250, 10% (v/v) acetic acid and 45% (v/v) ethanol and destained in 5% (v/v) ethanol, 7.5% (v/v) acetic acid.

Protein concentrations were estimated using Bradford's method with BSA as a standard [70].

Results and Discussion

Previously characterised active site variants have altered MD in key regions

Previous studies have shown that Arg-37 and Asp-186 (and the equivalent residues in galactokinase from other species) are critical to human galactokinase's function. However, this could be due to loss of key functionality in the active site or destabilisation of the protein. To gain further insight into this, we conducted molecular dynamics studies on R37K, R37E, D186N and D186A (variants previously characterised biochemically and shown to have reduced activity [47]).

All of the active site variants studied showed a propensity for change in two regions adjacent to the active site (residues 174-179 and 231-240). This indicates that these regions are susceptible to change when wild type (WT) human galactokinase is altered at the active site. In WT the 231-240

region occurs primarily as alpha helix (65%) but in R37E, R37K and D186A the region exists for the highest percentage of time as a 3_{10} helix (88%, 68% and 50% respectively) and D186N primarily as a turn (90.7%) (Figure 2a). All variants show an increase in the presence of a turn in region 174-179 (Figure 2b). The proximity of these two regions to the catalytically important residues Glu-174 and Arg-228 is particularly interesting. The distance between Glu-174 and Mg^{2+} increases from 4.3 Å in most prevalent WT structure to approximately 6.4 Å in the presence of the turn (in the region 174-179). The distance between Arg-228 and the Mg^{2+} increases from 4.8 Å to 5.4 Å when the adjacent region is a turn or decreases to 4.5 Å if the same region is a 3_{10} helix (Figure 2c,d).

These predicted structural and dynamic changes in galactokinase strongly suggest that alteration of either R37 or D186 results in more than changes to active site chemistry. This is consistent with previous data that showed that alteration of these residues resulted in proteins that were less stable to denaturation by urea [47]. The MD simulation also suggests that the two regions identified here may be important in catalysis and for maintaining wild-type structure and dynamics.

Arginine 105 is implicated in ATP binding

Alteration of Arg-105 to methionine reduced the turnover number (k_{cat}) approximately twofold (Table 1; Supplementary Figure S3) and increased $K_{m,ATP}$ by approximately the same factor. Consequently, the specificity constant ($k_{cat}/K_{m,ATP}$) was also reduced. However, $K_{m,gal}$ was relatively unaffected. This suggests that this residue has a role in the initial interaction between the enzyme and ATP, consistent with the predictions from previous computational studies [49]. The removal of the positive charge at this position has no major changes on the overall, predicted dynamics of the protein (Figure 3). DSF indicated no significant alteration of T_m (55.4±0.2 °C) compared to WT (55.7±0.3 °C) indicating that stability towards thermal denaturation was not altered by this variation (Table 2; Supplementary Figure S4). R105M also showed similar resistance to denaturation by urea to the wild-type (Supplementary Figure S5).

Arg-105 was previously shown to form an electrostatic interaction with Glu-174 in a molecular dynamics simulation [49]. The simulation of the neutral alteration from arginine to methionine at this position predicted that this variant would cause drifting of all ligands and eventual covalent bonding of Glu-174 to Mg^{2+} [49]. MD carried out in this study indicated that region 231-240 has similar structural propensities to WT whereas the presence of a turn between 174-179 increased by 60% (Figure 3a,b). This was reflected by increased root mean square fluctuation (RMSF) and increased RMSD of normal modes in these two regions (Figure 3c,d), however increased or

decreased flexibility in these two regions does not affect stability of the protein. The MD simulation carried out for this study provided no evidence for an electrostatic interaction between Arg-105 and Glu-174 suggesting the effects on ATP binding are not due to interaction with Glu-174 (Figure 3e).

Alteration of arginine 228 impacts on substrate interactions

When Arg-228 was altered to either lysine or methionine there was a modest increase in turnover number and an increase in $K_{m,ATP}$ (Table 1; Supplementary Figure S3). In the case of R228M, there was a substantial increase in $K_{m,gal}$, however, the Michaelis constant for this substrate remains essentially unchanged in R228K. Both variants are significantly destabilised compared to the wild-type (Table 2; Supplementary Figure S4, S5). Similar to the Arg-105 mutants, the presence of the turn from 174-179 showed an increase to around 80% in both Arg228 variants (Figure 4a). The region 231-240 is of particular note in R228K where the percentage of time as a turn increases to 93% for residues 233 and 234. However, there is a large increase in the percentage of time, during which the residues 235 and 236 form no secondary structure. This may partly explain to the significant decrease in T_m (Figure 4b; Table 2). Another factor which may contribute to the decrease in T_m (particularly for R228M) is the increase in flexibility of multiple regions i.e. around residues 100 and 350 (Figure 4c).

A negative charge at position 174 is critical for catalysis

Alteration of Glu-174 to aspartate resulted in an enzyme with steady state kinetic parameters very similar to the wild-type (Table 1, Supplementary Figure S3). However, when the residue was altered to asparagine there was no detectable activity. This suggests that the presence of the negative charge at this position is critical to the function of the enzyme. However, the exact distance of this charge from the backbone is less critical (aspartate has a shorter side chain, compared to glutamate). E174D was not significantly destabilised compared to the wild-type; however, E174Q became unstable (Table 2; Supplementary Figure S4, S5).

Despite being very similar to the wild-type, MD simulations of E174D indicated that this alteration introduces α-helix into the region 174-179 (Figure 5a). The significant thermal destabilisation of E174Q is not explained by RMSF, which is generally similar to WT enzyme (figure 5c) but may be explained by the loss of structure in the 231-240 region: approximately 99.6% of the time this region

formed no defined secondary structure compared to 0% in the WT simulation (Figure 5b). Despite alteration in the propensity of secondary structures in the 231-240 and the 174-179 regions and alteration to RMSF per residue, E174D is not significantly different to WT in terms of kinetics or T_m which suggests that as long as the 231-240 region remains mostly structured the enzyme is stable.

Conclusions

These data demonstrate the critical role of Glu-174 in catalysis by galactokinase, supporting the predictions made in our previous computational chemistry study. Our findings are consistent with both the active site base and direct attack mechanisms of catalysis. However, they do highlight the fact that active site residues hypothesised to be involved in the active site base mechanism are also required for wild-type protein stability. Therefore, loss of activity on alteration of these residues is insufficient evidence that they are vital for catalysis.

Our findings further highlight the importance of the conformational dynamics of two regions in controlling the activity of this enzyme. This knowledge may be important in the design of novel galactokinase inhibitors and engineered forms of the enzyme with novel specificities. Compounds which induce structural and dynamic transitions similar to those seen in the more unstable forms of the protein are more likely to be successful inhibitors of the enzyme. Manipulation of the dynamics of these regions may result in altered activity or specificity. One previous observation is that increased promiscuity of human galactokinase is generally accompanied by reduced activity [27]. Ensuring that the structural and dynamic behaviour of these regions remains close to wild-type (perhaps by introducing compensatory changes) may enable the identification of variant enzymes with broad specificity and high activity.

Acknowledgements

MM thanks the Department of Employment and Learning, Northern Ireland (DELNI, UK) for a PhD studentship and Queens University Belfast's high performance computing department for use of the Dell Cluster. MH would like to acknowledge the financial support from InvestNI (Northern Ireland, UK; grant number: RD0314092).

Figures

Figure 1: **Residues considered in this study are part of, or close to, the active site**: Arg-37, Red; Arg-105 blue; Glu-174; yellow; Asp-186 purple and Arg-228 green.

Figure 2: **Molecular dynamics simulations of five galactokinase variants previously shown to have reduced activity.** (a) Secondary structure of residues 231-240 showing percentage time during the 1 ns of the stable trajectory (as determined from RMSD values) each secondary structure feature occurs. (b) Secondary structure of residues 174-179 (which are adjacent to active site) showing percentage time during the 1 ns of the stable trajectory (as determined from RMSD values) each secondary structure feature occurs. (c) Structures representing changes observed in residues 174-179 (red) and 231-240 (yellow) and the distance from Glu-174 and Arg-228 to the Mg^{2+} ion. (i) The unstructured region at 174-179 and alpha helix at 231-240. (ii) Turn at 174-179 and 231-240. (iii) Turn at 174-179 and 3_{10} helix at 231-240.

Figure 3: **Molecular dynamics simulation of galactokinase R105M**. (a) and (b) The secondary structures of residues 174-179 and 231-240 respectively. (c) RMSF of R105M by residue compared to the WT enzyme. (d) RMSD by residue (calculated using NMWiz plugin of VMD) Normal modes calculated by the principle component method and subtracted from values obtained for Wild-Type enzyme under the same conditions to allow easy comparison. (e) The distance from Arg-105 to Glu-174 throughout the MM trajectory for the Wild-type enzyme.

Figure 4: **Molecular dynamics simulation of galactokinase R228K and R228M**. (a) and (b) The secondary structured of residues 174-179 and 231-240 respectively. (c) RMSF by residue for R228K and R228M compared to WT galactokinase.

Figure 5: **Molecular dynamics simulation of galactokinase E174D and E174Q.** (a) and (b) The secondary structures for residues 174-179 and 231-240 respectively. (c) RMSF by residue for E174D and E174Q compared to WT galactokinase.

References

[1] C.E. Cardini, L.F. Leloir, Enzymic phosphorylation of galactosamine and galactose, Arch Biochem Biophys, 45 (1953) 55-64.
[2] H.M. Holden, J.B. Thoden, D.J. Timson, R.J. Reece, Galactokinase: structure, function and role in type II galactosemia, Cell Mol Life Sci 61 (2004) 2471-2484.
[3] P.A. Frey, The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose, FASEB J, 10 (1996) 461-470.
[4] A.A. Haritos, M.A. Rosemeyer, Purification and physical properties of hexokinase from human erythrocytes, Biochim Biophys Acta, 873 (1986) 335-339.
[5] D.J. Timson, The molecular basis of galactosemia - Past, present and future, Gene, 589 (2016) 133-141.
[6] J.L. Fridovich-Keil, J.H. Walter, Galactosemia, in: D. Valle, A.L. Beaudet, B. Vogelstein, K.W. Kinzler, S.E. Antonarakis, A. Ballabio (Eds.) The Online Metabolic and Molecular Bases of Inherited Diseases, McGraw-Hill, Place Published, 2008.
[7] D. Stambolian, Y. Ai, D. Sidjanin, K. Nesburn, G. Sathe, M. Rosenberg, D.J. Bergsma, Cloning of the galactokinase cDNA and identification of mutations in two families with cataracts, Nat Genet, 10 (1995) 307-312.
[8] D.J. Timson, R.J. Reece, J.B. Thoden, H.M. Holden, Galactokinase Deficiency, in: F. Lang (Ed.) Encyclopedia of Molecular Mechanisms of Disease, Springer, 2009, pp. 679-680.
[9] A.M. Bosch, H.D. Bakker, A.H. van Gennip, J.V. van Kempen, R.J. Wanders, F.A. Wijburg, Clinical features of galactokinase deficiency: a review of the literature, J Inherit Metab Dis, 25 (2002) 629-634.
[10] G.N. Donnell, W.R. Bergren, G. Perry, R. Koch, Galactose-1-phosphate in galactosemia, Pediatrics, 31 (1963) 802-810.
[11] T. Slepak, M. Tang, F. Addo, K. Lai, Intracellular galactose-1-phosphate accumulation leads to environmental stress response in yeast model, Mol Genet Metab, 86 (2005) 360-371.
[12] K. Lai, M.B. Boxer, A. Marabotti, GALK inhibitors for classic galactosemia, Future Med Chem, 6 (2014) 1003-1015.
[13] L. Liu, M. Tang, M.J. Walsh, K.R. Brimacombe, R. Pragani, C. Tanega, J.M. Rohde, H.L. Baker, E. Fernandez, B. Blackman, J.M. Bougie, W.H. Leister, D.S. Auld, M. Shen, K. Lai, M.B. Boxer, Structure activity relationships of human galactokinase inhibitors, Bioorg Med Chem Lett, 25 (2015) 721-727.
[14] S. Odejinmi, R. Rascon, M. Tang, H. Vankayalapati, K. Lai, Structure-activity analysis and cell-based optimization of human galactokinase inhibitors, ACS Med Chem Lett, 2 (2011) 667-672.
[15] M. Tang, S.I. Odejinmi, H. Vankayalapati, K.J. Wierenga, K. Lai, Innovative therapy for Classic Galactosemia - tale of two HTS, Mol Genet Metab, 105 (2012) 44-55.
[16] M. Tang, K. Wierenga, L.J. Elsas, K. Lai, Molecular and biochemical characterization of human galactokinase and its small molecule inhibitors, Chem Biol Interact, 188 (2010) 376-385.
[17] K.J. Wierenga, K. Lai, P. Buchwald, M. Tang, High-throughput screening for human galactokinase inhibitors, J Biomol Screen, 13 (2008) 415-423.
[18] H. Kristiansson, D.J. Timson, Galactokinases: potential biotechnological applications as biocatalysts, Curr Biotech, 1 (2012) 148-154.
[19] S.M. Howard, M.R. Heinrich, The anomeric specificity of yeast galactokinase, Arch

Biochem Biophys, 110 (1965) 395-400.
[20] D.J. Timson, R.J. Reece, Sugar recognition by human galactokinase, BMC Biochem, 4 (2003) 16.
[21] P. Thomas, E.M. Bessell, J.H. Westwood, The use of deoxyfluoro-D-galactopyranoses in a study of yeast galactokinase specificity, Biochem J, 139 (1974) 661-664.
[22] D. Hoffmeister, J. Yang, L. Liu, J.S. Thorson, Creation of the first anomeric D/L-sugar kinase by means of directed evolution, Proc Natl Acad Sci USA, 100 (2003) 13184-13189.
[23] D. Hoffmeister, J.S. Thorson, Mechanistic implications of *Escherichia coli* galactokinase structure-based engineering, ChemBioChem, 5 (2004) 989-992.
[24] J. Yang, L. Liu, J.S. Thorson, Structure-based enhancement of the first anomeric glucokinase, ChemBioChem, 5 (2004) 992-996.
[25] J. Yang, X. Fu, J. Liao, L. Liu, J.S. Thorson, Structure-Based Engineering of *E. coli* Galactokinase as a First Step toward In Vivo Glycorandomization, Chem Biol, 12 (2005) 657-664.
[26] J. Yang, X. Fu, Q. Jia, J. Shen, J.B. Biggins, J. Jiang, J. Zhao, J.J. Schmidt, P.G. Wang, J.S. Thorson, Studies on the substrate specificity of *Escherichia coli* galactokinase, Org Lett, 5 (2003) 2223-2226.
[27] H. Kristiansson, D.J. Timson, Increased promiscuity of human galactokinase following alteration of a single amino acid residue distant from the active site, ChemBioChem, 12 (2011) 2081-2087.
[28] M. McAuley, H. Kristiansson, M. Huang, A.L. Pey, D.J. Timson, Galactokinase promiscuity: a question of flexibility?, Biochem Soc Trans, 44 (2015) 116-122.
[29] L. Lee, S. Kinoshita, H. Kumagai, T. Tochikura, Galactokinase of *Bifidobacterium bifidum*, Agric Biol Chem, 44 (1980) 2961-2966.
[30] L.P. Conway, J. Voglmeir, Functional analysis of anomeric sugar kinases, Carbohydr Res, 432 (2016) 23-30.
[31] L. Caputi, M. Rejzek, T. Louveau, E.C. O'Neill, L. Hill, A. Osbourn, R.A. Field, A one-pot enzymatic approach to the *O*-fluoroglucoside of *N*-methylanthranilate, Bioorg Med Chem, 21 (2013) 4762-4767.
[32] J. Yan, X. Chen, F. Wang, H. Cao, Chemoenzymatic synthesis of mono- and di-fluorinated Thomsen-Friedenreich (T) antigens and their sialylated derivatives, Org Biomol Chem, 11 (2013) 842-848.
[33] L. Li, Y. Liu, W. Wang, J. Cheng, W. Zhao, P. Wang, A highly efficient galactokinase from *Bifidobacterium infantis* with broad substrate specificity, Carbohydr Res, 355 (2012) 35-39.
[34] L. Li, Y. Liu, T. Li, W. Wang, Z. Yu, C. Ma, J. Qu, W. Zhao, X. Chen, P.G. Wang, Efficient chemoenzymatic synthesis of novel galacto-*N*-biose derivatives and their sialylated forms, Chem Comm, 51 (2015) 10310-10313.
[35] M. Chen, L.L. Chen, Y. Zou, M. Xue, M. Liang, L. Jin, W.Y. Guan, J. Shen, W. Wang, L. Wang, J. Liu, P.G. Wang, Wide sugar substrate specificity of galactokinase from *Streptococcus pneumoniae* TIGR4, Carbohydr Res, 346 (2011) 2421-2425.
[36] Y. Zou, W. Wang, L. Cai, L. Chen, M. Xue, X. Zhang, J. Shen, M. Chen, Substrate specificity of galactokinase from *Streptococcus pneumoniae* TIGR4 towards galactose, glucose, and their derivatives, Bioorg Med Chem Lett, 22 (2012) 3540-3543.
[37] X. Zhao, Y. Zou, M. Xue, Z. Ma, S. Wang, P.G. Wang, M. Chen, A one-pot approach to bio-synthesize globotriose and its derivatives from simpler substrates, Eur J Med Chem, 80 (2014) 423-427.
[38] Y. Zou, M. Xue, W. Wang, L. Cai, L. Chen, J. Liu, P.G. Wang, J. Shen, M. Chen, One-pot

three-enzyme synthesis of UDP-Glc, UDP-Gal, and their derivatives, Carbohydr Res, 373 (2013) 76-81.
[39] M.M. Muthana, J. Qu, Y. Li, L. Zhang, H. Yu, L. Ding, H. Malekan, X. Chen, Efficient one-pot multienzyme synthesis of UDP-sugars using a promiscuous UDP-sugar pyrophosphorylase from *Bifidobacterium longum* (BLUSP), Chem Comm, 48 (2012) 2728-2730.
[40] L.A. Reinhardt, J.B. Thoden, G.S. Peters, H.M. Holden, W.W. Cleland, pH-rate profiles support a general base mechanism for galactokinase (*Lactococcus lactis*), FEBS Lett, 587 (2013) 2876-2881.
[41] J.B. Thoden, H.M. Holden, The molecular architecture of human N-acetylgalactosamine kinase, J Biol Chem, 280 (2005) 32784-32791.
[42] P. Bork, C. Sander, A. Valencia, Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases, Protein Sci, 2 (1993) 31-40.
[43] D.J. Timson, GHMP kinases - structures, mechanisms and potential for therpeutically relevant inhibition, Curr Enz Inhib, 3 (2007) 77-94.
[44] S.S. Krishna, T. Zhou, M. Daugherty, A. Osterman, H. Zhang, Structural basis for the catalysis and substrate specificity of homoserine kinase, Biochemistry, 40 (2001) 10810-10818.
[45] T. Zhou, M. Daugherty, N.V. Grishin, A.L. Osterman, H. Zhang, Structure and mechanism of homoserine kinase: prototype for the GHMP kinase superfamily, Structure, 8 (2000) 1247-1257.
[46] T.K. Harris, G.J. Turner, Structural basis of perturbed pKa values of catalytic groups in enzyme active sites, IUBMB Life, 53 (2002) 85-98.
[47] C.F. Megarity, M. Huang, C. Warnock, D.J. Timson, The role of the active site residues in human galactokinase: Implications for the mechanisms of GHMP kinases, Bioorg Chem, 39 (2011) 120-126.
[48] X. Chu, N. Li, X. Liu, D. Li, Functional studies of rat galactokinase, J Biotech, 141 (2009) 142-146.
[49] M. Huang, X. Li, J.W. Zou, D.J. Timson, Role of Arg228 in the Phosphorylation of Galactokinase: The Mechanism of GHMP Kinases by Quantum Mechanics/Molecular Mechanics Studies, Biochemistry, 52 (2013) 4858-4868.
[50] N.M. Goodey, S.J. Benkovic, Allosteric regulation and catalysis emerge via a common route, Nat Chem Biol, 4 (2008) 474-482.
[51] M. McAuley, D.J. Timson, Modulating Mobility: a Paradigm for Protein Engineering?, Appl Biochem Biotechnol, (2016) In press. doi: 10.1007/s12010-016-2200-y
[52] M.J. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson, Gaussian 09, revision A. 1, Gaussian Inc., Wallingford, CT, 2009.
[53] D. Case, T. Darden, T. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, M. Crowley, R. Walker, W. Zhang, AMBER 10 University of California, San Francisco, 2008.
[54] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J Comp Chem, 25 (2004) 1157-1174.
[55] P.K. Weiner, P.A. Kollman, AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions, J Comp Chem, 2 (1981) 287-303.
[56] R.W. Pastor, B.R. Brooks, A. Szabo, An analysis of the accuracy of Langevin and molecular

dynamics algorithms, Mol Phys, 65 (1988) 1409-1419.
[57] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, J Chem Phys, 98 (1993) 10089-10092.
[58] J. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J Comp Phys, 23 (1977) 327-341.
[59] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J Mol Graph, 14 (1996) 33-38, 27-38.
[60] A. Bakan, L.M. Meireles, I. Bahar, ProDy: protein dynamics inferred from theory and experiments, Bioinformatics, 27 (2011) 1575-1577.
[61] D.J. Timson, R.J. Reece, Functional analysis of disease-causing mutations in human galactokinase, Eur J Biochem, 270 (2003) 1767-1774.
[62] W. Wang, B.A. Malcolm, Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis, BioTechniques, 26 (1999) 680-682.
[63] A. Platt, H.C. Ross, S. Hankin, R.J. Reece, The insertion of two amino acids into a transcriptional inducer converts it into a galactokinase, Proc Natl Acad Sci USA, 97 (2000) 3154-3159.
[64] B.L. Horecker, A. Kornberg, The extinction coefficients of the reduced band of pyridine nucleotides, J Biol Chem, 175 (1948) 385-390.
[65] L. Michaelis, M.L. Menten, Kinetics of invertase action, Biochem Z, 49 (1913) 333-369.
[66] D. Marquardt, An algorithm for least squares estimation of nonlinear parameters, SIAM J App Math, 11 (1963) 431-441.
[67] A. Cornish-Bowden, Fundamentals of Enzyme Kinetics, Portland Press, Place Published, 2004.
[68] D.J. Timson, Quantitative enzymology, Curr Enz Inhib, 11 (2015) 12-31.
[69] T.J. McCorvie, Y. Liu, A. Frazer, T.J. Gleason, J.L. Fridovich-Keil, D.J. Timson, Altered cofactor binding affects stability and activity of human UDP-galactose 4'-epimerase: implications for type III galactosemia, Biochim Biophys Acta, 1822 (2012) 1516-1526.
[70] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal Biochem, 72 (1976) 248-254.

Insight into the mechanism of galactokinase: role of a critical glutamate residue and helix/coil transitions

Margaret McAuley[1], Meilan Huang[2] and David J. Timson[1,3*]

[1] School of Biological Sciences Queen's University Belfast, Medical Biology Building, 97 Lisburn Road, Belfast, BT9 7BL. UK.

[2]School of Chemistry and Chemical Engineering, Queen's University Belfast, David Keir Building, Stranmillis Road, Belfast, BT9 5AG. UK.

[3] School of Pharmacy and Biomolecular Sciences, University of Brighton, Huxley Building, Lewes Road, Brighton, BN2 4GJ. UK.

* Author to whom correspondence should be addressed.

School of Pharmacy and Biomolecular Sciences, University of Brighton, Huxley Building, Lewes Road, Brighton, BN2 4GJ. UK.

Telephone +44(0)1273641623

Fax +44(0)1273642090

Email d.timson@brighton.ac.uk

Abstract

Galactokinase, the enzyme which catalyses the first committed step in the Leloir pathway, has attracted interest due to its potential as a biocatalyst and as a possible drug target in the treatment of type I galactosemia. The mechanism of the enzyme is not fully elucidated. Molecular dynamics (MD) simulations of galactokinase with the active site residues Arg-37 and Asp-186 altered predicted that two regions (residues 174-179 and 231-240) had different dynamics as a consequence. Interestingly, the same two regions were also affected by alterations in Arg-105, Glu-174 and Arg-228. These three residues were identified as important in catalysis in previous computational studies on human galactokinase. Alteration of Arg-105 to methionine resulted in a modest reduction in activity with little change in stability. When Arg-228 was changed to methionine, the enzyme's interaction with both ATP and galactose was affected. This variant was significantly less stable than the wild-type protein. Changing Glu-174 to glutamine (but not to aspartate) resulted in no detectable activity and a less stable enzyme. Overall, these combined *in silico* and *in vitro* studies demonstrate the importance of a negative charge at position 174 and highlight the critical role of the dynamics in to key regions of the protein. We postulate that these regions may be critical for mediating the enzyme's structure and function.

Keywords: active site; molecular dynamics; protein flexibility; biocatalysis; galactosemia; GALK1

Introduction

Galactokinase (EC 2.7.1.6) catalyses the ATP-dependent site and stereospecific phosphorylation of the hexose monosaccharide galactose [1, 2]. This reaction is critical since it forms the first committed step of the Leloir pathway of galactose metabolism [3]. This pathway is required for the conversion of galactose into the glycolytic intermediate glucose 6-phosphate since hexokinase (EC 2.7.1.1; the first enzyme of the glycolytic pathway) has negligible activity with galactose [4]. The Leloir pathway has particular significance in young mammals since the main sugar present in milk is lactose, a disaccharide of glucose and galactose. Reduced activity of any of the enzymes in the Leloir pathway can result in the inherited metabolic disease galactosemia [5]. This disease has a wide spectrum of manifestations ranging from alterations in blood chemistry with only mild effects for the patients to death in childhood [6]. Mutations in the gene encoding galactokinase (*GALK1*) can result in type II galactosemia in humans and other mammals [7, 8]. This is considered to be the mildest form of galactosemia with early onset cataracts being the most important manifestation. These cataracts can be managed by diet or surgery and there appear to be few long-term complications for the patients [9]. In more severe forms of the disease (types I and III galactosemia) more significant pathology is thought to be linked to the build up of excess galactose 1-phosphate, the product of the reaction catalysed by galactokinase [10, 11]. Therefore, considerable efforts have been made to identify specific inhibitors of galactokinase in order to block the reaction it catalyses [12-17]. In effect, galactokinase inhibition would convert the more severe forms of galactosemia into the more manageable type II [9].

Galactokinase has also attracted some interest as a potential biocatalyst [18]. The site-specific modification of monosaccharides is chemically challenging. However, galactokinase targets C_1-OH for phosphorylation [19]. The high specificity of the human and yeast enzymes towards α-D-galactose (and some closely related sugars) limits their application in the synthesis of a wider range of sugar 1-phosphates [19-21]. Consequently, a number of studies have focussed on identifying galactokinases with broader specificity or broadening the specificity of the enzyme [22-29]. Many bacterial galactokinases have broader specificity than the human or yeast enzymes [30]. For example, the *Escherichia coli* enzyme catalyses the phosphorylation of a range of galactose derivatives, but is inactive with D-glucose [26]. Its activity with fluorinated galactose derivatives enabled its use in the enzymatic synthesis of the *O*-fluoroglucoside of *N*-methylanthranilate and fluorinated Thomsen–Friedenreich (T) antigens [31, 32]. The promiscuity of the *E. coli* enzyme has been expanded by protein engineering methods to include a wider range of D- and L-sugars [22, 23, 25]. Galactokinase from *Bifidobacterium infantis* is active with D-galacturonic acid and can also use a variety of

phosphate donors instead of ATP including dATP, GTP, dGTP, ITP and dTTP [33]. This versatility has enabled the use of the enzyme in the synthesis of Galacto-*N*-biose derivatives [34]. *Streptococcus pneumoniae* TIGR4 galactokinase is active with D-glucose, some L-monosaccharides and, unusually, with *N*- acetyl-D-galactosamine [35, 36]. This enzyme has been applied in the synthesis of complex oligosaccharides (e.g. globotriose) and UDP-sugars [37-39].

The catalytic mechanism of galactokinase is generally accepted to involve the initial removal of a proton from the C_1-OH and it is believed that this is done by an aspartate residue acting as a base in the active site [2, 40]. This generates a highly nucleophilic, negatively charged species which attacks the oxygen atom bridging the β- and γ-phosphates of ATP. The γ-phosphate is thus transferred to the galactose molecule and ADP is released. The initial state of the enzyme is regenerated by loss of the proton to water. This mechanism is not universally accepted [41]. Galactokinase is a member of the GHMP kinase family of enzymes (named from some of the enzymes originally assigned to the family: galactokinase, homoserine kinase, mevalonate kinase and phosphomevalonate kinase) [42, 43]. While the majority of these enzymes have an aspartate (or glutamate) residue in a structurally equivalent position to the putative active site base in galactokinase, at least one member does not. In homoserine kinase (EC 2.7.1.39) the equivalent residue is an asparagine, a residue which lacks the capacity to act as a base in this manner [44, 45]. It has been proposed, based on crystal structures and enzyme kinetic analysis, that homoserine kinase catalyzes the reaction partly by stabilizing the transition state [44]. It is possible that, in this enzyme, negatively charged oxygens on the γ-phosphate of ATP abstract the proton from homoserine facilitating a direct transfer of the phosphate group [2]. Furthermore, in order to act as a base capable of removing a proton from the monosaccharide, the aspartate in galactokinase would need a pK_a value considerably higher than that of the free amino acid (4.8 [46]) or the pK_a of the C_1-OH would need to be substantially reduced. Such alterations are possible in the interior of proteins. However, previous computational chemistry work suggests that the pK_a of this active site aspartate (Asp-186) in human galactokinase is 5.3-6.3 [47]. While this is displaced from the free solution value, it may not be sufficient to facilitate transfer of the proton. Nevertheless, a number of studies have demonstrated that this aspartate is critical for the function of galactokinase from a variety of species [16, 40, 47, 48]. In addition, an adjacent arginine residue (Arg-37 in human galactokinase) is also believed to be important and may play a role in modifying the pK_a of the aspartate or stabilising the negative charge that develops on the sugar as it loses the proton [40]. Experiments in which active site residues are altered using site-directed mutagenesis are always compounded by the possibility that they may also cause alterations to the protein's structure or stability. In the case of Asp-186 in human galactokinase, alteration to either asparagine or alanine decreased the protein's resistance to denaturation by urea [47]. Therefore, it

is difficult to deduce that the loss of enzyme activity resulting from these alterations is caused entirely by the loss of chemical functionality at the active site.

In an attempt to resolve some of these mechanistic issues, we previously conducted a detailed quantum mechanics/molecular mechanics (QM-MM) study of galactokinase [49]. This predicted that Asp-186 does not, directly, participate in the reaction and that there is direct transfer of the phosphate group from ATP to galactose. Its role may be to polarise and weaken the oxygen-hydrogen bond in C_1-OH facilitating the direct attack by ATP [47, 49]. This study also identified a number of other residues which may play a key role in the enzyme's mechanism. Arg-105 and Glu-174 were predicted to cooperate in the formation of a hydrogen bonding network which restricts the mobility of ATP in the active site. Arg-228 may stabilise the transient negative charge which develops on the oxygen bridging the β- and γ-phosphates of ATP during bind breakage [49].

A greater understanding of this enzyme's mechanism is important for the development of selective inhibitors of galactokinase and the fully exploiting its potential as a biocatalyst. Catalytically important residues could be sterically hindered or modified by drug-like molecules. The dynamic behaviour of enzymes is critical to their functions, including specificity and catalysis [50, 51]. Knowing the structural and dynamic requirements for catalysis may enable further engineering of the enzyme's specificity. Therefore, we conducted a combined molecular dynamics (MD) and enzymological study to further elucidate the roles of Arg-105, Glu-174 and Arg-228.

Materials and Methods

Molecular dynamics simulations

Protein Preparation and Molecular dynamics simulations were carried out as described by Huang et al, 2013 [49]. Chain A of the GALK crystal structure (1WUU) was altered using Biovia Discovery studio (Dassault Systèmes), Ser230 and Leu231 were added, selenomethionine residues were varied back to native methionine, AMP.PNP was altered to ATP and Mg^{2+} and the two coordinating water molecules were added to the active site. Variants were produced using the mutate feature in Biovia Discovery studio. Partial and RESP charges of galactose, Mg^{2+} and ATP were determined as previously described using the Gaussian 09 package [52] and antechamber encoded in Amber 10 [53] respectively. All MD simulations were carried out using the Amber 10 package and the Amber Parm 99 forcefields [54]. Prior to simulation the protein, ATP and galactose were soaked in a TIP3P water box using tleap encoded in Amber 10. The dimensions of the box for all proteins were 71.071 Å X 80.431 Å X 87.517

Å and the minimum distance to the boundary of the protein was set to 8 Å. Sodium ions were added to neutralise the system WT. E174D, R37K and R228K required seven ions, E174Q, D186A and D186N required 6 ions, both R105M and R228M required eight ions and R37E required nine. Minimisation and simulation steps were carried out as described by Huang et al [49]. The system was subjected to 1250 steps of first steepest descent minimisation then conjugate gradient minimisation, the latter was carried out with a 0.5 kcal mol^{-1} $Å^{-2}$ [55]. The system was then heated from 0K to 300K for 50ps and with a collision frequency of 5.0 ps^{-1}, using the Langevin dynamics method [56]. Equilibration was then carried out using an NVT ensemble, a periodic boundary was applied and the system maintained at 300K for 50 ps. Production simulation was then carried out for 7 ns with a time step of 1 fs, reference pressure of 1 atm and at 300 K. A cut-off distance for Van der Waals interactions was set at 10 Å and these long range interactions were calculated using the particle mesh Ewald method [57]. Hydrogen convalent bonds were constrained via the SHAKE method [58].

Analysis of simulations

Simulations were monitored using perl encoded in Amber and resulting graphs produced using GraphPad Prism version 5.03 for Windows (GraphPad Software, San Diego California USA) (Supplementary Figure S1). Cluster analysis, secondary structure analysis and RMSF calculations were performed using the ptraj command in Amber 10 and structures visualised using Biovia Discovery Studio. VMD software [59] was used to visualise trajectories, the timeline plugin was used to visualise secondary structure and Normal mode wizard was used to carry out principal component analysis (PCA) both were calculated using 1 ns of equilibrated trajectory, determined by monitoring RMSD over time (Supplementary Figure S2) [60].

Expression and purification of human galactokinase

Recombinant human galactokinase was expressed in, and purified from, *Escherichia coli* HMS174(DE3) as previously described except that cobalt agarose resin (His-Select, Sigma, Poole, UK) was used in place of nickel agarose [47, 61]. Site-directed mutations were carried out by the QuikChange method [62] and the mutated DNA sequences were verified (GATC Biotech, London, UK). Variant proteins were expressed and purified using the same protocol as used for the wild-type protein. Purified proteins were stored in buffer A (50 mM Hepes-OH, pH 7.5, 150 mM NaCl, 10% (v/v) glycerol, 1 mM DTT), frozen at -80 °C in aliquots of 20-100 µl until required.

Steady-state kinetic analysis of galactokinase

Galactokinase activity was measured by coupling the reaction to those catalysed by pyruvate kinase (EC 2.7.1.4) and lactate dehydrogenase (EC 1.1.1.27) [63]. Rates were measured (in triplicate) at 37 °C in 96 well plates by monitoring the decline in absorbance at 340 nm in a Thermo Scientific Multiskan spectrum platereader. The total reaction volume was 160 µl and each reaction contained 50 mM Hepes-OH, pH 7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.4 mM phospho*enol*pyruvate, 1mM NADH, 10 %(v/v) glycerol. Reactions were initiated by the addition of enzyme (125-690 nM depending on the variant), monitored for 30 min and the linear portion of the reaction extracted by visual examination. Rates were converted to molar units using the extinction coefficient for NADH (6220 l mol^{-1} cm^{-1} [64]) and a standard curve to correct for the reduced pathlength in the 96 well plate.

Typically reactions were arrayed in a eight by ten grid in which ATP concentration was varied along one axis and galactose along the other [20, 61]. This enabled the extraction of ten datasets with varying ATP concentration and constant galactose concentrations along with eight datasets with varying galactose concentration and constant ATP concentrations. For each of these 18 datasets, rates were plotted against the variable substrate concentration and these data fitted to the Michaelis-Menten equation (1) using non-linear curve fitting as implemented in GraphPad Prism 6.0 (GraphPad Software, CA, USA) [65, 66]:

$$v = \frac{V_{max,app}[\mathrm{S}]}{K_{m,app}+[\mathrm{S}]} \qquad (1)$$

where v is the initial rate, $V_{max,app}$ is the apparent maximal rate, $K_{m,app}$ is the apparent Michaelis constant and [S] is the concentration of the variable substrate. Note that these are apparent kinetic parameters since the experiment was carried out in sub-saturating concentrations of substrates. Since human galactokinase has an ordered ternary complex mechanism [61], the steady state constants were estimated by plotting the $V_{max,app}$ values obtained for the various constant concentrations of ATP against these ATP concentrations and fitting these data to equation (2) [67, 68]:

$$V_{max,app} = \frac{V_{max}[\mathrm{ATP}]}{K_{m,ATP}+[\mathrm{ATP}]} \qquad (2)$$

where $K_{m,ATP}$ is the Michaelis constant for ATP. A similar process was used to determine $K_{m,gal}$, the Michaelis constant for galactose.

Analytical methods

Differential scanning fluorimetry (DSF) was used to estimate the thermal unfolding temperature (or "melting" temperature, T_m) essentially as previously described [69]. Galactokinase (6.8 µM) was mixed with Sypro Orange (Sigma; 5 ×; manufacturer's concentration definition) in a total volume of 20 µl. Fluorescence was monitored as a function of temperature (25 °C to 90 °C 1°C per step, 5 seconds each step) using a RotorGeneQ qPCR machine (Qiagen). Data were analysed using RotorGeneQ software.

Native gel electrophoresis was used to show the resistance of the protein to denaturation by the chaotropic compound urea. A discrete band represents folded protein which becomes less discrete and more smeared as the urea concentration increases. The protocol was based on that of Megarity et al [47]. Protein (2.55 µM) was incubated with increasing concentrations of Urea (0-2 M) in 10 mM HEPES-OH, pH 7.0 and in the presence of native loading buffer (125 mM Tris-HCl, pH8.8, 20%(v/v) glycerol, 1%(w/v) DTT and 0.0002%(w/v) bromophenol blue) for 30 min at 37 °C. Samples were then immediately loaded onto 15% native polyacrylamide gels (378 mM Tris-HCl, pH 8.8) and electrophoresed for 3 h at 20 mA. Gels were then stained using 0.25% (w/v) Coomassie Blue R250, 10% (v/v) acetic acid and 45% (v/v) ethanol and destained in 5% (v/v) ethanol, 7.5% (v/v) acetic acid.

Protein concentrations were estimated using Bradford's method with BSA as a standard [70].

Results and Discussion

Previously characterised active site variants have altered MD in key regions

Previous studies have shown that Arg-37 and Asp-186 (and the equivalent residues in galactokinase from other species) are critical to human galactokinase's function. However, this could be due to loss of key functionality in the active site or destabilisation of the protein. To gain further insight into this, we conducted molecular dynamics studies on R37K, R37E, D186N and D186A (variants previously characterised biochemically and shown to have reduced activity [47]).

All of the active site variants studied showed a propensity for change in two regions adjacent to the active site (residues 174-179 and 231-240). This indicates that these regions are susceptible to change when wild type (WT) human galactokinase is altered at the active site. In WT the 231-240

region occurs primarily as alpha helix (65%) but in R37E, R37K and D186A the region exists for the highest percentage of time as a 3_{10} helix (88%, 68% and 50% respectively) and D186N primarily as a turn (90.7%) (Figure 2a). All variants show an increase in the presence of a turn in region 174-179 (Figure 2b). The proximity of these two regions to the catalytically important residues Glu-174 and Arg-228 is particularly interesting. The distance between Glu-174 and Mg^{2+} increases from 4.3 Å in most prevalent WT structure to approximately 6.4 Å in the presence of the turn (in the region 174-179). The distance between Arg-228 and the Mg^{2+} increases from 4.8 Å to 5.4 Å when the adjacent region is a turn or decreases to 4.5 Å if the same region is a 3_{10} helix (Figure 2c,d).

These predicted structural and dynamic changes in galactokinase strongly suggest that alteration of either R37 or D186 results in more than changes to active site chemistry. This is consistent with previous data that showed that alteration of these residues resulted in proteins that were less stable to denaturation by urea [47]. The MD simulation also suggests that the two regions identified here may be important in catalysis and for maintaining wild-type structure and dynamics.

Arginine 105 is implicated in ATP binding

Alteration of Arg-105 to methionine reduced the turnover number (k_{cat}) approximately twofold (Table 1; Supplementary Figure S3) and increased $K_{m,ATP}$ by approximately the same factor. Consequently, the specificity constant ($k_{cat}/K_{m,ATP}$) was also reduced. However, $K_{m,gal}$ was relatively unaffected. This suggests that this residue has a role in the initial interaction between the enzyme and ATP, consistent with the predictions from previous computational studies [49]. The removal of the positive charge at this position has no major changes on the overall, predicted dynamics of the protein (Figure 3). DSF indicated no significant alteration of T_m (55.4±0.2 °C) compared to WT (55.7±0.3 °C) indicating that stability towards thermal denaturation was not altered by this variation (Table 2; Supplementary Figure S4). R105M also showed similar resistance to denaturation by urea to the wild-type (Supplementary Figure S5).

Arg-105 was previously shown to form an electrostatic interaction with Glu-174 in a molecular dynamics simulation [49]. The simulation of the neutral alteration from arginine to methionine at this position predicted that this variant would cause drifting of all ligands and eventual covalent bonding of Glu-174 to Mg^{2+} [49]. MD carried out in this study indicated that region 231-240 has similar structural propensities to WT whereas the presence of a turn between 174-179 increased by 60% (Figure 3a,b). This was reflected by increased root mean square fluctuation (RMSF) and increased RMSD of normal modes in these two regions (Figure 3c,d), however increased or

decreased flexibility in these two regions does not affect stability of the protein. The MD simulation carried out for this study provided no evidence for an electrostatic interaction between Arg-105 and Glu-174 suggesting the effects on ATP binding are not due to interaction with Glu-174 (Figure 3e).

Alteration of arginine 228 impacts on substrate interactions

When Arg-228 was altered to either lysine or methionine there was a modest increase in turnover number and an increase in $K_{m,ATP}$ (Table 1; Supplementary Figure S3). In the case of R228M, there was a substantial increase in $K_{m,gal}$, however, the Michaelis constant for this substrate remains essentially unchanged in R228K. Both variants are significantly destabilised compared to the wild-type (Table 2; Supplementary Figure S4, S5). Similar to the Arg-105 mutants, the presence of the turn from 174-179 showed an increase to around 80% in both Arg228 variants (Figure 4a). The region 231-240 is of particular note in R228K where the percentage of time as a turn increases to 93% for residues 233 and 234. However, there is a large increase in the percentage of time, during which the residues 235 and 236 form no secondary structure. This may partly explain to the significant decrease in T_m (Figure 4b; Table 2). Another factor which may contribute to the decrease in T_m (particularly for R228M) is the increase in flexibility of multiple regions i.e. around residues 100 and 350 (Figure 4c).

A negative charge at position 174 is critical for catalysis

Alteration of Glu-174 to aspartate resulted in an enzyme with steady state kinetic parameters very similar to the wild-type (Table 1, Supplementary Figure S3). However, when the residue was altered to asparagine there was no detectable activity. This suggests that the presence of the negative charge at this position is critical to the function of the enzyme. However, the exact distance of this charge from the backbone is less critical (aspartate has a shorter side chain, compared to glutamate). E174D was not significantly destabilised compared to the wild-type; however, E174Q became unstable (Table 2; Supplementary Figure S4, S5).

Despite being very similar to the wild-type, MD simulations of E174D indicated that this alteration introduces α-helix into the region 174-179 (Figure 5a). The significant thermal destabilisation of E174Q is not explained by RMSF, which is generally similar to WT enzyme (figure 5c) but may be explained by the loss of structure in the 231-240 region: approximately 99.6% of the time this region

formed no defined secondary structure compared to 0% in the WT simulation (Figure 5b). Despite alteration in the propensity of secondary structures in the 231-240 and the 174-179 regions and alteration to RMSF per residue, E174D is not significantly different to WT in terms of kinetics or T_m which suggests that as long as the 231-240 region remains mostly structured the enzyme is stable.

Conclusions

These data demonstrate the critical role of Glu-174 in catalysis by galactokinase, supporting the predictions made in our previous computational chemistry study. Our findings are consistent with both the active site base and direct attack mechanisms of catalysis. However, they do highlight the fact that active site residues hypothesised to be involved in the active site base mechanism are also required for wild-type protein stability. Therefore, loss of activity on alteration of these residues is insufficient evidence that they are vital for catalysis.

Our findings further highlight the importance of the conformational dynamics of two regions in controlling the activity of this enzyme. This knowledge may be important in the design of novel galactokinase inhibitors and engineered forms of the enzyme with novel specificities. Compounds which induce structural and dynamic transitions similar to those seen in the more unstable forms of the protein are more likely to be successful inhibitors of the enzyme. Manipulation of the dynamics of these regions may result in altered activity or specificity. One previous observation is that increased promiscuity of human galactokinase is generally accompanied by reduced activity [27]. Ensuring that the structural and dynamic behaviour of these regions remains close to wild-type (perhaps by introducing compensatory changes) may enable the identification of variant enzymes with broad specificity and high activity.

Acknowledgements

MM thanks the Department of Employment and Learning, Northern Ireland (DELNI, UK) for a PhD studentship and Queens University Belfast's high performance computing department for use of the Dell Cluster. MH would like to acknowledge the financial support from InvestNI (Northern Ireland, UK; grant number: RD0314092).

Figures

Figure 1: **Residues considered in this study are part of, or close to, the active site**: Arg-37, Red; Arg-105 blue; Glu-174; yellow; Asp-186 purple and Arg-228 green.

Figure 2: **Molecular dynamics simulations of five galactokinase variants previously shown to have reduced activity.** (a) Secondary structure of residues 231-240 showing percentage time during the 1 ns of the stable trajectory (as determined from RMSD values) each secondary structure feature occurs. (b) Secondary structure of residues 174-179 (which are adjacent to active site) showing percentage time during the 1 ns of the stable trajectory (as determined from RMSD values) each secondary structure feature occurs. (c) Structures representing changes observed in residues 174-179 (red) and 231-240 (yellow) and the distance from Glu-174 and Arg-228 to the Mg^{2+} ion. (i) The unstructured region at 174-179 and alpha helix at 231-240. (ii) Turn at 174-179 and 231-240. (iii) Turn at 174-179 and 3_{10} helix at 231-240.

Figure 3: **Molecular dynamics simulation of galactokinase R105M**. (a) and (b) The secondary structures of residues 174-179 and 231-240 respectively. (c) RMSF of R105M by residue compared to the WT enzyme. (d) RMSD by residue (calculated using NMWiz plugin of VMD) Normal modes calculated by the principle component method and subtracted from values obtained for Wild-Type enzyme under the same conditions to allow easy comparison. (e) The distance from Arg-105 to Glu-174 throughout the MM trajectory for the Wild-type enzyme.

Figure 4: **Molecular dynamics simulation of galactokinase R228K and R228M**. (a) and (b) The secondary structured of residues 174-179 and 231-240 respectively. (c) RMSF by residue for R228K and R228M compared to WT galactokinase.

Figure 5: **Molecular dynamics simulation of galactokinase E174D and E174Q.** (a) and (b) The secondary structures for residues 174-179 and 231-240 respectively. (c) RMSF by residue for E174D and E174Q compared to WT galactokinase.

References

[1] C.E. Cardini, L.F. Leloir, Enzymic phosphorylation of galactosamine and galactose, Arch Biochem Biophys, 45 (1953) 55-64.
[2] H.M. Holden, J.B. Thoden, D.J. Timson, R.J. Reece, Galactokinase: structure, function and role in type II galactosemia, Cell Mol Life Sci 61 (2004) 2471-2484.
[3] P.A. Frey, The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose, FASEB J, 10 (1996) 461-470.
[4] A.A. Haritos, M.A. Rosemeyer, Purification and physical properties of hexokinase from human erythrocytes, Biochim Biophys Acta, 873 (1986) 335-339.
[5] D.J. Timson, The molecular basis of galactosemia - Past, present and future, Gene, 589 (2016) 133-141.
[6] J.L. Fridovich-Keil, J.H. Walter, Galactosemia, in: D. Valle, A.L. Beaudet, B. Vogelstein, K.W. Kinzler, S.E. Antonarakis, A. Ballabio (Eds.) The Online Metabolic and Molecular Bases of Inherited Diseases, McGraw-Hill, Place Published, 2008.
[7] D. Stambolian, Y. Ai, D. Sidjanin, K. Nesburn, G. Sathe, M. Rosenberg, D.J. Bergsma, Cloning of the galactokinase cDNA and identification of mutations in two families with cataracts, Nat Genet, 10 (1995) 307-312.
[8] D.J. Timson, R.J. Reece, J.B. Thoden, H.M. Holden, Galactokinase Deficiency, in: F. Lang (Ed.) Encyclopedia of Molecular Mechanisms of Disease, Springer, 2009, pp. 679-680.
[9] A.M. Bosch, H.D. Bakker, A.H. van Gennip, J.V. van Kempen, R.J. Wanders, F.A. Wijburg, Clinical features of galactokinase deficiency: a review of the literature, J Inherit Metab Dis, 25 (2002) 629-634.
[10] G.N. Donnell, W.R. Bergren, G. Perry, R. Koch, Galactose-1-phosphate in galactosemia, Pediatrics, 31 (1963) 802-810.
[11] T. Slepak, M. Tang, F. Addo, K. Lai, Intracellular galactose-1-phosphate accumulation leads to environmental stress response in yeast model, Mol Genet Metab, 86 (2005) 360-371.
[12] K. Lai, M.B. Boxer, A. Marabotti, GALK inhibitors for classic galactosemia, Future Med Chem, 6 (2014) 1003-1015.
[13] L. Liu, M. Tang, M.J. Walsh, K.R. Brimacombe, R. Pragani, C. Tanega, J.M. Rohde, H.L. Baker, E. Fernandez, B. Blackman, J.M. Bougie, W.H. Leister, D.S. Auld, M. Shen, K. Lai, M.B. Boxer, Structure activity relationships of human galactokinase inhibitors, Bioorg Med Chem Lett, 25 (2015) 721-727.
[14] S. Odejinmi, R. Rascon, M. Tang, H. Vankayalapati, K. Lai, Structure-activity analysis and cell-based optimization of human galactokinase inhibitors, ACS Med Chem Lett, 2 (2011) 667-672.
[15] M. Tang, S.I. Odejinmi, H. Vankayalapati, K.J. Wierenga, K. Lai, Innovative therapy for Classic Galactosemia - tale of two HTS, Mol Genet Metab, 105 (2012) 44-55.
[16] M. Tang, K. Wierenga, L.J. Elsas, K. Lai, Molecular and biochemical characterization of human galactokinase and its small molecule inhibitors, Chem Biol Interact, 188 (2010) 376-385.
[17] K.J. Wierenga, K. Lai, P. Buchwald, M. Tang, High-throughput screening for human galactokinase inhibitors, J Biomol Screen, 13 (2008) 415-423.
[18] H. Kristiansson, D.J. Timson, Galactokinases: potential biotechnological applications as biocatalysts, Curr Biotech, 1 (2012) 148-154.
[19] S.M. Howard, M.R. Heinrich, The anomeric specificity of yeast galactokinase, Arch

Biochem Biophys, 110 (1965) 395-400.
[20] D.J. Timson, R.J. Reece, Sugar recognition by human galactokinase, BMC Biochem, 4 (2003) 16.
[21] P. Thomas, E.M. Bessell, J.H. Westwood, The use of deoxyfluoro-D-galactopyranoses in a study of yeast galactokinase specificity, Biochem J, 139 (1974) 661-664.
[22] D. Hoffmeister, J. Yang, L. Liu, J.S. Thorson, Creation of the first anomeric D/L-sugar kinase by means of directed evolution, Proc Natl Acad Sci USA, 100 (2003) 13184-13189.
[23] D. Hoffmeister, J.S. Thorson, Mechanistic implications of *Escherichia coli* galactokinase structure-based engineering, ChemBioChem, 5 (2004) 989-992.
[24] J. Yang, L. Liu, J.S. Thorson, Structure-based enhancement of the first anomeric glucokinase, ChemBioChem, 5 (2004) 992-996.
[25] J. Yang, X. Fu, J. Liao, L. Liu, J.S. Thorson, Structure-Based Engineering of *E. coli* Galactokinase as a First Step toward In Vivo Glycorandomization, Chem Biol, 12 (2005) 657-664.
[26] J. Yang, X. Fu, Q. Jia, J. Shen, J.B. Biggins, J. Jiang, J. Zhao, J.J. Schmidt, P.G. Wang, J.S. Thorson, Studies on the substrate specificity of *Escherichia coli* galactokinase, Org Lett, 5 (2003) 2223-2226.
[27] H. Kristiansson, D.J. Timson, Increased promiscuity of human galactokinase following alteration of a single amino acid residue distant from the active site, ChemBioChem, 12 (2011) 2081-2087.
[28] M. McAuley, H. Kristiansson, M. Huang, A.L. Pey, D.J. Timson, Galactokinase promiscuity: a question of flexibility?, Biochem Soc Trans, 44 (2015) 116-122.
[29] L. Lee, S. Kinoshita, H. Kumagai, T. Tochikura, Galactokinase of *Bifidobacterium bifidum*, Agric Biol Chem, 44 (1980) 2961-2966.
[30] L.P. Conway, J. Voglmeir, Functional analysis of anomeric sugar kinases, Carbohydr Res, 432 (2016) 23-30.
[31] L. Caputi, M. Rejzek, T. Louveau, E.C. O'Neill, L. Hill, A. Osbourn, R.A. Field, A one-pot enzymatic approach to the *O*-fluoroglucoside of *N*-methylanthranilate, Bioorg Med Chem, 21 (2013) 4762-4767.
[32] J. Yan, X. Chen, F. Wang, H. Cao, Chemoenzymatic synthesis of mono- and di-fluorinated Thomsen-Friedenreich (T) antigens and their sialylated derivatives, Org Biomol Chem, 11 (2013) 842-848.
[33] L. Li, Y. Liu, W. Wang, J. Cheng, W. Zhao, P. Wang, A highly efficient galactokinase from *Bifidobacterium infantis* with broad substrate specificity, Carbohydr Res, 355 (2012) 35-39.
[34] L. Li, Y. Liu, T. Li, W. Wang, Z. Yu, C. Ma, J. Qu, W. Zhao, X. Chen, P.G. Wang, Efficient chemoenzymatic synthesis of novel galacto-*N*-biose derivatives and their sialylated forms, Chem Comm, 51 (2015) 10310-10313.
[35] M. Chen, L.L. Chen, Y. Zou, M. Xue, M. Liang, L. Jin, W.Y. Guan, J. Shen, W. Wang, L. Wang, J. Liu, P.G. Wang, Wide sugar substrate specificity of galactokinase from *Streptococcus pneumoniae* TIGR4, Carbohydr Res, 346 (2011) 2421-2425.
[36] Y. Zou, W. Wang, L. Cai, L. Chen, M. Xue, X. Zhang, J. Shen, M. Chen, Substrate specificity of galactokinase from *Streptococcus pneumoniae* TIGR4 towards galactose, glucose, and their derivatives, Bioorg Med Chem Lett, 22 (2012) 3540-3543.
[37] X. Zhao, Y. Zou, M. Xue, Z. Ma, S. Wang, P.G. Wang, M. Chen, A one-pot approach to bio-synthesize globotriose and its derivatives from simpler substrates, Eur J Med Chem, 80 (2014) 423-427.
[38] Y. Zou, M. Xue, W. Wang, L. Cai, L. Chen, J. Liu, P.G. Wang, J. Shen, M. Chen, One-pot

three-enzyme synthesis of UDP-Glc, UDP-Gal, and their derivatives, Carbohydr Res, 373 (2013) 76-81.
[39] M.M. Muthana, J. Qu, Y. Li, L. Zhang, H. Yu, L. Ding, H. Malekan, X. Chen, Efficient one-pot multienzyme synthesis of UDP-sugars using a promiscuous UDP-sugar pyrophosphorylase from *Bifidobacterium longum* (BLUSP), Chem Comm, 48 (2012) 2728-2730.
[40] L.A. Reinhardt, J.B. Thoden, G.S. Peters, H.M. Holden, W.W. Cleland, pH-rate profiles support a general base mechanism for galactokinase (*Lactococcus lactis*), FEBS Lett, 587 (2013) 2876-2881.
[41] J.B. Thoden, H.M. Holden, The molecular architecture of human N-acetylgalactosamine kinase, J Biol Chem, 280 (2005) 32784-32791.
[42] P. Bork, C. Sander, A. Valencia, Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases, Protein Sci, 2 (1993) 31-40.
[43] D.J. Timson, GHMP kinases - structures, mechanisms and potential for therpeutically relevant inhibition, Curr Enz Inhib, 3 (2007) 77-94.
[44] S.S. Krishna, T. Zhou, M. Daugherty, A. Osterman, H. Zhang, Structural basis for the catalysis and substrate specificity of homoserine kinase, Biochemistry, 40 (2001) 10810-10818.
[45] T. Zhou, M. Daugherty, N.V. Grishin, A.L. Osterman, H. Zhang, Structure and mechanism of homoserine kinase: prototype for the GHMP kinase superfamily, Structure, 8 (2000) 1247-1257.
[46] T.K. Harris, G.J. Turner, Structural basis of perturbed pKa values of catalytic groups in enzyme active sites, IUBMB Life, 53 (2002) 85-98.
[47] C.F. Megarity, M. Huang, C. Warnock, D.J. Timson, The role of the active site residues in human galactokinase: Implications for the mechanisms of GHMP kinases, Bioorg Chem, 39 (2011) 120-126.
[48] X. Chu, N. Li, X. Liu, D. Li, Functional studies of rat galactokinase, J Biotech, 141 (2009) 142-146.
[49] M. Huang, X. Li, J.W. Zou, D.J. Timson, Role of Arg228 in the Phosphorylation of Galactokinase: The Mechanism of GHMP Kinases by Quantum Mechanics/Molecular Mechanics Studies, Biochemistry, 52 (2013) 4858-4868.
[50] N.M. Goodey, S.J. Benkovic, Allosteric regulation and catalysis emerge via a common route, Nat Chem Biol, 4 (2008) 474-482.
[51] M. McAuley, D.J. Timson, Modulating Mobility: a Paradigm for Protein Engineering?, Appl Biochem Biotechnol, (2016) In press. doi: 10.1007/s12010-016-2200-y
[52] M.J. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson, Gaussian 09, revision A. 1, Gaussian Inc., Wallingford, CT, 2009.
[53] D. Case, T. Darden, T. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, M. Crowley, R. Walker, W. Zhang, AMBER 10 University of California, San Francisco, 2008.
[54] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, J Comp Chem, 25 (2004) 1157-1174.
[55] P.K. Weiner, P.A. Kollman, AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions, J Comp Chem, 2 (1981) 287-303.
[56] R.W. Pastor, B.R. Brooks, A. Szabo, An analysis of the accuracy of Langevin and molecular

dynamics algorithms, Mol Phys, 65 (1988) 1409-1419.
[57] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, J Chem Phys, 98 (1993) 10089-10092.
[58] J. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J Comp Phys, 23 (1977) 327-341.
[59] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J Mol Graph, 14 (1996) 33-38, 27-38.
[60] A. Bakan, L.M. Meireles, I. Bahar, ProDy: protein dynamics inferred from theory and experiments, Bioinformatics, 27 (2011) 1575-1577.
[61] D.J. Timson, R.J. Reece, Functional analysis of disease-causing mutations in human galactokinase, Eur J Biochem, 270 (2003) 1767-1774.
[62] W. Wang, B.A. Malcolm, Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis, BioTechniques, 26 (1999) 680-682.
[63] A. Platt, H.C. Ross, S. Hankin, R.J. Reece, The insertion of two amino acids into a transcriptional inducer converts it into a galactokinase, Proc Natl Acad Sci USA, 97 (2000) 3154-3159.
[64] B.L. Horecker, A. Kornberg, The extinction coefficients of the reduced band of pyridine nucleotides, J Biol Chem, 175 (1948) 385-390.
[65] L. Michaelis, M.L. Menten, Kinetics of invertase action, Biochem Z, 49 (1913) 333-369.
[66] D. Marquardt, An algorithm for least squares estimation of nonlinear parameters, SIAM J App Math, 11 (1963) 431-441.
[67] A. Cornish-Bowden, Fundamentals of Enzyme Kinetics, Portland Press, Place Published, 2004.
[68] D.J. Timson, Quantitative enzymology, Curr Enz Inhib, 11 (2015) 12-31.
[69] T.J. McCorvie, Y. Liu, A. Frazer, T.J. Gleason, J.L. Fridovich-Keil, D.J. Timson, Altered cofactor binding affects stability and activity of human UDP-galactose 4'-epimerase: implications for type III galactosemia, Biochim Biophys Acta, 1822 (2012) 1516-1526.
[70] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding, Anal Biochem, 72 (1976) 248-254.

Tables

Table 1: **Steady-state kinetic parameters of human galactokinase and variants.** For experimental conditions, see Materials and Methods.

Protein	$K_{m,ATP}$ (μM)	$K_{m,Gal}$ (μM)	k_{cat} (s^{-1})	$k_{cat}/K_{m,ATP}$ (s^{-1} mM^{-1})	$k_{cat}/K_{m,Gal}$ (s^{-1} mM^{-1})
Wild-type	4.0 ± 1.1	330 ± 70	1.7 ± 0.1	430 ± 110	10.1 ± 1.2
E174D	3.9 ± 1.1	220 ± 70	1.6 ± 0.1	340± 74	7.6 ± 1.9
R228M	9.4 ± 1.3	670 ± 130	2.3 ± 0.1	240 ± 30	3.3 ± 0.4
R228K	14.9 ± 2.8	310 ± 100	3.7 ± 0.1	250 ± 40	19.6 ± 2.9
R105M	8.3 ± 0.1	260 ± 130	0.7 ± 0.1	98 ± 9	2.7 ± 1.1

Values are shown ± the standard error as determined by the fitting procedure.Table 2: **Thermal stability of human galactokinase and variants.** T_m values were determined using differential scanning fluorimetry. Statistical significance was determined by ANOVA.

Protein	T_m (°C)
Wild-type	55.72 ± 0.34
R105M	55.42 ± 0.15^{ns}
E174D	55.13 ± 0.35^{ns}
E174Q	$52.50 \pm 0.76^{***}$
R228K	$51.30 \pm 0.23^{***}$
R228M	$51.68 \pm 1.33^{***}$

Values are the mean of three determinations and are shown ± the standard errors of these means. One-way ANOVA was used to determine if T_m was significantly different from that determined for the wild-type enzyme. ns, not significant ($p>0.05$); *** significant $p<0.01$.

Figure 1

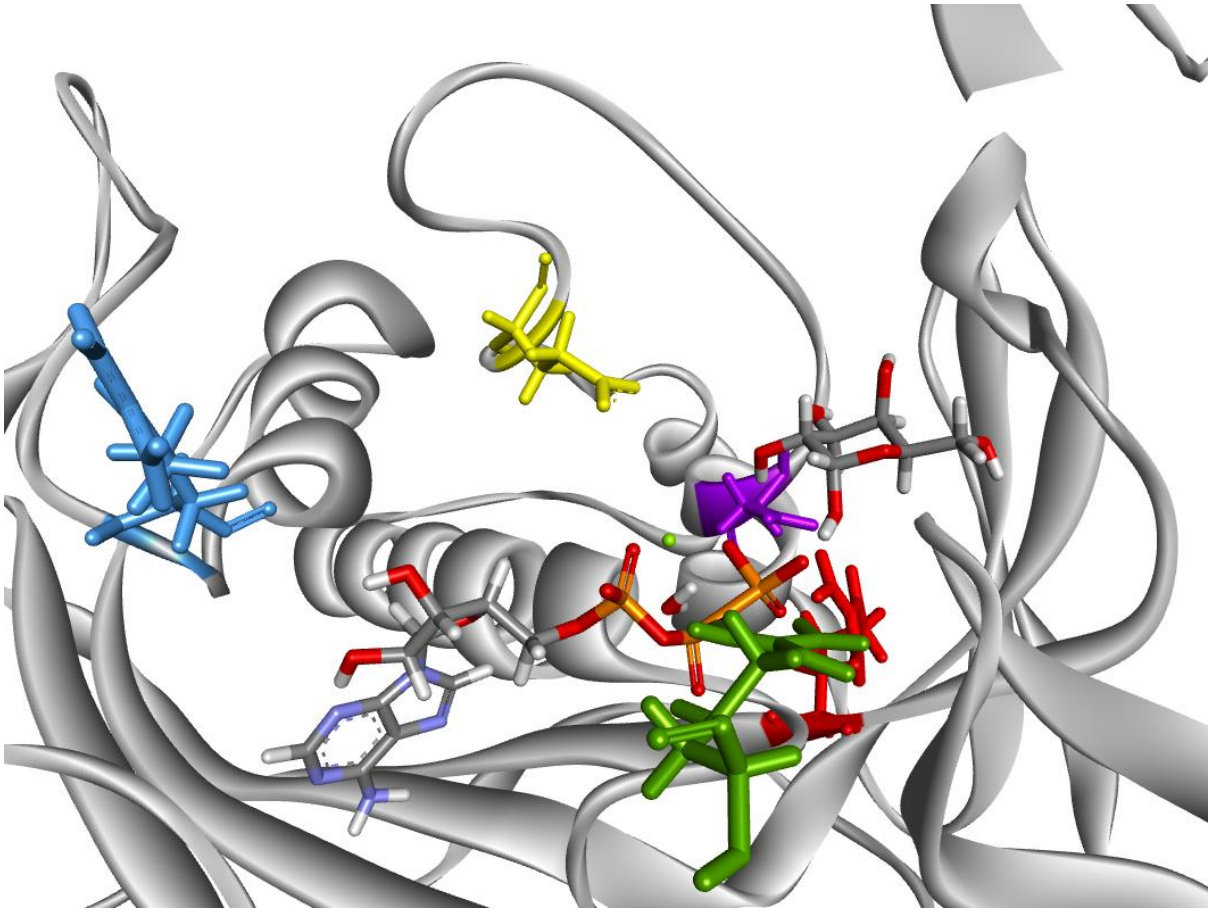
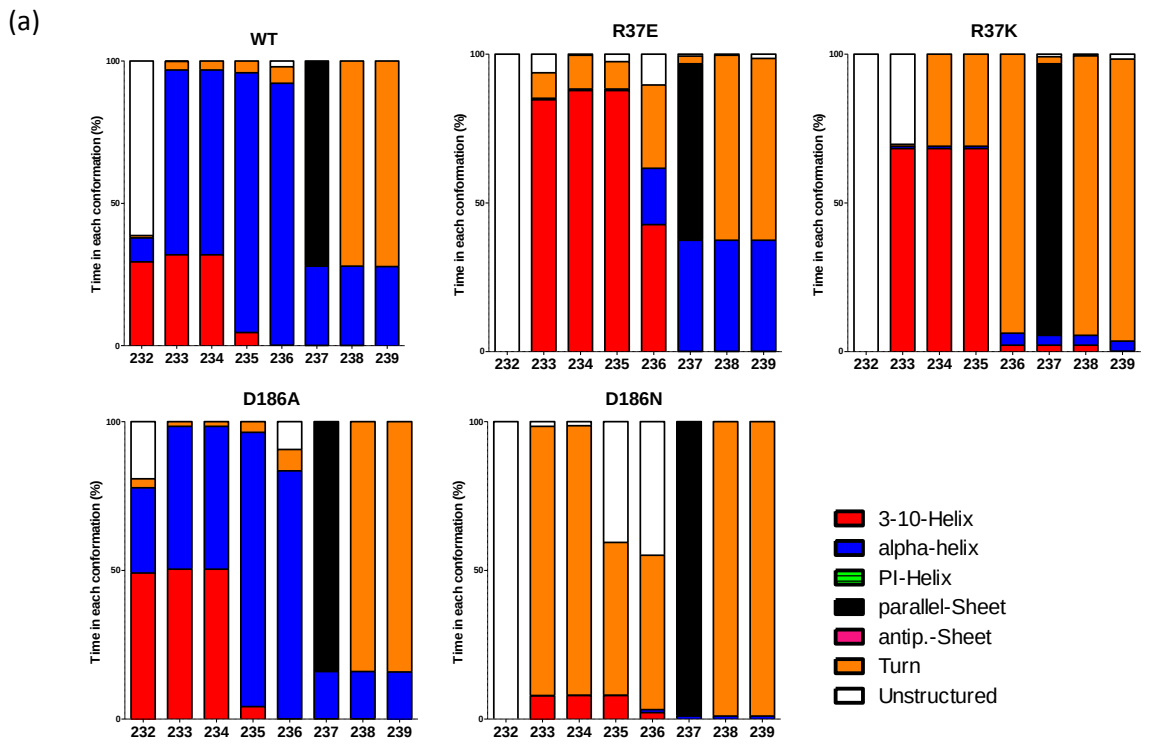
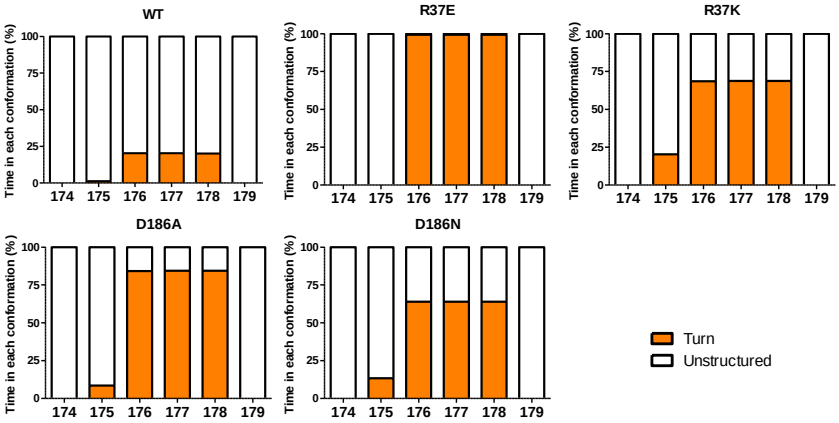


Figure 2


(a)
WT
R37E
R37K
D186A
D186N
Time in each conformation (%)
232 233 234 235 236 237 238 239
3-10-Helix
alpha-helix
PI-Helix
parallel-Sheet
antip.-Sheet
Turn
Unstructured


(b)
WT
R37E
R37K
D186A
D186N
Time in each conformation (%)
174 175 176 177 178 179
Turn
Unstructured

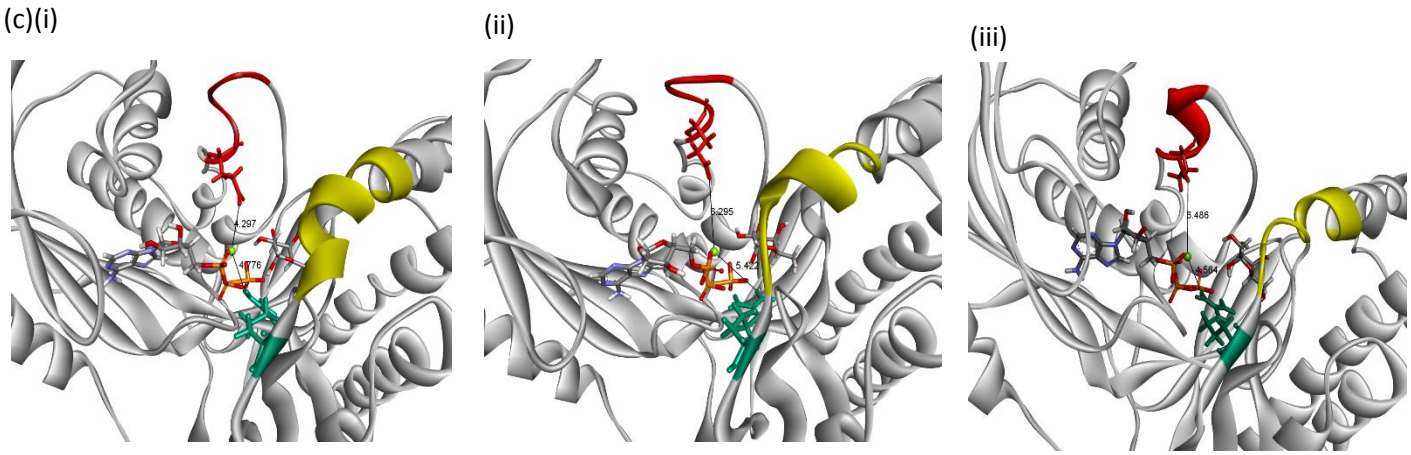

(c)(i)
(ii)
(iii)

Figure 3

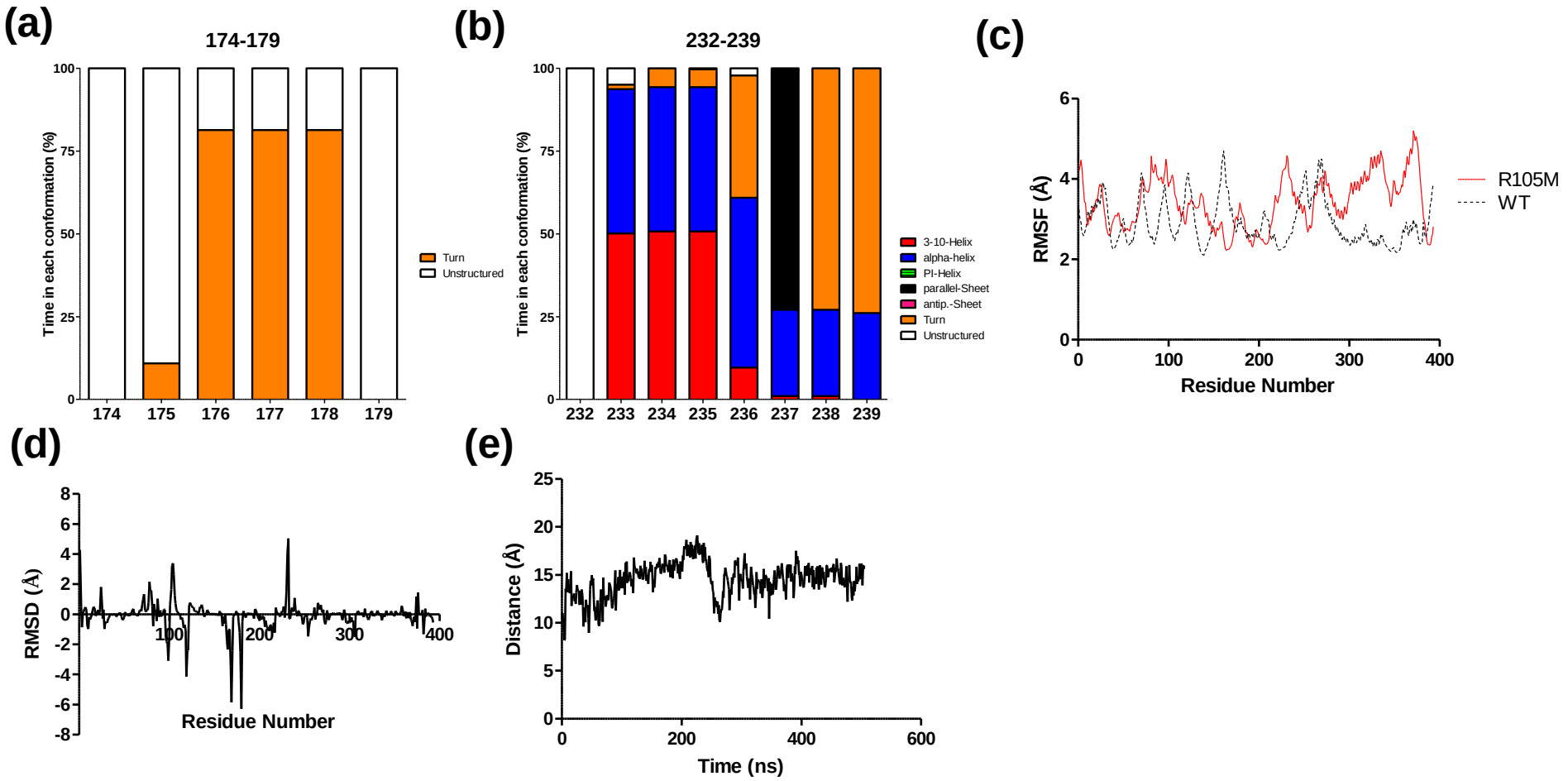

(a)
174-179
Time in each conformation (%)
100
75
50
25
0
174
175
176
177
178
179
Turn
Unstructured
(b)
232-239
Time in each conformation (%)
100
75
50
25
0
232
233
234
235
236
237
238
239
3-10-Helix
alpha-helix
PI-Helix
parallel-Sheet
antip.-Sheet
Turn
Unstructured
(c)
RMSF (Å)
6
4
2
0
0
100
200
300
400
Residue Number
R105M
WT
(d)
RMSD (Å)
8
6
4
2
0
-2
-4
-6
-8
100
200
300
400
Residue Number
(e)
Distance (Å)
25
20
15
10
5
0
0
200
400
600
Time (ns)

Figure 4

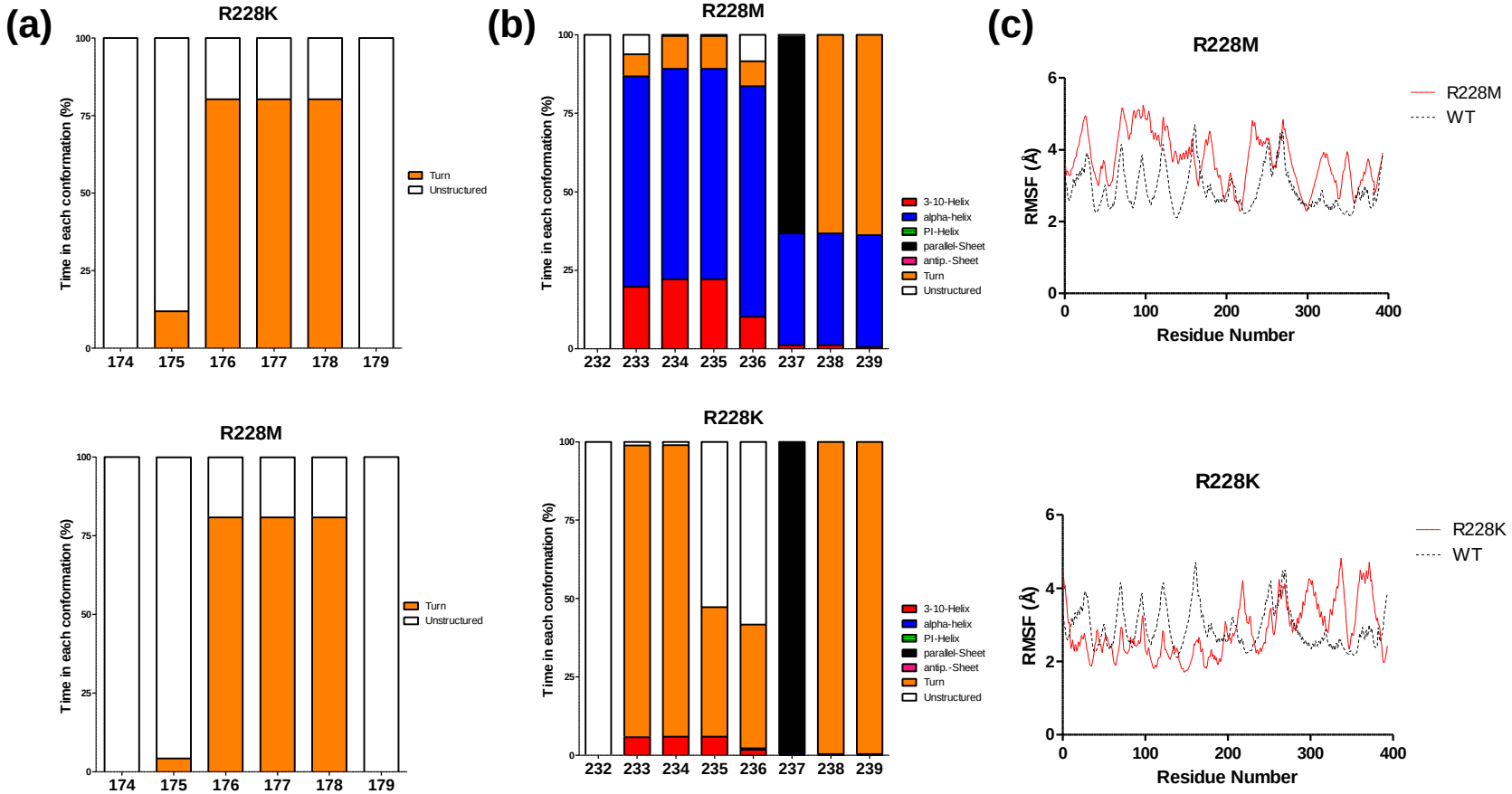
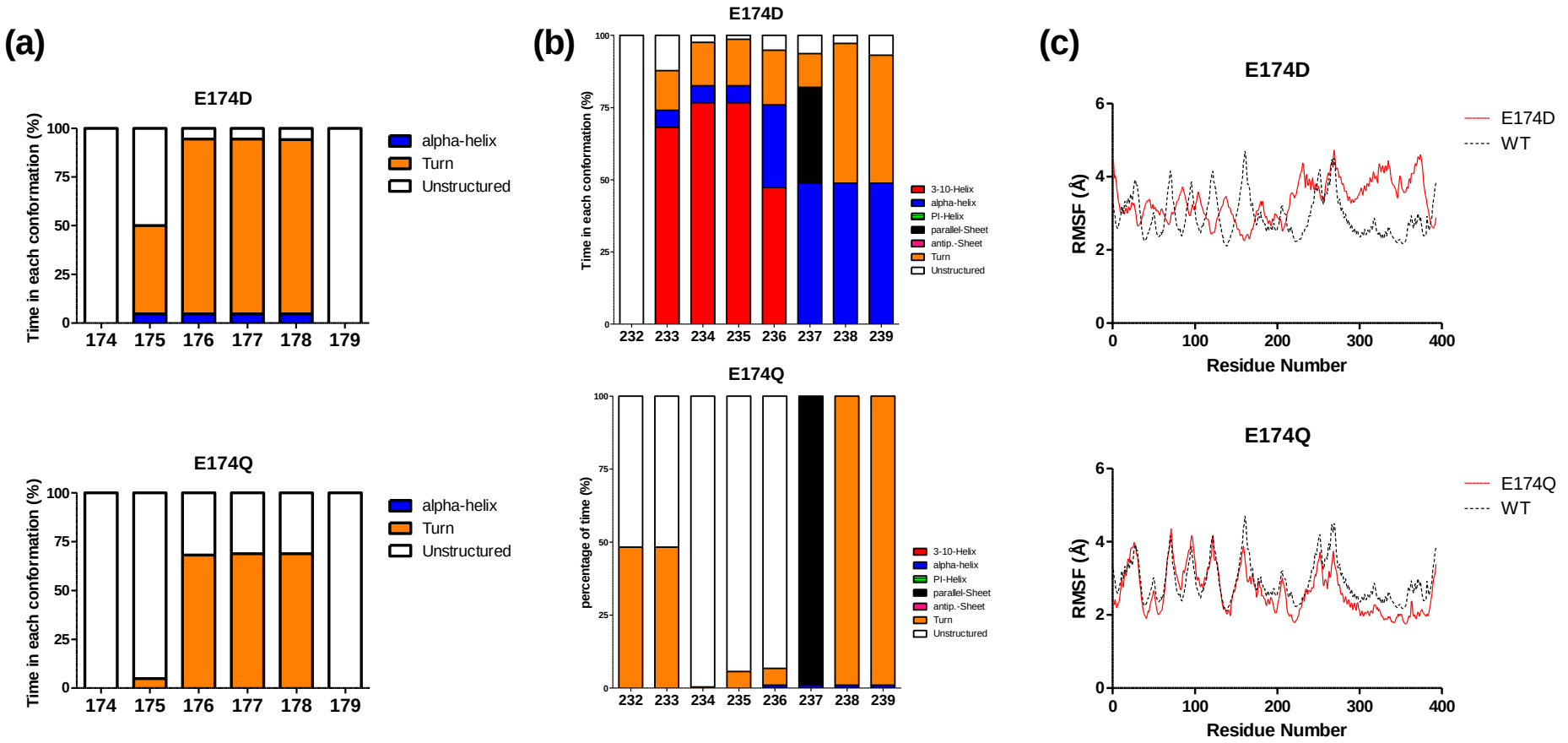
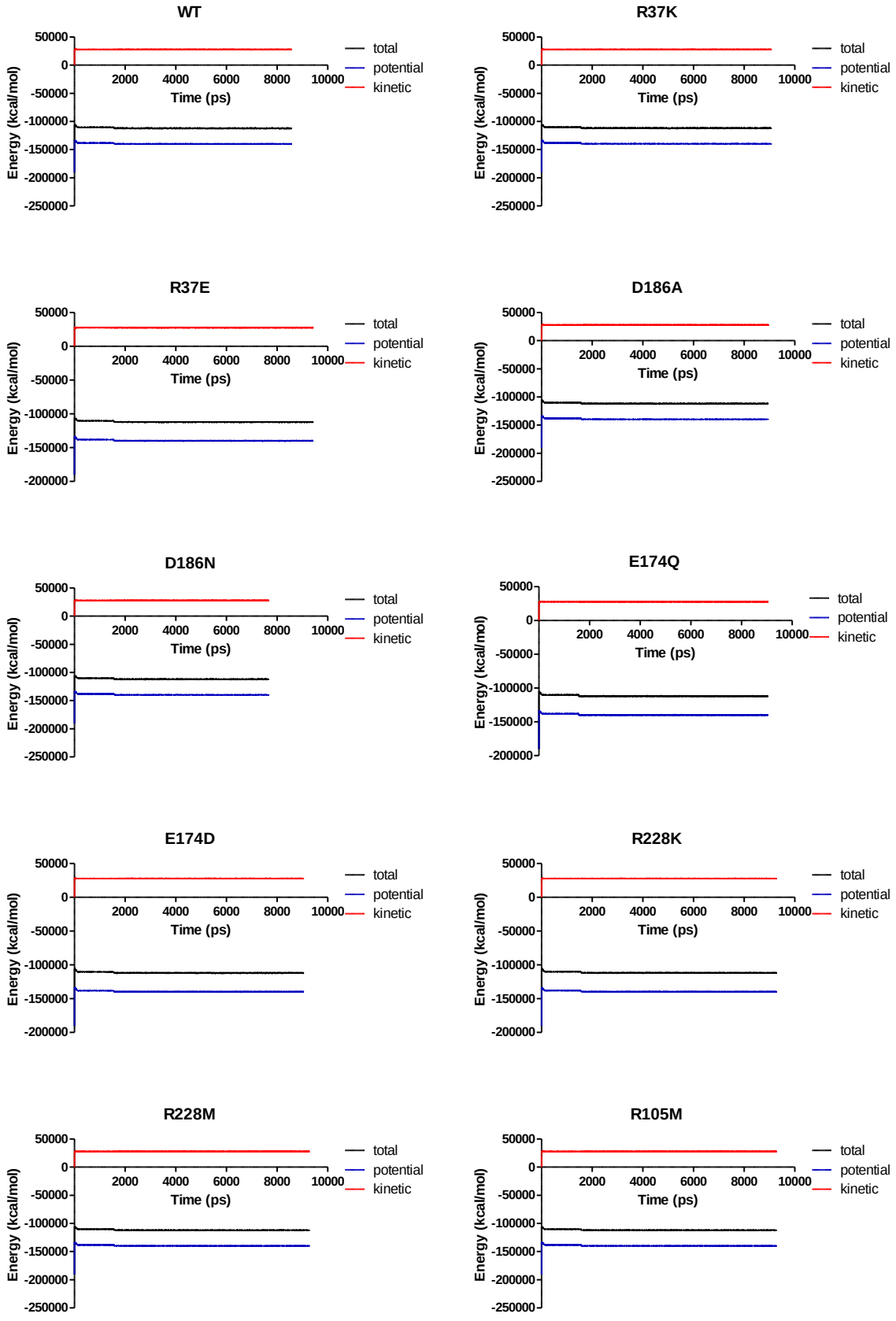

(a)
R228K
Time in each conformation (%)
100
75
50
25
0
174
175
176
177
178
179
Turn
Unstructured
R228M
Time in each conformation (%)
100
75
50
25
0
174
175
176
177
178
179
Turn
Unstructured
(b)
R228M
Time in each conformation (%)
100
75
50
25
0
232
233
234
235
236
237
238
239
3-10-Helix
alpha-helix
PI-Helix
parallel-Sheet
antip.-Sheet
Turn
Unstructured
R228K
Time in each conformation (%)
100
75
50
25
0
232
233
234
235
236
237
238
239
3-10-Helix
alpha-helix
PI-Helix
parallel-Sheet
antip.-Sheet
Turn
Unstructured
(c)
R228M
RMSF (Å)
6
4
2
0
0
100
200
300
400
Residue Number
R228M
WT
R228K
RMSF (Å)
6
4
2
0
0
100
200
300
400
Residue Number
R228K
WT

Figure 5


(a)
E174D
Time in each conformation (%)
100
75
50
25
0
174
175
176
177
178
179
alpha-helix
Turn
Unstructured
E174Q
Time in each conformation (%)
100
75
50
25
0
174
175
176
177
178
179
alpha-helix
Turn
Unstructured
(b)
E174D
Time in each conformation (%)
100
75
50
25
0
232
233
234
235
236
237
238
239
3-10-Helix
alpha-helix
PI-Helix
parallel-Sheet
antip.-Sheet
Turn
Unstructured
E174Q
percentage of time (%)
100
75
50
25
0
232
233
234
235
236
237
238
239
3-10-Helix
alpha-helix
PI-Helix
parallel-Sheet
antip.-Sheet
Turn
Unstructured
(c)
E174D
RMSF (Å)
6
4
2
0
0
100
200
300
400
Residue Number
E174D
WT
E174Q
RMSF (Å)
6
4
2
0
0
100
200
300
400
Residue Number
E174Q
WT

Supplementary Figure S1

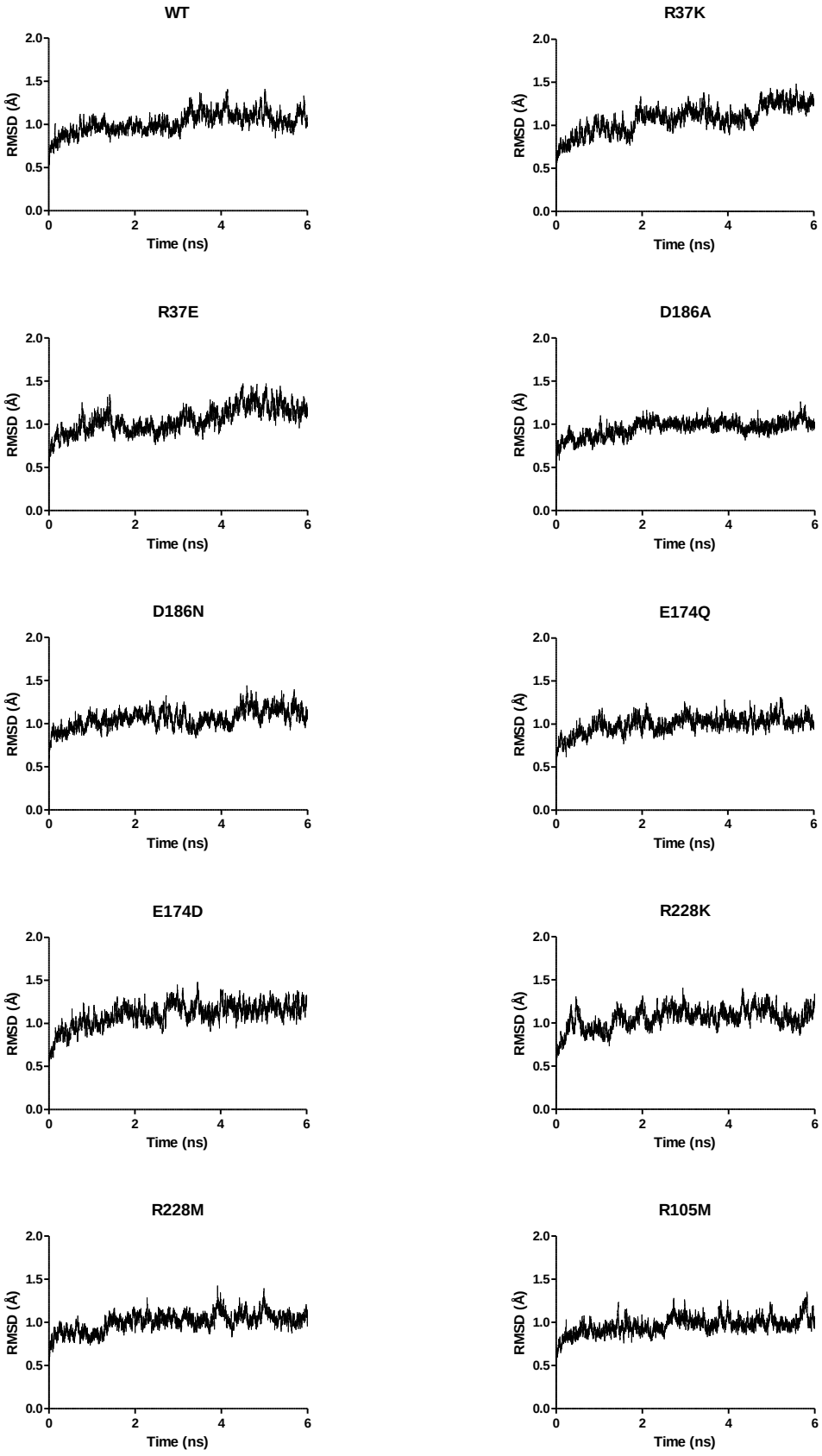
Click here to download Supplementary Material (for online publication): Suppl Fig S1 energy.docx



Supplementary Figure S1: Energy calculated per picosecond of simulation using the perl command in AmberTools.

Supplementary Figure S2

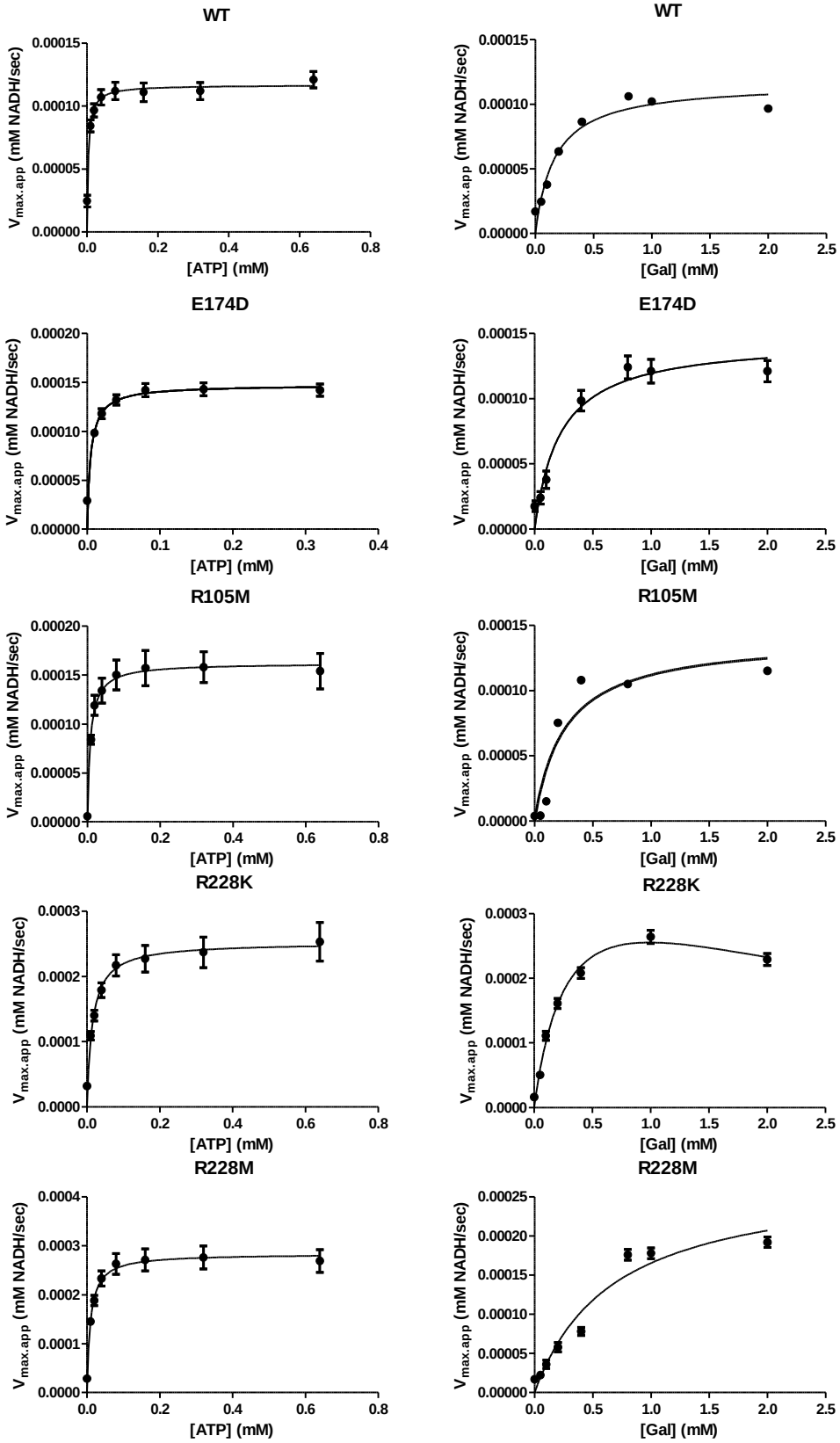
Click here to download Supplementary Material (for online publication): Suppl Fig S2 rmsd.docx



Supplementary Figure S2: Root mean square deviation (rmsd) measured throughout the trajectory compared to the initial minimised structure.

Supplementary Fig S3 (rev)

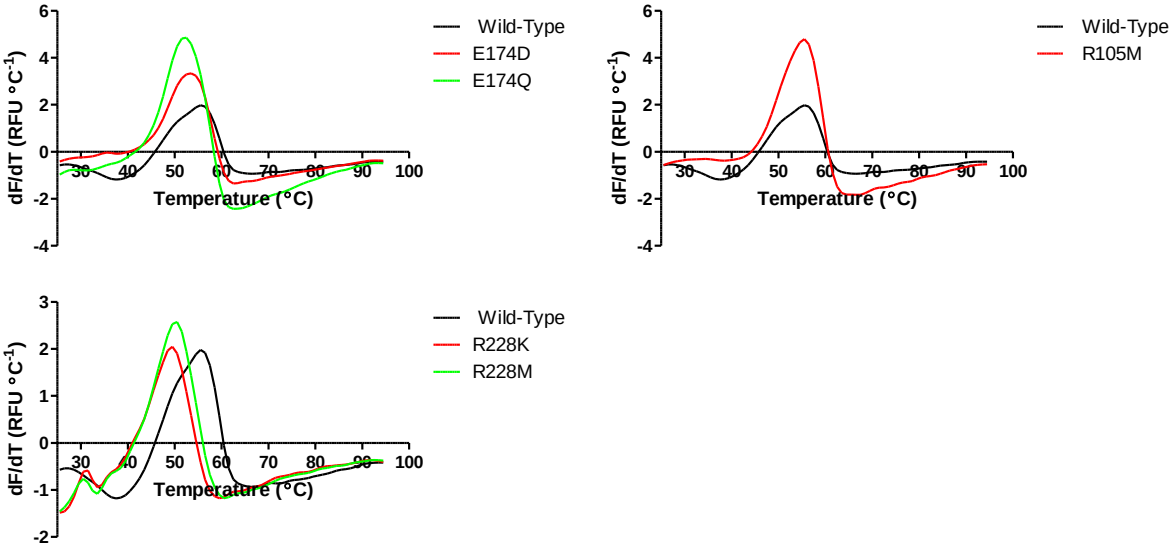
Click here to download Supplementary Material (for online publication): Suppl Figure S3 kinetics R1.docx



Supplementary Figure S3: Steady state kinetics for wild-type human galactokinase and each of the variants studied here. Points are the mean of three determinations and error bars the standard errors of these means. The lines were obtained by non-linear curve fitting as described in Materials and Methods.

Supplementary Figure S4

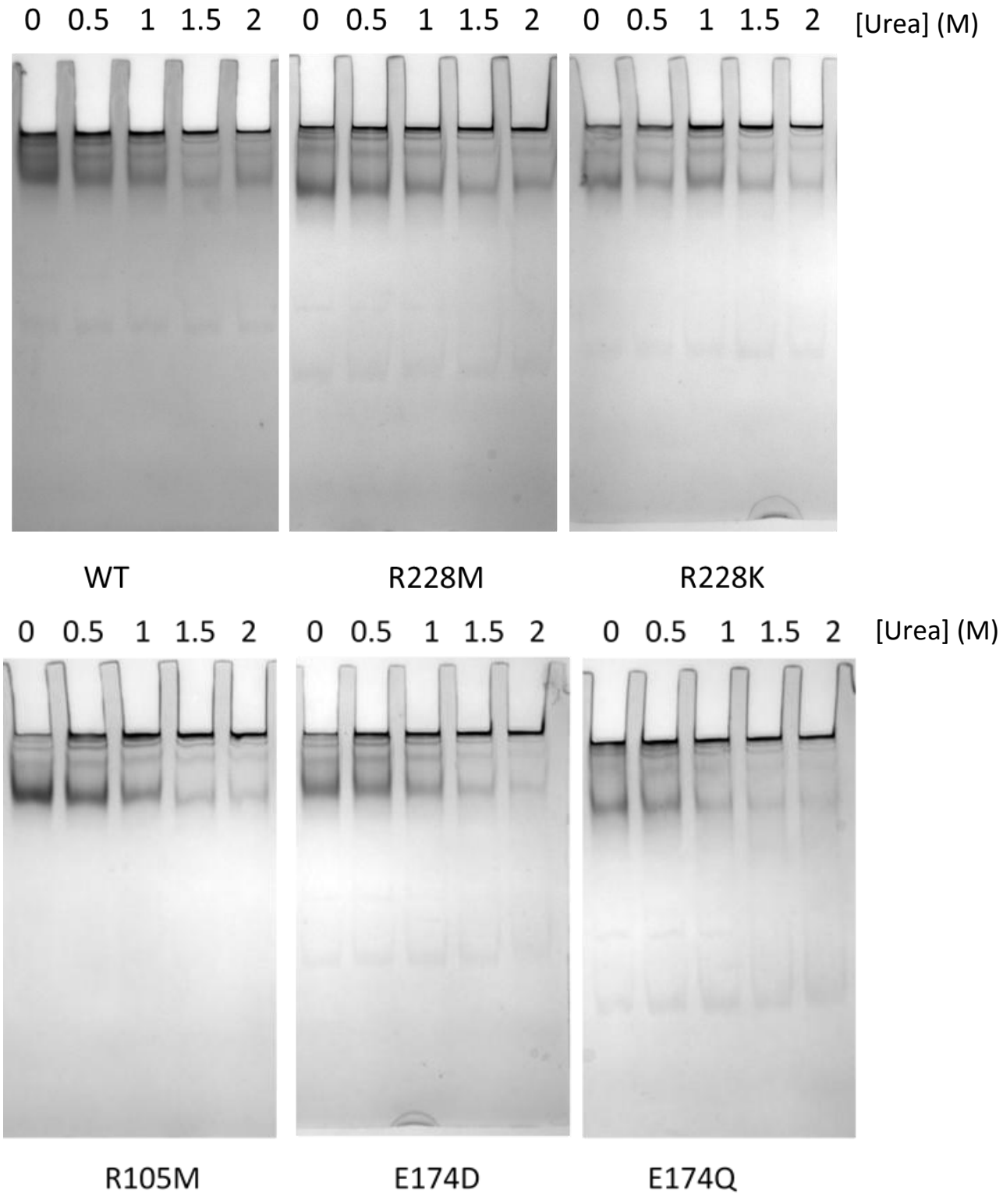
Click here to download Supplementary Material (for online publication): Suppl Figure S4 DSF.docx



Supplementary Figure S4: First derivative of differential scanning flourimetry data for human galactokinase and the variants studied here. Values for T_m are reported in Table 2.

Supplementary Figure S5

Click here to download Supplementary Material (for online publication): Suppl Fig S5 urea gels.docx



Supplementary Figure S4: Native PAGE with increasing concentration of urea. For details of conditions etc, please see Materials and Methods