Journal Pre-proof

Dynamic origins of substrate promiscuity in bacterial galactokinases

Margaret McAuley, Meilan Huang, David J. Timson

PII: S0008-6215(19)30420-3

DOI: <https://doi.org/10.1016/j.carres.2019.107839>

Reference: CAR 107839

To appear in: *Carbohydrate Research*

Received Date: 16 July 2019

Revised Date: 15 October 2019

Accepted Date: 15 October 2019

Please cite this article as: M. McAuley, M. Huang, D.J. Timson, Dynamic origins of substrate promiscuity in bacterial galactokinases, *Carbohydrate Research* (2019), doi: <https://doi.org/10.1016/j.carres.2019.107839>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier Ltd.

Motion in this α -helix helps control specificity at the active site

L. lactis

E. coli

Human

Dynamic origins of substrate promiscuity in bacterial galactokinases

Margaret McAuley¹, Meilan Huang² 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 catalyses the ATP-dependent phosphorylation of galactose and structurally related sugars. The enzyme has attracted interest as a potential biocatalyst for the production of sugar 1-phosphates and several attempts have been made to broaden its specificity. In general, bacterial galactokinases have wider substrate ranges than mammalian ones. The enzymes from *Escherichia coli* and *Lactococcus lactis* have received particular attention and a number of variants with increased promiscuity have been identified. Here, we present a molecular dynamics study designed to investigate the molecular causes of the wider substrate ranges of these enzymes and their variants with particular reference to protein mobility. Some regions close to the active site of the enzyme have different structures in the bacterial enzymes compared to the human one. Alterations known to increase the substrate range (e.g. Y371H in the *E. coli* enzyme), tend to alter the conformation of a key α -helical region (residues 216-232 in the *E. coli* enzyme). The equivalent helix in the human enzyme has previously been predicted to be altered in variants which affect catalytic activity or protein stability. This helix appears to be a key region in galactokinases from a range of species and may represent an interesting target for future attempts to broaden the specificity of galactokinases.

Keywords: GHMP kinase; molecular dynamics; biocatalyst; protein mobility; enzyme specificity; sugar 1-phosphate

1. Introduction

Galactokinase (EC 2.7.1.6) has attracted attention as a possible biocatalyst [1, 2]. This enzyme catalyses the site- and stereo-specific phosphorylation of α -D-galactose at position 1 [3-5]. The human enzyme is highly specific: α -D-galactose and the structurally related monosaccharides 2-deoxy- α -D-galactose, 2-fluoro-2-deoxy- α -D-galactose, α -D-galactosamine and α -D-talose are the only known substrates [6-9]. All these compounds differ from the physiological substrate at position 2 of the pyranose ring and human galactokinase does not appear to tolerate changes at any other position. Bacterial galactokinases have a more relaxed specificity. For example, in addition to 2-deoxy- α -D-galactose and α -D-galactosamine, the *Escherichia coli* galactokinase can also catalyse the phosphorylation of 3-deoxy- α -D-galactose and 6-deoxy- α -D-galactose (α -D-fucose) [10].

Galactokinase functions in the Leloir pathway of galactose metabolism, catalysing the first committed step of this pathway [3, 11, 12]. The reactions of this pathway are required to convert α -D-galactose into α -D-glucose 1-phosphate which is then isomerised into the glycolytic intermediate D-glucose 6-phosphate. In humans, mutations which lead to changes in the sequences of the Leloir pathway enzymes are associated with the inherited metabolic disease, galactosemia [13, 14]. This disease has a very wide range of manifestations from altered blood chemistry with little or no pathological consequences to death in childhood [15, 16]. Mutations in the gene encoding human galactokinase (*GALK1*) are associated with type II galactosemia (OMIM #230200) which is a milder form of the disease [17-20]. Knock-down of expression of the galactokinase gene inhibits the growth of liver cancer cells demonstrating that the enzyme may have potential as an anticancer drug target [21]. In some microbes (e.g the budding yeast, *Saccharomyces cerevisiae* and the gut bacterium *E. coli*), the Leloir pathway enzymes are under tight transcriptional control; they are only produced at high levels when galactose is present and, often, when more favourable carbon sources have been exhausted [22-24].

Galactokinases form part of the GHMP small molecule kinase family which is named after four members – galactokinase, homoserine kinase, mevalonate kinase and phosphomevalonate kinase [25, 26]. X-ray crystallographic studies show that these enzymes form compact globular, structures with two lobes and the active sites located in the cleft between these lobes [26-34]. Galactokinase is generally believed to function by an acid-base mechanism. The most commonly accepted mechanism involves the abstraction of a proton from the monosaccharide substrate by an aspartate residue acting as a base, assisted by an arginine residue (Table 1) [2, 35-39]. An alternative mechanism postulates that the aspartate increases the polarisation of the O-H bond facilitating the direct transfer of the phosphate group to the monosaccharide [2, 40]. Recent *in silico* and mutagenesis studies have identified a number of additional residues in human galactokinase which are likely to play a role in catalysis (Glu-174 and Arg-228; Table 1) [40, 41]. These studies, combined with molecular dynamics (MD) simulations of human galactokinase, have also identified two key region of

the protein (residues 174-179 and 231-240) which appear to control activity and specificity (Table 1) [41, 42]. The dynamic behaviour of these regions is altered in variants which have the active site aspartate or arginine altered and in many of the “return-to-consensus” variants which stabilise the overall structure of the enzyme (Table 1) [7].

The biotechnological interest in galactokinase stems from its ability to catalyse the formation of sugar 1-phosphates, in contrast to hexokinase (and related enzymes) which produce sugar 6-phosphates [1]. Sugar 1-phosphates are important in the process of *in vitro* glycorandomisation, which can be used to generate diverse, glycosylated drug-like molecules [43]. Such molecules have potential as future antibiotics. However, the high specificity of galactokinase is a barrier to its use as a biocatalyst. In response to this, several attempts have been made to broaden the specificity of the enzyme. Thorson and co-workers identified, through random mutagenesis and directed evolution, a residue which expands the substrate range of galactokinase from *E. coli*. This residue, Tyr-371, when changed to histidine broadened the substrate range to include D-galacturonic acid, L-altrose, 6-Amino-D-galactose and L-glucose (Table 1) [44]. Interestingly, this residue is distant from the active site, inferring that the change results in some form of structural and/or dynamic change which is propagated through the enzyme to the active site. The wild-type *E. coli* enzyme is intolerant of changes at C₄-OH in the monosaccharide substrate. However, alteration of an active site tyrosine (Tyr-223) to phenylalanine or tryptophan enabled the phosphorylation of 4-deoxy-D-galactose [45]. Further enhancement was possible by alteration of a methionine residue (Met-173). This residue lies close to the active site and the authors noted that the equivalent residue in *Lactococcus lactis* galactokinase was a leucine [27, 46]. The *L. lactis* enzyme has a greater tolerance of changes at C₄-OH [47]. Thus, it was reasoned that combining changes at Met-173 and Tyr-371 might further expand the substrate range. The M173L/Y371H variant has an impressive substrate range: at least 28 monosaccharides can be phosphorylated by this enzyme with reasonable activity (Table 1) [46]. The substrate range of the *L. lactis* galactokinase can be further expanded by alteration of Tyr-385 (equivalent to Tyr-371 in the *E. coli* enzyme) to histidine (Table 1) [47]. This suggests that this tyrosine residue may control specificity in galactokinase from a variety of sources. In the human enzyme, the equivalent tyrosine is residue 379. Alteration of this residue to histidine dramatically reduced the enzyme's activity with its natural substrates, but did not broaden the specificity. However, alteration of this tyrosine to other residues (notably cysteine, serine, lysine, arginine and tryptophan) did broaden the specificity to include D-mannose and D-fructose (Table 1) [48]. This increased promiscuity came at the price of reduced activity. Typically, turnover numbers (k_{cat}) were reduced 10 to 100-fold in comparison to the wild-type enzyme with α -D-galactose. Alteration of the methionine residue equivalent to the Met-173 in the *E. coli* enzyme did not broaden the specificity of the human enzyme [7].

Taken together these studies suggest two key differences between mammalian and bacterial galactokinases.

First, bacterial galactokinases have a naturally wider substrate range. Second, the bacterial enzymes when engineered for greater promiscuity retain a higher level of activity. We have previously undertaken a number of *in silico* and experimental studies on human galactokinase [49]. These aimed to understand its mechanism and to engineer the enzyme to make it a more viable biocatalyst. Here, we present a molecular dynamics study on the two most studied bacterial galactokinases in order to provide explanations for the differences in the enzymes from these two different Kingdoms of life.

2. Results and Discussion

2.1 Comparison of the bacterial and human galactokinase structures

Although experimental structures exist for human and *L. lactis* galactokinase, no structure has been published of the *E. coli* enzyme. To address this, we first built a homology model for this protein (Figure 2). As expected, the predicted structure is highly similar to the *L. lactis* one (rmsd of 1.226 Å over 1878 equivalent atoms) and to the human enzyme (rmsd of 1.203 Å over 1673 equivalent atoms). The model has 96.4% of residues in the allowed region of the Ramachandran plot (compared to 98% expected), 2.6% in the allowed region (2% expected) and 0.9% in the outlier region (data not shown). Analysis with ProSA resulted in a Z-score of -9.81, which lies clearly in the range of experimentally determined structures of similar sizes with no regions showing positive energy scores (window size of 40 residues). Validation with PSVS showed no close contact, no bond lengths deviating by more than 6.0 rmsd of the standard values and no bond angles deviating by more than 6.0 rmsd of the standard values. The model was, therefore, judged to be of acceptable quality to proceed with further analysis.

The *L. lactis* galactokinase protein sequence is 36% identical (51% similar) to the human one. However, the structures differ at several points (Figure 3a). The ATP binding cleft is wider in the *L. lactis* enzyme. However, this may be due to the crystal structures capturing different stages in the catalytic cycle which is expected to involve some opening and closing of this cleft as substrates enter and products leave. Note that the *L. lactis* enzyme was crystallised in the presence of galactose, whereas the human enzyme structure has ATP and galactose bound at the active site. (The crystal structure had the slow-hydrolysing ATP analogue AMP.PNP bound; however, this was altered as part of our preparation of the structure for MD studies.) Residues 151-168 of the *L. lactis* enzyme form an α -helix whereas the equivalent region in the human enzyme adopts an elongated, more random coil structure. Residues 356-371 form a longer α -helix compared to the equivalent region of the human protein. Two other groups of residues (81-93 and 216-224) are also predicted to adopt different conformations in the two proteins. These adopt random coil structures in both proteins, but with markedly different folds. The *L. lactis* enzyme also lacks a flexible region (residues 1-15) which is present at the N-terminus of the human enzyme. This region is not expected

to influence the specificity or activity of the enzyme.

E. coli galactokinase has 45% protein sequence identity (57% similarity) with the human enzyme. However, as with the *L. lactis* enzyme there are some key differences (Figure 3b). Residues 151-167 are more likely to adopt a helical conformation in the *E. coli* enzyme compared to the human one. Residues 64-83 and 89-103 are displaced compared to the equivalent ones in the human enzyme. These regions are close to the ATP binding site and so these differences may have an impact on catalysis. Region 342-362 is shortened and displaced in the *E. coli* enzyme compared to the human one. This region is close to the residue which affects substrate specificity (Tyr-371) and so may impact the overall flexibility of this part of the protein. Like the *L. lactis* galactokinase, the *E. coli* enzyme lacks ten, largely unstructured N-terminal residues which are present in the human protein.

2.2 Comparison of variant structures with wild-type galactokinases

Comparison of the equilibrated, predicted structures of *E. coli* galactokinase variants revealed some interesting features (Figure 4). The double variant M173L/Y371H was more similar to the wild-type structure than the single variant Y371H (rmsd of 1.960 Å compared to 2.611 Å). In Y371H, the α -helix comprising residues 90-104 is shortened and the β -sheet alongside it is displaced. This region lies on the surfaces of the protein and forms the outside of the ATP binding pocket. The loop 331-340, which lies beside the sugar binding pocket is also displaced. It is possible that this displacement is one cause of the increased substrate range seen in this variant. An α -helix (151-165) close to the active site is also displaced, which may also be a cause of the expanded substrate range. Another α -helix (214-224) is shortened (Figure 4a). Alteration of the equivalent helix in human galactokinase has also been predicted in some variants (e.g. R37K, R37E, E174Q, E174D, D186N, D186A and R228K) [41]. In the M173L/Y371H double variant, some of these differences are also seen: the α -helix close to the active site (151-165) is displaced, as is the 331-340 loop. The shortening of the 214-224 α -helix is also predicted in this variant, highlighting the critical importance of this region. The alteration of a bulky methionine residue to the smaller leucine side chain also creates more space in the active site potentially enabling the binding of a wider range of sugar substrates. Thus, this variant retains the alterations close to the sugar binding site present in Y371H and also increases the available space in the active site (Figure 4b).

The Y385H variant in *L. lactis* galactokinase is more similar to the wild-type enzyme (rmsd of 1.38 Å) than the equivalent variant of the *E. coli* enzyme (Figure 5). There is a loss of a short α -helix (residues 80-86) which displaces a loop close to the ATP binding site. A loop close to the sugar binding site (338-347) is also displaced (Figure 5). This is the structurally equivalent region to the 331-340 loop in *E. coli* galactokinase and it can be assumed that this alteration has similar effects on specificity in the two bacterial enzymes.

2.3 Dynamic effects on flexibility

It has been hypothesised that the alteration of the tyrosine residue affects the distant active site of human galactokinase via altered the flexibility of the substrate site thus increasing promiscuity [48]. To examine if this hypothesis might also be relevant to the bacterial galactokinases, the flexibility of the proteins was determined from the MD trajectories (Figure 6a). The *E. coli* Y371H variant and, to a lesser extent, M173L/Y371H showed increases in root mean squared fluctuation (RMSF) around residues 217-255 (Figure 6a). This region includes the α -helix adjacent to the active site which commonly had altered mobility in previously studied variants of the human enzyme [7, 41, 42]. This change to the RMSF of this region is common to the *L. lactis* Y385H variant but considering that this enzyme is natively more promiscuous it is likely that this region is already more flexible in the wild-type enzyme. Displacement of the loop, 338-347 region in *L. lactis* and 331-340 in *E. coli*, occurs alongside an increase in RMSF in all variants compared to their WT counterparts, suggesting that this may be one reason the active site can accommodate a wider range of monosaccharides.

Principal component analysis was then carried out to determine the directionality of changes to flexibility (Figure 6b). The most notable finding from this analysis was that a region close to the active site (the 232-240 helix) is altered in all promiscuous variants. It is likely that the alteration to the β -strand (269-373 in the *E. coli* enzyme) resulting from variation of Tyr-371 impacts the packing of this β -sheet resulting in a more flexible local region including the second strand of the β -sheet which leads to displacement of the loop close to the sugar binding pocket (Figure 6c).

2.4 Propensity for secondary structure formation in key regions

Analysis of secondary structure of a key region (residues 214-224) identified in the bacterial galactokinases highlighted the loss of structure at the C-terminal end of this helix in *E. coli* galactokinase variants. Both Y371H and M173L/Y371H show reduced likelihood of forming secondary structure in residues 217, 218 and 219. This would result in increased flexibility of this region (Figure 7a). The *L. lactis* Y385H variant shows loss of a turn at residue 222 but formation of a turn at residues 220-221 for 44% of the trajectory analysed. It appears that displacement of the loop 331-340 is the result of this alteration to the secondary structure. In *E. coli* galactokinase, the structurally equivalent turn is at residues 333 and 334 and is increased in propensity to form a turn from 5% in the wild-type to 57% in Y371H and 28% in M173L/Y371H (Figure 7b). The propensity for the formation of antiparallel β -sheet by residue 332 is reduced from 18% in the wild-type enzyme to 4% Y371H and 3% in M173L/Y371H. Both these changes in secondary structure are likely to

result in the displacement of the loop close to the sugar binding pocket. Interestingly, the results for the *L. lactis* enzyme differed. The propensity for formation of turn at residues 339 and 340 is decreased from 16% in the wild-type enzyme to 9% in Y385H. However, the propensity for formation of antiparallel β -sheet by residue 338 is decreased from 55% to 33% in this variant, which may contribute to displacement of the active site loop.

2.5 Dynamic effects on promiscuous variants of human galactokinase

We have demonstrated previously that the variants of human galactokinase with increased substrate range (e.g. Y379W) are predicted to disrupt an α -helix around residues 232-234 (equivalent to residues 219-221 in the *E. coli* enzyme). In addition, in Y379W, a short piece of α -helix is created around residues 176-179 (equivalent to residues 164-167 in the *E. coli* enzyme) [7]. There is no clear pattern in the overall flexibility of the five more promiscuous variants of the human enzyme (Y379C, Y379S, Y39K, Y379R and Y379W). Y379K was predicted to be less flexible across much of the length of the polypeptide chain, whereas Y379C was predicted to be generally more flexible than the wild-type (Supplementary Figure S1). The Y379W variant is quite similar in flexibility to the wild-type for the first 200 residues. After this point, it shows three peaks of considerably increased flexibility centred on residues 217, 298 and 337. It is also more flexible at the C-terminus (residues 359-367). Similar results, albeit to a lesser extent, were observed in Y379H (Supplementary Figure S1). This variant was not shown to have an expanded substrate range in our previous work [48].

3. Conclusions

The role of mobility in protein function, including catalysis, is well-established [50, 51]. Our MD simulations highlight how alterations in the primary sequence of bacterial galactokinases result in changes to the proteins' dynamic behaviour with consequent effects on catalysis. They further highlight the critical importance of a helical region in the protein: residues 216-232 in *E. coli* galactokinase, 226-242 in *L. lactis* and 231-245 in the human enzyme. Our previous work has implicated this region in catalysis and stability [7, 41, 42]. It seems likely from the work presented here that it also plays a role in controlling the specificity of the enzyme from a variety of species.

The greater catalytic promiscuity of some bacterial galactokinases most likely arises from several causes. The presence of a leucine (rather than a methionine) residue close to the active site (Leu-182 in the *L. lactis* enzyme) reduces steric hindrance and facilitates the binding of substrates with a wider range of groups at C₄

of the monosaccharide [47]. In addition, dynamic changes to the structures of regions close to the active site (as identified here) also play a role. These most likely result in greater plasticity of the active site, enabling a wider range of monosaccharide substrates to be accommodated. The cause of substantially reduced activity in the human enzyme when alterations are made to Tyr-379 is less clear. Changes to the equivalent residue in the two bacterial enzymes did not greatly reduce their activity [44, 47]. It should be noted that the human enzyme has been shown to be intolerant of single amino acid changes in many cases. Several of the variants associated with type II galactosemia result in insoluble protein during recombinant expression in *E. coli*; similar results have been observed following alteration of residues in the protein's active site [8, 52, 53]. MD studies on one of the disease-associated variants (P28T) suggested that local unfolding in the region around the affected residue results in destabilisation of the overall protein structure [54]. Two wider studies of galactokinase variants suggested that disease-associated forms of the human enzyme were generally likely to be less stable than the wild-type, but there was no clear pattern in terms of predicted structural changes [55, 56].

Overall, our results suggest that the search for more promiscuous variants of galactokinase may be more successful if they focus on bacterial enzymes rather than ones from higher eukaryotes. This is supported by the discovery in recent years of a variety of bacterial galactokinases with unusual specificities [57-63]. The results also suggest that further enhancement of the enzyme's substrate range may be possible through focussing on the region which was repeatedly identified in this and previous studies, namely the α -helix comprising residues 216-232 (*E. coli* numbering). For example, alteration of residues in this region to glycine or serine (the two residues which result in the greatest flexibility of the polypeptide chain [64]) may be useful. Alteration of Arg-216, the first residue of the helix to glycine or serine, would be likely shorten the helix and extend the random coil region comprised by residues Asn-211 to Lys-215. The helix is predicted to bend slightly, approximately in the middle around residues Asn-224 or Thr-225. These residues also lie close to the galactose binding site. Targeting these residues may cause the helix to "melt", converting it to flexible joint. These changes are likely to increase the ability of the active site to accommodate a wider range of monosaccharide structures. However, combining rational (structure-based) and random mutagenesis with *in silico* studies to monitor the effects of changes on global and local mobility is likely to yield the best results.

4. Materials and Methods

4.1 Homology model of *E. coli* galactokinase

In the absence of an experimental structure for *E. coli* galactokinase, a homology model was generated using SWISSMODEL [65-68]. The best alignment was found to be human galactokinase chain A (PDB: 1WUU;

[28]) with a root mean squared deviation (RMSD) of 2.5 Å, 44.2% sequence identity and 97% coverage.

Ramachandran plots were constructed using RAMPAGE

(<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>) [69]. ProSA was used to assess the quality of the model compared to experimentally determined structures [70, 71]. The protein structure validation software suite (PSVS) was used to further validate the model [72].

4.2 Molecular dynamics (MD) simulations and their analysis

The MD simulations of the *L. lactis* galactokinase structure were based on the crystal structure of this enzyme (PDB: 1PIE; [27]). The homology model of *E. coli* galactokinase was used to simulate the dynamics of this enzyme. The preparation and MD simulations of human galactokinase were based on the crystal structure of the enzyme (PDB: 1WUU; [28]). The structure was modified to alter selenomethionine residues back to methionine, change AMP.PNP to ATP and to fill gaps as previously described [7, 40, 41].

All MD simulations were carried out using the Amber 10 package and the Amber Parm 99 forcefields [73]. 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, the number of these required varied depending on the protein and variant being studied. Minimisation and simulation steps were carried out as described previously except simulations were run for 7 ns [40]. An initial minimisation step was carried out with Cartesian restraints (imin1). Energy minimisation was carried out for 2500 cycles, 1250 of conjugate gradient method and 1250 of deepest decent, at constant volume with residues fixed using harmonic potential with a 10 Å cut-off for non-bonded atoms. A second minimisation step was carried out on the entire system using the same settings as imin1, except that no restraints were used. The system was then heated from 0 to 300K using the script md 1 using Langevin dynamics and a collision frequency of 1 ps⁻¹. Bonds with hydrogens were constrained and bond interactions with hydrogens omitted, since the simulation was run at constant volume. These settings were used for 30,000 steps with a time step of 0.002 ps. After heating the system was subjected to an equilibration step (md2). This step of the simulation was run for 750,000 steps with a time interval of 0.002 ps at 300 K, constant pressure, using Langevin dynamics and constraints on bonds with hydrogen. The remainder of the simulation was considered 'production simulation' and was run using repeats of the same script (md3). This script was the same as md2 except there were no restraints on the atoms and there were 1,000,000 steps with a time step of 0.001 ps.

Physical characteristics of simulations were monitored using perl encoded in Amber and resulting graphs produced using GraphPad Prism version 5.03 for Windows (GraphPad Software, San Diego CA, USA). Cluster

analysis was performed using the 'clust' command of ptraj, in AmberTools, coordinates from 2 ns of simulation trajectory were clustered according to RMSD of α - carbons into 5 groups and an average and representative pdb file created. Cluster structures were visualised using Biovia Discovery Studio and molecular overlay used to compare and highlight altered regions. This software was also used to generate the images shown in the Figures. The proportion of time spent in each secondary structure was determined using the 'secstruct' command and a sample of the trajectory, every 10 frames. The summaries were then converted to percentages in Microsoft Excel and graphs produced using GraphPad prism. RMSD and RMSF calculations were also carried out using ptraj using the commands 'rms' and 'atomicfluct' respectively. RMSD was calculated over the entire production trajectory. RMSF was calculated over the 2 ns used for cluster analysis so that sampling was of the same timeframe. VMD software was used to visualise trajectories, the timeline plugin was used to visualise secondary structure and the prody plugin of Normal mode wizard was used to carry out principal component analysis (PCA) on 1 ns of stable trajectory [74, 75].

Acknowledgements

MM thanks the Department of Employment and Learning, Northern Ireland (DELNI, UK) for a PhD studentship. MH acknowledges support from InvestNI (grant number: RD0314092). We thank Queen's University Belfast for access to its high performance computer cluster on which the simulations were conducted.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Figure legends

Figure 1: Equilibration of molecular dynamics trajectories. RMSD over the entire trajectory of bacterial galactokinases and their more promiscuous variants. RMSD values were calculated using ptraj in AmberTools by comparing protein structure to initial structure of each variant.

Figure 2: The predicted structure of *Escherichia coli* galactokinase. A homology model of *E. coli* galactokinase was created using SWISS-MODEL and alignment with human wild-type galactokinase chain A which had the best protein sequence alignment with 44.2% sequence identity and 97% coverage. Helices are represented by spirals and sheets by slabs. The protein is coloured in a rainbow representation with red at the N-terminus and violet at the C-terminus.

Figure 3: Comparison of the bacterial and human galactokinase structures. (a) Top, molecular overlay of *L. lactis* galactokinase compared with the human enzyme. This was carried out using Biovia Discovery Studio; the modal cluster of the *L. lactis* simulation (c3) was compared with that of the human enzyme (c0). Human galactokinase is shown in grey and the *L. lactis* structure in red. Bottom, Regions of structural difference between human and *L. lactis* homologues of galactokinase. *L. lactis* in pink/red and WT in grey/yellow; numbering refers to *L. lactis* protein. Left to right: altered regions close to ATP binding site; region 151-168 adjacent to the active site; altered helix 342-362; the tyrosine residue previously shown to affect substrate promiscuity alongside the lack 15 residues at the N-terminus when compared to human enzyme. (b) Top, molecular overlay of *E. coli* galactokinase modelled structure with the human structure. The modal cluster of the *E. coli* simulation (c3) was compared with that of the human enzyme (c0). The human enzyme is shown in grey and *E. coli* one in red. Bottom, Regions of structural difference between human and *E. coli* galactokinases with the *E. coli* enzyme shown in pink/red and the human one in grey/yellow; number refers to the *E. coli* enzyme. Left to right: region 151-167 adjacent to the active site; regions 64-83 and 89-103 which are displaced and close to the ATP binding site; helix 342-362; tyrosine previously implicated in altering promiscuity alongside the lack of ten residues at the N-terminus when compared to human enzyme.

Figure 4: Comparison of wild-type and more promiscuous variants of *E. coli* galactokinase. (a) Top, Molecular overlay of the Y371H variant of *E. coli* galactokinase with the wild-type protein. This was carried out using Biovia Discovery Studio; the modal cluster of the Y371H simulation (c2) was compared with that of the wild-type (c3). The wild-type is shown in yellow and variant in red. Bottom, Regions of structural difference between Y371H variant and wild-type *E. coli* galactokinase. Y371H in pink/red and WT in grey/blue. Left to right: altered residues 65-111 close to ATP binding site; altered helix 331-340; region 151-165 adjacent to the active site; α -helix 214-244. (b) Top, molecular overlay of the M173L/Y371H variant of *E. coli* galactokinase with the wild-type enzyme. The modal cluster of the M173L/Y371H simulation (c4) was

compared with that of the wild-type (c3). The wild-type is shown in yellow and variant in red. Bottom, Regions of structural difference between the M173L/Y371H variant and wild-type *E. coli* galactokinase. M173L/Y371H is shown in pink/red and wild-type in grey/blue. Left to right: region 151-168 adjacent to the active site; shortening of the helix at 214-224; altered loop 331-340; residue 173 of wild-type (upper) and M173L/Y371H (lower).

Figure 5: Comparison of wild-type and more promiscuous variants of *L. lactis* galactokinase. Top, Molecular overlay of the Y385H variant of *L. lactis* galactokinase and the wild-type protein. This was carried out using Biovia Discovery Studio; the modal cluster of the Y385H simulation (c2) was compared with that of the wild-type (c1). The wild-type is shown in grey and the Y385H variant in purple. Bottom, regions of structural difference between the Y385H variant and wild-type *L. lactis* galactokinase. The Y385H variant is shown in green and the wild-type in grey/ blue. Left to right: shortening of the helix at residues 214-224; loss of a short helix between residues 80-86; altered loop 338-347.

Figure 6: Flexibility of wild-type and variant bacterial galactokinases. (a) RMSF values for each residue from *E. coli* and *L. lactis* galactokinases and their more promiscuous variants were calculated using 2 ns of stable trajectory with the 'atomicfluct' command of 'ptraj' of AmberTools. Of particular interest is the increase in RMSF around residues 214-224 (indicated by the arrow) in the both variants of the *E. coli* enzyme. A corresponding increase is not seen in the *L. lactis* enzyme. This may be because the wild-type *L. lactis* protein already has higher RMSF (note, different y-axes scales). (b) Squared fluctuation of *E. coli* and *L. lactis* galactokinases and their variants using principal component analysis and 1 ns of equilibrated trajectory. Calculations were carried out using an aligned protein and the normal mode wizard plugin of VMD software. All the variants result in increased flexibility between residues 200 and 250 (indicated with the arrows), suggesting that this contains a critical region for controlling activity and specificity. In (a) and (b) the wild-type trace is shown in red on the variant graphs for comparison. (c) Image showing local impact of alterations which result in increased promiscuity. The region shown in red has increased flexibility and the residue in blue is the tyrosine residue which when altered increases substrate promiscuity. The region in red is adjacent to sugar binding pocket.

Figure 7: Secondary structure propensity in wild-type and variant bacterial galactokinases. The histograms show the percentage time a particular residue spends in various forms of secondary structure during the molecular dynamics simulations. (a) Propensity for formation of secondary structure between residues 214-224. This is the region shown in red in Figure 6c. (b) Propensity for formation of secondary structure between residues 331-340 in *E.coli* and 338-347 in *L. lactis* galactokinase. This is the loop region shown in yellow in Figure 6c. In general, in the *L. lactis* enzyme both sets of residues tend to spend more time in an unstructured (i.e. no recognisable secondary structure). In the *E. coli* enzyme, both the Y371H and M173L/Y371H variants tend to reduce the amount of defined secondary structure, particularly α -helix.

The effect is less noticeable in the *L. lactis* enzyme, perhaps because these regions are less structured in the wild type enzyme.

Journal Pre-proof

Table

Table 1: A summary of the galactokinase variants which have been used to investigate mechanism and specificity. A comma (,) indicates distinct, separate variants and a forward slash (/) indicates compound variants.

Species/Variant	Outcome	Conclusion	Reference(s)
Human D186A, D186N	Loss of detectable activity and protein stability. Altered dynamics in residues 174–179 and 231–240	May be directly involved in catalysis. Required for optimum stability.	[36, 41]
Rat D186A, D186N, D186E	Loss of detectable activity	May be directly involved in catalysis.	[76]
<i>L. lactis</i> D183A, D183N	Loss of detectable activity. Likely to be responsible for group with $pK_a=6.9$.	May be directly involved in catalysis.	[39]
Human R37A, R37E	Loss of detectable activity and protein stability. Altered dynamics in residues 174–179 and 231–240	May be directly involved in catalysis. Required for optimum stability.	[36, 37, 41]
Human R37K	Partial loss of activity	Positive charge required at this position.	[36, 37]
<i>L. lactis</i> R36A, R36N	Partial loss of activity	Positive charge required at this position.	[39]
Human R105M, E174D	Modest reduction in activity. No change in stability. Altered dynamics in residues 174–179 and 231–240	Important, but not critical for mechanism.	[41]
Human R228M	Loss of protein stability. Some changes in interactions with substrates. Altered dynamics in residues 174–179 and 231–240	Critical for structure, but not directly required for mechanism.	[41]
Human E174Q	Inactive and unstable. Altered dynamics in residues 174–179 and 231–240.	Negative charge required at this position for both catalysis and structure.	[41]
Human M60V/M180V/A334S/D268E/R336Q/G373S (and some other combinations of these variants).	Increased stability and increased catalytic turnover. Altered dynamics in residues 174–181 and 231–240.	These “return to consensus” variants improve the performance of the enzyme.	[7]
Human L231S, N242S, E244S, E245S	Small changes in catalytic activity. In some cases,	This region is important in	[42]

	loss of detectable activity with 2-deoxy-galactose	controlling specificity.	
<i>E. coli</i> Y371H	Broadened specificity with good activity	This residue, despite being distant from the active site is involved in controlling specificity.	[44]
<i>E. coli</i> M173L, Y223F, Y223W, M173L/Y371H	Further expansion of the substrate range	These residues are also important in controlling specificity	[10, 45, 46]
<i>L. lactis</i> Y385H	Vastly expanded substrate range	<i>L. lactis</i> galactokinase is more promiscuous than <i>E. coli</i> or human – and has greater potential for engineering novel specificities.	[47]
Human Y379C, Y379K, Y379R, Y379S, Y379W	Expanded substrate range, but decreased activity. Dynamic and structural changes to residues 232-234.	This residue is important in determining specificity, but the human enzyme is less tolerant of sequence changes than the bacterial ones.	[7, 48]

References

- [1] H. Kristiansson, D.J. Timson, Galactokinases: potential biotechnological applications as biocatalysts, *Current Biotechnology*, 1 (2012) 148-154.
- [2] H.M. Holden, J.B. Thoden, D.J. Timson, R.J. Reece, Galactokinase: structure, function and role in type II galactosemia, *Cellular and molecular life sciences : CMLS*, 61 (2004) 2471-2484.
- [3] R.E. Trucco, R. Caputto, L.F. Leloir, N. Mittleman, Galactokinase, *Archives of biochemistry*, 18 (1948) 137-146.
- [4] S.M. Howard, M.R. Heinrich, The anomeric specificity of yeast galactokinase, *Archives of Biochemistry and Biophysics*, 110 (1965) 395-400.
- [5] F.J. Ballard, Kinetic studies with liver galactokinase, *The Biochemical journal*, 101 (1966) 70-75.
- [6] M. Sorensen, K.S. Mikkelsen, K. Frisch, L. Bass, B.M. Bibby, S. Keiding, Hepatic galactose metabolism quantified in humans using 2-¹⁸F-fluoro-2-deoxy-D-galactose PET/CT, *J Nucl Med*, 52 (2011) 1566-1572.
- [7] M. McAuley, N. Mesa-Torres, A. McFall, S. Morris, M. Huang, A.L. Pey, D.J. Timson, Improving the Activity and Stability of Human Galactokinase for Therapeutic and Biotechnological Applications, *Chembiochem*, 19 (2018) 1088-1095.
- [8] D.J. Timson, R.J. Reece, Sugar recognition by human galactokinase, *BMC biochemistry [computer file]*, 4 (2003) 16.
- [9] C.E. Cardini, L.F. Leloir, Enzymic phosphorylation of galactosamine and galactose, *Archives of Biochemistry and Biophysics*, 45 (1953) 55-64.
- [10] 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, *Organic letters*, 5 (2003) 2223-2226.
- [11] H.M. Holden, I. Rayment, J.B. Thoden, Structure and function of enzymes of the Leloir pathway for galactose metabolism, *Journal of Biological Chemistry*, 278 (2003) 43885-43888.
- [12] R. Caputto, L.F. Leloir, R.E. Trucco, C.E. Cardini, A.C. Paladini, The enzymatic transformation of galactose into glucose derivatives, *Journal of Biological Chemistry*, 179 (1949) 497-498.
- [13] D.J. Timson, *Molecular Genetics of Galactosaemia*, eLS, John Wiley & Sons, Ltd, Place Published, 2017.
- [14] D.J. Timson, The molecular basis of galactosemia - Past, present and future, *Gene*, 589 (2016) 133-141.
- [15] L. Welling, L.E. Bernstein, G.T. Berry, A.B. Burlina, F. Eyskens, M. Gautschi, S. Grunewald, C.S. Gubbels, I. Knerr, P. Labrune, J.H. van der Lee, A. MacDonald, E. Murphy, P.A. Portnoi, K. Ounap, N.L. Potter, M.E. Rubio-Gozalbo, J.B. Spencer, I. Timmers, E.P. Treacy, S.C. Van Calcar, S.E. Waisbren, A.M. Bosch, N. Galactosemia, International clinical guideline for the management of classical galactosemia: diagnosis, treatment, and follow-up, *J Inherit Metab Dis*, 40 (2017) 171-176.
- [16] M.E. Rubio-Gozalbo, M. Haskovic, A.M. Bosch, B. Burnyte, A.I. Coelho, D. Cassiman, M.L. Couce, C. Dawson, D. Demirbas, T. Derks, F. Eyskens, M.T. Forga, S. Grunewald, J. Haberle, M. Hochuli, A. Hubert, H.H. Huidekoper, P. Janeiro, J. Kotzka, I. Knerr, P. Labrune, Y.E. Landau, J.G. Langendonk, D. Moslinger, D. Muller-Wieland, E. Murphy, K. Ounap, D. Ramadza, I.A. Rivera, S. Scholl-Buergi, K.M. Stepien, A. Thijs, C. Tran, R. Vara, G. Visser, R. Vos, M. de Vries, S.E. Waisbren, M.M. Welsink-Karssies, S.B. Wortmann, M. Gautschi, E.P. Treacy, G.T. Berry, The natural history of classic galactosemia: lessons from the GalNet registry, *Orphanet J Rare Dis*, 14 (2019) 86.
- [17] 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, *Journal of inherited metabolic disease*, 25 (2002) 629-634.

- [18] R. Gitzelmann, Hereditary Galactokinase Deficiency, a Newly Recognized Cause of Juvenile Cataracts, *Pediatric research*, 1 (1967) 14-23.
- [19] W.R. Pickering, R.R. Howell, Galactokinase deficiency: clinical and biochemical findings in a new kindred, *The Journal of pediatrics*, 81 (1972) 50-55.
- [20] R. Gitzelmann, Deficiency of erythrocyte galactokinase in a patient with galactose diabetes, *Lancet*, 2 (1965) 670-671.
- [21] M. Tang, E. Etokidem, K. Lai, The Leloir Pathway of Galactose Metabolism - A Novel Therapeutic Target for Hepatocellular Carcinoma, *Anticancer Res*, 36 (2016) 6265-6271.
- [22] C.A. Sellick, R.N. Campbell, R.J. Reece, Chapter 3 Galactose metabolism in yeast - structure and regulation of the Leloir pathway enzymes and the genes encoding them, *International review of cell and molecular biology*, 269 (2008) 111-150.
- [23] D.J. Timson, Galactose metabolism in *Saccharomyces cerevisiae*, *Dynamic Biochemistry, Process Biotechnology and Molecular Biology*, 1 (2007) 63-73.
- [24] M.J. Weickert, S. Adhya, The galactose regulon of *Escherichia coli*, *Mol Microbiol*, 10 (1993) 245-251.
- [25] 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 science : a publication of the Protein Society*, 2 (1993) 31-40.
- [26] D.J. Timson, GHMP kinases - structures, mechanisms and potential for therapeutically relevant inhibition, *Current Enzyme Inhibition*, 3 (2007) 77-94.
- [27] J.B. Thoden, H.M. Holden, Molecular structure of galactokinase, *Journal of Biological Chemistry*, 278 (2003) 33305-33311.
- [28] J.B. Thoden, D.J. Timson, R.J. Reece, H.M. Holden, Molecular structure of human galactokinase: implications for type II galactosemia, *Journal of Biological Chemistry*, 280 (2005) 9662-9670.
- [29] Z. Fu, M. Wang, D. Potter, H.M. Mizioro, J.J. Kim, The structure of a binary complex between a mammalian mevalonate kinase and ATP: insights into the reaction mechanism and human inherited disease, *Journal of Biological Chemistry*, 277 (2002) 18134-18142.
- [30] S.S. Krishna, T. Zhou, M. Daugherty, A. Osterman, H. Zhang, Structural basis for the catalysis and substrate specificity of homoserine kinase, *Biochemistry (John Wiley & Sons)*, 40 (2001) 10810-10818.
- [31] M.J. Romanowski, J.B. Bonanno, S.K. Burley, Crystal structure of the *Streptococcus pneumoniae* phosphomevalonate kinase, a member of the GHMP kinase superfamily, *Proteins*, 47 (2002) 568-571.
- [32] T. Wada, T. Kuzuyama, S. Satoh, S. Kuramitsu, S. Yokoyama, S. Unzai, J.R. Tame, S.Y. Park, Crystal structure of 4-(cytidine 5'-diphospho)-2-C-methyl-D-erythritol kinase, an enzyme in the non-mevalonate pathway of isoprenoid synthesis, *Journal of Biological Chemistry*, 278 (2003) 30022-30027.
- [33] J.B. Bonanno, C. Edo, N. Eswar, U. Pieper, M.J. Romanowski, V. Ilyin, S.E. Gerchman, H. Kycia, F.W. Studier, A. Sali, S.K. Burley, Structural genomics of enzymes involved in sterol/isoprenoid biosynthesis, *Proceedings of the National Academy of Sciences of the United States of America*, 98 (2001) 12896-12901.
- [34] L. Miallau, M.S. Alphey, L.E. Kemp, G.A. Leonard, S.M. McSweeney, S. Hecht, A. Bacher, W. Eisenreich, F. Rohdich, W.N. Hunter, Biosynthesis of isoprenoids: crystal structure of 4-diphosphocytidyl-2C-methyl-D-erythritol kinase, *Proceedings of the National Academy of Sciences of the United States of America*, 100 (2003) 9173-9178.
- [35] S. Roy, M. Vivoli Vega, N. Harmer, Carbohydrate Kinases: A Conserved Mechanism Across Differing Folds, *Catalysts*, 9 (2019) 29.
- [36] 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, *Bioorganic Chemistry*, 39 (2011) 120-126.
- [37] M. Tang, K. Wierenga, L.J. Elsas, K. Lai, Molecular and biochemical characterization of

- human galactokinase and its small molecule inhibitors, *Chemico-biological interactions*, 188 (2010) 376-385.
- [38] X. Chu, N. Li, X. Liu, D. Li, Functional studies of rat galactokinase, *Journal of Biotechnology*, 141 (2009) 142-146.
- [39] 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 letters*, 587 (2013) 2876-2881.
- [40] 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.
- [41] M. McAuley, M. Huang, D.J. Timson, Insight into the mechanism of galactokinase: Role of a critical glutamate residue and helix/coil transitions, *Biochim Biophys Acta*, 1865 (2017) 321-328.
- [42] M. McAuley, M. Huang, D.J. Timson, Modulation of the mobility of a key region in human galactokinase: Impacts on catalysis and stability, *Bioorg Chem*, 81 (2018) 649-657.
- [43] J. Yang, D. Hoffmeister, L. Liu, X. Fu, J.S. Thorson, Natural product glycorandomization, *Bioorganic & medicinal chemistry*, 12 (2004) 1577-1584.
- [44] D. Hoffmeister, J. Yang, L. Liu, J.S. Thorson, Creation of the first anomeric D/L-sugar kinase by means of directed evolution, *Proceedings of the National Academy of Sciences of the United States of America*, 100 (2003) 13184-13189.
- [45] D. Hoffmeister, J.S. Thorson, Mechanistic implications of *Escherichia coli* galactokinase structure-based engineering, *Chembiochem*, 5 (2004) 989-992.
- [46] 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, *Chemistry & biology*, 12 (2005) 657-664.
- [47] J. Yang, L. Liu, J.S. Thorson, Structure-based enhancement of the first anomeric glucokinase, *Chembiochem : a European journal of chemical biology*, 5 (2004) 992-996.
- [48] H. Kristiansson, D.J. Timson, Increased promiscuity of human galactokinase following alteration of a single amino acid residue distant from the active site, *Chembiochem : a European journal of chemical biology*, 12 (2011) 2081-2087.
- [49] 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.
- [50] M. McAuley, D.J. Timson, Modulating Mobility: a Paradigm for Protein Engineering?, *Appl Biochem Biotechnol*, 181 (2016) 83-90.
- [51] N.M. Goodey, S.J. Benkovic, Allosteric regulation and catalysis emerge via a common route, *Nature chemical biology*, 4 (2008) 474-482.
- [52] D.J. Timson, R.J. Reece, Functional analysis of disease-causing mutations in human galactokinase, *European journal of biochemistry / FEBS*, 270 (2003) 1767-1774.
- [53] F. Sangiuolo, M. Magnani, D. Stambolian, G. Novelli, Biochemical characterization of two GALK1 mutations in patients with galactokinase deficiency, *Human mutation*, 23 (2004) 396.
- [54] B. Jojart, M. Szori, R. Izsak, I. Marsi, A. Laszlo, I.G. Csizmadia, B. Viskolcz, The effect of a Pro(28)Thr point mutation on the local structure and stability of human galactokinase enzyme-a theoretical study, *Journal of molecular modeling*, 17 (2011) 2639-2649.
- [55] A. d'Acierno, B. Scafuri, A. Facchiano, A. Marabotti, The evolution of a Web resource: The Galactosemia Proteins Database 2.0, *Hum Mutat*, 39 (2018) 52-60.
- [56] S. P. E.A. Ebrahimi, S.A. Ghazala, T.K. D, S. R, C.G. Priya Doss, H. Zayed, Structural analysis of missense mutations in galactokinase 1 (GALK1) leading to galactosemia type-2, *J Cell Biochem*, 119 (2018) 7585-7598.
- [57] 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.
- [58] L.P. Conway, F.F. Liu, Q. Li, J. Voglmeir, The *Shewanella woodyi* galactokinase pool phosphorylates glucose at the 6-position, *Carbohydr Res*, 455 (2017) 39-44.
- [59] K. Huang, F. Parmeggiani, E. Pallister, C.J. Huang, F.F. Liu, Q. Li, W.R. Birmingham, P. Both,

- B. Thomas, L. Liu, J. Voglmeir, S. Flitsch, Characterisation of a bacterial galactokinase with high activity and broad substrate tolerance for chemoenzymatic synthesis of 6-aminogalactose-1-phosphate and analogues, *Chembiochem*, 19 (2018) 388-394.
- [60] 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.
- [61] 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, *Carbohydrate research*, 346 (2011) 2421-2425.
- [62] T. Keenan, R. Mills, E. Pocock, D. Budhadev, F. Parmeggiani, S. Flitsch, M. Fascione, The characterisation of a galactokinase from *Streptomyces coelicolor*, *Carbohydr Res*, 472 (2019) 132-137.
- [63] L.P. Conway, J. Voglmeir, Functional analysis of anomeric sugar kinases, *Carbohydr Res*, 432 (2016) 23-30.
- [64] F. Huang, W.M. Nau, A conformational flexibility scale for amino acids in peptides, *Angewandte Chemie (International ed.in English)*, 42 (2003) 2269-2272.
- [65] K. Arnold, L. Bordoli, J. Kopp, T. Schwede, The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling, *Bioinformatics (Oxford, England)*, 22 (2006) 195-201.
- [66] M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. Gallo Cassarino, M. Bertoni, L. Bordoli, T. Schwede, SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information, *Nucleic Acids Res*, 42 (2014) W252-258.
- [67] F. Kiefer, K. Arnold, M. Kunzli, L. Bordoli, T. Schwede, The SWISS-MODEL Repository and associated resources, *Nucleic Acids Res*, 37 (2009) D387-392.
- [68] N. Guex, M.C. Peitsch, T. Schwede, Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective, *Electrophoresis*, 30 Suppl 1 (2009) S162-173.
- [69] S.C. Lovell, I.W. Davis, W.B. Arendall, 3rd, P.I. de Bakker, J.M. Word, M.G. Prisant, J.S. Richardson, D.C. Richardson, Structure validation by $\text{C}\alpha$ geometry: ϕ, ψ and $\text{C}\beta$ deviation, *Proteins*, 50 (2003) 437-450.
- [70] M.J. Sippl, Recognition of errors in three-dimensional structures of proteins, *Proteins*, 17 (1993) 355-362.
- [71] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, *Nucleic Acids Res*, 35 (2007) W407-410.
- [72] A. Bhattacharya, R. Tejero, G.T. Montelione, Evaluating protein structures determined by structural genomics consortia, *Proteins*, 66 (2007) 778-795.
- [73] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, *Journal of computational chemistry*, 25 (2004) 1157-1174.
- [74] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *Journal of Molecular Graphics*, 14 (1996) 33-38, 27-38.
- [75] A. Bakan, L.M. Meireles, I. Bahar, ProDy: protein dynamics inferred from theory and experiments, *Bioinformatics*, 27 (2011) 1575-1577.
- [76] X. Chu, D. Li, Cloning, expression, and purification of His-tagged rat mevalonate kinase, *Protein expression and purification*, 27 (2003) 165-170.

E. coli Wild-type

E. coli Y371H

E. coli M173L/Y371H

L. lactis Wild-type

L. lactis Y385H

(a)

(b)

(a)

(b)

(a)

(b)

(c)

(a)

(b)

- Bacterial galactokinases are typically more promiscuous than mammalian ones
- A key α -helix in human galactokinase is important in controlling specificity
- The mobility of the equivalent helix in the bacterial enzyme was investigated
- This region is predicted to be important in the wide specificity of these enzymes
- Therefore, it could be altered to further expand the substrate range

Dynamic origins of substrate promiscuity in bacterial galactokinases

Margaret McAuley¹, Meilan Huang² 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.

Conflict of interest statement: The authors have no conflicts of interest to declare.