

Improving the activity and stability of human galactokinase for therapeutic and biotechnological applications

Margaret McAuley^a, Noel Mesa-Torres^b, Aisling McFall^a, Sarah Morris^a, Meilan Huang^c, Angel L. Pey^b and David J. Timson^{a,d,*}

Abstract: Galactokinase catalyses the site- and stereospecific phosphorylation of α -D-galactose. As such it has attracted interest as a biocatalyst for the introduction of phosphate groups into monosaccharides. However, attempts to broaden the substrate range of human galactokinase have generally resulted in substantially reduced activity. The enzyme also has biotechnological potential in enzyme replacement therapy (ERT) for type II galactosemia. The return-to-consensus approach can be used to identify residues which can be altered to increase protein stability and enzyme activity. This approach identified six residues of potential interest in human galactokinase. Some of the single consensus variants increased the catalytic turnover of the enzyme (M60V, D268E, A334S and G373S), but none resulted in improved stability. When all six changes were introduced into the protein (M60V/M180V/D268E/A334S/R366Q/G373S) thermal stability was increased. Molecular dynamics simulations suggested that these changes altered the protein's conformation at key sites. The number of salt bridges and hydrogen bonds was also increased. Combining the six consensus variations with Y379W (a variant with greater substrate promiscuity) increased the stability of this variant and its turnover towards some substrates. Thus, the six consensus variants can be used to stabilise catalytically interesting variants of human galactokinase and may also be useful if the protein were to be used in ERT.

Introduction

Galactokinase (GALK1; EC 2.7.1.6) catalyses the ATP-dependent site- and stereo-specific phosphorylation of the hexose monosaccharide galactose^[1]. Physiologically, this reaction is required as the first committed step of the Leloir pathway of galactose metabolism^[2]. This reaction converts galactose (which cannot be efficiently phosphorylated by the first enzyme of glycolysis, hexokinase) to galactose 1-phosphate. This compound subsequently reacts with UDP-glucose, generating UDP-galactose and glucose 1-phosphate in a reaction catalysed by galactose 1-phosphate uridylyltransferase. Glucose 1-phosphate is isomerised to glucose 6-phosphate, a glycolytic intermediate by phosphoglucosmutase. UDP-glucose is regenerated from UDP-galactose by the action of UDP-galactose 4'-epimerase. The Leloir pathway also has a role in maintaining the pools of UDP-sugars which are used in the biosynthesis of glycoproteins and glycolipids.

Sugar 1-phosphates are useful precursors in the chemical synthesis of more complex molecules which contain monosaccharide moieties, including aminoglycoside antibiotics^[3]. Although phosphorylation of monosaccharides can be achieved chemically using reagents such as phosphoric acid, the presence of multiple hydroxyl groups means that a wide range of products are normally produced. Consequently, enzymes such as galactokinase which catalyse the site-specific phosphorylation of monosaccharides are of interest in biocatalysis^[4].

Galactokinase demonstrates high selectivity, catalysing the phosphorylation of the hydroxyl group at position 1 in the α -anomeric configuration only. The human enzyme is also highly specific for its sugar substrate: α -D-galactose and the closely related sugars α -D-galactosamine and 2-deoxy- α -D-galactose are the only known substrates of the wild-type enzyme^[5]. In

- [a] Dr Margaret McAuley, Aisling McFall, Sarah Morris, Dr David J Timson
School of Biological Sciences
Queen's University, Belfast
Medical Biology Centre, Lisburn Road, Belfast, BT9 7BL. UK
- [b] Dr Noel Mesa-Torres, Dr Angel L Pey
Department of Physical Chemistry
University of Granada
Av. Fuentenueva s/n, 18071, Granada. Spain
- [c] Dr Meilan Huang
School of Chemistry and Chemical Engineering
Queen's University, Belfast
David Keir Building, Stranmillis Road, Belfast, BT9 5AG. UK
- [d] Prof David J Timson
School of Pharmacy and Biomolecular Sciences
University of Brighton
Huxley Building, Lewes Road, Brighton, BN2 4GJ. UK
E-mail: d.timson@brighton.ac.uk

general, bacterial galactokinases seem to have more relaxed specificity, both in terms of the range of substrates and the sites which can be phosphorylated; however the substrate range is still limited^[6]. Considerable efforts have been made to expand the substrate range of *Escherichia coli* and *Lactococcus lactis* galactokinases. A combination of rational and random mutagenesis has identified two sites in bacterial galactokinases (Met-173 and Tyr-371 in the *E. coli* enzyme) which, when altered, greatly broaden the substrate range of these enzymes^[6a, 7]. Parallel studies have also been conducted using the human enzyme. Alteration of the equivalent tyrosine residue (Tyr-379) also broadens the specificity range of the enzyme, albeit with considerable reduction in activity^[8]. This tyrosine residue forms part of a β -sheet structure which is distant from the active site^[9]. It has been postulated that alteration of the tyrosine residue results in changes to the dynamics of active site residues which enable a greater range of molecules to bind, but also reduces catalytic efficiency^[8, 10].

Galactokinase is also implicated in human disease^[11]. Mutations in the gene encoding human galactokinase are associated with the inherited metabolic disease type II galactosemia (OMIM #230250)^[11b, 12]. Compared to types I and III galactosemia, type II is generally considered relatively mild^[13]. The only well-documented manifestations of this disease are early onset cataracts which can be resolved by elimination of galactose (and its precursors such as lactose) from the diet or by surgery^[14]. However, some reports suggest that there may be more serious, long term consequences of this disease including mental retardation and failure to thrive, especially where dietary compliance has been poor^[15]. Other than dietary restriction of galactose, no other treatments are available for type II galactosemia. Some inherited metabolic diseases can be treated by enzyme replacement therapy (ERT). This treatment involves the supplementation with highly purified, recombinant enzyme^[16]. A common problem in ERT is short protein half-lives in the plasma or in cells.

More structurally robust forms of galactokinase would have advantages in biocatalysis. Such modified forms of the enzyme would be more likely to withstand increased temperatures and also to remain active over longer periods. If ERT was considered desirable and

viable for the treatment of type II galactosemia, then more stable forms of the enzyme may be desirable. One method for improving the overall stability of proteins is the so-called consensus approach. This involves identification (through sequence alignments) of residues which deviate from the consensus in a target sequence (e.g. the human orthologue). These deviations are likely to result from natural changes in gene sequences over evolutionary timescales. Such changes will be tolerated so long as they do not damage the overall fitness of the organism. Altering the affected residues back to the consensus often results in increased overall protein stability and, sometimes, enhanced activity^[17]. This approach is also useful to identify second-site suppressors of disease phenotypes^[18]. Here we report the identification of six sites in human galactokinase which deviate from the consensus and the effects of restoring the consensus on the stability and activity of the enzyme. Molecular dynamics simulations were used to help understand the mechanisms of any changes in stability. Some of the variants reported here have improved overall robustness and so may be valuable in the development of biocatalysts.

Results and Discussion

Identification of six key sites which deviate from the consensus in human galactokinase

Consensus variants for human GALK1 were identified from the analysis of a sequence alignment of animal sequences (Supplementary Figure S2A). This analysis provided twelve different sites as potential consensus variants (Supplementary Figure S2B). Five variants (K69T, Q172K, V238E, H363R and H371K) were discarded since similar analyses using only sequences from mammals revealed a high frequency of the human amino acid at these positions (Supplementary Figure S2B). The C391S variant was also discarded due to the appearance of 12 different amino acids (out of 26 sequences at this position), indicating high degeneracy. Therefore, a set of six consensus variants (M60V (2), M180V (3), D268E (4), A334S (5), R366Q (6) and G373S (7)) were selected for further experimental characterisation (Figure 1).

The six individual return-to-consensus variants were expressed in, and purified from, *E. coli*. In addition, two double variants (M60V/M180V (**8**) and M60V/A334S (**9**)), one triple variant (M60V/M180V/A334S (**10**)), one quadruple variant (M60V/M180V/A334S/G373S (**11**)), one penta-variant (M60V/M180V/A334S/D268E/G373S (**12**)) and one hexa-variant incorporating all the return-to-consensus changes identified in this study (M60V/M180V/A334S/D268E/R336Q/G373S (**13**)) were synthesised. All six proteins were in the soluble fraction following disruption of the bacterial cells by sonication (Supplementary Figure S3). Typical yields were in the range 1-2 mg purified GALK1 per litre of bacterial cell culture.

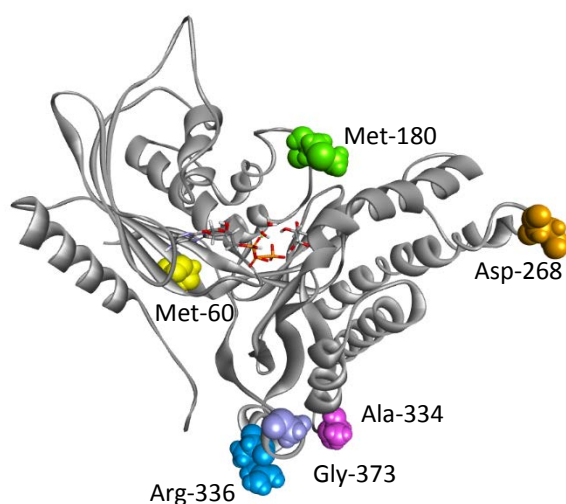


Figure 1. Location of the consensus residues identified in this study. The structure of human galactokinase (PDB 1WUU^[9]) is shown in grey. The substrates (galactose and ATP) are shown in stick format and the consensus residues in space-filling format, highlighted in colour.

Some individual return-to-consensus variants increase catalytic turnover, but none increase stability

The six return-to-consensus variants were first tested individually. Four of these (M60V (**2**), D268E (**4**), A334S (**5**) and G373S (**7**)) had statistically significantly increased catalytic turnover (k_{cat}). One (M180V (**3**)) had a significantly decreased turnover and one (R336Q (**6**)) was unchanged. However, none of these changes were greater than three-fold. (Table 1; Supplementary Figure S4). In general, where there was an increase in the turnover number, there was a corresponding

increase in the Michaelis constants. Consequently, there were no significant changes in the specificity constants for either galactose or ATP for any of these six variants.

Two of these variants (D268E (**4**) and A334S (**5**)) caused a significant decrease in the thermal stability (T_m); the remainder had no significant effect (Table 2). D268E (**4**) is also less stable towards chaotropic denaturation. In native gel electrophoresis, the wild-type protein ran largely as one band in the absence of urea. This band was still detected (albeit at lower intensity) up to 2 M urea. In contrast, D268E (**4**) in 0.5 M urea runs largely as a faster migrating form which is present up to 2 M urea (Figure 2).

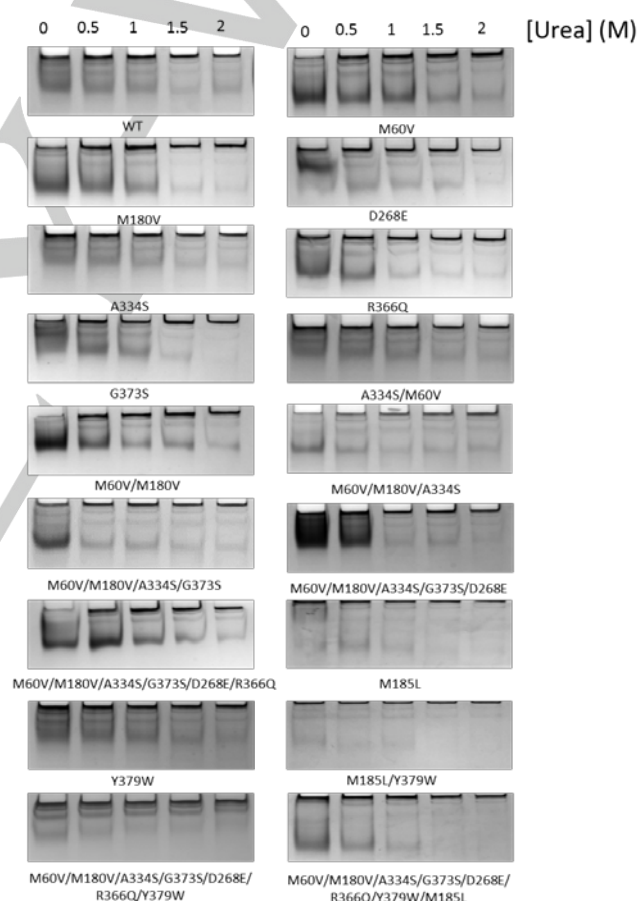


Figure 2. Stability of variant galactokinases to denaturation by the chaotropic agent, urea. Enzymes (4.5 μ M) were incubated at 37°C for 30 min with increasing concentrations of urea (shown above the gel) before loading onto a 15% native polyacrylamide gel (pH 8.8) and electrophoresed at 20 mA for 3 h. Proteins were visualised by staining with Coomassie blue. In general, non-denatured proteins run as a darker band (sometimes with slower migrating species) which becomes more blurred and/or lighter as the urea concentration increases. Variant galactokinases which become blurred/lighter at lower concentrations of urea are more susceptible to chaotropic denaturation.

Table 1. Kinetic parameters of consensus variants of human galactokinase

Entry number	Galactokinase variant	$K_{m, ATP}$ (μM)	$K_{m, Gal}$ (μM)	k_{cat} (s^{-1})	$k_{cat}/K_{m, ATP}$ ($mM^{-1}s^{-1}$)	$k_{cat}/K_{m, Gal}$ ($mM^{-1}s^{-1}$)
1	Wild type	3.9 ± 1.1	170 ± 25	1.7 ± 0.05	430 ± 108	10.1 ± 1.2
2	M60V	18 ± 4.7**	220 ± 90	4.8 ± 0.3***	270 ± 58	21.8 ± 7.5
3	M180V	5.4 ± 0.84	99 ± 36	1.0 ± 0.03**	185 ± 27	10.1 ± 3.3
4	D268E	13 ± 4.3*	580 ± 91**	3.2 ± 0.2***	240 ± 68	6.0 ± 7.0
5	A334S	7.9 ± 3.0	280 ± 98	4.3 ± 0.3***	550 ± 188	15.4 ± 4.3
6	R366Q	7.7 ± 4.2	530 ± 170*	2.1 ± 0.2	270 ± 130	4.0 ± 1.0
7	G373S	13 ± 4.3*	90 ± 47	2.4 ± 0.2**	190 ± 56	25 ± 12*
8	M60V/M180V	1.0 ± 0.3	190 ± 53	1.5 ± 0.02	1430 ± 410***	8.8 ± 2.1
9	M60V/A334S	4.2 ± 2.9	480 ± 180*	2.6 ± 0.01***	620 ± 230	5.4 ± 1.5
10	M60V/M180V/A334S	0.82 ± 1.60	78 ± 32	1.7 ± 0.04	2000 ± 240	20 ± 8
11	M60V/M180V/A334S/G373S	0.57 ± 0.66	230 ± 70	1.7 ± 0.1	3000 ± 1000	7.3 ± 1.9
12	M60V/M180V/A334S/D268E/G373S	2.5 ± 0.6	270 ± 103	1.7 ± 0.04	680 ± 141	6.3 ± 2.0
13	M60V/M180V/A334S/D268E/R366Q/G373S	3.8 ± 0.6	150 ± 20	1.6 ± 0.05	390 ± 58	10.9 ± 1.3
14	M185L	34.5 ± 6.3***	1190 ± 270	8.1 ± 0.4***	236 ± 34	8.0 ± 1.2
15	Y379W	5.0 ± 0.7	1320 ± 430	6.8 ± 1.5**	1360 ± 180	5.2 ± 0.6
16	M185L/Y379W	11.4 ± 2.9	1120 ± 400	2.5 ± 0.1	220 ± 50	2.4 ± 0.5*
17	M60V/M180V/D268E/A334S/R366Q/G373S/Y379W	8.1 ± 1.8	310 ± 85	5.0 ± 0.3*	550 ± 110	16.1 ± 3.6
18	M60V/M185L/M180V/D268E/A334S/R366Q/G373S/Y379W	12.4 ± 5.8	3278 ± 487***	12.9 ± 0.9***	1000 ± 350	3.9 ± 0.3

All measurements taken using 70 nM enzyme as described in Materials and Methods. Parameters were estimated by fitting graphs of $V_{max,app}$ plotted against substrate concentration with Michaelis-Menten equation using nonlinear curve fitting and errors shown are standard errors calculated in this process. Significance calculated via one-way ANOVA with Dunnett's post-test analysis to compare with the wild-type enzyme's parameters. * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

Molecular dynamics simulations of these variants revealed some interesting features when compared to the wild-type proteins. In M60V (**2**), alterations were seen in a largely α -helical region (residues 232-240) which has previously been implicated in catalysis ^[19]. Another key change was observed around residue 156, resulting in shortening of an α -helix and consequent displacement of residues 163-177 (Figure 3a). Similar changes were seen in M180V (**3**), D268E (**4**) and A334S (**5**). In M180V (**3**), residues 260-270 are also displaced and in A334S (**5**), a tighter turn around residues 257 and 258 displaces a similar region (residues 262-270). R336Q (**6**) and G373S (**7**) showed fewer changes in residues 232-240 compared to the other four variants. A334S (**5**) and G373S (**7**) showed similar changes around residues 257-270. The loop comprised of residues 257 and 258 is more flexible in R336Q (**6**). In this variant, the effect is the displacement of residues 261-264 and the lengthening of an α -helix (residues 274-290). M60V (**2**) has a predicted increase in the number of salt bridges and hydrogen bonds compared to the wild-type protein (Table 3). However, these do not translate into increased thermal stability (Table 3).

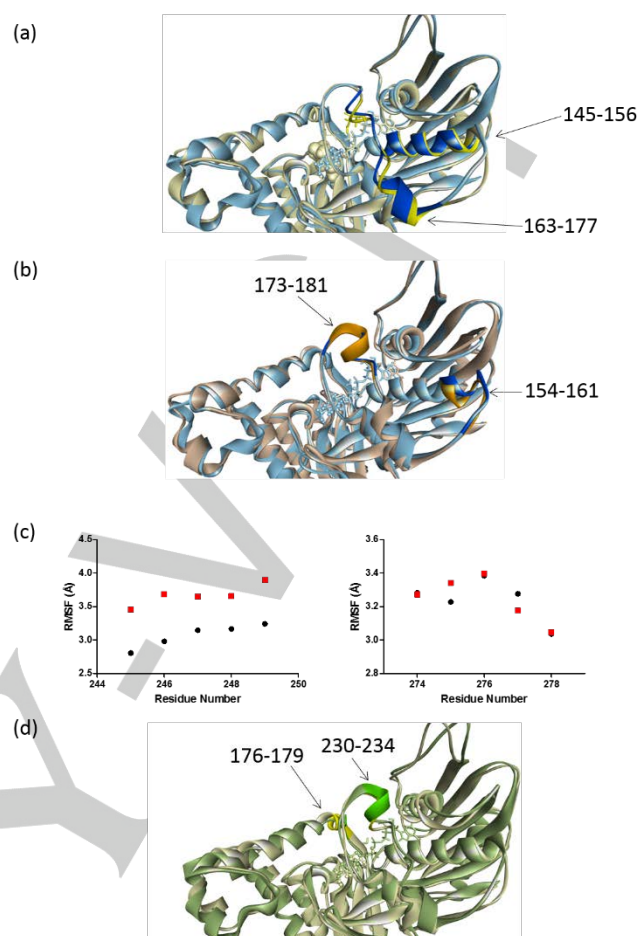


Figure 3. Molecular dynamics reveal changes in the protein's conformation. (a) A comparison of the wild-type protein (**1**) (yellow) and the M60V (**2**) variant (blue) showing predicted changes in two linked α -helices. (b) A comparison of the M60V (**2**) variant (orange) with the M60V/A334S (**9**) double variant showing two key regions which differ between the variants. (c) Graphs showing the root mean square fluctuation (RMSF) for two regions in the vicinity of salt bridges predicted to form in the hexa-variant (M60V/M180V/A334S/G373S/D268E/R336Q (**13**); black circles) compared to the wild type (**1**) (red squares). (d) A comparison of the wild-type protein (yellow) (**1**) and the Y379W variant (**15**) (green) showing two predicted structural changes.

Combinations of return-to-consensus variants increase stability

The greatest increase in turnover number was seen in the M60V (**2**) and A334S (**5**) variants. Combining these two variants (M60V/A334S (**9**)) also resulted in a significant increase in the turnover number, but this increase was not as great as in M60V (**2**) alone (Table 1; Supplementary Figure S4). This double variant (**9**) had reduced stability towards thermal denaturation (Table 2). Combining M60V and M180V (**7**) resulted in a modest, but significant increase in the melting

temperature, but no change in the turnover number (Tables 1, 2). Interestingly, this was the only variant studied with a significantly increased (three-fold) specificity constant for ATP (Table 1). When the three variants (M60V/M180V/A334S (**10**)) were combined, none of the steady state kinetic parameters were significantly altered. However, this triple variant did have significantly increased thermal stability. Combining four (M60V/M180V/A334S/G373S (**11**)), five (M60V/M180V/A334S/G373S/D268E (**12**)) or six (M60V/M180V/A334S/G373S/D268E/R336Q (**13**)) variants resulted in a similar increase in thermal stability, but no significant change in the kinetic parameters (Tables 1, 2). The combination of all six variants also appeared to result in a protein which was more stable towards denaturation by urea (Figure 2).

Melting temperature values (T_m) of variant galactokinases (4.5 μ M) were determined by differential scanning calorimetry. Significance was determined by comparing means of each dataset ($n=9$) to the wild-type value using one-way ANOVA. ns, not significant; * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

MD simulations showed that the M60V/A334S (**9**) and M60V/M180V (**8**) double variants are predicted to vary in the region 232-240 and also in the turn comprised of residues 173-181 which became tighter. In M60V/A334S (**9**) an α -helix (residues 154-161) is also shortened (Figure 3b). Interestingly, the changes to residues 154-161 and 173-181 were not observed in the triple variant (M60V/M180V/A334S (**10**)). The quadruple variant (M60V/M180V/A334S/G373S (**11**)) showed all three changes as did the hexa-variant (M60V/M180V/A334S/G373S/D268E/R336Q (**13**)). The main changes observed in the penta-variant (M60V/M180V/A334S/G373S/D268E (**12**)) were in residues 173-181. All of the multiple variants were predicted to have an increased number of salt bridges compared to wild-type and all except the hexa-variant were predicted to have an increased number of hydrogen bonds (Table 3). Of particular interest, are the salt bridges between Glu245-Arg249 and Glu274-Arg278. These are predicted to occur in all the variants with increased thermal stability. It seems likely that these residues contribute to the reduced flexibility around residue 270 resulting in increased global protein stability. Analysis of the root mean squared fluctuation (RMSF) of these two regions showed that the first region had a reduction in the hexa variant (**13**) compared to the wild-type. However, the second region did not (Figure 3c).

Table 2. Thermal stability of galactokinase variants

Entry number	Variant	T_m ($^{\circ}$ C)
1	Wild type	55.74 \pm 0.13
2	M60V	55.93 \pm 0.29 ^{ns}
3	M180V	55.17 \pm 0.29 ^{ns}
4	D268E	52.31 \pm 0.30 ^{***}
5	A334S	54.41 \pm 0.17 [*]
6	R366Q	55.81 \pm 0.10 ^{ns}
7	G373S	55.01 \pm 0.16 ^{ns}
8	M60V/M180V	57.7 \pm 0.35 ^{***}
9	A334S/M60V	52.31 \pm 0.12 ^{***}
10	M180V/M60V/A334S	58.08 \pm 0.08 ^{***}
11	M180V/M60V/A334S/G373S	58.11 \pm 0.28 ^{***}
12	M180V/M60V/A334S/G373S/D268E	58.09 \pm 0.35 ^{***}
13	M60V/M180V/A334S/D268E/R336Q/G373S	58.12 \pm 0.50 ^{***}
14	M185L	50.51 \pm 0.34 ^{***}
16	Y379W	52.07 \pm 0.22 ^{***}
16	M185L/Y379W	50.33 \pm 0.71 ^{***}
17	M60V/M180V/D268E/A334S/R366Q/G373S/Y379W	57.14 \pm 0.35 ^{***}
18	M60V/M185L/M180V/D268E/A334S/R366Q/G373S/Y379W	53.83 \pm 0.87 ^{***}

Table 3. Structural properties of the variant galactokinases predicted from molecular dynamics simulations

Entry number	Variant	Number of predicted salt bridges	Number of predicted hydrogen bonds
1	Wild type	32	87.3 ± 7.4
2	M60V	40	89.8 ± 7.9***
3	M180V	31	86.3 ± 7.4*
4	D268E	33	88.2 ± 7.6 ^{ns}
5	A334S	36	87.5 ± 7.3 ^{ns}
6	R366Q	38	87.3 ± 7.6 ^{ns}
7	G373S	36	88.1 ± 7.5 ^{ns}
8	M160V/M180V	40	89.1 ± 7.8***
9	A334S/M60V	39	92.1 ± 7.6***
10	M60V/M180V/A334S	37	90.8 ± 7.5***
11	M60V/M180V/A334S/G373S	33	89.2 ± 7.5***
12	M60V/M180V/D268E/A334S/G373S	35	89.5 ± 7.3***
13	M60V/M180V/D268E/A334S/R366Q/G373S	37	86.7 ± 7.3 ^{ns}

The number of salt bridges present in each variant was calculated using 1 ns of stable trajectory and the 'salt bridges' plugin of VMD. A 3.2 Å cut-off distance from oxygen to nitrogen was used. Mean numbers of hydrogen bonds present in each consensus variant were calculated using a 1 ns trajectory of stable simulation using the 'Hydrogen bonds' plugin in VMD. Error values shown are the standard deviations. Significance was determined using a one-way ANOVA and significance compared to wild-type determined using Dunnett's post-test analysis. ns, not significant; * $p \leq 0.05$; *** $p \leq 0.001$.

Consensus changes stabilise variants with altered catalytic properties

In some bacterial galactokinases, alteration of a methionine residue to leucine increases the substrate range^[7d]. The equivalent residue in the human enzyme is Met-185. An M185L (**14**) variant of human galactokinase had statistically significantly increased catalytic turnover, with corresponding increases in the Michaelis constants for both substrates (all approximately 8 to 10-fold). Consequently, the specificity constants for galactose and ATP were not significantly changed (Table 1). This variant was significantly less thermally stable than the wild-type protein and was less resistant to denaturation by urea (Table 2; Figure 2). An initial screen with 16 monosaccharides did not reveal any novel substrates for the M185L variant (**14**) (data not shown) suggesting

that this variant does not increase the promiscuity of the human enzyme. This screen did reveal that, like the wild-type enzyme, the M185L variant (**14**) is active with 2-deoxy-D-galactose and D-galactosamine. In addition, the wild-type (**1**) and M185L variant protein (**14**) are both active with D-talose, a sugar which has not been tested previously as a substrate of human galactokinase (Table 4).

It has been shown previously that alteration of Tyr-379 can increase the substrate range of human galactokinase^[8]. Of the alterations at Tyr-379, Y379W (**15**) was the most promiscuous, demonstrating activity with the greatest range of monosaccharide substrates^[8]. MD studies predicted that this change disrupted the α -helix between residues 232 and 234 (Figure 3d). Similar changes were seen in the other variants at this site (data not shown). In addition, a short section of α -helix is created in Y379W (**15**) between residues 176 and 179 (Figure 3d). This is close to the active site and the catalytically important residue, Glu-174. Like M185L (**14**), this variant is thermally destabilised compared to the wild-type (Table 2). However, when combined with the six return-to-consensus variants (M60V/M180V/A334S/G373S/D268E/R366Q/Y379W (**17**)), the protein was significantly more thermally stable than the wild-type or Y379W alone (Table 2). This variant also had a modest decrease in turnover number for galactose (compared to Y379W (**15**)), but a three-fold increase in the catalytic efficiency (as reflected in the specificity constant) for galactose which results largely from the reduction in the Michaelis constant (Table 1). Combining M185L and Y379W (M185L/Y379W (**16**)) did not result in an increase in the substrate range (data not shown). The resulting protein had a four-fold decreased specificity constant for galactose (Table 1) and reduced thermal stability (Table 2) compared to the wild-type. However, it did have a significantly increased turnover and specificity constant for D-galactosamine (Table 4). Combining these two alterations with the six consensus changes (**18**) resulted in an enzyme with significantly increased (six-fold) turnover number for galactose (Table 1), but this effect did not occur with 2-deoxy-D-galactose or D-galactosamine as substrates. However, the catalytic efficiency for galactose ($k_{cat}/K_{m,gal}$) was reduced three-fold largely due to the

Table 4. Kinetic parameters of selected variants with substrates other than galactose

ENTRY NUMBER	SUGAR	VARIANT	$K_{M,ATP}$ (μM)	$K_{M,SUGAR}$ (μM)	K_{CAT} (s^{-1})	$K_{CAT}/K_{M,ATP}$ ($\text{MM}^{-1}\text{s}^{-1}$)	$K_{CAT}/K_{M,SUGAR}$ ($\text{MM}^{-1}\text{s}^{-1}$)
1	2-DEOXY-D-GALACTOSE	WILD-TYPE	43.2 ± 5.6	970 ± 130	6.8 ± 0.4	190 ± 10	7.1 ± 0.6
14		M185L	17.3 ± 15.5*	7400 ± 7300	0.4 ± 0.1***	23 ± 15**	0.06 ± 0.01***
15		Y379W	16.0 ± 3.6**	2200 ± 830	0.8 ± 0.1***	50 ± 8*	0.39 ± 0.06***
17		M60V/M180V/D268E/A334S/R366Q/G373S/Y379W	17.4 ± 6.0*	2000 ± 890	6.8 ± 0.1	390 ± 94**	3.1 ± 0.6***
18		M60V/M185L/M180V/D268E/A334S/R366Q/G373S/Y379W	4.0 ± 4.0***	300 ± 90	0.7 ± 0.1***	19.4 ± 4.0	2.3 ± 0.5***
1	D-GALACTOSAMINE	WILD-TYPE	16.2 ± 5.3	710 ± 180	1.7 ± 0.2	110 ± 25	2.9 ± 0.5
15		Y379W	3.7 ± 0.8	220 ± 55	0.2 ± 0.0	50 ± 10	0.3 ± 0.1
16		M185L/Y379W	2.7 ± 1.0	270 ± 54	25 ± 1***	950 ± 310**	91 ± 14***
17		M60V/M180V/D268E/A334S/R366Q/G373S/Y379W	6.6 ± 5.7	4800 ± 1000***	4.0 ± 0.9	600 ± 420	0.9 ± 0.1
18		M60V/M185L/M180V/D268E/A334S/R366Q/G373S/Y379W	4.4 ± 1.8	620 ± 170	0.3 ± 0.0	67 ± 24	0.4 ± 0.1
1	D-TALOSE	WILD-TYPE	170 ± 86	370 ± 370	2.0 ± 0.7	110 ± 15	2.6 ± 0.8
17		M60V/M180V/D268E/A334S/R366Q/G373S/Y379W	13.4 ± 8.3*	2200 ± 780	0.2 ± 0.1*	17.6 ± 7.6	0.2 ± 0.02**

All measurements taken using 70 nM enzyme as described in Materials and Methods. Parameters were estimated by fitting graphs of $V_{\text{max,app}}$ plotted against substrate concentration with Michaelis-Menten equation using nonlinear curve fitting and errors shown are standard errors calculated in this process. Significance calculated via one-way ANOVA with Dunnett's post-test analysis to compare with the wild-type enzyme's parameters. * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

increase in the Michaelis constant (Table 4). While its thermal stability is lower than the wild-type, it is higher than M185L/Y379W (**16**) (Table 2). This demonstrates that the combination of six consensus variants can be used to increase the thermal stability (and, in some cases, the catalytic activity) of catalytically interesting forms of the enzyme. However, it is interesting to note that the effects appear to be quite substrate-dependent (compare Table 1 and Table 4).

Conclusions

The return-to-consensus variants identified in this study increase the stability of human galactokinase when combined together. Thus, the consensus variants identified here are likely to be useful in the stabilisation of the broad specificity variants which may be discovered in future work. Since the enzyme containing all six return-to-consensus variants (**13**) is more stable than the wild-type (**1**) without compromising the catalytic activity, this variant would have potential for the development of enzyme replacement therapies for type II galactosemia. Interestingly each individual variant has either no effect on, or reduces, the thermal stability demonstrating that the variants have synergistic effects on the protein's structure. These results are in contrast with those obtained with other proteins, in which the effect of consensus mutations' effects on the stability and activity are often shown to be additive (i.e. single mutants have an effect, and the overall effect in a multiple mutant is essentially additive)^[17a, 18a].

Interestingly, the opposite effect of multiple consensus mutations is found for GALK activity. Some variants individually increased the turnover number of the enzyme, but combinations generally had little effect. Where turnover number was increased, this was generally accompanied by corresponding increases in the Michaelis constants. Therefore, these variants would require higher concentrations of substrates in order to achieve the increased turnover. The use of these variants in biocatalysis would be restricted to applications where high turnover was the priority and it was economically and chemically feasible to maintain

high substrate concentrations. The lack of additivity in the variants was surprising when compared to previous studies on consensus variants of other enzymes (e.g. ^[17a, 20]). This currently remains unexplained, but we hypothesise that the high sensitivity of human galactokinase to variation may be part of the cause. We also observed poor correlation between increased stability (as judged by resistance to thermal or chaotropic denaturation) and increased catalytic turnover. The mobility of enzymes is closely related to their catalytic properties ^[21] and it does not automatically follow that more stable proteins will be more active. Too little flexibility at key sites may hinder activity. Nevertheless, we propose that, based on this work, it may be possible to use the more stable variants as starting points for the identification of more catalytically active or promiscuous galactokinases.

Molecular dynamics studies provide some partial explanations of the experimental data. Two key salt bridges appear to be critical to the increased stability and further alternation of the four residues involved might also generate variants with altered stability. The MD studies also re-emphasise the critical role of residues 232-240 on the structure, dynamics and catalytic activity of the enzyme. We previously identified this as a critical region in the catalytic mechanism of the enzyme ^[19]. We hypothesise that direct alteration of these residues may affect the activity or specificity of the kinase. In particular we propose that alterations designed to modulate flexibility of the protein in this region will impact on catalytic activity.

Overall, the data presented here demonstrate that return-to-consensus variants of human galactokinase can improve the stability or the turnover of the enzyme. Previous studies which have recapitulated disease-associated variants or have directly altered active site residues have often resulted in insoluble and/or inactive protein ^[5, 12b, 22]. While alteration of Tyr-379 did expand the substrate range of the enzyme, there were also substantial reductions in activity ^[8]. The consensus approach opens up the possibility of designing new variants by introducing further changes into human galactokinase against the background of the six return-to-consensus variants. This is more likely

to result in soluble protein with reasonable activity. Furthermore, a similar approach could be applied to bacterial galactokinases which appear to have naturally greater substrate ranges.

Experimental Section

Consensus analysis of human GALK1

For consensus analyses, 25 protein sequences which were highly homologous to that of human GALK1 (P51570, NP_000145.1) belonging to animals were retrieved online using BLASTP and aligned using ClustalW^[23]. This alignment of 26 sequences was used to identify consensus variants, as those that place the most common amino acid at a given position in the alignment (the consensus amino acid) instead of the human amino acid.

Expression, purification and mutagenesis of human GALK1

Recombinant human wild-type galactokinase was expressed in, and purified from, *E. coli* as previously described^[19, 24]. For site-directed mutagenesis, the plasmid was mutated using the QuikChange method and the full sequence verified (GATC Biotech, London, UK)^[25]. The resulting variant proteins were expressed and purified using the same methods as for the wild-type. Purifications were monitored using SDS-PAGE.

Galactokinase kinetics

The rate of reaction was measured by coupling the production of ADP to the reactions catalysed by pyruvate kinase and lactate dehydrogenase^[26]. This results in the stoichiometric consumption of one molecule of NADH for each sugar molecule phosphorylated, which can be monitored by measuring the decrease in absorbance at 340 nm. Reactions were carried out in triplicate at 37 °C in 50 mM Hepes-OH buffer (pH 7.5) containing 150 mM NaCl, 5 mM MgCl₂, 0.4 mM phosphoenolpyruvate, 1 mM NADH, 10%(v/v) glycerol, pyruvate kinase (7.5 U), lactate dehydrogenase (10 U) and varying, sub-saturating amounts of ATP and galactose. Enzyme (70 nM) was added to initiate the reaction which was monitored using a Thermo Scientific Multiskan Spectrum plate reader and the total reaction volume was 160 µl. Kinetic constants were determined by arraying reactions in an eight by ten grid in order to determine apparent Michaelis constants ($K_{m,app}$) and apparent turnover numbers ($k_{cat,app}$) over a range of sub-saturating concentrations of ATP and galactose using non-linear curve fitting to the Michaelis-Menten equation as implemented in GraphPad Prism 6.0 (GraphPad Software, CA, USA). These were then used to determine the absolute values of the turnover number (k_{cat}), the Michaelis constant for galactose ($K_{m,gai}$) and the Michaelis constant for ATP ($K_{m,ATP}$) as previously described^[19, 27]. In brief, plots of $k_{cat,app}$ against substrate concentration were fitted using equations 1a and 1b to determine absolute values.

$$k_{cat,app} = \frac{k_{cat}[ATP]}{K_{m,ATP} + [ATP]} \quad (1a)$$

$$k_{cat,app} = \frac{k_{cat}[Galactose]}{K_{m,gai} + [Galactose]} \quad (1b)$$

The specificity constants for galactose and ATP ($k_{cat}/K_{m,gai}$ and $k_{cat}/K_{m,ATP}$) were estimated directly in Prism as part of the curve-fitting process. A similar method was used for sugars other than galactose.

Analytical methods

Protein concentrations were estimated by the method of Bradford using bovine serum albumin as a standard^[28]. The thermal stability of galactokinase was measured by differential scanning fluorimetry (DSF) using Sypro Orange as the fluorescent probe as previously described^[19]. Native gel electrophoresis was used to determine the proteins' resistance to chaotropic denaturation as previously described^[19, 24].

Molecular dynamics simulations

Molecular dynamics (MD) simulations were performed essentially as previously described^[19, 29]. Chain A from the structure of human galactokinase (1WUU;^[9]) was prepared by adding Ser230 and Leu231, altering selenomethionines to methionine and AMP.PNP to ATP. RESP charges of ligands were calculated using the Gaussian 09 package and antechamber present in Amber 10^[30]. Prior to simulation the protein and ligands were soaked in a TIP3P solvent box and a maximum distance boundary of 8 Å set. All preparations and simulations were carried out using the Amber 10 software package and the Amber Parm 99 forcefield, the variants were generated using Biovia Discovery studio^[31]. The system was neutralised by addition of sodium ions prior to minimisation. Minimisation was carried out using Cartesian restraints and the system subjected to 1250 cycles of conjugate gradient method followed by 1250 steps of deepest decent method at 0.5 kcal mol⁻¹ Å⁻²^[31b]. The system was then heated from 0 K to 300 K over 50 ps with a collision frequency of 0.5 ps⁻¹ according to the Langevin dynamics method^[32]. Equilibration was carried out at 300 K for 50 ps using an NVT ensemble. Production simulations were performed as described by^[19]. Each simulation was executed for 7 ns at 1 atm and 300 K. The timestep was set a 1 fs, the cut off for Van der Waals interactions at 10 Å, monitored using the particle mesh Ewald method, and the bond lengths involving hydrogen atoms were constrained using the SHAKE method^[33].

Molecular dynamics analysis

The simulations were analysed as described in^[19]. Physical characteristics of the simulation were monitored using the perl feature in Amber and graphs produced using GraphPad prism version 5.03 for windows (GraphPad Software, San Diego California USA) (Supplementary Figure S1). The ptraj command of Amber was used to perform cluster analysis, secondary structure analysis and RMSF calculations. Structures were visualised using Biovia Discovery studio and trajectories in VMD^[34]. The timeline feature and normal mode wizards in VMD were used to visualise secondary structure and carry out principal component analysis (PCA) respectively. Both methods were carried out on 1 ns of equilibrated trajectory, determined by RMSD over time^[35].

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). ALP acknowledges the financial support of Junta de Andalucia (grant number: P11-CTS-07187).

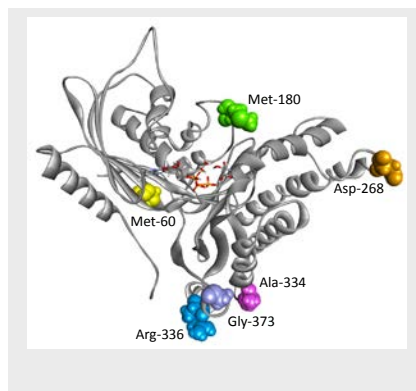
Keywords: biocatalysis • molecular dynamics • protein engineering • GHMP kinase • return-to-consensus variant

- [1] H. M. Holden, J. B. Thoden, D. J. Timson, R. J. Reece, *Cell Mol Life Sci* **2004**, *61*, 2471-2484.
- [2] a) R. Caputto, L. F. Leloir, R. E. Trucco, C. E. Cardini, A. C. Paladini, *J Biol Chem* **1949**, *179*, 497-498; b) C. E. Cardini, L. F. Leloir, *Arch Biochem Biophys* **1953**, *45*, 55-64.
- [3] J. Yang, D. Hoffmeister, L. Liu, X. Fu, J. S. Thorson, *Bioorg Med Chem* **2004**, *12*, 1577-1584.
- [4] H. Kristiansson, D. J. Timson, *Curr Biotech* **2012**, *1*, 148-154.
- [5] D. J. Timson, R. J. Reece, *BMC Biochem* **2003**, *4*, 16.
- [6] a) J. Yang, X. Fu, Q. Jia, J. Shen, J. B. Biggins, J. Jiang, J. Zhao, J. J. Schmidt, P. G. Wang, J. S. Thorson, *Org Lett* **2003**, *5*, 2223-2226; b) L. Lee, S. Kinoshita, H. Kumagai, T. Tochikura, *Agric Biol Chem* **1980**, *44*, 2961-2966; c) L. Li, Y. Liu, W. Wang, J. Cheng, W. Zhao, P. Wang, *Carbohydr Res* **2012**, *355*, 35-39; d) 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, *Chembiochem* **2017**, *In press*; e) L. P. Conway, F. F. Liu, Q. Li, J. Voglmeir, *Carbohydr Res* **2017**, *455*, 39-44; f) L. Caputi, M. Rejzek, T. Louveau, E. C. O'Neill, L. Hill, A. Osbourn, R. A. Field, *Bioorg Med Chem* **2013**, *21*, 4762-4767; g) J. Yan, X. Chen, F. Wang, H. Cao, *Org Biomol Chem* **2013**, *11*, 842-848; h) 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, *Carbohydr Res* **2011**, *346*, 2421-2425; i) Y. Zou, W. Wang, L. Cai, L. Chen, M. Xue, X. Zhang, J. Shen, M. Chen, *Bioorg Med Chem Lett* **2012**, *22*, 3540-3543.
- [7] a) D. Hoffmeister, J. S. Thorson, *Chembiochem* **2004**, *5*, 989-992; b) D. Hoffmeister, J. Yang, L. Liu, J. S. Thorson, *Proc Natl Acad Sci USA* **2003**, *100*, 13184-13189; c) J. S. Thorson, W. A. Barton, D. Hoffmeister, C. Albermann, D. B. Nikolov, *Chembiochem* **2004**, *5*, 16-25; d) J. Yang, X. Fu, J. Liao, L. Liu, J. S. Thorson, *Chem Biol* **2005**, *12*, 657-664; e) J. Yang, L. Liu, J. S. Thorson, *Chembiochem* **2004**, *5*, 992-996.
- [8] H. Kristiansson, D. J. Timson, *Chembiochem* **2011**, *12*, 2081-2087.
- [9] J. B. Thoden, D. J. Timson, R. J. Reece, H. M. Holden, *J Biol Chem* **2005**, *280*, 9662-9670.
- [10] M. McAuley, H. Kristiansson, M. Huang, A. L. Pey, D. J. Timson, *Biochem.Soc.Trans.* **2015**, *44*, 116-122.
- [11] a) D. Stambolian, *Ophthalmology* **1987**, *94*, 91-92; b) D. Stambolian, Y. Ai, D. Sidjanin, K. Nesburn, G. Sathe, M. Rosenberg, D. J. Bergsma, *Nat Genet* **1995**, *10*, 307-312; c) D. Stambolian, V. Scarpino-Myers, R. C. Eagle, Jr., B. Hodes, H. Harris, *Invest Ophthalmol Vis Sci* **1986**, *27*, 429-433; d) D. J. Timson, *Curr Enz Inhib* **2007**, *3*, 77-94; e) O. Thalhammer, R. Gitzelmann, M. Pantlitschko, *Pediatrics* **1968**, *42*, 441-445.
- [12] a) R. Gitzelmann, *Pediatr Res* **1967**, *1*, 14-23; b) D. J. Timson, R. J. Reece, *Eur J Biochem* **2003**, *270*, 1767-1774.
- [13] D. J. Timson, *Gene* **2016**, *589*, 133-141.
- [14] A. M. Bosch, H. D. Bakker, A. H. van Gennip, J. V. van Kempen, R. J. Wanders, F. A. Wijburg, *J Inherit Metabol Dis* **2002**, *25*, 629-634.
- [15] J. B. Hennermann, P. Schadewaldt, B. Vetter, Y. S. Shin, E. Monch, J. Klein, *J Inherit Metabol Dis* **2011**, *34*, 399-407.
- [16] E. Beutler, *Trends Biochem Sci* **1981**, *6*, 95-97.
- [17] a) N. Mesa-Torres, C. Yunta, I. Fabelo-Rosa, J. M. Gonzalez-Rubio, J. M. Sanchez-Ruiz, E. Salido, A. Albert, A. L. Pey, *The Biochemical journal* **2014**, *462*, 453-463; b) M. Lehmann, L. Pasamontes, S. F. Lassen, M. Wyss, *Medchimica et biophysica acta* **2000**, *1543*, 408-415.
- [18] a) E. Medina-Carmona, J. E. Fuchs, J. A. Gavira, N. Mesa-Torres, J. L. Neira, E. Salido, R. Palomino-Morales, M. Burgos, D. J. Timson, A. L. Pey, *Hum Mol Genet* **2017**, *26*, 3531-3544; b) I. G. Munoz, B. Morel, E. Medina-Carmona, A. L. Pey, *FEBS Lett* **2017**, *591*, 2826-2835.
- [19] M. McAuley, M. Huang, D. J. Timson, *Biochim Biophys Acta* **2017**, *1865*, 321-328.
- [20] a) N. Amin, A. D. Liu, S. Ramer, W. Aehle, D. Meijer, M. Metin, S. Wong, P. Gualfetti, V. Schellenberger, *Protein Eng Des Sel* **2004**, *17*, 787-793; b) B. J. Sullivan, T. Nguyen, V. Durani, D. Mathur, S. Rojas, M. Thomas, T. Syu, T. J. Magliery, *J Mol Biol* **2012**, *420*, 384-399.
- [21] a) M. McAuley, D. J. Timson, *Appl Biochem Biotechnol* **2016**, *181*, 83-90; b) N. M. Goodey, S. J. Benkovic, *Nat Chem Biol* **2008**, *4*, 474-482.
- [22] a) F. Sanguolo, M. Magnani, D. Stambolian, G. Novelli, *Hum Mutat* **2004**, *23*, 396; b) B. Jojart, M. Szori, R. Izsak, I. Marsi, A. Laszlo, I. G. Csizmadia, B. Viskolcz, *J Mol Model* **2011**, *17*, 2639-2649.
- [23] U. Bodenhofer, E. Bonatesta, C. Horejs-Kainrath, S. Hochreiter, *Bioinformatics* **2015**, *31*, 3997-3999.
- [24] C. F. Megarity, M. Huang, C. Warnock, D. J. Timson, *Bioorg Chem* **2011**, *39*, 120-126.
- [25] W. Wang, B. A. Malcolm, *BioTechniques* **1999**, *26*, 680-682.
- [26] a) A. Platt, H. C. Ross, S. Hankin, R. J. Reece, *Proc Natl Acad Sci USA* **2000**, *97*, 3154-3159; b) M. R. Heinrich, *J Biol Chem* **1964**, *239*, 50-53.
- [27] D. J. Timson, R. J. Reece, *Biochimie* **2002**, *84*, 265-272.
- [28] M. M. Bradford, *Anal Biochem* **1976**, *72*, 248-254.
- [29] M. Huang, X. Li, J. W. Zou, D. J. Timson, *Biochemistry* **2013**, *52*, 4858-4868.
- [30] M. J. Frisch, G. Trucks, H. Schlegel, G. Scuseria, M. Robb, J. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. Petersson, in *Gaussian Inc., Wallingford, CT, 2009*.
- [31] a) D. Case, T. Darden, T. Cheatham, C. Simmerling, J. Wang, R. Duke, R. Luo, M. Crowley, R. Walker, W. Zhang, in *San Francisco, 2008*; b) P. K. Weiner, P. A. Kollman, *J Comp Chem* **1981**, *2*, 287-303; c) J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J Comp Chem* **2004**, *25*, 1157-1174.
- [32] R. W. Pastor, B. R. Brooks, A. Szabo, *Mol Phys* **1988**, *65*, 1409-1419.
- [33] a) T. Darden, D. York, L. Pedersen, *J Chem Phys* **1993**, *98*, 10089-10092; b) J. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J Comp Phys* **1977**, *23*, 327-341.
- [34] W. Humphrey, A. Dalke, K. Schulten, *J Mol Graph* **1996**, *14*, 33-38, 27-38.
- [35] A. Bakan, L. M. Meireles, I. Bahar, *Bioinformatics* **2011**, *27*, 1575-1577.

Entry for the Table of Contents

FULL PAPER

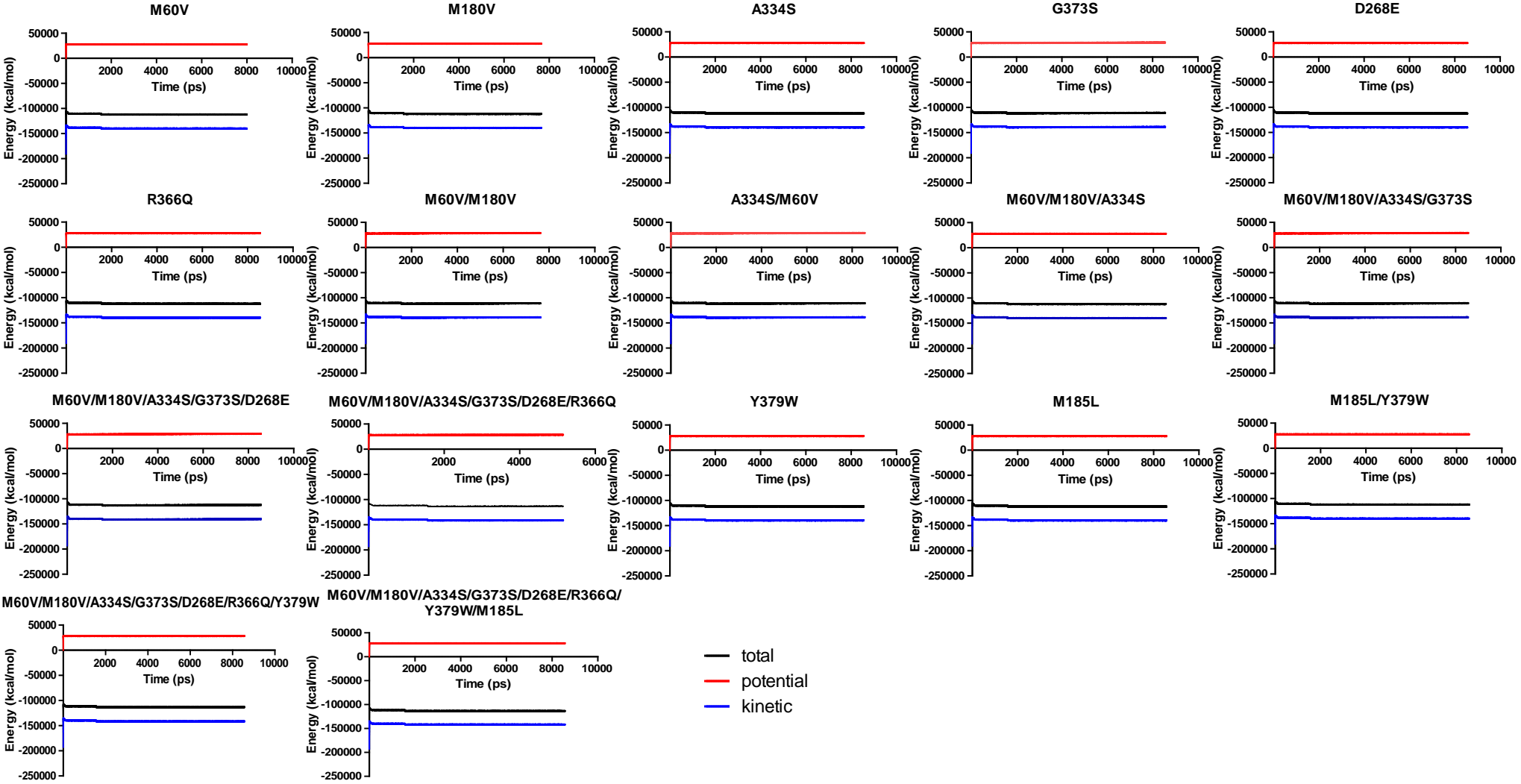
Return-to-consensus variants often have improved catalytic properties. Here, we show that six variants in human galactokinase improve the stability and turnover of the enzyme. These consensus variants will be useful to stabilise more promiscuous forms of the enzyme or if the protein were to be used in enzyme replacement therapy to treat type II galactosemia.



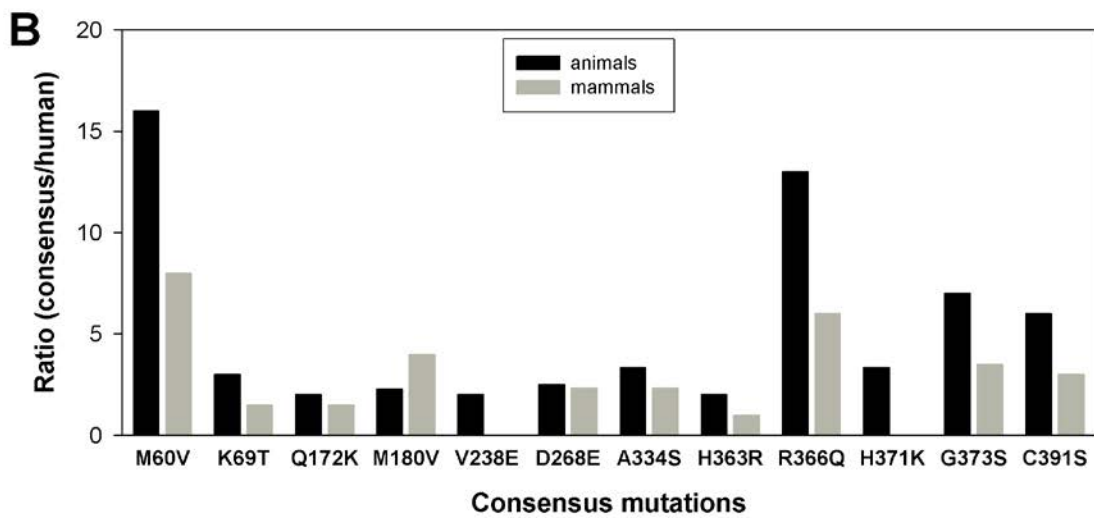
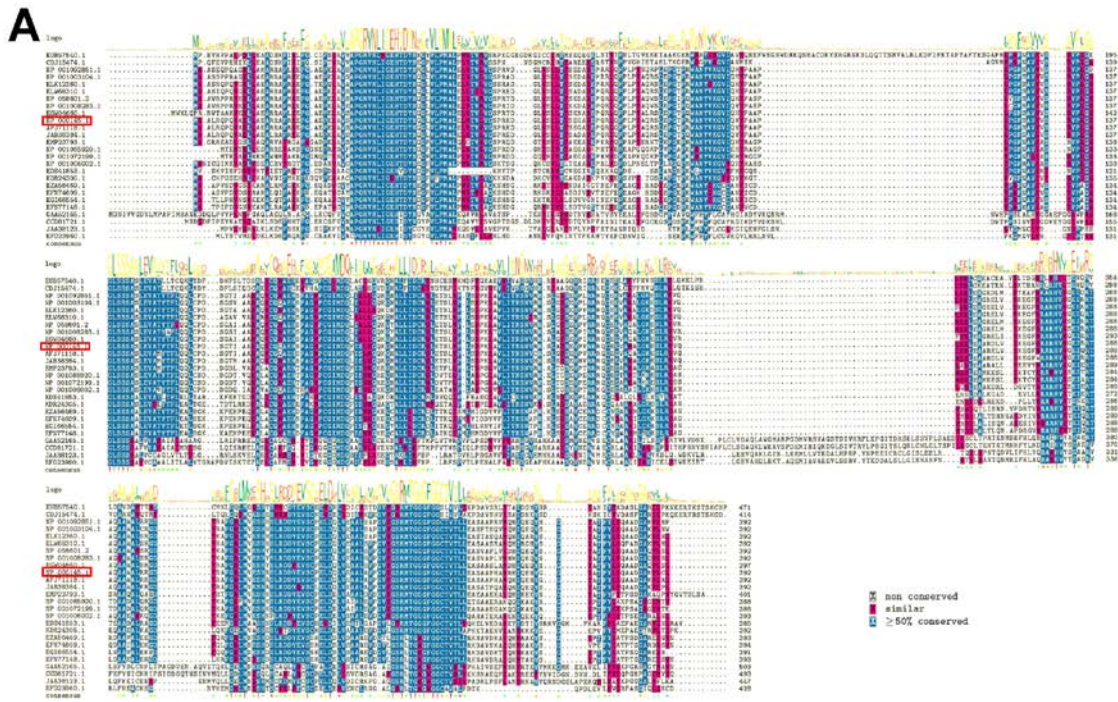
Margaret McAuley, Noel Mesa-Torres, Aisling McFall, Sarah Morris, Meilan Huang, Angel L. Pey and David J. Timson*

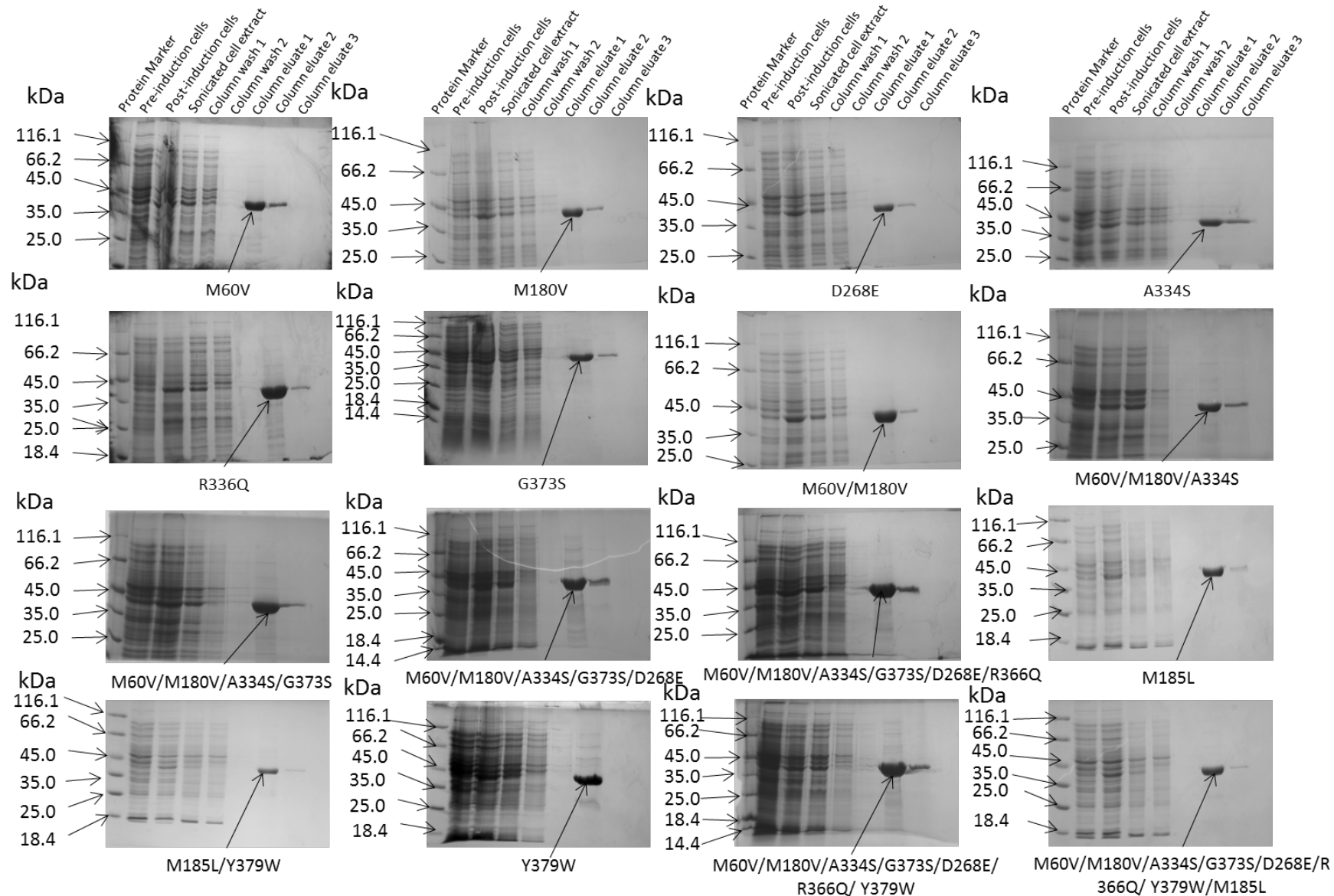
Page No. – Page No.

Improving the activity and stability of human galactokinase for therapeutic and biotechnological applications

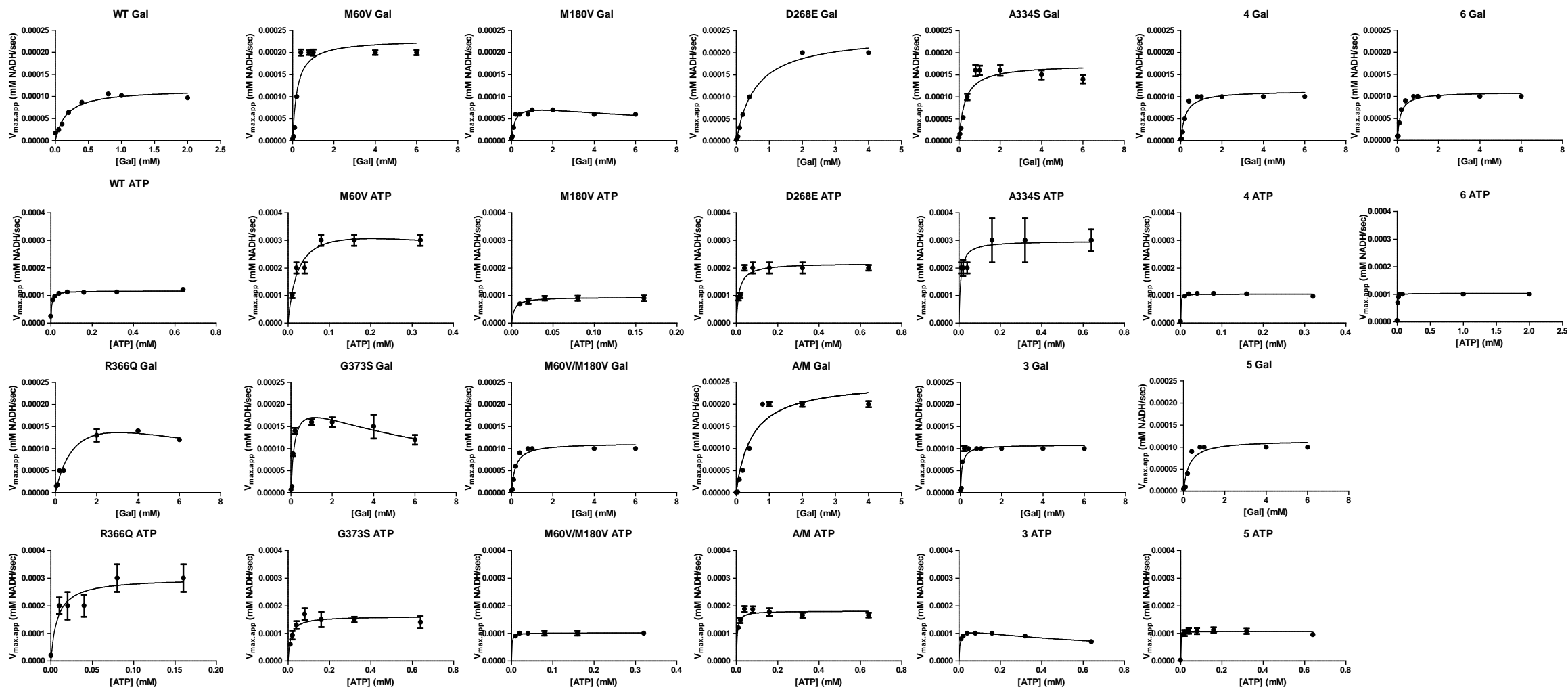


Supplementary Figure S2: Identification of consensus mutations for human galactokinase. (A) Sequence alignment of animal galactokinase sequences. Human galactokinase is indicated with a red box. The consensus amino acids are indicated as coloured logos, and the size of the amino acid corresponding to the relative frequency in the alignment. (B) Consensus mutations identified using the alignment of animal sequences. For comparison, the consensus ratio using sequences from animals or mammals is displayed.





Supplementary Figure S3: Expression and purification of consensus variants analysed by 10% SDS-PAGE. Samples are labelled above the gels, lane 1- protein markers (masses shown to the left of the gels in kDa): lane 2- samples of cell grow up; lane 3- sample taken after induction with IPTG and prior to harvesting by centrifugation; lanes 4 and 5- flow-through of washes 1 and 2 respectively; lanes 6, 7 and 8- flow-through of elutions 1,2 and 3 respectively through the column. All galactokinase variants were approximately 43 kDa as expected.



Supplementary Figure S4: Steady state Michaelis-Menten kinetics for the galactokinase variants. Apparent V_{max} values were plotted against substrate concentration for each variant (0.07 μ M enzyme) as described in Materials and Methods.