In silico prediction of the effects of mutations in the human triose phosphate isomerase gene: towards a predictive framework for TPI deficiency

Conor Oliver[1,2] and David J. Timson[1,3*]

[1] School of Biological Sciences and [2]School of Medicine, Dentistry and Biomedical Sciences, Queen's University Belfast, Medical Biology Building, 97 Lisburn Road, Belfast, BT9 7BL. 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

Triose phosphate isomerase (TPI) deficiency is a rare, but highly debilitating, inherited metabolic disease. Almost all patients suffer severe neurological effects and the most severely affected are unlikely to live beyond early childhood. Here, we describe an *in silico* study into well-characterised variants which are associated with the disease alongside an investigation into 79 currently uncharacterised TPI variants which are known to occur in the human population. The majority of the disease-associated mutations affected amino acid residues close to the dimer interface or the active site. However, the location of the altered amino acid residue did not predict the severity of the resulting disease. Prediction of the effect on protein stability using a range of different programs suggested a relationship between the degree of instability caused by the sequence variation and the severity of the resulting disease. Disease-associated variations tended to affect well-conserved residues in the protein's sequence. However, the degree of conservation of the residue was not predictive of disease severity. The majority of the 79 uncharacterised variants are potentially associated with disease since they were predicted to destabilise the protein and often occur in well-conserved residues. We predict that individuals homozygous for the corresponding mutations would be likely to suffer from TPI deficiency.

Keywords: TPI deficiency; *in silico* prediction; inherited metabolic disease; protein stability; disease-associated variant; glycolytic enzyme

Introduction

Triose phosphate isomerse (TPI; EC 5.3.1.1) catalyses a key step in the glycolytic pathway, namely the interconversion of dihydroxyacetone phosphate (DHAP) and glyceraldehyde 3-phosphate (GLA3P), the two products of the splitting of fructose 1,6-*bis*phosphate catalysed by adolase (EC 4.1.2.13). This step is necessary, since only GLA3P can be processed by the subsequent reactions in the pathway. Thus, TPI ensures that the three carbon atoms in DHAP can be utilised in glycolysis and subsequent pathways. In addition to this metabolic role it has been proposed that TPI also functions in sperm-egg recognition in mammals and may have non-metabolic roles in nerve cells (Auer et al., 2004; Petit et al., 2014; Roland et al., 2013; Roland et al., 2016).

TPI is an example of a 'kinetically perfect' enzyme since its activity is limited only by the rates of diffusion of substrate and product molecules (Stroppolo et al., 2001). In humans, TPI is encoded by a single gene (*TPI1*) located at chromosome 12p13 and is expressed in all tissues. TPI is expressed in almost every organism from bacteria to humans and its amino acid sequence is highly conserved. In healthy individuals, this enzyme exists as a stable homodimer of two 27 kDa subunits each of which consists of 248 amino acids. Although each monomer has a full set of catalytic residues, TPI is only active its dimeric form (De La Mora-De La Mora et al., 2013). TPI monomers have been shown to be thermodynamically unstable, demonstrating the crucial role of the dimeric form in both enzyme stability and catalytic function (Mainfroid et al., 1996).

First described in 1964, TPI deficiency (OMIN #615512) is a rare, multisystem disorder with autosomal recessive inheritance and is the most severe of the known glycolytic enzymopathies. However, clinical presentation varies between patients (Schneider et al., 1964; Schneider et al., 1965). The most common phenotypes include nonspherocytic haemolytic anaemia (often presenting as neonatal jaundice) and, in all but two reported cases, progressive neuromuscular impairment that almost inevitably culminates in the death of the patient in infancy or early childhood (Sarper et al., 2013). Other common phenotypes include cardiomyopathy and an increased susceptibility to infection (Schneider, 2000). These physical characteristics are generally accompanied by a decrease in enzyme activity in all tissues and an increase in the levels of DHAP in erythrocytes (Orosz et al., 2006). Although prenatal detection has become available, there has been no effective therapy developed for TPI deficiency and treatment is generally only supportive (Pekrun et al., 1995; Poinsot et al., 1986; Rosa et al., 1986). A number of studies have reported some limited success with enzyme replacement therapy (Ationu and Humphries, 1998; Ationu et al., 1997; Ationu et al., 1999a; Ationu et al., 1999b).

In total, fewer than 50 cases of TPI deficiency have been reported in the literature. Intriguingly, population studies have revealed a surprisingly high frequency of healthy heterozygote carriers - much

higher than would be suggested by the comparative scarcity of TPI deficiency cases (Eber et al., 1984). The frequency of heterozygosity has been estimated at 9 in 1,713 among Caucasians and 7 in 168 among African Americans (Watanabe et al., 1996). One possible explanation for this apparent discrepancy is that many foetuses homozygous or compound heterozygous for TPI deficiency mutations often die *in utero*. This is supported by studies performed on mice that demonstrated that mutations which resulted in TPI variants which were catalytically inactive (null alleles) caused early prenatal lethality in homozygotes for such mutations (Merkle and Pretsch, 1989). This theory is further substantiated by anecdotal evidence that mothers of TPI deficiency cases often have a personal and/or family history of frequent miscarriages (Valentin et al., 2000). Alternatively, it may be the case that heterozygotes are less fertile than wild-type homozygotes. Some evidence suggests that the tail regions of human spermatozoa are dependent on glycolysis for the majority of their ATP generation (du Plessis et al., 2015). It is also possible that there are phenotypes associated with homozygosity which are not currently known to be associated with TPI deficiency.

The underlying molecular basis of TPI deficiency is still a matter of some debate. Early propositions that the reduced TPI activity would lead to a reduction in the rate of glycolysis, and subsequently the rate of ATP synthesis, have been largely disproven due to the relative paucity of metabolic abnormalities other than a markedly increased DHAP level (Hollan et al., 1993; Orosz et al., 1996; Schneider, 2000). It was postulated that when DHAP accumulated to such high levels as a result of decreased TPI activity, it would inhibit the action of myo-inositol-3-phosphate synthase, which would contribute to the clinical manifestations of this disease (Shi et al., 2005). Although there has been very little indication that DHAP itself is toxic, it does decompose non-enzymatically to form methylglyoxal, a potent glycating agent that can modify proteins and DNA to form advanced glycation end products (AGEs) (Li et al., 2008). The accumulation of these AGEs can lead to oxidative stress, DNA damage and, consequently, to apoptosis. Indeed methylglyoxal has already been shown to be neurotoxic and therefore may contribute to the neurological impairment seen in many TPI deficiency patients (Ahmed et al., 2003; de Arriba et al., 2006; de Arriba et al., 2007). Abnormal dimerisation behaviour of disease-associated TPI variants has been suggested as a potential contributing factor in the pathogenesis of TPI deficiency (Ralser et al., 2006; Rodriguez-Almazan et al., 2008; Roland et al., 2016). In recent years the suggestion has arisen that this may be as a consequence of the formation of toxic protein aggregates in the brain due to the misfolding of the variant TPI (Olah et al., 2002; Olah et al., 2005)). This "conformational disease" hypothesis of TPI deficiency is supported by a growing body of evidence.

A bioinformatics investigation was undertaken to better understand the functional and structural contribution of molecular alterations in TPI and how these correlate to the associated severity. The results of these investigations were then compared against the relative severity of their associated

clinical patterns to develop a predictive framework for TPI deficiency. The parameters of such predictions were then applied to a number of uncharacterised mutants from the NCBI dbSNP database in order to attempt to predict their severity if they were to present in a homozygote/compound heterozygote in the future.

Such methods have already been employed, successfully, in order to create predictive frameworks for a number of other diseases, including type I and type III galactosaemia (Facchiano and Marabotti, 2010; McCorvie and Timson, 2013), hyperargininemia (Carvalho et al., 2012)(Carvalho *et al.*, 2012), mevalonate kinase deficiency (Browne and Timson, 2015) and apparent mineralocorticoid excess (Manning et al., 2010) among others (Timson, 2015). The development of an effective, accurate and reliable predictive framework could aid in providing genetic counselling to heterozygote carriers of previously uncharacterised mutations or in estimating prognoses for patients suffering from TPI deficiency.

Materials and Methods

Datasets

The identification and gathering of information on the known, investigated missense mutations associated with TPI deficiency were carried out by searching the literature on the NCBI PubMed database (http://www.ncbi.nlm.nih.gov/pubmed). Where available, the associated clinical symptoms seen in the relevant patient(s), along with pertinent biochemical results, including TPI activity and DHAP levels measured in red blood cells, and genetic background of the propositus was recorded for each mutant (Table 1; Supplementary Table S1; see also Table 1 in (Orosz et al., 2006)). Reported missense mutants from exome sequencing that had not been previously investigated were identified from the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/).

Structural analysis

Structural studies were based on the crystal structure of human TPI (PDB: 4POC; (Roland et al., 2015; Roland et al., 2013)) obtained from the RCSB protein data bank (http://www.rcsb.org/pdb/home/home.do)(Berman et al., 2002; Berman et al., 2000). Secondary structural elements were defined according to Supplementary Table S2. The structures were visualised using PyMol (http://www.pymol.org) and were computationally solvated and energy minimised using YASARA (http://www.yasara.org) (Krieger et al., 2009). Sequence variations were

introduced *in silico* using the mutagenesis function in PyMol. These variant forms of the protein were subsequently energy minimised using YASARA. Thus, where possible, all structural comparisons were made using minimised structures.

Multiple sequence alignment and residue conservation of TPI deficiency mutants

Multiple sequence alignment was carried out using ClustalOmega (Sievers et al., 2011). TPI sequences were identified from the NCBI protein database (http://www.ncbi.nlm.nih.gov/protein). A total of 119 sequences were obtained including 40 from mammals, 12 from bacteria, ten from birds, ten from insects, ten from plants, ten from fungi, ten from protists, eight from reptiles, seven from fish and two from amphibians (Supplementary Table S3). This resulting alignment was then used in combination with the Scorecons server (http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/valdar/scorecons_server.pl) (Valdar, 2002) and was used to calculate the degree of conservation based on the Valdar01 score.

Prediction of biochemical effects of each missense mutation on TPI deficiency

Prediction of the effect of individual mutations on protein stability and/or the likelihood of it leading to disease was carried out using a variety of web servers which are freely available. I-mutant 3.0 (http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi) reports the predicted effect of a specific mutation on protein stability using either a structure file (pdb file) or the amino acid sequence (Capriotti et al., 2005; Capriotti et al., 2008). It also reports the relative solvent accessibility value and the reliability index of the reported result from 1 to 10 with 10 being very reliable and 1 being very unreliable. A pH of 7.5 and a temperature of 37 ^{o}C were used in order to replicate human cellular conditions. The mutation Cutoff Scanning Matrix (mCSM) (http://bleoberis.bioc.cam.ac.uk/mcsm/) (Pires et al., 2014) reports the protein stability change (ΔΔG) in kcal.mol^{-1} for each mutation on one chain of the protein (in this case chain A) using the energy minimised wildtype pdb file. The SDM server (http://mordred.bioc.cam.ac.uk/~sdm/sdm.php) reports the pseudo ΔΔG in kcal.mol^{-1}, that predicts its effect on protein stability ranging from highly stabilising to neutral to highly destabilising and a number of degrees in between as well as whether or not the mutation is predicted to be disease-associated (Worth et al., 2011). A meta-predictor of disease causing variants, Meta-SNP (Capriotti et al., 2013) (http://snps.biofold.org/meta-snp/) reports the results from a number of prediction servers. PANTHER, PhD-SNP and SNAP all report a normalised value between zero and one and if this value is greater than 0.5 the mutation is predicted to cause

disease. SIFT reports a positive value which, if greater than 0.05, is predictive of a mutation being neutral. From these values, Meta-SNP then calculates a meta-result between zero and one which, again, if greater than 0.5 predicts that the mutation will be disease associated. Predict-SNP 1.0 (Bendl et al., 2014) (http://loschmidt.chemi.muni.cz/predictsnp/) reports the qualitative results (deleterious or neutral), along with the percentage predicted accuracy of the result, for eight prediction tools - MAPP (Stone and Sidow, 2005), PhD-SNP (Capriotti et al., 2006), PolyPhen-1, PolyPhen-2 (Adzhubei et al., 2010), SIFT (Kumar et al., 2009), SNAP (Bromberg and Rost, 2007), nsSNP-Analyser (Bao et al., 2005) and PANTHER (Thomas et al., 2003) - and then combines the results for the six best performing tools (the former six in this list) into a consensus classifier in the form of Predict-SNP. Using sequence information only, MUpro (http://mupro.proteomics.ics.uci.edu/) (Cheng et al., 2006) qualitatively predicts protein stability changes (Increase/Decrease) for single amino acid mutations using two methods - support vector machines and neural networks - along with the associated confidence score. The Cologne University protein stability analysis tool (CUPSAT) (http://cupsat.tu-bs.de/) reports the predicted ΔΔG for each mutation affecting one chain as well as the predicted effect of the mutation on overall stability (stabilising/destabilising/no change) and torsion (favourable/unfavourable/no change) (Parthiban et al., 2006). BeAtMuSiC (http://babylone.ulb.ac.be/beatmusic/) evaluates the impact of mutations on protein-protein binding affinity based on protein structure and also reports the solvent accessibility in complex (Dehouck et al., 2013). nsSNPAnalyser (http://snpanalyzer.uthsc.edu) also qualitatively predicts if a mutation is disease-associated or neutral (Bao et al., 2005). TANGO predicts the likelihood of a protein to aggregate and reports this prediction using a variety of scores (Fernandez-Escamilla et al., 2004b). The value deemed most relevant to this study was the AGG Score, which represents the tendency for β-sheet aggregation. A summary spreadsheet of the key findings from each server for each disease-associated variant is provided as Supplementary Table S1.

The aggregation potential of the wild-type and variant proteins was predicted using several servers. TANGO (http://tango.crg.es/) (Fernandez-Escamilla et al., 2004a; Linding et al., 2004; Rousseau et al., 2006) determines an aggregation score based on the protein sequence, with higher scores indicating a greater tendency for aggregation. This program also returns a prediction of the likelihood of forming amyloid. Predictions were carried out with unmodified N- and C-termini, at pH 7.5, temperature of 298 K, ionic strength of 150 mM and a protein concentration of 10 μM. CamSol (http://www-mvsoftware.ch.cam.ac.uk/index.php/camsolintrinsic) also makes predictions about the likely soluble or aggregated state of a protein, returning an intrinsic solubility score; the higher this score, the more soluble the protein (Sormanni et al., 2015). Zyggregator (http://www-mvsoftware.ch.cam.ac.uk/index.php/zyggregator) estimates the sequence-dependent properties of

proteins returning a score; a high score indicates increased, predicted amyloid propensity (Tartaglia et al., 2008). Both CamSol and Zyggregator were run at pH 7. AGGRESCAN (http://bioinf.uab.es/aggrescan/) identifies the number of aggregation "hotspots" in proteins and also returns an "area above the threshold" with higher values suggesting a greater likelihood of aggregation (Conchillo-Sole et al., 2007). FoldAmyloid (http://bioinfo.protres.ru/fold-amyloid/) predicts the number of regions in a protein with amyloid-forming potential and also returns an overall score for the protein (Garbuzynskiy et al., 2010). It was deployed using the default parameters, i.e. an averaging frame of five residues and a threshold score of 21.4. The scores for each disease-associated variant from the aggregation predictions were normalised against the wild-type score such that a number higher than 1.0 indicated increased likelihood of aggregation. The mean, normalised score was then calculated for each variant and compared by ANOVA (using GraphPad Prism 6.0). A similar treatment was used with the amyloid propensity scores.

Results

TPI deficiency variants can be arranged into four groups depending on severity of their associated clinical patterns

Two variants (Ile170Val and Phe240Leu) were classified as "pathological (less severe)" since these were the only two which are associated with patients who did not develop neurological symptoms - the main cause of death in TPI deficiency (Arya et al., 1997; Chang et al., 1993; Schneider, 2000; Valentin et al., 2000). Phe240Leu is a somewhat unusual case because two brothers were reported to have TPI deficiency resulting from this variant. Although one brother did display some neuromuscular impairment, because the impairment was relatively minor and did not develop until the age of 12 (relatively late compared to more 'stereotypical' forms of the disease), as well as the fact that both brothers have survived well into adulthood (Chang et al., 1993), this mutation was classified as "pathological (less severe)".

Two variants (Glu104Asp and Phe240Ser) were classified as "pathological (severe)" due to the severity of the associated symptoms. Case reports for Glu104Asp homozygote patients, as well as the one case report for the Phe240Ser/Glu104Asp compound heterozygote, have described haemolytic anaemia, early neuromuscular impairment (often within the first six months) and increased susceptibility to infection. Patients often die in infancy or early childhood and no case has survived past the age of six (Fermo et al., 2010; Orosz et al., 2006, 2009; Schneider, 2000).

Two variants (Cys41Tyr and Val231Met) were classified as "pathological (intermediate)" as they exhibited many of the stereotypical clinical phenotypes of TPI deficiency including haemolytic anaemia and neuromuscular impairment but patients with these variants survived longer than the typical TPI deficiency patient. In contrast to the more severe TPI deficiency cases, three of these patients were eight years of age and one was aged 15 at the time their cases were described in the literature (Arya et al., 1997; Bardosi et al., 1990; Orosz et al., 2006; Serdaroglu et al., 2011; Wilmshurst et al., 2004).

Finally a group of variants (Ala62Asp, Gly72Ala, Val154Met, Gly122Arg (TPI Manchester)) were classified as "unknown severity" (Manco and Ribeiro, 2007; Perry and Mohrenweiser, 1992; Watanabe et al., 1996). While these have been identified in the human population, no disease state has been associated with them. This is most likely because they have only been found in patients who have one wild-type allele alongside these mutations. In addition to variants for which there has been clinical and/or biochemical studies, a further 79 TPI variants were identified from database searches. To date, none of these have been associated with disease and must be presumed to result from some of the natural sequence diversity within the human *TPI1* gene. These variants were classified as "uncharacterised".

Analysis of the locations of the altered residues within the crystal structure of human TPI only reveals a tentative link with severity

The crystal structure of human TPI (PDB: 4POC) (Roland et al., 2015) was visualised and the disease associated variants were located (Figure 1). The positions of these residues within secondary structure elements (α-helix or β-sheet) along with their relationships to points of interest, primarily the substrate-binding site and dimer interface, were recorded (Figure 1, Table 1, Supplementary Table S1). This revealed that, as noted in previous publications (Li et al., 2013; Rodriguez-Almazan et al., 2008), the Glu104Asp (pathological – severe) variation is very closely related to the dimer interface suggesting a link between severity and the impaired ability of the mutant protein to form a dimer. This residue is also closely related to residue 95, which plays a key role in the formation of the substrate-binding site of the enzyme. The residue of the variant Ile170Val (pathological – less severe) is located on the flexible interconnecting loop which forms the lid of the TIM barrel of the substrate binding site, explaining the very low TPI activity reported in this patient (6%) but is not closely related to the dimer interface potentially explaining why this patient was neurologically intact (Arya et al., 1997; Roland et al., 2015).

Comparison of the average scores from servers predicting ΔΔG values show some relationship between the severity of the effects of the variation and protein stability

Although there is not a perfect correlation between disease severity and protein stability, the results from the six different servers (I-Mutant 3.0 (Structure and Sequence), mCSM, SDM, CUPSAT and BeAtMuSiC) all show that the average ΔΔG values reported for the pathological (less severe) variants are significantly lower than the corresponding values reported for the pathological (severe) variants (Table 1; Figure 2a). The average of five of these servers (BeAtMuSiC was not included as it reports values differently) revealed that the pathological (less severe) variants had an average ΔΔG value of -0.512 kcal/mol and the pathological (severe) variants had an average ΔΔG value of -2.435 kcal/mol (Table 1).

Certain servers predicted much more dramatic stability changes than others - for example Phe240Ser had a reported ΔΔG value of -1.67 kcal/mol in I-mutant 3.0 structure compared to -6.12 kcal/mol in CUPSAT. In order to rectify this, and to allow for greater ease of comparison between servers, the results of the servers were normalised. This was achieved by dividing each value by the highest value reported in each server for a disease associated variant. It is important to note that in order to use this approach, it must be assumed that each server is equally reliable/accurate. Analysis of the normalised results from those servers which report ΔΔG values shows that, on average, the pathological (severe) mutations have a more dramatic effect on protein stability than the pathological (less severe) mutations - 0.202 for the mild mutations compared to 0.631 for the severe mutations.

This contrast can be largely attributed to the much more significant contribution of Phe240Ser than the other pathological (severe) variant, Glu104Asp. The Phe240Ser variant, in the vast majority of cases, is predicted to have the most significant effect on protein stability and/or is reported as being disease-associated with the highest confidence score/reliability index/expected accuracy (Table 1). In those few instances where this is not the case, the respective values are still very high. In contrast to this, the results for Glu104Asp are rather surprising. Although it is, phenotypically, the most severe so far reported, the protein stability results place it in the moderate group. Indeed, a number of servers reported that this variation had among the smallest predicted effect on protein stability.

There was a statistically significant (two way ANOVA with Tukey's correction for multiple comparisons, $p<0.05$) difference between the normalised ΔΔG values for the pathological (less severe) and pathological (severe) groups (Figure 2a). However, there was not a significant difference between the pathological (intermediate) group and either of the other two other pathological groups. Therefore, the pathological (intermediate) and pathological (severe) groups were considered together for the analyses of uncharacterised variants.

The amalgamation of these two groups allowed us to set parameters by which an initial prediction of the likely effect of uncharacterised variants could be made. To do this, it was necessary to define boundaries between three groups of variants: non-pathological, "pathological (less severe)" and pathological (intermediate/severe). To do this, we focussed on the range of ΔΔG values associated with the middle group, pathological (less severe). The upper and lower 95% confidence interval values of the average ΔΔG values for this group -0.13 and -0.90 kcal/mol respectively and these were set as the boundaries. Thus, if the average ΔΔG value of an uncharacterised variant was greater than -0.13 kcal/mol it was predicted to be non-pathological. For a prediction of pathological (less severe) the average ΔΔG value was less than or equal to -0.13 kcal/mol but greater than or equal to -0.90 kcal/mol. If the average ΔΔG value was less than -0.90 kcal/mol a prediction of "pathological intermediate/severe" was made.

Application of these parameters to the 79 uncharacterised variants, resulted in a prediction that eight are non-pathological, 34 are pathological (less severe) and 37 are pathological (intermediate/severe) (Table 2; Figure 2b). These assignments are based on data from a relatively small number of clinically and biochemically characterised variants. Therefore, it is possible that they may change as novel variants are characterised and/or the quality of stability prediction improves. Nevertheless these boundaries serve as an important first step towards the development of a predictive framework for TPI deficiency. Furthermore, the good statistical discrimination between the three categories considered (Figure 2b) suggests that the boundaries are robust.

Interestingly, the application of these parameters to the ΔΔG scores of those variants found in healthy heterozygotes also predicted that all four were pathological. It is likely that these variants would result in very low or zero TPI activity if they were to be present in a homozygote or compound heterozygote, as they were associated with a 50% reduction in TPI activity in the heterozygote carrier (similar to the parents of TPI deficiency cases) (Watanabe et al., 1996). The positions of the corresponding residues of the 79 uncharacterised mutations, within the crystal structure of TPI, were then compared against their predicted disease severity (data not shown). Although this did not reveal any obvious correlation between these two variables, one interesting point to note is that six of the seven residues associated with variation which are predicted to be non-pathological (Ser3, Thr27, Thr89, Gln179, Ser194, Val196) are situated peripherally on the protein and are not part of the substrate binding site or the dimer interface. However a seventh residue (which is associated with two uncharacterised mutations that are predicted to be non-pathological), Asn 71, does form part of the dimer interface.

Aggregation and amyloid formation is not predicted to increase in the disease-associated variants

Recently, it has been demonstrated experimentally that human TPI can form amyloid *in vitro* (Carcamo-Noriega and Saab-Rincon, 2016). Aggregation has also been shown to be a causative factor in other inherited metabolic diseases (e.g. (Bang et al., 2009; McCorvie and Timson, 2013)). However, comparison of the six disease-associated variants with the wild type using a range of aggregation prediction servers provided no evidence that these variants were more likely to aggregate than the wild-type protein. There was no statistically significant difference between the normalised scores for any of the variants and the wild-type (Supplementary Figure S1a). Similarly, when amyloid propensity was predicted using a range of different servers, there was no significant difference between the wild-type and any of the variants (Supplementary Figure S1b).

Residue conservation analysis predicts disease association, but not severity

Disease associated missense mutations tend to alter residues that are evolutionarily conserved (Miller and Kumar, 2001). A combination of a sequence alignment and conservation scoring system allowed the identification of those residues which are highly conserved. All but one of the disease associated variants (Cys41Tyr) affects a residue with a conservation score above 90% and two of the six mutations were 100% conserved across the TPI sequences considered (Supplementary Table S4). There was no significant difference between the three groups (less severe, intermediate and severe) when the conservation scores based on all species where considered or when the analysis was restricted to just mammals (Supplementary Figure S2).

Discussion

Protein stability is an important factor in predicting disease-association and severity of a given mutation but is not the sole factor

In this study, it has been clearly shown that disease-associated variant proteins associated with severe clinical patterns of TPI deficiency are predicted to be significantly more unstable than those associated with comparatively milder progressions. It can, therefore, be said that decreased protein stability is predicted to be associated with increased severity in TPI deficiency. The results from the variants associated with milder forms of the disease, and for Ile170Val in particular, also seem to support the conformational disease hypothesis of TPI deficiency (Orosz et al., 2006), due to the observed link between the absence of neurological disease and the presence of a comparatively stable protein.

However, the imperfect correlation of stability and severity at higher levels, particularly between moderate and severe mutants, the surprisingly low predicted effect of the Glu104Asp variation on protein stability and the apparent lack of a correlation between residue conservation and disease severity all suggest that other factors (e.g. diet, environment, quality of medical care), play a role in the development of this enigmatic disease. Indeed data from a *Drosophila* study has indicated the existence of an isomerase-independent function of TPI which opens up new avenues of investigation that may prove crucial in developing our understanding of this enzyme in the conservation of normal neuronal function and, therefore, in TPI deficiency pathogenesis (Roland et al., 2015; Roland et al., 2013). Other non-metabolic functions of TPI have also been reported in a number of organisms, mostly pathogens. For instance, the potential role of TPI in adhesin function in the fungal pathogen *Paracoccidioides brasiliensis*, has been suggested (Pereira et al., 2007). Such interactions are likely to involve the surface of the protein and, therefore, a significant perturbation of its structure would be potentially affect these processes more than catalysis. Thus, the "moonlighting" roles of TPI may also have a role in disease since they are also likely to affected by altered protein stability and folding. Currently, the molecular-level details of these processes have not been explored in detail. It is, therefore, difficult to incorporate them into the predictive model at present.

Towards a predictive framework - key postulates:

Based on the conservation analysis, aggregation predictions and analysis of the stability predictions, four key postulates were made:

1. Variant proteins associated with pathological (intermediate/severe) phenotypes are significantly more unstable than those associated with a pathological (less severe) phenotype.
2. Although the wild-type protein has been shown to form amyloid, the likelihood of aggregation or amyloid formation is not increased in any of the variants. Therefore, these factors do not predict disease-association.
3. Disease-associated variants tend to occur in highly conserved residues. However their severity does not correlate with the degree of conservation.
4. Pathological (intermediate) mutations should be combined with the pathological (severe) group due to the similarities seen in these two groups.

Conclusions

Overall, our work supports the prevailing view that the pathology of TPI deficiency results largely from alterations to the protein's sequence which alter that stability of the protein with consequent effects of dimerisation and catalysis. From these results we were able to identify a number of variants which are present in the human population which are likely to be associated with pathology. That these variants have not yet been identified as such in patients suggests that these mutations are largely recessive in nature and that disease would only occur in individuals homozygous for these mutations or individuals who are compound heterozygotes for two of these mutations. While our work will not replace the need for laboratory testing in cases of suspected TPI deficiency the approach of predicting protein stability may be useful should new disease-associated variants be discovered. Finally, by reinforcing the concept that protein misfolding lies at the core of this disease we suggest that it may be possible to treat TPI deficiency through the use of small molecules which selectively stabilise TPI ("pharmacological chaperones").

Acknowledgements

This work received no financial support. CO conducted the stability and sequence conservation analyses. DJT conceived the study, conducted the aggregation predictions and supervised CO. CO and DJT co-wrote the paper.

References

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. Nature methods 7(4), 248-249.

Ahmed, N., Battah, S., Karachalias, N., Babaei-Jadidi, R., Horanyi, M., Baroti, K., Hollan, S., Thornalley, P.J., 2003. Increased formation of methylglyoxal and protein glycation, oxidation and nitrosation in triosephosphate isomerase deficiency. Biochim Biophys Acta 1639(2), 121-132.

Arya, R., Lalloz, M.R., Bellingham, A.J., Layton, D.M., 1997. Evidence for founder effect of the Glu104Asp substitution and identification of new mutations in triosephosphate isomerase deficiency. Hum Mutat 10(4), 290-294.

Ationu, A., Humphries, A., 1998. The feasibility of replacement therapy for inherited disorder of glycolysis: triosephosphate isomerase deficiency (review). International journal of molecular medicine 2(6), 701-704.

Ationu, A., Humphries, A., Bellingham, A., Layton, M., 1997. Metabolic correction of triose phosphate isomerase deficiency in vitro by complementation. Biochem Biophys Res Commun 232(2), 528-531.

Ationu, A., Humphries, A., Lalloz, M.R., Arya, R., Wild, B., Warrilow, J., Morgan, J., Bellingham, A.J., Layton, D.M., 1999a. Reversal of metabolic block in glycolysis by enzyme replacement in triosephosphate isomerase-deficient cells. Blood 94(9), 3193-3198.

Ationu, A., Humphries, A., Wild, B., Carr, T., Will, A., Arya, R., Layton, D.M., 1999b. Towards enzyme-replacement treatment in triosephosphate isomerase deficiency. Lancet 353(9159), 1155-1156.

Auer, J., Camoin, L., Courtot, A.M., Hotellier, F., De Almeida, M., 2004. Evidence that P36, a human sperm acrosomal antigen involved in the fertilization process is triosephosphate isomerase. Mol Reprod Dev 68(4), 515-523.

Bang, Y.L., Nguyen, T.T., Trinh, T.T., Kim, Y.J., Song, J., Song, Y.H., 2009. Functional analysis of mutations in UDP-galactose-4-epimerase (GALE) associated with galactosemia in Korean patients using mammalian GALE-null cells. The FEBS journal 276(7), 1952-1961.

Bao, L., Zhou, M., Cui, Y., 2005. nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. Nucleic acids research 33(Web Server issue), W480-482.

Bardosi, A., Eber, S.W., Hendrys, M., Pekrun, A., 1990. Myopathy with altered mitochondria due to a triosephosphate isomerase (TPI) deficiency. Acta neuropathologica 79(4), 387-394.

Bendl, J., Stourac, J., Salanda, O., Pavelka, A., Wieben, E.D., Zendulka, J., Brezovsky, J., Damborsky, J., 2014. PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. PLoS computational biology 10(1), e1003440.

Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D., Zardecki, C., 2002. The Protein Data Bank. Acta Crystallogr D Biol Crystallogr 58(Pt 6 No 1), 899-907.

Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E., 2000. The Protein Data Bank. Nucleic Acids Res 28(1), 235-242.

Bromberg, Y., Rost, B., 2007. SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic acids research 35(11), 3823-3835.

Browne, C., Timson, D.J., 2015. *In Silico* Prediction of the Effects of Mutations in the Human Mevalonate Kinase Gene: Towards a Predictive Framework for Mevalonate Kinase Deficiency. Ann Hum Genet 79(6), 451-459.

Capriotti, E., Altman, R.B., Bromberg, Y., 2013. Collective judgment predicts disease-associated single nucleotide variants. BMC genomics 14 Suppl 3, S2-2164-2114-S2163-S2162. Epub 2013 May 2128.

Capriotti, E., Calabrese, R., Casadio, R., 2006. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. Bioinformatics (Oxford, England) 22(22), 2729-2734.

Capriotti, E., Fariselli, P., Casadio, R., 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Nucleic acids research 33(Web Server issue), W306-310.
Capriotti, E., Fariselli, P., Rossi, I., Casadio, R., 2008. A three-state prediction of single point mutations on protein stability changes. BMC bioinformatics 9 Suppl 2, S6-2105-2109-S2102-S2106.
Carcamo-Noriega, E.N., Saab-Rincon, G., 2016. Identification of fibrillogenic regions in human triosephosphate isomerase. PeerJ 4, e1676.
Carvalho, D.R., Brand, G.D., Brum, J.M., Takata, R.I., Speck-Martins, C.E., Pratesi, R., 2012. Analysis of novel *ARG1* mutations causing hyperargininemia and correlation with arginase I activity in erythrocytes. Gene 509(1), 124-130.
Chang, M.L., Artymiuk, P.J., Wu, X., Hollan, S., Lammi, A., Maquat, L.E., 1993. Human triosephosphate isomerase deficiency resulting from mutation of Phe-240. Am J Hum Genet 52(6), 1260-1269.
Cheng, J., Randall, A., Baldi, P., 2006. Prediction of protein stability changes for single-site mutations using support vector machines. Proteins 62(4), 1125-1132.
Conchillo-Sole, O., de Groot, N.S., Aviles, F.X., Vendrell, J., Daura, X., Ventura, S., 2007. AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinformatics 8, 65.
de Arriba, S.G., Krugel, U., Regenthal, R., Vissiennon, Z., Verdaguer, E., Lewerenz, A., Garcia-Jorda, E., Pallas, M., Camins, A., Munch, G., Nieber, K., Allgaier, C., 2006. Carbonyl stress and NMDA receptor activation contribute to methylglyoxal neurotoxicity. Free Radic Biol Med 40(5), 779-790.
de Arriba, S.G., Stuchbury, G., Yarin, J., Burnell, J., Loske, C., Munch, G., 2007. Methylglyoxal impairs glucose metabolism and leads to energy depletion in neuronal cells - protection by carbonyl scavengers. Neurobiol Aging 28(7), 1044-1050.
De La Mora-De La Mora, I., Torres-Larios, A., Mendoza-Hernandez, G., Enriquez-Flores, S., Castillo-Villanueva, A., Mendez, S.T., Garcia-Torres, I., Torres-Arroyo, A., Gomez-Manzo, S., Marcial-Quino, J., Oria-Hernandez, J., Lopez-Velazquez, G., Reyes-Vivas, H., 2013. The E104D mutation increases the susceptibility of human triosephosphate isomerase to proteolysis. Asymmetric cleavage of the two monomers of the homodimeric enzyme. Biochim Biophys Acta 1834(12), 2702-2711.
Dehouck, Y., Kwasigroch, J.M., Rooman, M., Gilis, D., 2013. BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations. Nucleic acids research 41(Web Server issue), W333-339.
du Plessis, S.S., Agarwal, A., Mohanty, G., van der Linde, M., 2015. Oxidative phosphorylation versus glycolysis: what fuel do spermatozoa use? Asian journal of andrology 17(2), 230-235.
Eber, S.W., Dunnwald, M., Heinemann, G., Hofstatter, T., Weinmann, H.M., Belohradsky, B.H., 1984. Prevalence of partial deficiency of red cell triosephosphate isomerase in Germany--a study of 3000 people. Hum Genet 67(3), 336-339.
Facchiano, A., Marabotti, A., 2010. Analysis of galactosemia-linked mutations of GALT enzyme using a computational biology approach. Protein engineering, design & selection : PEDS 23(2), 103-113.
Fermo, E., Bianchi, P., Vercellati, C., Rees, D.C., Marcello, A.P., Barcellini, W., Zanella, A., 2010. Triose phosphate isomerase deficiency associated with two novel mutations in TPI gene. European journal of haematology 85(2), 170-173.
Fernandez-Escamilla, A.M., Rousseau, F., Schymkowitz, J., Serrano, L., 2004a. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 22(10), 1302-1306.
Fernandez-Escamilla, A.M., Rousseau, F., Schymkowitz, J., Serrano, L., 2004b. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nature biotechnology 22(10), 1302-1306.
Garbuzynskiy, S.O., Lobanov, M.Y., Galzitskaya, O.V., 2010. FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. Bioinformatics 26(3), 326-332.
Hollan, S., Fujii, H., Hirono, A., Hirono, K., Karro, H., Miwa, S., Harsanyi, V., Gyodi, E., Inselt-Kovacs, M., 1993. Hereditary triosephosphate isomerase (TPI) deficiency: two severely affected brothers one with and one without neurological symptoms. Hum Genet 92(5), 486-490.

Krieger, E., Joo, K., Lee, J., Lee, J., Raman, S., Thompson, J., Tyka, M., Baker, D., Karplus, K., 2009. Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: Four approaches that performed well in CASP8. Proteins 77 Suppl 9, 114-122.
Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nature protocols 4(7), 1073-1081.
Li, Y., Cohenford, M.A., Dutta, U., Dain, J.A., 2008. In vitro nonenzymatic glycation of guanosine 5'-triphosphate by dihydroxyacetone phosphate. Analytical and bioanalytical chemistry 392(6), 1189-1196.
Li, Z., He, Y., Liu, Q., Zhao, L., Wong, L., Kwoh, C.K., Nguyen, H., Li, J., 2013. Structural analysis on mutation residues and interfacial water molecules for human TIM disease understanding. BMC Bioinformatics 14 Suppl 16, S11.
Linding, R., Schymkowitz, J., Rousseau, F., Diella, F., Serrano, L., 2004. A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. J Mol Biol 342(1), 345-353.
Mainfroid, V., Terpstra, P., Beauregard, M., Frere, J.M., Mande, S.C., Hol, W.G., Martial, J.A., Goraj, K., 1996. Three hTIM mutants that provide new insights on why TIM is a dimer. J Mol Biol 257(2), 441-456.
Manco, L., Ribeiro, M.L., 2007. Novel human pathological mutations. Gene symbol: TPI1. Disease: triosephosphate isomerase deficiency. Hum Genet 121(5), 650.
Manning, J.R., Bailey, M.A., Soares, D.C., Dunbar, D.R., Mullins, J.J., 2010. *In silico* structure-function analysis of pathological variation in the HSD11B2 gene sequence. Physiological genomics 42(3), 319-330.
McCorvie, T.J., Timson, D.J., 2013. *In silico* prediction of the effects of mutations in the human UDP-galactose 4'-epimerase gene: Towards a predictive framework for type III galactosemia. Gene 524(2), 95-104.
Merkle, S., Pretsch, W., 1989. Characterization of triosephosphate isomerase mutants with reduced enzyme activity in *Mus musculus*. Genetics 123(4), 837-844.
Miller, M.P., Kumar, S., 2001. Understanding human disease mutations through the use of interspecific genetic variation. Human molecular genetics 10(21), 2319-2328.
Olah, J., Orosz, F., Keseru, G.M., Kovari, Z., Kovacs, J., Hollan, S., Ovadi, J., 2002. Triosephosphate isomerase deficiency: a neurodegenerative misfolding disease. Biochemical Society transactions 30(2), 30-38.
Olah, J., Orosz, F., Puskas, L.G., Hackler, L., Jr., Horanyi, M., Polgar, L., Hollan, S., Ovadi, J., 2005. Triosephosphate isomerase deficiency: consequences of an inherited mutation at mRNA, protein and metabolic levels. The Biochemical journal 392(Pt 3), 675-683.
Orosz, F., Olah, J., Ovadi, J., 2006. Triosephosphate isomerase deficiency: facts and doubts. IUBMB life 58(12), 703-715.
Orosz, F., Olah, J., Ovadi, J., 2009. Triosephosphate isomerase deficiency: new insights into an enigmatic disease. Biochimica et biophysica acta 1792(12), 1168-1174.
Orosz, F., Vertessy, B.G., Hollan, S., Horanyi, M., Ovadi, J., 1996. Triosephosphate isomerase deficiency: predictions and facts. Journal of theoretical biology 182(3), 437-447.
Parthiban, V., Gromiha, M.M., Schomburg, D., 2006. CUPSAT: prediction of protein stability upon point mutations. Nucleic acids research 34(Web Server issue), W239-242.
Pekrun, A., Neubauer, B.A., Eber, S.W., Lakomek, M., Seidel, H., Schroter, W., 1995. Triosephosphate isomerase deficiency: biochemical and molecular genetic analysis for prenatal diagnosis. Clin Genet 47(4), 175-179.
Pereira, L.A., Bao, S.N., Barbosa, M.S., da Silva, J.L., Felipe, M.S., de Santana, J.M., Mendes-Giannini, M.J., de Almeida Soares, C.M., 2007. Analysis of the *Paracoccidioides brasiliensis* triosephosphate isomerase suggests the potential for adhesin function. FEMS yeast research 7(8), 1381-1388.
Perry, B.A., Mohrenweiser, H.W., 1992. Human triosephosphate isomerase: substitution of Arg for Gly at position 122 in a thermolabile electromorph variant, TPI-Manchester. Hum Genet 88(6), 634-638.

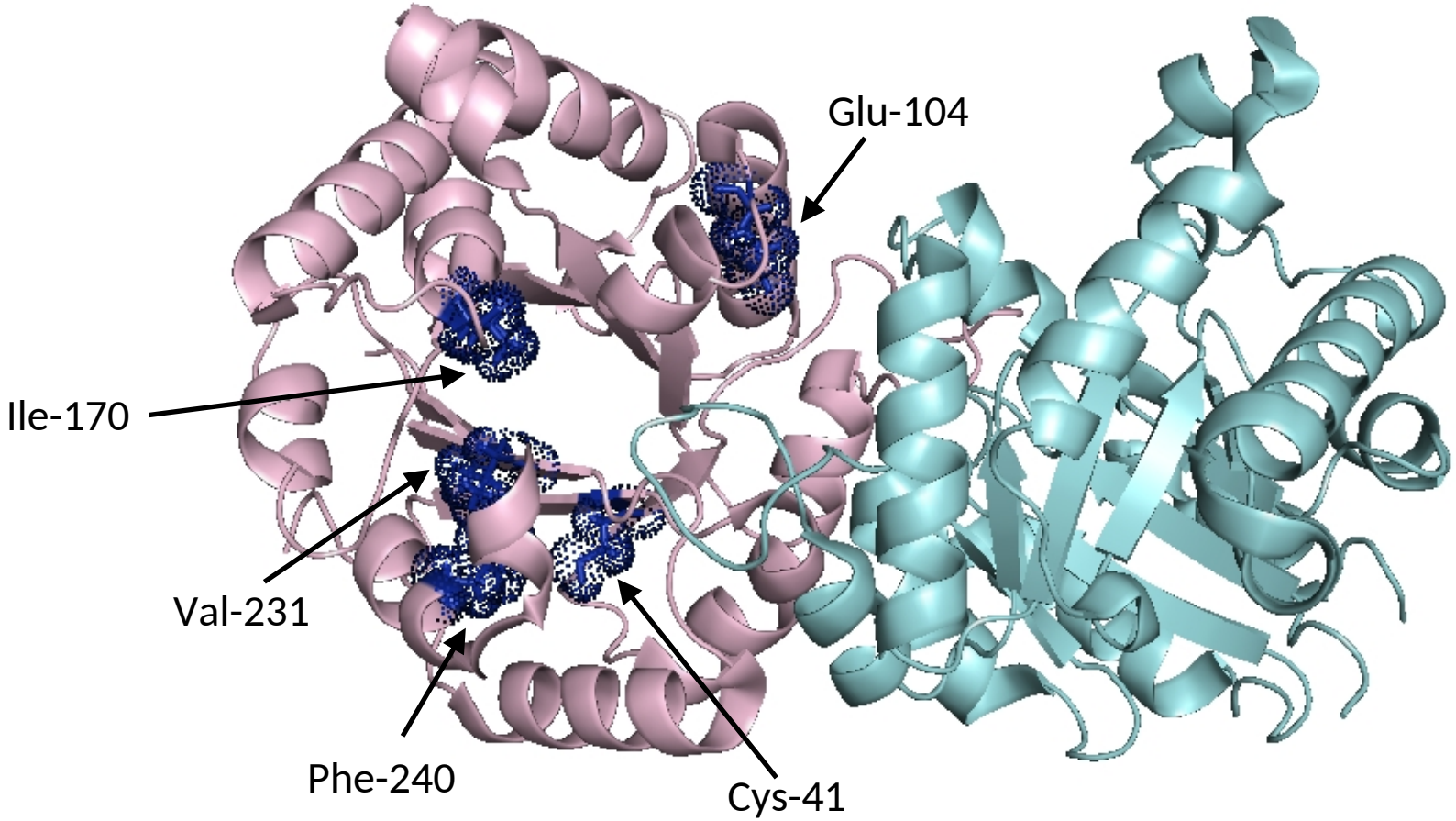
Petit, F.M., Serres, C., Auer, J., 2014. Moonlighting proteins in sperm-egg interactions. Biochem Soc Trans 42(6), 1740-1743.
Pires, D.E., Ascher, D.B., Blundell, T.L., 2014. mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics (Oxford, England) 30(3), 335-342.
Poinsot, J., Parent, P., Alix, D., Toudic, L., Castel, Y., 1986. A case of congenital non-spherocytic hemolytic anemia caused by triose phosphate isomerase deficiency. Prenatal diagnosis. Journal de genetique humaine 34(5), 431-437.
Ralser, M., Heeren, G., Breitenbach, M., Lehrach, H., Krobitsch, S., 2006. Triose phosphate isomerase deficiency is caused by altered dimerization--not catalytic inactivity--of the mutant enzymes. PloS one 1, e30.
Rodriguez-Almazan, C., Arreola, R., Rodriguez-Larrea, D., Aguirre-Lopez, B., de Gomez-Puyou, M.T., Perez-Montfort, R., Costas, M., Gomez-Puyou, A., Torres-Larios, A., 2008. Structural basis of human triosephosphate isomerase deficiency: mutation E104D is related to alterations of a conserved water network at the dimer interface. The Journal of biological chemistry 283(34), 23254-23263.
Roland, B.P., Amrich, C.G., Kammerer, C.J., Stuchul, K.A., Larsen, S.B., Rode, S., Aslam, A.A., Heroux, A., Wetzel, R., VanDemark, A.P., Palladino, M.J., 2015. Triosephosphate isomerase I170V alters catalytic site, enhances stability and induces pathology in a Drosophila model of TPI deficiency. Biochim Biophys Acta 1852(1), 61-69.
Roland, B.P., Stuchul, K.A., Larsen, S.B., Amrich, C.G., Vandemark, A.P., Celotto, A.M., Palladino, M.J., 2013. Evidence of a triosephosphate isomerase non-catalytic function crucial to behavior and longevity. J Cell Sci 126(Pt 14), 3151-3158.
Roland, B.P., Zeccola, A.M., Larsen, S.B., Amrich, C.G., Talsma, A.D., Stuchul, K.A., Heroux, A., Levitan, E.S., VanDemark, A.P., Palladino, M.J., 2016. Structural and Genetic Studies Demonstrate Neurologic Dysfunction in Triosephosphate Isomerase Deficiency Is Associated with Impaired Synaptic Vesicle Dynamics. PLoS Genet 12(3), e1005941.
Rosa, R., Prehu, M.O., Calvin, M.C., Daffos, F., Forestier, F., 1986. Possibility of prenatal diagnosis of hereditary triose phosphate isomerase deficiency. Prenat Diagn 6(3), 231-234.
Rousseau, F., Schymkowitz, J., Serrano, L., 2006. Protein aggregation and amyloidosis: confusion of the kinds? Curr Opin Struct Biol 16(1), 118-126.
Sarper, N., Zengin, E., Jakobs, C., Salomons, G.S., Wamelink, M.M., Ralser, M., Kurt, K., Kara, B., 2013. Mild hemolytic anemia, progressive neuromotor retardation and fatal outcome: a disorder of glycolysis, triose-phosphate isomerase deficiency. Turkish Journal of Pediatrics 55, 198-202.
Schneider, A., Valentine, W., Hattori, M., Heins, H., 1964. A new erythrocyte enzyme defect with hemolytic anemia-triosephosphate isomerase (TPI) deficiency. Blood 24(6), 855-856.
Schneider, A.S., 2000. Triosephosphate isomerase deficiency: historical perspectives and molecular aspects. Bailliere's best practice & research. Clinical haematology 13(1), 119-140.
Schneider, A.S., Valentine, W.N., Hattori, M., Heins, H.L., Jr., 1965. Hereditary hemolytic anemia with triosephosphate isomerase deficiency. N Engl J Med 272, 229-235.
Serdaroglu, G., Aydinok, Y., Yilmaz, S., Manco, L., Ozer, E., 2011. Triosephosphate isomerase deficiency: a patient with Val231Met mutation. Pediatr Neurol 44(2), 139-142.
Shi, Y., Vaden, D.L., Ju, S., Ding, D., Geiger, J.H., Greenberg, M.L., 2005. Genetic perturbation of glycolysis results in inhibition of de novo inositol biosynthesis. J Biol Chem 280(51), 41805-41810.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J.D., Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular systems biology 7, 539.
Sormanni, P., Aprile, F.A., Vendruscolo, M., 2015. The CamSol method of rational design of protein mutants with enhanced solubility. J Mol Biol 427(2), 478-490.
Stone, E.A., Sidow, A., 2005. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. Genome Res 15(7), 978-986.
Stroppolo, M.E., Falconi, M., Caccuri, A.M., Desideri, A., 2001. Superefficient enzymes. Cell Mol Life Sci 58(10), 1451-1460.

Tartaglia, G.G., Pawar, A.P., Campioni, S., Dobson, C.M., Chiti, F., Vendruscolo, M., 2008. Prediction of aggregation-prone regions in structured proteins. J Mol Biol 380(2), 425-436.
Thomas, P.D., Kejariwal, A., Campbell, M.J., Mi, H., Diemer, K., Guo, N., Ladunga, I., Ulitsky-Lazareva, B., Muruganujan, A., Rabkin, S., Vandergriff, J.A., Doremieux, O., 2003. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic Acids Res 31(1), 334-341.
Timson, D.J., 2015. Value of predictive bioinformatics in inherited metabolic diseases. World Journal of Medical Genetics 5, 46-51.
Valdar, W.S., 2002. Scoring residue conservation. Proteins 48(2), 227-241.
Valentin, C., Cohen-Solal, M., Maquat, L., Horanyi, M., Inselt-Kovacs, M., Hollan, S., 2000. Identical germ-line mutations in the triosephosphate isomerase alleles of two brothers are associated with distinct clinical phenotypes. C R Acad Sci III 323(3), 245-250.
Watanabe, M., Zingg, B.C., Mohrenweiser, H.W., 1996. Molecular analysis of a series of alleles in humans with reduced activity at the triosephosphate isomerase locus. Am J Hum Genet 58(2), 308-316.
Wilmshurst, J.M., Wise, G.A., Pollard, J.D., Ouvrier, R.A., 2004. Chronic axonal neuropathy with triosephosphate isomerase deficiency. Pediatr Neurol 30(2), 146-148.
Worth, C.L., Preissner, R., Blundell, T.L., 2011. SDM - a server for predicting effects of mutations on protein stability and malfunction. Nucleic acids research 39(Web Server issue), W215-222.

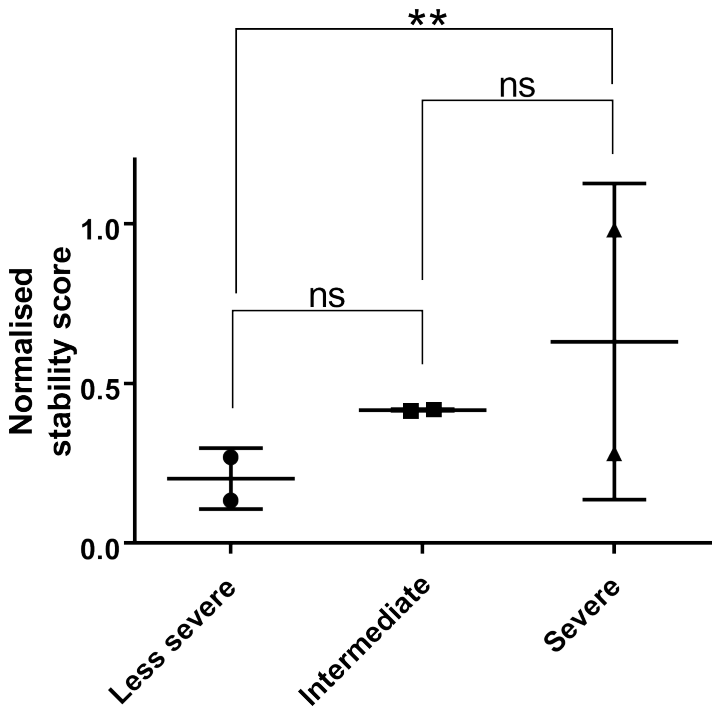
Figure legends

Figure 1: The three dimensional structure of dimeric human wild-type TPI (PDB: 4POC) (Roland et al., 2015), visualised using PyMol, with residues associated with known, disease-associated variants highlighted in blue and shown in stick format overlaid with space-filling spheres. The protein consists of two identical monomers, shown in light purple and cyan. Disease-associated residues are only highlighted on the left hand subunit. Note that Val-231 and Phe-240 are in contact with each other.

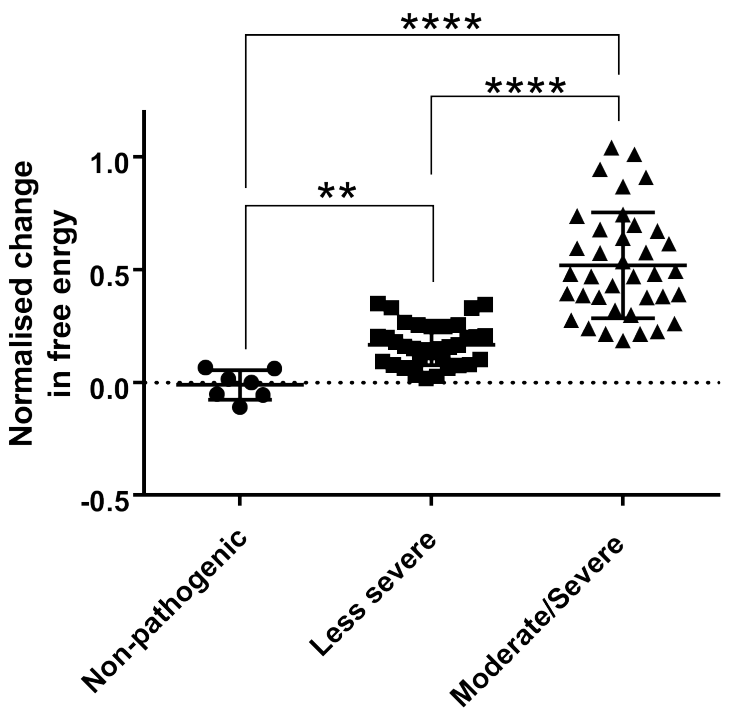
Figure 2: Summary of stability predictions by group. (a) The well-characterised, disease-associated variants were grouped by clinical features. The lack of significant difference between the pathological (intermediate) and pathological (severe) group resulted in the combination of these two groups it subsequent analyses. (b) The uncharacterised variants were grouped by their predicted severity. ns, not significant; ** $p<0.01$; **** $p<0.0001$ (two way ANOVA with Tukey's correction for multiple comparisons). Each point represents a normalised mean value (derived from six different prediction methods) for a single variant. The wide bar represents the mean of all values in the group and the narrower bars one standard deviation of this mean.


Glu-104
Ile-170
Val-231
Phe-240
Cys-41

(a)


**
ns
ns
Normalised stability score
1.0
0.5
0.0
Less severe
Intermediate
Severe

(b)


****
****
**
Normalised change in free enrgy
1.0
0.5
0.0
-0.5
Non-pathogenic
Less severe
Moderate/Severe

Tables

Table 1: Results of conservation analysis, structural analysis, and the average of those results from ΔΔG prediction servers for characterised variants and those variants reported in healthy heterozygotes.

Variant	Average ΔΔG Value (kcal mol^{-1})	Degree of Conservation	Position of the altered residue	Associated/predicted severity
Cys41Tyr	-2.60	43.8%	Towards C-terminal end of 2nd β-sheet	Pathological (moderate/severe)
Ala62Asp	-2.08	56.4%	Middle of 3rd β-sheet	PREDICTED: Pathological (moderate/severe)
Gly72Ala	-0.38	83.4%	Middle of interconnecting loop between 3rd β-sheet and 3rd α-helix	PREDICTED: Pathological (less severe)
Glu104Asp	-1.41	100%	Middle of interconnecting loop between 4th and 5th α-helix	Pathological (moderate/severe)
Gly122Arg	-0.13	50.1%	N-terminal end of 5th β-sheet	PREDICTED: Pathological (less severe)
Val154Met	-1.31	70.1%	N-terminal end of interconnecting loop between 7th α-helix and 6th β-sheet	PREDICTED: Pathological (moderate/severe)
Ile170Val	-0.43	100%	Middle of interconnecting loop between 6th β-sheet and 8th α-helix	Pathological (less severe)
Val231Met	-1.95	92.5%	C- terminal end of 8th β-sheet. Part of substrate binding site	Pathological (moderate/severe)
Phe240Leu	-0.59	93.9%	Towards N-terminal end of 12th α-helix	Pathological (less severe)
Phe240Ser	-3.46	93.9%	Towards N-terminal end of 12th α-helix	Pathological (moderate/severe)

Table 2: Results of conservation analysis, structural analysis, and the average of those results from ΔΔG prediction servers for each of the 79 "uncharacterised" variants from the dbSNP database, along with their predicted severity.

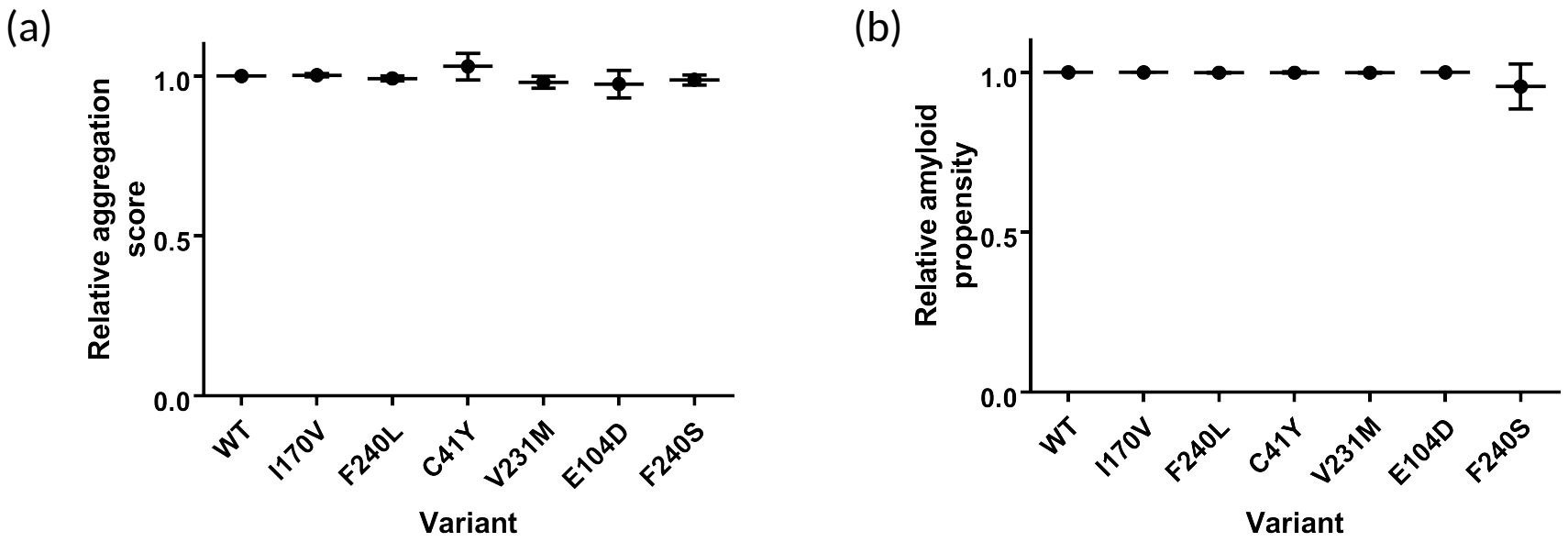
Variant	Average ΔΔG Value (kcal mol^{-1})	Degree of Conservation	Position of the altered residue	Prediction of severity
Ser3Ala	0.18	34.0%	N-terminal of first loop before 1st β-sheet	Non-Pathological
Arg4Gly	-1.93	64.9%	Middle of initial loop before 1st β-sheet	Pathological (moderate/severe)
Phe6Ile	-0.93	51.6%	N-terminal end of 1st β-sheet	Pathological (moderate/severe)
Phe6Leu	-1.05	51.6%	N-terminal end of 1st β-sheet	Pathological (moderate/severe)
Val8Ala	-2.04	71.2%	Middle of 1st β -sheet. Near to substrate binding site	Pathological (moderate/severe)
Asn11Asp	-0.15	77.8%	C-terminal end of 1st β-sheet. Key residues of the substrate binding site.	Pathological (less severe)
Asn11Lys	-0.43	77.8%	C-terminal end of 1st β-sheet. Key residue of substrate binding site.	Pathological (less severe)
Trp12Gly	-1.60	63.1%	N- terminal end of interconnecting loop between 1st β-sheet and 1st α-helix. Near substrate binding site	Pathological (moderate/severe)
Met14Ile	-0.92	53.9%	Middle of interconnecting loop between 1st β-sheet and 1st α-helix. Near substrate binding site and part of dimer interface	Pathological (moderate/severe)
Gly16Val	-0.31	62.5%	Towards C-terminal end of interconnecting loop between 1st β-sheet and 1st α-helix. Part of dimer interface	Pathological (less severe)
Ser20Asn	-0.20	35.5%	Towards N-terminal of 1st α-helix. Near dimer interface	Pathological (less severe)
Leu24Val	-1.14	52.6%	Middle of 1st α-helix.	Pathological (moderate/severe)
Thr27Ile	0.88	40.5%	Towards C-terminal end of 1st α-helix	Non-Pathological
Val33Leu	-0.92	48.9%	Middle of interconnecting loop between 1st α-helix and 2nd β-sheet	Pathological (moderate/severe)
Thr37Ala	-0.39	47.9%	N-terminal end of 2nd β-sheet.	Pathological (less severe)

Glu38Asp	-0.88	59.0%	Towards N-terminal end of 2nd β-sheet	Pathological (less severe)
Glu38Lys	-0.58	59.0%	Towards N-terminal end of 2nd β-sheet	Pathological (less severe)
Val39Leu	-0.22	65.7%	Middle of 2nd β-sheet	Pathological (less severe)
Ala46Val	-0.46	42.8%	Towards c-terminal end of interconnecting loop between 2nd β-sheet and 2nd α--helix. Near dimer interface	Pathological (less severe)
Ala51Thr	-1.91	52.5%	Middle of 2nd α-helix. Near dimer interface	Pathological (moderate/severe)
Arg52Gln	-0.51	52.8%	Middle of 2nd α-helix. Near dimer interface	Pathological (less severe)
Lys58Arg	-0.55	41.0%	Near C-terminal end of interconnecting loop between 2nd α-helix and 3rd β-sheet	Pathological (less severe)
Ile59Met	-0.68	53.5%	C-terminal end of interconnecting loop between 2nd α-helix and 3rd β-sheet	Pathological (less severe)
Ala63Ser	-1.60	71.6%	C-terminal end of 3rd β-sheet. Near dimer interface	Pathological (moderate/severe)
Val69Met	-0.88	37.1%	Part of interconnecting loop between 3rd β-sheet and 3rd α-helix. Part of dimer interface	Pathological (less severe)
Asn71Thr	0.17	42.0%	Part of interconnecting loop between 3rd β-sheet and 3rd α-helix. Part of dimer interface	Non-Pathological
Asn71Lys	-0.08	42.0%	Part of interconnecting loop between 3rd β-sheet and 3rd α-helix. Part of dimer interface	Non-Pathological
Ala88Pro	-0.27	57.8%	Middle of interconnecting loop between 3rd α-helix and 4th β-sheet. Near dimer interface	Pathological (less severe)
Thr89Met	0.13	46.7%	C-terminal end of interconnecting loop between 3rd α-helix and 4th β-sheet	Non-Pathological
Val91Ala	-2.54	83.5%	Middle of 4th β-sheet. Near substrate binding site	Pathological (moderate/severe)
Arg99Ser	-1.71	95.4%	Middle of 4th α-helix. Near substrate binding site and dimer interface	Pathological (moderate/severe)

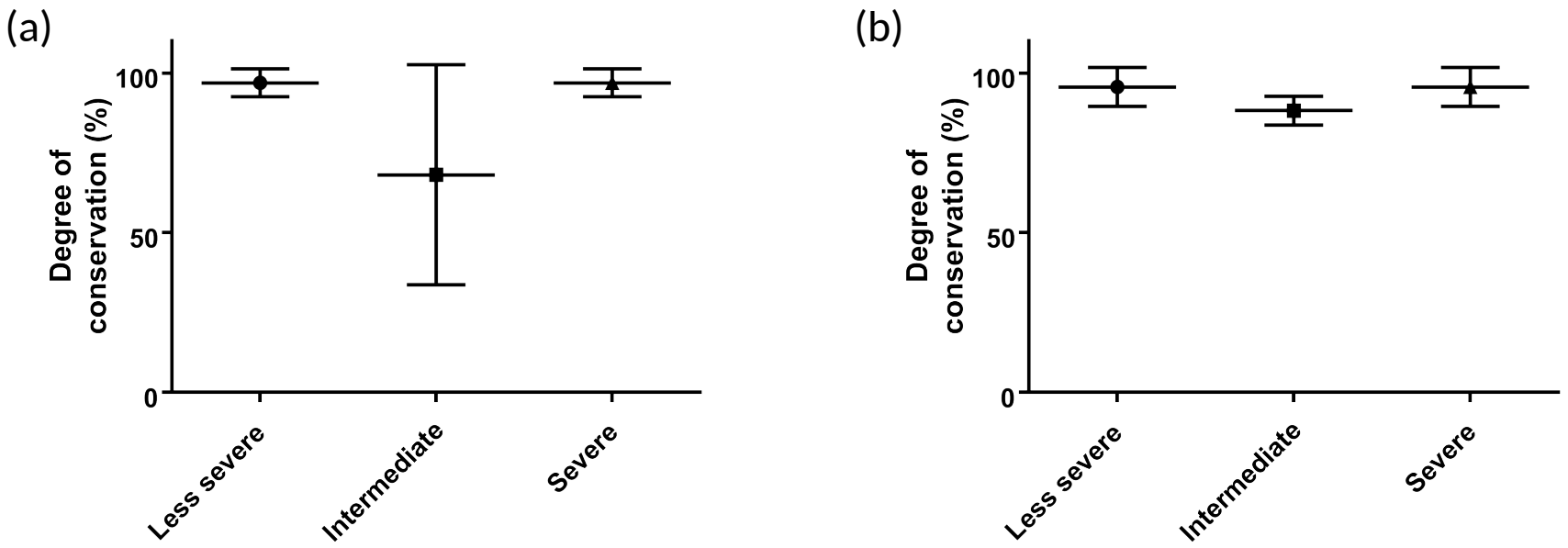
Ala118Thr	-0.76	49.1%	C-terminal end of 5th α-helix	Pathological (less severe)
Val123Ala	-2.74	78.0%	N-terminal end of 5th β-sheet. Near substrate binding site	Pathological (moderate/severe)
Val123Ile	-0.75	78.0%	N-terminal end of 5th β-sheet. Near substrate binding site	Pathological (less severe)
Ala125Thr	-1.71	64.1%	Middle of 5th β-sheet. Near substrate binding site	Pathological (moderate/severe)
Ile127Asn	-3.30	80.1%	C-terminal end of 5th β-sheet. Near substrate binding site	Pathological (moderate/severe)
Val143Leu	-0.37	76.4%	Part of the 7th α-helix.	Pathological (less severe)
Phe144Ser	-2.38	46.1%	Part of the 7th α-helix.	Pathological (moderate/severe)
Gln146His	-0.41	94.9%	Part of the 7th α-helix.	Pathological (less severe)
Thr147Arg	-1.18	61.0%	Part of the 7th α-helix.	Pathological (moderate/severe)
Val149Phe	-0.66	58.8%	Near c-terminal end of 7th α-helix.	Pathological (less severe)
Ile150Met	-1.45	67.0%	Near c-terminal end of 7th α-helix.	Pathological (moderate/severe)
Asn153Lys	-0.61	50.5%	C-terminal end of 7th α-helix.	Pathological (less severe)
Val154Ala	-2.12	70.1%	N-terminal end of interconnecting loop between the 7th α-helix and the 6th β-sheet	Pathological (moderate/severe)
Trp157Gly	-2.57	68.6%	Middle of interconnecting loop between the 7th α-helix and the 6th β-sheet	Pathological (moderate/severe)
Lys159Glu	-0.20	61.2%	N-terminal end of interconnecting loop between the 7th α-helix and the 6th β-sheet	Pathological (less severe)
Val161Ile	-0.40	91.3%	Near N-terminal end of 6th β-sheet. Near substrate binding site	Pathological (less severe)
Ala163Ser	-2.29	100.0%	Part of 6th β-sheet. Near substrate binding site	Pathological (moderate/severe)
Ala176Glu	-0.95	98.4%	Towards C-terminal end of interconnecting loop between 6th β-sheet and 8th α-helix. Near substrate binding site	Pathological (moderate/severe)

Pro178Thr	-0.26	81.6%	N-terminal end of 8th α-helix	Pathological (less severe)
Gln179Lys	0.10	59.1%	Towards N-terminal end of 8th α-helix	Non-Pathological
Ala181Gly	-2.39	90.5%	Towards N-terminal end of 8th α-helix.	Pathological (moderate/severe)
Gln182Arg	-0.49	87.2%	Part of 8th α-helix.	Pathological (less severe)
His185Gln	-1.09	88.1%	Part of 8th α-helix	Pathological (moderate/severe)
Glu186Lys	-0.53	52.7%	Part of 8th α-helix	Pathological (less severe)
Glu186Gly	-1.43	52.7%	Part of 8th α-helix	Pathological (moderate/severe)
Trp191Cys	-1.29	63.2%	Part of 8th α-helix	Pathological (moderate/severe)
Leu192Gln	-2.92	75.3%	Part of 8th α-helix	Pathological (moderate/severe)
Ser194Phe	-0.06	46.3%	Towards C-terminal end of 8th α-helix	Non-Pathological
Val196Ile	-0.10	62.8%	C-terminal end of 8th α-helix	Non-Pathological
Ala199Val	-0.60	49.5%	Towards N-terminal end of 9th α-helix	Pathological (less severe)
Val200Gly	-2.65	76.2%	Part of 9th α-helix	Pathological (moderate/severe)
Thr204Ile	-0.28	60.5%	Interconnecting loop between 9th α-helix and 7th β-sheet	Pathological (less severe)
Thr204Ser	-1.84	60.5%	Interconnecting loop between 9th α-helix and 7th β-sheet	Pathological (moderate/severe)
Arg205His	-0.64	85.9%	Interconnecting loop between 9th α-helix and 7th β-sheet	Pathological (less severe)
Arg205Cys	-2.72	85.9%	Interconnecting loop between 9th α-helix and 7th β-sheet	Pathological (moderate/severe)
Tyr208Cys	-0.31	100.0%	C-terminal end of 7th β-sheet. Near substrate binding site	Pathological (less severe)
Val212Met	-0.95	97.1%	Middle of interconnecting loop between 7th β-sheet and 10th α-helix. Near substrate binding site	Pathological (moderate/severe)
Glu219Asp	-0.64	77.7%	Middle of 10th α-helix.	Pathological (less severe)
Leu220Val	-1.30	91.0%	Middle of 10th α-helix	Pathological

				(moderate/severe)
Ser222Gly	-1.00	49.5%	C-terminal end of 10th α-helix	Pathological (moderate/severe)
Asp225Gly	-1.54	84.5%	Interconnecting loop between 10th α-helix and 8th β-sheet	Pathological (moderate/severe)
Val231Ala	-2.77	92.5%	C- terminal end of 8th β-sheet. Part of substrate binding site	Pathological (moderate/severe)
Lys237Arg	-0.37	82.2%	Interconnecting loop between 11th and 12th α- helices. Near dimer interface	Pathological (less severe)
Pro238Ser	-0.65	63.9%	N-terminal end of 12th α-helix	Pathological (less severe)
Glu239Lys	-1.11	67.3%	Towards N-terminal end of 12th α-helix	Pathological (moderate/severe)
Asp242Tyr	-0.16	61.7%	Middle of 12th α-helix	Pathological (less severe)
Ile243Val	-1.11	94.1%	Middle of 12th α-helix.	Pathological (moderate/severe)
Asn245Ser	-0.53	68.4%	C-terminal end of 12th α-helix.	Pathological (less severe)



Supplementary figure S1: Comparison of the normalised aggregation scores (a) and amyloid propensity scores (b) for wild-type human triose phosphate isomerase and six disease-associated variants. No statistically significant difference was detected between the score for the wild-type and any of the variants (one way ANOVA comparing each variant with the wild-type). Each point and long bar represents the mean score (normalised against the score for WT) from the prediction servers. The shorter bars represent one standard deviation of these means.



Supplementary figure S2: Comparison of the degree of conservation of the three groups of disease-associated variants. The comparison was based on (a) all species considered and (b) mammals. No significant difference between the groups was detected (one way ANOVA with Tukey's correction for multiple comparisons).

Supplementary Table S1: Summary of the primary data from the predictions made on	Column1	Column2	Column3	Column4	Column5
Mutation	Phenotype	I-Mutant 3.0 Structure (ddg in Kcal/mol)	I-Mutant 3.0 Sequence	mCSM (Kcal/mol)	SDM
Ile170Val	(Mild). Compound heterozygote with E104D HA. No NMI. Alive at age 8	RSA: 33.0 / 0.06 (Neutral)/ RI:1	Decrease -0.56 RI: 5	-0.805	-0.01. Neutral. Non-disease associated
Phe240Leu	(Mild-Moderate) 2 brothers - 1 neurologically intact, other brother had some NMI. Compound heterozygotes with E145X Both had HA and severe TPID (3%). Still alive in adulthood	RSA: 0.0 / -0.85 Large Decrease / RI:4	Decrease -1.01 RI: 6	-1.285	-0.17. Neutral. Non-disease associated
Cys41Tyr	(Mild - Moderate). 3 cases - compound heterozygotes with E104D. HA, NMI. Still alive at age 8	RSA: 0.0 / -0.99 Large Decrease / RI: 5	Neutral/Small increase 0.05 RI:1	-1.074	-4.08 Highly destabilising. Protein malfunction and disease
Val231Met	(Moderate) 2 cases. Homozygotes. Similar to mitochondrial myopathy. Patients alive at 8 and 15 years old. Haemolytic anaemia and neuromuscular impairment present	RSA: 1.3 / -0.62 Large Decrease / RI: 3	Decrease -1.28 RI: 8	-1.019	-0.88. Slightly destabilising. Non-disease associated
Phe240Ser	(Moderate - Severe) 1 case - compound heterozygote E104D. HA, NMI, increased susc. Infection. Death age 6 years	RSA: 0.0 / -1.67 Large Decrease / RI:7	Decrease -1.64 RI: 8	-3.142	-4.75. Highly destabilising. Protein malfunction and disease
Glu104Asp	(Severe) Most common. HA, NMI, Cardiomyopathy and increased susc. to infection. Death in early childhood	RSA: 9.2 / -0.82 Large Decrease / RI:8	Neutral/small decrease -0.40 RI:3	-1.328	-1.40. Destabilising. Non-disease associated
		RSA = Relative solvent accesible area, RI = reliabiliity index			
Uncharacterised variants/ those found in healthy heterozygotes					
Ala62Asp	Uncharacterised	RSA: 0.0 / -1.40 - Large Decrease / RI:8	Decrease -0.72 RI: 7	-2.697	-2.91. Highly Destabilising. Protein malfunction and disease.
Gly72Ala	Found in healthy heterozygote	RSA: 69.9 / -0.33 Neutral or Small Decrease / RI: 2	Decrease -0.77 RI: 8	-0.77	1.26. Stabilising. Non-disease associated
Gly122Arg	TPI Manchester. Healthy heterozygote	RSA:1.3 / -0.93 Large Decrease / RI:5	Neutral/small decrease -0.41 RI: 4	-0.895	0.80. Slightly stabilising. Non-disease associated
Val154Met	Found in healthy heterozygote	RSA: 9.8 / -0.84 Large Decrease/ RI:5	Decrease -0.62 RI: 5	-0.301	-0.56. Slightly destabilising. Non-disease associated

Mutation	Column6	Column7	Column8	Column9	Column10	Column11	Column12
	PANTHER (0-1)(>0.5 = Disease)	PhD-SNP (0-1)(>0.5=Disease)	SIFT (>0.05 = Neutral)	SNAP (0-1)(>0.5 = Disease)	Meta-SNP (0-1)(>0.5 = Disease)	MAPP (% expected accuracy)	Predict SNP 1.0 (% expected accuracy)
Ile170Val	Neutral 0.405	Neutral 0.491	0.010 Disease	Disease 0.740	Disease 0.608 RI = 2	Deleterious - 63%	Deleterious - 87%
Phe240Leu	Disease 0.642	Disease 0.786	0.000 Disease	Disease 0.710	Disease 0.737 RI = 5	Deleterious - 76%	Deleterious - 87%
Cys41Tyr	Disease 0.696	Disease 0.763	0.000 Disease	Disease 0.705	Disease 0.773 RI = 5	Deleterious - 56%	Deleterious - 87%
Val231Met	Disease 0.790	Disease 0.623	0.000 Disease	Disease 0.805	Disease 0.797 RI =6	Deleterious - 78%	Deleterous - 87%
Phe240Ser	Disease 0.776	Disease 0.893	0.000 Disease	Disease 0.715	Disease 0.782 RI = 6	Deleterious - 81%	Deleterious - 87%
Glu104Asp	Disease 0.841	Disease 0.704	0.030 Disease	Disease 0.710	Disease 0.636 RI = 3	Deleterious - 62%	Deleterious - 61%
Ala62Asp	Disease 0.870	Disease 0.864	0.000 Disease	Disease 0.740	Disease 0.830 RI=7	Deleterious -88%	Deleterious - 87%
Gly72Ala	Disease 0.622	Disease 0.777	0.010 Disease	Disease 0.765	Disease 0.737 RI = 5	Deleterious - 77%	Deleterious - 87%
Gly122Arg	Neutral 0.215	Disease 0.621	0.320 Neutral	Neutral 0.410	Neutral 0.229 RI = 5	Neutral - 65%	Neutral - 63%
Val154Met	Disease 0.584	Neutral 0.383	0.020 Disease	Disease 0.585	Disease 0.640 RI = 3	Neutral - 64%	Deleterious - 55%

Mutation	Column13	Column14	Column15	Column16	Column17	Column18	Column19
	MuPro - SVM - sequence only	MuPro Neural Network -sequence only	CUPSAT	BeAtMuSiC (DDG-bind in Kcal/mol)	nsSNPAnalyser	PolyPhen-2 HumDiv	PolyPhen-2 HumVar
Ile170Val	Decrease stability - CS: -1.00	Decrease stability - CS: -0.796	-0.84Kcal/mol. Destabilising. Torsion Favourable	0.18 - Decreases binding affinity. SAiC = 21.55% (Part of interface)	Neutral	Probably damaging 0.985	Possibly damaging 0.821
Phe240Leu	Decrease stability - CS: -0.844	Decrease stability - CS: -0.701	0.35Kcal/mol. Stabilising. Torson Favourable.	0.14 - Decreases binding affinity. SAiC = 0.00%	Disease	Probably damaging 0.998	Possibly damaging 0.901
Cys41Tyr	Decrease stability - CS: -0.618	Increase stability - CS: 0.522	-6.9Kcal/mol. Destabilising. Torsion Favourable	-0.39 - Increases binding affinity. SAiC = 0.00%	Disease	Benign 0.197	Possibly damaging 0.514
Val231Met	Decrease stability - CS: -0.402	Decrease stability - CS: -0.739	-5.94Kcal/mol. Destabilising. Torsion - No Change	-0.06 - Increases binding affinity. SAiC = 0.3%	Neutral	Probably Damaging 1.000	Probably damaging 0.985
Phe240Ser	Decrease stability - CS: -1.00	Decrease stability - CS: -0.977	-6.12Kcal/mol. Destabilising. Torsion favourable	1.51 - Decreases binding affinity. SAiC = 0.00%	Disease	Probably damaging 1.000	Probably damaging 0.995
Glu104Asp	Decrease stability - CS: -0.250	Decrease Stability - CS: -0.741	-3.08Kcal/mol. Destablising. Torsion Favourable	-0.32 - Increases binding affinity. SAiC = 6.44%	Neutral	Probably Damaging 1.000	Probably damaging 1.000
	CS = Confidence Score	CS = Confidence Score	had to use 4POC-not energy minimised	Mutations in Chain AB (therefore results may be for homodimers. SAiC = Solvent accesibility in Complex. For some reason ddg values are reported in opposite manner to previously (minus values mean increased affinity??).			
Ala62Asp	Increase Stability - CS: 0.779	Decrease Stability - CS: -0.578	-2.65Kcal/mol. Destabilising. Torsion Unfavourable	0.65 - Decreases binding affinity. SAiC = 0.00%	Disease	Probably Damaging 0.984	Probably damaging 0.971
Gly72Ala	Decrease stability - CS: -0.738	Decrease Stability - CS: -0.712	-1.3Kcal/mol. Destablising. Torsion Unfavourable	2.29 - Decreases binding affinity. SAiC = 14.19% (Part of interface)	Disease	Probably damaging 1.000	Probably damaging 1.000
Gly122Arg	Increase Stability - CS: 0.413	Increase stability - CS: 0.760	1.51Kcal/mol. Stabilising. Torsion Unfavourable	-0.79 - Increases binding affinity. SAiC = 5.68%	Disease	Possibly damaging 0.929	Benign 0.220
Val154Met	Decrease stability - CS: -0.810	Decrease stability - CS: -0.933	-4.23Kcal/mol. Destabilising. Torsion Unfavourable	-0.20 - Increases binding affinity. SAiC = 7.6%	Neutral	Probably damaging 1.000	Probably damaging 0.977

Supplementary Table S2: Assignment of secondary structure in human TPI (PDB: 4POC) as determined by PyMol

Residue Number	Secondary Structure
3-5	Interconnecting Loop before 1st β-sheet
6-11	1st β-sheet
12-17	Interconnecting loop between 1st β-sheet and 1st α-helix
18-30	1st α-helix
31-36	Interconnecting loop between 1st α-helix and 2nd β-sheet
37-42	2nd β-sheet
43-47	Interconnecting loop between 2nd β-sheet and 3rd α-helix
48-54	2nd α-helix
55-59	Interconnecting loop between 2nd α-helix and 3rd β-sheet
60-63	3rd β-sheet
64-79	Interconnecting loop between 3rd β-sheet and 3rd α-helix
80-86	3rd α-helix
87-89	Interconnecting loop between 3rd α-helix and 4th β-sheet
90 - 93	4th β-sheet
94 - 95	Interconnecting loop between 4th β-sheet and 4th α-helix
96 - 102	4th α-helix
103 - 105	Interconnecting loop between 4th α-helix and 5th α-helix
106 - 119	5th α-helix
120 - 121	Interconnecting loop between 5th α-helix and 5th β-sheet
122 - 127	5th β-sheet
128 - 130	Interconnecting loop between 5th β-sheet and 6th α-helix
131 - 136	6th α-helix
137	Interconnecting loop between 6th α-helix and 7th α-helix
138 - 153	7th α-helix
154 - 159	Interconnecting loop between 7th α-helix and 6th β-sheet
160 - 164	6th β-sheet
165 - 177	Interconnecting loop between 6th β-sheet and 8th α-helix
178 - 196	8th α-helix
197	Interconnecting loop between 8th and 9th α-helices
198 - 203	9th α-helix
204 - 205	Interconnecting loop between 9th α-helix and 7th β-sheet
206-208	7th β-sheet
209 - 216	Interconnecting loop between 7th β-sheet and 10th α-helix
217 - 222	10th α-helix
223 - 227	Interconnecting loop between 10th α-helix and 8th β-sheet
228 - 231	8th β-sheet
232	Interconnecting loop between 8th β-sheet and 11th α-helix
233 - 236	11th α-helix
237	Interconnecting loop between 11th and 12th α-helices
238 - 245	12th α-helix
246 -248	Final loop

Supplementary Table S3: Sequences used to construct sequence alignments.

Species	Accession Number
Homo Sapiens	P60174.3
Ailuropoda melanoleuca	XP_002922306.2
Aotus nancymaae	XP_012328257.1
Bison bison bison	XP_010843235.1
Bos taurus	NP_001013607.1
Canis lupus familiaris	NP_001183983.1
Cavia porcellus	XP_003463291.1
Ceratotherium simum simum	XP_004438751.1
Cercocebus atys	XP_011909761.1
Chinchilla lanigera	XP_005378895.1
Condylura cristata	XP_012588025.1
Dasypus novemcinctus	XP_004448557.1
Equus caballus	XP_001497522.2
Felis catus	XP_006933527.1
Fukomys damarensis	XP_010609974.1
Heterocephalus glaber	XP_004869515.1
Ictidomys tridecemlineatus	XP_005338375.1
Jaculus jaculus	XP_004668721.1
Loxodonta africana	XP_003410703.2
Macaca nemestrina	XP_011743949.1
Mesocricetus auratus	XP_005066069.1
Microcebus murinus	XP_012615900.1
Mus musculus	NP_033441.2
Mustela putorius furo	XP_004766701.1
Myotis brandtii	XP_005873734.1
Ochotona princeps	XP_004596449.1
Octodon degus	XP_004643583.1
Odobenus rosmarus divergens	XP_004413981.1
Orcinus orca	XP_004279116.1
Oryctolagus cuniculus	P00939.1
Otolemur garnettii	XP_012661797.1
Pan troglodytes	NP_001065250.1
Propithecus coquereli	XP_012493283.1
Pteropus vampyrus	XP_011364118.1
Rattus norvegicus	NP_075211.2
Saimiri boliviensis boliviensis	XP_010331175.1
Pongo abelii	NP_001126005.1
Sorex araneus	XP_004610688.1
Sus scrofa	NP_001032228.1
Tupaia chinensis	XP_006170851.1
Xenopus laevis	NP_001080476.1
Xenopus (*Silurana*) *tropicalis*	NP_001107706.1
Brachybacterium faecium DSM 4810	YP_003154941.1

Chlamydia trachomatis 434/Bu	YP_001654656.1
Enterobacter aerogenes KCTC 2190	YP_004591499.1
Escherichia coli O104:H4 str. 2011C-3493	YP_006781500.1
Listeria monocytogenes EGD-e	NP_463876.1
Mesorhizobium ciceri biovar biserrulae WSM1271	YP_004144467.1
Mycoplasma pneumoniae M129	NP_110318.1
Neisseria gonorrhoeae FA 1090	YP_207198.1
Rhodopirellula baltica SH 1	NP_867626.1
Streptococcus mutans UA159	NP_721138.1
Streptomyces coelicolor A3(2)	NP_624890.1
*Thermanaerovibrio acidaminovorans*DSM 6589	YP_003317406.1
Anas platyrhynchos	XP_005016819.1
Aquila chrysaetos canadensis	XP_011585873.1
Corvus cornix cornix	XP_010394421.1
Falco cherrug	XP_005434378.1
Falco peregrinus	XP_005235418.1
Haliaeetus leucocephalus]	XP_010565882.1
Meleagris gallopavo	XP_010716134.1
Pseudopodoces humilis	XP_005523884.1
Pygoscelis adeliae	XP_009327088.1
Gallus gallus	P00940.2
Esox lucius	NP_001290932.1
Haplochromis burtoni	XP_005951088.1
Ictalurus punctatus	NP_001187543.1
Maylandia zebra	XP_004569774.1
Oreochromis niloticus	XP_003452735.1
Pundamilia nyererei	XP_005745225.1
Xiphophorus maculatus	NP_001273230.1
Apis mellifera	NP_001090623.1
Athalia rosae	XP_012257760.1
Bombus impatiens	XP_012244746.1
Bombyx mori	NP_001119730.1
Linepithema humile	XP_012220627.1
Megachile rotundata	XP_003700534.1
Microplitis demolitor	XP_008557282.1
Monomorium pharaonis	XP_012533797.1
Orussus abietinus	XP_012288390.1
Papilio xuthus	XP_013179213.1
Alligator mississippiensis	XP_006273455.1
Alligator sinensis	XP_006037108.1
Anolis carolinensis	XP_003227154.1
Chelonia mydas	XP_007069697.1
Chrysemys picta bellii	XP_005312065.1
Pelodiscus sinensis	XP_006124207.1
Python bivittatus	XP_007420484.1

Thamnophis sirtalis	XP_013926209.1
Arabidopsis thaliana	NP_191104.1
Bathycoccus prasinos	XP_007510733.1
Erythranthe guttata	XP_012858305.1
Glycine max	NP_001237472.1
Medicago truncatula	XP_013450029.1
Morus notabilis	XP_010111131.1
Theobroma cacao	XP_007016406.1
Zea mays	NP_001147215.1
Cicer arietinum	XP_004487007.1
Jatropha curcas	XP_012075954.1
Capronia epimyces CBS 606.96	XP_007733361.1
Fomitiporia mediterranea MF3/22	XP_007261753.1
Fusarium graminearum PH-1	XP_011326337.1
Nematocida parisii ERTm1	XP_013059541.1
Neurospora crassa OR74A	XP_959063.1
Paracoccidioides brasiliensis Pb18	XP_010762134.1
Pneumocystis murina B123	XP_007874541.1
*Pseudogymnoascus destructans*20631-21	XP_012742655.1
Schizosaccharomyces japonicus yFS275	XP_002175311.1
Wickerhamomyces ciferrii	XP_011273813.1
Cryptosporidium muris RN66	XP_002142258.1
Fonticula alba	XP_009498316.1
Gregarina niphandrodes	XP_011129766.1
Leishmania infantum JPCM5	XP_001465915.1
Plasmodium falciparum 3D7	XP_001348552.1
Plasmodium reichenowi	XP_012765443.1
Salpingoeca rosetta	XP_004990529.1
Tetrahymena thermophila SB210	XP_001008794.2
Thecamonas trahens ATCC 50062	XP_013760960.1
Trypanosoma cruzi CL Brener	XP_818253.1

Supplementary Table S4: Degree of conservation of the disease-associated residues

Variant	ScoreCons All species (n=119)	ScoreCons - Mammals only (n=40)
Ile170Val	100%	100%
Phe240Leu	93.9%	91.4%
Cys41Tyr	43.8%	85.0%
Val231Met	92.5%	91.4%
Phe240Ser	93.9%	91.4%
Glu104Asp	100%	100%