A theoretical study to establish the relationship between the three-dimensional structure of triose-phosphate isomerase of Giardia duodenalis and point mutations in the respective gene

Author

Nolan, Matthew J, Hofmann, Andreas, Jex, Aaron R, Gasser, Robin B

Published

2010

Journal Title

Molecular and Cellular Probes

DOI

10.1016/j.mcp.2010.06.001

Molecular and Cellular Probes

(~2600 words)

A theoretical study to establish the relationship between the three-dimensional structure of triose-phosphate isomerase of *Giardia duodenalis* and point mutations in the respective gene

Matthew J. Nolan[a1], Andreas Hofmann[b1], Aaron R. Jex[a] and Robin B. Gasser[a*]

[a]*Department of Veterinary Science, The University of Melbourne, Werribee, Victoria 3030, Australia*
[b]*Structural Chemistry Program, Eskitis Institute for Cell & Molecular Therapies, Griffith University, Brisbane, Queensland 4111, Australia*

* *Correspondence:* Professor Robin B. Gasser, Department of Veterinary Science, The University of Melbourne, 250 Princes Highway, Werribee, Victoria 3030, Australia.
E-mail: Robin B. Gasser (robinbg@unimelb.edu.au).
Fax: +61-3-97312366.

[1] Equal contributions to this work.

Abbreviations: ***tpi***, triose-phosphate isomerase gene; **TPI**, triose-phosphate isomerase

Keywords:
Giardia
Triose-phosphate isomerase
Nucleotide polymorphism

ABSTRACT

Predicting how point mutations in genes alter the tertiary and quarternary structure of proteins is central to a number of areas of molecular biology and has implications in relation to the function and evolution of molecules. In the present study, we theoretically assessed the effects of 20 point mutations detected previously in a region of the triose-phosphate isomerase gene (*tpi*) of the protozoan *Giardia duodenalis* on the three-dimensional structure of the 'wild-type' protein (TPI). Amino acid substitutions arising from codon variations were mainly located at surface-accessible sites or in hydrophobic pockets of TPI. None of the substitutions was predicted to exert a significant change to the fold or functionality of the enzyme, with the exception of one alteration (Arg100). Almost all substitutions were either conservative or semi-conservative, and retained or even improved the expected stability of the fold. Overall, the findings provide support for the "neutral theory", which contends that evolution at the molecular level is not solely shaped by "Darwinian selection but also by random drift of selectively neutral or nearly neutral mutants".

1. Introduction

Giardia (Zoomastigophorea) is a genus of parasitic, flagellate protozoa that infect the small intestine of vertebrates . In humans, *Giardia* infections are predominantly caused by *G. duodenalis* (syn. *G. lamblia* and *G. intestinalis*) assemblages 'A' or 'B' . Infections are usually acquired following the ingestion of cysts (= infective stage) from contaminated food, water or *via* direct human-to-human contact . Although patent infections may be asymptomatic, the clinical manifestation of giardiasis in humans includes diarrhoea, bloating, colic, nausea, headache and/or weight loss . Clinical signs can vary significantly among individuals and have been proposed to be linked to infective dose , particular genotypes (or 'strains') of *Giardia* and/or host factors, such as age, and nutritional and immune status . Giardiasis is often short-lived in immunocompetent hosts and the infection is eliminated following the development of an effective immune response . However, chronic giardiasis can occur in some individuals, such as the elderly and immuno-compromised or -suppressed patients , causing malabsorption and/or severe weight loss .

The accurate diagnosis of giardiasis, including the identification and characterisation of *Giardia* species and/or assemblages, is central to the prevention and control of this disease in humans and a wide range of animals . Significant limitations in traditional methods of identification/diagnosis (including microscopy and immunodetection) do not allow the assignment of *G. duodenalis* to assemblages (A-G) . Given that only some of these assemblages are known to infect humans (*i.e.*, 'A' or 'B'), the genetic characterisation of *Giardia* isolates is imperative in evaluating their zoonotic potential. In addition to allowing the specific and sensitive detection and/or identification of cysts, molecular tools offer insights into *Giardia* populations and epidemiology that support the prevention and control of giardiasis .

Recently, Nolan et al. carried out a comprehensive molecular study of *Giardia* infecting young sheep on pasture-based farms in southern Australia, with a particular focus on the genetic characterisation of *G. duodenalis* with the potential to infect humans. A PCR-coupled restriction endonuclease fingerprinting (REF)-sequencing-phylogenetic analysis of a portion of the triose-phosphate isomerase (TPI or TIM; ref. ; EC 5.3.1.1) gene was undertaken. This single-copy gene was utilized because it has been reported to be well suited and informative for population genetic investigations and because substantial sequence data are available in public databases for comparative analysis. Fourteen different sequence variants representing two distinct assemblages of *G. duodenalis* (A and E) were identified . Polymorphism at one to five nucleotide positions (external to the those representing the active site of TPI; PROSITE code PS00171; ref.) were detected in seven of 14 p*tpi* sequence types, representing all of the variation detectable in this locus among 43 *G. duodenalis* samples . As there was no evidence of any artefacts produced in PCR or

sequencing, the polymorphism observed was interpreted to reflect variation within or among *Giardia* individuals as a consequence of evolutionary divergence (populations with mixed lineages within one host) or mutations occurring during the growth and replication of the parasite within the host. Therefore, given that the nucleotide polymorphism detected by Nolan et al. was consistent with but greater than that detected in previous studies , we focused on addressing the basic question as to whether polymorphisms at individual nucleotide positions (separately or in combination) in the *tpi* gene cause alterations to the three-dimensional structure of the 'wild-type' protein TPI employing a theoretical approach. This enzyme, TPI, is present within the cytoplasm , catalyses the reversible isomerization between dihydroxyacetone phosphate (DHAP) and (R)-glyceraldehyde 3-phosphate (GAP) in the glycolytic pathway, and is essential for energy production .

2. Materials and methods

Previously , we detected point mutations in seven p*tpi* sequences from *G. duodenalis* from sheep (n = 43 samples) and subsequently appraised the literature for polymorphic nucleotide positions in other p*tpi* sequences representing *Giardia*. Polymorphism was identified in sequences from samples classified as belonging to assemblage A or B from a variety of hosts (*i.e.*, human, barbary macaque, chimpanzee, mantled guereza, common eider and herring gull) ; however, much of the data were not available in public gene databases. Epidemiological information pertaining to the p*tpi* sequences that are available from previous studies are listed in Table 1, and the polymorphic positions in each nucleotide sequence are given in Table 2. The effects of individual amino acid mutations on the three-dimensional structure of wild-type TPI were inferred based on an inspection of the *G. duodenalis* TPI crystal structure (PDB accession code 2dp3) described previously . Models for single mutations were generated and visualised with O .

3. Results

Using the amino acid sequences inferred by translation of the open reading frames (ORFs) of individual sequences, substitutions were identified by comparative analysis against a reference ('wild-type') sequence of the protein TPI for *G. duodenalis*. The resultant mutations were mapped to the three-dimensional structure of this protein. Most of the amino acid variations were either conservative in nature or added to the overall stability of the molecule (see Table 3). Topologically, amino acid substitutions, due to genetic polymorphism detected in the current datasets (a total of 20 nucleotide positions), occurred mainly in the α-β-loop areas and α-helices at surface-accessible sites as well as in internal hydrophobic pockets. The only significant variations identified were

those of Arg100Cys and Ser110Gly in the sequence represented by GenBank accession number GQ444451. The side chain of Arg100 in helix α4 was inferred to be involved in stabilising the fold by forming a hydrogen bond interaction with the backbone carbonyl groups of Gly95 (loop 4), Cys127 (β5) and the carboxylate side chain of Asp107 (α4). A replacement of Arg100 by Cys leads to the loss of this stabilising interaction. The side chain hydroxyl group of Ser110 (α4) is hydrogen-bonded with the carboxylate of Glu105 (α4) and the main chain carbonyl of Thr106 (α4). Thus, the replacement Ser110Gly has the potential to cause only minor destabilisation, since all of these interactions occur within one secondary structure element.

Non-local hydrogen bonds between main chain amide hydrogen atoms and polar side chain acceptors have been identified as a characteristic stabilising feature of TIM barrel proteins . Such βα-hairpin clamps bracket consecutive β–α secondary structure elements for which the main chain amide group at the start of the β-strand interacts with a polar side chain located in the C-terminal region of the consecutive α-helix. Interestingly, such hairpin clamps are hardly observed in *Giardia* TPI. Only three non-local main chain-side chain interactions were identified, similar to the βα-hairpin clamps reported previously (see Table 4) . Furthermore, the amino acid substitutions inferred in this study do not alter these stabilising βα-brackets.

4. Discussion

The homodimeric enzyme triose-phosphate isomerase catalyses the reversible isomerization of glyceraldehyde 3-phosphate to the ketone dihydroxyacetone phosphate, without need for a metal ion or a cofactor . The enzyme is relatively conserved throughout the animal and/or plant kingdoms, and has been well characterised . It plays a key role in metabolic pathways and energy production. TPIs adopt the $(\beta\alpha)_8$ TIM barrel fold, which is one of the most common protein folds, enabling a variety of essential catalytic functions . Eight central β-strands are surrounded by eight α-helices, which are joined to the preceding and subsequent β-strand by loops. It is generally believed that the βαβ motif, comprising a pair of adjacent parallel β-strands and the interspersed anti-parallel α-helix, forms the minimal folding unit and serves as the building block for several βα-repeat structures, such as the Rossman, leucine-rich and flavodoxin folds .

Typically, TPIs form dimers through intimate contacts of loops 1 and 3 that connect β-strand 1 and α-helix 1, and β-strand 3 and α-helix 3, respectively. TPIs from two thermophilic organisms have been found to exist as tetramers . Previously, TPI from *Giardia* has been shown to exist in dimeric and oligomeric states where the latter species are formed in a RedOx-dependent manner .

The catalytic residues are strictly conserved in all known TPIs, and conserved residues in the vicinity of the active site are believed to correctly position the catalytic triad . However, structure-function studies of TPI have revealed that conservation measured as amino acid sequence identity may be a poor indicator of the actual level of conservation. For example, although TPIs from *Trypanosoma brucei* and *T. cruzi* (both Zoomastigophorea) share high amino acid sequence identity (73.7% across 251 amino acid residues; not shown), they have been shown to possess differential reactivity to biochemical reagents , susceptibility to proteolytic degradation , and folding stability . Several regions of TPI are of major importance for the proper functioning of the enzyme: (i) the substrate binding site with its catalytic triad (Lys13, His96, Glu170) , (ii) the dimer interface (loops 1, 3) , (iii) active site loop (loop 6) and nearby loop 7 , and, (iv) fold-stabilising elements (*e.g.* βα-hairpin clamps; ref).

In the present study, we theoretically evaluated effects of nucleotide polymorphisms (n = 20) reported in previous investigations on the molecular structure of TPI. Amino acid substitutions arising from codon variations were mainly located at surface-accessible sites or in hydrophobic pockets. Based on the evaluation, neither of the substitutions was inferred to exert a significant change to the fold or general functionality of the enzyme, with the exception of Arg100. None of the amino acid variations occurred in any of the four regions referred to as being critical (i-iv; see above). Most of the substitutions were either conservative or semi-conservative, and retained or even improved the expected stability of the fold. The catalytic triad remained conserved, and residues in loops 1 and 3 (dimer interface) as well as loops 6 and 7 (regulation of catalytic activity) were also not substituted. In addition, general stabilising features, such as βα-brackets, remained unchanged.

The only amino acid substitution predicted to cause an alteration of the conformation of *Giardia* TPI and enzymatic activity was Arg100Cys from the sequence represented by accession number GQ444451. Given the recent report about RedOx-dependent oligomerization and activity of *Giardia* TPI , this amino acid substitution may have far-reaching consequences. First, the side chain of Cys100 will be accessible on the surface, and the residue might thus engage in inter-molecular dithioether-linkages. The proposed model suggested that *Giardia* TPI forms dimers *via* the canonical dimerization interface (loops 1 and 3), thereby presenting the surface-accessible Cys202 at diametral positions of the dimer. These residues can engage in disulfide formation to yield tetramers and higher order oligomers . Although it is anticipated that the accessibility of Cys202 is greater than that of Cys100, it cannot be excluded that the latter residue might also engage in intermolecular disulfide linkages. Second, Cys127(β5) is positioned not too far from residue 100(α4). At a first glance, it seemed unlikely that Cys100 and Cys127 could engage in an intra-molecular dithioether-bond, since the distance between Cβ(100) and Cβ(127) was inferred to be 7.7

Å. Formation of a covalent bond between both cysteine residues would thus require a significant conformational change in the core of protein, pulling together α4 and β5. The covalent link would also force a re-arrangement of the N-terminus of helix α4, in which the catalytic residue His96 is located.

Reyes-Vivas et al. suggested, based on results of enzyme assays and mutagenesis, that native *Giardia* TPI possesses an intra-molecular disulfide-bridge formed by residues Cys222 and Cys228. In this context, it is interesting that the Cβ-Cβ distance of 10.8 Å between these two residues is even greater than that between Arg100 and Cys127. Additionally, Cys228 is positioned 1.5 turns higher on helix α7 and points in the opposite direction to Cys222. The formation of an intramolecular disulfide link between these two residues requires an almost complete unfolding of helix α7, which is also visible in the model obtained using a molecular dynamic simulation conducted previously . If an internal disulfide link is indeed formed between Cys100 and Cys127, it is very likely that this oxidised form of the *Giardia* TPI Arg100Cys variant would possess a greatly reduced catalytic activity.

5. Conclusion

The neutral theory contends that evolution at the molecular level is not solely shaped by "Darwinian selection but also by random drift of selectively neutral or nearly neutral mutants". This theory does not deny the role of natural selection, rather it assumes that only a small proportion of changes are adaptive, whereas the "majority of phenotypically silent molecular substitutions exert no significant influence on survival and reproduction and drift randomly through a species" . Structural properties associated with protein evolution are of substantial interest, because they connect genetic polymorphism with functional implications. The property, having received most attention since the early days of structural biology, is the level of solvent exposure, based on the repeated observation that 'exposed' residues are more likely to undergo evolutionary substitution than residues 'buried' in the core of a protein . However, it has also been pointed out that a simple classification of exposed *versus* buried residues might oversimplify the situation ; the molecular context at a residue level has been acknowledged recently as a contributing factor influencing the nature and rate of mutational change.

The amino acid variations due to polymorphic nucleotides in a portion of the triose-phosphate isomerase gene of *Giardia* assessed in this study support some of the current theories of molecular evolution . It is noteworthy that amino acid substitutions also occur in the non-solvent exposed areas of TPI, but the substitutions retain the local 'packing effects' .

The surprising finding in this study was the replacement Arg100Cys (*cf.* nucleotide sequence with GenBank accession number GQ444451), which was inferred to cause a loss of stability, because three hydrogen bonds between the side-chains of Arg100 in helix α4, with the backbone carbonyl groups of Gly95 (loop 4), Cys127 (β5) and the carboxylate side chain of Asp107 (α4) are lost. Thus, one might speculate that this amino acid substitution adds a particular functionality to TPI, which may comprise the intra-molecular disulfide formation between Cys100 and Cys127. Due to a concomitant conformational change, the catalytic activity of TPI is likely to be affected. The intriguing similarity between Cys100 and Cys127, with the previously described intra-molecular disulfide bond between Cys222 and Cys228 and the inferred suppression of TPI catalytic activity raises the hypothesis that RedOx regulation of TPI catalysis is a general mechanism (*i.e.*,).

Acknowledgements

Funding support from Melbourne Water Corporation and the Australian Research Council project is gratefully acknowledged.

References

Table 1
Relevant information for *Giardia duodenalis* samples (represented by GenBank accession numbers) for which polymorphic nucleotide positions have been reported for a portion of the triose-phosphate isomerase (p*tpi*) gene.

Species Inferred assemblage	Host or environmental origin	Locality	GenBank accession nos.	References
Giardia duodenalis				
A	Sheep (lamb) (*Ovis aries*)	Australia	GQ444448, GQ444451	
	Common eider (*Somateria mollissima*)	USA	EU518540	
	Herring gull (faeces) (*Larus argentatus*)	Canada	EU518535, EU518537	
B	Chimpanzee (*Pan troglodytes*)	Belgium	FJ890953	
	Herring gull (from faeces)	Canada	EU518581	
	Mantled guereza (*Colobus guereza*)	Belgium	FJ890959, FJ890960	
E	Sheep	Australia	GQ444454, GQ444458-GQ444460, GQ444462	

Table 2

Fourteen partial triose-phosphate isomerase (p*tpi*) gene sequences (represented by GenBank accession numbers) in which polymorphic nucleotide positions have been reported, and the amino acid residue and codon position at which the polymorphism occurs [compared to Uniprot sequence P36186 (http://www.uniprot.org/uniprot/P36186)].

	Polymorphic sites																				
Amino acid	24	35	37	44	56	63	70	72	100	110	121	121	122	128	141	143	160	164	166	225	
Codon position	3	3	1	3	3	3	3	3	1	1	2	3	1	2	1	3	1	2	1	3	References
Sequence																					
GQ444448														Y							[16]
GQ444451									Y	R					R						[16]
GQ444460		K									R	K									[16]
GQ444462		K																R			[16]
GQ444454	Y	K	R															R			[16]
GQ444459	Y	K	R	K									R								[16]
GQ444458	Y																				[16]
EU518535																				Y	[21]
EU518537																			R		[21]
EU518540																				Y	[21]
EU518581																	R				[21]
FJ890953					Y		R														[22]
FJ890959						R		Y													[22]
FJ890960						R		Y								R					[22]

Table 3
A structural evaluation of amino acid variation as a consequence of individual and/or combined nucleotide polymorphism in the p*tpi* gene.

Variation	TPI sequence	Topological element	Location	Effect
D35E	GQ444454, GQ444458 - GQ444460, GQ444462,	α1β2	Surface accessible	Conservative
V37I	GQ444454, GQ444459	α1β2	Hydrophobic pocket	Conservative
R100C	GQ444451	α4	Surface accessible	Destabilising; possible dithioether link with Cys127?
S110G	GQ444451	α4	Not fully surface accessible	Maybe destabilising; interactions with E105 and T106-CO lost
G121E	GQ444460	α4 β5	Surface accessible	Stabilising; hydrophilic interactions with solvent
G121D	GQ444460	α4 β5	Surface accessible	Stabilising; hydrophilic interactions with solvent
M122V	GQ444459	α4 β5	Hydrophobic pocket	Semi-conservative
V128A	GQ444448	β5	Hydrophobic pocket	Conservative
M141V	GQ444451	α5	Surface accessible	Semi-conservative
E157K	EU518581	α5	Surface accessible	Change of electrostatic potential, but hydrophilic interactions conserved
K163R	GQ444454, GQ444462	α5 β6	Surface accessible	Conservative
V166I	EU518537	β6	Hydrophobic pocket	Conservative

Note: 'βn' denotes the n-th beta-strand, 'αn' the helix following the n-th alpha-helix, and 'αnβ(n+1)' denotes the linker peptide between the n-th alpha-helix and the subsequent beta-strand.

Table 4
βα brackets in *Giardia* TPI (PDB accession code 2dp3)

Topological elements	Interaction	Distance (Å)
$b\bar{5}a5$	G129-NH - Q147	2.9
$b\bar{8}a8$	I231-CO - R256	2.7
$a\bar{8}b1$	S240 - N11-NH	3.6

Figure legend

Fig. 1. Mapping of amino acid substitutions, due to nucleotide polymorphism, onto the three-dimensional fold of *G. duodenalis* TPI (PDB accession code 2dp3) . The positions of amino acid variations listed in Table 3 are indicated using the following colour code: conservative substitution (green), stabilising substitution (blue), destabilising substitution (red). The variation of Arg100 to Cys is indicated in yellow. Catalytic residues are drawn explicitly and coloured orange. The loops involved in dimer formation (loop 1, left; loop3, right) are shown in magenta. Figure prepared with PyMOL

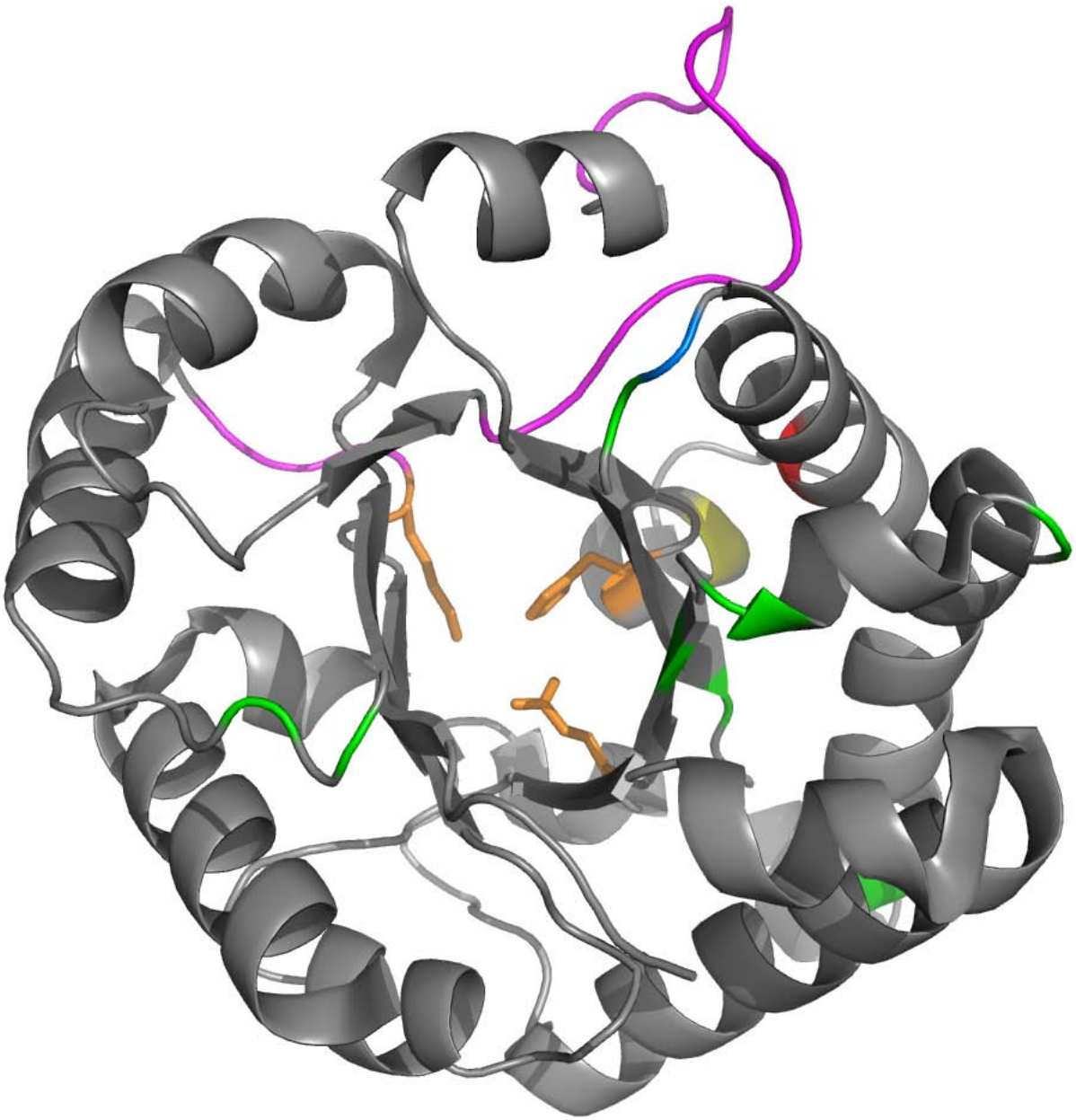

Fig. 1 – Nolan et al.