

**Biochemical characterisation of unusual glycolytic
enzymes from the human intestinal parasite
Blastocystis hominis.**

Submitted by

Sheera Abdulla to the University of Exeter as a thesis for the degree of

Doctor of Philosophy in Biological Sciences in June 2016

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: Sheera Abdulla

Abstract

Blastocystis is an important parasite that infects humans and a wide range of animals like rats, birds, reptiles, etc. infecting a sum of 60% of world population. It belongs to the Stramenopiles, a Heterologous group that includes for example the *Phytophthora infestans* the responsible for the Irish potato famine. Previous work had reported the presence of an unusual fusion protein that is composed of two of the main glycolytic enzymes; Triosephosphate isomerase-glyceraldehyde-3-phosphate dehydrogenase (TPI-GAPDH). Little is known about this protein. *Blastocystis* TPI-GAPDH and *Blastocystis* enolase were both characterized biochemically and biophysically in this project. The phylogenetic relationships of those two proteins among other members of either Stramenopiles, or other members of the kingdom of life were examined and found to be grouping within the chromalveolates. Our studies revealed that those two proteins, *Blastocystis* enolase and *Blastocystis* TPI-GAPDH, had a peptide signal targeting them to the mitochondria. This was an unusual finding knowing that text books always referred to the glycolytic pathway as a canonical cytoplasmic pathway. Structural studies had also been conducted to unravel the unknown structure of the fusion protein *Blastocystis* TPI-GAPDH. X-ray crystallography had been conducted to solve the protein structure and the protein was found to be a tetrameric protein composed of a central tetrameric GAPDH protein flanked with two dimmers of TPI protein. Solving its structure would be the starting point towards reviling the role that TPI-GAPDH might play in *Blastocystis* and other organisms that it was found in as well. Although a fusion protein, the individual components of the fusion were found to contain all features deemed essential for function for TPI and GAPDH and contain all expected protein motifs for these enzymes.

Contents

Abbreviations.....	17
Abstract.....	20
Chapter 1. Introduction.....	21
1.1. <i>Blastocystis</i>	21
1.1.1. History and background.....	21
1.1.2. Nomenclature.....	24
1.1.3. Structure of morphological forms and proposed functions.....	25
1.1.4. Epidemiology and prevalence.....	30
1.1.5. Genetic diversity and specific correlation with the host.....	31
1.1.6. Mode of transmission, people with high risk and clinical symptoms.....	32
1.1.7. Insight into <i>Blastocystis</i> pathogenesis.....	34
1.1.8. Linkage with irritable bowel syndrome (IBS).....	36
1.1.9. Treatment.....	38
1.1.10. Mitochondrial derived organelles.....	40
1.1.10.1. Blastocystis mitochondrial organelle (MLO).....	40
1.2. Glycolysis.....	41
1.2.1. Enolase (Phosphopyruvate hydratase).....	44
1.2.1.1. Enolase structure.....	45
1.2.1.2. Enolase function and moon lighting.....	46
1.2.2. Glyceraldehyde-3-phosphatedehydrogenase (GAPDH).....	47
1.2.3. Trioephosphate isomerase (TPI).....	51
1.2.3.2. Triosephoashate isomerase physiological roles.....	55
1.2.4. Moonlighting proteins pros and cons.....	56
1.2.5. Aim of the project.....	58
Chapter 2. Materials and methods.....	59
2.1. Materials.....	59
2.1.1. Microbiological work.....	59

2.1.1.1. Bacterial and Protist strains.....	59
2.1.1.2. Bacterial and protist media and solutions.....	60
2.1.1.3. Protist media and solutions	63
2.1.2. Molecular biology.....	67
2.1.2.1. Primers list.....	67
2.1.2.2 Plasmids list	69
2.1.2.3. Molecular biology Solutions.....	70
2.1.2.3.1. DNA work solutions	70
2.1.2.3.2. Protein work solutions	71
2.1.2.3.2.1. Sodium dodecyl sulphate (SDS) gel and solutions	71
2.1.2.3.2.2. Native gel solutions	74
2.1.2.3.2.3. Western blotting solutions	76
2.1.2.3.2.4. Colony hybridization solutions (colony blot)	77
2.1.2.3.2.5. Enzyme activity assay solutions.....	77
2.1.2.3.2.5.1. Blastocystis enolase activity assay solution.....	77
2.1.2.3.2.6. Protein purification solutions.....	78
2.1.2.3.2.7. Crystallography work.	80
2.1.2.3.2.8. Differential scanning fluorimetry (DSF) buffers.....	80
2.2 Methods	82
2.2.1. Molecular biology methods	82
2.2.1.1 DNA extraction.....	82
2.2.1.2. PCR reactions	83
2.2.1.2.1. PCR cycling conditions	83
2.2.1.2.2 Mutagenesis PCR conditions	85
2.2.1.3. DNA electrophoresis	86
2.2.1.4. Isolation and purification of DNA fragments.....	87
2.2.1.5. Cloning of PCR products.	87
2.2.1.5.1. Ligation into cloning vector.....	87
2.2.1.5.2. Transformation of competent cells with plasmid DNA	88

2.2.1.5.3. Plasmid DNA isolation.....	88
2.2.1.5.4. Sanger sequencing.....	89
2.2.1.5.5. Restriction enzyme digestion of DNA	89
2.2.1.5.6. Cloning of DNA fragment into expression vector.....	90
2.2.2. Microbiological methods.....	92
2.2.2.1. Growth measurement.....	92
2.2.2.2. E. coli growth	92
2.2.2.3. Long-term storage of bacterial clones.....	92
2.2.2.4. Bacterial harvesting	92
2.2.2.5. Blastocystis purification and harvesting.....	92
2.2.2.6. Long-term preservation of Blastocystis sp. strains	93
2.2.2.7. Recovering cryo-preserved Blastocystis cells	93
2.2.2.8. Blastocystis ST cultivation method.....	94
2.2.2.8.1. Xenic culturing method	94
2.2.2.8.2. Axenization trials.....	94
2.2.3. Biochemistry methods	95
2.2.3.1. Induction of gene expression	95
2.2.3.2. Cell lysis.....	95
2.2.3.3. Protein purification	96
2.2.3.3.1. Immobilized metal affinity chromatography (IMAC).....	96
2.2.3.3.2. Ion Exchange Chromatography (IEC)	97
2.2.3.3.3. Gel Filtration chromatography (GF column)	97
2.2.3.4. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)	99
2.2.3.5. Native gel.....	100
2.2.3.6. Second dimension SDS-PAGE.....	102
2.2.3.7. Western blotting.....	102
2.2.3.8. Colony hybridization (colony blot).....	103
1.2.3.9. Blastocystis enolase assay	105

1.2.3.10.	Biochemical and biophysical characterisation of the proteins	107
2.2.3.10.1.	BSA Standard curve.....	107
2.2.3.10.2.	Dynamic Light scattering, DLS	107
2.2.3.10.3.	Differential Scanning Fluorimetry, DSF	108
2.2.3.10.4.	Protein parameter determination:.....	110
2.2.3.10.5.	Analytical ultra-centrifugation:.....	111
2.2.3.10.6.	Crystallization work	112
2.2.3.10.6.1.	Initial crystal trials	112
2.2.3.10.6.2.	Crystal harvesting and X-ray data collection.....	113
2.2.3.10.6.3.	Crystallization optimization	114
2.2.3.10.7.	Enzyme activity assay.....	114
2.2.3.10.8.	Homology modelling.....	114
2.2.3.10.8.1.	Template selection	115
2.2.3.10.8.2.	Target -template alignment.....	115
2.2.3.10.8.3.	Model building and inspection	115
Chapter 3.	<i>Bioinformatic analysis of the Blastocystis glycolytic enzyme enolase</i>	116
3.1.	Enolase super family.....	117
3.2.	Pre-sequence and signature motif	117
3.2.1.	Enolase pre-sequence analysis	120
3.2.2.	<i>Blastocystis</i> enolase signature motif and amino acid discrepancies analysis 125	
3.3.	Homology modelling.....	130
3.4.	Phylogentic relationships of <i>Blastocystis</i> enolase and other enolases	134
3.4.1.	Sequence alignment and phylogenetic analysis.....	134
3.4.1.1.	Placing of <i>Blastocystis</i> enolase in the kingdom of life	139
3.5.	Conclusion	142
Chapter 4.	Biochemical characterization of enolase.....	143
4.1.	General background (introduction)	143

4.2. Recombinant <i>Blastocystis</i> Enolase was successfully cloned and over-expressed in <i>E. coli</i>	145
4.2.1. Cloning and amplification of recombinant <i>Blastocystis</i> enolase	145
4.2.2. Recombinant <i>Blastocystis</i> enolase is deposited in the inclusion bodies.	147
4.3. Identification of recombinant <i>Blastocystis</i> enolase by mass spectrometry	149
4.4. <i>Blastocystis</i> enolase expression and purification	150
4.5. Biochemical characterization	156
4.5.1. Molecular weight determination.....	156
4.5.1.1. Analytical gel filtration.....	156
4.5.1.2. Immunodetection using a heterologous antibody	158
4.5.1.3. Oligomeric status determination of the native enolase	159
4.5.1.4. Size distribution profile using Dynamic Light Scattering (DLS).....	161
4.5.2. Fluorescent-based thermal shift assay (DFS: differential scanning fluorimetry)	165
4.6. Activity assay and kinetic parameters	167
4.7. Summary	172
Chapter 5. <i>In silico</i> characterisation of recombinant <i>Blastocystis</i> TPI-GAPDH fusion protein.....	176
5.1. Individual proteins super families.....	177
5.1.1. TPI super family	177
5.1.2. GAPDH super family	180
5.2. <i>In silico</i> characterization of <i>Blastocystis</i> TPI-GAPDH; pre-sequence and signature motif.....	182
5.2.1. TPI-GAPDH Pre-sequence analysis.....	182
5.3. Phylogentic relationships of <i>Blastocystis</i> TPI-GAPDH and other organisms	186
5.4. Conclusion	190
Chapter 6. Biochemical characterisation of the unusual TPI-GAPDH fusion protein from <i>Blastocystis</i>	192
6.1. General background (introduction)	192

6.2. Recombinant <i>Blastocystis</i> TPI-GAPDH was successfully cloned and over-expressed in <i>Escherichia coli</i>	195
6.2.1. Cloning and amplification of <i>Blastocystis</i> TPI-GAPDH (B.TPI-GAPDH)	195
6.2.2. Recombinant <i>Blastocystis</i> TPI-GAPDH Expression trials.....	197
6.3. <i>Blastocystis</i> TPI-GAPDH expression and purification.....	201
6.4. Biochemical characterization	209
6.4.1. Molecular weight determination.....	209
6.4.1.1. Analytical GF.....	209
6.4.1.2. Immune-detection (Western Blotting) proved the identity of the purified protein to be TPI-GAPDH.....	211
6.4.1.3 Oligomeric status determination of the native TPI-GAPDH.....	212
6.4.1.4. Size distribution profile using Dynamic Light Scattering	215
6.4.1.5. Analytical ultracentrifugation (AUC)	217
6.4.1.5.1. Complementary work on triosephosphate isomerase (TPI) and Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)	220
6.4.2. Fluorescent-based thermal shift assay (DSF: differential scanning fluorimetry)	224
6.4.3. Site directed mutagenesis on recombinant fusion <i>Blastocystis</i> TPI-GAPDH protein.....	227
6.5. Discussion.....	231
Chapter 7. Crystallization and structure determination of the fusion recombinant <i>Blastocystis</i> TPI-GAPDH protein.....	234
7.1. Introduction	234
7.1.1. Background to protein crystallization.....	234
7.1.2. X-ray crystallography	238
7.2. Materials and methods	245
7.2.1. Crystallization	245
7.2.1.1. Protein preparation.....	245
7.2.1.2. Pre – crystallization test: PCT.....	245
7.2.1.3. Initial crystal trials	248

7.2.1.4.	Optimization of crystallization	249
7.2.1.5.	Co – Crystallization	250
7.2.2.	Preparing crystals for data collection.....	250
7.2.2.1.	Cryo-cooling straight from droplet using cryo-protectant.....	250
7.2.3.	X-ray data collection.....	254
7.2.3.1.	Synchrotron data collection.....	254
7.2.4.	Structure determination	255
7.2.4.1.	Data processing.....	255
7.2.4.2.	Phase determination	255
7.2.4.3.	Model building and refinement	255
7.2.4.4.	Structure validation.....	256
7.3.	Results	256
7.3.1.	Crystallization of the <i>Blastocystis</i> TPI-GAPDH fusion protein.....	256
7.3.2.	Crystallization of the <i>Blastocystis</i> GAPDH protein	258
7.3.3.	Structure determination of the fusion protein	260
7.3.3.1.	X–ray data collection.....	260
7.3.3.2.	Structure solution.....	261
7.3.3.3.	Model building and validation	262
7.3.3.3.1.	Apoenzyme	262
7.4.	Discussion	264
7.4.1.	Secondary structure analysis:.....	264
7.4.2.	Active sites and signature motifs analysis	266
7.4.3.	Tertiary structure analysis	268
7.4.3.1.	Triosephosphate isomerase (TPI).....	268
7.4.3.1.1.1.	Active site: catalytic residues and catalytic region	268
7.4.3.2.	Glyeraldehyde 3- phosphate dehydrogenase (GAPDH)	274
7.4.3.3.	TPI-GAPDH quaternary structure	277
7.5.	Discussion	282
Chapter 8.	Conclusion	287

Table of figures

<i>Figure 1.1 Morphological forms of Blastocystis sp. subtype 4 by phase-contrast microscopy</i>	28
<i>Figure 1.2 Simple representation of the Embden-Mayerhof-Parasan, the complete glycolytic pathway</i>	44
<i>Figure 1.3 Schematic representation of triosephosphate isomerase structure</i>	53
<i>Figure 1.4 Cartoon representation of the conformational changes of loop 6 in triosephosphate isomerase protein when changing between holo and apo enzyme of the protein</i>	54
<i>Figure 2.1 Bioline Hyperladder I DNA marker</i>	71
<i>Figure 2.2 Spectra Multicolor Broad Range Protein Ladder Fermentas</i>	73
<i>Figure 2.3 The general cloning vector pGEM-T-Easy showing the multiple cloning site (MCS)</i> 87	
<i>Figure 2.4 The gene expression plasmid pET-14b featuring a backbone size of 4.6 kb</i>	91
<i>Figure 2.5. Unstained native protein standard NativeMark™ for native gel from Life Technologies</i>	102
<i>Figure 2.6 Schematic diagram representing the theory of thermal shift assay</i>	109
<i>Figure 3.1 Predicted domain structure of Blastocystis enolase by Pfam</i>	117
<i>Figure 3.2 Multiple alignment of deduced amino acid sequences of Blastocystis enolase with proteins from taxonomically diverse organisms</i>	119
<i>Figure 3.3 Analysis of the N-terminal region of Blastocystis enolase in comparison to other organisms</i>	122
<i>Figure 3.4 Sequence alignment of the amino terminal region of enolases from stramenopiles</i>	122
<i>Figure 3.5 Alignment of the amino acid sequences of the enolases</i>	130
<i>Figure 3.6 Cartoon representation of the modelled overall structure of Blastocystis enolase, shown as a dimer</i>	132
<i>Figure 3.7 Blastocystis enolase model aligned with the template human neuron specific enolase (hNSE) (PDB 3UCD), showing almost identical overall topology between the two proteins</i>	132
<i>Figure 3.8 A close up in Blastocystis enolase model structure built using human neuron specific enolase (hNSE) (PDB 3UCD)</i>	133
<i>Figure 3.9 A close up in Blastocystis enolase model structure built using human neuron specific enolase (hNSE) (PDB 3UCD)</i>	133
<i>Figure 3.10 A close up in Blastocystis enolase model structure built using human neuron specific enolase (hNSE) (PDB 3UCD) as a template</i>	133

Figure 3. 11 Relationship of Blastocystis enolase among Eukaryotes and Prokaryotes as determined by phylogenetic analysis of protein sequence.....	140
Figure 3. 12 Phylogenetic tree (Maximum-likelihood and confirmed by MrBayes) of Blastocystis enolase and enolases from Stramenopila group.	141
Figure 4. 1 Agarose gel electrophoresis analysis of PCR-amplified enolase, from Blastocystis ST4.	145
Figure 4. 2 Agarose gel electrophoresis analysis of PCR-amplified Blastocystis enolase with added NdeI, BamHI restriction sites.	146
Figure 4. 3 Schematic representation of Blastocystis enolase-pET-14b construct of 5,968 kb size with the restriction sites added.....	147
Figure 4. 4 Identification of recombinant enolase Blastocystis by mass spectrometry.....	150
Figure 4. 5 Expression trials for Blastocystis enolase in different cell lines under several conditions.....	153
Figure 4. 6 Elution profile of Blastocystis enolase from a HisTrap column (GE Healthcare).	154
Figure 4. 7 Coomassie-stained SDS-PAGE gel of recombinant Blastocystis enolase purified by His-trap column.....	154
Figure 4. 8 Elution profile of Blastocystis enolase from a Superdex 200 gel filtration (GF) chromatography column.....	155
Figure 4. 9 Coomassie-stained SDS-PAGE gel of gel filtrated Blastocystis enolase.....	155
Figure 4. 10 Elution profile of Blastocystis enolase analysed on an analytical Superdex 75 GF chromatography column in triplicate.....	157
Figure 4. 11 Western blot of Blastocystis total protein lysate and recombinant Blastocystis enolase with anti- histidine monoclonal antibodies and anti- enolase antibodies.....	159
Figure 4. 12 Native gels of gel filtration – purified Blastocystis enolase.	160
Figure 4. 13 Second dimension SDS-PAGE representing a strip cut from blue native gel PAGE of expected recombinant Blastocystis enolase band.	160
Figure 4. 14 Western blot analysis of the predicted Blastocystis enolase from native PAGE. ...	161
Figure 4. 15 Size distribution (by intensity) of Blastocystis enolase. The sharp peak indicates the presence of a single species component with no evidence of aggregation.	163
Figure 4. 16 A Heterologous bimodal size distribution histogram (regularization graph) of the gel filtration- purified Blastocystis enolase. A total of two distinct populations were detected..	164
Figure 4. 17 Thermal shift stability assay of Blastocystis enolase.....	166
Figure 4. 18 The stability of Blastocystis enolase versus buffer composition, and salt addition.	167
Figure 4. 19 Schematic representation of recombinant Blastocystis enolase assay in which one molecule of NADPH ⁺ is oxidised into NAD ⁺	168
Figure 4. 20 Steady state kinetics of recombinant soluble Blastocystis enolase.....	170

Figure 4. 21 Rate of the recombinant <i>Blastocystis</i> enolase activity reaction and its dependence on protein concentration.	171
Figure 4. 22 Recombinant <i>Blastocystis</i> enolase activity using varied substrate concentration which varied from 1 mM to 100 mM.....	171
Figure 5. 1 Comparison of the amino acid sequences of triosephosphate isomerase part of the <i>Blastocystis</i> TPI-GAPDH fusion protein with various organisms.	179
Figure 5. 2 The active site region of <i>Blastocystis</i> TPI-GAPDH and other glycolytic GAPDH proteins from different species.	181
Figure 5. 3 Conserved region in <i>Blastocystis</i> TPI-GAPDH and other glycolytic GAPDH proteins from different species.....	182
Figure 5. 4 GAPDH protein sequence alignments using the <i>Blastocystis</i> TPI-GAPDH as query sequence on BLAST search.	185
Figure 5. 5 Fusion protein aligned with other fusions using <i>Blastocystis</i> TPI-GAPDH as a query.	186
Figure 5. 6 Relationship of <i>Blastocystis</i> TPI-GAPDH fusion protein among eukaryotes and prokaryotes as determined by phylogenetic analysis of protein sequence representing triosephosphate isomerase only.....	188
Figure 5. 7 Relationship of <i>Blastocystis</i> TPI-GAPDH fusion protein among a group of organisms containing GAPDH protein only determined by phylogenetic analysis of protein sequence.....	189
Figure 5. 8 Relationship of <i>Blastocystis</i> TPI-GAPDH fusion protein among a group of organisms containing the complete fusion protein only determined by phylogenetic analysis of protein sequence.	189
Figure 6. 1 Agarose gel electrophoresis analysis of PCR-amplified TPI-GAPDH from <i>Blastocystis</i> ST4.	196
Figure 6. 2 Agarose gel electrophoresis analysis of PCR-amplified <i>Blastocystis</i> TPI-GAPDH with pOPIN E and pOPIN F primers	197
Figure 6. 3 Expression trials for <i>Blastocystis</i> TPI-GAPDH in <i>E. coli</i> cell lines under several conditions.	199
Figure 6. 4 SDS gel of several expression trials of <i>Blastocystis</i> TPI-GAPDH pET-14b construct in <i>E. coli</i> BL21.....	200
Figure 6. 5 Elution profile of <i>Blastocystis</i> TPI-GAPDH on a His-trap column (GE Healthcare).	202
Figure 6. 6 A Coomassie-stained SDS-PAGE gel of recombinant <i>Blastocystis</i> TPI-APDH purified by nickel affinity chromatography.	202
Figure 6. 7 Ion exchange chromatography of recombinant <i>Blastocystis</i> TPI-GAPDH.	206
Figure 6. 8 Gel filtration chromatography of recombinant <i>Blastocystis</i> TPI-GAPDH.	207

Figure 6. 9 A) Final purification of <i>Blastocystis</i> TPI-GAPDH using gel filtration size exclusion chromatography. B) Elution profile of the purified <i>Blastocystis</i> TPI-GAPDH using the fractions collected from the highest peak to check if the gradient peak obtained in first purification	208
Figure 6. 10 Purified recombinant <i>Blastocystis</i> TPI-GAPDH.	209
Figure 6. 11 Gel filtration chromatography of <i>Blastocystis</i> recombinant TPI-GAPDH.	211
Figure 6. 12 Gel filtration chromatography of recombinant <i>Blastocystis</i> TPI-GAPDH.	211
Figure 6. 13 Western blot analysis detecting <i>Blastocystis</i> TPI-GAPDH protein	212
Figure 6. 14 Native gel electrophoresis of <i>Blastocystis</i> TPI-GAPDH following gel filtration chromatography..	213
Figure 6. 15 Second dimension SDS-PAGE following native gel electrophoresis.....	214
Figure 6. 16 Dynamic light scattering analysis of the <i>Blastocystis</i> TPI-GAPDH fusion protein.	216
Figure 6. 17 Sedimentation equilibrium analytical ultracentrifugation of <i>Blastocystis</i> TPI-GAPDH.....	220
Figure 6. 18 Purification of <i>Blastocystis</i> recombinant GAPDH and TPI.....	222
Figure 6. 19 Sedimentation equilibrium analytical ultracentrifugation of recombinant <i>Blastocystis</i> GAPDH.....	223
Figure 6. 20 Melt curve plot of TPI-GAPDH showing the thermal denaturation profiles of the <i>Blastocystis</i> TPI-GAPDH fusion protein purified in different buffering conditions.....	226
Figure 6. 21 Interaction between <i>Blastocystis</i> TPI-GAPDH and NAD ⁺ using thermal shift assay.	227
Figure 6. 22 Site-directed mutagenesis of the <i>Blastocystis</i> TPI-GAPDH fusion gene.	229
Figure 6. 23 Comparison of <i>Blastocystis</i> TPI-GAPDH gene expression levels.	230
Figure 7. 1 Phase diagram representing the phases that a protein goes through until it forms a crystal.....	236
Figure 7. 2 Conditions that satisfy Bragg's law. Figure adapted from (Asherie 2004).	240
Figure 7. 3 Ewald's sphere. A three dimensional illustration of Bragg's law. Adapted from (Dauter 1999).	241
Figure 7. 4 Crystallization process flow showing the steps to obtain the fusion protein structure from setting crystal trials to diffracting the obtained crystal in the synchrotron.....	258
Figure 7. 5 Examples of protein crystals obtained in several commercial buffer screens for both <i>Blastocystis</i> TPI-GAPDH and <i>Blastocystis</i> GAPDH proteins, with the droplet conditions they grew in.....	259
Figure 7. 6 The secondary structure of the protein <i>Blastocystis</i> TPI-GAPDH fusion protein showing that the protein consists of four identical chains.....	265
Figure 7. 7. Flat figure showing the amino acid sequence of the PDB entry file of TPI-GAPDH fusion protein adorned with secondary structure elements presented on top (helices with squiggles, β -strands with arrows and turns with TT letters).	266

Figure 7. 8. Schematic representation of <i>Blastocystis</i> TPI-GAPDH protein with active sites and main domains found in the structure as expected on Prosite.Expasy.....	267
Figure 7. 9 Cartoon representation of the catalytic site residues of the TPI subunit of the fusion <i>Blastocystis</i> TPI-GAPDH.....	269
Figure 7. 10 Cartoon representation of lid loop 6 in TPI subunit of several proteins.....	272
Figure 7. 11 A) Stick representation of the active site residues in a TPI dimer of the <i>Blastocystis</i> TPI-GAPDH structure. B) Superimposition of loop 6 in a TPI subunit of <i>Blastocystis</i> TPI-GAPDH with the model protein coleopteran TPI (PDB entry 2i9e).....	273
Figure 7. 12 Superimposition of active site loop 6 in both the studied protein <i>Blastocystis</i> TPI-GAPDH and TIM from rabbit muscle	273
Figure 7. 13 Secondary structure of GAPDH in the fusion protein <i>Blastocystis</i> TPI-GAPDH....	275
Figure 7. 14. Cartoon representation of <i>Blastocystis</i> TPI-GAPDH in cyan colour showing four NAD binding sites for the four GAPDH subunits in red and magenta.	276
Figure 7. 15. Superimposition of the study protein <i>Blastocystis</i> TPI-GAPDH (cyan) with the model used for building the structure (human GAPDH, 1u8f) (in lime green).....	277
Figure 7. 16. 1) Cartoon representation of the surface of the fusion protein <i>Blastocystis</i> TPI-GAPDH, superimposed with the coleopteran TPI protein (PDB entry 2i9e) with a surface quality of 1 . 2) Transparent representation of the fusion protein <i>Blastocystis</i> TPI-GAPDH.....	279
Figure 7. 17. 1) Cartoon representation of the recombinant fusion protein <i>Blastocystis</i> TPI-GADH. 2) Stereo view of superposition of the GAPDH tetramer on itself showing a displacement of TPI dimmers indication a less rigid TPI structural organization.	280
Figure 7. 18 total protein structure as interpreted by PISA, EMBL-EBI.....	282

List of tables

<i>Table 2. 1 List of bacterial and protist strains used in this study.</i>	59
<i>Table 2. 2 Primers used in this study. Restriction sites are italicized when applicable.</i>	68
<i>Table 2. 3 Plasmids used in this study.</i>	69
<i>Table 2. 4 Composition of SDS polyacrylamide gels.</i>	72
<i>Table 2. 5 Composition of native gels with different concentrations.</i>	74
<i>Table 2. 6 Buffers used in different protein purification steps.</i>	79
<i>Table 2. 7 Buffers used to study protein stability using Differential Scanning Fluorimetry (DSF).</i>	81
<i>Table 2. 8 PCR cycling condition using PfuTaq DNA polymerase and Gotaq master mix.</i>	84
<i>Table 2. 9 PCR cycling condition using PfuTaq DNA polymerase and Gotaq master mix.</i>	84
<i>Table 2. 10 PCR cycling parameters using Quick change lightning Site-directed mutagenesis kit.</i>	86
<i>Table 2. 11 Forward and reverse PCR primers used for Sanger sequencing of pGEM-T-Easy, pET-14b and pJET1.2 plasmids.</i>	89
<i>Table 2. 12 Fermentas Fast Digest reaction composition for both plasmid and PCR products.</i> .90	
<i>Table 2. 13 Proteins used as native molecular weight marker.</i>	101
<i>Table 3. 1 Possible mitochondrial targeting signal on enolase sequences from several organisms. Prediction performed using TargetP (http://ipsort.hgc.jp/), Mitoprot (Claros 1995), iPSort (Nielsen, Engelbrecht et al. 1996, Emanuelsson, Nielsen et al. 2000) and Predator (https://urgi.versailles.inra.fr/predotar/predotar.html).</i>	124
<i>Table 3. 2. Special features in Blastocystis enolase sequence as annotated by Uniprot server, http://www.uniprot.org/blast/. The table shows the important regions in the protein sequence for its activity and classification.</i>	127
<i>Table 3. 3 Taxa used in large-scale Blastocystis enolase protein phylogeny including accession numbers, and the super-groups.</i>	134
<i>Table 4. 1 Dynamic light scattering parameters obtained from 10 measurements.</i>	162
<i>Table 4. 2 Michaelis-Menten constant value for different organisms including Blastocystis sp4 (this study)</i>	170

<i>Table 6. 1 Dynamic light scattering data of Blastocystis TPI-GAPDH.....</i>	<i>216</i>
<i>Table 6. 2 Analytical ultra-centrifugation analyses of Blastocystis TPI-GAPDH.....</i>	<i>219</i>
<i>Table 7. 1 Pre-crystallization test (PCT) results and the recommended actions (https://hamptonresearch.com/documents/product/hr00559_9_2-140_142_user_guide.pdf)</i>	<i>247</i>
<i>Table 7. 2 Commercial screens used in this study.....</i>	<i>249</i>
<i>Table 7. 3 Cryo – protectants and conditions used for Blastocystis GAPDH crystals.....</i>	<i>252</i>
<i>Table 7. 4 Cryo – protectants and conditions used for Blastocystis TPI - GAPDH crystals.....</i>	<i>253</i>
<i>Table 7. 5 Cryo-cooling and X-ray data collection conditions for both Blastocystis TPI-GADH and Blastocystis GAPDH crystals.</i>	<i>254</i>
<i>Table 7. 6 Data processing statistics and solution of the recombinant protein Blastocystis TPI- GAPDH.....</i>	<i>261</i>
<i>Table 7. 7 The final X – ray data refinement statistics for Blastocystis TPI – GAPDH structure.</i>	<i>264</i>
<i>Table 7. 8 Summary of several triosephosphate isomerase flexible loop conformations in several organisms.</i>	<i>271</i>
<i>Table 7. 9 interfacing monomers of Blastocystis TPI-GAPDH structure, obtained from PISA service at EMBL-EBI.....</i>	<i>281</i>

Abbreviations

APS	Ammonium persulphate
AUC	Analytical ultracentrifuge
BAM	Benzamidine
BLAST	Basic Local Alignment Search Tool
bp	base pairs
CCP4	Collaborative Computational Project, number 4
ddH ₂ O	double distilled H ₂ O
DHAP	Dihydroxyactone phosphate
DSF	Differential Scanning Fluorimetry
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxyribose nucleotides
DPG	D-phosphoglycerate
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetraacetic acid (disodium salt)
Eno	Enolase
GAP	Glyceraldehyde-3-phosphate
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GF	Gel filtration
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HI	Heat inactivated
HIC	Hydrophobic interaction chromatography

His-tag	Poly histidine tag
hNSE	human neuron specific enolase
HS	Horse serum
IBs	Inclusion bodies
IBS	Irritable Bowel Syndrome
IPTG	Isopropyl β -D-galactopyranoside
IMAC	Immobilized Metal Affinity Column
kbp	Kilo base pair
kDa	Kilo Dalton
LB	Luria-Bertani
mTP	mitochondrial targeting peptide
MWCO	molecular weight cut off
NAC	Nickel affinity chromatography
NAD	Nicotine amide dinucleotide
NADH	Nicotine amide dinucleotide protonated
OPPF	Oxford Protein Production Facility
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PDB	Protein Databank
PEG	Poly Ethylene Glycol
PGK	Phospho Glycero Kinase
PK/LDH	Pyruvate Kinase / Lactate Dehydrogenase
PMSF	Phenylmethylsulfonyl fluoride
Q FF	Q Fast flow
RT	Room temperature
SDS	Sodium dodecyl sulphate

SE	Sedimentation equilibrium
Sp.	Species
SRS	Synchrotron Radiation Source
STY	Subtype
TBST	Tris-buffered saline and tween 20
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TE	Tris-HCl pH 8.0, 1 mM EDTA buffer
TEA	Triethanolamine
TEMED	N,N,N,N-tetramethylethylene diamide
TPI	Triosephosphate isomerase
Tris	Tris [hydroxymethyl] aminomethane

Aknowledgment

Firstly I would like to thank my project supervisor Dr. Mark van der Giezen for all the support, encouragement, freedom and independence he offered. This gave me the opportunity to explore my own ideas and make the project my own, for that I am indebted.

I would like to thank everybody who has contributed to this project in one form or another; however, I must notably thank Dr. Clive Butler my second supervisor, Oxford Protein Production Facility (OPPF) lab members and specially Anil Verma, for giving me the chance to use their equipments and help with their advices and follow up. And to Dr. David Scott for his help in the AUC experiment. I would also like to thank all CCCP4 group and Diamond in Harwell science campus, for the crystallography work performed there and for all what I have learnt with them.

A special thanks to Dr. Michail Isupov for all his support and guidance in crystallography work.

I am also grateful to Damascus University and the University of Exeter, who gave me the opportunity to work in the exciting scientific field with great and inspiring people.

I cannot forget our group member Dr. Maria Siegesmund who was such an inspirational person and helpful in the lab. A special thanks to Kim Payne who made my start such an enjoyable in spite of all the challenges, and for also being such a good friend who taught lots of the nice British culture.

A big thank for Dr. Kate McIntosh, for being such a nice person who always remembered me and my family in Christmas and birthdays times with her nice cards, making us feel loved in those nice days.

Thanks to Syria which without I wouldn't have been in this place. I hope will be able to pay back in the near future.

Thanks to my family, mum and dad, sister and brothers for making me enjoy times away from science, and for their encouragement to always do better.

The biggest thank to my angels Azzam and Hala for their love and support, patience and understanding. For accepting me in all my bad and good times that made it possible.

To the most important person in my life, MUM, I love you and I wish you were there now.....

Chapter 1. Introduction

1.1. *Blastocystis*

1.1.1. History and background

More than hundred years ago *Blastocystis* had been described in the literature, and drawings of it were published together with other parasites examined in a stool sample. It was first described in 1849 by London and Swayne who called it cholera bodies, and annular cells respectively, in almost every insect, bird, reptile and mammal studied. They first considered it as the pathogen responsible for cholera disease which was a serious health concern at the time, knowing that the first pandemic started in the 1817 and moved to hit Europe in 1826-1837. It later transpired that their work actually related to *Ascaris lumbricoides* ova, the giant roundworm. It is therefore more likely that *Blastocystis* was actually discovered in 1899, as it appeared in a written description by Perroncito, unfortunately, this paper did not include any drawings. This work was followed by a description of the same organism by Perroncito's assistant in 1901 ([Zierdt 1991](#)). In 1911, the first description including an appropriate nomenclature was introduced by Alexieff who called it *Blastocystis enterocola*. He however classified it as yeast but perhaps some shared morphological features led to this erroneous classification. In light microscopy, and similar to yeast, *Blastocystis* has a glistening appearance, due to its central vacuole, but lacks pseudopodia and is immotile ([Zierdt 1991](#)). *Blastocystis* has several morphological forms which has hampered the proper classification of this organism. It has moved repeatedly across the eukaryotic tree of life due to various different classifications. For a long time, it had been classified as a cyst of a yeast but it has also been considered as fungus,

a flagellate and a plant ([Zierdt 1991](#)). Finally it was classified by Zierdt as a protist in 1967. This classification was due to several features such as the presence of one or more nuclei, Golgi bodies, mitochondria-like organelles, and its failure to grow on fungal media and resistance to anti-fungal drugs. Moreover, it was found to be sensitive to anti-protozoal drugs such as metronidazole (Flagyl) and emetine ([Zierdt 1988](#), [Zierdt, Donnelly et al. 1988](#), [Zierdt 1991](#)). *Blastocystis* has finally found its home in the stramenopiles in 1996 by the work of Silberman and colleagues after electron microscopic studies failed to place *Blastocystis* in any particular group. Silberman *et al* sequenced the small sub-unit of the ribosomal RNA genes of two *Blastocystis* subtypes and a thorough phylogenetic analysis clearly placed *Blastocystis* in the Stramenopiles ([Silberman, Sogin et al. 1996](#)). Stramenopila are a large grouping of a seemingly heterogeneous assemblage of organisms that includes unicellular as well as cellular organisms. It contains heterotrophic and photosynthetic organisms. Examples include many algae (including the giant kelp), water moulds and diatoms. One of the best known members of the Stramenopiles is *Phytophthora infestans* which was the causative agent of the infamous Irish Potato Famine (also called the Great Hunger) which led to the death of 30% of the population and displacement and emigration of about 2,400,000 from their homes (1849-1854) in the middle of the 19th Century. *Blastocystis* is thought to cause blastocystosis, a disease manifested by abdominal and gastrointestinal symptoms. Although there are several pathogens in the stramenopiles, *Blastocystis* is generally considered the only human pathogen of this whole group ([Silberman, Sogin et al. 1996](#)).

Because it is often found in human stool samples, it was called *Blastocystis hominis*, because of this; additional *Blastocystis* isolates were named after their

host. For example, a *Blastocystis* isolated from rat became *Blastocystis rattis*. This 'explosion' of *Blastocystis* species, based on host, lacked any founding in evolutionary relationships and once molecular markers were used to unravel the relationships between all these isolates, it soon became clear the system was completely flawed. This led to an international effort to rectify the *Blastocystis* taxonomy and this system will from now on be used in this thesis ([Clark, Stensvold et al. 2007](#)).

Blastocystis is an obligatory anaerobic protist ([Zierdt 1991](#)) which inhabits the human intestines. In addition to humans, *Blastocystis* can be found in a range of different hosts such as pig, rat, and birds ([Zierdt 1991](#)). *Blastocystis* has a large geographic spread. The prevalence of this parasite varies from 5-10% in developed countries to up to 60% in developing countries. Higher prevalence is linked to animal handling or living in close proximity to animals such as at farms. There is also a reported increase in incidence in immuno-compromised patients ([Tan 2008](#)).

In addition to the long controversy regarding its classification, the role of *Blastocystis* in disease is still being debated. For over three quarters of a century it was believed that *Blastocystis* was a harmless commensal yeast which was confused with *Entamoeba histolytica*, the causative agent of amoebiasis ([Zierdt 1983](#)). This is mainly due to the fact that *Blastocystis* is found in both symptomatic as well as in asymptomatic cases. In addition, when present in symptomatic cases, *Blastocystis* is always accompanied by other organisms that might also be blamed for causing the symptoms. It was always considered as a gut commensal parasite, but it has been proven this parasite to be a true pathogen ([Hussein, Hussein et al. 2008](#), [Tan 2008](#)).

1.1.2. Nomenclature

As mentioned above, *Blastocystis* has gone through various different classifications. The additional naming of isolates based on the host species gave the impression of a tight host specificity ([Stenzel and Boreham 1996](#)). The description of different morphological phenotypes further contributed to the confusion and these phenotypes were thought to be related to the different hosts the organism was found in. Developments in sequencing techniques indicated that *Blastocystis* does not show any host specificity. In addition, many discrepancies were noticed in the literature regarding *Blastocystis* and some of these seem attributable to different protocols or equipment used to identify species types and variations ([Carbajal, DelCastillo et al. 1997](#)). There was an obvious need for standardization of *Blastocystis* research to avoid a continued confusion in the literature and an associated inability to assign hosts and or even the label 'pathogen' to this organism. Based on small subunit ribosomal RNA gene analyses of as many isolates as possible, a nomenclature for *Blastocystis* was proposed. A total of 17 genetically different *Blastocystis* subtypes were identified ([Clark, Stensvold et al. 2007](#)).

Nine of these subtypes were found to be associated with humans and might therefore potentially be real human pathogens ([Clark, van der Giezen et al. 2013](#)). No relationship was found between subtypes and host specificity ([Stensvold 2013](#)). Furthermore, it seems that *Blastocystis* subtypes can be transmitted between humans, between animals and also between animals and humans. So, after over a century of confusion, a consensus nomenclature was established for *Blastocystis*. It was decided to collapse the expanded naming for the various supposedly host-specific names and to call all isolates *Blastocystis* sp. together

with its subtype number which currently ranges from 1-17 ([Clark, Stensvold et al. 2007](#)).

1.1.3. Structure of morphological forms and proposed functions

Although *Blastocystis* sp. belongs to the stramenopiles, which is a large and heterogeneous group, it is nonetheless considered as an atypical member of this group. Despite the heterogeneous nature of the stramenopiles overall, there are common features that the members of this group share. For example, tripartite tubular hairs on the cell surface is considered a standard stramenopile feature. A long interior flagellum is also considered standard. However, *Blastocystis* sp. contains neither. They do however contain mitochondria with tubular cristae. *Blastocystis* generally has a spherical cell shape, though irregular shapes are reported as well. It can have a large central vacuole which leaves a very thin layer of cytoplasm squeezed near the outside of the cell. *Blastocystis* cell has a carbohydrate surface coat with a variety of carbohydrates ([Tan 2008](#)) and is covered by a bacterial layer in most of reported cases ([Dunn, Boreham et al. 1989](#), [Lanuza, Carbajal et al. 1996](#)). The presence of this layer of bacteria has been suggested to provide a nutritional role or to be important for survival ([Tan 2008](#)). As Tan study pointed out to the fact that *Blastocystis* surface coat thins out during laboratory cultivation, while it is at its thickest once freshly isolated from faeces. This was thought to be due to the coat role of capturing bacteria for nutritional purposes, which is not necessary during laboratory cultivation with nutritious media provided. Several nuclei have been reported (up to four), and those were always localized at one end of the cell along with other organelles ([Dunn, Boreham et al. 1989](#)). As mentioned, *Blastocystis* has several morphological forms which confused proper taxonomic classification but also

diagnosis in a clinical setting and often it has been misdiagnosed with other organisms such as *Entamoeba histolytica*. This might possibly explain the very high reported prevalence of *Blastocystis* ([Zierdt 1983](#)). *Blastocystis* diagnosis is traditionally based on direct examination using the light microscope of wet mount faeces culture and *in vitro* cultivation. More recently, polymerase chain reaction (PCR) is being used. Studies using scanning and transmission electron microscopy helped revealing the several forms of *Blastocystis*. The main four forms reported are the vacuolar, granular, amoeboid and cyst form ([Zierdt 1991](#), [Stenzel and Boreham 1996](#)), (see Figure 1.1.) No specific role has been assigned for each form, but several studies proposed a reproductive ([Zierdt 1988](#), [Dunn, Boreham et al. 1989](#), [Zierdt 1991](#)) and metabolic ([Dunn, Boreham et al. 1989](#), [Stenzel and Boreham 1996](#)) role of the vacuole in the vacuolar and the granular form. In the Dunn *et al.* study about ultrastructural variation in *Blastocystis* cell types, they showed the presence of cytoplasmic budding towards the central vacuole, and also showed budding from the vacuole towards the cytoplasm. They hypothesized that this might reflect a system of nutrient transport and expelling waste products or a means of storing. All of these would indicate a metabolic role of the vacuole but none have been proven yet. The same study also posited that the vacuole might also play a reproductive role similar to the schizogony proposed by ([Zierdt 1991](#)). However, only binary fission seen in electron and light microscopy was put forward as supporting this notion ([Dunn, Boreham et al. 1989](#)).

The central body, previously called the vacuole, occupies a large fraction of the whole cell volume, ranging from 50-90% of the total cell. As mentioned earlier, in the most extreme cases, this leaves a thin rim of cytoplasm and a large central

vacuole ([Dunn, Boreham et al. 1989](#)). Another form is the granular form which gets its granulated appearance due to the high concentration of mitochondria in the cytoplasm. In a study that tried to identify a possible link between the vacuolated and the granular form, it was found that the central vacuole arises from the coalescing of small granules which, over time, form the large central vacuole. This suggests that perhaps those morphological forms are sequential stages of cell degeneration where the central body presents the degenerative form of the parasite ([Vdovenko 2000](#)). The cystic form is another morphology that can be observed for *Blastocystis*. Zierdt considered that there was no cyst form for *Blastocystis* ([Zierdt 1991](#)). However, *Blastocystis* does have a form that is resistant to the environment. This is considered the infectious form that is transmitted through the faecal–oral route and causing infection ([Mehlhorn 1988](#)). Subsequent ultrastructural studies showed both a thick and thin cyst wall ([Dunn, Boreham et al. 1989](#)) and a large glycogen content which is assumed to play a role in energy storage during the encystation stage ([Zaman, Howe et al. 1995](#)). There are also multi–vacuolar and amoebic forms of *Blastocystis* which are proposed to develop from the vacuolar form through extrusion of the outer membrane ([Zierdt, Donnelly et al. 1988](#), [Clark, van der Giezen et al. 2013](#)). It was originally presumed that these different shapes were artefacts due to several environmental factors or due to crowding of cells in the culture ([Dunn, Boreham et al. 1989](#), [Tan, Howe et al. 2001](#)). However, the amoeboid form is also reported in symptomatic cases in diarrheal fluid. It is less often, or not at all, seen in *in vitro* cultures where it is thought to revert to the vacuolar form. The amoeboid form is often mistaken for leukocytes. However, a way to distinguish between leukocytes and *Blastocystis* is to use the Gram stain as the former will stain but the latter will

lyse ([Zierdt 1991](#)). A study by Dunn *et al* investigating the ultrastructural variation between *Blastocystis* stages reported that the amoeboid and vacuolar forms are phagocytic. They also showed that amoeboid buds contain lysosomes, lipid droplets and degraded bacteria. This might suggest that *Blastocystis* pseudopods have a role in bacterial phagocytosis ([Dunn, Boreham et al. 1989](#)). A recent microscopic study studied living *Blastocystis* in their natural microbial environment ([Nagel, Gray et al. 2015](#)).

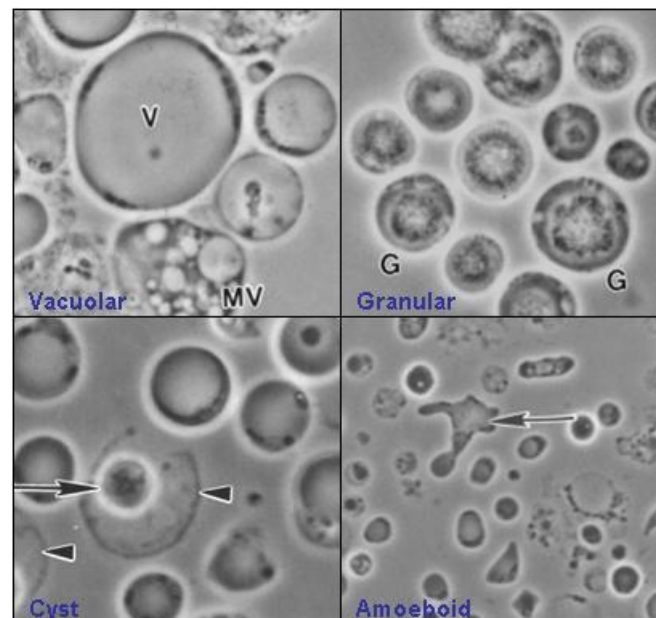


Figure 1.1 Morphological forms of *Blastocystis* sp. subtype 4 by phase-contrast microscopy. Vacuolar (V), multi-vacuolar (MV), granular (G) form with distinct granular inclusions within the central vacuole, Amoeboid forms occasionally seen in culture showing pseudo-pod-like cytoplasmic extensions, and Cyst form. Bar, 10 μ m. Credits to ([Tan 2008](#)) in this figure.

All previously discussed microscopic observations were based on dead/fixed cells. However, no differences were found between the forms apart from the observation that binary fission occurs in both vacuolar and granular forms ([Nagel, Gray et al. 2015](#)). The role of the granules in the granular and vacuolar form was

investigated and it was proposed to have metabolic or reproductive function ([Zierdt 1991](#)). The fluorescent DNA-specific dye (DAPI) only gave a strong signal for the nucleus and less so for the mitochondria but no staining was found in the granules or the vacuole, negating the proposed reproductive roles of the vacuole or the granules within it ([Stenzel and Boreham 1996](#)). The vacuolar form is most common in both culture and stool, while the granular form can only be seen in culture when the serum concentration is high. It is rarely seen in stool samples ([Dunn, Boreham et al. 1989](#)). The amoeboid form is most often found in faeces and in symptomatic patients with diarrhoea ([Stenzel and Boreham 1996](#), [Tan and Suresh 2006](#), [Casero, Mongi et al. 2015](#)). Few studies have suggested that the amoeboid form could be a result of microbial imbalance in the gut and actually not as a cause of it. Katsarou *et al* suggested that *Blastocystis* upsets gut homeostasis in their case study of a Caucasian male who reported with acute urticaria. The amoeboid form made 7% only of *Blastocystis* cells reported ([Katsarou-Katsari, Vassalos et al. 2008](#)). While in Tan et al a percentage of 2% amoeboid form of *Blastocystis* ST2 and ST4 to 13.7% of ST1 *Blastocystis* in the amoeboid form was noticed. A peak of 28% was seen in day 3-6 of cultivation of ST1, declining after that ([Tan and Suresh 2006](#)). On the other hand, those amoeboid forms were found in ST3 of symptomatic patients, or asymptomatic who developed symptoms later ([Vassalos, Spanakos et al. 2010](#)). The correlation between the presence of the amoeboid forms and symptomatic patients was negated by Souppart et al, who reported asymptomatic cases with amoeboid forms [Souppart, Sancier et al. \(2009\)](#). It has also been suggested that the central vacuole might play a role as a reservoir of cysteine proteases in *Blastocystis* ([Puthia, Lu et al. 2008](#)).

Several others less common morphological forms have also been reported which adds to the difficulties in diagnosis solely based on basic light microscopic methods. Some of these less common forms were multi-vacuolar, rod shaped, and other irregular forms rarely reported ([Zierdt 1991](#), [Tan, Howe et al. 2001](#)).

1.1.4. Epidemiology and prevalence

Due to its high prevalence *Blastocystis* has been considered by the World Health Organization (WHO) as one of the main causes for diarrhoea in both developed and developing countries. The highest prevalence was reported in developing countries due to poor hygienic conditions and the fact that it is a water-borne disease. In the developing countries the prevalence of *Blastocystis* was reported to fall between 0.4 -18% compared to *Dientamoeba fragilis* which was 0.4-6.3% ([Fletcher, Stark et al. 2012](#)). This clearly shows the high impact this parasite exerts on public health systems. We should also not forget the effects on domestic animals as well. Anecdotal evidence suggests that everyone has *Blastocystis* in their gut ([Zierdt 1983](#)). However, this is likely an exaggeration, yet it does reflect the very high occurrences of this parasite which was reported many decades ago. Casero *et al* study including 270 patients showed that 24% of them were *Blastocystis* positive. The population was picked from Cordoba city in Argentina, and the study revealed the high percentage of this parasite in adults in comparison to the number in children. Urticaria was more prevalent in symptomatic adults (56%) compared with to asymptomatic children (33%). Gastrointestinal symptoms tended to be higher in children (67%) than in adults ([Casero, Mongi et al. 2015](#)). Another study was carried out on Philippino children in a group of residential, street community where a poor sanitation and water quality was reported. In this study a significant number of children were positive

for *Blastocystis* reaching up to 40% of the study group ([Baldo, Belizario et al. 2004](#)).

1.1.5. Genetic diversity and specific correlation with the host

There has been a surprising amount of confusion and uncertainty linked to *Blastocystis*. There have been issues regarding its classification, its pathogenicity, and its morphology. As mentioned earlier, several studies have reported the diverse cell-types of *Blastocystis*. The literature until the late 1990s assumed these differences were due to adaptation to the metabolic status of the host cell. This morphological adaptability was apparently such, that the parasite was able to display these even in *in vitro* culture ([Dunn, Boreham et al. 1989](#)). Several studies using different *Blastocystis* stocks have found morphological differences between these cultures. Surprisingly, this variation was maintained even after laboratory culture. This would indicate that differences in metabolic status of the host cannot be the sole reason that explains the morphological flexibility. As discussed earlier, another source of confusion was the supposed host specificity of *Blastocystis*. This problem was clearly resolved with the development of sequencing approaches to distinguish between isolates. This also made the huge genetic diversity within *Blastocystis* apparent. A still outstanding question is whether *Blastocystis* is a primary pathogen and whether this ability is strain specific. A study which included only one patient reported a case of mixed inter and intra infection of *Blastocystis* STs in the same individual. They identified ST3, ST4 and ST2 respectively, and in each identified subtype several clones were also reported by PCR using STs specific primers. The mainly identified subtype was ST3 making 76% of the three subtypes found ([Meloni, Poirier et al. 2012](#)). This study showed that ST3 clones exhibited a large intra ST

variation due to (89-100%) variation in the identity of ST3, while ST4 on the other hand showed a higher degree of homogeneity ([Meloni, Poirier et al. 2012](#)). Due to the limited scope of this study (only based on one patient), these results can hardly be called conclusive and much larger cohort studies are required to shed light on the pathogenicity of *Blastocystis*.

A study by Yoshikawa *et al* performed on patients of a major hospital in north eastern Thailand, showed that the most prevalent form is ST3 followed by ST1 then ST6 and ST7 were less available and considered minor in this group of search, knowing that they are considered more avian subtype ([Jantermtor, Pinlaor et al. 2013](#)). Several studies have been done and most consider subtype 3 as the type most often associated with humans. For example, the Casero *et al* study on the population of Cordoba in Argentina found a prevalence of 71.6% for ST3, followed by 14.9% for ST1, 7.5% for ST6 and 5.9% for ST2 ([Casero, Mongi et al. 2015](#)).

1.1.6. Mode of transmission, people with high risk and clinical symptoms

As with many other parasitic infections, more incidents with *Blastocystis* are recorded within the vulnerable population, including immune-compromised patients, patients with cancer and children. There is an increased risk of being infected in low hygiene situations, when lacking hygienic habits and for people in contact with animals such as on farms ([Wawrzyniak, Poirier et al. 2013](#)).

For decades it has been reported that *Blastocystis* is clinically associated with symptoms related to the gastro-intestinal tract. Nausea, vomiting, abdominal

pain, constipation, diarrhoea, urticaria, flatulence and irritable bowel syndrome (IBS) have been linked to *Blastocystis* ([Kaya, Cetin et al. 2007](#), [Casero, Mongi et al. 2015](#)) . Although another case-controlled study performed in Denmark on an adult population from internet-based research institute, have ruled out the role of the suspect (*Blastocystis* and *Dientamoeba fragilis*) in IBS. This study has found that a greater proportion of control than patients carried the parasites ([Krogsgaard, Engsbro et al. 2015](#)). There have also been reports that *Blastocystis* is related with cases of acute and chronic gastroenteritis ([Kaya, Cetin et al. 2007](#)).

Most importantly, *Blastocystis* has often been considered as an important candidate as a causative agent for irritable bowel syndrome (IBS). IBS is a functional gastrointestinal disorder with changes in bowel habits, and abdominal pain. Several studies on IBS have reported infection with *Blastocystis* ST2. However, *Blastocystis* is not the only organism that has been put forward as possible causative agent for IBS ([Krogsgaard, Engsbro et al. 2015](#)). In order to be able to prove a role for *Blastocystis* in IBS, studies have to link the concomitant eradication of the parasite with the relief of IBS symptoms.

In addition to gastrointestinal symptoms, *Blastocystis* has also been linked to cutaneous symptoms ([Wawrzyniak, Poirier et al. 2013](#), [Casero, Mongi et al. 2015](#)). Urticaria has been reported in a case of Caucasian male reported with itchy wheals over his body and extremities. After the failure of the treatment with antihistamines, and having normal blood tests, the stool was examined and he was diagnosed with *Blastocystis* ST3 and specifically the amoeboid form was found in 7% of the total *Blastocystis* morphological manifestations ([Katsarou-Katsari, Vassalos et al. 2008](#)). The urticaria was limited after treating the patient with metronidazole for 10 days and discontinuing the antihistamines. It was

suggested that the amoeboid form adheres to the intestinal epithelia and causes an inflammatory response in the host which results in the urticaria ([Valsecchi, Leghissa et al. 2004](#)).

1.1.7. Insight into *Blastocystis* pathogenesis

Clinical studies have correlated *Blastocystis* with intestinal inflammation and with modulating the immune response. *Blastocystis* has also been considered as an opportunistic parasite as it was reported in AIDS cases, cancer and immunocompromised patients, where those patients were diagnosed with *Blastocystis* associated diarrhoea ([Tan, Ong et al. 2009](#)). However, the molecular mechanisms by which this parasite is supposed to cause inflammation are not clear and there is no clarity about possible virulence factors expressed of this organism. Although pathogenicity is not yet well elucidated in *Blastocystis*, several studies have reported a correlation between certain subtypes and symptoms, mainly *Blastocystis* ST3 and ST4 in Europe ([Clark, van der Giezen et al. 2013](#)), and ST1 in Asia and the Middle East. Hussein *et al* studied the effect of Egyptian *Blastocystis* isolates and their effect on rats where they observed a role in pathogenesis for subtype 1, while they did not find relation between symptoms and infection with ST2. Subtype 3 and 4 both manifested in symptomatic and asymptomatic cases in this study ([Hussein, Hussein et al. 2008](#)). It is therefore more likely that variations in clinical manifestation are due to genetic diversity and geographic distribution ([Clark, van der Giezen et al. 2013](#)).

Several studies have been conducted to investigate the pathology of this parasite. Most commonly studied were cysteine proteases which play various roles in cells, ranging from cell size regulation to pathogenicity. Other studies have shown that

proteases play important roles in protozoan virulence with a clear difference between pathogenic and non-pathogenic infection in, for example, *Entamoeba histolytica*. Here, it was found that some of the serine proteases that are linked with pathogenicity are more abundant in pathogenic strains than in the non-pathogenic Rahman strain of *E. histolytica* ([Davis, Schulze et al. 2007](#)). These proteases also played a role in inducing the pro-inflammatory cytokines of the host cells. *Blastocystis* cysteine protease have been found to cause degradation of human secretory Ig A ([Puthia, Vaithilingam et al. 2005](#)) and an ability of inducing the interleukin IL-8 response in colonic epithelial cells ([Puthia, Lu et al. 2008](#)). It was found that there is a correlation between cell size, generation time and cysteine protease levels in *Blastocystis* ([Mirza and Tan 2009](#)). In this study they compared *Blastocystis* in avian (subtype 7) and rodent (subtype 4) strains in terms of cell size, generation times and cysteine protease levels. An increased level of cysteine protease activity was associated with increased cell size ([Mirza and Tan 2009](#)). However, clinical observations indicated that the rodent *Blastocystis* strain was symptomatic while the avian strains were asymptomatic. It was suggested that this might be due to intra-variations between similar subtypes isolated, they suggested that more work would be required to investigate a possible qualitative difference between cysteine protease activity ([Mirza and Tan 2009](#)). *Blastocystis* infections have been linked with certain vulnerable groups. In one study, it was reported in cancer patients that 7.7% of the sample group (24 patients) were infected with *Blastocystis*. However, a total of 49 (19.8%) of HIV/AIDS patients were co-infected with *Blastocystis*. In this study, subtype 3 was the most prevalent subtype followed by subtype 4 and 1. None of the subtypes 5-7 were found ([Tan, Ong et al. 2009](#)). Nonetheless, a

conclusion regarding *Blastocystis* pathogenicity is still inconclusive and controversial. To clearly demonstrate that this is linked to certain subtypes is dependent on more diagnosis which should be based on both clear clinical diagnosis linked to molecular sub-typing ([Tan, Singh et al. 2002](#), [Clark, van der Giezen et al. 2013](#))

1.1.8. Linkage with irritable bowel syndrome (IBS)

In recent years, epidemiological studies have highlighted a high prevalence of *Blastocystis* cases in patients with IBS compared to healthy individuals or patients with gastrointestinal disorders. IBS is a functional gastrointestinal disorder with symptoms showing as abdominal pain and changes in bowel habits (constipation, diarrhoea or both). The prevalence of this disease (IBS) is 5 to 24% in industrial countries ([Poirier, Wawrzyniak et al. 2012](#)). Diagnosis of IBS is performed based on clinical criteria (symptom-based diagnosis) based on what is called the Rome III criteria; the recurrence of abdominal symptoms for at least three days per month for the last three months associated with at least two of the following: improvement with defecation, commencement associated with change of frequency of defecation, and commencement related with change of stool appearance ([Bellini, Gambaccini et al. 2015](#)). Several factors have been investigated in an attempt to understand the disease pathology. A bacterial imbalance was reported and thought to play a role in this disease due to the change of gas and metabolite production such as short chain fatty acids. These fatty acids play a role in gas production and the increase of this production might be the cause of bloating; the most predominant symptom of IBS. Intestinal pathogens such as *Giardia intestinalis* and *Entamoeba histolytica* were found to disrupt the intestinal mucosal integrity by using proteinases which were found to

induce the immune response in host cells ([Buret 2007](#)). Such proteases were also used as an explanation of *Blastocystis* pathogenicity in several studies as previously discussed in 1.1.7 paragraph of insights into *Blastocystis* pathogenesis.

Studies regarding potential causative mechanisms were focusing on modifications in the microbiota environment, postulating that the gastrointestinal movement and dysbiosis of the gastrointestinal microbiota might be the causative factor. Recently, it was noted that there is a co-occurrence of IBS cases with *Blastocystis* in faecal samples. This might suggest a role of the parasite in the pathophysiology of IBS. However, presence of these parasites will manifest in symptoms close to IBS symptoms and this might mislead an IBS diagnosis with a parasitic infection. Other studies had indeed disproved this role. A retrospective study on IBS patients in Denmark (Where the IBS prevalence is 16% compared to 11% globally) has revealed that more cases of infection with *Blastocystis* and *Dientamoeba fragilis* were reported in control cases of symptomatic patients compared to IBS patients (identified with Rome III criteria), negating the proposed role of parasites in causing the disease ([Krogsgaard, Engsbro et al. 2015](#)). However, this study has several limitations. Firstly, a lack of clinical evaluation, and thus possibility of miss-diagnosis. Secondly, only men and people of old age provided stool samples resulting in a possible gender or age bias. Finally, information whether patients were on metronidazole and/or antibiotic treatment was lacking. Usage of these medications would reduce chances of observing possibly causative microbes. Another study has found linkage between IBS symptoms and *Blastocystis* incidence to be significant in men only, with a prevalence of *Blastocystis* in IBS patients of 36.8% of the study population

([Nourrisson, Scanzi et al. 2014](#)). The same study also observed that *Blastocystis* affects the natural intestinal bacteria community. They found a significant decrease in *Bifidobacterium* sp. in patients infected with *Blastocystis*.

However, considering the wide range of genetic subtypes that *Blastocystis* has, in addition to its wide geographical distribution, it will be hard to blame *Blastocystis* for being the causing agent of IBS. As Clark et al concluded in their recent review about the developments in *Blastocystis* research ([Clark, van der Giezen et al. 2013](#)), more studies need to be performed that include molecular sub-typing of IBS samples in addition to the clinical diagnosis. These studies will have to be performed in different parts of the world as well in order to be able to rule out or condemn certain *Blastocystis* subtypes of being pathogens or not ([Poirier, Wawrzyniak et al. 2012](#)).

1.1.9. Treatment

Zierdt had reported the use of metronidazole in *Blastocystosis* infections and suggested that treatment failure was caused by the inadequacy of the dose used in these cases. He has also examined other drugs in terms of their efficacy and discussed the benefits of using the following in terms of importance from higher to less effective: emitine, metronidazole, furazolidone, trimethoprim-sulfamethoxazole ([Zierdt 1983](#)). Metronidazole is found to induce programmed cell death, while keeping the plasma membrane integrity. This will in turn keep the parasite's internal proteases in the cell and not released to the host. ([Puthia, Lu et al. 2008](#)).

In addition to its efficacy, metronidazole is the drug of choice due to its low cost and general availability. Yet, despite its efficacy, metronidazole also showed cases of resistance or failure of treatment. This might be due to several factors. One of the main factors that might play a role in drug resistance is the fact of the known high genetic diversity within *Blastocystis* sub types. Even within the same subtype genetic diversity can vary to up to 5% in most of the cases ([Clark, van der Giezen et al. 2013](#)).

Antimicrobial eradication has also showed to be effective as well in *Blastocystis* eradication. This might be due to the fact that the microbial flora is essential for *Blastocystis* survival. Others suggest that the responsiveness or the clearance of symptoms after treatment with either metronidazole or cotrimoxazole (trimethoprim-sulfamethoxazole) could be due to the clearance of other co-infectors rather than *Blastocystis* ([Coyle, Varughese et al. 2012](#)). In another case, paromomicine was successfully used for 10 days in urticaria case for a patient who proved positive for *Blastocystis*. Pruritis decreased a week after end of paromomicine therapy, two months later stool samples were negative of *Blastocystis*. Allergy or other causes were ruled out by blood tests ([Valsecchi, Leghissa et al. 2004](#)). It is obvious that more studies have to be performed to better understand the role of these drugs and to find a specific cure for blastocystosis.

1.1.10. Mitochondrial derived organelles

1.1.10.1. *Blastocystis mitochondrial organelle (MLO)*

To add to the paradox and enigma surrounding *Blastocystis*, this parasite is a strict anaerobic organism. However, it does contain several mitochondria in its cytoplasm, it is the predominant organelle compared to other organelles in this parasite ([Zierdt 1991](#)). *Blastocystis* has mitochondria with cristae and the organelles look similar to those in closely related organisms. However, these mitochondria have lost many of their classical features: *Blastocystis* has no cytochromes ([Zierdt 1986](#)), it lost the activities of typical mitochondrial enzymes such as the pyruvate dehydrogenase complex, the ketoglutarate dehydrogenase complex, isocitrate dehydrogenase, glutamate dehydrogenase, and cytochromes c oxidase ([Zierdt 1991](#)). According to Zierdt, it is unlikely that these 'crippled' mitochondria are non-functional organelles. In his 1991 paper he presumed that a non-functional organelle would not be maintained by an organism. In addition, they stain brightly with active dyes that indicate the presence of a physiologically active organelle. They take the typical rosette shape. Their numbers also seem to be regulated as there is a varied number of mitochondria present in the different cell forms, ranging from 2-4 in the vacuolated form and increasing in numbers in the rapidly growing forms during log growth phase, while they might reach hundreds in the granular form ([Zierdt 1991](#)).

The *Blastocystis* mitochondrial genome is highly reduced compared to other members of stramenopiles. This might be due to the transfer of mitochondrial genome to the nucleus preceded by a loss of unnecessary genes in the mitochondria ([Clark, Perez-Brocal et al. 2010](#)). It is clear that *Blastocystis* has an uncommon mitochondrial like organelle (MLO) that sits somewhere in

between classic mitochondria and hydrogenosomes. This MLO contains a classical a hydrogenosomal enzyme (hydrogenase), yet it is not demonstrated that it actually does produce molecular hydrogen ([Stechmann, Hamblin et al. 2008](#)).

1.2. Glycolysis

The central carbon metabolism contains glycolytic and pentose phosphate pathways. Glycolysis is the most common and conservative pathway in eukaryotes, widely conserved in prokaryotes as well. Glycolysis is the biochemical pathway by which glucose is converted to pyruvate with the generation of 2 mole of ATP/ mole of glucose, providing part of the energy utilized by most cells ([Stryer 1995](#)).

Glycolysis is a defined sequence of ten enzyme-catalysed reactions. The intermediates of these reactions are considered as entry points to glycolysis, and hence many mono-saccharides such as fructose can be converted to one of those intermediates.

Glycolysis is oxygen independent process, though the molecular oxygen might be used to dispose some intermediates, such as NADH, pyruvate and H^+ forming what is called aerobic glycolysis, while otherwise called anaerobic glycolysis. This process is widely abundant in both eukaryotes and prokaryotes, indicating how ancient it is. Usually this process takes place in the cytosol. The most common type is Embden-Myerehof-Parnas. This pathway is divided into two steps; the preparatory phase where energy is consumed, and hence called the investments phase, and the payoff phase where energy is produced.

The preparatory phase starts by the conversion of glucose into glucose-6-phosphate (G6P) by the help of hexokinase enzyme and the consumption of one ATP molecule. G6P will be rearranged into fructose-6-phosphate (F6P) by phosphohexose isomerase. This in turn will be phosphorylated by phosphofructokinase 1 (F1,6BP) into fructose 1,6 bisphosphate (F1,6BP) with the consumption of another energy molecule ATP. And the last step in the preparatory phase involves the breakdown of the (F1,6BP), a six carbon molecule, into two 3-carbon units i.e. glyceraldehydes-3-phosphate (GAP), and dihydroxyacetone phosphate (DHAP), which both can interconvert to each other by the help of triosephosphate isomerase (TPI). This step is the fourth step in glycolysis and last step in the preparatory phase which ended by consuming two ATP molecules in total.

The second half of glycolysis is composed of six steps. It starts by the dehydrogenation of GADP into 1,3 bisphosphoglycerate (1,3 BPG), using two inorganic phosphate molecules and producing two molecules of NADH and H⁺, performed by the enzyme glyceraldehydes-3-phosphate dehydrogenase (GAPDH).

1,3BPG is then dephosphorelated by phosphoglycerate kinase into 3-phosphoglycerate (3PG) and producing two molecules of ATP. 3PG is then isomerized into 2-phosphoglycerate (2PG) by phosphoglycerate mutase in a reversible reaction. 2PG is enolized by enolase enzyme into phosphoenolpyruvate (PEP), losing two molecules of water. And finally PEP is dephosphorelated by pyruvate kinase producing two molecules of ATP, a diagram representing the whole glycolytic process is shown in Figure 1.2.

I have sought in this project to study part of the glycolysis enzymes, namely the ones in the payoff phase i.e. enolase, GAPDH and TPI, for complying with the project objectives of unravelling unusual features in some of the glycolytic enzymes, especially in the payoff phase.

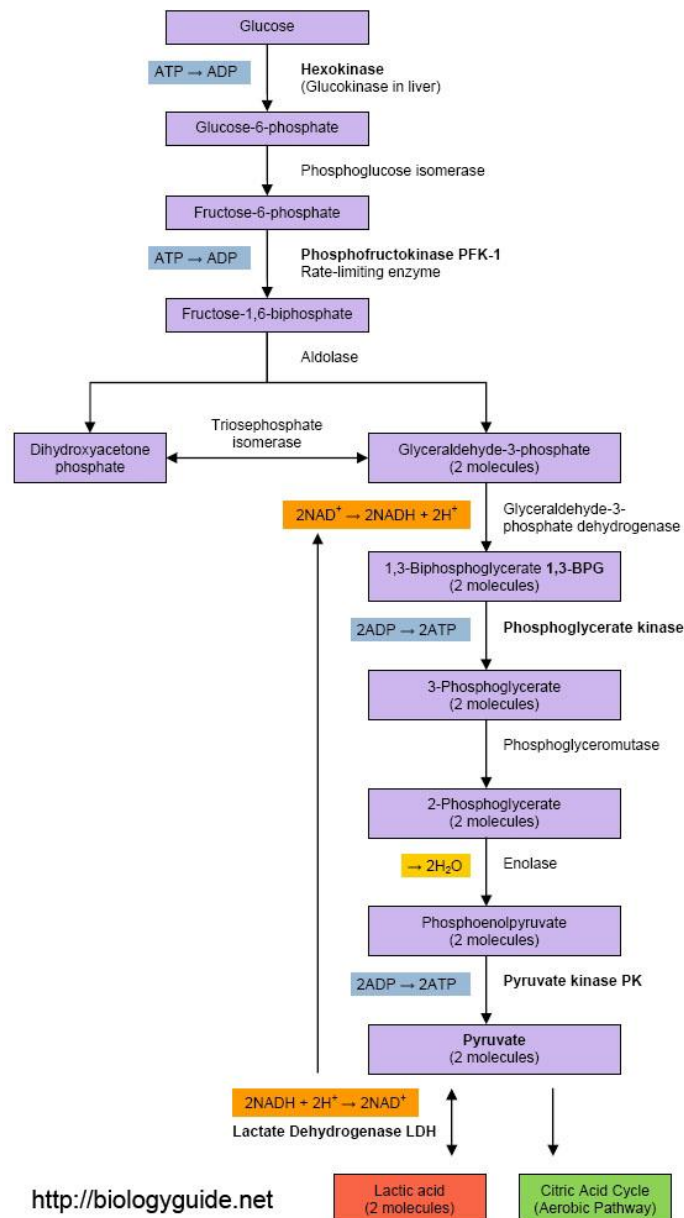


Figure 1. 2 Simple representation of the Embden-Meyerhof-Parasan, the complete glycolytic pathway. Credits of this picture go to <http://biologyguide.net>

1.2.1. Enolase (Phosphopyruvate hydratase)

Enolase (EC number 4.2.1.11), also called phosphopyruvate hydratase, is a glycolytic enzyme that participates in the step before the last in glycolysis. It is a member of the big enolase family, which consists of two main domains, enolase N (which is the N – terminal domain) and enolase C (which is the C terminal

domain). It is a metalloenzyme, that catalysis the conversion of 2-phosphoglycerate (2PG) to phosphoenolpyruvate (PEP). It has also been involved in other roles rather than glycolysis only, making it a very good example of what is known as a moonlighting protein. Enolase has traditionally been considered as an enzyme with no regulatory properties. However, it has been suggested in some cases that enolase expression varies according to the cell metabolic or developmental conditions.

1.2.1.1. Enolase structure

Enolase is a highly conserved protein, for this it is considered as a molecular marker as it highly conserved at the sequence level. It is found in eukaryotes and prokaryotes. In mammals, enolase has three subunits, i.e. α , β and γ , each encoded by a distinct gene, and each has specific localization in the body cells ([Joseph, Cruz-Sanchez et al. 1996](#)). Enolase is usually active as a dimeric protein. And those subunits in mammals, are found to form both homodimers as well as heterodimers formed of those three subunits as seen in the mouse muscle-specific β -enolase ([Merkulova, Lucas et al. 1997](#)). Those are tissue specific. In other organisms, enolase is usually found as a dimer ([Huther, Psarros et al. 1990](#)), except in some bacterial cases reported to be found as an octamer protein which was reported in hyper thermophilic bacterium *Thermotoga maritima* ([Schurig, Rutkat et al. 1995](#)).

Enolase is a metalloprotein, that it needs a metal molecule for its activity, (i.e. Mg^{2+} Zn^{2+} and Mn^{2+}). Those divalent metal ions can on the other hand have inhibitory effect if used over certain concentration ([Schreier and Hocker 2010](#)). Enolase activity requires two Mg^{2+} in general. The first divalent metal ion is

required for the correct conformation of the active site, while the other one will interact directly with the substrate PGA.

Enolase is a very highly conserved protein, and shares a large identity between species in general. The most conserved regions are the ones involved in metal binding, substrate binding, and catalysis activation sites.

The most important residues for enolase activity in general are His 172, Glu 224, Lys 358, His 386, Arg 387, Ser 388 (in our study) which are the essential five active sites catalytic residues.

Being a member of the enolase family, it shares enolase family motif. The enolase family motif takes the following sequence (SHRSGETED) usually located at the C-terminal of the protein sequence. It also possesses another conserved region called enolase signature which shares the following sequence (LLLKVNQIGSVTES).

Other conserved regions, but not found in all organisms, is the plasminogen binding site, which follows the (TYDENKKQY) sequence. It was found that some enolases have the property of plasminogen binding which helps as a virulence factor in some organisms as seen in *Clonorchis sinensis* enolase ([Wang, Chen et al. 2011](#)). This plasminogen binding capacity was found to be correlated to fibrinolytic activity which helps in host-tissue invasion ([Sun 2006](#)).

1.2.1.2. Enolase function and moon lighting

Enolase is a glycolytic protein that plays an important role in the payoff phase of glycolysis. It catalyzes the conversion of 2-phosphoglycerate into phosphoenolpyruvate. In addition to its role in energy conversion and glucose metabolism, enolase was found to play several other roles, changing the

traditional assumption about enolase being a house keeping enzyme, it is now a widely known multifunctional enzyme, making a very good example of a phenomenon called protein moonlighting. Those proteins were found to possess non-enzymatic activity in both prokaryotes and eukaryotes, an example of those proteins is aldolase which was found to have actin-binding ability, this in turn plays a role facilitating the parasite to glide across surfaces, and penetrate the host tissues by forming a kind of a bridge between certain protein and the skeleton, enabling the parasite for its gliding across the membrane ([Collingridge, Brown et al. 2010](#)). Enolase was found to be a very good example of a multifunctional protein. It plays a role as a plasminogen binding property as reported in mouse muscle-enolase and *Schistosoma japonicum* enolase ([Merkulova, Lucas et al. 1997](#), [Yang, Qiu et al. 2010](#)). In human brain enolase was found to play a role in tumour cell's growth. Being over-expressed in many tumour cells, suggests that this might enhance aerobic glycolysis of cancer cells, and thus enhancing the disease malignancy ([Liu 2007](#)).

Other un-defined and possible functions that enolase might do was suggested by ([Takishita, Patron et al. 2005](#)). They referred to the event of enolase forming a fusion protein with GAPDH. This was suggested to play a role different from glycolysis-gluconeogenesis, basically because another version of those proteins (enolase and GAPDH) was localized to the cytosol as a standalone version.

1.2.2. Glyceraldehyde-3-phosphatedehydrogenase (GAPDH)

The glycolytic protein glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12.) (GAPDH) is an enzyme in the glycolytic pathway that catalyses the oxidative phosphorylation of glyceraldehyde-3-phosphate (G3P) to form 1,3-bisphosphoglycerate. This reaction happens in two steps; the first one we find

that the active site cysteine residue attaches to G3P forming thiohemiacetal intermediate, which transfers a hydrid ion to NAD forming thioacyl enzyme. The second step includes the phosphorylation of the resulting thioester and forming the end product. This happens via a nucleophilic attack of an inorganic phosphate ion on the carbonyl atom of the thiacyl group, leading to the formation of 1,3 – bisphosphoglycerate.

1.2.2.1. GAPDH structure

GAPDH protein in general contains 340 amino acids in its polypeptide chain. It is usually found as homo tetramer, composed of four identical subunits, to form the active protein. It makes about 150 kDa, with each subunit making almost 37 kDa. The active protein has four active sites, binding four NAD⁺ in each active site of the four subunits ([Sirover 1999](#)). A common feature between GAPDHs structures is that it has two main regions; NAD⁺ binding domain and the catalytic domain. The catalytic domain binds the substrate glycerol-3-phosphate and is usually found at the N-terminal region of the protein. While the NAD⁺ binding domain (also contains Rossmann fold) is located at the C-terminal region. There are two critical residues in GAPDH structure for its glycolytic function; cysteine and histidine which are usually found at 149, and 176 respectively, in addition to its tetrameric structure ([Sirover 1999](#), [Colell A 2009](#)).

The NAD⁺ binding domain consists of beta sheets connected by alpha helices. This region, NAD⁺ binding region, (contains Rossmann fold, also called $\beta\alpha\beta$ fold) is not only important for its role in GAPDH activity, it also important for other dehydrogenases activity. Rossmann fold is probably playing an important role in

other functions of GAPDHs than the glycolytic role such as its RNA and glutathione binding ability which was located within the N-terminal region of GAPDHs ([Sirover 1999](#)).

The catalytic region is composed of antiparallel beta sheets and parallel helices. GAPDH active site has a conserved cysteine residue corresponding to Cys139 in *Sulfolobus solfataricus* ([Isupov, Fleming et al. 1999](#)), Cys141 in *Thermococcus kodakarensis* ([Jia, Linh le et al. 2011](#)), and Cys152 in humans has a conserved His179 residue and a conserved hydrogen bond with Cys 149

GAPDH is a very abundant protein in the cell making about 10-20% of the total cell content and a highly conserved protein in the tree of life ([Sirover 1999](#)). This protein occurs universally through Bacteria, Eukaryotes, and Archaea domains of life. And although it is highly conserved protein, the archaeal one shares less identity with the other life domains, 16-20%, and shows dual cofactor specificity for both NAD⁺ and NADP⁺ ([Littlechild and Isupov 2001](#)), while others have only NAD as a cofactor ([Isupov, Fleming et al. 1999](#)). A special sequence signature was found to predict the binding ability of GAPDH either to NAD⁺ or NADP⁺, which is the presence of the GXGXXG, replacing glycine by alanine would make the protein (GAPDH) bind NADP⁺ instead of NAD⁺ ([Hanukoglu 2015](#)).

1.2.2.2. Functional diversity of GAPDH

GAPDH has always been considered as a housekeeping gene used as internal marker in when doing DNA, RNA or protein work ([Zhang, Zhang et al. 2015](#)). In addition to GAPDH role in glycolysis pathway, several studies have proposed other function that GAPDH might play than being a solely glycolytic enzyme. It

was demonstrated that GAPDH is functioning in diverse other functions as transcription, nuclear tRNA export, neuronal apoptosis, DNA replication and repair, endocytosis and exocytosis ([Sirover 2005](#), [Guo, Liu et al. 2013](#)).

More evidence now has been found to support the role of glycolytic pathway in carcinogenic events, and one of its important components was GAPDH. GAPDH was found to be up-regulated in cancer cases as in lung, colorectal and renal cancer, which made it a potential marker for cancer progression and invasiveness ([Guo, Liu et al. 2013](#)). This could be explained by the need of energy supply by the cancerous cells which use glucose as their source of survival. So GAPDH role in cancer could either be explained by its role in glycolysis or by its role in cell cycle events ([Guo, Liu et al. 2013](#)). It was found that GAPDH is up-regulated in hypoxia, a condition of low oxygen levels, encountered by cancer cells. This up-regulation of GAPDH would offer an increase in the aerobic glycolysis, the main source of energy for cancerous cells, and hence providing a way for the tumour to proliferate ([Zhang, Zhang et al. 2015](#)). GAPDH is regulated by several cancer related factor such as insulin, nitric oxide (NO), hypoxia and p53, which why it is used as a tumour prognosis or progression marker ([Zhang, Zhang et al. 2015](#)). Other studies has proposed that GAPDH is a pro-apoptotic agent, other said it is down regulated in cancer helping in controlling the cell death, all of which makes the role of GAPDH in tumour cell fate still inconsistent issue ([Colell A 2009](#)).

Several theories have been put forward explaining GAPDH protein multi-functionality. One hypothesized that the protein multi-functionality is attributed to different localization in the cell ([Sirover 1999](#)). Other related its multi-functionality to the post-translational modifications this protein undergoes. Like in the case of

being phosphorylated by hypoxia in cancer, making it up-regulated in the cancer cases.

GAPDH was also found to play several other physiological roles in the cells from apoptosis to other roles in neurological disease as in cases like huntingtin disease (HD) or Alzheimer's disease, where they found that the protein's activity is reduced in those diseases. In HD the protein is found to be of higher molecular weight that is supposed to be due to interactions between GAPDH and huntingtin protein ([Mazzola and Sirover 2002](#)).

A study made by Kolln et al showed the decreased activity of this enzyme in the cerebrospinal fluid (CSF) where anti GAPDH IgGs were found in high concentration in the intrathecal of multiple sclerosis (MS) patients compared to other neurological diseases (OND) ([Kolln, Zhang et al. 2010](#)).

1.2.3. Triosephosphate isomerase (TPI)

Triosephosphate isomerase (EC 5.3.1.1), (TPI), also called D-glyceraldehyde-3-phosphate ketol-isomerase, catalyses the fifth reaction of the glycolytic pathway. It is an essential enzyme in the energy production of many pathogens. TPI catalysis the isomerisation between glyceraldehyde-3-phosphate (G3P) and dihydroxyacetone phosphate (DHAP). It is a ubiquitous enzyme, reported in eukaryotes, prokaryotes, and seen basically as a homodimer, although cases of tetramer were reported in Archaeal *Methanothermus fervidus* and *Pyrococcus woesei* TPI where it is suggested that this tetrameric structure is a nature's strategy to stabilize the native protein in extreme temperatures ([Kohlhoff, Dahm et al. 1996](#)).

1.2.3.1. *Triosephosphate isomerase structure*

TPI is usually a homodimer protein of two subunits, 27 kDa each subunit, and 248 amino acids each, TPI spatial structure belongs to the TIM barrel enzyme family. This TIM barrel fold is composed of eight β sheets alternate in sequence with eight α helices, taking the following pattern $(\beta\alpha)_8$ ([Orosz, Olah et al. 2009](#)). Those will form a parallel β sheets closed in a cylindrical topology, Figure 1.3, leaving a pocket where the ligand is located in the active form of the enzyme. This TIM-barrel is found in several other enzymes with different functions like enolase and other enzymes from functionally distinct protein families ([Tiwari and Reuter 2016](#)). The active site is almost always located at the carboxyl end of the barrel ([Tiwari and Reuter 2016](#)).

The active site loop in TPI, loop 6, is essential for the protein activity. This loop takes one of two conformations; either open loop when the protein is in unliganded (apo enzyme), and closed conformation when the protein is bound to its ligand (holo enzyme), see Figure 1.4. The change in conformational position of this loop gave it other names such as: the lid loop where it rests on border residues that function as hinges. catalytic residues , and also called the flexible loop, as it is taking a swung out conformation between the open and closed in about 8Å with a concomitant shift of the main catalytic residue by 2Å ([Pareek, Samanta et al. 2016](#)). Loop 6 is composed of a conserved amino acid sequence AYEPVWAIGTGKTA which contains the catalytic base E167 ([Jimenez, Vibanco-Perez et al. 2000](#)). Other active site residues are N10, K12, H94 and E97 are also highly conserved ([Pareek, Samanta et al. 2016](#)). And finally the highly conserved hexapeptide helical segment GHSERR, which contains a highly conserved serine residue.

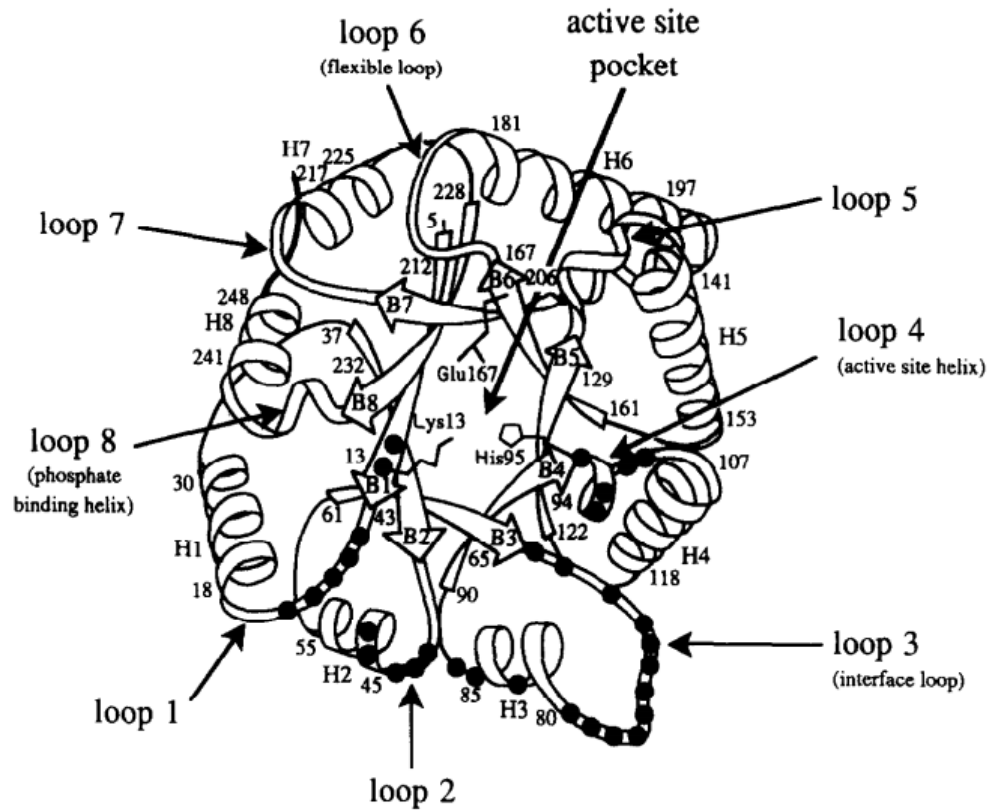


Figure 1. 3 Schematic representation of triosephosphate isomerase structure, showing the alpha helices and beta sheets taking the TIM barrel fold ($\beta\alpha$)₈. β sheets are referred to as β_x , and α helices are referred to as H_x , where x is number (1-8). Loops are also referred to in number between 1-8. A pocket formed in the middle of TIM barrel can also be seen in the representation. The 32 black dots indicate the location of the interface residues with the other TPI subunit, credits to [\(Wierenga, Noble et al. 1991\)](#).

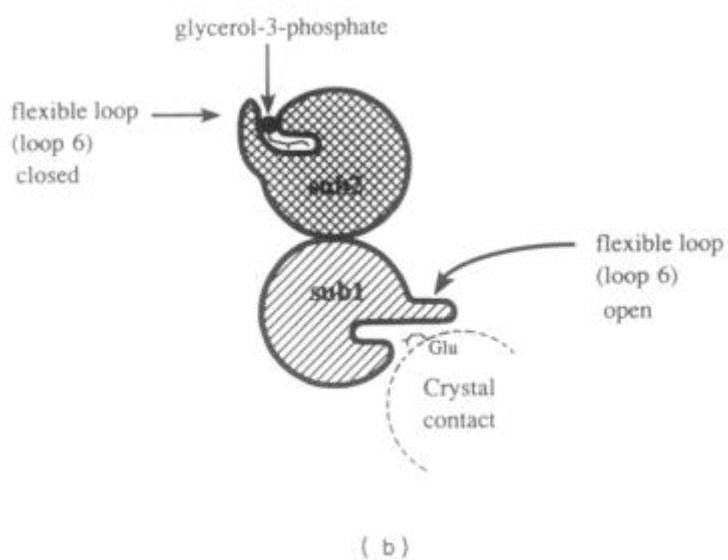
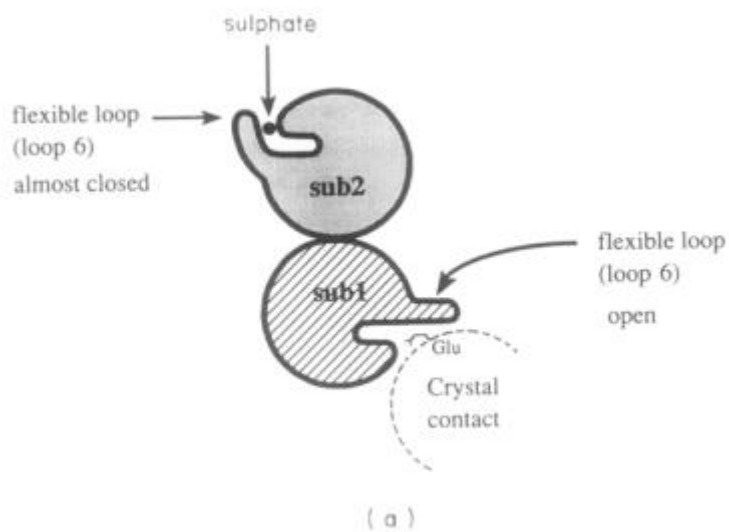


Figure 1. 4 Cartoon representation of the conformational changes of loop 6 in triosephosphate isomerase protein when changing between holo and apo enzyme of the protein (close and open conformation respectively). Adapted from [\(Wierenga, Noble et al. 1991\)](#).

1.2.3.2. *Triosephosphate isomerase physiological roles*

In addition to its role in glycolytic pathway, TPI was found to play several physiological roles in pathogens. Its main role as a glycolytic pathway, and the fact that several pathogens rely on glycolysis as their only source of energy, made TPI a highly targeted protein for drug delivery. Several vaccination trials have been conducted targeting TPI and other glycolytic enzymes of parasites causing tropical disease, such as *Trypanosoma brucei* ([Coley, Dodson et al. 2011](#)), the causative agent of sleeping sickness, others were to find a vaccine against schistosomiasis ([Wright, Davern et al. 1991](#)), knowing that glycolysis is their main source of energy.

In addition to finding cure of disease, TPI deficiency in humans was found to cause genetic disease of autosomal recessive disorder developmental retardation and manifested by neuromuscular dysfunction, cardiomyopathy and haemolytic anaemia, symptoms usually fatal ([Orosz, Olah et al. 2006](#), [Orosz, Olah et al. 2009](#)).

Recent studies have found that TPI can play other roles than glycolysis. Recent study has demonstrated that surface-exposed TPI of *Staphylococcus aureus* (for example) binds plasminogen and to a less but not negligible degree with fibronectin. This was suggested to play a role in host-pathogen interaction, as such protein can play a role in adherence and invasiveness ([Furuya and Ikeda 2011](#)). Both TPI and GAPDH were reported in *Lactobacillus plantarum* to be functioning as adhesion, and thus playing an important role in colonization ([Ramiah, van Reenen et al. 2008](#)).

A study performed by Bao, Chen *et al* has also shown a possible role for TPI in cyto-adherence of *Mycoplasma gallisepticum* to the host cell, where antibodies raised against this protein in glass inhibited the cell adherence. This adds to its possible role in host-pathogen interaction ([Bao, Chen et al. 2015](#)).

1.2.4. Moonlighting proteins pros and cons

Moonlighting proteins are those protein found to perform two or more function, operated by the same poly peptide chain. Gene products that perform more than one function could also be referred to as gene sharing, multitasking proteins, or moon-lightening. It is a reasonably newly characterised phenomenon, only discovered in the late 1980s, when lens crystallins were found to be the same as metabolic enzyme ([Wistow and Piatigorsky 1987](#)). Several factors can imply an enzyme to be multifunctional or not; cell type, cell compartment, ligand concentration or protein localization in the cell. More than 280 proteins have now been assigned to be moonlighting, and the number is increasing, but it is only dependant on total serendipity. It was first noticed that those multitasking proteins belong or have tendency to belong to the Intrinsically Disordered Proteins (IDP) class but this was negated afterwards as seen that there is no such tendency ([Hernández 2012](#)).

Several explanation were attempted to explain this phenomenon in the cells. One of which is that it is a way of economizing in the cell, as by this way the cell can perform so many functions in a small number of genes ([Hernandez, Ferragut et al. 2014](#)). That was found in proteins that perform several functions in addition to their traditional function as seen in previously mentioned cases of enolase,

GAPDH and TPI. Hernandez has also suggested that this phenomenon has kind of evolved through evolution where unused spaces became more used offering new binding sites. Moonlighting proteins seem to offer new way of virulence, by facilitating interaction with the host as reported in *Candida albicans*. GAPDH was found to be benefiting from the fact that it is in the host cytosol which helped in severity of invasion ([Karkowska-Kuleta and Kozik 2014](#)). In *Trichomonas vaginalis* infection, GAPDH was highly expressed after exposing to iron and is up-regulated during growth and multiplication, and it was also found on the surface of trichomonads. It is speculated that GAPDH plays a role in the virulence of trichomonas, though this work had shown that there might be other surface binding proteins that are to be found which help in the virulence as well ([Lama, Kucknoor et al. 2009](#), [Karkowska-Kuleta and Kozik 2014](#)). TPI was also found to play a role in pathogenesis of *Schistosoma mansoni* as it was located on the surface membrane of the pathogen larva making it as a target of vaccine development.

Moonlighting phenomenon is relatively new and might be exploited for developing drug targets as those proteins new roles might play a role in virulence, hence drugs and vaccines might be developed against the pathogen. On the negative side, moonlighting phenomenon, has shed doubts on phenotypes predicted for gene expressing those proteins. So moonlighting proteins which has been considered for long time as housekeeping gene should now be taken with scepticism.

1.2.5. Aim of the project

Blastocystis is a wide spread parasite in both developed and developing countries. Understanding this parasite would have an economic impact globally by minimizing number of cases seeking medical help, as will help developing drugs against this parasite. Little is known about this parasite's metabolism and bioinformatic analysis had shown special features in *Blastocystis* glycolysis pathway. Therefore, this thesis aims to understand the metabolic glycolysis pathway, and specifically three of its enzymes, i.e. enolase, triosephosphate isomerase and glyceraldehyde 3-phosphate dehydrogenase. Enolase was chosen as a model for the project since it is widely studied protein, while the fusion was chosen due to its specificity and the fact that it wasn't approached in details in literature before.

Specific thesis objectives include:

- 1) Biochemically characterize enolase, and the fusion protein TPI-GAPDH.
- 2) In silico analyses of those enzymes seeking any subtle differences from the host's version of those enzymes, which might help developing drug target.
- 3) Solve the structure of the fusion protein TPI-GAPDH, which will be the first report to interpret this structure after being mentioned before in work of other Stramenopils.

Chapter 2. Materials and methods

2.1. Materials

2.1.1. Microbiological work

2.1.1.1. Bacterial and Protist strains

Various bacterial and protist species and strains have been used in this study.

Table 2.1 lists all strains used and made during this study.

Table 2. 1 List of bacterial and protist strains used in this study.

Name of Strain	Source	Experiment
<i>Escherichia coli</i> strains		
α -select silver efficiency	Bioline	All cloning experiments
<i>BL21 pLysY</i>	Bioline	Enolase, TPI-GAPDH, PGK, TPI, and GAPDH, recombinant gene expression
<i>Rosetta 2 (DE3) pLysS</i>	Novagen	Enolase, TPI-GAPDH, recombinant gene expression
Arctic express	Agilent Technologies	Enolase, TPI-GAPDH, recombinant gene expression
XL10-goldSuper competent cell	Agilent	Mutant TPI-GAPDH recombinant gene expression
<i>Blastocystis ST</i> strains		
ST4, ST8	Dr. Graham Clark, London School of Hygiene and Tropical Medicine, UK	All experiments

2.1.1.2. Bacterial and protist media and solutions

Media and solutions used in this study for growth of bacteria and protists and chemicals utilized in all experiments are all listed in the following sections.

Luria-Bertani (LB) broth, 1,000 ml

- 10 g Bacto-tryptone
- 5 g Yeast extract
- 10 g NaCl

Adjust to 1 L final volume using dH₂O and sterilize by autoclaving for 20 minutes at 120°C.

LB agar, 1000 ml

As for LB broth with addition of 10 g agar.

SOC medium a super optimal broth (SOB) supplied with added glucose/sucrose for catabolite repression, or what is known as the glucose effect)

- 20 g Tryptone (Melford Ltd)
- 5 g Yeast extract
- 0.5 g NaCl
- 0.186 g KCl
- 3.6 g Glucose
- 0.95 g MgCl₂

Adjust to 1L final volume using dH₂O and sterilize by autoclaving for 20 minutes at 120°C.

ZYM-5052, To make 400 ml, add the following:

- 1% N-Z-amine (tryptone or other casein digest)
- 0.5% yeast extract

Add demineralised water to 380 ml and autoclave. To the autoclaved mixture the following sterile stocks are added:

- 8 ml 50 X M solution
- 8 ml 50 X 5052 solution
- 800 μ l 1 M MgSO_4 (400 μ l 2M MgSO_4)
- 80 μ l 1000 X metals

Antibiotics added according to the cells and plasmids used.

50 X M Solution

- 25 mM Na_2HPO_4
- 25 mM KH_2PO_4
- 50 mM NH_4Cl
- 5 mM Na_2SO

50 X 5052 Solution

0.5% glycerol = 54 mM

0.05% glucose = 2.8 mM

0.2% α -lactose = 5.6 mM

1000 X metals, in 100 ml

- 36 ml sterile H₂O
- 50 ml 0.1 M FeCl₂ in ~ 0.12 M HCl
- 2 ml 1 M CaCl₂
- 1 ml 1 M MnCl₂·4H₂O
- 1 ml 1 M ZnSO₄·7H₂O
- 1 ml 0.2 M CoCl₂·6H₂O
- 2 ml 0.1 M CuCl₂·2H₂O
- 1 ml 0.2 M NiCl₂·6H₂O
- 2 ml 0.1 M Na₂MoO₄·2H₂O
- 2 ml 0.1 M Na₂SeO₃·5H₂O
- 2 ml 0.1 M H₃BO₃
- 1 ml 0.2 M CoCl₂·6H₂O
- 2 ml 0.1 M CuCl₂·2H₂O

NZY + Broth (per Litre)

- 10 g of NZ amine (casein hydrolysate)
- 5 g of yeast extract
- 5 g of NaCl

Add deionised water to a final volume of 1 litre, adjust to pH 7.5 using NaOH, Autoclave. Before use, add the following filter sterilised supplements:

12.5 ml of 1 M MgCl₂

12.5 ml of 1 M MgSO₄

20 ml of 20% (w/v) glucose (or 10 ml of 2 M glucose)

12.5 ml of 1 M MgCl₂

Ampicillin (100 mg/ml)

Ampicillin powder (Sigma-Aldrich) was dissolved in deionised water, filter sterilised through a 0.2 µm filter, and stored at - 20°C. The final working concentration is 100 µg/ml.

Tetracycline hydrochloride (5 mg/ml)

Tetracycline hydrochloride salt was dissolved in ethanol and used in a final working concentration of 50 µg/ml.

X-Gal (20 mg/ml)

X-Gal was dissolved in dimethylformamide and used at a final concentration of 80 µg/ml and stored at - 20°C.

Isopropyl-beta-D-thiogalactopyranoside (IPTG), 1 M stock solution

IPTG was dissolved in deionised water, filter sterilised through a 0.2 µm filter and stored at - 20°C.

2.1.1.3. Protist media and solutions

Blastocystis was the only protist used in this project, and this parasite could be grown xenically and axenically. Most often *Blastocystis* was grown xenically in LYSGM ([Clark and Diamond 2002](#)). Several attempts to grow *Blastocystis* axenically were also performed by several labs and here in this project using several different media ([Tan, Singh et al. 1996](#), [Tan, Singh et al. 1996](#)). The media used to grow *Blastocystis* are listed below.

Xenic culture media

LYSGM Base

1.4 g dibasic potassium phosphate (19.7 mM)

0.2 g monobasic potassium phosphate (3.3 mM)

3.75 g NaCl (128.3 mM)

0.25 g liver extract

1.25 g yeast extract

Water up to 475 ml

Divide media into 95 ml aliquots and add 0.1 g gastric mucin per 100 ml bottle. Autoclave at 121°C for 15 min. Add heat inactivated adult bovine serum before use in a final concentration 5%.

Axenic culture solution and media

Iscove's Modified Dulbecco's Media IMDM

This media was used for axenic cultivation trials of *Blastocystis* and purchased ready made from PAA (GE Healthcare). IMDM consists of several metals, trace metals and trace elements supplemented with 10 % heat inactivated horse serum. For a detailed description of the IMDM media composition please refer to Appendix 1.

Diphasic media Axenic cultivation method was first described by Zierdt and Williams ([Zierdt and Williams 1974](#)). In 2002 Clark and Diamond reviewed the modified method developed by Lanuza *et al.* ([Lanuza, Carbajal et al. 1997](#)) ([Clark and Diamond 2002](#)).

Liquid phase

- 5 g glucose
- 16 g sodium chloride
- 0.4 g calcium chloride
- 0.4 g potassium chloride
- 0.02 g magnesium chloride
- 4 g sodium phosphate dibasic
- 0.8 g sodium bicarbonate
- 0.6 g potassium phosphate monobasic

The liquid phase ingredients were mixed and dissolved in a litre of distilled water.

A. Solid phase

- 12.5 ml Luck's solution
- 45 ml whole egg

The solid phase components were mixed and homogenized in a blender, filtered through gauze into a flask and put under vacuum to remove bubbles. The mixture is subsequently dispensed in 5 ml portions each in culture glass tubes. This was autoclaved for 10 min at 100°C with a slant position. The mixture is then cooled down and overlaid with the liquid phase, and autoclaved at 121°C for 15 min. After autoclaving, caps were tightened and the slant tubes were refrigerated up to 6 months.

Adult bovine serum

Purchased from PAA, and heat inactivated at 55°C for 30 minutes.

Horse-serum heat inactivated

Purchased from PAA.

Soft agar plates, consists of a mixture of IMDM, horse serum (HS), and Bacto™ agar in the following proportions:

- 2 X IMDM filtered and supplemented with 10% HS
- 0.72% Bacto™ agar dissolved in distilled water and autoclaved at 121°C for 15 min

Mix equal volume of both components and pour in sterile Petri dishes.

Soft agar plates with sodium thioglycollate

- 2 X IMDM filtered and supplemented with 10% HS.
- 0.05% sodium thioglycollate
- 0.72% Bacto™ agar

Dissolve the suspension of Bacto™ agar and thioglycollate in distilled water then autoclave at 121°C for 15 min. Mix equal volumes of the suspension and 2 X IMDM in a Petri dish.

Solid agar plates

- 2 X IMDM filtered and supplemented with 10 % HS
- 2% Bacto™ agar (autoclave at 121°C for 15 minutes)

Plates were made by mixing equal volumes of molten (40°C) 2% Bacto™ agar and 2 X IMDM-HS (warmed at 40°C). Leave the plates to solidify at room temperature, pre-reduce in an anaerobic jar for 24 hours before use.

10 X ampicillin /streptomycin for axenization experiment

40 mg/ml streptomycin - 10 mg/ml ampicillin

2.1.2. Molecular biology

2.1.2.1. Primers list

Many primers were used in this study to amplify the genes of interest. Some primers were designed to amplify genes from genomic DNA or cDNA and several new sets of primers containing restriction sites were designed to incorporate the gene of interest into an expression vector, mainly pET-14b in this study. Other primers were designed to introduce point mutations in genes, using the QuickChange Lightning Site directed mutagenesis kit (Agilent Technologies). Sequencing primers were designed for the purpose of checking that mutations were introduced at the correct site. The table below lists all the primers used in this project and the restriction sites are highlighted in italics. See Table 2.2

Table 2. 2 Primers used in this study. Restriction sites are italicized when applicable.

Primer Sets	Sequence 5' - 3'	Amplification product
Whole gene amplification primers		
BhEno F	ATGCTTTCTCGTCTTTCTAC	Enolase
BhEno R	TTAACAACGATCCGGAGTAC	
Recombinant gene expression primers		
BhEno pET F	AGAAGACATAATGGCCTCTACGATTACTG	Enolase with <i>NdeI</i> , <i>BamHI</i> restriction sites
BhEno pET R	CCGTTAAC TCTTCTGGATCCTTAACAACGATCCGGA GTACG	
TPI-GAPDH pET F	AGAAGACATAATGTTTCGTCGGTGGCAATT	TPI-GAPDH with <i>NdeI</i> restriction site
TPI-GAPDH pET R	GGAAGTGCAA TCTTCTGGATCCTTAAGAGCGATCCACC TTCGCCA	TPI-GAPDH with <i>BamHI</i> restriction site
Sequencing primers		
FTPGA1401	CTACAACATCATCCCGTCCTCCA	Reading along the whole mutant constructs to check that no unwanted mutations have been introduced.
FTPGA901	TTCATCACGCCGACTACAT	
RTPGA400	TCAGACCGTTCTTCTGAGCGTAAG	
TPI-GAPDHmut check F	GACCCGAAGAAGATCCAGTG	Internal primers used to check that the mutation was inserted and the rest of the gene is still correct.
TPI/GAPDHmut check R	GGACGGGATGATGTTGTAGC	
Mutation insertion primers		
RMM073	GTCGAACGCGTCCGCCACGACGAACTG	Mutant TPI-GAPDH in GAPDH C416A 5' and 3'
RMM074	CAGTTCGTCGTGGCGGACGCGTTTCGAC	

2.1.2.2 Plasmids list

Constructs made in this project were produced in several vectors divided between expression and cloning vectors. For general cloning purposes pGEM-T-Easy (Promega) was used. Some trials were also performed using pJET 1.2 (Fermentas), but work has not been continued with this vector because of the satisfactory results achieved with pGEM-T-Easy. For gene expression purposes pET-14b was the vectors used throughout the study. Table 2.3 lists the plasmids used, the constructs, their sizes and the suppliers of these plasmids.

Table 2. 3 Plasmids used in this study.

Plasmid	Information	Size (kb)	Source
pJET 1.2	Blunt end cloning vector	2.97	Fermentas
pGEM®-T_Easy	T overhangs for easy cloning of PCR products	3.015	Promega
pET-14b	Expression vector adds His-tag to N or C-terminus of protein to be expressed.	4.7	Novagen
Eno/pGEM®-T_Easy	BhEno PCR product cloned into pGEM®-T Easy.	4.365	Made in this study
Eno/pET-14b	BhEno PCR product cloned into pET-14b	6.05	Made in this study
TPI-GAPDH/pJET1.2	TPI-GAPDH PCR product cloned into pJET 1.2	4.78	Made in this study
TPI-GAPDH/pET-14b	TPI-GAPDH PCR product cloned into pET-14b	6.5	Made in this study
mTPI-GAPDH/pET-14b	Mutant TPI-GAPDH PCR product cloned into pET-14b	6.5	Made in this study

2.1.2.3. Molecular biology Solutions

2.1.2.3.1. DNA work solutions

50 X Tris-acetic acid-EDTA (TAE) electrophoresis buffer, 1000 ml

- 18.6 g Ethylenedcciaminetetraacetic acid (disodium salt) (EDTA)
- 57.1 ml of glacial acetic acid
- 242 g Tris-base

Dissolve in distilled water to a final volume of 1,000 ml (pH 8.0)

DNA isolation solutions

CTAB lysis buffer, 100 ml

- 3.72 g EDTA
- 0.25 g SDS

Instructions as above, pH to 8.0

10% CTAB in 0.7 M NaCl, 100 ml

10 g CTAB

Up to 100 ml with 0.7 M NaCl

Proteinase K, 2 mg/ml

Dissolve Proteinase K powder in distilled water then filter sterilise through a 2 µm filter. The final concentration used is 0.1 mg/ml.

DNA marker

Bioline Hyperladder I was used unless otherwise stated (see Fig 2.1)

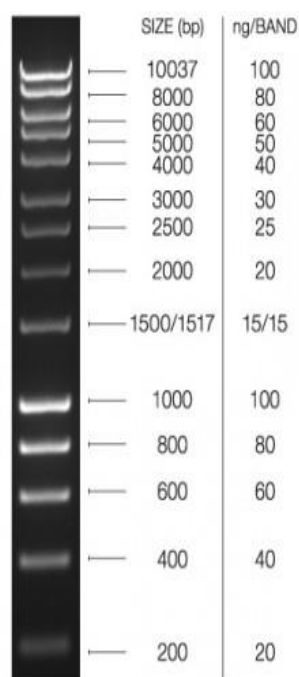


Figure 2. 1 Bioline Hyperladder I DNA marker. The left lane of the above image shows DNA bands stained with ethidium bromide. The intensity of the bands corresponds to the amount of DNA as indicated by the fluorescence, corresponding to 5 μ l overall load. Sizes are indicated in bp and corresponding amounts of DNA as ng/band. Image was taken from the Bioline website.

2.1.2.3.2. Protein work solutions

2.1.2.3.2.1. Sodium dodecyl sulphate (SDS) gel and solutions

SDS gels used in this project were 12 and 15% gels. 15% SDS gels were used for PGK, TPI and GAPDH protein experiments, while 12% gels were used for recombinant *Blastocystis* enolase, *Blastocystis* TPI-GAPDH and *Blastocystis* mTPI-GAPDH recombinant proteins experiments. Table 2.4 details the composition of the gels.

Table 2. 4 Composition of SDS polyacrylamide gels.

	6% stacking gel	12% resolving gel	15% resolving gel
H₂O	8.7 ml	8.6 ml	7.1 ml
40% acrylamide	2.25 ml	6 ml	7.5 ml
1.5 M Tris-HCl, pH 8.8	-	5 ml	5 ml
0.5 M Tris-HCl, pH 6.8	3.75 ml	-	-
10% SDS	150 µl	200 µl	-
10% APS	150 µl	200 µl	200 µl
TEMED	15 µl	20 µl	20 µl

SDS loading buffer, 100 ml

- 2.42 g Tris base
- 6.17 g DTT
- 8 g SDS
- 0.4 g Bromophenol blue
- 40 ml Glycerol

pH to 6.8, H₂O up to 100 ml.

SDS running buffer

- 30.29 g Tris base
- 150 g Glycine
- 100 g SDS

H₂O up to 1000 ml

Coomassie Blue stain

- 0.025% Coomassie Brilliant Blue
- 40% Methanol
- 14% Acetic acid

H₂O up to 100 ml

Destaining solution I

- 40% Methanol
- 7% Acetic acid

H₂O up to 100 ml

Destaining solution II

- 5% Methanol
- 7% Acetic acid

H₂O up to 100 ml

Protein marker

Fermentas Spectra Multicolor Broad Range Protein Ladder was used (See Fig. 2.2).

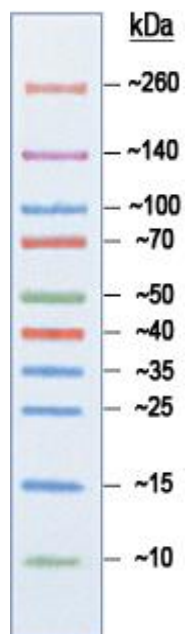


Figure 2. 2 Spectra Multicolor Broad Range Protein Ladder Fermentas. The sizes of the shown protein bands are given in kDa. Image was taken from the Fermentas website.

2.1.2.3.2.2. Native gel solutions

Native gels were used to determine the native molecular weight of target proteins of this study and thereby estimate their oligomeric status. Table 2.5 details the composition of the native gels used in this study.

Table 2. 5 Composition of native gels with different concentrations.

	5% stacking gel (ml)	10% resolving gel (ml)	12% resolving gel (ml)	15% resolving gel (ml)
H₂O	11.1	14.7	13.2	10.95
40% acrylamide	1.875	7.5	9	11.25
1.5 M Tris-HCl, pH 8.8	-	7.5	7.5	7.5
1 M Tris-HCl, pH 6.8	1.875	-	-	-
10% APS	0.15	0.3	0.3	0.3
TEMED	0.015	0.012	0.012	0.012

Separating gel buffer 1.5 M Tris HCl, pH = 8.8.

Stacking gel buffer 1 M Tris HCl, pH = 6.8

Sample buffer (5x)

- 15.5 ml of 1 M Tris-HCl pH 6.8
- 1% solution Bromophenol blue 2.5 ml
- H₂O 7.0 ml
- Glycerol 25 ml

Electrophoresis buffer

- 3.0 g Tris base
- 14.4 g Glycine

Dissolve in water to a final volume of 1 litre (PH 8.3).

In the case of blue native gels the solutions used are as follows:

Cathode buffer

- Bis-tris 15 mM
- Tricine 50 mM
- Coomassie blue G250 0.02%

Prepare 1 litre as a 10 x stock, adjust pH to 7.0 with HCl, and store at 4°C. Dilute before use with deionised water.

Anode buffer

- Bis-tris 50 mM

Prepare 1 litre as a 10 x stock, adjust pH to 7.0 with HCl, and store at 4°C. Dilute before use with deionised water.

Protein stain and de-stain solutions

A standard Coomassie stain / de-stain was used to stain native gels.

2.1.2.3.2.3. Western blotting solutions

Towbin transfer buffer, 1000 ml

- 3.028 g Tris base
- 15.14 g Glycine
- 100 ml methanol

Dissolve all components in H₂O up to a final volume of 1,000 ml.

5 x TBST, 1000 ml

- 12.112 g Tris base
- 12.86 g NaCl
- 50 ml Tween-20

Dissolve all components in H₂O up to a final volume of 1,000 ml, adjust pH to 7.6.

1 x TBST + 5% milk, 500 ml

- 100 ml 5 X TBST
- 25 g fat-free dried milk powder

Dissolve the components in H₂O up to 500 ml final volume.

Ponceau S, 100 ml

- 1.5 g Ponceau S
- 5 ml acetic acid

Dissolve the components in H₂O up to 100 ml final volume.

2.1.2.3.2.4. Colony hybridization solutions (colony blot)

Solution A	10% SDS
Solution B	0.5 N NaOH, 1.5 M NaCl
Solution C	0.5 M Tris-Cl, 1.5 M NaCl pH 7.4
Solution D	2 X SSC
20 X SSC	173 g NaCl (3 M) 88.2 g Sodium citrate (0.3 M)

Adjust pH to 7 and the final volume to 1 L with distilled water, sterilise by autoclaving.

2.1.2.3.2.5. Enzyme activity assay solutions

2.1.2.3.2.5.1. Blastocystis enolase activity assay solution

A. Triethanolamine buffer, TEA

18.565 mg/ml Triethanolamine Hydrochloride in H₂O, adjust to pH 7.4 with 1 M NaOH at 25°C

B. 2-Phosphoglycerate solution (DPG)

14.112 mg/ml D (+) 2-phosphoglyceric acid in water.

C. β-Nicotinamide adenine dinucleotide, reduced form solution (β-NADH)

4.9658 mg/ml β-nicotinamide adenine dinucleotide dissolved in water, prepare fresh.

D- Magnesium sulphate with potassium chloride solution (MgSO₄/KCl)

300.92 mg/5 ml magnesium sulphate

745.5 mg/5 ml potassium chloride

Dissolved in water to a final volume of 5 ml.

E- Adenosine 5'-diphosphate solute on (ADP)

8.544 mg/ml adenosine 5'-diphosphate, sodium Salt, in water, prepare fresh.

F- PK/LDH mixed enzymes

- 900-1400 units/ml lactic dehydrogenase
- 600-1000 units/ml pyruvate kinase

G- Tris-HCl with bovine serum albumin (enzyme diluent)

- 1.817 g/ml Trizma base
- 0.020 mg/ml albumin, bovine

Adjust to pH 7.4 at 25°C with 1 M HCl.

H- Enolase enzyme solution

Immediately before use, prepare a solution containing 0.25-0.5 unit/ml of enolase in cold Reagent G.

2.1.2.3.2.6. Protein purification solutions

Proteins studied in this project were purified throughout the work using different techniques. All purifications started by using nickel affinity columns for His-tagged proteins. For some proteins, ion exchange chromatography columns, which depend on the isoelectric point (PI) of the protein, were used (data not shown). A

size exclusion chromatography (SEC) column was used as a final step of protein purification. This technique depends basically on the size of the molecule and the molecular weight of the protein. The table below, Table 2.6 lists all the buffers used in the work.

Table 2. 6 Buffers used in different protein purification steps.

Buffer	Component
SP XL buffer A	0.05 M HEPES, Na ⁺ , pH 7.0–8.0
SP XL buffer B	0.05 M HEPES, Na ⁺ 1 M NaCl, pH 7.0–8.0
Q FFbuffer A	0.1 M Tris-HCl, pH 8.2
Q FF buffer B	0.1 M Tris-HCl 1 M NaCl ,pH 8.2
Nickel affinity chromatography (NAC) buffer A	0.1 M Tris-HCl 0.1 M NaCl cOmplete Protease Inhibitor Cocktail Tablets (Roche), pH 8.2
Nickel affinity chromatography (NAC) buffer B	0.1 M Tris-Hcl 0.1 M NaCl 0.5 M Imidazole cOmplete Protease Inhibitor Cocktail Tablets (Roche), pH 8.2
Gel filtration chromatography (GF) buffer	0.02 M Tris-Hcl 0.1 M NaCl 0.2 5 mM EDTA, pH 8.2

Recharging Nickel beads solution:

Stripping buffer

100 mM EDTA

Charging buffer

- 100 mM NiCl₂
- 500 mM NaCl / 20 mM Tris, pH 8.0

2.1.2.3.2.7. Crystallography work.

After protein purification, and for the purpose of growing crystals, the protein is incubated in a multitude of conditions varying in buffer composition, pH, precipitant composition, salt concentration, and plate incubation temperature. We can screen hundreds of conditions for crystals formation using commercially available screens supplied by different companies. In this project we used several screen including PEG-ION and Index (Hampton Research Company), Emerald Wizard III+IV (Emerald), and PACT premier, JCSG, Morpheus (Molecular Dimensions).

2.1.2.3.2.8. Differential scanning fluorimetry (DSF) buffers.

A set of screening buffers were used in this experiment, with different pH degrees, and different salt concentration ([Niesen, Berglund et al. 2007](#)). Table 2.7 lists all buffers used during differential scanning fluorimetry.

Table 2. 7 Buffers used to study protein stability using Differential Scanning Fluorimetry (DSF).

number	Buffer name	Buffer abbreviation
1	400 mM Sodium acetate-AcOH, pH 5, 600 mM NaCl	AcOH
2	400 mM Bicin, pH 8.3, 600 mM NaCl	Bicin
3	400 mM Bis-Tris-HCl, pH 6, 600mM NaCl	Bis-tris
4	400 mM MES-NaOH, pH 6, 600 mM NaCl	MES
5	400 mM Imidazole-HCl, pH 7, 600 mM NaCl	Imi
6	400 mM PIPES-NaOH, pH 7, 600 mM NaCl	PIPS
7	400mM MOPS-NaOH, pH7.2, 600 mM NaCl	MOPS
9	400 mM Tricine-NaOH, pH 8, 600 mM NaCl	Tricine
10	400 mM Tris-HCl, pH 8.1, 600 mM NaCl	Tris-Hcl
12	400 mM Glycine-NaOH, pH 9, 600 mM NaCl	Gly
14	400 mM CHES-NaOH, pH 9.5, 600 mM NaCl	CHES
15	400 mM Ethanolamine-HCl, pH 9.5, 600 mM NaCl	Etha
16	400 mM CAPS-NaOH, pH 10, 600 mM NaCl	CAPS
17	400 mM HEPES-NaOH, pH 7.5	HEPES
18	400 mM HEPES-NaOH, pH 7.5, 200 mM NaCl	HEPES1
19	400 mM HEPES-NaOH, pH 7.5, 600 mM NaCl	HEPES2
21	400 mM HEPES-NaOH, pH 7.5, 2 M NaCl	HEPES3
22	400 mM HEPES-NaOH, pH 7.5, 600 mM NaCl, 20 % glycerol	HEPES, gly1
23	400 mM HEPES-NaOH, pH 7.5, 600 mM NaCl, 40 % glycerol	HEPES, gly2
24	400 mM HEPES-NaOH, pH 7.5, 2 M NaCl, 20 % glycerol	HEPES, gly3

2.2 Methods

2.2.1. Molecular biology methods

2.2.1.1 DNA extraction

DNA was extracted from *Blastocystis* ST4 and ST8 using the cetyl-trimethyl-Ammonium Bromide (CTAB) method. The cells were harvested as described in 2.2.2.4, and the pelleted cells were re-suspended in 250 μ l CTAB lysis buffer and proteinase K (0.1 mg/ml). Suspended cells were incubated for 1 h at room temperature. After 1 hour incubation, NaCl was added to a final concentration of 0.7 M followed by addition of 10% CTAB to 1%. The mixture was mixed well and incubated at 65°C for 15 min.

Equal volume of chloroform: isoamyl alcohol (24:1) was mixed and centrifuged at 14,000 X g for 10 min. The resulting phase was then extracted with equal volume of phenol: chloroform (1:1) and centrifuged as before, the supernatant was precipitated with 2.5 volumes ethanol, incubated for 10 min and centrifuged as before. The pellet was washed once in 70% ethanol and centrifuged for an additional 5 minutes at 14,000 X g. After removing the liquid, the pellet was left to air dry before being rehydrated in small amount of sterile water (50 μ l H₂O).

2.2.1.2. PCR reactions

A Techne TC-312 PCR machine was used for all standard PCR reactions. All primers were ordered from Eurofins MWG Operon, and primers were re-suspended in distilled water to a final concentration of 100 pmol/ μ l and stored at -20°C.

2.2.1.2.1. PCR cycling conditions

PCR reactions were executed using two sets of polymerases in this study; in amplification reactions where high fidelity was not needed, GoTaq Green Master Mix was utilized, as in colony PCR experiments. While *Pfu* DNA polymerase was used when high accuracy was needed. *Pfu* DNA polymerase was also used when the addition of adenine overhang was needed for TA cloning. In this instance, a mixture of two different polymerases was used: GoTaq polymerase (for A tailing) and *Pfu Taq* polymerase 5'-3' (for proof-reading ability) (both Promega). Table 2.8 sets a comparison between the two reactions' components.

Table 2. 8 PCR cycling condition using PfuTaq DNA polymerase and Gotaq master mix.

<i>PCR reaction</i>	<i>Pfu Taq/GoTaq reaction</i>	<i>GoTaq Green Master Mix</i>
<i>Reaction components</i>		
Buffer	5 µl GoTaq buffer	12.5 µl Green Master Mix
MgCl ₂	2.5 µl MgCl ₂	-
dNTPs	0.5 µl 10 mM dNTPs	-
forward primer	1 µl 20 pM forward primer	0.5 µl 20 pM forward primer
reverse primer	1 µl 20 pM reverse primer	0.5 µl 20 pM reverse primer
DNA template	1 µl template DNA	1 µl template DNA
<i>pfu taq</i> polymerase	0.3 µl <i>Pfu taq</i> polymerase	-
GoTaq polymerase	0.2 µl GoTaq polymerase	-
Purite water	13.5 µl Purite water	10.5 µl Purite water

Table 2. 9 PCR cycling condition using PfuTaq DNA polymerase and Gotaq master mix.

Cycle step	Number of cycles	Temperature	Time
PCR cycling condition for <i>PfuTaq/</i> regular <i>Taq</i>			
Initial denaturation	1	95 °C	2 minutes
Denaturation	33 cycles	95 °C	45 sec
Annealing		48-65 °C*	45 sec
Extension		72 °C	X minutes**
Final extension	1	72 °C	5 minutes
PCR cycling condition for GoTaq green mastermix			
Initial denaturation	1	95 °C	5 minutes
Denaturation	30 cycles	95 °C	45 sec
Annealing		48-65 °C*	45 sec
Extension		72 °C	X minutes**
Final extension	1	72 °C	5 minutes

* Annealing temperature differs according to the primers melting temperature

** Extension temperature varies according to the amplified amplicon size and it is 1min/kb for GoTaq and 2min/kb for *Pfu*.

2.2.1.2.2 Mutagenesis PCR conditions

QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent technologies) was used for introducing mutations in a desired site. Manufacturer's instructions were followed. The basic reaction uses the whole double stranded plasmid with the insert of interest, two sets of primers with the desired mutation and a derivative of the high fidelity DNA polymerase (*PfuUltra* high-fidelity (HF) DNA polymerase). Extension of the designed oligonucleotides resulted in the formation of the mutated plasmid. The resulted mutant plasmid was then treated with a newly optimized *DpnI* enzyme to remove template plasmid.

The reaction composition was as follow:

5 μ l of 10 X reaction buffer

1 μ l DNA (up to 100 ng)

1 μ l *PfuUltra* high-fidelity (HF) DNA polymerase

X μ l (100 ng each primer for 1–3 primers; 50 ng each primer for 4–5 primers)

The reaction mix was topped up with double distilled water to a final volume of 25 μ l.

Details about the conditions used for setting the PCR reaction for mutagenesis reactions for site directed mutagenesis can be found in Table 2.10.

Table 2. 10 PCR cycling parameters using Quick change lightning Site-directed mutagenesis kit.

Cycle step	Number of cycles	Temperature	Time
PCR cycling condition			
Initial denaturation	1	95°C	2 minutes
Denaturation	18 cycle	95°C	20 sec
Annealing		60°C	10 sec
Extension		68°C	30 sec/kb of plasmid length
Final extension	1	68°C	5 minutes

2.2.1.3. DNA electrophoresis

Agarose gels were used to analyse DNA fragments and check for PCR products according to the DNA band sizes. Agarose gels were made in this project using 1% agarose dissolved in TAE buffer by boiling. Ethidium bromide was used to visualize DNA bands as a nucleic acid stain with a concentration of 0.04 µl/ml of gel. DNA fragments were mixed with the 5 X running buffer, and loaded on the gel, with 5 µl of DNA HyperLadder™ 1 (Bioline) and electrophoreses at 100 Volts for different times according to the fragment being analysed, allowing the estimation of DNA bands of sizes ranging from 200 bp and 10 kb, (See Fig2.1.).

2.2.1.4. Isolation and purification of DNA fragments

DNA fragments were excised from agarose gels and purified using The QIAquick Gel Extraction (Qiagen), following manufacturer's instructions. Otherwise, a QIAquick PCR Purification (Qiagen) was used to clean PCR products when no other products are observed. All steps were performed at room temperature (R.T.). These two kits remove primers, primer dimers, dNTPs and restriction enzymes to aid downstream processes.

2.2.1.5. Cloning of PCR products.

2.2.1.5.1. Ligation into cloning vector.

Primary PCR products were cloned in the pGEM®-T-Easy vector, which is a linear vector that contains thymidine tri-phosphate overhangs allowing the hybridization of the PCR product and the vector, see Fig2.3. Vector pJET1.2 was also used in this study. This vector has blunt ends allowing the cloning process without any further processing for PCR products apart from purification.

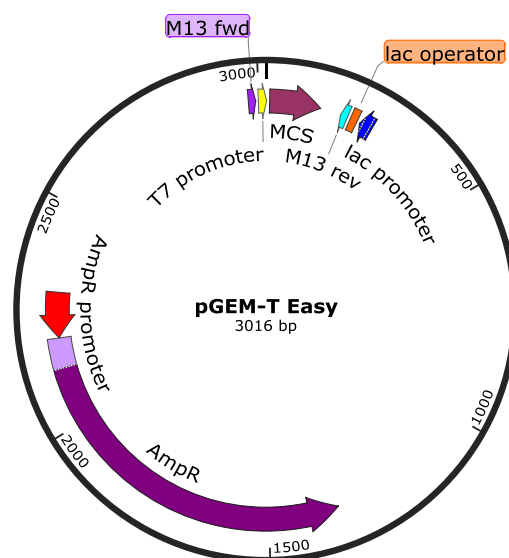


Figure 2. 3The general cloning vector pGEM-T-Easy showing the multiple cloning site (MCS) that comes after the T7 promoter site, antibiotic resistance site, and the sequencing primers: T7, SP6, M13 forward and reverse. Figure drawn using SnapGene.

2.2.1.5.2. Transformation of competent cells with plasmid DNA

The transformation procedure of any construct into *E. coli* competent cells was carried out by aliquoting 50 µl of the desired competent cells. Those cells were then kept on ice. 1-8 µl (1-50 ng of DNA) of the plasmid were added to the Eppendorf tube containing the competent cells, and mixed gently by flicking the tube. This mixture was placed on ice for 20 minutes, and then moved to a preheated water bath at 42°C for 45 seconds. After this step of heat-shock to the cells, the tube was re-placed on ice for another 2 minutes. A 1 ml of LB medium or SOC medium were added before incubating the mixture in a water bath at 37°C for 1 hour. Finally 100 µl and 50 µl of the mixture were plated on LB agar plates containing the required antibiotic, with IPTG, and XGAL for cells that needs blue-white screening, and those plates were incubated at 37°C overnight. The transformation is then checked by colony PCR and/or digestion of the plasmid after plasmid isolation from the growing colonies.

2.2.1.5.3. Plasmid DNA isolation

10 ml overnight culture of the transformed cells was grown at 37°C incubator shaking at 220 rpm. 5 ml of this culture was used for plasmid extraction using a QIAprep Spin Miniprep Kit (Qiagen). The manufacturer's instructions were followed, and the whole procedure was performed at RT. This procedure depends on alkaline lysis of the cells; the lysed cells were neutralized and centrifuged to remove proteins and cells debris. The final step includes loading the DNA on silica membrane to remove impurities and elution with elution buffer. DNA was stored until use at - 20°C.

2.2.1.5.4. Sanger sequencing

To check that the plasmid construct has the desired insert, constructs were sent to Eurofins MWG for sequencing. Inserts in pET-14b vector were sequenced using T7 forward and reverse primers, while those ligated in pGEM-T-Easy were sequenced using M13 forward and reverse primers. Constructs in pJET1.2 used proprietary primers from Fermentas (see Table 2.1)

Table 2. 11 Forward and reverse PCR primers used for Sanger sequencing of pGEM-T-Easy, pET-14b and pJET1.2 plasmids.

Primer	Sequence
T7 Forward	5'-TAATACGACTCACTATAGGG-3'
T7 Reverse	5'-GCTAGTTATTGCTCAGCGG-3'
M13 Forward	5'-GTAAAACGACGGCCAGT-3'
M13 reverse	5'-AACAGCTATGACCATG -3'
pJET1.2for	CGA CTC ACT ATA GGG AGA GCG GC
pJET1.2rev	AAG AAC ATC GAT TTT CCA TGG CAG

2.2.1.5.5. Restriction enzyme digestion of DNA

Fermentas Fast Digest Enzymes were used in this study. Fermentas' protocol was followed. Table 2.12 shows the digestion reaction contents. After brief centrifugation of the digestion mix, it was placed on heat block for 10 minutes at 37°C.

Table 2. 12 Fermentas Fast Digest reaction composition for both plasmid and PCR products.

	Plasmid DNA	PCR product
Nuclease free water	15 μ l	17 μ l
10xFastDigest Green Buffer	2 μ l	2 μ l
DNA	1 μ l	10 μ l
FastDigest Enzyme	1 μ l	1 μ l
Total volume	20 μ l	30 μ l

2.2.1.5.6. Cloning of DNA fragment into expression vector

The expression vector used in this study was pET-14b (Novagen). Ligation of the PCR product into this vector necessitates digesting of both vector and insert using the same restriction enzymes. *Nde*I and *Bam*HI restriction sites were added to the 5' and 3' ends of the genes in this study. The pET-14b vector has a poly-His tag that will flank the gene on the N-terminus region after being digested with *Bam*HI and *Nde*I. The gene was inserted in the *Bam*HI and *Nde*I position of the pET-14b plasmid in frame with the N-terminal His-tag. This tag helps in further purification steps of the protein. The digestion products were cleaned using a QIAquick Gel Extraction (Qiagen) or QIAquick PCR Purification (Qiagen). These products were subsequently ligated using a Takara Mighty DNA Ligation kit, using the 3:1 ratio of DNA: plasmid, and incubated at 25°C for 30 minutes. The ligation mix was transformed to *E. coli* expression cells as described in section 2.2.1.5.2.

Gene expression was driven by a T7 promoter and ampicillin resistance allowed for selection of positive clones, which was further confirmed by colony PCR, digestion and sequencing.

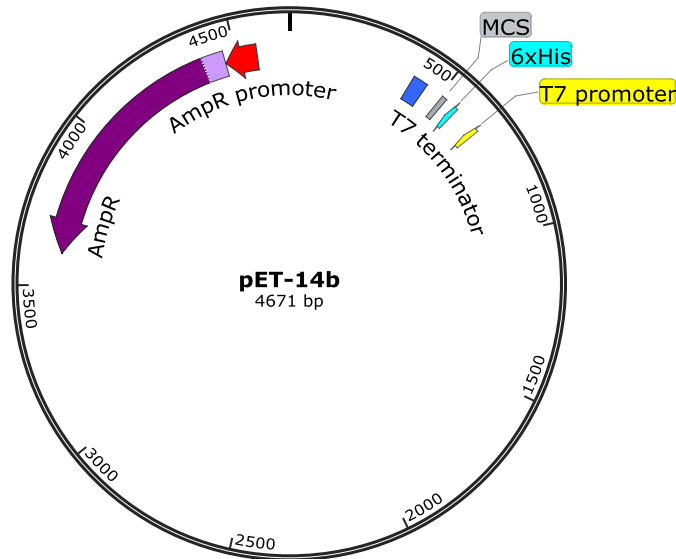


Figure 2. 4The gene expression plasmid pET-14b featuring a backbone size of 4.6 kb, the multiple cloning sites with *NdeI* and *BamHI* restriction sites used in this study, antibiotic resistance site and the T7 forward sequencing primer. Figure drawn using SnapGene.

2.2.2. Microbiological methods

2.2.2.1. Growth measurement

Cell growth was measured by aliquoting 1 ml of growing cells in disposable plastic cuvettes, and the optical density was recorded at 595 nm using a WPA CO 8000 Biowave Cell Density Meter.

2.2.2.2. E. coli growth

Starter cultures of 5-10 ml were grown overnight at 37°C from glycerol stocks, or preferably from single colonies on a previously streaked agar plate containing the appropriate antibiotic. This starter culture was then inoculated in 1 L LB medium to be induced later, or in ZYM medium for autoinduction experiments.

2.2.2.3. Long-term storage of bacterial clones

Cells were preserved by mixing equal volumes of the re-suspended cell pellet in fresh LB medium, with 50% cooled-autoclaved glycerol (glycerol in LB). This was vortexed to mix and snap frozen in liquid nitrogen and stored at -80 °C.

Note: glycerol concentration varies from 15 - 25% for long term storage.

2.2.2.4. Bacterial harvesting

All bacterial cells were harvested by centrifugation at 4,700 X g for 30 minutes at 4°C.

2.2.2.5. Blastocystis purification and harvesting

Cells were harvested at 800 X g for 5 minutes and re-suspended in 100 µl of LYSGM medium. To remove bacterial flora from xenic cultures, re-suspended cells were layered carefully on top of 10 ml Histopaque 1077 and centrifuged for further 30 minutes at 2,000 X g. A thin layer of *Blastocystis* cells is noticeable 1

cm from the top of the multiple-layers solution, this was pipetted out and centrifuged and washed with PBS ([Parkar, Traub et al. 2007](#)). This is either re-inoculated in IMDM for axenization experiment ([Ho, Singh et al. 1993](#)) or in LYSGM for further xenic cultivation ([Clark and Diamond 2002](#)), or collected via centrifugation at 275 X g for cryopreservation.

2.2.2.6. Long-term preservation of *Blastocystis* sp. strains

Blastocystis cells were centrifuged at 275 X g for 5 minute and the supernatant was decanted. Two tubes were prepared as follows: five ml of the culturing medium (LYSGM) with five ml heat inactivated adult bovine serum in the first tube, then 5 ml of the first tube was transferred to the second tube with the addition of 750 µl of DMSO to reach a final concentration of 15% DMSO.

The cells were re-suspended with two ml of tube one (having 50% serum and 50% media). Then 500 µl of the cells suspension were transferred to each of the cryo-tubes. To this suspension 500 µl of tube two that contains 15% DMSO were added. This mixture was then left in 36°C incubator for about 15 minutes, allowing *Blastocystis* cells to absorb the DMSO. Then cells were rapid-cooled by transferring them into a pre-chilled isopropanol bath at -80°C for 48 hours (isopropanol bath should be pre-chilled at least 24 hours before use).

After two days the cryo-tubes were moved to a liquid nitrogen container (Dewar). Cells should be stored for seven days before reviving cells for checking for viability.

2.2.2.7. Recovering cryo-preserved *Blastocystis* cells

Cryo-preserved cells were rapidly thawed in a water bath at 36° C, and left without agitating for 10-15 minutes. These cells were then transferred to tubes containing

10 ml of pre-warmed LYSGM and incubated in 36° C incubator. After incubating for 2-3 days and visual inspection, the medium was decanted and replaced with fresh medium. This was checked after a further 72 hours and sub-culturing continued as normal.

2.2.2.8. Blastocystis ST cultivation method

2.2.2.8.1. Xenic culturing method

Blastocystis cells were grown xenically in LYSGM medium. The cells were cultured twice a week in borosilicate tubes containing pre-warmed 10 ml liquid medium supplemented with 5% heat inactivated adult bovine serum. The fresh media tubes were inoculated with 100 µl of 3-4 days old culture and kept vertically in 37°C incubator. The cells were sub-cultured every 3-4 days (twice a week).

2.2.2.8.2. Axenization trials

Several trials were performed for axenizing *Blastocystis* cells, using two-phase (egg slant medium using modified Luck's solution) and single-phase medium IMDM.

2.2.3. Biochemistry methods

2.2.3.1. Induction of gene expression

10 ml overnight cultures were inoculated by a single colony from a particular strain at 37°C at 220 rpm for 12-16 hours. This overnight culture was inoculated in 1 L LB medium or ZYM for autoinduction experiments.

To check for over-expression, cells were exposed to different conditions. Cells induced with IPTG were grown at different temperatures: 37°C, 28°C and 10°C and for different induction times from (2 hours to overnight). Several IPTG concentrations were used as well (1 µM-1 mM).

An auto-induction method was also used for protein production purpose ([Studier 2005](#)), these were grown at 20°C for 24 hours or 48 hours at 220 rpm.

A sample of the over-expressed culture was taken, centrifuged to collect cells and analysed on SDS-PAGE.

2.2.3.2. Cell lysis

The induced cells were harvested by centrifugation at 6,000 X g for 30 min. The bacterial cell pellet was re-suspended using a suitable homogenisation buffer containing protease inhibitor (cOmplete Protease Inhibitor Cocktail Tablets (Roche) (1 tablet/50 ml buffer). Cell suspensions were lysed using the sonication method (sonoporation), to disrupt cell membrane and release cell content. The solution was sonicated using a Soniprep 150 Sonicator (Sanyo), at a frequency of 25 kHz and 60% amplitude on ice for a total of 5 pulses (cycle), with 15 seconds on/25 seconds off cycles to prevent overheating and degradation of the enzymes or aggregation of the targeted protein. Lysed cells were centrifuged at

20,000 X g for 30 minutes (or until having a clear soluble fraction) to pellet cells debris and insoluble fractions.

2.2.3.3. Protein purification

All proteins studied were His-tagged, which enabled the use of the Immobilized Metal Affinity Chromatography (IMAC) technique. This was done manually or using the ÄKTA purifier (GE Healthcare). When required, protein was further purified by ion exchange (IEX) chromatography. Proteins were also purified using size exclusion chromatography columns (Gel Filtration, GF). Table 2.7 summarises all buffers used in this study for purifying proteins. Those buffers were vacuum filtered through 0.2 µm nylon membrane. The protein concentration was determined by measuring the absorbance at 280 nm using the ÄKTA column or by using a Bradford assay using bovine serum albumin (BSA) standard curve.

2.2.3.3.1. Immobilized metal affinity chromatography (IMAC)

The crude soluble fraction of the lysed cells was loaded on 1 ml nickel column (column loaded with X ml of metal chelating cellulose charged with Ni²⁺) using a “Super loop”. A 1 ml nickel affinity column (HisTrap column/ GE Healthcare) was equilibrated with 5 column volume of nickel affinity column (NAC) buffer A, detailed in Table 2.6. Protein was loaded using the “super loop” (GE Health care) and the unbound protein was washed out using 5 X column volume of NAC buffer A. The desired protein was eluted using 5 column volume of NAC buffer B (Table 2.6.) using a step gradient of buffer B up to 100% of this buffer. A further washing of the column was performed using 2 X column volume of buffer B. The fractions were collected in 1 ml fractions.

The fractions representing the desired protein were collected and concentrated using a 30 kDa cut-off membrane (Millipore, Amicon® Ultra-15 Centrifugal Filter Unit) at 4,000 X g, at 4°C until a final volume of 1 - 5 ml was collected to be used in further purification steps.

2.2.3.3.2. Ion Exchange Chromatography (IEC)

Protein purified using HisTrap columns was, for further purification, isolated using HiTrap Q Sepharose FF (Q FF) fast flow column for anion exchange or an SP XL column for cation exchange. The buffers used in this experiment are listed in Table 2.6. Protein was loaded on the Q FF column on ÄKTA using “Super loop”. This column was equilibrated using 5 X column volume. Unbound protein was washed off with 3 column volumes of Q FF buffer A, while the target protein was eluted with 5 column volumes of Q FF buffer B. The fractions were 1 ml unless otherwise stated. The eluted protein was checked on SDS gel, and the fractions corresponding to the target protein were pooled and concentrated to 1-5 ml final volume using a 30 kDa cut-off membrane filters (Millipore, Amicon® Ultra-15 Centrifugal Filter Unit) at 400 X g at 4°C. This fraction was used for the gel filtration (GF) purification step afterwards.

2.2.3.3.3. Gel Filtration chromatography (GF column)

The final protein purification step was through gel filtration. A Superdex 200 HiLoad 16/60 (GE Healthcare) (or 75 for analytical column used for molecular weight and oligomeric status determination) gel filtration column was first equilibrated with the chosen elution buffer mentioned in Table 2.6. The equilibration step was performed using one column volume at a flow rate of 1

ml/min. The sample was loaded by injection into loops of different size depending on the pooled concentrated protein. The protein was eluted with 1 column volume of GF buffer, with elution fractions of 1 ml unless otherwise stated. The protein concentration was measured at 280 nm, and analyzed using SDS-PAGE. Fractions corresponding to the target protein were pooled and concentrated for further experiments.

The Superdex 200 (or 75) gel filtration column was calibrated with the protein standards mentioned in Table 2.13 in materials and methods, in triplicates. Those protein were analysed on the column equilibrated using the same buffer as used in protein elution. The void volume was determined by measuring the elution volume of the protein standard dextran (2,000 kDa). The calibration curve obtained from plotting log MW against K_{av} was used to estimate the molecular weight of the native protein. Void volume value for GF column Superdex 200 was 38.89 ml.

Where K_{av} is calculated using the following equation:

$$K_{av} = \frac{V_e - V_o}{V_t - V_o} \quad \text{Equation 2.1}$$

For details of the calibration curve refer to Appendix 2.

2.2.3.4. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were analysed by SDS PAGE gels, a method based on Laemmli (1970), for separating proteins according to their molecular weight. Gels were prepared in concentrations depending on the size of protein studied, where large proteins were analysed with 12% SDS PAGE separating gels, and smaller size protein were analysed using 15% SDS PAGE separating gels. Stacking gel concentration was always 6%. The protein molecular weight was usually determined visually by Fermentas Spectra Multicolor Broad Range Protein Ladder molecular weight marker (Fig 2.2). See Table 2.5 for detailed composition of different gel concentrations. The gel was exposed to an electric current of 25 mA per gel, using a Bio-Rad mini-protean-Tetra cell (Bio-Rad Laboratories Ltd).

2.2.3.5. Native gel

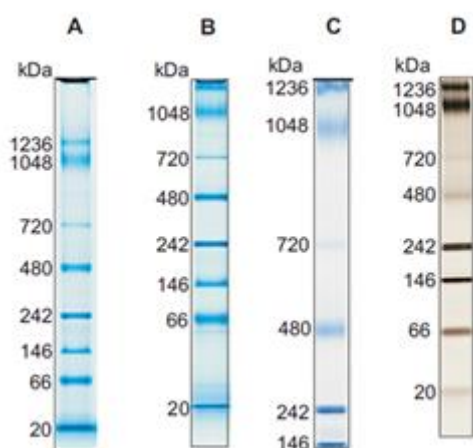
Native gel electrophoresis was used to determine the oligomeric status of proteins. These gels are prepared in non-denaturing buffers, which maintain the protein in its secondary structure, and therefore proteins separate according to their charge-to-mass ratio. The protein usually has an acidic or slightly basic charge, and by applying a current, the protein migrates to the negative charge. In this technique the gel filtrated protein sample is incubated with sample buffer for 15 minutes on ice. After loading the protein sample and the markers (20 µl each), and in the case of blue native gel, overlay the sample well and each well with a sample cathode buffer. The inner chamber is filled with cold cathode buffer, and the outer chamber is filled with cold anode buffer. The gel is then run by applying a constant current of 10 - 20 mA for about 3 hours (100 V), or at 150 V for 1 hour or until the dye reaches the running front of the gel (in the case of blue native gel), the work is all performed at 4°C ([Wittig, Braun et al. 2006](#), [Fiala, Schamel et al. 2011](#)). For more details about the expanded composition of the native gels see Table 2.6. The markers used in this method were a native protein marker (Sigma-Aldrich). See Table 2.13 for marker composition.

Table 2. 13 Proteins used as native molecular weight marker.

Protein Marker	Concentration mg/ml	Molecular weight kDa
Carbonic anhydrase	3	30
Albumin	10	66.5
Alcohol dehydrogenase	5	150
B amylase	4	200
Apoferritin	10	440
Thyroglobuline	8	660

Another marker was used in this study, NativeMark™ Unstained Protein

Standard (LifeTechnologies), see Fig. 2.5.



Protein Band	kDa
IgM Hexamer	1,236
IgM Pentamer	1,048
Apoferitin band 1	720
Apoferitin band 2	480
B-phycoerythrin	242
Lactate Dehydrogenase	146
Bovine Serum Albumin	66
Soybean Trypsin Inhibitor	20

Figure 2. 5. Unstained native protein standard NativeMark™ for native gel from Life Technologies. Protein marker is shown under different running conditions. The table shows the proteins representing each band with their corresponding molecular weight.

2.2.3.6. Second dimension SDS-PAGE

To check for the identity of native proteins obtained in native gel electrophoresis, the whole lane that contains the protein of interest was excised, placed in 2 X SDS sample buffer, left shaking for about 10 minutes at room temperature and boiled for 20 seconds in the microwave. The band was then incubated in the hot SDS sample buffer for another 15 minutes before being analysed on standard SDS gel. The excised band was loaded by inserting the excised lane in a pre-prepared SDS gel containing only one large well, overlaid with SDS sample buffer and subsequently electrophoresed as normal ([Fiala, Schamel et al. 2011](#)). This will enable the determination of the molecular weight of the proteins of interest.

2.2.3.7. Western blotting

After SDS gel electrophoresis, the gel was disassembled and proteins transferred to nitrocellulose membrane via Western blotting set from Bio-Rad and using Mini Trans-Blot® Cell. The gel, membrane, and blotting papers were equilibrated in Towbin buffer by gently rocking at room temperature for 10 minutes prior to

transfer. The transfer cassette used was assembled as follows, placing the white side of the cassette at the bottom submerged in the transfer buffer (facing the positive electrode):

Sponge

Blotting paper (thick)

Nitrocellulose membrane

SDS PAGE gel

Blotting paper (thick)

Sponge

This sandwich of layers was put on top of each other and a test tube was gently rolled over the assembly to remove any air bubbles, which might affect the transfer. The cassette was closed with the black side which will face the negative electrode later. The transfer chamber was covered with Towbin buffer to the top and protein transfer was performed for 1 hour and 15 minutes, at 400 mA, with the whole transfer unit was kept in ice to protect the protein against denaturation during transfer.

The quality of transfer was checked by briefly staining the nitrocellulose membrane with Ponceau S for several minutes. Unbound dye was removed by washing in water leaving the transferred protein visible as red stained bands.

2.2.3.8. Colony hybridization (colony blot)

Colony blot was used to identify clones that expressed the gene of interest and to have an estimation of the variation in expression levels between colonies. This

was performed after transformation, which enables identifying clones to be selected for subsequent expression experiments. To start a replicate plate of the transformation plate was grown up overnight at 37°C. The replicate was removed out of the incubator to let the condensation dry off for about 10 minutes. A dry nitrocellulose membrane was placed on top of the colonies (by holding it with forceps and place it gently starting from the middle of the plate to the edges to minimize air between the membrane and the cells). On this membrane three points were pierced on the edges of this filter through the agar using a needle, which help in the alignment of plate and membrane after the detection step. The membrane was then peeled off the membrane in one movement and placed on another AMP-LB-Agar plate containing 250 µM IPTG to induce gene expression in the cells. The attachment was made so that the first contact point between the membrane and the plate in the middle, and lower from that point to the sides. The plate was incubated for about 4 hours at 37°C, while the master plate was incubated at 30°C incubator to allow the master colonies grow again. Four plates were hence prepared for next steps. Each Petri dish contained Whatman paper soaked in each of the solutions mentioned in 2.1.2.3.2.4 successively, and marked as A, B, C, and D corresponding to the solution they contain. Excess of those solutions were discarded leaving them moist but not wet. The nitrocellulose membrane was incubated for a further 10 minutes in solution A, five minutes in solution B, five minutes in solution C twice, and finally 15 minutes in solution D, as follows:

Solutions in order	Incubation time
Solution A	10 minutes
Solution B	5 minutes
Solution C	5 minutes, two times
Solution D	15 minutes

The membrane was subsequently developed in the same way as explained in the Western blotting protocol (2.2.3.7).

It is advised to have the following plates as controls with the main experiment:

- * A plate of empty host bacteria without the expression plasmid.
- * A plate with host cells with the empty expression vector.

1.2.3.9. *Blastocystis enolase assay*

The reaction was measured in 4 ml quartz-cuvettes over a period of 5 minutes. The first minute was used to allow for NADH auto-degradation, until absorbance was stabilised. At which point the protein was added to initiate the reaction.

The final assay mixture contained the following: 81 mM triethanolamine, 1.9 mM 2-phosphoglycerate, 0.12 mM β -nicotinamide adenine dinucleotide reduced form, 25 mM magnesium sulphate, 100 mM potassium chloride, 1.3 mM adenosine 5'-diphosphate, 7 units pyruvate kinase, 10 units L-lactic dehydrogenase and 0.025-0.05 unit Enolase. Where one enzyme unit is defined as the amount that will convert 1 μ mole of the substrate (2- phosphoglycerate) to

phosphoenolpyruvate per minute at pH 7.4 at 25°C, see Equation 2.2 below for calculations.

$$\text{Units/ml enzyme} = \frac{(\Delta A_{340\text{nm}}/\text{min Test} - \Delta A_{340\text{nm}}/\text{min Blank})(3)(\text{df})}{(6.22)(0.1)} \quad \text{Equation 2.2}$$

$$\text{Units/mg protein} = \frac{\text{units/ml enzyme}}{\text{mg protein/ml enzyme}}$$

$$\text{Units/mg solid} = \frac{\text{units/ml enzyme}}{\text{mg solid/ml enzyme}}$$

Where:

3= volume (in millilitres) of assay

df= dilution factor

6.22= millimolar extinction coefficient of β -NADH at 340 nm

0.1= volume (in millilitre) of enzyme used

For the determination of kinetic parameters (Michaelis-Menten, K_m and V_{max}), 0.5 unit of the recombinant *Blastocystis* enolase in GF buffer was used with substrate concentrations ranging from 0.025-10 mg/ ml. The Michaelis-Menten Equation 2.3 was followed.

$$V_0 = V_{max} * [S] / K_m + [S] \quad \text{Equation 2.3}$$

Where V_0 is the initial velocity of the enzyme, while V_{max} is the maximal velocity.

S is the substrate (2-phosphoglycerate) concentration used in the assay and K_m

is Michaelis-Menten constant representing the substrate concentration at which the reaction velocity is half the maximal velocity. SigmaPlot version 7 was used to calculate the K_m and V_{max} .

1.2.3.10. Biochemical and biophysical characterisation of the proteins

All activity assays and standard curves were performed using biological and technical triplicates

2.2.3.10.1. BSA Standard curve

A serial dilution of Bovine Serum Albumin (BSA) was made in concentration range from 0.25-2 mg/ml. The concentration was measured by mixing 20 μ l of BSA of the serial dilutions with 980 μ l Bradford reagent after being brought to room temperature and measuring the colour intensity using a spectrophotometer at 590 nm. A standard curve was produced and used for the measurement of the concentration of other unknown proteins studied in this project.

2.2.3.10.2. Dynamic Light scattering, DLS

Light passing through biological solutions will be scattered by the molecules it encounters through its path. In static light scattering, we analyse the intensity of the light, while in dynamic light scattering we measure the light's fluctuation which is due to the Brownian motion of the particles. Dynamic light scattering measures the diameter of a particle that moves the same way as the studied protein. In this method there is no need to know the concentration of the molecules, it only needs enough light to be scattered. Light scattering measures the diffusion of the molecule in the solution and the statistics of the scattered signal is analysed in relation with a correlator and inverted to size distribution measurements

([Nobmann, Connah et al. 2007](#)). This method is a good non-invasive method, that requires small amounts of the protein, and able to detect a trace amount of aggregation in the sample.

The protein sample was centrifuged for 20 min at 16,000 X g at 4°C. 18 µl of the purified centrifuged protein was injected in a special, cleaned, quartz cuvette. Several measurements (10 measurements in this case), with 10 acquisitions each measurement and 20 seconds each acquisition time were recorded at 25°C using DynamPro Titan (Wyatt Technology Corporation). Data was analysed using the DYNAMICS programme (protein solution, USA).

2.2.3.10.3. Differential Scanning Fluorimetry, DSF

Aggregation might be a critical bottleneck towards drug development and protein production experiments. Many factors which were found experimentally and via observations, play an important role in protein stability, mainly the composition of the solution in which the protein is purified or crystallized in. Thermal shift assay is one of the valuable methods used to estimate the stability of the protein in different conditions in addition to the ability of having an insight into the protein-ligand interactions ([Senisterra, Chau et al. 2012](#))

DFS was developed by Johnson and Johnson to overcome the problems met when developing new drugs faced through aggregation problems encountered when using different ligands, buffers or substrates. In this experiment the protein is mixed with a dye (SYPRO Orange in this case) that fluoresces once the protein is unfolded or aggregated. This extrinsic fluorescence signal is measured in terms of increments by 1°C/min, when a rise of temperature is introduced to the mix

between 25 - 95°C. The signal produced by this procedure is amplified by the machine and the plate fluorescence is captured using a CCD camera.

The raw data is fitted according to standard equations to measure fluorescence intensity vs. temperature. Any changes in the composition of the protein solution would have a direct effect on the melting temperature of the protein, and hence affecting its stability ([Mezzasalma, Kranz et al. 2007](#)). A schematic representation of the thermal shift status that a protein undergoes is depicted in Figure 2.6.

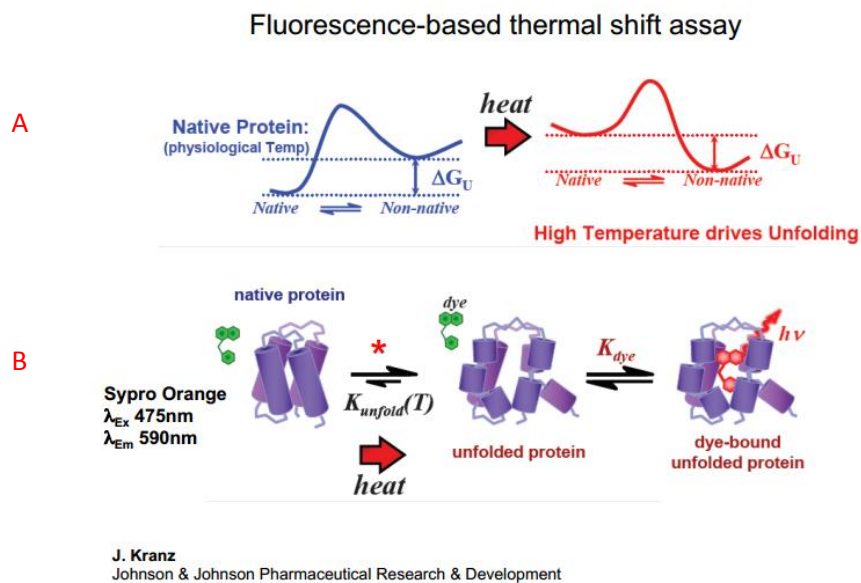


Figure 2. 6 Schematic diagram representing the theory of thermal shift assay. A: the shift in absorption is noticeable after exposure to high temperature. B: the mechanism by which SYPRO Orange get quenched by the protein after being heat- denatured (from Johnson & Johnson Pharmaceutical Research & Development).

To estimate the thermodynamic stability of the protein, there is a need of estimating the equilibration between folded and unfolded protein ([Niesen, Berglund et al. 2007](#)) whether this denaturing is induced by a denaturant or by heat.

Several dyes have previously been used in this context. Of which SYPRO Orange (excitation at 465 nm and emission at 610 nm) was proven to have several

beneficial characteristics: bright fluorescence, chemical stability, high fluorescence quantum yield and spectral properties suitable for RT-PCR instruments ([Shi, Semple et al. 2013](#)). The main benefit of using SYPRO Orange is its high signal to noise ratio ([Senisterra and Finerty 2009](#)). SYPRO Orange (Invitrogen) was used in this project, which binds to the unfolded/denatured protein after being heated using a qPCR machine (StepOne™ Real time PCR) and results were analysed using PROTEIN THERMAL SHIFT™ software, version 1.0 (Applied Biosystem, UK).

Buffer-stability was determined in the same experiment and was assessed on 24 buffers in duplicates. The reaction mixture was as follows:

For 24 wells a master mix was prepared on ice, with a total volume of 450 μ l containing:

Protein up to 0.133 mg/ml

5,000 X SYPRO Orange 0.96 μ l

Water up to 450 μ l

18 μ l of the master mix was used in each well with 2 μ l of the buffer screen or the studied compound ([Vivoli, Novak et al. 2014](#)).

2.2.3.10.4. Protein parameter determination:

Protein characteristics were analyzed using the online program ProtParam tool, <http://expasy.org/tools/protparam.html>. This facility helps calculating various protein characteristic parameters such as the iso-electric point of the protein pI,

its theoretical stability, estimated half-life, extinction coefficient, in addition to its theoretical molecular weight. All analyses are based on the provided amino acid sequence of the protein.

2.2.3.10.5. Analytical ultra-centrifugation:

The analytical ultra-centrifugation (AUC) experiment was conducted using facilities at the Research Complex at Harwell (RCH), Oxford.

Analytical ultra-centrifugation enables detecting the molecular weight and the shape of macromolecules. Velocity sedimentation (VS) gives an idea about the shape and mass of the macromolecules studied while equilibrium sedimentation (ES) only provides the mass of the macromolecule ([Cole, Lary et al. 2008](#)).

In analytical ultra-centrifugation the movement of a molecule in a solution is detected under the application of high speed produced by centrifugal forces. AUC establishes a balance between the sedimentation of the molecule under the centrifugal forces and the diffusion of the material in the other direction. Ultimately, a point is reached where there is no migration of material in the unit cell.

This detected movement is measured, recorded and analyzed using software integrated in the system that helps controlling the entire experiment from initialization, running, and data collection. Data analysis software analyses the data and provides information about the hydrodynamic and thermodynamics of the macromolecule calculating parameters such as sedimentation and diffusion coefficients, and provides an estimate of the heterogeneity of the sample.

The RCH facility offers the use of the Beckman Coulter ProteomeLab XL-1 Protein Characterization System, which has an AN-60 Ti rotor that rotates up to

60,000 x rpm (about 290,000 x g). This high speed is used in sedimentation velocity (SV) experiments while lower speeds are used when performing sedimentation equilibrium (SE) experiments. This system comprises two integrated detection systems: Rayleigh Interference and Scanning UV/Visible optics equipped with a CCD camera with a high resolution (2048 X 96 pixels). The detection rate is conducted at a high scanning rate of roughly once every 5 seconds.

The other system used is Windows-user interface software that allows initiating the experiment, and also recording and analyzing the data and calculating the parameters reflecting the hydrodynamics and thermodynamics properties of the macromolecules.

The AUC machine was operated by Professor David Scott (University of Oxford) at the Research Complex at Harwell (RCH). To conduct this experiment, three different concentrations of the protein were used: 1, 0.2 and 0.05 mg/ml of the protein using 400 µl of each sample. The protein was analysed by analytical centrifugation at 40,000 rpm speed generating about 129,000 X g. Each protein sample was centrifuged in a separate well, after being calibrated versus water. The procedure was conducted at 4°C using the rotor mentioned above, and AUC parameters were computed using the integrated software.

2.2.3.10.6. Crystallization work

2.2.3.10.6.1. *Initial crystal trials*

Proteins were purified using different purification methods as explained in section 2.2.3.3. Purified proteins were initially concentrated using Amicon® Ultra Millipore filters with varied cut-offs according to the protein's molecular weight. A final

concentration of about 12 mg/ml was generally used to conduct the initial crystallization trials. Before starting crystallization a pre-crystallization test Kit (PCT™, Hampton Research) was used according to the manufacturer's instructions. The pre-crystallization kit gives an idea of the concentrations to be used for setting crystallization dishes, and thus saving time and money. In this experiment if a precipitate forms, it indicates that the protein is too concentrated and therefore likely that the screen will result in many amorphous precipitates. On the other hand, if the PCT kit produces clear drops (no precipitate at all) this would likely also be the case in the actual crystal screen. A set of screens were used in this project from different companies as mentioned in 2.1.2.3.2.7. The protein under study was pelleted by centrifugation at 12,000 X g for about 20 minutes. A total of 50 µl was used for each screen. An automated robot was used in a high-throughput crystallization pipeline in Oxford Protein Production Facility (OPPF) at the University of Oxford. This Hydra instrument allows for loading 96-well Greiner plates with both the screen blocks and the protein solution using nanolitre-scale droplets (usually 100 nl protein/100 nl screen). These are subsequently kept in a Formulatrix storage and imaging system at 21°C. Plates are checked at automatically at regular intervals via a remote web-interface facility equipping stereo-microscopes with Pixera imaging cameras.

2.2.3.10.6.2. Crystal harvesting and X-ray data collection

Remotely captured images of the crystal screens were regularly checked for any crystal formation. Crystals generally take between several hours to several weeks to form. Once detected, a crystal was left and repeatedly checked for further growth. Once it reached a satisfactory size, the crystal was harvested by the help of my colleague Dr. Michail Isupov in our Biocatalysis Centre in Exeter. The

crystal was harvested using a cryoprotectant prepared according to the conditions that the crystal grew in. Harvested crystals were subsequently stored in liquid nitrogen until sent to the Diamond Light Synchrotron Radiation Source (www.diamond.ac.uk) where data collection took place using beamlines I03 and I04-1 with the help of Dr. Isupov who assisted in data collection and analysis.

2.2.3.10.6.3. Crystallization optimization

Optimization steps were performed on the conditions that produced crystals with good diffraction potential after X-ray exposure for obtaining a better resolution. The optimization was performed using Hampton 96 well plates micro-batch utilizing an Oryx 6 automated crystallization robot (Douglas research, UK). This system contains a programme that helps setting an optimization screen with varying buffer, precipitant, or protein concentrations.

2.2.3.10.7. Enzyme activity assay

All enzyme activities were performed on gel filtration (GF) purified proteins in biological and technical triplicates. Details of activity assay are described in the main *Blastocystis* enolase biochemical characterization chapter.

2.2.3.10.8. Homology modelling

Homology modelling was used to predict the 3D structure of a protein with known sequence based on comparison with another protein with a solved 3D structure (the template protein). Homology modelling allows the prediction of possible

function of the studied protein in addition to possibly identifying the binding site, provided that residues of conserved pocket are intact.

2.2.3.10.8.1. Template selection

The template protein is chosen after performing a BLASTP search against the PDB database using the amino acid sequence of the protein of interest as the query sequence. The template should usually share more than 30% similarity with the studied protein and ideally would be a protein with a solved structure.

2.2.3.10.8.2. Target -template alignment

After selecting the most suitable (identical) sequence, multiple alignment analysis is performed using ClustalW ([Larkin, Blackshields et al. 2007](#)). Alignments are manually inspected to remove gaps and make sure known important residues are aligned.

2.2.3.10.8.3. Model building and inspection

Model building was conducted using the automated server at <https://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html>. The quality of the model was checked and confirmed by superimposing the model structure on the template structure using PyMOL (<http://www.pymol.org>) and by manually inspecting the model (especially the active site).

Chapter 3. *Bioinformatic analysis of the Blastocystis glycolytic enzyme enolase*

The development of genomics and bioinformatics was a boon for the field of biology and molecular biology. Understanding the genomic potentials and capacities of an organism will enable a defined understanding of the metabolism and physiology of a pathogen. This knowledge has the potential to make a profound difference in drug development and therapeutics in order to possibly eradicate those pathogens. An important part of this process is to understand the critical proteins/enzymes that play important roles in the pathogen's metabolism. In this chapter, an initial *in silico* analysis was initiated in order to predict several structural and functional characteristics of *Blastocystis* enolase, a key glycolytic enzyme. Several publically available tools were utilised in this study, using the amino acid sequences of the protein and other proteins from different organisms. Bioinformatics analyses and structural modelling methods were used. This included identifying signature motifs of the enolase super family, checking for possible targeting signals, identifying a suitable homolog with a solved crystal structure to build a 3D model, predicting protein domains, etc. Phylogenetic analysis was performed on the *Blastocystis* enolase sequence to assess its possible evolutionary path.

3.1. Enolase super family

Blastocystis enolase belongs to the enolase super family. This is a highly diverse and broad super family containing over 8,000 species from members of all domains of life. *Blastocystis* enolase sequence was run on Pfam database, this came back with two Pfam domains: enolase N (which is the N-terminal domain) and enolase C (which is the C-terminal domain) (see Fig. 3.1)([Gerlt, Babbitt et al. 2012](#)).



Figure 3. 1 Predicted domain structure of *Blastocystis* enolase by Pfam. Showing *Blastocystis* enolase to belong to the enolase super family, which consists of two main Pfam domains: Enolase N; the N terminal domain and Enolase C; the C terminal TIM barrel domain.

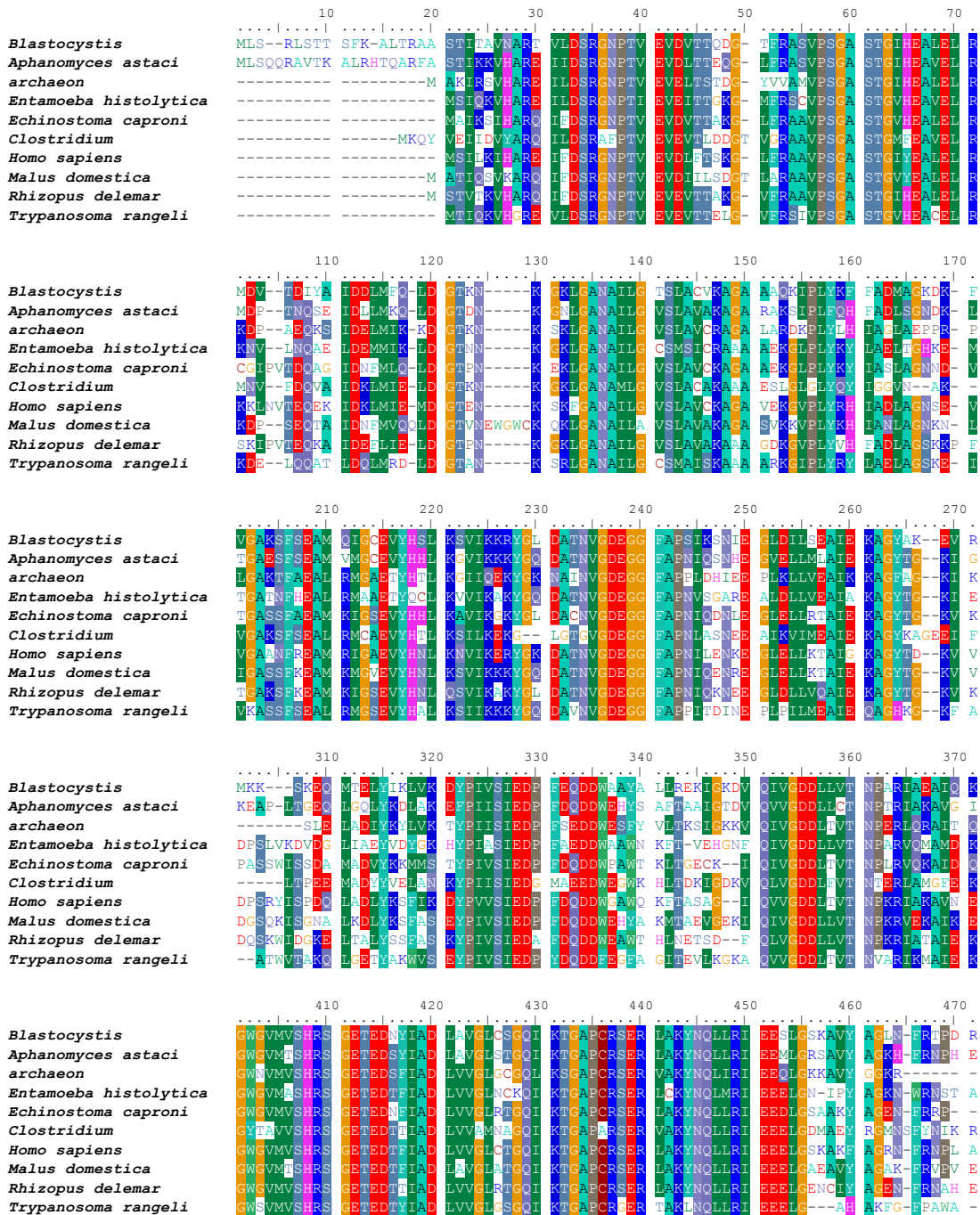
3.2. Pre-sequence and signature motif

The DNA sequence encoding enolase has an open reading frame of 1350 nucleotides, indicating that *Blastocystis* enolase is composed of 449 amino acids with predicted molecular mass of 48 kDa and predicted iso-electric point (PI) of 6.56.

Blastocystis enolase homologues were found using BLAST in the NCBI database. A BLASTP search of the translated *Blastocystis* enolase gene sequence was performed using the NCBI's web interface using the non-redundant protein sequence data base to identify possible homologues. An initial BLAST search of *Blastocystis* enolase sequences showed the high abundance of this protein in the tree of life, specifically *Plantae*. Homologues of *Blastocystis* enolase were identified in this research using a representative of each main group

including; *Plantae* (*Malus domestica*), Excavates (*Trypanosome rangeli*), Fungi (*Rhizopus delemar*), Amoebozoa (*Entamoeba histolytica*), Chromaveloata (*Aphanomyces astaci*), Animalia (*Echinostoma caproni*) as well as in Archaea (Archeon) and Bacteria (*Clostridium* and *Escherichia coli* ([Dannelly and Reeves 1988](#))). *Blastocystis* enolase exhibited high identity to previously characterized enolases from the previously mentioned groups in accordance with the fact that this is an essential enzyme and a highly conserved protein participating in several vital processes specifically glycolysis ([Piaś, Kustrzeba-Wojcicka et al. 2005](#)). The multiple sequence alignment included homologues with similarities of at least 60% except for bacteria and amoeba which was 57, and 58% respectively. Selected sequences and the *Blastocystis* query sequence were aligned using ClustalW ([Larkin, Blackshields et al. 2007](#)). The alignment shows that the primary structure conservation is not restricted to specific parts of the enolase sequence; it extends from N- to C-terminus, with highest similarity at the C-terminal, Figure 3.1.

This protein and its homologous share two conserved domains or two families; Enolase N forming the N terminus domain and enolase C forming the C terminus domain forming the TIM barrel domain. Enolase N is found at the N-terminus of the catalytic TIM-barrel like domain in enolase and other enzymes. Enolase C is found in enzymes that adopt TIM barrel fold.



3.2.1. Enolase pre-sequence analysis

A more detailed BLAST search was performed using more representatives of several groups of life. All sequences that lacked an obvious start at their amino-terminus were excluded manually. Aligned sequences were inspected by eye and sequences containing aberrant sequence regions that were highly divergent and ambiguously aligned positions were removed manually before the subsequent phylogenetic analysis.

Although there is no consensus sequence for mitochondrial targeting signals, the N-terminal amino acids usually shows a bias in distribution forming a kind of a pattern that was related to a mitochondrial targeting signal. This pattern seems to be associated with cleavage of precursor protein by a certain protease. In literature, the mitochondrial targeting peptides (mTP) were found to be rich in basic, hydrophobic and hydroxylated residues and they lack acidic amino acids. Using those features helped in building predictions for the mitochondrial targeting sequence ([von Heijne 1986](#)). Possible targeting signals in this project were detected using the following four programmes; Mitoprot ([Claros 1995](#)), Predator ([Small, Peeters et al. 2004](#)), TargetP ([Emanuelsson, Brunak et al.](#)), and iPSORT ([Bannai, Tamada et al. 2002](#)) as indicated in Table 3.1 which shows the prediction signal with cleavage site and sequence when applicable. Each sequence was checked for possible targeting signals and Figure 3.2 shows the N-terminal region of several enolase proteins which indicate a possible mitochondrial targeting signal. Sequences with predicted mitochondrial targeting signal were rich in basic residues with no acidic amino acids at all.

It is suggested that a kind of sequence specificity might be available for mitochondrial targeting peptides. This was found to be true as the mitochondrial

targeting signals were later found to be sharing the following features: arginine rich in position -2, -3, -10 and -11 in relation to the cleavage site, or at least a small and neutral residue close to the cleavage site, for the cleavage to occur ([Nielsen, Engelbrecht et al. 1996](#)). Another feature was found for mTPs, which was a bias towards having hydrophobic residues in position -8 that have arginine in position -10 of the mitochondrial targeting peptide (mTP) ([Gavel and von Heijne 1990](#)).

In this particular project, those predicted targeting signals followed the description expected of having a net positively charged mitochondrial targeting residues, with a length between 17-70 (typically 30), enriched with arginine residues specially at the cleavage sites R-2, R-3, R-10, apart from few who had different localization of the arginine residues, Figure 3.3. And as mentioned in Heijne *et al* analysis of the mTP, the R-10 had the conserved R-X (F-I-L), and it was isoleucine (I) in this instance. A similar alignment was made between *Blastocystis* and members of the stramenopiles group, Figure 3.4. Those spotted patterns around the cleavage site, could be an indicator of a single protease responsible for the cleavage of mTP in R-2 and R-3, as well as the first cleavage of R-10 ([Gavel and von Heijne 1990](#)).

Analysis of the targeting signal in this project suggests that the *Blastocystis* enolase possibly contains a mitochondrial targeting signal with high support values. It seems that stramenopiles in general also contain a mitochondrial targeting signal on their enolase which is quite interesting finding. Most stramenopiles With the exception of *Saprolegnia parasitica*, *Plasmopara halstedii* and *Aureococcus anophagefferens*, seem to contain these pre-sequences as well. Those studied expressed high tendency for targeting the mitochondria with

a signal prediction of 0.93 and more, except for *Phaeodactylum tricornutum* which had a value of 0.65 for signalling the mitochondria.

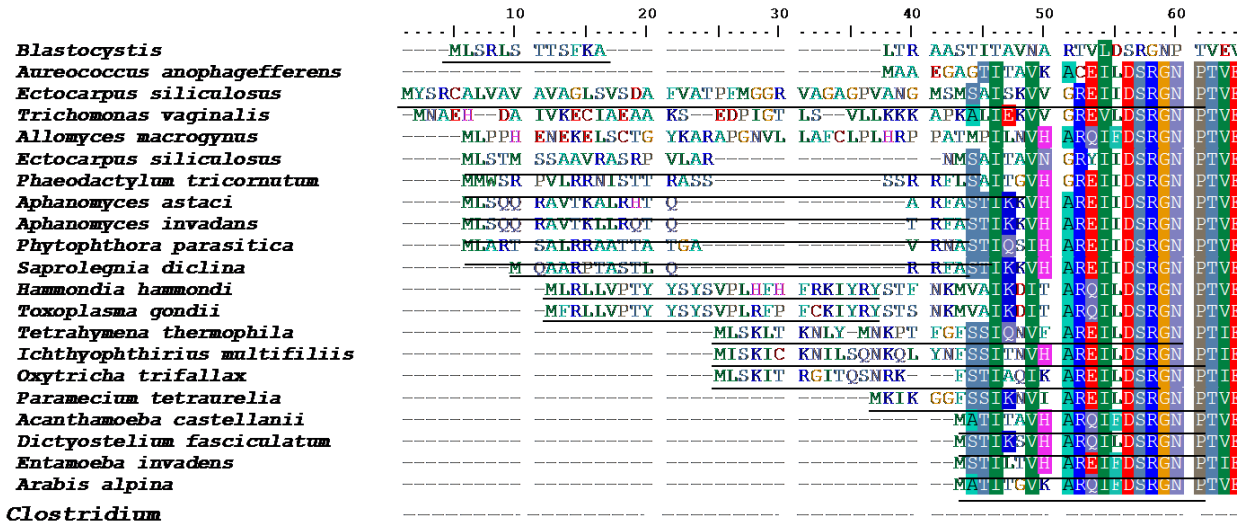


Figure 3.3 Analysis of the N-terminal region of *Blastocystis* enolase in comparison to other organisms, indicating a possible mitochondrial targeting signal depending on MitoProt (Claros 1995). Targeting signals were predicted using four programs; MitoProt, Prodeator, iPSORT, and TargetP. The predicted targeting signals are underlined in the figure in black colour. The figure shows no targeting signal for bacteria; *Clostridium* (WP_010294244.1) and other organisms used have the following accession numbers respectively. *Blastocystis* of the study, *Aureococcus anophagefferens* (EGB09778.1), *Ectocarpus siliculosus* (CBJ32586.1), *Trichomonas vaginalis* (EAY10059.1), *Allomyces macrogynus* (KNE63834.1), *Ectocarpus siliculosus* (CBN78148.1), *Phaeodactylum tricornutum* (EEC42834.1), *Aphanomyces astaci* (ETV81397.1), *Aphanomyces invadans* (ETV95487.1), *Phytophthora parasitica* (ETO77406.1), *Saprolegnia diclina* (EQC41869.1), *Hammondia hammondi* (KEP66655.1), *Toxoplasma gondii* (EPR64867.1), *Tetrahymena thermophila* (EAR94453.1), *Ichthyophthirius multifiliis* (EGR28782.1), *Oxytricha trifallax* (EJY82089.1), *Paramecium tetraurelia* (CAK85009.1), *Acanthamoeba castellanii* (ELR15398.1), *Dictyostelium fasciculatum* (EGG24313.1), *Entamoeba invadans* (BAN40624.1), *Arabis alpina* (KFK32725.1).

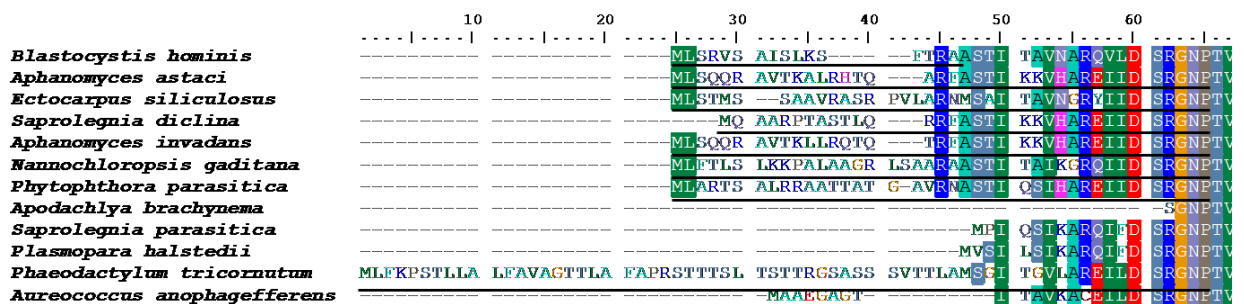


Figure 3.4 Sequence alignment of the amino terminal region of enolases from stramenopiles. The underlined sequence represents the predicted targeting signal of *Blastocystis* sp. Accession numbers from top to bottom; *Blastocystis* (KNB43609.1), *Aphanomyces astaci* (ETV81397.1), *Ectocarpus siliculosus* (CBJ32586.1), *Saprolegnia diclina* (EQC41869.1), *Aphanomyces invadans* (ETV95487.1), *Nannochloropsis gaditana* (EWM28602.1), *Phytophthora parasitica* (ETO77406.1), *Albugo candida* (CCI42145.1), *Thalassiosira pseudonana* (EED92063.1), *Phaeodactylum tricornutum* (EEC42834.1), *Aureococcus anophagefferens* (EGB09778.1).

These finding that enolase is targeting the mitochondria is quite interesting and is in accordance with Liaud *et al* (2000). In their study they demonstrated that some glycolytic proteins of stramenopiles are targeting the mitochondria, in addition to Takishita *et al* work that mentioned the same hypothesis ([Liaud, Lichtle et al. 2000](#), [Nakayama, Ishida et al. 2013](#)). Another interesting remark in this project is the finding that *Blastocystis* enolase targeting signal was significantly shorter (16 amino acid) than other mitochondrial targeting signals detected. The same remark was noticed in stramenopiles' putative mitochondrial targeting sequences, though not as short as *Blastocystis* enolase targeting signal sequence.

Previous studies have mentioned the short mitochondrial targeting signal such as those found on the *Trichomonas vaginalis* hydrogenosomal proteins which ranged between 5-14 amino acids ([Bradley, Lahti et al. 1997](#)). Those studies have indicated to the possibility that these short signals might need further assistance by other internal signals ([Burri and Keeling 2007](#)). According to Keeling and Burri (2007) pre-sequences are generally larger in yeast and mammals and can be as big as 20-80 amino acids, while they were reported to be shorter in parasites such as *T. vaginalis* (5-14 amino acids) and (20-30 amino acids) in fungal hydrogenosomes such as *Neocallimastix frontalis* ([vanderGiezen, Rechanger et al. 1997](#)).

Table 3. 1 Possible mitochondrial targeting signal on enolase sequences from several organisms. Prediction performed using TargetP (<http://ipsort.hgc.jp/>), Mitoprot ([Claros 1995](#)), iSort ([Nielsen, Engelbrecht et al. 1996](#), [Emanuelsson, Nielsen et al. 2000](#)) and Predator (<https://urgi.versailles.inra.fr/predotar/predotar.html>).

Organism	TargetP	MitoProt probability/ targeting signal	iSort	Predotar
<i>Blastocystis hominis</i>	0.831	0.9829 MLSRVSAISLKSFTRA	MLSRV SAISL KSFTR AASTI TAVNA RQVLD	0.82
<i>Aphanomyces astaci</i>	0.823	0.9849 MLSQQRAVTKALRHTQARF ASTIKKVHAREIIDSRGNP	MLSQQ RAVTK ALRHT QARFA STIKK VHARE	0,50
<i>Ectocarpus silicicolocus</i>	0.908	0.9871 MLSTMSSAAVRASRPVLAR NMSAITAVNGRYIIDSRGNP	MLSTMSSAAVRASRPVL ARNMSAITAVNGR	0.57
<i>Saprolegnia diclina</i>	0.791	0.9913 MQAARPTASTLQRRFASTIK KVHAREIIDSRGNP	MQAAR PTAST LQRRF ASTIK KVHAR EIIDS	0,16
<i>Aphanomyces invadans</i>	0.828	0.9930 MLSQQRAVTKLLRQTQTRF ASTIKKVHAREIIDSRGNP	MLSQQ RAVTK LLRQT QTRFA STIKK VHARE	0,57
<i>Nannochloropsis gaditana</i>	0.912	0.9615 MLFTLSLKKPALAAGRLSAA RAASTITAIKGRQIIDSRGNP	MLFTL SLKKP ALAAG RLSAA RAAST ITAIK	0,94
<i>Phytophthora parasitica</i>	0.922	0.9970 MLARTSALRRAATTATGAVR NASTIQSIHAREIIDSRGNP	MLART SALRR AATTA TGAVR NASTI QSIHA	0,71
<i>Plasmopara halstedii</i>	0.246	0.1958 MVSILSIKARQIFDSRGNP	No signal or mts	0,37
<i>Phaeodactylum tricornutum</i>	mTP 0.287 SP 0.777	0.6506 MLFKPSTLLALFAVAGTTLAF APRSTTTSLTSTTRGSASSS VTTLAMSGITGVLAREILDSR GNP	Signal peptide, not mitochondrial MLFKP STLLA LFAVA GTTLA FAPRS TTTSL	0.99 ER 0.21 mit
<i>Aureococcus anophagefferens</i>	0.082	Not predictable 0.1038	No	No mit No ER

3.2.2. *Blastocystis* enolase signature motif and amino acid discrepancies analysis

A BLAST search of the Pubmed database with the deduced amino acid sequence of *Blastocystis* enolase revealed a strong identity with the homologous genes of several organisms representing the tree of life. The organisms and their accession number according to GeneBank database were as follows; *Trypanosoma rangeli* (ESL05219.1), *Aphanomyces astaci* (XP_009829255.1), *Archaeon* (KHO45292.1d), *Entamoeba histolytica* (XP_649161.1), *Echinostoma caproni* (ABI26619.1), *Clostridium* (WP_010294244.1), *Homo sapiens* (BAD96912.1), *Malus domestica* (XP_008391739.1), *Rhizopus delemar* (EIE87278.1), *Leishmania mexicana* (XP_003873490.1). Functional domain and key sites were estimated in two ways, either by comparing the amino acid sequence with already studied protein of different organisms, or by using automated servers that predicted the function based on the amino acid sequence of the query protein as well, like the new tool offered by embi-ebi, InterProScan (<http://www.ebi.ac.uk/InterProScan/>).

By analysing the aligned sequences it was noticed that the primary sequence revealed the conservation of several structural features of enolase. Most importantly, *Blastocystis* enolase sequence included the four potential Mg²⁺ binding sites, which were in *Blastocystis* enolase as follows: Ser 57, Asp 259, Glu 306, Asp 333. Those residues were indicated to by black arrowheads in Figure 3.5. Mg²⁺ is a divalent cation important for activating the enzyme, and works as its natural cofactor. Residues for potential substrate binding site which are crucial for enzyme activity were all located and indicated in the same figure with black star at (His 172, Glu 224, Lys 358, His 386, Arg 387, Ser 388). Those four metal

binding residues and the six substrate binding residues are important for enzyme function and dimerization of enolase, and were reported in other studies ([Gan, Zhao et al. 2010](#), [Carabarin-Lima, Rodriguez-Morales et al. 2014](#)). This comparison reveals that residues responsible for both substrate and metal binding are invariably present in *Blastocystis* enolase sequence and other sequences as well.

Moreover, enolase signature was predicted at the C-terminus amino acid sequence covering the region 355-368 (LLLKVNQIGSVTES) as predicted by PROSITE for motif scan detection tool. Enolase signature was underlined in Fig 3.5. In *Blastocystis* enolase deduced sequence, a characteristic enolase family fingerprint motif was also detected at 385-393 (SHRSGETED) (located at the C-terminal of the protein sequence), and this was clearly highly conserved in the organisms studied in this research and in other organisms studied to date ([Quinones, Pena et al. 2007](#)). A plasminogen-binding site which was previously detected in *Leishmania mexicana* was not conserved in *Blastocystis* enolase sequence (not included in this search), while it appeared in *Trypanosoma rangeli* sequence included in this project search (TYDENKKQY). This indicates that no plasminogen-binding activity is expected in *Blastocystis* enolase. *Blastocystis* enolase amino acid sequence was submitted to Uniprot server and the annotations revealed by its Blast search are summarized in Table 3.2. <http://www.uniprot.org/blast/>.

Table 3. 2. Special features in *Blastocystis* enolase sequence as annotated by Uniprot server, <http://www.uniprot.org/blast/>. The table shows the important regions in the protein sequence for its activity and classification.

Feature key	Position(s)	Length	Residue	Description
Function				
Binding site	171-171	1	K	Substrate
Binding site	181-181	1	E	Substrate
Active site	224-224	1	E	Proton donor
Binding site	306-306	1	E	Substrate
Binding site	333-333	1	D	Substrate
Active site	358-358	1	K	Proton acceptor
Binding site	409-409	1	K	Substrate
Family and domains				
Region	385-388	4	SHRS	Substrate binding

	10	20	30	40	50

<i>Blastocystis</i>	MLS--RLSTT	SFK-ALTRAA	S ITAV ART	LD SRGNPTV	EV TTQDG-
<i>Trypanosoma rangeli</i>	-----	-----	I KVHGRE	LD SRGNPTV	EVE TTELG-
<i>Aphanomyces astaci</i>	MLSQQRAVTK	ALRHTQARFA	S I KVHARE	LD SRGNPTV	EV TTEQG-
<i>archaeon</i>	-----	-----M	AKI SVHARE	LD SRGNPTV	EVE T TDG-
<i>Entamoeba histolytica</i>	-----	-----	I KVHARE	LD SRGNPTV	EVE TTGKG-
<i>Echinostoma caproni</i>	-----	-----	AI S HAR	LD SRGNPTV	EV TTAKG-
<i>Clostridium</i>	-----	-----MKQY	EIIDV AR	LD SRAFPTV	EVE TLDDGT
<i>Homo sapiens</i>	-----	-----	ILK HARE	LD SRGNPTV	EV FTSKG-
<i>Malus domestica</i>	-----	-----M	A I SVKAR	LD SRGNPTV	EV ILSDGT
<i>Rhizopus delemar</i>	-----	-----M	S TKVHAR	LD SRGNPTV	EVE TTAKG-
<i>Leishmania mexicana</i>	-----	-----	PI KVHARE	LD SRGNPTV	EVE MTEAG-

	60	70	80	90	100

<i>Blastocystis</i>	TFRA VPSCA	STG HEALEL	RDG GKA-YNG	KCVTKAV NV	KTITGPH CG
<i>Trypanosoma rangeli</i>	FR IVPSGA	STG HEACEL	RDD KRRY G	KGC HAVKNV	ND APAL G
<i>Aphanomyces astaci</i>	FRA VPSCA	STG HEA EL	RDG GKR-Y G	KGV AV NV	KNII GPKLEG
<i>archaeon</i>	YVVMVPSCA	STG HEALEL	RDG GKR-YNG	GV KAVKN	NTI IAPL KG
<i>Entamoeba histolytica</i>	FR CVPSGA	STG HEA EL	RDG KCRYG	KCV KAV NV	NTI IGPAL G
<i>Echinostoma caproni</i>	FRAAVPSGA	STG HEALEL	RDG PPG-Y G	KGV KAVSNV	NQI IAPGL K
<i>Clostridium</i>	GRAAVPSGA	STG FEA EL	RDG KAVYNG	KGV KAV NV	NET IAEEL G
<i>Homo sapiens</i>	FRAAVPSGA	STG EALEL	RDN KTRY G	KGVSKAV	NKT IAPAL S
<i>Malus domestica</i>	ARAAVPSGA	STG EALEL	RDG GKD-Y G	KGVSKAV NV	NTI IGPAL G
<i>Rhizopus delemar</i>	FRAAVPSGA	STG HEALEL	RDG KSQY G	KCVTKAV NV	NS IAPAL E
<i>Leishmania mexicana</i>	FR AVPSGA	STG HEACEL	RDG KARYCG	AGC AVKNV	NET IAPAL G

	110	120	130	140	150

<i>Blastocystis</i>	MD -- IYA	ID LMF -LD	GTKN-----K	GKLGANAILG	TSLACVKAGA
<i>Trypanosoma rangeli</i>	KDE--LQOAT	D LMRD-LD	GTAN-----K	S LGANAILG	CS A KAAA
<i>Aphanomyces astaci</i>	MDP-- QSE	ID LLMK -LD	GTDN-----K	GNLGANAILG	VSLAVAKAGA
<i>archaeon</i>	KDP--A OKS	ID LM -KD	GTKN-----K	SKLGANAILG	VSLAVC AGA
<i>Entamoeba histolytica</i>	K --L QAE	D M -LD	GTNN-----K	GKLGANAILG	CS C AAA
<i>Echinostoma caproni</i>	CG PV QAG	IDNFM -LD	GTPN-----K	EKLGANAILG	VSLAVKAGA
<i>Clostridium</i>	M --F QVA	ID LM -LD	GTKN-----K	GKLGANA LG	VSLACAKAAA
<i>Homo sapiens</i>	KK NV QEK	ID LM -D	GTEN-----K	SKFCANAILG	VSLAVCKAGA
<i>Malus domestica</i>	KDP-- QTA	IDNFM QLD	GTVNEWGWCK	QKLGANAILA	VSLAVAKAGA
<i>Rhizopus delemar</i>	SK PV QKA	ID F -LD	GTPN-----K	GKLGANAILG	VSLAVAKAAA
<i>Leishmania mexicana</i>	KDE-- QAG	D MC -LD	GTKN-----K	SKLGANAILG	CS A KAAA

	160	170	180	190	200

<i>Blastocystis</i>	AA K PLY	FA AGKD -	FVLPVP MNV	INGGKHAGNA	LAFOEFM P
<i>Trypanosoma rangeli</i>	ARKG PLY	A LAG K -	RLPVPCFNV	INGGKHAGNA	LPFOEFMIAP
<i>Aphanomyces astaci</i>	RAKS PL	FA L G D -	VLPVP FNV	INGG HAGN	LAFOEFMI P
<i>archaeon</i>	LA DKPLYL	IAGLAEPFR-	FVLPVPFFNV	INGGRHAGN	ID QE MIAP
<i>Entamoeba histolytica</i>	AEKG PLY	IA LTGHK -	T VPVPCFNV	INGGAHAGNA	LA QEFMICP
<i>Echinostoma caproni</i>	AEKG PLY	IASLAG N -	VLPVP FNV	INGG HAGN	LA QEFMI P
<i>Clostridium</i>	ESLG GLY	IGG N--A -	-VLPVPMN	INGGKHADNN	D QEFMI P
<i>Homo sapiens</i>	VEKG PLY	IA LAG S -	LPVPAFNV	INGG HAGN	LA QEFMI P
<i>Malus domestica</i>	VKK PLY	IA LAG KN-	VLPVPAFNV	INGG HAGN	LA QEFMI P
<i>Rhizopus delemar</i>	GDKG PLYV	FA LAG K P	FVLPVPAFNV	INGG HAGN	LA QEFMI P
<i>Leishmania mexicana</i>	AKAG PLY	IAALAGTK -	RLPVPCFNV	INGGKHAGNV	LPFOEFMIAP

	210	220	230	240	250

<i>Blastocystis</i>	CAKSF EAM	GCEVYHSL	K VIK YGL	DATNVGDEGG	FAP I SNIE
<i>Trypanosoma rangeli</i>	KASSF EA	RMG EVYHAL	K IK KYG	DAVNVGDEGG	FAPPITDINE
<i>Aphanomyces astaci</i>	TGAESF EAM	VMGCEVYHHL	KGVIK KYG	DATNVGDEGG	FAPNI SNHE
<i>archaeon</i>	GAK FAEA	RMGAETYHTL	KG I KYG	AINVGDEGG	FAPP DHIEE
<i>Entamoeba histolytica</i>	TGAT FHEA	RMAAETYQCL	KVVIKAKYG	DATNVGDEGG	FAPN SGARE
<i>Echinostoma caproni</i>	TGASSFAEAM	G EVYHHL	KAVIKGKYGL	DACNVGDEGG	FAPNI DNLE
<i>Clostridium</i>	GAKSF EA	RMCAEVYHTL	K K KG--	LGTGVGDEGG	FAPN ASNEE
<i>Homo sapiens</i>	CAA FREAM	RGAEVYHNL	K VIK YG	DATNVGDEGG	FAPNI LENKE
<i>Malus domestica</i>	GASSFKEAM	MGEVYHNL	K VIK KYG	DATNVGDEGG	FAPNI ENRE
<i>Rhizopus delemar</i>	TGAKSFKEAM	G EVYHNL	VIKAKYGL	DATNVGDEGG	FAPNI KNEE
<i>Leishmania mexicana</i>	TKATSFREA	RMG EVYHAL	KV IKNKYG	DAVNVG CG	FAPPI HIDE

260 270 280 290 300

Blastocystis GL LSEATE KAGYAK-- RTAMDCAASE FVT- DGYVD IMFKQGG-GM
Trypanosoma rangeli PIP L EATE AG KG--KF AICMDC AASE T DENKKQY LTFK PN---
Aphanomyces astaci G LL LAITE KAGYTG--K GIGMDVA SE FVT- SKHYD LNFK PG--S
archaeon PL LL EAI KAG AG--K KIAMDAAASE FVVE KKYD LQK VD---
Entamoeba histolytica AL LL EATA KAGYTG--K EIAMDC AASE FVNE TKKYD LGKKI PADKK
Echinostoma caproni GL LLRTAIE KAGYTG--K KIAMD SAASE FVK- GK-YD LDFK PK--S
Clostridium A EATE KAGYKAGE F A DPA SE F E--DGKY LSG GRI---
Homo sapiens GL LLKTAIG KAGYTD--K VIGMDVAASE F R-SGK-YD LDFK P---D
Malus domestica GL LLKTAIE KAGYTG--K VIGMDVAASE FVG-SDKTYD LNFKEEK--N
Rhizopus delemar GL LL AIE KAGYTG--K KIGMDCAASE FVK- GK-YD LDFK PN--S
Leishmania mexicana PIP L EATE KAG KN--KF AICMDC AASE AADA RKMV LTFK PE---

310 320 330 340 350

Blastocystis KK---- K T YIKL K YPIVSIEDP F QDDWAA A L REK GKD
Trypanosoma rangeli --ATW AK LG TYAK YPIVSIEDP DDD EG A G TEV KGKA
Aphanomyces astaci KEAP- G LGQ Y DLAK PI SIEDP FDQDDWEH S AFTAA GTD
archaeon -----SL LA Y YL K TYPI SIEDP FS DDWE Y V TKS GKK
Entamoeba histolytica DP LVKDV G LIAEYVD GK HYPIASIEDP FA DDWAA N KFT-VEHGNF
Echinostoma caproni PA SW S A A Y KM TYPIVSIEDP FDQDDWPA T K TGECK--
Clostridium ----- P A YVELA YPI SIEDG MA DWEG K H TDK GDK
Homo sapiens DP RY P LA Y S K YPIVSIEDP FDQDDWGA Q KFTASAG--
Malus domestica DG QK GNA LK Y S A YPIVSIEDP FDQDDWEH A K TAE GEK
Rhizopus delemar DQ KW DGK LTA YSS A YPIVSIEDA FDQDDWEA T H NETSD--F
Leishmania mexicana --PTY AA LQATY R YP VSIEDP FA D E S A TKA AGRA

360 370 380 390 400

Blastocystis QIVGDDLLVT NPAR AEAI KKACNALLLK VNQIG VTES QACLD AA
Trypanosoma rangeli QIVGDDLLVT NPAR KMAI KKACN LLLK VNQIGVTEA IDASK CMAN
Aphanomyces astaci QIVGDDLLCT NPAR AKA G IKACNALLLK VNQIGVTES IQAVKDA AA
archaeon QIVGDDLLVT NPAR QRAIT K ATALLLK VNQIGVTES QAAA C GH
Entamoeba histolytica QIVGDDLLVT NPAR QMA KNACN L K VNQIGTTE F TIK A EK
Echinostoma caproni QIVGDDLLVT NPAR QKAI KACNLLLK VNQIG VTES IQACK A A
Clostridium QIVGDDLFVT NT R AMGF K VAN L K VNQIGTTE INAI ASRK
Homo sapiens QIVGDDLLVT NP R AKA N K CNLLLK VNQIG VTES QACK A AN
Malus domestica QIVGDDLLVT NP R EKAI K CNALLLK VNQIG VTES IAV KA
Rhizopus delemar QIVGDDLLVT NP R ATAI KK CNALLLK VNQIGT TES IQAAKD AA
Leishmania mexicana QIVGDDLLVT NV R RMAI KSACN LLLK VNQIGT ES IAAAK CMEN

410 420 440 450

Blastocystis GWGVMVSHRS GETEDN IAD LAVGLC GQI KTGAPCRSER IAKYNQLLRI
Trypanosoma rangeli GWSVMVSHRS GETED IAD LVVGLG GQI KTGAPCRGER TAKLNQLLRI
Aphanomyces astaci GWGVMVSHRS GETED IAD LAVGLSTGQI KTGAPCRSER IAKYNQLLRI
archaeon GWNVMVSHRS GETED IAD LVVGLGCGQI KTGAPCRSER IAKYNQLLRI
Entamoeba histolytica GWGVMVSHRS GETED IAD LVVGLNCKQI KTGAPCRSER ICKYNQLLRI
Echinostoma caproni GWGVMVSHRS GETEDN IAD LVVGLRTGQI KTGAPCRSER IAKYNQLLRI
Clostridium G TA VSHRS GETED ITIAD LVVA NAGQI KTGAPARSER IAKYNQLLRI
Homo sapiens GWGVMVSHRS GETED IAD LVVGLCTGQI KTGAPCRSER IAKYNQLLRI
Malus domestica GWGVMVSHRS GETED IAD LAVGLATGQI KTGAPCRSER IAKYNQLLRI
Rhizopus delemar GWGVMVSHRS GETED ITIAD LVVGLRTGQI KTGAPCRSER IAKYNQLLRI
Leishmania mexicana GWSVMVSHRS GETED IAD LSVGLGTGQI KTGAPCRGER TAKLNQLLRI

470

Blastocystis EESLGSKAVY AGLN-FRTPD RC-
Trypanosoma rangeli EEELG---A AKFG-FPAWA ---
Aphanomyces astaci EEMLGRSAVY AGK -FRNPH EA-
archaeon EE L GK KAVY GCKR-----
Entamoeba histolytica EEELGN-IPY AGKN- RNST A--
Echinostoma caproni EE L GSAAKY AGEN-FRRP- ---
Clostridium EEELGDMAEY RCMNSFYNIK R--
Homo sapiens EEELGSKAK AGRN-FRNPL AK-
Malus domestica EEELGAEAVY AGAK-FRVPV EPY
Rhizopus delemar EEELGENCIY AGEN-FRNAH EL-
Leishmania mexicana EEE G---SA AAYG-FPGWA ---

Figure 3. 5 Alignment of the amino acid sequences of the enolases. *Trypanosoma rangeli* (ESL05219.1), *Aphanomyces astaci* (XP_009829255.1), *Archaeon* (KHO45292.1d), *Entamoeba histolytica* (XP_649161.1), *Echinostoma caproni* (ABI26619.1), *Clostridium* (WP_010294244.1), *Homo sapiens* (BAD96912.1), *Malus domestica* (XP_008391739.1), *Rhizopus delemar* (EIE87278.1), *Leishmania mexicana* (XP_003873490.1). Brackets show GeneBank accession numbers. The dashed lines demonstrate gaps introduced between sequences. Residues indicating Mg²⁺ binding sites are shown in black arrowheads, residues forming the enolase signature motif are underlined, while residues motif for substrate binding are indicated with a black star. The enolase fingerprint motif is boxed. Coloured amino acids indicate conservative amino acid regions. The letters below the alignment mark residues involved in binding to the phosphate (P), carboxyl (C) group of substrate, and in metal binding (M).

3.3. Homology modelling

Function of proteins can be inferred in most of the times from the protein structure, although the relationship between function and fold is complicated. Some cases one fold of protein can perform several functions, while in other cases; it is found that one function can be achieved by several folds. The complexity of fold-function relation necessitates different approaches. This gave way to inspecting special sites, motifs, signatures, and other conserved residues ([Jones and Thornton 2004](#)). Homology modelling is a method used to predict the three dimensional structure of a protein of known sequence by modelling it on a known 3D structure of a related protein sequence (template sequence). This will give an insight of a probable function of the studied protein in light with an already assigned structure. Finding a homologous protein is only feasible with templates of 30% or more similarity. Modelling the protein will allow us to assess the function of the protein and its binding sites of study in comparison to the known structure of the template.

Structures of enolases from different species were available from protein data bank (PDB) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) seeking potential template-homologues proteins for model construction. The template protein used to build a 3D structure model for enolase was chosen as described in 2.2.3.10.8.1, a

multiple sequence alignment between the target and the template protein was run using pair-wise sequence alignment tool which allows manipulation of the sequences parameters such as gaps, which will help avoiding errors.

The model built using the template chosen (3UCD) and the SwissModel server (<http://swissmodel.expasy.org/interactive>) came with the model protein shown in Fig 3.6 as a dimer. The structure was built as described in both 2.2.3.10.8.2 and 2.2.3.10.8.3. The chosen model was human neuron specific gamma enolase (hNSE) (PDB 3UCD) ([Qin, Chai et al. 2012](#)) which shares 64.24% identity with *Blastocystis* enolase protein sequence and belongs to *Homo sapiens*. The quality of the model was checked by superimposing the model protein with the template chosen using PyMol program and inspecting the active sites manually and visually. The query protein and the template both have a conserved active sites and binding sites. Both proteins (the model and the template) have virtually the same overall topology with nearly identical secondary structural elements (Figure 3.7). The overall structure of the model was found to be highly conserved when compared to the template 3UCD in regards to the enolase signature motif, enolase family motif, Figure 3.8, and figure 3.9. tow loops were found to take different position between the template and the model. Those loops were covering the region 259-266 in each subunit, see Figure 3.10.

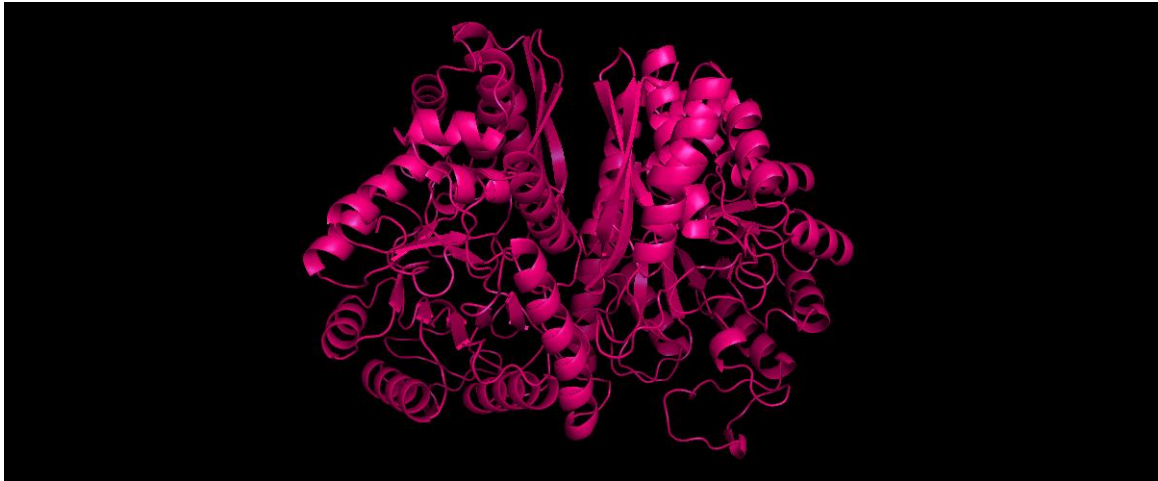


Figure 3. 6 Cartoon representation of the modelled overall structure of *Blastocystis* enolase, shown as a dimer. The model was built using the neuron human enolase protein template (hNSE) (3UCD) as a template. Molecule is seen looking down the twofold axis. A greater portion of the polypeptide backbone adopts α -helical conformation. The two monomers are shown in hot pink. The image was generated using PyMOL (<http://www.pymol.org>).



Figure 3. 7 *Blastocystis* enolase model aligned with the template human neuron specific enolase (hNSE) (PDB 3UCD), showing almost identical overall topology between the two proteins. The image was created using PyMOL.

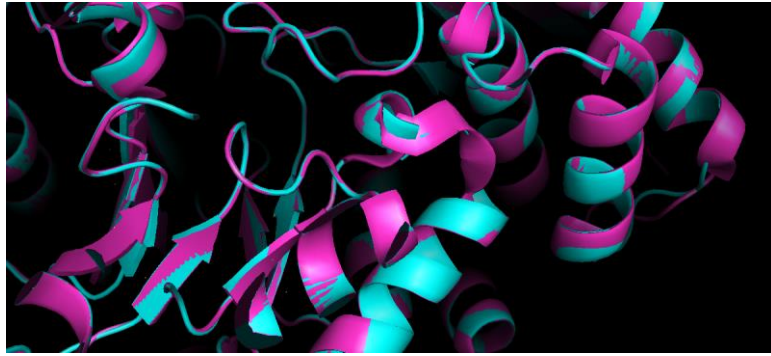


Figure 3. 8 A close up in *Blastocystis* enolase model structure built using human neuron specific enolase (hNSE) (PDB 3UCD) as a template showing the highly conserved enolase signature motif covering the amino acid region (355-368) (LLLKVNQIGSVTES). This was made using PyMol.

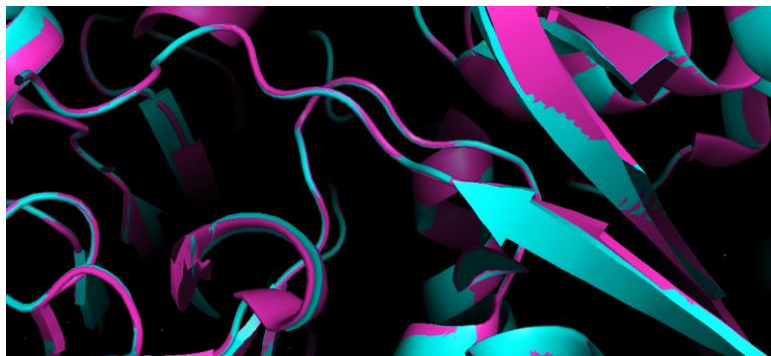


Figure 3. 9 A close up in *Blastocystis* enolase model structure built using human neuron specific enolase (hNSE) (PDB 3UCD) as a template. It shows the highly conserved enolase finger print region covering amino acids (385-393) (SHRSGETED). This was made using PyMol.

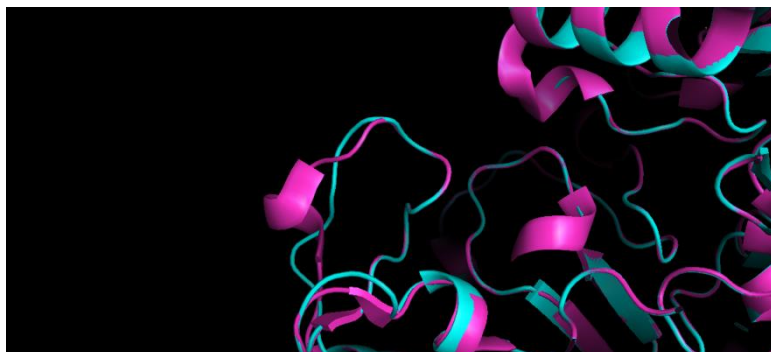


Figure 3. 10 A close up in *Blastocystis* enolase model structure built using human neuron specific enolase (hNSE) (PDB 3UCD) as a template. It shows a loop converted from the template orientation, covering amino acids (259-266), an area which was responsible for asymmetry between the unit interfaces in (hNSE) structure. This was made using PyMol.

3.4. Phylogentic relationships of *Blastocystis* enolase and other enolases

In order to confirm the previous results of the mitochondrial targeting nature of *Blastocystis* enolase phylogenetic analyses including extended sampling were conducted. Sequences previously used in signature motif and mitochondrial targeting signal analyses were analysed by phylogenetic reconstructions.

3.4.1. Sequence alignment and phylogenetic analysis.

Two sets of enolase sequence data were constructed. The first set comprised a large sample of protein sequences (87) that included enolase protein sequences from all super-groups, the second set of sequences was only an expanded stramen opiles searches. A detailed list of organisms used to execute this phylogenetic analysis can be found in Table 3. 3.

Table 3. 3 Taxa used in large-scale *Blastocystis* enolase protein phylogeny including accession numbers, and the super-groups.

Subgroup	Name of organism	Accession number
Stramenopiles		
	<i>Blastocystis hominis</i>	(XP_012894576.1)
	<i>Aphanomyces astaci</i>	(XP_009829255.1)
	<i>Ectocarpus siliculosus</i>	(CBJ32586.1)
	<i>Saprolegnia diclina</i>	(XP_008604438.1)
	<i>Phytophthora parasitica</i>	(ETO77406.1)
	<i>Apodachlya brachynema</i>	(AAR97547.1)
	<i>Saprolegnia parasitica</i>	(XP_012196295.1)
	<i>Plasmopara halstedii</i>	(CEG45103.1)

	<i>Phaeodactylum tricornutum</i>	(XP_002185511.1)
	<i>Aureococcus anophagefferens</i>	(XP_009035814.1)
	<i>Aphanomyces invadans</i>	(XP_008875680.1)
	<i>Nannochloropsis gaditana</i>	(EWM28602.1)
Amoebozoa		
	<i>Acanthamoeba castellanii</i>	(XP_004337411.1)
	<i>Dictyostelium fasciculatum</i>	(XP_004362164.1)
	<i>Polysphondylium pallidum</i>	(EFA79459.1)
	<i>Entamoeba invadens</i>	(BAN40624.1)
Plantaea		
Chromophytes	<i>Arabis alpine</i>	(KFK32725.1)
	<i>Phaseolus vulgaris</i>	(XP_007162849.1)
Algae	<i>Chondrus crispus</i>	(XP_005716244.1)
	<i>Galdieria sulphuraria</i>	(XP_005706995.1)
	<i>Cyanidioschyzon merolae</i>	(XP_005536545.1)
	<i>Prionitis lanceolata</i>	(AAL05461.1)
	<i>Mastocarpus papillatus</i>	(AAL05459.1)
Plantae	<i>Solanum pennellii</i>	(XP_015088120.1)
	<i>Nicotiana tabacum</i>	(ACR56690.1)
	<i>Selaginella moellendorffii</i>	(XP_002984551.1)
	<i>Glycine max</i>	(KRG96124.1)
	<i>Oryza sativa</i>	(NP_001049556.1)
	<i>Phaseolus vulgaris</i>	(XP_007162849.1)
The Opisthokonta		

Animalia	<i>Clonorchis sinensis</i>	(GAA51601.1)
	<i>Echinostoma caproni</i>	(ABI26619.1)
	<i>Fasciola hepatica</i>	(CAK47550.1)
	<i>Lingula anatine</i>	(XP_013386675.1)
	<i>Taenia solium</i>	(AHB59732.1)
	<i>Hydra vulgaris</i>	(XP_012564321.1)
	<i>Echinococcus granulosus</i>	(ACY30465.1)
	<i>Ictalurus punctatus</i>	(NP_001187631.1)
	<i>homo sapiens</i>	(BAD96912.1)
Choanoflagellata	<i>Salpingoeca rosetta</i>	(XP_004994770.1)
Fungi	<i>Rhizopus delemar</i>	(EIE87278.1)
	<i>Mucor circinelloides</i>	(EPB85979.1)
	<i>Parasitella parasitica</i>	(CEP19658.1)
	<i>Lichtheimia ramosa</i>	(CDS14531.1)
	<i>Conidiobolus lamprauges</i>	(ALX18691.1)
	<i>Allomyces macrogynus</i>	(KNE63834.1)
	<i>Blastocladiella emersonii</i>	(ABK34894.1)
	<i>Neocallimastix frontalis</i>	(AFJ73493.1)
	<i>Mortierella verticillata</i>	(KFH71943.1)
	<i>Rozella allomycis</i>	(EPZ33681.1)
Opisthokonta	<i>Opisthorchis viverrini</i>	(XP_009174927.1)
	<i>Rhizopus delemar</i>	(EIE80761.1)
	<i>Echinostoma caproni</i>	(CAK47551.3)
	<i>Lichtheimia corymbifera</i>	(CDH56669.1)

The Excavata		
Diplomonads	<i>Spironucleus salmonicida</i>	(EST42221.1)
	<i>Giardia lamblia</i>	(EFO60949.1)
	<i>Hexamita inflata</i>	(AAD45328.1)
	<i>Hexamita inflata</i>	(AAD45329.1)
Heterolopbosea	<i>Naegleria gruberi</i>	(XP_002680773.1)
Kinetoplastides	<i>Trypanosoma congolense</i>	(CCC93490.1)
	<i>Bodo saltans</i>	(CUG84387.1)
	<i>Trypanosoma cruzi</i>	(EKF32804.1)
	<i>Angomonas deanei</i>	(EPY25410.1)
	<i>Strigomonas culicis</i>	(EPY28299.1)
	<i>Leptomonas pyrrocoris</i>	(KPA77322.1)
	<i>Leishmania panamensis</i>	(XP_010697345.1)
Parabasalids	<i>Trichomonas vaginalis</i>	(XP_001322282.1)
	<i>Pseudotriconympha grassii</i>	(BAF48677.1)
	<i>Trichomonas vaginalis</i>	(XP_001315627.1)
Chromalveolates		
Apecomplexa	<i>Plasmodium falciparum</i>	(XP_001347440.1)
	<i>Gregarina niphandrodes</i>	(XP_011131468.1)
	<i>Babesia bigemina</i>	(XP_012768952.1)
	<i>Theileria annulata</i>	Ankara(XP_953341.1)
	<i>Babesia bovis</i>	(XP_001611923.1)
	<i>Hammondia hammondi</i>	(XP_008882278.1)
	<i>Toxoplasma gondii</i>	(XP_002365578.1)

	<i>Eimeria tenella</i>	(XP_013229351.1)
	<i>Toxoplasma gondii</i>	(XP_002365579.1)
	<i>Cryptosporidium parvum Iowa</i>	(XP_626138.1)
Ciliata	<i>Paramecium tetraurelia</i>	(XP_001452406.1)
	<i>Tetrahymena thermophila</i>	(XP_001014643.1)
	<i>Ichthyophthirius multifiliis</i>	(XP_004030018.1)
	<i>Oxytricha trifallax</i>	(EJY82089.1)
Diatoms	<i>Thalassiosira pseudonana</i>	(XP_002290311.1)
Oomycetes	<i>Aphanomyces astaci</i>	(XP_009829255.1)
Plants and higher plants	<i>Solanum pennellii</i>	(XP_015088120.1)
Phayophytes	<i>Aureococcus anophagefferens</i>	(XP_009035814.1)
	<i>Ectocarpus siliculosus</i>	(CBJ32586.1)

Initially, database searches (BLASTp) over a wide taxonomic spectrum were performed using *Blastocystis* enolase protein sequence. *Blastocystis* enolase was aligned with other enolase sequences retrieved from NCBI data bank. The sequences were picked, aligned and manually refined by eye then aligned using MUSCLE ([Edgar 2004](#)). MUSCLE will detect any sequence repetitions and gaps and will remove them automatically. Regions of gaps or those that are not clearly alignable in all sequences were excluded from the phylogenetic analysis by masking the alignment manually. A phylogenetic tree is then constructed using those edited sequences.

The enolase phylogenies were calculated using the maximum likelihood (ML) approaches with the PhyML programme using (phylogeny.fr) web server tools ([Castresana 2000](#), [Chevenet, Brun et al. 2006](#), [Dereeper, Guignon et al. 2008](#), [Dereeper, Audic et al. 2010](#)). *Blastocystis* enolase phylogeny was additionally analysed using the Bayesian analysis program MrBayes ([Huelsenbeck and Ronquist 2001](#), [Ronquist, Teslenko et al. 2012](#)) on the same web server (phylogeny.fr). Bootstrap re-sampling was performed on ML trees with 100 replications for all data sets. Phylogenetic reconstruction of Blastocystis enolase using MrBayes were run using a mixed amino acid. One hundred thousand generations were calculated and trees sampled every 100 generation. The model stabilised rapidly and 250 trees were discarded as burnin.

3.4.1.1. *Placing of Blastocystis enolase in the kingdom of life*

The large scale phylogenetic reconstruction reveals a wide diversity within the enolase family and a general distribution within tight systematic groups. The tree generated from amino acid sequences of enolase (see Figure 3.11), clearly showed that *Blastocystis* enolase is clustering with other members of chromalveolates and particularly stramenopiles forming a monophylum with a high bootstrap value of 95%. It showed a group that clusters with organisms of high mitochondrial targeting signal of over 95% probability to target the mitochondria. Other organisms have also grouped together with high bootstrap value such as Plantae (98%), and Opisthokonta (89%). Another interesting value was the notice of a weakly supported relationship between Amoebozoa and all of the following (Plantae, Stramenopile and Opisthokonta), while those themselves

(Plantae, Stramenopile and Opisthokonta) showed a strongly supported relationship demonstrated by a bootstrap value of 87% between themselves. A far relationship was found between Bacteria and Chromalveolates, although strongly supported with a root bootstrap of 100% and showed that *Blastocystis* enolase is evolved after bacteria. This might explain the relationship between the mitochondrial targeting sequence found in *Blastocystis* and other members of Stramenopiles, and the fact that this mitochondria might be of a bacterial ancestor.

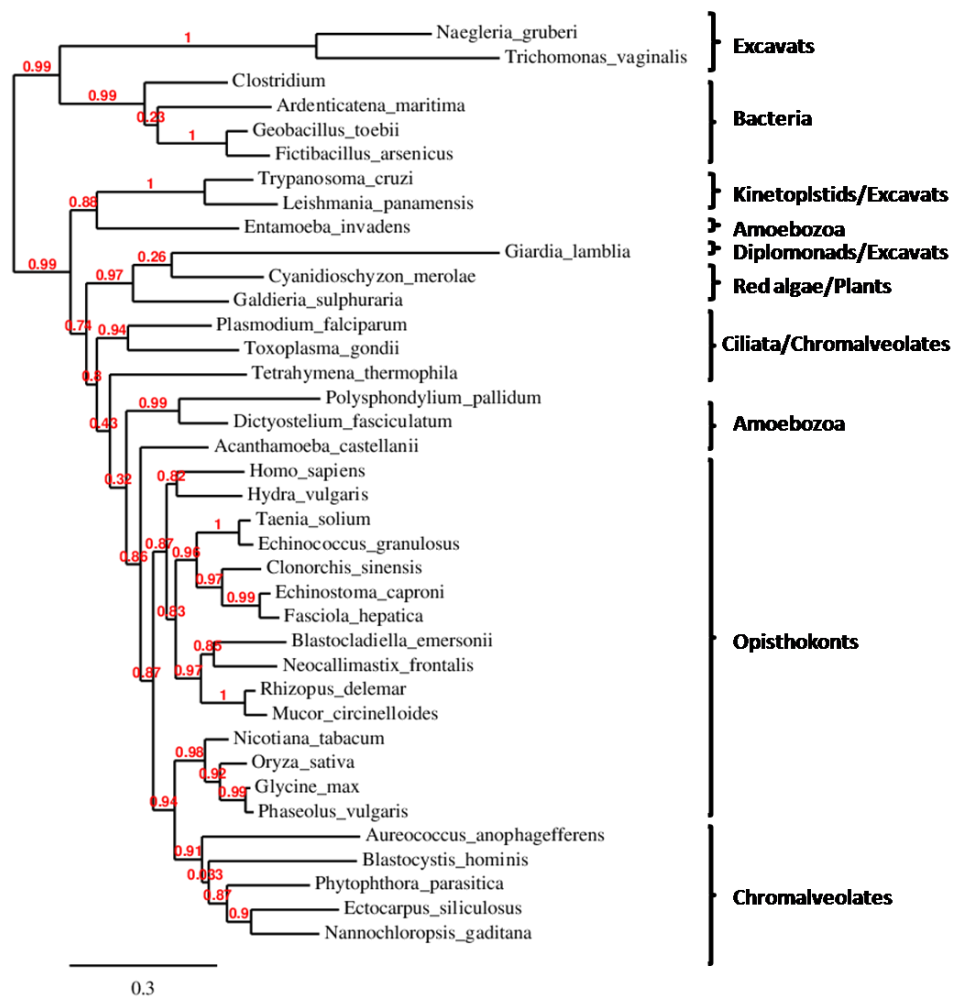


Figure 3. 11 Relationship of *Blastocystis* enolase among *Eukaryotes* and *Prokaryotes* as determined by phylogenetic analysis of protein sequence. Shown is a protein maximum-likelihood topology constructed by the PhyML program. Bootstrap values and posterior probability values were calculated using PhyML and MrBayes. Phylogenetic reconstruction using maximum-likelihood (and: neighbouring-joining or parsimony analyses) shows that *Blastocystis* enolase is embedded within the *Stramenopila* lineages. The dataset contained 11 from all super-groups of *Eukaryotes* and *Prokaryotes*. Numbers at nodes are bootstrap P-values ($\geq 50\%$). A complete list of taxa and their accession numbers can be found in Table 3.3.

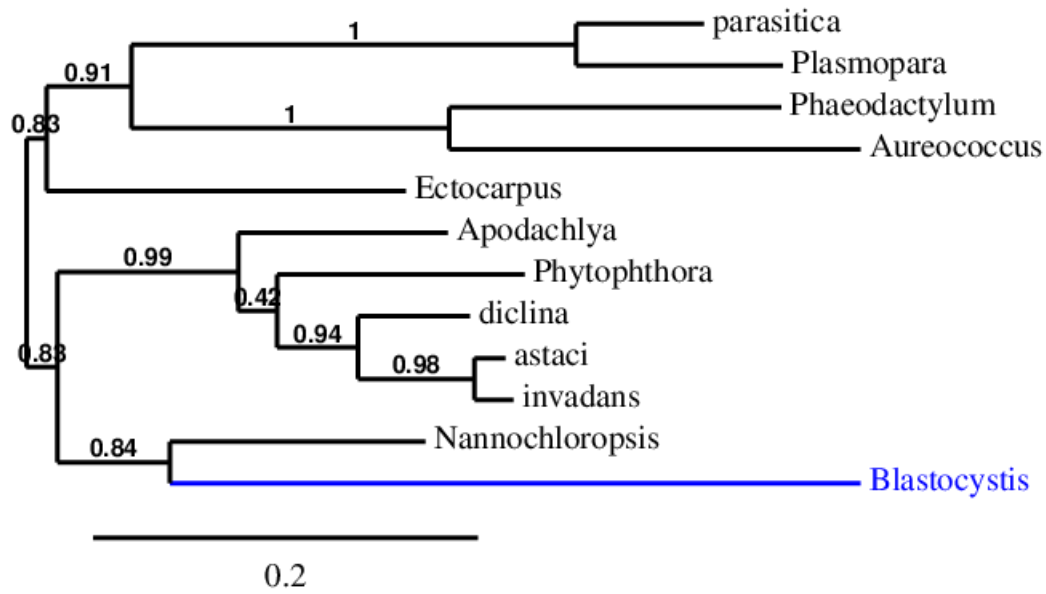


Figure 3. 12 Phylogenetic tree (Maximum-likelihood and confirmed by MrBayes) of *Blastocystis* enolase and enolases from Stramenopila group. Numbers at nodes are bootstrap P-values ($\geq 50\%$). A complete list of taxa and their accession numbers can be found in Table 3.3.

3.5. Conclusion

The conducted comparison of *Blastocystis* enolase amino-acid sequences, demonstrated a high degree of amino acid conservation in enolases. This was the case when compared to representative members of the kingdom of life. It is highly conserved with respect to residues implicated in substrate binding and catalysis active site. The enzyme was also predicted to be a dimer as reported in other enolases, apart from cases where some bacterial enolases had shown octameric enolase ([Karbassi, Quiros et al. 2010](#)).

Of the features that were spotted in some enolases was its plasminogen binding activity ([Vanegas, Quinones et al. 2007](#)), which was not conserved in this instance.

A special finding was that *Blastocystis* enolase posses a mitochondrial targeting signal, which is unusual as glycolysis is canonically taking place in the cytosol. This is an interesting finding as it might imply a special role for enolase being targeted to the mitochondria.

This analysis has shown that *Blastocystis* enolase is like other enolases, a highly conserved enzyme, apart from its unusual mitochondrial targeting signal. More work is to be done to unravel the possible role of this unusual localization in this organism.

Chapter 4. Biochemical characterization of enolase

4.1. General background (introduction)

Enolase (EC 4.2.1.11), also called (phosphopyruvate hydratase or 2-phospho-D- glycerate hydrolase) is a glycolytic enzyme that belongs to the Enolase family (PF03952) (including both the Enolase N and Enolase C family) and mediates the reversible conversion of 2-phosphoglycerate (2-PG) into phosphoenolpyruvate (PEP). It requires magnesium as a cofactor, which helps in catalysis and for stabilizing the dimer protein ([Castillo-Romero, Davids et al. 2012](#)). It is one of the most abundant enzymes expressed in the cytosol of a wide range of organisms ([Castillo-Romero, Davids et al.](#) , [Brewer and Collins 1980](#), [Dannelly, Duclos et al. 1989](#), [Bjerkehagen, Fossa et al. 1994](#), [Agarwal, Hammerschmidt et al. 2012](#), [Azapagasi, Alehan et al. 2012](#), [Bae, Kim et al. 2012](#), [Bruna, Carolina et al. 2012](#), [Chen, Yuan et al. 2012](#)). It participates in many roles and functions, from biological to pathophysiological. It is well known through its role in glycolysis. However, it is also involved in tissue invasion in brain tumors ([Liu 2007](#)) and in parasite virulence by facilitating invasiveness of the host ([Wang, Chen et al. 2011](#)). It was also been found to be involved with cyst formation in *Naegleria fowleri* ([Chavez-Munguia, Segovia-Gamboa et al. 2011](#)). Multifunctional proteins such as enolase behave differently depending on their cellular location, so-called moonlighting where the protein has variant non-enzymatic functions in different hosts and different tissues ([Collingridge, Brown et al. 2010](#)). In vertebrates, there are three isoforms called α , β , and γ of this enzyme, and these are tissue specific. While α isoform can be found in most tissues (but mainly in the liver), the β isoform is principally found in muscle tissue. The γ isoform is neuron specific. All these forms of enolase are found as a

homodimers. Heterodimer forms also do exist, for example $\alpha\beta$ and $\beta\gamma$ ([Keller, Scarna et al. 1981](#), [Joseph, Cruz-Sanchez et al. 1996](#), [Liu 2007](#)). However, enolase generally exists as a homodimer. The molecular mass of the monomers ranges from 22 kDa, as in *Streptococcus rattus* FA-1 ([Huther, Psarros et al. 1990](#)), to 45 kDa in *Streptococcus. pyogenes* forming an octameric enolase of 450 kDa ([Karbassi, Quiros et al. 2010](#)).

Enolase is a highly conserved protein that belongs to the enolase super family. It contains two conserved domains; the N terminal enolase domain and the C terminal enolase domain. Enolase can be divided into clusters according to sequence similarity with a BLASTP E value <10⁻⁹⁰ corresponding to > 35% sequence identity ([Gerlt, Babbitt et al. 2012](#)). The amino terminal (Enolase N) domain covers the amino acid residues 3-131, and the carboxy terminal domain (Enolase C) (which adopt a TIM barrel fold that contains metal binding site) covers amino acid residues 143-431, ([Asuncion, Blankenfeldt et al. 2002](#)). See chapter 3 for a more detailed discussion.

In this chapter, I will discuss the cloning, purification and the biochemical characterization of the recombinant *Blastocystis* enolase to further study its activity and localization into *Blastocystis* cells for understanding the possible role of this enzyme in the parasite cells.

4.2. Recombinant *Blastocystis* Enolase was successfully cloned and over-expressed in *E. coli*

4.2.1. Cloning and amplification of recombinant *Blastocystis* enolase

The gene encoding *Blastocystis* enolase was amplified from *Blastocystis* ST₄ genomic DNA (Figure 4.1) using primers EnoF and EnoR mentioned in Table 2.2. A mix of enzymes was used in the PCR reaction including proofreading *Pfu* DNA polymerase and *Taq* polymerase to minimize any chance of introducing mutations during amplification. The PCR product was incorporated in cloning vector pGEM-T-Easy and subsequently transformed into *E. coli* cells. This construct was confirmed by the DNA sequencing facility at MWG-Eurofins using M13F and M13R primers of pGEM-T-Easy plasmid referred to in Figure 2.3.

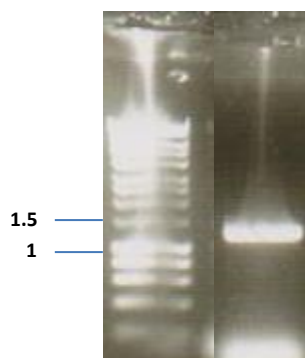


Figure 4. 1 Agarose gel electrophoresis analysis of PCR-amplified enolase, from *Blastocystis* ST₄. The gel shows a band of the amplified *Blastocystis* enolase gene with an expected size of 1.3 kb. DNA HyperLadder I is on the left of the PCR product. Sizes of relevant marker are indicated in kb (Bioline).

The *Blastocystis* enolase was amplified using PCR primers containing *Nde*I, and *Bam*HI restriction sites, mentioned in Table 2.2 for directional cloning into the gene expression plasmid pET-14b to incorporate an N-terminal His-tag for a later affinity purification step. The gene was directly cloned into this expression vector; Figure 4.2 is an agarose gel of the amplified *Blastocystis* enolase gene with the

new restriction sites, and Figure 4.3 a schematic representation of the final construct.

The construct was validated through DNA sequencing using the MWG-Eurofins sequencing facility (Germany). Three strains of *E.coli* cells were used to express this gene: Rosetta2 (DE3) pLyS (Novagen) and BL21 pLysY (Bioline) and Arctic Express cells (Agilent Technologies). *Rosetta* cells were initially used and showed a high expression level of enolase. However, most of the recombinant *Blastocystis* enolase was present in inclusion bodies, and therefore not soluble. Insoluble protein cannot be used for downstream experiment such as activity assay. Therefore the *Blastocystis* enolase gene was transformed into BL21 cells. This resulted in good expression levels in the soluble fraction which was subsequently used for downstream experiments.

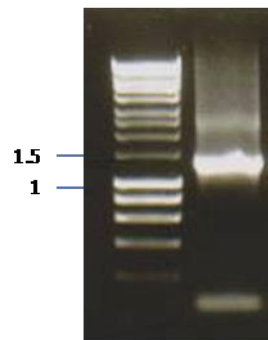


Figure 4. 2 Agarose gel electrophoresis analysis of PCR-amplified *Blastocystis* enolase with added *Nde*I, *Bam*HI restriction sites. The gel shows a band of the expected size 1.46 kb. DNA HyperLadder I is on the left side lane of the PCR product. The sizes of the gene of interest and the relevant reference DNA are indicated next to the band.

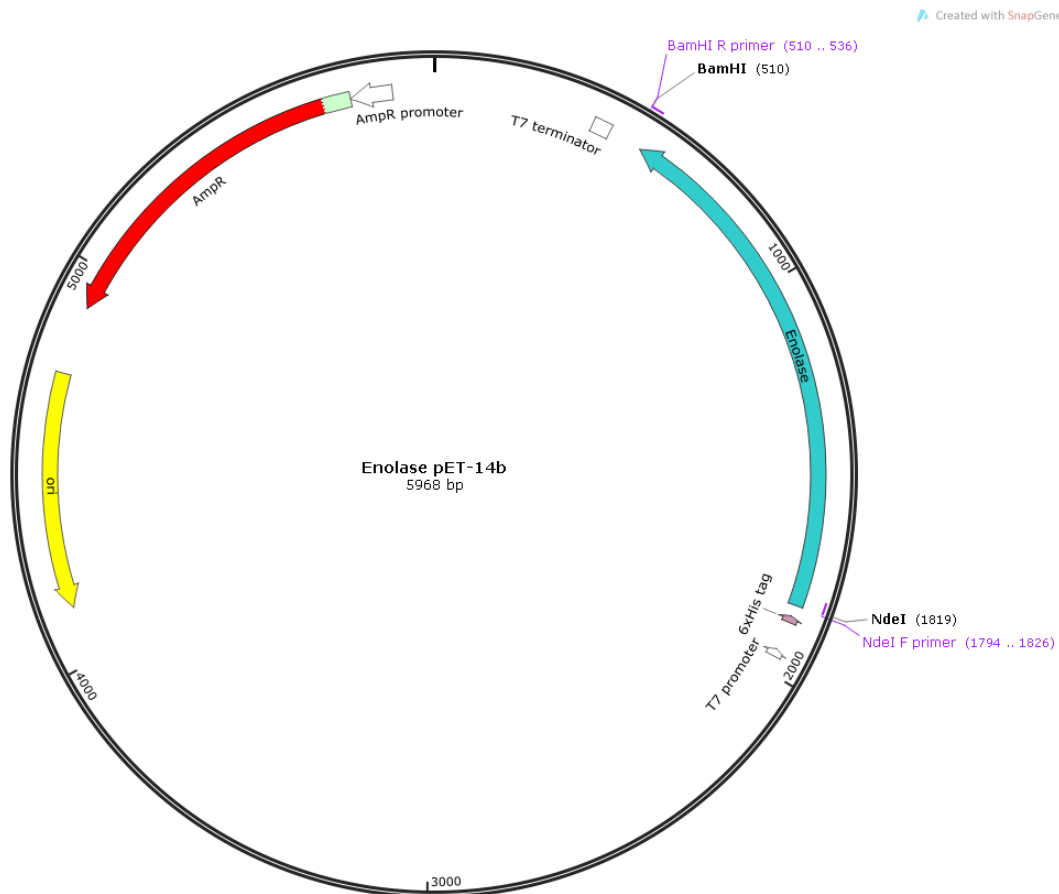


Figure 4. 3 Schematic representation of *Blastocystis* enolase-pET-14b construct of 5,968 kb size with the restriction sites added. The map shows the location of the gene under the influence of T7 RNA polymerase promoter, and showing the ampicillin resistance marker (AmpR).

4.2.2. Recombinant *Blastocystis* enolase is deposited in the inclusion bodies of *E. coli* Rosetta cells but soluble in BL21 and Arctic Express cells.

Analytical-scale expression experiments were performed to determine which *E. coli* strain would produce the highest level of soluble protein. Strains included: BL21, Rosetta 2 and Arctic Express cells. Various parameters were tested to identify optimal expression conditions. Induction times ranged between 2, 4, 5, overnight and 48 hours. Growth temperature tested over 10, 18 or 37°C, and cultures were induced using different concentration of IPTG (1 μ M-1 mM). Different growth media (LB and ZYM-5052) were tested as well, (for media

composition see section 2.1.1.2). Finally expression trials varied between IPTG inductions and auto-induction technique.

The *Blastocystis* enolase pET-14b construct was initially transferred into *E. coli* Rosetta cells which resulted in a high amount of the protein ending up in inclusion bodies. A possible explanation might be that the prokaryotic machinery might have difficulty folding the eukaryotic polypeptide chain and depositing this unfolded or partially folded protein into the inclusion bodies. These are formed by insoluble protein exposing their hydrophobic regions ([Sabate, de Groot et al. 2010](#), [Upadhyay, Murmu et al. 2012](#))

However, transferring the construct into *E. coli* Arctic Express cells and induction at the lower 10°C for 24 hours (after growing for 3-4 of hours at 37°C to boost initial biomass) resulted in very low levels of recombinant protein.

A final trial was performed by using the auto-induction technique which depends on the fact that the expression strain will grow to a high density until inhibiting factors, such as glucose, are depleted from the medium. After that has happened, the lactose will automatically induce the expression of the gene in what is called auto-induction ([Studier 2005](#)). The protein produced using this method gave a high yield in comparison to the other over-expression conditions. This was achieved by growing an overnight culture at 37°C, which was subsequently used to inoculate a 500 ml culture of ZYM-5052 and incubated at 37°C for a period of 2-3 hours or until the OD was less than 1. The culture was subsequently incubated at 20 °C with shaking at 220 rpm for 24 - 48 hours. This method gave a yield about 10 times higher than traditional expression experiments (OD~7-10) and it can sometimes reach an OD of 20. Figure 4.5 gives examples of the main

expression trials at different conditions discussed above and shown on 12% SDS gels.

4.3. Identification of recombinant *Blastocystis* enolase by mass spectrometry

In order to identify the produced protein, a mass spectrometry analysis was performed, for proteomic analysis, by sending a gel slice of the putative recombinant *Blastocystis* enolase, to the protein sequencing facility at the University of Exeter Biosciences, by Dr. Hannah Florance. Protein sample that was sent for analysis was of insoluble enolase found in inclusion bodies as seen in 4.2.2 paragraph. The resulting mass spectrometry data (Figure 4.4) was checked against the protein of interest and clearly showed a very good hit. 69 spectral files were submitted to search and 67 were classed as hits. All hits were attributable to *Blastocystis* enolase, with 64% coverage of the whole protein. The mass spectrometry data also confirmed that there is no contamination from other proteins in the sample which indicated that there was no need to search against the non-redundant species database. See Figure 4.4 for detailed proteomics data of the purified recombinant *Blastocystis* enolase protein.

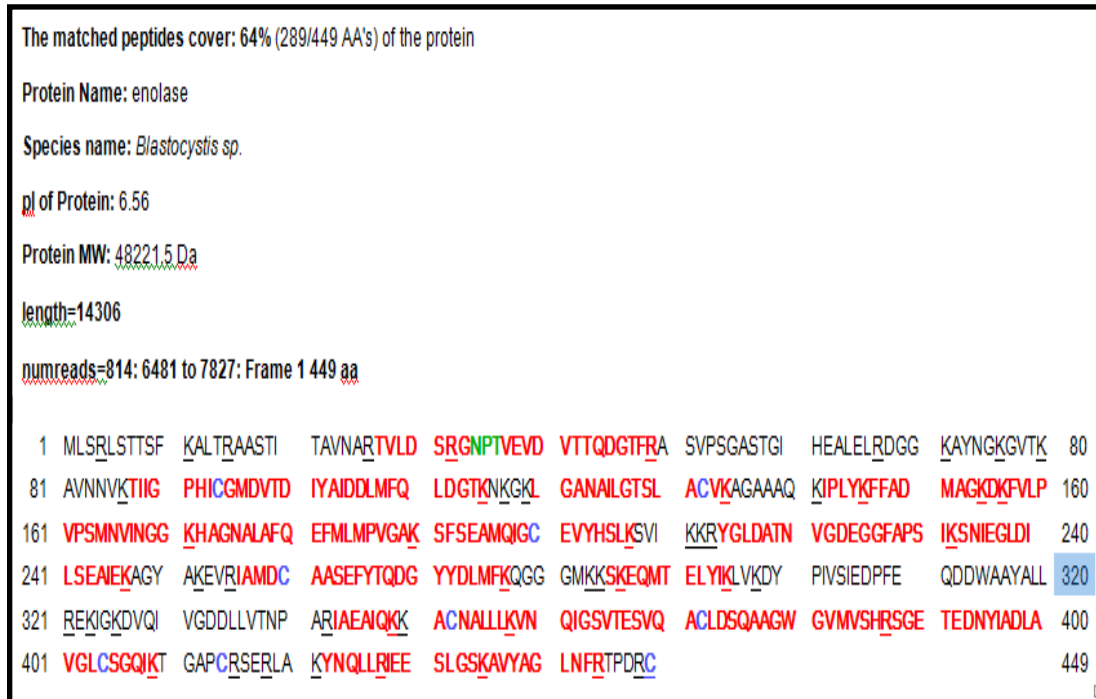


Figure 4. 4 Identification of recombinant enolase *Blastocystis* by mass spectrometry. Matched peptides are shown in red.

4.4. *Blastocystis* enolase expression and purification

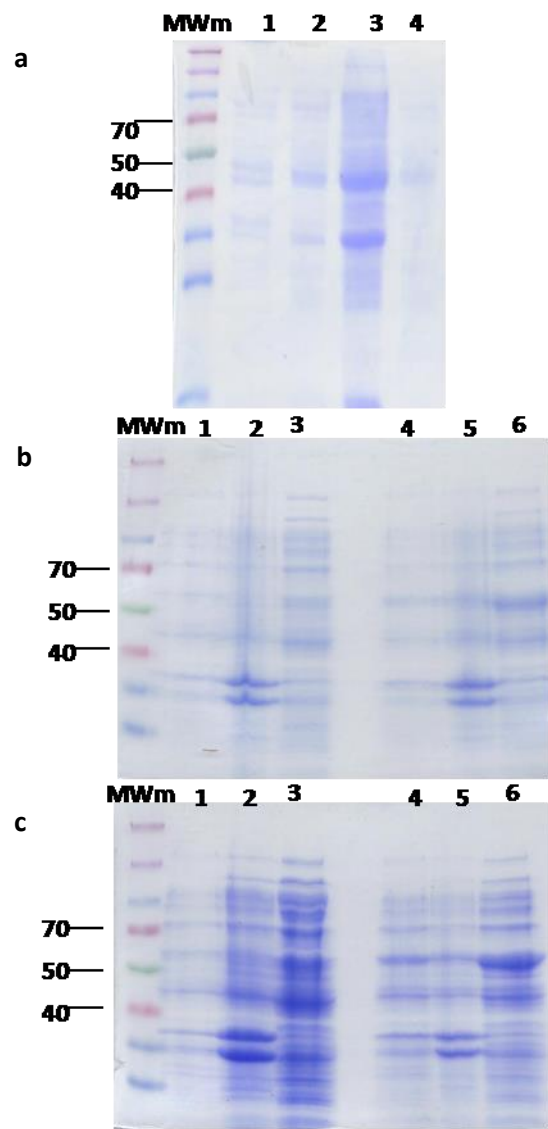
The *E. coli* BL21 pLysY cells harbouring His-tagged recombinant *Blastocystis* enolase in the pET-14b were plated on LB plates containing a final concentration of 100 µg/ml ampicillin and grown overnight at 37°C. A single colony was picked to inoculate a 10 ml LB media containing 100µg/ml ampicillin and incubated overnight at 37°C in a 200 rpm shaking incubator. The 10 ml overnight culture was then used to inoculate 500 ml ZYM-5052 media, which was either grown at 37°C for a few hours until growth had visually started and then transferred to 20°C overnight, or was incubated at 20°C for 48 hours. Cells were harvested at 6000 g for 30 minutes and lysed following the protocol described in methods 2.2.3.2.

A nickel affinity column (NAC) was equilibrated using 5 x column volume of NAC buffer A described in Table 2.6, with a flow rate of 1 ml/min. The crude soluble

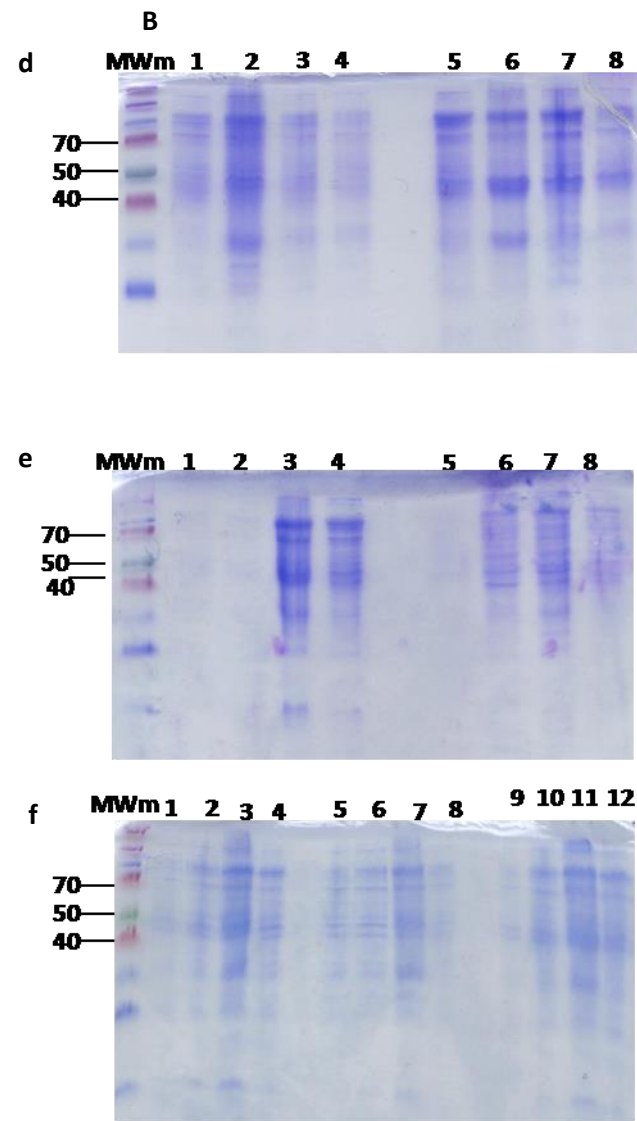
clarified cell extract was loaded onto the column via an injection into the “Superloop”. The unbound protein was washed off using buffer A and the bound protein was eluted using 5 x column volumes of buffer B, (see Table 2.6), using a step gradient up to 100% buffer B. Eluted fractions were all collected. Fractions containing protein gave absorbance at 280 nm. Figure 4.6 represents the elution profile of *Blastocystis* enolase purified by NAC. Fractions containing the recombinant protein were analysed by 12% SDS-PAGE showing the protein of interest with a purity of about 95%, see Figure 4.7.

The fractions representing the putative *Blastocystis* enolase were pooled in one fraction and concentrated to final volumes between 1-5 ml using Amicon® Ultra Millipore filters, with a 30 kDa cut-off value. The concentrated fractions were loaded manually on a previously equilibrated GF column using 1.5 x column volume of GF buffer at a flow rate of a 1 ml/min. Fractions were collected and absorbance measured at 280 nm. Figure 4.8 represents the elution profile of the *Blastocystis* enolase on GF column. The peak containing the *Blastocystis* enolase was also checked on 12% SDS-PAGE, (Figure 4.9). The recombinant *Blastocystis* enolase protein showed a high purity up to about 99 % and was used for all subsequent experiments.

A



B



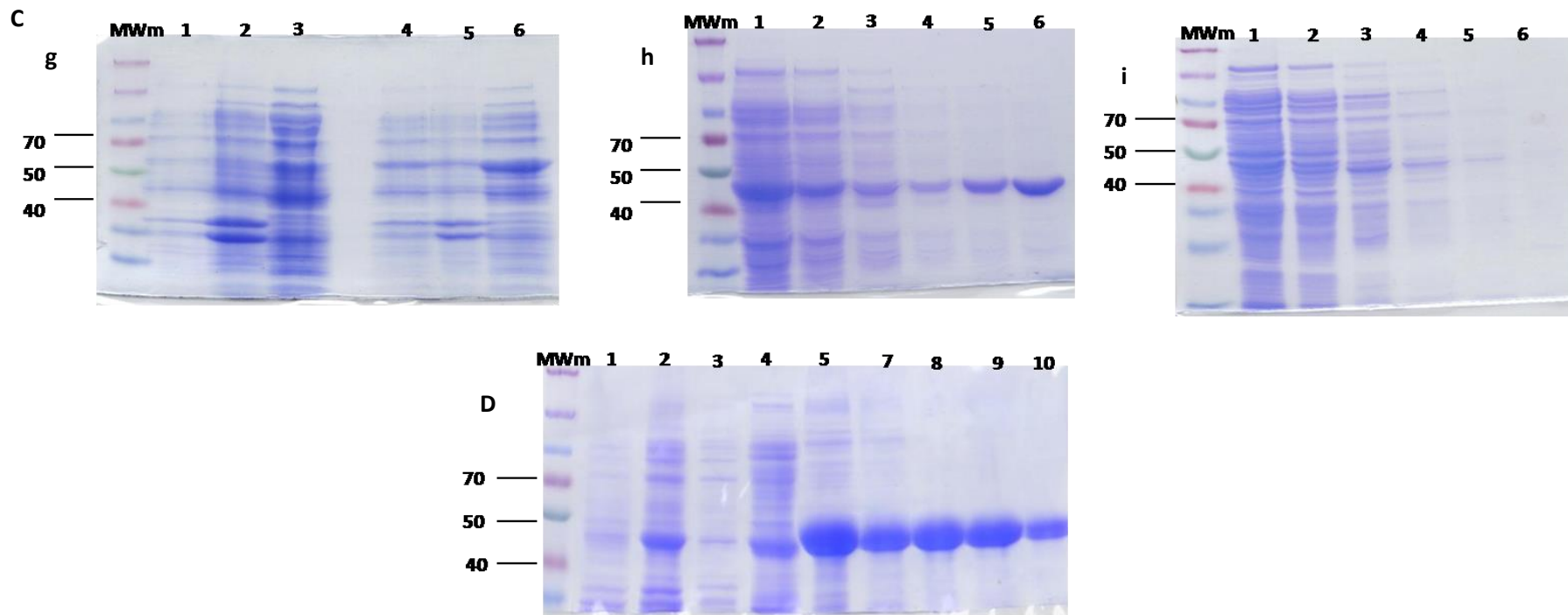


Figure 4.5 Expression trials for *Blastocystis* enolase in different cell lines under several conditions. A) *Blastocystis* enolase cloned into pET-14b in *E. coli* Arctic Express cells was induced using: a) 100 mM IPTG for 24 h at 10°C. Lane 1: uninduced fraction, 2: total cell lysate, 3: insoluble fraction, 4: soluble fraction. b) 1 mM IPTG for 24 h at 10°C where the lanes 1-3 represent the empty vector and lanes 4-6 represent enolase construct, both constructs lanes are in the following sequence: uninduced, insoluble soluble. c) 400 mM IPTG for 24 h at 10°C where the lanes 1-3 represent *Blastocystis* enolase construct and lanes 4-6 represent the empty vector, both constructs lanes are in the following sequence: uninduced, insoluble soluble. B) *Blastocystis* enolase cloned in pET-14b in *E. coli* BL21 cells, which were induced at 37°C for 4 h using: d) 1 mM IPTG, where lanes 1-4 represent the empty vector while lanes from 5-8 represent the *Blastocystis* enolase construct in *E. coli* BL21, e) 400 mM IPTG. Lanes 1-4 represent the enolase construct in *E. coli* BL21 while lanes 5-8 represent the empty vector and f) 10 mM IPTG where lanes 1-4 represent the *Blastocystis* enolase construct in *E. coli* BL21 and lanes from 5-8 represents the empty vector while lanes 9-12 show the expression of *Blastocystis* enolase construct in BL21 using 100 mM IPTG. The lanes pattern of lanes 1-4, 5-8, 9-12 all are in the following sequence: uninduced, induced, insoluble, soluble consecutively. C) *Blastocystis* enolase constructs in *E. coli* BL21 autoinduced at 20°C for 48 h. g) The expression pattern of autoinduced *Blastocystis* enolase construct (1-3) compared to the empty pET-14b vector (4-6). Lanes 1-3 and 4-6 are as follows consecutively: uninduced, insoluble, soluble. h) *Blastocystis* enolase purification with different imidazole concentrations from (10 - 500 mM) after being autoinduced compared to i) autoinduced empty vector construct purified also using nickel affinity column. Lanes 1-6 in h and i are the elution gradient flow throw, 25 mM, 50 mM, 100 mM, 250 mM and 500 mM respectively D) *Blastocystis* enolase construct in pET-14b in *E. coli* Rosetta2 cells induced with 1mM IPTG over night at 18°C. Where lane 1: uninduced fraction. Lane 2: insoluble, lane 3: soluble, lane 4: flow through, 5-10 imidazole concentration gradient from 10-500 mM. MWm: molecular weight marker. Molecular weights are indicated in kDa on the left of all the previous gels.

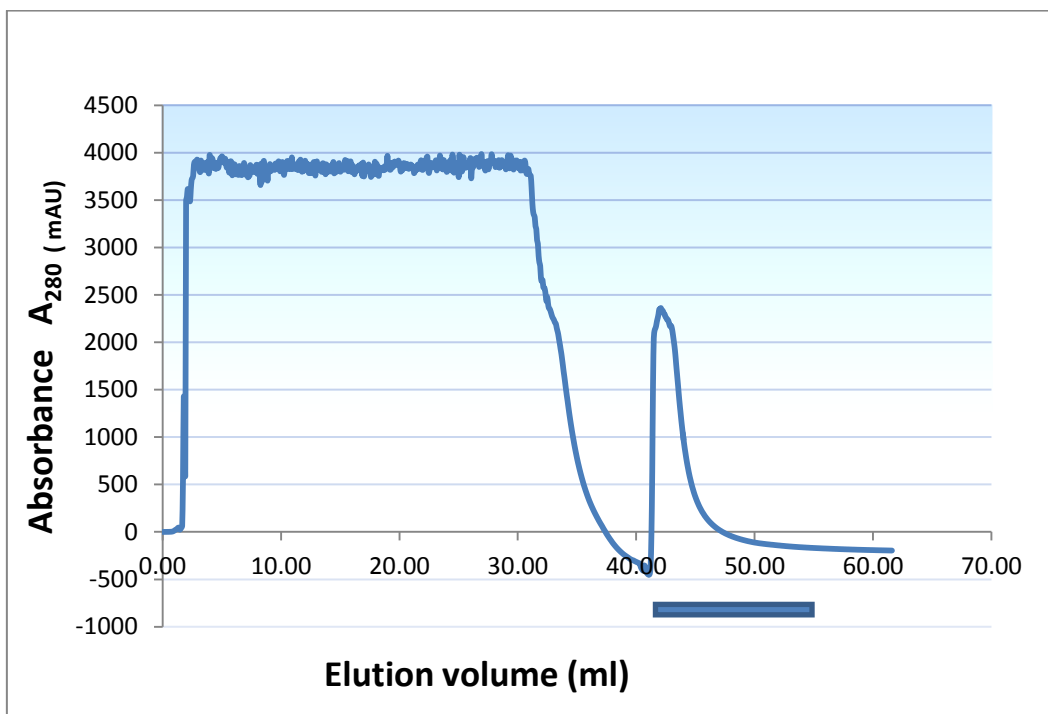


Figure 4. 6 Elution profile of *Blastocystis* enolase from a HisTrap column (GE Healthcare). 10 ml fractions were collected across the entire elution. The A280 trace is seen in blue. The y-axis represents the absorbance at 280 nm. The x-axis shows elution volumes. Protein was loaded on the column from 0-40 ml. Fraction from 45-60 ml represent eluted protein and the bar represents protein analysed by SDS-PAGE (Figure 4.7).

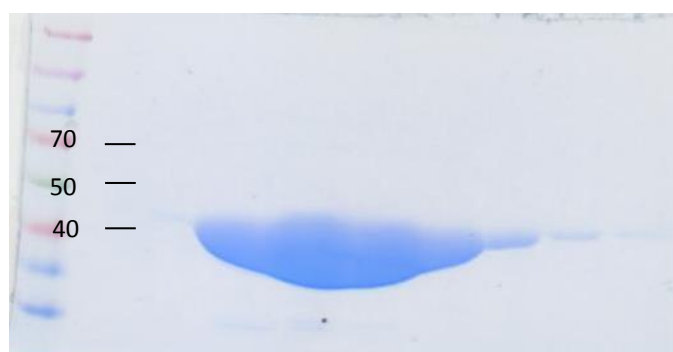


Figure 4. 7 Coomassie-stained SDS-PAGE gel of recombinant *Blastocystis* enolase purified by His-trap column. Fractions represent (47-60) ml, (see Fig. 4.6). The protein bands representing enolase (~48kDa) shows a large amount of protein with purity of more than 95%. Lane 1 shows the molecular weight marker (Fermentas Spectra Multicolor Broad Range). Molecular weights are indicated in kDa on the left.

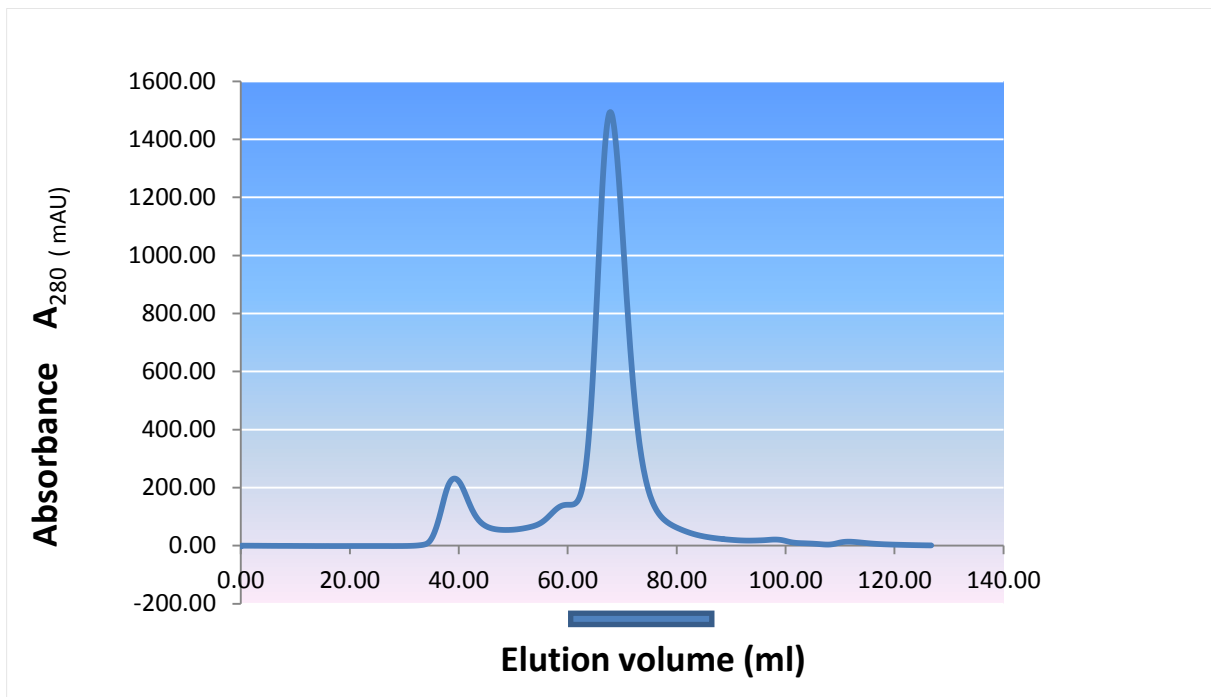


Figure 4. 8 Elution profile of *Blastocystis* enolase from a Superdex 200 gel filtration (GF) chromatography column. The A₂₈₀ trace is seen in blue, and *Blastocystis* enolase eluted in the fractions between 60-80 ml, and the bar represents protein fractions analysed by SDS-PAGE (Figure 4.8). The x axis represents the absorbance of the protein at 280 nm. The Y axis shows the elution volumes of the protein. The peak at about 40 ml represents high molecular weight and aggregated protein, which elutes in the void volume. The peak at about 65 ml represents the eluted target protein, *Blastocystis* enolase. See Fig 4.9.

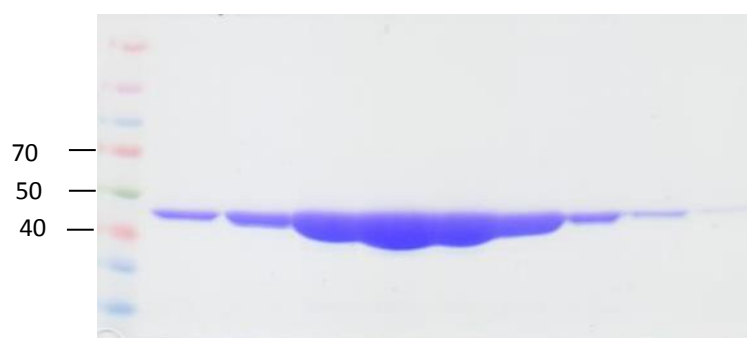


Figure 4. 9 Coomassie-stained SDS-PAGE gel of gel filtrated *Blastocystis* enolase. Fractions representing 60 to 80 ml eluted protein from Superdex 200 (gel filtration) GF chromatography column, (see Fig 4.8). Lane 1 shows the molecular weight marker used from (Fermentas Spectra Multicolor Broad Range). Molecular weights are indicated in kDa on the left.

4.5. Biochemical characterization

4.5.1. Molecular weight determination

4.5.1.1. Analytical gel filtration

In order to determine the experimental molecular weight of the recombinant native protein. The affinity-purified recombinant *Blastocystis* enolase (Figure 4.7) was pooled and analysed on a Superdex 200 HiLoad GF chromatography column which has a separation range for molecules of molecular weight between 10-600 kDa. The recombinant *Blastocystis* enolase monomer has a theoretical molecular weight, as estimated by ExPasy, of 48 kDa.

The Superdex 200 GF chromatography column can be used to estimate the size and oligomeric status of native proteins. *Blastocystis* enolase purification was performed in triplicate and the molecular weight was determined using the gel filtration standard curve shown in Appendix 2. *Blastocystis* enolase was eluted with an elution volume V_e of 68.37 ml. This was used to calculate the molecular weight using the equation 4.1. After applying the trend line equation of the calibration curve, the calculated molecular weight for *Blastocystis* enolase was 74.13 kDa. The eluted fractions were analysed on both a 12% SDS-PAGE gel confirming the size of a monomer of the protein, and on a native gel to estimate the native molecular weight of the protein. All calculations were based on triplicate experiments.

The experimental results differ from the theoretical value and is closer in size to a dimer of a protein. The discrepancies between the expected theoretical molecular weight and the measured experimental molecular weight might be due to experimental errors, or to factors that the calculating programs do not take in consideration while calculating. The molecular weight as measured by

gel filtration was reproducible (performed in triplicate). The possibility of a difference in protein shape between a standard protein and the *Blastocystis* enolase might have contributed to a deviation from the ideal gel filtration behaviour. Protein travel in gel filtration columns not only according to their molecular weight but also according to their volume and size. Hence, a possible difference in the shape might result in an unexpected size estimation.

To be able to determine the molecular composition of the *Blastocystis* enolase, additional experiments were performed. Dynamic light scattering and native gel electrophoresis was performed to get a better understanding of the nature of the native protein.

Measure: $K_{av} = \frac{V_e - V_o}{V_t - V_o}$ Equation 4.1

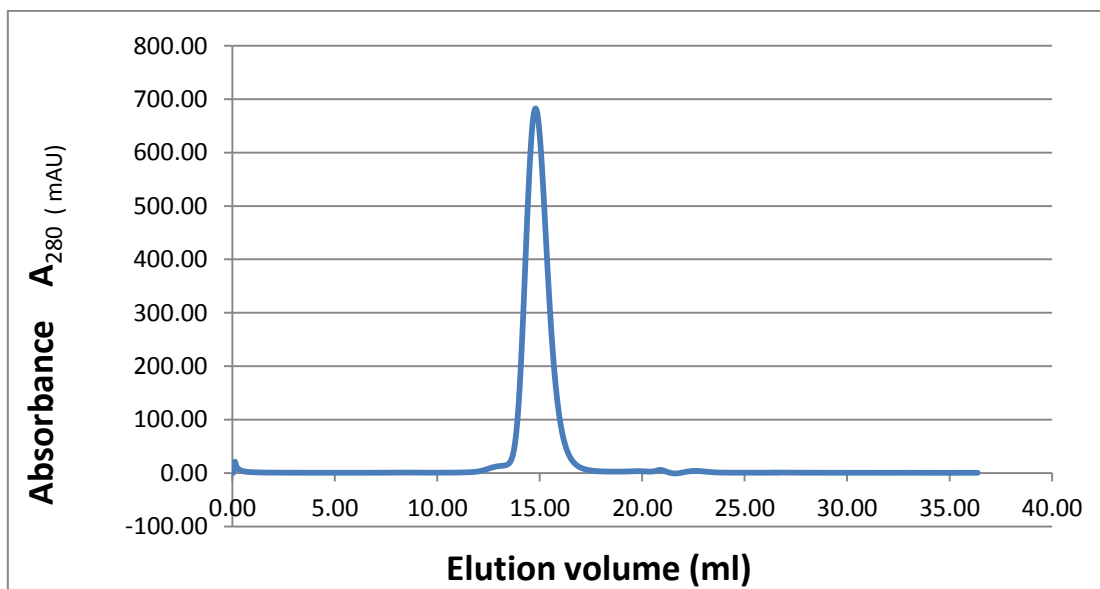


Figure 4. 10 Elution profile of *Blastocystis* enolase analysed on an analytical Superdex 75 GF chromatography column in triplicate. The single peak obtained represents *Blastocystis* enolase. The y axis represents the absorbance at 280 nm. The x axis shows elution volumes in ml.

4.5.1.2. Immunodetection using a heterologous antibody

The purified recombinant *Blastocystis* enolase was immuno detected in a step used to prove identity of the purified protein described previously. The protein was loaded and electrophoresed on a 12% SDS gel and transferred to nitrocellulose membrane as previously detailed in chapter 2, section 2.2.3.7. The transferred proteins were detected using two different antibodies. The first one was a His tag antibody (Qiagene) targeting histidine residues on the recombinant protein. The antibody against this affinity tag was used in a concentration of 1:1,000 followed by the secondary antibody (1:5,000). Figure 4.11/A shows a band corresponding to the calculated theoretical size (~ 48 kDa) of *Blastocystis* enolase.

A heterologous polyclonal antibody against *Trypanosoma brucei* enolase was also used (kind gift from Prof. Julius Lukeš, Czech Republic). This heterologous antibody was used in a concentration of 1:1,000 for the primary antibody and 1:5,000 for the secondary antibody. Immuno-detection using the anti-enolase antibody also showed a band of 48 kDa, the expected size of enolase, see Figure 4.11/B.

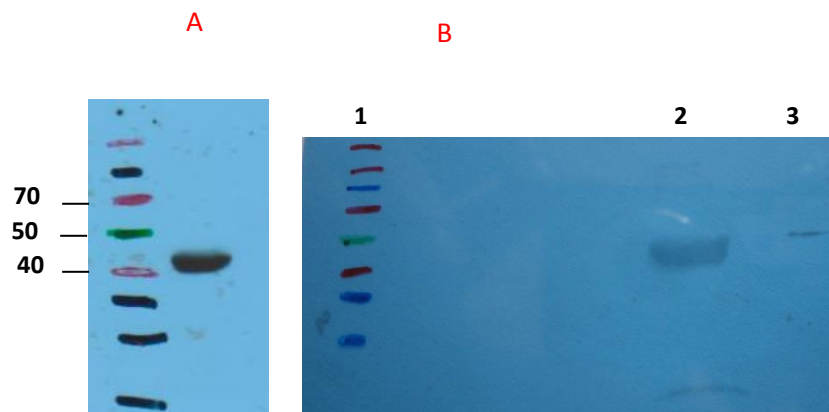


Figure 4. 11 Western blot of *Blastocystis* total protein lysate and recombinant *Blastocystis* enolase with anti- histidine monoclonal antibodies and anti- enolase antibodies. A: only one band corresponding to enolase probed with His Abs as the predicted molecular weight is about 48 kDa. Lane 1: molecular weight marker, (Fermentas Spectra Multicolor Broad Range). Lane 2: Gel filtrated purified recombinant *Blastocystis* enolase purified enolase. B: Western blot of *Blastocystis* enolase using monoclonal enolase antibodies (raised against *Trypanosoma brucei*, kindly donated by Dr. Julius Lukeš, Czech Republic) showing lane 1 of a molecular weight corresponding to *Blastocystis* enolase, Lane 2: *Blastocystis* total cell lysate, lane 3: *Blastocystis* enolase.

4.5.1.3. Oligomeric status determination of the native enolase

Blastocystis enolase purified using GF column was analysed by native gel electrophoresis using native condition as described in methods 2.2.3.5. This gel allows the determination of the molecular weight and also oligomeric status of native protein electrophoresed alongside a set of six standard native proteins mentioned in Table 2.13 ([Olga Randelj 2007](#)). *Blastocystis* enolase was analysed on 12% Clear Native Polyacrylamide Gel Electrophoresis (CN-PAGE) and Blue Native – PAGE (BN-PAGE) for about 3 hour or until the running front reached the front side of the gel, of 100 V at 4°C to avoid denaturing the protein due to the heat produced by the running current (Figure 4.12 A/B). This gel showed a band below the 100 kDa marker, in agreement with the dimeric status of enolase found

in the literature ([Lee, Kang et al. 2006](#)). The band produced by the native gel was cut out and analysed on SDS gel in a technique called 2D electrophoresis, (Figure 4.13). The SDS gel produced a band corresponding to the expected size of *Blastocystis* enolase. This band was tested using western blot analysis with anti-His antibodies (Figure 4.14).

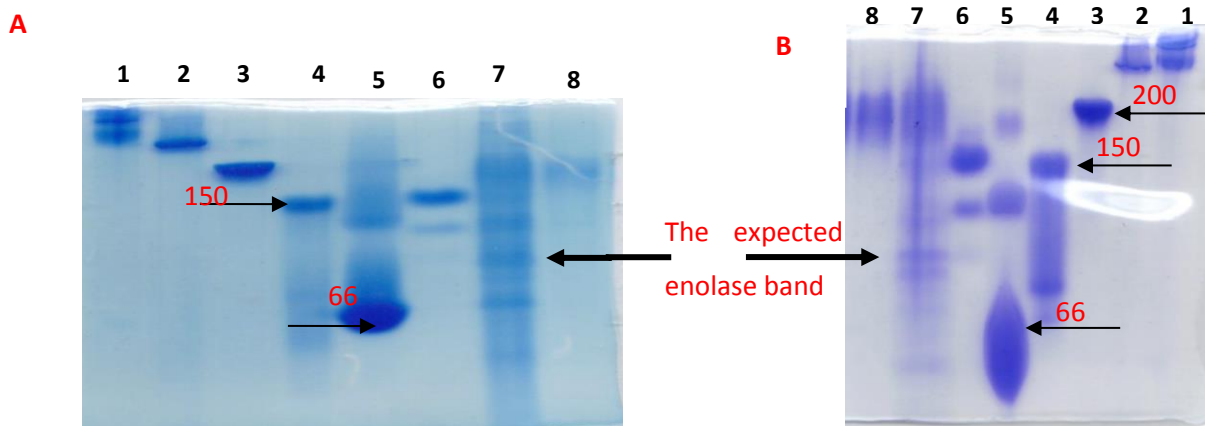


Figure 4. 12 Native gels of gel filtration – purified *Blastocystis* enolase. A. Blue native (BN) gel where the bands from 1 - 6 represents the following standard proteins: 1: Thyroglobuline (660 kDa), 2: Appoferritine (440 kDa), 3: Beta amylase (200 kDa), 4: Alcohol dehydrogenase (150 kDa), 5: Albumin (66 kDa), 6: carbonic anhydrase (30 kDa). B. Clear native (CN) gel. The native gel (CN and BN/PAGE) of recombinant *Blastocystis* enolase alongside 6 protein standard the lanes from 1-6 as in A. The big arrows in the figure refers to the corresponding expected enolase band.

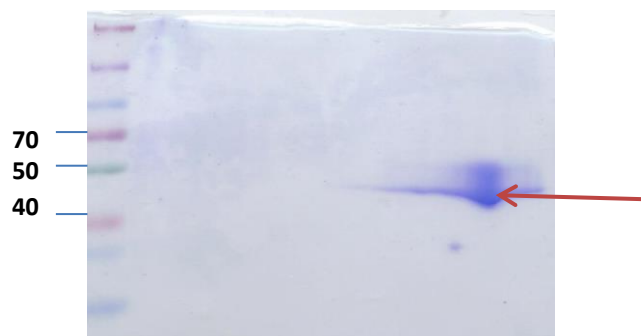


Figure 4. 13 Second dimension SDS-PAGE representing a strip cut from blue native gel PAGE of expected recombinant *Blastocystis* enolase band. The red arrow shows the band representing the expected enolase with a molecular weight of about 48 kDa. Molecular weights are indicated in kDa on the left.



Figure 4. 14 Western blot analysis of the predicted *Blastocystis* enolase from native PAGE. Anti – His antibody was used to decorate the protein.

4.5.1.4. Size distribution profile using Dynamic Light Scattering (DLS)

The *Blastocystis* recombinant enolase was analysed using dynamic light scattering (DLS) as described in methods, chapter 2, section 2.2.3.9.2 after being centrifuged for 20 min at 16,000 X g at 4°C. The analyses was repeated for about 10 measurements and each measurement had a total of 10 acquisitions with 20 seconds each acquisition time recorded at room temperature using DynamPro Titan (Wyatt Technology Corporation) and the analysis programme DYNAMICS (protein solution, USA). Table 4.1 shows the medium measurements for recombinant *Blastocystis* enolase sample parameters such as intensity, diameter, and molecular weight. The mean molecular weight of recombinant *Blastocystis* enolase 77.99 ± 1.3 kDa is in agreement with the analytical column results. The Pd value was 0.2 ± 0.3 nm, where any value below 0.1 indicates a sample near-monodisperse. The percentage of polydispersity is also $< 20\%$ which also indicates a monodisperse sample.

These results suggest the presence of a Heterologous bimodal sample comprising of two peaks which mean that there are two different species with a difference of five times or more in their parameters. The first peak is considered to be negligible due to the very low percentage of mass (0%), where any value of less than 1% of the relative amount of the value of the peak, is considered

negligible. The second peak is composed of two pins, and the intensity of the pin close to the expected size of the *Blastocystis* enolase is 97.9% resulting in a molecular weight of 79 kDa. The other bin intensity comprises only 2.1% of the mass (see Figure 4.16). The presence of a single sharp peak as shown in Figure 4.15 indicates the presence of a homogenous species in the tested solution. This is confirmed by Figure 4.16 which shows the presence of 2 species; however one of these is a very low molecular weight molecule which is likely to be a salt as the intensity is only 2.1% of the whole solution, while the main molecule of 77.99 kDa makes 97.9 % of the solution composition.

Table 4. 1 Dynamic light scattering parameters obtained from 10 measurements and each measurement is a mean of 10 readings. R: diameter of the scattered molecule, Pd: Poly dispersity, MW-R is the molecular weight measured in kDa.

Item	Time (s)	Temp (C)	Intensity (Cnt/s)	R (nm)	Pd (nm)	%Pd	MW-R (kDa)
Measurement 1	200	18	109123	3.8	0.1	3.3	76
Measurement 2	409	18	108058	3.8	0.2	5.3	77
Measurement 3	629	18	108015	3.9	0.9	22.4	79
Measurement 4	833	18	107007	3.8	0.9	22.3	78
Measurement 5	1037	18	105749	3.8	0.2	5.5	78
Measurement 6	1239	18	107264	3.9	0.2	4.2	80
Measurement 7	1441	18	106161	3.8	0.9	22.5	77
Measurement 8	1698	18	102997	3.8	0.1	3.2	75
Measurement 9	1902	18	103438	3.8	0.1	2	77
Measurement 10	2104	18	103426	3.8	0.3	8.7	78

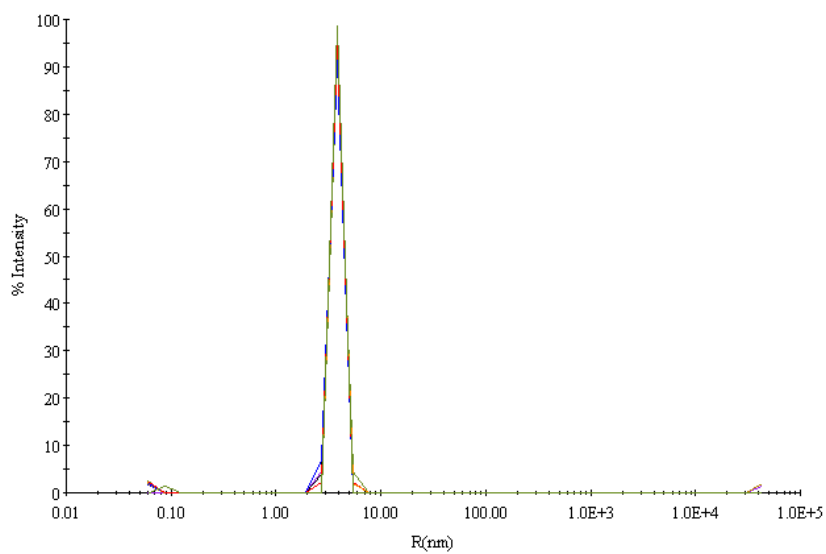
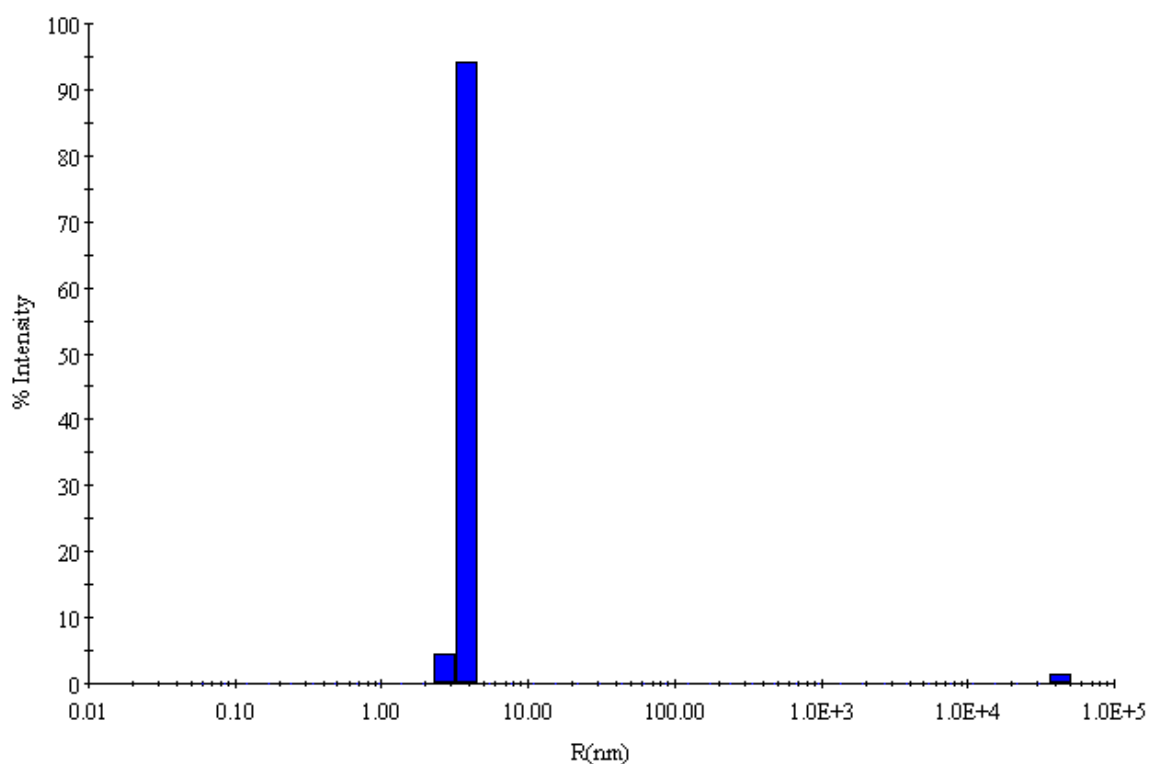


Figure 4. 15 Size distribution (by intensity) of *Blastocystis* enolase. The sharp peak indicates the presence of a single species component with no evidence of aggregation.



Item	R (nm)	Pd (nm)	%Pd	MW-R (kDa)	%Int	%Mass
<input checked="" type="checkbox"/> Peak 1	3.8	0.2	6.1	75.804	98.6	100.0
<input checked="" type="checkbox"/> Peak 2	43110.6	0.0	0.0	235060000000.000	1.4	0.0

Figure 4. 16 A Heterologous bimodal size distribution histogram (regularization graph) of the gel filtration- purified *Blastocystis* enolase. A total of two distinct populations were detected. The graph displays the distribution of particle size from a regularization analysis of the data selected. The observed mass peaks from dynamic light scattering on recombinant *Blastocystis* enolase protein shows a total of two peaks, and the second peak comprises of two bins. The intensity of the scattered light of the second peak was 97.9% with a molecular weight of 79 kDa along with small molecules of 2.1% intensity. The first peak is considered negligible in these results. The results table summary under the histogram shows an estimated relative amount of mass (concentration) (% mass) of 100% of the second peak. The x axis represents the discrete particle size in nm. The y axis represents the relative amount of light scattered by each bin (% intensity).

4.5.2. Fluorescent-based thermal shift assay (DFS: differential scanning fluorimetry)

The thermal stability of a protein in a solution is influenced by several factors such as buffer identity, pH capacities, ligands, metal-ions, salts and concentrations. Any change in these components will result in changes in the protein melting temperature, T_m . To screen for the optimal condition at which *Blastocystis* enolase shows the highest stability, the protein is made in contact with a fluorescent dye such as SyprOrange®, which is quenched in the solution. By exposing the protein to increasing temperatures, it will undergo thermal unfolding, whereby the hydrophobic sites will be exposed. The dye becomes unquenched by attaching to the hydrophobic sites of the protein. The procedure is performed using StepOne™ Real-time PCR and the results are analyzed using PROTEIN THERMAL SHIFT™ software, version 1.0 (applied Biosystems, UK). Temperature will be plotted against the fluorescence signal obtained. The midpoint temperature towards protein unfolding represents the T_m which is the optimal temperature at which the protein is stable. Figure 4.17 represents the calculated melting temperature of the *Blastocystis* enolase. Thermal stability assays were done in duplicates.

In this project a screen of different buffers at different pH values was used. A list of all the buffers used in this experiment is listed in Table 2.7. A thermoFluor® instrument was programmed to increase the temperature from 25°C to 94°C by 1°C/min. Fluorescence was measured using a CCD camera. A comparison between the melting temperatures of the protein in different buffers at different pHs and different salt concentrations was performed and can be seen in Figure 4.18. *Blastocystis* enolase showed the highest stability in HEPES buffers and even more when HEPES was combined with glycerol. This is reassuring as

HEPES1 was used for enolase purification in GF columns. The stability of enolase was also reasonably good with imidazole, which is used in NAC. An increase in stability was noticed with increasing ionic strength. See for example the case of HEPES with varied concentrations of sodium chloride and glycerol. In the case of a complicated buffer mix, the stability of the protein in HEPES and gly 1 was the highest. Some additives such as glycerol are known to enhance the stability of the protein by forming glycerol – water bonds, these bonds form a coat around the protein molecule. Also as glycerol is a poly-hydric alcohol, this decreases the rupturing nature of hydrogen bonds in addition to the hydrating effect of glycerol on the molecule’s surface. However, a decrease in stability was noticed when the glycerol percentage was doubled to 40% in the buffer composition. A similar effect was reported when the salt concentration was increased up to 2 M. ([Gekko and Timasheff 1981](#)).

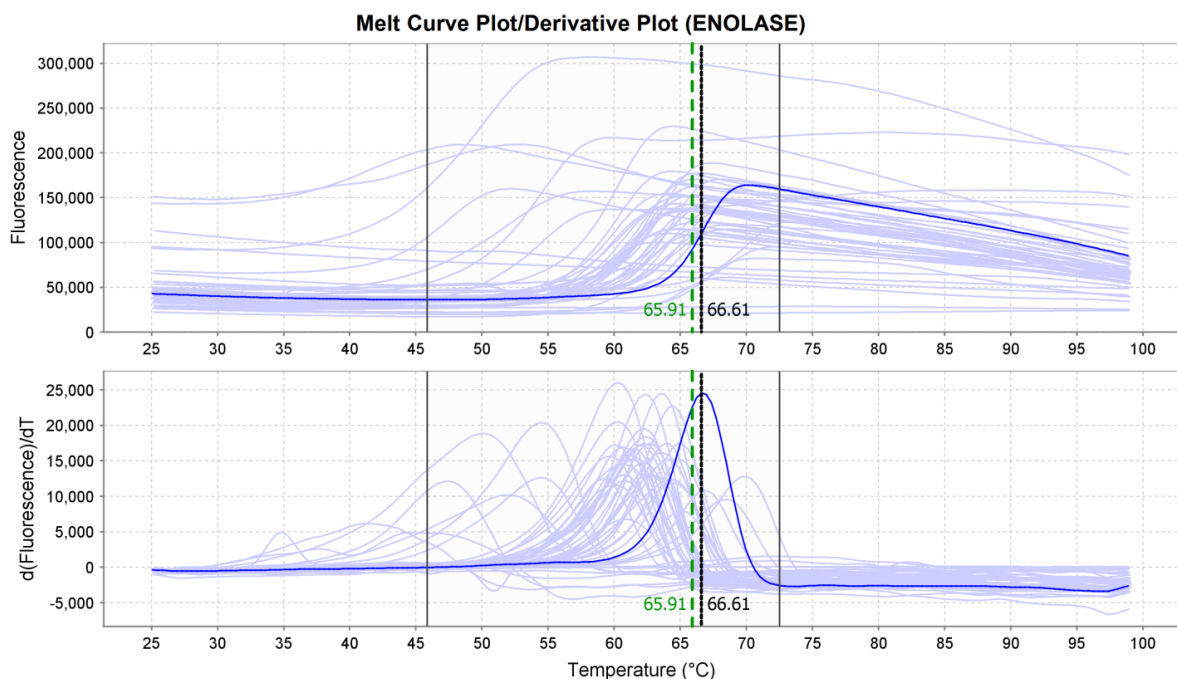


Figure 4. 17 Thermal shift stability assay of *Blastocystis* enolase. The optimal melting temperature of *Blastocystis* enolase is 65.91°C. The x axis shows the melting temperature of the protein-buffer mix. The Y axis shows the fluorescence signal detected.

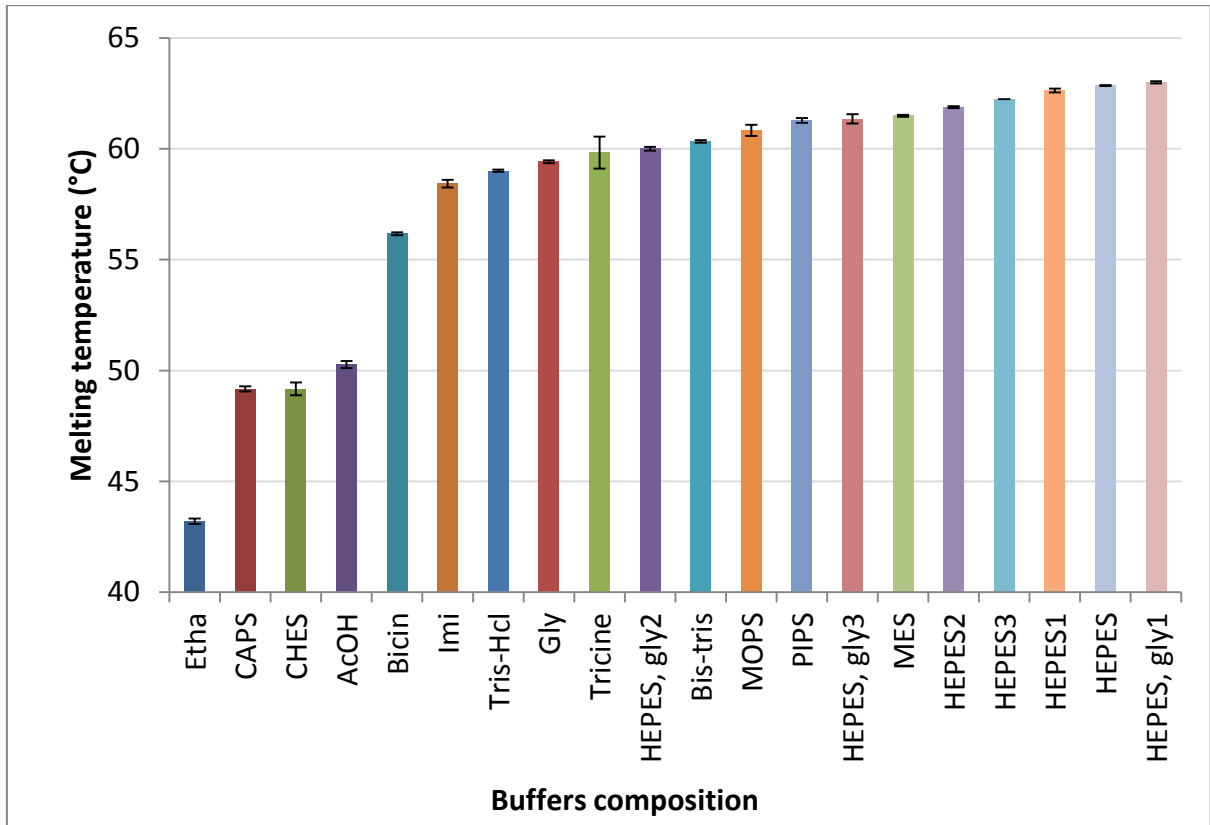


Figure 4. 18 The stability of *Blastocystis* enolase versus buffer composition, and salt addition. Variations in protein melting temperature with different buffer's composition and varying pHs. Buffers names in this experiment are shown in abbreviation detailed in Table 2.7.

4.6. Activity assay and kinetic parameters

The principle reaction for detecting recombinant *Blastocystis* enolase activity is explained in Figure 4.19 where enolase mediates the 9th reaction of glycolysis and catalyses the reverse reaction of 2-phosphoglycerate into phosphoenolpyruvate. Pyruvate kinase then converts phosphoenolpyruvate into pyruvate, a three carbon molecule, with the production of ATP. NADH should then be re-oxidized to NAD and this happens in the mitochondria in aerobic conditions, while in anaerobic conditions pyruvate is reduced to form lactic acid

in animal cells. In yeast, the reduction of pyruvate in anaerobic conditions results in carbon dioxide and methanol instead of lactic acid.

Recombinant *Blastocystis* enolase activity was determined spectrophotometrically by measuring the conversion of NADH⁺ into NAD⁺ (the reduction of absorption at A₃₄₀ nm). Enzyme kinetics was determined by carrying out the standard enolase assay ([Bergmeyer 1965](#), [Florian Muller 2012](#)), and measuring the continuous rate changes over a period of 5 minutes for each measurement. Main measurements were carried out at 25°C (Section 2.3.8.2).

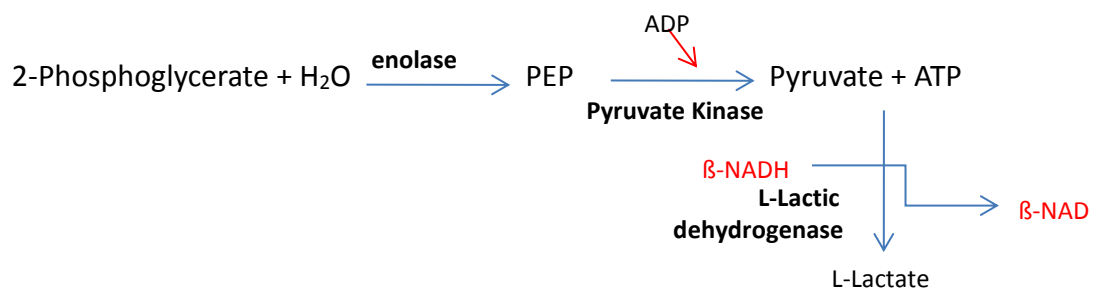


Figure 4. 19 Schematic representation of recombinant *Blastocystis* enolase assay in which one molecule of NADPH⁺ is oxidised into NAD⁺. 2-phosphoglycerate is converted into phosphoenolpyruvate (PEP) in the presence of enolase, thereafter it is converted by pyruvate kinase into pyruvate producing one molecule of ATP. Pyruvate is then reduced into lactate by lactate dehydrogenase. NADPH⁺ absorbance is monitored at 340 nm.

Recombinant *Blastocystis* enolase activity detection was performed at 25°C at increasing substrate concentrations while the enzyme concentration was fixed using 0.05 units in all reactions, see Figure 4.20. The increased substrate concentration was accompanied by an increase of the velocity of the reaction, suggesting the presence of a genuine enolase activity. The velocity of the reaction reached a plateau at about 56 mM of the substrate concentration. According to the Michaelis-Menten plot, the maximum velocity of the recombinant

Blastocystis enolase was 0.03808 mmol/min, as calculated by SigmaPlot while the K_m value was calculated as 16.8 mM. K_m gives an idea about the affinity of the enzyme to its substrate, the higher K_m value the lower the affinity of the protein to the substrate. It is clear from the calculated Michaelis-Menten value that recombinant *Blastocystis* enolase has a low affinity for its substrate in comparison to other previously studied enolases ([Merkulova, Lucas et al. 1997](#), [Quinones, Pena et al. 2007](#), [Yang, Qiu et al. 2010](#)). Table 4.2 shows examples of enolases purified for other organisms, showing a high K_m of 16.8 mM for *Blastocystis* recombinant enolase while it is 0.078 ± 1.1 mM in mouse muscle-specific enolase ([Merkulova, Lucas et al. 1997](#)), 1.99 mM in *Schistosoma japonicum*, 0.08 mM in *Leishmania mexicana* (it has to be noted that this study used a different assays method) ([Quinones, Pena et al. 2007](#)), A similar low affinity of the enzyme for its substrate of 7.2 mM can be seen for *Ricinus communis* ([Miernyk and Dennis 1984](#)) and the highest 43 μ M in *Oryctolagus cuniculus* ([Yang, Qiu et al. 2010](#)). One possible explanation about these data discrepancies from literature could be attributed to the odd point in the data shown in Figure 4.20. Deleting this point might lead to changing the kinetics to fit with enolase kinetics reported in other organisms.

Table 4. 2 Michaelis-Menten constant value for different organisms including *Blastocystis* sp4 (this study)

Organism	K_m value (mM)	Study
<i>Blastocystis</i> sp 4	16.8	This study
<i>Leishmania mexicana</i>	0.08	Quinones, Pena. 2007
<i>Schistosoma japonicum</i>	1.99	Yang, Qiu. 2010
Mouse muscle	0.078	Merkulova, Lucas. 1997
<i>Ricinus communis</i>	7.2	Miernyk, 1984

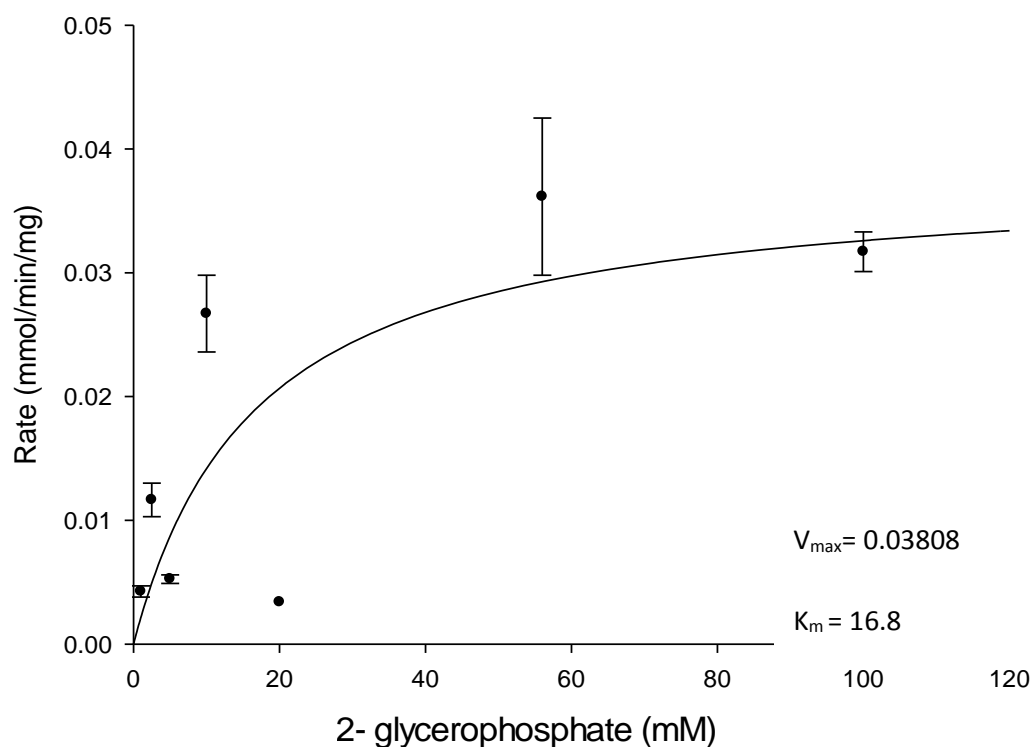


Figure 4. 20 Steady state kinetics of recombinant soluble *Blastocystis* enolase showing an increase in the reaction rate with increasing substrate concentration, reaching a plateau at about 56 mM 2-glycerophosphate. X-axis: substrate concentration (2GP: 2-phosphoglycerate), Y-axis: rate of reaction in m.mol/min/mg. 0.5 mg/ml of enolase was used in this assay. Error bars show standard deviated calculated from duplicate measurements.

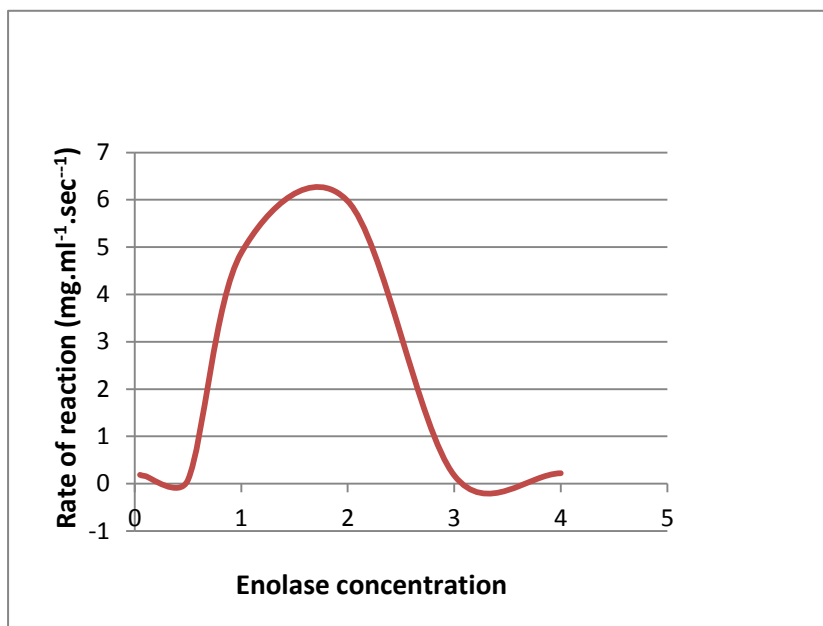


Figure 4. 21 Rate of the recombinant *Blastocystis* enolase activity reaction and its dependence on protein concentration. Protein concentrations varied between 0-4 mg/ml. X axis shows the enolase concentration in mg/ml. Y axis shows the rate of the reaction in mg.ml⁻¹ .sec⁻¹

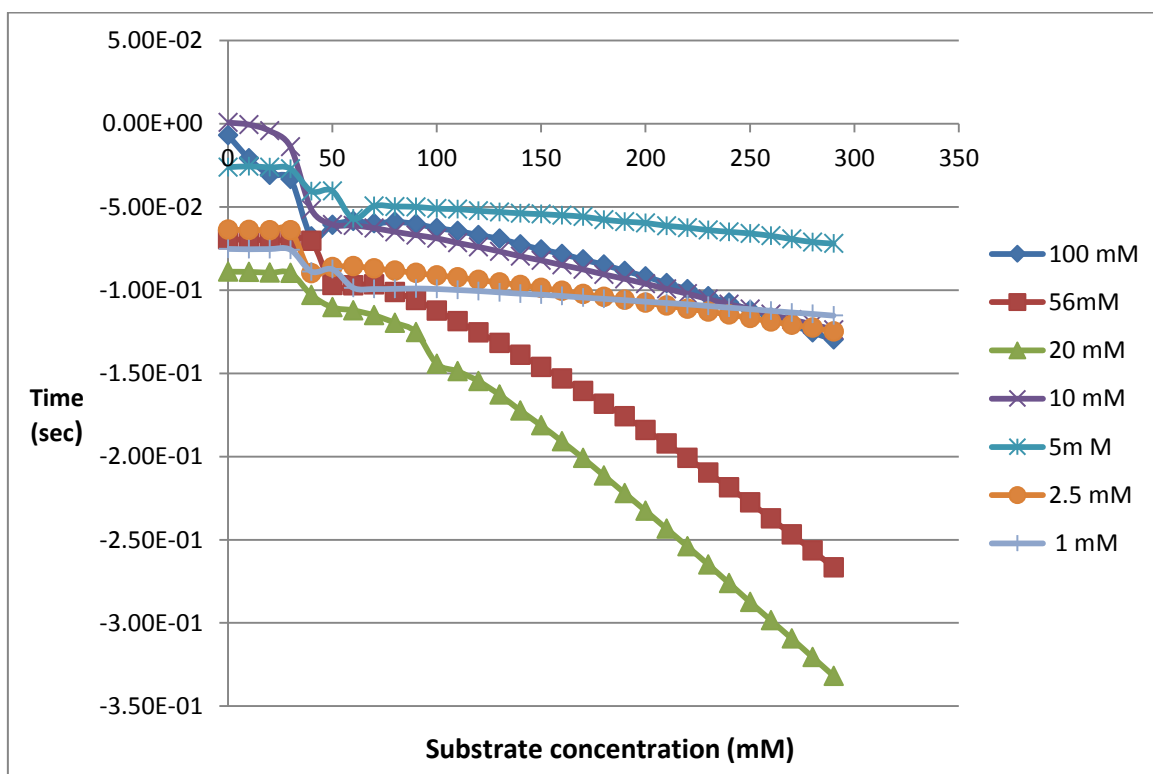


Figure 4. 22 Recombinant *Blastocystis* enolase activity using varied substrate concentration which varied from 1 mM to 100 mM. The substrate concentration is coloured varied. The X axis shows the protein concentration in mM. The Y axis shows the time consumed by the protein in sec.

When comparing the effect of substrate concentration on the reaction rate, it was noticed that the reaction was faster with substrate concentrations such as 20 and 56 mM, while it was almost unnoticed when using low concentration as 1 mM. The higher substrate concentration of 100 mM resulted in a very low rate as well (see Figure 4.21/ Figure 4.22).

4.7. Summary

Here I have shown that the *Blastocystis* enolase was successfully cloned into different expression systems, of which *E.coli* BL21 gave the best yield when autoinduced at 20°C for 48h. Cloning into *E.coli* Rosetta2 cells gave a high yield of recombinant *Blastocystis* enolase as well, but this was deposited in inclusion bodies, which suggests unfavourable conditions that forced the protein to target the inclusion bodies. *Blastocystis* enolase was purified using two basic columns. A nickel affinity column, utilising a His tag flanked on the N-terminus on the recombinant protein. This column resulted in a high purity of the protein of about 95%. A SEC column resulted in a final purity of about 99%.

The recombinant protein *Blastocystis* enolase identity was checked using both His tag and a heterologous enolase specific antibodies from *Trypanosoma brucei*.

The recombinant *Blastocystis* enolase was shown to be a homodimer after using mass spectrometry, native gel electrophoresis, analytical gel filtration and dynamic light scattering experiments, in accordance with previously published data for other enolases. The calculated theoretical molecular weight is 48 kDa, while the measurements were about 74 kDa according to analytical gel filtration, and 78 kDa according to dynamic light scattering. Those data suggest a monomer of about 37 kDa. This difference between the theoretical and calculated molecular

weight could be attributed to experimental errors, or to possible differences in the shape of the protein standards used for generating the standard curve of gel filtration. Other methods are advised to be used for the calculation of the molecular weight of enolase. Since gel filtration and native gel do not give a precise value due to the nature of the experiment, and same for dynamic light scattering. The addition of more steps is recommended such as using analytical ultra-centrifuge, which is more precise, but less approachable due to the need for specialist equipment and also because it is more time consuming (approximate 4 hours compared to 10-20 minutes using dynamic light scattering). Analytical ultra-centrifugation not only gives a precise number of the protein's molecular weight, but it can also give an idea about the oligomeric status of the protein. Another method that could be used is size exclusion chromatography-multi angle laser scattering (SEC –MALS) which is a combination of light scattering and gel filtration.

The kinetics parameters; V_{max} and K_m of the soluble recombinant *Blastocystis* enolase were lower than those of previously studied enolases in different organism and show a low affinity to the substrate. This high K_m relative to the physiological concentration of the substrate indicates that the protein is not normally saturated with the substrate and its activity depends highly on the substrate concentration and hence the product depends on substrate availability. The high K_m value gives an indication of low affinity of the enolase to its substrate 2-phosphoglycerate. Which means that the product is more readily available in the environment. A study performed by Liu *et al* was looking to characterize proteins depending on their Kinetic modelling, has referred to

to enolase, GAPDH and PGK as critical enzymes in system biology, i.e. their deletion will influence the metabolic system robustness ([Li and Liu 2012](#)). This means that any change of these enzymes would make a systemic biology impact on distant areas and exert a persistent impact on system structure. Therefore, enolase has a system-level role on the physiology not only the local effect studied by the protein's biochemical and biophysical properties.

Therefore and from what was mentioned above, the unusual value of enolase K_m compared to other enolases could suggest that *Blastocystis* enolase exhibits different biochemical properties than enolase studied in other organisms, shown by a low affinity to its substrate. In an attempt to explain this, several theories might be utilised to achieve this; first it was shown in chapter 3 that enolase may localize differently than other text book enolases. It showed to localize to the mitochondria and not the cytosol of the cells. This might indicate a different role played by *Blastocystis* enolase, falling into the moonlighting group of proteins which exhibits several other functions that the protein might play. Being a moonlighting protein might mean that the gene coding for *Blastocystis* enolase is a duplicated gene of the ancestral enolase gene. This might result in two possibilities; either that the duplicated copy gene has lost its original function through what is called loss-of-function mutation which means that they both share the same sequence but the other copy is un-functional, otherwise it had adapted a new function. This process of losing functionality might take millions of years after having the duplicated gene and before losing the function. Or it might have lost its functionality, and is in the process of gene resurrection which happens due to stressful new environmental changes that challenges the organism, in an event known as gene conversion ([Koch 1972](#)). But this not likely to be correct since the *Blastocystis* enolase had shown activity as a glycolytic protein.

The other suggestion is that the *Blastocystis* enolase gene is duplicated gene that could be still functional but performing other functions, such as other moonlighting enolases reported in literature. And here we face two scenarios; the gene might become a paralog gene adopting a new function forming what is called a *neofunctionalisation* where one copy retains the main function and the other copy adapts a new function, which would be the moonlighting phenomenon. Or that the paralog copy goes in what called quantitative *subfunctionalisation* where both copies retain part of the function and hence the ancestral function is partitioned between the two paralogs ([Espinosa-Cantu, Ascencio et al. 2015](#)).

On the other hand, and looking at the data points in Figure 4.20, an odd point is protruding out of the curve. This point might be the reason behind different kinetics of *Blastocystis* enolase of the study compared to other enolases.

Further work is to be performed to find the possible role of enolase in *Blastocystis* virulence, and find the impact of its localization to the mitochondria on its role.

Chapter 5. *In silico* characterisation of recombinant *Blastocystis*

TPI-GAPDH fusion protein

Glycolysis is an important process that offers energy supply to all living cells. Text books always refer to glycolysis process to be a cytosolic process, with some exceptions as in trypanosomatids (*Trypanosome* spp and *Leishmania* spp) whose glycolysis only take place in peroxisomes like organelles called glycosomes.

However in a paper published 2000 by Liaud et al, a non-canonical localization of glycolytic proteins was reported in diatoms ([Liaud, Lichtle et al. 2000](#)). They showed the presence of unusual glycolytic protein TPI-GAPDH to possess a mitochondrial targeting signal. They had also reported the presence of other glycolytic enzymes, not only TPI-GAPDH, and basically a tendency to be those taking place in the second half of the glycolytic pathway, i.e. the payoff phase.

Later in 2012 another group published an *in silico* analysis of the glycolytic proteins and they showed that this phenomenon is spreading to a wide group of organisms that belong to Stramenopiles ([Nakayama, Ishida et al. 2013](#)).

In *Blastocystis*, a similar event was found, and an unusual glycolytic protein TPI-GAPDH was also found to be fused in one gene, with a possible targeting to the mitochondria.

Hence, it is sought in this project to analyse the *Blastocystis* TPI-GAPDH sequence, and find where this protein localizes in the living system, giving an insight about the evolutionary relation of this protein and other mitochondrial protein reported as well.

5.1. Individual proteins super families

The *Blastocystis* TPI-GAPDH fusion protein is composed of two proteins coded on one gene and being translated as a 603 amino acid protein, and assembled in the mitochondria as a tetramere. No isoforms were found in *Blastocystis* cytosol as a stand-alone proteins, either TPI or GAPDH, making it the only version of those two glycolytic proteins. For this, a separate investigation of those two proteins is done to find how conserved are those proteins, and if they kept their active sites and residues.

To perform this, *Blastocystis* TPI-GAPDH fusion protein sequence was used as a query search against all main group of living system, including prokaryotes and eukaryotes. Due to the speciality of this sequence, the search always came back with a mix of both TPI and GAPDH results. Therefore, the results were separated into those two.

5.1.1. TPI super family

Performing an NCBI protein search using *Blastocystis* TPI-GAPDH fusion protein as a query came back with results from all members in the living kingdom, since it is a widely abundant protein.

Triosephosphate isomerase belongs to the ($\beta\alpha$) 8 barrel enzyme family (TIM barrel family), a wide group of enzymes performing several functions. This enzyme doesn't need a metal ion or co-factor for its catalysis, its catalysis is acid-base catalysis, which means special residues are responsible for the catalytic reaction to happen ([Nagano, Orengo et al. 2002](#)). The TIM barrel family has remarkable residues important for the proteins activity. When comparing the sequence of *Blastocystis* TPI-GAPDH fusion protein with other TPI proteins, it was found that the fusion protein has kept all TIM barrel special sequences and

active site. Three residues are important for the catalytic activity, and the fourth is important for the right orientation of the lysine residue in the active site ([Knobeloch, Schmidt et al. 2010](#)). The four residues [the substrate binding asparagine (N) and lysine (K) in addition to the Histidine (H) and glutamic acid (E)] playing an important role in substrate binding and active site catalysis were found to be conserved in all the search proteins, see Figure 5.1. Another important feature, loop 6 connecting β strand 6 with α helix 6 and forming a lid that covers the active site residues during the catalytic reaction, taking the sequence AYEPIWAIGTG which is important in protein activity was also highly conserved in *Blastocystis* and the other search organisms, Figure 5.1. Organisms used in this search are all mentioned in Figure 5.1 with their corresponding accession numbers or PDB codes when available.

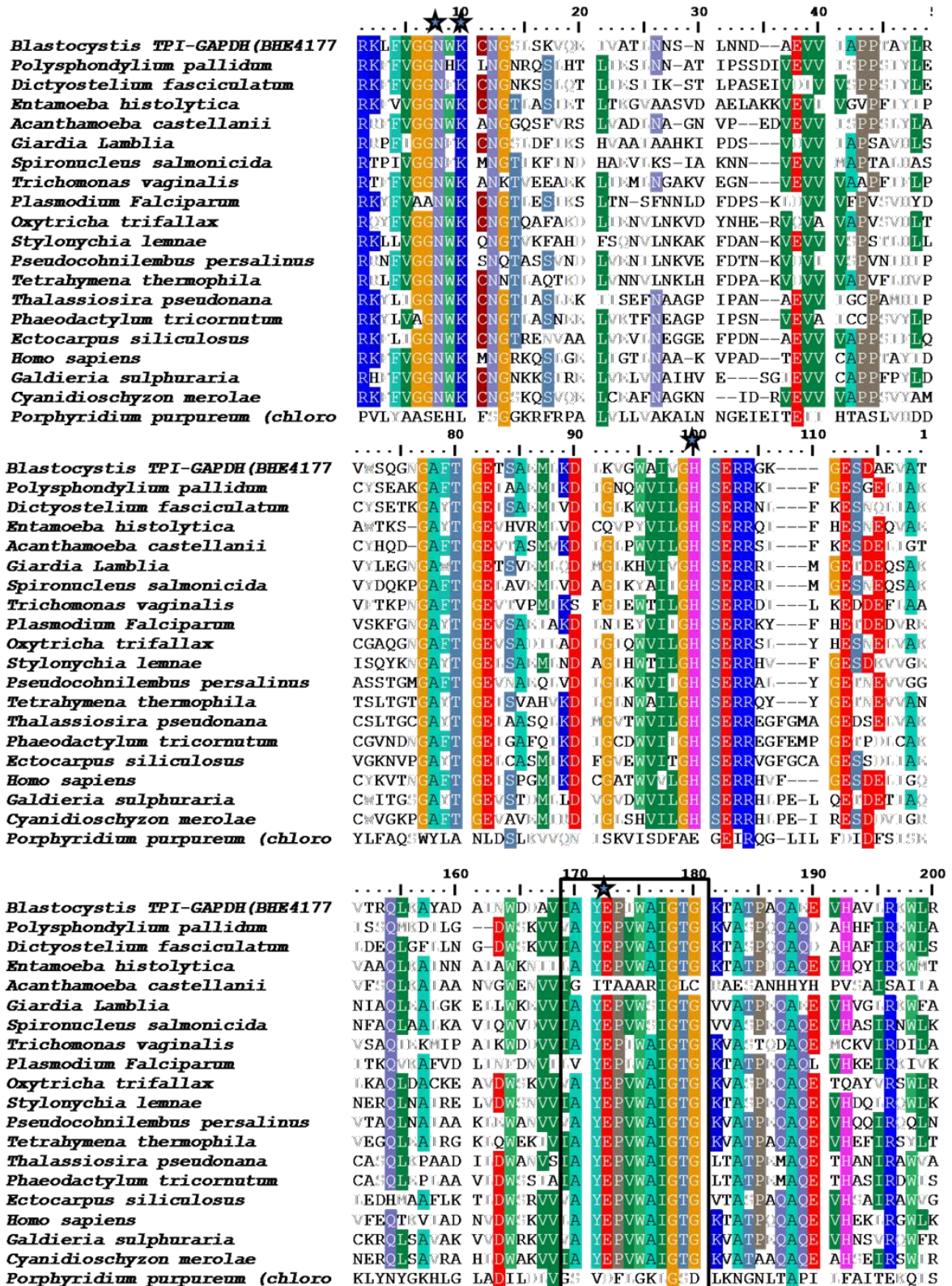


Figure 5.1 Comparison of the amino acid sequences of triosephosphate isomerase part of the *Blastocystis* TPI-GAPDH fusion protein with various organisms. The asterisks refer to the amino acids playing important role in protein activity. Region in box refers to the amino acid sequence covering the flexible loop region. Accession numbers or PDB codes are as follows from top to bottom: EFA82625.1, XP_004357462.1, BAN38145.1, XP_004333360.1, pdb|4B17|A, EST45141.1, XP_001320336.1, pdb|4Z0J|A, EJY68394.1, CDW88838.1, KRX01127.1, XP_001008794.2, XP_002295570.1, XP_002181761.1, CBJ32973.1, NP_000356.1, XP_005706061.1, XP_005538108.1 and YP_008965787.1.

5.1.2. GAPDH super family

Same search was performed using *Blastocystis* TPI-GAPDH fusion protein sequence on NCBI server against representatives of members of the living system. The result came back with mixed TPI and GAPDH due to the nature of the query. For this reason, only results containing the stand alone GAPDH protein were chosen to perform the comparison.

Protein sequences were aligned using ClustalW multiple alignment ([Larkin, Blackshields et al. 2007](#)). The alignment was investigated by eye, and vaguely aligned regions were excluded. The alignment showed that GAPDH region of the fusion protein is highly conserved among other species and *Blastocystis*. The main regions detected were two. The first one is the active site region taking the amino acid sequence (ASCTTNCL). This sequence covers the active site of GAPDH, and it was conserved as Figure 5.2 shows. Another important feature in GAPDH sequence is a signature motif indicating the protein binding ability. GAPDH Rossmann fold can either bind the cofactor NAD⁺ or NADP⁺ to its active site, in what is called dual co-substrate specificity. A consensus N-terminal sequence, which is Glycine-rich phosphate binding loop, was found to play an important role in the determination of cofactor binding. It takes the following sequence GXGXXGX. Replacing Glycine with alanine will change the binding from NAD⁺ to NADP⁺ ([Hanukoglu 2015](#)). In the *Blastocystis* TPI-GAPDH fusion

protein it was clear that the cofactor is NAD⁺, see Figure 5.3.

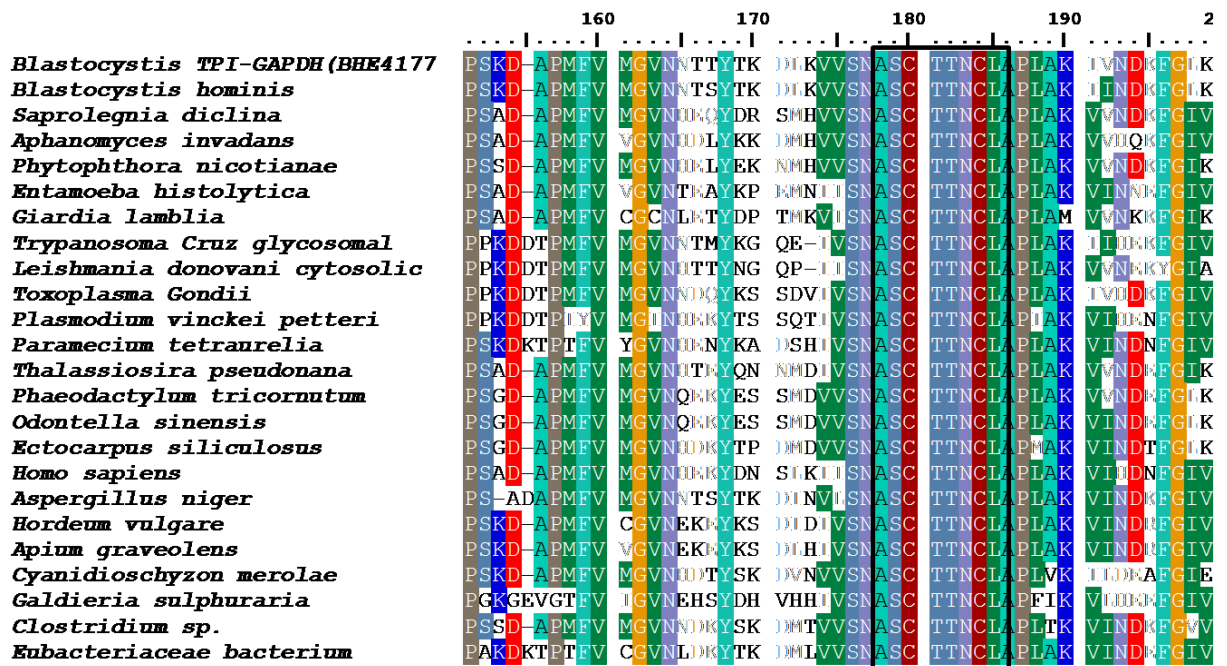


Figure 5. 2 The active site region of *Blastocystis* TPI-GAPDH and other glycolytic GAPDH proteins from different species. The residues in the box show the sequence motif indicative of loop 6 area. The organism's accession number or PDB codes are as follows from top to bottom. Query protein, XP_012894401.1, XP_008620095.1, XP_008861346.1, ETO79873.1, BAN38604.1, EFO63427.1, PDB 4LSM, XP_003865363.1, PDB 3STH, EUD71497.1, XP_001445566.1, XP_002290129.1, XP_002177670.1, AAU81890.1, CBJ32822.1, CAA25833.1, GAQ46613.1, BAJ90709.1, AMT91855.1, XP_005537058.1, XP_005704189.1, WP_004033988.1 and CZT56408.1.

	10	20	30	40
<i>Blastocystis TPI-GAPDH (BHE4177)</i>	NIINANS--T	VQNAGPVS	VG I I G F G R I G R L	VFRASQPNPL ----VNVVA
<i>Blastocystis hominis</i>	NIINANS--T	VKNAGPVS	VG I I G F G R I G R L	VFRASQPNPL ----VNVVA
<i>Saprolegnia diclina</i>	NIINCQNDAA	KSVGGPVNV	VG I I G F G R I G R L	VLRAAEKNPL ----INVVA
<i>Aphanomyces invadans</i>	HIINAQGGGT	TNAGGPVNV	VG I I G F G R I G R L	VLRAAEANPL ----INVVA
<i>Phytophthora nicotianae</i>	HIINAQN-PT	TNVGAVNV	VG I I G F G R I G R L	VLRAAAKNPL ----INIVA
<i>Entamoeba histolytica</i>	-----	---MSIKVG	I I G F G R I G R L	VARVLER-K D---FKIVA
<i>Giardia lamblia</i>	-----	---MPRIIG	I I G F G R I G R L	ALRAALNIEG ----VQVVA
<i>Trypanosoma Cruz glycosomal</i>	MAHHHHHMT	EGKKVTIKIG	I I G F G R I G R L	VFRAAQHRND ----IKTVG
<i>Leishmania donovani cytosolic</i>	-----	---MVKVG	I I G F G R I G R L	VFRVAQTRPD ----IKTVG
<i>Toxoplasma Gondii</i>	HMGTTLEAQTQ	GP GSMVCKIG	I I G F G R I G R L	VFR AAMERGD ----VNVVA
<i>Plasmodium vinckei petteri</i>	-----	---MAIKVG	I I G F G R I G R L	VFRVAQERSD ----INIVA
<i>Paramecium tetraurelia</i>	VFWLSKYKYL	LSLQMSVTIG	I I G F G R I G R L	VLRAAIENNK ---PVTVKA
<i>Thalassiosira pseudonana</i>	-----	---MSINIG	I I G F G R I G R L	VMRAAAKNPN ----VNTVA
<i>Phaeodactylum tricornutum</i>	-----	---MSINVG	I I G F G R I G R L	VMRAAQKNPN ----IKTVA
<i>Odontella sinensis</i>	-----	---MSINIG	I I G F G R I G R L	VMRAAKGRSD ----INVVA
<i>Ectocarpus siliculosus</i>	-----	---MPVNV	VG I I G F G R I G R L	VMRAAKNNKN ----INIVA
<i>Homo sapiens</i>	-----	---MGKVKVG	V I G F G R I G R L	VTRAAFNS-G K---VNVVA
<i>Aspergillus niger</i>	-----	---MAPKVG	I I G F G R I G R L	VFRNAINHGK ----VNVVA
<i>Hordeum vulgare</i>	-----	---MAPIKIG	I I G F G R I G R L	VARVALQCPD ----VNTVA
<i>Apium graveolens</i>	-----	---MAPIKIG	I I G F G R I G R L	VARVALQRDD ----VNTVA
<i>Cyanidioschyzon merolae</i>	-----	---MAIPKVG	I I G F G R I G R L	VLRAAVEKGT ----VNVVA
<i>Galdieria sulphuraria</i>	LSSKTANQR	WVMQAKVNV	VG I I G F G R I G R L	FLRCWATRPS S---HIIVVA
<i>Clostridium sp.</i>	-----	---MVKVG	I I G F G R I G R L	VFRASLSRTD ----INVVA
<i>Eubacteriaceae bacterium</i>	-----	---MGIKVG	I I G F G R I G R L	VFRAAVAQPK K---FNVVA

Figure 5. 3 Conserved region in *Blastocystis* TPI-GAPDH and other glycolytic GAPDH proteins from different species. The residues in the box show the GXGXXGX sequence motif indicative of GAPDH binding NAD⁺. The organism's accession number or PDB codes are as follows from top to bottom. Query protein, XP_012894401.1, XP_008620095.1, XP_008861346.1, ETO79873.1, BAN38604.1, EFO63427.1, PDB 4LSM, XP_003865363.1, PDB 3STH, EUD71497.1, XP_001445566.1, XP_002290129.1, XP_002177670.1, AAU81890.1, CBJ32822.1, CAA25833.1, GAQ46613.1, BAJ90709.1, AMT91855.1, XP_005537058.1, XP_005704189.1, WP_004033988.1 and CZT56408.1.

5.2. *In silico* characterization of *Blastocystis* TPI-GAPDH; pre-sequence and signature motif

5.2.1. TPI-GAPDH Pre-sequence analysis

The *Blastocystis* TPI-GAPDH fusion protein is composed of an intact moiety of TPI at its N-terminal, attached with a small hinge region with a complete intact GAPDH moiety. TPI's C-terminal is attached to GAPDH'S N-terminal immediately. And the whole fusion protein has an N-terminal which is a targeting signal. Liaud *et al*/ has shown that this protein has a mitochondrial targeting signal

in diatoms, and they proved it with electron microscopy experiments. In their papers, ([Liaud, Lichtle et al. 2000](#)) and ([Nakayama, Ishida et al.](#)) both hypothesized that this is widely spread in the Stramenopiles. Nakayama *et al* have referred to the presence of these mitochondrial proteins in several other glycolytic proteins, including the second half of glycolysis pathway.

To test this hypothesis in *Blastocystis* and since it only possess the fused version of the protein, the targeting signal of *Blastocystis* TPI-GAPDH and same proteins of several groups were evaluated.

A BLAST search was performed using *Blastocystis* TPI-GAPDH fusion protein, and was divided into two groups as mentioned before in 5.1.1. and 5.1.2. sections.

This search resulted in three groups; proteins representing TPI stand-alone version only, protein representing GAPDH stand-alone version only, and the third group was limited to the organisms which showed a fusion protein as the protein of query.

The mitochondrial targeting signal was evaluated in each group using four predicting programs; MitoProt ([Claros 1995](#)), iPSORT ([Bannai, Tamada et al. 2002](#)), Predotar ([Small, Peeters et al. 2004](#)) and TargetP ([Emanuelsson, Brunak et al. 2007](#)). Only those organisms whose protein showed to be targeting the mitochondria in at least three out of the four programs, with a high value (> 0.7%) were considered to be mitochondrial localized proteins to avoid false positive results.

The group containing TPI proteins only had shown that only *Blastocystis* TPI-GAPDH is targeting the mitochondria. The rest were all negative for targeting the mitochondria.

On the other hand, the search results representing GAPDH proteins had shown almost a similar result; proteins that belong to the Chromalveolates were all positive towards having a mitochondrial targeting signal, with a high support value. Inspecting those sequences by eye revealed that those proteins were all a full version of the fusion protein, and not only GAPDH as the NCBI blast search showed. In addition to the full TPI-GAPDH proteins, were unusual localization reported in *Homo sapiens* and members of diatoms as well including (*Thalassiosira pseudonana*, *Phaeodactylum tricornutum*, *Odontella sinensis*), Figure 5.4.

The BLAST search using the fusion protein as a query resulted in a group of fusion proteins as well. Those proteins mostly belonged to the chromalveolates represented by diatoms, Oomycetes, Ciliata and Phaeophytes, Figure 5.5.

These results are in agreement with what was reported in Liaud paper, of spotting fusion protein of TPI-GAPDH in diatoms. It also supports the findings of Nakayama of a wider spread of the fusion protein in stramenopiles

Those mitochondrial targeting signals were all basic, positive residues, rich in arginine residues specially at R-3 and R-2 from the cleavage site, short signals ranging between (12-22) which all in accordance with the mitochondrial targeting signal specifications as discussed in details in chapter 3, section 3.2.1. and in agreement with what was already published about these kind of signals ([Gavel and von Heijne 1990](#), [vanderGiezen, Rechinger et al. 1997](#)).

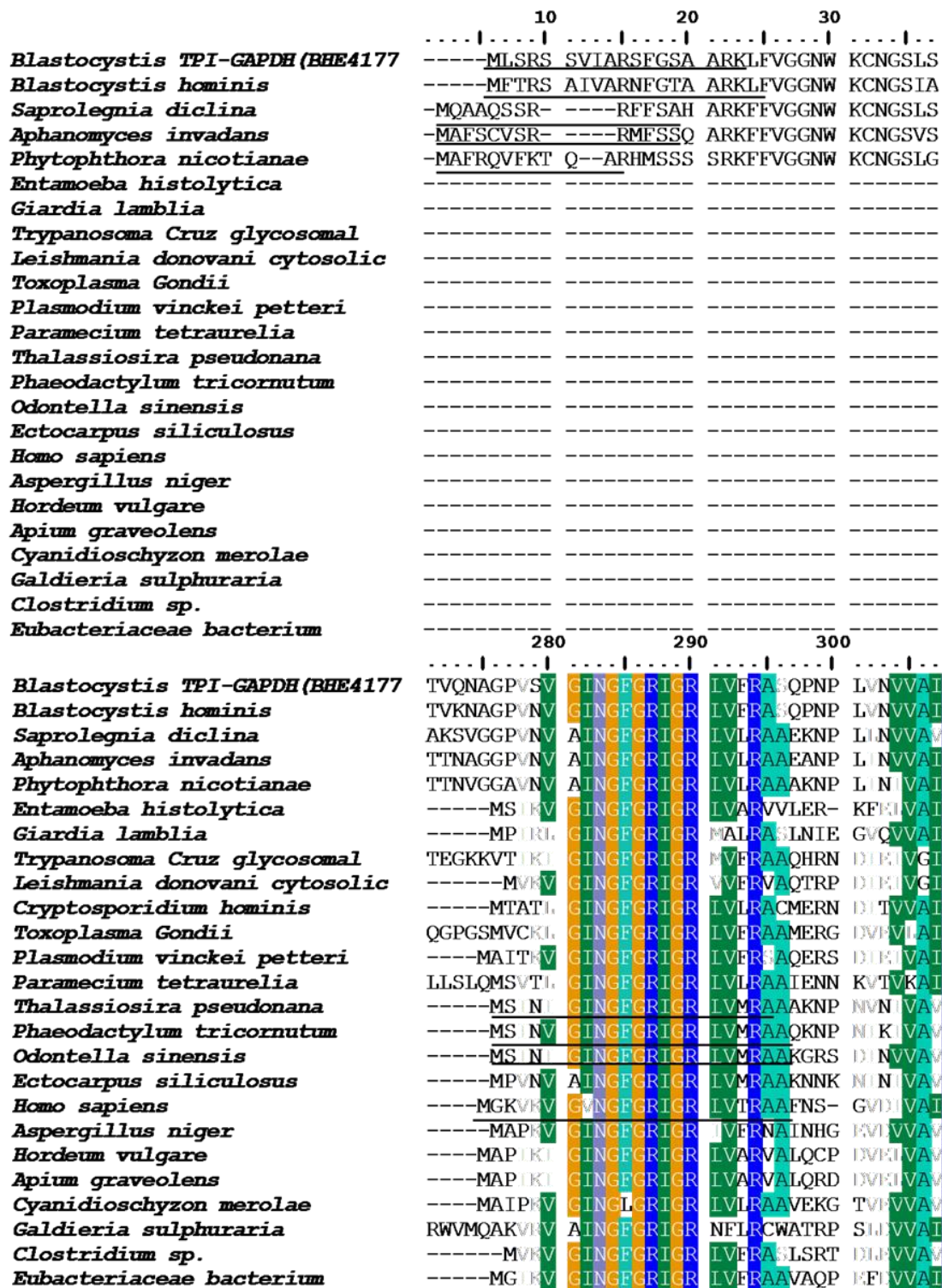


Figure 5. 4 GAPDH protein sequence alignments using the *Blastocystis* TPI-GAPDH as query sequence on BLAST search, and representing members covering all groups of life. Mitochondrial targeting signals are underlined. Accession numbers and PDB codes are as follow from top to bottom: XP_012894401.1, XP_008620095.1, XP_008861346.1, ETO79873.1, BAN38604.1, EFO63427.1, 4LSM, XP_003865363.1, PDB 3STH, EUD71497.1, XP_001445566.1, XP_002290129.1, XP_002177670.1, AAU81890.1, CBJ32822.1, CAA25833.1, GAQ46613.1, BAJ90709.1, AMT91855.1, XP_005537058.1, XP_005704189.1, WP_004033988.1 and CZT56408.1.

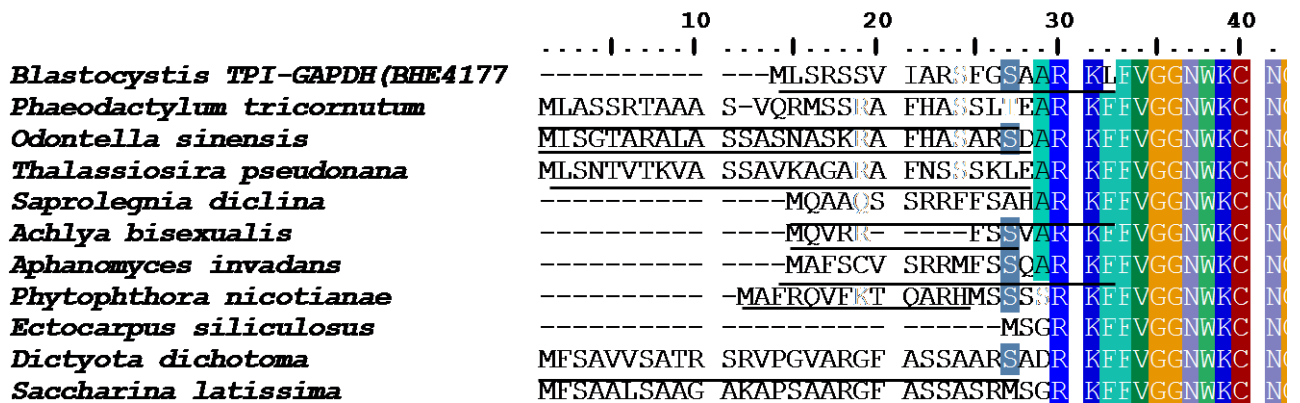


Figure 5. 5 Fusion protein aligned with other fusions using *Blastocystis* TPI-GAPDH as a query. Showing the wide spread of the fusion proteins in chromalveolates. The mitochondrial targeting signals are underlined. Accession numbers and PDB codes are as follow from top to bottom: XP_012894401.1, AAF34330.1, AAF34328.1, XP_002290574.1, XP_008620095.1, AAF44720.1, XP_008861346.1, ETO79873.1, CBJ30340.1, ABU96664.1, and ABU96661.1.

5.3. Phylogentic relationships of *Blastocystis* TPI-GAPDH and other organisms

Blastocystis TPI-GAPDH fusion protein was aligned with protein sequences of TPI and GAPDH proteins, separately, from a broad spectrum of eukaryotes and bacteria. Another alignment was also made between stramenopiles members for both proteins as well, and for the fusion proteins reported by the BLAST search. Highly diverged and vaguely aligned sequences were removed manually from the alignment before the phylogenetic analysis. A complete list of all organisms used in this analysis is found in. Each set of sequences was aligned using MUSCLE (Edgar 2004). MUSCLE will detect any sequence repetitions and gaps and will remove them automatically. Regions of gaps or those that are not clearly alignable in all sequences were excluded from the phylogenetic analyses by masking the alignment manually. A phylogenetic tree is then

constructed using those edited sequences. Each alignment was analyzed using PhyML (maximum likelihood).

The phylogenetic search was divided into three groups in this part due to the nature of the sequences submitted and to the fact that the BLAST search return with a mixed results containing TPI proteins, GAPDH proteins, and a smaller group containing the fusion protein.

Phylogeny tree constructed from either TPI or GAPDH only, were of a low support value. In the case of tree constructed using TPI protein results only, Figure 5.6 shows that *Blastocystis* TPI-GAPDH fusion protein is not clustering in a group with other TPI protein-containing organisms, and that it is evolving earlier than the most neighbouring organisms.

The same procedure was performed using BLAST results containing GAPDH only, and as Figure 5.7 demonstrating, *Blastocystis* TPI-GAPDH fusion protein is clustering in the same group with organisms belonging to the diatoms, oomycetes, and phaeophytes. Interestingly, it is found that *Blastocystis* TPI-GAPDH is branching closely to eubactericae as well. Finally, the tree constructed using the fusion protein results only, showed clearly that the organisms containing this fusion belong to the chromalveolates and specially diatoms, oomycetes, ciliata and phaeophytes, with a very high support value, Figure 5.8.

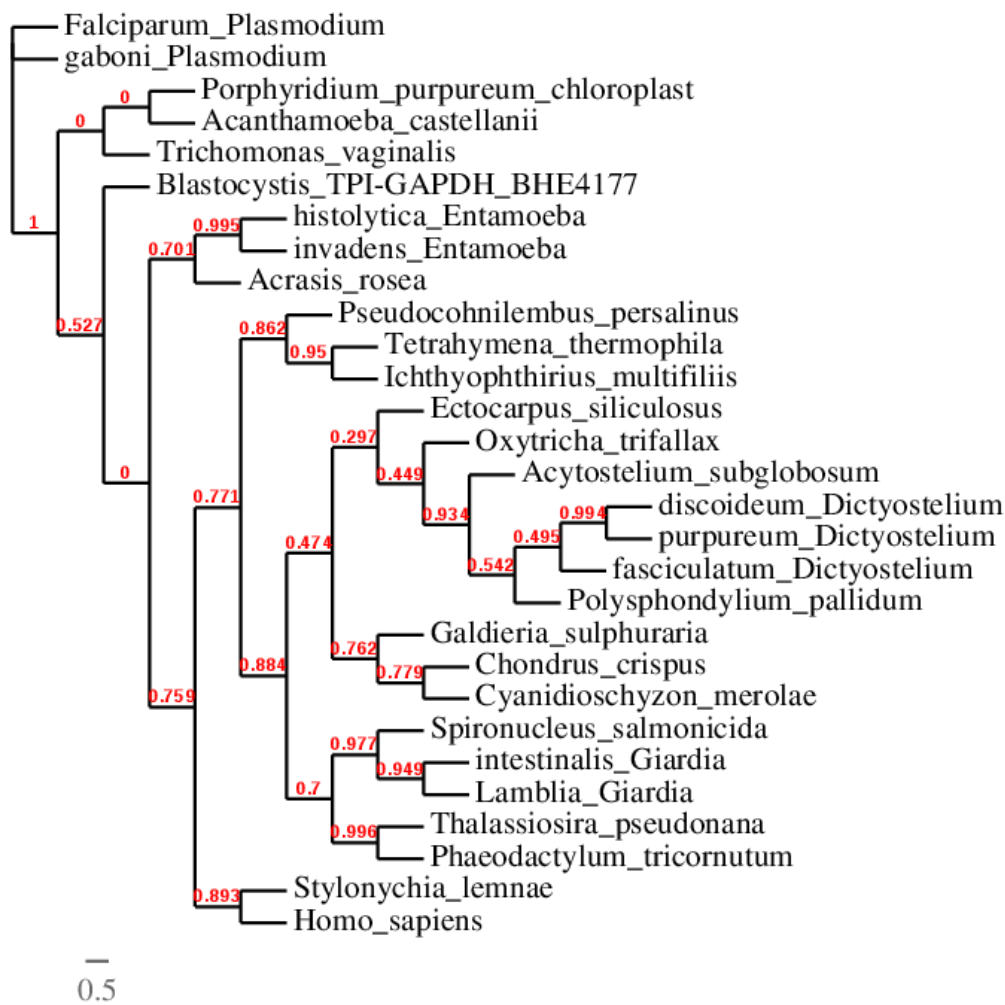


Figure 5.6 Relationship of *Blastocystis* TPI-GAPDH fusion protein among eukaryotes and prokaryotes as determined by phylogenetic analysis of protein sequence representing triosephosphate isomerase only. Shown is a protein maximum-likelihood topology constructed by the PhyML program. Bootstrap values and posterior probability values were calculated using PhyML. The dataset contained 29 from all super-groups of eukaryotes and prokaryotes. A complete list of taxa and their accession numbers can be found in Table 3.3.

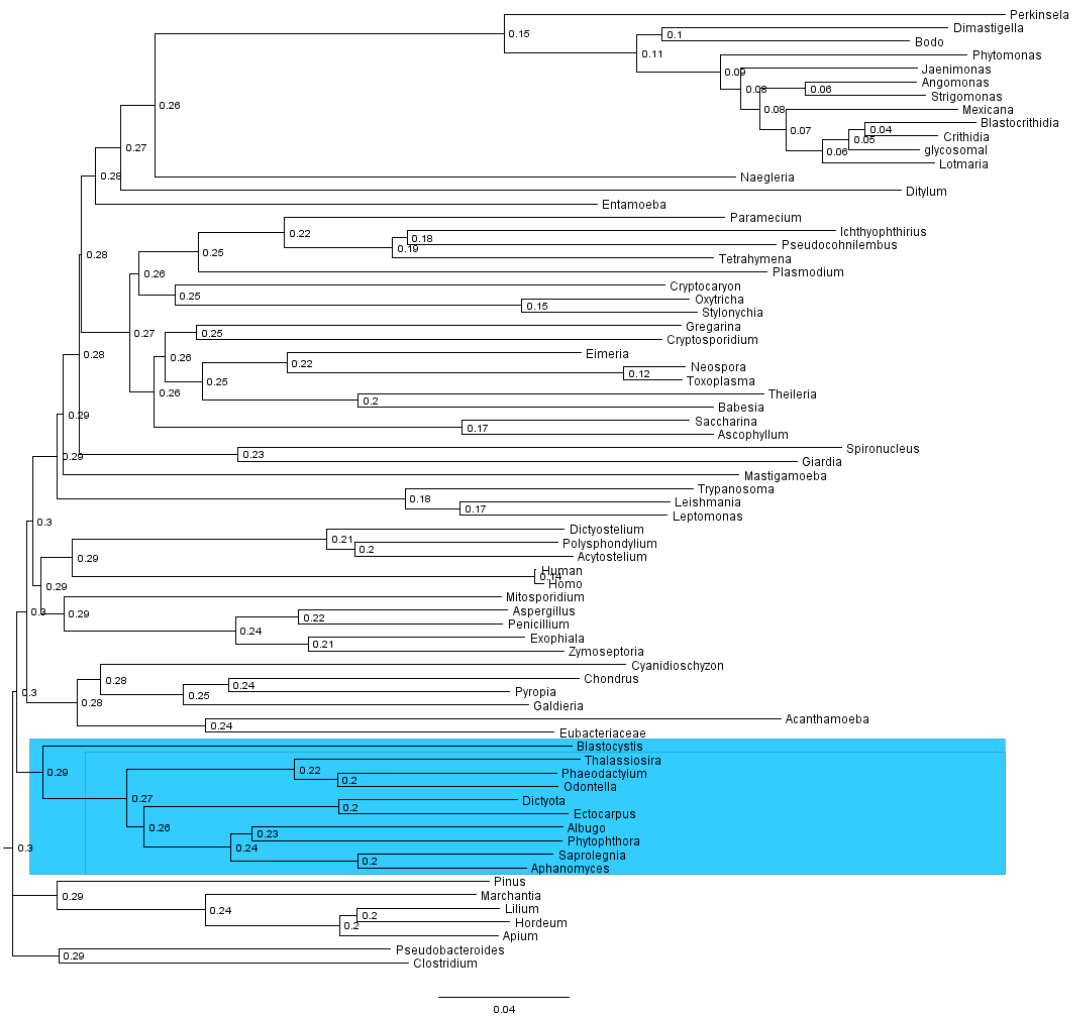


Figure 5. 7 Relationship of *Blastocystis* TPI-GAPDH fusion protein among a group of organisms containing GAPDH protein only determined by phylogenetic analysis of protein sequence. Shown is a protein maximum-likelihood topology constructed by the PhyML program. The dataset contained 13 organisms. A complete list of taxa and their accession numbers can be found in Table 3.3.

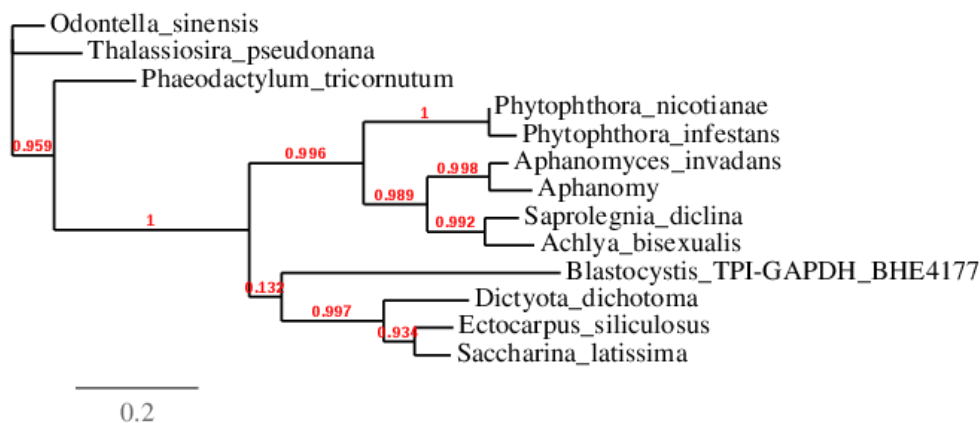


Figure 5. 8 Relationship of *Blastocystis* TPI-GAPDH fusion protein among a group of organisms containing the complete fusion protein only determined by phylogenetic analysis of protein sequence. Shown is a protein maximum-likelihood topology constructed by the PhyML program. The dataset contained 13 organisms. A complete list of taxa and their accession numbers can be found in Table 3.3.

5.4. Conclusion

In silico analysis of the *Blastocystis* TPI-GAPDH fusion protein clearly shows a highly conserved protein in general. The sequences analysis of the two parts of the protein have presented the important features and signatures identifying both TIM barrel family, to which triosephosphate isomerase (TPI) protein belongs to, and the same was noticed for the glyceraldehydes-3-phosphate dehydrogenase (GAPDH) part of the fusion, which presented with the Rossmann fold with its features.

Interestingly this unusual protein has shown a special targeting signal. Unlike all glycolytic enzymes, the fusion protein in this study had presented with a mitochondrial targeting peptide with a high support value. This targeting signal was found in other organisms containing the fusion protein as well. Those organisms belonged basically to the stramenopiles, which is quite an interesting finding. The same phenomenon was previously reported in the work of Liaud *et al* where they noticed the presence of the mitochondrial targeting signal in a group of diatoms ([Liaud, Lichtle et al. 2000](#)).

Phylogenetic analysis was performed to find where this protein originates from and how did it evolved. Interestingly, the tree constructed using the fusion protein sequence only, has shown a clear bias of those proteins to belong to the stramenopiles as well. This has been reported before in the work of Nakayama *et al*, in their paper 2012, they reported a broad distribution of this unusual enzyme in stramenopiles as well.

It is difficult to trace the origin of this fusion protein, since neither TPI nor GAPDH trees give a clear resolution of this idea. However and quite interestingly, *Blastocystis* TPI-GAPDH had been very close to the bacterial GAPDH in the tree

constructed using GAPDH sequences only. This suggests that they are closely related and they share the same origin. This notion supports the hypothesis that the modern-day mitochondria are originating from α -proteobacterial ([Cavalier-Smith 1987](#)) that has lost a lot of its genes through evolution both gene loss and nuclear transfer ([Martin, Hoffmeister et al. 2001](#)). Many of those transferred genes encode proteins that target the mitochondria. However there is no strong proof that those enzymes studied so far are of α -proteobacterial ancestry, but the existence of those glycolytic enzymes targeting signal to the mitochondria and others indicated in ([Nakayama, Ishida et al.](#)) gives a large motivation to investigate this more in organisms and study the localization *in vivo* as well as *in silico*.

Chapter 6. Biochemical characterisation of the unusual TPI-GAPDH fusion protein from *Blastocystis*

6.1. General background (introduction)

The conservation of energy is a vital process for all organisms, and several pathways are made for this purpose. The glycolysis pathway is one of the most common ways to conserve energy, and glycolytic proteins are one of the most ubiquitous enzymes in eukaryotes. This pathway usually takes place in the cytosol where its end product pyruvate is transferred to the mitochondria to be utilized in Krebs cycle. This scenario is quite common that it is presented in most of the text books when explaining the glycolytic process. However, some organisms have a different localization of their glycolysis pathway. Trypanosomes have special organelles called glycosomes which are micro body-like organelles containing glycolytic enzymes ([Misset, Bos et al. 1986](#), [Michels 1988](#)). These trypanosomal glycosomal-localized glycolytic enzymes have been reported to possess special properties that are not found in the cytosolic Trypanosomal enzymes, and not even in the enzymes of other organisms and that are not localized to a special compartment. This made these proteins interesting as potential targets for drug development targeting the glycosomal glycolytic proteins.

Recently, another unusual glycolysis cellular localization was discovered, taking place in a group of eukaryotes called stramenopiles. It was suggested that at least the second half of the glycolytic enzymes are targeted to the mitochondria, which is the first unusual event reported in this context. Yet fewer researchers had drawn the attention to the presence of an unusual protein formed from the fusion of two consecutive proteins, TPI-GAPDH. Unlike what was previously

reported about glycolysis of the presence of fusions between non-consecutive enzymes like enolase and GAPDH ([Takishita, Patron et al. 2005](#)).

Triosephosphate isomerase (TPI) is a widely studied enzyme and it is well known for its role in the catalysis of the isomerization between glyceraldehydes-3-phosphate (GAP) and dihydroxyacetone phosphate (DHAP) which are formed from fructose 1, 6-bisphosphate that is produced by the aldolase enzyme. GAP is then utilized in the glycolysis pathway by glyceraldehydes-3-phosphate dehydrogenase (GAPDH) which is the following enzyme on this pathway. DHAP can be used in the production of glycerine phosphate in lipid synthesis. GAPDH utilizes the other product of the TPI reaction, GAP, as a substrate in the reversible reaction converting GAP to 1, 3-diphosphoglycerate.

Both TPI and GAPDH are well studied enzymes. Although almost universal, they do show some degree of divergence in their sequences, making them interesting candidates for drug development studies.

In this instance, it was reported that TPI and GAPDH, together form a fusion protein that is coded by one gene, where the C-terminal of TPI is upstream the N-terminal of GAPDH, with no stop codon at the end of TPI. This fusion protein was reported abundantly in protists and specifically stramenopiles such as diatoms, photosynthetic protists that have plastids to which the fusion protein is targeted ([Liaud, Lichtle et al. 2000](#), [Nakayama, Ishida et al.](#)).

Genomic studies in *Blastocystis* have shown the presence of a similar fusion containing TPI and GAPDH on the same open reading frame (ORF). This gene seems to have a mitochondrial targeting signal, which makes it an interesting subject to be investigated for two reasons: firstly because it defies the general notion that glycolysis is localised in the cytosol, and secondly because

Blastocystis is an anaerobic organism with mitochondria and mitochondrial glycolytic enzymes might suggest unexpected metabolic adaptations of this parasite which could perhaps be exploited as drug targets..

The TPI-GAPDH fusion protein is not fully characterized yet, apart from a paper in 2000 that investigated its molecular weight and oligomeric status using molecular biology techniques ([Liaud, Lichtle et al. 2000](#)). In 2014 an MSc report was written at the University of Bedfordshire which looked into the kinetics of the recombinant *Blastocystis* TPI-GAPDH fusion protein (using clones provided by our group, the same as used in this study). That study somewhat inconclusive regarding the oligomeric status of the fusion protein ([Evans 2014](#)).

We anticipate that the TPI-GAPDH fusion protein performs a function similar to the sum of the individual glycolytic enzymes. However, no conclusive data to support this notion currently exists. As TPI and GAPDH normally have different multimeric stoichiometries (TPI is normally a homodimer while GAPDH is a homotetramer), it would be interesting to understand the multimeric arrangement of this unusual enzyme. In addition, is there any benefit from having a fusion enzyme compared to having two individual enzymes?

In this project I sought to characterize this unusual fusion protein biochemically and biophysically in order to address these questions. Any difference with the host enzymes would make the *Blastocystis* TPI-GAPDH fusion a potential candidate for drug development against *Blastocystis* in the future.

Evans et al (2014) suggested from their preliminary size exclusion chromatography data that the fusion protein is composed of a dimer of TPI and a dimer of GAPDH ([Evans 2014](#)). On the other hand, Liaud et al suggested the fusion to be a tetramer of GAPDH and two dimers of TPI ([Liaud, Lichtle et al.](#)

[2000](#)). The work in this chapter was sought to biochemically characterize the *Blastocystis* fusion protein. In order to do this the protein needed to be purified to a high level. This might ultimately lead to being able to crystalize this unusual fusion and help to unravel the three-dimensional structure of this interesting fusion protein.

6.2. Recombinant *Blastocystis* TPI-GAPDH was successfully cloned and over-expressed in *Escherichia coli*

6.2.1. Cloning and amplification of *Blastocystis* TPI-GAPDH (B.TPI-GAPDH)

The gene encoding the recombinant *Blastocystis* TPI-GAPDH fusion protein was previously cloned in pET-14b and stored in our group stocks. The gene was originally amplified using TPI-GAPDH pET F and TPI-GAPDH pET R primers mentioned in Table 2.2 which introduce *Nde*I and *Bam*HI restriction sites. Cloning in these sites introduced an N-terminal His-tag and allows for expression under the control of the T₇ promoter. The gene was amplified from clones of the previously stored construct (see Figure 6.1). A mix of enzymes was used in the PCR reaction including proofreading *Pfu* DNA polymerase and *Taq* polymerases as mentioned in chapter 5, section 4.2.1. The gene construct was confirmed by the DNA sequencing facility at MWG - Eurofins using T7 and T7 term primers of the pET-14b plasmid referred to in Figure 2.4 and which sequence is mentioned in Table 2.11.

Genes incorporated in pET-14b vector were transferred into several *E. coli* cell lines, including *BL21*, Arctic express and Rosetta2 cells to identify optimal protein production conditions.

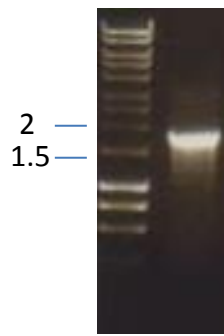


Figure 6. 1 Agarose gel electrophoresis analysis of PCR-amplified TPI-GAPDH from *Blastocystis* ST4. Adding *Nde*I and *Bam*HI restriction sites. The gel shows a band of the amplified recombinant *Blastocystis* TPI-GAPDH with an expected size of 1.86 kb. DNA HyperLadder I is on the left side lane of the PCR product. Sizes of relevant markers are indicated in kb (Bioline).

Different cloning approaches were used at the Oxford Protein Production Facility laboratories in an attempt to examine the effects of changing the vector and amplification enzymes on the final protein product. New vectors and amplifying enzymes were tried. A new vector was tried to assess its impact on gene expression. The standard vector adopted by Oxford University/Oxford Protein Production Facility (OPPF) department is developed by them. These are pOPIN F 16417 and pOPIN E 16418. Primers used were kindly offered by the OPFF lab. The recombinant *Blastocystis* TPI-GAPDH gene was amplified using the Phusion Flash enzyme (Life Technologies) and the KOD Xtreme™ Hot Start DNA Polymerase (Merck Mellipore), as both give higher fidelity than the Phusion enzyme (Figure 6.2). Constructs made in pOPIN E and F were not further used afterwards in this project and only the recombinant fusion *Blastocystis* TPI-GAPDH pET-14 construct in *E. coli* BL21 was used for downstream experiments.

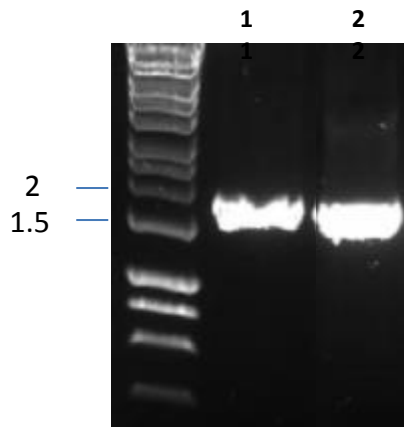


Figure 6. 2 Agarose gel electrophoresis analysis of PCR-amplified *Blastocystis* TPI-GAPDH with pOPIN E and pOPIN F primers, using both Phusion flash (lane 1) and KOD Xtreme (lane 2) amplifying enzymes respectively. The gel shows a band of the expected size of 1.86 kb. DNA HyperLadder I (Bioline) is on the left side lane of the PCR product. The sizes of the relevant reference DNA are indicated next to the band.

6.2.2. Recombinant *Blastocystis* TPI-GAPDH Expression trials; recombinant fusion protein TPI-GAPDH was soluble in all construct and in abundance in auto-induced BL21 cells.

In order to identify the optimal gene expression conditions that would result in maximum amounts of soluble protein, several parameters were varied. All *Blastocystis* TPI-GAPDH constructs were incorporated in the pET-14b vector. Several *E. coli* cell lines were used including BL21, Rosetta 2 and Arctic Express cells. Expression conditions have also been varied. The IPTG concentration has been varied and auto-induction has been used as well. In addition, incubation temperatures, media used, and time of induction have been changed.

The TPI-GAPDH pET-14b construct was transferred into *E. coli* Rosetta, BL21 and Arctic Express cells and induced at 37, 30, 25 and 18°C for Rosetta and BL21 *E. coli* cells and 10 °C for Arctic express *E. coli* cells. All of which gave a very low yield of the protein using different IPTG concentration of (1-400 mM). A few examples of the several expression trials conducted to find the most favourable conditions for gene expression and yield are shown in Figure 6.3.

As using various conditions of IPTG to induce protein did not result in satisfactory yields, auto-induction was used in the hope this would solve the poor yield obtained thus far. Interestingly, the gene was highly expressed under this condition (20°C, auto-induction for 48 hours in *E. coli* BL21 cells). This method depends on the role of lactose in the media that will induce the growth of cells to a much higher level than that achieved in normal induction over expression methods using IPTG. This role will start only after the high density media is depleted from the inhibiting factors, ([Studier 2005](#)). For further explanation of this approach please see chapter 4, section 4.2.2.

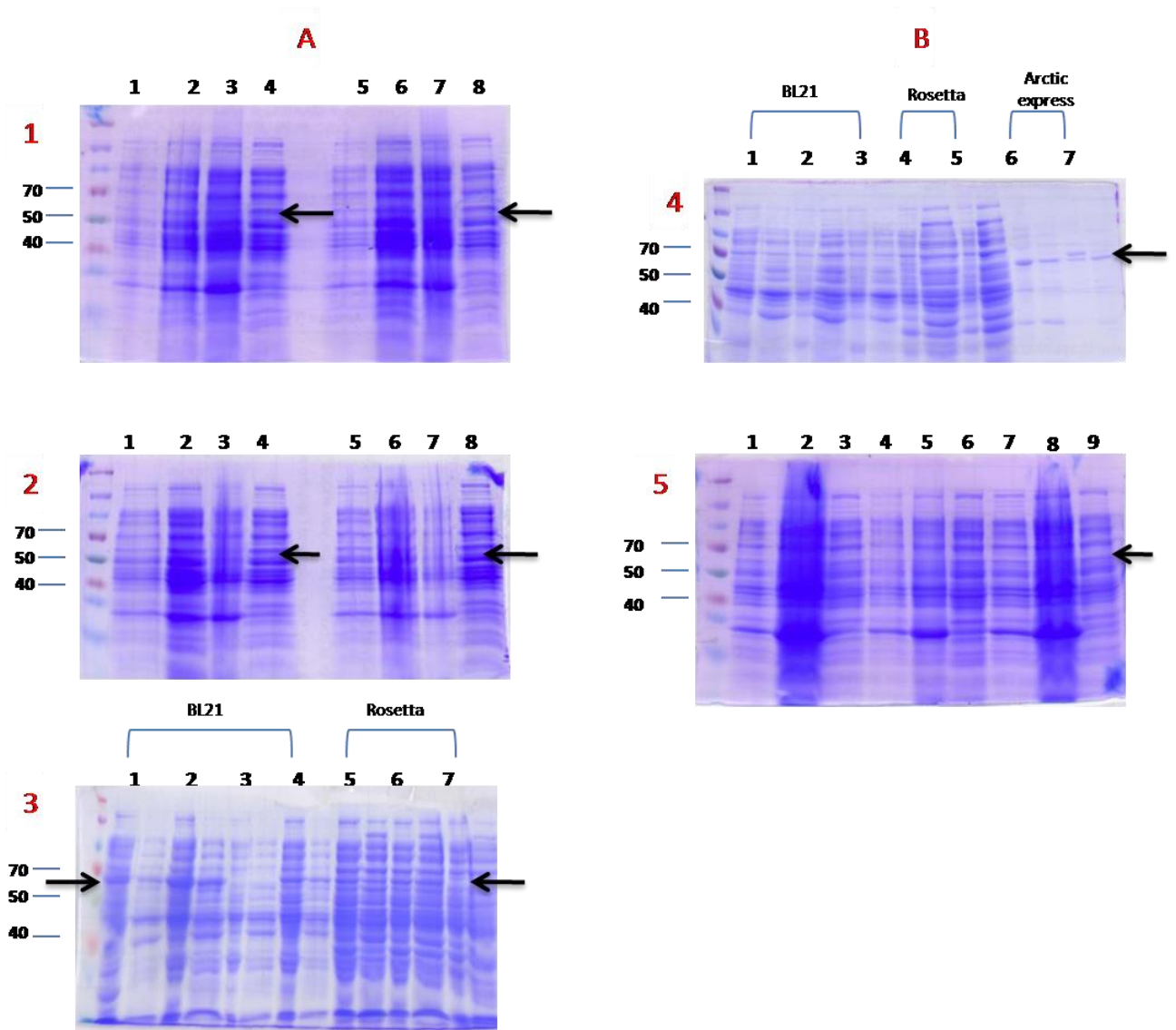


Figure 6.3 Expression trials for *Blastocystis* TPI-GAPDH in *E. coli* cell lines under several conditions. A) TPI-GAPDH cloned into pET-14b in BL21 *E. coli*, was induced at 37°C at different optical densities and different IPTG concentrations where:

- 1) Gene expression induced with 1 mM IPTG at ODs values of 0.8 and 1 respectively. The lanes are as follows respectively: 1, 5. uninduced, 2, 6. induced cells cooled down for 4-5 hours for better folding of the protein (4⁺°C), 3,7 insoluble and 4,8. Soluble fractions.
- 2) Induction with 1 mM IPTG at ODs values of 0.4 and 0.6 respectively. The lanes are as follows respectively: 1, 5. uninduced, 2, 6. induced cells cooled down for 4-5 hours for better folding of the protein (4⁺°C), 3, 7 insoluble and 4, 8 soluble fractions.
- 3) Induction with 400 mM for 4 hours, where: 1: OD=1, 2: OD= 1.2, 3: OD= 1.3, 4: OD= 0.75, 5: OD= 1, 6: OD= 1.2, 7: OD= 1.3. Each number is composed of both insoluble and soluble fractions.

B) TPI-GAPDH cloned in pET-14b in different cell lines:

- 4) 400 mM IPTG at 30°C with different ODs: (0.6, 0.8, 1) respectively from 1-3 in B21, while 4 and 5 shows TPI-GAPDH pET-14b Rosetta2, 400 mM IPTG at 30°C with different ODs: (0.6, 0.8) respectively, and finally 6 and 7 TPI-GAPDH pET-14b Arctic express at 10°C degrees for 36 h. at ODs: (0.6, 0.8) respectively
- 5) (1-3) TPI-GAPDH pET-14b *E. coli* BL21, (4-5): in *E. coli* Rosetta2 cells, (6-7) arctic express cells. Each two lanes represent Un-induced, soluble pattern respectively for the same condition.

Molecular weights are indicated in kDa on the left.

A comparison of the expression levels of *E. coli* BL21 cells harbouring *Blastocystis* TPI-GAPDH pET-14b construct and using different expression methods including auto-induction in comparison with an empty vector is shown in Figure 6.4. It shows the difference in expression, with the highest level achieved using the auto-induction method. It also clearly shows the lack of expression in the empty vector.

The gene induced via auto-induction resulted in a high level of protein compared to other methods used. The overnight culture was inoculated in 500 ml culture of ZYM-5052 (for composition see materials 2.1.1.2) and incubated at 37°C for a period of 4-5 hours or until the OD is less than 1, ideally 0.6-0.8. The culture is subsequently incubated at 20°C with shaking at 220 rpm for 24 - 48 hours.

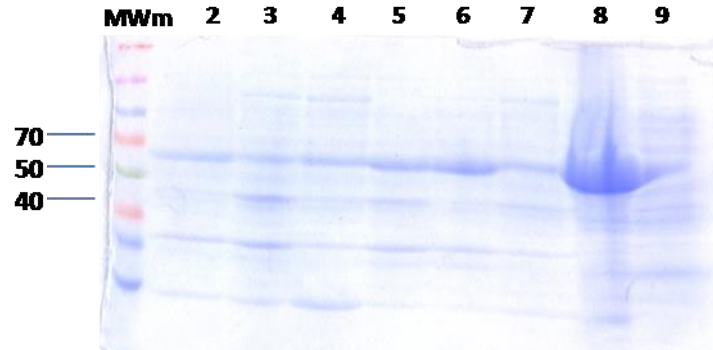


Figure 6. 4 SDS gel of several expression trials of *Blastocystis* TPI-GAPDH pET-14b construct in *E. coli* BL21. Lane1: molecular weight marker. Lanes 2-6: the same construct induced at different temperatures and with different IPTG concentration. Lanes 2-4: the same construct induced at 37°C with 400 mM IPTG, un-induced, induced total, soluble, consecutively. Lanes 5-7 induction with 400 mM overnight at 30°C, where 5-7: un-induced, induced total, soluble consecutively. Lane 8 auto-induction of *Blastocystis* TPI-GAPDH pET-14b construct in *E. coli* BL21. Lanes 9 auto-induced empty vector construct.

6.3. *Blastocystis* TPI-GAPDH expression and purification

As mentioned above, the *Blastocystis* TPI-GAPDH gene gave a high expression yield when grown under auto-induction conditions. This method was used subsequently to carry out the rest of purification experiments of this chapter. Because the protein has a His-tag at the N-terminus, a nickel affinity column was the first step of the purification process. The recombinant *Blastocystis* TPI-GAPDH protein was eluted from the nickel affinity column by elution using imidazole concentrations ranging from 10 - 500 mM, in 10 mM Tris-HCl.

The clarified cell extract was applied to a 1 ml pre-equilibrated nickel column with 3 X column volumes using a “super loop”. The protein was loaded by injecting the sample in the “super loop”. Any unbound protein was removed by applying 3 X column volumes of nickel affinity column (NAC) buffer A mentioned in Table 2.6. A gradient of NAC buffer B was applied ending with 100% of this buffer. This purification step was always followed by washing the column with 3 X 100% of NAC buffer B (Figure 6.5). Fractions were collected on a 96 well plate and the fractions were analysed by measuring the absorbance at 280 nm and by using SDS-PAGE as described in methods 2.2.3.4.

Nickel affinity purification resulted in several additional bands in addition to a protein of the expected size (Fig. 6.6). To improve homogeneity, ion-exchange chromatography was attempted as well; both cationic and anionic affinity chromatography. Unfortunately, both resulted in a very low yield of protein. In addition, the purification steps were reversed to assess the effect on protein purification. So, ion exchange chromatography was performed first, followed by nickel affinity chromatography. This also resulted in poor yields unfortunately. Purification attempts using ion exchange chromatography (both cation as well as

anion chromatography) are shown in Figure 6.7. Unfortunately, protein yields were so poor that any significant protein was barely visible on SDS gels. No corresponding gels for ion exchange chromatography are shown. As a consequence, it was decided not to include ion exchange chromatography in further purification attempts.

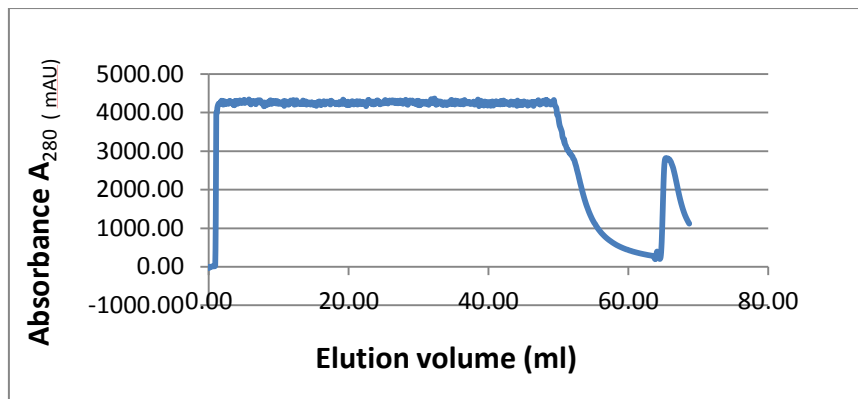


Figure 6. 5 Elution profile of *Blastocystis* TPI-GAPDH on a His-trap column (GE Healthcare). The A_{280} trace is shown in blue. The y-axis represents the absorbance of the protein at 280 nm. The x-axis shows the elution volumes of the protein. Recombinant protein was loaded from 0-55 ml. Elution of the recombinant *Blastocystis* TPI-GAPDH occurred from 55-75 ml.

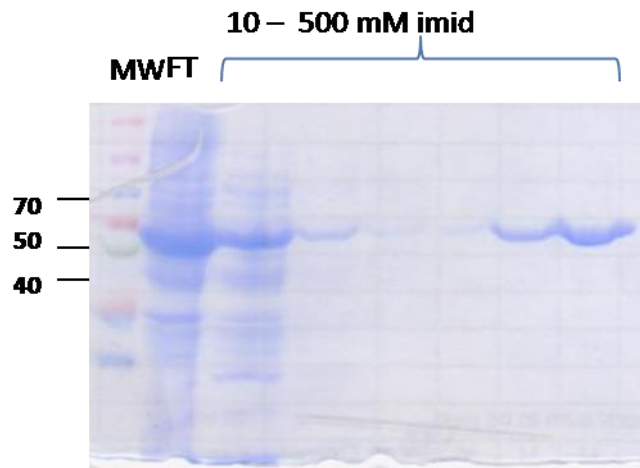


Figure 6. 6 A Coomassie-stained SDS-PAGE gel of recombinant *Blastocystis* TPI-APDH purified by nickel affinity chromatography. The *Blastocystis* TPI-GAPDH pET-14b construct in *E. coli* BL21 was auto-induced at 20°C for 48 h and purified using a nickel column with 500 mM final concentration of imidazole. Lane 1 represents the molecular weight marker (MW) (Fermentas Spectra Multicolor Broad Range), FT: flow through of the total protein lysate incubated with nickel beads and washed with imidazole buffer. The last six lanes represent gradient concentration of imidazole from 10 mM-500 mM. Molecular weights are indicated in kDa on the left.

The additional purification steps might have resulted in protein degradation. For this reason it was finally decided to restrict purification steps to nickel affinity chromatography followed by gel filtration chromatography.

Initially, protein purification resulted in either aggregated or inactive protein. Phosphate buffer was therefore used in both nickel and gel filtration chromatography to aid protein stability. The use of phosphate buffer was inspired by the fact that this buffer is used in the activity assay of this protein. It was therefore reasonable to assume that the protein would be more stable in this buffer. The buffer used for nickel affinity chromatography was: 50 mM phosphate pH 7.5, 500 mM NaCl, 30 mM imidazole, 0.2% Tween 80.

Different fractions containing the eluted protein from nickel affinity chromatography were pooled, concentrated and loaded onto a gel filtration column. This step is important to further purify the protein and to remove the imidazole buffer and exchange for a more suitable buffer for downstream analyses. Similar to nickel affinity chromatography, several buffers were tested for gel filtration chromatography. The most commonly used buffers (TRIS-HCl and HEPES) both resulted in a protein which showed a tendency to elute in the void volume (around 38.9 ml for the column used). This suggests the formation of aggregates of the protein. Protein aggregation was confirmed by performing activity assays which showed no activity indeed. In addition, thermal shift assays were performed on the eluted protein which indicated a miss-folded, aggregated or inactive form of the protein (shown later in paragraph 6.4.3).

Other buffers which contained several additives we also tested for better protein stability. These additives were added either to prevent aggregation by keeping

the protein in solution, for example by using detergents, or to enhance the stability of the protein, by adding reducing or chelating agents. Detergents such as Glycine and Tween 80 were added to stabilize the protein and prevent aggregation by breaking protein-protein and protein-lipid interactions. The detergents anti-aggregation effect is achieved by their surfactant property by disturbing the cell surface and hence allowing the proteins to be released in a soluble material ([Johnson 2014](#)). Here, a purification buffer mix of: 50 mM Tris /HCl (pH 7.8), 200 mM NaCl, 2 mM DTE, 1 mM EDTA, and 0.2 mM NAD was used ([Liaud, Lichtle et al. 2000](#)). This buffer mix helped stabilizing the protein, by using both a reducing agent (DTE) which protects the protein from damaging oxidative reactions, and a metal chelating agent (EDTA), which also helps in reducing oxidative processes, and chelates metal ions. However, unfortunately, no protein crystals formed in downstream crystallization experiments.

Ultimately, changing the buffer used in both the nickel affinity and the gel filtration purification step from Tris-HCl to phosphate buffer resulted in improved purification and also helped in crystallizing the protein in further experiments.

Therefore, the final buffer used in the gel filtration column was: 20 mM phosphate buffer with 200 mM NaCl, 1 mM of Tris (2-carboxyethyl) phosphine hydrochloride (TCEP) as a reducing agent, at a pH of 8.2. This resulted in a shift of the elution pattern of the protein from the void volume to about 60 ml elution volume, as shown in Figure 6.8 and 6.9. As mentioned above, these conditions also favoured crystal formation. Using this buffer has resulted in a protein with a high purity of almost 95% (Fig. 6.10). In order to check whether the TPI-GAPDH fusion protein obtained using gel filtration chromatography is of a single oligomeric status, or is in equilibrium with other forms, an additional experiment was conducted.

Fractions containing the highest peak were collected and analyzed using another gel filtration chromatography step. This resulted in a single peak profile, suggesting that each peak shown in Figure 6.9 A consists of one single form of the fusion protein and that these are not in equilibrium with other forms.

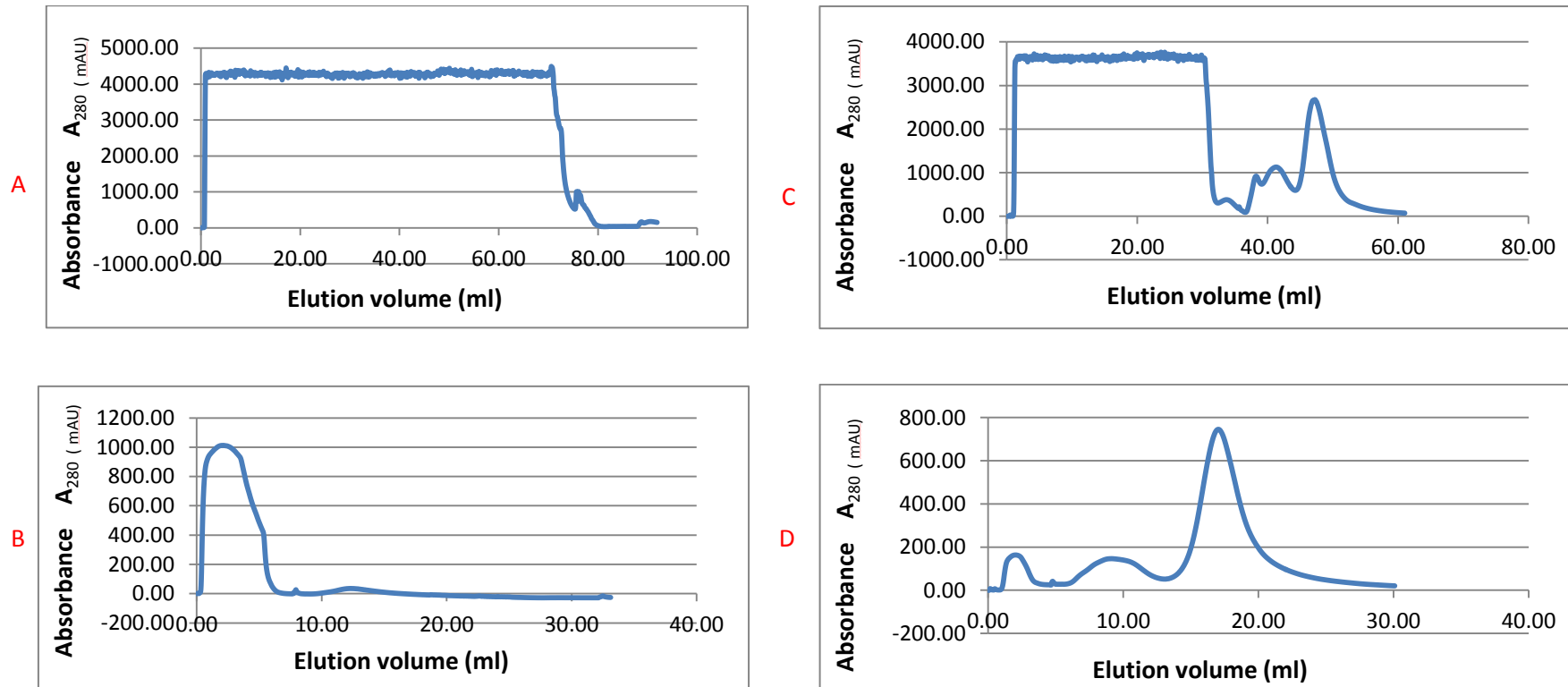


Figure 6. 7 Ion exchange chromatography of recombinant *Blastocystis* TPI-GAPDH. Elution profiles representing purification trials of *Blastocystis* TPI-GAPDH using ion exchange (IEX) chromatography are shown. A: Elution profile of TPI-GAPDH starting from crude material and applied on a cation exchange column SP XL (GE Healthcare). Less than 10 ml fractions were collected across the entire elution. B: Elution profile of *Blastocystis* TPI-GAPDH from the SP XL (cation exchange column) (GE Healthcare) after nickel affinity chromatography. About 5 ml fractions were collected across the entire elution. C: Elution profile of *Blastocystis* TPI-GAPDH from an anion exchange column from crude material. (The protein is presented in the low peak between 30-40ml). D: elution profile of the fusion protein using a QFF column after nickel affinity chromatography, showing very low yield of the protein with a peak around 10 ml. All protein resulting from these experiments did not result in any detectable band on SDS gels run afterwards. The A_{280} trace is shown in blue. The y-axis represents the absorbance of the protein at 280 nm. And x-axis shows the elution volumes of the protein in ml.

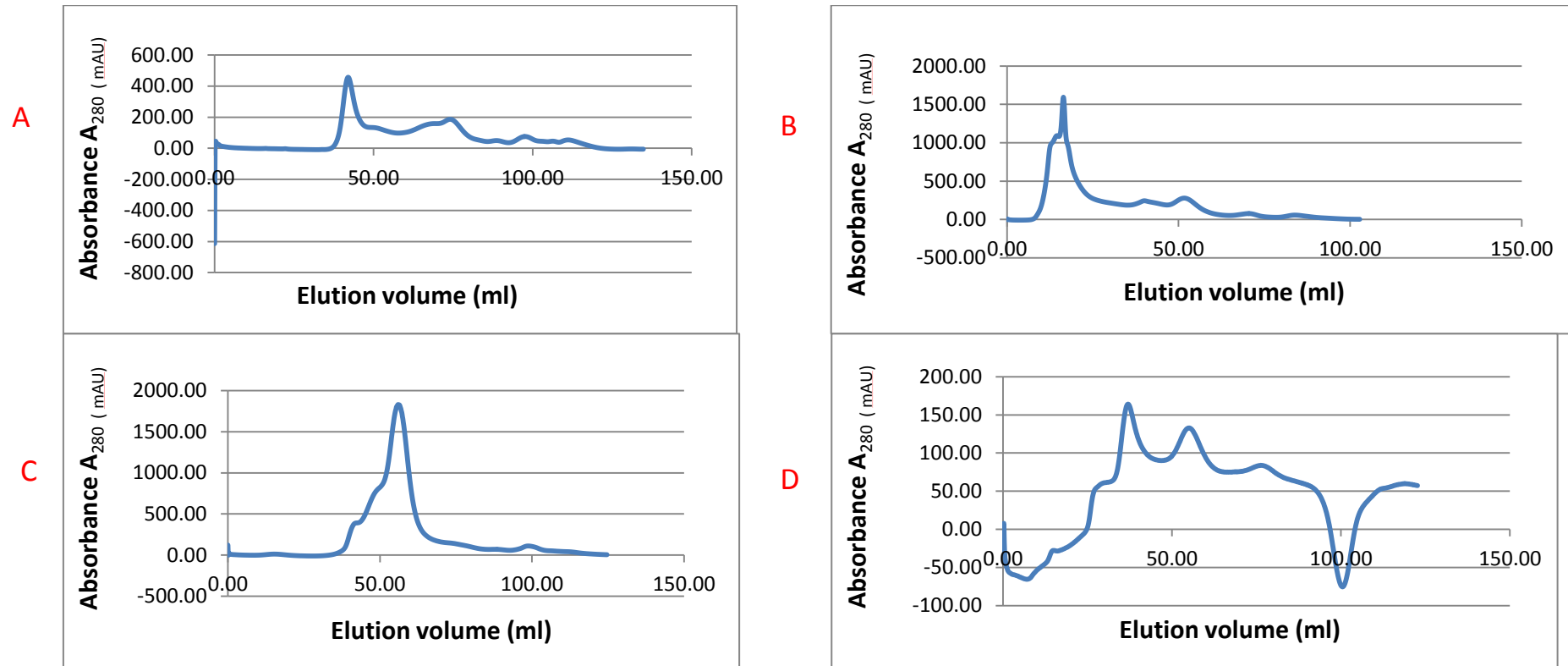
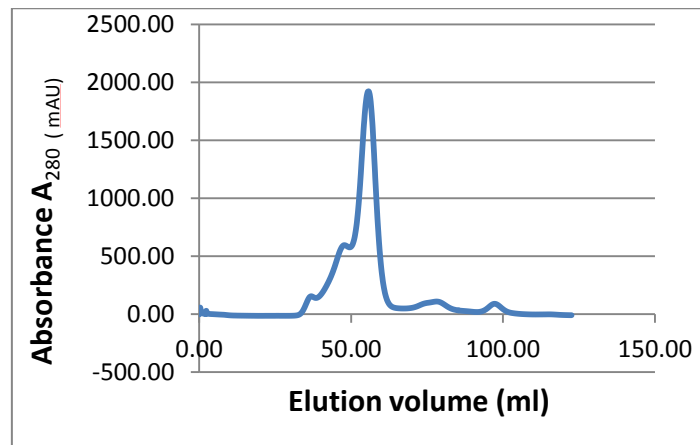


Figure 6.8 Gel filtration chromatography of recombinant *Blastocystis* TPI-GAPDH. Elution profiles of *Blastocystis* TPI-GAPDH are shown from a Superdex 200 gel filtration (GF) column using different buffers and detergents. A: *Blastocystis* TPI-GAPDH is eluted in the void volume using a Tris-HCl buffer. B: *Blastocystis* TPI-GAPDH also eluted in the void volume using a HEPES buffer. C: The *Blastocystis* TPI-GAPDH fusion protein elution profile shifted using a 50 mM Tris buffer with detergent (200 mM NaCl, 2 mM DTE, 1 mM EDTA, and 0.2 mM NAD⁺). D: Elution profile of TPI-GAPDH with a Tris buffer containing the co-factor NAD⁺. The A₂₈₀ trace is shown in blue. The y-axis represents the absorbance of the protein at 280 nm. The x-axis shows the elution volumes of the protein in ml.

A



B

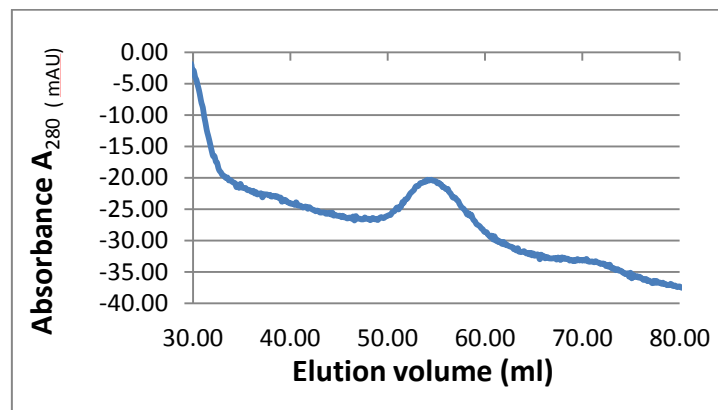


Figure 6. 9 A) Final purification of *Blastocystis* TPI-GAPDH using gel filtration size exclusion chromatography. *Blastocystis* TPI-GAPDH was eluted in phosphate buffer from a Superdex 200 gel filtration (GF) chromatography column. The A₂₈₀ trace is shown in blue. The fusion protein is eluted in the fractions between 40 - 60 ml. Y axis represents the absorbance of the protein at 280 nm. X axis shows the elution volumes of the protein in ml. Peak at about 55 ml represents the eluted target protein. B) Elution profile of the purified *Blastocystis* TPI-GAPDH using the fractions collected from the highest peak to check if the gradient peak obtained in first purification is in equilibrium between different oligomeric state proteins. Showing a single peak, indicating that the peak in A is not an equilibrium between several oligomeric status proteins.

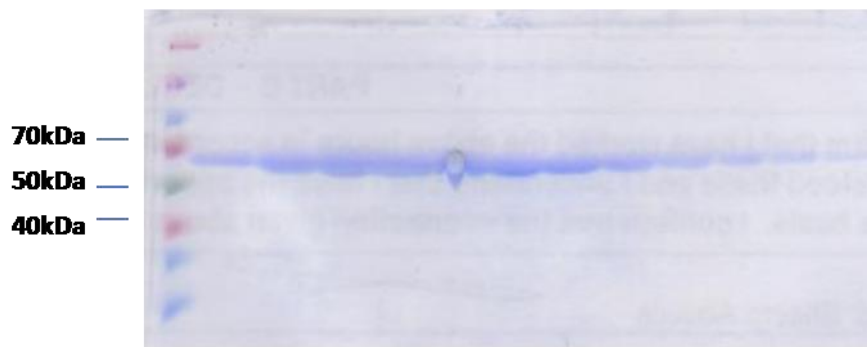


Figure 6. 10 Purified recombinant *Blastocystis* TPI-GAPDH. A Coomassie-stained SDS-PAGE gel of the *Blastocystis* TPI-GAPDH fusion protein obtained after gel filtration chromatography using a GF Superdex 200 gel filtration chromatography column. *Blastocystis* TPI-GAPDH protein bands correspond to all the fractions between 40 - 60 ml. The first lane represents the molecular weight marker. (Fermentas Spectra Multicolor Broad Range). Molecular weights are indicated in kDa on the left.

6.4. Biochemical characterization

6.4.1. Molecular weight determination

6.4.1.1. Analytical GF

Gel filtration chromatography is a method that can be used for purifying proteins but which can also be used as a tool for estimating the molecular weight of proteins when used in conjunction with reference proteins of known molecular weight. In order to determine the molecular weight of the recombinant *Blastocystis* TPI-GAPDH, the protein was loaded onto a HiLoad 16/60 Superdex 200 gel filtration column and run as described in methods 2.2.3.3.3. To obtain the molecular weight of the native TPI-GAPDH, the protein was run in triplicate. A typical TPI-GAPDH elution profile is shown in Figure 6.11 and Figure 6.12. TPI-GAPDH always eluted in the same pattern having three main adjacent peaks. For Figure 6.11, the first peak eluted at 37 ml, a second larger peak eluted at 47.80 ml, the third main peak elutes around 55.67 ml, the fourth small peak eluted at 78.12 ml and the last small peak eluted at 97.65 ml. The molecular weights were then calculated based on the calibration curve in Appendix 2, and using 38.92 ml

as a void volume of the used column. The following results are the estimated molecular weights of the peaks:

Peak 1: VE = 37.00 ml	log Mw = 3.2	Mw = 1447.8 kDa
Peak 2: VE = 47.80 ml	log Mw = 2.7	Mw = 508.2 kDa
Peak 3: VE = 55.67 ml	log Mw = 2.4	Mw = 236 kDa
Peak 4: VE = 78.12 ml	log Mw = 1.4	Mw = 26.9 kDa
Peak 5: VE = 97.65 ml	log Mw = 0.6	Mw = 4.7 kDa

Based on the predicted molecular weight of a TPI-GAPDH monomer of 65 kDa, such an elution profile suggests the following. The very high molecular weight initial peak most likely represent aggregated proteins. The second peak of 508 kDa is about eight times the molecular weight of the monomer which would suggest an octameric protein. Analysing this protein on SDS-PAGE shows a band corresponding to the expected molecular weight of the fusion protein indicating this high-molecular weight peak does indeed consist of TPI-GAPDH. The third peak would represent a tetramer of the fusion protein (~ 240 kDa) and this has also been confirmed by SDS-PAGE as above. The fourth peak might be a degraded fusion protein as it is corresponding to the molecular weight of triosephosphate isomerase alone (TPI Mw = 27 kDa). Finally, the last peak most likely represents small molecules that always tend to elute at the end of gel filtration chromatography runs. These are probably salts that are washed out at the end.

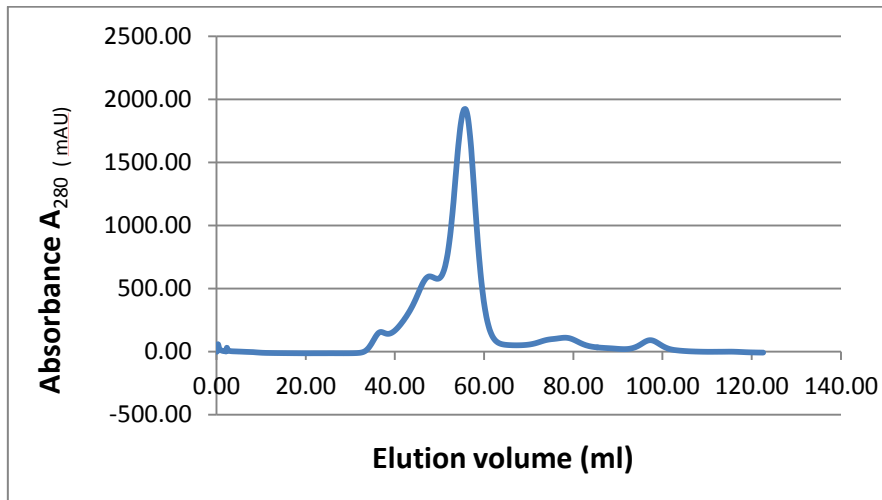


Figure 6.11 Gel filtration chromatography of *Blastocystis* recombinant TPI-GAPDH. Elution profile for *Blastocystis* TPI-GAPDH run on HiLoad 16/60 Superdex 200 gel filtration chromatography column in triplicates, the single peak obtained represents Blastocystis TPI-GAPDH. The Y axis represents the absorbance of the protein at 280 nm. The X axis shows the elution volumes of the protein in ml.

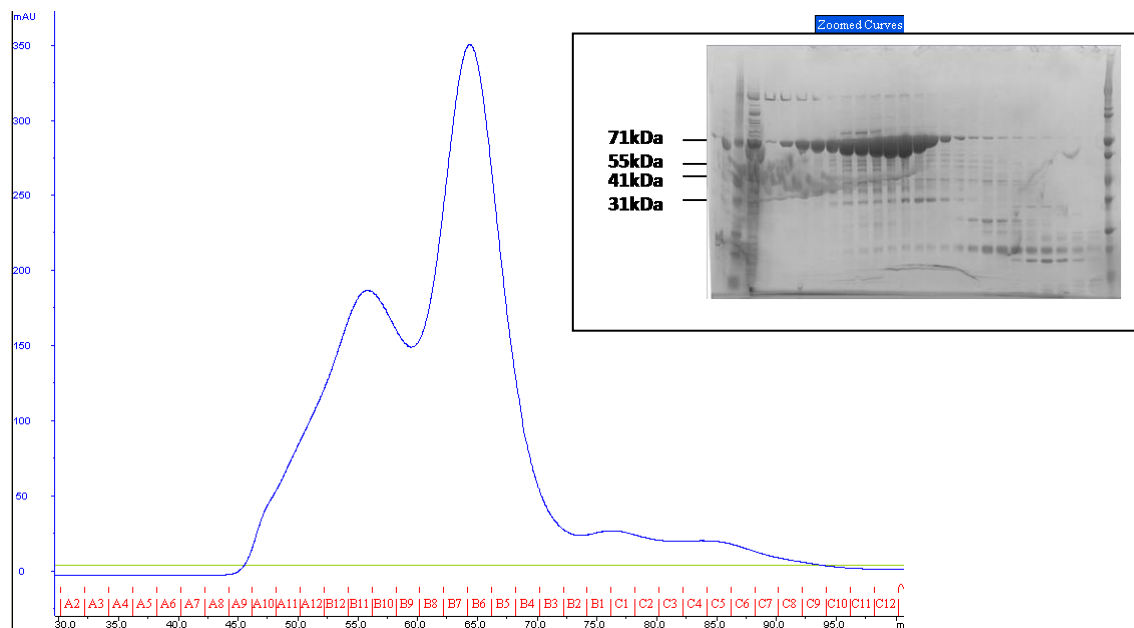


Figure 6.12 Gel filtration chromatography of recombinant *Blastocystis* TPI-GAPDH. Elution profile of *Blastocystis* TPI-GAPDH on HiLoad 16/60 Superdex 200 gel filtration column and SDS-PAGE of the corresponding fractions. This work was repeated while at the Oxford protein production facility (OPPF) and confirmed work done previously.

6.4.1.2. Immune-detection (Western Blotting) proved the identity of the purified protein to be TPI-GAPDH

In order to confirm the identity of the protein purified in paragraph 6.3, an immune detection approach was used (see methods 2.2.3.7). The recombinant

Blastocystis TPI-GAPDH which was purified using gel filtration chromatography was used in a Western blotting experiment using two different antibodies: an anti-His antibody (Qiagen) targeting histidine residues on the N-terminus of the recombinant protein, and using antibodies raised specifically against the protein in *Blastocystis* (unpublished work Dr. Matthew Rogers, University of Exeter). Both experiments identified a protein with a molecular weight of 65 kDa which is in agreement of the expected molecular weight of the recombinant TPI-GAPDH protein (see Figure 6.13). TPI-GAPDH antibodies were used in the following concentrations: 1:2,000 for the primary and 1:10,000 for the secondary antibodies. While it was 1:1,000 and 1:5,000 for primary and secondary antibodies when using Anti His Abs.

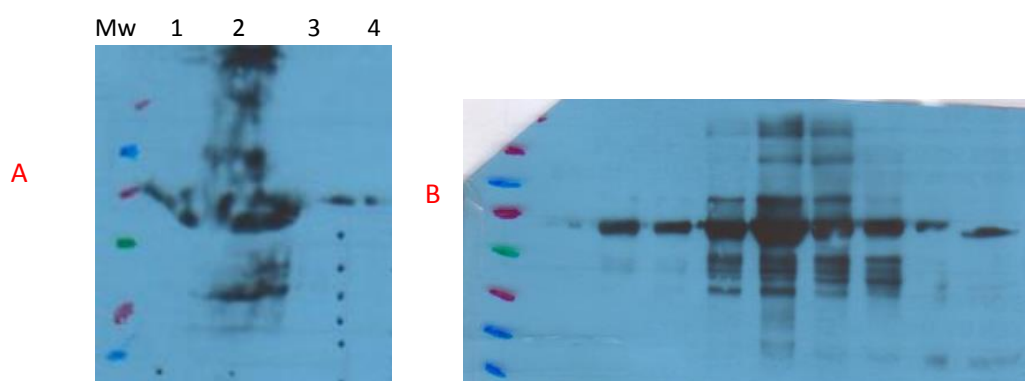


Figure 6.13 Western blot analysis detecting *Blastocystis* TPI-GAPDH protein. A. Western blot analysis using a homologous antibody raised in guinea pig against *Blastocystis* TPI-GAPDH protein. Only one band detected corresponding to the fusion protein probed with Anti-guinea pig TPI-GAPDH Abs corresponding to the predicted molecular weight of about 65 kDa. MW lane is the molecular weight marker (Fermentas Spectra Multicolor Broad Range). Lane 1: empty pET14b plasmid used as negative control. Lane 2: represents the total cell lysate. Lanes 3 and 4 show gel filtration-chromatography purified TPI-GAPDH. B: Western blot analysis using anti-His antibodies targeting the His-tag introduced in *Blastocystis* TPI-GAPDH. This represents a sample of each of the three peaks obtained after purification, confirming the homo content of the 3 peaks which is the fusion protein TPI-GAPDH is. In B other bands appear which indicates a degradation and/or contamination with other proteins after purification. Antibody concentrations used: anti His: primary antibody: 1:1,000 and secondary: 1:5,000. For anti-TPI-GAPDH: 1:2,000 for the primary and 1:10,000 for the secondary antibody.

6.4.1.3 Oligomeric status determination of the native TPI-GAPDH

The recombinant *Blastocystis* TPI-GAPDH fusion protein, purified by gel filtration chromatography, was analyzed using native gel electrophoresis to determine its oligomeric status. As this type of gel electrophoresis does not contain SDS, it

allows for the determination of the molecular weight of proteins under native conditions. For details of this technique please see methods 2.2.3.5. In order to be able to determine the native molecular weight, the protein under study is analysed alongside a set of standard proteins mentioned previously in Table 2.13 (except for Dextran). The *Blastocystis* TPI-GAPDH fusion protein was always located high up in the gel as can be seen in Figure 6.14. This suggests it is either aggregated or exists as a hexadecamer. The band corresponding to the purified TPI-GAPDH is close in size to 1048 kDa which is 16 x 65 kDa.

To further analyse and confirm the native TPI-GAPDH fusion protein, the strip containing the TPI-GAPDH band was cut out of the native gel and analysed under reducing conditions on SDS-PAGE. This experiment confirmed the presence of the *Blastocystis* TPI-GAPDH fusion protein as a band of 65 kDa was observed on the Coomassie stained SDS-PAGE gel, see Figure 6.15.

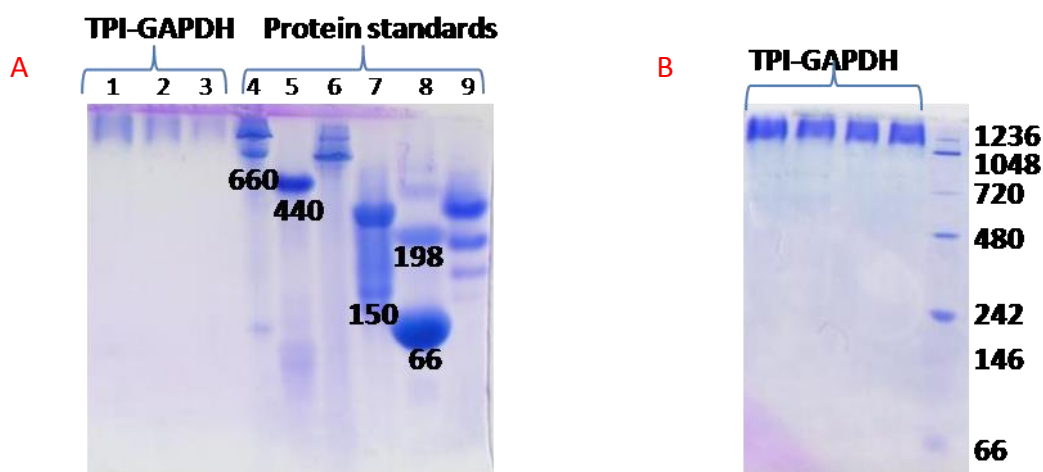


Figure 6. 14 Native gel electrophoresis of *Blastocystis* TPI-GAPDH following gel filtration chromatography. A: Blue native gel electrophoresis of TPI-GAPDH. The fusion protein is present in lanes 1-3. Protein standards located as follows: 4) Thyroglobulin (660 kDa), 5) apoferritin (440 kDa), 6) beta amylase (200 kDa), 7) alcohol dehydrogenase (150 kDa), 8) albumin (66 kDa) 9) and carbonic anhydrase (30 kDa). B: Native gel electrophoresis of TPI-GAPDH alongside the native protein standard NativeMark™ (Life Technologies, see Figure 2.5). The protein marker is located in the last lane. Molecular weights are indicated in kDa.

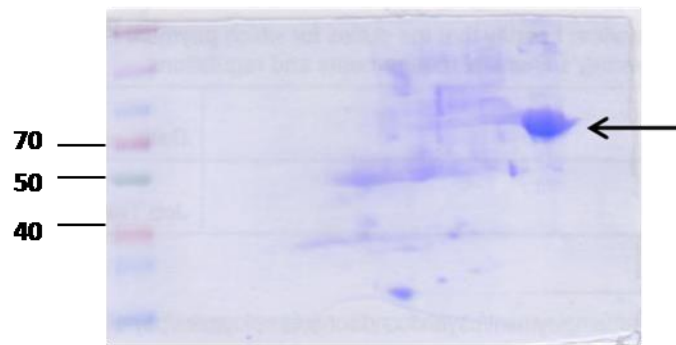


Figure 6. 15 Second dimension SDS-PAGE following native gel electrophoresis. A strip was cut from a blue native gel containing TPI-GAPDH. The black arrow indicates the major band present representing the expected fusion protein with a molecular weight of about 65 kDa. Molecular weights are indicated in kDa on the left.

Unfortunately, as can be seen from Figures 6.14 and 6.15, the oligomeric status of the purified fusion protein could not be properly determined. However, this experiment did confirm the identity of the high molecular weight protein to be the *Blastocystis* TPI-GAPDH fusion protein and that therefore this protein does seem to form a higher order oligomeric state.

6.4.1.4. Size distribution profile using Dynamic Light Scattering

In order to determine the oligomeric state of the *Blastocystis* TPI-GAPDH protein, the gel filtration chromatography purified protein was centrifuged for 20 minutes at 4°C at 16,000 x g before being analysed on the dynamic light scattering machine as described in paragraph 2.2.3.9.2. Ten measurements were taken; each measurement had ten acquisition times of 20 seconds acquisition, recorded at room temperature using a DynamPro Titan (Wyatt Technology Corporation). Data was analysed using DYNAMICS (Protein Solution, USA).

Unfortunately, the results of these experiments only suggested that there is a protein aggregation as a likely cause of the un-interpretable results. It is shown in Figure 6.16 that the protein always had some very high molecular weight components in the several thousand of kilo Daltons. This happened irrespective of the various purification trials conducted using different buffers, including the phosphate buffer (see section 6.2.2). Some of the results suggested degradation of the protein into its two building units as seen with the reported 38 kDa, representing a monomeric GAPDH, or the 78 kDa which is close to a dimeric GAPDH. The molecular weight of TPI alone is 28 kDa and it is normally found as a dimer (56 kDa). The molecular weight of GAPDH is 36.5 kDa and is naturally occurring as a tetramer (144 kDa). Table 6.1 represents some of those results.

Table 6. 1 Dynamic light scattering data of *Blastocystis* TPI-GAPDH. Data obtained from six measurements, each measurement is a mean of ten readings. R: diameter of the scattered molecule in nm, Pd: poly dispersity, MW-R is the molecular weight measured in kDa

Item	Intensity (%)	R (nm)	Pd (nm)	Pd (%)	MW-R (kDa)
Meas 1	53.3	0.2	0.0	16.4	78.836
Meas 2	9.7	45442.2	0.0	10.1	265891000000
Meas 3	95	0.1	0.0	28.3	0.012
Meas 4	4	3.8	0.0	0.0	78.8364
Meas 5	26.3	2.8	0.3	11.0	38.302
Meas 6	7.4	3.1	0.5	17.1	47.700

Nonetheless, DLS results when experimented on *Blastocystis* TPI-GAPDH showed a poly dispersity of the purified protein, and this indicates less homogeneity, Figure 6.16 shows examples of the results obtained from running the fusion protein on the DynamPro Titan machine.

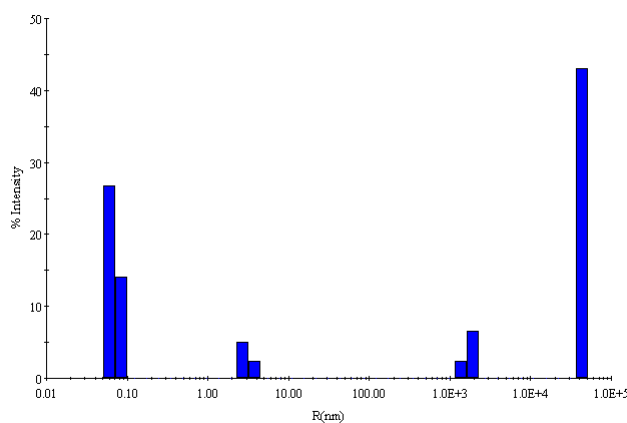


Figure 6. 16 Dynamic light scattering analysis of the *Blastocystis* TPI-GAPDH fusion protein. The fusion protein shows varied intensities of several molecular weights ranging from very low as 0.005 kDa of 51% intensity to 26% intensity of 38 kDa, with the presence of very large aggregated molecules of several thousand kDa with 18-21% intensity. X axis represents the measured diameter of the measured molecules. Y axis represents the percentage of the intensity of the molecule detected in the solution.

6.4.1.5. Analytical ultracentrifugation (AUC)

A solute in a solution undergoes two forces, sedimentation and buoyant force. When using centrifugation, a solute starts precipitating at the bottom of the cell due to the gravitational effect resulting from the centrifuge forces. Hence, this would result in an increase of solute concentration near the bottom of the tube. This in turn will be opposed by another force, the distribution force which helps the solute to float (buoyant force). A final force is the frictional force resulting from the movement of the molecule in a solvent. After a certain time these forces will balance, resulting in a net force, which will help making the concentration of the solute the same in each part of the solution. Measuring concentration distribution after reaching equilibrium represents the sedimentation equilibrium (SE) ([Cole, Lary et al. 2008](#)).

In the case of analytical ultra centrifugation, the sedimentation equilibrium represents the balance between those two already mentioned forces, the action of piling the molecules at the bottom of the cell due to gravity, and the diffusion of the molecule due to buoyant forces. Therefore, the concentration will be the same and it will only depend on the molecular weight of the solute. When we measure the concentration distribution after the balance has been reached, we are measuring equilibrium sedimentation ([Cole, Lary et al. 2008](#)). Sedimentation equilibrium offers a very accurate way of measuring the molecular weight and consequently the oligomeric status of molecule in biological solution.

This is the best method to determine the molecular weight of a substance, that no other method can surpass it, in accuracy and its wide coverage. Analytical ultra-centrifugation covers a wide range of molecular weights from few hundred Daltons to several hundred millions Daltons. Another factor that makes it superior to other methods such as dynamic light scattering (DLS) is the very small size

and concentration of the protein used in sedimentation equilibrium method (20–120 μ l) and (0.01-1 g/l) sequentially. As for the applicability sedimentation equilibrium can be used for any substance whose absorbance is different from the solvent it is found in. It could be used to detect both the purity and molecular weights of proteins, carbohydrates, nucleic acids, etc. Sedimentation equilibrium also gives an idea about the purity of the solution and the homogeneity of the molecules in it.

In this project, analytical ultra-centrifugation was performed to precisely determine the molecular weight and the oligomeric status of the unusual TPI-GAPDH fusion protein from *Blastocystis*. In addition to TPI-GAPDH, the purified *Blastocystis* GAPDH protein (details of this work is found in chapter 7) was also submitted to Oxford Protein Purification Facility. The results obtained from *Blastocystis* GAPDH will help elucidate whether this protein is found as a tetramer like most known GAPDH proteins studied ([Jia, Linh le et al. 2011](#), [Baker, Shi et al. 2014](#)), or whether it is taking a different conformation as suggested in the MSc work by Evans *et al* ([Evans 2014](#)) that suggested that the *Blastocystis* GAPDH is a dimers. The sedimentation equilibrium experiment was performed using three different protein concentrations of each protein as explained in section 2.2.3.9.2. These were 1, 0.2 and 0.05 mg/ml of TPI-GAPDH in a total volume of 400 μ l each. Each sample was loaded in a separate analytical cell cleaned properly with the gel filtration buffer used for purification and equilibrated against water, and centrifuged at the same run on Beckman Coulter ProteomeLab XL-I Protein Characterization System.

The protein was run using an AN-60 Ti rotor (Beckman) at 40,000 X rpm generating about 128,793 X g at 4°C.

The data of the three concentrations were fitted to a range of models. The results indicate a homogeneous sample. These results also showed that *Blastocystis* TPI-GAPDH sediments at 9.89 S and this was consistent for all 3 concentrations used, demonstrating that it is larger and heavier than GAPDH on its own, which sedimented at 7.5 S (see Figure 6.17). The data also indicated the presence of some higher order aggregates in the sample with a very low degree. Most of the samples therefore appear to be homogeneous, *i.e.* they are a mono disperse species, see Table 6.2.

Table 6. 2 Analytical ultra-centrifugation analyses of *Blastocystis* TPI-GAPDH. Data was obtained using three different protein concentrations as explained in the text.

Concentration (mg /ml)	Percentage of signal %	Molecular weight (kDa)
0.05	65	186
0.2	72	235
1	79	260

The signal obtained from analytical ultracentrifugation gave results varying in the expected molecular weight. The values ranged between 235-260 kDa forming 72-79% of the signal obtained for each concentration measured (Table 6.2). A frictional ratios less than one is considered unrealistic in AUC results. The frictional ration obtained in this experiment was 1.49 and this corresponds to a tetramer. The molecular weight was calculated by fitting for the frictional ratio which resulted in ~160 kDa for GAPDH and ~240 kDa for the TPI-GAPDH fusion protein.

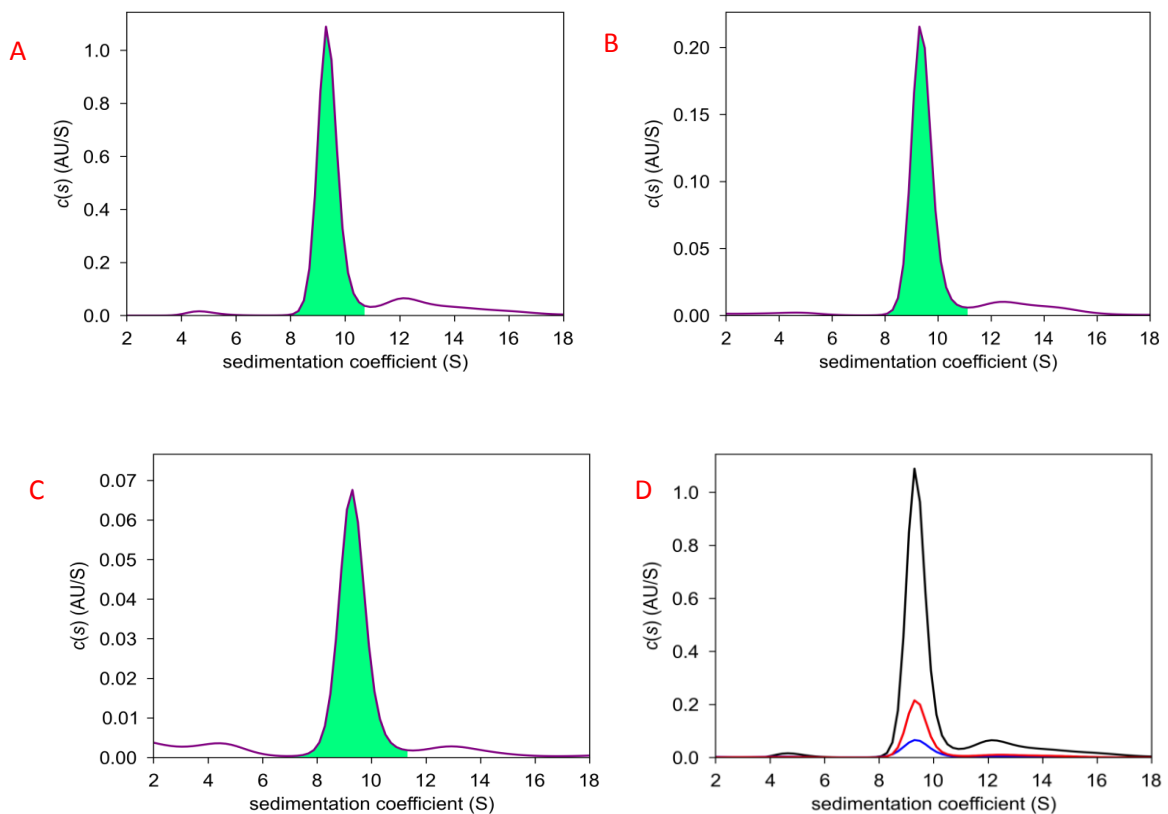


Figure 6.17 Sedimentation equilibrium analytical ultracentrifugation of *Blastocystis* TPI-GAPDH. Distribution of sedimentation coefficient using three different concentrations of TPI-GAPDH. Size distribution profiles are shown in three different concentrations for the same protein, in: A: 1 mg/ml B: 0.2 mg/ml C: 0.05 mg/ml. D the 3 different concentrations aligned together and showing concentric peaks indicating the dominance of one form of the protein. X axis represents the sedimentation coefficient. Y axis represents concentration of the protein (thanks to Dr David Scott from the Oxford lab for experimental help, analyses and interpretation).

6.4.1.5.1. Complementary work on triosephosphate isomerase (TPI) and Glyceraldehyde-3-phosphate dehydrogenase (GAPDH)

Because the fusion protein is composed of two units that are normally present as the separate proteins triosephosphate isomerase (TPI) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH), it is important to study the components of this fusion protein. The main purpose of this part of the study was to determine the molecular weight and oligomeric state of the individual proteins comprising the building units of the fusion protein. To achieve this, both proteins were individually expressed from a construct of each gene in pET-14b vector in *E. coli*

BL21 cells. The constructs were all stored in the Dr van der Giezen group stocks and previously made by Dr Matthew Rogers.

The *Blastocystis* artificially separated TPI and GAPDH genes were each cloned into *E. coli* BL21 cells in the pET-14b vector. The genes were expressed at 20°C using the auto-induction approach. They were thereafter over-expressed and the respective protein was subsequently purified using nickel affinity chromatography followed by gel filtration chromatography. Figure 6.18 A shows the purified recombinant *Blastocystis* GAPDH protein elution profile which was further analysed on SDS PAGE. Gel filtration chromatography resulted in a single large peak with very small shoulders, see Figure 6.18 A. When subsequently analysed on SDS PAGE, the recombinant GAPDH was clearly visible as a clear band of the expected size of 36 kDa (Figure 6.18 A). In addition, the purified recombinant GAPDH protein was analysed using AUC to determine its molecular weight and oligomeric state of about 160 kDa. AUC confirmed the presence of GAPDH in a tetrameric form (Figure 6.19).

Similarly, the recombinant *Blastocystis* TPI was also expressed using auto-induction. The crude cell lysate was subsequently processed using nickel affinity chromatography and gel filtration chromatography, successively. The GF elution profile of TPI, in addition to the SDS PAGE analyses indicate that the recombinant TPI is degraded (Figure 6.18 B).

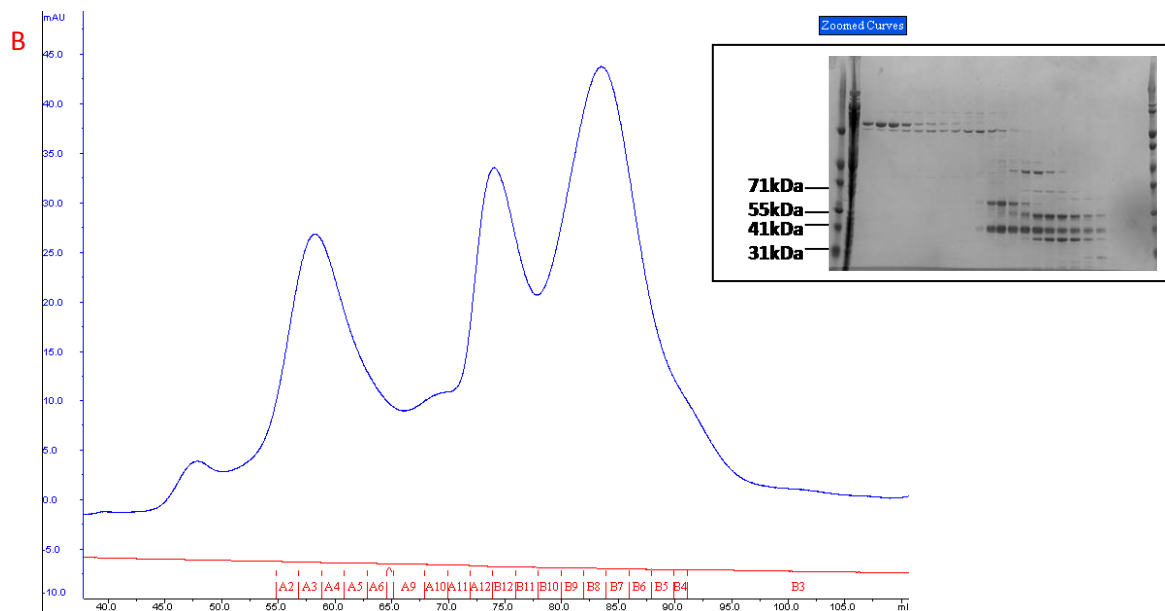
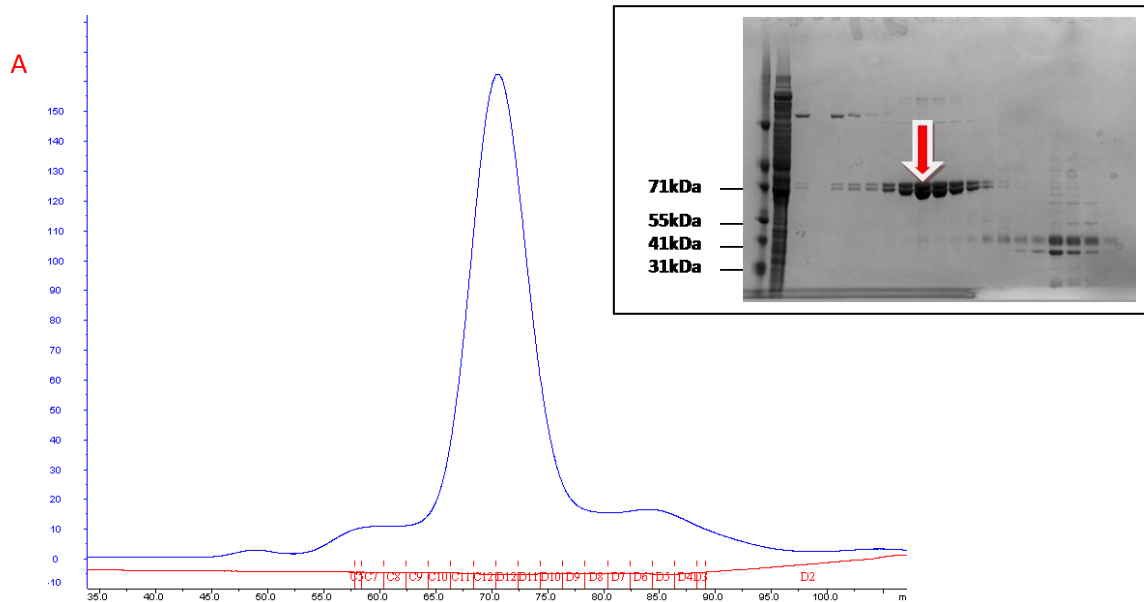


Figure 6. 18 Purification of *Blastocystis* recombinant GAPDH and TPI. A. Elution profile of recombinant GAPDH during gel filtration chromatography. Inset shows SDS PAGE analysis of fractions taken from the elution profile indicated underneath the trace. Molecular weight marker indicated on the left. The marker used in OPF is HiMark™ Pre-Stained Protein Standard. The red arrow indicates *Blastocystis* GAPDH of 36 kDa. B. Elution profile of recombinant TPI during gel filtration chromatography. Inset shows SDS PAGE analysis of fractions indicated underneath elution profile.

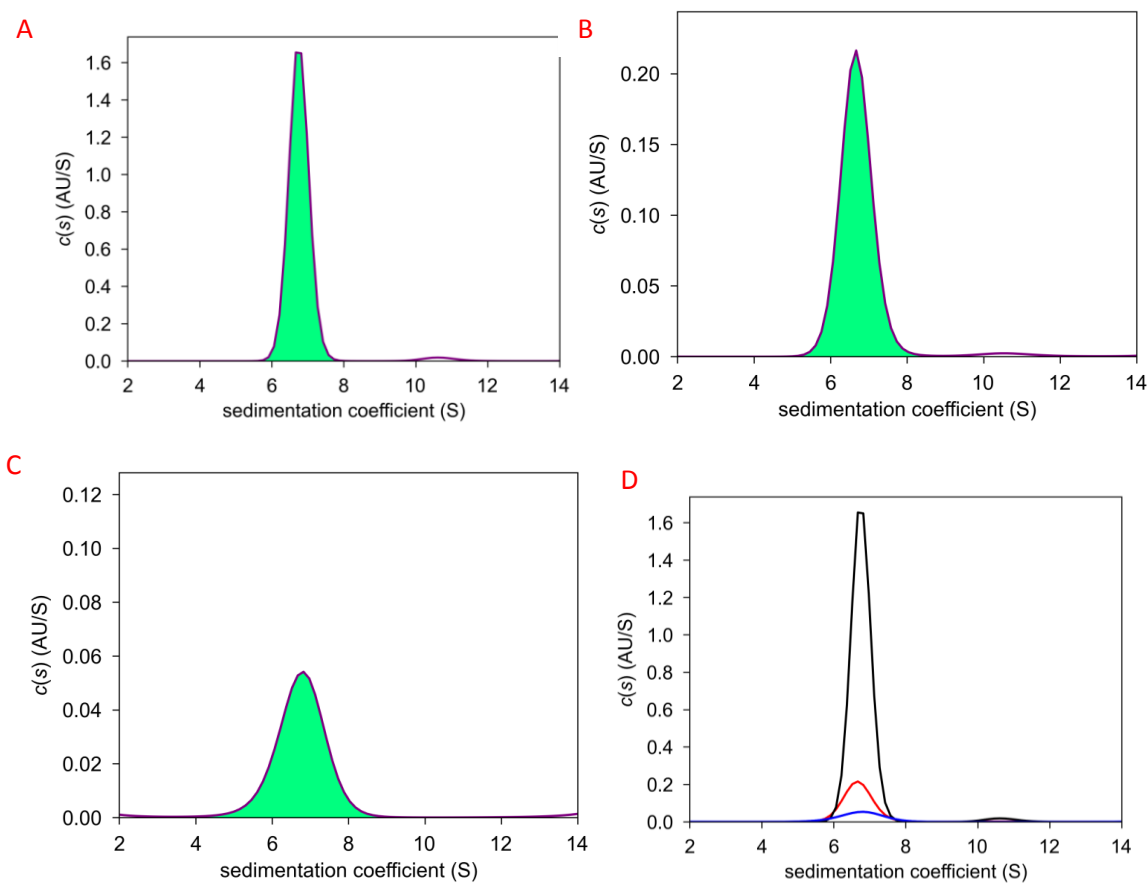


Figure 6. 19 Sedimentation equilibrium analytical ultracentrifugation of recombinant *Blastocystis* GAPDH. Distribution of sedimentation coefficient using three different concentrations of GAPDH. Size distribution profiles are shown in three different concentrations for the same protein A. 1 mg/ml B: 0.2 mg/ml C: 0.05 mg/ml. D the three different concentrations aligned together and showing a concentric peaks indicating the dominance of one form of the protein. X axis shows the sedimentation coefficient values. Y axis shows the protein concentration (thanks to Dr David Scott from the Oxford lab for experimental help, analyses and interpretation).

6.4.2. Fluorescent-based thermal shift assay (DSF: differential scanning fluorimetry)

Differential scanning fluorimetry, or DSF, is a powerful technique to analyze protein stability in the presence or absence of ligands or other small molecules. It has played important roles in drug development. In addition, it can be used to provide information about the stability of a protein in different buffers. This can provide crucial information when choosing which buffer to use when purifying a protein. DSF is further explained in Materials and Methods 2.2.3.10.3.

Here, the recombinant *Blastocystis* TPI-GAPDH fusion protein was analysed using differential scanning fluorimetry initially to check for buffer suitability. The protein was first purified using gel filtration chromatography as explained before and subsequently used for DSF analyses. Several different buffers were tested for this analyses and a full list of these buffers can be found in materials and methods Table 2.7.

When the TPI-GADPH fusion protein was purified in a Tris-HCl buffer, a very high DSF signal was obtained suggesting that the protein is either inactive or aggregated (Figure 6.20 A). Changing the purification buffer to a phosphate buffer resulted in a much a lower DSF signal (Figure 6.20 B) which supports the assumption that the protein indeed behaves better when present in phosphate buffer. It was also noticed that using a phosphate buffer resulted in a moderate shift to a higher melting temperature (55.60°C) in comparison to the original Tris-HCl buffer (49.92°C). This change in the melting temperature of 5.7°C suggests a better stability of the protein.

Changing the purification buffer from Tris-HCL to phosphate buffer was also accompanied by the addition of the ligand NAD⁺ to increase stability. This combination resulted in rather unusual results for the recombinant protein TPI-GAPDH thermal shift profile. Usually in DSF experiments a simple denaturation profile is observed with a single peak indicating the optimal temperature at which the protein is stable. However, adding the ligand NAD⁺ resulted in rather unusual biphasic pattern of melting (Figure 6.20 B). This could be explained by the tendency of the protein to melt at either ligand-free temperature or the fully bound protein melting temperature. A higher concentration of the ligand would result in shifting this biphasic profile to a single peak profile towards the higher melting temperature.

DFS can also be used to get insight into protein/ligand interaction. It will provide information on how tightly a ligand is bound to its protein through measuring the dissociation constant K_d . An initial screen of the ligand (NAD⁺) concentration suggested that the likely K_d is to be in the range of $93 \pm 19 \mu\text{M}$ (Figure 6.21). These observations suggest a high tendency of the ligand to dissociate from its protein. In other words, it indicates a lower affinity between TPI-GAPDH and its ligand NAD⁺. This might explain the results presented in the following crystallography chapter when solving the structure of the TPI-GAPDH fusion protein. It was hard to see the ligand bound to the active site of the fusion protein. The lower affinity of the ligand for its protein suggests that the ligand might be easily washed off the protein during the various purification and washing steps.

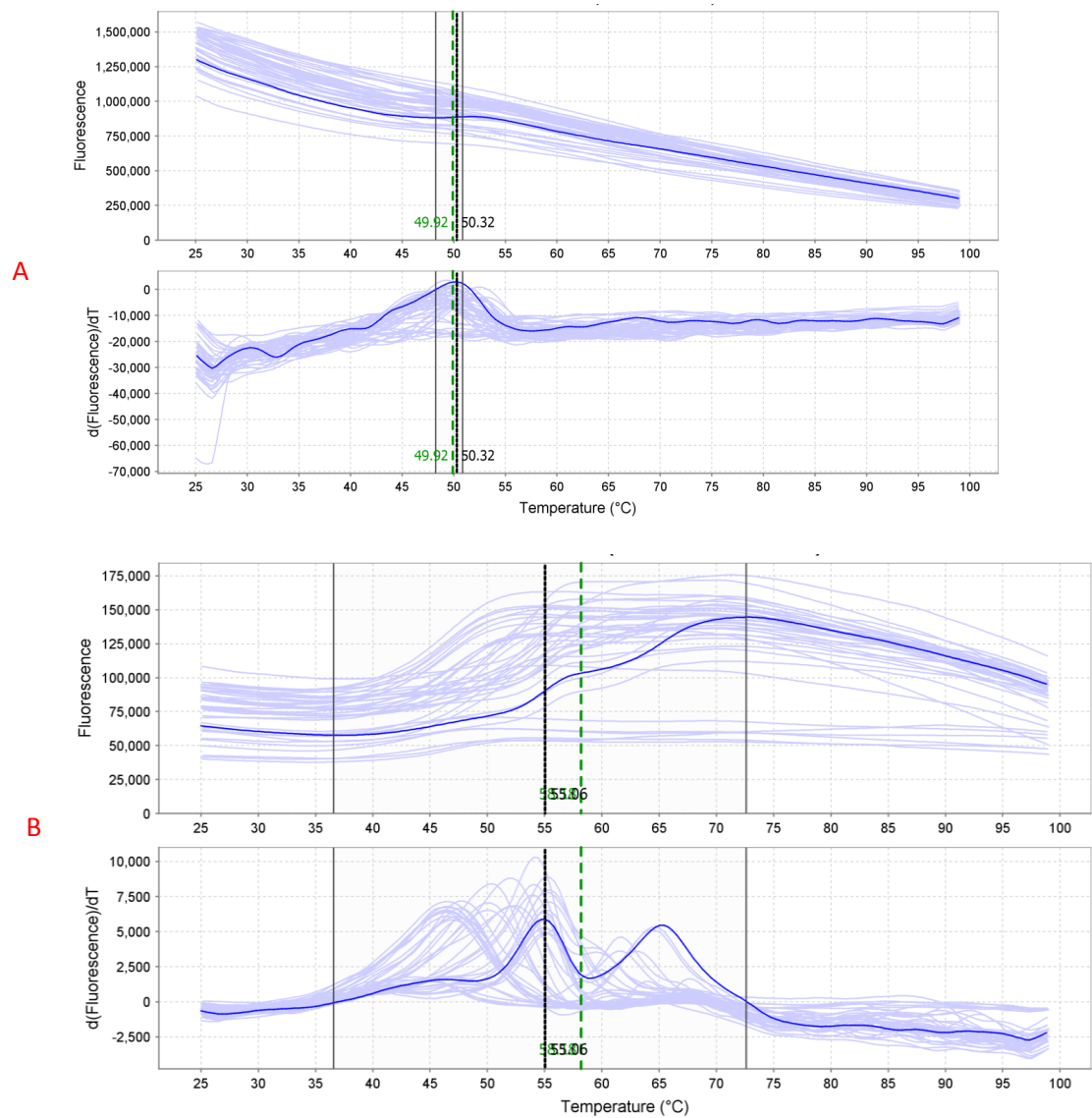


Figure 6.20 Melt curve plot of TPI-GAPDH showing the thermal denaturation profiles of the *Blastocystis* TPI-GAPDH fusion protein purified in different buffering conditions. A: Purification of the fusion protein in HEPES. B: Purification of TPI-GAPDH in phosphate buffer with the addition of the ligand NAD⁺ for better stability. It is clear from these data that the different buffer results in increased stability of the TPI-GAPDH fusion protein. The second profile shows a progressive increase in the fluorescence signal to reach a maximum, followed by a shallow decline. This is accompanied by a biphasic melting peaks, indicating a ligand free and ligand bound states of the protein in the first derivative of the fluorescence. Observed melting temperature is given on the X-axis, and the fluorescence signal is given on the Y-axis.

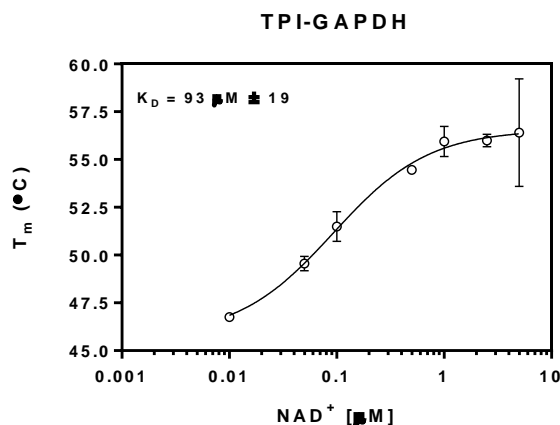


Figure 6. 21 Interaction between *Blastocystis* TPI-GAPDH and NAD⁺ using thermal shift assay. Initial testing using a wide range of NAD⁺ concentration (0.01-100 µM), suggesting that the dissociation constant K_d is more likely to be in the range of 84-112 µM. Images were prepared using GraphPad. The X axis represents the ligand NAD⁺ concentration. The Y axis is representing the melting temperature of the protein in the presence of the ligand. $R = 0.9425$ (95% confidence intervals)

6.4.3. Site directed mutagenesis on recombinant fusion *Blastocystis* TPI-GAPDH protein

In order to study the effect of crucial amino acid residues on protein function, site-directed mutagenesis was used to produce mutated versions of the *Blastocystis* TPI-GAPDH fusion protein. The QuickChange Lightning Site-Directed Mutagenesis Kit (Agilent Technologies) was used to introduce a point mutation in the wild type (WT) TPI-GAPDH gene at a single position. This mutation was introduced in the TPI part of the gene by mutating a cysteine residue (C 416) into an alanine. The whole construct was amplified according to the PCR conditions mentioned in Table 2.10 in paragraph 2.2.1.2 in materials and methods using the mutation primers mentioned in Table 2.2. The final product of this PCR reaction was cloned into pET-14b. This plasmid was transformed into *E. coli* BL21 cells following the detailed method mentioned in 2.2.1.5.2. Putative clones containing the mutated version of the gene were sent for sequencing to MWG-Eurofins in Germany to confirm the identity of the clones.

Primers used were FTPGA 1401/ FTPGA901/ RTPGA400/ TPI-GAPDHmut check F/ TPI-GAPDHmut check R/ and finally RMM073 and RMM074 for

inserting the mutation at position 417 to change Cysteine in the wild type gene into alanine. Sequences of these primers can be found in Table 2.2 in materials and methods. Figure 6.22 A shows a schematic representation of the sequencing procedure to check the mutation was inserted at the right place. A close up of this region is presented in Figure 6.22 B.

This mutated clone was subsequently checked for proper expression levels as explained in methods 2.2.3.8. This experiment will help identifying whether there is any expression at the colony level and will also give an idea of the variation in gene expression between individual colonies. Differences in expression level between different colonies grown on the same plate after transforming the mutant construct into *E. coli* BL21 cells can be seen in Figure 6.23. A clear difference can be seen between two different plates and also between colonies on the same plate in terms of intensity of gene expression. For further experiments, the clone showing the higher expression level was isolated. This colony was picked and stored in the -80°C freezer after mixing with glycerol to a final percentage of

30%.

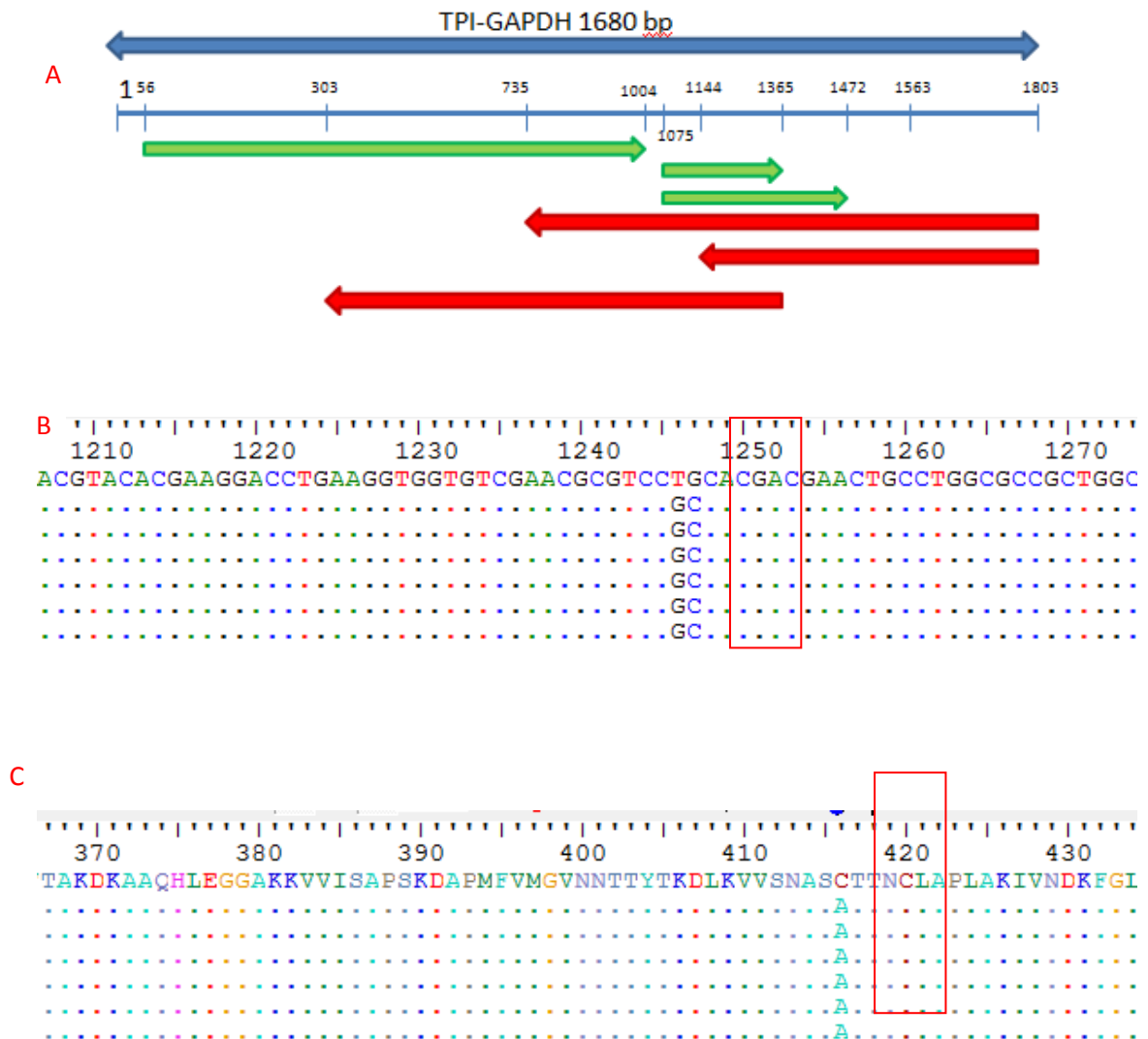


Figure 6. 22 Site-directed mutagenesis of the *Blastocystis* TPI-GAPDH fusion gene. A) Schematic representation of the sequencing strategy to confirm the entire mutated gene. N and C terminal parts were also read from the plasmid to get the double stranded coverage B) Alignment of the mutated region. Top row shows the wild type sequence and subsequent rows show individual clones sent for sequences. The mutated site is clearly visible C) As in B. but now the amino acid sequence is shown where a cysteine is converted to alanine. The point mutation is indicated in red box for both the nucleotide and amino acid sequences. The sequence mutated is indicated to by a red box.

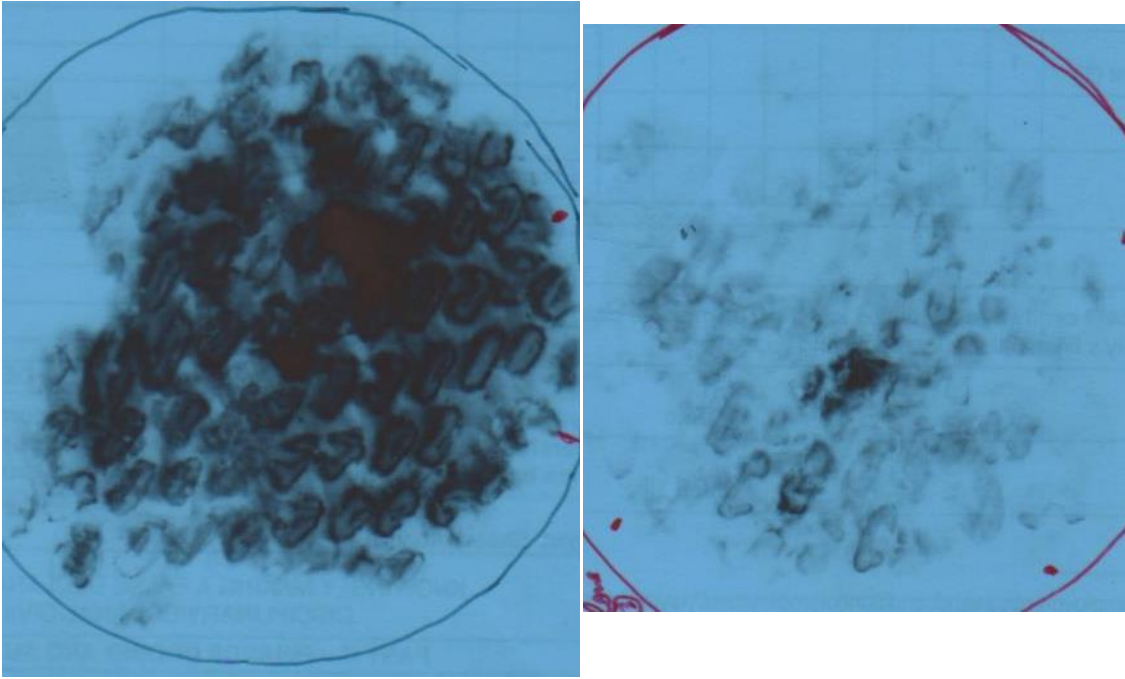


Figure 6. 23 Comparison of *Blastocystis* TPI-GAPDH gene expression levels. Expression levels were compared using anti-his antibody after colony western blotting to identify clones with the highest level of expression of the mutated *Blastocystis* TPI-GAPDH fusion gene. It is clear that differences in expression levels exist between clones but also within clones on the same agar plate.

6.5. Discussion

This chapter aimed to characterize the unusual TPI-GAPDH fusion protein found in *Blastocystis*. Because this protein normally exists as two individual proteins in other organisms, I aimed to characterise this peculiar fusion protein using a set of physical and biochemical approaches, seeking any difference between this fusion protein in *Blastocystis* and other organisms, especially human glycolytic enzymes. As no structure has been solved for such a fusion protein, it was crucial to identify a protein purification strategy which would result in high yields of clean protein folded in its native state. This would enable, or facilitate, the production of good quality crystals that could be used to solve the protein structure. The gene encoding the TPI-GAPDH fusion protein was cloned in several cell lines to screen for the best expression yield. The best of these were *E. coli* BL21 cells, which gave a very high yield when using the auto-induction technique.

Purifying the fusion protein was a challenging process because of the protein's tendency to elute in the void volume during gel filtration chromatography. Several trials have been performed to modify this elution behaviour. Several chromatography approaches were employed, including ion exchange chromatography, nickel affinity chromatography and gel filtration chromatography. Attempts have been made changing the order of purification steps to investigate if that had any role in the final yield. In the end it was decided not to include the ion exchange chromatography step as it proved to be minimizing the yield of the final protein.

After overcoming the problems of the low yield, several crystallization screens were set up to identify ideal crystallization conditions of the fusion protein. Unfortunately, none of these produced any protein crystals which necessitated

further investigation using thermal stability assays. Thermal stability assay experiments gave an indication that the fusion protein was either present as aggregates or in its inactive state. We hypothesized that it might be possible that the fusion protein might be negatively affected by the buffer it was purified in. As protein activity assays for TPI and GAPDH are performed in phosphate buffers, the purification of this protein was repeated using a phosphate buffer ([Bergmeyer 1965](#), [Ganea and Harding 2000](#), [Evans 2014](#)).

This phosphate buffer-purified protein was tested for its aggregated status by DSF experiments. A huge shift in the DFS signal was observed. This suggests that the protein is more stable and possibly in its native form using phosphate buffer.

It is essential to obtain a protein in its native state to be able to solve its structure. For this reason several biochemical and biophysical experiments were performed on this newly purified fusion protein in addition to the single *Blastocystis* GAPDH protein. Most of performed experiments suggested that the polymeric status of the *Blastocystis* TPI-GAPDH protein is a tetramer. *Blastocystis* GAPDH seems to be following the norm as seen for other GAPDH proteins in archaea, bacteria and eukaryotes where it is normally a tetrameric protein, ([Moras, Olsen et al. 1975](#), [Isupov, Fleming et al. 1999](#), [Malay, Bessho et al. 2009](#), [Ayres, Schormann et al. 2014](#)). Both results of the fusion TPI-GAPDH and the stand alone GAPDH were in agreement with the work of Liaud et al who suggested that the TPI-GAPDH fusion protein as found in the diatom *Phaeodactylum tricornutum*, has a tetrameric oligomeric status, similarly to its GAPDH ([Liaud, Lichtle et al. 2000](#)). Their results were based on fast protein liquid chromatography (FPLC) experiments using a Superdex 200 (separation range 10-600 kDa) column for

calculating the molecular weight of the fusion protein. Evans, in her recent MSc work, had concluded that the fusion protein is dimeric, and not a tetramer protein ([Evans 2014](#)). She based her results on high performance liquid chromatography (HPLC) experiments. Nevertheless, Evans did mention, inter alia, the presence of tetrameric TPI-GAPDH, but ignored its significance due to the small percentage, and considered it as an aggregation. However, both groups haven't based their results on structure, and were merely depending on approximations using chromatography columns. My work clearly shows that the protein can be best explained as a tetramer but a more conclusive result will be based on structural biology work. This would clearly show what the tertiary structure of this unusual fusion protein is.

The protein obtained in this chapter was used for the crystallography work explained in detail in chapter seven.

More work will need to be performed to fully characterise the individual proteins that make up this fusion protein, and also to study the actual enzyme kinetics of the fusion protein.

Chapter 7. Crystallization and structure determination of the fusion recombinant *Blastocystis* TPI-GAPDH protein

7.1. Introduction

7.1.1. Background to protein crystallization

Protein crystallization was historically used in the laboratories for many years to demonstrate the purity of the protein and as a method of its purification.

Obtaining macromolecular crystal is only achieved when the solution of the molecules is in a supersaturated state. Crystallization only occurs when the protein concentration is higher than its solubility limits in the solution. This is achieved by manipulating the conditions under which the protein-containing drop, such as precipitant concentration or pH. While this process is still highly experimental and depends mainly on both the patience of the scientist and trials and errors; trying different conditions and screens to get a well ordered protein crystal and hence obtain a good resolution data. This was especially accurate at the start of macromolecular crystallography work, when it was considered as the 'art of crystallization', while scientist efforts to turn this art into science started 15 years ago, and hence the new terminology of crystallogenesis emerged ([Chayen 2004](#)).

Crystallization is a process of nucleation followed by crystal growth. For the nucleation stage a higher protein concentration is required than for the subsequent crystal growth state. The whole crystallization procedure is better explained by the phase diagram shown in Figure 7.1. From the diagram it is clear that the protein state is described as a function of the variables in the crystallization conditions, such as precipitant's concentration, pH, temperature,

additives etc. The conditions of the protein can be classified into under-saturated and super-saturated phases. The protein in the under-saturated state is in its soluble state and will never crystallize in this phase. While the super saturated phase can be divided into three zones; the high super saturation where the protein will precipitate and will not form crystals; the moderate super saturation zone also called the labile zone and the lower super saturation zone. In the moderate super saturation phase the protein nucleation will occur. While in the lower super saturation zone, protein crystals will grow although nucleation is unlikely to happen ([Asherie 2004](#), [Chayen 2004](#)). Both protein crystal growth and nucleation can occur in the labile zone only, while the high supersaturated state can only produce protein precipitants, ([Hausmann, Aguilar Netz et al. 2005](#)). In order to obtain protein crystals we need to get into the lower super saturation zone after some nucleation occurs or crystals seeds are added. There best crystals can be obtained.

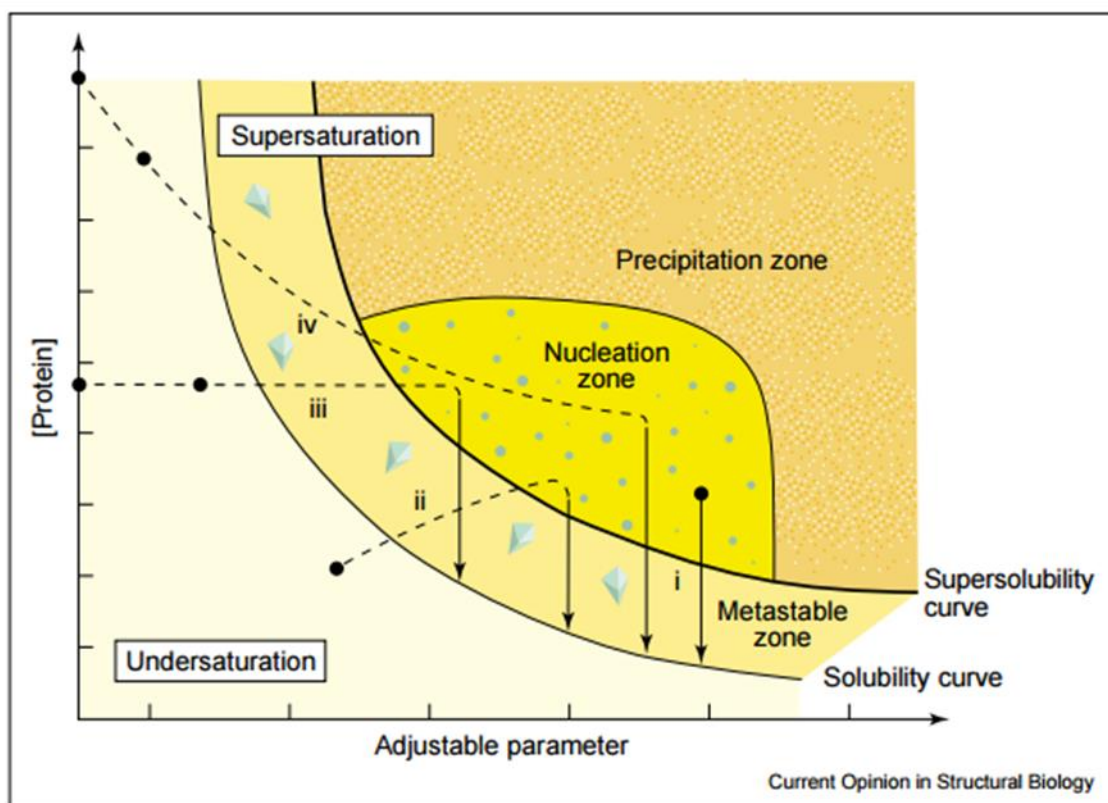


Figure 7. 1 Phase diagram representing the phases that a protein goes through until it forms a crystal. The lines represent the equilibrium between the solid concentration and the free molecule phase. The phases are divided into two major phases of under-saturation and super-saturation. The under saturation phase is the one that the protein is at its maximum soluble state. The super-saturation phase which itself is divided into two regions; the metastable where the crystals start to form; and the labile phase where nucleation takes place. The last phase which is a highly super-saturated state which is called the precipitation state where the protein precipitates out of the solution. Crystals only grow in the first phases of the supersaturated phase. This diagram is taken from Current Opinion in Structure Biology ([Chayen 2004](#)).

Several factors are tested during the process of producing crystals, in an attempt to be able to get the protein to crystallize. These factors include changing the protein concentration, precipitants, additives, pH and temperature.

Naturally, protein crystals grow in a confined environments. As obtaining a crystal is such a challenging process, it is generally limited to modifying the conditions in order to get the final crystal. The procedure of obtaining crystals can be divided into two main steps, the first step is to find the appropriate conditions (physical and chemical conditions) at which the crystals are formed. The second step relies

basically on varying those conditions by making small alternations to the components of the environment in which that initial crystal was formed, this is the optimization step. The hardest step is finding the original conditions, which is generally rather laborious. It has to be taking into account that some proteins simply cannot be crystallized. The optimization step is time consuming as well, but often quite important and critical to enhance the quality of the data obtained to solve the protein structure. Specially that the quality of the crystal is very important when performing the X-ray studies which are aimed at producing best possible data for solving the structure afterwards by getting a better diffraction pattern.

The quality of the diffraction pattern and the crystal resolution limit are related to the degree of internal order of the protein in the crystal. The higher purity and uniformity of the crystal and the more precise its periodic arrangement will result in better diffraction to the higher the resolution. Having a better determination of the atomic position is closely related to the quality of the crystal diffraction. The higher the crystal order and resulting diffraction resolution, the more accurate the final structure would be.

Knowing that the average protein crystal is composed of 50% of its total volume of water and solvent is important; since that those two play an important role in the final data quality. Those liquid channels play a double sword role. On one hand, the liquid channels and cavities are the cause of poor diffraction pattern. This is mainly because the solvent and water make the adjacent molecules far away from each other and separated, resulting in a weak lattice forces, and thus weakening the forces holding them together. On the other hand those solvents help keeping the atoms in their natural environment, which is considered very

essential by biochemists. It also worth mentioning that the kinetics of the crystal growth is three times slower than the kinetics of growth of the small molecules like salt or ion crystals. This is mainly due to the large protein molecule size compared to the small molecule.

Several technical methods are used to obtain crystals; the two main are vapor phase diffusion and the batch method. The vapor phase diffusion method is based on achieving equilibrium between two reservoirs; one containing the precipitating agent and the other reservoir contains the protein droplet mixed with small volume of the precipitant either in a sitting or hanging drop. Over time, water diffuses from the protein to the precipitant reservoir. This decrease in the droplet volume, accompanied by the increase in protein and precipitant concentration, brings the protein into the super saturation state, which might start the crystal nucleation. Once the crystal starts to form, the protein concentration decreases and the solution is brought into metastable growth phase.

In the batch method, the protein and the precipitants are already mixed in one droplet and left undisturbed. Here to obtain a crystal, the protein should already be in a high concentration to start nucleation. The droplet should be covered with oils such as silicon and paraffin oils, allowing slow evaporation resulting in increased concentration of protein and precipitant in the crystal droplet.

7.1.2. X-ray crystallography

X-rays were first discovered in 1901 by Röntgen, who was the first Nobel Prize winner in Physics. Since then X-rays were routinely used for medical work. In 1912, William Henry Bragg and his son William Lawrence Bragg started a different approach for using X-rays; they used diffraction of the X-rays passing

through the crystal to determine the structure and arrangement of molecules comprising the crystal. Bragg the father had adapted a standard microscope into a new one by replacing the normal light source with an X-ray source. After adding a focusing tube, the crystal replaced the prism or the optical grating on the rotating table, and a rotating ionization chamber for detecting the diffracted X-rays replaced the optical telescope ([Jenkin 2001](#)). While on the other hand, William Bragg the son directed his interests into structure solving using his father's- developed microscope, where he thereafter managed to determine the structure of sodium chloride crystal using the new technique in 1913. Nowadays, crystallography is used in a wide range of disciplines, such as physics, materials science, chemistry and medicine. Now a days, X-ray crystallography is a tool used to determine the atomic or molecular structure of the crystal with different properties of the X-rays used in each. The X-rays are also used in medicine and material science since they are absorbed differently by varying tissues and materials, but these has less intensity than used in crystallography. In X-ray crystallography, the X-rays are aimed at the crystal, which will in turn be diffracted by the crystal depending on its crystal symmetry and the contents of its unit cell. The diffraction pattern is registered thereafter and recorded and the images obtained are used afterwards to calculate the three dimensional structure of the protein. X-ray wavelengths used in solving molecular structure range between 0.5-1.5Å, the same region where the sizes of the spaces between atoms lies in. Protein crystal is composed of regular repetition of unit cells which amplifies the diffraction pattern. The unit cell is designated by the three lengths (a, b, and c) and the three angles (alpha, beta and gamma). The more unit cells in a single crystal the stronger the resulting diffraction pattern. The diffraction occurs on parallel planes running through the crystal; those planes have different orientation

and spacing between them. The diffraction pattern is made up of a number of spots (individual reflections), which correspond to the diffraction from the crystal planes.

Bragg's law (Bragg 1913) represents the conditions need to be fulfilled to get a diffraction pattern from the crystal, equation 7.1.

$$n\lambda=2d\sin\theta$$

Equation 7. 1 Bragg's law, n is an integer, λ is the wavelength of the radiation, d is the distance between the lattice planes and θ is the angle of the incidence of the X-ray beam.

Looking at Bragg's law, one can see that all the planes running through crystal are considered as mirror image as long as the interplanar distance equals an integer number of the X-ray wavelength. Each plane gives a reflection spot. The X-rays will add up constructively if they start in phase and reach the crystal plane in phase, Figure 7.2.

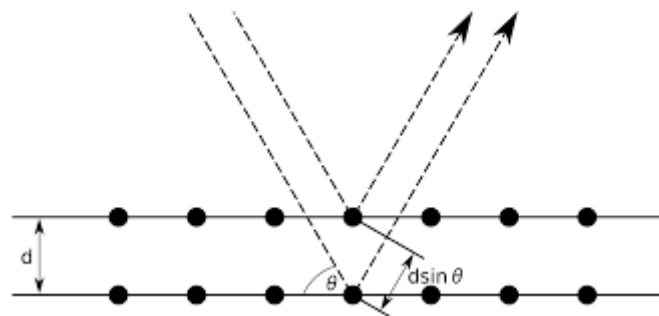


Figure 7. 2 Conditions that satisfy Bragg's law. Figure adapted from (Asherie 2004).

Bragg's law shows that the shorter the spacing between the planes, the greater the angle of diffraction. Reflections are described by a set of indices called Miller

indices (h, k, l), which defines the spacing between the lattice planes of the crystal in reciprocal space. Measuring the spacing will allow calculating the dimensions of the unit cell.

During data collection, crystals are rotated until reaching a plane that satisfies Bragg's law, which will produce diffraction thereafter. This rotation allows collecting data from planes. This is best explained by what is called Ewald's sphere, which explains Bragg's law in three dimensions. Ewald's sphere places the crystal at the center, and has a diameter of $1/\lambda$. Reciprocal lattice points lie on Ewald's sphere and satisfy Bragg's law and hence producing a diffraction spot, Figure 7.3.

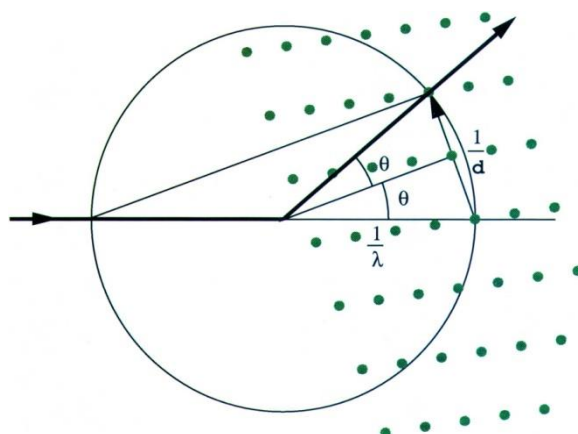


Figure 7. 3 Ewald's sphere. A three dimensional illustration of Bragg's law. Adapted from ([Dauter 1999](#)).

The X-ray diffraction data is used to calculate the electron density throughout the unit cell. To calculate the electron density, both the amplitude $|F_h|$ and the phase of the diffracted X-rays waves need to be calculated. The amplitude can be calculated from the experimental information as it represents the square root of

the intensities. The Phase information is lost during data collection and can't be calculated directly using data collected; instead they are calculated using number of different methods, which will be described later in this paragraph.

Once the phase and the amplitude information are calculated for each of the Miller indices (h, k, and l), they are all combined to obtain the structure factor (F). The total structure factor is a combination of all the atomic structure factors (f_j) described by the Miller indices, Equation 7.2. Structure factor and electron density are related by a Fourier transform.

$$F_{hkl} = |F_h| \exp[i\alpha h] = \sum_j f_j \exp 2\pi i (hx_j + ky_j + lz_j)$$

Equation 7. 2 The structure function equation. F_{hkl} is the structure factor of all Miller indices, f_j is the atomic structure factor, |F_h| is the amplitude of the reflections, and [iαh] is the phase of the reflections.

After calculating the inverse Fourier transform for all Miller indices, we can calculate the electron density for all positions in the unit cell, Equation 7.3.

$$p(x, y, z) = \frac{1}{V} \sum_{hkl} F_{hkl} \exp[-2\pi i (hx + ky + lz)]$$

Equation 7. 3 Electron density calculation in a unit cell. P(x,y,z) is the electron density of all positions in the unit cell. F_{hkl} is the structure factor.

The method used in this project to determine the phase problem that was lost during data collection is the molecular replacement using a model that has a

sequence homology of 25% or more of the unknown structure. To choose a good model, one aims at having a model with the lowest r.m.s.d. (indicated by high sequence identity), lowest deviation from the target structure and a high accuracy, which means that it models a high proportion of the scattering from the target structure ([Evans and McCoy 2008](#)). The chosen model is used to make an estimation of the initial phase. In this method we try all possible orientations and positions of the model in the unknown structure and try to find the best match of the predicted diffraction with the observed diffraction ([Evans and McCoy 2008](#)).

An electron density map that minimizes the bias of the calculated structure factor for the model is being created. This is achieved by subtracting the calculated structure factors from a multiple of the observed amplitudes. This map is referred to as the $2F_o-F_C$ map, Equation 7.4. where F_o-F_C represents the subtraction of the calculated structure factor amplitudes from the observed amplitudes, and is used to highlight regions where an electron density has been observed in the data collected (but has not been observed or computed in the model), and the regions where the model has been built but no electron density has been observed. Once the electron density map has been calculated, the building of the amino acid chain into the map can start using both the calculated and the observed density map.

$$p(x, y, z) = \frac{1}{V} \sum_{hkl} (2 |F_{obs}| - |F_{calc}|) \exp[i\alpha_{calc}(h)]$$

Equation 7. 4 The electron density $2F_o-f_C$ map summation. F_{obs} is the amplitude observed from the model. F_{calc} is the amplitude calculated from the model.

Each protein chain building cycle is always followed by refinement. Each refinement should result in improvement in the R factor. Structural refinement includes applying geometrical restraints on the bonds lengths, angles, and planarity of the bonds. These refinements will be accompanied by alteration in the position of atoms, B-factors and occupancy, which will in turn improve the phases thereafter.

The R-factor is the statistical measurement of how well the calculated amplitudes to match the observed amplitudes in the diffraction pattern, Equation 7.5.

The smaller the R-factor value and the closer it is to zero, the better the calculated density map. Since R-factor is calculated using structure factors used in both structure refinements, is a biased value. Therefore, another statistical R-factor is calculated, called the R_{free} is calculated to overcome the biased problem of R-factor. To calculate R_{free} , the amplitudes of (2-5%) of the reflection spots are used which are not included in the structure refinement process.

$$R = \frac{\sum_h || F_{obs} | - | F_{calc} ||}{\sum_h | F_{obs} |}$$

Equation 7. 5 The R-factor equation. . F_{obs} is the amplitude observed from the model. F_{calc} is the amplitude calculated from the model.

When no further decrease in the R-factor is observed, the refinement is stopped. Thereafter, a validation of the structure is performed using parameters such as bond lengths, angles, and stereo-chemical parameters, where the structure is thought to be correct when those parameters falls within the calculated parameters.

The aim of this chapter is describe solution of the structure of the recombinant fusion protein *Blastocystis* triosephospate isomerase/glyceraldehydes-3-phosphatedehydrogenase (TPI-GAPDH) using X-ray crystallography. *Blastocystis* glyceraldehydes-3-phosphatedehydrogenase (GAPDH) was also crystallized separately and X-ray data were collected in a step aimed at solving its structure and comparing it to the fusion protein. This is the first work reported to solve the structure of the unusual fusion protein *Blastocystis* TPI-GAPDH. The work was performed with the help of Dr Michail Isupov in the Biocatalysis department who helped in data collection and data processing and crystal structure solution.

7.2. Materials and methods

7.2.1. Crystallization

7.2.1.1. Protein preparation

The protein obtained from size exclusion chromatography (gel filtration chromatography: GF) was pooled and concentrated to 60 mg/ml in the case of the *Blastocystis* TPI-GAPDH and 20 mg/ml for *Blastocystis* GAPDH. The concentrated protein is centrifuged at 15,000 x g at 4°C for 10 min before being used in crystallization trials. This centrifugation step is very important as it removes any precipitants or aggregates.

7.2.1.2. Pre – crystallization test: PCT

A very important factor to be considered when setting crystallization trials is the protein concentration. Commercial kits have been developed to test both the homogeneity and the suitability of the protein concentration for the crystal trials. In this study, a commercial kit called PCT™ (Pre crystallization test) from Hampton Research was used. In this experiment, the protein mixed in a certain

concentration with equal volumes of the buffers provided. The outcome of this is important for identifying the next steps. If the protein concentration is very high, an amorphous precipitation will be observed resulting in precipitation and a blurred solution in the mixed well. If after adding the protein a very clear solution is observed, then the protein is too much diluted or in a very low concentration. The optimal case is to have a light precipitation when the protein is mixed with the buffer. The various possibilities that might be encountered and which actions to take are summarised in Table 7.1.

Table 7. 1 Pre-crystallization test (PCT) results and the recommended actions (https://hamptonresearch.com/documents/product/hr00559_9_2-140_142_user_guide.pdf)

Buffer A ₁ */ B ₁ ***	Buffer A ₂ **/ B ₂ ****	Action taken
Heavy precipitation	Heavy precipitation	Dilute the protein
Heavy precipitation	Clear	Perform PCT with B ₁ /B ₂ and check for homogeneity
Heavy precipitation	Light granular precipitate	Set the crystal screen
clear	Heavy precipitation	Perform PCT with B ₁ /B ₂ and check for homogeneity
Light granular precipitate	Heavy precipitation	Set the crystal screen
clear	Clear	Concentrate the protein to half the volume
clear	Light granular precipitate	Set the crystal screen
Light granular precipitate	Clear	Set the crystal screen

* A₁: 0.1M Tris hydrochloride pH 8.5, 2.0 M ammonium sulphate

** A₂: 0.1M Tris hydrochloride pH 8.5, 0.2 M magnesium chloride hexahydrate, 30% w/v Polyethylene glycol 4,000

*** B₁: 0.1M Tris hydrochloride pH 8.5, 1.0 M ammonium sulphate

**** B₂: 0.1M Tris hydrochloride pH 8.5, 0.2 M magnesium chloride hexahydrate, 15% w/v Polyethylene glycol 4,000

Performing this pre – crystallization test is extremely useful and will save time, and most importantly, the precious protein samples as well. In this study the fusion protein *Blastocystis* TPI-GAPDH gave the desired results at 35 mg/ml, while *Blastocystis* GAPDH protein was accepted at around 16 mg/ml.

In the case a protein always results in a heavy precipitation no matter how diluted it is (if it goes beyond 3 mg/ml), then it is advised to check the protein homogeneity using methods such as DLS, native gel electrophoresis, etc. mentioned in details in the previous chapter.

7.2.1.3. Initial crystal trials

Commercial crystal screens were tried for both the *Blastocystis* fusion protein TPI-GAPDH and GAPDH. A list of the screens used is shown in Table 7.2.

Screens were set at the University of Exeter's facility using the micro-batch method and were prepared using an Oryx 6 crystallization robot (Douglas Instruments). The crystal plates which were set using this method were stored at 4 °C and 18°C. Another set of screens were set up using the vapour diffusion method at the Oxford Protein Production Facility (OPPF) at the University of Oxford. The OP PF uses a Hydra robot which transfers screen buffers to the plates where crystallization screening would take place. It is also used for transferring nanoliter amounts of droplets using the dispensing Cartesian Microsys instrument. The set crystal plates were stored in a hotel system (Formulatrix Storage) supplied with its own imaging system that records the growth of the crystals at either 4 or 21°C. These plates were checked afterwards using an automated coding system installed at the crystallography unit at the OP PF. Protein crystals obtained in this study were the ones stored at 21°C.

Table 7. 2 Commercial screens used in this study

Screen name (Hampton research)
JCSG-plus HT-96
PACT premier
Morpheus
Wizard $\frac{3}{4}$
PEG/ion
Index

7.2.1.4. Optimization of crystallization

After obtaining crystals that gave a good diffract under x-ray, further trials were performed in order to maximize the data obtained and to enhance the quality of these data and images collected. The aim of this was to achieve a higher resolution of the data obtained. In addition, the original dataset did not allow getting data of the holo protein, so additional data acquisition was required. For this, Xstep optimization software (Douglas instrument, UK) was used to set new conditions for crystallography. This software varies the concentration of the protein, precipitant used, buffer strength (pH) and any other additives in the primary trial conditions by about $\pm 30\%$. In this work, only the micro-batch method was used. The robot used to set the optimization trial plates was the Oryx 6 crystallization robot (Douglas Instruments, UK).

Each droplet was covered by an oil mixture (Al's oil) (50: 50 of silicon oil and paraffin). The plates were stored at 18°C and checked regularly using a light microscope.

7.2.1.5. Co – Crystallization

The *Blastocystis* fusion protein TPI-GAPDH was tried for co-crystallization with its cofactor; reduced nicotinamide adenine dinucleotide (NADH⁺). The cofactor was incorporated in two ways; in the first instance the cofactor NADH⁺ was added during the final purification step (gel filtration) but no crystal yield was obtained in this instance. The other way of incorporating the cofactor was in the optimization step. Here, the cofactor NADH⁺ was added to the final purified protein, left in contact with the gel filtration- purified protein on ice for 2-3 hours. The combined mixture was subsequently used for setting new crystallization plates. The cofactor was used in concentration of the dissociation factor value 10x K_d) calculated in chapter 6, section 6.4.3. To set the crystal plates a final protein concentration of 35mg/ml was used. The protein-cofactor mixture was centrifuged at 15,000x g before setting the crystal plates.

7.2.2. Preparing crystals for data collection

7.2.2.1. Cryo-cooling straight from droplet using cryo-protectant

The crystals obtained in this study were mounted directly from the crystallization droplet under the light microscope and placed in the cryo-protectant of choice. This cryo-protectant was contained the purification buffer components. The crystal was picked up from a cryoprotectant mixture using cryo-loop was then plunged into liquid nitrogen to be cryo-cooled and ready for the next step, i.e. shooting by x- rays. For a full list of the cryo-protectants used with *Blastocystis*

GAPDH crystals and *Blastocystis* TPI- GAPDH crystal analyzed in this study see Table 7.3 and Table 7.4, respectively. A final cryo-cooler composed of (200mM NaCl, 200mM K₂HPO₄, 16% PEG 3350, 30% PEG 400, pH 8.6) was used for the *Blastocystis* TPI - GAPDH crystal which gave the best X-ray diffraction data.

Table 7. 3 Cryo – protectants and conditions used for *Blastocystis* GAPDH crystals, showing various shapes and relative sizes of the crystals.

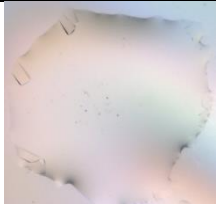
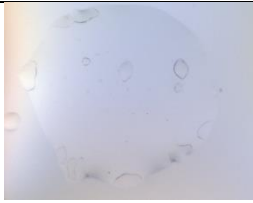
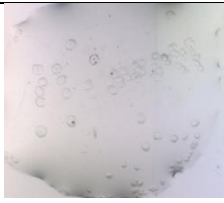
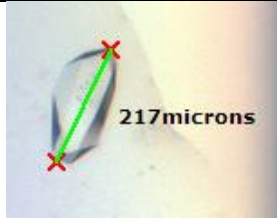
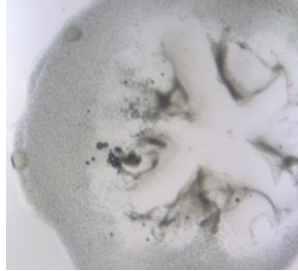
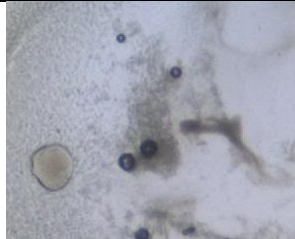
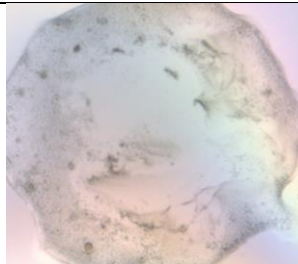
Crystal plate conditions	Screen name	Cryoprotectant used	crystal
0.1 M HEPES 1.1 M sodium malonate 0.5% v/v jeffamine ED-2001 Reagent	JCSG-plus HT-96	Same as the well conditions, exclude jeffamine pH 7-8	
0.1 M TRIS pH=8.50 0.2 M lithium sulphate 1.26 M ammonium sulphate	JCSG-plus HT-96		
0.1 M sodium acetate pH 4.60 2.0 M ammonium sulphate	JCSG-plus HT-96	1.8 M ammonium sulphate 30% glycerol	

Table 7. 4 Cryo – protectants and conditions used for *Blastocystis* TPI - GAPDH crystals, showing various shapes and relative sizes of the crystals.

Crystal plate conditions	Screen name	Cryoprotectant used	crystal
0.2 M potassium phosphate dibasic 20% W/V PEG 3350 pH 9.2	PEG/ION	200 mM NaCl 200 mM K ₂ HPO ₄ 16% PEG 3350 30% PEG 400 pH 8.6	
20% PEG 3350 0.2 M ammonium nitrate	JCSG-plus HT-96	Same as the well condition, pH 5.6.	
30% v/v jeffamine ED-2001 Reagent, pH 7.0, 0.1 M HEPES pH 7.0	Index	Same as well conditions, delete jeffamine. Added 35% PEG44	
20% w/v polyethylene Glycol monomethyl ether 5000 0.1 M bis-Tris pH 6.5	Index	20% PEG 5000 0.1 M Bis-Tris pH 6.5	

7.2.3. X-ray data collection

7.2.3.1. Synchrotron data collection

Data were collected on several beam lines for several crystals at the Diamond Light Source (Oxford, UK). The data presented in this study for the recombinant *Blastocystis* TPI-GAPDH protein were obtained in beam line I03 at Diamond Synchrotron light source using an ADSC Q315 CCD detector. Data were collected in cryo-cooled conditions (100 K under a stream of gaseous nitrogen). A high resolution dataset was collected at wavelength of 0.976Å at a distance of 611.9mm. 900 frames were collected at an oscillation of 0.2° for each frame. For detailed information about the data collection conditions see Table 7.5.

Table 7. 5 Cryo-cooling and X-ray data collection conditions for both *Blastocystis* TPI-GAPDH and *Blastocystis* GAPDH crystals.

Crystal	<i>Blastocystis</i> TPI- GAPDH	<i>Blastocystis</i> GAPDH
Cryo-protectant	200 mM NaCl 200 mM K ₂ HPO ₄ 16% PEG 3350 30% PEG 400 pH 8.6	1.8 M ammonium sulphate 30% glycerol
Cryo-cooling	Straight from drop in addition to the purification buffer	Straight from drop in addition to the purification buffer
Beam	Diamond synchrotron I03	Diamond synchrotron I03
Wavelength (Å)	0.976 (SAD)	0.976(SAD)
Distance (nm)	611.9	611.9
No. frames	900	900
Oscillation (°)	0.2	0.2

7.2.4. Structure determination

7.2.4.1. Data processing

Data were processed using the XDS ([Kabsch 2010](#), [Kabsch 2010](#)), Pointless (CCP4), and xia2 (CCP4) programs ([Collaborative Computational Project 1994](#), [Evans 2006](#)). This part of the work kindly performed by Dr Michail Isupov.

7.2.4.2. Phase determination

Phases were determined using molecular replacement using MOLREP program ([Vagin and Teplyakov 2010](#)), and REFMAC5 ([Murshudov, Skubak et al. 2011](#)) for refinement cycles after every amino acid building cycle. This procedure is discussed in more details later on. Two homologous models were used for this tetramer model method. A human placental GAPDH structure (PDB IU8F) as a GAPDH tetramer model with 67.5% identity ([Jenkins and Tanner 2006](#)) and the mealworm (*Tenebrio molitor*) TPI structure (PDB 2I9E) as a TPI dimer model with 53.2% identity ([Knobeloch, Schmidt et al. 2010](#)). Both were used for molecular replacement methods and details can be found in the results of the *Blastocystis* TPI-GAPDH structure solution later in this chapter. This part and all the following parts of molecular replacement, model building, refinement and validation were performed under the kind help and supervision and guidance of Dr Michail Isupov.

7.2.4.3. Model building and refinement

Electron density maps were calculated and the structure was fitted to best fit the $2F_0-F_c$ and F_0-F_c maps. The protein was built using the program COOT ([Emsley and Cowtan 2004](#)). After each session of model rebuilding performed using COOT, a maximum likelihood refinement was performed using REFMAC 5.2 ([Murshudov, Skubak et al. 2011](#)). Phases were improved by four-fold averaging

using density modification (DM) ([Cowtan 2010](#)). Crystallographic origin ambiguity was resolved using Zanuda ([Lebedev and Isupov 2014](#)).

7.2.4.4. Structure validation

The final structure after refinement was checked for quality using the program PROCHECK ([Laskowski, MacArthur et al. 1993](#)).

7.3. Results

The work in this chapter was divided into two parts. The main goal was to solve the structure of the unusual *Blastocystis* TPI-GAPDH fusion protein. The recombinant *Blastocystis* GAPDH was also studied alongside this work to be used as a reference as well as for comparing the structure of the fusion protein with the structure of the stand-alone GAPDH protein(truncated from the fusion protein) of the same organism. Unfortunately, the *Blastocystis* GAPDH structure was not solved due to the poor data quality obtained with a resolution of about 10 Å, which doesn't allow for solving the structure. Thereafter, the *Blastocystis* GAPDH structure solution work was ceased at the data collection step. In the following sections I will describe the details of work done on both proteins, with detailed explanations on only the recombinant *Blastocystis* TPI-GAPDH fusion protein structure solution.

7.3.1. Crystallization of the *Blastocystis* TPI-GAPDH fusion protein

The gel filtration purified recombinant *Blastocystis* TPI-GAPDH fusion protein was concentrated to a final concentration of 35 mg/ml using Amicon Ultra-15 centrifugal filter units with a cut off 30kDa. This concentration was determined

after being tested for optimal concentration to be utilized using the PCT kit as mentioned in chapter 7, section 7.2.1.2. Crystallization trials of the protein were conducted using the commercial screens mentioned in Table 7.2. The screening plates were set by the Hydra instrument under the vapor diffusion method. Initial results produced several crystals in different screen buffers, yielding varying sizes and shapes. Only one crystal diffracted well at a moderate resolution value of 2.6Å in the synchrotron (Figure 7.4), under the following conditions: 200 mM sodium chloride (NaCl), 200mM potassium phosphate dibasic (K₂HPO₄), 16% PEG 3350, 30% PEG 400, pH 8.6.

Other crystals have shown some diffraction but not of good quality (i.e. low resolution), (see Figure 7.5 A, B, and C). The low resolution value might be attributed to either the fact the protein is not ordered enough or might be related to the cryoprotectant choice. In the case that the cryoprotectant choice was not suitable, ice rings would form. Ice rings in crystals are known for their negative effect as they destroy the protein structure, and hence lower the quality of the collected data. Based on the initial performed screening, further screening trials were performed aiming at optimizing the primary conditions. Optimization conditions varied from the initial screen in the percentages of both the buffers and the precipitants where a crystal growth was reported. The optimization experiments were performed using the micro-batch method, see Appendix 3 for optimization conditions. Several types of crystals were obtained, harvested to be prepared for analyses at the synchrotron. Unfortunately, none of these crystals diffracted indicating that they were not protein crystals.

From the data collected, no electron density in the active site was obtained, so the ligand could not be detected in the final structure. Analyzing the recombinant

Blastocystis TPI-GAPDH protein structure using the data obtained from the first crystal demonstrated that the crystal is lacking the cofactor in the active site of the GAPDH protein. Hence, a new approach was taken in the following optimization steps. Optimization was aimed at enhancing the data quality and to get data with electron density around the cofactor NADH⁺. To achieve this, the cofactor was added by using one of two approaches. During the final gel filtration purification step, where there is a chance that NADH⁺ has been washed off during the consecutive purification steps. So NADH⁺ was added during the final GF purification step. The other approach was to add the NADH⁺ cofactor after all the purification steps were performed. This was performed by a simple incubation step by mixing the NADH⁺ cofactor with the protein on ice in the proportions as mentioned in chapter 7, section 7.2.1.5. Both approaches resulted in crystals with varied sizes and shapes, but unfortunately, both failed to yield any diffraction.

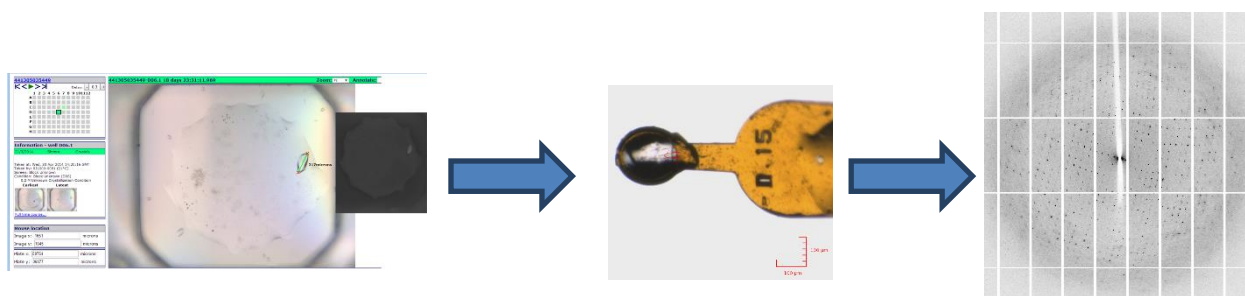


Figure 7.4 Crystallization process flow showing the steps to obtain the fusion protein structure from setting crystal trials to diffracting the obtained crystal in the synchrotron.

7.3.2. Crystallization of the *Blastocystis* GAPDH protein

Blastocystis GAPDH was expressed, purified and prepared for further crystallization work. The protein sample was treated as described in chapter 7, section 7.2.1.3. The same set of screening buffers used in the crystallization trials of the *Blastocystis* TPI-GAPDH fusion protein was applied for *Blastocystis* GAPDH as well. Several protein crystals were harvested from this work and exposed to X-ray in the synchrotron at Diamond Light Source. Data obtained

were all with low resolution and hence no data could be processed for structure solution. The low resolution of the data could be enhanced by repeating the experiments with choosing better cryo-coolers since most crystals (different sizes and varied buffers as presented in Figure 7.5/D, E, F) have shown crystal ice rings after being exposed to X-rays. Different approaches might be investigated to enhance the quality of data obtained.

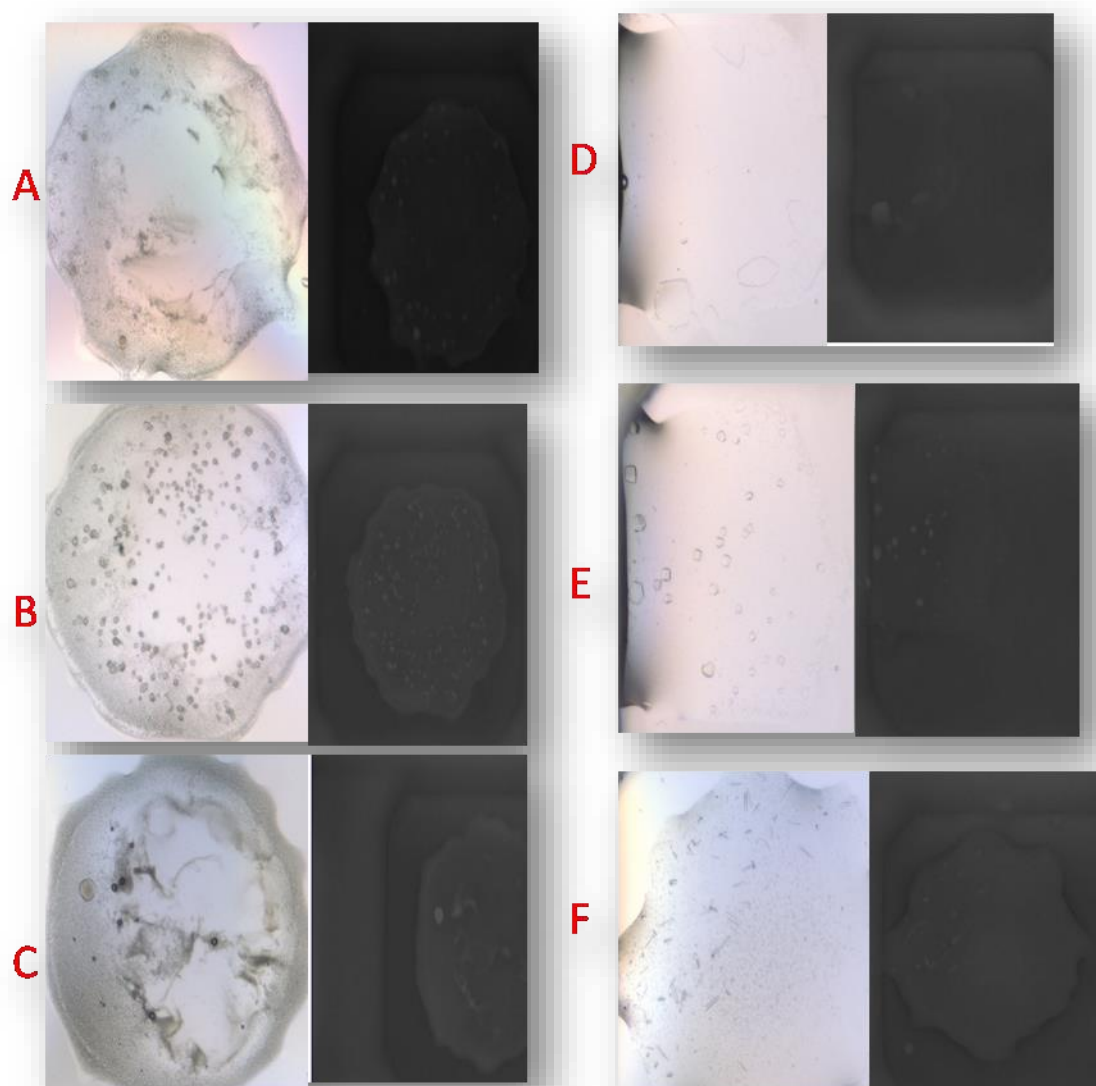


Figure 7. 5 Examples of protein crystals obtained in several commercial buffer screens for both *Blastocystis* TPI-GAPDH and *Blastocystis* GAPDH proteins, with the droplet conditions they grew in. 1) *Blastocystis* TPI-GAPDH A: Index (20% w/v Polyethylene Glycol Monomethyl Ether 5000, 0.1 M bis-Tris, and pH 6.5. B: PACT premier (0.1 M MES, pH 6.0, 20% w/v Polyethylene Glycol 6000, 0.2 M Ammonium Chloride), C: Index (30% v/v Jeffamine ED-2001 Reagent pH 7.0, 0.1 M HEPES pH 7.0) . 2) *Blastocystis* GAPDH D: Wizard 3/4 (20% (v/v) 1,4-butanedio, 100mM HEPES/ Sodium hydroxide pH 7.5, 200 mM Sodium chloride). E: Wizard 3/4 (30% (w/v) PEG 3000, 100 mM Tris base/ Hydrochloric acid pH 7.0, 200 mM Sodium chloride). F: PEG / ion (0.2 M Sodium phosphate dibasic dehydrate, 20% w/v Polyethylene glycol 3,350, pH 9.1).

7.3.3. Structure determination of the fusion protein

7.3.3.1. X-ray data collection

Several crystals have been harvested and tested at the synchrotron for diffraction. Most of the crystals showed the formation of ice rings which negatively affects the diffraction and hence the data quality. Fortunately, one crystal of the *Blastocystis* TPI-GAPDH fusion protein gave a good diffraction results and this diffracted to 2.6Å. Data were collected at Diamond Light Source (Oxford, UK) as mentioned in methods 7.2.3.1. Indexing and scaling were performed and revealed that the crystal belongs to the space group C 1 2 1, with the following cell parameters ($a = 179.5$, $b = 77$, $c = 212.3$ Å / $\alpha = \gamma = 90$, $\beta = 102.78^\circ$). The ice ring regions obtained from frozen parts of the crystal were excluded from data processing, and to obtain a higher completeness of the data the loop was bent 360° to minimize the blind region. Data completeness was 99.7%. Table 7.6 describes the data processing statistics for the *Blastocystis* TPI-GAPDH fusion protein.

Table 7. 6 Data processing statistics and solution of the recombinant protein *Blastocystis* TPI-GAPDH.

Crystal	Recombinant <i>Blastocystis</i> TPI-GAPDH
X – ray source	I03
Wavelength (Å)	0.97625 (SAD)
Resolution range (Å)	2.6 -2.7 (103.52)
Space group	C 1 2 1
Cell parameter	A= 179.5, b= 76.9, c= 212.3 α = 90, β = 102.7, γ = 90
No. parameters in asymmetric unit	1
No. of measured reflections	279228
No. unique reflections	83654
Completeness (%)	99.7
Multiplicity	3.3
	70.5
Redundancy	0.049
(1)/ σ (I)	12.2 (1.7)
Rmerge † (%)	0.049 (0.682)

7.3.3.2. Structure solution

A dimeric *coleopteran* triosephosphate isomerase (TPI) from *Tenebrio molitor* (yellow mealworm beetle) (PDB 2I9E) and a tetrameric glyceraldehydes-3-phosphate dehydrogenase (GAPDH) (PDB 1U8F) were used for molecular replacement using the program MOLREP ([Vagin and Teplyakov 2010](#)).

The crystal packing suggests four TPI-GAPDH monomers in the asymmetric unit of the crystal. The molecular replacement has positioned a single GAPDH tetramer and two TPI dimers in the new crystal cell with high contrast. Single

monomer of each protein was chosen to carry out the structure building, and then extended by non-crystallographic symmetry operators to have a full tetramer.

The correct rotation function peaks for the tetrameric GAPDH model calculated at 3Å resolution were in 27-28 σ range. The rotation function peaks for the dimeric TPI model were 12.8 and 12.5 σ . Translation function was calculated at 4Å resolution. The correlation coefficient for the correct position of GAPDH tetramer was 47% while the correlation coefficient was 25% for the wrong solution. With GAPDH tetramer fixed the correlation was 53% for the first TIM dimer positioned and 56% for the second TIM dimer. Each subunit was built separately and the model was subjected to rigid – body refinement implemented in MOLREP. The final resulting model was refined using REFMAC ([Murshudov, Vagin et al. 1997](#)). The resulting model did not refine well due to the pseudo-origin problem, since native Patterson function contained strong (40% of origin) peaks. The program ZANUDA resolved origin ambiguity and further refinement and model rebuilding went smoothly.

7.3.3.3. Model building and validation

7.3.3.3.1. Apoenzyme

Since the link between the GAPDH and TPI parts is flexible, the tetramer did not follow a strict 222 molecular symmetry. The GAPDH component did follow this symmetry, Since rebuilding of four copies of each chain was too time consuming, one copy of each chain was originally rebuilt on graphics and then it was multiplied four times for each protein chain. The structure solution was manually built in COOT (Emsley and Cowtan, 2004) and refined using REFMAC version 5.6 ([Murshudov, Vagin et al. 1997](#)). The R-factor and Rfree were 17.215% and

23.72% respectively. No density was observed in the active site of the GAPDH structure. The final refinement statistics for the native TPI-GAPDH and validation results are shown in Table 7.6 and the Ramachandran plot can be accessed in Appendix 4.

Finally the protein was built of four subunits; each subunit consists of 603 amino acids forming the monomer unit of TPI-GAPDH.

The N terminal amino acids (1-18) containing His-tag and part of the signal peptide were not visible in the electron density and were not modeled. 135 water and several phosphate molecules were modeled. REFMAC gave an estimated r.m.s. value of 0.922 in the coordinates based on the R_{free} value. The most favoured regions of the Ramachandran plots were determined by PROCHECK contain (93.04%) the overall G factor.

The statistics for the final round of refinement of the structure are given in Table 7.7. The Ramachandran plot is displayed in Appendix 5.

There are 18044 atoms in the molecule in the final structure analysis. The final structure showed four chains: chain A running from amino acid number 18 until 603, and same for chain C. chains B and D both ran from amino acid 20 till 603. Correlation coefficient F_o-F_c : 0.965, and correlation coefficient F_o-F_c free: 0.928.

Table 7. 7 The final X – ray data refinement statistics for *Blastocystis* TPI – GAPDH structure.

Crystal	<i>Blastocystis</i> TPI – GAPDH
Refinement	
Resolution (Å)	2.6
R factor/ Rfree (%)	0.17/0.23
No. of protein atoms	17772
No. of ligand atoms	0
No. of water atoms	135
R.m.s. deviation from ideality	0.92
Bond lengths (Å)	0.012
Bond angels (°)	1.675
Ramachandran plot (%)	
Most favored regions	93%
Additional allowed regions	5.24%
Generously allowed regions	1.89%
Average B values (Å²)	74.4
Overall	
Protein atoms	2340
Ligand atoms	0

7.4. Discussion

7.4.1. Secondary structure analysis:

The PDB file generated for the *Blastocystis* TPI-GAPDH fusion protein was submitted at several structure analysis servers to analyze its secondary structure and special features. Submitting the PDB file to PolyView 2D, showed that the protein comprises four chains, each chain contains alpha helices and beta sheets, Figure 7.6. Looking into this result file of the secondary structure of the

protein, one can see a pattern of $(\beta\alpha)_4$ in the first 200 amino acids corresponding to the TPI sequence. This is in accordance with the universal structure of TIM barrel which TPI belongs to, confirming the TIM barrel of TPI. And the Rossmann fold for GAPDH which is $(\beta-\alpha-\beta)$ fold at the nucleotide binding domain (NBD) and the α/β fold at the catalytic domain (CAD). The results were confirmed using another tool using ESPrpt3 program, Figure 7.7 which shows that the protein consists of four units, each composed of 24 β sheets and either 19 or 20 α helix, making a total of 96 β sheets and 78 α helix in the whole protein unit structure.

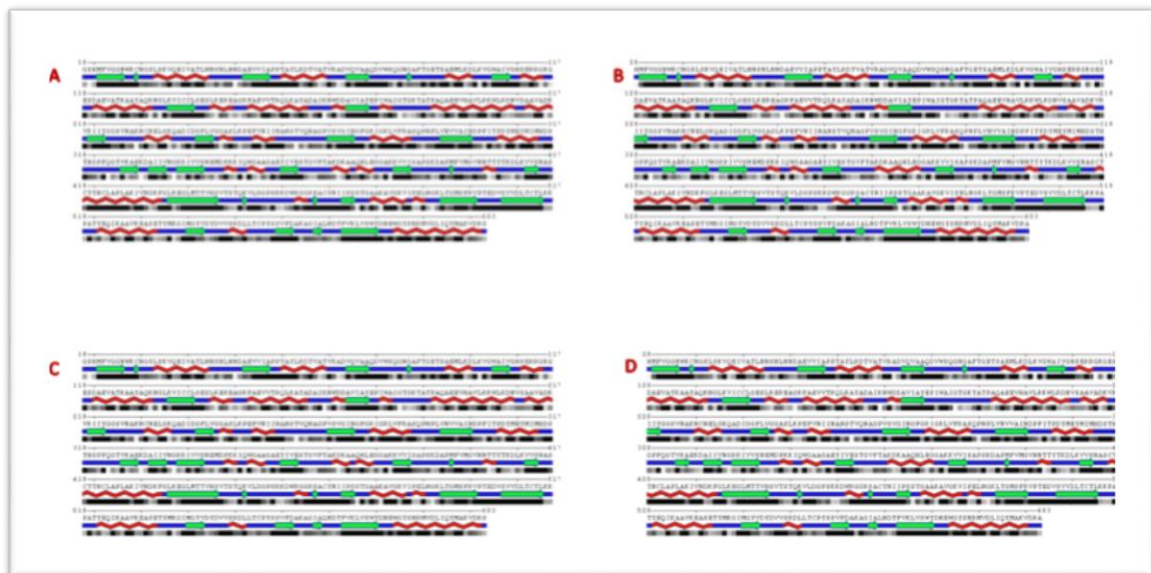


Figure 7. 6 The secondary structure of the protein *Blastocystis* TPI-GAPDH fusion protein showing that the protein consists of four identical chains. The figure obtained by POLYVIEW-2D program.

The green arrow represents beta strands. The red zigzag represents alpha helices.



Figure 7.7 Flat figure showing the amino acid sequence of the PDB entry file of TPI-GAPDH fusion protein adorned with secondary structure elements presented on top (helices with squiggles, β -strands with arrows and turns with TT letters). This secondary structure only shows two subunits since the other two are identical. Solvent accessibility is rendered by a first bar below the sequence (blue is accessible, cyan is intermediate, white is buried) and hydropathy scale per residue showed by a second bar below (Pink is hydrophobic, white is neutral, and cyan is hydrophilic). The figure is produced by ESPrpt 3 program.

7.4.2. Active sites and signature motifs analysis

The overall structure of the TPI-GAPDH fusion protein is similar to the typical proteins TPI and GAPDH that makes its building body. The TPI sequence covered the region between amino acid residues 20-260. GAPDH was divided into two main domains, a GAPDH N (nucleotide binding domain) (270-418) and a GAPDH C (catalytic domain) (423-580) according to a Pfam sequence search. The TPI region showed significant residues playing an important role in substrate

binding and active site catalysis; asparagine (N) and lysine (K) at position 26 and 28 respectively were detected as substrate binding positions. Histidine (H) and glutamic acid (E) at position 110 and 179, respectively, were identified as an electrophile and a proton acceptor, respectively, with the region 177-187 (AYEPIWAIGTG) identified as a catalytic site on TPI region as well which covers loop 6 region whose position is important in activity and substrate binding.

In the GAPDH region cysteine was identified as a nucleophilic amino acid in the active site region of GAPDH (416-423) (ASCTTNCL). An important motif was also detected in this structure, a consensus N-terminal Glycine-rich phosphate-binding loop with a consensus sequence GXGXXG which in *Blastocystis* is (GFGRIG) which corresponds to a Rossmann fold motif indicating an NAD binding property of the protein. This binding site is located at the N-terminal of GAPDH. All the main features of TPI-GAPDH sequence are presented in detail in Figure 7.8 in a schematic representation of the fusion protein sequence.

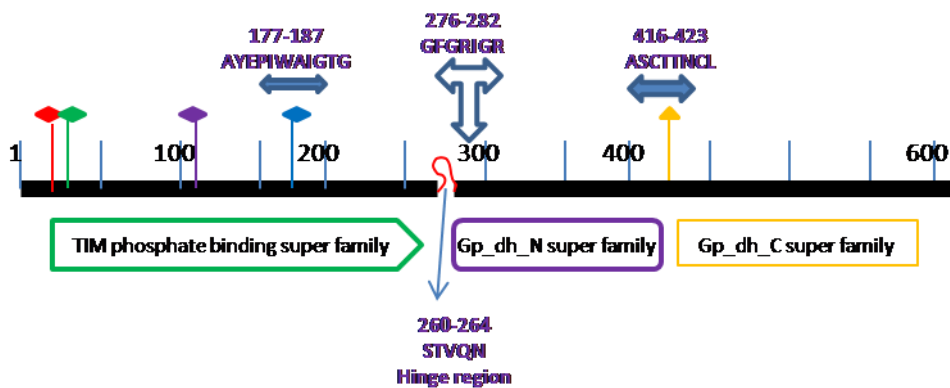


Figure 7. 8 Schematic representation of *Blastocystis* TPI-GAPDH protein with active sites and main domains found in the structure as expected on Prosite.Expasy. Three main domains were detected, TIM phosphate binding, GAPDH N, and C super family. The diamond shape represents active sites belonging to the TPI domain. The triangle shape represents active site belonging to the GAPDH domain. Red: Amino acid position 26 (N), green: Amino acid position 28 (K), purple: amino acid position 110 (H), blue: amino acid position 179 (E), and orange: amino acid position 418 (C). The double sided arrow indicates the active site region in: TPI (177-187) (AYEPIWAIGTG), GAPDH (416-423) (ASCTTNCL). The GAPDH motif (GXGXXG) is also presented on the figure covering (276-282) (GFGRIGR). The red loop represents the joint region in the fusion protein covering five amino acids in the sequence.

7.4.3. Tertiary structure analysis

The *Blastocystis* TPI-GAPDH fusion protein structure solved in this project was identified to be a tetramer. It consists of two triosephosphate isomerase (TPI) dimers and a tetramer of glyceraldehydes-3-phosphate dehydrogenase (GAPDH).

To discuss the three dimensional structure of the fusion protein, I will divide the section into two sections, one for TPI and one for GAPDH.

7.4.3.1. Triosephosphate isomerase (TPI)

The three dimensional structure of TPI was reported in prokaryotes, unicellular eukaryotes, higher eukaryotes and Archaea. It is generally found that catalytically active TPI is usually a dimer, but cases of different oligomeric state have been described such as a monomer (mutants and not biologically active one), a homotetramer in Archaeal *hyperthermophiles Thermoproteus tenax* and *Methanocaldococcus jannaschii* in addition to a human tetrameric TPI ([Walden, Taylor et al. 2004](#), [Walden, Taylor et al. 2004](#), [Kinoshita, Maruki et al. 2005](#), [Gayathri, Banerjee et al. 2007](#)). The protein belongs to the TIM family and takes the $(\beta\alpha)_8$ -barrel assembly. TPI is a highly conserved protein, especially in the amino acids that plays a role in the ligand binding, catalysis, subunit interactions or stabilizing the protein structure.

7.4.3.1.1.1. Active site: catalytic residues and catalytic region:

As triosephosphate isomerase is considered as a highly conserved protein, its active site region is one of the main features that need to be inspected closely to check for activity of the protein.

Eye inspection of the superimposed structure of the studied protein in comparison with the model used to build the structure; Coleopteran TPI (PDB entry 2I9E) ([Knobeloch, Schmidt et al. 2010](#)) shows the conservation of the main residues in the active site of TPI subunits: Asn10, Lys12, His94, and Glu164 in comparison to the corresponding residues; Asn26, Lys28, His110, and Glu179 in *Blastocystis* TPI-GAPDH. Loop 6 residues constitutes the amino acids from Pro180- Ala190; (PIWAIGTGTKA), which are mostly hydrophobic amino acids. This is in agreement with the fact that once the loop closes it needs to be buried inside the active site pocket inside the protein core, and hence should be hydrophobic. A close up of the conserved catalytic residues in the *Blastocystis* TPI monomer active site domain is shown in Figure 7.9 in a cartoon representation showing the main conserved residues of the active site in addition to the 6th lid-loop conformation formed.

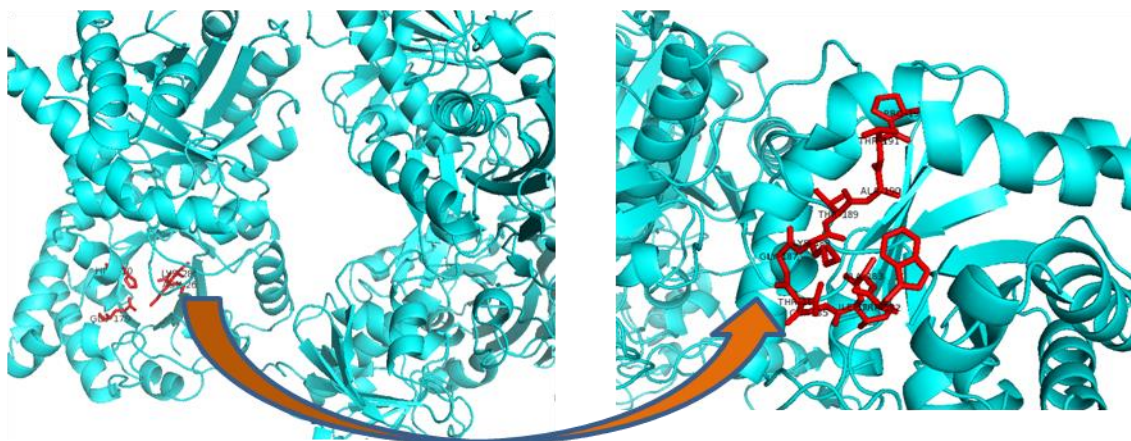


Figure 7. 9 Cartoon representation of the catalytic site residues of the TPI subunit of the fusion *Blastocystis* TPI-GAPDH. The active site residues are displayed as red sticks on the left side figure to highlight the location of the active site. Residues representing the active site are: Asn26, Lys28, His110, and Glu179 respectively. The right side figure is a close up cartoon presentation of the ten residues forming the lid-loop 6 region P-A (covering 180-190).

To determine the conformation of loop 6 in TPI fraction of *Blastocystis* fusion protein TPI-GAPDH, structural alignment was carried out with other TPI proteins

studied so far, particularly the rabbit and the trypanosomal TPI protein as well. The rabbit triosephosphate isomerase protein showed unusual conformation of loop 6 in closed and open conformation in the same protein crystal (1R2R) (PDB entry: 1R2R, 1R2S, 1R2T) ([Aparicio, Ferreira et al. 2003](#)). While *Trypanosoma brucei brucei* TPI protein was unique in showing another conformation which was an almost closed conformation in one of the subunits in addition to an open conformation in the other subunit (PDB entry 5TIM) ([Wierenga, Noble et al. 1991](#)), see Table 7.8 for a summary of all the models studied and their flexible loop conformation.

Superimposing triosephosphate isomerase from both the *Trypanosoma Brucei Brucei* (PDB entry 5TIM) and the fusion protein TPI-GAPDH of *Blastocystis* shows two possible conformation of the flexible loop (loop 6). According to Wierenga study, TPI has shown two conformations; open and an almost closed conformation. Aligning with *Blastocystis* TPI gave also two conformations, which means that it is either open or almost closed. To conclude which conformation *Blastocystis* TPI is taken, it was aligned with 1R2T from rabbit, and showed different conformation from it. Knowing that 1R2T had shown an open conformation only, we can conclude that TPI from *Blastocystis* TPI-GAPDH is not taking the open conformation, as shown in Figure 7.10/A and B (1,2,3). This was also confirmed by superimposing with the model protein used to solve the structure (2i9e), which also takes only the open conformation, Figure 7.11. Those two comparisons lead to the conclusion that TPI from *Blastocystis* TPI-GAPDH is taken an almost closed conformation. Superimposing both; loop 6 in TPI from *Blastocystis* TPI-GAPDH and the TIM structure of the rabbit muscle TPI, shows positioning of the lysine residue (K186) in a swung out conformation, Figure 7.12

A thorough comparison between several TPI studied proteins and the protein of study, confirms that TPI in *Blastocystis* fusion protein takes the almost closed confirmation, when compared to all possible solutions of open/ closed/ almost closed loop lids of previously studied proteins,

This is quite surprising result for two reasons. First, no substrate was added in the crystal plate solution or in purification steps. Second, the electron density obtained didn't show any density for a cofactor in the active site. Knowing that the open conformation of loop 6 in the active site of triosephosphate isomerase is usually found in apo structures, while the close conformation is found in the holo structure, where the protein forms complexes with substrate analogous, phosphate containing ligands, all add to the uniqueness of this finding. All the possible solutions for loop 6 lid conformation were superimposed to TPI unit of *Blastocystis* TPI-GAPDH, and also superimposed between both the rabbit and the trypanosomal TPI having shown special features and differences in their subunits ([Wierenga, Noble et al. 1991](#), [Aparicio, Ferreira et al. 2003](#)).

Table 7. 8 Summary of several triosephosphate isomerase flexible loop conformations in several organisms.

Protein name	PDB entry	Flexible loop conformation
<i>Coleopteran</i> TPI	2I9E	Open
Rabbit <i>Oryctolagus cuniculus</i> TPI	1R2R	Open and closed
	1R2S	Open
	1R2T	Open
<i>Trypanosomal Brucei Brucei</i> TPI	5TIM	Open and almost closed
<i>Blastocystis</i> TPI (study protein)	-	Almost closed

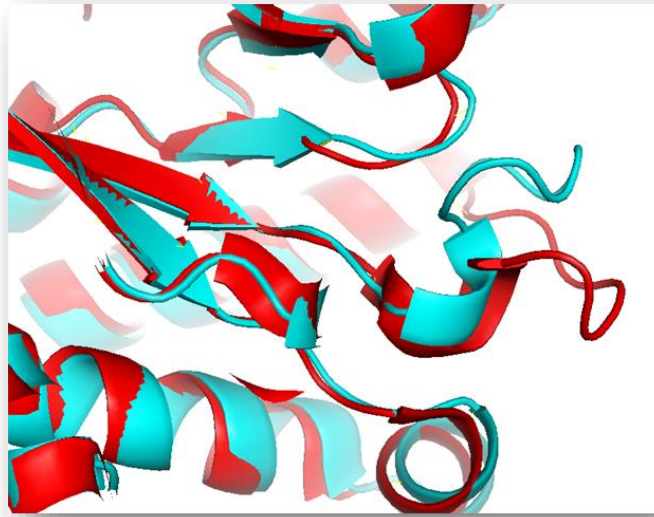
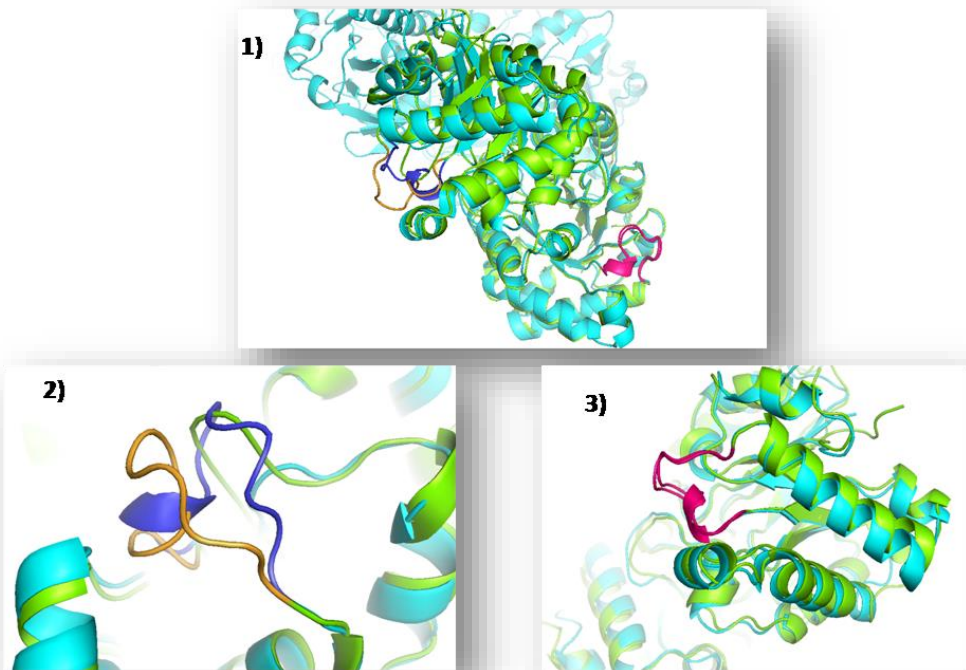
A**B**

Figure 7. 10 Cartoon representation of lid loop 6 in TPI subunit of several proteins including rabbit TPI (PDB entry 5TIM), trypanosomal TPI (PDB entry: 1R2R, 1R2S, 1R2T) and the TPI subunit of the protein of study *Blastocystis* TPI-GAPDH protein. TPI of *Blastocystis* TPI-GAPDH protein was always represented in cyan. A) 1R2T (open) (hot pink) superimposed to TPI of study, showing a closed or almost closed conformation of the loop. B) 5TIM (green) superimposed to *Blastocystis* TPI of study (1), showing both open (2) and almost closed (3) conformation of the flexible loop (loop 6).

A



B

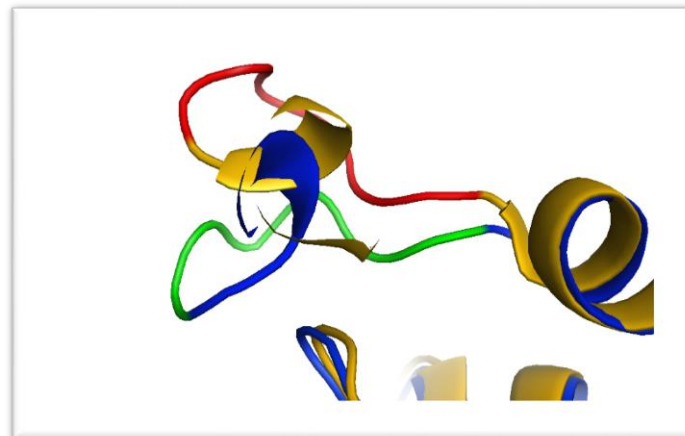


Figure 7. 11 A) Stick representation of the active site residues in a TPI dimer of the *Blastocystis* TPI-GAPDH structure, showing the localization of the active site in the interface between the two molecules. B) Superimposition of loop 6 in a TPI subunit of *Blastocystis* TPI-GAPDH with the model protein coleopteran TPI (PDB entry 2i9e). The fusion protein is in blue and its loop is green while the coleopteran is in yellow and its loop in red.

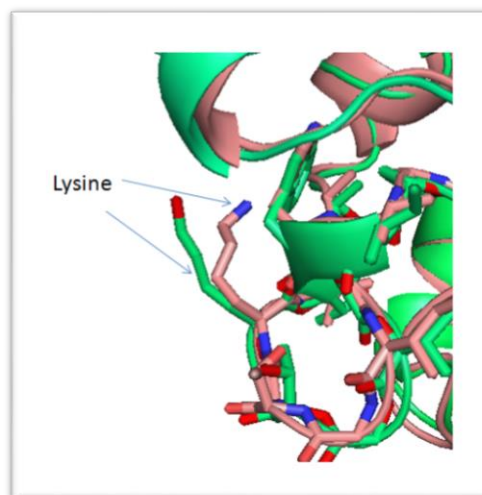


Figure 7. 12 Superimposition of active site loop 6 in both the studied protein *Blastocystis* TPI-GAPDH and TIM from rabbit muscle. The studied protein is shown in green and the model (PDB entry 5TIM) is presented in pink colour, with only lysine residue getting different orientation (swung out conformation).

From what previously discussed, the overall structure of TPI (in *Blastocystis* TPI-GAPDH protein), takes almost a similar conformation to other usual TPI structures known so far as expected, with one phosphate group found in the active site of each unit (shown later in figure 7.15/2). Only one remarkable difference was found which is the unusual conformation of the flexible loop 6. Several other cases were also reported for a closed conformation of loop 6 in an apo enzyme. Those cases of conformational variability were attributed to an inherent flexible nature of the loop, and to changes in the crystallographic environment.

7.4.3.2. Glyeraldehyde 3- phosphate dehydrogenase (GAPDH)

GAPDH structure was reported in a wide range of the kingdom of life, including eukaryotes, bacteria and *archaea* ([Markos, Miretsky et al. 1993](#), [Isupov, Fleming et al. 1999](#), [Malay, Bessho et al. 2009](#), [Zhao, Liu et al. 2011](#)). This protein is usually reported to be as a tetramer. GAPDH sequence of the *Blastocystis* TPI-GAPDH fusion protein is composed of three distinct domains (as analyzed by ExPASy- PROSITE); the nucleotide binding domain covering the region 270-418 (NBD), the catalytic domain (CAD) (423-580) and finally the relatively short domain; C-terminal domain (CTD) covering (580-603), consisting of one alpha helices only, knowing that this short domain used to be considered as part of the NBD. The N-terminal consists of parallel sheets surrounded by helices forming a classical Rossmann fold. While the catalytic domain consists of anti-parallel beta sheets surrounded by helices. This domain also contains the active site residues and the S-loop, Figure 7.13/ A, B, C, shows the total secondary structure of GAPDH, the N-terminal and the catalytic domain respectively.

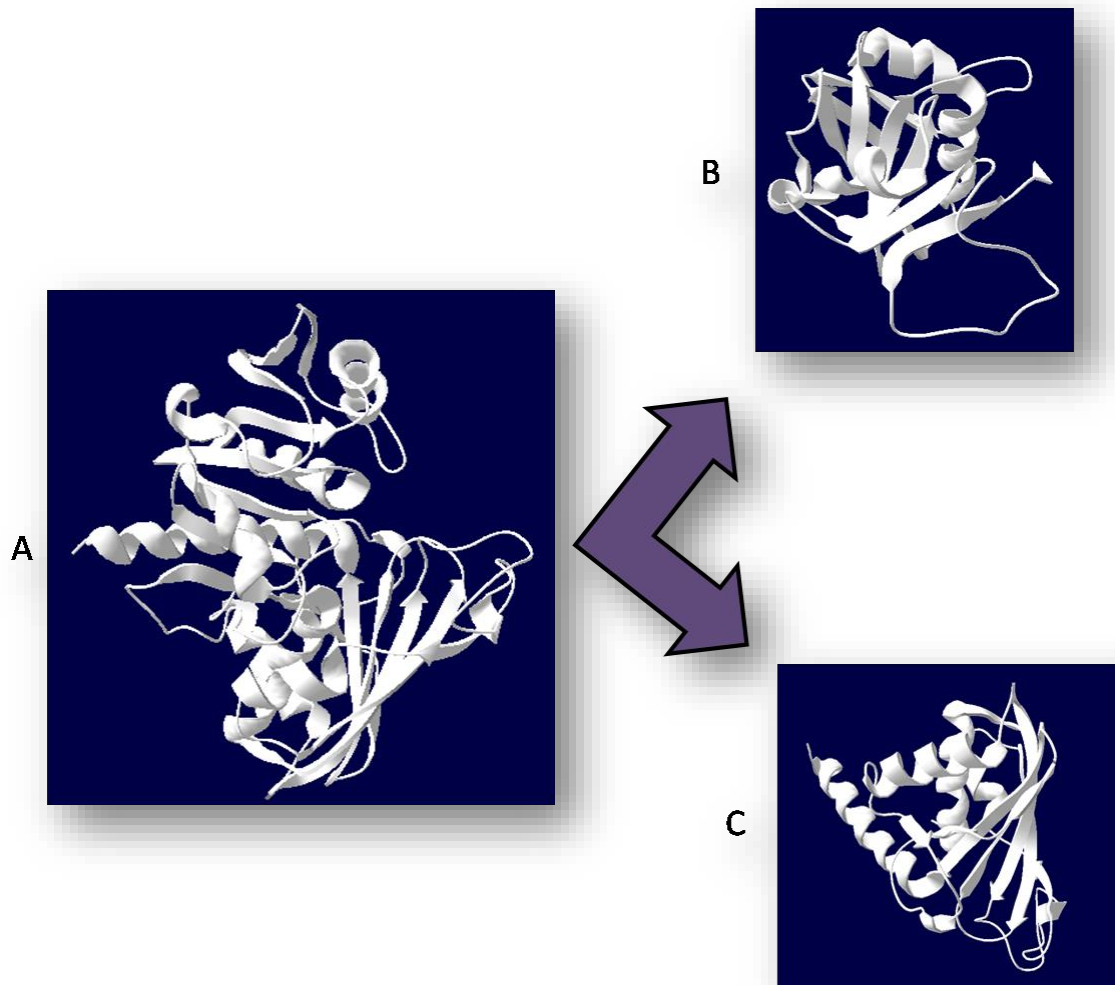


Figure 7. 13 Secondary structure of GAPDH in the fusion protein *Blastocystis* TPI-GAPDH. A) The secondary structure of GAPDH in the fusion protein *Blastocystis* TPI-GAPDH. B) The secondary structure of the nucleotide binding domain (NBD) consisting of six parallel β sheets surrounded by α helices. The 7th β sheet is anti-parallel. The NBD takes the typical Rossmann fold structure. C) The catalytic domain (CAD) consisting of anti-parallel β sheets and α helices.

GAPDH active site showed the motif signature that is found in Rossmann structure proteins; GXGXXG. The sequence of the signature motif in this specific project it was GFGRIG covering the amino acid sequences between 276 -281. The protein clearly adapts a Rossmann fold motif binding NAD not NADP as the signature motif shows ([Hanukoglu 2015](#)). The electron density obtained didn't show NAD electron density, but the presence of clear electron density of two phosphate groups in the active site of GAPDH might be an indication of the possibility of having NAD in this site as a cofactor. The specific details of the GAPDH subunit active site are shown in a cartoon representation of GAPDH

structure in Figure 7.14. In this figure a clear indication to the active site is indicated by the amino acids and the phosphate group.

The GAPDH obtained in this study was superimposed to another GAPDH whose structure was studied in the human placenta with a very high resolution of 1.75Å. This human placental GAPDH (PDB entry 1u8f) has shown a clear electron density of only three NAD molecules ([Jenkins and Tanner 2006](#)). This superimposition shows the expected geometric position of the NAD groups in the studied structure, which is expected to be washed off from the protein of study during the purification steps as indicated earlier in this chapter, Figure 7.15.

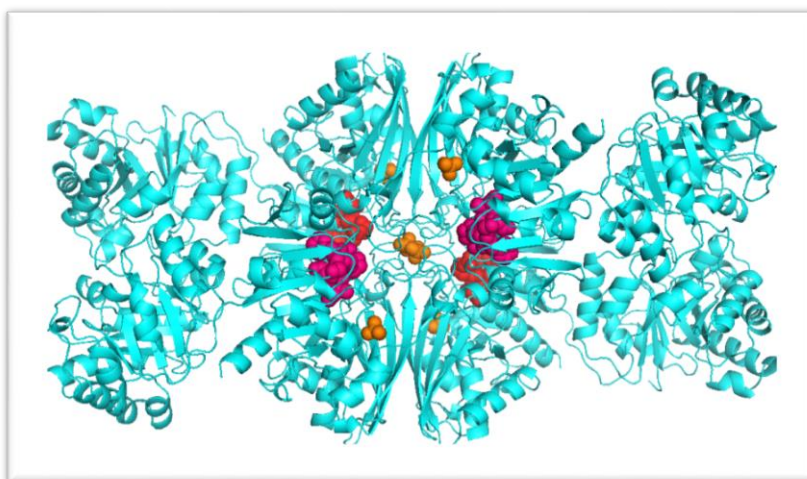


Figure 7. 14 Cartoon representation of *Blastocystis* TPI-GAPDH in cyan colour showing four NAD binding sites for the four GAPDH subunits in red and magenta. And showing six phosphate molecules in orange; one in each subunit and two in the middle pocket of the GAPDH tetramer.

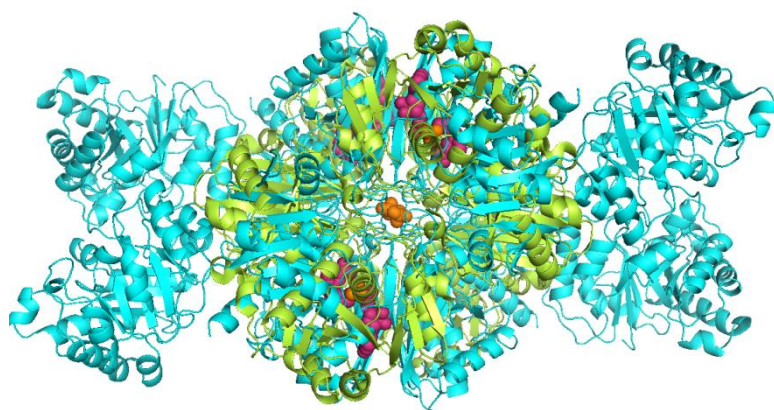


Figure 7. 15 Superimposition of the study protein *Blastocystis* TPI-GAPDH (cyan) with the model used for building the structure (human GAPDH, 1u8f) (in lime green). The figure shows the phosphate groups that belong to *Blastocystis* TPI-GAPDH in orange and three NAD molecules belonging to the human GAPDH in magenta.

7.4.3.3. TPI-GAPDH quaternary structure

From all what was previously explained, one can see that the unusual fusion protein *Blastocystis* TPI-GAPDH adapts a tetrameric structure of the protein; composed of four subunits of GAPDH and four subunits of TPI. GAPDH localizes in the centre of the whole protein molecule, forming a tetramer. TPI is presented in this study as a usual dimer, with the fusion protein having two dimers in its total structure separated by a tetrameric GAPDH in the centre. A different presentations of the whole structure of the fusion protein was shown in Figure 7.16/1, which is a surface representation of the protein superimposed with the model proteins used to solve the structure; and also transparent representation showing a cartoon fancy helices representation of the TPI and GAPDH inside the protein molecule, supplemented with stick presentation of the active sites and the phosphate molecules appeared in the electron density data, Figure 7.16/2. The fusion protein *Blastocystis* TPI-GAPDH was also superimposed to both the model proteins used to build the structure, (Figure 7.17). When superimposing GAPDH to itself, a displacement of TPI dimmers was noticed, indicating a less rigid TPI structural organization.

Analyzing the secondary structure of the fusion *Blastocystis* TPI-GAPDH protein has shown that the protein structure contains 92 alpha helices (20 in A and C and 19 in B and D) and 98 beta sheets (24 in each subunit), Figure 7.6/ 7.7.

The structure coordinates was also tested using PISA interface on the EMBL-EBI server ([Krissinel and Henrick 2007](#)). It was found that the final structure contained a total of: four monomer chains, each is composed of monomer TPI and monomer GAPDH, Figure 7.18, and ten phosphate ligand. For details of the interfacing monomers statistics see Table 7.9.

The interface analysis on PISA showed that the protein has no disulfide or covalent bonds. The protein is getting its stability through the presence of 19 salt bridges and 45 hydrogen bonds found in (interface, PISA) which are known to stabilize the total structure of the protein, especially that no di-sulphide or covalent bonds are found in this structure. A total of 4433 atom, 586 residues in each chain (contains monomer of TPI and a monomer of GAPDH).

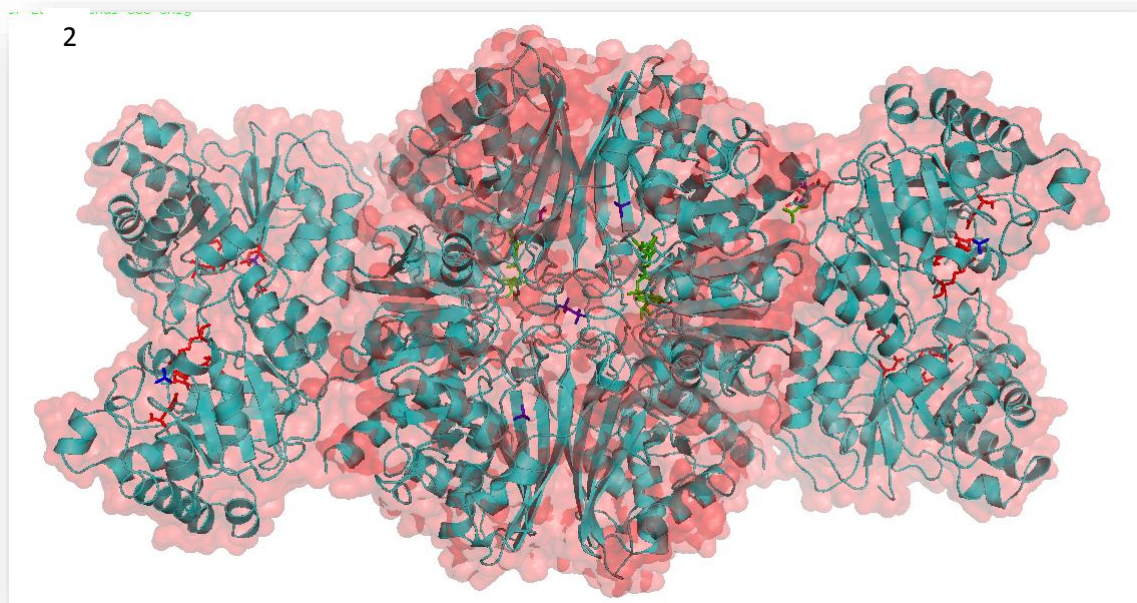
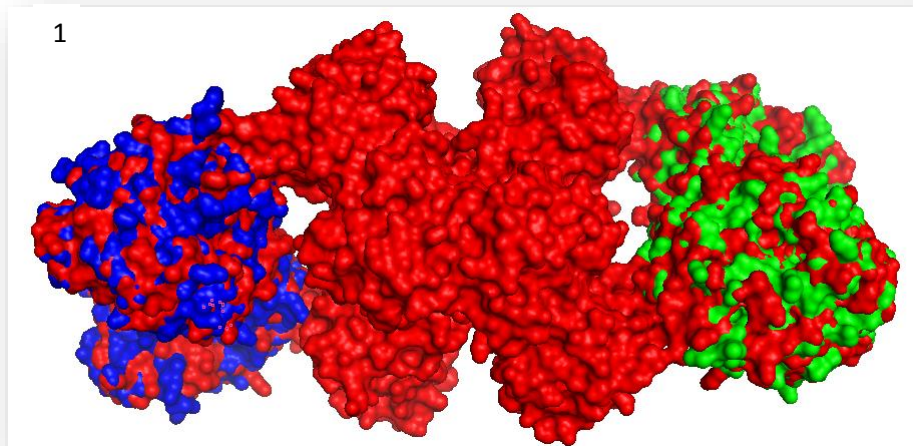


Figure 7. 16 1) Cartoon representation of the surface of the fusion protein *Blastocystis* TPI-GAPDH, superimposed with the coleopteran TPI protein (PDB entry 2i9e) with a surface quality of 1 . 2) Transparent representation of the fusion protein *Blastocystis* TPI-GAPDH showing the inside structure of the protein in cartoon fancy helices with surface transparency 0.7. TPI active sites belonging to the fusion protein structure are presented as sticks and colored red. The NAD molecules were colored in green and they belong to the human placental GAPDH protein (PDB entry 1u8f). Nine phosphate molecules presented in the electron density are colored in dark blue, and presented in stick presentation.

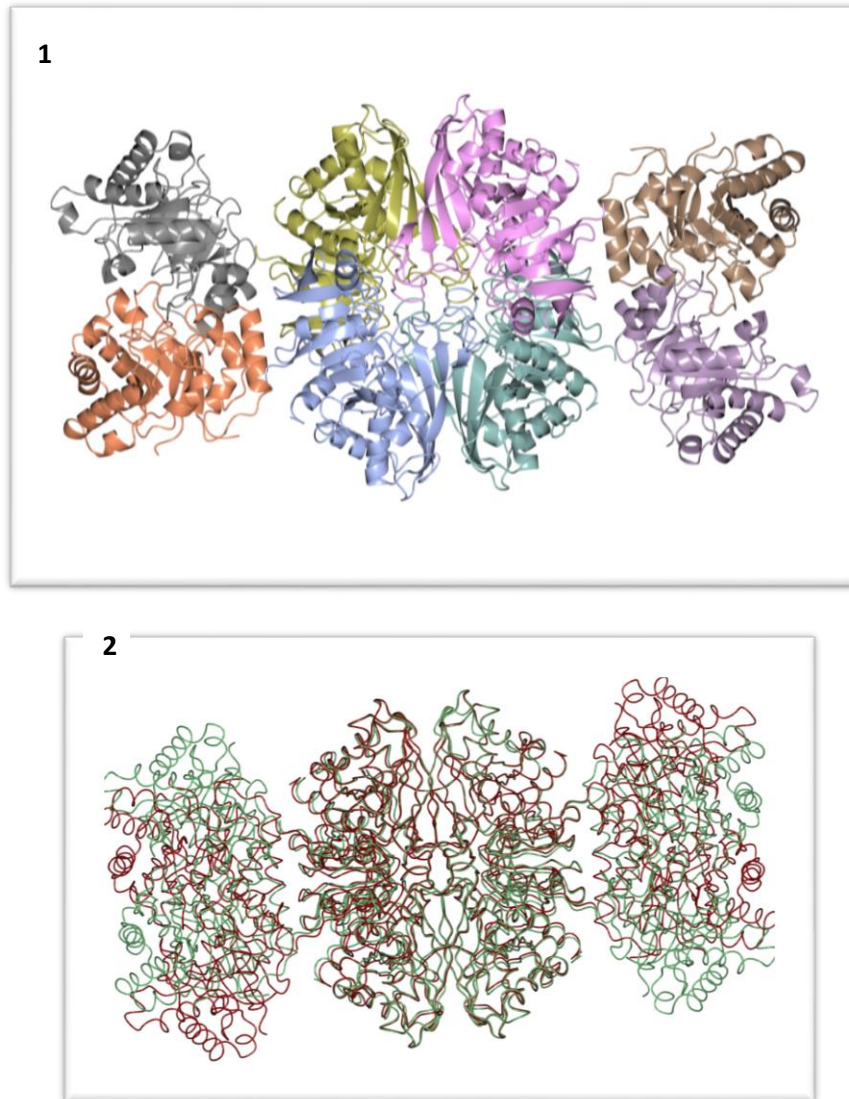


Figure 7. 17 1) Cartoon representation of the recombinant fusion protein *Blastocystis* TPI-GADH. The cartoon illustrates the quaternary structure of the protein represented by a tetrameric GAPDH in the core of the protein surrounded by two dimeric TPI proteins flanking the GAPDH subunits. 2) Stereo view of superposition of the GAPDH tetramer on itself showing a displacement of TPI dimers indication a less rigid TPI structural organization. The following homologous proteins were used as models; a Coleopteran TPI dimer from *Tenebrio molitor* (yellow mealworm beetle (PDB entry 2i9e) of 53.2% identity and the human placental GAPDH (PDB entry iu8f) of 67.5% identity. The figures were produced using Pymol.

Table 7. 9 interfacing monomers of *Blastocystis* TPI-GAPDH structure, obtained from PISA service at EMBL-EBI

Range	Class	Structure		Surface		
		N. atoms	N. residues	N. atoms	N. residues	Area, Å
A	Protein	4433	586	2402	523	25653.9
C	Protein	4433	586	2404	523	25669.0
	Average	4433	586	2403	523	25661.5
B	Protein	4422	584	2390	525	25395.2
D	Protein	4422	584	2388	525	25399.9
	Average	4422	584	2389	525	25397.6
[PO4]E:1	Ligand	5	1	5	1	189.1
[PO4]F:1	Ligand	5	1	5	1	188.8
[PO4]G:1	Ligand	5	1	5	1	187.9
[PO4]H:1	Ligand	5	1	5	1	188.2
[PO4]L:1	Ligand	5	1	5	1	188.1
[PO4]T:1	Ligand	5	1	5	1	187.0
[PO4]V:1	Ligand	5	1	5	1	188.6
[PO4]X:1	Ligand	5	1	5	1	187.9
[PO4]Y:1	Ligand	5	1	5	1	188.5
[PO4]Z:1	Ligand	5	1	5	1	188.6
	Average	5	1	5	1	188.3

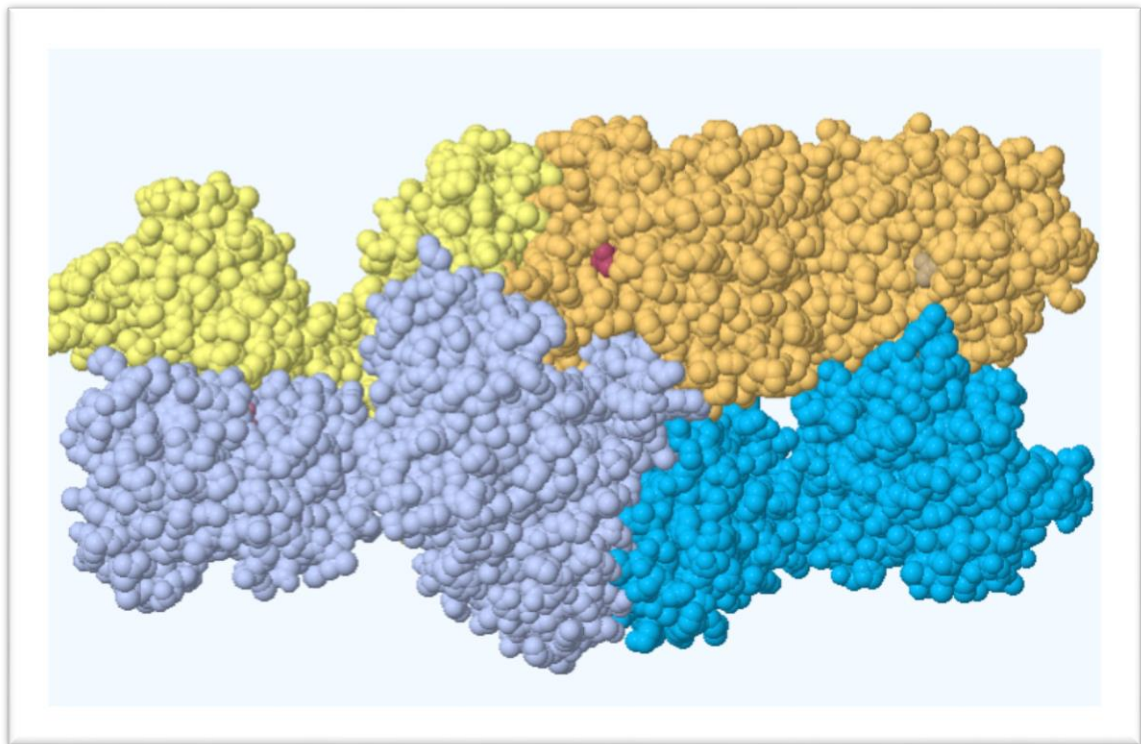


Figure 7. 18 total protein structure as interpreted by PISA, EMBL-EBI. The figure shows the protein as four monomers each presented by separate colour.

7.5. Discussion

In the aim of solving the structure of the unusual protein *Blastocystis* TPI-GAPDH, each component of this fusion protein was studied separately. For this purpose the fusion protein *Blastocystis* TPI-GAPDH, *Blastocystis* GAPDH and *Blastocystis* TPI were all expressed and tried for purification and crystallization. Expression trials of TPI failed in this attempt and more trials of expression are advised to be experimented in the future, which couldn't be done in this project due to the lack of time. Both *Blastocystis* TPI-GAPDH and *Blastocystis* GAPDH were expressed and purified to a high concentration as previously mentioned.

Blastocystis GAPDH was sent for data collection but all crystals grown in this project gave a very low resolution which prevented further study of its individual structure. To enhance the data quality it worth's considering changing the cryo-preservant used, which might enhance the data quality in future work.

The fusion protein *Blastocystis* TPI-GAPDH has diffracted to a moderately good resolution of 2.6Å. The protein structure was solved using the molecular replacement method and the protein final structure showed to be a tetrameric protein of *Blastocystis* TPI-GAPDH. The protein subunits were arranged in a way that GAPDH occupies the core of the protein and two dimers of TPI are flanking the GAPDH subunits. Searching for interfaces in the data base similar to the monomer unit (TPI-GAPDH) came back with negative result, using PISA EMBLE-EBI server. This shows that there is no protein yet in the data base of similar structure to the studied protein (TPI-GAPDH), which makes this the first study to solve the structure of the fusion *Blastocystis* TPI-GAPDH in any organism.

The final protein structure obtained after refinement and model building was lacking the first eighteen amino acids of the electron density of chains A, C, and twenty amino acid of chains B and D which represent the N terminal targeting signal.

Electron density was not high enough to show holo protein, and NAD molecules didn't appear in the final solved structure. Indicators of the presence of NAD were the presence of phosphate molecules in the center of each TPI subunit and in the core of the tetrameric GAPDH. The signature motif of GAPDH indicating to Rossmann fold was another marker of having NAD as a cofactor, as previous study of Israel Hanukoglu has shown that the sequence motif GXGXXG has another version of GXGXXA where the replacement of Glycine residue by Alanine changes the enzyme's cofactor from NAD/FAD to NADP/FADP ([Hanukoglu 2015](#)).

No significant differences were found in the structure of the protein in either TPI or GAPDH subunits, compared to the previously studied structures, in both their

catalytic residues and active sites. The only striking difference was the conformation taken by the flexible loop (loop 6), which was indicating a structure of a holo enzyme. This could either be a positive result and the protein solved in this structure is a holo enzyme, confirmed by the phosphate groups found in TPI active site and could be from the ligand (2GP). Although this is a weak possibility especially that no ligand was added in the purification steps, nor in the crystallization trials. Another possibility, is that the protein solved is an apo enzyme and then conformation of loop 6 in an almost closed position could be explained differently. As previous studies showed, closed or almost closed conformation of loop 6 in an apo enzyme could be an indicator of heterogeneity in the solution and due to an inherent flexibility of loop 6 in TPI structure. Added to the remark of unusual DLS results found in chapter 6 of this project, the possibility of heterogeneity is the most probable explanation for the flexible loop conformation. The conformational heterogeneity of TPI-GAPDH in solution might be the origin of the observed almost closed loop in the apo form of the enzyme.

Same case for GAPDH, no cofactor was in the active site, yet it was indicated to by the presence of phosphor group in the electron density. This could be washed off during purification steps; especially that DSF experiment in Chapter 5 has shown low affinity to the ligand by this enzyme.

Although no significant differences were found in the structure of the *Blastocystis* TPI-GAPDH fusion protein, yet this work is the first reported to confirm the tetrameric structure of the fusion protein that was proposed for the TPI-GAPDH fusion protein. The same fusion was found also in diatoms in Liaud et al work who suggested a similar structure lay-out of the protein through size exclusion chromatography work ([Liaud, Lichtle et al. 2000](#)). As seen in the *in silico* analysis

of the fusion protein *Blastocystis* TPI-GAPDH, it seems that this protein forms an ancestral form or evolutionary indication of its alphaprotobacterial origin, and hence it is conserving its structure and function. Yet, it would be interesting to investigate if this structure could be a potential drug target as being restricted to stramenopiles in a wide spread and few other organisms as indicated in Nakayama work ([Nakayama, Ishida et al.](#)) and confirmed in the *in silico* chapter of this work.

Most studies that seek designing species-specific drugs target any structural differences in the active site between homologous enzymes which are crucial in the organism's metabolism, for drug development. Since glycolysis is a universal pathway, it is quite conserved between species along with the enzymes of this pathway. Usually the structure is quite conserved of those proteins between species, same for the active site of those enzymes. Nonetheless, residues might vary. To circumvent the problem of conserved catalytic regions, several studies had oriented their search towards regions that are not in the active site but close to it and still make a difference in enzyme activity and stability. In *Trypanosma brucei*, *Giardia* and *Plasmodia* Cys15 or Cys14 are residues that are located in the dimer interface close to the active site of TPI. Those residues were always a target for drug development, because they are replaced in human TPI by Met instead of Cys ([Gomez-Puyou, Saavedra-Lira et al. 1995](#)). In *Blastocystis* this residue is Cys 29 found in the interface between the dimmers and forming hydrogen bonds, as shown in Table 7.9, so this can be a target as well for Blastocystosis remedies. Several studies have made mutations in this specific residue and showed a drastic reduction of TPI activity upon this change ([Gomez-Puyou, Saavedra-Lira et al. 1995](#)). Another cysteine residue that was also a target in drug delivery is Cys222, which is Cys231 in the protein of this study.

Although Cys222 is a non-conserved, non-catalytic residue, studies in *Giardia lamblia* has found that its derivatization with a small non-polar thiomethyl group of methylmethane thiosulfonate (MMTS) resulted in the reduction of TPI enzymatic activity by half. This loss of activity was restored by replacing MMTS by a negatively charged chemical, which suggest that this loss of activity is dependent on the physiochemical properties of the derivatizing agent. This might be due to the fact that mutating this residue would perturb loop 7 which is important as being a region of the enzyme that interacts with the catalytic loop 6 and hence it is important in ligand binding and enzyme activity ([Hernandez-Alcantara, Torres-Larios et al. 2013](#)). Another study by the same group found that this loss of activity is due to the decrease of *Giardia* TPI affinity to its ligand 2-phosphoglycerate (2-PG) ([Enriquez-Flores, Rodriguez-Romero et al. \(2011\)](#)). All those residues are conserved in *Blastocystis* TPI-GAPDH and can be a very good target for further inhibition studies for the sake of drug development against *Blastocystis*.

Chapter 8. Conclusion

Blastocystis and *Giardia* are the most common protists diagnosed worldwide in almost every examined stool sample when investigating the causes of diarrhoea and gastrointestinal symptoms. *Blastocystis* is responsible for up to 60% of the cases worldwide. Adding to the challenge in treating Blastocystosis is the fact that it could be found both symptomatic and asymptomatic in the detected cases which resulted in the current controversy whether *Blastocystis* actually is a pathogenic or non-pathogenic protist.

The traditional way of treating pathogens, such as targeting DNA, protein, or cell wall synthesis might result in unwanted side effects. This is especially the case for protists, being eukaryotes themselves, as they might share common features with the host, whether this is a human, an animal or a plant. Therefore, targeting a specific and unique aspect of the pathogen will be a better approach that will minimize any possible side effects.

From this point it was thought to look at unique features in *Blastocystis*. Several drugs have been approached in this way.

Blastocystis, is a widespread parasite that is found in the stool of humans, animals, birds, etc. This organism belongs to the *Stramenopiles* and one of its unique features is that it is an anaerobic parasite, yet seems to have normal mitochondria. This has driven lots of research to be conducted investigating the specifications of this organelle. Leading to the discovery that this parasite has an unusual fusion protein of two proteins that are an essential part of the glycolysis pathway. Being an essential pathway, this unusual fusion protein might be a potential drug target. If this unusual fusion protein has indeed specific characteristics, or an unusual three-dimensional structure, then disabling this

protein might be a good way of depriving this organism from essential glycolytic intermediates without affecting the host.

In *Blastocystis*, the TPI-GAPDH fusion protein is structured with TPI upstream of the GAPDH part. The C-terminus of TPI is attached to the N-terminus of GAPDH. Interestingly, it was found that the fusion protein has an N-terminal signal which resembles mitochondrial targeting signals. This is quite interesting considering the fusion protein consists of glycolytic enzymes. Glycolysis is canonically known to be a cytosolic pathway.

Very few studies have been conducted into such a fusion protein. Liaud et al studied the characteristics of a similar fusion protein in the algae *Phaeodactylum tricornutum* and concluded that this protein is a tetramer of GAPDH with two TPI dimers on both sides ([Liaud, Lichtle et al. 2000](#)). They assumed the protein maintained the same function as the glycolytic version of the stand-alone proteins. Work by an MSc student in a collaborating laboratory has suggested that this protein is forming a dimer of TPI and a dimer of GAPDH ([Evans 2014](#)), with the latter being unusual as this would be the first time GAPDH to be reported not to be a tetramer.

Three glycolytic enzymes were studied in this thesis, enolase, triosephosphate isomerase and glyceraldehydes-3-phosphate dehydrogenase. The last two were in the form of a fusion protein in the studied organism.

Therefore, I sought in this project to characterize this fusion protein in *Blastocystis* both chemically and physically, seeking any special features that differentiate it from its host. The crystallographic structure of the fusion protein have also been solved for the first time since the fusion protein TPI-GAPDH was reported. Enolase was also studied and characterized both chemically and physically.

Enolase, also called phosphopyruvate hydratase, responsible for the ninth reaction in glycolysis pathway, which is the conversion of 2-phosphoglycerate (2-PG) to phosphoenolpyruvate (PEP). In this study, enolase has shown to be a highly conserved protein in the organism of this study, *Blastocystis*, presented by its special sequence features that followed the enolase family general characteristics. It follows the general characteristic of other enolases, from its molecular weight to its general sequence. The *Blastocystis* enolase has kept the essential sequence features for enzyme activity, including residues for substrate binding and residues important for protein activity. On the other hand, *Blastocystis* enolase was found not to contain the plasminogen binding site which is thought to help in host invasion as found in some other enolases from other parasites. In phylogenetic analyses, the *Blastocystis* enolase had clearly grouped within the chromalveolates. Interestingly, the protein has a putative mitochondrial targeting signal at its amino-terminus as shown by *in silico* characterization of the protein. Biochemical characterization of *Blastocystis* enolase showed a similar conventional molecular weight comparable to other enolases. This study revealed the *Blastocystis* enolase to be a homo-dimer with a native subunit molecular weight of 78 kDa as confirmed by several experiments. However, when comparing its enzymatic activity to other studied enolases, it was suggested that the *Blastocystis* enolase has a lower affinity to its substrate, or the data had an odd point that might have played a role in the odd results of the protein kinetics. In addition to biochemical characterizations, enolase was also looked for from a biophysical side by employing biophysical methodologies, and those have shown the protein to be an active dimer protein as reported in literature. *Blastocystis* enolase was found to be stable at 66°C and that it performs best in a HEPES-

Glycine buffer. While its catalytic activity has shown that *Blastocystis* enolase is an active enzyme with a putatively lower affinity to its substrate.

Triosphosphate isomerase (TPI), catalyses the reversible interconversion of dihydroxyacetone phosphate and D-glyceraldehyde-3-phosphate. While glyceraldehydes-3-phosphate dehydrogenase (GAPDH) catalyses the sixth step of glycolysis pathway which is the reversible conversion of glyceraldehydes-3-phosphate to D-glycerate 1, 3-bisphosphate. Those two proteins were found to be fused in *Blastocystis* and the fusion enzyme was characterized chemically and physically. The fusion protein was found to be following the general characteristics studied so far for the stand-alone version of the protein. The unusual TPI-GAPDH fusion protein has a putative mitochondrial targeting signal at its amino terminus as well. Phylogenetic analysis of *Blastocystis* TPI-GAPDH fusion protein indicates that it clusters within the stramenopiles. Conserving all features essential for function of TPI and GAPDH and containing all expected protein motifs for these enzymes. The structure of the fusion protein was not solved or studied before, and this project shows the structure of the solved protein for the first time in literature. The fusion protein was found to form a tetrameric protein of tetraeric GAPDH flanked with two dimeric TPI. The fusion protein's components were also following the general characteristics of the text book enzymes.

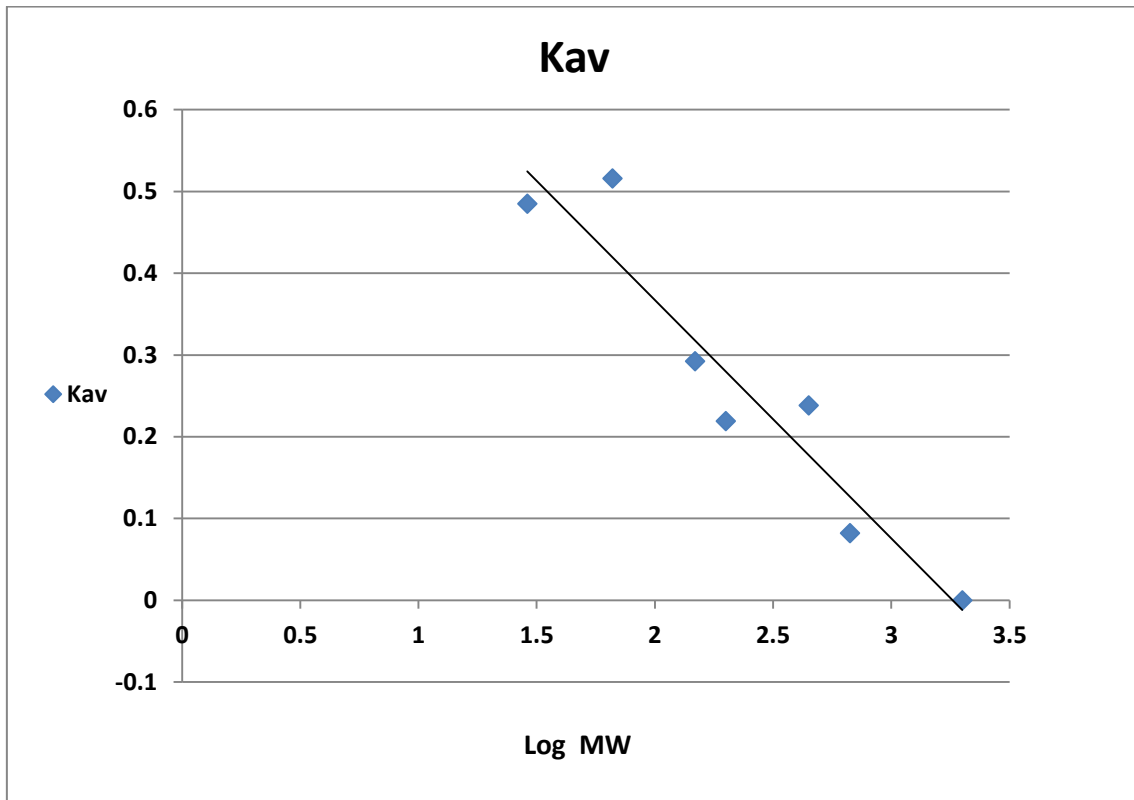
Finally, data obtained in this thesis suggests that at least part of glycolysis in *Blastocystis* takes place in the mitochondrion.

Appendices

Appendix 1 IMDM formula as described on PAA company.

Inorganic salts mg/m	Amino acids mg/ml	Vitamins mg/ml	Other components mg/ml
Calcium Chloride.2 H ₂ O 218.97	L-Alanine 25.00	D-Biotin 0.018	D-Glucose 4500.00
Magnesium Sulphate 97.67	L-Arginine. HCl 84.00	Choline Chloride 4.00	HEPES 5958.00
Potassium Chloride 330.00	L-Asparagine . H ₂ O 28.40	Folic Acid 4.00	Phenol Red. 15.00
Potassium Nitrate 0.067	L-Aspartic Acid 30.00	<i>myo</i> -Inositol 7.20	Sodium pyruvate 110.00
Sodium hydrogen carbonate (only in liquid solution) 3024.00	L-Cystine 70.00	Niacinamide 4.00	
Sodium Chloride 4505.00	L-Glutamic Acid 75.00	D-calcium pantothenate (Vitamin B5) 4.00	
Sodium Phosphate Monobasic.2H ₂ O 108.68	L-Glutamine 584.00	Pyridoxal. HCl 4.00	
Sodium Selenite. 5H ₂ O 0.017	Glycine 30.00	Riboflavin 0.40	
	L-Histidine. HCl.H ₂ O 42.00	Thiamine. HCl 4.00	
	L-Isoleucine 105.00	Vitamin B12 0.013	
	L-Leucine 105.00		
	L-Lysine. HCl 146.00		
	L-Methionine 30.00		
	L-Phenylalanine 66.00		
	L-Proline 40.00		
	L-Serine 42.00		
	L-Threonine 95.00		
	L-Tryptophan 16.00		
	L-Tyrosine. 2Na 89.47		
	L-Valine 94.00		

Appendix 2 Gel filtration standard curve



Appendix 3 Crystallization optimization conditions for *Blastocystis* TPI-GAPDH protein.

Project: C:\DOCUMENTS AND SETTINGS\ORYX6\DESKTOP\SHEERA\TPIGAPDH OPTIMIZATION 130614.XDB
 Experiment: Experiment 1

Design : 5 Channel, Droplet Only
 Plate : Vapor_Batch
 Printed : 10:07 26-Jun-2014
 Executed : 16:09 13-Jun-2014

Display : Concentration Mode
 SubWell : DEFAULT
 Modified : 16:07 13-Jun-2014
 Saved : 10:05 26-Jun-2014

33.00 mg/ml Protein	pH: 7.55	6A \$	14.143	6B \$	14.143	6C \$	14.143	6D \$	14.143	6E \$	14.143	6F \$	14.143	6G \$	14.143	6H \$	14.143
60.00 % PEG 3350			20.000		17.778		15.873		13.651		11.429		9.206		7.302		5.079
3.00 M Phosphate	pH: 9.74		0.206		0.190		0.159		0.143		0.079		0.048		0.032		0.000
1.00 M Phosphate	pH: 7.20		0.000		0.032		0.048		0.079		0.111		0.143		0.190		0.206
Nett Solution pH			7.68		7.29		7.26		7.24		7.23		7.22		7.22		7.21
Total Buffer Concentration			0.21		0.23		0.21		0.23		0.20		0.20		0.23		0.21
Volume			1.00		1.00		1.00		1.00		1.00		1.00		1.00		1.00
33.00 mg/ml Protein	pH: 7.55	5A \$	19.381	5B \$	19.381	5C \$	19.381	5D \$	18.333	5E \$	17.460	5F \$	16.587	5G \$	16.413	5H \$	15.540
60.00 % PEG 3350			9.841		9.841		9.841		9.206		8.889		8.254		8.254		7.619
3.00 M Phosphate	pH: 9.74		0.492		0.397		0.365		0.270		0.175		0.095		0.063		0.000
1.00 M Phosphate	pH: 7.20		0.000		0.079		0.127		0.196		0.254		0.312		0.333		0.392
Nett Solution pH			7.80		7.26		7.24		7.22		7.22		7.21		7.21		7.21
Total Buffer Concentration			0.50		0.49		0.50		0.48		0.44		0.42		0.41		0.40
Volume			1.00		1.00		1.00		1.00		1.00		1.00		1.00		1.00
33.00 mg/ml Protein	pH: 7.55	4A \$	19.381	4B \$	19.381	4C \$	19.381	4D \$	18.333	4E \$	17.460	4F \$	16.587	4G \$	16.413	4H \$	15.540
60.00 % PEG 3350			9.841		9.841		9.841		9.206		8.889		8.254		8.254		7.619
3.00 M Phosphate	pH: 9.74		0.492		0.397		0.365		0.270		0.175		0.095		0.063		0.000
1.00 M Phosphate	pH: 7.20		0.000		0.079		0.127		0.196		0.254		0.312		0.333		0.392
Nett Solution pH			7.80		7.26		7.24		7.22		7.22		7.21		7.21		7.21
Total Buffer Concentration			0.50		0.49		0.50		0.48		0.44		0.42		0.41		0.40
Volume			1.00		1.00		1.00		1.00		1.00		1.00		1.00		1.00
33.00 mg/ml Protein	pH: 7.55	3A \$	19.381	3B \$	19.381	3C \$	19.381	3D \$	18.333	3E \$	17.460	3F \$	16.587	3G \$	16.413	3H \$	15.540
60.00 % PEG 3350			9.841		9.841		9.841		9.206		8.889		8.254		8.254		7.619
3.00 M Phosphate	pH: 9.74		0.492		0.397		0.365		0.270		0.175		0.095		0.063		0.000
1.00 M Phosphate	pH: 7.20		0.000		0.079		0.127		0.196		0.254		0.312		0.333		0.392
Nett Solution pH			7.80		7.26		7.24		7.22		7.22		7.21		7.21		7.21
Total Buffer Concentration			0.50		0.49		0.50		0.48		0.44		0.42		0.41		0.40
Volume			1.00		1.00		1.00		1.00		1.00		1.00		1.00		1.00
33.00 mg/ml Protein	pH: 7.55	2A \$	19.032	2B \$	19.556	2C \$	19.905	2D \$	19.905	2E \$	19.905	2F \$	19.905	2G \$	19.905	2H \$	19.905
60.00 % PEG 3350			19.048		17.460		15.873		13.651		11.429		9.206		7.302		5.079
3.00 M Phosphate	pH: 9.74		0.206		0.175		0.159		0.143		0.079		0.048		0.032		0.000
1.00 M Phosphate	pH: 7.20		0.000		0.016		0.048		0.079		0.143		0.159		0.190		0.206
Nett Solution pH			7.65		7.37		7.28		7.25		7.23		7.22		7.22		7.22
Total Buffer Concentration			0.22		0.20		0.22		0.23		0.23		0.22		0.23		0.22
Volume			1.00		1.00		1.00		1.00		1.00		1.00		1.00		1.00
33.00 mg/ml Protein	pH: 7.55	1A \$	5.063	1B \$	5.063	1C \$	5.063	1D \$	5.063	1E \$	5.063	1F \$	5.063	1G \$	5.063	1H \$	5.063
60.00 % PEG 3350			20.000		17.778		15.873		13.651		11.429		9.206		7.302		5.079
3.00 M Phosphate	pH: 9.74		0.206		0.190		0.159		0.143		0.079		0.048		0.032		0.000
1.00 M Phosphate	pH: 7.20		0.000		0.032		0.048		0.079		0.143		0.159		0.190		0.206
Nett Solution pH			7.99		7.25		7.23		7.22		7.21		7.21		7.21		7.20
Total Buffer Concentration			0.21		0.23		0.21		0.23		0.23		0.21		0.23		0.21
Volume			1.00		1.00		1.00		1.00		1.00		1.00		1.00		1.00

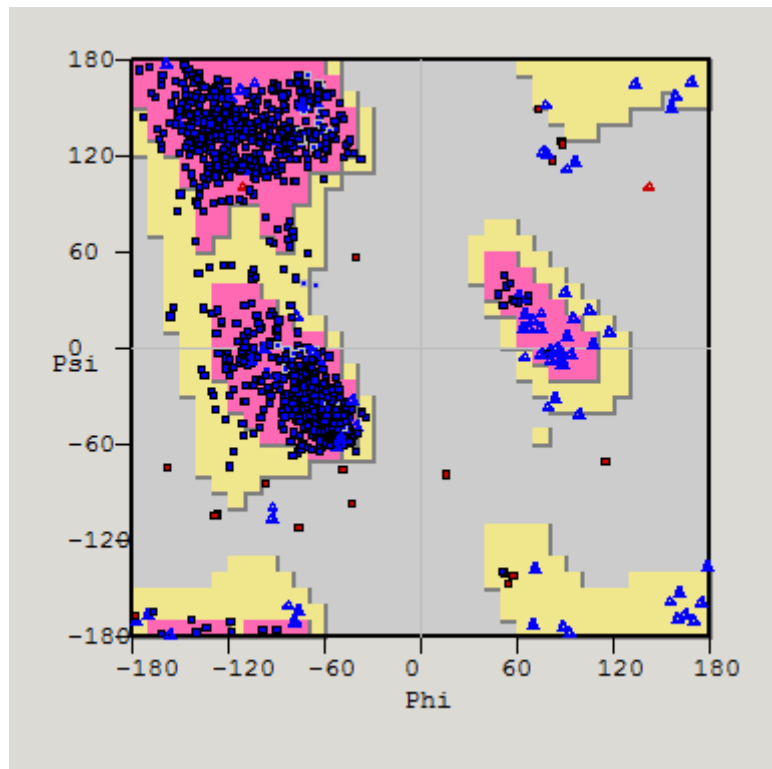
Project: C:\DOCUMENTS AND SETTINGS\ORYX6\DESKTOP\SHEERA\TPIGAPDH OPTIMIZATION 130614.XDB
Experiment: Experiment 1

Design : 5 Channel, Droplet Only
 Plate : Vapor_Batch
 Printed : 10:07 26-Jun-2014
 Executed : 16:09 13-Jun-2014

Display : Concentration Mode
 SubWell : DEFAULT
 Modified : 16:07 13-Jun-2014
 Saved : 10:05 26-Jun-2014

33.00 mg/ml Protein pH: 7.55	12A \$	12B \$	12C \$	12D \$	12E \$	12F \$	12G \$	12H \$
60.00 % PEG 3350	19.032	20.079	20.079	20.079	20.079	20.079	20.079	20.079
3.00 M Phosphate pH: 9.74	19.048	17.778	15.873	13.651	11.429	9.206	7.302	5.079
1.00 M Phosphate pH: 7.20	0.206	0.190	0.159	0.143	0.111	0.079	0.048	0.032
Nett Solution pH	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Total Buffer Concentration	7.65	7.63	7.62	7.61	7.60	7.58	7.57	7.56
Volume	0.22	0.20	0.17	0.16	0.12	0.09	0.06	0.04
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
33.00 mg/ml Protein pH: 7.55	11A \$	11B \$	11C \$	11D \$	11E \$	11F \$	11G \$	11H \$
60.00 % PEG 3350	19.032	20.079	20.079	20.079	20.079	20.079	20.079	20.079
3.00 M Phosphate pH: 9.74	19.048	17.778	15.873	13.651	11.429	9.206	7.302	5.079
1.00 M Phosphate pH: 7.20	0.206	0.190	0.159	0.143	0.111	0.079	0.048	0.032
Nett Solution pH	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Total Buffer Concentration	7.65	7.63	7.62	7.61	7.60	7.58	7.57	7.56
Volume	0.22	0.20	0.17	0.16	0.12	0.09	0.06	0.04
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
33.00 mg/ml Protein pH: 7.55	10A \$	10B \$	10C \$	10D \$	10E \$	10F \$	10G \$	10H \$
60.00 % PEG 3350	19.032	20.079	20.079	20.079	20.079	20.079	20.079	20.079
3.00 M Phosphate pH: 9.74	19.048	17.778	15.873	13.651	11.429	9.206	7.302	5.079
1.00 M Phosphate pH: 7.20	0.206	0.190	0.159	0.143	0.111	0.079	0.048	0.032
Nett Solution pH	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Total Buffer Concentration	7.65	7.63	7.62	7.61	7.60	7.58	7.57	7.56
Volume	0.22	0.20	0.17	0.16	0.12	0.09	0.06	0.04
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
33.00 mg/ml Protein pH: 7.55	9A \$	9B \$	9C \$	9D	9E	9F	9G	9H
60.00 % PEG 3350	19.032	12.571	5.063					
3.00 M Phosphate pH: 9.74	19.048	20.000	20.000					
1.00 M Phosphate pH: 7.20	0.206	0.206	0.206					
Nett Solution pH	0.000	0.000	0.000					
Total Buffer Concentration	7.65	7.70	7.99					
Volume	0.22	0.21	0.21					
	1.00	1.00	1.00					
33.00 mg/ml Protein pH: 7.55	8A \$	8B \$	8C \$	8D \$	8E \$	8F \$	8G \$	8H \$
60.00 % PEG 3350	19.032	19.556	19.905	19.905	19.905	19.905	19.905	19.905
3.00 M Phosphate pH: 9.74	19.048	17.460	15.873	13.651	11.429	9.206	7.302	5.079
1.00 M Phosphate pH: 7.20	0.206	0.175	0.159	0.143	0.079	0.048	0.032	0.000
Nett Solution pH	0.000	0.016	0.048	0.079	0.111	0.143	0.190	0.206
Total Buffer Concentration	7.65	7.37	7.28	7.25	7.24	7.23	7.22	7.22
Volume	0.22	0.20	0.22	0.23	0.20	0.20	0.23	0.22
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
33.00 mg/ml Protein pH: 7.55	7A \$	7B \$	7C \$	7D \$	7E \$	7F \$	7G \$	7H \$
60.00 % PEG 3350	17.111	17.111	17.111	17.111	17.111	17.111	17.111	17.111
3.00 M Phosphate pH: 9.74	20.000	17.778	15.873	13.651	11.429	9.206	7.302	5.079
1.00 M Phosphate pH: 7.20	0.206	0.190	0.159	0.143	0.079	0.048	0.032	0.000
Nett Solution pH	0.000	0.032	0.048	0.079	0.111	0.143	0.190	0.206
Total Buffer Concentration	7.66	7.30	7.27	7.25	7.23	7.22	7.22	7.22
Volume	0.22	0.23	0.22	0.23	0.20	0.20	0.23	0.22
	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Appendix 4 Ramachandran plot of the recombinant *Blastocystis* TPI-GAPDH protein



In Preferred Regions: 2166 (93.04%)
In Allowed Regions: 122 (5.24%)
Outliers: 40 (1.72%)

Bibliography:

Agarwal, V., S. Hammerschmidt, S. Malm, S. Bergmann, K. Riesbeck and A. M. Blom (2012). "Enolase of *Streptococcus pneumoniae* Binds Human Complement Inhibitor C4b-Binding Protein and Contributes to Complement Evasion." Journal of Immunology **189**(7): 3575-3584.

Aparicio, R., S. T. Ferreira and I. Polikarpov (2003). "Closed conformation of the active site loop of rabbit muscle triosephosphate isomerase in the absence of substrate: evidence of conformational heterogeneity." Journal of Molecular Biology **334**(5): 1023-1041.

Asherie, N. (2004). "Protein crystallization and phase diagrams." Methods **34**(3): 266-272.

Asuncion, M., W. Blankenfeldt, J. N. Barlow, D. Gani and J. H. Naismith (2002). "The structure of 3-methylaspartase from *Clostridium tetanomorphum* functions via the common enolase chemical step." Journal of Biological Chemistry **277**(10): 8306-8311.

Ayres, C. A., N. Schormann, O. Senkovich, A. Fry, S. Banerjee, G. C. Ulett and D. Chattopadhyay (2014). "Structure of *Streptococcus agalactiae* glyceraldehyde-3-phosphate dehydrogenase holoenzyme reveals a novel surface." Acta Crystallographica Section F
STRUCTURAL BIOLOGY COMMUNICATIONS **70**(Pt 10): 1333-1339.

Azapagasi, E., F. Alehan, S. Saygi, N. Bayraktar and A. C. Yazici (2012). "Serum concentrations of neuron-specific enolase in pediatric migraine." Turkish Journal of Pediatrics **54**(2): 150-155.

Bae, S., H. Kim, N. Lee, C. Won, H. R. Kim, Y. I. Hwang, Y. W. Song, J. S. Kang and W. J. Lee (2012). "alpha-Enolase Expressed on the Surfaces of Monocytes and Macrophages Induces Robust Synovial Inflammation in Rheumatoid Arthritis." Journal of Immunology **189**(1): 365-372.

Baker, B. Y., W. Shi, B. Wang and K. Palczewski (2014). "High-resolution crystal structures of the photoreceptor glyceraldehyde 3-phosphate dehydrogenase (GAPDH) with three and four-bound NAD molecules." Protein Science **23**(11): 1629-1639.

Baldo, E. T., V. Y. Belizario, W. U. De Leon, H. H. Kong and D. I. Chung (2004). "Infection status of intestinal parasites in children living in residential institutions in Metro Manila, the Philippines." Korean Journal of Parasitology **42**(2): 67-70.

Bannai, H., Y. Tamada, O. Maruyama, K. Nakai and S. Miyano (2002). "Extensive feature detection of N-terminal protein sorting signals." Bioinformatics **18**(2): 298-305.

Bao, S., D. Chen, S. Yu, H. Chen, L. Tan, M. Hu, X. Qiu, C. Song and C. Ding (2015). "Characterization of triosephosphate isomerase from *Mycoplasma gallisepticum*." FEMS Microbiology Letters **362**(17): fnv140.

Bellini, M., D. Gambaccini, P. Usai-Satta, N. De Bortoli, L. Bertani, S. Marchi and C. Stasi (2015). "Irritable bowel syndrome and chronic constipation: Fact and fiction." World Journal of Gastroenterology **21**(40): 11362-11370.

Bergmeyer, H. U. (1965). Methods of enzymatic analysis.

Bjerkehegen, B., S. D. Fossa, N. Raabe, R. Holm and J. M. Nesland (1994). "Transitional cell carcinoma of the renal pelvis and its expression of p53 protein, c-erbB-2 protein, neuron-specific enolase, Phe 5, chromogranin, laminin and collagen type IV." European Urology **26**(4): 334-339.

Bradley, P. J., C. J. Lahti, E. Plumper and P. J. Johnson (1997). "Targeting and translocation of proteins into the hydrogenosome of the protist *Trichomonas*: similarities with mitochondrial protein import." EMBO Journal **16**(12): 3484-3493.

Brewer, J. M. and K. M. Collins (1980). "Studies of the role of catalytic and conformational metals in producing enzymatic activity in yeast enolase." Journal of Inorganic Biochemistry **13**(2): 151-164.

Bruna, S. P., M. G. Carolina, S. Laura, M. C. Keila, P. K. Flavio and S. G. Clarissa (2012). "Altered levels of Neuron Specific Enolase (NSE) in patients with bipolar disorder and schizophrenia." Bipolar Disorders **14**: 64-65.

Buret, A. G. (2007). "Mechanisms of epithelial dysfunction in giardiasis." Gut **56**(3): 316-317.

Burri, L. and P. J. Keeling (2007). "Protein targeting in parasites with cryptic mitochondria." International Journal for Parasitology **37**(3-4): 265-272.

Carabarin-Lima, A., O. Rodriguez-Morales, M. C. Gonzalez-Vazquez, L. Baylon-Pacheco, P. A. Reyes, M. Arce-Fonseca and J. L. Rosales-Encina (2014). "In silico approach for the identification of immunological properties of enolase from *Trypanosoma cruzi* and its possible usefulness as vaccine in Chagas disease." Parasitology Research **113**(3): 1029-1039.

Carbajal, J. A., L. DelCastillo, M. D. Lanuza, J. Villar and R. Borrás (1997). "Karyotypic diversity among *Blastocystis hominis* isolates." International Journal for Parasitology **27**(8): 941-945.

Casero, R. D., F. Mongi, A. Sanchez and J. D. Ramirez (2015). "Blastocystis and urticaria: Examination of subtypes and morphotypes in an unusual clinical manifestation." Acta Tropica **148**: 156-161.

Castillo-Romero, A., B. J. Davids, T. Lauwaet and F. D. Gillin "Importance of enolase in *Giardia lamblia* differentiation." Molecular and Biochemical Parasitology **184**(2): 122-125.

Castillo-Romero, A., B. J. Davids, T. Lauwaet and F. D. Gillin (2012). "Importance of enolase in *Giardia lamblia* differentiation." Molecular and Biochemical Parasitology **184**(2): 122-125.

Castresana, J. (2000). "Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis." Molecular Biology and Evolution **17**(4): 540-552.

Cavalier-Smith, T. (1987). "The simultaneous symbiotic origin of mitochondria, chloroplasts, and microbodies." Annals of the New York Academy of Sciences **503**: 55-71.

Chavez-Munguia, B., N. Segovia-Gamboa, L. Salazar-Villatoro, M. Omana-Molina, M. Espinosa-Cantellano and A. Martinez-Palomo (2011). "Naegleria fowleri: Enolase is Expressed During Cyst Differentiation." Journal of Eukaryotic Microbiology **58**(5): 463-468.

Chayen, N. E. (2004). "Turning protein crystallisation from an art into a science." Current Opinion in Structural Biology **14**(5): 577-583.

Chen, N., Z. G. Yuan, M. J. Xu, D. H. Zhou, X. X. Zhang, Y. Z. Zhang, X. W. Wang, C. Yan, R. Q. Lin and X. Q. Zhu (2012). "Ascaris suum enolase is a potential vaccine candidate against ascariasis." Vaccine **30**(23): 3478-3482.

Chevenet, F., C. Brun, A. L. Banuls, B. Jacq and R. Christen (2006). "TreeDyn: towards dynamic graphics and annotations for analyses of trees." BMC Bioinformatics **7**: 439.

Clark, C. G. and L. S. Diamond (2002). "Methods for cultivation of luminal parasitic protists of clinical importance." Clinical Microbiology Reviews **15**(3): 329-+.

Clark, C. G., V. Perez-Brocal and R. Shahar-Golan (2010). "A Linear Molecule with Two Large Inverted Repeats: The Mitochondrial Genome of the Stramenopile *Proteromonas lacertae*." Genome Biology and Evolution **2**: 257-266.

Clark, C. G., C. R. Stensvold, G. K. Suresh, K. S. W. Tan, R. C. A. Thompson, R. J. Traub, E. Viscogliosi and H. Yoshikawa (2007). "Terminology for *Blastocystis* subtypes - a consensus." Trends in Parasitology **23**(3): 93-96.

Clark, C. G., M. van der Giezen, M. A. Alfellani and C. R. Stensvold (2013). "Recent developments in *Blastocystis* research." Advances in Parasitology **82**: 1-32.

Claros, M. G. (1995). "MitoProt, a Macintosh application for studying mitochondrial proteins." Computer Applications in the Biosciences **11**(4): 441-447.

Cole, J. L., J. W. Lary, P. M. T and T. M. Laue (2008). "Analytical ultracentrifugation: sedimentation velocity and sedimentation equilibrium." Methods in Cell Biology **84**: 143-179.

Colell A, G. D., Ricci JE. (2009). "Novel roles for GAPDH in cell death and carcinogenesis." Cell Death and Differentiation **16**: 1573-1581.

Coley, A. F., H. C. Dodson, M. T. Morris and J. C. Morris (2011). "Glycolysis in the african trypanosome: targeting enzymes and their subcellular compartments for therapeutic development." Molecular Biology International **2011**: 123702.

Collaborative Computational Project, N. (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallographica Section D Biological Crystallography **50**(Pt 5): 760-763.

Collingridge, P. W., R. W. Brown and M. L. Ginger (2010). "Moonlighting enzymes in parasitic protozoa." Parasitology **137**(9): 1467-1475.

Cowtan, K. (2010). "Recent developments in classical density modification." Acta Crystallographica Section D Biological Crystallography **66**(Pt 4): 470-478.

Coyle, C. M., J. Varughese, L. M. Weiss and H. B. Tanowitz (2012). "Blastocystis: to treat or not to treat." Clinical Infectious Diseases **54**(1): 105-110.

Dannelly, H. K., B. Duclos, A. J. Cozzone and H. C. Reeves (1989). "Phosphorylation of Escherichia coli enolase." Biochimie **71**(9-10): 1095-1100.

Dannelly, H. K. and H. C. Reeves (1988). "Purification and Characterization of Enolase from Escherichia-Coli." Current Microbiology **17**(5): 265-268.

Dauter, Z. (1999). "Data-collection strategies." Acta Crystallographica Section D Biological Crystallography **55**(Pt 10): 1703-1717.

Davis, P. H., J. Schulze and S. L. Stanley, Jr. (2007). "Transcriptomic comparison of two Entamoeba histolytica strains with defined virulence phenotypes identifies new virulence factor candidates and key differences in the expression patterns of cysteine proteases, lectin light chains, and calmodulin." Molecular and Biochemical Parasitology **151**(1): 118-128.

Dereeper, A., S. Audic, J. M. Claverie and G. Blanc (2010). "BLAST-EXPLORER helps you building datasets for phylogenetic analysis." BMC Evolutionary Biology **10**: 8.

Dereeper, A., V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J. F. Dufayard, S. Guindon, V. Lefort, M. Lescot, J. M. Claverie and O. Gascuel (2008). "Phylogeny.fr: robust phylogenetic analysis for the non-specialist." Nucleic Acids Research **36**(Web Server issue): W465-469.

Dunn, L. A., P. F. L. Boreham and D. J. Stenzel (1989). "Ultrastructural Variation of Blastocystis-Hominis Stocks in Culture." International Journal for Parasitology **19**(1): 43-56.

Edgar, R. C. (2004). "MUSCLE: a multiple sequence alignment method with reduced time and space complexity." BMC Bioinformatics **5**: 113.

Emanuelsson, O., S. Brunak, G. von Heijne and H. Nielsen (2007). "Locating proteins in the cell using TargetP, SignalP and related tools." Nature Protocols **2**(4): 953-971.

Emanuelsson, O., H. Nielsen, S. Brunak and G. von Heijne (2000). "Predicting subcellular localization of proteins based on their N-terminal amino acid sequence." Journal of Molecular Biology **300**(4): 1005-1016.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallographica Section D Biological Crystallography **60**(Pt 12 Pt 1): 2126-2132.

Enriquez-Flores, S., A. Rodriguez-Romero, G. Hernandez-Alcantara, J. Oria-Hernandez, P. Gutierrez-Castrellon, G. Perez-Hernandez, I. de la Mora-de la Mora, A. Castillo-Villanueva, I. Garcia-Torres, S. T. Mendez, S. Gomez-Manzo, A. Torres-Arroyo, G. Lopez-Velazquez and H. Reyes-Vivas (2011). "Determining the molecular mechanism of inactivation by chemical modification of triosephosphate isomerase from the human parasite *Giardia lamblia*: a study for antiparasitic drug design." Proteins **79**(9): 2711-2724.

Espinosa-Cantu, A., D. Ascencio, F. Barona-Gomez and A. DeLuna (2015). "Gene duplication and the evolution of moonlighting proteins." Frontiers in Genetics **6**: 227.

Evans, K. (2014). Characterisation of the novel TPI-GAPDH fusion enzyme in *Blastocystis hominis* MSc by research thesis A thesis submitted to the University of Bedfordshire, in partial fulfillment of the requirements for the degree of Master of Science by research., University of Bedfordshire.

Evans, P. (2006). "Scaling and assessment of data quality." Acta Crystallographica Section D Biological Crystallography **62**(Pt 1): 72-82.

Evans, P. and A. McCoy (2008). "An introduction to molecular replacement." Acta Crystallographica Section D Biological Crystallography **64**(Pt 1): 1-10.

Fiala, G. J., W. W. Schamel and B. Blumenthal (2011). "Blue native polyacrylamide gel electrophoresis (BN-PAGE) for analysis of multiprotein complexes from cellular lysates." Journal of Visualized Experiments (JoVE)(48).

Fletcher, S. M., D. Stark, J. Harkness and J. Ellis (2012). "Enteric protozoa in the developed world: a public health perspective." Clinical Microbiology Reviews **25**(3): 420-449.

Florian Muller, Elisa Aquilanti & Ronald DePinho (2012). "In vitro enzymatic activity assay for ENOLASE in mammalian cells in culture." Protocol Exchange.

Furuya, H. and R. Ikeda (2011). "Interaction of triosephosphate isomerase from *Staphylococcus aureus* with plasminogen." Microbiology and Immunology **55**(12): 855-862.

Gan, W., G. Zhao, H. Xu, W. Wu, W. Du, J. Huang, X. Yu and X. Hu (2010). "Reverse vaccinology approach identify an *Echinococcus granulosus* tegumental membrane protein enolase as vaccine candidate." Parasitology Research **106**(4): 873-882.

Ganea, E. and J. J. Harding (2000). "alpha-crystallin assists the renaturation of glyceraldehyde-3-phosphate dehydrogenase." Biochemical Journal **345**: 467-472.

Gavel, Y. and G. von Heijne (1990). "Cleavage-site motifs in mitochondrial targeting peptides." Protein Engineering Design and Selection **4**(1): 33-37.

Gayathri, P., M. Banerjee, A. Vijayalakshmi, S. Azeez, H. Balaram, P. Balaram and M. R. Murthy (2007). "Structure of triosephosphate isomerase (TIM) from *Methanocaldococcus jannaschii*." Acta Crystallographica Section D Biological Crystallography **63**(Pt 2): 206-220.

Gekko, K. and S. N. Timasheff (1981). "Mechanism of protein stabilization by glycerol: preferential hydration in glycerol-water mixtures." Biochemistry **20**(16): 4667-4676.

Gerlt, J. A., P. C. Babbitt, M. P. Jacobson and S. C. Almo (2012). "Divergent Evolution in Enolase Superfamily: Strategies for Assigning Functions." Journal of Biological Chemistry **287**(1): 29-34.

Gomez-Puyou, A., E. Saavedra-Lira, I. Becker, R. A. Zubillaga, A. Rojo-Dominguez and R. Perez-Montfort (1995). "Using evolutionary changes to achieve species-specific inhibition of enzyme action--studies with triosephosphate isomerase." Chemistry & Biology **2**(12): 847-855.

Guo, C., S. Liu and M. Z. Sun (2013). "Novel insight into the role of GAPDH playing in tumor." Clinical and Translational Oncology **15**(3): 167-172.

Hanukoglu, I. (2015). "Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites." Biochemistry and Molecular Biology Education **43**(3): 206-209.

Hausmann, A., D. J. Aguilar Netz, J. Balk, A. J. Pierik, U. Muhlenhoff and R. Lill (2005). "The eukaryotic P loop NTPase Nbp35: an essential component of the cytosolic and nuclear iron-sulfur protein assembly machinery." Proceedings of the National Academy of Sciences of the United States of America **102**(9): 3266-3271.

Hernandez-Alcantara, G., A. Torres-Larios, S. Enriquez-Flores, I. Garcia-Torres, A. Castillo-Villanueva, S. T. Mendez, I. de la Mora-de la Mora, S. Gomez-Manzo, A. Torres-Arroyo, G. Lopez-

Velazquez, H. Reyes-Vivas and J. Oria-Hernandez (2013). "Structural and functional perturbation of Giardia lamblia triosephosphate isomerase by modification of a non-catalytic, non-conserved region." PLoS One **8**(7): e69031.

Hernández, S. (2012). "Do Moonlighting Proteins Belong to the Intrinsically Disordered Protein Class?" Journal of Proteomics and Bioinformatics **05**(11): 262-264.

Hernandez, S., G. Ferragut, I. Amela, J. Perez-Pons, J. Pinol, A. Mozo-Villarias, J. Cedano and E. Querol (2014). "MultitaskProtDB: a database of multitasking proteins." Nucleic Acids Research **42**(Database issue): D517-520.

Ho, L. C., M. Singh, G. Suresh, G. C. Ng and E. H. Yap (1993). "Axenic culture of Blastocystis hominis in Iscove's modified Dulbecco's medium." Parasitology Research **79**(7): 614-616.

Huelsenbeck, J. P. and F. Ronquist (2001). "MRBAYES: Bayesian inference of phylogenetic trees." Bioinformatics **17**(8): 754-755.

Hussein, E. M., A. M. Hussein, M. M. Eida and M. M. Atwa (2008). "Pathophysiological variability of different genotypes of human Blastocystis hominis Egyptian isolates in experimentally infected rats." Parasitology Research **102**(5): 853-860.

Huther, F. J., N. Psarros and H. Duschner (1990). "Isolation, characterization, and inhibition kinetics of enolase from Streptococcus rattus FA-1." Infection and Immunity **58**(4): 1043-1047.

Isupov, M. N., T. M. Fleming, A. R. Dalby, G. S. Crowhurst, P. C. Bourne and J. A. Littlechild (1999). "Crystal structure of the glyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic archaeon Sulfolobus solfataricus." Journal of Molecular Biology **291**(3): 651-660.

Jantermtor, S., P. Pinlaor, K. Sawadpanich, S. Pinlaor, A. Sangka, C. Wilailuckana, W. Wongsena and H. Yoshikawa (2013). "Subtype identification of Blastocystis spp. isolated from patients in a major hospital in northeastern Thailand." Parasitology Research **112**(4): 1781-1786.

Jenkin, J. (2001). "A unique partnership: William and Lawrence Bragg and the 1915 Nobel Prize in Physics." Minerva **39**(4): 373-392.

Jenkins, J. L. and J. J. Tanner (2006). "High-resolution structure of human D-glyceraldehyde-3-phosphate dehydrogenase." Acta Crystallographica Section D Biological Crystallography **62**(Pt 3): 290-301.

Jia, B., T. Linh le, S. Lee, B. P. Pham, J. Liu, H. Pan, S. Zhang and G. W. Cheong (2011). "Biochemical characterization of glyceraldehyde-3-phosphate dehydrogenase from *Thermococcus kodakarensis* KOD1." Extremophiles **15**(3): 337-346.

Jimenez, L., N. Vibanco-Perez, L. Navarro and A. Landa (2000). "Cloning, expression and characterisation of a recombinant triosephosphate isomerase from *Taenia solium*." International Journal for Parasitology **30**(9): 1007-1012.

Johnson, M. (2014). "Detergents: Triton X-100, Tween-20, and More." Materials and Methods **3**.

Jones, S. and J. M. Thornton (2004). "Searching for functional sites in protein structures." Current Opinion in Chemical Biology **8**(1): 3-7.

Joseph, J., F. F. Cruz-Sanchez and J. Carreras (1996). "Enolase activity and isoenzyme distribution in human brain regions and tumors." Journal of Neurochemistry **66**(6): 2484-2490.

Kabsch, W. (2010). "Integration, scaling, space-group assignment and post-refinement." Acta Crystallographica Section D Biological Crystallography **66**(Pt 2): 133-144.

Kabsch, W. (2010). "Xds." Acta Crystallographica Section D Biological Crystallography **66**(Pt 2): 125-132.

Karbassi, F., V. Quiros, V. Pancholi and M. J. Kornblatt (2010). "Dissociation of the octameric enolase from *S. pyogenes*--one interface stabilizes another." PLoS One **5**(1): e8810.

Karkowska-Kuleta, J. and A. Kozik (2014). "Moonlighting proteins as virulence factors of pathogenic fungi, parasitic protozoa and multicellular parasites." Molecular Oral Microbiology **29**(6): 270-283.

Katsarou-Katsari, A., C. M. Vassalos, K. Tzanetou, G. Spanakos, C. Papadopoulou and N. Vakalis (2008). "Acute urticaria associated with amoeboid forms of *Blastocystis* sp. subtype 3." Acta Dermato-Venereologica **88**(1): 80-81.

Kaya, S., E. S. Cetin, B. C. Aridogan, S. Arikan and M. Demirci (2007). "Pathogenicity of *Blastocystis hominis*, a clinical reevaluation." Türkiye Parazitoloji Dergisi **31**(3): 184-187.

Keller, A., H. Scarna, A. Mermet and J. F. Pujol (1981). "Biochemical and immunological properties of the mouse brain enolases purified by a simple method." Journal of Neurochemistry **36**(4): 1389-1397.

Kinoshita, T., R. Maruki, M. Warizaya, H. Nakajima and S. Nishimura (2005). "Structure of a high-resolution crystal form of human triosephosphate isomerase: improvement of crystals using the gel-tube method." Acta Crystallographica Section F Structural Biology and Crystallization Communications **61**(Pt 4): 346-349.

Knobeloch, D., A. Schmidt, P. Scheerer, N. Krauss, H. Wessner, C. Scholz, G. Kuttner, T. von Rintelen, A. Wessel and W. Hohne (2010). "A coleopteran triosephosphate isomerase: X-ray structure and phylogenetic impact of insect sequences." Insect Molecular Biology **19**(1): 35-48.

Koch, A. L. (1972). "Enzyme evolution. I. The importance of untranslatable intermediates." Genetics **72**(2): 297-316.

Kohlhoff, M., A. Dahm and R. Hensel (1996). "Tetrameric triosephosphate isomerase from hyperthermophilic Archaea." FEBS Letters **383**(3): 245-250.

Kolln, J., Y. Zhang, G. Thai, M. Demetriou, N. Hermanowicz, P. Duquette, S. van den Noort and Y. Qin (2010). "Inhibition of glyceraldehyde-3-phosphate dehydrogenase activity by antibodies present in the cerebrospinal fluid of patients with multiple sclerosis." Journal of Immunology **185**(3): 1968-1975.

Krissinel, E. and K. Henrick (2007). "Inference of macromolecular assemblies from crystalline state." Journal of Molecular Biology **372**(3): 774-797.

Krogsgaard, L. R., A. L. Engsbro, C. R. Stensvold, H. V. Nielsen and P. Bytzer (2015). "The prevalence of intestinal parasites is not greater among individuals with irritable bowel syndrome: a population-based case-control study." Clinical Gastroenterology and Hepatology **13**(3): 507-513 e502.

Lama, A., A. Kucknoor, V. Mundodi and J. F. Alderete (2009). "Glyceraldehyde-3-phosphate dehydrogenase is a surface-associated, fibronectin-binding protein of *Trichomonas vaginalis*." Infection and Immunity **77**(7): 2703-2711.

Lanuza, M. D., J. A. Carbajal and R. Borrás (1996). "Identification of surface coat carbohydrates in *Blastocystis hominis* by lectin probes." International Journal for Parasitology **26**(5): 527-532.

Lanuza, M. D., J. A. Carbajal, J. Villar and R. Borrás (1997). "Description of an improved method for *Blastocystis hominis* culture and axenization." Parasitology Research **83**(1): 60-63.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins (2007). "Clustal W and Clustal X version 2.0." Bioinformatics **23**(21): 2947-2948.

Laskowski, R. A., M. W. MacArthur, D. S. Moss and J. M. Thornton (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." Journal of Applied Crystallography **26**(2): 283-291.

Lebedev, A. A. and M. N. Isupov (2014). "Space-group and origin ambiguity in macromolecular structures with pseudo-symmetry and its treatment with the program Zanuda." Acta Crystallographica Section D Biological Crystallography **70**(Pt 9): 2430-2443.

Lee, J. H., H. K. Kang, Y. H. Moon, D. L. Cho, D. Kim, J. Y. Choe, R. Honzatko and J. F. Robyt (2006). "Cloning, expression and characterization of an extracellular enolase from *Leuconostoc mesenteroides*." FEMS Microbiology Letters **259**(2): 240-248.

Li, R. D. and L. Liu (2012). "Characterizing criticality of proteins by systems dynamics: *Escherichia coli* central carbon metabolism as a working example." BMC Systems Biology **6 Suppl 1**: S11.

Liaud, M. F., C. Lichtle, K. Apt, W. Martin and R. Cerff (2000). "Compartment-specific isoforms of TPI and GAPDH are imported into diatom mitochondria as a fusion protein: evidence in favor of a mitochondrial origin of the eukaryotic glycolytic pathway." Molecular Biology and Evolution **17**(2): 213-223.

Littlechild, J. A. and M. Isupov (2001). "Glyceraldehyde-3-phosphate dehydrogenase from *Sulfolobus solfataricus*." Methods in Enzymology **331**: 105-117.

Liu, K.-J. (2007). "The role of enolase in tissue invasion and metastasis of pathogens and tumor cells." Journal of Cancer Molecules.

Malay, A. D., Y. Bessho, M. J. Ellis, S. V. Antonyuk, R. W. Strange, S. S. Hasnain, A. Shinkai, B. Padmanabhan and S. Yokoyama (2009). "Structure of glyceraldehyde-3-phosphate dehydrogenase from the archaeal hyperthermophile *Methanocaldococcus jannaschii*." Acta Crystallographica Section F Structural Biology and Crystallization Communications **65**(Pt 12): 1227-1233.

Markos, A., A. Miretsky and M. Muller (1993). "A glyceraldehyde-3-phosphate dehydrogenase with eubacterial features in the amitochondriate eukaryote, *Trichomonas vaginalis*." Journal of Molecular Evolution **37**(6): 631-643.

Martin, W., M. Hoffmeister, C. Rotte and K. Henze (2001). "An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle." Biological Chemistry **382**(11): 1521-1539.

Mazzola, J. L. and M. A. Sirover (2002). "Alteration of nuclear glyceraldehyde-3-phosphate dehydrogenase structure in Huntington's disease fibroblasts." Molecular Brain Research **100**(1-2): 95-101.

Mehlhorn, H. (1988). "Blastocystis hominis, Brumpt 1912: are there different stages or species?" Parasitology Research **74**(4): 393-395.

Meloni, D., P. Poirier, C. Mantini, C. Noel, N. Gantois, I. Wawrzyniak, F. Delbac, M. Chabe, L. Delhaes, E. Dei-Cas, P. L. Fiori, H. El Alaoui and E. Viscogliosi (2012). "Mixed human intra- and inter-subtype infections with the parasite *Blastocystis* sp." Parasitology International **61**(4): 719-722.

Merkulova, T., M. Lucas, C. Jabet, N. Lamande, J. D. Rouzeau, F. Gros, M. Lazar and A. Keller (1997). "Biochemical characterization of the mouse muscle-specific enolase: developmental changes in electrophoretic variants and selective binding to other proteins." Biochemical Journal **323 (Pt 3)**: 791-800.

Mezzasalma, T. M., J. K. Kranz, W. Chan, G. T. Struble, C. Schalk-Hihi, I. C. Deckman, B. A. Springer and M. J. Todd (2007). "Enhancing recombinant protein quality and yield by protein stability profiling." Journal of Biomolecular Screening **12**(3): 418-428.

Michels, P. A. (1988). "Compartmentation of glycolysis in trypanosomes: a potential target for new trypanocidal drugs." Biology of the Cell **64**(2): 157-164.

Miernyk, J. A. and D. T. Dennis (1984). "Enolase isozymes from *Ricinus communis*: partial purification and characterization of the isozymes." Archives of Biochemistry and Biophysics **233**(2): 643-651.

Mirza, H. and K. S. Tan (2009). "Blastocystis exhibits inter- and intra-subtype variation in cysteine protease activity." Parasitology Research **104**(2): 355-361.

Misset, O., O. J. Bos and F. R. Opperdoes (1986). "Glycolytic enzymes of *Trypanosoma brucei*. Simultaneous purification, intraglycosomal concentrations and physical properties." European Journal of Biochemistry **157**(2): 441-453.

Moras, D., K. W. Olsen, M. N. Sabesan, M. Buehner, G. C. Ford and M. G. Rossmann (1975). "Studies of asymmetry in the three-dimensional structure of lobster D-glyceraldehyde-3-phosphate dehydrogenase." Journal of Biological Chemistry **250**(23): 9137-9162.

Murshudov, G. N., P. Skubak, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin (2011). "REFMAC5 for the refinement of macromolecular crystal structures." Acta Crystallographica Section D Biological Crystallography **67**(Pt 4): 355-367.

Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallographica Section D Biological Crystallography **53**(Pt 3): 240-255.

Nagano, N., C. A. Orengo and J. M. Thornton (2002). "One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions." Journal of Molecular Biology **321**(5): 741-765.

Nagel, R., C. Gray, H. Bielefeldt-Ohmann and R. J. Traub (2015). "Features of *Blastocystis* spp. in xenic culture revealed by deconvolutional microscopy." Parasitology Research.

Nakayama, T., K. Ishida and J. M. Archibald (2013). "Broad Distribution of TPI-GAPDH Fusion Proteins among Eukaryotes: Evidence for Glycolytic Reactions in the Mitochondrion?" Plos One **7**(12): e52340.

Nielsen, H., J. Engelbrecht, S. Brunak and G. v. Heijne (1996). "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites." Protein Engineering **10**: 1-6.

Niesen, F. H., H. Berglund and M. Vedadi (2007). "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability." Nature Protocols **2**(9): 2212-2221.

Nobbmann, U., M. Connah, B. Fish, P. Varley, C. Gee, S. Mulot, J. Chen, L. Zhou, Y. Lu, F. Shen, J. Yi and S. E. Harding (2007). "Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies." Biotechnology and Genetic Engineering Reviews **24**: 117-128.

Nourrisson, C., J. Scanzi, B. Pereira, C. NkoudMongo, I. Wawrzyniak, A. Cian, E. Viscogliosi, V. Livrelli, F. Delbac, M. Dapoigny and P. Poirier (2014). "Blastocystis is associated with decrease of fecal microbiota protective bacteria: comparative analysis between patients with irritable bowel syndrome and control subjects." PLoS One **9**(11): e111868.

Olga Randelj, J. R., and Christian Motz (2007). Protein Targeting Protocols. M. v. d. Giezen. Totowa, New Jersey, Humana press: 417-427.

Orosz, F., J. Olah and J. Ovadi (2006). "Triosephosphate isomerase deficiency: facts and doubts." International Union of Biochemistry and Molecular Biology Life (IUBMB Life) **58**(12): 703-715.

Orosz, F., J. Olah and J. Ovadi (2009). "Triosephosphate isomerase deficiency: new insights into an enigmatic disease." Biochimica et Biophysica Acta (BBA) - Bioenergetics **1792**(12): 1168-1174.

Pareek, V., M. Samanta, N. V. Joshi, H. Balaram, M. R. Murthy and P. Balaram (2016). "Connecting Active-Site Loop Conformations and Catalysis in Triosephosphate Isomerase: Insights from a Rare Variation at Residue 96 in the Plasmodial Enzyme." ChemBioChem **17**(7): 620-629.

Parkar, U., R. J. Traub, S. Kumar, M. Mungthin, S. Vitali, S. Leelayoova, K. Morris and R. C. Thompson (2007). "Direct characterization of Blastocystis from faeces by PCR and evidence of zoonotic potential." Parasitology **134**(Pt 3): 359-367.

Piast, M., I. Kustrzeba-Wojcicka, M. Matusiewicz and T. Banas (2005). "Molecular evolution of enolase." Acta Biochimica Polonica **52**(2): 507-513.

Poirier, P., I. Wawrzyniak, C. P. Vivares, F. Delbac and H. El Alaoui (2012). "New insights into *Blastocystis* spp.: a potential link with irritable bowel syndrome." *PLOS Pathogens* **8**(3): e1002545.

Puthia, M. K., J. Lu and K. S. Tan (2008). "Blastocystis ratti contains cysteine proteases that mediate interleukin-8 response from human intestinal epithelial cells in an NF-kappaB-dependent manner." *Eukaryotic Cell* **7**(3): 435-443.

Puthia, M. K., A. Vaithilingam, J. Lu and K. S. Tan (2005). "Degradation of human secretory immunoglobulin A by Blastocystis." *Parasitology Research* **97**(5): 386-389.

Qin, J., G. Q. Chai, J. M. Brewer, L. L. Lovelace and L. Lebioda (2012). "Structures of asymmetric complexes of human neuron specific enolase with resolved substrate and product and an analogous complex with two inhibitors indicate subunit interaction and inhibitor cooperativity." *Journal of Inorganic Biochemistry* **111**: 187-194.

Quinones, W., P. Pena, M. Domingo-Sananes, A. Caceres, P. A. Michels, L. Avilan and J. L. Concepcion (2007). "Leishmania mexicana: molecular cloning and characterization of enolase." *Experimental Parasitology* **116**(3): 241-251.

Ramiah, K., C. A. van Reenen and L. M. Dicks (2008). "Surface-bound proteins of *Lactobacillus plantarum* 423 that contribute to adhesion of Caco-2 cells and their role in competitive exclusion and displacement of *Clostridium sporogenes* and *Enterococcus faecalis*." *Research in Microbiology* **159**(6): 470-475.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard and J. P. Huelsenbeck (2012). "MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space." *Systematic Biology* **61**(3): 539-542.

Sabate, R., N. S. de Groot and S. Ventura (2010). "Protein folding and aggregation in bacteria." *Cellular and Molecular Life Sciences* **67**(16): 2695-2715.

Schreier, B. and B. Hocker (2010). "Engineering the enolase magnesium II binding site: implications for its evolution." *Biochemistry* **49**(35): 7582-7589.

Schurig, H., K. Rutkat, R. Rachel and R. Jaenicke (1995). "Octameric enolase from the hyperthermophilic bacterium *Thermotoga maritima*: purification, characterization, and image processing." Protein Science **4**(2): 228-236.

Senisterra, G., I. Chau and M. Vedadi (2012). "Thermal denaturation assays in chemical biology." ASSAY and Drug Development Technologies **10**(2): 128-136.

Senisterra, G. A. and P. J. Finerty, Jr. (2009). "High throughput methods of assessing protein stability and aggregation." Molecular BioSystems **5**(3): 217-223.

Shi, S., A. Semple, J. Cheung and M. Shameem (2013). "DSF method optimization and its application in predicting protein thermal aggregation kinetics." Journal of Pharmaceutical Sciences **102**(8): 2471-2483.

Silberman, J. D., M. L. Sogin, D. D. Leipe and C. G. Clark (1996). "Human parasite finds taxonomic home." Nature **380**(6573): 398-398.

Sirover, M. A. (1999). "New insights into an old protein: the functional diversity of mammalian glyceraldehyde-3-phosphate dehydrogenase." Biochimica et Biophysica Acta (BBA) - Bioenergetics **1432**(2): 159-184.

Sirover, M. A. (2005). "New nuclear functions of the glycolytic protein, glyceraldehyde-3-phosphate dehydrogenase, in mammalian cells." Journal of Cellular Biochemistry **95**(1): 45-52.

Small, I., N. Peeters, F. Legeai and C. Lurin (2004). "Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences." Proteomics **4**(6): 1581-1590.

Souppart, L., G. Sancier, A. Cian, I. Wawrzyniak, F. Delbac, M. Capron, E. Dei-Cas, K. Boorom, L. Delhaes and E. Viscogliosi (2009). "Molecular epidemiology of human *Blastocystis* isolates in France." Parasitology Research **105**(2): 413-421.

Stechmann, A., K. Hamblin, V. Perez-Brocal, D. Gaston, G. S. Richmond, M. Van der Giezen, C. G. Clark and A. J. Roger (2008). "Organelles in *Blastocystis* that blur the distinction between mitochondria and hydrogenosomes." Current Biology **18**(8): 580-585.

Stensvold, C. R. (2013). "Blastocystis: Genetic diversity and molecular methods for diagnosis and epidemiology." Tropical Parasitology **3**(1): 26-34.

Stenzel, D. J. and P. F. Boreham (1996). "Blastocystis hominis revisited." Clinical Microbiology Reviews **9**(4): 563-584.

Stryer, L. (1995). Biochemistry Stanford University, United states of America, W. H. Freeman and company New York.

Studier, F. W. (2005). "Protein production by auto-induction in high density shaking cultures." Protein Expression and Purification **41**(1): 207-234.

Sun, H. (2006). "The interaction between pathogens and the host coagulation system." Physiology (Bethesda) **21**: 281-288.

Takishita, K., N. J. Patron, K. Ishida, T. Maruyama and P. J. Keeling (2005). "A transcriptional fusion of genes encoding glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and enolase in dinoflagellates." Journal of Eukaryotic Microbiology **52**(4): 343-348.

Tan, K. S. (2008). "New insights on classification, identification, and clinical relevance of Blastocystis spp." Clinical Microbiology Reviews **21**(4): 639-665.

Tan, K. S., J. Howe, E. H. Yap and M. Singh (2001). "Do Blastocystis hominis colony forms undergo programmed cell death?" Parasitology Research **87**(5): 362-367.

Tan, K. S., M. Singh and E. H. Yap (2002). "Recent advances in Blastocystis hominis research: hot spots in terra incognita." International Journal for Parasitology **32**(7): 789-804.

Tan, S. W., M. Singh, K. T. Thong, L. C. Ho, K. T. Moe, X. Q. Chen, G. C. Ng and E. H. Yap (1996). "Clonal growth of Blastocystis hominis in soft agar with sodium thioglycollate." Parasitology Research **82**(8): 737-739.

Tan, S. W., M. Singh, E. H. Yap, L. C. Ho, K. T. Moe, J. Howe and G. C. Ng (1996). "Colony formation of Blastocystis hominis in soft agar." Parasitology Research **82**(4): 375-377.

Tan, T. C., S. C. Ong and K. G. Suresh (2009). "Genetic variability of Blastocystis sp isolates obtained from cancer and HIV/AIDS patients." Parasitology Research **105**(5): 1283-1286.

Tan, T. C., S. C. Ong and K. G. Suresh (2009). "Genetic variability of Blastocystis sp. isolates obtained from cancer and HIV/AIDS patients." Parasitology Research **105**(5): 1283-1286.

Tan, T. C. and K. G. Suresh (2006). "Predominance of amoeboid forms of *Blastocystis hominis* in isolates from symptomatic patients." Parasitology Research **98**(3): 189-193.

Tiwari, S. P. and N. Reuter (2016). "Similarity in Shape Dictates Signature Intrinsic Dynamics Despite No Functional Conservation in TIM Barrel Enzymes." PLOS Computational Biology **12**(3): e1004834.

Upadhyay, A. K., A. Murmu, A. Singh and A. K. Panda (2012). "Kinetics of inclusion body formation and its correlation with the characteristics of protein aggregates in *Escherichia coli*." PLoS One **7**(3): e33951.

Vagin, A. and A. Teplyakov (2010). "Molecular replacement with MOLREP." Acta Crystallographica Section D Biological Crystallography **66**(Pt 1): 22-25.

Valsecchi, R., P. Leghissa and V. Greco (2004). "Cutaneous lesions in *Blastocystis hominis* infection." Acta Dermato-Venereologica **84**(4): 322-323.

vanderGiezen, M., K. B. Rechinger, I. Svendsen, R. Durand, R. P. Hirt, M. Fevre, T. M. Embley and R. A. Prins (1997). "A mitochondrial-like targeting signal on the hydrogenosomal malic enzyme from the anaerobic fungus *Neocallimastix frontalis*: Support for the hypothesis that hydrogenosomes are modified mitochondria." Molecular Microbiology **23**(1): 11-21.

Vanegas, G., W. Quinones, C. Carrasco-Lopez, J. L. Concepcion, F. Albericio and L. Avilan (2007). "Enolase as a plasminogen binding protein in *Leishmania mexicana*." Parasitology Research **101**(6): 1511-1516.

Vassalos, C. M., G. Spanakos, E. Vassalou, C. Papadopoulou and N. Vakalis (2010). "Differences in clinical significance and morphologic features of *Blastocystis* sp subtype 3." American Journal of Clinical Pathology **133**(2): 251-258.

Vdovenko, A. A. (2000). "Blastocystis hominis: origin and significance of vacuolar and granular forms." Parasitology Research **86**(1): 8-10.

Vivoli, M., H. R. Novak, J. A. Littlechild and N. J. Harmer (2014). "Determination of protein-ligand interactions using differential scanning fluorimetry." Journal of Visualized Experiments (JoVE)(91): 51809.

von Heijne, G. (1986). "Mitochondrial targeting sequences may form amphiphilic helices." EMBO J **5**(6): 1335-1342.

Walden, H., G. Taylor, H. Lilie, T. Knura and R. Hensel (2004). "Triosephosphate isomerase of the hyperthermophile *Thermoproteus tenax*: thermostability is not everything." Biochemical Society Transactions **32**(Pt 2): 305.

Walden, H., G. L. Taylor, E. Lorentzen, E. Pohl, H. Lilie, A. Schramm, T. Knura, K. Stubbe, B. Tjaden and R. Hensel (2004). "Structure and function of a regulated archaeal triosephosphate isomerase adapted to high temperature." Journal of Molecular Biology **342**(3): 861-875.

Wang, X. Y., W. J. Chen, F. Y. Hu, C. H. Deng, C. H. Zhou, X. L. Lv, Y. X. Fan, J. T. Men, Y. Huang, J. F. Sun, D. Hu, J. F. Chen, Y. B. Yang, C. Liang, H. Q. Zheng, X. C. Hu, J. Xu, Z. D. Wu and X. B. Yu (2011). "Clonorchis sinensis enolase: Identification and biochemical characterization of a glycolytic enzyme from excretory/secretory products." Molecular and Biochemical Parasitology **177**(2): 135-142.

Wawrzyniak, I., P. Poirier, E. Viscogliosi, M. Dionigia, C. Texier, F. Delbac and H. E. Alaoui (2013). "Blastocystis, an unrecognized parasite: an overview of pathogenesis and diagnosis." Therapeutic Advances in Infectious Disease **1**(5): 167-178.

Wierenga, R. K., M. E. Noble, G. Vriend, S. Nauche and W. G. Hol (1991). "Refined 1.83 Å structure of trypanosomal triosephosphate isomerase crystallized in the presence of 2.4 M-ammonium sulphate. A comparison with the structure of the trypanosomal triosephosphate isomerase-glycerol-3-phosphate complex." Journal of Molecular Biology **220**(4): 995-1015.

Wistow, G. and J. Piatigorsky (1987). "Recruitment of enzymes as lens structural proteins." Science **236**(4808): 1554-1556.

Wittig, I., H. P. Braun and H. Schagger (2006). "Blue native PAGE." Nature Protocols **1**(1): 418-428.

Wright, M. D., K. M. Davern and G. F. Mitchell (1991). "The functional and immunological significance of some schistosome surface molecules." Parasitology Today **7**(2): 56-58.

Yang, J., C. Qiu, Y. Xia, L. Yao, Z. Fu, C. Yuan, X. Feng and J. Lin (2010). "Molecular cloning and functional characterization of Schistosoma japonicum enolase which is highly expressed at the schistosomulum stage." Parasitology Research **107**(3): 667-677.

Zaman, V., J. Howe and M. Ng (1995). "Ultrastructure of Blastocystis hominis cysts." Parasitology Research **81**(6): 465-469.

Zhang, J. Y., F. Zhang, C. Q. Hong, A. E. Giuliano, X. J. Cui, G. J. Zhou, G. J. Zhang and Y. K. Cui (2015). "Critical protein GAPDH and its regulatory mechanisms in cancer cells." Cancer Biology and Medicine **12**(1): 10-22.

Zhao, Y., Q. Liu, X. Wang, L. Zhou, Q. Wang and Y. Zhang (2011). "Surface display of Aeromonas hydrophila GAPDH in attenuated Vibrio anguillarum to develop a Noval multivalent vector vaccine." Marine Biotechnology (New York, N.Y.) **13**(5): 963-970.

Zierdt, C. H. (1983). "Blastocystis hominis, a protozoan parasite and intestinal pathogen of human beings." Clinical microbiology newsletter **5**(9).

Zierdt, C. H. (1986). "Cytochrome-free mitochondria of an anaerobic protozoan--Blastocystis hominis." Journal of Protozoology **33**(1): 67-69.

Zierdt, C. H. (1988). "Blastocystis-Hominis, a Long-Misunderstood Intestinal Parasite." Parasitology Today **4**(1): 15-17.

Zierdt, C. H. (1991). "Blastocystis hominis--past and future." Clinical Microbiology Reviews **4**(1): 61-79.

Zierdt, C. H., C. T. Donnelly, J. Muller and G. Constantopoulos (1988). "Biochemical and Ultrastructural-Study of Blastocystis-Hominis." Journal of Clinical Microbiology **26**(5): 965-970.

Zierdt, C. H. and R. L. Williams (1974). "Blastocystis hominis: axenic cultivation." Experimental Parasitology **36**(2): 233-243.