OPEN ACCESS Freely available online PLoS one

Metabolic Profiling of the Protozoan Parasite *Entamoeba invadens* Revealed Activation of Unpredicted Pathway during Encystation

Ghulam Jeelani[1,2], Dan Sato[3], Afzal Husain[1,4¤], Aleyla Escueta-de Cadiz[1,4], Masahiro Sugimoto[3], Tomoyoshi Soga[3], Makoto Suematsu[2], Tomoyoshi Nozaki[1,5]*

1 Department of Parasitology, National Institute of Infectious Diseases, Shinjuku, Tokyo, Japan, **2** Department of Biochemistry and Integrative Medical Biology, School of Medicine, Keio University, Shinjuku, Tokyo, Japan, **3** Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, Japan, **4** Department of Parasitology, Graduate School of Medicine, Gunma University, Maebashi, Japan, **5** Graduate School of Life and Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan

Abstract

Encystation, which is cellular differentiation from the motile, proliferative, labile trophozoite form to the dormant, resistant cyst form, is a crucial process found in parasitic and free-living protozoa such as *Entamoeba*, *Giardia*, *Acanthamoeba*, and *Balamuthia*. Since encystation is an essential process to deal with the adverse external environmental changes during the life cycle, and often integral to the transmission of the diseases, biochemical understanding of the process potentially provides useful measures against the infections caused by this group of protozoa. In this study, we investigated metabolic and transcriptomic changes that occur during encystation in *Entamoeba invadens*, the reptilian sibling of mammal-infecting *E. histolytica*, using capillary electrophoresis-tandem mass spectrometry-based metabolite profiling and DNA microarray-based expression profiling. As the encystation progressed, the levels of majority of metabolites involved in glycolysis and nucleotides drastically decreased, indicating energy generation is ceased. Furthermore, the flux of glycolysis was redirected toward chitin wall biosynthesis. We found remarkable temporal increases in biogenic amines such as isoamylamine, isobutylamine, and cadaverine, during the early period of encystation, when the trophozoites form large multicellular aggregates (precyst). We also found remarkable induction of γ-aminobutyric acid (GABA) during encystation. This study has unveiled for the first time the dynamics of the transcriptional and metabolic regulatory networks during encystation, and should help in better understanding of the process in pathogenic eukaryotes, and further development of measures controlling infections they cause.

Citation: Jeelani G, Sato D, Husain A, Escueta-de Cadiz A, Sugimoto M, et al. (2012) Metabolic Profiling of the Protozoan Parasite *Entamoeba invadens* Revealed Activation of Unpredicted Pathway during Encystation. PLoS ONE 7(5): e37740. doi:10.1371/journal.pone.0037740

Editor: John Parkinson, Hospital for Sick Children, Canada

Received November 21, 2011; **Accepted** April 23, 2012; **Published** May 25, 2012

Funding: This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (18GS0314, 18050006, 18073001, 20390119, 23390099), a Grant-in-Aid for Innovative Scientific Research from MEXT (23117001), a grant for research on emerging and re-emerging infectious diseases from the Ministry of Health, Labour and Welfare of Japan (H20-Shinkosaiko-ippan-016, H23-Shinkosaiko-ippan-014), and a grant for research to promote the development of anti-AIDS pharmaceuticals from the Japan Health Sciences Foundation (KAA1551, KHA1101) to TN. GJ and TN were supported in part by the Global Center of Excellence Program for Human Metabolomic System Biology of MEXT. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: nozaki@nih.go.jp

¤ Current address: Department of Immunology and Genomic Medicine, Kyoto University Graduate School of Medicine, Yoshida, Sakyo-ku, Kyoto, Japan

Introduction

Differentiation or developmental stage conversion is ubiquitous in all living organisms. For many pathogenic protozoa, it is essential for their survival and transmission. Several parasitic protozoa such as *Entamoeba histolytica* [1] and *Giardia lamblia* [2], which cause diarrhea, and free-living protozoa including *Balamuthia mandrillaris*, which causes granulomatous encephalitis [3], and *Acanthamoeba castellanii*, responsible for amoebic keratitis and encephaltitis [4], present two morphologically distinct stages in their life cycle: the motile, proliferative, labile trophozoite form, which inhabits the mammalian hosts and is responsible for pathogenesis, and the resistant cyst, which is protected by a cell wall that allows survival under the adverse external environment, and responsible for transmission [5]. The encystation process has been attracting attention not only from the viewpoint of disease transmission, but also as a model for differentiation. Therefore, elucidation of the encystation process, both at the metabolic and transcriptomic levels should lead to a better understanding of the process and, therefore, to the possibility of better control measures against infectious diseases the parasites cause.

The intestinal parasite *Entamoeba histolytica*, the causative agent of amoebiasis, is estimated to infect 50 million people annually, mainly in developing countries, where it is a major source of morbidity and mortality [6]. No axenic *in vitro* system currently exists for studying encystation of the human-infecting parasite *Entamoeba histolytica*; however, such a system exists for *E. invadens*, a parasite of reptiles. These two organisms have the same two-stage life cycle and pathogenic potential toward their vertebrate hosts [7], and form quadrinucleated, chitinaceous cysts with osmotically resistant cyst walls [8]. *E. invadens* will readily encyst *in vitro* in

response to carbon source deprivation [9], hypoosmotic shock [10], or a combination of the two stimuli [11].

As most of current drugs against protozoa target metabolism, it is critical to understand the structure and dynamics of the parasite metabolic network during encystation. Indirect approaches to reconstructing the metabolic network, by comparative genomics and enzymological studies of individual enzymes, are at the best incomplete and face major obstacles in highly divergent organisms such as parasitic protozoa. Global metabolomics is a new and powerful technology that provides a relatively complete picture of the metabolism in biological systems and has recently been applied to a wide variety of important problems [12–15]. We decided to apply this approach to understand the basis of the changes in cellular metabolism that occur during encystation. To better understand the relationship between gene expression and metabolites levels, we also analyzed the mRNA expression profile of the enzymes involved in the formation or utilization of these metabolites.

Results and Discussion

Overall strategy of metabolome and transcriptome analyses of encystation

In vitro encystation of *E. invadens* was carried out using the 47% LG medium lacking glucose [11]. Under these conditions, approximately 80% of the trophozoites differentiated into the sarkosyl-resistant cysts within 120 h (Figure 1A). We also verified the cyst formation by calcofluor staining, and showed that the percentage of the sarkosyl-resistant amoebae and that of the calcofluor-stained amoebae were comparable (data not shown). Metabolite extracts were prepared from the cell harvested at different time points during encystation (0, 0.5, 8, 24, 48, and 120 h). The capillary electrophoresis-time-of-flight mass spectrometry (CE-TOFMS) systems in cation and anion modes were used to identify the peaks [16]. The main peaks were identified and quantified with metabolite standards by matching the closest m/z values and normalized migration times for further statistical comparisons and interpretations. All data presented were normalized by cell number (per 10^6 cells) as it is the method commonly used and practically accepted in most of studies [17,18]. In addition, it is not practically possible to normalize metabolite data with cell volume, because the population during encystation is polymorphic and heterogeneous (i.e., a mixture of trophozoites and cysts with different proportions at different time points). However, one should note that the trophozoites and the cyst slightly differ in size (the diameter of usual trophozoites and cysts of *E. histolytica* ranges 13–20 and 11–14 μm, respectively [19]). Therefore, our data need to be carefully interpreted as the metabolite concentrations in cysts tend to be underestimated (potentially >2 fold). However, most, if not all, changes in metabolites presented here largely reflect changes in intracellular concentrations, but not in cell volume. The identified metabolites and their quantities are listed in Dataset S1. We identified more than 100 intermediary metabolites, which include amino acids, nucleotides, biosynthetic precursors, and central carbon metabolism intermediates (Figure 1B and C). To validate the reproducibility of the results, we compared the metabolomic and transcriptomic data from the two biological replicates at different time points during encystation. A nearly perfect correlation between the first and second replicates was observed (Figure 2).

Using hierarchical clustering analysis [20], we grouped metabolites by their temporal profiles and identify those that exhibited similar changes in the profile during encystation (Figure 1B and C). For example, clustering analysis revealed that the metabolites that are involved in chitin biosynthetic pathway displayed highly similar profiles (see below). Moreover, other metabolites known to be in common biochemical pathways such as polyamine metabolism showed similar temporal abundance profiles (see below). To correlate changes in metabolite levels with changes in transcript levels, we also globally examined gene expression by DNA microarray analysis (Dataset S2).

Glycolysis and chitin biosynthetic pathway

Glucose metabolism plays a pivotal role in cellular metabolism to produce energy and precursors of nucleotides and fatty acids [21]. Several lines of evidence suggest that this parasite relies solely on glycolysis for ATP supply, as it is devoid of the Krebs cycle and oxidative phosphorylation [22,23]. We observed that the levels of various glycolytic pathway intermediates, including glucose-6-phosphate and fructose-6-phosphate, were significantly depleted during encystation (Figure 3A). While transcript levels of many genes involved in glycolysis, and other pathways also changed during encystation (Dataset S2), the metabolites involved in chitin biosynthesis were remarkably increased after 8 h of induction, suggesting that the flux of the glycolysis was redirected toward the chitin biosynthetic pathway (Figure 3A). Chitin is the major component of the cyst walls of protozoan parasites [24] including *E. histolytica* [25], *G. lamblia* [26], *B. mandrillaris* [27], and *A. castellanii* [28]. As chitin does not occur in vertebrates, its synthetic pathway represents an excellent parasite-specific target to develop new chemotherapeutics, as proposed [29]. We found that gene expression of all the enzymes involved in chitin wall biosynthesis also increased (Figure S1).

The transcript level of glucosamine-fructose-6-phosphate aminotransferase (GFAT), which is the first and rate-limiting enzyme [30] of the chitin biosynthetic pathway, increased during encystation, whereas that of glucosamine-6 phosphate isomerase (GN6PI) remained unchanged (Figure S1), suggesting that *Entamoeba* prefers glutamine to NH_3 as an ammonia donor for the synthesis of chitin wall. Interestingly, glutamine synthase (GS) is also upregulated during encystation (Figure S1). GS catalyses the formation of glutamine from ammonium ion and glutamate. Thus, in order to meet the demands of glucosamine for chitin biosynthesis, GS seems to be co-up-regulated with GFAT to provide enough amino group donors for GFAT. The co-up-regulation of GFAT and GS may be linked to a different hypothetical role of the peritrophic matrix, namely elimination of toxic ammonium ions, which are released by utilization of arginine during encystation [5]. As reported previously, GFAT activity was sensitive to feedback inhibition by UDP-GlcNAc, the end product of the hexosamine pathway, and is modulated by cAMP-dependent protein kinase A [31]. GFAT has attracted the interest of several research groups [32] as it offers potential target for antibacterial and antifungal agents.

Expression of one of three genes encoding chitin deacetylase (CDA) (EIN_058630) was upregulated during encystation (Dataset S2), which led us to hypothesize that this protein deacetylates chitin to form chitosan (Figure 3A), which is a mixture of *N*-acetylglucosamine and glucosamine. Chitosan, which has a positive charge, is a major component of spore walls of *Saccharomyces cerevisiae* and lateral walls of *Mucor rouxii* [33]. The chitosan also has another important feature: the ability to inhibit the glycolytic pathway in cancerous cells [34].

We also found that the transcripts of the enzymes involved in glycogen/starch biosynthesis decreased, whereas those involved in their decomposition increased (Figure S1). This observation led us to propose that during encystation the glucose which is stored in the form of glycogen/starch may be redirected toward chitin

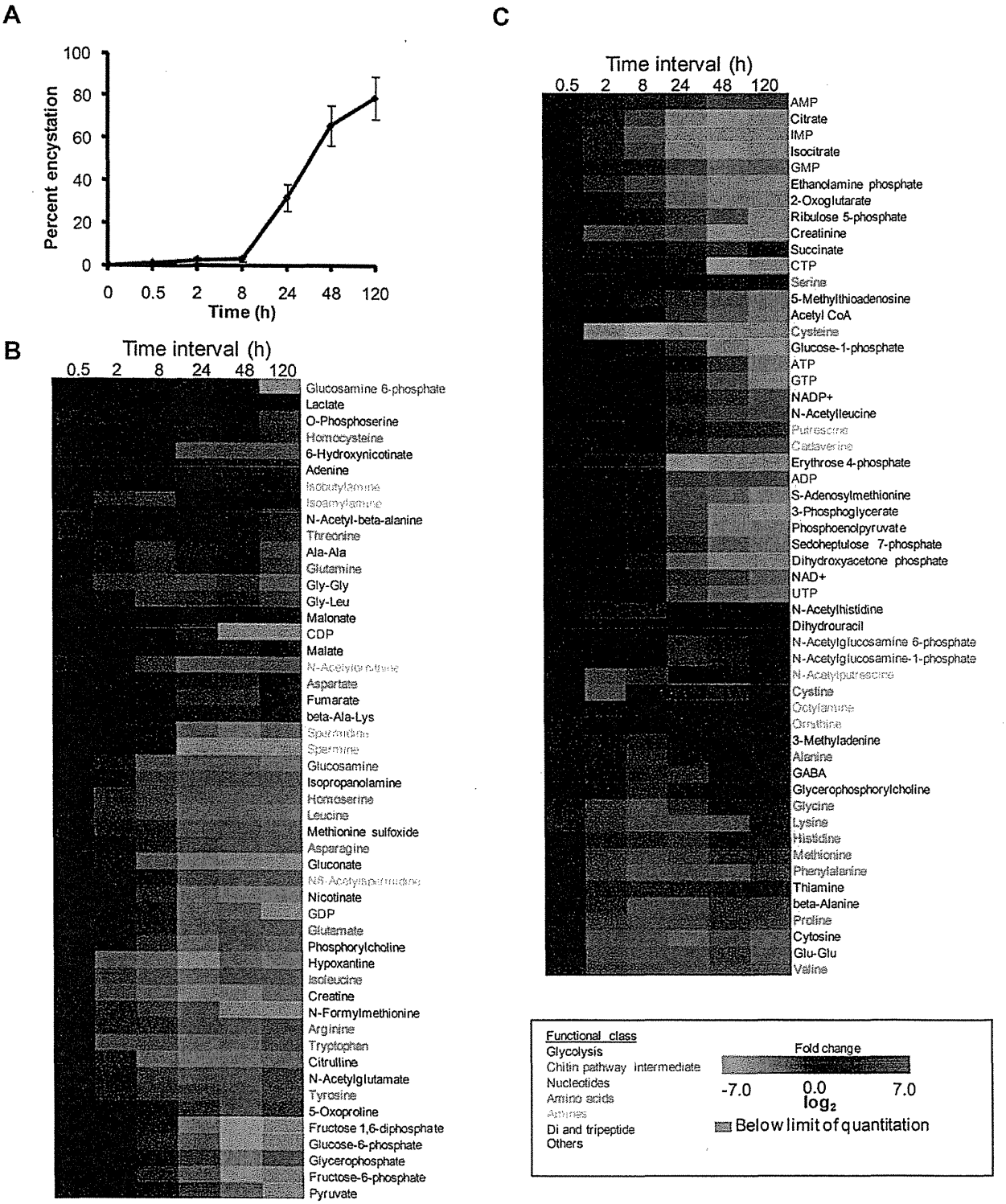


Figure 1. Change in metabolites during encystation. (**A**) Kinetics of encystation. The percentages of the amoebae resistant to 0.05% sarkosyl during encystation are shown. (**B and C**) Heat map produced by hierarchical clustering of metabolites profiles obtained from CE-TOFMS analysis. Rows correspond to metabolites and columns correspond to time intervals. Shown are 104 metabolites detected during encystation. Metabolites levels are expressed as $\log_2$ of the fold change with respect to time 0 h. Shades in red and green indicate an increase and decrease of metabolites, respectively, according to the scale bar shown at the bottom.
doi:10.1371/journal.pone.0037740.g001

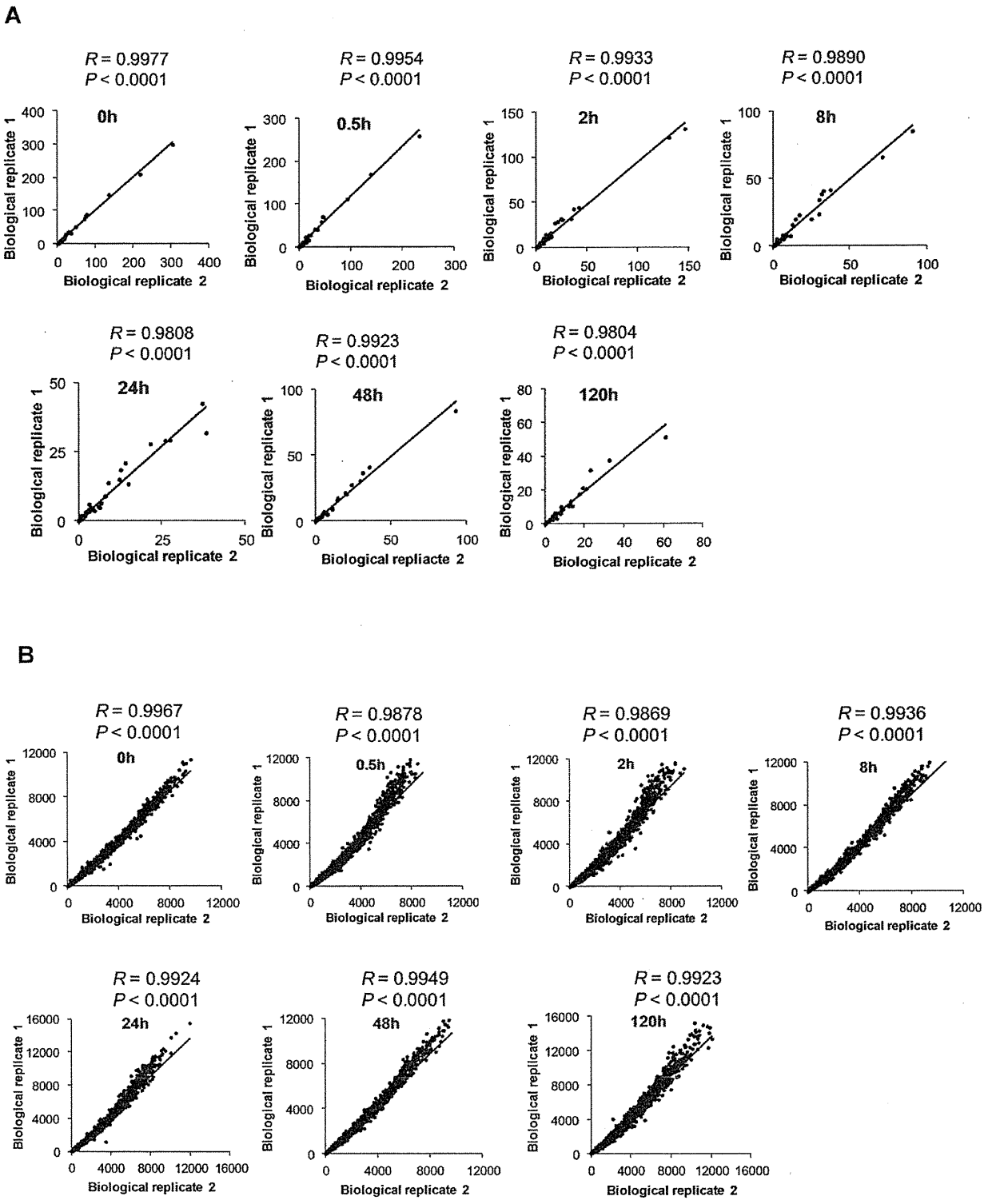


Figure 2. Correlation between two biological replicates. The correlation of the levels of metabolites in metabolomic analysis (**A**) and transcripts in DNA microarray analysis (**B**) between first and second biological replicates at different time points during encystation is shown. The Pearson correlation coefficient and their P-values (two tailed) were calculated using GraphPad prism version 5.04.
doi:10.1371/journal.pone.0037740.g002

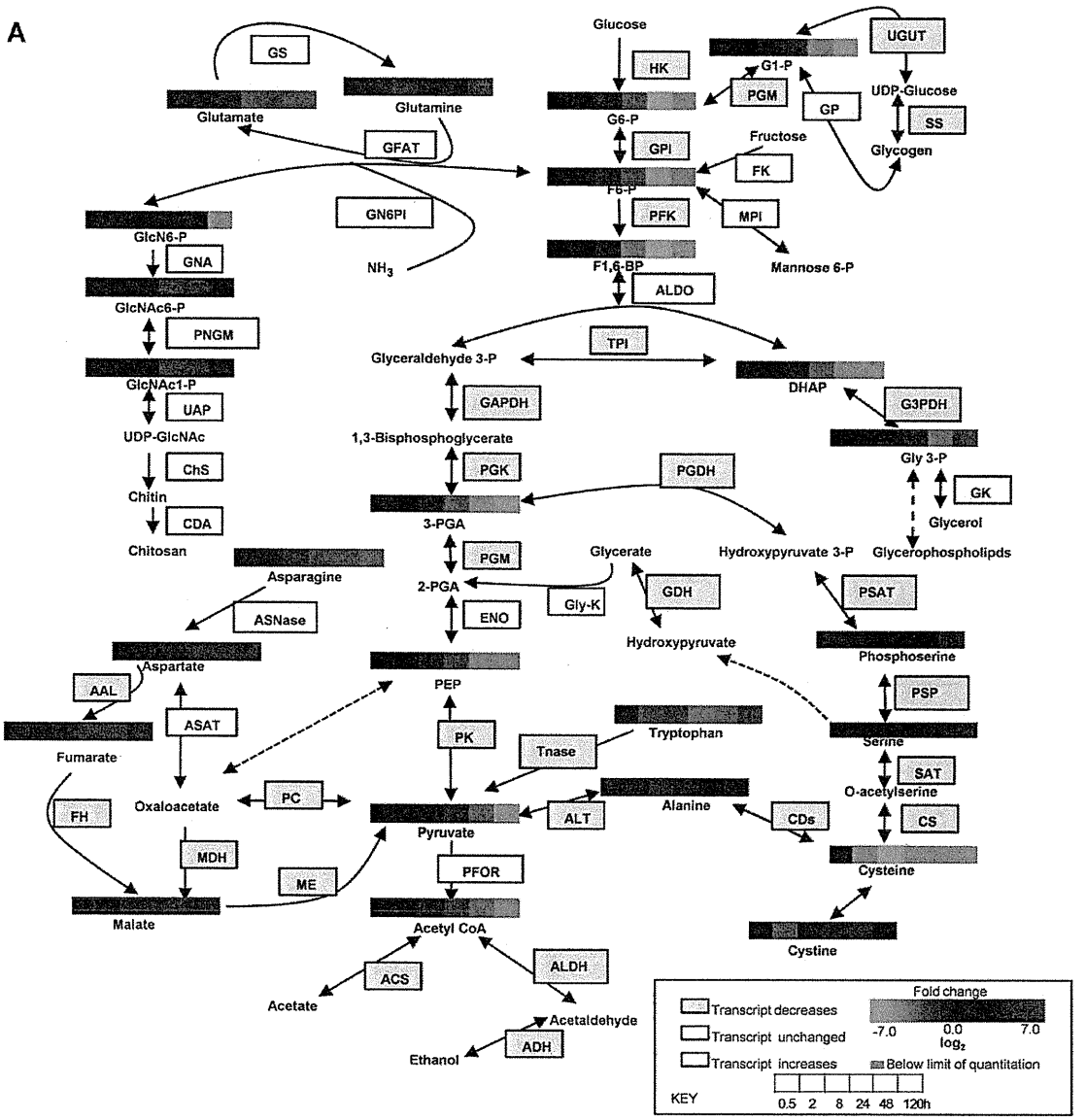

A
GS
Glutamate
Glutamine
GFAT
GN6PI
NH3
Glucose
HK
G6-P
GPI
Fructose
FK
F6-P
PFK
MPI
Mannose 6-P
F1,6-BP
ALDO
UGUT
G1-P
PGM
GP
UDP-Glucose
SS
Glycogen
GlcN6-P
GNA
GlcNAc6-P
PNGM
GlcNAc1-P
UAP
UDP-GlcNAc
ChS
Chitin
CDA
Chitosan
TPI
Glyceraldehyde 3-P
DHAP
GAPDH
G3PDH
1,3-Bisphosphoglycerate
Gly 3-P
PGK
PGDH
GK
3-PGA
Glycerol
Glycerophospholipds
Asparagine
PGM
Glycerate
Hydroxypyruvate 3-P
2-PGA
Gly-K
GDH
PSAT
ASNase
ENO
Hydroxypyruvate
Phosphoserine
Aspartate
PEP
AAL
PSP
Tryptophan
ASAT
PK
Serine
Fumarate
Tnase
SAT
O-acetylserine
Oxaloacetate
PC
Alanine
FH
ALT
CDs
CS
Pyruvate
MDH
Cysteine
ME
PFOR
Malate
Acetyl CoA
Cystine
ACS
ALDH
Acetate
Acetaldehyde
ADH
Ethanol
Fold change
Transcript decreases
Transcript unchanged
Transcript increases
-7.0
0.0
7.0
log2
Below limit of quantitation
KEY
0.5 2 8 24 48 120h
B
nmole/1×10^6 cells
0h
0.5h
2h
8h
24h
48h
120h
ATP
GTP
UTP
CTP
ADP
GDP
AMP
GMP
C
Energy charge
Time (h)
0 0.5 2 8 24 48 120

Figure 3. Central carbon metabolism during encystation. (**A**) Alterations in the levels of metabolites involved in central carbon metabolism during encystation. The metabolite levels represented by heat map strip and the transcript levels shown in rectangles during encystation are superimposed on a metabolic pathway map that includes glycolysis, chitin biosynthesis, serine biosynthesis, and other related pathways. Metabolites levels are expressed as $\log_2$ of the fold change with respect to time 0 h. Shades in red and green indicate an increase and decrease of metabolites, respectively, according to the scale bar. The enzymes are shown in yellow or blue in case where their transcript levels increased or decreased by >3 fold relative to time 0 h, respectively. The entire quantitative metabolite data, as well as the indicator of statistical significance are given in Dataset S1. Abbreviations are: G 6-P, glucose 6-phosphate; G1-P, glucose 1-phosphate; F6-P, fructose 6-phosphate; F1,6-BP, fructose 1,6-biphosphate; DHAP, dihydroxy acetone phosphate; Gly 3-P, glycerol 3-phosphate; 3-PGA, 3-phosphoglycerate; PEP, phosphoenolpyruvate; GlcN6P, glucosamine 6-phosphate; GlcNAc6-P, N-acetylglucosamine 6-phosphate; GlcNAc1-P, N-acetyl glucosamine 1-phosphate; UDP-GlcN6P, UDP-glucosamine-6-phosphate. The full names of the enzymes denoted by abbreviations, the basal expression levels at time 0 h, and the fold changes relative to 0 h at different time points during encystation can be found in supplementary dataset S2. (**B**) The average concentration of the nucleotides at various time points during encystation. Data are depicted as means ± S. D. Statistical comparisons were made by Student's t test (* $P<0.05$, ** $P<0.01$). (**C**) Adenylate energy charge of the cell, which is calculated by the equation, [(ATP)+1/2(ADP)]/[(ATP)+(ADP)+(AMP)] during encystation. Data are shown as means ± S.D. Statistical comparisons were made by Student's t test (* $P<0.05$, ** $P<0.01$).
doi:10.1371/journal.pone.0037740.g003

synthesis. Apart from the chitin wall, we also found that the level of O-phosphoserine, which is an intermediate of serine biosynthesis, increased at 24 h (Figure 3A).

Amino acid and nucleotides

Metabolomic analysis showed that most amino acids decreased during encystation, except alanine (Dataset S1). The observation is consistent with the premise that during encystation when all the glucose is used for the synthesis of chitin, various amino acids were used as an alternative energy source [35]. Interestingly, *Giardia* [36] and *Trichomonas* [37] were shown to produce alanine as a major end product of carbohydrate metabolism under anaerobic conditions [38], suggesting that carbohydrate metabolism during encystation in *Entamoeba* partially resembles that in *Giardia* and *Trichomonas*.

Expression of one of two ASAT (aspartate aminotransferase), which is involved in the conversion of aspartate to oxaloacetate (Dataset S2) (EIN_146870), increased by >400 fold during encystation, suggesting that asparagine as well as aspartate are used as an energy source when glucose is not available. It was previously reported that amino acids such as aspartate, asparagines, and arginine also serve as an energy source in anaerobic parasitic protists [39].

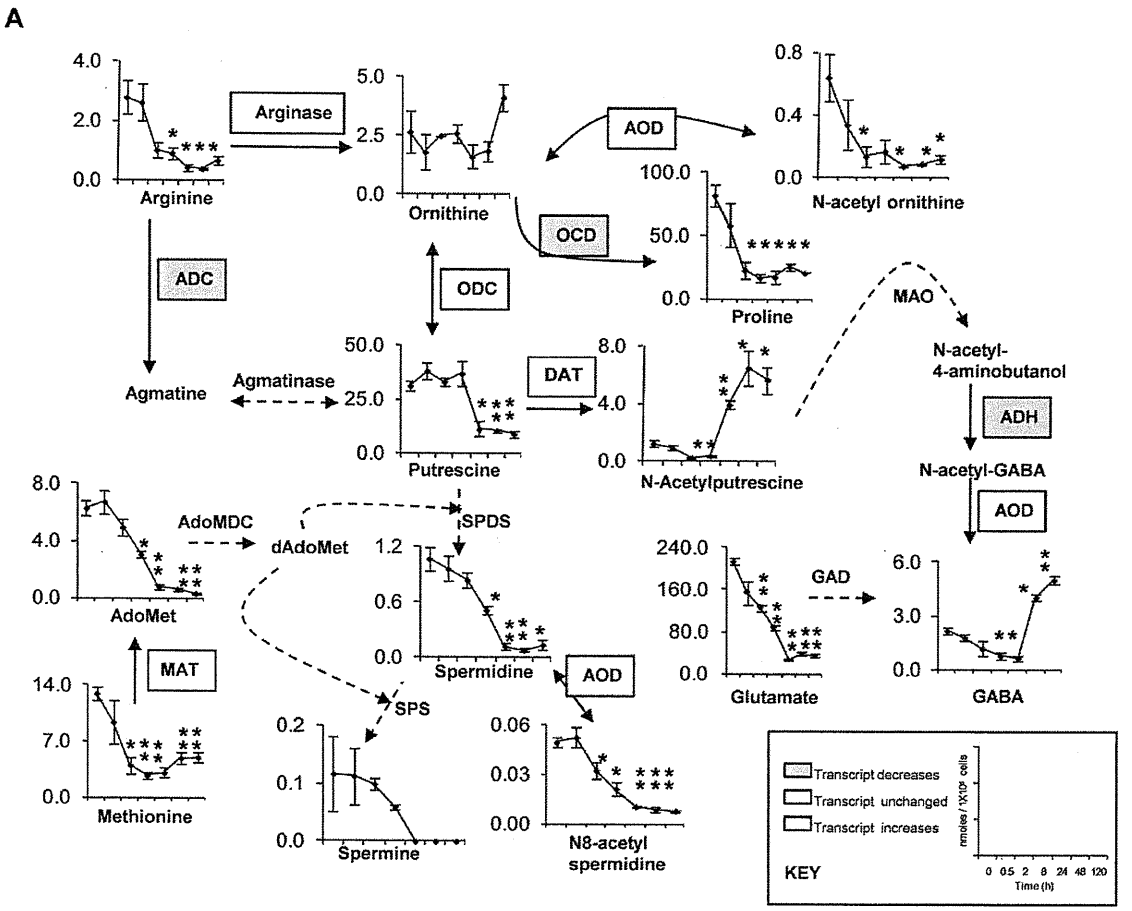
Consistent with the decrement of glycolytic intermediates and amino acids, energy content of the cell was drastically decreased during encystation. The levels of most nucleotides such as ATP, GTP, UTP, and CTP were transiently increased at 2 h, after which the level of all the nucleotides drastically decreased (Figure 3B). We also examined the average adenylate energy charge [40], which is calculated by the equation: [(ATP)+1/2(ADP)]/[(ATP)+(ADP)+(AMP)]. The energy charge in a growing organism is normally stabilized in the range of 0.8 to 0.95, and it is not greatly affected by the wide variations [41]. The energy charge of the cell notably decreased after 48 h during encystation (Figure 3C). The decrease in most purines and pyrimidines after 8 h was consistent with the previous observation for *Escherichia coli* that the nucleotide pools continuously decreased as cells moved from the exponential to stationary growth phase [42].

Polyamine metabolism during encystation

Aliphatic polyamines putrescine, spermidine, and spermine, occur ubiquitously and have important functions in the stabilization of cell membranes, biosynthesis of informing molecules, cell growth, and differentiation, as well as adaptation to osmotic, ionic, pH and thermal stress [43,44]. The inhibition of polyamine metabolism has important pharmacological and therapeutic implications for the control of physiological processes, reproduction, cancer, and parasitic diseases [45,46]. Enzymes for the synthesis of longer polyamines, spermidine and spermine, were not identified in the genome of *Entamoeba*. This is puzzling as spermidine is known to be required for the formation of hypusine on the essential initiation factor eIF5A in eukaryotes [47]. This amino acid modification is likely present in *E. histolytica* as deoxyhypusine synthase, the enzyme that uses spermidine to create deoxyhypusine on eIF5A, was identified in the genome [48]. Spermidine is therefore likely to be an essential trace nutrient for *E. histolytica* and therefore spermidine synthesis by *E. histolytica* cannot be ruled out. Our metabolomic data also revealed that beside putrescine, other polyamines like spermidine, spermine, and N^8-acetylspermidine were present in proliferating trophozoites and the level of these metabolites decreased as encystation proceeded. These finding indicate that the amoeba utilizes these polyamines during encystation (Figure 4A). Our result is also consistent with the previous finding that polyamine levels decreased during encystment in *A. castellanii* [49]. The role of N^8-acetylspermidine in polyamine metabolism remains unclear. N^8-acetylspermidine is known to be converted back into spermidine by a deacetylating enzyme [50], suggesting a reversible sequestration of excessive spermidine. The level of N-acetylputrescine, which is derived from putrescine, increased during encystation (Figure 4A). Polyamine acetylation appears to be a component of a cellular mechanism involved in polyamine turnover and excretion [51], which could function as a means of reducing the intracellular polyamine concentration. In addition, storage of the acetylated (metabolically inactive) form of polyamines in the dormant cells (cysts) could provide the newly activated cells with a readily available source of polyamines for growth purposes. Another consequence of polyamine acetylation is to decrease the positive charge, which helps to displace them from anionic binding sites as well as to increase their lipophilic properties that may aid transport processes [51].

Induction of GABA during encystation

Surprisingly, we found that at the later stage of encystation (48 and 120 h) the level of γ-aminobutyric acid (GABA) increased (Figure 4A). GABA is made from L-glutamate in a single reaction catalyzed by the enzyme glutamate decarboxylase, which is missing in the *Entamoeba* genome. However, a number of amino acid decarboxylases are encoded in the genome, and it is possible that some of these decarboxylases convert glutamate to GABA. It has been demonstrated that in other organisms, GABA is also made from an alternative putrescine pathway [52]. The concentration profile of N-acetylputrescine during encystation was similar to that of GABA; the increase of N-acetylputrescine slightly preceded that of GABA (Figure 4A). These data are consistent with the premise that GABA may be synthesized from N-acetylputrescine during encystation. However, further labeling experiments are required to verify the formation of GABA from


A
Arginase
Arginine
Ornithine
AOD
N-acetyl ornithine
OCD
Proline
ADC
ODC
MAO
Agmatine
Agmatinase
DAT
Putrescine
N-Acetylputrescine
N-acetyl-
4-aminobutanol
ADH
N-acetyl-GABA
AOD
AdoMDC
dAdoMet
SPDS
AdoMet
Spermidine
GAD
Glutamate
GABA
MAT
SPS
Methionine
Spermine
AOD
N8-acetyl
spermidine
Transcript decreases
Transcript unchanged
Transcript increases
KEY
Time (h)

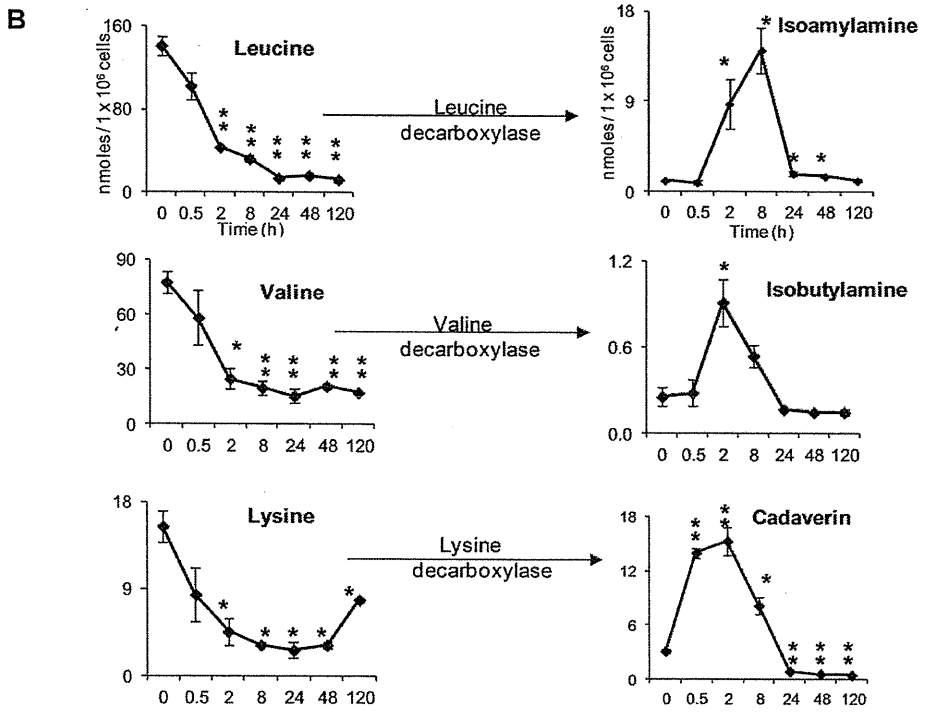

B
Leucine
Leucine
decarboxylase
Isoamylamine
Valine
Valine
decarboxylase
Isobutylamine
Lysine
Lysine
decarboxylase
Cadaverin
nmoles / 1 x 10⁶ cells
Time (h)

Figure 4. Polyamines and biogenic amines levels during encystation. (**A**) Metabolome and transcriptome data of the pathway involved in biosynthesis and catabolism of polyamines and related metabolites from arginine. The average concentrations of each metabolite (nmol/1×10^6 cells) at time intervals (0, 0.5, 2, 8, 24, 48 and 120 h) are shown. The enzymes are shown in yellow or blue in case where their transcript levels increased or decreased by >3 fold relative to time 0 h, respectively. Abbreviations are: GABA, γ-aminobutyric acid; AdoMet, S-Adenosyl methionine; dAdoMet, S-adenosylmethioninamine. Solid lines represent the steps catalyzed by the enzymes whose encoding genes are present in the genome, whereas dashed lines indicate those likely absent in the genome. Abbreviations are: MAO, monoamine oxidase; AdoMDC, S-adenosylmethionine decarboxylase; SPDS, spermidine synthase; SPS, spermine synthase; GAD, glutamate decarboxylase. Data are shown as means ± S.D. Statistical comparisons were made by Student's t test (* $P<0.05$, ** $P<0.01$). (**B**) Level of some biogenic amines and its precursors during encystation. X-axis represents time in hours, whereas Y-axis represents the concentration of metabolites in nmoles/1×10^6 cells. Data are represented as means ± S.D. Statistical comparisons were made by Student's t test (* $P<0.05$, ** $P<0.01$).
doi:10.1371/journal.pone.0037740.g004

either glutamate or putrescine. The role of GABA in *Entamoeba* is unclear. GABA has also been reported in other protozoan parasites like *Plasmodium falciparum* [53]. GABA is the major inhibitory neurotransmitter in mammalian central nervous system. This neurotransmitter has an ancient heritage as an intercellular signal and has been reported to induce terminal differentiation (sporulation) of *Dictyostelium discoideum*, a soil-living amoeba, through a $GABA_B$ receptor [54]. In plants, GABA is thought to play a role in various stress responses, like heat stress, which leads to the six to ten-fold accumulation of GABA compared to an unstressed plant [55], sudden decrease in temperature [56], and water stress [57]. Rapid GABA accumulation in response to wounding was also indicated to play a role in plant defense against insects [58].

Biogenic amines increases during early period of encystation

Biogenic amines are naturally occurring compounds, ubiquitous in animals and plants. They are low molecular weight organic bases, aliphatic (putrescine, cadaverine, isoamylamine, isobutylamine), heterocyclic (histamine and tryptamine), or aromatic (tyramine and phenylethylamine) [59]. These compounds are known to play important roles in normal mammalian physiology, like cell proliferation and differentiation [60]. We also found remarkable changes in biogenic amines such as cadaverine, isoamylamine, and isobutylamine (Figure 4B), which increased during the early period of encystation (0.5 to 8 h), when the trophozoites formed large multicellular aggregates (precyst), and then decreased when the precyst differentiated to the cyst. These three biogenic amines showed distinct kinetics during encystation, suggesting that these biogenic amines could play a specific role in distinct processes of encystation. Biogenic amines play an important role in induction of encystation in *Hartmannella vermiformis*, a nonpathogenic free-living amoeba that is the natural reservoir of *Legionella pneumophila*, the causative agent of Legionnaire's disease [61]. Biogenic amines such as cadaverine, isoamylamine and isobutylamine are the decarboxylated product of amino acids such as lysine, leucine, and valine. It was previously suggested that *E. invadens* encystation and *E. histolytica* cyst-like structure formation are induced by CO_2 [62]. Therefore, it is conceivable that the CO_2, which is released during the formation of biogenic amines, may also participate in inducing the encystation process.

Conclusion and future directions

In summary, our metabolome and transcriptome analyses provided, for the first time in eukaryotes, global changes of metabolisms during a major differentiation process from trophozoites to cysts. Genomic reconstructions, which generally map genomic information of an individual organism onto the metabolic networks of well-studied model organisms, must be informed by direct experimental metabolic evidence. Otherwise, they likely fail to identify the best candidate pathways for drug targets. We found increase in the level of some biogenic amines as well as γ-aminobutyric acid (GABA) during encystation. Finally, further works are needed to characterize key metabolites and their functions responsible for the signal transduction pathways triggering encystation via the formation of GABA and biogenic amines.

Materials and Methods

E. invadens culture and encystation

Trophozoites of the *E. invadens* IP-1 strain were cultured axenically in BI-S-33 medium at 26°C. To induce encystation, 2-week-old *E. invadens* cultures were passaged in 47% LG medium lacking glucose [11] at approximately 6×10^5 cells/ml. Amoebae were collected at various time points, and the formation of cysts was assessed by virtue of the resistance to 0.05% sarkosyl using 0.22% trypan blue to selectively stain dead cells. Cysts were also verified by cyst wall staining by incubating amoebae with calcofluor white (fluorescent brightener; Sigma-Aldrich) at room temperature.

Metabolic extraction

Intracellular metabolites were extracted as described previously with some modifications [63]. Approximately 1.5×10^6 *E. invadens* cells were harvested after 0, 0.5, 2, 8, 24, 48, and 120 h cultivation in 47% LG medium lacking glucose. The cells were immediately suspended in 1.6 ml of −75°C methanol to quench metabolic activity. Given the speed of metabolic reactions, quickly removing medium and immediately quenching metabolism is essential for obtaining reliable results. The selection of cold methanol as an extraction solvent was based on systematic studies in bacteria and yeast, which point to its providing relatively good extraction of a broad spectrum of metabolites [64–65], while avoiding marked metabolite decomposition and associated formation of decomposition products, which can themselves mimic metabolites [66]. To ensure that experimental artifacts such as ion suppression did not lead to misinterpretation of metabolite levels, internal standards, 2-(*N*-morpholino) ethanesulfonic acid, methionine sulfone, and D-camphor- 10-sulfonic acid were added to every sample. The samples were than sonicate for 30 second and then mixed with 1.6 ml of chloroform and 640 µl of deionized water. After vortexing, the mixture was centrifuged at 4,600× *g* at 4°C for 5 min. The aqueous layer (1.6 ml) was filtrated using an Amicon Ultrafree-MC ultrafilter (Millipore Co.) and centrifuged at 9,100× *g* at 4°C for ~2 h. The filtrate was dried and preserved at −80°C until mass spectrometric analysis [67]. Prior to the analysis, the sample was dissolved in 20 µl of deionized water containing reference compounds (200 µM each of 3-aminopyrrolidine and trimesic acid).

Instrumentation of capillary electrophoresis-time-of-flight mass spectrometry (CE-TOFMS)

CE-TOFMS was performed using an Agilent CE Capillary Electrophoresis System equipped with an Agilent 6210 time-of-flight mass spectrometer, Agilent 1100 isocratic HPLC pump, Agilent G1603A CE-MS adapter kit, and Agilent G1607A CE-ESI-MS sprayer kit (Agilent Technologies, Waldbronn, Germany). The system was controlled by Agilent G2201AA ChemStation software for CE. Data acquisition was performed by Analyst QS software for Agilent TOF (Applied Biosystems, CA, USA;MDS Sciex, Ontario, Canada).

CE-TOFMS conditions for cationic metabolite analysis

Cationic metabolites were separated in a fused-silica capillary (50 μm i.d.×100 cm total length) filled with 1 M formic acid as the electrolyte [68]. Sample solution (~3 nL) was injected at 50 mbar for 3 s, and a positive voltage of 30 kV was applied. The capillary and sample trays were maintained at 20°C and below 5°C, respectively. Sheath liquid composed of methanol/water (50% v/v) that contained 0.1 μM hexakis (2,2-difluorothoxy) phosphazene was delivered at 10 μL/min. ESI-TOFMS was operated in the positive ion mode. The capillary voltage was set at 4 kV and a flow rate of nitrogen gas (heater temperature 300°C) was set at 10 psig. For TOFMS, the fragmenter voltage, skimmer voltage, and octapole radio frequency voltage (Oct RFV) were set at 75, 50, and 125 V, respectively. An automatic recalibration function was performed using two reference masses of reference standards; protonated ^{13}C methanol dimer (m/z 66.063061) and protonated hexakis (2,2-difluorothoxy) phosphazene (m/z 622.028963), which provided the lock mass for exact mass measurements. Exact mass data were acquired at the rate of 1.5 cycles/s over a 50 to 1,000 m/z range.

CE-TOFMS conditions for anionic metabolite analysis

Anionic metabolites were separated in a cationic-polymer–coated COSMO(+) capillary (50 μm i.d.×110 cm) (Nacalai Tesque) filled with 50 mmol/L ammonium acetate solution (pH 8.5) as the electrolyte [69]. Sample solution (~30 nL) was injected at 50 mbar for 30 s and a negative voltage of −30 kV was applied. Ammonium acetate (5 mmol/L) in 50% methanol/water (50% v/v) that contained 0.1 μmol/L hexakis (2,2-difluorothoxy) phosphazene was delivered as sheath liquid at 10 μL/min. ESI-TOFMS was operated in the negative ion mode. The capillary voltage was set at 3.5 kV. For TOFMS, the fragmenter voltage, skimmer voltage, and Oct RFV were set at 100, 50, and 200 V, respectively [70]. An automatic recalibration function was performed using two reference masses of reference standards: deprotonated ^{13}C acetate dimer (m/z 120.038339) and acetate adduct of hexakis (2,2-difluorothoxy) phosphazene (m/z 680.035541). The other conditions were identical to those used for the cationic metabolome analysis.

CE-TOFMS data processing

Raw data were processed using the in-house software Masterhands [71]. The overall data processing flow consisted of the following steps: noise-filtering, baseline-removal, migration time correction, peak detection, and integration of peak area from a 0.02 m/z-wide slice of the electropherograms. This process resembled the strategies employed in widely used data processing software for LC-MS and GC-MS data analysis, such as MassHunter (Agilent Technologies) and XCMS [72]. Subsequently, accurate m/z values for each peak were calculated by Gaussian curve fitting in the m/z domain, and migration times were normalized using alignment algorithms based on dynamic programming [73]. All target metabolites were identified by matching their m/z values and normalized migration times with those of standard compounds in the in-house library. All data presented were normalized by cell number (per 10^6 cells).

RNA isolation and affymetrix microarray hybridization

E. invadens cells were harvested after 0, 0.5, 2, 8, 24, 48 and 120 h cultivation in 47% LG medium lacking glucose and washed twice with phosphate buffer saline. Total RNA was isolated from harvested trophozoites using trizol reagent (Invitrogen, Carlsbad, CA, U.S.A.) according to the manufacturer's protocol. The RNA was quantified and checked for purity by comparison of absorbance at 260 and 280 nm in the nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA). Integrity of isolated RNA was verified by using Bio-Rad's automated electrophoresis system experion (RNA StdSens analysis kit). All reagents and protocols followed those described in the affymetrix manuals. Briefly, total RNA (5 μg) was reverse transcribed using T7-Oligo (dT) primer in the first strand cDNA synthesis. After second strand synthesis, the double-stranded cDNA template was used for in vitro transcription, in the presence of biotinylated nucleotides to produce labelled cRNA. The cRNA was purified, quantified, fragmented, and hybridized for 16 h at 45°C to custom-generated affymetrix platform microarray (49-7875) with probe sets consisting of 11 probe pairs representing 9,327 *E. histolytica* (Eh_Eia520620F_Eh) and 12,385 *E. invadens* open reading frames (Eh_Eia520620F_Ei). After hybridization, the arrays were washed and stained with streptavidin–phycoerythrin using a GeneChip® Fluidics Station 450 (Affymetrix, Santa Clara, CA, USA), according to the recommendations of the manufacturer. After washing and staining, the GeneChip® arrays were then scanned using the Hewlett-Packard Affymetrix Scanner 3000 (Affymetrix, Santa Clara, CA, USA), and the probe intensities were extracted using Affymetrix® GeneChip® Command Console™ (Affymetrix, Santa Clara, CA, USA).

Analysis of microarray data

A minimum of two arrays were used for each condition and each time point. Raw Mas5 gene expression data were imported into the GeneSpring GX 10.0.2 program and normalized expression values for each probe set were obtained from raw probe intensities in R 2.7.0 downloaded from the BioConductor project (http://www.bioconductor.org) using robust multiarray averaging with correction for oligosequence (gcRMA). Standard correlation coefficients were calculated using GeneSpring GX 10.0.2 [74]. The data presented in this publication have been deposited in NCBIs Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE33312.

Statistical Analysis

For each experimental condition, two independent biological replicates were made and for each biological replicate, three technical replicates were made. All data are shown as means ± S.D. for the indicated number of experiments. Statistical comparisons were made by Student's t test. Heat maps of metabolite levels were generated using hierarchical clustering based on Pearson correlation coefficients using the MultiExperiment Viewer (MeV) software (Institute for Genomic Research [20]). The Pearson correlation coefficients and their P-values (two-tailed) were calculated using GraphPad Prism version 5.04 (GraphPad Software, Inc., San Diego, CA). For microarray data one way ANOVA analysis with Tukey's post hoc test was

performed to extract differentially expressed genes. The P-values were calculated using Welch's test, and were corrected by Benjamini-Hochberg method (Dataset S2).

Supporting Information

Dataset S1 The quantitative metabolites data during encystation. Normalized average metabolites data of two biological replicate at 0, 0.5, 2, 8, 24, 48 and 120 h are shown. SD, standard deviation of means; N.D, the metabolite concentrations were below the detection limit of the analysis. All the P values were evaluated by the student's t-test.
(XLSX)

Dataset S2 Transcriptomic data analyzed in this study.
(XLSX)

Figure S1 Heat map representation of microarray expression data of genes involved in glycolysis, serine/cysteine, chitin, glycogen, and polyamine pathways, with the fold change in expression relative to time 0 h at each time point being indicated by different colors boxes. The color scale is also shown below.
(TIF)

Acknowledgments

We thank Kumiko Nakada-Tsukui and all other members of our laboratory for the technical assistance and valuable discussions. We would also like to thank Akiyoshi Hirayama (Keio University) for the technical assistance of CE-MS.

Author Contributions

Conceived and designed the experiments: GJ DS TN. Performed the experiments: GJ AEdC DS. Analyzed the data: GJ M. Sugimoto AH. Contributed reagents/materials/analysis tools: TS M. Suematsu TN. Wrote the paper: GJ TN.

References

1. Barrón-González MP, Villarreal-Treviño L, Reséndez-Pérez D, Mata-Cárdenas BD, Morales-Vallarta MR (2008) *Entamoeba histolytica*: Cyst-like structures in vitro induction. Exp Parasitol 118: 600–603.
2. Lauwaet T, Davids BJ, Reiner DS, Gillin FD (2007) Encystation of *Giardia lamblia*: a model for other parasites. Curr Opin Microbiol 10: 554–559.
3. Siddiqui R, Jarroll EL, Khan NA (2010) *Balamuthia mandrillaris*: Role of galactose in encystment and identification of potential inhibitory targets. Exp Parasitol 126: 22–27.
4. Chagla AH, Griffiths AJ (1974) Growth and encystation of *Acanthamoeba castellanii*. J Gen Microbiol 85: 139–145.
5. Eichinger D (2001) Encystation in parasitic protozoa. Curr Opin Microbiol 4: 421–426.
6. Ximénez C, Morán P, Rojas L, Valadez A, Gómez A (2009) Reassessment of the epidemiology of amebiasis: state of the art. Infect Gen & Evol 9: 1023–1032.
7. Meerovitch E (1961) Infectivity and pathogenicity of polyxenic and monoxenic *Entamoeba invadens* to snakes kept at normal and high temperatures and the natural history of reptile amoebiasis. J Parasitol 47: 791–794.
8. McConnachie EW (1955) Studies on *Entamoeba invadens* Rodhain, 1934, in vitro, and its relationship to some other species of *Entamoeba*. Parasitology 45: 452–481.
9. Vazquezdelara-Cisneros LG, Arroyo-Begovich A (1984) Induction of encystation of *Entamoeba invadens* by removal of glucose from the culture medium. J Parasitol 70: 629–633.
10. Avron B, Stolarsky T, Chayen A, Mirelman D (1986) Encystation of *Entamoeba invadens* IP-1 is induced by lowering the osmotic pressure and depletion of nutrients from the medium. J Protozool 33: 522–525.
11. Sanchez L, Enea V, Eichinger D (1994) Identification of a developmentally regulated transcript expressed during encystation of *Entamoeba invadens*. Mol Biochem Parasitol 67: 125–135.
12. Schnackenberg LK, Beger RD (2006) Monitoring the health to disease continuum with global metabolic profiling and systems biology. Pharmacogenomics 7: 1077–1086.
13. Lawton KA, Berger A, Mitchell M, Milgram KE, Evans AM, et al. (2008) Analysis of the adult human plasma metabolome. Pharmacogenomics 9: 383–397.
14. Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, et al. (2009) Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. Nature 457: 910–914.
15. Boudonck KJ, Rose DJ, Karoly ED, Lee DP, Lawton KA, et al. (2009) Metabolomics for early detection of drug-induced kidney injury: review of the current status. Bioanalysis 1: 1645–1663.
16. Soga T, Baran R, Suematsu M, Ueno Y, Ikeda S, et al. (2006) Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. J Biol Chem 281: 16768–16776.
17. Bölling C, Fiehn O (2005) Metabolite profiling of *Chlamydomonas reinhardtii* under nutrient deprivation. Plant Physiol 139: 1995–2005.
18. Jozefczuk S, Klie S, Catchpole G, Szymanski J, Cuadros-Inostroza A, et al. (2010) Metabolomic and transcriptomic stress response of *E. coli*. Mol Sys Biol 6: 364.
19. Bruckner DA (1992) Amebiasis. Clin Microbiol Rev 5: 356–369.
20. Saeed AI, Sharov V, White J, Li J, Liang W, et al. (2003) TM4: a free, open source system for microarray data management and analysis. Biotechniques 34: 374–378.
21. Saavedra E, Encalada R, Pineda E, Jasso-Chávez R, Moreno-Sánchez R (2005) Glycolysis in *Entamoeba histolytica*. Biochemical characterization of recombinant glycolytic enzymes and flux control analysis. FEBS J 272: 1767–1783.
22. Reeves RE (1984) Metabolism of *Entamoeba histolytica* Schaudinn, 1903. Adv Parasitol 23: 105–142.
23. McLaughlin J, Aley S (1985) The biochemistry and functional morphology of the *Entamoeba*. J Protozool 32: 221–240.
24. Chávez-Munguía B, Omaña-Molina M, González-Lázaro M, González-Robles A, Cedillo-Rivera R, et al. (2007) Ultrastructure of cyst differentiation in parasitic protozoa. Parasitol Res 100: 1169–1175.
25. Arroyo-Begovich A, Cárabez-Trejo A (1982) Location of chitin in the cyst wall of *Entamoeba invadens* with the colloidal gold tracers. J Parasitol 68: 253–258.
26. Ward HD, Alroy J, Lev BI, Keusch GT, Pereira ME (1985) Identification of chitin as a structural component of *Giardia* cysts. Infect Immun 49: 629–634.
27. Siddiqui R, Khan NA, Jarroll EL (2009) The cyst wall carbohydrate composition of *Balamuthia mandrillaris*. Parasitol Res 104: 1439–1443.
28. Chávez-Munguía B, Omaña-Molina M, González-Lázaro M, González-Robles A, Bonilla P, et al. (2005) Ultrastructural study of encystation and excystation in *Acanthamoeba castellanii*. J Eukaryot Microbiol 52: 153–158.
29. Spindler KD, Spindler-Barth M, Londershausen M (1990) Chitin metabolism: a target for drugs against parasites. Parasitol Res 76: 283–288.
30. Badet B, Vermoote P, Haumont P, Lederer F, Le Goffic F (1987) Glucosamine synthetase from *Escherichia coli* : purification, properties, and glutamine-utilizing site location. Biochemistry 26: 1940–1948.
31. Zhou J, Huynh QK, Hoffman TR, Crook DE, Danniels CM, et al. (1998) Regulation of glutamine:fructose-6-phosphate amidotransferase by cAMP dependent protein kinase. Diabetes 47: 1836–1840.
32. Badet-Denisot MA, Rene L, Badet B (1993) Mechanistic investigations on glucosamine-6-phosphate synthase. Bull Soc Chim Fr 130: 249–255.
33. Tsigos I, Martinou A, Kafetzopoulos D, Bouriotis V (2002) Chitin deacetylases: new, versatile tools in biotechnology. Trends Biotechnol 18: 305–312.
34. Struszczyk, H (1987) Microcrystalline chitosan. I. Properties and preparation. J Appl Polym Sci 33: 177–187.
35. Muller M (1991) Energy metabolism of anaerobic parasitic protists. In: Biochemical Protozoology Coombs GH, North MJ, eds. pp 80–91. Taylor and Francis, London.
36. Edwards MR, Gilroy FV, Jimenez MB, O'Sullivan WJ (1989) Alanine is a major product of metabolism by *Giardia lamblia*: a proton magnetic resonance study. Mol Biochem Parasitol 37: 19–26.
37. Chapman A, Linstead DJ, Lloyd D, Williams J (1985) 13C-NMR reveals glycerol as an unexpected major metabolite of the protozoan parasite *Trichomonas vaginalis*. FEBS Lett 191: 287–292.
38. Ali V, Nozaki T (2007) Current therapeutics, their problems, and sulfur-containing amino acid metabolism as a novel target against infections by "amitochondriate" protozoan parasites. Clin Microbiol Rev 20: 164–187.
39. Zuo X, Coombs GH (1995) Amino acid consumption by the parasitic, amoeboid protists *Entamoeba histolytica* and *E. invadens*. FEMS Microbiol Lett 130: 253–258.
40. Atkinson DE (1969) Regulation of enzyme function. Annu Rev Microbiol 23: 47–68.
41. Chapman AG, Fall L, Atkinson DE (1971) Adenylate energy charge in *Escherichia coli* during growth and starvation. J. Bacteriol 108: 1072–1086.
42. Buckstein MH, He J, Rubin H (2008) Characterization of nucleotide pools as a function of physiological state in *Escherichia coli*. J Bacteriol 190: 718–726.
43. Slocum RD, Kaur-Sawhney R, Galston AW (1984) The physiology and biochemistry of polyamines in plants. Arch Biochem Biophys 235: 283–303.
44. Tabor CW, Tabor H (1985) Polyamines in microorganisms. Microbiol Rev 49: 81–99.
45. Sjoerdsma A, Schechter PJ (1984) Chemotherapeutic implications of polyamine biosynthesis inhibitors. Clin Pharmacol Therapy 35: 287–300.
46. Shukla OP, Kishore P, Gupta S (1990) Polyamine metabolism of protozoan *Acanthamoeba culbertsoni* as a target for chemotherapy. Biol Membranes 16: 6–17.

47. Park MH, Lee YB, Joe YA (1997) Hypusine is essential for eukaryotic cell proliferation. Biological Signals 6: 115–123.
48. Loftus B, Anderson I, Davies R, Alsmark UC, Samuelson J, et al. (2005) The genome of the protist parasite *Entamoeba histolytica*. Nature 433: 865–868.
49. Zhu CM, Cumaraswamy A, Henney HR, Jr. (1989) Comparison of polyamine and S-adenosylmethionine contents of growing and encysted *Acanthamoeba* isolates. Mol Cell Biochem 90: 145–153.
50. Blankenship J, Marchant PE (1984) Metabolism of N1-acetylspermidine and N8-acetylspermidine in rats. Proc Soc Biol Med 177: 180–187.
51. Seiler N (1987) Functions of polyamine acetylation. Can J Physiol Pharmacol 65: 2024–2035.
52. Jakoby WB, Fredericks J (1959) Pyrrolidine and putrescine metabolism: gamma-aminobutyraldehyde dehydrogenase. J Biol Chem 234: 2145–2150.
53. Teng R, Junankar PR, Bubb WA, Rae C, Mercier P, et al. (2009) Metabolite profiling of the intraerythrocytic malaria parasite *Plasmodium falciparum* by (1)H NMR spectroscopy. NMR Biomed 22: 292–302.
54. Anjard C, Loomis WF (2006) GABA induces terminal differentiation of *Dictyostelium* through a $GABA_B$ receptor. Development 133: 2253–2261.
55. Mayer RR, Cherry JH, Rhodes D (1990) Effects of heat shock on amino acid metabolism of cowpea cells. Plant Physiol 94: 796–810.
56. Wallace W, Secor J, Schrader LE (1984) Rapid accumulation of gamma-aminobutyric acid and alanine in soybean leaves in response to an abrupt transfer to lower temperature, darkness, or mechanical manipulation. Plant Physiol 75: 170–175.
57. Rhodes D, Handa S, Bressan RA (1986) Metabolic changes associated with adaptation of plant cells to water stress. Plant Physiol 82: 890–903.
58. Ramputh A-I, Bown AW (1996) Rapid gama-aminobutyric acid synthesis and the inhibition of the growth and development of oblique-banded leaf-roller larvae. Plant Physiol 111: 1349–1352.
59. Lounvaud-Funel A (2001) Biogenic amines in wines: Role of lactic acid bacteria. FEMS Microbiol Lett 199: 9–13.
60. Bauza T, Blaise A, Teissedre PL, Cabanis JC, Kanny G, et al. (1995) Les amines biogenes du vin: Metabolisme et toxicite. Bullettin de L'OIV. pp 42–67.
61. Verma AK, Raizada MK, Murti CK (1974) Effect of bioamines on the cellular differentiation of *Hartmannella culbertsoni*. Biochem Pharmacol 23: 57–63.
62. Morales-Vallarta M, Villarreal-Treviño L, Guerrero Medrano L, Ramírez-Bon E, et al. (1997) *Entamoeba invadens* differentiation and *E. histolytica* cyst-like formation induced by CO_2. Arch Med Res 28: 150–151.
63. Husain A, Sato D, Jeelani G, Mi-ichi F, Ali V, et al. (2010) Metabolome analysis revealed increase in S-methylcysteine and phosphatidyl isopropanolamine synthesis upon L-cysteine deprivation in the anaerobic protozoan parasite *Entamoeba histolytica*. J Biol Chem 285: 39160–39170.
64. Ohashi Y, Hirayama A, Ishikawa T, Nakamura S, Shimizu K, et al. (2008) Depiction of metabolome changes in histidine-starved *Escherichia coli* by CE-TOFMS. Mol Biosyst 4: 135–147.
65. Maharjan RP, Ferenci T (2003) Global metabolite analysis: The influence of extraction methodology on metabolome profiles of *Escherichia coli*. Anal Biochem 313: 145–154.
66. Villas-Boas SG, Hojer-Pedersen J, Akesson M, Smedsgaard J, Nielsen J (2005) Global metabolite analysis of yeast: Evaluation of sample preparation methods. Yeast 22: 1155–1169.
67. Kimball E, Rabinowitz JD (2006) Identifying decomposition products in extracts of cellular metabolites. Anal Biochem 358: 273–280.
68. Soga T, Heiger, DN (2000) Amino acid analysis by capillary electrophoresis electrospray ionization mass spectrometry. Anal Chem 72: 1236–1241.
69. Soga T, Ueno Y, Naraoka H, Ohashi Y, Tomita M, et al. (2002) Simultaneous determination of anionic intermediates for *Bacillus subtilis* metabolic pathways by capillary electrophoresis electrospray ionization mass spectrometry. Anal Chem 74: 2233–2239.
70. Soga T, Igarashi K, Ito C, Mizobuchi K, Zimmermann HP, et al. (2009) Metabolomic profiling of anionic metabolites by capillary electrophoresis mass spectrometry Anal Chem 81: 6165–6174.
71. Sugimoto M, Wong DT, Hirayama A, Soga T, Tomita M (2010) Capillary electrophoresis mass spectrometry-based saliva metabolomics identified oral, breast and pancreatic cancer-specific profiles. Metabolomics 6: 78–95.
72. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem 78: 779–787.
73. Baran R, Kochi H, Saito N, Suematsu M, Soga T, et al. (2006) MathDAMP: a package for differential analysis of metabolite profiles. BMC Bioinformatics 7: 530.
74. Husain A, Jeelani G, Sato D, Nozaki T (2011) Global analysis of gene expression in response to L-cysteine deprivation in the anaerobic protozoan parasite *Entamoeba histolytica*. BMC Genomics 12: 275.

Biochimie 94 (2012) 1676–1686

Contents lists available at SciVerse ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Research paper

Novel protein–protein interactions between *Entamoeba histolytica* D-phosphoglycerate dehydrogenase and phosphoserine aminotransferase

Vibhor Mishra [a,*], Ashutosh Kumar [b], Vahab Ali [c], Tomoyoshi Nozaki [d], Kam Y.J. Zhang [b], Vinod Bhakuni [a]

[a] *Division of Molecular and Structural Biology, CSIR-Central Drug Research Institute, MG Marg, Lucknow 226001, India*
[b] *Zhang Initiative Research Unit, Advanced Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan*
[c] *Department of Biochemistry, Rajendra Memorial Research Institute of Medical Sciences, Agam Kuan, Patna 800007, India*
[d] *Department of Parasitology, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-Ku, Tokyo 162-8640, Japan*

ARTICLE INFO

Article history:
Received 10 August 2011
Accepted 18 February 2012
Available online 24 February 2012

Keywords:
Protein–protein interaction
Entamoeba histolytica
D-phosphoglycerate dehydrogenase
Phosphoserine aminotransferase

ABSTRACT

Physical interactions between D-phosphoglycerate dehydrogenase (EhPGDH) and phosphoserine aminotransferase (EhPSAT) from an enteric human parasite *Entamoeba histolytica* was observed by pull-down assay, gel filtration chromatography, chemical cross-linking, emission anisotropy, molecular docking and molecular dynamic simulations. The protein–protein complex had a 1:1 stochiometry with a dissociation constant of 3.453×10^{-7} M. Ionic interactions play a significant role in complex formation and stability. Analysis of the energy minimized average simulated model of the protein complex show that the nucleotide binding domain of EhPGDH specifically interacts with EhPSAT. Denaturation studies suggest that the nucleotide binding domain (Nbd) and substrate binding domain (Sbd) of EhPGDH are independent folding/unfolding units. Thus the Nbd-EhPGDH was separately cloned over-expressed and purified to homogeneity. Fluorescence anisotropy study show that the purified Nbd interacts with EhPSAT. Forward enzyme catalyzed reaction for the EhPGDH-PSAT complex showed efficient Km values for 3-phosphoglyceric acid as compared to only EhPGDH suggesting a possibility of substrate channelling in the protein complex.

1. Introduction

Proteins often carry out their cellular functions through their concerted interactions with other proteins. Such protein–protein interactions can be classified as permanent or non-obligatory depending on the ability of the proteins to exist independently [1,2].

In the last twenty years it has been demonstrated that metabolic pathways are often orchestrated by sequential enzymes which operate in perfect synchrony by interacting with each other in the form of highly organized complexes [1]. The term "metabolon" was introduced by Paul Srere to describe such enzyme–enzyme complexes [2,3]. Such molecular complexes provide highly efficient, regulated and targeted metabolite flux through a particular pathway [4]. These supramolecular complexes have been extensively characterized in the mitochondrion, especially between the sequential enzymes of the tricarboxylic acid (TCA) cycle [5–8].

Protists possess the most simplified versions of metabolic pathways which involves a very limited set of enzymes. Over the past decade serine and its related metabolism has been extensively explored in an enteric human parasite *Entamoeba histolytica* [9]. Phosphorylated serine biosynthesis pathway serves as a major source of L-serine flux in *E. histolytica*. D-phosphoglycerate dehydrogenase (EhPGDH) and phosphoserine aminotransferase (EhPSAT) catalyze the first two and only reversible steps of the pathway. *E. histolytica* possess a unique D-Phosphoglycerate dehydrogenase (EhPGDH) which is a homodimer of 70 kDa with two domains the substrate binding domain (Sbd) and the nucleotide binding domain (Nbd). EhPGDH lacks the Trp-139 residue which is considered to be crucial for the tetramerization of the protein [9,10]. Most interestingly the protein is insensitive to feedback inhibition by L-serine because of the absence of a regulatory domain, a complete contrasting feature from its counterparts in *E. coli* and *Mycobacterium tuberculosis* [9].

Abbreviations: SEC, size exclusion chromatography; ESI-MS, electro spray ionization mass spectroscopy; EhPGDH, *Entamoeba histolytica* D-phosphoglycerate dehydrogenase; EhPSAT, *Entamoeba histolytica* phosphoserine aminotransferase.

* Corresponding author. Tel.: +91 0 522 2232955; fax: +91 0 522 223405/223938.
E-mail addresses: vibhor6strings@rediffmail.com, mishravibhor@gmail.com (V. Mishra).

0300-9084/$ – see front matter © 2012 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.biochi.2012.02.028

In the present study we have investigated unique structural features of EhPGDH especially novel interactions between EhPGDH and EhPSAT using *in vitro* protein–protein interaction techniques. Furthermore, for understanding the role of the two domains of EhPGDH in protein complex formation, interaction studies were performed with isolated Nbd. Our results provide an evidence for substrate channelling between EhPGDH and EhPSAT. To the best of our knowledge this is a first report of novel interactions between these two proteins of phosphorylated serine biosynthesis pathway.

2. Materials and methods

2.1. Materials

All chemicals used in the study were purchased from Sigma–Aldrich chemical company St Louis, USA. All chromatography columns/matrices were purchased from GE Healthcare Biosciences. Ni-NTA agarose was purchased from Qiagen. Milli Q water was used for all buffer preparations.

2.2. Methods

2.2.1. Over-expression and purification of EhPGDH and EhPSAT

The proteins, EhPGDH-His_6 and EhPSAT-His_6 were over-expressed and purified by previously described methods [9,11,12].

2.2.2. Cloning, over-expression and purification of GST-EhPGDH

EhPGDH protein encoded by an 897 base pair gene fragment was PCR amplified from EhPGDH gene harboured in a pET15b vector (pET15b-EhPGDH). PCR was performed with primers (forward-5′-GGATCCAAGATAGTTGTGATAACCGA-3′ and reverse-5′-CTCGAGTTAGAACTTATTGACTTGGAA-3′) with BamHI and XhoI restriction sites. PCR conditions used were: 1 × 94 °C for 5 min, 30 × 94 °C for 45 s, 55 °C for 45 s and 72 °C for 1 min and finally 1 × 72 °C for 10 min. The amplified fragment was cloned in pGEX-4T1 (GE healthcare life sciences) vector between BamHI and XhoI restriction sites in the same orientation as the LacZ promoter to produce GST-EhPGDH. The resultant construct was transformed into *E coli* BL-21 competent cell for expression. A single colony of BL-21 LB ampicillin plate was inoculated into 5 mL of LB broth (HI media) having ampicillin concentration of 100 μg/mL and was allowed to grow overnight at 37 °C. It was then sub-cultured in a 500 mL LB broth at 37 °C until A_{600} of 0.6 was achieved. The culture was then induced with 1 mM of isopropyl-1-thio-β–D-galactopyranoside and was further incubated at 20 °C for 16 h. The cells were harvested at 8000 rpm for 10 min and the resultant pellet was suspended in 50 mM potassium phosphate (pH 8) buffer containing 300 mM NaCl, 2 mM PMSF and 10% glycerol. Cells were disrupted using a probe type sonicator and centrifuged at 13,500 rpm for 30 min. The supernatant was applied on glutathione sepharose-4B (GE healthcare life sciences) affinity column pre-equilibrated with 50 mM potassium phosphate (pH 8) buffer along with 300 mM NaCl. The column was subsequently washed with two column volumes of the same buffer. The protein was eluted using 20 mM reduced glutathione in the same buffer. The eluted protein was extensively dialyzed at 4 °C in 50 mM potassium phosphate pH 7.5 and was stored with 10% glycerol at 4 °C. Finally, the eluted protein tested for purity by SDS-PAGE and ESI-MS was found to be > 95% pure. The purified protein existed in a properly folded conformation as observed by far-UV CD spectrum and aromatic amino acid fluorescence emission spectrum (supplementary Fig. 1).

2.2.3. Cloning, over-expression and purification of nucleotide binding domain (Nbd) of EhPGDH

The Nucleotide binding domain (Nbd) (Arg^{103}–Gly^{265}) of EhPGDH encoded by a 162 base pair gene fragment was PCR amplified from the full length gene of EhPGDH (pET15b-EhPGDH). PCR was performed with primers (forward-5′-GCTAGCATGAGGAATGGTGTCG-CAGAACTTTGT-3′ and reverse-5′-AAGCTT TCCTATTTTAATTGGTGTTG-CAAAGAATCTTCC-3′) with NheI and HindIII restriction sites. PCR conditions used were: 1 × 94 °C for 5 min, 30 × 94 °C for 45 s, 51 °C for 45 s and 72 °C for 2 min and finally 1 × 72 °C for 10 min. The amplified fragment was cloned in pTZ57R/T (InsTAclone™ PCR cloning kit, Fermentas) cloning vector. The resultant construct was subsequently double digested with respective restriction enzymes and the obtained DNA fragment was further sub-cloned into pET23a (+) vector (Novagen) between NheI and HindIII sites. The resultant construct was transformed into *E. coli* Rosetta™ (DE3) pLysS competent cell for expression. A single colony of LB ampicillin plate was inoculated into 5 mL of LB broth (HI media) having ampicillin at a concentration of 100 μg/mL and allowed to grow overnight at 37 °C. It was then sub-cultured in a 500 mL LB broth at 37 °C until A_{600} of 0.6 was achieved. The culture was then induced with 1 mM of isopropyl-1-thio-β-D-galactopyranoside and was further incubated at 20 °C for 7 h. The cells were harvested at 8000 rpm for 10 min and the resultant pellet was suspended in 50 mM potassium phosphate (pH 8) buffer containing 300 mM NaCl, 2 mM PMSF, 5 mM imidazole, 1 mM DTT and 10% glycerol. Cells were disrupted using a probe type sonicator and centrifuged at 13,500 rpm for 30 min. The supernatant was applied on nickel nitrilotriacetic acid (Ni-NTA) agarose affinity column pre-equilibrated with 50 mM potassium phosphate (pH 8) buffer along with 300 mM NaCl and 5 mM imidazole. The column was subsequently washed with two column volumes of the same buffer containing 5 mM and one column volume each of 20 mM and 40 mM imidazole, respectively. The protein was eluted using 500 mM imidazole. The eluted protein sample tested for purity by SDS-PAGE and ESI-MS was found to be > 95% pure.

2.2.4. Spectroscopy

Fluorescence spectra were recorded on Perkin Elmer LS50b luminescence spectrometer at 25 °C. For aromatic amino acid fluorescence an excitation wavelength of 280 nm was used and emission was recorded between 300 nm and 400 nm. Tryptophan fluorescence anisotropy for the protein–protein complex was measured at a fixed excitation and emission wavelengths of 295 nm and 335 nm, respectively.

CD experiments were performed with JASCO J810 spectropolarimeter (equipped with peltier temperature controller system) in a 0.2 cm cell at 25 °C. The far-UV CD spectra were measured at a protein concentration of 3 μM for EhPGDH-His_6, GST-EhPGDH and Nbd, respectively. Readings obtained were normalized by subtracting the baseline observed for the buffer.

Coupled enzyme catalyzed reactions of protein–protein complex were assayed as described earlier for the single EhPGDH protein in phosphate buffer (pH 7.5) on Shimadzu UV 1650PC spectrophotometer at 25 °C [9]. 5 μg of EhPGDH (single protein) or EhPGDH-PSAT complex was used to measure the enzyme catalyzed reactions.

2.2.5. Pull-down assay

Purified GST-EhPGDH and EhPSAT-His_6 were incubated for 30 min in 0.5 mL reaction mixture at 4 °C, for complex formation in phosphate buffer (pH 7.5). Mice anti-GST antibody (Sigma) was then added into the milieu and the reaction was further incubated for 30 min. After addition of 5 mg of protein A-Sepharose beads (Sigma), the proteins were incubated for an additional 1 h. The immune-complex was precipitated and washed with ice-cold

reaction buffer three times, suspended in loading buffer, and resolved on 12% SDS-PAGE.

2.2.6. Size exclusion chromatography (SEC)

Gel filtration experiments were carried out on a Superdex™ 200 and Superdex™ 75 10/300GL columns on AKTA FPLC (GE Health care) pre-calibrated with standard molecular weight markers. The columns were pre-equilibrated and run with 50 mM potassium phosphate buffer pH 7.5 at 25 °C at a flow rate of 0.3 mL/min, with detection at 280 nm and 340 nm for EhPGDH and Nbd, respectively. For protein–protein complex equimolar amounts of EhPGDH and EhPSAT were incubated for 30 min for complex formation. Subsequently, gel filtration experiments were carried out on a Superdex™ 200 10/300GL column on AKTA FPLC (GE Health care) pre-calibrated with standard molecular weight markers. The columns were pre-equilibrated and run with 50 mM potassium phosphate buffer pH 7.5 at a flow rate of 0.05 mL/min, with detection at 280 nm.

2.2.7. Glutaraldehyde cross-linking

The cross-linking of native EhPGDH, Nbd and the EhPGDH-PSAT complex was carried out by method described earlier and analyzed on 12% SDS-PAGE [11]. For the protein–protein complex the cross-linked samples were analyzed on 8% SDS-PAGE.

2.2.8. Molecular docking

The homology models of EhPGDH and EhPSAT were generated with the help Modeller program [13]. These models were used as starting structures for computational docking experiments. The model of the bound protein–protein complex structure from its unbound components was generated in a two-step docking procedure. Initially, Patchdock [14] was used to perform a global search to specifically predict protein–protein interaction site. Protein–protein interaction information was then used to predict EhPGDH-PSAT complex using RosettaDock [15]. In the first stage of the algorithm, RosettaDock employs a rigid-body Monte Carlo search, translating and rotating one docking partner around the surface of the other, using residue-scale interaction potentials [15–17]. After this low-resolution search, explicit side chains are added to the protein backbones using a backbone-dependent rotamer packing algorithm [18]. A Monte Carlo-plus-minimization scheme then efficiently samples a set of local minima in a small region of docking conformation space by simultaneously optimizing the side chain conformations and the rigid-body position. The search procedure is repeated from different random starting orientations to create $o(10)$ structures, which are then ranked using an energy function dominated by van der Waals interactions, an implicit solvation model and an orientation-dependent hydrogen bonding potential [19,20]. In total, 1000 protein–protein complex decoys were generated at the end of the high-resolution search. These 1000 protein–protein decoys were then clustered on the basis of pair-wise rmsd using a Rosetta's clustering algorithm. The lowest energy structure in the cluster with the most population was selected as the final docking complex to run the MD simulations.

2.2.9. Molecular dynamics simulation

A 10 ns Molecular dynamics (MD) simulation was carried out for the lowest energy structure of most populous cluster produced by RosettaDock. The complex was immersed in the rectangular truncated octahedron filled with 12 Å TIP3P water molecules and neutralized by adding 12 Cl^- ions. The system was minimized by 500 steps of steepest descent followed by 2000 steps of conjugate gradient. After the minimization, the system was gradually heated in the canonical ensemble from 0 to 300 K over 50 ps and additional MD simulation was performed for 500 ps for equilibration. Finally, a 10 ns MD simulation was performed under a constant temperature of 300 K. Partical Mesh Ewald (PME) was employed to deal with the long-range electrostatic interactions [21]. SHAKE procedure was employed to constrain all hydrogen atoms and the time step was set to 2 fs [22]. Trajectory snapshots were taken at each 1 ps, which were finally used for analysis. All simulations were performed by the SANDER module of Amber10 [23] with amber forcefield (ff03). Analysis was performed Ptraj module of Amber tools [23]. A Similar methodology was used to generate an energy minimized average model of individual EhPGDH protein after 10 ns molecular dynamic simulations.

2.2.10. GdnHCl and thermal denaturation

3 μM proteins (EhPGDH and Nbd) was dissolved in potassium phosphate buffer (50 mM, pH 7.5) in presence of increasing concentration of GdnHCl and incubated overnight at 25 °C before measurements were made. Under all conditions studied overnight incubation time at different denaturant concentrations was sufficient for the reaction to achieve equilibrium. Far-UV CD spectra were taken on a JASCO J810 spectropolarimeter in a 0.2 cm cell at 25 °C. The denaturation graph was plotted for the loss of CD signal at 222 nm as a function of increasing GdnHCl concentration. Thermal denaturation study was carried out by monitoring changes in the CD ellipticity at 222 nm as a function of temperature with constant rate of 1 °C/minute.

3. Results and discussion

3.1. EhPGDH production and oligomeric state

EhPGDH was over-expressed and purified by method described earlier [9]. The purified protein was homogenous as indicated by a single protein band on SDS-PAGE and a single peak in ESI-MS of molecular mass ~35 kDa (data not shown). The protein was eluted with a retention volume of 14.3 mL on a Superdex™ 200 10/300GL column which corresponded to a molecular mass of ~70 kDa as compared with the standard molecular weight markers which shows that the protein exists as a homodimer under non-dissociating conditions (Fig. 1A). This result was further confirmed by glutaraldehyde mediated chemical cross-linking experiment, where a single protein band corresponding to a molecular mass of ~70 kDa was observed on 12% SDS-PAGE (Fig. 1B).

3.2. Structural features of EhPGDH

3.2.1. Secondary structure

EhPGDH possesses a typical α/β type secondary structure as observed by the far-UV CD scan of the native protein (Fig. 2A) and analysis of the CD data using K2D software (not discussed) [24].

3.2.2. Molecular modelling and molecular dynamic simulations

Since no crystal structure of EhPGDH was reported, a three dimensional model of EhPGDH dimer, based on four crystal structure templates of phosphoglycerate dehydrogenases from (*Escherichia coli*, Protein Data bank code 2PA3, *M. tuberculosis*, PDB code 1YGY, *Pyrococcus horikoshii*, PDB code 1WWK and *Homo sapiens*, PDB code 2G76) was generated and subsequently subjected to energy minimization followed by a 10 ns molecular dynamics simulation by method as described under the section "materials and methods"(2.10). The results show that EhPGDH exists as a homodimer and each subunit comprises of two domains. The nucleotide binding domain (Nbd) which consists of residues Arg103–Gly265 contains six parallel β strands interconnected with

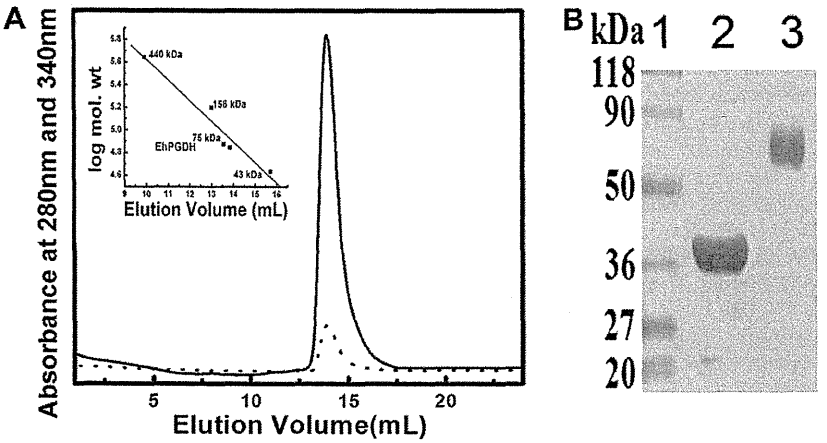


Fig. 1. EhPGDH is stabilized as a homodimer. (A) SEC profile of EhPGDH on a Superdex 200 10/300GL column. The inset shows a log plot of standard molecular markers. (B) 12% SDS-PAGE chemical cross-linking profile of EhPGDH. Lane 1–3 shows molecular weight markers, EhPGDH protein and cross-linked dimer of EhPGDH, respectively.

seven α-helices. This is a typical Rossman fold type arrangement characteristic for dehydrogenases (Fig. 2B). The substrate binding domain (Sbd) is adjacent to the Nbd and is composed of residues Val4–Asn102 and Ala266–Phe299. The active site pocket is present at the interface of the two domains which harbours the signature nucleotide binding site fingerprint sequence (Gly–Xaa–Gly–Xaa2–Gly–Xaa17–Asp) located between amino acids 139–162 at the domain interface.

3.2.3. *Fluorescence resonance energy transfer (FRET)*

Fig. 2C summarizes the result of FRET experiment on EhPGDH. For fluorescence scan of native the protein, two emission maxima

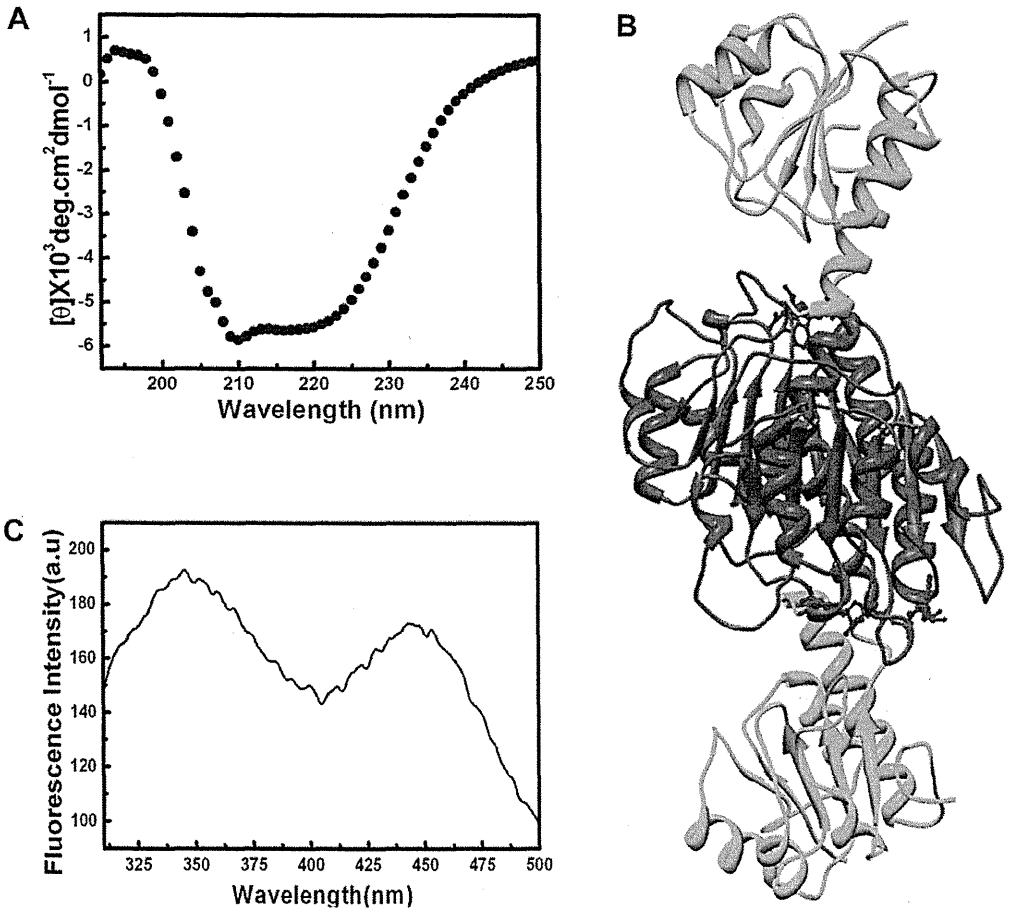


Fig. 2. Structural features of EhPGDH. (A) far-UV CD spectrum of EhPGDH. (B) Energy minimized average simulated model of EhPGDH. The nucleotide binding domain (Nbd), substrate binding domain (Sbd) and the cofactor NADH is shown in purple, green and blue colours, respectively. (C) Fluorescence resonance energy transfer spectrum of EhPGDH. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

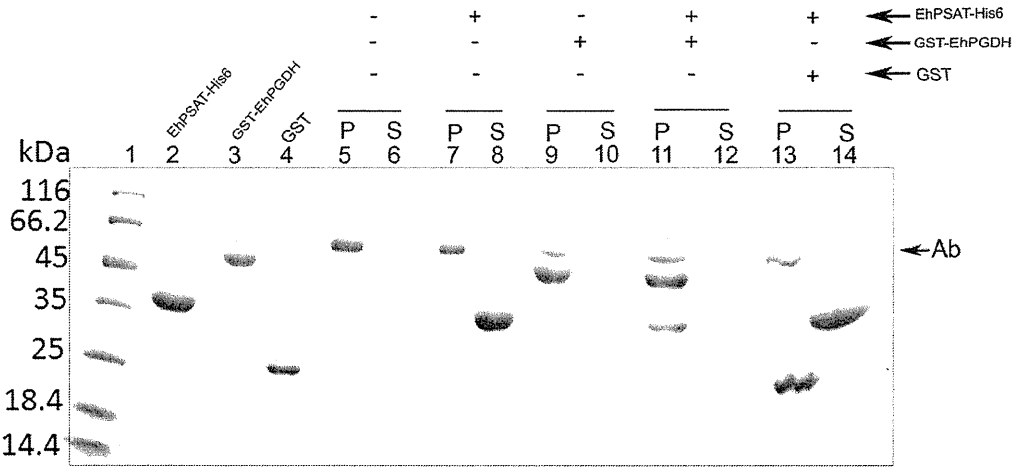


Fig. 3. EhPGDH physically interacts with EhPSAT. SDS-PAGE profile showing binding of GST-EhPGDH specifically with EhPSAT-His_6 assessed by the pull-down assay using anti-GST antibodies. Lanes 1–4 are the molecular weight marker and the respective purified proteins used during the pull-down experiment; Lane 2, EhPSAT-His_6 (43 kDa); Lane 3, GST-EhPGDH (66 kDa); Lane 4, Glutathione-S-transferase (GST)(26 kDa). Lanes 5–14 are the proteins after pull-down using anti-GST antibodies. "P" represents proteins on protein A-sepharose beads; "S" represents unbound proteins in the supernatant. The arrow Ab, marks the antibody band.

one at 336 nm (corresponding for the aromatic amino acids) and another at 470 nm (corresponding to NADH) resulting from resonance energy transfer were observed. However, unfolding of the protein on incubation with 6 M GdnHCl resulted in a single emission maximum at 351 nm (data not shown). These observations demonstrate that in the native conformation of EhPGDH the cofactor NADH is closely associated with the protein molecule. On denaturation, the protein conformation was perturbed resulting in loss of FRET as the two fluorescent moieties move apart from each other (data not shown).

3.3. *EhPGDH physically interacts with EhPSAT under in vitro conditions*

3.3.1. *Pull-down assay*

To analyze presence of physical interaction between the two proteins EhPGDH and EhPSAT were cloned in pGEX-4T1 and pET15b vectors with GST and His_6-tags, respectively. The proteins were over-expressed and purified to homogeneity. Furthermore, to establish affinity between the two interacting partners they were incubated under *in vitro* conditions. Mice anti-GST antibody was used for the pull-down experiment. The antibody was able to pull both the proteins when added to the interacting mixture. However, it was unable to precipitate the non-cognate protein in the absence of the cognate partner. In a control experiment with purified Glutathione-S-transferase (GST) protein and EhPSAT, only GST was pulled out from the reaction mixture and EhPSAT was left in the milieu (Fig. 3). These results indicate presence of specific protein–protein interactions between EhPGDH and EhPSAT. To further substantiate these results we carried out gel filtration and chemical cross-linking experiments.

3.3.2. *Gel filtration*

To analyze the molecular mass of the EhPGDH-PSAT protein complex His_6-tag proteins were used for gel filtration experiments on a Superdex™ 200 10/300GL column (manufacturer protein exclusion limit of 600 kDa). The column was equilibrated and run in 50 mM potassium phosphate buffer at pH 7.5 and was pre-calibrated using ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa) and albumin (43 kDa) as molecular weight markers and blue dextran as a marker of the excluded (void) volume. A 0.5 mL sample containing equimolar amount of the interacting proteins was incubated for complex formation prior of being applied on the column. The enzyme complex eluted at a retention volume of 13.1 mL (Fig. 4A). The molecular weight of the protein complex estimated with reference to the above mentioned

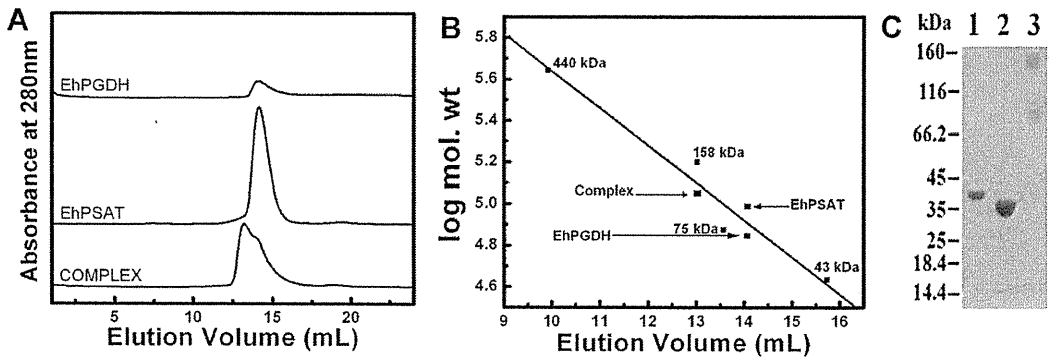


Fig. 4. Characterization of the EhPGDH-PSAT complex. (A) SEC profile of EhPGDH-PSAT complex and individual PGDH and PSAT proteins on a Superdex 200 10/300GL column. (B) Log plot of standard molecular markers ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa) and albumin (43 kDa). (C) 8% SDS-PAGE chemical cross-linking profile of EhPGDH-PSAT complex. Lane 1–3 shows EhPSAT, EhPGDH protein and cross-linked protein–protein complex, respectively.

standard molecular weight markers was ~155 kDa (Fig. 4B). Fractions eluted were assayed for the presence of both aminotransferase and dehydrogenase activities, respectively. However a shoulder (hump) was also observed on the latter half of the eluted peak of the complex which corresponded to the fraction of unreacted protein molecules.

3.3.3. Chemical cross-linking

The eluted protein–protein complex fraction from the SEC experiment was subjected to glutaraldehyde mediated chemical cross-linking and was resolved on an 8% SDS-PAGE. A protein band of ~155 kDa was obtained corresponding for the EhPGDH-PSAT complex (Fig. 4C). The above results suggest that EhPGDH and EhPSAT interact with each other forming a protein–protein complex.

3.3.4. Tryptophan fluorescence anisotropy

To analyze the stoichiometry of the interacting proteins and the binding constant of the EhPGDH-PSAT complex we carried out fluorescence studies. Anisotropic studies are ideally suited for studying such interacting systems where one of the components has an intrinsic fluorophore or is tagged with a fluorescent probe [25,26]. Tryptophan is an intrinsic fluorophore which has been frequently used for monitoring changes in the tertiary microenvironment of the proteins as a consequence of physical interactions [25]. Interestingly, EhPGDH has no tryptophan moiety however; EhPSAT has three tryptophan residues present in the large PLP-binding domain. Fig. 5A shows the increase in fluorescence anisotropy when a sample of EhPSAT-His_6 at a concentration of 1 μM was allowed to interact with increasing concentrations of EhPGDH-His_6 at pH 7.5 in 50 mM phosphate buffer.

Initially the fluorescence readings increased with increasing concentration of EhPGDH with a maximum value observed at 1 μM protein concentration, the anisotropy values stabilized on further increasing the EhPGDH concentration even up to 4 μM. Thus the anisotropy plot showed that the two proteins interacted in a 1:1 stoichiometry. The individual molecular weight of EhPGDH and EhPSAT is 70 kDa and 86 kDa respectively and the molecular weight of the protein complex as observed by gel filtration and chemical cross-linking is ~155 kDa which is in close approximation with the collective molecular weight of the two proteins i.e. ~157 kDa. Thus the results show the EhPGDH and EhPSAT specifically interact in a 1:1 ratio to form a stable protein complex. The apparent dissociation constant for a 1:1 protein–protein complex calculated from fluorescence anisotropy readings was 3.453×10^{-7} M.

3.3.5. Complex docking and molecular dynamics simulation

To further understand the mode of interaction which stabilize the EhPGDH-PSAT complex, we generated a model for the complex between EhPGDH and EhPSAT using the rigid-body Monte Carlo search employed by protein–protein docking program, RosettaDock which combines Monte Carlo-plus-minimization scheme with Van der Waals interactions, implicit solvation, and hydrogen bonding potential dominated energy function [15]. As the three dimensional structure of both EhPGDH and EhPSAT were not yet known hence we generated homology models for both the proteins using *E. coli*, Protein Data bank (PDB) code 2PA3, *M. tuberculosis*, PDB code 1YGY, *P. horikoshii*, PDB code 1WWK and *H. sapiens*, PDB code 2G76 as crystal structure templates for EhPGDH and *E. coli*, PDB code 1BJN, *Campylobacter jejuni*, PDB code 3M5U, *Bacillus circulans* var. *alkalophilus*, PDB code 1BT4, and *H. sapiens*, PDB code 3E77 as templates for EhPSAT homology model. Homo-dimeric structures of both EhPGDH and EhPSAT were prepared by superimposing on their corresponding template dimers.

Initial starting structure for Rosetta docking was generated using an alternative docking program, Patchdock based on a global scan of translational and rotational space [14]. The output structures from Patchdock run were refined using Firedock and the top ranked complex model was used as input for Rosetta run [14]. From the docking using RosettaDock 1000 structures were first generated, which were then clustered by Rosetta's clustering algorithm. Clustering of 1000 Rosetta protein–protein decoys resulted in 101 clusters including 27 clusters with more than 10 members. The number of members in three biggest clusters is 46, 39 and 23. The

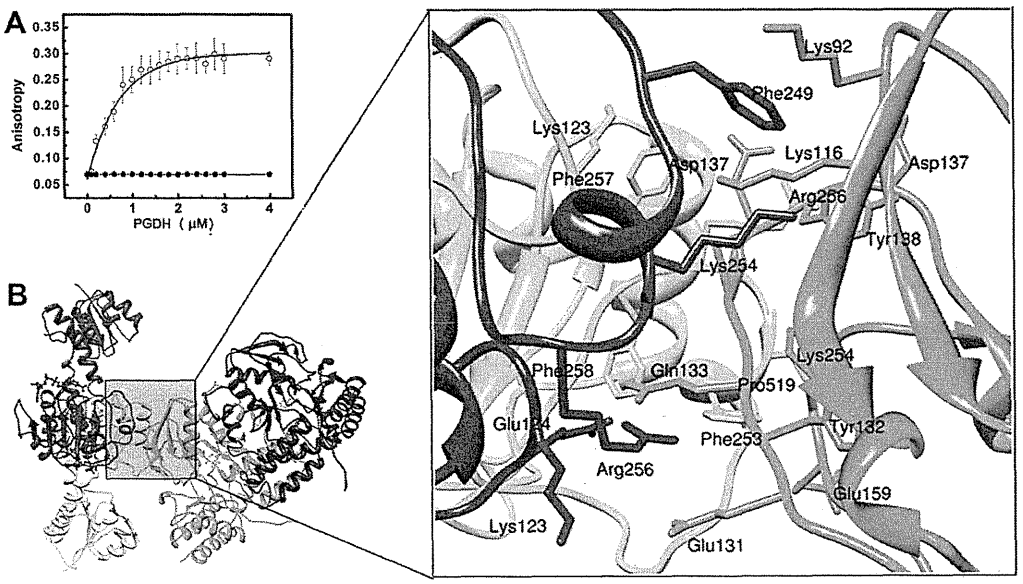


Fig. 5. Fluorescence anisotropy, homology modelling, protein–protein docking and computer simulation of the EhPGDH-PSAT complex. (A) Tryptophan fluorescence anisotropy profile of EhPGDH-PSAT complex. (○) Shows interaction between the two proteins in absence of salt in phosphate buffer pH 7.5 and (●) shows loss of the protein–protein interaction in presence of 200 mM NaCl. (B) Energy minimized average simulated model of EhPGDH-PSAT complex. The enlarged protein–protein interaction interface shows the amino acids residues of both the interacting partners which are involved in complex formation.

members in the biggest clusters were than ranked by Rosetta-all-atom total energy and 10 top ranked members with lowest energy are shown in supplementary Fig. 2A.

Relatively small deviation in structures of most populous cluster and a well-defined docking funnel obtained after plotting the Rosetta total score and RMSD with lowest ranked structure as seen in supplementary Fig. 2B indicates that the EhPGDH-PSAT predicted complex is a well-defined one.

In the lowest energy structure, 99.5% of the residues were in most favoured and allowed regions on Ramachandran plot which was analyzed by using the program Procheck [27], 0.3% residues are present in the generously allowed region and 0.2% in the disallowed region, suggesting good quality of the predicted complex model. As Cluster 1 contains the largest number of members (46), the lowest energy structure from this cluster with Rosetta-all-atom total energy of −2387.98 kcal/mol was selected to represent model of EhPGDH-PSAT complex.

Subsequently this complex model was subjected to a 10 ns molecular dynamic simulation by method described under the section "materials and methods"(2.10). Fig. 5B shows energy minimized average simulated model of the EhPGDH-PSAT protein complex. To assess the quality of our MD simulations the energetic and structural properties were monitored along the entire MD trajectory of the complex. Supplementary Fig. 3 shows plots of the total energy of the entire system as a function of simulation time for EhPGDH-PSAT complex. The average total energy is found to be −369619.6862 kcal/mol and the average fluctuation from the average total energy was less than 2000 kcal/mol.

It is clear from the supplementary Fig. 3 that total energy is oscillating around the average energy, showing the stability of system at a given temperature of 300 K. Stability of the complex structure can also be reflected by its deviation from the initial structure in terms of root mean square deviation (RMSD) of the backbone atoms of a protein over the course of a simulation and may be used as a measure of the conformational stability of a protein structure or model during that simulation. Therefore, we monitored the RMSD of backbone atoms throughout the trajectory from their initial configuration. It can be seen from supplementary Fig. 4 that initially RMSD from starting structure rose from 1 to 5 Å till 8 ns. This initial rise is commonly observed in protein simulations of this much size and is thought to reflect the relaxation of the protein once constraints applied at the time of heating and equilibrations are removed. After 8 ns the relative fluctuation for the RMSD is small (less than 1 Å) as system has been stabilized. Thus, from the structural and energetic analysis of the energy minimized simulated protein complex we observed that the system existed in a stable conformation.

The MM-PBSA and MM-GBSA approach was then used to calculate the free energy change due to the formation of EhPGDH-PSAT complex. The snapshot structures used for MM-GBSA and MM-PBSA calculation were extracted from molecular dynamics simulation trajectory and last 2 ns data was used for the calculation of binding free energy. Detail account of contribution from various energy components based on the MM-PBSA and MM-GBSA methods is given in Table 1. As can be seen in Table 1 the major contributions to the binding free energy arose from the polar contribution to solvation energy or electrostatic solvation term $\Delta E_{gbsapolar}$ and $\Delta E_{pbsapolar}$, as calculated by MM-GBSA and MM-PBSA approach (−67.6851 and −89.6863 kcal/mol). Highly favourable solvation energy probably reflects large number of favourable interactions from the positively and negatively charged residues at the binding interface. Apart from this, the intermolecular Van der Waals interactions with a value of −50.1361 kcal/mol also provide the driving force for binding and formation of the protein–protein complex. The nonpolar interactions were also found to make favourable contribution to the binding energy with of $\Delta E_{gbsa,nonpolar}$ of −6.5842 kcal/mol and $\Delta E_{pbsa,nonpolar}$ of −5.9895 kcal/mol based on the MM-GBSA and MM-PBSA methods, respectively.

Furthermore, in order to gain a clear picture about the contribution of each residue at the complex interface that make significant contributions to the binding free energy associated with complex formation, the total binding free energy was decomposed into per residue contributions, using MM/GBSA free energy decomposition analysis. The residue wise contribution and the amino acid residues that are most important contributors for the interaction between EhPGDH and EhPSAT are shown in supplementary Fig. 5.

In general, Glu131, Gln133, Pro135, Asp137, Tyr138, Asp139, Glu159, Ser160, Pro161 of EhPSAT and Phe122, Lys123, Phe249, Phe253 and Arg256 of EhPGDH showed most significant and favourable contributions for complex formation (Fig. 5B).

3.4. Ionic interactions play a vital role in EhPGDH-PSAT complex formation and stability

In silico analysis showed presence of ionic residues at the complex interface (Fig. 5B) which indicates that ionic residues might play a major role in complex formation and its stability. To

Table 1
Components of the binding free energy of the EhPGDH-PSAT complex calculated from the trajectory.

	EhPGDH-EhPSAT		EhPSAT		EhPGDH		Delta	
Contribution	Mean	Std	Mean	Std	Mean	Std	Mean	Std
ΔE_{vdw}	−10,425.6617	49.7578	−5882.7384	38.7040	−4492.7872	29.4199	−50.1361	5.1049
ΔE_{elec}	−98,381.5237	210.4441	−53,648.1356	146.5179	−44,864.1590	159.9214	130.7708	66.7448
$\Delta E_{gbsa,polar}$	−15,914.2873	170.7201	−7960.1279	128.5757	−7886.4744	130.9012	−67.6851	62.4015
$\Delta E_{gbsa,nonpolar}$	442.2818	3.9515	226.6367	3.2593	222.2293	2.3709	−6.5842	0.7713
ΔE_{gbsa}	−124,279.1909	119.2186	−67,264.3651	89.1104	−57,021.1913	76.9788	6.3655	7.7763
$\Delta E_{pbsa,polar}$	−17,443.4239	162.8883	−8738.3619	118.7700	−8615.3756	128.9706	−89.6863	62.3469
$\Delta E_{pbsa,nonpolar}$	335.3100	1.7046	168.3451	1.5341	172.9543	1.3699	−5.9895	0.5968
ΔE_{pbsa}	−125,915.2992	122.8136	−68,100.8907	95.8278	−57,799.3675	74.8465	−15.0411	8.6207

ΔE_{vdw} = Van der waals contribution to binding free energy from molecular mechanics forcefield.
ΔE_{elec} = Electrostatics contribution to binding free energy from molecular mechanics forcefield.
$\Delta E_{gbsa,polar}$ = Polar contribution to solvation energy from generalized born model.
$\Delta E_{gbsa,nonpolar}$ = Nonpolar contribution to solvation energy from surface area calculations.
ΔE_{gbsa} = Total binding free energy according to MM-GBSA approach.
$\Delta E_{pbsa,polar}$ = Polar contribution to solvation energy by solving poisson Boltzmann equation.
$\Delta E_{pbsa,nonpolar}$ = Nonpolar contribution to solvation energy from surface area calculations.
ΔE_{pbsa} = Total binding free energy according to MM-PBSA approach.
All values are given in kcal/mol.

further validate this we performed interaction assays with EhPGDH and EhPSAT in presence of increasing concentration of NaCl. The two interacting partners stayed with each other at low NaCl concentrations however; at ≥200 mM NaCl the interactions were lost as observed by fluorescence anisotropy (Fig. 5A). The pull-down assay in 200 mM NaCl also showed loss of interactions between the two proteins with presence of EhPGDH and EhPSAT in separate pellet and supernatant fractions on SDS-PAGE (data not shown). Similar results were observed with gel filtration and chemical cross-linking experiments (data not shown) in presence of 200 mM NaCl. Thus ionic interactions are vital for complex formation. On increasing NaCl concentration the electrostatic interactions are compromised resulting in loss of complexation.

Furthermore, the *in silico* analysis of the protein–protein complex also shows that the Nbd of EhPGDH specifically interacts with the EhPSAT. To further understand the role of the structural domains of EhPGDH in complex formation we studied the structural, stability and complex forming features of separated Nbd. Prior to that independent existence of the two domains i.e. Nbd and Sbd were studied by analyzing the structural and stability features of the EhPGDH.

3.5. The two structural domains of EhPGDH fold/unfold independently

3.5.1. Thermal denaturation

The thermal unfolding of EhPGDH was characterized by monitoring the loss of secondary structure of enzyme with temperature. Fig. 6A summarizes the changes in CD ellipticity at 222 nm of EhPGDH with increasing temperature. Two distinct transitions between 25 and 100 °C were observed for the protein. These transitions fall in the temperature regions from 35 to 60 °C and 65–100 °C and centred at about 45 and 65 °C, respectively.

3.5.2. GdnHCl unfolding

The guanidine hydrochloride-induced changes in the structural properties of EhPGDH were studied using CD spectroscopy. The effect of GdnHCl on the secondary structure of EhPGDH was studied by monitoring changes in the far-UV CD profiles of the enzyme. Fig. 6B summarizes the effect of increasing GdnHCl concentration on loss of ellipticity at 222 nm for EhPGDH. The denaturation profile shows two well separated transitions. First transition between 0.4 and 0.8 M GdnHCl, and second transition between 0.9 and 3 M GdnHCl were observed. Furthermore, only about 30% loss of secondary structure was found to be associated with the first transition.

Thermal and GdnHCl unfolding curves of EhPGDH with two transitions indicates a possible presence of two independent folding/unfolding structural domains. To further test this hypothesis we isolated and studied the structural features of the Nbd-EhPGDH, separately.

3.6. Nucleotide binding domain (Nbd) of EhPGDH

The results from the thermal and chemical unfolding studies of EhPGDH suggest that the two domains the Nbd and the Sbd probably unfold independent of each other. The MD simulations analysis clear suggests that EhPGDH interacts exclusively through its Nbd with EhPSAT. To check the results of protein–protein docking and simulation, the Nbd of EhPGDH was cloned, over-expressed, purified to homogeneity. The purified domain was biophysically characterized before performing interaction studies. The purified protein has 19 kDa of molecular mass as observed from the SDS-PAGE analysis (Fig. 7A). Under non-dissociating conditions the domain exist as a dimer as observed by SEC on a Superdex™ 75 10/300GL column (Fig. 7B). Similar results were also observed by glutaraldehyde mediated protein cross-linking where a single band of approximately 38 kDa was obtained on a 12% SDS-PAGE (Fig. 7B inset).

3.6.1. Structural characterization of Nbd

The recombinant domain was properly folded as observed by its far-UV CD spectrum which showed presence of a typical α/β type secondary structure (Fig. 7C). The purified Nbd was functionally inactive as the active site pocket of the protein is situated at the interface of the two domains i.e. the Nbd and the Sbd. Thermal unfolding of purified Nbd as monitored by the loss of secondary structure of the domain with increasing temperature show a single transition with a Tm of approximately 65 °C (Fig. 6A). The denaturation profile overlapped with the second transition of the thermal denaturation curve of the native EhPGDH. These observations suggest that the second transition observed during the thermal unfolding of EhPGDH corresponds to melting of Nbd of the EhPGDH.

GdnHCl-induced unfolding of the Nbd also showed a single transition, as monitored by the loss of CD ellipticity at 222 nm (Fig. 6B). This transition overlapped the second phase transition of the GdnHCl-induced unfolding curve of the native EhPGDH. The Cm

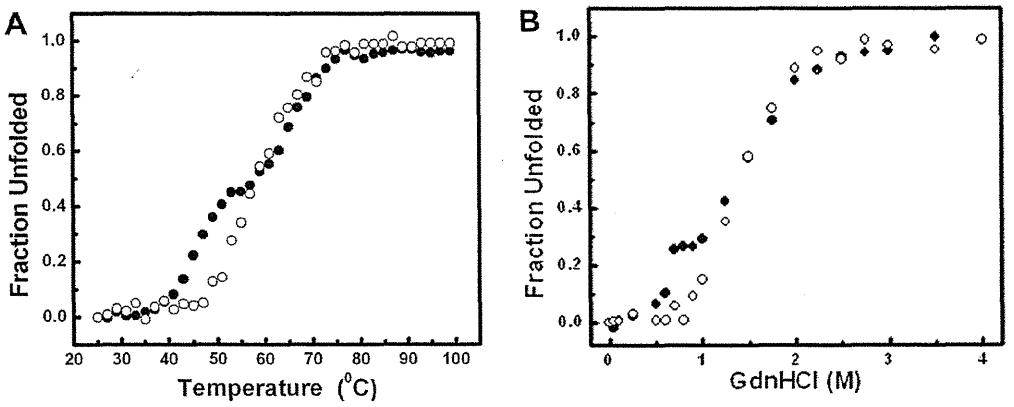


Fig. 6. Denaturation profile of EhPGDH. (A) Thermal denaturation curve of EhPGDH (●) and nucleotide binding domain (Nbd) (○) monitored by loss of CD ellipticity at 222 nm as a function of increasing temperature. (B) GdnHCl denaturation profile of EhPGDH (●) and nucleotide binding domain (Nbd) (○) monitored by loss of CD ellipticity at 222 nm as a function of increasing denaturant concentration.

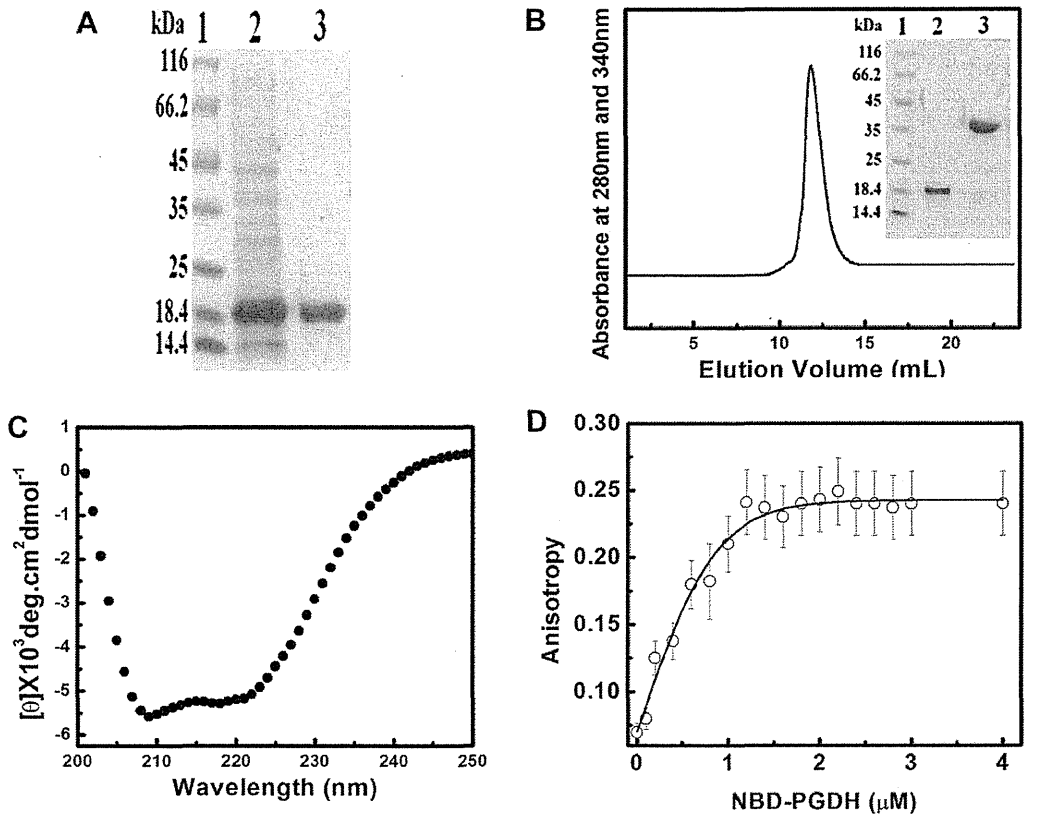


Fig. 7. Over-expression and characterization of the nucleotide binding domain (Nbd) of EhPGDH. (A) 12% SDS-PAGE profile showing over-expression and purification of Nbd. Lane 1–4 represents molecular weight markers, un-induced, induced *E coli* (BL-21 RPL) cell lysate and purified domain, respectively. (B) SEC profile of Nbd monitored on a Superdex 75 10/300GL column. Inset shows a 12% SDS-PAGE chemical cross-linking profile of Nbd. Lanes 1–3 represents molecular weight markers, uncross-linked and cross-linked domain, respectively. (C) Far-UV CD spectrum of Nbd. (D) Tryptophan fluorescence anisotropy profile of Nbd-PSAT complex.

value for the domain was about 1.8 M, which was quite comparable to the Cm observed for the second transition of the biphasic curve during urea denaturation of the EhPGDH.

These results support our earlier suggestion that the two domains of EhPGDH are independent folding/unfolding units. Thus the Nbd maintains its native like conformation even when it's been isolated.

3.6.2. The isolated Nbd interacts with EhPSAT

The energy minimized model of the protein complex suggests that the Nbd of EhPGDH specifically interacts with the PLP-binding domain of EhPSAT. Subsequently interaction study was performed between the Nbd and EhPSAT using fluorescence anisotropy. Incubation of increasing concentration of Nbd with EhPSAT showed an enhancement in the anisotropy values and followed the trend very similar to native EhPGDH-PSAT anisotropy experiment (Fig. 7D). Based on these results the dissociation constant calculated was 3.491×10^{-7} M which was quite similar to that observed for the native EhPGDH-PSAT protein complex (as discussed above). The experiments clearly suggest that the Nbd of the EhPGDH specifically interacts with the EhPSAT to form a protein–protein complex.

3.7. Substrate channelling

Enzymes efficiently work in close association with each other particularly the enzymes that work within a metabolic pathway interact with each other to form organized enzyme complexes [2]. Such a micro-compartmentation leads to channelling of a metabolite from one catalytic site to another; the intermediate produced in the process is efficiently transferred between catalytic sites of the interacting partners without being completely mixed with the bulk phase [28–30]. Such substrate channelling is observed for enzymes involved in metabolic pathways and the phenomenon ensures optimal metabolon activity without diffusion of the pathway intermediates in milieu. Ever since Paul Srere have coined the term "metabolon", a large number of cases of such enzyme complexes have been reported [6,31–34]. Such complexes are generally a part of large metabolite chain, in which the product of one enzyme catalyzed reaction serves as a substrate for the next enzyme. A considerable body of evidence has been presented which demonstrate these interactions and shows differences in kinetics of enzymes which catalyze sequential metabolic reactions when they are in close proximity compared to the kinetics of the same enzymes in random solution [35,36].

Among the mitochondrial enzymes of the Krebs cycle, interactions between sequential enzymes have been demonstrated between six of eight possible interactions [3,37]. Also, high molecular weight complexes of the Krebs cycle enzymes have been isolated and kinetic effects have been observed in some sequential reactions within these complexes compared to corresponding free enzymes [38–40]. A problem in demonstrating the *in situ* existence of enzyme complexes as well as demonstrating possible advantages of these complexes is that during isolation many of them tend to dissociate due to dilution effects. A number of model systems have therefore been used to investigate the effects of close proximity of