

Insight into the 3D structure of ADP-glucose pyrophosphorylase from rice (*Oryza sativa* L.)

Chhavi Dawar · Sunita Jain · Sudhir Kumar

Received: 14 February 2013 / Accepted: 4 April 2013 / Published online: 15 May 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract ADP-glucose pyrophosphorylase (E.C. 2.7.7.27; AGPase) is a key regulatory enzyme that catalyzes the rate-limiting step of starch biosynthesis in higher plants. AGPase consists of pair of small (SS) and large (LS) subunits thereby constituting a heterotetrameric structure. No crystal structure of the native heterotetrameric enzyme is available for any species, thus limiting the complete understanding of structure–function relationships of this enzyme. In this study, an attempt was made to deduce the heterotetrameric assembly of AGPase in rice. Homology modeling of the three-dimensional structure of the LS and SS was performed using the Swiss Model Server, and the models were evaluated and docked using GRAMM-X to obtain the stable heterodimer orientation (LS as receptor and SS as ligand) and then the heterotetrameric orientation. The initial heterotetrameric orientation was further refined using the RosettaDock Server. MD simulation of the representative heterodimer/tetramer was performed using NAMD, which indicated that the tail-to-tail interaction of LS and SS was more stable than the head-to-head orientation, and the heterotetramer energy was also minimized to $-767,011$ kcal mol⁻¹. Subunit–subunit interaction studies were then carried out using the programs NACCESS and Dimplot. A total of 57 interface residues were listed in SS and 63 in LS. The residues plotted by Dimplot were similar to those listed by NACCESS. Multiple sequence alignment of the sequences of LS and SS from potato, maize and rice validated the interactions inferred in the study. RMSD of 1.093 Å was obtained on superimposition of the deduced heterotetramer on the template homo-tetramer (1YP2), showing the similarity between the two structures.

Keywords ADP-glucose pyrophosphorylase · Homology modeling · Docking · MD simulation · Subunit–subunit interactions

Introduction

ADP-glucose pyrophosphorylase (E.C. 2.7.7.27; AGPase) is a major enzyme controlling starch biosynthesis. It catalyzes the regulatory step of ADP-glucose formation, which provides a glucosyl donor for elongation of the α -1-4-glycosidic chain in the presence of divalent metal Mg²⁺ [34]. Thus, this enzyme plays a key role in the modulation of photosynthetic efficiency in source tissues and also determines the level of storage starch in sink tissues, thereby influencing overall crop yield potential [21]. Regulation of almost all higher AGPase depends on the ratio of 3-phosphoglyceric acid to inorganic phosphate (3PGA/Pi). While 3-PGA functions as the main stimulator, Pi inhibits the activity of the enzyme [36].

Plant AGPases consist of a pair of small subunits (SS) and large subunits (LS), which constitute a heterotetrameric structure. AGPase is encoded by two genes named Shrunken-2 (*Sh2*) and Brittle-2 (*Bt2*). The *Sh2* gene encodes the less conserved (50 %–60 % identity) larger subunit responsible for regulating the allosteric properties of SS, and the *Bt2* gene encodes the highly conserved (85 %–95 % identity) smaller subunit with catalytic and regulatory functions [35]. In addition, specific regions from both subunits are important for subunit association and thus enzyme stability and functioning in potato [31]. Using chimeric maize and potato small subunits, Cross et al. [8] observed a polymorphic motif in the SS that is critical for subunit interaction, and reported that a 55-amino acid region between residues 322

C. Dawar · S. Jain · S. Kumar (✉)
Bioinformatics Section, CCS Haryana Agricultural University,
Hisar 125004, India
e-mail: sudhir@hau.ernet.in

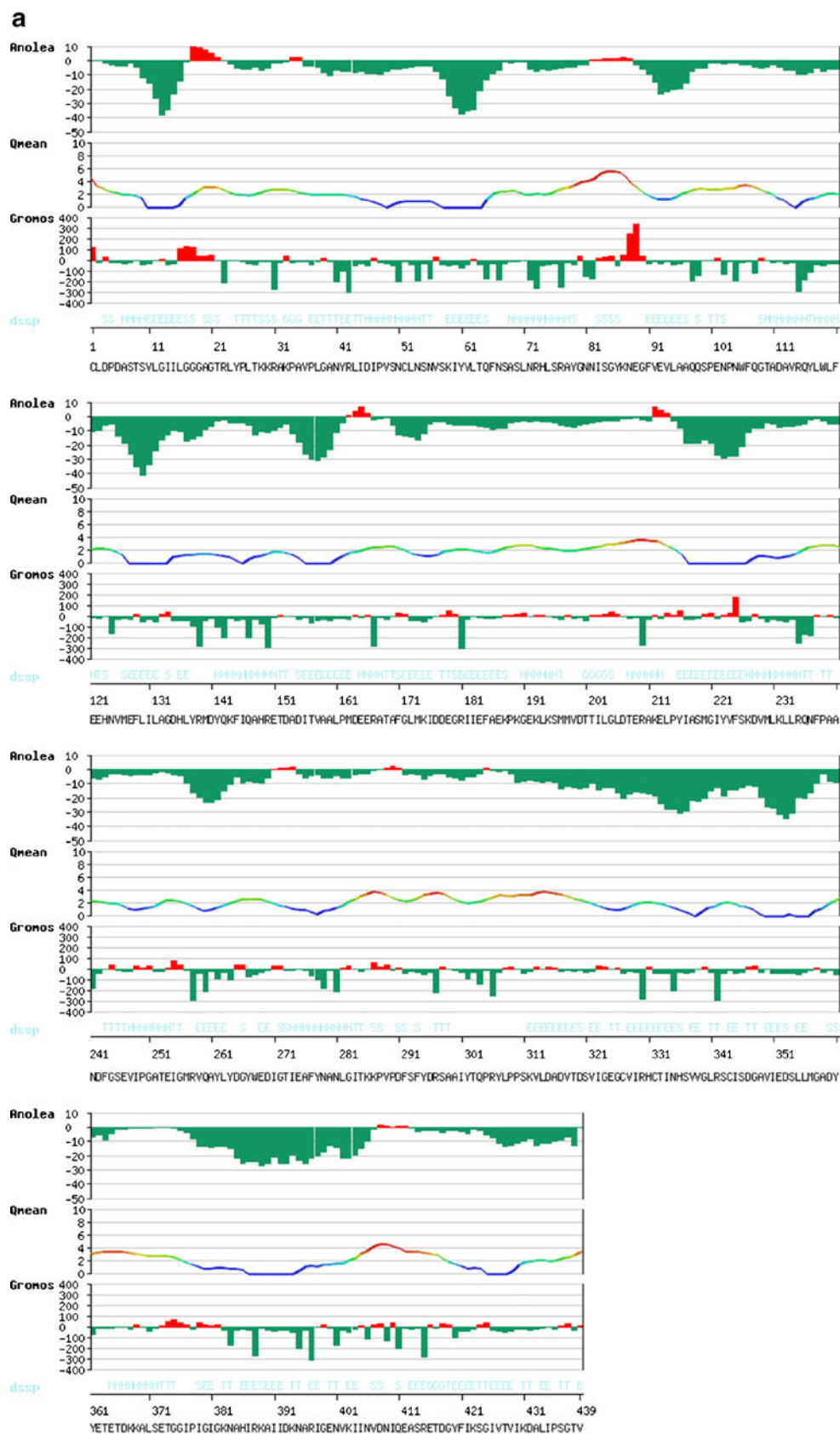


Fig. 1 Assessment for Anolea, Qmean and Gromos force fields of **a** small (SS) and **b** large (LS) subunits of rice ADP-glucose pyrophosphorylase (AGPase) structures. *Green* favorable energy environment, *red* unfavorable energy environment

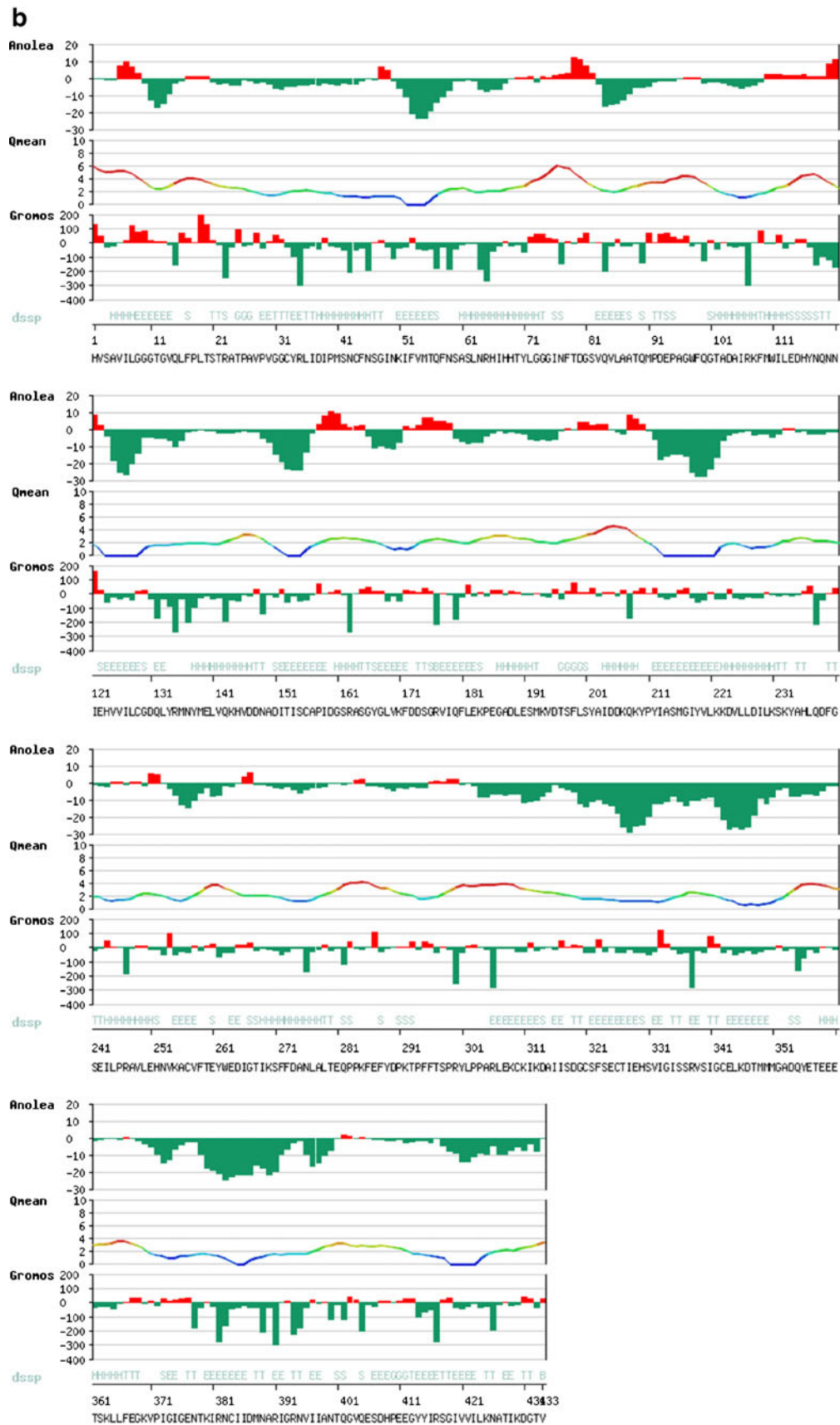


Fig. 1 (continued)

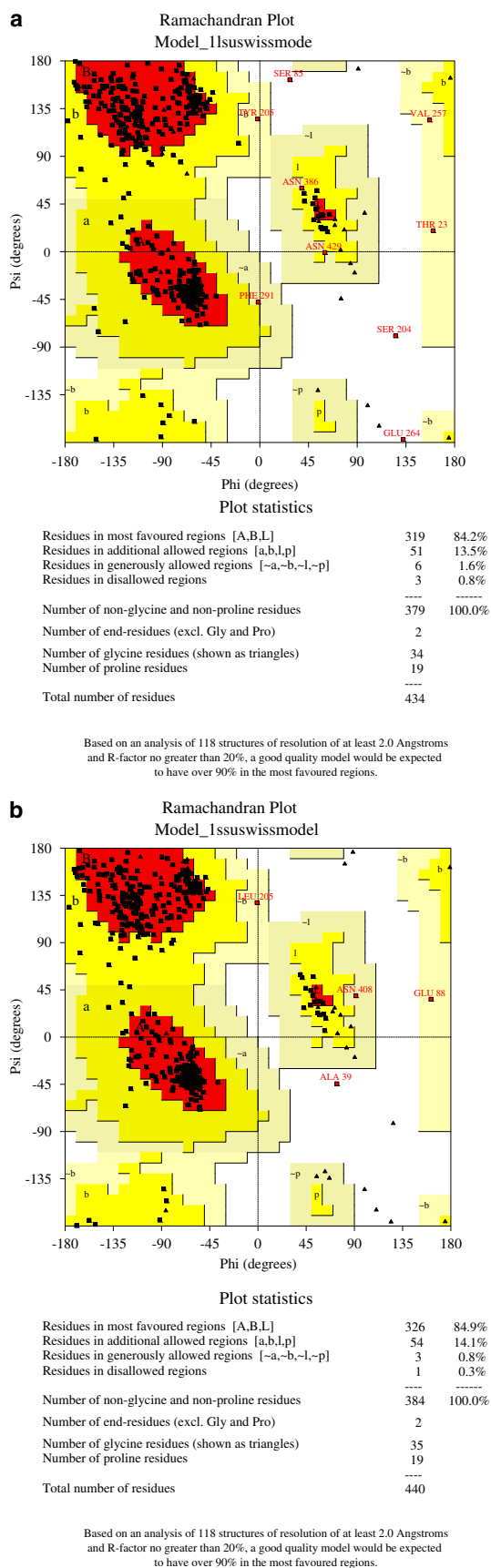


Fig. 2 Ramachandran plot for **a** LS and **b** SS of AGPase of rice

and 376 interacts directly with LS and contributes significantly to overall enzyme stability. Several chemical modification studies were also carried out by Frueauf et al. [11] to determine the role of conserved residues in potato tuber.

Analysis of the allosteric regulatory properties of these enzymes indicates that, although the LS is required for optimal activation by 3-PGA and resistance to Pi, the overall allosteric regulatory and kinetic properties are specified by both subunits. The results of Hwang et al. [19] showed that the regulatory and kinetic properties of AGPase are not simply due to the LS modulating the properties of the SS but, instead, are a product of synergistic interaction between the two subunits. Additionally, phylogenetic analysis of AGPase subunits revealed a role of subunit interfaces in the allosteric properties of the enzyme inferring that large subunits have undergone more duplication events than small subunits [12].

Several attempts have been made to obtain AGPase in a crystalline state allowing its structural characterization but have failed to date due to the difficulty of obtaining the enzyme in a pure state. The structures of neither the LS nor the heterotetrameric AGPase have been solved yet. This strongly limits a complete understanding of the structure–function relationships of the enzyme and also manipulation of the enzyme for increased starch production in grains. Recently, the first atomic resolution structure of AGPase from potato tuber (SS homotetramer) was obtained [20], following which several computational approaches involving homology modeling were applied to obtain an idea of the heterotetramer structure in potato as well as in other species. Tuncel et al. [40] modeled the large subunit of potato AGPase using homology modeling techniques and proposed a model for the heterotetrameric AGPase. Baris et al. [2] examined the AGPase model to identify important residues mediating the interactions between LS and SS using both computational and experimental techniques. Based on the molecular mechanics generalized Born surface area (MM-GBSA) method, two distinct LS domains were found to be involved in LS–SS subunit interactions.

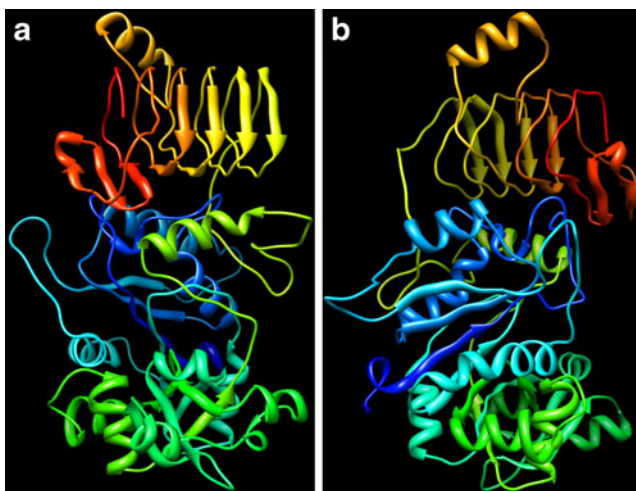
Manipulations of starch biosynthesis have already been shown by several studies. The expression of a cytoplasmically localized AGPase mutant gene from *Escherichia coli* in rice endosperm resulted in enhanced starch synthesis and, in turn, higher seed weight. The transgenic plants showed up to 13-fold higher AGPase activity [30]. Also, an in vivo, site-specific mutagenesis system was used in maize [13, 33] to create defined mutations within the gene encoding the large subunit of the endosperm-specific starch-synthesizing enzyme

Table 1 Evaluation of models of large (LS) and small (SS) subunits of ADP-glucose pyrophosphorylase (AGPase) of rice

Model	PROCHECK	ERRAT	VERIFY 3D	STATUS
LS	Core region	Overall Quality factor – 77.70	98.85 % of residues Averaged 3D- 1D Score >0.2	PASS
	84.2 %			
	Allowed region			
	13.5 %			
	Generously allowed region			
1.6 %				
SS	Disallowed region	Overall quality factor – 83.565	96.83 % of the residues had an averaged 3D-1D Score >0.2	PASS
	0.8 %			
	Core region			
	84.9 %			
	Allowed region			
14.1 %				
Generously allowed region				
0.8 %				
Disallowed region				
0.3 %				

AGPase. This work identified an important protein motif of this subunit and one resulting change that lead to an increase in seed weight of 11–18 %. These results complemented those in potato [37], showing that the AGPase reaction is one of the limiting factors of starch biosynthesis and that carbon flux into starch can be enhanced by increasing the net catalytic activity of this enzyme.

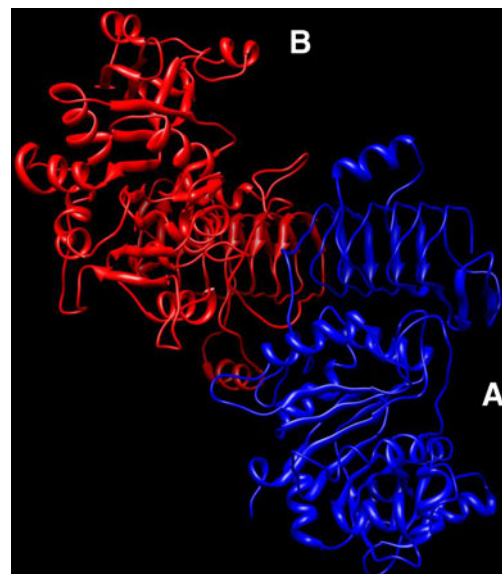
Therefore, it is critical to reveal the native heterotetrameric AGPase structure and identify the key residues taking part in subunit–subunit interactions to obtain a more detailed picture of the enzyme. In the present study, an attempt has been made to obtain an insight into the structure of two subunits (LS and SS) of heterotetramer AGPase in rice—one of the most important cereal plants—as well as the interactions of the two subunits to form the heterotetramer.

**Fig. 3** Predicted structures of **a** LS and **b** SS of AGPase of rice

Materials and methods

Homology modeling of the large and small subunit

The sequences of LS and SS of rice AGPase were retrieved from the NCBI protein sequence database (<http://www.ncbi.nlm.nih.gov/protein>) and a template was identified using PSI-BLAST ([1], <http://blast.ncbi.nlm.nih.gov/BLAST.cgi>) against the RCSB protein databank (PDB) ([5], <http://www.pdb.org>). The three dimensional (3D) structures of both the protein subunits were build using the Swiss Model

**Fig. 4** Best orientation as obtained from the GRAMM-X server of the heterodimer of AGPase of rice. *Blue* chain A, LS; *red* chain B, SS

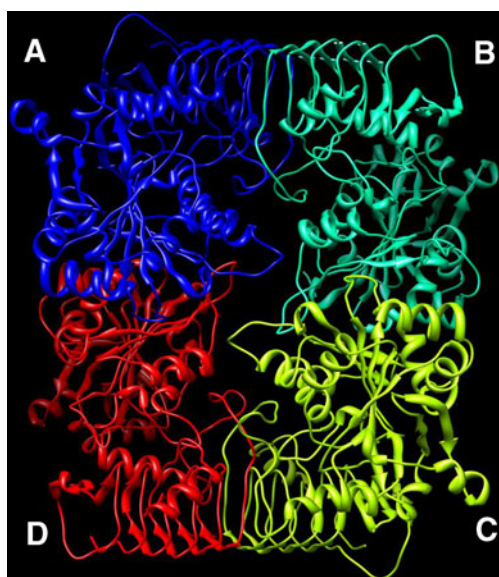


Fig. 5 Most stable orientation of the heterotetramer of AGPase of rice as obtained from the GRAMM-X Server. *Blue* Chain A, LS; *green* chain C, LS; *cyan* chain B, SS; *red* chain D, SS

Server ([32], <http://swissmodel.expasy.org/>). The modelled structures were assessed using the Protein structure and model assessment tools at the Swiss Model server, which utilizes various local and global quality estimation parameters. The models were further analyzed and verified using programs PROCHECK [23], VERIFY3D [10] and ERRAT [7] at the Structure Analysis and Verification Server (SAVES) (<http://nihserver.mbi.ucla.edu/SAVES>). The results were analyzed and models were further improved.

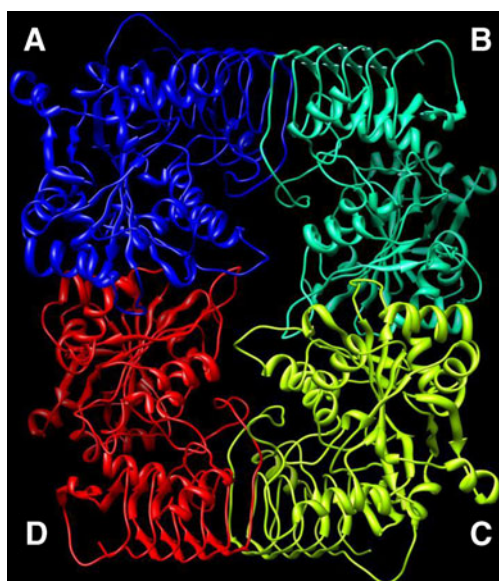


Fig. 6 Refined heterotetramer structure of AGPase of rice from the RosettaDock server with minimum total score of $-863.023 \text{ kJ mol}^{-1}$. *Blue* Chain A, LS; *green* chain C, LS; *cyan* chain B, SS; *red* chain D, SS

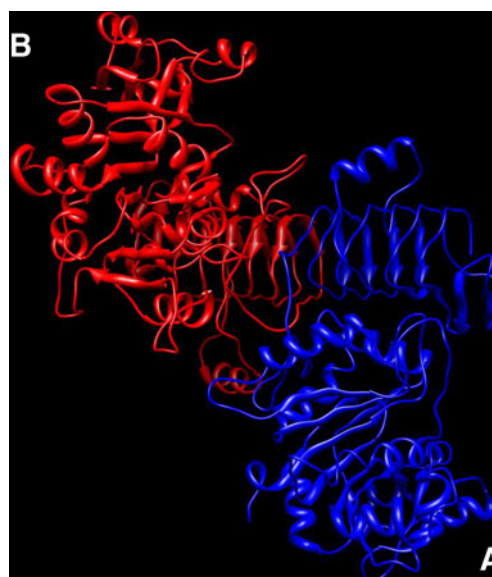


Fig. 7 Molecular dynamics (MD)-minimized structure of dimer 1 ($-437,986 \text{ kcal mol}^{-1}$) with tail-to-tail subunit interactions. Chain A LS and chain B SS

Protein–protein docking studies

The homology modeled LS and SS of rice AGPase were submitted to the GRAMM-X docking server ([39], <http://vakser.bioinformatics.ku.edu/resources/gramm/grammx>) to perform a rigid body docking using fast Fourier transformation methods by applying smoothed Lennard-Jones potential, knowledge-based and refinement stage scoring, which gives rise to the best surface match. The best dimer orientation was again fed to the GRAMM-X server to obtain initial heterotetramer orientations. The RosettaDock server ([24], <http://rosettaserver.graylab.jhu.edu/>) was used to refine the initial

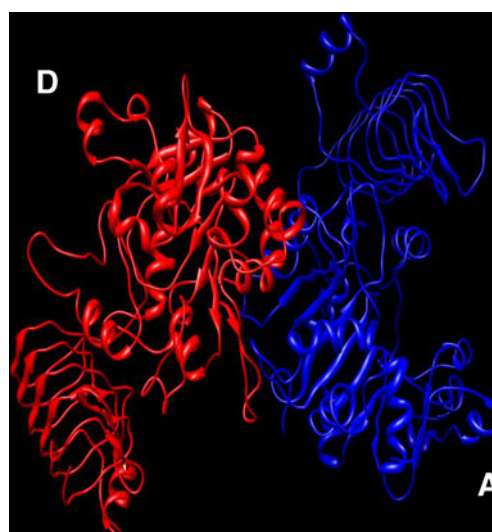


Fig. 8 MD-minimized structure of dimer 2 ($-396,097 \text{ kcal mol}^{-1}$) with head-to-head subunit interactions. Chain A LS and chain D SS

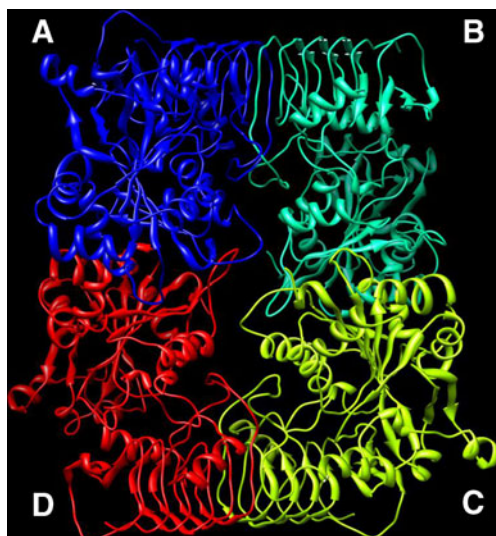


Fig. 9 MD-minimized structure ($-767011 \text{ kcal mol}^{-1}$) of the heterotetramer. *Blue* Chain A, LS; *green* chain C, LS; *cyan* chain B, SS; *red* chain D, SS

orientation obtained from GRAMM-X. The best orientation PDB files with energy scores were retrieved and the minimum energy structure was taken for further reference. Idealization of bond geometry and removal of unfavorable non-bonded contacts was performed by energy minimization with force field GROMOS96 [42] with the Swiss-Pdb Viewer ([14], <http://spdbv.vital-it.ch>).

Molecular dynamic simulations

Explicit solvent molecular dynamic (MD) simulation for the representative structure of AGPase heterodimer/heterotetramer from RosettaDock server was performed using NAMD ([28], <http://www.ks.uiuc.edu/Research/namd/>). MD was performed using the CHARMM force field [25] parameters with 2-fs time steps and constant pressure conditions on an 8-node Linux cluster operating at 0.1 TFLOPS. The starting structures

were solvated in a rectangular box of TIP3P (three-site models have three interaction sites, corresponding to the three atoms of the water molecule). The distance between the edge of the box and the closest solute molecule was 10 Å. Counter ions (Na^+) were added to neutralize the system, which was minimized using 10^4 steps for fixed backbone atoms and then another 10^4 steps with relaxed atoms to remove bad contacts with conjugate gradient algorithm. Finally the .coord, .xst, .dcd, .log and .vel files for the system were retrieved. The final minimized structures for the heterodimers and the heterotetramer were obtained using VMD ([18], <http://www.ks.uiuc.edu/Research/vmd/>).

Subunit–subunit interaction studies

To study subunit–subunit interactions, the MD minimized structures were used as input to two different programs namely NACCESS ([17], <http://www.bioinf.manchester.ac.uk/naccess/>) and Dimplot ([44], <http://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/>).

The sequences from potato tuber, maize and rice for the protein subunits were aligned to validate the interface residues listed in subunit–subunit interaction by the two programs. The MSA was performed using ClustalW ([6, 16], <http://www.ebi.ac.uk/Tools/msa/clustalw2/>).

Superimposition of the deduced heterotetrameric structure on the template

The heterotetramer generated by the study was superimposed on the template 1YP2, i.e., the crystal structure of the potato tuber homotetramer available from the Protein Data Bank. The superimposition and RMSD calculation was performed by the UCSF Chimera ([27], <http://www.cgl.ucsf.edu/chimera/download.html>) to determine the accuracy of the modeled complex.

Table 2 Interface residues in the SS of rice with relative positions in the structure of AGPase as listed by NACCESS

Residue	Position	Residue	Position	Residue	Position	Residue	Position	Residue	Position	Residue	Position
CYS	1	TYR	85	PRO	289	TYR	306	VAL	317	TYR	360
GLY	38	GLU	92	PHE	293	LEU	307	THR	318	TYR	361
ALA	39	GLN	115	TYR	294	PRO	308	ASP	319	THR	363
ASN	40	LYS	143	ARG	296	PRO	309	SER	320	GLU	364
TYR	41	PHE	184	ALA	299	SER	310	VAL	321	LYS	367
SER	75	ILE	283	ILE	300	LYS	311	ILE	322	ILE	404
ARG	76	THR	284	TYR	301	VAL	312	GLU	324	ASP	417
ALA	77	LYS	285	THR	302	LEU	313	VAL	327		
GLY	79	LYS	286	PRO	304	ASP	314	ARG	329		
ASN	80	VAL	288	ARG	305	ALA	315	ARG	341		

Table 3 Interface residues in the LS with relative positions in the structure of AGPase of rice as listed by NACCESS

Residue	Position	Residue	Position	Residue	Position	Residue	Position	Residue	Position	Residue	Position
GLY	34	THR	73	ALA	90	PRO	297	ARG	309	ILE	320
GLY	35	LEU	75	ALA	91	PHE	298	LEU	310	SER	321
CYS	36	GLY	76	GLN	93	PHE	299	GLU	311	ASP	322
TYR	37	GLY	77	LEU	282	THR	300	LYS	312	ASP	357
SER	63	ILE	79	THR	283	PRO	302	CYS	313	GLN	358
ALA	64	ASN	80	GLU	284	ARG	303	LYS	314	TYR	359
SER	65	THR	82	GLN	285	TYR	304	ILE	315	THR	361
ASN	67	SER	85	PRO	287	LEU	305	LYS	316	GLU	362
ARG	68	GLN	87	GLU	290	PRO	306	ASP	317		
HIS	71	VAL	88	TYR	292	PRO	307	ALA	318		
HIS	72	LEU	89	PRO	294	ALA	308	ILE	319		

Results

Homology modeled structure of LS and SS

The target sequences of the LS and SS of rice AGPase were retrieved from the NCBI protein database with accession numbers ACJ71342.1 and ACY56071.1, having 518 and 502 amino acids, respectively. 1YP2 (Chain A) was identified as template for both subunits of AGPase in rice using PSI-BLAST. It had 50 % homology with the LS and 87 % with the SS of rice. The model was built using the Swiss Model server. The model for the smaller subunit of AGPase was built from residues 63 to 502, and the large subunit from residues 81 to 518. The starting 62 and 80 residues were removed (leader sequence) from SS and LS, respectively, as they have no homology with the template sequence. The structures generated were assessed using the Protein Structure and Model Assessment Tools at the Swiss Model server, which includes Z-score [3], QMEAN6 score [4] and Dfire_energy [46]. The predicted Z-scores were -0.214 and -1.28 , QMEAN6 scores were 0.748 and 0.659 and Dfire_energy was -631.78 and $-585.02 \text{ kJ mol}^{-1}$ for SS and LS, respectively.

The graphs for SS (Fig. 1a) and LS (Fig. 1b) were obtained for the atomic empirical mean force potential Anolea [26], which assesses the packing quality of the models, the Qmean a composite scoring function for both the estimation of the global quality of the entire model as well as the local per-residue analysis of different regions within a model, and GROMOS—a general-purpose molecular dynamics computer simulation package applied to the analysis of conformations obtained by experiment or by computer simulation. Negative energy values (in green in Fig. 1) represent a favorable energy environment whereas positive values (in red) depict unfavorable energy environments for a given amino acid.

Model evaluation

Models were evaluated with SAVES using its PROCHECK, ERRAT and VERIFY3D options. PROCHECK, which checks the stereochemical quality of a protein structure by analyzing residue-by-residue geometry and overall structure geometry, showed 84.2 % core regions in the LS and 84.9 % core regions in SS. The Ramachandran plot for both subunits (Fig. 2a, b.) depicts the PROCHECK results. ERRAT analyzes the statistics of non-bonded interactions between different atom types; it gave an overall quality factor of 77.70 and 83.565 for LS and SS, respectively. VERIFY3D determines the compatibility of an atomic model (3D) with its own amino acid sequence (1D), and showed 98.85 % and 96.83 % of the LS and SS residues, respectively, to have an average 3D-1D Score >0.2 (Table 1). A status of PASS was obtained at the SAVES, giving the green light to use these models for further studies.

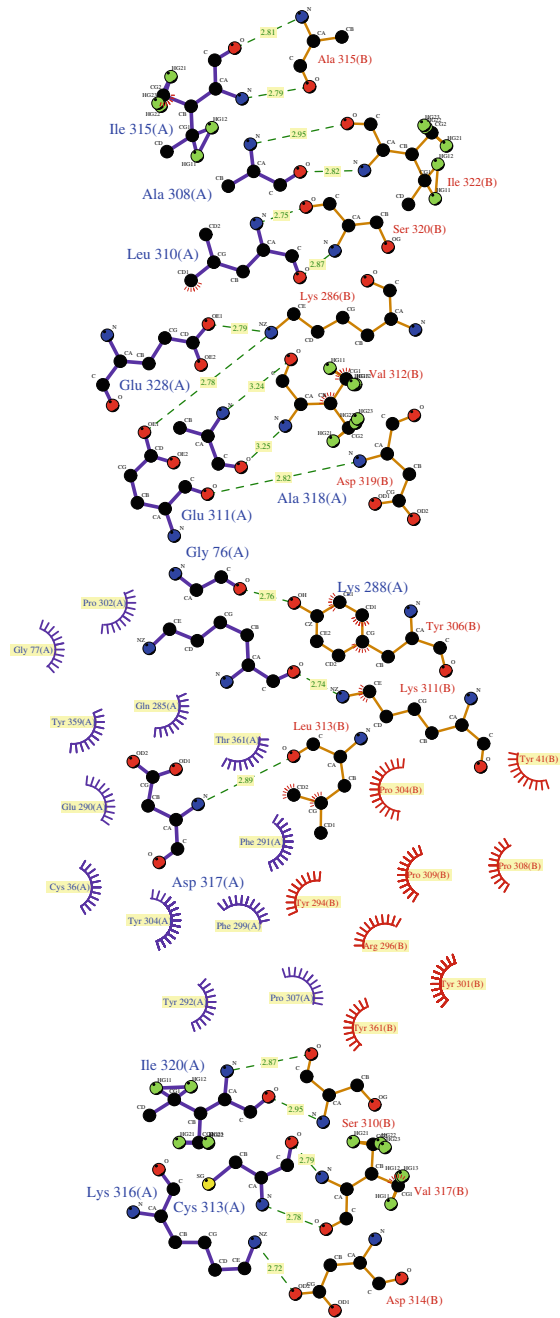
The models generated for the LS and SS were visualized and examined in the UCSF Chimera. The LS (Fig. 3a) had 19 β -strands and 15 α -helices, whereas the SS (Fig. 3b) had 18 β -strands and 15 α -helices.

Docking studies to obtain heterodimer/tetramer orientations

The two subunits were docked with the GRAMM-X docking server. The server performs a rigid body docking and returned an initial heterodimeric orientation (Fig. 4) on docking the LS as a receptor and SS as ligand. Also, the heterodimer was used both as receptor and ligand to obtain an initial heterotetrameric orientation (Fig. 5).

The initial orientation of the heterotetramer from GRAMM-X was refined using the RosettaDock server. The server performs a local docking search. It requires upload of a reasonable starting position, placing the protein partners in near contact (but not overlapping) with the

a



Key

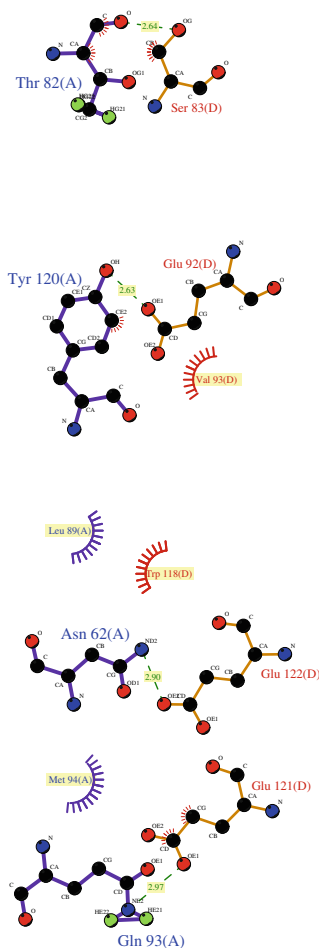
- Residues of first surface
- Residues of second surface
- Hydrogen bond and its length
- ⌋ His 53 Residues involved in hydrophobic contact(s)
- Corresponding atoms involved in hydrophobic contact(s)

tet: chains A and B

Fig. 10a–d Hydrogen and hydrophobic interactions as plotted by Dimplot. **a** Between chain A(LS) and chain B(SS), **b** between chain A(LS) and chain D(SS), **c** between chain B(SS) and chain C(LS), **d**

between chain C(LS) and chain D(SS) of the AGPase heterotetramer. *Dashed lines* Hydrogen bonds, *arcs* hydrophobic interactions

b



Key

- | | | | |
|--|------------------------------|--|--|
| | Residues of first surface | | Residues involved in hydrophobic contact(s) |
| | Residues of second surface | | Corresponding atoms involved in hydrophobic contact(s) |
| | Hydrogen bond and its length | | |

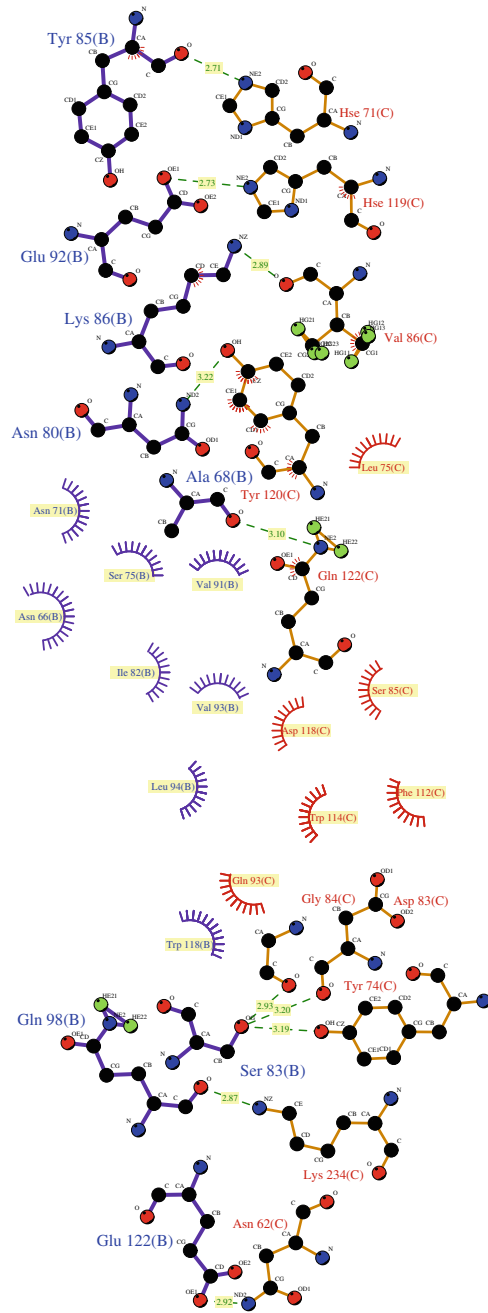
tet: chains A and D

Fig. 10 (continued)






relevant patches of the proteins facing each other. RosettaDock's local perturbation includes $\sim \pm 3$ Å in the direction between the two proteins, ~ 8 Å in the directions sliding the proteins relative to each other along their surfaces, $\sim 8^\circ$ of tilt of the proteins, and a complete 360° spin around the axis between the centers of the two proteins. The server performed 1,000 independent simulations. From this range of random positions, the server returned the ten best-scoring structures from the run in rank order by energy. The

best orientation PDB files with energy scores were retrieved and the minimum energy score structure was taken for further reference (Fig. 6). The bond geometry idealization and removal of unfavorable non-bonded contacts was performed by energy minimization using the GROMOS96 force field in Swiss-Pdb Viewer. The energy of the heterotetramer obtained from docking studies was minimized to $-69,853$ kJ/mol $^{-1}$ using the GROMOS96 force field.

c



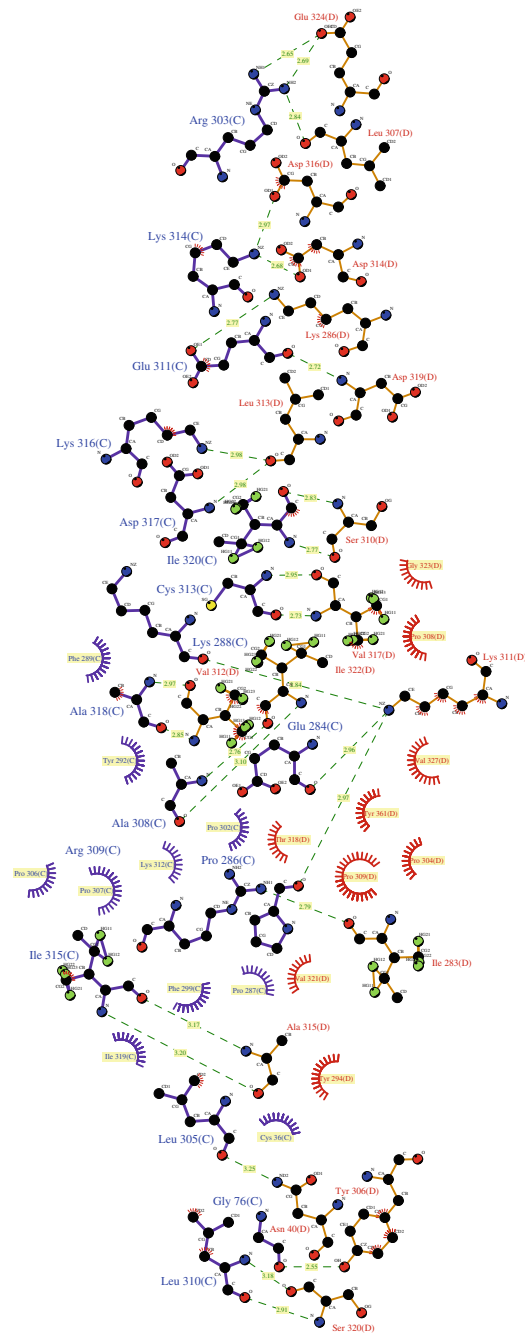
Key

-  Residues of first surface
-  Residues of second surface
-  Hydrogen bond and its length
-  His 53 Residues involved in hydrophobic contact(s)
-  Corresponding atoms involved in hydrophobic contact(s)






tet: chains B and C

Fig. 10 (continued)

d



Key

- | | | | |
|---|------------------------------|---|--|
|  | Residues of first surface |  | His 53 Residues involved in hydrophobic contact(s) |
|  | Residues of second surface |  | Corresponding atoms involved in hydrophobic contact(s) |
|  | Hydrogen bond and its length | | |

tet: chains C and D

Fig. 10 (continued)

Table 4 Predicted hydrogen bonds and hydrophobic interactions in dimer–dimer interface studies in AGPase of rice

Chains	Hydrogen bonds	Hydrophobic contacts
AB	19	28
AD	4	15
BC	10	19
CD	27	38

MD simulations of possible heterodimers/tetramer and analysis of interactions between SS and LS of potato AGPase

An explicit solvent MD simulation for the representative structure of AGPase heterodimer/heterotetramer from the RosettaDock server was performed using NAMD software employing the CHARMM force field. Two heterodimer orientations are possible in the heterotetramer, one in which the two subunits have a head-to-head interaction (Dimer 2) as in chains A and D, and the second having tail-to-tail subunit interactions (Dimer 1) as in chains A and B of the heterotetramer as shown in Fig. 5. MD simulations were performed on both heterodimers and on the heterotetramer from the RosettaDock server results. The energy profiles from the .log file of the simulation were extracted and TOTAL energy was obtained. Dimer 1 (Fig. 7) was found to be more stable with less energy ($-437,986 \text{ kcal mol}^{-1}$) than Dimer 2 (Fig. 8; $-396,097 \text{ kcal mol}^{-1}$). The heterotetramer (Fig. 9) was also minimized to a minimum energy state of $-767,011 \text{ kcal/mol}^{-1}$ by the end of the simulation run. These minimized energy structures were used to study interactions in LS and SS. The NACCESS listed the residues having $>1 \text{ \AA}^2$ decrease in the accessible surface area (ASA) value on complex formation as interface residues, numbering a total of 57 residues in SS and 63 in LS (Tables 2, 3). The results showed that the LS had 31 hydrophobic and 32 hydrophilic residues at the interface, whereas SS had 30 hydrophobic and 27 hydrophilic residues. Dimplot plotted the hydrogen bonds and hydrophobic contacts between the four chains of the heterotetramer. The resulting postscript files were retrieved and analyzed (Fig. 10a–d). Table 4 shows the summary of predicted interactions.

The residues involved in interactions as plotted by DIMPLOT were also listed by NACCESS. These residues were validated by multiple sequence alignment (MSA) of sequences for the two subunits from potato tuber, rice and maize using ClustalW. The MSA results (Tables 5, 6) showed that almost all the residues of the interacting motifs in the two subunits were conserved in the three species.

Superimposition and RMSD calculation

The heterotetramer generated by the study was superimposed on the template 1YP2; crystal structure of the potato tuber homotetramer available at PDB. The superimposition (Fig. 11) showed a root mean square deviation (RMSD) of 1.093 \AA .

Discussion

A homology modeling approach was applied to obtain the structure of rice AGPase LS and SS using the structure of potato SS AGPase as a template, since they have a sequence identity over 50 %. The similarity between the template and target, which was 50 % for LS and 87 % for SS, was found to be optimum for homology modeling. If the target and the template sequence share more than 50 % sequence identity, predictions are of very good to high quality and have been shown to be as accurate as low-resolution X-ray predictions [22, 43]. The model for the smaller subunit of AGPase was built from residues 63 to 502, and that of the large subunit used residues 81 to 518; the starting residues were removed as they have no homology with the template sequence. Previous studies have indicated these sequences to be the leader sequences [12]. Tuncel et al. [40] also modeled the LS and SS of potato AGPase from 11th and 71st position, respectively, thus also removing the leader sequence with no homology to the template.

Protein–protein docking is the only computational approach that directly models physical interactions between proteins [38, 41]. High computational complexity restricts the flexible docking algorithms and is rarely applicable to

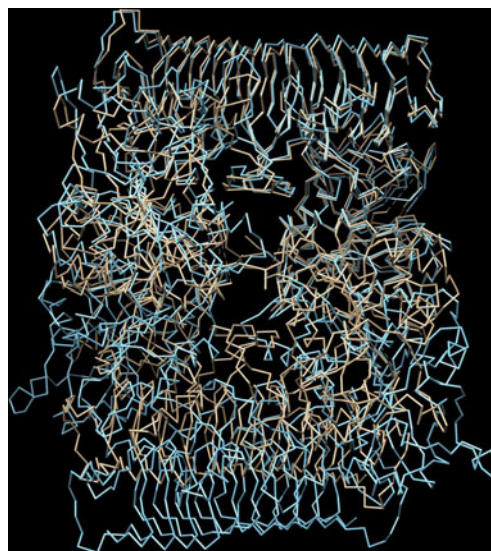


Fig. 11 Superimposed structures of the predicted heterotetramer (blue) and 1YP2 (brown). The root mean square deviation (RMSD) of 1.093 \AA indicated that the structures are similar

Table 5 Multiple sequence alignment (MSA) of sequences of SS of AGPase of potato, rice and maize using ClustalW

CLUSTAL 2.1 multiple sequence alignment		
POTATO SMALL SUBUNIT	-----	
RICE SMALL SUBUNIT	MAMMAMGAASWAPIAPAPARAAAFFYPGRDLAAARRRRGAAAAARRPFVFT	50
MAIZE SMALL SUBUNIT	-----MDMALASKASPPPSNATTAEQLI	23
POTATO SMALL SUBUNIT	-MAVSDSQNSQTCCLDPDASRSVLGIILGGGAGTRLYPLTKKRAKPAVPLG	49
RICE SMALL SUBUNIT	PRAVSDRSRSQTCCLDPDASTSVLGIILGGGAGTRLYPLTKKRAKPAVPLG	100
MAIZE SMALL SUBUNIT	PKRDKAAANDSTYLNPPQAHDSVLGIILGGGAGTRLYPLTKKRAKPAVPLG	73
	. : . . . * * : * : * *****	
POTATO SMALL SUBUNIT	ANYRLIDI PVSNCNLSNISKIYVLTQFNASLNRHLSRAYASNMGYKNE	99
RICE SMALL SUBUNIT	ANYRLIDI PVSNCNLSNVSKIYVLTQFNASLNRHLSRAYGNINISGYKNE	150
MAIZE SMALL SUBUNIT	ANYRLIDI PVSNCNLSNISKIYVLTQFNASLNRHLSRAYGNSNIGGYKNE	123
	***** : ***** * * * * * : . . . : *****	
POTATO SMALL SUBUNIT	GFVEVLAQQSPENPDWFQGTADAVRQYLWLFEEHTVLEYLILAGDHLYR	149
RICE SMALL SUBUNIT	GFVEVLAQQSPENPNWFQGTADAVRQYLWLFEEHNVMFLILAGDHLYR	200
MAIZE SMALL SUBUNIT	GFVEVLAQQSPDNPNWFQGTADAVRQYLWLFEEHNVMFLILAGDHLYR	173
	***** * * * * * : * : ***** * * * * * . : * : *****	
POTATO SMALL SUBUNIT	MDYEKFIQAHRETDADITVAALPMDEKRATAFGLMKIDEEGRIEEEFAEKP	199
RICE SMALL SUBUNIT	MDYQKFIQAHRETDADITVAALPMDEERATAFGLMKIDDEGRIEEEFAEKP	250
MAIZE SMALL SUBUNIT	MDYEKFIQAHRETNADITVAALPMDEKRATAFGLMKIDEEGRIEEEFAEKP	223
	*** : ***** : ***** : ***** : *****	
POTATO SMALL SUBUNIT	QGEQLQAMKVDTTILGLDDKRAKEMPFIASMGIYVISKDVMLNLLRDKFP	249
RICE SMALL SUBUNIT	KGEKLSMMVDTTILGLDTERAKELPYIASMGIYVFSKDVMLKLLRQNF	300
MAIZE SMALL SUBUNIT	KGEQLKAMVDTTILGLDDVRAKEMPFYIASMGIYVFSKDVMLQLLREQFP	273
	: * : * : * ***** * * * * * : * : ***** : ***** : * * : * *	
POTATO SMALL SUBUNIT	GANDFGSEVIPGATSLGMRVQAYLYDGYWEDIGTIEAFYANANLGITKKPV	299
RICE SMALL SUBUNIT	AANDFGSEVIPGATEIGMRVQAYLYDGYWEDIGTIEAFYANANLGITKKPV	350
MAIZE SMALL SUBUNIT	EANDFGSEVIPGATSIGKRQAYLYDGYWEDIGTIAAFYANANLGITKKPM	323
	***** : * ***** ***** ***** :	
POTATO SMALL SUBUNIT	PDFSFYDRSAPIYTPRYLPPSKMLDADVTDSVIGEGCVIKNCKIHHSV	349
RICE SMALL SUBUNIT	PDFSFYDRSAAIYTPRYLPPSKVLDADVTDSVIGEGCVIRHCTINHSV	400
MAIZE SMALL SUBUNIT	PDFSFYDRFAPIYTPRHLPPSKVLDADVTDSVIGEGCVIKNCKINHSV	373
	***** * . ***** : ***** : ***** : * : * : * * *	
POTATO SMALL SUBUNIT	GLRSCISEGAIIEDSLLMGADYETDADRKLLAAKGSVPPIGIGKNCHIKR	399
RICE SMALL SUBUNIT	GLRSCISDGAVIDESLLMGADYYETDCKALSETGGIPIGIGKNAHIRK	450
MAIZE SMALL SUBUNIT	GLRSCISEGAIIEDSLLMGADYYETDADKLLAEKGGIPIGIGKNCSIRR	423
	***** : * : ***** : * : * * : * : * : * * * * * . * : :	
POTATO SMALL SUBUNIT	AIIDKNARIGDNVKIINKDNVQEAARETDGYFIKSGIVTVIKDALIPSGI	449
RICE SMALL SUBUNIT	AIIDKNARIGENVKIINVDNIQEASRETDGYFIKSGIVTVIKDALIPSGT	500
MAIZE SMALL SUBUNIT	AIIDKNARIGDNVKVFQTDVPTVKNINLDFSLGYYIILDMPFCPLIFNIY	473
	***** : * * : * : * : * : * : * : * : * * .	
POTATO SMALL SUBUNIT	II	451
RICE SMALL SUBUNIT	VI	502
MAIZE SMALL SUBUNIT	F-	474
	.	
Blue: leader sequence; red residue: listed in literature; pink: as listed from our study. * conserved residues.		

practical protein docking at present. This problem can be overcome by using a rigid body docking algorithm [45]. Thus, LS as receptor and SS as ligand were docked with the GRAMM-X server, returning an initial orientation of the

heterodimer as well as the heterotetramer, which were further refined.

Simulations aid our understanding of biochemical processes and give a dynamic dimension to structural data. MD

Table 6 MSA of sequences of LS of AGPase of potato, rice and maize using ClustalW

CLUSTAL 2.1 multiple sequence alignment		
POTATO	-----NLIKPGVAVSVI	12
RICE	MQFMPLDTNACAQPMRRAGEGAGTERLMERLNI GGMTQE KALRKR CFGDGV TGTARCVF	60
MAIZE	--MQFALALDTNSGPHQIRSCEGDGIDRLEKLSIGGRKQE KALRNR CFGGRVAATTQCIL	58
	. . . : :	
POTATO	TTENDTQTVFVDMRPLERRRANPKDVAAVILGGEGTKLFPLTS R TATPAVPVGG CY RLI	72
RICE	TSDADRDTPLHRTQSSRKNY ADASHVSAVILGGGTGVQLFPLTSTRATPAVPV GGCY RLI	120
MAIZE	TSDACPETLHSQTQSSRKNYADANRVSAIILGGGTGSQLFPLTSTRATPAVPVGGCYRLI	118
	:: : . . : * : : * : * : * : * : * : * : * : * : * : * : *	
POTATO	DIPMSNCINSAINKIFVLTY NSAPLN RHI ARTYF GNV SFGDGF VE VLAA T QTP GEAGK	132
RICE	DIPMSNCFNSGINKIFVMTQ FNSASL NRH HHTY L GGGIN F TDGSV Q VLAA T QMP DEP-A	179
MAIZE	DIPMSNCFNSGINKIFVMSQFNSTSLNRH I HRTYLEGGINFADGS VQ VLAA TQMP EEP-A	177
	*****:*.*****:*.**:.***** :*: .*: * BB * :***** B *	
POTATO	KWFQGTADAVRK F I WV FEDA AKN -KNIENIVLSGDHLYRMDYELVQNHDNRADITLSC	191
RICE	GWFQGTADAIKRFMWILEDHYNQNNIEHVILCGDQLYRMNYMELVQKHVDDNADITISC	239
MAIZE	GWFQGTADSIRKFIWVLEDYSHKSIDNIVILSGDQLYRMNYMELVQKHVEDDADITISC	237
	*****:***:*.**:** . :*:**:* :*.**:***:*****:* : : *****:*	
POTATO	APAEDSRASDFGLVKIDSRGRVQFAEKPKGFDL KAMQ VD TTLV GLSPQD AKKSP YI AS M	251
RICE	APIDGSRASGYGLVKFDDSRGVIQFLEKPEGADLESMKVDTSFLSYAIDDKQKYPIYI AS M	299
MAIZE	APVDESRAKNGLVKIDHTGRVLQFF EKPKG ADLNSMRVETN FLSYA IDDAQKYPIY AS M	297
	** : **** *****:* **:* ** **:* * * :*:***:* : : : * : * * : * : * : *	
POTATO	GVYVFKTDVLLKLLKWSYPTSNDFGSEIIPAAIDYIN VQ AYIFKDYWEDIGTIKSFF YN AS	311
RICE	GIYVLKKDVLLDILKSKY AHLQ DFGSEI LP RAVLEHNVKACV FT EYWEDIGTIKSFF D AN	359
MAIZE	GIYVFKKDALDLKSKY TQL HDFGSEI LP RAVLDH SVQ ACI FT GYWEDVGTIKSFF D AN	357
	*:***:*.*:**:* :* . * :*****:* * : :*:*** :* . *****:*****:* :	
POTATO	L ALT Q E FP EFQ F Y DPK TP FP Y T SR F L P P T K ID N CK I K D A I I S H G C F L R D C S V E H S I V G E R	371
RICE	L ALT Q P PP K F E F Y DPK T PF FT S P R L PP AR L E K CK I K D A I I S D G C S F S E C T I E H S V I G I S	419
MAIZE	LALTEQ PS K F D F YDPK T PF T AP R CL P P T Q L D K CK M K Y A F I S D G CL L REC N I E H S V I G V C	417
	*****: . *:*****:*:* ***: : : : : **:* * :* . * : : * : * : * : *	
POTATO	SRLDCGV L KDT F MMG ADY Y Q T E SEIAS L LA E G K V P IG I GEN T K I R K C I I D NA I G K N V	431
RICE	SRV S I G CEL K D T MM G AD Q Y T E E TS K LL F E G K V P I G I GEN T K I R N C I I D M N A R I G R N V	479
MAIZE	SRVSSGCEL K DSVMMG AD TY E TE E E A SK L LL L AG K V P V G I G R N T K I R N C I I D M N A R I G K N V	477
	:. * ***:***** *:*:* :*.** *****:***.*****:***** **:*:***	
POTATO	SIINKD G VQEAD R PEEGFYIRSGII I L E KATIR D G T VI	470
RICE	IIANTQ G VQE S D H PEEGYIRSGIVV I L K NATIK D G T VI	518
MAIZE	VITNS K GIQE A D H PEEGYIRSGIVV I L K NATIN D G S VI	516
	* * . * : * : * : * : * : * : * : * : * : * : * : * : * : * : *	
Blue: leader sequence; red residue: listed in literature; pink: as listed from our study. * conserved residues.		

simulation is used to bring biomolecular structures alive, giving insights into natural dynamics on different timescales of biomolecules in solution [15]. Thus MD simulations were performed on both the heterodimers and the heterotetramer from the RosettaDock server results using NAMD. Two possible heterodimer orientations are possible in the heterotetramer: one in which the two subunits have a head-to-head interaction and a second having tail-to-tail subunit interactions. MD studies revealed the tail-to-tail

interaction to be more stable compared to the head-to-head interacting dimer. A similar result was also obtained by Baris et al. [2] for AGPase dimers in potato using a similar approach. Further, the energy-minimized heterotetramer was also obtained from MD simulations with a minimized energy of $-767,011 \text{ kcal mol}^{-1}$. Subunit–subunit interactions were predicted. NACCESS listed the interface residues with the criteria that the residues showing a decrease of $> 1 \text{ \AA}^2$ in the ASA on complex

formation are present at the interface of the subunits, as taken into consideration by Tuncel et al. [40] for similar work on potato AGPase. Dimplot, part of the Ligplot v.4.5.3 suite, was used to plot interactions between the four chains of the heterotetramer on a Linux machine. Dimplot studies the dimer interface and plots the hydrophobic interactions and hydrogen bonds. Chains A and B showed 19 hydrogen bonds and 28 hydrophobic contacts, whereas chains A and D had only 4 hydrogen bonds and 15 hydrophobic contacts between them. Similarly, chains B and C had 10 hydrogen and 19 hydrophobic interactions. The maximum interactions of 27 hydrogen bonds and 38 hydrophobic contacts were observed between chains C and D. A similar approach was used by Danishuddin et al. [9] to study interactions in wheat AGPase. The results of NACCESS and Dimplot were similar. Next the listed residues were validated by MSA of the LS and SS sequences from potato tuber, rice and maize. The results of MSA showed that a lot of residues predicted at the interface were similar to those predicted by Tuncel et al. [40] for the LS and Jin et al. [20] for the SS, and almost all the residues of the interacting motifs in the two subunits were conserved in the three species. As the criteria for the accuracy of the determined structures was proposed [29], for proteins with close homologous templates, most predicted structures have a RMSD of 1–2 Å from the experimental structure, which in some cases achieves the accuracy of medium-resolution NMR or low-resolution X-ray structures. Consequently, the final structure was superimposed on the template structure (1YP2) and the RMSD was calculated to be 1.093 Å, illustrating the structures to be similar.

Conclusions

In the present study, we predicted the structure of the LS and SS of rice AGPase using comparative modeling. Further, the orientation of the dimer and the heterotetramer were determined using protein–protein docking studies, which were refined using MD simulations. Interactions leading to assembly of the AGPase heterotetramer in rice were predicted using two approaches and validated by MSA. The structure of the AGPase heterotetramer of rice deduced using various modeling and docking tools along with MD simulations illustrated that the model obtained in rice was similar to the AGPase homotetramer crystal structure of potato (template) as evident by the RMSD value of superimposition of the two. Subunit interaction studies by NACCESS and Dimplot listed the residues at the interface, which were found to be conserved in the AGPases of all three crops. Identification of these critical amino acids between LS and SS will pave the way to further engineering of the enzyme to improve plant starch yield.

Acknowledgments The authors gratefully acknowledge the help rendered by Aytug Tuncel and Ibrahim Halil Kavakli, Department of Chemical and Biological Engineering, College of Engineering, Koc University, Rumeli Feneri Yolu, Istanbul, Turkey for MD simulations. The funding provided by the Department of Biotechnology, New Delhi and Indian Council of Agricultural Research, New Delhi for computing facilities is acknowledged.

References

- Altschul SF, Madden TL, Schaffer AA (1997) Gapped BLAST and PSIBLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402 <http://blast.ncbi.nlm.nih.gov/BLAST.cgi>
- Baris I, Tuncel A, Ozber N, Keskin O, Kavakli IH (2009) Investigation of the interaction between the large and small subunits of potato ADP-glucose pyrophosphorylase. *PLoS Comput Biol* 5:e1000546
- Benkert P, Biasini M, Schwede T (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* 27(3):343–350
- Benkert P, Tosatto SCE, Schomburg D (2008) QMEAN: a comprehensive scoring function for model quality assessment. *Proteins Struct Funct Bioinformatics* 71(1):261–277
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242 <http://www.pdb.org>
- Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res* 31:3497–3500 <http://www.ebi.ac.uk/Tools/msa/clustalw2/>
- Colovos C, Yeates TO (1993) Verification of protein structures: patterns of non-bonded atomic interactions. *ProteinSci* 2:1511–1519 <http://nihserver.mbi.ucla.edu/SAVES>
- Cross JM, Clancy M, Shaw JR, Boehlein SK, Greene TW (2005) A polymorphic motif in the small subunit of ADP-glucose pyrophosphorylase modulates interactions between the small and large subunits. *Plant J* 41:501–511
- Danishuddin M, Chatrath R, Singh R (2011) Insights of interaction between small and large subunits of ADP-glucose pyrophosphorylase from bread wheat (*Triticum aestivum* L.). *Bioinformation* 6(4):144–148
- Eisenberg D, Luthy R (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol* 277:396–404 <http://nihserver.mbi.ucla.edu/SAVES>
- Frueauf JB, Ballicora MA, Preiss J (2003) ADP-glucose pyrophosphorylase from potato tuber: site-directed mutagenesis of homologous aspartic acid residues in the small and large subunits. *Plant J* 33:503–511
- Georgelis N, Shaw JR, Hannah CL (2009) Phylogenetic analysis of ADP-glucose phosphorylase subunits reveals a role of subunit interfaces in the allosteric properties of the enzyme. *Plant Physiol* 151:67–77
- Giroux MJ, Shaw J, Barry G, Cobb BG, Greene T, Okita T, Hannah LC (1996) A single gene mutation that increases maize seed weight. *Proc Natl Acad Sci USA* 93:5824–5829
- Guex N, Peitsch MC (1996) Swiss-PdbViewer: a fast and easy-to-use PDB viewer for Macintosh and PC. *Protein Data Bank Q Newslett* 77: 7 <http://spdbv.vital-it.ch>
- Hansson T, Oostenbrink C, van Gunsteren W (2002) Molecular dynamics simulations. *Curr Opin Struct Biol* 12(2):190–196
- Higgins D, Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific

- gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680 <http://www.ebi.ac.uk/Tools/msa/clustalw2/>
17. Hubbard SJ, Thornton JM (1993) ‘NACCESS’, Computer Program. University of College London, London; <http://www.bioinf.manchester.ac.uk/naccess/>
 18. Humphrey W, Dalke A, Schulten K (1996) VMD—Visual Molecular Dynamics. *J Mol Graphics* 14: 33–38; <http://www.ks.uiuc.edu/Research/vmd/>
 19. Hwang SK, Salamone PR, Okita TW (2005) Allosteric regulation of the higher plant ADP-glucose pyrophosphorylase is a product of synergy between the two subunits. *FEBS Lett* 579:983–990
 20. Jin XS, Ballicora MA, Preiss J, Geiger JH (2005) Crystal structure of potato tuber ADP-glucose pyrophosphorylase. *EMBO J* 24:694–704
 21. Kavakli IH, Park JS, Slattery CJ, Salamone PR, Frohlick J, Okita TW (2001) Analysis of allosteric effectors binding sites of potato ADP-glucose pyrophosphorylase through reverse genetics. *J Biol Chem* 276(44):40834–40840
 22. Kopp J, Schwede T (2004) Automated protein structure homology modeling: a progress report. *Pharmacogenomics* 5(4):405–416
 23. Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 26: 283–291 <http://nihserver.mbi.ucla.edu/SAVES>
 24. Lyskov S, Gray JJ (2008) The RosettaDock server for local protein-protein docking. *Nucleic Acids Res* 36: W233–W238; <http://rosettaserver.graylab.jhu.edu/>
 25. MacKerell AD, Bashford D, Bellott M et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616
 26. Melo F, Feytmans E (1998) Assessing protein structures with a non-local atomic interaction energy. *J Mol Biol* 277(5):1141–1152
 27. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 25: 1605–1612; <http://www.cgl.ucsf.edu/chimera/download.html>
 28. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E (2005) Scalable molecular dynamics with NAMD. *J Comput Chem* 26: 1781–1802; <http://www.ks.uiuc.edu/Research/namd/>
 29. Read RJ, Chavali G (2007) Assessment of CASP7 predictions in the high accuracy template-based modeling category. *Proteins* 69:27–37
 30. Sakulsingharoj C, Choi SB, Hwang SK, Bork J, Meyer CR, Edwards GE (2004) Engineering starch biosynthesis for enhanced rice yields: the role of the cytoplasmic ADP-glucose pyrophosphorylase. *Plant Sci* 167:1323–1333
 31. Salamone PR, Greene TW, Kavakli IH, Okita TW (2000) Isolation and characterization of a higher plant ADP-glucose pyrophosphorylase small subunit homotetramer. *FEBS Lett* 482:113–118
 32. Schwede T, Ko J, Guex N, Peitsch MC (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res* 31:3381–3385; <http://swissmodel.expasy.org/>
 33. Shaw JR, Hannah LC (1992) Genomic nucleotide sequence of a wild-type *shrunken-2* allele of *Zea mays*. *Plant Physiol* 98:1214–1216
 34. Slattery CJ, Kavakli IH, Okita TW (2000) Engineering starch for increased quantity and quality. *Trends Plant Sci* 5:291–298
 35. Smith-White BJ, Preiss J (1992) Comparison of proteins of ADP-glucose pyrophosphorylase from diverse sources. *J Mol Evol* 34:449–464
 36. Sowokinos JR, Preiss J (1982) Pyrophosphorylases in *Solanum tuberosum*. III. Purification, physical, and catalytic properties of ADP-glucose pyrophosphorylase in potatoes. *Plant Physiol* 69:1459–1466
 37. Stark DM, Timmerman KP, Barry GF, Preiss J, Kishore GM (1992) Role of ADP-glucose pyrophosphorylase in regulating starch levels in plant tissues. *Science* 258:287–292
 38. Sternberg MJE, Gabb HA, Jackson RM (1998) Predictive docking of protein-protein and protein-DNA complexes. *Curr Opin Struct Biol* 8:250–256
 39. Tovchigrechko A, Vakser IA (2006) GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res* 34: W310–W314; <http://vakser.bioinformatics.ku.edu/resources/gramm/grammx>
 40. Tuncel A, Kavakli IH, Keskin O (2008) Insights into subunit interactions in the heterotetrameric structure of potato ADP-glucose pyrophosphorylase. *Biophys J* 95:3628–3639
 41. Vajda S, Sil M, Novotny J (1997) Empirical potentials and functions for protein folding and binding. *Curr Opin Struct Biol* 7:222–228
 42. van Gunsteren WF, Billeter SR (1996) Biomolecular simulations: the GROMOS96 manual and user guide. ETHZ, Zurich
 43. Vitkup D, Melomud E, Moulton J, Sander C (2001) Completeness in structural genomics. *Nat Struct Biol* 8(6):559–566
 44. Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Prot Eng* 8: 127–134; <http://www.ebi.ac.uk/thornton-srv/software/LIGPLOT/>
 45. Wenfan, Huang (2005) Rigid Body Protein Docking by Fast Fourier Transform. Honor year project report, Department of Computer Science, School of Computing, National University of Singapore
 46. Zhou H, Zhou Y (2002) Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Sci* 11:2714–2726