

Structure Prediction of Potato Large Subunit ADP – Glucose Pyrophosphorylase

Aytug Tuncel

Graduate Program in
Computational Sciences and
Engineering, Koc
University, Rumelifeneri
Yolu, Sariyer, 34450,
Istanbul, Turkey

I.Halil Kavakli*

Department of Chemical and
Biological Engineering, Koc
University, Rumelifeneri
Yolu, Sariyer, 34450,
Istanbul, Turkey

Ozlem Keskin*

Department of Chemical and
Biological Engineering, Koc
University, Rumelifeneri
Yolu, Sariyer, 34450,
Istanbul, Turkey

*Corresponding Authors

Abstract - *ADP-glucose pyrophosphorylase, a key enzyme involved in the biosynthesis of starch in higher plants, is composed of pairs of large (LS) and small subunits (SS). Its catalytic activity is activated by 3-phosphoglyceric acid (3-PGA) and inhibited by inorganic phosphate (Pi). Current evidence indicates that the two subunit types play distinct roles in enzyme function. The large subunit is involved mainly in the allosteric regulation through its interaction with the catalytic small subunit. Recently, crystal structure of small subunit homotetramer has been solved. Comparison of LS and SS primary sequences reveals that, there is a 53% identity between LS and SS. Such a high homology between the subunits allows us modeling the 3-D structure of potato AGPase LS and to identify the possible interaction sites between the subunits.*

Keywords: ADP-glucose pyrophosphorylase, subunit-subunit interaction, homology modeling

1 Introduction

ADP-glucose pyrophosphorylase (AGPase) plays a dominant role in the biosynthesis of starch both in photosynthetic and non photosynthetic plant tissues. It catalyzes the formation of ADP-glucose using ATP and Glucose-1 phosphate as substrates. Later ADP-glucose will serve as a glucosyl donor for the next enzyme, starch synthase, in the pathway. Plant AGPase is a heterotetramer, composed of two distinct subunits, large (LS) and small (SS) subunits, encoded by different gene [1-3]. Comparison of cDNA of SSs from various plants showed that SSs are highly conserved (80-90%), whereas LSs from different plant species are less conserved (50-60%) [4,5]. Although the SS and LS share considerable identity at the primary sequence level, current evidence suggests that these

subunits play distinct roles in AGPase function. A missense mutation in the LS gene in *Arabidopsis* leaves causes only a partial reduction of AGPase activity and starch level [6]. The isolated enzyme, which is composed of only the SSs, required much higher levels of 3-PGA for activation compared to wild type (WT) *Arabidopsis* leaf AGPase [6]. The different roles of the LS and SS in AGPase function are also supported by both random and site directed mutagenesis [7-10]. In these studies cDNA of both potato LS and SS are subjected to mutagenesis and characterized at biochemical level. Analysis of the mutants for both subunits indicated that the LS serves primarily as a regulator, while the SS plays a major role in catalytic activity of the enzyme.

Biochemical characterizations of all AGPases that are located in chloroplast and some others, located in nonphotosynthetic tissues, show that they are allosterically regulated by the levels of 3-phosphoglyceric acid (3-PGA) and inorganic phosphate (Pi) [3]. AGPase in maize endosperm [11], potato tuber [12], rice endosperm [13] and *Arabidopsis* leaf [14] are activated by 3-PGA and inhibited by Pi. On the other hand it has been reported that AGPase from barley and wheat appear to be insensitive to both 3-PGA and Pi.

Recently, recombinant SS of ADP-glucose pyrophosphorylase purified from *E.coli* has been crystallized [15]. Its structure has been solved within the resolution range 2.1 to 2.6 Å in the absence effectors and substrate and also in the presence of ATP and glucose-1 phosphate. Since there is a 53% homology between LS and SS, we have used the computational tools to predict the 3-D structure of potato LS AGPase and possible interaction sites with the SS. First, we made a three dimensional model of the potato LS AGPase using Swiss-Model. RMSD value between the predicted LS and SS obtained from x-ray crystallography is 1.35 Å. This value suggests that our predicted structure of LS is within the confident range to

probe interaction sites with SS. Next, we used NACCESS software to identify the key residues that are important in subunit - subunit interaction. The hypothetical model of the tetramer (composed of two LSs and two SSs) is further discussed.

2 Materials and Methods

2.1 Sequence Alignment

The sequence alignment of the LS and SS was obtained using CLUSTAL W (1.83) [16] with default parameters (Alignment algorithm is full, Kimura correction is off, score type is percent, ignore gaps in alignment is off). Due to the presence of chloroplast target sequences at the N-terminus of both subunits, alignment started from the 11th position (Val) and ended with the 470th amino acid (Ile) for LS and started from 71st position (Lys) and ended with 521st position (Ile) for SS.

2.2 RMSD Calculations

RMSD values were calculated by VMD [25] software using the backbone atoms. During the 3D model construction of the LS some regions from its N terminus were truncated. Differences in the number of residues between the LS and SS made it necessary to first align the sequences. Prior to calculation of RMSD values the two structures were superimposed and the pairs of residues to be superimposed were obtained from pairwise sequence alignment. The alignment was done by CLUSTAL W (1.83). Then, RMSD values were calculated based on the superimposed structures.

2.3 3D Models

To construct a 3-D structure of potato tuber LS AGPase following servers were used; CPH models 2.0 [17], EpyPred3D [18], Geno3D [19], Swiss Model [20-22] and Robetta [23, 24]. The methods used by these servers for construction of a model are mainly based on homology modeling. Since the SS and LS share 53 % identity [5], the SS subunit of the potato tuber AGPase used as a template to construct the 3D model of the potato tuber LS AGPase. RMSD values between the models and SS were calculated and no significant differences were found. In order to find the best fitting model with SS we followed such a routine: RMSD calculations between the hypothetical LS models and SS were done and the model that has the closest RMSD value to the average was chosen. First the average RMSD values of G3D – 1 and G3D – 2 and five of the Robetta models were calculated to provide

equal contribution of models to the total average RMSD. Then, the average of these five values was calculated as 1, 62. Therefore, we have taken Swiss Model structure as our 3-D model for further calculations, since it has the closest RMSD (1.35) to total average.

2.4 Construction of the Tetramer Model

The crystal structure of potato tuber SS₄ AGPase has a resolution of 2.1 Å with an approximate 222 symmetry [15]. The catalytic domain of the enzyme consists of β sheets covered by α helices which is also present in LS. The β sheet region between the amino acids 321 and 334 makes close contact with the nearby identical subunit in homotetrameric structure. The similarity between the predicted LS models and the SS, based on the RMSD values, suggests a similar dimerization. For construction of the heterotetrameric (α₂β₂) enzyme we first aligned the monomeric LS with chain A of homotetrameric SS₄ structure and deleted the chain A using VMD software. Same procedure was applied to the other chains, B, C and D, to obtain the possible interaction ways of LS and SS.

2.5 Identification of Interface Residues

Identification of the surface was performed by using the CONSURF [26] server. The server takes a PDB formatted file and searches for the homologues and aligns them. By constructing a phylogenetic tree it calculates the conservation score for each amino acid, assigning an integer value to score between 1 and 9 (the higher the score the more conserved is the residue). Maximum Likelihood method was used and the homologues templates were collected from Swiss-Prot. Conservation score was adjusted to 9 in order to obtain highly conserved residues. Number of Psi-Blast iterations was 1 and E value cut off was 0.001. Then NACCESS [27], software that calculates both the atomic and residue level surface accessible area based on the method developed by Lee & Richards [28], was used to identify the interface residues. Parameters were default, 1.4 Å probe radius and 0.05 Å z-slices. For each dimeric interaction possibility, right-left, diagonal-diagonal and top-bottom, we calculated the accessibility of the residues for LS, SS and LS-SS complex separately. Any residue that showed >1Å² decrease in absolute accessibility [29] and >5% decrease in relative accessibility [30] when dimerized compared to its monomeric form, was considered to be at interface.

3 Results and Discussion

It was shown previously that the crystal structure of some proteins may serve as good templates for modeling structurally unknown proteins, which they share significant amount of homology. It was therefore thought that a similar approach can be applied for modeling the structure of potato tuber AGPase LS using the structure of potato SS AGPase as a template, since they have a sequence identity over 50% [4, 5].

Based on the high sequence identity between the LS and SS primary sequences, the crystal structure of SS (PDB ID 1yp2) is used as a template for modeling LS. For the modeling of the LS five different modeling servers were used, namely: CPH, EasyPred, Geno3D, Swiss Model and Robetta and ten different theoretical models were obtained from these servers. The results for the comparison of RMSD values of the structures with each other and with the crystal structure of SS (1yp2, 2.11 Å resolution) (Table 1) show that all the models are very similar to each other (they all have RMSD values less than 4.0 Å based on the backbone atoms). Therefore, all the 3D structures of SS from

these different servers returned similar models, which are in broad agreement with the structure of SS, obtained by crystallographic methods.

The predicted LS structure obtained from Swiss Model server was superimposed with SS as shown in Figure 1. Both the high sequence similarity and low RMSD values between the predicted LS structure and the SS along with superimposed data validate the predicted LS structure for further analysis. Comparisons of two structures reveal that there are two regions that are mainly different (shown in Figure 1). These regions, region 1 and region 2, correspond to the loop structures. Primary sequence alignment has indicated that these regions are less conserved between the LS and the SS in amino acid level (Table 2) as well. Differences in these regions in both structural and residue level may be important in communication between the LS and SS upon 3-PGA binding to the LS.

Table 1 RMSD values in Å calculated by VMD software. Ten different 3D models were obtained from five different servers. G3D – 1 and G3 – 2 stands for the two models taken from Geno3D server and R1, 2, 3, 4, 5 stands for the five models taken from Robetta server. 1yp2_A is the A chain of the x-ray crystallographic homotetrameric structure of small subunit. Structural differences are considered in the superimposition.

	CPH	Easy Pred	G3D – 1	G3D – 2	Swiss Model	R1	R2	R3	R4	R5	1yp2_A
CPH	-	1.95	2.37	2.28	1.61	2.27	2.62	2.56	2.24	2.67	0.94
EasyPred		-	2.52	2.42	1.77	2.53	3.17	2.71	2.09	2.87	1.08
G3D – 1			-	2.07	2.21	2.86	3.23	2.93	2.75	2.89	2.06
G3D – 2				-	2.26	2.86	3.06	2.82	2.63	2.98	1.92
SwissModel					-	2.26	2.93	2.61	2.40	2.89	1.35
R1						-	3.25	3.49	3.89	3.09	2.71
R2							-	2.77	3.38	3.49	3.17
R3								-	3.00	3.54	3.13
R4									-	2.49	2.30
R5										-	2.51
1yp2_A											-

Table 2 Amino acid sequences of the two different regions, shown in Figure 2, based on the sequence alignment by Clustal W (1.83). Region 2 and 3 are loop structures whereas region 1 corresponds to alpha helix followed by a coil in LS.

	Region 1	Region 2
LS	114 – 127 TQTPGEAGKKWFQG	141 - 145 DAKNK
SS	108 – 119 QQSPENPDWFQG	133 - 134 EH

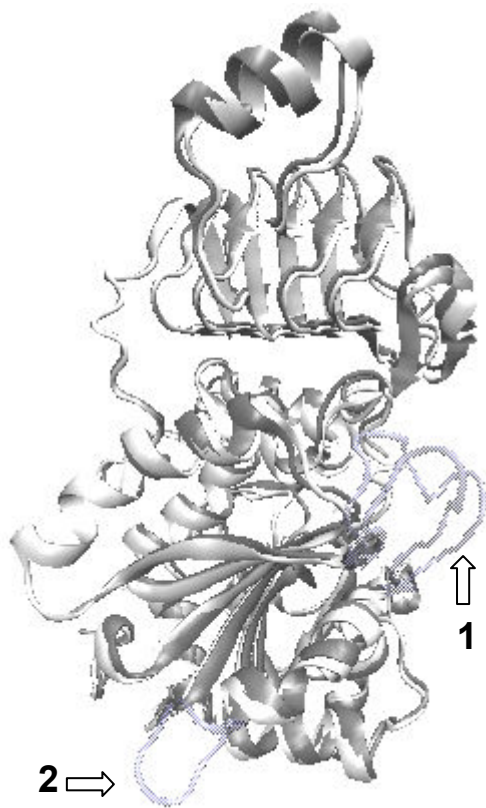


Figure 1 Superimposed structures of SS (1yp2_A) and LS showing the two different regions indicated by arrows and transparent color in both subunits. The pairs of residues to be superimposed were obtained from pairwise sequence alignment. LS is gray in color and SS is the white structure. Region 1 and region 2, indicated by arrows in the figure, correspond to loop structures in the predicted structure of the LS. There is a gap in PDB data for SS (chain A) between the amino acids 1-9, 27-32 and 91-98.

To identify residues (important for the interaction with small subunit) on the predicted potato tuber AGPase LS, putative heterotetrameric structure has been built with possible combinations of SS and LS as shown in *Figure 2*. Chains of LS were superimposed on the small subunit monomer, based on the x-ray crystallographic homotetrameric structure (PDB accession code is 1yp2). The corresponding SS chain was then deleted to obtain heterotetrameric enzyme model. Once the 3-D heterotetrameric structure has been established we have identified interface amino acids on the LS that may be important for interaction with SS using CONSURF and NACCESS. Some of these amino acids were common with the residues identified at the homotetrameric structure by Jin et al. [15]. In addition to those, additional amino acids, that are listed in *Table 3*, were identified in the interfaces of the LS based on the criteria of $>1 \text{ \AA}^2$ decrease in absolute accessibility [29] and $>5\%$ decrease in relative accessibility [30] when dimerized. The

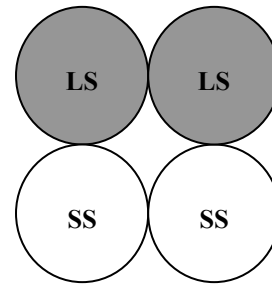


Figure 2a
Right - Left Interaction

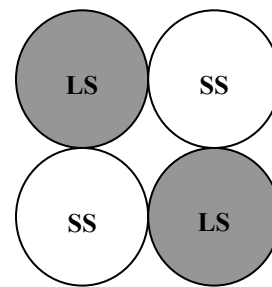


Figure 2b
Diagonal interaction

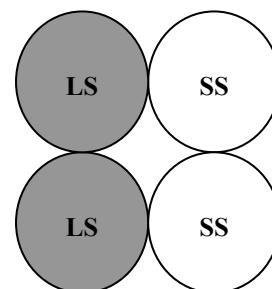


Figure 2c
Top - Bottom interaction

Figure 2 Predicted heterotetrameric structure of potato tuber AGPase with different interactions between the LS and the SS.

resulting amino acids at interface region make close contact with the small subunit as shown in *Figure 3*. In the interface of LS we obtained 33 residues for right to left, 6 residues for diagonal and 25 residues for top to bottom LS-SS interactions. Also, the residues Pro329 and Tyr381, found in the LS interface region of right to left interaction has a conservation score of 9, which makes these amino acids more critical. These sites most probably play crucial roles in the assembly of the functional heterotetrameric structure or in the regulation of the enzyme in response to the activators or inhibitors such as 3-PGA, ATP or inorganic phosphate (P_i).

Interface amino acids
Right-Left
GLY57, CYS58, TYR59, SER102 LEU304, THR305 , GLU307, PHE313 TYR314 , PRO319, TYR321 , THR322 , PRO324, PHE326, LEU327, PRO328 PRO329 , THR330 , LYS331, ILE332 ASP333 , ASN334, LYS336, ILE337 LYS338, ASP339, ILE341, ILE342 SER343, HIS344, ARG349, TYR381 THR383
Diagonal - Diagonal
ARG21, LYS26, SER102, GLN167 ASP171,LYS317
Top-Bottom
PHE78, ASN84 , SER85, ALA86 , ASN89 , ARG90 ALA93, ARG94 , PHE97,GLY99 VAL101,ASP105 PHE107,GLU109 LEU111, GLN115 , THR116,PRO117 LYS134 ,PHE135 TRP137,VAL138 LYS143,ASN144, LYS145

Table 3 Positions of the interface residues of LS interacting with SS. Residues dark in color correspond to interacting amino acids that were also found in the crystal structure of the SS.

4 Conclusions

In this study, we predicted the structure of potato tuber AGPase LS using computational methods and proposed 3 different models for the heterotetrameric enzyme. Based on these models we examined the subunit interactions between the LS and SS at dimeric level and identified the amino acids at interface regions of LS. Further analysis of the possible heterotetrameric models by using molecular dynamics (MD) simulations can provide the best model among these three structures. This will help us to identify other residues that play critical roles in the enzyme. Then, using site directed mutagenesis technique, experimental verification of computationally found residues will be performed. Identification of these critical amino acids between LS and SS will enable us to engineer the enzyme for improving the plant starch yield.

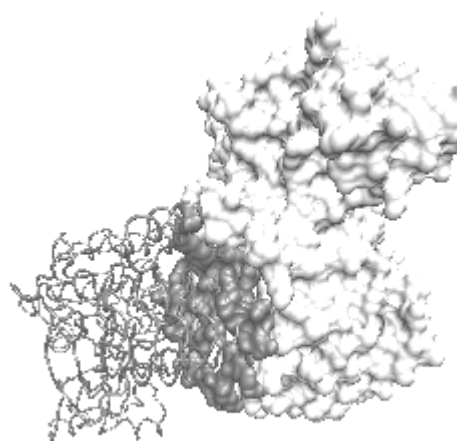


Figure 3a (Right – Left interaction)

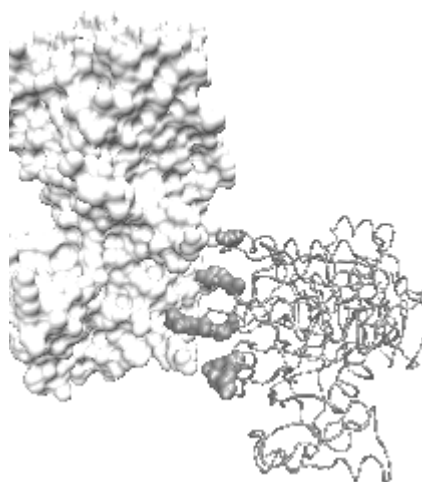


Figure 3b (Diagonal interaction)

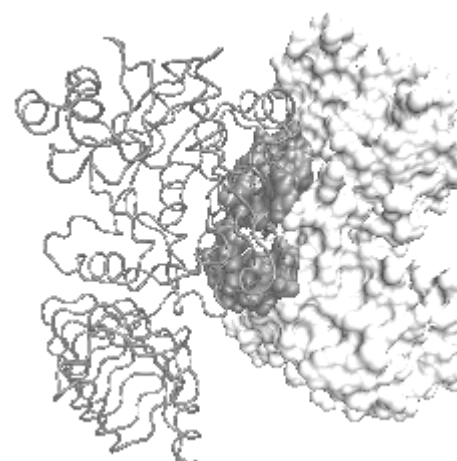


Figure 3c (Top – Bottom Interaction)

Figure 3 Interface regions of LS interacting with SS in three different ways at dimeric level. Structure white in color is the SS and gray in color is the LS.

4 References

- [1] C. J. Slattery, I. H. Kavakli, T. W. Okita, "Engineering starch for increased quantity and quality," *Trends in Plant Science*, 5, 291-298, 2000
- [2] I. H. Kavakli, J. C. Slattery, H. Ito, T. W. Okita, "The conversion of carbon and nitrogen into starch and storage organs: an overview," *Australian Journal of Plant Physiology*, 27(6), 561-570, 2000
- [3] M. A. Ballicora, A. A. Iglesias, J. Preiss, "ADP-glucose Pyrophosphorylase; a Regulatory enzyme for Plant Starch Synthesis," *Photosynthesis Research*, 79, 1-24, 2004
- [4] P. A. Nakata, T. W. Greene, J. M. Anderson, B. J. Smith-White, T. W. Okita, J. Preiss, "Comparison of the primary sequences of two potato tuber ADP-glucose pyrophosphorylase subunits," *Plant Mol. Biol.*, 17, 1089-1093, 1991
- [5] B. J. Smith-White, J. Preiss, "Comparison of Proteins of ADP-Glucose Pyrophosphorylase from diverse sources," *J.Mol.Evol.*, 34, 449-464, 1992
- [6] T. P. Lin, T. Caspar, C. R. Somerville J. Preiss, "Isolation and characterization of a starch less mutant of Arabidopsis thaliana (L) Henyh lacking ADPglucose pyrophosphorylase activity," *Plant Physiology*, 86, 1131-1135, 1998
- [7] Y. B. Fu, M. A. Ballicora, J. Preiss "Mutagenesis of the Glucose-1-phosphate Binding Site of Potato Tuber ADP-Glucose Pyrophosphorylase," *Plant Physiology*, 117, 989-996, 1998
- [8] H. I. Kavakli, J. S. Park, C. J. Slattery, P. R. Salamone, J. Frohlick, T. W. Okita, "Analysis of allosteric effector binding sites of potato ADP-glucose pyrophosphorylase," *Journal of Biological Chemistry*, 276, 40834-40840, 2001
- [9] P. R. Salamone, T. W. Greene, I. H. Kavakli, T. W. Okita, "Isolation and characterization of a higher plant ADP-glucose pyrophosphorylase small subunit homotetramer," *Febs Letters*, 482, 113-118, 2000
- [10] P. R. Salamone, I. H. Kavakli, C. J. Slattery, T. W. Okita, "Directed Evolution of ADP-glucose pyrophosphorylase of potato ADP-glucose pyrophosphorylase by reversion genetics," *Proceedings of the National Academy of Sciences of the United States of America*, 99, 1070-1075, 2002
- [11] W. C. Plaxton, J. Preiss, "Purification and Properties of Nonproteolytic Degraded ADPglucose Pyrophosphorylase from Maize Endosperm," *Plant Physiology* 83, 105-112, 1987
- [12] M. A. Ballicora, M. J. Laughlin, Y. B. Fu, T. W. Okita, G. F. Barry, J. Preiss, "ADPglucose from potato tuber. Significance of the N-terminal of the small subunit for catalytic properties and heat stability," *Plant Physiology* 109, 245-251, 1995
- [13] V. K. Sikka, S. B. Choi, I. H. Kavakli, C. Sakulsingharoj, S. Gupta, H. Ito, T. W. Okita, "Subcellular compartmentation and allosteric regulation of the rice endosperm ADPglucose pyrophosphorylase," *Plant Science* 161, 461-468, 2001
- [14] I. H. Kavakli, C. Kato, S. B. Choi, K. H. Kim, P. R. Salamone, H. Ito, T. W. Okita, "Expression, characterization of *A.thaliana* wildtype and up-regulated mutants cDNA in *E.coli*," *Planta* 215, 430-439, 2002
- [15] X. S. Jin, M. A. Ballicora, J. Preiss, J. H. Geiger, "Crystal structure of potato tuber ADP-glucose pyrophosphorylase," *Embo Journal*, 24, 694-704, 2005
- [16] R. Chenna, H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, J. D. Thompson, "Multiple sequence alignment with the Clustal series of program," *Nucleic Acids Research*, 31, 3497-3500, 2003
- [17] O. Lund, M. Nielsen, C. Lundegaard, "in Abstract at the CASP5 conference A102," 2002
- [18] C. Lambert, N. Leonard, X. De Bolle, E. Depiereux, "ESyPred3D: Prediction of proteins 3D structures. *Bioinformatics*," 18, 1250-1256, 2002
- [19] C. Combet, M. Jambon, G. Deleage, C. Geourjon, "Geno3D: automatic comparative molecular modeling of protein," *Bioinformatics* 18, 213-214, 2002
- [20] T. Schwede, J. Kopp, N. Guex, M. C. Peitsch, "SWISS-MODEL: an automated protein homology-modeling server," *Nucleic Acids Research* 31, 3381-3385, 2003
- [21] M. C. Peitsch, "Protein Modeling by E-mail *Bio-Technology*," 13, 658-660, 1995
- [22] N. Guex, M. C. Peitsch, "SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling," *Electrophoresis* 18, 2714-2723, 1997
- [23] D. E. Kim, D. Chivian & D. Baker, "Protein structure prediction and analysis using the Robetta server," *Nucleic Acids Research* 32, W526-W531, 2004

- [24] D. Chivian, D. E. Kim, L. Malmstrom, P. Bradley, T. Robertson, P. Murphy, C. E. M. Strauss, R. Bonneau, C. A. Rohl, D. Baker, "Automated prediction of CASP-5 structures using the Robetta server," *Proteins-Structure Function and Genetics* 53, 524-533, 2003
- [25] W. Humphrey, A. Dalke, K. Schulten, "VMD - Visual Molecular Dynamics," *J. Molec. Graphics* 14, 33-38, 1996
- [26] F. Glaser, T. Pupko, I. Paz, R. E. Bell, D. Bechor-Shental, E. Martz, N. Ben-Tal, "ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information," *Bioinformatics*, 19, 163-164, 2003
- [27] S. J. Hubbard, J. M. Thornton, "'NACCESS', Computer Program, Department of Biochemistry and Molecular Biology, University College London," 1971
- [28] S. K. Lee, F. M. Richards, "Interpretation of Protein Structures - Estimation of Static Accessibility," *J.Mol.Evol.* 55, 379-400, 1993
- [29] S. Jones, J. M. Thornton, "Analysis of Protein-protein Interaction Sites using Surface Patches," *J.Mol.Biol.* 272, 121-132, 1997
- [30] S. Miller, J. Janin, A. M. Lesk, C. Chothia, "Interior of surface of monomeric proteins," *J. Mol. Biol.*, 196, 641-656, 1987