Research Article

In Silico Studies of C₃ Metabolic Pathway Proteins of Wheat (Triticum aestivum)

Muhammad Kashif Naeem,¹ Sobiah Rauf,² Hina Iqbal,² Muhammad Kausar Nawaz Shah,¹ and Asif Mir²

¹ Department of Plant Breeding and Genetics, PMAS-Arid Agriculture University, Rawalpindi, Pakistan
² Department of Bioinformatics and Biotechnology, FBAS, International Islamic University, Islamabad, P.O. Box 44000, Pakistan

Correspondence should be addressed to Asif Mir; mir77uspk@gmail.com

Received 23 August 2012; Revised 10 November 2012; Accepted 26 November 2012

Academic Editor: Ravindra N. Chibbar

Copyright © 2013 Muhammad Kashif Naeem et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Photosynthesis is essential for plant productivity and critical for plant growth. More than 90% of plants have a C₃ metabolic pathway primarily for carbon assimilation. Improving crop yields for food and fuel is a major challenge for plant biology. To enhance the production of wheat there is need to adopt the strategies that can create the change in plants at the molecular level. During the study we have employed computational bioinformatics and interactomics analysis of C₃ metabolic pathway proteins in wheat. The three-dimensional protein modeling provided insight into molecular mechanism and enhanced understanding of physiological processes and biological systems. Therefore in our study, initially we constructed models for nine proteins involving C₃ metabolic pathway, as these are not determined through wet lab experiment (NMR, X-ray Crystallography) and not available in RCSB Protein Data Bank and UniProt KB. On the basis of docking interaction analysis, we proposed the schematic diagram of C₃ metabolic pathway. Accordingly, there also exist vice versa interactions between 3PGK and Rubisco. Future site and directed mutagenesis experiments in C₃ plants could be designed on the basis of our findings to confirm the predicted protein interactions.

1. Introduction

Photosynthesis is arguably the most important energy conversion process on earth because the chemical energy it yields is the base of food chains that sustain the overwhelming majority of other life forms. Plants utilize atmospheric CO₂ to liberate oxygen and synthesize carbohydrates during photosynthesis. It is an event where radiant energy of sunlight is utilized to convert carbon dioxide into photosynthetic by-products. On the basis of 1st product of photosynthesis or modifications in Calvin-Benson Cycle plants are grouped into three categories: C₃ plants, C₄ plants, and CAM plants. More than 90% of plants have C₃ metabolic pathway primarily for the carbon assimilation, whereas only 3% plants utilized C₄ metabolic pathway [1], and crassulacean acid metabolism (CAM) is found in sixteen thousand species of plants [2].

C₃ plants are less photosynthetically efficient and cannot grow in hot areas because of the activity of RuBisCO enzyme. RuBisCo has the ability to fix both carbon dioxide and oxygen that result in the loss of carbon, nitrogen, and energy from plants [3]. Bread wheat (Triticum aestivum L.), durum wheat (Triticum turgidum L.), and barley (Hordeum vulgare L.) are the most outstanding C₃ crops in terms of cultivated area and food source, from the Holocene time up to present [4]. Improving crop yields for food and fuel is a major challenge for plant biology which is necessary to meet the requirements of a rapidly increasing world population [5].

Wheat is a common staple cereal food crop in all parts of the world and contributes to 28% of the world’s edible dry matter (DM) and up to 60% of the daily calorie intake in several developing countries [6].

To enhance the production of wheat necessitates adopting strategies that can create the change in plants at the molecular level through one of two approaches: wet lab or computational. The study utilized computational bioinformatics and interactomics analysis of C₃ metabolic pathway proteins in wheat. Structural bioinformatics is concerned with the
prediction, analysis, and visualization of the 3D structure of proteins. Besides NMR and X-ray Crystallography that entail enormous experimental costs, time, and laborious procedure, another recent and attractive approach built for the analysis of protein structure has recently emerged in structural bioinformatics. There are three approaches primarily utilized to predict 3D structures of studied proteins: (1) homology modeling or comparative modeling [7, 8], (2) threading or fold recognition [9–11], and (3) ab-initio prediction [12–15]. Proteins interact with each other as well as with other macromolecules to accomplish their functions within cell; therefore protein-protein interactions are complex and play a crucial role in most biological processes to determine the actual functioning of protein. Wet lab approaches such as tandem affinity purification mass spectrometry [16], yeast two hybrid [17, 18], and some others are found to enable the mapping of complex protein interactions. It is difficult to take advantage of these experimental techniques due to the complexity of biological molecules compared to computational protein-protein docking that offer more benefits [19]. More recently, a high-throughput docking approach for protein-protein interaction has been reported [20]. The aim of this study is to improve the understanding of C₃ metabolic pathway proteins through in silico analysis and to study protein interaction in wheat. Each protein have has an important role, either in CO₂ fixation to regenerate ribulose 1, 5-bisphosphate or to synthesize starch and sucrose. This study selected only nine of proteins involved in the C₃ Pathway. The nine selected protein structures have not yet been determined through wet lab experiment (NMR, X-ray Crystallography) nor available in RCSB Protein Data Bank and UniProt KB.

2. Materials and Methods

A flowchart representing the use of tools in a sequential manner for study of nine important proteins, involved in C₃ metabolic pathway of wheat, is given in Figure 1. Sequences of these proteins were retrieved from NCBI (National Center for Biotechnology Information) in FASTA format, and it was used as a query for further analysis [21].

In order to generate three-dimensional models homology, threading and ab-initio approach was applied.

**Homology Modeling.** The modeling of the 3D structure of the proteins was executed by Swiss-Modeler (http://swiss-model.expasy.org/) program [22, 23].

**Threading Approach.** SAM-T08 (http://compbio.soe.ucsc.edu/SAM_T08/T08-query.html) [24]. SAM-T08 web-based program for the modeling of three-dimensional structures of all selected protein sequences was used.

**Ab-Initio Prediction. Iterative Threading Assembly Refinement (I-TASSER)** [25, 26] online web server for predicting 3D protein structure was used.

**Model Visualization and Evaluation.** For model visualization the ViewerLite software was used. Rampage server (http://mordred.bioc.cam.ac.uk/~rapper/rampage.php) was used for evaluating and assessing the accuracy of the model [27, 28].

**Protein-Protein Interaction.** Protein-protein interaction is generated by the help of HEX 6.1. HEX 6.1 program is used to speed up docking estimations; the method provides a feasible orientation rapidly and precisely [29].

3. Results

3.1. Retrieval of Target and Recognition of Template Protein. For this study, selected proteins involved in C₃ metabolic pathway were retrieved from NCBI (having maximum residual length), and their protein sequences in FASTA were used for a query sequence similarity search. PSI-BLAST identified the potential template proteins with their respective PDB-ID with maximum similarity and E value. The BLAST results of nine target proteins are summarized in Table 1. As per our findings Rbc1 and ATP synthase template showed 100% similarity with target sequence. GAPN and F-base demonstrated similarity more than 90%. 3PGK and WAXY showed percent similarity more than 80%, and remaining 3 proteins demonstrated the sequence similarity from 70 to 48%. After the selection of potential templates 3D model is generated by the use of Swiss-modeler program.
Table 1: Summary of selected nine proteins involved in C₃ metabolic pathway of wheat and their BLAST results.

| Number | Target proteins                                      | Gene name | Gene ID number | Residues | PDB ID   | % MI*       | E value* |
|--------|------------------------------------------------------|-----------|----------------|----------|-----------|-------------|----------|
| 1      | RuBisCO large subunit                                | RbcL      | BAB 47042.1    | 477      | 4RUB_A   | 100%        | 0.0      |
| 2      | Phosphoglycerate kinase                              | 3PGK      | CAA 51931.1    | 480      | 1VPE_A   | 80%         | 2e−128   |
| 3      | NADP-dependent glyceraldehyde-3-phosphate dehydrogenase | GAPN      | Swiss Prot Q8LK61.2 | 496   | 1EUH_A   | 95%         | 1e−137   |
| 4      | Sucrose phosphate synthase II                        | SPS II    | ADK 11932.1    | 626      | 2R60_A   | 50%         | 7e−50    |
| 5      | Fructose-bisphosphate aldolase                       | Fba       | CBH 32613.1    | 385      | 3KX6_A   | 90%         | 1e−103   |
| 6      | ATP synthase alpha subunit                            | ATP A     | AAA 84725.1    | 504      | 1FX0_A   | 100%        | 0.0      |
| 7      | Starch branching enzyme I                            | Sbe1      | CAA 72987.1    | 830      | 3O7Y_A   | 70%         | 4e−40    |
| 8      | Granule-bound starch synthase 1                      | WAXY      | Swiss Prot P27736.1 | 615   | 3D1J_A   | 83%         | 1e−59    |
| 9      | Phosphoribulokinase                                  | PRK       | Swiss Prot P26302.1 | 404   | 2JEO_A   | 48%         | 4e−16    |

% MI: percent maximum similarity; E value: energy value.

determined the length and alignment accuracy of generated model by ViewerLite. Five proteins (Rbcl, 3PGK, GAPN, Fba, and ATP A) showed the model length percentage more than 80%. While remaining four proteins (SPS II, Sbe I, WAXY, and PRK) have less than 80% (Table 2). All 3D full length models were generated by SAM-T08 (based on HMM) except Sbe I. Modeling of Sbe I protein is done by I-TASSER (using ab-intio approach) as shown in given figures (Figures 2, 3, 4, 5, 6, 7, 8, 9, and 10).

3.2 Model Validation. Three-dimensional protein models validation is last step of protein modeling. It evaluates the significance and accuracy by using Ramachandran plot in Procheck web-based server. Results of Ramachandran plot demonstrate that eight proteins have more than 90% number of residues in favored region which shows the accuracy of these models. One protein (Sbe1) has residues in favored region <90% as 83.6% and the lower level significance of Sbe I model. Model evaluation is done by Rampage program, and results are described in Ramachandran plot (Table 2). The summary of Ramachandran plot results is presented in Table 3. ATP A protein showed maximum residues 97.8% in favored region among all protein. 3PGK protein depicted the least residues 0.3% in outlier region, and PRK protein demonstrated the maximum residues 1.8% in outlier region.

3.3 Protein-Protein Interaction. Protein-protein interaction describes the actual functioning of protein and their role in metabolic pathway. We used HEX-6.1 program for all our selected proteins. Docking results for each protein are summarized in Table 4. Docking results predicted that Rbcl has strong interaction with 3PGK indicated by minimum E value −709.8. 3PGK showed strong linkage with two proteins that is Fba (−664.83) and GAPN (−653.19). GAPN showed
Table 2: Summary and comparison of ViewerLite results of selected proteins with various web servers.

| Proteins | Number of residues (amino acids) | Swiss-model | % model length | SAM-T08 | I-TASSER |
|----------|----------------------------------|-------------|----------------|---------|----------|
| Rbcl     | 477                              | 1-464       | 97.2%          | 1-477   | —        |
| 3PGK     | 480                              | 1-398       | 82.9%          | 1-480   | —        |
| GAPN     | 496                              | 1-474       | 95.5%          | 1-496   | —        |
| SPS II   | 626                              | 1-456       | 72.89%         | 1-626   | —        |
| Fba      | 385                              | 1-358       | 92.87%         | 1-385   | —        |
| ATP A    | 504                              | 1-477       | 94.6%          | 1-504   | —        |
| Sbe I    | 830                              | 1-612       | 73.73%         | —       | 1-830    |
| WAXY     | 615                              | 1-405       | 65.85%         | 1-615   | —        |
| PRK      | 404                              | 1-215       | 53.2%          | 1-404   | —        |

4. Discussion

Over 90% of plants have C₃ metabolic pathway. C₃ photosynthetic pathway present in major crops such as wheat, barley, and rice. Wheat is a major staple food crop all over the world. Decrease in productivity of wheat due to C₃ metabolic pathway involved oxygenation reactions and results in loss of energy 25–30%. Improve the efficiency of C₃ metabolic pathway needs to know the functions of proteins involved in C₃ pathway. Consequently, it is very crucial to be familiar with the functions of these proteins.
with the 3D structure of proteins which gave insight into molecular functioning, which enhance the better understanding of physiological processes and biological systems. Therefore in our study, eight protein sequences were used and modeled via threading approach, as these proteins have >700 amino acids, and models are constructed by SAM-T08. For one protein, that is, Sbe I (830 aa in length), I-TASSER program generated full length model (Figure 8). Full length generated models are significant to use for future studies. Significance of our modeling studies for nine proteins was depicted by confirming that there is no crystalline structure present in protein databank till date.

On the basis of docking result we proposed the schematic diagram of C₃ metabolic pathway (Figure 11). As per our analysis, Rbcl showed strong interaction with 3PGK in C₃ metabolic pathway, as we obtained minimum docking value (Table 4), while Rbcl as receptor indicated that it has no interaction with GAPN, Fba, WAXY, and PRK. Our docking results also described that 3PGK as a receptor demonstrated the maximum interaction value with Rbcl. It is predicted that 3PGK regenerate the Rubisco enzyme. 3PGK showed strong linkage with two proteins, that is, Fba and GAPN. 3PGK interaction with Fba leads the process to starch biosynthesis, while its interaction with GAPN produces sucrase, as final products. Previous in vitro study depicted 3PGK (EC 2.7.2.3) and GAPN (EC 1.2.1.13) association [30].

Glyceraldehyde-3-P dehydrogenase (GAPN) is essential for plant metabolism and development. GAPN depicted strong interaction with SPS II in our analysis. SPS II are

| Proteins | Favored region | Allowed region | Outlier region |
|----------|----------------|----------------|---------------|
| Rbcl     | 96.8%          | 2.7%           | 0.4%          |
| 3PGK     | 95.8%          | 3.9%           | 0.3%          |
| GAPN     | 95.7%          | 3.3%           | 1.0%          |
| SPS II   | 96.0%          | 3.0%           | 1.0%          |
| Fba      | 96.1%          | 3.1%           | 0.8%          |
| ATP A    | 97.8%          | 1.8%           | 0.4%          |
| Sbe I    | 83.6%          | 10.4%          | 6.0%          |
| WAXY     | 94.6%          | 4.6%           | 0.8%          |
| PRK      | 96.6%          | 1.6%           | 1.8%          |
carbon compound and are indispensable for growth and development of the plant [31, 32]. In vivo it is shown that phosphorylation of SPS II changes the activity of SPS II and carbon partitioning [33]. In this regard SPS II enzyme is very essential for carbon partitioning, and also SPS II synthesizes sucrose essential for plant development and growth. Sucrose is also synthetase CO2 fixation pathway. So there also exist vice versa interactions of 3PGK with RbcL.

(iii) In biochemical pathways PRK interacts with RbcL to complete the cycle, but according to our predicted pathway interaction between PRK with RbcL is mediated through 3PGK.

(iv) Future site directed mutagenesis experiments in C3 plants could be designed on the basis of our findings to confirm the predicted protein interactions.

### References

[1] R. F. Sage and R. K. Monson, *C4 Plant Biology*, Academic Press, New York, NY, USA, 1999.

[2] A. N. Dodd, A. M. Borland, R. P. Haslam, H. Griffiths, and K. Maxwell, "Crassulacean acid metabolism: plastic, fantastic," *Journal of Experimental Botany*, vol. 53, pp. 569–580, 2002.

[3] A. R. Portis Jr and M. A. J. Parry, "Discoveries in Rubisco (ribulose 1,5-bisphosphate carboxylase/oxygenase): a historical perspective," *Photosynthesis Research*, vol. 94, no. 1, pp. 121–143, 2007.

[4] L. T. Evans, *Feeding the Ten Billion: Plants and Population Growth*, Cambridge University Press, Cambridge, UK, 1998.

[5] M. D. Edgerton, "Increasing crop productivity to meet global needs for feed, food, and fuel," *Plant Physiology*, vol. 149, no. 1, pp. 7–13, 2009.

[6] B. S. Gill, R. Appels, A. M. Botha-Oberholster et al., "A workshop report on wheat genome sequencing: international genome research on wheat consortium," *Genetics*, vol. 168, no. 2, pp. 1087–1096, 2004.

[7] A. Fiser, R. K. Do, and A. Salì, "Modeling of loops in protein structures," *Protein Science*, vol. 9, no. 9, pp. 1753–1773, 2000.

[8] M. A. Martí-Renom, A. C. Stuart, A. Fiser, R. Sánchez, F. Melo, and A. Šali, "Comparative protein structure modeling of genes and genomes," *Annual Review of Biophysics and Biomolecular Structure*, vol. 29, pp. 291–325, 2000.

[9] Y. Xu and D. Xu, "Protein threading using PROSPECT: design and evaluation," *Proteins*, vol. 40, no. 3, pp. 343–354, 2000.

[10] R. David, M. J. Korenberg, and I. W. Hunter, "3D-1D threading methods for protein fold recognition," *Pharmacogenomics*, vol. 1, no. 4, pp. 445–455, 2000.

[11] J. Skolnick, D. Kihara, and Y. Zhang, "Development and large scale benchmark testing of the PROSPECTOR 3.0 threading algorithm," *Proteins*, vol. 56, no. 3, pp. 502–518, 2004.

### Table 4: Summary of protein docking results.

| Ligands | RbcL | 3PGK | GAPN | SPS II | Fba | Sbe I | WAXY | PRK | ATP A |
|---------|------|------|------|--------|-----|-------|------|-----|-------|
| RbcL    | —    | —    | —    | —      | 0.00| —     | —    | 0.00| —     |
| 3PGK    | -731.7| —    | —    | —      | 0.00| —     | —    | 0.00| —     |
| GAPN    | 0.00  | -697.7| —    | —      | 0.00| —     | —    | 0.00| —     |
| SPS II  | -668.63| -563.23| —    | —      | 0.00| —     | —    | 0.00| —     |
| Sbe I   | -540.88| 445.36| —    | —      | 0.00| —     | —    | 0.00| —     |
| WAXY    | 0.00  | -684.00| —    | —      | 0.00| —     | —    | 0.00| —     |
| PRK     | 0.00  | -614.7 | —    | —      | 0.00| —     | —    | 0.00| —     |
| ATP A   | -550.1 | -621.99| —    | —      | 0.00| —     | —    | 0.00| —     |
