MATERIALS AND METHODS

In order to achieve the assigned objectives following materials and methods have been followed. The proposed research work is carried out in the Bioinformatics Centre and Biotechnology laboratory at Indian Institute of Wheat and Barley Research, Karnal where all the facilities for the proposed research work are available.

3.1 Comparative analysis of wheat AGPase with other well studied crops using various bioinformatics tools

3.1.1 Physico-chemical characterization

Protein sequences of small subunit (SS) and large subunit (LS) of AGPase from wheat, rice, maize, potato and Arabidopsis were retrieved from protein database of National Center for Biotechnology Information (NCBI; http://www.ncbi.nlm.nih.gov/protein/) in FASTA format. The ProtParam tool (http://web.expasy.org/protparam/) of ExPASy was used to compute amino acid composition (%), molecular weight, theoretical isoelectric point (pl), number of positively and negatively charged residues, extinction coefficient, instability and aliphatic index and grand average of hydropathy (GRAVY).

3.1.2 Secondary structure and functional properties

Secondary structural properties of the protein including alpha helix, 3_10 helix, Pi helix, beta bridge, extended strand, beta turns, bend region, random coil, ambiguous and other states were computed by using SOPMA (Self Optimized Prediction Method with Alignment; http://npsapbil.ibcp.fr/cgibinpsa_automat.pl?page=/NPSA/npsa_sopma.html) tool of NPS (Network Protein Sequence Analysis). For functional analysis, the motifs of the AGPase protein sequences were identified by using Expasy’s prosite tool (http://prosite.expasy.org/). Input data type was in FASTA format and motifs were scanned against prosite patterns.
3.1.3 **Identification of Signature Logo**

Graphical representation of the pattern within a multiple sequence alignment of AGPase was generated using Web Logo tool (http://weblogo.berkeley.edu/). In sequence logo total height of the stack implies the sequence conservation at that position, while the heights of the symbols within the stack indicate the relative frequency of each amino acid at that position.

3.1.4 **Phylogenetic tree construction**

Sequences of both large and small subunits of wheat, rice, maize, barley, potato and *Arabidopsis* were aligned using ClustalW. Based on the alignment score, phylogenetic tree was build using Neighbor Joining (NJ) method using MEGA6 program (Tamura et al., 2007).

3.1.5 **Identification of the 5’ regulatory region**

Nucleotide sequences of AGPase were scanned for the presence of putative cis-acting regulatory elements with or similar to the recorded in Plant CARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/cgi-bin/CallMatIE55.html), and PLACE (http://www.dna.affrc.go.jp/PLACE/signalscan.html) databases.

3.2 **In silico study of interaction between large and small subunit of wheat AGPase enzyme**

3.2.1 **Modeling of the large and small subunit**

The protein sequences of large and small subunit of wheat AGPase were retrieved from the NCBI protein sequence database (http://www.ncbi.nlm.nih.gov/protein) and a template was predicted using PSI-BLAST (Altschul et al., 1990, http://blast.ncbi.nlm.nih.gov/BLAST.cgi) against the RCSB protein data bank (Berman et al., 2000, http://www.pdb.org). The three dimensional (3D) structures of both the protein subunits were build using the Swiss Model Server (Schwede et al., 2003,
In addition to Swiss Model server, Modeller9v10 was also used to build the protein model of wheat AGPase (TaAGPase). The assessments of protein structure was done using the Protein structure and model assessment tools integrated in Swiss Model server, which employs various local and global quality evaluation parameters. The models were additionally analyzed and verified using programs PROCHECK (Laskowski et al., 1993) VERIFY3D (Eisenberg and Luthy 1997) and ERRAT (Colovos and Yeates 1997) at the Structure Analysis and Verification Server (SAVES) (http://nihserver.mbi.ucla.edu/SAVES). The outcomes were analyzed and models were further improved.

### 3.2.2 Docking studies of AGPase protein

Three approaches were used for constructing the model of heterotetrameric enzyme of TaAGPase i.e. monomeric, dimeric and homodimer of LS and SS. In monomeric approach firstly model of LS and SS were built individually then its dimerization and tetramerization was done through GRAMM-X docking server (Tovchigrechko and Vakser 2006). In dimeric approach, primarily dimer of LS-SS was built using Yasara server then tetramerization was carried out. In the last approach, dimer of LS-LS and SS-SS was created then its tetramerization was done through docking. The homology modeled LS and SS of wheat AGPase were submitted to the GRAMM-X docking server (Tovchigrechko et al., 2008) to carry out a rigid body docking using fast Fourier transformation approaches applying smoothed Lennard-Jones potential, knowledge-based and refinement stage scoring, which gives rise to the finest surface match. The most excellent dimer orientation was again submitted to the GRAMM-X server to obtain initial heterotetramer orientations. Idealization of bond geometry and elimination of unfavorable non-bonded contacts was accomplished by energy minimization with force field GROMOS96 (Van Gunsteren and

3.2.3 Molecular simulation studies

Explicit solvent molecular dynamic (MD) simulation for the representative structure of AGPase heterotetramers were performed using YASARA server. MD was performed using the CHARMM force field (MacKerell et al., 1998) parameters with 1 nanosecond (ns) time Yasara sever (www.yasara.org/minimizationserver.htm) was used to run the simulation on three types of heterotetrameric complexes generated by different approaches. Thus, MD simulations were performed on final heterotetrameric complexes and 40 snapshots were taken for each simulation. The energy profiles from the .tab file of the simulation were extracted and total energy was obtained. The final minimized structures for the heterotetramer were obtained and visualized in chimera. The model was selected based on energy and Z score.

3.2.4 Subunit-subunit interaction studies

To predict subunit-subunit interactions, the MD minimized structures were used as input to a program namely Dimplot (Wallace et al., 1995). The sequences from rice, maize and wheat for the protein subunits were aligned to validate the interface residues listed in subunit-subunit interaction by this program. The multiple sequence alignment was carried out using ClustalW, (http://www.ebi.ac.uk/Tools/msa/clustalw2/). The heterotetramer generated was superimposed on the template 1YP2, i.e., the crystal structure of the potato tuber homotetramer available from the PDB. The superimposition and root mean square deviation (RSMD) calculation was performed by the UCSF Chimera 1.6.1 (Pettersen et al., 2004) to determine the accuracy of the modeled tertameric complex.
3.3 Mining of SNPs

SNP identification through computational approach makes use of large number of ESTs available in public database. This method is considered to be quick and economical than experimental procedures. To fulfill this objective following strategies were adopted (Figure 3.1)

![Diagram](image)

**Figure 3.1:** Schematic diagram delineate the computational approach for identification of SNPs in wheat AGPase.

3.3.1 Data description

We have searched all the available wheat EST databases for AGPase such as Gramene, TIGR, Cerealdb etc. However, in these databases either the information about the sequence was partial or not available. Therefore ESTs sequences of wheat AGPase gene were downloaded from Unigene (http://www.ncbi.nlm.nih.gov/unigene) at NCBI (http://www.ncbi.nlm.nih.gov/).
3.3.2. EST Pre-processing

EST preprocessing steps consist of number of essential steps to minimize the chance to cluster unrelated sequences. ESTs sequences were derived by sequencing of cloned cDNA libraries of wheat. In first step, vector fragments were clipped before ESTs analysis. VecScreen web server (http://www.ncbi.nlm.nih.gov/V...screen.html) was used for detection of ESTs containing vector sequence contamination. Bioedit software (http://www.mbio.ncsu.edu/bioedit/bioedit.html) was used to crop the vector region. At first, detection of any vector contaminated segments can escape wrong inference about the biological meaning of the sequence. Polyadenosine tail [poly (A)] and poly (T) stretches in ESTs were clipped with the help of the EST-trimmer Perl program available online (http://www.pgrc.ipk-gatersleben.de/ misa/ download /est-trimmer.pl). EST sequences being too short (less than 100 bp), or too long (over 800 bp) were discarded. Moreover, the interspersed repeats and low complexity region present in ESTs were screened using Repeat masker server (http://www.repeatmasker.org/) which yields a masked query sequence.

3.3.3. EST Assembly of AGPase

The cleaned ESTs sequences of AGPase were assembled to produce consensus sequences or contigs using Seqman module of DNASTAR's Lasergene sequence analysis software (Burland, 2000) and CAP3 program (Huang and Madan, 1999). Any EST sequences which were not assembled were considered as singletons. Assembly of these cleaned ESTs were carried out using stringent criteria like (Match size = 40, sequence length = 100, maximum expected coverage = 40 and match percentage = 95).

3.3.4. Mining of SNPs using automated method

EST repetition is very useful for mining SNPs. If different ESTs from the same gene have alignment mismatches, they may be SNPs. To avoid the detection of sequencing
errors, a redundancy of two was chosen for a given mismatch in order to be considered a candidate SNP (Cox et al., 2001; Picoult-Newberg et al., 1999). Only contigs comprising four overlapping sequences or more were selected using perl program and further analyzed. Though contigs with too many sequences (>500 ESTs) were difficult to view and edit and its analysis could lead to the detection of false positive candidate SNPs (Batley et al., 2003). Therefore, it was removed from the analysis. In silico SNPs were predicted using following software programs:

3.3.4.1 AutoSNP

AutoSNP version.1.0 (Barker et al., 2003) tool was used to detect the candidate SNPs from contigs. AutoSNP requisite input as ace or fasta format files generated by CAP3 program. For each predicted candidate SNP, two scores, the redundancy of the polymorphism at a SNP locus and the weighted co-segregation score of the candidate SNP were calculated. AutoSNP is written in perl and run from the Linux command line.

3.3.5 DNASTAR's Lasergene software

SeqMan module of DNAsstar’ Lasergene software version 6.0 (Burland, 2000) was used to detect the putative SNPs of AGPase LS and SS in the assembled sequence. The criteria for putative SNP prediction were as given below:

a) SNP score >40%

b) Analyzing contigs comprising 5 or more ESTs from different cultivars.

3.3.6 Designing of gene based SNP primers

The SNP primers, hereafter named as allele specific marker (ASM), were designed from the position of base transition. The underlying principle in designing the ASM was based on the idea that the 3’-terminal positions must be exclusive among the known wheat genomic sequences. Consensus sequence of contigs were used by Primer3 program (http://frodo.wi.mit.edu/cgi bin/primer3/primer3_www.cgi) using criterion with
Materials & Methods

length of 18 to 28 bases, annealing temperature of 50-60°C, product sizes ranging from 100 to 400 bp, and G-C content from 40% to 60% (Rozen and Skaletsky 2000). List of designed ASM primers are mentioned in Table 3.1.

Table 3.1: Details of designed gene based primers*

<table>
<thead>
<tr>
<th>S.No</th>
<th>Oligos Name</th>
<th>SNPs position</th>
<th>Primer sequence</th>
<th>Tm (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AgpL/C11/S-398F, AgpL/C11/S-398R</td>
<td>398</td>
<td>F 5’-TCCAGGAGGATGTGCAAC-3’, R 5’-GAGGTAGGTGCGGTGAATGT-3</td>
<td>58</td>
</tr>
<tr>
<td>4</td>
<td>AgpS/C8/S-937F, AgpS/C8/S-937 R</td>
<td>937</td>
<td>F 5’-TACGACCATACTGGCCGGCAGAC-3’, R 5’-GCTTGTACCCTGACAGCAGA-3’</td>
<td>57</td>
</tr>
<tr>
<td>5</td>
<td>AgpS/C6/S-192F, AgpS/C6/S-192R</td>
<td>192</td>
<td>F 5’-TCCTTCTCTTTGCGCTGTT-3’, R 5’-TGCGCAGTAGGTTGCTGCT-3’</td>
<td>56</td>
</tr>
<tr>
<td>6</td>
<td>AgpS/C14/S-630F, AgpS/C14/S-630R</td>
<td>630</td>
<td>F 5’-TCACGCAGTTCAACTCAGCA-3’, R 5’-CAGCATCGCATCCCATGCAG-3’</td>
<td>57</td>
</tr>
</tbody>
</table>

*F→ forward primer; R→ reverse primer; Tm→ annealing temperature

3.3.7 Molecular analysis of AGPase

3.3.7.1 Plant Material

A set of diverse bread wheat genotypes was procured from Germplasm unit of Indian Institute of Wheat and Barley Research, Karnal. Total DNA extraction of all the genotypes was carried out from 3 week old seedlings by Cetyl trimethylammonium bromide (CTAB) based method (Allen et al., 2006). Harvested 50mg of fresh leaves were immediately frozen and were ground to a fine powder with liquid nitrogen in mortar with pestle and transferred to 1.5ml sterilized centrifuge tubes. Later, 15ml of 1.33 x
CTAB buffer was added to the fine powdered leaf tissue, contents were mixed to form an emulsion, which was incubated by shaking at 65°C for 30 minutes in water bath. After incubation, the emulsion was allowed to cool down at room temperature and equal volume of Chloroform: Isoamylalcohol (24:1) mixture was added. The tubes were capped and gently swirled to mix the contents for 30 min. The tubes were centrifuged at 4000 rpm for 15 minutes at room temperature in a centrifuge. The aqueous layer formed after centrifugation was taken carefully and transferred to a fresh 1.5 ml tubes. To this, an equal volume of chilled Isopropyl alcohol was added and mixed by gently inverting the tubes. The tubes were again centrifuged at 12000 rpm for 15 minutes at room temperature in a centrifuge (Eppendorf 5430R). The aqueous layer formed was discarded and DNA pellet so obtained was washed twice suspending in 500 µl of 70% ethanol, pelleted at 10,000 rpm for 5 minutes at 4°C and supernatant discarded. DNA pellet was air dried by keeping the tubes in an inverted position for about 30 minutes. Care was taken to remove all the traces of ethyl alcohol from the DNA pellet for longer storage of DNA. The dried DNA pellet was dissolved in 50 µl of TE (10mM tris – HCL pH 8.0 1mM EDTA pH 8.0) buffer. The DNA samples were treated with RNase to remove residual RNA. For this 5µl of RNase (10mg/m) was added to each DNA sample and incubated at 37°C for 45 minutes.

3.7.2 DNA Quantification

For quantification of DNA, 0.8%(w/v) agarose gel was prepared by dissolving 0.8g of agarose Powder in 100 ml 1 x TAE buffer heated in a microwave oven till fully melted. Ethidium bromide (EtBr) was added to gel (final concentration 0.5µg/ml) to enable visualization of DNA after electrophoresis. The solution was poured into a casting tray comprising a sample comb and permitted to solidify at room temperature. After solidification of the gel, the comb was detached and the gel in the plastic tray was placed in the electrophoresis chamber. The electrophoresis chamber was filled with 1 x TAE
buffer so that it covered the gel. Extracted genomic DNA (1µl) mixed with loading dye (6x) was loaded into the sample wells along with known dilutions of DNA ladder (50 ng, 100 ng, 200ng and 500 ng). The DNA concentration of the sample was assessed by visual comparison of the band with known dilutions of DNA ladder. Electrophoresis was carried out at constant 100 volts for 1 hour. The gels were then examined under Gel Documentation System (Alpha Innotech). Purity and concentration of extracted DNA for each sample was also checked by Nano drop spectrophotometer (Thermo Scientific Co.). One microliter of TE buffer was used to calibrate the Nano drop spectrophotometer at 260 and 280 nm wavelengths. One microliter of DNA was loaded. The concentration (ng/µl) and optical density (OD) of DNA samples were recorded. The A260/A280 ratio around 1.9 (1.85- 1.95) shows best quality of DNA.

3.3.8. Polymerase chain reaction analysis (PCR)

PCR reaction was followed as; 2.5µl of 10X PCR buffer, 2.0 µl of dNTPs (2.5mM each dNTP) (Banglore Genei) and 0.5 µl each of forward and reverse primers synthesized (Eurofins), 1.0 unit of Taq DNA polymerase (Bangalore Genei) and 100 ng of DNA. PCR amplification was carried out as described in Table 3.2.

Table 3.2: Thermocycler Program for the PCR-based marker amplification

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Time</th>
<th>No. of cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>94°C</td>
<td>4 min</td>
</tr>
<tr>
<td>Denaturation</td>
<td>94°C</td>
<td>1 min</td>
</tr>
<tr>
<td>Annealing</td>
<td>57°C to 59°C</td>
<td>1 min 35 cycles</td>
</tr>
<tr>
<td>Elongation</td>
<td>72°C</td>
<td>5 min</td>
</tr>
<tr>
<td>Final Extension</td>
<td>72°C</td>
<td>10 min</td>
</tr>
<tr>
<td></td>
<td>4°C</td>
<td>∞</td>
</tr>
</tbody>
</table>
The amplified PCR products for SNP markers of AGPase were resolved on to 2% agarose gel (Low EEO Hi-Media) in 1x TAE buffer and was photographed in a Gel Documentation unit (Alpha Innotech) visualized under UV transilluminator.

### 3.9 Plant materials used for enzyme extraction

For enzymatic studies 36 released Indian wheat genotypes (1967-2006), representing in different agro-climatic regions of India were used (Table 3.1). The data on thousand grain weight (TGW) and grain yield were taken from IIWBR, Karnal (Kundu et al., 2010). At appearance of flag leaf, a close watch was kept for the onset of anthesis. Ears were marked and tagged corresponding to the date of anthesis. At a stage of 15 and 21 days after anthesis (DAA) the ears of each variety were harvested and brought to laboratory on ice for the enzymatic studies on developing wheat grains. Data was recorded on 2 plants of each variety and post harvest observations about grain parameters were recorded (Sikka and Kanchan 2001).

### 3.10 Enzyme Extraction

Ears with developing grains at a stage of 15 and 21 DAA were used for preparation of grain extract. The buffer used was having a composition as follows: Extraction buffer- 50 mM MOPS pH-7.4, 2 mM MgCl₂, 1mM EDTA and 2 mM DTT. Approximately 15-20 developing grains amounting to 0.5 g of the grains were removed randomly from ear heads and were hand homogenized in a prechilled pestle and mortar at 4°C with 2ml of buffer on ice. The supernatant obtained was used as the AGPase enzyme extract (Sikka and Kanchan 2001).

### 3.11 Enzyme Assay

AGPase was assayed in the reverse direction by customized method of Kleczkowski et al. (1993). The reaction mixture contained-

50 mM MOPS (pH-7.4) 300 µl
7.5 μ mole MgSO₄·7H₂O 100 µl
Enzyme extract 100 µl
3 μ mole 3 PGA 100 µl
0.5 μ mole NADP⁺ 100 µl
0.5 μ mole ADP-glucose 100 µl
2 units Phosphoglucomutase 3 µl
2 units Glucose-6-phosphate dehydrogenase 2 µl

The reaction was initiated by the addition of 200 µl of sodium pyrophosphate (2.5 μ mole). The pyrophosphorolytic activity of AGPase was assayed spectrophotometrically by monitoring the change in O.D. due to conversion of NADP to NADPH at 340 nm.

3.4. Development of AGPaseDB: A Comprehensive database of ADP Glucose Prophosphorylase of crop plants

Relational database AGPaseDB for various fields comprising physical properties, secondary structure, subcellular localization and physiological and biochemical parameters has been created for easy retrieval of information related to AGPase of various crops.

3.4.1 Collection of data

Data collection was the primary focus and most important task for development of AGPaseDB; A comprehensive database on ADP glucose pyrophosphorylase of crop plants. The physical properties, secondary structure information was predicted from Expasy’s protparam and information related to other parameters was retrieved from primary database such as NCBI (http://www.ncbi.nlm.nih.gov/) and scientific literatures linked to PubMed and PubMed Central. Overlapping information was removed to generate the non-redundant dataset.
3.4.2 Technology used

Attributes are grouped into one-to-one or one-to-many relations by interpreting the attributes. One-to-many relations are further decomposed using normalization process. C# ASP.NET has been used for the main front design development of the database (Walther, 2007). The storage system in the database is SQL Server 2005, a relation database management system running on windows server using ISI sever. Further to reduce time lag or to increase the responsiveness of web application AJAX withASP.NET was incorporated (Garrett, 2005). The architecture of database is described in Figure 3.2.

![Figure 3.2: Architecture of AGPaseDB. Request made by user is directed to logical interface which parse the user request to database where all the information is stored physically and displaying the information.](image-url)