# CHAPTER V

## GENERAL DISCUSSION

The AGPase enzyme plays a pivotal role in starch biosynthesis for both photosynthetic and non-photosynthetic plant tissues. The catalytic activity and allosteric regulation of this enzyme has significantly contributed to the overall yield potential of many crop plants. In the present study, a detailed *in silico* sequence, structure and evolutionary analysis of the enzyme AGPase was performed to understand the mechanism of enzyme regulation in selected crop species.

A total of 37 SS and 87 LS mature protein and nucleotide sequence of AGPase was retieved from NCBI. Sequence analysis suggested that molecular weight of both the subunit ranges between 42.7 and 57 kDa. In case of monocots, SS's molecular weight ranges between 47.2 and 51.7 kDa and in dicots it ranges between 42.7 and 52.6 kDa. However LS of monocots are ranging between 50 and 52 kDa and dicots ranges between 47 and 57 kDa. It is evident from the molecular weight that SSs are smaller in size in comparison to the LS of AGPases. Physio-chemical properties computed for both the subunit signifies almost the similar behaviour. It was computed that almost all the SSs of AGPases were moderately acidic in nature, however, almost an equal percentage of the LS of AGPases were moderately acidic and basic in nature. The aliphatic index (AI) of both the subunits was very high indicating their thermostability. In comparison, SS showed slightly higher thermostability than the LS. Computed low GRAVY indices for both SS and LS of AGPase indicated that the enzyme has high affinity towards water.

Protein domain boundaries and architecture knowledge is essential for understanding and characterizing of protein function. Detection of protein domain and architecture in the absence of 3D structure benefits many areas of protein science, such as protein engineering and protein structure prediction [Kong and Ranganathan, 2004]. Putative conserved domains families, and superfamilies possessed by both the subunits of AGPase were predicted based on sequence similarity search with its closest orthologous family members. Various on line servers equivocally predicted similar domain architecture for both SS and LS. Both the subunits of AGPases were composed of an N-terminal catalytic ADP-glucose pyrophosphorylase domain and a C-terminal left-handed parallel beta helix domain. The N-terminal catalytic domain is approximately 250 residues and was structurally similar to Rossmann fold, which is typically present in nucleotide-binding domains. The C-terminal domain was composed of approximately 125 residues and is involved in co-operative allosteric regulation and oligomerization. Analysis of both the domain architecture and boundaries in a multiple alignment showed a high percentage of sequence conservation within both SS and LS of AGPase along with their structural homologue. This reflected the conservation of domains throughout the evolutionary period and suggested their conserved role in enzyme mechanism.

Overall functionality and efficiency of multiple domain proteins are affected by linker sequences. Cooperation and interaction between domains are affected by linker sequences which are flexible in 3D space, nonglobular, unstructured, or low complexity segment [Wootton 1994]. They keep the domains apart and provide great extent of flexibility to move individually. This phenomenon is a part of their catalytic function. The linker sequence joining the discrete domains of AGPases were inferred manually and

the amino acid propensities and order in linkers were examined. It was observed that both the SS and LS of AGPase had a linker peptide of approximately 37 amino acid residues which joined both the the domains present in it. Multiple sequence alignment of the linker region showed a high percentage of sequence identity between SS and LS of AGPase. Polar charged and uncharged hydrophilic residues and nonpolar hydrophobic residues occured in equal propensities in the linker region. The number of proline residues was not adequate for rigidity of the linker which keeps apart the discrete domains present in AGPase.

To determine the function of a protein, the 3D structure of the same is essential. In the absence of experimental 3D structures, comparative modeling of protein is considered as one of the most accurate method of model building and is often measured fundamental for understanding their function [Melo and Feytmans, 1998]. This approach provides reasonable result based on the assumption that the tertiary structure of two proteins will be similar if they share high percentage of sequence similarity [Bodade et. al., 2010]. It is widely being used when there is a clear relationship of homology between the target protein sequences and at least with an experimental (XRD or NMR) protein structure.

Based upon the BLASTP search against the PDB, it was observed that *Solanum tuberosum*, PDB id: 1YP3 (C Chain) shared the most sequence identity with a good query coverage and alignment quality with both the subunits of AGPases and was selected as the best template for model building. Various model building approaches were employed to obtain the most accurate model of SS and LS. Results from model building tools were analyzed and compared using different model quality assessment tools and the best models predicted by these tools were used for further work. PROCEHCK server was used

to qualify the stereo-chemical properties of the model based on Ramachandran plot statistics. A Ramachandran plot provides the position of the torsion angles phi (φ) and psi (ψ) between Cα-C and N-Cα atoms of the residues contained in a peptide. The PROCHECK analysis of both the subunit showed that Φ, Ψ angles of more than 84% residues belonged to the most favoured region of the Ramachandran plot and a very low percentage of residues (0-1.3%) falled in the disallowed region. Although there should not be any residues in the disallowed region of the Ramachandra plot, detailed analysis of the structure reflected that these residues belonged to the loop region of the structure and can be accepted for further analysis. Non-bonded interactions between different atom types in both the subunits were assessed by ERRAT tool which in return provides the overall quality of the protein. The acceptability of the predicted models for SS and LS was also confirmed with a very high ERRAT score of 69.47. Models with more than 50 ERRAT score are considered to be reliable. Compatibility of the 3D model with its primary sequence was validated through Verify_3D and a considerable amount of the side chain residues (average 87.76%) maximally lied above 0.2 as evident from Verify_3D program. Therefore, the side chain environment of the model was acceptable as residues with a score above 0.2 is generally considered dependable. Moreover, native protein folding energy of the predicted models were checked by ProSA tool. The Z score computed by ProSA was on an average -9.63 for SS and LS assuring good model quality. These cascade of model qualifiers checked the stereo chemical quality, nonbonded interactions of the residues, the compatibility of the side chain environment, packing quality and the energy profile of the predicted models and equivocally authenticated high

quality, reliability, acceptability and reproducibility of the proposed models that can be used for further analysis.

Theoretical 3D models of both SS and LS of AGPases belonging to different monocot and dicot species were analyzed extensively to have a wide spectrum on the 3D structure and the role of key residues responsible for catalytic and inhibitory functions. All the predicted structures of AGPase were composed of 12-18 helices (~22.3- 22.9%) and 22-27 strands (~24.5-31.4%). Moreover, other secondary structural elements i.e. beta alpha beta unit, beta hairpins, psi loop, beta bulges, helix-helix interactions, beta turns and gamma turns of both the subunit did not differ much from each other. Computed RMSD of the C$\alpha$ and backbone atom pairs for all the models were very low indicating that their secondary structural elements were conserved and had a common folding pattern.

After getting the structure of both SS and LS, different combinations of SS and LS of AGPases were docked to identify the most stable initial SS/SL/LL homo- or hetero-dimer. Most stable dimer was again docked with each other to obtain the complete heterotetrameric assembley i.e. α2β2. From the initial docking analysis it was found that SS and LS binds each other side by side to form the initial hetero-dimer and later SS and LS binded each other up and down to form the hetero-dimer 2. Later these two dimers binded complementary with each other to form the final heterotetrameric assembly. Comparative study of all the residues participated in the protein-protein interface interaction were mostly similar and parallel to those predicted by Tuncel and co-workers (2008) for the complete heterotetrameric assembly and Jin et al. (2005) for the SS, and almost all the residues of the interacting motifs in the two subunits were conserved in the selected species. All the surface interacting residues of allosterically regulated and non-

regulated SSs were conserved. However, surface interacting residue of allosterically regulated LS from endosperm (eg. *Zea mays*) at position 69, conserved Ala69 was mutated to Thr69 in all the allosterically non-regulated AGPases. Moreover, conserved Gly82 was mutated to Val82/Phe82. Conserved Lys317 was mutated to Arg317/Glu317 and Arg322 was mutated conservedly to Lys322 in all the allosterically non-regulated AGPases. Howevr other mutations found were not conserved.

Later the modeled 3D structure of both SS and LS of AGPase were selected for docking study to understand the inhibitor binding mechanism of both SS and LS. Previous study by Jin and co-workers (2005) on potato tuber AGPase SS reported that sulphate molecule which is an alalogoue of Pi acts as an inhibitor of the protein. Taking together in the present study, sulphate, which acted as an inhibitor of potato tuber AGPase was docked in to the active sites of both SS and LS of AGPases to elucidate its structural and functional relevance in terms of inhibitor binding. Three sulphate molecules were successively docked into the active site of both SS and LS of AGPase using CDOCKER algorithm. The docking of inhibitor into the active site of the modeled subunit revealed that in almost all cases of both SS and LS, Arg32 (equivalent to Arg 41 in potato tuber), Arg44, Lys395 and Lys432 were directly involved in the interaction with strong hydrogen and hydrophobics bonding to the first sulphate molecule of AGPase. Tyr135 which is an aromatic, partially hydrophobic, amino acid was substituted to polar Asn126 in most of the SS and LS, played a major role in binding with the second sulpfate molecule along with Lys60 and His125 via H-bonding. The third sulphate molecule binded with Arg44, Gln305 and Arg307 in almost all the SS and LS of AGPases, whereas H75 also showed its binding efficiency for sulpfate in some of the selected species. The

comparative docking study of sulphate inhibitor into the binding cavity of SS and LS of AGPases reflected a similar mode of binding specificity and contribution of similar residues in the interaction. Moreover structural superimposition of all the AGPase subunits showed that the secondary structure elements are superposed well and the key residues of allosteric regulation i.e. Arg32 equivalent to Arg41 in potato tuber, Arg44/Arg53, Lys60/Lys69, His75/His84, His125/His134, Gln305/Gln314, Arg307/Arg316, Lys395/Lys404 and Lys432/Lys441 were almost conserved and were allosterically significant. This findings strongly suggest a similar regulator binding mechanism of AGPases in both SS and LS in relation to the template protein.

Tyr135 which was a key residue within the active site of potato tuber SS for allosteric regulation was substituted to Asn126 in almost all the SS and LS of AGPases. Previous study on this enzyme reports that GXGXRL loop, PAVP motif and residue equivalent to Arg33 in potato tuber SS plays a key role for ATP binding and it has been demonstrated by mutagenesis study. Structural superimposition showed the strong conservation of GXGXRL loop positioning 20-25, PAVP motif positioning 35-38 and conservation of Arg24 (equivalent to Arg33 of potato tuber AGPase) in most of the SS and LS of AGPases firmly reflected the similar mode of action in this family of enzymes. However a certain number of the LS possess some mutations at the respective positions which might corresponded to their inability to catalysis. Amino acid residues R33 and K43 (potato tuber numbering) have been shown to be critical for catalysis [Ballicora et al., 2005]. These two amino acids were conserved in all SSs but not abundant in LS. Arg33 was mutated to Lys, His, and Gln in the respective positions of most of the LS AGPases under current investigation. Moreover catalytic Lys43 was replaced by Thr43 in a

number of LS. Because of these two mutations LS of AGPase were thought to be modulatory rather than catalytic. However depending upon these two mutations, Ventriglia and co-workers (2008) demonstrated that two *Arabidopsis* AGPase LS (APL1 and APL2) were catalytic as they contain these two key residues (Arg and Lys) in the substrate binding regions and propsed the vital role of these two residues. Taking together, the availability of these two residues were screened in all the LS selected under the present investigation. The presence of these residues were found in a number selected LS. Moreover structural superimposition of this region with the catalytically active SS showed strong similiarity within the binding pocket with very low RMSD difference. Therefore it may be that the LS having Arg33 and Lys43 in the substrate binding pocket were catalytic in nature although it needs further confirmatory biochemical investigations.

Previous study by Jin and co-workers (2005) concluded that metal mediated catalytic mechanism is also used by AGPase. Residues equivalent to Asp145 and Asp280 (in potato tuber) chelates the metal ion and plays a crucial role in metal-mediated catalytic mechanism. In several organisms the absolute need of a metal ion for AGPase has been biochemically demonstrated. Taken together, the present study checked the binding specificity of these residues in the selected models to have a better confidence about the importance and involvement of these residues in metal mediated catalytic mechanisms across the family. Structural superimposition of the modeled proteins with the template protein revealed that these two residues were strongly conserved in almost all the SS and LS of AGPases, and followed common folding pattern with a very low RMSD difference which may prompt for metal mediated catalytic mechanism.

A detailed phylogenetic analysis of the enzyme AGPase was performed to gain insights into its evolutionary history. With the sequence homology search it was evident that AGPase and other NDP-sugar PPase enzymes are derived from a common ancestor as they contain a similiar pyrophosphorylase domain with a similar fold. However the main difference between AGPase and NDP-sugar PPase enzymes is that the formers are allosterically regulated. Moreover the N and C-terminus ends of AGPases are bigger and extended in comparision to others. Whereas, in case of other NDP-sugar PPases this extended C-terminus end is either absent or a part of it forms a completely different domain to become a bifunctional enzyme.

A phylogenetic tree of the selected AGPase sequences provided information about the origin of the sub-units. In multicellular organisms, SS and LS formed two and four distinct groups respectively in the phylogenetic tree. The phylogentic tree of the higher plant SS was simple and can be divided in to two subgroups: one corresponding to monocots and a second one corresponding to dicots. Whereas, AGPase LS has narrower tissue specificity than SS, and the LS phylogeny appeared more complex (with four major clades some of which include both monocots and dicots) than the SS phylogeny. The first LS group, generally expressed in photosynthetic tissues (leaves), comprised LS from monocots and dicots. Second phylogenetic group displayed a broader expression pattern and are expressed both in source and sink tissues. The group 3 genes are expressed in sink tissues (these genes are subdivided into group 3a in dicots and group 3b in monocots), and group 4 reffered to a clade of that have not been characterized yet in terms of function and expression patterns. Comparison of SS and LS tree results

emphasize that the LS underwent a larger number of duplications than did the SS and that only LS duplications began before the divergence of monocots and dicots.

The modulatory and catalytic variants of LS can not be ascertained at this point only by phylogenetic analysis as this is based upon primary structure rather than function. However, with structure-function relationship information, theoretically trace function or the absence of it in the tree can be predicted. Both groups 3(a) and 3(b) lacked the residues Arg33 and Lys43 (potato tuber numbering) which are critical in substrate binding. Group 3(a) has Lys, His, and Gln, and group 3(b) has Gln, respectively, in place of a homologous Arg33. A distinctive characteristic in both the groups is that a Thr43 replaces the catalytic Lys43. Therefore with the phylogenetic and structure function relationship we can conclude that the modulatory LS from group 1, 2 and 4 of the phylogenetic tree may be catalytic too as they contain in their sequence the important residues (Arg33 and Lys43) for catalysis, and not much alteration in the substarte binding pocket was observed.

Sequence comparison, structural analysis, docking study and molecular evolutionary analysis reflected a small number of amino acid residues to be critical for enzyme regulations. Based on the docking study it was evident that Arg32 (equivalent to Arg41 in potato tuber) Arg44, Lys60, His75, His125, Gln305, Arg307, Lys395 and Lys432 are crucial for regulator binding. Therefore to investigate the importance of these residues in regulator binding these reidues were mutated one by one to alanine at their corresponding positions and monitored the change in binding energy between the wild and mutated ones. Upon mutations, it was evident that Arg_44_Ala, Lys_60_Ala and Arg_307_Ala are chiefly responsible for decreasing the binding energy of $1^{st}$, $2^{nd}$ and $3^{rd}$ sulphate

binding to the receptor molecule. Moreover, Arg44 was equally responsible for binding with 1$^{st}$ as well as 2$^{nd}$ sulphate molecule. However, all other residues may be presumed to be essential for the stability and interaction of the regulator. Thus, the results of the present study suggest that Arg44, Lys60 and Arg307 are the key amino acids for inhibiting regulator binding and further biochemical investigations of the proposed residues might help in development of a custom designed AGPase for greater starch yield.

## CONCLUSIONS

Importance of starch is immense both in case of plants and humans and increase in starch yield is higly recommended to meet the socio-economic and environmental demands. Genetic manipulation of enzymes have been useful as a strategy to increase crop yield that in turn enhances starch production. AGPase, a rate limiting heterotetrameric enzyme is being studied extensively for increasing starch production as it determines the rate of starch synthesis in plants.

This study suggests no difference in the SS of allosterically regulated and non-regulated AGPases, however four strikingly conserved mutations have been observed in the LS of allosterically regulated and non-regulated AGPases. These four mutations hold tremendous scope for future investigations and it may be prosposed that they are critical for being allosterically regulated and non-regulated isoforms in the edosperm of the selected species. This proposition however needs further wet lab validations.

Comparative sequence, structure and docking study suggested that GXGXRL loop, PAVP motif, Arg24 and Lys34 (equivalent to Arg33 and Lys43 of potato tuber AGPase) and Arg32, Arg44, Lys60, His75, His125, Asp126, Gln305, Arg307, Arg361, Lys395 and Lys432 are the key residues and motifs responsible for regulator and substrate binding. These are highly conserved in both the SS and LS of AGPases in majority of the cases, Some LS are catalytic and have been grouped together in the phylogenetic tree. The findings of the present study holds scope towards development of a custom designed AGPase for achieving greater starch yield.