CHAPTER IV

RESULTS

Objective 1: Molecular modeling of both small and large subunits of AGPase using available crystal structures.

Following methodology was applied to complete the 1st objective.

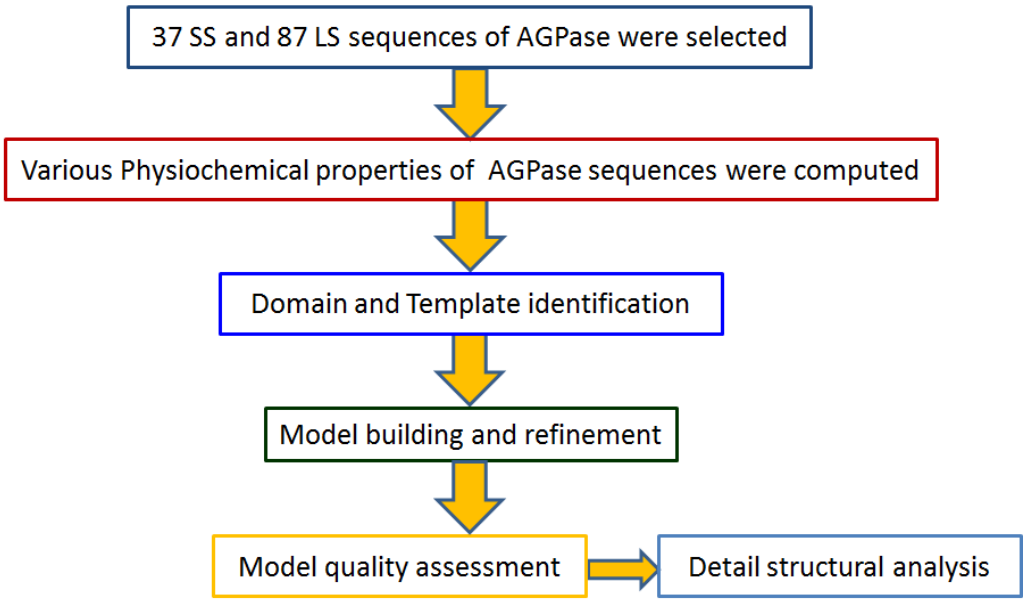


Figure 4: Step by step workflow of Objective 1.

Sequence retrival

Fasta formatted amino acid sequence of both SS and LS of AGPase from different monocot and dicot crop plants were retrieved from the NCBI. Brief details about the sequences that were investigated in the present study are tabulated in Tables 1 and 2.

Table 1: Small subunit of AGPase sequences analyzed in this study. GenBank accession no., organism name, total number of amino acids and their sub-cellular locations are mentioned.

GenBank Accession no.	**Source (organism name)**	**No. of amino acids**	**Sub-cellular location**
MONOCOTS			
1143500	*Hordeum vulgare subsp. vulgare*	472	Seed endosperm
27464770	*Hordeum vulgare subsp. vulgare*	501	Seed endosperm
1707940	*Hordeum vulgare subsp. vulgare*	513	Unknown
50944557	*Oryza sativa*	479	Seed endosperm
13508485	*Oryza sativa*	500	Seed endosperm, leaf
125987830	*Oryza sativa*	514	Unknown
242048788	*Sorghum bicolor*	510	Unknown
7340287	*Triticum aestivum*	473	Endosperm, leaf
52430025	*Triticum aestivum*	498	Endosperm
33340600	*Zea mays*	475	Seed endosperm, leaf
14582768	*Zea mays*	517	Leaf
13892040	*Zea mays*	510	Seed embryo
DICOTS			
1575754	*Arabidopsis thaliana*	520	Leaf, stem, inflorescence
556622	*Beta vulgaris subsp. vulgaris*	489	Unknown
7688095	*Brassica napus*	520	Leaf, seed
13487709	*Brassica rapa subsp. pekinensis*	519	Leaf
16950559	*Cicer arietinum*	505	Seed, weak in leaf
13487787	*Cicer arietinum*	516	Seed, weak in leaf
2642636	*Citrullus lanatus subsp. vulgaris*	526	Fruit
5917789	*Citrus unshiu*	445	Fruit, leaf

2625084	*Cucumis melo*	525	Leaf, early fruit, weak in root and stem
41350641	*Fragaria x ananassa*	521	Unknown
603565	*Ipomoea batatas*	427	Tuber, leaf
1771495	*Ipomoea batatas*	523	Tuber, stem, leaf
7671230	*Perilla frutescens*	523	Stem, root, seed, leaf
7671232	*Perilla frutescens*	520	Stem, root, seed, leaf
29421116	*Phaseolus vulgaris*	515	Seed
1237082	*Pisum sativum*	507	Seed, weak in leaf, stem, root
1237080	*Pisum sativum*	516	Seed, leaf
1325984	*Solanum lycopersicum*	521	Leaf, fruit
21475	*Solanum tuberosum*	521	Unknown
232164	*Solanum tuberosum*	521	Tuber, leaf
27819107	*Solanum tuberosum*	521	Unknown
77416911	*Solanum tuberosum*	521	Unknown
633678	*Spinacia oleracea*	444	Leaf
440593	*Vicia faba var. minor*	508	Seed
440595	*Vicia faba var. minor*	512	Leaf, seed

Table 2: Large subunit of AGPase sequences analyzed in this study. GenBank accession no., organism name, total no. amino acids and their sub-cellular locations are mentioned.

GenBank Accession no.	**Source (organism name)**	**No. of amino acids**	**Sub-cellular Location**
MONOCOTS			
2627058	*Oryza sativa*	519	Seed endosperm
1565308	*Oryza sativa*	514	Seed endosperm

50582723	*Oryza sativa*	511	Leaf
34900728	*Oryza sativa*	524	Stem, leaf, seed
2735840	*Sorghum bicolor*	517	Seed endosperm
242088961	*Sorghum bicolor*	519	Unknown
242053733	*Sorghum bicolor*	517	Unknown
242033053	*Sorghum bicolor*	507	Unknown
995746	*Triticum aestivum*	522	Seed endosperm
32812836	*Triticum aestivum*	522	Unknown
89277026	*Triticum aestivum*	503	Unknown
1707930	*Triticum aestivum*	522	Unknown
121293	*Triticum aestivum*	500	Unknown
1707924	*Zea mays*	516	Seed endosperm
162463875	*Zea mays*	514	Unknown
558365	*Zea mays*	521	Embryo
162460455	*Zea mays*	505	Unknown
189027076	*Zea mays*	516	Unknown
806638577	*Zea mays*	518	Unknown
1279513	*Hordeum vulgare*	523	Seed endosperm
2105137	*Hordeum vulgare*	503	Leaf, weak in endosperm
445623	*Hordeum vulgare*	527	Unknown
357132398	*Brachypodium distachyon*	522	Unknown
721619253	*Brachypodium distachyon*	512	Unknown
357116651	*Brachypodium distachyon*	504	Unknown
22347636	*Oncidium hybrid cultivar*	517	Leaf, flower
DICOTS			
297812109	*Arabidopsis lyrata subsp. lyrata*	522	Unknown
297845724	*Arabidopsis lyrata subsp. lyrata*	518	Unknown
297797902	*Arabidopsis lyrata subsp. lyrata*	521	Unknown
297821353	*Arabidopsis lyrata subsp. lyrata*	520	Unknown

20258949	*Arabidopsis thaliana*	518	Low level of expression
20465645	*Arabidopsis thaliana*	523	Sink tissues
3893079	*Arabidopsis thaliana*	521	Sink tissues
4586350	*Arabidopsis thaliana*	522	Leaf
556624	*Beta vulgaris subsp. vulgaris*	522	Leaf, weak in hypocot, root
13487711	*Brassica rapa*	570	Leaf
13487783	*Cicer arietinum*	525	Leaf
13487785	*Cicer arietinum*	521	Seeds, weak in leaf, stem and root
2642640	*Citrullus lanatus subsp. vulgaris*	481	Fruit
2642638	*Citrullus lanatus subsp. vulgaris*	526	Fruit
5917791	*Citrus unshiu*	531	Fruit, leaf
2625086	*Cucumis melo*	525	Stem, root, early fruit, weak in leaf
2625088	*Cucumis melo*	518	Stem, root, early fruit, weak in leaf
41350643	*Fragaria x ananassa*	507	Unknown
571559815	*Glycine max*	529	Unknown
356563435	*Glycine max*	523	Unknown
356562361	*Glycine max*	524	Unknown
356553863	*Glycine max*	523	Unknown
356545193	*Glycine max*	528	Unknown
356538761	*Glycine max*	530	Unknown
359807441	*Glycine max*	520	Unknown
356518710	*Glycine max*	531	Unknown
356517038	*Glycine max*	519	Unknown
356509672	*Glycine max*	530	Unknown
356508352	*Glycine max*	519	Unknown

571446804	*Glycine max*	529	Unknown
3211989	*Ipomoea batatas*	517	Tuber, stem
922376662	*Medicago truncatula*	518	Unknown
922340593	*Medicago truncatula*	528	Unknown
922364294	*Medicago truncatula*	521	Unknown
357473317	*Medicago truncatula*	528	Unknown
7671234	*Perilla frutescens*	527	Stem, cotyledon
29421114	*Phaseolus vulgaris*	525	Seed
1237086	*Pisum sativum*	510	Sink tissues: seed, pod, seed coat
566203612	*Populus trichocarpa*	530	Unknown
224103389	*Populus trichocarpa*	527	Unknown
224100249	*Populus trichocarpa*	528	Unknown
566173136	*Populus trichocarpa*	528	Unknown
566166858	*Populus trichocarpa*	526	Unknown
224062107	*Populus trichocarpa*	528	Unknown
255552303	*Ricinus communis*	533	Unknown
255548169	*Ricinus communis*	528	Unknown
255543725	*Ricinus communis*	531	Unknown
255538708	*Ricinus communis*	523	Unknown
5923897	*Solanum habrochaites*	520	Unknown
1947084	*Solanum lycopersicum*	524	Stems, roots, early fruit
1840116	*Solanum lycopersicum*	516	Leaf
1840114	*Solanum lycopersicum*	518	Fruit, root
1778434	*Solanum lycopersicum*	516	Unknown
1778436	*Solanum lycopersicum*	518	Unknown
426476	*Solanum tuberosum*	483	Tuber
400489	*Solanum tuberosum*	519	Leaf, tuber
21473	*Solanum tuberosum*	470	Tuber, weak in leaf

225458219	*Vitis vinifera*	524	Unknown
359479997	*Vitis vinifera*	514	Unknown
225432564	*Vitis vinifera*	520	Unknown
225428422	*Vitis vinifera*	527	Unknown

Primary sequence analysis

Primary structural analysis of the retrieved protein sequences from both SS and LS of AGPase were done by computing various physico-chemical properties such as molecular weight, isoelectric point, instability index, aliphatic index and grand average hydropathy (GRAVY) using ProtParam tool. Results are shown in Tables 3 and 4.

Table 3: Physico-chemical properties of small subunit of AGPases.

GenBank Accession no.	**Molecular weight**	**Theoretical pI**	**(Asp + Glu)**	**(Arg + Lys)**	**Instability index**	**Aliphatic index**	**GRAVY**
1143500	52010.4	5.59	60	48	33.54	93.43	-0.176
27464770	54779.4	5.79	60	54	38.81	89.02	-0.174
1707940	56049.2	6.11	60	55	36.52	91.72	-0.118
50944557	52950.4	5.87	61	55	37.92	92.65	-0.238
13508485	54745.5	6.06	58	54	37.17	89.44	-0.124
125987830	56104.0	6.58	59	57	42.56	90.37	-0.159
242048788	56011.9	6.24	60	57	42.73	90.71	-0.153
7340287	52061.3	5.53	61	49	34.98	90.97	-0.214
52430025	54371.8	5.67	60	53	39.48	89.18	-0.160
33340600	52197.7	5.46	59	50	33.55	91.87	-0.164
14582768	56482.4	6.48	60	57	38.92	88.74	-0.202
13892040	55936.9	6.07	59	55	42.08	89.55	-0.133
1575754	56598.4	5.89	60	54	35.21	92.85	-0.129
556622	53796.2	5.59	59	52	38.06	91.19	-0.171
7688095	57044.8	5.86	63	57	37.39	91.38	-0.215
13487709	57087.8	5.87	63	57	37.46	91.56	-0.232
16950559	55309.9	6.20	60	56	34.80	91.19	-0.186

13487787	56368.2	6.49	59	57	39.54	90.19	-0.161
2642636	57257.1	7.62	60	61	37.60	89.60	-0.180
5917789	49296.3	5.66	58	51	32.72	92.54	-0.191
2625084	57435.4	7.59	60	61	38.20	91.07	-0.175
41350641	57111.2	7.07	60	60	37.00	88.00	-0.225
603565	47300.2	6.13	52	48	36.29	97.12	-0.119
1771495	57179.2	8.02	61	63	36.64	90.52	-0.183
7671230	57474.6	6.69	63	62	39.71	88.45	-0.198
7671232	57140.1	6.44	61	59	31.77	87.62	-0.171
29421116	56212.0	6.43	58	56	34.93	90.52	-0.162
1237082	55566.4	6.69	59	58	37.32	91.58	-0.172
1237080	56330.0	6.13	61	57	41.66	89.81	-0.178
1325984	57370.4	6.49	62	60	44.37	91.96	-0.201
21475	57240.3	6.73	61	60	44.18	91.21	-0.196
232164	57240.3	6.73	61	60	44.18	91.21	-0.196
27819107	57175.2	6.29	62	59	43.54	93.26	-0.191
77416911	57277.4	6.73	61	60	43.53	92.32	-0.193
633678	48980.8	5.20	58	46	34.86	95.38	-0.152
440593	55627.4	6.43	58	56	37.94	91.79	-0.160
440595	56059.6	6.19	61	57	40.56	90.4	-0.196

Table 4: Physico-chemical properties of large subunit of AGPases.

GenBank Accession no.	**Molecular weight**	**Theoretical pI**	**Asp + Glu**	**Arg + Lys**	**Instability index**	**Aliphatic index**	**GRAVY**
2627058	57547.6	6.48	62	59	35.39	83.60	-0.204
1565308	57153.4	6.37	60	56	38.89	86.15	-0.187
50582723	55427.0	7.01	60	60	35.55	88.26	-0.129
34900728	57472.8	7.89	58	60	39.24	85.29	-0.085
2735840	57096.8	5.76	61	51	34.03	84.68	-0.202
242088961	57570.7	6.40	61	58	39.91	83.04	-0.165
242053733	57096.8	5.76	61	51	34.03	84.68	-0.202
242033053	55381.1	8.33	59	62	37.03	89.86	-0.132
995746	57808.7	6.12	65	60	42.86	80.52	-0.253
32812836	57706.5	5.89	66	59	43.58	79.58	-0.270

89277026	54774.4	8.31	58	61	33.32	87.69	-0.133
1707930	57808.7	6.12	65	60	42.86	80.52	-0.253
121293	55560.0	6.61	59	57	43.12	75.90	-0.356
1707924	57071.0	6.16	60	55	38.88	85.06	-0.207
162463875	56182.3	7.07	55	55	37.81	84.49	-0.072
558365	57988.5	8.37	57	62	43.98	80.86	-0.218
162460455	55292.1	8.57	59	63	36.53	89.82	-0.137
189027076	57071.0	6.16	60	55	38.88	85.06	-0.207
806638577	57450.6	6.40	61	58	39.92	82.45	-0.171
1279513	57932.9	6.17	65	60	43.13	80.75	-0.242
2105137	54849.4	8.49	58	62	34.95	87.12	-0.148
445623	58188.4	7.54	61	62	48.55	79.01	-0.257
357132398	57901.0	6.06	65	59	40.50	85.19	-0.153
721619253	56340.5	6.73	56	55	42.49	84.92	-0.118
357116651	54899.3	7.06	61	61	37.65	88.85	-0.141
22347636	57080.2	8.34	59	62	34.76	89.83	-0.163
297812109	57611.6	8.48	61	65	40.83	90.44	-0.189
297845724	57359.7	7.12	56	56	38.55	87.51	-0.153
297797902	58181.2	8.52	61	66	37.89	81.63	-0.302
297821353	58093.4	8.37	59	63	38.07	82.52	-0.223
20258949	57469.8	7.15	57	57	36.06	87.32	-0.176
20465645	58205.4	8.36	59	63	36.46	80.92	-0.257
3893079	58029.3	8.66	61	68	37.95	82.00	-0.290
4586350	57701.8	8.30	61	64	43.09	92.87	-0.176
556624	57716.5	6.05	62	57	31.03	81.99	-0.208
13487711	63145.1	9.04	64	75	47.60	87.42	-0.246
13487783	58012.3	7.60	62	63	38.38	89.73	-0.160
13487785	57795.8	6.77	58	57	32.59	84.05	-0.193
2642640	53522.9	5.78	59	53	27.75	85.97	-0.195
2642638	58375.8	7.55	58	59	34.26	86.39	-0.169
5917791	58918.5	8.47	63	68	32.82	85.40	-0.212
2625086	58454.0	8.64	56	62	30.37	83.73	-0.198
2625088	57725.4	5.56	66	57	38.86	82.10	-0.289
41350643	56446.2	5.91	59	54	37.68	84.99	-0.179
571559815	58501.3	8.72	50	58	38.08	88.85	-0.053
356563435	58147.3	5.52	60	64	38.56	87.07	-0.240

356562361	58054.1	7.07	57	57	38.17	80.40	-0.206
356553863	57882.1	7.92	56	58	38.77	82.79	-0.188
356545193	58501.9	8.54	58	63	34.05	87.52	-0.202
356538761	58584.0	8.23	57	60	35.79	89.77	-0.153
359807441	57785.9	8.66	60	65	37.97	86.46	-0.254
356518710	58768.0	8.50	59	64	31.57	81.17	-0.229
356517038	57001.9	6.39	59	56	34.59	90.92	-0.140
356509672	58915.1	8.39	59	63	36.62	82.60	-0.235
356508352	57070.9	5.98	61	55	34.10	91.31	-0.145
571446804	58352.1	8.56	51	57	40.08	88.11	-0.062
3211989	57686.9	7.55	63	64	38.55	86.92	-0.234
922376662	57246.2	6.80	58	57	28.53	86.58	-0.194
922340593	58560.3	8.86	52	62	40.05	86.23	-0.120
922364294	57533.7	7.62	63	64	36.63	86.10	-0.216
357473317	58257.3	6.68	60	59	31.56	84.19	-0.175
7671234	58394.9	8.09	62	65	33.63	86.26	-0.193
29421114	57834.5	6.33	62	58	34.23	86.00	-0.240
1237086	56166.0	7.07	58	58	30.63	86.80	-0.197
566203612	59043.8	8.95	55	64	38.81	83.70	-0.173
224103389	58606.7	7.07	63	63	33.52	81.23	-0.238
224100249	58264.8	8.89	57	64	36.62	88.84	-0.154
566173136	58325.7	8.55	56	61	27.85	84.37	-0.181
566166858	58255.3	6.58	61	59	33.56	84.73	-0.192
224062107	58209.6	7.03	58	58	26.95	84.96	-0.133
255552303	58578.8	8.73	58	63	35.84	88.59	-0.152
255548169	58723.1	7.20	58	58	38.54	86.06	-0.182
255543725	58893.3	7.56	61	62	37.57	85.20	-0.227
255538708	58184.7	7.46	59	60	31.20	83.90	-0.158
5923897	57833.1	7.95	65	67	31.67	85.13	-0.269
1947084	58213.6	6.64	66	65	31.54	86.74	-0.209
1840116	57096.3	8.32	60	63	36.21	90.52	-0.156
1840114	57292.7	8.41	55	59	31.96	83.38	-0.145
1778434	57414.6	7.01	64	64	32.27	88.06	-0.223
1778436	57353.7	8.19	55	58	32.75	83.19	-0.143
426476	53602.3	8.92	55	62	36.44	88.03	-0.214
400489	57399.8	6.43	59	57	31.71	83.60	-0.130

21473	52252.7	6.09	59	56	33.90	87.55	-0.188
225458219	58330.8	8.49	59	64	34.46	86.51	-0.206
359479997	57351.0	6.84	60	59	41.77	87.28	-0.140
225432564	57954.4	8.88	61	68	35.46	86.10	-0.217
225428422	58541.2	8.51	61	66	35.31	86.98	-0.202

Domain analysis

To have a better confidence on various putative domains, families and superfamilies possessed by both the SS and LS of AGPases, different online computational programs were used to have a consensus outcome. Moreover the linker between the different domains and the domain boundaries were delineated with manual inspection. Different domain possessed by both SS and LS AGPases along with the domain boundary information are shown in Tables 5 and 6 respectively.

Template identification, model refinement and structural assessment

Based upon the BLASTP of all the query sequences against the PDB, *Solanum tuberosum*, PDB id: 1YP3 (C Chain) was selected as the best template for model building of both SS and LS of AGPases. Keeping 1YP3_C as the template, different model building approaches were employed to predict the theoretical 3D modeling of both SS and LS of AGPases. Prior to modeling, the amyloplast target sequence of both SS and LS of AGPase at the N-terminal end was removed to exclude the random coil fragment at this region and to achieve a better global superimposition with the template. Template identified for model building of both SS and LS AGPases are shown in Tables 5 and 6 along with their sequence identity with the template and model building coverage and identity.

Table 5: Respective domains are identified in each small subunit along with their sequence positions. Sequence identity of the query sequence with the template and the coverage and confidence of the predicted 3D model is shown.

Accession	Species	Domain1	Domain2	Sequence identity with the closest homologue	Coverage and confidence	Template identified for model building
21475	*Solanum tuberosum*	102-351	389-515	100	435 residues (83% of the query sequence) have been modeled with 100.0% confidence	1YP3C
232164	*Solanum tuberosum*	102-351	389-515	100	435 residues (83% of the query sequence) have been modeled with 100.0% confidence	1YP3C
440593	*Vicia faba var. minor*	89-338	376-502	94	435 residues (86% of the query sequence) have been modeled with 100.0% confidence	1YP3C
440595	*Vicia faba var. minor*	93-342	380-506	94	435 residues (85% of the query sequence) have been modeled with 100.0% confidence	1YP3C
556622	*Beta vulgaris subsp. vulgaris*	82-331	369-483	90	422 residues (86% of the query sequence) have been modeled with 100.0% confidence	1YP3C
603565	*Ipomoea batatas*	1-257	295-421	89	419 residues (98% of the query sequence) have been modeled with 100.0% confidence	1YP3C
633678	*Spinaciao leracea*	25-274	312-438	95	435 residues (98% of the query sequence)	1YP3C

					have been modeled with 100.0% confidence	
1143500	*Hordeum vulgare subsp. vulgare*	53-302	340-465	92	432 residues (92% of the query sequence) have been modeled with 100.0% confidence	1YP3C
1237080	*Pisum sativum*	97-346	384-510	93	435 residues (84% of the query sequence) have been modeled with 100.0% confidence	1YP3C
1237082	*Pisum sativum*	88-337	375-501	94	435 residues (86% of the query sequence) have been modeled with 100.0% confidence	1YP3C
1325984	*Solanum lycopersicum*	102-351	389-515	98	435 residues (83% of the query sequence) have been modeled with 100.0% confidence	1YP3C
1575754	*Arabidopsis thaliana*	101-350	388-514	93	435 residues (84% of the query sequence) have been modeled with 100.0% confidence	1YP3C
1707940	*Hordeum vulgare subsp. vulgare*	94-343	381-506	92	435 residues (85% of the query sequence) have been modeled with 100.0% confidence	1YP3C
1771495	*Ipomoea batatas*	104-353	391-517	94	435 residues (83% of the query sequence) have been modeled with 100.0% confidence	1YP3C
2625084	*Cucumi smelo*	107-355	393-519	94	435 residues (83% of the query sequence)	1YP3C

					have been modeled with 100.0% confidence	
2642636	*Citrullus lanatus subsp. vulgaris*	107-356	394-520	94	435 residues (83% of the query sequence) have been modeled with 100.0% confidence	1YP3C
5917789	*Citrus unshiu*	96-345	383-509	94	432 residues (84% of the query sequence) have been modeled with 100.0% confidence	1YP3C
7340287	*Triticum aestivum*	54-303	341-466	92	432 residues (91% of the query sequence) have been modeled with 100.0% confidence	1YP3C
7671230	*Perilla frutescens*	95-353	391-517	93	435 residues (83% of the query sequence) have been modeled with 100.0% confidence	1YP3C
7671232	*Perilla frutescens*	99-350	388-514	87	434 residues (83% of the query sequence) have been modeled with 100.0% confidence	1YP3C
7688095	*Brassica napus*	101-350	388-514	94	435 residues (84% of the query sequence) have been modeled with 100.0% confidence	1YP3C
13487709	*Brassica rapa subsp. pekinensis*	100-349	387-513	93	435 residues (84% of the query sequence) have been modeled with 100.0% confidence	1YP3C
13487787	*Cicer arietinum*	97-346	384-510	93	435 residues (84% of the query sequence)	1YP3C

					have been modeled with 100.0% confidence	
13508485	*Oryza sativa*	81-330	368-494	88	435 residues (87% of the query sequence) have been modeled with 100.0% confidence	1YP3C
13892040	*Zea mays*	91-340	378-504	88	435 residues (85% of the query sequence) have been modeled with 100.0%	1YP3C
14582768	*Zea mays*	98-347	385-510	91	435 residues (84% of the query sequence) have been modeled with 100.0% confidence	1YP3C
16950559	*Cicer arietinum*	86-335	373-499	94	435 residues (86% of the query sequence) have been modeled with 100.0% confidence	1YP3C
27464770	*Hordeum vulgare subsp. vulgare*	82-331	369-495	87	435 residues (87% of the query sequence) have been modeled with 100.0% confidence	1YP3C
27819107	*Solanum tuberosum*	102-351	389-515	99	435 residues (83% of the query sequence) have been modeled with 100.0% confidence	1YP3C
29421116	*Phaseolus vulgaris*	96-345	383-509	92	435 residues (84% of the query sequence) have been modeled with 100.0% confidence	1YP3C
33340600	*Zea mays*	56-305	343-468	90	434 residues (91% of the query sequence) have been modeled with 100.0%	1YP3C

					confidence	
41350641	*Fragaria x ananassa*	93-351	389-513	91	434 residues (83% of the query sequence) have been modeled with 100.0% confidence	1YP3C
50944557	*Oryza sativa*	60-309	60-309	91	432 residues (90% of the query sequence) have been modeled with 100.0% confidence	1YP3C
52430025	*Triticum aestivum*	79-328	366-492	87	435 residues (87% of the query sequence) have been modeled with 100.0% confidence	1YP3C
77416911	*Solanum tuberosum*	102-351	389-515	99	435 residues (83% of the query sequence) have been modeled with 100.0% confidence	1YP3C
125987830	*Oryza sativa*	95-344	382-507	92	435 residues (85% of the query sequence) have been modeled with 100.0% confidence	1YP3C
242048788	*Sorghum bicolor*	91- 340	378-504	88	435 residues (85% of the query sequence) have been modeled with 100.0% confidence	1YP3C

Table 6: Respective domains are identified in each large subunit along with their sequence positions. Sequence identity of the query sequence with the template and the coverage and confidence of the predicted 3D model is shown.

Accession	Species	Domain1	Domain2	Sequence identity with the closest homologue	Coverage and confidence	Template identified for mode building
21473	*Solanum tuberosum*	40-301	338-464	52	429 residues (91% of the query sequence) have been modelled with 100.0% confidence	1yp3C
121293	*Triticum aestivum*	73-334	371-494	49	422 residues (84% of the query sequence) have been modeled with 100.0% confidence	1yp3C
400489	*Solanum tuberosum*	91-350	387-513	52	427 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
426476	*Solanum tuberosum*	51-314	351-477	54	424 residues (88% of the query sequence) have been modeled with 100.0% confidence	1yp3C
445623	*Hordeum vulgare*	97-358	395-521	49	432 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
556624	*Beta vulgaris subsp. vulgaris*	91-353	390-516	54	427 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
558365	*Zea mays*	89-351	388-515	47	427 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
995746	*Triticum aestivum*	92-353	390-516	49	426 residues (83% of the query sequence) have been modeled with 100.0% confidence	1yp3C
1237086	*Pisum sativum (Pea)*	81-341	378-504	54	425 residues (83% of the query sequence) have been modeled with 100.0%	1yp3C

					confidence	
1279513	*Hordeum vulgare*	93-354	391-517	50	425 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
1565308	*Oryza sativa*	88-349	394-504	47	409 residues (80% of the query sequence) have been modeled with 100.0% confidence	1yp3C
1707924	*Zea mays*	86-347	384-510	49	426 residues (83% of the query sequence) have been modeled with 100.0% confidence	1yp3C
1707930	*Triticum aestivum*	92-353	390-516	49	425 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
1778434	*Solanum lycopersicum*	87-347	384-510	52	426 residues (84% of the query sequence) have been modeled with 100.0% confidence	1yp3C
1778436	*Solanum lycopersicum*	88-349	386-512	53	431 residues (83% of the query sequence) have been modeled with 100.0% confidence	1yp3C
1840114	*Solanum lycopersicum*	88-349	386-512	55	428 residues (83% of the query sequence) have been modeled with 100.0% confidence	1yp3C
1840116	*Solanum lycopersicum*	84-347	384-510	56	425 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
1947084	*Solanum lycopersicum*	94-355	392-518	52	425 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
2105137	*Hordeum vulgare*	72-334	371-497	57	432 residues (86% of the query sequence) have been modeled with 100.0% confidence	1yp3C
2625086	*Cucumis melo*	96-356	393-519	52	425 residues (81% of the query sequence)	1yp3C

					have been modeled with 100.0% confidence	
2625088	*Cucumis melo*	89-349	386-512	53	424 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
2627058	*Oryza sativa*	89-350	387-513	50	425 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
2642638	*Citrullus lanatus subsp. vulgaris*	97-357	394-520	52	425 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
2642640	*Citrullus lanatus subsp. vulgaris*	51-312	349-475	57	426 residues (91% of the query sequence) have been modeled with 100.0% confidence	1yp3C
2735840	*Sorghum bicolor*	87-348	385-511	48	427 residues (83% of the query sequence) have been modeled with 100.0% confidence	1yp3C
3211989	*Ipomoea batatas*	87-348	385-511	51	426 residues (84% of the query sequence) have been modeled with 100.0% confidence	1yp3C
3893079	*Arabidopsis thaliana*	91-352	389-515	50	429 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
4586350	*Arabidopsis thaliana*	93-353	390-516	56	428 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
5917791	*Citrus unshiu*	97-362	399-525	52	421 residues (79% of the query sequence) have been modeled with 100.0% confidence	1yp3C
5923897	*Solanum habrochaites*	90-351	388-514	51	428 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C

7671234	*Perilla frutescens*	97-358	395-521	52	425 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
13487711	*Brassica rapa*	88-372	409-564	56	428 residues (75% of the query sequence) have been modeled with 100.0% confidence	1yp3C
13487783	*Cicer arietinum*	94-356	393-519	56	430 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
13487785	*Cicer arietinum*	90-352	389-515	55	426 residues (83% of the query sequence) have been modeled with 100.0% confidence	1yp3C
20258949	*Arabidopsis thaliana*	88-349	386-512	57	426 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
20465645	*Arabidopsis thaliana*	93-354	391-517	51	426 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
22347636	*Oncidium hybrid cultivar*	86-348	385-511	56	425 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
29421114	*Phaseolus vulgaris*	96-356	393-519	53	425 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
32812836	*Triticum aestivum*	92-353	390-516	48	425 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
34900728	*Oryza sativa*	79-355	392-518	54	426 residues (86% of the query sequence) have been modeled with 100.0% confidence	1yp3C
41350643	*Fragaria x ananassa*	86-346	383-488	50	413 residues (81% of the query sequence) have been modeled with 100.0%	1yp3C

					confidence	
50582723	*Oryza sativa*	80-243	379-505	57	426 residues (85% of the query sequence) have been modeled with 100.0% confidence	1yp3C
89277026	*Triticum aestivum*	72-334	371-497	57	426 residues (87% of the query sequence) have been modeled with 100.0% confidence	1yp3C
162460455	*Zea mays*	74-336	373-499	57	426 residues (87% of the query sequence) have been modeled with 100.0% confidence	1yp3C
162463875	*Zea mays*	84-345	382-508	55	426 residues (83% of the query sequence) have been modeled with 100.0% confidence	1yp3C
189027076	*Zea mays*	86-347	384-510	49	426 residues (83% of the query sequence) have been modeled with 100.0% confidence	1yp3C
224062107	*Populus trichocarpa*	98-359	396-522	54	432 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
224100249	*Populus trichocarpa*	97-359	396-522	55	428 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
224103389	*Populus trichocarpa*	98-358	395-521	51	425 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
225428422	*Vitis vinifera*	98-358	395-521	53	425 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
225432564	*Vitis vinifera*	89-351	388-514	56	428 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
225458219	*Vitis vinifera (wine*	93-355	392-518	55	425 residues (81% of the query sequence)	1yp3C

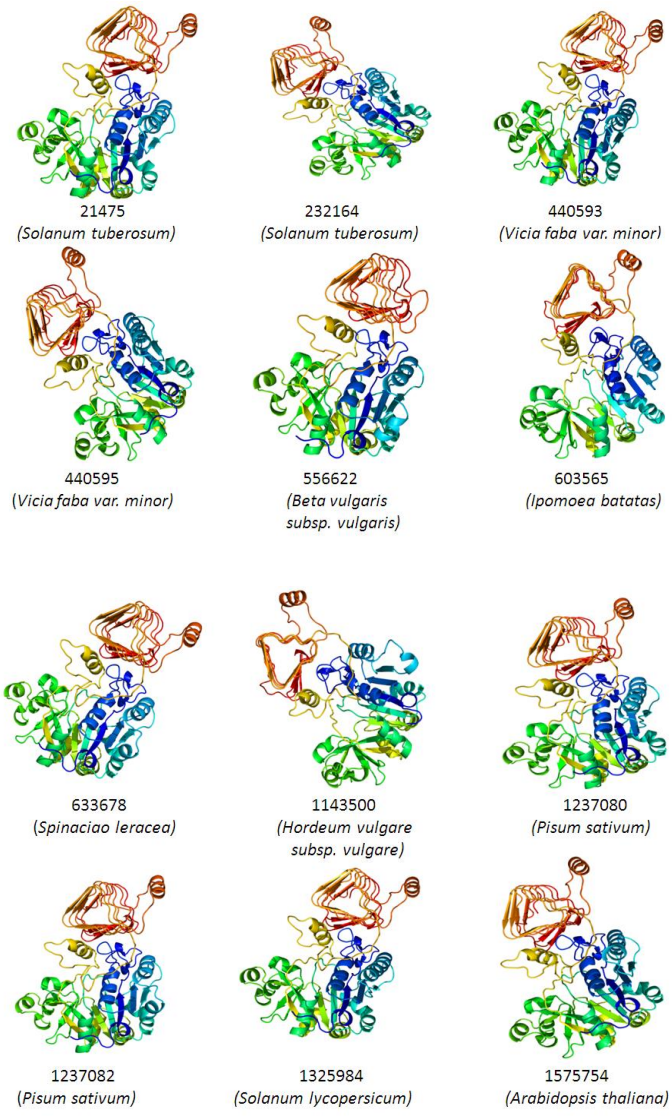
	grape)				have been modeled with 100.0% confidence	
242033053	*Sorghum bicolor*	76-338	375-501	56	431 residues (85% of the query sequence) have been modeled with 100.0% confidence	1yp3C
242053733	*Sorghum bicolor*	87-348	385-511	48	427 residues (83% of the query sequence) have been modeled with 100.0% confidence	1yp3C
242088961	*Sorghum bicolor*	89-350	387-513	49	427 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
255538708	*Ricinus communis*	93-354	391-517	54	426 residues (83% of the query sequence) have been modeled with 100.0% confidence	1yp3C
255543725	*Ricinus communis*	102-362	399-525	53	426 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
255548169	*Ricinus communis*	98-359	396-522	55	425 residues (80% of the query sequence) have been modeled with 100.0% confidence	1yp3C
255552303	*Ricinus communis*	102-364	401-527	55	425 residues (80% of the query sequence) have been modeled with 100.0% confidence	1yp3C
297797902	*Arabidopsis lyrata subsp. lyrata*	91-352	389-515	49	429 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
297812109	*Arabidopsis lyrata subsp. lyrata*	93-353	390-516	57	425 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
297821353	*Arabidopsis lyrata subsp. lyrata*	90-351	388-514	52	425 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C

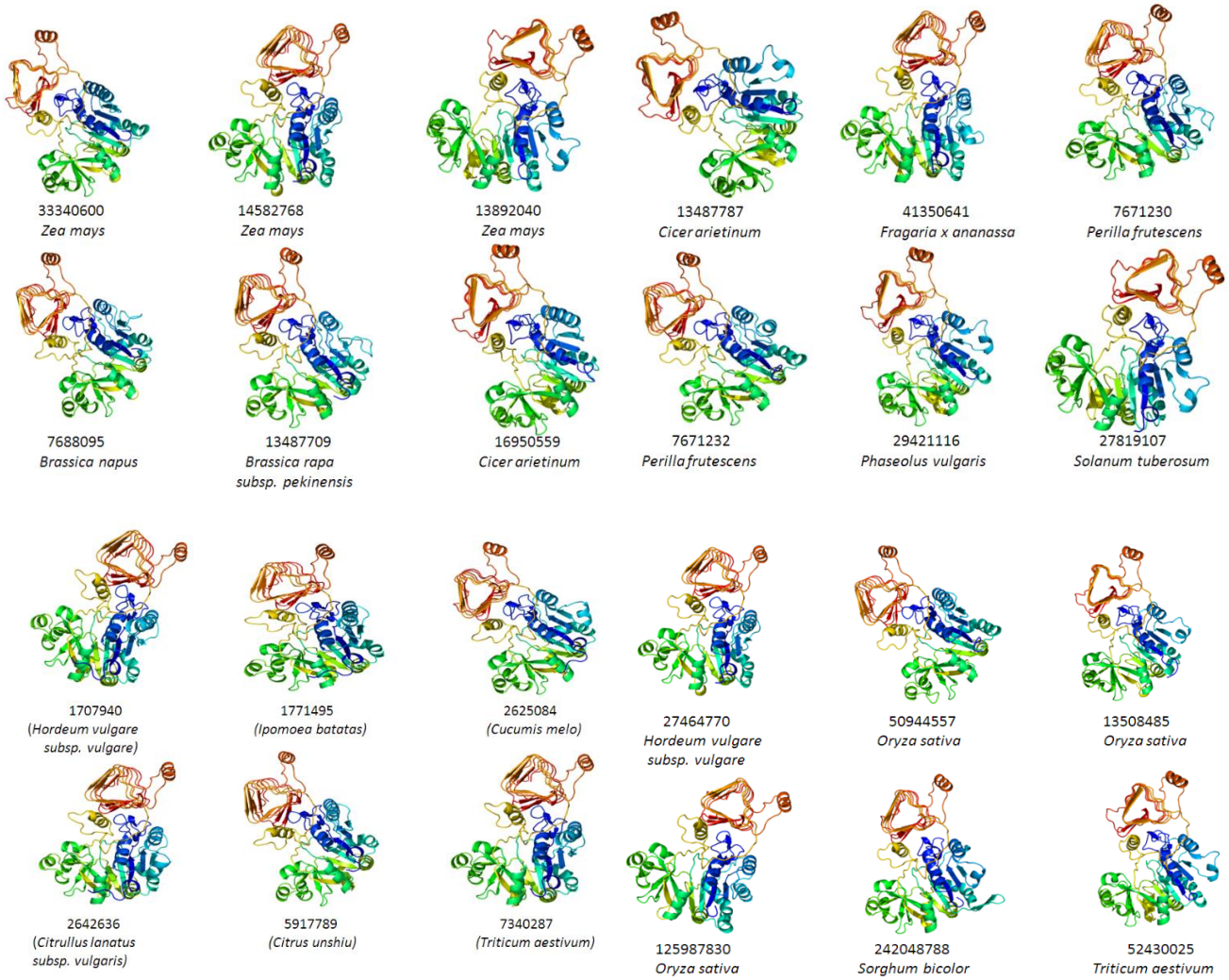
297845724	*Arabidopsis lyrata subsp. lyrata*	88-349	386-512	56	426 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
356508352	*Glycine max*	90-350	387-513	53	426 residues (83% of the query sequence) have been modeled with 100.0% confidence	1yp3
356509672	*Glycine max*	100-361	398-524	56	425 residues (80% of the query sequence) have been modeled with 100.0% confidence	1yp3C
356517038	*Glycine max*	90-350	387-513	53	426 residues (84% of the query sequence) have been modeled with 100.0% confidence	1yp3C
356518710	*Glycine max*	101-362	399-525	55	430 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
356538761	*Glycine max*	101-361	398-524	52	432 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
356545193	*Glycine max*	99-359	396-522	53	427 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
356553863	*Glycine max*	92-354	391-517	53	425 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
356562361	*Glycine max*	93-355	392-518	53	425 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
356563435	*Glycine max*	92-354	391-517	57	430 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
357116651	*Brachypodium distachyon*	73-335	372-498	56	429 residues (85% of the query sequence) have been modeled with 100.0%	1yp3C

357132398	*Brachypodium distachyon*	92-353	390-516	49	425 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
357473317	*Medicago truncatula*	99-359	396-522	50	425 residues (80% of the query sequence) have been modeled with 100.0% confidence	1yp3C
359479997	*Vitis vinifera*	84-345	382-508	57	424 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
359807441	*Glycine max*	89-351	388-514	57	428 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
566166858	*Populus trichocarpa*	97-357	394-520	52	425 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
566173136	*Populus trichocarpa*	98-359	396-522	55	425 residues (80% of the query sequence) have been modeled with 100.0% confidence	1yp3C
566203612	*Populus trichocarpa*	99-361	398-524	55	425 residues (80% of the query sequence) have been modeled with 100.0% confidence	1yp3C
571446804	*Glycine max*	99-360	397-523	55	429 residues (81% of the query sequence) have been modeled with 100.0% confidence	1yp3C
571559815	*Glycine max*	99-360	397-523	56	425 residues (80% of the query sequence) have been modeled with 100.0% confidence	1yp3C
721619253	*Brachypodium distachyon*	82-343	380-506	56	433 residues (85% of the query sequence) have been modeled with 100.0% confidence	1yp3C
806638577	*Zea mays*	88-349	386-512	50	427 residues (82% of the query sequence) have been modeled with 100.0%	1yp3C

					confidence	
922340593	*Medicago truncatula*	98-359	396-522	56	425 residues (80% of the query sequence) have been modeled with 100.0% confidence	1yp3C
922364294	*Medicago truncatula*	90-352	389-515	56	429 residues (82% of the query sequence) have been modeled with 100.0% confidence	1yp3C
922376662	*Medicago truncatula*	89-349	386-512	54	425 residues (82% of your sequence have been modeled with 100.0% confidence	1yp3C

Generated models were refined by a number of model refinement algorithms to have the most energetically stable conformation as stated above. Quality of the models were assessed by various model quality assessment programs. Out of different models built by various model building tools for a single query sequence was selected based on their model quality score. Best model for all the SS and LS AGPases are shown in Figures 5 and 6, respectively. The result of the model quality assessment programs are shown in Tables 7 and 8, respectively.




33340600
Zea mays
14582768
Zea mays
13892040
Zea mays
13487787
Cicer arietinum
41350641
Fragaria x ananassa
7671230
Perilla frutescens
7688095
Brassica napus
13487709
Brassica rapa
subsp. pekinensis
16950559
Cicer arietinum
7671232
Perilla frutescens
29421116
Phaseolus vulgaris
27819107
Solanum tuberosum
1707940
(Hordeum vulgare
subsp. vulgare)
1771495
(Ipomoea batatas)
2625084
(Cucumis melo)
27464770
Hordeum vulgare
subsp. vulgare
50944557
Oryza sativa
13508485
Oryza sativa
2642636
(Citrullus lanatus
subsp. vulgaris)
5917789
(Citrus unshiu)
7340287
(Triticum aestivum)
125987830
Oryza sativa
242048788
Sorghum bicolor
52430025
Triticum aestivum

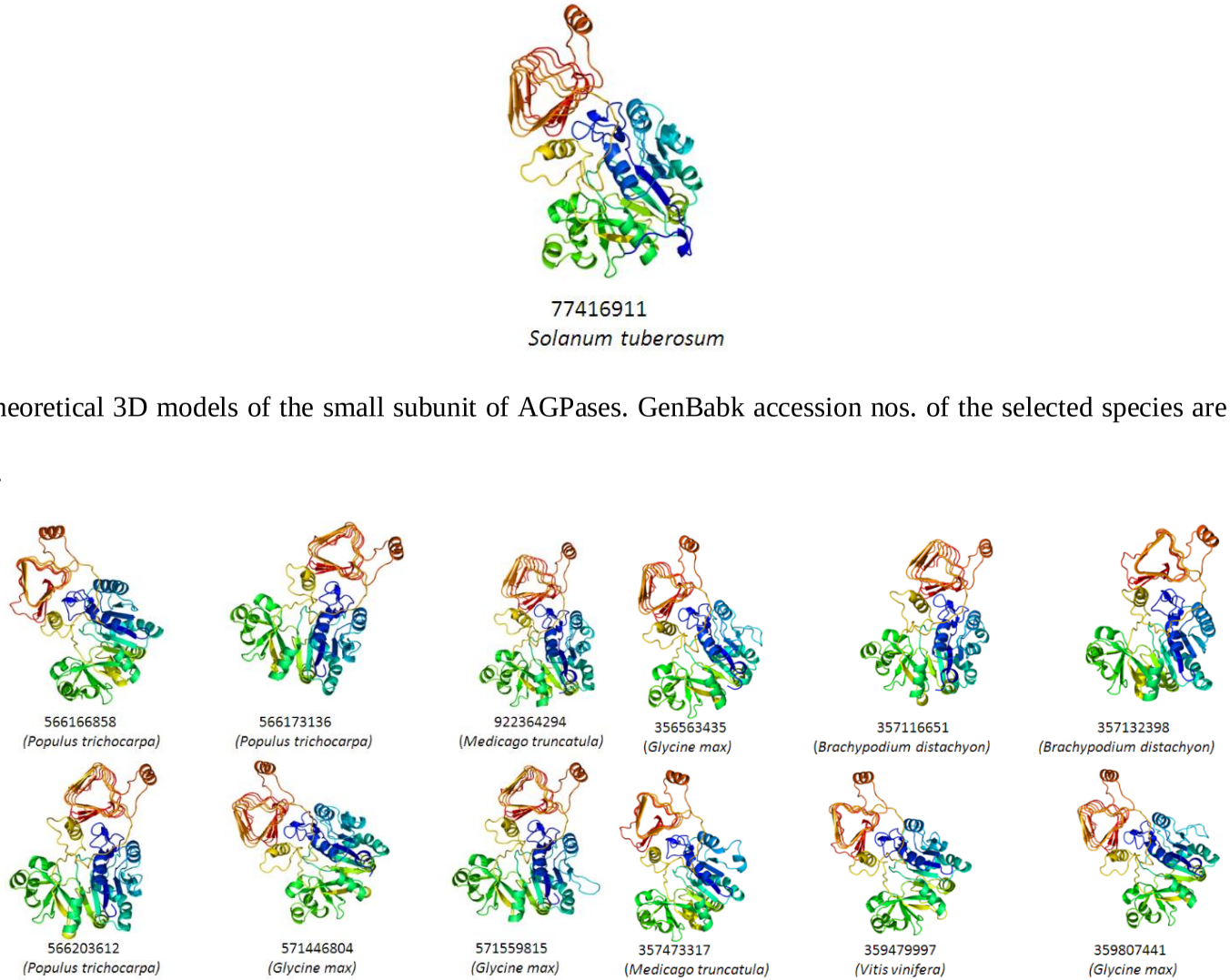
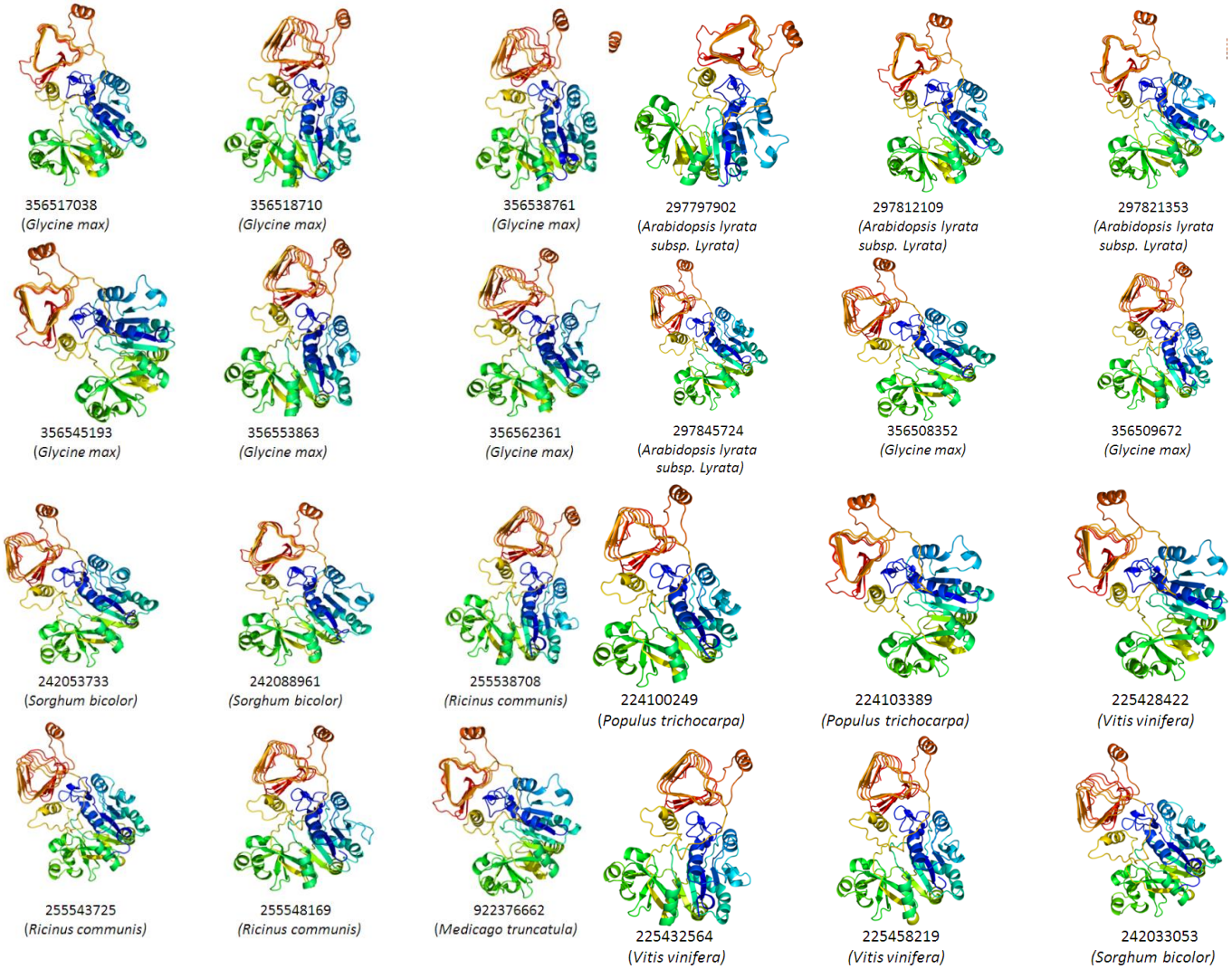
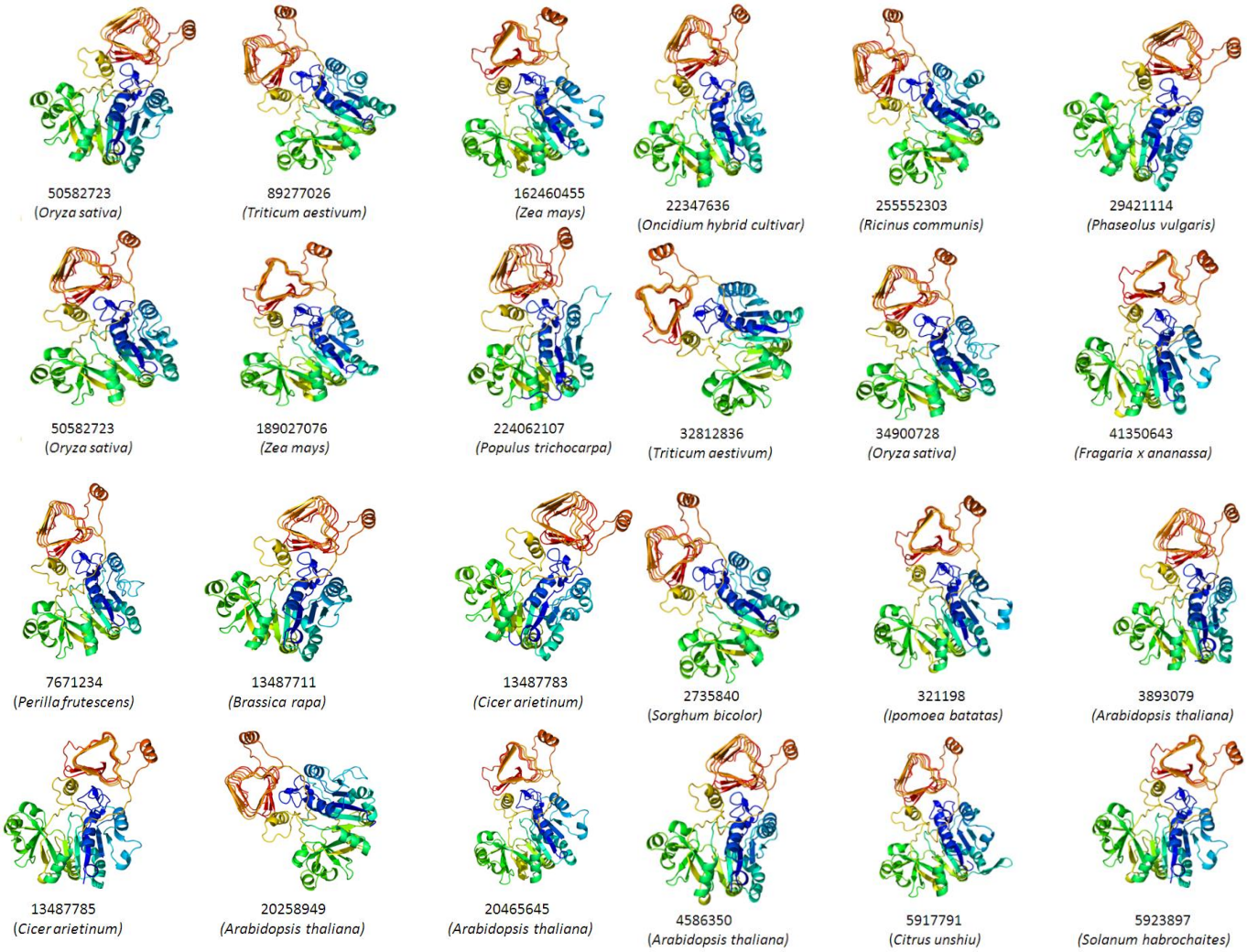
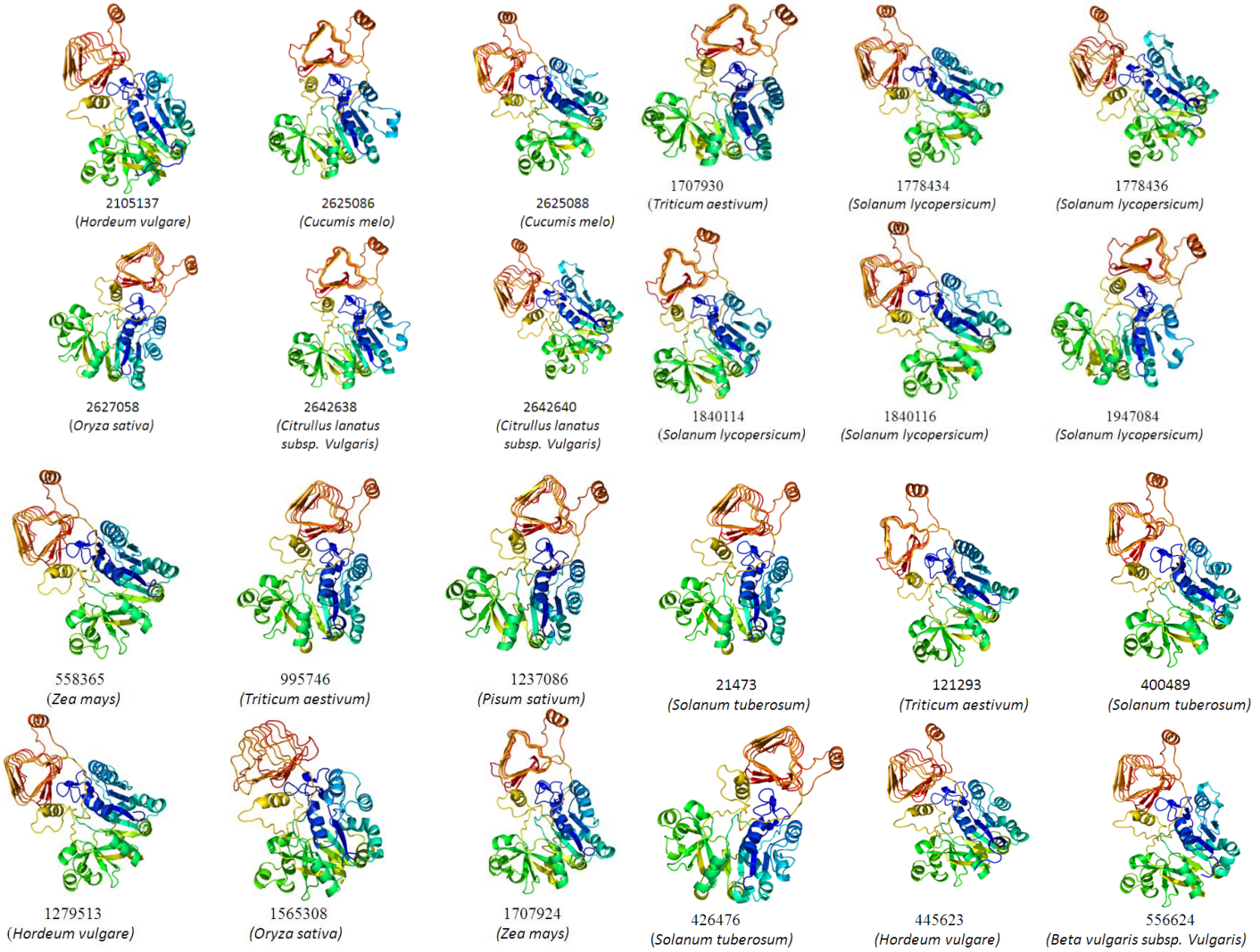


Figure 5. Theoretical 3D models of the small subunit of AGPases. GenBabk accession nos. of the selected species are also given in the structure.


356517038
(Glycine max)
356518710
(Glycine max)
356538761
(Glycine max)
297797902
(Arabidopsis lyrata subsp. Lyrata)
297812109
(Arabidopsis lyrata subsp. Lyrata)
297821353
(Arabidopsis lyrata subsp. Lyrata)
356545193
(Glycine max)
356553863
(Glycine max)
356562361
(Glycine max)
297845724
(Arabidopsis lyrata subsp. Lyrata)
356508352
(Glycine max)
356509672
(Glycine max)
242053733
(Sorghum bicolor)
242088961
(Sorghum bicolor)
255538708
(Ricinus communis)
224100249
(Populus trichocarpa)
224103389
(Populus trichocarpa)
225428422
(Vitis vinifera)
255543725
(Ricinus communis)
255548169
(Ricinus communis)
922376662
(Medicago truncatula)
225432564
(Vitis vinifera)
225458219
(Vitis vinifera)
242033053
(Sorghum bicolor)


50582723
(Oryza sativa)
89277026
(Triticum aestivum)
162460455
(Zea mays)
22347636
(Oncidium hybrid cultivar)
255552303
(Ricinus communis)
29421114
(Phaseolus vulgaris)
50582723
(Oryza sativa)
189027076
(Zea mays)
224062107
(Populus trichocarpa)
32812836
(Triticum aestivum)
34900728
(Oryza sativa)
41350643
(Fragaria x ananassa)
7671234
(Perilla frutescens)
13487711
(Brassica rapa)
13487783
(Cicer arietinum)
2735840
(Sorghum bicolor)
321198
(Ipomoea batatas)
3893079
(Arabidopsis thaliana)
13487785
(Cicer arietinum)
20258949
(Arabidopsis thaliana)
20465645
(Arabidopsis thaliana)
4586350
(Arabidopsis thaliana)
5917791
(Citrus unshiu)
5923897
(Solanum habrochaites)


2105137
(Hordeum vulgare)
2625086
(Cucumis melo)
2625088
(Cucumis melo)
1707930
(Triticum aestivum)
1778434
(Solanum lycopersicum)
1778436
(Solanum lycopersicum)
2627058
(Oryza sativa)
2642638
(Citrullus lanatus subsp. Vulgaris)
2642640
(Citrullus lanatus subsp. Vulgaris)
1840114
(Solanum lycopersicum)
1840116
(Solanum lycopersicum)
1947084
(Solanum lycopersicum)
558365
(Zea mays)
995746
(Triticum aestivum)
1237086
(Pisum sativum)
21473
(Solanum tuberosum)
121293
(Triticum aestivum)
400489
(Solanum tuberosum)
1279513
(Hordeum vulgare)
1565308
(Oryza sativa)
1707924
(Zea mays)
426476
(Solanum tuberosum)
445623
(Hordeum vulgare)
556624
(Beta vulgaris subsp. Vulgaris)

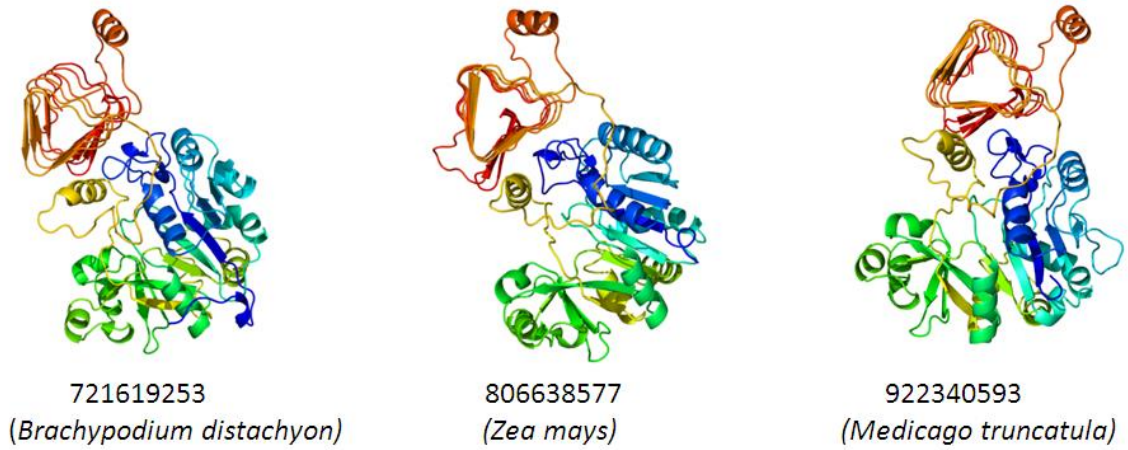


Figure 6. Theoretical 3D models of the LS of AGPases. GenBabk accession nos. of the selected species are also given in the structure.

Table 7: Model quality assessment scores of the best model of AGPase small subunit.

Accession no.	**Most favoured regions**	**Additional allowed regions**	**Generously allowed regions**	**Disallowed regions**	**Errat**	**Verify_3D**	**Prosa (Z Score)**	**MolProbity**
21475	87.8	11.9	0.0	0.3	88.018	97.51	-9.63	1.901
232164	87.8	11.9	0.0	0.3	88.018	97.51	-9.63	2.246
440593	88.8	10.9	0.0	0.3	89.862	97.74	-9.48	2.313
440595	89.1	10.7	0.0	0.3	89.862	99.55	-9.58	1.850

556622	86.1	12.9	0.3	0.8	89.100	93.02	-8.2	1.951
603565	87.7	11.2	0.5	0.5	80.668	94.85	-8.69	1.813
633678	90.1	9.4	0.3	0.3	79.439	97.71	-9.53	1.984
1143500	88.7	11.1	0.0	0.3	90.023	97.72	-9.53	1.984
1237080	89.1	10.6	0.3	0.0	72.748	98.64	-9.6	1.923
1237082	88.6	11.2	0.0	0.3	88.940	98.19	-9.26	1.866
1325984	88..5	11.2	0.3	0.0	81.585	91.53	-9.56	1.890
1575754	88.1	11.7	0.0	0.3	88.018	97.51	-9.6	1.958
1707940	90.7	8.5	0.8	0.0	81.250	97.50	-9.36	1.908
1771495	88.8	10.9	0.0	0.3	89.862	97.74	-9.42	1.956
2625084	87.8	11.9	0.0	0.3	88.018	97.51	-8.92	1.866
2642636	88.2	11.1	0.3	0.5	75.227	97.77	-9.38	1.861
5917789	88.6	11.2	0.0	0.3	88.940	98.19	-9.11	1.993
7340287	89.7	9.7	0.5	0.0	73.427	94.74	-9.42	1.861
7671230	89.1	10.7	0.0	0.3	89.862	99.55	-9.39	1.902
7671232	90.7	8.5	0.8	0.0	81.250	97.50	-8.92	1.997
7688095	88.9	10.3	0.5	0.3	73.843	97.95	-9.49	2.014
13487709	90.1	9.4	0.3	0.3	79.439	97.71	-9.41	2.001
13487787	90.0	9.2	0.8	0.0	77.907	97.03	-9.2	1.963
13508485	87.8	11.7	0.0	0.5	85.253	96.83	-9.25	1.859

13892040	87.8	12.0	0.0	0.3	85.714	95.70	-9.18	1.971
14582768	88.5	11.0	0.0	0.5	77.804	97.02	-9.39	2.095
16950559	88.9	10.6	0.3	0.3	80.831	93.65	-9.22	1.982
27464770	88.6	10.9	0.0	0.5	83.871	97.06	-9.13	2.160
27819107	88.5	11.0	0.0	0.5	77.804	97.02	-9.51	1.885
29421116	90.1	9.6	0.0	0.3	89.631	96.61	-9.39	1.993
33340600	88.5	10.8	0.8	0.0	80.886	97.25	-9.18	1.849
41350641	87.8	11.9	0.0	0.3	88.018	97.51	-9.57	1.862
50944557	89.1	10.6	0.3	0.0	72.748	98.64	-9.54	2.011
52430025	87.0	12.7	0.0	0.3	85.714	99.55	-9.27	1.886
77416911	88.9	10.3	0.5	0.3	73.843	97.95	-9.42	1.896
125987830	88..5	11.2	0.3	0.0	81.585	91.53	-9.48	1.982
242048788	88.9	10.3	0.3	0.5	82.217	96.60	-9.51	1.962

Table 8: Model quality assessment scores of the best model of AGPase large subunit.

Accession no.	**Most favoured regions**	**Additional allowed regions**	**Generously allowed regions**	**Disallowed regions**	**Errat**	**Verify_3D**	**Prosa (Z Score)**	**MolProbity**
21473	89.6	9.6	0.5	0.3	75.000	99.77	-9.18	2.161

121293	88.1	11.3	0.3	0.3	75.765	94.46	-9.02	2.303
400489	88.1	10.6	0.8	0.5	83.217	91.08	-9.13	2.222
426476	89.0	10.2	0.8	0.0	72.727	96.80	-9.02	2.569
445623	88.9	10.6	0.3	0.3	82.989	96.16	-9.8	2.336
556624	87.7	11.0	1.3	0.0	73.550	92.03	-9.07	2.468
558365	86.4	12.3	0.8	0.5	78.009	92.27	-9.23	2.447
995746	88.5	11.0	0.3	0.3	77.336	92.43	-9.54	2.201
1237086	90.2	8.7	0.8	0.3	80.607	98.17	-9.38	2.082
1279513	89.0	10.2	0.5	0.3	80.374	96.56	-9.71	2.160
1565308	84.8	12.1	1.8	1.3	76.706	87.76	-8.34	2.811
1707924	89.6	9.9	0.3	0.3	74.126	96.11	-9.81	2.088
1707930	88.5	11.0	0.3	0.3	77.336	92.43	-9.54	2.201
1778434	89.8	9.7	0.3	0.3	81.499	95.40	-9.1	2.266
1778436	88.8	10.2	0.8	0.3	74.483	98.19	-9.25	2.279
1840114	88.5	10.5	0.8	0.3	76.852	95.00	-9.17	2.283
1840116	90.0	8.9	1.0	0.0	79.487	95.19	-9.31	2.308
1947084	89.8	9.7	0.3	0.3	81.776	99.31	-9.39	2.030
2105137	89.2	9.8	0.5	0.5	81.193	97.30	-9.67	2.071
2625086	87.9	11.5	0.3	0.3	77.804	97.48	-8.6	2.099
2625088	87.1	12.1	0.3	0.5	76.402	94.72	-8.69	2.414

2627058	87.7	11.8	0.3	0.3	85.514	99.31	-9.85	2.151
2642638	88.7	10.0	1.0	0.3	83.645	94.72	-8.46	2.186
2642640	88.6	10.4	1.0	0.0	78.009	95.00	-9.52	2.172
2735840	88.3	10.6	0.8	0.3	80.930	94.06	-9.95	2.088
3211989	90.1	9.2	0.5	0.3	76.869	97.94	-9.31	2.126
3893079	89.4	10.1	0.0	0.5	80.324	95.45	-9.38	2.090
4586350	90.9	8.6	0.3	0.3	81.903	98.63	-9.23	2.087
5917791	89.4	9.4	0.8	0.5	83.102	97.95	-9.03	2.342
5923897	89.8	9.4	0.5	0.3	79.814	99.32	-9.16	2.152
7671234	86.5	11.2	1.6	0.8	78.505	94.04	-8.83	2.410
13487711	89.8	9.4	0.5	0.3	79.070	97.95	-9.41	2.111
13487783	89.2	10.0	0.5	0.3	82.488	95.48	-9.29	2.353
13487785	88.6	10.4	0.3	0.8	82.093	98.63	-9.46	2.294
20258949	89.2	10.0	0.8	0.0	75.581	96.35	-9.26	2.295
20465645	87.9	11.6	0.3	0.3	79.021	96.80	-9.08	2.205
22347636	89.6	9.7	0.8	0.0	78.089	98.63	-9.59	2.445
25552303	90.4	8.8	0.5	0.3	81.352	98.17	-9.61	2.283
29421114	88.4	10.5	0.5	0.5	81.075	96.79	-9.33	2.192
32812836	86.6	12.9	0.3	0.3	80.140	94.04	-9.39	2.125
34900728	90.1	9.2	0.5	0.3	76.224	95.88	-9.28	2.129

41350643	87.7	11.5	0.5	0.3	69.471	94.10	-9.55	1.908
50582723	89.6	9.6	0.8	0.0	81.481	99.32	-8.59	2.446
89277026	89.8	8.9	1.0	0.3	86.311	96.58	-9.64	2.425
162460455	88.9	10.1	1.0	0.0	80.460	96.84	-9.23	2.005
162463875	90.0	9.2	0.8	0.0	77.907	97.03	-9.46	2.190
189027076	89.6	9.9	0.3	0.3	74.126	96.11	-8.99	2.513
224062107	89.5	9.8	0.8	0.0	76.484	96.41	-9.81	2.088
224100249	89.8	9.4	0.8	0.0	79.167	98.18	-8.99	2.115
224103389	90	9.5	0.3	0.3	78.738	97.48	-8.63	2.431
225428422	90.0	9.7	0.0	0.3	82.243	92.66	-9.38	2.376
225432564	88.9	10.3	0.5	0.3	73.843	97.95	-8.85	2.024
225458219	89.3	9.7	0.5	0.5	78.322	94.74	-9.86	2.377
242033053	89.3	9.7	0.8	0.3	82.989	99.32	-8.91	2.442
242053733	87.0	12.2	0.5	0.3	80.233	93.84	-9.78	2.458
242088961	89.5	9.7	0.5	0.3	82.326	90.87	-9.94	2.234
255538708	90.3	8.9	0.5	0.3	82.751	96.57	-9.72	2.097
255543725	91.0	8.4	0.3	0.3	78.505	94.95	-8.83	2.144
255548169	88.5	10.8	0.8	0.0	80.886	97.25	-9.41	2.156
297797902	87.3	12.4	0.0	0.3	83.565	96.82	-8.99	2.171
297812109	90.1	9.4	0.3	0.3	79.439	97.71	-9.38	2.207

297821353	88.9	10.3	0.3	0.5	77.336	96.79	-9.56	2.096
297845724	89.2	10.3	0.5	0.0	73.256	98.86	-8.8	1.993
356508352	90.8	8.2	0.8	0.3	83.879	98.39	-9.19	2.455
356509672	89.5	9.7	0.8	0.0	75.524	92.91	-9.36	1.977
356517038	88.5	10.5	0.8	0.3	83.023	97.49	-8.9	2.357
356518710	88.8	10.1	0.8	0.3	69.585	97.06	-9.05	2.475
356538761	87.9	11.6	0.5	0.0	82.569	98.65	-9.5	2.222
356545193	89.7	9.7	0.5	0.0	81.860	98.86	-9.24	1.969
356553863	88.3	10.9	0.5	0.3	82.517	98.86	-9.25	2.174
356562361	90.6	8.9	0.3	0.3	80.00	94.28	-9.11	2.250
356563435	88.9	10.6	0.5	0.0	80.645	97.29	-8.75	2.352
357116651	88.9	10.6	0.3	0.3	80.831	93.65	-9.35	2.300
357132398	88.7	10.7	0.3	0.3	82.944	89.45	-9.65	2.154
357473317	89.7	9.8	0.3	0.3	78.505	97.94	-9.35	1.998
359479997	88.5	11.0	0.0	0.5	77.804	97.02	-8.92	2.357
359807441	90.7	8.5	0.8	0.0	81.250	97.50	-9.01	2.267
566166858	91.1	8.1	0.5	0.3	81.075	97.25	-9.48	2.133
566173136	89.7	9.5	0.8	0.0	80.886	99.08	-9.22	2.399
566203612	89.7	9.7	0.5	0.0	73.427	94.74	-9.02	2.389
571446804	89.1	10.6	0.3	0.0	72.748	98.64	-8.8	2.445

571559815	87.1	12.1	0.8	0.0	82.051	97.48	-8.97	2.269
721619253	88.2	11.1	0.3	0.5	75.227	97.77	-9.29	2.408
806638577	85.9	12.8	1.0	0.3	80.698	89.50	-9.57	2.313
922340593	88..5	11.2	0.3	0.0	81.585	91.53	-8.57	2.217
922364294	88.9	10.3	0.3	0.5	82.217	96.60	-9.26	2.237
922376662	90.5	8.7	0.3	0.5	83.178	97.25	-9.36	1.937

After having the best models for each of the SS and LS of AGPase, detailed structural analysis such as number of sheets, strands, helices their percentages etc. are calculated using ProFunc server. Details about the structural statistics of both SS and LS of AGPases are shown in Tables 9 and 10, respectively.

Table 9: Secondary structure statistics of AGPase small subunit.

Accession no.	Total residues	Sheets	beta alpha beta unit	beta hairpins	Psi loop	Beta bulges	Strands	Helices	Helix-helix interacs	Beta turns	Gamma turns	Strand	Alpha helix	3_{10} helix	Other
1143500	439	5	1	4	1	10	25	17	9	35	5	138 (31.4%)	106 (24.1%)	13 (3.0%)	182 (41.5%)
27464770	442	5	2	4	1	10	24	15	9	39	4	136 (30.8%)	107 (24.2%)	6 (1.4%)	193 (43.7%)
13508485	442	5	1	4	1	10	25	16	10	34	5	134 (30.3%)	108 (24.4%)	10 (2.3%)	190 (43.0%)
52430025	442	5	1	4	1	10	25	16	8	32	4	138 (31.2%)	111 (25.1%)	13 (2.9%)	180 (40.7%)
1575754	442	5	2	4	1	10	24	16	8	33	5	136 (30.8%)	109 (24.7%)	13 (2.9%)	184 (41.6%)
556622	430	5	1	4	1	10	22	15	9	37	6	126 (29.3%)	97 (22.6%)	10 (2.3%)	197 (45.8%)
603565	427	6	1	4	1	9	24	13	7	38	6	117 (27.4%)	104 (24.4%)	7 (1.6%)	199 (46.6%)
29421116	442	5	1	4	1	10	25	16	8	32	5	134 (30.3%)	108 (24.4%)	13 (2.9%)	187 (42.3%)
1237082	442	5	1	4	1	11	25	16	11	34	7	136 (30.8%)	108 (24.4%)	10 (2.3%)	188 (42.5%)
21475	442	5	2	4	1	10	25	15	8	35	5	130 (29.4%)	109 (24.7%)	10 (2.3%)	193 (43.7%)
232164	442	5	2	4	1	10	24	15	8	35	5	130 (29.4%)	109 (24.7%)	10 (2.3%)	193 (43.7%)
440593	442	5	1	4	1	11	25	15	8	34	5	136 (30.8%)	109 (24.7%)	10 (2.3%)	187 (42.3%)
440595	442	5	1	4	1	10	25	15	8	35	4	138 (31.2%)	108 (24.4%)	10 (2.3%)	186 (42.1%)
1707940	439	6	1	4	1	10	25	14	9	39	1	127 (29.1%)	102 (23.3%)	7 (1.6%)	201 (46.0%)
50944557	442	5	1	4	1	11	25	15	8	36	3	134	107	10	185

												(30.7%)	(24.5%)	(2.3%)	(42.4%)
125987830	442	5	1	4	1	9	25	16	10	36	4	127 (29.1%)	109 (24.9%)	9 (2.1%)	192 (43.9%)
242048788	442	5	1	4	1	9	25	15	9	36	4	126 (28.8%)	100 (22.8%)	10 (2.3%)	202 (46.1%)
7340287	442	5	1	4	1	11	25	17	9	33	5	132 (30.1%)	100 (22.8%)	16 (3.6%)	191 (43.5%)
33340600	430	5	1	4	1	10	25	12	7	42	5	132 (30.3%)	105 (24.1%)	4 (0.9%)	195 (44.7%)
14582768	427	5	1	4	1	10	25	14	8	37	3	134 (30.7%)	107 (24.5%)	7 (1.6%)	188 (43.1%)
7688095	439	5	1	4	1	10	25	13	7	42	5	134 (30.7%)	100 (22.9%)	7 (1.6%)	195 (44.7%)
13487709	442	5	0	4	1	9	25	15	12	38	4	121 (27.7%)	111 (25.4%)	7 (1.6%)	198 (45.3%)
16950559	442	5	1	4	1	10	25	17	12	35	5	134 (30.5%)	109 (24.8%)	10 (2.3%)	187 (42.5%)
13487787	442	5	1	4	1	10	25	16	9	40	3	134 (30.2%)	98 (22.1%)	13 (2.9%)	198 (44.7%)
2642636	442	5	1	4	1	10	25	15	8	37	3	125 (28.6%)	98 (22.4%)	10 (2.3%)	204 (46.7%)
5917789	430	5	1	4	1	9	25	17	11	32	2	128 (28.7%)	105 (23.5%)	13 (2.9%)	200 (44.8%)
2625084	427	6	1	4	1	10	25	13	8	44	3	128 (29.3%)	103 (23.6%)	4 (0.9%)	202 (46.2%)
41350641	442	5	2	4	1	10	25	15	8	35	5	130 (29.4%)	109 (24.7%)	10 (2.3%)	193 (43.7%)
1771495	442	5	2	4	1	10	24	15	8	35	5	130 (29.4%)	109 (24.7%)	10 (2.3%)	193 (43.7%)
7671230	442	5	1	4	1	11	25	15	8	34	5	136 (30.8%)	109 (24.7%)	10 (2.3%)	187 (42.3%)
7671232	430	5	1	4	1	10	25	15	8	35	4	138 (31.2%)	108 (24.4%)	10 (2.3%)	186 (42.1%)
1237080	427	6	1	4	1	10	25	14	9	39	1	127 (29.1%)	102 (23.3%)	7 (1.6%)	201 (46.0%)

1325984	442	5	2	4	1	10	25	15	8	35	5	130 (29.4%)	109 (24.7%)	10 (2.3%)	193 (43.7%)
27819107	437	5	2	4	1	10	24	15	8	35	5	130 (29.4%)	109 (24.7%)	10 (2.3%)	193 (43.7%)
77416911	427	5	1	4	1	11	25	15	8	34	5	136 (30.8%)	109 (24.7%)	10 (2.3%)	187 (42.3%)
633678	440	5	1	4	1	10	25	15	8	35	4	138 (31.2%)	108 (24.4%)	10 (2.3%)	186 (42.1%)
13892040	442	5	1	4	1	11	25	14	8	35	4	138 (31.2%)	108 (24.4%)	7 (1.6%)	189 (42.8%)

Table 10: Secondary structure statistics of AGPase large subunit

Accession no.	**Total residues**	**Sheets**	**Beta alpha beta unit**	**Beta hairpins**	**Psi loop**	**Beta bulges**	**Strands**	**Helices**	**Helix-helix interacs**	**Beta turns**	**Gamma turs**	**Strand**	**Alpha helix**	**3-10 helix**	**Other**
5923897	439	5	1	4	1	10	25	16	9	38	3	135 (30.8%)	102 (23.2%)	10 (2.3%)	192 (43.7%)
1237086	436	5	0	4	1	10	24	14	10	41	4	123 (28.2%)	108 (24.8%)	4 (0.9%)	201 (46.1%)
3211989	436	5	1	4	1	10	25	13	7	35	3	133 (30.5%)	110 (25.2%)	7 (1.6%)	186 (42.7%)
2642638	436	5	1	4	1	10	25	14	8	40	2	134 (30.7%)	99 (22.7%)	7 (1.6%)	196 (45.0%)
29421114	436	5	0	4	1	10	24	14	8	43	7	124 (28.4%)	100 (22.9%)	7 (1.6%)	205 (47.0%)
224103389	436	5	1	4	1	11	25	14	8	38	2	134 (30.7%)	107 (24.5%)	7 (1.6%)	188 (43.1%)
225428422	436	5	1	4	1	10	25	16	10	36	4	133 (30.5%)	107 (24.5%)	10 (2.3%)	186 (42.7%)
356517038	438	5	1	4	1	10	25	15	8	35	5	134 (30.6%)	102 (23.3%)	10 (2.3%)	192 (43.8%)

356538761	444	5	1	4	1	10	24	15	9	39	5	124 (27.9%)	106 (23.9%)	7 (1.6%)	207 (46.6%)
297821353	436	5	1	1	1	10	25	15	9	38	4	128 (29.4%)	106 (24.3%)	7 (1.6%)	195 (44.7%)
356545193	438	5	1	4	1	12	25	16	12	38	4	135 (30.8%)	109 (24.9%)	7 (1.6%)	187 (42.7%)
297845724	438	5	1	4	1	10	25	15	8	37	3	126 (28.8%)	104 (23.7%)	10 (2.3%)	198 (45.2%)
13487785	438	5	1	4	1	10	25	16	10	35	5	129 (29.5%)	111 (25.3%)	10 (2.3%)	188 (42.9%)
41350643	424	5	1	4	1	11	24	13	8	39	5	134 (31.6%)	97 (22.9%)	4 (0.9%)	189 (44.6%)
566166858	436	5	1	4	1	10	25	14	7	35	6	130 (29.8%)	109 (25.0%)	10 (2.3%)	187 (42.9%)
20465645	437	5	1	4	1	10	25	15	9	38	3	134 (30.7%)	106 (24.3%)	10 (2.3%)	187 (42.8%)
1840114	440	5	1	4	1	11	25	16	10	38	2	131 (29.8%)	102 (23.2%)	10 (2.3%)	197 (44.8%)
356562361	437	6	1	4	1	10	25	14	9	39	1	127 (29.1%)	102 (23.3%)	7 (1.6%)	201 (46.0%)
357473317	436	5	1	4	1	11	25	15	8	36	3	134 (30.7%)	107 (24.5%)	10 (2.3%)	185 (42.4%)
356509672	437	5	1	4	1	9	25	16	10	36	4	127 (29.1%)	109 (24.9%)	9 (2.1%)	192 (43.9%)
20258949	438	5	1	4	1	9	25	15	9	36	4	126 (28.8%)	100 (22.8%)	10 (2.3%)	202 (46.1%)
556624	439	5	1	4	1	11	25	17	9	33	5	132 (30.1%)	100 (22.8%)	16 (3.6%)	191 (43.5%)
255543725	436	5	1	4	1	10	25	12	7	42	5	132 (30.3%)	105 (24.1%)	4 (0.9%)	195 (44.7%)
922376662	436	5	1	4	1	10	25	14	8	37	3	134 (30.7%)	107 (24.5%)	7 (1.6%)	188 (43.1%)
356508352	436	5	1	4	1	10	25	13	7	42	5	134 (30.7%)	100 (22.9%)	7 (1.6%)	195 (44.7%)
400489	437	5	0	4	1	9	25	15	12	38	4	121	111	7	198

												(27.7%)	(25.4%)	(1.6%)	(45.3%)
297797902	440	5	1	4	1	10	25	17	12	35	5	134 (30.5%)	109 (24.8%)	10 (2.3%)	187 (42.5%)
1778436	443	5	1	4	1	10	25	16	9	40	3	134 (30.2%)	98 (22.1%)	13 (2.9%)	198 (44.7%)
255538708	437	5	1	4	1	10	25	15	8	37	3	125 (28.6%)	98 (22.4%)	10 (2.3%)	204 (46.7%)
224062107	446	5	1	4	1	9	25	17	11	32	2	128 (28.7%)	105 (23.5%)	13 (2.9%)	200 (44.8%)
356553863	437	6	1	4	1	10	25	13	8	44	3	128 (29.3%)	103 (23.6%)	4 (0.9%)	202 (46.2%)
356518710	442	5	1	4	1	9	25	16	10	35	4	127 (28.7%)	108 (24.4%)	9 (2.0%)	198 (44.8%)
225458219	437	5	1	4	1	9	25	17	8	36	3	129 (29.5%)	98 (22.4%)	15 (3.4%)	195 (44.6%)
2625088	436	5	1	4	1	8	26	15	9	44	2	121 (27.8%)	94 (21.6%)	9 (2.1%)	212 (48.6%)
1778434	435	5	1	4	1	11	25	17	11	35	2	129 (29.7%)	111 (25.5%)	10 (2.3%)	185 (42.5%)
558365	440	5	2	3	1	10	23	13	6	41	3	129 (29.3%)	99 (22.5%)	6 (1.4%)	206 (46.8%)
1279513	436	5	1	4	1	11	25	14	7	34	5	134 (30.7%)	109 (25.0%)	10 (2.3%)	183 (42.0%)
2625086	436	5	1	4	1	10	25	15	8	38	3	134 (30.7%)	97 (22.2%)	10 (2.3%)	195 (44.7%)
3893079	440	5	1	4	1	11	25	15	8	35	4	132 (30.0%)	109 (24.8%)	10 (2.3%)	189 (43.0%)
1707924	437	5	2	4	1	12	24	14	7	35	3	133 (30.4%)	116 (26.5%)	5 (1.1%)	183 (41.9%)
2735840	438	5	0	4	1	10	24	14	9	38	4	124 (28.3%)	110 (25.1%)	7 (1.6%)	197 (45.0%)
189027076	437	5	2	4	1	12	24	14	7	35	3	133 (30.4%)	116 (26.5%)	5 (1.1%)	183 (41.9%)
242088961	438	5	1	4	1	10	25	14	10	37	4	134 (30.6%)	100 (22.8%)	7 (1.6%)	197 (45.0%)

21473	440	5	1	4	1	10	25	14	8	39	3	133 (30.2%)	103 (23.4%)	7 (1.6%)	197 (44.8%)
1565308	433	7	0	5	1	8	26	14	7	47	4	106 (24.5%)	99 (22.9%)	8 (1.8%)	220 (50.8%)
7671234	436	5	1	4	1	11	25	14	7	42	7	129 (29.6%)	99 (22.7%)	10 (2.3%)	198 (45.4%)
5917791	440	5	2	0	4	9	25	15	7	34	3	137 (31.1%)	115 (26.1%)	7 (1.6%)	181 (41.1%)
426476	437	5	1	4	1	10	25	16	9	38	3	122 (27.9%)	103 (23.6%)	9 (2.1%)	203 (46.5%)
1840116	437	5	1	4	1	10	25	14	7	35	4	127 (29.1%)	101 (23.1%)	10 (2.3%)	199 (45.5%)
2105137	444	5	1	4	1	9	25	18	9	31	2	134 (30.2%)	107 (24.1%)	17 (3.8%)	186 (41.9%)
4586350	439	5	1	4	1	10	25	15	9	36	4	131 (29.8%)	105 (23.9%)	10 (2.3%)	193 (44.0%)
13487711	438	5	1	4	1	10	24	15	8	39	3	133 (30.4%)	102 (23.3%)	11 (2.5%)	192 (43.8%)
13487783	442	5	1	4	1	10	25	15	9	39	2	130 (29.4%)	108 (24.4%)	7 (1.6%)	197 (44.6%)
22347636	437	5	1	4	1	10	25	15	8	35	4	125 (28.6%)	108 (24.7%)	9 (2.1%)	195 (44.6%)
25552303	437	5	0	4	1	9	24	16	8	32	3	124 (28.4%)	110 (25.2%)	13 (3.0%)	190 (43.5%)
50582723	440	5	1	4	1	9	25	16	11	37	3	127 (28.9%)	101 (23.0%)	9 (2.0%)	203 (46.1%)
89277026	439	5	1	4	1	10	25	16	10	34	3	130 (29.6%)	101 (23.0%)	10 (2.3%)	198 (45.1%)
162460455	443	5	1	4	1	11	25	17	12	37	1	131 (29.6%)	114 (25.7%)	9 (2.0%)	189 (42.7%)
224100249	440	5	1	4	1	10	25	16	12	36	3	126 (28.6%)	109 (24.8%)	7 (1.6%)	198 (45.0%)
225432564	440	5	1	4	1	10	25	15	10	34	4	128 (29.1%)	110 (25.0%)	7 (1.6%)	195 (44.3%)
24203305	443	5	1	4	1	10	25	18	9	36	3	128	106	14	195

3												(28.9%)	(23.9%)	(3.2%)	(44.0%)
297812109	436	5	1	4	1	10	25	13	7	36	6	131 (30.0%)	106 (24.3%)	7 (1.6%)	192 (44.0%)
356563435	442	5	1	4	1	9	25	14	11	41	2	124 (28.1%)	106 (24.0%)	4 (0.9%)	208 (47.1%)
357116651	441	5	1	4	1	10	25	16	12	43	2	128 (29.0%)	102 (23.1%)	7 (1.6%)	204 (46.3%)
359807441	440	5	1	4	1	10	25	14	8	37	3	128 (29.1%)	103 (23.4%)	7 (1.6%)	202 (45.9%)
922364294	441	5	1	4	1	10	25	15	9	37	2	130 (29.5%)	108 (24.5%)	7 (1.6%)	196 (44.4%)
34900728	437	5	1	4	1	11	25	14	8	36	2	134 (30.7%)	104 (23.8%)	7 (1.6%)	192 (43.9%)
162463875	438	5	1	4	1	10	25	17	8	37	2	127 (29.0%)	100 (22.8%)	16 (3.7%)	195 (44.5%)
255548169	437	5	2	4	1	10	24	13	9	40	4	123 (28.1%)	99 (22.7%)	4 (0.9%)	211 (48.3%)
359479997	436	5	1	4	1	9	25	17	8	34	2	122 (28.0%)	99 (22.7%)	15 (3.4%)	200 (45.9%)
566203612	437	5	1	4	1	11	25	16	8	35	3	130 (29.7%)	99 (22.7%)	12 (2.7%)	196 (44.9%)
571446804	441	5	1	4	1	9	25	16	8	38	3	128 (29.0%)	105 (23.8%)	10 (2.3%)	198 (44.9%)
571559815	437	6	1	5	1	10	27	14	7	38	6	132 (30.2%)	97 (22.2%)	10 (2.3%)	198 (45.3%)
721619253	448	5	1	4	1	9	25	14	8	40	3	123 (27.5%)	100 (22.3%)	7 (1.6%)	218 (48.7%)
922340593	437	5	1	4	1	10	25	13	8	39	3	126 (28.8%)	101 (23.1%)	4 (0.9%)	206 (47.1%)
121293	433	5	1	4	1	9	25	14	7	37	2	124 (28.6%)	99 (22.9%)	10 (2.3%)	200 (46.2%)
445623	443	5	0	4	1	10	24	16	9	33	3	124 (28.0%)	115 (26.0%)	10 (2.3%)	194 (43.8%)
1707930	436	5	1	4	1	11	25	14	7	34	2	133 (30.5%)	106 (24.3%)	10 (2.3%)	187 (42.9%)

32812836	436	5	1	4	1	11	25	15	7	35	3	134 (30.7%)	100 (22.9%)	13 (3.0%)	189 (43.3%)
2627058	436	5	0	4	1	10	24	14	9	35	4	124 (28.4%)	103 (23.6%)	7 (1.6%)	202 (46.3%)
357132398	436	5	1	4	1	11	25	13	7	37	4	134 (30.7%)	105 (24.1%)	6 (1.4%)	191 (43.8%)
242053733	438	5	0	4	1	11	24	13	9	40	5	126 (28.8%)	109 (24.9%)	4 (0.9%)	199 (45.4%)
995746	436	5	1	4	1	11	25	14	7	34	2	133 (30.5%)	106 (24.3%)	10 (2.3%)	187 (42.9%)
806638577	438	5	1	4	1	11	25	14	7	40	4	134 (30.6%)	99 (22.6%)	9 (2.1%)	196 (44.7%)
1947084	436	5	1	4	1	10	25	15	8	34	4	131 (30.0%)	106 (24.3%)	10 (2.3%)	189 (43.3%)
2642640	440	5	1	4	1	11	25	15	8	40	2	134 (30.5%)	102 (23.2%)	11 (2.5%)	193 (43.9%)
566173136	437	5	1	4	1	9	25	17	8	32	3	126 (28.8%)	108 (24.7%)	15 (3.4%)	188 (43.0%)

Objective 2: Hetero-tetrameric assembly of AGPase subunits through docking

For completion of the Objective 2, predicted models for both SS and LS of AGPase were used to assemble the complete hetero-tetrameric structure of AGPase. Following steps were used to accomplish the objective.

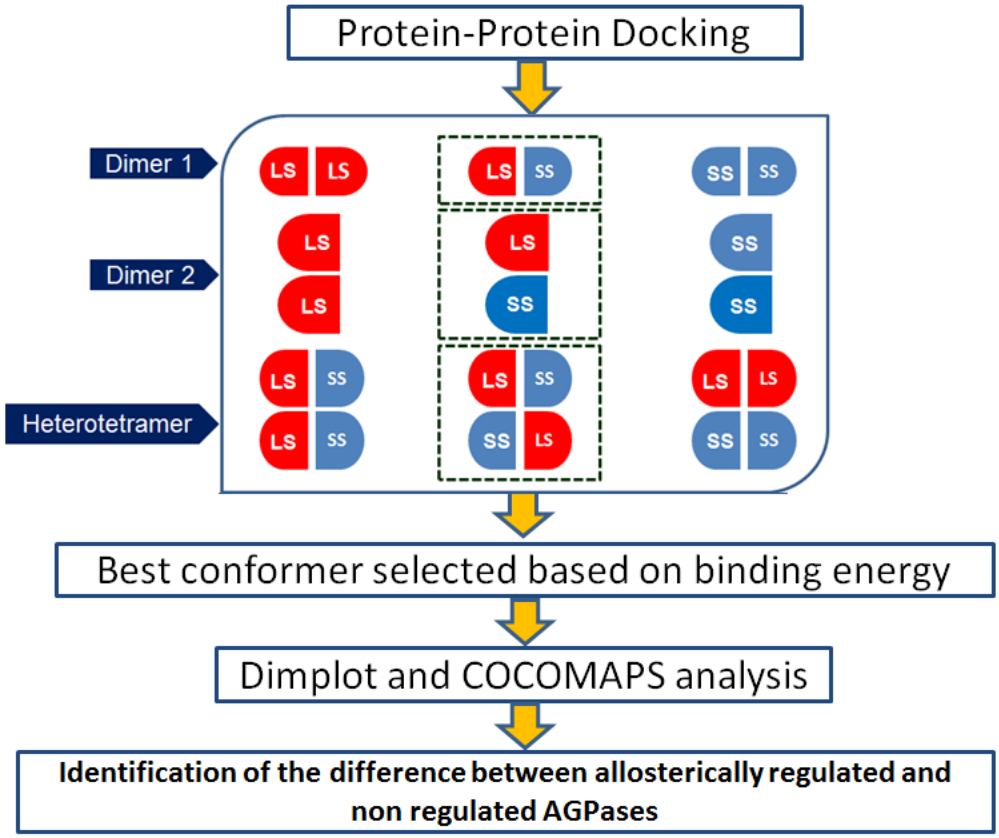


Figure 7: Step by step workflow of Objective 2.

Different combinations of SS and LS of AGPases were docked to identify the most stable initial SS/SL/LL homo or heterodimer. Respective SS and LS of AGPase from selected cereal crops were submitted to the HADDOCK server 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. Several dimers were formed by the HADDOCK server and the dimer orientation with the best HADDOCK score was selected and was again fed to the RosettaDock server (http://rosettaserver.graylab.jhu.edu/) to refine the initial orientation obtained from HADDOCK. The best conformation and lowest energy score PDB files were taken for further analysis. Bond geometry idealization and removal of unfavourable non-bonded contacts from the lowest energy PDB files were energy minimized in Swiss-Pdb Viewer (http://spdbv.vital-it.ch) using GROMOS96 force field. It was found that side by side and up and down interactions between the LS and SS are the most energetically stable conformations. Graphical representation of most stable conformation of the hetero-dimersis shown in Figure 8.

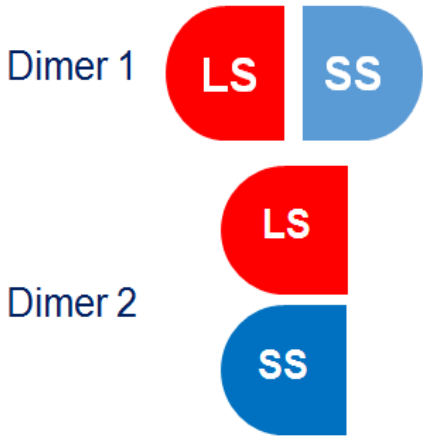


Figure 8: Graphical representation of the most stable hetero-dimer conformation of AGPase.

Later these two conformations were again submitted to HADDOCK server to form the heterotetrameric assembly and similar approach as stated above was conceived to attain the most energetically stable heterotetrameric assembly. Graphical representation of most stable conformation is shown in Figure 9.

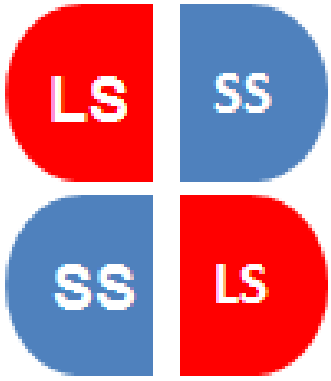


Figure 9: Graphical representation of the most stable heterotetramer conformers of AGPase.

The residues involved in subunit-subunit interaction are inspected using COCOMAPS server and to have a more confidence about the interactions, DIMPLOT tool was also used to plot the interacting residues. COCOMAPS marked the interacting residues within subunits if there is a decrease of >1 $Å^2$ accessible surface area (ASA) upon complex formation. Dimplot, which is a part of the Ligplot suite, studies the subunit-subunit interface and plots the hydrophobic and H-bonds. The results of both the tools were similar and the dimer and heterotetramer structure along with the interacting residues are shown in Figures 10-21.

(a)

(b)

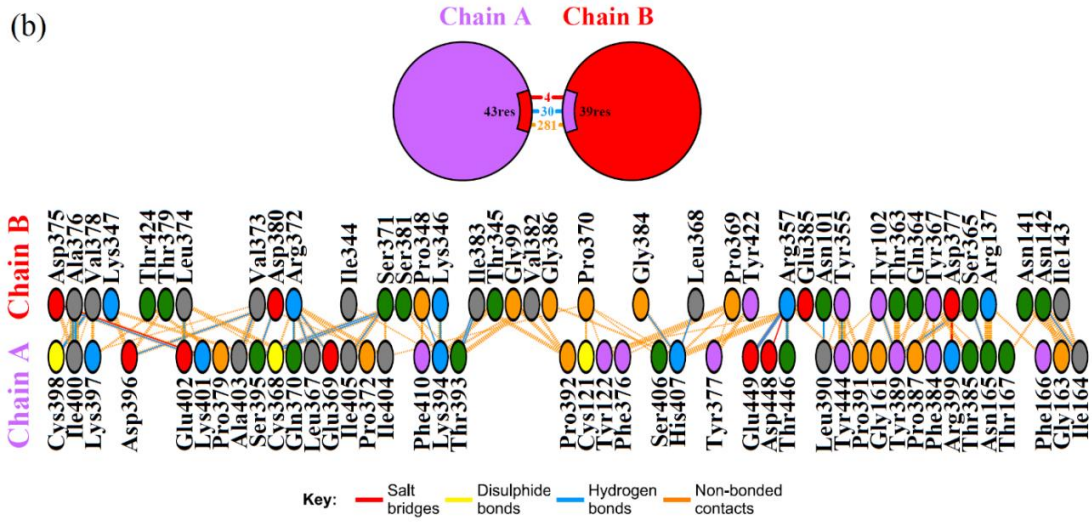


Figure 10(a): Side by side conformation between LS and SS of *Hordeum vulgare* to form dimer1. LS is coloured in green and SS is coloured in red. 10(b): Side by side interacting residues between LS and SS of *Hordeum vulgare* to form dimer1.

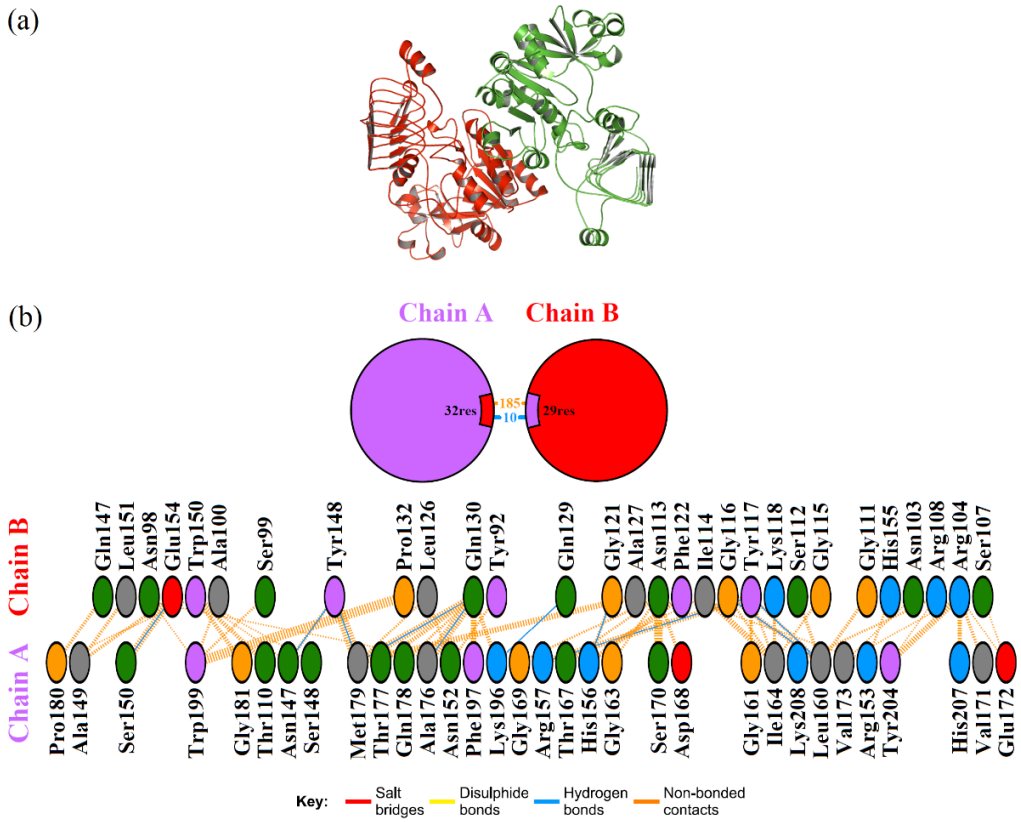


Figure 11(a): Up and down conformation between LS and SS of *Hordeum vulgare* to form dimer2. LS is coloured in green and SS is coloured in red. 11(b): Up and down interacting residues between LS and SS of *Hordeum vulgare* to form dimer2.

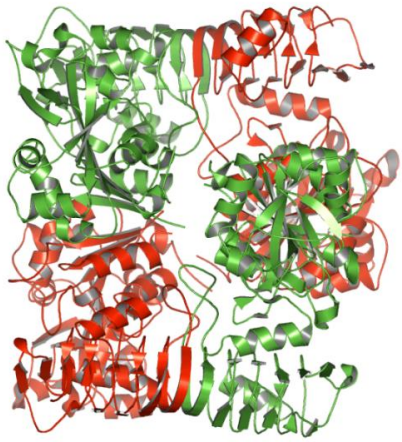


FIGURE 12: Heterotetrameric assembly of *Hordeum vulgare*. LS is coloured in green and SS is coloured in red.

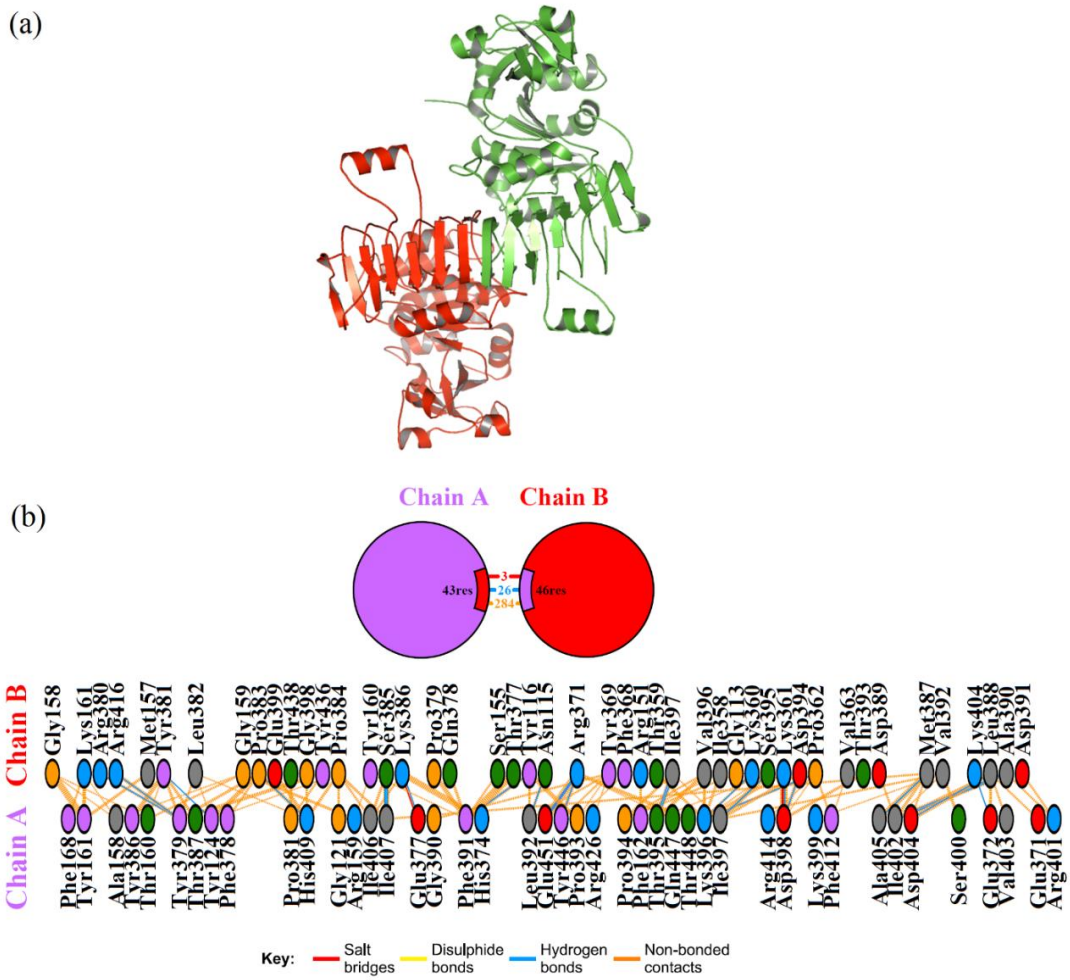


FIGURE 13(a): Side by side conformation between LS and SS of *Phaseolous vulgaris* to form dimer1. LS is coloured in green and SS is coloured in red. 13(b): Side by side interacting residues between LS and SS of *Phaseolous vulgaris* to form dimer1.

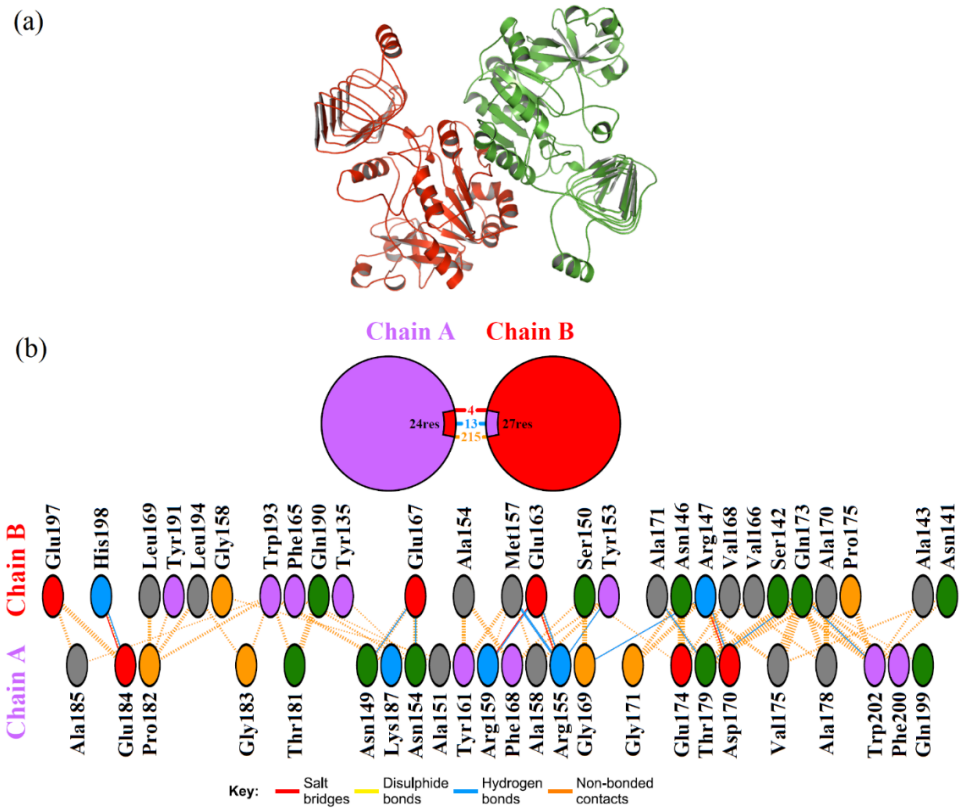


FIGURE 14(a): Up and down conformation between LS and SS of *Phaseolous vulgaris* to form dimer2. LS is coloured in green and SS is coloured in red. 14(b): Up and down interacting residues between LS and SS of *Phaseolous vulgaris* to form dimer2.

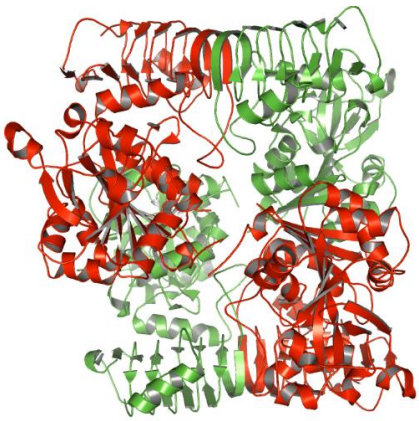


FIGURE 15: Heterotetrameric assembly of *Phaseolous vulgaris*. LS is coloured in green and SS is coloured in red.

(a)

(b)

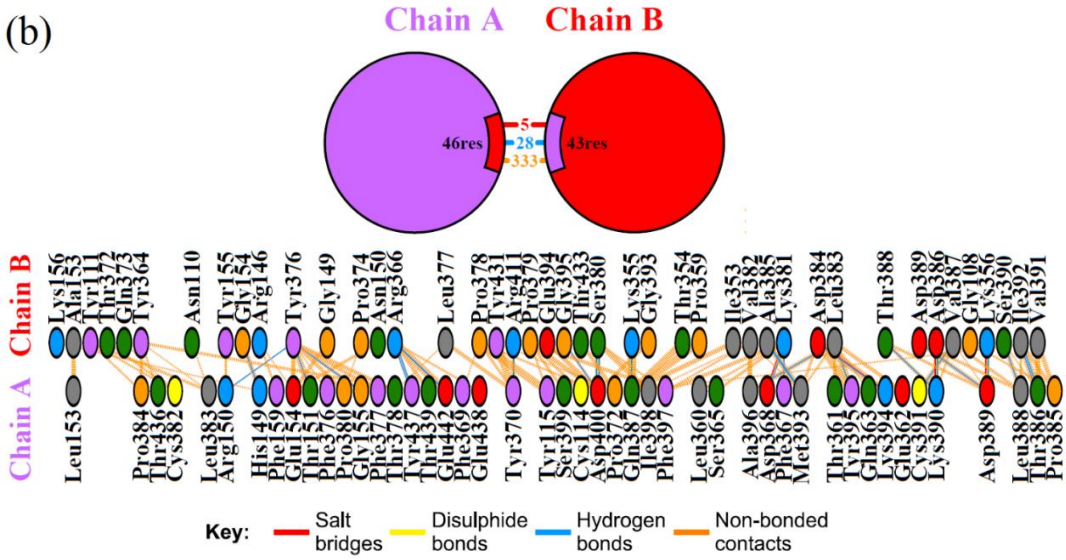


Figure 16(a): Side by side conformation between LS and SS of *Zea mays* to form dimer1. LS is coloured in green and SS is coloured in red. 16(b): Side by side interacting residues between LS and SS of *Zea mays* to form dimer1.

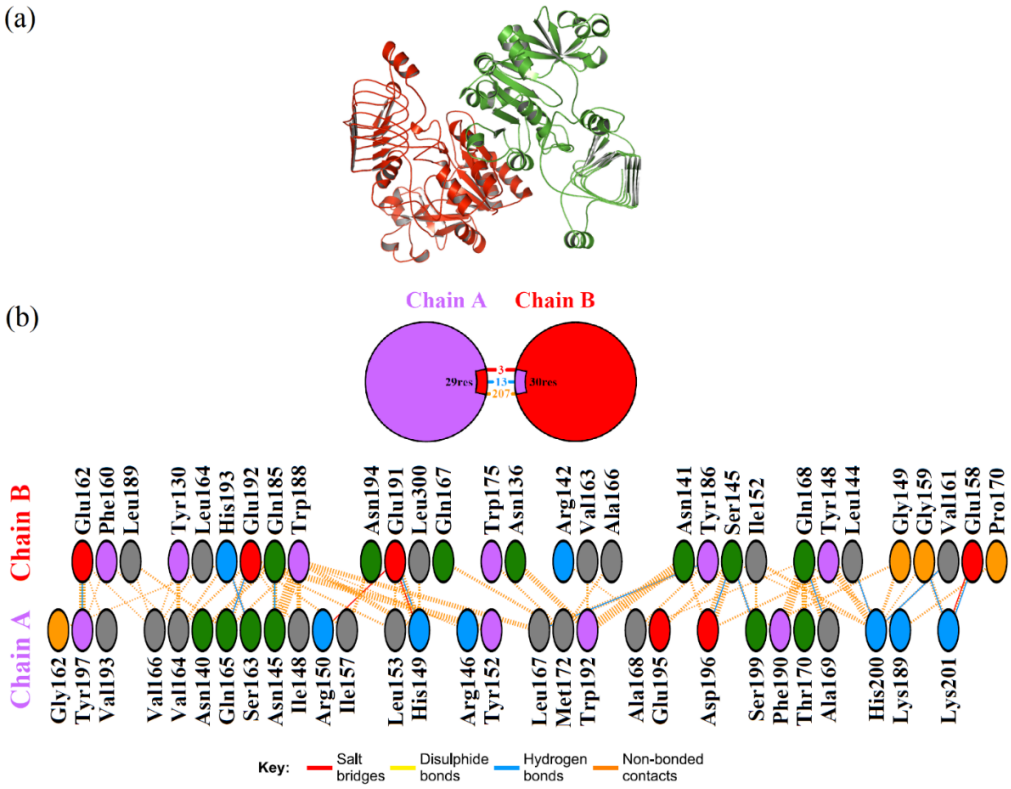


Figure 17(a): Up and down conformation between LS and SS of *Zea mays* to form dimer2. LS is coloured in green and SS is coloured in red. 17(b): Up and down interacting residues between LS and SS of *Zea mays* to form dimer2.

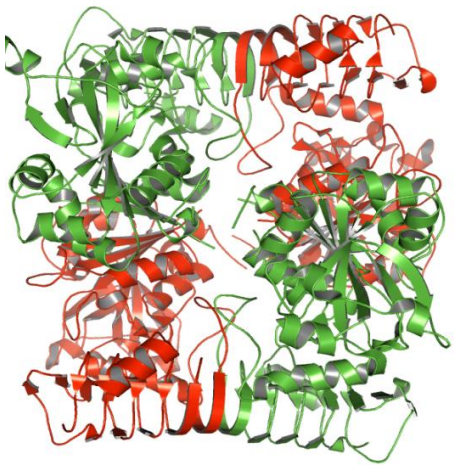


FIGURE 18: Heterotetrameric assembly of *Zea mays*. LS is coloured in green and SS is coloured in red.

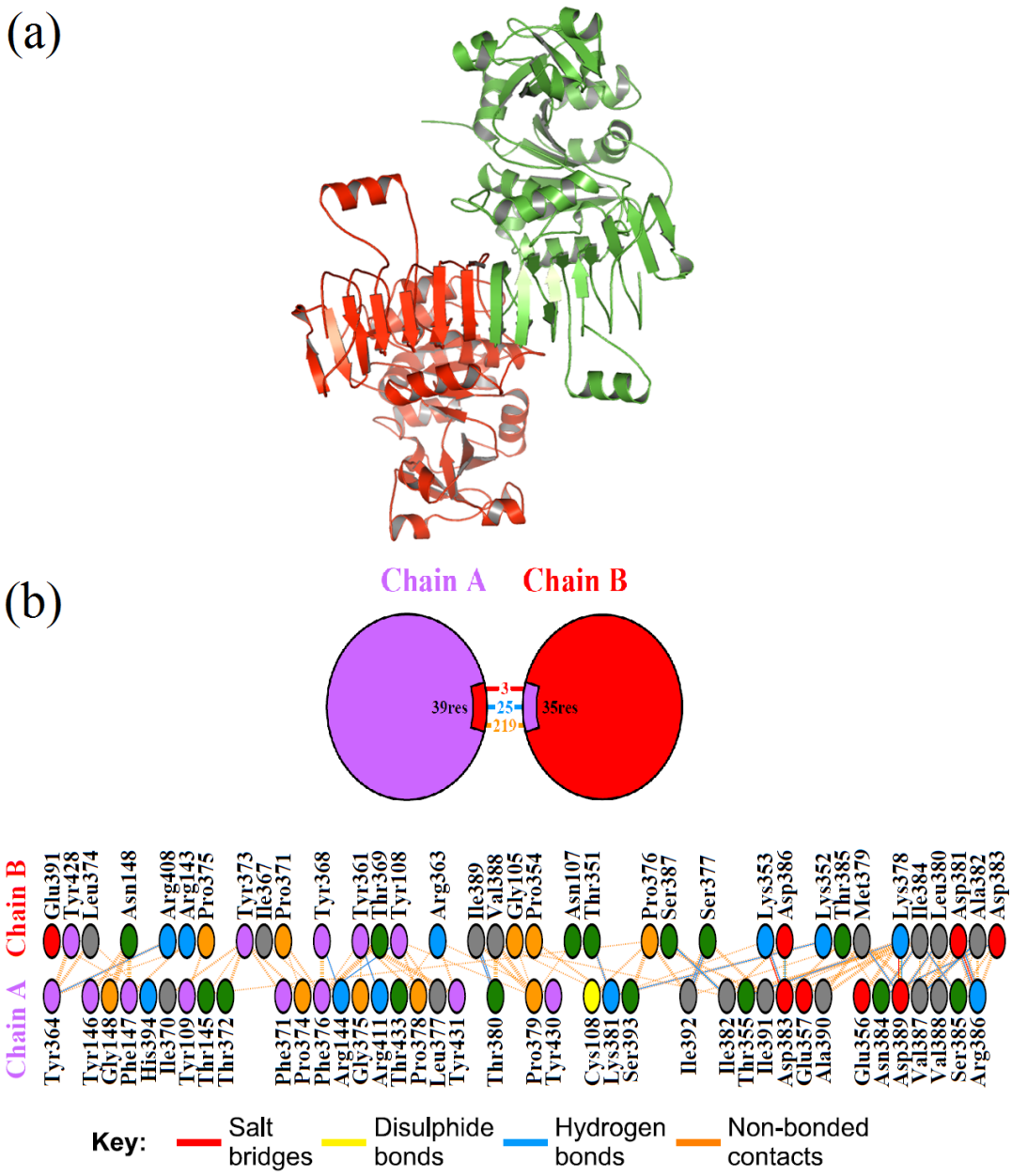


FIGURE 19(a): Side by side conformation between LS and SS of *Pisum sativum* to form dimer1. LS is coloured in green and SS is coloured in red. 19(b) Side by side interacting residues between LS and SS of *Pisum sativum* to form dimer1.

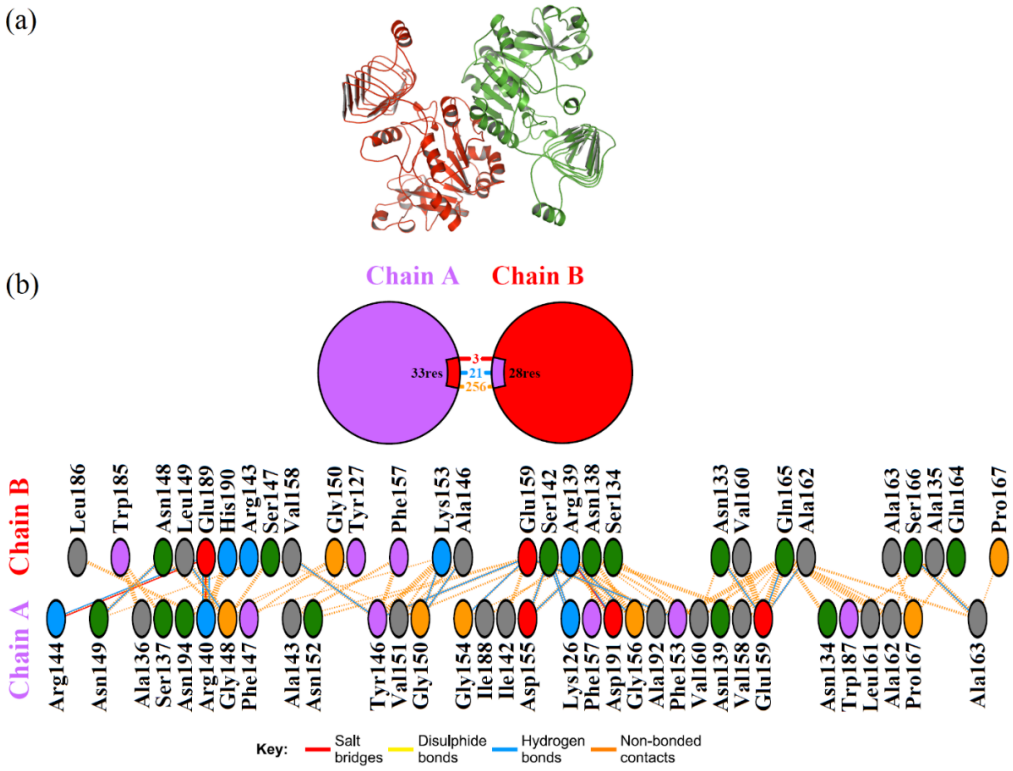


FIGURE 20(a): Up and down conformation between LS and SS of *Pisum sativum* to form dimer2. LS is coloured in green and SS is coloured in red. 20(b): Up and down interacting residues between LS and SS of *Pisum sativum* to form dimer2.

Figure 21: Heterotetrameric assembly of *Pisum sativum*. LS is coloured in green and SS is coloured in red.

Interacting residues among side by side and up and down interactions of some of the selected crop species identified by COCMAPS and DIMPLOT were studied with the help of structural alignment. Figures 22 and 23 show the structural alignment of the surface interacting residues of allosterically regulated and non-regulated SS and LS of AGPases, respectively.

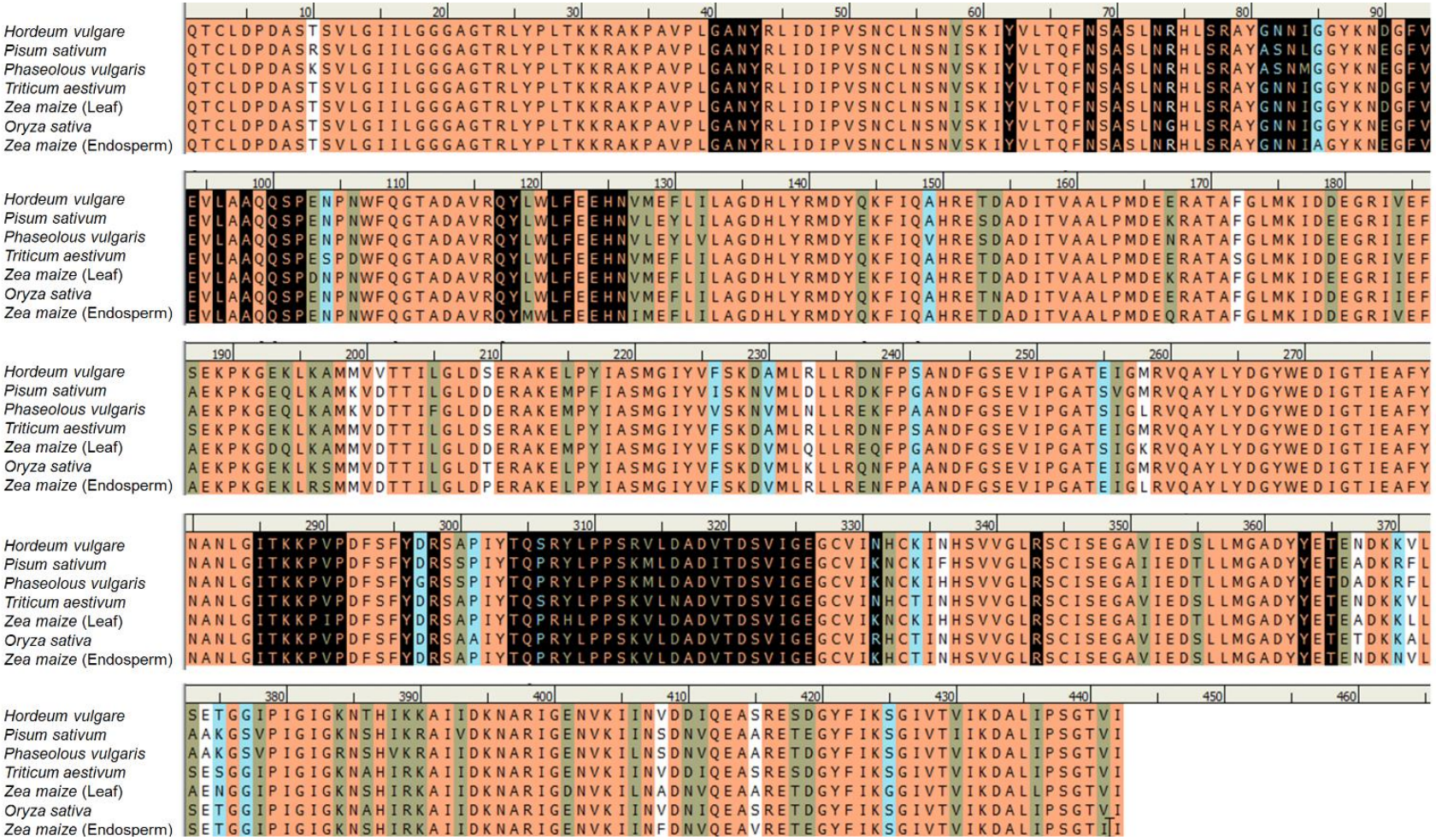


Figure 22: Structural alignment of the surface interacting residues of allosterically regulated and non-regulated SS of AGPase. Interacting residues are highlighted in black colour.

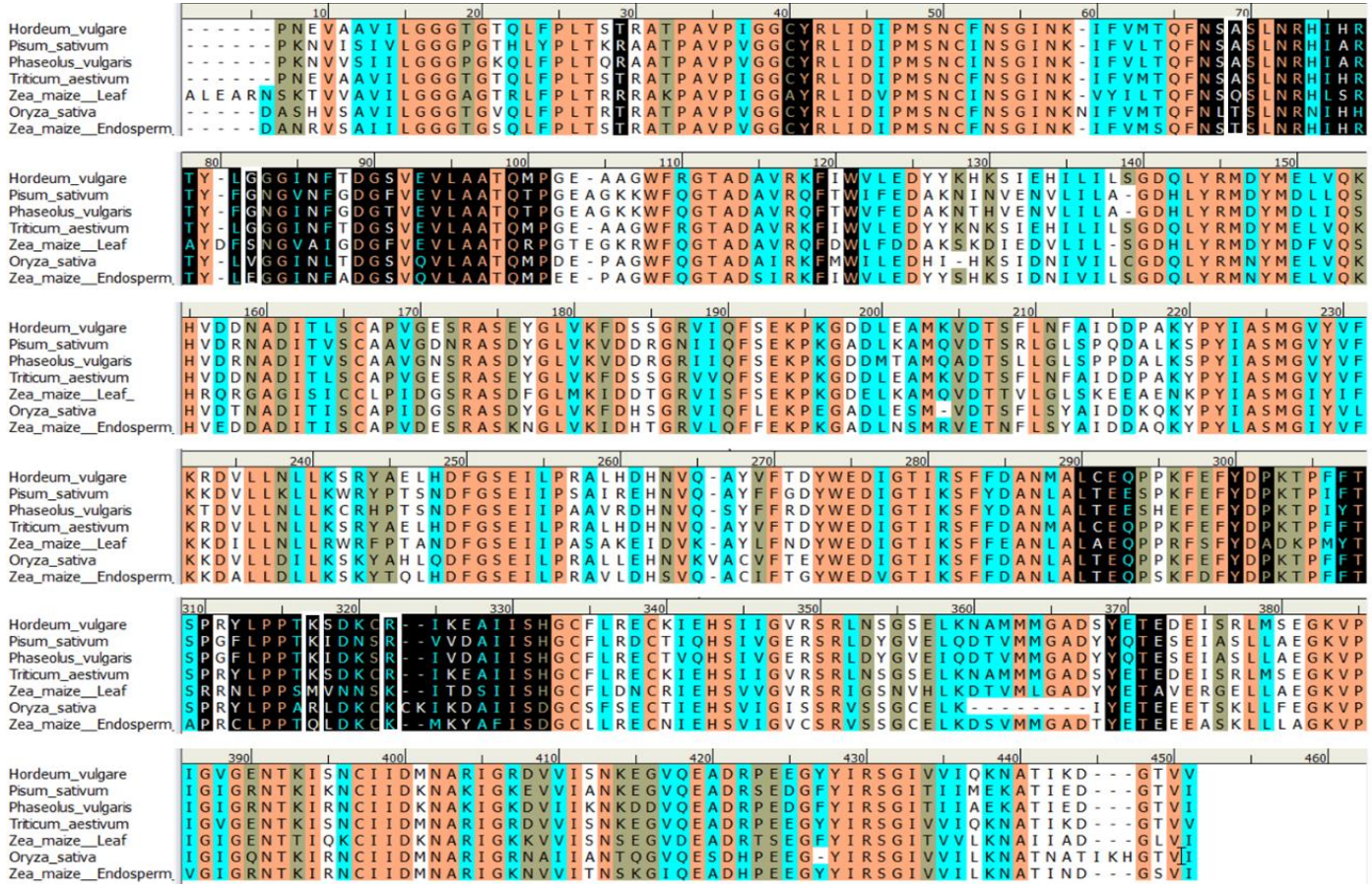


Figure 23: Structural alignment of the surface interacting residues of allosterically regulated and non-regulated LS of AGPase. Interacting residues are highlighted in black colour. Four conserved mutations are labelled with a white box.

Objective 3: Identification and characterization of critical amino acid residues or particular motif involved in allosteric regulation, enzyme activity and thermostability of AGPase.

For completion of the objective 3, following steps were followed.

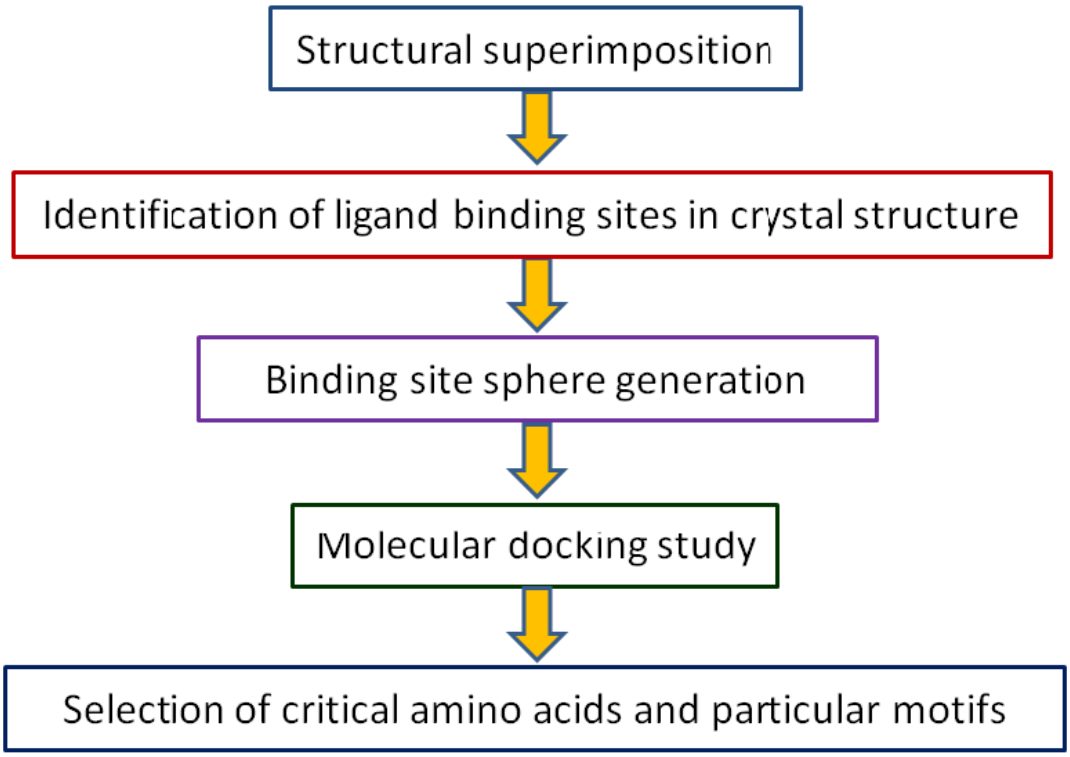


Figure 24: Step by step workflow of Objective 3**.**

Modeled 3D structure of both SS and LS of AGPase were selected for docking study followed by an extensive comparative structural analysis. Previous study by Jin and co-workers (2005) on potato tuber AGPase SS reported that sulphate molecule acts as an inhibitor of the protein. Taking together in the present study, three sulphate molecules were successively docked into the active site of both SS and LS of AGPase using

CDOCKER algorithm to elucidate its structural and functional relevance in terms of inhibitor binding. Structural superimposition of the sulphate binding pockets are shown in Figure 25.

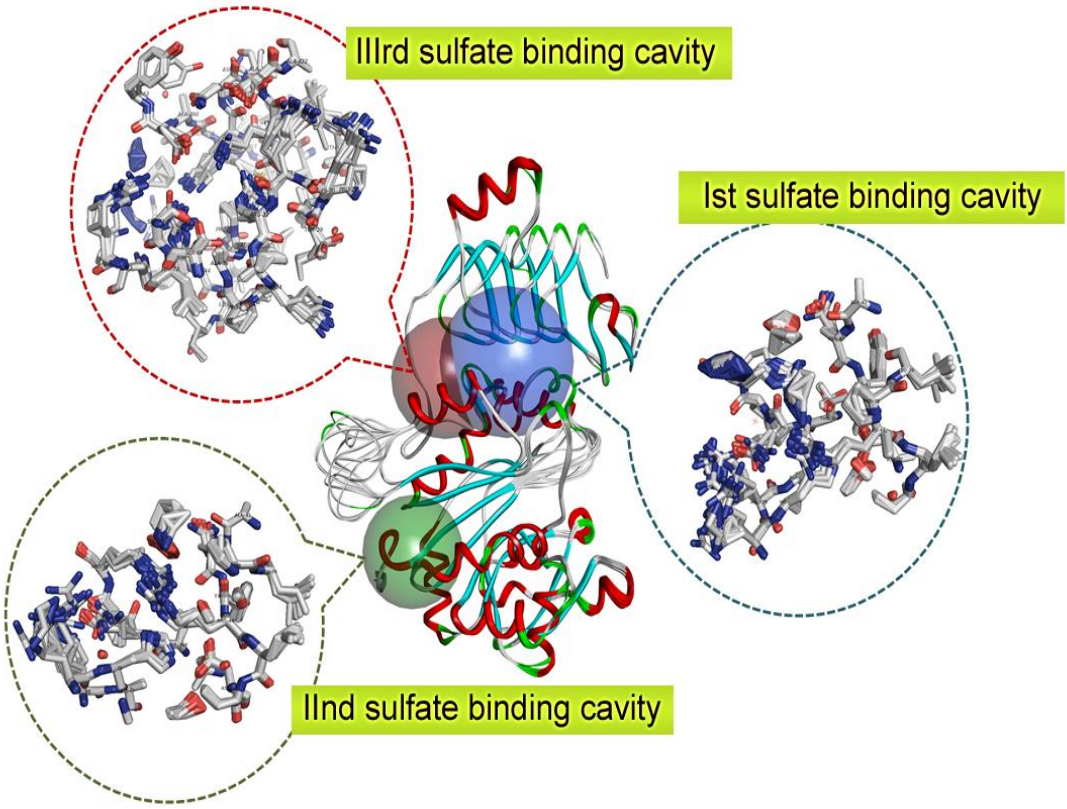
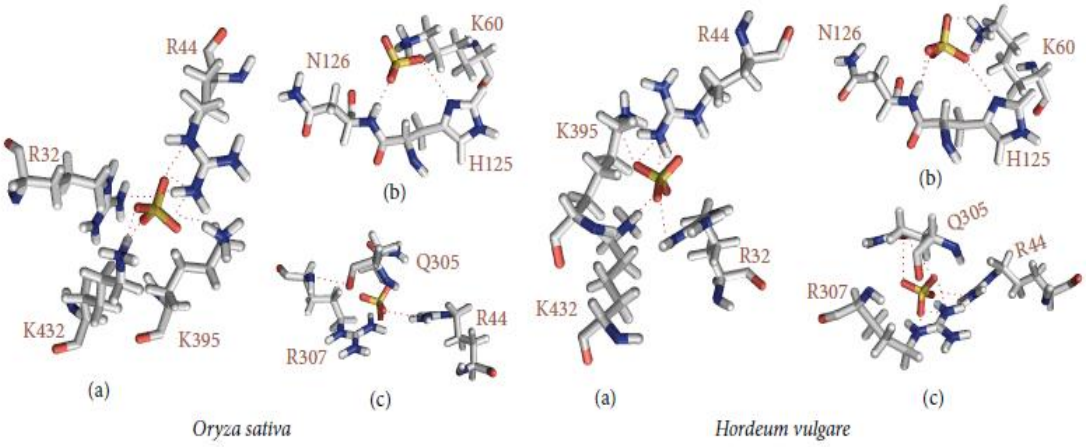
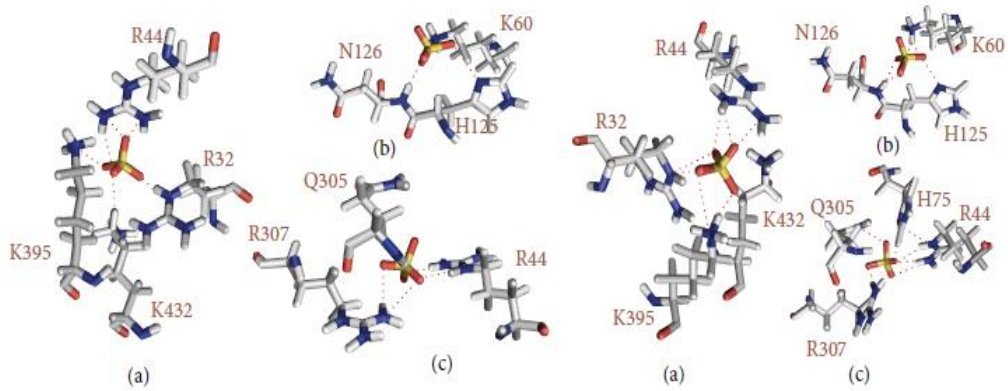


Figure 25: Structural superimposition of sulphate binding pocket forming residues are shown with blue, green and red spheres. Blue sphere shows the 1st sulphate binding site, green sphere shows the 2nd sulphate binding site and red sphere shows the 3rd sulphate binding site.

The most stable docked poses were analyzed with the help of Pymol and some of the docking poses are shown in Figure 26.


R44
N126
K60
R32
H125
(b)
Q305
K432
K395
R44
R307
(a)
(c)
Oryza sativa
R44
N126
K60
K395
H125
(b)
R32
Q305
K432
R44
R307
(a)
(c)
Hordeum vulgare


R44
N126
K60
H125
(b)
R32
Q305
K395
R307
R44
K432
(a)
(c)
Phaseolous vulgaris
R44
N126
K60
R32
H125
(b)
K432
Q305
H75
R44
K395
R307
(a)
(c)
Pisum sativum

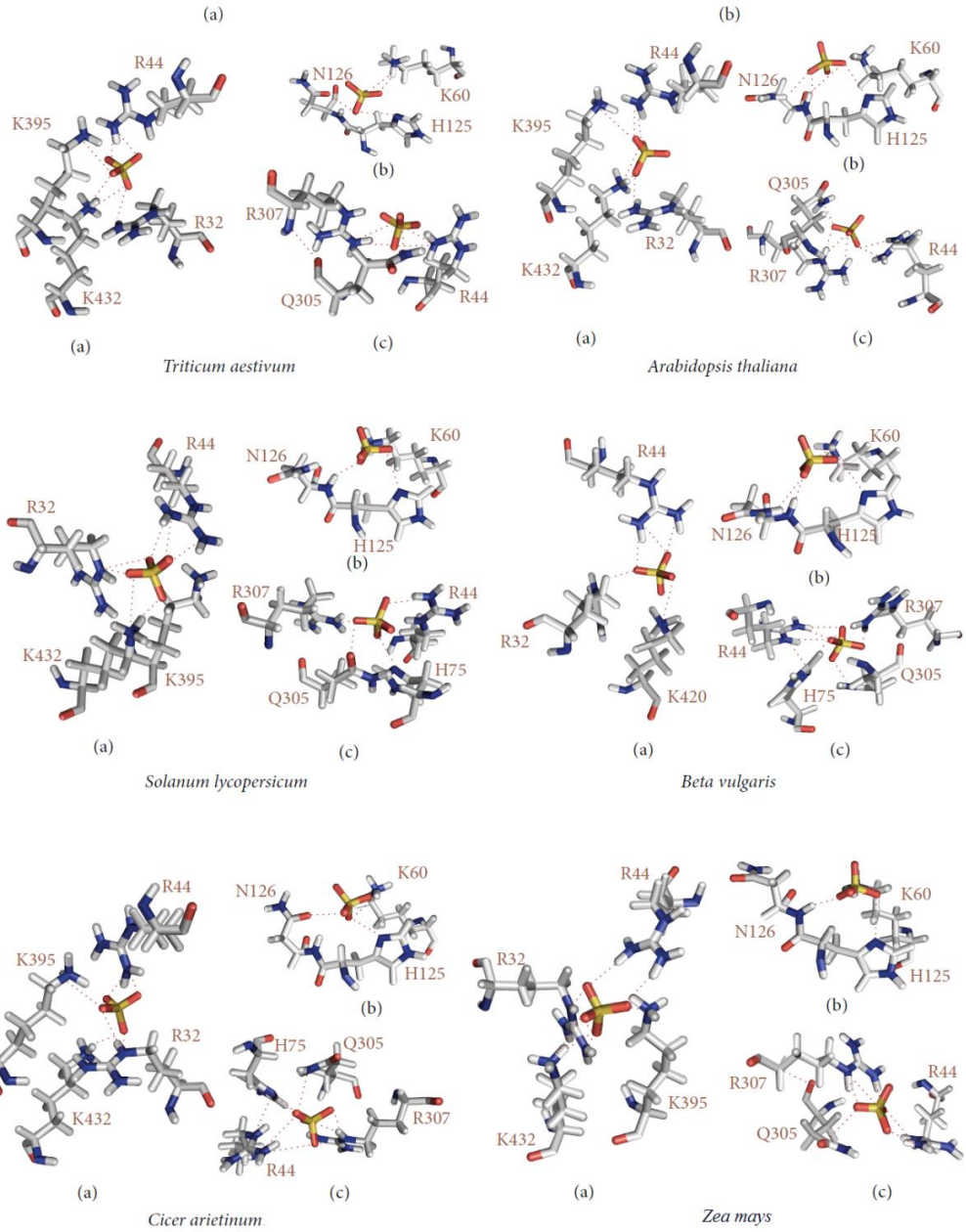
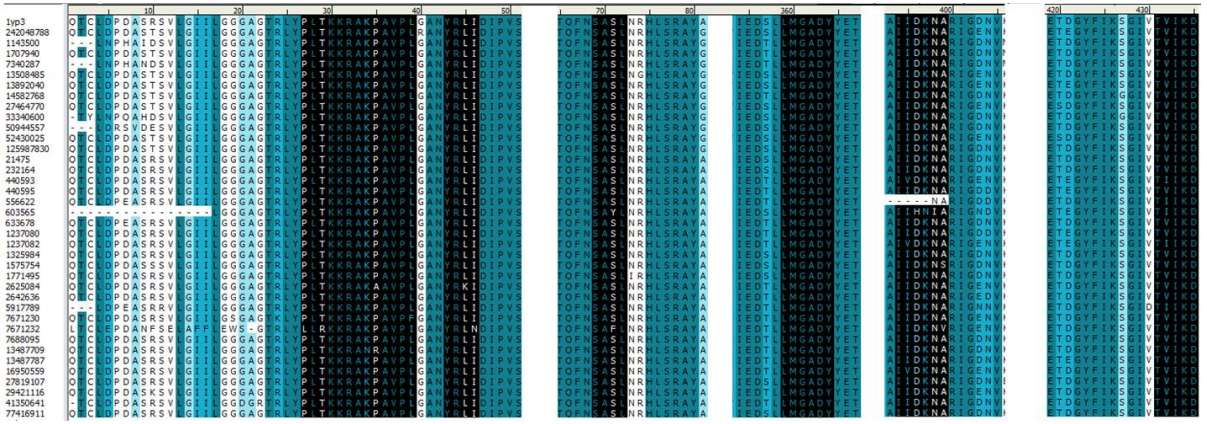
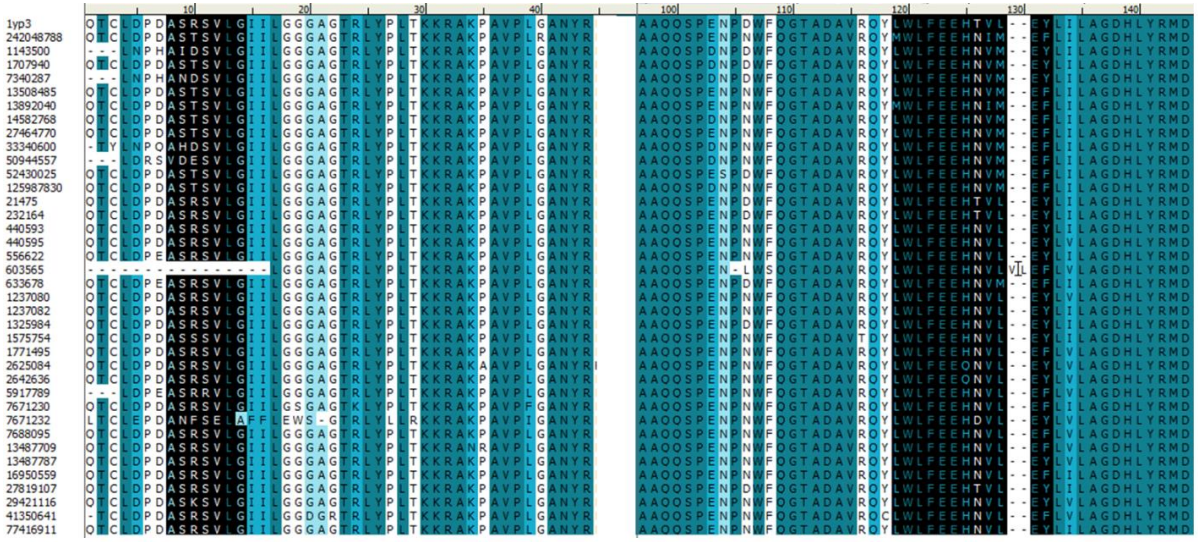


Figure 26: Stick representation of the docking poses of AGPase with sulphate: (a), (b) and (c) represents the 1st, 2nd and 3rd sulphate binding residues. Hydrogen bonds are shown with red dotted lines.

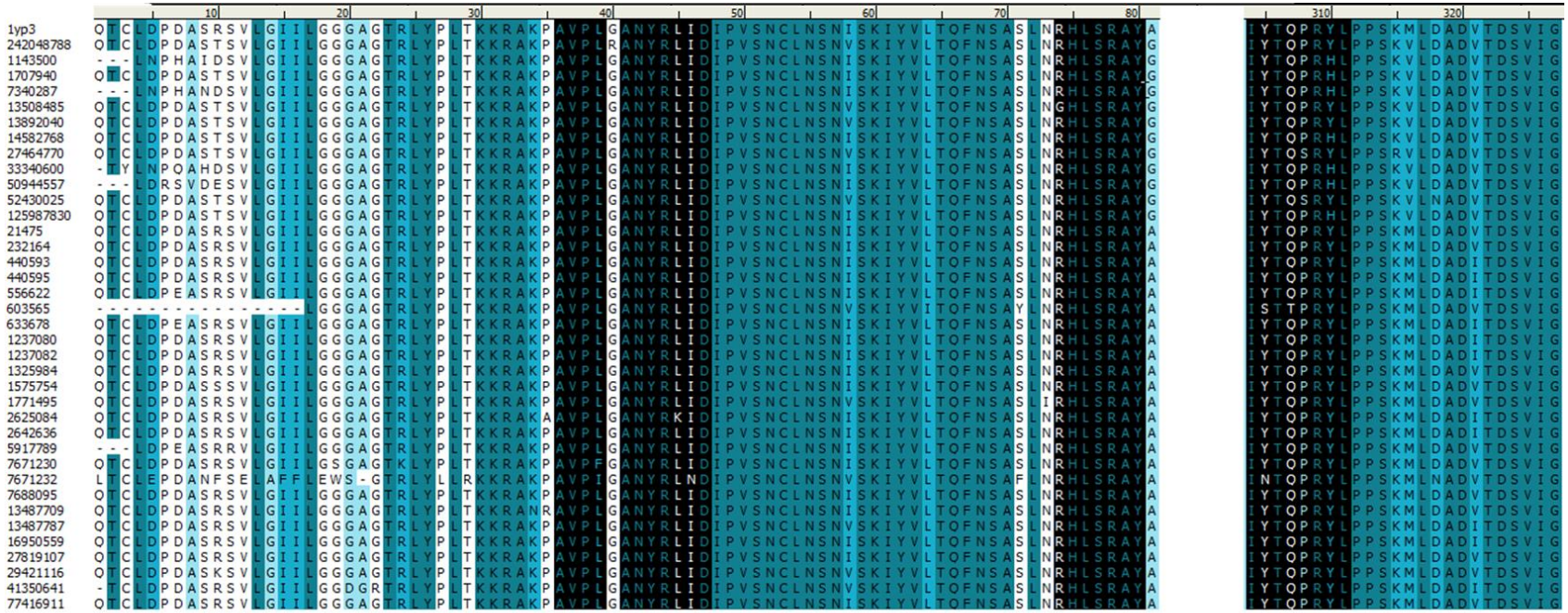
After extensive analysis of the docked poses, sulphate binding cavity forming residues were superimposed to find for its conservedness across different species in both SS and LS. Structural superimposition of the sulphate binding site forming residues of both SS and LS are shown in Figures 27 and 28, respectively. Binding cavity forming residues are coloured in black. This analysis suggested that in case of all the sulphate binding sites, binding cavity forming residues are highly conserved and follows common folding pattern.



27(a)



27(b)



27(c)

Figures 27(a), 28(b) and 28(c) show the structural alignment of the 1st, 2nd and 3rd sulphate binding cavity forming residues all the SS under present investigation. Participating residues are highlighted in black colour. Sequences are represented with their corresponding accession no. For better visibility in the alignment we inserted gaps with white spaces.

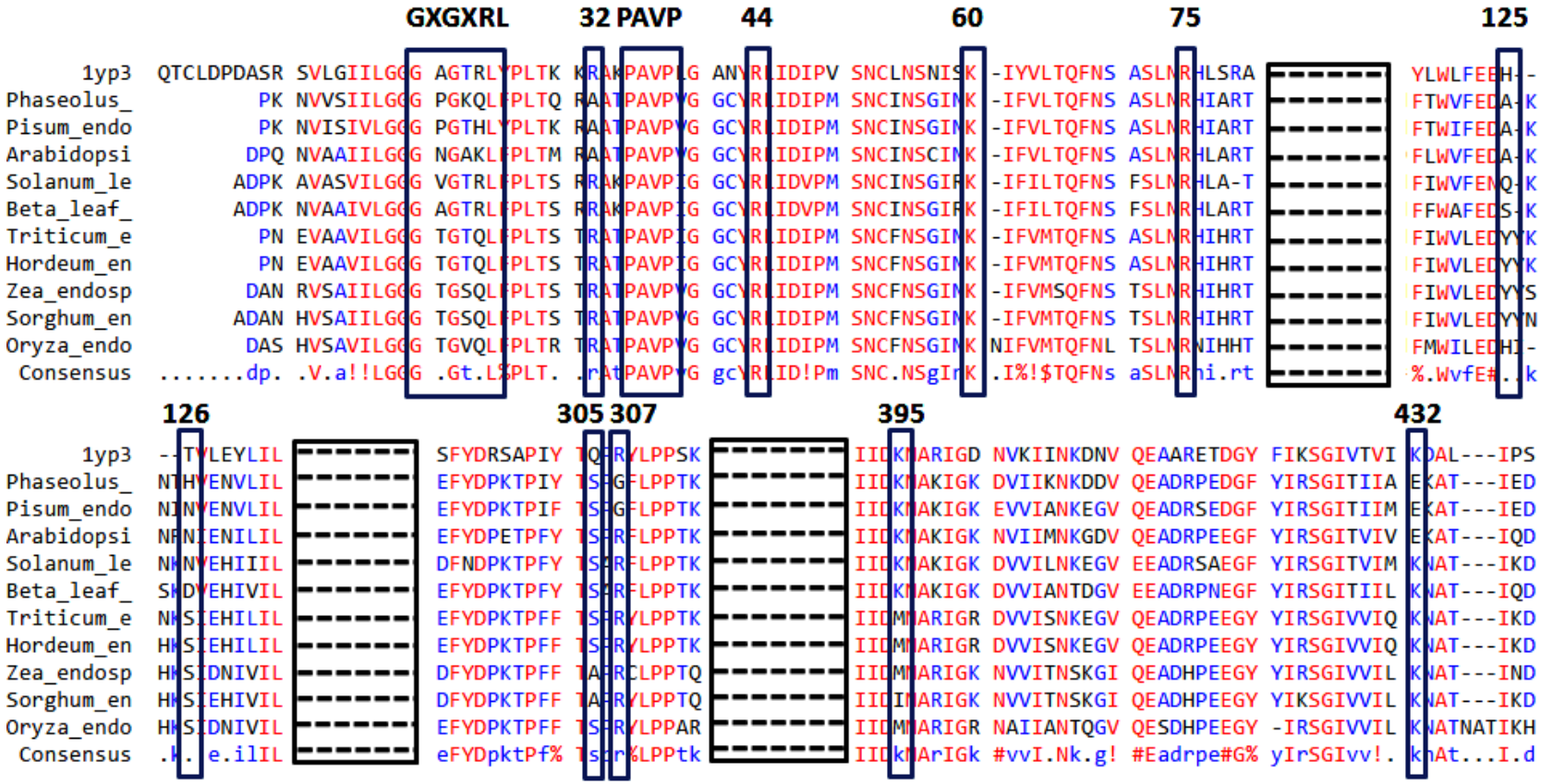


Figure 28: Structural alignment of the LS of AGPase showing the conservation of active site residues. Black boxes represent the key residues responsible for sulphate and substrate binding along with their corresponding positions. For better visibility in the alignment we inserted gaps with white black boxes separated with dahes.

Similarly, substrate binding sites of both SS and LS were superposed to find for conservedness of substrate binding residues across different species in both the subunit. Substrate binding sites of both the subunits are shown in Figures 28-30.

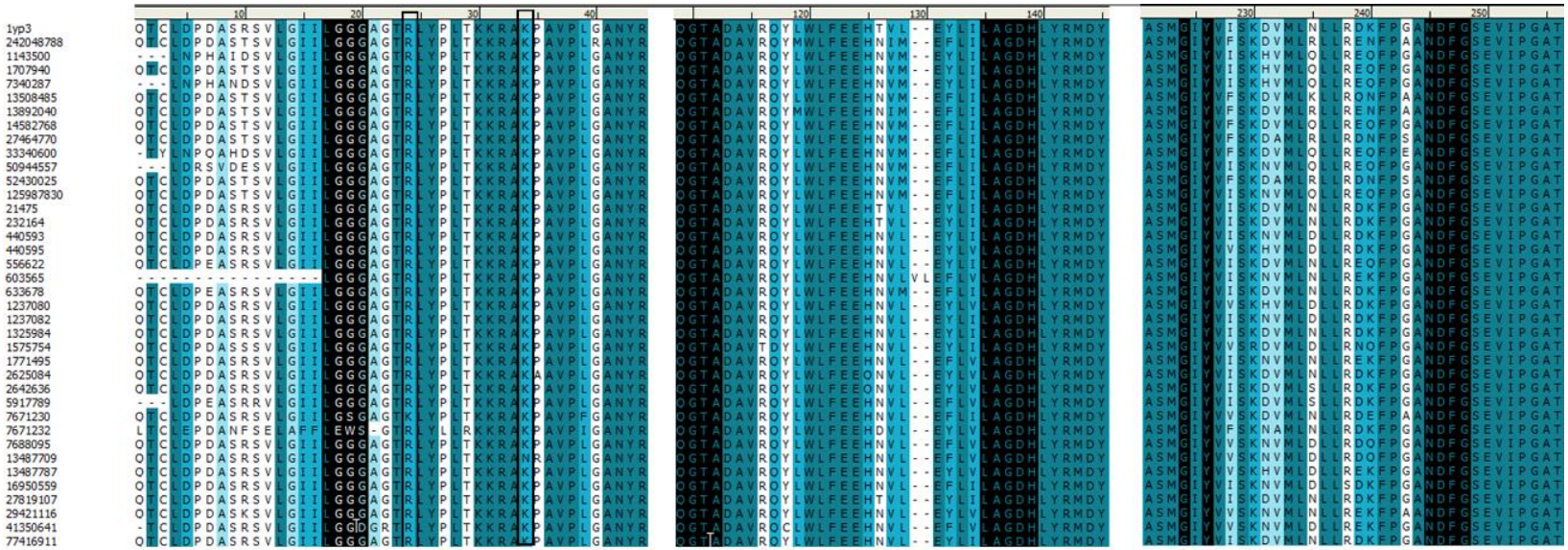


Figure 29: Substrate binding cavity forming residues of the SS. Participating residues are highlighted in black and conserved Arg and Lys are shown with a black box. Sequences are represented with their corresponding accession no. For better visibility in the alignment we inserted gaps with white spaces.

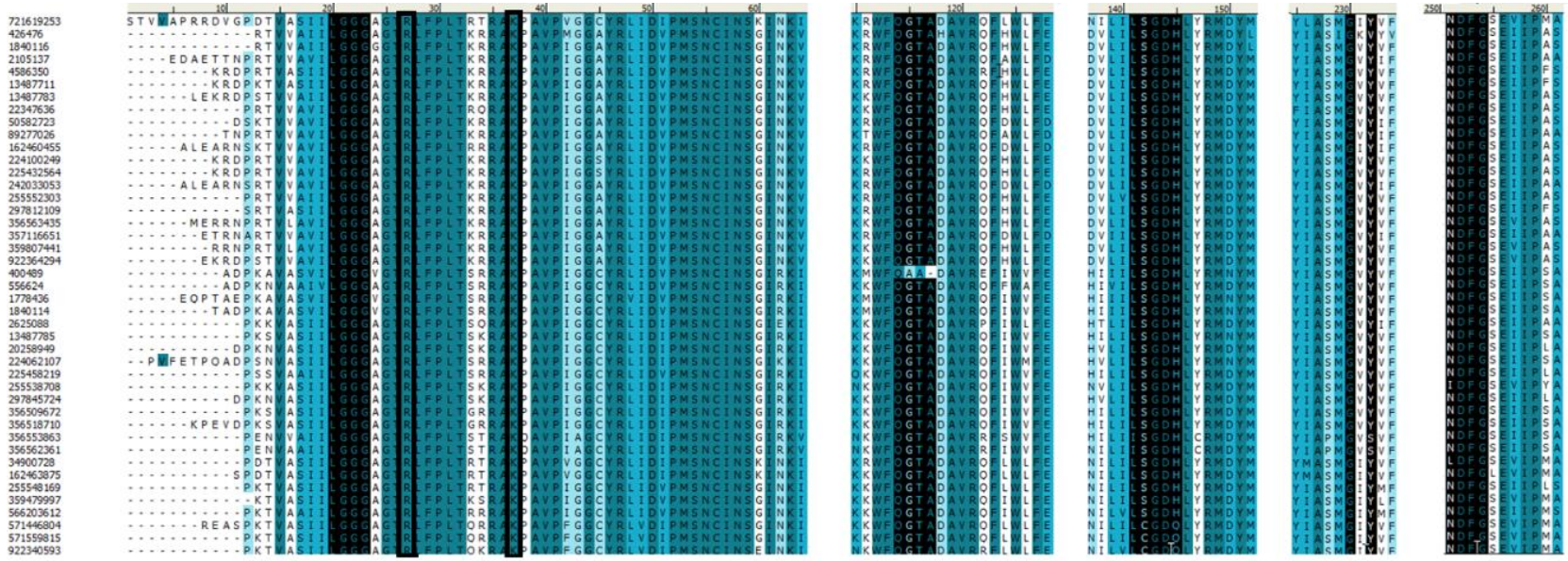


Fig. 30(a)

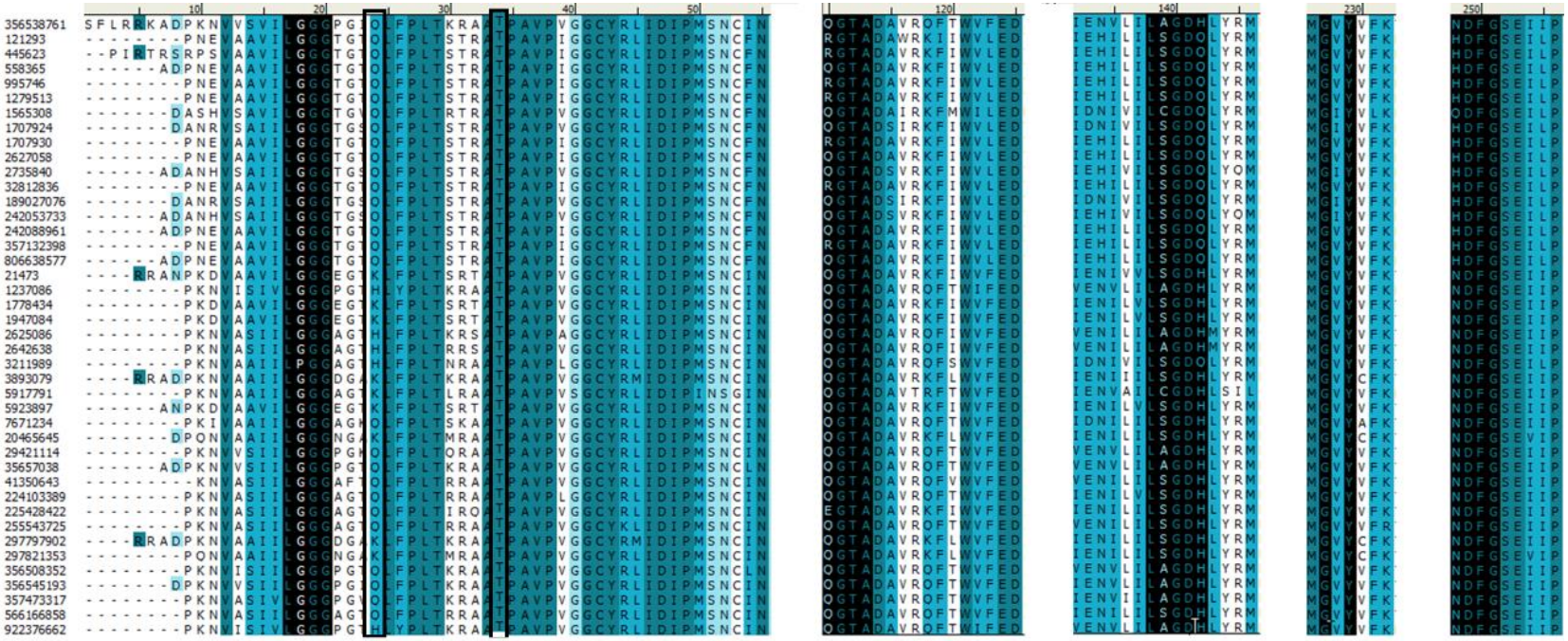


Fig. 30(b)

Figure 30(a) LS substrate binding cavity forming residues from the phylogenetic group 1, 2 and 4. Conserved Arg and Lys residues are shown with a black box. 30(b): Large subunit substrate binding cavity forming residues from the phylogenetic group 3(a) and 3(b). Participating resides are coloured in black and mutations at the conserved Arg and Lys positions are shown with a black box. Sequences are represented with their corresponding accession no. For better visibility in the alignment we inserted gaps with white spaces.

Objective 4: To elucidate the evolutionary relationship, construct the phylogenetic tree of both small and large subunit of AGPase.

For completion of the Objective 4, following steps were followed

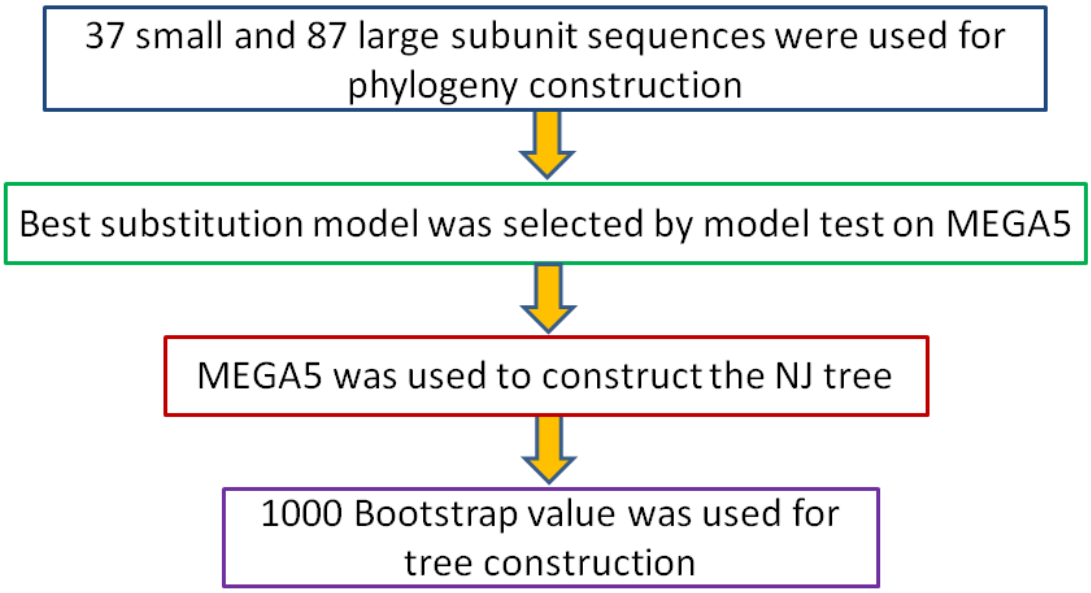


Figure 31: Step by step workflow of Objective 4.

For constructing the phylogenetic tree of AGPase a total of 126 full length nucleotide sequences of both SS and LS were used. Trees were constructed for both the subunits individually. Phylogenetic analysis of the SS AGPase inferred that the higher plant SS group can be divided in to two subgroups: one corresponding to monocots and a second one corresponding to dicots. The phylogenetic tree of SS AGPase is shown in Figure 32.

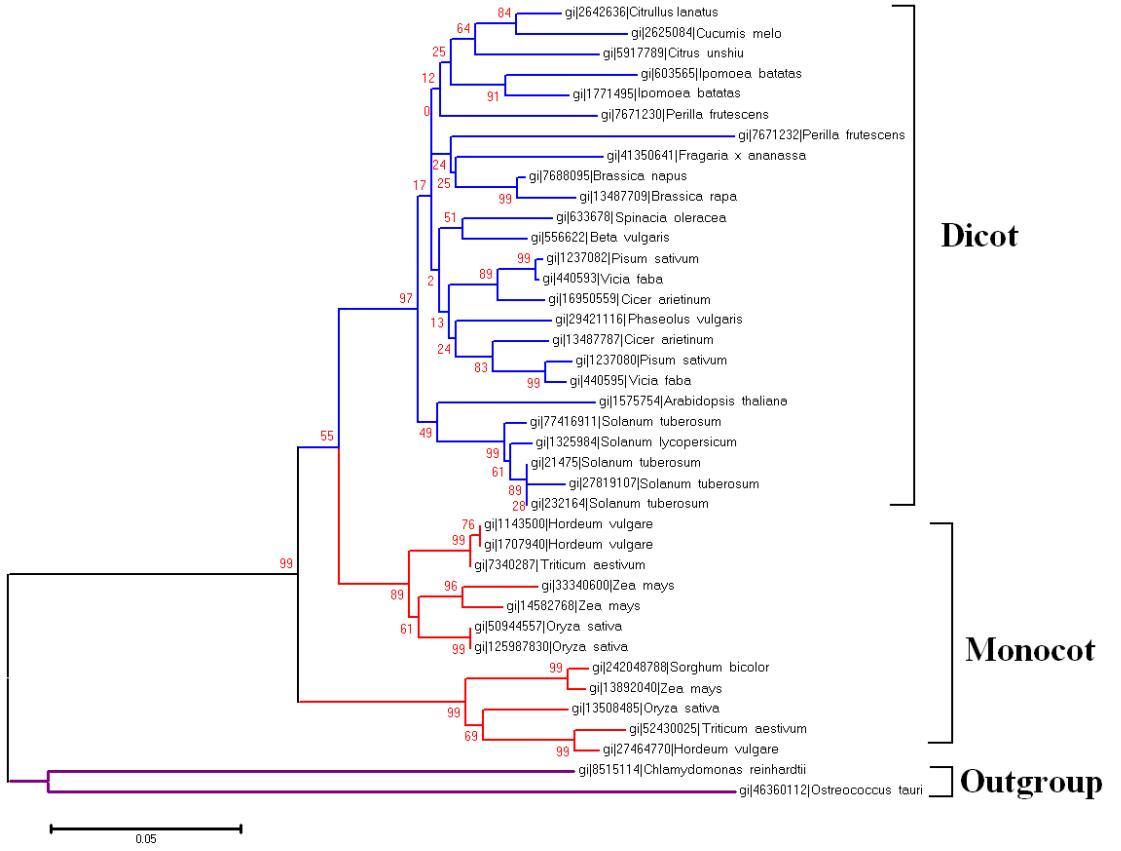


Figure 32: The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 1.73979255 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). The analysis involved 39 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 407 positions in the final data set. Evolutionary analyses were conducted in MEGA5.

In contrast to the SS, LS AGPases were grouped into four major clades some of which included both monocots and dicots. The phylogenetic tree of the LS is shown in Figure 33.

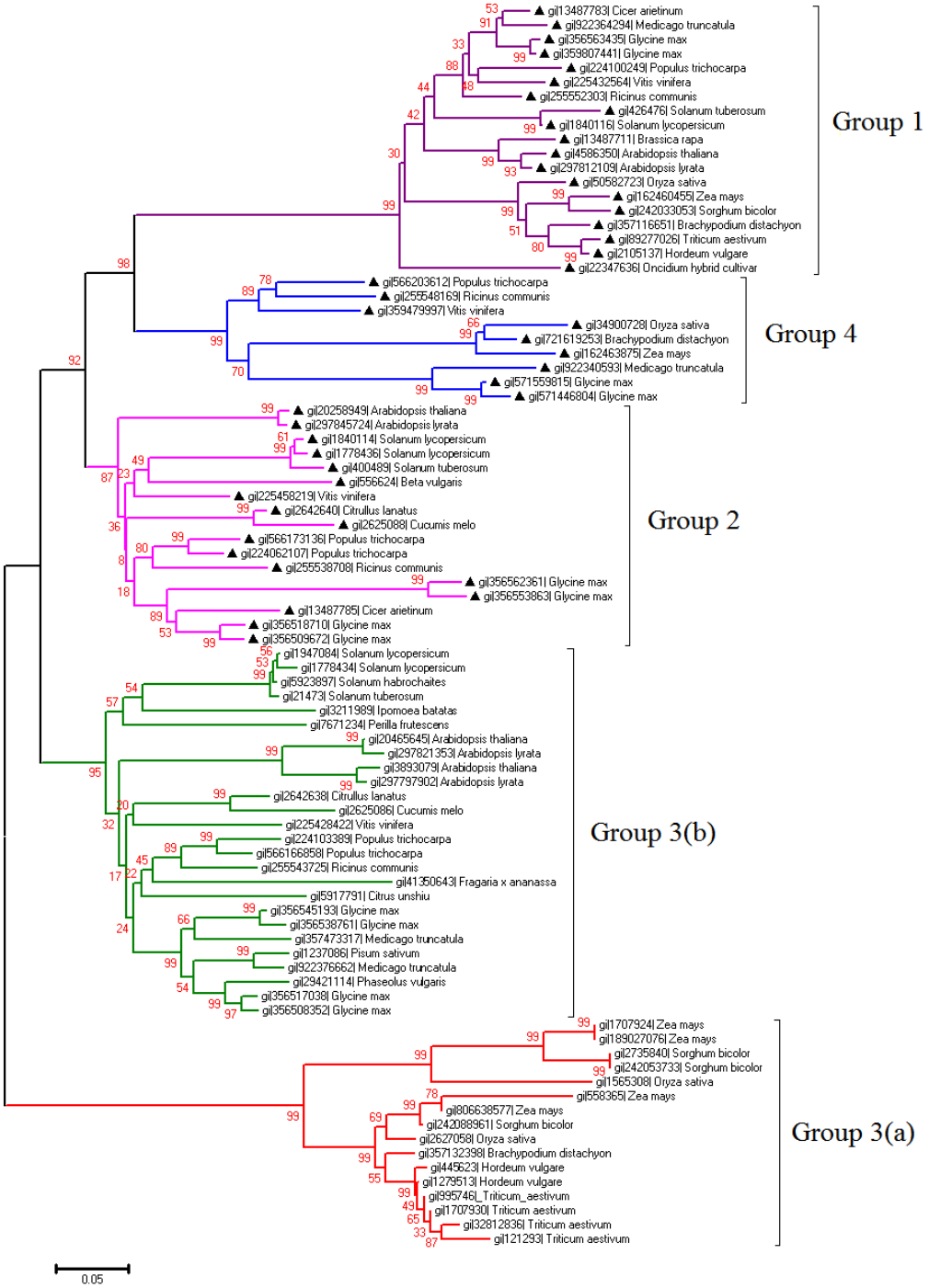


Figure 33: The evolutionary history was inferred using the Neighbor-Joining method. The optimal tree with the sum of branch length = 5.96798444 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the JTT matrix-based method and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1). The analysis involved 87 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 407 positions in the final data set. Evolutionary analyses were conducted in MEGA5.

Objective 5: *In-silico* site-directed mutagenesis of specific amino acid residues or particular motifs responsible for enzyme activity.

Following steps were followed to fulfill the Objective 5.

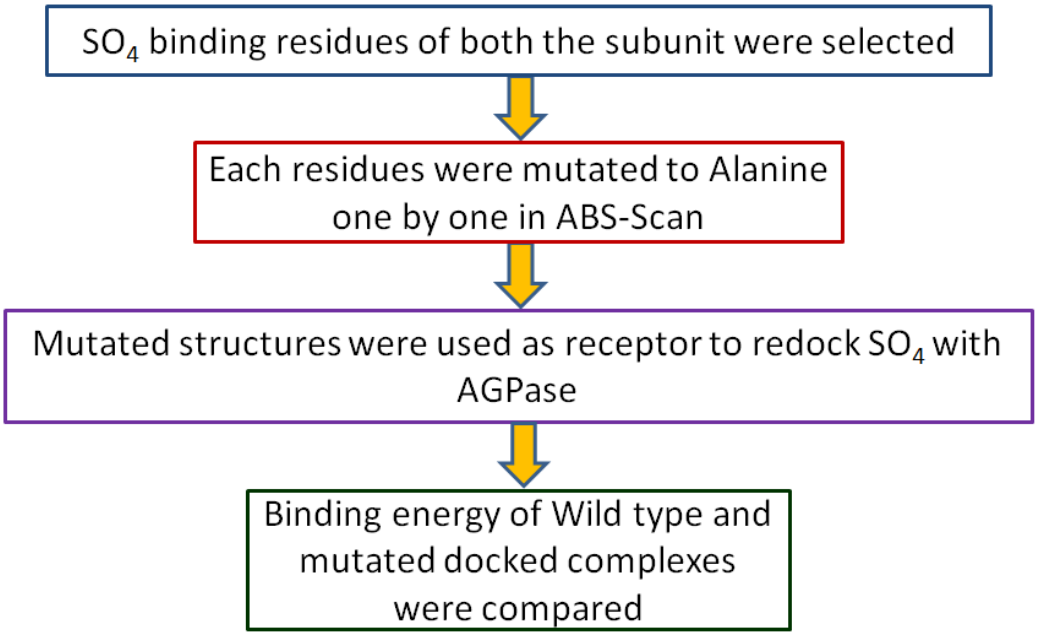


Figure 34: Step by step workflow of Objective 4.

In order to validate the importance of each of the predicted key residues binding with the regulator molecule i.e. sulphate, ASM was performed. ASM was carried out by replacing the side chain ligands of the important AGPase binding residues using ABS-Scan server which later implements AutoDock 4.1 software for binding energy calculation. Binding site distance cut-off was set to 4.5Å. From the docking study it was evident that Arg32, Arg44, Lys395 and Lys432 were involved in the interaction with the first sulphate

molecule of AGPase and accordingly these residues were mutated one by one to alanine in their respective positions. Binding energy of the mutated and the wild type AGPases were calculated for all the SS and LS docked complexes and the average binding energy of all the residues in sulphate binding is shown in Figure 35.

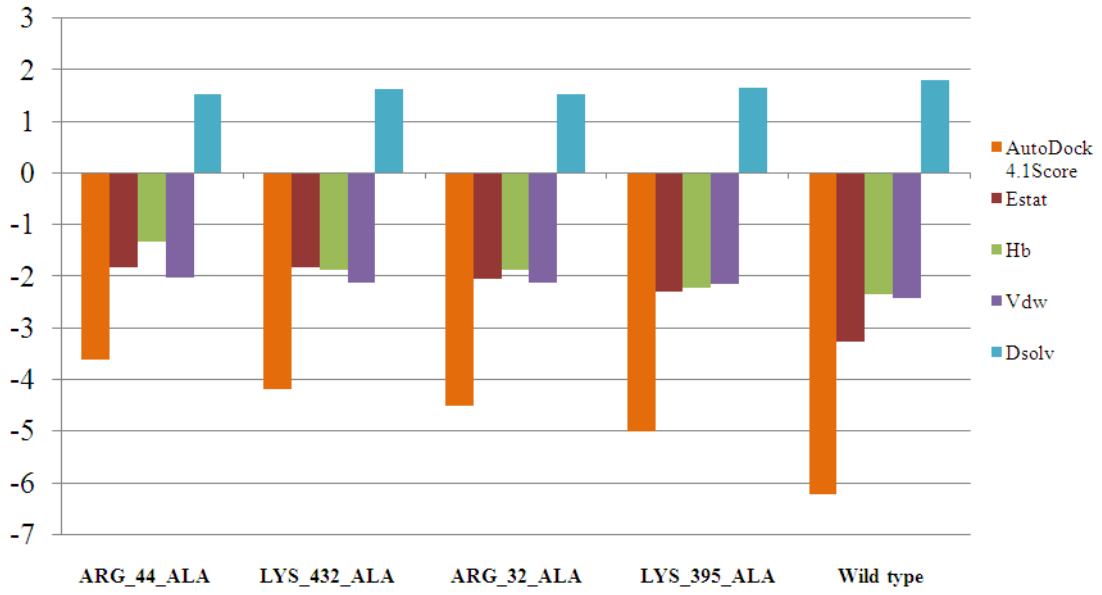


Figure 35: Average alanine scanning mutagenesis score of the 1st sulphate binding site. X axis represents the mutant variant and Y (-Y) axis represents the binding energy in terms of AutoDock score, electrostatic energy, hydrogen bond energy, van der Walls force and desolvation energy.

In most of the cases, second sulphate molecule is attached to Lys60, His125 and Asn126. These three amino acids were accordingly mutated to respective alanine and their binding energy was compared with the wild typ AGPase and the average score is reported in Figure 36.

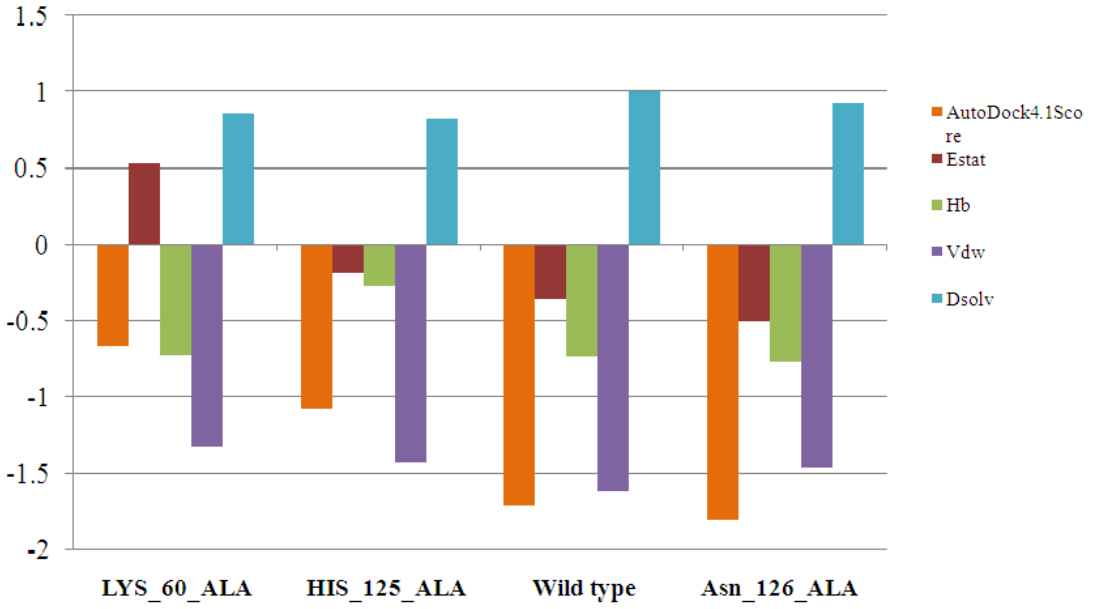


Figure 36: Average alanine scanning mutagenesis score of the 2nd sulphate binding site. X axis represents the mutant variant and Y (-Y) axis represents the binding energy in terms of AutoDock score, electrostatic energy, hydrogen bond energy, van der Walls force and desolvation energy.

The third sulphate molecule binded with Arg44, Gln305 and Arg307 in case of most of the AGPases, however in some cases H75 also showed its binding efficiency with the sulphate molecule. These residues were accordingly mutated one by one to Ala in their respective position. Binding energy of the mutated and the wild type AGPases were calculated for all the SS and LS and the average binding energy of all the residues in sulphate binding is shown in Figure 37.

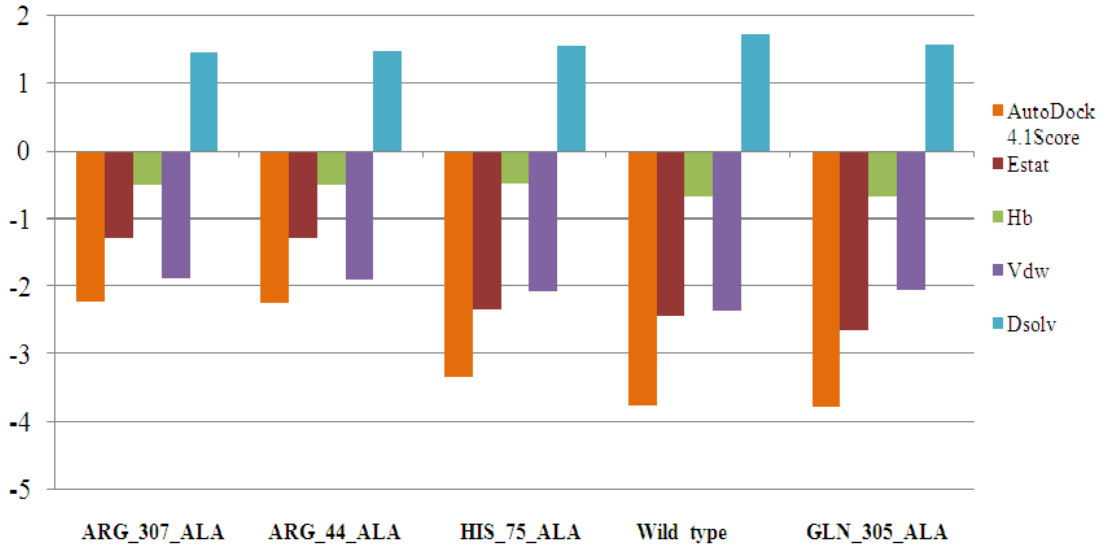


Figure 37: Average alanine scanning mutagenesis score of the 3rd sulphate binding site. X axis represents the mutant variant and Y (-Y) axis represents the binding energy in terms of AutoDock score, electrostatic energy, hydrogen bond energy, van der Walls force and desolvation energy.