This may be treated as Project Completion Report.
Anupam
21/10/16

Progress Report for R&D Projects [Year 2013-2016]

Section-A: Project Details

A1. Project Title: In- silico design and evaluation of sequences for γD- crystallin protein

A2. DBT Sanction Order No. & Date: BT/353/NT/TBP/2012 dated January 4, 2013

A3. Name of Principal Investigator: Dr Anupam Nath Jha

Name of Co-PI/Co-Investigator: Dr Saraswathi Vishveshwara

A4. Institute: Tezpur University, IISc, Bangalore

A5. Address with Contact No. (Landline & Mobile) & Email:

Tezpur University	IISc, Bangalore
Dr Anupam Nath Jha	Dr Saraswathi Vishveshwara
Assistant Professor	Professor
MBBT,	Molecular Biophysics Unit,
Tezpur University	IISc , Bangalore
Contact - 03712-275416, 8876236119	Contact -+91-80 -2293 2611
E-mail - anjha@tezu.ernet.in	E-mail- sv@mbu.iisc.ernet.in

A6. Total Cost: 33.44 Lakhs (Tezpur University) and 15.28 Lakhs (IISc, Bangalore)

A7. Duration: Three year

A8. Approved Objectives of the Project:

(a) Design new sequences for γD crystallin structure

(b) Modeling the three-dimensional structure of the designed sequences

(c) Structural and functional evaluation of the designed sequences for γD crystalline

A9. Specific Recommendations made by the Task Force (if any): Nil

Section-B: Scientific and Technical Progress Report (Jan 2013 –Jul 2016)

B1. Progress made against the Approved Objectives, Targets & Timelines

In-Silico design and evaluation of sequences for γD crystallin protein

DBT Sanction Order No. & Date: BT/353/NT/TBP/2012 dated January 4, 2013

Approved objectives, Targets and timelines

Period of study	Achievable targets
6 Months	Literature survey for different mutated structure of γD crystallin protein and its family and other basic information
12 Months	Topological parameter extraction and evaluation, Setting up the cluster, assessment of its performance with softwares
18 Months	Designing the set of sequences and energy optimization
24 Months	Model building for the selected sequences
30 Months	Molecular dynamics simulation for chosen models
36 Months	Simulation data analysis to obtain the new protein structure

Consolidated progress report

1. Equipment purchasing & Manpower recruitment

a) A computer workstation had been purchased and installed successfully.

b) Printer and AC has also been purchased and installed successively.

c) HPC (High performing cluster) with UPS was procured. It has three nodes (AMD Opteron Processor) with forty cores. Gromacs software along with all the required libraries (C, C++, MPI, Fortran, FFTW, cmake) on linux platform (64 bit) was installed and tested.

d) Software - MATLAB software with different toolbox (Bioinformatics, Statistics, Optimization, Global Optimization, Curve fitting and Parallel computing) has been purchased and installed successfully.

e) 3 JRF (Ms Sewali Bora, Mr Swaroop Chakrabarty, Ms Nikita Bora) had been recruited at different time period of the project.

2. Visit to collaborating institution

Dr A N Jha has visited IISc (during summer break of Tezpur University) to discuss the collaborative work.

3. Attending workshop

a) Dr A N Jha had given a talk in Advances in Mathematical & Computational Biology (AMCB) on "**An in-silico approach for HopS2 effector protein structure prediction**" at IIT-Ropar during 21-22 May 2016.

b) Dr A N Jha attended Annual Meeting of the Indian Biophysical Society on Molecules in "**Living Cells: Mechanistic basis of function**" in IISc, Bangalore (08-10 Feb 2016) and presented a poster entitled: "An in-silico approach for HopS2 effector protein structure prediction".

c) Dr A N Jha has attended a workshop on "**Data Deluge in Biology: Use of High-Performance Grid & Cloud computing**" at Jorhat Medical College Hospital campus, Jorhat during 19-20 December, 2013.

4. Research work

Cataract is one of the leading causes of blindness around the world generally caused due to aggregation of the water soluble eye proteins. This results in the cloudiness of the lens and vision impairment. It disrupts the refractive index of the protein packed lens eventually leading to scattering of light with the loss of lens transparency (opacification). The possible causes include exposure to UV radiation, mutation as well as lifestyle related factors. The major proteins expressed in the eye lens are the crystallins (upto 95%). The β and γ families are the class of crystallin proteins that ensures optimum refraction and minimal light scattering (Benedek, G. B., 1971).

The gamma crystallin proteins have been considered to be the major eye lens proteins constituting 25% of the total crystallin content in the human lens and are one of the three major γ- crystallins required for transparency of the human lens. It is present in high concentrations in the lens nucleus. γD-crystallins are one of the longest-lived proteins. Identification of the γ-crystallin precursors to aggregates is crucial for developing strategies to prevent and reverse cataract (Das et al., 2011). Many forms of congenital hereditary cataract are associated with mutations in the crystallin genes. Therefore a fuller elucidation of the molecular mechanism of crystallin aggregation is essential for understanding cataract formation. Mature onset cataract is associated with the aggregation of partially unfolded or damaged proteins in the lens, which accumulate as an individual ages (Goulet R. D et al., 2011).

Human gamma D-crystallin (HgammaD-Crys) is a monomeric eye lens protein composed of two highly homologous beta-sheet domains. The domains interact through inter domain side chain contacts forming two structurally distinct regions, a central hydrophobic cluster and peripheral residues. The hydrophobic cluster contains Met43, Phe56, and Ile81 from the N-terminal domain (N-td) and Val132, Leu145, and Val170 from the C-terminal domain (C-td) (Flaugh L. S. et al., 2005). Single amino acid substitutions of HγD-Crys are associated with juvenile-onset cataracts. Features of the interface between the two domains conserved among γ-crystallins are a central six-residue hydrophobic cluster, and two pairs of interacting residues flanking the cluster. In HγD-Crys these pairs are Gln54/Gln143 and Arg79/Met147 (Flaugh L. S. et al., 2005).

According to World Health Organization, it is estimated that the cost for surgical procedures and drug services is 6.8 billion dollars which even multiplies further when comes to a developing country like India. Hence, this study aims at designing such crystallin protein sequences (Gamma D crystallin) that could fold and function like the ones we already have in our eye lens, but do not cause aggregation or cataract.

MUTATIONAL ANALYSIS

Literature study was carried out to document all the reported mutations and their effects, in the concerned protein. Most of these mutations have been reported giving rise to Nuclear, Coralliform cataract except R58H which causes aculeiform cataract (Rao V. P., et.al 2013). All of these have also been well attributed with lack of solubility or thermal stability or both.

Table 1- List of 17 mutations (in red) found in human γD crystallin

G	K	I	T	L		E	D	R	G	*10*
F	Q	G	R14	H	Y	E	C	S	S	*20*
D	H	P23	N	L	Q	P	Y	L	S	*30*
R	C	N	S	A35	R36	V	D	S	G	*40*
C	W42	M43	L44	Y	E	Q	P	N	Y	*50*
S	G	L53	Q	Y55	F	L	R58	R	G60	*60*
D	Y	A	D	H	Q	Q	W	M	G	*70*
L	S	D	S	V	R76	S	C	R	L	*80*
I	P	H	S	G	S	H	R	I	R	*90*
L	Y	E	R	E	D	Y	R	G	Q	*100*
M	I	E	F	T	E106	D	C	S	C	*110*
L	Q	D	R	F	R	F	N	E	I	*120*
H	S	L	N	V	L	E	G	S	W	*130*
V	L	Y133	E	L	S	N	Y	R139	G	*140*
R	Q	Y	L	I	M	P	G	D	Y	*150*
R	R	Y	Q	D	W156	G	A	T	N	*160*
A	R	V	G164	S	L	R	R	V	I	*170*
D	F	S								*173*

STRUCTURAL AND FUNCTIONAL ANALYSIS

Every protein has some crucial amino acid positions for maintaining its structural and function. Perturbation of such positions may change the protein structure, in turn which may affect the function. Table 2 is showing the list of such amino acids in the selected protein. It is important to note that table 1 is a subset of table 2. All the reported mutations which cause disease are also play an important role in either protein structure or function.

Table 2: List of structurally and functionality important residue of γD crystallin (in red)

G1	K2	I	T4	L	Y6	E7	D	R9	G
F11	Q12	G	R14	H15	Y16	E17	C	S	S
D21	H	P23	N	L	Q26	P	Y28	L	S30
R	C	N33	S	A35	R36	V	D	S	G
C	W42	M43	L44	Y45	E46	Q47	P	N49	Y50
S51	G	L53	Q54	Y55	F56	L	R58	R	G60
D61	Y62	A	D64	H65	Q66	Q67	W68	M69	G
L71	S72	D73	S	V	R76	S	C	R79	L
I81	P	H	S84	G	S	H87	R88	I	R
L91	Y92	E	R94	E95	D96	Y97	R98	G	Q100
M	I102	E	F104	T105	E106	D107	C	S	C110
L111	Q112	D113	R	F115	R	F117	N118	E119	I120
H	S122	L123	N124	V125	L126	E127	G	S129	W130
V	L132	Y133	E134	L135	S136	N137	Y138	R139	G
R141	Q142	Y143	L144	I	M	P147	G	D	Y150
R	R	Y153	Q154	D	W156	G	A158	T159	N160
A161	R162	V163	G	S	L166	R	R168	V169	I170
D	F172	S							

COMPUTATIONAL ANALYSIS OF THE WILD TYPE AND MUTANT FORMS

Computational analysis for the gamma D crystallin protein and the mutant forms were carried out. The mutant structures for the gamma D crystallin protein available from the Protein Data Bank are 2KFB, 1H4A, 4GR7 and 2G98 and the normal form of the protein available is 1HK0. Structures were downloaded from the PDB database. Compared to the normal structure, in the 4JGF structure Proline at the 23rd position has been substituted by Threonine at the 24th position.

While for the structure 1H4A, the Argininge amino acid residue has been substituted by the Histidine residue at the 58th position. For the structure 4GR7 mutation occurred at the 42nd position where a Tryptophan residue was found to be substituted by an Arginine residue. Similarly for the structure 2G98, the mutation was seen at position 36 where an Arginine residue was substituted by a serine residue.

1) Structure comparison

Structural differences of the γD proteins were analyzed by superimposing the native γD crystallin protein to the mutant forms. It was observed that the structures were very much similar to each other with a RMSD value of 0.336Å. Without any significant change in the structure, the mutant forms of the proteins are responsible for causing aggregation and disease. This leads to the further study of the mutant forms of the protein and factors governing the aggregation process.

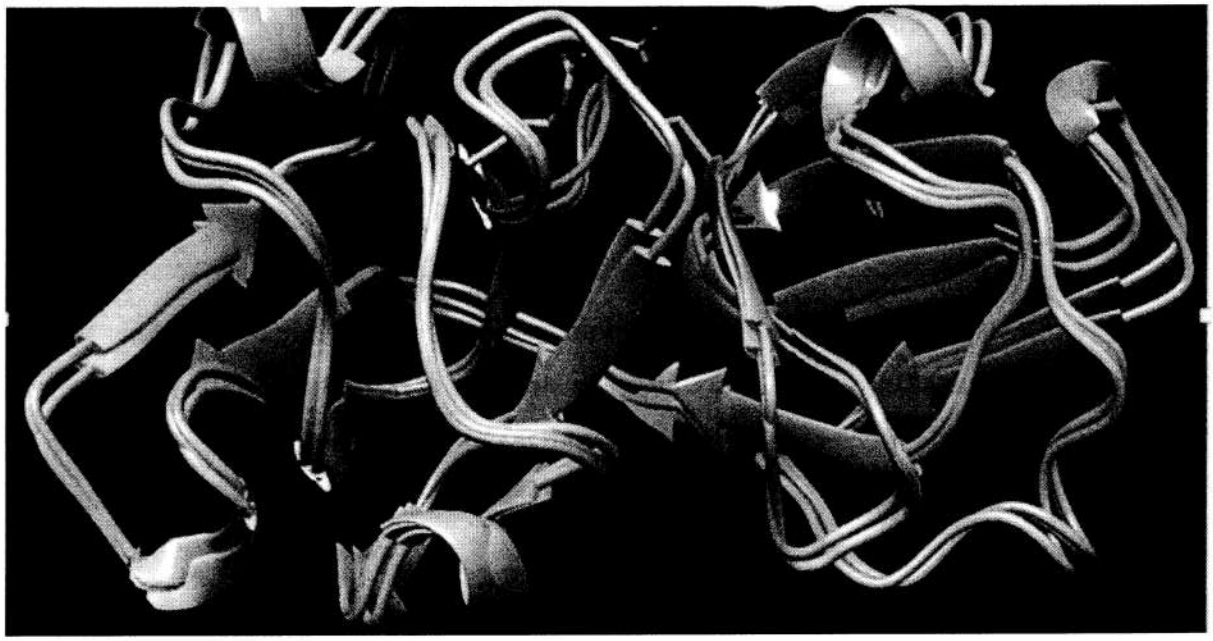


Figure1: The superposition of wild type and mutant structure of γD crystalline

2) Sequence based thermo-dynamical analysis

It has been a routine practice to check the protein stability by different parameters. Here we have used a web-sever MuStab (Mutant protein Stability change), which helps in predicting the change in protein stability after amino acid substitutions (Teng, S, et. al. 2009). A single mutant may change the structural stability of protein. This sever includes the biochemical features Co (overall amino acid composition) and P (polarity), the structural features A (conformational parameter for alpha-helix), B (conformational parameter for beta-sheet) and Aa (average area buried on transfer from standard state to folded protein), and the other biological feature No (number of codons for an amino acid). We have used this web-sever for predicting the change in protein stability because of different mutants.

Table 3: Predicting the change in Protein stability due to mutational change

S.No	Mutation	Change in Stability	Confidence score (in %)
1	R14C	+	26.25
2	R14S	-	78.57
3	P23T	+	26.25
4	P23S	-	81.25
5	P23V	+	28.75
6	A35P	-	92.32

7	R36S	-	84.46
8	R36P	-	85.54
9	W42R	-	92.50
10	L44P	-	90.71
11	L53P	-	90.36
12	M43V	+	24.64
13	R58H	-	90.71
14	G60C	-	79.64
15	R76S	-	88.21
16	E106A	-	91.07

(Here "+" and "-"sign indicates increase and decrease in stability)

It has been found that most of the mutations are decreasing the stability with high confidence score whereas the score is very low (less than 30%) for the opposite ones. It indicates that all the diseases causing mutation are reducing the protein stability and hence eye vision becomes affected by protein aggregation.

We have further validated this data by using another web-sever (DUET http://bleoberis.bioc.cam.ac.uk/duet), which predict the effects of mutations on protein stability via an integrated computational approach (Douglas et.al 2013). The server gives the change in folding free energy upon mutation (ΔΔG in kcal/mol) and the effects of mutation have been categorized in the basis of ΔΔG value. The table 4 has shown the relation between ΔΔG values and the effect.

Table 4: ΔΔG values and the effect of mutation

ΔΔG (in kcal/mol)	**Effect of Mutation**
<= 2.0 and >1.0	Stabilizing
<= 1.0 and >0.5	Slightly stabilizing
<= 0.5 and >= -0.5	Neutral
< -0.5 and>= -1.0	Slightly destabilizing
< -1.0 and >= -2.0	Destabilizing

The results obtained from this server have been shown in Table 5. It is interesting to see here that almost all the mutations destabilize the protein. All this information helped us in making rules of designing the new sequence for de novo protein.

Table 5- Prediction of folding free energies (ΔΔG) and their effects upon mutation

S.No	Mutation Position	ΔΔG (in kcal/mol	Effect of mutation
1	R14C	+	↓
2	R14S	-	↓
3	P23T	+	↓
4	P23S	-	↓
5	P23V	+	↓
6	A35P	-	↑
7	R36S	-	↓
8	R36P	-	↓
9	W42R	-	↓
10	L44P	-	↓
11	L53P	-	↓
12	M43V	+	↓
13	R58H	-	↓
14	G60C	-	↓
15	R76S	-	↓
16	E106A	-	↓

3) Analysis of H-bonding patterns

Hydrogen bonds are very important component in maintaining the conformation and properties of the biomolecules. Analysing the hydrogen bonding patterns in the proteins reveals the underlying protein folding and unfolding process. In the gamma-D crystallin proteins, H bonding study showed the pattern of Hydrogen bond formed by the residue in the wild type protein and the mutated residue in the mutant form of the proteins.

The downloaded structures were visualized in Pymol and their respective bonding patterns were analyzed. Changes in the bonding patterns were observed in the mutant forms of the proteins from the native bonding patterns which might govern the aggregation process of the gamma D crystallin proteins associated with other factors.

Table 6: List of Hydrogen Bonds in the native and the mutated forms

1HK0 (main chain-side chain)	1H4A (main chain-side chain)	2G98 (main chain-side chain)	4GR7 (main chain-side chain)
Arg58-Tyr62 (2 H Bond)	His58-Tyr62 (2 H Bond)		
Arg58-Arg59 (1 H Bond)			
Arg36-Tyr6 (1 H Bond)			
Arg36-Asp61 (1 H Bond)		Ser36-Asp61 (1 H Bond)	
		Ser36-Ala35 (1 H Bond)	
Gly1- Trp42 (2 H Bond)			Gly1-Arg42 (1 H Bond)
			Ser20-Arg42 (1 H Bond)
			Cys18-Arg42 (1 H Bond)

3) Solvent accessible surface area (SASA):

The solubility of the wild type and the mutant proteins were studied through the calculation of the solvent accessible surface area (SASA) which gives the surface area of the protein that is accessible to a solvent. In this study the solvent accessible surface area was calculated through the stand alone program Naccess. It was observed that with respect to the wild type protein 1HK0, the solvent accessible surface area for the proteins 2G98, 1H4A, 4GR7 gradually decreased. This could lead to a decrease in the solubility of the proteins and further

towards aggregation. For the protein 4JGF, the solvent accessibility surface area was calculated to be the least which might predict a higher rate of aggregation of the mutated protein.

Table 7: Solvent Accessibility Surface Area of the native and the mutant forms

S. No	PROTEINS	SASA (Angstorm^2)
1	1HK0 (native)	8659.3
2	1H4A	8583.1
3	4GR7	8517.3
4	2G98	8513.2
5	4JGF	8480.1

4) Protein Interactions analysis

Molecular associations of a protein help in maintaining the functions of the protein. The gamma D protein interactions were calculated through the PIC server (Protein Interactions Calculator) which recognizes various kinds of interactions; such as disulphide bonds, hydrophobic interactions, ionic interactions, hydrogen bonds, aromatic- aromatic interactions, aromatic-sulphur interactions and cation -π interactions within a protein or between proteins in a complex (http://pic.mbu.iisc.ernet.in/ .

The interactions of the residues prone to mutation in the 1HK0 protein were analyzed and the interactions of their specific mutated residues in the mutant proteins were analyzed. It was observed that the interaction pattern changes in every form of the mutant proteins compared to the wild type protein leading towards a deviation from the native functions. This change in interaction pattern might govern the changes associated with the aggregation pattern

Table8: Intramolecular interactions of the native and the mutant forms

Proteins	HI	MM	MS	SS	IONIC	AA	AS	CP	TOTAL
1HK0(WILD TYPE)	126	127	89	76	21	9	5	8	461
4GR7	121	124	87	60	22	8	3	10	435
4JGF	125	122	68	62	25	8	4	5	419
2G98	123	119	68	48	18	8	6	4	394
1H4A	131	142	83	71	22	11	7	4	471

*HI-Hydrophobic Interactions, DB-Disulfide Bridges, MM-main chain-main chain hydrogen bonds, MS- main chain-side chain hydrogen bonds, SS- Side chain-side chain Hydrogen bonds, Ionic interactions, AA- aromatic-aromatic interactions, AS-aromatic-sulphur interactions, CP- cation-pi interactions

A gradual decrease in the number of interactions in the mutant forms of the proteins 2KFB, 2G98 and 4GR7 was observed compared to the wild type protein 1HK0. While for the protein 1H4A the number of interactions is slightly increased compared to the wild type protein.

5) SEQUENCE ANALYSIS

Multiple sequence alignment of the Y D sequences from five species (Rattus, Mus, Canis, Bos and Homo) was carried out to identify the conserved residues. The important amino acids positions have been further verified by doing the sequence analysis. Figure shows the conserved region within the selected sequences. The total number of conserved residues is 132 (76.3%). Most of the important residues (shown in table 2) are conserved and it validates the idea of keeping them at their respective positions in the designed sequence.

```
Rattus          MGKITFYEDRGFQGRHYECSTDHSNLQPYFSRCNSVRVDSGCWMLYEQPNFTGCQYFLRR
Mus             MGKITFYEDRGFQGRHYECSTDHSNLQPYFSRCNSVRVDSGCWMLYEQPNFTGCQYFLRR
Canis           MGKITFYEDRGFQGHHYECSSDHSNLQPYFSRCNSVRVDSGCWMLYEQPNYTGGQYFLRC
Bos             MGKITFYEDRGFQGRHYECSSDHSNLQPYLGRCNSVRVDSGCWMIYEQPNYLGPQYFLRR
Homo            MGKITLYEDRGFQGRHYECSSDHPNLQPYLSRCNSARVDSGCWMLYEQPNYSGLQYFLRR
                *****:********:*****:**.*****:.****.********:*****: * *****

Rattus          GDYPDYQQWMGFSDSVRSCRLIPHAGSHRIRLYEREDYRGQMVEFTEDCPSLQDRFHFNE
Mus             GDYPDYQQWMGFSDSVRSCRLIPHAGSHRIRLYEREEYRGQMIEFTEDCPSLQDRFHFNE
Canis           GDYPDYQQWLGLSDSVRSCRLIPHAGAHRIRLYEREDYRGQMMEFTEDCSSLQDRFHFNE
Bos             GDYPDYQQWMGLNDSVRSCRLIPHAGSHRLRLYEREDYRGQMIEITEDCSSLQDRFHFNE
Homo            GDYADHQQWMGLSDSVRSCRLIPHSGSHRIRLYEREDYRGQMIEFTEDCSCLQDRFRFNE
                ***.*:***:*:.***********:*:**:******:*****:*:****..*****:***

Rattus          IYSLNVLEGCWVLYEMTNYRGRQYLLRPGEYRRYHDWGAMNARVGSLRRVMDFY
Mus             IYSLNVLEGCWVLYDMTNYRGRQYLLRPGEYRRYHDWGAMNARVGSLRRVMDFY
Canis           IYSLHVLEGSWVLYELPNYRGRQYLLRPGDYRRYHDWGGTSARVGSLRRVMDYY
Bos             IHSLNVLEGSWVLYELPNYRGRQYLLRPGEYRRYHDWGAMNAKVGSLRRVIDIY
Homo            IHSLNVLEGSWVLYELSNYRGRQYLLMPGDYRRYQDWGATNARVGSLRRVIDFS
                *:**:****.****::.********* **:****:***. .*:********:*
```

Figure 2: Multiple Sequence Alignment of the selected species

6) SEQUENCE DESIGNING

New sequences were designed which takes up the structure of the gamma D crystallin restraining the functional and the conserved amino acid positions of the native protein. The topological information of the native protein structure will be encoded into the amino acid positions and the sequences will be designed with the constraint of maintaining the functional regions.

The amino acid sequences were randomly generated where the substitution was restricted to the non-conserved region. The rand () function was incorporated and the complete code was written in C programming language. The program takes a sequence as an input and generates random sequences obeying certain constraints, into an output file. The amino acid residues are identified by the program as ASCII characters ranging from 65-90.

The program was further extended to include a scoring matrix i.e. Blosum62 to evaluate the biological significance of generated sequences. As the matrix is simply a table of values assigned to every amino acid substitution to infer whether a particular substitution is tolerable or detrimental for a protein's structure and function, these values were extracted and assigned to score each generated sequence. The greater the score is, the better the biological significance. Further validation was done with the Blast+ program as it also incorporates the same similarity matrix for scoring.

A number of sequences were generated based on the above program out of which sequences with the best score were screened out for further modeling and evaluation.

7) MOLECULAR DYNAMICS STUDY

When new sequences for γD crystallin are generated, one of the challenges is the classification of near native/decoy structures from the native-like protein structures. We have developed a method based on Protein Structure Networks (PSNs) to distinguish native structures from decoys. The network parameters generated for a large number of native structures has been used as positive data set and they have been optimally combined with Support Vector Machines (SVM), to derive a general metric to distinguish decoys from the native protein structures. The model thus generated gives an accuracy of 94.11% when applied on the test datasets. The method [PSN-QA-Rank] also ranks the modelled structures and distinguishes good quality models from bad structures as shown in Figure 1. This methodology will be used to select the designed sequences of γD crystallin, which would yield good quality structures.

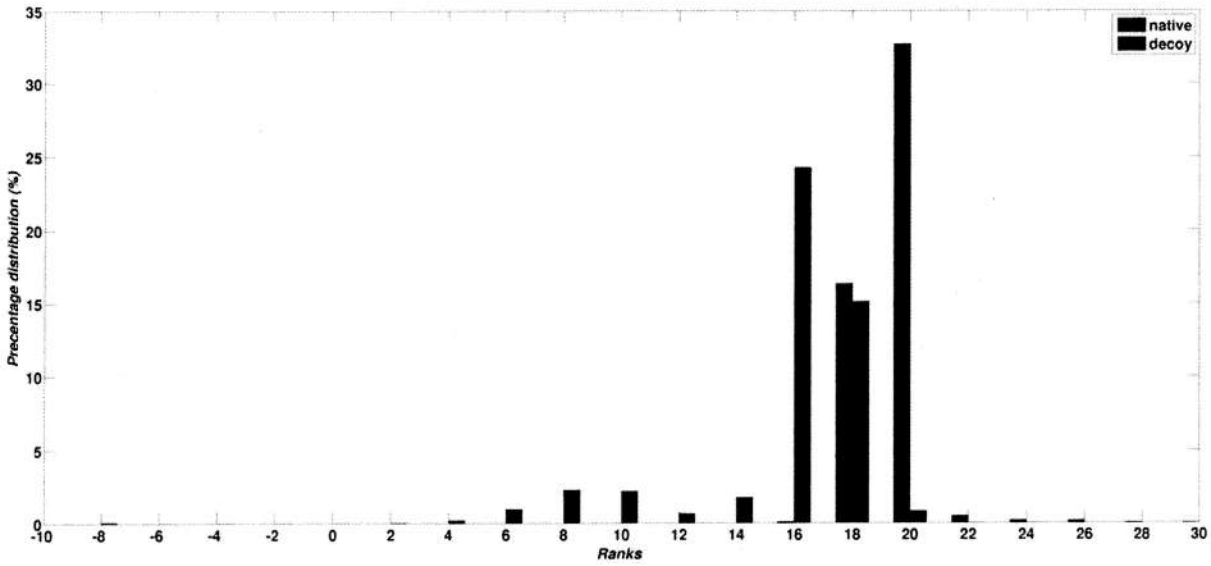


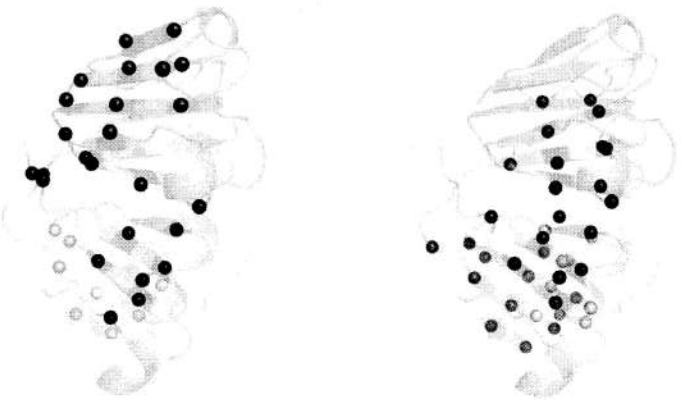
Figure3: The figure shows the percentage distribution of ranks for the 5422 native structures (blue) and 29543 decoy/modelled structures (red) of proteins from different sources. X-axis represents ranks while Y-axis represents the percentage distribution. It is clear that native structures have higher ranks (> 16) as compared to the decoy/modelled structures, a parameter which can be used to evaluate the structures modelled for designed sequences of γD crystallin.

The molecular dynamics simulation trajectory of γD crystallin was obtained and the quality of structures simulated was assessed using PSN-QA-Rank. The rank obtained for a few simulated snapshots are provided in Table 1, which clearly shows that the simulated structures are native-like. However, the differences in the ranks indicate that there are subtle variations in the structures. The difference in the structures are shown [fig 4] by network parameters –hubs and clusters of the starting and the final structures of simulations. These investigations clearly show that the protein-structure-network methodology can be applied to select good quality structures from a pool of structures obtained from designed sequences and also to capture the differences in the structures at detailed molecular level.

Table 9: PSN-QA ranks of the snapshots from simulation trajectory of γD crystallin

Snapshots	(0-9)	(10-19)	(20-25)
1	16.6932	17.0035	16.6815
2	16.9805	16.9425	17.8137
3	17.1035	17.0908	16.9437
4	16.6008	17.3892	16.7432
5	17.0756	17.007	16.8713
6	16.5686	17.4875	16.8704
7	17.4993	17.0259	
8	17.0659	17.1687	
9	16.7048	17.5507	
10	17.2908	17.1087	

Step0: Clusters Step25: Clusters



Step0: Hubs Step25: Hubs

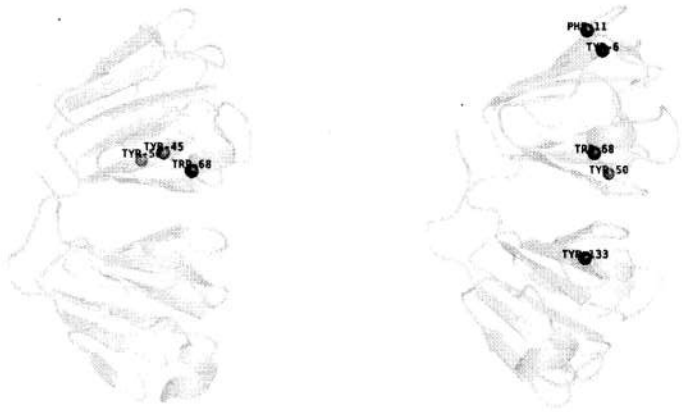


Figure4: The network parameters, Clusters and Hubs shown on the starting and the ending structures of γD crystallin, obtained from MD simulations.

B2. Summary and Conclusions of the Progress

Scientific progress

(1) Literature survey- detailed survey about different type of cataracts, role of crystallin proteins in cataract and importance of γ-crytallin protein has been done. γD- crystallin protein structure and its mutant protein structure have been reviewed to find the important structural changes, which cause the cataract disease

(2) Sequence analysis - amino acid composition of different crystallin proteins has been calculated to find the presence of important amino acids. Multiple sequence alignment analysis gives sequence similarity between different crystallin proteins.

(3) Structural study- the representative structure from each type of crystallin proteins have been analyzed to evaluate the similarity and to find the functionally important region of these proteins. The structural differences between mutant γD- crystalline proteins are important in further analysis.

(4) Topological parameters: The HγD- crystallin structure (1HK0) coordinates has been used to extract the topologically important positions in protein conformation. These positions will be analyzed to generate optimized amino acid sequences.

(5) Molecular Dynamics simulation: A small simulation run of modeled HγD- crystallin structure has been performed for testing the installed software and hardware from the current project. The analysis of the obtained simulation trajectories is under process.

Output

A complete analysis of HγD- crystallin protein at the sequence, structure and evolutionary level has been done. All amino acids of the selected protein have been categorized in structurally and functionally important classes and at the same time it has been shown that the disease causing mutations are a part of these important amino acid residues. It has given a list of positions (110) which can be fixed in the generation of optimized sequence.

A variation in the hydrogen bonding pattern and the intramolecular interactions has been observed in the wild type protein and the mutated form of the proteins.

Also, a large scale simulation of human Gamma D crystallin and a thorough analysis at protein structure network level of the simulation trajectories have been carried out. This has led to the identification of important amino acid residues, pair-wise residue interactions and higher order connectivity like cliques. This, in combination with mutations affecting the structure or function has provided valuable information, which can be used to constrain during the generation of new sequences to the protein. Additionally the simulation and analysis will also throw light on the equilibrium dynamics of the protein, which can serve as a basis to analyze the effect of mutations on the structural dynamics.

A robust method was required to validate the structures modelled from the designed sequences of γD crystallin and also structures generated from molecular dynamics simulations. A general network based validation and ranking scheme [PSN-QA] has been developed in the lab for this purpose. The method has been validated on the simulation snapshots of γD crystallin obtained Network parameters such as hubs and clusters have also been evaluated, which shows subtle differences in the simulated structures.

References:

1. Ageing and Vision: Structure, stability and function of lens crystallins, Hans Bloemendal, Wilfried de Jong, Rainer Jaenicke Nicolette H. Lubsen, Christine Slingsby and Annette Tardieu, Progress in Biophysics and Molecular Biology, 2004, 86, 407-485
2. Mutation analysis of CRYAA, CRYGC, and CRYGD associated with autosomal dominant congenital cataract in Brazilian families, A. Santana, M. Waiswol, E. Saran, JP Cabral de Vasconcellos and M Barbosa de Melo, Molecular Vision, 2009, vol.15, 793-800
3. Novel roles for α-crystallins in retail function and disease, Ram Kannan, P.G Sreekumar and David R. Hinton, Prog Retin Eye Res , 2012, vol.31(6), 576-604
4. High-resolution X-ray crystal structures of human γD-crystallin (1.25A°) and the R58H mutant (1.15A°) associated with aculeiform cataract, A Basak, O Bateman, C Slingsby, A Pande, N Asherie, O Ogun, GB Benedek and J Pande, J. Mol. Biol, 2003, 328, 1137-1147

5. A novel deletion variant of γD-crystallin responsible for congenital nuclear cataract, Li-Yun Zhang, Gary Hin- Fai Yam, Fan DS, Tam PO, Lam DS, Pang CP, Molecular vision, 2007, 13, 2096-2104

6. Structural and aggregation behavior of the human γD-crystallin mutant E107A, associated with congenital nuclear cataract, Venkata Pulla Rao Vendra, Dorairajan Balasubramanian, Molecular Vision 2010, 16, 2822-2828

7. Cataract-associated mutant E107A of human gammaD-crystallin shows increased attraction to alpha-crystallin and enhanced light scattering, Banerjee PR, Pande A, Patrosz J, Thurston GM, Pande J., PNAS, 2011, 108(2), 574-579

8. Contributions of aromatic pairs to the folding and stability of long-lived human γD-crystallin, Fanrong Kong and Jonathan King, Protein Sci., 2011, 20(3), 513–528.

9. The Mutation V42M Distorts the Compact Packing of the Human Gamma-S-Crystallin Molecule, Resulting in Congenital Cataract, Venkata Pulla Rao Vendra, Sushil Chandani, Dorairajan Balasubramanian, PLoS ONE, 2012, 7 (12)

10. Structural analysis of the mutant protein D26G of human γS-crystallin, associated with Coppock cataract, Srinivasu Karri, Ramesh Babu Kasetti, Venkata Pulla Rao Vendra, Sushil Chandani, Dorairajan Balasubramanian, Molecular Vision 2013, 19, 1231-1237

11. Protein sequence design based on the topology of the native state structure, Jha AN, Ananthasuresh GK, Vishveshwara S, J Theor Biol. 2007, 248(1), 81-90

12. A search for energy minimized sequences of proteins, Jha AN, Ananthasuresh GK, Vishveshwara S, PLoS One. 2009, 4(8), e6684

B3. Details of New Leads Obtained, if any: none

B4. Details of Publications & Patents, if any: one publication

1) Ghosh S and Vishveshwara S (2014) Ranking the quality of protein structure models using sidechain based network properties [v1; ref status: indexed, http://f1000r.es/2eu] F1000Research 2014, 3:17 (doi: 10.12688/f1000research.3-17.v1)

Section-C: Details of Grant Utilization

C1. Equipment Acquired or Placed Order with Actual Cost:

(1) MATLAB (software) – Rs 1,71,495.00

(2) Cluster with UPS – Rs 14,95,090.00

(3) AC – Rs 49,800

(4) Computer Workstation and Printer – Rs 2,67,615

C2. Manpower Staffing and Expenditure Details:

(1) Ms Sewali Bora (JRF 30/04/13 to 31/10/13) – Rs 72,400

(2) Mr. Swarup Chakrabarty (JRF 01/06/15 to 31/12/15 – Rs 84,000

(3) Ms Nikita Bora (JRF 18/01/2016 to 30/06/2016) – Rs 1,36,291

C3. Details of Recurring Expenditure:

S No	Recurring head	Amount (Rs in lakhs)
1	Consumables	2.60064
2	Travel	1.26192
3	Contingency	0.5950
4	Overhead	1.42017

C4. Financial Requirements for the Next Year with Justifications: Nil

#Grant utilization details (UC&SE, Assets Certificate & manpower details) also required to be submitted separately as per the prescribed format

[Signature(s) of the Investigator(s)]

Instructions:

(i) *All the information needs to be provided; otherwise the Progress Report will be treated as incomplete. In case of 'Nil' / 'Not Applicable' information, the same may be indicated.*

(ii) *In case of multicentre project, a combined Progress Report should be submitted incorporating the progress of all components. The Project Co-coordinator/ PI will be responsible for this.*

(iii) **Please indicate the reporting period [i.e. Year 1/2/3/4/5].*

(iv) *Submission of Progress Report by the end of the 11th month of grant sanction is linked with further continuation of the project and timely release of funds for the next year.*

Appendix-B

Utilisation Certificate

(for the financial year 2016 -2017)

(Rs. in Lakhs)

1.	Title of the Project/Scheme: **In-Silico design and evaluation of sequences for γD crystallin protein**	
2.	Name of the Organisation:	Tezpur University, Napam, Tezpur, Assam
3.	Principal Investigator:	Dr Anupam Nath Jha
4.	Dept. of Biotechnology sanction order No. & date of sanctioning the project:	**BT/353/NT/TBP/2012 dated January 4, 2013**
5.	Amount brought forward from the previous financial year quoting DBT letter No. & date in which the authority to carry forward the said amount was given:	**2.88335** Lakhs
6.	Amount received from DBT during the financial year (*please give No. and dates of sanction orders showing the amounts paid)*:	**Nil**
7.	Other receipts/interest earned, if any, on the DBT grants (till March 2016):	**0.10092** Lakhs
8.	Total amount that was available for expenditure during the financial year (Sl. Nos. 5,6 and 7):	**2.98427** Lakhs
9.	Actual expenditure (excluding commitments) incurred during the financial year (statement of expenditure is enclosed):	**1.98597** Lakhs
10.	Unspent balance refunded, if any *(Please give details of cheque No. etc.)*:	**NA**
11.	Balance amount available:	**0.99830** Lakhs
12.	Amount allowed to be carried forward to the next financial year vide letter No. & date:	**NA**

Finance Officer
Tezpur University

Dr. Anupam Nath Jha
...sor
...technology
...784028
Dept. of ...
Tezpur University, Tezpur ...

1. Certified that the amount of **Rs 1.98597** mentioned against col. 9 has been utilised on the project/scheme for the purpose for which it was sanctioned and that the balance of **Rs. 0.99830** will be returned.

2. Certified that I have satisfied myself that the conditions on which the grants-in-aid was sanctioned have been duly fulfilled/are being fulfilled and that I have exercised the following checks to see that the money was actually utilised for the purpose for which it was sanctioned.

Kinds of checks exercised:

1. (Cash Book)
2. (Ledgers)
3. (Vouchers)
4. (Bank Statements)
5. Any other

(PROJECT INVESTIGATOR)

(FINANCE OFFICER)

Finance Officer
Tezpur University

Dr. Anupam Nath Jha
Assistant Professor
Dept. of Molecular Biology & Biotechnology
Tezpur University, Tezpur - 784028

(HEAD OF THE INSTITUTE)

Registrar
Tezpur University

(To be countersigned by the DBT Officer-in-charge)

Appendix-C

Statement of Expenditure referred to in para 9 of the Utilisation Certificate

Showing grants received the Department of Biotechnology and the expenditure incurred during the financial year 2016 - 2017

(Rs. in lakhs)

Items	Unspent balance carried forward from previous year	Grants received from the DBT during the year	Other receipts/interest earned if any, on the DBT grants	Total of col. (2+3+4)	Expenditure (Excluding commitments incurred during the year)	Balance (5-6)	Remarks
1	2	3	4	5	6	7	8
1. Non-Recurring (i) Equipments	0.00	Nil	Nil	0.00	0.00	0.00	
Sub Total (I)	0.00	Nil	Nil	0.00	0.00	0.00	
II. Recurring (i) Human Resource	1.87374	Nil	Nil	1.87374	0.97065	0.90309	
(ii) Consumables	0.53294	Nil	Nil	0.53294	0.53358	-0.00064	
(iii) Travel	0.06856	Nil	Nil	0.06856	0.23048	-0.16192	
(iv) Contingency	0.16	Nil	Nil	0.16	0.005	0.15500	
(v) Overheads	0.24811	Nil	Nil	0.24811	0.24626	0.00185	
Sub Total (II)	2.88335	Nil	Nil	2.88335	1.98597	0.89738	
Interest (till March 2016)	Nil	Nil	0.10092	0.10092	Nil	0.10092	
Grand Total (I+II)	**2.88335**	**Nil**	**0.10092**	**2.98427**	**1.98597**	**0.99830**	

(PROJECT INVESTIGATOR)

(FINANCE OFFICER)

Finance Officer
Tezpur University

(HEAD OF THE INSTITUTE)

Registrar
Tezpur University

Consolidated Statement of Expenditure
(for the financial year 2012-2013 to 2016-2017)

Sr No	Sanctioned Heads	Grants received from DBT during the year		Expenditure Incurred						Balance, if any
		2012- 2013	2015-2016	2012- 2013	2013- 2014	2014- 2015	2015- 2016	2016- 2017	Total	
1. (i)	Non-Recurring Equipment's	19.84	0.00	Nil	18.12505	1.71495	0.00	0.00	19.84	0.00
Sub Total (I)		19.84	0.00	Nil	18.12505	1.71495	0.00	0.00	19.84	0.00
2. (i)	Recurring Human Resource	2.11	1.72	Nil	0.724	Nil	1.23226	0.97065	2.92691	0.90309
(ii)	Consumables	1.00	1.60	Nil	1.00	Nil	1.06706	0.53358	2.60064	-0.00064
(iii)	Travel	0.50	0.60	Nil	0.50	Nil	0.53144	0.23048	1.26192	-0.16192
(iv)	Contingency	0.50	0.25	0.13275	0.36725	Nil	0.09	0.005	0.5950	0.1550
(v)	Overheads	1.0	0.20	Nil	0.16174	0.77197	0.24020	0.24626	1.42017	-0.22017
Sub Total (II)		5.11	4.37	Nil	2.75299	0.77197	3.16096	1.98597	8.67189	0.67536
*Interest		0.22202 (March 2015) 0.10092 (March 2016)		Nil	Nil	Nil	Nil	Nil	Nil	0.32294
Grand Total (I+II)		29.64294		0.13275	20.87804	2.48692	3.16096	1.98597	28.64464	0.99830

*Interest of Rs 0.2202 lakhs for the year 2014-2015 have been re-appropriated to overhead

(Project Investigator)

10.8.16

(Finance Officer)

Finance Officer
Tezpur University

(HEAD OF THE INSTITUTE)

Registrar
Tezpur University