African Journal of Biotechnology Vol. 10 (2), pp. 244-252, 10 January, 2011
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB10.619
ISSN 1684–5315 © 2011 Academic Journals

Full Length Research Paper

Bioinformatic analysis of Rp1 gene causing visual disparity in humans

Sana Zahra and Hamid Rashid*

Department of Bioinformatics, Mohammad Ali Jinnah University, Islamabad, Pakistan.

Accepted 5 November, 2010

Retinitis pigmentosa (RP) is a group of inherited diseases that damage rod and cone cells located in human retina. A nonsense mutation R677X has been identified in RP1 gene which not only causes mRNA degradation but also results in truncated protein production leading towards visual disparity in humans. Secondary structure of RP1 gene was determined in order to elucidate the structural changes conferred due to nonsense mutation R677X. The structural differences among non mutated and mutated RP1 gene range from 23 to 43%. Similarly, the truncated protein also resulted in the loss of certain functional as well as active sites which were identified by predicting motifs. A detailed comparison between non mutated and mutated RP1 gene revealed the significance of R677X mutation causing significant structural (helix, turn, sheet and coil) as well as functional loss. Four domains of RP1 gene were predicted using ModWeb and SWISS – MODEL Comparative Modeling Server. The 3D structures of the domains were determined based upon the crystal structure of the homologous templates. The 3D structures were then verified using PROCHECK protein structure validation and verification tool. The quality of the structures obtained was good. This is also useful for future work for annotating the functions of protein using their structures.

Key words: Retinitis pigmentosa, secondary structure, comparative modeling.

INTRODUCTION

Retinitis pigmentosa (RP) is a group of inherited diseases that smash up the light-sensitive rods and cones positioned in the retina. Rods, which are endowed with side (peripheral) and night vision, are exaggerated more than the cones that provide colour and clear central vision. RP first affects the rods in the human eye. As the rods are damaged, vision in low light decreases and peripheral vision constricts (Riazuddin et al., 2005). Further progression of the disease then affects the cones, leading to central vision loss and often leads towards blindness in humans (Liu et al., 2002). Typical symptoms include night blindness, progressive visual field constriction and eventually legal blindness (Daiger et al, 2007).

*Corresponding author. E-mail: drhamid@jinnah.edu.pk. Tel: 111 – 87 – 87 – 87.

Abbreviations: adRP, Autosomal dominant retinitis pigmentosa; **RP,** retinitis pigmentosa.

Approximately one out of 3,000 to 4,000 individuals is affected with RP (Haim 2002). RP can be passed on by all types of inheritance, 20 - 25% is autosomal dominant, 15 to 20% is autosomal recessive and 5 to10% is X linked, while the remaining 45 to 50% is found in patients without any known affected relatives (Guillonneau et al., 1999). RP is most commonly found in isolation, but it can be associated with systemic disease. The most common systemic association is hearing loss (up to 30% of patients). Many of these patients are diagnosed with Usher syndrome. Other systemic conditions also demonstrate retinal changes identical to RP (Riazuddin et al., 2005).

Retinitis pigmentosa can result from mutations in more than 45 genes (Hartong et al., 2006). RP1 is the fourth identified gene causing adRP. The other three include rhodopsin, RSD-peripherin and NRL (Guillonneau et al., 1999). All the introns and exons of RP1 gene were confirmed by sequencing the cDNA products of RT- PCR using human retinal mRNA as a template (Guillonneau et al., 1999; Schwartz et al., 2003). The nonsense mutation

R677X has been identified which causes degradation of mRNA and leads towards the production of truncated protein lacking ~70% of its original length (Guillonneau et al., 1999; Schwartz et al., 2003).

In this study, a detailed analysis of RP1 gene was carried out using different bioinformatics online tools. PSIPRED method (Jones, 1999) was used for predicting secondary structure of Retinitis pigmentosa 1 protein. Secondary structure enabled us to monitor in detail structural responses due to mutations, thus the entire work revealed the impact of missense mutation not only on the structure of the protein encoded by RP1 gene but also on its function.

Motifs, which highlighted the presence of different functional as well as active sites within the RP1 gene were predicted using PROSITE method (Hulo et al., 2004). Domains were predicted using SWISS-MODEL comparative modeling method (Arnold et al., 2006) and 3D models of the domains were built using ModWeb Comparative Modeling server (Pieper et al., 2009). The 3D structures of the domains were determined based upon the crystal structure of the homologous templates. The 3D structures were then verified using PROCHECK protein structure validation and verification tool (Laskowski et al., 1993). The quality of the structures obtained was good. This is also useful for future work for annotating the functions of protein using their structures.

MATERIALS AND METHODS

Retrieval of gene sequence

The amino acid sequence of RP1 (ref: NP_006260.1) gene was obtained from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov). Different online Bioinformatics tools were used for the analysis and structure prediction of RP1 gene.

Secondary structure prediction

Methods predicting protein secondary structure improved substantially in the 1990s through the use of evolutionary information taken from the divergence of proteins in the same structural family. Secondary structure predictions are increasingly becoming the work horse for numerous methods aimed at predicting protein structure and function. PROF, PSIPRED and SSpro are the three very well known methods for predicting secondary structure with the accuracy rate of 77.0, 76.7 and 76.3%, respectively. However, such methods allow monitoring in detail, structural responses to mutations. Further more, if the function/structure of protein A is known and to infer whether B shares this function/structure, a similarity in the local secondary structure may help substantially (Rost, 2001).

In this study, PSIPRED method (Jones, 1999) was used for predicting secondary structure of Retinitis pigmentosa 1 protein and further certain predictions were made in order to elucidate the structural differences due to nonsense mutation R677X.

Motif prediction

Function is anything that occurs to or through a protein. Predicting function of a protein is done using either structure or sequence and it is based upon homology that is, similar sequence similar structure, similar sequences similar function. But still doubts occur for the second fact similar sequences similar function. In this study PROSITE method (Hulo et al., 2004) was used for predicting motifs or patterns for Retinitis pigmentosa 1 protein.

Comparative modelling

Comparative modelling is predicting 3D structure of protein using sequence. Let proteins A and B have homologous 3D structures, the coordinates of protein A are known than the structure of protein B is modelled using protein A.

Three-dimensional (3D) protein structures are of great interest for the rational design of many different types of biological experiments, such as site-directed mutagenesis or structure-based discovery of specific inhibitors. However, the number of structurally characterized proteins is small as compared with the number of known protein sequences. Various computational methods for modeling 3D structures of proteins have been developed to overcome this limitation. Homology modeling has proven to be the method of choice to generate a reliable 3D model of a protein from its amino acid sequence (Arnold et al., 2006).

In this study, SWISS-MODEL (Arnold et al., 2006) and ModWeb Comparative Modeling methods (Pieper et al., 2009) were used for 3D model building. It enabled us to identify not only the two already existing domains of the Retinitis pigmentosa 1 protein but also three new domains. The 3D structures were then verified using PROCHECK protein structure validation and verification tool (Laskowski et al., 1993). Identification of these domains is useful for future work determining the crystal structure of the domains and annotating the functions of proteins using their structures.

RESULTS

Secondary structure prediction

The secondary structure of Retinitis pigmentosa 1 protein was predicted in order to monitor in detail structural responses due to mutation(s).The residual comparison based upon secondary structure prediction enabled us to predict the extent of structural loss due to nonsense mutation R677X. The residual comparison mark the point of structural differences in non mutated and mutated RP1 gene which is further explained in Table 1.

Motif prediction

Motif prediction is useful in function annotation by finding a motif or a pattern in functionally characterized family and searching for the same motif or pattern in a new protein and finally transfer function annotation to the new protein. The nonsense mutation R677X which lead towards the production of truncated protein lacking ~70% of its original length (Guillonneau et al., 1999; Schwartz et al., 2003) also results in the loss of certain motifs in Retinitis pigmentosa 1 protein a comparison of which is

Table 1. Residual comparison of predicted secondary structure of non mutated and mutated RP1 gene.

Amino acid sequence	Position of amino acid	Secondary structure (non mutated)	Secondary structure (mutated)
Coils converting in beta sheet			
111	K	Coil	Sheet
154	R	Coil	Sheet
266,267	HM	Coil	Sheet
279	S	Coil	Sheet
387,388	FS	Coil	Sheet
430,431	VD	Coil	Sheet
470	V	Coil	Sheet
493	E	Coil	Sheet
509	S	Coil	Sheet
613 to 616	THFS	Coil	Sheet
628	E	Coil	Sheet
657	L	Coil	Sheet
Beta sheets converting into coils			
90,91	HS	Sheet	Coil
107	S	Sheet	Coil
131	R	Sheet	Coil
134	S	Sheet	Coil
309 to 311	NLP	Sheet	Coil
345	E	Sheet	Coil
354	S	Sheet	Coil
391	V	Sheet	Coil
436	Q	Sheet	Coil
445 to 447	RFY	Sheet	Coil
463 to 465	VIG	Sheet	Coil
505	C	Sheet	Coil
588	N	Sheet	Coil
Helices converting into coils			
100	D	Helix	Coil
239,240	LP	Helix	Coil
367	M	Helix	Coil
454	L	Helix	Coil
462	S	Helix	Coil
624,625	KN	Helix	Coil
660	N	Helix	Coil
666	S	Helix	Coil
Coils converting into helices			
296,297	PE	Coil	Helix
361-363	NDE	Coil	Helix
380	L	Coil	Helix
428,429	AT	Coil	Helix
432	T	Coil	Helix
491,492	SG	Coil	Helix

Table 1. Contd.

Amino acid sequence	Position of amino acid	Secondary structure (non mutated)	Secondary structure (mutated)
Beta sheets converting into helices			
262,263	KI	Sheet	Helix
412,413	MT	Sheet	Helix
435	I	Sheet	Helix
443,444	KH	Sheet	Helix
486 to 488	SEE	Sheet	Helix
575	I	Sheet	Helix
587	D	Sheet	Helix
590	T	Sheet	Helix
648 to 651	EFAQ	Sheet	Helix
322	S	Helix	Sheet
457 to 461	VRQKK	Helix	Sheet
482	Q	Helix	Sheet
549 to 552	VIEI	Helix	Sheet
Probability Determination			
No. of Coils converting into Beta sheets			12
No. of Beta sheets converting into Coils			14
No. of Helices converting into Coils			8
No. of Coils converting into Helices			6
No. of Helices converting into Beta sheets			4
No. of Beta sheets converting into Helices			9
Probability of Coils converting into Beta sheets			(12/53) = 0.226 ~ 0.23
Probability of Coils converting into Helices			(6/53) = 0.11
Probability of Helices converting into Coils			(8/53) = 0.15
Probability of Helices converting into Beta sheets			(4/53) = 0.075 ~ 0.08
Probability of Beta sheets converting into Coils			(14/53) = 0.26
Probability of Beta sheets converting into Helices			(9/53) = 0.169 ~ 0.17

shown in Table 2.

Comparative modeling

Building a homology model comprises four main steps: (1) Identification of structural template(s), (2) alignment of target sequence and template structure(s), (3) model building and (4) model quality evaluation. These steps can be repeated until a satisfying modeling result is achieved. Each of the four steps require specialized software as well as access to up-to-date protein sequence and structure databases (Arnold et al., 2006).

The two already existing Doublecortin domains were identified using SWISS-MODEL comparative modeling server. The two 3D structures of the doublecortin obtained by SWISS-MODEL comparative modeling server have been reported earlier. The ModWeb Comparative Modeling methods (Pieper et al., 2009) were further used for predicting 3D structures of other domains of RP1 gene. The 3D structures of three new domains were identified and have been shown in Figures 1 - 3.

The predicted 3D models of three new domains (Figures 1–3) have not been reported earlier and hence are the theoretical models of RP1 gene. A brief description of the three new domains of RP1 gene has been shown in Table 3.

The 3D models of the domains were built based upon the templates that show maximum sequence identity with the target region. Hence three models were predicted and their coordinates file was obtained using ModWeb Comparative Modeling methods (Pieper et al., 2009). The models are theoretical models and are not experimentally verified. The quality of the predicted models was further verified using PROCHECK protein structure validation and verification tool (Laskowski et al., 1993). The validation of the models was carried out using Ramachandran plot calculations computed with PROCHECK. The phi and psi distributions of the Ramachandran plots of non-glycine, non-proline residues is summarized in Tables 4 to 6 and

Table 2. Comparison of predicted motifs in mutated and non mutated RP1 gene.

Predicted site	No. of predicted motif (mutated)	No. of predicted motifs (non mutated)
N-glycosylation site	9	28
cAMP- and cGMP-dependent protein kinase phosphorylation site.	4	9
Protein kinase C phosphorylation site.	16	42
Protein Casein kinase II phosphorylation site.	12	52
Tyrosine kinase phosphorylation site.	0	3
N-myristoylation site	5	21
Amidation site	2	3
Cell attachment sequence	0	2
Leucine zipper pattern	0	1
Nucleoside diphosphate kinases active site	0	1

Table 3. Details of the three new domains of RP1 gene.

Domain	Target Region	Template	Template Region	Sequence identity	E- value
I	1624 to 1677	3ec1A	172 to 221	40 %	0.489
II	1553 to 1845	1k90A	394 to 696	22%	0.320
III	624 to 783	3bigA	144 to 322	27%	0.400

Table 4. Ramachandran plot calculations on 3D model of predicted domain I of RP1 gene.

Parameter	No. of residue	%
Most favored regions [A,B,L]	39	84.8
Additionally allowed regions [a,b,l,p]	6	13.0
Generously allowed regions [~a,~b,~l,~p]	0	0.0
Disallowed regions [XX]	1	2.2
Non glycine, non proline residues	46	100
End residues (excl. glycine, proline)	1	-
Glycine residues	4	-
Proline residues	3	-
Total no. of residues	54	-

Table 5. Ramachandran plot calculations on 3D model of predicted domain II of RP1 gene.

Parameter	No. of residue	%
Most favored regions [A,B,L]	220	85.3
Additionally allowed regions [a,b,l,p]	31	12.0
Generously allowed regions [~a,~b,~l,~p]	4	1.6
Disallowed regions [XX]	3	1.2
Non glycine, non proline residues	258	100
End residues (excl. glycine, proline)	2	-
Glycine residues	17	-
Proline residues	16	-
Total no. of residues	293	-

Table 6. Ramachandran plot calculations on 3D model of predicted domain III of RP1 gene.

Parameter	No. of residue	%
Most favored regions [A,B,L]	127	85.8
Additionally allowed regions [a,b,l,p]	16	10.8
Generously allowed regions [~a,~b,~l,~p]	3	2.0
Disallowed regions [XX]	2	1.4
Non glycine, non proline residues	148	100
End residues (excl. glycine, proline)	2	-
Glycine residues	7	-
Proline residues	3	-
Total no. of residues	160	-

Figure 1. Predicted 3D structure of domain I of RP1 gene.

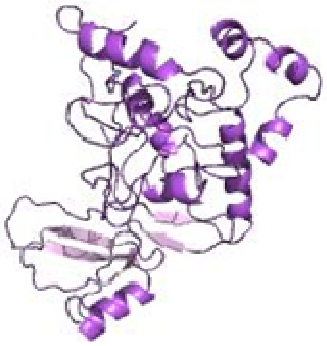

Figure 2. Predicted 3D structure of domain II of RP1 gene.

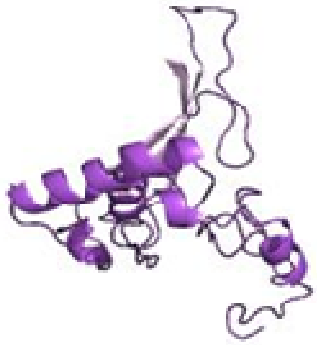

Figure 3. Predicted 3D structure of domain III of RP1 gene.

Figures 4 to 6.

The percentage of residues within the most favoured regions is 84.8, additionally allowed regions is 13.0 and generously allowed regions is 0.0 therefore more than 90% residues lie within the allowed regions and hence they form a good quality model. The percentage of residues within the most favoured regions is 85.3, additionally allowed regions is 12.0 and generously allowed regions is 1.6 therefore more than 90% residues lie within the allowed regions and hence they form a good quality model. The percentage of residues within the most favored regions is 85.8, additionally allowed regions is 10.8 and generously allowed regions is 2.0 therefore more than 90% residues lie within the allowed regions and hence they form a good quality model.

DISCUSSION

The comparison clearly supports the fact that missense mutation R677X causes degradation of mRNA and leads towards the production of truncated protein lacking ~70% of its original length (Guillonneau et al., 1999; Schwartz et al., 2003). There exists a loss of 1518 amino acids which not only reduces the important functional sites within RP1 gene but also results in its improper functioning. The detailed residual comparison showed that 23% coils converted into beta sheets and 11% of them converted to helices. Fifteen percent (15%) helices converted into coils and 8% converted into beta sheets. Twenty six percent (26%) beta sheets converted into coils and 17% into helices. So there exists 34% loss in

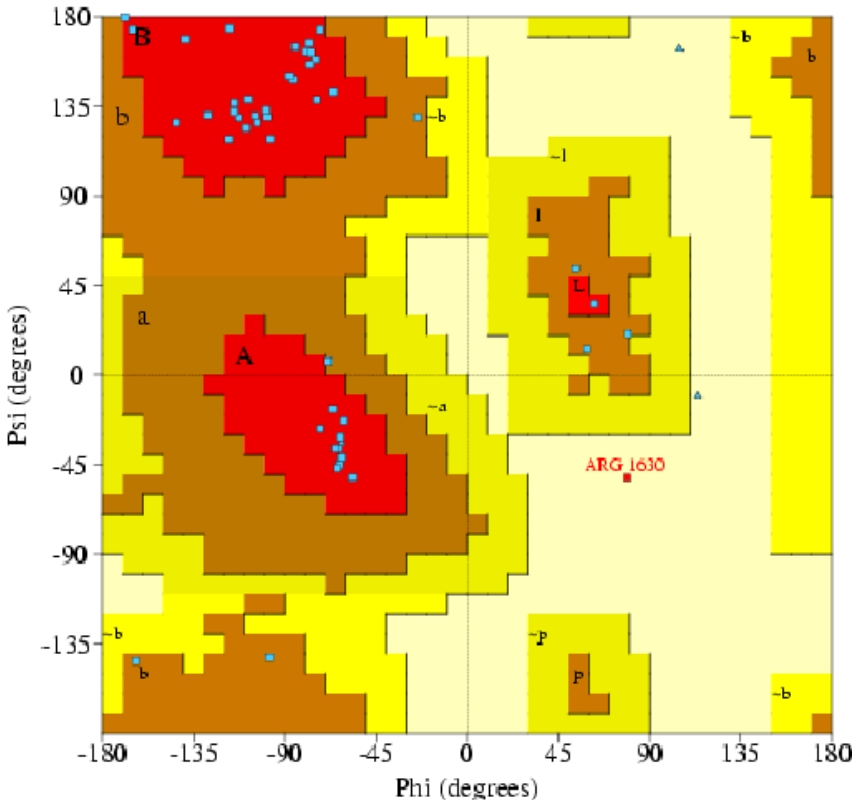

Figure 4. Ramachandran plot for predicted domain I of RP1 gene.

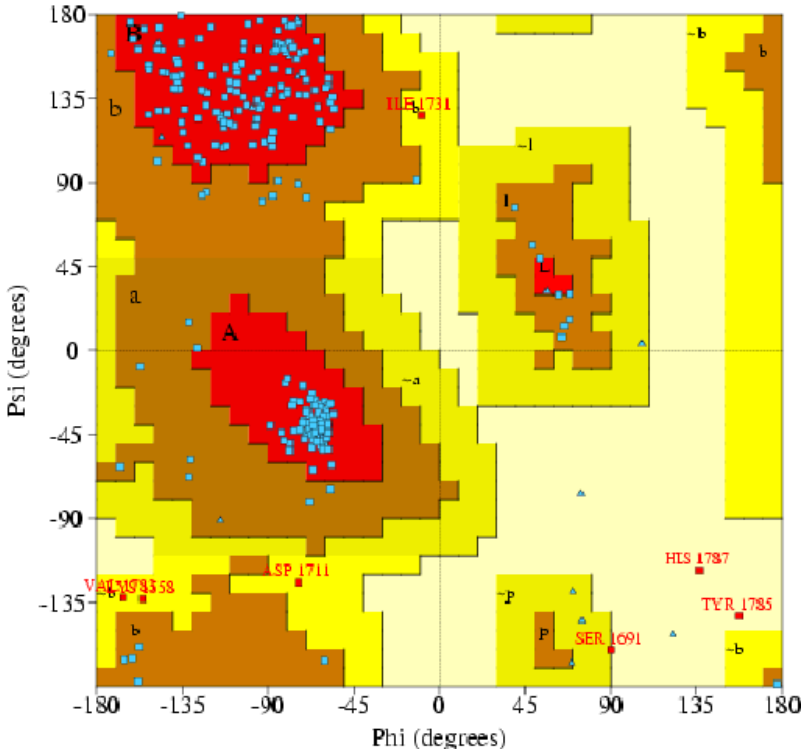

Figure 5. Ramachandran plot for predicted domain II of RP1 gene.

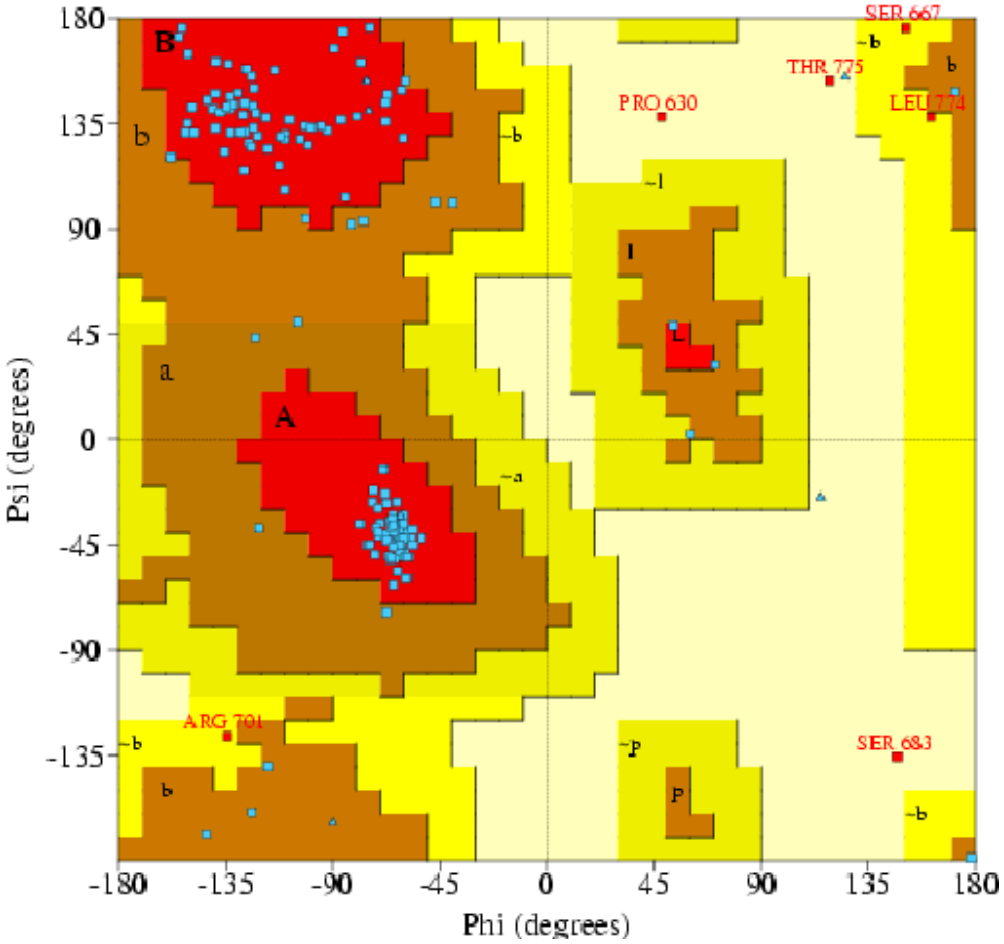

Figure 6. Ramachandran plot for predicted domain III of RP1 gene.

coils, 23% loss in helices and 43% loss in beta sheets in RP1 gene. The nonsense mutation R677X which leads towards the production of truncated protein lacking ~70% of its original length (Guillonneau et al., 1999; Schwartz et al., 2003) also results in structural loss of about 23 to 43% in RP1 gene. Therefore the detailed residual comparison highlighted the structural changes due to nonsense mutation R677X.

The total number of N-glycosylation sites in wild type RP1 gene were 28, cAMP- and cGMP-dependent protein kinase phosphorylation sites 9, protein kinase C phosphorylation sites 42, protein casein kinase II phosphorylation sites 53, tyrosine kinase phosphorylation sites 3, N-myristoylation sites 21, amidation sites 3, cell attachment sequences 2, leucine zipper pattern 1 and nucleoside diphosphate kinases active site 1. The total no. of N-glycosylation sites in mutant type RP1 gene were 9, cAMP- and cGMP-dependent protein kinase phosphorylation sites 4, protein kinase C phosphorylation sites 16, protein casein kinase II phosphorylation sites 12, no tyrosine kinase phosphorylation site, N-myristoylation sites 5, amidation sites 2, no cell attachment sequence, no leucine zipper pattern 1 and no nucleoside diphosphate kinases active site. There exists significant loss in N-glycosylation, cAMP- and cGMP-dependent protein kinase C phosphorylation, protein casein kinase II phosphorylation and N-myristoylation sites where as tyrosine kinase phosphorylation, cell attachment, leucine zipper and nucleoside diphosphate kinase active sites were completely lost due to missense mutation R677X.

The 3D structures of the domains were (Figures 1 – 3) determined based upon the crystal structure of the homologous templates and were then verified using PROCHECK protein structure validation and verification tool (Laskowski et al., 1993). The quality of the structures obtained was good. This is also useful for future work for annotating the functions of protein using their structures.

REFERENCES

Arnold K, Lorenza B, Ju¨ rgen K, Torsten S (2006). The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics, 22: 195-201.

Daiger SP, Bowne SJ, Sullivan LS (2007). Perspective on genes and

mutations causing retinitis pigmentosa. Arch. Ophthalmol. 125: 151-158.
Guillonneau X, Natik IP, Michael D, Christine AK, Artur VC, Samuel GJ, Debora BF (1999). A nonsense mutation in novel gene is associated with retinitis pigmentosa in a family linked to the RP1 locus. Hum. Mol. Genet. 8: 1541-1546.
Haim M (2002). Epidemiology of retinitis pigmentosa in Denmark. Acta Ophthalmol. Scand. 233: 1-34.
Hartong DT, Berson EL, Dryja TP (2006). Retinitis pigmentosa. Lancet, 368: 1795-809.
Hulo N, Christian JAS, Virginie LS, Petra SLG, Lorenza B, Alexandre G, Edouard D CPB, Amos B (2004). Recent improvements to the PROSITE database. Nucleic Acids Res. 32: D134-D137.
Jones DT (1999). Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 292: 195-202.
Laskowski RA, Malcolm WM, David SM, Janet MT (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Cryst. 26: 283-291.
Liu Q, Jie Z, Stephen PD, Debora BF, John RH, Julie ES, Lori SS, Jian Z, Ann HM, Eric AP (2002). Identification and subcellular localization of RP1 protein in humans and mouse. Invest. Ophthalmol. Vis Sci. 43: 22-32.
Pieper U, Narayanan E, Ben MW, David E, Libusha K, David TB, Hannah C, Parminder M, Rachel K, Marc AM, Fred PD, Andrej S (2009). MODBASE, a database of annotated comparative protein structure models and associated resources. Nucleic Acids Res. 37: D347-D354.
Riazuddin SA, Fariha Z, Qingjiong Z, Yuri VS, Zaheeruddin AQ, Tayyab H, Rafael C, Sheikh R, Paul AS, Fielding JH (2005). Autosomal recessive retinitis pigmentosa is associated with mutations in RP1 in three consanguineous Pakistani families. Invest. Ophthalmol. Vis Sci. 46: 2264-2270.
Rost B (2001). Protein secondary structure prediction continues to rise. J. Struct. Biol. 134: 204-18.
Schwartz SB, Tomas SA, Artur VC, Anand S, Samuel GJ, Edwin MS (2003). De Novo Mutation in the RP1 Gene (Arg677ter) Associated with Retinitis Pigmentosa. Invest. Ophthalmol. Vis Sci. 44: 3593-3597.