Int.J.Curr.Microbiol.App.Sci (2016) Special Issue-3: 50-58

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Special Issue-3 (February-2016) pp. 50-58**
Journal homepage: http://www.ijcmas.com

Original Research Article

Bioinformatical tools on identification of deleterious Mutations in OCA2 Gene of Albinism

M. Vijayalakshmi, R. Anuradha and A. M.Thafshila Aafrin*

Research Department of Biochemistry, Sengamala Thayaar Educational Trust Women's College, Sundarakkottai, Mannargudi- 614 016, Tamil Nadu, India
**Corresponding author*

Keywords

OCA2, Missense mutations, Single amino acid polymorphism

A B S T R A C T

A computational study of this work is to identify the most detrimental missense mutations of P protein causing Oculocutanaeous albinism type 2(OCA2). 10/55 missense mutations reported in swissprot, where commonly found less stable, deleterious, damaging and aggregation prone regions by I-Mutant 2.0, SIFT, Polyphen2 and Aggregation programs respectively. Subsequently, the detrimental variants identified, were submitted to the programs Pmut and DisEMBL to find the pathological mutation and discovered regions of a protein respectively. On 10 mutations detrimental mutation, 7 mutation were predicted to be pathological and 2 mutation to be in disordered regions and it is observed that these mutations are highly conserved. Based on our investigation, of was found that two potential candidate variants which are responsible for causing OCA2 disorders.

Introduction

Albinism - Oculocutaneous (Ocular albinism or Hermansky - pudlak syndrome)

Albinism is a recessively inherited disease of disturbed melanin synthesis and/or melanin distribution in the epidermis, scalp and retinal pigment epithelium. In albinism, melanocytes fail to synthesize or distribute melanins properly, which results in oculocutaneous albinism (OCA) which is a group of four autosomal recessive disorders caused by either a complete lack or reduction of melanin biosynthesis in the melanocytes (Markus Preising *et al*., 2011).

Oculocutaneous albinism type 2 (OCA2 or P gene albinism) results from a genetic defect in the P protein that helps the tyrosinase enzyme to function. Individuals with OCA2 make a minimal amount of melanin pigment and can have hair colour ranging from very lightbond to brown (Karen Gronskov *et al.*, 2007). It is located in the chromosome 15q11-q13 (Murray Brilliant *et al.*, 2001).

P protein is involved in the transport of tyrosine,the precursor to melanin synthesis, within the melanocyte. It regulates the pH of melanosome and the melanosome maturation. The P protein has been localized

to the melanosome membrane. Melanosome within the melanocytes have an adenosine triphosphate (ATP-driven proton pump) that helps to generate an acidic lumen. To compensate for the charge of these protons, anions must also be transported to the lumen of the melanosome.

Oculocutaneous albinism is a form of albinism involving the eyes "Oculo, skin-cutaneous", and according to some definitions, the hair as well. Oculocutaneous albinism is a group of conditions that affect coloring (pigmentation) of the skin, hair and eyes ("Oculocutaneous albinism- Genetics Home Reference", March 2007). Long term sun exposure greatly increases the risk of skin damage and skin cancers, including an aggressive form of skin cancer called melanoma, in people with this condition. Oculocutaneous albinism also reduces the pigmentation of colored part of the eye (the iris) and the light sensitive tissue at the back of the eye (the retina). People with this condition usually have vision problems such as reduced sharpness; rapid involuntary eye movements (nystagmus); and increased sensitivity to light (photophobia). Four types of oculocutaneous albinism have been described, all caused by a distruption of melanin synthesis and all autosomal recessive disorders (Gronskov *et al.,* 2007).

Types and Genes Related to Oculocutaneous Albinism

The four types of oculocutaneous albinism are designated as type 1 (OCA1) through type 4 (OCA4). These four types of oculocutaneous albinism each result from mutations in a single gene: TYR, OCA2, TYRP1 or SLC45A2. Changes in the TYR gene cause type 1. OCA1 is caused by an alteration of the tyrosinase gene, and can occur in two variations. The first is OCA1a, and means that the organism cannot develop pigment at all. The hair is usually white (often translucent) and the skin very pale. Vision usually ranges from 20/200 to 20/400. The second is OCA1b, which has several subtypes itself. Some individuals with OCA1b can tan and also develop pigment in the hair (Richard King *et al.*, 2007).

Till date, the structure of wild and mutant P protein and the function by which mutant cause an autosomal recessive OCA2 is not known. Hence, here we present a study for the insilico analysis of P protein variants based on sequence based approaches with the aim of overcoming the associated pathological effects of OCA2 variants.

Materials and Methods

Data Sets

The protein sequence and variants (single amino acid polymorphisms/ missense mutations/ point mutations) for P protein were obtained from swissprot database, available at http://www.expasy.ch/sprot to find out the detrimental point mutants. SwissVar is a portal to search variants in Swissprot entries of the UniProt Knowledge base (UniProtKB), and gives direct access to the swissprot variant pages. The subsection of each swissprot entry provides information on polymorphic variants, in which some polymorphic variants, may be diseases associated with defects in a protein; most of them are nsSNPs (nonsynonyms SNPs) in gene sequence and SAPs (Single Amino acid Polymorphisms) in protein sequence.

Consurf Analysis

The development of consurf, an web based tool for the identification of functionally important regions in proteins by surface mapping of the level of evolutionary

conversation at each amino acid site. We recently developed algorithmic tools for the identification of functionally important regions in proteins of known three dimensional structure by estimating the degree of conversation of the amino acid sites among their close sequence homologues. Projecting the conversation grades onto the molecular surface of these proteins reveals patches of highly conserved (or occasionally highly variable) residues that are often of important biological function. We present a new web server, consurf, which automates these algorithmic tools. Consurf may be used for high throughout characterization of functional regions in proteins. AVAILABILITY: The consurf web server is available at: http://consurf.tau.ac.il. SUPPLEMENTARY INFORMATION: A set of examples is available at http://consurf.tau.ac.il under "GALLERY" (Fabian glasser *et al.,* 2002). The protein sequence is the input. The server identifies the regions of a protein conserved over a wide phylogenetic range very quickly and easily. Consurf 2010: Evolutionary conservation in sequence and structure of proteins and nucleic acids. A web server for calculating Bayesian inference, starting from protein structure and sequence, respectively.

Identification of Detrimental Mutations

Predicting Stability Change by I-Mutant 2.0

The program I-Mutant2.0 available at http://biocomp.unibio.it/chi/predictors/I-Mutant2.0/I-Mutant2.0.cgi was used. The I-Mutant2.0 is a support vector machine (SVM) based tool for the automatic prediction of protein stability changes upon single point mutation; I-Mutant2.0 predictions are performed starting either from the protein structure or more importantly from the protein sequence (Capriotti *et al.,* 2005). This program was trained and tested on a dataset derived from ProTherm (Bava *et al.,* 2004), which is presently the most comprehensive available database of thermodynamic experimental data of free energy changes of protein stability upon mutation under different conditions. The output file shows the predicted free energy change value or sign (DDG), which is calculated from unfolding Gibbs free energy value of the mutated protein minus the unfolded Gibbs free energy value of the native type (kcal/mol). Positive DDG values mean that the mutated protein possesses highly stability and vice versa.

When only the protein sequence is available the required inputs are:

Protein Sequence: The protein sequence in raw format and one letter code;

Position: The position number in the sequence of the residue that undergoes mutation;

Temperature: The temperature value in celcius degrees [0-100];

pH: The negative logarithm value of H^+ concentration [0-14].

Predicting Deleterious Mutations by SIFT

The program SIFT (Ng and Henicoff 2003) was used, which is specifically to detect the deleterious single amino acid polymorphism, available at http://blocks.fhcrc.org/sift/SIFT.html. SIFT is a sequence homology based tool, which presumes that important amino acids will be highly conserved in the protein family. The underlying principle of this program is that, SIFT takes a query sequence and uses multiple alignment information to predict

tolerated and deleterious substitution for every position of the query sequence. SIFT is a multiple step procedure that, given a protein sequence.

Searches for similar sequences,

Chooses closely related sequences that may share similar function,

Obtains the multiple alignment of those chosen sequences, and

Calculates normalized probabilities for all possible substitutions at each position from the alignment.

Substitutions at each position with normalized probabilities less than a chosen cut off are predicted to be tolerated (Ng and Henicoff 2003). The cutoff value in SIFT program is tolerance index of C 0.05. Higher the tolerance index, less functional impact a particular amino acid substitution is likely to have (Ngak-Leng Sim *et al.,* 2012).

Predicting Damaging Mutations by Polyphen2

Analyzing the damaged point mutations at the structural level is considered to be very important to understand activity of the concerned protein. The server polyphen (Ramensky *et al.,* 2002), which is available at http://genetics.bwh.harvard.edu/pph2/ was used for this purpose. We submitted the query in the form of protein sequence with mutational position and two amino acid variants. Sequence based characterization of the substitution site, profile analysis of homologous sequences and mapping of the substitution site to known protein three dimensional structures are the parameters taken into account by the polyphen2 server to calculate the score. It calculates the position specific independent count (PSIC) scores for each of the two variants, and then computes the PSIC scores difference between them. Higher the PSIC score difference, higher is the functional impact a particular amino acid substitution is likely to have (Adzhubei *et al.,* 2010) (Sunyae *et al.,* 1999).

Protein Aggregation by Aggrescan

Aggrescan is based in the aggregation propensity scale for natural amino acids derived from invivo experiments and the assumption that short and specific sequence stretches modulate protein aggregation (Oscar Conchillo-sole *et al.,* 2007). The query is in the form of protein sequence. Aggregation prone segments in the given in the protein sequence will be displayed as output. This shall facilitate the identification of potential therapeutic targets for antidepositional strategies in conformational diseases (bioinf.uab.es/aggrescan/).

Prediction of Pathological Effect of Detrimental Variants by PMut

P-Mut software which is available at http://mmb2.pcb.ubes:8080/PMut/PMut.jsp allows the fast and accurate prediction (80% success rate in humans) of the pathological character of single point amino acidic mutations based on the use of neural networks. Given a mutation happening at a specific location in a protein sequence, P-Mut predicts whether it is pathological (it can lead to disease) or neutral (no effect). The first input to P-Mut is either the sequence of the protein to its swiss prot/ tremble code. The user has to select the mutation site and whether to analyze a single mutation (default) or to perform a complete mutation scan at this position. The input was the protein sequence with all the single point mutation specified. It provides a very simple output: a yes/no answer and a reliability index.

Prediction of Disordered Regions by DisEMBL

Disordered regions in proteins often contain short linear peptide motifs (sif3-ligands and targeting signals) that are important for protein function. DisEMBL, a computational tool for disordered/unstructured regions within a protein sequence available at http://dis.embl.de/ is used. The web interface is fairly straight forward use, the user can paste a sequence or enter the SWISS-PROT/SWALL access or entry code. The probability of disorder is shown graphically and the web server only allows predictions on one sequence at a time. Furthermore, the method is highly accurate, predicting more than 60% of hot loops with less than 2% false positives (Rune Linding *et al.*, 2003).

Results and Discussion

Consurf Analysis

Conservation analysis of the P protein sequence through consurf program classified the amino acid conservation into highly conserved, average and variable conservation and as the position of the amino acid into buried or exposed region.

Identification of Functional Variants by I-Mutant 2.0

Out of 55 variants reported, 45 variants were found to be less stable from the I-Mutant 2.0 server. Among the 45 variants: 6 variants R290G, 1370T, L440H, I473S, V519A, I722T showed a DDG value of >-2.0, 17 variants G27R, P198L, P211L, P241R, V350M, R419Q, L440F, V443I, A481T, R560H, W652R, W679R, L688F, P743L, A773T, Q799H showed >-1.0 DDG value and remaining all showed a DDG value of less than -1.0 (Table 2).

Codes of Amino acids

A – Alanine P – Proline	G – Glycine
R – Arginine S – Serine	H – Histidine
N – Aspargine T – Threonine	I – Isoleucine
D – Aspartic acid W – Tryptophan	L – Leucine
C – Cysteine Y – Thyroxine	K – Lysine
Q – Glutamine V – Valine	M – Methionine
E – Glutamate	F – Phenylalanine

Deleterious single point mutant by SIFT

Protein sequence and its 55 variants were submitted to the SIFT program to check its tolerance index. Among the 55 variants, 37 variants were found to bedeleterious having the tolerance index score of <=0.05.

On deleterious variants obtained, 22 variants: R226W, A334V, M394I, M395L, T404M, L440F, L440H, I473S, N476D, I617L, W652R, E678K, W679C, R720C, A724P, S736L, P743L, G775R, A787V, G795R, Q799H showed a highly deleterious tolerance index score of 0.00.

7 variants: P198L, R305W, A368V, T592I, K614N, H615R, L688F with a tolerance index score of 0.01.

3 variants: V350M, I370T, R419W with a tolerance index score of 0.02.

3 variants: P211L, R290G, N489D with a tolerance index score of 0.03.

2 variants: T387M, V443I with a tolerance index score of 0.04.

31 and 37 deleterious variants according to SIFT were interestingly seen to be less stable by the mutant 2.0 server (Table 3).

Damaged Single Point Mutant by Polyphen2 Server

Protein sequence with mutational position and amino acid variants associated with 55 single point mutants, investigated in this work were submitted as inputs to the polyphen2 server. Polyphen2 server generates two PSIC score difference (Humavar) of 0.5 and above is considered to be damaging. On 55 variants, 41 variants were considered to be damaging by polyphen. Also these 41 variants, exhibited a Humavar score between 0.5-1.0. On 41 variants which were considered damaging by Polyphen, 31 variants were commonly found less stable and deleterious by I-Mutant 2.0 and SIFT server respectively (Table 4).

Aggrescan

On 31detrimental variants identified by I-Mutant 2.0, SIFT and ployphen, 21/31 variants were found in the aggregation prone segments of a protein. They are V350M, I370T, M394I, M395L, R419W, L440H, L440F, V4431, I473S, N489D, W652R, E678K, W679C, W679R, R720C, A787V, G795R, S736L, G775R, Y827H, L688F.

The 21/55 missense mutations reported in swissprot, were commonly found less stable, deleterious, damaging and aggregation prone by I-Mutant 2.0, SIFT, polyphen2 and aggresean programs respectively are italicized with star mark in the table 5.

On 21/55 detrimental mutations, 10 were in the transmembrane regions and 11 were in the topological domain (cytoplasmic and extracellular regions) of a protein. The 11 mutations: V350M, R419W, V443I, I473S, N489D, E678K, W79C, W679R, R720C, A787B, G795R which are in the cytoplasmic and extracellular regions (soluble) are taken into account for further studies.

Pathological Effect of Detrimental Variants by p-mut

Protein sequence and the 11 detrimental mutations were submitted to the p-mut server. P-mut predicted 7 variants to be pathological with the scores between 0.5-1.0. The 7 pathological variants are:R419W, W679R, R720C, 1473S, A787V AND G795R (Table 6).

Disordered Regions by DisEMBL

Protein sequence in FASTA format was submitted as input to DisEMBL. The Hot loops(disorderd) regions were depticted graphically, 23/55 mutational sites were found to be disordered and on 11 mutational sites (which is in the topological regional),only two were found to be disordered (N489D,A787D). The 7/11 detrimental mutations were predicted to be pathological by P-Mut are italicized and the 2/11 disordered regions predicted by DisEMBL are given with star mark in Table 7. Oculocutaneous albinism type 2 (OCA2 or P gene albinism) results from a genetic defect in the P protein that helps the tyrosinase enzyme to function. Individuals with OCA2 make a minimal amount of melanin pigment and can have hair colour ranging from very lightbond to brown (Karen Gronskov *et al.*, 2007). It is located in the chromosome 15q11-q13.

Table.1 List of P Protein and its 55 Variants

R10W	G27R	S86R	C112F	P198L
P211L	P241R	A257V	R266W	R290G
R305W	A334V	A336V	V350M	A368V
I370T	F385I	T387M	M394I	M395L
T404M	R419W	R419Q	L440H	L440F
V443I	M446V	I473S	N476D	A481T
N489D	V519A	H549Q	R560H	T592I
K614E	K614N	H615R	I617L	W652R
E678K	W679C	W679R	L688F	R720C
I722T	A724P	S736L	P743L	A773T
G775R	A787V	G795R	Q799H	Y827H

Table.2 Highly Conserved Amino Acid

Buried (B)/Exposed (E)	Total Number of B/E
B	26
E	12

Table.3 Average Conservation of Amino Acids

Variants of average conservation	Residue variety	Buried (B)/Exposed (E)
I370T	I	B
F385I	F	B
R419W	R	E
R419Q	T	E
T592I	R	E
I722T	I	B

Table.4 Variable Conservation of Amino Acids

Variants of average conservation	Residue variety	Buried (B)/Exposed (E)
R10W	R	E
S86R	S	E
C112F	C	B
P198L	P	E
P211L	P	E
P241R	P	E
A257D	A	E
R266W	R	E
R290G	R	E
R305W	R	E
A336V	A	B

Table.5 Detrimental Mutations

Position	I-Mutant (DDG)	SIFT (Tolerance index)	Polyphen2 (PSIC SD)
*V350M	-1.66	0.02	0.984
*I370T	-3.54	0.02	0.038
*M394I	-0.67	0.00	0.993
*M395L	-0.82	0.00	0.988
*R419W	-0.85	0.02	0.999
*L440H	-2.70	0.00	1.000
*L440F	-1.58	0.00	1.000
*V443I	-1.51	0.04	0.998
*I473S	-2.98	0.00	0.992
*N489D	-0.30	0.03	0.999
*W652R	-1.19	0.00	0.995
*E678K	-0.72	0.00	0.998
*W679C	-1.19	0.00	1.000
*W679R	-1.36	0.00	1.000
*L688F	-1.28	0.01	1.000
*R720C	-0.36	0.00	0.997
*G775R	-0.88	0.00	0.983
*A787V	-0.58	0.00	0.995
*G795R	-0.49	0.00	0.996
*Y827H	-0.58	0.00	0.792

Table.6 Potential Candidate Variants

Position	P-Mut score	Character
V350M	0.3410	Neutral
R419W	**0.9551**	**Pathological**
V443I	0.0576	Neutral
I473S	**0.7095**	**Pathological**
*N489D	0.1789	Neutral
E678K	0.4182	Neutral
W679C	**0.9327**	**Pathological**
W679R	**0.9698**	**Pathological**
R720C	**0.9248**	**Pathological**
***A787V**	**0.5808**	**Pathological**
G795R	**0.7583**	**Pathological**

Table.7 Mutation Summary

Gene region	WT amino acid	Variant amino acid	Amino acid prediction	Nucleotide change	Effect on coding sequence
Exon-13	R	W	419	c.1255 C>T	p.Arg419Trp
Exon-14	I	S	473	c.1418 T>G	p.Ile473Ser
Exon-14	N	D	489	c.1465 A>G	p.Asn489Asp
Exon-19	W	C	679	c.2037 G>C	p.Trp679Cys
Exon-19	W	R	679	c.2035 T>C	p.Trp679Arg
Exon-21	R	C	720	c.2158 C>T	p.Arg720cys
Exon-23	A	V	787	c.2360 C>T	p.Ala787Val
Exon-23	G	R	795	c.2383 G>C	p.Gly795Arg

21/55 missense mutations reported in Swiss-prot, were commonly found less stable, deleterious, damaging and aggregation prone by I-Mutant 2.0, SIFT, Polyphen2 and Aggrescan respectively. On 21/55 detrimental mutations, 10 were in the transmembrane regions and 11 were in the topological domain (cytoplasmic and extracellular regions) of a protein. The 11 mutations which are in the cytoplasmic and extracellular regions (soluble) are taken into account for further studies. 7/11 detrimental mutations were predicted to be pathological by P-Mut and 2/11 were in the disordered regions predicted by DisEMBL.

It can concluded from present study that the potential candidate variants (R419W, I473S, N489D, W679R, R720C, A787V, G795R) for in oculocutaneous albinism type 2 disorders.

References

Abdulla Bava, K. Michael Gromiha, M. Hatsuho Uedaira, Koji Kitajima, and Akinori Sarai thermodynamic database for proteins and mutants doi: 10.1093/nar/gkh082 Nucleic Acids Res. 2004 Jan 1; 32.

Capriotti E[1], Fariselli P, Calabrese R, Casadio R. Predicting protein stability changes from sequences using support vector machines. 2005 Sep 1;21 Suppl 2:ii54-8.

Fabian Glaser1, Tal Pupko2, Inbal Paz1, Rachel E. Bell1, Dalit Bechor-Shental1, Eric Martz3 and Nir Ben-Tal1, ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. Oxford Journals, Science & Mathematics, Bioinformatics, Volume 19 Issue 1,Pp. 163-164.

Karen Grønskov, Jakob Ek and Karen Brondum-Nielsen, Oculocutaneous albinism *Orphanet Journal of Rare Diseases* 2007. 2:43

King, Richard A., *et al*. "Facts about Albinism." *International Albinism Center*, *University of* Minnesota.

Markus N.Preising, Hedwig Forster, Miriam Gonser, Birgit Lorenz, 2011. Screening of TYR, OCA2, GPR143 and MCIR in patients with congenital nystagmus, macular hypoplasia and fundus hypopigmentation indicating albinism. Molecular Vision 17: 939-948.

Ngak-Leng *et al*., 2012. Ocular Manifestations of Albinism, eMedicine. Murray H.Brilliant, (2001). The Mouse p (pink-eyed dilution) and Human p genes.

Oculocutaneous albinism – genetic home reference. 2007. Online mendelian inheritance in man database at Johns Hopkins University.

Oscar Conchillo-Sole *et al.*, 2007. AGGRESCAN: a server prediction and evaluation of “hot spots” of aggregation in polypeptides. BMC Bioinformatics 8-65.

Rune Linding *et al.*, 2003. Protein disorder prediction: implications fot structural proteomics, structure, Vol.11, 1453-1459.

Shamil R.Sunyadev *et al.*, PSIC: profile extraction from sequence alignments with position specific counts of independent observations, Protein Engineering vol.12 no.5 pp.387-394, 1999.