



Insilico Annotation and Prediction of Mutational Hotspots of Optineurin (OPTN), a Novel Approach for Structural Analysis and Peptide Designing

¹Pritam Kumar Panda*, ²Suman Mishra, ³Samaya Chakra Pradhan, ⁴Biswajit Satapathy, ⁵Subhashree Hota, ⁶Piyus Pritam, ⁷Desu Jyothsna

^{1,2,3,4,5,6} BJB Autonomous College, Utkal University, ⁷Bioaxis DNA Research Centre Pvt Ltd

Email ID: *pritamkp15@gmail.co,² sumanmishra2615@gmail.com

Received: March 3rd 2014, Accepted: March 12th 2014, Published: May 1st 2014

Abstract:

One of the major causes of blindness is primary open-angle glaucoma, which affects millions of elderly people worldwide. *Optineurin (OPTN)* has recently been linked to glaucoma, a major cause. Mutations in OPTN have been reported in patients, particularly those with normal pressure glaucoma. The purpose of analysis was to determine the potential regions of Optineurin protein that have maximum probability of mutations based on their structural properties. The identification of disordered regions was performed using *RONN* and *GLOBPLOT*. From the results of the disordered prediction the regions are identified which are further processed for the identification of mutation hotspot regions. The hotspots selected were further screened for the presence of 19 alternative amino acids along with the calculation of the stability changes due to these alterations using *Imutant & Polyphen*. The complete mutational sites annotation and structure prediction was carried out. From the mutational sites determined, various Peptides have been designed using *ARGUSLAB*. Geometry Optimisation was performed to stabilize the designed peptides. Mutations in the receptor Optineurin are associated with the pathogenesis of glaucoma, but the underlying molecular and structural basis is poorly understood, to trace back the structure of this protein Multiple Template Modelling approach was employed using *MODELLER 9v12*. The structure was further validated using *SAVES SERVER* which includes the parameters like *PROCHECK*, *PROVE* and can be further analysed by using other parametric calculations using *RAMPAGE*, *QMean*, *Annolea*, *Gromos Evaluation*.

Keywords: Mutational Analysis, Optineurin(OPTN), modeller9v12, SavesServer, Imutant, Polyphen, Peptide Designing. Swiss Pdb Viewer, RAMPAGE. SWISS MODEL.

Introduction:

Glaucoma is a group of eye diseases causing optic nerve damage. **Glaucoma** is a term describing a group of ocular disorders with multi-factorial etiology united by a clinically characteristic

intraocular pressure-associated optic neuropathy. The optic nerve carries images from the retina, which is the specialized light sensing tissue, to the brain so we can see. In glaucoma, eye pressure plays a role in damaging the delicate nerve fibers of the optic nerve. When a significant number of nerve fibers are damaged, blind spots develop in the field of vision. Once nerve damage and visual loss occur, it is permanent. Most people don't notice these blind areas until much of the optic nerve damage has already occurred. If the entire nerve is destroyed, blindness results.

Glaucoma is a leading cause of blindness in the world, especially in older people. Early detection and treatment by your ophthalmologist are the keys to preventing optic nerve damage and vision loss from glaucoma. This can permanently damage vision in the affected eye(s) and lead to blindness if left untreated. It is normally associated with increased fluid pressure in the eye (aqueous humour). The term "ocular hypertension" is used for people with consistently raised intraocular pressure (IOP) without any associated optic nerve damage. Conversely, the term 'normal tension' or 'low tension' glaucoma is used for those with optic nerve damage and associated visual field loss, but normal or low IOP.

Glaucoma has been called the "**silent thief of sight**" because the loss of vision often occurs gradually over a long period of time, and symptoms only occur when the disease is quite advanced. Once lost, vision cannot normally be recovered, so treatment is aimed at preventing further loss Worldwide.

Materials and methods:

Sequence Retrieval of PTN from GENECARDS:

GENECARDS is a database of human genes that provides genomic, proteomic, transcriptomic, genetic and functional information on all known and predicted human genes. Using the selected Boolean Operator search to study the behaviour of OPTN protein in Optical cells and for the sequence retrieval Uniprot Knowledge based Database has been used.

Sequence Homology using BLASTp:

To find out the homologous sequence of the query protein OPTN Blastp analysis was done and twelve homologous sequences have been selected for the conservation study based upon their similarity and BLOSUM62 Matrix which uses local alignment approach of BLAST Algorithm.

Conservation studies using TEXSHADE and BOXSHADE (SDSC WORKBENCH):To study the conservation pattern of the protein with respect to other organisms TEXSHADE and BOXSHADE tools has been employed that would align the sequences using both local and global alignment respectively. This tool specifically performs multiple sequence alignment in order to analyse the conservation pattern and evolutionary strong sites present within the amino acid sequences.

Domain identification using SMART:

S.M.A.R.T. (Self-Monitoring, Analysis and Reporting Technology) is a monitoring system to detect and report on various indicators of reliability, in the hope of anticipating failures which performs internal Blast with the sequences in the CDD to identify the total number of domains with their associative functions and it can further provide information regarding low sequence identity and signal peptides.

Structural analysis:

The physico-chemical parameters such as the molecular weight, isoelectric point (pI), instability index (Ii), aliphatic index (Ai), and grand average of hydropathicity (GRAVY), -R: No. of negative residues and +R: No. of positive residues were determined using the ProtParam tool of the ExPasy proteomics server. Secondary structure prediction was done using PsiPred Server.

Disorder Analysis using Insilco tools:

Disorder Analysis was done using several tools like DISEMBL and GLOBPLOT which are computational tools for prediction of disordered/unstructured regions within a protein sequence that allows the user to plot the tendency within the query protein for order/globularity and disorder. It successfully identifies inter-domain segments containing linear motifs, and also apparently ordered regions that do not contain any recognised domain.

Regional Order Neural Network software (RONN) will not only specify the regions of mutability but also provides the disorder probability for each residue based upon which the main target site has the maximum mutability.

Analysis effect of Mutations on the stability and Functionality of Proteins:

IMutant and Polyphen were employed to analyze

the effect of the substitutions of wild type amino acid with the new amino acid. One tool would study the functional effect whereas the other is useful to calculate the effect of substitution mutations on the energy and stability of the protein sequence. Based upon the results of the disorder prediction tools mentioned above the regions of mutability were identified followed by the selection of the sites responsible for major mutations. Once the sites are identified the effect of substitution of these sites with any of the other 19 amino acids was studied by the I mutant and PolyPhen tools.

Construction of the peptides using ARGUS LAB and Energy Minimisation:

The selected peptides are constructed in the ARGUS LAB software. Here the peptides are constructed based on the known sequence of the amino acids. The constructed peptides are further made stable by the process of optimization. Once the optimization is performed the energy evaluation is performed using Swiss Pdb Viewer.

Template identification for 3-Dimensional Modelling of Protein:

As per the structural analysis tools, the structural templates obtained showed less than 35% identity with Optineurin Protein. Using BLASTp against PDB Database the templates obtained are **2LO4, 2V4H, 2ZVN, 2ZVO, 3JSV** with less than 35% identity to show structural homologues. From the overall structural analysis and literature study it was confirmed that the structure of Optineurin protein could not be found in the database. Thus the homology modelling approach was used to model its structure. The method used is Multiple Template Modelling. The sequence of the Optineurin protein was subjected to BLASTp against PDB Database using BLOSUM62 matrix with default Threshold value 10. The templates were retrieved from RCSB PDB database based on their identity and query coverage parameters.

3-Dimensional structure prediction using Modeller9v12 and their Validation:

Modeller 9V12 is a Homology Modelling Tool used for Comparative protein structure modelling by satisfaction of Spatial Restraints. The predicted structures generated were selected based on the low molpdf score against DOPE score. Validation of predicted structure was done by using SAVES SERVER which includes PROCHECK. Ramachandran plot analyses were accomplished with aid of RAMPAGE server.

Results and Discussions:

The protein sequence of OPTN Optineurin of Homo Sapiens has been retrieved from the

UNIPROT Database using specific Boolean Operator search and the length of the sequence was found to be 577 amino acids.

Fig1: Conservation Study using TEXSHADE:

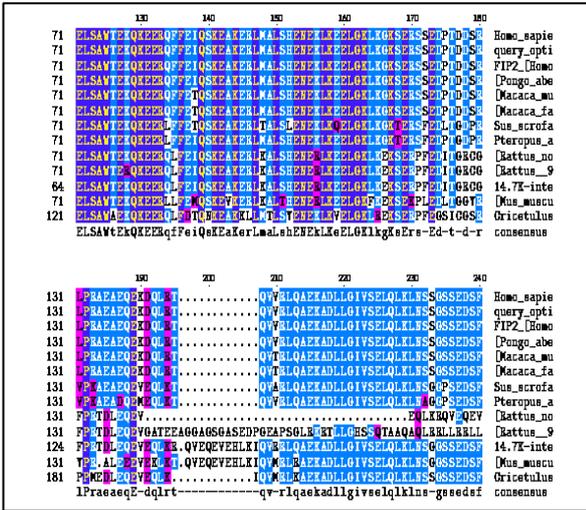


Fig1: The Domain QKEER is conserved in all species with no substitution inferred from the Functional Analysis of the query protein.

Fig2: Conservation Study using BOXSHADE:

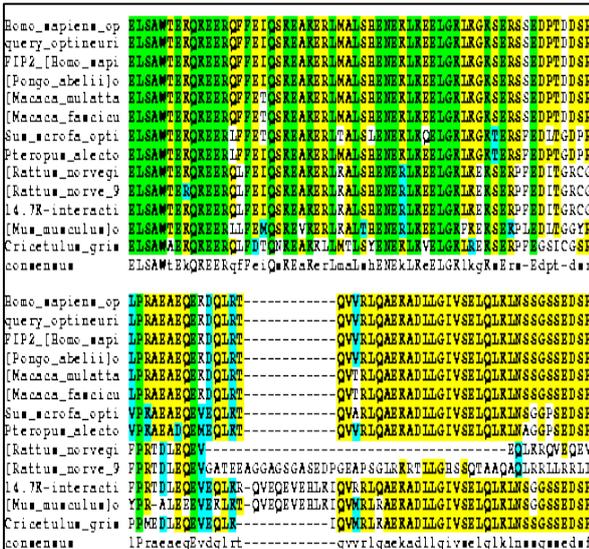


Fig2: The QKEER pattern is conserved throughout all the sequences among various species of organisms considered for mutational study.

Sequences:

Fig4: Chromosomal Location:

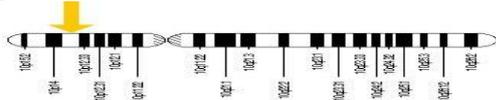


Fig3::Calculation of Identities among the Functional analysis using SMART:

1	Homo_sapiens_optineurin	100.0%
2	query_optineurin_[Homo_sapie	100.0%
3	FIP2_[Homo_sapiens]	99.3%
4	[Pongo_abelii]optineurin	88.7%
5	[Macaca_mulatta]Optic_neurop	96.7%
6	[Macaca_fascicularis]Optic_n	96.9%
7	Sus_scrofa_optineurin	82.9%
8	Pteropus_alecto_Optineurin	85.9%
9	[Rattus_norvegicus]Optn	73.7%
10	[Rattus_norvegicus]optineuri	72.7%
11	14.7K-interacting_protein-2	77.7%
12	[Mus_musculus]optineurin__is	70.8%
13	Cricetulus_griseus]Optineuri	70.7%

Based on the results of smart it can be observed that there is only one functional domain named NEMO extending from 37 to 104 positions in the amino acid sequence. The major function of this NEMO domain is a regulatory protein which is part of the IKK complex along with the catalytic IKKalpha and beta kinases which phosphorylates IkappaB and beta kinases which phosphorylates IkappaB targeting it for degradation which results in the release of NF-kappaB which initiates the inflammatory response, cell proliferation or cell differentiation NEMO activates the IKK complex's activity by associating with the unphosphorylated IKK kinase C termini.

Fig 5: Picture showing the domain region of the Optineurin protein as displayed by SMART tool.



Domain Sequence:-
PEELLQQMKELLTENHQLKEAMKLNNQA
MKGRFEELSAWTEKQKEERQFFEIQSKEA
KERLMALSHEN

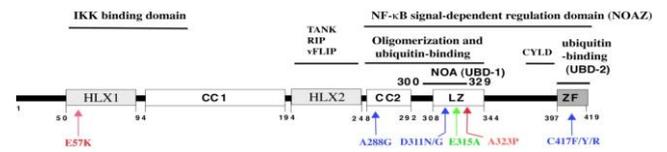
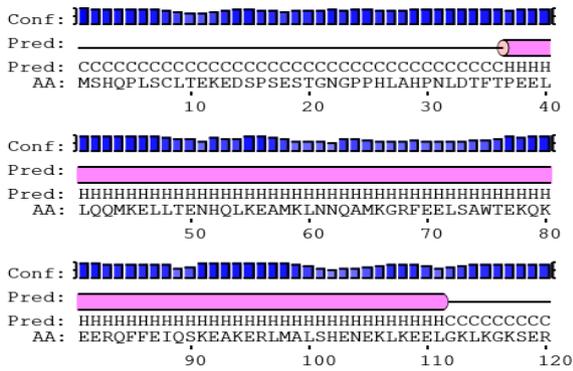


Table 1: Physico-Chemical Parameters:

Number of amino acids	Molecular weight	Theoretical pI	Extinction coefficient	Estimated half-life	Instability index	Aliphatic index	hydropathicity (GRAVY)
577	65922.0	5.12	18825	30 hours	52.16	75.77	-0.924

From the above primary structure analysis it is concluded that the protein is unstable.

Fig 6. Secondary Structure Prediction of OPTN Domain Region using PsiPred Server:



Disorder analysis using DISEMBL, GLOBPLOT and RONN:

From the results of the above mentioned tools the Region 79-83 was found to be common among the regions shown by all the tools and has been selected for further analysis as it has fallen in the conserved domain pattern.

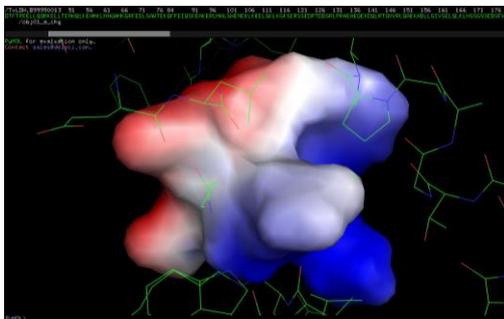
Table 2: Summarized Results of Disorder prediction

TOOLS	POSITIONS
DISEMBL	37-113, 137-167,177-184, 229-265, 297-339,351-501,511-518, 530-537, 567-577
GLOBPLOT	1-32, 57-140, 155-217, 232-306,316-576

Table 4: Effect of substitutions on stability and function Screened results from Imutant and Polyphen.

POSITION	WILDTYPE	NEW	STABILITY(IMUTANT)	EFFECT (POLYPHEN)	RI VALUE
79	Q(Gln)	G(Gly)	Decrease	Deleterious	7
80	K(Lys)	G(Gly)	Decrease	Deleterious	6
	K(Lys)	H(His)	Decrease	Deleterious	6
81	E(Glu)	K(Lys)	Decrease	Deleterious	7
82	E(Glu)	K(Lys)	Decrease	Deleterious	8
83	R(Arg)	G(Gly)	Decrease	Deleterious	7
	R(Arg)	H(His)	Decrease	Deleterious	7

Fig 7. Showing the Mutable site (hot spot site) on the 3D structure of the OPTN protein



RONN	1-33,52-61,68-150,174-200,201-228,253-301,314-354,362-395,396-457,478-495,502-543
------	---

The maximum probability of disorder was found to be in the region QKEER from 79-83 positions.

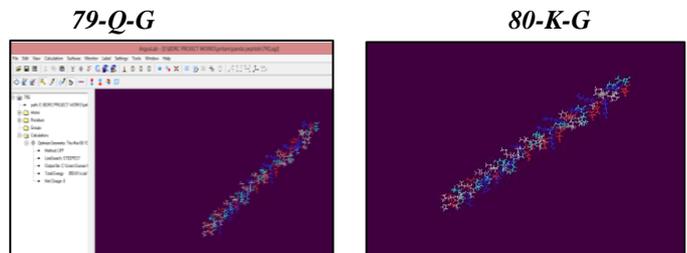
Table 3: Disorder probability for the selected region 79-83.

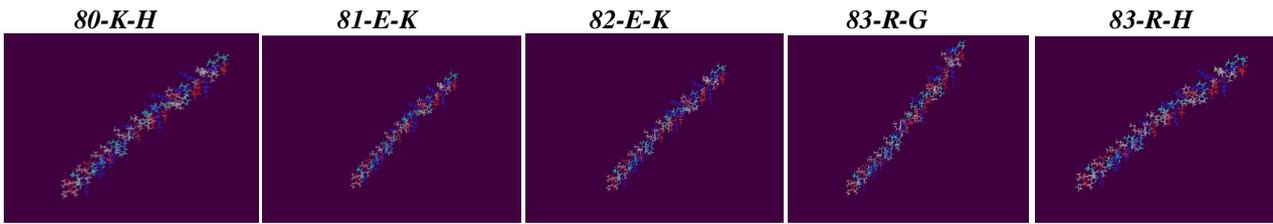
AMINOACID	POSITION	DISORDER PROBABILITY
Q (Gln)	79	0.62
K (Lys)	80	0.63
E (Glu)	81	0.64
E (Glu)	82	0.64
R (Arg)	83	0.64

Effect of Substitution mutations at 79-83 positions on Function and Stability of the protein:

Using the online tools like I Mutant and PolyPhen the effect of the substitutions on the Stability and Functionality has been calculated.

Fig 8: Construction of Mutant and Wild peptides using Argus Lab:





All the alternative peptides with the positions 79-83 replaced by the other amino acids were constructed in Argus lab and the energies were calculated. The energy obtained for the peptide with wild type amino acid is 847.15 kcal/mol

where as all the other amino acids in that position gave a high increase in the energy value. This indicates that the mutant peptides have low stability than the wild type amino acid.

Table 5: Energy Minimisation and Geometry Optimisation

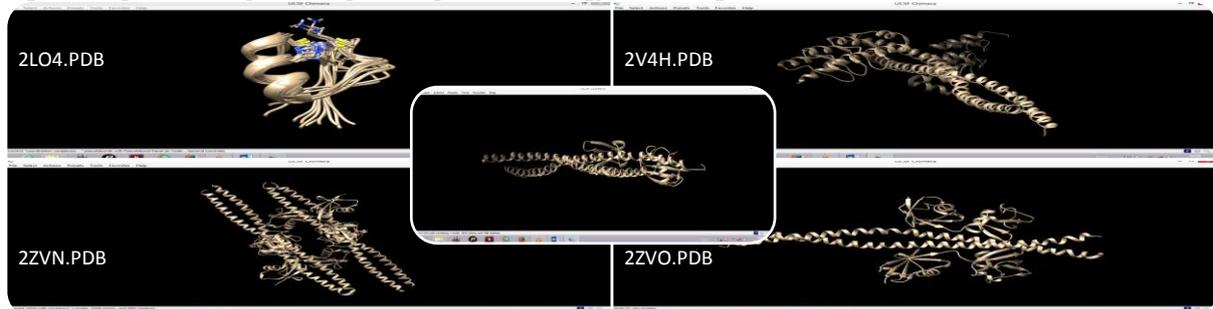
POSITION	AMINOACID	SUBSTITUTED AMINOACID	GEOMETRY OPTIMISATION ENERGY	ENERGY MINIMISATION
Wild peptide	QKEER	Wild peptide	847.15 kcal/mol	-4060.602
79	Q(Gln)	G(Gly)	850.81 kcal/mol	-3769.803
80	K(Lys)	G(Gly)	854.78 kcal/mol	-3723.155
	K(Lys)	H(His)	911.77 kcal/mol	-4116.362
81	E(Glu)	K(Lys)	889.84 kcal/mol	-4276.536
82	E(Glu)	K(Lys)	890.31 kcal/mol	-4294.975
83	R(Arg)	G(Gly)	972.47 kcal/mol	-4341.723
	R(Arg)	H(His)	909.00 kcal/mol	-3759.429

Homology Modelling (Multiple Template Modelling approach):

Based on the given protein sequence, this method aims to predict the protein structure in native

confirmation using spatial restrains parameters using multiple templates.50 models were generated using python environment scripts.

Fig 9: Templates used for Multiple Template Homology Modelling:



TvLDH.B99990013.pdb 4786.60547 -20114.19141

The lowest molpdb score has been considered as the best model among the 50 generated models. The TvLDH.B99990013.pdb was the best model generated having molpdb score is 4786.60547 and DOPE Score is 20114.19141.

Fig 10: Best predicted optineurin model generated using Multiple Template Homology Modelling:

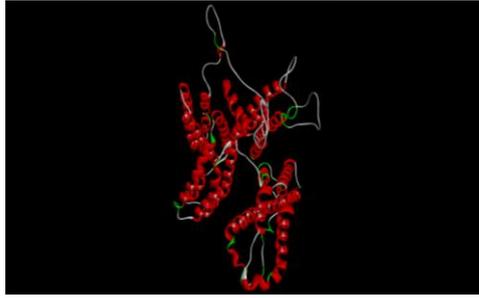
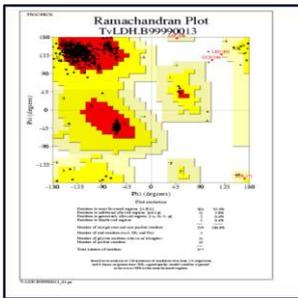
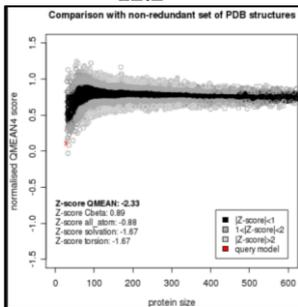


Fig 11: Ramachandran Plot

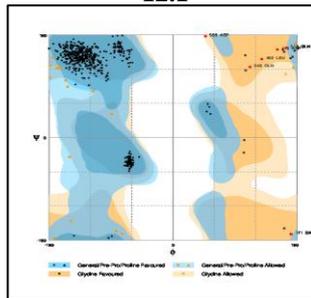


12.2



Ramachandran plot showing Residues in favourable region is 93.9% which is an indication for the accuracy of the structure designed.

Fig 12: Rampage Server 12.1



12.3

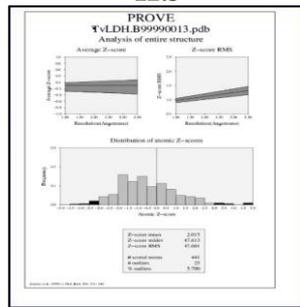
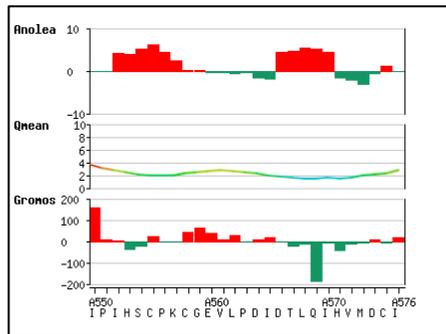


Fig 12: SwissModel Server Analysis Fig 13:PsiPred Evaluation: Fig 14:PROVE



Conclusion:

The current work aimed to characterize and annotate the Optineurin, an important protein in Glaucoma. The Sequence, Functional and Structural characterization was performed using several tools like BLASTp,

SDSC BIOLOGY WORKBENCH. The possible mutational changes along with their effect on the structure and function was evaluated using POLYPHEN2, IMUTANT2.0, RONN, DISEMBL, GlobPlot. A potential hotspot for the probable onset of mutation was identified in the pattern QKEER at the position 79-GlnQ, 80-LysK, 81-GluE, 82-GluE, 83-ArgR in the sequence. The peptide was designed using ARGUS LAB and Geometry Optimization approach was made to stabilize the mutated peptide. As the 3D structure of the protein was not available in the databases Modelling approach was employed to develop the structure of the protein. Modeller9v12 was used for the same. The structure thus designed was further evaluated using Saves server such as PROCHECK and RAMPAGE.

Bibliography:

1. <http://www.ncbi.nlm.nih.gov/>
2. <http://blast.ncbi.nlm.nih.gov/Blast.cgi>
3. <http://www.genecards.org/> Developed at the Crown Human Genome Center, Department of Molecular Genetics, the Weizmann Institute of Science
4. <http://www.uniprot.org/>
5. <http://workbench.sdsc.edu/CPHmodels> 2.0: X3M a Computer Program to Extract 3D Models. O. Lund, M. Nielsen, C. Lundegaard, P. Worning Abstract at the CASP5 conference A102, 2002.
6. <http://bioinf.cs.ucl.ac.uk/psipred/>
7. <http://smart.embl-heidelberg.de/>
8. GlobPlot: exploring protein sequences for globularity and disorder Nucleic Acid Res 2003 - Vol. 31, No.13 (OpenAccess)
9. R. Linding, L.J. Jensen, F. Diella, P. Bork, T.J. Gibson and R.B. Russell Protein disorder prediction: implications for structural proteomics Structure Vol 11, Issue 11, 4 November 2003
10. Yang, Z.R., Thomson, R., McMeil, P. and Esnouf, R.M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins Bioinformatics. 3369-3376
11. <http://salilab.org/modeller/>
12. <http://www.arguslab.com/arguslab.com/ArgusLab.html>
13. <http://nihserver.mbi.ucla.edu/SAVES/>
<http://mordred.bioc.cam.ac.uk/~rapper/rampage.php>
