

---

## Mutation Hotspot Prediction, Structural Analysis & Stability Changes of Prion Protein Involved in Causing CJD

---

Ankita Barik

BJB College, Orissa

E mail: [ankitasuchi258@gmail.com](mailto:ankitasuchi258@gmail.com)

Received: 23<sup>rd</sup> March 2017, Accepted: 30<sup>th</sup> March 2017, Published: 1<sup>st</sup> may 2017

### Abstract

Creutzfeldt-Jakob disease (CJD) is a rare, degenerative, invariably fatal brain disorder. It affects about one person in every one million people per year worldwide. Creutzfeldt-jakob disease is the most frequently seen type of prion diseases. The study on the role of prion protein and its involvement in the occurrence of prion protein is the point to be considered. In the current work the protein structural analysis and mutation prediction has been undertaken for prion protein. Focus was to identify the disorder regions which have maximum probability of mutation which was followed by the effect of mutation on these sites. The work included a series of steps, collection of the protein sequence from NCBI followed by domain analysis using SMART & GLOB-PLOT. STRING is a database of known and predicted protein-protein interaction which is used to identify the networking proteins for prion protein. The disordered/unstructured regions within a protein sequence were identified by DisEMBL. RONN is used for the prediction of protein disorder with the disorder probability predicted for each amino acid. I Mutant and Polyphen were the two major tools used for the study of effect of mutations on the stability and functions of the protein. For primary structure and secondary structure prediction PROTPARAM and SOPMA were applied. The three-dimensional structure of protein sequence was analysed using Phyre 2.

### Keywords

CJD, Prion, SMART, GLOBPLOT, RONN, STRING, PROTPARAM, SOPMA

### Introduction

The term "prion" is derived from proteinacious particle and refers to the pathogen that causes transmissible spongiform encephalopathy (TSEs). This small infectious particle is a disease-causing form a protein called cellular prion protein (PrpC) PRNP (PRION PROTEIN) is the human gene encoding for the major prion protein prp (for prion protein or protease-resistance protein). It is also known as CD230 (cluster of differentiation 230). Prions are also unique and they don't contain nucleic acid, unlike bacteria, fungi, viruses and pathogens. The normal prion protein is thought to be made up of flexible coils referred to as alpha helices, but in the abnormally folded form, these helix are stretched out into densely packed structures called beta sheets.

### Prion discovery:

Mark Rogers in the department of zoology and the Biotechnology center at University College, Dublin: The term "prion" was coined by Stanley B. Prusiner of the University of California, School of medicine at San Francisco in 1982 to distinguish the infectious agent that causes scrapie in sheep, creutzfeldt-jakob disease (CJD) in humans and bovin spongiform encephalopathy (BSE) in cattle from other, more typical infectious agents. The prion hypothesis postulates that these diseases are caused not by a conventional virus or bacterium but by a protein that has adopted an abnormal form. Prusiner won the Nobel Prize in physiology or medicine 1997 for his research into prion.

### Diseases Caused by Prion Protein

There are several diseases of both man and animals whose onset is either directly or indirectly linked to this special protein called prion protein. Some of the diseases caused by it include Scrapie, Bovine Spongiform encephalopathy, Transmissible milk encephalopathy, Chronic wasting disease (CWD), FSE, EUE, CJD etc.

Scrapie was transferred to Scotland in the 1940s. Similar prion diseases are known to affect other animals, e.g., mink, cats, deer and moose. Bovine Spongiform Encephalopathy (BSE) - Mad cow disease is a prion disease that has recently received a great deal of publicity. In England BSE was transmitted to cows through feedstuff supplemented with offals from scrapie-infected sheep. The BSE epidemic first became evident in 1985. Due to the long incubation time the epidemic did not peak until 1992. In this year alone roughly 37,000 animals were affected.

**Creutzfeldt-Jakob disease (CJD)** is a Neurodegenerative Disease that is incurable and invariably fatal "(CJD (Creutzfeldt-Jakob Disease, Classic)". Centers for Disease Control and Prevention. 2008-02-26. Retrieved 2009-06-20.) A new variant of CJD that may have arisen through BSE-transmission. Since 1995 about 20 patients have been identified that exhibit CJD-like symptoms. Psychological symptoms with depression have dominated, but involuntary muscle contractions and difficulties to walk are also common. CJD is caused by an infectious agent called prion. Prions are misfolded proteins that replicate by converting their properly folded counterparts, in their

host, to the same misfolded structure they possess. CJD causes the brain tissue to degenerate rapidly, and as the disease destroys the brain develops and the texture changes to resemble that of a kitchen sponge.

## **Materials and Methods**

### **Sequence Retrieval from NCBI**

The NCBI database has been used in the study to collect the basic information about the prion protein and most important is to retrieve the protein sequence for further analysis. NCBI data base id a free website launched by NLB. It has a direct collection of several genes, proteins, SNP'S, chemicals, Journals etc. It is available at [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)

### **SMART prediction of protein domains**

SMART is a freely available online tool used for the study of the localized regions on the protein called domains. The tool has an inbuilt database containing the information about the protein families and their functional regions. The submission of a query sequences causes a search engine to search for the similar sequence possessing the domains. Than the result is displayed as both a picture and the data. SMART tool is available at <http://smart.embl-heidelberg.de/>

### **STRING**

It is a tool used for collecting the protein network for a specific protein. The network indicates a list of all the other protein that have a direct interaction to the study protein. It would display a network picture showing all the interactions along with the functions and the information. This is used for the study of functional protein association network. This tool is available at [http://string-db.org/cgi/input.pl?UserId=N9knn8qPE2nj&sessionId=PZZKLwsqjEvE&input\\_page\\_show\\_search=on](http://string-db.org/cgi/input.pl?UserId=N9knn8qPE2nj&sessionId=PZZKLwsqjEvE&input_page_show_search=on)

### **GLOBPLOT**

As the name indicates it refers to the tool used for the identification of globular regions (domains) in the protein. Apart from this the tool also identifies the mutation prone regions in the protein sequence. This provides a graphical representation of the domains and the disorder regions. In addition to the graph the tool also provides the complete protein sequence with its disorder prone regions in a different colour. This is a product of EMBL. <http://globplot.embl.de/>

### **DisEMBL**

It is another tool from EMBL. It serves to identify the protein disorders for the submitted sequence. Though it serves the same purpose as Globplot it runs on a different algorithm. The disorder prone regions are highlighted in the submitted sequence and an additional graphical representation is also provided to display the disorder regions along with their disorder probability cut off. The tool is available at <http://dis.embl.de/>

### **RONN**

This is a specialized online tool for the identification of disorder probability of each amino acid of a protein so as to identify the regions of highest disorder probability. The cut off value for the disorder probability for an amino acid is said to be 0.5 above which it is prone to disorder. The tool is freely available at <https://www.strubi.ox.ac.uk/RONN>

### **IMUTANT**

It is also an online tool used to detect the effect of single point mutations on protein stability changes. It works on the neural network based algorithm. This was first developed by Emidio Capriotti, Piero Fariselli, Rita Casadio. Another version of Imutant exists called I mutant 2.0 which is an updated tool that runs on the Support Vector Machines based Predictor for the protein stability changes. Both of the tolls accept the protein sequence along with the required mutation as an input. The tool is available at <http://folding.biofold.org/i-mutant/>

### **POLYPHEN-2**

This too uses the physical and comparative considerations to detect the functional effect on a protein upon single nucleotide change. The prediction lies on the total features of sequence, phylogenetic and structural information that depends on the substitution. For a single change in the amino acid specified the tool, would extract several sequence and structure based properties and feed them to the classifier based on probability.

### **PROTPARAM**

This is a tool used for the characterization of the physical and chemical properties of the protein. This runs on a theoretical calculation. The tool provides the information about the protein molecular weight, stability, isoelectric point, polarity, half-life etc. This is available at <http://web.expasy.org/protparam/>

### **Structure Modelling using ITASSER**

ITASSER is a software that specifically works on modelling approach to develop the structures of un structured proteins, or whose data is unavailable in PDB bank. ITASSER is available at <http://zhanglab.ccmb.med.umich.edu/I-TASSER/>

## **Results and Discussion**

### **Sequence of Prion Protein**

The prion protein sequence has been retrieved from NCBI protein data base. The sequence retrieved is as follows:

```
MANLGCWMLVLFVATWSDLGLCKKRPKPGGWNTGGSRY
PGQGSPGGNRYPPQGGGGWGQPHGGGWGQPHGGGWGQP
HGGGWGQPHGGGWGQGGGTHSQWNKPSKPKTNMKHMA
GAAAAGAVVGGGLGGYVLGSAMSRPIIHFSGSDYEDRYREN
MHRYPNQVYYRPMDEYSNQNNFVHDCVNITIKQHTVTTTT
KGENFTETDVKMMERVVEQMCITQYERESQAYYKRGSSM
VLFSSPPVILLISFLIFLIVG
```

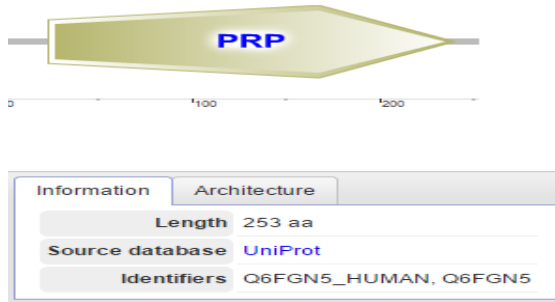
**Inference**

From the above sequence it can be inferred that the sequence is of 253 amino acids, it is a linear peptide with a signal peptide from 1 to 28 amino acids and the domain includes the region from 29 to 253 amino acids.

**SMART for domain identification**

For the given sequence submitted SMART has provided the following result.

**Fig 1: SMRT domain**



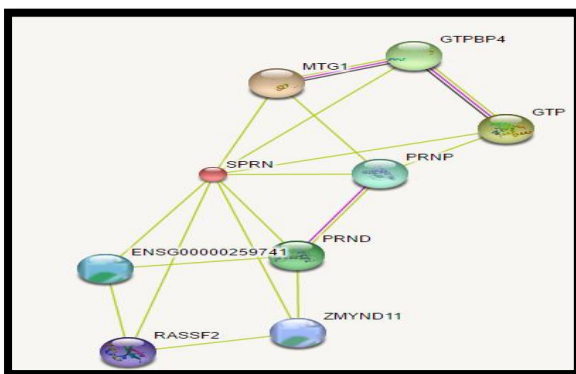
The above result of SMART shows that a total number of one domain named PRP is present in the Prion protein. This domain was extended from 23 to 240 position on the amino acid sequence. The description provided for this protein was as follows

The prion protein is a major component of scrapie-associated fibrils in Creutzfeldt-Jakob disease, kuru, Gerstmann-Straussler syndrome and bovine spongiform encephalopathy.

**STRING**

The following network chart has been obtained from String database for the query protein

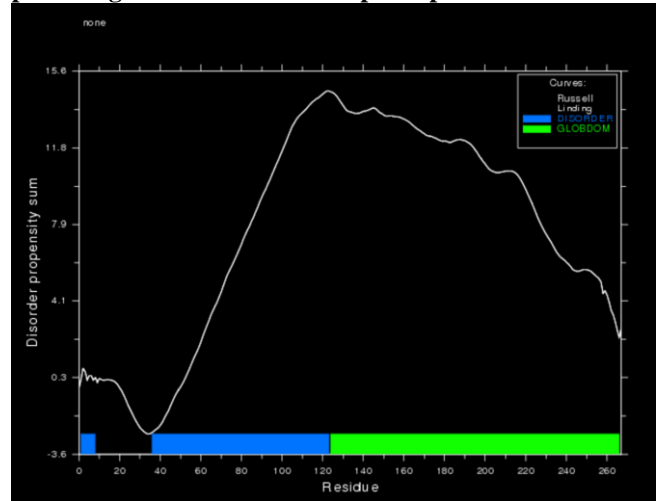
**Fig 2: The network of protein interacting with the Prion Protein**



From the above result of String it can be inferred that the prion protein would interact with several other proteins like SPRN, PRND, RASSF2, MTG1 etc.

**GLOBPLOT for the prediction of domain and mutation regions**

**Fig 3: Shows the globular regions and the disorder prone regions identified in the prion protein**



From the above graph it can be inferred that the disordered regions lie from 1-8 and 36 to 123.

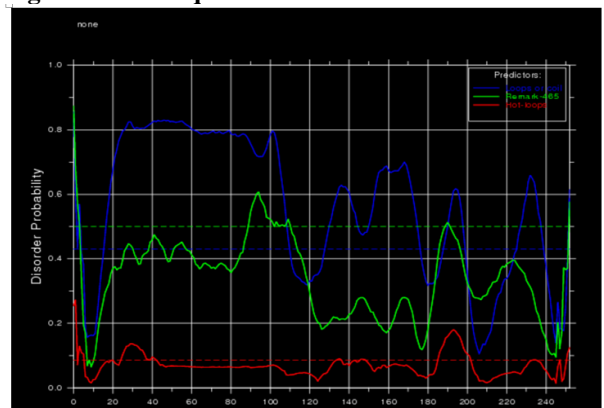
**Fig 4: The disorder region in the sequence:**

```
>none Disorder 1-8, 36-123
AGPRINHMsa piensmanlg cwmlvlvat wsdlgLCKKR PKPGGWN TGG SRYPGQGSPG
GNRYPPQGGG GWGQPHGGG GQPHGGGWGQ PHGGGWGQPH GGGWGQGGT HSQWNKPSKP
KTNmkmaga aaagavvggl ggyvlgsams rpiihfgsdy edryyrenmh rypngvyyrp
mdeysnqnnf vhdvntik ghtvttttkg enftetdvkm mervveqnci tgyeresqay
yqrgssmvlf ssppvilllis flifliivg
```

From the above picture it can be clearly depicted that the disorder region one lies in the signal peptide region and the other lies in the domain region. The amino acids shown in the upper case letters indicate the disordered regions.

**DisEMBL for disorder prediction:**

**Fig 5: DisEMBL graph showing the disordered regions of the sequence:**



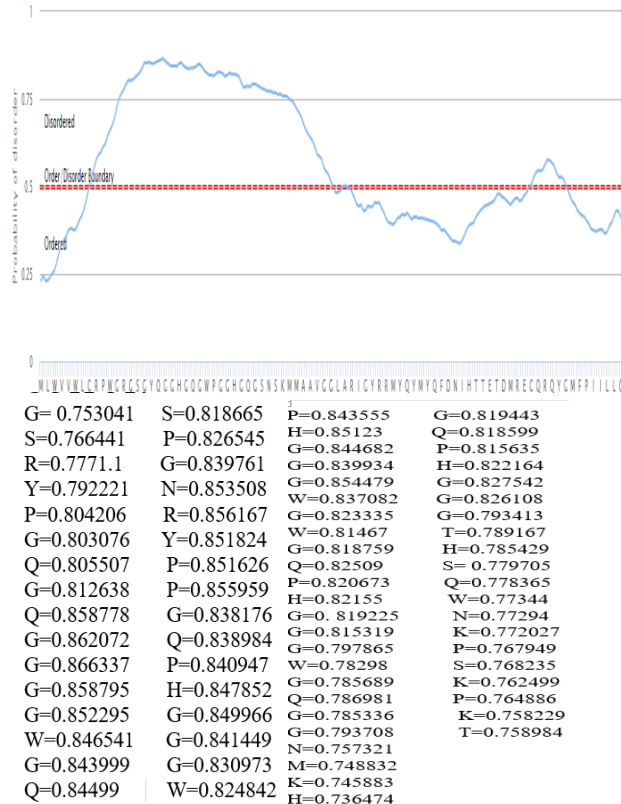
>none LOOPS 17-110, 129-176, 189-199, 227-239  
 manlgcwmlv lfvatwSDLG LCKKRKPPGG WNTGGSRYPG  
 QGSPGGNRYR PQGGGGWGQP HGGGWGQPHG GGWQPHGGG  
 WQPHGGGGWG QGGGTHSQWN KPSKPKTNMK hmaqaaaaaga  
 vvgglggyVL GSAMSRPIIH FGSDYEDRYR RENMHRYPNQ  
 VYYRPMDEYS NQNNFVhdcv nitikghtVT TTTKGENFTe tdkmmervv  
 eqmcitqyer esgayyQRGS SMVLFSSPPv illisflifl ivg

From the above result of it can be identified that the disorder regions identified by DisEMBL were 17-110, 129-176, 189-199, 227-239. The regions of both the tools slightly differ as they both work on a different algorithm and principle.

**RONN (Regional Order Neural Network):**

RONN has been specially used to identify the disorder probability of individual amino acids and identification of the most disordered region or site. The following results have been obtained.

**Fig 6: The disordered regions shown by RONN**



The above data shows the individual probability of disorderness.

**Table 1: Summarizing the above result**

Tool	Regions
GlobPlot	1-8, 36-123
DisEMBL	17-110, 129-176, 189-199, 227-239
Highest probability region RONN	PPQGGGG

**Table 2: The Disorder probability of the final amino acids**

Amino Acid	Position	Disorder Probability
P	50	0.851626
P	51	0.855959
Q	52	0.858778
G	53	0.862072
G	54	0.854479
G	55	0.844682
G	56	0.839934

From the above analysis it can be inferred that the maximum disorder prone region for prion protein is PPQGGGG however the highest mutation showing point is the first G in the series with probability 0.862072 at the position 53.

**Analysis of Mutational Effect on Function and Stability of protein**

**Table 3: Imutant to detect the stability changes of the protein:**

Position	WT	NEW	Stability	RI	pH	T
53	G	V	Increase	7	7.0	25
53	G	L	Increase	4	7.0	25
53	G	I	Increase	6	7.0	25
53	G	M	Increase	5	7.0	25
53	G	F	Increase	4	7.0	25
53	G	W	Increase	5	7.0	25
53	G	Y	Increase	6	7.0	25
53	G	A	Increase	8	7.0	25
53	G	P	Increase	6	7.0	25
53	G	S	Increase	6	7.0	25
53	G	T	Increase	3	7.0	25
53	G	C	Increase	4	7.0	25
53	G	H	Increase	3	7.0	25
53	G	R	Increase	4	7.0	25
53	G	K	Increase	3	7.0	25
53	G	Q	Increase	4	7.0	25
53	G	E	Increase	7	7.0	25
53	G	N	Increase	4	7.0	25
53	G	D	Increase	3	7.0	25

From the I mutant 2 results it can be identified that the mutation at the position 53 replacing the wild allele G with any of the other 19 amino acids causes either the increase or decrease in the stability of the protein. Those that increase the stability can be termed as useful mutation were as the one reducing the stability are pathogenic mutations. The above results shows that the SNP at 53 replacing the any of the amino acid causes only an increase in the stability of the protein, thus it would not be pathogenic as per the stability.

**Polyphen Analysis for functional Annotation of protein**

Polyphen is a tool is used to analyse the effect of the mutations on those sites of the protein which were



selected to be mutational hotspots. The effect of mutation on the 53<sup>rd</sup> position with any of the 19 amino acids is analysed below.

**Table 4: The results of polyphen showing the effect of mutation on functionality**

S. No	Wild	Mutant	Functional Consequence	Score
1	G	V	Probably Damaging	0.999
2	G	L	Probably Damaging	0.999
3	G	I	Probably Damaging	0.999
4	G	M	Probably Damaging	1.000
5	G	F	Probably Damaging	0.999
6	G	A	Probably Damaging	0.990
7	G	Y	Probably Damaging	1.000
8	G	W	Probably Damaging	1.000
9	G	P	Probably Damaging	0.999
10	G	S	Probably Damaging	0.996
11	G	T	Probably Damaging	0.999
12	G	C	Probably Damaging	1.000

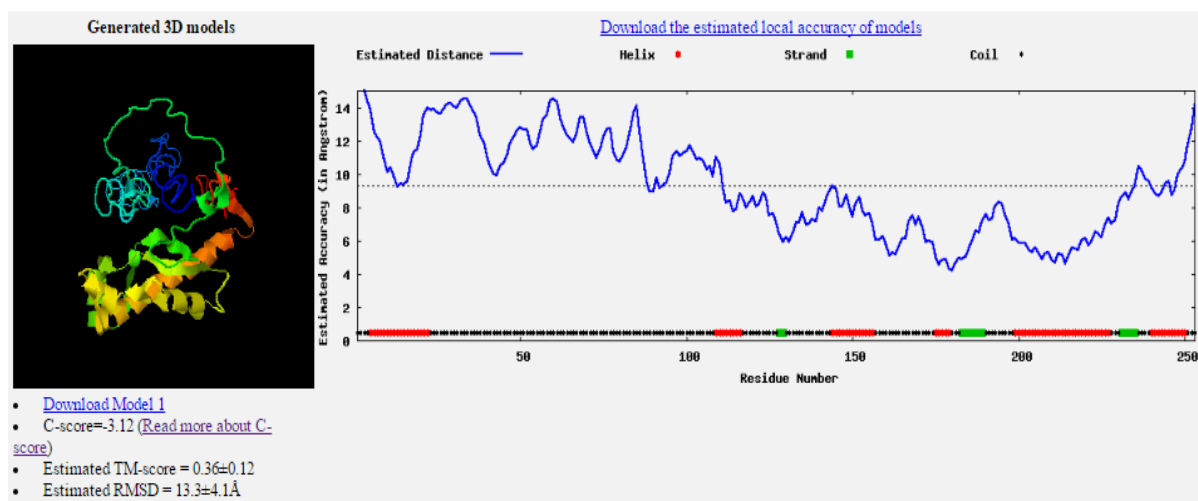
13	G	H	Probably Damaging	1.000
14	G	R	Probably Damaging	0.999
15	G	K	Probably Damaging	0.999
16	G	Q	Probably Damaging	0.999
17	G	E	Probably Damaging	0.999
18	G	N	Probably Damaging	0.999
19	G	D	Probably Damaging	0.999

From the above results of polyphen it can be a very clear indication that any point mutation at the position 53 replacing the original allele G causes a probably damaging effect to the protein. This in turn may be the cause of several diseases.

**Structure prediction of Prion Protein using ITASSER**

ITASSER was used to develop the structure of prion protein as there was no proper structure available in PDB. Five best structures were developed by ITASSER from which 1st being the suitable structure for the protein. Depicted in fig 7.

**Fig 7 showing the Structure 1 generated by ITASSER along with the structural parameters**



**Conclusion:**

The above work was aimed to study in detail the sequence properties and the mutational possibility of the prion protein which is a cause of several diseases. The study included the identification of the most

Mutable site on the protein i.e. the position 53 in the sequence. The effect of mutations to these site were also studied functionally and stability based. Functionally polyphen proved that any mutation to this site causes a probably damaging effect. Thus making it important. Further the 3D structure of the

protein was predicted using ITASSER, which produced 5 different structures out of which model 1 being best was selected and shown above.

**prediction.** Nature Methods, 12: 7-8 (2015).  
([Download the PDF file](#)).

#### References

HL True, SL Lindquist “**A yeast prion provides a mechanism for genetic variation and phenotypic diversity**”, Nature 2000

S Chen, SP Yadav, WK Surewicz , “**S Chen, SP Yadav, WK Surewicz - Journal of Biological Chemistry, 2010 – ASBMB**” Journal of Biological Chemistry, 2010 – ASBMB

N Vassallo, J Herms, “**Cellular prion protein function in copper homeostasis and redox signalling at the synapse**”, Journal of neurochemistry, 2003 - Wiley Online Library

MS Palmer, AJ Dryden, “**Homozygous prion protein genotype predisposes to sporadic Creutzfeldt-Jakob disease**”, Nature, 1991 - search.proquest.com

K Hsiao, Z Meiner, E Kahana, C Cass, “**Mutation of the prion protein in Libyan Jews with Creutzfeldt–Jakob disease**”, England Journal of Medicine, Volume 324

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A.; **Protein Identification and Analysis Tools on the Expasy Server** (In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press (2005). pp. 571-607

**Yang, ZR, Thomson, R, McNeil, P, and Esnouf, RM (2005).** RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins.

Rune Linding,<sup>a</sup> Robert B. Russell, Victor Neduva, and Toby J. Gibson, **GlobPlot: exploring protein sequences for globularity and disorder**, Nucleic Acid Research, 2003

Emidio Capriotti, Piero Fariselli, and Rita Casadio: “**I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure**”, Nucleic Acid Research, 2005

Ivan Adzhubei,<sup>1</sup> Daniel M. Jordan,<sup>1,2</sup> and Shamil R. Sunyaev<sup>1</sup>, **Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2**, Curr Protoc Hum Genet. 2013 Jan; 0 7: Unit7.20.

J Yang, R Yan, A Roy, D Xu, J Poisson, Y Zhang. **The I-TASSER Suite: Protein structure and function**