Analysis of transmembrane and globular protein depending on their solvent energy

Sachin Wakadkar

Master's dissertation

University of Skövde

22-12-2008

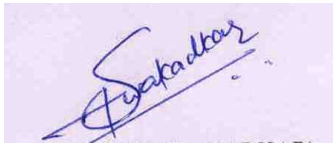
Analysis of transmembrane and globular protein depending on their solvent energy

Sachin Wakadkar

Submitted by Sachin Wakadkar to the University of Skövde as dissertation towards the degree of Master by examination and dissertation in the School of Humanities and Informatics.

22-12-2008

I certify that all material in this thesis which is not my own work has been identified and that no material is included for which a degree has previously been conferred on me.



_____

Sachin Wakadkar

Acknowledgments

This thesis was performed in the School of Humanities and Informatics, University of Skovde under the supervision of Dr. Dan Lundh, Ph D. It has been one of the wonderful experiences of my life. Firstly I would like to thank my supervisor Dan Lundh for his informative and valuable guidance.

I would like to express my sincere gratitude to Bjorn Olsson, Ph D. I would like to thank him for his Kind support. Also I must pay my regards to Kim Laurio, Angelica Lindlof and Jane Synnergren for their informative lectures.

I would also like to thank my best friends like Sameer, Rohit, Anant, Sanjay and Alper for their support. I thank to my family for their encouragement and love. Finally, I would like to thank University of Skovde for accepting me as its part and making my dream true.

Abstract

The number of experimentally determined protein structures in the protein data bank (PDB) is continuously increasing. The common features like; cellular location, function, topology, primary structure, secondary structure, tertiary structure, domains or fold are used to classify them. Therefore, there are various methods available for classification of proteins. In this work we are attempting an additional method for making appropriate classification, i.e. solvent energy. Solvation is one of the most important properties of macromolecules and biological membranes by which they remain stabilized in different environments. The energy required for solvation can be measured in term of solvent energy. Proteins from similar environments are investigated for similar solvent energy. That is, the solvent energy can be used as a measure to analyze and classify proteins.

In this project solvent energy of proteins present in the Protein Data Bank (PDB) was calculated by using Jones' algorithm. The proteins were classified into two classes; transmembrane and globular. The results of statistical analysis showed that the values of solvent energy obtained for two main classes (globular and transmebrane) were from different sets of populations. Thus, by adopting classification based on solvent energy will definitely help for prediction of cellular placement.

Table of Contents

1. Introduction

Proteins are large macromolecules made up by combining the 20 different amino acids to long peptide chains. These amino acids are joining together by peptide bonds. Proteins are made up of single or multiple chains of amino acids. Proteins are essential parts of the living organism and play important roles in every cellular function. Proteins are performing both chemical and mechanical functions in the cell. Many proteins are enzymes and catalyse chemical reactions. Proteins (peptide chains) are folded into unique three-dimensional structures and the functions of the proteins are dependent on their structures. To understand the functions of proteins, it is very essential to understand their structures first (Branden and Tooze, 1998).

Classification of proteins has always been a difficult challenge for the bioinformaticians. The total number of experimentally determined protein structures has reached more than 50,000 entries in the PDB database. However, efficient and accurate classification has not yet been achieved. Probably, classification is good way to understand this huge dataset. The most important task is to classify experimental data into groups or more precisely, partitioning the data into groups in such a way that the data set within the same group is highly similar (sharing a feature) while the data set in a different group is different. Here, classification means grouping of proteins by taking one or more properties into consideration like: cellular location, function, topology, primary structure, secondary structure, tertiary structure, domains or fold. Proteins sharing these properties can be placed under one class. However, the basic problem of classification is not intuitive. Suppose we try to classify proteins depending on four of the above mentioned properties, one protein can share the first two properties with the reference class and another protein can share the next two properties with the reference class. Then it becomes difficult to classify the two proteins in the same class. Another problem is to decide at which level structure classification should be started. Domain is the generally used level. The reason behind this is that the fold pattern of proteins is a deeper level than structural classification. It is more complicated and domains are independent of the evolutionary process. Sub-domain levels can also be used to start the classification (Shen and Chou, 2008). Classification of proteins based on domains as a basic unit is the mostly used in the databases. Various databases have been developed e.g., SCOP (Murzin, 1995) and CATH (Orengo *et al*, 1997), which are widely used. The outcome of the classification of protein structures provides valuable information that can be used to understand protein function and their evolutionary relationships (Kim and Patel, 2006).

Even after experimental structure determination of proteins, it is difficult to classify them on the basis of their structural and functional relationship. A structural similarity could arise because of homology or convergent evolution. Two proteins might not have similar sequences, but could be similar in three dimensional structures or fold and might be performing the same or different kinds of functions in different organisms. It is believed that during evolution three-dimensional structures remain more conserved than sequence itself (Murzin, 1998). It is difficult to say whether different folds share a common ancestor or evolved separately. Events like inserts/deletes, gene duplication, partial

deletion and fusion can lead to a change in the topology of the protein fold (Lupas *et al*, 2001). Under these circumstances it is very difficult to state the exact relationship between protein structures and these proteins are very difficult to classify (Murzin, 1998). These reasons make classification of proteins difficult and sometimes inaccurate.

Proteins could be classified by using other parameters. In this project we have used solvent energy or energy of solvation for the same. Solvation is the process of stabilization of solute molecules by solvent molecules using physical or chemical forces. The term solvation can also be applied to the insoluble molecules, where some functional groups of solutes are stabilized or complexes are stabilized by solvent molecules (Hirata, 2003). Biological membranes and macromolecules are considered to have implicit solvation properties. Implicit solvation is also known as continuum solvation. It is a method of representing solvent as a continuous medium instead of individual explicit molecules. The solvation process can be explained briefly as follows; macromolecules like proteins, lipids and carbohydrates remain in specific orientations with each other. These macromolecules also remain in specific orientations with biological membrane in such a way that its polar parts get chemically or physically bound with polar parts and non-polar parts get bound to the non-polar parts of each other. Solvation is playing very important roles in the various biological processes like protein folding, conformational changes of DNA, RNA and polysaccharides, protein-protein interaction and protein-ligand interaction. The term solvent energy is used to calculate the Gibbs free energy of these various biological processes (Dill, 1990).

Solvent energy of a protein is the energy required for the association of amino acids of the protein with lipid molecules or water molecules. The solvation process is very complex and mainly consists of electrostatic, van der Waals, hydrogen-bonding interactions (Baumeister and Cordes, 2004). There are basically two methods for calculation of energy of solvation; one is based on ASA (Accessible Surface Area) and the other is based on electrostatics models. In this work we have used the ASA-based Jones' algorithm (Jones *et al*, 1992). ASA is the area of the solute, which is accessible to the solvent. In the ASA methods, a linear relationship is calculated between the surface area of the solute and the Gibbs free energy of transfer. The solvent energy of the solute molecule can be calculated by equation 1:

$$\Delta G_{sol} = \sum \sigma \times ASA \qquad (1)$$

where, σ is the solvation parameter of the solute molecule, that is a contribution to the free energy of solvation of the particular solute molecule per surface unit area, and ASA is the solvent accessible surface area of the solute molecule.

There are basically two types of proteins found in the living organism, transmembrane proteins and globular proteins. The sub-cellular location of transmembrane and globular proteins is shown in Figure 1. Transmembrane proteins span the biological membrane completely. They are lipophilic in nature and present in lipophilic environments. Globular proteins are present in globelike structures. They are hydrophilic in nature and also present in aqueous environments. A solvation effect from their respective

environments stabilizes both types of proteins. The environment of proteins plays a key role in their behavior and function.

Transmembrane proteins are surrounded by lipid bilayers and their solvation is much more complicated than solvation of globular proteins because they are solvated by lipids as well as by water. A hydrophobic single chain protein is easily inserted and solvated into a lipid membrane. The surface of the multiple chain proteins (complex) is rough and contains pockets of different dimensions. For efficient solvation of a transmembrane protein complex, it is necessary for lipids to enter in the pockets of the protein complex. The extent of insertion of lipid chains is depending on the physiological condition, which allows proteins to change their conformation (Carney *et al*, 2007). Not only lipophilic peptides are solvated by membrane, but small polar peptides can also be easily solvate. Buried water molecules are playing important role in their solvation. However larger polar peptides are difficult to solvate, they are solvated by a shielding effect in which polar peptides are surrounded by non-polar peptides in helical form.

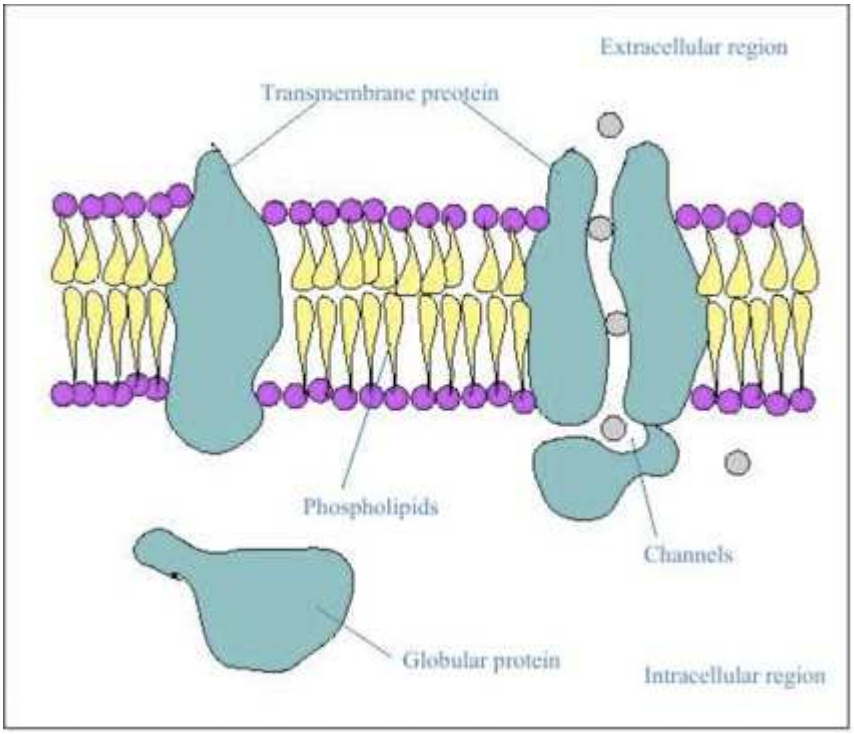


Figure 1: Placement of transmembrane and globular protein in phospholipids bilayer and cytoplasm respectively.

Globular proteins are solvated by water molecules. Water molecules can be found on the surface and bulk of the globular proteins. Globular proteins are exposed to the aqueous environment. Most of the amino acids present in globular protein are polar in nature and can be easily solvated by the water molecules. If any non-polar chain is present in the globular protein, it is protected by polar chains by surrounding it from outside. The information of solvation properties of peptides has been suggested as a good resource to study protein folding (Lindahl and Johansson, 2008).

In this thesis we are trying to classify proteins into the classes of transmembrane and globular proteins depending on the solvent energy. Transmembrane and globular proteins have in general opposite solvent energy. Transmembrane proteins usually have positive solvent energy and globular proteins usually have negative solvent energy. The implication of this is that the cellular placement with respect to solvent energy may indicate if the protein is transmembrane or globular.

In this thesis we also address the environment of proteins and placements of the polypeptide chains in the protein complex. Transmembrane proteins may consist of one or several polypeptide chains. If only one polypeptide chain is present in the protein, it's here proposed that the placement could be predicted based on solvent energy. If several polypeptide chains are present, it is here proposed that the outer (surface) chains' placement could be predicted on the basis of solvent energy and that the inner chains (protected chains) should be similar to the globular chains. In this way a more exact location of the polypeptide chain could be predicted. Likewise, globular proteins can also consist of one or several polypeptide chains. If only one polypeptide chain is present in the protein, its solvent energy should be negative. If several polypeptide chains are present in the globular protein, it is here suggested that the placement could be predicted on the basis of solvent energy. The negative solvent energy will indicate the outer surface of the globular protein.

For example, 1Q90 is a Cytochrome b6f (Cyt b6f) complex. It is a transmembrane protein. It contains 9 polypeptide chains, chain A, B, C, D, G, L, M, N and R. In the structure of Cyt b6f, chain N and M are completely protected and chain A and D are partially protected by other chains, see figure 2 and figure 3.

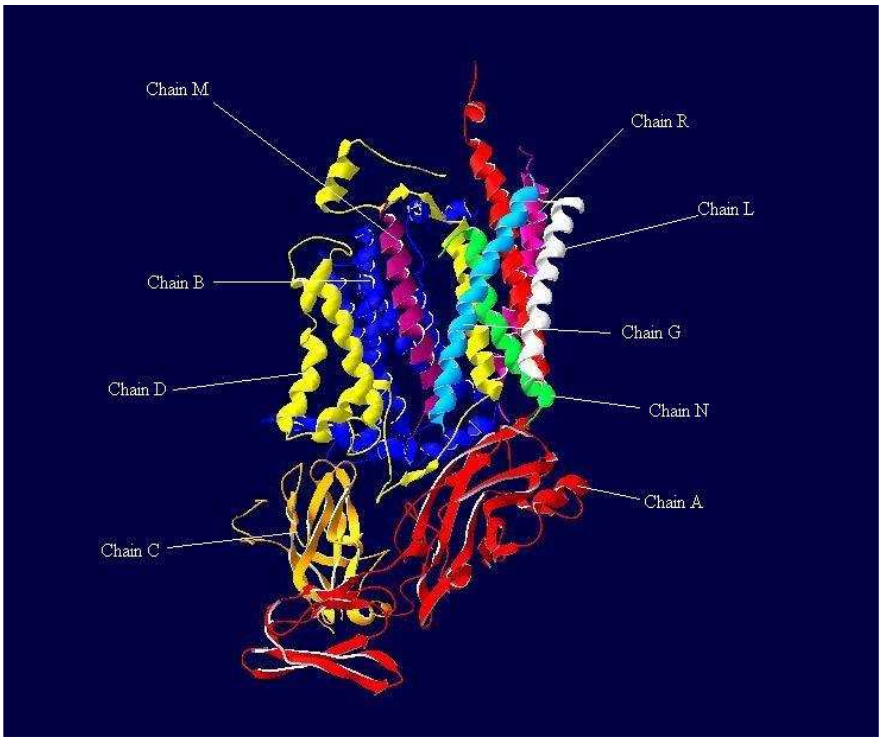


Figure 2: Side view of protein 1Q90. It shows 9 chains like chain A, B, C, D, G, L, M, N and R. All chains are shown in different colors, chain A-red, chain B-blue, chain C-magmata, chain D-yellow, chain L-white, chain M-purple, chain N-green and chain R-pink. (The image was generated by Swiss pdb viewer (spdbv) by using 1Q90.pdb file)

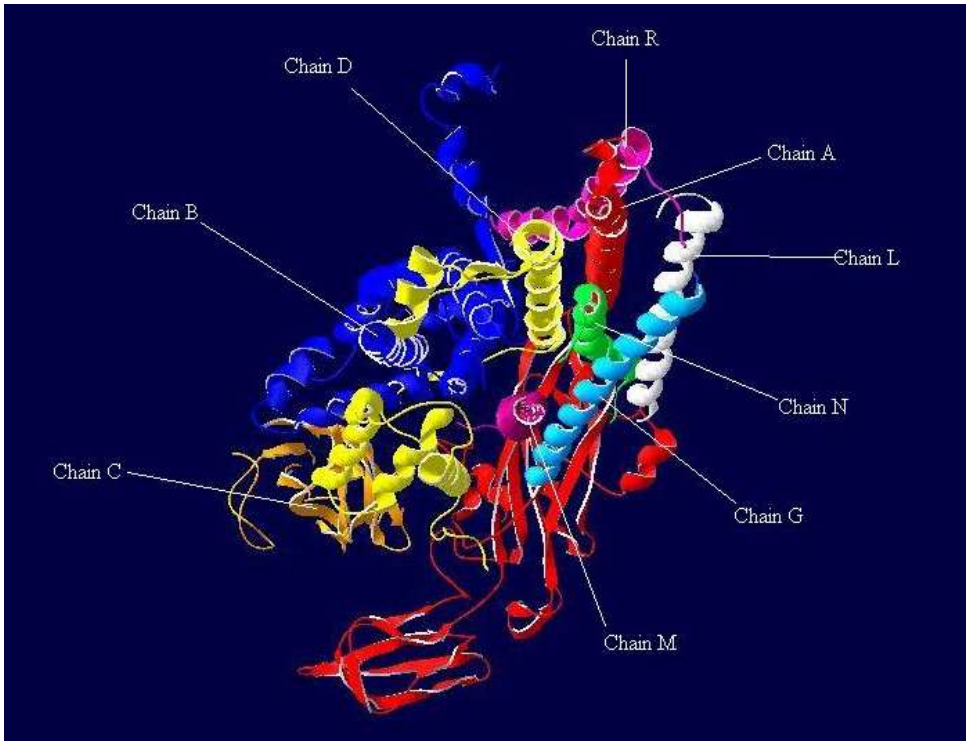


Figure 3: Top view of 1Q90. Chains N and M are completely protected and chains A and D are partially protected. (The image was generated by spdbv by using 1Q90.pdb file)

In this work all proteins in the Protein Data Bank, PDB (RSCB, 2006) were tentatively classified into two main classes, transmembrane and globular depending on their cellular location. The solvent energy of the all proteins was calculated by using Jones' algorithm (Jones *et al*, 1992). This algorithm is working on the basis of solvent accessible surface area. Using the solvent energy for the all proteins, an analysis was made to confirm the difference, if any, present between two classes. In order to check the significance of the difference between the two classes, a statistical test was performed for the values obtained for each class. The unpaired T test was selected for this purpose based on two reasons. Firstly; it was assumed that the two datasets were unpaired and having Gaussian distributions. Secondly, the two datasets were different in sample size. As the unpaired T test was the most appropriate test we have not tried any other statistical test. There are two types of unpaired T tests; one is based on equal variance and the second is based on unequal variance. In order to carry out an unpaired T test it was necessary to calculate the variance. Variance was calculated by using a F test. Based on equal or unequal variance, the respective unpaired T test was performed. However, it would be better to try other statistical tests for this work. The statistical analysis confirmed that a significant difference is present between the two classes. The classification of proteins was justified from the results of the statistical analysis. However, there were some exceptional cases in

each class. Some transmembrane proteins have negative solvent energy and some globular proteins have positive solvent energy. All exceptional cases were further investigated for possible reasons by taking solvent energy of the individual chains into consideration. Exceptions were classified into a third 'erroneous' class. Probable reasons for proteins in the erroneous class were investigated and outlined, see discussion chapter.

2. Background

As mentioned in the introduction, solvation is playing a crucial role in the stabilizing three-dimensional structures of the proteins. Scientists have used solvent energy to check various aspects of the proteins and it is explained below.

2.1 Solvent energy to evaluate folds

Liisa Holm and Chris Sander used solvent energy to identify correct folds amongst the incorrect folds (Holm and Sander, 1992). As the stability of the globular proteins depends on the interaction of solvent molecules, they used a solvent contact model to evolve atomic solvation from known protein structures in the database. Holm and Sander (1992) used solvent preference to discriminate between correct and incorrect three-dimensional structures of proteins for a given sequence for which the backbone was taken from proteins with known or hypothetical models. They also used solvent energy to identify the correct placement of the chains in the protein. They tested the capability of solvation preference to recognize correct sequence-structure pairs on misfolded models that were generated in three ways; first, by keeping a native fold static and shifting a sequence along the structure. Second, by keeping a native fold static and replacing the sequence by an unrelated sequence and vice versa. Third, by giving a sequence and secondary structure that generated an alternative packing of helices and sheets. In all cases solvation preference successfully identified the correct structure for a given sequence or the correct sequence placement among possible alternatives.

2.2 Solvent energy for hypothetical structure prediction

Themis Lazaridis and Martin Karplus used solvation in the hypothetical structure prediction (Lazaridis and Karplus, 1998). They used an energy function in which a Gaussian model for solvation energy was combined with the CHARMM (Chemistry at HARvard Macromolecular Mechanics) vacuum potential for the discrimination of the native models from the misfolded protein models. CHARMM is a versatile program that is used for molecular simulation. It can be applied to many particles systems. It is used to study biological macromolecules like lipids, proteins, carbohydrates and nucleic acids as they occur in the membrane environment, solution or crystal (CHARMM, 2008). Lazaridis and Karplus have calculated solvent energy by using equation 2.

$$\Delta G^{solv} = \sum_i \Delta G_i^{solv} ; \Delta G_i^{solv} = G_i^{ref} - \sum_j f_i(f_{ij}) V_j \quad (2)$$

where,

G_i^{ref} : Solvation free energy of group *i* in reference compound,

V_j : The volume of group j,

$f_i(f_{ij})V_j$: The solvation free energy density of the group i at distance r_{ij}.

They found that the solvation model had a significant advantage over other models that did not involve the calculation of accessible surface area.

2.3 Solvent energy in protein folding

Solvent energy is also used in the prediction of protein folding for the evolution of the models. First Jones *et al* (1992) started the application of solvent energy in protein fold recognition (in 1992), where he used solvent energy to evaluate folds threaded with an unknown sequence. In a further development Jones *et al* (1992) used solvent energy in the software application GenTHREADER (McGuffin and Jones, 2003), which is new and fast method for fold recognition and in its first step it is using a traditional algorithm for sequence alignment. In its second step it performs calculation of pair potentials and solvent energy. The solvation energy for each residue is calculated by using:

$$\Delta E_{solv}^{a}(r) = -RT \ln \frac{f^{a}(r)}{f(r)} \quad (3)$$

where,

E_{solv}^{a}: The solvation potential of *a*, r is the degree of residue burial. Residue burial is describes a protein residue's exposure to the solvent and neighboring atom.

$f^{a}(r)$: The frequency of occurrence of residue *a* with burial *r*

$f(a)$: The frequency of occurrence of all residues with burial *r*.

R*:* The universal gas constant (R = 8.314472(15)$JK^{-1}mol^{-1}$).

T: The absolute temperature

In the third step all models are evaluated by a neural network (David Jones, 1998). In this way Jones' algorithm was evolved, however the solvent energy term in this algorithm has always looked the same. For this project the Jones' solvent energy algorithm (program) was obtained from Dr. Dan Lundh. This version of the algorithm was implemented in C. The implemented algorithm was crosschecked and validated by calculating the solvent energy of protein by using implemented algorithm, Threader (3.5) (Jones *et al*, 1992) and GenTHREDER (McGuffin and Jones, 2003). Their outputs were compared with each other and it was found that output were similar. In this way implemented algorithm was

evaluated on more than 100 protein cases and for all cases output were similar. The program was crosschecked and validated against a local implementation of Threader (3.5) (Jones *et al*, 1992) and GenTHREDER (McGuffin and Jones, 2003).

2.4 Solvent energy in protein interaction

Jackson and Sternberg (1995) studied the effect of the solvent on the interaction of proteins. They studied the proteins that were interacting with each other in presence of high dielectric medium (polar solvent). For this purpose they used a Poisson-Boltzman equation, which is generally used for the approximation of solvent effect on macromolecular structure and interaction. In their study Jackson and Sternberg calculated electrostatic energy of protein-protein interaction in three different components: a) the change in solvent energy of the protein after binding, b) the change in solvent energy of the inhibitory protein after binding and c) interaction between these two proteins in presence of solvent. That is, they considered solvation energy of bound and unbound conformations of proteins for correct prediction of interaction and final conformation of associated proteins (Jackson and Sternberg, 1995). Their study showed that minor fluctuations in the atomic structure only affect desolvation and hydrophobicity marginally. However, the effect of both desolvation and hydrophobicity was much more contributing than the electrostatic potential in protein-protein interaction.

2.5 Solvent energy in peptide structure prediction

Scott *et al* (2008) used solvation potential and a rotamer library, dependent on the backbone, for the prediction of peptide structures. To predict the structure of the protein Scott *et al* used three different versions of a genetic algorithm with different force fields. The first contained only a *Van der waals* term. In the second version they added a electrostatic potential. In the third a force field was used in which solvation was added to the *Van der waals* term and the electrostatic potential. Scott *et al* used a method based on solvent accessible surface area to calculate the solvation energy. Their study concluded that the performance of the genetic algorithm was improved after inclusion of solvation potential and that the solvation potential was very important for efficient predictions.

All different methods explained above must be evaluated against Jones' algorithm for better performance. However, there are no specific tests available that can compare Jones' algorithm with other methods. One can see outlines of comparisons as far as the Threader algorithm is concerned of which solvent energy is part (McGuffin and Jones, 2003).

The efficiency of our method of classification of transmembrane and globular proteins can be confirmed from a hydropathy plot. A hydrophobic plot is a graph in which the hydrophobicity of the amino acids is plotted against their respective positions in the polypeptide. Hydrophobicity plots can be used in the discrimination of transmembrane and globular proteins. A hydrophobicity plot is usually used to find regions in a polypeptide that span the biological membrane. Hydrophobicity plots are also used for

the identification of the different domains of the proteins (Hydrophathy plot, 2008). Tools for predictions of transmembrane topology, e.g. TopPred, use this method. A hydrophobicity pattern can be used to identify transmembrane regions. If the hydrophobicity plot shows transmembrane regions one can expect the protein to be a transmembrane protein, likewise if there are no transmembrane regions its likely that it is not a transmembrane protein, i.e. it is likely a globular protein. Transmembrane proteins are suggested to have high solvent energy and globular supposed to have low solvent energy, implying that the solvent energy are indirectly coupled with hydrophobicity plot, which can be very useful while determining the cellular placement of regions in a protein.

3. Material & Methods

The solvent energy of all proteins present in PDB was calculated by using Jones' algorithm (Jones *et al*, 1992). PDB is protein database that contains three-dimensional structural data of proteins as well as DNA that was obtained by X-ray crystallography and NMR spectroscopy. Previously, Jones' algorithm was evaluated on smaller databases. Jones' algorithm was evaluated on 102 (Jones *et al*, 1992) and 18 (Jones *at al*, 1995) proteins separately. McGuffin, and Jones (2003) also evaluated Jones' algorithm on 2727 proteins.

In this project the solvent energy of each chain in the protein was calculated separately. The total number of proteins was 51,447; out of which 57 showed infinite solvent energy due to containing artificial amino acids. Those chains were eliminated from the list. The remaining 51,390 were classified into two classes; transmembrane and globular. Transmembrane proteins were 1554 in numbers and the rest were globular proteins. Transmembrane proteins were further aberrantly classified into TM alpha (392), TM alpha 1 (300), TM alpha 2 (224), TM alpha 3 (524), TM alpha buried (19) and TM beta (95). These classifications were made randomly.

After classification of transmembrane proteins into subclasses, a significance test was used to identify possible differences between transmembrane subclasses. As the sample size of each class was different, an unpaired T test was used. For simplicity, two subclasses were selected each time for comparison and all subclasses were compared with each other. In order to perform the unpaired T test a calculation of the variance of the two samples was necessary. An F-test was used to calculate the variance for each sample. Based on the variance calculation a "T-Test: Two-Sample Assuming Unequal Variances" was selected. The confidence degrees chosen for this study were 95%, 99% and 99.9%. In our study a comparison was also made between globular and transmembarne proteins. The globular proteins were divided into subclasses containing 500 proteins in each subclass. All subclasses of globular proteins were compared with all subclasses of transmembrane protein by applying the unpaired T test. The main aim of this comparison was to investigate if these two classes were different from each other with respect to solvent energy. Furthermore, by dividing the transmembrane class into subclasses following topology and membrane association an analysis might reveal

significant or non-significant differences for a particular subclass

Exceptional cases of transmembrane (transmembrane protein with negative solvent energy) and globular (globular protein with positive solvent energy) were found and further classified on the basis of their sub-cellular location. These exceptional cases, of both classes, were investigated from PDB (http://www.rcsb.org/pdb/home/home.do) and PDBsum (http://www.ebi.ac.uk/pdbsum/). Each exceptional protein was screened for function and sub-cellular location by searching it into PDB and PDBsum database manually. Proteins related to the exceptional protein were also observed to identify the possible reasons for the unusual value of the solvent energy.

4. Result and Discussion

As mentioned in material and methods, transmembrane proteins were classified into subclasses like TM-alpha, TM-alpha 1, TM-alpha 2, TM-alpha 3, TM-alpha buried and TM beta and these classes were compared with each other by using an unpaired T test. Statistical analysis was performed to check the significance of the difference between them. The results of the statistical analysis are shown in Table 1.

	TM-Alpha 1	TM-Alpha 2	TM-Alpha 3	TM-Alpha Buried	TM-Beta
TM-Alpha	***	***	**	***	
TM-Alpha 1		***	***	***	***
TM-Alpha 2			***	***	***
TM-Alpha 3				***	
TM-Alpha Buried					***

Table 1: The significance if the difference between subclasses of transmembrane proteins: TM Alpha, TM-Alpha 1, TM-Alpha 2, TM-Alpha 3, TM-Alpha Buried and TM-Beta. A single asterisk (*) denotes a P value less than 0.05, two asterisks (**) denotes a $P < 0.01$, and three asterisks (***) denotes $P < 0.001$).

From the result it can be concluded that classification of transmembrane proteins into subclasses was justified because there was a significant difference between all pairs of classes except two pairs; one is TM-Alpha and TM-Beta and second is TM-Alpha 3 and TM-Beta. No significant difference between them suggesting that they might be similar in nature.

Similarly to the transmembrane proteins, globular proteins were divided into subclasses in which each subclass contains 500 globular proteins. These subclasses were derived randomly. Each subclass was compared with all subclasses of transmembrane proteins by

using the unpaired T test. The results of statistical analysis are shown in Appendix 1.

It can be seen in Appendix 1, that all subclasses of the globular proteins were significantly different from subclasses of the transmembrane proteins (excluding buried transmembrane alpha helical proteins). The statistical analysis showed that the values of solvent energy obtained for two main classes (globular and transmembrane) are different from each other. So, there classification on the basis of solvent energy had justified.

Here we have also tried to derive a hypothetical threshold for each class. For this purpose we have calculated the confidence interval for transmembrane and globular proteins. The confidence interval was calculated with 95%, 99% and 99.99%. The confidence interval for transmembrane proteins is calculated as follows,

Confidence interval (95%) = mean $\pm 1.960*(\sigma/\sqrt{10})$
$= 7.85 \pm 5.15$

Confidence interval (99%) = mean $\pm 2.576*(\sigma/\sqrt{10})$
$= 7.85 \pm 6.78$

Confidence interval (99.99%) = mean $\pm 3.29*(\sigma/\sqrt{10})$
$= 7.85 \pm 8.65$

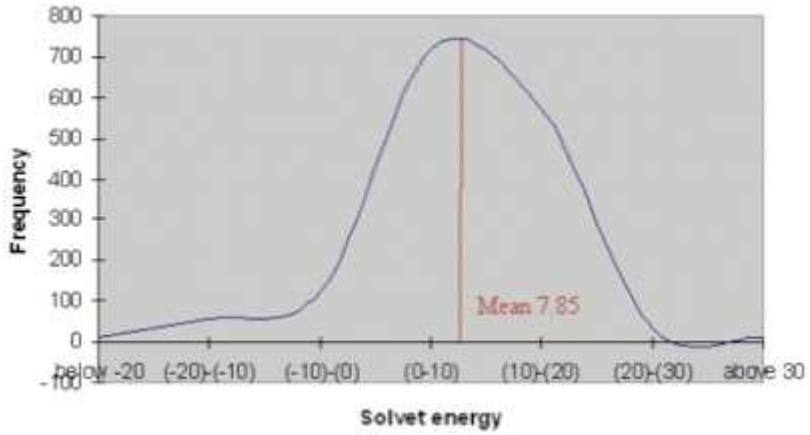


Figure 4: Distribution of solvent energy of the transmembrane proteins.

By this way we can say that 7.85 ± 8.65 is the 99.99 % assured threshold of solvent energy for transmembrane proteins. If any unknown protein shows a solvent energy between 7.85 ± 8.65, it can be consider as a transmembrane protein.

Likewise, we have also calculated the confidence interval for globular proteins and it is as follows,

Confidence interval (95%) = mean $\pm 1.96*(\sigma/\sqrt{10})$
$= -11.69 \pm 5.95$

Confidence interval (99%) = mean ±2.576*(σ/√10)
= -11.69 ± 7.85

Confidence interval (99.99%) = mean ± 3.29*(σ/√10)
= -11.69 ± 10.02

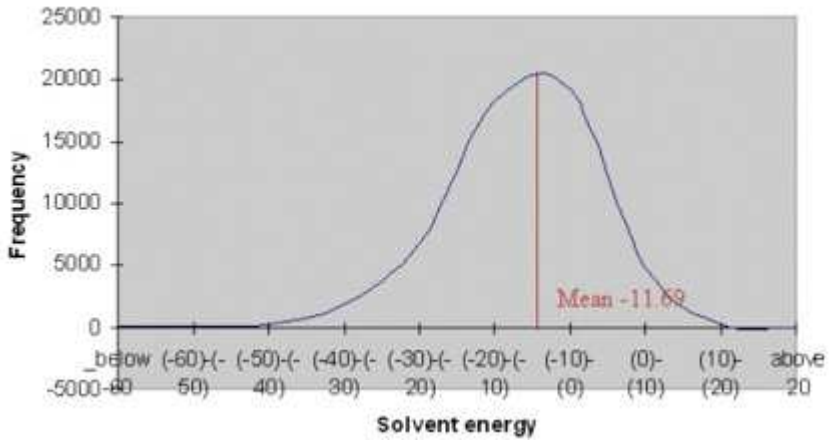


Figure 5: Distribution of solvent energy of the globular proteins.

By this way we can say that (-19.69) ± 10.2 is the 99.99 % assured threshold of solvent energy for globular proteins. If any protein shows a solvent energy between (-11.69) ± 10.2, it can be consider as a globular protein.

This hypothetical threshold could help to understand the nature of unknown proteins at the beginning level. If there is no data available for a particular protein, the solvent energy can be a first step of studying it.

Sensitivity and specificity

We had also carried out sensitivity and specificity test for our method. We had performed these tests by considering 99 % confidence interval for transmembrane (7.85 ± 6.78) and globular (-11.69 ± 7.85) proteins.

The total numbers of transmembrane proteins were 1554.
True positive (TP): 1030
True negative (TN): 194
False positive (FP): 0
False negative (FN): 330

$$\textbf{Sensitivity} = \frac{\textbf{True Positive}}{\textbf{True Positive} + \textbf{False Negative}}$$

$$= \frac{1030}{1030 + 330} = 0.757$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False positive}}$$

$$= \frac{194}{194 + 0} = 1$$

Total numbers of globular proteins were 49,836.
True positive (TP): 32,082
True negative (TN): 3692
False positive (FP): 0
False negative (FN): 14,082

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negaive}}$$

$$= \frac{32{,}080}{32{,}082 + 14{,}082} = 0.694$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False positive}}$$

$$= \frac{3692}{3692 + 0} = 1$$

Analysis of the solvent energy of the proteins revealed exceptional cases in the globular (5053) and transmembrane (36) protein datasets, i.e. proteins (chains) having opposite solvent energy. All proteins that had exceptional values with respect to their solvent energy and cellular placement (globular or transmembrial) were further investigated to explore possible reasons for having an opposite value of the solvent energy.

The 5053 exceptional cases of globular proteins were first classified into subclasses depending on their sub-cellular location such as membrane (inner or outer membrane), nucleus, cytoplasm and secreted and signal. The last subclass, the virion caspid protein subclass, was a separate subclass for proteins present in viral caspid. The exceptional cases and subclasses of the globular proteins are shown in Appendix 2.

The 765 proteins in the membrane subclass were either present in the inner or outer part

of the cell membrane. This could be one possible reason for their positive solvent energy. The 57 underlined proteins in the membrane subclass in Appendix 2 were special cases to this subclass because they are present in the membrane, cytoplasm and nucleus as well. The proteins present in the nucleus, cytoplasm and signal and secreted subclasses were believed to have negative solvent energy as they are dealing with an aqueous environment. These three classes must be investigated further by studying their amino acid sequences to find possible reasons for their exception in solvent energy. The last subclass contained viral caspid proteins. This class also contained 59 special cases of proteins, which were present in capsid, cytoplasm, cytoplasmic vesicle, membrane, and virion. These proteins were underlined in Appendix 2.

Similar to the globular protein class, the transmembrane proteins also contained 36 exceptional cases. The exceptional transmembrane proteins were present in the inner and outer part of the cell membrane. All of the proteins in the membrane class that showed exceptions with respect to solvent energy were investigated in detail for explanations of exceptional solvent energy (see Table 2). The underlined proteins illustrated in Table 2 were proteins with polypeptide chains present in the membrane as well as in the aqueous environment. It means that most of their protein structure is exposed to an aqueous environment. This could be a probable reason for the negative solvent energy. Exceptional cases of transmembrane proteins are shown in Table 2.

Membrane, inner or outer membrane	Secreted and signal,	Nucleus
1HSA, 1QO3, 1DDH, 1PTH, 1CQE, 1PGE, 1PRH, 1PGF, 1PGG, 1EHK, 1EK9, 2CUA, 1CYX, 1HWG, 1HWH, 3HHR, 1A22, 1AXI, 1B12, 1OCC, 2OCC, 3BCC, 1MPS, 1RGN, 1RHZ, 1RG5, 1F50, 1GFO, 1H2S, 1C3W, 1QHJ, 1FFT, 1AIJ	1BII, 7AHL	1E0P

Table 2: Sub-classification of exceptional cases of the transmembrane proteins.

5. Future work

While studying exceptional cases we came across prions. There are around 64 human prion proteins present in PDB. Prions are proteins that can flip their configuration from a helical bundle to beta sheets. Prion proteins have two isoforms that are made up of the same string of amino acids but have different three-dimensional structures. These forms are PrP-sen (sensitive prion or normal prion) and PrP-res (resistant prion or abnormal). PrP-sen is produced mainly in normal and healthy brain cells and susceptible to protease. The function of PrP-sen is not exactly known, however it is believed that it might be

involved in neuron communication. The second form is PrP-res, which is responsible for conditions like Creutzfeldt-Jacob disease (CJD), Alzheimer's disease (AD), bovine spongiform encephalopathy (BSE) and ovine scrapie (Axelsson, 2001). This form is highly resistant to protease. Once the organism is infected with PrP-res, PrP-res keep on replicating. It is believed that when PrP-sen comes in contact with PrP-res it gets converted into PrP-res. How PrP-res cause misfolding of PrP-sen is not yet understood. The conversion or misfolding of PrP-sen into PrP-res is suggesting that PrP-sen might not be in its stable configuration at physiological conditions and gets converted to a stable confirmation by PrP-res. PrP-res have a tendency to stick to each other and form amyloidal fibers. These amyloid fibers are very toxic and kill neural cells. Astrocytes remove dead cells leaving holes behind. In this way prions degrade the neuron. Prion proteins are present either on the membrane of the Golgi apparatus or membrane. On the basis of their cellular placement they must show positive solvent energy. However, most of them show negative solvent energy. Their solvent energy is statistically different from globular and transmembrane proteins, suggesting that prions might have a very specific solvent energy. Therefore, in the future, more work should be done to study the solvent energy of prions. It would be interesting and helpful to investigate more information about the solvent energy of prions as amyloidal fiber forming proteins in future.

Another future prospective of this project is to study the exceptional cases in more detail. To identify better the reasons underlying exceptional behavior, these cases should be studied on all possible levels like primary structure, secondary and tertiary structure, placement and functions. To study the primary structure would be very important because most of the exceptional proteins are small with only one peptide chain, which is different from our prediction of multiple chains. Our prediction was; peptide chains of exceptional solvent energy are protected by the rest of the peptide chains in the protein complex. However, this is not the case always and it has been proved by small peptides with exceptional solvent energy. By understanding the placement of each amino acid in sequences, e.g. the frequency of amino acids in certain domains/regions, it might be possible to reveal the exceptional behavior. That is, the answer for their exceptional behavior might lie in their amino acid sequence. Studies of the secondary and tertiary structure might increase our understanding of the folding pattern of these cases. It could be interesting to see whether these proteins have a specific pattern of folding or not? It is also important to study the exact cellular locations of exceptional proteins. The cellular placement of exceptional proteins can give important information about them because some of them are partially buried in the membrane; either in the inner or outer membrane. Some are present in the membrane, cytoplasm and nucleus also. It would also be very interesting to see which fold they adopt to perform their function and to cope with aqueous as well as lipophilic environments.

On the basis of this work, in the future it would be possible to predict the sub-cellular placement and also to some extent the molecular function of unknown proteins on the basis of accurate measurement of energy of solvation. The reason behind this is that the function of proteins is mainly depending on their three dimensional structures, i.e. folds, and identical folds are performing similar functions in the same or different organisms. Hence, it is obvious that identical folds can have identical solvent energy. This type of

prediction would be very useful for proteins like prions that change their secondary structures. However; prediction of molecular function would not be that accurate, as there is a wide number of proteins in living organisms and different folds can have identical solvent energy. In order to be able to predict the sub-cellular placement and function, this project must continue. The next step of this project should be the creation of a new database for all proteins in the PDB with a classification on the basis of identical solvent energy and similar function and a program, which will calculate the solvent energy of a query protein from its amino acid sequence. The program must also show the other proteins that are having identical solvent energy with the query protein. In this way this method can help in studying unknown proteins.

6. References

Branden C. and Tooze J. Introduction to Protein Sciences. 2nd ed. New York: Garland Publishing, Inc.

Shen HB and Chou KC, Predicting protein fold pattern with functional domain and sequential evolution information. *J Theor Biol*., In Press, Corrected Proof, Available online 19 October 2008.

Murzin AC: SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures. *J Mol Biol* 1995, 247:536-540.

Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM: CATH – a hierarchic classification of protein domain structures. *Structure* 1997, 5:1093-1108.

Y. J. Kim, J. M. Patel, A framework for protein structure classification and identification of novel protein structures, *BMC Bioinformatics*, *BMC Bioinformatics* 2006, 7:456

Murzin AG., How far divergent evolution goes in proteins *Curr. Opin. Struct Bio*l. 1998 Jun;8(3):380-387.

Andrei N. Lupas, Chris P. Ponting, Robert B. Russell, On the Evolution of Protein Folds, *Journal of Structural Biology* 134, 191–203 (2001)

Fumio Hirata, Molecular Theory of Solvation. Vol. 24, Tokyo: Springer, 2003.

K.A. Dill, Dominant forces in protein folding, Biochemistry, 29:7133-7155, 1990.

T. Baumeister, F. Cordes, A new Model for the Free Energy of Solvation and its Application in Protein Ligand Scoring, ZIB-Report 04-51 (October 2004)

Carney, Joanne, East, J Malcolm, Lee, Anthony G, *BiophysicsJournal*, May 15, 2007

(Lindahl and Johansson, 2008), Lindahl E. & Johansson A. C. V., Position-resolved freeenergy of solvation for amino acids in lipid membranes from molecular dynamicssimulations, *Proteins*, 70(4), 1332 –1344, 2008

Holm, L. and C. Sander, Evaluation of protein models byatomic solvation preference. *J Mol Biol*, 250(2), 258-275. 1992.

http://www.charmm.org/

T. Lazaridis, M. Karplus, Discrimination of thenative from misfolded protein models with an energy function including implicit solvation, *J. Mol. Biol.* (1998) 288, 477-487

D. T. Jones, W. R. Taylor, J. M. Thornton, A new approach to protein fold recognition, *Nature*, 358, 86-89, 1992

D.T. Jones, R.T. Miller, J.M. Thorton, "Successful Protein Fold Recognition by Optimal Sequence Threading Validated by Rigorous Blind Testing", PROTEINS: Structure, Function and genetics 23:387-397.

L.J. McGuffin D.T. Jones, "Improvement of the GenThreader method for genomic fold recognition", Bioinformatics 7:874-881, 2003.

D. T. Jones, GenTHREADER: An Efficient and Reliable Protein Fold Recognition Method for Genomic Sequences, J. Mol. Biol. 287, 797-815, 1999.

Research Collaboratory for Structural Bioinformatics) RCSB's PDB, http://www.rcsb.org/pdb/home/home.do

R. M. Jackson, M. J. E. Sternberg, A Continuum Model for Protein–Protein Interactions: Application to the Docking Problem, J. Mol. Biol. 250, 258–275, 1995.

(Scott *et al*, 2008), Luis P.B. Scott a, Jorge Chahine b, Jose R. Ruggiero, Use of genetic algorithms and salvation potential to study peptide structure, Applied Mathematics and Computation 195, 515–522, 2008.

http://cancerweb.ncl.ac.uk/cgi-bin/omd?hydropathy+plot

The basic reality of mind and spongiform diseases. *Med Hypotheses*., 57(5):549-54, 2001.

7. Appendix

Appendix 1: Significance of the differences between subclasses of transmembrane proteins and subclasses of globular proteins with each other. A single asterisk (*) denotes

P < 0.05, two asterisks (**) denotes P < 0.01 and three asterisks (***) denotes P < 0.001).

Globular	TM-Alpha	TM-Alpha 1	TM-Alpha 2	TM-Alpha 3	TM-Alpha Buried	TM-Beta
Globular-1	***	***	***	***		***
Globular-2	***	***	***	***		***
Globular-3	***	***	***	***		***
Globular-4	***	***	***	***		***
Globular-5	***	***	***	***		***
Globular-6	***	***	***	***		***
Globular-7	***	***	***	***		***
Globular-8	***	***	***	***		***
Globular-9	***	***	***	***		***
Globular-10	***	***	***	***		***
Globular-11	***	***	***	***		***
Globular-12	***	***	***	***	**	***
Globular-13	***	***	***	***	*	***
Globular-14	***	***	***	***		***
Globular-15	***	***	***	***	*	***
Globular-16	***	***	***	***	*	***
Globular-17	***	***	***	***		***
Globular-18	***	***	***	***	*	***
Globular-19	***	***	***	***	*	***
Globular-20	***	***	***	***		***
Globular-21	***	***	***	***		***
Globular-22	***	***	***	***		***
Globular-23	***	***	***	***		***
Globular-24	***	***	***	***		***
Globular-25	***	***	***	***		***
Globular-26	***	***	***	***		***
Globular-27	***	***	***	***		***
Globular-28	***	***	***	***	**	***
Globular-29	***	***	***	***		***
Globular-30	***	***	***	***		***
Globular-31	***	***	***	***		***
Globular-32	***	***	***	***	*	***
Globular-33	***	***	***	***	*	***
Globular-34	***	***	***	***		***
Globular-35	***	***	***	***		***
Globular-36	***	***	***	***		***
Globular-37	***	***	***	***		***
Globular-38	***	***	***	***		***
Globular-39	***	***	***	***		***
Globular-40	***	***	***	***		***

Globular-41	***	***	***	***	**	***
Globular-42	***	***	***	***		***
Globular-43	***	***	***	***	**	***
Globular-44	***	***	***	***	*	***
Globular-45	***	***	***	***	**	***
Globular-46	***	***	***	***		***
Globular-47	***	***	***	***	*	***
Globular-48	***	***	***	***		***
Globular-49	***	***	***	***		***
Globular-50	***	***	***	***		***
Globular-51	***	***	***	***		***
Globular-52	***	***	***	***	*	***
Globular-53	***	***	***	***	*	***
Globular-54	***	***	***	***		***
Globular-55	***	***	***	***		***
Globular-56	***	***	***	***		***
Globular-57	***	***	***	***		***
Globular-58	***	***	***	***		***
Globular-59	***	***	***	***		***
Globular-60	***	***	***	***		***
Globular-61	***	***	***	***		***
Globular-62	***	***	***	***	**	***
Globular-63	***	***	***	***		***
Globular-64	***	***	***	***		***
Globular-65	***	***	***	***		***
Globular-66	***	***	***	***	*	***
Globular-67	***	***	***	***	*	***
Globular-68	***	***	***	***		***
Globular-69	***	***	***	***	*	***
Globular-70	***	***	***	***		***
Globular-71	***	***	***	***	*	***
Globular-72	***	***	***	***	*	***
Globular-73	***	***	***	***		***
Globular-74	***	***	***	***	**	***
Globular-75	***	***	***	***	**	***
Globular-76	***	***	***	***		***
Globular-77	***	***	***	***		***
Globular-78	***	***	***	***		***
Globular-79	***	***	***	***		***
Globular-80	***	***	***	***	**	***
Globular-81	***	***	***	***		***
Globular-82	***	***	***	***	*	***
Globular-83	***	***	***	***		***
Globular-84	***	***	***	***		***
Globular-85	***	***	***	***		***

Globular-86	***	***	***	***	*	***
Globular-87	***	***	***	***		***
Globular-88	***	***	***	***		***
Globular-89	***	***	***	***		***
Globular-90	***	***	***	***		***
Globular-91	***	***	***	***		***
Globular-92	***	***	***	***		***
Globular-93	***	***	***	***		***
Globular-94	***	***	***	***		***
Globular-95	***	***	***	***		***
Globular-96	***	***	***	***	*	***
Globular-97	***	***	***	***	**	***
Globular-98	***	***	***	***	*	***
Globular-99	***	***	***	***	**	***
Globular-100	***	***	***	***		***
Globular-101	***	***	***	***		***
Globular-102	***	***	***	***	*	***
Globular-103	***	***	***	***		***

Appendix 2: Further classification of exceptional cases of globular proteins into Membrane, Nucleus, Cytoplasm, Signal and Secreted and Viral caspid proteins.

Membrane, Inner or outer membrane	**Nucleus**	**Cytoplasm / Globular**	**Secreted and Signal,**	**Virion Caspid protein**	**De-novo and synthetic**
1QD5,1V54,1AR1, 1K4C, 1BTR,1OCR,1OCZ,1D2V ,1OCO,1RH5,1CDC, 1C8S,1R3L,1VF5,1R3L, 1VF5,1HO9,1ATY,1FI8, 2AT9,1N2D, 1S5L, 1FLC, 1R3K,1D7W,1NHD,1JVM ,1S5H,1NKZ,1A64,1RHZ, 1NHW,1V55,1O6A,1BL8, 1NNU,1NKZ,1Q90,1JVM, 1NHG,1K4D,1NHD,1PPJ, 1KZU,1NHW,1A64,1BGY , 1RKL,1BHA,1JB0, 1BHB,1F0C,1BE3, 1NTZ, 1GRM,1PP9, 2BCC, 1LGH,1S5L,1BCC,1NU1, 1IJD,2BRD,1K24,1BNX, 1JY2,1NTK,1KYO,1BDE, 1H6W,1AP9,1JO5,1QLE, 1KB9,1EZV,1P4T,1J95, 1FW3,1S5L,1NTM,1ILD, 1M57,1A0T,1OH2, 1QD6,	7MSF,1HBW, 1R8U, 1L0L, 1IBT, 1DSK, 1GXS, 5LIP, 2LIP, 3LIP, 1L0N, 1IM0, 1W0K,1NYH, 1P3A, 1F66, 1GK6, 1P34, 1P4Q, 1B35, 1N6J, 1LOI, 1AJ1, 1L8C, 1HN3, 1IPP, 1A92, 1PAR, 1EV0, 1F36, 1AAF, 1OMA,1CYQ, 1BDT, 1CE9, 1S5R, 1T2V, 1P3K, 1FM6, 1EVX, 1ISQ, 1A73, 1KX4,	1EPT, 6RLX, 1D5L, 1LOC, 1N13, 1ZDJ, 1QJ0, 1IFD, 1IK7, 1L0N, 1HH0, 1O9Y, 1A7B, 1QJ0, 1N1C, 1QHH, 1L0N, 1H7D, 1A0S, 1A0S, 1WDC, 1A0S, 1SCM, 1SXJ, 1PYA, 1AI0, 1EGW, 1OJH, 1M93, 1P3G, 1LJ2, 1IBU, 1ECM, 1Q7L, 1IBV, 1P3Q, 1R6R, 1IBV, 1C8O, 2HIU, 1IBW, 1H7J, 1WRT, 1MT1,	1B2E,1G7A,1OFS,1FU2, 1J73,1BQP,1TYM,1FUB, 1AIY, 1G7B, 1PID, 1G7B, 1OS4, 1EV6, 1B2C, 1RIN, 1LGC, 1LOD, 1RIN, 1LOF, 1LOB, 1HTV, 1SDB, 1LPH, 1FE6, 1B2F, 1B2F, 1B19, 1B2A., 1B2D, 1LOA,1GAV, 1XGL, 1TRZ, 1JCO, 1KFP, 1XDA, 1B18, 7INS, 1LOE, 1B2B, 1WAV, 1B17, 1UYO, 1QIZ, 1LGB, 1ZNJ, 1SF1, 1BEN, 1EV3, 1LOG, 1JCA, 1Q4V, 1LYW, 1QIY, 1ZEH, 1FI8, 1MA5, 1ZEG, 1MT1, 2AIY, 1IZA, 1KVE, 1RY3, 1JJO, 1LYW, 1KVE, 1HE0, 1KVD, 1WAV, 1QIZ, 1EBO, 2HNT, 1JY3, 1MA2, 1N73, 1PG1, 1M4F, 1LWU,1RPC, 1OIL, 1M4E, 1HVZ, 1LWU, 1HQD, 4LIP, 1A91, 1MM5,	1H8J, 1G2C, 1SVF, 2CPB, 5MSF,1DWN, 1HE6, 1BMS, 1FR5, 1NH4, 1MST, 1IFI, 1HDW, 1ZDI, 1E7X, 1ZDH, 1EBO, 1DZS, 1UNA, 1HGZ, 1IFJ, 1MVA, 1GKV, 1DSJ,	1JY6,1KCN, 1MPV, 1Q4F, 1JY4, 1LT1, 1BYZ, 1IN2, 1N09, 1BYZ, 1L4X, 1OVR, 1SOP, 1K43, 1D1E, 1AL1, 1IJ3, 1LT1, 1M02, 1BCV, 1JM0, 1OVU, 1SBU, 1Q2F,

1M56, 1PI7,1P84,1NYU, 1G90,1BZK,3IFM,1IFN, 4IFM,1V54,1S5L,1M57, 1NTK,1FW3,3MRA,1ILZ, 1ORM,1VF5,1EZV,1KQF, 1BRD,1OCR,1JB0,1KKD, 1V55,1FQY,1UYN,1OH2, 1QD6,1FW2,1KYO,2IFO, 1FFT,1KYO,1OCZ,1P84, 1A0T,1JB0,1EHK,1KB9, 1OKC,1AFO,1JDM,1O7D ,1MAL,1IJP,1NTM,1B9U, 1OED,1KF6,1MPN,1FAV, 1R2C,1C99,1MPR,2MPR , 1IIJ,1L6T,1S00,1KFY, 1MPO,1MM4,1FFT,1H2S ,1N7L,1DXR,1RHZ,1MP Q,5PRC, 1MXM, 1RWT, 1MPM,7PRN,1QLA,1RVJ ,1PO3,1AF6,1QLB,1ORS ,6PRC,3PRC,2PRN,1AIJ, 2PRC,1L9J,7PRC,1E54, 1BXW,1RZH,1MSL,1RY5 ,1JGY,1DS8,1H6S,1E14, 4PRC,5PRN,1RQK,1E7P , 1NEK,1NEN,1RZZ, 1PRN,1K6L,1AIG, 3PRN, 1K6N,1RGN,1EYS,6PRN ,1FNQ,1E6D,1M3X,1FJP, 1DV3,1SDZ,1BCT,1QOV, 1M57,1KBY,1JGX, 1JGZ, 1DV6,1PRC,1PCR,1M56, 4RCR,1F6N,1KMP,1FNP ,1MPS,1L9B,1UMX,1PST ,1PNZ,1Q16,1PSS,1IH5, 1SIW,1JGW,1L9B,1KMO, 1UJW,1RG5,1NQE,1OR Q,2RCR,1YST,1NQH,1B Y3,1NQF,1S7B,1BY5,1Q KC,1FCP,1P84,1FI1,2FC P,1QFF,1QJQ,1FEP,1PF 4,1Q9F,1H8E,1DEI,1JEK ,1OSG,1L6K,1E1Q,1PHO ,1OSX,1MAB,1P7B,1RP Q,1E79,1HVV,1KIL,1JTH ,1H8H,1RPQ,1D7W,1L4A ,1BTT,1FLC,1SMZ,1HOF ,1O6A,1DNW,1CXP,1QJ 9,1DNU,1RFL,1AIK,1DX Z,1R3J,1R3I,1OM9,1E0A , 1VYT,1M27,1D2J,1HE7, 4HVP,1BYY,1FVY,1OGT, 1B33,1ICF,1H15,1IWQ, 1KTL,1KBG,1LCJ, 1LEK, 1OEI,1JPF,1P1Z,1MHE, 1GFN,1N0X,1PTQ,1OHH	1M18, 1NVP, 1ID3, 1CE9, 1KX5, 1H3O, 1JN7, 1P3L, 1HWU, 1P94, 1NGM,1EQZ, 1S32, 1JMT, 1P3F, 1ZTA, 1K2M, 1FHR, 1GO4, 1AXC, 1N0W,1GD2, 1ZIM, 1AOI, 1HJI, 1FIP, 1M19, 1RP3, 1YCR,1MNM ,1OL5,1DML, 1MYL, 1N4M, 3ERD, 1P3P, 1OT7, 1LGQ, 3FIS, 1PD7, 1RJK, 1ZIK, 1IJ0, 1RB4, 1M1A, 1RK3, 1RB1, 1KZ0, 1KZ2, 1OSV, 1FMH, 2PJR, 1HCW, 1ZIJ, 1IJ2, 1IJ1, 1DL6, 1TFC, 4FIS, 2PRG, 2ZTA, 1PD3, 1LQB, 1P9D, 1NRL, 1RB5, 1ZIL, 1RB6, 1CZ0, 1IRQ, 1K7L, 1O9K, 1SWI, 1PDQ, 1ZIL, 1KV6, 1RB6, 1NRL, 1MV9,1GUX, 1ETO,1MZN, 1N5G, 1P3M, 1P8D, 1G3J, 1RB5, 1L2I, 1CYQ, 1G39, 2ZNF, 1YTF, 1UHL,1MM3, 1IQ5, 1OV9, 1N4H, 1NQ7, 1FYB, 1KZ5, 1IPP, 1PFT, 1Q08, 1DPU, 1KGB, 1PIQ, 1K4W, 1K74, 1FM9, 1TAF, 1GNG, 1PYI,	1IBT, 1HQ6, 1J1D, 1CO0, 1FZP, 1LYA, 1MI7, 1KMI, 1MHL, 1IEO, 1MYP, 1H8B, 1D4T, 1PSB, 1JPP, 1J55, 1BMQ, 1BT6, 1ICE, 1FPR, 1UKL, 1NIW, 1HS5, 1P3B, 1FT8, 1OQP, 1L6O, 1RFO,1JPP, 1YCQ, 1CTP, 1S6N, 1N4R, 1V9U, 1MQ1, 1L1K, 1PSB, 1C9I, 2BBN, 1DOW, 1MFG, 1P3B, 1K5K, 1MV4, 1DP3, 1F36, 1N69, 1P9C, 1SMH, 1DKD, 1N7S, 1QTX, 1SYQ,1EJH, 1L3E, 1P0L, 1Q61, 1Q8W, 1CDK, 1K1F, 2MLP, 1D8E, 1APM, 1ATP, 1Q24, 1Q62, 1Q8T, 1YDR, 1YDS, 1Q8U, 1MNF, 1K1F, 1RDQ, 1FMO, 1YDT, 1S3S, 1HQ3, 1MXE, 1I7W, 1JBP, 1L3R, 1G8E, 1I51, 1OZS, 1MBY, 1IBC, 1OK7, 1P0J, 1OW8, 1LXF, 1VH6, 1FS1, 1I4O, 1JPW, 1MXL, 1TMZ, 1RE1, 1N0D, 1UHD, 1JLU, 1CQG, 1CDM, 1RHQ, 1AVO, 1MHM, 1C26, 1AW8, 1RCS, 1CM4, 1GK7,	4LIP, 1L0V, 1QGE, 1SPF, 1JH0, 1PO0, 1ORY,1ZWG, 1J1E, 1HIS, 1ZNJ, 1B9E, 2TCI, 1CFG, 1JLP, 1TYL, 1GUJ, 1HIT, 1MSO, 1LKQ, 1QOW, 1L6L, 1MZ9, 1PID, 3MTH, 7INS, 1FZE, 1G96, 1OS3, 1IZA, 1TRZ, 1UYA, 1ZNI, 1MHI, 1MQX, 3CAA, 1M5A, 1H6I, 1JCO, 4CAA, 1DFN,1MQZ,1LES,1QMO, 3INS, 1EDN, 1HIQ, 1SJT, 1MHJ, 2ACH, 1KMF,1QVH, 1ETM, 4INS, 2INS, 1BON, 1AS4, 1IOG, 1V6R, 1CW6, 1MA4, 1HLS, 1IZB, 1AS5, 1PQR, 1I78, 1BPH, 1P9J, 1LNP, 1MA6, 1HUL, 9INS, 1IOH, 1BCR, 1B2G, 1A7F, 1B18, 1APH, 1CPH, 1DPH, 1CZ6, 1LQ8, 1XGL, 1HUI, 1B17, 1B2B, 1IZA, 1MPJ, 1R1G, 1HKD, 1AV1, 2LTN, 1TMB, 1JAC, 1JUI, 1KJG, 1UGW,1UGY,1RU5,3MON, 1PW9, 1BTG, 1BJR, 1XTC, 1PPB, 1RO4, 1ID5, 1A92, 1T37, 1F8P, 1CXO, 1WFA, 2STA, 1SGF, 1JAC, 1IJU, 1DQC, 1RTF, 1N7T, 1TAB, 1JQ8, 1B0N, 1RU7, 1G37, 1WFB,1C9P,1DE7, 1EWS, 1XY2, 1DW4, 1EFE, 1QKY, 1AZJ,1HI6, 1MOT, 1IJV, 1AZ6,1KUN,1MHW,1MYU, 1KCP,1RGX,1CFW, 1DCD, 1RUU, 1DU9, 1ARQ, 1JYI, 1UGW,1UGY,3BTW,1DHG ,1UCY,1EGP,1VWB,1VWD ,1VWE,1VWM,1SHP,1KU8, 1FAK,1PSB,1BH4,1O8Y, 1JOJ,1ETR,1M26,1MUJ, 1AGQ,1DXG,1ALX,1P22, 1JBL,1DW5,1EX9,1BBR, 1JLZ, 1MPE, 2CTI, 1GUR, 1BH0, 1ABI, 1IRR, 1PMZ, 1YCP, 1PC6, 1N0X, 1ETH, 2STB, 1FIW, 1CO7, 1CMK, 1ETT, 1S7P, 1TOC, 1BZB, 6INS, 1JYI, 1UGX, 1ALZ, 1M3D, 1O8Z, 1JBN, 1SFI, 2PKA, 1FVN, 1ZEI, 2BTC, 1ANS, 1RJK, 1HD9, 1CRF, 4ER4, 1WM8,1GD2, 1QK7, 1O8T, 1ONT, 1H8T, 1OQE, 3ERD, 1FIZ, 1EYO, 1A0M,	1MVB, 1HE6, 1Q4V, 1AQ4, 1ZDH, 1MST, 1MVA, 1E6T, 1GKV, 1EBO, 1KUO, 1BMS, 1ZDJ, 1HGV, 1A6P, 1DZS, 1ZDK, 1IFK, 1AQ3, 1E7X, 1GKW, 1FRS, 1HDW, 1ENV, 2MS2, 1ZDI, 1MSC, 1JY3, 1QL2, 1QL1, 1C0V, 2IFN, 2IFM, 1IFM, 1PFI, 1IFP, 1FDM, 1GAV, 2EBO, 1FR5, 1JAU, 1RVZ, 1QR8, 1RHI, 1RV0, 1EJO, 2MIP, 1JSI, 1MQT, 1KKE, 1HTM, 1F23, 1VWB, 1RVX, 1JPX, 2BBV, 1R1A, 2HWD,	1QA5, 1JC8, 1Q2J, 1C94, 1LT1, 1D0W, 1P9I, 1IN3, 1XBH, 1NVO, 1S9Z, 1N0C, 1M3W, 2MAG, 1F0H, 1RH4, 1IMW, 1JBF, 1KVF, 1PEF, 1I93, 1CZQ, 1R9I, 1LB7, 1KVG, 1I98, 1T6A, 1FUV, 1FUL, 1LQV, 1G1Z, 1J8N, 1A1P, 1HZ3, 1I8E, 1K8D, 1IHQ, 1NEI, 1ICL, 1KDD, 1BKV, 1HQJ, 1T3U, 1JMB, 1HQJ, 1IHQ, 1OVV, 1EC5, 1BHX, 1COI, 1PBZ, 1LE0, 1NEI, 1FOZ, 1KDD, 1KD9,1N0 A, 1LE1, 1ICO, 1COS,

,1HCF,1JGJ,1EEO,1ZTO ,1GL2,1G4Y,1UAZ,1WW A,1LF8,1AQG,1HZX, 1OQE,1KUJ,1KUZ,1EPJ, 2CKB,1A81,1IVO,1PY1, 1UJK,1KPR,1EEY,1LQS, 1EFX,1TME,1S4X,1S7M, 1OF2,1BEV,1UJL,1K8D, 1HES,1SFC,1D4M,1JD5, 1QBZ,1JCP,1NPQ,1UCQ ,1T0J,1L4A,1KUP,1JL9, 1PV7,1NYS,1JGD,1D9L, 1PLP,1BQH,2MHA,2VAA ,1FZM,1JWG,2SIV,1QCE ,1CDL,1S4T,1B4G,1IW6, 1EES,1J9V,1UEF,1JDK, 1RKG,1QGW,1DZE, 1EEZ,1Q68,1P34,1NYU, 1OSZ,1UTC,1GFP,1PON ,1T3L,1KLG,1QCM,1QSF ,1QR1,2SIV,1Q5L,1DKX, 1DKZ,1OM2,1D9M,1F4V, 1R5W,1P8D,1T1L,4PRC, 1NJ0,2CLR,1A03,1LK2, 1PSS,1QU7,1BX2,1D9O, 1CWQ,1QM8,1CE0, 1KMC,1KPU,1NAM,1FZJ ,1I7T,1PQ1,1Q69,1HXU, 1OQW,1M25,1INQ,1C4D ,1BT9,1S8L,1EGT,1GFQ, 2CLR,1I7U,1K36,1ED3, 1G0Y,1VPU,1LP9,1DUY, 1OQD,1QSE,1MVK, 5PRC,1UHE,1KGB,1J8Z, 1KN7,1OSM,1M45,1JUF, 1G04,2CDX,1HZX,1I4M, 1JOY,1BD2,7PRC,1JHT, 1I78,1MHC,2PRC,1FCT, 2RCR,1N64,1OGA,1EBA ,1UEF,1CKY,1JO6,1G6R ,1M8O,1QLS,2POR, 1QLO,1BRV,1P1P,1BJC, 1CF4,1I5Y,1Q2I,3PRC, 6PRC,1DUZ,1HHK,1PRC ,1JF1,1KWE,1CKZ,1L9H, 1CKW,1BRX, 1FGE, 1M23,1DVW,1DXR,1F88, 1P23,1AQD,1EBP,1VF6, 1QD6,1DLB,1TMR,1GCL ,1CKX,3LDH,4RCR,1P8I, 1R2N,1QRN,1T1L,1JV6, 1D9P,1G1J,3POR,1M57, 1HHI,1AML,1L6E,1BE3, 1GZL,1HVV,1NBM,1KG8 ,1FGD,1M56,1CE4,1DXP ,1PST,1SRQ,1JK8,1GK5,	1GWQ,1B28, 1ETW, 1N1J, 1E91, 1JB6, 1MVC, 1JFI, 1K7L, 1G2Y, 1ZNM,1SRS, 1ZII, 1Q09, 1B6Q, 1FV5, 1KKQ, 1GP2, 1Q0A,1GMG, 1F93, 1IRQ, 1H2L,1GWR, 1GG2, 1KJY, 1IK9, 1GCM, 1ETQ, 1PET, 1G2Z,1M36,1 UFI, 1FIA, 1P3O, 1GCL, 1S5Q, 1G1E, 1G2Z, 1H2K, 1ETK, 1HV2, 1GCL,1ETY, 1S4Z, 1MVF, 1P3I, 1ETV, 1DU2, 2HIO, 1P65, 1ETY, 1G1I, 1PCG, 1ETX, 1BH9, 1PL5, 1JEN, 1I7B,1TBC, 1AE2,2A93, 1N9S,1F2I, 1LPQ,1SZD, 1MUL,1JN5, 1A7W,1KNA, 1KNE,1BAZ, 1N9S,1GUS, 1OAI,1QEY, 1BY9,1BW5, 1BHI,1G6G, 1N9R,1F2I, 1ZME,1RPV, 1RK8,18VP, 1NLW,1MDY, 3ULL,2WRP, 2ADR,1UMU, 1HTA,1LM8, 1PZR,2GN5, 1LMV,2A93, 1EBK,1FG8, 1Q5V,1UPG, 1BDV,1SUT, 1F59,1EVW, 1NC8,1PFB, 1H89,1EA4, 1HBX,1EBK,	1AIE, 1NMR, 1RHJ, 1FZF, 1DD4, 1OMW, 1S0Y, 1RHM, 1TIH, 1KNZ, 1GFW, 1NME, 1OX3, 1DS5, 1RHR, 1P93, 1P0O, 1RHK, 1PYU, 1DD3, 1GK4, 1S0Y, 1JYU, 1AOJ, 1H2M, 1SKV, 1HUC, 1EJ4, 1MT1, 1E0Q, 1CMX, 1LI1, 1KZS, 1IHR, 1I79, 1S0Y, 1BZG, 1I7C, 1I7M, 1MT1, 1NHO, 1BH8, 1RQT,1NZV, 1H26, 1MTN, 3SAK,1GUG, 1GUO,1JCS, 1DZ7,1PZR, 4OTA, 1PAV, 1FCH,1JJG, 1O6K, 1KBH, 1N6E,1HR9, 1M2Z,1CZZ, 1D0A,1ELR, 1EAI, 1P6Z, 1TRR,1BJP, 1O6L,1EE4, 1V65,4OTB, 1NU2,1DW9, 1OKU,1FCH, 1OQN,1GYY, 1H28,1RYJ, 1NNR,4OTC, 1EMU,1MJM, 1TRO,1DWK, 1NAQ,1UND, 1LEW,1RLP, 1HCT,4OTA, 1QZ7,1MED, 1HR9,1KSO, 1P83,1OZB, 1MZW,1MA3, 1O6O,1FPH, 1QLP,1F9E, 1F47,1I7X, 1VKE,1UCN, 1F9E,1KR4, 1TZE, 1CMI,	1F58, 1G89, 1RKC, 1LJV , 1JYC, 1LWU, 1HIA, 1RFB, 1UDK, 1N8O1DKC, 1VIT, 1ETS, 1RVS, 1TBQ, 1F0E, 1SLD, 1RVX, 1EJA, 2CK0, 1PPE, 1P2J, 1N9U, 1SKV, 2KAI, 1I51, 1TER, 1OAW, 1LU0, 1KQH, 1IAK, 1J5B, 1JAR, 1KMC,1CDT,1FWO, 1SP4, 1D1F, 1K64, 1FZG, 1PCO,1MXP,1MXN, 1BQF, 1BMP, 1L4V, 1TBR, 1PEN, 1GL0, 1JAA, 1HRP, 1JK4, 1JUI, 1DG2, 1QXQ, 1LUP, 1FYR, 1MM0, 1B98, 1B5F, 1CFS, 1ER8, 1F0F, 1KG1, 1HY9, 1JMO, 2CCX, 1IVA. 4TGF, 1A13, 1HTR, 1F0D, 1QDP, 1R7H, 1HCN, 1B03, 1LU0, 1LGP, 1SP7, 1JOU, 1DF6., 1PFT, 3CTI, 1OEG, 1I02, 1CDT, 1HTH, 1LX5, 1M5J, 1JPY, 1C98, 1SRB, 1QGM, 1RHU,2BI6, 1SBW, 8TFV, 1DVA, 1DUM, 1ILK, 2ILK, 1HO0, 1OEF, 1M5M, 1CM1, 1C9A, 1N2Y, 1AVG, 1N1L, 1CCV, 1KJ2, 3MON, 1LXI, 1NBJ, 1CSB, 1G8C, 1GCN, 1FL7, 1G9I, 1KX6, 1I8X,1CFH,1FI8,1I8Y,1CB3 , 1MCT,8API,1GOE, 1TGX, 1BX8, 7API, 1B8K, 4SGB, 1LPB, 2H1P, 2NBT, 1L5B, 1FRY,1P1P, 1DGR, 1AKG, 1M2C, 1VPP, 1PJV, 1D9J, 9API, 1D7N, 1GIB, 1IM3, 1GMN,1HYK, 1F2S, 1SOH, 1KRL, 1OHM, 1EZX,1NPO, 1SG1, 1S4G, 1RE3, 1UT3, 1BKU, 1SCY, 1ZWF,1HPY, 1D0R, 1NOT, 1F3K, 1I07, 1A1R, 1CCQ, 1W0J,1OAU, 1WHS, 1M7L, 1N98,1EMX, 1UL2, 1OYV, 1XGA, 1FZA, 1E74, 1MII, 1XGC, 1KMR, 1JLO, 1H0J, 1QMW, 1G9P, 1ICA, 1BYV, 1IEN, 1HJE, 1F0G, 2CRS, 1KV4, 1OG7, 1OHN, 1FFJ, 1RF1, 4MON, 1XGB, 1MTQ, 1N86, 1LTJ, 1G2G,4MON,1AGG, 1AQ5, 1HU6, 1DGW, 1IOD, 1LFC, 1LT9. 1D5S, 1N86, 1ALG, 1G1I, 1M57, 1HXL, 1B8M, 1M3D,1QFA,1D6X, 1DGW,	2HWE, 2HWF, 1UPN, 1LVB, 1CN3, 1FPN, 1WDG, 1QA4, 1JQ0, 1LCX, 1BMX, 1BM4, 1EV1, 1MEC, 1QBE, 1MZI, 1AL0, 1AYN, 1NOV, 1WDF, 1GFF, 1SZT, 1ND2, 1AYM, 1QJU, , 1QJY, 1K34, 1ND3, 1F8V, 1C8M, 1NCR, 1NS3, 2MEV, 1KWD, 1DF4, 1FR5, 1JXP, 1DY8, 1DY9, 1I5X, 1IBN, 1QQP, 2EZP., 1FMD, 1FOD, 1KZV, 1KZT, 1BBT,1MO F, 1IBO, 1KJ4,1HXS ,1VBB, 1PIV, 1PVC, 1JMU,1VB E, 1RVZ, 1NG7, 1K5M,	1FI0, 1NAK,1KY C, 1JMB, 1F90, 1QHR, 3AL1,

1JB0,1FAQ,1QXX,1R84, 1MGY,1EFR,1BRR, 1QG9,1L2W,1COW,1A11 ,1AA0,1FQJ,1CKK,1FKN, 1RZX,1MK7,1JPL,1CNO, 1F83,1WWB,1BDV,1G1H , 1CWE, 1JEG, 1I1F, 1VFY,1PD0,1RSO,1C4U, 1Q3P,1JOH, 1CNO, 1NOH,1SOR,1PV6,1AIG, 1P1L,1T16,1DS8,1GUT, 4OTA,1RQQ,1B9Y, 1GUG,1OCY,1F4Z,1O6N ,1EAW,1CI6,1M7E,1NO4 ,1Q94,1HXT,1BZ9, 1WWC,1MK9,1AAP, 1PCX, 1LKL,1RFI, 1JUQ, 1EYX,1PEI,1MIZ,1B4I, 2PDZ,1B08,1N3N,1S7V, 1P9U, 1TBG,1JTH,1J8H, 1JWS,1PXR,1KJV,1MFL, 1IHJ,1GYO,1CG9,2SEB, 1NX0,1PBY,1M6Z,1EGF, 1I7R,1PXR,1DNY,1QGC, 1K0T,1PXS,1EVW,1UK4, 1IYM,1G1S,1JWM,1R5I, 1RVT,1M6O,1HHJ,1LO5, 1P4U,1OPF,1GFM,1H68, 1NX1,1X11,1G5Q,1N5A, 1LDR,1I4F,2OMF,1PTU, 1HXY,1HTP,1P7Q,1NX0, 1E27,1JWU,1FYT,1A3P, 1UHB,1B9X,1QKP,1VIT, 1PD1,1GUN,1BII,2R1R, 3R1R,1GUE,1SPS, 1TOR, 1MQS, 1APP, 1IQ1,1KG0, 1DLH,1NHL, 1PNB,1OBY,1AUT,1JPL, 1KO6, 1S7U,1FFN,1FFP, 1RUZ,1JMZ,1I31,1BW8, 1GZM,1RV0,1S7X,1CAV, 1LDP,1HPW,2TRC,1K33, 2PHK,1R5V,1IR3, 1HHG, 1M4H,1KME,1FSB,1S7Q ,1N59,1C9L,1JSD,1DF5, 1FO0,1NAN,1GOT,1AIK, 1Q5I,1F3J,1DM4,1CR8, 1N7F,1S7R,1A0R,1I1Y, 1Q5B,1S7W,1IM9,1QQD, 1PTR, 1FFO, 1QVO, 1H6E,1HQR,1QJ6,1LE3, 1JI4,1D9K,1BM1,1BXX, 1S7S,1S7T,1P8V,1N4Q, 1TOS,1OW6,1KZO, 1AMC,1IW9,1VQE,1BJB, 1M46,1EEN,1SE0,1RXZ,	1PIL,1O9S, 1J4X,1O6P, 1RFF,1F59, 1TVT,484D, 1RXM,1N87, 1BDT,1JUU, 1KJ4,1PZL, 2GNK,1H24, 1UL3,1LEZ, 1OLH,1JGN, 1DGC,1PAA, 1BAZ,1K2N, 2DRP,1PZW, 1YHB,1A02, 1VQJ,1TFE, 1R2B,1FOS, 1VQB,1OJ5, 2DGC,1SB0, 1JM4, 1VQI, 1HRZ,1VQH, 1VQD,1QPM, 1MDY,1VQG, 1PAR,1GXC, 1NOP,1CQT, 1VQF,1F3U, 1CD3,1A79, 1AE3,2PII, 1GKH,1BDV, 1VQA,1CT6, 1FU9,1AFT, 1K6O,2NMB, 1CS9,1B8H, 1K3Q,1LLM, 1TVS,	1MN3,1F4V, 1WRP,1QDU, 1CMB,1MJK, 1C4E,1R1R, 4R1R,1QJA, 1CTD,1UN0, 1SAE,1S4H, 1WRS,1Q5W, 1RO3,1QS7, 1KZZ,1A0A, 1F1W,1HR8, 1NJ3,1G8E, 1UR6,1LQ1, 1Q4Q,1SHA, 1BMB,1P4B, 1JMX,1JWY, 1JX2, 1EZJ, 1SAF,1SHD, 1D4W,1HFE, 1KJ7,1QSC, 1O9U, 1OGY, 1PUB,1IID, 1OKV,1OW7, 1QGR,1UEO, 1JD6,1STC, 1VRK,1DKY, 1P0G,1EGS, 1WD2,1RRZ, 1JH4,1MDK, 1MDJ,	1QMB, 1ZWD,1LTA, 1KJH, 1BB1, 1CIX, 1G1J, 1HTL, 1A7F, 1LTT, 1LTS, 1LTB, 1RE4, 1FZB, 1HQQ, 1LTG, 1BF2, 1KQI, 1E76, 1LTI, 1HLT, 1HUC, 1RF1, 1LNP, 1ZWB, 1IM1, 1PNH, 1CNL, 1WWW,1FLE,1HPH,1VDF, 1AZH, 1DTC, 1VTX, 1IMI, 1LXG, 1PMX, 1FZC, 1BCS, 1CB9, 1BWX, 1HXZ,1CQ4, 1HLE, 1K3B, 1QX9, 1G26, 1K8V, 1MR0, 1Q71, 1VKT, 1LTJ, 1K3M, 1UYB, 1IDH, 2DTB, 1ZWE, 1CIR, 1JJO, 1COU, 1K09, 1HR1, 1JBU, 1HY2, 1HUI, 1CIQ, 1RY1, 1BH1, 1PP5, 1E75, 1FBM, 1HUL, 1IOH, 1I5J, 1LYA, 1LT9, 1ETL, 1WHT, 1FU3, 1LI1, 1PDG, 1P0T, 1G1P, 1ETN,1A37,1MGX,1GVX, 1QG7,1AE8,1BCU,1QJ7, 1NY2,1TOM,1G30,1G32, 1H8D,1H8I,1BET,1Q3M, 1KC2,1M8A,1KYJ,1C1U, 1C1V,1C1W,1C5O,1GHW, 1GJ4,1GJ5, 1GVU, 1NRN, 1FEV,1GMD,1GL1,2PTC, 1SB1,1LOA,1NRS,1CCM, 1CFT,1AAL,4TPI,1KTS, 1KTT,1QHR,1QJ1,1QJ6, 1C5L,1C5N, 2HRP,1LHC, 1LHD,1LHE,1LHF,1LHG, 1KLJ,1HH6,1NB5,1ABJ, 1C9T,1Q3J,1B7X, 1JXX, 1BIP,1AWF, 1DW9,1MUE, 1CCD,1AI8,3TPI,1NB1, 1CBW,2IGF,1KBF,1PYO, 1GG6,1IY6,7PTI,1NU7, 1TFG,1OB4,1OB6,1OB7, 1UVT,1A46,1QM7,1A3E, 1JKZ,1SKG,1MF4,1CHL, 1YPC,1R2M,1PWV,3BTK, 1EB1,1CTX,1B08, 2CWG, 1JXW,1EJM,1GGD,1LQO, 1DIT,1C5M,2PLE,1CA0, 1BDW,1NT1,1TA2,1TA6, 1BMN,1JVQ,1VIT,1MIC, 1NIP,1HIN,2PLV,1PI2, 1BA8,1OKX,1LV4,1DP5, 1TMB,1HJA,1LYN,1CRN, 1P2Q,1I25,1AAL,1YPB, 3BTD,1D3Q,1NRO,3BTE, 1HAO,1HAP,1SGI, 1SFQ, 2F5B,3HAT,1FF4,1E4R,	1DSQ, 1NCQ, 2BPA, 1R08, 1R09,2R04 , 2R06, 4RHV, 1RMU, 2RMU, 2RR1, 2RS1, 2RS3, 2RS5, 1RUC, 1RUD, 1RUE, 1RUF, 1RUG, 1RUH, 1RUI, 1RUJ, 1VRH, 1HRV, 2HWB, 2HWC, 1PON, 1VBD, 3CK0, 1NA1, 1OOP, 1PO2, 1AR6, 1AL2, 1AR9, 1RVF, 1AR7, 1ASJ, 1M06, 1AR8, 1JSH, 1TMF, 1QR9,	

1AJJ,1RZZ,1MPF,1CWD, 1QKO,1LVZ,1J9V,1M0L, 1N2R,1QEW,1NJQ,1A6A ,1K3A,1NQL,1KB8,1OM9 , 1FG2, 1A28,1IXF,1UJJ, 1M05,1KG9,1SHB, 1VAC,1K5N, 1KLU,			1RSO,1AVF,1CA8,1HGT, 1E1H,1IDL,1RON,1K21, 1K22,1C4V,1TBZ,1PJU, 1CKS,1CVW,1CBW,1AL4, 1D3T,1FI2,1D4P,1SG8, 4VGC,1D3D,1D3P,1K6U, 1AB9,1N9D,1MVJ,1VGC, 1P2N,1F83,1K5C,1P2O, 1AN1, 1GCT,3GCT,1AHL, 1VWH,1VWP,4HTC,1BFZ, 1K9B,1C9T,1B5G,1DWD, 1UVU,1SSC,1WBR,2HPQ, 1FPC,1A3B,3BTG,1IMX, 1SHH, 1RJT,1JQ9,1BXZ, 1OYT,1SKZ,1HDT,1DFY, 1JRR,1UVS,1DFZ,1HXF, 1MU8, 1C9T,1AHT,1A5G, 1AIX,1OXG,1SLE,1PBI, 1GMK,1YUF,1E4X,1HBT, 1D6W,1HXE,1MU6,1THS, 1QUR,1VR1,2SOC,1VLK, 2VGC,1DWE,1TMU,1SSB, 1FB9,1G0V,1AKS,1PFX, 1DWB,1DWC,1TMT,1HIM, 1RVT,1AD8,1CNO,1PCE, 1UMA,1HGT,2HPP,1AWH, 1NRN,1BR8,1MVI,1G6X, 1CTL,1TT3,2LET,1NTV, 1LD5,1GL1,1TG4,1ECI, 1SMF,1TDV, 1PZ5,1ODP, 1HAG,1HAH,1SRN,1ATE, 1EL6,1PYO,1FEO,1ALF, 3BTH,1SSA,1LDT,3BTT, 3BTQ,1ODR,1GEA,2UTG, 1KCS,1C5W,1C5X,1C5Y, 1C5Z,1GJS,1IHS,1EOF, 2SH1,1YPA,1ZFO,1KC4, 1C4Y,1D9I,1THR,1DX5, 1HX2,1YUG,1SHH,1PWB, 1DLK,1HUT,5GDS,1TFS, 1DW9,1MVU,1PLW,1PLX, 1JOT,1AG7,1RD3,1HAI, 1BTH,1HAJ,1O0D,1VWO, 2CCO,1ATD,1NXN,1ZWA, 1I8K,3EZM,1ORX,1SHI, 1DM4,1H34,7HVP,1BNB, 1A2C,1H8U,1JW6,4THN, 1PIP,1NZQ,1AVP,1NLN, 1HOY,3F58,1EPM,1A3B, 1AFE,1DWB,1C5N,1KAT, 2THF,3SRN,1UGL,2F58, 1SDX,1LBJ,4SRN,7KME, 8KME,1ROO,1QJ1,1SJU, 1FRG,1I3Z,1GHV,1LSL, 1CZZ,1GJC,1WHE,1HIT, 1UH0,1UH1,1DSR,1O3P,		

			1GJB,1B1V,1BXP,1FPC, 1C5L,1TTK,1KJ6,1P9Z, 1ONU,1GQ0,1KUW,1O2G, 1P2K, 1P8B, 1TTL,1GI7, 1GI8,1GI9,1BX7,4MON, 1CPI,1EAI,1E4S,1P9G, 1CZ0,1GHX,1RST,1HEF, 1SB1,1BVN,1AFQ,1GM2, 1A2X,1Q1J,1BXJ,1J7V, 1ICY,1JP5,2TGI,1IYC, 2REL,1GHY,1TK2,1KTZ, 1Q2K,1VQC,1N12,1PW9, 1OMN,1GGI,3BTF,1OX1, 1AY6,1LPA,1FQQ,3BTM, 1LOM,1HRT,1FYG, 1GJ7, 1GJ9,1UVQ,1MKW,1LVM, 1JGK,1BIK,1OOK,1ODQ, 1CXR,1HRL,1B9Q,1NO9, 1I8I,1R8O,1OMG,1MKX, 1N6T, 1AV3,1E4T,1PAL, 1JGE,		

Appendix 3

A: Significance of the differences between subclasses of transmembrane proteins and amyloidal fiber forming proteins. A single asterisk (*) denotes $P < 0.05$, two asterisks (**) denotes $P < 0.01$ and three asterisks (***) denotes $P < 0.001$).

	TM Alpha	TM Alpha 1	TM Alpha 2	TM Alpha 3	TM Alpha buried	TM Beta
AFP	***	***	***	***		***

B: Significance of the differences between subclasses of globular proteins and amyloidal fiber forming proteins. A single asterisk (*) denotes $P < 0.05$, two asterisks (**) denotes $P < 0.01$ and three asterisks (***) denotes $P < 0.001$).

	AFP
Globular 1	***
Globular 2	***
Globular 3	***
Globular 4	***
Globular 5	***
Globular 6	***
Globular 7	***
Globular 8	***
Globular 9	***

Globular 10	***
Globular 11	***
Globular 12	***
Globular 13	***
Globular 14	***
Globular 15	***
Globular 16	***
Globular 17	***
Globular 18	***
Globular 19	***
Globular 20	***
Globular 21	***
Globular 22	***
Globular 23	***
Globular 24	***
Globular 25	***
Globular 26	***
Globular 27	***
Globular 28	***
Globular 29	***
Globular 30	***
Globular 31	***
Globular 32	***
Globular 33	***
Globular 34	***
Globular 35	***
Globular 36	***
Globular 37	***
Globular 38	***
Globular 39	***
Globular 40	***
Globular 41	***
Globular 42	***
Globular 43	***
Globular 44	***
Globular 45	***
Globular 46	***
Globular 47	***
Globular 48	***
Globular 49	***
Globular 50	***
Globular 51	***
Globular 52	***
Globular 53	***
Globular 54	***

Globular 55	***
Globular 56	***
Globular 57	***
Globular 58	***
Globular 59	***
Globular 60	***
Globular 61	***
Globular 62	***
Globular 63	***
Globular 64	***
Globular 65	***
Globular 66	***
Globular 67	***
Globular 68	***
Globular 69	***
Globular 70	***
Globular 71	***
Globular 72	***
Globular 73	***
Globular 74	***
Globular 75	***
Globular 76	***
Globular 77	***
Globular 78	***
Globular 79	***
Globular 80	***
Globular 81	***
Globular 82	***
Globular 83	***
Globular 84	***
Globular 85	***
Globular 86	***
Globular 87	***
Globular 88	***
Globular 89	***
Globular 80	***
Globular 91	***
Globular 92	***
Globular 93	***
Globular 94	***
Globular 95	***
Globular 96	***
Globular 97	***
Globular 98	***
Globular 99	***

Globular 100	***