IDENTIFICATION OF NEW NEURODEGENERATIVE DISEASES

Author: Marta Diaz Caballero marta.diazc@e-campus.uab.cat - Tutor: Salvador Ventura - Biochemistry Degree
Universitat Autònoma de Barcelona (UAB)

1.Introduction:

In recent years, neurodegenerative diseases are related to prionic aggregates formation. This work discusses how to predict prion-like domains from protein amino acidic composition and then, use bioninformatic algorithms to select a couple of proteins and search in bibliography what are the knowledge about them and their relation with disease or relate them to disease which could be involved with.

2.Prion-like domains and their prediction:

A prion is an infectious protein or an infectious protean fragment, called prion-like domain capable to make aggregates involved in transmissible neurodegenerative diseases. These are low-complexity regions, Q, N and hydrophobic amino acids rich regions, and also lack in charges and P.

Their conservation suggest they can also accomplish important functions inside the cell, like transcription, translation or non-membrane cell compartments formation.

Various investigation groups have develop specific algorithms that allow differentiate these regions forming prionic aggregates form those sequences which generates non-prionic amyloid aggregates. The algorithms used in this work were PAPA and PrionScan. Both predict these prion-like domains from the compositional protein characteristic of complete proteomes. We started from 824 human proteins previously selected and then we compare the results obtained in the follow graphics (Figures 1 and 2).

3.Correlation graphics:

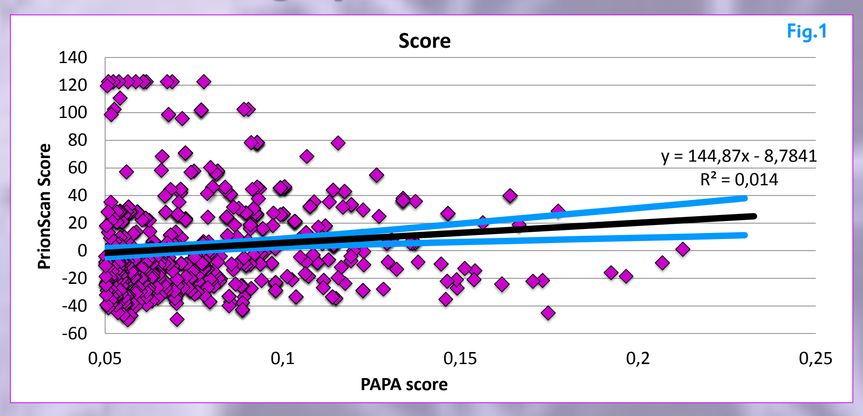

Fig.1

Figure 1: This graph shows algorithms score correlation. As is visible in the graph, both algorithms do not have any correlation for score punctuation. Discussion of these results is showed in the next paragraph. Blue lines indicates statistical analysis with 95% of confidence.

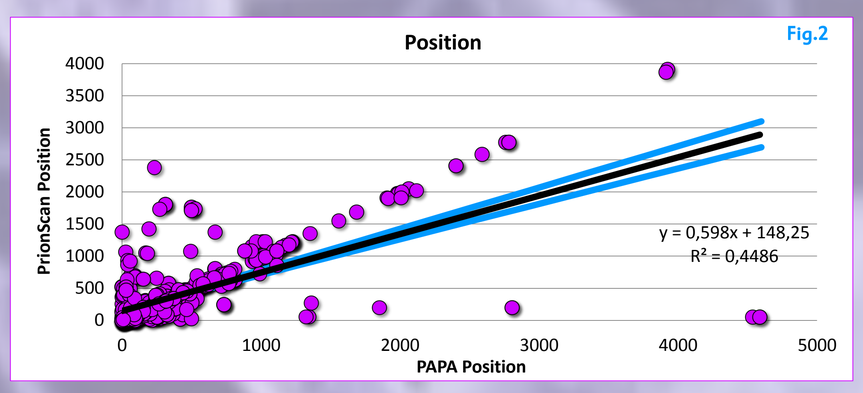

Fig.2

Figure 2: This graph shows algorithms position correlation. At first appearance, it seems there is no correlation between both algorithms, but considering only central points and avoiding peripheral ones, the observation shows there is correlation in reference to position. Blue lines indicates statistical analysis with 95% of confidence.

4. Correlation score and position discussion:

The main difference between these algorithms falls into LOr values for each amino acid. Table 1 shows both libraries. For PAPA, the most important amino acids are the hydrophobic, inclusive aromatics, whereas for PrionScan most important amino acids are Q and N. So, PAPA predicts regions that form non-prionic aggregates like prionic domains, results that are false positives. Otherwise, PrionScan cannot detect these regions as prion-like domains but these that have elevate contain in Q/N. In other words, PAPA is an algorithm with higher sensibility, but detects more false positives, and is less precise. However, PrionScan has lower sensibility but detects less false positives, so it is more precise.

Despite these differences, I have chose two proteins, TDP43 and MED15 (punctuations shown in Table 2).

Table 1

Residue	Prion domains		Prion domain (Library 1)	
	Odds ratio	LOr	Odds ratio	LOr
A	0.675	−0.568	0.670	−0.578
C	0.071	−3.807	1.520	0.604
D	0.352	−1.507	0.280	−1.837
E	0.147	−2.766	0.550	−0.862
F	0.718	−0.478	2.310	1.208
G	1.028	0.040	0.960	−0.059
H	0.913	−0.131	0.760	−0.396
I	0.350	−1.515	2.260	1.176
K	0.271	−1.883	0.210	−2.252
L	0.340	−1.556	0.960	−0.059
M	1.125	0.170	1.960	0.971
N	5.700	2.511	1.080	0.111
P	1.170	0.227	0.300	−1.737
Q	4.125	2.044	1.070	0.098
R	0.436	−1.196	0.670	−0.578
S	1.662	0.733	1.140	0.189
T	0.830	−0.268	0.890	−0.168
V	0.304	−1.716	2.260	1.176
W	0.091	−3.459	1.950	0.963
Y	1.724	0.786	2.180	1.124

Table 1: Libraries from PrionScan (Prion Domains) and PAPA (Prion Domains Library 1). There are shown Odds ratio and LOr values for each amino acid.

5. Algorithms results:

Algorithms / Proteins	PrionScan		PAPA		
	Score	Position	Score	Position	Fold Index
TDP43	32,822	341	0,04	366	0
MED15	78,334	159	0,09	175	-0,3

Table 2: Algorithms results. In this table are shown algorithms punctuations for score and the prion-like predicted position. The predicted positions prove both proteins contain this prion-like domain, and Fold Index obtained with PAPA indicates that are unfolded regions. Score differences are explained above. The final results differs because algorithms use different formulas to calculate these scores.

5. TD43:

TDP43 is a 414 residues and 43 kDa human protein that belongs to TARDBP family. This protein is capable to bind DNA and RNA and is involved in very important functions in all the cells like general RNA regulation, alternative splicing, RNA nuclear to cytoplasm exportation, transcription and translation. In neurons, this protein has specific functions like RNA and microRNA transport from neuronal soma to axonal terminal. There are evidences which prove TDP43 prion-like domain has functional implications.

TDP43 are involve in amyotrophic lateral sclerosis (ALS) and fronto-temporal lobar disease (FTLD). In both, TDP43 forms prionic-aggregates which can cause toxicity and neuron death. TDP43 aggregates contains truncated-TDP43 (corresponting to C-terminus fragments (CTF)) fragments by caspase-3 and full-length-TDP43. Aggregates also are hyperphosphorylated and ubiquitylated. Despite all of these possibilities, prion-like mechanism initiation, and their implications with diseases still unclear.

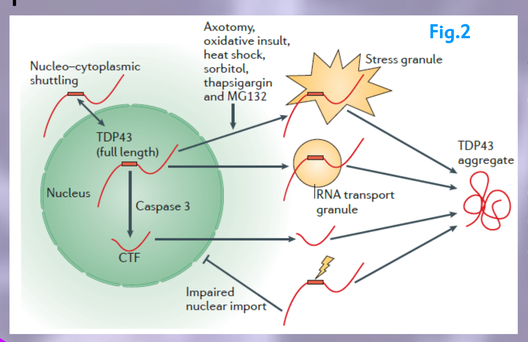

Fig.2

Figure 3: TDP43 modification, stability and turnover. In this figure is shown TDP43 possible mechanism to form aggregates. Stress conditions and diverse stimulus induces stress granules formation, increasing aggregates formation ability. Also mutations in NLS and disease-associated mutations (shown by a yellow lightning bolt) can increase this ability. Truncated CTF-fragments are other possible mechanism to initiate aggregates formation.

6.MED15:

MED15 is a 788 residues and 87 kDa human protein. MED15 belongs to MED family, and are one of RNA polymerase II complex mediator members. This complex has a very important function: to transmit information DNA to RNA polymerase II.

MED15 has been involved in diseases like DiGeorge syndrome, obesity, type-2 diabetes and cardiovascular diseases. Despite the implication in these implications, there still are not evidences of MED15-aggregation in these diseases. An hypothetical evidence of prion-like current functions is this domain is localized in the middle of ARC105 domain, which is essential to DNA-protein and protein-protein binding. The prion-like domain presence and conservation between species may think this domain could be involved into forming these interactions. Mutations and deletions could increase prion aggregates formation and could disregulate the cell translation system.

7.Conclusions:

- Scientist try to reveal these possible prion-like domains using algorithms. The main problem is to find the correct combinations of punctuations, because each amino acid has different contribution in prionic aggregates. In order to obtain better predictions, the aim is to design an algorithm that gives the correct weight to each amino acid considering compositional features of these prion-like regions.
- TDP43 has clear evidence to be involved into prion aggregates formation. The investigation has to be focus on find the initiation aggregation mechanism and aggregates implication in disease. MED15 has to be investigated in order to relate it with diseases with prionic aggregates presence.
- Prion-like domains could be a good tool in cell interacting mechanisms and also in adaptive processes, where could proportionate adaptive advantages. These could be good reasons to their conservation along the evolution of species. But the ability and the implication in neurodegenerative and other types of diseases is the main point of the issue which has to follow under investigation.

References:

[1] Malinovska, L., Kroschwald, S., & Alberti, S. (2013). Protein disorder, prion propensities, and self-organizing macromolecular collectives. *Biochimica et Biophysica Acta*, **1834**(5), 918–31. doi:10.1016/j.bbapap.2013.01.003
[2] Toombs, J. a, Petri, M., Paul, K. R., Kan, G. Y., Ben-Hur, A., & Ross, E. D. (2012). De novo design of synthetic prion domains. *Proceedings of the National Academy of Sciences of the United States of America PNAS*, **109**(17), 6519–24. doi:10.1073/pnas.1119366109
[3] Espinosa Angarica, V., Ventura, S., & Sancho, J. (2013). Discovering putative prion sequences in complete proteomes using probabilistic representations of Q/N-rich domains. *BMC Genomics*, **14**, 316. doi:10.1186/1471-2164-14-316
[4] Lee, E. B., Lee, V. M.-Y., & Trojanowski, J. Q. (2012). Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. *Nature Reviews Neuroscience*, *13*(1), 38–50. doi:10.1038/nrn3121
[5] Yang, F., Vought, B. W., Satterlee, J. S., Walker, A. K., Jim Sun, Z.-Y., Watts, J. L., DeBeaumont, R., Saito, R. M., Hyberts, S. G., Yang, S., Macol, C., Iyer, L., Tijan, R., van den Heuvel, S., Hart, A. C., Wagner, G., & Näär, A. M. (2006). An ARC/Mediator subunit required for SREBP control of cholesterol and lipid homeostasis. *Nature*, *442*(7103), 700–4. doi:10.1038/nature04942