ProADD: A database on Protein Aggregation Diseases

Ramesh Shobana & Eswari PJ Pandaranayaka*

Centre of Excellence in Bioinformatics, School of Biotechnology, Madurai Kamaraj University, Madurai -625021,India; Eswari PJ Pandaranayaka – Email: eswari@mkustrbioinfo.com; *Corresponding author

Abstract:
ProADD, a database for protein aggregation diseases, is developed to organize the data under a single platform to facilitate easy access for researchers. Diseases caused due to protein aggregation and the proteins involved in each of these diseases are integrated. The database helps in classification of proteins involved in the protein aggregation diseases based on sequence and structural analysis. Analysis of proteins can be done to mine patterns prevailing among the aggregating proteins.

Availability: http://bicmku.in/ProADD

Keywords: protein, aggregation, intrinsic disorder, disease, database, ProADD

Background:
The potential of a protein to assume a functional conformation determines the ability of the protein to perform its fundamental function in a cell. Proteins often appear to be misfolded in protein aggregation diseases [1]. There may be several kinds of aggregates, including disordered or 'amorphous' aggregates but amyloid fibrils are most prevalent. Increase in the knowledge of diseases like Alzheimer disease (AD), Parkinson disease (PD), Huntington disease (HD), amyotrophic lateral sclerosis (ALS) and prion diseases has helped to realize that common cellular and molecular mechanism such as protein aggregation exists among them [2]. Yet the mechanism is not well understood. Recently there are many reports on intrinsically disordered regions in many proteins which do not fold into stable three dimensional (3D) structures under physiological conditions. These regions occur in clusters of non-cooperative interchanging conformations. The atom coordinates and the backbone Ramachandran angles in these regions vary with time and has no specific equilibrium values. [3]. Along with aggregation, looking into the intrinsic disorderness of the proteins involved in these diseases might help in elucidating the mechanism of the disease.

Methodology:
Data collection:
Data about the proteins involved in aggregation diseases were collected through literature search and UniProt [4] keyword search. The database integrates data on 12 protein aggregation diseases with around 600 proteins involved in them. The basic information regarding the proteins involved in aggregation diseases such as protein name, gene name, protein length, protein family and availability of three dimensional structures were collected from UniProt for each protein.

Database interface:
The front end of the database was designed with HyperText Markup Language (HTML) for creating web pages and Cascading Style Sheets (CSS), a stylesheet language, for enriching the look and format of the web pages. MySQL was used to create the back end of the database owing to its cross platform accessibility, high-performance and scalable web-

BIOINFORMATION

based and embedded database applications. Hypertext Preprocessor (PHP) was used to generate dynamic page content because of its ability to be used as a general purpose scripting language, especially suited for web applications and as it can be embedded in HTML. The database was developed with

XAMPP (Linux Apache MySQL PHP) package. The web interface can be accessed through the home page of database **(Figure 1A)**. The database also contains a list of aggregation diseases **(Figure 1B),** proteins involved in each disease (**Figure 1C**) and so on.
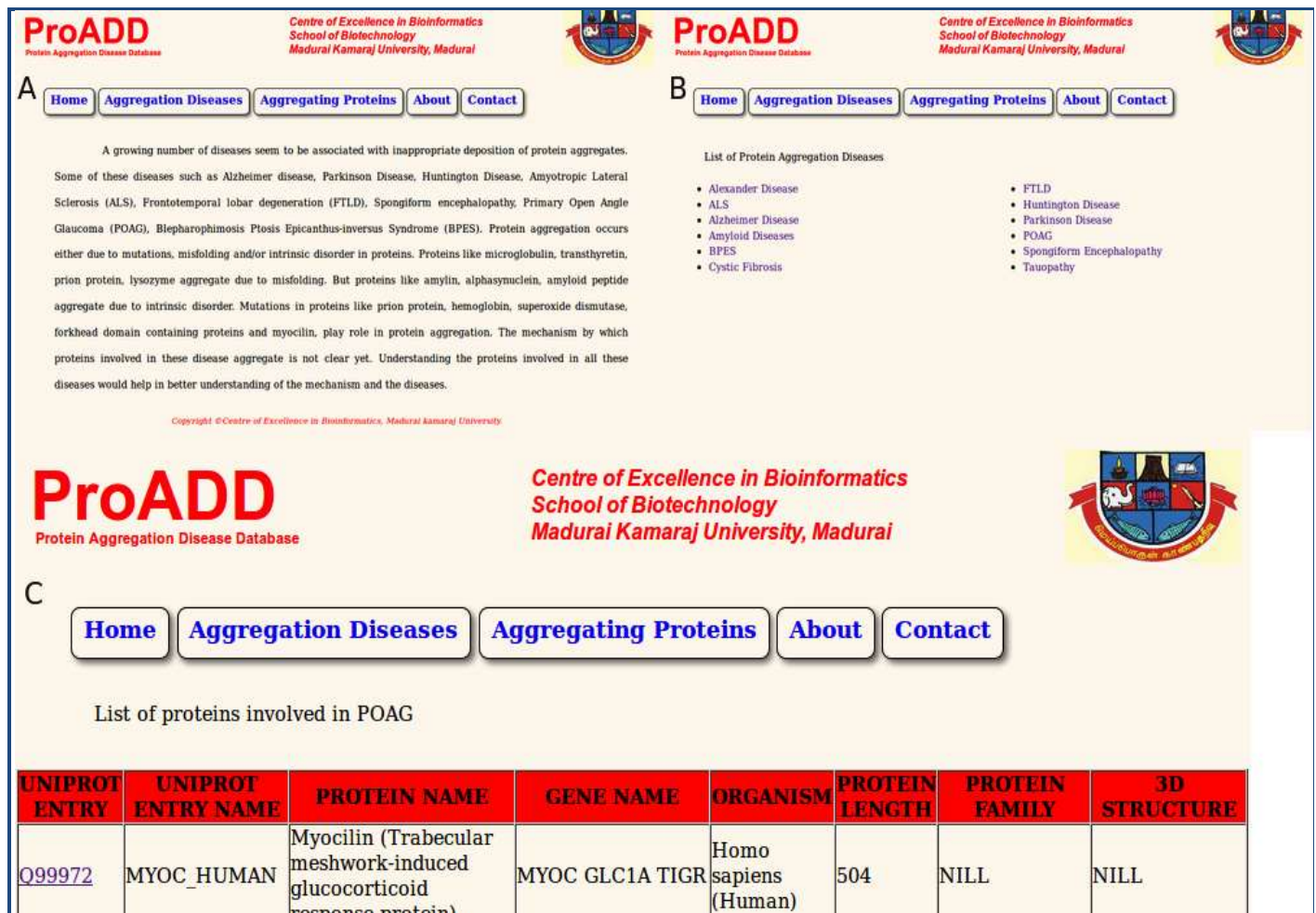

Figure 1: A) Home page of ProADD; **B)** List of aggregation diseases; **C)** Proteins involved in one of the aggregation diseases POAG

Prediction of aggregating proteins:
Proteins involved in one of the aggregation diseases, Primary Open Angle Glaucoma (POAG), were considered for the prediction. The proteins are classified into various groups. The protein sequences were clustered based on their identity score using CD-HIT suite **[5]** and were found to be highly varied. The sequences were further analysed for the prediction of aggregation-prone segments using Normalised Hot Spot Area (NHSA) of Aggrescan **[6].** Intrinsic Aggregation Scale **(**IAS**),** the propensity scale used to predict the regions prone to constructive aggregation was also used **[7].** The ratio between the NHSA value and the intrinsic aggregation scale value of POAG proteins was calculated and found that three proteins are prone to aggregation.

Prediction of intrinsic disordered proteins:
DisProt **[8]** was used to find the intrinsically disordered regions in a protein based on the amino acid compositions. Further the structures of the proteins were extracted from Protein Data Bank (PDB) **[9].** The structural details were analysed for finding the missing residues. Studies have shown that in general

regions of missing electron density are short in the PDB, as long flexible regions hinder the process of crystallization **[10].** Thus twenty two proteins were found to have intrinsic disorder regions

Conclusion:
The proteins known to be involved in POAG do not show any similarity at the sequence level, as also seen in the case of amyloid disease caused due to protein aggregation **[11].** Based on further sequence analysis, oculomedin, caveolin1 and caveolin2 are predicted to be aggregating proteins and 22 proteins were predicted to have intrinsic disorder regions. Tank binding kinase, neurotrophin, apolipoprotein are predicted to be intrinsic disorder proteins based on structural analysis. Detailed results will be published elsewhere. Further in depth computational and experimental analysis of these proteins may help in understanding their role in the disease.

Features:
There is no database available related to protein aggregation diseases till date. The database was developed with the aim of

BIOINFORMATION

integrating data available for various aggregation diseases to further enhance the study of the diseases. In addition to this, the classification of proteins to various groups namely aggregating proteins, intrinsically disordered proteins will be made available.

Further Development:

The primary interactors of the proteins involved in the aggregation diseases would be integrated and analysed further to study about the interactions prevailing among the proteins involved in aggregation diseases. It will be useful in studying the molecular interactions involved in the aggregation diseases. Searching of queries need to be included for easy retrieval of data. Various online resources like PDB, STRING, UniProt would be linked to make the database a complete web source for aggregation diseases.

Acknowledgement:

References:

[1] Uversky VN, *Cell Mol Life Sci.* 2003 **60**: 1852 [PMID: 14523548]

[2] Ross CA & Poirier MA, *Nat Med*. 2004 **10:** S10 [PMID: 15272267]

[3] Click TH *et al. Int J Mol Sci*. 2010 **11**: 5292 [PMID: 21614208]

[4] http://www.uniprot.org/

[5] http://weizhong-lab.ucsd.edu/cdhit_suite/cgi-bin/index.cgi?cmd=cd-hit/

[6] http://bioinf.uab.es/aggrescan/

[7] Mary S *et al. J Prot Proteomics*. 2014 **5**: 49

[8] Van der Lee R *et al. Chem Rev*. 2014 In press [PMID: 24773235]

[9] http://www.pdb.org/

[10] Habchi J *et al. Chem Rev* 2014 In press [PMID: 24739139]

[11] Koo EH *et al. Proc Natl Acad Sci U S A*. 1999 **96**: 9989 [PMID: 10468546]