Virus Research xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Predicting the aggregation propensity of prion sequences

Alba Espargaró[a,b], Maria Antònia Busquets[a,b], Joan Estelrich[a,b], Raimon Sabate[a,b,*]

[a] Department of Physical Chemistry, Faculty of Pharmacy, University of Barcelona, Avda. Joan XXIII 27-31, E-08028 Barcelona, Spain
[b] Institute of Nanoscience and Nanotechnology of the University of Barcelona (IN²UB), Spain

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Amyloid prediction
Prion prediction
β-Sheet prediction
Amyloid algorithm
Hot-spot

ABSTRACT

The presence of prions can result in debilitating and neurodegenerative diseases in mammals and protein-based genetic elements in fungi. Prions are defined as a subclass of amyloids in which the self-aggregation process becomes self-perpetuating and infectious. Like all amyloids, prions polymerize into fibres with a common core formed of β-sheet structures oriented perpendicular to the fibril axes which form a structure known as a cross-β structure. The intermolecular β-sheet propensity, a characteristic of the amyloid pattern, as well as other key parameters of amyloid fibril formation can be predicted. Mathematical algorithms have been proposed to predict both amyloid and prion propensities. However, it has been shown that the presence of amyloid-prone regions in a polypeptide sequence could be insufficient for amyloid formation. It has also often been stated that the formation of amyloid fibrils does not imply that these are prions. Despite these limitations, *in silico* prediction of amyloid and prion propensities should help detect potential new prion sequences in mammals. In addition, the determination of amyloid-prone regions in prion sequences could be very useful in understanding the effect of sporadic mutations and polymorphisms as well as in the search for therapeutic targets.

1. Introduction

Cells contain a high concentration of macromolecules including complex sugars, nucleic acids and proteins. In this crowded environment, protein folding is extraordinarily complicated: constant collisions between molecules, the high protein concentration and the delicate balance between expression and folding all affect the outcome (Ellis and Minton, 2003; Jahn and Radford, 2008). In this complex environment, protein folding and aggregation are competing pathways. A lack of regulation of the folding–aggregation equilibrium may lead to the formation of well-organized protein aggregates termed amyloids, which present a highly ordered and repetitive conformation with a common core formed of β-sheet structures oriented perpendicular to the fibril axes that form a characteristic structure known as a cross-β structure (Chiti and Dobson, 2006). Amyloid formation has increasingly been associated with severe diseases grouped under the term "conformational diseases". There are currently more than 40 proteins, without any common amino acid sequence, native structure or function, which have been linked to these diseases. In humans, the pathologies can be classified according to the location of the aggregates as non-neuropathic amyloidosis (either localized or systemic) *e.g.* type II diabetes and cataracts; or neurodegenerative human disorders *e.g.* Alzheimer's disease and Parkinson's disease (Chiti and Dobson, 2006; Selkoe, 2003). Interestingly, the large number of proteins capable of folding into the cross-β structure suggests that these structures may be omnipresent and linked to intrinsic properties of the polypeptide chains (Dobson, 2003). It is already known that these conformations have specific functions in a variety of organisms, from bacteria to humans (Berson et al., 2003; Chapman et al., 2002; Fowler et al., 2007; Kelly and Balch, 2003; Maji et al., 2009). Indeed, if amyloid folding is omnipresent, the number of human genetic illness associated with misfolding and aggregation could be much larger than previously thought.

Prions are infectious proteins which, after conformational conversion, self-assemble into amyloid structures that are capable of self-replicating their conformation *in vivo*, thus becoming neurotoxic and protein-based genetic elements in mammals and fungi, respectively (Aguzzi and Calella, 2009; Prusiner, 2001; Uptain and Lindquist, 2002). In mammals, prions are related to the onset of neurodegenerative diseases known as transmissible spongiform encephalopathies (TSEs) including scrapie in sheeps and goats, and Creutzfeldt–Jakob disease in humans (Prusiner, 1982). These pathologies are caused by the accumulation of prion protein (PrP) that can be found in two different conformations: the physiological form (PrP^C), rich in α-helical structure; and the infectious form

* Corresponding author at: Departament de Fisicoquímica, Facultat de Farmàcia, Universitat de Barcelona, Avda. Joan XXIII 27-31, E-08028 Barcelona, Spain.
Tel.: +34 93 402 4559; fax: +34 93 403 5987.
E-mail address: rsabate@ub.edu (R. Sabate).

http://dx.doi.org/10.1016/j.virusres.2015.03.001
0168-1702/

Table 1
Amyloid and prion prediction methods.

β-Aggregation prediction algorithms	
Empirical methods	
Chiti and Dobson	Equation
DuBay et al.	Equation
Pawar et al.	Equation
Zyggregator	Server: (http://www-vendruscolo.ch.cam.ac.uk/zyggregator.php)
TANGO	Server: (http://tango.crg.es/)
Tartaglia et al.	Equation
AGGRESCAN	Server: (http://bioinf.uab.es/aggrescan)
SALSA	Equation
Pafig	Server: (http://www.mobioinfor.cn/pafig/)
Structured-based methods	
PASTA	Server: (http://biocomp.bio.unipd.it/pasta/)
FoldAmyloid	Server: (http://antares.protres.ru/fold-amyloid/)
NetCSSP	Server: (http://cssp2.sookmyung.ac.kr)
BETASCAN	Server: (http://betascan.csail.mit.edu)
AmyloidMutants	Server: (http://amyloid.csail.mit.edu/)
STITCHER	Server: (http://stitcher.csail.mit.edu)
Amyloid prediction methods	
ZipperDB	Server: (http://services.mbi.ucla.edu/zipperdb/)
PreAmyl	Server: (ftp://mdl.ipc.pku.edu.cn/pub/software/pre-amyl)
WALTZ	Server: (http://waltz.switchlab.org/)
Consensus methods	
Hamodrakas et al.	Equation
AmylPred2	Server: (http://biophysics.biol.uoa.gr/AMYLPRED2)
Prion prediction algorithms	
Alberti et al.	Equation
PLAAC	Server: (http://plaac.wi.mit.edu/)
PAPA	Server: (http://combi.cs.colostate.edu/supplements/papa/)
PrionScan	Server: (http://webapps.bifi.es/prionscan)
pWALTZ	Server: (http://bioinf.uab.es/pWALTZ)

(PrP^{Sc}), enriched in β-sheet structure. The structural transition from the former to the latter is essential for prion perpetuation (Caughey and Baron, 2006; Chien et al., 2004; Pan et al., 1993; Prusiner, 2001). Meanwhile, fungal prions, ranging from yeast prions such as [PSI^+] and [URE3] from *Saccharomyces cerevisiae* (Uptain and Lindquist, 2002; Wickner et al., 2011) to the [Het-s] prion from the filamentous fungus *Podospora anserina* (Coustou et al., 1997; Maddelein et al., 2002), are related to cellular functions (Uptain and Lindquist, 2002). Importantly, since prions and amyloids share a same macroscopic structure and properties, the differentiation between prionic and amyloid non-prionic sequences is nowadays blurry. This fact enlightened the possibility that some of known amyloid-prone proteins linked to human diseases could really show certain infection capacity. However, since none of known and newly amyloid-prone proteins have yet been shown to be infectious or contagious under natural conditions, these proteins are recently grouped under the term prionoid; showing their potential prion capacity (Ashe and Aguzzi, 2013).

Advances in our understanding of the amyloid structure as well as that of the key factors which promote and stabilize amyloids have led to the construction of over 20 algorithms to predict the amyloid and prion propensity of primary polypeptide sequences (see Table 1) (Belli et al., 2011).

2. β-Aggregation and amyloid prediction algorithms

It is generally accepted that although much of the primary sequence of a protein is involved in the formation of the amyloid core, there are some short amino acid sequences, known as hot-spots (HS), that are more likely to aggregate than others and which can induce aggregation (Lopez de la Paz and Serrano, 2004; Sanchez de Groot et al., 2005). These regions are rich in hydrophobic, aliphatic and aromatic residues (Ivanova et al., 2004; Monsellier et al., 2008; Rousseau et al., 2006). Normally the HS are flanked by specific residues that curtail the aggregation propensity by increasing folding efficiency. The regions composed of these residues are termed "gatekeepers" and are enriched in polar charged residues and prolines (Monsellier et al., 2008; Rousseau et al., 2006). The accumulation of knowledge of the amyloid structures and the forces that induce and stabilize their formation have led to the creation of several algorithms that are capable of identifying amyloid-prone regions in protein sequences, thereby predicting their aggregation propensity (Belli et al., 2011; Castillo et al., 2011; Villar-Pique et al., 2014).

Current amyloid prediction methods have usually divided in two main groups (Belli et al., 2011; Villar-Pique et al., 2014): empirical and structure-based methods. Of the empirical methods, based on experimental or theoretical analysis of amino acid properties and their contribution to amyloid formation, the following algorithms should be highlighted: Chiti and Dobson (Chiti et al., 2003), DuBay et al. (DuBay et al., 2004), TANGO (Fernandez-Escamilla et al., 2004), Tartaglia et al. (Tartaglia et al., 2004), Pawar et al. (Pawar et al., 2005), AGGRESCAN (Conchillo-Sole et al., 2007), SALSA (Zibaee et al., 2007), Zyggregator (Tartaglia and Vendruscolo, 2008) and Pafig (Tian et al., 2009). The structure-based methods take into account the three-dimensional (3D) structures of well-defined fibrillar conformations of some proteins or peptides. These latter algorithms include: Zipper DB (Thompson et al., 2006), PASTA (Trovato et al., 2007; Walsh et al., 2014), FoldAmyloid (Galzitskaya et al., 2006), NetCSSP (Kim et al., 2009; Yoon et al., 2007), PreAmyl (Zhang et al., 2007), BETASCAN (Bryan et al., 2009), WALTZ (Maurer-Stroh et al., 2010), AmyloidMutants (O'Donnell et al., 2011) and STITCHER (Bryan et al., 2012). Importantly, although some of the previously enumerated algorithms were initially designed to check amyloid propensity, nowadays would to be considered that most of them really detect β-aggregation (including amorphous β-aggregates). Thus, we consider that only 3D Profile method (including Zipper DB and PreAmyl) and WALTZ (both structured-based methods) are focused to detect sequences with specific amyloid propensity. For this reason, we divide the previously mentioned algorithms in two main categories: β-aggregation prediction algorithms (subdivided in empirical and structure-based methods) and amyloid prediction algorithms. Additionally, as some of the algorithms are improvements of previous algorithms, these ones are grouped in families to favour the reader understanding. In addition, Hamodrakas et al. (Hamodrakas et al., 2007) and AmylPred2 are amyloid prediction tools, which integrate the results of several of the methods mentioned above, thereby providing consensus predictions (Frousios et al., 2009; Tsolis et al., 2013).

2.1. β-Aggregation prediction algorithms: empirical methods

2.1.1. Chiti and Dobson

It has been stated that single-base mutations in human muscle acylphosphatase (AcP) and other amyloid-prone proteins influence its aggregation provided the aggregation process is carried out under conditions in which they are unstructured. Analysis of the mutation set shows a high degree of correlation between the changes in aggregation rates due to a single amino acid mutation in the protein and the intrinsic properties of the enzyme. These observations lead to identify the factors that affect the intrinsic propensities of proteins to aggregate, predicting that the aggregation propensity of a protein is depending on the hydrophobicity and charge of its sequence. Thus, sequences showing high hydrophobicity and low charge display high aggregation propensity (Chiti et al., 2003).

Please cite this article in press as: Espargaró, A., et al., Predicting the aggregation propensity of prion sequences. Virus Res. (2015), http://dx.doi.org/10.1016/j.virusres.2015.03.001