

WALTZ-DB 2.0: an updated database containing structural information of experimentally determined amyloid-forming peptides

Nikolaos Louros^{1,2}, Katerina Konstantoulea^{1,2}, Matthias De Vleeschouwer^{1,2},
Meine Ramakers^{1,2}, Joost Schymkowitz^{1,2,*} and Frederic Rousseau^{1,2,*}

¹VIB Center for Brain & Disease Research, Switch Laboratory, Leuven, 3000, Belgium and ²KU Leuven, Department of Cellular and Molecular Medicine, Leuven, 3000, Belgium

Received July 14, 2019; Revised August 13, 2019; Editorial Decision August 16, 2019; Accepted August 30, 2019

ABSTRACT

Transition of soluble proteins into insoluble amyloid fibrils is driven by self-propagating short sequence stretches. However, accurate prediction of aggregation determinants remains challenging. Here, we describe WALTZ-DB 2.0, an updated and significantly expanded open-access database providing information on experimentally determined amyloid-forming hexapeptide sequences (<http://waltzdb.switchlab.org/>). We have updated WALTZ-DB 2.0 with new entries, including: (i) experimental validation of an in-house developed dataset of 229 hexapeptides, using electron microscopy and Thioflavin-T binding assays; (ii) manual curation of 98 amyloid-forming peptides isolated from literature. Furthermore, the content has been expanded by adding novel structural information for peptide entries, including sequences of the previous version. Using a computational methodology developed in the Switch lab, we have generated 3D-models of the putative amyloid fibril cores of WALTZ-DB 2.0 entries. Structural models, coupled with information on the energetic contributions and fibril core stabilities, can be accessed through individual peptide entries. Customized filtering options for subset selections and new modelling graphical features were added to upgrade online accessibility, providing a user-friendly interface for browsing, downloading and updating. WALTZ-DB 2.0 remains the largest open-access repository for amyloid fibril formation determinants and will continue to enhance the development of new approaches focused on accurate prediction of aggregation prone sequences.

INTRODUCTION

Protein folding is a crucial process during which polypeptide chains adopt a thermodynamically stable three-dimensional structure that is pivotal for most cellular functions. Proteins that misfold or fail to retain their native tertiary structure are prone to forming amyloid fibril aggregates (1). Amyloids are linked to a growing number of widespread debilitating diseases, including type II diabetes (T2D), atherosclerosis, systemic amyloidosis and capital neurodegenerative diseases, such as Alzheimer's and Parkinson's disease (2,3). On the other hand, recent studies also suggest that toxicity may precede the formation of large fibrous deposits (3). Phase separation has emerged as an alternative mechanism and has been proposed for several proteins associated to neurodegenerative diseases (4,5), suggesting that protein or peptide molecules with specific sequence properties may self-assemble into oligomeric granular modules with increased toxicity (6). At the same time, amyloid formation also serves as a natural scaffold for the formation of molecular superstructures with impressive functional, protective or structural properties, both in humans and other organisms (7). Amyloid aggregation propensity is encoded in the primary structure of protein molecules, hidden within harboured short sequence segments (8–10). These aggregation prone stretches mediate self-assembly of proteins into ordered perpetuating intermolecular β -sheet assemblies known as 'cross- β ' spines, which protrude in parallel orientation to the amyloid fibre axis (11). This conformation comprises an extensive network of backbone hydrogen bonds and a set of laterally inter-fitted side chains excluding water molecules, yielding, thus, a tightly packed and energetically favourable amyloid fibril core (12). Aggregation prone regions are usually integral parts buried within the hydrophobic core of the protein native fold and consequently are often enriched with residues favouring β -strand formation, increased hydrophobicity and low charge content (13). Considering

*To whom correspondence should be addressed. Tel: +32 16372570; Fax: +32 16372751; Email: frederic.rousseau@kuleuven.vib.be
Correspondence may also be addressed to Joost Schymkowitz. Email: Joost.Schymkowitz@kuleuven.vib.be

such sequence propensities, several computational tools have been developed over the years in an effort to accurately predict aggregation potential from polypeptide sequences (8,14–16). This increasing interest has manifested to a considerable growth in experimental data regarding protein self-assembly regions. Putative aggregation mechanisms based on the notion of aggregation hot spots have been proposed for proteins associated to the formation of both functional and disease-associated amyloid fibrils (17–19). Furthermore, synthetic peptide analogues have been developed as novel strategies for the production of antibacterial or anti-tumoural agents (20,21), for the development of transgenic plants with growth phenotypes (22), or as a new source for the development of potent nanomaterials with various applications (23,24). Following this demand, here we describe the fully updated and significantly expanded WALTZ-DB 2.0, the largest publicly available repository for experimentally determined amyloid-forming peptide sequences.

EXPANDED CONTENT AND FEATURE IMPROVEMENTS

New peptide entries and database statistics

WALTZ-DB 2.0 is currently updated to store 1416 hexapeptide entries, divided into nine distinct subsets of origin. In total, 512 peptides have experimentally determined amyloid-forming properties, whereas 904 peptides self-assemble into aggregates with amorphous morphological characteristics. Two novel peptide subsets were added during this update, containing individual peptide mutation screens of known aggregation prone stretches derived from tau and apolipoprotein A-I, which are known amyloid-forming proteins associated to neurodegeneration (25) and atherosclerosis (26). The subsets, designated as *tau mutant set* and *apoAI mutant set*, are composed of 114 and 115 hexapeptides, respectively, and were systematically developed following a single mutation strategy along all available residue positions. Hexapeptide additions were classified into the database as amyloid or non-amyloid-forming sequences when corresponding morphologies were identified, using electron microscopy or by producing positive Thioflavin-T (Th-T) binding spectra. In detail, to characterize a peptide as amyloidogenic, we followed the general convention for amyloid-like morphology. Amyloid fibrils are typically long and unbranched with a diameter ranging between 8 and 10 nm and often tend to coalesce laterally forming superhelices or proto-fibrillar ribbons (27). Once bound to the surface of a β -rich amyloid fibril structure, the benzothiazole Th-T dye displays enhanced fluorescence intensity (28). Peptides producing spectra with increased fluorescence maxima at 480 nm were also considered as amyloid-forming sequences. Following the example of the previous version of the database, all experimental data are available online in order to allow users to independently conclude on the amyloid classification of every hexapeptide entry. Complementary to the above, WALTZ-DB 2.0 also includes an additional number of 98 hexapeptides with detailed annotated amyloid-forming properties which were mined from literature, manually curated and added to the previous *Literature* subset of the database. To summarize, this major up-

date included the storage of 327 new peptide entries, out of which 268 hexapeptides were classified as amyloid-forming and 59 were judged as non-amyloidogenic sequences. Peptide data entries remain stored in a MySQL database available through a web server built with the Drupal content management system. This provides the required infrastructure to keep WALTZ-DB 2.0 regularly up to date, as well as to ensure fast and secure access to the stored data. Finally, a refreshed version of the methods is mentioned on the help page of the website.

WALTZ-DB 2.0 novel features

For clarity and in an effort to assist non-experienced users with the evaluation of scoring aggregation properties for peptide entries, predicted aggregation propensities in WALTZ-DB 2.0 are highlighted by specific colouring schemes (Figure 1). In detail, TANGO and WALTZ predictions are shown with a red-to-blue colouring gradient, using thresholds that have been previously reported to provide high specificity (8,15). For sequence hydrophobicity, secondary structure and parallel or antiparallel β -strand formation propensities, positive predictions are shown in blue and negatives in red, respectively. Detailed information on the corresponding threshold values for every individual field can be retrieved from the help page available online.

Every peptide entry stored in the expanded WALTZ-DB 2.0 now contains detailed structural information regarding the putative amyloid fibril core. Utilizing a structural prediction methodology for amyloidogenic sequences developed in our lab, we have analysed all 1416 peptide entries and now provide a structural steric zipper prediction model. Users can download the corresponding models in a PDB format through links that are accessible in the peptide entry pages. Alternatively, WALTZ-DB 2.0 also provides a novel molecular graphics interface panel that is accessible online on every peptide page. This new feature allows users to actively manipulate and investigate peptide model structures on the spot, whilst browsing the online database. The JSmol plugin interface provides several options available for style effects (cartoon, ball and stick, ribbon etc), structural colouring (by secondary structure, residue, element etc) and surface or cavity representation, selection for hydrogen bond or disulphide bridge annotations, as well as modules for structural rotation or background colouring. Adding to this, following structural analysis of the stability of the models and manual curation, all peptide page entries contain information on the energy contributions for the steric zipper models. Energies of all major interactions, such as the contribution of electrostatics, hydrogen bond networking between backbone groups or side chains, solvation energies of residues with hydrophobic or polar properties, as well as Van der Waals packing interactions or potential clashes are attributed to each model entry.

The WALTZ-DB 2.0 database page lists a paged updated table and new filtering steps aiming to make browsing a more user-friendly process. Building on the options of the previous version, the database now offers selections for filtering peptide entries based on availability of Th-T spectral data, as well as energetic stability of the structural models. Using this feature, users can now filter the database and

WALTZ-DB 2.0 DATABASE OF AMYLOID FORMING PEPTIDES

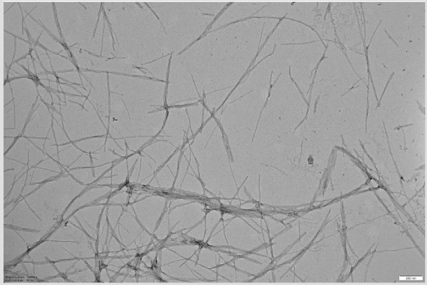
Peptide VQIVYH HOME ABOUT HELP DATABASE CONTACT

WALTZ-DB Login

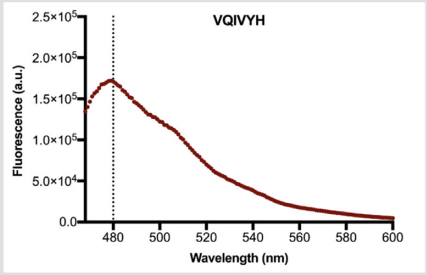
Properties of peptide VQIVYH

Sequence	Class	TEM Staining	Th-T Binding	FTIR Peaks	Proteostat binding	WALTZ	TANGO	PASTA (P)	PASTA (AP)	Hydrophobicity	CF Helix	CF Strand	References
VQIVYH	amyloid	yes	yes	N.A.	N.A.	0.9	0	-4.98	-2.87	66.7	100	140.7	Not published

Electron Microscopy image



Th-T binding



Protein information of peptide VQIVYH

Sequence	UniProt AC	Position	Mutant	Amyloid structures	Homology model	Predicted class
VQIVYH	TAU_HUMAN	623-628	K628H	No structures	Download	1

Energy information of calculated model for peptide VQIVYH

Stability	Backbone Hbond	Sidechain Hbond	VdW	Electrostatics	Solvation Polar	Solvation Hydrophobic	VdW Clashes	Entropy Sidechain	Entropy Mainchain	Electrostatic kon
-32.7	-56.5	-13.6	-71	3.6	87.2	-102.7	1.5	43.4	70.2	2.7

Display Options

- Style: **Ball and Stick**
- Color: **Rainbow**
- Surface: **None**
- H-Bonds
- Rotation
- SS Bonds
- Black Background

```
>sp|P10636|TAU_HUMAN
MAEPFQEFVEMDHAGTYGLGDRKDGQGYTMHQDQEGDTDAGLKESEPLQPTEDGSEEPGSETSDAKSTPTAEVDVAPLIVDEGAPGQAAAQPHTEIPEG
TTAEAGIGDTPSLEDEAAGHVQEPESGKVVQGFLEPGPPGLSHQLMSGMPGALLPEGPREATRQPSGTGPEDETEGRRHAPPELLKQLLGLDGHQEG
PPLKAGAGKERPGSKEEVEDRVDDESSQDPSFSAQDGRPPQTAAREATSIIPGFAEGALPLVDLFSKVSTEIPASEPDGSPVGRAGQDAPLE
FTFHVEITPNVQKEQAHSEELGRAAFPGAPGEGEARGPSLGEDTKEADLPEPSEKQPAAPRGKPVSRVPLKARHVSXKSDGTGSDDKKAKTSTRSS
AKTLKNRCLSPKHPTGSSDPLIQPSSPAVCEPPSPKYSVTSRTGSSGAKEMKLGADGKTIATPRGAAPPQKQANATRIPAKTPPAKTPP
SSGEPFKSGDRSGYSSPGSGPTGSRSLPLPTPTREPKVAVVRTPKSPSASRSLQAPVMPDLKXVSKIGSTENLKHQPGGGKVIINKKLD
LSNVQSKCGSKNDIKHVPGGGSVQIVYKPVVDSLKVTSCOSLGNIIHKPGGGQVEVSEKLDKDRVQSKIGSLDNIITVPPGGGKNIETHKLTFRNAK
AKTDHGAIEVYKSPVVGDTSPRHLNVSSTGSDIMVDSPLQATLAEVDSASLAKQGL
```

Figure 1. Example output of a peptide entry page on WALTZ-DB 2.0. An upper table contains information on the specific peptide sequence and corresponding predicted propensities. Negative predictions are shown in red, whereas positives are indicated in blue. Electron micrographs and Th-T binding spectra are available online for all new peptide entries. A mutation field indicating the position and transitional mutation for peptide entries has been added in the protein information table, along with a field highlighting the steric zipper class for the model prediction. A new table has been added including detailed energy contributions and the overall stability for the structural model. Finally, a JSmol JmolApplet is used to provide an integrated molecular graphics interface displaying the 3D-structural models of the corresponding hexapeptide entries.

isolate individual sequences by searching within a specific range of overall structural stability energies for predicted steric zippers. Notably, a major disadvantage of the previous version was that users could only download the entire database locally. WALTZ-DB 2.0 now allows users to create and access specific entry datasets by combining any of the provided filters and subsequently downloading the resulting list in a CSV, Excel or JSON format, using buttons that are available at the bottom of the listed table.

MATERIALS AND METHODS

Peptide synthesis

Hexapeptides of the new subsets were synthesized using an in-house Intavis Multiprep RSi solid phase peptide synthesis robot capable of parallel synthesis of 24–384 peptides. RP-HPLC purification protocols were used to ensure high levels of peptide purification (>90%). Peptide stock solutions were prepared by dissolving in milli-Q water to a final concentration of 1 mM. Dimethyl Sulfoxide (DMSO) traces (<5%) were used to assist with peptide solubility. The peptide solutions were incubated for 2 weeks at 25°C with shaking prior to analysis of amyloid-forming properties.

Determination of amyloid fibril properties

Transmission electron microscopy was performed to track the morphological properties of the peptide aggregates. Suspensions (5 µl) of peptide aliquots were adsorbed for 1 min to formvar film coated 400-mesh copper grids (Agar Scientific Ltd., England), following a short glow discharge step to improve adsorption. Grids were subsequently washed in 50 µl of milli-Q water and stained with uranyl acetate (2% w/v) for 60 s. Excess stain was removed by blotting with a filter paper. The grids were examined using a JEM-1400 120 kV transmission electron microscope (JEOL, Japan) operated at 80 keV. Amyloid formation was also monitored using Thioflavin-T binding assays. Thioflavin-T (Th-T) is a rotor dye that acts as an efficient reporter of amyloid fibril formation, since it increases its fluorescence when binding to cross-β rich aggregates (28). Th-T (Sigma) was added in low volume black 384-well microplates at a final concentration of 20 µM. Peptide concentration was set to 30 µM. Fluorescence intensity was measured in triplicates, through a ClarioStar plate reader (BMG Labtech, Germany), using an excitation filter at 440 nm and by recording an emission spectrum ranging between 468 and 600 nm. Emission spectra were corrected by subtracting Thioflavin-T - only spectra as background and binding was evaluated by measuring the intensity peak emitted at 480 nm.

Structural models and energy calculations

To provide a structural characterization for the database peptide entries, we have followed the structural topologies of steric zippers introduced by the Eisenberg lab (10,12,29). Representative 3D-model structures were generated utilizing a structural prediction methodology, developed by the Switch lab. Briefly, this pipeline comprises a large dataset of steric zipper hexapeptide fragment templates that have

been extracted from the Protein Data Bank (30). Following implementation of the FoldX energy force field (31), hexapeptide sequences are threaded against all templates, stability energies are calculated and subsequently fed into a random forest classifier. This non-linear classifier then provides a probability estimation of aggregation propensity as a non-trivial function of the corresponding input energies. The threaded structure producing the optimal stability predicted is finally selected as a putative 3D-model representation of the amyloid fibril core.

LINKS TO OTHER DATABASES

Every WALTZ-DB 2.0 peptide entry is linked to a Uniprot ID (<http://www.uniprot.org/>) when information of the parental protein is available (32). For peptide entries directly mined from literature, a corresponding reference link connecting to the PUBMED literature portal (<http://www.ncbi.nlm.nih.gov/pubmed/>) is maintained. Finally, we also provide useful links to other related databases and web servers on protein aggregation (14,29).

SUMMARY

The former release of WALTZ-DB has served as the largest available repository for amyloid-forming short sequence stretches containing experimental annotation (33). It has been used extensively for the development or as a carefully annotated validation set of several high performing predictors of aggregation propensity (15,16,34–40) and has also been utilized as major source of information incorporated in related databases of amyloid aggregation, such as CPAD (41), AmyLoad (42) and AmyPro (43). Following the above, in the current release we have opted to significantly expand the content of the database by simultaneously doubling the coverage of amyloid-forming peptide sequences (512 amyloid sequences compared to 244 previously available), thus providing an improved and more balanced dataset of entries. Furthermore, structural data and novel online features have been added to promote online access to WALTZ-DB 2.0 as a more user-friendly experience and to provide a new layer of information to the users. Finally, we encourage users to help us keep the database up to date by submitting newly identified aggregation-prone hexapeptide sequences using the contact form available online (<http://waltzdb.switchlab.org/contact>).

FUNDING

European Research Council under the European Union's Horizon 2020 Framework Programme ERC Grant Agreement [647458 (MANGO) to J.S.]; Flanders institute for Biotechnology (VIB); University of Leuven; Fund for Scientific Research Flanders (FWO); Fund for Scientific Research Flanders Post-doctoral Fellowship (FWO) [12P0919N to N.L.].

Conflict of interest statement. None declared.

REFERENCES

1. Dobson, C.M. (2003) Protein folding and misfolding. *Nature*, **426**, 884–890.

2. Benson, M.D., Buxbaum, J.N., Eisenberg, D.S., Merlini, G., Saraiva, M.J.M., Sekijima, Y., Sipe, J.D. and Westermark, P. (2018) Amyloid nomenclature 2018: recommendations by the International Society of Amyloidosis (ISA) nomenclature committee. *Amyloid*, **25**, 215–219.
3. Chiti, F. and Dobson, C.M. (2006) Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.*, **75**, 333–366.
4. Babinchak, W.M., Haider, R., Dumm, B.K., Sarkar, P., Surewicz, K., Choi, J.K. and Surewicz, W.K. (2019) The role of liquid-liquid phase separation in aggregation of the TDP-43 low-complexity domain. *J. Biol. Chem.*, **294**, 6306–6317.
5. Wegmann, S., Eftekharzadeh, B., Tepper, K., Zoltowska, K.M., Bennett, R.E., Dujardin, S., Laskowski, P.R., MacKenzie, D., Kamath, T., Commins, C. *et al.* (2018) Tau protein liquid-liquid phase separation can initiate tau aggregation. *EMBO J.*, **37**, e98049.
6. Elbaum-Garfinkle, S. (2019) Matter over mind: Liquid phase separation and neurodegeneration. *J. Biol. Chem.*, **294**, 7160–7168.
7. Fowler, D.M., Koulov, A.V., Balch, W.E. and Kelly, J.W. (2007) Functional amyloid—from bacteria to humans. *Trends Biochem. Sci.*, **32**, 217–224.
8. Fernandez-Escamilla, A.M., Rousseau, F., Schymkowitz, J. and Serrano, L. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.*, **22**, 1302–1306.
9. Lopez de la Paz, M. and Serrano, L. (2004) Sequence determinants of amyloid fibril formation. *PNAS*, **101**, 87–92.
10. Teng, P.K. and Eisenberg, D. (2009) Short protein segments can drive a non-fibrillizing protein into the amyloid state. *Protein Eng. Des. Sel.*, **22**, 531–536.
11. Nelson, R., Sawaya, M.R., Balbirnie, M., Madsen, A.O., Riek, C., Grothe, R. and Eisenberg, D. (2005) Structure of the cross-beta spine of amyloid-like fibrils. *Nature*, **435**, 773–778.
12. Sawaya, M.R., Sambashivan, S., Nelson, R., Ivanova, M.I., Sievers, S.A., Apostol, M.I., Thompson, M.J., Balbirnie, M., Wiltzius, J.J., McFarlane, H.T. *et al.* (2007) Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature*, **447**, 453–457.
13. Buck, P.M., Kumar, S. and Singh, S.K. (2013) On the role of aggregation prone regions in protein evolution, stability, and enzymatic catalysis: insights from diverse analyses. *PLoS Comput. Biol.*, **9**, e1003291.
14. Conchillo-Sole, O., de Groot, N.S., Aviles, F.X., Vendrell, J., Daura, X. and Ventura, S. (2007) AGGRESCAN: a server for the prediction and evaluation of “hot spots” of aggregation in polypeptides. *BMC Bioinformatics*, **8**, 65.
15. Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., Lopez de la Paz, M., Martins, I.C., Reumers, J., Morris, K.L., Copland, A., Serpell, L., Serrano, L. *et al.* (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. *Nat. Methods*, **7**, 237–242.
16. Walsh, I., Seno, F., Tosatto, S.C. and Trovato, A. (2014) PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Res.*, **42**, W301–W307.
17. Louros, N.N., Bolas, G.M.P., Tsiolaki, P.L., Hamodrakas, S.J. and Iconomidou, V.A. (2016) Intrinsic aggregation propensity of the CsgB nucleator protein is crucial for curli fiber formation. *J. Struct. Biol.*, **195**, 179–189.
18. Louros, N.N., Chrysina, E.D., Baltatzis, G.E., Patsouris, E.S., Hamodrakas, S.J. and Iconomidou, V.A. (2016) A common ‘aggregation-prone’ interface possibly participates in the self-assembly of human zona pellucida proteins. *FEBS Lett.*, **590**, 619–630.
19. Tsiolaki, P.L., Louros, N.N. and Iconomidou, V.A. (2018) Hexapeptide tandem repeats dictate the formation of silkworm chorion, a natural protective amyloid. *J. Mol. Biol.*, **430**, 3774–3783.
20. Gallardo, R., Ramakers, M., De Smet, F., Claes, F., Khodaparast, L., Khodaparast, L., Couceiro, J.R., Langenberg, T., Siemons, M., Nystrom, S. *et al.* (2016) De novo design of a biologically active amyloid. *Science*, **354**, aah4949.
21. Khodaparast, L., Khodaparast, L., Gallardo, R., Louros, N.N., Michiels, E., Ramakrishnan, R., Ramakers, M., Claes, F., Young, L., Shahrooei, M. *et al.* (2018) Aggregating sequences that occur in many proteins constitute weak spots of bacterial proteostasis. *Nat. Commun.*, **9**, 866.
22. Betti, C., Vanhoutte, I., Coutuer, S., De Rycke, R., Mishev, K., Vuylsteke, M., Aesaert, S., Rombaut, D., Gallardo, R., De Smet, F. *et al.* (2016) Sequence-specific protein aggregation generates defined protein knockdowns in plants. *Plant Physiol.*, **171**, 773–787.
23. Mankar, S., Anoop, A., Sen, S. and Maji, S.K. (2011) Nanomaterials: amyloids reflect their brighter side. *Nano Rev.*, **2**, doi:10.3402/nano.v2i0.6032.
24. Onur, T., Yuca, E., Olmez, T.T. and Seker, U.O.S. (2018) Self-assembly of bacterial amyloid protein nanomaterials on solid surfaces. *J. Colloid Interface Sci.*, **520**, 145–154.
25. Bloom, G.S. (2014) Amyloid-beta and tau: the trigger and bullet in Alzheimer disease pathogenesis. *JAMA Neurol.*, **71**, 505–508.
26. Louros, N.N., Tsiolaki, P.L., Griffin, M.D., Howlett, G.J., Hamodrakas, S.J. and Iconomidou, V.A. (2015) Chameleon ‘aggregation-prone’ segments of apoA-I: A model of amyloid fibrils formed in apoA-I amyloidosis. *Int. J. Biol. Macromol.*, **79**, 711–718.
27. Kodali, R. and Wetzel, R. (2007) Polymorphism in the intermediates and products of amyloid assembly. *Curr. Opin. Struct. Biol.*, **17**, 48–57.
28. Biancalana, M., Makabe, K., Koide, A. and Koide, S. (2009) Molecular mechanism of thioflavin-T binding to the surface of beta-rich peptide self-assemblies. *J. Mol. Biol.*, **385**, 1052–1063.
29. Goldschmidt, L., Teng, P.K., Riek, R. and Eisenberg, D. (2010) Identifying the amyloids, proteins capable of forming amyloid-like fibrils. *PNAS*, **107**, 3487–3492.
30. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
31. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res.*, **33**, W382–W388.
32. UniProt, C. (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.*, **47**, D506–D515.
33. Beerten, J., Van Durme, J., Gallardo, R., Capriotti, E., Serpell, L., Rousseau, F. and Schymkowitz, J. (2015) WALTZ-DB: a benchmark database of amyloidogenic hexapeptides. *Bioinformatics*, **31**, 1698–1700.
34. Antonets, K.S. and Nizhnikov, A.A. (2017) Predicting amyloidogenic proteins in the proteomes of plants. *Int. J. Mol. Sci.*, **18**, 2155.
35. Burdukiewicz, M., Sobczyk, P., Rodiger, S., Duda-Madej, A., Mackiewicz, P. and Kotulska, M. (2017) Amyloidogenic motifs revealed by n-gram analysis. *Sci. Rep.*, **7**, 12961.
36. Chen, M., Schafer, N.P., Zheng, W. and Wolynes, P.G. (2018) The Associative memory, water mediated, structure and energy model (AWSEM)-amyloimeter: predicting amyloid propensity and fibril topology using an optimized folding landscape model. *ACS Chem. Neurosci.*, **9**, 1027–1039.
37. Niu, M., Li, Y., Wang, C. and Han, K. (2018) RFamyloid: a web server for predicting amyloid proteins. *Int. J. Mol. Sci.*, **19**, 2071.
38. Prabakaran, R., Goel, D., Kumar, S. and Gromiha, M.M. (2017) Aggregation prone regions in human proteome: Insights from large-scale data analyses. *Proteins*, **85**, 1099–1118.
39. Roche, D.B., Villain, E. and Kajava, A.V. (2017) Usage of a dataset of NMR resolved protein structures to test aggregation versus solubility prediction algorithms. *Protein Sci.*, **26**, 1864–1869.
40. Sankar, K., Krystek, S.R. Jr, Carl, S.M., Day, T. and Maier, J.K.X. (2018) AggScore: Prediction of aggregation-prone regions in proteins based on the distribution of surface patches. *Proteins*, **86**, 1147–1156.
41. Thangakani, A.M., Nagarajan, R., Kumar, S., Sakthivel, R., Velmurugan, D. and Gromiha, M.M. (2016) CPAD, curated protein aggregation database: a repository of manually curated experimental data on protein and peptide aggregation. *PLoS One*, **11**, e0152949.
42. Wozniak, P.P. and Kotulska, M. (2015) AmyLoad: website dedicated to amyloidogenic protein fragments. *Bioinformatics*, **31**, 3395–3397.
43. Varadi, M., De Baets, G., Vranken, W.F., Tompa, P. and Pancsa, R. (2018) AmyPro: a database of proteins with validated amyloidogenic regions. *Nucleic Acids Res.*, **46**, D387–D392.