Construction of Amyloid PDB Files Database

Stanković, Ivana; Hall, Michael B.; and Zarić, Snežana D.

Abstract: ***Amyloids are insoluble proteins of a cross-β structure found as deposits in many diseases. They are largely examined structurally, but there is a lack of a unique structural database for amyloid proteins resolved with atomic resolution. We present a constructed amyloid database made on keyword criterion as well as structural features of amyloids described in literature. The searching filter performed by python programming, gave the total number of 109 structures. This database can help further structural general and statistical analysis of amyloids, that can lead to understanding of disease mechanisms related to amyloid proteins.***

Index Terms: Amyloid, Database, PDB

1. *Introduction*

AMYLOIDS are insoluble proteins of a cross-β structure found as deposits in many diseases like Alzheimer's, Parkinson's, Creutzfeldt–Jakob's, type II diabetes, etc. They are also found in normal tissues (nails, spider net, silk) because of their strong fibrilar nature. Among functional nanostructured materials of a significant impact in nanotechnology and biological environments, amyloid fibrils have attracted great attention because of their unique architectures and exceptional physical properties.

All amyloid proteins, independently of their sequence, form very similar structure, the cross-β structure, made of parallel arrays of β-strands, Figure 1. Short polypeptides, of minimum 4 amino acids [1], are self-assembled into β-sheets via backbone hydrogen atoms. Several β-sheets interact with each other in a parallel fashion via polypeptide side chains forming long linear unbranched protofilaments with an axis nearly perpendicular to a polypeptide strand. Several protofilaments, the number being specific to the particular amyloid protein, form fibrils. These structures are different only in the inter-sheet spacing which depends on the side chain size, and in the fibril morphology [2].

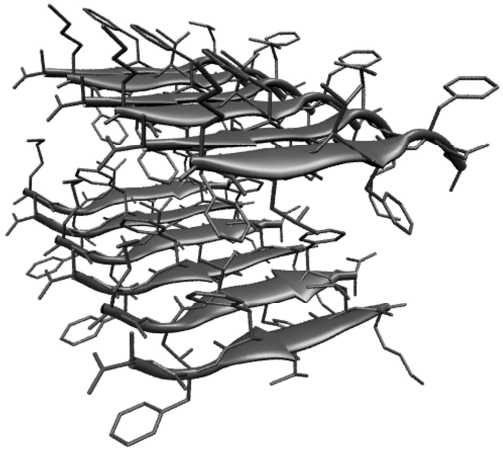

Figure 1: An amyloid structure, PDB ID: 3OW9. Protein β-strands are represented in arrows and the side chains in bars.

Amyloids are largely examined structurally and individually [3, 4, 5], but there is no systematic structural analysis of all resolved structures so far in the literature. There is a lack of a unique structural database for amyloid proteins resolved with atomic resolution. The Protein Data Bank (PDB) consists of nearly 120 000 3D shapes of proteins, nucleic acids and complex assemblies [6, 7]. The PDB contains amyloid structures, but they are hard to find by a simple one criterion search.

PDB files are often not uniform about the amyloid keyword. The molecules in .pdb files are often labeled by another name referring to an amyloid precursor name or a disease name, while the word 'amyloid' could be mentioned within description such as publication title, publication keywords, title section, etc. One example is given in Table 1.

Table 1: Example of a .pdb file header, PDB ID: 2OKZ. The keyword is bolded.

HEADER PROTEIN FIBRIL 18-JAN-07 2OKZ
TITLE MVGGVV PEPTIDE DERIVED FROM ALZHEIMER'S A-BETA

Manuscript received June 27, 2016. This work was supported by the Serbian Ministry of Education, Science and Technological Development [grant number 172065] and NPRP grant from the Qatar National Research Fund (a member of the Qatar Foundation) [grant number NPRP8-425-1-087]. The HPC resources and services used in this work were partially provided by the IT Research Computing group in Texas A&M University at Qatar. IT Research Computing is funded by the Qatar Foundation for Education, Science and Community Development (http://www.qf.org.qa).

I. Stanković is with the ICTM, University of Belgrade, Serbia. M. B. Hall is with the Department of Chemistry, Texas A&M University at Qatar. S. Zarić is with the Department of Chemistry, University of Belgrade, Serbia and the Department of Chemistry, Texas A&M University at Qatar (e-mail: szaric@chem.bg.ac.rs).

COMPND MOL_ID: 1;
COMPND 2 MOLECULE: PEPTIDE FROM ALZHEIMER'S A-BETA;
COMPND 3 CHAIN: A, B;
COMPND 4 FRAGMENT: RESIDUES 35-40;
COMPND 5 ENGINEERED: YES
SOURCE MOL_ID: 1;
SOURCE 2 SYNTHETIC: YES;
SOURCE 3 OTHER_DETAILS: THIS SEQUENCE CORRESPONDS TO RESIDUES 35-40
SOURCE 4 OF HUMAN A-BETA PEPTIDE
KEYWDS STERIC ZIPPER, PROTEIN FIBRIL
EXPDTA X-RAY DIFFRACTION
AUTHOR M.R.SAWAYA, S.SAMBASHIVAN, D.EISENBERG
REVDAT 3 24-FEB-09 2OKZ 1 VERSN
REVDAT 2 05-JUN-07 2OKZ 1 JRNL
REVDAT 1 30-JAN-07 2OKZ 0
JRNL AUTH M.R.SAWAYA, S.SAMBASHIVAN, R.NELSON, M.I.IVANOVA,
JRNL AUTH 2 S.A.SIEVERS, M.I.APOSTOL, M.J.THOMPSON, M.BALBIRNI
JRNL AUTH 3 J.J.WILTZIUS, H.T.MCFARLANE, A.O.MADSEN, C.RIEKEL,
JRNL AUTH 4 D.EISENBERG
JRNL TITL ATOMIC STRUCTURES OF **AMYLOID** CROSS-BETA SPINES
JRNL TITL 2 REVEAL VARIED STERIC ZIPPERS.
JRNL REF NATURE V. 447 453 2007
JRNL REFN ISSN 0028-0836
JRNL PMID 17468747
JRNL DOI 10.1038/NATURE05695
….

Another difficulty in constructing amyloid database is that amyloid proteins exist in different conformations depending on conditions. They might exist in nonamyloidal conformation in solutions when they form helical or random coil secondary structure with no parallel fragments forming fibrils [8].

Here, we present a constructed amyloid database made based on keyword criterion as well as structural criteria.

2. METHODOLOGY

2.1 Online Search

The online search was performed on the website http://www.rcsb.org/pdb/home/home.do by amyloid and 38 amyloid precursor names keyword criterion, and the list of PDB codes was generated online. This search gives the list of all the potential amyloid structures.

2.2 Excluding Helical Structures

Next, we excluded the helical structures, leaving the β-sheets and coil structures in the first step of the filtration. The filtration was made using TCL scripting language [9] command *get structure* incorporated in the VMD software [10].

2.3 Excluding Non-parallel Structures

In the final filtration step, we excluded structures which contain more than one non-parallel fragment for each fragment. Firstly, flat fragments were defined according to the Ramachandran backbone torsion angles of φ= (-180˚, -103˚) and ψ= (104˚, 180˚). Furthermore, a fragment must be of minimum 4 amino acids length.

The criterion for the parallelism of fragments was the maximal difference in the distances between two Cα atoms belonging to two parallel fragments of 1.5 Å.

For the purpose of this final structural filtration, the .pdb files were downloaded from http://files.rcsb.org/pub/pdb/data/structures/divided/pdb, and the .pdb1 files containing information in biological assembly were downloaded from http://files.rcsb.org/pub/pdb/data/biounit/coordinates/divided. This is important because both translating a crystallographic unit cell in all the three directions, and completing the biological assembly structure must be done in order to complete the amyloid structure and find all the parallel fragments.

Homemade scripts for the downloading and structural filtration were programmed in Python programming language (http://www.python.org/) and for PDB file parsing MDAnalysis python library has been used [11].

3. RESULTS AND DISCUSSION

Amyloid protein 3D structures were searched in Protein Data Bank (PDB) and in Cambridge Structural Database (CSD). The searching criteria for the CSD were any 4 residue long acyclic polypeptide with nearly β-sheet structure. Eight structures were found, database was made by searching the PDB for the keyword *amyloid* and precursor names. Only the β secondary structures or extended ones were taken. There are 109 structures found in PDB, resolved by X-ray crystallography, solid state or solution NMR.

The online search on the website http://www.rcsb.org/pdb/home/home.do gave us a list of PDB IDs of potential amyloid structures according to the name keyword. Getting the ID list is represented in the Figure 2.

The search was done by picking every structure in which the desired keyword appears. The keyword was simply *amyloid* and 38 amyloid precursor names. The precursor names were published recently in the editorial of Amyloid, The Journal of Protein Folding Disorders [12]. These are all known naturally occurring amyloids. By searching for files that contain the keyword *amyloid*, we included all the synthetic amyloids as well, described by the keywords *amyloid-like*, *amyloid-related*, *amyloidogenic* etc.

We got 1218 structures in total. It is difficult to separate all the amyloid structures, but not pick the non-amyloid ones. Not every structure mentioning the amyloid keyword, is in fact an amyloid, hence filtration about structural features

of amyloids was further used.

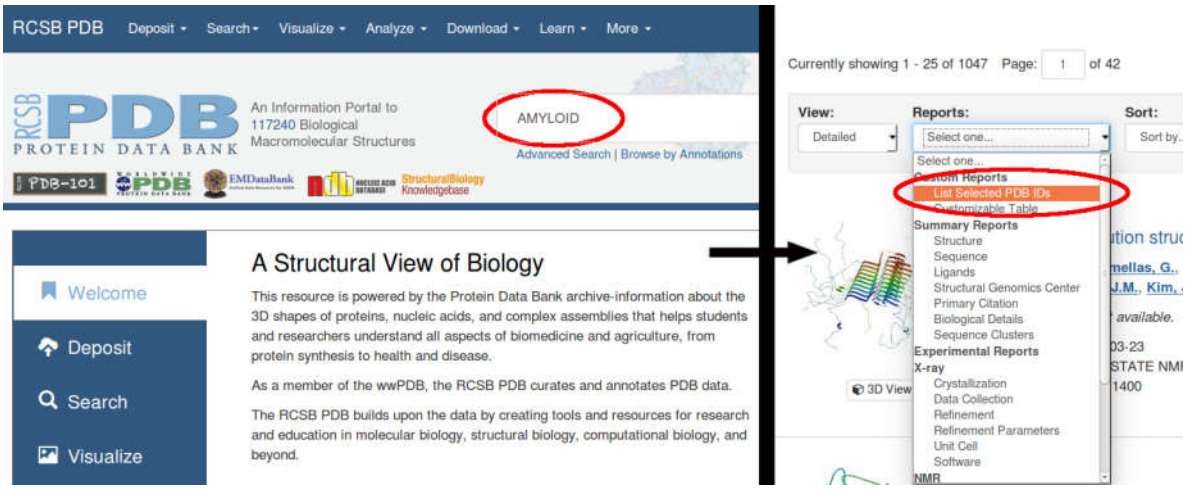

Figure 2: Online search by keyword and making PDB ID list on the PDB site http://www.rcsb.org/pdb/home/home.do

It appears that amyloids are not exclusively β structures, there are also coil and extended peptides which pack in a parallel manner forming long fibrils perpendicular to the peptide axis, Figure 3.

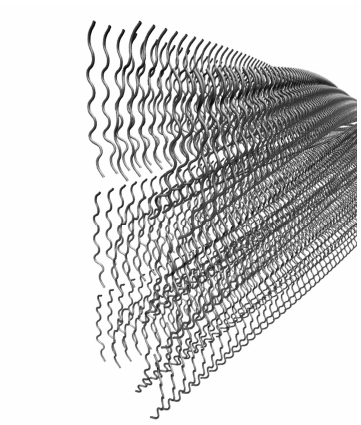

Figure 3: There are amyloid structures with no β-sheets as in this one made of coils, still parallel ones, PDB ID 2M5M.

This is why we excluded only the helical structures, leaving the β-sheets and coil in the first step of the filtration. The result was total of 241 structures. This is still not the ready database, as it contains non-parallel, globular protein arrangements.

We defined amyloid structure as a structure which does not contain more than one non-parallel peptide fragment for every fragment in the whole structure. This is because there are highly ordered structures with alternating parallel and tilted fragments, as in PDB ID: 4UBZ [13], Figure 4, thus amyloids could contain non-parallel fragments. On the other hand, there are parallel fragments in non-amyloidogenic structures, as they may contain β-sheets made of parallel β-strands. But they are mostly globular proteins. We distinguished them from amyloids as structures that contain more than one non-parallel fragment for each fragment.

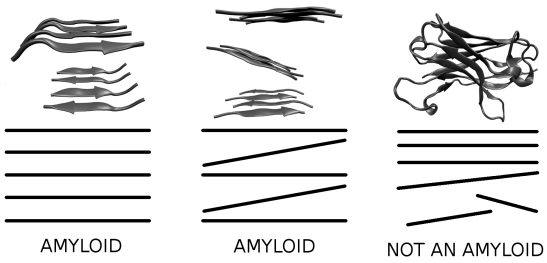

Figure 4: Criterion for distinguishing amyloid structures from non-helical structures: an amyloid possesses maximum one nonparallel fragment for each fragment in the whole structure.

Flat fragments were defined according to the Ramachandran backbone torsion angles found in structures of eight amyloid-β fragments published by [14]. Among these structures, there are β-sheets as well as curved coil fragments with the total torsion angles scope of (-156˚, -103˚) for the φ angle, and (104˚, 154˚) for the ψ angle. This range was expanded by the fully extended peptide conformation, (φ, ψ) = (-180˚, 180˚), so the final range was φ=(-180˚, -103˚), ψ=(104˚, 180˚). Furthermore, a fragment must be of minimum 4 amino acids length.

The criterion for the parallelism of fragments was also taken from the eight structures in [14]. In these structures the maximal difference in the distances between two Cα atoms belonging to two parallel fragments is 1.5 Å, Figure 5.

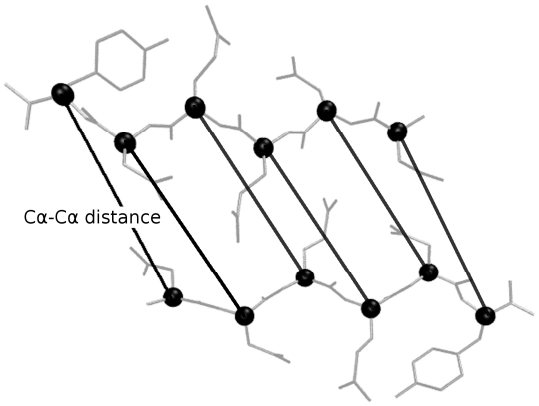

Figure 5: Criterion for parallel fragments: the distance between two Cα atoms belonging to two parallel fragments must differ maximally 1.5 Å, as found in amyloid-β structures resolved by [14].

The resulting database consists of 109 structures represented in Table 2.

Table 2: Amyloid structures found on the basis of keyword criterion as well as structural features, 109 structures.

PDB IDs of amyloid structures				
1YJO	2OKZ	3FTH	3Q9H	4RIL
1YJP	2OL9	3FTK	3Q9I	4RP6
2BEG	2OLX	3FTL	3Q9J	4RP7
2BFI	2OMM	3FTR	3SGS	4TUT
2E8D	2OMP	3FVA	3ZPK	4UBY
2KIB	2OMQ	3HYD	4E0K	4UBZ
2KJ3	2ON9	3LOZ	4E0L	4W5L
2LBU	2ONA	3MD4	4E0M	4W5M
2LMN	2ONV	3MD5	4E0N	4W5P
2LMO	2ONW	3NHC	4E0O	4W5Y
2LMP	2ONX	3NHD	4NIN	4W67
2LMQ	2RNM	3NVE	4NIO	4W71
2LNQ	2Y29	3NVF	4NIP	4WBU
2M5K	2Y2A	3NVG	4NP8	4WBV
2M5M	2Y3J	3NVH	4OLR	4XFN
2M5N	2Y3K	3OVJ	4ONK	4XFO
2MPZ	2Y3L	3OVL	4Q8D	4ZNN
2MVX	3DG1	3OW9	4QXX	5E5V
2MXU	3DGJ	3PPD	4R0P	5E5X
2N0A	3FOD	3PZZ	4R0U	5E5Z
2N1E	3FPO	3Q2X	4R0W	5E61
2NNT	3FR1	3Q9G	4RIK	

The database was confirmed by visual inspection of the 241 non-helical structures found by TCL scripting search.

According to the geometric parameters we considered, flat fragments weather as β-sheets or coils, and number of nonparallel fragments of each fragment, there are five classes of amyloid PDB structures: U-shape with β-sheets connected by unstructured coils (18 structures), β-sheets packed in a flat fashion (23 structures), β-sheets packed in a tilted fashion (7 structures), coil structure packed in a flat fashion (57 structures) and coil structure packed in a tilted fashion (4 structures). Figure 6 represents examples of these five classes.

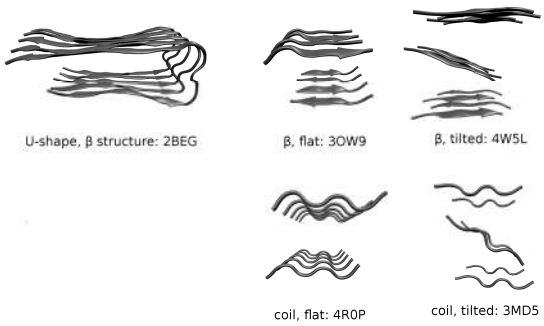

Figure 6: Result of the PDB search: five classes of amyloid structures according to the fragment secondary structure type and their relative angle. The figure shows representative examples for each class.

The found classes of amyloid structure arrangements are all described in the review on amyloid states [15] according to the facial and directional alignment of the interacting β-sheets.

4. CONCLUSION

An amyloid atomic resolution structural data bank was made by searching the Protein Data Bank. The criteria were based on both amyloid name keyword and structural features of amyloid described in literature. The total number of structures is 109 on the 27th of June of 2016. This number will grow as new amyloid structures will be resolved by X-ray crystallography and by NMR spectroscopy. This database can help further structural general and statistical analysis of amyloids, since structural data at molecular level can lead to understanding of disease mechanisms related to amyloid proteins.

REFERENCES

[1] Lakshmanan A., Cheong D. W., Accardo A., Di Fabrizio E., Riekel C., Hauser C. A., "Aliphatic peptides show similar self-assembly to amyloid core sequences, challenging the importance of aromatic interactions in amyloidosis", Proc. Natl. Acad. Sci. U.S.A., 2013, 110, pp. 519-524.

[2] Harrison R. S., Sharpe P. C., Singh Y., Fairlie D. P., "Amyloid peptides and proteins in review", Physiol Biochem Pharmacol ,2007, 159: pp.1-77

[3] Nielsen J. T., Bjerring M., Jeppesen M. D., Pedersen R. O., Pedersen J. M., Hein K. L., Vosegaard T., Skrydstrup T., Otzen D. E., and Nielsen N. C., "Unique Identification of Supramolecular Structures in Amyloid Fibrils by Solid-State NMR Spectroscopy", Angew. Chem. Int. Ed., 2009, 48, pp. 2118-2121

[4] Davis C. H., and Berkowitz M. L., "Interaction Between Amyloid-β (1-42) Peptide and Phospholipid Bilayers: A Molecular Dynamics Study", Biophysical Journal 96, 2009, pp. 785-797

[5] Das P, Kang S., Temple S, Belfort G., "Interaction of Amyloid Inhibitor Proteins with Amyloid Beta Peptides: Insight from Molecular Dynamics Simulations" PLoS ONE 9(11), 2014, e113041.

[6] Berman H. M., Henrick K., Nakamura H. "Announcing the worldwide Protein Data Bank" *Nature Structural Biology* 10 (12), 2003, pp. 980.

[7] www.wwpdb.org

[8] Calamai M., Chiti F., and Dobson C. M., "Amyloid Fibril Formation Can Proceed from Different Conformations of

a Partially Unfolded Protein," Biophysical Journal 89, 2005, pp. 4201-4210
[9] http://www.tcl.tk
[10] Humphrey W., Dalke A., Schulten K., "VMD-Visual Molecular Dynamics" J Molec Graphics 14, 1996, pp. 33-38.
[11] Michaud-Agrawal N., Denning E. J., Woolf T. B., and Beckstein O., "MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations", J. Comput. Chem. 32, 2011, pp. 2319-2327.
[12] Nomenclature 2014: Amyloid fibril proteins and clinical classification of the amyloidosis Amyloid, 2014; 21(4): pp. 221-224, Editorial
[13] Yu L., Lee S., and Yee V. C., "Crystal Structures of Polymorphic Prion Protein β1 Peptides Reveal Variable Steric Zipper Conformations", Biochemistry 2015, 54, pp. 3640-3648
[14] Colletier J. P., Laganowsky A., Landau M., Zhao M., Soriaga A. B., Goldschmidt L., Flot D., Cascio D., Sawaya M. R. and Eisenberg D., "Molecular basis for amyloid-β polymorphism", PNAS, 2011, 108, pp. 16938-16943.
[15] Eisenberg D., and Jucker M., "The Amyloid State of Proteins in Human Diseases", Cell 148, 2012