Book of Abstracts

Belgrade BioInformatics Conference 2016

20-24 June 2016, Belgrade, Serbia

UNIVERSITY OF BELGRADE
FACULTY OF MATHEMATICS

Nenad Mitić, editor

Belgrade BioInformatics Conference 2016

Book of abstracts

Belgrade, June 20th-24th

The conference is organized by the Bioinformatics Research Group, University of Belgrade - Faculty of Mathematics (http://bioinfo.matf.bg.ac.rs).

Coorganizers of the conference are: Faculty of Agriculture, Faculty of Biology, Faculty of Chemistry, Faculty of Physical Chemistry, Institute for Biological Research "Siniša Stanković", Institute for General and Physical Chemistry, Institute for Medical Research, Institute of Molecular Genetics and Genetic Engineering, Vinča Institute of Nuclear Sciences, Mathematical Institute of SASA, Belgrade, and COST - European Cooperation in Science and Technology

The conference is financially supported by

- Ministry of Education, Science and Technological Development of Republic of Serbia
- Central European Initiative (CEI)
- Telekom Srbija
- SevenBridges Genomic
- RNIDS - Register of National Internet Domain Names of Serbia
- Genomix4Life

Publication of this Book of abstracts is financed by the Ministry of Education, Science and Technological Development of Republic of Serbia

Publisher: Faculty of Mathematics, University of Belgrade
Printed in Serbia, by DonatGraf, Belgrade

Serbian National Library Cataloguing in Publication Data
Faculty of Mathematics, Belgrade
Book of Abstracts: Belgrade BioInformatics Conference 2016, 20-24 June 2016.–
Book of abstracts
Nenad Mitić, editor. XIX+151 pages, 24cm.

ISBN: 978-86-7589-108-6

Number of copies printed: 200

International Advisory Committee

Vladik Avetisov	The Semenov Institute of Chemical Physica, RAS Moscow, Russia
Vladimir Brusić	School of Medicine and Bioinformatics Center, Nazarbayev University, Kazakhstan and Department of Computer Science, Metropolitan College, Boston University, USA
Michele Caselle	Department of Physics, Torino University, Torino, Italy
Radu Constantinescu	Department of Physics, University of Craiova, Craiova, Romania
Oxana Galzitskaya	Group of bioinformatics, Institute of Protein Research of the RAS, Russia
Madhavi Ganapathiraju	Department of Biomedical Informatics, University of Pittsburgh, USA
Mikhail Gelfand	A.A. Kharkevich Institute for Information Transmission Problems, RAS, Faculty of Bioengineering and Bioinformatics, M.V. Lomonosov Moscow State University, Moscow, Russia
Ernst Walter Knapp	Fachbereich Biologie, Chemie, Pharmazie/Institute of Chemistry and Biochemistry, Freie Universitt Berlin, Germany
Sergey Kozyrev	Steklov Mathematical Institute, Moscow, Russia
Zoran Obradović	Center for Data Analytics and Biomedical Informatics, Temple University, USA
Yuriy L. Orlov	Institute of Cytology and Genetics SB RAS, Novosibirsk State University, Russia
George Patrinos	Department of Pharmacy, University of Patras, Greece
Nataša Pržulj	Department of Computing , Imperial College London, UK
Paul Sorba	Laboratory of Theoretical Physics and CNRS, Annecy, France
Bosiljka Tadić	Department of Theoretical Physics, Jozef Stefan Institute, Ljubljana, Slovenia
Peter Tompa	VIB Structural Biology Research Center, Flanders Institute for Biotechnology (VIB), Belgium
Silvio Tosatto	Department of Biomedical Sciences, University of Padova, Italy
Edward Trifonov	Weizmann Institute of Science, University of Haifa, Haifa, Israel
Matthias Ullmann	Structural Biology/Bioinformatics Universitt Bayreuth, Germany
Bane Vasić	The University of Arizona, Department of Electrical and Computer Engineering, Bios Institute for Collaborative Bioresearch, USA
Sergey Volkov	Bogolyubov Institute for Theoretical Physics, Kiev, Ukraine
Ioannis Xenarios	SIB Swiss Institute of Bioinformatics, Switzerland

International Programme Committee

Miloš Beljanski	Institute for General and Physical Chemistry, University of Belgrade, Serbia
Erik Bongcam-Rudloff	Division of Molecular Genetics, Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Sweden
Antonio Cappuccio	Immunity and Cancer, Institut Curie, France
Oliviero Carugo	Faculty of Science, University of Pavia, Italy
Boris Delibašić	Faculty of Organizational Sciences, University of Belgrade, Serbia
Zsuzsanna Dosztanyi	Department of Biochemistry Eötvös Loránd University, Budapest, Hungary
Branko Dragovich	Institute of Physics, Mathematical Institute SANU, Belgrade, Serbia
Marko Djordjević	Faculty of Biology, University of Belgrade, Serbia
Olgica Djurković-Djaković	Institute for Medical Research, University of Belgrade, Serbia
Lajos Kalmar	Department of Veterinary Medicine, Cambridge Veterinary School, Cambridge, UK
Eija Korpelainen	CSC IT Center for Science, Finland
Ilija Lalović	Faculty of Natural Sciences and Mathematics, Banja Luka, Bosnia and Herzegovina
Nenad Mitić	Faculty of Mathematics, University of Belgrade, Serbia
Mihajlo Mudrinić	Vinča Institute of Nuclear Sciences, University of Belgrade, Serbia
Zoran Ognjanović	Mathematical Institute SANU, Serbia
Gordana Pavlović-Lažetić	Faculty of Mathematics, University of Belgrade, Serbia
Marco Punta	Pierre and Marie Curie University, France
Predrag Radivojac	Department of Computer Science and Informatics, Indiana University, USA
Ana Simonović	Institute for Biological Research Siniša Stanković, Belgrade, Serbia
Jerzy Tiuryn	Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland
Andrew Torda	Center for Bioinformatics, University of Hamburg, Germany
Alessandro Treves	SISSA-Cognitive Neuroscience, Trieste, Italy
Nevena Veljković	Institute for Nuclear Sciences VINCA, University of Belgrade, Serbia
Igor V. Volovich	Department of Mathematical Physics, Steklov Mathematical Institute, RAS, Moscow, Russia
Snežana Zarić	Faculty of Chemistry, University of Belgrade, Serbia

Local Organizing Committee

Bojana Banović	Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia
Miloš Beljanski	Institute for General and Physical Chemistry, University of Belgrade, Serbia
Branko Dragovich	Co-Chair, Institute of Physics, Mathematical Institute SANU, Belgrade, Serbia
Marko Djordjević	Faculty of Biology, University of Belgrade, Serbia
Olgica Djurković-Djaković	Institute for Medical Research, University of Belgrade, Serbia
Jelana Guzina	Faculty of Biology, University of Belgrade, Serbia
Jovana Kovačević	Faculty of Mathematics, University of Belgrade, Serbia
Saša Malkov	Faculty of Mathematics, University of Belgrade, Serbia
Mirjana Maljković	Faculty of Mathematics, University of Belgrade, Serbia
Vesna Medaković	Faculty of Chemistry, University of Belgrade, Serbia
Nenad Mitić	Co-Chair, Faculty of Mathematics, University of Belgrade, Serbia
Ivana Morić	Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia
Mihajlo Mudrinić	Vinča Institute of Nuclear Sciences, University of Belgrade, Serbia
Vesna Pajić	Faculty of Agriculture, University of Belgrade, Serbia
Mirjana Pavlović	Institute for General and Physical Chemistry, University of Belgrade, Serbia
Gordana Pavlović-Lažetić	Co-Chair, Faculty of Mathematics, University of Belgrade, Serbia
Jelena Samardzić	Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, Serbia
Ana Simonović	Institute for Biological Research Siniša Stanković, Belgrade, Serbia
Miomir Stanković	Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, Serbia
Biljana Stojanović	Faculty of Mathematics, University of Belgrade, Serbia
Aleksandra Uzelac	Institute for Medical Research, University of Belgrade, Serbia

Construction of Amyloid PDB Files Database

Ivana Stanković and Snežana Zarić

[1] ICTM, University of Belgrade, Njegoševa 12, Belgrade, Serbia
ivana_stankovic@chem.bg.ac.rs
[2] Department of Chemistry, University of Belgrade, Studentski trg 12-16, Belgrade, Serbia
szaric@hem.bg.ac.rs

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. Here, we present a constructed amyloid database made based on keyword criterion as well as structural features of amyloids described in literature. The searching filter was performed by python programming. The total number of structures is 109. This database can help further structural general and statistical analysis of amyloids, as we know the molecular basis can lead to understanding of disease mechanisms related to amyloid proteins.

Keywords: database, protein structure, amyloid

1. Introduction

Amyloids are insoluble proteins of a cross-β structure found as deposits in many diseases like Alzheimer's, Parkinson's, CreutzfeldtJakob's, type II diabetes etc. They are also found in normal tissues (nails, spider net, silk) because of their strong fibrillar 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.

Short polypeptides, of minimum 4 amino acids [1], are self-assemblied into β-sheets via backbone hydrogen atoms, then 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. All amyloid proteins, independently of their sequence, form very similar structure, the cross-β structure, made of parallel arrays of β-strands. These structures are different only in the inter-sheet spacing which depends on the side chain size, and in a morphology of a fibril [2].

Amyloids are largely examined structurally [3–5] individually, 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]. 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.

Another difficulty in constructing amyloid database is that amyloid proteins exist in different conformations depending on conditions. They might exist in non-amyloidal conformation in solution when they form helical or random coil secondary structure with no parallel fragments forming fibrils [7].

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

2. Methodology

Amyloid protein 3D structures were searched in Protein Data Bank (PDB) and in Cambridge Structural Database (CSD). The searching criteria for the CSD was any 4 residue long acyclic polypeptide with nearly β-sheet structure. 8 structures were found, but with no proof of self-assembly in the published papers.

Amyloid PDB subdatabase 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.

2.1. Online Search

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.

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 precursors names were published recently in the editorial of Amyloid, The Journal of Protein Folding Disorders, Tables 1, 2 and 3 in [8]. These are all known naturally occurring amyloids. By searching by files that contain the keyword *amyloid*, we include 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. Further filtration in the next sections will deal with structural features of amyloids.

2.2. Excluding Helical Structures

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

2.3. Excluding Non-parallel Structures

We defined amyloid structure as a structure which does not contain more than 1 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 [11], thus amyloids could contain non-parallel fragments. On the other hand, there are parallel fragments in non-amyloidogenic structures, as they may contain -sheets made of parallel β-strands. But they are mostly globular proteins. We distinguished them from amyloids as structures which contain more than 1 non-parallel fragment for each fragment, Fig. 1.

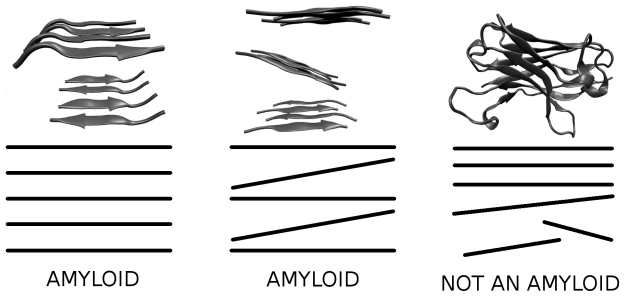

Fig. 1. Criterion for distinguishing amyloid structures from non-helical structures: an amyloid possesses maximum 1 nonparallel fragment for each fragment in the whole structure.

Flat fragments were defined according to the Ramachandran backbone torsion angles found in structures of 8 amyloid-β fragments published by [12]. Among these structures, there are β-sheets as well as curved coil fragments with the total torsion angles scope of (-156°, -103°) for the angle, and (104°, 154°) for the ψ angle. We expanded this scope by the fully extended peptide conformation, (φ, ψ) = (-180°, 180°), so the final scope was φ=(-180°, -103°), ψ=(104°, 180°). Furthermore, a fragment must be of minimum 4 amino acids length.

The criterion for the parallelity of fragments was also taken from the 8 structures in [12]. In these structures the maximal difference in the distance between two $C\alpha$ atoms belonging to two parallel fragments is 1.5$\mathring{A}$, Fig. 2.

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

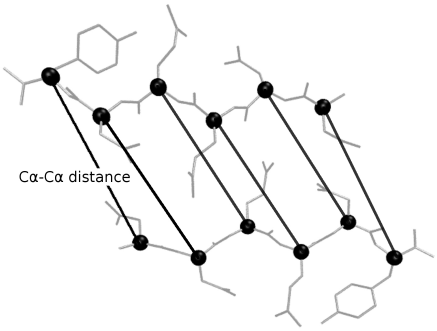

Fig. 2. Criterion for parallel fragments: the distance between two $C\alpha$ atoms belonging to two parallel fragments must differ maximally 1.5$\mathring{A}$, as found in amyloid-β structures resolved by [12].

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 [13] and for PDB file parsing MDAnalysis python library has been used [14].

3. Results and Discussion

The resulting database consists of 109 structures. 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 5 classes of amyloid PDB structures: U-shape with β-sheets connected by unstructured coils, β-sheets packed in a flat fashion, β-sheets packed in a tilted fashion, coil structure packed in a flat fashion and coil structure packed in a tilted fashion. These arrangements of amyloid structures are all found 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 25^{th} of March of 2016. This number will grow as new amyloid structures are resolved crystallographically and by NMR spectroscopy.

This database can help further structural general and statistical analysis of amyloids, as we know the molecular basis can lead to understanding of disease mechanisms related to amyloid proteins.

References

1. Lakshmanan, A. and Cheong, D. W. and Accardo, A. and Di Fabrizio, E. and Riekel, C. and 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., 110, 519–524. (2013)
2. Harrison, R. S. and Sharpe, P. C. and Singh, Y. and Fairlie, D. P.: Amyloid peptides and proteins in review. Physiol Biochem Pharmacol, 159:1–77. (2007)
3. Jakob T. Nielsen and Morten Bjerring and Martin D. Jeppesen and Ronnie O. Pedersen and Jan M. Pedersen and Kim L. Hein and Thomas Vosegaard and Troels Skrydstrup and Daniel E. Otzen and Niels C. Nielsen: Unique Identification of Supramolecular Structures in Amyloid Fibrils by Solid-State NMR Spectroscopy. Angew. Chem. Int. Ed., 48, 2118–2121. (2009)
4. Charles H. Davis and Max L. Berkowitz: Interaction Between Amyloid-β (1–42) Peptide and Phospholipid Bilayers: A Molecular Dynamics Study. Biophysical Journal 96, 785–797. (2009)
5. Das, P. and Kang, S-g. and Temple, S. and Belfort, G.: Interaction of Amyloid Inhibitor Proteins with Amyloid Beta Peptides: Insight from Molecular Dynamics Simulations. PLoS ONE 9(11): e113041. (2014)
6. Berman, H. M. and Henrick, K. and Nakamura, H.: Announcing the worldwide Protein Data Bank Nature Structural Biology 10 (12): 980. (2003)
7. Martino Calamai and Fabrizio Chiti and Christopher M. Dobson: Amyloid Fibril Formation Can Proceed from Different Conformations of a Partially Unfolded Protein. Biophysical Journal 89, 4201–4210. (2005)
8. Nomenclature 2014: Amyloid fibril proteins and clinical classification of the amyloidosis Amyloid, 21(4): 221–224, Editorial. (2014)
9. http://www.tcl.tk/
10. Humphrey, W. and Dalke, A. and Schulten, K.: VMD-Visual Molecular Dynamics. J Molec Graphics 14, 33–38. (1996)
11. Lu Yu and Seung-Joo Lee and Vivien C. Yee: Crystal Structures of Polymorphic Prion Protein β1 Peptides Reveal Variable Steric Zipper Conformations. Biochemistry, 54, 3640–3648. (2015)
12. Jacques-Philippe Colletier and Arthur Laganowsky and Meytal Landau and Minglei Zhao and Angela B. Soriaga and Lukasz Goldschmidt and David Flot and Duilio Cascio and Michael R. Sawaya and David Eisenberg: Molecular basis for amyloid-β polymorphism, PNAS, 108, 16938-16943. (2011)
13. http://www.python.org/
14. Michaud-Agrawal, N. and Denning, E. J. and Woolf, T. B. and Beckstein, O.: MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. J. Comput. Chem. 32, 2319–2327. (2011)
15. David Eisenberg and Mathias Jucker: The Amyloid State of Proteins in Human Diseases, Cell 148. (2012)