TASK QUARTERLY 18 No 3, 289–300

COMPUTATIONAL STUDIES OF TTR RELATED AMYLOIDOSIS: EXPLORATION OF CONFORMATIONAL SPACE THROUGH PETRI NET-BASED ALGORITHM

RAFAL JAKUBOWSKI[1], ANNA GOGOLINSKA[1,2], LUKASZ PEPLOWSKI[1], PIOTR SKRZYNIARZ[1] AND WIESLAW NOWAK[1]

[1] *Institute of Physics, Faculty of Physics, Astronomy and Informatics Nicolaus Copernicus University Grudziadzka 5, 87-100 Torun, Poland*

[2] *Faculty of Mathematics and Computer Science, Nicolaus Copernicus University Chopina 12/18, 87-100 Torun, Poland*

(Paper presented at the CBSB14 Conference, May 25–27, 2014, Gdansk, Poland)

Abstract: Amyloidosis, a serious and widespread disease with a genetic background, manifests itself through the formation of dangerous fibrils in various organs. Apart from the polluted environment and an unhealthy lifestyle, genetic factors may accelerate this process leading in some cases to lethal damages to the body. Recently, a growing interest in amyloidogenic protein research has been observed. Transthyretin (TTR) is a tetrameric protein that transports thyroid hormone thyroxine and retinol binding protein in plasma and the cerebral fluid. Sometimes TTR breaks apart and forms fibrils. Several single point mutations, having destabilizing impact on the TTR complex, are involved in the amyloidogenic TTR cascade. Problems with the TTR tetramer stability and conformational space characteristics of the protein have not been addressed computationally before. We present selected results of our molecular dynamics (MD, $\sim 2\,000$ ns) and steered MD simulations (SMD) of three variants of TTR: Wild Type (WT), V30M and L55P. SMD has been used to enforce the dissociation of TTR. Conformational spaces of WT TTR and its amyloidogenic variants have been investigated using a novel "*One Place One Conformation*" (**OPOC**) algorithm based on a graph technique called Petri net (PN) formalism. While the PN approach alone does not permit a direct identification of protein regions with reduced stability, it gives quite a useful tool for an effective comparison of complex protein energy landscapes explored during classical and/or SMD steered molecular dynamics simulations.

Keywords: transthyretin, amyloidosis, molecular dynamics, Petri nets, clustering, conformational space, graphs

1. Introduction

Numerous diseases have a clear molecular basis. Amyloidosis belongs to such medical conditions and, due its social impact, it has been widely studied in recent years. The common feature of amyloidogenic pathologies is the formation of spurious unnatural biological fibrils in the affected organs, for example, in peripheral nerves or in the heart. Amyloid deposits are made of proteins mainly. Interestingly enough, precursors of those toxic materials are often WT natural proteins having their sequences affected in a few places only.

Familial polyneuropathy (FAP) is amyloidosis caused by numerous deposits of fibrils formed from transthyretin (TTR) [1]. TTR is a tetrameric, 55 kDa, protein which transports thyroid hormones and retinol binding protein in blood and the cerebral fluid. Kelly *et al.* [2] have suggested that the decay of a TTR tetramer into monomers is a key step in such a TTR based disease. There are over 100 dangerous TTR mutations reported in the literature; the particular TTR mutation and the site of amyloid deposition determine the clinical manifestation of the disease. Mutations usually increase the probability of unwanted TTR tetramer destruction [3]. Even such simple mutations, like the L55P substitution, may have a long-term dangerous impact on human health.

Estimates of TTR protein stability will contribute to better understanding of FAP. Before the energetics of TTR is studied in detail and new stabilizing drugs are designed, it is worth characterizing a conformational space of this complex protein. To this end, we have performed computer modeling and checked whether any substantial differences between the WT TTR and amyloidogenic mutants in terms of accessible structures can be observed. Studies of destabilization caused by single point mutations and based on free energy changes are rather challenging and computationally expensive, despite the great progress in the computational methodology. It is easier to obtain exploratory estimates of the conformational freedom of protein, especially using classical molecular dynamics computer simulations. Therefore, we have taken this approach in this work.

TTR has been studied computationally in several papers, however, in contrast to our work, the sole subject of investigations were TTR monomers [4]. On rare occasions when the tetramer was modeled, like in the work of Rodrigues *et al.* [5], artificial environment promoting pathogenesis (500K temperature) was applied as a destabilizing factor, thus, the usefulness of their results is limited. In order to asses to what extent the TTR structure was affected by the dynamics in physiological conditions we performed large scale (>900 ns of trajectories) MD simulations of a human TTR model and models of two widely studied variants, V30M and L55P. Moreover, we calculated the forces required to split the TTR tetramer into two dimers using the steered MD method (SMD) [6]. To the best of our knowledge, this has been the first such study of TTR.

The analysis of MD trajectories is a crucial but sometimes difficult task and typically it is performed using various approaches [7]. For example, in the paper by J. Shao *et al.* [7], eleven different clustering algorithms are compared, based

on top-down splitting and bottom-up aggregating. In particular algorithms such as: single-linkage edge joining, centroid-linkage, average-linkage, complete-linkage, centripetal, and centripetal complete self-organizing maps and tree splittings are discussed.

In numerous applications the Principal Component Analysis (PCA) technique is used. In PCA the diagonalization of the atomic movement covariance matrix and ordering of the eigenvalues allows the system dimension to be significantly reduced and helps detect the most important fluctuations [8]. More recently "dihedral" variants of PCA have been becoming popular: the use of internal dihedral angles, instead of Cartesian coordinates, facilitates construction of free energy landscapes of biomolecules [9]. Alternatively, more advanced methods, such as the Functional Mode Analysis, may be employed to obtain interesting correlations between the system's behavior and quasi-functional quantity [10]. Those methods have established popularity, however, they are not easy to use [7]. Every method of processing multi-gigabyte protein trajectories into simpler, yet meaningful graphs has a potential of giving new insights into the MD analysis process and expands the arsenal of computational biophysics. Graphs developed in a systematic way may help, for example, select or point out those features in protein conformational spaces which are distinct during equivalent simulations for WT molecules and their mutated variants. Such an approach is certainly "more automatic" than tedious visual inspection of individual frames, and hopefully may be more informative than the traditional clustering methods [7].

Here we adapted a Petri net approach [11] to the analysis of MD and SMD trajectories. A special algorithm was developed for generation of a representation Petri net from a given MD output data set – we called it the "*One Place One Conformation*" (**OPOC**) algorithm. The **OPOC** algorithm has been tested on MCP-1 chemokine, perhaps related to Autism Spectrum Disorders [12]. In this paper we present new data on TTR and its variants related to FAP dynamics and further exploit the **OPOC** algorithm using TTR MD trajectories. Our goal is to unearth possible differences in conformational spaces of TTR and its variants using simple PN graphs and to provide means for better selection of TTR structures suitable for a future effective anti-FAP drug design.

2. Methods

2.1. Molecular Dynamics Simulations

All simulations were carried out using the NAMD package [13] with the CHARMM27 force field [14]. The human TTR structure 1ICT from the PDB database was used [15] (see Figure 1). The protein was immersed in a TIP3P water box, at least 8 Å layer was kept in each-direction. NaCl ions were added at the level of 0.15 mmol/L. The MD simulations were preceded by water equilibration, minimization and heating from 0 to 300 K stages. The 1 fs timestep was used and the Ewald summation method was employed for calculation of electrostatic

interactions [16]. Mutants were generated from WT protein using the VMD code [17].

Each SMD simulation [6] was preceded by at least 10 ns of classic MD and cutting of the water box to a droplet. The SMD pulling directions were determined by vectors connecting the centers of mass of dimers (in the TTR tetramer dissociation study) or monomers (in the dimer dissociation investigation). In the SMD Cα atoms of TTRC-TTRD (the tetramer case, see Figure 1) or TTRD (the dimer study) segments were fixed and the Cα atoms of TTRA-TTRB or TTRA segments were pulled, respectively. SMD forces were attached to all Cα atoms of the pulled fragment. 3 trajectories 200 ns long for WT, V30M and L55P were split into four equal parts each (p1-p4) and the p1+p2 parts were further analyzed using our PN code for the **OPOC** algorithm written in Java.

2.2. Petri Net Representation of Conformational Space

Below we present a short synopsis on Petri nets. Petri nets (PN) belong to the mathematical tools created to describe distributed systems [11]. They were introduced by Carl Adam Petri in 1939 for modeling chemical reactions. A Petri net has the form of a bipartite graph with two kinds of nodes: places and transitions.

Definition 1.
A Petri net graph is a 4-tuple (P,T,F,W), where:

- P is a finite set of places.
- T is a finite set of transitions (or actions), such that $P \cap T = \emptyset$
- F is a set of directed arcs, satisfying: $F \cap (P \times P) = F \cap (T \times T) = \emptyset$ (the place may be connected with the transition or the transition with the place; two places or two transitions cannot be connected).
- $W : F \rightarrow \{1,2,3,\ldots\}$ is a weight function assigned to arcs. The weight of one is assigned to an arc as a default.

On the network plot places are represented by circles, transitions by squares and arcs by arrows. Weights are represented by numbers labeling arcs. No label is displayed near an arrow if its weight is the default (the default: 1).

We do not have any actions in a PN created according to Def. 1 – it is a steady framework only. We need tokens to have actions. Distribution of tokens over places of a net is called marking.

Definition 2.
Net N marking is mapping $M : P \rightarrow \{0,1,2,3,\ldots\}$.

Definition 3.
A Petri net is a quintuple $(P,T,F,W,M0)$, where $M0$ is initial marking, and P, T, F, W are as in Def. 1.

Definition 4.
For each element $t \in T$ we can define the set of input places $\bullet t = \{p \in P; (p,t) \in P\}$

which is the set of all places from which an arc runs to transition t and the set of output places $t\bullet = \{p \in P; (t,p) \in P\}$ – the set of all places to which arcs run from transition t.

Definition 5.
Transition t may fire (such transition is called firabled or enabled) in marking M, if the number of tokens in every input place p of transition t is equal or greater than weight $W(p)$ assigned to the arc between p and t in marking M.

Transition t consumes tokens from its input places p and puts them into output places q – the number of tokens transferred is described by the weights of arcs. Firing of a transition changes the marking.

*2.2.1. One Place One Conformation (**OPOC**) algorithm*

In the **OPOC** algorithm one PN place will represent one conformation of a molecule – a point in the conformational hyperspace. The conformation of a molecule is represented here by positions of all Cα atoms. Conformations are sampled from a classical, or steered, MD trajectory. The sampling frequency depends on the user.

Thus, one PN place in the **OPOC** algorithm corresponds to the positions of every Cα atom. Transitions represent possible changes between conformations. The token marks the current conformation. As the molecule can be just in one conformation at a given time point, one token only is present in our PN. The network is generated using a special code developed in house and it depends on the conformational space visited by the molecule during the simulation (or simulations) used in the PN generation step. The "spatial" resolution of the network, and therefore its complexity, depends on the RMSD distance cutoff used and the parameter (R_d) that determines the "granularity" of the analyzed conformational space. One PN may be generated from a larger number of trajectories of the same protein.

2.2.2. **OPOC** *protocol*

The input data for the **OPOC** algorithm is: an MD trajectory (or multiple trajectories), and the R_d threshold – its value is given by the user. Values from the range of 0.75–5 Å were tested here. During the PN generation the **OPOC** algorithm reads data from every input file, frame by frame. Every conformation of the protein is represented by positions of all its Cα atoms. A place of a newly generated PN is linked to the conformation identified as the first one forming a new distinct point (or a cluster) in the reduced conformational space. When the next frame is read-in, we check if the place appropriate for that structure has already been obtained. The assignment of all conformations (frames) to PN places is performed in chronological order. The RMSD measure and structural alignment are used to compare the current conformation (from the current frame) and previous conformations (from previous frames), which are already represented by places in our growing PN. The current frame is aligned with every PN place generated so far and the smallest resulting RMSD is compared, whether it is

smaller than the threshold R_d or not. If it is smaller, the current conformation is assigned to the respective place in the PN. Otherwise, a new place and a new transition are added to the PN.

The PN generated using the **OPOC** algorithm is usually small. The size of the network depends, of course, on the number and size of input trajectories and the value of the RMSD threshold parameter. However, the main information contained in these simple graphs is the span of the conformational space subject to inspection. Such networks clearly represent "the journey" of a molecule across the conformation space. Sequences of conformations as well as the relationship between different trajectories are presented in the net as every place can be associated with an MD trajectory frame, or a cluster of similar frames. For bigger R_d thresholds the Petri net is strongly reduced and the **OPOC** analysis results are similar to those obtained by standard clustering methods. It is expected that all similar structures/fragments of trajectories will be assigned to one part of the Petri net.

3. Results and Discussion

3.1. MD and SMD simulations

The structures examined are presented in Figure 1 (a). One can see that the V30M and L55P mutations are located on the "outskirts" of the main protein core and not on the interfaces between TTR monomers. Thus, the modulation of TTR tetramer stability induced by these mutations is a long-range indirect effect. We selected initial X-ray, random and distant (in terms of the RMSD) structures from the 100 ns TTR trajectory (see the RMSD plot in Figure 2). These are overlaid in Figure 1 (b). One can see that during our simulations the tetramer did not exhibit any substantial conformational changes. The VMD inspection of V30M and L55P trajectories gave a similar picture. Thus, it was not possible to identify any large scale (or functional) motions/rearrangements that might be linked with modified stability of the tetramer. The **OPOC** algorithm is used in the next step of analysis of the TTR conformational space. The results are described in the next section.

SMD is a useful technique for searching mechanical clamps in forced unfolding of proteins [18]. Our studies of autism-related neural proteins [12, 19, 20] have given valuable hints about unfolding paths and intermediates. The method may be also used to check the mechanical stability of protein complexes. Our preliminary results (to be published elsewhere) show that the maximum force for WT TTR tetramer enforced dissociation into AB+CD and AD+BD dimers is 3719 pN and 2520 pN, respectively. This trend is identical with experimental observations [21]. The results for both mutants are similar, but not identical (for V30M: 3500 pN vs 3643 pN for AB+CD case, preliminary data), Hopefully, better statistics of SMD will show distinct stability in these complexes.

3.2. Petri Net representation of the TTR conformational space

The **OPOC** algorithm was used to create simplified graphical representations of the conformational space sampled during 100 ns MD simulations of WT

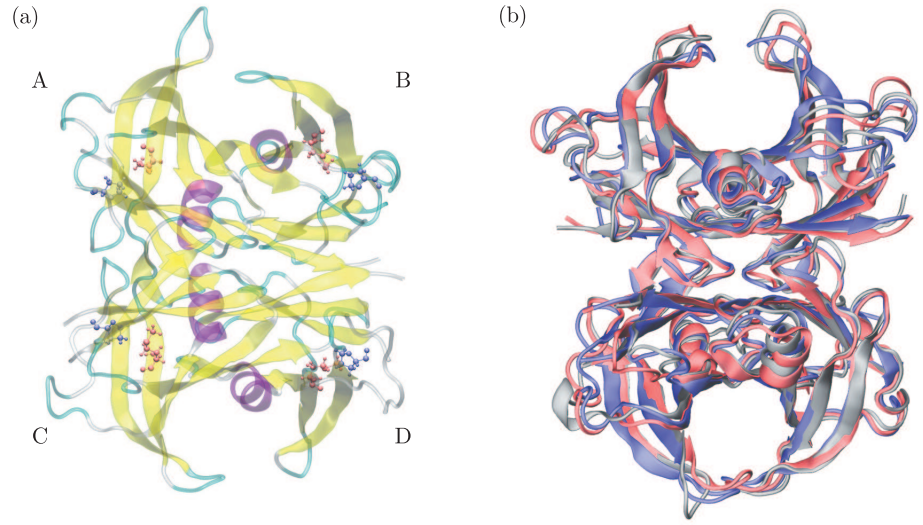

Figure 1. (a) Structure of TTR and localization of V30M and L55P mutations; (b) X-ray structure, a random representative, one the most deformed structures in the 100 ns TTR WT trajectory

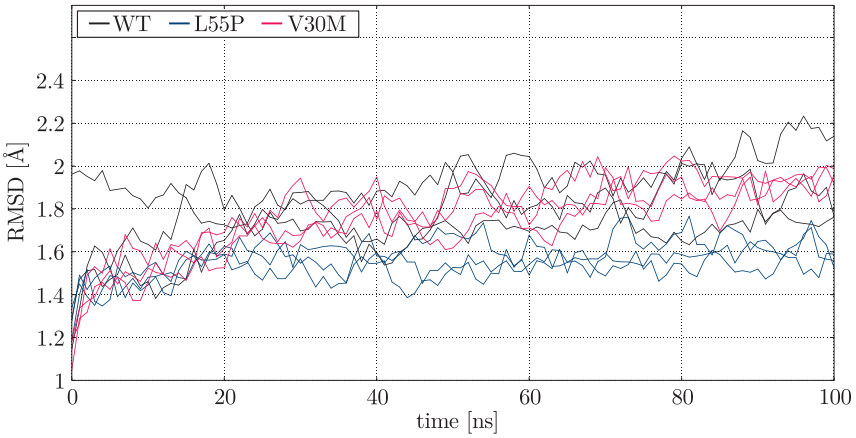

Figure 2. Root mean square distances (RMSD, in Å) between the examined structures and pdb-based minimized initial structures; the distances were calculated for Cα atoms only

TTR and both mutants. The graphs show increasing complexity when the distance radius parameter R_d used for the allocation of frames to a given place is getting increasingly smaller. The real data generated for 3 WT 100 ns MD trajectories (total 3000 frames) is shown in Figure 3. It is up to the user to decide what level of PN complexity is best suited for further analysis. Our preliminary experience is that, for complexes similar to TTR, the R_d value of 2 Å gives sufficiently complex graphs grasping main clusters and distributions of the representative states.

The PN graphs may be compared in terms of numbers of places p and numbers of transitions t resulting from a given protein, the sampling protocol

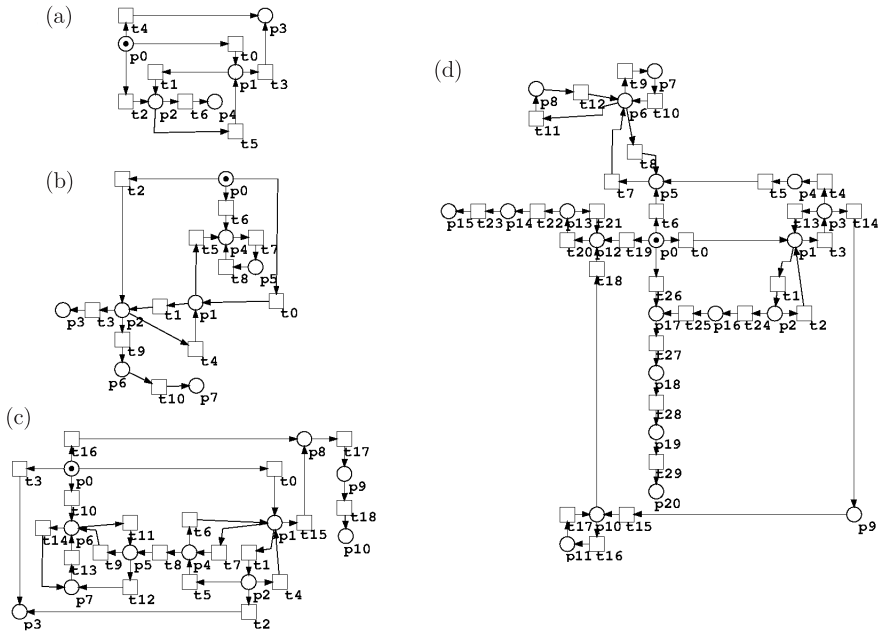

Figure 3. Comparison of PN representation of 300 ns WT TTR MD trajectories generated with **OPOC** algorithm using the decreasing R_d parameter of (a) 5 Å, (b) 4 Å, (c) 3 Å, (d) 2 Å

and the simulation length. Such p/t parameters may be easily extracted from the **OPOC** output. In Table 1 we present data for WT TTR and both V30M and L55P mutants. The higher the number of places for the same protein, the more complex (extensive) its conformational space. The PN shape and the distribution of transitions give another new characteristics of the regions allowed for the examined system. This data itself describe the flexibility of the molecular complex in a new and very concise way.

From the data presented in Table 1 we see that the p/t ratios for the richest set (row 6) for WT and L55P mutants have very similar values: 240/323, 242/314. However, a distinct (in terms of p) net is generated: $p/t = 263/313$ for the V30M mutant. Thus, the number of places is higher by 20 (10%) here. This means that the conformational heterogeneity of the V30M mutant is clearly higher than in case of the WT TTR and L55P systems. This observation alone does not explain the amyloidogenic propensity of mutants, but it gives a useful hint for a further detailed MD study of the particular V30M variant. When large sets of data are analyzed, such PN based fast analysis helps focus further more elaborate studies on a small subset of mutations and save computational time.

In Figure 4 we show representative WT TTR and V30M mutant PNs, together with representative structures from selected p sites. A very careful inspection of overlapped structures (Figure 4 a b) is required to notice minute changes in heterogeneity of the left parts of WT and V30M structures, probably induced by a mutation. On the other hand, the PN graphs shown in Figure 4 (c)

Table 1. Dependence of p/t (number of places p and transitions t) on the threshold (R_d, in Å) used in the PN generation; data for an increasing number of frames is presented (#Input files 2 has 1000 frames, #Input files 6 has 3000 frames)

WT

#Input files	$R_d = 1$	$R_d = 2$	$R_d = 3$	$R_d = 4$	$R_d = 5$
2	79/106	9/13	4/4	4/4	3/3
4	159/210	16/24	8/15	6/9	4/5
6	240/323	21/30	11/19	8/11	5/7

V30M

#Input files	$R_d = 1$	$R_d = 2$	$R_d = 3$	$R_d = 4$	$R_d = 5$
2	91/105	10/15	5/8	3/4	3/3
4	179/212	18/28	8/14	4/6	4/4
6	263/313	27/43	11/18	6/9	5/6

L55P

#Input files	$R_d = 1$	$R_d = 2$	$R_d = 3$	$R_d = 4$	$R_d = 5$
2	86/98	11/19	5/11	3/4	3/4
4	166/207	15/34	7/14	4/5	4/5
6	242/314	22/50	10/19	6/8	5/6

WT SMD

#Input files	$R_d = 1$	$R_d = 2$	$R_d = 3$	$R_d = 4$	$R_d = 5$
1, 10 ns	35/35	14/13	9/8	7/6	6/5

and Figure 4 (d) are easy to interpret and one can see immediately that the plot of the mutant (Figure 4 d) is more complex than that of WT (Figure 4 c) since it contains more places and transitions.

We compared our OPOC data with a standard clustering method based on the Cα distance criterion. The number of clusters found grows substantially with increasingly smaller RMSD cutoff used for clustering, making a comparison of all three trees (data not shown) almost impossible. Therefore, in Figure 5 we show only a part of this dependence. One can see that the classical clustering (1.3–2.0 Å RMSD range) predicts the number of clusters in mutants to be slightly higher than in WT TTR. We inferred that mutations had led to more scattered structures. This result is in perfect agreement with our OPOC-based observations.

One should note that a PN graph alone does not allow direct assessment of the stability of a given TTR variant (or another complex). However, when the same methodology is applied to numerous protein variants, a comparison of graphs helps classify the examined systems in groups of those having large heterogeneity of the conformational space (low energy barriers between very distinct structures) and those with conformations clustered into small basins of stable conformations.

The p/t fraction helps control whether the conformational space was exhaustively sampled in a given MD study. If new trajectories added to the

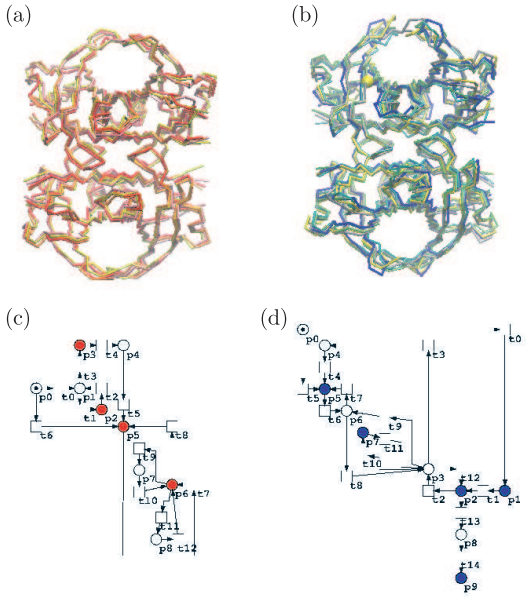

Figure 4. Comparison of selected structures from WT (a) and V30M (b) TTR trajectories and PN graphs ($R_d = 2$ Å) for conformational spaces of these proteins (c) WT, (d) V30M; places representing structures are marked in color

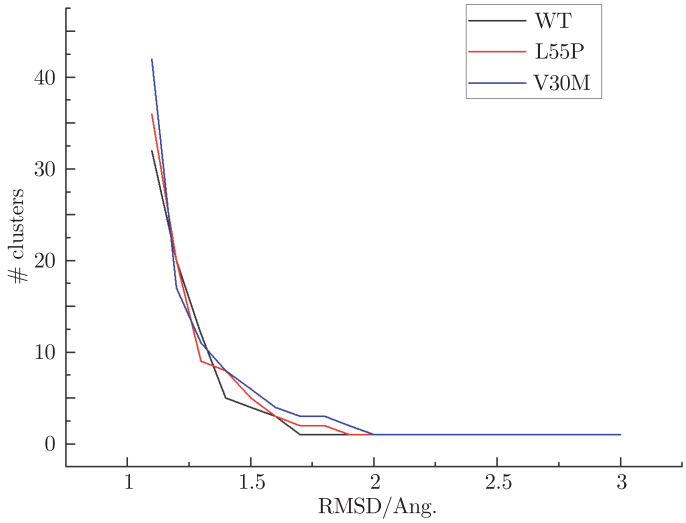

Figure 5. Results from classical clustering of TTR and its variant MD trajectories; the dependence of the number of clusters found (#clusters) on the root mean square distance (RMSD) cutoff is shown

OPOC input change substantially the p/f fraction, it means that the part of the conformational space visited during simulations increases and those additional calculations bring new data (for example, crucial for free energy calculations).

If p/t in the PN remains almost constant, despite an increased number of frames used for the PN generation, it means that the system remains trapped in the same conformational region.

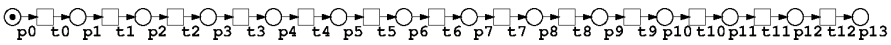

Figure 6. PN generated for 10 ns SMD trajectory describing forced dissociation of WT TTR

Comparisons of different molecule/calculation schemes are even more useful. For example, in Figure 6 we show a PN of a WT TTR tetramer stretched by SMD into two dimers. Clearly the trivial shape of this graph is somehow anticipated since the RMSD rises in a systematic way in the SMD protocol and nothing but just new places are expected to arise in the course of such simulation. This example shows that **OPOC**-generated PNs correctly grasp the major features of molecular conformational space.

4. Conclusions

TTR is a protein complex of high medical relevance. We performed extensive MD and SMD studies of this tetrameric molecule and its two amyloidogenic variants: V30M and L55P. Over 1000 ns trajectories were analyzed. A new method was proposed to facilitate the comparison of conformational spaces of WT protein and its mutants. The Petri Nets – concise bipartite graphs were generated using a newly developed **OPOC** algorithm. The method, initially used for an analysis of autism-related chemokine MCP-1 [12] was used here to pinpoint major differences in the conformational flexibility of three TTR systems. We have found that the complexity of PN critically depends on the R_d distance parameter used for PN generation. In our opinion, R_d of 2 Å gives a reasonable compromise between the graph complexity and its informational content (a useful signal) encoded in the graph. The inspection of PN indicates that the conformational space visited by the V30M variant on 1000 ns MD simulation scale is richer than that of WT and L55P systems. This observation is consistent with more troublesome classical clustering of MD trajectory frames. This shows that further modeling of TTR systems is a promising research area.

We postulate that the PN and the **OPOC** algorithm may be useful tools for fast scrutiny of conformation sets generated during biophysical computer simulations.

Acknowledgements

This work was supported by NCN Grant No. N202 262083 (WN) and Faculty of Physics, Astronomy and Informatics, NCU Grant No. 1623-F (RJ). AG thanks "Krok w przyszłość-stypendia dla doktorantów V edycja" fellowship. The authors thank R. Adamczak, Ph. D for reference clustering and M. Michalski, Ph. D for reviewing the manuscript.

References

[1] Hagen G A, Elliott W J 1973 *The Journal of Clinical Endocrinology & Metabolism* **37** 415
[2] Sekijima Y, Wiseman R L, Matteson J, Hammarstrom P, Miller S R, Sawkar A R, Balch W E, Kelly J W 2005 *Cell* **121** 73
[3] Jakubowski R, Skrzyniarz P, Peplowski L, Nowak W 2014 *Biophysical journal* **106**, 661a
[4] Yang M, Lei M, Huo S 2003 *Protein science* **12** 1222
[5] Rodrigues J R, Simões C J, Silva C G, Brito R M 2010 *Protein science* **19** 202
[6] Grubmüller H, Heymann B, Tavan P 1996 *Science* **271** 997
[7] Shao J, Tanner S W, Thompson N, Cheatham T E 2007 *Journal of Chemical Theory and Computation* **3** 2312
[8] Amadei A, Linssen A, Berendsen H J 1993 *Proteins: Structure, Function, and Bioinformatics* **17** 412
[9] Altis A, Otten M, Nguyen P H, Hegger R, Stock G 2008 *The Journal of Chemical Physics* **128** 245102
[10] Hub J S, De Groot B L 2009 *PLoS computational biology* **5**, e1000480
[11] Peterson J L 1977 *ACM Comput. Surv.* **9** 223
[12] Gogolinska A, Nowak W 2013 *Journal of Molecular Modeling* **19** 4773
[13] Phillips J C, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R D, Kale L, Schulten K 2005 *Journal of computational chemistry* **26** 1781
[14] Brooks B R, Brooks III C, Mackerell Jr A, Nilsson L, Petrella R, Roux B, Won Y, Archontis G, Bartels C, Boresch S 2009 *Journal of computational chemistry* **30** 1545
[15] Wojtczak A, Neumann P, Cody V 2001 *Acta Crystallographica Section D: Biological Crystallography* **57** 957
[16] de Leeuw S W, Perram J W, Smith E R 1980 *Proceedings of the Royal Society of London, A. Mathematical and Physical Sciences* **373** 27
[17] Humphrey W, Dalke A, Schulten K 1996 *Journal of molecular graphics* **14** 33
[18] Lee W, Zeng X, Zhou H-X, Bennett V, Yang W, Marszalek P E 2010 *Journal of Biological Chemistry* **285** 38167
[19] Mikulska K, Pepłowski Ł., Nowak W 2012 *Journal of Molecular Modeling* **17** 2313
[20] Mikulska K, Strzelecki J, Nowak W 2014 *Journal of Molecular Modeling* **20** 1
[21] Trivella D B, dos Reis C V, Lima L M T, Foguel D, Polikarpov I 2012 *Journal of structural biology* **180** 143