



## Research Trip Summary Report

### Task 2. Foreign mobility of WUST doctoral students

#### I. Data of the doctoral student

1. Full name: Daniel Wiczew
2. Year of studies: 3
3. Educational discipline: Biomedical Engineering

#### II. Foreign research trip (research visit)

1. Research institute in which the foreign research was implemented: LPCT (Laboratoire de Physique et Chimie Théoriques) from Université de Lorraine
2. Name and surname of the host person (mentor): prof. Mounir Tarek
3. Dates of the research trip: 02.06.2022 – 02.07.2022W
4. Title and date of a seminar delivered during the research trip: Markov State Models in molecular dynamics protein simulations, 24.06.2022.
5. Description of work carried out during the research trip:

Dimensionality reduction often dictates the performance of the other steps required to build a Markov State Model (MSM) from a molecular dynamics trajectory. Several of them were developed specifically for it, like Time-Independent independent component analysis (TICA), SRV (State-Free Reversible VAMPNet), VAMP (Variational Approach for Markov Process).

According to currently available literature the best performing is SRV. Nevertheless, most of the time research was done either on toy models or low-dimensional data (small proteins with not many degrees of freedom). Here I tried to check whether it will work on high dimensional data (e.g. a large proteine). Further, if it can be improved by improving function approximation architecture.

During the research covered by this report, SRV was assessed with regard to baseline method (TICA) on two protein systems, a small pentapeptide and the RFAH (Reversible fold-switching) protein which requires higher dimensional features to be described. The molecular dynamics trajectories for those two proteins were sampled enough, so that error introduced during the comparison wasn't the source of differences. Then the results were assessed by calculating VAMP2 score and by convergence of the Markov state model for each system.



## InterDocSchool Project

### 6. Description of the main results obtained:

There wasn't any working implementation of SRV before this NAWA mobility, one was created and made publicly available (<https://github.com/DanielWicz/easysrv>).

With regard to scientific results, we discovered that the original proposed method over-fits significantly, even with very large datasets. Original method uses fully connected neural network as a function approximation, which are very prone to over-fitting. To solve the problem, a different function approximator was used, namely convolutional neural networks (CNNs). Since features from proteins are often highly correlated and very high dimensional, theoretically they should work correctly. Furthermore, the convolutional neural networks have far less parameters than feed forward neural networks, which makes them significantly harder to over-fit, and easier to generalize.

In the present results, the SRV with CNN appeared to be either as good as the baseline (TICA) method or they perform better. In contrast, SRV with fully connected NNs perform either worse than the baseline or at best the same way. Nevertheless, this questions usage of the deep models, as they "simply" over fit too much due to high number of parameters. They may work for limited cases, but are not as robust as simple linear models like TICA.

### 7. Future collaborations (if applicable):

Further collaboration is expected with the host Université de Lorraine.

8. Title and date of a seminar presenting the results of the trip to be delivered at Wrocław University of Science and Technology after returning from the research trip:

*Dimensionality reduction for molecular trajectories of proteins*, 10:00, 07.07.2022.

## III. Doctoral student's signature



(Date) 02.07.2022  
(doctoral student's signature)

#### **IV. Confirmation and information from the host**

1. Confirmation of compliance of the information contained in the report: I **CONFIRM** / ~~DO NOT CONFIRM~~. (*In justified cases, the confirmation of the host may be sent by e-mail to the Dean's Office of the Doctoral School email: [interdocschool@pwr.edu.pl](mailto:interdocschool@pwr.edu.pl)*)

2. Additional information and comments

Daniel was present at Université de Lorraine, Nancy, France between 02.06.2022 – 02.07.2022. Furthermore, he actively attended weekly group meetings, presented research during the meetings and communicated with other members of the group and other groups.

(Date) July 2<sup>nd</sup> 2022  
(signature(s) of Host)