Modelling Biologically Relevant Internal Motions of Protein Molecules with Ensemble-based Approaches

Zita Harmat

Thesis Book of the PhD Dissertation

Pázmány Péter Catholic University
Faculty of Information Technology and Bionics

Supervisor:
Zoltán Gáspári, PhD

Budapest, 2022

1 Introduction

Proteins belong to the most important biomolecules. They determine cellular shape and motility, catalyze the reactions, which are most important for life, make transport of materials possible between cells and their environments, and they have important roles in several other molecular processes. Describing their functioning in a precise way is crucial for understanding processes of life and for designing their modulation for example for therapeutic purposes.

Proteins can be divided into four groups based on their structure: globular, fibrillar, transmembrane and intrinsically disordered proteins. Proteins having a well defined, stationary three-dimensional structure belong to the first three groups. The local structure, so called secondary structure of the globular proteins is versatile, they contain more α-helical and / or β-strand regions, and their shape is approximately spherical. Proteins having long, thread-like structures belong to the fibrillar type. The transmembrane proteins are similar in many ways to the globular proteins, but their feature is, that they reach through the biological membranes. The intrinsically disordered proteins - opposite to the first three groups - have no single characteristic structure, but due to their high degree of internal mobility, they are in constant exchange between several conformations.

The diversity of protein shapes is crucial for the determination of the versatile biological role of proteins. It is well known today, that all proteins, not only the intrinsically disordered ones, are dynamic objects and their internal motions along with their structure organically contribute to the fulfilling of their molecular function. Atomic level picture was obtained about proteins with roentgen crystallography, which described a static structure in a crystalline state. But later other experimental methods could be applied to analyze proteins, such as liquid state NMR measurements. These experiments reported about the internal dynamics, and motions, on many different timescales. This internal dynamics can be described by a model consisting of an ensemble of structures.

Because experimental discovering of the structure and internal dynamics of proteins is time- and money exhaustive, *in silico* protein structure and dynamics calculations are on the rise. But it is not neat, when they are not connected to any experimental data, because it makes guaranteeing and checking the accuracy - correspondence to reality - of the model hard. Becuase of this, these days the combination of *in silico* modelling with experimental data is an important trend. The models help to understand and calibrate experimental data, and in turn the experimental data make the mod-

els more realistic. This computational modelling, which uses experimental models to build up atomistic protein structural models is a state of the art method is widely used since years. It can be remarked, that actually this method is not new in the sense, that the traditional structure determination methods are data sparse, this means, that the number of coordinates to be determined is significantly more than the number of experimentally measured parameters, and hence they heavily rely on *a priori* knowledge, primarily geometrical data (bond lengths, bond angles, etc.), which are built into the protocol in a similar way as in the modelling procedures. This can be catched out mostly in structur calculation protocols using NMR specroscopy data. A feature of the nowadays trending "hybrid" procedures is that the weights of the experimental data is less than the weight of the forcefield used in the modelling, making its role more pronounced, and the experimental data can be interpreted mostly in the case of already known structures. An example used in this work is the S^2 order parameter or the experimental restraints used in protein complex model building with docking approaches.

A model has two main criteria: precision and accuracy. Its accuracy is expressing, how much the model corresponds to reality, and precision describes, how exact is the experimental data used, how much experimental error is included. A model which is accurate, but not precise has a big standard deviation in its data, but their average corresponds to reality. In a precise, but not accurate model the standard deviation is small, but the average is far from the real values. From these two cases, the accurate, but not precise model can be used in a limited way, but the precise but not accurate model is unfortunately useless. We can check the accuracy of a model, if we use experimental data from multiple experimental methods, because it is very unlikely, that the same error is made in different experiments. Also it can occur, that when building our model, experimantal data is misinterpreted or it is processed the wrong way. In this case for example this mistake can be spotted if our model corresponds to the data from one of the experiments, but it does not correspond to the data from another experiment, so we have problems with the accuracy of our model [1].

In the dissertation, I present the production and analysis of ensemble models of three different kinds of proteins: the protein gastrotropin represents the group of globular proteins, the myosin VI single α-helical (SAH) domain represents the fibrillar group and the cytoplasmic domain of Cd3ϵ belongs to the intrinsically disordered proteins. In the case of the gastrotropin protein I performed restrained molecular dynamics simulations, whereas in the case of the other two proteins I produced a preliminary conformer library, from which I produced a subensemble corresponding to the experi-

mental data by selection. Production of the conformer libraries was made with the program DIPEND (DIsordered Protein Ensembles from Neighbor-dependent Distri-butions), which I developed. In the case of each protein for the analysis of the correspondence of the ensembles to the experimental data was performed by CONSENSX$^+$ program, which was developed by other members of our research group.

The procedure of protein structure modelling employed in the case of the myosin VI and cd3ϵ proteins has two steps. In the first step, an ensemble of many members is made, which diversely covers a large part of the conformational space using the DIPEND program. These has to evenly distribute along the conformational space, covering as much of it as possible. In the second step we choose a subensemble from these with the CONSENSX$^+$ program, which corresponds to the experimental data. Separating these two steps is advantegous, because with this approach, both steps can be arbitrarily shaped, taking into consideration the properties of the protein to be modeled.

In accordance with what is written above, along the course of this work, experimental data derived from the scientific literature were used for the validation of the structural models.

2 Methods

2.1 Components of the DIPEND program

I wrote the program mostly in Python3. This is a programming language created by Guido van Rossum at the end of the eighties (https://www.python.org/doc/). It is interpreted, high level and very popular. Object oriented, but supports other styles of programming as well. It results in a well readable, transparent and development friendly code. Its special feature is, that program blocks are separated with a different level of indentation instead of brackets.

The Chimera molecule visualization and molecular modelling program is developed at the UCSF university [2], its new version is ChimeraX [3] (https://www.rbvi.ucsf.edu/chimerax/download.html) which is a well maintained, continuously developed program, available free of charge upon registration, and is versatile, excellently suitable for molecular modelling. From the point of view of my program, its big advantage is, that it can be called from the command line with parameters without any user intervention. The DIPEND program uses the protein model building module, which

builds based on the sequence and a given dihedral angle pair, it also uses the dihedral angle setter module and the steric clash check module of ChimeraX.

The SCWRL4 program developed by the Dunbrack lab [4] (http://dunbrack.fccc.edu/SCWRL3.php/) is free to use under a licence. This enables the building of an atomic model of not sterically clashing protein side chains for a fixed backbone. It can be called in the command line, it is very fast and uses a PDB file as input, and likewise a PDB file for output, keeping the chain and amino acid numbering. It uses a backbone dependent rotamer library, for which it calculates energies based on their frequency and closeness to the other sidechains. It avoids steric clashes by using k-discrete polytopes, which are 3-dimensional shapes. It chooses from the many possibilities with its own graph tree decomposition method and makes the choice faster by applying heuristics, which always gives results quickly, but it is not sure, that the conformer having minimal energy is found.

2.2 Molecular dynamics simulations

We performed molecular dynamics calculations using GROMACS version 4.5.5. [5], [6] modified to handle S^2 order parameters as well as pairwise averaging of NOE distance restraints over replicas [7], as proposed for the MUMO (Minimal Under-restraining Minimal Over-restraining) approach [8]. We used the OPLS-AA force field [9] and the TIP3P water model [10] for all molecular dynamics simulations described below.

For modeling the apo structure of gastrotropin, we chose model 7 of PDB entry 1O1U [11] based on its highest PRIDE-NMR score [12] among the deposited models. The PRIDE NMR is a method developed in our research group in 2007 [12] and it selects the structure from a structural ensemble having the best NOE correspondence by comparing the distribution of the atomic distances with the NOE data using contingency analysis. It derives a score in a 0 and 1 range, which is a similarity probability. It regards the distribution of the experimental data and based on our experience it reliably selects the starting model best representing the experimental data. As an initial model of the holo structure we used model 1 of the PDB entry 2MM3. We generated ligand topologies for glycocholic acid (GCA, PDB ligand ID: GCH) and glycochenodeoxycholic acid (GCDA, PDB ligand ID: CHO) with the TopolGen script (http://www.mdtutorials.com/gmx/complex/02_topology.html) and corrected manually for atom types where it was necessary as well as with an in-house Perl script to reassign hydrogen atoms to the charge groups de-

fined by the heavy atoms they are connected to.

NOE restraints were only available for the *holo* protein (PDB ID: 2MM3). For the *apo* form, we used restraints from the 2MM3 list that were unviolated in the deposited 1O1U structure as checked with the CoNSEnsX server. Restraints were modified by the removal of stereospecificity and rounding the restrained distance up to the next integer Å , creating 1Å wide 'bins' from 4 to 10 Å .

Chemical shifts for the *apo* structure were obtained from BMRB [13] (BMRB ID: 19843) and for the *holo* structure directly from the authors. S^2 values for the *apo* and *holo* structures measured at 283, 291, 298, and 313 K were taken from [14].

After generating a topology using the OPLS-AA force field and TIP3P water model, the molecule was put into a cubic box, followed by energy minimization with conjugate gradient method for 5000 number of steps with 0.001 ps step length. We set the maximum force to 200 $kJmol^{-1}nm^{-1}$. In the next step we solvated the molecule and then replaced one of the water molecules by a Na+ ion to ensure the neutrality of the system. After that, we performed another energy minimization using the same parameters, but including the water molecules. In the last step, a short MD simulation was performed using position restraints of 1000 $kJmol^{-1}nm^{-2}$ on the heavy atoms of the protein for 2500 steps with 0.002 ps step size using the LINCS algorithm [15].

For the production runs, we simulated eight replicas in parallel with the OpenMPI environment [16]. We applied backbone S^2 order parameter restraints on the full ensemble and NOE distance restraints were averaged between neighboring replicas, as it is performed in the MUMO (Minimal Under-restraining, Minimal Over-restraining) protocol [8]. we made the simulations at four temperatures: 283 K, 291 K, 298 K, and 313 K using S^2 restraints measured at the corresponding temperatures. With LINCS constraining on bond lengths, a timestep of 2 fs was used to generate runs of 2 ns and 6 ns, totaling 16 and 48 ns for the 8 replicas combined, respectively. We performed control simulations with the same parametrization but without restraints.

In order to generate a larger pool of possible conformations to further explore the conformational space, we performed molecular dynamics simulations with only one type of restraint (NOE or S^2) or without any restraints. Accelerated Molecular Dynamics and short (500 ps) Targeted Molecular Dynamics simulations were also performed on the *apo* structure using the chemical shifts of the *holo* structure and vica versa in order to achieve transition from one form to the other.

We made 1 μs all-atom simulation on Cd3ϵ in an explicit SPC/E water model with GROMACS (version 2020) using the Amber ff99SB forcefield. After neutralization and a short energy minimization, the production run was preceded by 1 ns NVT and NPT equilibrations. We run the simulation at 300 K using GPU acceleration. We used structures taken at every 50 ps to obtain an ensemble with 20,001 conformers. we retained all structures in order not to reduce conformational variability by omitting the more extended structures at the start of the simulation.

2.3 Principal Component Analysis

The Principal Component Analysis was performed with the Pythn module named ProDy [17], visualization was made with the NMWiz module of the VMD molecular visualization program [18].

2.4 Analysis of the S2 and chemical shift data of gastrotropin

In the case of the S^2 parameters we compared the values of the PDB small ensemble with our ensembles made with the MUMO approach. We excluded points from the analysis where the difference was more than 0.2. In the case of the MUMO simulations it was a maximum of 5 points.

Chemical shift values were back-calculated from the structures with the Shiftx2 program [19].

2.5 Examination of the amide chemical shift differences between gastrotropin structures

For each structure, backbone ^{15}N chemical shifts were estimated with Shiftx [19]. For each conformation in the large conformer pool, the absolute value of the difference of the predicted chemical shifts relative to those in each calculated *apo* structure in the MUMO ensembles was calculated. These differences were then compared to experimental $|\Delta\omega(^{15}\mathrm{N})|$ data derived from CPMG relaxation dispersion NMR measurements for each residue for which it was available [14]. Both correlation and RMSD measures were calculated after normalization to the 0–1 range. As there are $|\Delta\omega|$ values available for three temperatures and the conformational pool is of a heterogeneous source with no well-defined temperature, the correlation and RMSD values were calculated for all three temperatures and then were averaged for each structure investigated. The structures with highest correlation and lowest RMSD values were selected for analysis.

3 New scientific results

3.1 First thesis gorup

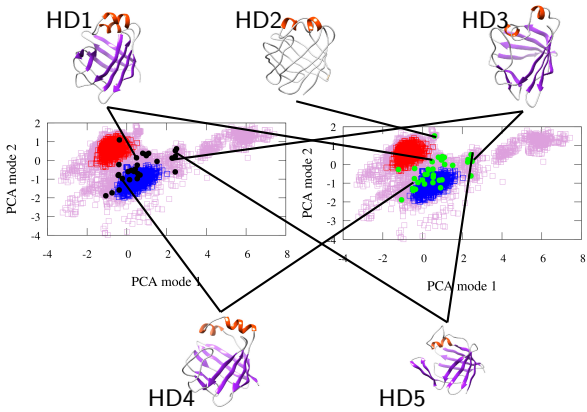

Figure 1: **PCA scatter plot of the apo MUMO (red dots) and holo MUMO (blue dots) ensembles along with the conformer pool (purple hollow squares) used to select the structures best corresponding to the NMR-derived invisible state.** Structures with a mean correlation between ^{15}N values and $|\Delta\omega(^{15}\mathrm{N})|$ calculated chemical shift differences above a threshold of 0.35 are shown with black dots (left panel). Structures with an RMSD between $|\Delta\omega(^{15}\mathrm{N})|$ values and calculated chemical shift differences lower than 0.00603 are depicted with green dots (right panel). Selected structures are also depicted and linked to their corresponding points in the PCA scatter plots. These hidden conformations are termed HD1-HD5.

Thesis 1.1.A: **I generated structural ensembles of the protein gastrotropin with molecular dynamics simulations restrained by S^2 and NOE parameters using the MUMO approach on four temperatures, which ensembles correspond to the chemical shift parameters not included in the model**

Chemical shifts from the MUMO generated ensembles showed better correspondence to the chemical shift parameters not included in the model, compared to the small ensembles found in the PDB database (table 1).

	Ensemble size	**Tempera-ture (K)**	**Main chain RMSD (Å)**	**Main chain S^2 correlation**	**Corrected main chain S^2 correlation**	**Violated NOE restraints % ($r^{-6} > 0.5$A)**	**Amide N chemical shift**	**Amide H chemical shift**
1O1U pdb	10	305	0.66+-0	0.502[a]	0.857	0.00	*0.709*	0.596
apo MUMO	168	283	1.01+-0.03	0.675	0.873	0.29	**0.832**	**0.72**
apo MUMO	168	291	1.12+-0.04	0.8	0.841	0.29	**0.828**	**0.729**
apo MUMO	168	298	1.27+-0.06	0.797	0.883	0.12	**0.834**	**0.738**
apo MUMO	168	313	1.45+-0.04	0.963	0.972	-	**0.834**	**0.705**
apo unrestr	48	283	1.46+-0.09	0.307	-	-	**0.827**	**0.724**
apo unrestr	48	291	1.45+-0.05	0.351	-	-	**0.832**	**0.703**
apo unrestr	48	298	1.53+-0.03	0.301	-	-	**0.833**	**0.731**
apo unrestr	48	313	1.5+-0.07	0.6	-	-	**0.829**	**0.733**
2MM3 pdb	10	293	0.52+-0	0.297[b]	0.315	0.00	0.773	0.593
holo MUMO	168	283	1.28+-0.03	0.726	0.756	0.41	**0.784**	*0.548*
holo MUMO	168	291	1.31+-0.07	0.544	0.731	0.46	**0.78**	*0.589*
holo MUMO	168	298	1.46+-0.15	0.884	0.930	0.36	*0.767*	*0.552*
holo MUMO	168	313	1.5+-0.08	0.724	0.870	0.36	*0.758*	*0.551*
holo unrestr	48	283	1.92+-0.64	0.073	-	-	**0.782**	*0.531*
holo unrestr	48	291	2.34+-0.65	0.296	-	-	**0.777**	*0.537*
holo unrestr	48	298	1.82+-0.07	0.484	-	-	**0.772**	*0.493*
holo unrestr	48	313	2.59+-2.09	0.272	-	-	**0.775**	*0.544*

Table 1: RMSD values, S^2 parameters, NOE correspondence, amide N and H chemical shift correlations

[a] S^2 data of 313 K (highest correlation from the 4 datasets)
[b] S^2 data of 298 K (highest correlation from the 4 datasets)

Thesis 1.1.B: **I generated using different kinds of molecular dynamics simulations a structural ensemble of many structures for modelling the protein gastrotropin. By using Principal Component Analysis I demonstrated that the ensemble comprises a big part of the conformational space.**

This ensemble is diverse and wanders along a big conformational space, as it is shown in figure 1.

Thesis 1.2.A: **I showed on the structural ensemble generated with S^2 and NOE restrained molecular dynamics simulations, that two kind of motions dominate in the *apo* and the *holo* form, both are resulting in the opening of the barrel.**

These motions take place along the first two modes of the Principal Component Analysis (PCA) (figure 2), the E-F and G-H loops taking part mostly in the first motion, while in the second motion the C-D loop and the helix moves the most.

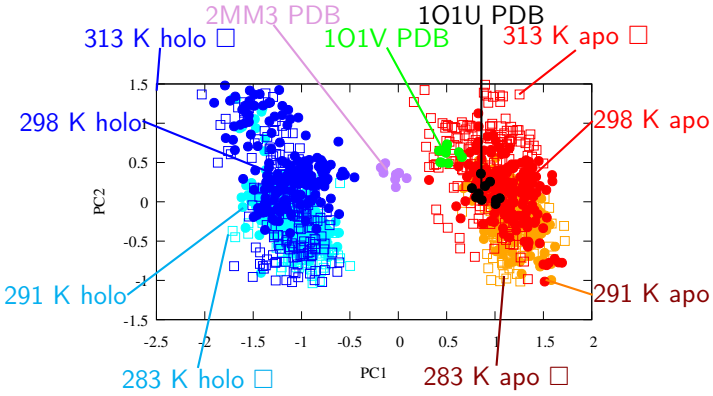

Figure 2: PCA (Principal Component Analysis) scatter plot of the simulated and experimentally determined conformer pool.

Thesis 1.2.B: **The group of amino acids taking part in the Type II motion based on the PCA anslysis are are greatly overlapping with the group of amino acids taking part in the slow timescale motion measured by Orsolya Tőke and her research group.**

See figure 3. We could not investigate this connection for every amino acid, because they could not measure every single amino acid.

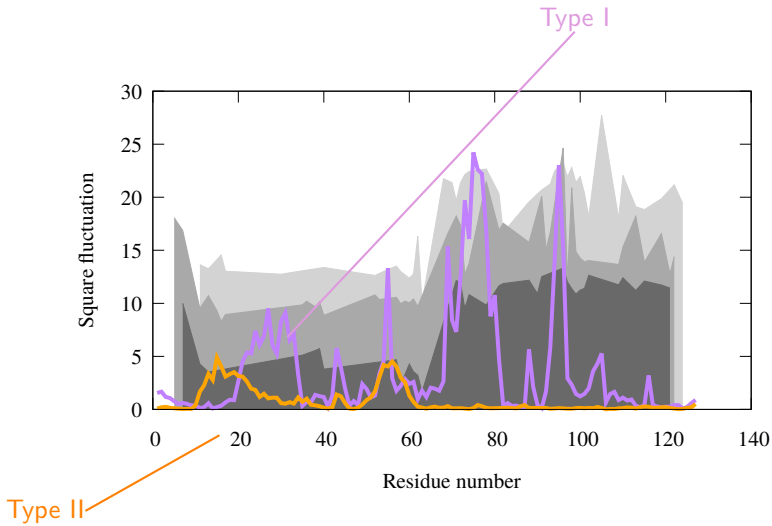

Figure 3: **Square fluctuation of Cα atoms in the two PCA modes: PC1 (Type I motion, purple), PC2 (Type II motion, orange).** The previously measured experimental k_{ex} values indicating two distinct clusters of residues involved in slow conformational exchange processes are depicted as different gray areas corresponding to the three different temperatures (283 K, 287 K, 291 K) of the measurements. As only about 30–40 amino acids have displayed ms timescale motion with measurable k_{ex} values [14], a continuous depiction is used to guide the eye to highlight the regional differences.

Thesis 1.3: **I determined a mechanism for the entry of the ligands into the binding sites, in which the helical cap is partially unfolding.**

Taking the chemical shift difference between each structure and the few hidden energy state structures determined by Orsolya Tőke and her research group, the most corresponding structures are half unfolded, mainly in the helical cap (figure 4). Although the forcefield can contribute to this as well.

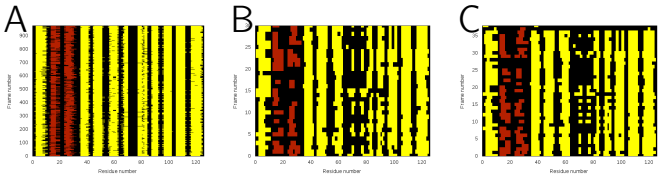

Figure 4: **Secondary structure of the conformations inferred from our simulations (rows).** Each column represents one amino acid. Extended β-strands are colored yellow, α-helices are brown, the rest of the residues are colored black. (A) All of the conformations. (B) The high correlation conformations (subset of conformations of panel A). (C) Conformations with lowest RMSD (another subset of panel A). The analysis was performed with DSSPCont [20]. Note the shortening of secondary structure elements in some structures, especially in B) and C).

3.2 Second thesis group

Thesis 2.1: **I created the DIPEND program suitable for building an atomic level ensemble model of protein segments**

The program using a probabilistic neighbor depenent model solves the problem of creating an ensemble of protein conformers, which is a good sampling of the conformational space around relevant conformations of the protein segment (figure 5). The additional distribution, which can be added with a weight can mirror the *a priori* known conformational preferences. The program can reliably build protein segments up to about 100 amino acid length, the probability of steric clashes increases in the case of segments longer than 100 amino acids. Hopefully this can be extended using the modes with custom weighting and the unknotting module, if the user gives a distribution resulting in an extended conformation. But this increases the running time. On the whole the program can generate a useful starting ensemble, from which a subensemble can be selected using other programs based on experimental data.

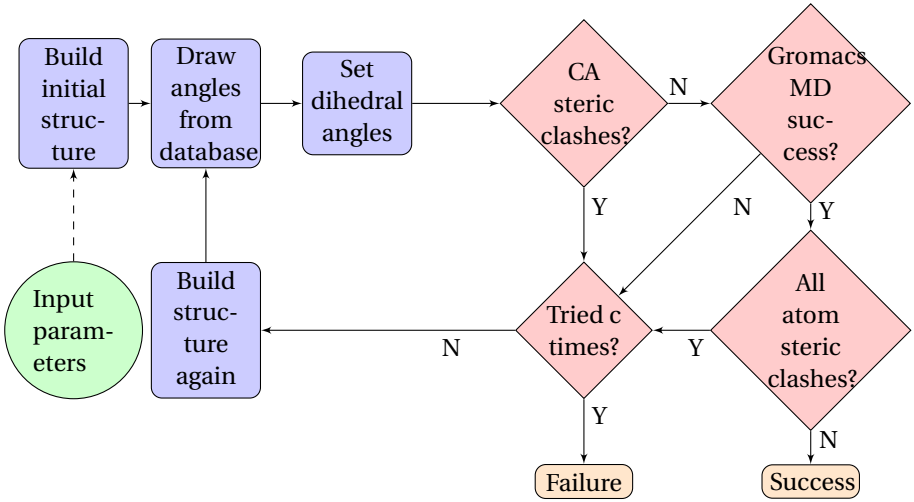

Figure 5: Flowchart of the steps of the DIPEND (DIsordered Protein Ensembles from Neighbor-dependent Distributions) program

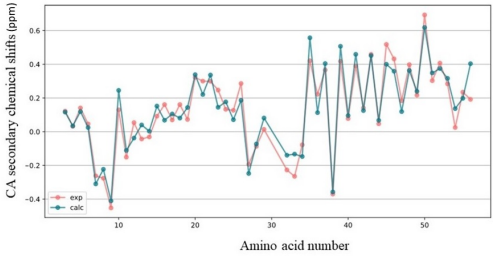

Figure 6: Observed and calculated CA secondary chemical shifts for the CD3 ensemble.

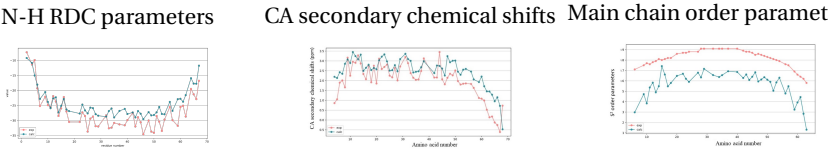

Figure 7: Measured (red) and calculated (blue) values of some NMR parameters for the selected SAH ensemble. Calculated values were obtained with CoNSensX+.

Thesis 2.2: **I applyied the program to the intrinsically disordered region of the CD3ϵ protein and I demonstrated that the conformational ensemble gemerated by the DIPEND program covers greater part of the conformational ensemble than the ensemble derived from other methods. I also demonstrated that the DIPEND-generated ensemble corresponds better to the experimental data.**

See figure 6.

Thesis 2.3: **I applied the DIPEND program to the SAH domain of the myosin VI protein and I showed that the ensemble generated by the DIPEND program is able to sample relevant parts of the conformational space also in the case of a protein having well defined previously known structural preferences and from this conformational ensemble a subensemble can be selected which corresponds to the experimental data.**

The correlation of the correspondence of the generated ensemble is greater than for the ensemble derived from the PDB database (figure 7).

4 Publications

This dissertation is based on the following publications:

- Harmat Z, Szabó AL, Tőke O, Gáspári Z (2019) Different modes of barrel opening suggest a complex pathway of ligand binding in human gastrotropin. *PLOS ONE* 14(5): e0216142. https://doi.org/10.1371/journal.pone.0216142

- Harmat Z, Dudola D., Gáspári Z (2021) DIPEND: An Open-Source Pipeline to Generate Ensembles of Disordered Segments Using

Neighbor-Dependent Backbone Preferences. *Biomolecules* 11(10): 1505. https://doi.org/10.3390/biom11101505

More publications:

- Harmat, Z. (2017) Generation and analysis of dynamic structural ensembles of human gastrotropin PhD PROCEEDINGS ANNUAL ISSUES OF THE DOCTORAL SCHOOL FACULTY OF INFORMATION TECHNOLOGY & BIONICS 2017, HU ISSN 2064-7271

- Harmat, Z. (2018) Characterisation of the Subunit Composition, Stochiometry and Structure of Calcium Channels Involved in Neurotransmission PhD PROCEEDINGS ANNUAL ISSUES OF THE DOCTORAL SCHOOL FACULTY OF INFORMATION TECHNOLOGY & BIONICS 2018, HU ISSN 2064-7271

- Harmat, Z. (2019) Building the three-dimensional structural model of the postsynaptic protein GKAP PhD PROCEEDINGS ANNUAL ISSUES OF THE DOCTORAL SCHOOL FACULTY OF INFORMATION TECHNOLOGY & BIONICS 2019, HU ISSN 2064-7271

- Harmat, Z. (2020) Development and Optimization of a Pipeline to build Structural Ensemble Models of Intrinsically DisorderedProtein Segments PhD PROCEEDINGS ANNUAL ISSUES OF THE DOCTORAL SCHOOL FACULTY OF INFORMATION TECHNOLOGY & BIONICS 2020, HU ISSN 2064-7271

References

[1] Philip E. Bourne (Editor) Jenny Gu (Editor). *Structural Bioinformatics, 2nd Edition*. J. Wiley & Sons, 2 edition, 2009.

[2] Eric F Pettersen, Thomas D Goddard, Conrad C Huang, Gregory S Couch, Daniel M Greenblatt, Elaine C Meng, and Thomas E Ferrin. UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem*, 25(13):1605–1612, Oct 2004.

[3] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Elaine C. Meng, Gregory S. Couch, Tristan I. Croll, John H. Morris, and Thomas E. Ferrin. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci*, 30(1):70–82, Jan 2021.

[4] Georgii G Krivov, Maxim V Shapovalov, and Roland L Dunbrack. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins*, 77(4):778–795, Dec 2009.

[5] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E. Mark, and Herman J. C. Berendsen. Gromacs: Fast, flexible, and free. *Journal of Computational Chemistry*, 26(16):1701–1718, 2005.

[6] Sander Pronk, Szilárd Páll, Roland Schulz, Per Larsson, Pär Bjelkmar, Rossen Apostolov, Michael R. Shirts, Jeremy C. Smith, Peter M. Kasson, David van der Spoel, Berk Hess, and Erik Lindahl. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, 02 2013.

[7] Ádám Fizil, Zoltán Gáspári, Terézia Barna, Florentine Marx, and Gyula Batta. "Invisible" conformers of an antifungal disulfide protein revealed by constrained cold and heat unfolding, CEST-NMR experiments, and molecular dynamics calculations. *Chemistry*, 21(13):5136–5144, Mar 2015.

[8] Barbara Richter, Joerg Gsponer, Péter Várnai, Xavier Salvatella, and Michele Vendruscolo. The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. *J Biomol NMR*, 37(2):117–135, Feb 2007.

[9] D Kony, W Damm, S Stoll, and WF Van Gunsteren. An improved OPLS-AA force field for carbohydrates. *J Comput Chem*, 23(15):1416–1429, Nov 2002.

[10] William L Jorgensen, Jayaraman Chandrasekhar, Jeffry D Madura, Roger W Impey, and Michael L Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[11] Michael Kurz, Volker Brachvogel, Hans Matter, Siegfried Stengelin, Harald Thüring, and Werner Kramer. Insights into the bile acid transportation system: the human ileal lipid-binding protein-cholyltaurine complex and its comparison with homologous structures. *Proteins*, 50(2):312–328, February 2003.

[12] Annamária F. Ángyán, Balázs Szappanos, András Perczel, and Zoltán Gáspári. Consensx: an ensemble view of protein structures and nmr-derived experimental data. *BMC Structural Biology*, 10(1):39, Oct 2010.

[13] Eldon L. Ulrich, Hideo Akutsu, Jurgen F. Doreleijers, Yoko Harano, Yannis E. Ioannidis, Jundong Lin, Miron Livny, Steve Mading, Dimitri Maziuk, Zachary Miller, Eiichi Nakatani, Christopher F. Schulte, David E. Tolmie, R. Kent Wenger, Hongyang Yao, and John L. Markley. BioMagResBank. *Nucleic Acids Research*, 36(suppl_1):D402–D408, 11 2007.

[14] Gergő Horváth, Orsolya Egyed, and O Toke. Temperature dependence of backbone dynamics in human ileal bile acid-binding protein: implications for the mechanism of ligand binding. *Biochemistry*, 53(31):5186–5198, Aug 2014.

[15] Berk Hess, Henk Bekker, Herman JC Berendsen, and Johannes GEM Fraaije. Lincs: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[16] Edgar Gabriel, Graham E. Fagg, George Bosilca, Thara Angskun, Jack J. Dongarra, Jeffrey M. Squyres, Vishal Sahay, Prabhanjan Kambadur, Brian Barrett, Andrew Lumsdaine, Ralph H. Castain, David J. Daniel, Richard L. Graham, and Timothy S. Woodall. Open mpi: Goals, concept, and design of a next generation mpi implementation. In Dieter Kranzlmüller, Péter Kacsuk, and Jack Dongarra, editors, *Recent Advances in Parallel Virtual Machine and Message Passing Interface*, pages 97–104, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.

[17] Ahmet Bakan, Lidio M Meireles, and Ivet Bahar. ProDy: protein dynamics inferred from theory and experiments. *Bioinformatics*, 27(11):1575–1577, Jun 2011.

[18] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: visual molecular dynamics. *J Mol Graph*, 14(1):33–38, Feb 1996.

[19] Beomsoo Han, Yifeng Liu, Simon W Ginzinger, and David S Wishart. SHIFTX2: significantly improved protein chemical shift prediction. *J Biomol NMR*, 50(1):43–57, May 2011.

[20] Claus AF Andersen, Arthur G Palmer, Seren Brunak, and Burkhard Rost. Continuum secondary structure captures protein flexibility. *Structure*, 10(2):175–184, Feb 2002.