NKFIA project 104198

Generation and analysis of dynamic conformational ensembles of biomolecules

Final report

Establishment of a novel research group at PPCU FIT

The aim of the NF scheme was to establish a novel research group by an early-career researcher. I have used the grant period to build a group in the field of protein structural studies at the Faculty of Information Technology and Bionics of Pázmány Péter Catholic University. After a relatively slow start - the first new PhD students joined in September 2014 -, and many unavoidable fluctuations - one PhD student left for the US and a number of undergraduates continued their studies elsewhere or had detours during their undergraduate years -, the group at the closing date of the grant period consisted of two postdoctoral researchers, six PhD students and over 5 BSc and MSc students. The majority of these group members work on a topic directly related to dynamic protein ensembles, but the overall scope is somewhat wider and includes experimental work also. Overall, the NKFIA funding provided a substantial contribution to the establishment of the group as it exists now.

Methodological developments

An important aspect of the work undertaken was the development and improvement of several methods used for the generation and analysis of dynamic structural ensembles and related aspects of protein structure and flexibility.

GROMACS additions

We have implemented several additions to the freely available GROMACS molecular dynamics package. The majority of our code, except for the parts still under testing, are freely available on our webpage at *users.itk.ppke.hu/~gaszo*. The version 4.5.5. Used as a basis for the extensions provides force fields and other features like implicit solvent support that are routinely used in our ensemble-based calculations. Additions implemented during the course of the project are:

- Support for applying NOE-derived distance restraints with pairwise averaging over the replicas (as described in the MUMO protocol by Richter et al. 2008).
- Support for S^2 order parameter restraints (Best & Vendruscolo 2004), with enabling local fit for parameter calculation for molecules that can not be described by a single overall rotational correlation time like multidomain proteins with flexible linkers. S^2 parameters can also be applied to subpopulations of the full simulated ensemble.
- Support for the AMD (accelerated molecular dynamics) scheme (Hamelberg) for the dihedral restraint energy term to enhance conformational sampling.
- Experimental support for targeted molecular dynamics based on CA-CA distances as restraints implemented as half-harmonic potential depending on a progress variable. The concept is highly similar to that applied for S^2 order parameter restraints.
- Experimental support for B-factor restraining (Clore & Schwieters 2006) to assess the dynamics in the crystal state.

All novel features are implemented in a user-friendly way by allowing parameterization and restraint definition using mdp and topology files according to the existing logic of GROMACS. The most complete description of the implemented stable features can be found in [IV].

The CoNSEnsX$^+$ webserver

We have put serious efforts to completely redesign and reimplement the CoNSEnsX service capable of evaluating dynamic structural ensembles with relation to NMR experimental parameters. To our knowledge, this is the only web service dedicated specifically to the analysis of protein structural ensembles reflecting internal dynamics. Such ensembles can be much more diverse than conventional NMR ensembles. Our server is designed to handle file formats that can be directly obtained from public databases such as BMRB.

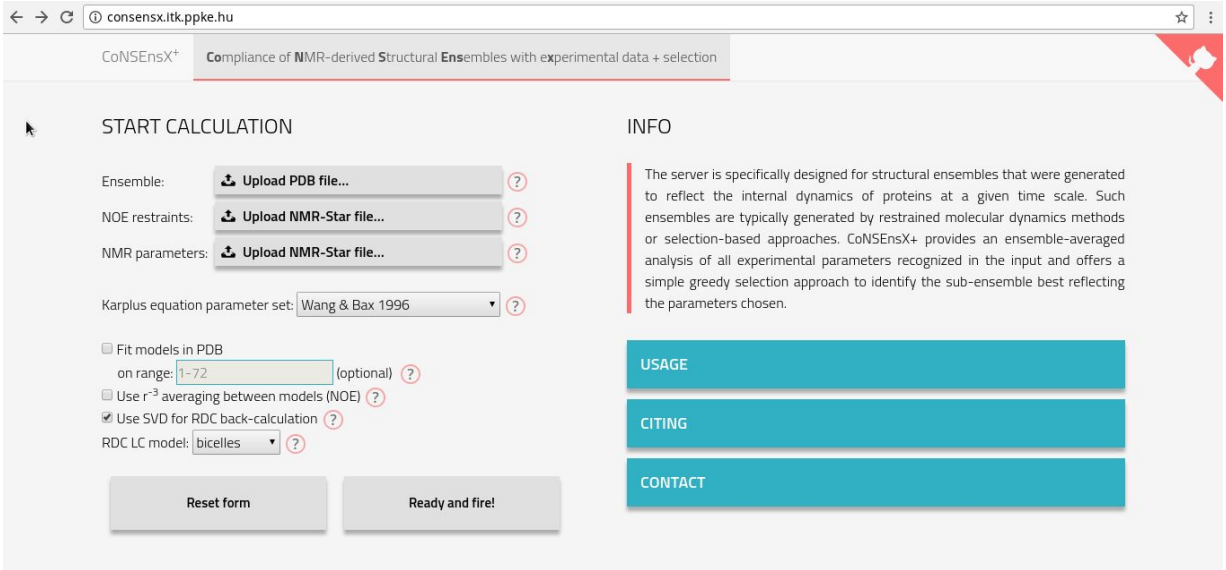
The main novel features of the CoNSEnsX$^+$ service are:

- a complete redesign with a python core performing the calculations
- added support for side-chain S^2 parameters
- support for multiple Karplus equation parameter sets for the back-calculation of scalar couplings
- independent handling of multiple residual dipolar coupling experiments with different molecular alignments
- a greedy selection algorithm to select a sub-ensemble best corresponding to the chosen parameters according to a selected measure (correlation, RMSD or Q-factor)
- enhanced user interface with vector graphics and principal component analysis plot comparing the originally submitted and selected ensembles.

CoNSEnsX$^+$ is freely available at *consensx.itk.ppke.hu*, and is listed in the home page of the West-Life project (west-life.eu) as a 3rd party service. The source code written by our group is also fully available from GitHub (*github.com/PPKE-Bioinf/consensx.itk.ppke.hu*).

Applications of the present form of the CoNSEnsX$^+$ webserver can be, but are not necessarily limited to:

- to evaluate the correspondence of structural ensembles generated with any method, especially to perform cross-validation, i.e. examine the correspondence of the ensemble to parameters not used in its generation
- select a sub-ensemble to assess potential overfitting (i.e. whether the number of conformers in the ensemble are larger than minimally needed to account for the experimental observations)
- generate an ensemble corresponding to specified parameters using a large conformer library submitted (created by any means of enhanced conformational sampling)



Home page of the CoNSEnsX$^+$ web server at *consensx.itk.ppke.hu*

The CoNSEnsX$^+$ server is described with example applications in [X].

Enhanced validation of the geometry of structural ensembles

Structure validation is an important process in all structure determination pipelines. For dynamic structural ensembles, it is also highly important that all conformers in the ensemble should be realistic. However, given that many such ensembles are generated with restraining, deviations from ideal geometry can occur and in some cases can seriously affect the correspondence to parameters and undermine the reliability of the ensemble. This can be especially the case for amide N-H groups as many NMR parameters (S^2, RDC, NOE) and thus restraints can relate to the orientation of these groups. The CoNSENsX$^+$ service contains a module (not yet available from the web interface) capable of assessing the planarity of amide groups with special emphasis of out-of-plane position of the amide hydrogen atoms.

Another aspect of structural ensemble validation is Ramachandran analysis. In this respect, many currently available tools restrict the number of structures that can be analyzed, and are thus not applicable to large structural ensembles. In addition, the analysis should take into account recent results concerning the neighbour-dependent conformational preferences of residues, as well as the unusual behaviour of proline and glycine residues. Both aspects are especially crucial for the ensembles of intrinsically disordered proteins. Therefore, we set out to implement an analysis pipeline based on a relatively recently described neighbour-dependent distribution (Ting et al. 2010). Large-scale analysis of a number of PDB.deposited structures and comparison of the results obtained with traditional classification (Morris et al. 1992), as well as applications on dynamic structural ensembles generated in our group is under way. Integration of this analysis into the CoNSEnsX$^+$ service is planned. The code is still under development, a test version is currently available at *pev-webapi.azurewebsites.net/swagger/*.

Tools for the analysis of NOE restraints

In conjunction with the CoNSEnsX$^+$ project, we developed a set of tools specifically for the analysis of structural ensembles in terms of their correspondence to NOE-based distance restraints. A series of scripts for the conversion and detailed analysis of NMR restraint lists available from the PDB have been developed in order to provide a standardized pipeline for such analysis. Special care is taken of the stereochemical aspects of restraints and the different nomenclatures used for prochiral protons. A comprehensive analysis of more than 7000 structures from the PDB is currently under way and detailed statistics of the correspondence of restraints to structures is generated. A number of scripts are already available from the description page of the CoNSEnsX$^+$ server and new/updated ones will be added gradually. Full integration of the analysis onto the CoNSEnsX$^+$ web interface will be considered depending on our experience with the analysis in progress.

Structure comparison using Ramachandran angles on FPGA platform

We have initiated the development of a protein structure comparison approach based on backbone dihedral angles. The method is similar to a previously described one but with the important difference that we calculate the similarity scores based on the distance of the angles in the Ramachandran space. As the method is currently implemented on FPGA, we will be able to perform optimization of the exact comparison parameters by running comprehensive calculations on very large data sets. The method is expected to be useful as a preliminary filtering step before more accurate algorithms as well as a useful approach to identify evolutionarily related proteins with structural rearrangements. The current implementation of the method is described in [XI].

Update to the CSAH server

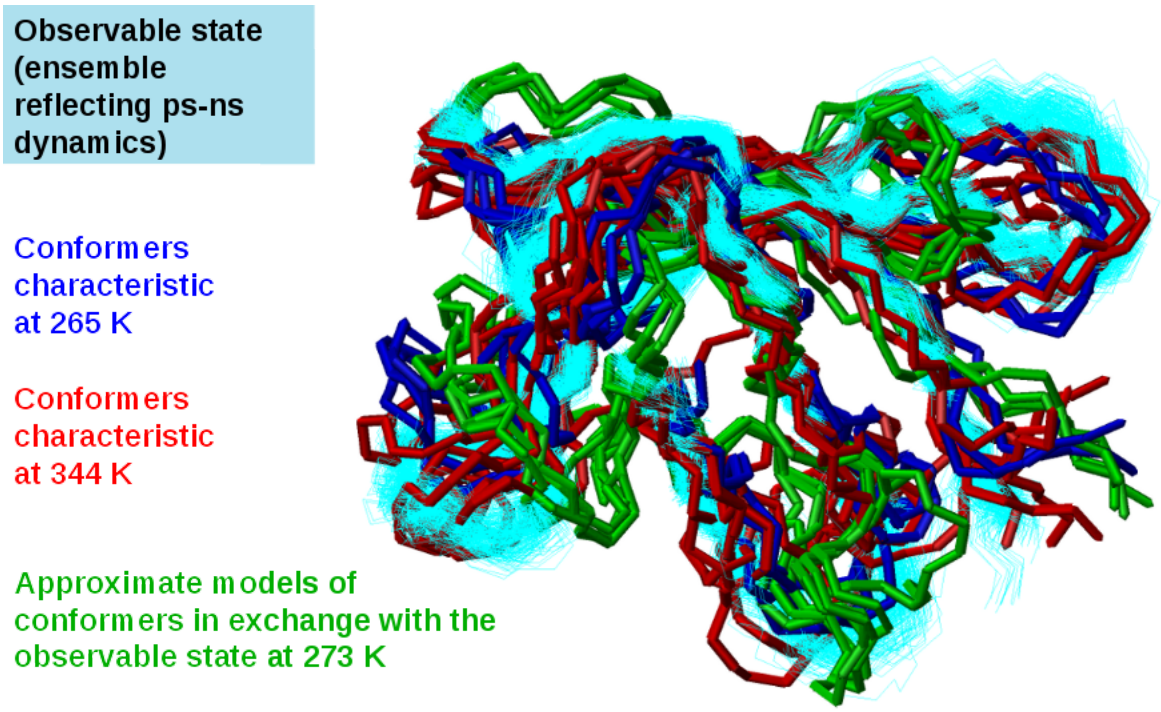
Charged single alpha-helices are a rare structural motif formed by highly charged sequences with a characteristic alternating charge pattern (Süveges et al. 2009). Such segments are often predicted to be intrinsically disordered, and thus their correct prediction is crucial in our understanding of protein disorder and mobility. To this end, we have updated our web server capable of predicting charged single alpha helices from protein sequences. The main concept of the server is to be highly specific even if it compromises its sensitivity, as for a novel and rare structural motif we intended to avoid false positive predictions as much as possible. The server, originally established in 2012 (Gáspári et al. 2012), uses two algorithms, SCAN4CSAH and FT_CHARGE, both based on identifying the characteristic charge pattern of CSAH segments regardless of their exact amino acid distribution. During the update, besides relocating the server to PPCU FIT and giving it a new interface, we have added a filter based on simple Chou-Fasman secondary structural preferences to reduce the number of segments that are unlikely to be alpha-helical despite having the charge pattern characteristic of CSAHs. With this filter, several hits with relatively high proline content have been successfully eliminated from the result lists. The server is available at *csahserver.itk.ppke.hu* and its detailed usage is described in [VII]

Applications on proteins

The methods developed have been applied to a number of proteins to get insight into the role of their internal motions during their function. While we attempted the analysis of more systems than listed below, we chose to summarize only those with results already published or being relatively close to publication, the corresponding manuscripts expected to be submitted within the next 4-6 months.

Disulfide-rich cationic antifungal proteins

Investigation of disulfide-rich cationic antifungal proteins is performed in collaboration with Dr. Gyula Batta from the University of Debrecen. Using data measured in his laboratory, we have described the hot and cold unfolding of PAF (Penicillium AntiFungal protein) by comparing the structural ensemble representing its state observable at room temperature with conformers selected to conform to the parameters measured at the hot and cold unfolded states. Our results revealed that although hot and cold unfolding primarily affects the loop regions of the protein, the two states are not identical to each other. In addition, we have also generated an approximate model from very sparse data obtained for a low-populated state observed by N^{15} CEST measurements. The conformers obtained represent yet another different state and hint at differences at the terminal regions of an otherwise compact molecule. Our calculations helped to propose a model explaining the unexpected observation that a large portion of PAF molecules is not observable even at the temperature with the largest maximum stability population.Our results are published in [IV].



Structural models of PAF at various states as deduced from temperature-dependent NMR measurements as well as CEST experiments

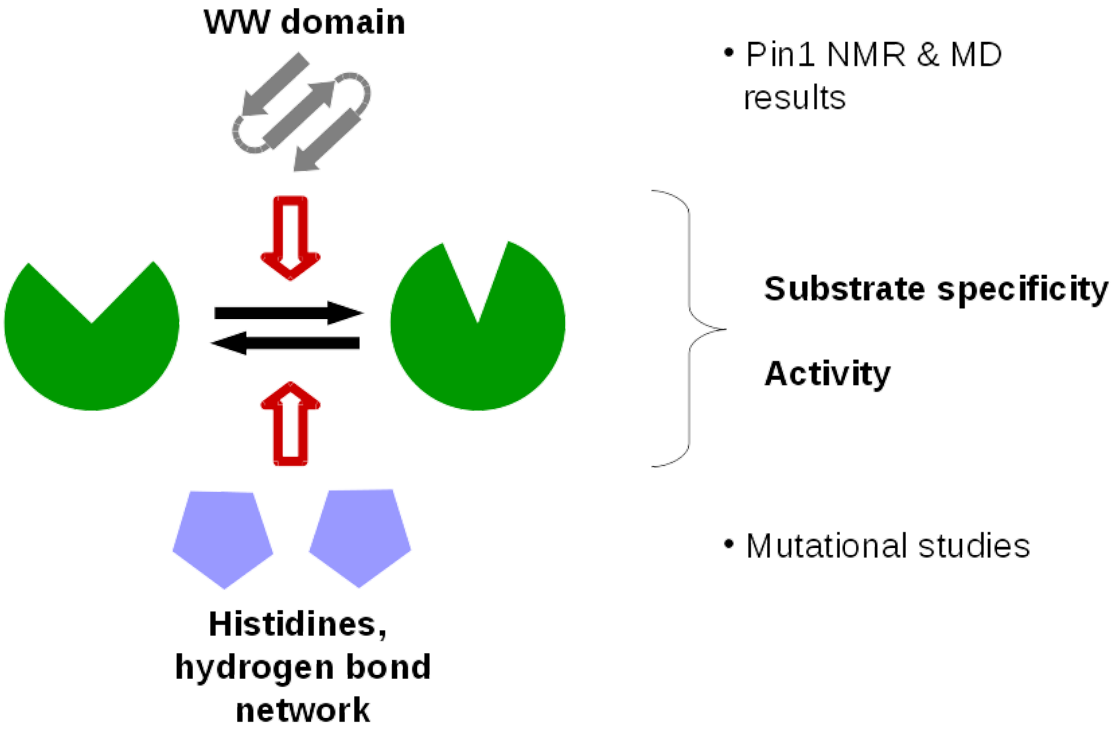
The D19S variant of PAF was shown to possess diminished antifungal activity compared to the wild type molecule. NMR measurements revealed that the overall fold of the molecule is largely unaffected by the mutation. However, detailed analysis of the surface charge distribution revealed characteristic differences between PAF and PAF^{D19S}, pointing to a possible source of the functional differences. The charge redistribution has been speculated to arise from the removal of a negative surface charge forcing a redistribution of the surface lysine side chains according to the repulsive interactions between their amino groups. We have performed preliminary calculations to obtain structural ensembles to reflect the internal dynamics of PAF^{D19S} and compared it to the ensemble available for PAF. As these ensembles are much more diverse than those obtained by conventional structure calculations, they can be used to rule out that specific side-chain conformation patterns were observed by biases in the original structure determination procedure. The ensembles were largely in accordance with the charge redistribution scenario proposed on the basis of the originally determined structures. In summary, the observed changes in the activity are likely due to the different surface charge properties of the two variants. Our results are published in [VIII].

Most recently, together with Dr. Batta's group, we have performed structure analysis of PAFB, another member of the disulfide-rich cationic antifungal protein family. Again, we could point out the different surface charge properties of PAFB relative to PAF. More detailed investigations of the observed surface properties with the aid of structural ensembles of PAFB are currently under way. Part of our results on the structural analysis of PAFB are included in a manuscript submitted for publication [XII].

Parvulin-type peptidyl-prolyl cis-trans isomerases

Parvulins are among of the most intensively studied protein families due to their size and biological importance, including, by some family members, modulation of molecular recognition processes by isomerization of prolines next to serine and threonine residues prone to phosphorylation. The best described parvulin-type prolyl isomerase is Pin1. In search for proteins with available NMR dynamics parameters, we have identified three parvulins, TbPin1, CsPinA and SaPrsA from three different organisms. Importantly, each of these contains only the isomerase domain and no WW domain that is characteristic to Pin1 and many other parvulins. Using the available NOE and backbone S^2 data we generated structural ensembles of all three proteins and performed a detailed comparative analysis of these with each other and other known parvulin structures using, among others, multiple structure alignment and principal component analysis performed on the CA atoms common to all analyzed molecules as identified in the structure alignment. As a result, we have identified characteristic differences in the openness of the substrate-binding site of different parvulins, which was more pronounced in the dynamic ensembles than in the originally deposited ones. Next, we noticed that a characteristic structural difference, corresponding to the 2^{nd} principal component in the comparison of all parvulin domains, is located at the hinge region between the two lobes flanking the substrate binding site. This region is very close to the residues for which the largest chemical shift differences were observed in Pin1 during interaction with its WW domain (Wang et al. 2015). Based on these observations we suggest that the opening-closing motion of the substrate-binding site is key in the activity and its regulation of parvulins and interactions with the WW domain modulates this motion. We further investigated the relative positions of five residues, including two largely conserved histidines, forming a hydrogen-bond

network in one of the lobes flanking the active site. The positions of these residues relative to each other show a weak but detectable connection with the openness of the substrate-binding cleft. Altogether we proposed a simple unified model of parvulins where the extent and dynamics of binding site opening can be modulated by an internal network of largely conserved residues via their hydrogen bonding pattern as well as by interactions with a WW domain, if present. We note that this model is compatible with experimental findings published after our paper was accepted (Wang et al. 2017). Our results are published in [IX] and were presented at a number of conferences.



A unified schematic model for parvulin activity and its regulation based on our comparative analysis of dynamic structural ensembles

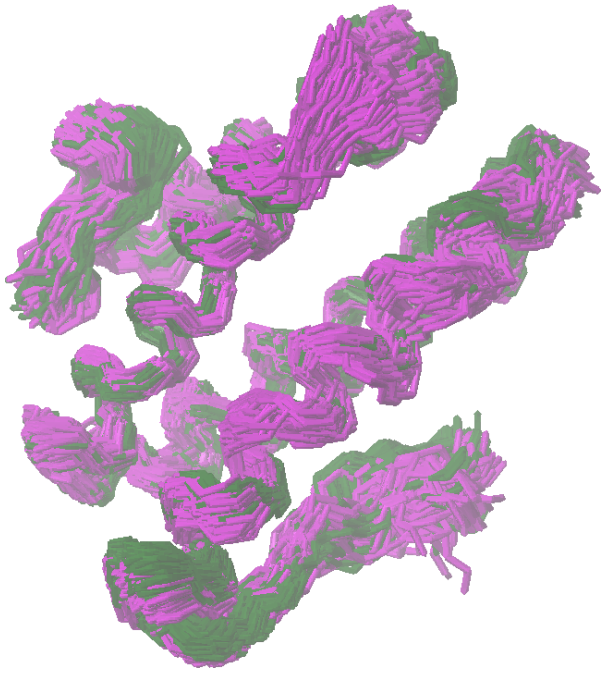
PDZ domains

The postsynaptic density protein 95 (PSD-95) is an intracellular synaptic protein located in the postsynaptic site. In humans, it is encoded by the discs large homolog 4 (*DLG4*) gene. Along with a number of its homologs (e.g. SAP97, PSD-93 or SAP-102), it is a member of the membrane-associated guanylate kinase (MAGUK) superfamily. The first two PDZ domains of PSD-95 are organized in tandem: they are connected to each other through a short flexible linker. In PDZ tandems, close proximity is required for interdomain communication, as well as for proper function and ligand specificity which is usually different from the individual domains. It was shown earlier that ligand binding enhances interdomain mobility in the PDZ1-PDZ2 tandem of PSD-95. The exact mechanism of interdomain communication and allostery, however, remains poorly understood.

We have generated dynamic structural ensembles of the PDZ1-PDZ2 tandem module pair of PSD95 using NOE and backbone S^2 order parameter restraints. Order parameter restraints were applied separately for the two domains, as allowed by our GROMACS extension, accounting for their largely independent overall motions. Detailed analysis of the liganded and unliganded ensembles and their structural and dynamical features allowed us the identification of structural rearrangements possibly responsible for the differences in the interdomain mobility between the liganded and unbound forms. Validation of these rearrangements in a broader context of other PDZ domains with known structure is currently under way. The initial results of the work have been presented at several Hungarian and international conferences.
The 3rd PDZ domain of PSD95 contains an unusual helical extension that was suggested to contribute to intradomain allostery by modulating the conformational entropy of the domain. We have initiated a comprehensive analysis of structural ensembles of this domain with and without the helical extension, both in the ligand-bound and unliganded forms in order to get more insight to the structural factors responsible for the differences between the variants. We have performed a number of entropy calculations on the ensembles with different parametrizations. We currently have preliminary results with some clear trends emerging between the studied systems. Detailed validation of these including assessment of the dependence of the observed trends on ensemble size is under way.

Rev1CT

The Rev1 polymerase is part of the translesion synthesis process whereby specialized DNA polymerases help rescue DNA replication blocked by damage sites. The C-terminal (CT) domain of Rev1 binds Rev1-interacting regions (RIR) of other DNA polymerases and Rev7 protein via two distinct binding regions on the opposite faces of the domain. The current view is that these binding processes do not significantly alter the 3D structure of the protein according to currently available data. Using NOE and backbone S^2 order parameters, we have performed extensive restrained molecular dynamics calculations on Rev1-CT both in its free form and bound to a RIR-peptide derived from its partner polymerase η. Our calculations with different restraint combinations and setups revealed subtle but clearly detectable structural changes occurring upon binding. The analysis of the robustness of our results is in progress.



Calculated ensembles of Rev1CT using NOE and backbone S^2 data. Green: unbound conformers, magenta: bound conformers.

Human ileal bile acid binding protein

The protein gastrotropin, also known as Intestinal bile acid-binding protein, binds bile salts in the cytoplasm of the epithelial cells of the ileum. Gastrotropin plays an important part in the enterohepatic circulation of bile salts, which is an important part of digestion, because it ensures that only a very small part of bile salts must be synthesized de novo (Smathers & Petersen 2011) (Horváth et. al. 2012). The protein binds its ligands in a cooperative manner (Tochtrop et al. 2002). NMR experiments revealed that in the apo state there are two conformers in slow exchange with each other, one of them making up only a small fraction of the population of states and being similar to the holo form (Horváth et. al. 2014) identifying the ligand binding mode as conformational selection.

We have generated a number of large dynamic conformational ensembles of gastrotropin in the free and bound states using NMR-derived parameters as restraints and performed a comprehensive analysis of these and other available structures of gastrotropin-like proteins. With these analyses we have identified two types of characteristic concerted internal motions possibly related to ligand binding. The first motion corresponds to a difference already described between the unbound and bound structures but is more pronounced. The second one hints at a different motion that we suspect is also related to ligand entry into the internal binding sites. This possibility is currently under investigation by comparing the geometry features of conformers representing different states along this motional mode with the set of residues involved in exchange processes at different time scales as identified by NMR. In addition. we performed extensive docking studies using Schrödinger Glide to assess how the motions might affect the ligand-binding properties of the molecule. Using all these results, we are currently devising a model that could account for the role of internal dynamics of gastrotropin in ligand recognition.

The initial results of the work have been presented at several conferences.

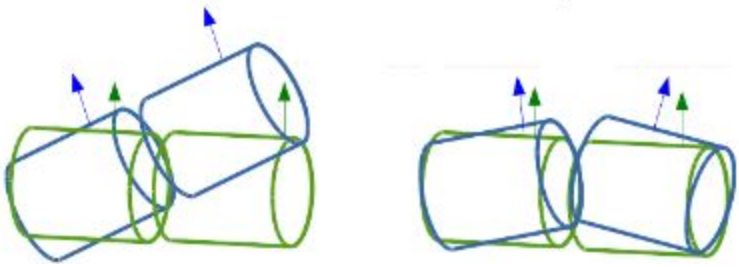
Nonmuscle myosin IIA

Nonmuscle myosin IIA (NMIIA) is an eukaryotic a motor protein. Assembly of NMIIA molecules into filaments is regulated by the Ca^{2+}-binding protein S100A4. The S100A4-NMIIA interaction is subject to

intensive investigations as mutations in S100A4 are implicated to have a role in tumorigenesis. S100A4 specifically binds the C-terminal disordered segment of NMIIA, adjacent to its coiled-coil forming region. Using NMR data determined by Andrea Bodor and coworkers, we have generated a number of structural ensembles of this fragment with different methodologies including ENSEMBLE (Krzeminski et al. 2013) and CoNSEnsX^{+}, and compared them to the bound conformers. Currently we are evaluating whether the enrichment of the initial conformer library in structures similar to the bound conformer influences the selected ensembles. This is necessary to ensure the relevance of the structural divergence of the structural ensemble. In addition, we performed preliminary docking studies to assess the compatibility of selected conformers with the binding site, this work is still in progress.

Galectin-1, a homodimeric carbohydrate-binding protein

Human Galectin-1 takes part, among other functions, in the induction of apoptosis of T cells and thus is an important modulator of immune response. Galectin-1 is expression has also been observed in several cancers. In order to investigate the structural background of ligand-binding and possible interdependence of the binding sites, we have generated dynamic structural ensembles (using experimental data kindly provided by Dr. Tamás Martinek, University of Szeged) of the protein. The simulation of this homodimeric system posed some challenges as there are multiple possibilities for the handling of S^2 restraining depending on the possible reorientations of the two subunits relative to each other, even if the dimer tumbles as a single unit in solution. Therefore, we have tested a number of different restraining schemes with local (i.e. subunit-wise) and global treatment of S^2 restraints. Initial evaluation of relative domain orientations revealed subtle changes, and the more detailed analysis with a python script developed specifically for this purpose is in progress.



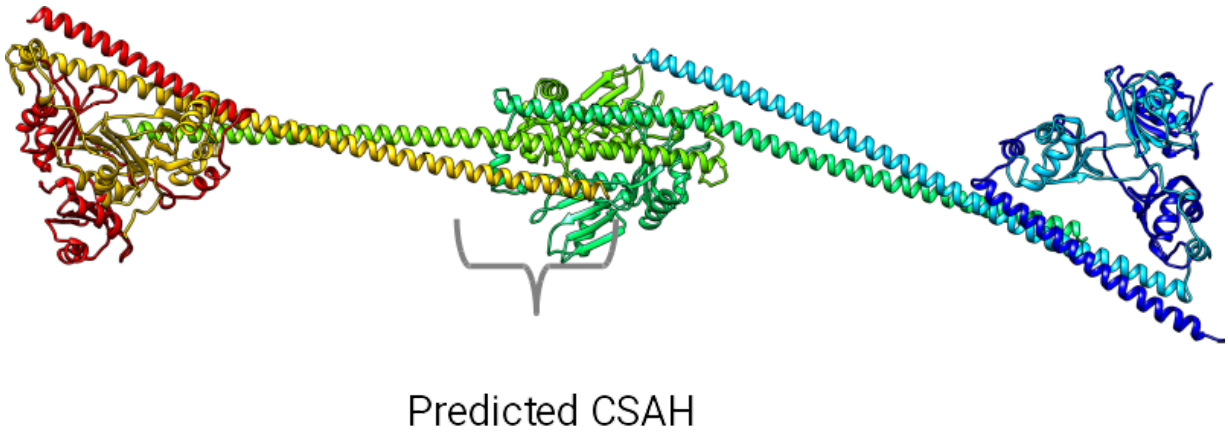
Different treatments of the subunits in a homodimeric protein. With respect to S^2 order parameter restraining. The two subunits can be treated as an entirely rigid unit, where their relative orientation is fixed, or, as tightly bound entities where specific kinds of relative subunit movements are nevertheless possible along the dimer interface. Choice of treatment might affect S^2 parameters back-calculated from the simulated conformations.

The initial results of the work have been presented at the National Student"s Research Conference (OTDK) in Debrecen, 2017.

Modeling of protein multimers in paraspeckles

Using our method to detect charged single alpha-helices, we have identified CSAHs in a number of RNA-binding proteins including a protein family associated with the formation of paraspeckles, a type of

relatively recently described subnuclear particles. We have investigated a number of protein sequences with domain composition similar to known paraspeckle proteins and concluded that the presence of a CSAH segment is characteristic of a subgroup of these proteins with a specific arrangement of a long coiled coiled coil and two RRM domains. We have developed a coiled coil modeling pipeline similar in concept to CCBuilder (Wood et al. 2014), a web service published while our work was still in progress (but after our pipeline was ready). Our pipeline is capable of modeling heterodimeric coiled coils with the architecture observed in truncated dimers of the PSPC1-NONO paraspeckle proteins and also hypothesized to be relevant in their higher-order organization. We have modeled the coiled coil regions of NONO and PSPC1 beyond those observed in crystal structures of the core PSPC1:NONO dimers (Passon et al., 2012) and proposed an atomic-level architectural model for the higher-order complexes. In these models, the predicted CSAH segments are a direct continuation of the coiled coil regions and their position suggests that they could act as a steric ruler, guiding protein assembly on the lncRNA molecules acting as the scaffolds of paraspeckles. The presence of a long single alpha-helical segment in paraspeckle proteins is compatible with crystallographic observations published just before our manuscript has been submitted (Lee et al. 2015) Our results are published in [VI].



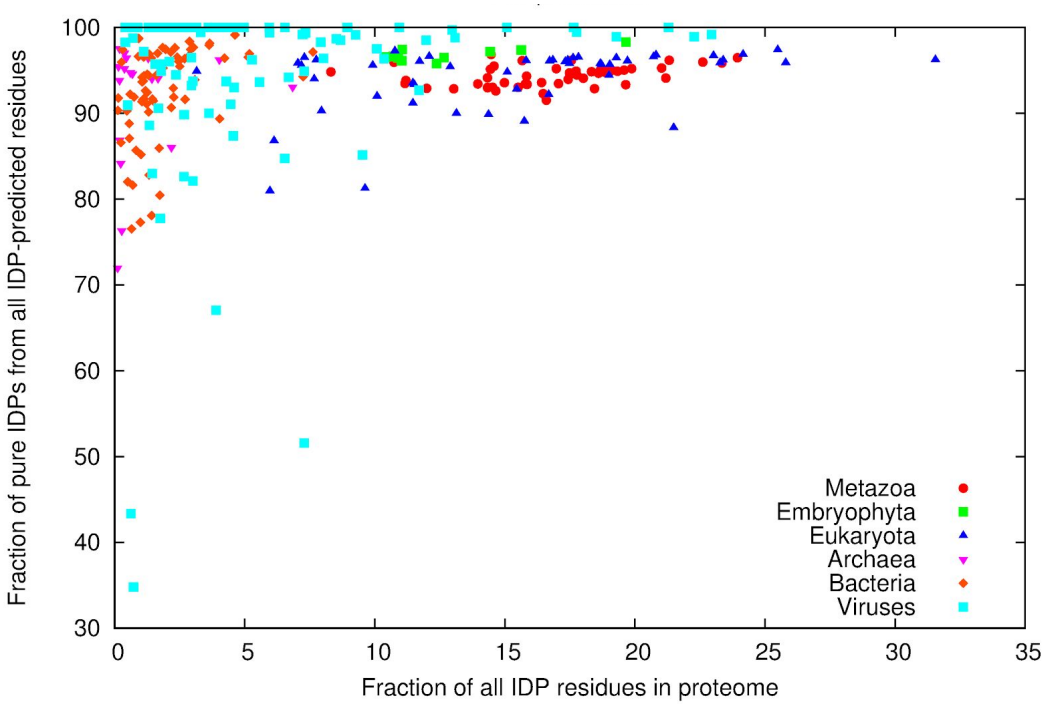
Model of the (PSPC1:NONO)3 hexamer, a putative part of the large paraspeckle complex. Chains A (red), C (green), and E (cyan) correspond to PSPC1, whereas chains B (yellow), D (blue–green), and F (blue) to NONO. The position of the CSAH segment of chain B is indicated. The models do not contain the genuinely disordered N- and C-terminal regions of the molecules and only contain the 'core' region for chains A and F.

Applications on proteomes

Prediction of intrinsic disorder, coiled coils, and collagen segments in proteomes

One of our observations during CSAH detection and analysis was the that several protein segments could be predicted both to be intrinsically disordered and to form coiled coils. Although this has been mentioned sporadically in the literature, we conducted a systematic analysis of overlapping predictions in 2010 and concluded that they are widespread and depend on the actual prediction tools used. However, it was still not clear how this phenomenon affects proteome-wide analyses. Therefore, we performed a comprehensive analysis on reference proteomes available in 2013. As an additional aspect, we included

collagen predictions also, as the high glycine and proline content of such segments were suggested to cause their recognition by disorder prediction methods. Our analysis revealed that there can be large differences between proteomes with similar overall predicted disorder content in the sense that what fraction of this content is actually attributable to coiled coil and/or collagen-like segments. The variance in this respect varies with taxonomic groups and also with proome size, the most variable ones being viral and bacterial proteomes. Thus, simply using a prediction method to detect protein intrinsic disorder without more detailed analysis of the nature of the predicted regions can yield to misleading results in computational proteomics. Detailed description of the results can be found in [II].



Fraction of residues of 'genuine' disorder relative of all predicted to be disordered, plotted as a function of the fraction of all disorder-predicted residues in different proteomes. Each point represents a proteome and proteomes of different taxonomic groups are colored differently. 'Genuine' or 'pure' disordered residues are those that are predicted to be disordered but not to form coiled coil or collagen-like structures.

Early genetic codes and protein disorder

We have conducted a detailed study on proposed earlier stages of genetic codes to assess the basic structural properties of proteins that might have been encoded by them. We have used a number of prediction algorithms for protein disorder, aggregation and for the presence of transmembrane segments. We have focused on the consensus of these methods and it should be stressed that we have only used the results in a comparative context, i.e. we did not assume that any of the results would necessarily be relevant for any particular sequence, rather the trends between predictions were evaluated. We concluded that prediction methods, optimized for proteins built from all 20 presently encoded amino acids, can yield profoundly different results. We have also concluded that the simplest proposed early codes are unlikely

to encode for structured proteins, at least when considering typical present-day intracellular conditions. Thus, either conditions were different if these codes existed, or the evolution of the code was more complex than these models suggest. Our results are published in [III].

Establishment of a protein expression laboratory

During the course of the project we have initiated the establishment of a protein expression laboratory at PPCU FIT, with the aim to be able to produce samples for experimental structural and dynamic investigations. This is considered a major step to overcome the limitations of using only published parameters for systems that independent groups chose to investigate. Establishment of the laboratory makes close collaboration with experimentalists a real possibility. We have already produced several recombinant proteins in quantities sufficient for structural investigations and will begin NMR measurements at collaborating laboratories in a few months. Then, the techniques developed with the aid of this grant will be used to interpret the measured structural and dynamical parameters. The fact that laboratory work is carried out at PPCU also means that these results, in turn, can be used to design variants suitable for experimental testing hypotheses about the role of internal dynamics of the proteins investigated.

References

Best, R.B., Vendruscolo, M. (2004) Determination of protein structures consistent with NMR order parameters. *J. Am. Chem. Soc. 2004 126: 8090-8091.*

Clore, G.M., Schwieters, C.D. (2006) Concordance of residual dipolar couplings, backbone order parameters and crystallographic B-factors for a small α/β protein: a unified picture of high probability, fast atomic motions in proteins. *J. Mol. Biol. (2006) 355, 879–886.*

Gáspári, Z., Süveges, D., Perczel, A., Nyitray, L., & Tóth, G (2012) Charged single alpha-helices in proteomes revealed by a consensus prediction approach. *Biochem. Biophys. Acta - Proteins and Proteomics 1824:637-646.*

Hamelberg, D., Mongan, J., McCammon, J.A. (2004) Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. *J. Chem. Phys. 120:11919-11929.*

Horváth, G., Király, P., Tárkányi, G. & Toke, O. (2012) Internal motions and exchange processes in human ileal bile acid binding protein as studied by backbone (15)N nuclear magnetic resonance spectroscopy. *Biochemistry 51*: 1848-1861.

Horváth, G., Egyed, O. & Toke, O, (2014) Temperature dependence of backbone dynamics in human ileal bile acid-binding protein: implications for the mechanism of ligand binding. *Biochemistry 53:5186-5198.*

Krzeminski, M., Marsh, J.A., Neale, C., Choy, W.Y., & Forman-Kay, J.D. (2013) Characterization of disordered proteins with ENSEMBLE. *Bioinformatics 29: 398-399*.

Lee, M. et al. (2015) The structure of human SFPQ reveals a coiled-coil mediated polymer essential for functional aggregation in gene regulation. *Nucleic Acids Res 43: 3826–3840.*

Morris, A.L, MacArthur, M.W., Hutchinson, E.G., & Thornton, J.M. (1992) Stereochemical quality of protein structure coordinates. *Proteins 13:345-364.*

Passon, D.M. et al. (2012) Structure of the heterodimer of human NONO and paraspeckle protein component 1 and analysis of its role in subnuclear body formation. *Proc. Natl. Acad. Scii. USA 109:4846-4850.*

Richter, B., Gsponer, J., Várnai, P., Salvatella, X., Vendruscolo, M. (2007) The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. *J. Biomol. NMR. 37: 117-135*

Smathers, R.L., & Rebecca L. (2011) The human fatty acid-binding protein family: evolutionary divergences and functions. *Human Genomics, 5:170-191.*

Süveges, D., Gáspári, Z., Tóth, G., & Nyitray, L. (2009) Charged single α-helix: a versatile protein structural motif *Proteins 74:905-916*.

Ting, D. et al. (2010) Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. *PLoS Comput Biol 6(4): e1000763*

Tochtrop, G. P., Richter, K., Tang, C., Toner, J. J., Covey, D. F. & Cistola, D. P. (2002) Energetics by NMR: Site-specific binding in a positively cooperative system. *Proc. Natl. Acad. Sci. U.S.A. 99, 1847−1852.*

Wang, X., Mahoney, B.J., Zhang, M., Zintsmaster, J.S., & Peng, J.W. (2015) Negative regulation of peptidyl-prolyl isomerase activity by interdomain contact in human Pin1. *Structure 23, 2224-2233.*

Wang, J., Kawasaki, R., Uewaki, J. ichi, Rashid, A. U. R., Tochio, N., & Tate, S. ichi. (2017). Dynamic allostery modulates catalytic activity by modifying the hydrogen bonding network in the catalytic site of human Pin1. *Molecules*, *22*(6).

Wood, C.L. et al. (2014) CCBuilder: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies. *Bioinformatics.30:3029-3035.*

Publications describing the results the project so far

Peer-reviewed papers & chapters (including reviews)

I Annamária F. Ángyán, Zoltán Gáspári: Ensemble-based interpretations of NMR structural data to describe protein internal dynamics. *Molecules (2013) 18:10548-10567.*

II Zoltán Gáspári: Is five percent too small? Analysis of the overlaps between disorder, coiled coil and collagen predictions in complete proteomes. *Proteomes (2014) 2:72-83*

III Annamária F. Ángyán, Csaba Ortutay, Zoltán Gáspári: Are proposed early genetic codes capable of encoding viable proteins? *J. Mol. Evol. (2014) 78:263-274.*

IV Ádám Fizil, Zoltán Gáspári, Terézia Barna, Florentine Marx, Gyula Batta: "Invisible" conformers of an antifungal disulfide protein revealed by constrained cold & heat unfolding, CEST-NMR experiments and molecular dynamics calculations. *Chem. Eur. J. (2015) 21:5136-5144.*

V László Gráf, Tamás Molnár, József Kardos, Zoltán Gáspári, Gergely Katona: The role of structural flexibility and stability in the interaction of serine proteases with their inhibitors *Curr. Prot. Pept. Sci. (2015) 6:521-531*

VI László Dobson, László Nyitray, Zoltán Gáspári: A conserved charged single alpha-helix with a putative steric role in paraspeckle formation. *RNA (2015) 21:2023-2029.*

VII Dániel Dudola, Gábor Tóth, László Nyitray, Zoltán Gáspári: Consensus prediction of charged single alpha-helices with CSAHserver Zhou, Kloczkowski, Faraggi, Yang (eds): Prediction of Protein Secondary Structure. *Methods Mol Biol Vol. 1847, Springer, 2017, pp. 25-34.*

VIII Christoph Sonderegger, Ádám Fizil, Laura Burtscher, Dorottya Hajdu, Alberto Muńoz, Zoltán Gáspári, Nick D. Read, Gyula Batta, Florentine Marx: D19S mutation of the cationic, cysteine-rich protein PAF: novel insights into its structural dynamics, thermal unfolding and antifungal function. *PLoS ONE (2017) 12:e0169920*

IX András Czajlik, Bertalan Kovács, Perttu Permi, Zoltán Gáspári: Fine-tuning the extent and dynamics of binding cleft opening as a potential general regulatory mechanism in parvulin-type peptidyl prolyl isomerases. *Sci. Rep. (2017) 7:44504*

X Dániel Dudola, Bertalan Kovács, Zoltán Gáspári: The CoNSEnsX^{+} web server for the analysis of protein structural ensembles reflecting experimentally determined internal dynamics. *J. Chem. Inf. Model. (2017) 57:1728-1734.*

XI Árpád Goretity, Zoltán Nagy, Zoltán Gáspári: Acceleration of a protein structure comparison algorithm on FPGA. *(Proceedings of the) 2017 European Conference on Circuit Theory and Design (ECCTD), in press.*

XII. Anna Huber, Dorottya Hajdu, Doris Bratschun-Khan, Zoltán Gáspári, Mihayl Varbanov, Stéphanie Philippot, Ádám Fizil, András Czajlik, Zoltán Kele, Christoph Sonderegger, László Galgóczy, Andrea Bodor, Florentine Marx, Gyula Batta: New antimicrobial potential and structure properties of PAFB: a cationic, cysteine-rich protein from *Penicillium chrysogenum* Q176. *Submitted for publication*

XIII. Sándor Boros, Zoltán Gáspári, Gyula Batta: Accurate NMR determinations of proton-proton distances. *Annu Rep NMR, invited chapter, to be submitted by November 2017*

Conference lectures & posters

Nóra Epresi, András Czajlik, Annamária F. Ángyán, Zoltán Gáspári: Protein structural ensembles reflecting experimentally determined dynamics: applications and developments (oral lecture), *5th European Conference: Chemistry For Life Sciences, 10-12 June 2013, Barcelona, Spain*

Zoltán Gáspári, Nóra Epresi: Dynamic conformational ensembles: challenges of multiple time scales and multidomain proteins (poster) *Gordon Research Conference Computational Aspects: Biomolecular NMR, June 2-7, 2013 2013, Mount Snow Resort, West Dover, VT, USA*

Zoltán Gáspári, Nóra Epresi, András Czajlik, Annamária F. Ángyán: Dynamic protein structural ensembles based on NMR data: developments and applications (oral lecture), *KeMoMo–QSAR symposium, April 29-30 2013, Szeged, Hungary*

Nóra Epresi, András Czajlik, Annamária F. Ángyán, Zoltán Gáspári: Tweaking GROMACS to generate NMR-based dynamic structural ensembles of proteins (oral lecture), *annual meeting of the NMR workgroup of the Hungarian Academy of Sciences, 2013, Pécs, Hungary*

Zoltán Gáspári, Ádám Fizil, Gyula Batta: Using measurements and calculations to reveal the secret life of an antifungal protein (oral lecture), *KeMoMo–QSAR 2014 Symposium, May 22-23, Szeged, Hungary*

László Dobson, András István Seres, László Nyitray, Zoltán Gáspári: When disorder is deception: from predictions to structural modeling (oral lecture), *Annual meeting of the peptide chemistry workgroup of the Hungarian Academy of Sciences, May 28-30 2014, Balatonszemes, Hungary*

András Czajlik, Comparative analysis of dynamic structural ensembles of peptidyl-prolyl cis-trans isomerases (oral lecture), *Annual meeting of the peptide chemistry workgroup of the Hungarian Academy of Sciences, May 28-30 2014, Balatonszemes, Hungary*

Zoltán Gáspári: Protein conformational ensembles generated with molecular dynamics simulations incorporating NMR-based parameters (oral lecture), *Meeting of the NMR workgroup of the Hungarian Academy of Sciences in honor of prof. Attila Szabó, June 18 2014, Budapest, Hungary*

László Dobson, László Nyitray, Zoltán Gáspári: The charged single alpha-helix in paraspeckle-forming proteins (poster) *Annual meeting of the Hungarian Biochemical Society, August 24-27 2014, Debrecen, Hungary*

Zoltán Gáspári, Annamária F. Ángyán, András Czajlik, Anett Korai: Visualizing and understanding experimentally determined protein dynamics at the atomic level: methods and applications (poster) *Annual meeting of the Hungarian Biochemical Society, August 24-27, 2014 Debrecen, Hungary*

Bertalan Kovács, Zoltán Gáspári: Dynamic structural ensembles of the tandem PDZ1-PDZ2 domains of PSD-95. (poster) *Hungarian Molecular Life Sciences, March 27-29, 2015, Eger, Hungary*

Bertalan Kovács, Anett Hinsenkamp, Zoltán Gáspári: Dynamic structural ensembles of PDZ domains of PSD-95. (poster) *Gordon Research Conferences, Computational Aspects - Biomolecular NMR, June 7-12, 2015, Il Ciocco, Lucca (Barga), Italy*

Gyula Batta. Ádám Fizil, Zoltán Gáspári, Terézia Barna, Florentine Marx: “Invisible” conformers of the antifungal disulfide protein PAF. (poster) *Gordon Research Conference Computational Aspects: Biomolecular NMR, June 7-12, 2015, Il Ciocco, Lucca (Barga), Italy*

Zoltán Gáspári, Dániel Dudola: Can we capture multi-timescale dynamics in a single simulation? (poster). *Gordon Research Conference Computational Aspects: Biomolecular NMR, June 7-12 2015, Il Ciocco, Lucca (Barga), Italy*

Dániel Dudola, Zoltán Gáspári, How well can we model protein internal dynamics with ensemble-based methods? (oral lecture), *Annual meeting of the peptide chemistry workgroup of the Hungarian Academy of Sciences, April 24-25 2015, Balatonszemes, Hungary*

Bertalan Kovács, Zoltán Gáspári:Dynamic structural ensembles of the tandem PDZ1-PDZ2 domains of PSD-95. (oral lecture), *Annual meeting of the peptide chemistry workgroup of the Hungarian Academy of Sciences, April 24-25 2015, Balatonszemes, Hungary*

László Dobson, László Nyitray, Zoltán Gáspári: Charged single alpha-helices: prediction, analysis and modeling (oral lecture) *Hungarian Molecular Life Sciences Conference, March 27-29 2015, Eger, Hungary.*

Dániel Dudola, Zoltán Gáspári: Assessing the compliance of dynamic protein structural ensembles with experimental NMR data (poster) *Hungarian Molecular Life Sciences Conference, March 27-29 2015, Eger, Hungary.*

Zoltán Gáspári: Ensemble-based representation of protein dynamics using NMR-derived parameters (invited lecture). *Finnish NMR symposium, June 12-15 2016, Peurunka, Finland*

Zoltán Gáspári: Protein structural ensembles reflecting internal dynamics: generation, evaluation and biomedical relevance (distinguished lecture). *BJMT Applied Mathematics Conference, June 1-3 2016, Győr, Hungary*

Zoltán Gáspári: Is the NOE satisfied and/or stereospecific? (oral lecture) *Annual meeting of the NMR workgroup of the Hungarian Academy of Sciences, May 20-21 2016, Debrecen, Hungary*

Zita Harmat, András Szabó, Zoltán Gáspári: Investigation of the free and ligand.bound forms of gastrotropin using molecular dynamics calculations with experimental restraints. (oral lecture) *Annual meeting of the NMR workgroup of the Hungarian Academy of Sciences, May 20-21 2016, Debrecen, Hungary*

András Czajlik, Bertalan Kovács, Zoltán Gáspári: Analysis of the internal dynamics of parvulin-type PP isomerases based on simulations combined with experimental data (oral lecture). *KeMoMo–QSAR symposium, May 12-13 2016, Miskolc, Hungary*

Zoltán Gáspári, Bertalan Kovács, Zita Harmat: Dynamic protein structural ensembles based on NMR data: methodology & applications (poster). *Annual meeting of the Hungarian Biochemical Society, August 28-31 2016, Szeged, Hungary*

Dániel Dudola, Zoltán Gáspári: Selection of dynamic ensembles to achieve maximum correspondence to experimental NMR data – an extension for the CoNSEnsX server (poster). *Annual meeting of the Hungarian Biochemical Society, August 28-31 2016, Szeged, Hungary*

Anett Hinsenkamp, Zoltán Gáspári: Dynamic structural ensembles of the third PDZ domain of PSD-95 (poster). *Annual meeting of the Hungarian Biochemical Society, August 28-31 2016, Szeged, Hungary*

Bertalan Kovács, Zoltán Gáspári: Internal motion of the PDZ1-PDZ2 tandem of PSD-95 represented by dynamic structural ensembles. (poster) *Annual Conference of the Hungarian Biochemical Society 2016, Szeged, Hungary*

Zoltán Gáspári: Protein structural ensembles reflecting experimental parameters (oral lecture). *Meeting of the Foundation for the Hungarian Peptide and Protein Research on the occasion of the 20th anniversary of its establishment, November 14 2016, Budapest, Hungary*

Bertalan Kovács, András Czajlik, Perttu Permi, Zoltán Gáspári: Dynamic structural ensembles of parvulins based on experimental NOEs and order parameters. (poster) *Magnetic Moments in Central Europe, March 8-12 2017, Budapest, Hungary*

Zita Harmat, Abdrás L. Szabó, Orsolya Tőke, Zoltán Gáspári: Investigating the internal dynamics of a lipid-binding protein with molecular dynamics simulations using NMR data as restraints (poster). *Magnetic Moments in Central Europe, March 8-12 2017, Budapest, Hungary*

Bertalan Kovács, András Czajlik, Perttu Permi, Zoltán Gáspári: Internal motions of parvulin-type peptidyl-prolyl cis-trans isomerases. (oral lecture) *Hungarian Molecular Life Sciences, March 31- April 2 2017, Eger, Hungary*

Bertalan Kovács, András Czajlik, Perttu Permi, Zoltán Gáspári: Analysis of the internal motions of parvulin-type peptidyl-prolyl isomerases by generation of dynamic structural ensembles. (poster) *Gordon Research Conferences, Computational Aspects - Biomolecular NMR, June 11-16 2017, Newry, MN, USA*

Bertalan Kovács, András Czajlik, Perttu Permi, Zoltán Gáspári: Generation and assessment of dynamic structural ensembles of parvulin-type peptidyl-prolyl isomerases by externally restrained molecular dynamics simulation. (poster) *Conformational Ensembles from Experimental Data and Computer Simulations, August 25-29 2017, Berlin, Germany*