Final report for OTKA grant no. F68079

In the project, we have achieved progress in diverse but related areas of protein dynamics and its investigation by NMR. A significant portion of our work has focused on the generation and analysis of dynamic structural ensembles, i.e. a set of conformers whose diversity reflects the internal dynamics of the target protein at a specific time scale. To this end, we have used our own implementation of the MUMO (Minimal Under-restraining Minimal Over-restraining) approach (originally descibed by Richter et al. in 2007) in the free molecular dynamics software GROMACS. This implementation is first described in our paper on PAF (Penicillium Antifungal Protein) (**Batta et al. 2009 FEBS J**). In addition, a Monte-Carlo based approach for generating highly diverse intrinsically unstructured segments is under development. We have also provided computational tools for the analysis of such ensembles (e.g. CoNSEnsX, **Ángyán et al 2010 BMC Struct Biol**). Using dynamic structural ensembles as well as other NMR-derived information we have also investigated protein:protein interactions involving canonical serine protease inhibitors and proteins of the complement system. Besides proposing an universal selective isotope labeling system for protein NMR spectroscopy, we have participated in the identification and characterization of charged single α-helices (CSAHs), a recently identified and possibly dynamic protein structural motif. In addition, several aspects of protein evolution in connection with protein internal dynamics were also investigated both experimentally and theoretically, some of these is still in progress.
More specifically, our most important results are as follows:

<u>Generation and analysis of protein structural ensembles</u>

- Implementation of the MUMO approach into GROMACS (version 3.3.1). Our extension can be parametrized like 'standard' GROMACS features and is freely available (**Batta et al, 2009 FEBS J**)
- Generation of dynamical structural ensembles of the antifungal protein PAF with two possible disulfide pairing and without dusulides. As all ensembles correspond equally well to experimental data used for cross-validation, this did not allow us to identify disulfide pairing (similarly to the other known meber of the respective protein family) (**Batta et al. 2009 FEBS J**)
- We have extended our PAF structure calculations to include RDC data. Evaluation of the resulting ensembles is in progress.
- Generation of dynamical structural ensembles of the small serine protease inhibitors SGCI and SGTI. We have shown that although these ensembles are more diverse than those

obtained by 'conventional' structure calculations, they correspond to experimental data better and have a more restricted structure-stabilizing interaction between key side chains (**Gáspári et al. 2010 FEBS Lett**)

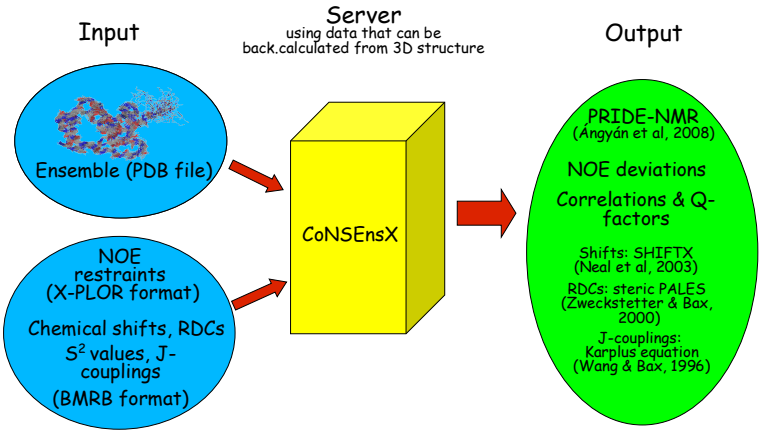
- We have started the development of an ensemble Monte Carlo-based method to generate diverse structural ensembles of intrinsiclly unstructured protein segments that correspond to RDC data. We are currently testing the present implementation on the disordered N- and C-terminal tail regions of *Drosophila* dUTPase.

- We have caculated a structural ensemble of SGTI-PO2, a variant that is a potent inhibitor of bovine trypsin in contrast with wild-type SGTI. The structure revealed an unexpected structural rearrangement the relevance of which is still under investigation.

Development of tools for the analysis of protein structures in the light of NMR-derived experimental data

- We have developed PRIDE-NMR, a method capable of finding protein structures from a database that correspond to a set of NOE data given as input. Our results sindicate that NOE-derived distances, although scarce, capture the essence of a fold well and can be successfully used to quickly relate structures to experimental data. (**Ángyán et al. 2008 Bioinformatics**)

- Using PRIDE-NMR, thid-party software (PALES & SHIFTX) and in-house calculation methods we have set up the CoNSEnsX web server, specifically designed for the analysis of dynamically restrained protein ensembles. The server reports the correspondence of the ensemble as a whole to experimental data (**Ángyán et al. 2010 BMC Struct Biol**; web server available at: *http://consensx.chem.elte.hu*)

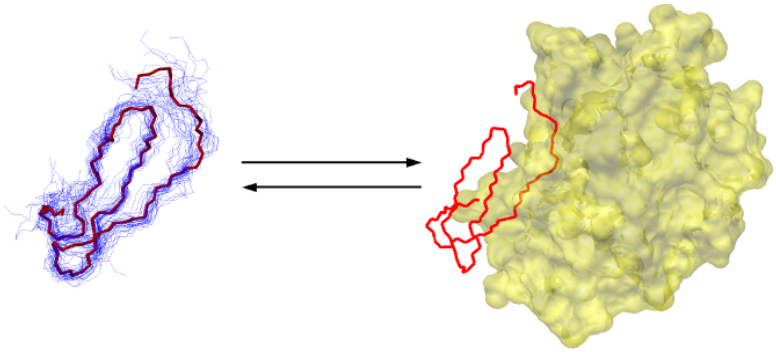


Scheme of the CoNSEnsX server for the evaluation of dynamic conformational ensembles

- We have shown that the diversity of dynamic structural ensembles heavily influences bioinformatic analyses performed on the individual conformers (**Gáspári et al. 2010 Curr Prot Pep Sci**)

Investigating protein:protein interactions

- Using the dynamic conformational ensemble of SGCI and SGTI we have shown that the enzyme-bound state is accessible in solution for these inhibitors. This finding reconcilies the widely held rigid lock-and-key model for canonical serine protease inhibitor action with NMR-based evidence for the remarkable flexibility of their protease binding loop (**Gáspári et al. 2010 FEBS Lett**).



Scheme of the interation of SGCI with chymotrypsin (yellow) via *conformer selection. The conformation in the bound form (red) is sampled during the ps-ns time-scale internal dynamics of SGCI in solution (blue ensemble)*

- Performing data-driven docking with NMR chemical shift perturbation information revealed a novel type of interacion between the consecutive complement control (CCP) modules of the human C1r protein, a component of the complement system. The interaction identified in solution differs slightly from that seen in the crystal structure and interconversion of the two states is plausible (**Láng et al. 2010 FEBS Lett**).

- We have excessively reviewed methods for assessing protein internal dynamics by NMR spectroscopy and propsed the extension of structure-activity relationships to account for the dynamics of the partners (dynamic SAR, DSAR) (**Gáspári & Perczel, 2010, Annu Rep NMR**)

- We have proposed a novel selective isotope labeling system for protein NMR. The method is based on a possible extension of the genetic code to code for the labeled and unlabeled form of each of the 20 amino acids. We have enumerated the foreseeable difficulties of the development of such a system and have proposed solutions for them. We stress that even a syste capable of incorporating 3-5 types of amino acids both in labeled and unlabeled form

would be highly useful (**Gáspári et al. 2008 BioEssays**; *Gáspári & Pál, 2010, Élet és Tudomány*)

- We have written a popular paper on the role of protein internal dynamics in biomolecular interactions (*Gáspári & Perczel, 2011*, *Természet Világa*)

Internal dynamics and protein evolution

- We have developed computational methods to identify charged single α-helices (CSAHs) in proteins and have shown that CSAHs occur in diverse proteins and organisms and might have highly versatile functions. CSAHs most likely represent dynamic structural motif that can adopt features remiscent to both coiled coils and unstructured segments (**Süveges et al. 2009 Proteins**)
- We have designed a consensus web server based on the methods described previously to detect CSAHs from proteun sequences. We have also performed a compehensive analysis of CSAH segment in diverse proteomes. Detailed evaluation of the results is in progress (the web server is functional and available at: *http://csahserver.chem.elte.hu*)
- We have investigated the occurrence of cross-predictions between coiled coils and intrinsically disordered segments by a number of algorithms available for their detection. We found that coiled coils are often recognized as unstructured regions, while the reverse is not true, and that many segments predicted to be both coiled coil and unstructures exhibit amino acid distribution resembling CSAHs (**Szappanos et al. 2010 FEBS Lett**)
- We have reviewed the development of the SBASE domain library capable of detecing atypical members of protein familes that have diverged substantially from 'representative' ones during evolution (**Dhir et al. 2010 Curr Pot Pept Sci**)
- We have designed and successfully produced double mutants of the Tc5b miniprotein that show characteistics reminiscent of intrinsically unstructured proteins. Our preliminary NMR and CD measurements on two of the mutants expressed in 15N-labeled form suggest that our design strategy is successful and appropriate to understand structural evolutionary tranistions between folded and unfolded proteins. Production of more variants and their characterization is under way.
- We have reviewed the potential use of coiled coil segments as models for evolutionary structural transition studies. Coiled coils are highly variable in sequence, structure, stability and dynamics and show versatile functions, making them ideal targets for modelling stepwise changes during evolutionary fold transitions (**Gáspári & Nyitray, submitted**)