Membrane mimetics and peptide interactions, effect of the lipid environment
Final report

In our project we aimed to characterize the membrane-protein systems from the side of each participant. To do so, we applied a synergy of methods, based mostly on NMR spectroscopy, SAXS and CD spectroscopy. As a result, we revealed important basic properties/relations regarding folded proteins and intrinsically disordered proteins (IDPs); we gave new insights into the treatment of bicelle shape characterizations; we developed novel NMR pulse sequences with increased resolution and sensitivity; we proved that the peptides we proposed to investigate, have remarkable cell penetration properties and we characterized their internalization, and localization. The main results can be summarized as follows:

Empirical D-M relations for proteins.
Protein conformation and compactness are global parameters that affect the biological function; thus, characterization of such properties is essential in the investigation of a biochemical system.
(1) The translational diffusion coefficient *D* is capable to discern between the different shapes/conformations. Based on this property, using a set of well-known representative globular proteins and IDPs we could formulate reliable log*D* – log*M* linear correlations (Dudás, 2019, Anal.Chem.). At 288K the relationships are:
$\log D\ (\text{globular}) = -0.381 \cdot \log M - 8.499$
$\log D(\text{IDP}) = -0.507 \cdot \log M - 8.169$
From the Stokes-Einstein equation for a spherical molecule the power of *M* is -0.33. Analysing our data the -0.381 value for folded proteins is in correlation with a spherical/distorted spherical shape; while the -0.507 value resulting for IDPs shows a resemblance with values obtained for synthetic polymers. Further on, we proved, that for typical biomolecular conditions the applicability of the Stokes – Einstein relation. The determined correlations can be used for various bioanalytical applications.
(2) We established hydrodynamic radii – number of aminoacid residues r_H – *N* empirical relations, and performed a comparison to other literature results, showing the benefit of using the same experimental conditions and method - in our case. On the other hand we believe that this interpretation is less appropriate, as deriving the hydrodynamic radius from the translational diffusion coefficient assumes a spherical geometry, and no shape factors are considered.
(3) The presented *D-M* equations were determined at 288K, and we showed that - in case the protein does not change the shape - by applying corresponding temperature and viscosity corrections the empirical relations are valid at other temperatures (Szabó, 2022, Anal.Chem.)
(4) In chaotropic solvents (urea, DMSO) proteins unfold, and using our relations we showed how can one follow protein unfolding via the translational diffusion measurements. We investigated the effects of unfolding and draw the attention on the influence of disulphide bridges (Figure 1). In DMSO the unfolding of lyzosime starts at 60% DMSO content and can be considered complete in 100% DMSO. In 8M urea, pH assisted unfolding of ubiquitin – with no disulphide bridges - is complete at pH < 5. Lysozyme - containing 3 disulphide bonds - shows partial denaturation at pH < 3, addition of TCEP at pH 2 causes the disruption of disuplhide bonds and brings a fast and complete denaturation. BPTI has 4 disulphide bonds and

even at pH 2 we see no unfolding. Addition of TCEP causes slow denaturation and a final unfolding in more than 1 day.

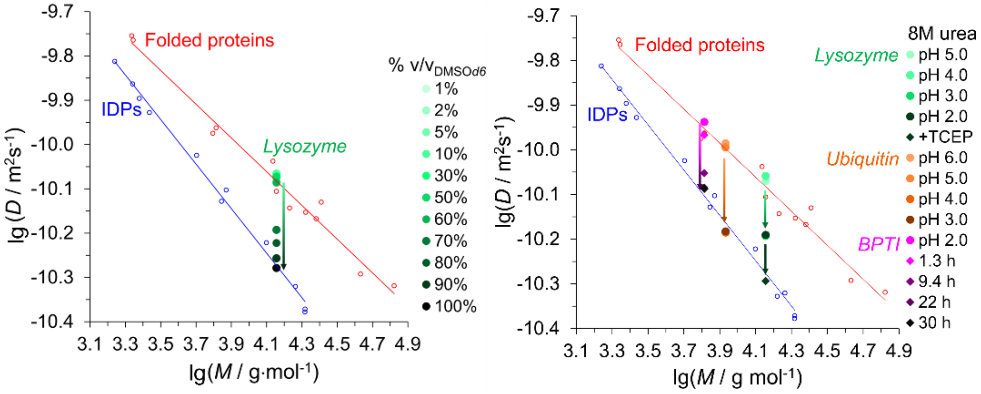

Figure 1. Monitoring protein unfolding for different proteins at 298K, in DMSO-d_6 (left) and 8M urea (right). The empirical lg *D*- lg *M* relations are the continuous red (folded proteins) and blue (IDPs) lines, while corresponding circles represent the studied proteins used for the establishment of these relations. The unfolding of lysozyme (green) in DMSO-d_6 is investigated for 1-100% DMSO-d_6 content. The unfolding of lysozyme (green), ubiquitin (orange) and BPTI (magenta) is followed in 8M urea at different pH values.

(5) Working with pharmacologically important H_2O-DMSO solvent mixtures we showed, that while the protein hydrodynamic properties are not changed, the atomic environment might alter. As an example, we showed that already small amounts of DMSO can alter the binding of NAG (N-acetyl-D-glucosamine) to lysozyme.
(6) We investigated the pitfalls of evaluating diffusion measurements regarding (*i*) the limitations of the usage of a small molecule as internal reference (generally dioxane) (*ii*) the importance of viscosity in molecular weight determination for biomolecules - used also for aggregation studies and association in case of membrane mimetics.
(7) Results from global characterization and IDP behavior, as well as secondary structural propensity assessment were summarized review articles (Bodor, 2021, "Amino acids, peptides, proteins" book series; Kovács, 2023, RSC.Adv.)

Bicelle membrane mimetics – peptide interactions.
Physico-chemical characterization of a micelle/bicelle membrane mimetic system is of great importance, as quite often the in vitro laboratory experimental model setup constitutes the basis for the more complex analysis. Protein-membrane interactions influence the global properties - shape, size, curvature of the studied system. As we can reliably handle the diffusion data, we aimed to combine NMR and SAXS measurements to characterize the shape variations for membrane-peptide complexes. In this respect we used peptides with different topology: either surface bound (melittin) or transmembrane (KALP23). Neutral DHPC/DMPC (PC) and negatively charged DHPC/DMPC/DMPG (PC:PG) bicelles were investigated in free form and loaded with model peptides (Dudás, 2020, BBA Biomembranes).
(1) Using the translational diffusion measurements, we determined the DHPC *cmc* values under different conditions and analyzed the effects of ionic strength and temperature on these values.
(2) We showed that using 1D ^{1}H spectra, analysis of the well separated -CH_3 resonances belonging to DHPC and DMPC lead to a proper analysis of the system composition: free DHPC, micelle, bicelle (Figure 2). Moreover, an exact q value - being the ratio between the short and

long chain lipid – can be calculated, which is an important parameter for fitting SAXS scattering curves.

(3) Using our empirical log$D$ - log$M$ relations we determined the aggregation numbers in the PC bicelles having different composition.

(4) The samples used for NMR measurements were used also for SAXS investigations, keeping the experimental conditions constant. The scattering curve was fitted using a model that features an ellipsoidal core with two semi-axes (a, b). These will loosely correspond to the hydrophobic interior of the bicelle, while the lipid headgroups are part of the outer shell with thickness ta and tb. We proved the applicability of core-shell lentil model for the bicelles and showed that for all our studied system an oblate shape is obtained.

(5) SAXS analysis of the bicelle - model peptides showed, that irrelevant of the topology the rim became narrower while the shape became more elongated. This can be accompanied by a migration of lipid molecules between the rim and bilayer regions. The increased oblateness should also mean increase in r_H, a result indicated by the NMR measurements, too.

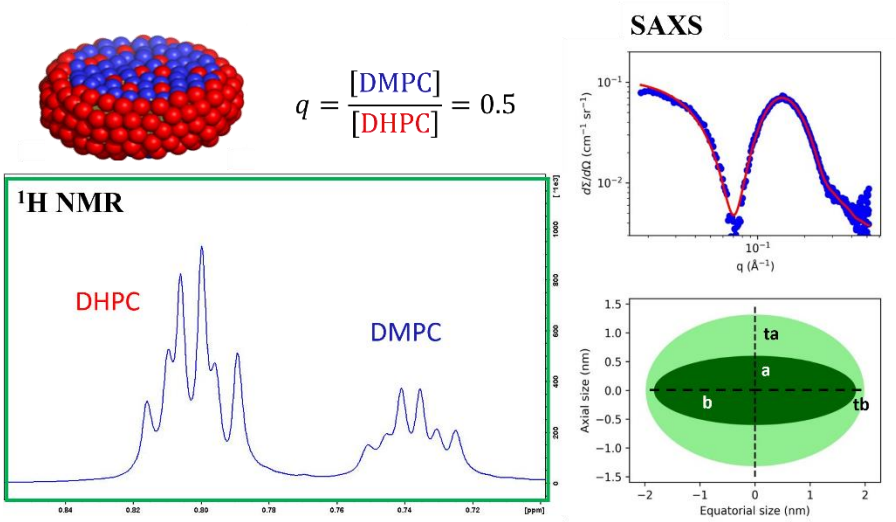

Figure 2. Schematic view of the studied bicelles and the used composition; the -CH_3 region of the ^{1}H NMR spectrum at 700 MHz; a typical SAXS scattering curve: blue experimental, red: fitted using the below presented ellipsoidal core-shell model with different electron densities, and the denominations of the semi-axises.

(5) In conclusion, according to our aims we succeeded to introduce a combined NMR-SAXS approach - used for the first time in case of bicelle systems - that provides a full description of solution composition, molecular size and shape determination. The presence of peptides alters several global parameters: the transmembrane KALP23 induces elongation and size increase, while the mostly surface-active melittin does not significantly perturb PC bicelles. Careful data interpretation is needed for the negatively charged PC/PG bicelles, as the net charge of the peptide influences as well. We also found that it is not advisable to directly compare the results originating from different peptides, instead comparison should be done with respect to the empty bicelle system. Our study can represent a good starting point for analyzing peptides with unknown behavior in their interaction with various bicelles.

(6) Along the new research we concluded that there is a need for a comprehensive physico-chemical investigation of the neat bicelles, having different composition, but keeping q=0.5. In a still ongoing study, we consider the experimentally measured viscosities, and set the limits of the Stokes-Einstein relation. Besides the NMR derived translational diffusion coefficients of the bicelle components we evaluate values obtained for the water and try to determine

obstruction factors. In the light of all these we intend to try to give an even more exact fitting of the SAXS scattering curve.

Novel NMR pulse sequences.

Peptide/bicelle systems are best to be studied at near-neutral pH. If the changes occurring on the peptide side need to be monitored, the commonly used $^{1}H_{N}$-detected experiments will fail due to the ^{1}H-exchange between the backbone -NH- and H_2O, that causes line broadening of resonances in the ^{1}H,^{15}N-HSQC spectra, often below detection limit. Another drawback of this approach is that proline residues do not give signal in these measurements. Therefore, a different way has to be chosen. We tried to exploit the possibilities of ^{1}Ha-detection, focusing on the -HaCa- backbone environment. This is not affected by pH, and all residues are detectable, drawback is, that the solvent water resonance falls in this region and must be taken care of.

(1) For isotopically labeled molecules an extensive homo- and heteronuclear decoupling is needed. In this respect we developed the so-called BASEREX selective inversion scheme - achieved by a combination of the BIRD^{d} element with the band-selective refocusing on the X-nucleus (Haller, 2019, JMR).

(2) We introduced the novel HaCa selective-HSQC 2D measurement (SHACA), including real-time homodecoupling via the BASEREX selective inversion scheme (Figure 3). Comparing the performance of this novel pulse-sequence to other "work horse" 2D measurements: ^{1}H,^{15}N-HSQC for $^{1}H_{N}$ detection and CON for ^{13}C detection we proved that the *sensitivity and resolution* of SHACA is superior, and the water peak does not represent an obstacle (Bodor, 2020, Anal.Chem.). For IDPs the gain in intensity - if no internal dynamics is involved – can be close or reach the theoretical value. Nevertheless, minor species of 2-3μM concentration can be detected in about 40mins, or faster, if non-uniform sampling is applied. An example for a SHACA-HSQC for a 1mM IDP sample (EZH2, 105aa) in 10% D_2O is shown on Figure 3. The water peak suppression is excellent, peaks are well-resolved, as seen for the different types of aminoacid residues (Szabó, 2022, IJMS).

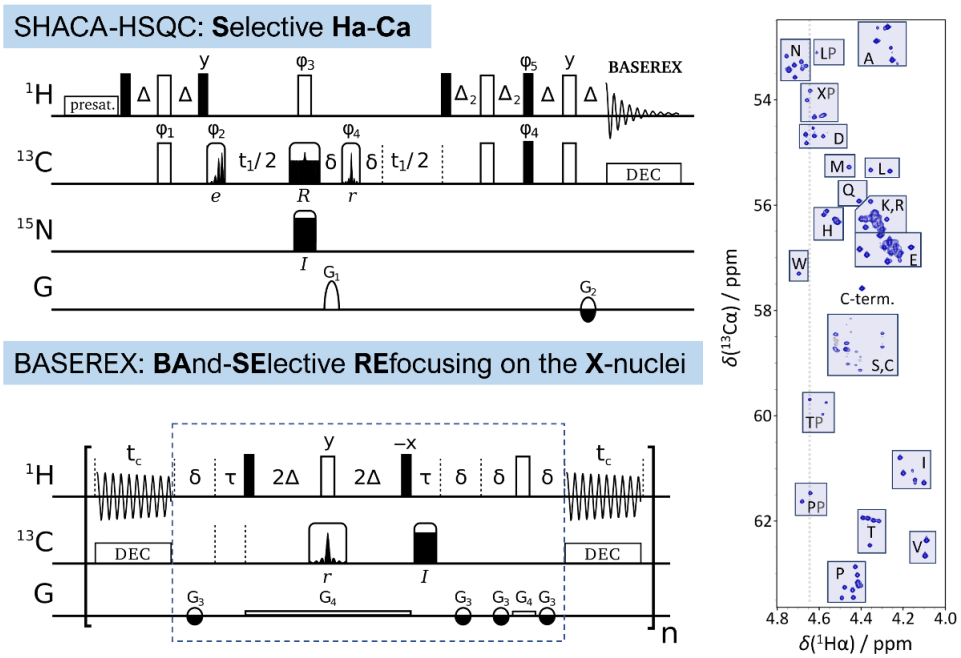

Figure 3. The newly developed SHACA-HSQC pulse sequence and the BASEREX acquisition scheme; the SHACA of 1.0mM EZH2, 105 residues long IDP at pH 7, 700 MHz, 10% D_2O. The vertical line shows the water peak resonance.

(3) The developed sequences can have numerous applications. They can be useful to monitor labeled compounds, to detect minor components in IDPs, to follow the effect of posttranslational modifications in proteins (Sebák, 2022, Angew.Chem.). At the same time, they represent useful building blocks for 3D measurements (Szabó, 2022, IJMS; Sebák, 2023, Biomol.NMR.Assign.)
(4) As prolines are not observable via $^{1}H_{N}$-detected measurements, we introduced two novel ^{1}Ha-detected proline selective pulse sequences: Pro-HCANCACBCAHA, and Pro-HCGCBCAHA. These experiments allow assignment of proline Cb and Cg chemical shifts, that bear important information regarding the proline isomerization. The sensitivity and the resolution of these experiments allowed us to highlight and fully characterize 13 minor proline species in the disordered transactivation domain of p53(1-60), and to give an atomic level proof of how phosphorylation can shift the proline *cis-trans* equilibrium – a deed relevant in biological investigations (Sebák, 2022, Angew.Chem.).
(5) We introduced a real-time pure shift acquisition scheme for the detection of amide protons which is based on ^{13}C-BIRD^{r,X}. The scheme was incorporated in FHSQC and BEST-TROSY experiments, giving exceptional results both for folded and disordered proteins. For p53(1-60) the line width of 12-13Hz could be reduced to 4.0Hz (Haller, 2022, J.Biomol.NMR).

Structural characterization and cell penetration abilities of lysine rich peptides.
We proposed to investigate the bicelle interaction of selected biologically relevant protein fragments: K-segments of the plant dehidrin protein ERD14 (ERDA, ERDB, ERDC) and the C-terminal part of the vertebrate specific, metastasis associated S100A4 (S100).
(1) We performed an in-depth NMR investigation, giving the full assignment of these peptides using 2D homo- and heteronuclear correlation experiments, using peptides having the natural isotope abundance. In free form all systems were highly mobile, without any secondary structural preference.
(2) We decided to deviate from the initial plan, and test the cell-penetrating properties of these peptides. All of them are easily water soluble and have satisfactory logP values – therefore they might be able to carry hydrophobic drug candidates. For intracellular localization studies we used carboxyfluorescein (Cf) as an optimal peptide conjugate, which has (5)- and (6)-substituted forms. Using an extensive NMR spectroscopic investigation, we showed that the Cf conjugation/modification can influence the structural features of the attached peptide that might alter the cellular uptake and, consequently, the drawn conclusions (Sebák, 2021, ACS Omega).
(3) We found that the small structural differences resulting from the NMR investigations can/might influence the cellular uptake. The determined order of the uptake tendency is: ERD-B> ERD-A>ERDC>S100. The uptake mechanism for all peptides mainly occurs through energy-dependent paths and can be inhibited chemically.
(4) All four peptides are taken up by the A431 cells without any in vitro cytotoxic effects. Based on CLSM studies, the Cf peptides have different intracellular localization. Thus, the ERDA is localized mainly in the cytoplasm, ERDB has partial colocalization with lysosomal staining, and ERDC is mainly localized in the lysosomes. The S100 peptide also has partial colocalization with lysosomal staining (Figure 4).
(5) We investigated the effect of a proline residue in the peptide amino acid sequence. NMR characterization showed that depending on the type of the neighboring amino acid side chain, varying amounts of *cis* proline minor forms are present. It is still to be tested, whether this has any influence on the cellular uptake and localization (Sebák, 2023, FBL).

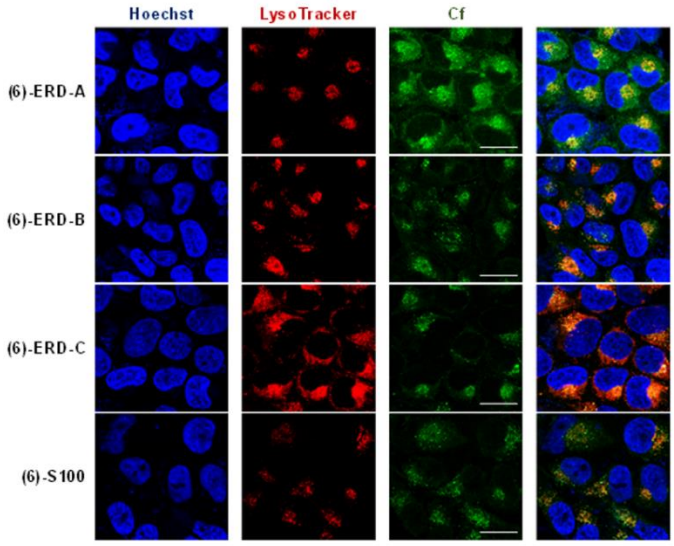

Figure 4. In vitro intracellular localization of Cf-peptides visualized by CLSM. Cells were treated for 90 min with Cf-peptides (25 μM, green), lysosomes (red), nuclei (blue).

(6) Our experimental proof regarding the cell penetration abilities of the studied peptides opened new routes regarding possible applications of these systems. In the framework of new collaborations, we test the antimicrobial properties as well as delivery properties with selected drug candidates.

Concluding our achievements, we believe, that besides the system-specific results, the strength of our work is the general applicability regarding several aspects. Thus, (*i*) we formulated empirical D-M relations that can be used for diverse bioanalytical purposes: differentiation between folded or unfolded protein; monitoring protein unfolding; following aggregation; determining the aggregation numbers for various bicelle systems; (*ii*) we developed novel NMR-SAXS based approach for characterizing bicelle systems. Using this combined method the exact solution composition, and the shape of the system can be determined; the influence of peptides/proteins interacting with the bicelle can be described; (*iii*) we introduced novel ^{1}Ha-detected NMR pulse sequences with increased resolution and sensitivity, applicable under physiological conditions. The SHACA-HSQC can be used to follow metabolic paths for labeled molecules, the effect of post-translational modifications; and serves as basis for novel 3D experiments. The proline selective 3D measurements are prone to determine the *cis/trans* isomeric form of the minor components in peptides/IDPs; (*iv*) we proved that our lizine rich peptides (the plant derived ERD segments and the human S100 C-terminus) have remarkable cell penetration properties and we characterized their internalization, and localization.

The results achieved are part of 4 PhD works. 18 papers were published in these topics, regarding direct results and follow-up applications of our developments, 3 more manuscripts are under review/waiting editorial decision. The results were presented and discussed at many national and international conferences. Amongst others I was invited speaker at ENC, Orlando, USA, 2022; at the 43rd FGMR NMR meeting, Karlsruhe, Germany 2022; I was an online speaker at the ISMAR, Osaka, 2021. We participated at the EUROMAR in Berlin, Utrecht, Glasgow; we presented the results at the Hungarian NMR meeting, Peptide meeting, Biologicum-analitics meetings.