

NMRlipids databank enables machine learning models for lipid bilayers

Dr. Samuli Ollila

Senior scientist

Bioanalytics and biological data science

Industrial biotechnology and food

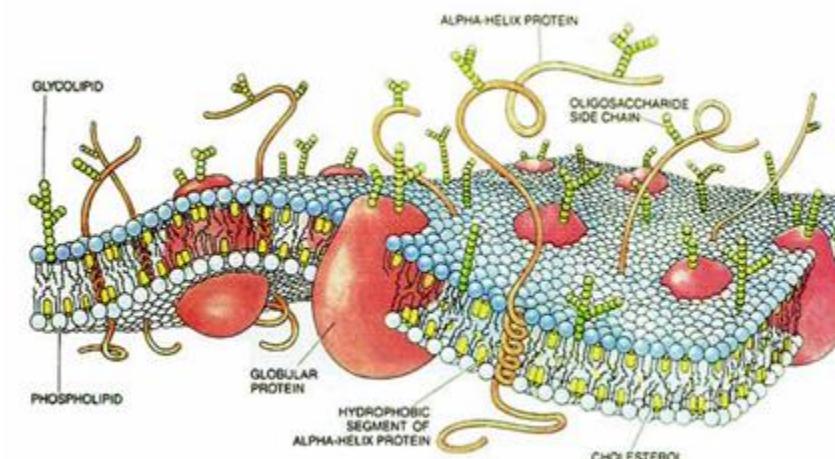
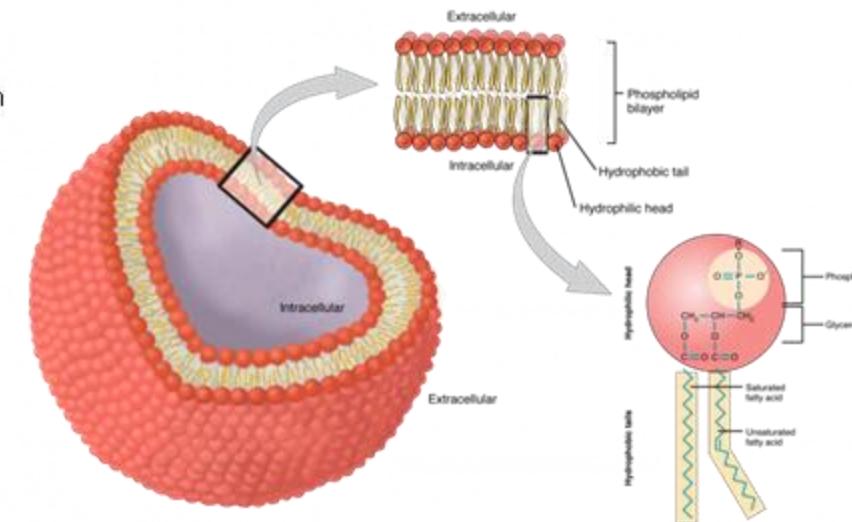
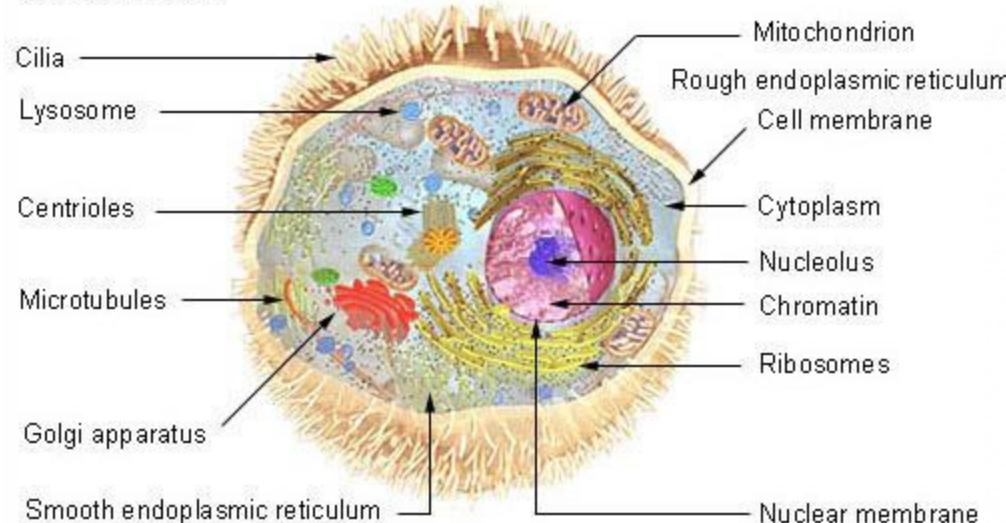
VTT Technical Research Centre of Finland

Structural biology aims to understand structure and dynamics of biological material at molecular level

VTT

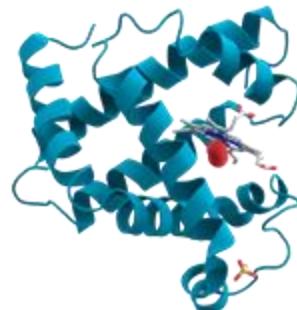
We aim to understand how membranes regulate biological functions by regulating permeability of molecules and interactions with proteins

Cell Structure

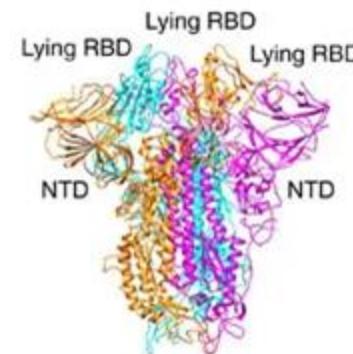


Standard structural biology methods can be used to study molecules with well defined structure

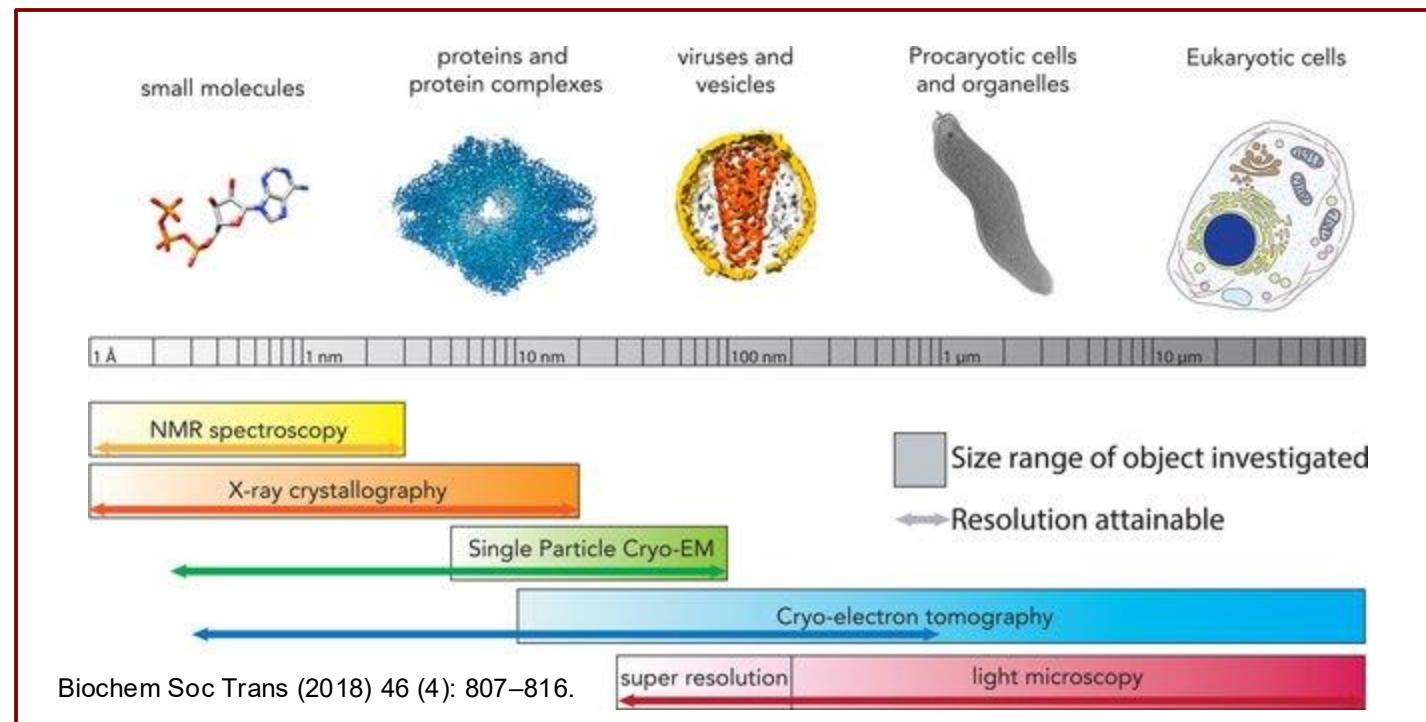
VTT



Myoglobin



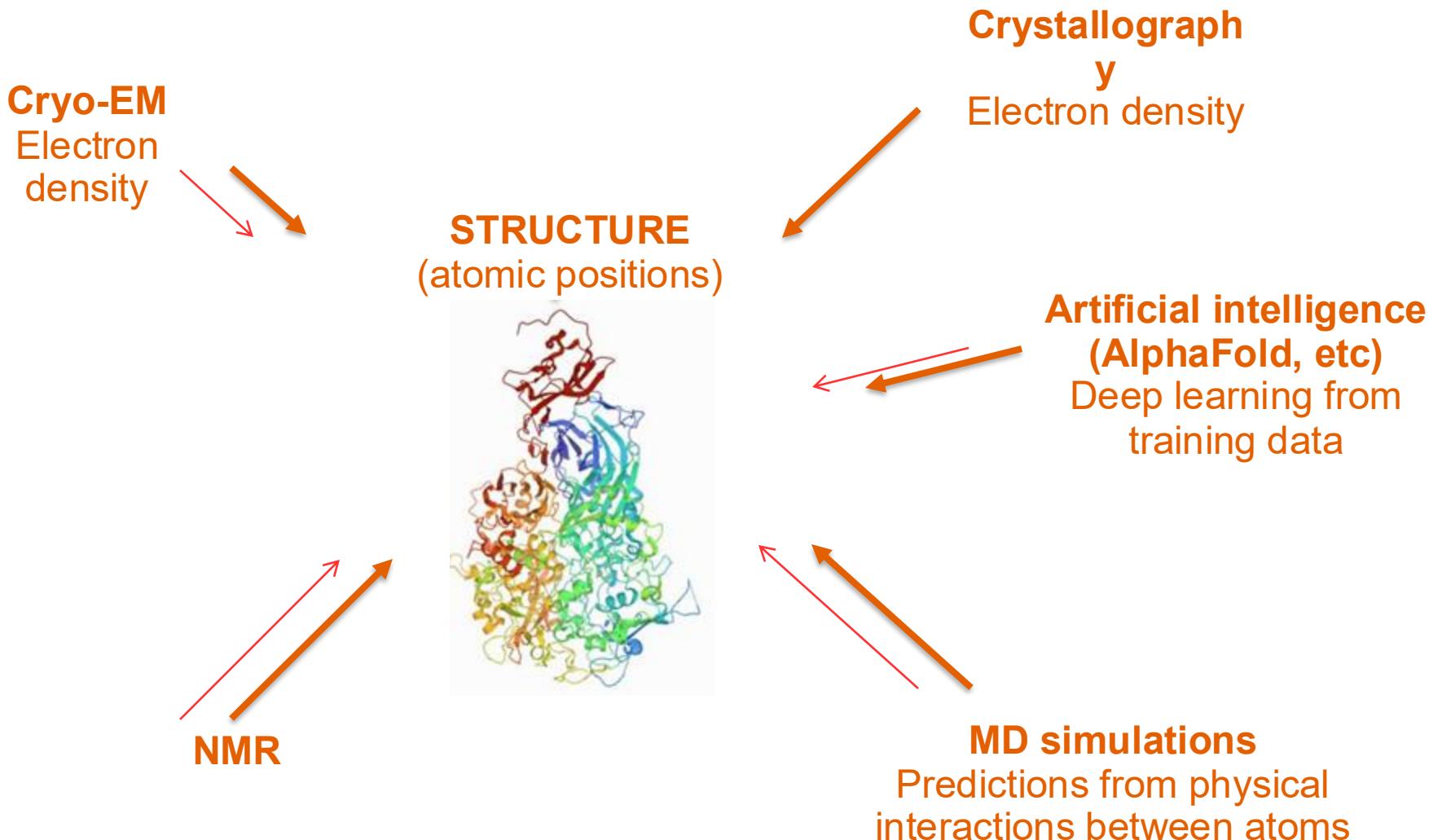
Protein from surface of SARS-CoV



Impressive toolbox available to characterize and understand Folded proteins

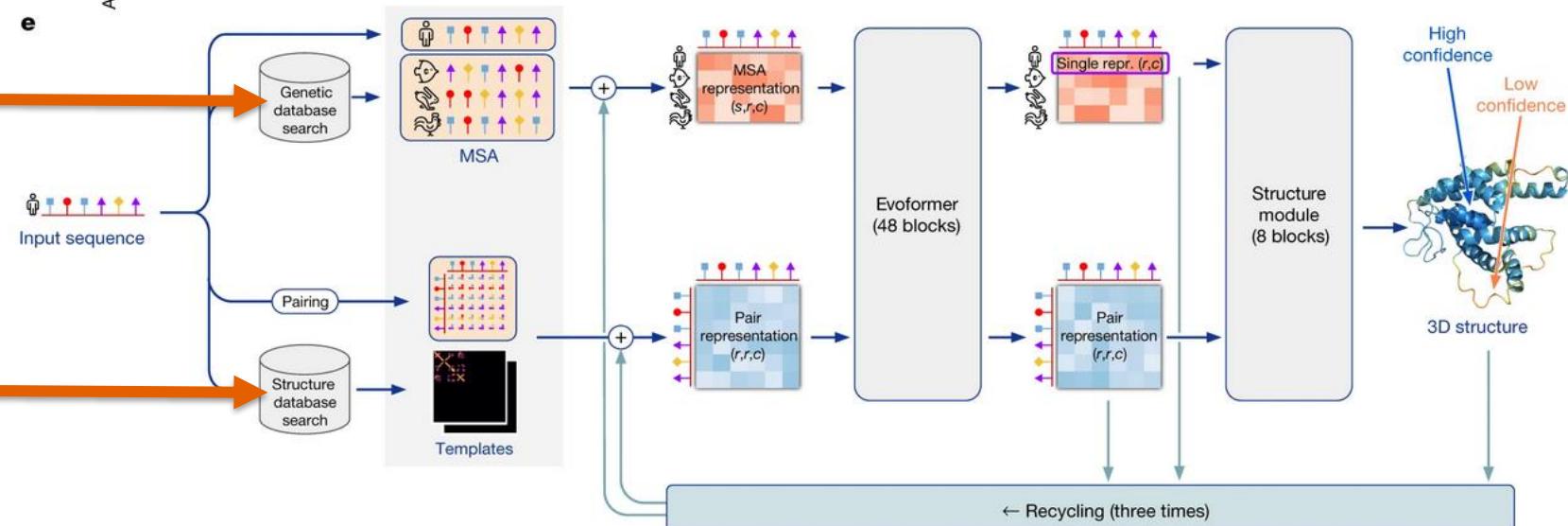
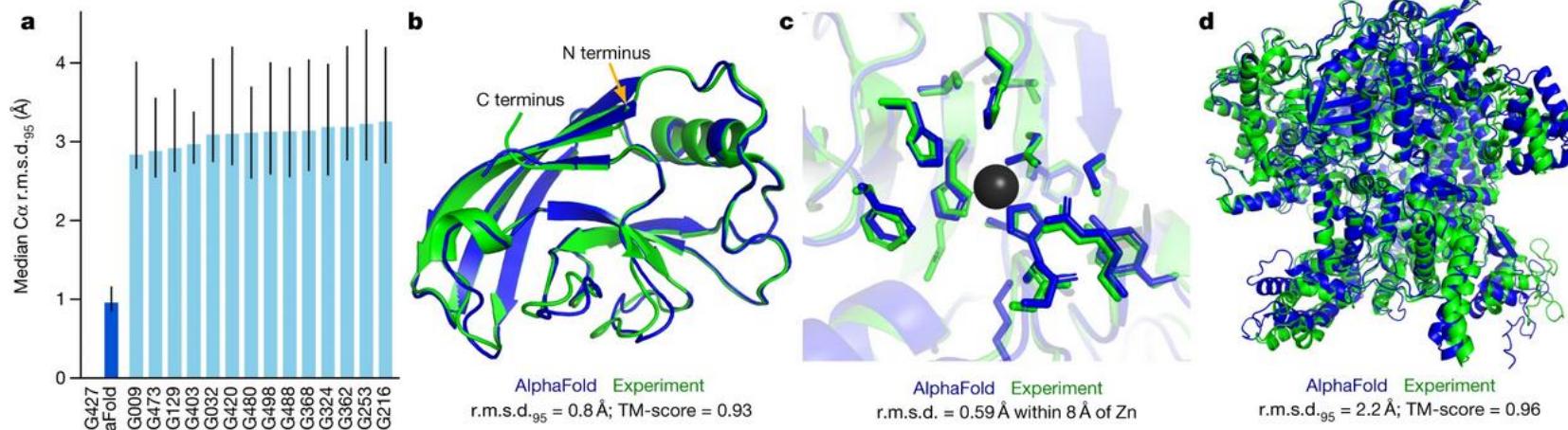
VTT

Wide range of applications from drug design to protein engineering in biotechnology

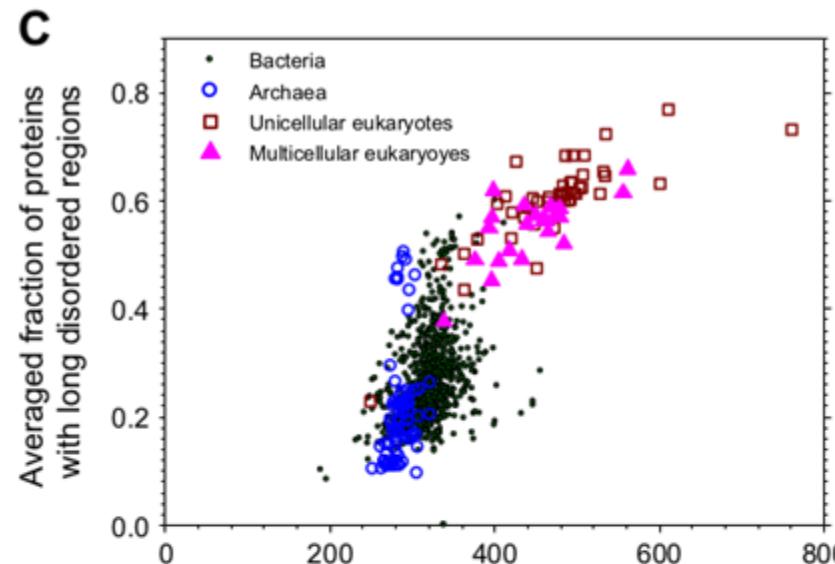


We have alphafold because we have databases for folded proteins

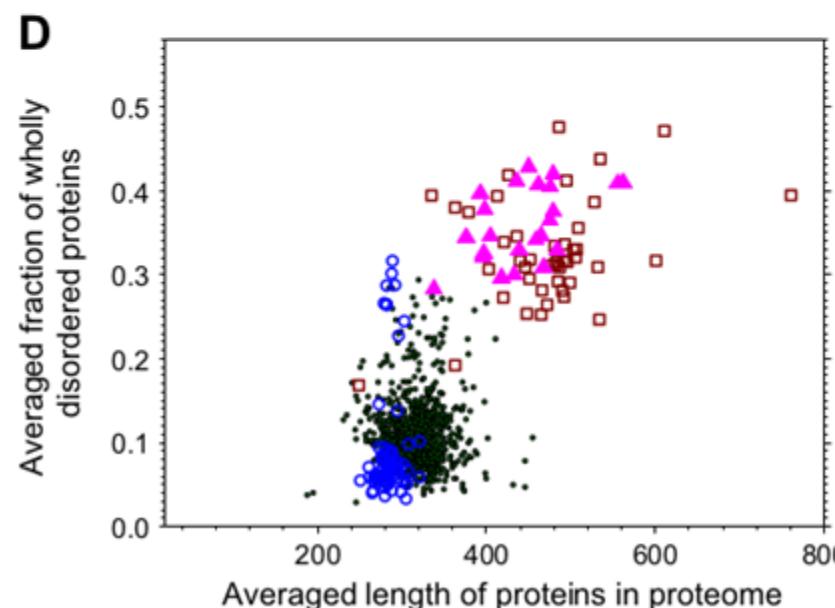
VTT



How about proteins that do not fold?



Large fraction of proteome consists of disordered or partially disordered proteins



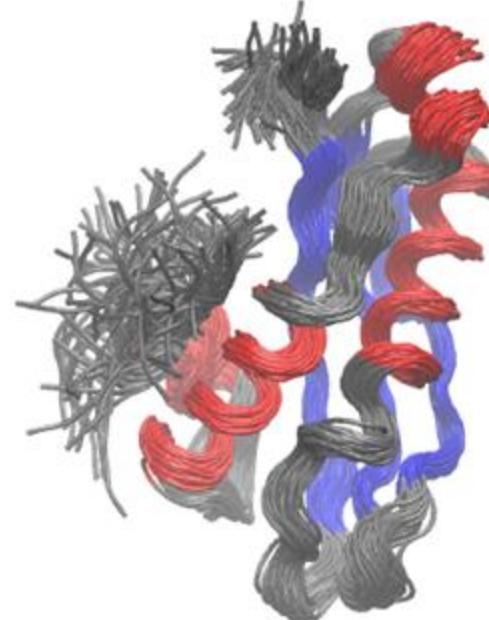
(if $50\% > 0$ disordered, wholly disordered here)

33 % of long (>30 residue) eukaryotic proteins, lipids and sugars are disordered biomolecules that lack the rigid 3D structure

$$p \propto e^{-\frac{U}{k_B T}}$$

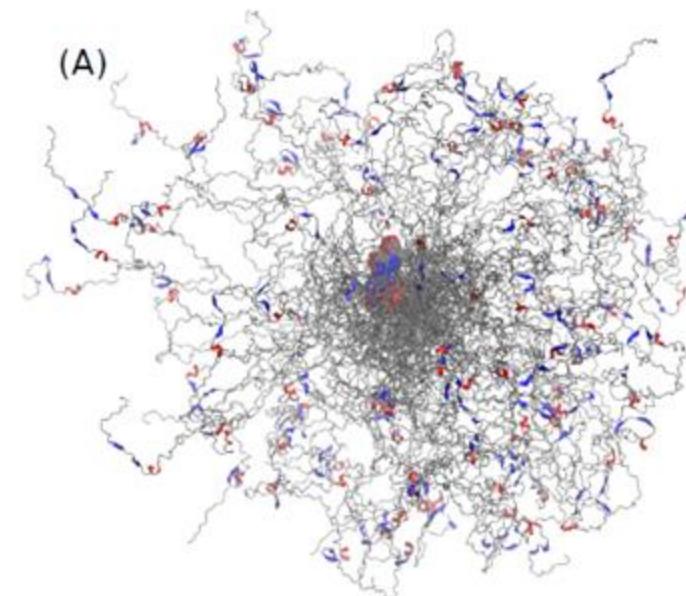
p, probability of a conformation
U, potential energy
 k_B , Boltzmann constant
T, temperature

Folded protein
(AlphaFold)



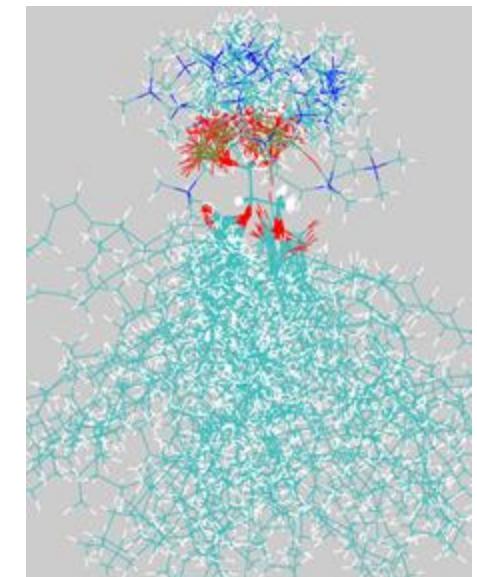
<http://dx.doi.org/10.1021/acs.jpcb.8b02250>
J. Phys. Chem. B 2018, 122, 6559–6569

Partially disordered protein



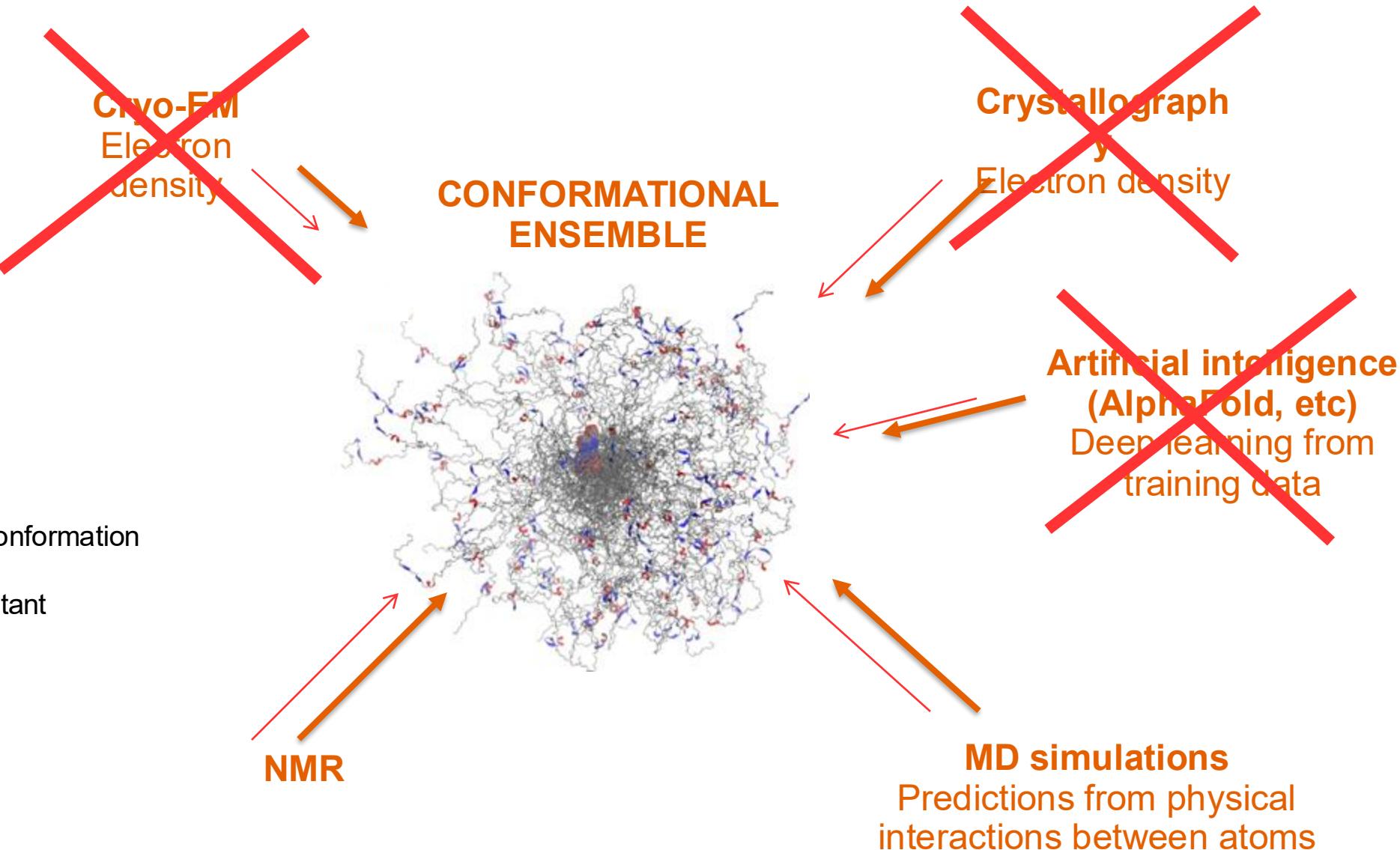
<https://doi.org/10.1039/D0CP03473H>
Phys. Chem. Chem. Phys., 2020, 22, 21185

Lipid molecule in a bilayer



<https://doi.org/10.1038/s41467-024-45189-z>
Kiirikki et al. Nat. Comm. 2024

Most methods are not applicable for disordered proteins

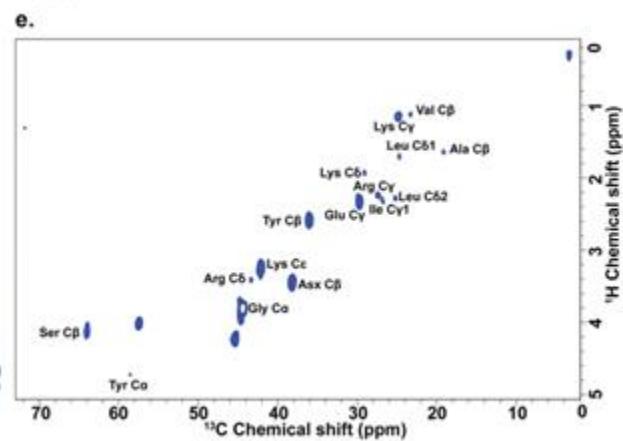
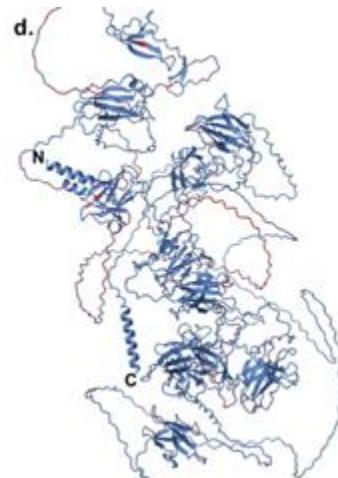
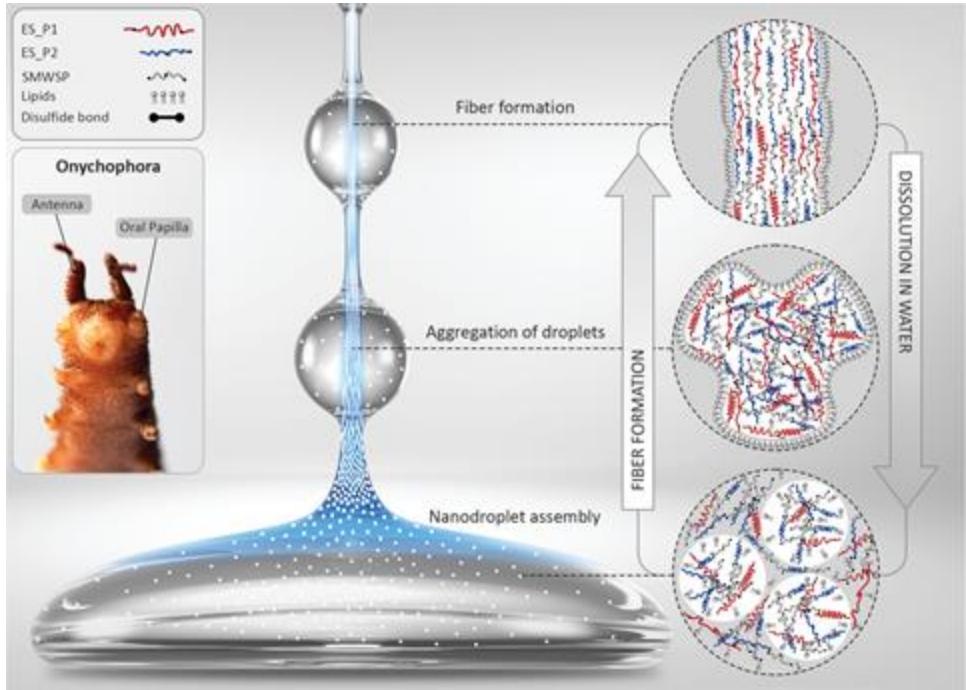


Applications in materials?

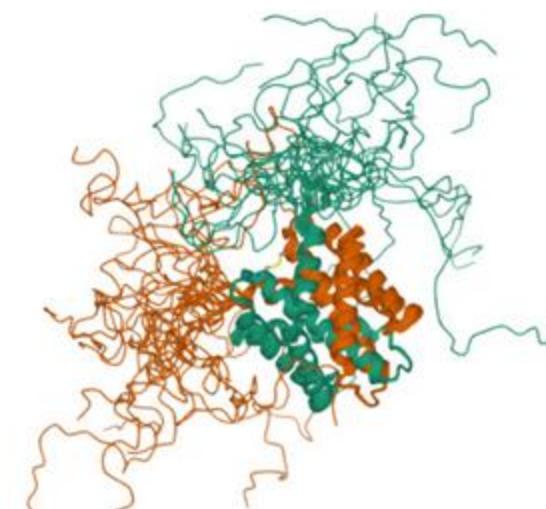
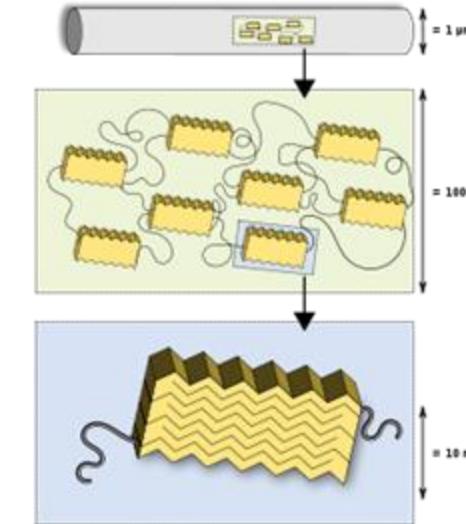
VTT

Velvet Worm Slime Proteins

<https://doi.org/10.1002/advs.202201444>



Spider Silk Proteins

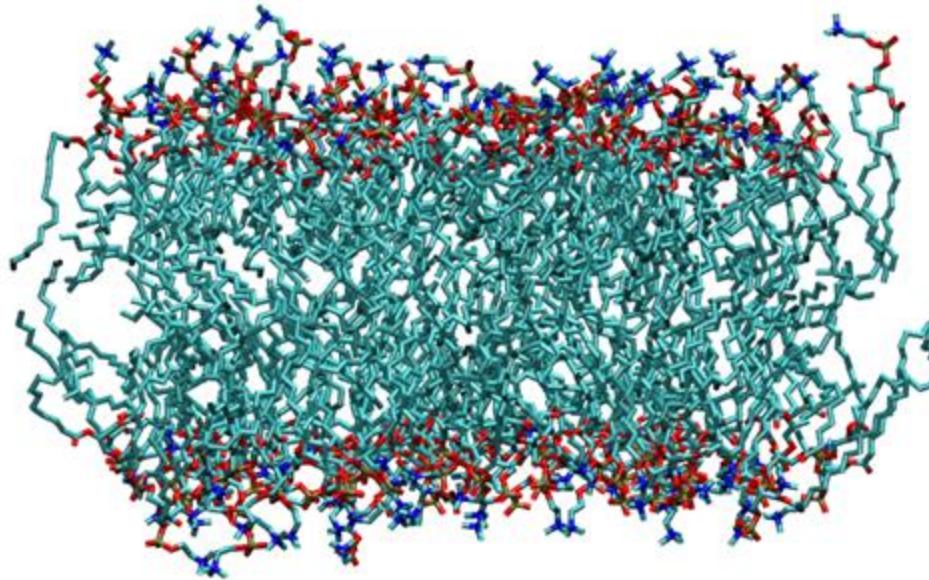


<https://doi.org/10.1038/nature08936>

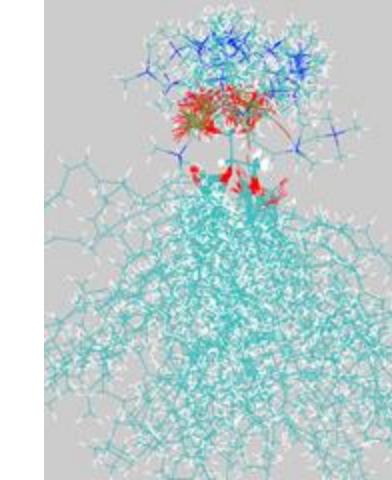
Lipids forming cellular membranes are in disordered liquid state in physiological conditions

VTT

Lipid bilayer

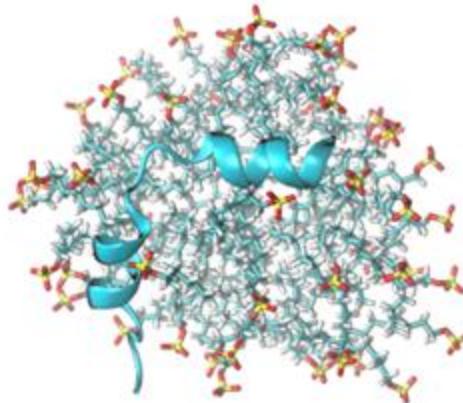


Individual PC lipid molecule in membrane

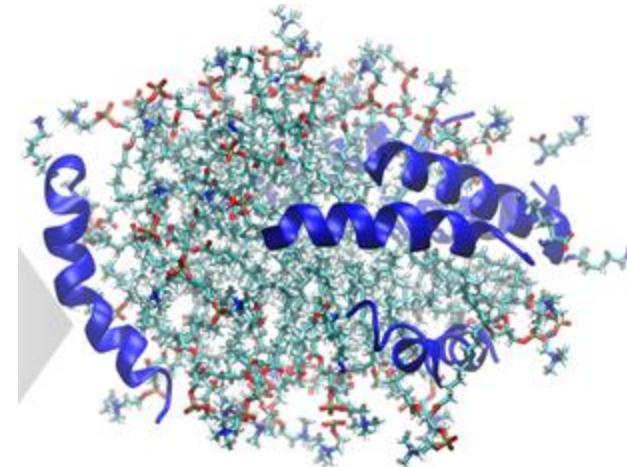


Other types of lipid aggregates

Peptide in a micelle

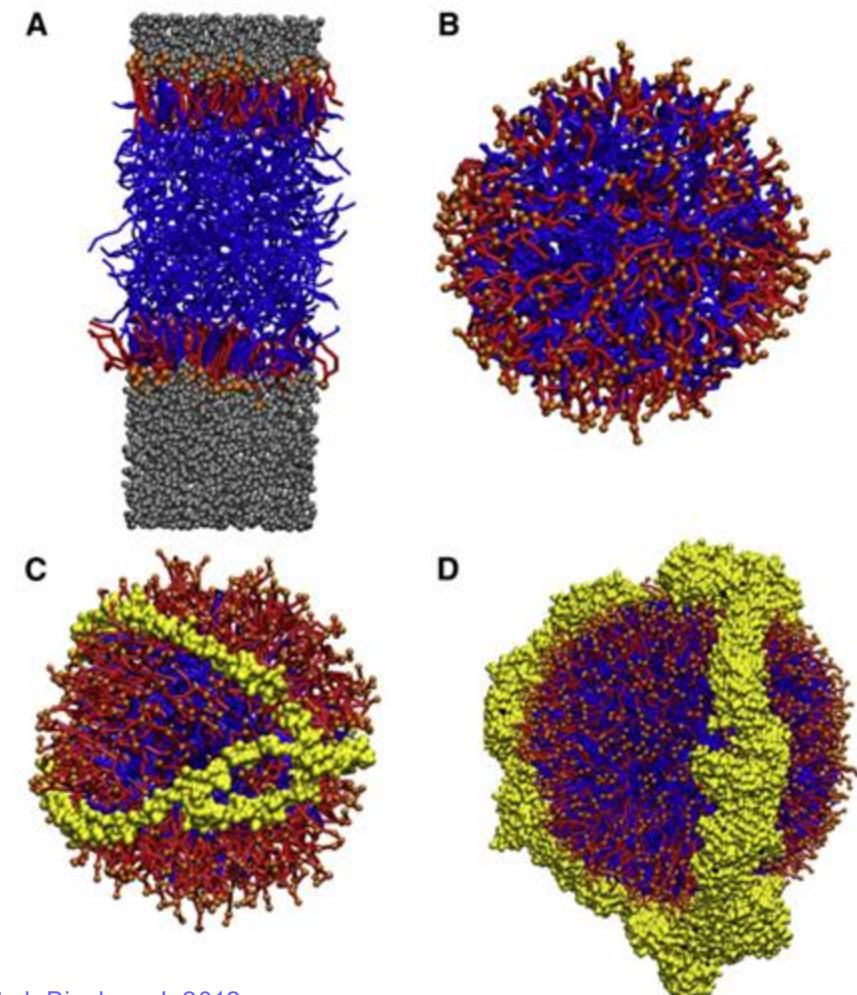


Nanodisc stabilized by peptides



Nencini et al. Comm. Chem. 2024, <https://doi.org/10.1038/s42004-024-01115-4>

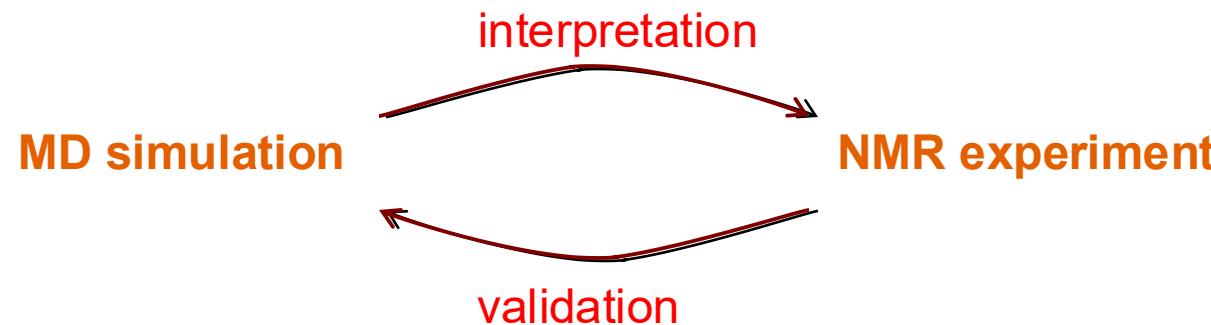
Lipid droplets, HDL and LDL



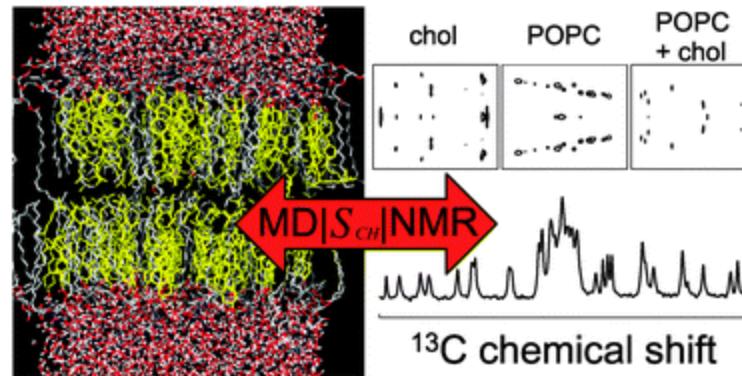
Ollila et al. Biophys. J. 2012
<http://dx.doi.org/10.1016/j.bpj.2012.08.023>

Combined approach for disordered biomolecules

VTT



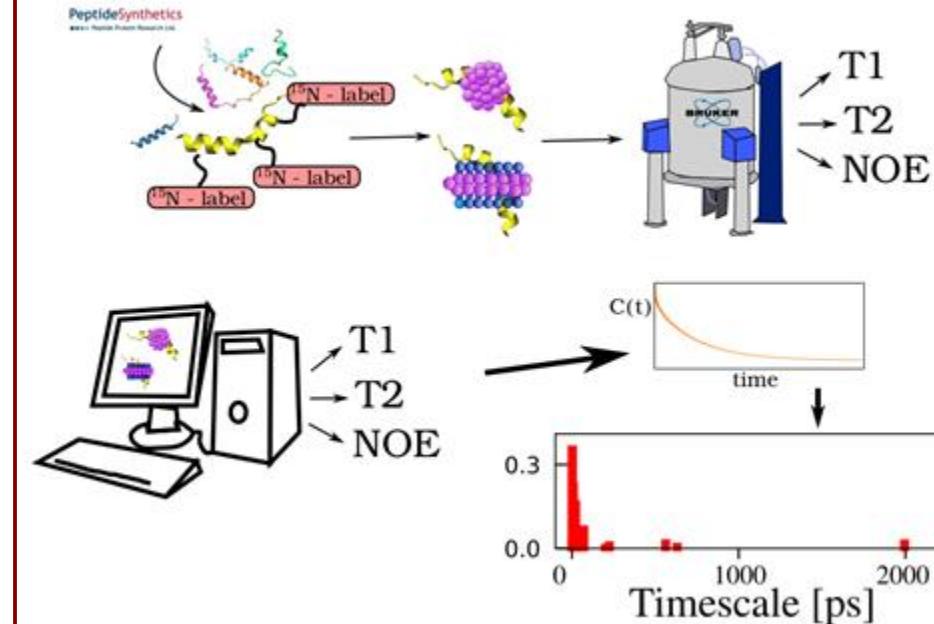
Solid state NMR for membranes



<https://doi.org/10.1039/C2CP42738A>

Ferreira et al. Phys. Chem. Chem. Phys., 2013, 15, 1976-1989

Solution state NMR for disordered proteins and small aggregates



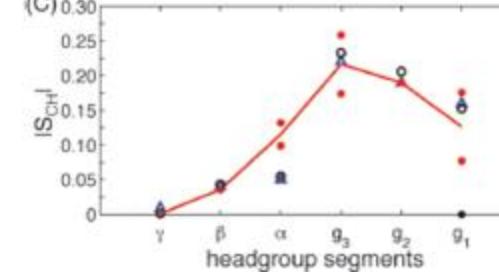
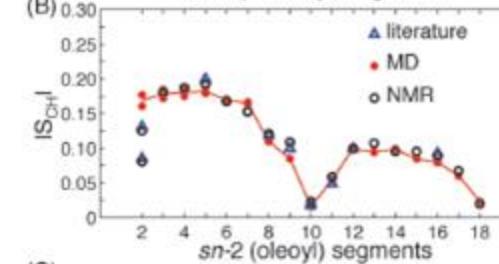
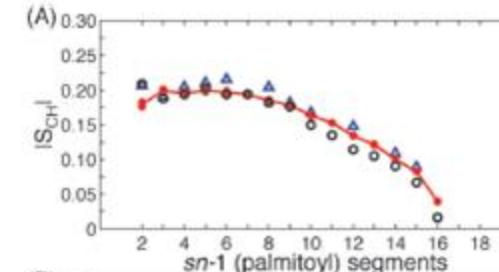
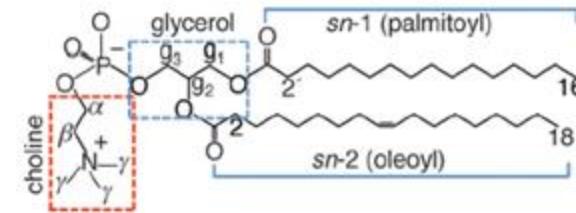
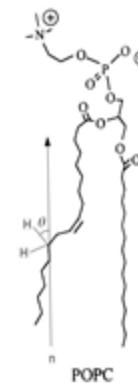
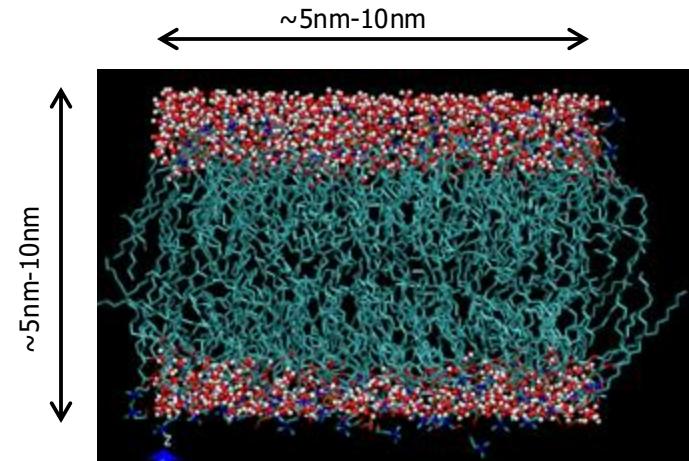
Solid state NMR and MD simulations for membranes

VTT



Tiago Ferreira
Halle, Germany

Lipid membrane systems (solid state NMR)



Order parameter:

$$S_{\text{CH}} = \frac{1}{2} \langle 3 \cos^2 \theta - 1 \rangle$$

Molecular Dynamics (MD) simulations of biomolecules

- Newton's equation of motion

$$m_i \frac{d^2 \mathbf{x}_i}{dt^2} = \mathbf{F}_i \equiv -\nabla_{\mathbf{r}_i} V(\mathbf{r}_1, \dots, \mathbf{r}_N)$$

- Forces between all atoms are calculated from force field (CHARMM, Amber, OPLS, Gromos, etc):

$$\begin{aligned} V(\mathbf{r}) &= \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 \\ &\quad + \sum_{\text{dihedrals}} K_\chi (1 + \cos(n\chi - \delta)) \\ &\quad + \sum_{\text{nonbonded-pairs, } i, j} \left[\frac{q_i q_j}{4\pi e_0 r_{ij}} - \varepsilon_{ij} \left\{ \left(\frac{R_{\min ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\min ij}}{r_{ij}} \right)^6 \right\} \right] \end{aligned}$$

Energy dependencies on:

1. Bond length
2. Bond valence angle
3. Bond dihedral angle
4. Non-bonded electrostatic interactions
5. Non-bonded van-der Waals interactions

Molecular Dynamics (MD) simulations of biomolecules

- Time resolved trajectories for individual atoms
- Dynamical pictures of biomolecule with unprecedent details
- Time (μ s) and lenght (20nm) scales comparable with experiments
 - High potential for applications in wide range of fields (biochemistry, drug design, structural biology, material science)
 - Are these simulations sufficiently accurate to provide interpretation of NMR experiments with atomistic resolution?**

NMRLipids open collaboration project

2013-2023

VTT

- Aims to find and distribute MD simulations of lipid bilayers that describe membrane properties in agreement with experiments using open collaboration
- Open collaboration inspired by Linux and Polymath
- All the data, methods and contributions are publicly available all the time. Contributors are offered an authorship in the produced publications. The decision of the authorship is based on self-assessment. Authors are alphabetical order.

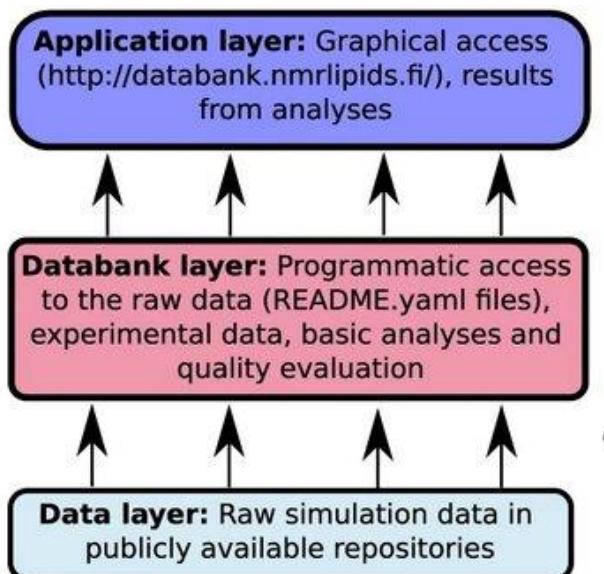


Markus Miettinen
University of Bergen

<http://nmrlipids.blogspot.fi/>
<https://github.com/nmrlipids/>
<https://nmrlipids.github.io/>

NMRlipids databank: Toward machine learning models for disordered biomolecules

- Automatic quality evaluation and ranking of membrane simulations against experimental data
- Programmatic access to simulation data for machine learning: <https://nmrlipids.github.io/>
- NMRlipids databank-GUI: <https://databank.nmrlipids.fi/>



nature communications

Article

<https://doi.org/10.1038/s41467-024-45189-z>

Overlay databank unlocks data-driven analyses of biomolecules for all

Received: 2 June 2023

Accepted: 17 January 2024

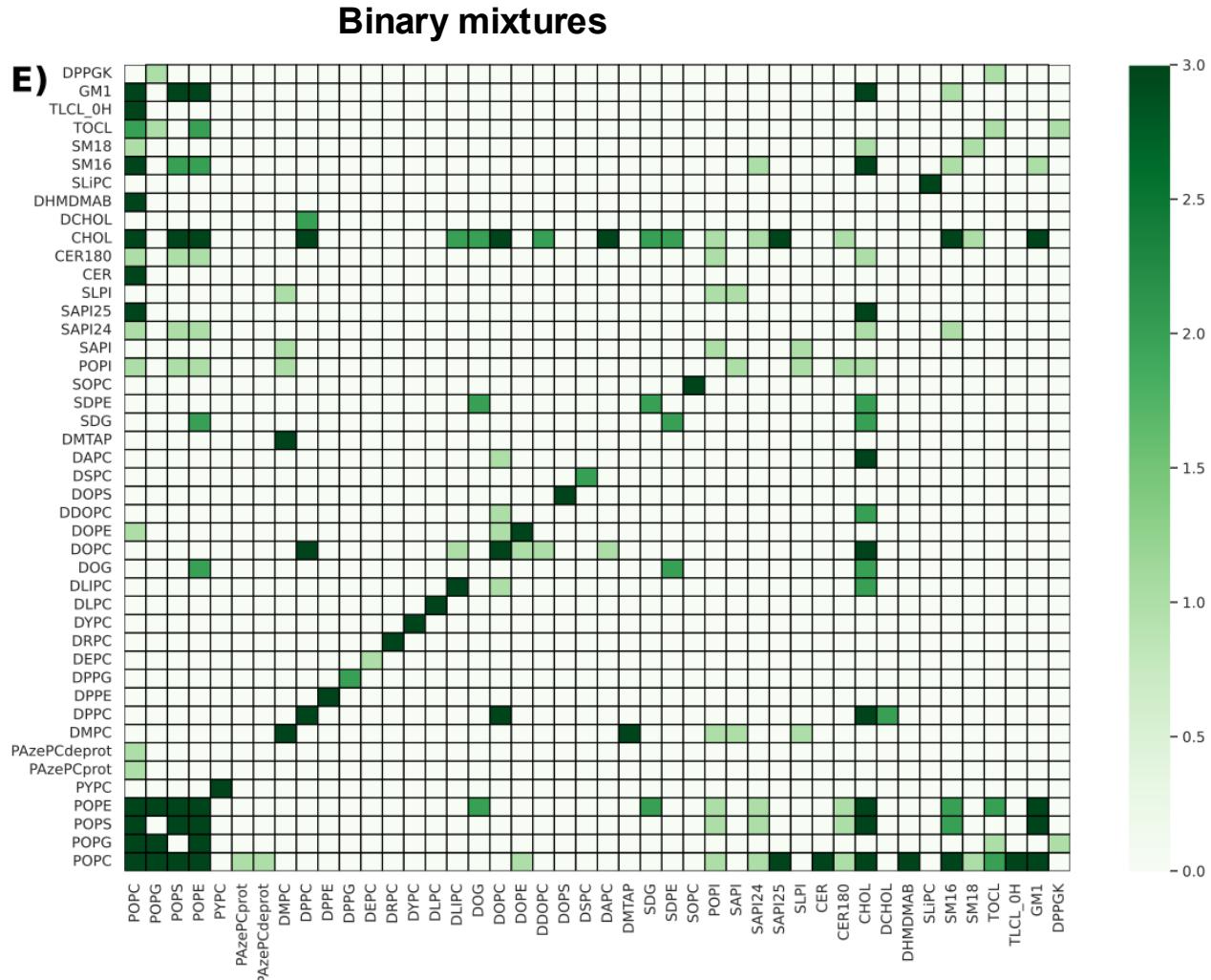
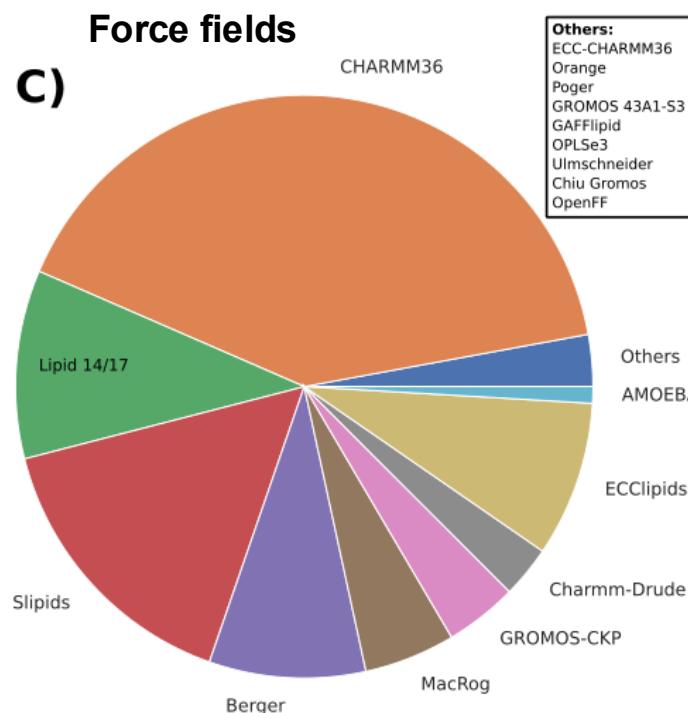
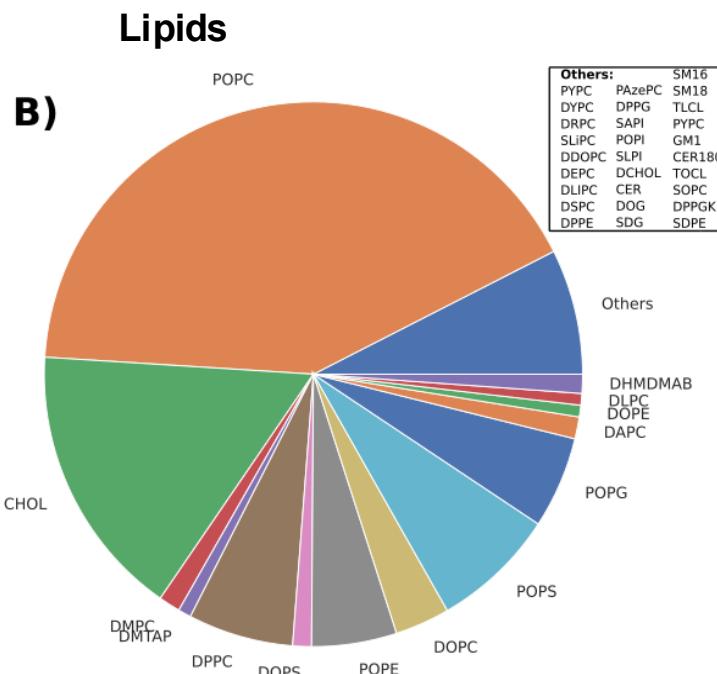
Published online: 07 February 2024

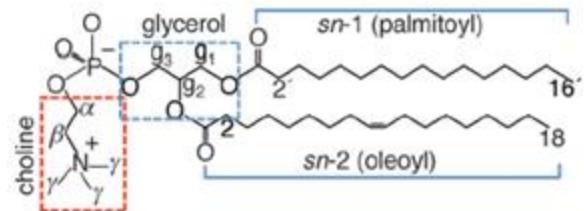
Check for updates

Anne M. Kiirikki¹, Hanne S. Antila^{2,3}, Lara S. Bort^{2,4}, Pavel Buslaev⁵, Fernando Favela-Rosales ⁶, Tiago Mendes Ferreira⁷, Patrick F. J. Fuchs ^{8,9}, Rebeca Garcia-Fandino¹⁰, Ivan Gushchin, Batuhan Kav ^{11,12}, Norbert Kučerka¹³, Patrik Kula¹⁴, Milla Kurki ¹⁵, Alexander Kuzmin ¹⁶, Anusha Lalitha ¹⁶, Fabio Lolicato ^{17,18}, Jesper J. Madsen ^{19,20}, Markus S. Miettinen ^{2,21,22}, Cedric Mingham²³, Luca Monticelli ^{24,25}, Ricky Nencini^{1,26}, Alexey M. Nesterenko^{21,22}, Thomas J. Piggot²⁷, Ángel Piñeiro²⁸, Nathalie Reuter^{21,22}, Suman Samantray ^{11,29}, Fabián Suárez-Lestón ^{10,28,30}, Reza Talandashti^{21,22} & O. H. Samuli Ollila ^{1,31}✉

NMRlipids databank content

The VTT logo consists of the letters "VTT" in a bold, white, sans-serif font, centered on an orange rectangular background.



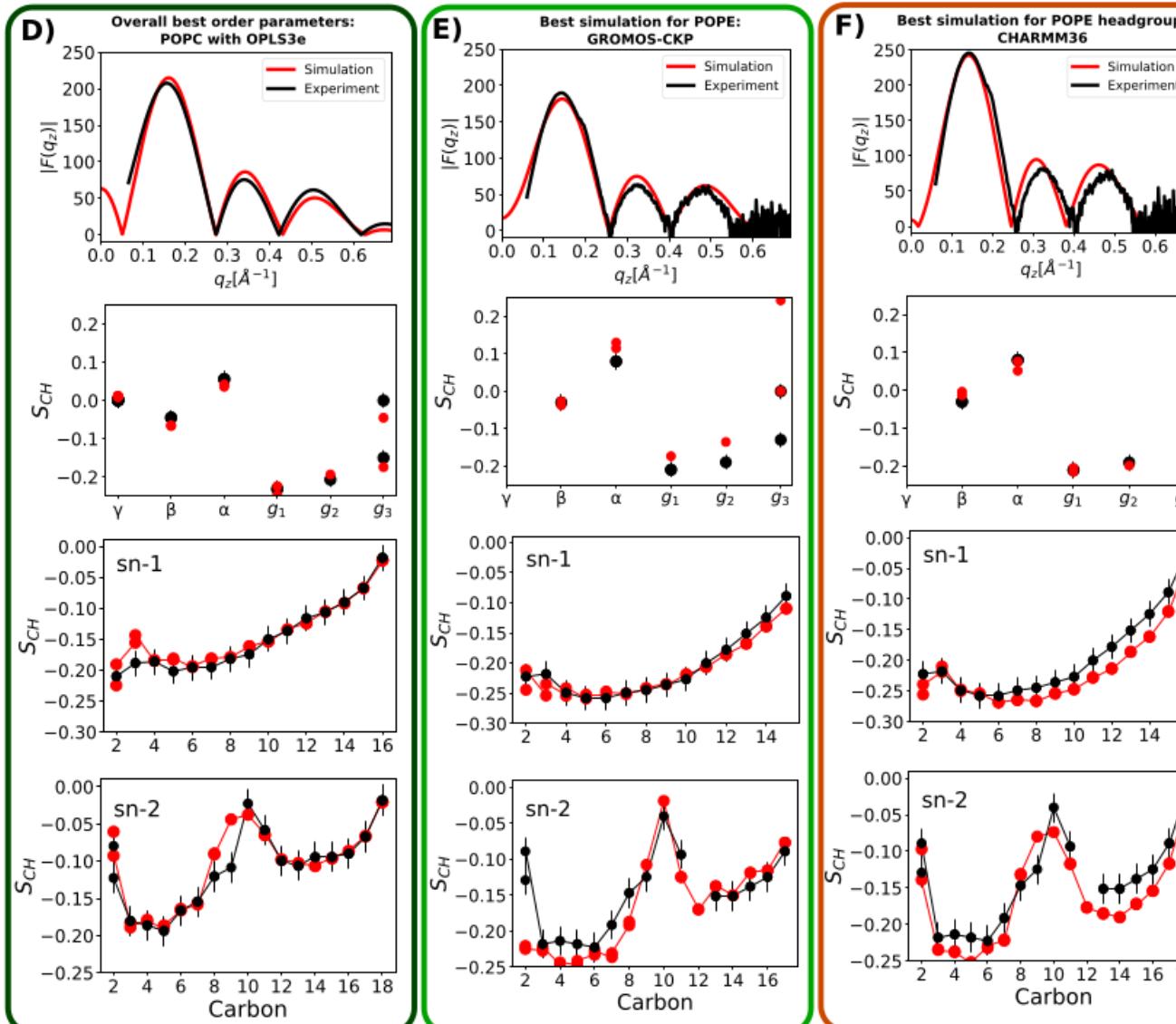


Ranking simulation quality against experiments

<https://hmrlipids.github.io/exampleAndTutorials.html>

VTT

C-H bond order parameters and x-ray scattering



Top 25 simulations (probability within experiments)

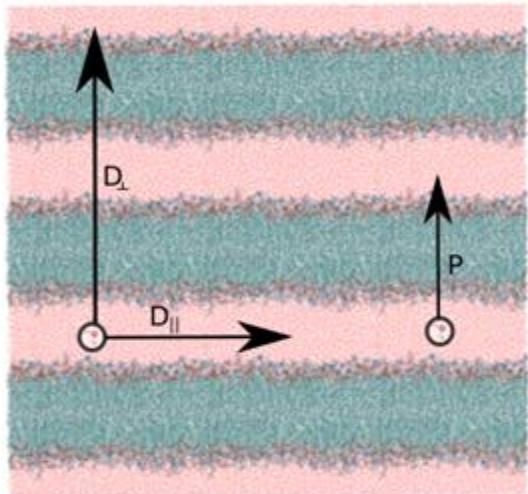
	P_{tails}	P^{hg}	P^{total}	FF_q	τ_{rel}	Force field	Molecules	Temperature	ID
1	0.86	0.76	0.83	0.15	0.4	OPLS3e	POPC:SOL (200:8859)	300.00	1
2	0.83	0.61	0.76	0.45	7.4	Slipids	POPC:SOL (512:23943)	298.00	617
3	0.81	0.58	0.73	0.45	0.4	Slipids	POPC:SOL (1024:51200)	298.15	696
4	0.73	0.70	0.72	0.65	0.3	MacRog	POPC:SOL (1024:51200)	298.15	658
5	0.73	0.63	0.69	0.55	0.5	MacRog	POPC:SOL (128:5120)	300.00	457
6	0.76	0.54	0.69	0.55	0.4	Slipids	POPC:SOL (256:12800)	298.15	708
7	0.69	0.67	0.68	0.55	0.3	MacRog	POPC:SOL (256:12800)	298.15	675
8	0.71	0.62	0.68	0.55	0.3	MacRog	POPC:SOL (64:3200)	298.15	674
9	0.69	0.54	0.64	0.76	0.4	Slipids	POPC:SOL (64:3200)	298.15	664
10	0.60	0.64	0.61	0.55	7.7	MacRog	POPC:SOL (288:14400)	298.00	63
11	0.58	0.65	0.60	0.55	1.0	ECC-lipids	POPC:SOL (128:6400)	300.00	573
12	0.66	0.47	0.60	0.15	0.2	Lipid17	POPC:SOL (64:3200)	298.15	715
13	0.57	0.60	0.58	0.6		AMOEBA	DOPC:SOL (72:2880)	303.00	742
14	0.60	0.53	0.58	0.25	0.2	Lipid17	POPC:SOL (256:12800)	298.15	657
15	0.86	0.01	0.58	0.95	0.4	Berger	POPC:SOL (256:10342)	300.00	115
16	0.58	0.54	0.57	0.25	0.2	Lipid17	POPC:SOL (1024:51200)	298.15	684
17	0.81	0.02	0.55	1.05	1.7	Berger	POPC:SOL (128:7290)	298.00	497
18	0.49	0.65	0.54	0.76	1.1	ECC-lipids	POPC:SOL (128:6400)	300.00	43
19	0.65	0.29	0.53	0.4	1.3	GROMOS-CKP	POPE:SOL (500:25000)	310.00	400
20	0.78	0.01	0.52	0.1	1.3	Slipids	POPE:SOL (500:25000)	310.00	414
21	0.72	0.10	0.51		0.7	Slipids	POPE:SOL (336:13460)	310.00	29
22	0.71	0.10	0.51		0.8	Slipids	POPE:SOL (336:13460)	310.00	74
23	0.66	0.16	0.50	0.4	0.2	ECClipids	POPS:SOL:SOD (72:3600:72)	298.00	443
24	0.38	0.67	0.48	1.26	0.7	CHARMM36	POPC:SOL (256:9767)	300.00	558
25	0.38	0.67	0.47	1.26	0.3	CHARMM36	POPC:SOL (1024:51200)	298.15	701

Water diffusion in membrane matrix from the NMRlipids databank

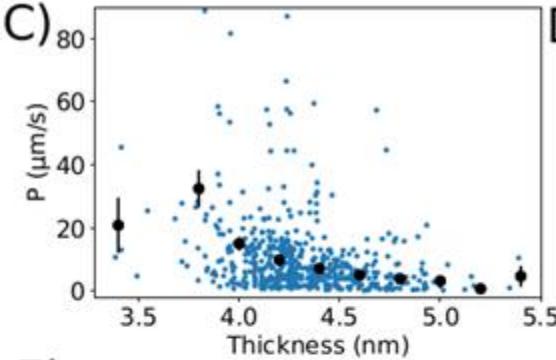
VTT

- Important parameters for MRI imaging and pharmacokinetics
- Good statistics is difficult to collect from standard MD due to slow rate

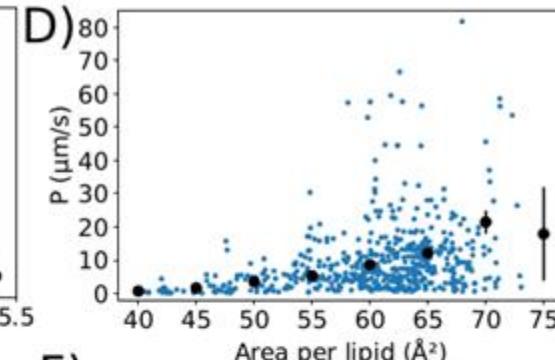
A)



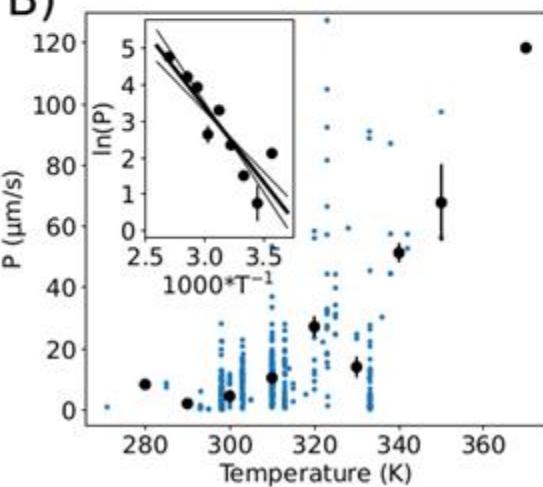
C)



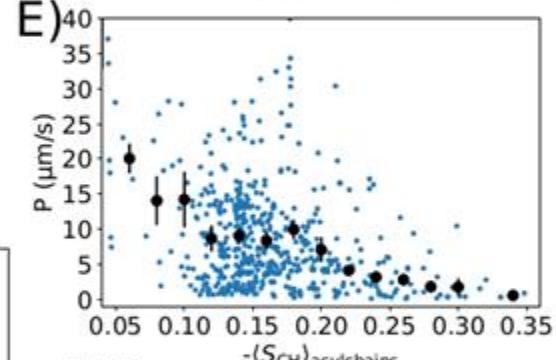
D)



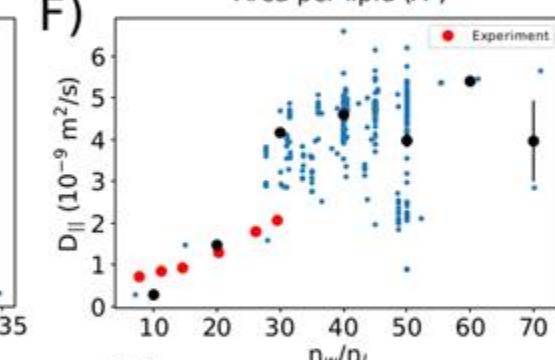
B)



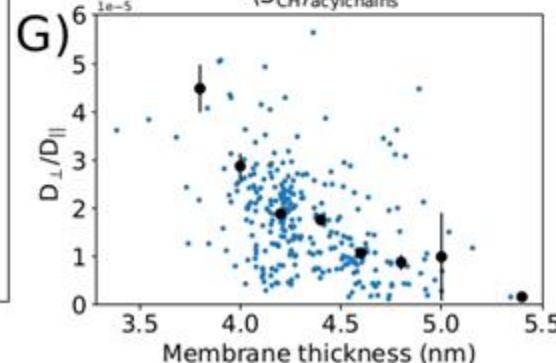
E)



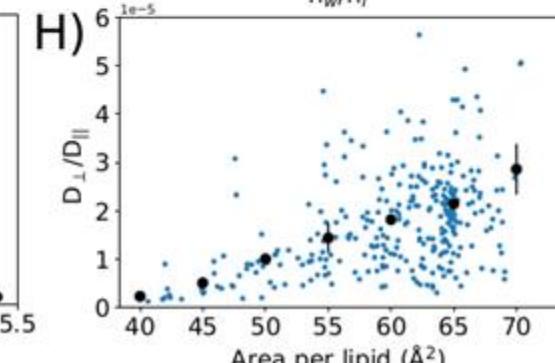
F)



G)



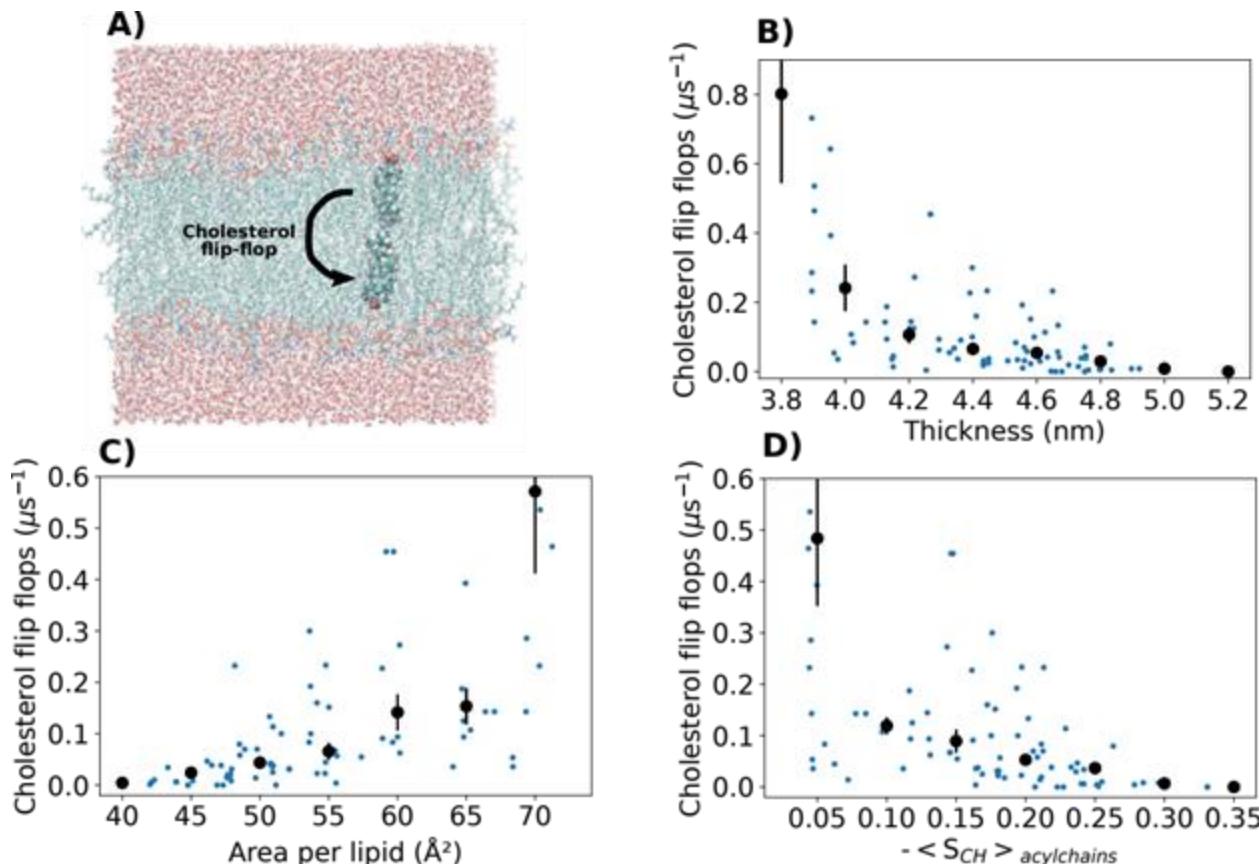
H)



Cholesterol flip-flop rates from the NMRlipids databank

VTT

- Flip-flops across leaflets regulate lipid trafficking and membrane properties
- Good statistics is difficult to collect from standard MD due to slow rate (less than once per microsecond)
- Analysis from the NMRlipids databank reveals non-linear dependence on Membrane thickness and packing



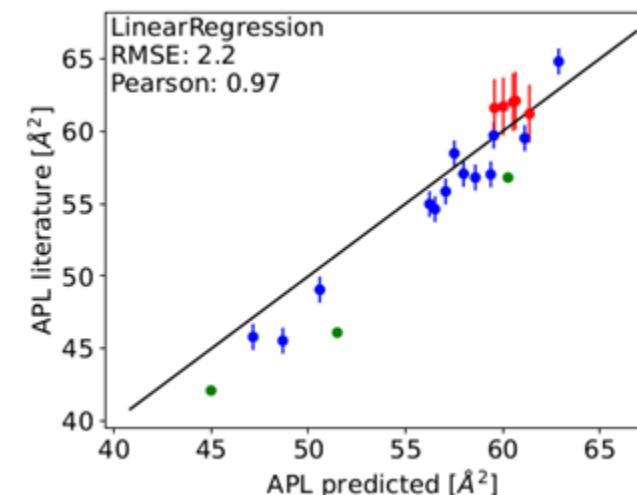
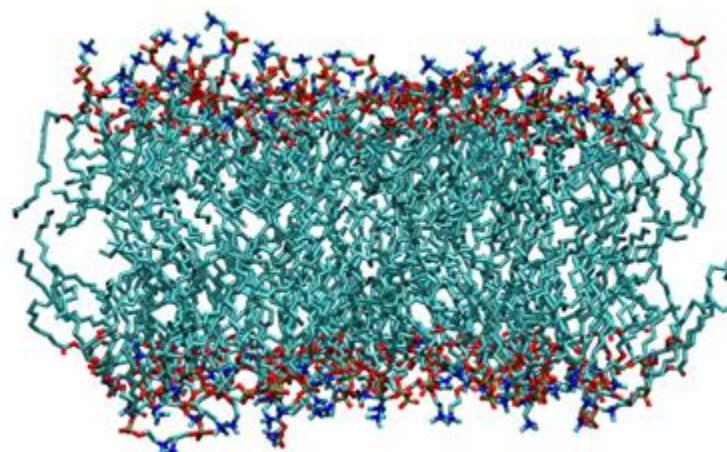
Machine learning model predicting complex membrane properties

Table 1 | Areas per lipid (APL) and membrane thicknesses predicted by the linear regression model trained using the NMRLipids Databank for membrane compositions corresponding different biological membranes

Membrane	Composition [lipid(%-fraction)]	APL [\AA^2]	Thickness [nm^2]
Mitochondria	POPC(37) : POPE(31) : PI(6) : CL(22) : CHOL(4) ^{7,70,71}	74 (66) ^a	4.4
Bacterial	POPC(20) : POPE(35) : POPG(35) : CL(5) : CHOL(5) ⁷²	61 (60) ^a	4.2
ER	POPC(54) : POPE(20) : POPS(4) : PI(11) : SM(4) : CHOL(8) ^{7,70,71}	57	4.6
Golgi	POPC(36) : POPE(21) : POPS(6) : PI(12) : SM(7) : CHOL(18) ^{7,70,71}	52	4.8
Plasma	POPC(23) : POPE(11) : POPS(8) : PI(7) : SM(17) : CHOL(34) ^{7,70,71}	45	5.0
Synaptic	POPC(28) : POPE(20) : POPS(9) : PI(2) : SM(4) : CHOL(37) ⁷³	45	4.8
Influenza	POPC(5) : POPE(32) : POPS(15) : SM(9) : CHOL(40) ^{74,75}	42	4.8

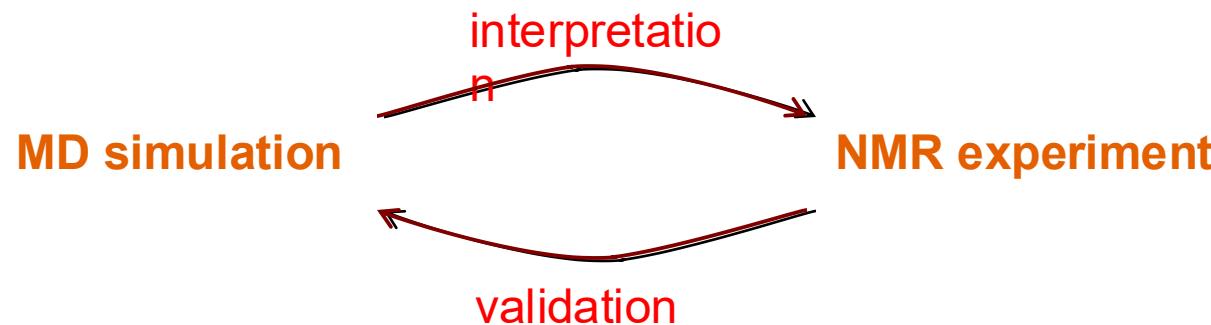
Systems are sorted in the order of increasing membrane packing. Compositions of different membranes are estimated based on references given in the composition column.

^aValue in parenthesis is the area per acyl chain taking into account that cardiolipin has four chains per molecule while other lipids have two.

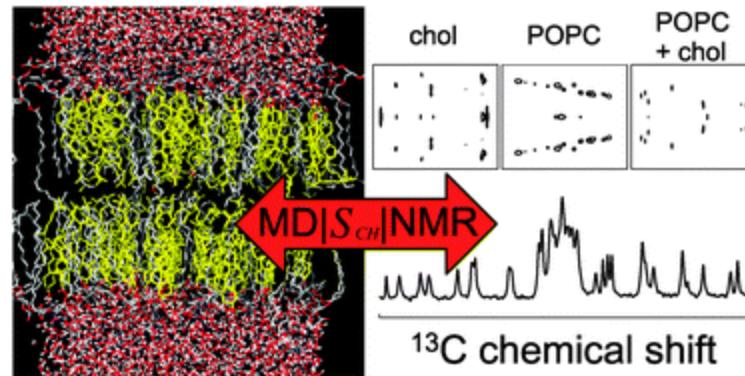


Combined approach for disordered biomolecules

VTT



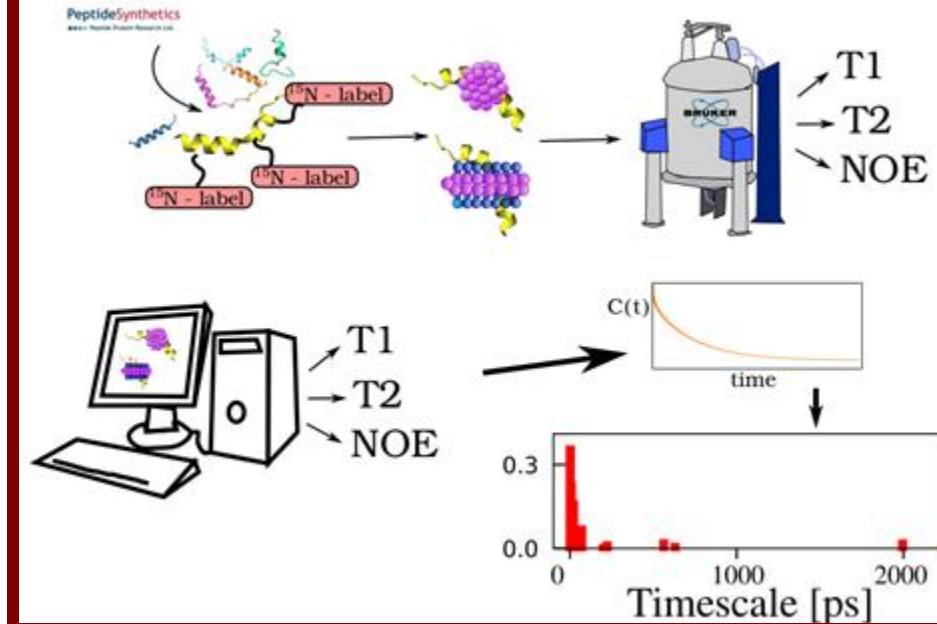
Solid state NMR for membranes



<https://doi.org/10.1039/C2CP42738A>

Ferreira et al. Phys. Chem. Chem. Phys., 2013, 15, 1976-1989

Solution state NMR for disordered proteins and small aggregates

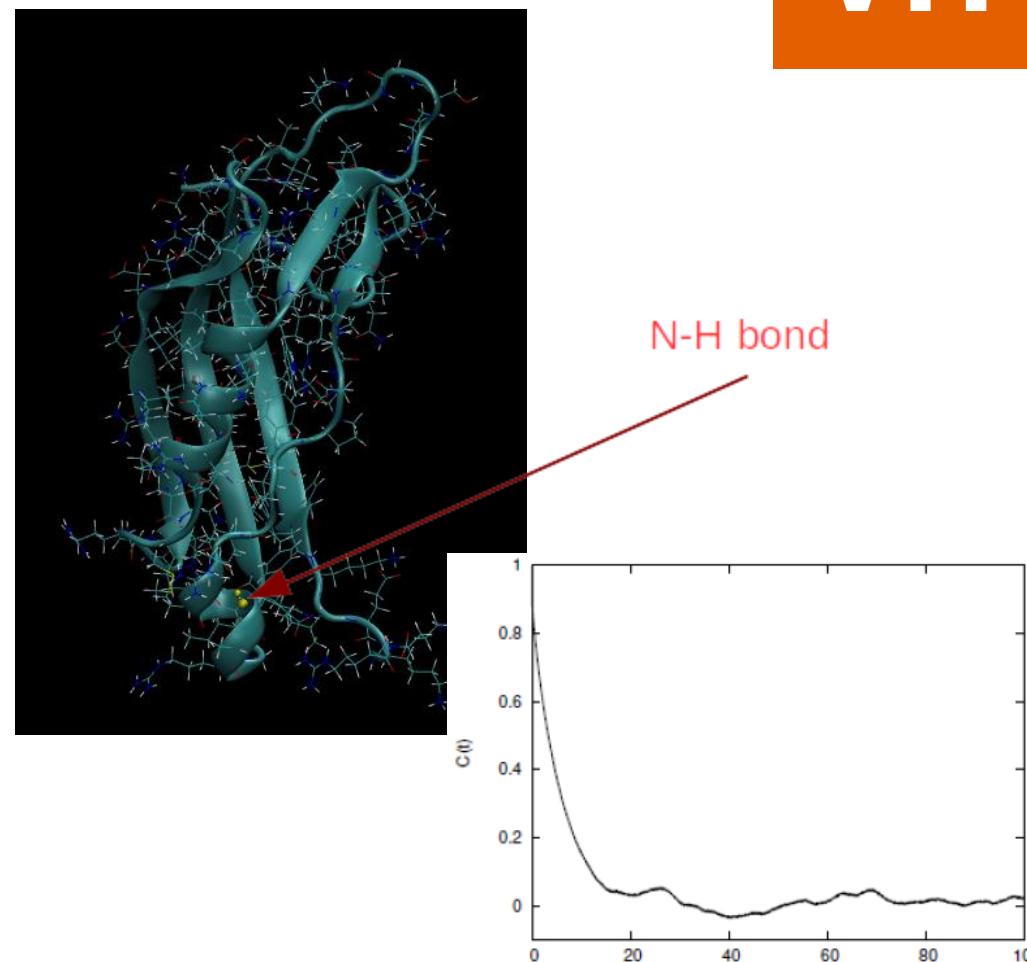


NMR ^{15}N spin relaxation times detect protein backbone dynamics

$$\frac{1}{T_1} = R_1 = d_{oo}[3J(\omega_N) + J(\omega_{H-N}) + 6J(\omega_{N+H})] + c_{oo}\omega_N^2 J(\omega_N)$$

$$\frac{1}{T_2} = R_2 = \frac{1}{2}d_{oo}[4J(0) + 3J(\omega_N) + J(\omega_{H-N}) + 6J(\omega_H) + 6J(\omega_{N+H})] + \frac{1}{6}c_{oo}\omega_N^2[4J(0) + 3J(\omega_N)]$$

$$\text{NOE} = 1 + \frac{\gamma_H}{\gamma_N} d_{oo} T_1 [6J(\omega_{H+N}) - J(\omega_{H-N})]$$

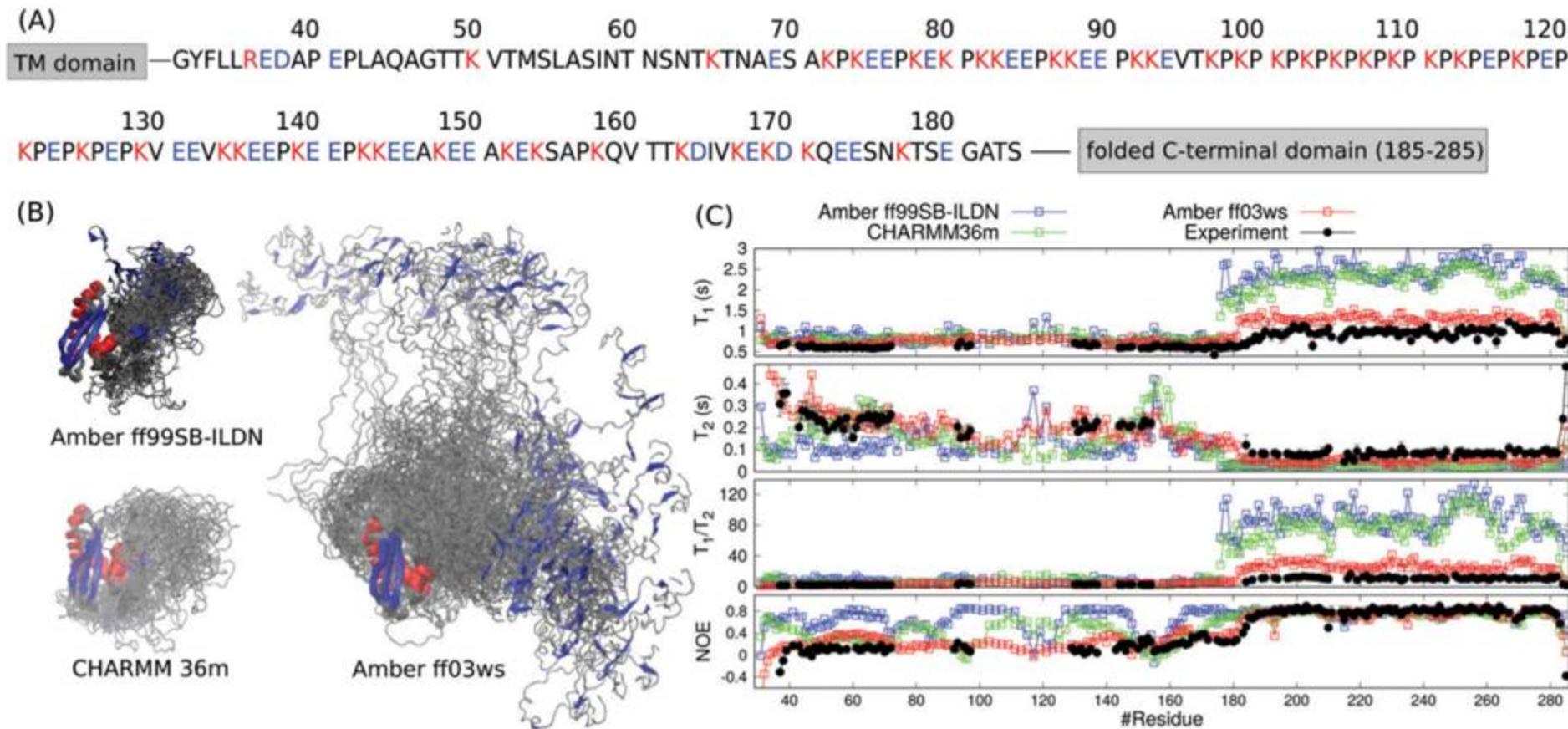


$$J(\omega) = 2 \int_0^\infty C(t) \cos(\omega t) dt$$

$$C(t) = \langle P_2[\mathbf{e}_{\text{LF}}(\tau) \mathbf{e}_{\text{LF}}(\tau + t)] \rangle$$

Backbone ^{15}N spin relaxation are sensitive also to conformational ensemble

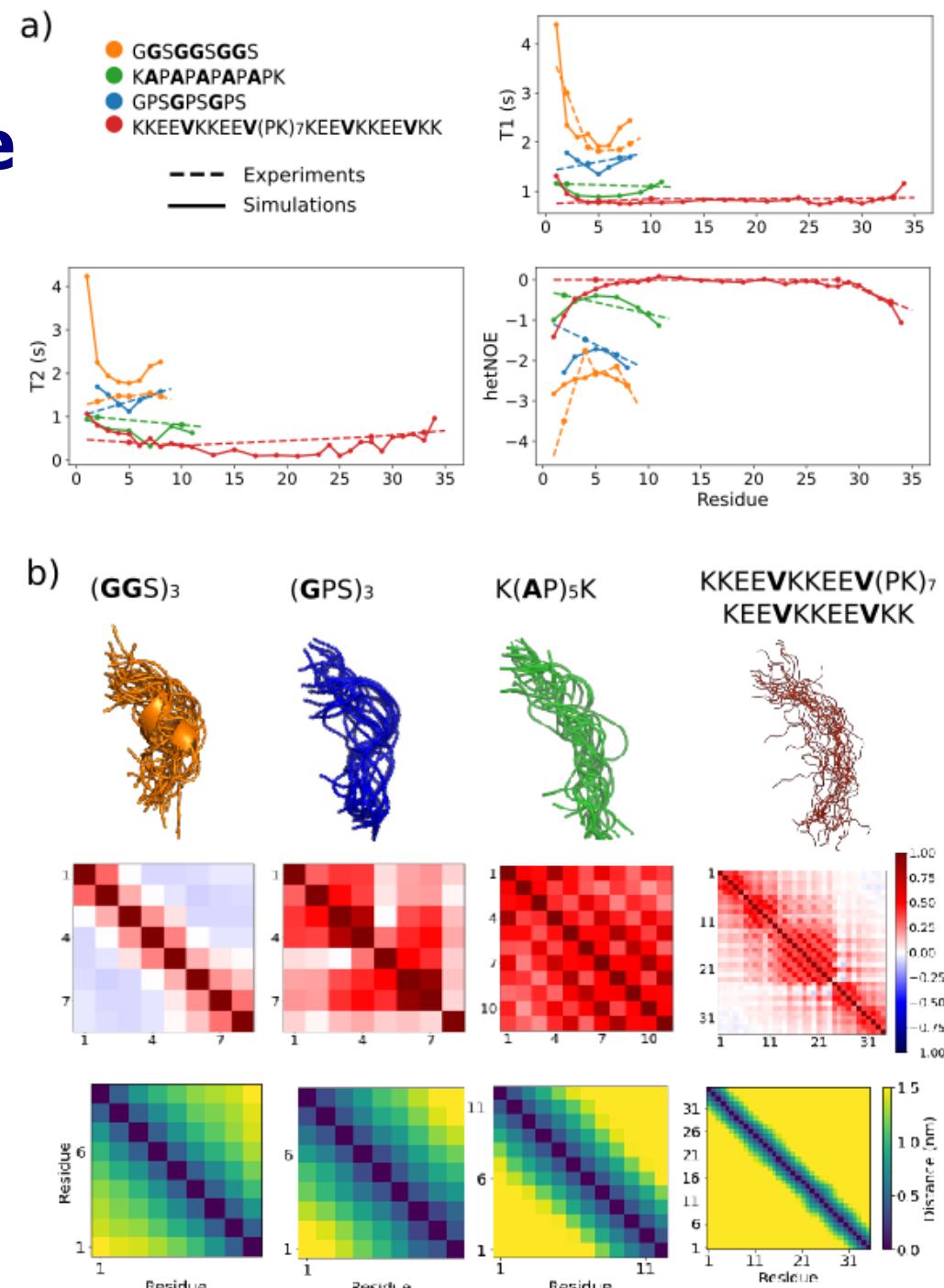
VTT



- Standard force fields (Amber ff99SB-ILDN and CHARMM36m) give too collapsed conformational ensemble
- Amber ff03ws is significantly better

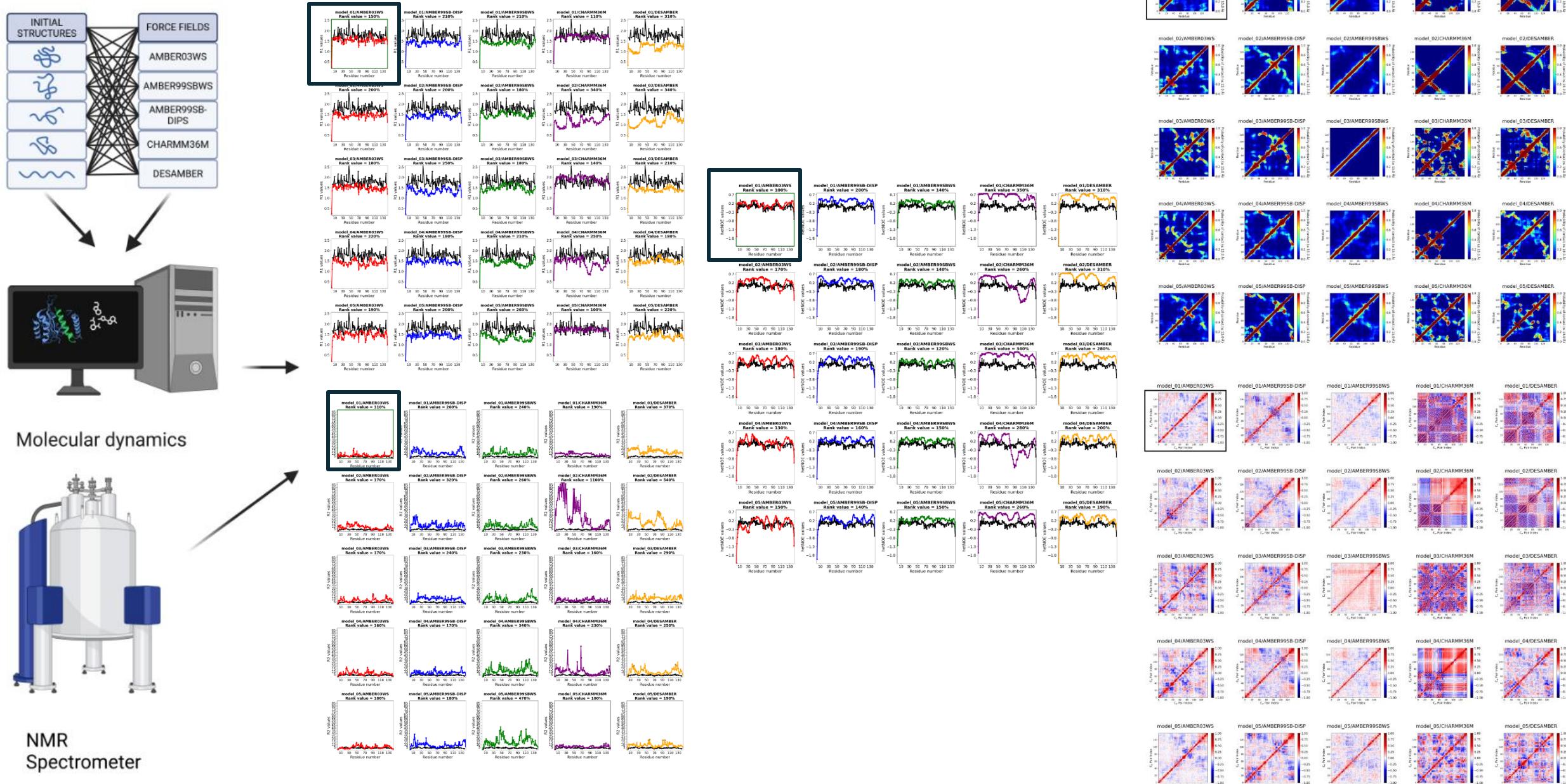
Backbone ^{15}N spin relaxation are sensitive to backbone rigidity

Efstathia Manzari, Cajsa Malm, Ricky Nencini,
Amanda Sandelin, Ollila (unpublished)



QEBSS: Quality Evaluation Based Simulation Selection for IDPs

Cajsa Malm, Mykhailo Girych, Ollila (<https://doi.org/10.26434/chemrxiv-2025-m5m0p>)

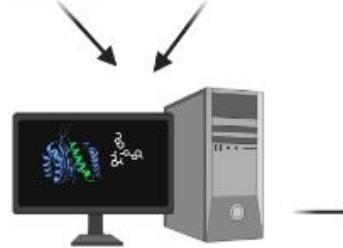
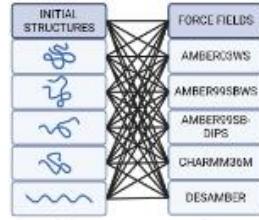


Ensembles of IDPs with different backbone rigidities resolved with QEBSS

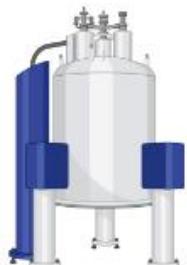
VTT

Cajsa Malm, Mykhailo Girych, Ollila (<https://doi.org/10.26434/chemrxiv-2025-m5m0p>)

Generate diverse set of simulations and evaluate against NMR data

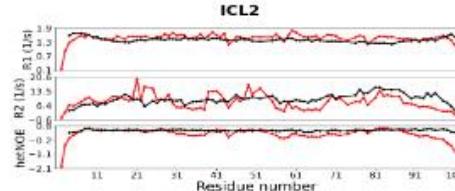
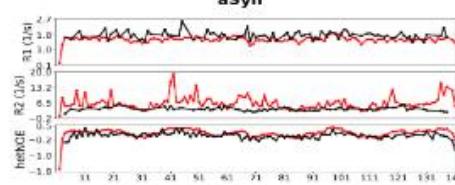
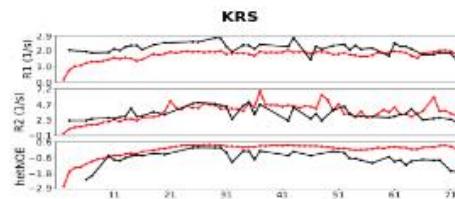
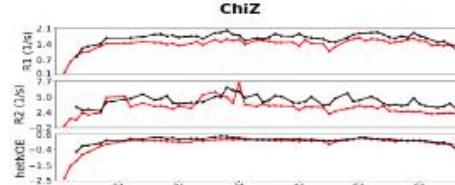


Molecular dynamics

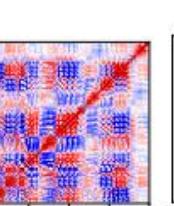
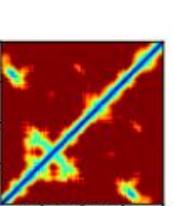
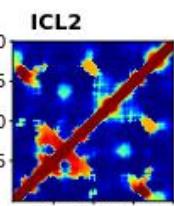
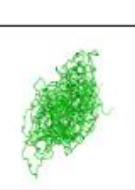
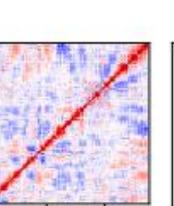
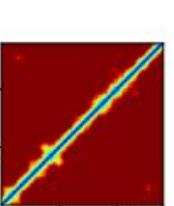
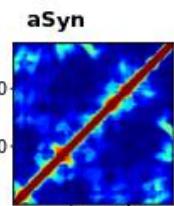
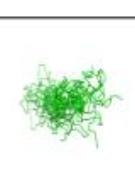
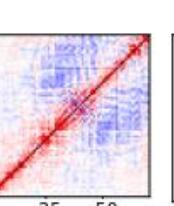
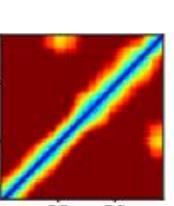
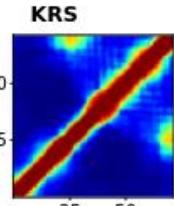
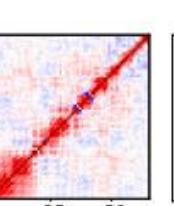
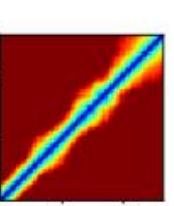
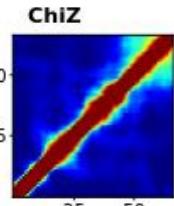
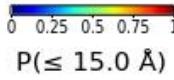


NMR Spectrometer

Average relaxation times of simulations in good agreement with experiment



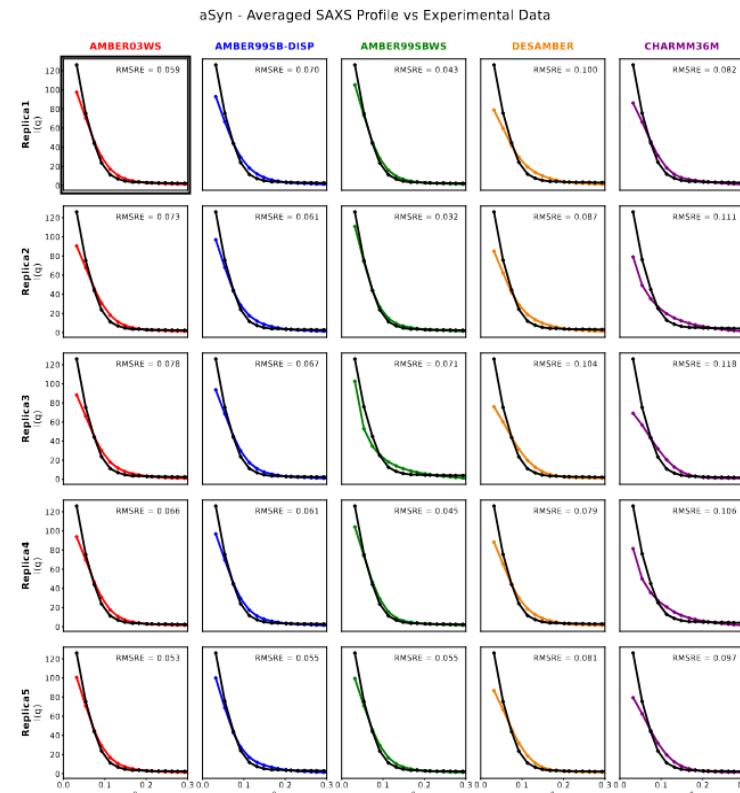
Analysis of distance, contact and backbone correlation maps of selection average



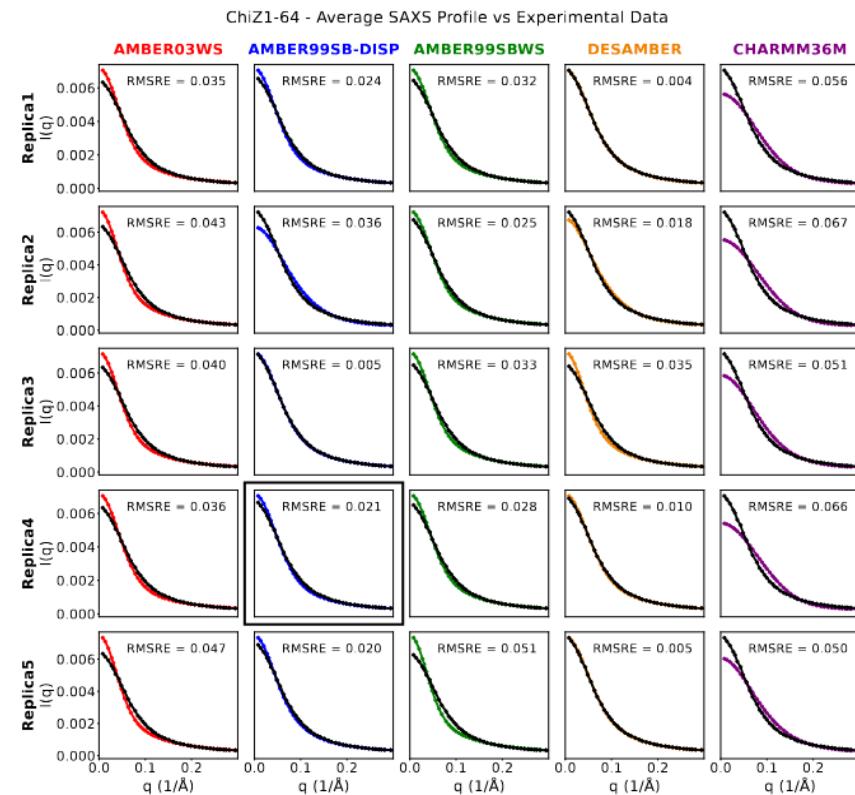
QEBSS cross validated against SAXS and PRE data

Spin relaxation times seem to be more sensitive to the ensemble than SAXS or chemical shifts

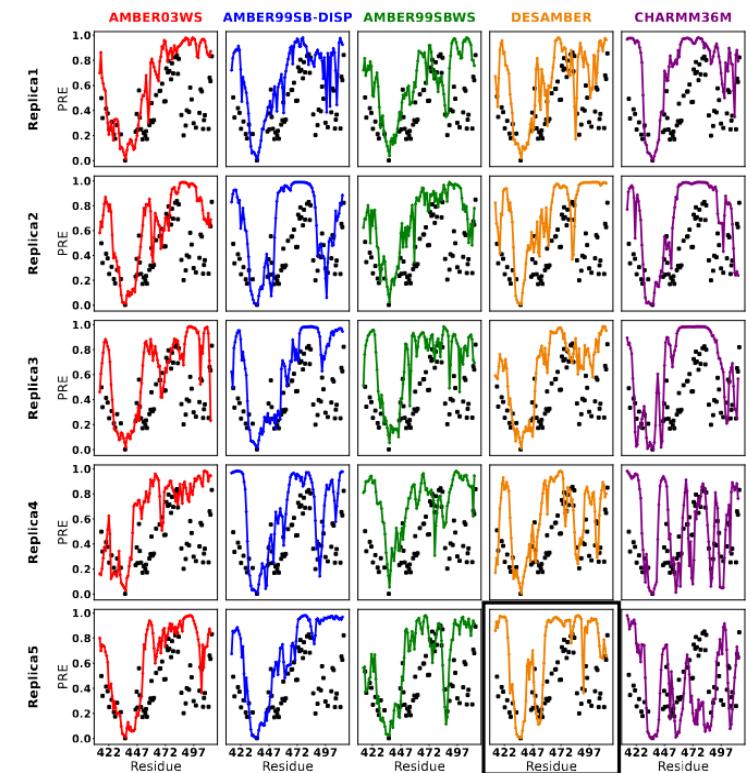
aSyn



ChiZ

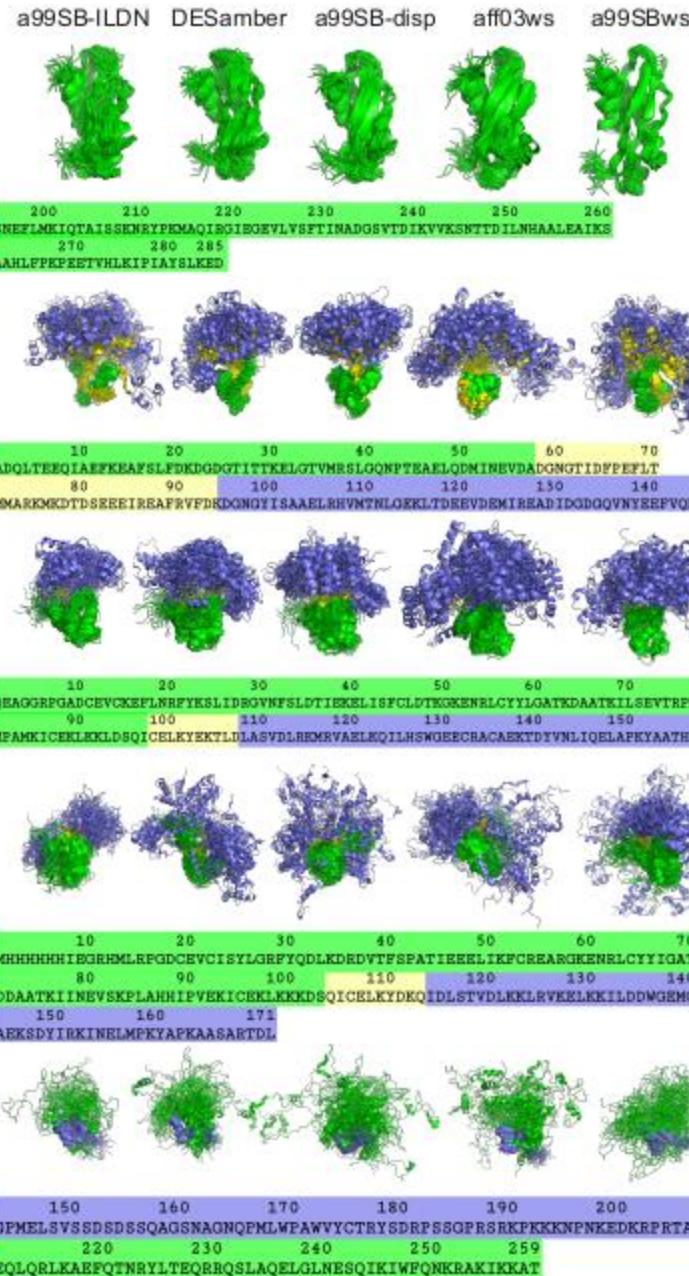


ICL2 (PRE)



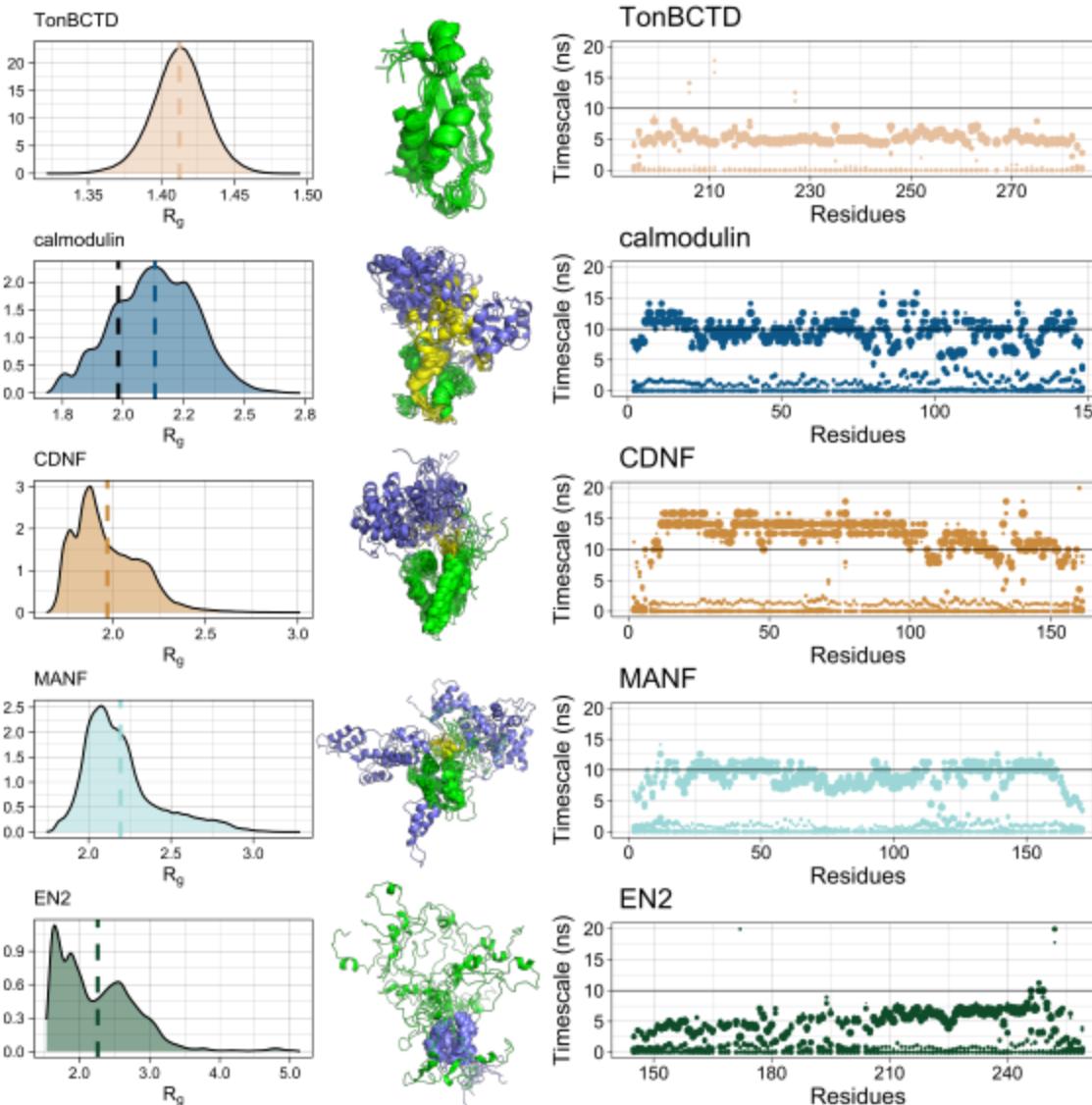
Cajsa Malm, Mykhailo Girych, Ollila (<https://doi.org/10.26434/chemrxiv-2025-m5m0p>)

Two domain protein dynamical ensembles solved with QEBSS



Amanda Sandelin, Nencini, Yasar, Fudo, Stratoulias, Kajander and Ollila
(<https://doi.org/10.26434/chemrxiv-2024-h3pmt-v2>)

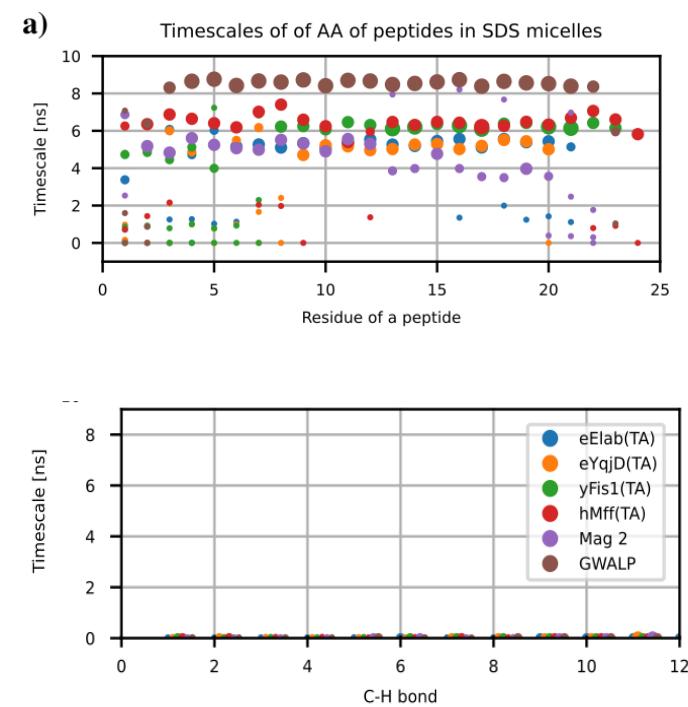
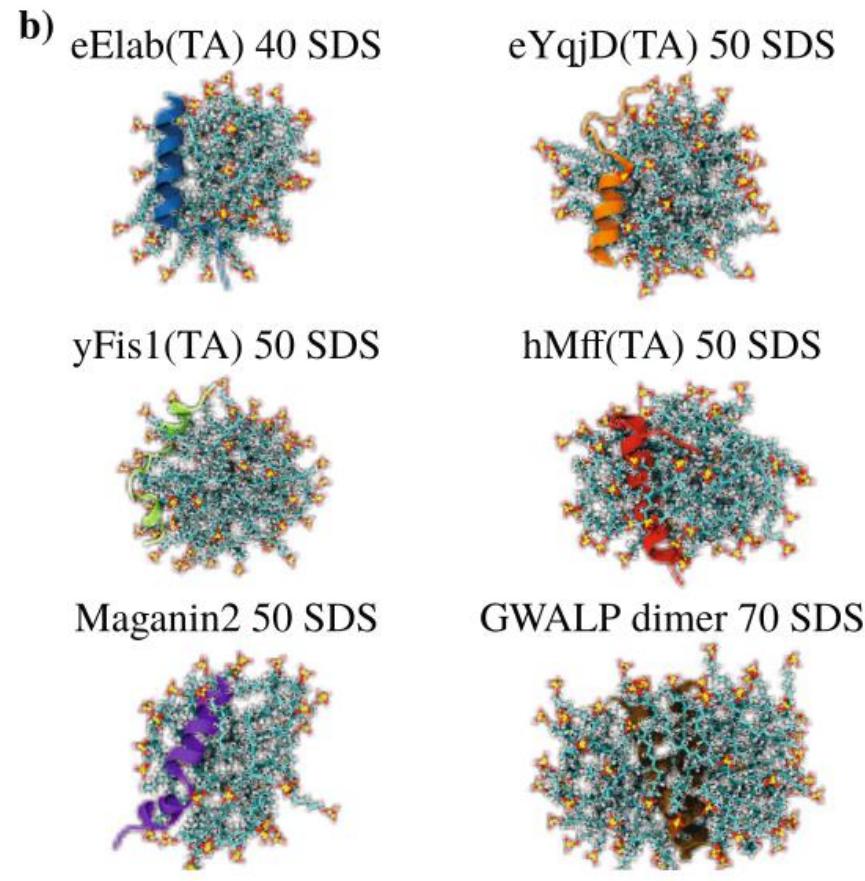
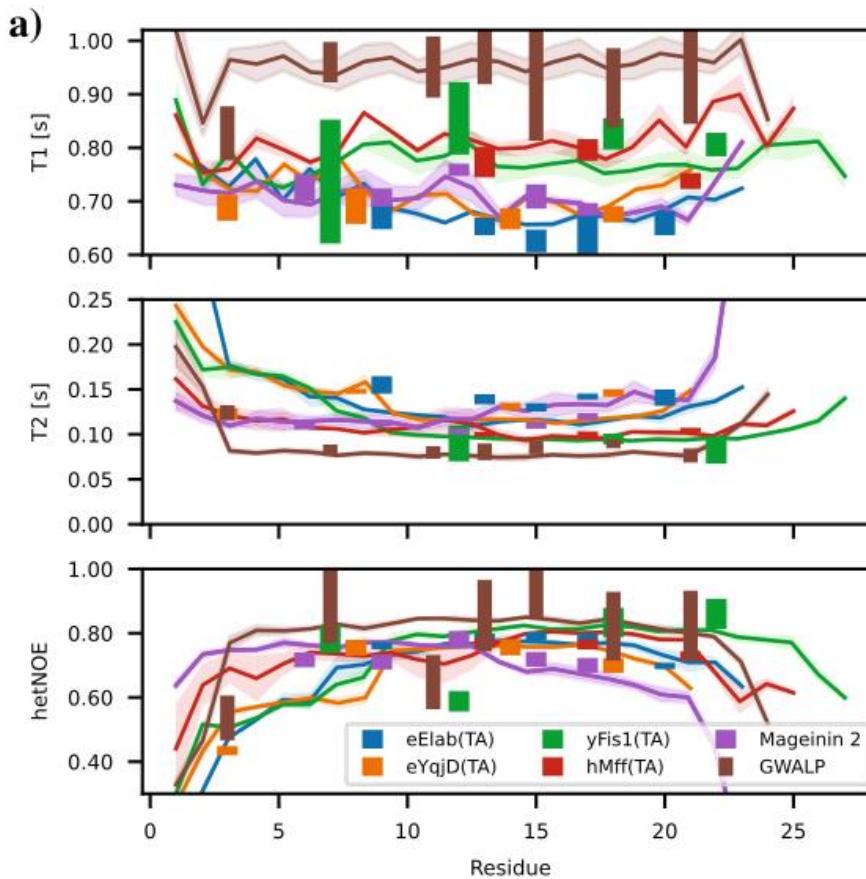
QEBSS



Dynamic landscape of peptides in micelles

Ricky Nencini, Morgan Regnier, Sofia Backlund, Efstathia Manzari, Cory Dunn, Ollila (Comm. Chem. 2024)

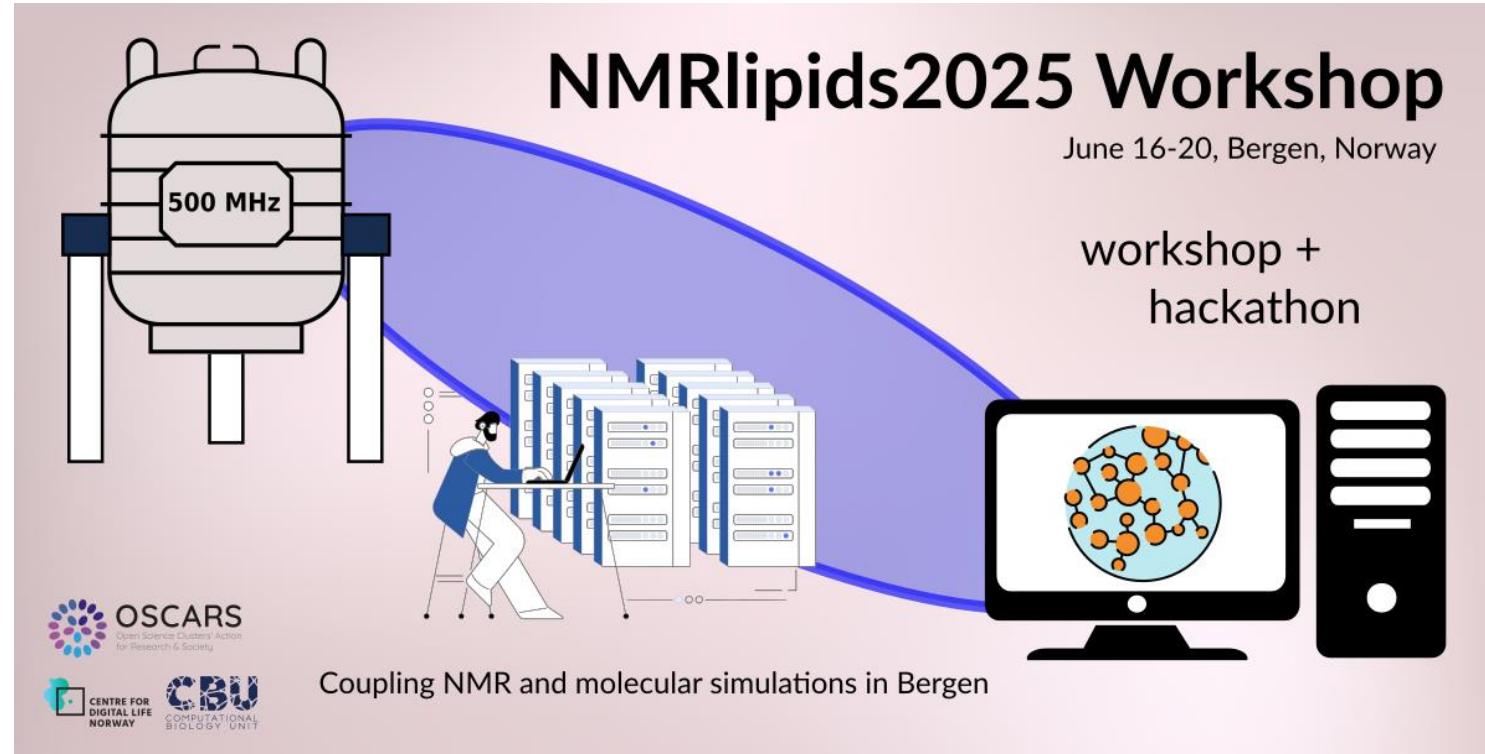
Peptides rotate in viscous environment of surfactants independently on individual surfactant molecules



FAIRMD - Extending the NMRLipids Databank to IDPs

VTT

<https://oscars-project.eu/projects/fairmd-disorder-order-streamlining-biomolecule-simulation-re-use-fair-nmrlipids-database>



<https://www.digitallifenorway.org/events/nmrlipids-workshop-2025.html>

THANK YOU FOR YOUR ATTENTION!

Practicalities and hands on session

NMRlipids Databank

- Databank containing quality evaluated molecular dynamics (MD) simulations of lipid bilayers with atomic resolution
- Overlay databank with programmatic access
- Initiated from the NMRlipids project (nmrlipids.blogspot.fi)
- Open for submissions

Documentation: <https://nmrlipids.github.io/>
GitHub: <https://github.com/NMRLipids/Databank>
GUI: www.databank.nmrlipids.fi

NMRlipids databank: Expected applications



Force field evaluation: What is the best force field for my application?



Analysis of bilayer properties from large datasets: For example, correlations between lipid bilayer properties.



Reference simulations: For example, reference pure bilayer simulations for membrane-protein interaction studies.



Exercise and example for sharing simulation data: “PDB” for simulations?

NMRlipids databank: Key features



Overlay databank: NMRlipids databank contains links to the raw data, but data is stored in public repositories (currently in Zenodo).



Programmatic access: Programmatic access enables flexible analysis of the content (<https://nmrlipids.github.io/>).



Quality evaluation: Automatic quality evaluation is made against available NMR order parameters and x-ray scattering form factors, and the results are stored in the databank.

Overlay structure of the NMRLipids databank

For example

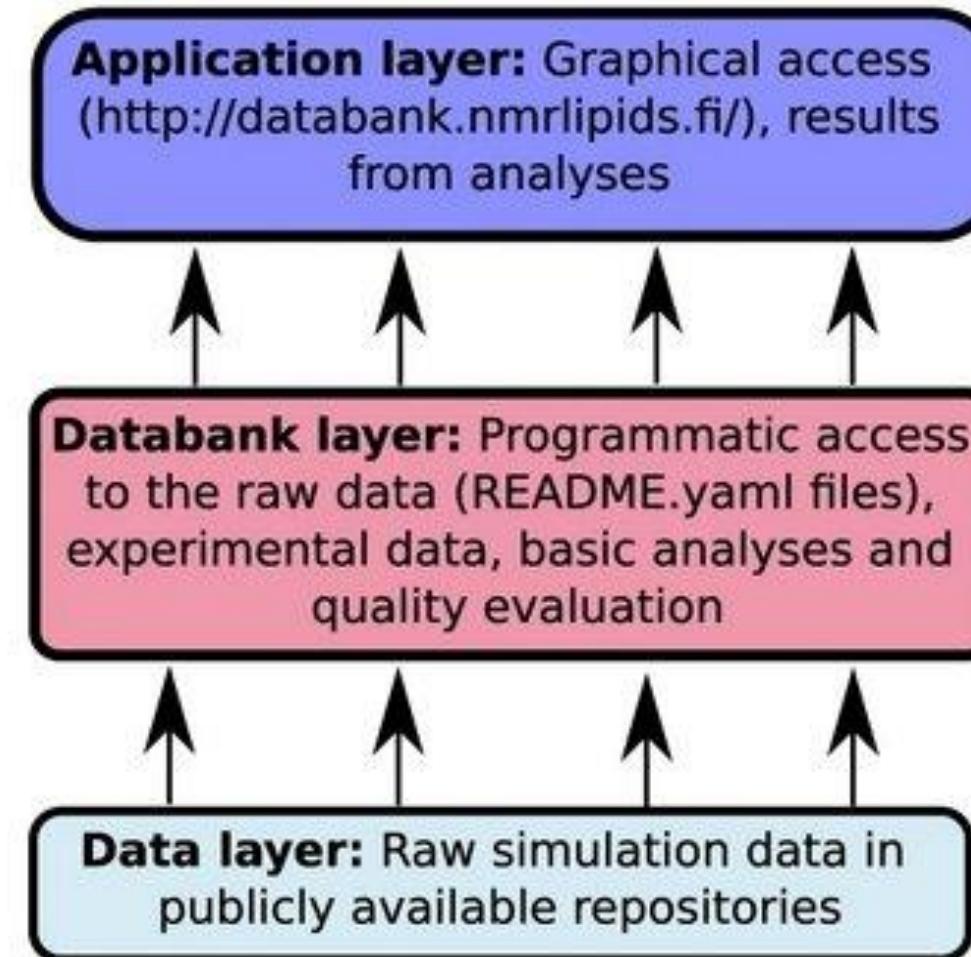
GUI: <https://databank.nmrlipids.fi/>
API: <https://nmrlipids.github.io>

The "NMRLipids databank"

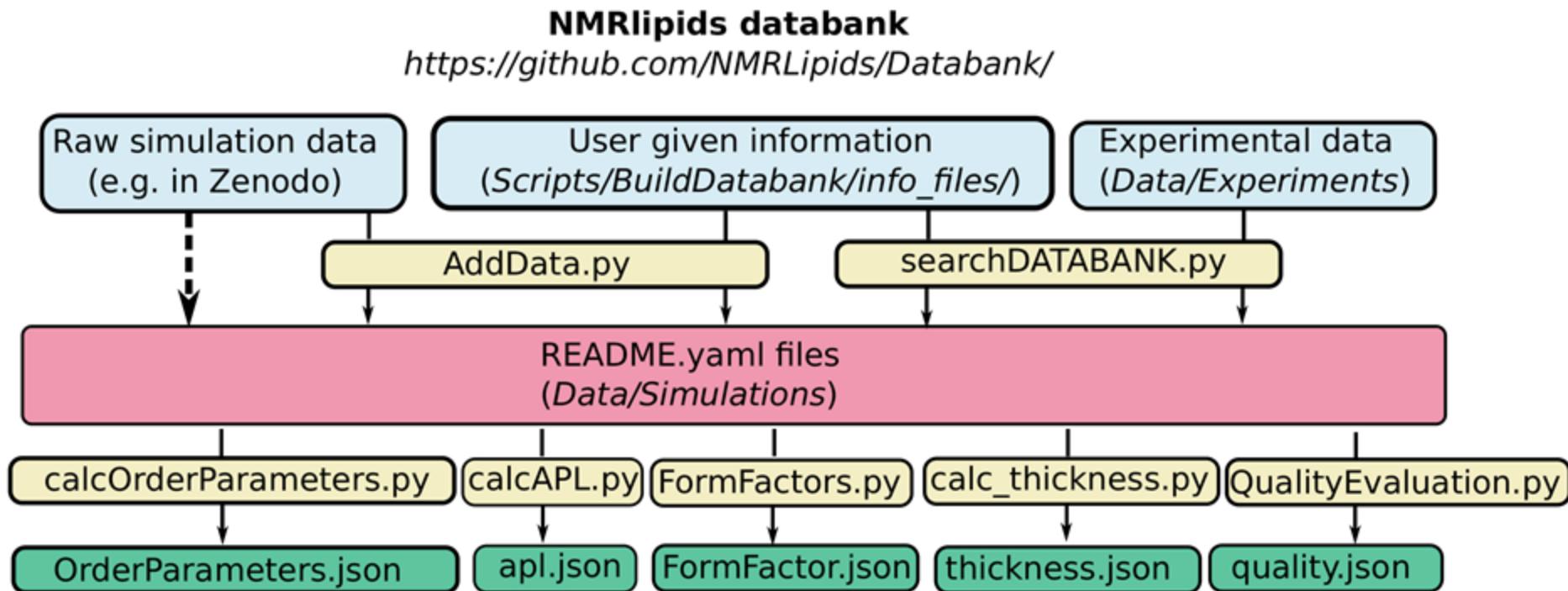
<https://github.com/NMRLipids/Databank>

Currently Zenodo

(CERN hosted file sharing service)



Overlay structure of the NMRLipids databank



Simulations in the NMRLipids databank

VTT

- Each subfolder in <https://github.com/NMRLipids/Databank/tree/main/Data/Simulations> corresponds one simulation
- Folders are named according to the hash of trajectory and tpr file
- **README.yaml in each folder contains all the relevant information on the simulation!**

HANDS ON EXERCISE 1:

- Investigate fields README.yaml files describing the simulations. Contents described in here: <https://nmrlipids.github.io/listOfFiles.html> and <https://nmrlipids.github.io/READMEcontent.html>
- Plot basic properties of selected simulations using notebook (preferably in Colab) <https://github.com/NMRLipids/databank-template/blob/main/scripts/plotSimulation.ipynb>

Experimental data in the NMRLipids databank

- Each folder in [*https://github.com/NMRLipids/Databank/tree/main/Data/experiments*](https://github.com/NMRLipids/Databank/tree/main/Data/experiments) corresponds one experimental dataset
- Folders are named according to the DOI of experimental data
- **All the relevant information to connect experimental and simulation datasets are found from README.yaml files within these folders**

Quality evaluation in the NMRlipids databank

- Simulations (*Data/Simulations/*) are paired with experimental data (*Data/experiments*) when:
 - - temperature is the same within ± 2 degrees
 - - molar concentrations are within ± 5 percentage units
 - - counterions are the same
- If a pair is found, the path to the experimental data is added to the “EXPERIMENT” key in the simulation README.yaml (*searchDATABANK.py*)
- Quality measures are calculated for simulations paired with experimental data (*QualityEvaluation.py*)
- Quality is defined against C-H bond order parameters from NMR and x-ray scattering form factors

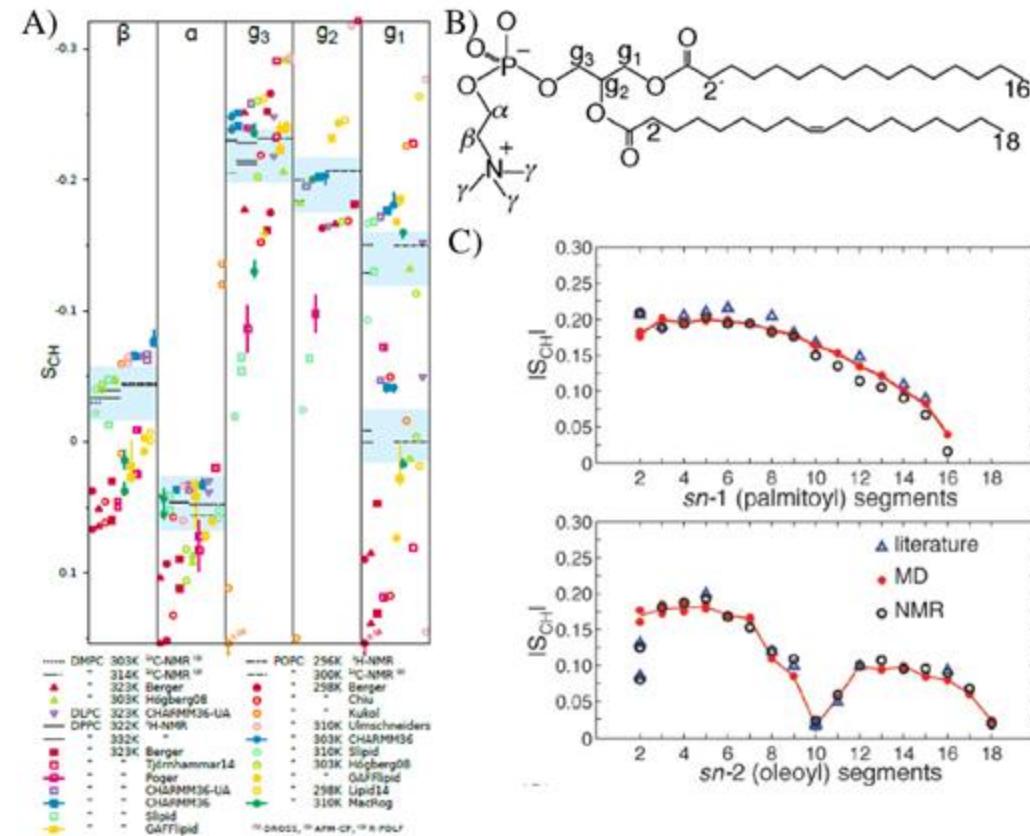
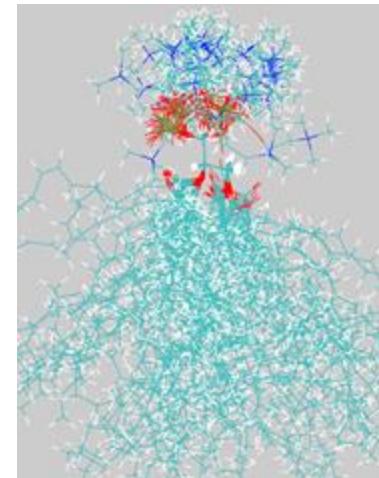
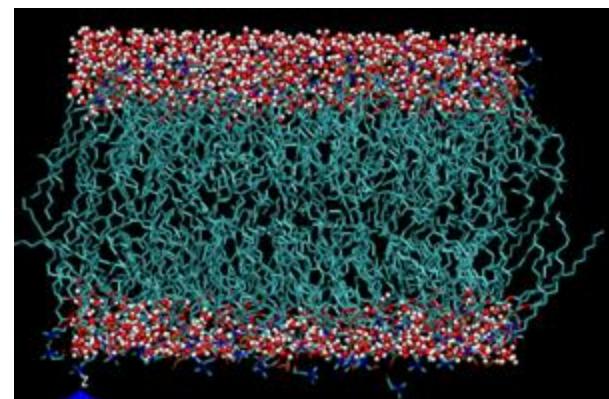
Quality evaluation: Order parameters

- Order parameters are sensitive to the conformational ensembles of individual lipids
- Acyl chain order correlates with lipid packing (area per lipid)

$$S_{CH} = \frac{1}{2} \langle 3 \cos^2 \theta - 1 \rangle$$

θ = angle between C-H bond and membrane normal
 $\langle . \rangle$ = average over conformational ensemble

↔ ~5nm-10nm

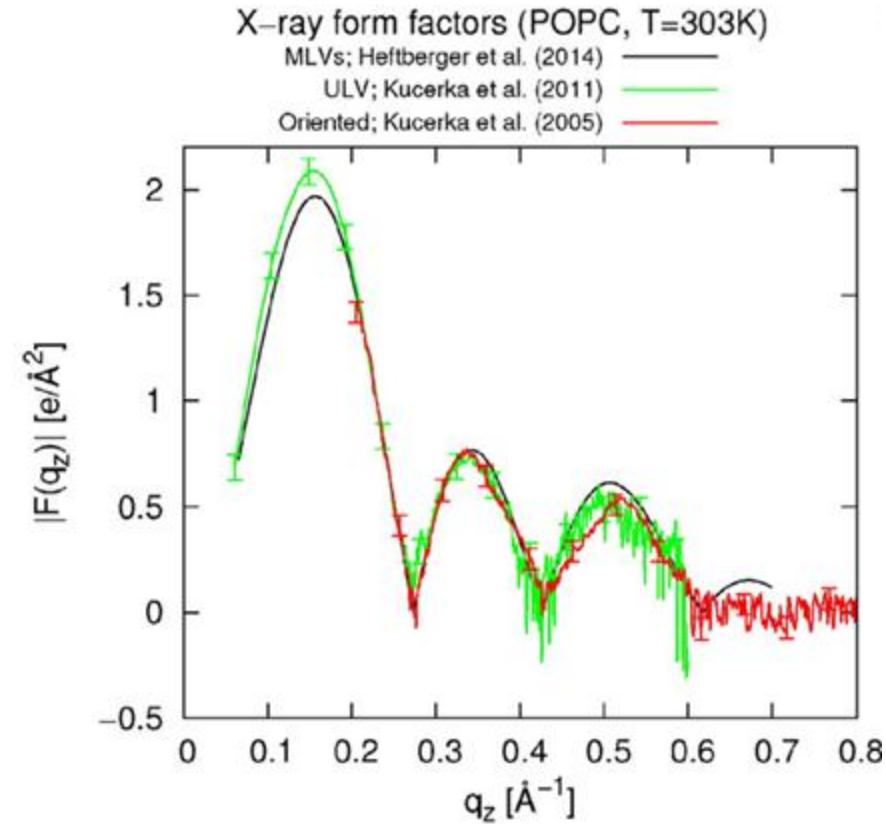
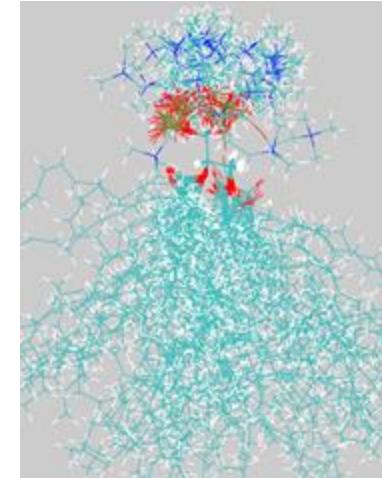
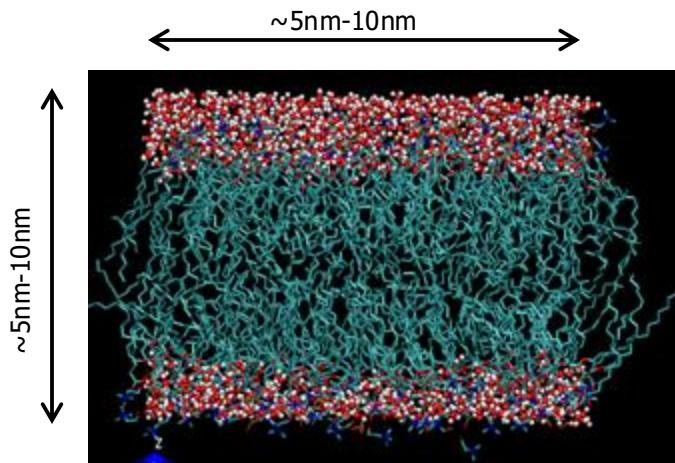


Quality evaluation: Form factor

- Scattering form factors are sensitive to membrane dimensions (electron density profile, thickness and area per molecule)

$$F(q) = \int_{-D/2}^{D/2} \Delta\rho_e(z) \cos(zq_z) dz$$

$\rho_e(z)$ = electron density difference with respect to bulk water
 $D/2$ = beginning of bulk water region



NMRlipids quality measure of lipid conformational ensemble

- Probability for each order parameter to locate within experimental error bars calculated from the Student's t-distribution

$$P = f \left(\frac{S_{\text{CH}} - (S_{\text{exp}} + \Delta S_{\text{exp}})}{s / \sqrt{n}} \right) - f \left(\frac{S_{\text{CH}} - (S_{\text{exp}} - \Delta S_{\text{exp}})}{s / \sqrt{n}} \right)$$

P = probability for simulation value to locate within experimental error bars

S_{exp} = experimental order parameter

ΔS_{exp} = error bars of experimental order parameter (0.02)

S_{CH} = order parameter from simulation

s = standard deviation of order parameter from simulation

n = number of lipids in the simulation

f = Student's t-distribution

NMRlipids quality measure of lipid conformational ensemble

- Quality of fragments of each lipid estimated by averaging over order parameters
- Divided by the fraction of available order parameters to penalize from missing data

$$P^{\text{frag}}[\text{lipid}] = \frac{\langle P[\text{lipid}] \rangle_{\text{frag}}}{p_{\text{frag}}[\text{lipid}]}$$

frag = headgroup, *sn*-1 chain, *sn*-2 chain, or total (all order parameter in a molecule)

p_{frag} = fraction of experimentally available order parameters for the fragment

- The overall quality of fragments in each simulation is estimated by averaging over lipids weighted by molar fraction

$$P^{\text{frag}} = \sum_{\text{lipid}} \chi_{\text{lipid}} P^{\text{frag}}[\text{lipid}]$$

X_{lipid} = molar fraction of a lipid in the simulation

- Form factor quality evaluated as in SIMtoEXP program

$$\chi^2 = \frac{\sqrt{\sum_{i=1}^{N_q} (|F_s(q_i)| - k_e |F_e(q_i)|)^2 / (\Delta F_e(q_i))^2}}{\sqrt{N_q - 1}}$$

F_s = form factor from a simulation

F_e = form factor from experiment

ΔF_e = error of form factor from experiment

(1,...,N_q) = Experimental datapoints

$$k_e = \frac{\sum_{i=1}^{N_q} \frac{|F_s(q_i)||F_e(q_i)|}{(\Delta F_e(q_i))^2}}{\sum_{i=1}^{N_q} \frac{|F_e(q_i)|^2}{(\Delta F_e(q_i))^2}}$$

HANDS ON EXERCISE 2:

- Investigate rankings of simulations in NMRLipids databank against experiments using notebook (preferably in Colab):
<https://github.com/NMRLipids/databank-template/blob/main/scripts/plotQuality.ipynb>
- Plot and investigate highly and poorly ranked simulations using notebook (preferably in Colab) <https://github.com/NMRLipids/databank-template/blob/main/scripts/plotSimulation.ipynb>

- Properties analyzed from all simulations and stored to the databank ([Data/Simulations](#)):
 - - C-H bond order parameters
 - - X-ray scattering form factors
 - - Area per lipid as a function of time
 - - Membrane thickness from intersection of water and lipid densities
- **Further analyses can be done with Python**

HANDS ON EXERCISE 3:

- Investigate template for API applications of analyzing NMRLipids databank:
<https://github.com/NMRLipids/databank-template/blob/main/scripts/template.ipynb>

Demonstrates the usage of API by three examples. **1)** Selects a random simulation and prints the related databank content in human readable format. **2)** Shows the readily analyzed properties for the selected random simulation (area per lipid, membrane thickness, relative equilibration times, X-ray scattering form factors, and C-H bond order parameters). **3)** Selects a random simulation with the trajectory size below 100Mb and calculates P-N vector angle with respect to membrane normal for all lipids for which P and N atoms are available in headgroup

HANDS ON EXERCISE 4:

- Train a machine learning model that predicts area per lipid for membranes from a composition using the notebook (preferably using Colab version):
https://github.com/NMRLipids/DatabankExercises/blob/master/cecam_290525/APLpredictor.ipynb
- Published in this article: <https://doi.org/10.1038/s41467-024-45189-z>
- Compare results from different models trained. What is the best model and why?

HANDS ON EXERCISE 5 (extra):

- Train a machine learning model that predicts lipid bilayer electron density from an x-ray scattering form factor composition using the notebook (preferably using Colab version):
- https://github.com/NMRLipids/databank-template/blob/main/scripts/FormFactor_to_TotalDensity_model_training_walkthrough.ipynb
- Use the developed model to calculate electron density from experimental form factor

Practical implementation of exercises

- Notebooks and these slides available at:
 - <https://nmrlipids.github.io/exampleAndTutorials.html>
 - <https://github.com/NMRLipids/databank-template/tree/main/scripts>
 - https://github.com/NMRLipids/DatabankExercises/tree/master/cecam_290525
- Using Colab (links on top of notebooks) is recommended
- Local usage should be also possible
 - Follow instructions here: <https://github.com/NMRLipids/Databank>
 - Set up kernel for Jupyter-notebook:
 - pip install ipykernel
 - python -m ipykernel install --user --name databank --display-name "Python (myenv)"

HANDS ON EXERCISE 1: Familiarise with the NMRLipids databank content

- Investigate content of README.yaml files describing the simulations:
 - <https://nmrlipids.github.io/listOfFiles.html>
 - <https://nmrlipids.github.io/READMEcontent.html>
- Plot basic properties of selected simulations using notebook (preferably in Colab):
 - <https://github.com/NMRLipids/databank-template/blob/main/scripts/plotSimulation.ipynb>

HANDS ON EXERCISE 2: Investigate highly and lowly ranked simulations

- Investigate rankings of simulations in NMRLipids databank against experiments using notebook (preferably in Colab):
 - <https://github.com/NMRLipids/databank-template/blob/main/scripts/plotQuality.ipynb>
- Plot and investigate highly and poorly ranked simulations using notebook (preferably in Colab):
 - <https://github.com/NMRLipids/databank-template/blob/main/scripts/plotSimulation.ipynb>

HANDS ON EXERCISE 3: Using NMRLipids databank for analyses

- Investigate template for API applications of analyzing NMRLipids databank:
<https://github.com/NMRLipids/databank-template/blob/main/scripts/template.ipynb>
Demonstrates the usage of API by three examples. **1)** Selects a random simulation and prints the related databank content in human readable format. **2)** Shows the readily analyzed properties for the selected random simulation (area per lipid, membrane thickness, relative equilibration times, X-ray scattering form factors, and C-H bond order parameters). **3)** Selects a random simulation with the trajectory size below 100Mb and calculates P-N vector angle with respect to membrane normal for all lipids for which P and N atoms are available in headgroup

HANDS ON EXERCISE 4: Using NMRLipids databank for machine learning

- Train a machine learning model that predicts area per lipid for membranes from a composition using the notebook (preferably using Colab version):
 - https://github.com/NMRLipids/DatabankExercises/blob/master/cecam_290525/APLpredictor.ipynb
- Published in this article: <https://doi.org/10.1038/s41467-024-45189-z>
- Compare results from different models trained. What is the best model and why?

HANDS ON EXERCISE 5 (extra): Using NMRLipids databank for machine learning

- Train a machine learning model that predicts lipid bilayer electron density from an x-ray scattering form factor composition using the notebook (preferably using Colab version)
 - https://github.com/NMRLipids/databank-template/blob/main/scripts/FormFactor_to_TotalDensity_model_training_walkthrough.ipynb
- Use the developed model to calculate electron density from experimental form factor