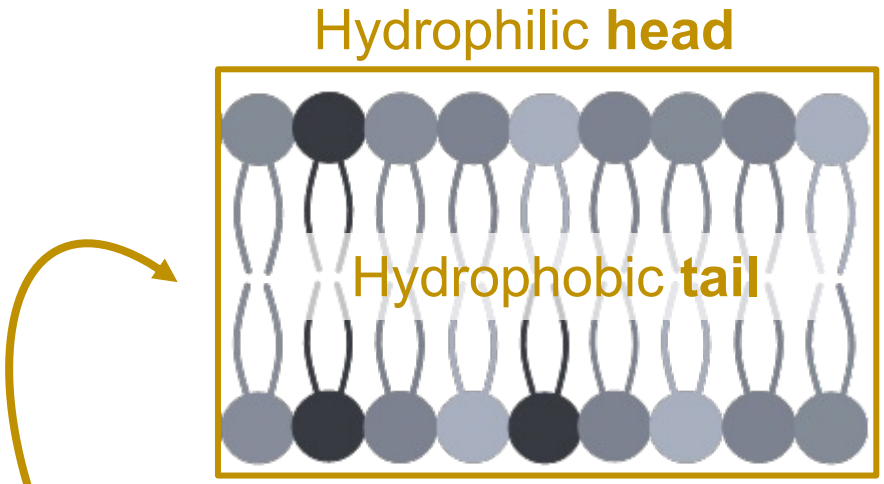
Pushing the boundaries of molecular simulations with LUMI: multi-million atoms simulations on a realistic dyed cell membrane

Charlotte Bouquiaux, Benoît Champagne, and Pierre Beaujean

Why cell membranes?

- **Fundamental functions** in the human body: compartmentalization, energy storage, external signaling, etc.
- **Large (chemical and compositional) diversity** of lipids, which impacts the structural and thus the function of the membrane.
- An abnormal composition of lipid is reported in various diseases such as cancer, and others

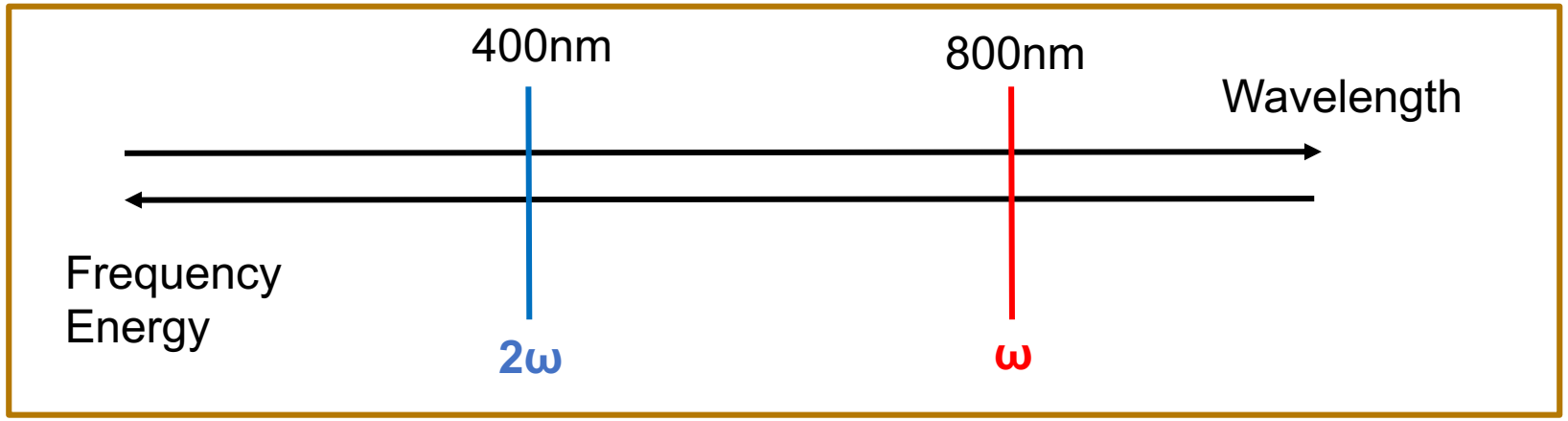


Schematic representation of a membrane bilayer

Change in head or tail → chemical diversity → change in **properties**

How to detect those changes? The Second Harmonic Generation (SHG) response

Spectroscopic method that exploits a nonlinear optical (NLO) phenomenon where the detected photons have twice the frequency of the incident ones. A dedicated **chromophore** is generally required to enhance the contrast.

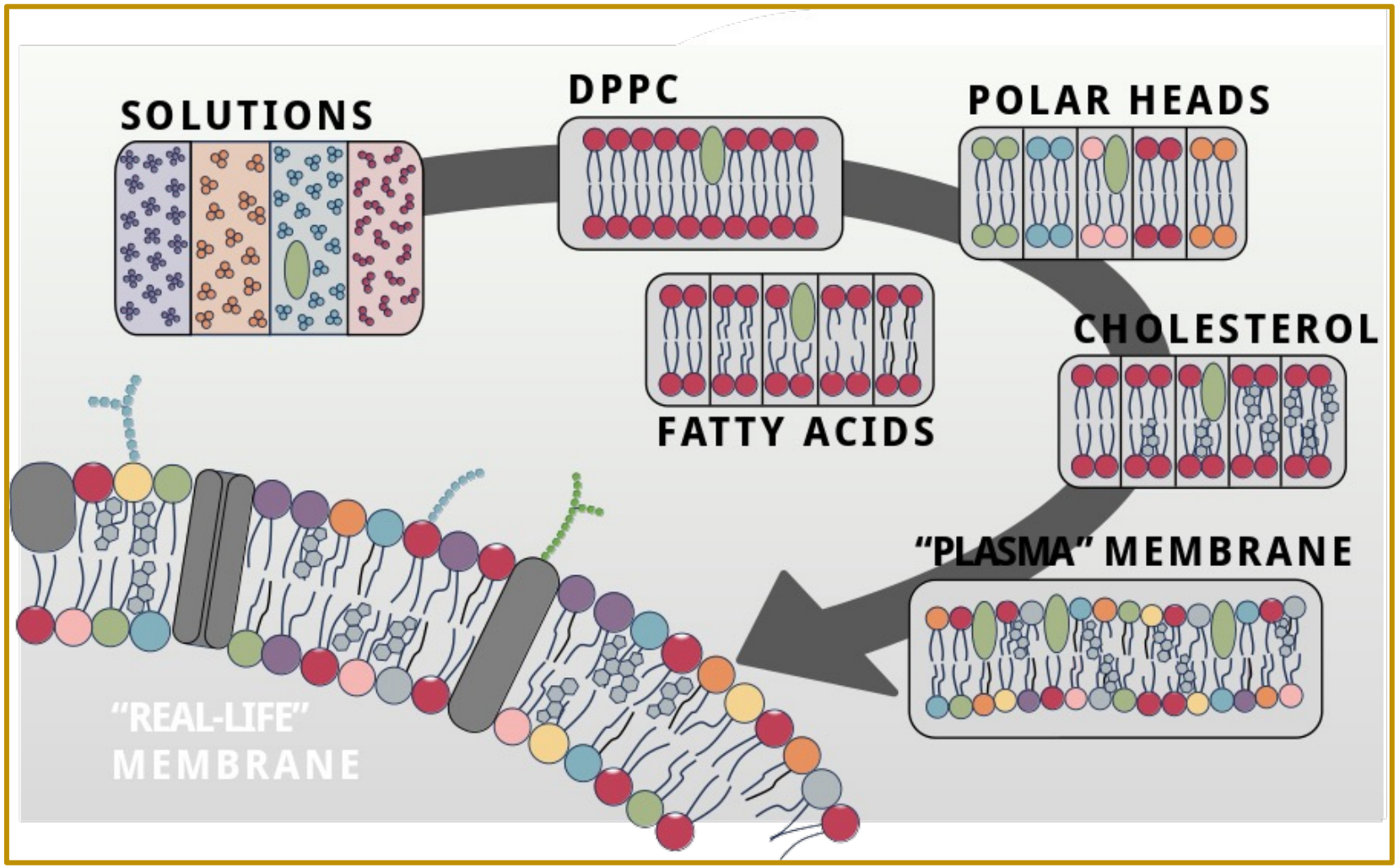


Energetic diagram of incident (red) and detected (blue) photons.

Advantage: less energetic incident photons → less tissue damages

Strategy

- Not a lot of theoretical studies for chromophores embedded in a membrane
- Lacking systematic studies on the impact of the environments
- Low-level calculations of the NLO response



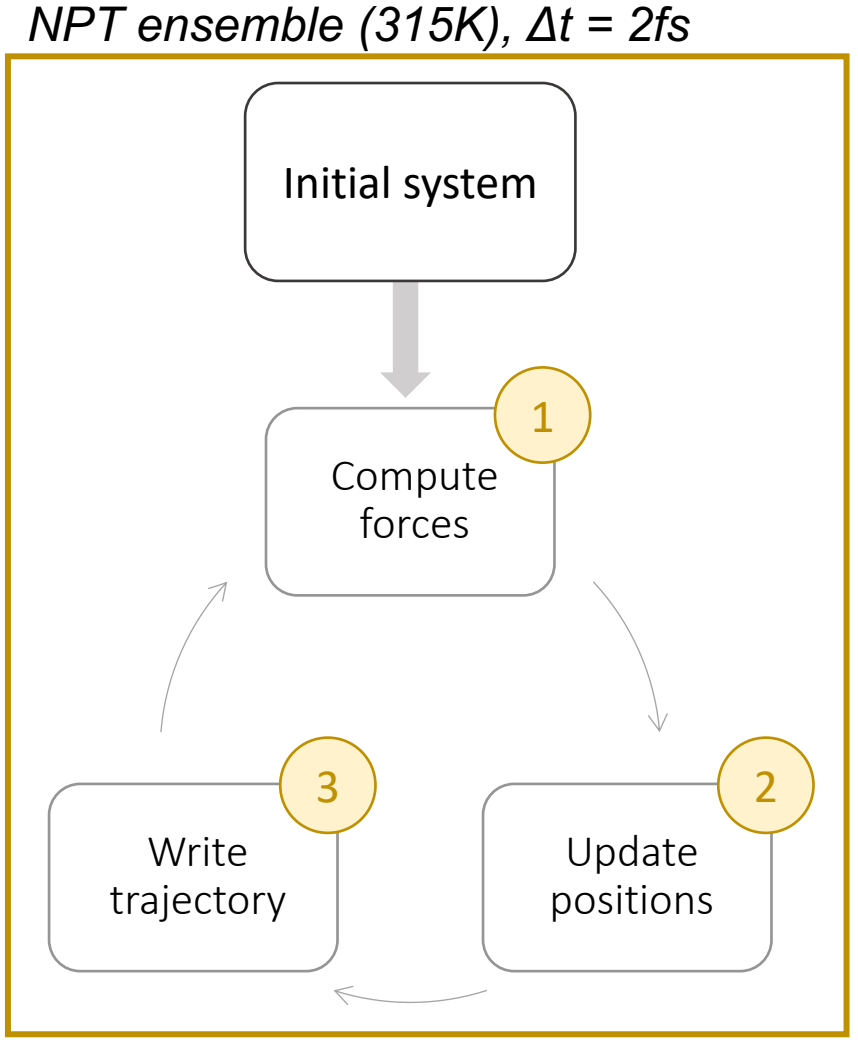
Strategy used to study the insertion of a chromophore in more and more realistic membranes.

Our approach: sequential MD/QC

1. **Molecular dynamics (MD) simulation:** solving the Newton equation of motion to get the evolution of the system over time**.** The energies and forces are obtained from tailor-made empirical force fields (i.e., “ball and spring” models).

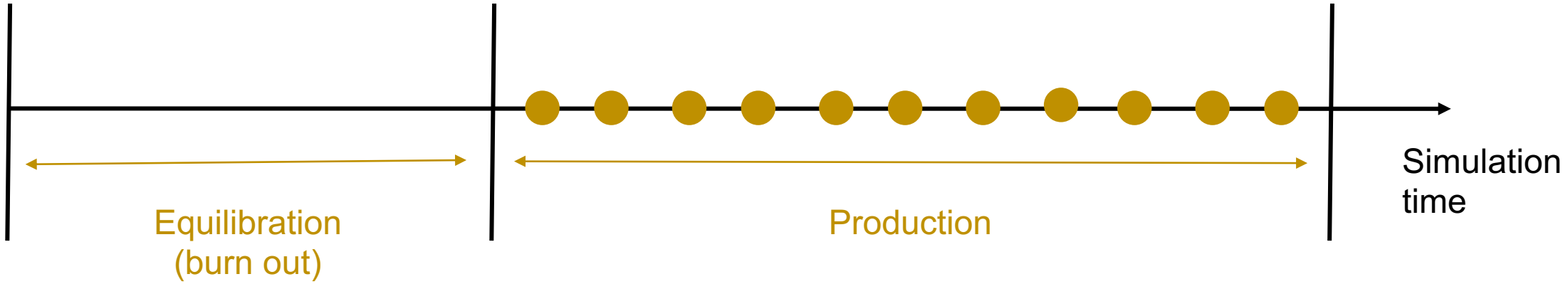
NAMD&VMD
Molecular Dynamics Packages

2. **Quantum chemistry (QC) calculations** to compute the SHG responses of the chromophores embedded in the lipidic environment.



Steps of a molecular dynamic simulation

Equilibration vs production



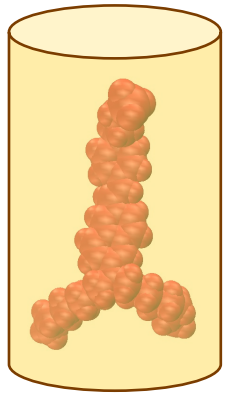
Since we do not necessarily start from thermal equilibrium, we need to wait a certain amount of time (**equilibration**) before we can extract data that are statistically relevant (**production**). Then, we cannot perform QC calculation on each frame, so we need long simulation time to obtain enough **uncorrelated** data (>200).

Trajectory

NAMD&VMD
Molecular Dynamics Packages

Trajectory over time

https://www.pinterest.com/pin/focus-on-spectacular-image-sequences--546624473494461488/



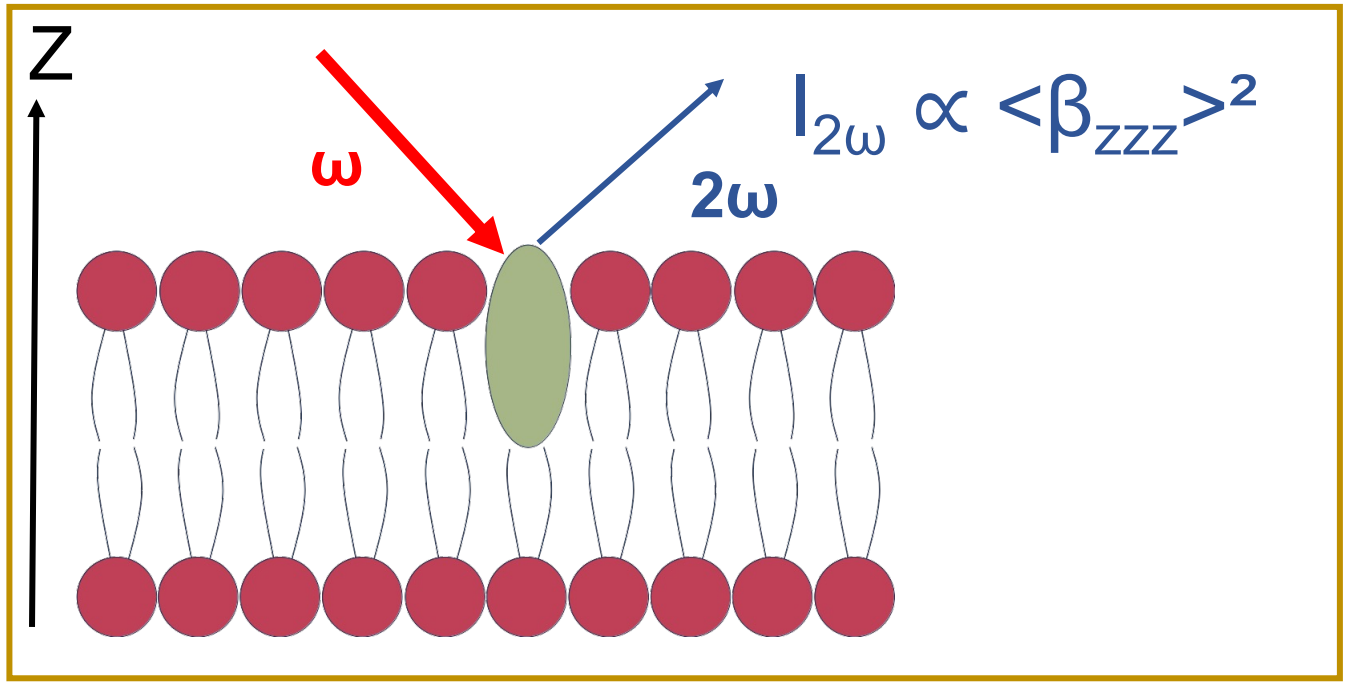
Snapshot: positions at a given time.

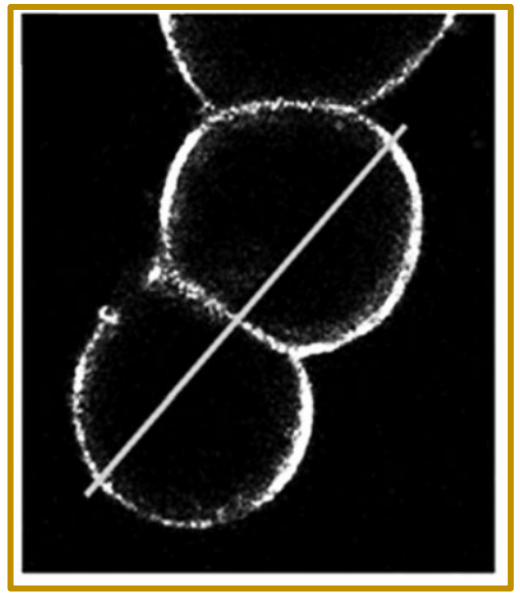
Geometries of the chromophore are extracted (with point charges to model the surroundings) and the NLO responses are computed (see next slide).

Nonlinear Optical (NLO) response of a chromophore in a membrane

*TD-DFT/ M06-2X/6-311+G**

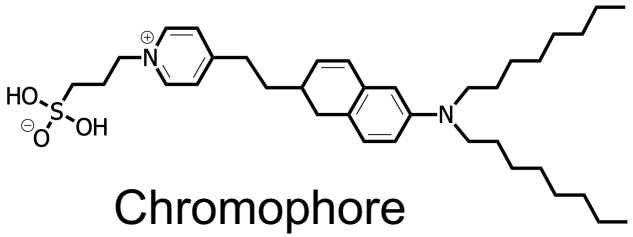
At the microscopic level, the NLO response of a molecule is described by a tensor, **β**, since it is orientation dependent. In particular, we are interested in the response normal to the surface of the bilayer (Z).



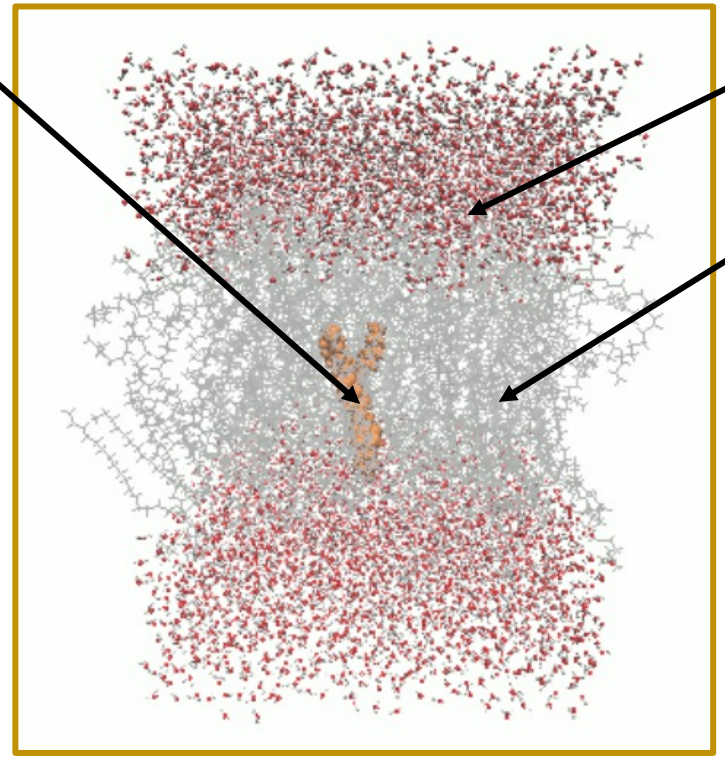


Yan, P. *et al. J. Am. Chem. Soc.* **2006**, 128, 11030–11031.

Model systems



Chromophore



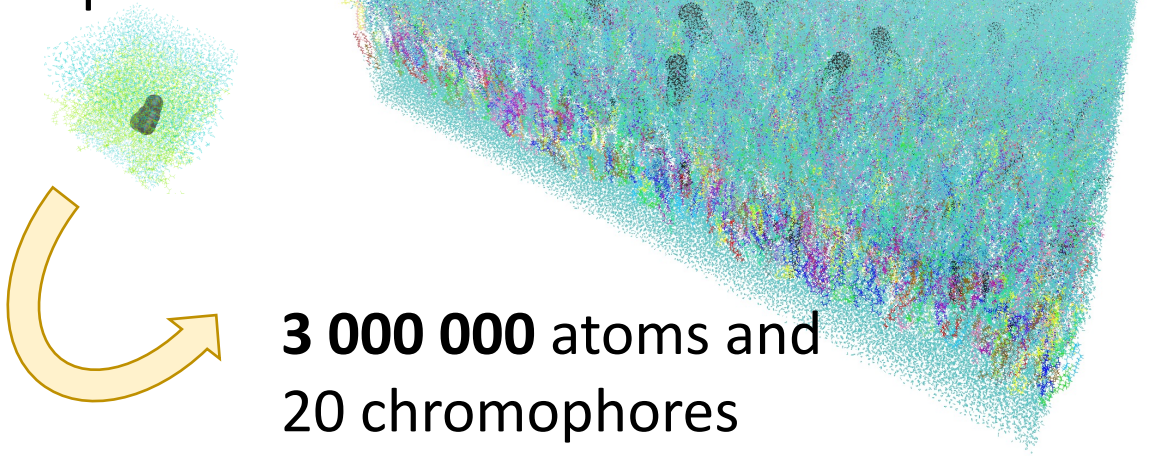
Model bilayer

- Choice of the force field, adaptation of its parameters for the chromophore (di-8-ANEPPS).
- Use of appropriate tools to build the system (otherwise, long equilibration time!)
- Model systems: 1 chromophore, 1 kind of lipids → ~35 000 atoms during 100 000 000 steps (200 ns)
- Then, complexification: 3 publications so far to explore the diversity of lipids
  - C. Bouquiaux, C. Tonnelé, F. Castet et B. Champagne *J. Phys. Chem. B* **2020**, *124*, 2101–2109
  - C. Bouquiaux, F. Castet et B. Champagne *J. Phys. Chem. B* **2021**, *125*, 10195–10212
  - C. Bouquiaux, F. Castet et B. Champagne *J. Phys. Chem. B* **2023**, 127, 528-541.

Towards a realistic system

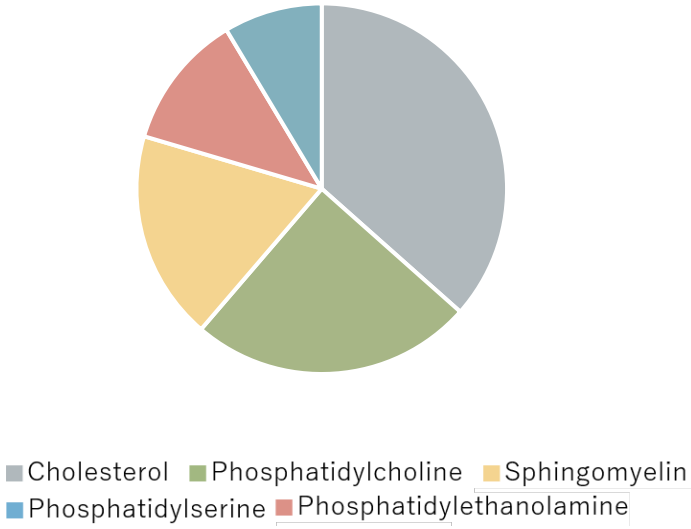
50 kinds of lipids (5 families) → Tier-2 & 1 are not sufficient.

35 000 atoms and 1 chromophore



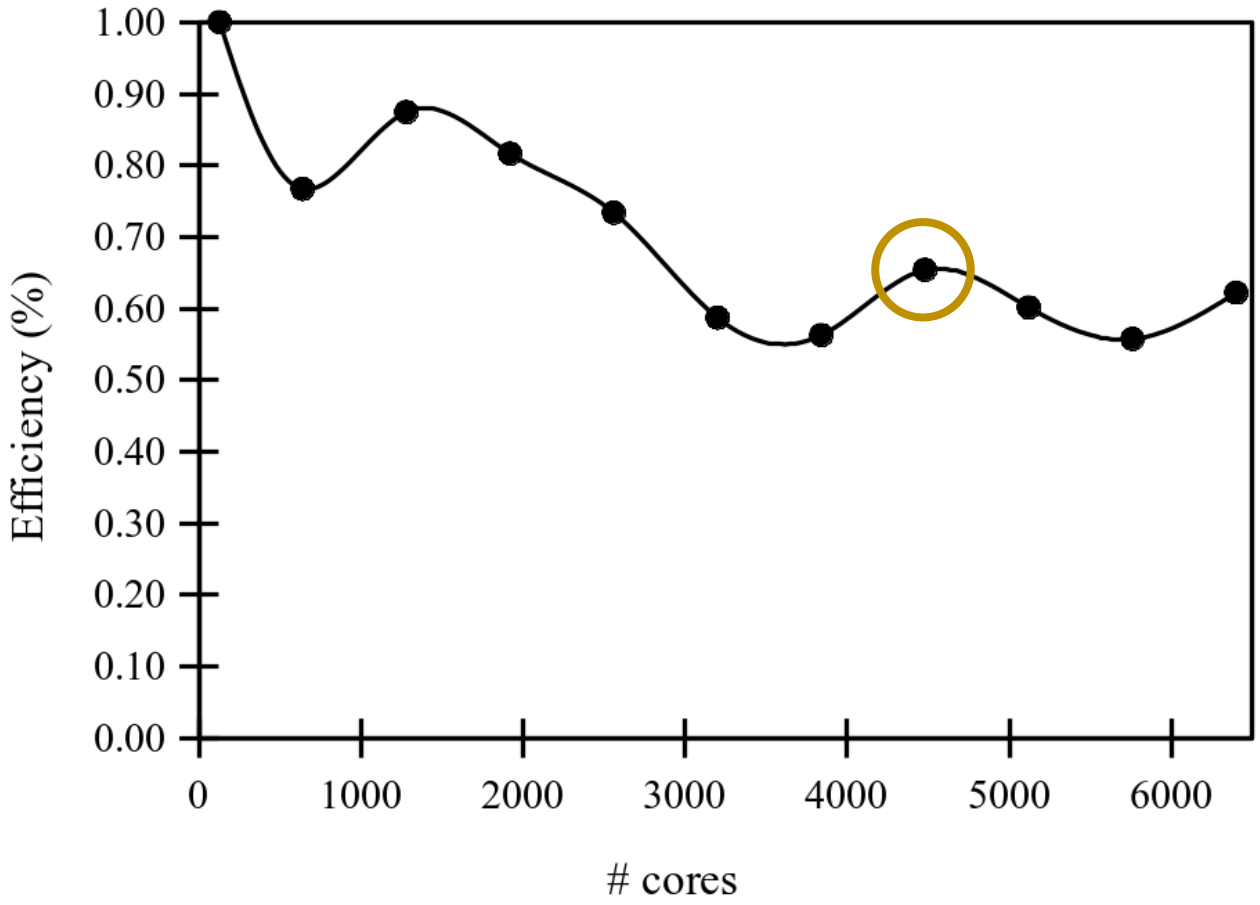
3 000 000 atoms and 20 chromophores

Model system (left) and realistic system (right)



Example of family of lipids found in a cell membrane

Scaling tests



- NAMD scales with the number of “patches”, which is the unit of spatial decomposition[a] distributed over the cores
- Here, using 35 nodes (4480 cores), we almost match the number of patches, for a 65% of efficiency

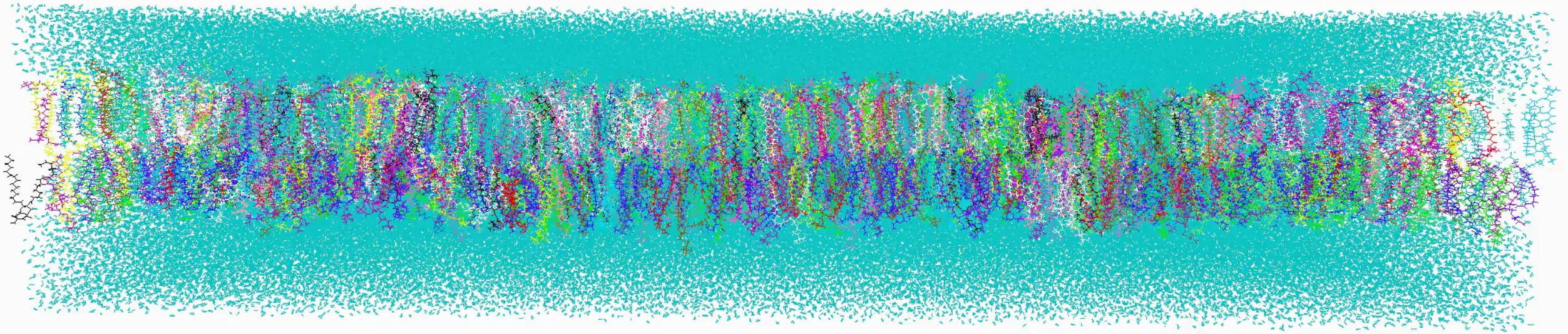
  With that, we were able to perform ~10 000 000 steps (20 ns) of simulation per day

  → To achieve long enough simulation times, two successive calls (~ 10 Mio core hours) were actually required.
- The amount of memory and scratch (~0.2 TB hours) is minimal compared to other kinds of calculations.

[a]: https://www.ks.uiuc.edu/Research/namd/wiki/?NamdPerformanceTuning

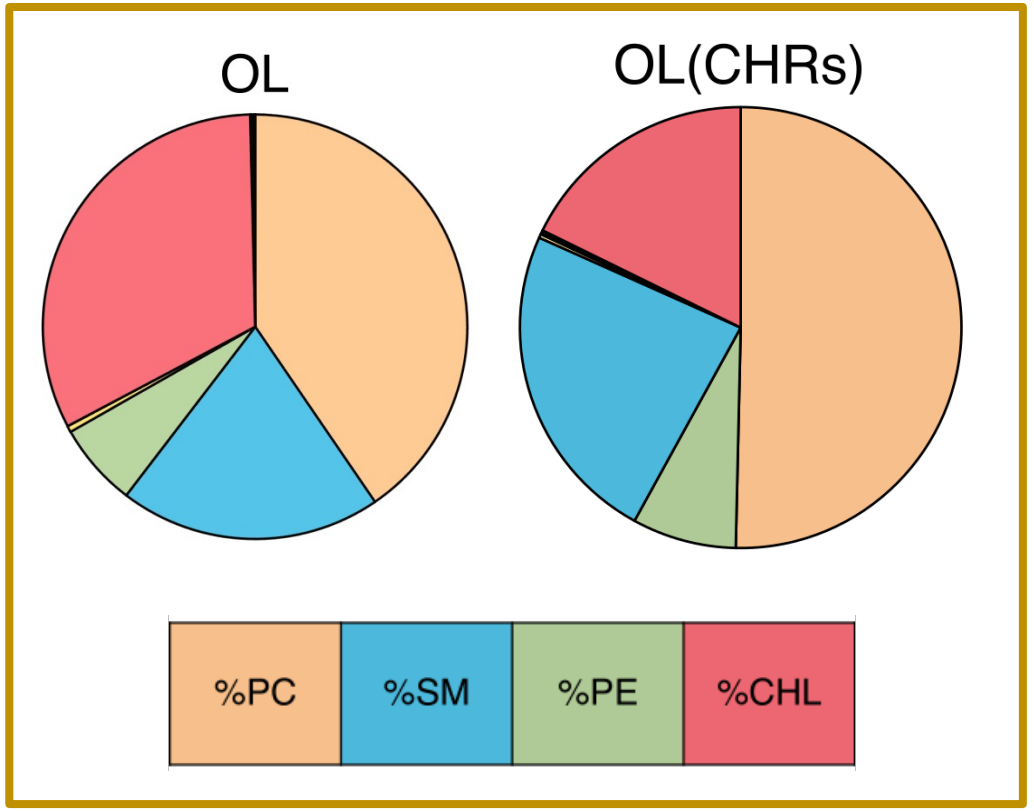
Structural parameters

- At the end, **325 ns** of (production) data, from wich we extracted **1000 geometries** of chromophore.
- No formation of domain (lipids of the same kind do not aggregate)
- Dynamic curvature fluctuations, but no variation of the membrane thickness (45.2 ± 0.1 Å)



Di-8-ANEPPS surroundings

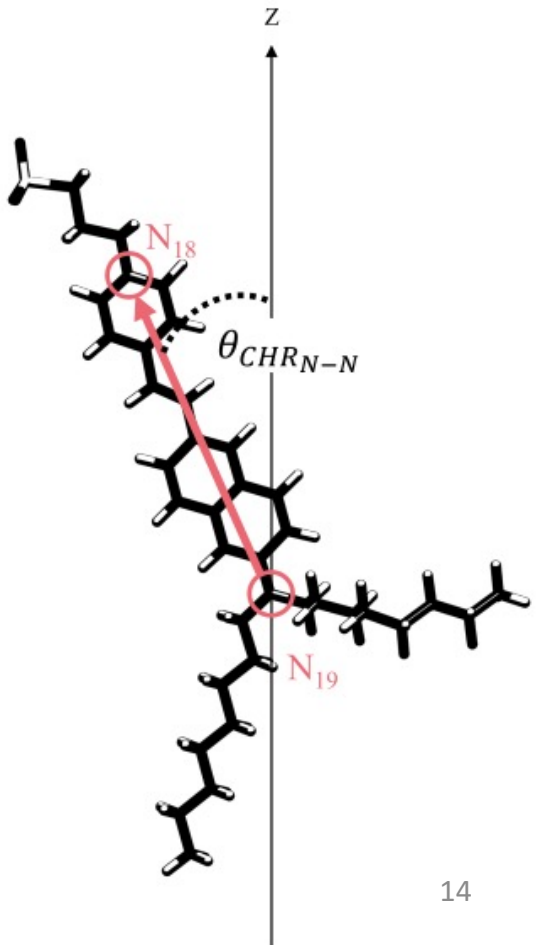
- No aggregation of the chromophores (very small diffusion constant in the membrane)
- Differences between the global distribution of lipids and the distribution near the chromophore: in particular, lower concentration of **cholesterol** around chromophore.



Composition of the membrane (4 most important families of lipids, based on the nature of the polar head), and modification of this composition near the chromophores (CHRs)

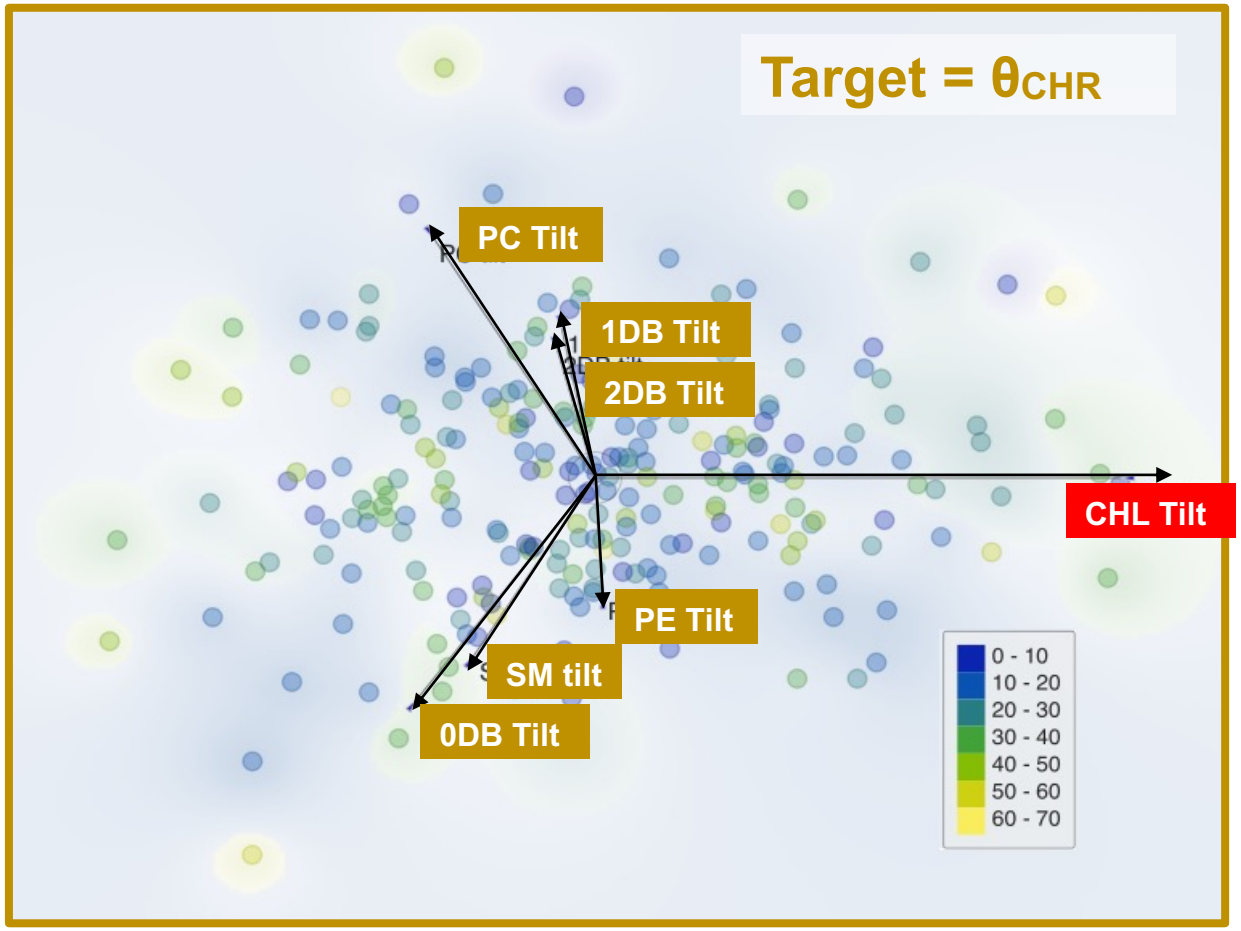
Structure-activity relationships

- In order to have a powerful detection tool, we want to get the link between the **lipid composition** near the chromophore and:
  - **the orientation** of the chromophore w.r.t. the normal (θ_{CHR}), and
  - its **SHG response.**
- A lot of data → **machine learning tools** to the rescue?



FreeViz representation[a]

- The **importance** of the feature is indicated by the length of the vectors
- **Correlation** between features indicated by their relative orientation (the farther apart they are, the less correlated)
- Colored circles are the data

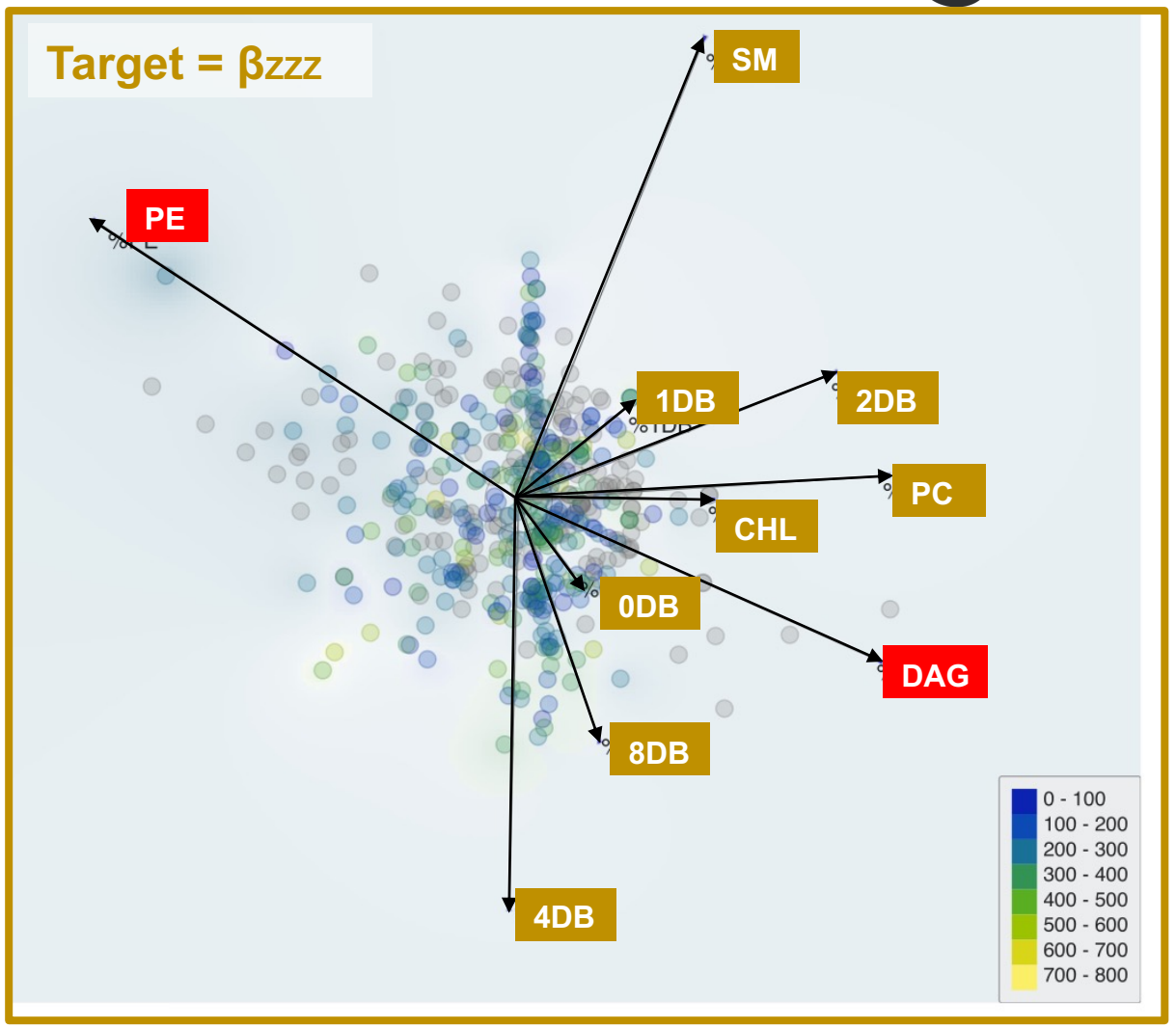


Important features to explain θ_{CHR} (DB = double bonds).

[a]: Demšar, J. et al. *J. Biomed. Inform.* **2007**, *40*, 661–671.

FreeViz representation

- The impact of each feature is not easy to disentangle
- This is the model of an **healthy tissue**! We expect larger differences in the case of diseases



Important features to explain the NLO response (βzzz).

Returns on the experience

- I have participated, in one way or another, to all the calls for the *Belgian share.* The procedure has clearly improved since the first time.
- The documentation is very nice, and the *easyconfigs* repository is a great plus.
- Frequent node failures/SLURM issues: a non-negligible amount of the allotted hours has been lost.
- NAMD GPU version not yet public, and, in general, slow movement toward AMD GPU.
- Overall, a positive experience. Therefore, I participated in the last call, in view of studying reactions at interfaces.

Conclusions

LUMI

- With the Belgian share of LUMI, we were able to simulate the first **realistic** dyed membrane composed of 3 millions atoms:
  - Thanks to a reparameterized force field, we were able to obtain correct geometries for the chromophore in this complex environment;
  - Using our sequential MD/QC approach, accurate NLO responses have been computed;
  - Finally, using machine learning tools, it was possible to extract the important features for structure-activity relationships.
  - One has to keep in mind that this model is the one of an **healthy cell membrane**.
- An article has been submitted with those results.
- Hands on process, but a very positive experience overall!

https://twitter.com/EuroCC_Belgium/status/1711675956866195836