A machine learning assessment of the two states model for lipid bilayer phase transitions

Vivien WALTER
Department of Chemistry
King's College London
Britannia House, 7 Trinity Street,
SE1 1DB, London, United Kingdom

Céline RUSCHER
Stewart Blusson Quantum Matter Institute
University of British Columbia
Vancouver BC V6T 1Z1, Canada

Olivier BENZERARA, Carlos M MARQUES, and Fabrice THALMANN*
Institut Charles Sadron
CNRS and University of Strasbourg
23 rue du Loess, F-67034 Strasbourg, France
(Dated: July 15, 2019)

We have adapted a set of classification algorithms, also known as Machine Learning, to the identification of fluid and gel domains close to the main transition of dipalmitoyl-phosphatidylcholine (DPPC) bilayers. Using atomistic molecular dynamics conformations in the low and high temperature phases as learning sets, the algorithm was trained to categorize individual lipid configurations as fluid or gel, in relation with the usual two-states phenomenological description of the lipid melting transition. We demonstrate that our machine can learn and sort lipids according to their most likely state without prior assumption regarding the nature of the order parameter of the transition. Results from our machine learning approach provides strong support in favor of a two-states model approach of membrane fluidity.

Machine Learning Techniques

Naive Bayes

Probability Basics

- Prior, conditional and joint probability for random variables
  - Prior probability: $P(x)$
  - Conditional probability: $P(x_1 \mid x_2), P(x_2 \mid x_1)$
  - Joint probability: $\mathbf{x} = (x_1, x_2), P(\mathbf{x}) = P(x_1, x_2)$
  - Relationship: $P(x_1, x_2) = P(x_2 \mid x_1)P(x_1) = P(x_1 \mid x_2)P(x_2)$

- Bayesian Rule

$$P(c \mid \mathbf{x}) = \frac{P(\mathbf{x} \mid c)P(c)}{P(\mathbf{x})}$$

$$Posterior = \frac{Likelihood \times Prior}{Evidence}$$

K-Nearest Neighbors

- Based on **feature similarity**
- **KNN is non-parametric, instance-based and used in a supervised learning**

Support Vector Machines

Phospholipids

Previous Studies

- Cubuk et al. used support vector machines to localize plastic flow regions in amorphous structures

- Carrasquilla and Melko revealed the strong aptitude of neural networks models to recognize various spin ordering regimes in condensed matter systems

- Le and Tran succeeded in predicting the polymorphism of complex lipid mixtures given a set of structural, chemical and composition parameters, by means of an artificial neural network approach

.

DPPC

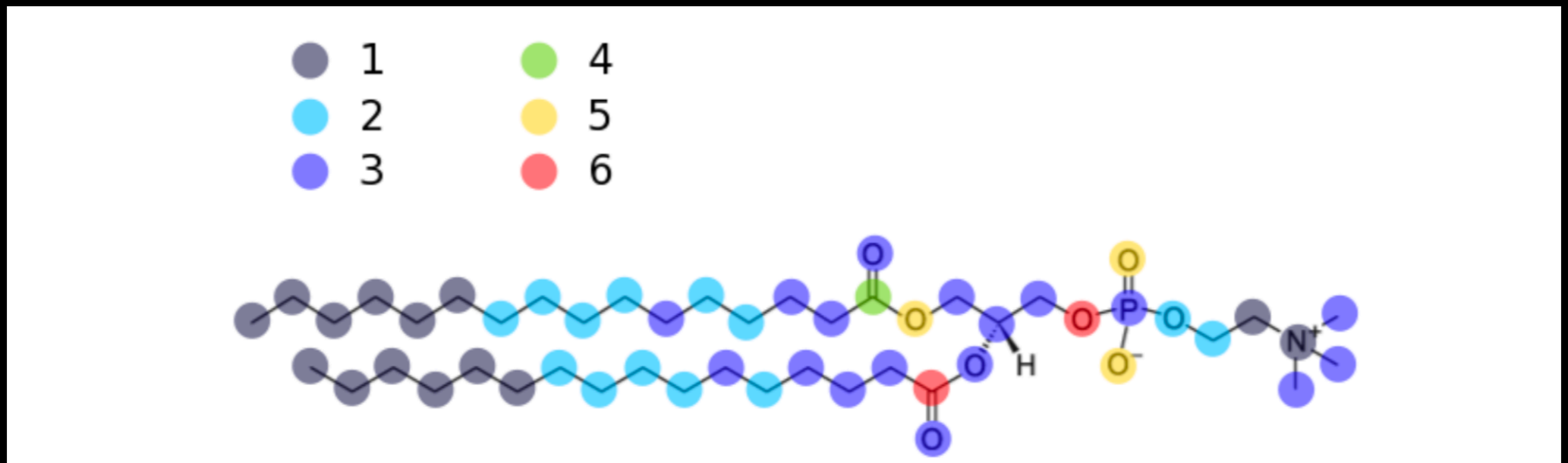
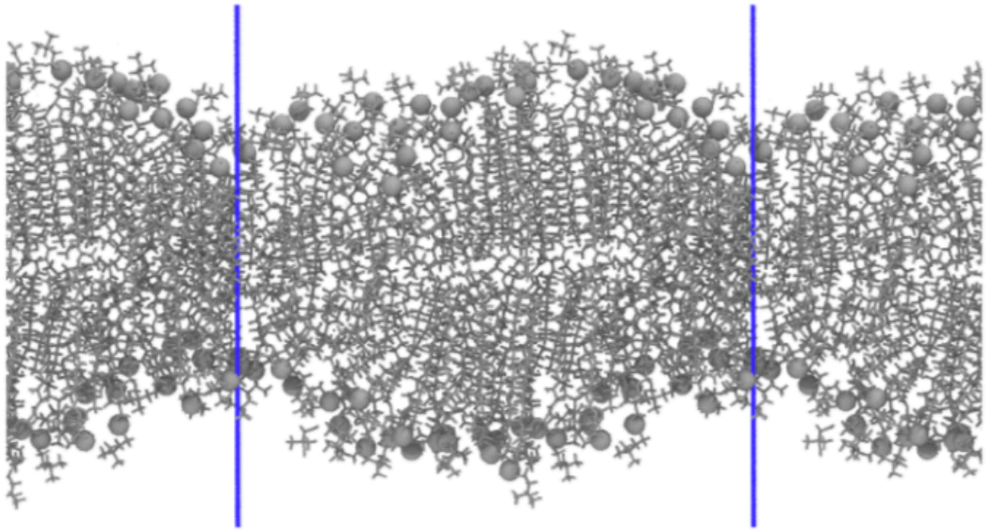


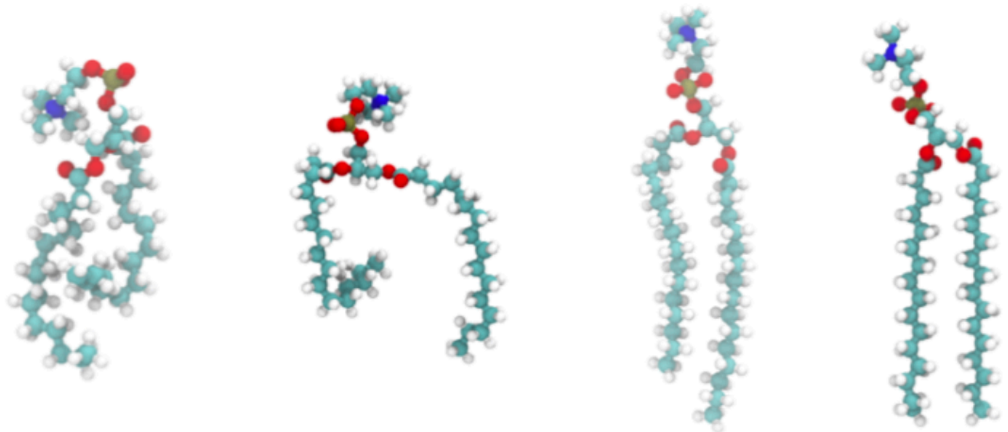
FIG. S2: Skeletal structure of a DPPC molecule. The atoms are colored according to the number of pairs formed between this atom and other atoms separated by exactly 6 covalent bonds along the molecule chain. The highest number of neighbors for a single atom was found to be 6. The total number of pairs is 61.

S_{mol}

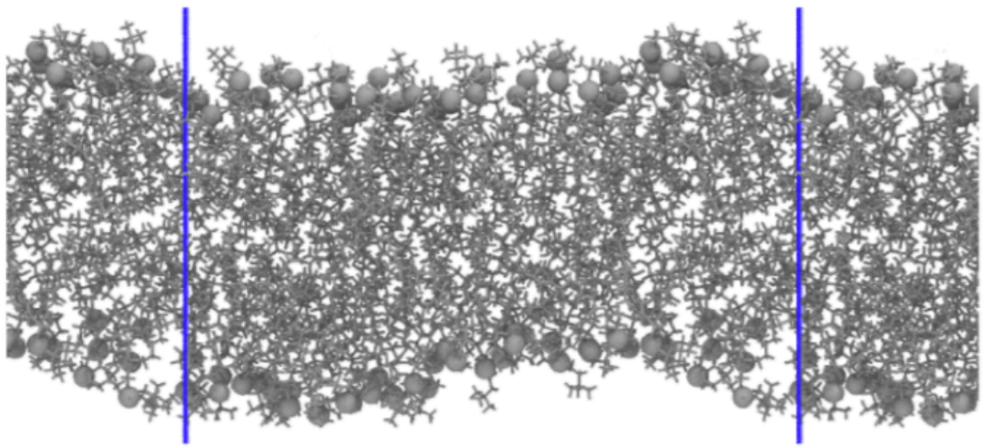
288K



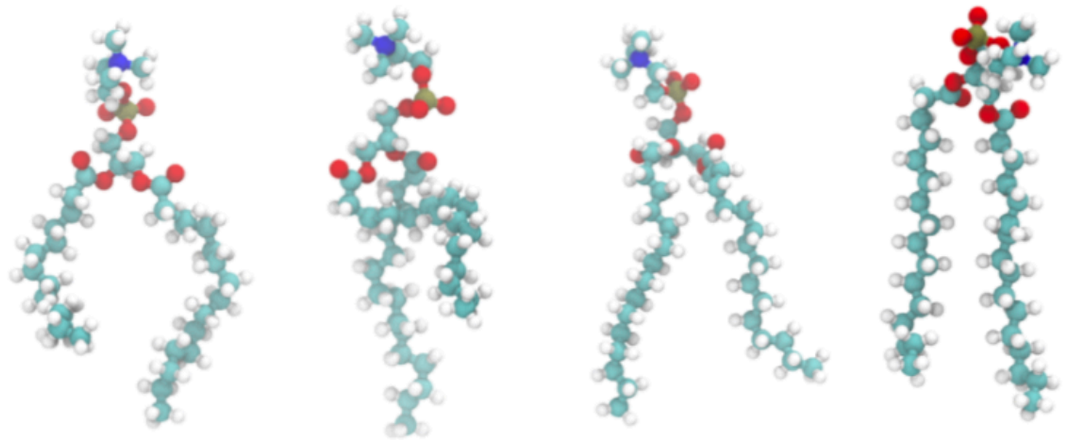
0.17 0.40 0.87 0.95



358K



0.17 0.22 0.48 0.61



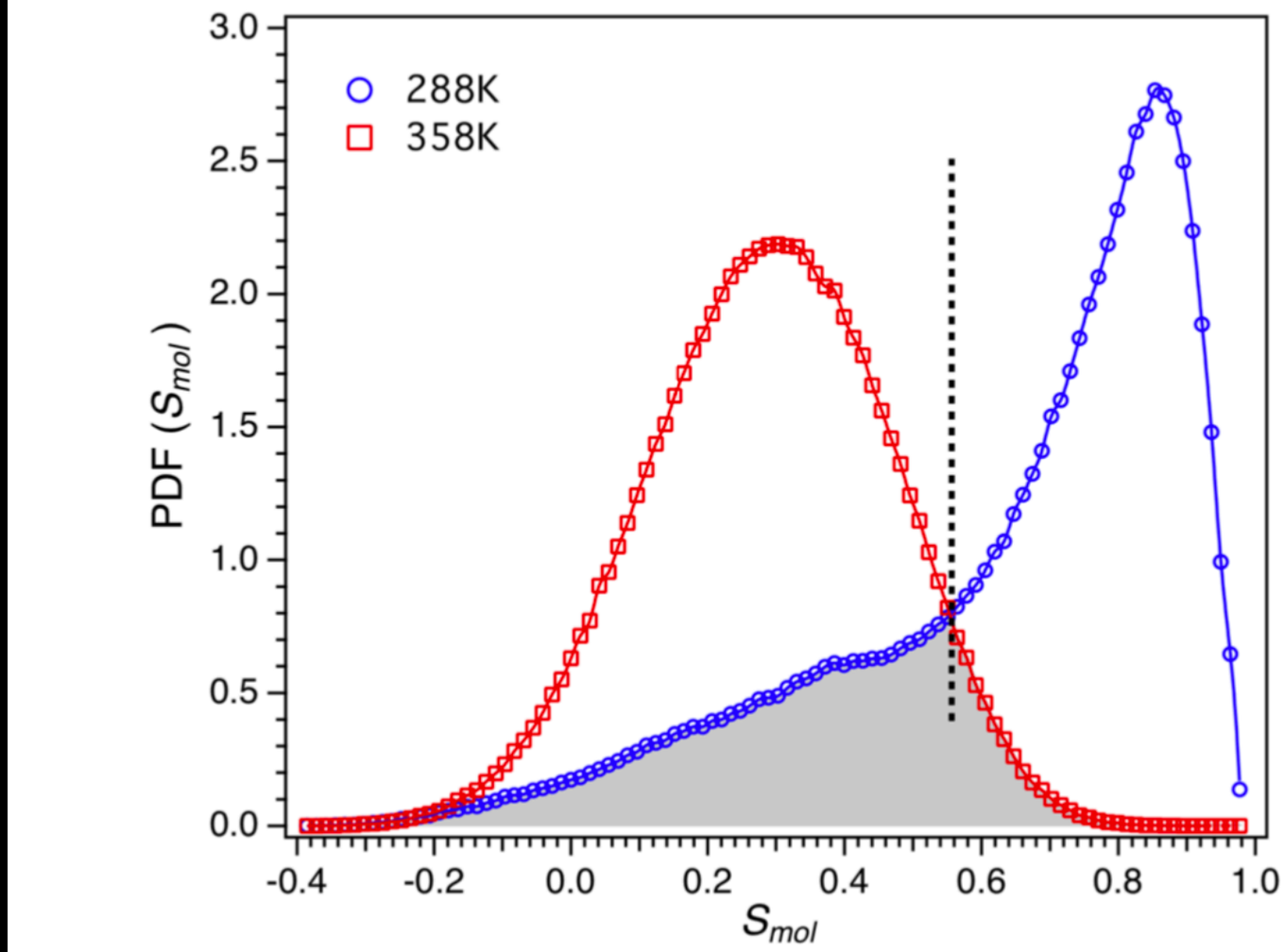


FIG. 2: Distribution of the molecular order parameters S_{mol} obtained from MD trajectories at 288 and 358 K respectively (using 10^6 lipid conformations). The dashed black line shows the threshold value that should be set to best determine the internal lipid states. The fraction of the population which would be incorrectly classified by using this threshold is represented by the shaded area in gray.

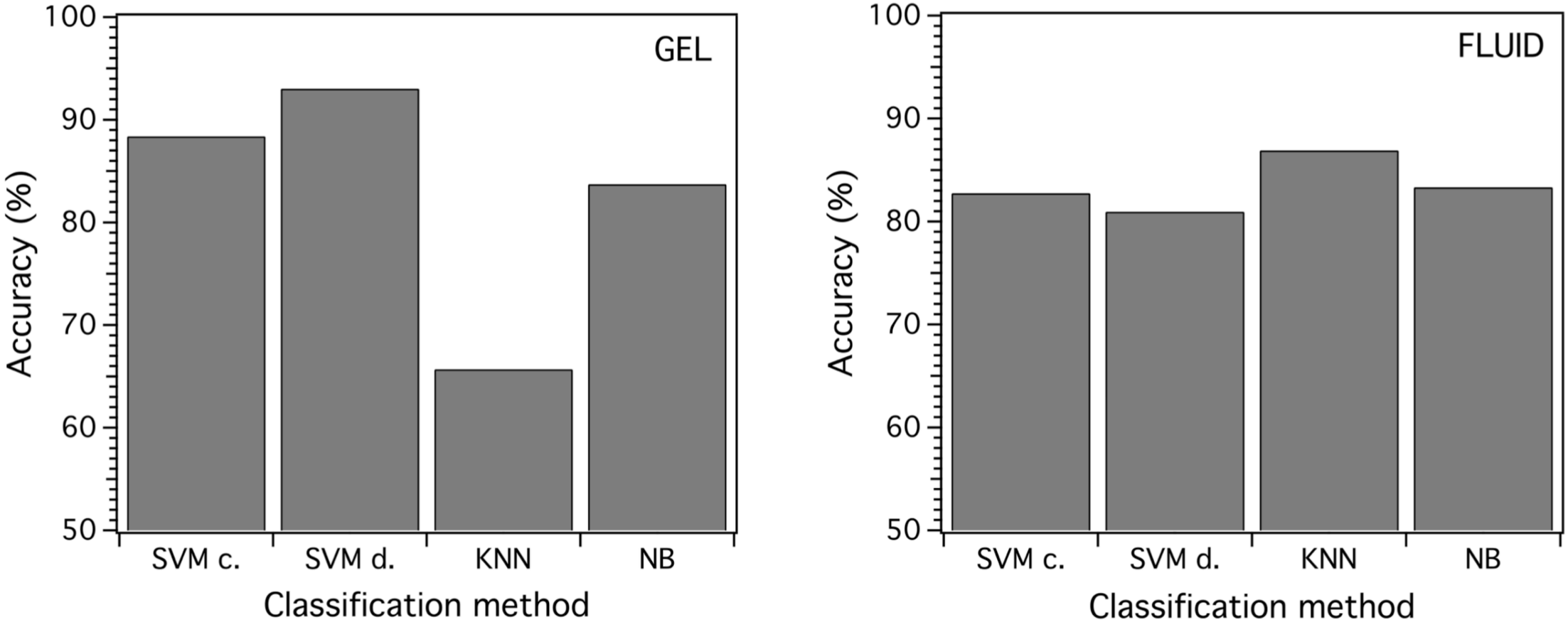


FIG. S3: Prediction scores of the Machine Learning classification methods for lipids in the gel phase (Left) and in the fluid phase (Right). Besides the Naive Bayes method (NB), all methods have an important asymmetry in their accuracy between each phase. We call these asymmetries the "expertises" of the methods.

Thank You