# Assessing the quality of reduced representations of biomolecules

Marco Giulini

Physics Department, University of Trento, via Sommarive, 14 I-38123 Trento, Italy

## Introduction

All-atom simulations of biological systems are computationally expensive and provide us with a huge amount of data: extracting relevant information is not trivial.
*Coarse Graining* (CG): effective reduction of the degrees of freedom of a complex system. It requires a reduced representation (*mapping*) of the high-resolution system and the definition of *effective interactions* among the chosen degrees of freedom.

## Issue of Mapping

$$M = \{\mathbf{M}_I(\mathbf{r}),\ \ I = 1, ..., N\}$$

$$\mathbf{M}_I(\mathbf{r}) = \sum_{i=1}^{n} c_{Ii}\mathbf{r_i} = \mathbf{R_I}$$

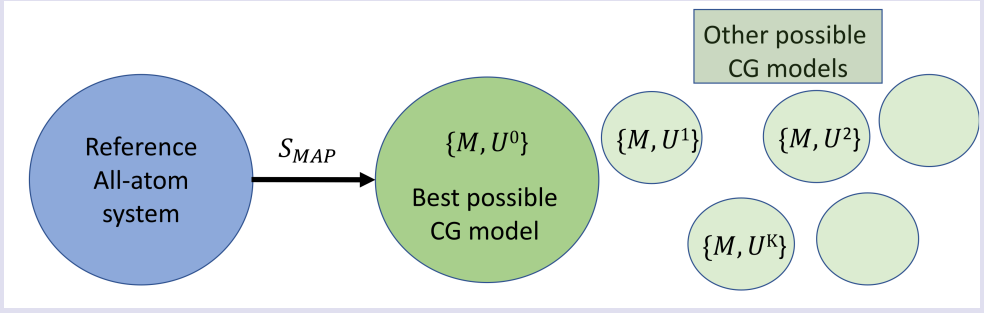
$N$ ($n$) is the number of CG sites (atoms). The accuracy of any CG model depends on the selection of degrees of freedom operated by the Mapping.

## Decimation Mapping

$\mathbf{M}_I(\mathbf{r}) = \sigma_i\mathbf{r}_i$, $\sigma_i = 1$ for one $I$, 0 otherwise,
$\sum_{i=1}^{n} \sigma_i = N$.

- preserves an intuitive representation of the protein in terms of atoms;
- decreases the number of possible CG representations to $2^n$.

## Mapping Entropy




$\mathbf{U^0}$ is the potential of mean force, i.e. the most accurate set of interactions that can be introduced in a system given a mapping $\mathbf{M}$.

$S_{map}$ quantifies the amount of information lost upon any particular choice of mapping with respect to the full atomistic system:

$$S_{map} = k_B \int d\mathbf{r}\ p_r(\mathbf{r}) \ln\left[\frac{V^n}{V^N}\frac{p_r(\mathbf{r})}{p_R(\mathbf{M}(\mathbf{r}))}\right] \geq 0$$

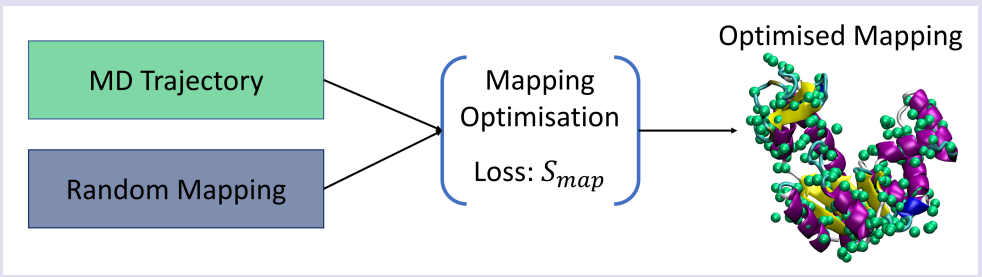
where $p_r(\mathbf{r})$ ($p_R(\mathbf{M}(\mathbf{r}))$) is the probability of the high (low) resolution system. In the case of *decimation mapping*, $S_{map}$ of a CG macrostate can be approximated with the variance of energies of AT configurations that map onto it.

$$S_{map} \simeq k_B \int d\mathbf{R}\ p_R(\mathbf{R}) \frac{\beta^2}{2} \langle (U - \langle U \rangle_{\beta|\mathbf{R}})^2 \rangle_{\beta|\mathbf{R}}$$

## Numerical Implementation

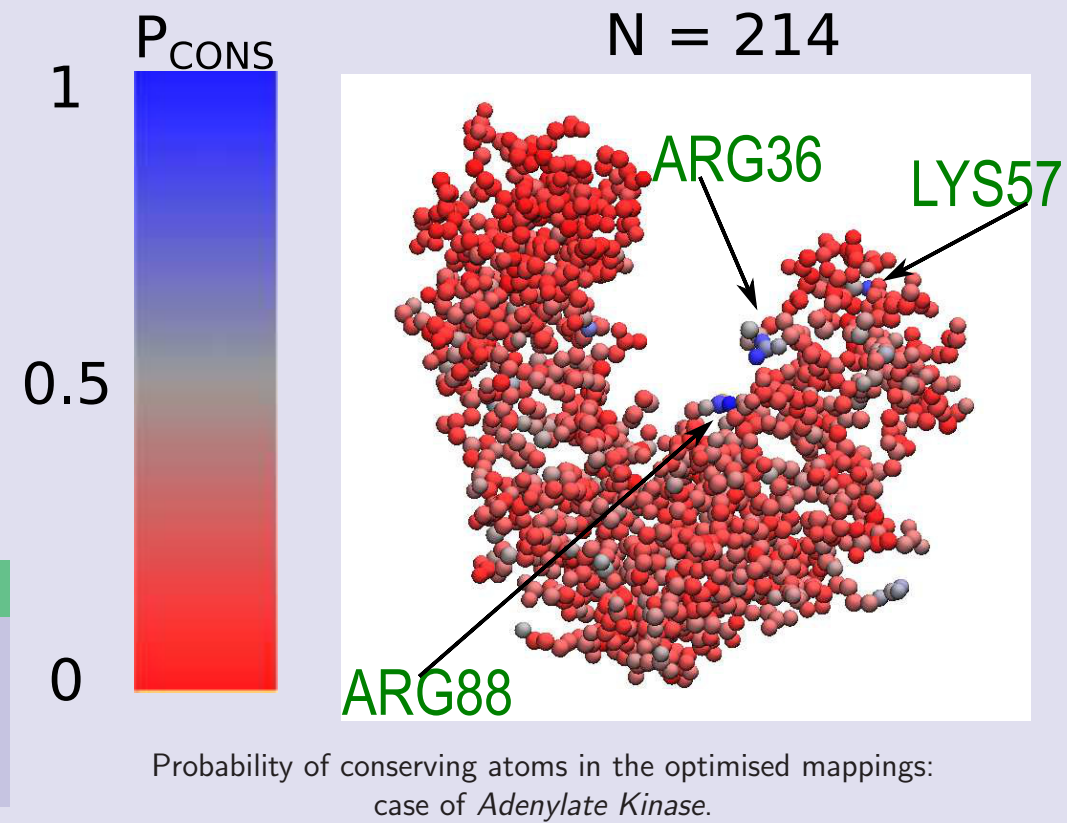
Ingredients:

- atomistic Molecular Dynamics simulation to extract configurations on which we compute $S_{map}$;
- optimisation algorithm: Monte Carlo simulated annealing. At each step two atoms are swapped in the mapping and the move is accepted using Metropolis rule.




### Results: more than Coarse-Graining

- $C_\alpha$ and Backbone mapping: local maxima of $S_{map}$. Neglecting the side chains is detrimental;
- atoms with important biological function are conserved with high probability by the optimisation.




Probability of conserving atoms in the optimised mappings: case of *Adenylate Kinase*.

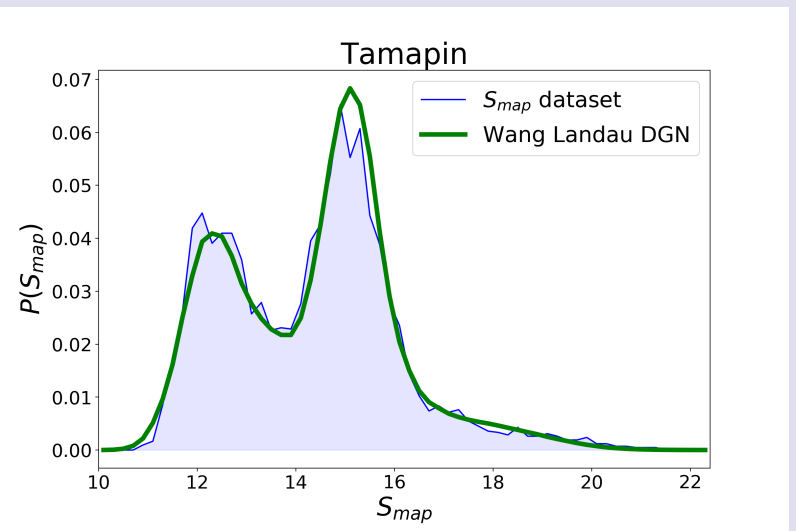
## Deep learning-enhanced sampling approach

### Deep Graph Networks

*Challenge*: both the atomistic simulation and the optimisation process are computationally heavy => machine learning can be leveraged to extract the value of mapping entropy in a much shorter amount of time.
*Our approach*: we treat the static structure of the protein as a graph and we employ a Deep Graph Network (DGN) to infer the value of $S_{map}$. The DGN proves to be accurate and remarkably efficient (speed-up $\sim 10^5$).

### Wang Landau Sampling

We incorporate the trained network into the Wang Landau sampling scheme to reconstruct the mapping entropy landscape of proteins.




*Tamapin's* $P(S_{map})$: algorithmic (blue) *vs* DGN (green).

## References

[1] M. Giulini, R. Menichetti, M. S. Shell, and R. Potestio, "An information-theory-based approach for optimal model reduction of biomolecules," *Journal of Chemical Theory and Computation*, vol. 16, no. 11, pp. 6795–6813, 2020.

[2] F. Errica, M. Giulini, D. Bacciu, R. Menichetti, A. Micheli, and R. Potestio, "Accelerating the identification of informative reduced representations of proteins with deep learning for graphs," 2020.

## Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme .

marco.giulini@unitn.it

http://variamols.physics.unitn.eu