Department of Physics,
University of Trento,
Povo (TN), Italy

entropy 2021 — **The scientific tool of the 21st century**

VAriable **R**esolut**I**on **A**lgorithms for macro**MOL**ecular **S**imulations — VARIAMOLS

Measurement and Minimisation of the Mapping Entropy of a Coarse-Grained Biomolecular System

Marco Giulini, Roberto Menichetti, Raffaello Potestio

All-atom Molecular Dynamics (MD):

- Accurate **first-principles theory** for simulating biomolecules
- Numerical solution of Newton's equation of motion
- Biologically relevant **time and length scales** are still hard to reach with plain MD

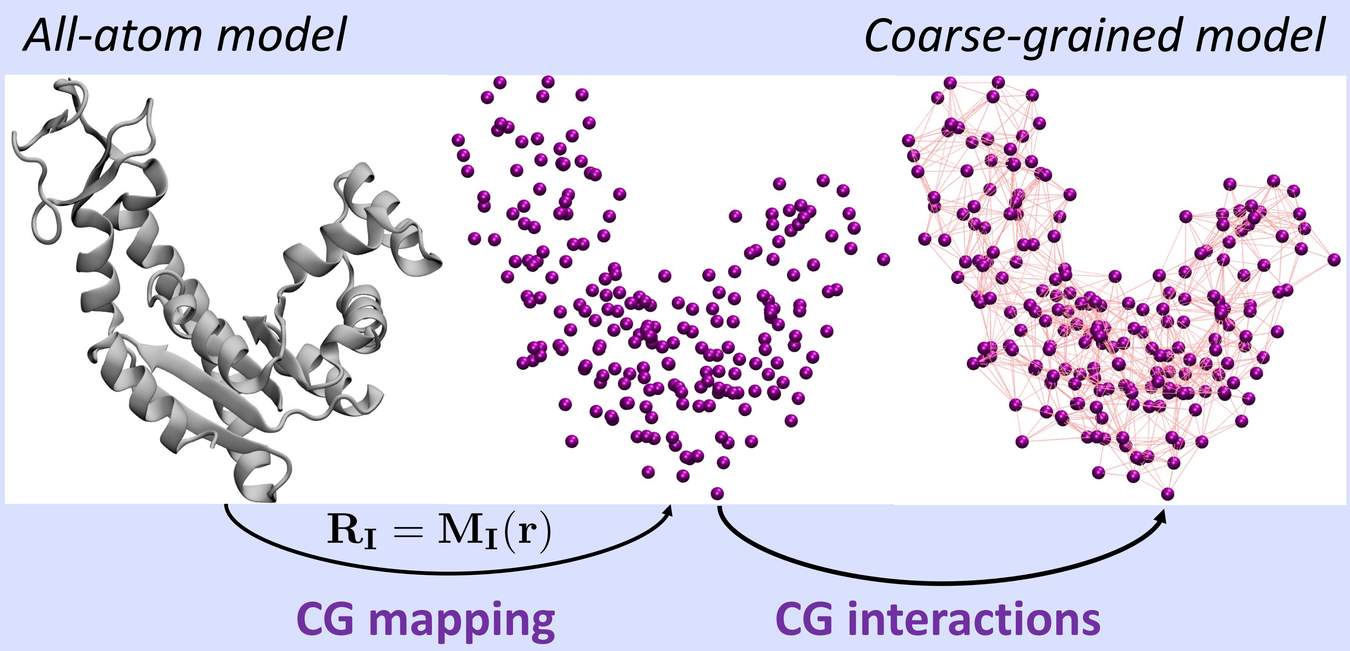

Fig.1 The two-step process of Coarse-graining.

Coarse-grained (CG) modelling:

- Effective **reduction of degrees of freedom** of a biomolecule
- **CG mapping** retains a subset of the original atoms
- Effective **CG interactions** take into account the implicit presence of the removed atoms

Target of CG: Many-body potential of mean force

$$U^0 = -k_B T \ln\left(V^N p_R(\mathbf{R})\right) + const.$$

$$p_R(\mathbf{R}) = \int d\mathbf{r}\ p_r(\mathbf{r})\ \delta(\mathbf{M}(\mathbf{r}) - \mathbf{R})$$

U^0 samples the CG probability distribution **that would be sampled** by the reference system **observed through the CG mapping**.

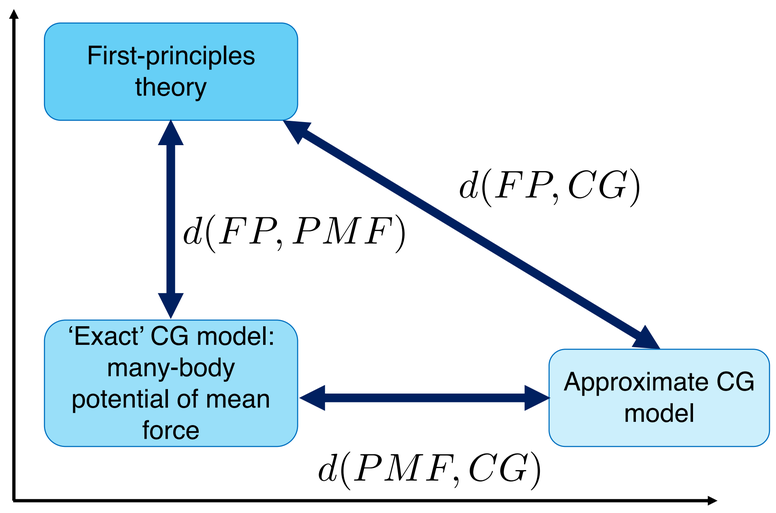

Fig.2 First-principles theory, approximate CG model and MB-PMF form a right triangle in the space of models.

How to construct good CG models?

Most CG methods focus on minimising $d(PMF, CG)$

BUT the quantity we are interested in is the **distance between the 'reality' (MD) and the model:** $d(FP, CG)$

SO we need an appropriate measure for the **quality of CG mapping**, that is $d(FP, PMF)$

The mapping entropy

- **Kullback-Leibler divergence** between the all-atom probability distribution and the one generated by the PMF
- Measures the error we make **when we reconstruct** the all-atom description from a coarse-grained representation

$$S_{map} = k_B \int d\mathbf{r}\ p_r(\mathbf{r}) \ln\left[\frac{p_r(\mathbf{r})}{\bar{p}_r(\mathbf{r})}\right]$$

$$\bar{p}_r(\mathbf{r}) = p_R(\mathbf{M}(\mathbf{r}))/\Omega_1(\mathbf{M}(\mathbf{r}))$$

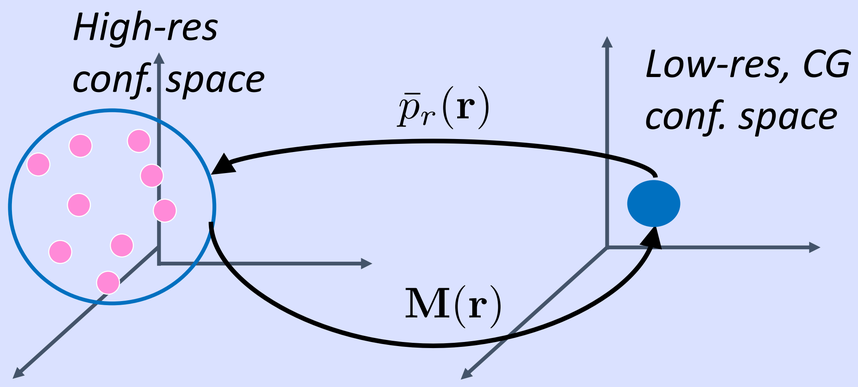

Fig.3 $\bar{p}_r(\mathbf{r})$ tries to reconstruct the correct high-res probability from the CG conf. space

Mapping optimization

- The space of CG mappings is huge ($> 10^{1000}$ elements for common proteins)
- Decimation mappings are discrete quantities
- Simulated Annealing scheme

Output: **Informative CG representations**

Results

The atoms that are more likely conserved during the minimization of the mapping entropy are those that are **essential for the biological role of the molecule.**

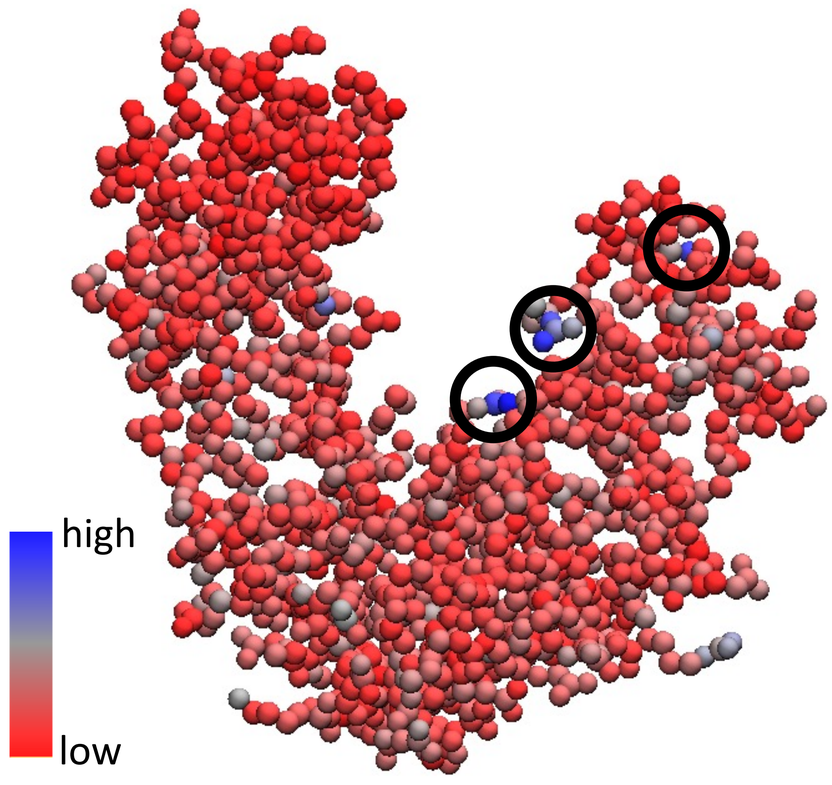

Fig.4 Probability of conserving atoms in the optimized mappings.

Conclusions

1. The mapping entropy quantifies the **distance between a first-principles theory and the most accurate effective theory** at low resolution
2. A **change of paradigm** in CG modelling: the CG mapping is an outcome of a Coarse-graining procedure
3. Mapping entropy minimization can be used as a tool for the **unsupervised analysis** of MD simulations

Ongoing and Future work

Deep learning algorithms are employed to guarantee a substantial speed-up of the calculations **[2]**. Ongoing work involves the analysis of the relationship between the mapping entropy and other information theoretical quantities. Additionally, the dependency on the conformational space sampled by MD is under investigation.

References

[1] **An Information Theory-Based Approach for Optimal Model Reduction of Biomolecules**
M Giulini, R Menichetti, MS Shell, R Potestio
J. Chem. Theory Comput. 2020, 16, 11, 6795–6813

[2] **A deep graph network-enhanced sampling approach to efficiently explore the space of reduced representations of proteins**
F Errica, M Giulini, D Bacciu, R Menichetti, A Micheli, R Potestio
Frontiers in Molecular Biosciences 8, 136.

Acknowledgements

This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (Grant Agreement 758588).

European Commission — erc