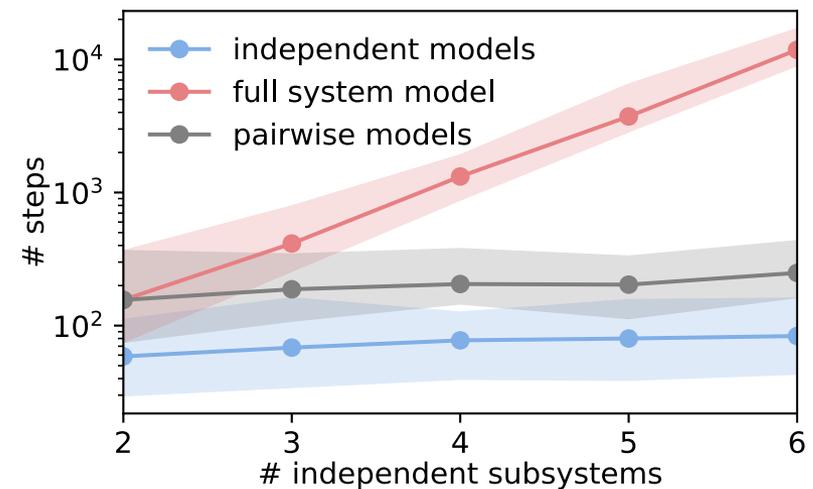


# Independent Markov Decomposition

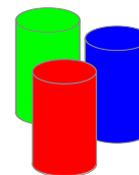
# Breaking the law

- Markov modeling requires reversible sampling
- sampling scales with the combinatorial number of states in a system
- splitting large molecules a-priori into regions of interest keeps sampling constant

sampling necessary for reversibly connecting  $n$  independent 3 state toy models



*Can we do it?*



# Recap MSMs & VAMP

The MSM transition matrix  $P$  propagates probability distributions:

$$\mathbf{p}_{t+\tau}^\top = \mathbf{p}_t^\top \mathbf{P}(\tau)$$

It is usually estimated from data by using discrete basis functions

The variational approach to Markov processes (VAMP) states that for finding the optimal model, we need to maximize the VAMP- $n$  score [1], for MSMs:

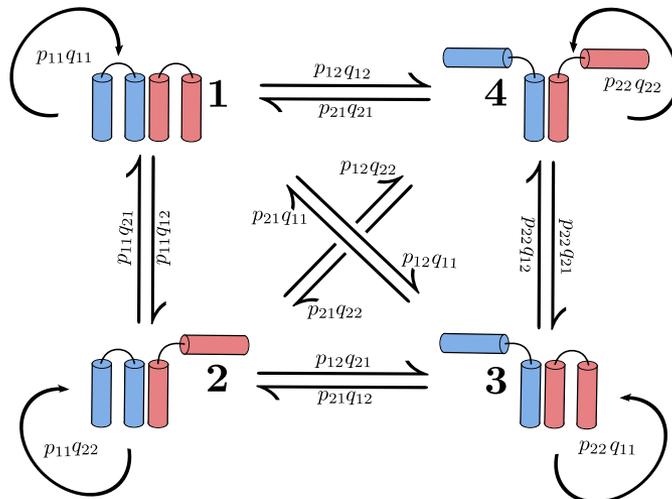
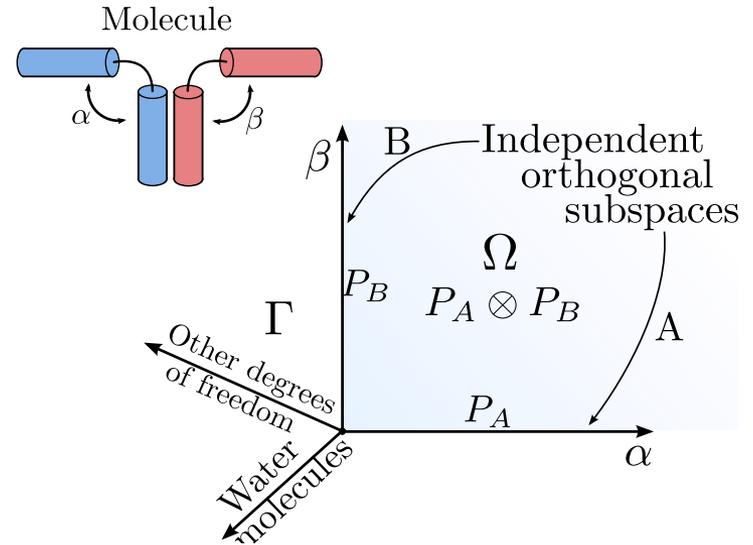
$$R_n(\mathbf{P}) = \|\mathbf{P}(\tau)\|_n^n$$

„The sum of eigenvalues“ ( $n=1$ )

[1] Wu H, Noé F (2019) Variational Approach for Learning Markov Processes from Time Series Data. *J Nonlinear Sci.*

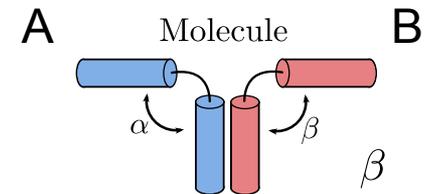
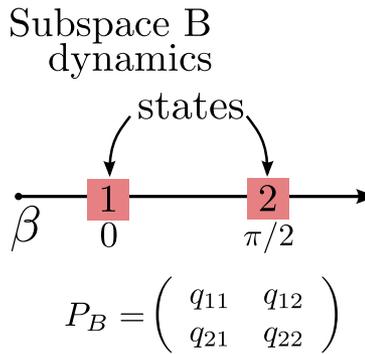
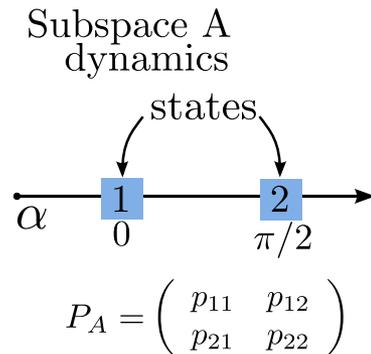
# Decomposition

- Two independent systems (blue, red)
  - embedded in a larger state space
  - live in orthogonal subspaces



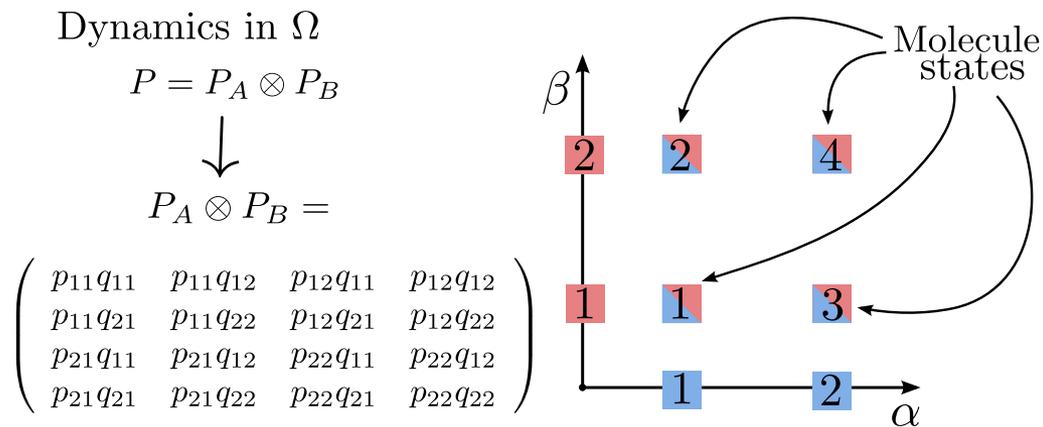
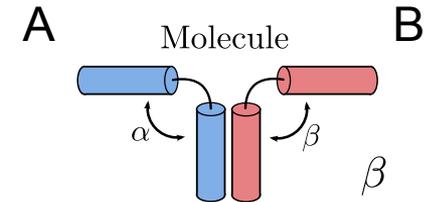
# Decomposition

Defining two disjoint sub-regions A, B of phase space, the dynamics in those regions can be independent of each other. Then



# Decomposition

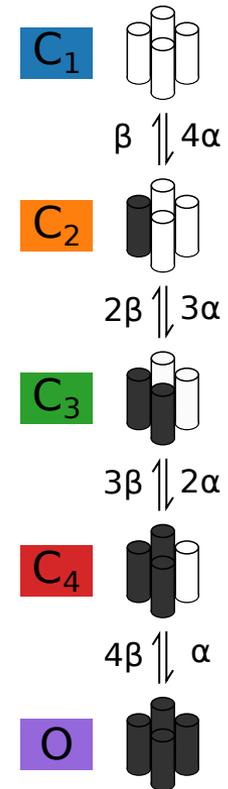
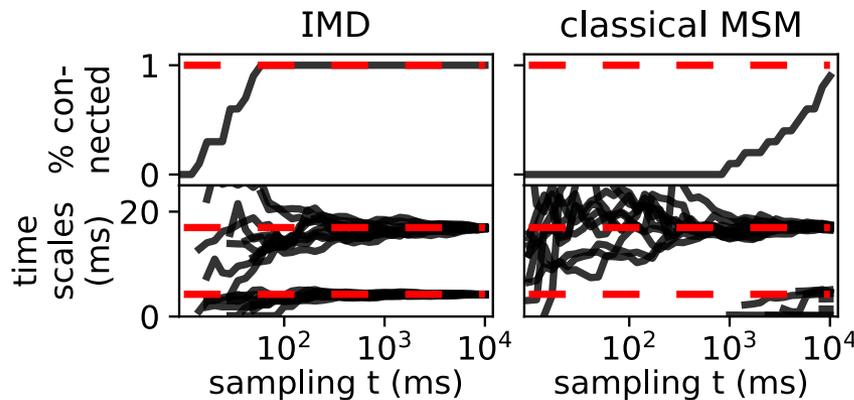
The dynamics in joint (global) space can be computed from the Kronecker product



Hempel, T.; del Razo, M. J.; Lee, C. T.; Taylor, B. C.; Amaro, R. E.; Noé, F. Independent Markov Decomposition: Toward Modeling Kinetics of Biomolecular Complexes. *Proc Natl Acad Sci USA* **2021**, *118* (31), e2105230118. <https://doi.org/10.1073/pnas.2105230118>.

# Example of 4-gated ion channel

IMD can approximately reproduce conductance measurements with  $\sim 2$  orders of magnitude less sampling as compared with a classical MSM.

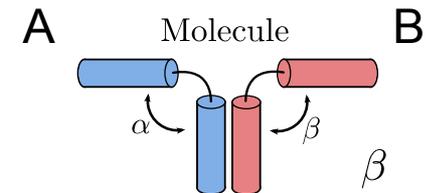


# Scoring dependency

By simply applying Kronecker product properties of matrix norms, we find (for VAMP-2 score):

$$\|T\|_F^2 = \|T_A\|_F^2 \cdot \|T_B\|_F^2.$$

Equivalent results apply to all VAMP- $p$  scores.



One can define the *dependency* score between two subsystems using the discrepancy

$$d(A, B) = \left| \|T_{A,B}\|_F^2 - \|T_A\|_F^2 \cdot \|T_B\|_F^2 \right|$$

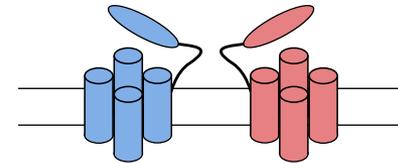
# Optimal partition toy model

**Q:** How to define subsystems in an unknown system?

**A:** Optimize dependency score!

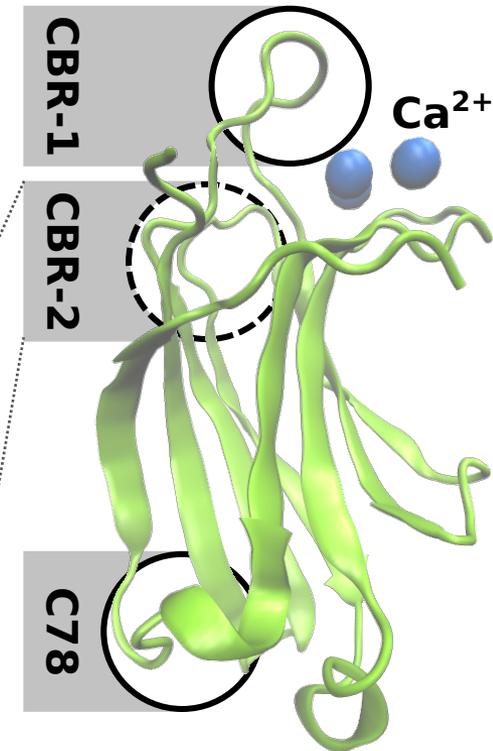
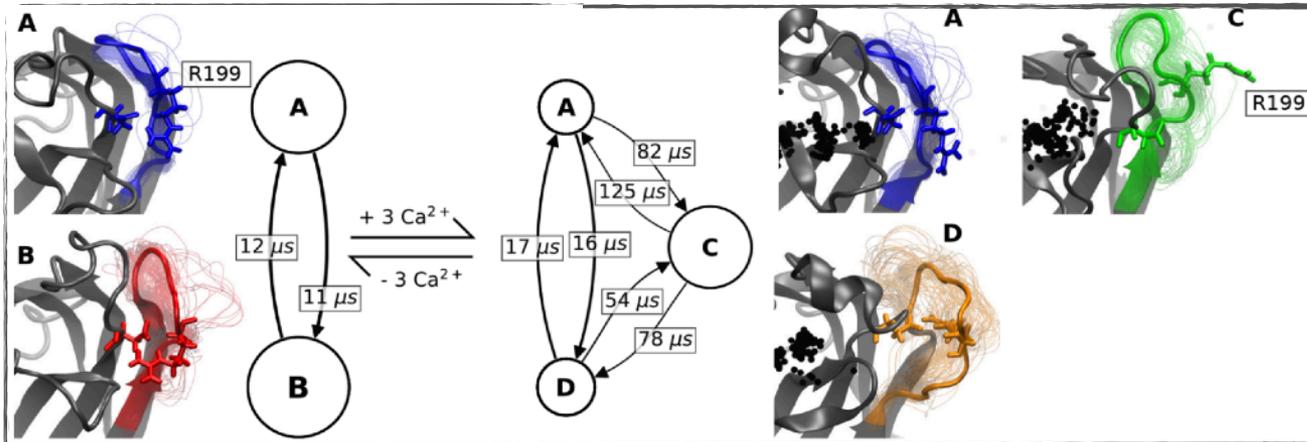
Task for you.

- You will compute the *dependency* score for all pairs of subsystems and assess which ones are strongly coupled.
- You will model one of the independent systems that you find.



# Syt-C2A: hand-selected subsystems

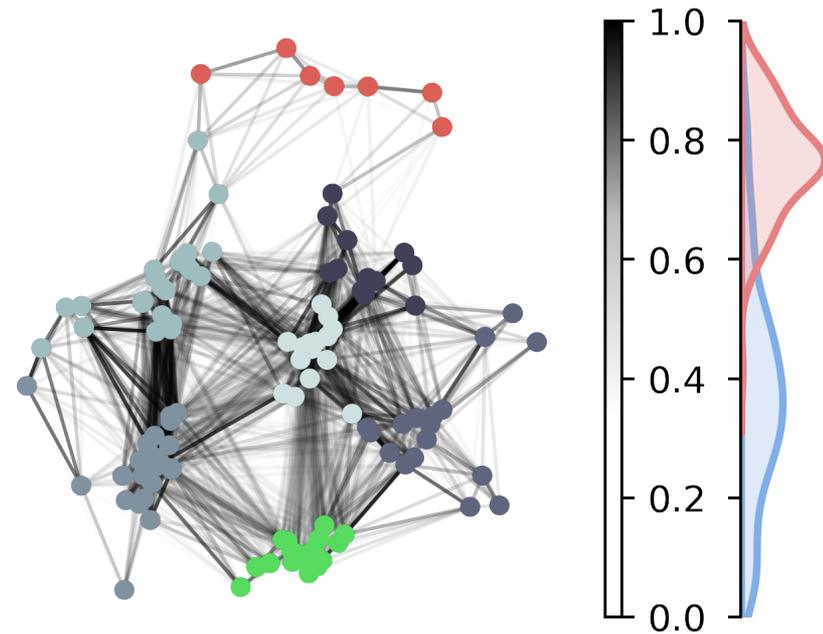
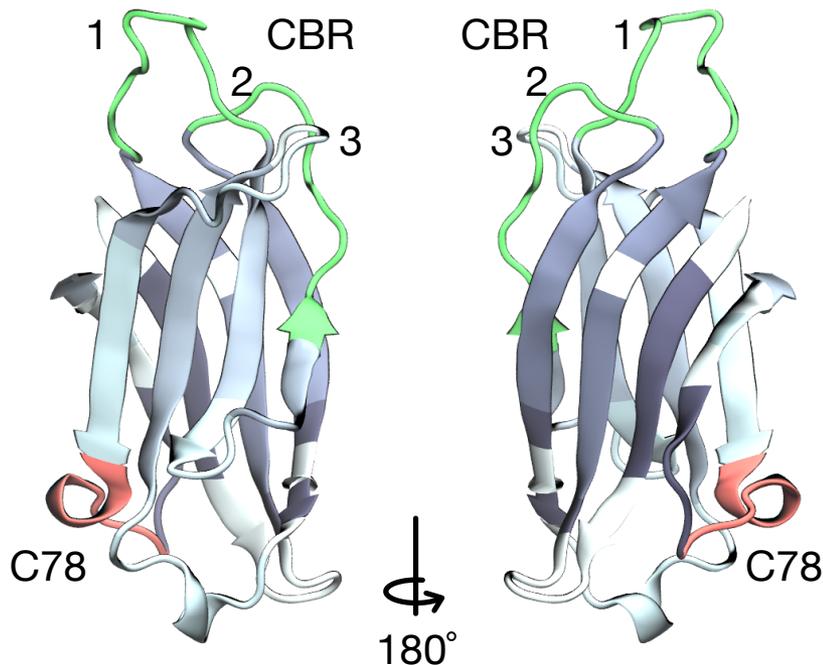
- Syt-C2A is a calcium switch in the neurotransmitter release machinery
- 180  $\mu\text{s}$  MD data (apo form)



Can we automatize this task with what we learnt?

# Syt-C2A: IMD network analysis

- Prototypical analysis with Fruchterman-Reingold projection
- We find clusters that correspond to structural features of protein



# Thanks for your attention

-> Notebook