Testing subspace overlap implementation

In order to test the implementation of the covariance overlap equation, which we wanted to use to measure the convergence of mutual information matrices, we calculate the time-averaged subspace overlap of positional covariance matrices of a 10 ns simulation of deca-alanine peptide. We tested the accuracy of the result by calculating the same measure using the program 'gmx anaeig' in Gromacs-5.1. After fixing several indexing errors we obtained the same covariance overlap result as that calculated using the gromacs tools: We obtained a perfect agreement with the covariance overlap calculated us-
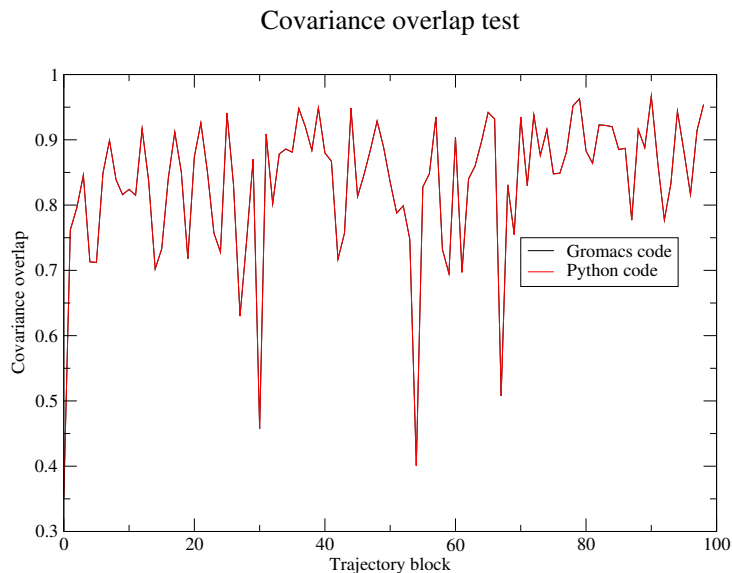
Covariance overlap test

Figure 1: Comparison of time-averaged covariance overlap measurement with Python code and with Gromacs tools

ing the Gromacs tools, giving us confidence that we had implemented the algorithm accurately.

Measurements of positional and allosteric convergence in monomeric PKM2

Next we investigated the degree of convergence obtained from explicit solvent simulations of monomeric PKM2 in various liganded states. For each liganded state: apo-, FBP-bound, Tepp-46-bound and L-Phe-bound- PKM2, we split the trajectory into blocks of increasing length at a fixed time window

of 20 ns. From the trajectory blocks we computed the positional covariance matrix and the structural alphabet mutual information matrix. To assess the degree of convergence of positional fluctuations and allosteric transitions, respectively, we measured the covariance overlap of these two measures.

First, we investigated the effect of using increasing numbers of eigenmodes in the calculation of subspace overlap. For a 1 μs explicit solvent MD simulation of monomeric PKM2 + Tepp-46, we calculated the covariance overlap for time-averaged structural alphabet mutual information. Qualitatively, we

(a) $E = 1$ (b) $E = 2$ (c) $E = 3$ (d) $E = 4$

(e) $E = 5$ (f) $E = 6$ (g) $E = 7$ (h) $E = 8$
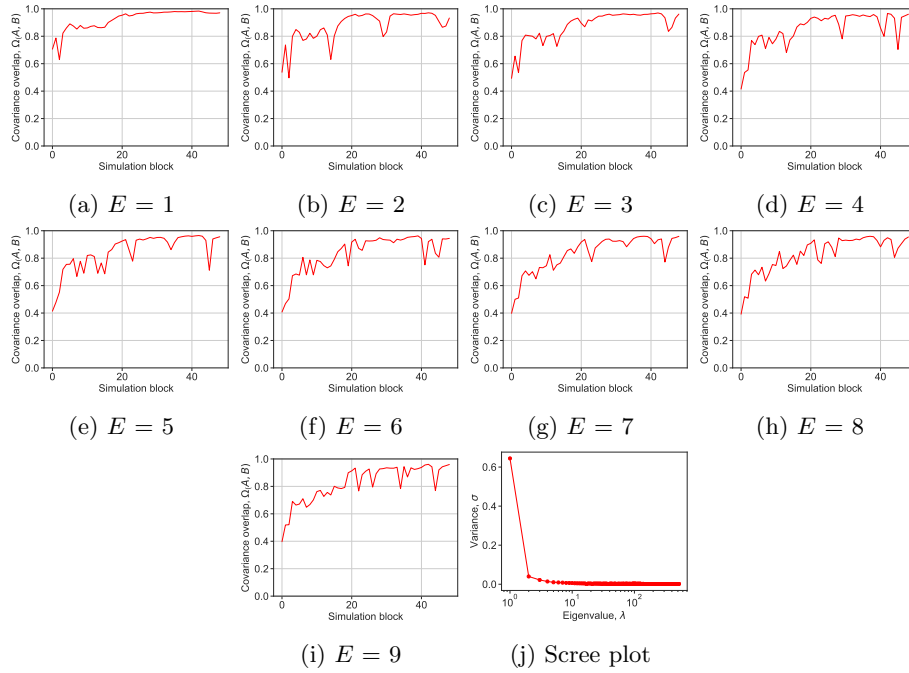
(i) $E = 9$ (j) Scree plot

Figure 2: The numerical behaviour of covariance overlap with different eigenmodes for monomeric PKM2 + Tepp-46. The MD trajectory was split into 50 sliding windows with a length of 20 ns. The subspace overlap was calculated for each blocks i_n and i_{n-1}

observed a convergent effect on the shape of the covariance overlap curve with increasing number of eigenmode, though a Scree plot suggested that only the first eigenmode contributed any appreciable variance following the spectral decomposition. To see whether inclusion of different numbers of eigenmodes in the time-averaged covariance overlap measurement we fitted the curves to a two-exponental regression. Autocorrelation functins of prin-

2

cipal components can be fitted with a double exponential function using a fast and slow correlation time. Within a conformational state there is a short correlation time as the local well(s) are explored. The transition between different conformations occurs on slower timescales, which is captured by the second term. We found no meaningful correlation between either rate terms and the eigenmode number. It should be noted that the time-averaged covariance overlap for $\lambda = 2$ was not possible to fit to the exponental regression. To determine whether the convergence of positional fluctuations

Table 1: Fitting of two-exponential curve to covariance overlap for different eigenmodes, mutual information matrix for monomeric PKM2 + Tepp-46.

λ	t_1 (ns)	t_2 (ns)
1	28.88	209.2
2	N.A	N.A
3	17.13	184.7
4	18.56	164.34
5	20.40	212.2
6	22.18	161.38
7	16.81	175.32
8	16.87	200.60
9	13.74	214.80

captured by the C_α covariance matrix coinsided with the convergence of the structural alphabet mutual information matrix, we compared the time-averaged covariance overlap for these two measures for a $1\mu s$ simulation of monomeric PKM2 + Tepp-46. A two-exponential fit of the curves in figure 3a and 3b revealed quantitative differences between the two measures of structural relaxation. The covariance overlap of the mutual information matrix yielded a fast relaxation time (t_1) of 28.8 ns and a slow relaxation time (t_2) of 174.2 ns. Both the (t_1) and (t_2) values for the C_α positional covariance matrix were significantly shorter 0.1 ns and 90.1 ns, respectively.

Longer relaxation rates estimated from the structural alphabet mutual information matrix compared to the C_α positional covariance matrix suggests that backbone conformational changes, captured by the structural alphabet, take much longer to equilibrate than the fast vibrations of the protein backbone. The long relaxation rate (t_2) may also provide an indication of the timescale on which allosteric transitions occur in PKM2 upon Tepp-46

(a) C_α positional covariance   (b) Mutual information   (c) Entropy
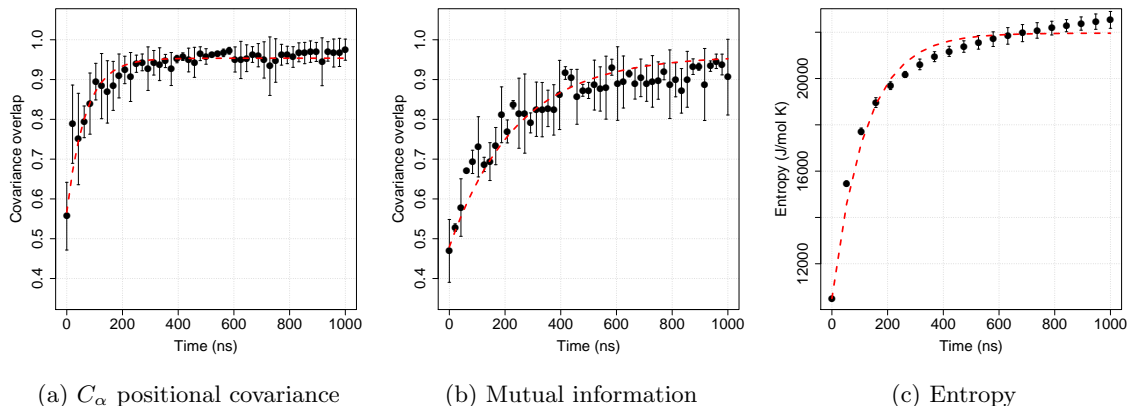
Figure 3: Energetic and structural features of monomeric PKM2 + Tepp-46 equilibration. Comparison of the time-averaged covariance overlap for the (a) C_α covariance matrix and the (b) structural alphabet mutual information matrix. The first ten eigenmodes were used in calculating covariance overlap for both (a) and (b). (c) Time-averaged configurational entropy measured over the course of the MD simulation.

binding. Still longer relaxation kinetics were observed when the energetic evolution of the MD trajectory was considered with configurational entropy: $t_1 = 41.3 \ ns$ and $t_2 = 272.4 \ ns$. These three measurements reveal three distinct timescales of protein dynamics; fast-timescale C_α positional vibrations, backbone conformational, and configurational energetic changes.

From the rate constants derived from time-averaged measurements of the covariance overlap of the C_α covariance matrix and the structural alphabet mutual information matrix, along with the time-averaged configurational entropy, it appears that the separation of distinct timescales of protein dynamics is present in each of the simulated liganded states of PKM2. From these rate constants we can assign a numerical confidence to sampling. Additionally, we are able to assign a numerical rate constant to the relaxation of independent modes of protein dynamics.

Identifying allosteric microstates in MD simulations of PKM2

To extract amino acids for directed mutagenesis it is necessary to average over parts of the MD simulations. Although the simulation may reach con-

4

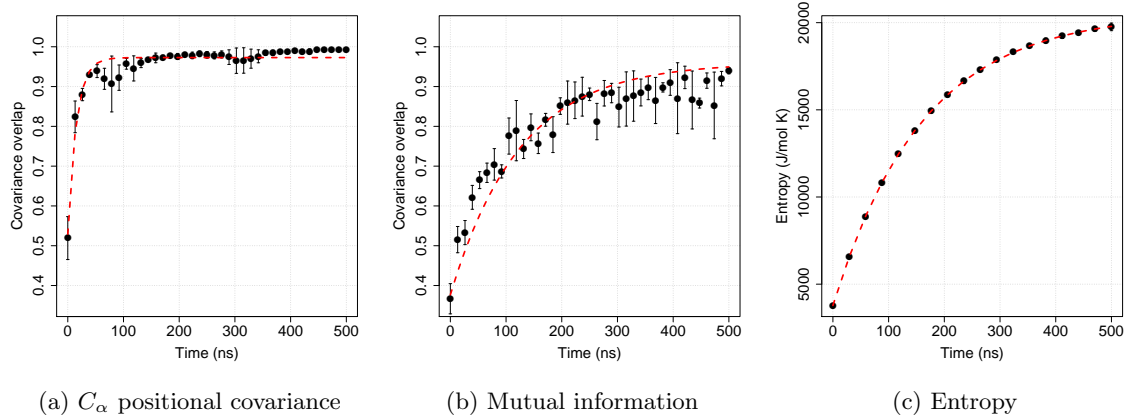(a) C_α positional covariance   (b) Mutual information   (c) Entropy

Figure 4: Energetic and structural features of monomeric PKM2 + fructose-1,6-bisphosphate equilibration. Comparison of the time-averaged covariance overlap for the (a) C_α covariance matrix and the (b) structural alphabet mutual information matrix. The first ten eigenmodes were used in calculating covariance overlap for both (a) and (b). (c) Time-averaged configurational entropy measured over the course of the MD simulation.
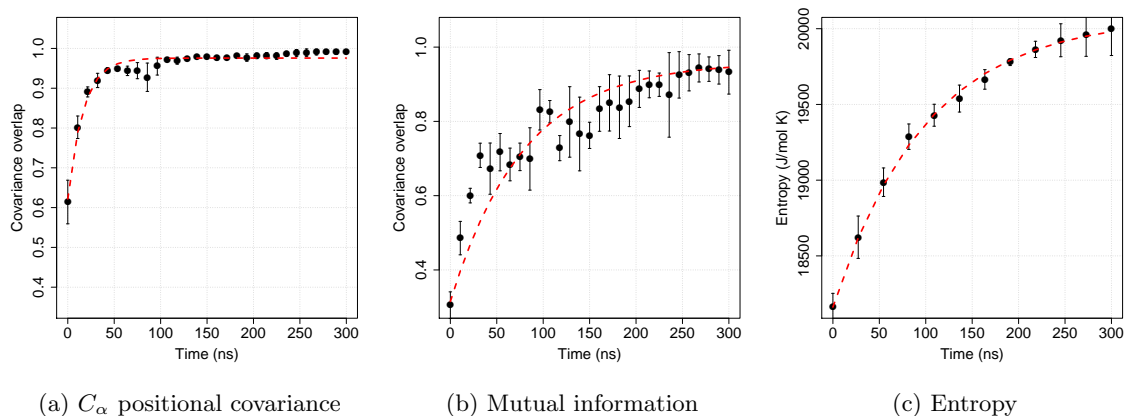

(a) C_α positional covariance   (b) Mutual information   (c) Entropy

Figure 5: Energetic and structural features of apo monomeric PKM2 equilibration. Comparison of the time-averaged covariance overlap for the (a) C_α covariance matrix and the (b) structural alphabet mutual information matrix. The first ten eigenmodes were used in calculating covariance overlap for both (a) and (b). (c) Time-averaged configurational entropy measured over the course of the MD simulation.

5

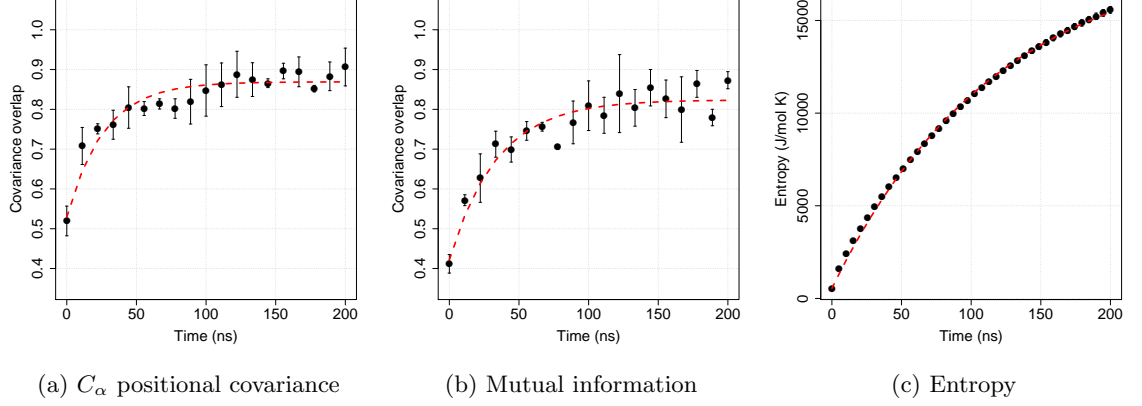(a) C_α positional covariance (b) Mutual information (c) Entropy

Figure 6: Energetic and structural features of monomeric PKM2 + L-phenylalanine equilibration. Comparison of the time-averaged covariance overlap for the (a) C_α covariance matrix and the (b) structural alphabet mutual information matrix. The first ten eigenmodes were used in calculating covariance overlap for both (a) and (b). (c) Time-averaged configurational entropy measured over the course of the MD simulation.

vergence with respect to the covariance overlap of the mutual information matrices in a time-averaged way, it was not clear whether the simulation sampled microstates over the course of the simulation. This has the possibility of scewing the signal extracted from the simulation. In order to reliably select sections of the MD trajectories to average over, we computed the covariance overlap of the mutual information matrix of time-averaged blocks of the MD trajectories of PKM2.
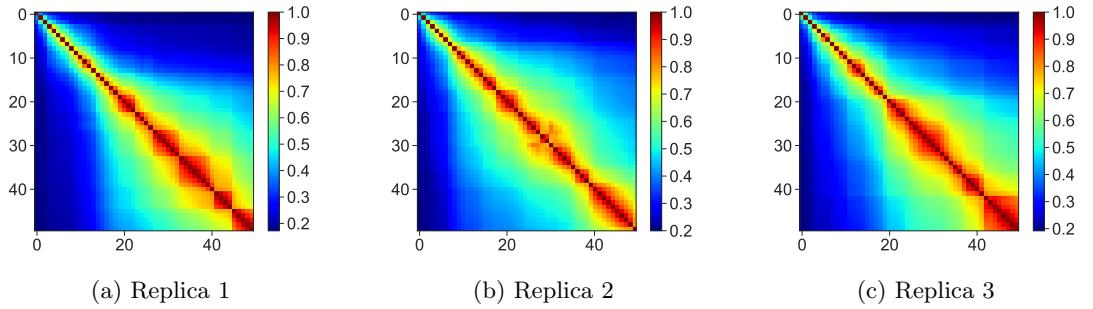

(a) Replica 1 (b) Replica 2 (c) Replica 3

Figure 7: Time-averaged covariance overlap of the structural alphabet mutual information for monomeric PKM2 + Tepp-46.

6

Table 2: Equilibration kinetics of MD simulations fit using exponential curves

System	Measure	k_1 (ns^{-1})	t_1 (ns)
Apo mPKM2 (300 ns)	Covariance	0.060	11.56
	Mutual information	0.012	57.15
	Entropy	0.010	101.30
mPKM2 + FBP (500 ns)	Covariance	0.065	10.29
	Mutual information	0.011	62.13
	Entropy	0.006	111.30
mPKM2 + Tepp-46 (1000 ns)	Covariance	0.012	57.53
	Mutual information	0.006	123.70
	Entropy	0.003	272.40
mPKM2 + L-Phe (200 ns)	Covariance	0.038	18.41
	Mutual information	0.029	23.67
	Entropy	0.008	81.84