Supplementary Information

Fahimeh Baftizadeh,[†] Xevi Biarnes,[‡] Fabio Pietrucci,[¶] Fabio Affinito,[§] and Alessandro Laio[*,†]

SISSA, Via Bonomea 265, Trieste, Italy, Institut Quimic di Sarria Universitat Ramon Llull, Barcelona, Spain, Centre Europeen de Calcul Atomique et Moleculaire, Lausanne, Switzerland, and CINECA, Bologna, Italy

E-mail: laio@sissa.it

Procedure to build the templates used in the definition of CVs

In order to obtain templates 1(2) in Figure 1, we built a system of 4 chains of VAL8 in vacuum in anti-parallel(parallel) beta-sheet configuration, using Amber99 force field. After 1 ns of equilibration at 350 K using GROMACS, we took the coordinates of the two strands in the middle as a perfect template for anti-parallel(parallel) beta-sheet of VAL8 system. Similarly, to obtain template 3, we repeated the same procedure for a system of 8 chains of VAL8 in double layer (4 strands in each layer in anti-parallel beta-sheet form, strands in the different layers were also in anti-parallel form) and we took the coordinates of the two strands in the middle of the double-layer structure as a perfect template for frontal packing of VAL8 system.

[*]To whom correspondence should be addressed
[†]SISSA, Via Bonomea 265, Trieste, Italy
[‡]Institut Quimic di Sarria Universitat Ramon Llull, Barcelona, Spain
[¶]Centre Europeen de Calcul Atomique et Moleculaire, Lausanne, Switzerland
[§]CINECA, Bologna, Italy

The criterion to choose the relevant CVs for cluster analysis.

In order to start analyzing the BE-META trajectories, it is necessary to reach the convergence in the direction of the CVs. in Figure 1, the metadynamics potentials for all the CVs in our simulation are presented. From this picture, it is clear that, we arrived to a reasonable convergence for all the CVs.

The next important step is to choose a set of CVs that are necessary to describe the thermodynamics of the system. Of course, not all the CVs are relevant to describe the process, as some of the CVs might be correlated. In our simulation, as it is clear from the definition of the CVs (see Methods), CV-B, CV-D and CV-E describe anti-parallel packing of beta strands and therefore they are correlated. The same applies for CV-C and CV-F which are describing the parallel packing of beta strands. Finally CV-G and CV-H are also correlated. Considering the level of convergence in our simulation, we decided to keep only one of the CV in each of the three sets of correlated CVs that are necessary to describe different structures. Obviously, now the question is which one between each set of correlated CV should one keep? To answer this question we performed the cluster analysis with 20 different sets of 3 CVs, with different combination and various bins size for clustering (see table 1). In all different setups, we monitored three aspects: 1- how many metastable states (basins) we obtain? 2- are the clusters well connected and covered all the space of interest? 3- are the clusters well populated? the best setup was the one that can provide us with the highest number of metastable states, which are connected in the space of CVs with well populated clusters. For our simulation, this setup was a combination of CV-C, CV-E and CV-H with grid size equal to 4.18, 1.9 and 1.64 respectively (setup 20 in the table 1) which leads to almost 500 structures.

Table 1: Different setups chosen for cluster analysis, the bins number in the direction of each CV is presented. The connectivity of the clusters in the space of CVs is monitored. Good connectivity means that all the relevant states are connected with each other. Moreover the basins that have been found in each setup is reported. *D* means disordered basin that is shown in Fig. 4 with number 1 and *A* means ordered or amyloid basin that is shown in Fig. 4 with number 3.

setup	CV-A	CV-B	CV-C	CV-D	CV-E	CV-F	CV-G	CV-H	connectivity	found basins
1	-	-	-	20	-	20	-	30	Bad	D
2	-	-	-	30	-	40	-	50	Bad	D
3	-	-	-	14	-	20	-	20	Good	D
4	-	-	-	14	-	14	-	20	Good	D
5	-	-	-	10	-	10	-	15	Good	D
6	-	-	-	-	40	20	20	-	Bad	D
7	-	-	-	-	60	40	40	-	Bad	D
8	-	-	-	-	50	30	30	-	Bad	D
9	-	-	-	-	14	14	20	-	Good	D
10	-	-	-	-	20	12	12	-	Bad	D
11	-	-	-	-	20	10	12	-	Bad	D
12	-	-	-	-	14	8	12	-	Bad	D
13	-	-	50	-	30	-	-	30	Bad	A and D
14	-	-	30	-	20	-	-	20	Bad	A and D
15	-	-	30	-	20	-	-	30	Bad	A and D
16	-	-	40	-	20	-	-	20	Bad	A and D
17	-	-	20	-	12	-	-	12	Good	A and D
18	-	-	30	-	12	-	-	12	Good	A and D
19	-	-	20	-	14	-	-	14	Good	A and D
20	-	-	16	-	10	-	-	14	Good	A and D

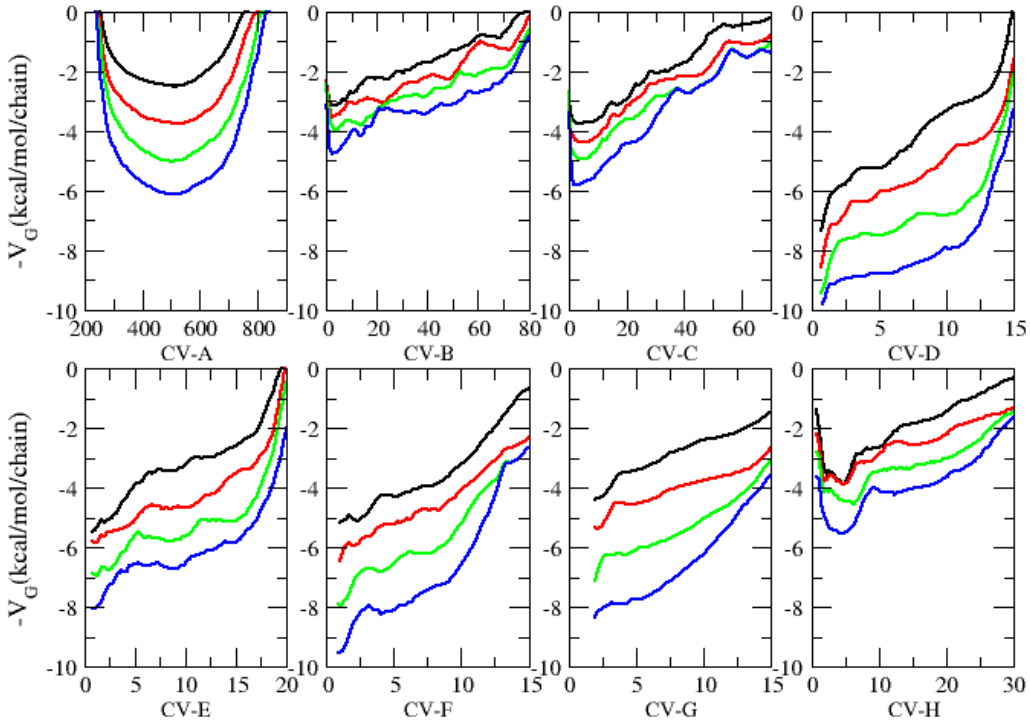

Figure 1: The Metadynamics potential growing with time as a function of CV-A. Coordination number between C_α, CV-B. antibetarmsd CV-C. Parabetarmsd CV-D. number of anti-parallel beta sheets (with $rdist_0 = 0.1$ in the switching function), CV-E. number of anti-parallel beta sheets (with $rdist_0 = 0.2$ in the switching function), CV-F. number of parallel beta sheets, CV-G. number of anti-parallel steric zipper (with $rdist_0 = 0.1$ in the switching function), and CV-H. number of anti-parallel steric zipper (with $rdist_0 = 0.2$ in the switching function). Black,red, green and blue lines are the average of $-V_G$ after 320 ns,390 ns,460 ns and 530 ns.