Electronic Supplementary Material (ESI) for RSC Advances.

A Dynamical Coarse-Grained Model to disclose Allosteric control of misfolding ββ-Microglobulin

O. Carrillo-Parramon*,[a] G. Brancolini*[b] and S. Corni[b]

Molecular Dynamics Simulations

Fully atomistic MD simulation of the free-standing β2m in water are started from one of the 18 available NMR native structure of the unbound protein (PDB code: 1JNJ, model 1). Protonation of residues has been conducted to mimic the presence of pH=6.8. The protein has been solvated by SPC/E water molecules in a 5.5x6.5x5.5 nm^3 box and described by the OPLS/AA force field at T=300 K. [50] After an energy minimization, the simulation has been started in the NVT ensemble and equilibrated for 2 ns. The following 50 ns have been collected and only the trajectories from the last 20 ns have been used to extract essential dynamics eigenvectors and eigenvalues [44]. All simulations have been performed with the Gromacs 4.5.5 package [51]. The lengths of bonds have been constrained with the LINCS algorithm. Periodic boundary conditions and the Particle-Mesh-Ewald algorithm were used. A 2 fs integration time step was set. At the end of the equilibration, the trajectory was reasonably stable in terms of density, temperature and potential energy. The RMSD analysis of the last 20 ns revealed a stable

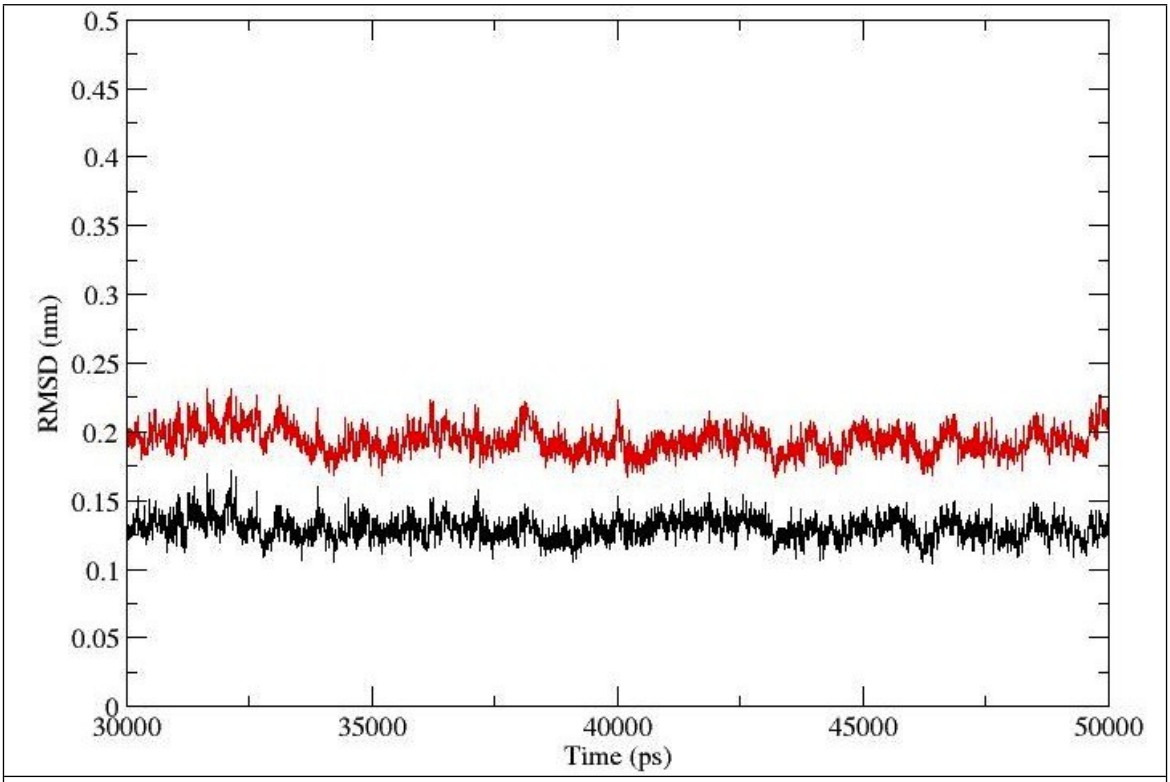

Fig.1SI Classical MD: All atoms root mean square deviation (red) and backbone root mean square deviation (black) respect to the equilibrated structure of the selected model 1JNJ-1 at each snapshot along the last 20 ns of MD trajectory.

trajectory over time (Fig. 1SI).

Comparison of MD and CG results

The quality of the CG model was assessed through the comparison between the last 20 ns of the fully-atomistic MD simulation and the LD-CG counterpart. In case of validation, both methods are expected to provide equivalent results, as discussed in ref. (44). Here, LD has been run using a force constant of κ=150 kcal/molÅ^2 between all pair of beads of the CG model. Essential Dynamics was applied to both MD and LD-CG simulations in order to extract the corresponding sets of eigenvectors (or modes) and eigenvalues characteristic of each dynamics. The principal component analysis was used to facilitate the comparison between the two independent simulations. Additionally, in Table 2 we report the similarity index γ computed by including the first ten modes for both MD and LD eigenvectors. The same value was also computed in the case of 90% and 95% of variance, respectively.

	10 first modes	**90%**	**95%**
γ	0.39	0.73	0.80

Table 1SI. Similarity index is used to compare the Essential Dynamics modes of MD simulations by using AMBER with the OPLSAA force-field and LD simulations by considering Kovacs' CG-method. Comparison is done for the first 10 essential modes, the first 34 modes (providing 90% of the variance) and for the 59 first essential modes (providing 59% of the variance). Although the first eigenvectors of both dynamics are not similar, the general motion is well described by LD.

Our results indicate a lower value of the index a in the presence of only 10 modes, but a clear increase of the index at higher values of the variance, where the similarity index becomes larger than 0.6 indicating an equivalent flexibility pattern between the two cases. Similar conclusions can be extracted from Fig. 2SI, where eigenvalues are represented as a function of their mode number.

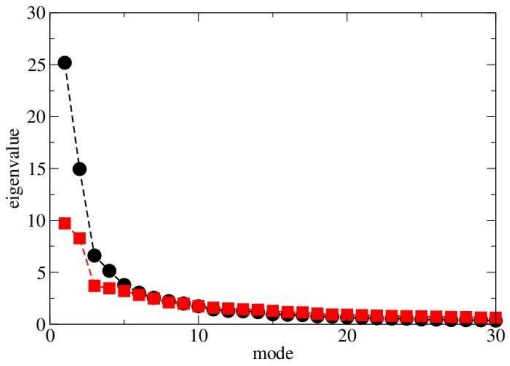

Fig.2SI Representation of the eigenvalues (in Å^2) of the Essential Dynamics from MD (in black) and from CG (in red), versus the principal mode. Discrepancies are found only in the autocorrelations terms of the first modes as expected from the similarity index results.

Computed eigenvectors for the MD (in black) and the CG (in red) are sharing a similar trend and the data tend to the same curve. Larger discrepancy are observed for the autocorrelation terms of the first few computed principal modes, in line with the previous discussion on the similarity index. β-factors were also computed and compared in Fig. 3SI, the plot is showing the differences between the two simulations

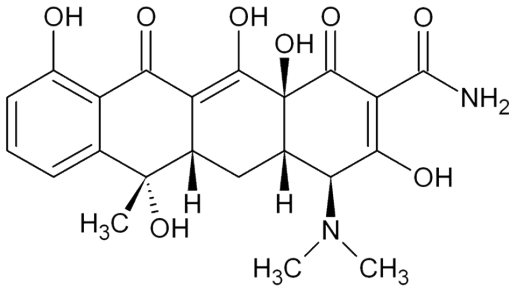

Fig.3SI Chemical Structure of Doxycicline Drug

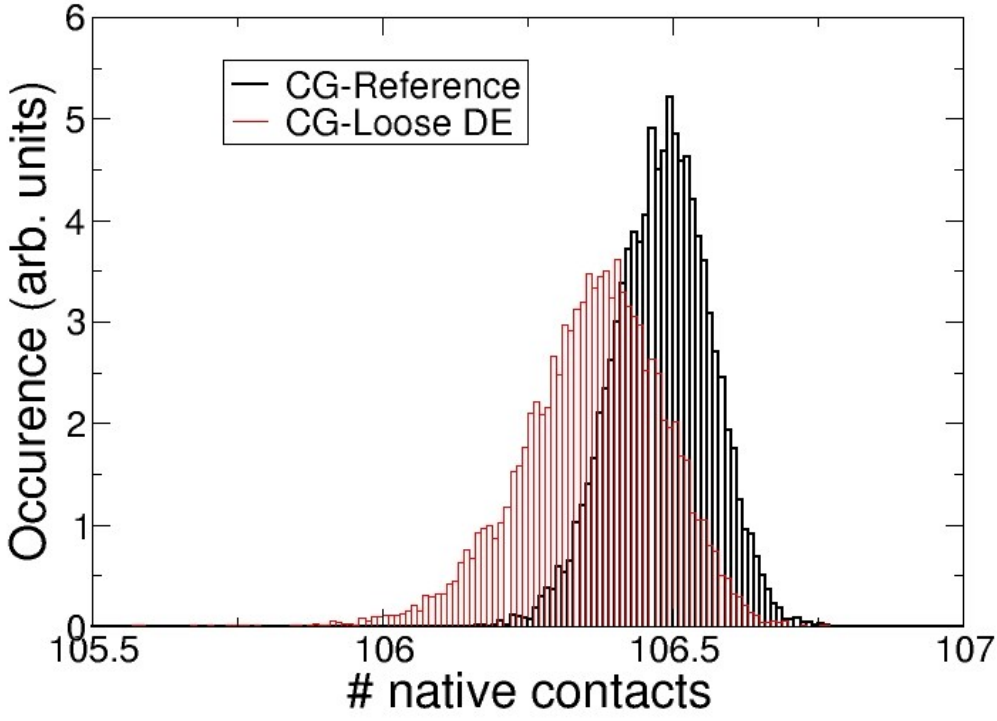

Fig. 4SI. Occurrence of native contacts for the CG dynamics in the original ("CG-Reference") and K_{DE} perturbed ("CG-Loose DE") simulations. The results indicate that in the perturbed simulations the number of native constant is, on the average, smaller than for the unperturbed one. The native contacts are defined as in Robert B. Best, Gerhard Hummer, and William A.

Eaton PNAS 2013 vol. 110, 17874-17879

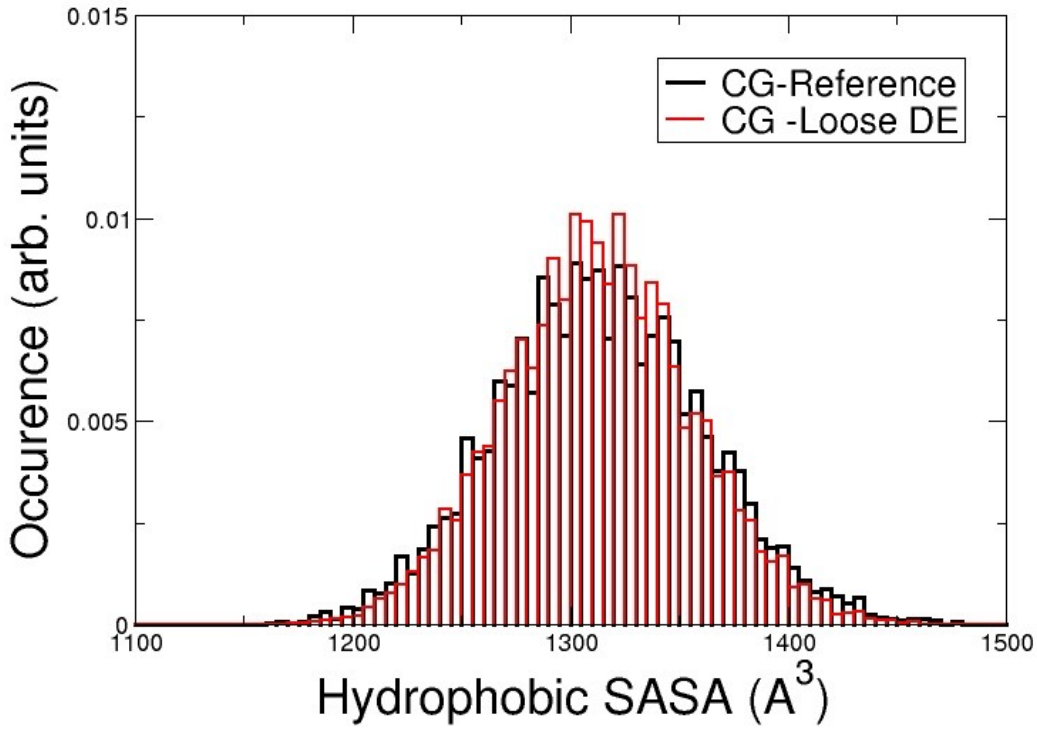

Fig. 5SI Occurrence of values for the protein Surface Accessible Surface Area pertaining to hydrophobic residues, as calculated with VMD measure command, for the non-perturbed and the perturbed CG simulations. A radius of 4.5Å has been used for each residue.

Comparison of CG results with an allosteric path: the swapped dimer

The GC results on protein structure were compared with the MM characterization of the Potential Energy Surface of an allosteric path in which the final structure was characterized experimentally. More in details, we used as the final product a swapped dimer recently characterized by X-ray crystallography (Domanska, K. *et al* **2011**, PNAS 108 (4), 1314-1319), that has been speculated to be an early amyloidogenic intermediate of β2m-DN6 self-association. In the swapped dimer, two extended hinge loops are unmasked and fold into a new two stranded antiparallel β-sheet. The β-strands of this sheet are prone to self-associate and stack perpendicular to the direction of the strands to build large intermolecular β-sheets that run parallel to the axis of growing oligomers, providing an elongation mechanism by self-templated growth. For consistency with the CG simulations we have used PDB:1JNJ as a starting structure i.e. reactant (devoided by its first six terminal residue) and PDB:2X89 as the final structure i.e. product, the X-ray structure of the ΔN6β2m swapped dimer.

All of the molecular structures, intermediates and transition states, reactants and products through the path, were fully optimized using a full matrix Newton Rapson algorithm (FMNR) within the OPLSAA molecular mechanics procedure as implemented in the Tinker program (Dudek, M.J.; Ponder, J.W. *J. Comput. Chem.* **1995**, *16*, 791-816). The presence of water solvent was simulated using the GB/SA implicit solvent model (Mohamadi, F *et al. J. Comput. Chem.* **1990**,*11*, 440-467).

This description of the mechanism involved in the barrier-crossing event followed the variation of two selected geometrical parameter. The selected parameters were the detachment of F and G strands and the distance between strand D and E which was evaluated as the distance between C_alpha of residue SER51 (strand D) and SER55 (strand E). Our results in Fig.6SI, show that the distance between F and G strands (given by the PDB:2X89) is correlated with the distance between D and E strands, described as the increasing CA-CA distance between residues (51,55) belonging to the DE hinge. The results generate the allosteric transformation which permit the protein the hopping from a local minimum to another and the corresponding conformational changes are supporting the crucial role of DE strand and DE loop in those changes.

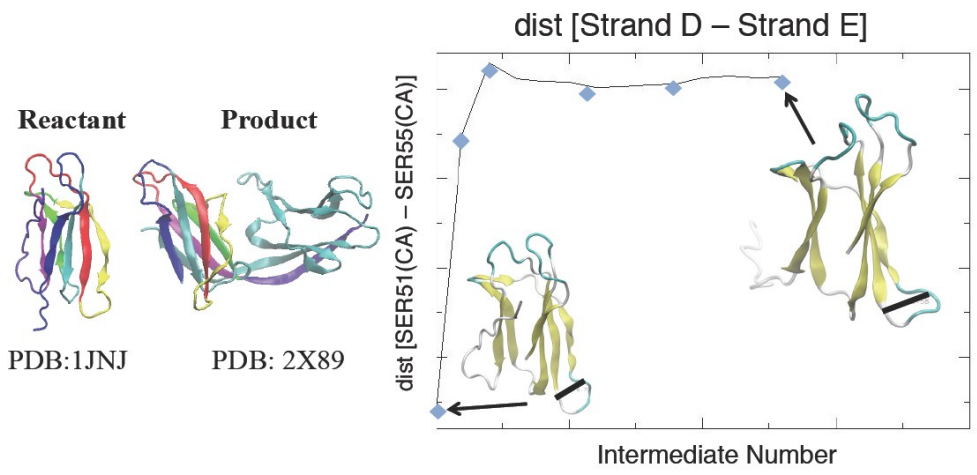

Fig. 6SI Occurrence of values for the protein CA-CA distance between residue SER51 (strand D) and SER55 (strand E) along the allosteric path leading the native protein (PDB:1JNJ) to its structure in the swapped dimer (PDB:2X89).

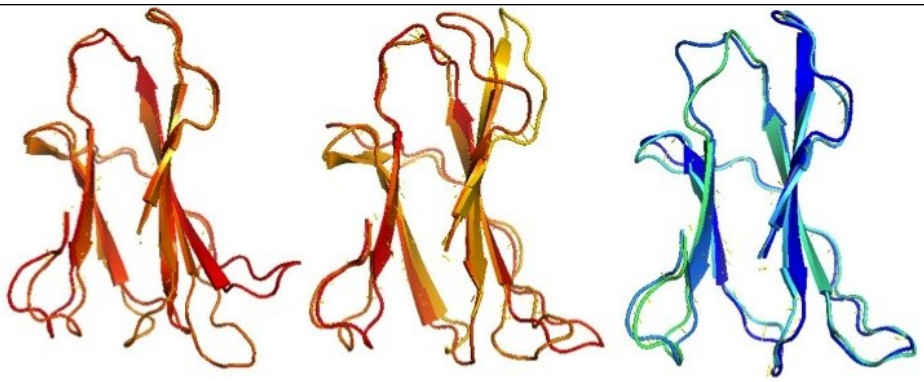

Fig. 7SI Principal Component Analysis: first three dominant fluctuations of the protein in solvent from 20 ns of T-REMD with 32 replicas. Largest collective motions of atoms were localized at DE-loop regions.