Supporting Information

Exploring the folding process of the human βB2-crystallin using multiscale molecular dynamics and Markov state model

José-Luis Velasco-Bolom[a] and Laura Domínguez*[,a]

[a] Facultad de Química, Departamento de Fisicoquímica, Universidad Nacional Autónoma de México, Mexico City 04510, Mexico

Corresponding Author: E-mail: lauradd@unam.mx

Table of contents

Contents

Table S1. Systems of MD simulations. Illustrating the number of replicas of each MD production and the length of each replica.

Systems All-atom	Replicas	Length (µs) of each replica
EM	5	1
	15	0.2
CM	5	1
	5	0.2
ED	2	0.5
	3	0.2
CD	2	0.5
	3	0.2
Systems Coarse-grain	Replicas	Length (µs) of each replica
EM	50	5
ED	50	5
CD	50	5

Table S2. Stationary probabilities and relative free energies of each macrostate described for HβB2C.

Macrostate	Stationary probability (π)	Free energy (k_BT)
S1	0.006	5.137
S2	0.024	3.748
S3	0.010	4.636
S4	0.014	4.253
S5	0.015	4.204
S6	0.018	4.015
S7	0.012	4.418
S8	0.017	4.093
S9	0.013	4.303
S10	0.021	3.840
S11	0.016	4.154
S12	0.022	3.832
S13	0.128	2.059
S14	0.025	3.430
S15	0.032	3.690
S16	0.077	2.563
S17	0.124	2.089
S18	0.103	2.270
S19	0.099	2.308
S20	0.070	2.661
S21	0.154	1.870
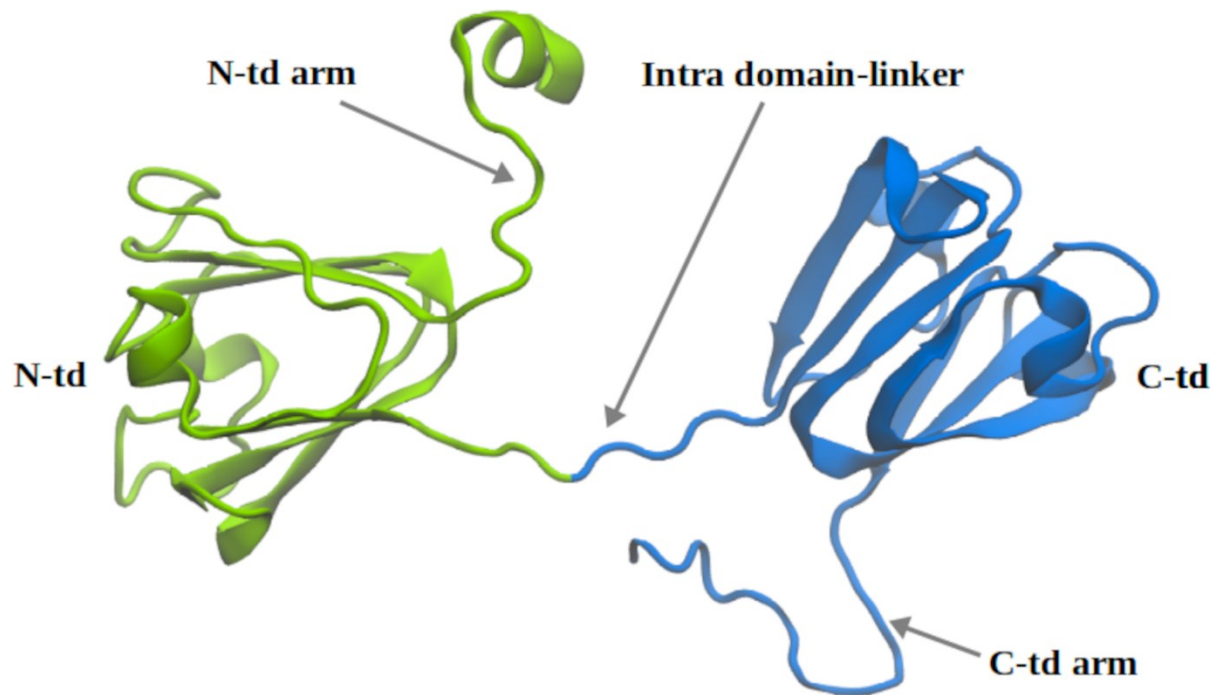
Figure S1. General depiction of the Human βB2-Crystallin in an extended conformation. This protein exhibits a N-terminal domain (N-td), C-terminal domain (C-td), N-td and C-td arms, and the intra-domain linker.

001 sp|P07316|CRGB_HUMAN
002 sp|P26998|CRBB3_HUMAN
003 sp|P05813|CRBA1_HUMAN
004 sp|P07315|CRGC_HUMAN
005 sp|P62697|CRBB2_RAT
006 sp|P11844|CRGA_HUMAN
007 sp|P53673|CRBA4_HUMAN
008 sp|Q8WXF5|CRGN_HUMAN
009 sp|P53674|CRBB1_HUMAN
010 sp|P07320|CRGD_HUMAN
011 sp|P22914|CRYGS_HUMAN
012 sp|P43320|CRBB2_HUMAN

h=Residues conserved at the hydrophobic
intra-domain interface
W=Tryptophan conserved
Y=Tyrosine conserved

1 2 3 4 5 6 7 8 9
Variable   Average   Conserved

Figure S2. Sequence alignment of twelve members of βγ-crystallin family. The sequence alignment was made with ClustalW server and colored according to residue conservation calculated with ConSurf server.
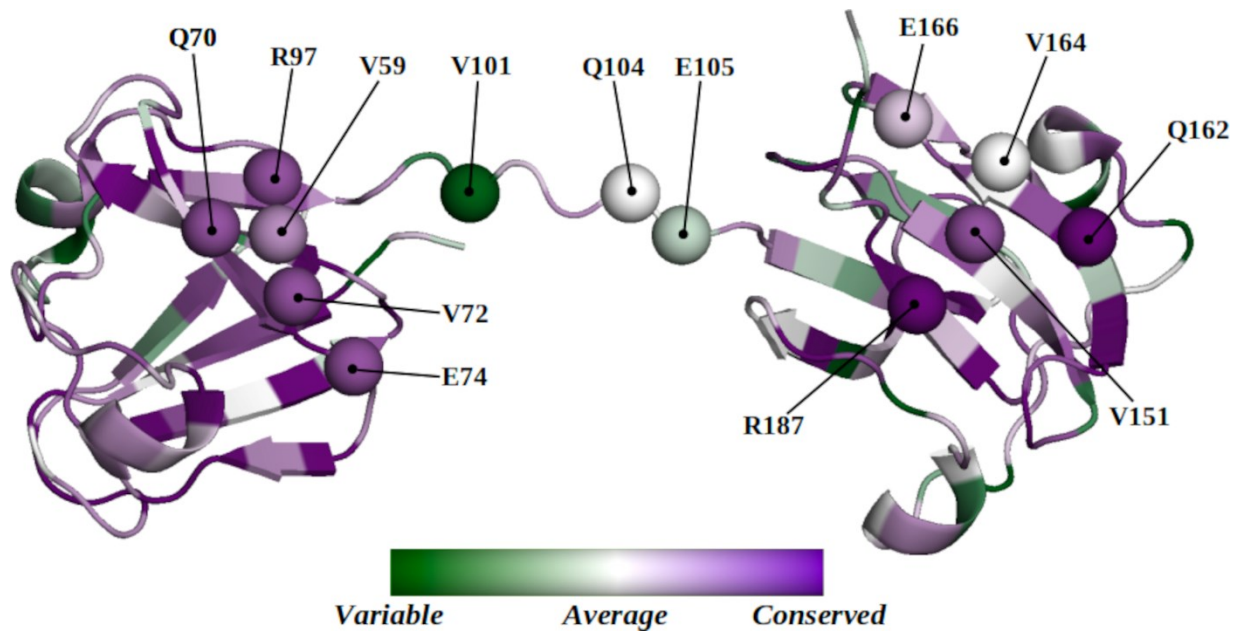
Figure S3. Projection into the one monomer of the crystallographic structure of the Human βB2-Crystallin (PDB ID: 1YTQ) of the conserved residues, calculated with ConSurf server.
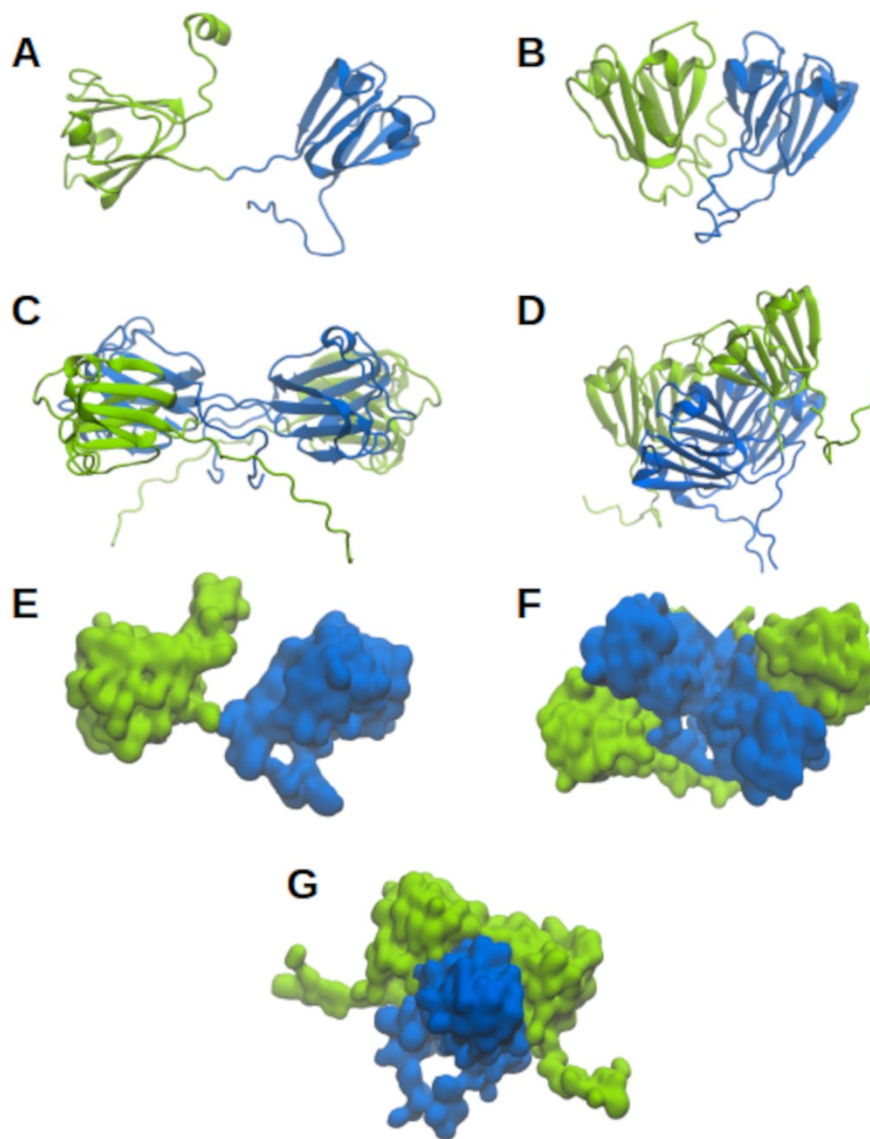
Figure S4. Starting structures for each MD simulations systems. (A) Extended monomeric conformation (EM) of all-atom systems. (B) Closed monomeric conformation (CM) of all-atom systems. (C) Extended dimeric conformation (ED) of all-atom systems. (D) Closed dimeric conformation (CD) of all-atom systems. (E) Extended monomeric conformation (EM) of coarse-grained systems. (F) Extended dimeric conformation (ED) of coarse-grained systems. (G) Closed dimeric conformation (CD) of coarse-grained systems.
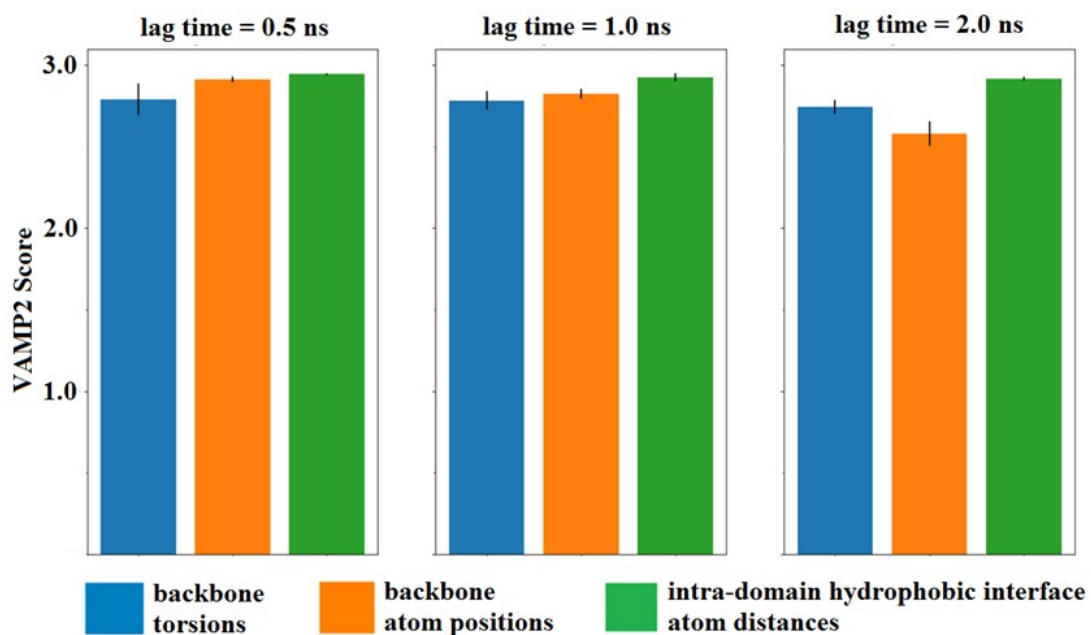
Figure S5. Three features were evaluated with VAMP2 Score at different lag-times. The variational approach for Markov processes (VAMP2) provides a systematic means to quantitatively compare multiple features of the simulation data. From our trajectory analysis, the VAMP2 Score indicates that the featurization that involves the distances between Cα atoms of the residues located at the intra-domain hydrophobic interface (V59, V72, E74, R97, V151, L164, E166, and R187) are sufficient to characterize the conformational changes. This feature contains more kinetic variance than the backbone torsions and backbone atom positions.

Figure S6. Implied time scales of the EM systems of HβB2C. Shown twenty implied time scales of the twenty-one microstates, plott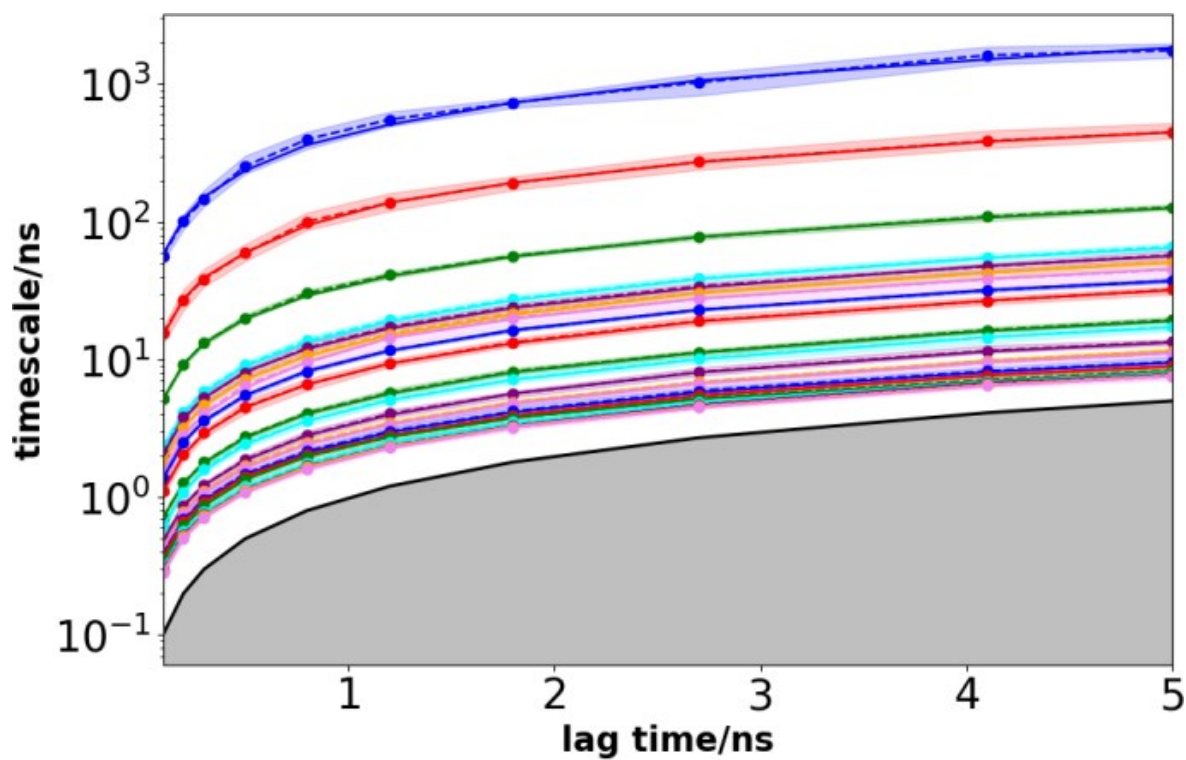ed as a function of lag time. The solid lines correspond to Implied time scales of maximum likelihood, while the means are plotted as dashed lines, and the 95% confidence intervals of the means are depicted as shaded regions. The gray area

indicates the region where timescales horizon below which the MSM cannot resolve process. The implied time scales converged at lag time of 2 ns.
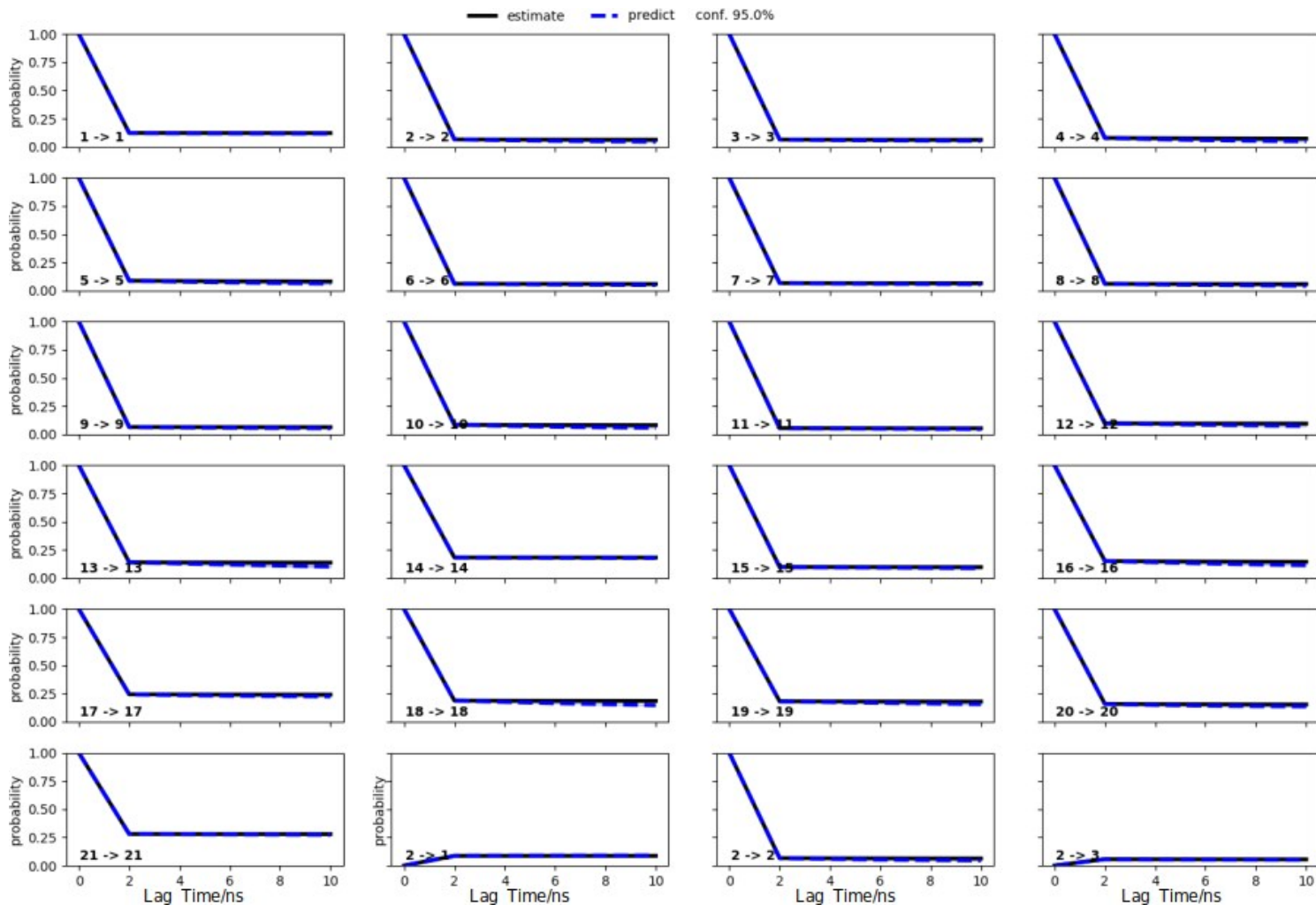

Figure S7. Chapman-Kolmogorov test of the EM systems of HβB2C comparing the probabilities of remaining within each of the twenty-one macrostates as a function 2 ns of lag time.
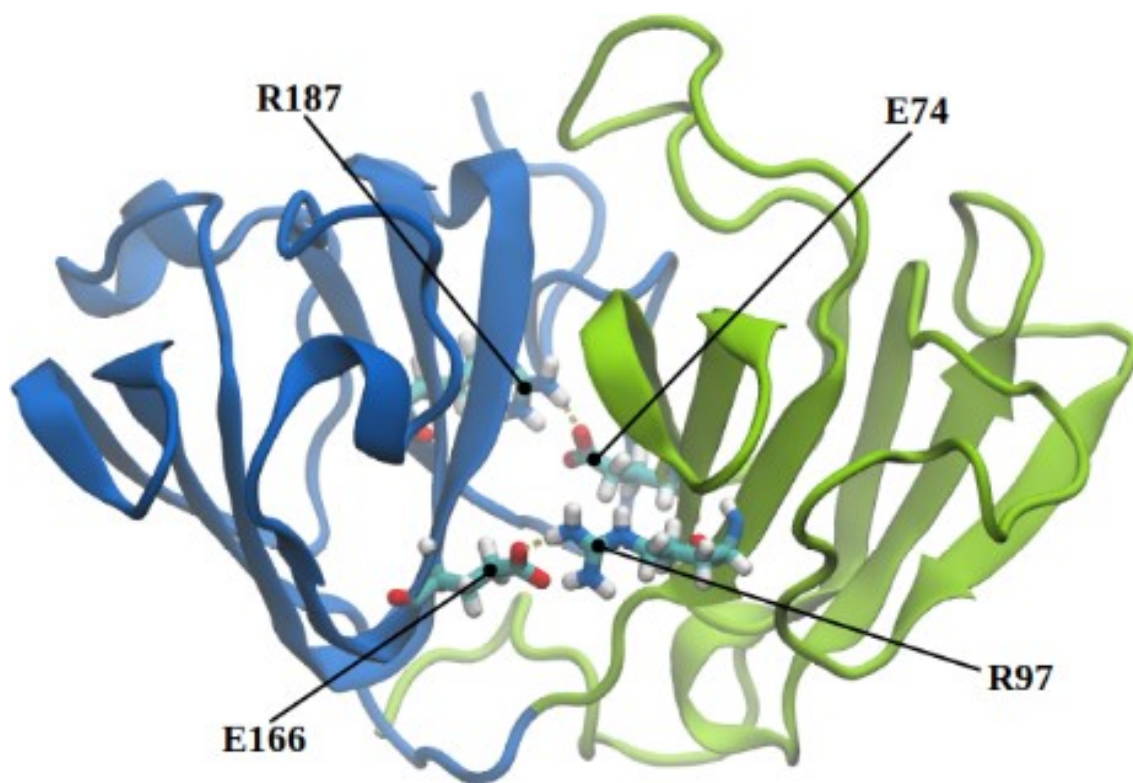
Figure S8. Formation of two salt bridges between residue-pairs E74-R187 and R97-E166 of HβB2C.
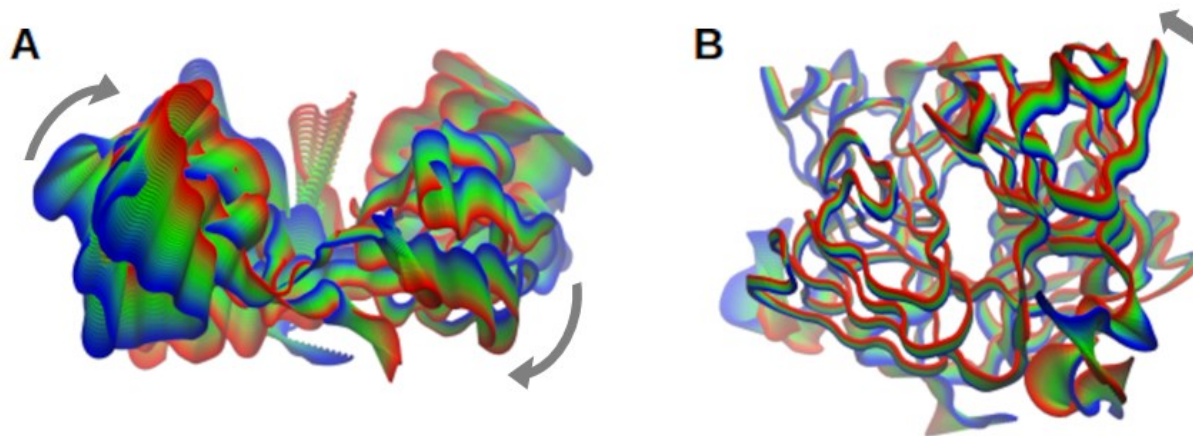
Figure S9. The 3D-projection of the first eigenvector obtained from PCA analysis. (A) Extended dimeric (ED) all-atom system. (B) Closed dimeric (CD) all-atom system. The gray arrows depict domain motions in each system. The extended dimeric all-atom system (A) exhibits a hinge bending of terminal domains with up and down movements. The closed dimeric all-atom system (B) shows a small amplitude domain motions indicating high stability conformation.