

Supporting Information

Exploring the folding process of the human β B2-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	Page
Length and number of replicas of each MD production	S2
Stationary probabilities and relative free energies	S3
General depiction of the Human β B2-Crystallin	S4
Sequence alignment of twelve members of $\beta\gamma$ -crystallin family	S5
Conserved residues projected into the crystallographic structure of the H β B2C	S6
Starting structures for each MD simulations systems	S7
Three features were evaluated with VAMP2 Score at different lag-times	S8
Implied time scales of the EM systems of H β B2C	S9
Chapman-Kolmogorov test of the EM systems of H β B2C	S10
Formation of two salt bridges in H β B2C	S11
The 3D-projection of the first eigenvector obtained from PCA analysis	S12

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

Systems	Replicas	Length (μs)
All-atom		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	Replicas	Length (μs)
Coarse-grain		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 H β B2C.

Macrostate	Stationary probability (π)	Free energy ($k_B T$)
<i>S1</i>	0.006	5.137
<i>S2</i>	0.024	3.748
<i>S3</i>	0.010	4.636
<i>S4</i>	0.014	4.253
<i>S5</i>	0.015	4.204
<i>S6</i>	0.018	4.015
<i>S7</i>	0.012	4.418
<i>S8</i>	0.017	4.093
<i>S9</i>	0.013	4.303
<i>S10</i>	0.021	3.840
<i>S11</i>	0.016	4.154
<i>S12</i>	0.022	3.832
<i>S13</i>	0.128	2.059
<i>S14</i>	0.025	3.430
<i>S15</i>	0.032	3.690
<i>S16</i>	0.077	2.563
<i>S17</i>	0.124	2.089
<i>S18</i>	0.103	2.270
<i>S19</i>	0.099	2.308
<i>S20</i>	0.070	2.661
<i>S21</i>	0.154	1.870

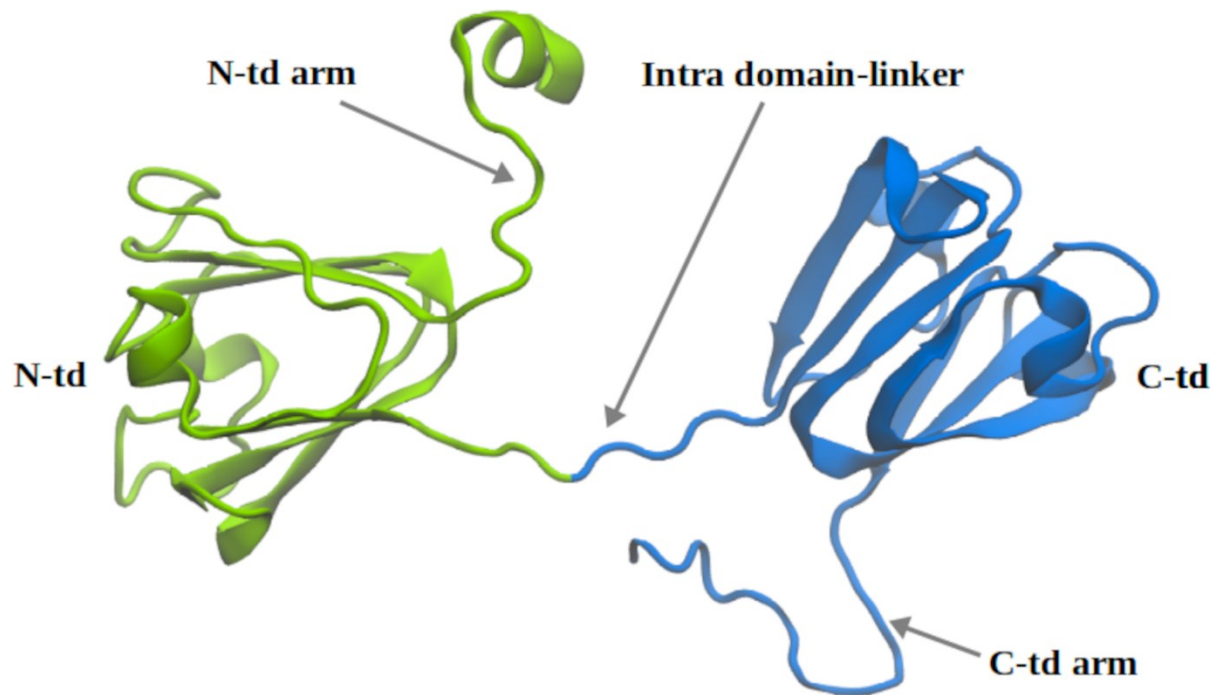


Figure S1. General depiction of the Human β B2-Crystallin in an extended conformation. This protein exhibits a N-terminal domain (N-td), C-terminal domain (C-td), N-td and C-td arms, and the intra-domain linker.

```

001 sp|P07316|CRGB_HUMAN  - - - - - M G K I T F Y E D R A F Q G R S Y E C T T D C P N L Q P - - Y F S R C N S I R V E S G
002 sp|P26998|CRBB3_HUMAN  S H G D L G G S Y K V I L Y E L E N F Q G K R C E L S A E C P S L T D S - L L E K V G S I Q V E S G
003 sp|P05813|CRBA1_HUMAN  P T P G S L G P W K I T I Y D Q E N F Q G K R M E F T S S C P N V S E R - S F D N V R S L K V E S G
004 sp|P07315|CRGC_HUMAN  - - - - - M G K I T F Y E D R A F Q G R S Y E T T D C P N L Q P - - Y F S R C N S I R V E S G
005 sp|P62697|CRBB2_RAT    P Q - - - P L N P K I I F E Q E N F Q G H S H E L S G P C P N L K E T - G M E K A G S V L V Q A G
006 sp|P11844|CRGA_HUMAN  - - - - - M G K I T F Y E D R D F Q G R C Y N C I S D C P N L R V - - Y F S R C N S I R V D S G
007 sp|P53673|CRBA4_HUMAN  Q C T K S A G P W K M V V W D E D G F Q G R R H E F T A E C P S V L E L - G F E T V R S L K V L S G
008 sp|Q8WXF5|CRGN_HUMAN  - - - H A Q R S G K I T L Y E G K H F T G Q K L E V F G D C D N F Q D R G F M N R V S I H V E S G
009 sp|P53674|CRBB1_HUMAN  A A E L P P G N Y R L V F E L E N F Q G R R A E F S G E C S N L A D R - G F D R V R S I I V S A G
010 sp|P07320|CRGD_HUMAN  - - - - - M G K I T L Y E D R G F Q G R H Y E C S S D H P N L Q P - - Y L S R C N S A R V D S G
011 sp|P22914|CRYGS_HUMAN  - - - H S K T G T K I T F Y E D K N F Q G R R Y D C D C D C A D F H T - - Y L S R C N S I K V E G G
012 sp|P43320|CRBB2_HUMAN  P Q - - - S L N P K I I F E Q E N F Q G H S H E L N G P C P N L K E T - G V E K A G S V L V Q A G

```

```

001 sp|P07316|CRGB_HUMAN  W Y h h h Y W h
002 sp|P26998|CRBB3_HUMAN  C W H I Y E R P N Y Q G H Q Y F L R R G E Y P D Y Q Q W M G - - - L S D S I R S C C L I P P - H S
003 sp|P05813|CRBA1_HUMAN  P W L A F E S R A F R G E Q F V L E K G D Y P R W D A W S - - N S R D S D S L L S L R P L N I D S P
004 sp|P07315|CRGC_HUMAN  C W M L Y E R P N Y Q G Q Q Y F L R R G E Y P D Y Q Q W M G - - - L S D S I R S C C L I P - Q T
005 sp|P62697|CRBB2_RAT    P W V G Y E Q A N C K G E Q Y V F E K G E Y P R W D S W T - - S S R R T D S L S L R P I K V D S Q
006 sp|P11844|CRGA_HUMAN  C W M L Y E R P N Y Q G H Q Y F L R R G K Y P D Y Q H W M G - - - L S D S V Q S C R I I P - H T
007 sp|P53673|CRBA4_HUMAN  A W V G F E H A G F Q G Q Q Y I L E R G E Y P S W D A W G G N T A Y P A E R L T S C R P A A C A N H
008 sp|Q8WXF5|CRGN_HUMAN  A W V C F N H P D F R G Q Q Y F L E H G D Y P D F F R W N S - - - H S D H M G S C R P V G - M H G
009 sp|P53674|CRBB1_HUMAN  P W V A F E Q S N F R G E M F I L E K G E Y P R W N T W S - - S S Y R S D R L M S F R P I K M D A Q
010 sp|P07320|CRGD_HUMAN  C W M L Y E R P N Y S G L Q Y F L R R G D Y A D H Q Q W M G - - - L S D S V R S C R L I P - H S
011 sp|P22914|CRYGS_HUMAN  T W A V Y E R P N F A G L Y M I L P Q G E Y P E Y Q R W M G - - - L N D R L S S C R A V A H L P S G
012 sp|P43320|CRBB2_HUMAN  P W V G Y E Q A N C K G E Q Y V F E K G E Y P R W D S W T - - S S R R T D S L S L R P I K V D S Q

```

```

001 sp|P07316|CRGB_HUMAN  G A Y R M K I Y D R D E L R G Q M S E L T D D C I S V Q D R F H L T - - E I H S L N V L E G S W I L
002 sp|P26998|CRBB3_HUMAN  - H K L H L F E N P A F S G R K M E I V D D D V P S L W A H G F Q - D R V A S V R A I N G T W V G
003 sp|P05813|CRBA1_HUMAN  K E S K M T I F E K E N F I G R Q W E I S - D D Y P S L Q A M G W F N N E V G S M K I Q S G A W V C
004 sp|P07315|CRGC_HUMAN  V S H R L R L Y E R E D H K G L M M E L S E D C P S I Q D R F H L S - - E I R S L H V L E G C W V L
005 sp|P62697|CRBB2_RAT    - E H K I I L Y E N P N F T G K K M E I V D D D V P S F H A H G Y Q - E K V S V R V Q S G T W V G
006 sp|P11844|CRGA_HUMAN  S S H K L R L Y E R D D Y R G L M S E L T D D C A C V P E L F R L P - - E I Y S L H V L E G C W V L
007 sp|P53673|CRBA4_HUMAN  R D S R L T I F E G E N F L G K K G E L S - D D Y P S L Q A M G W E G N E V T S F H V H S G A W V C
008 sp|Q8WXF5|CRGN_HUMAN  E H F R L E I F E G C N F T G Q C L E F L E D S P F L Q S R G W V K N C V N T I K V Y G D G A A W S
009 sp|P53674|CRBB1_HUMAN  - E H K I S L F E G A N F K G N T I E I Q G D D A P S L W V Y G F S - D R V G S V K V S S G T W V G
010 sp|P07320|CRGD_HUMAN  G S H R I R L Y E R E D Y R G Q M I E F T E D C S C L Q D R F R F N - - E I H S L N V L E G S W V L
011 sp|P22914|CRYGS_HUMAN  G Q Y K I Q I F E K D F S G Q M Y E T T E D C P S I M E Q F H M R - - E I H S C K V L E G V W I F
012 sp|P43320|CRBB2_HUMAN  - E H K I I L Y E N P N F T G K K M E I T D D D V P S F H A H G Y Q - E K V S V R V Q S G T W V G

```

```

001 sp|P07316|CRGB_HUMAN  Y Y h h h h
002 sp|P26998|CRBB3_HUMAN  Y E F P G Y R G R Q Y L L R P - - - G E Y R R F L D W G - - A P N A K V G S L R R V M D L Y - -
003 sp|P05813|CRBA1_HUMAN  Y E F P G Y R G R Q Y V F E R - - - G E Y R H W N E W D - - A S Q P Q L Q S V R R I R D Q K W H K
004 sp|P07315|CRGC_HUMAN  Y E L P N Y R G R Q Y L L R P - - - Q E Y R R C Q D W G - - A M D A K A G S L R R V V D L Y - -
005 sp|P62697|CRBB2_RAT    Y Q Y P G Y R G L Q Y L L E K - - - G Y K D N S D F G - - A P H P Q V G S V R R I R D M Q W H Q
006 sp|P11844|CRGA_HUMAN  Y E M P N Y R G R Q Y L L R P - - - G D Y R R Y H D W G - - G A D A K V G S L R R V T D L Y - -
007 sp|P53673|CRBA4_HUMAN  S Q F P G Y R G F Q Y V L E C D H H S G D Y K H F R E W G S H A P T F Q V Q S I R R I Q Q - -
008 sp|Q8WXF5|CRGN_HUMAN  P R S F G A E D F Q L S S L Q S D Q G P E E A T T K P A T T Q P P F L T A N L - - - - -
009 sp|P53674|CRBB1_HUMAN  Y Q Y P G Y R G R Q Y L L E P - - - G D F R H W N E W G - - A F Q P Q M Q S L R R L R D K Q W H L
010 sp|P07320|CRGD_HUMAN  Y E L S N Y R G R Q Y L L M P - - - G D Y R R Y Q D W G - - A T N A R V G S L R R I I D F S - -
011 sp|P22914|CRYGS_HUMAN  Y E L P N Y R G R Q Y L L D K - - - K E Y R K P I D W G - - A A S P A V Q S F R R I V E - -
012 sp|P43320|CRBB2_HUMAN  Y Q Y P G Y R G L Q Y L L E K - - - G D Y K D S S D F G - - A P H P Q V Q S V R R I R D M Q W H Q

```

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

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

Figure S2. Sequence alignment of twelve members of β -crystallin family. The sequence alignment was made with ClustalW server and colored according to residue conservation calculated with ConSurf server.

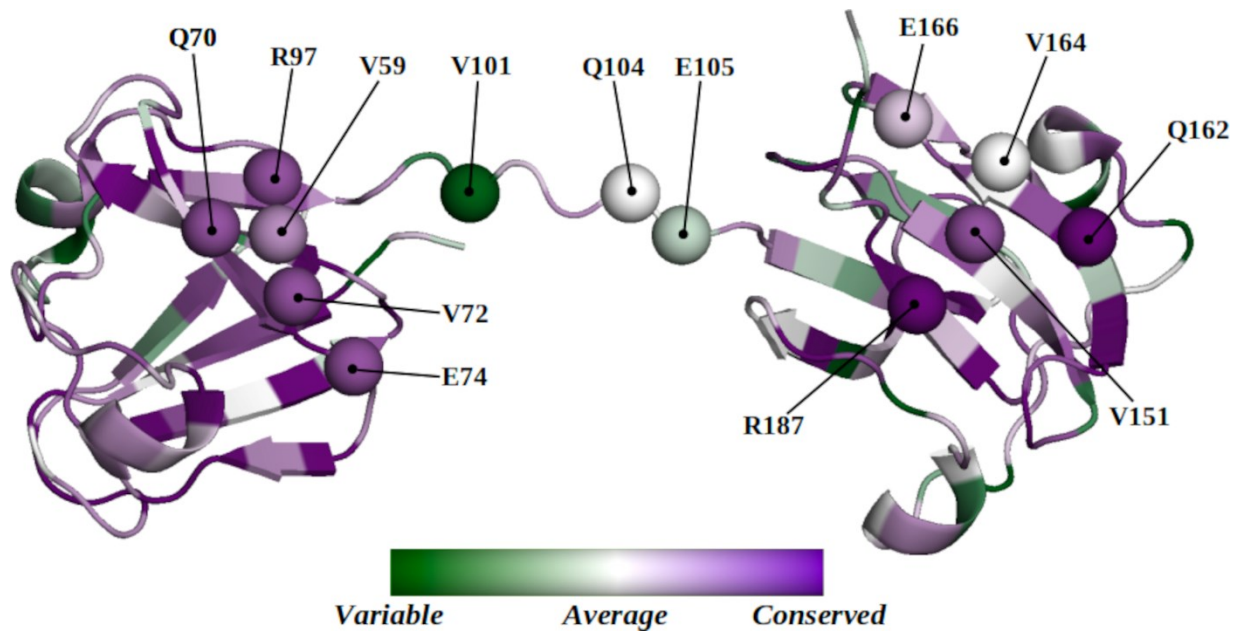


Figure S3. Projection into the one monomer of the crystallographic structure of the Human β B2-Crystallin (PDB ID: 1YTQ) of the conserved residues, calculated with ConSurf server.

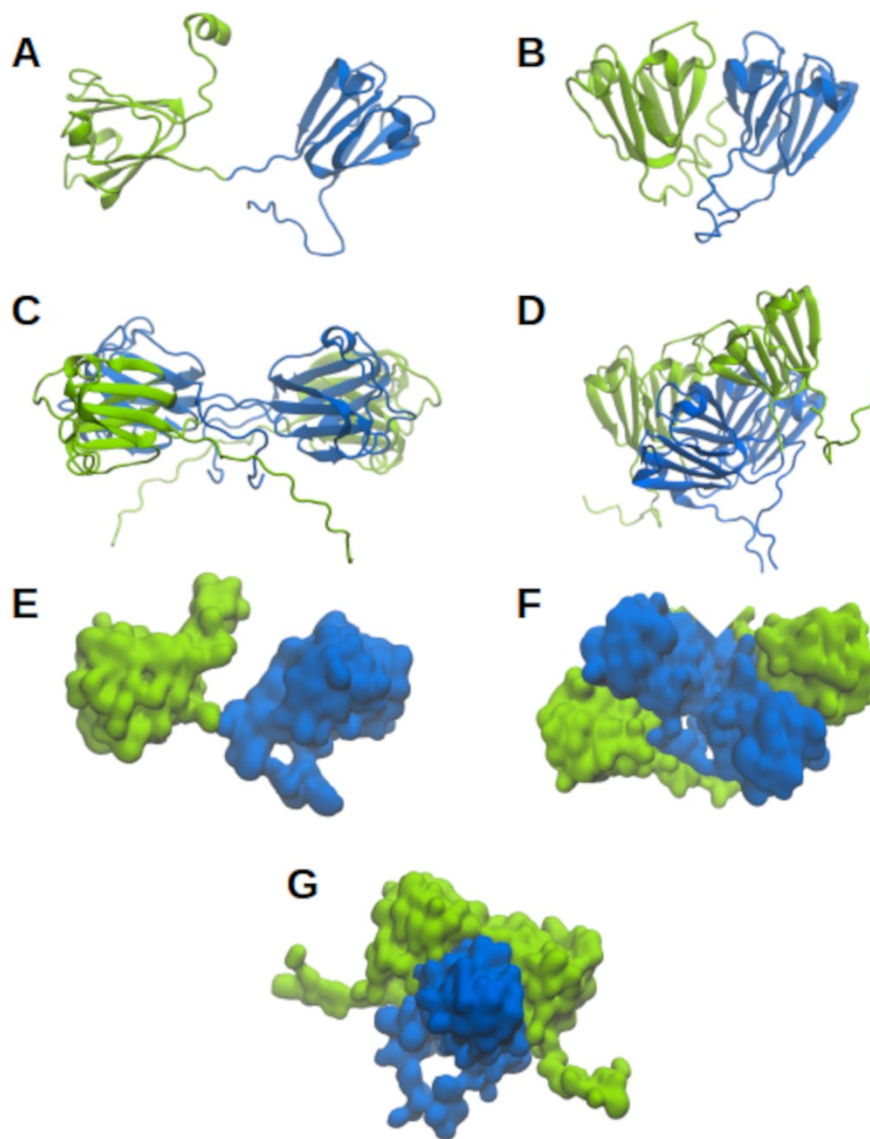


Figure S4. Starting structures for each MD simulations systems. (A) Extended monomeric conformation (EM) of all-atom systems. (B) Closed monomeric conformation (CM) of all-atom systems. (C) Extended dimeric conformation (ED) of all-atom systems. (D) Closed dimeric conformation (CD) of all-atom systems. (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.

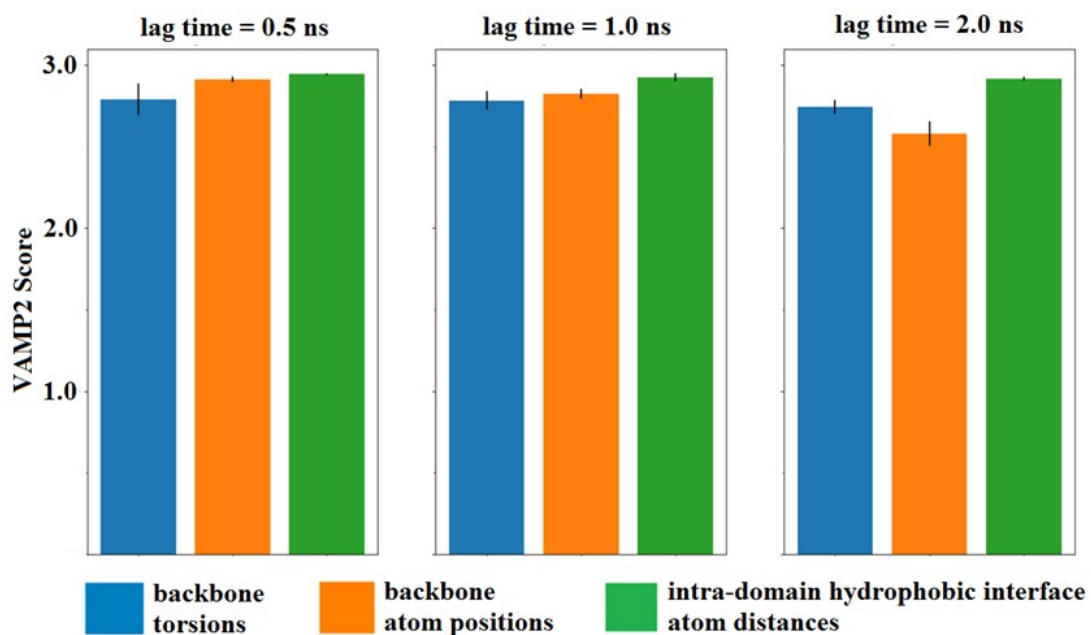


Figure S5. Three features were evaluated with VAMP2 Score at different lag-times. The variational approach for Markov processes (VAMP2) provides a systematic means to quantitatively compare multiple features of the simulation data. From our trajectory analysis, the VAMP2 Score indicates that the featurization that involves the distances between C α 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.

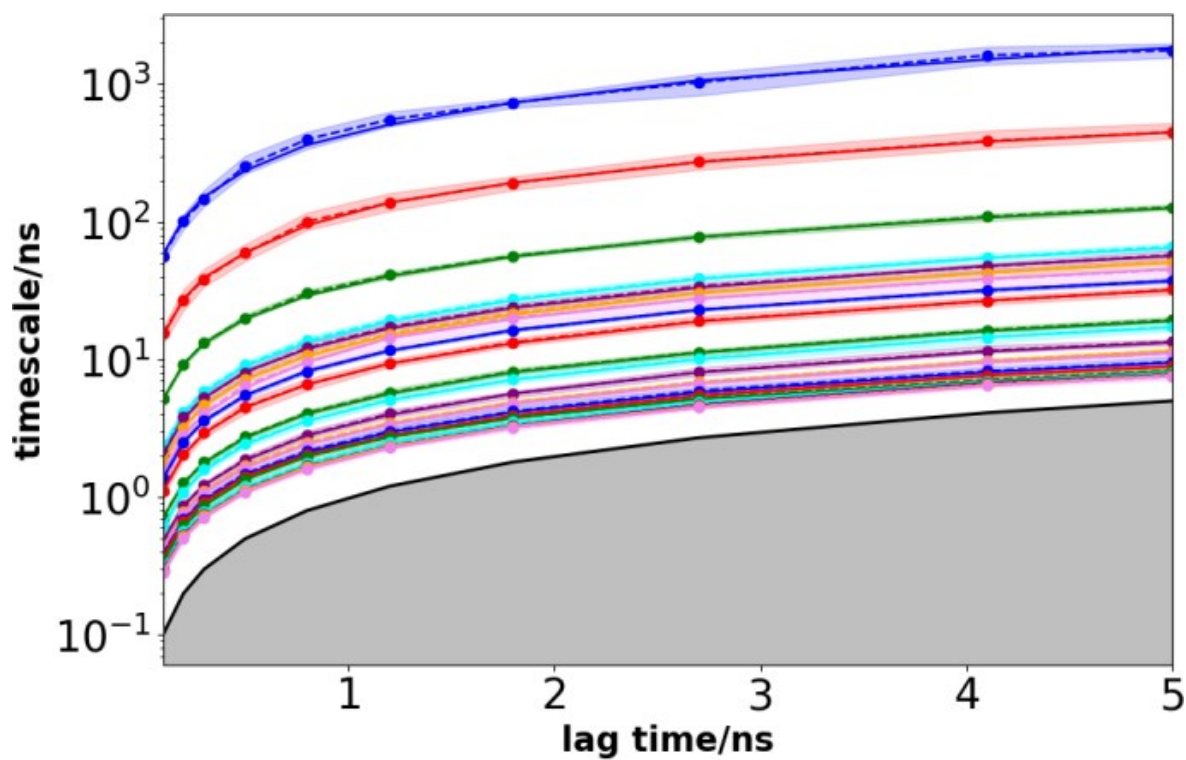


Figure S6. Implied time scales of the EM systems of H β B2C. Shown twenty implied time scales of the twenty-one microstates, plotted as a function of lag time. The solid lines correspond to Implied time scales of maximum likelihood, while the means are plotted as dashed lines, and the 95% confidence 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.

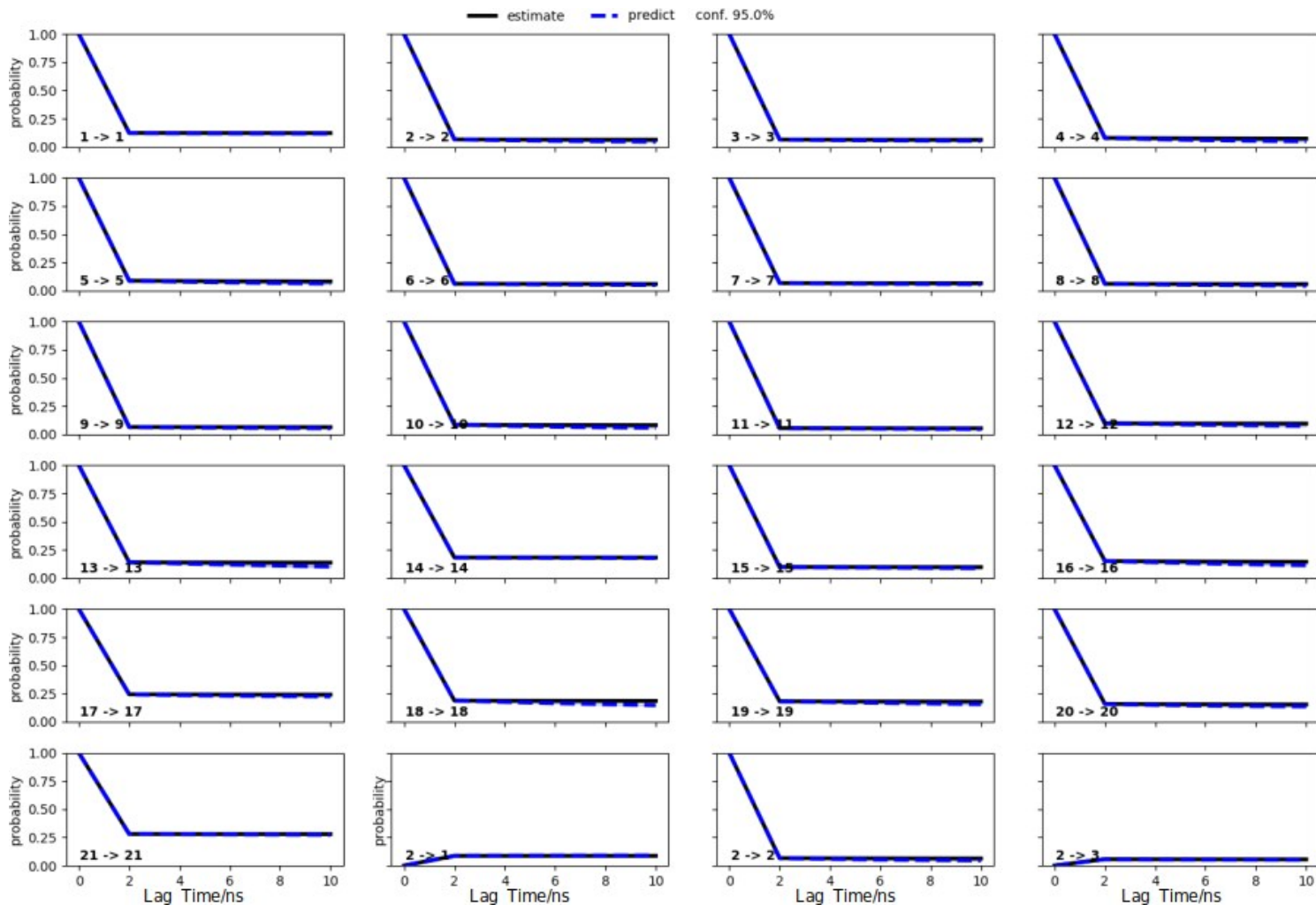


Figure S7. Chapman-Kolmogorov test of the EM systems of H β B2C comparing the probabilities of remaining within each of the twenty-one macrostates as a function 2 ns of lag time.

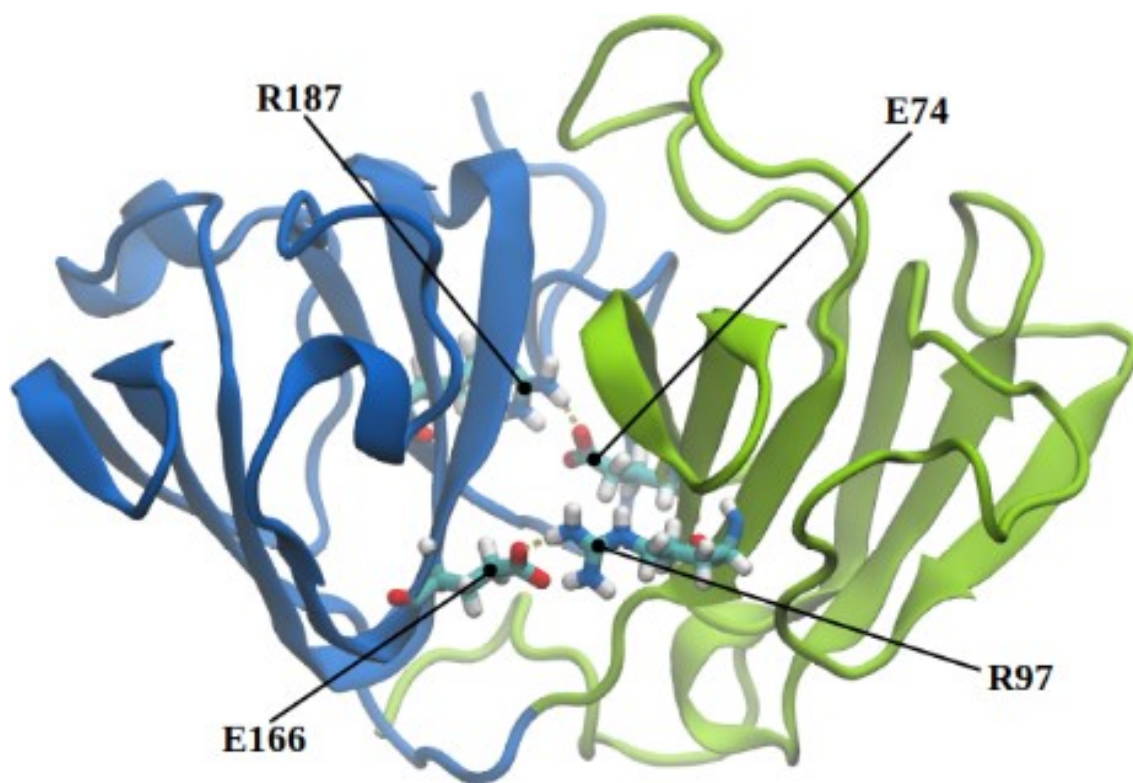


Figure S8. Formation of two salt bridges between residue-pairs E74-R187 and R97-E166 of HβB2C.

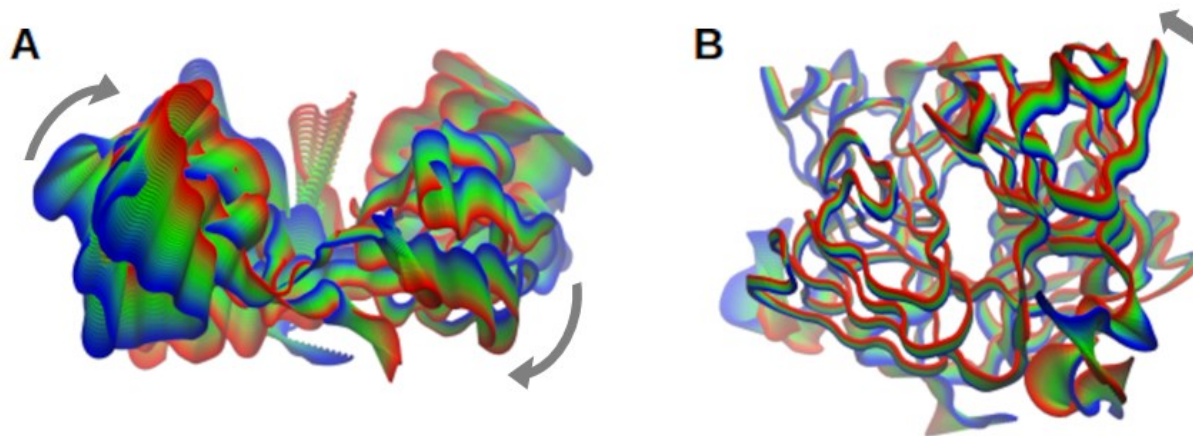


Figure S9. The 3D-projection of the first eigenvector obtained from PCA analysis. (A) Extended dimeric (ED) all-atom system. (B) Closed dimeric (CD) all-atom system. The gray arrows depict domain motions in each system. The extended dimeric all-atom system (A) exhibits a hinge bending of terminal domains with up and down movements. The closed dimeric all-atom system (B) shows a small amplitude domain motions indicating high stability conformation.