Supplementary information

Molecular dynamics analysis of the structural properties of the transglutaminases of Kutzneria albida and Streptomyces mobaraensis

Deborah Giordano [1], Cassiano Langini [2], Amedeo Caflisch [2], Anna Marabotti [3], Angelo Facchiano [2,*]

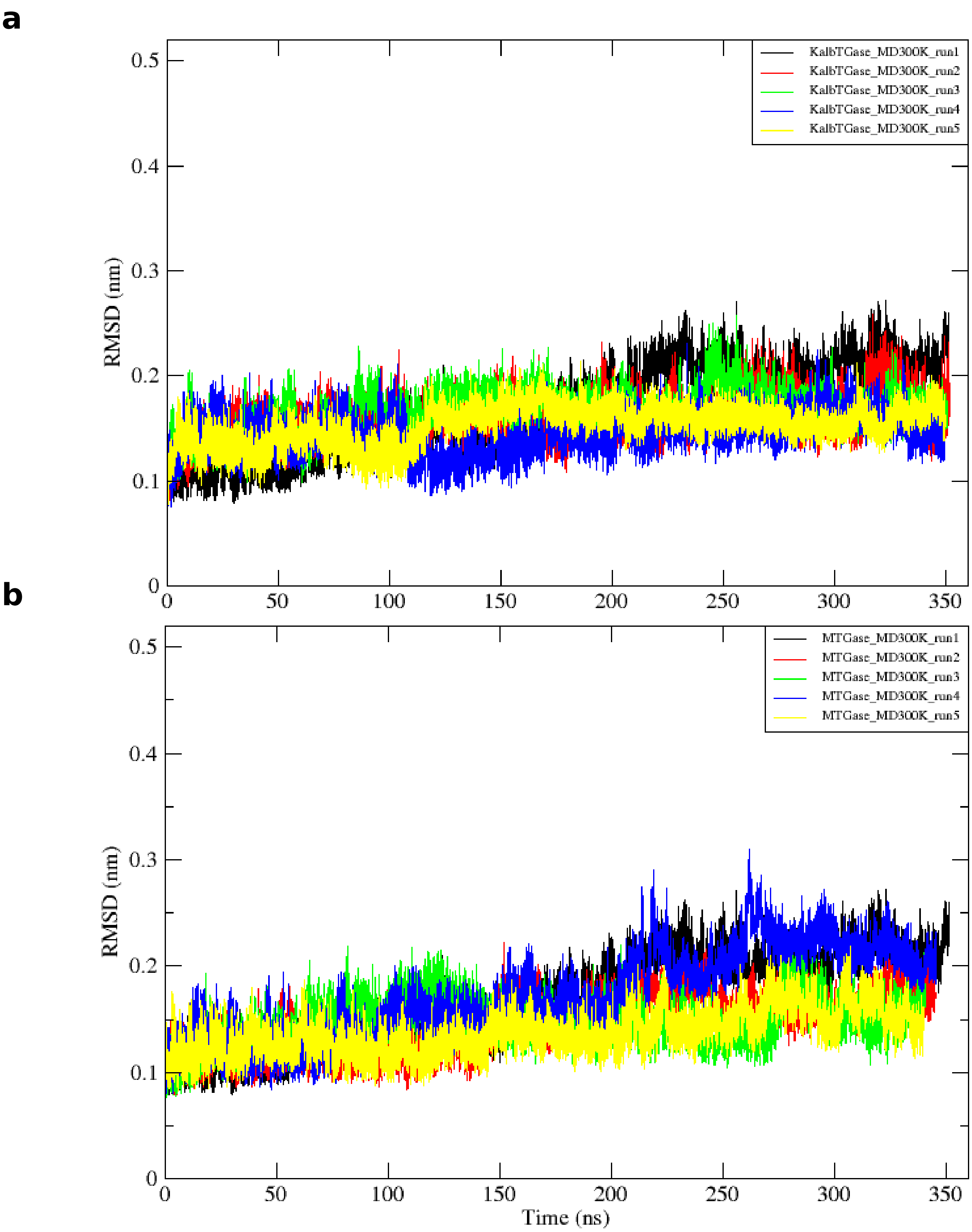
[1] Istituto di Scienze dell'Alimentazione, CNR, via Roma 64, 83100, Avellino, Italy
[2] Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
[3] Department of Chemistry and Biology "A. Zambelli", University of Salerno, via Giovanni Paolo II 132, 84084 Fisciano (SA), Italy

[*] Corresponding Author – angelo.facchiano@isa.cnr.it

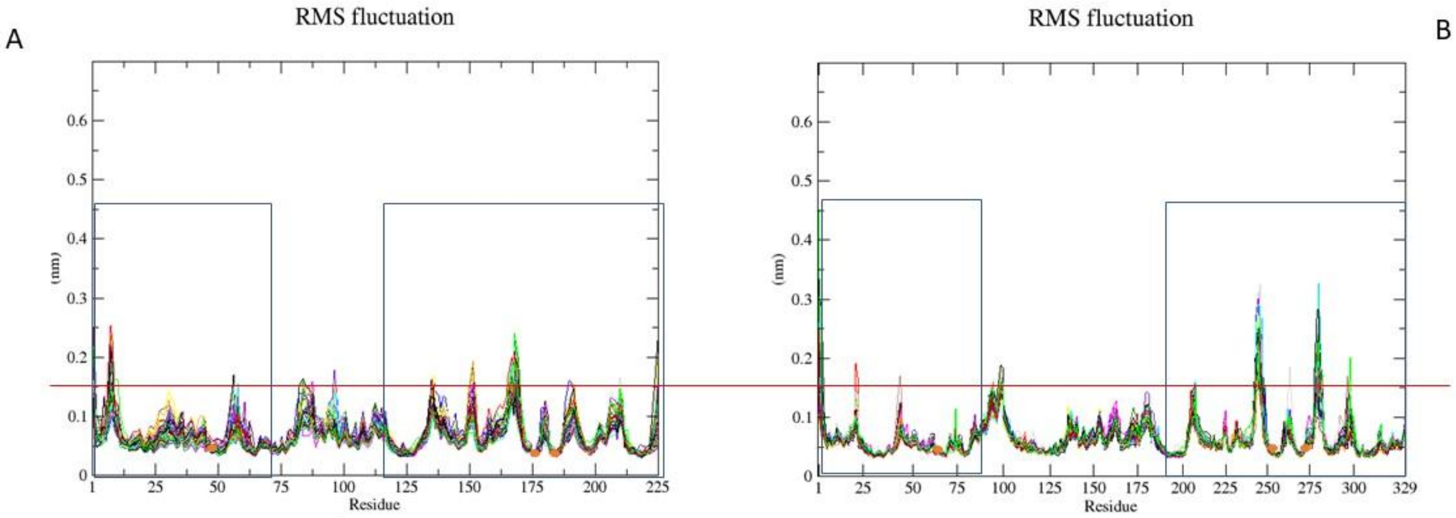
Supplementary Table S1

List of residues whose phi and psi dihedral angles were used as starting features for the conformational analysis. Residue numbering is reported as in the PDB file. Catalytic residues are underlined.

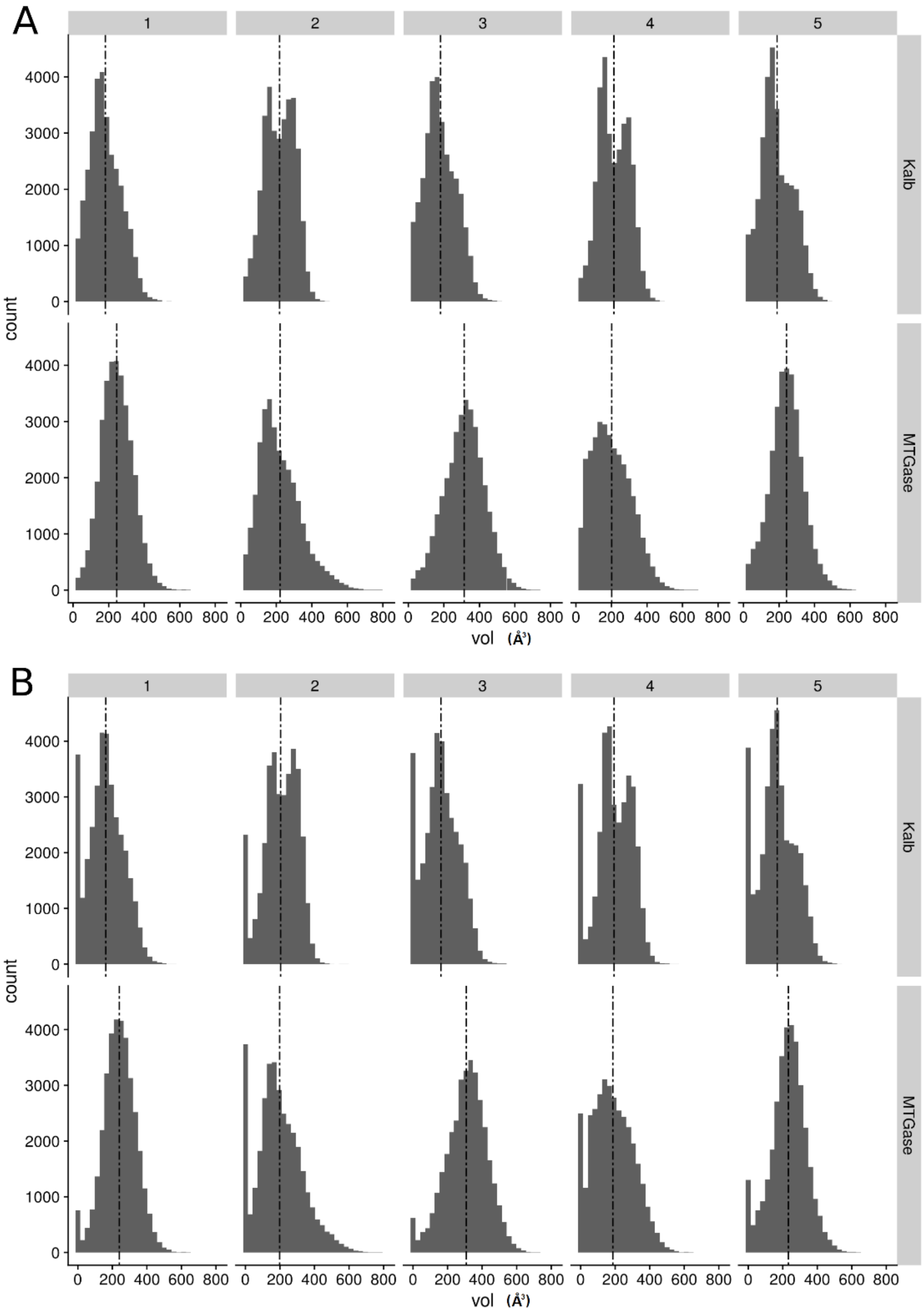
3IU0 (MTGase)				**5M6Q (KalbTGase)**			
residue	position	residue	position	residue	position	residue	position
SER	69	PHE	300	ASP	11	TYR	176
TYR	70	ASP	301	ARG	12	GLY	177
GLY	71	TYR	302	SER	13	TRP	178
ARG	72	GLY	303	TYR	14	PHE	179
ALA	73	TRP	304	ARG	15	ASN	184
GLU	74	PHE	305	TRP	17	THR	185
THR	75	LYS	315	TYR	22	TRP	186
TYR	80	VAL	317	ASN	41	TRP	187
TRP	105	TRP	318	LEU	42	HIS	188
LEU	106	THR	319	ALA	43	ALA	189
SER	107	HIS	320	ARG	44	ASN	190
TYR	108	GLY	321	GLY	45	HIS	191
GLY	109	ASN	322	CYS	46	GLU	192
CYS	110	HIS	323	ILE	47	GLU	193
VAL	111	TYR	324	GLY	48	MET	196
GLY	112	HIS	325	VAL	49	VAL	197
VAL	113	MET	334	THR	50	VAL	198
THR	114	HIS	335	VAL	51	TYR	199
TRP	115	VAL	336	ASN	53	SER	201
ASN	117	TYR	337	PRO	62	THR	202
TYR	121	GLU	338	PHE	128	LEU	203
PRO	122	SER	339	SER	129	HIS	205
ASN	124	LYS	340	LYS	130	TYR	206
TYR	244	ASN	343	ARG	131	SER	207
SER	245	TRP	344	PHE	132	ARG	208
LYS	246	TYR	348	TYR	133	LEU	210
HIS	247	ASP	350	TYR	163	ASP	212
PHE	248	PHE	351	VAL	172	PHE	213
TRP	249	ARG	353	ASN	173	GLU	215
VAL	298	GLY	354	PHE	174	GLN	216
ASN	299	ALA	355	ASP	175	VAL	217



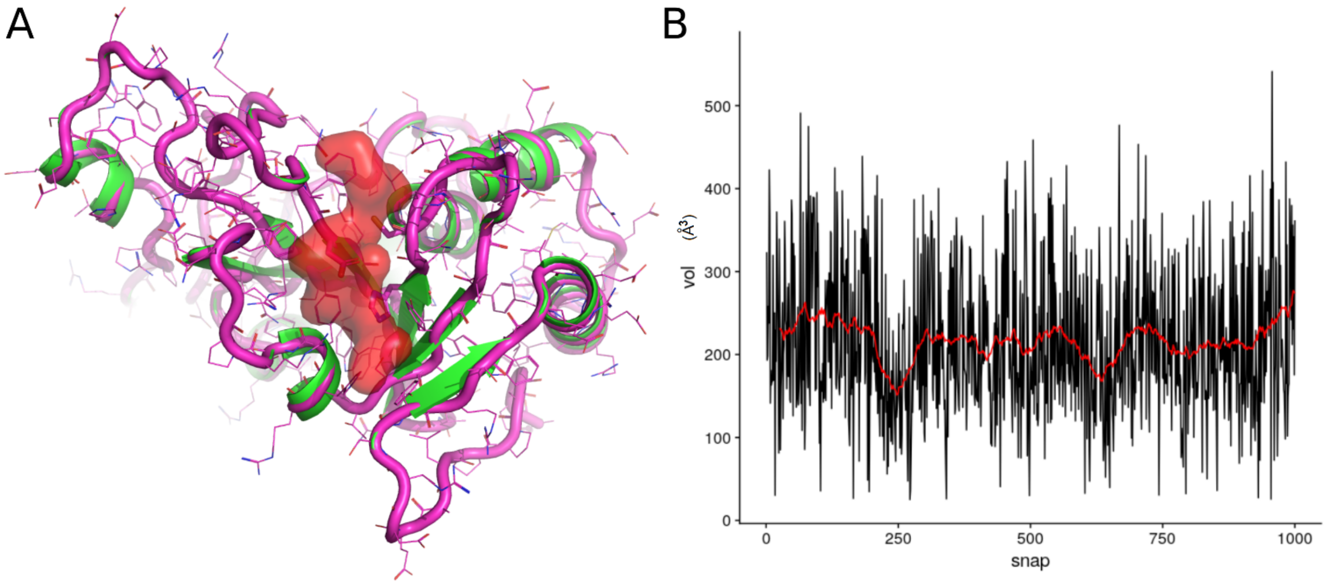
Supplementary Figure S1. RMSD plot of the five different simulation performed at 300K on KalbTGase (a) and MTGase (b).



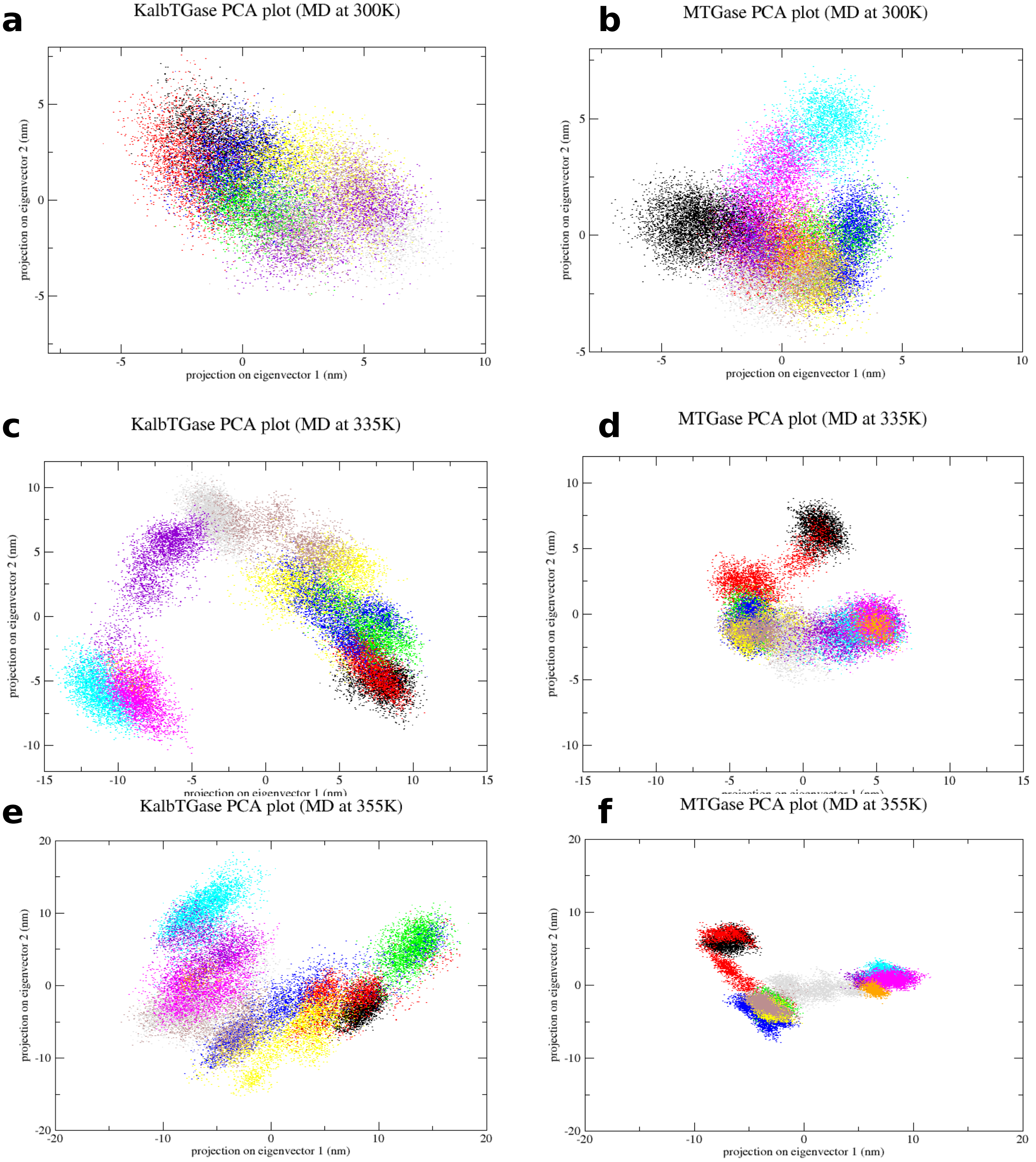
Supplementary Figure S2. Conformational variability of KalbTGase and MTGase at 300 K. RMSF plots for KalbTGase (panel A) and MTGase (panel B) are reported for the last 100 ns of simulation. Each colored line corresponds to a window of 2 ns from the simulation. Blue boxes show the part which can be compared between the two proteins due to a correspondence in secondary structures. Orange dots highlight the position of the catalytic residues in the sequences. Residue numbers assigned by the software starts from the first residue in the PDB file.



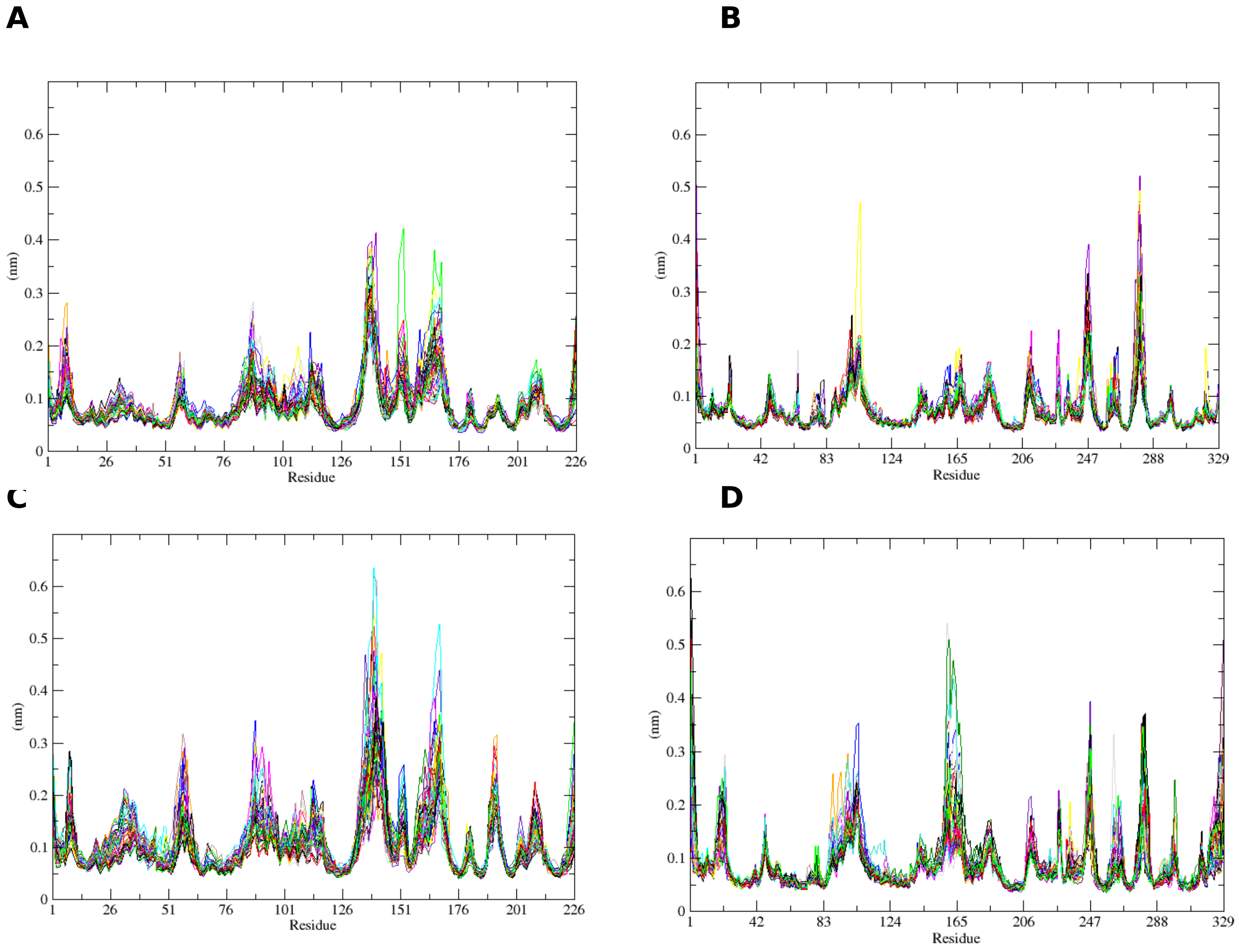
Supplementary Figure S3. distributions for the volume of the catalytic pocket at 300K. A Distributions of the volume of the catalytic pocket calculated with MDpocket on the 5 runs of KalbTGase (top row) and MTGase (bottom row) at 300K. The mean value is reported as a vertical dashed line. Snapshots with a null volume output were excluded (see main text for a detailed discussion on this issue). **B** Same as panel A, but without the removal of the snapshots with null volume. An additional peak at zero volume can be distinguished in every distribution. This is most probably an artifact.



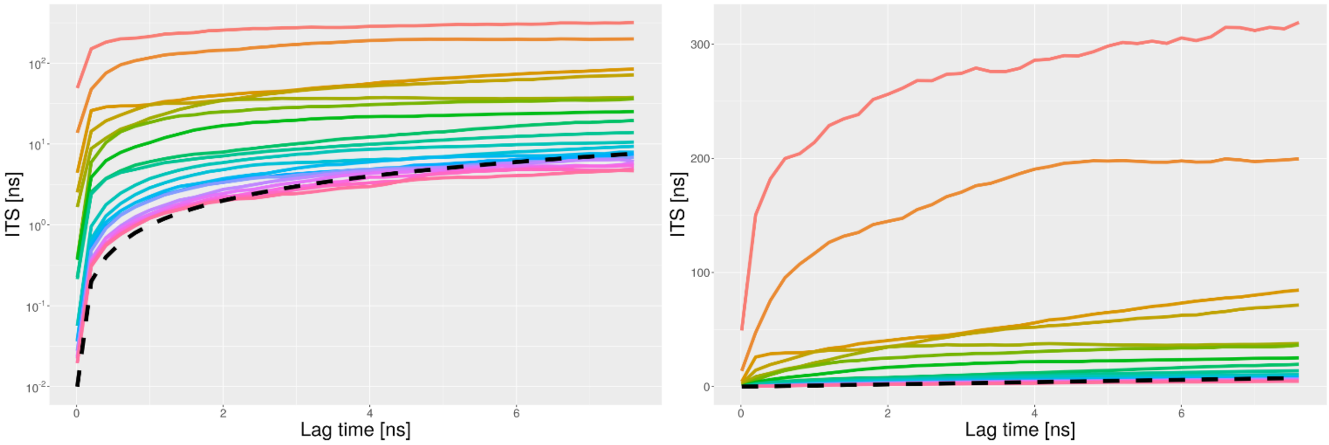
Supplementary Figure S4. Volume calculation with MDpocket. A: Structure of an arbitrary snapshot (magenta) of KalbTGase superimposed to the crystal structure (green). The red surface encloses the active site pocket as identified by MDpocket for the snapshot in magenta. **B:** Time trace of the volume of the active site for the first 1000 snapshots of an MD run on KalbTGase (black line). The moving average (red line) with a window of 50 snapshots, *i.e.* 0.5 ns is used to reduce the noise for visualizing the volume in the SAPPHIRE plots.



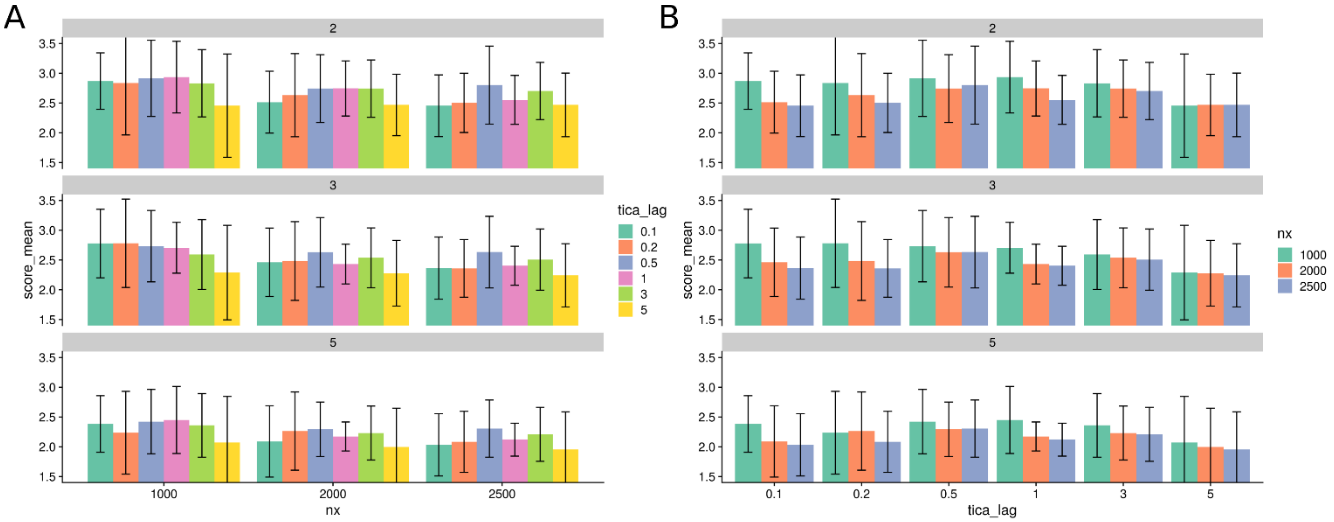
Supplementary Figure S5. PCA plots of MD simulations performed at 300 K, 335 K 355 K of MTGase (panels a, c, e) and of KalbTGase (panels b, d, f). Different colors correspond to different frame cluster collected in a time dependent manner, following the order: black-red-green-blue-yellow-brown-gray-violet-cyan-magenta-orange.



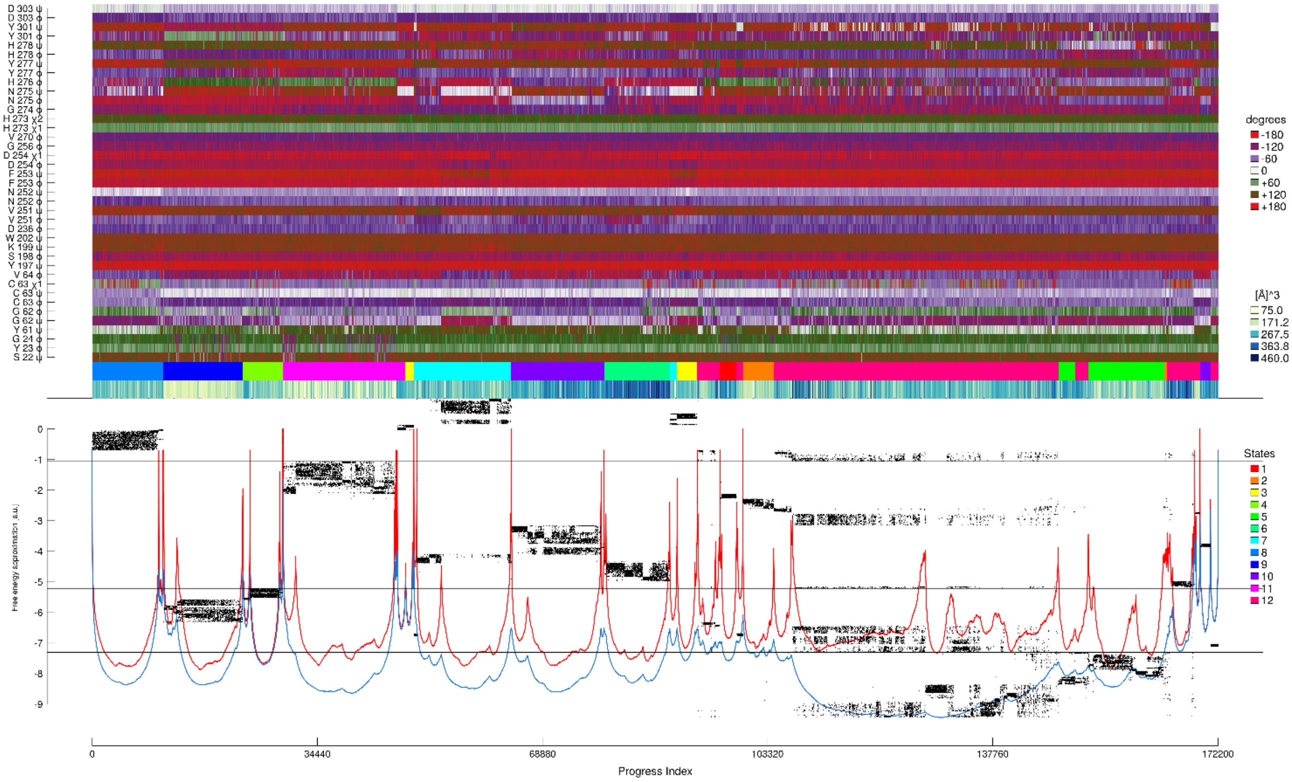
Supplementary Figure S6. RMSF plot of the last 100ns of MD simulations of KalbTGase and MTGase. On the x-axis the protein residues, on the y-axis the RMSF values in nm for KalbTGase MD simulation at 335 K (panel A) and at 355 K (panel C) and for MTGase MD simulation at 335 K (panel B) and at 355 K (panel D). RMSF is reported for the last 100ns of simulation. Each colored line corresponds to an analyzed time window of 2 ns of the simulation. Residue numbers assigned by the software starts from the first residue in the PDB file.



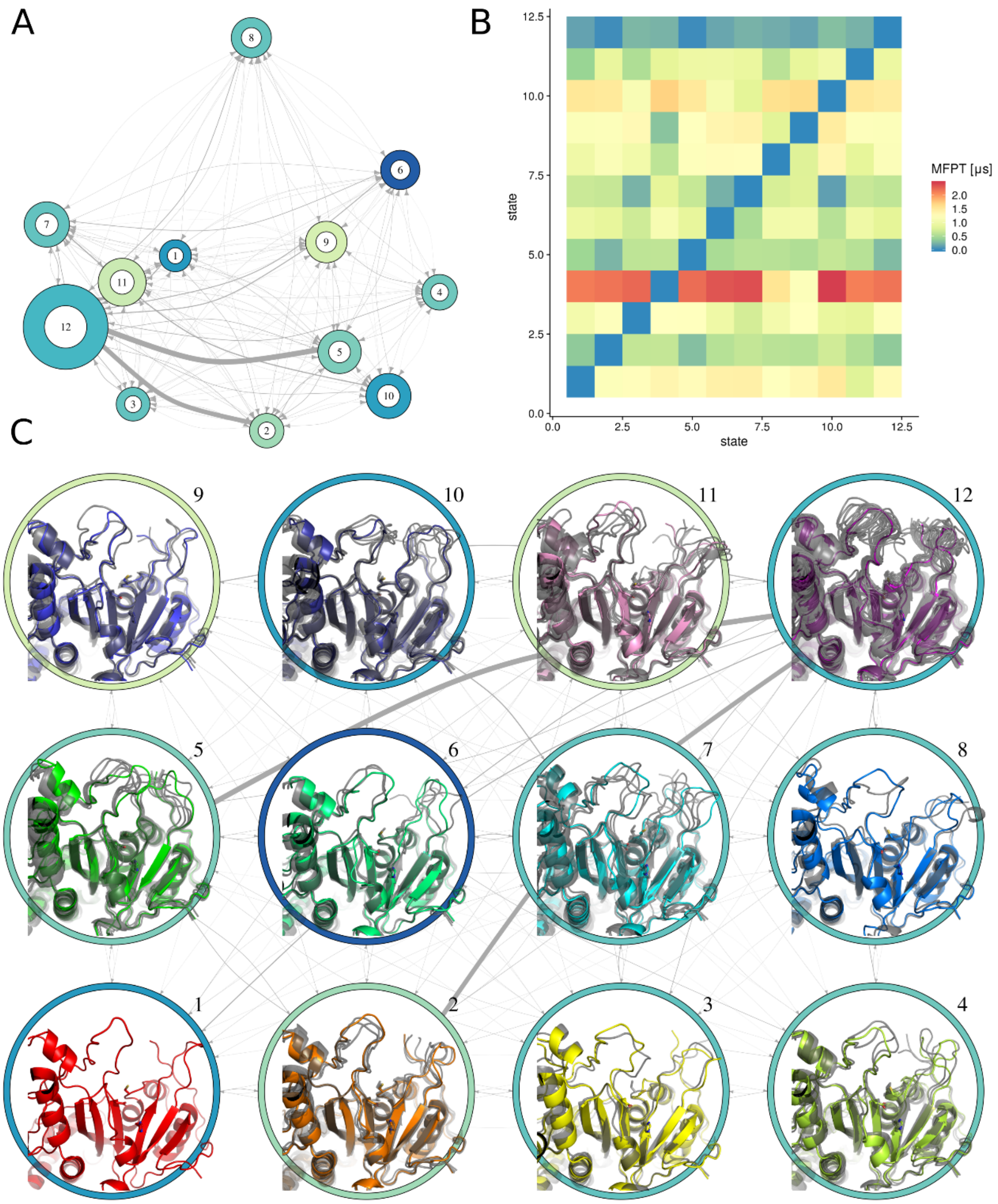
Supplementary Figure S7. Implied timescales. Implied timescales as a function of the MSM lag-time for KalbTGase with logarithmic *y* axis (left) and linear *y* axis (right). The identity line (dashed black) represents the fastest timescale that can be resolved by the model at the different lag-times.



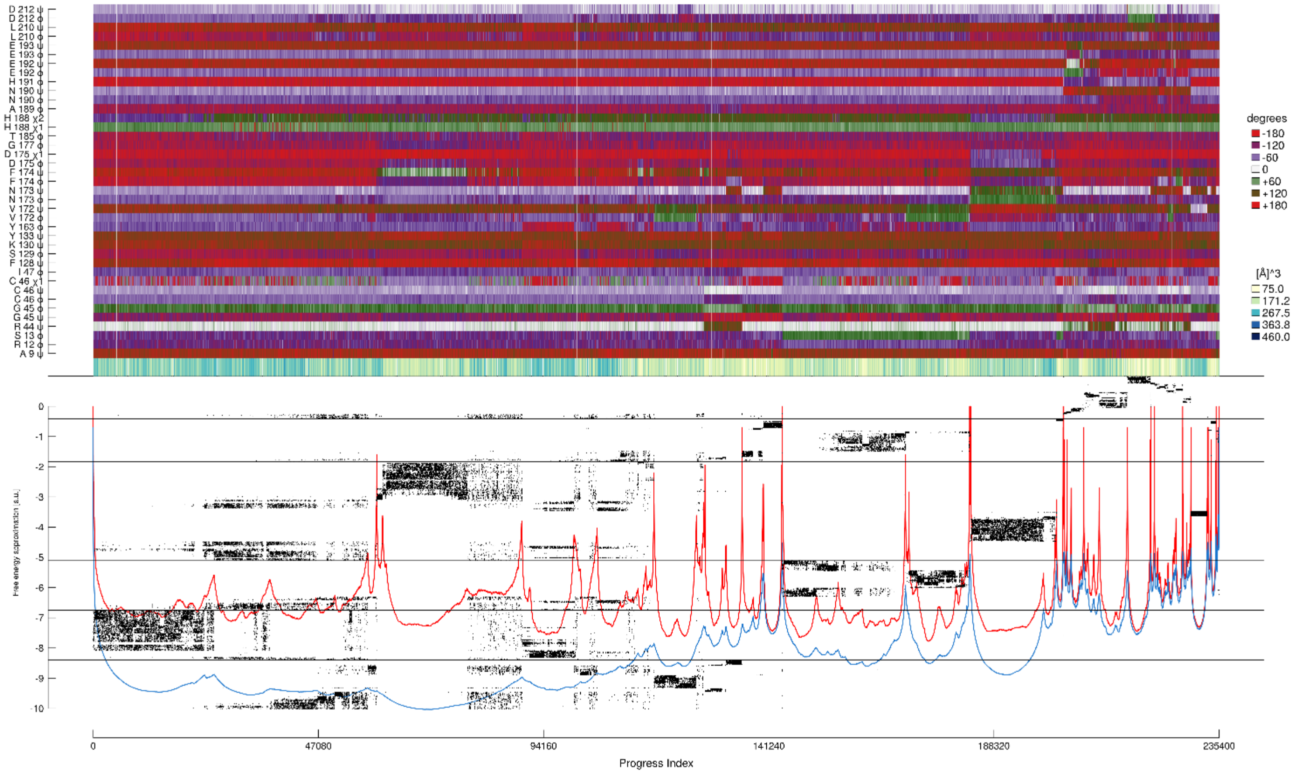
Supplementary Figure S8. VAMP-2 scores for KalbTGase at 300K. Grid search of two hyperparameters (tICA lag-time and *nx*) for building the MSM for KalbTGase at 300K, using the VAMP-2 score. Scores are calculated with 5-fold cross-validation on the 5 runs. The VAMP-2 score method cannot be used to optimize the MSM lag-time, so values for three different MSM lag-times (2, 3, and 5 ns) are reported from top to bottom. Calculated VAMP-2 scores are presented grouped by *nx* and colored by tICA lag-time (panel A), or grouped by tICA lag-time and colored by *nx* (panel B).



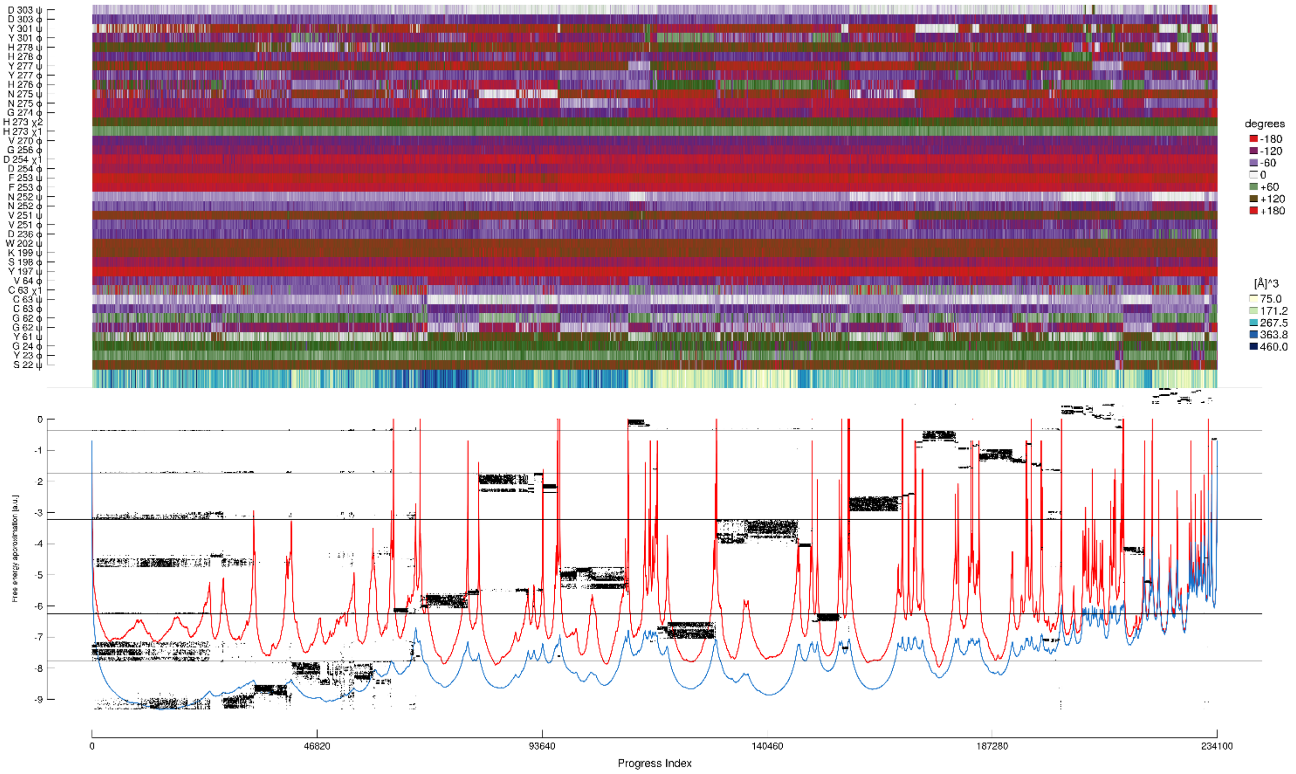
Supplementary Figure S9. SAPPHIRE plot for MTGase at 300K. SAPPHIRE plot for all the sampling of MTGase at 300K. The annotations are the same as in the equivalent SAPPHIRE plot for KalbTGase (Fig. 4 in the main text). Residue numbers assigned by the software starts from the first residue in the PDB file.



Supplementary Figure S10. conformational network for MTGase at 300K. A Conformational network for MTGase at 300K. Node and edge properties are the same as in the corresponding network for KalbTGase (Fig. 5 in the main text). **B** Coarse-grained MFPTs between PCCA+ states. **C** Conformational network with a grid layout. Each node contains the cartoons of the SbC cluster representatives belonging to the relative macrostate. The centroid of the largest cluster is in color, the rest are in gray. Sidechains of the catalytic residues are in sticks. The ring color is the same as in panel A.



Supplementary Figure S11. SAPPHIRE plot of the cumulative sampling for KalbTGase. SAPPHIRE plot with the cumulative sampling of KalbTGase at all temperatures. Trajectories were concatenated in order of increasing temperature.



Supplementary Figure S12. SAPPHIRE plot of the cumulative sampling for MTGase. SAPPHIRE plot with the cumulative sampling of MTGase at all temperatures. Trajectories were concatenated in order of increasing temperature. Residue numbers assigned by the software starts from the first residue in the PDB file.