Computational and Structural Biotechnology Journal 20 (2022) 3924–3934

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

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

Deborah Giordano [a], Cassiano Langini [b], Amedeo Caflisch [b], Anna Marabotti [c], Angelo Facchiano [a,*]

[a] National Research Council, Institute of Food Science, via Roma 64, 83100 Avellino, Italy
[b] Department of Biochemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
[c] Department of Chemistry and Biology "A. Zambelli", University of Salerno, via Giovanni Paolo II 132, 84084 Fisciano, SA, Italy

ARTICLE INFO

Article history:
Received 12 April 2022
Received in revised form 11 July 2022
Accepted 11 July 2022
Available online 20 July 2022

Keywords:
Acyltransferase
Transglutaminase
Protein flexibility
Structure-function
Computational biology
Bioinformatics

ABSTRACT

The microbial transglutaminase (TGase) from *Streptomyces mobaraensis* (MTGase) is widely used for industrial applications. However, in the last decades, TGases from other bacteria have been described. We focused our attention on TGase, from *Kutzneria albida* (KalbTGase), recently characterized as more selective than MTGase and proposed for applications in drug delivery. By comparison of the crystallographic structures, the volume of the catalytic site results smaller in KalbTGase. We compared KalbTGase and MTGase structural flexibility by molecular dynamics (MD) simulations at different conditions. KalbTGase is more rigid than MTGase at 300 K, but the catalytic site has a preserved conformation in both structures. Preliminary studies at higher temperatures suggest that KalbTGase acquires enhanced conformational flexibility far from the active site region. The volume of the catalytic active site pocket of KalbTGase at room temperature is smaller than that of MTGase, and decreases at 335 K, remaining stable after further temperature increase. On the contrary, in MTGase the pocket volume continues to decrease as the temperature increases. Overall, the results of our study suggest that at room temperature the enhanced specificity of KalbTGase could be related to a more closed catalytic pocket and lower flexibility than MTGase. Moreover, by preliminary results at higher temperature, KalbTGase structural flexibility suggests an adaptability to different substrates not recognized at room temperature. Lower adaptability of MTGase at higher temperature with a reduction of the catalytic pocket, instead, suggests a reduction of its activity.

1. Introduction

Transglutaminase (TGase) is a class of enzymes widespread in plants, microorganisms, invertebrates, and vertebrates. TGase generally catalyzes acyl transfer reactions from an acyl donor to an acyl acceptor such as from glutamines to amines, or from glutamines to water when no acyl acceptor is present [1–3]. Among all these acyl-transfer reactions, the most exploited one, from the industrial point of view, is the ability to catalyze the formation of cross-links between γ-carboxamide group of glutamine residues (acyl donor) and ε-amino group of lysine residues (acyl acceptor). This reaction allows the use of TGase in the food industry, in the textile industry and in the pharmacological and biotechnological fields. Indeed, cross-links production allows the formation of protein-DNA/protein/polymer conjugates that are employed for research and biotechnology purposes and that induce significant changes in the functional properties of the food matrix, such as gelation, emulsification, foam formation, viscosity, and water retention capacity [3–4].

Although the research on TGases, particularly the ones of microbial origin, is very active, to date the only TGase enzyme widely used for industrial applications is the microbial TGase discovered in 1989 and extracted from *Streptomyces mobaraensis* (MTGase), because of its facility of expression and purification [5].

We reported a classification of microbial TGases, aimed to help the finding of novel forms of this enzyme with potential applications [6]. In particular, a novel microbial TGase extracted from the organism *Kutzneria albida* (KalbTGase), more selective than MTGase, has been described and proposed for applications in drug delivery [7]. Both enzymes are produced in an inactive form, in which the helix of the pro-peptide segment occludes the active site pocket groove, and probably, they also share the same activation mechanism, during which some extracellular proteases cut this

* Corresponding author.
E-mail address: angelo.facchiano@isa.cnr.it (A. Facchiano).

https://doi.org/10.1016/j.csbj.2022.07.024

segment setting the groove free [7]. The mature form of KalbTGase has a molecular weight of 26.4 kDa and is 100 residues smaller than MTGase, whose molecular weight in the active form is 38 kDa.

Although KalbTGase and MTGase have only 28% sequence identity, they are structurally conserved, as it is shown in Fig. 1. The superposition of MTGase and KalbTGase structures highlights a conserved core domain composed of a central β-sheet flanked by α-helices. In the active site pocket, located in a surface depression, it is possible to notice the structural conservation of the active site residues Cys-Asp-His. It is also possible to notice that KalbTGase has very short surface loops, looking more compact than MTGase.

Despite the preservation of the catalytic triad and of the core domain structure, the differences in terms of amino acids composition and compactness affect the specificity of these two protein molecules, which are hence very different. KalbTGase, although possessing basic microbial TGase activity (1.65 units/mg), has low or undetectable activity with many substrates recognized by conventional MTGase, resulting in a higher selectivity. The motif YRYRQ seems to be the best glutamine substrate and the motif RYESK the best lysine substrate of KalbTGase [7]. Due to this high selectivity, it was suggested to use KalbTGase for the production of therapeutic antibody-drug conjugates for enhanced drug delivery [7].

Here we present a computational study investigating differences and similarities between MTGase and KalbTGase structures, based on multiple runs of molecular dynamics (MD) simulations at 300 K. We evaluate the flexibility of the two enzymes, the variations of the volume of the active site and compare their conformational dynamics. Two additional runs at 335 K and 355 K are used to explore in a preliminary way any sensitive temperature-dependent deviations from the observations at 300 K. Our simulations suggest that at room temperature KalbTGase is less flexible than MTGase, and that its catalytic pocket is narrower than the one of MTGase, in agreement with KalbTGase higher specificity. At higher temperature, differences between the two enzymes could be reduced, thus operating conditions might tune the reactivity of KalbTGase towards different substrates.

2. Methods

2.1. MD simulation parameters

MD simulations were performed on the structures of both KalbTGase (PDB code: 5M6Q) [7] and MTGase (PDB code: 3IU0) using GROMACS 5.0 [8]. The structure of MTGase has been modified to remove the amino acids numbered from 9 to 33 in the pdb file, belonging to the signal peptide, thus removing an isolated fragment, considering that the PDB structure lacks further amino acids up to the number 48. The two starting pdb files were prepared for the submission using CHARMM-GUI Simulation Input Generator [9–10] and, after visual inspection of the molecules, by manual editing. The two proteins were then solvated in a cubic box (box volume: 930.934 nm^3 and 557.752 nm^3 for MTGase and KalbTGase, respectively), filled with water (29,992 molecules for MTGase, 17,809 for KalbTGase) and neutralized with chloride and potassium ions placed randomly in the simulation box (final concentration: 150 mM). Ions and water molecules are positioned far from the active sites and do not interact stably with the proteins. Potassium and chloride ions are commonly added to the simulation box in the quantity necessary to compensate the system charges and approximate an average ionic strength resembling

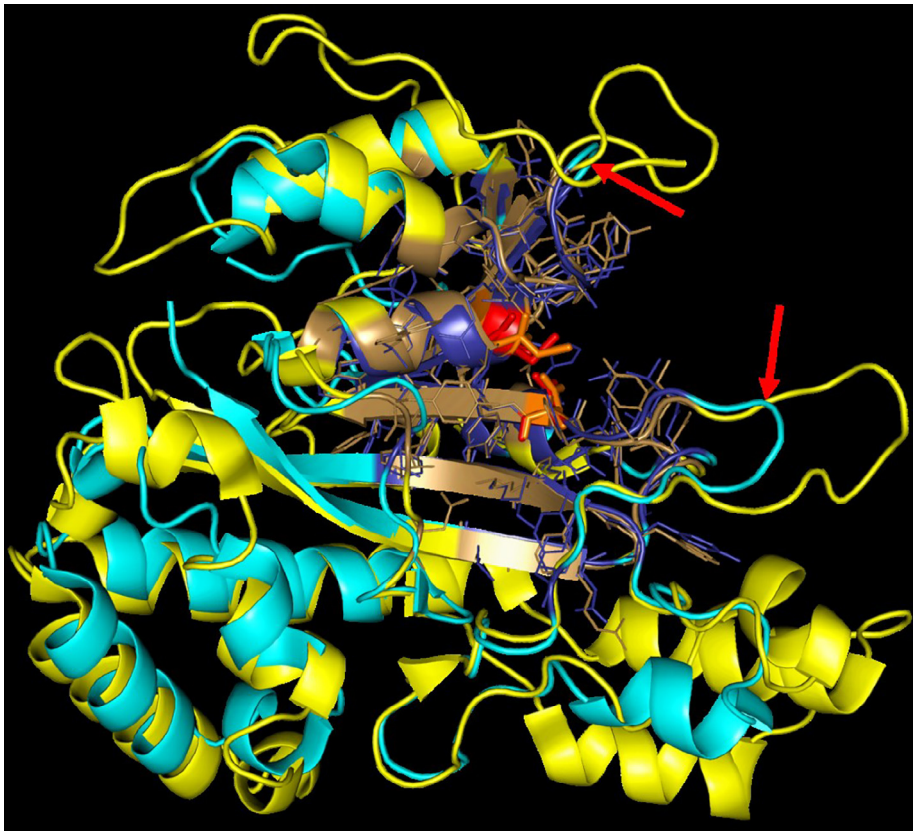

Fig. 1. Structural overlap of KalbTGase and MTGase. KalbTGase and MTGase are represented in cyan and yellow cartoons, respectively. The active site pockets residues used for the analyses are represented as lines (deep blue for KalbTGase, sand for MTGase) with the catalytic triad shown in sticks (orange in KalbTGase and red in MTGase). Red arrows point to the surface loops flanking the active site; the latter are shorter in KalbTGase than MTGase. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the cellular environment. The possible effects of counter-ions' size are considered not relevant for this study, as we are analyzing the differential behaviors of two systems under similar simulation conditions (see under Results and Discussion). For all the preparation and production runs we used the CHARMM36 [11] force field with its modified TIP3P water model [12], chosen for compatibility with the CHARMM force field and according to suggestion of its manuals and documentation. The CHARMM36 force field was used to parametrize the system. The systems were first minimized by applying steepest descent minimization, setting the cut-off for short-range electrostatic and van der Waals interactions to 1.2 nm. Default parameters were used for vdwtype (=Cut-off), vdw-modifier (=Potential-shift-Verlet), coulombtype (=PME), coulomb-modifier (=Potential-shift-Verlet). Minimization stopped when the maximum force reached a value lower than 1000.0 kJ/mol/nm. Equilibration steps with position-restrained MD simulations were run in NVT and NPT conditions for 500 ps, respectively. For the NVT equilibration, the Berendsen thermostat [13] was applied; for NPT equilibration, the Berendsen barostat [13] was added. At the end of the equilibration, for each system, we performed five independent MD simulations of 350 ns in NPT conditions at a temperature of 300 K. Additionally, one 300 ns-long simulation in NPT conditions at the temperature of 335 K and one 300 ns-long simulation in NPT conditions at a temperature of 355 K were also performed to explore in a preliminary way the behavior of the proteins in the wider range of industrial conditions. During the production runs, the Berendsen barostat was replaced with the Parrinello-Rahman barostat [14], and the temperature coupling was obtained using the velocity rescaling thermostat [15]. The equations of motion were integrated using the leap-frog algorithm [16], keeping all bonds constrained with the LINCS algorithm [17]. Long range electrostatic interactions were evaluated using the Particle Mesh Ewald [18] method.

2.2. Analyses of fluctuations

The root mean square fluctuations (RMSF) of atomic coordinates were calculated with GROMACS tool *gmx rmsf* over non-overlapping time windows of 2 ns length. The average value per residue at 300 K was reported on the crystal structures (Fig. 2) using PyMOL open-source [19]. RMSF profiles were plotted using the XMGrace software [20] (supplementary Fig. S2).

The deviation of each run from the starting structure was monitored by calculating the root mean square deviation (RMSD) of atomic coordinates for each snapshot after optimal alignment to the initial structure. Both RMSF and RMSD were calculated only on the Cα atoms. Principal component analysis (PCA) was run on the coordinates of Cα atoms using GROMACS.

2.3. Active site pocket volume analysis

The analysis of the volume of the active site in the crystallographic structures has been performed with the POCASA server (https://altair.sci.hokudai.ac.jp/g6/Research/POCASA_e.html) [21] with the following settings: Probe radius: 2 Å; Single-Point Flag (SPF): 16; (Protein-Depth Flag (PDF): 18; Grid size: 1.0 Å. The analysis of the volume variation of the catalytic pocket during all the MD simulations was tested with two different web servers: Fpocket [22] and MDpocket [23] and finally calculated on the whole trajectory with the desktop version of MDpocket [24]. Fpocket was used to perform the initial pocket detection and its output was a useful reference to prepare the input pocket file needed for the analysis performed by MDpocket on the trajectories. The first step of the analysis consists in the definition of a bounding box surrounding a specific region of the structure, which gets inspected to identify pockets on the protein surface of the reference structure. The identified pockets are tracked along the trajectory by MDpocket, which calculates their volume and other geometric properties (like the solvent accessible surface area or an index of hydrophobicity). It is possible to apply these tools on the whole surface in a completely unsupervised way, but, depending on the protein size, the computational slow-down can be relevant; also, we are only interested in determining the plasticity of the active site. The initial region selection was done by manually selecting a list of residues encompassing the catalytic site of both targets separately. In order for MDpocket to be able to keep track of a specific pocket along the dynamics, the trajectory had to be aligned to the initial reference structure. The α carbon atoms of the whole protein residues were used for alignment. The output of MDpocket was very noisy and this seemed to be related to the small fluctuations of the sidechains that modified the protein surface (see supplementary Fig. S4). This resulted in a large standard deviation of the volume values and there were many snapshots in which the pocket was not recognized and it was assigned a volume of 0.0 $Å^3$. A further discussion about this issue is presented in the result section.

The volume means, mean ranks, and distributions were compared between enzymes and runs using the following statistical tests: Student's *t*-test, Welch's *t*-test, Mann-Whitney *U* test, and Kolmogorov-Smirnov test, with a significance level of 5%. Block averaging of the volume trajectories was used for the comparison of runs. Different tests were used as they have different assumptions, as detailed in the Results and Discussion section.

2.4. Featurization and conformational analysis

In order to have a comprehensive view of the protein conformations, the snapshots from the trajectories at room temperature were analyzed using a SAPPHIRE plot [25] and a Markov State Model (MSM) [26] built on top of it.

The general steps required to construct an MSM are the following: featurization, dimensionality reduction, and clustering [27]. The first step is the choice of a set of features to describe the system under study. As we are mainly interested in the dynamics of the active site, our starting features were the phi and psi backbone dihedral angles (separated into sin and cos components) of a set of 62 residues encompassing the catalytic site, for a total of 248 features. The residues were chosen based on their distance from the catalytic triad and are depicted in Fig. 2, panel B (non-gray residues). The full list is given in Table S1.

The initial features were transformed using time-lagged independent component analysis (tICA) [28–30], with a lag-time of 500 ps, and scaled according to a kinetic mapping [31]. As KalbTGase and MTGase show a similar extent of conformational variation on the single-trajectory timescale (175 ns), in both cases only the first 13 tICA components (accounting for 61% of the total kinetic variance for KalbTGase and 57% for MTGase) were kept for the next steps of the analysis. These roughly account for dynamic modes up to the first large gap in the eigenvalue spectrum of the tICA transform.

In order to identify the residues participating in the slowest conformational transitions, from the tICA decomposition we calculated the absolute value of the correlation between features and tICA independent components (TIC). We considered only correlations with the 13 components that were kept after dimensionality reduction. Additionally, for each feature, correlation absolute values were weighted by the corresponding eigenvalue of the tICA transformation matrix and summed. For each residue we obtained four values (sin and cos components of phi and psi), which were summarized by taking the maximum of the sin and cos values and averaging between phi and psi. The calculated residue-level feature-TIC correlation values could be encoded as B-factors and mapped onto the structure of the enzymes (Fig. 2, panel B).

The tICA features, after dimensionality reduction, were selected to construct the progress index (PI), using a published approximate algorithm [32] based on the pre-clustering of the trajectory snapshots with a tree-based algorithm [33]. Starting from an arbitrary snapshot (in our case the centroid of the largest cluster), the PI consists of a reordering of a trajectory by sequential addition of the snapshot that is closest to any of the already added ones; the pairwise distance between snapshots was calculated as the Euclidean distance between features. The PI groups together snapshots that are in the same free energy basin, and it can be used to get a comprehensive overview of all the configurations sampled. The approximate PI algorithm used 6000 maximum search attempts for the next neighbor in the construction of the minimum spanning tree, with a search depth covering all the 16 levels of the clustering. Additionally, the folding of the leaves was set to 3 in order to prevent snapshots of fringe regions to accumulate towards the end of the basin [34].

The SAPPHIRE plot [25] consists of a collection of different annotations that are plotted along the PI-ordered trajectory (see Fig. 3 and supplementary Fig. S9). The dihedral angles chosen for the geometric annotation of the SAPPHIRE plot were those with the largest absolute value of correlation with the selected tICA components (we considered the maximum between sin and cos components); as for the feature-TIC correlation, each value is weighted by the corresponding eigenvalue of the tICA transformation matrix. The set of dihedrals for annotation is determined separately for KalbTGase and MTGase, but the two selections are merged into a single list of largely corresponding dihedrals (according to the structural alignment), which makes it easier to compare the SAPPHIRE plots of the two enzymes. Additionally, sidechain dihedrals of the catalytic triad are also shown.

To build the MSM, we used the recently developed SAPPHIRE-based clustering (SbC) method [35], which turns the visual notion of SAPPHIRE plot basins into a quantitative clustering algorithm. The number of bins *nx* and *ny*, along the *x* and *y* axis respectively, which determines the smallest basins that can be distinguished, was set to 1000, with *nx* = *ny*, resulting in 45 clusters for KalbTGase and 66 for MTGase. The corresponding discretized trajectory was used to infer the transition matrix of an MSM; the sliding-window method to count the transitions was used and detailed balance was imposed by naïve symmetrization of the count matrix, as implemented in the CAMPARI software package (keyword CADDLINKMODE set to 4). The MSM lag-time was set to 2 ns after monitoring the implied timescales (supplementary Fig. S7).

The lag-time for tICA construction and the number of bins *nx* for running the SbC were chosen by optimizing the variational approach for Markov processes (VAMP)-2 [36] score through grid search. This score gives an indication on how well the model approximates the slow modes of the true propagator. For different MSM lag-times (2 ns, 3 ns, and 5 ns) VAMP-2 scores were calculated by doing leave-one-out cross-validation on the five trajectories at room temperature. Results are shown in supplementary Fig. S8, grouped by either tICA lag-time or number of bins *nx*. The MSM construction is robust with respect to the choice of the latter hyperparameters. However, the chosen value of 0.5 ns (50 snapshots) for the tICA lag-time and 1000 for *nx* seem to provide the best VAMP scores across the considered MSM lag-times. The choice of lag-times and hyperparameters was determined on the trajectories of KalbTGase at 300 K, but then kept for the analyses at 335 K and 355 K, for both KalbTGase and MTGase.

SbC clusters were grouped into larger macrostates using the PCCA+ algorithm [37]. The number of macrostates (which is a hyperparameter and was set to 11 and 12 for KalbTGase and MTGase, respectively) was chosen by looking at the spectral gap in the eigenvalues of the MSM transition matrix. Macrostates were used to generate a network depiction summarizing the whole sampling at 300 K (see Fig. 4 and supplementary Fig. S10). The cartoons shown for each state of the network are the representative conformations of the SbC clusters contained in the same macrostate. The representative of the largest SbC cluster is in color, with the catalytic triad in sticks, and the representatives of all other clusters are in transparent grey. A SbC representative was chosen as the snapshot that is closest to the average conformation across the cluster, where the average was calculated on the chosen features describing the system. The edges of the network are proportional to the inverse of the mean first passage time (MFPT) between PCCA+ states. The microscopic MFPT between SbC clusters was calculated from the estimated transition matrix using the formulas for Markov chains [38] and it was then coarse-grained [39] to get the MFPT between macrostates.

tICA featurization, VAMP scores, and PCCA+ regrouping were calculated with PyEMMA v2 [40]; the PI and MSM transition matrix were evaluated using CAMPARI v4 (http://campari.sourceforge.net/) and the SAPPHIRE-based clustering and plots were done with the companion CampaRi R package [41–42]; network layouts were generated with the *igraph* R package [43].

3. Results and discussion

3.1. Fluctuations of the catalytic site

From the comparison of their RMSD plots (Supplementary Fig. S1, panels a-b), the five MD simulations of KalbTGase and MTGase at 300 K seem to have a very similar evolution, and no particular alterations are detected, suggesting that both systems are quite stable on the timescale of 350 ns with no major conformational rearrangements and no significant differences among the runs.

The RMSF analyses performed on KalbTGase highlight that the most flexible regions are the small peripheral loops joining the secondary structure elements, some of which are directly connected to the active site residues (Fig. 2, panel A on the right, and Fig. S2, panel A). The core of the protein, comprising the catalytic site, instead, seems to be very rigid.

In MTGase, the most flexible regions largely correspond to the homologous ones in KalbTGase and are predominantly the peripheral loops connected to the active site residues (Fig. 2, panel A on the left, and Fig. S2, panel B), but since these loops are longer in MTGase, their movements probably do not allow the closure of the active site. Moreover, from the comparison of the two RMSF profiles, it is possible to see that the detected fluctuations are higher in MTGase than in KalbTGase, with a maximum average RMSF of 0.20 nm versus 0.12 nm, respectively, at the level of a 2 ns time window. An exception is represented by the small loop in front of the catalytic Asp (residues 208–214 in KalbTGase and 299–305 in MTGase), which shows similar fluctuations in both enzymes. Overall, these results suggest that MTGase has an enhanced flexibility with respect to KalbTGase in the regions connected to the active site, and this could explain its wider specificity and lower selectivity.

3.2. Analysis of the volume of the catalytic site pocket

In order to investigate in more details the reasons that lead to enhanced specificity of KalbTGase with respect to MTGase, a deeper analysis on their active site pockets has been performed. The volume of the catalytic sites in the crystallographic structure of KalbTGase and MTGase resulted 133 and 213 $Å^3$, respectively, with a difference of 80 $Å^3$ representing more than 1/3 of the MTGase catalytic site volume. The analyses carried out on the MD simulations performed at 300 K demonstrate that, despite the fluctua-

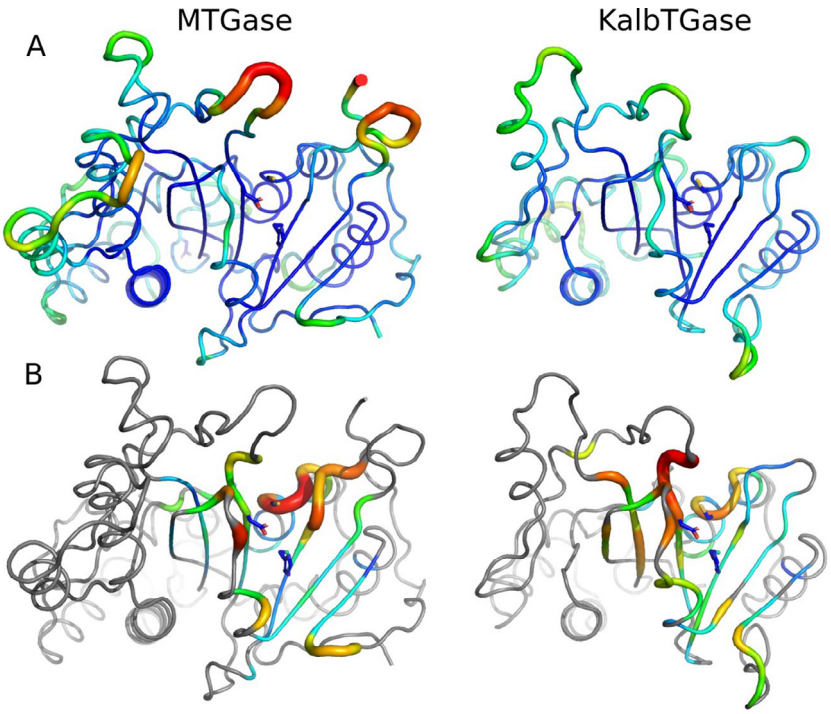

Fig. 2. RMSF and feature-TIC correlations mapped on the crystal structure at 300 K. A Average RMSF profiles of MTGase (left) and KalbTGase (right) at 300 K. Average RMSF values (supplementary Fig. S3), from small to large, are mapped to the color (blue-green–red color scale) and the tube width of the cartoon (from small to large). Sidechains of catalytic residues are in sticks. B Feature-TIC correlation values (see methods section for details) are mapped on the crystal structure of MTGase (left) and KalbTGase (right). Small to large values are encoded by the color (blue-green–red) and the tube width (small to large). Sidechains of the catalytic residues are in sticks. Residues for which no correlation value was calculated are in grey. For both panels, numerical values were encoded in the PDB as B-factors and visualized with Pymol. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

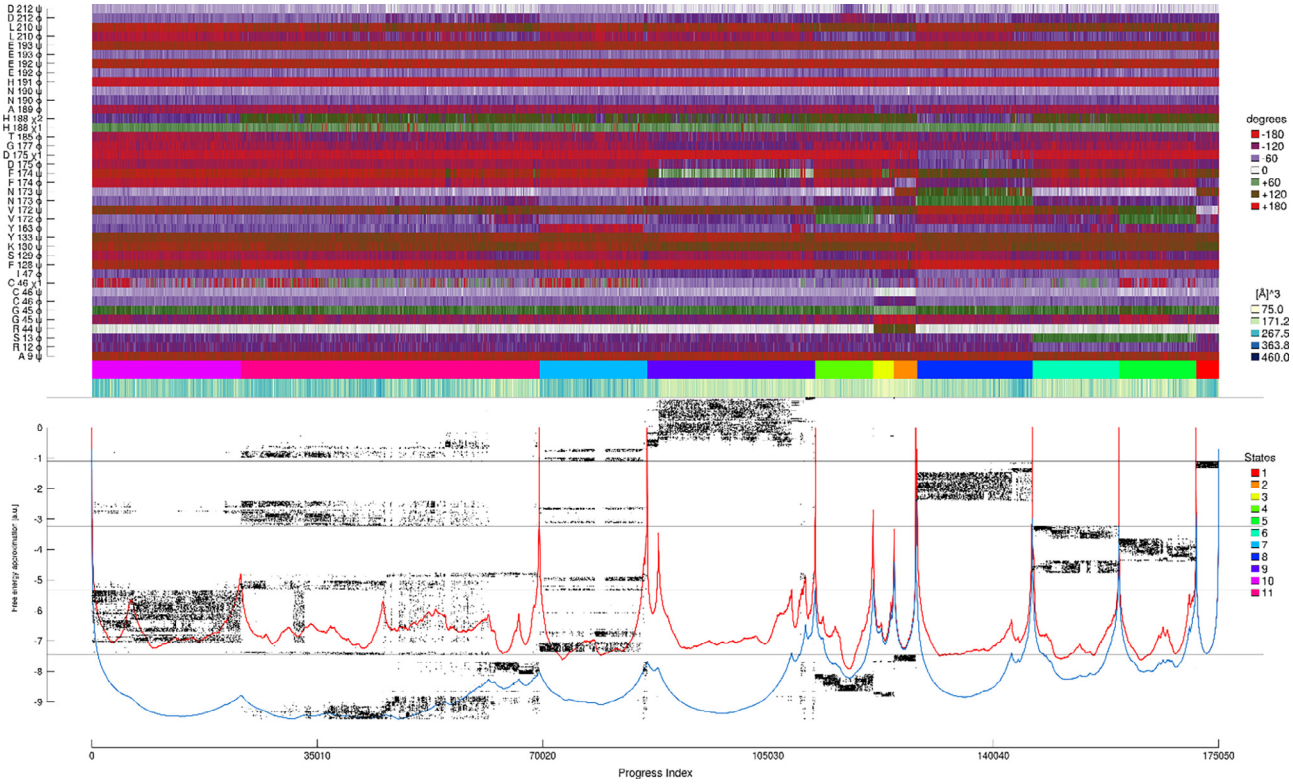

Fig. 3. SAPPHIRE plot for KalbTGase at 300 K. Snapshots are reordered according to the progress index (PI). The global and localized cut (blue and red lines) represent a pseudo-free energy profile and separate the snapshots in different basins. The dot pattern reports (vertical axis) for each PI value, the occurrence time along the real trajectory. The five runs at 300 K were concatenated and are separated by horizontal black lines. The upper part of the plot reports the following annotations as heatmaps, from bottom to top: volume of the catalytic pocket; PCCA+ assignment to 11 metastable states; selected set of relevant dihedral angles (see *materials and methods* section for the choice criteria). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

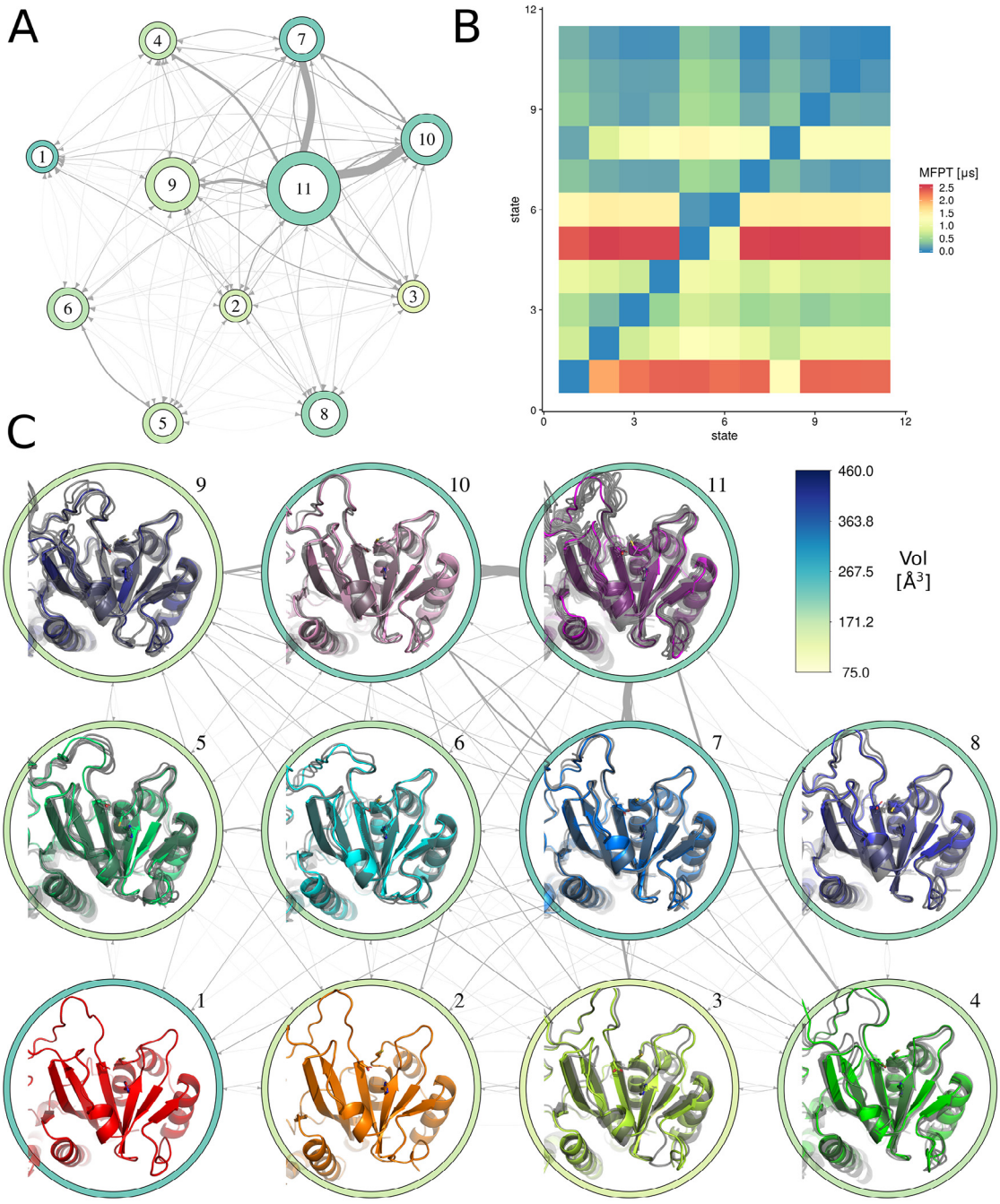

Fig. 4. Conformational network for KalbTGase at 300 K. A Each node corresponds to a PCCA+ state and the node size is proportional to the number of snapshots. The edge width reflects the inverse of the MFPT between states. The ring color is the average volume of the catalytic site in the state. The color scale is reported in panel C and it is the same as in the SAPPHIRE plot of Fig. 4. 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.

tions, the active site of KalbTGase is roughly 50 Å^3 smaller than that of MTGase (Table 1 and supplementary Fig. S3). Indeed, considering the total sampling at 300 K, it is possible to estimate that KalbTGase active site has an average volume equal to 196.96 Å^3 (median value = 186.00 Å^3), whereas MTGase catalytic pocket has an average volume equal to 245.75 Å^3 (median value = 240.8 2 Å^3). Moreover, the range of values reached by the volume of the active site of KalbTGase (25% percentile at 132.63 Å^3 and 75% percentile at 266.62 Å^3) is also smaller than the one of MTGase (25% percentile at 165.01 Å^3 and 75% percentile at 318.89 Å^3).

In order to assess whether the difference in the volumes of the two active sites is significant, we ran several statistical tests.

Specifically, we used Mann Whitney *U* test, Welch's *t*-test and Student's *t*-test, with the alternative hypothesis of KalbTGasehaving a smaller binding site volume than MTGase. The volume distributions are not normal and do not always have the same shape (supplementary Fig. S3, panel A), motivating the use of Mann Whitney *U* test, which compares the mean ranks (and not the median) of the distributions. Moreover, even though the volumes are themselves not normally distributed, for sufficiently large samples

Table 1
Analysis of volumes of the active site pocket in MD simulations.

Protein	300 K		335 K		355 K	
	mean*	sd*	mean	sd	mean	sd
KalbTGase	197	88	153	83	161	86
MTGase	246	111	200	96	180	91

	KalbTGase		MTGase	
	mean	sd	mean	sd
RUN1	181	88	245	88
RUN2	218	84	222	117
RUN3	181	88	314	112
RUN4	214	85	201	108
RUN5	189	89	242	93
MEAN**	197	18	246	43

* The reported values are calculated for each enzyme on the whole sampling at 300 K (5 runs).
** The reported mean and standard deviation values at 300 K are calculated as the mean and standard deviation of the means of the five runs.

we can still run a Welch's *t*-test and a Student's *t*-test to compare the means if the central limit theorem holds [44]. Welch's *t*-test assumes unpaired samples with unequal variance; however, as the variances of the single runs are close to each other (Table 1), we also ran a *t*-test, which assumes unpaired samples with equal variance.

The biggest obstacle to a rigorous analysis comes from the fact that we are dealing with MD snapshots, thus volume values are time-correlated, and this is incompatible with most statistical methods, which assume independence of the samples. For this reason, instead of using the raw volume data, we first used as independent samples the means of the five trajectories; these values can be considered independent, possibly except for the fact that the trajectory starting points are the same. All three types of tests confirmed that the volume of the catalytic site of KalbTGase is smaller than MTGase at a 0.05 significance level (p-values between 1.6 and 3.2%).

To further support our claim, we also repeated the tests using samples of block-averaged values calculated on each trajectory, this time comparing every possible combination of runs of MTGase and KalbTGase. With a block size of 200 snapshots (non-overlapping blocks of 2 ns), using the Mann-Whitney *U* test, the null hypothesis could be always rejected except for the comparisons involving combinations of runs 2 and 4 of both MTGase and KalbTGase and additionally in runs number 3 and 5 of KalbTGase versus run number 4 of MTGase. Using larger blocks (3500 snapshots, which equals to 35 ns and roughly 10 samples per trajectory), more tests could not be rejected at the chosen significance level (0.05), but all of the latter involved comparisons to trajectory 2 and 4 of MTGase, which have a lower mean with respect to the other MTGase trajectories (Table 1). The same tests were also run on the raw volume data without block averaging, yielding similar results, although hampered by sample dependence.

Rather than comparing the means, the two-sample Kolmogorov-Smirnov test compares the distributions and requires in first instance the independence of the distribution estimates, which is the case for independent trajectories of different enzymes. Using this test on the whole volume distributions (*i.e.* combining the values of the five runs for each enzyme) also agreed to the hypothesis that the volume of KalbTGase is smaller than MTGase (p-value < 1e−15).

There are two important caveats to this analysis: first, snapshots, for which MDpocket failed to identify the pocket (returning a null volume), were discarded; indeed, by looking at the volume distributions in panel B of supplementary Fig. S3, the null values look like an artifact. However, treating these values as true zeros does not change the results significantly and the outcomes of the statistical tests are comparable. In particular, both the Mann-Whitney *U* test on the means and the Kolmogorov-Smirnov test on the whole sampling accept the alternative hypothesis that the volume of the catalytic site of KalbTGase is smaller than the one of MTGase.

Second, both the Welch's *t*-test and the Student's *t*-test assume normality of the distribution of the sample mean; this is often assessed by verifying normality of the sample population, which is hardly the case here (supplementary Fig. S3), although these tests are relatively robust with respect to deviations from normality [45]. The two tests give indeed results comparable to the Mann-Whitney *U* test. Moreover, for large enough samples, or for testing directly using the independent trajectory mean values, the normality of the distribution of the sample mean derives from the central limit theorem [44].

It is important to point out that the similar variance of the volume distributions for MTGase and KalbTGase might indeed be related to the comparable size of the volume oscillations returned by MDpocket. The changes in volume of the active site throughout an MD run are governed by modifications of the protein surface, which are ultimately determined by fluctuations of sidechains and the slow conformational rearrangements of the binding site. In the conformational analysis paragraph, we try to qualitatively correlate the changes in volume with conformational changes. In order to extract more signal from the volume time trace, the values shown in the SAPPHIRE plot annotation are smoothed using a moving average filter with a window of 0.5 ns, excluding zero values.

Our analyses suggest that KalbTGase has a smaller active site, suitable to explain its higher selectivity in the choice of substrates; on the other hand, the larger size of the active site of MTGase may guarantee more adaptability to different substrates. It is likely that the enhanced fluctuations of the MTGase loops closest to the active site region affect the distance of the catalytic residues during all the MD simulation.

3.3. Feature-TIC correlations

Correlations between the dihedral angles describing the catalytic pocket and the 13 tICA components considered are useful to identify the residues involved in the slowest modes of the dynamics. The residue-level absolute values of the feature-TIC correlations are mapped on the crystal structures of the two enzymes in Fig. 2, panel B. It should be stressed that, differently from panel A, where the values of RMSF have the same scale, here the values of the correlations refer to different tICA transformations and dynamic modes, hence they cannot be compared between the two enzymes. However, they still give valuable information about the regions of the active site undergoing backbone conformational changes. Values are reported only for the residues used to calculate the tICA transformation (all the others are in grey).

For MTGase the regions with the major changes are Tyr 301 and the terminal α-helical loop containing the catalytic residue Cys 63. Tyr 301 is placed in front of the catalytic site and its movement might affect the volume and accessibility of the latter. Moreover, changes directly affecting the catalytic residue Cys 63 might also affect the enzyme function.

For KalbTGase, the largest conformational variability is located on the loop connected to catalytic Asp 175 (residues 172–174); this is also the loop showing some of the largest fluctuations in the RMSF profile. Indeed, by comparing the feature-TIC correlations (Fig. 2B) with the RMSF profiles (Fig. 2A), we can see that the regions of the catalytic pocket undergoing backbone changes do not show large flexibility and fluctuations and the two quantities measure different types of motion. Note that loops with large RMSF values might also show large backbone rearrangements but these are not considered in the featurization as they are further away from the catalytic cleft.

3.4. Conformational analysis

The conformational analysis is used to gain a structural insight and corroborate the previous observations. The SAPPHIRE plot in Fig. 3 gives an overview of the cumulative sampling for KalbTGase at 300 K. The plot consists of a reordering of the trajectory snapshots based on geometric similarity (and assuming the latter corresponds to kinetic vicinity). The cut functions (blue and red lines in the lower part of the plot) represent pseudo-free energy profiles that help identify the barriers between different basins, and the dot pattern reports (on the *y* axis) the actual time of occurrence of the reordered snapshots; the 5 runs are separated by horizontal black lines. The upper panel of the plot features several heatmap annotations, which are snapshot-based and help identify the differences among the basins. Specifically, the following annotations are reported starting from the bottom: the volume of the catalytic pocket (after applying a moving average filter, see 3.2 section); the PCCA+ metastable state the snapshot is assigned to; the value of the most relevant backbone dihedral angles; additionally, sidechain dihedrals of the catalytic residues are also included.

A coarse-grained network at the level of the PCCA+ metastable states (Fig. 4, for KalbTGase) is useful to visualize the conformations identified by the SAPPHIRE plot. It should be noted that the PCCA+ algorithm returns the most metastable grouping of the SbC clusters into the specified number of states, viz. 11 for KalbTGase and 12 for MTGase. This does not mean that states are necessarily homogeneous, but clusters in the same state have a shorter kinetic distance than clusters assigned to different states.

The size of the network vertices in panel A is proportional to the number of snapshots and the color reflects the average volume of the catalytic site in that state. The width of the edges accounts qualitatively for the kinetic distance as it is proportional to the inverse of the mean first passage time (MFPT) between two states. All the coarse-grained MFPT values are reported as a heatmap in panel B. Panel C rearranges the network on a grid layout adding the cartoons of the centroids of the SbC clusters composing the state. The centroid of the largest SbC cluster is in color, the others are in grey to suggest the heterogeneity of PCCA+ states. State numbering and coloring is the same in the SAPPHIRE and the network plot.

The analysis identifies one large heterogeneous basin (state 11, in magenta), encompassing approximately one fourth of the sampling (from PI ∼25,000 to ∼70,000). This basin shows recurrence in all the 5 runs and includes many of the starting snapshots of each simulation. Although it is very heterogeneous and many sub-basins can be distinguished, all snapshots are assigned to the same state. State 11 is kinetically close to other large basins, in particular to 7 and 10, from and to which frequent direct transitions are sampled. There are only minor differences among them, which are localized on neighboring regions of the binding pocket (residues 210–212 and 128–133). Moreover, state 10 (pink), that is separated from 11 by a low barrier, is characterized by a 180° flip of the CHI_2 angle of His 188. The same flip is observed in state 8 (PI ∼140,000, dark blue), where notably also Asp 175 has a unique reorientation, no longer pointing towards His 188. In this way, the possibility of a hydrogen bond between the carboxyl group of Asp 175 and the imidazole group of His 188 is lost.

The rest of the SAPPHIRE plot shows smaller basins that are often sampled in single runs. These states are in general more compact and homogeneous, as it is demonstrated by the low number of SbC clusters and by the dihedral and volume annotations. The latter shows indeed some partitioning with the basins, although it is not a geometric variable directly employed for the generation of the PI. This means that the overall volume is largely determined by the backbone conformation of the catalytic pocket. State 1 and 5 are kinetically the most distant from any other states, with estimated MFPTs to reach them between 2 and 2.5 μs. This is because they are characterized by slow conformational changes which are sampled only once throughout the trajectory, as it is visible from the dot pattern in Fig. 3.

The analogous SAPPHIRE and network plots for MTGase at 300 K are presented in supplementary Fig. S9 and S10 for comparison. The SAPPHIRE plot for MTGase appears more structured and includes many unique sub-basins with little to no recurrence. SbC identifies more clusters; however, the number of larger basins sampled is comparable between the two enzymes. The dihedral annotations on the SAPPHIRE plot of KalbTGase (Fig. 3) and MTGase (Fig. S9) mainly report corresponding angles for comparison. Overall, dihedral angles show the same average pattern, which means the two enzymes share the same fold; an exception to this is represented by the first three dihedrals which are part of the N-terminal loop region of the structures and differ substantially.

As it was already observed from the feature-TIC correlations of Fig. 2, the regions with most of the conformational changes differ between the two enzymes. For KalbTGase, most of the changes involve the catalytic Asp 175 and the loop next to it. Residues 210–212, which also exhibit some changes, are spatially adjacent to the former loop. The rest of the structure does not undergo any large variations, except for states 2 and 3, where the short turn including Cys 46 rearranges (Fig. 4, panel C). The latter conformational change is also visible in states 7 and 3 of MTGase (supplementary Fig. S10).

MTGase shows even more conformational variability of the latter region containing catalytic Cys 63 (residues 61–64), whereas the loop region next to the catalytic Asp 254 (residues 251–254) keeps the same conformation throughout most of the sampling. These very few residues also have lower RMSF values than the corresponding ones in KalbTGase (despite the rest of the loop reaching higher fluctuations). The other region showing most of the conformational plasticity for MTGase is the loop downstream of His 273; its movement might be coupled to the conformational changes of Cys 63, to which it is spatially close. Tyr 301 in MTGase and the corresponding Leu 210 in KalbTGase are also involved in transitions; interestingly, the sidechains of these two residues are placed on top of the catalytic site.

Counter-ions play a crucial role in MD simulations and they are commonly added to the simulation box to compensate the system charges and approximate an average ionic strength resembling the cellular environment. The size of the ions (van der Waals radius) affects their behavior in solution, balancing inter-ion electrostatic interactions and partial dehydration. In our study, we compared the results obtained for MTGase and KalbTGase under similar

conditions of simulations. Therefore, any possible effect on the results due to the choice of the counter-ions can be considered a methodological and systematic effect, not influencing the differential observation we made.

As concerns the sidechains of the catalytic residues, the CHI_1 of cysteine (Cys 46 in KalbTGase and Cys 63 in MTGase) is allowed to spin quite freely. The sidechains of the aspartate and histidines (Asp 175, His 188 for KalbTGase; Asp 254, His 273 for MTGase) are locked in the same conformation throughout the MTGase simulations; on the contrary, they see some structural changes in KalbTGase.

The volume of the catalytic site of MTGase also partitions quite well with the basins and it shows a much larger range of values than for KalbTGase.

3.5. Extension of the results to higher temperatures

In order to perform a preliminary analysis to the wide range of temperatures used in industrial applications, we ran two 300 ns-simulations at 335 K and 355 K and compared them to the ones at 300 K.

By looking at the RMSD time trace (Fig. 5), the run on MTGase at 335 K (red line) stays close to the corresponding curve at 300 K, meaning that it possibly samples similar regions of the conformational space. On the contrary, the RMSD of the run on KalbTGase at 335 K (yellow line) diverges more, sampling conformations that are geometrically further away from the crystal structure. The PCA analysis (supplementary Fig. S5, panels C-D) also shows more overlapping states for MTGase as a function of time.

At 355 K the RMSD curves for both enzymes (blue and pink lines) show a marked increase, with onset at the beginning of the run, and they reach values of RMSD that are twice as large as the ones at 300 K. The PCA plot confirms the temporal evolution through separate states in the reduced space of the first two components.

For both enzymes, the RMSF analyses at 335 K show an increment of the fluctuations of the same regions already highlighted in the simulations at 300 K (Supplementary Fig. S6, panel A-B). In MTGase, the RMSF seems to be more affected than the RMSD by the rising of the temperature. From these results, it is possible to predict that the rising of temperature mainly affects the longer loops, whereas the catalytic core remains stable. In fact, it has been recently observed that the increase in temperature may affect the lifetime of H-bond interactions, while the total number of H-bond interactions is less influenced [46]. This means that longer loops, lacking characteristic interactions of the secondary structures and being intrinsically flexible [47], are more suitable to increasing their flexibility. The RMSF profiles at 355 K (supplementary Fig. S6, panel C-D) show that, as expected, the fluctuation peaks are even higher than the peaks related to the MD simulation at 335 K and are still associated with the most flexible regions of the molecule already identified. Moreover, these regions reach higher RMSF values in KalbTGase than MTGase, at both temperatures.

We also performed the active site pocket volume calculations on the runs at 335 K and 355 K, with additional statistical tests to compare the means, mean ranks and distributions of the volume values between the two proteins and with respect to the runs at 300 K (for the latter we considered the cumulative sampling of the 5 runs). The mean volume and standard deviations, excluding the null volume structures, are reported in Table 1. The standard deviation is comparable between the runs and does not seem to be influenced by the temperature rise, thus it should largely be attributed to the noise in the determination of the pocket. For KalbTGase, the increase of temperature at 335 K induces a marked decrease of the active site pocket volume. A further temperature increase to 355 K only results in a minor increase of the mean volume with respect to the value at 335 K. On the other hand, for MTGase the volume decreases constantly as temperature rises.

In order to have a collective view of the sampling of each enzyme we generated a SAPPHIRE plot including the runs at all temperatures (supplementary Fig. S11 for KalbTGase and S12 for MTGase). From the dot pattern, it can be seen that the run of KalbTGase at 335 K has some recurrence with the runs at 300 K at the beginning and in the middle of the simulation. Additionally, it discovers two relatively large unique states (around PI 140,000).

The run on MTGase at 335 K mainly explores new territory, but it also discovers only a few large states (see PI values around 180,000).

On the contrary, for both proteins the run at 355 K discovers a multitude of very small states that are added at the end of the PI. This confirms that they are further away from the initial structures and such a high temperature increases the speed of conformational transitions.

The volume annotation still shows partitioning with the basins and at higher temperature, on average it samples conformations with a smaller volume of the catalytic pocket. The SAPPHIRE plot confirms that there is almost no recurrence at high temperature; hence the observations about the trends of volume variation with temperature might be hindered by the limited sampling at 335 K and 355 K.

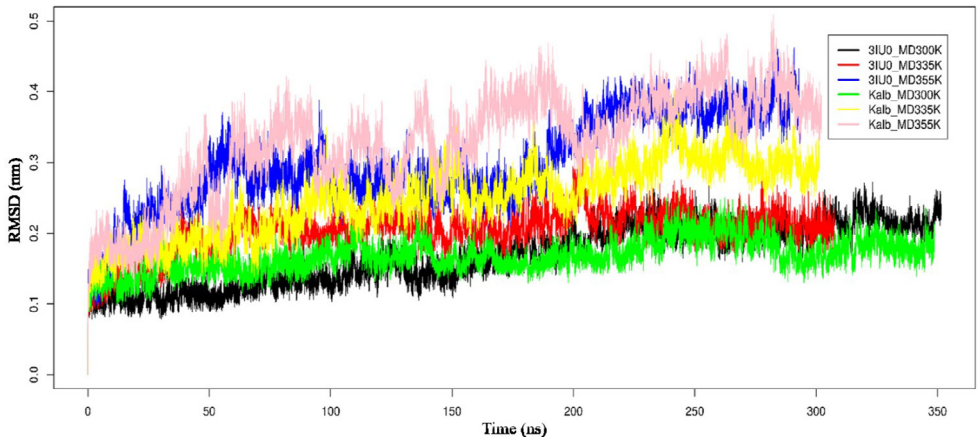

Fig. 5. RMSD comparison between MD simulations of MTGase and KalbTGase. The RMSD is calculated on the Cα for the MD simulation of MTGase at 300 K (black line), at 335 K (red line) and at 355 K (blue line), and compared with the MD simulation of KalbTGase at 300 K (green line), at 335 K (yellow line) and at 355 K (pink line). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. Conclusion

MD simulations performed on KalbTGase and MTGase show that at 300 K both proteins preserve the conformation of the catalytic site and its closest areas, and their flexibility is mainly focused in the peripheral loops. In MTGase, maximum fluctuation values are higher than in KalbTGase, due to the presence of longer loops. Some fluctuations have been detected also in the linking regions between the catalytic His and Asp, and in particular they reach higher values in MTGase than in KalbTGase. Thus, from these results, it appears that KalbTGase at room temperature is a more rigid protein than MTGase.

The conformational analysis shows that slow conformational changes also involve regions of the binding pocket which are not flexible. This can be due primarily to two reasons: first, the separation of the timescales that are considered by the two types of analyses. Second, the overall fold of the catalytic region is maintained, hence neighboring dihedrals often undergo changes that compensate each other, not resulting in large fluctuations. Conformational changes also involve regions adjacent to the catalytic residues and can perturb them, as it happens for the short turn containing the Cys and the loop next to the Asp. In general, in both enzymes, the catalytic Cys is relatively free to rotate, whereas His and Asp assume metastable orientations. In particular, for MTGase no conformational changes involving the sidechains of the catalytic His and Asp are sampled.

The average volume of the catalytic pocket is smaller for KalbTGase than for MTGase, and for both proteins, it is largely determined by the overall backbone fold. However, a sequence component due to the sidechains is also plausible, and it would also partially explain the volume oscillations.

From MD simulations performed at higher temperatures, both proteins show enhanced flexibility. As a possible consequence of the increased disorder in KalbTGase, a modification of the catalytic site could result in enhanced catalysis of different substrates with respect to room temperature. The volume of the catalytic pocket shows a general trend to shrink for both proteins and, although the volume of KalbTGase is consistently smaller than the volume of MTGase, the difference becomes narrower.

Our analyses suggests that increasing temperatures might tune the activities of these two enzymes in a different way, making KalbTGase less specific, whereas MTGase could become less active and/or more specific. Further studies will be necessary to prove it, but this study paves the way for a scenario in which both proteins could be used for broader applications.

Author contributions

D.G. performed research, analyzed data, wrote the manuscript; C.L. analyzed data, contributed analytic tools, wrote the manuscript; A.C. designed research, contributed analytic tools, reviewed and edited the manuscript; A.M. designed research, analyzed data, reviewed and edited the manuscript; A.F. designed research, analyzed data, reviewed and edited the manuscript, supervised the work.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

D.G. and A.F. work is in the framework of ELIXIR, the European Research Infrastructure for life sciences data. D.G. is supported by a post-doc fellowship under the project framework "CIR01_00017-' 'CNRBiOmics–Centro Nazionale di Ricerca in Bioinformatica per le Scienze "Omiche"–Rafforzamento del capitale umano" funded by MUR, CUP B56J20000960001. A.M. is supported by the Italian Ministry of University and Research (FFABR 2017 program, and PRIN 2017 program, grant number: 2017483NH8), by University of Salerno (FARB 2019-2020-2021 programs), and by funds from BANCA D'ITALIA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2022.07.024.

References

[1] Beninati S, Piacentini M. The transglutaminase family: an overview: minireview article. Amino Acids 2004;26:367–72.
[2] Facchiano F, Facchiano A, Facchiano AM. The role of transglutaminase-2 and its substrates in human diseases. Front Biosci 2006;11:1758–73.
[3] Camolezi Gaspar AL, Pedroso de Góes-Favoni S. Action of microbial transglutaminase (MTGase) in the modification of food proteins: A review. Food Chem 2015;171:315–22.
[4] Strop P. Versatility of microbial transglutaminase. Bioconjug Chem 2014;25:855–62.
[5] Santhi D, Kalaikannan A, Malairaj P, Arun Prabhu S. Application of microbial transglutaminase in meat foods: A review. Crit Rev Food Sci Nutr 2017:572071–576.
[6] Giordano D, Facchiano A. Classification of Microbial Transglutaminases by evaluation of evolution trees, sequence motifs, secondary structure topology and conservation of potential catalytic residues. Biochem Biophyis Res Commun 2018;509:506–13.
[7] Steffen W, Ko FC, Patel J, Lyamichev V, Albert TJ, Benz J, et al. Discovery of a microbial transglutaminase enabling highly site-specific labeling of proteins. J Biol Chem 2017;292:15622–35.
[8] Abraham MJ, Murtolad T, Schulzb R, Pálla S, Smith JC, Hessa B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 2015;1–2:19–25.
[9] Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. J Comput Chem 2008;29:1859–65.
[10] Lee J, Cheng X, Swails JM, Yeom MS, Eastman PK, Lemkul JA, et al. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations using the CHARMM36 Additive Force Field. J Chem Theory Comput 2016;12:405–13.
[11] Best RB, Zhu X, Shim J, Lopes PE, Mittal J, Feig M, et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ, ψ and side-chain $\chi 1$ and $\chi 2$ dihedral angles. J Chem Theory Comput 2012;8:3257–73.
[12] Durell SR, Brooks BR, Ben-Naim A. Solvent-induced forces between two hydrophilic groups. J Phys Chem 1994;98:2198–202.
[13] Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684–90.
[14] Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. J Appl Phys 1981;52:7182–90.
[15] Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. J Chem Phys 2007;126:014101.
[16] Hockney RW. The potential calculation and some applications. Methods Comput Phys 1970;9:135–211.
[17] Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. J Comp Chem 1997;18:1463–72.
[18] Darden T, York D, Pedersen L. Particle mesh Ewald: An N•log(N) method for Ewald sums in large systems. J Chem Phys 1993;98:10089–92.
[19] PyMOL Molecular Graphics System. https://sourceforge.net/projects/pymol/
[20] Turner, P.J. (2005) XMGRACE, Version 5.1.19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR.
[21] Yu J, Zhou Y, Tanaka I, Yao M. Roll: A new algorithm for the detection of protein pockets and cavities with a rolling probe sphere. Bioinformatics 2010;26:46–52.
[22] Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: An open source platform for ligand pocket detection. BMC Bioinf 2009;10:168.
[23] Schmidtke P, Le Guilloux V, Maupetit J, Tuffery P. Fpocket: online tools for protein ensemble pocket detection and tracking. Nucleic Acids Res 2010;38:W582–9.
[24] Schmidtke P, Bidon-Chanal A, Luque FJ, Barril X. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. Bioinformatics 2011;27:3276–85.
[25] Blöchliger N, Vitalis A, Caflisch A. High-resolution visualisation of the states and pathways sampled in molecular dynamics simulations. Sci Rep 2014;4:6264.

[26] Prinz JH, Wu H, Sarich M, Keller B, Senne M, Held M, et al. Markov models of molecular kinetics: Generation and validation. J Chem Phys 2011;134:174105.
[27] Husic BE, McGibbon RT, Sultan MM, Pande VS. Optimized parameter selection reveals trends in Markov state models for protein folding. J Chem Phys 2016;145:194103.
[28] Molgedey L, Schuster HG. Separation of a mixture of independent signals using time delayed correlations. Phys Rev Lett 1994;72:3634.
[29] Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, Noé F. Identification of slow molecular order parameters for Markov model construction. J Chem Phys 2013;139:07B604_1.
[30] Schwantes CR, Pande VS. Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. J Chem Theory Comput 2013;9:2000–9.
[31] Noé F, Clementi C. Kinetic distance and kinetic maps from molecular dynamics simulation. J Chem Theory Comput 2015;11:5002–11.
[32] Blöchliger N, Vitalis A, Caflisch A. A scalable algorithm to order and annotate continuous observations reveals the metastable states visited by dynamical systems. Comput Phys Commun 2013;184:2446–53.
[33] Vitalis A, Caflisch A. Efficient construction of mesostate networks from molecular dynamics trajectories. J Chem Theory Comput 2012;8:1108–20.
[34] Vitalis A. (2020) An Improved and Parallel Version of a Scalable Algorithm for Analyzing Time Series Data. arXiv preprint arXiv:2006.04940.
[35] Cocina F, Vitalis A, Caflisch A. Sapphire-based clustering. J Chem Theory Comput 2020;16:6383–96.
[36] Wu H, Noé F. Variational approach for learning Markov processes from time series data. J Nonlinear Sci 2020;30:23–66.
[37] Röblitz S, Weber M. Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. Adv Data Anal Classif 2013;7:147–79.
[38] Sheskin TJ. Computing mean first passage times for a Markov chain. Int J Math Educ Sci Technol 1995;26:729–35.
[39] Plattner N, Noé F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. Nat Commun 2015;6:1–10.
[40] Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, et al. PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. J Chem Theory Comput 2015;11:5525–42.
[41] Garolini D, Cocina F, Langini, C. (2019) CampaRi: an R package for time series analysis. doi:10.5281/zenodo.3428933.
[42] Garolini D, Vitalis A, Caflisch A. Unsupervised identification of states from voltage recordings of neural networks. J Neurosci Methods 2019;318:104–17.
[43] Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, complex systems 2006;1695:1–9.
[44] Lumley T, Diehr P, Emerson S, Chen L. The importance of the normality assumption in large public health data sets. Annu Rev Public Health 2002;23:151–69.
[45] Sawilowsky SS, Blair RC. A more realistic look at the robustness and type II error properties of the t test to departures from population normality. Psychol Bull 1992;111:352.
[46] Alves ED, de Andrade DX, de Almeida AR, Colherinhas G. Atomistic molecular dynamics study on the influence of high temperatures on the structure of peptide nanomembranes candidates for organic supercapacitor electrode. J Mol Liq 2021;334:116126.
[47] Ragone R, Facchiano F, Facchiano A, Facchiano AM, Colonna G. Flexibility plot of proteins. Protein Eng 1989;2:497–504.