



# Molecular dynamics and solvation structures of the $\beta$ -glucosidase from *Humicola insolens* (BGHI) in aqueous solutions containing glucose

Felipe Cardoso Ramos, Leandro Martínez \*

Institute of Chemistry and Center for Computing in Engineering and Science - CCES, Universidade Estadual de Campinas (UNICAMP), Brazil

## ABSTRACT

The  $\beta$ -glucosidase enzyme is a glycosyl hydrolase that breaks down the  $\beta$ -1,4 linkage of cellobiose. It is inhibited by glucose at high concentrations due to competitive inhibition. However, at lower glucose concentrations, the glucose-tolerant  $\beta$ -glucosidase from *Humicola insolens* (BGHI) undergoes stimulation. Proteins, in aqueous sugar solutions, tend to be preferentially hydrated, which generally promotes their stabilization. Thus, solvation phenomena may contribute to both glucose tolerance and stimulation processes. We have performed atomistic classical Molecular Dynamics (MD) simulations of BGHI at different glucose concentrations to mimic the conditions found in the catalytic experiments. A detailed examination of the solvent environment through the calculation of minimum distance distribution functions (MDDFs) and Kirkwood-Buff (KB) integrals was performed. The enzyme is preferentially hydrated in the presence of glucose at all concentrations. Nevertheless, the hydration does not prevent the glucose from directly interacting with the BGHI surface or from entering the active site. Based on the obtained results, we hypothesize that preferential hydration is beneficial for enzyme activity. At the same time, product inhibition has little effect at lower concentrations of glucose, and at higher glucose concentrations, competition for the active site becomes predominant and the enzyme is primarily inhibited.

## 1. Introduction

In industrial bioprocesses, it is vital to address the issue of enzyme activation or inactivation [1]. This is especially relevant when considering the employment of biocatalysis in non-conventional media [2], meaning conditions that are not similar to those in which the enzyme is found naturally. Organic cosolvents, for example, can confer many modifications to enzymatic processes [3–6], especially if it is also a product of the enzymatic reaction [7]. This is the case of the enzyme  $\beta$ -glucosidase (BG) enzyme, a glycosyl hydrolase (GH) that breaks down the  $\beta$ -1,4 linkage of cellobiose [8]. In BGs, the catalytic reaction is carried out by two conserved glutamate residues, which perform the hydrolysis or transglycosylation reaction by a retention mechanism [9].

As a key enzyme in the biomass saccharification process, BG is widely present in numerous hydrolysis cocktails derived from microorganisms [10–12]. The BG is a crucial enzyme because it performs the final step of the lignocellulosic biomass deconstruction [13]. However, as occurs with many industrial GHs, the BG is inhibited by coproducts or byproducts of biomass hydrolysis. The main inhibitory compound of BG is normally its reaction product, glucose [12,14,15].

Some BGs, however, only experience product inhibition at high glucose concentrations and so are classified as glucose-tolerant. The enzyme BG from *Humicola insolens* (BGHI) is a glucose-tolerant enzyme from family GH1 that, in fact, is stimulated by glucose and other

monosaccharides, such as xylose, when they are present in low quantities in the reaction medium [7,9,16,17]. The maximum stimulation of BGHI occurs at 50–125 mM glucose and, above this concentration, the BGHI activity is inhibited as glucose starts to compete with cellobiose for the active site [9,16].

Experimental and computational investigations have postulated a variety of molecular mechanisms to explain glucose tolerance and stimulation in GH1 GBs [8,9,12,14,18–22]. The two phenomena appear to be interrelated and can be dependent on glucose accessibility [8,12], allosteric interactions [9,18], glucose competition to nonproductive substrate binding sites [14,19], or molecular events that lead the glucose to be removed from the substrate cavity [20] – here we prefer to use “cavity” instead of “channel” for reasons that will become clear later.

In the present work, we would like to draw attention to phenomena of another nature: solvation-induced effects. It is common knowledge that water plays a fundamental role in the folding and dynamics of proteins so that altering the composition of the aqueous environment can directly or indirectly affect the protein stability in solution [23,24]. Previous studies have shown that osmolytes, sugars in particular, can alter the hydration layer around proteins, potentially affecting key biochemical interactions and enzymatic activities [25–27]. Small carbohydrates such as glucose are excluded from the protein surface in aqueous solutions. The protein is preferentially hydrated, and the net effect is that these osmolytes act as structural protectors [28–30].

\* Corresponding author.

E-mail address: [lmartine@unicamp.br](mailto:lmartine@unicamp.br) (L. Martínez).

Preferential hydration, in turn, can significantly impact enzyme function, allowing the enzyme to operate in the presence of organic solvents or modulating biochemical reactions [31–33].

Even with all of this prior knowledge, there are still some unanswered questions, particularly regarding how sugar-induced modifications to hydration layers affect the dynamics of protein structure in aqueous solution, which is challenging to investigate both theoretically and experimentally [34–36]. Addressing this gap, the present study aims to investigate the solvation structure and hydration properties of the  $\beta$ -glucosidase from *Humicola insolens* (BGHI) in glucose-containing solutions, contributing new insights into enzyme behavior in saccharide-rich environments.

Simulation techniques such as Molecular Dynamics (MD) allow to recreate the thermodynamic conditions in which the reaction occurs and provide atomistic details that are often challenging to obtain experimentally [18,20,37–39]. Therefore, MD studies have been used to investigate solvent-induced effects on the structural dynamics of enzymes [36,40–43]. Here, we have performed atomistic classical MD simulations of the BGHI enzyme at different glucose concentrations (0 mM, 125 mM, 250 mM, and 1000 mM) in order to mimic the conditions found in the experiments. We have conducted a detailed examination of the solvent environment based on the calculations of minimum-distance distribution functions (MDDFs) and the Kirkwood-Buff (KB) theory of solvation. The former allowed us to determine how the solvent is structured around the enzyme, the main types of interactions that occur between the BGHI and the cosolvent, as well as the main glucose-interacting residues in the substrate cavity. The latter was crucial to precisely estimate the protein preferential solvation (or interaction) parameters [44–46]. In addition, we have carried out several analyses of the BGHI in water and in the presence of glucose to evaluate the protein flexibility at different cosolvent concentrations, as well as other properties dependent on the interactions between glucose and the active site and glucose accessibility to the substrate cavity.

The obtained results support experimental findings on the structure and function of BGHI under the evaluated conditions, aligning with the hypothesis that solvation effects may also contribute to glucose stimulation and tolerance.

## 2. Methods

### 2.1. Molecular dynamics simulations

The initial coordinates for the studied protein was obtained from the Protein Data Bank (PDB: 4MDO [8]). The protein was modeled with the CHAMM36m force field [47], water with the TIP3P model [48], and the  $\beta$ -glucose (BGL) cosolvent with the modified force field CHARMM36 created by Cloutier et al. [49]. This later force field was developed to reproduce experimental KB integral values obtained for ternary protein/sugar/water solutions. We performed all steps involving energy minimization and MD simulations by using the software GROMACS (v2019.4) [50]. The system minimization was done by performing up to 50,000 steepest descent steps, or until the total force is  $<10.0 \text{ kJ mol}^{-1} \text{ nm}^{-1}$ . Before the production simulation, we carried out the three system equilibration steps: (i) 100 ps in the NVT ensemble, followed by two additional steps of (ii) 500 and (iii) 1000 ps, both in the NPT ensemble. Harmonic position restraints with a force constant of  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$  were imposed on the protein backbone in steps i and ii. The production step was performed under the same conditions as step iii.

Production simulations were performed at a constant pressure of 1 bar using the Parrinello-Rahman [51] algorithm with a relaxation time of 2 ps and isothermal compressibility of  $4.5 \times 10^{-5} \text{ bar}^{-1}$ . A stochastic velocity-rescaling thermostat was used to control the temperature with a 0.1 ps period [51]. Periodic boundary conditions were applied, and a cutoff of 1.2 nm was used for short-range interactions. Long-range electrostatic interactions were calculated by Particle-Mesh Ewald (PME) [52] summation method with a fourth-order interpolation and a

**Table 1**

Number of water and glucose molecules, ions, and total particles for each simulated system. The glucose bulk concentrations are indicated in mM.

	0 mM	125 mM	250 mM	500 mM	1000 mM
Water	41,658	40,754	39,847	38,044	34,447
Glucose	0	93	187	374	747
Na <sup>+</sup>	9	9	9	9	9
Particles	132,472	131,992	131,527	130,606	128,767

grid spacing of 0.16 nm. All bonds involving hydrogen atoms were constrained with the LINCS algorithm [53]. To integrate the equations of motion, the leap-frog algorithm [54] was used with a time step of 2 fs.

All MD simulations were carried out at 323.15 K (50 °C) and the protonation state of the BGHI residues was set to reproduce the pH of 6.0 by using PypKa [55]. Such values were chosen because 50 °C and pH 6.0 are the optimum conditions of BGHI according to the reference experimental assays [7,9].

For each glucose concentration, and in water, four independent simulations of 1000 ns each were performed (refer to Supporting Information Fig. S1 for an overview of the systems simulated). The four glucose concentrations studied were 125 mM, 250 mM, 500 mM, and 1000 mM, to mimic the same conditions found in the biochemical assays reported in the literature [7,9,16,17]. The Packmol [56,57] program was used to solvate BGHI in water and at different concentrations of cosolvent, as well as to add ions to achieve charge neutrality in the system (9 Na<sup>+</sup> ions were added to each system for this purpose). The simulated systems have around 130,000 particles (see Table 1), inside a cubic box with an edge of 110 Å.

We used the PackmolInputCreator tool from the Julia [58] `MolSimToolkit.jl` package to estimate the number of glucose molecules to insert in the simulation box in order to reproduce the bulk concentration. This estimation was based on density values obtained for aqueous solutions of glucose at room temperature (298 K) [59,60]. The target and effective glucose bulk concentrations are discretized in Table S1.

We employed the post-simulation analysis tools of GROMACS for the time-dependent RMSD analysis, using the crystal structure of BGHI as a reference for the structural fitting. The flexibility of the protein residues was assessed using the MDLovoFit software [61], which performs a robust alignment that classifies the regions of the protein as rigid or mobile [62,63], and the alignment is based on the rigid subset only. We considered rigid all  $\alpha$  that could be aligned with an RMSD smaller than 1.0 Å. In MDLovoFit, if the residue is found in the rigid part (RMSD  $<1.0$  Å) of the protein sequence, it is assigned an occupancy value of 1.0. On the contrary, it has an occupancy value of 0.0 and is considered as flexible. For the RMSD analysis, we employed 1000 frames equally spaced from 0 to 1000 ns for each replica (4000 frames for each system). For the other analysis, we used 10,000 frames equally spaced from 0 to 1000 ns for each replica (40,000 frames for each system).

We used the Visual Molecular Dynamics (VMD) [64] software to visualize the MD trajectories and produce images containing molecular structures. The Julia `MolSimToolkit.jl` package was used to assign secondary structures according to the STRIDE algorithm [65,66]. To process and represent the data obtained from MD analysis we used mainly Julia packages such as `Plots.jl`, `KernelDensity.jl`, and `StatsPlots.jl`. Distance distributions were computed with the aid of the `CellListMap.jl` library [67].

### 2.2. Minimum-distance distribution functions and solvation thermodynamics

In order to characterize the protein-solvent interactions, we use here minimum-distance distribution functions (MDDFs), Kirkwood-Buff (KB) integrals, and the preferential interaction parameters, which were computed with the `ComplexMixtures.jl` package [45]. In the present study, we are investigating ternary solutions containing the protein

(species  $p$ ), water (species  $w$ ), and glucose as cosolvent (species  $c$ ). The protein is treated as a solute at infinite dilution and the molar concentrations of water and cosolvent are, respectively,  $\rho_w$  and  $\rho_c$ . The solvent distribution around the protein can be described in terms of the average density of molecules at a distance  $r$  from the protein surface,  $n(r)$ , relative to the density at the same distance but in absence of protein-solvent interactions (ideal-gas distribution),  $n^*(r)$ , expressed as follows in Eq. (1):

$$g_{pi}^{md}(r) = \frac{n_i(r)}{n_i^*(r)} \quad (1)$$

where  $i$  represents the species for which the distribution is being calculated ( $w$  or  $c$ ) and  $r$  is the distance between the protein and the cosolvent. The  $md$  index indicates that here we are using the minimum distance between any atom of the protein and any atom of the solvent, which defines the minimum-distance distribution function (MDDF) [44].

MDDFs are powerful tools for the study of macromolecules in solution. Despite the asymmetric shape and irregular surface of proteins and other biopolymers, MDDFs retain a clear structural meaning [44,45]. In practical terms, the MDDFs describe how the protein impacts the arrangement of solvent around it, by estimating the deviation from ideality throughout the protein's region of influence (also called the "protein domain"). Distances for which the MDDF value is  $>1.0$  represent regions where the density of a solvent component has increased due to the presence of the protein, and indicate the formation of solvation shells. The MDDFs can, furthermore, be decomposed into the contributions of each atom or group of atoms of the system, providing valuable information about the protein-solvent interactions; this will be exemplified later in this text.

Last but not least, MDDFs can be used, when properly normalized, to compute KB integrals and preferential interaction parameters [44], connecting the microscopic aspects of the protein solvation with macroscopic thermodynamic observables [31,68]. A series of examples of the use and versatility of MDDFs can be found in the ComplexMixtures.jl package documentation, at <https://m3g.github.io/ComplexMixtures.jl>, and additional applications and technical details about their implementation can be found in previous publications [44,46,69,70].

The KB integrals can be computed from minimum-distance counts, by

$$G_{pi}(R) = \frac{1}{\rho_i} [N_{pi}(R) - N_{pi}^*(R)] \quad (2)$$

where  $\rho_i$  is the density of component  $i$ , and  $N_{pi}(R)$  and  $N_{pi}^*(R)$  are, respectively, the number of minimum distances between the protein and the solvent species smaller than  $R$ , and the number of equivalent distances within  $R$  in a system with ideal gas distribution (i.e. in the absence of solute-solvent interactions), with the same density as that of the bulk solvent [44]. These ideal distributions of the solvent have to be explicitly generated in the case of MDDFs, and this is implemented in the ComplexMixtures.jl package [45].

The integral expressed in Eq. (2) stands for the excess volume occupied by the cosolvent in the protein domain, relative to the volume that the cosolvent would occupy if there were no solute-solvent interactions [71]. Favorable protein-solvent interactions increase the KB integral value. On the other hand, if the protein-solvent interactions are unfavorable, the solvent's concentration in the protein domain will be smaller in comparison to bulk, reducing the KB integral.

The preferential solvation (or interaction) parameter,  $\Gamma_{pi}$ , in turn, is a thermodynamic quantity that can be experimentally measured by techniques like equilibrium dialysis and vapor pressure osmometry [72,73], associated to the change in the chemical potential of the protein,  $p$ , in response to a variation in the concentration of a given solvent component ( $i$ ), in this case the water ( $w$ ) or the cosolvent ( $c$ ) [31,74].

The preferential solvation parameter can be computed from the difference of KB integrals of the solvent components and gives the information of which component is preferentially bound to the protein surface [75,76]. As previously stated, here we are dealing with ternary solutions (a protein solute infinitely diluted in an aqueous solution of glucose). In this case, the preferential interaction parameter of the cosolvent relative to the water can be approximated by

$$\Gamma_{pc}(r) \approx \rho_c [G_{pc}(R) - G_{pw}(R)] \quad (3)$$

where  $G_{pc}(R)$  and  $G_{pw}(R)$  are the KB integrals of the cosolvent,  $c$ , and of water,  $w$ , computed up to a finite distance  $R$  from which solute-solvent interactions are negligible.

A positive value for  $\Gamma_{pc}(R)$  means that the cosolvent accumulates preferentially relative to water in the protein domain, in other words, the protein is preferentially solvated by the cosolvent. On the other hand, a negative value for  $\Gamma_{pc}(R)$  means that the protein interacts preferentially with the water molecules, that is, it is preferentially hydrated.

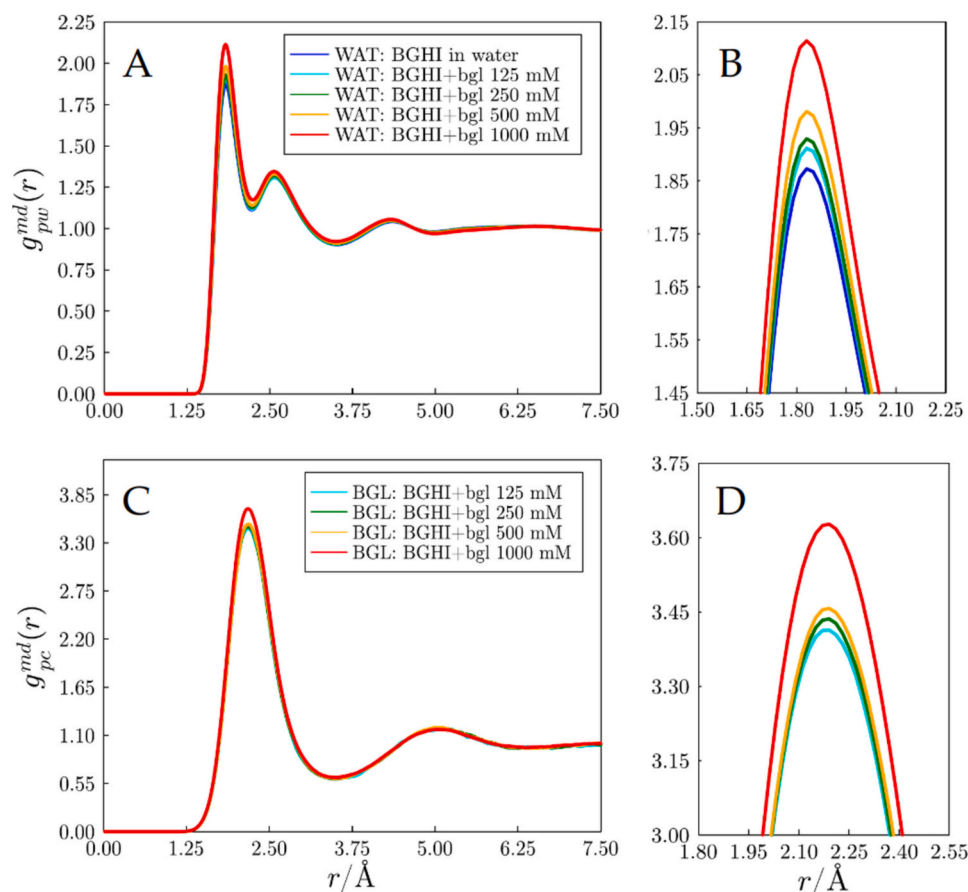
In this study, specifically, we have calculated the MDDFs using a discretized version of Eq. (1). To compute the MDDFs, minimum-distance histograms of the true and ideal-gas distributions were evaluated with a 0.1 Å bin-size. We employed 40,000 frames from our simulations (for each system - 10,000 frames per replica) in the calculation of both MDDFs and KB integrals. The KB integrals and preferential interaction parameters for the cosolvent were computed according to Eqs. (2) and (3), respectively, and using  $R = 15$  Å. The computational protocol employed here resulted in an excellent convergence for the KB integrals and  $\Gamma_{pc}$ , as seen in Section. 3.1.

### 2.3. Selection and analysis of glucose dimers

We have processed 40,000 frames for each concentration and extracted the frames containing at least 2 molecules of glucose within 6.0 Å from E166 and E377. The glucose molecules found within this region were classified into BGL1–4, according to the distances from the catalytic dyad. Tables S6 and S7 provide, respectively, the number of BGL molecules and the average distance of each BGL for each system. With the extracted frames, we produced trajectories with only the protein and glucose molecules, using the rigid part of the BGHI as reference for the structural alignment. The glucose dimer was then defined as the two glucoses (BGL1 and BGL2) closest to the catalytic residues. After that, we applied a distance analysis to select dimers that resembled cellobiose. The linkage between the glucose monomers in cellobiose occurs between carbons C1 and C4, so we computed the distances BGL1: C1-BGL2:C4 ( $d_1$ ) and BGL2:C4-BGL1:C1 ( $d_2$ ) for all the frames selected. Furthermore, we monitored the distance,  $d_3$ , between the centers of mass of the catalytic residues ( $CM_1$ ) and the glucose dimer ( $CM_2$ ) (see Fig. 9). The distances  $d_1$ ,  $d_2$  and  $d_3$  are represented in Fig. 9B. To account for the three distances simultaneously, we defined the following score function:  $S = 1000 / (d_1 + d_2 + d_3)$ ; the outcomes are shown in Figure 10 and S7, as well as in Table S3.  $S < 50 \text{ \AA}^{-1}$  was considered as the threshold for best-scoring glucose dimers. This  $S$  value is associated with an average distance of  $<7$  Å, a typical distance used to evaluate contacts in protein structures, and which is consistent with the two molecules being close to each other and in the vicinity of the active site.

### 2.4. Glucose residence time and active site occupancy

For both the estimation of glucose residence time and active site occupancy, we applied the intermittent autocorrelation analysis, first described by Rapaport [77] to obtain hydrogen bond lifetimes from water simulations and later employed by Garcia et al. [78] and Makarov et al. [79] to estimate the residence time of water around protein groups. By using VMD [64], we first verified the presence of glucose molecules in the BGL1 sites – the closest glucose from E166 and E377 within a radius



**Fig. 1.** A and B) MDDF of water (WAT) and C and D) of glucose (BGL) relative to BGHI at different concentrations. The hydrogen bonding peaks increase both for water and glucose with the increase in the concentration of the sugar.

of 6.0 Å in each frame. Next, we obtain the probability of finding a glucose molecule in the BGL1 site at step  $t_0 + t$  given that it was present at step  $t_0$ , which we denote  $P(t)$ . This procedure can be expressed in terms of the following autocorrelation function,

$$P(t) = \frac{1}{N_G} \langle S(t_0) \cdot S(t_0 + t) \rangle_t \quad (4)$$

where  $N_G$  is the number of glucose molecules, and  $S$  is a binary vector of length  $N_G$ , whose elements are 1 if the glucose is at the BGL1 distance (see Table S7), and 0 otherwise. The  $\langle \dots \rangle_t$  indicates an average over all frames with a time-interval of  $t$  along the simulations. We emphasize that an intermittent autocorrelation function evaluates the glucose binding and unbinding not considering events that occur at time intervals smaller than  $t$  (the smallest time-interval here was in this case  $t = 100$  ps, determined by the sampling frequency of frames of the simulation). This analysis was carried out using the same set of frames as in the glucose dimer selection step mentioned in the previous section. The correlation calculation was implemented in the intermittent correlation function of [MolSimToolkit.jl](#) v1.22.0, and used to compute  $P(t)$  up to a maximum of 300 ns. We calculated  $P(t)$  for the glucose-free active (intermittent survival probability) site using the same dataset but accounting for the frames in which  $S_j = 0$ , that is when there are no glucose molecules within 6.0 Å from the catalytic residues.

## 2.5. Cavity volume estimation

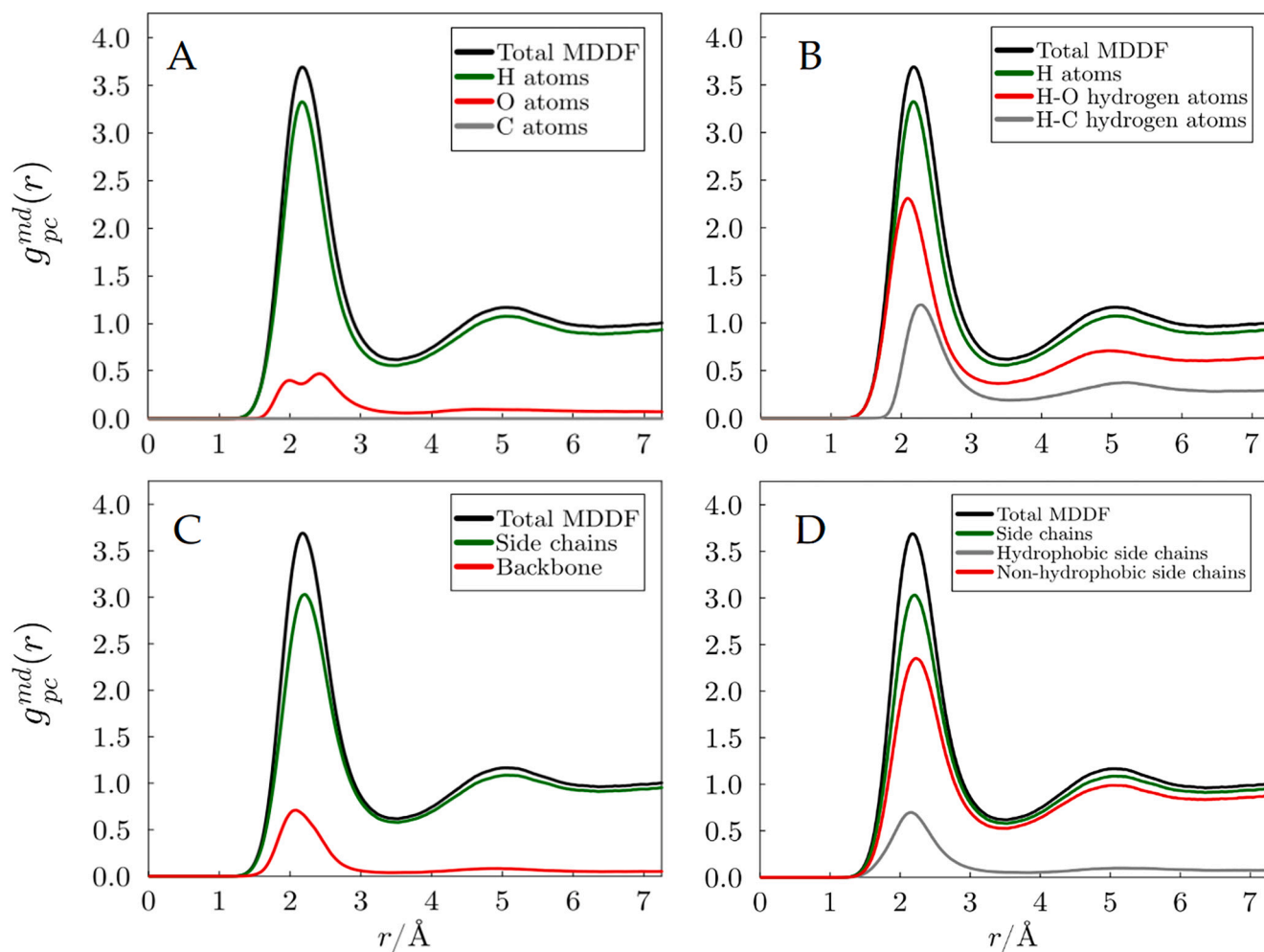
The volume of the cavity containing the BGHI active site was calculated using the Epoch software [80] as a graphical user interface plugin for VMD. We defined a spherical inclusion zone of 15 Å centered

on residues H120, W121, E166, C169, L172, L173, W168, N235 and E377, in which E166, and E377 are the catalytic dyad. The other selected residues are nearby residues E166 and E377 and have bulky side chains. Some of them (e.g. L173 and W168) have been described previously as gatekeeper side chains that contribute significantly to the volume and shape of the substrate cavity [8,20]. The convolution of spheres with a radius of 4.0 Å was used to determine the volume of the substrate cavity within the inclusion region. Exclusion zones were also defined to prevent spheres from being created in other smaller cavities within the inclusion region. They consisted of 8 spherical zones with a radius of 6.0 Å centered on the C $\alpha$  atoms of the residues D122, F134, PHE164, S170, G174, P200, F259 and F264. For each glucose concentration, we analyzed 40,000 frames selected from the MD trajectories (from 0 to 1000 ns). With the extracted frames, we produced trajectories with only the protein aligned on the rigid part of the BGHI. Only residues from the rigid part were taken into consideration while defining the inclusion and exclusion zones, preventing these zones from shifting over the aligned structures. See Fig. S5 for a graphical explanation of the cavity volume estimation.

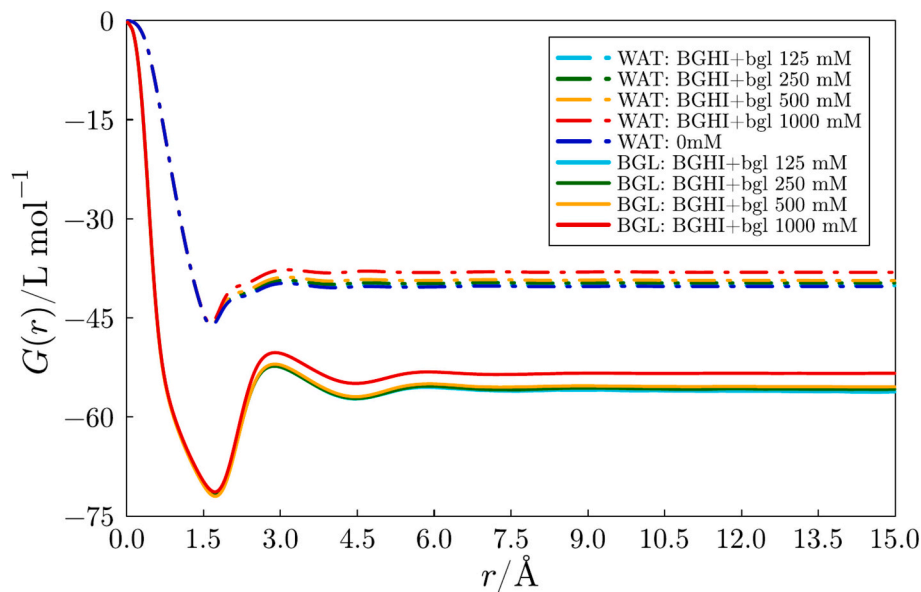
## 3. Results and discussion

### 3.1. Protein solvation structure and thermodynamics

The understanding of the structure and thermodynamics of protein solvation can be achieved by means of distribution functions, here minimum-distance distribution functions (MDDFs), and the Kirkwood-Buff (KB) theory, to connect the molecular distribution with the macroscopic thermodynamics of the solution. In this session, we review the key findings from those analyses. Fig. 1 shows the MDDFs for water



**Fig. 2.** Decompositions of the MDDFs of glucose relative to the protein. The total glucose MDDF is shown in black. (A) Contributions from the glucose atoms: hydrogen (green), oxygen (red), and carbon (gray). (B) Contributions from all glucose hydrogen atoms (green), as well as from H-O (red) and H-C (gray) hydrogen atoms. (C) Contributions from protein side chains (green) and backbone (red) atoms. (D) Contributions from protein side chains (green), as well as from hydrophobic (gray) and non-hydrophobic side chains (red).



**Fig. 3.** KB integrals for water (WAT) and glucose (BGL) at different concentrations (125–1000 mM) from 0 to 15 Å. The glucose concentration of 0 mM indicates the BGHI in water.

**Table 2**

Preferential solvation parameters for the systems containing glucose at the specified concentrations.

	125 mM	250 mM	500 mM	1000 mM
$\Gamma_{pc}$	-2.038	-4.115	-8.302	-16.041
STD	0.001	0.002	0.004	0.001

and glucose in various concentrations.

The presence of the protein and its interaction with the solvent molecules results in the formation of three and two distinct solvation shells for water and glucose, respectively. The width of the first peak (1.5–3.0 Å) for the MDDF of glucose indicates that it makes both specific (hydrogen bonds) and non-specific (van der Waals or indirect) interactions with the BGHI surface (see Fig. 1C). While the former is characterized by peaks relatively close to the solute (typically 1.5–2.0 Å - for the distance between the polar Hydrogen and the electronegative partner), the latter affects MDDFs at longer distances. Fig. 1B and D provide a zoomed-in view of the first peak of the MDDFs for water and glucose. As can be observed,  $g^{md}(r)$  increases gradually in accordance with the glucose concentration. This indicates that water and glucose interact directly with the protein and that both substances compete for interaction sites on the BGHI surface. The increase of the hydrogen-bonding peaks reflects some degree of stabilization of these specific interactions, which may be due to the greater viscosity of the solutions.

As introduced in Section 2.2, it is possible to obtain the contribution of different atoms (or groups of atoms) to the total MDDF of a system component. In Fig. 2A, we show the decomposition of the total glucose-protein MDDF into types of elements of glucose. It is possible to see that hydrogen atoms contribute more to the total MDDF of glucose than do carbon and oxygen atoms. The peaks related to oxygen atoms are due to the interactions that the oxygen from the pyran ring makes with the protein. The same MDDF decomposed into the types of hydrogen contributions (Fig. 2B) shows that H-O hydrogen atoms (those linked to hydroxyl groups) have the greatest contribution to the total glucose-protein MDDF. Following the same approach, we also decomposed the contributions of protein atoms to the glucose-protein MDDF (Fig. 2C and D). We found that the largest contributions come from non-hydrophobic side chains, with a small contribution from the backbone at shorter distances. Thus, these analyses indicate that polar interactions of glucose with the protein are the dominant local short-ranged interactions that attach the glucose to the protein surface.

The KB integrals for glucose and water indicate that the BGHI is preferentially hydrated in all cosolvent concentrations, as water KB integrals are always greater than glucose KB integrals, such that  $\Gamma_{pc} < 0.0$  in all systems (Fig. 3 and Table 2).  $\Gamma_{pc}$  decreases with the glucose

concentration. The KB integrals vary only slightly among the concentrations, which implies that the preferential hydration is roughly proportional to the concentration of the cosolvent. As introduced before, this is expected for proteins in aqueous solutions of low molecular weight osmolytes such as glucose [31,68,71,81,82]. A possible structural consequence of preferential hydration is discussed in Section 3.6, where we show that the increase in glucose concentration induces the relative closure of the substrate cavity.

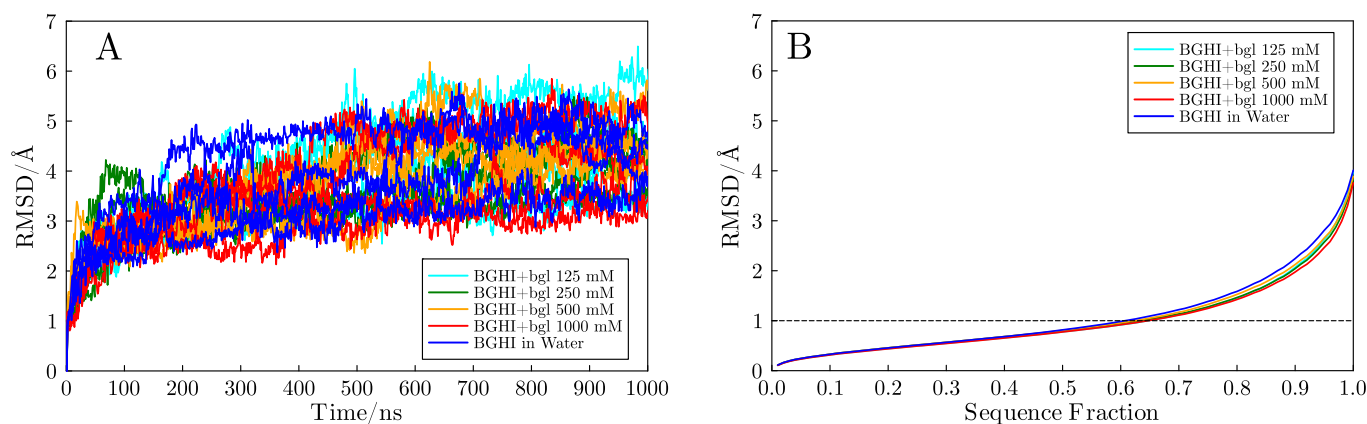
### 3.2. Protein flexibility at different glucose concentrations

Another goal of our research was to attempt to link the findings of solvation analysis to the structural dynamics of BGHI. However, our objective here was not to scrutinize the conformational dynamics of BGHI in solution; an effort in this direction was made by Lima et al. [38]. Instead, we focused on investigating the flexibility of BGHI residues and whether glucose concentration affects the flexibility pattern along the protein sequence. This flexibility analysis was critical for standardizing the conformational analysis of glucose dimers at the active site, as well as the cavity volume analysis, reported in Sections 3.4 and 3.6, respectively. The results of the BGHI structural dynamics analysis are summarized in Figs. 4 and 5.

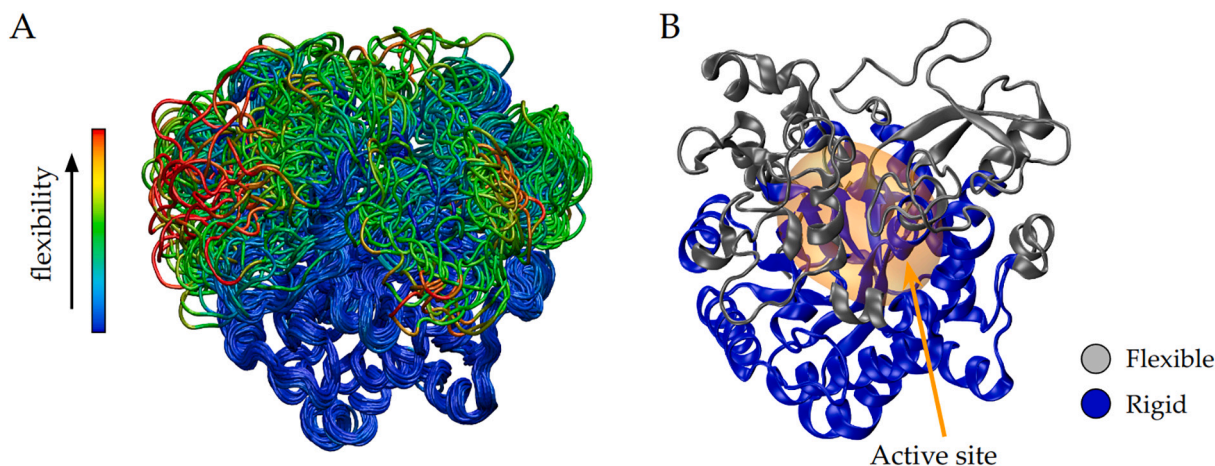
Using the MDLovoFit software [61], we found that the BGHI structure is on average 40 % flexible, meaning that in the time scale of our simulations, roughly 60 % of the protein can be aligned to Root Mean Square Deviation (RMSD) value smaller than 1.0 Å (Fig. 4B). It was possible to characterize the flexibility of BGHI residues individually, for each cosolvent concentration. The flexibility pattern of BGHI is very similar for all simulated conditions. In both water and in glucose, the most flexible regions are those at the entrance of the substrate cavity that comprise the active site (Fig. 5) — from now on, whenever we say “BGHI cavity” or simply “cavity” we will be referring to the main cavity of the BGHI, which holds the catalytic residues E166 and E377.

Here, we will refer generically as “flexible” to the BGHI segments composed mainly of flexible residues that dominate the structural dynamics of the enzyme. They are represented in gray colour in Figs. 5B and, and are, in fact, the opposite selection of the “rigid” segments common to all conditions we have simulated, which contains only rigid residues (see Figs. S3 and S4). Within the flexible part of the BGHI, we have found highly flexible segments that we call as “F segments”. They are composed of five individual segments (F1–5) that surround the BGHI cavity.

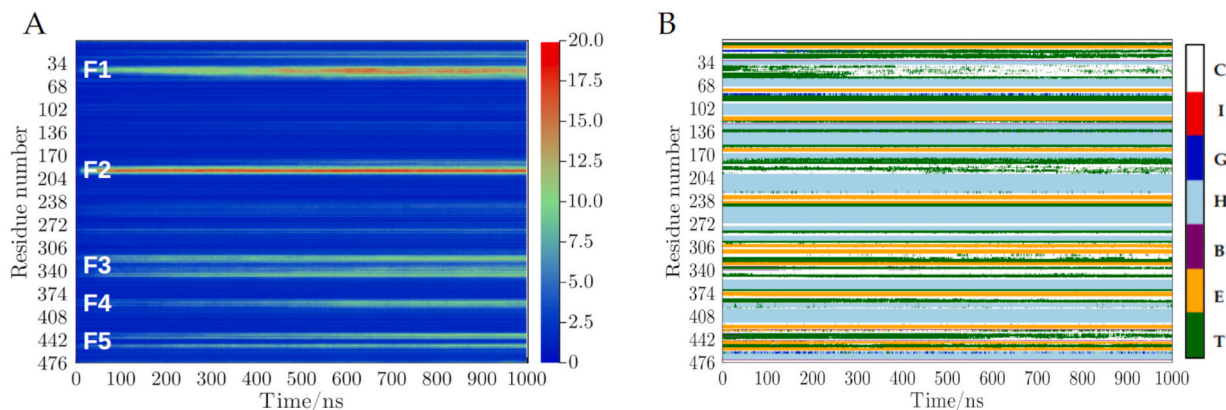
We observed that the crystallographic model of BGHI is more compact than the conformations obtained during simulations, meaning that an increased RMSD value generally indicates the opening of the enzyme cavity (see Fig. S14). The average RMSD values of the F



**Fig. 4.** RMSD analysis of BGHI backbone atoms with respect to the crystal structure. (A) Time-dependent RMSD plot for the BGHI in different glucose concentrations (125–1000 mM) and in water, indicating overall displacements in the 2–5 Å range. (B) RMSD versus sequence fraction of BGHI. The plot shows that it is possible to align 60 % of the BGHI structure with an RMSD < 1.0 Å.



**Fig. 5.** (A) Structural superposition of frames obtained from MD simulations of BGHI in water and in the presence of glucose (125–1000 mM) after robust alignment with MDLovoFit. The gradient indicates the flexibility with respect to the crystal structure of the BGHI. High RMSD values are indicated in red, and low RMSD values are indicated in blue. BGHI exhibited a very similar flexibility pattern for all conditions investigated. (B) Flexible (blue) and rigid (red) regions of the BGHI are projected over the enzyme crystal structure. The flexibility of the residues was determined from MD simulations by using the MDLovoFit software [61]. Here, “Flexible” means the opposite selection of the rigid part and is mostly composed of flexible residues (see Supporting Information Figs. S3 and S4 for details).



**Fig. 6.** (A) RMSD per residue analysis. The values represent the average over all simulated systems from 0 to 1000 ns (similar figures for independent concentrations are shown in the Supporting Information). The regions F1–5 are also indicated. (B) Secondary structure (SS) analysis according to the STRIDE algorithm. The colors indicate the SS types: namely turn (T, green),  $\beta$  strand (E, yellow),  $\beta$  bridge (B, purple),  $\alpha$  helix (H, light blue), 310 helix (G, blue),  $\pi$  helix (I, red) and random coil (C, white). The values represent the mode for each frame from 0 to 1000 ns, taking all simulations together. See Figs. S10 and S11 for the results obtained individually for each system.

segments in pure water are larger than those in the presence of glucose, suggesting that the preferential hydration may favor the structures with more closed cavities. Later on, in Section 3.6, we will revisit this topic as we discuss the BGHI cavity volume during the MD simulations.

We also evaluated the content of the secondary structure in the different conditions studied. And, as were found for the residue flexibility, very similar results were obtained for all systems. This allowed us to summarize the representation of the results in Fig. 6, which shows the average RMSD value and the mode for the discrete SS values, both obtained for each frame between 0 and 1000 ns (considering all the simulations together). For the results obtained for each system separately, see Figs. S10 and S11.

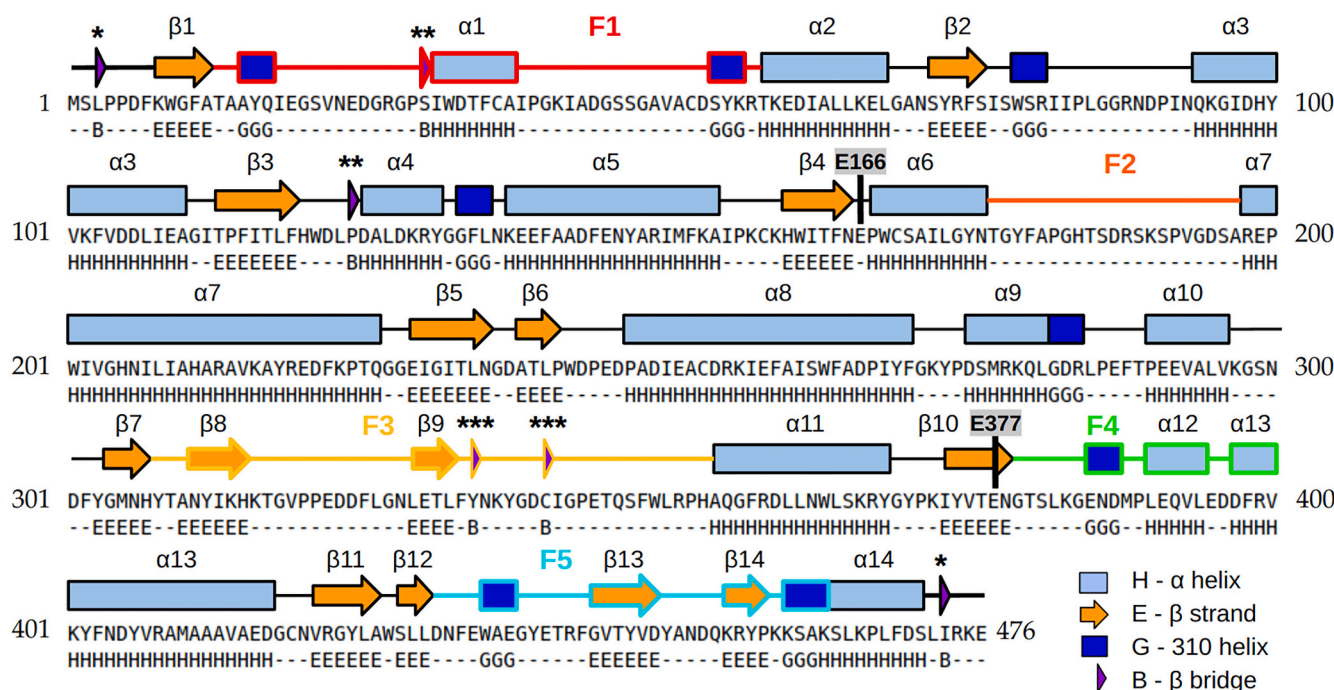
Comparing the RMSD and SS maps of Fig. 6, we note the F segments (F1–5) are mainly formed of turns and random coils. The dynamics of the F segments may have an influence on the volume of the BGHI cavity, as will be discussed in Section 3.6, but has little impact on the total SS content. The percentages of SS obtained in the MD simulations are very similar to those of the BGHI crystal structure (Fig. S15), except for the fact that the crystal has a greater 310 helices content and less random coil content. Most of the 310 helices are found within the F segments and

exhibit remarkable flexibility. Throughout the simulations, the unstable 310 helices are primarily replaced by turn structures and  $\alpha$  helices. Furthermore, the highest content of random coils in MD structures is due to the destabilization of turn structures, located especially in segments F1 and F2.

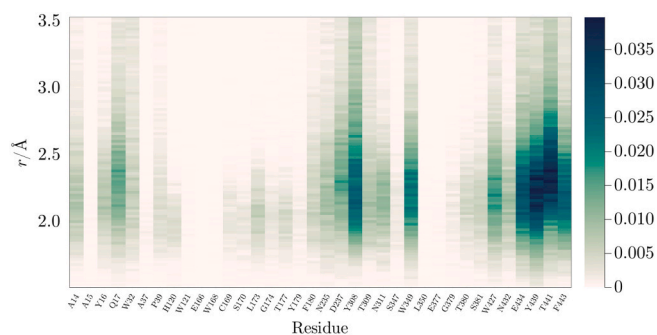
The SS contents reported in the present study are similar to those estimated experimentally by Souza et al. [7,17], which performed the biophysical characterization of both the native and recombinant BGHI through far-ultraviolet circular dichroism (UV-CD) spectroscopy. According to the authors, the BGHI in buffer solution presented 29–32 % of  $\alpha$  helix and 12–18 % of  $\beta$  strand, which is close to the results reported in Fig. S15 and Table S2.

### 3.3. Interaction between the glucose and the active site residues

The presence of glucose molecules was inspected at the enzyme active site along the simulations. We took advantage of the fact that glucose eventually enters with the active site to map the main glucose-interacting protein residues. By using VMD, we selected all frames with at least one glucose molecule in the BGHI active site (defined as 6.0



**Fig. 7.** Secondary structure map of the BGHI crystal structure. For clarity, both turn and random coils are represented by dashes (-). We also indicate the regions of high flexibility (F1–5) and the position of the catalytic residues E166 and E377. The stars (\*) denote  $\beta$  bridges that are connected to each other.



**Fig. 8.** Contributions of the active site BGHI residues to the glucose MDDF. The gradient indicates the increase in glucose density due to the presence of the respective protein residues. In the figure, we report the results for 1000 mM of glucose. The contributions vary from 0.0 to 0.04 (from white to dark green). Similar patterns were obtained in other glucose concentrations (Fig. S6). The figure was generated with [ComplexMixtures.jl](#) [45] and [Plots.jl](#) [83].

Å from the E166 and E377 residues). The presence of a maximum of 4 glucose molecules in the active site was verified in 11 of 40,000 frames analyzed (2 frames from the simulation at 500 mM and 9 at 1000 mM). Thus, the simulations indicate that the interaction sites close to the catalytic residues can accommodate a maximum of 4 glucose molecules.

We selected all protein residues within 3.0 Å from the 4 glucose molecules in the 11 frames extracted previously, resulting in a list of 36 potentially glucose-interacting residues at the active site. MDDF density maps were then constructed with the contributions of these 36 protein residues to the distribution functions, for each concentration of glucose. These density maps, as that shown in Fig. S6, illustrate the residues that are most likely to interact with the cosolvent. The maps plotted for 125 mM and 250 mM resulted in very noisy densities due to the small number of glucose molecules in these systems. Nevertheless, the interaction profiles do not depend qualitatively on the glucose concentration. The map for 1000 mM, shown in Fig. 8, provides an accurate representation of the interactions between the glucose and BGHI active site

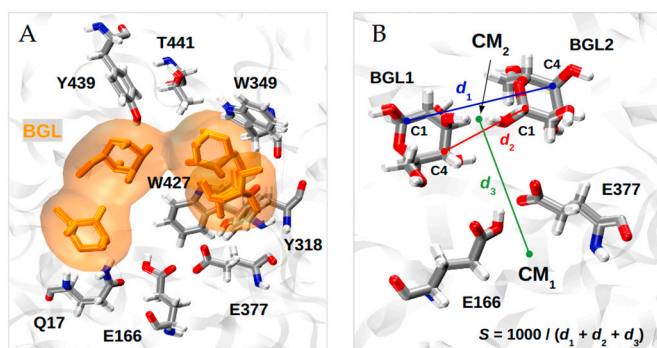
residues at all concentrations evaluated.

Fig. 8 shows the contributions of the active site residues of BGHI to the MDDF for 1000 mM between 1.5 and 3.0 Å. Surprisingly, E166 and E377 are not among the residues with greater nearby glucose density, which means that glucose interacts much more frequently with other residues close to the active site, so that E166 and E377 are “over-shadowed” in the density map. It is possible to observe highly dense regions centered on residues Q17, Y308, W349, W427, Y439, and T441. Other high-density areas can be attributed to the presence of residues A14, P39, H120, S170, L173, T177, F180, N235, N311, T380, S381 and F443. The densities obtained in this analysis mainly represent the interaction between glucose and the side chains of the residues since the interaction with the backbone has little contribution to the cosolvent total MDDF (Fig. 2C). However, regions of medium or low density below 2.0 Å obtained for nonpolar residues such as A14, P39, L173, F180, and F443 indicate hydrogen-bonding interactions with the protein main chain.

Significant densities at greater distances (beyond 3.5 Å), were obtained for residues W32, Y308, N311, Y439 and F443 (Supporting Information Fig. S7), indicating that these residues have a long-range influence in the cosolvent organization, which may seem unlikely for residues that are located deep within the active site. However, it is worth noting that the cavity structures obtained from MD are not as compact as that found in the crystal structure. The BGHI structure relaxes along the simulations, causing the active site cavity to open up. In fact, this is the reason why we decided to call it a “cavity” instead of a “channel”, as it is generally referred to in other studies. As a result, active site residues that strongly interact with the cosolvent can eventually contribute to the second solvation shell formation (at ~5.0 Å), similar to those on the BGHI surface.

Some glucose-interacting residues, highlighted here, are described in the literature as having functional relevance. Residues Q17, L173, Y308, W349, and N235 have already been reported previously by Costa et al. [20] as important for the interaction with both glucose and cellobiose. According to the authors, such residues play a role in a series of molecular events that aid glucose release from the BGHI cavity while cellobiose remains inside it. Also, according to Moleiro et al. [9],





**Fig. 9.** (A) Random configuration from MD simulation at 1000 mM of glucose showing the active site occupied by four glucose molecules (BGL, steaks and surfaces in orange). To facilitate visualization, the glucose hydrogen atoms were hidden. Residues that exhibit strong interactions with glucose (according to Fig. 8 analysis) are also indicated. (B) Glucose dimer from a random MD frame obtained at 1000 mM of glucose. The image also shows the catalytic dyad (E166 and E377), the glucose molecules BGL1 (closer to the catalytic dyad), and BGL2 (furthest from the catalytic dyad). The distances  $d_1$ ,  $d_2$ , and  $d_3$  used in the scoring analysis are represented in blue, green, and red, respectively. CM1 and CM2 stand for the two centers of mass used for the calculation of  $d_3$  (see Section 2.3).

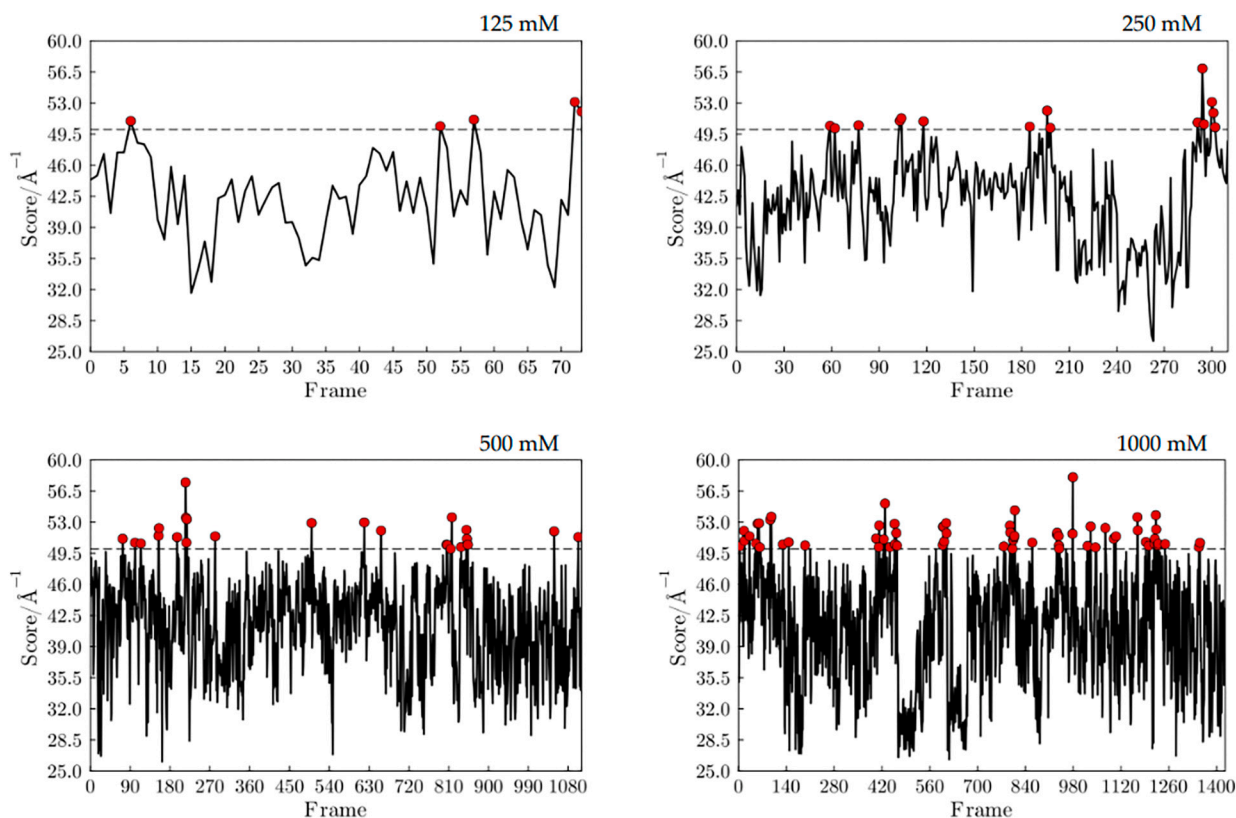
mutations in residues N235 and D237 can have a direct effect on the balance between the transglycosylation and hydrolysis reactions performed by BGHI. The same authors highlight the fact that the side chains Y308, N235, and D237 constitute highly conserved positions, suggesting them to be critical for protein function.

### 3.4. Glucose dimers in the active site: obtaining cellobiose-like configurations from MD

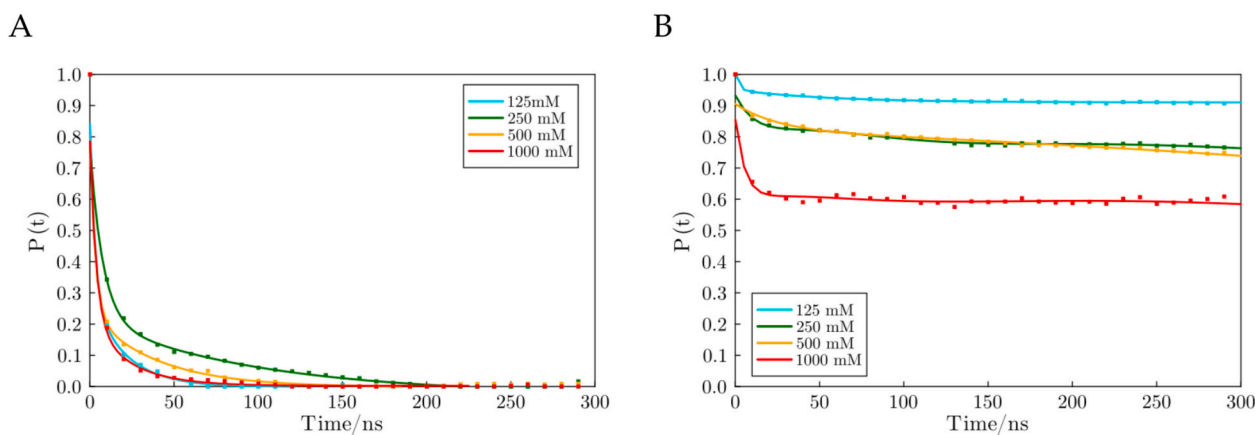
In this section, we discuss the results of the selection of glucose dimers in complex with BGHI. The glucose molecules in the dimer were called BGL1 (closest to the catalytic dyad) and BGL2 (furthest from the catalytic dyad). As described in Section 2.3, the distances used in the scoring analysis were defined as BGL1:C1-BGL2:C4 ( $d_1$ ), BGL2:C4-BGL2:C1 ( $d_2$ ) and CM<sub>1</sub>-CM<sub>2</sub> ( $d_3$ ), where CM<sub>1</sub> and CM<sub>2</sub> are, respectively, the centers of mass of the catalytic residues and of the glucose dimer (Fig. 9).

As mentioned before, we defined the active site as a region within 6.0 Å from E166 and E377. The average active site occupancy is close to one glucose molecule, regardless of concentration (Table S4). Nevertheless, configurations with two or more glucose molecules (necessary for dimer formation) are more frequent at higher concentrations. For this reason, we were able to obtain a better sampling for  $d_1$ ,  $d_2$ , and  $d_3$  at 1000 mM. Consequently, at 1000 mM we also obtained a greater number of frames with a distance-score  $S \geq 50 \text{ \AA}^{-1}$  (see Methods). We observed 5 (125 mM), 15 (250 mM), 29 (5000 mM), and 83 (1000 mM) frames with  $S \geq 50 \text{ \AA}^{-1}$ . The distances  $d_1$ ,  $d_2$ , and  $d_3$  are plotted individually for each concentration in Fig. S8. The scoring analysis is shown in Fig. 10. At 125 mM, the concentration at which the enzyme is predominantly stimulated by glucose, we found a greater proportion of frames with  $S \geq 50 \text{ \AA}^{-1}$  (about 7 % of the selected frames, see Table S3).

The current analysis attempts to demonstrate that the active site conformations derived from our simulations are compatible with those of the enzyme in association with the substrate, and hence capable of carrying out the hydrolysis or transglycosylation reaction. As illustrated in Fig. 10, frames with  $S \geq 50 \text{ \AA}^{-1}$  were observed at all concentrations. This provides insight into the BGHI active site's integrity, which appears to remain functional even at high concentrations of the cosolvent. Such findings are consistent with experimental studies demonstrating that,



**Fig. 10.** Scoring analysis for glucose dimers near the catalytic residues E166 and E377 performed for each system containing the cosolvent. The score function was defined as  $S = 1000 / d_1 + d_2 + d_3$ . See Fig. 9 for distance definitions. The distances  $d_1$ ,  $d_2$ , and  $d_3$  are plotted individually in Fig. S8. Red dots indicate the dimers with  $S \geq 50 \text{ \AA}^{-1}$ .



**Fig. 11.** (A) Intermittent residence probabilities for glucose molecules closest to catalytic residues (BGL1 sites). (B) Intermittent survival probability of the glucose-free active site. The dots in both figures constitute a representative sampling of the original dataset, composed of 3000 frames from 0 to 300 ns for each concentration. The curves in both figures represent the average for the time series among replicas for each system (see Figs. S12 and S13). Bi-exponential fits are shown (See Supplementary Tables S8 and S9).

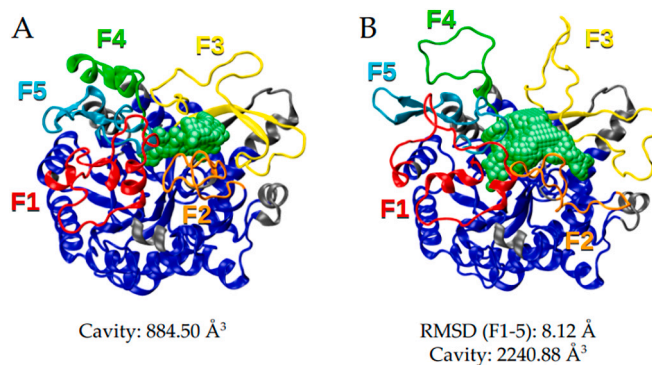
although glucose inhibits the enzyme, it retains residual glucosidase activity even at high glucose concentrations [9,16,17].

Furthermore, from a methodological point of view, here we demonstrate how it is possible to sample the protein conformation at a specific glucose concentration and, at the same time, obtain different configurations for the reaction product in the active site. Such structures can be used to analyze enzyme-substrate interactions and reaction processes through hybrid quantum mechanics/molecular mechanics (QM/MM) calculations. We have made the selected structures available to the public through the link: [https://github.com/m3g/2024\\_Ramos\\_Martinez\\_BGHI](https://github.com/m3g/2024_Ramos_Martinez_BGHI).

### 3.5. Residence time profiles for glucose and active site occupancy

Residence time probabilities for glucose and active site occupancy were computed to probe if at high concentrations glucose enters the active site more frequently, or if the increase in concentration is related to glucose residence time in interaction sites near the catalytic residues. Residence time probabilities from our simulations for glucose at BGL1 sites (the closest to catalytic residues, see Fig. 11) through intermittent autocorrelation analysis, as described in Section 2.4. The results show that the residence time for glucose is most likely independent of the enzyme dynamics in the four different cosolvent concentrations. The apparent longer glucose residence time at 250 mM can be attributed to the occasional increase in  $P(t)$  in simulation replica R3 (see Supplementary Fig. S12). This finding suggests that, over the range of concentrations examined, the presence of the cosolvent had little effect on the nature of glucose-active site dynamics.

Fig. 11B, in turn, shows the intermittent survival probability of the glucose-free active site; the curves converge to the average probability of the active site with no glucose molecules: 0.91 (125 mM), 0.76 (250 mM), 0.74 (500 mM) and 0.58 (1000 mM). As expected,  $P(t)$  decreases faster with increasing glucose concentration in the bulk solution (the occupancy of the active site increases with increasing glucose concentration). The occupancy of the active site at 250 mM is higher than expected since, following the linear trend of increasing cosolvent concentration,  $P(t)$  should converge to approximately 0.82 instead of 0.76. Although the  $P(t)$  varied significantly in different MD replicas (see Fig. S13), particularly for the larger concentrations, these results may suggest that the affinity of glucose to the active site at 250 mM is higher than expected and that larger concentrations can lead to competitive inhibition even if the residence time of glucose in the active site remains unchanged.

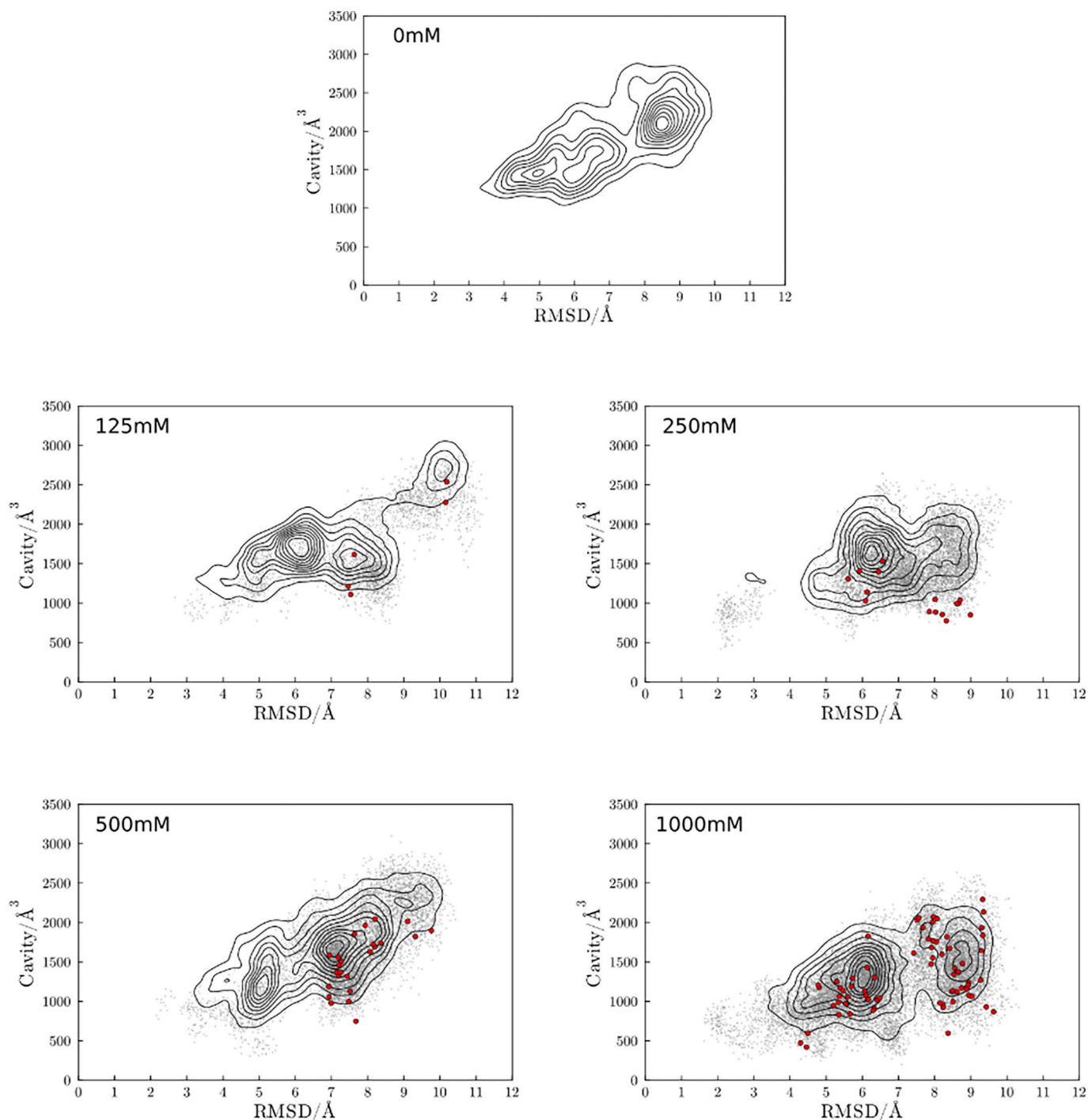


**Fig. 12.** Volumes obtained for the BGHI cavity (green spheres). The regions of the BGHI structure (rigid, flexible, and F1–5), are indicated following the same colour definitions as in Fig. 7. (A) Cavity estimated for the BGHI crystal structure (884.50 Å<sup>3</sup>). (B) Cavity estimated for a random frame extracted from simulations at 125 mM (2240.88 Å<sup>3</sup>). The indicated RMSD (8.12 Å) refers to the value obtained for the backbone of the F1–5 regions using the crystal structure as a reference.

### 3.6. BGHI cavity size and the accessibility of glucose to the active site

The relationship between the shape of the BGHI cavity and the accessibility of glucose to the active site has been a subject of debate [8,12]. Here we have explored the conformational space based on the RMSD value of the F1–5 regions. As we described previously, the F regions are highly flexible regions of BGHI, mainly composed of random coils and turn structures, which delimit the active site cavity. Its mobility can impact both the size and shape of the BGHI cavity. We found that the BGHI cavity can adopt compact and narrow structures, similar to the crystal structure of BGHI, but also open structures comparable to the broad pockets found in the GH3 BGs [8] (Fig. 12).

The RMSD versus cavity volume relationship at different glucose concentrations is shown in the contour map of Figs. 13 and S9. In general, they indicate that at higher glucose concentrations the enzyme tends to adopt less accessible conformations. For 0 mM, the most populated conformations had a cavity volume of about 2250 Å<sup>3</sup>, whereas for 125 mM, 250 mM, and 500 mM it is around 1750 Å<sup>3</sup>. The wide-open conformations are much less common at 1000 mM; at this concentration, the region with the highest density on the map in Fig. 13 is situated around 1250 Å<sup>3</sup>. Such relative closure can reflect the enzyme's preferential hydration. When the protein is preferentially



**Fig. 13.** Distribution of BGHI cavity volumes as a function of RMSD calculated for the F1-5 segments with respect to the crystal structure at each glucose concentration (0 mM stands for the simulation in water). Gray dots indicate frames with the presence of at least one molecule of glucose within 6 Å from the catalytic dyad (E166 and E377). Red dots indicate the glucose dimers with  $S \geq 50 \text{ \AA}^{-1}$  (see Fig. 10).

hydrated, the hydrophobic core of the molecule is less exposed and, as a consequence, the protein tends to prefer collapsed conformations [25,84,85]. Since preferential hydration of BGHI is proportional to the glucose concentration in the medium, the protein is expected to present more compact cavity structures at high concentrations of glucose.

Nonetheless, even the narrow cavity conformations observed at 1000 mM can easily accommodate glucose molecules, so the dependence between cavity shape and glucose accessibility was not verified in our simulations. The gray dots in Fig. 13, which indicate the presence of glucose within 6.0 Å to E166 and E377, are well dispersed throughout

the contour maps. The frames with the best-scored glucose dimers ( $S \geq 50 \text{ \AA}^{-1}$ , red dots) also exhibit this behavior; they simply seem to be more localized at 125 mM, 250 mM, and 500 mM, as opposed to 1000 mM due to limited sampling at these concentrations. As was discussed in Section 3.4, dimer formation around catalytic residues becomes less frequent at low glucose concentrations. However, as the cosolvent concentration increases, the distribution of frames with glucose in the active site and with  $S \geq 50 \text{ \AA}^{-1}$  more closely resembles the total distribution.

#### 4. Conclusions

In this study, we investigated the solvation of the BGHI enzyme in water and at different glucose concentrations (125 mM, 250 mM, 500 mM, and 1000 mM). We have employed minimum-distance distribution functions (MDDFs) and Kirkwood-Buff (KB) integrals as analysis tools. Through MDDFs, we verified that water and glucose interact directly with the protein and compete for sites on the BGHI surface. Using MDDFs it was also possible to determine how water and glucose are structured around the enzyme in different solvation shells, elucidate what types of interactions occur preferentially between the enzyme and the cosolvent, and to predict key interactions between the cosolvent and residues in the active site. The KB integrals revealed that the higher the concentration of the cosolvent, the greater the preferential hydration of the enzyme.

The dynamics of the F segments control the volume of the BGHI cavity. The BGHI cavity tends to adopt more compact structures as the cosolvent concentration increases, which may be related to the preferential hydration of the enzyme in the presence of glucose. However, we found that the ability of glucose to access the active site is not clearly associated with the shape of the BGHI cavity. Also, the enzyme active site tends to be more occupied at high glucose concentrations, which is consistent with a competitive inhibition mechanism. However, the glucose residence time at the active site of BGHI does not seem to depend on the cosolvent bulk concentration. Additionally, we performed a structure selection analysis for glucose dimers close to residues E166 and E377 and were able to extract glucose dimers from the simulations that resemble cellobiose. These structures were made available to the public through the link available in Section 3.4, as well as in the Supporting Information.

Here, we hypothesize that the preferential hydration may contribute to the BGHI stimulation at low glucose concentrations (125–175 mM). At higher concentrations, however, competitive inhibition becomes increasingly significant, until the enzyme is completely inhibited by its reaction product. This could help to explain the activity profile observed experimentally in the enzymatic assays for the BGHI in the presence of small monosaccharides.

#### CRedit authorship contribution statement

**Felipe Cardoso Ramos:** Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Leandro Martínez:** Writing – review & editing, Supervision, Software, Methodology, Data curation, Conceptualization.

#### Declaration of competing interest

The authors state that there are no conflicts of interest related to the research discussed in this publication. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

#### Acknowledgments

The authors acknowledge the financial support of FAPESP (2022/04810-6, 2013/08293-7, 2019/17874-0). Research developed with the help of CENAPAD-SP (National Center for High Performance Processing in São Paulo), project UNICAMP/FINEP-MCTI.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ijbiomac.2024.138210>.

#### References

- [1] X. Li, S.H. Chang, R. Liu, Industrial applications of Cellulases and Hemicellulases, in: X. Fang, Y. Qu (Eds.), *Fungal Cellulolytic Enzymes: Microbial Production and Application*, Springer, Singapore, 2018, pp. 267–282.
- [2] V. Stepankova, J. Damborsky, R. Chaloupkova, Hydrolases in non-conventional media: Implications for industrial biocatalysis, in: P. Grunwald (Ed.), *Industrial Biocatalysis*, Pan Stanford Publishing Pte. Ltd, 2015, pp. 583–629.
- [3] A. Gupta, S.K. Khare, Enzymes from solvent-tolerant microbes: useful biocatalysts for non-aqueous enzymology, *Crit. Rev. Biotechnol.* 29 (2009) 44–54.
- [4] N. Doukyu, H. Ogino, Organic solvent-tolerant enzymes, *Biochem. Eng. J.* 483 (2010) 270–282.
- [5] H. Chen, X. Jin, L. Zhu, Y. Lu, Z. Ma, S. Liu, X. Chen, Glycosyl hydrolase catalyzed glycosylation in unconventional media, *Appl. Microbiol. Biotechnol.* 104 (2020) 9523–9534.
- [6] V. Stepankova, S. Bidmanova, T. Koudelakova, Z. Prokop, R. Chaloupkova, J. Damborsky, Strategies for stabilization of enzymes in organic solvents, *ACS Catal.* 3 (2013) 2823–2836.
- [7] F.H.M. Souza, R.F. Inocentes, R.J. Ward, J.A. Jorge, R.P.M. Furriel, Glucose and xylose stimulation of a  $\beta$ -glucosidase from the thermophilic fungus *Humicola insolens*: a kinetic and biophysical study, *J. Mol. Catal. B Enzym.* 94 (2013) 119–128.
- [8] P.O. de Giuseppe, T. de A.C.B. Souza, F.H.M. Souza, L.M. Zanphorlin, C.B. Machado, R.J. Ward, J.A. Jorge, R. dos P.M. Furriel, M.T. Murakami, Structural basis for glucose tolerance in GH1  $\beta$ -glucosidases, *Acta Crystallogr. D Biol. Crystallogr.* 70 (2014) 1631–1639.
- [9] L.P. Meleiro, J.C.S. Salgado, R.F. Maldonado, S. Carli, L.A.B. Moraes, R.J. Ward, J.A. Jorge, R.P.M. Furriel, Engineering the GH1  $\beta$ -glucosidase from *Humicola insolens*: insights on the stimulation of activity by glucose and xylose, *PLoS One* 12 (2017) e0188254.
- [10] C.A. Santos, L.M. Zanphorlin, A. Crucello, C.C.C. Tonoli, R. Ruller, M.A.C. Horta, M.T. Murakami, A.P. de Souza, Crystal structure and biochemical characterization of the recombinant ThBgl, a GH1  $\beta$ -glucosidase overexpressed in *Trichoderma harzianum* under biomass degradation conditions, *Biotechnol. Biofuels* 9 (2016) 71.
- [11] D. Toyama, M.A.B. de Moraes, F.C. Ramos, L.M. Zanphorlin, C.C.C. Tonoli, A. F. Balula, F.P. de Miranda, V.M. Almeida, S.R. Marana, R. Ruller, M.T. Murakami, F. Henrique-Silva, A novel  $\beta$ -glucosidase isolated from the microbial metagenome of Lake Poraquê (Amazon, Brazil), *Biochim. Biophys. Acta Protein Proteomics* 2018 (1866) 569–579.
- [12] C.A. Santos, M.A.B. Moraes, O.M. Terrett, J.J. Lyczakowski, L.M. Zanphorlin, J. A. Ferreira-Filho, C.C.C. Tonoli, M.T. Murakami, P. Dupree, A.P. Souza, An engineered GH1  $\beta$ -glucosidase displays enhanced glucose tolerance and increased sugar release from lignocellulosic materials, *Sci. Rep.* 9 (2019) 4903.
- [13] J.L. Salgado, L.R. Meleiro, S. Carli, R.J. Ward, Glucose tolerant and glucose stimulated  $\beta$ -glucosidases – a review, *Bioresour. Technol.* 267 (2018) 704–713.
- [14] Y. Yang, X. Zhang, Q. Yin, W. Fang, Z. Fang, X. Wang, X. Zhang, Y. Xiao, A mechanism of glucose tolerance and stimulation of GH1  $\beta$ -glucosidases, *Sci. Rep.* 5 (2015) 17296.
- [15] C. Bohlin, E. Praestgaard, M.J. Baumann, K. Borch, J. Praestgaard, R.N. Monrad, P. Westh, A comparative study of hydrolysis and transglycosylation activities of fungal  $\beta$ -glucosidases, *Appl. Microbiol. Biotechnol.* 97 (2013) 159–169.
- [16] F.H.M. Souza, C.V. Nascimento, J.C. Rosa, D.C. Masui, F.A. Leone, J.A. Jorge, R.P. M. Furriel, Purification and biochemical characterization of a mycelial glucose- and xylose-stimulated  $\beta$ -glucosidase from the thermophilic fungus *Humicola insolens*, *Process Biochem.* 45 (2010) 272–278.
- [17] F.H.M. Souza, L.P. Meleiro, C.B. Machado, A.L.R.L. Zimbardi, R.F. Maldonado, T.A. C.B. Souza, D.C. Masui, M.T. Murakami, J.A. Jorge, R.J. Ward, R.P.M. Furriel, Gene cloning, expression and biochemical characterization of a glucose- and xylose-stimulated  $\beta$ -glucosidase from *Humicola insolens* RP86, *J. Mol. Catal. B Enzym.* 106 (2014) 1–10.
- [18] T.L.R. Corrêa, J.P.L. Franco Cairo, J. Cota, A. Damasio, L.C. Oliveira, F.M. Squina, A novel mechanism of  $\beta$ -glucosidase stimulation through a monosaccharide binding-induced conformational change, *Int. J. Biol. Macromol.* 166 (2021) 1188–1196.
- [19] S. Kuusk, P. Våljamäe, When substrate inhibits and inhibitor activates: implications of  $\beta$ -glucosidases, *Biotechnol. Biofuels* 10 (2017) 7.
- [20] L.S.C. Costa, D.C.B. Mariano, R.E.O. Rocha, J. Kraml, C.H. da Silveira, K.R. Liedl, R. C. de Melo-Minardi, L.H.F. de Lima, Molecular dynamics gives new insights into the glucose tolerance and inhibition mechanisms on  $\beta$ -glucosidases, *Molecules* 24 (2019) 3215.
- [21] S. Konar, S.K. Sinha, S. Datta, P.K. Ghorai, Probing the dynamics between the substrate and the product towards glucose tolerance of *Halothermothrix orenii*  $\beta$ -glucosidase, *J. Biomol. Struct. Dyn.* 39 (2021) 5438–5448.
- [22] S.K. Sinha, S. Das, S. Konar, P.K. Ghorai, R. Das, S. Datta, Elucidating the regulation of glucose tolerance in a  $\beta$ -glucosidase from *Halothermothrix orenii* by active site pocket engineering and computational analysis, *Int. J. Biol. Macromol.* 156 (2020) 621–632.
- [23] M.-C. Bellissent-Funel, A. Hassanali, M. Havenith, R. Henchman, P. Pohl, F. Sterpone, D. van der Spoel, Y. Xu, A.E. Garcia, Water determines the structure and dynamics of proteins, *Chem. Rev.* 116 (2016) 7673–7697.
- [24] E. Brini, C.J. Fennell, M. Fernandez-Serra, B. Hribar-Lee, M. Lukšič, K.A. Dill, How water's properties are encoded in its molecular structure and energies, *Chem. Rev.* 117 (2017) 12385–12414.

- [25] G.S. Sharma, S. Krishna, S. Khan, T.A. Dar, K.A. Khan, L.R. Singh, Protecting thermodynamic stability of protein: the basic paradigm against stress and unfolded protein response by osmolytes, *Int. J. Biol. Macromol.* 177 (2021) 229–240.
- [26] B. Adamczak, M. Wieczór, M. Kogut, J. Stangret, J. Czub, Molecular basis of the osmolyte effect on protein stability: a lesson from the mechanical unfolding of lysozyme, *Biochem. J.* 473 (2016) 3705–3724.
- [27] M. Levy-Sakin, O. Berger, N. Feibish, N. Sharon, L. Schnaider, G. Shmul, Y. Amir, L. Buzhansky, E. Gazit, The influence of chemical chaperones on enzymatic activity under thermal and chemical stresses: common features and variation among diverse chemical families, *PLoS One* 9 (2014) e88541.
- [28] T. Arakawa, S.N. Timasheff, The stabilization of proteins by osmolytes, *Biophys. J.* 47 (1985) 411–414.
- [29] D.W. Bolen, Protein stabilization by naturally occurring osmolytes, *Methods Mol. Biol.* 168 (2001) 17–36.
- [30] K.S. Negi, N. Das, T. Khan, P. Sen, Osmolyte induced protein stabilization: modulation of associated water dynamics might be a key factor, *Phys. Chem. Chem. Phys.* 25 (2023) 32602–32612.
- [31] S. Shimizu, C.L. Boon, The Kirkwood-buff theory and the effect of cosolvents on biochemical reactions, *J. Chem. Phys.* 121 (2004) 9147–9155.
- [32] S.N. Timasheff, Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components, *Proc. Natl. Acad. Sci. USA* 99 (2002) 9721–9726.
- [33] H. Cui, L. Zhang, L. Eltrouky, Q. Jiang, S.K. Korkunç, K.-E. Jaeger, U. Schwaneberg, M.D. Davari, Enzyme hydration determines resistance in organic Cosolvents, *ACS Catal.* 10 (2020) 14847–14856.
- [34] I. Kuroiwa, Y. Maki, K. Matsuo, M. Annaka, Protein preferential solvation in (sucralose + water) mixtures, *J. Phys. Chem. B* 128 (2024) 676–683.
- [35] S. Shimizu, N. Matubayasi, Preferential solvation: dividing surface vs excess numbers, *J. Phys. Chem. B* 118 (2014) 3922–3930.
- [36] B. Manna, A. Ghosh, Molecular insight into glucose-induced conformational change to investigate uncompetitive inhibition of GH1  $\beta$ -glucosidase, *ACS Sustain. Chem. Eng.* 9 (2021) 1613–1624.
- [37] D.C.B. Mariano, C. Leite, L.H.S. Santos, L.F. Marins, K.S. Machado, A.V. Werhli, L.H.F. Lima, R.C. de Melo-Minardi, Characterization of glucose-tolerant  $\beta$ -glucosidases used in biofuel production under the bioinformatics perspective: a systematic review, *Genet. Mol. Res.* 16 (2017) 10–4238.
- [38] L.H.F. de Lima, M.L. Fernandez-Quintero, R.E.O. Rocha, D.C.B. Mariano, R.C. de Melo-Minardi, K.R. Liedl, Conformational flexibility correlates with glucose tolerance for point mutations in  $\beta$ -glucosidases - a computational study, *J. Biomol. Struct. Dyn.* 39 (2021) 1621–1634.
- [39] R.E.O. Rocha, D.C.B. Mariano, T.S. Almeida, L.S. CorrêaCosta, P.H.C. Fischer, L.H. Santos, E.R. Caffarena, C.H. da Silveira, L.M. Lamp, M.L. Fernandez-Quintero, K.R. Liedl, R.C. de Melo-Minardi, L.H.F. de Lima, Thermostabilizing mechanisms of canonical single amino acid substitutions at a GH1  $\beta$ -glucosidase probed by multiple MD and computational approaches, *Proteins* 91 (2023) 218–236.
- [40] R. Radhakrishnan, B. Manna, A. Ghosh, Solvent induced conformational changes for the altered activity of laccase: a molecular dynamics study, *J. Hazard. Mater.* 423 (2022) 127123.
- [41] S. Goswami, B. Manna, K. Chattopadhyay, A. Ghosh, S. Datta, Role of conformational change and glucose binding sites in the enhanced glucose tolerance of agrobacterium tumefaciens 5A GH1  $\beta$ -glucosidase mutants, *J. Phys. Chem. B* 125 (2021) 9402–9416.
- [42] R. Gazi, S. Maity, M. Jana, Conformational features and hydration dynamics of proteins in cosolvents: a perspective from computational approaches, *ACS Omega* 8 (2023) 2832–2843.
- [43] N. Chéron, M. Naepels, E. Pluhařová, D. Laage, Protein preferential solvation in water: Glycerol mixtures, *J. Phys. Chem. B* 124 (2020) 1424–1437.
- [44] L. Martínez, S. Shimizu, Molecular interpretation of preferential interactions in protein solvation: a solvent-shell perspective by means of minimum-distance distribution functions, *J. Chem. Theory Comput.* 13 (2017) 6358–6372.
- [45] L. Martínez, ComplexMixtures.Jl: investigating the structure of solutions of complex-shaped molecules from a solvent-shell perspective, *J. Mol. Liq.* 347 (2022) 117945.
- [46] A.F. Pereira, V. Piccoli, L. Martínez, Trifluoroethanol direct interactions with protein backbones destabilize  $\alpha$ -helices, *J. Mol. Liq.* 365 (2022) 120209.
- [47] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B.L. de Groot, H. Grubmüller, A.D. MacKerell Jr., CHARMM36m: an improved force field for folded and intrinsically disordered proteins, *Nat. Methods* 14 (2017) 71–73.
- [48] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79 (1983) 926–935.
- [49] T. Cloutier, C. Sudrik, H.A. Sathish, B.L. Trout, Kirkwood-buff-derived alcohol parameters for aqueous carbohydrates and their application to preferential interaction coefficient calculations of proteins, *J. Phys. Chem. B* 122 (2018) 9350–9360.
- [50] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, *Comput. Phys. Commun.* 91 (1995) 43–56.
- [51] M. Parrinello, A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, *J. Appl. Phys.* 52 (1981) 7182–7190.
- [52] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems, *J. Chem. Phys.* 98 (1993) 10089–10092.
- [53] B. Hess, H. Bekker, H.J.C. Berendsen, J.G.E.M. Fraaije, LINCS: a linear constraint solver for molecular simulations, *J. Comput. Chem.* 18 (1997) 1463–1472.
- [54] W.F. Van Gunsteren, H.J.C. Berendsen, A leap-frog algorithm for stochastic dynamics, *Mol. Simul.* 1 (1988) 173–185.
- [55] P.B.P.S. Reis, D. Vila-Viçosa, W. Rocchia, M. Machuqueiro, PypKa: a flexible Python module for Poisson-Boltzmann-based pKa calculations, *J. Chem. Inf. Model.* 60 (2020) 4442–4448.
- [56] J.M. Martínez, L. Martínez, Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking, *J. Comput. Chem.* 24 (2003) 819–825.
- [57] L. Martínez, R. Andrade, E.G. Birgin, J.M. Martínez, PACKMOL: a package for building initial configurations for molecular dynamics simulations, *J. Comput. Chem.* 30 (2009) 2157–2164.
- [58] J. Bezanson, A. Edelman, S. Karpinski, V.B. Shah, Julia: a fresh approach to numerical computing, *SIAM Rev.* 59 (2017) 65–98.
- [59] C.A. Cerdeirina, E. Carballo, C.A. Tovar, L. Romaní, Thermodynamic properties of aqueous carbohydrate solutions, *J. Chem. Eng. Data* 42 (1997) 124–127.
- [60] A.F. Fucaloro, Y. Pu, K. Cha, A. Williams, K. Conrad, Partial molar volumes and refractions of aqueous solutions of fructose, glucose, mannose, and sucrose at 15.00, 20.00, and 25.00 °C, *J. Solut. Chem.* 36 (2007) 61–80.
- [61] L. Martínez, Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis, *PLoS One* 10 (2015) e0119264.
- [62] L. Martínez, R. Andreani, J.M. Martínez, Convergent algorithms for protein structural alignment, *BMC Bioinform.* 8 (2007) 306.
- [63] R. Andreani, J.M. Martínez, L. Martínez, F.S. Yano, Low order-value optimization and applications, *J. Glob. Optim.* 43 (2008) 1–22.
- [64] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph.* 14 (33–8) (1996) 27–28.
- [65] D. Frishman, P. Argos, Knowledge-based protein secondary structure assignment, *Proteins* 23 (1995) 566–579.
- [66] W. Kabsch, C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* 22 (1983) 2577–2637.
- [67] L. Martínez, CellListMap.jl: Efficient and customizable cell list implementation for calculation of pairwise particle properties within a cutoff, *Comput. Phys. Commun.* 279 (2022) 108452.
- [68] S. Shimizu, N. Matubayasi, Preferential hydration of proteins: a Kirkwood-buff approach, *Chem. Phys. Lett.* 420 (2006) 518–522.
- [69] V. Piccoli, L. Martínez, Correlated counterion effects on the solvation of proteins by ionic liquids, *J. Mol. Liq.* 320 (2020) 114347.
- [70] I.P. de Oliveira, L. Martínez, The shift in urea orientation at protein surfaces at low pH is compatible with a direct mechanism of protein denaturation, *Phys. Chem. Chem. Phys.* 22 (2019) 354–367.
- [71] D. Harries, J. Rösgen, A practical guide on how osmolytes modulate macromolecular properties, *Methods Cell Biol.* 84 (2008) 679–735.
- [72] P.E. Smith, Equilibrium dialysis data and the relationships between preferential interaction parameters for biological systems in terms of Kirkwood-buff integrals, *J. Phys. Chem. B* 110 (2006) 2862–2868.
- [73] E.S. Courtenay, M.W. Capp, C.F. Anderson, M.T. Record Jr., Vapor pressure osmometry studies of osmolyte-protein interactions: implications for the action of osmoprotectants in vivo and for the interpretation of “osmotic stress” experiments in vitro, *Biochemistry* 39 (2000) 4455–4471.
- [74] M.T. Record Jr., C.F. Anderson, Interpretation of preferential interaction coefficients of nonelectrolytes and of electrolyte ions in terms of a two-domain model, *Biophys. J.* 68 (1995) 786–794.
- [75] I.L. Shulgin, E. Ruckenstein, A protein molecule in a mixed solvent: the preferential binding parameter via the Kirkwood-buff theory, *Biophys. J.* 90 (2006) 704–707.
- [76] I.L. Shulgin, E. Ruckenstein, A protein molecule in an aqueous mixed solvent: fluctuation theory outlook, *J. Chem. Phys.* 123 (2005) 054909.
- [77] D.C. Rapaport, Hydrogen bonds in water, *Mol. Phys.* 50 (1983) 1151–1162.
- [78] A.E. Garcia, L. Stiller, Computation of the Mean Residence Time of Water in the Hydration Shells of Biomolecules, No. 11 (1993) 1396–1406.
- [79] V.A. Makarov, B. Kim Andrews, P.E. Smith, B. Montgomery Pettitt, Residence times of water molecules in the hydration sites of myoglobin, *Biophys. J.* 79 (2000) 2966–2974.
- [80] B. Laurent, M. Chavent, T. Cragnolini, A.C.E. Dahl, S. Pasquali, P. Derreumaux, M. S.P. Sansom, M. Baaden, Epock: rapid analysis of protein pocket dynamics, *Bioinformatics* 31 (2015) 1478–1480.
- [81] V. Pierce, M. Kang, M. Aburi, S. Weerasinghe, P.E. Smith, Recent applications of Kirkwood-buff theory to biological systems, *Cell Biochem. Biophys.* 50 (2008) 1–22.
- [82] S. Shimizu, D.J. Smith, Preferential hydration and the exclusion of cosolvents from protein surfaces, *J. Chem. Phys.* 121 (2004) 1148–1154.
- [83] S. Christ, D. Schwabeneder, C. Rackauckas, M.K. Borregaard, T. Breloff, Plots.Jl – A user extendable plotting API for the Julia programming language, *J. Open Res. Softw.* 11 (2023), <https://doi.org/10.5334/jors.431>.
- [84] M. Mukherjee, J. Mondal, Heterogeneous impacts of protein-stabilizing Osmolytes on hydrophobic interaction, *J. Phys. Chem. B* 122 (2018) 6922–6930.
- [85] S. Mojtabavi, N. Samadi, M.A. Faramarzi, Osmolyte-induced folding and stability of proteins: concepts and characterization, *Iran, J. Pharm. Res.* 18 (2019) 13–30.