
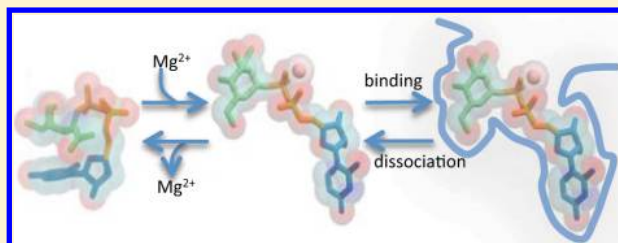


# Modulation of a Ligand's Energy Landscape and Kinetics by the Chemical Environment

Martin Held,<sup>†</sup> Petra Imhof,<sup>‡</sup> Bettina G. Keller,<sup>†</sup> and Frank Noé<sup>\*,†</sup><sup>†</sup>Institute of Mathematics, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany<sup>‡</sup>Institute of Physics, Freie Universität Berlin, Arnimallee 14, 14195 Berlin, Germany Supporting Information

**ABSTRACT:** Understanding how the chemical environment modulates the predominant conformations and kinetics of flexible molecules is a core interest of biochemistry and a prerequisite for the rational design of synthetic catalysts. This study combines molecular dynamics simulation and Markov state models (MSMs) to a systematic computational strategy for investigating the effect of the chemical environment of a molecule on its conformations and kinetics. MSMs allow quantities to be computed that are otherwise difficult to access, such as the metastable sets, their free energies, and the relaxation time scales related to the rare transitions between metastable states.

Additionally, MSMs are useful to identify observables that may act as sensors for the conformational or binding state of the molecule, thus guiding the design of experiments. In the present study, the conformation dynamics of UDP-GlcNAc are studied in vacuum, water, water +  $\text{Mg}^{2+}$ , and in the protein UDP-GlcNAc 2-epimerase. It is found that addition of  $\text{Mg}^{2+}$  significantly affects the conformational stability, thermodynamics, and kinetics of UDP-GlcNAc. In particular, the slowest structural process, puckering of the GlcNAc sugar, depends on the overall conformation of UDP-GlcNAc and may thus act as a sensor of whether  $\text{Mg}^{2+}$  is bound or not. Interestingly, transferring the molecule from vacuum to water makes the protein-binding conformations of UDP-GlcNAc first accessible, while adding  $\text{Mg}^{2+}$  further stabilizes them by specifically associating to binding-competent conformations. While  $\text{Mg}^{2+}$  is not cocrystallized in the UDP-GlcNAc 2-epimerase complex, the selectively stabilized  $\text{Mg}^{2+}$ /UDP-GlcNAc complex may be a template for the bound state, and  $\text{Mg}^{2+}$  may accompany the binding-competent ligand conformation to the binding pocket. This serves as a possible explanation of the enhanced epimerization rate in the presence of  $\text{Mg}^{2+}$ . This role of  $\text{Mg}^{2+}$  has previously not been described and opens the question whether “binding co-factors” may be a concept of general relevance for protein–ligand binding.



## INTRODUCTION

The chemical or biological roles of molecules depend on the conformations they can access and on the transition rates, that is, the kinetics between these conformations. The conformations and kinetics of biomolecules are strongly affected by their chemical environment, such as the type of solvent or the presence of ions. For example, it has been shown that (i) the stability of proteins is affected by small changes in the polarity of the solvent,<sup>1</sup> (ii) the folding kinetics of loop-forming peptides depend on the viscosity of the solvent,<sup>2</sup> and (iii) the  $\text{Mg}^{2+}$  concentration affects the folding dynamics,<sup>3,4</sup> as well as the conformation of the binding pocket<sup>5</sup> in ribozymes.

Understanding which effect the chemical environment has on the conformations and kinetics of biomolecules is one of the core interests of biochemistry and physical chemistry. Being able to model this relationship is crucial to comprehend how biochemical steps are driven in a cell. Here, we construct a computational strategy toward this end. In particular, we are interested in understanding how a substrate that specifically binds to an enzyme can be stabilized in its binding conformation. From a chemical point of view, such an

understanding may be useful to mimic efficient biological catalysis by designing a synthetic scaffold.

The conformations and kinetics of a molecule may be understood to arise from its energy landscape. Such an energy landscape may be “sculpted” by the chemical environment, thus, changing both the accessible conformations and the kinetics. These changes also affect experimental observables, such as intensities and relaxation time scales of fluorescence signals or scattering functions, and can thus be monitored with stationary and kinetic ensemble experiments.<sup>6–12</sup> Single-molecule experiments may even monitor changes of the transition rates between conformations, provided these conformations are distinguishable in the experimental observables employed.<sup>3,4,13–16</sup> While such an experimental indication is crucial, it is, however, indirect as it only traces changes in a particular experimental observable that may not be sensitive to all kinetic processes of the system. Hence, complementary computer simulation studies are needed in which chemical

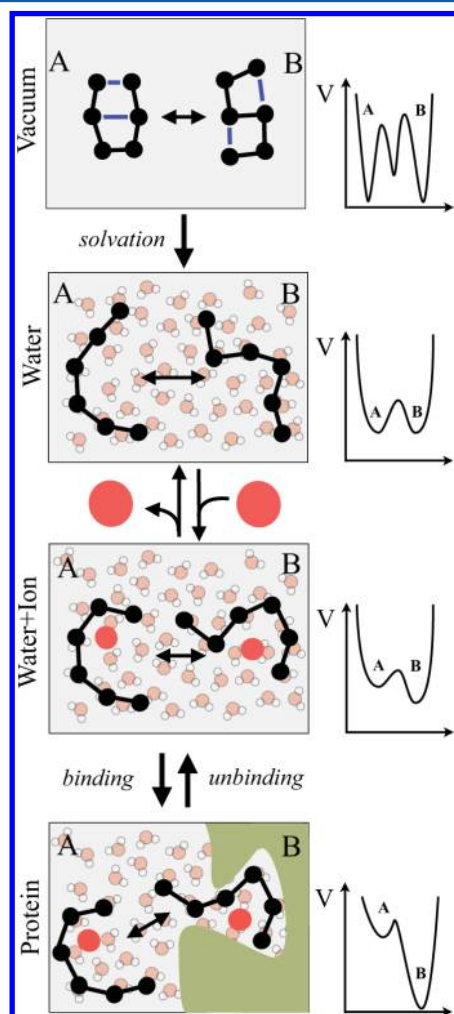
Received: January 19, 2012

Revised: September 28, 2012

Published: October 1, 2012

environmental effects on conformations and kinetics can be captured in a microscopic model. Such a microscopic model may prove useful to both understand the environmental effects onto conformational stability and kinetics in detail and also to guide experimental investigation by predicting interesting observables.

Consider the hypothetical scenario given in Figure 1 for an illustration of how the chemical environment may sculpt a free



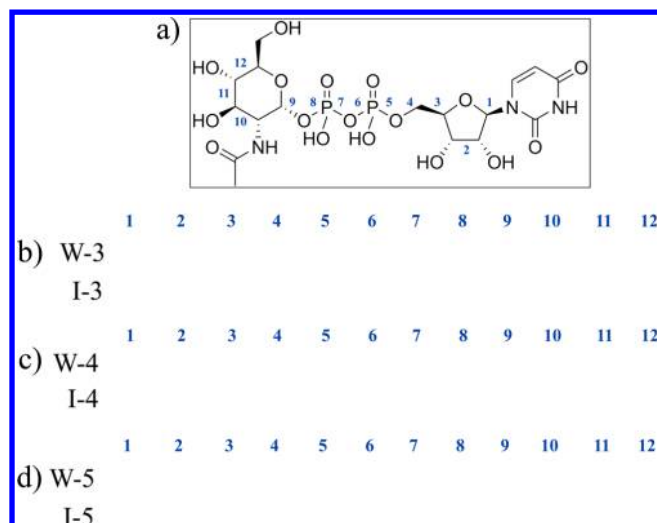
**Figure 1.** Scheme of different chemical environments and their influence on structural and energetic aspects of a hypothetical ligand molecule that can here attain two conformations A and B. The left-hand side shows a pathway that describes the solvation, cofactor association, and protein binding of the ligand molecule. Correspondingly, the right-hand side illustrates the hypothetical energy landscapes of the ligand molecule in the different environments. In vacuum the energy landscape is rough with a high barrier between the two conformations. When the molecule is solvated the landscape is smoothed and the barrier is reduced. The subsequent binding of a cofactor stabilizes conformation B with respect to A, resulting in a deeper energy well for conformation B. In the last step, the binding to a protein further stabilizes conformation B by specific ligand–protein interactions.

energy landscape. Here, a ligand molecule is depicted in four different chemical environments: vacuum, solvent, and solvent plus an ion, as well as bound to a macromolecule. In vacuum, strong electrostatic intramolecular interactions may stabilize compact ligand conformations.<sup>17,18</sup> In the associated free

energy landscape, this is reflected by well-defined minima and high energetic barriers between the states. Adding a polar solvent such as water effectively weakens electrostatic interactions by creating a reaction field and the availability of alternative hydrogen bonding partners. This effect increases the conformation space accessible by the ligand and lowering the free energy barriers between the states. The addition of some ion may stabilize specific conformations. In this way, the ion may be involved in conformational selection of binding competent ligand conformation prior to the specific binding of the ligand to a protein. The protein environment may further stabilize this conformation, here represented by a deeper energy well. Concepts such as binding by induced or selected fit<sup>19,20</sup> are usually discussed in terms of the influence of ligands on protein conformations. However, biologically relevant ligands can potentially form specific interactions with ions present in the cell (charged ligands, aromatic moieties, etc.). It is therefore quite likely that selection of ligand conformations by ions or small molecules does play a role in protein–ligand binding.

To arrive at a more thorough understanding of how the chemical environment sculpts the energy landscape of specific ligand molecules, thus affecting their conformations and kinetics, three components are needed: (i) a microscopic model of the molecule that allows changes of the chemical environment to be implemented, (ii) a way to map the interesting features of this model in a way that does not depend on subjective choices such as predefined reaction coordinates, and (iii) a way to link this model to experimental observables. As a microscopic model (i), we chose atomistic molecular dynamics simulations, which provide temporally resolved trajectories of all the atoms in a molecular system. Classical molecular dynamics force fields have been demonstrated to be able to match experimentally measurable quantities in many cases and are steadily improved.<sup>12,21–23</sup> However, given molecular dynamics data, extracting the relevant features of the system's energy landscape, such as metastable structures and dynamical processes can be a nontrivial task.<sup>24</sup> A rather objective analysis that largely avoids bias from user-defined reaction coordinates are Markov state models (MSMs). In MSMs, the molecular state space is clustered into many (e.g., several 1000) “microstates” of similar configurations, and the transition probabilities among these microstates are estimated from the MD trajectories.<sup>25–34</sup> See ref 34 for an overview of the state of the art in MSM methodology and refs 35 and 36 for software to construct MSMs. The MSM transition probability matrix can then be analyzed so as to yield the metastable states,<sup>25,30,37</sup> their thermodynamic properties,<sup>38</sup> the slowest relaxation processes,<sup>12</sup> transition pathways,<sup>39–41</sup> and so on. In particular, kinetic experimental observables and their uncertainties may often be computed in terms of the eigenvalues and eigenvectors of the transition matrix, thus, facilitating a comparison to experimental data and their interpretation.<sup>12,42–45</sup>

In this paper, a systematic MSM approach is presented to study how the energy landscape and kinetics of a small ligand molecule are affected by changes of its chemical environment. In particular we investigate uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc; Figure 2a), a key player in the sialic acid synthesis pathway whose products, sialic acids, are involved in a number of important biological processes, for example, cellular adhesion or glycoprotein stabilization.<sup>46</sup> UDP-GlcNAc is simulated in different chemical environments: vacuum, water, water with a  $Mg^{2+}$  ion, and in the binding pocket of an



**Figure 2.** (a) Structure of uridine diphosphate *N*-acetylglucosamine (UDP-GlcNAc) with dihedral annotation (see Table S1 for the definition of the dihedrals). (b–d) UDP-GlcNAc dihedral histograms of identified metastable sets that show a high microstate overlap. *W*-*i* indicates metastable set *i* of the water system, and *I*-*i* indicates the respective set for the water + Mg<sup>2+</sup> system. The range of each dihedral histogram is from  $-180$  to  $180^\circ$ , and the bin size is  $5^\circ$ .

epimerase protein where UDP-GlcNAc is specifically bound. MSMs are constructed for each of the setups to investigate the effects that changes in the chemical environment have on the ligand molecule's thermodynamics and kinetics. Given the generated MSMs we are able to access quantities that are otherwise difficult to compute, such as metastable conformations, experimentally measurable relaxation time scales as well as thermodynamic quantities such as free energies, internal energies, and entropies of the identified metastable conformations in each of the considered chemical environments. We find a number of unexpected effects of the chemical environment onto the conformations and kinetics of the UDP-GlcNAc molecule and identify a potential conformational selection mechanism by the interaction of UDP-GlcNAc with the divalent ion Mg<sup>2+</sup> that may be relevant for binding to the protein.

## METHODS

**System Preparation and Molecular Dynamics Simulations.** The parameters for UDP-GlcNAc were assembled from the Amber-Glycam04<sup>47,48</sup> parameters for the GlcNAc moiety and GAFF<sup>49</sup> based parameters for the UDP part. The charge at the UDP  $\beta$ -phosphate oxygen atom that binds to GlcNAc was adjusted so as to obtain a net charge of UDP-GlcNAc of  $-2$ . The corresponding Gromacs input files were generated using the Antechamber package<sup>50</sup> in conjunction with the amb2gmx<sup>51</sup> script. The protein structure coordinates for the UDP-GlcNAc 2-epimerase (*E. coli*) with bound UDP-GlcNAc were obtained from the Protein Data Bank (accession number 1VGV). Based on these coordinates the program pdb2gmx<sup>52</sup> was employed to generate simulation input files. The individual protein residue protonation was determined by automatic internal pdb2gmx procedures.

MD simulations of UDP-GlcNAc were carried out in vacuum (64 atoms), pure water (5902 atoms), pure water and one Mg<sup>2+</sup> ion (5900 atoms) and with the solvated protein system ( $\sim 47,000$  atoms). For the ion simulations the Mg<sup>2+</sup> ion was

positioned in the vicinity of the two phosphates by replacing one of the previously added water molecule that were the closest to the two phosphates. During the simulations the ion was not distance restrained. However, it was observed to reside close to the phosphate oxygens for the whole simulation time of  $4 \mu\text{s}$ . The following simulation protocol was applied for all four setups: The simulation program Gromacs 4.5.3,<sup>52</sup> the amber-99ff<sup>53</sup> force field, and TIP3P water model were used. All covalent bonds to hydrogens were constrained using the LINCS algorithm<sup>54</sup> (lincs\_order = 4, lincs\_warn\_angle = 30, lincs\_iter = 1), permitting an integration time-step of 2 fs. As integrator a Langevin integrator with a coupling constant of  $1 \text{ ps}^{-1}$  and a temperature of  $T = 300 \text{ K}$  was used (NVT ensemble). The volume of the watered simulation box was determined by running initial 1 ns NpT simulations (1 atm, Berendsen barostat, isotropic coupling). To handle electrostatic interactions PME<sup>55</sup> with a real space cutoff of 1.0 nm, interpolation order 4 and 1.2 nm grid spacing was applied. van der Waals interactions were cutoff at 1.0 nm. To study the effect of the absence of an ion in the pure water setup, no counterions were used. The nonzero net charge in the pure water setup is corrected by a virtual background charge in the Gromacs PME implementation. The trajectory data was stored every 5 ps. For each water and water + Mg<sup>2+</sup>,  $2 \times 2 \mu\text{s}$  were obtained; for vacuum,  $4 \times 2 \mu\text{s}$ ; and for the protein system, 90 ns, resulting in  $8 \times 10^5$ ,  $8 \times 10^5$ ,  $18 \times 10^3$ , and  $16 \times 10^5$  saved frames for each simulation.

**Markov Models: Construction, Analysis, and Validation. General Remarks.** To construct a Markov model, it is not necessary to have simulations that are longer than the longest relaxation time scale of the system. However, for the present system, these could be easily prepared and were preferred over doing many short simulations that would pose the additional difficulty of assuring that the starting points be drawn from a local equilibrium within the discrete states.<sup>34,39</sup> Nevertheless, the Markov model makes the statistical information of the simulation data more accessible than most direct trajectory analyses, as all transitions between conformational substates are taken into account, and the resulting transition matrix directly provides access to macroscopic information such as free energy differences or relaxation time scales.

Note that other approaches<sup>56</sup> use rate matrices rather than transition matrices. This is avoided here because a rate matrix is only defined for the slow processes between metastable states: The relationship between transition matrix  $T(\tau)$  and rate matrix  $K$ ,  $T(\tau) = \exp(\tau K)$ , is only uniquely invertible when the eigenvalues of  $T(\tau)$  are positive, and that is only true for the slow processes. Thus, a direct estimation of a rate matrix is not meaningful on a fine microstate discretization level. The fine discretization, on the other hand, is important to obtain a Markov model with high precision.<sup>57</sup> Therefore, the present Markov model was constructed using a transition matrix.

**State Space Discretization.** To capture the dynamics of UDP-GlcNAc in the different simulation environments Markov models were constructed using EMMA 1.2.<sup>36</sup> Many possible ways to generate a state space discretization exist, and any discretization that gives a good approximation of the slow dynamical processes is acceptable.<sup>34,57</sup> Here, a single state space discretization was defined for all chemical environments by clustering the trajectory data using regular space clustering with minimal RMSD metric using all 39 heavy atoms of UDP-GlcNAc and  $d = 0.15 \text{ nm}$  as threshold (EMMA command mm\_discretize). Regular space clustering iterates the simulation

data and adds a trajectory frame to the set of cluster centers whenever its distance to all existing cluster centers is greater than  $d$ . Subsequently, a Voronoi decomposition of state space is done by assigning each trajectory frame to the nearest cluster centers. This approach generates a clustering with an approximately uniform distribution of cluster centers across the visited part of state space. To obtain a single discretization, the clustering was performed on the unification of the trajectory data from all chemical environments. Ion and solvent molecules were not taken into account in the distance metric used for clustering.

**Markov State Model Estimation.** The all-atom trajectories of each environment were projected onto the discretized state space and from the resulting discretized trajectories an MSM was constructed for each environment (vacuum, water, water +  $\text{Mg}^{2+}$ ). The MSM lag time  $\tau = 2$  ns was identified by calculating the “implied” relaxation time scales<sup>26</sup> (mm\_time scales). Reversible transition matrices  $\mathbf{T}(\tau)$  were subsequently calculated for all of the three simulation environments (mm\_estimate).

**Time Scale Estimation.** When the relation  $t_i = \tau / (\ln \lambda_i)$  is used, the relaxation time scales,  $t_i$ , of the slowest processes were computed based on the eigenvalues  $\lambda_i$  of the individual MSMs (mm\_transitionmatrixAnalysis). These slowest processes were assigned to structural rearrangements by investigating the sign structure of the corresponding eigenvectors.<sup>34</sup>

**Metastable Sets.** Based on the estimated system time scales four slow relaxation processes could be identified for each of the chemical environments. The associated five metastable sets of states were extracted using the improved Perron cluster cluster analysis (PCCA;<sup>37</sup> mm\_pcca) separately on each simulation set.

**Chapman–Kolmogorov Tests.** To validate the estimated MSMs, it was tested whether the Chapman–Kolmogorov condition  $\mathbf{T}(n\tau) = \mathbf{T}^n(\tau)$  holds within statistical error. This was done using the procedure suggested in ref 34, where the system is assumed to start in each of the five metastable sets at time  $t = 0$ . The probability to stay in the starting metastable set is propagated to later times using the MSMs and compared to a distribution estimated directly from the trajectory data. Here this condition is tested for time ranges from 0 to 40 ns using the EMMA command mm\_chapman. The test results are presented in Supporting Information, Figure S1.

**Statistical Uncertainty Estimation.** To estimate the statistical uncertainties involved in state probabilities, energetics, and relaxation time scales, the transition matrix sampling algorithm proposed in ref 32 was employed as implemented by the EMMA command mm\_transitionmatrixSampling. As input, matrices of independent transition counts (using one count per lag time  $\tau$ ) were used in conjunction with a prior using 0 pseudocounts on diagonal matrix elements and on matrix elements where  $c_{ij} + c_{ji} > 0$ . The  $-1$  pseudocounts were added on all other matrix elements, effectively forcing the distribution in these elements to 0.<sup>34,39</sup>

**Energetics.** For each metastable set determined for the different chemical environments, the associated free energies, and their decompositions into internal energies and entropies were calculated. All energies are estimated up to an arbitrary additive constant, which renders the direct comparison of energies between different simulation environments impossible. However, energies can be compared between different metastable sets of the same simulation environment. The free energy of a metastable set  $k$  was calculated using the relation

$F_k = -k_B T \ln(\sum_{i \in S_k} \pi_i)$ , where  $k_B$  is the Boltzmann constant,  $T$  is the absolute temperature, and  $\sum_{i \in S_k} \pi_i$  denotes the sum of stationary weights over all states  $i$  in metastable set  $k$ . To evaluate the statistical uncertainty in  $F_k$ , the probability distribution of stationary distributions,  $\pi$ , was sampled using the reversible Monte Carlo sampling described in ref 32 (see above). For each sample, the corresponding estimate of  $F_k$  was calculated, and its statistical uncertainty is then given by the direct sampling estimate of the standard deviation of  $F_k$  from the Monte Carlo sampling.

The internal energy  $U_k$  of a metastable set  $k$  was computed as mean total potential energy of all simulation trajectory frames that are assigned to a metastable set. The statistical uncertainty of these values is calculated as standard error of the mean potential energy, using the number of assigned trajectory frames as sample size. For this calculation, it was validated that the potential energies of subsequently stored trajectory frames are nearly statistically independent in all chemical environments (the normalized autocorrelation function of the potential energy has dropped to 0.05 in subsequent frames).

The entropy  $S_k$  was computed using the relation  $F_k = U_k - TS_k$  using  $F_k$  and  $U_k$  as given above. The statistical uncertainty  $\text{SE}(S_k)$  is simply computed from the standard errors in  $U_k$  and  $F_k$  via  $\text{SE}(S_k) = (\text{SE}(F_k)^2 + \text{SE}(U_k)^2)^{1/2}$ .

## RESULTS AND DISCUSSION

### Structures and Metastable Sets. Chemical Environment Modulates the Size of the Accessible Conformation Space.

The conformation space was discretized using regular space RMSD clustering producing “microstates” of approximately equal diameter. Thus the number of visited microstates gives a rough indication of the size of conformation spaces and of the conformational flexibility of UDP-GlcNAc in the different chemical environments (see Table 1). The smallest conforma-

**Table 1. Available Simulation Data, Determined Number of Microstates and the Stationary Percent Fraction of the Dominant GlcNAc Pucker  ${}^4C_1$  (as Defined by the Metastable States That Are in the  ${}^4C_1$  State: See Dihedrals 10–12 in Figure S2)<sup>a</sup>**

system	total simulation time ( $\mu\text{s}$ )	No. of microstates	$p({}^4C_1)$ (%)
vacuum	8 ( $2 \times 4 \mu\text{s}$ )	40	98 ( $\pm 1.3\%$ )
water	4 ( $2 \times 2 \mu\text{s}$ )	2281	76.5 ( $\pm 1.1\%$ )
water + $\text{Mg}^{2+}$	4 ( $2 \times 2 \mu\text{s}$ )	1544	96.2 ( $\pm 1.5\%$ )

<sup>a</sup>All quantities are reported for the water, water +  $\text{Mg}^{2+}$ , and vacuum environments.

tional freedom of UDP-GlcNAc is found in vacuo and would likely also be found in nonpolar solvents. In vacuo, unshielded intramolecular electrostatic interactions result in a strong conformational confinement that manifests in only 40 populated microstates. In pure water these electrostatic interactions are effectively reduced by the reaction field created in the polar solvent, or, structurally speaking, by the availability of water molecules as alternative hydrogen bonding partners. This effect increases the conformational freedom to 2281 populated microstates. In the water +  $\text{Mg}^{2+}$  environment this conformational flexibility is reduced to 1544 populated microstates. Although this number cannot be transformed into a configurational entropy, it supports the visual impression

and the intuition that the ion restricts the access to some conformations compared to pure water solvent.

*In All Chemical Environments, Five Metastable Sets Exist on Time Scales of 4 ns or Slower.* A metastable set is characterized by fast kinetics within the state, that is, conformations within that state interconvert quickly, while transitions between different metastable sets are rare, thus, giving rise to slow kinetics. To facilitate a comparison of metastable sets in different chemical environment, the conformation spaces of UDP-GlcNAc in vacuum, water, and water + Mg<sup>2+</sup> were decomposed into five metastable sets. This selection corresponds to investigating the kinetics occurring on time scales of a few ns to  $\mu$ s (see Figure 5). The number of microstates pertaining to each metastable set quantifies the sizes of these sets and their relative probability (see Tables 2

**Table 2. Number of Microstates for Each Identified Metastable Set in Vacuum, Water, and Water + Mg<sup>2+</sup>**

	PCCA set				
	1	2	3	4	5
vacuum	3	7	3	11	16
water	158	78	933	388	724
water + Mg <sup>2+</sup>	148	15	554	201	626

and 3). Figure 3 illustrates the structures found in each metastable set. In vacuum, only two of the identified five metastable sets have populations above 10%. This changes to three of five metastable sets when the ligand is immersed in water. Addition of the Mg<sup>2+</sup> ion alters this scenario again resulting in mainly two populated metastable sets. Note how the flexibility (in terms of the number of accessible microstates) of each metastable set changes with addition of Mg<sup>2+</sup>: While there is almost no change in set 5, considering the overall difference in populated microstates, the flexibility of sets 3 and 4 are significantly reduced.

*Different GlcNAc Sugar Puckers Are in Different Metastable Sets.* The GlcNAc sugar is observed in two different pucker conformations: the more stable <sup>4</sup>C<sub>1</sub> chair and the less stable <sup>1</sup>C<sub>4</sub> chair. These puckers are located in different metastable sets, indicating that puckering is a slow process. The pucker is illustrated in Figure 3 and manifests in the distribution of UDP-GlcNAc dihedral angles 10–12 shown in Figure 2 and SI, Figure S2. In vacuum, the <sup>4</sup>C<sub>1</sub> chair is found in sets 4 and 5, which are clearly predominant with 98% ( $\pm 0.43\%$ ) of the population (Table 1). The water + Mg<sup>2+</sup> environment also has a predominance of <sup>4</sup>C<sub>1</sub> with 96% ( $\pm 1.5\%$ ; states 2, 3, and 5). These results are consistent with a previous study combining molecular dynamics and NMR on GlcNAc (without UDP) in aqueous solvent which predicted a 99.6% probability of <sup>4</sup>C<sub>1</sub>.<sup>58</sup> However, we find a very different fraction when UDP-GlcNAc is solvated in pure water, where the <sup>4</sup>C<sub>1</sub> probability (states 3 and 5) drops to 77% ( $\pm 1.1\%$ ). This observation is due

to compact structures found in the water environment in which GlcNAc and UDP form interactions that stabilize the otherwise unstable <sup>1</sup>C<sub>4</sub> chair, yielding 23% probability in states 1, 2, and 4 (see Figure 3e). When Mg<sup>2+</sup> is added, the ion is coordinated by the oxygen atoms of the diphosphate group and the associated stretching of the diphosphate backbone separates the GlcNAc and UDP, preventing this interaction (see Figure 3e). Thus, the populated water + Mg<sup>2+</sup> structures are more extended compared to those populated in water.

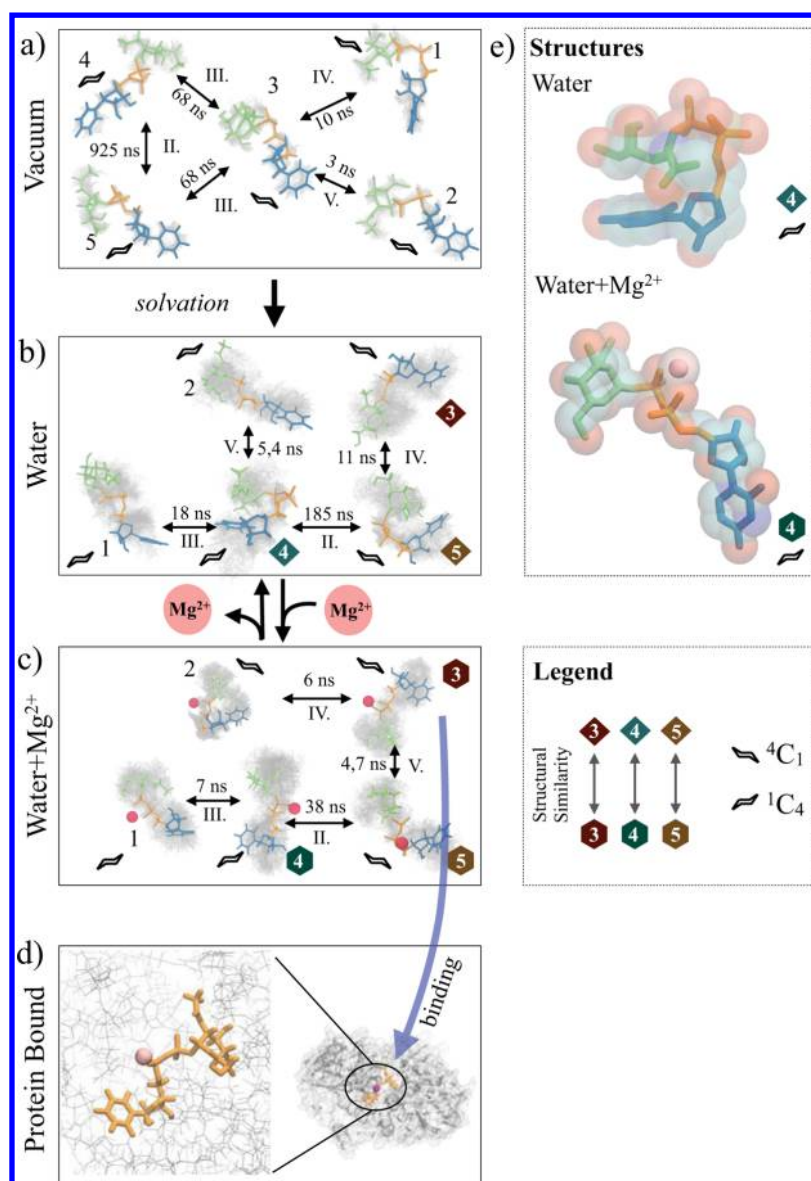
*Metastable Sets of UDP-GlcNAc in Water and Water + Mg<sup>2+</sup> Show Structural Similarities.* To facilitate comparison of metastable sets between chemical environments, the similarity of metastable sets of each chemical environment was evaluated in terms of the fraction of microstates common to both of the compared metastable sets (see SI, Figure S3). The metastable sets in vacuum do not have significant overlap with metastable sets of the solvated systems. However, water and water + Mg<sup>2+</sup> environments share similar UDP-GlcNAc structures, allowing metastable sets 3–5 to be roughly associated between these environments (see Figure 3 and SI, Figure S3). These metastable sets are the most probable ones in both environments (see Table 3). To get a more detailed impression of the structural similarity refer to Figure 2 where the dihedral histograms of the similar PCCA sets are compared. Besides the sugar pucker discussed above, the orientation of the Uracil ring is also the same in sets 3–5 of the water and the water + Mg<sup>2+</sup> environments. The structural difference between these environments mainly arises from the dihedrals of the phosphate link. In the pure water environment, various orientations are possible, while in the Mg<sup>2+</sup> setup, only straight orientations that coordinate the ion occur. This selective stabilization of UDP-GlcNAc conformations by Mg<sup>2+</sup> is effectively a conformational selection mechanism that will be of importance when studying the binding competence of UDP-GlcNAc (see below).

*Chemical Environment Changes the Flexibility within the Metastable Structures.* Representative structures of the five metastable sets with an indication of their flexibility are shown in Figure 3. The structures presented are randomly drawn representations of structures present in the respective metastable set, the gray cloud indicates the conformational flexibility found in the set. Note that the metastable sets found in vacuum are much more confined than the sets found for the solvated chemical environments. A more detailed impression of the structures present in each set can be obtained from the distribution of UDP-GlcNAc dihedral angles shown in Figure 2 and SI, Figure S2. Based on the dihedral histograms differences in conformational flexibility can be explained. The vacuum histograms show very well-defined structures by sharply peaked histograms. Solvation in water permits a high flexibility in the phosphate link (dihedrals 5–8). This flexibility is reduced by addition of a Mg<sup>2+</sup> ion that selects conformations that are suitable to coordinate the doubly positive charge, thus,

**Table 3. Stationary Probability Estimates for Each Identified Metastable Set in Vacuum, Water, and Water + Mg<sup>2+</sup>**<sup>a</sup>

	PCCA set				
	1	2	3	4	5
vacuum	0.002 ( $\pm 0.003$ )	0.002 ( $\pm 0.004$ )	0.01 ( $\pm 0.009$ )	0.61 ( $\pm 0.098$ )	0.37 ( $\pm 0.097$ )
water	0.032 ( $\pm 0.007$ )	0.005 ( $\pm 0.002$ )	0.197 ( $\pm 0.014$ )	0.195 ( $\pm 0.019$ )	0.569 ( $\pm 0.009$ )
water + Mg <sup>2+</sup>	0.015 ( $\pm 0.014$ )	0.005 ( $\pm 0.002$ )	0.192 ( $\pm 0.008$ )	0.023 ( $\pm 0.006$ )	0.765 ( $\pm 0.012$ )

<sup>a</sup>The error bars report the statistical uncertainty (68% confidence interval).



**Figure 3.** (a–d) Conformational dynamics of UDP-GlcNAc in different chemical environments. The structures presented are randomly drawn representations of structures present in the respective metastable set (1–5), the gray cloud indicates the conformational flexibility of that set. The pucker of the GlcNAc sugar is represented by a small icon. The arrows and time scales indicate the slowest relaxation processes of the system with the corresponding relaxation time scale calculated from the Markov model. Associated statistical uncertainties are given in Figure 5. Sets 3–5 in the water and water +  $\text{Mg}^{2+}$  environment are structurally similar and can thus be roughly associated with each other. Set 3 contains structures that are similar to protein-bound conformations. (e) representative structures from the predominant set of  ${}^1\text{C}_4$  conformations (metastable set 4) in water and 4 water +  $\text{Mg}^{2+}$ . In pure water, the  ${}^1\text{C}_4$  ring conformations is likely to be stabilized by interactions with the uracil ring.

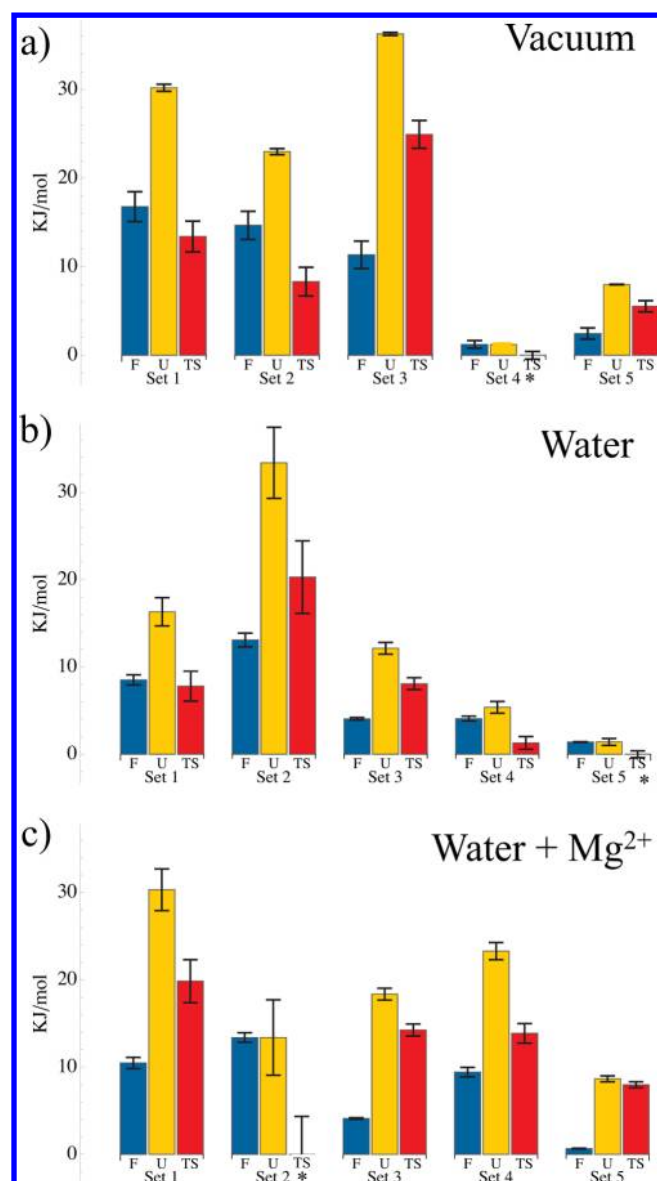
effectively “focusing” the metastable sets onto more well-defined subsets.

**Thermodynamics.** Having identified the metastable sets of structures in all chemical environments, we are now in a position to investigate their thermodynamic properties. To this end, free energies, internal energies, and entropies of the metastable sets were calculated in their chemical environments (see Figure 4). We focus on sets 3–5 in the solvated environments as 1 and 2 are only rarely populated due to their high free energies.

*Low Free Energy States of UDP-GlcNAc in Water Have Small Entropies Due to Electrostriction.* In pure water, sets 4 and 5 have the smallest free energies and also the smallest entropies. In contrast, set 3 also has a low free energy but a large internal energy that is compensated by an increased

entropy. The reason of the small entropies in sets 4 and 5 are likely due to electrostriction:<sup>59</sup> sets 4 and 5 have the charged phosphate oxygen atoms exposed, while in set 3 they are shielded by interactions with other solute atoms. Such a solute charge exposure has been shown to induce ordering in the surrounding solvent with an accompanying entropy decrease of the solute–solvent system.<sup>38</sup>

*$\text{Mg}^{2+}$  Binding Modulates the Energy Landscape and Reduces Electrostriction.* The conformational free energies of the most stable states 3–5 are more uniform in pure water than when  $\text{Mg}^{2+}$  is added (see Figure 4). Due to the high free energy of state 4 only a few realizations for this state were obtained which results in a relatively big error in the energy estimates. In contrast to the pure water environment, the state with the smallest free energy in the  $\text{Mg}^{2+}$  scenario is not the one having



**Figure 4.** Energy composition of each identified metastable set of the different chemical environments: (a) vacuum, (b) water, and (c) water +  $Mg^{2+}$ .  $F$  denotes the Helmholtz free energy,  $F_i = -kT \log p_i$  with probabilities  $p_i$  normalized,  $U$  is the internal energy, and  $TS$  is entropy multiplied by absolute temperature. Sets marked with \* are the reference set for the respective chemical environment. Within each environment a constant is subtracted such that  $TS$  of the \* marked set equals 0. The error bars report the statistical uncertainty (68% confidence interval).

the smallest entropy. It is possible that this can be attributed to the attached  $Mg^{2+}$  ion. It compensates the negative charges of the oxygen atoms, thus, presenting an effective net neutral charge to the environment. In a similar study of charge compensation,<sup>38</sup> it was shown that solvent molecules became less ordered when the charges were brought together, thus, reducing the electrostriction effect and increasing the entropy of the solute–solvent system.

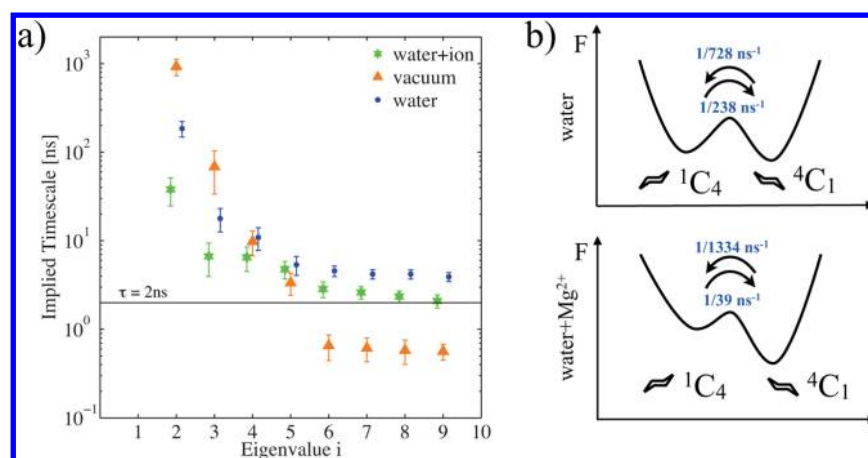
**Kinetics.** MSMs allow the slowest relaxation time scales of the system and the associated structural rearrangements to be computed in terms of the eigenvalues and eigenvectors of the MSM transition matrix.<sup>12,60</sup> This is of special interest from an experimental point of view as relaxation time scales can also be

probed via kinetic experiments providing an experimental observable is found that is able to pick up the structural changes indicated by the corresponding eigenvector.<sup>12,61</sup>

#### Ring Puckering and Isomerization of the Uracil Ring Conformation Are the Slowest Conformational Changes.

The structural rearrangements corresponding to the slowest relaxation time scales can be determined from the eigenvectors of the MSM transition matrices.<sup>34</sup> For the solvated environments (water and water +  $Mg^{2+}$ ), we find that the slowest process (second-largest eigenvalue) describes a transition between sets with the more stable  ${}^4C_1$  chair (sets 3 and 5) and sets with the less stable  ${}^1C_4$  chair (sets 1 and 4). Thus, the slowest process is found to be the GlcNAc sugar ring puckering. The second slowest process corresponds to isomerization of the Uracil ring. Refer to dihedrals 10–12 in SI, Figure S2 for an illustration of the ring puckering and to dihedral 1 in the same figure for the Uracil ring turning. This explains the structural diversity found in the metastable sets of the solvated systems: Isomerizations in the phosphate link induce large conformational changes but occur on time scales that are faster (ns to sub-ns) than the kinetics of the metastable sets.

**Chemical Environment Modulates Kinetics.** The slowest time scales for the three simulation setups (vacuum, water, and water +  $Mg^{2+}$ ) are shown in Figure 5. For the vacuum chemical environment the time scales are 1 order of magnitude slower than in the solvated environments. This indicates the presence of a rough energy landscape with high barriers that arise from unscreened electrostatics.<sup>17,18</sup> Once the solute is immersed in water, the polar solvent shields these strong interactions. This smooths the energy landscape and results not only in a larger conformation space but also in faster dynamics. Comparing the water and water +  $Mg^{2+}$  environments, the pure water environment gives rise to slower time scales. This finding is interesting, as one might expect the dynamics of the  $Mg^{2+}$  system to be slower due to conformational stabilization of the phosphate link by the  $Mg^{2+}$  ion. However, given the energetics (Figure 4) and the structural changes corresponding to the slowest kinetic process, that is, the process that switches between the pucker  ${}^4C_1$  and  ${}^1C_4$  (see Figure 3), this result can be understood in terms of a simple two-state rate theory argument. Note that this argument is qualitative, as the present system is not an ideal two-state system. Let  $A$  denote the more stable  ${}^4C_1$  chair and  $B$  denote the less stable  ${}^1C_4$  chair. The time scale  $\tau$  of a process that switches between the two states  $A$  and  $B$  is given by  $\tau = 1/(k_{AB} + k_{BA})$  with pucker rates  $k_{AB}$  and  $k_{BA}$ . In the present system, we find the maximum likelihood estimates of this slowest time scale to be  $\tau_{\text{water}} = 248$  ns for the pure water chemical environment and  $\tau_{\text{water}+Mg^{2+}} = 38$  ns for the water +  $Mg^{2+}$  environment. From detailed balance, we have the relation  $(k_{BA})/(k_{AB}) = \exp(-\Delta F_{AB}/(k_B T))$  with estimated free energy differences  $\Delta F_{AB} = 2.67$  kJ/mol for the water environment and  $\Delta F_{AB} = 8.75$  kJ/mol for the water +  $Mg^{2+}$  environment. The resulting estimated two-state pucker rates are for pure water ( $k_{AB} = 1/248$  s<sup>-1</sup> and  $k_{BA} = 1/728$  s<sup>-1</sup>) and for the water +  $Mg^{2+}$  environment ( $k_{AB} = 1/39$  s<sup>-1</sup> and  $k_{BA} = 1/1334$  s<sup>-1</sup>). Thus, the faster time scales in the water +  $Mg^{2+}$  environment are dominated by the increased transition rate from the less stable state  $\{4\}$  into the more stable state  $\{3,5\}$  and are thus a result of the remodeling of the energy landscape (see Figure 5b). Note that all time scales smaller than 2 ns are unreliable in the present Markov models that were parametrized at a lag time of 2 ns and are, hence, not further investigated.

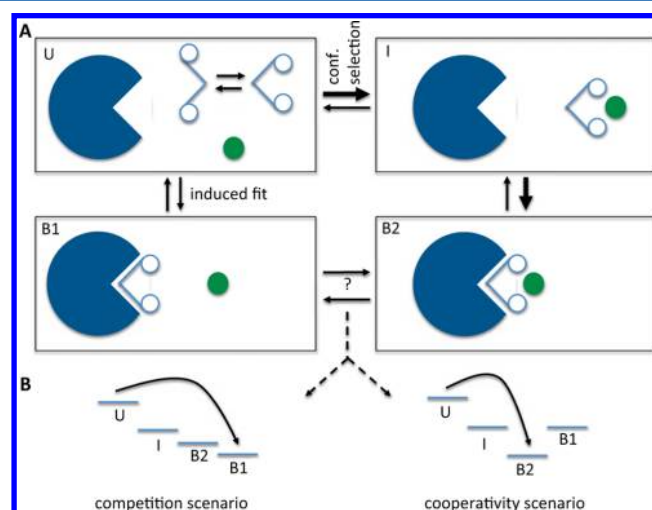


**Figure 5.** Kinetic properties of UDP-GlcNAc: (a) Implied time scales of Markov state models for vacuum, water, and water + Mg<sup>2+</sup>. The X-axis indicates the eigenvalue number, note that the first eigenvalue is 1 and, hence, has no associated time scale. The Y-axis shows the time scale associated to the respective eigenvalue ( $ITS_i = -(\tau/(\ln \lambda_i))$ ). The thick black line indicates the lag time  $\tau = 2$  ns at which the Markov state model was obtained. Time scales significantly below 2 ns are, thus, numerically unreliable. The error bars report the statistical uncertainty (68% confidence interval). (b) Schematic free energy landscapes of the GlcNAc puckering in the water and water + Mg<sup>2+</sup> setups.

**Slowest Time Scale Is a Sensor for Mg<sup>2+</sup> Binding.** As discussed above, the presence of Mg<sup>2+</sup> could be experimentally detected by measuring the population of <sup>4</sup>C<sub>1</sub> and <sup>1</sup>C<sub>4</sub>, which is significantly different with and without the Mg<sup>2+</sup> due to the different interactions between GlcNAc and UDP in the two scenarios. However, the results above indicate that also the kinetics can be used as a sensor for the interaction of UDP-GlcNAc with Mg<sup>2+</sup>: In pure water, the stabilization of the otherwise unstable <sup>1</sup>C<sub>4</sub> chair is predicted to increase the slowest relaxation time scale by a factor of 7. Kinetic experiments such as fluorescence correlation experiments of a fluorescent analog<sup>62</sup> or temperature-jump triggered time-resolved IR experiments are able to measure the slowest relaxation time scales of the molecule, provided an appropriate observable is available.<sup>63,64</sup> Thus, such kinetic experiments may be employed to measure the binding of UDP-GlcNAc to Mg<sup>2+</sup> via a change in the slowest relaxation time scale. It is likely such a sensor can also be constructed for other small molecules with ion binding sites.

**Selection of Protein-Binding States. Mg<sup>2+</sup> Binding Stabilizes Binding-Competent Structures.** Microstates containing structures that have an RMSD of less than 0.15 nm to the bound structure of UDP-GlcNAc in the UDP-GlcNAc-2-Epimerase protein were defined as binding-competent states. Binding-competent states comprise a <sup>4</sup>C<sub>1</sub> pucker in GlcNAc, a specific Uracil ring isomer, and a stretched conformation of the diphosphate backbone (see Figure 3d). In the vacuum environment, the binding-competent microstates have not been sampled at all. However, binding-competent structures are found in both pure water and water + Mg<sup>2+</sup>. Interestingly, their total population is higher in the water + Mg<sup>2+</sup> setup, indicating that an addition of the Mg<sup>2+</sup> cofactor brings the conformational ensemble of UDP-GlcNAc closer to the bound-like ensemble. In the context of protein–ligand binding, this can be interpreted as part of a conformational selection mechanism.<sup>19,65</sup> The energy landscape of the ligand is changed by addition of Mg<sup>2+</sup> such that the energy of the binding competent states is lowered, resulting in a higher population of these states (see state I in Figure 6).

**Binding Competent UDP-GlcNAc Structures are Found in a Single Metastable Set.** As discussed above, the Mg<sup>2+</sup>



**Figure 6.** Idealized scheme of ligand binding with conformational selection by an ion. A: Consider the conformational ensemble of protein, ligand and ion to be partitioned as follows: U: No interactions, the ligand exists in binding-competent and binding-incompetent conformations, I: The ion has associated to the ligand, stabilizing the binding-competent conformation, B2: Both ligand and ion bind to the protein, B1: The ligand binds to the protein, but the ion does not. In the present study, it is suggested that the binding mechanism would progress from U via I to B. B: The product state could be either B1 or B2, depending on the energetics. When ion and protein compete for binding the ligand, B1 is more stable than B2. When ion and protein bind the ligand cooperatively, B2 is the more stable state.

“focuses” metastable sets, making them narrower. This “focusing” is especially interesting when considering the binding-competent structures. Table 4 shows in which metastable sets the binding competent microstates (described above) are located. Interestingly, nearly all binding-competent conformations can be assigned to metastable set 3, in both the water and the water+Mg<sup>2+</sup> chemical environment. It is at first sight counterintuitive that the probability of metastable set 3 decreases when Mg<sup>2+</sup> is added, while the probability of binding-competent conformation increases. This is explained by the fact that the probability fraction of binding competent microstates



**Table 4. Number of Microstates in Each Metastable Set Where UDP-GlcNAc Is in a Binding Conformation<sup>a</sup>**

	1	2	3	4	5
vacuum	0	0	0	0	0
water	1 (0.005)	0	37 (0.056)	0	3 (0.0033)
water + Mg <sup>2+</sup>	1 (0.025)	0	32 (0.258)	0	2 (0.0037)

<sup>a</sup>The fraction of PCCA set probability that accounts for bound microstates is given in parentheses.

within the respective metastable water + Mg<sup>2+</sup> set is significantly larger than in pure water (see Table 3). Adding the Mg<sup>2+</sup> ion thus focuses the metastable set onto the binding-competent conformations, such that the surrounding energy barriers that prevent rapid exit out of the metastable set are much closer around the binding-competent conformations. Thus, addition of Mg<sup>2+</sup> may be understood as a conformational selection of the protein-binding conformation (see state I in Figure 6).

#### *Mg<sup>2+</sup>-UDP-GlcNAc Complex Could Bind to the Protein.*

There is no indication that the UDP-GlcNAc-2-Epimerase protein binds a divalent ion such as Mg<sup>2+</sup> as a cofactor in the UDP-GlcNAc binding site.<sup>66</sup> Is it nevertheless biologically significant that Mg<sup>2+</sup> stabilizes the binding-competent conformation of UDP-GlcNAc? In the water + Mg<sup>2+</sup> simulations, Mg<sup>2+</sup> is bound during the entire simulation time at a well-defined coordination site of the diphosphate group (see Figure 3e). When fitting this Mg<sup>2+</sup>-UDP-GlcNAc structure into the protein complex, it is found that the Mg<sup>2+</sup> would not hinder the binding of UDP-GlcNAc. Rather, Mg<sup>2+</sup> would be located outside of the protein, in the UDP-GlcNAc entrance channel. It is, thus, conceivable that Mg<sup>2+</sup> or other divalent ions interact with UDP-GlcNAc in the solute, saturating its charges and stabilizing the binding-competent conformation, and accompany the ligand until it binds specifically to a protein such as the UDP-GlcNAc-2-Epimerase. In ref 67 it was found that divalent cations are not necessary for UDP-GlcNAc epimerase activity, but do increase its rate under some conditions. It is conceivable that this increased rate is due to an increased binding rate resulting from such “ion-accompanied” binding. Figure 6 sketches the possible mechanism of ion-accompanied binding.

## CONCLUSIONS

In the present paper we present a general methodology to analyze the influence the chemical environment has on structure, thermodynamics, and kinetics of ligand molecules. These questions are investigated by using a combination of molecular dynamics simulations and Markov state models (MSMs). As an example, the ligand UDP-GlcNAc is analyzed in different chemical environments. The utilization of MSMs has permitted the systematic extraction of quantities that are otherwise difficult to access, such as the system's metastable sets and their thermodynamics, the relaxation time scales, and their link to structural rearrangements.

As expected, the conformational flexibility increases and the relaxation time scales reduce when the ligand is solvated in water. The reverse is observed when the ligand binds to the protein, where a specific binding conformation is stabilized by the binding pocket. Interesting changes occur when a Mg<sup>2+</sup> ion is added to the water solvent. On one hand, these changes are not dramatic, as the most populated metastable sets can be roughly associated in both scenarios. However, the metastable sets become smaller as the addition of Mg<sup>2+</sup> focuses the

conformations onto structures that are competent to interact with the ion. There is also a marked change in the conformational energetics: In the water environment, the most populated states have relatively low entropies, likely due to an ordering in the surrounding solvent molecules caused by exposed phosphate charges (electrostriction<sup>38,59</sup>). This effect is reduced when the Mg<sup>2+</sup> ion is attached, as the ion interacts with the negative phosphate charges, thus, effectively shielding them from the solvent.

Interestingly, the GlcNAc ring pucker is strongly affected by the presence of Mg<sup>2+</sup> because the interaction of Mg<sup>2+</sup> with the ligand stretches the phosphate backbone, thus, preventing an interaction between GlcNAc and UDP that stabilizes the otherwise unlikely <sup>1</sup>C<sub>4</sub> chair. Thus, the GlcNAc pucker may act as a sensor for Mg<sup>2+</sup> binding. In water solvent, the fraction of the <sup>1</sup>C<sub>4</sub> pucker is predicted to be 23%, which is large enough to be detected and quantified by NMR.<sup>58</sup> In water + Mg<sup>2+</sup>, the fraction is predicted to drop below 5%, which would be effectively invisible with current NMR techniques.

However, MSMs permit to explicitly calculate the system's kinetics, that is, its slowest relaxation time scales and the corresponding structural rearrangements. These can be linked using MSMs through the duality of eigenvalues and eigenvectors of the transition matrix. In the present system, the change of the pucker populations also has a dramatic effect on the system's kinetics: In presence of the Mg<sup>2+</sup> ion, the slowest relaxation time is predicted to be reduced by a factor of 7 compared to pure water solvent, mainly resulting from an increased rate of the <sup>1</sup>C<sub>4</sub> → <sup>4</sup>C<sub>1</sub> transition. In principle, all relaxation time scales of the system that can be theoretically calculated via MSMs are also experimentally measurable by kinetic experiments such as correlation experiments (e.g., fluorescence correlation, neutron/X-ray scattering) or perturbation-relaxation experiments (e.g., temperature jump, time-resolved IR).<sup>12,61</sup> The time scales that actually enter the experimental curve with significant amplitude, however, crucially depend on how well the experimental observable is able to trace changes along the corresponding eigenvectors. Therefore, experiments in which these observables can be controlled, for example, by site-specific labeling, are of special interest as they can be designed to specifically track relaxations predicted by an MSM analysis.<sup>12,61</sup> In the present case of a small ligand with time scales in the nanoseconds range, time-resolved IR spectroscopy may be a good candidate to complement simulation studies. The slowest time scales in water and water + Mg<sup>2+</sup> are in the range of tens to hundreds of nanoseconds and can thus be probed in terms of the IR spectrum relaxations after a sufficiently rapid temperature jump.<sup>68</sup> IR spectroscopy can be combined with site-specific isotope labeling, thus, offering the ability to select specific eigenvectors and measure specific time scales separately.<sup>69</sup>

Arguably the most interesting finding is how binding-competent ligand conformations are stabilized by the chemical environment. By adding water to the ligand, the energy landscape is changed such that binding competent conformations become accessible. By further adding Mg<sup>2+</sup>, these binding competent conformations are selectively stabilized since the Mg<sup>2+</sup> makes specific interactions with the diphosphate backbone of UDP-GlcNAc. At the same time, the nonbinding conformations that lie in the same metastable set are destabilized. Thus, the Mg<sup>2+</sup> ion narrows the energy well of the corresponding metastable set such that it “focuses” on the binding competent structures. This explains the surprising

finding that the binding-competent structures become more probable while at the same time the metastable set that contains these structures becomes less probable. Stabilization of this sort is mainly a kinetic effect: With an  $Mg^{2+}$  ion, the metastable set containing the binding-competent structures is less often visited, but when visited, the system spends more time in binding-competent structures than without an  $Mg^{2+}$  ion.

Association of divalent cations such as  $Mg^{2+}$  to phosphate groups is a well-known and important interaction in biomolecules.  $Mg^{2+}$  binds to pairs of phosphate groups in DNA and RNA and is important for the stabilization of the three-dimensional fold of RNA.<sup>3,5,70</sup> In ligands such as ATP and GTP,  $Mg^{2+}$  is often needed as a cofactor.  $Mg^{2+}$  cofactors have not only an electrostatic and structural role (compensation of the negatively charged phosphate groups), but is often needed for the catalytic reaction, hence, taking the role of a specific protein residue. The role of  $Mg^{2+}$  association to UDP-GlcNAc found in the present study is related, but different: There is no evidence that the UDP-GlcNAc-epimerase binds  $Mg^{2+}$  as a cofactor. The UDP-GlcNAc binding pocket of the epimerase is itself positively charged, hence,  $Mg^{2+}$  is probably not needed to coordinate or stabilize UDP-GlcNAc in the binding pocket. Nonetheless, it has been observed that  $Mg^{2+}$  has a positive effect on the catalytic rate of UDP-GlcNAc-2-Epimerase.<sup>67</sup> It is, thus, conceivable that binding-conformation selection of UDP-GlcNAc by  $Mg^{2+}$  is relevant before or during binding, but not after binding to the protein: The  $Mg^{2+}$  ion stabilizes the binding-competent conformation of UDP-GlcNAc and accompanies the ligand into its binding pocket. In this role,  $Mg^{2+}$  acts as a “binding co-factor”.

Note that we purely describe a possible conformational selection mechanism here and have not presented simulation data that would allow to conclude how the overall energetics of protein–ligand binding would be modulated based on a protein–ligand interaction. Figure 6 sketches two scenarios: a competitive scenario, in which protein and ion separate and compete for binding the ligand, thus, reducing the protein–ligand binding energy compared to the situation when no ion was present; and a cooperative scenario where protein, ligand, and ion bind together. We have discussed the mechanism based on the fact that, under sufficiently high  $Mg^{2+}$  conditions, we expect the transition  $U \rightarrow I$  to be preferred to direct binding but cannot decide whether B1 or B2 is more stable and which one of them should thus be considered as representative for the bound state.

Please note that from a biological point of view, the binding process of a single protein–ligand pair is not necessarily tuned such that their binding affinity is maximized.<sup>40</sup> Quite often in biological signaling processes, binding occurs at low affinities and it seems rather important to be able to go through entire binding-dissociation cycles sufficiently rapidly.

The role of the  $Mg^{2+}$  observed in this study needs to be confirmed by further simulation studies and experiments. In particular, it needs to be excluded that the observations made here are force field artifacts. For example, the reliability of the divalent cation treatment may be increased by using polarizable solvent methods such as amoeba.<sup>71,72</sup>

It remains to be investigated whether such an “ion-assisted” binding is a generally relevant mechanism for other protein–ligand pairs. Even if this is not the case, being able to understand how such cofactors may be used to bias the conformation of a ligand toward a desired state, for example,

the transition state of a substrate, is an important step toward creating a molecular toolbox for design of synthetic catalysts.

## ■ ASSOCIATED CONTENT

### 📄 Supporting Information

Results of Chapman Kolmogorov tests, histograms of all dihedral angles, and microstate overlap matrix. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [frank.noe@fu-berlin.de](mailto:frank.noe@fu-berlin.de).

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors wish to thank Werner Reutter, Stephan Hinderlich, and Christof Schütte (all FU Berlin) for enlightening discussions. We gratefully acknowledge funding from the German Science Foundation (DFG) through SFB 740 (M.H.) and Matheon (F.N.).

## ■ REFERENCES

- (1) Pace, C. N.; Treviño, S.; Prabhakaran, E.; Scholtz, J. M. *Philos. Trans. R. Soc., B* **2004**, *359*, 1225.
- (2) Neuweiler, H.; Löllmann, M.; Doose, S.; Sauer, M. *J. Mol. Biol.* **2007**, *365*, 856–869.
- (3) Kobitski, A. Y.; Nierth, A.; Helm, M.; Jäschke, A.; Nienhaus, G. U. *Nucleic Acids Res.* **2007**, *35*, 2047–59.
- (4) Qu, X.; Smith, G. J.; Lee, K. T.; Sosnick, T. R.; Pan, T.; Scherer, N. F. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 6602–6607.
- (5) Berezniak, T.; Zahran, M.; Imhof, P.; Jaeschke, A.; Smith, J. C. *J. Am. Chem. Soc.* **2010**, *132*, 12587–12596.
- (6) Doster, W.; Cusack, S.; Petry, W. *Nature* **1989**, *337*, 754–756.
- (7) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7220–7225.
- (8) Woenckhaus, J. *Biophys. J.* **2001**, *80*, 1518–1523.
- (9) Jäger, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M. *J. Mol. Biol.* **2001**, *311*, 373–393.
- (10) Kim, H. D.; Nienhaus, U.; Ha, T.; Orr, J. W.; Williamson, J. R.; Chu, S. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 4284–4289.
- (11) Neuweiler, H.; Doose, S.; Sauer, M. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16650–16655.
- (12) Noé, F.; Doose, S.; Daidone, I.; Löllmann, M.; Chodera, J. D.; Sauer, M.; Smith, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 4822–4827.
- (13) Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. *Science* **1997**, *276*, 1109–1112.
- (14) Schuler, B.; Lipman, E. A.; Eaton, W. A. *Nature* **2002**, *419*, 743–747.
- (15) Greenleaf, W. J.; Woodside, M. T.; Block, S. M. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 171–190.
- (16) Chodera, J. D.; Elms, P.; Noé, F.; Keller, B.; Kaiser, C. M.; Ewall-Wice, A.; Marqusee, S.; Bustamante, C.; Singhal-Hinrichs, N. Bayesian hidden Markov model analysis of single-molecule force spectroscopy: Characterizing kinetics under measurement uncertainty. <http://arxiv.org/abs/1108.1430>, 2011.
- (17) Miller, D. W.; Dill, K. A. *Protein Sci.* **1997**, *6*, 2166–79.
- (18) Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R. *Protein Sci.* **2000**, *9*, 10–9.
- (19) Lange, O. F.; Lakomek, N.-A.; Fares, C.; Schröder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; DeGroot, B. L. *Science* **2008**, *320*, 1471–1475.
- (20) Weikl, T. R.; von Deuster, C. *Proteins* **2009**, *75*, 104–10.
- (21) Snow, C. D.; Sorin, E. J.; Rhee, Y. M.; Pande, V. S. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 43–69.

- (22) Schaeffer, D. D.; Fersht, A.; Daggett, V. *Curr. Opin. Struct. Biol.* **2008**, *18*, 4–9.
- (23) van Gunsteren, W.; Dolenc, J.; Mark, A. *Curr. Opin. Struct. Biol.* **2008**, *18*, 149–153.
- (24) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 14766–14770.
- (25) Schütte, C.; Fischer, A.; Huisinga, W.; Deuffhard, P. *J. Comput. Phys.* **1999**, *151*, 146–168.
- (26) Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M. *J. Phys. Chem. B* **2004**, *108*, 6582–6594.
- (27) Rao, F.; Caflisch, A. *J. Mol. Biol.* **2004**, *342*, 299–306.
- (28) Singhal, N.; Pande, V. S. *J. Chem. Phys.* **2005**, *123*, 204909.
- (29) Chodera, J. D.; Dill, K. A.; Singhal, N.; Pande, V. S.; Swope, W.; Pitera, J. W. *J. Chem. Phys.* **2007**, *126*, 155101.
- (30) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. *J. Chem. Phys.* **2007**, *126*, 155102.
- (31) Noé, F.; Fischer, S. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154–162.
- (32) Noé, F. *J. Chem. Phys.* **2008**, *128*, 244103.
- (33) Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. *J. Am. Chem. Soc.* **2010**, *132*, 1526–1528.
- (34) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Fischbach, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. *J. Chem. Phys.* **2011**, *134*, 174105.
- (35) Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S. *J. Chem. Theory Comput.* **2011**, *7*, 3412–3419.
- (36) Senne, M.; Trendelkamp-Schroer, B.; Mey, A. S. J. S.; Schütte, C.; Noé, F. *J. Chem. Theory Comput.* **2012**, *8*, 2223–2238.
- (37) Deuffhard, P.; Weber, M. *ZIB Report* **2003**, 03–09.
- (38) Noe, F.; Daidone, I.; Smith, J. C.; di Nola, A.; Amadei, A. *J. Phys. Chem. B* **2008**, *112*, 11155–11163.
- (39) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 19011–19016.
- (40) Held, M.; Metzner, P.; Prinz, J.-H.; Noé, F. *Biophys. J.* **2011**, *100*, 701–710.
- (41) Bowman, G. R.; Voelz, V. A.; Pande, V. S. *J. Am. Chem. Soc.* **2011**, *133*, 664–667.
- (42) Chodera, J. D.; Noé, F. *J. Chem. Phys.* **2010**, *133*, 105102.
- (43) Sezer, D.; Freed, J. H.; Roux, B. *J. Chem. Phys.* **2008**, *128*, 165106.
- (44) Cui, W. Z. R. Z.; Silva, D.-A.; Huang, X. *J. Phys. Chem. B* **2011**, *115*, 5415–5424.
- (45) Voelz, V. A.; Jäger, M.; Zhu, L.; Yao, S.; Bakajin, O.; Weiss, S.; Lapidus, L. J.; Pande, V. S. *Biophys. J.* **2011**, *100*, 515a.
- (46) Schauer, R.; Kelm, S.; Reuter, G.; Roggentin, P.; Shaw, L. In *Biochemistry and Role of Sialic Acids*; Rosenberg, A., Ed.; Plenum Press: New York, 1995; pp 7–67.
- (47) Kirschner, K. N.; Woods, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10541–5.
- (48) Basma, M.; Sundara, S.; Çalgan, D.; Vernali, T.; Woods, R. J. *J. Comput. Chem.* **2001**, *22*, 1125.
- (49) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (50) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.
- (51) Mobley, D. L.; Chodera, J. D.; Dill, K. A. *J. Chem. Phys.* **2006**, *125*, 084902.
- (52) Spoel, D. V. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–18.
- (53) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (54) Hess, B.; Bekker, H.; Berendsen, H. J. C. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (55) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (56) Buchete, N.-V.; Hummer, G. *J. Phys. Chem. B* **2008**, *112*, 6057–69.
- (57) Sarich, M.; Noé, F.; Schütte, C. *SIAM Multiscale Model. Simul.* **2010**, *8*, 1154–1177.
- (58) Sattelle, B. M.; Almond, A. *Glycobiology* **2011**, *21*, 1651–62.
- (59) Dill, K. A. *Biochemistry* **1990**, *29*, 7133–55.
- (60) Prinz, J.-H.; Keller, B.; Noé, F. *Phys. Chem. Chem. Phys.* **2011**, *13*, 16912–16927.
- (61) Keller, B.; Prinz, J.-H.; Noé, F. *Chem. Phys.* **2012**, *396*, 92–107.
- (62) Bhattacharyya, S.; Kerzmann, A.; Feig, A. L. *Eur. J. Biochem.* **2002**, *269*, 3425–32.
- (63) Weiss, S. *Science* **1999**, *283*, 1676–1683.
- (64) Gerwert, K.; Souvignier, G.; Hess, B. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 9774–9778.
- (65) Boehr, D. D.; Nussinov, R.; Wright, P. E. *Nat. Chem. Biol.* **2009**, *5*, 789–796.
- (66) Campbell, R. E.; Mosimann, S. C.; Tanner, M. E.; Strynadka, N. C. *Biochemistry* **2000**, *39*, 14993–5001.
- (67) Hinderlich, S.; Sonnenschein, A.; Reutter, W. *BioMetals* **1998**, *253*–258.
- (68) Ballew, R. M.; Sabelko, J.; Reiner, C.; Gruebele, M. *Rev. Sci. Instrum.* **1996**, *67*, 3694–3699.
- (69) Ihalainen, J. A.; Paoli, B.; Muff, S.; Backus, E. H. G.; Bredenbeck, J.; Woolley, G. A.; Caflisch, A.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 9588–93.
- (70) Misra, V. K.; Draper, D. E. *J. Mol. Biol.* **2002**, *317*, 507–21.
- (71) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; et al. *J. Phys. Chem. B* **2010**, *114*, 2549–64.
- (72) Ren, P.; Wu, C.; Ponder, J. W. *J. Chem. Theory Comput.* **2011**, *7*, 3143–3161.