

Domain motions of hyaluronan lyase underlying processive hyaluronan translocation

Harshad V. Joshi,^{1,2} Mark J. Jedrzejewski,³ and Bert L. de Groot^{1*}

¹Computational Biomolecular Dynamics Group, Max-Planck-Institute for Biophysical Chemistry, Göttingen, Germany

²Department of Chemistry, University of Cincinnati, Cincinnati, Ohio 45221

³Center for Immunobiology and Vaccine Development, Children's Hospital Oakland Research Institute, Oakland, California 94609

ABSTRACT

Hyaluronan lyase (Hyal) is a surface enzyme occurring in many bacterial organisms including members of *Streptococcus* species. Streptococcal Hyal primarily degrades hyaluronan-substrate (HA) of the extracellular matrix. This degradation appears to facilitate the spread of this bacterium throughout host tissues. Unlike purely endolytic degradation of its other substrates, unsulfated chondroitin or some chondroitin sulfates, the degradation of HA by Hyal proceeds by processive exolytic cleavage of one disaccharide at a time following an initial endolytic cut. Molecular dynamics (MD) studies of Hyal from *Streptococcus pneumoniae* are presented that address the enzyme's molecular mechanism of action and the role of domain motions for processive functionality. The analysis of extensive sub-microsecond MD simulations of this enzyme action on HA-substrates of different lengths and the connection between the domain dynamics of Hyal and the translocation of the HA-substrate reveals that opening/closing and twisting domain motions of the Hyal are intimately linked to processive HA degradation. Enforced simulations confirmed this finding as the domain motions in SpnHyal were found to be induced by enforced substrate translocation. These results establish the dynamic interplay between Hyal flexibility and substrate translocation and provide insight into the processive mechanism of Hyal.

Proteins 2009; 76:30–46.
© 2008 Wiley-Liss, Inc.

Key words: molecular dynamics; simulation; processivity; sugar degradation; principle component analysis; force probe; essential dynamics; free energy; umbrella sampling.

INTRODUCTION

Hyaluronan lyase (Hyal) is one of the major surface proteins of many bacterial organisms including members of *Streptococcus* species, and is present on most strains of these organisms. The expression of Hyal appears to be directly related to the virulence of this organism.^{1,2} The enzyme facilitates bacterial spread among the host tissues, by degrading essential components of the host's extracellular matrix (ECM), hyaluronan (HA), unsulfated chondroitin (CH), and certain chondroitin sulfates (CHS). As a result, it also provides a source of carbon and energy for bacterial cells. Hyal is a four-domain protein consisting of an N-terminal domain, followed by a spacer domain, a catalytic domain (termed α -domain), and a C-terminal domain (termed β -domain) [Fig. 1(A)].^{3–5} The first of these domains is a substrate (HA/CH/CHS)-binding domain that acts to enhance the Hyal affinity for substrates and localize/concentrate the enzyme to its site of action of mostly aggregated HA/CH/CHS within the ECM; consequently, this domain directs presumably only one single polymeric substrate strand (out of an aggregated cluster of them) toward the catalytic domain.^{5,6} The second, small domain acts as a spacer to distance the catalytic α -domain from the substrate-binding domain at the extreme N-terminus and thus can facilitate the above described feeding of the substrate for catalysis by the α -domain.⁵ These two N-terminal domains do not appear to affect catalysis *in vitro* at all, and the data are consistent with their role *in vivo* as described earlier.⁵ The β -domain modulates the access of the substrate to the catalytic machinery of the α -domain. Therefore, most often only the two C-terminal

Additional Supporting Information may be found in the online version of this article.

Abbreviations: μ s, microsecond; CH, chondroitin; CHS chondroitin sulfate; COM, center of mass; ECM, extracellular matrix; EDSAM, essential dynamics sampling simulation; FPMD, force probe molecular dynamics; HA, hyaluronan; Hyal, hyaluronan lyase; MD, molecular dynamics; ns, nanosecond; PCA, principal component analysis; RMSD, root mean-square deviation; RMSE, root mean-square fluctuation; SagHyal, *Streptococcus agalactiae* hyaluronan lyase; SpnHyal, *Streptococcus pneumoniae* hyaluronan lyase.

Grant sponsor: Deutsche Forschungsgemeinschaft; Grant number: GR 2079/1.

*Correspondence to: Bert L. de Groot, Abt. 105-10, Computational Biomolecular Dynamics Group, Max-Planck-Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany.
E-mail: bgroot@gwdg.de.

Received 9 April 2008; Revised 12 September 2008; Accepted 16 October 2008

Published online 4 November 2008 in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/prot.22316

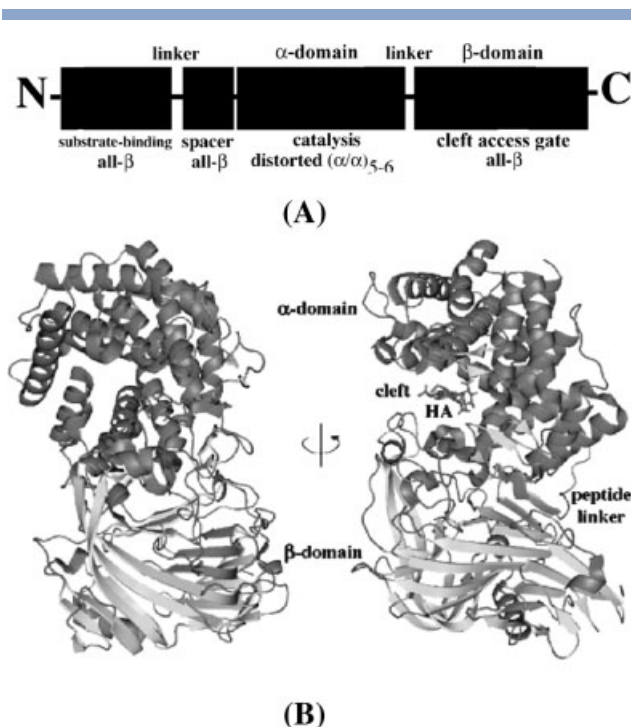


Figure 1

(A) A schematic representation of the domain blocks of the full *SpnHyal*. In the crystal structure, the first two domains at the N-terminus are autodegraded and the stable structure of the last two domains at the C-terminus with the linker is obtained. (B) Structural details of *SpnHyal*. Two views of the crystal structure of *SpnHyal* (pdb code: 1LOH) in cartoon representation. The α -helical domain (α -domain) and the β -sheet domain (β -domain) are marked. The small linker peptide can be seen in the right picture joining the two domains. The HA-binding cleft is clearly visible in the α -domain with HA in stick representation.

domains, the α - and β -domains, are studied as their catalytic and processivity activity *in vitro* is indistinguishable from the full-length enzyme^{1,7} [Fig. 1(B)]. In addition, the full-length four-domain streptococcal Hyals undergo degradation to smaller forms.^{7,8} The three-dimensional structure of the truncated and stable form, consisting of the α - and β -domains, of the pneumococcal enzyme has been elucidated by X-ray crystallography³ and was followed by a variety of structural,^{9–11} biochemical, molecular biology, and flexibility studies.^{9,11–15} The α -domain contains a cleft large enough to accommodate a hexasaccharide substrate and is responsible for substrate catalysis. These studies have led to the formulation of a proposed mechanism of catalysis for this enzyme termed as “proton acceptance and donation”³ and its mode of action during degradation of different substrates.^{9,16} The catalytically important residues as well as those involved in substrate binding have been identified.^{3,11–13} In addition, flexibility studies using CONCOORD¹⁷ and molecular dynamics (MD) simulations have allowed the

identification of significant flexibility of Hyal and identified three types of domain motions.

These domain motions (see Fig. 2) in order of decreasing amplitude are*: (i) an opening/closing of the substrate-binding cleft accompanied by the movement of catalytic amino acid residues and those responsible for the release of the product of degradation (termed Eigenvector 1 or *vec1*). This opening mode was confirmed by a recently resolved crystal structure with the cleft in an “open” conformation (referred to as PEGMME or *holo-SpnHyal*-OPEN), whereas all earlier structures had the catalytic cleft in a “closed” conformation; (ii) a rotation/twisting motion of the α -domain with respect to the β -domain that yields a ~ 10 Å movement along the cleft axis (termed Eigenvector 2 or *vec2*). The third movement (iii) involves a side-wise bending of the domains resulting in an effective opening and closing of the access/entrance to the cleft (termed Eigenvector 3 or *vec3*). These motions have been hypothesized to be directly related to the enzyme function: catalysis, endolytic “random-bite” degradation, and exolytic cleavage of one disaccharide at a time from the reducing to nonreducing direction in a processive manner.^{9,12,16,18}

Here, we test this hypothesis with an extensive analysis of a large set of free and enforced sub-microsecond MD simulations of Hyal with different lengths of the HA-substrate inside the cleft. This study establishes for the first time the dynamic relationship between the enzyme flexibility and the translocation phase of the HA-substrate, thereby providing direct insight into the role of domain flexibility in the processive mechanism of catalytic degradation of HA. Preliminary calculations to reconstruct free energy profile from different enforced simulations identify the rate-limiting intermediate states during the translocation process.

THEORY AND METHODS

Several free (“real time MD”) and enforced MD simulations of ligand-bound and ligand-free *SpnHyal* were carried out with the MD software package GROMACS (version 3.1.4, 3.2.1, and 3.3).^{19–21} The starting point for all the simulations was the crystal structure of the *SpnHyal*-hexasaccharide HA complex (*holo-SpnHyal*-CLOSED) with PDB code 1LOH.⁹

For the *SpnHyal*-*apo* simulation, the HA-substrate was removed from the structure manually. Other lengths of the HA-substrates were modeled inside the cleft as needed. The terminal residues (1Val and 721Leu) were completed and protonation states for histidines were determined using WHAT IF.²² The system was subsequently energy-minimized using a steepest-descent

*Previously, these domain motions were mentioned in the order according to the results from CONCOORD simulations (See Refs. 9 and 12). Here, we change it to one that is more consistent to the crystallographic differences and explicit MD simulations.

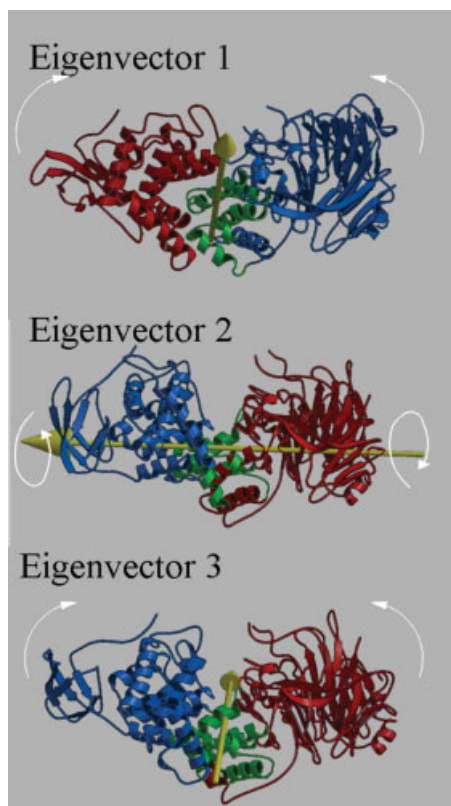


Figure 2

Principal modes of flexibility of the domain motions of *SpnHyal* as obtained from a PCA of crystallographic differences.¹² Eigenvector 1: opening closing of the cleft is the largest amplitude domain motion, Eigenvector 2: twisting of the α -domain with respect to the β -domain, and Eigenvector 3: bending of the domains thereby modulating cleft access.

algorithm, followed by solvation in a rectangular box of explicit water solvent. The 515 crystallographic water molecules were retained. The water model used was TIP4P.²³ For the protein, the OPLS-AA force field^{24,25} was used, and HA parameters were derived by modifying the molecule-specific dihedral parameters in the OPLS-AA force field for the carbohydrate (Joshi H and de Groot BL, 2007). Na^+ counter ions were added to neutralize the overall charge of the system. A typical simulation system consisted of $\sim 80,000$ – $100,000$ atoms (see Table II). See Figure 3(A) for a typical simulation setup.

Each of the simulation systems was subsequently energy-minimized again to obtain the starting configuration for the simulations. The solvent, HA, and the protein were separately coupled to an external temperature bath²⁶ of 300 K with a relaxation time of 0.1 ps unless otherwise stated. In all simulations, the system was weakly coupled to a pressure bath of 1 atm with isotropic scaling and a relaxation time constant $\tau_p = 1$ ps. Bond lengths were constrained to their equilibrium values using the LINCS algorithm.²⁷ This algorithm allows a 2

fs time step for the leap-frog integration scheme. For the Lennard–Jones interactions, a cutoff distance of 1 nm was applied. Electrostatic interactions between charge groups at a distance less than 1 nm were calculated explicitly, whereas long-range electrostatic interactions were calculated using the Particle-Mesh-Ewald method²⁸ with a grid spacing of 0.12 nm and a fourth-order spline

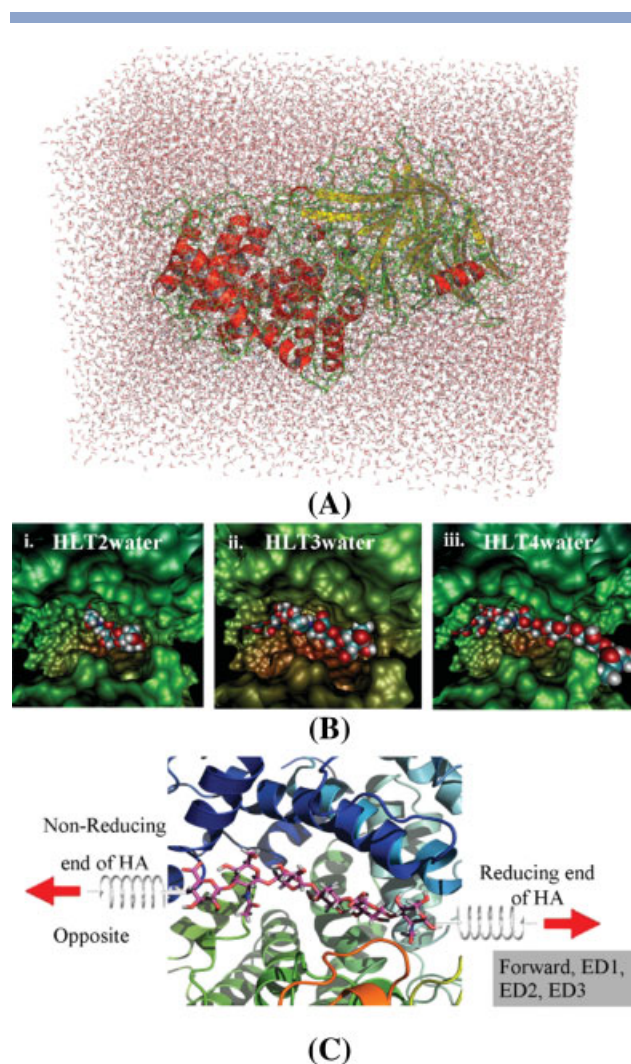


Figure 3

(A) A typical simulation set up for the *SpnHyal/SpnHyal*-HA system. The enzyme is shown in cartoon representation. Water molecules are shown in line representation. (B) Different lengths of HA modeled inside the cleft: (i) tetrasaccharide HA (T2) in a processive position (ready to translocate in a processive cycle) (ii) hexasaccharide HA (T3) in a processive position ready to translocate in a processive cycle, and (iii) octasaccharide HA (T4) in a catalytic position ready for the catalytic action by the enzyme. (C) Scheme of the FPMD simulations setup. The arrow at the “nonreducing” end (left) of the sugar depicts the pulling of HA in the “backward” direction (in the nonprocessive direction). Similarly, the arrow at the reducing end of the HA (right) denotes the pulling in the putative processive direction for four cases (pulling with protein free and pulling with each of the eigenvectors of the protein fixed).

Table I
Crystal Structures Referred to in this Work and Their Abbreviations

Abbreviation	Bacterium	Ligand	Resolution (Å)	PDB code
<i>SpnHyal</i> ³	<i>S. pneumoniae</i>	None	1.56	1EGU
<i>holo-SpnHyal</i> ^{9,10}		Hyaluronan-		
		Disaccharide/	1.70	1C82
		Tetrasaccharide/	1.53	1LXK
		Hexasaccharide	1.70	1LOH
<i>SpnHyal</i> -OPEN ¹⁴ [PEGMME]		None	2.80	2BRW
<i>SpnHyal</i> -MALONATE ¹⁴		None	3.30	2BRV
<i>SpnHyal</i> -vitamin C ¹⁸		Vitamin C	2.00	1F9G
<i>SagHyal</i> ⁴	<i>S. agalactiae</i>	None	2.10	1F1S
<i>SagHyal</i> - <i>holo</i> ¹²		Hyaluronan-		
		Hexasaccharide	2.20	1LXM

Starting structure for the simulations is *holo-SpnHyal*-CLOSED. *SpnHyal*-OPEN and *SpnHyal*-MALONATE refer to the recently solved *SpnHyal* structures while *SagHyal* refers to structures from different bacterial species.

interpolation. A 2 ns MD simulation was carried out with harmonic restraints on the protein heavy atoms with a force constant of $k = 1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ to equilibrate water and ions of the system. A subsequent MD simulation of 5–10 ns length was performed to equilibrate the whole system while the protein backbone RMSD was monitored. Finally, production runs were carried out for several nanoseconds (~ 100 ns), storing the coordinates of all the atoms at every picosecond for further analysis. Particular simulation systems were set up with simulation details as discussed later for each case.

A principal component analysis (PCA) was used to analyze the concerted fluctuations of domains from the MD trajectories. PCA is based on the principle of diagonalization of the covariance matrix of the atomic fluctuations calculated from an ensemble of structures yielding a set of eigenvectors and eigenvalues with the principal modes representing the directions of the largest collective fluctuations. The “essential dynamics” or the dynamics in the essential subspace consisting of the largest eigenvalue modes are frequently functionally relevant and in the focus of computational studies.^{29,30}

To compare the collective motions of the protein in different simulations, a PCA was first performed on the common set of C^α atoms of the available crystal structures (see Table I). This PCA represents the structural differences occurring in the crystallographic data. By directly projecting different simulations of the *SpnHyal*-HA onto the resulting set of eigenvectors, the dynamic exploration of these structural differences in the simulations is quantitatively assessed while allowing a direct comparison of the extent of flexibility to the available X-ray conformations.

Free simulations

The details of the free (unbiased or “real-time”) simulations carried out are shown in Table II. Two simulations already discussed in earlier work,¹⁶ namely *apo*-

SpnHyal and *holo-SpnHyal*-CLOSED, were extended to 100 ns. An additional simulation was performed with a similar configuration to that of *holo-SpnHyal*-CLOSED to enhance statistical significance. The latter simulation was also carried out for 100 ns but with different initial conditions of coordinates and velocities (coordinates of the previous *holo-SpnHyal*-CLOSED simulation at 10 ns as a starting structure). In subsequent simulations, the HA-substrate was modified such that it resembles the substrate just after the catalytic action has been performed (see later for details). That way, the HA-substrate in the starting structure would require to translocate by one disaccharide unit through the cleft to be ready for the next catalytic round.

Simulations including tetra-, hexa- and HA-octasaccharides were carried out as follows: for the simulations with a tetrasaccharide-HA-substrate, the first disaccharide of the hexasaccharide from the reducing end was manually removed and the remaining tetrasaccharide part was treated as a substrate [see Fig. 3(B)-i]. With this configuration, two types of simulations were performed. In the crystal structure (PDB code: 1LOH), the catalytic residue Tyr408 was mutated to Phe to deactivate the enzyme for studies of complexation with HA.⁹ This mutation concerns a catalytic residue and so presumably would not affect the processivity mechanism (the sliding phase of the HA-substrate).^{3,13} Nevertheless, to investigate the effect of this change, the original protein was obtained by substituting Phe408 back to the original residue Tyr, that is, F408Y, and both structures were used for subsequent simulations.

For the simulations with hexasaccharide-HA as the substrate, the first disaccharide from the reducing end was manually removed as described earlier and then another disaccharide was modeled at the nonreducing end. This resembles the *holo-SpnHyal*-CLOSED structure but shifted by one disaccharide unit toward the nonreducing end [see Fig. 3(B)-ii]. Simulations based on these structures were performed to investigate whether in

Table II

List of Simulation Systems with the Simulation Time in ns (the Equilibration Time in Brackets) and Total Number of Atoms in the System

No.	Simulation	Protein structure	HA substrate length and position	Time [ns]	Total System size [atoms]
1.	<i>apo-SpnHyal</i>	1loh	–	100 (5)	77,278 (7)
2.	<i>holo-SpnHyal</i> -CLOSED	1loh	Hexasaccharide ready for catalysis	100 (5)	83,542 (10)
3.	<i>holo-SpnHyal</i> -CLOSED2	1loh	Hexasaccharide ready for catalysis	100 (5)	83,542 (10)
4.	HLT2water1	1loh	Tetrasaccharide ready to slide	100 (5)	89,550 (9)
5.	HLT2water2	1loh	Tetrasaccharide ready to slide	85 (5)	89,167 (9)
6.	HLT2water3	1loh structure back mutated to F408Y	Tetrasaccharide ready to slide	95 (5)	89,551 (9)
7.	HLT2water4	1loh structure back mutated to F408Y	Tetrasaccharide ready to slide	40 (5)	89,460 (9)
8.	HLT3water1	1loh	Hexasaccharide ready to slide	100 (7.5)	100,681 (10)
9.	HLT3water2	1loh	Hexasaccharide ready to slide	100 (7.5)	100,682 (10)
10.	HLT3water3	1loh	Hexasaccharide ready to slide	100 (7.5)	100,486 (10)
11.	HLT4water1	1loh	Octasaccharide ready to slide	75 (10)	162,681 (11)
12.	HLT4water2	1loh	Octasaccharide ready to slide	46 (10)	165,341 (11)
13.	HLT4water3	1loh	Octasaccharide ready for catalysis	70 (10)	165,341 (11)

The number in the brackets of the last column relates to number of counter ions (Na^+) used to neutralize the system charge.

unbiased simulations, the HA-substrate would spontaneously slide through the cleft in the reducing end direction to yield a structure as observed in the *holo-SpnHyal*-CLOSED structure. Similarly, for two of the simulations (sim. nos. 11 and 12) with octasaccharide as substrate, the initial structures were also modeled such that the octasaccharide is at the start of the next round of the catalytic cycle or sliding phase. For sim. no. 13, with the octasaccharide as substrate, the HA was placed as in the *holo-SpnHyal*-CLOSED structure, that is, the substrate ready for catalytic action [see Fig. 3(B)-iii]. The details of each simulation system are given in Table II. Each series of simulations identical in structure differs in the initial conditions of coordinates and velocities. The total simulation time was more than 1 μs .

FPMD and essential dynamics sampling simulations

In Force Probe Molecular Dynamics (FPMD) simulations, the system can be forced to a predefined state by “pulling” one or more pull groups or parts of the system in a selected pulling direction specified in Cartesian coordinates. This mimics the situation in an AFM experiment by employing a harmonic potential (or a “virtual” spring), which acts as a flexible cantilever. This technique allows the system to explore the areas of the configurational space that it otherwise might not visit in the short time span of “real-time” MD simulations. FPMD simulations are becoming increasingly popular and find many applications.^{31–34}

Defining an appropriate reaction coordinate in the force-induced translocation of the HA-substrate is an important step. To ensure that the HA-substrate was “pulled” along an appropriate reaction coordinate in the putative catalytic/processivity direction, the first sugar ring at the reducing end of the HA-substrate was sub-

jected to a harmonic potential while keeping fixed the center-of-mass of the catalytic cleft of the enzyme. The forces were recorded every step. Maxima of these force profiles denote rupture forces corresponding to the phenomena where the sugar starts moving along the chosen reaction coordinate. A schematic representation of the FPMD simulations is depicted in Figure 3(C).

Choosing the speed or force constant is not a very critical issue in this type of simulation as long as it does not impose a change of the pathway and the force profiles are not friction dominated. This can be ensured by choosing the pulling velocities from the thermally activated regime (i.e., a logarithmic dependency of the pulling velocity on the rupture forces). From an initial set of simulations performed with different speeds and force constants, such a comparison (see Supp. Info. Fig. S2) was performed. From this comparison, a force constant of $\kappa = 750 \text{ kJ mol}^{-1} \text{ nm}^{-2}$ and a pulling velocity of $v = 1 \text{ nm ns}^{-1}$ was determined to be suitable.

In Essential Dynamics sampling (EDSAM), another efficiency-enhancing sampling technique,^{35,36} the reaction coordinate can be deduced from a PCA. The reaction coordinate represents one or several of the principal modes of motion resulting from the PCA of the system. Accordingly, the system is driven along this reaction coordinate. This way, external bias is minimized because now the reaction coordinate is an inherent property of the system.

To investigate the influence of the protein domain motions on the translocation of the HA-substrate, we used FPMD simulations alone as well as in conjunction with the EDSAM technique. In the latter case, EDSAM was used to constrain the protein in a specific position along each of the main domain modes to investigate their influence on sugar translocation. In simulations with EDSAM constraints, coordinates were reset every

time step to prevent motion along the respective eigenvector coordinate.

Construction of a free energy profile

When FPMD simulations or EDSAM simulations are carried out using a constraint force (in the harmonic sense this would correspond to an infinitely stiff spring), a free energy profile for the substrate translocation in the processive cycle can be constructed through integration over the mean force (averaged at several points along the pathway). Such a free energy profile (also termed potential of mean constraint force (PMF)) provides information on, for example, free energy barriers along the reaction pathway. An evaluation of the structures corresponding to such barriers then may reveal critical enzyme–substrate interactions.

EDSAM simulations force the system to move along the chosen principal modes. One method used in this work to estimate the free energy profile was to fix the essential coordinates at n equidistant points, spaced by dz (by use of “linfix” module in GROMACS package with a stepsize of zero). To obtain the equilibrium forces acting on the system at those coordinates, the remaining $3N - 1$ degrees of freedom, for an N atom system, were allowed to equilibrate for a period of 25 ns for each set. Excluding the initial equilibration period, one can then calculate the mean force at each of these n points along the principal mode. The PMF is given by the sum over the average forces of the n points,

$$\Delta G_{\text{PMF}} = \sum \langle F \rangle dz, \quad (1)$$

where dz is the distance between sampling points along the principal mode z and $\langle F \rangle$ is the force acting out the system along the remaining coordinates. Convergence can be slow because of equilibration effects and high force fluctuations. The convergence can be assessed by comparing PMFs obtained from windows of different lengths and parts of the trajectories.

We performed three sets of simulations to compute free energies using this method. To obtain the free energies, 20 equidistant positions of HA along the reaction coordinate were chosen for each of the three runs. By keeping the sugar “locked” in each of these positions, the system was relaxed for the duration of each window. The simulation time for each window was 25 (for run1), 10 (for run2), and 25 ns (for run3), and thus totaling a simulation time of more than 1.2 μs .

Umbrella sampling

Another widely used method to assess the free energy profile from simulated trajectories is umbrella sampling. In this method, many overlapping windows along the chosen reaction coordinate are simulated simultaneously.

An artificial harmonic biasing potential, called umbrella potential, is added to each window simulation at systematically spaced positions along the reaction coordinate. This umbrella potential forces the system to sample a biased distribution within that window. After sufficiently long simulations (equilibration per window) are carried out for each window, the umbrella bias is removed to obtain the final estimate of PMF. We used the Weighted Histogram Analysis method³⁷ (WHAM), which makes use of all the information in umbrella sampling and does not discard the overlapping regions. In particular, the WHAM technique computes the total unbiased distribution function as a weighted sum of the unbiased distribution functions. The weighting function can be expressed in terms of known biased distribution functions. We use this procedure to obtain the PMF along the reaction coordinate of sugar-translocation obtained by PCA. At each simulation window, the bias potential was introduced in practice by a harmonic potential applied to the collective sugar translocation coordinate as implemented in the conformational flooding technique.³⁸ Two sets of umbrella sampling simulations (US1 and US2) were carried out. In these simulations, structures at 50 equidistant windows for US1 and 60 equidistant windows for US2 were obtained from EDSAM simulations (see Theory and Methods). Umbrellas were created by biasing the simulations with a harmonic potential. The simulation was performed for 25 ns for set US1 and 20 ns for set US2 for each window, thus totaling almost 2.2 μs .

RESULTS AND DISCUSSION

Later, we discuss the results from the free simulations (Table II) with different lengths of the HA ligand inside the cleft and from the enforced simulations with the hexasaccharide HA inside the cleft. These simulations establish the dynamic relationship between the large-scale domain motions of the protein and the processive motion of the HA-substrate.

Equilibration and free dynamics/flexibility of the *Spn*Hyals

Large-scale conformational changes or domain flexibility in proteins are often related to function. To investigate the interplay between substrate presence and protein dynamics in bacterial Hyals, we studied the influence of the presence of HA-substrates on the flexibility of Hyals. As a measure of structural stability and flexibility, the root mean-square deviation (RMSD) from the crystal structure ILOH (i.e., the starting structure of the simulations) was monitored during the equilibration. For all simulations, after an initial equilibration period, the RMSD for the protein C^α atoms was found to level off between 1.25 and 2.5 \AA (see Supp. Info. Fig. S1). Coordi-

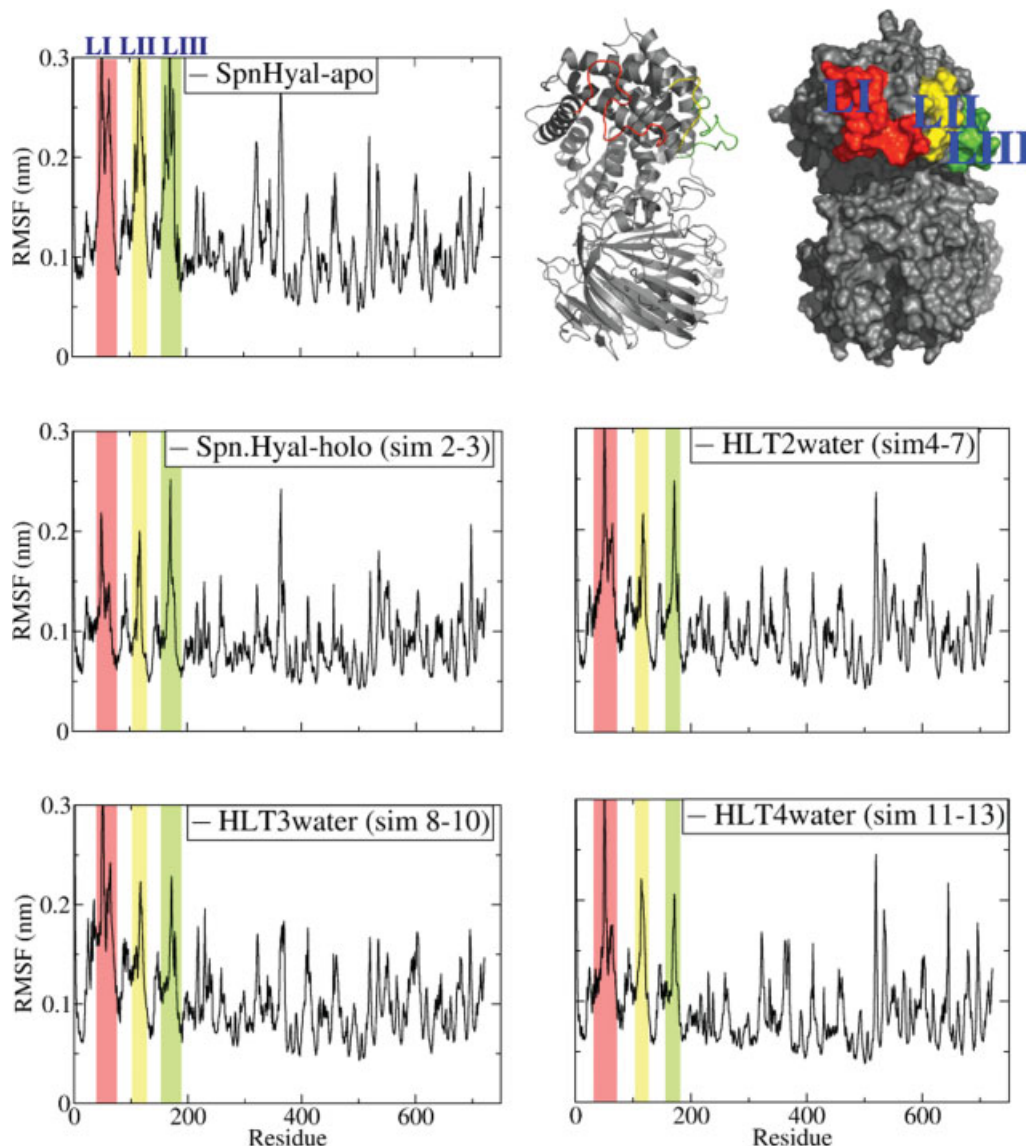


Figure 4

RMSF per residue for all the simulations. In the upper right corner, the protein is shown in cartoon and in surface representation. The three flexible loops are marked with LI: red, LII: yellow, and LIII: green. The RMSF for these loops are highlighted in the same colors in all the graphs. For the simulations with similar structure, the RMSF per residue is averaged for simplicity.

nates after this initial equilibration were used for any further analysis.

The root mean-square fluctuations (RMSF) per residue for the *apo*-*SpnHyal* simulation and all other simulations are shown in Figure 4. A striking feature is the lower RMSF for the three flexible loops around the catalytic cleft in the *holo*-simulations when compared with the *apo*-simulation. These loops are indicated as LI, LII, and LIII in Figure 4. In the *apo* simulation, the RMSF for these loops is very high, up to 4.0 Å. This fluctuation is lower (ranging from 2.0 to 3.0 Å) in all cases in which the HA molecule, longer than tetrasaccharide, is present/bound in the cleft, showing the influence of the

presence of the HA-substrate on these fluctuations. To the contrary, the RMSF for residues 27–43, which form a helix at the entry side of the cleft (just preceding loop LI) (see Fig. 4), is higher for most simulations with the HA-substrate present (simulations 4–11 in Table II). In the cases where the RMSF for this helix is high, the position of the HA-substrate is such that it is ready for the next round of catalysis, suggesting a possible role in the processive mechanism. Interestingly, mutation studies also indicate the residues from LII to LIII to be of critical influence in the overall processive mechanism of *SpnHyal*.¹¹ Loop LIII may play a role in product release.

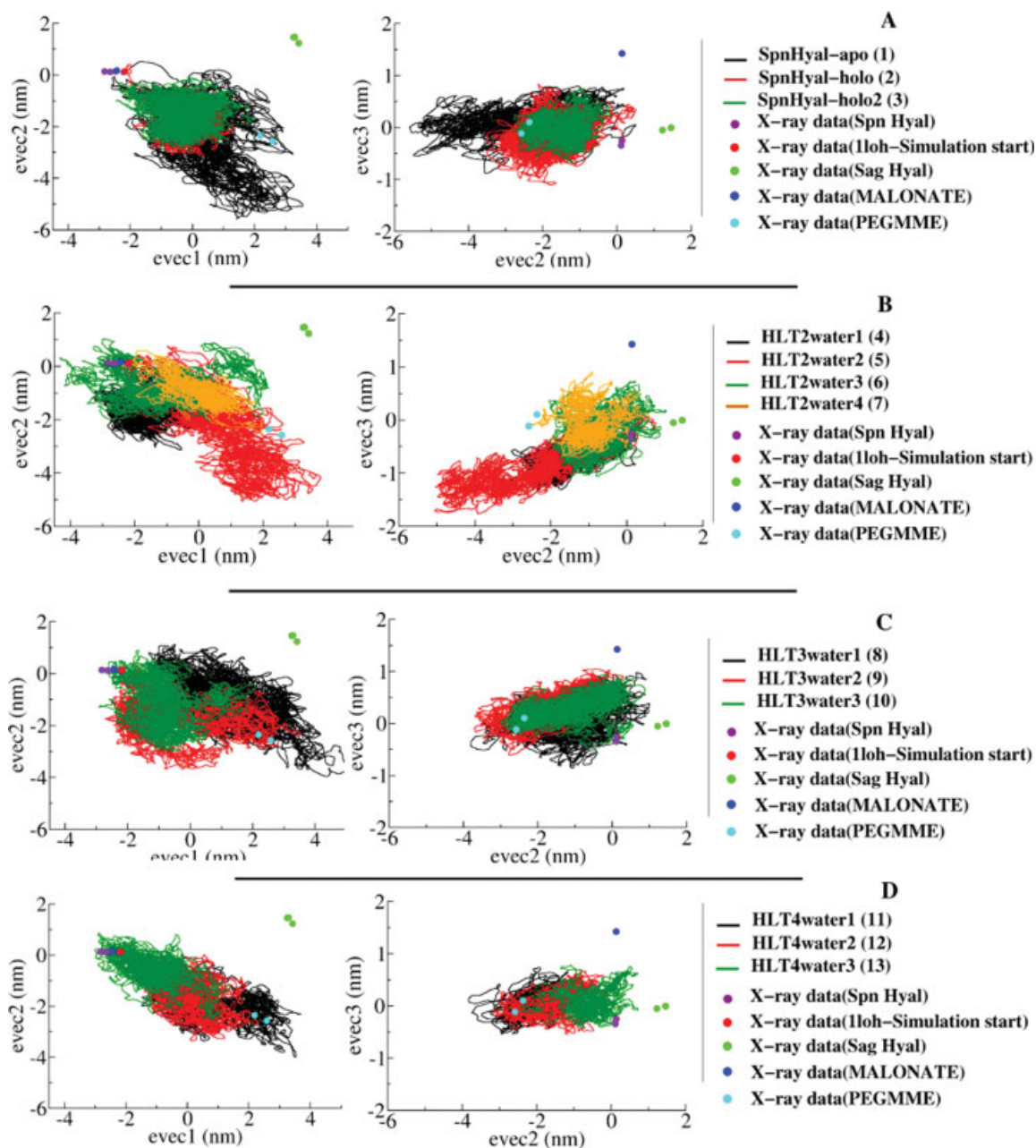


Figure 5

Two-dimensional projections of the simulation trajectories together with the X-ray structures (see Table I) projected on the essential subspace obtained from the PCA. Two different projections per set of the simulations comprising similar structures are shown on the left and the right panel. On the left hand side, MD trajectories with the available X-ray structures of Hyal (circles) are projected onto the first two principal modes of the PCA. The x -axis gives the extent of the opening/closing motion of the cleft, whereas the second eigenvector (y -axis) shows the twisting mode (twisting between α - and β -domain). On the right hand side panel, the projections on the second and third eigenvector are shown, with the third eigenvector describing the opening/closing motion of the entrance of the cleft (y -axis). (A) Projections of the MD trajectories of the simulations of Hyal-*apo* and two *holo*-Hyal simulations (simulations of 1LOH structure). Similar projections of the trajectories of simulations of SpnHyal with the (B) tetrasaccharide HA, (C) hexasaccharide HA, and (D) octasaccharide HA inside the cleft.

PCA and domain motions of the SpnHyal

Previous studies^{9,12,16} suggested a functional role of three types of domain motions of Hyal in the processive mechanism. Accordingly, the current simulation results

were analyzed in terms of these domain motions as derived from a PCA of the X-ray conformers (see Fig. 2) to test this hypothesis. The results are summarized in Figures 5 and 6 (see Supp. Info. Fig. S3). The two-

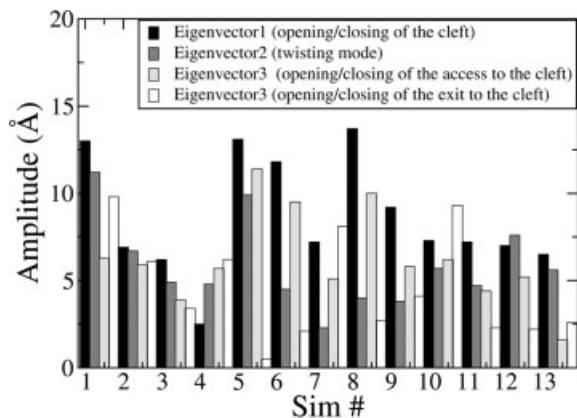


Figure 6

Bar-graph showing the extent of the dynamic domain motions in all the simulations. The order of the simulations on the x -axis is the same as in Table II. The measurement of the amplitude of each eigenmode is shown on y -axis (in Å). The measurement of these amplitudes and the choice of residues are described in detail in the text. For the first eigenvector, the distance between the C^α — C^α separations of Asn231 and Gly769 in the most open and least open trajectory snapshots gives the extent of the opening/closing motion of the cleft. The twisting extent (amplitude of the second eigenvector) is measured by the separation between the C^α atoms of Asp340 for the superimposed most twisted and the least twisted snapshots from the trajectory. The last two columns give the extent of the amplitudes for the third eigenvector showing entry-side opening and exit-side opening, respectively. For the entry site opening, the distance between the C^α — C^α separations of Asp211 and Ser771 is measured, whereas for the exit side opening, the distance between the C^α — C^α separations of Asn341 and Asn580 is measured to give the extent of the opening/closing of the access to the cleft. Simulation 1 (with no substrate) and simulations 4–7 (with tetrasaccharide inside the cleft) show the widest range of the fluctuations showing minimal influence of the substrate on the domain motions. Simulations 8–12 show in general lower amplitudes for the fluctuations indicating larger influence of the longer substrate. Simulations 2, 3, and 13 show the least fluctuations indicating the importance of the position of the substrate inside the cleft.

dimensional projections of both simulated trajectories and X-ray structures onto the corresponding first three eigenvectors are shown in Figure 5. In each panel A, B, C, and D, the graph on the left side shows the projections onto the first two eigenvectors. These graphs demonstrate the opening/closing motion of the cleft on the x -axis and the twisting motion along the y -axis. The green circles in the upper right corner represent the *S. agalactiae* X-ray conformations. All 13 simulations leave these conformations unsampled, confirming the structural divergence among bacterial Hyal structures as was speculated earlier. All simulations in which the HA is in a position similar to the *holo*-X-ray structure (ILOH) [both *holo*-*SpnHyal*-CLOSED simulations (sim. nos. 2 and 3, red and green traces from A-left) and HLT4water3 simulation (sim. no. 13, green trace from D-left), see also Fig. 6] sample the opening/closing motion of the cleft reversibly, but to a smaller extent than other simulations,

some of which sample this motion up to the recently solved “open” conformations (*holo*-*SpnHyal*-OPEN, cyan circles). These results suggest a correlation between the presence and location of the substrate and the Hyal domain motions. The simulations that sample the opening motion substantially also show a substantial sampling of the twisting mode (evect2). This is also seen in the graphs on the right in each panel of Figure 5 where the x -axis represents the extent of the twisting mode. In relating the twisting mode to the cleft opening/closing, an anti-correlation can be seen for at least six simulations (*apo* simulation,¹ HLT3water^{8–10} simulations, and two of the HLT2water simulations^{5,7}—red and orange traces). The twisting (second eigenvector) is seen to be at its extreme whenever the protein is in the “open” conformation (first eigenvector).

In general, the most open structures in the simulations are even more widely open than the *SagHyal* and the *holo*-*SpnHyal*-OPEN crystal structures for as many as six simulations (sim. nos. 1, 5, 6, 8, 9, 11). This is a remarkable result as all simulations started from the closed X-ray structure. The extent of the principal modes during the simulations was quantified as described earlier¹⁶ and is summarized in Figure 6. The maximal amplitude of the opening of the cleft of *apo*-*SpnHyal* is observed to be ~ 13.0 Å. This was measured as the difference between the C^α — C^α separations of Asn231 and Gly769 in the most open and least open trajectory snapshots. The residues were chosen because they belong to different domains (α and β) and are close to the catalytic site. For comparison, this distance difference between the “closed” and “open” crystallographic conformations was observed to be 7.3 Å.

In contrast, both simulations of the *holo*-*SpnHyal*-CLOSED system (sim. nos. 2 and 3) failed to sample the more open configurations to the same extent as the other simulations within the same time span of 100 ns, suggesting that the substrate in the catalytic position locks the protein in a more closed conformation. Here, the extent of the cleft-opening was observed to be only up to 6.9 and 6.2 Å, respectively (see Fig. 6). This observation was further supported by another simulation that had the *holo*-X-ray structure as a starting point (HLT4water3 simulation (sim. no. 13)), in which the opening was found to be similar (6.5 Å). All other simulations (except sim. no. 4, Fig. 6) show varying but generally larger amplitudes along this mode. These results suggest that the decreased flexibility of the protein in the presence of the substrate is caused by HA in the catalytic position. Together with the results obtained with HA in different positions as described earlier, this suggests that the *SpnHyal* yields a differential response to the HA-substrate depending on the position of the HA-substrate inside the cleft, possibly providing first hints of the binding mechanism of the *SpnHyal*-HA complex required for the processivity of this enzyme.

The second eigenvector represents a twisting mode of the two domains with respect to the cleft. This mode has therefore been speculated to play an important role in the effective translocation of the substrate along the cleft. As for the opening/closing mode, the motions observed in the simulations along this mode agree closely with the crystallographic structural differences. In the simulation of *apo-SpnHyal*, the α -domain is observed to be shifted by up to about 11 Å with respect to the β -domain because of this twisting mode. This extent was measured as the separation between the C $^{\alpha}$ atoms of Asp340 of the superimposed most twisted and least twisted snapshots from the trajectory. The residue was chosen to represent the maximum twisting of the α -domain. Strikingly, 11 Å is also approximately the length of a disaccharide unit of the HA-substrate.

Only one simulation from the HLT2water system (sim. no. 5) shows a twisting of the α -domain up to the same extent [compare red trace in Fig. 5(B)]. The simulations for both *holo* systems do not show as much twisting motion as the *apo*-simulation, providing another indication for the interplay between HA presence on one hand and position and the Hyal domain motions on the other.

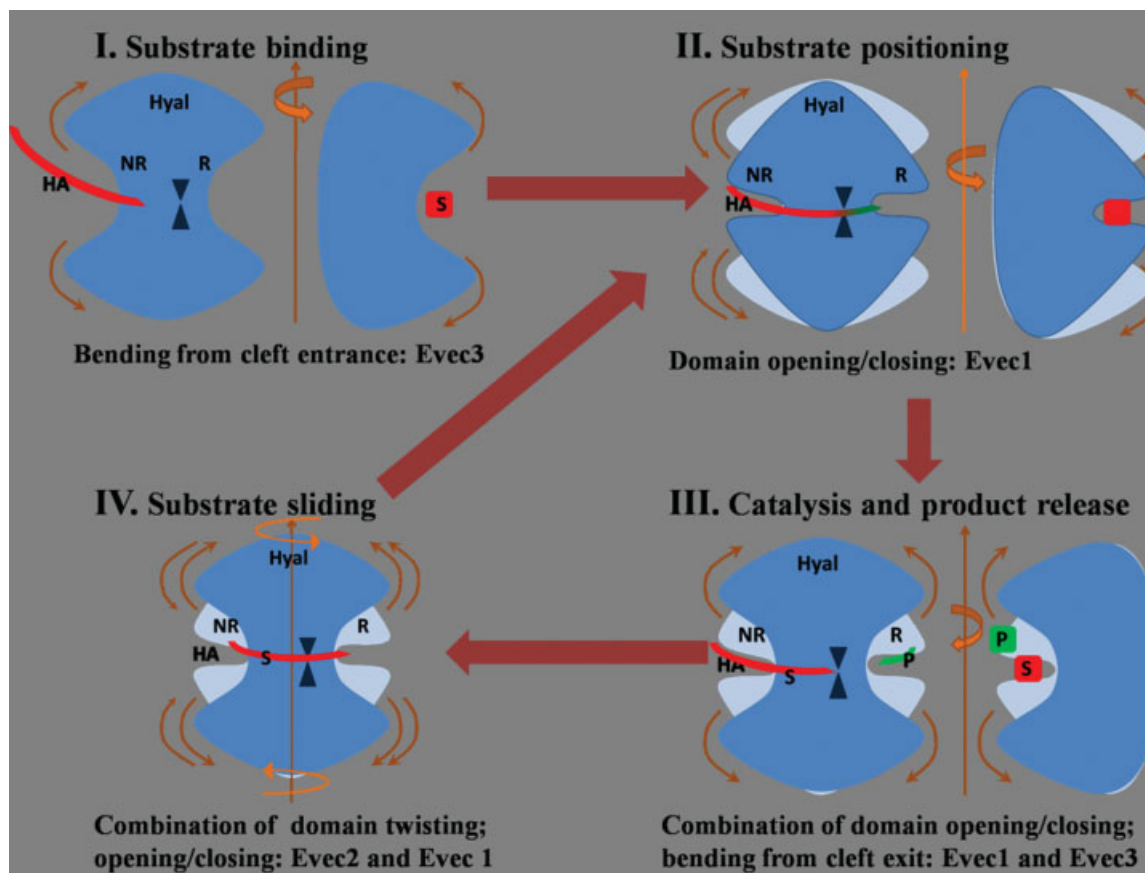
The third eigenvector from the PCA of the crystallographic structures is also the third-largest amplitude mode resulting from the MD simulations and corresponds to a bending of the two domains that influences the width of the cleft entrance/exit. Along this mode, the *holo-SpnHyal-OPEN* structure and *SagHyal* structures represent a conformation with the entrance opened and the exit closed, whereas the *holo-SpnHyal-CLOSED* represents the conformation with the entrance closed and the exit opened. The MD simulations demonstrate that this motion is sampled to the same extent on both sides of the *holo-SpnHyal-OPEN* structure and *SagHyal* structures, thereby enabling an opening the cleft entrance as well as an opening of the cleft exit. The amplitude of this motion was measured in both directions. The opening of the cleft entrance was measured as the difference between the C $^{\alpha}$ —C $^{\alpha}$ separations of Asp211 and Ser771 in the trajectory snapshots with the most open and least open cleft entrances, whereas the opening of the cleft exit was measured as the difference between the C $^{\alpha}$ —C $^{\alpha}$ separations of Asn341 and Asn580 for the same snapshots.

These pairs of residues were observed to be close to the HA-substrate in the *holo*-X-ray crystal structure (1LOH) while belonging to the two different domains (α and β). The opening of the cleft-entrance was found to be maximal when the complete length of the HA-substrate was inside the cleft (i.e., for tetrasaccharide and hexasaccharide). As the length of the cleft is long enough to fit about three disaccharide units of HA (a hexasaccharide substrate), the last disaccharide unit of HA in the HLT4water systems lies outside the cleft. In these cases (sim. nos. 11 and 12), the opening of the cleft access was

found to be minimal for the systems in a postcatalytic configuration in which HA is ready for sliding.

For the simulations in which the HA position is similar to the *holo*-X-ray structure (sim. nos. 2, 3, and 13), this opening was found to be even smaller. These observations suggest the binding mechanism of *SpnHyal* to HA during the catalysis process. The maximal opening of the cleft-exit was found to be correlated to the maximal closing of the entrance site with few exceptions. The cleft-exit was found to be maximally opened (9.8 Å) in the simulation of *apo-SpnHyal*, whereas for the rest of the simulations, it was seen to vary from 0.5 (sim. no. 5) to 9.3 Å (sim. no. 10). Thus, this motion is also significantly more explored in the *apo-SpnHyal* simulation compared with the simulations of Hyals with HA-substrate. Interestingly, for the last simulation (sim. no. 13) with an octasaccharide substrate in a position corresponding to the *holo*-X-ray structure, the opening of the cleft access is minimal. Comparing (a) the opening of the cleft in case of two *holo* simulations and three simulations where a hexasaccharide is ready for sliding, and (b) the opening of the cleft in a simulation of an octasaccharide in a position corresponding to the *holo*-X-ray conformation (sim. no. 13) and two other simulations with an octasaccharide substrate in a postcatalytic position, the cleft access opens. Once the substrate is in the position for the next step of catalytic action, the cleft access closes. This indicates a putative direct functional role in the processive mechanism of the motion described by the third eigenvector as guiding the substrate-binding after sliding.

In summary, the simulations presented here indicate a direct relationship between the domain modes of the enzyme and the presence and position of the HA-substrate. This coupling strongly suggests that the protein domain motions play an essential role in the processivity mechanism. Although the exact sequence of the processes involved in the processive mechanism is not known at present, the domain motion analyses would be consistent with the following putative mechanism: a relatively open structure is essential for the HA-substrate to (re)orient itself inside the cleft. The substrate can bind in a proper position for catalysis upon closing the cleft. Once the catalysis step is performed, the exit side of the cleft opens so that the product formed can be released. This step may also be coupled with the first domain motion of the cleft opening so that the binding of the remaining substrate becomes weak enough to slide along the cleft. After the product release, there is enough space inside the cleft for one more disaccharide, and so a combination of the domain motions occurs (opening of the cleft, twisting of the domains, and opening of the entry side enabled by the third domain motion) to facilitate the processive translocation of the HA-substrate by one disaccharide unit. The cycle can then be repeated. This mechanism is depicted schematically in Figure 7.

**Figure 7**

A schematic representation of the putative processive mechanism in *SpnHyal*-HA system and the associated domain motion. Step I consists of the binding of HA-substrate for which opening of the domains as well as bending of the domains from the entry side of the cleft is required. The combination of the opening/closing of the cleft positions the HA-substrate properly along the catalytic residues (step II). After the catalytic action, the domains perform an opening motion again along with the bending motion to release the product (step III). The remaining substrate slides along the cleft with the help of the combination of the opening/closing motion and the twisting motion of the domains (step IV). The cycle is then repeated until the whole HA is exhausted.

Processive motion of the sugar

To investigate in detail whether in any of the simulated cases the sugar spontaneously moved toward a reactive configuration (as in the *holo*-X-ray structure), thereby carrying out one processive cycle of sliding, the motion of the sugar relative to the cleft was monitored. As the residues inside the catalytic cleft constantly move because of the domain rearrangements, a simple distance measure does not serve the purpose. Hence, we projected the simulation trajectory onto the first eigenvector obtained from a PCA carried out on a set of two structures: the *holo*-CLOSED crystal structure and the starting structure of HLT3water1 simulation. The single nonzero eigenvalue eigenvector of this PCA reflects the effective translocation of the sugar toward the *holo*-X-ray structure that would correspond to the translocation in one processive cycle by one disaccharide unit. As can be seen from Figure 8, no significant sliding motion of the sugar was observed

except in the two HLT3water simulations (magenta and orange traces in the uppermost panel A corresponding to sim. nos. 9 and 10, respectively). Both *holo*-*SpnHyal*-CLOSED simulations show remarkable stability in the sugar position. Sim. no. 13, in which an octasaccharide was modeled similar to the position in the *holo*-X-ray structure (i.e., positioned correctly for catalytic action), shows a similar stability to both *holo*-*SpnHyal*-CLOSED simulations. These three simulations indicate that the sugar does not show spontaneous translocation from the catalytic position and might indicate that translocation is more likely to occur after the catalysis is completed. In all simulations with tetra- hexa- or octasaccharide HA-substrate except for sim. nos. 9 and 10 (magenta and orange traces in the uppermost panel), the sugar also remains stable at its original position. Surprisingly, in sim. no. 8 (HLT3water1 simulation, red trace in upper panel A), the sugar moves in the nonreducing direction.

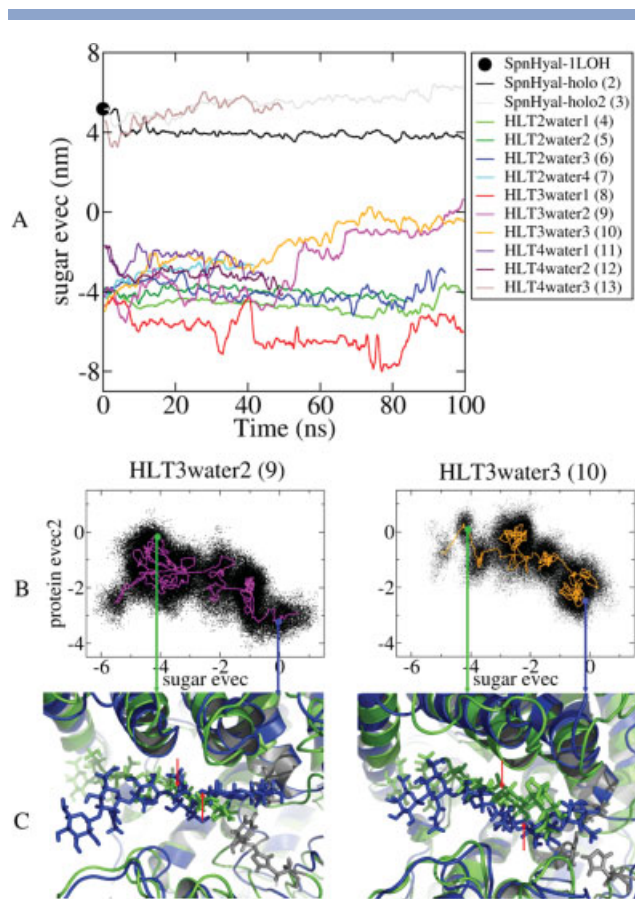


Figure 8

Motion of the sugar inside the cleft as depicted by the PCA performed on the sugar. Panel A (top) shows the motion of the sugar for all the 12 simulations with HA-substrate present inside the cleft. The black circle shows the position of the sugar in the *SpnHyal* structure. In an idealized processive cycle, the simulations ready for the sliding phase (sim. nos. 4–12) should move through the distance of one disaccharide unit and attain the *holo*-X-ray structure (move toward the black circle). Simulations 9 and 10 showed at least half-way translocation of the sugar in the processive direction and the second eigenvector of the enzyme showed a weak correlation with this motion. Panel B shows a detailed picture of this correlation by a scatter plot of the projection of the simulation trajectory onto the second eigenvector of the enzyme (evec2, y -axis) versus the sugar motion from Panel A (sugar evec, x -axis). The correlation is seen stronger in sim. no. 10. The running averages of these scatter plots are shown in orange (sim. no. 9) and pink (sim. no. 10). Panel C shows the overlays of the initial structure of simulations (in green, also denoted by green arrow on the panel B) and the structure at maximum sugar translocation (in blue, also denoted by blue arrow on the panel B) for each of the cases. The catalytic position is shown by the residues in gray stick representation. The red arrows show the position of the next glycosidic β -1,4 oxygen linkage that would be cleaved. The arrows pointing down show the initial position, whereas arrows pointing upward show the final position depicting the movement in the processive direction.

In this simulation, the cleft of the protein was relatively open at the start of the simulations. The backward propagation may imply the reversibility of this sliding step of the mechanism. Interestingly, the other two cases of HLT3water (sim. nos. 9 and 10) spontaneously showed

almost half a processive translocation toward the reducing end (i.e., to be ready for the next catalytic round).

Subsequently, we inspected the role of protein domain motions in relation to the spontaneous partial translocation as observed in sim. nos. 9 and 10. Interestingly, panel B in Figure 8 shows a weak correlation between the positions along the second eigenvector emerging from the PCA of the protein domain motions (twisting mode) with the translocation of the sugar. The decrease in the projection of this eigenvector (α -domain twisting toward exit side of the cleft) is seen to be correlated with the forward motion of the sugar (in the processive (i.e., HA reducing) direction). This finding is in agreement with the proposed role of this mode in the substrate translocation. A comparison of the time evolution of the sugar translocation with the opening mode suggests that the opening of the cleft seems to be a consequence rather than a cause of sugar translocation. Panel C shows the snapshots of the system near the maximum sugar translocation. The catalytic residues are shown. The oxygen from the glycosidic bond of the second sugar ring has to approach the catalytic residues to complete one processive cycle. The position of this oxygen is shown with red arrows in each picture of panel C. From these snapshots, it can be seen that the sugar in simulation 9 and 10 has moved approximately through one sugar ring (i.e., about halfway) toward the catalytic position (*holo*-X-ray structure). However, the lack of translocation by a complete disaccharide unit even on a 100 ns timescale may suggest the requirement of a proper reorientation between the *SpnHyal* domains and the HA-substrate, which occurs on a longer timescale, thereby putatively contributing to the main rate-limiting step in a complete processive cycle.

Two sets of simulations with the HA-substrate in different relative positions were compared to obtain residue-level interactions. The two sets contained sim. nos. 2 and 3 (in which the hexasaccharide HA is in the catalytic position) and sim. nos. 9 and 10 (in which the hexasaccharide HA is shifted by one disaccharide unit toward the nonreducing end). The differences in the interactions with different parts of the HA-substrate confirmed previously identified important residues^{3,9,11–13} and additionally identified residues Asp293, Arg462, Arg466, and Asn468 to form critical interactions with the substrate. It should therefore be interesting to investigate those residues by mutagenesis studies.

Protein flexibility in FPMD simulations

We performed FPMD simulations as well as combined FPMD-EDSAM simulations (Essential Dynamics sampling) to pull the sugar in the forward as well as in the backward direction to further address the interplay between the *SpnHyal* dynamics and the HA translocation. For this, two types of structures were chosen. Simulations were performed both with the “closed” (as in the *holo*-

X-ray structure 1LOH) and “open” structure of the protein. With this structure, the sugar was then pulled under three conditions:

1. Pulling of the sugar in the presumed processive direction (“forward”) with the protein dynamics fully free to relax.
2. Pulling of the sugar in the “forward” direction, while constraining individual PCA eigenvector coordinates of the protein. Here, constraining the first eigenvector resulted in the cleft “closed” at all times (ED1), constraining the second eigenvector did not allow the protein to twist (ED2), and constraining the third eigenvector resulted in keeping the entry/exit of the cleft locked in the original position (ED3).
3. Pulling the sugar backward in the nonprocessive direction such that it emerges from the cleft at the nonreducing end (“backward”). The pulling direction for these simulations was chosen as the vector connecting the center of mass (COM) of the last ring and the COM of the penultimate ring. These simulations were performed to obtain any knowledge on the directionality of the processive mechanism.

The principle of these simulations is depicted in Figure 3(C). In addition to simulations starting from the “closed” crystallographic conformation, simulations were also performed with the “open” structure of the protein. For this, the “open” structure configurations were first obtained by EDSAM simulations that drove the protein to the “open” configuration while keeping the sugar fixed as in the original configuration.

Simulations for the “closed” structure were performed on 10 different configurations, whereas simulations on five different “open” configurations were performed. Figure 9 shows the result for representative cases for both the “closed” and “open” scenarios. Panels A and B show the force curves for the simulations in each scenario. They show the development of the forces on the sugar until it starts moving toward the reducing end at which point the forces suddenly drop to low values. These rupture forces vary in different conditions depending on whether the pulling is performed in the forward or backward direction.

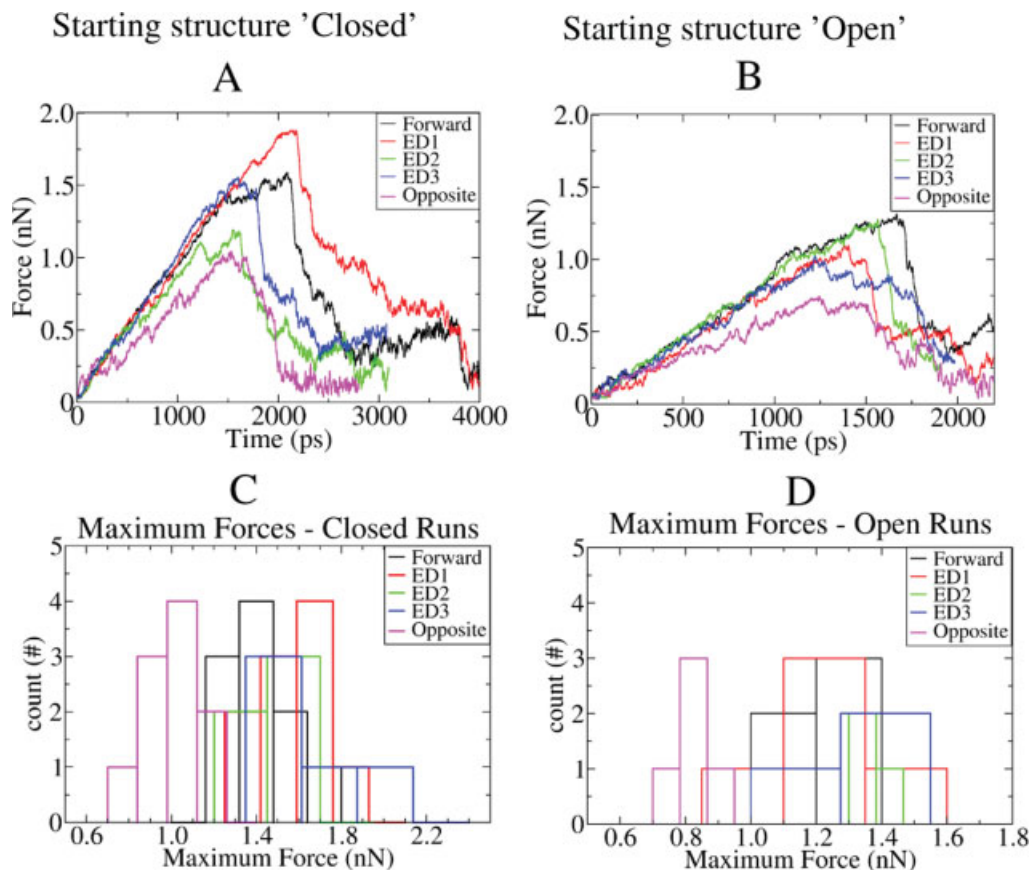
A major difference between the force profiles in panels A and B is the overall lower profile for the simulations starting with the “open” structure compared with those that start from “closed” structures. This shows that the opening of the protein cleft facilitates (and may be necessary for) the translocation of the sugar. This is supported by the FPMD simulations in which the opening mode (evec1) was not constrained, where the protein opens up significantly in all the simulations starting from the “closed” structure. For the simulations starting from the “open” structures, the protein remained open for the duration of the simulations. The opening in the simulations

starting from the closed structure was observed to occur faster than in *apo*-simulations that did not apply an FPMD restraint (free simulations—sim. nos. 1–13). Interestingly, in each scenario, the simulations with the position along the first eigenvector “fixed” also concur with this observation (in Fig. 9, panels A and B, ED1 depicts the forces for the pulling simulations with the opening mode constrained, red curve). For the simulation with the protein cleft kept closed and fixed at all times (A, red curve), larger forces were required for the translocation of the sugar than for the simulations performed by forcing the protein cleft “open” at all times. These results strongly support the functional role of the opening mode in the processive mechanism.

Although the relation between sugar mobility and protein conformation for the opening mode was found to be unambiguous and consistent in all the 13 simulations, the relation with the second eigenmode was observed to be less clear. In panel A of Figure 9, it can be seen that forces required in the simulation with the twisting of the protein constrained (ED2, green curve) were lower than those observed with the protein free (Forward, black curve). This, however, is probably not an indication that the twisting of the protein unfavors the sugar translocation, as in three of the other simulations in the “closed” protein scenario [Fig. 9(C)], the rupture forces for ED2 cases were larger than the “free-forward” simulations (for the others they were comparable).

Constraining the third eigenvector, implying a locked cleft access, resulted in a slight increase in the rupture force (ED3, blue curves). This might be correlated with the fact that the exit site has to open for the sugar translocation and matches the observed projection of the unconstrained trajectories onto this eigenvector (see Supp. Info. Figs. S3 and S5). Strikingly, a systematic increase in the projection onto eigenvector 3 is observed upon pulling the substrate in the putative processive forward direction [Fig. S5(B)], for the first time suggesting a putative role of the cleft entrance opening mode on the sugar translocation, in addition to the presumed role in the product-release mechanism. Thus, the FPMD simulations confirm the mechanical coupling between sugar translocation and protein domain motions.

Despite that a direct role of both domain motions described by eigenvectors 2 and 3 in the processive mechanism was strongly suggested from the unbiased simulations with different ligand occupancies and locations within the cleft as well as from a mechanical coupling between HA and domain motions in Hyal from FPMD simulations described earlier, the FPMD simulations do not reveal an unambiguous effect on the translocation force when these domain motions are constrained. The simulation results allow for a number of explanations for this observation. First, the pulling speeds employed in the FPMD simulations are likely too high to sufficiently allow the slow domain modes to sufficiently equilibrate

**Figure 9**

(A) Force curves for the pulling simulations with the starting structure of the protein “closed”: Forward—sugar pulling with protein free, ED1—sugar pulling with the first eigenvector of the protein kept fixed at all times, ED2—sugar pulling with the second eigenvector of the protein kept fixed at all times, ED3—sugar pulling with the third eigenvector of the protein kept fixed at all times, opposite—sugar pulling in the backward (nonreducing) direction. (B) Similar force curves for the pulling simulations with the starting structure of the protein as “open.” The overall lower force profiles and time-durations required for the translocation of the sugar indicate that opening of the cleft of the protein is favorable (and may be necessary) for the processivity. The lower forces for the backward pulling concur with the presence of a proposed energy barrier in both directions. (C, D): Histograms showing the maximum force rupture in all the pulling simulations discussed. For the simulations starting with “open” configuration, the overall decrease in the rupture forces show that opening of the cleft is favorable (or may be necessary) for the translocation.

in the FPMD simulations, thereby possibly overshadowing the true translocation force profile with high forces that may disappear at longer timescales. In addition, the current statistics, although involving a massive computational effort, may not be sufficient to detect significant differences between the ED1, ED2, ED3, and free sets of FPMD simulations. The longer simulation lengths and the larger variety of substrate lengths and positions probed in the unbiased simulations render the statistical significance of the involvement of domain movements observed in the unbiased simulations higher than those derived from the FPMD simulations.

The dependence of the protein domain motions on the HA presence and position was additionally independently investigated by a set of simulations in which the temperature of the HA-substrate for various lengths (tet-

rasaccharide, hexasaccharide, and octasaccharide) was set to a higher value than the protein and rest of the environment (Supp. Info. Fig. S4). This change in temperature resulted in the corresponding protein domain motions at an enhanced rate compared with the “free” simulations suggesting that modifications in the behavior of the sugar translated to the dynamics of the protein.

An interesting observation is the low rupture forces required to pull the sugar in the nonreducing (backward or nonprocessive) direction (see Fig. 9). This was consistent in all simulations irrespective of whether the protein is in a “closed” or “open” conformation. This might indicate a bidirectional pathway for the sugar to slide in both the forward and backward direction. It is important, however, to note the difference between (pulling) forces and (free) energies. For example, in the hypotheti-

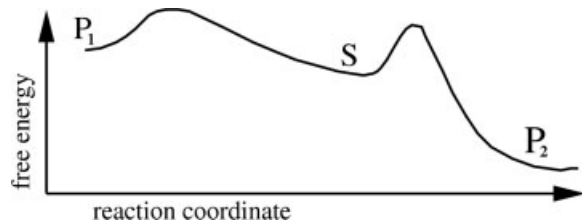


Figure 10

A hypothetical scenario of the energy landscape for the HA translocation compatible with the differences in the force profiles in “forward” and “backward” pulling simulations as observed in Figure 8. Starting from *S*, the overall lower force profile for the pulling in backward direction toward *P*₂ is associated with an energy barrier having a higher slope (larger rupture force) for the processive direction than for the nonprocessive direction toward *P*₁, while *P*₂ is lower in free energy than *P*₁.

cal scenario sketched in Figure 10, if we start at *S*, in an FPMD simulation the maximum force toward *P*₂ would be higher than in the direction of *P*₁, whereas *P*₁ is higher in free energy. The difference is due to the fact that forces are proportional to the slope of the free energy profile. Therefore, these FPMD simulations—even though they provided the first hints on the shape of the landscape—do not allow a direct estimate of the free energy profile underlying processive sugar translocation. In principle, a free energy estimate should be attainable by exploiting the Jarzynski equality.^{39,40} However, the statistics available from the presented simulations do not allow a reliable estimate. Therefore, for an accurate estimate of the free energy profile underlying the complete translocation process, an additional set of simulations using, for example, an umbrella framework is required and is presented in the next section.

Free energy profile for HA translocation

We employed the simulations to obtain a putative free energy profile for the HA translocation (see Theory and Methods). Intriguingly, the convergence was found to be slow, at timescales of tens of nanoseconds, the same time scale known to characterize the protein domain motions. Two sets of umbrella sampling simulations were performed (US1, US2). Free energy profiles computed from different window lengths taken from the US2 set illustrate the slow convergence [see Fig. 11(A)].

The resulting free energy profiles are shown in Figure 11(B). Here, the green curve shows the free energy profile from the umbrella sampling on the first set (for 25 ns simulations). US1 showed convergence at least in the initial parts and the deviation at the right end was found to be due to nonoverlapping of the neighboring umbrellas in that region, posing a problem for the WHAM procedure. The free energy profile obtained in a similar manner for US2 is shown in blue (the first 10 ns) and ma-

genta (for 10–20 ns). Profiles from both sets at this stage, despite not being fully converged, confirm the features obtained from the EDSAM simulations (black curve), indicating a free energy barrier of an estimated 10–20 kJ mol⁻¹. The exact intermolecular interactions underlying this energy barrier are still unknown, but the slow convergence of the profiles strongly suggest an active role of slowly relaxing collective protein degrees of freedom, once again underscoring the role of domain movements

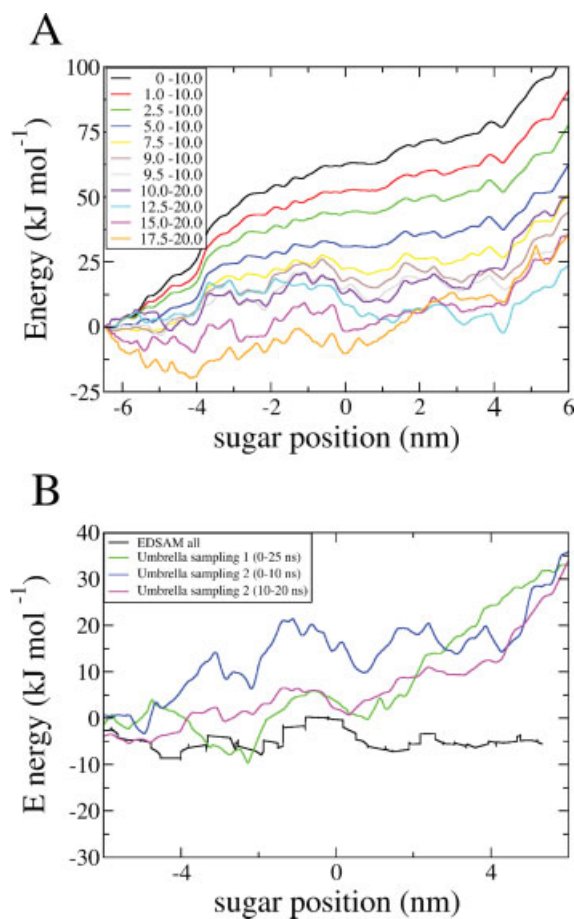


Figure 11

(A) Convergence for the umbrella sampling set US2. The profile converges very fast in the initial periods (for the first 10 ns) and for the next 10 ns it is slowly converging. (B) The umbrella sampling simulations performed for the two sets of runs (run1 and run2) from the earlier EDSAM simulations. The green curve shows the free energy profile obtained from the umbrella sampling for run1, sampling each of the 50 windows for 10 ns. The blue (first 10 ns) and magenta curves (10–20 ns) show similar estimates of the free energy from the simulations of 20 ns simulation of each of the 60 windows. The large deviation at the end is both due to poor overlap of the neighboring umbrellas and the simulations being far from equilibrium (because we drive the system fast to their starting position structure, large forces remain active on the system). The black curve shows the free energy estimation from the simulations of all the three sets of EDSAM simulations (run1, run2, and run3) with the sugar constrained at 21 different positions.

in the processive substrate translocation in the Hyal-HA complex.

Note that the free energy profile depicted here does not include the contribution from the leaving disaccharide product, which presumably entropically aids the process nonreversibly in the processive direction.

CONCLUSIONS

We have carried out MD simulations of the *Spn*Hyal-HA complex in which both the position and length of the HA-substrate inside the cleft were varied. The results strongly support the notion that bacterial Hyals are inherently flexible. A direct interplay between HA sugar substrate translocation and protein domain motions emerges from the simulations that provides a mechanistic picture of how three particular dynamic protein domain modes facilitate processive HA degradation. An opening/closing mode was found responsible for tight binding during catalysis in the closed state while facilitating substrate sliding when opened. The role of the twisting mode appears directly linked to the translocation of the sugar, whereas a third domain mode modulates opening of the cleft entrance and closing of the exit and vice versa. The fact that the simulation structures closely match the crystallographic conformers additionally confirms the validity of the simulation model.

Investigation of the mobility of the sugar inside the cleft showed that even for the ~ 100 ns simulations, a complete processive cycle by one disaccharide unit was not observed. This indicates that the process of sliding that occurs after catalysis is a main bottleneck or even time-limiting factor in the overall kinetics of the enzyme–ligand interaction. More detailed investigations are therefore required to estimate the translocation rate. Nevertheless, it is evident that specific interactions and/or fluctuations in the protein–ligand complex are closely correlated with the substrate movement. The interaction between protein domain modes and substrate translocation was further investigated by a series of FPMD simulations. These simulations independently underscored the direct interplay between HA presence and position on one hand and protein domain motions on the other.

The overall protein–ligand interplay and the active role of the protein domain motions in the processive mechanism as analyzed from the extended “free” or unbiased simulations as well as FPMD simulations were further underscored from the preliminary free energy profile constructed from umbrella and essential dynamics sampling simulations. Strikingly, the computed free energy profiles were found to be characterized by slow convergence, on the same time scale as observed for interdomain movements detected in unbiased simulations. Although not fully converged at the current accessible simulation timescales, the derived profiles provide the

first hints on the shape of the free energy landscape involved in the processive mechanism. Exhaustive simulations performed to obtain an estimate for the underlying free energy profile for the sliding phase of the processive mechanism concur the views of the previous observations, indicating a moderate free energy barrier in the processive direction that is low enough to be easily passed because of thermal energy, and can thus be expected to occur spontaneously. Based on the fact that no spontaneous translocation has been observed so far on the sub-microsecond timescale and the slow convergence of the free energy profile, it is tempting to speculate that coupling to slow collective domain motions poses the rate-limiting step for the translocation part of the processive mechanism.

In summary, it is evident that opening of the cleft of *Spn*Hyal favors (or may be even necessary for) the translocation of the HA-substrate and therefore plays a vital role in the processivity mechanism. The role of twisting and opening of the cleft entrance/exit appears to be directly involved with the translocation and the release of the disaccharide product, respectively.

ACKNOWLEDGMENTS

The authors thank Ira Tremmel and Katja Riebeseel for carefully reading the manuscript.

REFERENCES

- Berry AM, Lock RA, Thomas SM, Rajan DP, Hansman D, Paton JC. Cloning and nucleotide sequence of the *Streptococcus pneumoniae* hyaluronidase gene and purification of the enzyme from recombinant *Escherichia coli*. *Infect Immun* 1994;62:1101–1108.
- Linker A, Meyer K, Hoffman P. The production of unsaturated uro-nides by bacterial hyaluronidases. *J Biol Chem* 1956;219:13–25.
- Li S, Kelly SJ, Lamani E, Ferraroni M, Jedrzejewski MJ. Structural basis of hyaluronan degradation by *Streptococcus pneumoniae* hyaluronate lyase. *EMBO J* 2000;19:1228–1240.
- Li S, Jedrzejewski MJ. Hyaluronan binding and degradation by *Streptococcus agalactiae* hyaluronate lyase. *J Biol Chem* 2001;276:41407–41416.
- Rigden DJ, Jedrzejewski MJ. Genome-based identification of a carbohydrate binding module in *Streptococcus pneumoniae* hyaluronate lyase. *Proteins* 2003;52:203–211.
- Rigden DJ, Galperin MY, Jedrzejewski MJ. Analysis of structure and function of putative surface-exposed proteins encoded in the *Streptococcus pneumoniae* genome: a bioinformatics-based approach to vaccine and drug design. *Crit Rev Biochem Mol Biol* 2003;38:143–168.
- Jedrzejewski MJ, Mewbourne RB, Chantalat L, McPherson DT. Expression and purification of *Streptococcus pneumoniae* hyaluronate lyase from *Escherichia coli*. *Protein Expr Purif* 1998;13:83–89.
- Jedrzejewski MJ, Chantalat L, Mewbourne RB. Crystallization and preliminary x-ray analysis of *Streptococcus pneumoniae* hyaluronate lyase. *J Struct Biol* 1998;121:73–75.
- Jedrzejewski MJ, Mello LV, de Groot BL, Li S. Mechanism of Hyaluronan degradation by *Streptococcus pneumoniae* hyaluronate lyase: Structures of complexes with the substrate. *J Biol Chem* 2002;277:28287–28297.

10. Ponnuraj K, Jedrzejak MJ. Mechanism of hyaluronan binding and degradation: Structure of *Streptococcus pneumoniae* hyaluronate lyase in complex with hyaluronic acid disaccharide at 1.7 Å resolution. *J Mol Biol* 2000;299:885–895.
11. Nukui M, Taylor KB, McPherson DT, Shigenaga MK, Jedrzejak MJ. The function of hydrophobic residues in the catalytic cleft of *Streptococcus pneumoniae* hyaluronate lyase. Kinetic characterization of mutant enzyme forms. *J Biol Chem* 2003;278:3079–3088.
12. Mello LV, de Groot BL, Li S, Jedrzejak MJ. Structure and flexibility of *Streptococcus agalactiae* hyaluronate lyase complex with its substrate: Insight into the mechanism of processive degradation of hyaluronan. *J Biol Chem* 2002;277:36678–36688.
13. Kelly SJ, Taylor KB, Li S, Jedrzejak MJ. Kinetic properties of *Streptococcus pneumoniae* hyaluronate lyase. *Glycobiology* 2001;11:297–304.
14. Jedrzejak M, Chantalat L. Structural studies of *Streptococcus agalactiae* hyaluronate lyase. *Acta Crystallogr D Biol Crystallogr* 2000;56:460–463.
15. Li S, Taylor KB, Kelly SJ, Jedrzejak MJ. Vitamin C inhibits the enzymatic activity of *Streptococcus pneumoniae* hyaluronate lyase. *J Biol Chem* 2001;276: 15125–15130.
16. Rigden DJ, Littlejohn JE, Joshi HV, de Groot BL, Jedrzejak MJ. Alternate structural conformations of *Streptococcus pneumoniae* hyaluronan lyase: insights into enzyme flexibility and underlying molecular mechanism of action. *J Mol Biol* 2006;358:1165–1178.
17. de Groot BL, van Aalten DMF, Scheek RM, Amadei A, Vriend G, Berendsen HJC. Prediction of protein conformational freedom from distance constraints. *Proteins: Struct Funct Genet* 1997;29: 240–251.
18. Rigden DJ, Jedrzejak MJ. Structures of *Streptococcus pneumoniae* hyaluronate lyase in complex with chondroitin and chondroitin sulfate disaccharides. Insights into specificity and mechanism of action. *J Biol Chem* 2003;278:50596–50606.
19. Berendsen HJC, van der Spoel D, van Drunen R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput Phys Commun* 1995;91:43–56.
20. Lindahl E, Hess B, van der Spoel D. Gromacs 3.0: A package for molecular simulation and trajectory analysis. *J Mol Model* 2001;7: 306–317.
21. van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible and free. *J Comput Chem* 2005;26:1701–1718.
22. Vriend G. WHAT IF: a molecular modelling and drug design program. *J Mol Graph* 1990;8:52–56.
23. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–935.
24. Jorgensen WL, Tirado-Rives J. The OPLS potential functions for proteins—energy minimizations for crystals of cyclic peptides and crambin. *J Am Chem Soc* 1988;110:1657–1666.
25. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 2001;105:6474–6487.
26. Berendsen HJC, Postma JPM, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;81: 3684–3690.
27. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. LINCS: A linear constraint solver for molecular simulations. *J Comp Chem* 1997;18: 1463–1472.
28. Darden T, York D, Pedersen L. Particle Mesh Ewald—An N log N method for Ewald sums in large systems. *J Chem Phys* 1993; 98:10089–10092.
29. de Groot BL, Vriend G, Berendsen HJC. Conformational changes in the chaperonin GroEL: New insights into the allosteric mechanism. *J Mol Biol* 1999;286:1241–1249.
30. Tournier AL, Smith JC. Principal components of the protein dynamical transition. *Phys Rev Lett* 2003;91:208106.
31. Grubmüller H, Heymann B, Tavan P. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* 1996;271:997–999.
32. Heymann B, Grubmüller H. AN02/DNP unbinding forces studied by molecular dynamics AFM simulations. *Chem Phys Lett* 1999; 303:1–9.
33. Israelowitz B, Gao M, Schulten K. Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol* 2001;11: 224–230.
34. Gräter F, Shen JH, Jiang HL, Gautel M, Grubmüller H. Mechanically induced titin kinase activation studied by force-probe molecular dynamics simulations. *Biophys J* 2005;88:790–804.
35. de Groot BL, Amadei A, van Aalten DMF, Berendsen HJC. Towards an exhaustive sampling of the configurational spaces of the two forms of the peptide hormone guanylin. *J Biomol Struct Dyn* 1996;13:741–751.
36. Amadei A, Linssen ABM, de Groot BL, Aalten DMF, Berendsen HJC. An efficient method for sampling the essential sub-space of proteins. *J Biomol Struct Dyn* 1996;13:615–626.
37. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. The weighted histogram analysis method for free-energy calculations on biomolecules. I: The method. *J Comp Chem* 1992;13: 1011–1021.
38. Lange OF, Schaefer LV, Grubmüller H. Flooding in GROMACS: Accelerated barrier crossings in molecular dynamics. *J Comp Chem* 2006;27: 1693–1702.
39. Jarzynski C. Nonequilibrium equality for free energy difference. *Phys Rev Lett* 1997;78:2690–2693.
40. Jarzynski C. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Phys Rev E* 1997; 56:5018–5035.