Available online at www.sciencedirect.com
ScienceDirect
Carbohydrate Research 343 (2008) 1023–1033
Carbohydrate RESEARCH

Computational analyses of the conformational itinerary along the reaction pathway of GH94 cellobiose phosphorylase

Shinya Fushinobu,[a,*] Blake Mertz,[b] Anthony D. Hill,[b,†] Masafumi Hidaka,[c] Motomitsu Kitaoka[c] and Peter J. Reilly[b]

[a]*Department of Biotechnology, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan*
[b]*Department of Chemical and Biological Engineering, 2114 Sweeney Hall, Iowa State University, Ames, IA 50011, USA*
[c]*National Food Research Institute, 2-1-12, Kannondai, Tsukuba, Ibaraki 305-8642, Japan*

Received 10 October 2007; received in revised form 25 February 2008; accepted 26 February 2008
Available online 4 March 2008

Abstract—GH94 cellobiose phosphorylase (CBP) catalyzes the phosphorolysis of cellobiose into α-D-glucose 1-phosphate (G1P) and D-glucose with inversion of anomeric configuration. The complex crystal structure of CBP from *Cellvibrio gilvus* had previously been determined; glycerol, glucose, and phosphate are bound to subsites −1, +1, and the anion binding site, respectively. We performed computational analyses to elucidate the conformational itinerary along the reaction pathway of this enzyme. AUTODOCK was used to dock cellobiose with its glycon glucosyl residue in various conformations and with its aglycon glucosyl residue in the low-energy $^4C_1$ conformer. An oxocarbenium ion-like glucose molecule mimicking the transition state was also docked. Based on the clustering analysis, docked energies, and comparison with the crystallographic ligands, we conclude that the reaction proceeds from $^1S_3$ as the pre-transition state conformer (Michaelis complex) via E_3 as the transition state candidate to $^4C_1$ as the G1P product conformer. The predicted reaction pathway of the inverting phosphorylase is similar to that proposed for the first-half glycosylation reaction of retaining cellulases, but is different from those for inverting cellulases. NAMD was used to simulate molecular dynamics of the enzyme. The $^1S_3$ pre-transition state conformer is highly stable compared with other conformers, and a conformational change from $^4C_1$ to $^{1,4}B$ was observed.

Keywords: Cellobiose phosphorylase; Docking; Molecular dynamics; Phosphorolysis; Substrate conformation; Transition state

1. Introduction

The reaction mechanism of β-glycosidases is of interest because distortion of the β-glycoside glycon is thought to be required for efficient catalysis. A conformational change of the scissile glycosidic C-1–O_G bond into a pseudoaxial orientation removes potential steric hindrance from H-1, facilitating direct in-line nucleophilic attack.[1] Moreover, the antiperiplanar lone pair hypothesis (ALPH), a stereoelectronic concept, requires that the scissile bond be antiperiplanar to a lone pair of electrons on the ring oxygen atom to reach the transition state (TS).[2] The C-5, O-5, C-1, and C-2 atoms of the oxocarbenium ion-like TS must be planar because of the partial double-bond character of the O-5–C-1 bond.[3] Different conformational itineraries are adopted by different glycoside hydrolase (GH) families (Fig. 1).[1,4,5] The puckering coordinate system of a six-membered pyranose ring introduced by Cremer and Pople[6] is often used to describe sugar conformations. Spherical mapping of sugar conformations defined by the three coordinates (meridian angle ϕ, azimuthal angle θ, and radius Q) is shown in Figure 2. The 38 basic conformations in the IUPAC nomenclature[7] are used here to specify sugar ring conformations. A pseudorotational map of

* Corresponding author. Tel./fax: +81 3 5841 5151; e-mail: asfushi@mail.ecc.u-tokyo.ac.jp
† Present address: Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA.

0008-6215/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2008.02.026

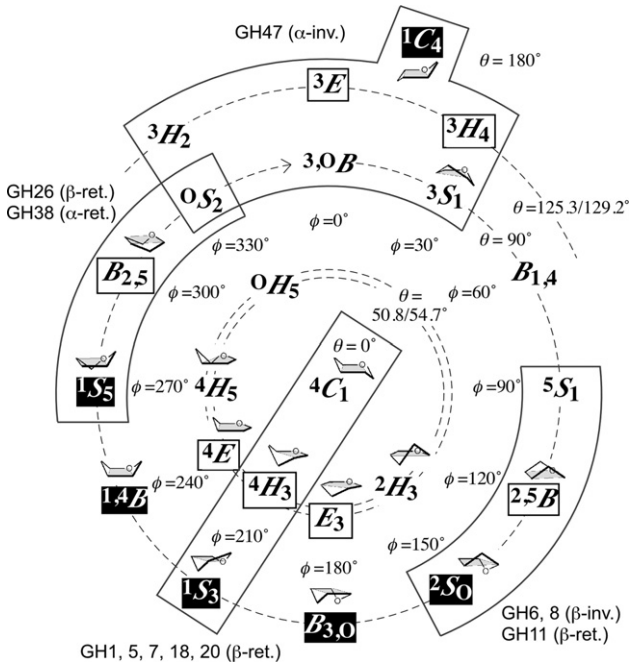

Figure 1. Pseudorotational map of pyranose ring conformations. Proposed reaction itineraries for various GH families (mostly reviewed in Ref. 1), including GH1,[5] GH38,[4] and GH47,[15] are boxed with gray lines. ALPH-compliant conformers of β-glycosides and TS-compliant conformers are highlighted and boxed, respectively. Schematic drawings of sugar rings are shown for the 15 conformers tested in the cellobiose docking part of this study.

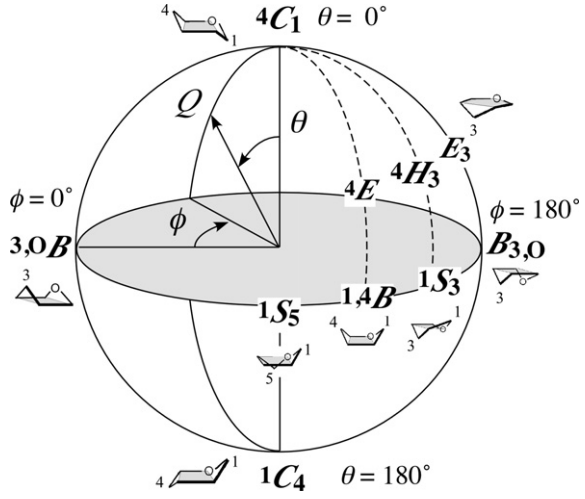

Figure 2. Spherical mapping of pyranose conformations represented by the Cremer–Pople polar coordinates (ϕ, θ, and Q). The closest basic conformation of the IUPAC name is used here to concisely specify sugar ring conformations.

pyranose ring conformations (Fig. 1) corresponds to the projection of the Cremer–Pople sphere from the North Pole.

Sugar phosphorylases catalyze glycosidic bond cleavage by transferring inorganic phosphate instead of water, generating glycosyl phosphates by phosphorolysis.[8] Because the reaction is reversible, phosphorylases can be employed to synthesize sugar chains. Sugar phosphorylases are categorized as glycosyltransferases (GTs) in the EC numbering system (EC 2.4.1.–). Amino acid sequences and three-dimensional structures of sugar phosphorylases indicate that they can be classified as retaining GT-type (GT4 and GT35), retaining GH-type (GH13), and inverting GH-type (GH65, GH94, and GH112) enzymes.[9,10] The crystal structures of two GH94 members (chitobiose phosphorylase, ChBP, EC 2.4.1.–, and cellobiose phosphorylase, CBP, EC 2.4.1.20) have been solved.[11,12] They share a similar $(\alpha/\alpha)_6$ barrel fold and active site structure topology with inverting GH15 glucoamylases. Therefore, it is assumed that the reaction mechanism of these inverting phosphorylases is similar to that of inverting GHs (Fig. 5 in Ref. 10). Enzymatic phosphorolysis begins with direct nucleophilic attack by phosphate on the anomeric C-1 atom, aided by a conserved catalytic acid residue (Asp490 in CBP). Interestingly, GH15 and GH94 enzymes act on α- and β-glycosidic bonds, respectively, so these bonds undergo nucleophilic attack from opposite sides. This mechanistic difference is qualified by the almost flipped glycon sugar in subsite −1 in these enzymes.[12,13] Because GH94 CBP and ChBP are enzymes acting on β-glycosidic bonds, the substrate should be distorted to fulfill steric and stereoelectronic compliance during catalysis. To the best of our knowledge, there are no studies on the conformational itinerary along the reaction pathways of sugar phosphorylases.

We have determined several structures of GH94 enzymes complexed with anions and sugars. Among them are CBP complexed with phosphate and with glycerol in subsite −1 and glucose in subsite +1 (PDB

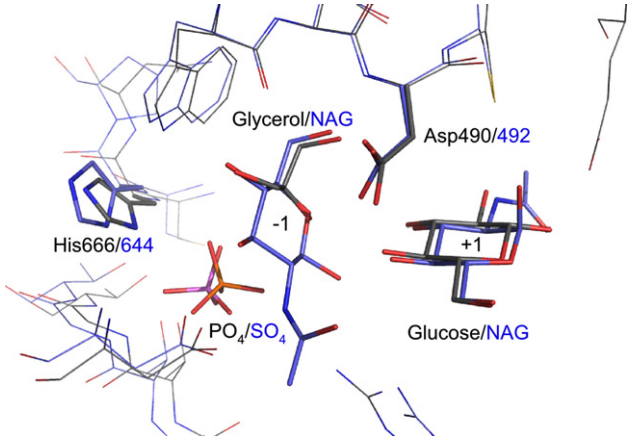

Figure 3. Comparison of the active sites of the GH94 enzymes CBP and ChBP. CBP complexed with glycerol, glucose, and PO_4 (2CQT, carbon atoms in gray), and ChBP complexed with two NAG molecules and SO_4 (1V7X, blue) are superimposed. Phosphorus and sulfur atoms are orange and purple, respectively. Ligand molecules, the catalytic aspartate residue, and the histidine residue forming a hydrogen bond with the anion are shown as sticks. This figure was prepared using PyMol.[38]

2CQT),[12] and ChBP complexed with sulfate and with two *N*-acetylglucosamine molecules (NAGs) in subsites −1 and +1 (PDB 1V7X) (Fig. 3).[11] NAG in ChBP and glucose in CBP almost overlap in subsite +1. The atoms of glycerol in CBP correspond well to the O-4 to O-6 and C-4 to C-6 atoms of NAG bound to ChBP in subsite −1. Therefore, the conformational state of cellobiose (β-D-glucopyranosyl-(1→4)-β-D-glucopyranose) before the phosphorolytic transition state (pre-TS) is postulated to have the same interactions as glycerol in subsite −1. The pre-TS should simultaneously satisfy the requirements for β-glycosidases, for example, a pseudoaxial glycosidic bond and ALPH compliance.

Recently, Mulakala et al. developed a computational docking method to elucidate the conformational pathway of the hydrolytic reaction of GH47 α-(1→2)-mannosidase.[14–16] In the present work, three conformers of oxocarbenium ion-like glucose as well as 15 glycon conformers of cellobiose were docked in the CBP active site, and the most stable and probable pre-TS and TS candidate conformers were determined. In addition, molecular dynamics (MD) was employed to investigate the dynamic aspect of CBP–ligand binding.

2. Computational methods

2.1. Automated docking

Computational docking with AUTODOCK 3.0[17] was performed with methods basically identical to those used earlier,[15] using PCModel (Serena Software, Bloomington, IN) to generate cellobiose and oxocarbenium ion-like glucose models. To produce each glycon conformer, the dihedral angles of atoms in the sugar ring were fixed and the total molecular conformation was minimized using the MM3 force field.[18] For example, the C-2′, C-4′, C-5′, and O-5′ atoms of the $^{1}S_{3}$ conformer of the glycon were fixed in a plane (primes signify glycon atoms), leading to Cremer–Pople parameters of $\phi = 210.7°$, $\theta = 88.5°$, and $Q = 0.72$ Å after MM3 minimization of the sugar ring. Partial ligand charges were generated using GAMESS[19] with the Mulliken charge distribution.[20] The total molecular charges were set to 0 and +1 for cellobiose and oxocarbenium ion-like glucose, respectively. Rotatable ligand bonds (12 in cellobiose, including the glycosidic bond torsional angles, and 5 in oxocarbenium ion-like glucose) were defined using the AUTOTORS module of AUTODOCK.

The crystal structure of CBP complexed with phosphate, glycerol, and glucose (PDB 2CQT)[12] was used for this docking study, and all water molecules were removed before docking. Glycerol and glucose were removed for cellobiose docking and glycerol was removed for oxocarbenium ion-like glucose docking, respectively. Addition of hydrogen atoms and other model corrections prior to docking were performed using the 'What_If' Web server.[21] Partial charges were assigned to protein atoms using all-atom charges of the AMBER force field.[22] Atomic solvation parameters and fragmental atom volumes were added using the ADDSOL program in AUTODOCK. The protonation/charge state of phosphate ion was defined as HPO_4^{2-} based on the optimal enzyme pH (7.6).[23] The Nδ1 atom of His666 forms a hydrogen bond with the phosphate ion.[12] Therefore, this atom and the neighboring phosphate oxygen atom were set to the unprotonated and protonated states, respectively.

Van der Waals and electrostatic energy grid maps were prepared using AUTOGRID 3.0,[17] with $61 \times 41 \times 41$ points spaced at 0.375 Å distances. The grid was centered on the O-4 atom of glucose bound to subsite +1, and it covers the entire binding site for cellobiose and phosphate. During the docking analysis, the conformational space of the ligand is searched, while the protein is assumed to be rigid. This assumption is reasonable because the crystal structure we used is almost the 'full' complex, occupied with both subsites and the anion binding site. For the Lamarckian genetic algorithm (LGA) search, the size of the initial random population was 50 individuals, the maximal number of energy evaluations was 20×10^5, the maximal number of generations was 500, the number of top individuals that survived into the next generation was 1, the rate of mutation was 0.02, the rate of crossover was 0.80, and the average of the worst energy was calculated over a window of 10 generations. The Solis and Wets method was used for local searches. There were a maximum of 15,000 iterations per local search, the size of the local search space for the sample was 1.0, the maximal number of consecutive successes or failures before doubling or halving the step size of the local search was 4, and the lower bound on the step size, 0.01, was the termination criteria for the local search. A total of 1000 dockings were performed in each docking run.

In analyzing the docked conformations, the clustering tolerance of the root mean square deviation (RMSD) was 1.0 Å. The clusters were ranked with the final docked energy (E_{Total}). The E_{Total} value reported by AUTODOCK is a sum of enzyme-interaction energy (E_{Inter}) and the ligand internal energy (E_{Intra}). As described previously, E_{Intra} calculated by AUTODOCK with a simple energy function formulation is insufficient to accurately estimate carbohydrate ligand internal energies.[15] Therefore, values of E_{Inter} with different ring puckering were compared first, and then detailed conformational landscape maps of E_{Intra} values for β-D-glucose based on Car–Parrinello molecular dynamics (CPMD)[24] or the MM3 molecular mechanics algorithm[25] were consulted. These values cannot substitute for AUTODOCK's E_{Intra} values, because scaling schemes are different. However,

the landscape maps give important information about β-D-glucose conformational energies. To obtain the docking result with the lowest E_{Inter} value and most probable docking conformation consistent with the crystallographic data, we selected the most probable optimal cluster according to the cluster size (number of docking results in the cluster) and the RMSD with the crystallographic ligands (glycerol and glucose), and then subjected this cluster to another local search 30 times in succession, each with 200 iterations.[15] The local search result with the lowest E_{Inter} value was selected as the final docking solution.

All docking calculations were run on Dell Precision 650n workstations containing dual 2.66-GHz Xeon processors with 2 GB of RAM running the Red Hat Enterprise Linux 5 operating system.

2.2. Molecular dynamics

MD simulations were run in NAMD 2.5.[26] The entire CBP dimer (PDB 2CQT), which is the native state in solution,[23] was used for MD analysis. The coordinates were checked with the 'What_check' module[27] on the 'What_If' Web server, and side-chains of several His/Glu/Asn residues were flipped to form proper hydrogen bonds. The catalytic acid residue (Asp490) was protonated, because this residue donates a proton during the catalysis. The phosphate ion was set to HPO_4^{2-}, as the assumed pH was 7.6. The glycerol and glucose molecules in PDB 2CQT were removed and the docking result of cellobiose in each conformer was modeled into the active sites in both subunits. Most of the water molecules observed in the crystal structure, except for those located within 5 Å of phosphate and modeled cellobiose, were kept to prepare the MD initial state. Additional water molecules were modeled using the SOLVATE package in the program VMD[28] to form a water box with a 5-Å cushion. The CHARMM22 force field[29] was used for protein parameters, and TIP3P parameters[30] were used for water molecules. The Carbohydrate Solution Force Field[31] was used to obtain cellobiose parameters. Prior to MD calculations, all models were subjected to 10,000 energy minimization steps to relieve geometric strain and close intermolecular contacts. MD simulations were run using periodic boundary conditions, with constant pressure and temperature (1.01325 bar at 310 K) applied using the Nosé–Hoover Langevin piston algorithm.[32,33] The particle-mesh Ewald algorithm was used for calculation of electrostatics. The integration time step was 1 fs. For short MD runs (⩽100 ps), the structure was recorded at 200-step (=0.2 ps in MD) intervals. For long MD runs (⩽2.5 ns), the recording interval was every 2000 steps (=2.0 ps in MD). All MD jobs were run on a 1024-node BlueGene supercomputer, containing dual-core PPC440 CPUs running at 700 MHz with 512 MB of RAM per node.

3. Results and discussion

3.1. Validation of the automated docking procedure with glucose

Our docking method was validated by docking glucose into the CBP active site. A model of β-D-glucose in the $^4C_1$ conformation was constructed and docked into subsite +1 of CBP after its crystallographic glucose molecule, which is also in the $^4C_1$ conformation, was removed. The cluster having conformers with lowest E_{Total} values contained 858 of the 1000 LGA runs. The final docking solution has an E_{Inter} value of −124.5 kcal/mol and an RMSD to the glucose crystallographic coordinates of 0.91 Å for 12 heavy atoms. The O-6 hydroxyl group was docked in a different conformation than that of the crystallographic glucose molecule; the RMSD value without the O-6 atom (11 atoms) is

Table 1. Cluster analysis of 1000 LGA docked $^1S_3$, $^1S_5$, and $^2S_0$ cellobiose glycon conformers

Cluster rank	Number in cluster	E_{Inter} (kcal/mol)		RMSD (Å)	
		Lowest	Mean	Subsite −1[c]	Subsite +1[d]
$^1S_3$					
1[a]	744	−158.3	−133.7	1.00	0.95
2	57	−141.3	−120.1	1.07	1.03
3	10	−130.5	−116.7	0.93	1.01
4	1	−113.6	−113.6	6.48	7.14
5	12	−69.3	−57.5	5.63	5.25
$^1S_5$					
1[b]	5	−130.4	−107.9	6.38	7.00
2	36	−97.2	−66.0	1.59	1.25
3[a,b]	338	−90.7	−63.0	1.03	1.08
4	17	−90.7	−66.4	7.00	6.53
5	77	−82.4	−50.8	1.58	1.08
6	1	−76.1	−76.1	6.53	6.77
7	13	−56.2	−30.4	1.76	1.30
8	3	−55.7	−49.0	1.38	1.12
9[b]	203	−54.2	−21.5	1.35	4.05
10	1	−47.8	−47.8	6.82	7.48
$^2S_0$					
1	16	−163.6	−120.4	3.41	0.93
2[a]	243	−132.8	−91.3	2.03	0.85
3	85	−117.2	−75.1	1.90	1.04
4	40	−113.1	−72.4	2.08	0.89
5	10	−108.1	−88.4	7.39	6.02
6	12	−103.4	−60.4	1.89	0.78
7	14	−100.1	−66.8	1.93	0.66
8	107	−99.1	−20.4	1.90	0.95
9	4	−88.9	−70.8	3.37	4.01
10	10	−77.2	−25.1	2.22	1.32

[a] Selected for subsequent local search.
[b] Shown in Supplementary Figure S1.
[c] RMSD between the crystallographic ligand (glycerol) and the glycon moiety of docked ligand with six atoms (O-4′ to O-6′ and C-4′ to C-6′).
[d] RMSD between crystallographic ligand (glucose) and the aglycon moiety of docked ligand with 12 atoms (O-1 to O-6 and C-1 to C-6).

0.51 Å. This agreement was sufficiently good to proceed with docking other molecules.

3.2. Cellobiose docking

Values of E_{Inter} and RMSDs to crystallographic ligands were compared for various glycon conformations in cellobiose optimally docked by their E_{Total} values. Fifteen conformers were selected (Fig. 1): (1) the pseudorotational series of possible pre-TS ALPH-compliant conformers ($^{2}S_0$, $B_{3,0}$, $^{1}S_3$, $^{1,4}B$, and $^{1}S_5$); (2) conformers adjacent to them ($^{2,5}B$, $B_{2,5}$, $^{2}H_3$, E_3, $^{4}H_3$, ^{4}E, and $^{4}H_5$); (3) the lowest-energy conformer in solution ($^{4}C_1$); and (4) two other low-energy conformers ($^{1}C_4$ and $^{3}S_1$).[25] Here we describe in detail the docking of three possible pre-TS conformers, $^{1}S_3$, $^{1}S_5$, and $^{2}S_0$.

A total of 744 structures from 1000 LGA runs of the $^{1}S_3$ conformer were clustered into the first rank (Table 1). The conformer of lowest E_{Inter} value from this cluster was subjected to local search. The final docked structure has an E_{Inter} value of −166.0 kcal/mol and overlaps the crystallographic ligands well. RMSDs of glucosyl residues at subsites −1 and +1 are 0.96 and 0.89 Å, respectively (Table 2 and Fig. 4A). The docked cellobiose conformation is suitable for in-line nucleophilic attack by phosphate on the anomeric C-1′ atom from behind the glycosidic bond (Fig. 4B). The distance between the nearest oxygen atom of PO_4 and the C-1′ atom is 3.6 Å, and the angle formed by this PO_4 oxygen atom and the C-1′ and O_G atoms of the ligand is almost linear (163°). The distance between the nearest oxygen atom of

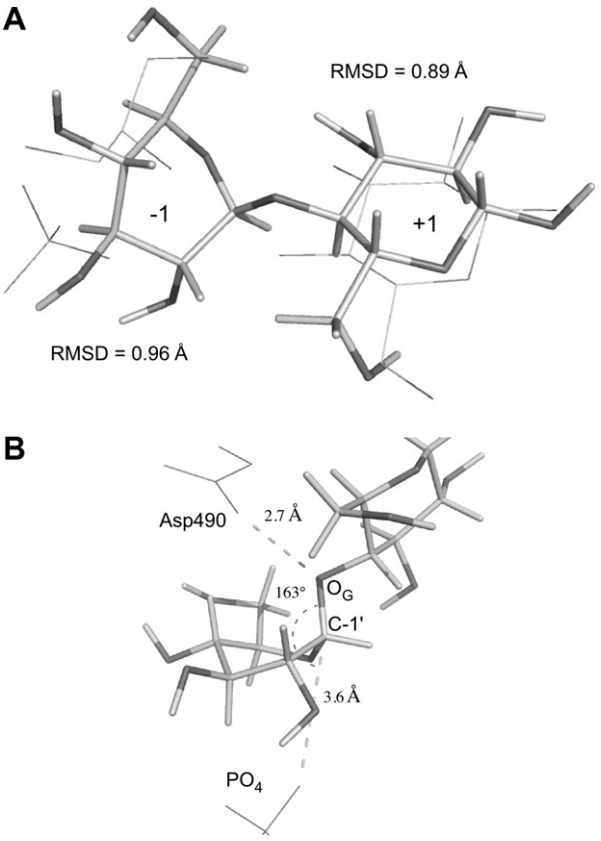

Figure 4. Final docking result of cellobiose with its glycon sugar in the $^{1}S_3$ conformation. (A) Comparison with crystallographic ligands, shown as thin lines. RMSD values in each subsite are indicated. (B) Interaction with catalytic components. Distances and angles with the nearest oxygen atoms of the phosphate group and the catalytic acid residue (Asp490) are shown.

Table 2. Summary of docking results of cellobiose in each conformer after local searches

Conformer[a]	Cremer–Pople parameters[c]			Derived from cluster analysis		Derived from conformers with lowest E_{Inter} values				
	θ (°)	ϕ (°)	Q (Å)	Cluster rank	Number in cluster	E_{Inter} (kcal/mol)	RMSD at subsite −1 (Å)[d]	RMSD at subsite +1 (Å)[e]	PO_4–C-1′ distance (Å)	PO_4–C-1′–O_G angle (°)
$^{4}C_1$	1.8	67.6	0.62	12	114	−118.1	1.75	1.00	4.2	148
$^{2}H_3$	50.3	159.0	0.58	4	155	−132.5	2.17	0.84	4.2	145
E_3	53.1	180.2	0.57	1	752	−149.6	1.87	0.95	4.0	148
$^{4}H_3$	50.6	195.1	0.57	5	201	−122.2	1.23	0.78	3.9	171
^{4}E	55.1	238.8	0.59	1	88	−172.5	1.41	0.91	4.2	153
$^{4}H_5$	52.9	253.2	0.59	1	110	−175.1	1.33	0.79	4.1	153
$^{3}S_1$	82.4	38.2	0.75	12	95	−126.4	2.10	1.10	3.9	130
$^{2,5}B$	80.1	120.7	0.68	9	537	−56.8	2.54	1.18	4.6	162
$^{2}S_0$[b]	85.1	146.9	0.73	2	668	−125.1	1.99	0.85	3.5	165
$B_{3,0}$[b]	70.3	180.1	0.64	2	446	−135.2	1.30	0.88	3.3	169
$^{1}S_3$[b]	88.5	210.7	0.72	1	744	−166.0	0.96	0.89	3.6	163
$^{1,4}B$[b]	85.4	239.8	0.72	1	555	−145.8	1.25	1.19	3.9	155
$^{1}S_5$[b]	89.5	270.3	0.74	3	338	−87.5	1.05	1.06	4.0	153
$B_{2,5}$	88.4	300.1	0.69	3	60	−134.4	1.64	0.91	4.2	124
$^{1}C_4$[b]	177.0	45.6	0.49	2	386	−110.0	1.28	0.91	3.6	164

[a] These conformers are listed according to their placement in Figure 1.
[b] ALPH-compliant conformers.
[c] Cremer–Pople parameters are calculated after MM3 energy minimization using PCModel.
[d] RMSD between the crystallographic ligand (glycerol) and the glycon moiety of docked ligand with six atoms (O-4′ to O-6′ and C-4′ to C-6′).
[e] RMSD between crystallographic ligand (glucose) and the aglycon moiety of docked ligand with 12 atoms (O-1 to O-6 and C-1 to C-6).

Asp490 and the O_G atom is 2.7 Å. These results indicate that the docked $^1S_3$ conformer is compliant as a possible pre-TS conformer.

In contrast to $^1S_3$, the docked structures of the $^1S_5$ conformer are less clustered (Table 1). Structures of lowest E_{Inter} in clusters 1 (5 conformers), 3 (338 conformers), and 9 (203 conformers) are shown in Supplementary Figure S1. The structures in cluster 1 are bound in reversed conformation, while those in cluster 9 have slipped out of subsite +1. RMSD values based on crystallographic ligands at each subsite of >3.0 Å indicate inappropriate docking conformations. Cluster 3 has the most reasonable conformation of the various $^1S_5$ clusters, because it has the largest number of cluster members and the lowest RMSD values. The structure with the lowest E_{Inter} value in cluster 3 was subjected to local search, giving E_{Inter} and RMSD values of the final docked structure (−87.5 kcal/mol, 1.05 Å, 1.06 Å), far worse than those of the $^1S_3$ conformer.

The $^2S_0$ conformer is also an unlikely pre-TS candidate. Structures after LGA docking were less clustered than those of $^1S_3$ and $^1S_5$. Cluster 2, with 243 members, was subjected to local search. The E_{Inter} value of the optimal docked conformer is −125.1 kcal/mol, and the RMSD to the crystallographic ligand at subsite −1, 1.99 Å, is much larger than that of $^1S_3$.

Table 2 summarizes the docking results of 15 conformers after local searches. Among them, $^1S_3$ has among the highest cluster numbers and lowest RMSD values. Moreover, its E_{Inter} value is lowest among all ALPH-compliant conformers. The ALPH-compliant pseudorotational series ranging from $^2S_0$ to $^1S_5$ form a long valley on the MM3 isocontour map of β-D-glucose.[25] Moreover, the free energy minima located at $^1S_3/B_{3,0}$ (M2) and at $B_{3,0}/^2S_0$ (M3) show significantly lower energies (difference: >2.5 kcal/mol) compared with other minima (M4–M9) on the CPMD conformational landscape map.[24] These facts indicate that $^1S_3$ is a relatively stable conformer with low E_{Intra}. The ^4E and $^4H_5$ conformers have more negative E_{Inter} values than that of $^1S_3$, but the numbers of members in their optimal clusters are lower than that of $^1S_3$. Moreover, these conformers are unstable because their E_{Intra} values are about 10 kcal/mol and 3–4 kcal/mol less negative on the CPMD[24] and MM3[25] free energy landscape maps, respectively, than is $^1S_3$. Therefore, docking of cellobiose suggests that its $^1S_3$ conformer is the most probable pre-TS candidate.

The torsional angles of the glycosidic bonds, φ = H-1′–C-1′–O_G (O-4)–C-4 and ψ = C-1′–O_G (O-4)-C-4–H-4, of the docked cellobiose in each conformer are illustrated in Figure 5A. The φ and ψ values range between −20° and 50°, and between −70° and 0°, respectively. This area corresponds to the lowest-energy region of cellobiose on the MM3 isocontour map of cellobiose.[34] Therefore, it is suggested that the glycosidic

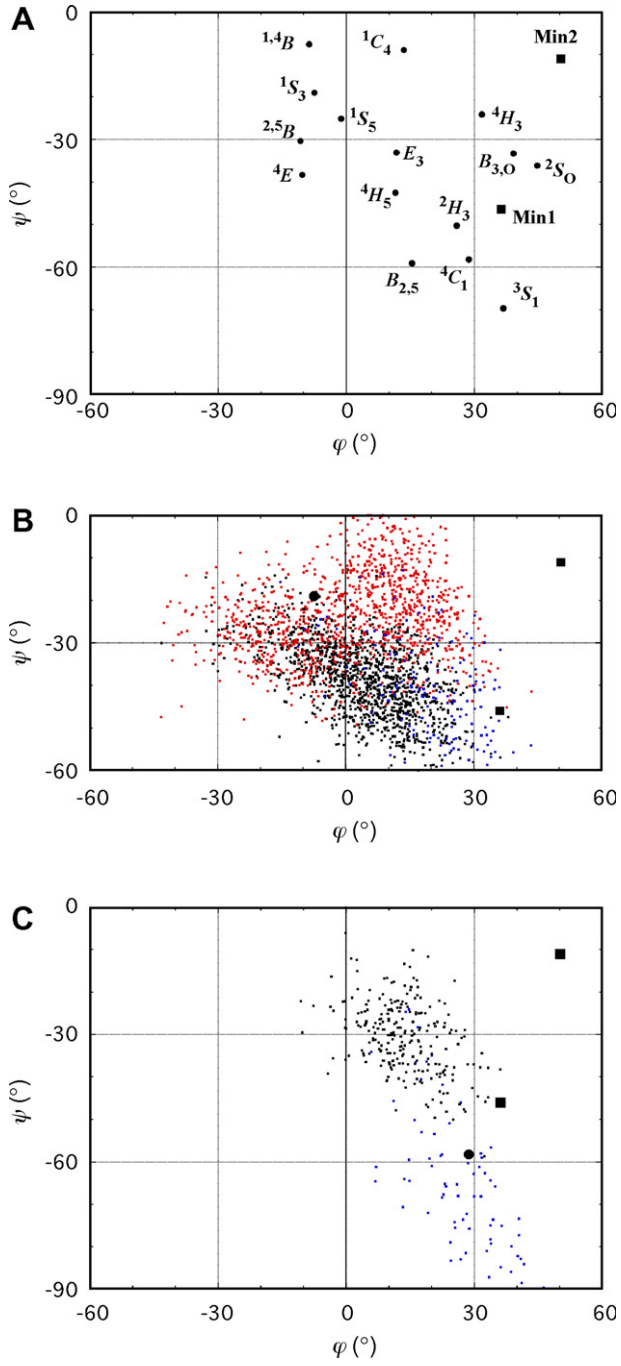

Figure 5. Torsional angles of the glycosidic bond of cellobiose (φ = H-1′–C-1′–O_G (O-4)–C-4 and ψ = C-1′–O_G (O-4)-C-4–H-4) (glycon atoms are primed and aglycon atoms are not primed). The global (Min1; φ, ψ = 36°, −46°) and the second (Min2; φ, ψ = 50°, −11°) minima of the MM3 isocontour map when both glucosyl rings are in the $^4C_1$ conformation[34] are marked in each panel (filled squares). (A) Results of cellobiose docking with various glycon conformers (filled circles). (B) Trajectories of a long (2.5 ns) MD run starting from the $^1S_3$ conformation. The values for cellobiose in the A subunit (0–2.5 ns in red) and B subunit (0–2.2 ns in black and 2.2–2.5 ns in blue) are plotted. The starting point (φ, ψ = −7.5°, −18.9°) is marked by filled circle. (C) Trajectory of a MD run starting from the $^4C_1$ conformation. The values for cellobiose in the B subunit (0–13 ps in blue and 13–50 ps in black) are plotted. The starting point (φ, ψ = 28.7°, −58.2°) is marked by filled circle.

bond of cellobiose bound to CBP is in a relaxed conformation regardless of the glycon sugar ring conformer.

3.3. Oxocarbenium ion-like glucose docking

The reaction product (α-D-glucose 1-phosphate, G1P) is most likely a $^4C_1$ conformer, as the conformation of G1P complexed with another GH94 enzyme, laminaribiose phosphorylase, is $^4C_1$ (Hidaka et al., unpublished results). If so, its C-1–P bond will have an axial configuration. Moreover, β-glycosidases starting from the pre-TS $^1S_3$ conformer are generally thought to produce $^4C_1$ conformers after the TS, because covalent intermediates (2-fluoro- or oxazoline analogs) of retaining GH enzymes (e.g., GH1, GH5, GH18, and GH20) are frequently observed in this conformation.[5,35–37] Therefore, the transition from $^1S_3$ to $^4C_1$ likely occurs via one of the three TS-compliant conformers (^4E, $^4H_3$, or E_3), because the reaction should proceed through an inversion at the anomeric C-1 atom. To investigate which conformer is nearest to the probable TS, we produced oxocarbenium ion-like glucose models in these three candidate conformers, and docked them into subsite −1 of CBP whose crystallographic glycerol molecule in that site was removed. The E_3 conformer had the best values among the three conformers in all the criteria we used (cluster rank, numbers in cluster, and E_{Inter} value) after local searches (Table 3). Notably, the RMSD at subsite −1 of the E_3 conformer is 0.88 Å, smaller than those of all docked cellobiose conformations listed in Table 2. Figure 6 shows the structure of the E_3 conformer of docked oxocarbenium ion-like glucose, as well as its interactions with phosphate and glucose in subsite +1. The distances from the C-1 atom of oxocarbenium ion-like glucose to two key atoms, the nearest oxygen atom of phosphate and the O-4 atom of glucose, are both 2.7 Å. The angles formed by these atoms with the C-1–H bond (PO_4–C-1–H and O-4–C-1–H) are both almost orthogonal (80° and 72°, respectively). Therefore, the docked structure of the E_3 conformer of oxocarbenium ion-like glucose represents the possible TS conformer well.

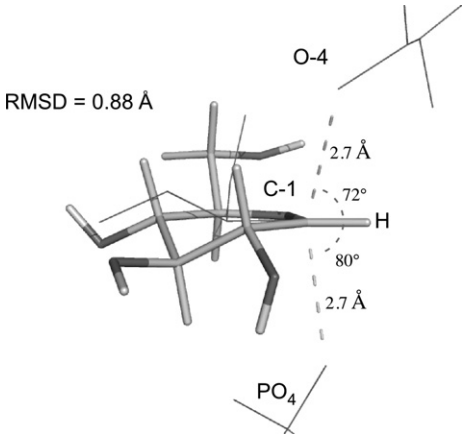

Figure 6. Final docking result of oxocarbenium ion-like glucose in the E_3 conformation. Crystallographic ligands are shown as thin lines, and interactions with them as well as the RMSD value are indicated.

3.4. Molecular dynamics

We performed MD simulation, starting from the complex structures produced by automated docking and using Cremer–Pople parameters as our primary indicators of the changes of sugar conformation. The complex of CBP with the $^1S_3$ conformer, which is inferred to be the pre-TS state, was the starting structure. Because CBP acts as a dimer in solution,[12,23] two chains in the asymmetric unit of the crystal structure were placed in a periodic box filled with water molecules. Following 10,000 energy minimization steps, MD simulation of 100 ps at 1 fs/step (=100,000 steps) was carried out (Supplementary Fig. S2). The ϕ angles of A and B subunits both start at 210.7°, and their averages during the MD run (500 recorded structures) are 217.8° and 216.7°, respectively. The conformation remains between $^1S_3$ and $^{1,4}B$, but nearer the former. The distance between the C-1′ atom of the cellobiose glycon and the nearest oxygen atom of the phosphate starts at 3.6 Å and

Table 3. Summary of docking results of oxocarbenium ion-like glucose in TS-compliant conformers after local searches

Conformer	Cremer–Pople parameters			Derived from cluster analysis		Derived from conformers with lowest E_{Inter} values					
	θ (°)	ϕ (°)	Q (Å)	Cluster rank	Number in cluster	E_{Inter} (kcal/mol)	RMSD at subsite −1 (Å)[a]	PO_4–C-1 distance (Å)	C-1–O-4 distance (Å)	PO_4–C-1–H angle (°)	O4–C-1–H angle (°)
E_3	53.5	180.8	0.51	1	933	−110.3	0.88	2.7	2.7	80	72
$^4H_3$	51.3	226.7	0.52	4	674	−104.0	1.04	2.6	2.6	99	57
^4E	54.0	239.0	0.52	3	720	−109.7	1.13	2.6	2.6	102	57

[a] RMSD between the crystallographic ligand (glycerol) and the docked ligand with six atoms (O-4 to O-6 and C-4 to C-6).

progresses during the MD run to averages of 3.97 and 4.05 Å in A and B subunits, respectively.

Next we performed a longer MD simulation (2.5 ns), again starting from the docked $^1S_3$ conformation (Fig. 7). Here the sugar ring in the B subunit remained in the $^1S_3$ conformation until 2.24 ns, when it changed to the $^1S_5$ conformation. On the other hand, the sugar ring in the A subunit changed to $^1S_5$ at earlier stages (described later). Values of ϕ, the distance between C-1′ and the nearest phosphate oxygen, and the distance between the O_G atom and the nearest carboxylic oxygen of Asp490 (the catalytic acid) during 0–2.0 ns in the B subunit appear in Table 4. The ϕ angle fluctuated between 199.3° and 243.9°, with an average of 217.1° (almost $^1S_3$). The lowest values of PO_4–C-1′ and Asp490–O_G distances were 3.23 and 2.88 Å, respectively. Because this simulation was recorded at every 2000 steps (=2.0 ps interval), these distances might have been less between the recorded values. The glycosidic bond torsional angles fluctuated within a relatively low-energy region during the MD run, but the area is narrower in the B subunit until 2.2 ns (Fig. 5B).

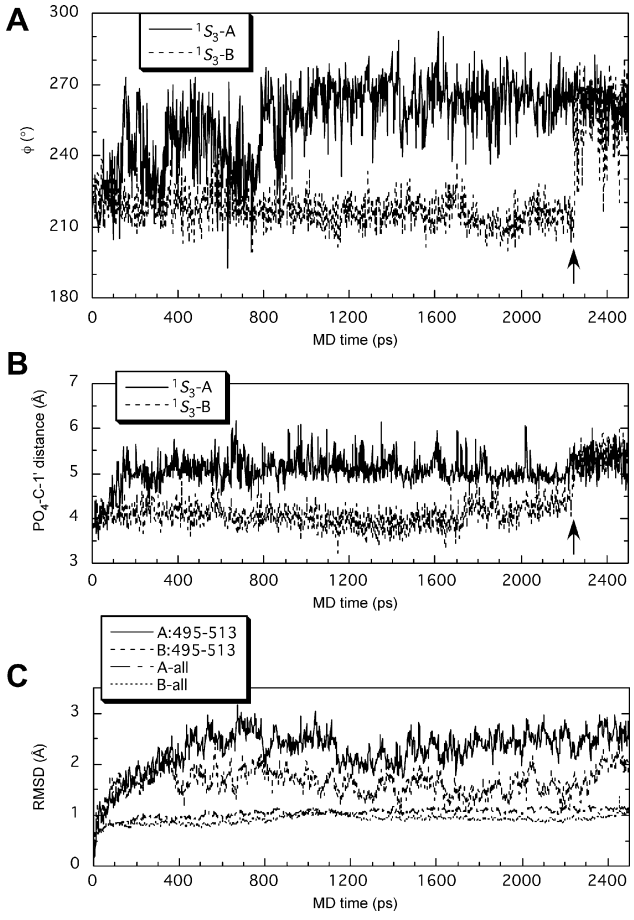

Figure 7. A long MD run (2.5 ns) starting from the $^1S_3$ conformation. (A) ϕ angle; (B) PO_4–C-1′ distance, and (C) RMSD value to the starting structure in each subunit are shown. On panel C, a loop region (495–513) covering the active site is compared with the whole polypeptide chain. Arrows in panels A and B indicate the point of transition from $^1S_3$ ($\phi \sim 210°$) to $^1S_5$ ($\phi \sim 270°$) in the subunit B.

Table 4. Parameter changes during a long MD run[a]

	ϕ (°)	PO_4–C-1′ distance (Å)	Asp490–O_G distance (Å)
Average ± S.D.[b]	217.1 ± 6.9	4.04 ± 0.24	3.80 ± 0.45
Minimum	199.3	3.23	2.88
Maximum	243.9	4.93	5.46

[a] Statistics of MD simulation during 0–2 ns in the chain B active site, starting from $^1S_3$.
[b] Standard deviation.

To summarize these MD results, an ALPH-compliant $^1S_3$ conformation is stable for at least 2 ns with its glycosidic bond in a relaxed conformation, and the key interactions for nucleophilic attack of PO_4 and proton donation from Asp490 are close during that time. However, the average turnover time (about 12 ms) estimated from the k_{cat} of CBP at 310 K[23] is far longer than 2.5 ns, the MD timescale here.

One of possible factors limiting the catalytic turnover of CBP is substrate binding and release. A loop fully covering the active site (495–513) is thought to be flexible due to its relatively high B-factor in the crystal structure, and it may open with substrate binding and release.[12] Figure 7C shows the RMSD values of the 495–513 loop region and the entire polypeptide chain compared with the starting structure during the long (2.5 ns) MD run. As expected, this loop is highly mobile, exhibiting high RMSDs. In this MD run, the cellobiose glycon in the B subunit changed its conformation to $^1S_5$ after 2.24 ns, whereas that in the A subunit changed to the $^1S_5$ conformation at earlier stages, and it was constant after 0.8 ns (Fig. 7A). In the $^1S_5$ state, the cellobiose substrate slipped out from the catalytically correct position in the active site and descended into a non-productive local minimum state. The distance between the phosphate oxygen and the C-1′ atom can be a good indicator for catalytic compliance (Fig. 7B); in the 'slipped out' $^1S_5$ state, this distance is about 5 Å. Departure of cellobiose from the active site seems to be linked with the opening of the 495–513 loop, because the RMSD is larger (>2 Å) in subunit A from the earlier stage, and it suddenly increased after 2.24 ns in subunit B (Fig. 7C).

In the next MD simulations, we started from $^4C_1$ as well as from neighboring conformers of $^1S_3$ ($^{1,4}B$, $^1S_5$, $B_{3,0}$, $^2S_0$, ^4E, $^4H_3$, and E_3), to explore the local energy minima around the conformer with the most favorable docking results. Five starting conformers ($^{1,4}B$, $^1S_5$, $B_{3,0}$, ^4E, and E_3) out of the eight that we tested immediately changed to the $^1S_3$ conformer during the energy minimization stage, and this conformer was stable for at least 20 ps (Supplementary Fig. S3). Therefore, the $^1S_3$ conformer is clearly more stable than neighboring

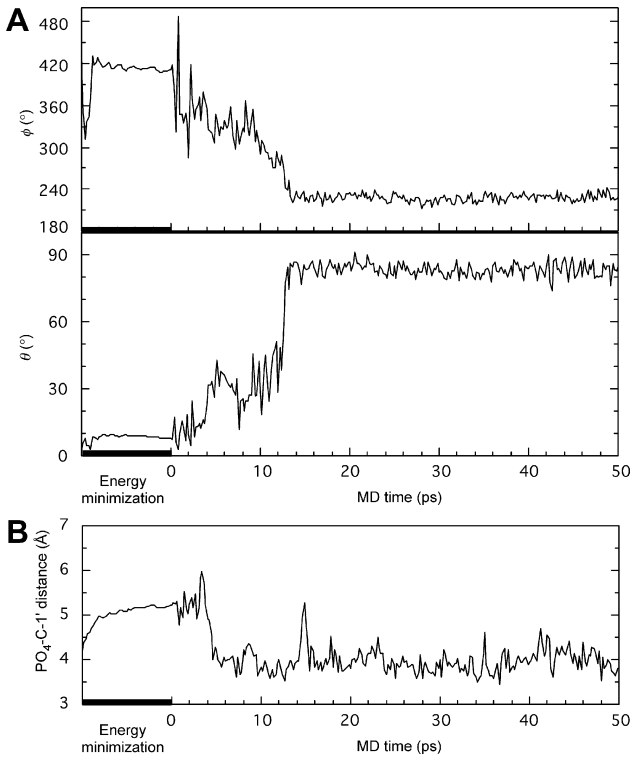

Figure 8. MD run starting from the $^{4}C_{1}$ conformation. (A) Cremer–Pople parameters and (B) PO_4–C-1′ distance in subunit B are shown. Note a conformational change to $^{1,4}B$–$^{1}S_{3}$ before 14 ps.

conformers, agreeing with the cellobiose docking results. In the MD simulation starting from the $^{2}S_{0}$ conformer, however, the sugar conformation did not change (data not shown). The PO_4–C-1′ distance immediately increased from 3.5 Å to about 5 Å during the energy minimization stage, indicating that this $^{2}S_{0}$ state is also a non-productive local minimum. From the MD simulation starting from the $^{4}H_{3}$ and $^{4}C_{1}$ conformers, a similar non-productive local minimum state was present around the $^{4}C_{1}$ conformation. Both subunits in the simulation starting from $^{4}H_{3}$, as well as subunit A starting from $^{4}C_{1}$, descended into this non-productive local minimum state, θ being <30° and the PO_4–C-1′ distance being about 6.0 Å (data not shown).

Interestingly, the $^{4}C_{1}$ conformer of subunit B quickly changed starting at 2–3 ps via the $^{0}H_{5}$-like conformer ($\phi \sim 300°$ and $\theta \sim 40°$) to a $^{1,4}B/^{1}S_{3}$ conformer, completing the transition at about 13 ps (Figs. 8A and 9). After the change to $^{1,4}B/^{1}S_{3}$ conformer, the PO_4–C-1′ distance stayed at about 4 Å (Fig. 8B), and the glycosidic bond torsional angles stayed in a narrow low-energy region (Fig. 5C). The change from $^{4}C_{1}$ to $^{1,4}B/^{1}S_{3}$ indicates that the former conformation is a quasi-stable (less stable) state when in the CBP active site. Because both sugar rings of cellobiose are expected to be $^{4}C_{1}$ conformers in solution, such conformational change to a catalytically compliant state may occur after substrate penetration, with opening of the loop shielding the active site.

4. Conclusions

We report here automated docking and MD analyses to reveal the conformational itinerary along the reaction pathway of GH94 CBP. We conclude that the reaction proceeds through the following steps: (1) the 495–513 mobile loop opens before binding of substrate, phosphate, and cellobiose in random order, because this enzyme exhibits a random-ordered Bi Bi mechanism;[12] (2) the cellobiose glycon changes its conformation to the pre-TS $^{1}S_{3}$ state, possibly via $^{0}H_{5}$ and $^{1,4}B$; (3) nucleophilic attack of phosphate oxygen to C-1′ atom and proton donation from Asp490 to the O_G atom leads to the E_3 TS candidate state; (4) after anomeric inversion, product G1P is in a $^{4}C_{1}$ conformation; and (5) opening of the 495–513 loop enables ordered product release, glucose first, then G1P.[23] A clear path with a lower energy difference between $^{1}S_{3}/B_{3,0}$ and $^{4}C_{1}$ via $E_3/^{4}H_{3}$ is present on the CPMD free energy landscape map for β-D-glucose,[24] indicating that conformational

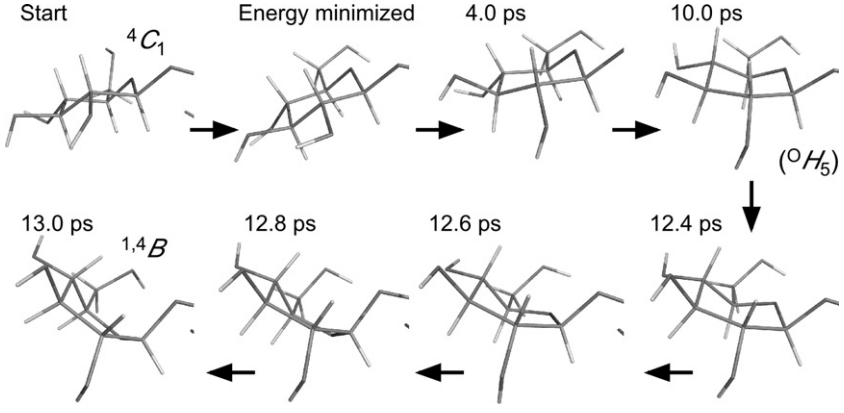

Figure 9. A conformational change to $^{1,4}B$–$^{1}S_{3}$ observed in subunit B of a MD run starting from $^{4}C_{1}$.

change from $^{1}S_3/B_{3,0}$ to $E_3/^{4}H_3$ is energetically favored for the substrate. Although we showed that the subsite −1 of CBP is suitable for binding the E_3-like conformer of the oxocarbenium ion-like glucose, detailed reaction mechanism of the phosphorolysis step is still to be elucidated (e.g., S_N2 vs S_N1).

The conformational itinerary of an inverting phosphorylase shown above ($^{4}C_1$ ground-state glycon, $^{1}S_3/^{1,4}B$ pre-TS, $E_3/^{4}H_3$ TS candidate, and $^{4}C_1$ ground-state product) is very similar to those proposed for the first-half glycosylation reactions of many retaining β-glycoside hydrolases, including GH1, GH5, GH7, GH18, and GH20 enzymes.[1] Interestingly, two inverting GH enzymes acting on β-glucosidic bonds (GH6 and GH8) are thought to adopt the distinct TS conformation near to $^{2,5}B$. Most of the current insights about the conformational itinerary of GH reactions are based on X-ray crystallography and inhibition experiments using mutant enzymes and/or substrate analogs, to trap a labile conformational state of the sugar ring. Computational analysis can complement these commonly-used methods because the conformations of the wild-type enzyme and canonical substrate are subjected to calculation. In this work, we combined MD analysis with previously established docking analysis to illustrate the dynamic motions of protein and substrate.

Acknowledgments

We thank Professor Arthur Olson (Scripps Research Institute) for donating AUTODOCK, and we thank Iowa State University for supporting our computational work. This work was supported by the Program for Promotion of Basic Research Activities for Innovative Biosciences (PROBRAIN), the Ministry of Education, Culture, Sports, Science and Technology (MEXT), and the US National Science Foundation.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.carres.2008.02.026.

References

1. Davies, G. J.; Ducros, V. M.-A.; Varrot, A.; Zechel, D. L. *Biochem. Soc. Trans.* **2003**, *31*, 523–527.
2. Deslongchamps, P. *Pure Appl. Chem.* **1993**, *65*, 1161–1178.
3. Sinnott, M. L. *Chem. Rev.* **1990**, *90*, 1171–1202.
4. Numao, S.; Kuntz, D. A.; Withers, S. G.; Rose, D. R. *J. Biol. Chem.* **2003**, *278*, 48074–48083.
5. Gloster, T. M.; Roberts, S.; Ducros, V. M.; Perugino, G.; Rossi, M.; Hoos, R.; Moracci, M.; Vasella, A.; Davies, G. J. *Biochemistry* **2004**, *43*, 6101–6109.
6. Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.
7. Schwarz, J. C. P. *J. Chem. Soc., Chem. Commun.* **1973**, 505–508.
8. Kitaoka, M.; Hayashi, K. *Trends Glycosci. Glycotechnol.* **2002**, *14*, 35–50.
9. Coutinho, P. M.; Henrissat, B. 1999. http://www.cazy.org/.
10. Fushinobu, S.; Hidaka, M.; Miyanaga, A.; Imamura, H. *J. Appl. Glycosci.* **2007**, *54*, 95–102.
11. Hidaka, M.; Honda, Y.; Kitaoka, M.; Nirasawa, S.; Hayashi, K.; Wakagi, T.; Shoun, H.; Fushinobu, S. *Structure* **2004**, *12*, 937–947.
12. Hidaka, M.; Kitaoka, M.; Hayashi, K.; Wakagi, T.; Shoun, H.; Fushinobu, S. *Biochem. J.* **2006**, *398*, 37–43.
13. Stam, M. R.; Blanc, E.; Coutinho, P. M.; Henrissat, B. *Carbohydr. Res.* **2005**, *340*, 2728–2734.
14. Mulakala, C.; Reilly, P. J. *Proteins* **2002**, *49*, 125–134.
15. Mulakala, C.; Nerinckx, W.; Reilly, P. J. *Carbohydr. Res.* **2006**, *341*, 2233–2245.
16. Mulakala, C.; Nerinckx, W.; Reilly, P. J. *Carbohydr. Res.* **2007**, *342*, 163–169.
17. Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
18. Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8565.
19. Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Gordon, M. S.; Nguyen, K. A.; Su, S.; Windus, T. L.; Elbert, S. T.; Montgomery, J.; Dupuis, M. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
20. Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833–1840.
21. Rodriguez, R.; Chinea, G.; Lopez, N.; Pons, T.; Vriend, G. *CABIOS* **1998**, *14*, 523–528.
22. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M. J.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
23. Kitaoka, M.; Sasaki, T.; Taniguchi, H. *Biosci. Biotechnol. Biochem.* **1992**, *56*, 652–655.
24. Biarnés, X.; Ardèvol, A.; Planas, A.; Rovira, C.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2007**, *129*, 10686–10693.
25. Dowd, M. K.; French, A. D.; Reilly, P. J. *Carbohydr. Res.* **1994**, *264*, 1–19.
26. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
27. Hooft, R. W. W.; Vriend, G.; Sander, C.; Abola, E. E. *Nature* **1996**, *381*, 272.
28. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph.* **1996**, *14*, 33–38.
29. MacKerell, J. A. D.; Bashford, D.; Bellot, M.; Dunbrack, J. R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, I. W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
30. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
31. Kuttel, M.; Brady, J. W.; Naidoo, K. J. *J. Comput. Chem.* **2002**, *23*, 1236–1243.

32. Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.
33. Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
34. Dowd, M. K.; French, A. D.; Reilly, P. J. *Carbohydr. Res.* **1992**, *233*, 15–34.
35. Davies, G. J.; Mackenzie, L.; Varrot, A.; Dauter, M.; Brzozowski, A. M.; Schülein, M.; Withers, S. G. *Biochemistry* **1998**, *37*, 11707–11713.
36. Mark, B. L.; Vocadlo, D. J.; Knapp, S.; Triggs-Raine, B. L.; Withers, S. G.; James, M. N. *J. Biol. Chem.* **2001**, *276*, 10330–10337.
37. van Aalten, D. M.; Komander, D.; Synstad, B.; Gaseidnes, S.; Peter, M. G.; Eijsink, V. G. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 8979–8984.
38. DeLano, W. L. The PyMOL Molecular Graphics System, 2002. http://www.pymol.org.