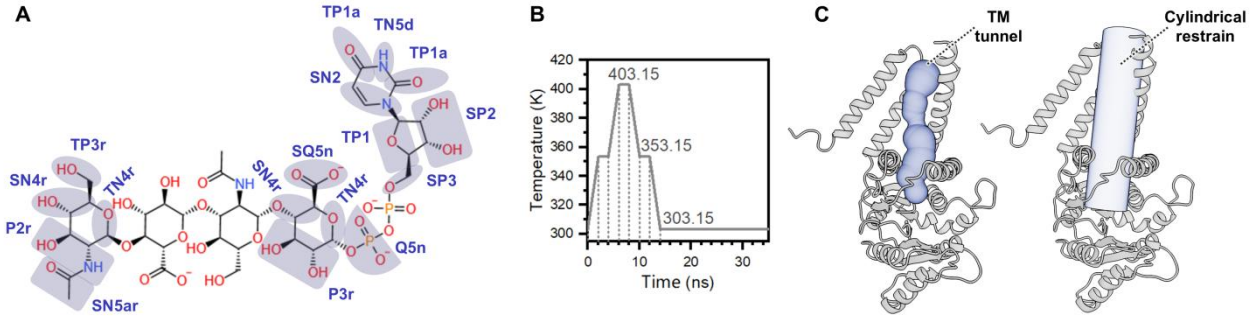
Supplementary Information for

An intramolecular energy metabolism fueling the chain elongation-translocation cycle in hyaluronic acid synthesis

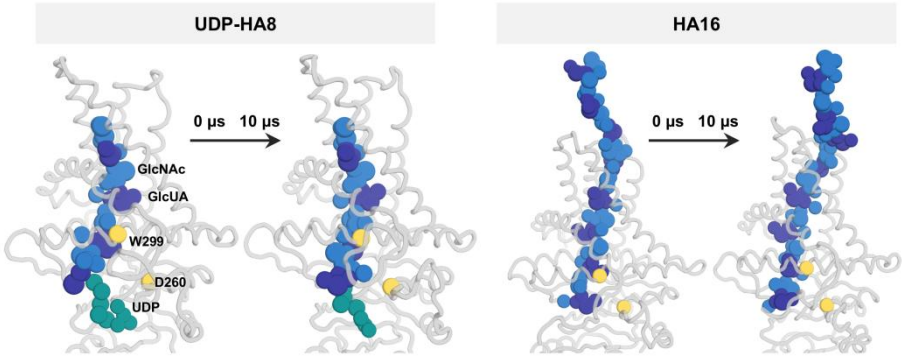
Kaiyi Zhu[1], Yupei Jian[1], Yiyao Dai[1], Guoqiang Jiang[1], Diannan Lu[1], Zheng Liu[1]*

[1] Key Lab of Industrial Biocatalysis, Ministry of Education, Department of Chemical Engineering, Tsinghua University, Beijing 100084, China

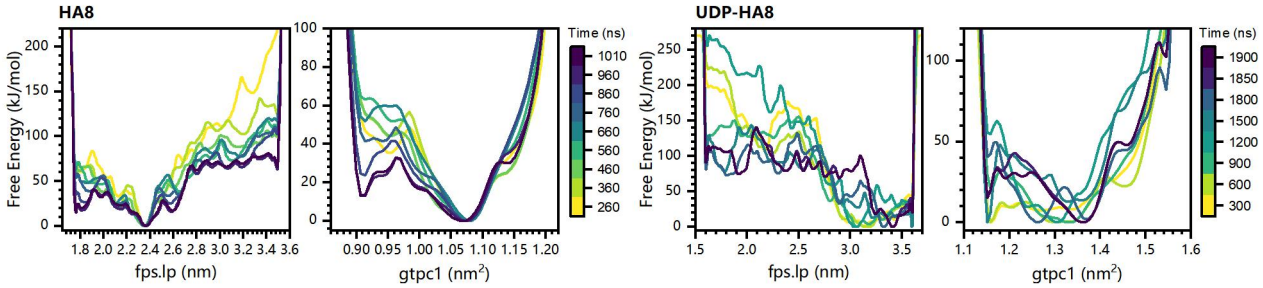
Contents



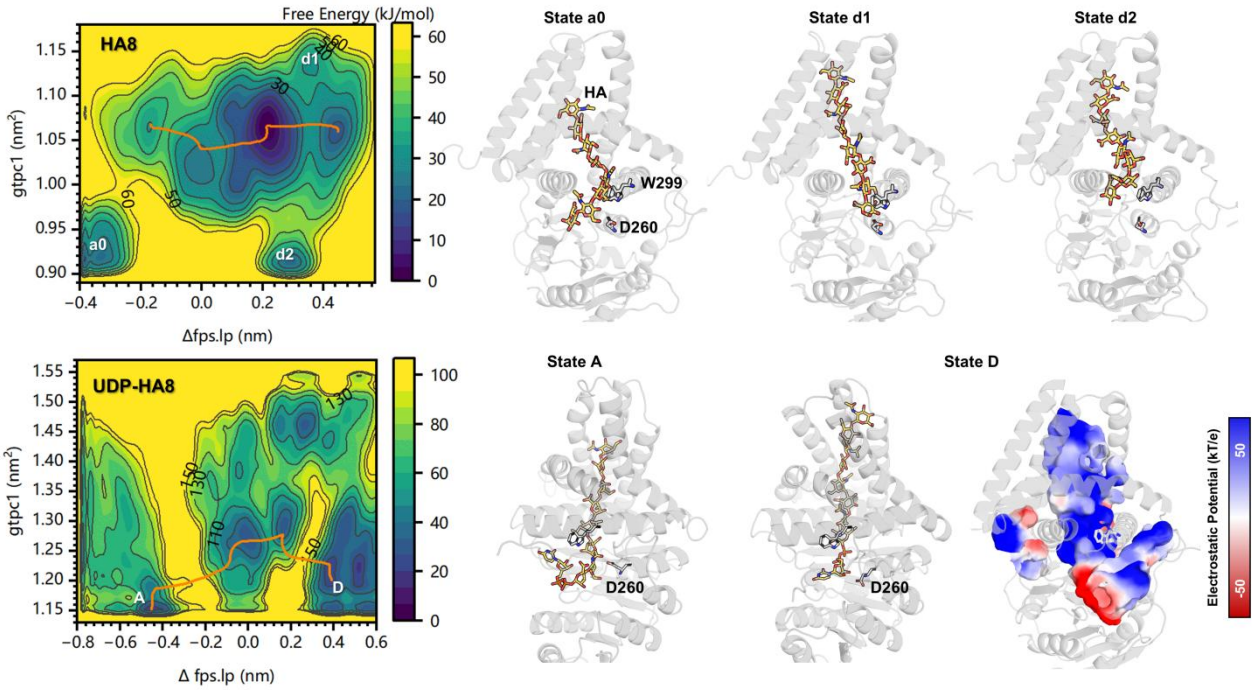
Supplementary Figure 1 MD simulation setup. (A) Coarse-grained mapping of UDP-HA. (B) Simulated-annealing procedure. (C) A cylindrical restrain (right) mimicking the TM tunnel identified by CAVER (left).



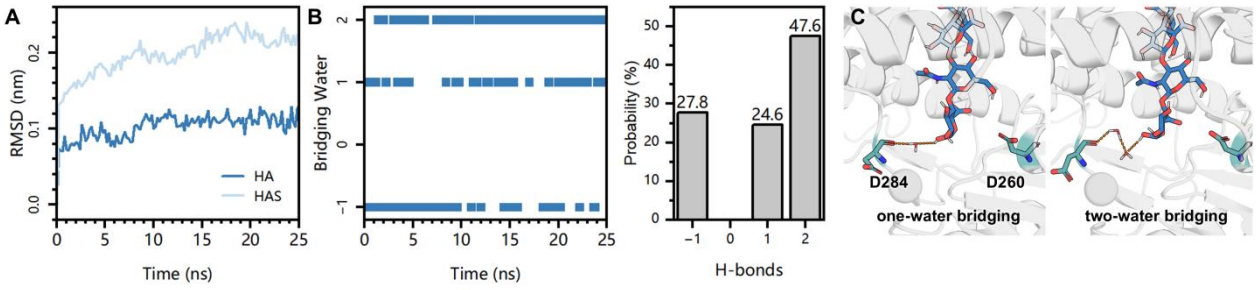
Supplementary Figure 2 HA translocation during 10 μs CG-MD simulations.



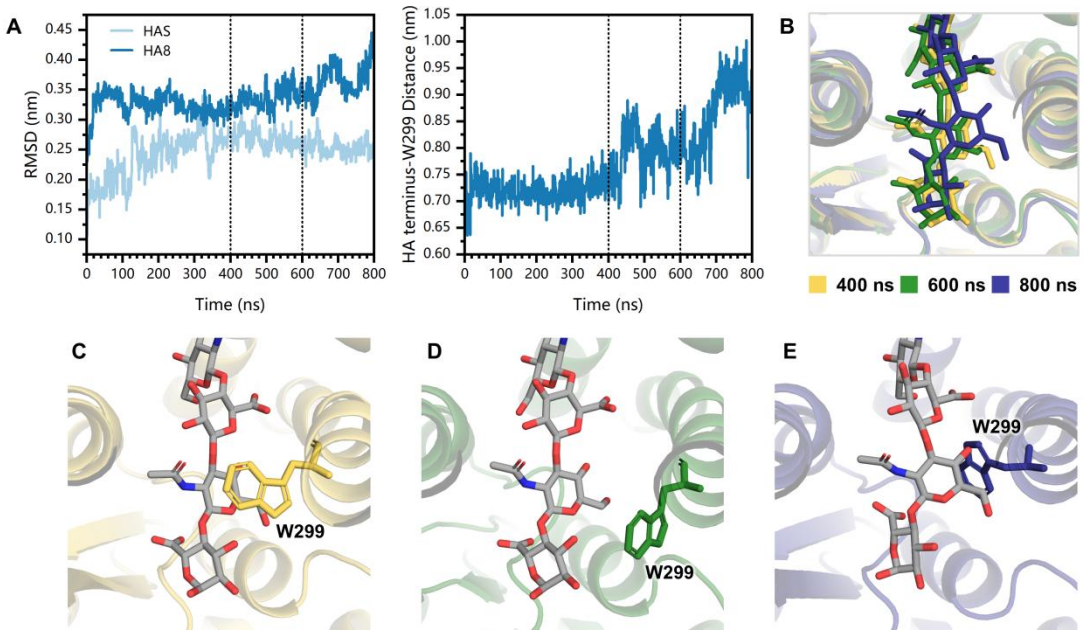
Supplementary Figure 3 Convergence of free energy profiles during metadynamics simulations.



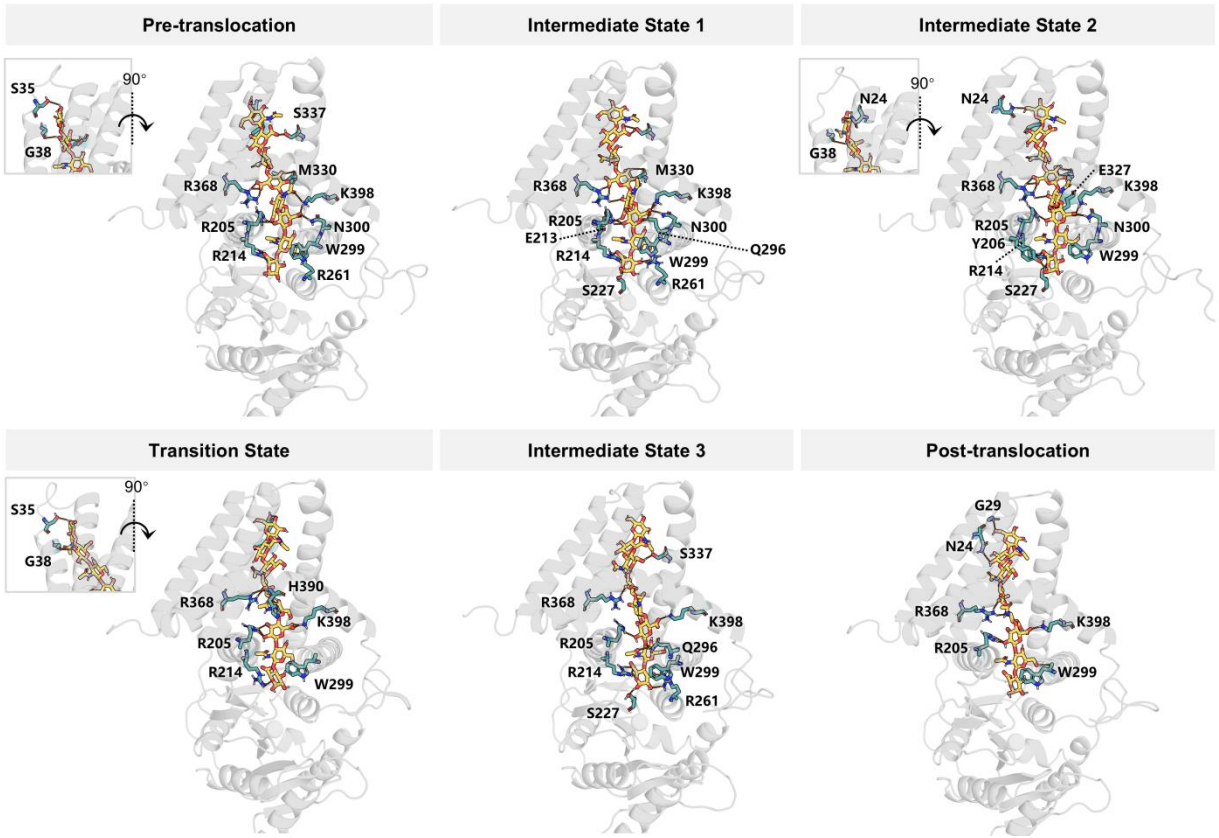
Supplementary Figure 4 Local free energy minima unrelated to HA translocation. State a0 and d2: HA8 excessively folded within the GT pocket (a0) or the TM tunnel (d2). State d1: HA8 overstretched in the TM tunnel, with the HA terminus not entering the tunnel. State A (UDP-HA8): UDP-HA8 withdrawn from the TM tunnel and accommodating in the GT pocket. State D: Half of the HA terminus enters the TM tunnel, UDP moiety attracted by the negatively charged TM vestibule. The vacuum electrostatic surface is generated by PyMOL.



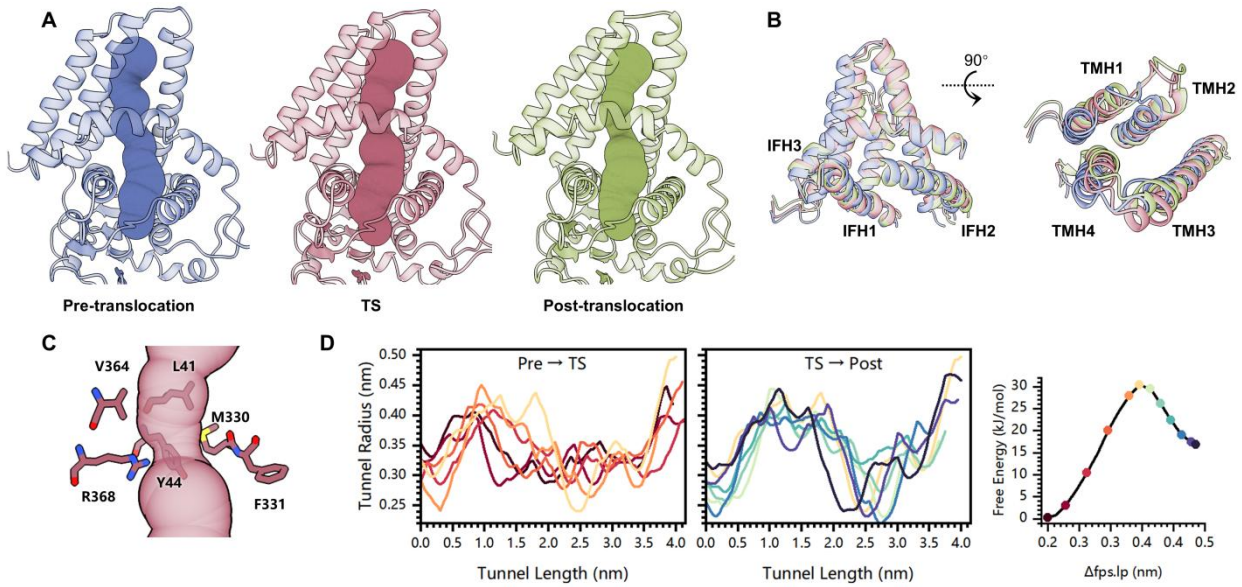
Supplementary Figure 5 Water-mediated H-bonding chain between D284 and HA during 25 ns MD simulation of the pre-translocation state. (A) RMSD of HA and HAS. (B) Bridging water counts. -1: no interaction, 0: direct H-binding between D284 and HA, 1: one-water-mediated H-bonding chain, 2: two-waters-mediated H-bonding chain. (C) D284 forms water-mediated H-bonding (orange dashes) with HA on the opposite side of D260.



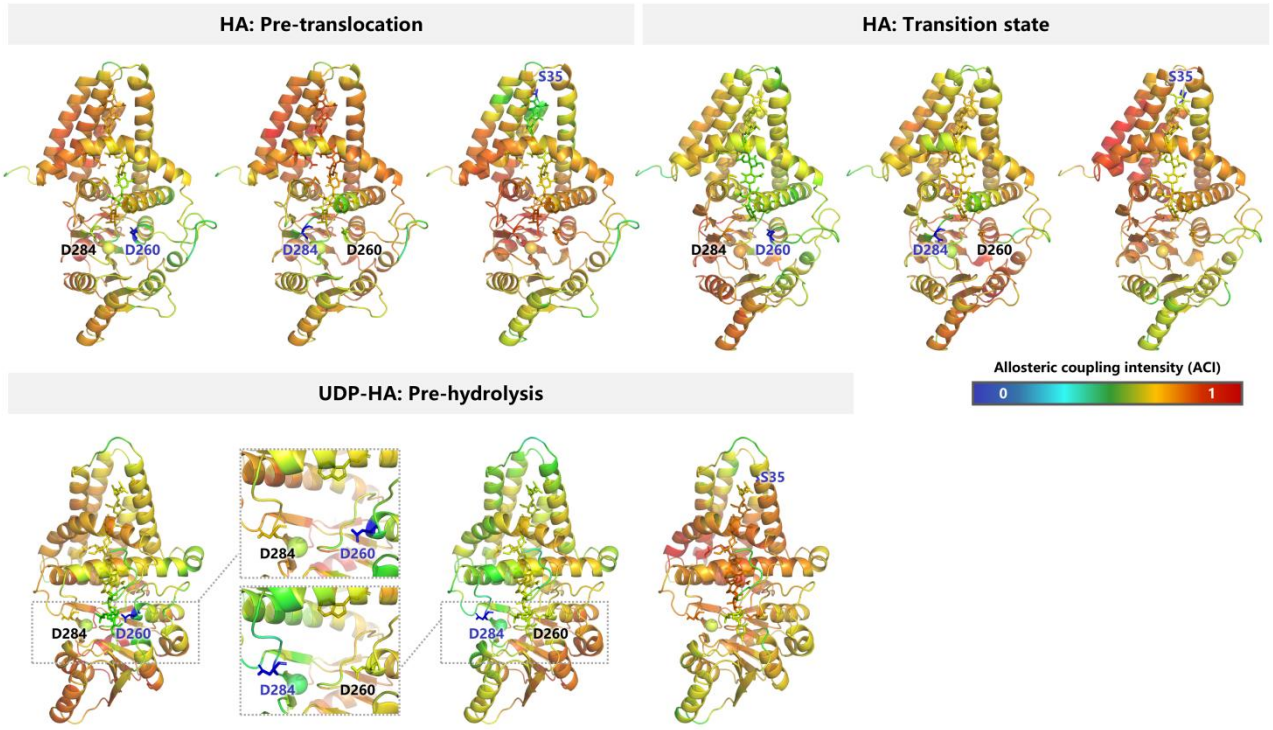
Supplementary Figure 6 Spontaneous HA8 translocation during 800 ns all-atom equilibrium MD simulation. (A) RMSD and HA position over time. (B) Conformational alignment of gradually translocated HA. (C) W299 at the TM tunnel entrance stabilizes GlcNAc-2 via CH-π interaction. (D) W299 fluctuates downward, releasing the constraint on HA. (E) W299 sways away, allowing HA translocation toward the TM tunnel.



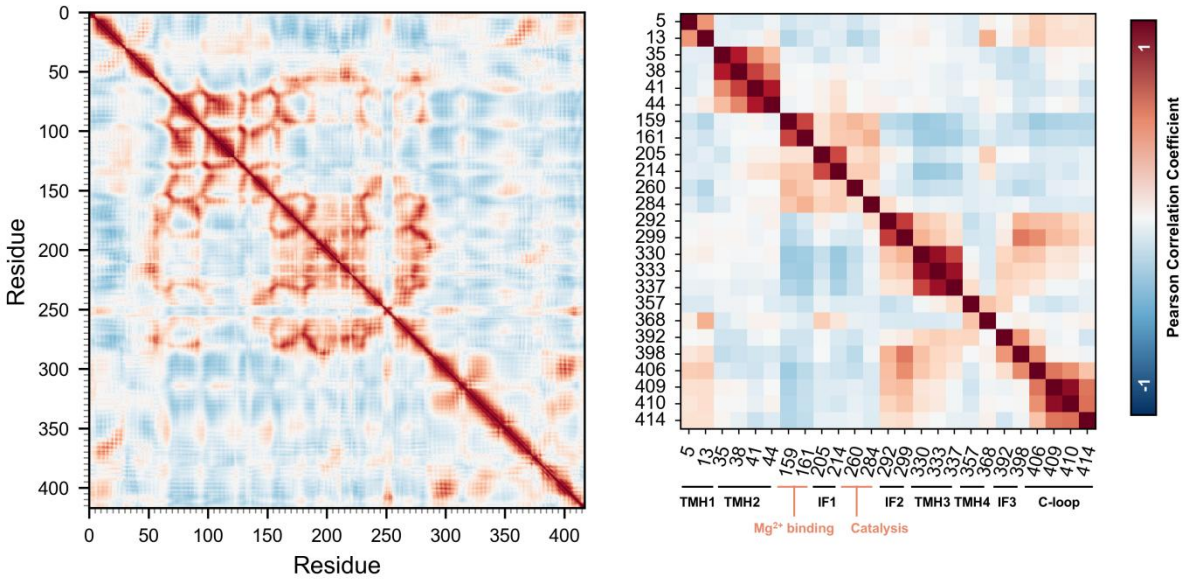
Supplementary Figure 7 Hydrogen bonds between HAS and HA8 during translocation



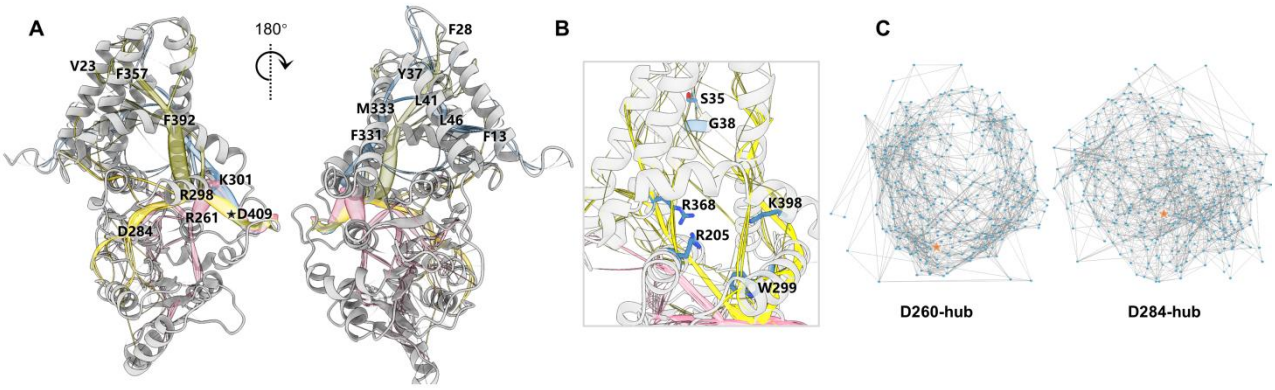
Supplementary Figure 8 TM tunnel variation during HA translocation. (A) TM tunnel geometry. (B) Conformational changes in the TM domain that define the tunnel geometry. (C) Bottleneck residues at the transition state. (D) Tunnel radius profiles of the intermediate states.



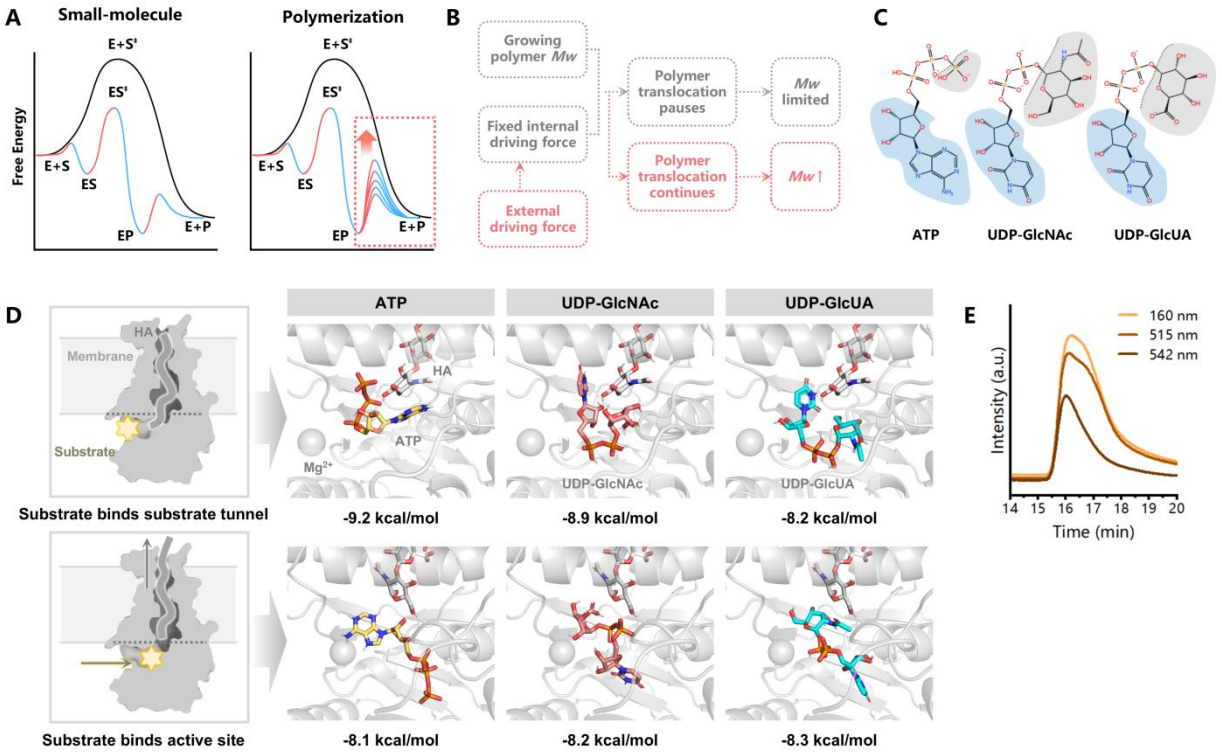
Supplementary Figure 9 Allosteric correlation analyzed by Ohm. The origins for dynamics propagation analysis are highlighted in blue.



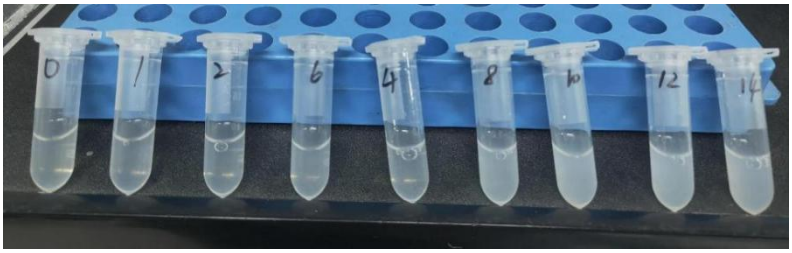
Supplementary Figure 10 Dynamic cross-correlation matrix of residue motions. Correlations between Cα atom positions are calculated using the Pearson method.



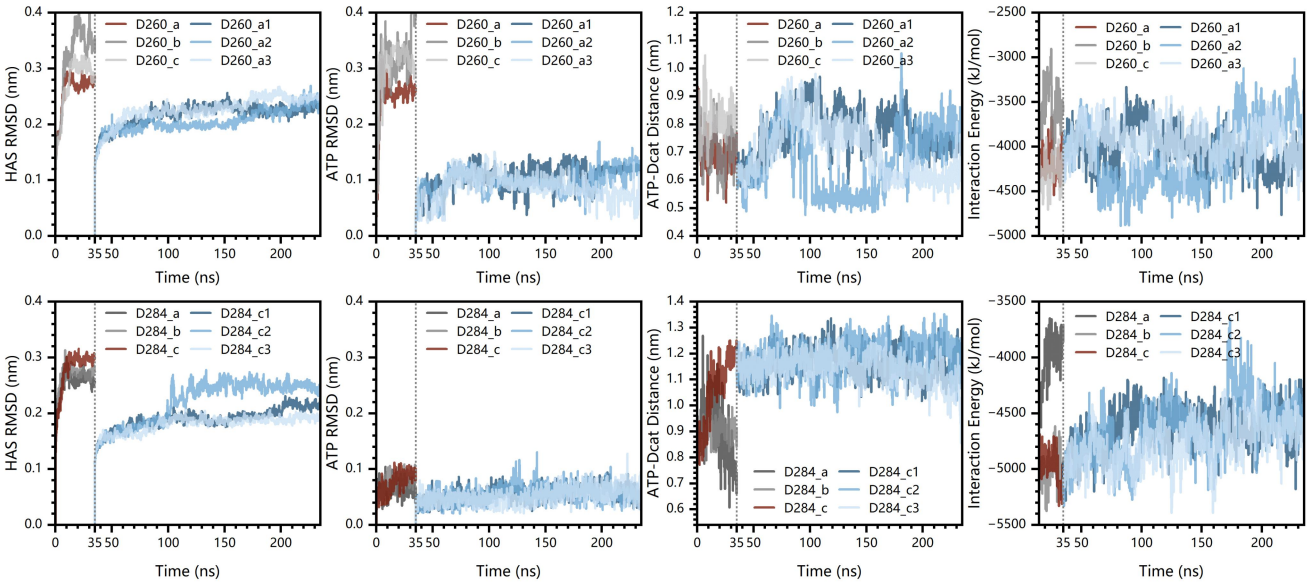
Supplementary Figure 11 Allosteric networks identified by MDPath. (A) D409-hub allosteric network. The top 500 pathways are clustered and color-coded. (B) Hotspots of the D260-hub allosteric network coincide with the driving residues. (C) Graphs of allosteric pathways. The hub residues are highlighted in orange.



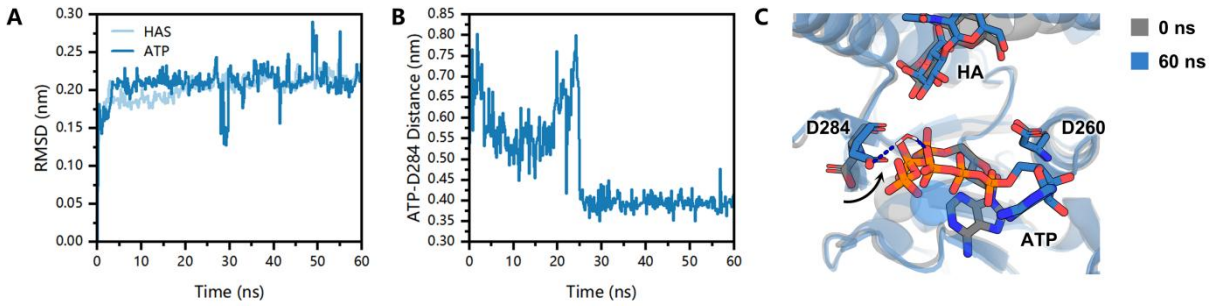
Supplementary Figure 12 Hypothesis of ATP energization increasing HA M_w. (A) Free energy profiles of small-molecule and polymerization reactions. E, S, P, and the superscript "‡" denote enzyme, substrate, product, and transition state, respectively. (B) Model of external energy input increasing M_w. (C) Structures of ATP, UDP-GlcNAc, and UDP-GlcUA. The blue region indicates the nucleoside subunit. S_N2 substitution occurs at the gray dashed line. (D) Molecular docking of ATP or substrate in the substrate tunnel or active sites of SeHAS. The SeHAS-HA complexes are equilibrated over 800 ns. Binding energy is shown below each conformation. (E) Effect of IMV size on HA M_w.



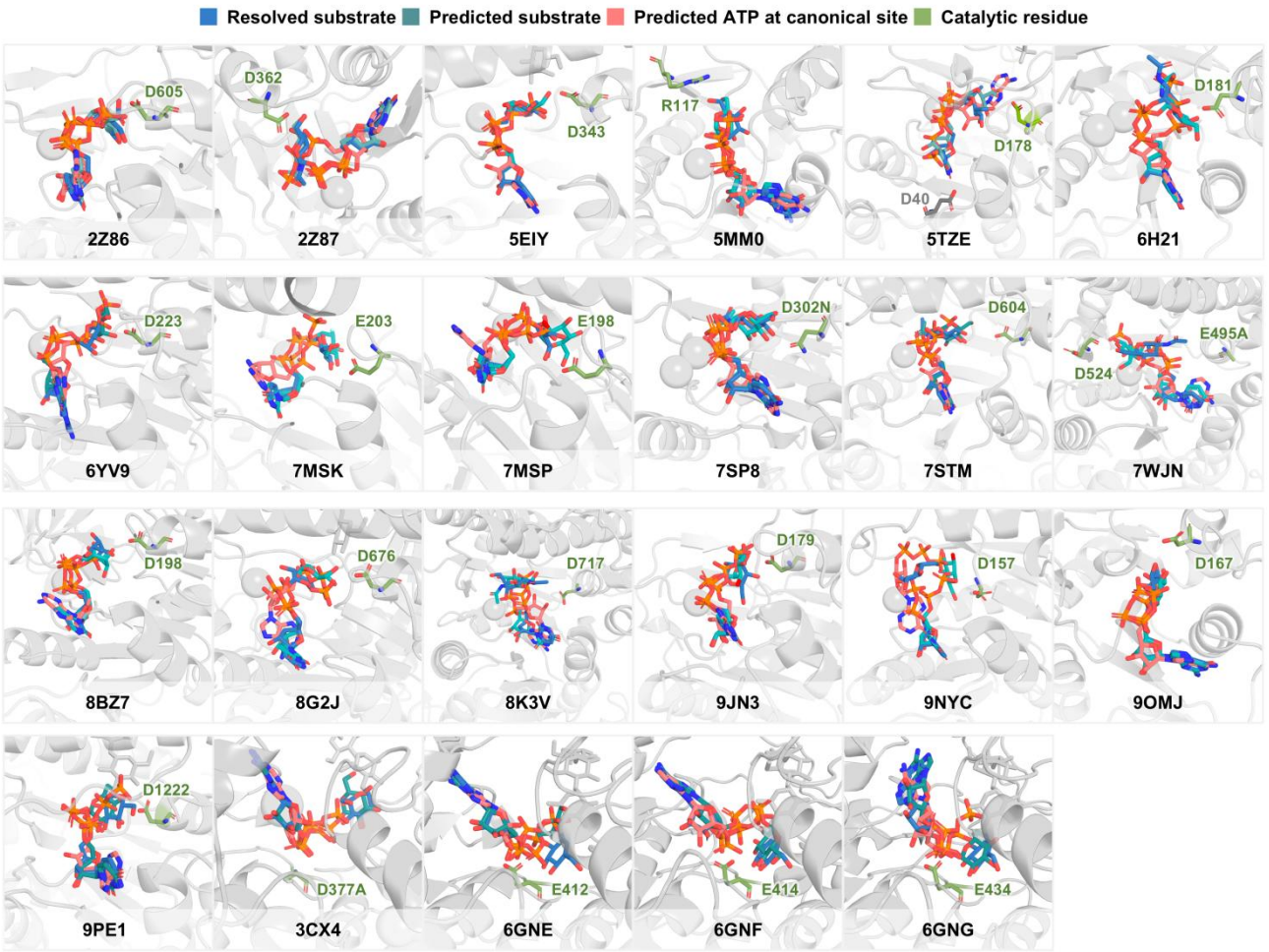
Supplementary Figure 13 IMV aggregation with increasing ATP concentration (mM, labeled on tubes).



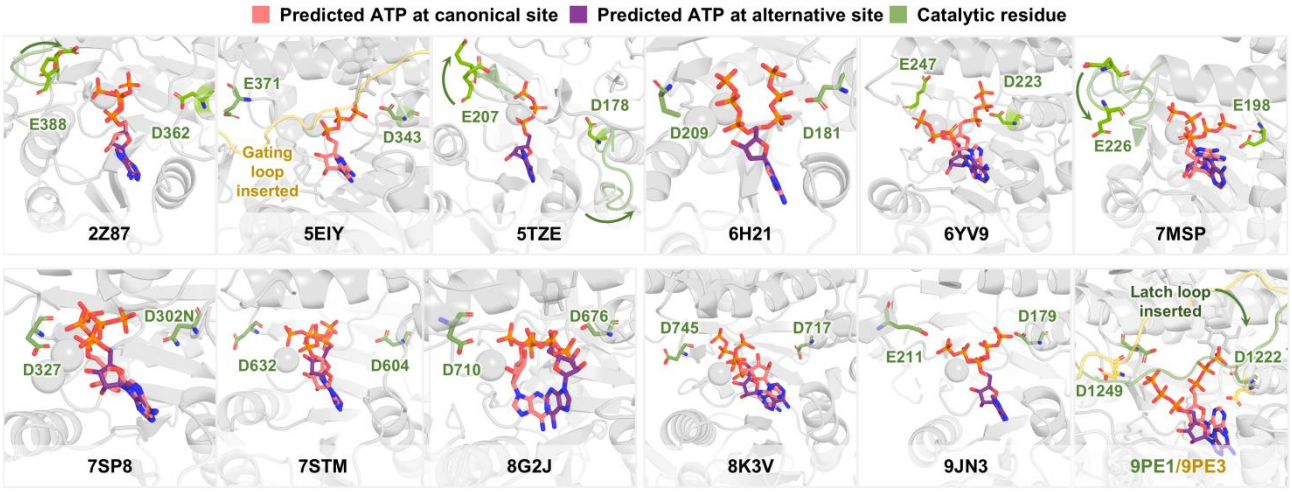
Supplementary Figure 14 Time profiles of ATP-HAS MD simulations. The simulated annealing procedure is conducted during the first 35 ns simulation, and the lowest-energy conformation (red) is selected for subsequent equilibration. ATP-Dcat distance: distance between the centroid of the γ-phosphate of ATP and the carboxyl group of the catalytic residue (D260 or D284). Interaction energy: sum of Lennard-Jones and Coulomb potentials between ATP and the HAS-Mg^{2+} receptor.



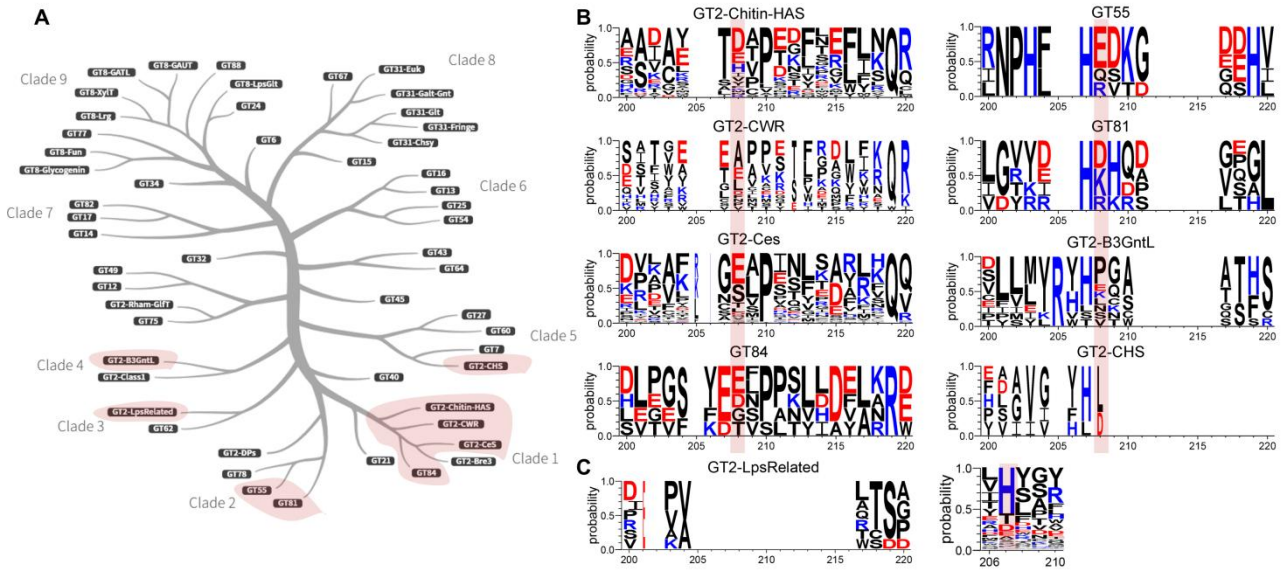
Supplementary Figure 15 D284 reorients towards ATP during MD simulation of the HAS-HA-ATP complex. (A) RMSD of HA and ATP. (B) Distance between the centroid of the γ-phosphate of ATP and the carboxyl group of D284. (C) The carboxyl group of D284 reorients toward ATP and forms water-bridge hydrogen bonds with the γ-phosphate of ATP. Hydrogen bonds are indicated by blue dashes. Simulation is initiated from the post-translocation state of the HAS-HA complex docked with ATP.



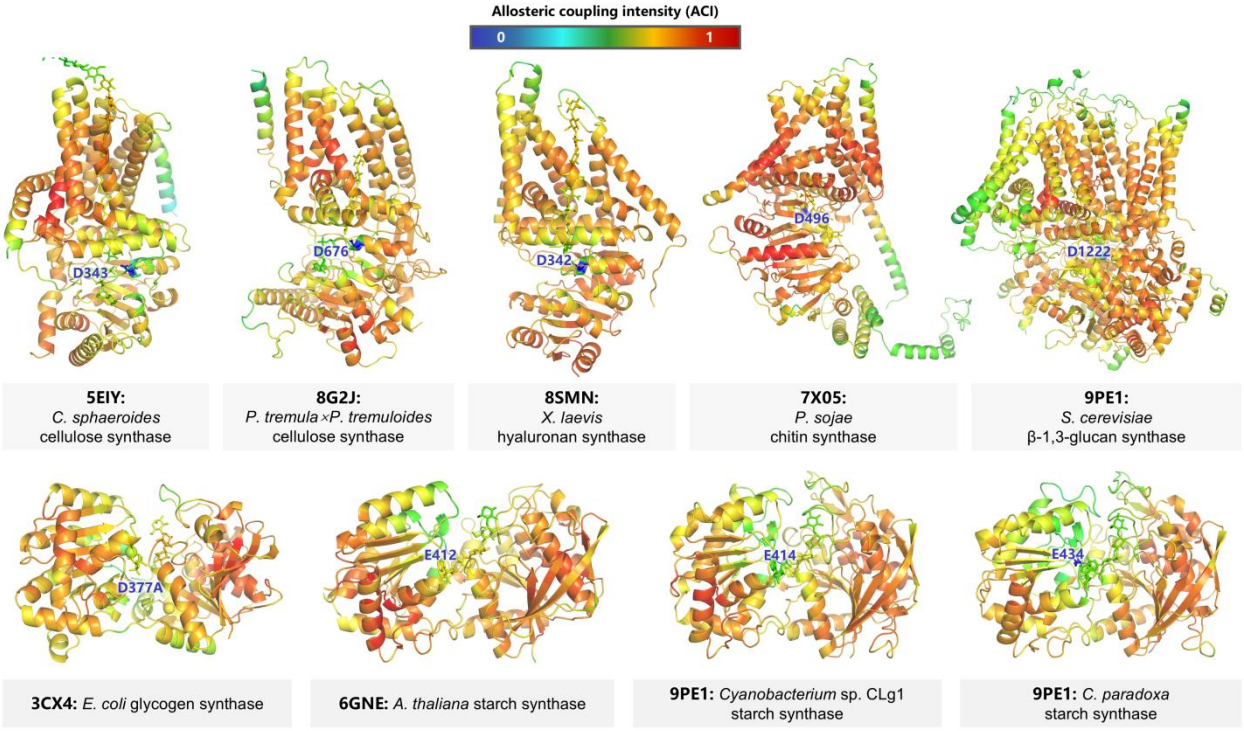
Supplementary Figure 16 ATP and substrate binding poses overlap in GT2, GT48, and GT5 enzymes as predicted by molecular docking. ATP in 5TZE flips towards D40.



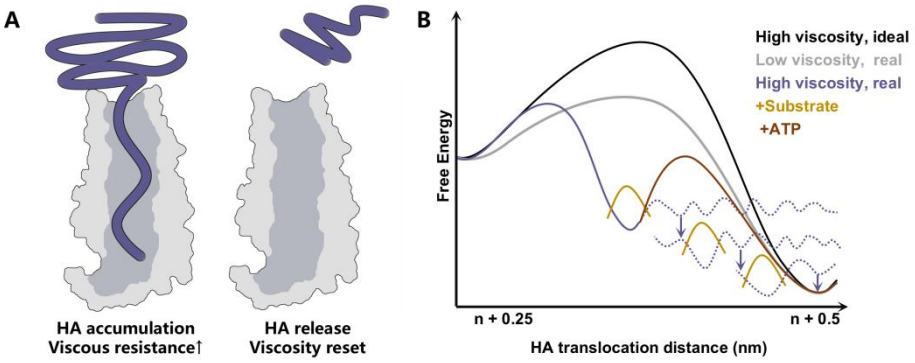
Supplementary Figure 17 Dual ATP binding sites in GT2 and GT48 enzymes predicted by molecular docking and AlphaFold3. Alternative ATP conformations in 2Z87, 5TZE, 7MSP and 6YV9 are predicted by AlphaFold3, as they require loop rearrangements (green arrows) to accommodate ATP. ATP in 6YV9 binds 1.1 nm from E247, suggesting further conformational adjustments for stable binding. In 5EIY, the inserted loop blocks the alternative site and no ATP binding pose is predicted. For the β-1,3-glucan synthase from GT48 family, 9PE1 (green) represents a post-elongation state with an inserted latch loop stabilizing UDP, the 9PE3 (yellow) is a basal state with latch loop open.



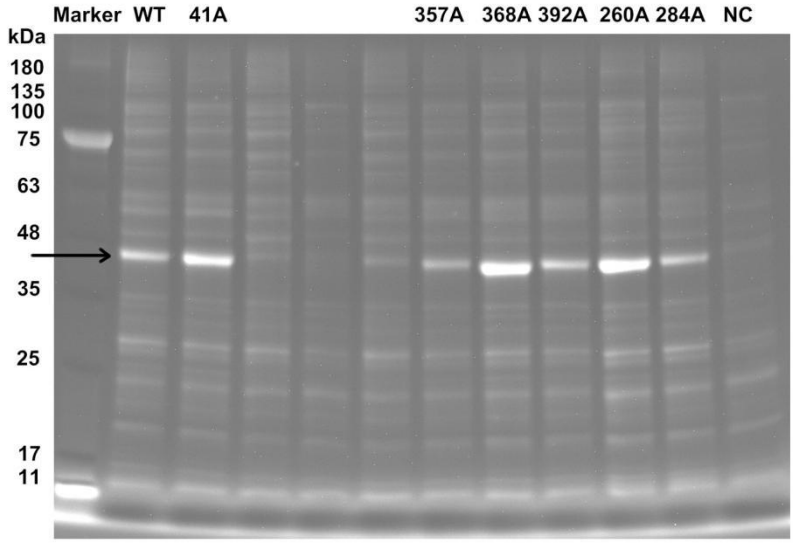
Supplementary Figure 18 Conservation of the alternative active site D/E 208 across the GT-A fold family. (A) Families with conserved D/E 208 (red) in the GTxplorer phylogeny tree. (B) Sequence conservation of the phylogeny tree sequences. D/E 208 is highlighted in red. (C) The GT2-LpsRelated family shows no conservation using the tree dataset (left), low conservation using random 500 sequences (right) or full alignment of >10,000 sequences (see GTxplorer webserver).



Supplementary Figure 19 Allosteric correlation of processive GT2, GT48, and GT5 enzymes analyzed by Ohm. The catalytic residue (blue) was defined as the origin of perturbation propagation. PDB structures with bound polysaccharide were selected for analysis.



Supplementary Figure 20 Effects of viscous resistance on HA translocation. (A) Viscosity increases with growing HA chain. (B) Mechanistic hypothesis for the boosting effects of substrate and ATP on HA translocation. Growing viscosity increases the energy barrier for translocatoin (gray to black lines). The fixed chemical energy only drives a semi-translocation of HA (blue line). Frequent substrate entry into the GT pocket biases the conformational distribution and allows thermal fluctuations (blue dashes) to drive translocation (blue arrows). ATP remains accessible for reaction due to its smaller size and fuels HA translocation via additional chemical energy.



Supplementary Figure 21 SDS-PAGE of SeHAS mutants. Full-length SeHAS bands are indicated by arrow. Rainbow pre-stained markers appear dark due to their non-fluorescent dyes after UV excitation. NC: negative control (WT without induction).

Supplementary Table 1 GT enzymes used for molecular docking

PBD	Substrate	Description	GT family[a]	Active site		Binding energy (kcal/mol)		
				Can.[b]	Alt.[c]	Substrate	ATP, can.	ATP, alt.
2Z86	UDP-GlcUA	Chondroitin polymerase (C-terminal domain)	GT31-Chsy	D605	-	-8.80	-7.81	-
2Z87	UDP-GalNAc	Chondroitin polymerase (N-terminal domain)	GT2-CHS	D362	E388	-10.29	-8.69	-[d]
5EIY	UDP-Glc	Cellulose synthase subunit A	GT2-Ces	D343	E371	-11.55	-9.69	-[e]
5MM0	GDP-Man	Dolichyl phosphate mannose synthase	GT2-DPs	R117	-	-9.63	-9.37	-
5TZE	UDP-GlcNAc	Wall teichoic acid β-glycosyltransferase	GT2-B3GntL	D178	E207	-11.39	-	-[d]
6H21	UDP-GlcNAc	Polyribitol phosphate β-1,3-N-acetylglucosaminyl-transferase	GT2-Class1	D181	D209	-7.79	-7.86	-8.04
6YV9	GDP-Man	Mannosyltransferase	GT2-CWR	D223	E247	-9.10	-9.09	-[d]
7MSK	UDP-Glc	Glycocin S/O-glucosyltransferase	GT2-LpsRelated	E203	-	-10.02	-8.33	-

Supplementary Table 1 (Continued) GT enzymes used for molecular docking

PBD	Substrate	Description	GT family[a]	Active site		Binding energy (kcal/mol)		
				Can.[b]	Alt.[c]	Substrate	ATP, can.	ATP, alt.
7MSP	UDP-Glc	Sublancin S-glycosyltransferase	GT2-LpsRelated	E198	E226	-9.89	-8.84	-[d]
7SP8	UDP-GlcNAc	Hyaluronan synthase	GT2-Chitin-HAS	D302N	D327	-9.11	-8.21	-8.37
7STM	UDP-GlcNAc	Chitin synthase 2	GT2-Chitin-HAS	D604	D632	-8.13	-8.38	-7.83
7WJN	UDP-GlcNAc	Chitin synthase 1 mutant E495A	GT2-Chitin-HAS	E495A	D524	-7.28	-[f]	-6.84
8BZ7	TDP-Rha	Wall teichoic acid rhamnosyltransferase	GT2-Rham-GlfT	D198	-	-10.05	-7.69	-7.97
8G2J	UDP-Glc	Cellulose synthase 8	GT2-Ces	D676	D710	-8.51	-7.69	-8.42
8K3V	UDP-GlcNAc	Chitin synthase 1	GT2-Chitin-HAS	D717	D745	-8.55	-7.83	-8.40
9JN3	UDP-Glc	Tubercidin/adenosine 3'-O-glucosyltransferase	GT2-Class1	D179	E211	-9.31	-8.04	-7.77
9NYC	UDP-Glc	Undecaprenyl-P β-glucosyltransferase	GT2-DPs	D157	-	-8.38	-7.40	-
9OMJ	GDP-β-Fuc	O-polysaccharide fucosyltransferase	GT2-Class1	D167	-	-9.36	-8.18	-
9PE1	UDP-Glc	β-1,3-glucan synthase	GT48	D1222	D 1249	-10.57	-8.77	-8.69[g]
3CX4	ADP-Glc	Glycogen synthase mutant E377A	GT5	E377A	-	-11.23	-9.38	-
6GNE	ADP-Glc	Starch synthase IV	GT5	E412	-	-10.89	-9.30	-
6GNF	ADP-Glc	Granule bound starch synthase	GT5	E414	-	-9.96	-9.32	-
6GNG	ADP-Glc	Granule bound starch synthase	GT5	E434	-	-9.92	-8.55	-

a. GT2 subfamilies were annotated based on sequence alignment with the GT-A phylogeny tree recorded in the GTxplorer webserver.
b. Canonical active site.
c. Alternative active site.
d. The alternative sites were identified by AlphaFold3 rather than docking, because ATP binding requires conformational changes such as loop opening (Supplementary Figure 17). Therefore, binding energy was not reported.
e. The alternative sites were inaccessible due to insertion of the gating loop into the GT pocket

(Supplementary Figure 17).

f. Ligands were adjacent to the alternative sites due to the E495A mutation.

g. Docked to the 9PE3 variant with an open GT pocket.

Supplementary Table 2 Reconstitution condition and IMV size

IMV size[a] (nm)	Lipid (mmol/L)	DDM (mmol/L)	Biobeads SM2 (g/L)
160	2.44	9.8	200
515	2.44	19.6	400
542	4.88	19.6	400

a. IMV size was measured by dynamic light scattering (DLS).