

**Manuscript version: Author's Accepted Manuscript**

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

**Persistent WRAP URL:**

<http://wrap.warwick.ac.uk/162619>

**How to cite:**

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

**Copyright and reuse:**

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

**Publisher's statement:**

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk).

# Structure, substrate-recognition, and initiation of hyaluronan synthase

Finn P. Maloney<sup>1†</sup>, Jeremi Kuklewicz<sup>1†</sup>, Robin A. Corey<sup>2</sup>, Yunchen Bi<sup>1,3</sup>, Ruoya Ho<sup>1</sup>,  
Lukasz Mateusiak<sup>4</sup>, Els Pardon<sup>5,6</sup>, Jan Steyaert<sup>5,6</sup>, Phillip J. Stansfeld<sup>7</sup>, Jochen  
Zimmer<sup>1\*</sup>

† These authors contributed equally

<sup>1</sup> Department of Molecular Physiology and Biological Physics, University of Virginia  
School of Medicine; 480 Ray C. Hunt Dr., Charlottesville, VA 22903, USA

<sup>2</sup> Department of Biochemistry, University of Oxford; Oxford, OX1 3QU, UK

<sup>3</sup> Present address: CAS and Shandong Province Key Laboratory of Experimental  
Marine Biology, Institute of Oceanology, Center for Ocean Mega-Science, Chinese  
Academy of Sciences; Qingdao, China, and Laboratory for Marine Biology and  
Biotechnology, Pilot National Laboratory for Marine Science and Technology (Qingdao);  
Qingdao, China

<sup>4</sup> Laboratory for In Vivo Cellular and Molecular Imaging, ICMI-BEFY, Vrije Universiteit  
Brussel; Laarbeeklaan 103, 1090 Brussels, Belgium

<sup>5</sup> VIB-VUB Center for Structural Biology, VIB; Pleinlaan 2, 1050 Brussels, Belgium

<sup>6</sup> Structural Biology Brussels, Vrije Universiteit Brussel; VUB, Pleinlaan 2, 1050  
Brussels, Belgium

<sup>7</sup> School of Life Sciences & Department of Chemistry, University of Warwick; Gibbet Hill  
Campus, CV4 7AL, UK

\* Corresponding author: jochen\_zimmer@virginia.edu

## Summary

Hyaluronan (HA) is an acidic heteropolysaccharide of alternating N-acetylglucosamine and glucuronic acid sugars that is ubiquitously expressed in the vertebrate extracellular matrix<sup>1</sup>. The high molecular weight polymer modulates essential physiological processes in health and disease, including cell differentiation, tissue homeostasis, and angiogenesis<sup>2</sup>. HA is synthesized by a membrane-embedded processive glycosyltransferase, HAS, that catalyzes the synthesis and membrane translocation of HA from UDP-activated precursors<sup>3,4</sup>. Here, we describe five cryo-electron microscopy structures of a viral HAS homolog at different states during substrate binding and initiation of polymer synthesis. Combined with biochemical analyses and molecular dynamics simulations, our data reveal how HAS selects its substrates, hydrolyzes the first substrate to prime the synthesis reaction, opens a HA-conducting transmembrane (TM) channel, ensures alternating substrate polymerization, and coordinates HA inside its TM pore. Our work suggests a detailed model for the formation of an acidic extracellular heteropolysaccharide and provides the first insights into the biosynthesis of one of the most abundant and essential glycosaminoglycans in the human body.

## Introduction

Hyaluronan (HA) is a ubiquitous acidic glycosaminoglycan of the vertebrate extracellular matrix (ECM), particularly enriched in connective tissues, the vasculature, and cartilage<sup>1</sup>. HA modulates a broad range of tissue remodeling processes, including wound healing, embryological development, angiogenesis, and tumorigenesis<sup>5-7,2</sup>. It is a linear polysaccharide of alternating  $\beta$ -1,3 and  $\beta$ -1,4-linked N-acetylglucosamine (GlcNAc) and glucuronic acid (GlcA) units, respectively, (Fig. 1a and b) that is typically megadaltons in size<sup>2,8</sup>. High molecular weight HA (>10<sup>6</sup> kDa) associates with healthy tissue homeostasis, while low molecular weight species exhibit pro-inflammatory and angiogenic properties<sup>2</sup>.

HA is synthesized from UDP-activated monosaccharides. Type-I HASs contain a single catalytic domain and secrete the nascent HA polymer during synthesis through a TM channel formed by their membrane-embedded segment<sup>9,3,4</sup>. Type-II HAS, however, is limited to select bacteria and is not membrane integrated. The enzyme uses two

catalytic domains to transfer either GlcA or GlcNAc in a non-processive manner<sup>10,11</sup>. Thus, Type-I HASs not only couple HA synthesis with translocation, but also utilize a single catalytic domain to transfer two different donor sugars by forming substrate-specific glycosidic linkages.

Vertebrates express three HAS isoforms (Extended Data Fig. 1) that differ in tissue expression and catalytic activity; HAS-2 is essential<sup>12,13</sup>. HAS contains a cytosolic glycosyltransferase (GT) domain flanked by two N-terminal and four C-terminal TM helices (TMH). How HAS selects its substrates, catalyzes regio- and stereospecific glycosyl transfer, secretes HA, and controls polymer length are important unresolved questions.

Chlorella viruses (Cv) contain HAS enzymes homologous to vertebrate HAS that synthesize and secrete authentic HA *in vivo* and *in vitro*<sup>14,15</sup>. Cv infect unicellular green algal endosymbionts of the ciliate *Paramecium bursaria*, and HA production by the infected algae may benefit the endosymbiotic relationship. Cv-HAS shares with human HAS-2 ~45% sequence similarity, the same predicted number and distribution of TMHs, as well as a highly conserved catalytic GT domain (Extended Data Fig. 1).

To reveal the mechanistic basis for HA biosynthesis, we determined cryo-electron microscopy (EM) structures of Cv-HAS in apo, UDP-bound, UDP-GlcNAc-bound, and GlcNAc-primed states. Our analyses provide the structural basis for substrate selectivity and alternating polymerization, demonstrate that only GlcNAc can prime HA biosynthesis, and suggest a model for processive HA synthesis and translocation. These insights are corroborated by functional studies and molecular dynamics (MD) simulations.

## Results

Cv-HAS was expressed in *E. coli* and purified as described in the Methods. To facilitate EM analysis of the ~65 kDa enzyme, Cv-HAS was bound to two camelid nanobodies produced *in vivo*, Nb872 and Nb881 (Fig. 1c and d)<sup>16</sup>. Nb872, binding to the

extracellular TMH5-6 loop, supports *in vitro* catalytic activity similar to un-complexed Cv-HAS and thermo-stabilizes the enzyme (Extended Data Fig. 2a and b). Nb881 binds a peripheral region of the GT domain and reduces HA biosynthesis for unknown reasons (Extended Data Fig. 2b-c).

Cryo-EM analyses of Cv-HAS were performed in lipid nanodiscs formed from *E. coli* total lipid extract and the MSP1E3D1 scaffold protein, see Methods. Samples were analyzed bound to UDP, a product and competitive inhibitor of Cv-HAS<sup>17</sup>, and the substrate UDP-GlcNAc. Three-dimensional sorting and variability analyses of the ‘UDP-GlcNAc-bound’ dataset revealed distinct states: nucleotide-free ‘apo’, substrate-bound, and monosaccharide-bound ‘primed’ (Extended Data Fig. 3-5, Supplementary Video 1). The cryo-EM maps generated in cryoSPARC<sup>18</sup> range in estimated resolutions from 3.1 - 2.7 Å (Supplementary Table 1) and resolve essentially all residues, with the exceptions of residues 1 – 37 (including TMH1), 452 – 464 connecting IF3 with TMH5, and 553 - 561 at the C-terminus.

#### *Architecture of Cv-HAS*

Cv-HAS consists of a cytosolic catalytic domain that adopts a GT-A fold<sup>19</sup> and packs against the five resolved TMHs (TMH2-6) as well as three amphipathic interface helices (IF1-3) near the water/lipid interface (Fig. 1e). The GT domain contacts the TMHs via IF1 and IF2 (residues 238-263 and 331-357, respectively), of which IF2 contains a QxxRW motif characteristic of membrane-embedded processive family-2 GTs, according to the CAZy classification<sup>20</sup> (residues 339-342). IF3, a C-terminal extension of TMH4, runs perpendicular and on top of the IF1/2 pair, thereby framing the entrance to the TM channel, discussed below.

The TM architecture resembles a teepee with helices straddling the cytosolic IF helices. TMH2 extends past the cytosolic water/lipid interface to interact with the GT domain. Past TMH6, 13 C-terminal residues (540-552) form a short  $\alpha$ -helix that packs into a

groove on the GT domain formed by its central  $\beta$ -sheet and helix 1, together with IF1 (Fig. 1e).

Cv-HAS' first TMH is disordered in all cryo-EM maps. A MapPred co-evolution analysis<sup>21</sup> suggests evolutionarily-coupled residues within TMH1 and TMH2 that position the helix in a groove between TMH2 and 4 (Extended Data Fig. 6a). Direct interactions with TMH2 are also supported by AlphaFold2<sup>22</sup> and RoseTTAfold<sup>23</sup> Cv-HAS models, although the predictions position TMH1 differently around TMH2 (Extended Data Fig. 6b and c).

TMH1 is also disordered when using two cytosolic Nbs to facilitate structure determination (Extended Data Fig. 6d), indicating that its flexibility is not induced by Nb872. The helix is not essential for function as TMH1-truncated Cv-HAS produces HA *in vitro*, albeit at reduced levels compared to the wild type enzyme (Extended Data Fig. 6e).

Helices 4 and 5 of the GT domain are connected by an 11-residue long 'priming loop' that forms one wall of the active site (residue 291–301). The loop precedes the putative base catalyst Asp302 within the GDD motif that facilitates glycosyl transfer. It contains a conserved cysteine at its tip (Cys297) and extends roughly along the GT's central  $\beta$ -sheet towards the nucleotide-binding pocket (Fig. 1e). As demonstrated initially by the structure of the soluble enzyme SpsA and later cellulose synthase, the catalytic pocket of family-2 GTs includes characteristic acidic motifs that form a single substrate binding pocket<sup>19,24,25</sup>. For Cv-HAS, these include the nucleotide binding 'DGD' motif (121-123), the cation binding 'DSD' motif (201-203), and the putative catalytic 'GDD' motif (300-302) (Fig. 1e, Extended Data Fig. 1).

#### *Cv-HAS forms a lipid-filled crescent-shaped channel*

Cv-HAS contains a short membrane-embedded region about 20 and 30 Å thick near TMH2 and above the active site (proximal to the TMH5-6 face), respectively (Fig. 1e and f). The TMHs form a crescent-shaped channel with a cytosolic entrance above the

active site formed by Trp342 of the QxxRW motif within IF2 (Fig. 1f). Halfway across the membrane, IF3, TMH5, and TMH6 create a lateral channel opening towards the lipid bilayer. Two lipid molecules, assigned as a CHS molecule and a phospholipid tail based on their shapes, occupy and seal the portal (Fig. 1f and Extended Data Fig. 7). Atomistic MD simulations in a POPE bilayer corroborate that lipid acyl chains entering the channel portal indeed prevent water flux across the membrane (Extended Data Fig. 7b).

### *Substrate-bound conformation of Cv-HAS*

Cv-HAS structures bound to UDP-GlcNAc were determined using an inactive enzyme in which Asp302 was replaced with Asn (Extended Data Fig. 1 and 8). Unexpectedly, the substrate was partially hydrolyzed during sample preparation, capturing views of the active site before and after priming (Fig 2a-c). Cv-HAS requires manganese for catalytic activity<sup>15</sup>, indicating that it binds  $Mn^{2+}$ -complexed substrates. Nucleotide binding displaces the priming loop, which retracts from the active site towards the cytosolic water-lipid interface (Fig. 2d, Extended Data Fig. 8d and f, and Supplementary Video 2). This conformation is stabilized by the loop's Phe292, interacting with Arg348 and Tyr352 of IF2, as well as Tyr299 that occupies a hydrophobic pocket at the interface of IF2, IF3 and TMH5. Cys297 fits into a crevice at the water/lipid interface formed by Thr298 and Lys344 as well as Arg348 of IF2. Although its biological role is unknown, Cys297 is critical for function; replacing it with Ala renders Cv-HAS inactive (Fig. 2e). Relative to the apo state, nucleotide binding induces a ~10 degree rigid body rotation of the GT domain towards the membrane, fostering the priming loop-IF2 interaction. This movement narrows the active site cleft, which likely facilitates proper substrate positioning (Fig. 2a, Supplementary Video 3).

The UDP and UDP-GlcNAc-bound structures reveal similar UDP coordinations (Extended Data Fig. 8a and b, Supplementary Discussion 1). Asp121 of the conserved

'DGD' motif contacts the uracil's N3 ring nitrogen, while the ribose's C3 hydroxyl group interacts with Glu93 from the conserved 'EDP' motif (Fig. 2b, Extended Data Fig. 1). The UDP-GlcNAc diphosphate hydrogen-bonds with Gln338 of the QxxRW motif while the associated  $Mn^{2+}$  ion is coordinated by Asp203 and Asp327. A second  $Mn^{2+}$  ion is coordinated by Glu93 as well as Asp327 and Asp203 (Fig. 2b). Replacing Glu93 or Asp327 with alanine renders Cv-HAS inactive, suggesting that both  $Mn^{2+}$  ions are functionally important (Fig. 2e). It is likely that the additional  $Mn^{2+}$  ion that is not directly implicated in diphosphate coordination performs a structural role by positioning the Dx D motif (Asp201-Asp203) for substrate binding. Mutating Asp94 of the EDP motif to Ala reduces catalytic activity to about 20% relative to wild type. This residue does not directly coordinate  $Mn^{2+}$ , but interacts with Arg549 near Cv-HAS' C-terminus (Fig. 2b).

In the substrate-bound state, the donor sugar ring oxygen is proximal to Trp342's N1 hydrogen (Fig. 2b), right below the acceptor-binding site. The GlcNAc acetamido group occupies a pocket formed by Asp203, Cys231, Tyr248, Pro271, Asp327, and Ser326, and its carbonyl oxygen hydrogen bonds with the side chain hydroxyl of Tyr248 (Fig. 2b). The sugar's C4 hydroxyl interacts with Asp301 and Asp302, while its C6 hydroxyl points towards Arg341, positioned within a positively charged pocket containing Lys177. This pocket likely accommodates the C6 carboxylate when the substrate is UDP-GlcA (see below).

Compared to the nucleotide-free or primed conformations (see below), UDP or substrate binding repositions a partially conserved CVGGP loop (switch loop, residues 267-271) at the back of the nucleotide-binding pocket (Fig. 2f). The loop moves toward the membrane interface upon nucleotide binding, likely stabilized by unresolved bridging water molecules with the ligand. Cys267 at the beginning of the switch loop is necessary for function (Fig. 2e, Supplementary Discussion 2).

### *Priming of Cv-HAS*

Substrate hydrolysis likely generated the GlcNAc monosaccharide primer that diffused to the acceptor-binding site at the entrance to the TM channel (Fig. 2a, Extended Data Fig. 8e, Supplementary Video 1). The primer's acetamido group sits in a pocket formed



by Pro271, Arg256, Cys231, Tyr248 as well as backbone atoms from the switch loop (Fig. 2c). The sugar ring is stabilized through CH- $\pi$  stacking interactions with Trp342 and is coordinated by the switch loop's Cys267, whose sulfhydryl group is  $\sim 4.2$  Å away from the ring oxygen. To facilitate this interaction, the switch loop returns to its 'down' position (Fig. 2f). Further, the conserved Ser345 (Extended Data Fig.1) is in hydrogen bonding distance to the primer's C5 hydroxyl, while the conserved Arg303 binds the C6 hydroxyl group. These interactions position the primer's C3 hydroxyl towards the active site and the putative base catalyst (D302N) (Fig. 2c).

In the primed state, the priming loop re-inserts into the catalytic pocket (Fig. 2d, Extended Data Fig. 8f) and the GT domain relaxes away from the membrane upon priming (Fig. 2a, Supplementary Video 3). While an inserted loop position is also observed in the apo state, the loop is better resolved in the primed state, suggesting that the primer (or HA polymer) is stabilizing. Accordingly, the priming loop's Tyr299, which is a conserved Phe in vertebrate HAS, rotates towards the primer-binding site in close proximity to GlcNAc's C3 hydroxyl group (Fig. 2c).

### *Substrate hydrolysis initiates HA biosynthesis*

The *in situ* generation of the GlcNAc primer suggests substrate hydrolysis as a general initiation mechanism of HAS-related enzymes. Accordingly, supplied monosaccharides should also prime HA biosynthesis. To test this model, wild type Cv-HAS was incubated with  $^{14}\text{C}$ -labeled GlcNAc or GlcA and unlabeled substrates. The obtained polysaccharide revealed that the GlcNAc monosaccharide is readily incorporated into HA, while GlcA fails to prime the synthesis reaction (Fig. 3a).

Additionally, we used substrate hydrolysis experiments to verify the substrate-hydrolysis initiation mechanism. Cv-HAS hydrolyzes its substrates when exposed to only one of them without forming a homo-polysaccharide (Fig. 3a, lane 1). This allows monitoring substrate hydrolysis rates in real time based on enzyme-coupled NADH oxidation (Fig. 3b and Extended Data Fig. 9)<sup>9</sup>.

We observe that saturating GlcNAc monosaccharide increases the UDP-GlcA hydrolysis rate ~30-fold and decreases UDP-GlcNAc hydrolysis ~2.5-fold. Conversely, hydrolysis rates of both substrates were insignificantly affected by comparable concentrations of GlcA (Fig. 3b and Extended Data Fig. 9e). Thus, a GlcNAc primer facilitates the binding and/or turnover of UDP-GlcA to form an HA disaccharide, while a GlcA primer does not provide the same enhancement of UDP-GlcNAc hydrolysis.

#### *HAS priming creates a continuous TM channel*

Strikingly, we resolve two different TM architectures of the primed state – one similar to the apo and substrate-bound states, and another with the N-terminus of TMH2 tilted away from TMH3 (Fig. 4 and Extended Data Fig. 8h-i, Supplementary Video 4), creating a continuous TM channel. The newly formed channel overlaps with the curved, lipid-plugged channel right above the active site to form a common vestibule, yet diverges halfway across the membrane (Fig. 4c). The newly created channel is lined with hydrophilic and apolar residues and its dimensions are sufficient to accommodate an HA polymer about 5 disaccharides long (Fig. 4d).

TMH2 kinks around its contact point with IF1, mediated by Gln57 and Phe60 of TMH2 and Val254, Ala258 and Leu261 of IF1 (Extended Data Fig. 8g). The conserved residues Arg256 and Gln259 on the opposite side of IF1 point towards the channel lumen and Arg256 contacts the primer's C1 hydroxyl group (Fig. 2c). TMH2 could be further stabilized in this 'open' conformation by the predicted location of TMH1 (Extended Data Fig. 6a-c). Thus, HAS priming creates an exit path for the nascent HA at the interface of TMH1-2 and 3-4, similar to cellulose synthase (Extended Data Fig. 7c)<sup>25</sup>.

#### *Dynamic coordination of HA inside the TM channel*

255 We performed atomistic MD simulations to gain insights into HA coordination within the  
256 TM channel of Cv-HAS. Simulations were performed with a HA hexasaccharide  
257 manually placed within the channel vestibule, with the terminal non-reducing end sugar  
258 occupying the acceptor-binding site next to Trp342 (Fig. 2c).

259 We identify significant register-dependent differences in the positional stability of HA.  
260 While oligosaccharides starting with GlcNAc at the acceptor site were stable over the  
261 course of 5 simulations, polymers starting with GlcA were pulled into the TM channel by  
262 roughly one sugar unit (Fig. 5a). After this motion, the GlcA-ending polymer is  
263 coordinated similarly compared to the GlcNAc-ending chain, yet one sugar unit farther  
264 inside the channel (Fig 5a-e, Supplementary Table 2).

265 The HA oligosaccharide primarily interacts with conserved polar and charged side  
266 chains. Arg303 and Arg256 bind GlcA carboxylates in the terminal or the second  
267 position (Fig 5b-e), and Arg247 and Arg442 coordinate GlcA at the third or fourth  
268 positions. The GlcNAc groups are mainly coordinated by polar and aromatic residues,  
269 including Tyr248, which coordinates GlcNAc, but not GlcA, when in the terminal  
270 position. Trp346 interacts with GlcNAc when at the second or third position. All of these  
271 HA-coordinating interactions are also seen in simulations seeded from different initial  
272 HA poses albeit with some localized variability (Supplementary Table 2).

273 MD simulations of UDP-GlcA and HA-bound Cv-HAS also provide insights into the likely  
274 coordination of the UDP-GlcA substrate. While the UDP moiety is coordinated similarly  
275 to the experimental UDP-GlcNAc-bound state (Fig. 2b), the substrate's carboxylate is  
276 coordinated by Lys177 as well as Arg341 of the QxxRW motif, which bridges the  
277 nucleotide's  $\alpha$ -phosphate and GlcA's carboxylate (Fig 5f and Supplementary Table 2).

278 HA biosynthesis experiments with Cv-HAS constructs carrying substitutions at HA  
279 coordinating residues reveal their functional importance. Replacing Arg247 with Ala or  
280 Lys and Arg256 with Lys inactivates the enzyme, supporting a critical role for GlcA  
281 coordination. Substituting Trp346 with Leu similarly abolishes activity, while an Ala  
282 substitution of Tyr248 supports activity at roughly 20% of wild type (Fig. 5g).

283

## 284 **Discussion**

285 Cellulose, chitin and HA synthases are membrane-embedded GTs that synthesize and  
286 secrete high molecular weight polysaccharides. While cellulose and chitin are  
287 homopolymers, HA is a heteropolysaccharide of alternating GlcNAc and GlcA units  
288 linked via different glycosidic linkages. Our structural and biochemical analyses provide  
289 molecular insights into how HAS combines these seemingly irreconcilable tasks.

290 Monomeric Cv-HAS elongates HA's non-reducing end, similar to cellulose and chitin  
291 synthases<sup>15,25,26</sup>. This reaction mechanism stands in contrast to bacterial HAS enzymes  
292 that function as obligate dimers and elongate HA's reducing end<sup>15</sup>. Considering the  
293 similarities of vertebrate HASs to Cv-HAS, we propose that Cv-HAS mechanistically and  
294 structurally represents the vertebrate enzymes. This is supported by the excellent  
295 agreement of AlphaFold2 and RoseTTAfold models of human HAS-2 with our  
296 experimental Cv-HAS structure (Extended Data Fig. 6a-c)<sup>22</sup>. The models agree well  
297 over the range of the catalytic GT domain, the membrane associated IF helices, as well  
298 as the arrangement of the five resolved TM helices. Discrepancies exist within the  
299 predicted location of TMH1, which appears to be flexible.

300 Catalytic activity of HAS may be modulated by post-translational modifications  
301 (Supplementary Discussion 3). Initiating HA biosynthesis requires the formation of a  
302 priming carbohydrate acceptor. We demonstrate biochemically and structurally that  
303 substrate hydrolysis is the most likely mechanism to generate this primer (Fig. 3). A  
304 similar mechanism has been proposed for cellulose synthase<sup>27,28</sup> and experimentally  
305 tested for chitin synthase<sup>29</sup>. The observation that GlcA fails to prime the biosynthesis  
306 reaction is consistent with its instability at the acceptor-binding site in MD simulations.  
307 Accordingly, when the nascent HA chain with a terminal GlcA moves forward in the  
308 channel, only UDP-GlcNAc is likely to bind (Fig. 5h).

309 HAS extends GlcNAc-ending HA polymers with GlcA and not GlcNAc to ensure an  
310 alternating HA sequence. Our 3D variability analysis of Cv-HAS incubated with UDP-

GlcNAc reveals that the enzyme either binds UDP-GlcNAc at the active site or GlcNAc at the acceptor position, but not both ligands simultaneously. This is likely due to clashes between the GlcNAc primer and substrate and corroborated by the reduced rate of UDP-GlcNAc hydrolysis in the presence of saturating GlcNAc monosaccharide (Extended Data Fig. 9 and 10). To investigate this phenomenon at the structural level, we performed atomistic MD simulations of HAS with GlcNAc both in the donor (UDP-GlcNAc) and acceptor (HA) positions. Quantifying the occupancy of the acetamido binding pocket reveals that either just the donor or acceptor GlcNAc, but never both, are able to bind the pocket (Extended Data Fig. 11).

We further propose that the accepting sugars (GlcNAc or GlcA) are differently coordinated and positioned at the acceptor site. This would ensure that only the C4 hydroxyl of GlcA and the C3 hydroxyl of GlcNAc serve as acceptors during glycosyl transfer (Fig. 5h). For a nucleophilic displacement reaction<sup>19</sup>, the accepting hydroxyl group must be within close distance to the base catalyst Asp302 of the GDD motif as well as the donor's C1 carbon. Future structural analyses of substrate-bound HAS at intermediate states during HA translocation will be required to test this model.

Cellulose synthase produces a ribbon-shaped amphipathic glucose polymer that is secreted through a pore lined with aromatic and hydrophilic residues. In contrast, HAS' TM channel is wider and rich in positively charged residues, reflecting HA's properties. Structural insights into cellulose synthase are limited to polysaccharide associated states, i.e. describing enzymes with an open TM channel, whereas our HAS analyses are predominantly of the closed state. The overall similarity of the HAS and cellulose synthase TM architectures (Extended Data Figure 7c and 10b) reinforce our assignment of the HAS channel exit at the TMH2-4 interface, and suggest similar conformational changes when cellulose synthase opens and closes its TM channel.

Our cryo EM structures do not resolve a conserved loop connecting IF3 with TMH5, right above the catalytic pocket (Fig. 1). The loop contains a conserved WGTR/SG motif (Extended Data Fig. 1), which is important for function of bacterial HAS<sup>30</sup>. It resembles cellulose synthase's gating loop, which transiently inserts into the catalytic pocket for

substrate binding (Extended Data Fig. 10b)<sup>27</sup>. A similar process in HAS could reposition the substrate towards the acceptor sugar to facilitate HA elongation.

Following glycosyl transfer, cellulose synthase translocates the extended polysaccharide upon binding of a new substrate through conformational changes of a conserved  $\alpha$ -helix of the GT domain<sup>31</sup>. The same helix exists in HAS, with the GDD motif capping its N-terminus. Thus, a similar step-by-step HA translocation mechanism seems plausible (Fig. 5h and i). Because HA is negatively charged, the membrane's electrochemical potential and the conserved positive charges lining the TM channel likely also contribute to translocation. Our MD simulations suggest that an unfavorable register with GlcA at the acceptor position may be a transient intermediate that facilitates polymer translocation upon UDP-GlcNAc binding (Fig. 5h and i).

#### **Acknowledgments:**

We are grateful for support from K. Dryden and M. Purdy from the MEMC at the University of Virginia as well as Adam Wier and support staff at the NCI-NCEF. We thank James Van Etten for providing *Chlorella* virus genomic DNA. E.P. and J.S. acknowledge the support of Instruct-ERIC, part of the European Strategy Forum on Research Infrastructures (ESFRI), and the Research Foundation - Flanders (FWO) for supporting the nanobody discovery and thank Eva Beke for technical assistance. J.Z. and F.M. were supported by NIH grant R21AI148853. R.H. was supported by NIH grant R01GM101001 (awarded to J.Z.). R.A.C. and P.J.S. are supported by Wellcome (208361/Z/17/Z). P.J.S.'s lab is supported by awards from the BBSRC (BB/P01948X/1, BB/R002517/1 and BB/S003339/1) and MRC (MR/S009213/1). P.J.S. acknowledges the University of Warwick Scientific Computing Research Technology Platform for computational access.

#### **Author contributions:**

Y.B. cloned Cv-HAS and generated materials for Nb production. E.P. and J.S. generated the Nbs. Y.B. and L.M. characterized the Nbs. L.M. performed thermo-stability assays. F.M. purified Cv-HAS-Nb complexes and produced nanodiscs. R.H. collected EM data,

370 F.M. and J.K. processed the data. F.M. determined the apo and UDP-bound Cv-HAS  
371 structures. J.K. determined the substrate-bound and primed Cv-HAS structures. R.A.C.  
372 performed all MD simulations. J.Z. wrote the first manuscript, F.M., J.K., R.A.C., P.J.S.,  
373 J.Z. edited the draft, and all authors commented on the manuscript.

374

375 **Competing interests:**

376 The authors declare no competing interests.

377

378 **Materials and correspondence:**

379 Correspondence should be addressed to: jochen\_zimmer@virginia.edu

380

381

382

## References:

- 1 Girish, K. S. & Kemparaju, K. The magic glue hyaluronan and its eraser hyaluronidase: a biological overview. *Life Sci* **80**, 1921-1943, (2007).
- 2 Cyphert, J. M., Trempus, C. S. & Garantzotis, S. Size Matters: Molecular Weight Specificity of Hyaluronan Effects in Cell Biology. *Int J Cell Biol* **2015**, 563818, (2015).
- 3 Vigetti, D. *et al.* Hyaluronan: biosynthesis and signaling. *Biochim Biophys Acta* **1840**, 2452-2459, (2014).
- 4 DeAngelis, P. Hyaluronan synthases: fascinating glycosyltransferases from vertebrates, bacterial pathogens, and algal viruses. *Cell Mol Life Sci* **56**, 670-682, (1999).
- 5 Sironen, R. *et al.* Hyaluronan in human malignancies. *Exp Cell Res* **317**, 383-391, (2011).
- 6 Tammi, R. H. *et al.* Hyaluronan in human tumors: pathobiological and prognostic messages from cell-associated and stromal hyaluronan. *Semin Cancer Biol* **18**, 288-295, (2008).
- 7 Itano, N. & Kimata, K. Altered hyaluronan biosynthesis in cancer progression. *Semin Cancer Biol* **18**, 268-274, (2008).
- 8 Cowman, M. K., Lee, H.-G., Schwertfeger, K. L., McCarthy, J. B. & Turley, E. A. The Content and Size of Hyaluronan in Biological Fluids and Tissues. *Front Immunol* **6**, 261, (2015).
- 9 Hubbard, C., McNamara, J., Azumaya, C., Patel, M. & Zimmer, J. The hyaluronan synthase catalyzes the synthesis and membrane translocation of hyaluronan. *J Mol Biol* **418**, 21-31, (2012).
- 10 Weigel, P. H. & Deangelis, P. L. Hyaluronan synthases: a decade-plus of novel glycosyltransferases. *J Biol Chem* **282**, 36777-36781, (2007).
- 11 DeAngelis, P. L., Jing, W., Drake, R. R. & Achyuthan, A. M. Identification and molecular cloning of a unique hyaluronan synthase from *Pasteurella multocida*. *J Biol Chem* **273**, 8454-8458, (1998).
- 12 Camenisch, T. D. *et al.* Disruption of hyaluronan synthase-2 abrogates normal cardiac morphogenesis and hyaluronan-mediated transformation of epithelium to mesenchyme. *J Clin Invest* **106**, 349-360, (2000).
- 13 Itano, N. *et al.* Three isoforms of mammalian hyaluronan synthases have distinct enzymatic properties. *J Biol Chem* **274**, 25085-25092, (1999).
- 14 DeAngelis, P., Jing, W., Graves, M., Burbank, D. & Van Etten, J. Hyaluronan synthase of chlorella virus PBCV-1. *Science* **278**, 1800-1803, (1997).
- 15 Blackburn, M. R. *et al.* Distinct reaction mechanisms for hyaluronan biosynthesis in different kingdoms of life. *Glycobiology* **28**, 108-121, (2018).
- 16 Pardon, E. *et al.* A general protocol for the generation of Nanobodies for structural biology. *Nat Protoc* **9**, 674-693, (2014).
- 17 Tlapak-Simmons, V. L., Baron, C. A. & Weigel, P. H. Characterization of the purified hyaluronan synthase from *Streptococcus equisimilis*. *Biochemistry* **43**, 9234-9242, (2004).
- 18 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, (2017).
- 19 Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: structures, functions, and mechanisms. *Annu Rev Biochem* **77**, 521-555, (2008).



427 20 Cantarel, B., Coutinho, P., Rancurel, C. & Bernard, T. The Carbohydrate-Active EnZymes  
428 database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **37**, D233-238,  
429 (2009).

430 21 Wu, Q. *et al.* Protein contact prediction using metagenome sequence data and residual  
431 neural networks. *Bioinformatics* **36**, 41-48, (2020).

432 22 Tunyasuvunakool, K., Adler, J., Wu, Z., Jumper, J. & Hassabis, D. Highly accurate protein  
433 structure prediction for the human proteome. *Nature*, <https://doi.org/10.1038/s41586-41021-03828-41581>, (2021).

434  
435 23 Baek, M. Accurate prediction of protein structures and interactions using a 3-track  
436 network. *bioRxiv* <https://doi.org/10.1101/2021.06.14.448402> (2021).

437 24 Charnock, S. J. & Davies, G. J. Structure of the nucleotide-diphospho-sugar transferase,  
438 SpsA from *Bacillus subtilis*, in native and nucleotide-complexed forms. *Biochemistry* **38**,  
439 6380-6385, (1999).

440 25 Morgan, J., Strumillo, J. & Zimmer, J. Crystallographic snapshot of cellulose synthesis  
441 and membrane translocation. *Nature* **493**, 181-186, (2013).

442 26 Purushotham, P., Ho, R. & Zimmer, J. Architecture of a catalytically active homotrimeric  
443 plant cellulose synthase complex. *Science (New York, N Y)* **369**, 1089-1094, (2020).

444 27 Morgan, J. L. W., McNamara, J. T. & Zimmer, J. Mechanism of activation of bacterial  
445 cellulose synthase by cyclic di-GMP. *Nature Struct Mol Biol* **21**, 489-496, (2014).

446 28 McManus, J., Yang, H., Wilson, L., Kubicki, J. & Tien, M. Initiation, Elongation, and  
447 Termination of Bacterial Cellulose Synthesis. *ACS Omega* **3**, 2690-2698, (2018).

448 29 Orlean, P. & Funai, D. Priming and elongation of chitin chains: Implications for chitin  
449 synthase mechanism. *Cell Surf* **5**, 100017, (2019).

450 30 Yang, J. *et al.* Key Role of the Carboxyl Terminus of Hyaluronan Synthase in Processive  
451 Synthesis and Size Control of Hyaluronic Acid Polymers. *Biomacromolecules* **18**, 1064-  
452 1073, (2017).

453 31 Morgan, J. L. *et al.* Observing cellulose biosynthesis and membrane translocation in  
454 crystallo. *Nature* **531**, 329-334, (2016).

455 32 PyMol. *PyMol*, <<https://pymol.org/2/>> (<https://pymol.org/2/>).

456 33 Ho, B. & Gruswitz, F. HOLLOW: generating accurate representations of channel and  
457 interior surfaces in molecular structures. *BMC Struct Biol* **8**, 49, (2008).

458 34 Studier, F. Protein production by auto-induction in high density shaking cultures. *Protein*  
459 *Expr Purif* **41**, 207-234, (2005).

460 35 Kelly, S. J., Taylor, K. B., Li, S. & Jedrzejewski, M. J. Kinetic properties of *Streptococcus*  
461 *pneumoniae* hyaluronate lyase. *Glycobiology* **11**, 297-304, (2001).

462 36 Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from  
463 electron micrographs. *J Struct Biol* **192**, 216-221, (2015).

464 37 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta*  
465 *Crystallogr D Biol Crystallogr* **60**, 2126-2132, (2004).

466 38 Adams, P. *et al.* PHENIX: a comprehensive Python-based system for macromolecular  
467 structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221, (2010).

468 39 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and  
469 analysis. *J Comput Chem* **25**, 1605-1612, (2004).

- 470 40 Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically  
471 disordered proteins. *Nature* **14**, 71-73, (2016).
- 472 41 Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface  
473 for CHARMM. *J Comput Chem* **29**, 1859-1865, (2008).
- 474 42 Lee, J. *et al.* CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM,  
475 and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J Chem*  
476 *Theo Comput* **12**, 405-413, (2016).
- 477 43 Sondergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved  
478 Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of  
479 pK(a) Values. *Journal of Chemical Theory and Computation* **7**, 2284-2295, (2011).
- 480 44 Olsson, M. H. M., Sondergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3:  
481 Consistent Treatment of Internal and Surface Residues in Empirical pK(a) Predictions.  
482 *Journal of Chemical Theory and Computation* **7**, 525-537, (2011).
- 483 45 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J*  
484 *Chem Phys* **126**, 014101, (2007).
- 485 46 Parrinello, M., Rahman, A. Polymorphic transitions in single crystals: A new molecular  
486 dynamics method. *J Appl Phys* **52**, 7182-7190, (1981).
- 487 47 Berendsen, H. J. C., van der Spoel, D., van Drunen, R. GROMACS: A message-passing  
488 parallel molecular dynamics implementation. *Comp Phys Comm* **91**, 43-56, (1994).
- 489 48 Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph*  
490 **14**, 33-38, 27-38, (1996).
- 491 49 Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Computing in Science &*  
492 *Engineering* **9**, 90-95, (2007).
- 493 50 Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit  
494 for the analysis of molecular dynamics simulations. *J Comp Chem* **32**, 2319-2327, (2011).
- 495 51 Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357-362, (2020).
- 496 52 Bernsel, A., Viklund, H., Hennerdal, A. & Elofsson, A. TOPCONS: consensus prediction of  
497 membrane protein topology. *Nucleic Acids Res* **37**, W465-468, (2009).
- 498 52 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J*  
499 *Chem Phys* **126**, 014101, (2007).

## Figure legends

**Fig. 1l Structure of hyaluronan synthase.** (a and b) HA GlcA-GlcNAc disaccharide repeat unit and representation of an HA polymer (blue square: GlcNAc, blue/white diamond: GlcA). (c) Overall volume of the Cv-HAS-Nb872/881 complex with Nb872 and 881 in salmon and yellow, respectively, Cv-HAS in green, with the nanodisc at lower contour represented as an outline. (d) as well as representative 2D class averages of nanobody-bound Cv-HAS. (e) Apo conformation of Cv-HAS. The glycosyltransferase (GT), interface (IF), and TM regions are colored green, light blue, and blue, respectively. Manganese ions are shown as purple spheres. Bottom panel: Schematic of Cv-HAS' domain organization and Nb interactions. (f) Surface electrostatics of Cv-HAS. The enzyme forms an electropositive crescent-shaped channel. Partially ordered lipid and detergent molecules occupying the membrane exposed channel's opening are shown as yellow ball-and-sticks. The electrostatic potential was calculated using the APBS plugin in PyMol (red – blue: -5 to +5 kT)<sup>32</sup>.

**Fig. 2l Substrate-bound and primed Cv-HAS conformations.** (a) 3D variability analysis of UDP-GlcNAc-incubated Cv-HAS particles reveals substrate-bound and primed states. Shown are representative volumes of the generated ensemble (Supplementary Video 1). Left panel: Substrate and acceptor locations mapped onto one structure; Right panel: Rotation of the GT domain towards the TM region upon substrate binding. (b and c) Substrate coordination at the active site. Dashed lines indicate distances from ~2.5 to ~4.2 Å. (d) Priming loop retraction in response to substrate binding. The loop is shown as a cartoon with the flanking residues shown as sticks and semi-transparent surfaces. (e) Contribution of conserved residues to catalytic activity normal to wild type (WT) levels and obtained by quantifying the production of <sup>3</sup>H-labeled HA. Inset: Western blot of the inverted membrane vesicles used for activity measurements. Error bars represent deviations from the means obtained from three independent replicas. (f) Conformational changes of the switch loop. The loop moves towards the membrane interface in the presence of UDP-GlcNAc. The primer and substrate are shown as ball-and-sticks in the same structure.

**Fig. 3l GlcNAc priming of HA biosynthesis.** (a) Shown is an autoradiogram of  $^{14}\text{C}$ -labeled HA after SDS-PAGE. (b) Quantification of substrate hydrolysis rates in the presence of increasing monosaccharide concentrations. Rates are normalized relative to reactions without monosaccharide and after background subtraction. Blue and Red: Hydrolysis of UDP-GlcNAc and UDP-GlcA, respectively. Error bars represent deviations from the means. See also Extended Data Figure 9e.

**Fig. 4l Structural rearrangements upon GlcNAc priming.** (a) Movement of TMH2 (green= substrate bound, blue= primed-open) upon substrate hydrolysis. (b) The outward movement of TMH2 creates a continuous TM channel. TMH1 is shown at its predicted location as a green cylinder. (c) Partial overlap of the lipid-plugged channel with the continuous TM channel formed upon priming (shown as an orange surface). The predicted location of TMH1 is indicated as a green cartoon helix. The channels were rendered in HOLLOW using a 1.2 Å probe<sup>33</sup>. (d) Aromatic, charged, and polar residues lining the TM-channel in the open state.

**Fig. 5l Dynamics of HA-bound HAS.** (a) MD simulations of Cv-HAS bound to either UDP-GlcNAc and a GlcA-ending HA hexasaccharide or UDP-GlcA and a GlcNAc-ending polymer. The GlcA-ending polymer migrates into the channel by one sugar unit. (b-c) Representative coordination of HA at different registers after 300 ns MD simulations. (d) Coordination of HA inside the TM channel in a UDP-GlcNAc bound state. Values represent contact probabilities in percent over 300 ns of simulation, averaged over 5 independent repeats. The N-acetyl binding pocket is highlighted in blue. (e) As in panel d but for an UDP-GlcA bound state and HA polymer with different register. (f) Coordination of UDP-GlcA at the active site. Shown is a representative pose from a 300 ns simulation. The contact analysis is shown in Supplementary Table 2. (g) Mutagenesis of HA coordinating residues and corresponding catalytic activity. Inset: Anti His-tag Western blot of IMVs used for activity assays. Error bars represent deviations from the means of three replicas. (h and i) Model of HA elongation and membrane translocation. (h) HA's terminal sugar unit determines the selectivity of the

catalytic pocket such that GlcNAc or GlcA acceptors only allow binding UDP-GlcA or UDP-GlcNAc substrate, respectively. (i) UDP-GlcNAc hydrolysis and outward movement of TMH2 to prime HAS. Right: Processive primer extension and translocation leads to HA secretion. HA translocation is likely coupled to substrate binding. The priming loop (shown in pink) retracts from the active site to enable substrate binding, while the GT contracts and relaxes upon substrate binding and HA-extension, respectively.

## Materials and Methods

### Construct Design and Mutagenesis

The gene for *Paramecium bursaria Chlorella virus CZ-2* hyaluronan synthase (Cv-HAS) was cloned as described<sup>15</sup> and sub-cloned after restriction enzyme digest with NcoI/XhoI into a pET28a vector modified to encode a C-terminal 10x histidine tag (pHAScz2). Mutagenesis was performed by PCR with overlapping primers (Supplementary Table 3) on pHAScz2 using KOD Hot Start polymerase (Novagen). Constructs for deletions of the N-terminal TMH were generated by amplifying the vector and remaining Cv-HAS gene with primers containing complementary overhangs.

### Protein Expression and Purification

Cv-HAS was expressed in *E.coli* C43. Cultures were grown in terrific broth (TB) supplemented with 4% v/v glycerol and 1XM buffer<sup>34</sup>. Each culture was inoculated from an overnight culture in lysogeny broth (LB), grown to an OD<sub>600</sub> of ~0.8 at 30 °C, cooled 1 hour to 20 °C, then expression was induced with 100 mg/L Isopropyl β-D-1-thiogalactopyranoside (IPTG). Constructs were expressed ~18 hours at 20 °C, then cell pellets were harvested by centrifugation.

Cell pellets were resuspended in a buffer of 10% glycerol, 100mM NaCl, and 20mM Tris pH 7.5 (RB), then incubated for 1 hour at 4 °C with 1 mg/mL lysozyme (Alfa Aesar). Cell suspensions were processed by 3 passes through a microfluidizer, then intact cells and aggregated material was removed by low-speed centrifugation for 25 min at 12.5 kRPM in a JA-20 rotor (Beckman). Membranes were isolated from the resulting lysates by centrifugation for 2 hours at 200k x g in a Ti45 rotor (Beckman), then harvested and flash-frozen in liquid N<sub>2</sub> and stored at -80 °C.

For *in surfo* samples, membranes were thawed and resuspended in 300 mM NaCl, 20 mM Tris 7.5, 10% glycerol, 5mM β-mercaptoethanol (BME) (SB), supplemented with 40 mM imidazole, 1% lauryl maltose neopentyl glycol (LMNG), and 0.2% cholesterol

hemisuccinate (CHS), then incubated 1 hour with agitation. Aggregated material was removed by centrifugation at 200k x g for 30 min, then the supernatant was incubated with Ni-NTA resin for 1 hour. Resin was washed with SB supplemented with 80 mM imidazole and the protein was eluted with SB supplemented with 320 mM imidazole. The eluted fraction was concentrated and purified by size-exclusion chromatography (SEC) using an S200-increase column equilibrated in RB. Peak fractions were harvested and concentrated, followed by glycerol removal by a second SEC in 100 mM NaCl and 20 mM Tris pH 8.0 (FB) (all steps performed at 4°C).

Purification of nanodisc samples was similar, except membranes were solubilized in SB supplemented with 40 mM imidazole, 1 % n-Dodecyl  $\beta$ -D-maltoside (DDM), and 0.1 % CHS, the Ni-NTA wash and elution buffers were supplemented with 0.02 % DDM/0.002 % CHS, and the first SEC buffer was 100 mM NaCl, 20 mM Tris pH 7.5, 0.02 % DDM and 0.002 % CHS. Following the first SEC, protein was concentrated and reconstituted into nanodiscs in a 1:3:30 molar ratio of HAS:MSPE3D1:*E.coli* total lipid extract (solubilized in 60 mM DDM). To produce nanobody complexes, selected nanobodies were added in two-fold molar excess compared to HAS to the nanodisc reconstitution mixture. After a 30 minute incubation, BioBeads were added stepwise to a final volume of 0.5 mL and incubated overnight to sequester detergent. Properly formed HAS nanodiscs were purified by SEC in FB, then fractions were screened by SDS-PAGE and negative stain EM for presence of both MSP and HAS and particle quality (all steps performed at 4°C).

### *In vitro* HA Synthesis Assays

HA synthesis was assayed by incorporation of  $^3\text{H}$ - or  $^{14}\text{C}$ -labeled sugars into HA produced by Cv-HAS in inverted membrane vesicles (IMVs), as described<sup>15</sup>. Wild type or mutated HAS constructs were expressed as described above. Lysates resulting from the low-speed spin were floated on a 2 M sucrose cushion, then spun for 2 hours at 200k x g in a Ti45 rotor. IMVs were harvested, diluted in RB, and centrifuged again at

200k x g for 1.5 hours. The resulting IMV pellet was harvested, homogenized with a dounce in RB, and flash-frozen in liquid N<sub>2</sub>.

Cv-HAS levels in each IMV isolate were quantified by Western blotting. 200 µL IMVs were solubilized by addition of 800 µL SB containing 1 % DDM, 0.1 % CHS, and 20 mM Imidazole, then incubated 1 hour with Ni-NTA resin. The resin was washed twice in 250 µL SB containing 0.03 % DDM, 0.003 % CHS, and 20 mM Imidazole. The washed resin was then resuspended in Laemmli buffer and 1 M imidazole, then analyzed on a 12.5 % polyacrylamide gel, and transferred at 100 V to a nitrocellulose membrane. The membrane was washed 5 min in dH<sub>2</sub>O, blocked twice for 15 minutes with 5 % milk in Tris-buffered saline-tween buffer (TBST), and incubated with primary antibody for 1 hour (all at room temperature). The membrane was rinsed 15 mins in TBST, then incubated for 1 hour with IRDye800-conjugated secondary antibody, then rinsed again for 15 mins in TBST. Blots were imaged using an Odyssey Licor scanner at 700 and 800nm, and analyzed using imageJ to determine relative protein content across multiple Cv-HAS variants.

Prior to each assay, IMVs were diluted to normalize Cv-HAS concentration. 10 µL of each IMV were mixed with 10 µL 2X reaction buffer, containing 40 mM MnCl<sub>2</sub>, 80 mM Tris pH 7.5, 150 mM NaCl, 10 mM TCEP, 10 mM UDP-GlcNAc, 10 mM UDP-GlcA, and 0.2 µCi of the appropriate radiolabeled sugar (listed below with each experiment). Reaction mixtures were incubated 2 hours at 30 °C then analyzed.

For descending paper chromatography, the radiolabel was <sup>3</sup>H-UDP-GlcNAc and each reaction was quenched with 2% SDS. 20 µL of each reaction mixture was spotted onto Whatman 2mm filter paper, developed with 65% 1 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>/35 % ethanol, and the origins were counted using a Beckman S Ls liquid scintillation counter. Activity of each variant was repeated in triplicate, from which background was subtracted as determined by a control reaction lacking UDP-GlcA.

For autoradiography comparison of HA sizes produced across Cv-HAS variants, the radiolabel was <sup>14</sup>C-UDP-GlcA. Reactions were quenched by addition of loading buffer, then run at 100 V for 2 hours on a 12.5 % polyacrylamide gel. The gel was fixed using



30 % methanol/3 % glycerol for 15 mins and dried onto a Whatman 2mm filter paper using a Biorad model 583 gel dryer for 2 hours. The dried gel was exposed for ~1 week to a Kodak Biomax MS film. Autoradiography of the initiation experiments was identical, except the radiolabels were <sup>14</sup>C-labeled GlcNAc or GlcA (not UDP activated). Hyaluronate lyase digestions were performed by adding 1 mg/mL final concentration of recombinantly expressed *S. pneumoniae* hyaluronate lyase<sup>35</sup> to the relevant reaction mixture.

## Nanobody generation

Camelid nanobodies against Cv-HAS were generated *in vivo* by immunizing a male llama weekly over a period of 6 weeks with purified wild type and catalytically inactive (D201A) Cv-HAS. Approximately 1 mg of total protein reconstituted into *E. coli* total lipid proteoliposomes was used for immunization. The isolation of high affinity Cv-HAS binders and nanobody cloning into the pMESy4 expression vector followed previously described procedures<sup>16</sup>. Briefly, Cv-HAS wild type, Cv-HAS wild type with synthesized hyaluronan, and Cv-HAS D201A reconstituted into *E. coli* total lipid proteoliposome were each solid phase coated and used as target for the biopanning in 25 mM MES pH 6.5, 0.2 M NaCl, 10 % glycerol, 5 mM beta-mercaptoethanol, 3.5 mM MnCl<sub>2</sub>. After rescue of the eluted phage, individual colonies were screened in ELISA using the same buffer conditions. Thirty-eight different Cv-HAS-specific nanobody families were discovered belonging to 28 families based on their CDR3 sequences.

## Nanobody purification

Constructs for each nanobody gene cloned in pMESy4 were expressed in *E. coli* WK6 cultures grown in TB supplemented with 4% v/v glycerol and 1XM buffer. Periplasmic fractions of the harvested cell pellets were extracted by osmotic shock using a Tris-sucrose-EDTA buffer (TES). Periplasmic extracts were incubated with Ni-NTA resin, then nonspecifically bound material was removed by successive washes in 20 mM Tris

pH 7.5 containing 1 M NaCl and 20 mM Imidazole, then 0.1 M NaCl and 40 mM Imidazole, and 0.1 M NaCl, 320 mM Imidazole to elute. Eluted protein was further purified by SEC using an S75 column equilibrated in FB, then concentrated and flash-frozen in liquid N<sub>2</sub>.

#### Thermo-stability assays

Stability measurements were based on monitoring Cv-HAS' enzymatic activity following incubation at elevated temperatures. The activity was estimated by quantifying the release of UDP in real time using an enzyme-coupled reaction that oxidizes NADH, as previously described<sup>9</sup>. Cv-HAS was individually incubated with the 38 nanobodies in a thermo-cycler at temperatures ranging from 30 to 66°C for 2 hours in 1:3 molar ratio. Biosynthesis reactions were performed in 120  $\mu$ L volumes containing 1 mM PEP, 0.75 mM NADH, and 1 U of pyruvate kinase and lactate dehydrogenase each.

The oxidation of NADH was monitored at room temperature based on its absorbance at 340 nm in a microplate spectrophotometer in 30 s intervals for 1 hour. Reaction rates were calculated based on Lambert-Beer's law using an NADH extinction coefficient at 340 nm of 6220 (M cm)<sup>-1</sup>. All experiments were performed in triplicate and error bars represent the deviations from the means.

#### Substrate hydrolysis assays

Hydrolysis of UDP-GlcNAc or UDP-GlcA was measured using an enzyme-coupled assay as described above, Fig S11 A-B. 1  $\mu$ M Cv-HAS in nanodiscs was incubated with 2 mM substrate and the indicated concentration of GlcNAc or GlcA monosaccharide. Background hydrolysis was measured by omitting Cv-HAS. For each combination of substrate and monosaccharide, Cv-HAS catalyzed and background hydrolysis rates were measured in triplicate (Extended Data Fig. 9c-e). Net hydrolysis rates were

normalized separately for each substrate based on reactions with no monosaccharide  
(Extended Data Fig. S9f).

## EM Sample Preparation and Data Collection

Protein samples were concentrated using a 50 kDa filter (Amicon) to 3-4 mg/mL. For relevant samples  $\text{MnCl}_2$  and UDP or UDP-GlcNAc were added to a final concentration of 2 mM. 3  $\mu\text{L}$  samples in detergent were applied to glow-discharged C-flat holey carbon grids (Cu 1.2/1.3, 300 mesh), blotted for 10 seconds at 4°C and 100 % humidity, then plunge frozen in liquid ethane using a Vitrobot Mark IV (FEI). 3  $\mu\text{L}$  samples in nanodiscs were applied to Quantifoil holey carbon grids (Cu 1.2/1.3, 300 mesh) glow discharged with amylamine, blotted for 4 s at 4 °C and 100 % humidity, then plunge frozen in liquid ethane.

Cryo-EM data were collected at the University of Virginia Molecular Electron Microscopy Core (MEMC) on a Titan Krios (FEI) 300-kV electron microscope using a Gatan Imaging Filter (GIF) and a K3 direct detection camera. Movies were collected at a magnification of 81,000x with an energy filter width of 10 eV, using counting mode with a total dose of 51  $\text{e}^-/\text{\AA}^2$  over 40 frames, and with a target defocus of -1.0 to -2.0  $\mu\text{m}$ , Table S1. Several earlier (unreported) datasets were collected under similar conditions at the National Cryo-Electron Microscopy Facility (NCEF) at the National Cancer Institute (NCI).

## EM Data Processing and Model Building

All data processing steps were performed in cryoSPARC<sup>18</sup>, unless otherwise noted. Movies were imported and gain corrected, then subjected to patch-based motion-correction and contrast transfer function estimation (CTFFIND4)<sup>36</sup>. Initially, particles were manually picked from accepted micrographs to generate references for template-based particle picking. References for later datasets were based on 2D classes from earlier sets. Particles identified from template-based picking were extracted using 2-fold

Fourier cropping for initial rounds of 2D classification and/or *ab initio* reconstruction followed by heterogeneous refinement to remove junk particles. Accepted particles were re-extracted at full scale then subjected to further rounds of classification. High-resolution reconstructions from clean particle sets were generated by non-uniform (NU) refinement, then masked local refinement (using NU algorithm), followed by local CTF estimation and refinement of the corrected particles. Finally, 3D variability analysis and subsequent clustering was used to separate populations within the high-resolution reconstructions belonging to distinct catalytic states. Clusters belonging to apo, UDP-bound, substrate bound, or primed states were individually pooled and used to generate final reconstructions from NU or local refinement. Two maps for the apo state were derived by 3D variability analyses from the dataset collected for wild-type Cv-HAS in the presence of UDP and from the dataset for D302N Cv-HAS collected in the presence of UDP-GlcNAc. The wild-type apo model (CC mask 0.68) and D302N apo model (CC mask 0.75) are virtually identical with an RMSD of 0.132 Å. Reported is the model derived from the UDP-GlcNAc dataset, which extends to 3.1 Å versus 3.3 Å.

The initial atomic model was built into a ~3.5 Å map, derived from collection on an apo Cv-HAS sample *in surfo*, starting from a poly-alanine model of the RsBcsA GT domain. The model was generated from iterative rounds of building in Coot<sup>37</sup> and real-space refinement in PHENIX<sup>38</sup>. Register was assigned based on well-resolved map regions and on the location of highly conserved catalytic motifs. TM helices were built *de novo* with register interpreted from density for bulky residues and based on connectivity to the GT domain. The initial model was refined and register errors corrected using subsequent higher resolution maps *in surfo* and in nanodiscs to generate the final models. All figures were prepared using PyMol and Chimera<sup>32,39</sup>.

## Molecular Dynamics Simulations

The coordinates of Cv-HAS, without nanobodies, obtained in nanodiscs were used for building atomistic MD simulations. Where included, the substrate coordinates were taken from the resolved cryo-EM densities for UDP-GlcNAc, and the terminal residue of

the HA chain placed according to the position of the primer monosaccharide. Three different HA poses were manually constructed in Coot<sup>37</sup> for each register, to provide a broad degree of HA conformational sampling. The systems were described with the CHARMM36m force field<sup>40</sup> and built into model POPE membranes with TIP3P waters and K<sup>+</sup> and Cl<sup>-</sup> to 150 mM, using CHARMM-GUI<sup>41,42</sup>. The bound Mn<sup>2+</sup> were swapped to Mg<sup>2+</sup>. The final systems had approximately 300 lipids, 150 ions, and 27500 waters, for a total of ca. 125,000 atoms and a box size of 10 x 10 x 12.5 nm. All titratable side chains were set to their default protonation state, based on analysis run with propKa3.1<sup>43,44</sup>.

Each system was minimized using the steepest descents method, then equilibrated in 6 rounds with restraints initially applied to protein backbone and sidechain atoms, lipid phosphate atoms and selected protein, lipid and sugar dihedrals, as per the standard CHARMM-GUI output. Where the HA chain was present, to optimise its positioning in the channel additional equilibration simulations of 90 ns were run with 50 kJ/mol/nm<sup>2</sup> positional restraints applied to the protein backbone and to the bound substrate, but not to the HA chain. Production simulations were then seeded using the output frame of this final equilibration, with 3 or 5 repeats seeded for each system. Simulations were run using 2 fs time steps, with V-rescale thermostat at 303.15 K with a tau t of 1 ps, and semi-isotropic Parrinello-Rahman pressure coupling with tau p of 5 ps and a reference of 1 bar<sup>42,45</sup>. All simulations were run in Gromacs 2019.4<sup>45-47</sup>. In total, ca. 9  $\mu$ s of data were generated, with full details in Supplementary Table 4.

Lipid and solvent densities in the apo MD simulations were computed over the 5 repeats for each system using the VolMap tool of VMD<sup>48,49</sup>. HA-residue contact analysis was determined for each frame of the simulation based on a distance cut-off of 0.4 nm between each residue in the system and the specified sugar, and was run using MDAnalysis<sup>49</sup>. Quantification of the occupancy of the binding pocket using in Extended Data Fig. 11 was performed by counting the number of frames that the acetamido group of the donor or acceptor GlcNAc was within 0.4 nm of any atom in Cys-231, and was run using the Gromacs tool gmx mindist. Analysis of the HA chain position in relation to the channel was calculated using the Gromacs tool gmx distance. The distance in the z-

797 dimension between the centre-of-mass of each sugar and the centre-of-mass of the  
798 protein was computed, and plotted using NumPy<sup>50</sup> and Matplotlib<sup>51</sup>.

799

800

801 **Data availability**

802 Raw EM movies and maps have been deposited at the PDB and EM data banks under  
803 accession codes 7SP7/EMD-25367, 7SP6/EMD-25366, 7SP8/EMD-25368, 7SP9/  
804 EMD-25369, and 7SPA/EMD-25370 for the UDP-bound, D302N apo, UDP-GlcNAc-  
805 bound, primed (closed), and primed (open) states, respectively.

806

807

## Extended Data Figure Legends

**Extended Data Fig. 1| Sequence alignment of HAS orthologs.** Comparison of HAS primary sequences from *Chlorella virus* (Cv), *Homo sapiens* (Hs) and *Streptococcus equisimilis* (Se). Naked mole rats (Nmr) contain an asparagine residue at Cv-HAS positions S165 and T289. Topology predictions were performed using TopCons<sup>52</sup>. Cylinders indicate secondary structure elements observed in Cv-HAS.

**Extended Data Fig. 2| Impact of nanobodies on Cv-HAS stability and catalytic activity.** (a) Increased melting temperature of Cv-HAS in the presence of Nb872. Protein melting was measured based on enzymatic activity detected by quantifying the release of UDP in real time. (b) HA biosynthesis in the presence of the indicated nanobodies and based on quantification of <sup>3</sup>H-labeled HA by scintillation counting. Data is normalized relative to product yields in the absence of nanobodies. (c) Autoradiography of <sup>14</sup>C-labeled HA produced in the presence of the indicated nanobodies. NC: Negative control in the absence of UDP-GlcNAc substrate (for panel b) or UDP-GlcA (for panel c). PC: Positive control in the absence of nanobody. Lyase: Hyaluronan lyase treatment prior to SDS-PAGE.

**Extended Data Fig. 3| Cryo-EM data collection and processing for UDP-bound and apo wild-type Cv-HAS.** This workflow produced the UDP-bound Cv-HAS structure.

**Extended Data Fig. 4| Cryo-EM data collection and processing of Cv-HAS D302N in the presence of substrate.** This workflow generated the apo, substrate-bound, primed, and primed with open channel Cv-HAS structures.

**Extended Data Fig. 5| Map quality and model building of UDP-bound Cv-HAS.** (a-d) Map overview, estimated resolution based on FSC, and particle orientation distribution. (e) Secondary structure elements and topology of Cv-HAS. (f-j) TM helices 2 to 6 of Cv-HAS. (k) TMH3-4 extracellular loop. (l) The extracellular TMH5-6 loop. (m) The QxxRW motif. (n) The C-terminal cytosolic helix. (o) The unresolved TMH5-IF3 loop. All maps are contoured at 7.0 $\sigma$ .

**Extended Data Fig. 6I Predicted location of TM helix 1.** (a) Relationship of evolutionarily coupled residues within Cv-HAS' TM and GT regions, generated in MapPred based on 65,535 sequences. TMH1 is shown at its predicted location as a violet cylinder. (b) RoseTTAfold models of full-length Cv-HAS. Cv-HAS is shown as a surface and its TMH 2 as a blue cylinder. TMH1 is shown as a cartoon at its predicted locations. (c) An AlphaFold2 predicted structure of human HAS-2 (colored blue to red from its N- to C-terminus) overlaid with the Cv-HAS structure shown as a gray cartoon and semi-transparent surface. (d) TMH1 remains disordered when two cytosolic Nbs are used for cryo-EM analyses. (e) Catalytic activity of TMH1 truncated Cv-HAS. Left: Western blot of IMVs used for *in vitro* activity measurements. Right: Catalytic activity of the indicated Cv-HAS mutants expressed relative to the wild type enzyme. The assay quantifies <sup>3</sup>H-labeled HA by scintillation counting. Control reactions in the absence of UDP-GlcA served as background and are subtracted. Error bars represent deviations from the means of three replicas.

**Extended Data Fig. 7I Lipids plug the lateral channel opening.** (a) Representative map regions for modeled lipids contoured at 7.0 $\sigma$  (from the UDP-GlcNAc bound set). (b) 2D slice from MD simulations of Cv-HAS (black area) within a POPE bilayer. Water and lipid densities are colored blue and green, respectively. Right panel: Lipid contact times with selected channel residues. (c) Comparison of the Cv-HAS (rainbow colored from the N- to C-terminus) and RsBcsA (gray, 4P00). Cellulose associated with BcsA is shown as black sticks. Helices are shown as cylinders except BcsA's N-terminal two TMHs, which are shown as coils.

**Extended Data Fig. 8I Details of substrate-binding and of priming-induced conformational changes.** (a and b) Map quality for UDP-GlcNAc, UDP, and Mn<sup>2+</sup> ligands. (c) Comparison of UDP and UDP-GlcNAc positions. (d) Map for the priming loop in nucleotide bound states. (e) Representative map for the GlcNAc primer. (f) Map for the priming loop in the primed states. (g) Contact point of TMH2 (open in blue, closed in gray) with IF1 in the primed state. (h and i) Map quality for TMH2 in a closed position (UDP-GlcNAc bound) and open position. All maps are contoured at 7.0 $\sigma$ .



**Extended Data Fig. 9| Effect of monosaccharides on substrate hydrolysis.** (a and b) Reaction schemes for UDP-GlcA and UDP-GlcNAc hydrolysis. (c and d) Raw absorbance measurements. (e) Quantification of hydrolysis rates in the presence of increasing monosaccharide concentrations. Blue and Red: Hydrolysis of UDP-GlcNAc and UDP-GlcA, respectively. Light and dark colors represent control reactions in the absence of enzyme. Error bars represent deviations from the means.

**Extended Data Fig. 10| Likely mechanism of alternating substrate polymerization and comparison with cellulose synthase.** (a) Superimposition of substrate-bound and primed Cv-HAS structures. The close distance between the primer and donor sugar is indicated by gray bars. (b) Cv-HAS is superimposed with the *Rhodobacter sphaeroides* (Rs) BcsA-B complex (PDB: 4P00) based on secondary structure matching. Rs-BcsA-B is colored gray and Cv-HAS is colored blue and green for its TM and GT domains. The cellulose polymer associated with Rs-BcsA-B is shown as black sticks.

**Extended Data Fig. 11| Modulation of substrate selectivity.** (Left) Coordination of UDP-GlcNAc at the substrate binding pocket (based on experimental data). (Middle) Coordination of a terminal GlcNAc unit at the acceptor binding site (based on experimental data). (Right) Simulated coordination of a GlcNAc donor and acceptor sugar. In this case, both GlcNAc units are less likely to bind C231, and exhibit very high variance regarding binding poses. Contact times are shown as percentage over the simulations (shown in gray).









