Molecular basis of hyaluronan synthesis

Ireneusz Paweł Górniak
Brzeg, Poland

MS Biotechnology, University of Wrocław, 2019

A Dissertation presented to the Graduate Faculty of the
University of Virginia in Candidacy for the
Degree of Doctor of Philosophy

Department of Molecular Physiology and Biological Physics

University of Virginia
July, 2024

Abstract

Hyaluronan (HA) is an essential component of the vertebrate extracellular matrix (ECM). It is enriched in soft connective tissues, the vitreous of the eye, and cartilage. HA affects important physiological processes, including cell migration, proliferation, and adhesion. Accordingly, many pathological conditions from autoimmune diseases to infertility and cancer correlate with altered HA production levels. Yet the molecular basis of how HA is synthesized and exerts its physiological functions remains poorly understood.

HA's physiological activities depend on its length. High molecular weight (HMW) HA contributes to wound healing and has anti-inflammatory properties, whereas low molecular weight (LMW) HA is involved in cell migration, and in high concentration, pathologies such as cancer metastasis. Large quantities of unusually HMW HA are believed to be the reason behind cancer resistance and longevity of some subterranean rodents.

HA is a heteropolysaccharide of alternating units of N-acetylglucosamine and glucuronic acid synthesized by HA-synthase (HAS). HAS is an unusual enzyme for three reasons. Firstly, it recognizes its two uridine-diphosphate (UDP)-activated substrates through a single binding pocket. Second, it catalyzes formation of HA by incorporating the substrates into the growing polysaccharide in a strictly alternating fashion. Third, HAS translocates HA during synthesis across the plasma membrane through a channel formed by its membrane-embedded region for integration into the ECM. Vertebrates express three HAS isoforms that produce HA of different sizes. HAS isoform 2 is essential.

Determining the molecular basis for HA synthesis and length control requires detailed structural and functional insights. In my graduate research, I characterized a heterologously expressed vertebrate HAS to determine factors controlling substrate polymerization and product length. I used cryogenic electron microscopy to determine the first structural snapshots of a vertebrate HAS homolog and provided insights into HA biosynthesis and translocation. I revealed the coordination of the UDP product by a conserved gating loop and captured the opening of the translocation channel to coordinate the translocating HA polymer. Using site-directed mutagenesis studies, I identified active site- and channel-lining residues that modulate HA product lengths. By integrating structural and biochemical analyses, my research provided unprecedented insights into the biosynthesis of one of the most abundant extracellular polysaccharides in the human body and established the molecular basis for understanding its function in health and disease.

Signature page

Dedication page

I dedicate this work to my biggest supporters – my parents – Roman and Małgorzata. Thank you both.

List of Abbreviations

β-ME, β-mercaptoethanol
CBM, carbohydrate-binding module
CHS, cholesteryl hemisuccinate
CMC, critical micelle concentration
COMPPAA, The Center on Membrane Protein Production and Analysis
CryoEM, cryogenic electron microscopy
CTF, contrast transfer function
DDM, dodecyl maltoside
DIBMA, diisobutylene-maleic acid
DM, decyl maltoside
DPM, disintegrations per minute
EB, elution buffer
ECM, extracellular matrix
ELISA, enzyme-linked immunosorbent assay
EM, electron microscopy
EPS, extracellular polysaccharides
FSEC, fluorescence-detected size exclusion chromatography
GAG, glycosaminoglycan
GDN, glyco-diosgenin
GFB, gel filtration buffer
GFP, green fluorescent protein
GlcA, glucuronic acid
GlcNAc, N-acetylglucosamine
GPCR, G-protein coupled receptor
GS-FCS, gold standard Fourier shell correlation
GT, glycosyltransferase
HA, hyaluronic acid
HAS, hyaluronic acid synthase
HC, heavy chain
HEK, human embryonic kidney
HMW HA, high molecular weight hyaluronic acid
IFH, interface helix
IMAC, immobilized metal ion affinity chromatography
IMV, inverted membrane vesicle
IPTG, isopropyl β-D-1-thiogalactopyranoside
JCIMPT, Joint Center for Innovative Membrane Protein Technologies
K_M, Michaelis-Menten constant
LC, light chain
LDAO, dodecyl-dimethylamine-oxide
LDH, lactate dehydrogenase
LMNG, lauryl-maltose-neopentyl-glycol
LMW HA, low molecular weight hyaluronic acid
MEMC, molecular electron microscopy core
MSP, membrane scaffold protein
NAD+, oxidized nicotinamide adenine dinucleotide
NADH, reduced nicotinamide adenine dinucleotide
ND, nanodisc
NiNTA, nickel-nitrilotriacetic acid

OG, octyl-glucoside
PCR, polymerase chain reaction
PDB, protein data bank
PEP, phosphoenolpyruvate
PK, pyruvate kinase
PL, proteoliposome
PMSF, phenylmethylsulphonyl fluoride
SDS, sodium dodecyl sulfate
SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC, size exclusion chromatography
Sf9, *Spodoptera frugiperda*
SMA, styrene-maleic acid
TCEP, tris(2-carboxyethyl)phosphine
TEV, tobacco etch virus
TM, transmembrane
TMH, transmembrane helix
TX-100, TritonX-100
UDP, uridine diphosphate
UTP, uridine triphosphate
WT, wild type

List of Figures

List of Supplementary Figures and Tables

Table of Contents

1. Introduction

1.1 Complex carbohydrates and glycosyltransferases

Complex carbohydrates are among the most abundant biomolecules on the planet and are involved in many essential biological processes across all kingdoms of life. Most complex carbohydrates are synthesized from simple nucleotide-activated monomers by specific glycosyltransferases (GTs) [1].

GTs can be broadly classified based on their structure, amino acid sequence, mechanism of action, cation dependence, and membrane localization (Figure 1A). The catalytic domains of GTs adopt GT-A, GT-B and GT-C fold to create highly specific binding sites for their substrates. GTs catalyze the transfer of an activated glycosyl unit, the donor, to a receiver molecule, the acceptor. The donor substrates can range from nucleotide-activated to lipid-linked and even more complex carbohydrates. Acceptors on the other hand, can include other monosaccharides, polysaccharides, proteins and lipids [2].

The GT-A fold represents the most common architecture and consists of a globular domain that contains an α/β/α Rossmann domain (Figure 1B). GT-B fold GTs contain two α/β/α Rossmann domains connected via a flexible linker with a substrate binding pocket in between (Figure 1C). GT-C fold enzymes are the least abundant and represent multi spanning membrane proteins that use lipid-linked sugar donors [3] (Figure 1D). On the basis of their amino acid sequence, GTs can be further divided into families, of which there are currently 110 [4]. GT1 is the best-characterized family with the largest number of solved structures reported. This family primarily adopts the GT-B fold and includes a large number of glycosylating enzymes, especially enriched in bacteria and plants where they contribute to antibiotic glycosylation (a common mechanism of antibiotic resistance in bacteria) and synthesis of glycosylated polyphenols [5]. GT2 is the biggest family based on the number of predicted open reading frames [4] and is involved in the synthesis of, among others, extracellular polysaccharides (EPSs).

EPSs are of special interest as they constitute diverse architectural components of the cell surface, contributing to cell communication, protection and adhesion via the formation of cell walls, coats and capsules [6]. GTs can catalyze polysaccharide formation in a processive or distributive manner. Processive GTs hold on to the polymeric product and keep extending it, thereby allowing the formation of extremely long EPSs. Distributive GTs, on the other hand, release the product after each glycosyl transfer reaction [7]. The most abundant EPSs, cellulose, chitin, alginate, and hyaluronan, are predominantly synthesized by membrane-embedded processive GT-2 enzymes. These GTs

are characterized by their GT-A fold that contains signature sequence motifs responsible for cation binding (DxD), positioning of the acceptor sugar (QxxRW), and catalyzing the glycosyl transfer reaction (GDD) [1].

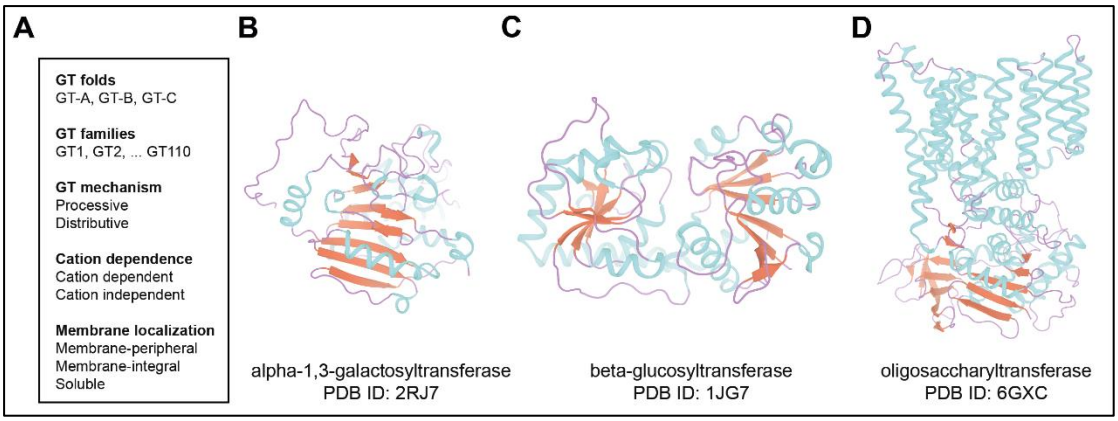


Figure 1 Glycosyltransferases and their classification

A – Different ways glycosyltransferases (GTs) can be classified. B – Structure of a representative GT-A fold GT – alpha-1,3-galactosyltransferase. C – Structure of a representative GT-B fold GT – beta-glucosyltransferase. D – Structure of a representative GT-C fold GT – oligosaccharyltransferase PglB. For panels B-D secondary structure elements are colored red, cyan and purple for α-helices, β-sheets and loops, respectively.

1.2 Hyaluronan is an important extracellular polysaccharide

Hyaluronan (hyaluronate, hyaluronic acid, HA) is ubiquitously produced in vertebrates where it forms an essential component of the extracellular matrix (ECM) of soft connective tissues, cartilage, as well as the endothelium. The acidic polymer is responsible for important physiological processes, such as cell migration, proliferation and adhesion [8]. Although the exact signaling mechanisms by which HA modulates physiological processes are unresolved to date, known cell surface HA receptors include CD44 and RHAMM, with established functions in hematopoiesis, organogenesis, and cell motility. Accordingly, a vast literature documents HA's impact on pathological conditions, including several forms of cancer, auto-immune diseases, chronic inflammations, and infertility [9-13]. Studies have also demonstrated many medicinal properties of HA injections aiding osteoarthritis symptoms, wound healing, eye surgeries, as well as a dermal filler [8].

HA is a linear extracellular polysaccharide of repeating disaccharide units consisting of N-acetylglucosamine (GlcNAc) and glucuronic acid (GlcA, Figure 2A). HA can reach ~7 MDa in size (roughly 40,000 sugar residues) and its biological functions correlate with polymer length. Healthy tissues usually contain polymers of several megadaltons in size. Inflammatory processes and osteoarthritis, however, associate with a decline in HA length to hundreds of kilodaltons [8]. How HA's length modulates its biological functions is currently unknown, yet receptor clustering on the cell surface may be one mechanism.

1.3 Hyaluronan synthases are a unique group of glycosyltransferases

Out of all the glycosaminoglycans (GAGs) present in the ECM, HA is unique because it is not synthesized in the Golgi apparatus on a protein core. Instead, HA is synthesized from uridine diphosphate-activated sugars (UDP-GlcNAc and UDP-GlcA) by a single plasma membrane-integrated GT-2 enzyme, called hyaluronan synthase (HAS) [14]. All vertebrate HASs are predicted to have 6 transmembrane helices (TMHs) and a cytosolic GT-A fold domain. HAS is unique among GTs because the enzyme utilizes two different substrates and can catalyze the formation of two different glycosidic linkages (beta-1,3 and beta-1,4 between GlcNAc-GlcA and GlcA-GlcNAc, respectively, Figure 2A, B). Furthermore, HAS secretes the HA polymer during synthesis through a membrane-spanning channel, such that HA synthesis is directly coupled to translocation across the plasma membrane [14] (Figure 2B).

This coupled synthesis and translocation process is fairly well understood for cellulose synthase, a protein that belongs to the same family of GTs [4]. Cellulose synthase produces a linear glucose polymer in which each glucosyl unit is connected to its neighbors by a beta-1,4 linkage. The enzyme elongates the polymer from UDP-glucose one sugar unit at a time and undergoes conformational changes after each elongation reaction that translocate the extended polymer into its transmembrane (TM) channel [15]. While cellulose is a stiff, ribbon-shaped polymer well suited for such a pushing mechanism, the physio-chemical properties of the HA heteropolysaccharide are fundamentally different. Thus, it remains to be determined whether HAS uses a similar polymer elongation and translocation mechanism.

It is worth noting that two other kingdoms of life have been found to synthesize HA, namely viruses and bacteria [14]. Viral HASs have so far only been identified in algal viruses where they induce the formation of a thick HA coat shortly after infection. The biological function of this coat is currently unknown, yet the enzymes closely resemble the vertebrate orthologs and likely share a similar reaction mechanism [16, 17] (Figure 2B). Structural and functional analyses of *Chlorella* virus HAS (CvHAS) provided important insights into the enzyme's architecture and how it initiates HA biosynthesis [18]. However, because the enzyme produces HA polymers substantially shorter than vertebrate HAS, as well as lacking a nascent HA chain in structural analyses, little was learned about how HAS coordinates the nascent HA polymer, translocates it between elongation steps, and regulates its size.

Bacterial HASs are much smaller, predicted to contain only 4 TMHs and are believed to function as obligate dimers, which co-assemble a single HA-secretion channel. Interestingly, these HASs utilize an HA elongation mechanism that is different from vertebrate and viral homologs, which elongate at the non-reducing end of the HA polymer. Bacterial HASs elongate the

reducing end of HA and each subunit is fulfilling distinct roles. One subunit is believed to retain UDP-terminated polymer, whereas the other binds another substrate molecule [6, 14]. In the bacterial biosynthesis mechanism, HA is covalently attached to a UDP molecule at its reducing end and the incoming UDP-sugar serves as the acceptor of the GT reaction. In the case of viral and vertebrate HA biosynthesis, HA is not attached to UDP and incoming UDP-sugar serves as the donor (Figure 2B). The main purpose of bacterial HA production is to form a non-immunogenic protective capsule against various stressors, as well as antimicrobial agents [19].

Viral and bacterial HAS are interesting model systems to study HA biosynthesis and secretion. However, vertebrate and especially the human enzymes are of particular interest for human physiology and health. Hence, detailed knowledge of vertebrate HAS structure and function, in particular in the context of post-translational modifications, can help us better understand the molecular mechanisms by which HA exerts its numerous biological functions, and how those processes drive diseases.

Three HAS isoforms have been identified in vertebrates, including humans (HAS1-3). Out of the three, isoform 2 is absolutely essential for development in mammals, as deletion of the *has2* gene in mice is lethal, embryos succumb right before the onset of skeletal muscle development [10]. While HAS1 and HAS2 produce HA polymers of several megadaltons, HAS3-derived products usually range from 10^5 to 10^6 Da in size [8, 12, 20, 21].

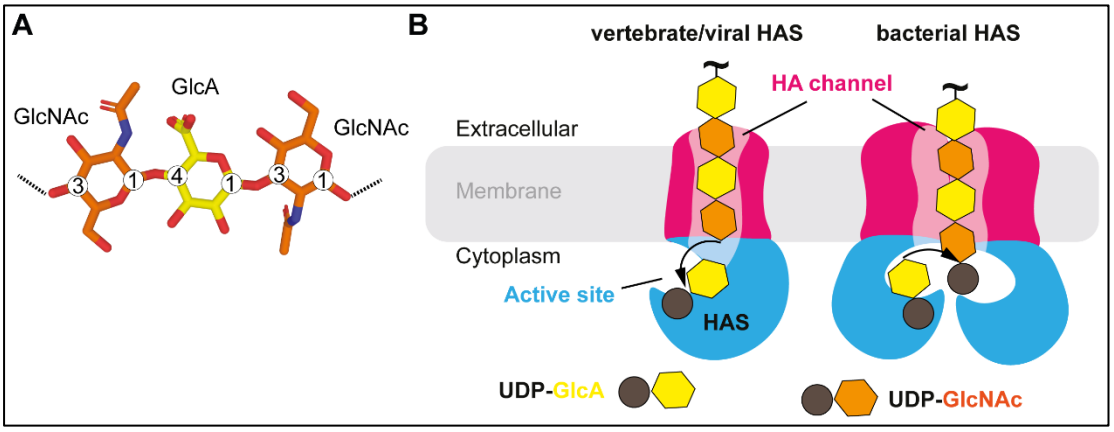


Figure 2 Dual activity of hyaluronan synthase
A – Structure of hyaluronan trisaccharide showing the alternating pattern of sugar units and glycosyl linkages. B – Schematic representation of hyaluronan synthase (HAS), a membrane-embedded GT that catalyzes the formation and transport of hyaluronan. Both monomeric vertebrate/viral-type HAS and dimeric bacterial-type HAS are depicted.

1.4 Goals of the dissertation

My goals for this project were to (1) express a suitable vertebrate HAS homolog in insect cells at high levels, (2) solubilize the protein from the membrane and purify it to homogeneity, while retaining its enzymatic activity, (3) biochemically

characterize the vertebrate HAS (substrate turnover, selectivity, HA product size distribution), (4) determine the first high-resolution structure of a vertebrate HAS homolog.

This project provided a lot of insights into the molecular mechanism of HA biosynthesis by vertebrate HAS and other HASs in general (Figure 3). The most interesting questions I wanted to address include but are not limited to: *How HAS selects its substrates? How is the HA size determined? Does HAS oligomerize? How is the membrane-spanning channel assembled? How is HA secreted through the channel? What is the gating mechanism of the HA secretion channel?*

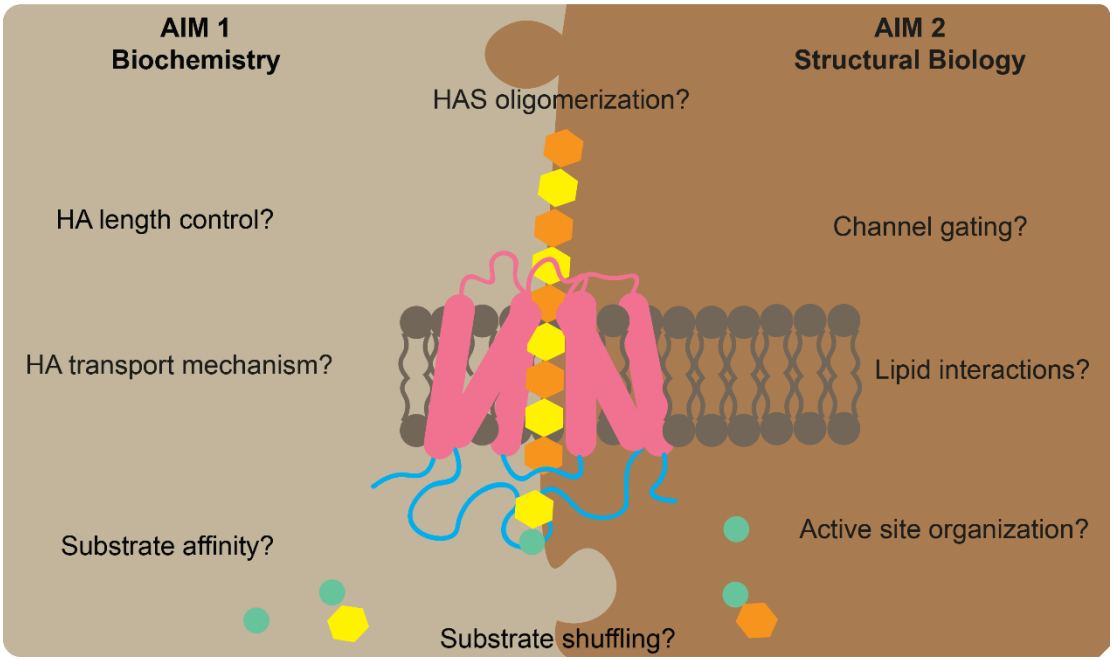


Figure 3 Goals of the dissertation
Schematic representation of the goals of the dissertation that can be divided into two types of aims – structural and biochemical.

Structural and functional analyses of integral membrane proteins include at least 5 steps. First, suitable expression conditions have to be identified enabling the production of sufficient amounts of the target protein. Second, solubilization conditions have to be established that allow extracting the protein of interest from the membrane in a functional form. Third, the protein has to be purified without losing catalytic activity. Fourth, the protein has to be prepared for structural and functional analysis by optimizing solubilizing detergents or reconstitution into a suitable membrane mimetic system. Fifth, electron microscopy or crystallographic data have to be collected and processed for model building.

2. Results

2.1 HAS homolog search

A successful structural biology project needs to start with obtaining high quantities of pure, and preferably, catalytically active material. A lot of initial work needs to be dedicated to find a target protein that can be produced in the desired expression host. Most eukaryotic membrane proteins are expressed in yeast (*Pichia pastoris*, *Saccharomyces cerevisiae*), insect (*Spodoptera frugiperda* - Sf9) or mammalian (human embryonic kidney cells - HEK293) cells [22-25]. Further, comprehensive assessments need to be performed in order to ensure stability of the target protein to facilitate successful in vitro manipulations, such as purification, enzymatic assays, crystallization, or cryogenic electron microscopy (cryoEM) analyses.

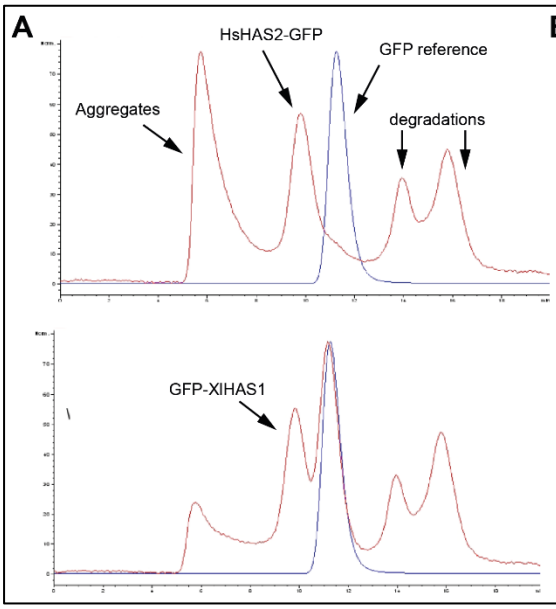
Quite often, the homolog from an organism of choice is not stable enough or sub-optimally expressed. This can be due to the protein translation machineries from certain expression hosts being incompatible with some target proteins, the need for target-specific post-translational modifications, toxic effects from protein overexpression, proteolytic degradation, differences in codon usage, among others [23, 25]. It is a common practice to resort to studying a more stable ortholog from a less evolutionarily advanced organism (e.g. fish instead of human) or a thermophilic bacterium instead of pathogenic strain of interest. Hence, the choice of homolog is often a tradeoff between getting as close as possible to solving a specific biological question and allowing robust, in-depth biochemical analyses.

2.1.1 Hight throughput membrane protein expression screening

In order to identify HAS homologs for biophysical studies, the Zimmer lab collaborated with The Center on Membrane Protein Production and Analysis (COMPPAA) to clone and express a large number of HASs for functional and structural studies. Brian Kloss at COMPPAA screened 10 vertebrate and 7 viral HASs fused to green fluorescent protein (GFP) for expression in Sf9 insect cells. Expression and detergent solubility were screened using fluorescence-detected size exclusion chromatography (FSEC). FSEC allows for a rapid assessment of membrane protein stability by running the detergent-solubilized membranes obtained from the expression host through a SEC column connected to a chromatography system with fluorescence detector. Monitoring green fluorescence coming from the GFP fused to the target protein makes this method highly sensitive and specific. Obtained FSEC profiles are highly informative as they report on both the relative expression levels and aggregation state of the target protein-GFP fusion (Figure 4A). A well-optimized FSEC setup, together with high throughput cloning and cell culturing techniques

allows for screening tens to hundreds of targets in a few weeks, depending on the desired expression host [24-26].

All vertebrate HAS targets were expressed in Sf9 insect cells, which constitute a cost-effective and relatively fast eukaryotic protein production platform. Brian obtained highly promising results for zebrafish (Dr), human (Hs), African frog (Xl and Xt), and mouse (Mm) HAS isoforms (Figure 4B). I carried on from there by assessing some of those homologs further for expression, stability and catalytic activity.



B

Organism	Protein	Tag	FSEC result
Zebrafish	DrHAS3	N-term GFP	OK
	DrHAS2	N-term GFP	OK
Human	HsHAS3	N-term GFP	OK
	HsHAS2	N-term GFP	OK
Mouse	MmHAS3	N-term GFP	OK
	MmHAS2	N-term GFP	OK
	XtHAS2	N-term GFP	OK
African frog	XlHAS1	N-term GFP	OK
	XlHAS-like	N-term GFP	OK
Zebrafish	DrHAS2	C-term GFP	OK
	HsHAS3	C-term GFP	OK
Human	HsHAS2	C-term GFP	OK
	HsHAS1	C-term GFP	OK
	XtHAS2	C-term GFP	BAD
African frog	XlHAS1	C-term GFP	BAD
	XlHAS-like	C-term GFP	BAD

Figure 4 Initial results of HAS homolog search

A – Representative fluorescence-detected size exclusion chromatography (FSEC) profiles for vertebrate HAS constructs screened by COMPPAA. Each HAS was fused to green fluorescent protein (GFP) and ran over FSEC upon solubilization in dodecyl maltoside. The HAS-GFP profile (red) is overlaid with the profile obtained for GFP alone (blue). Peaks corresponding to the target protein, aggregates and degraded material are indicated. B – Summary of FSEC results obtained for all vertebrate HAS homologs screened by COMPPAA.

2.1.2 HAS expression trials

Motivated by the promising FSEC results from COMPPAA, I carefully re-assessed expression of some of the HAS homologs in our lab. Due to high sequence identity between HASs from higher vertebrates (more than 90% in most mammals), I decided to focus on the most relevant target – HsHAS2. Additionally, I selected *Xenopus laevis* HAS1 as a fallback platform, since it's less similar and showed catalytic activity in raw membrane preparations in the past [14]. Subsequently, motivated by good results for XlHAS1 (see below), I decided to move on to the other isoforms from *Xenopus* – XlHAS2 and XlHAS3, as well as HAS1 from humans. HsHAS3 expression and stability was not investigated at that time.

The HAS genes were provided to us by COMPPAA or synthesized by a company. They were amplified by polymerase chain reactions (PCR) and cloned, as an N-terminal or C-terminal GFP fusion into the pACEBac-1 plasmid for bacmid generation. The engineered protease recognition site between the GFP tag and the protein of interest allows me to remove the GFP after protein purification. My GFP fusions are cleavable by the tobacco etch virus (TEV) protease, which has good detergent tolerance and can be produced in-house in large quantities. Removal of GFP might be important for optimal enzymatic activity. Additionally, I generated constructs carrying only N- and C-terminal His tags, instead of GFP. The final tag of choice and its position were determined for each HAS homolog based on FSEC results, in-house expression tests, as well as subsequent purification trials. For HAS2 and HAS3 isoforms, the tags were placed at their C-terminus. For HAS1s, the tag was positioned at the N-terminus (Figure 5A).

To generate a bacmid for Sf9 cell transfection, I transformed the pACEBac1-HAS shuttle vectors into *E. coli* DH10Bac cells, which generated the bacmid by transposition [27]. This bacmid was purified by alcohol precipitation and used to transfect Sf9 cells to generate P0, P1 and P2 virus samples according to the Joint Center for Innovative Membrane Protein Technologies (JCIMPT) protocol. Results of the expression were assessed by analyzing whole-cell fluorescence and anti-His Western Blots of the lysed cells. Overall, all homologs expressed well, with XlHAS1-3 being expressed at higher levels than human isoforms (Figure 5B). Confocal images of the Sf9 cells expressing HsHAS2 and XlHAS1 suggest proper integration with the host's membranes (Figure 5C).

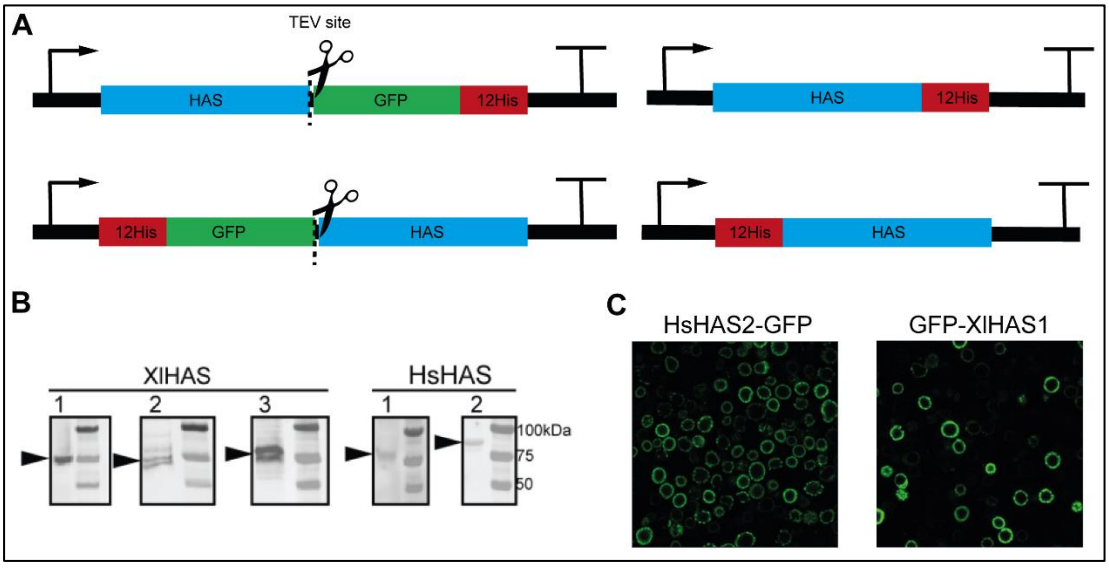


Figure 5 Initial HAS expression tests

A – Schematic representation of the vertebrate HAS constructs prepared for this project. HASs were expressed either as an N- or C-terminal His or GFP-His fusion. B – Western Blot analyses of the lysed Sf9 cells after expression of the indicated HAS homologs. C – Confocal microscopy

images of Sf9 cells expressing GFP-fused HsHAS2 and XlHAS1 indicating membrane localization of HAS.

2.1.3 Assessing the activity of the expressed HASs

HA biosynthesis can be detected by incorporating radiolabeled substrates, ^{14}C-labeled UDP-GlcA or ^{3}H-labeled UDP-GlcNAc, into the growing HA polymer. The acidic HA can then be run into an SDS-PAGE gel and visualized by ^{14}C autoradiography or purified using paper chromatography and quantified with ^{3}H liquid scintillation counting [14, 28]. These detection methods, combined with enzymatic degradation of the polymer, are highly specific and allow for high sensitivity functional analyses, even in raw membrane preparations.

XlHAS1 robustly synthesizes a polymer in vitro that can be detected by both approaches that is degradable with hyaluronidase, which indicates that it is indeed authentic HA (Figure 6A). The expressed HsHAS1, HsHAS2, XlHAS2 and XlHAS3 construct shows significantly less catalytic activity compared to XlHAS1 (Figure 6A). This could be due to issues with folding and trafficking in the expression host, intrinsic lower activity, or suboptimal reaction conditions used.

2.1.4 Detergent solubility tests

Previous work on HAS and related enzymes in our lab showed that in vitro activity can be sensitive to delipidation during solubilization and purification. Thus, a crucial part of a membrane protein characterization is finding suitable solubilizing detergents and/or establishing reconstitution protocols. Things to consider when choosing the detergent include (1) solubilization efficiency, (2) protein stability and activity, and (3) price. The number of chemically diverse detergents commercially available is large, hence finding the right one is usually just a matter of dedicating time and labor.

One of the most popular detergents currently used for purification of eukaryotic membrane proteins is glycol-diosgenin (GDN), a newly-developed synthetic substitute of digitonin [29]. Both of these detergents are considered to be among the mildest detergents available and have a good track record in cryoEM experiments, but are also among the most expensive [29, 30]. This forces the use of cheaper detergents, such as lauryl-maltose-neopentyl-glycol (LMNG) or dodecyl-maltoside (DDM) for solubilization as this step requires a lot of detergent. LMNG and DDM detergents are typically supplemented with the cholesterol mimetic, cholesteryl hemisuccinate (CHS), which greatly improves the stability of many membrane proteins. Notably, LMNG/CHS is not compatible with cryoEM experiments due to high background, whereas DDM/CHS is often too harsh. Hence, an effective strategy to satisfy all three

abovementioned requirements is to solubilize the protein in LMNG/CHS and exchange it into GDN during affinity chromatography.

Nonetheless, I decided to empirically test this approach and assess the overall stability of the two representative HAS homologs. Inverted membrane vesicles (IMVs) derived from Sf9 cells expressing either HsHAS2 or XlHAS1 were subjected to solubilization in a handful of popular detergent choices. One hour solubilization was followed by ultracentrifugation to remove insoluble material, including detergent-induced protein aggregates, and subjected to Western Blot analysis. XlHAS1 shows excellent solubility in all of the tested detergents, including the mild GDN and digitonin, as well as harsher DDM/CHS (Figure 6B). This is indicative of good protein stability, at least when exposed to 1h solubilization. By comparison, HsHAS2 did not solubilize as easily or aggregated in all of the detergents, except for sodium dodecyl sulfate (SDS), which is a strong, denaturing detergent not suitable for native protein purification (Figure 6C).

Additionally, detergent solubilized IMV fractions were used for activity assays in the presence of radioactive substrates. XlHAS1, in addition to being stable and easy to solubilize, has proven to remain active upon solubilization in a handful of tested detergents, including Digitonin, GDN, LMNG/CHS and, to lesser extent, in DDM/CHS (Figure 6D). This is extremely important, as solubilization is an unfortunate but crucial step in protein purification that often leads to stripping of the annular layer of lipids supporting the proper fold and the catalytic activity of a membrane protein [29].

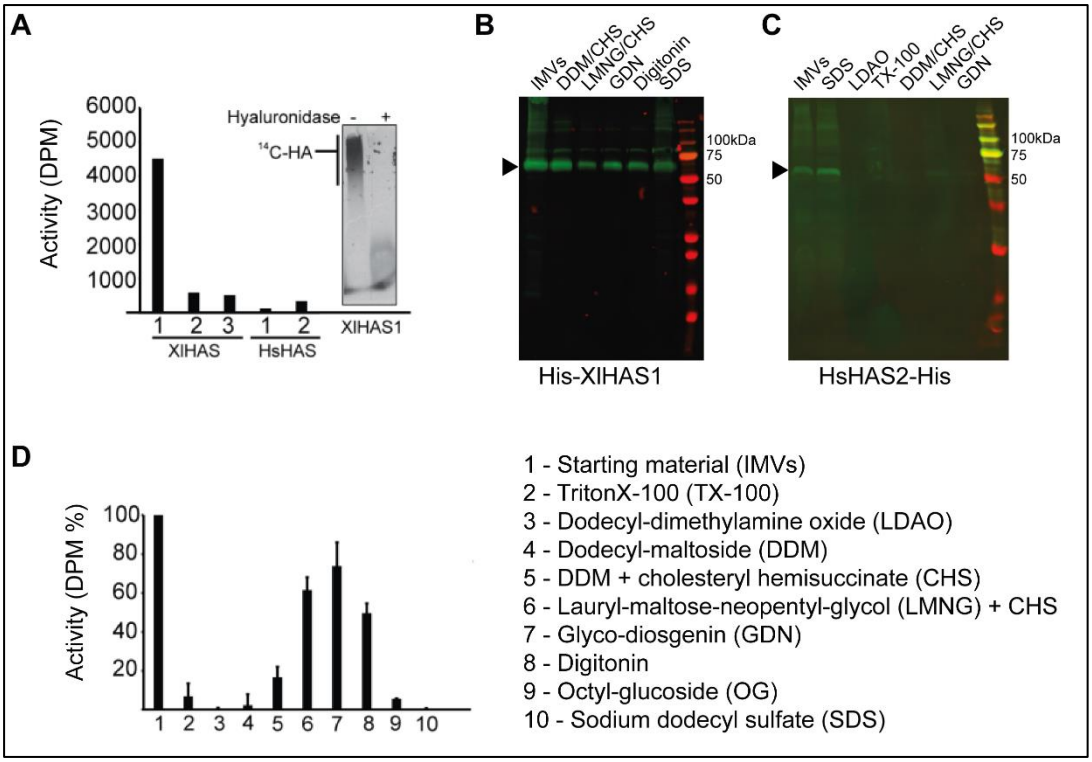


Figure 6 HAS stability tests

A – Results of the activity assay performed using Sf9 cell-derived inverted membrane vesicles (IMVs) expressing the indicated HAS constructs. The bulk HA production was quantified by measuring the number of disintegrations per minute (DPM) after ^{3}H-GlcNAc incorporation. The inset shows an autoradiograph of the synthesized HA after ^{14}C-GlcA incorporation, with and without hyaluronidase treatment. B – Results of the solubility screening for XlHAS1. IMVs from Sf9 cells were solubilized in the indicated detergents, subjected to ultracentrifugation and Western Blot analyses to assess the solubility. C – As panel B but for HsHAS2. D – Results of the activity assay conducted on Sf9 cell-derived IMVs expressing XlHAS1, solubilized using the indicated detergents. The activity was quantified as for panel A and expressed as percentage relative to starting material, the IMVs.

2.1.5 Protein purification trials

After identifying suitable solubilization conditions, the next step was to attempt purification of the proteins by affinity chromatography. The engineered His-tag allows for cost-effective purification by immobilized metal ion affinity chromatography (IMAC). IMAC sometimes suffers from substantial contaminations, in particular if the protein of interest is not stable or expressed at high levels.

All of the tested homologs were purified upon solubilization in LMNG/CHS, using standard IMAC approach followed by SEC. All IMAC and SEC buffers contained the milder GDN detergent, which substituted the LMNG/CHS during the later purification steps. XlHAS1 was again the only successful homolog that could be obtained in milligram quantities in a pure and catalytically active form (Figure 7A). All of the other homologs solubilized poorly and bound to IMAC column very inefficiently, resulting in a preparation that was heavily contaminated. A minor fraction of the target protein could typically be

detected in SDS-PAGE gels following Coomassie Brilliant Blue staining and Western Blotting, yet no in vitro HA synthase activity could ever be determined (Figure 7B-E).

On the basis of all of these experiments, XlHAS1 was chosen for subsequent characterization using cryoEM and biochemical methods.

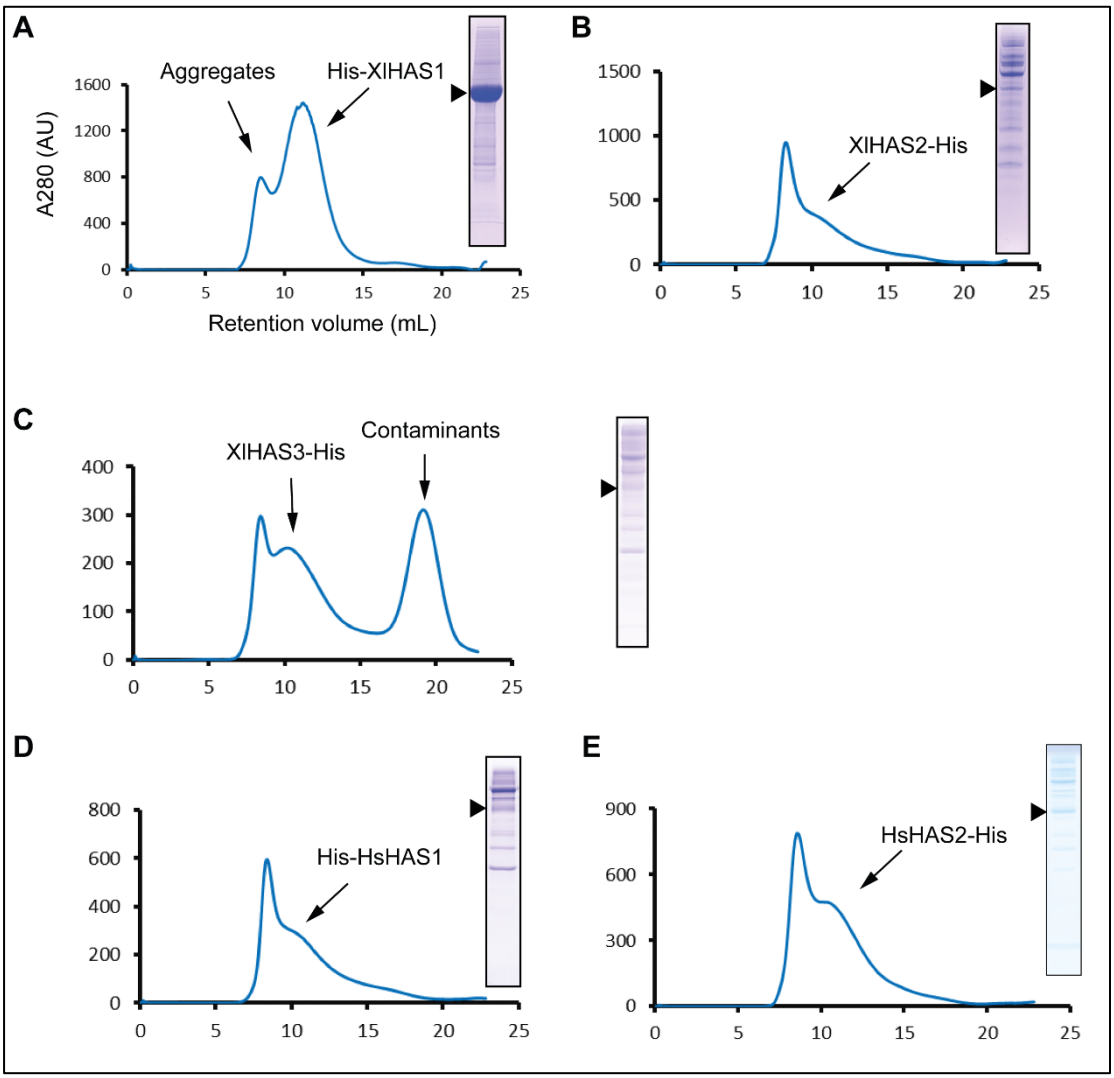


Figure 7 HAS purification trials

SEC results of the purification trials of the HAS homologs: XlHAS1 (panel A), XlHAS2 (panel B), XlHAS3 (panel C), HsHAS1 (panel D), HsHAS2 (panel E). Coomassie-stained gel insets for each peak fraction are shown and the target protein bands are indicated.

2.2 XlHAS1 production

Reasonably high overall expression levels and, more importantly, good stability of XlHAS1 allow for its relatively easy and cost-efficient purification using IMAC. Some optimization steps to the expression and purification procedures were performed to maximize protein yields and allow robust cryoEM and biochemical characterization.

2.2.1 XlHAS1 expression

XlHAS1 expression was revised substantially compared to the standard protocol developed by JCIMPT. The Sf9 cells are infected at a higher density than recommended - 2-3 million cells per mL instead of 1 million. Additionally, low quantities of the P2 virus stock are used for infection, 1-2% of the culture volume, depending on performance of the particular viral stock. The cultures are harvested as soon as the cell morphology indicates viral infection (enlarged cells, swollen nuclei), before dead cells start appearing, typically after 48 hours. At this point, the target gene expression is already quite high. More importantly, harvesting the culture sooner (as opposed to waiting for the protocol-recommended 70% viability) allows to obtain more intact cells, and consequently more protein per culture volume. Additionally, this approach simplifies the cell manipulation procedures, reduces the volume of media and P2 virus used, and saves time and money.

2.2.2 XlHAS1 solubilization

As mentioned, XlHAS1 tolerates solubilization in a handful of non-ionic detergents that can preserve a substantial fraction of its catalytic activity (Figure 6D). Throughout multiple purification trials, I have found that both LMNG/CHS and DDM/CHS effectively solubilize XlHAS1. Both detergents, however, are not tolerated if used throughout the entire purification and need to be substituted with GDN (Figure 8A).

2.2.3 XlHAS1 purification

XlHAS1 protocol was gradually optimized to maximize the protein yield. In addition to the abovementioned expression protocol revisions, some purification procedures were streamlined as well. Those included (1) omitting cell lysis using a microfludizer. This method of cell lysis is time consuming, needlessly harsh for eukaryotic cells, and often leads to unwanted heating of the sample. Instead, cycling the cells through a Dounce-style tissue grinder in the presence of the solubilizing detergent is sufficient. (2) Adjusting the amount of nickel-nitrilotriacetic acid (NiNTA) beads used for IMAC. The theoretical capacity of the commercial NiNTA resins is 5-50 mg of protein per 1 mL of the resin. In reality, membrane proteins don’t bind very well, even when long His-tags are used (e.g. 12xHis), likely due to steric hindrance of the detergent micelle and reduced expression compared to soluble proteins. The volume of affinity beads used constitutes a tradeoff between the capacity (the amount of protein) and the specificity (purity) of the binding. The optimal amount of beads was gradually determined for XlHAS1 to be at 12 mL of in-house recharged, 50% NiNTA suspension per 1 L of cell pellet. (3) Optimizing the imidazole concentration during IMAC. This step also determines the amount of protein that can be obtained and its purity. The amount of imidazole in the binding and

wash buffers needs to be adjusted for the level of expression of the target protein and the length of the His-tag. Efficient and specific XlHAS1 binding is achieved at 10 mM imidazole, which is gradually increased to 20 mM during IMAC washes. (4) Inclusion of 10% glycerol in the purification buffers to facilitate protein stability through kosmotropic effects. Omitting glycerol leads to substantial aggregation of XlHAS1 (Figure 8B, C). Notably, glycerol must be omitted from the final SEC running buffer for proper vitrification of cryoEM grids, which is well-tolerated by XlHAS1. (5) Adjusting the amount of GDN used to allow preparation of high quality cryoEM grids, while not compromising protein stability. XlHAS1 tolerates as little as 0.01% GDN. Lowering the GDN concentration further leads to protein aggregation, whereas higher concentrations contribute to background in cryoEM experiments.

Overall, these small tweaks of the production conditions allowed for a roughly two-fold increase in protein yield, as the starting yield of ~1 mg per 1 L of culture turned into 2-4 mg after optimization.

2.2.4 XlHAS1 reconstitution trials

Biochemistry is best done in proteoliposomes (PL) that create an enclosed compartment suitable to study the membrane translocation of HA using radioactive labels and centrifugation [28, 31]. Additionally, the presence of a lipid bilayer, preferably with close-to-native lipid composition, adds validity to biological conclusions drawn from biochemical experiments.

Structural studies may benefit from reconstitutions into nanodiscs (NDs). NDs are planar bilayers stabilized by membrane scaffold proteins (MSPs) that do not form solvent-excluding compartments. Their formation requires extensive optimizations of the reconstitution protocol [32, 33]. Factors to consider here include choosing an MSP variant that forms a nanodisc of suitable diameter, as well as the lipid:MSP:protein ratio, which is crucial for good reconstitution efficiency [32]. Nanodiscs are a great reconstitution system that allows membrane protein structure determination in a native or native-like environment.

The detergent in the sample needs to be gradually removed and substituted by lipids that form a vesicle (PL) or a planar membrane stabilized by the MSP belt (ND) during both reconstitution approaches. The most common detergent removal method is adding sorbent media, BioBeads, to the reconstitution mixture [34]. BioBeads is a polystyrene-based nanopore medium which efficiently binds organic compounds below 2000 Da of molecular weight. They were initially developed to extract organic chemicals from water, drugs from biological materials, polyphenols from plants, for example. Thus, BioBeads constitute a convenient method for detergent removal [35]. The need for detergent removal during reconstitution is problematic due to two reasons.

First, the rapid depletion of the detergent often leads to target protein destabilization and poor reconstitution efficiency. Second, the mildest detergents used for purifying fragile eukaryotic membrane proteins (GDN, LMNG, etc.) have extremely low critical micelle concentration (CMC) and are very difficult to remove due to low number of individual detergent monomers in solution being accessible to BioBeads or suitable for dialysis. This often prevents effective reconstitution.

XlHAS1 reconstitution into NDs was tested upon purification in several detergent combinations. Purification in LMNG alone or LMNG/CHS prevented efficient reconstitution, as it is well established in our laboratory that removal of LMNG is extremely difficult (Figure 8D). Reconstitution upon purification in DDM/CHS also failed, likely due to extended incubation time in DDM/CHS leading to aggregation of XlHAS1, as shown in Figure 8A. Reconstitution appeared successful when DDM/CHS was used for solubilization and exchanged to GDN for purification, followed by reconstitution (Figure 8D). The MSP variant, lipids of choice and MSP:lipid ratio were assessed under these conditions. The final protocol involved using MSP1D1, *E. coli* total lipids dissolved in sodium cholate detergent, and using a molar ratio of 1:4:80 of XlHAS1, MSP, lipids, respectively. The obtained sample was very homogenous (Figure 8E, F) and was used for Fab discovery, conducted by the Kossiakoff lab (see below and Methods), as well as cryoEM experiments. Whereas Fab selection was reasonably successful, cryoEM data suggests the existence of multiple species in this sample, including empty micelles/NDs of various sizes, and poor quality reconstructions of XlHAS1 embedded in an unknown vehicle (Figure 8G).

It is my interpretation that GDN cannot be removed efficiently from the reconstitution mixture due to its extremely low CMC (0.0021% according to the manufacturer, Anatrace), hence the obtained reconstitution mixture consists of empty nanodiscs, GDN-solubilized HAS and empty GDN micelles (class 1, 2, 3 on Figure 8G, respectively). However, since GDN facilitates extremely good stability and catalytic activity of XlHAS1, all of the subsequent biochemical and cryoEM characterizations were performed with GDN-solubilized XlHAS1.

Synthetic copolymers are of rising interest as an alternative to MSP-based NDs. Copolymers of Styrene-Maleic Acid (SMA), Diisobutylene Maleic Acid (DIBMA), and various modifications thereof allow for reconstitution following detergent solubilization, or more excitingly, direct extraction of the target protein from membranes [36-38]. This, in principle, allows imaging of membrane proteins in an almost native environment. So far, the use of copolymers has been limited to stable bacterial membrane proteins, whose structures have already been solved in detergents or classical NDs. Despite advances in the field, these compounds still show a lot of limitations, such as

poor solubilization/extraction efficiency, pH restrictions, intolerance of divalent cations, lack of control over the ND size and high price [37].

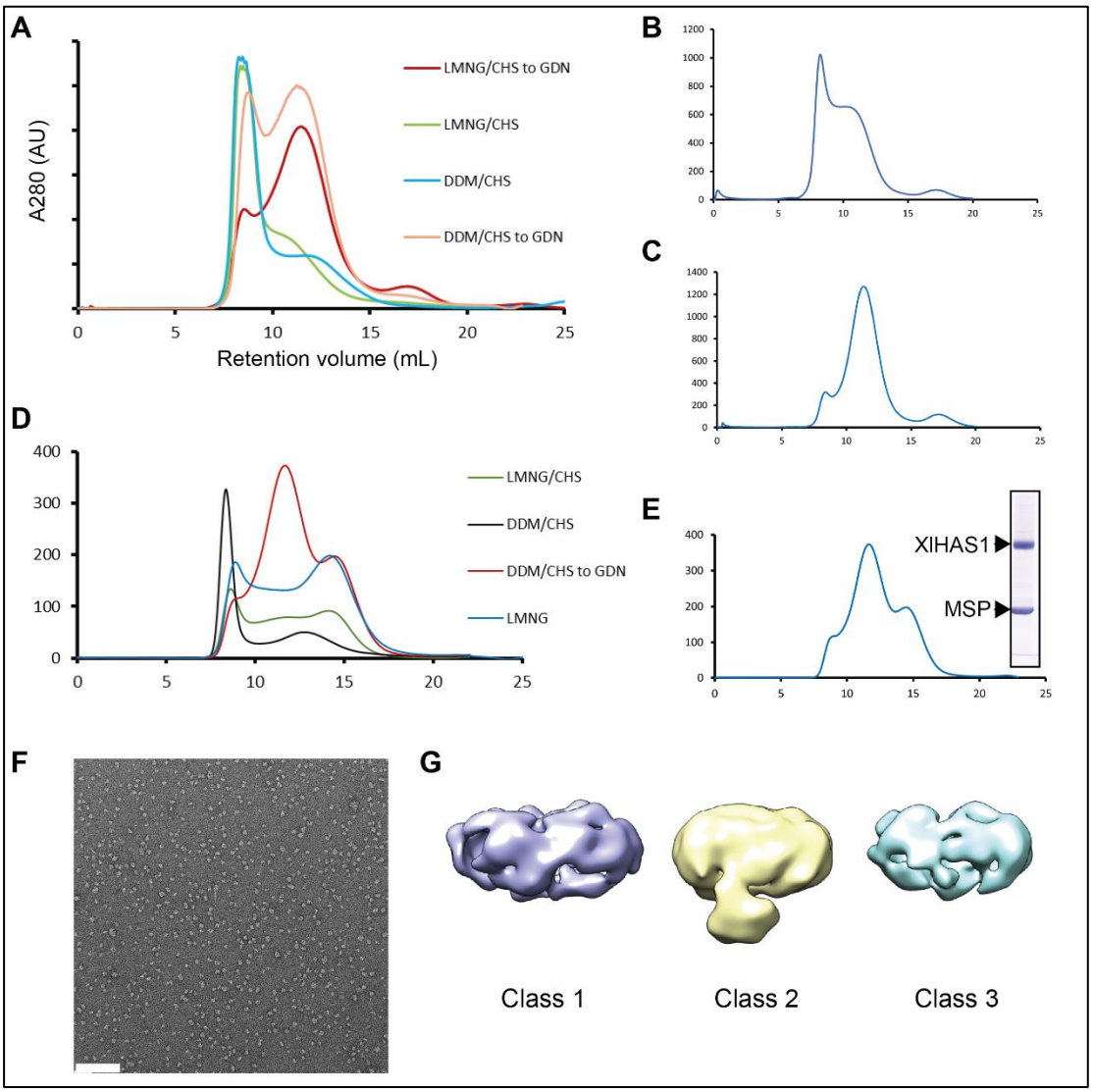


Figure 8 XlHAS1 purification optimization

A – SEC results of XlHAS1 purification schemes using various detergent combinations - LMNG/CHS or DDM/CHS used for solubilization, followed by exchange into GDN or keeping the solubilizing detergent throughout. B – XlHAS1 SEC profile upon purification in buffers without glycerol. C – XlHAS1 SEC profile upon purification in buffers with glycerol. D – SEC results for XlHAS1 nanodisc reconstitution trials after purification in the indicated detergent mix. E – SEC profile and a Coomassie-stained gel for the optimized XlHAS1 nanodisc preparation. F – Negative Stain electron micrograph of the optimized XlHAS1 nanodisc preparation. Scale bar on the micrograph corresponds to 100 nm. G – 3D reconstructions obtained from cryoEM data collected on the optimized XlHAS1 nanodisc.

2.3 Biochemical characterization of XlHAS1

HAS transfers the UDP-activated glycosyl unit (the donor sugar) to the non-reducing end of the nascent HA polymer (the acceptor sugar). This reaction generates UDP and an elongated HA as reaction products [39]. Catalytic activity can be assessed by monitoring the release of UDP in an enzyme-coupled reaction [28] and by analyzing HA, either radiometrically (as described

above and in ref. [14]) or by dye staining following electrophoresis (as described below and in ref. [40]).

A plethora of factors control the length of HA in the physiological setting, including substrate availability, activity of hyaluronidases, ECM composition, HA modifications, as well as expression of different, tissue-specific HAS isoforms [8, 41]. However, the specific molecular mechanism by which HAS itself synthesizes HA of specific length or modulates it, remains unknown. To gain insights into HA size control, I conducted a series of in vitro assays to test the size distribution of HA synthesized by XlHAS1 under various experimental conditions.

In order to characterize factors controlling HA length, I first need to determine the size of in vitro synthesized HA. I adapted a protocol used for analyzing extracellular GAG composition and size distribution in native tissues and tissue cultures. The protocol involves using polyacrylamide gel electrophoresis (for LMW GAGs), as well as agarose gel electrophoresis (for HMW GAGs) followed by staining with a Stains-All dye (3,3'-dimethyl-9-methyl-4,5,4'5'-dibenzothiacarbocyanine) [40, 42]. Stains-All reacts with many biomolecules and changes its color in various ways. It shows high sensitivity for DNA (stains blue), RNA (purple), proteoglycans (purple) and acidic polysaccharides (hyaluronan, alginate, pectinate, heparin, heparan sulfate, chondroitin sulfate, dermatan sulfate, all stained blue) [43]. The dye also stains anionic proteins (blue), cationic proteins (pink), and lipids (yellow), although with much lower sensitivity [44]. This method in combination with commercially available HA standards (“ladders”) [45], enables me to precisely determine the size of HA synthesized by my enzyme in vitro. Degradation of the polymer with a HA-specific hydrolase is a sensitive test for the formation of authentic HA and constitutes an important control when using a non-specific dye (the Stains-All). Those experiments are complemented well by the data obtained for paper chromatography analyses of ^{3}H-labelled HA, which allow for quick quantification of the overall HA synthase activity. In some cases, higher sensitivity of HA detection may be required. For this, autoradiography followed by introduction of ^{14}C-labelled GlcA into the polymer can be used (as shown above).

2.3.1 Determine the size of HA synthesized by vertebrate HAS

XlHAS1 exhibits robust catalytic activity in vitro. My initial trials of characterizing HA size in vitro showed that HA migrates as a smear, rather than a discrete band (Figure 9A, B, D-E). This suggests a highly heterogenous HA size distribution, which correlates well with data reporting that HA indeed is highly polydisperse in native tissues [40, 42]. The obtained polysaccharide is readily degraded by hyaluronidase, demonstrating that it is authentic HA (Figure 9A, B). Agarose gel electrophoresis of time points of an in vitro HA synthesis

reaction revealed increasing polymer lengths over a ~120 min synthesis reaction before stalling. XlHAS1's product size distribution broadens significantly during prolonged in vitro synthesis reactions, perhaps due to substrate depletion and/or product inhibition. Notably, at completion, XlHAS1 synthesizes HA polymers of a consistent length that migrate well above a 1.6 MDa HA marker (Figure 9A). This material is comparable to HA in vitro synthesized by bacterial HAS from *Streptococcus equisimilis* (Se). In contrast, CvHAS, produces polydisperse and low molecular weight HA within ~30-200 kDa range (Figure 9B).

2.3.2 Determine substrate affinity of vertebrate HAS

The cellular concentrations of HAS' substrates differ in native cells – UDP-GlcA typically occurs at 10-30% of the concentration of UDP-GlcNAc (around 1 mM), the most abundant UDP-activated sugar [46]. Hence, XlHAS1's affinity is likely different for the two substrates.

Kinetic analyses were performed using a UDP release assay, which allows for real-time quantification of the UDP released during glycosyl transfer reaction. A standard HA synthesis reaction is supplemented with lactate dehydrogenase (LDH), pyruvate kinase (PK), phosphoenolpyruvate (PEP) and reduced nicotinamide adenine dinucleotide (NADH). UDP released during HA synthesis is converted to uridine triphosphate (UTP) with PEP serving as phosphate donor. The resulting pyruvate is reduced to lactate in an NADH dependent reaction, which is converted to NAD+. The oxidation of NADH can be monitored based on its absorbance at 340 nm [28]. The amount of NADH consumed correlates directly with the amount of hydrolyzed UDP-substrate and can be quantified using NADH's extinction coefficient.

Michaelis-Menten constants (K_M) for each substrate were determined using the UDP release assay by titrating one substrate at a saturating concentration of the other. The titration curves were processed in GraphPad using nonlinear regression function and revealed K_M for UDP-GlcNAc and UDP-GlcA of about 470 and 370 µM, respectively (Figure 9C). The obtained K_M values indicate that HAS shows similar affinity for both substrates, which likely contributes to its fidelity as an HA-synthase, as opposed to GlcNAc polymerase. Slightly lower K_M (i.e. higher affinity) for UDP-GlcA is dictated by the lower abundance of this substrate in cells [46].

Monitoring the length distribution of the synthesized HA under similar substrate conditions (one limiting, the other in excess) shows increasing HA lengths up to a substrate concentration of 2.5 mM (Figure 9D). In agreement with the kinetic curves obtained from the UDP release assay (Figure 9C), both substrates must be present in abundance for proper HA formation. Based on the observed HA product lengths (~1.6 MDa produced within 90 min), the

electrophoretic analysis suggests a catalytic rate of about 60 substrate turnovers per minute.

2.3.3 UDP inhibition limits the size of HA produced in vitro

In vivo, HA biosynthesis could stall due to substrate depletion, competitive UDP inhibition, and/or HA accumulation. To analyze the impact of sustained high substrate concentrations during HA biosynthesis, an in vitro synthesis reaction was supplemented with both substrates two, four, and six hours after initiation. Similarly to the UDP release assay, this reaction mixture was supplemented with the LDH/PK enzyme mix and PEP to convert the released UDP to non-inhibitory UTP. These conditions further extend the synthesized HA polymers to products substantially exceeding the 1.6 MDa marker (Figure 9E, F). This indicates that, despite common belief [46], HAS1 can synthesize HMW HA under optimal reaction conditions.

Notably, HA extension is only observed when both substrates are supplied in excess (2.5 mM). If one substrate (UDP-GlcNAc) remains at a limiting concentration (0.1–0.5 mM), no increase in HA size is observed (Figure 9E). This is likely caused by HA release between substrate replenishing steps, such that most HASs reinitiate HA biosynthesis. This hints at a possible HA release mechanism - when not actively synthesizing HA, HAS will release HA, possibly due to spontaneous diffusion of the flexible HA polymer out of the HAS' secretion channel.

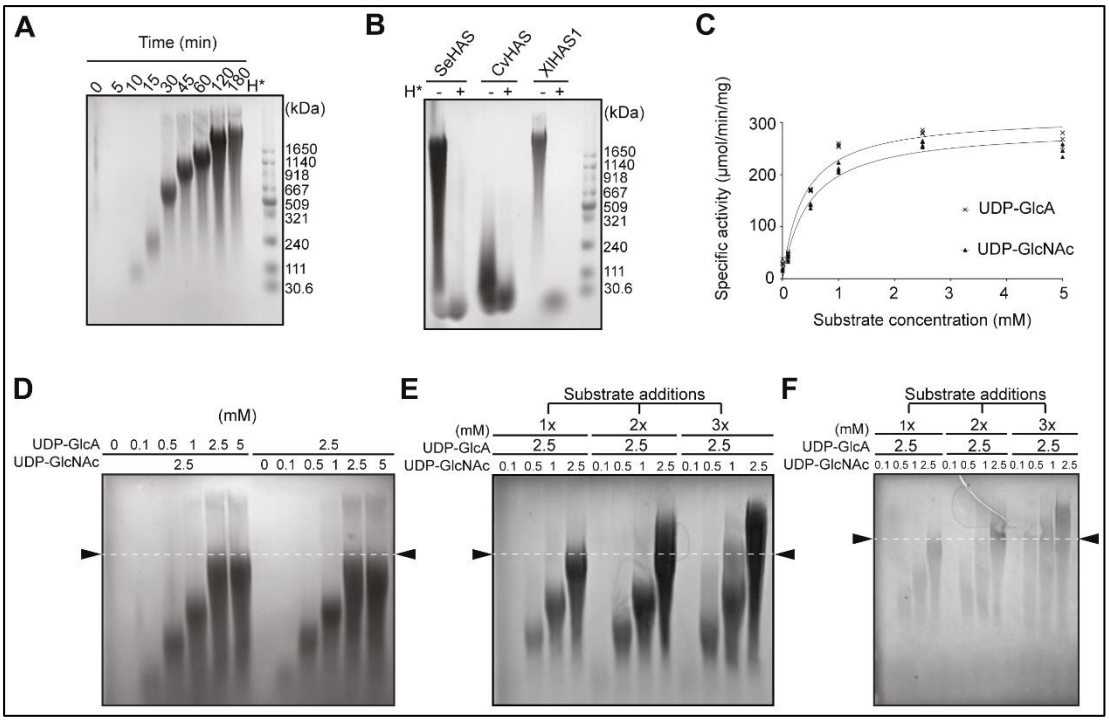


Figure 9 XlHAS1 biochemical characterization

A – Time-course of in vitro HA biosynthesis monitored by agarose gel electrophoresis and Stains-All staining. H*: Hyaluronidase digestion. The molecular weight marker represents HA

size standards. B – Comparison of HA synthesized in vitro by *Streptococcus equisimilis* (Se), *Chlorella* virus (Cv) and *Xenopus laevis* (Xl) HAS1. All proteins were assayed after purification in the GDN detergent. C – Catalytic activity depends on substrate concentration. One substrate concentration was at 2.5 mM while the other was varied as indicated. Activity is measured by quantifying the released UDP in an enzyme coupled reaction. D – The same as in panel C but by monitoring HA formation. E – Increasing HA molecular weight in the absence of product inhibition. A standard synthesis reaction was supplemented three times with 2.5 mM UDP-GlcA and the indicated amounts of UDP-GlcNAc while enzymatically converting UDP to UTP. F – The same as panel E, except UDP-GlcNAc concentration was kept constant (2.5 mM) and GlcA is varied. Arrowheads in panels D-F indicate maximum HA extension when both substrates are added once at 2.5 mM concentration.

2.4 Determining the structure of XlHAS1

A structural analysis of the vertebrate HAS provided unique insights into the mechanism of HA biosynthesis. It revealed the enzyme's overall architecture, the organization of its active site and the TM channel structure. The structure also informed further biochemical assays to analyze the mechanism of HA translocation and length control. Currently, a technique of choice for structure determination of membrane proteins, as well as proteins that are difficult to crystallize, is cryoEM [23, 47]. Unfortunately, at 63 kDa, XlHAS1 is quite small for cryoEM approaches [48]. A powerful strategy to overcome this obstacle is to conjugate a membrane protein of interest with a specific antibody fragment (Fab) [49]. Those antibody particles serve two purposes. First, they stabilize flexible regions of the protein improving the sample behavior and homogeneity. Second, they increase the overall protein mass, which improves signal to noise ratio in cryoEM and allows more precise 2D and 3D particle alignments.

2.4.1 Anti-XlHAS1 Fab discovery

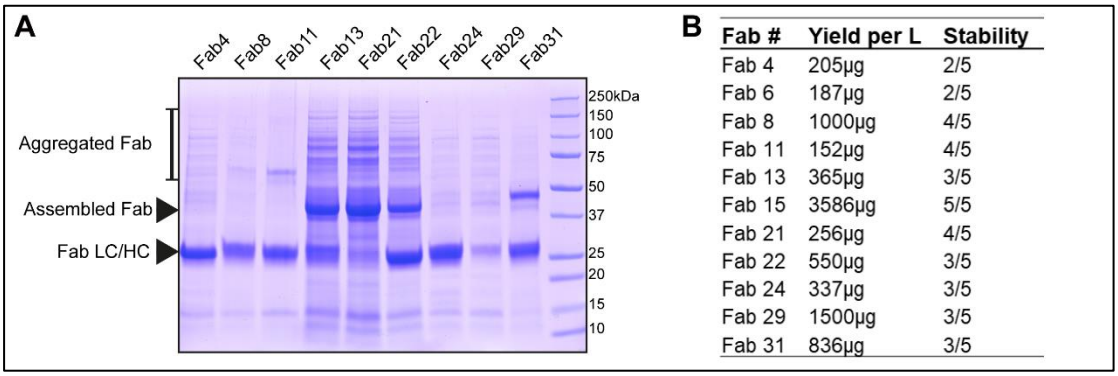
I collaborated with the Kossiakoff lab at the University of Chicago to identify XlHAS1-specific Fabs. Satchal Erramilli and Tomasz Gawda screened their lab's phage library E (consisting of 10^{10} different phages expressing Fabs on the surface [50]) by employing high throughput phage-display enzyme-linked immunosorbent assay (ELISA) against XlHAS1. They found 11 binders that seemed promising. I carefully screened those Fabs in-house to identify the best candidates as "cryoEM chaperones". An ideal binder must fulfill 3 important criteria: (1) the Fab itself, often rich in hydrophobic residues, must be stable enough to allow efficient expression, purification in milligram quantities, and not affect sample behavior in any negative way, (2) Fab-target protein interaction needs to be strong to allow efficient complex assembly and (3) the assembled complex should manifest a similar biological activity to that of the target protein alone.

2.4.2 Anti-XlHAS1 Fab purification

Fabs, unlike full antibodies, can be easily expressed in *E. coli*, which offers a great cost advantage. To this end, a standard lab strain of *E. coli* is

transformed with the RH2.2 plasmid carrying the Fab light chain (LC) and heavy chain (HC) genes [50]. The plasmid also contains a Lac promoter which allows induction of the Fab expression by the native *E. coli* RNA polymerase using lactose or its nonhydrolyzable derivative isopropyl-thiogalactoside (IPTG). The genes are expressed in conjunction with a periplasm-targeting signal sequences to facilitate disulfide bond formation between the LC and the HC. Properly formed Fab can be extracted from the bacterial periplasm using osmotic shock or the whole cells can be lysed using a microfluidizer as a more time-efficient strategy.

Fabs were purified using a specialized antibody purification resin with immobilized protein L, a bacteria-derived protein which shows high affinity for antibodies and fragments thereof (Figure 10A) [51]. The affinity of the resin is very high and the Fab needs to be eluted off the column using fairly harsh (pH=3.5) conditions and subjected to dialysis into a more compatible buffer, such as Tris-buffered saline. Finally, the obtained Fabs are concentrated to achieve highly concentrated stocks for subsequent uses. The majority of Fabs showed reasonable yields between 0.5-2 mg of protein out of 1 L of *E. coli* culture. Substantial precipitation of Fab4 was apparent during concentration, suggesting stability issues, which caused poor yields (200 µg out of 1 L of culture, Figure 10B). The same was observed with Fab6, Fab11 and Fab21, although precipitation of those was not as severe. Fab15 on the other hand, gave substantially better yields than others (3.5 mg per 1 L of culture, Figure 10B) and is one of the few Fabs that migrated as a well-defined peak on the SEC (see below). Overall, the purification results suggest that many Fabs are not stable, perhaps due to abundance of hydrophobic residues in their randomized variable loop regions.



Fab #	Yield per L	Stability
Fab 4	205µg	2/5
Fab 6	187µg	2/5
Fab 8	1000µg	4/5
Fab 11	152µg	4/5
Fab 13	365µg	3/5
Fab 15	3586µg	5/5
Fab 21	256µg	4/5
Fab 22	550µg	3/5
Fab 24	337µg	3/5
Fab 29	1500µg	3/5
Fab 31	836µg	3/5

Figure 10 Fab purification and characterization

A – Coomassie-stained reducing SDS-PAGE gel showing Fab purification results. B – Summary of protein yields obtained after Fab purifications and their stability. Stability score was assigned qualitatively based on purification yield, level of precipitation during protein concentration, and whether or not the Fab migrated as a well-defined peak on SEC.

2.4.3 Anti-XlHAS1-Fab complex activity profiles

As an initial binding assessment, the effects of the Fabs on XlHAS1's activity were tested. Expected binding should result in partial inhibition, suggesting interference with HAS' substrate binding site, or stimulation, perhaps induced by stabilization effects. This is an important selection criteria, as an ideal binder should not interfere with the catalytic activity of the target protein.

Purified Fabs were supplemented in an in vitro HA synthesis reaction with XlHAS1, the HA products were quantified by liquid scintillation counting, and analyzed for size using agarose gel electrophoresis. Interestingly, all but one of the tested Fabs showed activity-stimulating properties (Figure 11A), most likely due to thermal stabilization of XlHAS1's GT domain. Fab4 manifested poor solubility, which induced aggregation of the Fab-HAS complex contributing to partially abolished catalytic activity based on scintillation counting. Interestingly those results were corroborated by agarose gel staining of HA, which showed reduced HA lengths synthesized by XlHAS1 in the presence of Fab4 (Figure 11B). On the basis of this experiment, all of the Fabs were selected for subsequent binding assays.

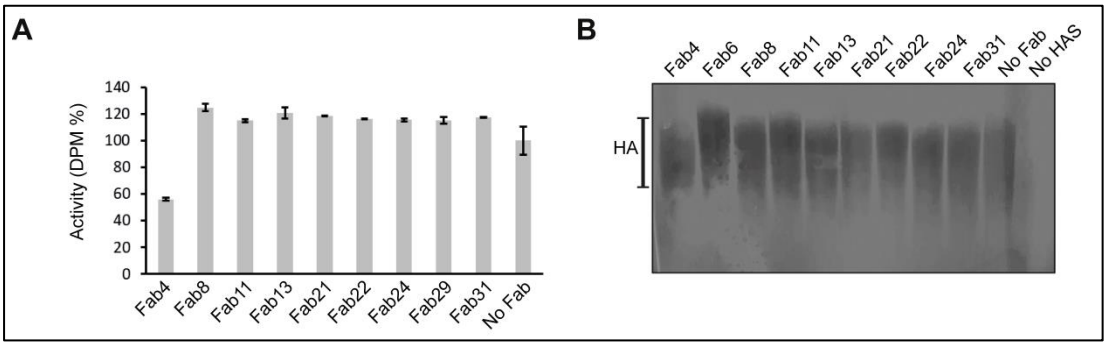


Figure 11 Fab influence on XlHAS1 activity

A – Results of the activity assay conducted on XlHAS1 incubated with the indicated Fab. The activity was quantified by liquid scintillation counting of the radioactive HA product, as described above. Experiments were performed in triplicates and error bars represent standard deviations. B – Same as panel A, except the radioactive label was omitted and the resulting HA product was subjected to agarose electrophoresis and Stains-All staining.

2.4.4 Anti-XlHAS1 Fab binding tests

Concentrated Fabs were incubated for 1h with the purified XlHAS1 at 5:1 molar ratio of Fab to XlHAS1, followed by a SEC run to separate XlHAS1 complexes from the unbound Fabs. A substantial number of the Fabs gave target peaks that were shifted to the left on SEC chromatograms, suggesting an increase in molecular weight (Figure 12A). Peak fractions corresponding to XlHAS1, possibly complexed with a Fab, were subjected to SDS-PAGE analyses to compare the band intensity and determine the strongest Fab binders.

Fabs bind XlHAS1 with varying efficiency, which is reflected by the Fab band intensities on the gels after SEC runs (Figure 12B). Based on these experiments, the library of 11 Fabs was short-listed to 3 that showed the most intense staining after co-elution with XlHAS1: Fab8, Fab15 and Fab22. Notably, out of the three, the excess Fab15 was the only one that eluted as a separate peak on SEC (Figure 12A), which, together with its substantial purification yield, (Figure 10B) suggests it is the most stable one.

SEC co-elution is not a definitive test for binding due to the possibility of Fabs accidentally co-eluting with the target protein due to aggregation. To validate the most promising binders, a pull-down assay was employed. Purified XlHAS1 was bound to NiNTA resin and incubated with one of the three Fabs. After extensive washing, the complexes were eluted and again analyzed using SEC and SDS-PAGE. The obtained results corroborate the SEC co-elution data as all of the tested Fabs appear to bind the NiNTA-immobilized XlHAS1 and the complexes survive stringent washing (Figure 11C).

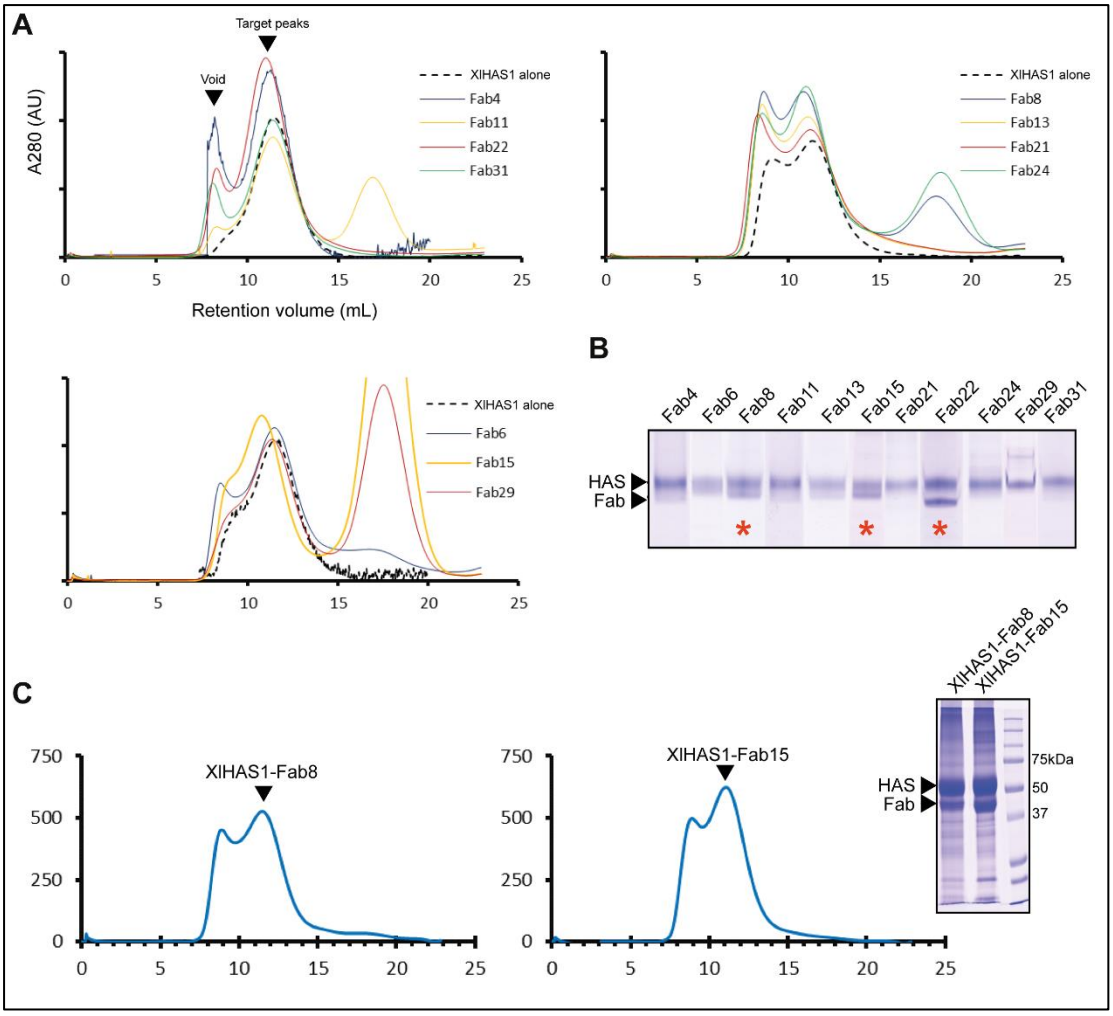


Figure 12 anti-XlHAS1 Fab binding characterization

A – Results of XlHAS1-Fab co-elution experiments. The SEC data was separated into three plots for clarity. Most Fabs cause XlHAS1 peak to increase and shift towards higher molecular weights, indicating binding. B – Coomassie-stained gel of the target peak fractions obtained

from panel A. Lanes marked with the red asterisk were determined to represent the strongest binders based on the intensity of the band that corresponds to the Fab. C – Results of the pull-down experiments. XlHAS1 was NiNTA-immobilized and exposed to a Fab. The resulting material was again subjected to SEC and analyzed on the gel.

2.4.5 XlHAS1-Fab complex cryoEM tests

The three most promising Fab binders were selected for cryoEM trials to determine their binding sites and their influence on the overall quality of the cryoEM sample. For sample preparation, similar procedures as described for the XlHAS1-Fab co-elution experiments were used, except scaled up to obtain enough sample for cryoEM experiments. The purified XlHAS1-Fab complexes for each combination were concentrated to 4 mg/mL and used for the preparation of 4-8 cryoEM grids. Grids were screened using the Glacios microscope at University of Virginia Molecular Electron Microscopy Core (MEMC) and the best grids were chosen for an overnight data collection. Subsequent cryoEM data processing revealed an outline of a detergent micelle with a small noisy density protruding from it, likely corresponding to the GT domain of XlHAS1. Intriguingly, cryoEM reconstructions showed no apparent density corresponding to the Fab8 (Figure 13A), while Fab22 appeared highly mobile and diffused in the dataset (Figure 13B). It is possible that those Fabs recognize flexible regions of XlHAS1 and do not contribute to the overall signal due to averaging. It is also possible that those Fab complexes broke down upon grid vitrification.

Excitingly, XlHAS1-Fab15 data revealed a well-ordered and rigid complex with readily detectable Fab15 attachment at the bottom of XlHAS1's GT domain (Figure 13C). This dataset, despite being collected on a sub-optimal microscope geared mostly toward screening purposes, already showed partial densities for most of the TMHs of XlHAS1 and allowed for confident docking of the Alphafold2 [52] XlHAS1 atomic model prediction (Figure 13D). This shows how substantially improved data can be obtained from XlHAS1 stabilized and enlarged by a suitable Fab.

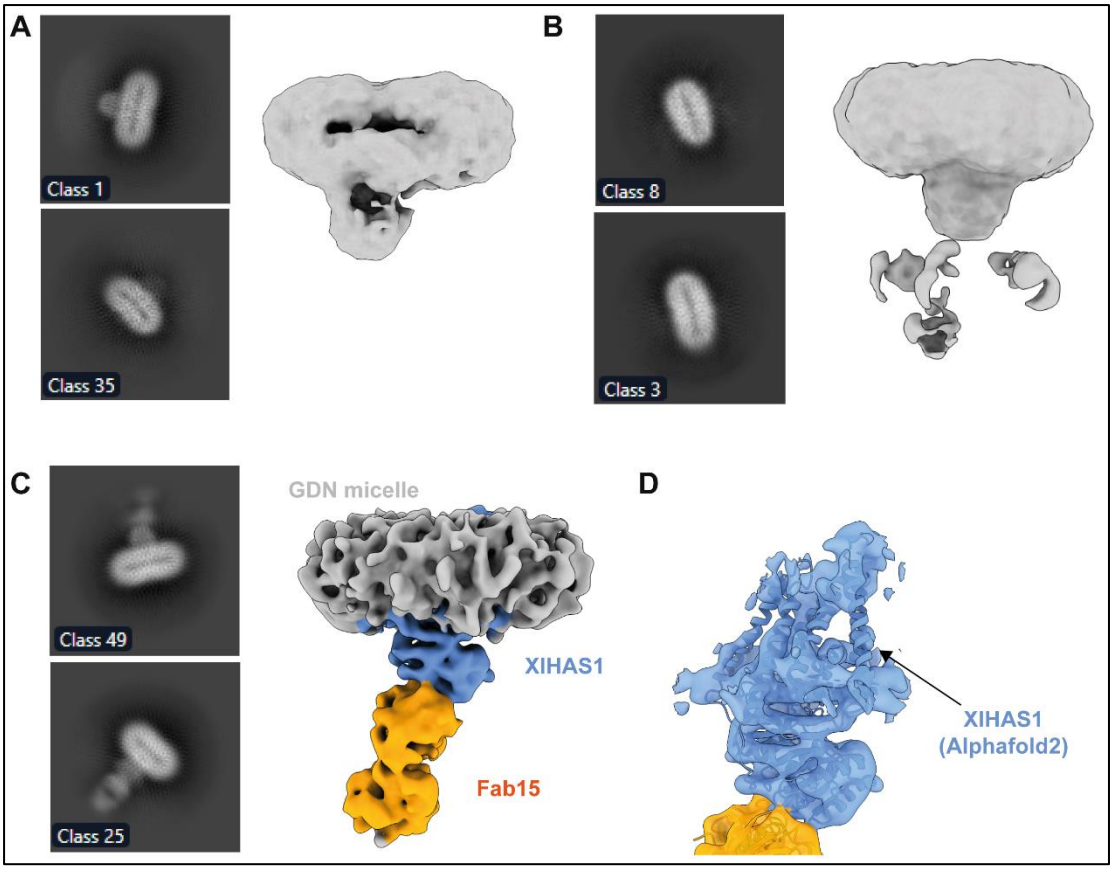


Figure 13 anti-XlHAS1 Fab cryoEM characterization

A – Results of cryoEM trials on the XlHAS1 complexed with the most potent binders. A – XlHAS1-Fab8 cryoEM 2D averages and a 3D reconstruction. B – XlHAS1-Fab22 cryoEM 2D averages and a 3D reconstruction. C – XlHAS1-Fab15 cryoEM 2D averages and a 3D reconstruction. D – Closeup on the XlHAS1-Fab15 map showing partial density for the TM region of XlHAS1. Alphafold2 XlHAS1 model prediction is rigid body docked into the map as a reference.

2.4.6 XlHAS1-Fab15 complex cryoEM data collection

With a stable XlHAS1-Fab15 complex in hand, I decided to optimize grid preparation conditions to obtain the highest possible sample quality for high resolution cryoEM analysis. One of the key optimization points was switching to a different cryoEM grid vendor – C-flat instead of Quantifoil. For detergent-free samples, such as nanodisc-reconstituted proteins, Quantifoil grids are usually the top choice. C-flat grids are typically not preferred due to their extremely thin carbon support film and susceptibility to cracking, forcing very careful handling of the grids. However, from our lab's experience, C-flat grids seem to facilitate better partitioning of detergent-solubilized protein particles into thin vitreous ice, which is crucial for obtaining cryoEM data with good contrast and high resolution. Additionally, to force protein particles into even thinner vitreous ice, the protein concentration for grid preparation was increased from 4 to 8 mg/mL. With good grids in hand, I collected data on the MEMC's Titan Krios, a top-of-the-line cryoEM microscope.

Due to the presence of the GDN detergent, the sample viscosity is low and the resulting vitreous ice thickness shows aggressive gradients within the grid holes. As a consequence, the coverage of particles over the grid holes is uneven with edges of the holes being more populated than the centers (Figure 14A). This necessitates the very careful, time-consuming, and often tedious process of target picking for automated cryoEM data acquisition within a very narrow window of grid holes with optimal ice thickness. Additionally, exposures are being collected closer to the edge of the holes where the particle coverage is typically the best. Even then, a lot of micrographs show only partial occupancy by protein particles. Although this approach yields a low number of particles per exposure, it generates data with very high contrast since most particles are embedded in the thinnest vitreous ice possible (Figure 14A, B).

2.4.7 XlHAS1-Fab15 cryoEM data processing and model building

A typical cryoEM data processing starts with importing the collected exposures into cryoSPARC, an advanced program package containing all necessary algorithms needed to process cryoEM data for model building [53]. The workflow starts with aligning the EM exposure stacks, correcting for drift, estimating contrast transfer function (CTF), and extracting images of individual protein particles (Figure 14C). Subsequently, the particles are averaged in a 2D classification job for sorting into different views and shapes, and for cleanup purposes. An additional consequence of the GDN detergent being present in the sample is heavy contamination of the cryoEM datasets with the empty detergent micelles. This enforces stringent curation of the data to obtain a set of high-quality particles suitable for high resolution reconstruction (Figure 14C). After careful sorting, the 2D-currated particle stack is subjected to an Ab initio job, aimed at 3D reconstructing reference-free volumes of the target protein. The job outputs a series of good quality volumes resembling the target protein, as well as a series of noisy volumes resulting from residual junk particles remaining in the dataset. Those volumes can be subsequently used for additional cleanup in 3D using Heterogenous refinement, which will funnel good particles into a good 3D volume class, and bad particles into a bad volume class (Figure 14C). This approach greatly complements 2D classification, as not all junk particles can be removed in 2D. Finally, high quality particles can be aligned with high precision using specialized 3D refinement jobs. Membrane proteins are characterized by a lot of disorder coming from the detergent micelle or other membrane mimetics. A specially designed algorithm, Non-uniform Refinement, is highly useful, as it is able to regularize 3D maps to account for spatial variability and yield higher resolution compared to classical refinements which assume that the entire structure is rigid [54]. Non-uniform refinement is greatly complemented by Local Refinement, which can additionally focus on user-specified areas of the map using targeted masks. By masking-out the micelle and the flexible domain of the Fab, I focused the Local Refinement on

the most ordered parts of XlHAS1-Fab complex, which further improved the quality of the map. Ultimately, I was able to obtain maps with estimated resolutions of roughly 3.2Å (Figure 14D). This resolution is sufficient to model most of the XlHAS1 and Fab15 residues into the map (Figure 14E).

Alphafold2 is a novel neural network-based algorithm that can predict protein structures based on sequence and previously solved homologous structures, with a never-seen-before level of accuracy [52]. The XlHAS1 structure was predicted using Alphafold2 and fitted into the EM map, which made the model building process more straight-forward compared to classical de novo approaches. The model was subsequently refined in an iterative fashion by alternating between refinement by hand in WinCoot [55] and automatically in Phenix [56], until a good map-model fit and chemical validation statistics were achieved (Figure 14E).

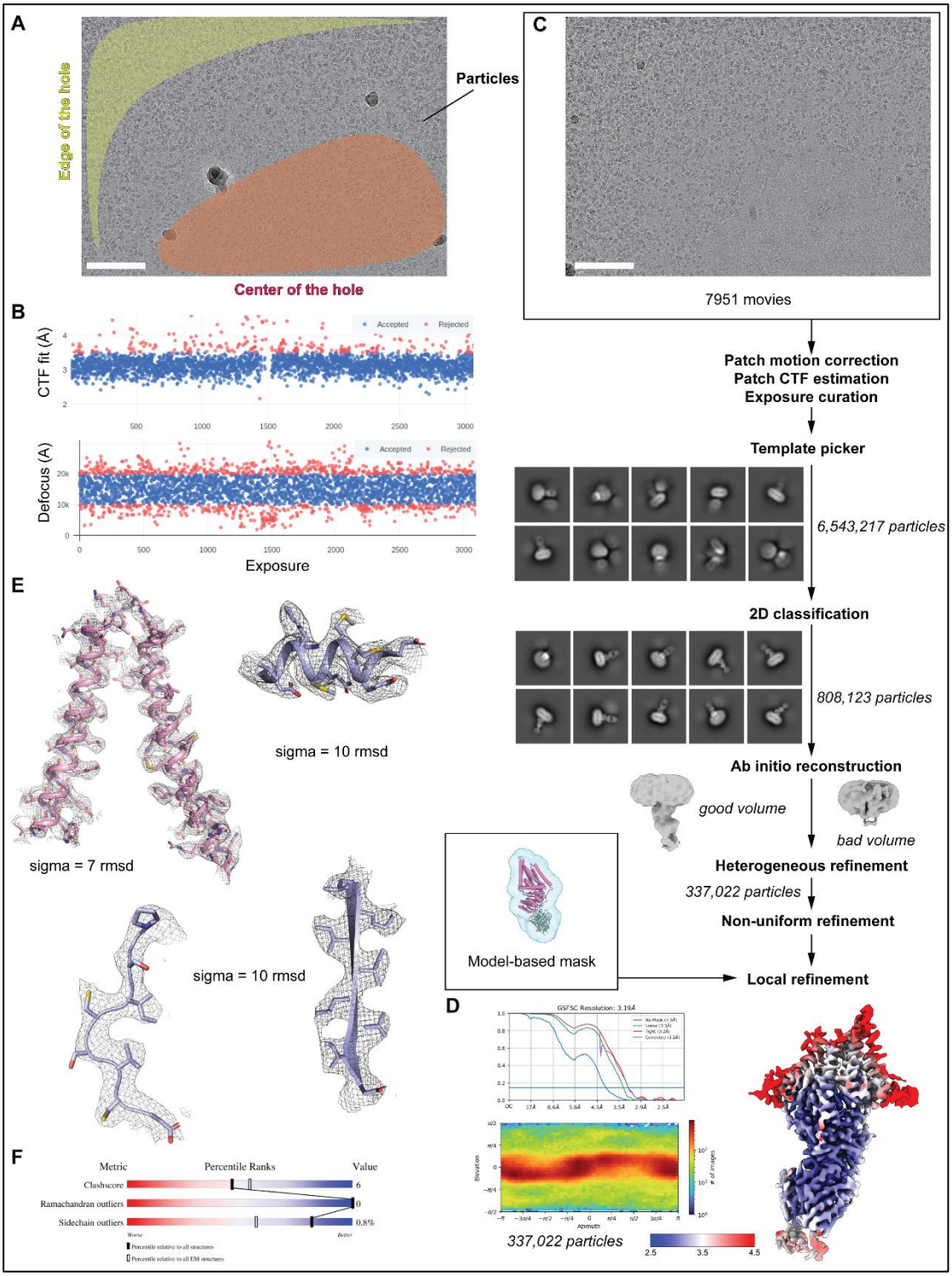


Figure 14 XlHAS1-Fab15 cryoEM data processing and model building

A – Example EM micrograph obtained for XlHAS1-Fab15 vitrified on C-flat grids. Edge, center of the holes, and XlHAS1-Fab15 particles are indicated. B – Data collection statistics showing average defocus and contrast transfer function (CTF) fit resolution of a subset of collected movies. Movies marked blue on the plots are within the target/acceptable range. C – Schematic representation of the cryoEM workflow for the XlHAS1-Fab15 complex. Scale bar on the micrograph corresponds to 200 nm. D – Quality of the final XlHAS1-Fab15 cryoEM map colored according to the local resolution. Gold Standard Fourier Shell Correlation (GS-FSC, with 0.143 cutoff) validation and particle viewing angle distribution plots are shown. E – Selected parts of the XlHAS1 model showing cryoEM data quality for α-helical, β-sheet and loop structural

elements of XlHAS1. CryoEM density is depicted as a mesh carved at the indicated sigma levels. F – Protein Data Bank (PDB) validation statistics.

2.4.8 XlHAS1 architecture

The overall architecture of XlHAS1 is similar to the recently solved structure of the viral HAS (CvHAS) published by our lab [18]. The membrane-embedded region consisting of 6 TMHs assembles into a teepee-like arrangement, which contains a channel sufficiently wide to accommodate a polysaccharide (Figure 15A-D). The entry to the putative HA channel is lined with residues contributed by three horizontal interface helices (IFHs 1-3). This is followed by a large cavity formed at the interface of TMH2, TMH3 and TMH4, which leads to two hypothetical channel exit points (Figure 15D). The lateral pathway is formed at the interface of TMH5 and TMH6 and opens widely towards the membrane bilayer. Interestingly, this interface is almost perpendicular to the membrane plane and imposes bilayer thickening such that the membrane is thicker near TMH5 and TMH6 (by about ~10Å) compared to the opposite, diagonal TM interface near TMH1 and TMH2. This could be attributed to an artifact caused by the presence of the GDN detergent micelle, yet the same observation was made for the CvHAS reconstituted into NDs [18]. The functional importance of this membrane distortion is unclear, but it may have mechanistic importance for HA secretion and deposition in the ECM. Notably, the structure revealed the long-debated position of TMH1, which is missing from all CvHAS reconstructions. TMH1 runs parallel to TMH3 and bundles with TMH2 and TMH3 at the putative top exit of the HA channel (Figure 15E).

The cytosolic domain of XlHAS1 contains a large GT-2 domain [57], which harbors the active site (Figure 15C). The active site is lined with highly conserved residues (some of which are also present in cellulose synthases [15, 58]), which are involved in coordinating the Mg^{2+} complexed substrates and mediate priming of the HA synthesis and translocation reaction, as described for CvHAS [18]. XlHAS1's cytosolic GT domain interacts with two N-terminal and four C-terminal TMHs as well as the three IFHs. IFH2 contains the conserved QxxRW motif (Figure S1) of which Trp382 is pivotal for positioning the acceptor sugar right at the entrance to the TM channel (Figure 15C).

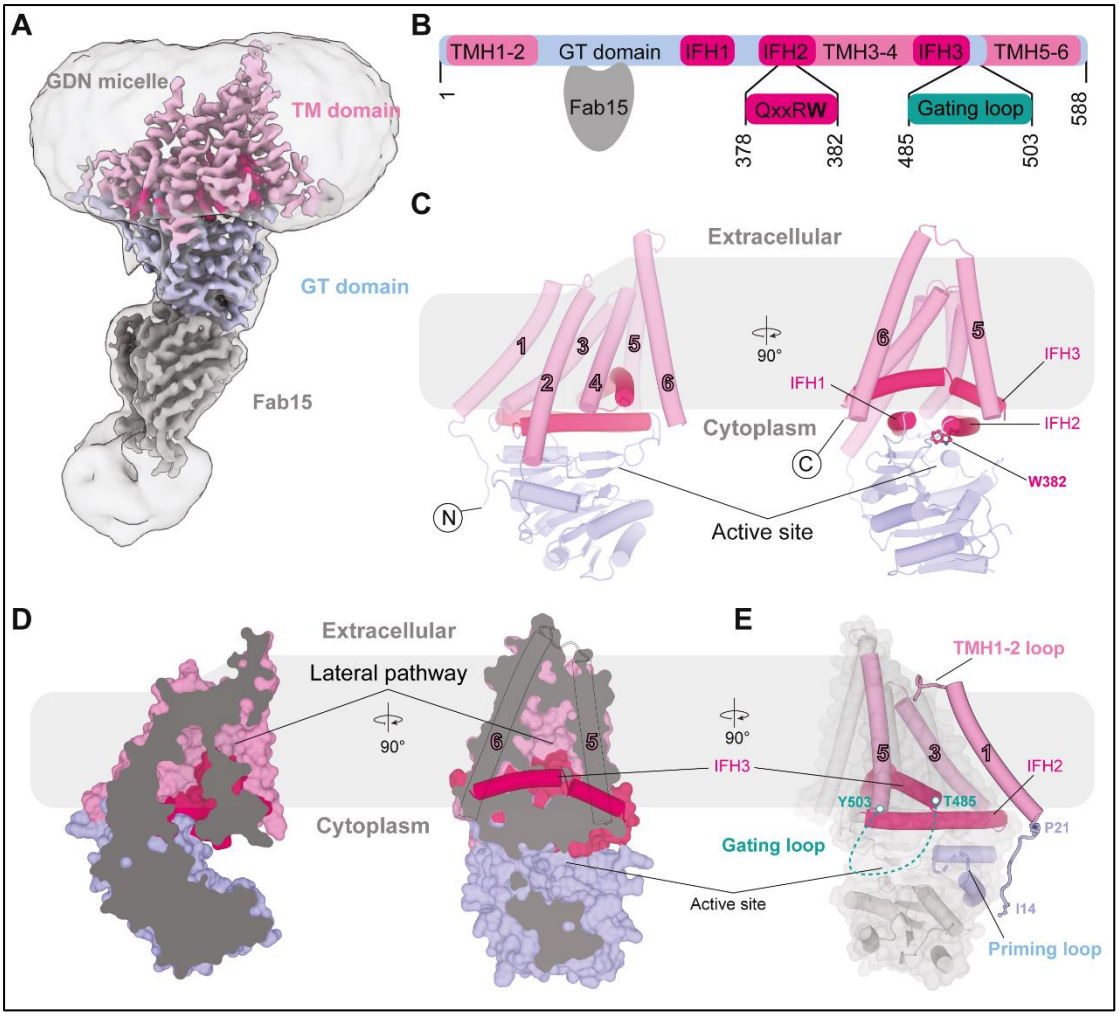


Figure 15 XlHAS1 model overview

A – CryoEM map of XlHAS1. The map is colored according to domain organization of XlHAS1 and was contoured at σ=9 r.m.s.d and overlaid with an Ab initio volume to visualize the detergent micelle and the flexible constant region of Fab15. B – Domain organization of XlHAS1 and Fab binding site. C – Architecture of XlHAS1. The catalytic domain, amphipathic interface helices (IFH), and transmembrane regions are colored blue, light pink, and pink, respectively. D – Surface representation of XlHAS1 indicating a curved channel with lateral pathway. E – Cartoon representation of XlHAS1 highlighting the positions of the TMH1 and the N-terminal extension, as well as the unresolved gating loop above the active site (dashed line).

Despite the excellent overall resolution of the cryoEM map (Figure 14D, Figure 15A), several parts of the XlHAS1 model are missing or remain poorly resolved. Vertebrate HASs contain a conserved WGTSGRR/K motif in a cytosolic loop connecting IFH3 with TMH5 (Figure S1). A similar motif in a loop close to the active site is found in chitin and cellulose synthases [59-62]. This loop has been shown to be important for the catalytic activity of bacterial HAS and is implicated in closing the active site with each substrate binding cycle in bacterial cellulose synthase [15]. Although the loop is unresolved in the apo XlHAS1 structure, its flanking residues Thr485 and Tyr503 at the C-terminus of IFH3 and beginning of TMH5, respectively, position it right above the catalytic pocket. It is unknown whether this loop fulfills a similar function in viral and

vertebrate HASs. In analogy to cellulose synthase and functional analyses described below, I refer to this loop as XlHAS1's gating loop (Figure 15E).

The other important region is the so-called priming loop (Figure 15E) that is conserved among HASs. Based on CvHAS analyses [18], the loop undergoes conformational changes with each substrate binding-hydrolysis-release cycle, which involves partial insertion of the loop into the catalytic pocket in the absence of substrate and its retraction after binding a new substrate molecule. No priming movement of the priming loop in XlHAS1 was apparent in cryoEM 3D variability analyses, perhaps due to the lack of substrates in the sample. The loop, although poorly resolved, seems to occupy mostly the retracted position.

Additionally, the loop that connects TMH1 and TMH2 and lids the exit of the putative HA translocation channel is highly flexible and poorly resolved (Fig 14E). The presence of the nascent HA polymer is expected to stabilize it in the open-channel conformation.

2.5 Determining the HA-bound conformation of XlHAS1

Capturing XlHAS1 in a nascent HA polymer-bound state revealed the open channel architecture, interactions of channel-lining residues with the HA, the exact HA secretion path, and provided insights into the HA translocation mechanism. The structure should also reveal insights into substrate coordination by XlHAS1, and the specific mechanism by which a strictly alternating sugar sequence of the HA polymer is achieved.

Since XlHAS1, unlike cellulose synthase, does not copurify with a nascent polysaccharide inside its TM channel, a method to reintroduce it after protein purification needed to be developed. In principle, with an adequate incubation time prior to cryoEM grid vitrification, XlHAS1 in the presence of both substrates should synthesize HA sufficiently long to span the length of the HA-secreting channel. A great advantage of cryoEM over X-ray crystallography is that it allows to sort out three-dimensional heterogeneity in the dataset. This allows us to solve several structures from a single dataset. In this case, with a sufficiently large and high-quality dataset, I should be able to sort out multiple conformational states that HAS goes through while synthesizing HA, such as the intermediate UDP-GlcNAc-bound and UDP-GlcA-bound states, as well as with progressively longer nascent HA polymers in the channel. Additionally, the presence of substrates in the active site is expected to stabilize the priming and gating loops, both adjacent to the active site and poorly resolved in the apo cryoEM map. Hence, such a dataset would provide a great wealth of knowledge and mechanistical insights into HA synthesis in vertebrates.

2.5.1 Attempts to trap XlHAS1-HA intermediate for cryoEM

The most straight-forward approach of preparing the XlHAS1-HA intermediate is to supplement the purified XlHAS1-Fab15 complex with substrates for HA synthesis (UDP-GlcA, UDP-GlcNAc, magnesium chloride) and prepare cryoEM grids after suitable incubation. Typically for cryoEM, sample handling is done at 4 °C to avoid protein aggregation. This creates a potential issue with much slower substrate turnover rates. Hence, sufficient time must be given to HAS to ensure the formation of a channel-spanning HA. Yet, the incubation time cannot extend for too long as the accumulation of HA in the sample may induce protein aggregation, alter the viscosity of the sample, or in other ways generate background noise in cryoEM.

My biochemical data shows that within 15 minutes of incubation at 37 ºC, XlHAS1 synthesizes HA polymers that surpass the length of the protein many times (Figure 9A). Hence, there is a possibility that the robust HA synthesis by XlHAS1 will induce protein aggregation and hinder the grid vitrification process. A fine tuning of incubation conditions was therefore required to achieve a good particle distribution on cryoEM grids. Reaction conditions that influence the rate of HA synthesis constituted the key optimization points and included substrate concentration, temperature, as well as incubation time.

Initial attempts to capture an HA-translocation intermediate of XlHAS1 failed due to rapid accumulation of HA indeed preventing preparation of a good quality cryoEM sample in 3 ways. (1) HA generated in the sample causes changes in viscosity, which greatly affects the process of vitrifying cryoEM grids. Once the sample is applied onto a cryoEM grid, the automatic grid plunger blots away excess liquid to ensure forming of suitably thin vitreous ice covering the holes of the grid. This is arguably the most important aspect of cryoEM grid preparation, as too thick of a vitreous ice reduces the signal to noise ratio, while too thin will not allow partitioning of the protein particles into the grid holes. HA accumulation greatly reduces the blotting efficiency causing the obtained vitreous ice to be overall much thicker. (2) XlHAS1-HA complexes will not partition into thin vitreous ice in the centers of the grid holes. When screening cryoEM grids, despite point (1), some grid holes with suitably thin ice could be found. Unfortunately, those contained very few protein particles (Figure 16A, B). The obtained cryoEM dataset revealed low resolution reconstructions of XlHAS1 bound to a single substrate and no polymer in the channel (Figure 16C). Presumably those reflect XlHAS1 particles that did not fully initiate HA synthesis during the allowed substrate incubation time (15 minutes at 4 °C), or that HA dissociated upon grid vitrification. Longer substrate incubation times (up to 45 minutes) made the effect of particle scarcity even more severe. (3) HA causes particle aggregation on cryoEM grids. Although the scarce particles that could be found using this approach were relatively well-separated and homogenous, imaging the carbon film surrounding the grid holes revealed big

clumps of aggregated material. It is unclear whether HA shows affinity for the carbon support film, or whether it causes HAS aggregation in solution and those aggregates preferentially partition to the carbon support film. Considering these points, an alternative approach of generating XlHAS1-HA intermediate needed to be developed.

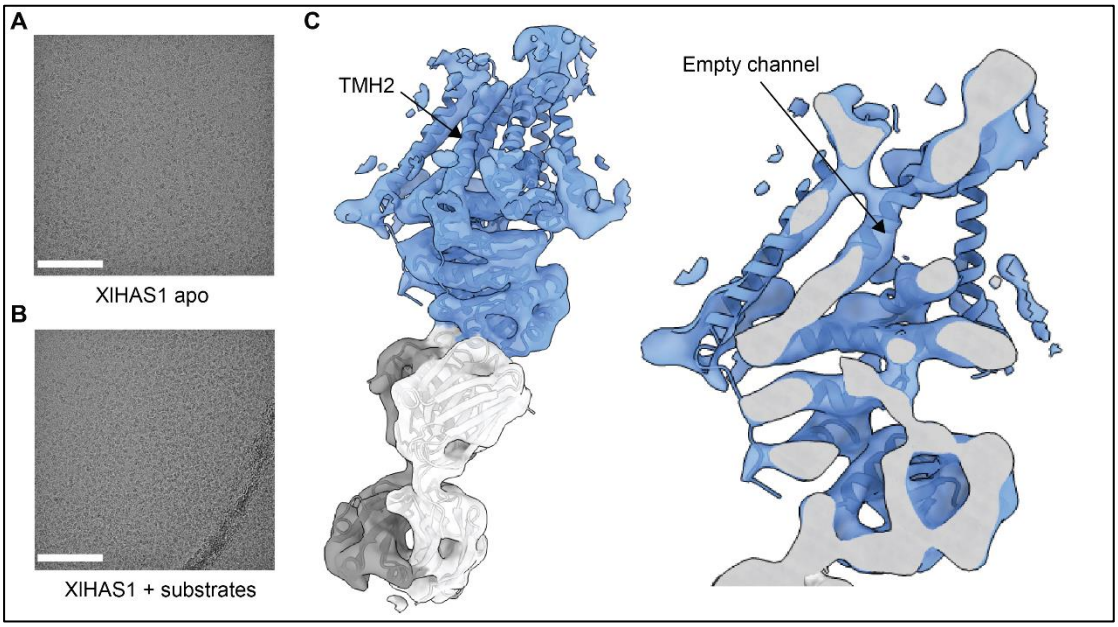


Figure 16 CryoEM of substrate-incubated XlHAS1

A – Representative cryoEM micrograph showing particle distribution for XlHAS1 apo sample. B – Representative cryoEM micrograph showing particle distribution for XlHAS1 incubated with substrates. Scale bar on the micrographs corresponds to 200 nm. C – 3D reconstruction obtained from cryoEM dataset after XlHAS1 was incubated with substrates.

To stabilize an HA-associated translocation intermediate, I performed an in vitro HA synthesis reaction in the presence of all necessary reaction components, as well as HA lyase (Figure 17A). In principle, only HA polymers that substantially protrude from HAS' channel and thus cause issues with grid vitrification will be accessible for digestion. Meanwhile, the crucial part of the HA that forms interactions with the channel will be protected. This approach eliminates time as an optimization point, since the HA synthesis reaction can simply be run overnight at 4 °C until all substrate is depleted, leaving a rich population of protein particles with HA polymers that are just long enough to span the channel. Additionally, since HA does not "stick out" of the synthase, it makes very little interactions with the surrounding solvent, making the XlHAS1-HA intermediate quite stable. The obtained species survive SEC, which can be utilized to separate them from other reaction components, such as the HA lyase, excess Fab15 or the released UDP (Figure 17A, B, D). Interestingly, performing this workflow without the HA lyase resulted in a similar SEC profile, except that a substantial fraction of the protein eluted in an unusually wide void peak, likely representing residual HMW HA-bound XlHAS1 (Figure 17C). Attempts of cryoEM analysis of that fraction failed due to a jelly-like consistency of the sample not allowing proper blotting of the cryoEM grids, resulting in

extremely dark grid squares (Figure 17E). Data collected for the fraction corresponding to the target peak revealed XlHAS1 in an apo conformation, likely due to HMW HA polymers being more prone to dissociate from HAS' channel during SEC (Figure 17F).

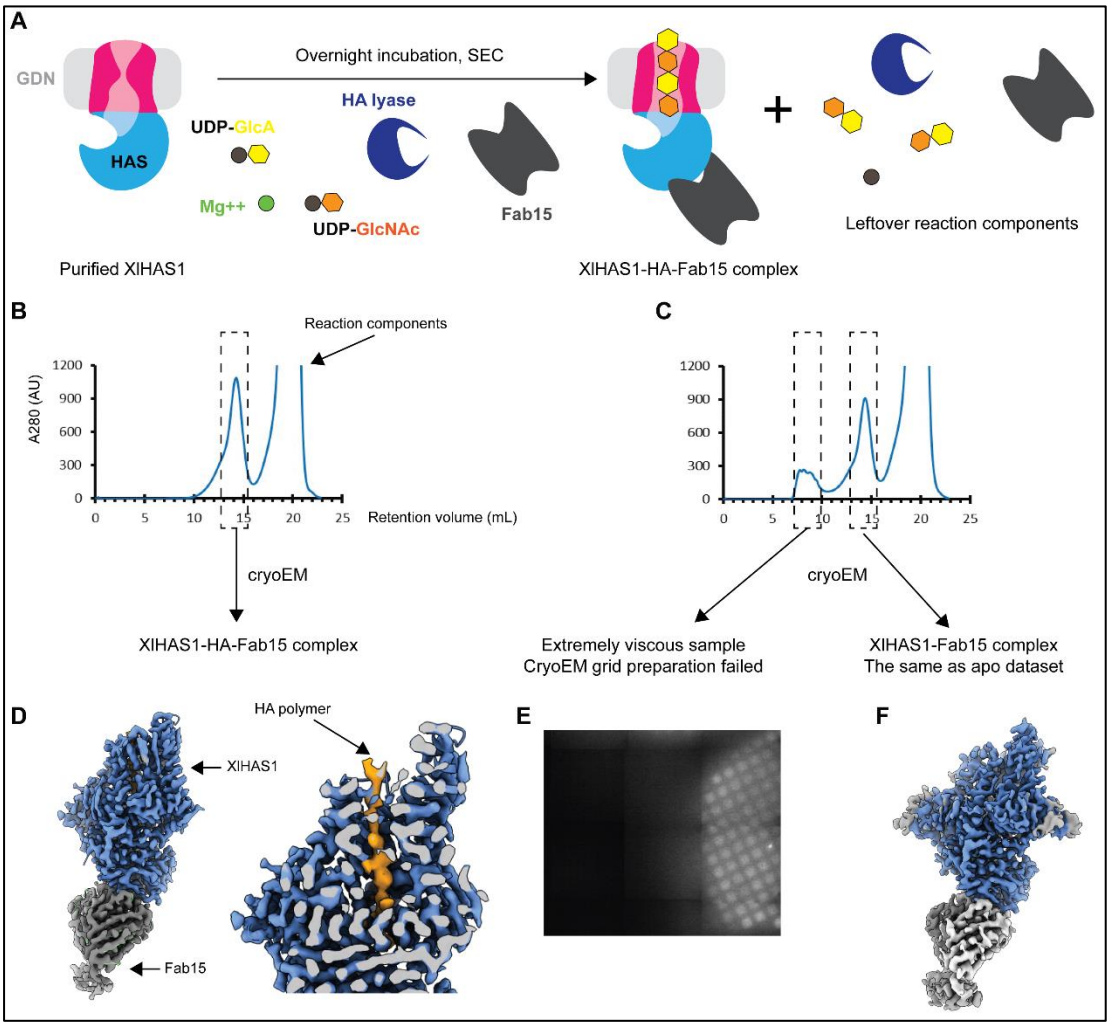


Figure 17 CryoEM sample preparation for HA-bound XlHAS1

A – Schematic representation of a sample preparation approach for obtaining XlHAS1-HA translocation intermediate. The purified XlHAS1 was incubated with excess Fab15, recombinant bacterial HA lyase, both substrates necessary for HA synthesis and Mg^{2+} ions. The mixture was incubated overnight at 4 °C and subjected to SEC purification to separate XlHAS1 with only a channel-spanning HA oligosaccharide bound to it. B – SEC chromatogram of the XlHAS1-HA intermediate prepared as described for panel A. C – SEC chromatogram of the XlHAS1-HA intermediate prepared as described for panel A but omitting the HA lyase. D – CryoEM map obtained for the sample shown in panel B. E – Atlas view of the cryoEM grid obtained for the void fraction of the sample shown in panel C. F – CryoEM map obtained for the target peak fraction of the sample shown in panel C.

2.5.2 CryoEM of XlHAS1-HA translocation intermediate

The described approach of preparing XlHAS1-HA intermediate by trimming the excess HA yielded a high quality cryoEM dataset (Figure 18A-C). The initial 3D reconstructions revealed high flexibility for some helices of XlHAS1's TM channel as well as the putative translocating HA. In addition to

extensive data cleanup using standard approaches (exposure curation, 2D classification, heterogenous refinement, as mentioned for the apo XlHAS1), a lot of additional 3D heterogeneity analyses needed to be performed.

3D variability analysis is a newly-developed algorithm that analyses conformational transitions within cryoEM datasets. As opposed to classical 3D refinements, the job outputs a continuous series of volumes which reveal conformational transitions within the target protein of interest [63]. 3D variability analyses of XlHAS1-HA cryoEM dataset revealed gating movements in XlHAS1's TMH1, TMH2, as well as the corresponding appearance of an extra density spanning the HAS channel (Figure 18A). Subsequent 3D classifications with focused masks covering those regions allowed me to refine a high resolution cryoEM map for a XlHAS1 channel-open, HA-bound state (Figure 18B,C).

The overall architecture of the XlHAS1-HA translocation intermediate resembles the apo XlHAS1. The main difference lies in the TM channel being open and occluded by the HA polymer (Figure 19A). No substrates were resolved in the cytosolic GT domain, as expected with this sample preparation approach (Figure 17A, see below).

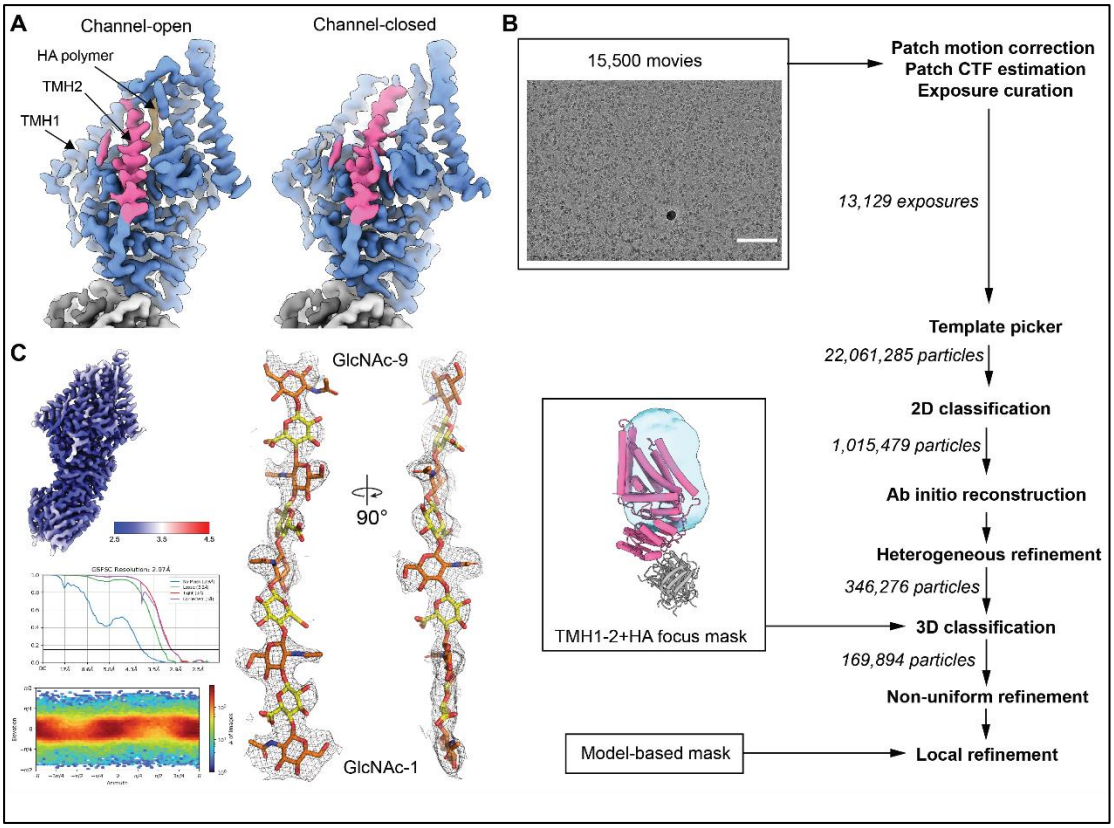


Figure 18 CryoEM workflow for HA-bound XlHAS1

A – Results of the initial 3D variability analyses showing the resolved channel-open and channel-closed states. TMH2 is colored pink and the putative HA density orange. B – Schematic representation of the cryoEM workflow for XlHAS1-HA intermediate. Scale bar on the micrograph corresponds to 200 nm. C – Final cryoEM map quality for the HA-bound XlHAS1

and the HA polymer inside the XlHAS1's translocation channel. HA density is depicted as a mesh carved at σ=6 rmsd.

2.5.3 HA modelling and register assignment

CryoEM maps of HA-bound XlHAS1 revealed a continuous non-proteinaceous density running from the catalytic pocket to its extracellular surface through the HAS' channel (Figure 17D, Figure 18A, C). The density accommodates nine glycosyl units, denoted 1–9 starting at the non-reducing end at the acceptor position (Figure 18C). The assignment of the HA register is based on the following observations. (1) The shape of the non-reducing end terminal glycosyl unit resembles a GlcNAc moiety and is consistent with maps of GlcNAc primed CvHAS [18, 64]. Following this register, the densities at the polymer's third and fifth glycosyl positions are also consistent with GlcNAc moieties, as expected for an alternating GlcNAc-GlcA repeat unit (Figure 18C). (2) This register is also consistent with previous molecular dynamics simulations of CvHAS demonstrating stable coordination of GlcNAc, but not GlcA, at the acceptor position [18]. Recent cryoEM data of GlcA-extended HA disaccharide-bound CvHAS [64] show that GlcA occupies the active site, allowing the polymer to maintain favorable GlcNAc-Trp382 stacking interaction. It is possible that there is a fraction of XlHAS1 particles with GlcA-terminated HA polymers in this dataset, yet the density protruding into the catalytic pocket (below the acceptor site) is extremely weak and cannot be interpreted with confidence. (3) The carboxylate substituents of the GlcA units are expected to be weak or even absent in the electron potential maps. Despite confidence in the register of the modeled HA chain, the exact orientation of the glycosyl units is less certain. Disaccharide repeat units of the HA polymer likely enter the translocation channel in two orientations with their acetamido and carboxylate groups pointing roughly in opposite directions. Thus, in addition to the modeled conformation (Figure 18C), I cannot exclude contributions from an alternatively oriented polymer to the observed cryoEM map.

2.5.4 HA channel gating transitions

HA translocation requires conformational changes of XlHAS1 to create a continuous TM channel (Figure 19A-C). Compared to the resting conformation described above, the N-terminal half of TMH2 near the extracellular water-lipid interface bends away from TMH4. Bending occurs around a conserved GLYG motif (residues 63 to 66, Figure S1) that places two Gly residues at the interface with TMH4 (Figure 19B). Furthermore, the extracellular helical turns of TMH1 and TMH2 unwind to form a short β-hairpin with residues of the TMH1-2 loop. Although only the loop's backbone is resolved in XlHAS1's apo conformation, it is evident that the loop lids the extracellular channel exit in the absence of HA. Upon channel opening, the loop flips towards the membrane and rotates by approximately 90º to run roughly

parallel to the membrane surface. Additionally, the N-terminal half of TMH6 bends by approximately 10º towards TMH2, around the conserved Gly547. In the new position, His543 is in hydrogen bonding distance to the backbone nitrogen of Ile58 (Figure 19B).

2.5.5 HA coordination inside the secretion tunnel

The first GlcNAc (GlcNAc-1) sits inside a collar of invariant residues, including Tyr288, Phe292, Cys307, Ser385, acceptor Trp382, as well as Asp342 and Arg343 of the conserved GD**DR** motif (Figure S1). Following the acceptor site, the conserved, positively charged Arg296 and Arg287 of IFH1 are in close proximity to GlcA-2 and likely stabilize the negatively charged GlcA units entering the channel. Past this point, the channel dimensions widen and HA's interactions with side chains are less extensive (Figure 19C).

About halfway across the membrane, the HA polymer is encircled by a Met-rich ring of hydrophobic residues, including Met69, Phe414, Ile418, Ile441, Met444, and Met472. Past this hydrophobic ring, the nascent HA chain is surrounded by moderately conserved hydrophilic and hydrophobic residues, including Tyr46, Glu49, Gln51, Ser61, Thr421, Leu425, Asn433, Trp436, and Cys440. This channel segment is strikingly devoid of positively charged residues, unlike the preceding section (Figure 19C).

2.5.6 HA orientation inside XlHAS1's channel

While HA's first three glycosyl units enter the channel in a co-planar conformation, the following GlcAc-4-GlcNAc-5 disaccharide is roughly 90° out of plane. The rotation is evident from the planes of the resolved glycopyranose rings (Figure 18C). It occurs at a central widening of the TM channel that can accommodate spontaneous structural rearrangements of the polysaccharide (Figure 19C), as observed in solution [65]. At this widening, His72 and Lys448 contribute to the channel's electropositive character (Figure 19C). Similarly, the next disaccharide unit (GlcA-6-GlcNAc-7) is rotated by about 45º relative to the preceding glycosyl units and the following disaccharide (GlcA-8 -GlcNAc-9) exhibits a similar rotation relative to the preceding pair (Figure 19C). Past GlcNAc-9, insufficient EM map quality prevents further interpretations.

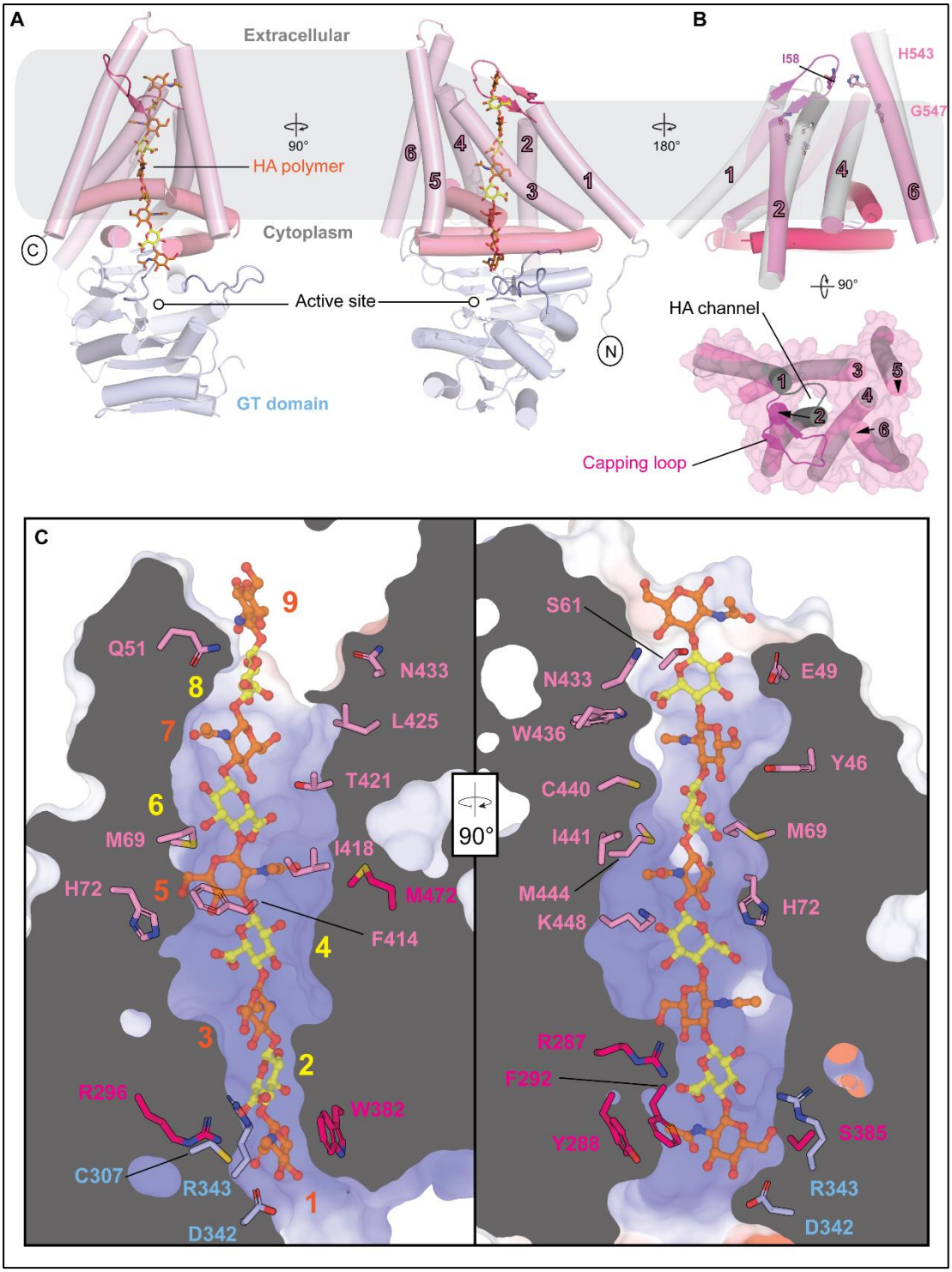


Figure 19 Overview of the HA-bound XlHAS1

A – Architecture of HA-bound XlHAS1. The catalytic GT domain, interface helices, and transmembrane helices are colored blue, light pink, and pink, respectively, as for Figure 15. GlcNAc substituents of the nascent HA (shown as sticks) are color orange for carbon atoms, whereas GlcAs are colored yellow. B – Gating transitions of the HA channel upon HA formation. Shown is a superimposition of the TM regions only in the presence (pink) and absence (white) of a translocating HA polymer. C – Stick representations of residues coordinating HA inside XlHAS1 TM channel. Channel surface is colored by the electrostatic potential calculated in PyMOL using the APBS plugin [66] (red to blue, -10 to 10 kT).

2.6 Determining the UDP-bound conformation of XlHAS1

The XlHAS1-HA intermediate described above was subjected to SEC purification before preparing cryoEM grids to separate it from other reaction components. Any residual substrates or UDP remaining in the active site are expected to diffuse out due to rapid dilution of the sample during the SEC run. Accordingly, this structure did not provide any insights into substrate coordination by XlHAS1, nor did it resolve additional conformational transitions within the priming loop or the gating loop. However, having trapped a stable and pure XlHAS1-HA intermediate provides an opportunity to attempt reintroducing one of the substrates to (1) generate both GlcA- and GlcNAc-terminated HA species for comparative studies and obtain unquestionable HA register within the channel, (2) determine UDP-substrate- (pre-hydrolysis) and UDP-bound (post hydrolysis) structures to gain insights into the step-wise HA extension and translocation as well as substrate coordination, and finally (3) stabilize the priming loop and the gating loop in alternative conformations. The latter will enable determining their exact roles in substrate coordination and transfer, and provide insights into their dynamics.

2.6.1 CryoEM sample preparation and processing

CryoEM sample preparation followed a similar approach as for the XlHAS1-HA structure, except the purified XlHAS1-HA intermediate (presumably with mixed terminating sugars) was incubated with one of the substrates (either UDP-GlcNAc or UDP-GlcA) briefly before preparing cryoEM grids. In principle, this should enrich a population of HA species terminated by this specific sugar (Figure 20A).

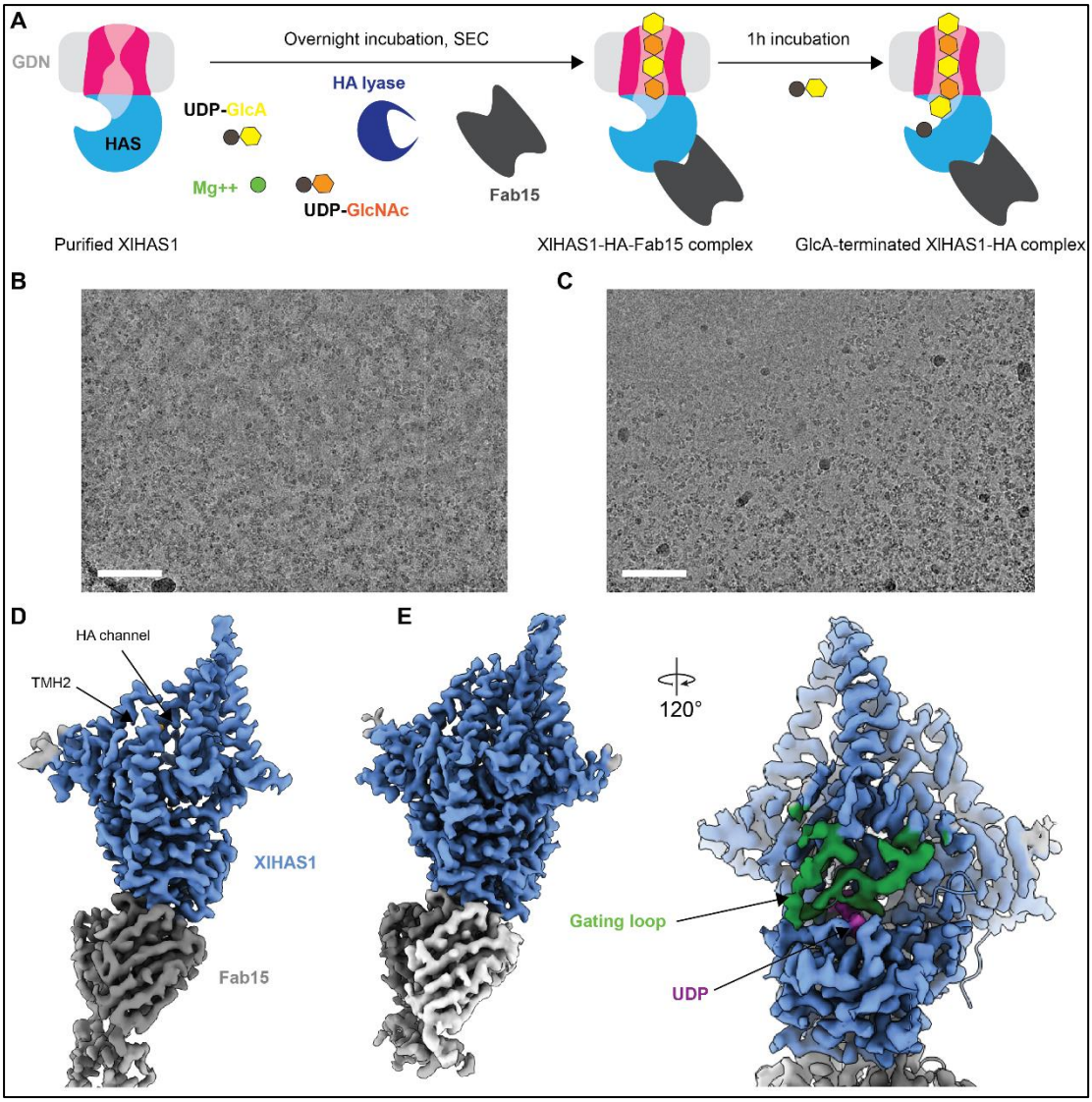


Figure 20 CryoEM sample preparation for UDP-bound XlHAS1

A – Schematic representation of a sample preparation approach for obtaining an XlHAS1-HA translocation intermediate terminated by a specific sugar residue. Depicted here is an attempt to introduce UDP-GlcA into the XlHAS1-HA translocation intermediate and obtain only species terminated by GlcA. UDP-GlcNAc introduction was also attempted in a similar fashion. B – Representative cryoEM micrograph for XlHAS1-HA intermediate incubated with UDP-GlcA. The presence of this substrate caused vitreous ice artifacts. C – Representative cryoEM micrograph of an XlHAS1-HA intermediate incubated with UDP-GlcNAc. Scale bar on the micrograph corresponds to 200 nm. D – 3D reconstruction of the GlcA-terminated XlHAS1-HA intermediate. E – 3D reconstruction of the GlcNAc-terminated XlHAS1-HA intermediate. View of the catalytic pocket with the UDP moiety and well-resolved gating loop is also shown. The channel-bound HA polymers for the datasets shown in panels D and E were lost for unknow reasons, which likely contributed to increased TMH2 flexibility.

CryoEM datasets were collected for both combinations (GlcNAc- and GlcA-terminated HA in the channel). GlcA addition caused vitreous ice artifacts which affected the signal to noise ratio. Substantially better data was obtained for GlcNAc-incubated sample (Figure 20B, C). For both datasets, cryoEM reconstructions revealed XlHAS1 with channel closed, polymer-free states (Figure 20D, E), suggesting that incubation with an abundance of a single substrate led to polymer release for unknow reason. The top half of the TMH2

was poorly resolved, suggesting some fraction of particles retained the polymer, which contributed to TMH2 flexibility and noisy refinements. Unfortunately, this fraction of particles was not sufficient for high resolution reconstruction of a polymer bound-state. Notably, the supplied substrate underwent hydrolysis and the resulting UDP was retained in the active site and is well resolved in both datasets. Consequently, the gating loop, positioned just above the catalytic pocket, was found to interact with the UDP and was sufficiently ordered to allow modeling (Figure 20E, Figure 21C).

CryoEM data processing was carried out using similar approaches as described for the XlHAS1-HA dataset. On order to obtain the best possible quality for the UDP ligand and the gating loop, 3D classification approaches were employed (Figure 21A-C).

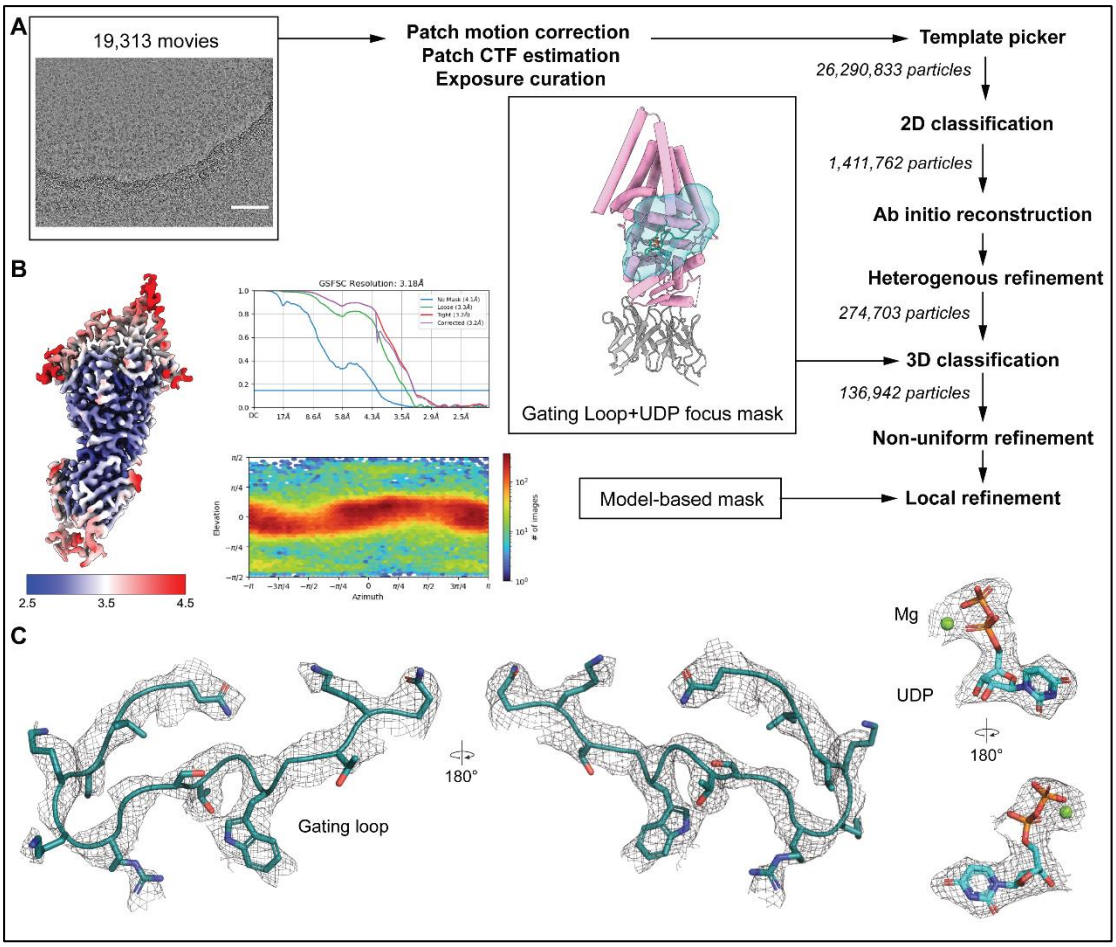


Figure 21 CryoEM workflow for UDP-bound XlHAS1

A – Schematic representation of the cryoEM workflow for XlHAS1-HA intermediate. Scale bar on the micrograph corresponds to 200 nm. B – Final cryoEM map quality for the UDP-bound XlHAS1. C – Gating loop and UDP density depicted as a mesh carved at σ=6 rmsd.

2.6.2 UDP-bound, gating-loop-inserted state of XlHAS1

The overall architecture of the UDP-bound XlHAS1 is consistent with the apo XlHAS1 structure (Figure 22A). The translocation channel is closed,

although some flexibility in TMH2 is apparent and the TMH1-TMH2 loop remains poorly resolved (Figure 21B). In this structure, UDP is coordinated via conserved motifs of the GT domain (Figure 22B), as previously described for CvHAS and other processive and non-processive GTs of the GT-A fold [3, 18, 59, 60, 67, 68] (Figure S1). Importantly, the structure shows XlHAS1's gating loop, connecting IFH3 with TMH5 (residues 485 to 503) and containing the conserved WGTSGRK/R sequence (residues 491 to 497). The loop inserts into the catalytic pocket and interacts with the nucleotide, similarly to related GTs (Figure 22C). In this conformation, Trp491 forms a cation-π interaction with Arg381 of the QxxRW motif in IFH2, which in turn forms a salt bridge with the nucleotide's diphosphate group. The indole ring of Trp491 runs approximately perpendicular to the uracil moiety, placing its Nε within hydrogen bonding distance to UDP's α-phosphate. This phosphate group is also contacted by the side chain of the following Thr493. Arg496, the penultimate residue of the WGTSGRK/R motif inserts into a negatively charged pocket formed by Glu109, Asp242, Glu367 and UDP's α-phosphate (Figure 22B).

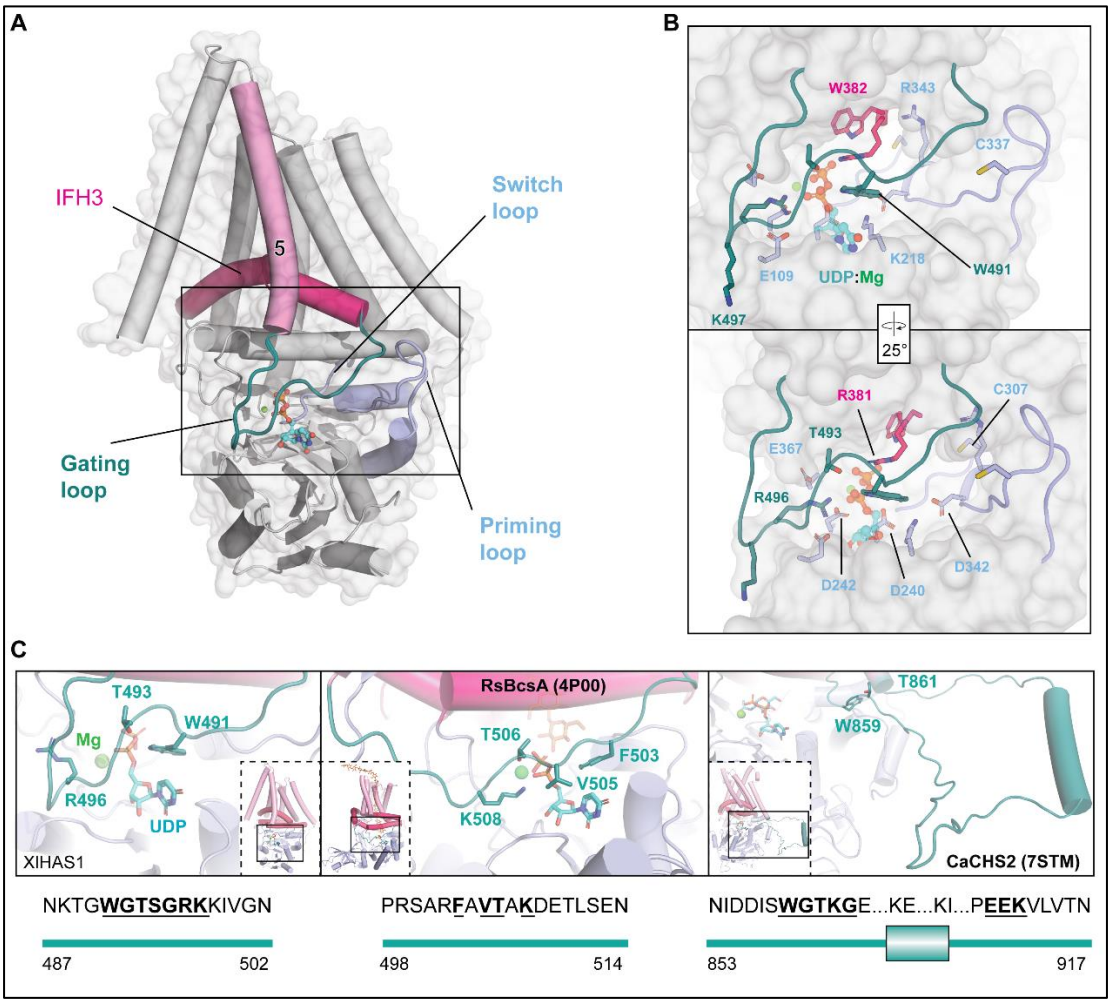


Figure 22 Overview of the UDP-bound XlHAS1

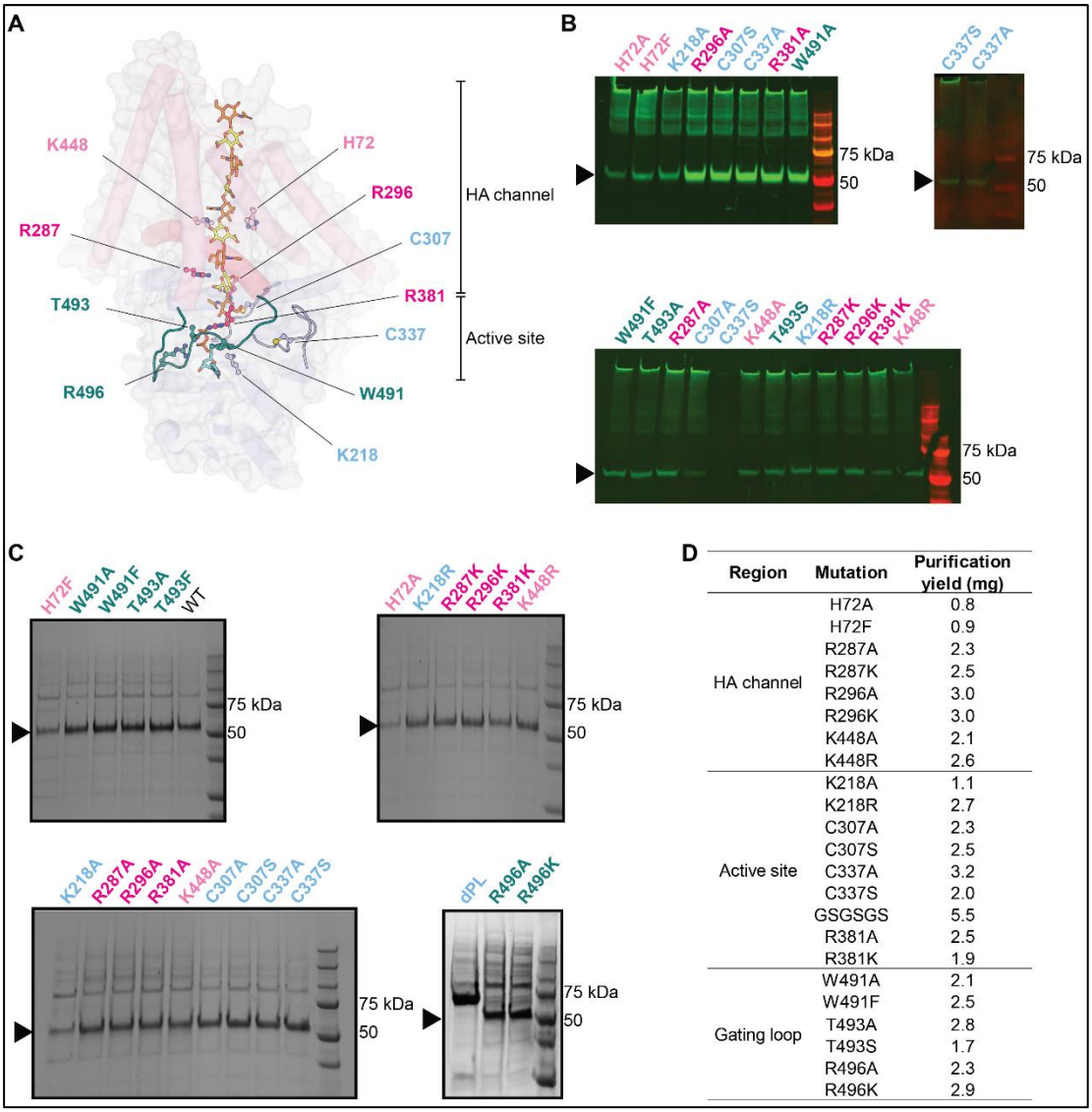
A – Overview of the UDP-bound XlHAS1 model. Active-site-proximal loops are depicted as thick teal and blue cartoon for gating loop, switch loop and priming loop, respectively. B – Closeup on the catalytic pocket of XlHAS1. Conserved residues contacting UDP are shown as sticks.

UDP is shown as ball-and-sticks colored cyan for carbon atoms. C – Gating loop comparison in XlHAS1, bacterial cellulose synthase (RsBcsA), and chitin synthase (CHS). Conserved motifs within the gating loop sequence for each GT representative are underlined.

2.7 Substrate processivity affects HA size in an unexpected way

Guided by the structures, I used site-directed mutagenesis to further assess the functional importance of various residues involved in substrate binding and hydrolysis in XlHAS1, as well as residues responsible for HA coordination (Figure 23A). Substitutions were introduced using standard mutagenesis approaches (see Methods). Amino acids of interest were replaced either with a similar residue that could fulfil the original function or an alanine. All XlHAS1 mutants were expressed and purified as described for the wild type (WT) XlHAS1 for in-depth biochemical analyses. Notably, all variants expressed reasonably well (Figure 23B) and most of them, except for a few channel mutants, could be obtained in WT amounts or larger, in a pure form (Figure 23C, D). This suggests that none of the substitutions had a substantial impact on the enzyme stability and folding.

All of the mutations were subsequently assessed using the HA gel electrophoresis assay to determine the amount and the size of the synthesized HA. HA accumulation was tested after a “short reaction” (1 hour at 37 °C), during which substrates are expected to remain in abundance. Additionally, effects of extended incubation time (“long reaction”, 8 hours at 37 °C) were also tested. This data is complemented well by the UDP release assay, which measures the rates of substrate hydrolysis (as mentioned above). Substrate turnover rates are expected to correlate with the HA accumulation, yet, as described below, that is not always the case.



D

Region	Mutation	Purification yield (mg)
HA channel	H72A	0.8
	H72F	0.9
	R287A	2.3
	R287K	2.5
	R296A	3.0
	R296K	3.0
	K448A	2.1
	K448R	2.6
Active site	K218A	1.1
	K218R	2.7
	C307A	2.3
	C307S	2.5
	C337A	3.2
	C337S	2.0
	GSGSGS	5.5
	R381A	2.5
	R381K	1.9
Gating loop	W491A	2.1
	W491F	2.5
	T493A	2.8
	T493S	1.7
	R496A	2.3
	R496K	2.9

Figure 23 Overview of XlHAS1 mutants generated for this study

A – XlHAS1 composite model showing mutagenized residues forming the active site (colored blue and teal for residues of the GT domain and the gating loop, respectively) or the HA channel (colored light pink and pink for residues of TMHs and IFHs, respectively). B – Western Blot analyses of Sf9 cell lysates after expression of the indicated XlHAS1 variants. Mutations are colored as for panel A. C – Coomassie-stained SDS-PAGE gels showing purity of all XlHAS1 mutants. D – Summary of generated XlHAS1 mutants and their purification yields expressed in milligrams per 1 L of Sf9 culture. Most mutants yielded wild type levels of protein yield.

2.7.1 Cysteines at the active site

XlHAS1's active site contains two conserved cysteines. Cys307 belongs to the so-called switch loop at the back of the active site, adjacent to the GlcNAc acceptor (Figure 23A). Replacing this Cys with Ala or Ser leads to the production of low amounts of polydisperse HA (Figure 24A). Accordingly, substrate turnover is substantially reduced to about 20% of that obtained for the WT enzyme (Figure 24B). This suggests that Cys307 is important for function, most likely to position the acceptor. The second Cys, Cys337, is part of the priming loop located to one side of the catalytic pocket (Figure 23A). This Cys

is the most conserved residue of the priming loop, yet it can be replaced with Ala and Ser, resulting in HA products similar to WT, with a slight reduction in HA length and substrate turnover rates for the C337S mutant (Figure 24A, B). It is unclear what the exact role of this conserved Cys337 is, leaving alone the function of the priming loop. Mutations targeting a wider residue range of the priming loop, aiming to make it more flexible, are also well-tolerated by the enzyme, yet no expected stimulation of activity was observed (Figure 24C, D).

2.7.2 Residues of the gating loop

The other batch of XlHAS1 mutants aimed at delineating the role of the gating loop in substrate turnover and consequently, HA size regulation. Replacing Trp491 of the **W**GTSGRK/R motif with Ala (W491A), which interacts with the UDP moiety (Figure 23A), abolishes HA biosynthesis and substrate turnover (Figure 24A, B). Partial activity can be retained by a more conservative substitution with Phe (W491F). This mutation results in a reduced yield of a lower molecular weight HA product in a short HA synthesis reaction, but the polymer becomes slightly longer than the WT product in a long synthesis reaction (Figure 24A). Accordingly, this variant shows reduced substrate turnover rates, compared to the WT enzyme (Figure 24B). Likewise, replacing the following Thr493 (WG**T**SGRK/R) with Ala or Ser reduces the catalytic rate to about 20–25% of that of the WT enzyme (Figure 24B). Strikingly, the T493A and T493S mutants generate HA exceeding the size of the WT product even during a short synthesis reaction. Further, replacing Arg496 (WGTSG**R**K/R) with Ala abolishes catalytic activity while an R496K mutant shows reduced substrate turnover rates and produces polymers of slightly increased length and greater polydispersity in short and long synthesis reactions (Figure 24A, B). The size differences are even more pronounced on a lower percentage agarose gel (Figure 24E). In vivo, similar effects may be achieved by limiting substrate availability and/or post-translational modifications of the enzyme [69, 70].

HAS' gating loop resembles the corresponding loops in cellulose and chitin synthases [60, 61] (Figure 22C). While the precise function of the loop is unclear, site directed mutagenesis experiments of HA demonstrate its profound importance for catalytic activity. Controlling its ability to interact with the substrate at the active site could be a regulatory mechanism, similar to bacterial cellulose synthase [62].

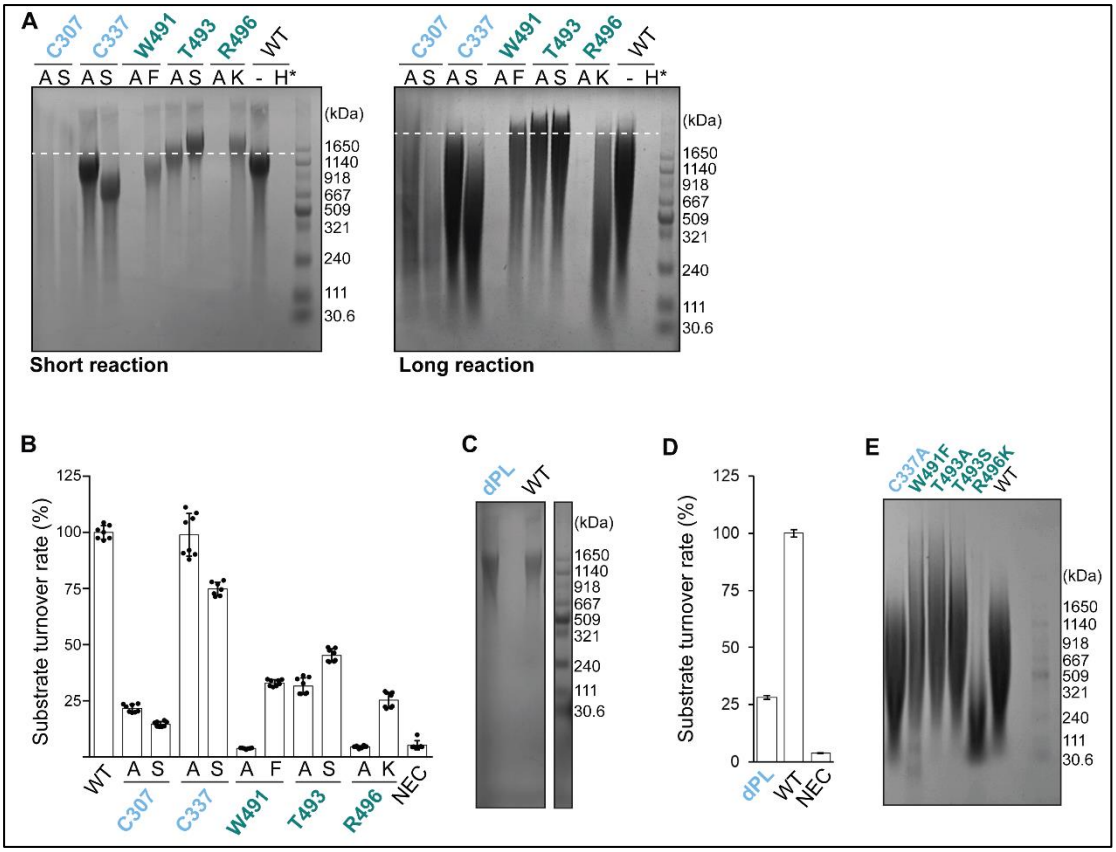


Figure 24 Mutagenesis of active site-lining residues

A – In vitro HA biosynthesis by the indicated XlHAS1 mutants for 1 h or 8 h (short and long reactions, respectively). WT - wild type, H* - hyaluronidase-treated samples. The dashed line roughly indicates the WT's highest molecular weight product. B – Catalytic rates of the mutants shown in panel A. Rates, normalized to WT, report substrate turnovers and were determined by quantifying UDP release in real time during HA synthesis. NEC - No Enzyme Control. Experiments were performed in eight technical replicates (n=8) and error bars represent standard deviations. C – In vitro HA biosynthesis by the XlHAS1 priming loop substitution mutant (GTYCTLGDDR sequence substituted by GSGSGSGDDR, referred to as "dPL"). D – Same as panel C but quantifying substrate hydrolysis. Experiments were performed in four replicates (n=4) and error bars represent standard deviations. E – Electrophoresis of HA species synthesized by selected active site-lining residue mutants, analyzed on a lower percentage agarose gel (0.5% instead of 1%) for better visualization of HMW HA species.

2.8 HA coordination is important for maintaining HA size

Processive HA biosynthesis requires sustained HAS-HA interactions between the elongation steps. To test how HA coordination affects the HA length distribution, I altered conserved, positively charged active site and channel-lining residues that likely contact HA in the translocation process and monitored HA size and UDP release (as described above).

XlHAS1's active site contains a conserved lysine (Lys218) and an arginine (Arg381) that likely contact the UDP moiety of the incoming substrate and/or stabilize the carboxyls of the incoming GlcA units (Figure 23A). Replacing Lys218 with Ala or Arg abolishes substrate turnover and HA synthesis (Figure 25A, B). Similarly, Arg381 of the Qxx**R**W motif, located in IFH2 right above the catalytic pocket, cannot be replaced with either Ala or Lys

(Figure 25A, B). Taken together, these results indicate that these residues of the active site are essential, most likely for substrate coordination.

Inside the TM channel, the polymer's GlcA-2 carboxyl group is positioned near Arg287 and Arg296 of IFH1. Substituting these residues with Ala reduces substrate turnover rates to about 15 and 30%, respectively, compared to WT. The resulting HA species are of very low molecular weight, not exceeding 100 kDa (Figure 25A). Interestingly, introducing more conservative substitutions (Lys) results in XlHAS1 mutants with similar or even increased substrate turnover rates compared to WT (Figure 25B). Despite substrate hydrolysis rates being largely unaffected, these mutants produce extremely polydisperse (R287K) and low molecular weight (R296K) HA products (Figure 25A). These effects are consistent between short and long HA synthesis reactions, suggesting that they arise mainly from reduced interactions between HA and XlHAS1, while substrate binding and hydrolysis are largely unaffected. This data suggests that these residues at the entrance to the TM channel are critical in stabilizing the nascent HA chain for elongation.

Farther inside the channel, His72 and Lys448 are adjacent to negatively charged GlcA-4 (Figure 25A). Compared to the WT enzyme, the H72A, K448A and K448R mutants display between ~50-75% and the H72F substitution about 10% of catalytic activity (Figure 25B). In short synthesis reactions, all mutants produce HA polymers of reduced length, with the H72A product being closest to WT. In a long synthesis reaction, however, the H72F and K448R variants produce polymers equivalent to or exceeding the HA length obtained from the WT enzyme (Figure 25A, C). This is likely due to the fact that Phe at this position is too bulky and hydrophobic, resulting in reduced HA transport rate, which in turn inhibits substrate hydrolysis. Similarity, the K448R mutant likely makes more extensive interactions with the HA polymer, again slowing down the transport and subsequently the substrate turnover. The H72F mutation was introduced as a harsher substitution compared to the H72A variant since CvHAS has a Phe at this position. It is unclear why CvHAS has a sub-optimal residue at this position, yet the necessity for lower processivity could be one of the reasons.

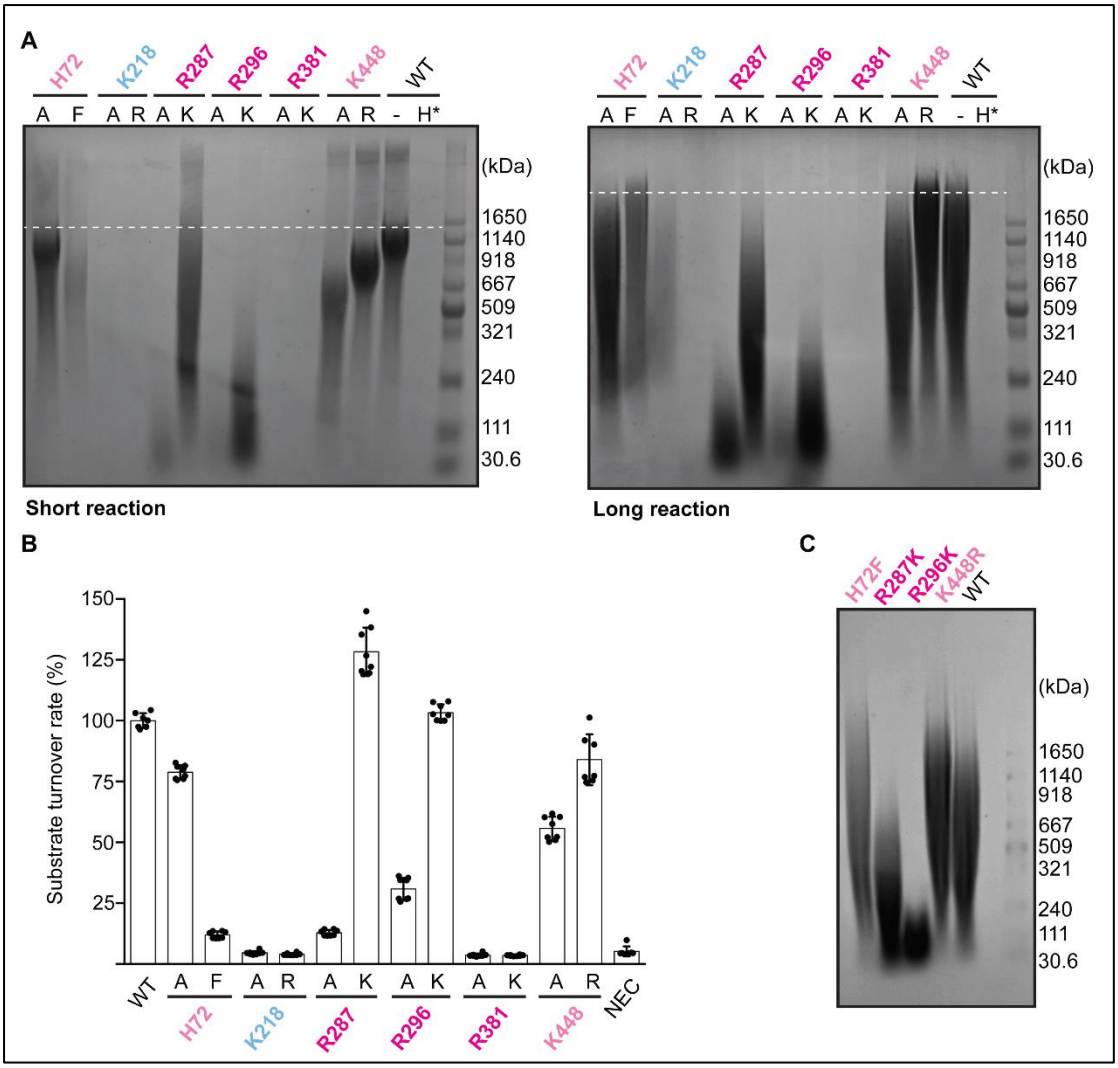


Figure 25 Mutagenesis of active site- and channel-lining residues

A – In vitro HA biosynthesis by the indicated XlHAS1 mutants for 1 h or 8 h (short and long reactions, respectively). WT – wild type, H* - hyaluronidase-treated samples. The dashed line roughly indicates the WT's highest molecular weight product. B – Catalytic rates of the mutants shown in panel A. Rates, normalized to WT, report substrate turnovers and were determined by quantifying UDP release in real time during HA synthesis. NEC – No Enzyme Control. All experiments were performed in eight replicates (n=8) and error bars represent standard deviations. C – Electrophoresis of HA species synthesized by selected channel-lining residue mutants, analyzed on a lower percentage agarose gel (0.5% instead of 1%) for better visualization of HMW HA species.

3. Discussion

3.1 HA size regulation

HA, cellulose and chitin synthases are multitasking enzymes that synthesize high molecular weight extracellular polysaccharides. Mechanisms by which those enzymes control product lengths and thus the polymers' physical properties are largely unresolved. The narrow size distribution of the XlHAS1-synthesized HA underscores high processivity of biosynthesis. XlHAS1's product size distribution broadens significantly during prolonged in vitro synthesis reactions due to substrate depletion and product inhibition. Further, substrate replenishing experiments show that HAS can release the HA during extended idle time, suggesting that, in vitro, HA release can be mediated by spontaneous diffusion of HA out of the secretion pore. In vivo, this could be modulated by various HA-interacting proteins present in the ECM, including HA-degrading hyaluronidases. Accordingly, throughout the human body, HA size is tissue-specific, with skin and synovial fluid HA typically reaching 5-6 MDa, whereas vitreous, amniotic fluid, lymph fluid and milk HA usually not exceeding 0.5 MDa. HA size distribution can also shift under oxidative stress, such as osteoarthritis-induced inflammation, which causes HA in synovial fluid to be degraded below 0.5 MDa species [20]. Additionally, recent analyses of HA biosynthesis in subterranean mammals pinpointed HAS expression levels and hyaluronidase activities as important determinants of HA size [71].

As shown by mutagenesis studies, processive HA biosynthesis requires a finely tuned network of HA-HAS interactions. Accordingly, modulating HA association inside HAS' translocation channel profoundly affects the HA length distribution. Two Arg to Lys substitutions within XlHAS1's IFH1 (R287K and R296K) abolish length control and lead to increased rates of HA diffusion out of the TM pore. These residues at the entrance to the TM channel are critical in stabilizing the nascent chain for elongation. Positive charges at these positions could be implicated in siphoning the nascent HA into the pore, as Ala mutations (R287A and R296A) show even greater reduction in substrate turnover and HA synthesis. On the other hand, replacing K448, about halfway across the translocation channel, with Arg leads to larger HA lengths, likely due to enhanced HAS-HA interactions that prevent robust transport and, consequently, reduce the enzyme's catalytic rates.

Notably, mutations reducing XlHAS1's catalytic rate while not affecting the HA secretion give rise to higher molecular weight HA. This underlines how important the high substrate availability is for maintaining constant HA synthesis and formation of HMW HA species. In vivo, similar effects may be achieved by limiting substrate availability and post-translational modifications of the enzyme [69, 70].

Collectively, the observed product size distributions synthesized by the ensemble of XlHAS1 mutants reflect the HA size range produced by vertebrate HAS isoenzymes [20, 72]. Although the residues analyzed in this study are conserved across HAS isoforms, disparities in channel dynamics and thus HA coordination may differ between isoforms (see future directions). Thus, physiological differences in HA size likely arise from variations in isoform-specific HA coordination, tissue-specific composition of HA-interacting ECM components including hyaluronidases, as well as substrate availability under specific metabolic states of cells and tissues, as previously discussed [73].

3.2 Insights into HA transport

The shape and electropositive character of XlHAS1's TM channel (Figure 19C) contrast the flat, acidic channel formed by cellulose synthase [15, 58]. Because the channel's gate is near the extracellular water–lipid interface, I estimate channel opening is induced by HA polymers exceeding 3-4 glycosyl units. In this case, the nascent chain and the channel's central hydrophobic ring likely prevent water flux across the membrane, similar to other polysaccharide secretion systems [74, 75].

As shown by the cryoEM data obtained for CvHAS bound to GlcA-terminated HA disaccharide [64], the HA register in the channel is most likely fixed. The polymer does not move through the channel when GlcA is incorporated. Instead, the newly introduced GlcA protrudes into the catalytic pocket and translocation only takes place upon GlcNAc addition. This way, the unfavorable conformation of GlcA stacking against the conserved acceptor site is extremely short-lived. Consequently, the polymer needs to move by two sugar units upon GlcNAc incorporation. A possible driving force may result from UDP-GlcNAc binding and can likely be assisted by the insertion of the gating loop.

Considering the narrow entry into the HA translocation channel, HA must enter the channel in a flat, ribbon-like conformation. Past the first three glycosyl units, the channel dimensions widen sufficiently to accommodate structural rearrangements of HA, allowing its relaxation. The observed rotation of the nascent HA chain inside the translocation channel may generate energetically favorable conformations contributing to HA translocation and preventing backsliding.

4. Preliminary data for future directions

4.1 Express all HAS isoforms for in vitro characterization

The observed product size distributions synthesized by the ensemble of XlHAS1 mutants reflect the HA size range produced by vertebrate HAS isoenzymes [20, 72]. Although the residues analyzed in this study are conserved across HAS isoforms (Figure S1), disparities in channel dynamics and thus HA coordination may differ between isoforms. Thus, physiological differences in HA size likely arise from variations in substrate availability, HA coordination, and/or metabolic states of the expressing cells and tissues, as previously suggested [73]. It is a common belief that one of the layers though which vertebrates regulate or adjust HA size is by expressing different HAS isoforms. HAS2 is believed to be most uniformly expressed and constitutes the main producer of HMW HA. Conflicting results are reported for HAS1, whereas HAS3 is thought to primarily synthesize LMW HA at different developmental stages and tissues [8, 10, 19, 20, 41, 46]. It could be highly informative to express and purify all three HAS isoforms, ideally from the same organism and characterize HA size distribution and substrate turnover rates for each. This would allow us to gain unprecedented insights into HA size distribution and kinetics for each HAS isoform in a chemically defined environment.

Since obtaining stable vertebrate HAS homologs from Sf9 insect cells failed for the most part, an alternative expression host is needed. The most common expression system for the large-scale mammalian protein expression are suspension-grown human embryonic kidney cells (HEK293S). The pEG BacMam vector [25] allows for a robust protein overexpression in mammalian cells (Figure 26A, B). The generated plasmid can be used for transfecting HEK cells using reagents such as Lipofectamine 2000. The downside of this approach is that it requires large quantities of plasmid DNA (up to milligrams for large cultures) prepared in aseptic conditions. An alternative approach is to generate a baculovirus and use it for introducing recombinant protein DNA into the HEK cells. To this end, a procedure for generating baculoviruses in Sf9 insect cells (as described above) can be used. Additionally, the pEG BacMam vector has a cloning site for insect-cell based expression, making it quite versatile [25] (Figure 26B).

Motivated by promising results for XlHAS1 but still wanting to obtain human homologs, I cloned all HAS isoforms from *Homo sapiens* and *Xenopus laevis* as GFP-fusion proteins into insect cell- and mammalian cell-based expression vectors. GFP is an extremely useful tag for 3 reasons. (1) Fusing the target protein with GFP typically boosts its expression and stability [24, 25, 78, 79]. (2) GFP fluorescence allows to quickly and easily visualize the target protein of interest within the cells using confocal microscopy. Additionally, GFP

remains fluorescent in SDS-PAGE gels, which allows for rapid detection of target protein in raw membrane fractions without the need for Western Blotting. (3) A high affinity nanobody binder is available against GFP, which can be used to generate a home-made high affinity column. Nanobodies, much like Fabs, can be expressed in *E. coli* in large quantities, often hundreds of milligrams, and immobilized on amine-reactive purification resins, such as NHS-activated Sepharose. The anti-GFP nanobody column allows for much tighter and specific binding than IMAC and hence even poorly expressed proteins can be purified to homogeneity. The caveat is that the binding is so strong that the only way to release the protein without denaturation is to remove the GFP tag using a specific protease, such as TEV protease. The obtained protein is tag-free, which makes it well-suited for downstream analyses.

Based on the anti-GFP purification approach, I devised a plan to comprehensively and relatively quickly compare insect- and mammalian cell-based expression systems for human and *Xenopus* HAS isoforms. The protocol involves lysing the cells in the presence of LMNG/CHS and purifying the HAS-GFP fusions using an anti-GFP column in GDN detergent (Figure 26C). The anti-GFP column has limited binding capacity and typically a lot of protein is lost in the unbound fraction. Yet, more protein can be recovered by subjecting the flowthrough to IMAC purification, which also constitutes a stability/expression test, as only well-expressed and stable protein will bind well to NiNTA (Figure 26C). Both IMAC- and AntiGFP-purified proteins can then be subjected to SEC and activity assays to further assess their stability and catalytic activity. This scheme is high throughput-oriented and omits a lot of optimization steps based on 3 assumptions. (1) Purification buffer conditions are well-optimized for XlHAS1 and involve the mildest possible detergent combinations, which have the highest chance of being suitable for other isoforms as well. Any homolog that does not survive these conditions is likely not a good candidate for robust structural biology analyses. (2) High affinity of the anti-GFP column allows for purification of even poorly expressed proteins. TEV protease digestion is an additional readout for protein stability, as only properly folded, non-aggregated HAS-GFP fusions can be digested efficiently. If no protein can be obtained using this approach, this again suggests that the HAS homolog is not suitable for in vitro work. (3) Only properly folded HAS-GFP fusion should manifest green fluorescence. Tracking the GFP fluorescence in all purification fractions (using in-gel fluorescence and visual inspection), allows for quick assessment of the target protein presence, concentration, degradation and stability.

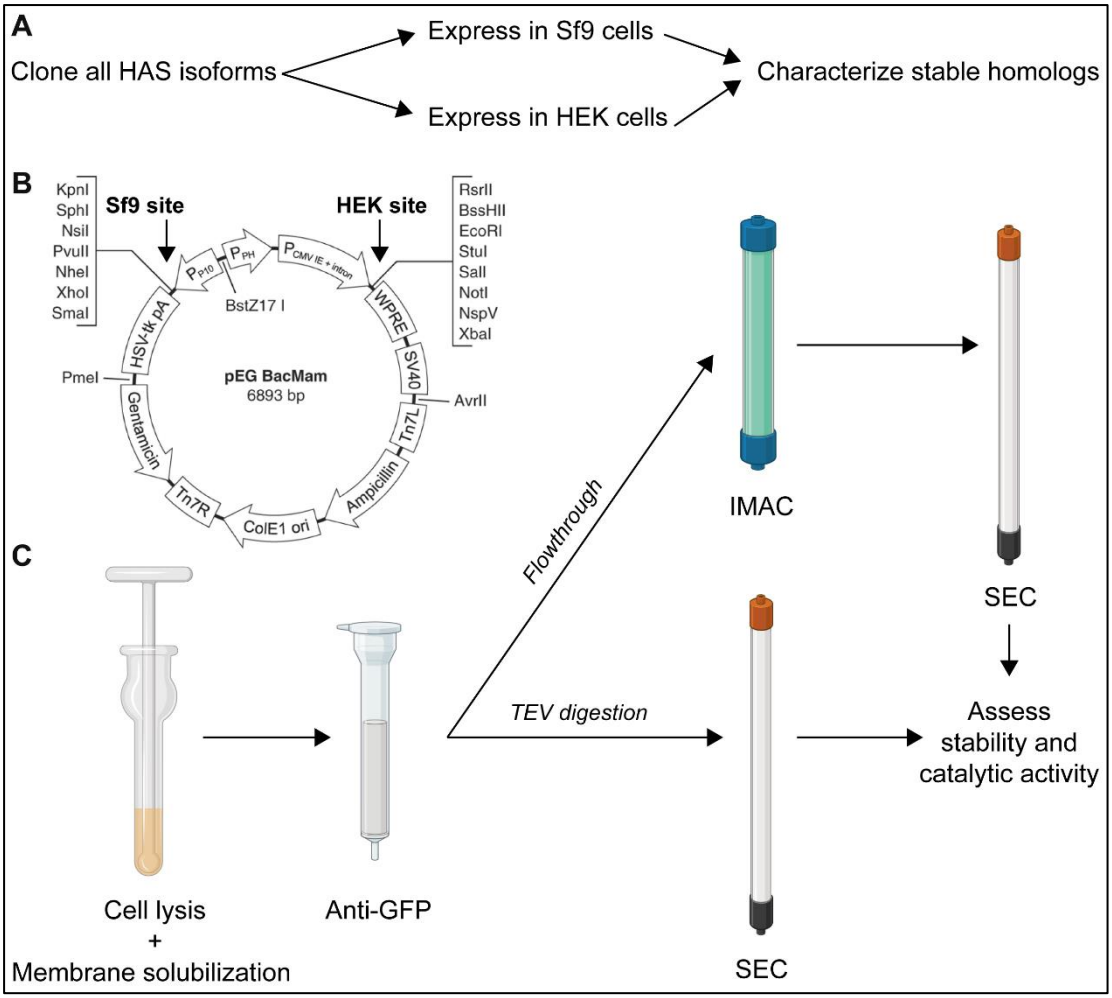


Figure 26 HAS isoform characterization workflow

A – Plan for expression of all HAS isoforms from humans and *Xenopus laevis*. All genes were cloned into mammalian and insect cell expression vectors as GFP fusions. Initial expression was assessed in Sf9 insect cells and purification attempts were performed for the most promising candidates. All unstable HASs can be subsequently expressed in HEK cells. B – pEG BacMam plasmid map showing the most relevant features. The plasmid has two cloning sites, one for insect cell expression and one for mammalian cell expression. Both cloning sites are flanked by appropriate promoter and terminator regions. Adapted from [25]. C – Purification scheme for characterizing HAS isoforms. Solubilized membranes are first subjected to purification using anti-GFP resin. The unbound fraction is used for standard IMAC purification (as described for XlHAS1), while bound HAS is released via TEV digestion. Protein fractions obtained from both columns are subsequently subjected to SEC for additional cleanup and assessing their homogeneity. Downstream biochemical analyses include HA gel and UDP release assays. The figure was prepared in part with BioRender.

While efforts to establish mammalian cell culture workstation in our lab are ongoing, I re-assessed expression of some of the vertebrate HAS homologs in a form of GFP fusions in insect cells, including the previously skipped HsHAS3. Although most of these attempts to re-express and re-purify all three HAS isoforms from human and *Xenopus laevis* failed, HsHAS3 and to some extent XlHAS3 showed some promise. Purification of HsHAS3 as a C-terminal GFP fusion using the anti-GFP column described above yielded a stable and catalytically active preparation (Figure 27A, B). Notably, the protein also showed good purity upon IMAC purification (Figure 27A), which makes it an

ideal candidate for subsequent structural and biochemical analyses. XlHAS3 purification was partially successful. The obtained protein eluted mostly in the void volume of the SEC column, did not manifest substantial fluorescence, could not be digested with TEV protease, and was heavily contaminated upon IMAC purification (Figure 27C, D).

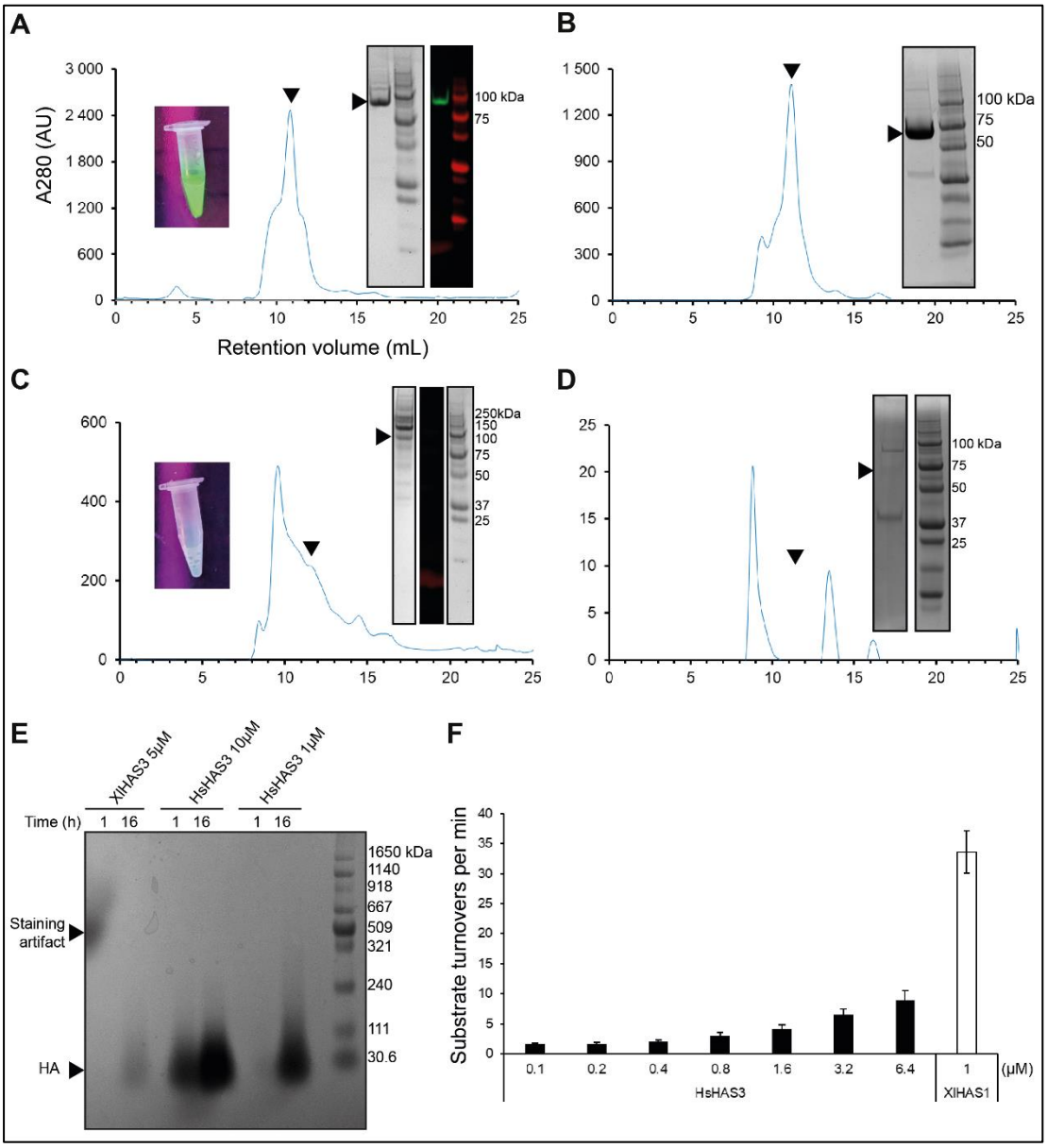


Figure 27 HAS3 purification attempts

A – SEC profile of HsHAS3 still fused to GFP after IMAC purification. Inset shows fluorescent fusion protein in the test tube under UV lamp. Gel inset shows Coomassie staining (left gel slice) and in-gel fluorescence (right slice). Coomassie staining shows high purity of the preparation and in-gel fluorescence indicates good stability. B – SEC profile of HsHAS3 after on-column TEV protease cleavage. Gel inset shows Coomassie staining revealing high purity and TEV protease digestion efficiency as the protein migrates as expected at ~63 kDa. C – SEC profile of XlHAS3 fused to GFP after IMAC purification. Inset shows poorly fluorescent fusion protein in the test tube under UV lamp. Gel inset shows Coomassie staining (left gel slice) and in-gel fluorescence (right slice). Coomassie staining shows poor purity of the preparation and in-gel fluorescence indicates lack of stability. D – SEC profile of XlHAS3 after on-column TEV protease cleavage. Gel inset shows Coomassie staining revealing poor TEV protease digestion efficiency as there is no protein band at the expected ~63 kDa. E – HA gel

assay showing activity of the purified HAS3 homologs. XlHAS3 was used at 5 μM, while HsHAS3 was used at 1 and 10 μM for the assay. HA synthesis reactions were incubated for 1 or 16 hours. The HA bands after electrophoresis reveal unusually LMW HA species not exceeding ~100 kDa for both HAS3 homologs. F – UDP release assay showing greatly reduced catalytic rates for HsHAS3 compared to XlHAS1. HsHAS3 is about 6 times slower than XlHAS1 and had to be used in a relatively high concentration range to obtain reaction rates well above background. The experiments were performed in triplicates (n=3) and error bars represent standard deviations.

Based on agarose gel analyses of in vitro synthesized HA, HsHAS3 makes only LMW HA that is less than 100 kDa (Figure 27E). These species can be reproducibly obtained for both GFP-fused and tag-less HsHAS3 in various protein to substrate ratios and incubation times. Importantly, partially successful XlHAS3-GFP preparation also makes HA within this size range, which excludes the possibility if in vitro HA synthesis conditions being suboptimal. Notably, HsHAS3 is much slower at substrate hydrolysis compared to XlHAS1 (Figure 27F). Since slower XlHAS1 mutants usually make HA of larger size than WT (Figure 24E), this suggests that HAS3s make LMW HA by limiting interactions between the HAS channel and the HA polymer, much like XlHAS1 R287/R296 mutants (Figure 25C). Most HA-interacting residues are conserved across all vertebrate HAS isoforms, including the crucial R287 and R296 (Figure S1). This leads me to hypothesize that the synthesis of varying lengths of HA by different HAS isoforms arises from differences in HA coordination by moderately conserved residues and HA channel dynamics. To delineate these nuanced differences, a high resolution structure of each HAS isoform-HA intermediate would be highly informative.

Vertebrate HASs are about 60 kDa membrane proteins, which makes them quite difficult targets for cryoEM. Although similarly small proteins can be solved using cryoEM, those typically require substantial optimization of sample preparation and grid vitrification procedures. The Fab discovery approach is time-consuming and does not always guarantee success, as many Fabs found via phage-display screening are false-positives, as shown by my Fab assessment trials (Figure 13A-C). The anti-XlHAS1 Fab15 was one of 11 Fabs and binds to a region of the GT domain with low sequence conservation (Figure 28A). Hence, it is very unlikely that Fab15 is a universal anti-HAS binder. Nonetheless I assessed Fab15 binding to HsHAS3 using SEC coelution experiment and found no evidence of binding (Figure 28B).

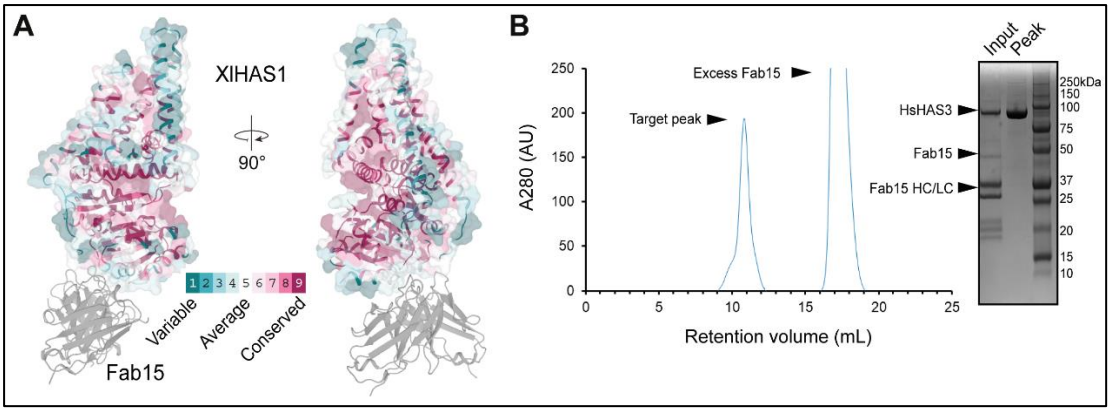


Figure 28 HAS-Fab interaction

A – XlHAS1-Fab15 model. Fab15 is show as a grey cartoon and XlHAS1 is shown as a surface colored according to residues conservation. The conservation map was calculated using the ConSurf server [80]. B – Results of HsHAS3 and Fab15 co-elution experiments - SEC co-elution chromatogram (left panel) and the corresponding Coomassie-stained SDS-PAGE gel (right panel).

An alternative approach would be to introduce an epitope that could be recognized by an existing Fab or a nanobody. Apocytochrome B562 (BRIL) is a 4-helix bundle (Figure 29A) developed as a fusion construct, initially to aid crystallization of various G-protein coupled receptors (GPCRs). Inserting the BRIL sequence into extracellular loops facilitated successful crystallization of several GPCRs [81]. BRIL also has a good track record in cryoEM experiments as it contributes 15 kDa of potentially rigid protein mass, provided a proper linker design. Additionally, an anti-BRIL Fab is available, which would add another 45 kDa to the total mass [81]. Hence, BRIL constitutes a universal marker that can be utilized to substantially increase the protein size for cryoEM analyses, provided that it does not interfere with protein folding. Alphafold2 [52] prediction of HsHAS3 hints at two loops within HsHAS3's cytoplasmic GT domain that appear to be flexible and distant from the active site that should be suitable insertion points for the BRIL sequence (Figure 29B). Insertion spot 1 is a loop that connects an α-helix at the bottom of the GT domain to its first β-strand. This loop appears to be flexible and is unusually long and disordered in vertebrate HAS1s, suggesting that it is not crucial for the stability of the GT domain (Figure 29B). Insertion spot 2 is a long loop that connects TMH2 to the GT domain and also appears quite flexible (Figure 29B). BRIL insertion into one of the 3 extracellular loops would be another possibility, yet this could affect the intrinsic dynamics of the HA channel and have major implications on HA size and/or HAS destabilization.

I made HsHAS3-GFP constructs with a single BRIL domain inserted into one of the 2 insertion spots using Gibson Assembly, a cloning technique that does not rely on the presence of restriction sites [82]. Both constructs appear to express well, based on Western Blot analyses of the Sf9 cell lysates (Figure 29C). Imaging in-gel fluorescence suggests that the protein is properly folded, yet purification trials need to be performed to further assess the stability and

the catalytic activity of these constructs. CryoEM sample preparation will involve generating an HA-bound HsHAS3 and attaching the anti-BRIL Fab to this construct (as described for XlHAS1, Figure 17A). The GFP present at the C-terminus of HsHAS3 can also aid cryoEM analyses, especially when attached to anti-GFP nanobody (Figure 29D). In the case that BRIL insertion destabilizes HsHAS3, a simple HsHAS3-GFP fusion with anti-GFP nanobody could be utilized as a fallback platform, provided the GFP adduct is at least somewhat rigid (Figure 29E).

If this approach is successful, it could be a suitable platform for high resolution cryoEM analyses of HsHAS3, as well as other vertebrate HASs successfully expressed and purified in the future.

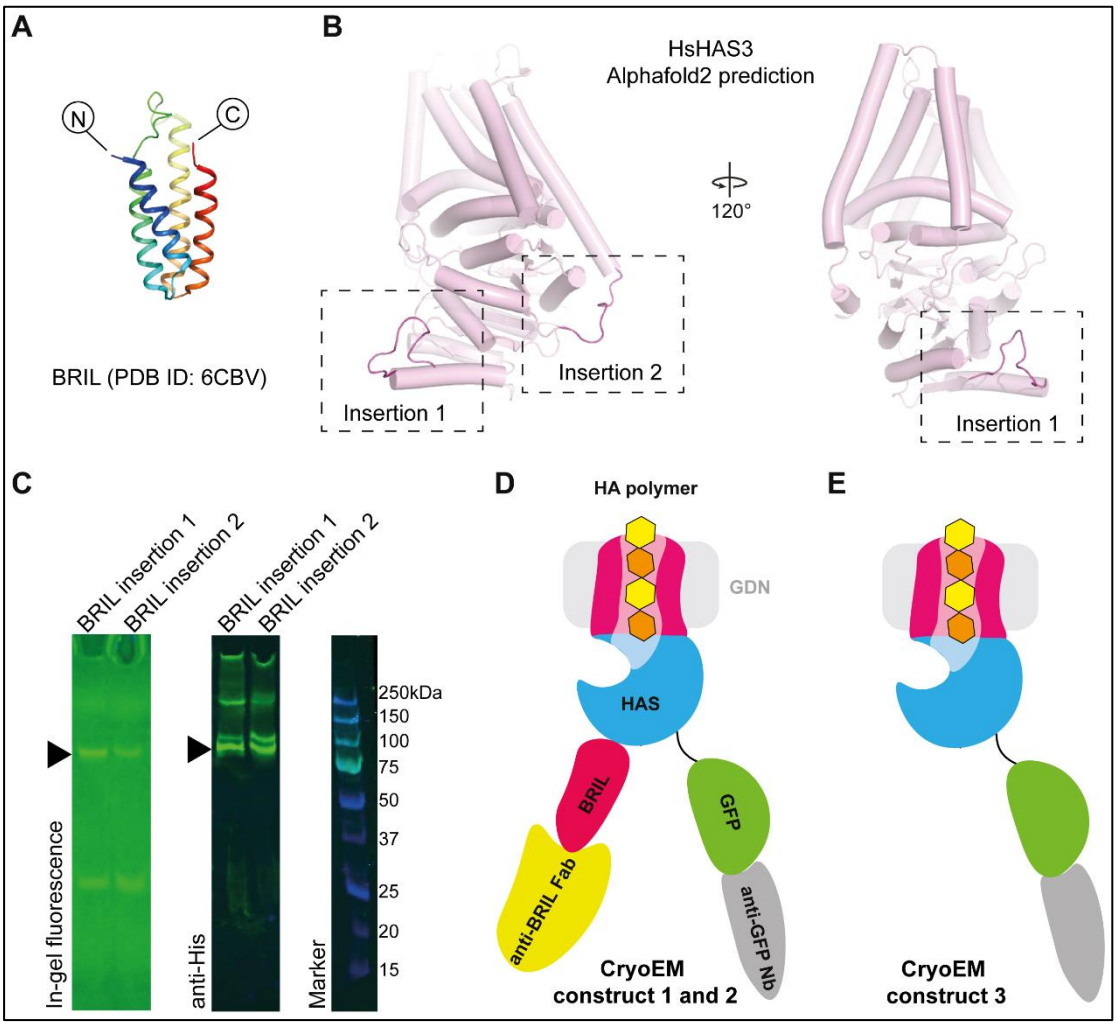


Figure 29 HsHAS3 cryoEM strategy

A – Structure of BRIL. B – Alphafold2 prediction of HsHAS3 showing loops that were selected as BRIL insertion spots. C – Result of expression tests of HsHAS3-GFP with BRIL insertions. In-gel fluorescence suggests proper folding of the target protein-GFP fusion (left), while anti-His Western Blot show expression of the proper length product (right). D – Schematic representation of HsHAS3-GFP BRIL insertion constructs (1 and 2 as two different BRIL insertions were generated) conjugated with anti-BRIL Fab and anti-GFP nanobody for cryoEM analyses. E – Schematic representation of the backup HsHAS3 construct only fused to GFP and conjugated with anti-GFP nanobody.

4.2 Characterize the contribution of each isoform to HA formation

Cell-based data from the Sf9 insect cells suggest that at least some HAS homologs are active in the expression hosts. Sf9 cells infected with baculoviruses harboring HAS-GFP genes show significant green fluorescence, easily detectable even by eye. Confocal microscopy analyses of those cells show a bright, green outline suggesting HAS expression in the cytoplasmic membrane (Figure 30A). However, this data needs to be interpreted with caution as the Sf9 cell morphology is disturbed upon viral infection, including swelling of the nuclei forcing all membranous organelles to the periphery of the cell. Interestingly, incubating those cells with HA-specific carbohydrate-binding module (CBM70) fused to mCherry reveals formation of HA coat on the surface of Sf9 cells expressing HsHAS2-GFP but not GFP-XlHAS1 (Figure 30B). If the cells are pre-treated with hyaluronidase, the mCherry signal disappears from the surface of HsHAS2-GFP-expressing cells (Figure 30B). It is unclear why a robust XlHAS1 homolog does not generate an HA coat on the cell surface. This could be due to its high demand for substrates [46]. Additionally, the green signal coming from GFP-XlHAS1 is more diffused compared to HsHAS2-GFP expressing cells (Figure 30A), suggesting possible issues with trafficking of XlHAS1 to the plasma membrane. This data correlates well with the results of HA ELISA showing high HA signal for Sf9 cells expressing HsHAS2 fused to either GFP or a His-tag alone, as well as *E. coli* BL21 cells expressing bacterial SeHAS, but not for non-infected Sf9 cells (Figure 30C).

This data could be further complemented by HA extraction followed by electrophoresis, as ELISA only reports on the quantity of HA with no information on its size [40, 83]. Similar approaches were recently utilized to characterize HAS2 from naked mole rats (NMRs). NMRs are famous for their longevity (exceeding 30 years) and resistance to cancers, both of which are attributed to formation of high amounts of unusually HMW HA [12]. Using transient transfection of nmrHAS2, followed by HA purification from cells and conditioned media, revealed a substantial increase in HA synthesized in HeLa cells by nmrHAS2, compared to mouse HAS2 [84].

Possible trafficking issues, lack of stability, and an evolutionarily distant host questions the validity of observations drawn from insect cell imaging (Figure 30A, B). With a well-established HEK cell expression system for HASs, it would be tempting to characterize HASs using similar cell-based assays. Subsequently, it would be worth to repeat these experiments in a genetically clean background where individual and combinations of HASs are removed by CRISPR-Cas9-mediated knock-outs.

Additionally, utilizing MINFLUX, a fluorescent nanoscopy technique capable of reaching 1 nm localization precision of fluorophores [85] could provide insights into the density and thickness of the HA coat on the cell surface

at unprecedented level of detail. To this end, a modified version of the CBM70 fused to the SNAP domain can be used. SNAP domain allows to load the CBM probe with blinking fluorophores required for MINFLUX, such as Alexa647 [74].

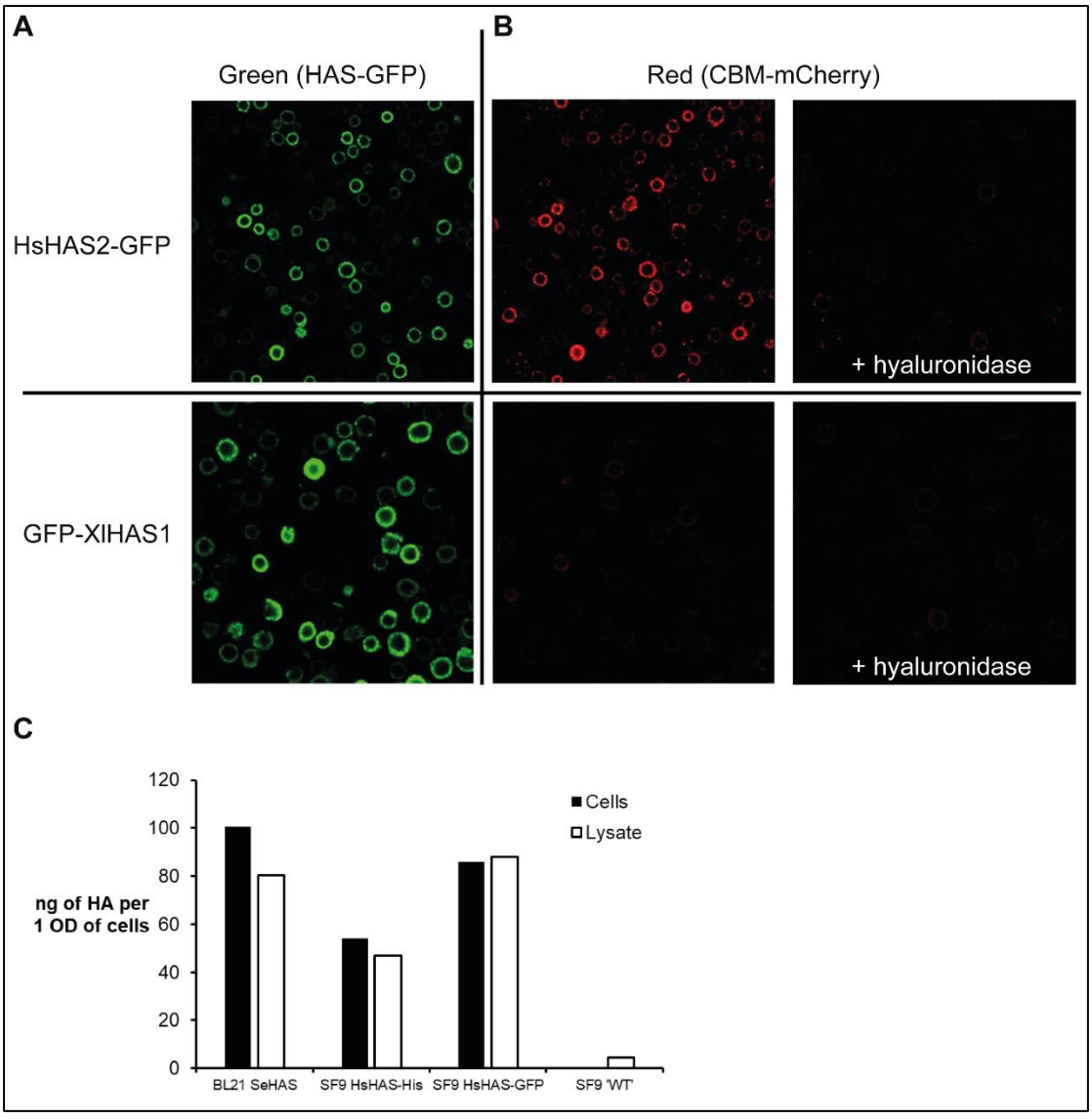


Figure 30 HAS cell-based characterization

Confocal micrographs showing Sf9 cells expressing HsHAS2-GFP and GFP-XlHAS1 (green signal comes from GFP) labelled with HA-specific CBM70 fused to mCherry (red signal). Panel A – Green channel. B left – Red channel of the same cells. B right – Red signal coming from cells pre-treated with hyaluronidase. C – Results of HA enzyme-linked immunosorbent assay (ELISA) showing HA accumulation in whole cell- and cell lysate-containing samples of Sf9 cells expressing various HsHAS2 constructs, uninfected Sf9 cells, as well as *E. coli* BL21 cells expressing *Streptococcus equisimilis* HAS (SeHAS).

4.3 Determine the role of ECM in controlling HA size

Agarose gel analyses of HA synthesized in vitro by XlHAS1 show that the size of HA appears to be more homogenous and relatively small when intact Sf9 cell-derived inverted membrane vesicles (IMVs) are used for the assay. If, however, the IMVs are pre-treated with detergent, the synthesized HA appears

to be much larger (Figure 31A). Hence, I speculate that one of the key regulatory points for HA size is the space in which the nascent HA polymer is secreted into. This could vary depending on the crowding of the ECM.

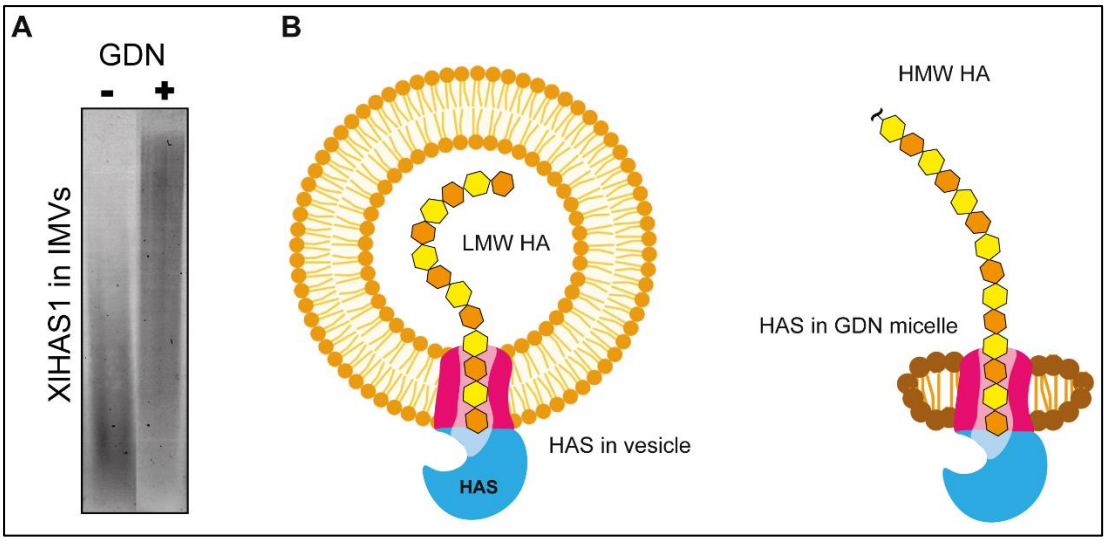


Figure 31 Influence of HA packing on HA secretion and size

A – HA size distribution upon synthesis by XlHAS1-containing IMVs (left lane) and XlHAS1 IMVs pre-treated with the GDN detergent. B – Possible explanation of lower molecular weight synthesis by XlHAS1 in IMVs – limiting vesicle lumen, compared to GDN-solubilized HAS, which can make unrestricted HA. Similar effects may occur in tight, dense spaces of the extracellular matrix. The figure was prepared in part with Biorender.

HA can reach up to 8 MDa in size [8] and it retains several times its mass in water to form a gel-like substance. In addition, the polymer is highly negatively charged [86]. I hypothesize that the high molecular weight and negative charge may be the reason why IMVs show accumulation of shorter HA species, compared to detergent-solubilized samples (Figure 31B). Similarly, in the physiological setting, the physical ECM volume may be one of the limiting factors that leads to stalling of HA secretion into the ECM and, consequently, HA biosynthesis. In addition, the negative charge of the polymer likely requires a presence of positively charged counter ions to facilitate a robust HA deposition in the ECM.

To test these points, these conditions can be simulated in vitro using proteoliposome-reconstituted XlHAS1. By generating XlHAS1-containing liposomes of specific sizes, it will be possible to assess if and how exactly the volume of the ECM limits HA size and if it leads to stalling of HA synthesis. Small vesicles should limit the HA size and stall the enzyme, even after replenishing the substrates (as shown on Figure 9). Additionally, with such a system, the effect of other factors may also be analyzed, for example the presence of other GAGs or HA as well as ions counteracting the negative charge of HA. I expect that adding excessive amounts of positively charged ions, such as Mg^{2+} and Ca^{2+}, as well as ionophores should counteract unfavorable energetics of loading the vesicles with negatively charged HA. On

the other hand, saturating the vesicles with exogenous HA should inhibit the HA synthesis by HAS.

The crucial part of this aim would be establishing a proteoliposome reconstitution protocol for XlHAS1. My preliminary trials show that removal of the GDN detergent (the only detergent I found so far to guarantee good stability of XlHAS1 during purification) is difficult using standard techniques, such as sorbent media (BioBeads) or dialysis, most likely due to its extremely low CMC, as discussed above. Additionally, XlHAS1 appears to lose stability and activity upon partial detergent removal. This could be due to suboptimal lipids used during initial reconstitution trials. Hence it will be important to find an optimal lipid extract suitable for XlHAS1, as well as to explore alternative protocols of incorporating the target protein into liposomes, for example by using pre-formed vesicles in the presence of minimal concentrations of detergents [31]. A good quality vesicle mixture can then be fractionated based on size using gradient centrifugation or extrusion, as shown for the Piezo1 channel, which yielded highly homogenous PL preparations suitable for cryoEM single particle analyses [87].

5. Methods

5.1 Molecular cloning

Genes encoding vertebrate HASs were cloned into the pACEBac-1 vector using BamHI, SacI and HindIII restriction sites. Most of the XlHAS1 mutants were generated using QuickChange mutagenesis method [84]. Some variants were difficult to obtain due to issues with PCR reaction failing or generating off-site mutations. As a remedy, an alternative method with non-overlapping primers was used according to manufacturer's (New England Biolabs) protocol [88, 89]. Cloning was confirmed by restriction analyses and DNA sequencing. Baculoviruses harboring each target gene were prepared as described [58]. Briefly, the XlHAS1 containing pACEBac-1 plasmid was transformed into *E. coli* DH10MultiBac cells. Bacmids for the WT and all XlHAS1 mutants were purified from 3 white colonies and transfected into Sf9 cells at $1x10^6$ cells/mL using the FuGene reagent (Promega). Cells were maintained in ESF921 medium (Expression Systems) at 27 ºC with mild shaking. The baculovirus was amplified to generate P2 virus stock, which was used at 1.5% culture volume to infect Sf9 cells at a density of $3x10^6$ cells/mL. Cells grown in 1 L media bottles (0.45 L per bottle) were pelleted by centrifugation 48 h post-infection and resuspended in 40 mL of Buffer A (40 mM Tris-HCl pH 7.8, 150 mM NaCl, 10 % glycerol, 5 mM $MgCl_2$) per 0.9 L of cell culture. The harvested cells were flash-frozen and stored at -80 °C for subsequent use.

5.2 Inverted membrane vesicle preparation

To prepare Sf9 cell-derived inverted membrane vesicles (IMVs), 0.45 L culture-worth of cells was resuspended in Buffer A supplemented with 5 mM β-mercaptoethanol (β-ME) and 1 mM phenylmethylsulphonyl fluoride (PMSF). Cells were lysed by three passes through a microfluidizer at 18,000 psi. Unbroken cells were pelleted by centrifugation at 20,000 x g for 25 minutes. Cleared lysate was carefully layered onto 40 mL of 2 M sucrose, followed by ultracentrifugation at 200,000 x g for 2 hours. The brown IMV ring was collected, diluted to 60 mL using Buffer A and subjected to another round of ultracentrifugation for 1 hour. The pelleted IMVs were resuspended in 2 mL of Buffer A, aliquoted, flash-frozen, and stored at -80 °C for subsequent use in activity assays.

5.3 HAS purification

All preparation steps were carried out at 4 °C unless stated otherwise. To purify vertebrate HAS homologs, including XlHAS1, typically 0.9 L-culture worth of cell suspension was thawed and diluted to 200 mL using Buffer A supplemented with 5 mM β-ME, 1 mM PMSF, 10 mM imidazole, 1% DDM and 0.2% CHS. Cells were lysed using a tissue homogenizer and rocked for one

hour at 4 °C, followed by ultracentrifugation at 200,000 x g for 30 minutes. The cleared lysate was mixed with 10 mL of 50% NiNTA (ThermoFisher) resin suspension equilibrated in Buffer A and subjected to batch binding for an hour. After that, the slurry was poured into a glass gravity flow column (Kimble), the flow-through was discarded, and the resin was washed three times with 50 mL of Buffer A supplemented with 0.03% GDN (wash 1) and 1 M NaCl (wash 2) or 20 mM imidazole (wash 3). XlHAS1 was eluted using Elution Buffer (EB) consisting of 25 mM Tris-HCl pH 7.8, 150 mM NaCl, 10% glycerol, 350 mM imidazole, 0.03% GDN in two steps. First, 15 mL of EB was added, followed by ~ 5 min incubation, draining, and addition of another 15 mL of EB and draining. The eluted sample was concentrated using 50 kDa MWCO Amicon centrifugal concentrator (Millipore) to <1 mL for SEC using a Superdex200 column (GE healthcare) equilibrated in Gel Filtration Buffer (GFB) consisting of 20 mM Tris-HCl pH 7.8, 150 mM NaCl and 0.02% GDN. The target peak fractions were pooled and concentrated as necessary for subsequent experiments.

5.4 Quantification of HAS activity by scintillation counting

All reagents for biochemical analyses were supplied by Sigma, unless stated otherwise. Reaction buffer consisted of 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 20 mM $MgCl_2$, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP), 0.02% GDN (for GDN-purified protein samples), 5 mM UDP-GlcNAc, 5 mM UDP-GlcA and 0.01 µCi/µL [^{3}H]-UDP-GlcNAc (Perkin Elmer). Reactions were carried out at 37 °C for the indicated times, depending on the experiment. Typically, 1 µM of GDN-purified HAS was used. For assessing HASs in IMVs, IMVs were added at half of the reaction volume. HA digest controls were performed by adding 75 U hyaluronidase (MP Biotech) and 2% DDM (for reactions containing IMVs) to the reaction and incubating for an additional 10 minutes at 30 °C. Reactions were terminated with 3% SDS, and product accumulation was quantified using descending paper chromatography and liquid scintillation counting as previously described [90].

5.5 Electrophoretic HA size determination

To assess the size of in vitro synthesized HA, similar reaction conditions were applied to the ones described above, except that the radioactive tracer was omitted. Synthesis reactions were mixed with SDS-PAGE loading dye and applied to a 1% agarose gel (Ultra-pure agarose, Invitrogen) casted in an Owl B2 system (ThermoFisher). All gels were subjected to electrophoresis at 100 V for 2 hours at room temperature to achieve comparable separation for each run. After the run, the gel was equilibrated in 50% ethanol for 1.5 h and subjected to staining in 0.05% Stains-all (Sigma) in 50% ethanol overnight under light-protection. Post-staining background was reduced by soaking the gel in 20% ethanol for 3-7 days in the dark. The migration of the synthesized HA species

was compared to HA standards from *Streptococcus equisimilis* (Sigma), as well as enzymatically generated ladders: HA LoLadder and HA HiLadder (Hyalose).

5.6 Substrate turnover rate quantification

UDP release during HA synthesis was quantified using an enzyme-coupled assay as previously described [18, 28]. Depletion of NADH was monitored at 340 nm every 60 s for 3 hours at 37 °C in a SpectraMax instrument. The raw data was processed in MS Excel and a linear phase of each reaction was determined. The rate of NADH depletion was converted to µmoles of UDP released using a UDP standardized plot for subsequent Michaelis-Menten constant determination using GraphPad Prism. For substrate turnover rate of XlHAS1 mutants, the µmoles of released UDP were presented relative to wild type activity. All experiments were performed three to eight times and error bars represent the standard deviations.

5.7 Cell surface HA imaging and quantification

Sf9 cells expressing GFP-fused HAS were resuspended in 200 µL of PBS at OD=2 and 5 µL of bovine hyaluronidase (at 20 mg/mL, MP Bio) or 5 µL of water was added. Cells were incubated with rocking at 4°C for 1 h. Next, the cells were washed by centrifugation at 1000 x g for 5 minutes and resuspending the pellet in 200 µL of PBS twice. The sample was subjected to HA ELISA according to manufacturer's protocol (Corgenix) directly or after lysing the cells using a sonicator. The data was analyzed in Microsoft Excel and converted to ng of HA based on the standard curve prepared using the HA standards provided with the HA ELISA kit.

For cell imaging, 1 mg/mL of mCherry-fused CBM70 [74] was added to the hyaluronidase- or water-treated cells. Cells were incubated with rocking at 4°C for 1 h and washed as above. After the final wash, a 50 µL drop was directly imaged on a cover slide without fixation using Zeiss LSM880 confocal microscope with an Airyscan detector and water-immersion objective. Cells were imaged in bright field. GFP-fused HAS (green) and mCHerry-CBM70 HA probe (red) channels were recorded sequentially using 488 nm and 587 nm excitation lasers, respectively. The images were analyzed in ImageJ.

5.8 XlHAS1 reconstitution

For Fab selection, XlHAS1 was reconstituted into *E. coli* total lipid nanodiscs [22] using MSP1D1 chemically biotinylated according to the protocol supplied by the manufacturer of NHS-biotin (ThermoFisher). The purified enzyme was mixed with MSP and sodium cholate-solubilized *E. coli* total lipids at 40 µM final concentration in 1 mL final volume according to 1:4:80 molar ratio of HAS:MSP:lipids. Detergent removal was initiated 1 hour after mixing all

components by adding 200 mg of BioBeads (BioRad) and mixing at 4 °C. After an hour, another batch of BioBeads was added, followed by mixing overnight. The following morning (after ~12 hours), the reconstitution mixture was transferred to a fresh tube and the last batch of BioBeads was added, followed by mixing for one hour and SEC using a Superdex200 column equilibrated with GFB lacking detergent.

A similar protocol was applied for proteoliposome reconstitution trials. Purified HAS was mixed with sodium cholate-solubilized *E. coli* total lipids at 10 µM final concentration in 1 mL final volume according to 1:200 molar ratio of HAS:lipids. Detergents were removed by BioBeads (as described for the nanodisc sample) and the final sample was subjected to two rounds of centrifugation. Firstly, the mixture was spun down at 10,000 x g to pellet precipitated protein, followed by ultracentrifugation at 250,000 x g to pellet fully-formed proteoliposomes. Proteoliposome pellet was resuspended in 1 mL of detergent-free GFB using tissue homogenizer.

5.9 Identification of XlHAS1-specific Fabs

Phage selection was performed as previously described [91, 92]. In the first round of selection, 400 nM of XlHAS1-loaded nanodiscs diluted in the selection buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 1% BSA) was immobilized on streptavidin paramagnetic beads (Promega). Beads were washed three times in the selection buffer with 5 mM d-desthiobiotin added during the first wash to block nonspecific binding. Fab phage library E [50] resuspended in selection buffer was added to the beads and incubated for 1 hour with gentle shaking. The beads were washed three times in the selection buffer and then transferred to log-phase *E. coli* XL1-Blue cells. Phages were amplified overnight in 2xYT medium with ampicillin (100 µg/ml) and M13-KO7 helper phage (10^9 pfu/mL). Four additional rounds of selection were performed with decreasing target concentration (200 nM, 100 nM, 50 nM, 25 nM) using a KingFisher magnetic beads handler (ThermoFisher). In every subsequent round, the amplified phage pool from the previous round was used as the input. Prior to being used for selection, each phage pool was precleared by incubation with 100 µl of streptavidin magnetic beads. Additionally, 2 µM of non-biotinylated MSP1D1 nanodiscs were present in the selection buffer to reduce the presence of non-specific binders during rounds two to five. In these rounds, selection buffer supplemented with 1% Fos-choline-12 was used to release the target and bound phages from the nanodiscs. Cells infected after the last round were plated on LB agar with ampicillin (100 µg/ml) and phagemids from individual clones were sequenced at the University of Chicago Comprehensive Cancer Center Sequencing Facility to identify unique binders. Single-point phage ELISA was used to validate the specificity of unique binders as described previously [92]. Fabs were expressed and purified as described [92] and used

for activity assays and SEC co-elution experiments with XlHAS1. The strongest non-inhibitory binders were chosen for cryoEM trails and one of those yielded well-structured projections of the HAS-Fab complex.

5.10 XlHAS1 cryoEM sample preparation

GDN-solubilized XlHAS1 was used for cryoEM experiments. The purified enzyme was mixed with the Fab at a 1:4 molar ratio and incubated overnight at 4 °C, followed by SEC using a Superose 6 column (Cytiva) equilibrated with GFB2 containing 0.01% GDN.

Initial attempts to prepare cryoEM grids of the purified XlHAS1-Fab complex in the presence of substrates failed to capture a HAS–HA intermediate due to rapid HA accumulation in the sample hampering grid vitrification. To generate a HA-associated sample, substrates (UDP-GlcA and UDP-GlcNAc, 2.5 mM each, Sigma), 20 mM $MgCl_2$ and recombinant HA lyase (0.01 mg/mL, purified as described previously [93]) were included during the overnight Fab incubation, followed by SEC. Post-SEC, the sample was concentrated to 8 mg/mL using a 100 kDa MWCO Amicon ultrafiltration membrane. Attempts to generate a GlcNAc-terminated and HA- and UDP-bound sample involved addition of 2.5 mM UDP-GlcNAc and 2.5 mM $MgCl_2$ after SEC and incubation on ice for 1 hour prior to cryoEM grid vitrification. 4 µL of XlHAS1 sample was applied onto the C-flat 1.2/1.3 grid, and glow-discharged for 45 seconds in the presence of 1 drop of amylamine. Grids were blotted for 4 seconds at a blot force of 4, at 4 ºC and 100% humidity and plunge-frozen in liquid ethane using Vitrobot Mark IV (FEI). This sample yielded the UDP-bound and gating loop inserted XlHAS1 structure.

5.11 CryoEM data collection and processing

All cryoEM datasets were collected on a Titan Krios equipped with a K3/GIF detector (Gatan) at the Molecular Electron Microscopy Core (University of Virginia School of Medicine). Forty-frame movies were recorded in counting mode at 81,000x nominal magnification, -2.0 to -1.0 µm target defocus, and 50 $e^-/Å^2$ total dose.

All datasets were processed in cryoSPARC [53]. Raw movies were subjected to patch motion correction and patch contrast transfer function (CTF) estimation. Particles were automatically selected by template picker and sorted by iterative cycles of 2D classification and heterogeneous refinement. To separate HASs bound to their respective ligands, 3D variability and 3D classification approaches were used. The final volumes were refined using non-uniform and local refinements to generate high-resolution maps at 3.0–3.2 Å average resolution.

5.12 Model building

To generate the XlHAS1 apo model, the Alphafold2 [52] prediction was docked into the EM map using Chimera [94] and the model was iteratively real-space refined in Coot [55] and Phenix [56] to good statistics (Table S1). UDP- and HA-bound structures were generated using the XlHAS1 apo model as a starting point and likewise refined (Table S1).

Across all my models, I was able to model most of XlHAS1 residues with the exception of the flexible loop of the GT domain (residues 172-193), the N-terminal (residues 1-14) and C-terminal (residues 569-588) extensions.

Data availability

Coordinates and EM maps have been deposited at the Protein Data Bank and Electron Microscopy Data Bank under accession codes 8SMM/EMD-40591, 8SMN/EMD-40594 and 8SMP/EMD-40598 for XlHAS1 apo, XlHAS1 HA-bound and XlHAS1 UDP-bound, respectively.

Acknowledgments

I would like to thank the following people who contributed to this work and my overall well-being and sanity during my PhD journey.

- All members of the Zimmer lab, past and present. For great scientific and working environment, their willingness to help and exchange ideas.
- Jochen Zimmer for mentorship, conceiving the project and providing me both scientific freedom and great ideas.
- Zachery Stephens for great collaboration on the HAS project, many stimulating scientific discussions and joint navigation through the hardships of the PhD program.
- Finn Maloney, Jeremi Kuklewicz and Louis Wilson for critical comments on my manuscript.
- Kelly Dryden and Michael Purdy of the Molecular Electron Microscopy Core at the University of Virginia and Ruoya Ho for help with cryoEM data collection.
- Paul DeAngelis of the University of Oklahoma Health Sciences Center for advice on HA detection using Stans-All and providing samples of HA ladders.
- Boehringer Ingelheim Fonds for giving me the honor of becoming a BIF fellow, great support, networking and workshop opportunities.
- My best friends Kasia, Louis, Jeremi, Zuzanna, Bartek and Weronika for endless emotional support and our many road trips and adventures in the US and Poland.
- My parents and brothers for always having my back and for understanding my missing most of important family events.

Supplementary Figures and Tables

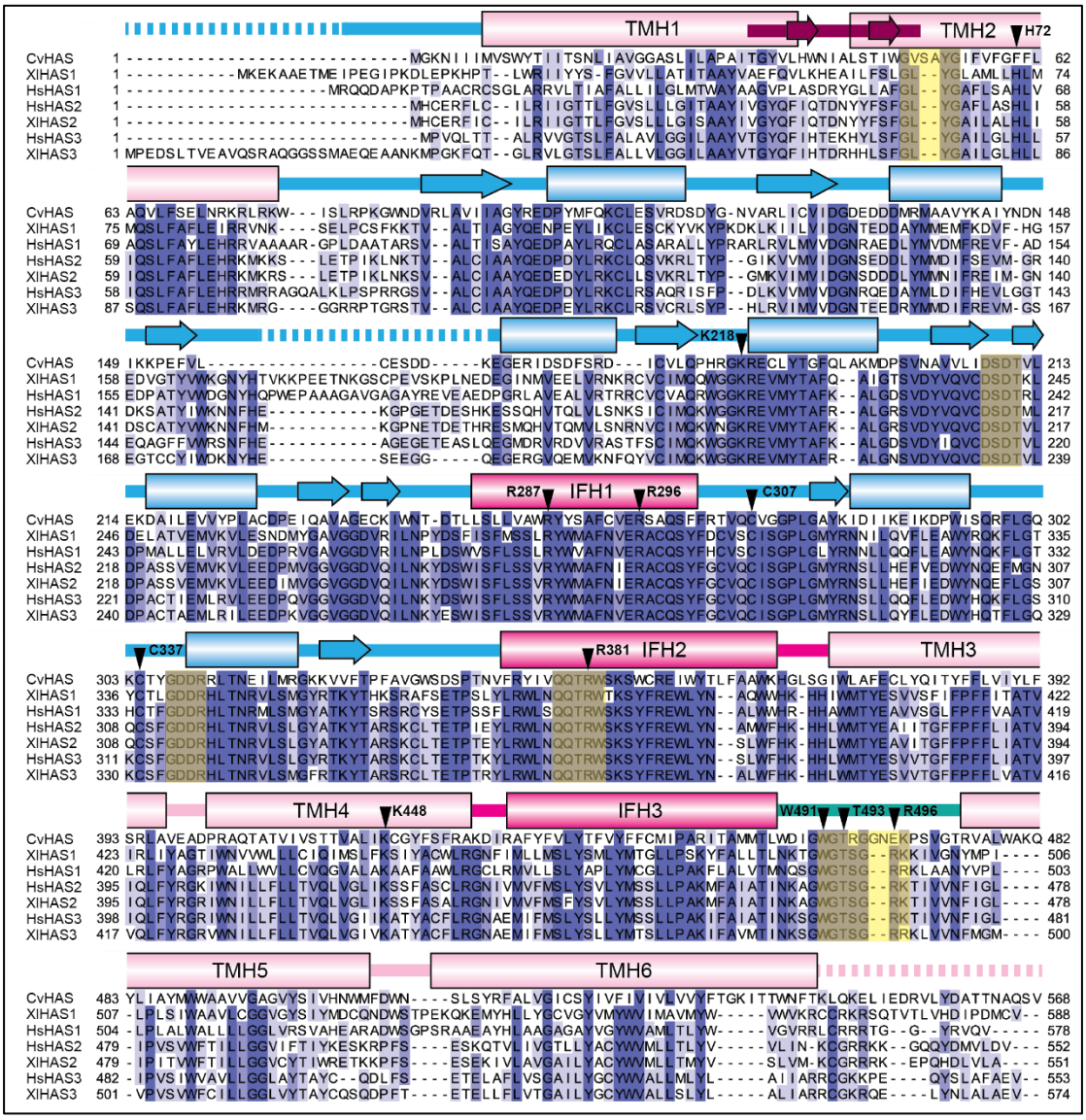


Figure S1 HAS sequence alignment

HAS sequence alignments (MUSCLE) [95] showing conserved motifs (GLYG, DxDT, GDDR, QxxRW and WGTSGRK/R, highlighted in yellow), as well as conserved residues used for mutagenesis studies (indicated with arrowheads). The bar on top of the sequences is colored according to structural domains of XlHAS1 (light pink: TM domain, purple: TMH1-TMH2 loop, pink: IF domain, blue: GT domain, teal: gating loop). Cylinders indicate α-helices, arrows β-sheets, lines loops, while dashed lines correspond to unstructured regions.

Table S1 CryoEM data collection, refinement and validation statistics

	Xl-HAS-1 Apo (EMDB-40591) (PDB 8SMM)	Xl-HAS-1 HA-bound (EMDB-40594) (PDB 8SMN)	Xl-HAS-1 UDP-bound (EMDB-40598) (PDB 8SMP)
Data collection and processing			
Magnification	81,000	81,000	81,000
Voltage (kV)	300	300	300
Electron exposure (e–/Å^2)	50	50	50
Defocus range (μm)	-2.0 to -1.0	-2.0 to -1.0	-2.0 to -1.0
Pixel size (Å)	1.08	1.08	1.08
Symmetry imposed	C1	C1	C1
Initial particle images (no.)	6,543,217	22,061,285	26,290,833
Final particle images (no.)	337,022	169,894	136,942
Map resolution (Å)	3.2	3.0	3.2
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	2.7-47.6	2.6-45.1	2.8-49.0
Refinement			
Initial model used (PDB code)	Alphafold2	8SMM	8SMM
Model resolution (Å)	3.2	3.0	3.2
FSC threshold	0.143	0.143	0.143
Map sharpening *B* factor (Å^2)	95.4	102.0	87.0
Model composition			
Non-hydrogen atoms	5930	6138	6097
Protein residues	738	749	757
Ligands	0	9	2
B factors (Å^2)			
Protein	75.6	35.98	73.11
Ligand	-	50.07	60.99
R.m.s. deviations			
Bond lengths (Å)	0.004	0.002	0.002
Bond angles (°)	0.660	0.519	0.526
Validation			
MolProbity score	1.63	1.48	1.64
Clashscore	6.97	4.61	6.29
Poor rotamers (%)	0.31	0.15	0.3
Ramachandran plot			
Favored (%)	96.28	96.35	95.72
Allowed (%)	3.72	3.65	4.15
Disallowed (%)	0.00	0.00	0.13

References

1. Lairson, L.L., et al., *Glycosyltransferases: structures, functions, and mechanisms.* Annu Rev Biochem, 2008. **77**: p. 521-55.
2. Rini, J.M., et al., *Glycosyltransferases and Glycan-Processing Enzymes*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022: Cold Spring Harbor (NY). p. 67-78.
3. Taujale, R., et al., *Deep evolutionary analysis reveals the design principles of fold A glycosyltransferases.* Elife, 2020. **9**.
4. Lombard, V., et al., *The carbohydrate-active enzymes database (CAZy) in 2013.* Nucleic Acids Res, 2014. **42**(Database issue): p. D490-5.
5. Zhang, P., et al., *Glycosyltransferase GT1 family: Phylogenetic distribution, substrates coverage, and representative structural features.* Comput Struct Biotechnol J, 2020. **18**: p. 1383-1390.
6. DeAngelis, P.L. and J. Zimmer, *Hyaluronan synthases; mechanisms, myths, & mysteries of three types of unique bifunctional glycosyltransferases.* Glycobiology, 2023. **33**(12): p. 1117-1127.
7. Yakovlieva, L. and M.T.C. Walvoort, *Processivity in Bacterial Glycosyltransferases.* ACS Chem Biol, 2020. **15**(1): p. 3-16.
8. Girish, K.S. and K. Kemparaju, *The magic glue hyaluronan and its eraser hyaluronidase: a biological overview.* Life Sci, 2007. **80**(21): p. 1921-43.
9. Cook, A.C., et al., *Osteopontin induction of hyaluronan synthase 2 expression promotes breast cancer malignancy.* J Biol Chem, 2006. **281**(34): p. 24381-9.
10. 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, 2000. **106**(3): p. 349-60.
11. Toole, B.P. and M.G. Slomiany, *Hyaluronan: a constitutive regulator of chemoresistance and malignancy in cancer cells.* Semin Cancer Biol, 2008. **18**(4): p. 244-50.
12. Tian, X., et al., *High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat.* Nature, 2013. **499**(7458): p. 346-9.
13. Singleton, P.A., *Hyaluronan regulation of endothelial barrier function in cancer.* Adv Cancer Res, 2014. **123**: p. 191-209.
14. Blackburn, M.R., et al., *Distinct reaction mechanisms for hyaluronan biosynthesis in different kingdoms of life.* Glycobiology, 2018. **28**(2): p. 108-121.
15. Morgan, J.L., et al., *Observing cellulose biosynthesis and membrane translocation in crystallo.* Nature, 2016. **531**(7594): p. 329-34.
16. DeAngelis, P.L., et al., *Hyaluronan synthase of chlorella virus PBCV-1.* Science, 1997. **278**(5344): p. 1800-3.
17. Van Etten, J.L., et al., *Chloroviruses Have a Sweet Tooth.* Viruses, 2017. **9**(4).
18. Maloney, F.P., et al., *Structure, substrate recognition and initiation of hyaluronan synthase.* Nature, 2022. **604**(7904): p. 195-201.
19. Agarwal, G., et al., *Biosynthesis of Hyaluronic acid polymer: Dissecting the role of sub structural elements of hyaluronan synthase.* Sci Rep, 2019. **9**(1): p. 12510.

20. Cowman, M.K., et al., *The Content and Size of Hyaluronan in Biological Fluids and Tissues.* Front Immunol, 2015. **6**: p. 261.
21. Spicer, A.P. and J.A. McDonald, *Characterization and molecular evolution of a vertebrate hyaluronan synthase gene family.* J Biol Chem, 1998. **273**(4): p. 1923-32.
22. Errasti-Murugarren, E., P. Bartoccioni, and M. Palacin, *Membrane Protein Stabilization Strategies for Structural and Functional Studies.* Membranes, 2021. **11**(2).
23. Pandey, A., et al., *Current strategies for protein production and purification enabling membrane protein structural biology.* Biochem Cell Biol, 2016. **94**(6): p. 507-527.
24. Drew, D., et al., *Optimization of membrane protein overexpression and purification using GFP fusions.* Nat Methods, 2006. **3**(4): p. 303-13.
25. Goehring, A., et al., *Screening and large-scale expression of membrane proteins in mammalian cells for structural studies.* Nat Protoc, 2014. **9**(11): p. 2574-85.
26. Kawate, T. and E. Gouaux, *Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins.* Structure, 2006. **14**(4): p. 673-81.
27. Hanson, M.A., et al., *Profiling of membrane protein variants in a baculovirus system by coupling cell-surface detection with small-scale parallel expression.* Protein Expr Purif, 2007. **56**(1): p. 85-92.
28. Hubbard, C., et al., *The hyaluronan synthase catalyzes the synthesis and membrane translocation of hyaluronan.* J Mol Biol, 2012. **418**(1-2): p. 21-31.
29. Breibeck, J. and A. Rompel, *Successful amphiphiles as the key to crystallization of membrane proteins: Bridging theory and practice.* Biochim Biophys Acta Gen Subj, 2019. **1863**(2): p. 437-455.
30. Kampjut, D., J. Steiner, and L.A. Sazanov, *Cryo-EM grid optimization for membrane proteins.* iScience, 2021. **24**(3): p. 102139.
31. Tonggu, L. and L. Wang, *Cryo-EM sample preparation method for extremely low concentration liposomes.* Ultramicroscopy, 2020. **208**: p. 112849.
32. Hagn, F., M.L. Nasr, and G. Wagner, *Assembly of phospholipid nanodiscs of controlled size for structural studies of membrane proteins by NMR.* Nat Protoc, 2018. **13**(1): p. 79-98.
33. Denisov, I.G. and S.G. Sligar, *Nanodiscs for structural and functional studies of membrane proteins.* Nat Struct Mol Biol, 2016. **23**(6): p. 481-6.
34. Wang, L. and L. Tonggu, *Membrane protein reconstitution for functional and structural studies.* Sci China Life Sci, 2015. **58**(1): p. 66-74.
35. Rigaud, J.L., et al., *Bio-Beads: an efficient strategy for two-dimensional crystallization of membrane proteins.* J Struct Biol, 1997. **118**(3): p. 226-35.
36. Oluwole, A.O., et al., *Formation of Lipid-Bilayer Nanodiscs by Diisobutylene/Maleic Acid (DIBMA) Copolymer.* Langmuir, 2017. **33**(50): p. 14378-14388.
37. Ravula, T., N.Z. Hardin, and A. Ramamoorthy, *Polymer nanodiscs: Advantages and limitations.* Chemistry and Physics of Lipids, 2019. **219**: p. 45-49.

38. Rouck, J.E., et al., *Recent advances in nanodisc technology for membrane protein studies (2012-2017).* FEBS Lett, 2017. **591**(14): p. 2057-2088.
39. Bi, Y., et al., *Insights into the structure and function of membrane-integrated processive glycosyltransferases.* Curr Opin Struct Biol, 2015. **34**: p. 78-86.
40. Lee, H.G. and M.K. Cowman, *An agarose gel electrophoretic method for analysis of hyaluronan molecular weight distribution.* Anal Biochem, 1994. **219**(2): p. 278-87.
41. Caon, I., et al., *Cell Energy Metabolism and Hyaluronan Synthesis.* J Histochem Cytochem, 2021. **69**(1): p. 35-47.
42. Cowman, M.K., et al., *Improved agarose gel electrophoresis method and molecular mass calculation for high molecular mass hyaluronan.* Anal Biochem, 2011. **417**(1): p. 50-6.
43. Volpi, N., F. Maccari, and J. Titze, *Simultaneous detection of submicrogram quantities of hyaluronic acid and dermatan sulfate on agarose-gel by sequential staining with toluidine blue and Stains-All.* J Chromatogr B Analyt Technol Biomed Life Sci, 2005. **820**(1): p. 131-5.
44. Goldberg, H.A. and K.J. Warner, *The staining of acidic proteins on polyacrylamide gels: enhanced sensitivity and stability of "Stains-all" staining in combination with silver nitrate.* Anal Biochem, 1997. **251**(2): p. 227-33.
45. Kayashima, Y., et al., *Reduction of Stabilin-2 Contributes to a Protection Against Atherosclerosis.* Front Cardiovasc Med, 2022. **9**: p. 818662.
46. Rilla, K., et al., *Hyaluronan synthase 1 (HAS1) requires higher cellular UDP-GlcNAc concentration than HAS2 and HAS3.* J Biol Chem, 2013. **288**(8): p. 5973-83.
47. Renaud, J.P., et al., *Cryo-EM in drug discovery: achievements, limitations and prospects.* Nat Rev Drug Discov, 2018. **17**(7): p. 471-492.
48. Fan, X., et al., *Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution.* Nat Commun, 2019. **10**(1): p. 2386.
49. Ashraf, K.U., et al., *Structural basis of lipopolysaccharide maturation by the O-antigen ligase.* Nature, 2022. **604**(7905): p. 371-376.
50. Miller, K.R., et al., *T cell receptor-like recognition of tumor in vivo by synthetic antibody fragment.* PLoS One, 2012. **7**(8): p. e43746.
51. Bjorck, L., *Protein L. A novel bacterial cell wall protein with affinity for Ig L chains.* J Immunol, 1988. **140**(4): p. 1194-7.
52. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* Nature, 2021. **596**(7873): p. 583-589.
53. Punjani, A., et al., *cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination.* Nat Methods, 2017. **14**(3): p. 290-296.
54. Punjani, A., H. Zhang, and D.J. Fleet, *Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction.* Nat Methods, 2020. **17**(12): p. 1214-1221.
55. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

56. Adams, P., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
57. Cantarel, B.L., et al., *The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics.* Nucleic Acids Res, 2009. **37**(Database issue): p. D233-8.
58. Purushotham, P., R. Ho, and J. Zimmer, *Architecture of a catalytically active homotrimeric plant cellulose synthase complex.* Science, 2020. **369**(6507): p. 1089-1094.
59. Chen, W., et al., *Structural basis for directional chitin biosynthesis.* Nature, 2022. **610**(7931): p. 402-408.
60. Ren, Z., et al., *Structural basis for inhibition and regulation of a chitin synthase from Candida albicans.* Nat Struct Mol Biol, 2022. **29**(7): p. 653-664.
61. Verma, P., et al., *Insights into substrate coordination and glycosyl transfer of poplar cellulose synthase-8.* Structure, 2023. **31**(10): p. 1166-+.
62. Morgan, J.L.W., J.T. McNamara, and J. Zimmer, *Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP.* Nature Struct Mol Biol, 2014. **21**(5): p. 489-496.
63. Punjani, A. and D.J. Fleet, *3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM.* J Struct Biol, 2021. **213**(2): p. 107702.
64. Górniak, I., et al., *Structural snapshots of hyaluronan formation reveal principles of length control and secretion.* bioRxiv, 2023: p. 2023.05.11.540447.
65. Almond, A., P.L. DeAngelis, and C.D. Blundell, *Hyaluronan: the local solution conformation determined by NMR and computer modeling is close to a contracted left-handed 4-fold helix.* J Mol Biol, 2006. **358**(5): p. 1256-69.
66. Jurrus, E., et al., *Improvements to the APBS biomolecular solvation software suite.* Protein Sci, 2018. **27**(1): p. 112-128.
67. Charnock, S.J. and G.J. Davies, *Structure of the nucleotide-diphospho-sugar transferase, SpsA from Bacillus subtilis, in native and nucleotide-complexed forms.* Biochem, 1999. **38**(20): p. 6380-6385.
68. Morgan, J., J. Strumillo, and J. Zimmer, *Crystallographic snapshot of cellulose synthesis and membrane translocation.* Nature, 2013. **493**(7431): p. 181-6.
69. Goentzel, B.J., P.H. Weigel, and R.A. Steinberg, *Recombinant human hyaluronan synthase 3 is phosphorylated in mammalian cells.* Biochem J, 2006. **396**(2): p. 347-54.
70. Vigetti, D. and A. Passi, *Hyaluronan synthases posttranslational regulation in cancer.* Adv Cancer Res, 2014. **123**: p. 95-119.
71. Zhao, Y., et al., *Evolution of high-molecular-mass hyaluronic acid is associated with subterranean lifestyle.* Nature Communs, 2023. **14**.
72. Itano, N., et al., *Three isoforms of mammalian hyaluronan synthases have distinct enzymatic properties.* J Biol Chem, 1999. **274**(35): p. 25085-92.
73. Vigetti, D., et al., *Metabolic control of hyaluronan synthases.* Matrix Biology, 2014. **35**: p. 8-13.

74. Kuklewicz, J. and J. Zimmer, *Molecular insights into capsular polysaccharide secretion.* Nature, 2024. **628**(8009): p. 901-909.
75. Caffalette, C.A., et al., *A lipid gating mechanism for the channel-forming O antigen ABC transporter.* Nat Commun, 2019. **10**(1): p. 824.
76. Knott, B.C., et al., *Simulations of cellulose translocation in the bacterial cellulose synthase suggest a regulatory mechanism for the dimeric structure of cellulose.* Chem Sci, 2016. **7**(5): p. 3108-3116.
77. McManus, J., et al., *Initiation, Elongation, and Termination of Bacterial Cellulose Synthesis.* ACS Omega, 2018. **3**(3): p. 2690-2698.
78. Drew, D., et al., *GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae.* Nat Protoc, 2008. **3**(5): p. 784-98.
79. Drew, D.E., et al., *Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli.* FEBS Lett, 2001. **507**(2): p. 220-4.
80. Yariv, B., et al., *Using evolutionary data to make sense of macromolecules with a "face-lifted" ConSurf.* Protein Sci, 2023. **32**(3): p. e4582.
81. Miyagi, H., et al., *Structural insight into an anti-BRIL Fab as a G-protein-coupled receptor crystallization chaperone.* Acta Crystallogr D Struct Biol, 2023. **79**(Pt 5): p. 435-441.
82. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nature Methods, 2009. **6**: p. 343.
83. Haserodt, S., M. Aytekin, and R.A. Dweik, *A comparison of the sensitivity, specificity, and molecular weight accuracy of three different commercially available Hyaluronan ELISA-like assays.* Glycobiology, 2011. **21**(2): p. 175-83.
84. Zhao, Y., et al., *Evolution of high-molecular-mass hyaluronic acid is associated with subterranean lifestyle.* Nat Commun, 2023. **14**(1): p. 8054.
85. Gwosch, K.C., et al., *MINFLUX nanoscopy delivers 3D multicolor nanometer resolution in cells.* Nat Methods, 2020. **17**(2): p. 217-224.
86. Theocharis, A.D., et al., *Extracellular matrix structure.* Adv Drug Deliv Rev, 2016. **97**: p. 4-27.
87. Yang, X., et al., *Structure deformation and curvature sensing of PIEZO1 in lipid membranes.* Nature, 2022. **604**(7905): p. 377-383.
88. Liu, H. and J.H. Naismith, *An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol.* BMC Biotechnol, 2008. **8**: p. 91.
89. Dickinson, D.J., et al., *Engineering the Caenorhabditis elegans genome using Cas9-triggered homologous recombination.* Nat Methods, 2013. **10**(10): p. 1028-34.
90. Hilton, M.A., et al., *Single-molecule investigations of single-chain cellulose biosynthesis.* Proc Natl Acad Sci U S A, 2022. **119**(40): p. e2122770119.
91. Fellouse, F.A., et al., *High-throughput generation of synthetic antibodies from highly functional minimalist phage-displayed libraries.* J Mol Biol, 2007. **373**(4): p. 924-40.
92. Kim, J., et al., *Structure and drug resistance of the Plasmodium falciparum transporter PfCRT.* Nature, 2019. **576**(7786): p. 315-320.

93. Jedrzejas, M.J., et al., *Expression and purification of Streptococcus pneumoniae hyaluronate lyase from Escherichia coli.* Protein Expr Purif, 1998. **13**(1): p. 83-9.
94. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.
95. Madeira, F., et al., *The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024.* Nucleic Acids Res, 2024.