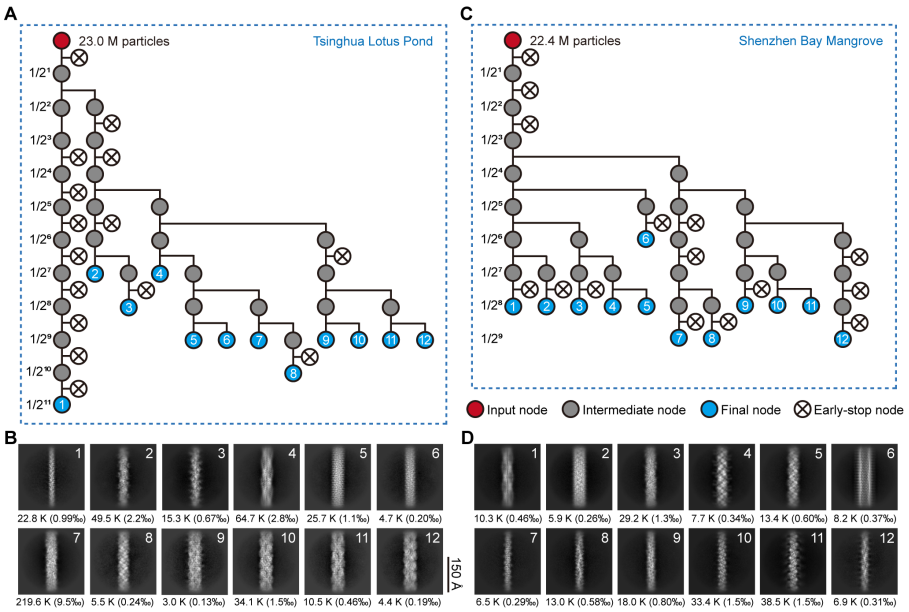
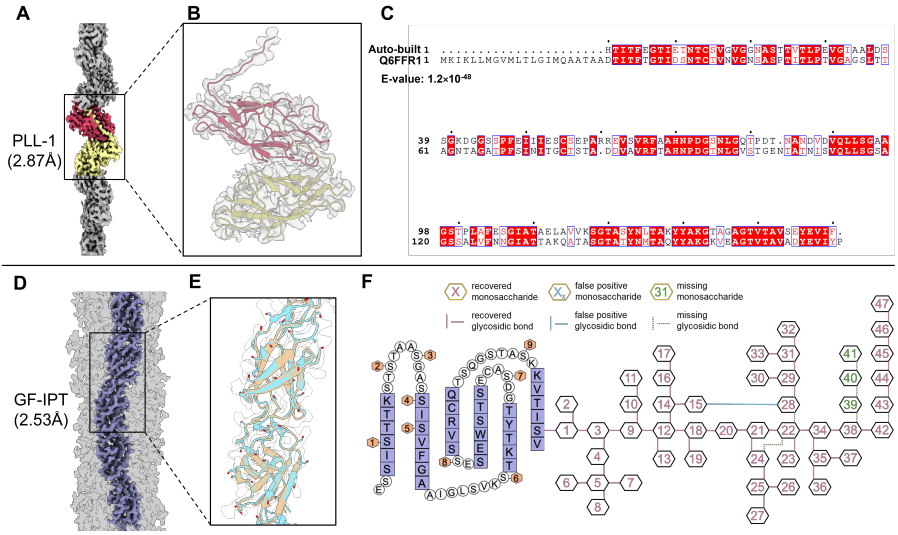
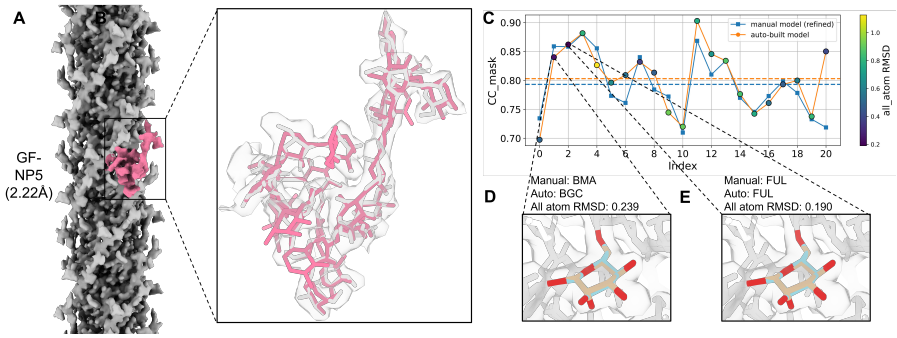


Supplementary Figure 1: The RBC strategy was applied to the 2D classification of particles derived from the water samples collected at Tsinghua lotus pond and the Shenzhen Bay mangrove. (A)-(B) A binary tree generated via the RBC strategy is presented alongside twelve characterized fibril structures. Input particles were picked from cryo-EM micrographs of samples collected at Tsinghua lotus pond. Final nodes of the binary tree comprise particles that facilitated the reconstruction of high-resolution density maps. The number of such particles, together with their proportion relative to the total input particles, is annotated at the bottom. **(C)-(D)** Similar to **(A)-(B)**, except that the water sample was collected at the Shenzhen Bay mangrove.



Supplementary Figure 2: **Automatic model building of PF-PLL1 and GF-IPT1 by EModelG. (A)** Cryo-EM density map of PLL-1 at 2.87Å resolution. **(B)** The EModelG-auto-built model, depicting two subunits (pink and yellow) fitted into the EM density (white transparent). **(C)** Sequence alignment of the EModelG (w/o seq) auto-built model identified Q6FFR1, a major fimbrial subunit protein from Acinetobacter baylyi, as the top hit in the UniProtKB database via BLAST. Following this alignment, Q6FFR1 was used to rebuild the final model for PF-PLL1. (D) Cryo-EM density map of GF-IPT1 at 2.87Å resolution. **(E)** Structural comparison between the EModelG-auto-built model (sky-blue) and the manually built reference model (tan). The side chains at the glycosylation site in the auto-built model are represented as sticks. **(F)** Schematic diagram of the glycosylation site (left) and the auto-built glycan structure against the manual model (right).



Supplementary Figure 3: **Automatic model building of GF-NP5 by EModelG. (A)** Cryo-EM density map of the protein-free glycofibril GF-NP5 at 2.2Å resolution, with a representative glycan segment boxed. **(B)** Zoomed-in view of the boxed region showing a comparison between the auto-built glycan model (pink) and the manually built reference model (white). **(C)** Per-residue map–model correlation coefficients CC_mask for the auto-built (orange) and manual (blue) models along the glycan chain; point colors encode the all-atom RMSD between the two models as indicated by the color bar, and horizontal dashed lines denote the chain-averaged CC_mask values for each model. **(D)-(E)** Zoomed-in comparison of some monosaccharide units in the manual model (tan) and the auto-built model (sky blue).

site	coordinate	description	location
A	116.3190°E, 40.0024°N	Tsinghua lotus pond water	Beijing, China
B	114.0096°E, 22.5246°N	Shenzhen Bay mangrove water	Guangdong, China
C	101.2700°E, 21.9161°N	Xishuangbanna botanical garden	Yunnan, China
D	109.4986°E, 18.2116°N	sea water from Sanya	Hainan, China
E	121.8342°E, 40.9050°N	coastal wetland in Panjin	Liaoning, China
F	109.4290°E, 22.9388°N	well water from Guigang	Guangxi, China
G	110.3183°E, 25.2167°N	karst cave in Guilin	Guangxi, China
H	98.4692°E, 25.4410°N	high-temperature spring water from Tengchong	Yunnan, China
I	124.0852°E, 41.3032°N	karst cave in Benxi	Liaoning, China
J	120.2304°E, 30.1565°N	Xiang Lake water from Hangzhou	Zhejiang, China
K	90.9345°E, 29.4409°N	Lhasa River	Tibet, China
L	104.7970°E, 25.4815°N	karst cave in Liupanshui	Guizhou, China
M	81.2445°E, 43.8969°N	Ili River	Xinjiang, China

Supplementary Table 1: The alphabetical identifiers and coordinates of the 13 water sampling sites in this study.

Name		Resolution (Å)	Absolute hand	Rise (Å)	Twist (°)
PF-PLL	1	2.87	measured	26.50	-156.18
	2	2.85	measured	86.86	99.71
	3	2.59	measured	10.85	85.08
	4	2.26	measured	10.05	103.69
	5	3.37	measured	10.08	100.90
	6	2.70	measured	5.27	109.05
	7	3.66	measured	10.40	112.27
	8	2.73	measured	11.74	-76.14
	9	2.35	measured	12.18	-78.17
	10	2.36	measured	11.48	-74.33
	11	3.44	measured	11.53	-75.17
	12	3.56	measured	11.56	-74.58
	13	3.73	measured	10.09	94.09
GF-IPT	1	2.53	measured	32.24	85.63
GF-1BH	1	2.54	unknown	7.04	+/-149.57
	2	3.11	unknown	7.07	+/-148.93
GF-3BH	1	2.35	measured	9.61	-11.54
	2	2.63	measured	9.63	-11.12
	3	2.66	measured	9.62	-10.88
GF-1L8R	1	3.00	inferred	24.28	83.23
GF-1L6R	1	3.56	inferred	17.30	-55.41
GF-1L4R	1	3.06	measured	12.05	-138.51
	2	2.60	measured	12.24	-138.50
	3	2.92	inferred	12.14	-136.36
	4	3.42	inferred	12.13	-134.91
	5	3.56	inferred	12.42	-133.93
	6	3.55	inferred	12.40	-138.60
	7	2.89	inferred	12.24	-138.36
	8	2.78	inferred	12.24	-138.50
GF-3L3R	1	3.18	unknown	2.95	+/-105.69
	2	3.14	unknown	2.96	+/-99.85
GF-1L2R	1	2.80	inferred	5.93	110.00
	2	2.58	measured	5.80	136.31
	3	2.90	inferred	6.03	104.11
	4	2.67	inferred	6.02	104.11
	5	2.53	measured	5.54	132.10
	6	2.83	measured	6.09	108.70
	7	2.74	inferred	6.00	106.97
	8	2.88	measured	6.16	93.77
GF-NP	1	3.48	unknown	7.72	+/-109.00
	2	2.83	unknown	8.21	+/-100.86
	3	2.97	unknown	26.30	+/-20.89
	4	2.98	unknown	26.05	+/-23.57
	5	2.22	unknown	4.38	+/-100.91
	6	3.33	unknown	7.64	+/-98.71
	7	3.23	unknown	2.44	+/-56.69
	8	2.33	measured	2.65	78.49
	9	2.40	inferred	2.65	79.14
	10	3.00	inferred	2.68	79.05
	11	2.66	unknown	1.15	+/-27.40

Supplementary Table 2: Resolution, absolute hand, rise, and twist of the 50 reported fibrils.

Site	A	B	Ca	Cb	D	E	Fa	Fb	Ga	Gb	H	I	J	K	L	M
PF-PLL1												2.87Å				
PF-PLL2												2.85Å				
PF-PLL3											2.59Å					
PF-PLL4											2.26Å					
PF-PLL5										3.37Å						
PF-PLL6													3.64Å		2.70Å	3.20Å
PF-PLL7										3.66Å						
PF-PLL8										2.73Å						
PF-PLL9	3.21Å			3.30Å					3.00Å	3.40Å			3.68Å	3.49Å		2.35Å
PF-PLL10	3.07Å			3.68Å					2.69Å	3.37Å			3.39Å			2.36Å
PF-PLL11	3.44Å															
PF-PLL12	3.56Å															
PF-PLL13										3.73Å						
GF-IPT1	3.03Å	3.14Å	2.25Å	3.77Å		3.64Å			2.85Å	2.77Å			3.85Å			
GF-1BH1			2.54Å			3.34Å										
GF-1BH2			3.11Å													
GF-3BH1					2.35Å	2.98Å										
GF-3BH2	2.86Å	2.83Å			2.63Å	2.98Å			2.82Å	2.93Å			3.18Å			
GF-3BH3					2.66Å											
GF-1L8R1								3.00Å								
GF-1L6R1			3.56Å													
GF-1L4R1		3.10Å								3.06Å						
GF-1L4R2			3.03Å	3.86Å				2.87Å	3.01Å	2.79Å		2.60Å	3.75Å	3.95Å	3.67Å	3.09Å
GF-1L4R3		2.92Å														
GF-1L4R4						3.42Å										
GF-1L4R5	3.56Å															
GF-1L4R6	3.55Å															
GF-1L4R7								2.89Å			3.15Å					
GF-1L4R8			2.88Å		3.04Å				3.24Å					3.21Å*		
GF-3L3R1	2.95Å															
GF-3L3R2						3.14Å										
GF-1L2R1		2.83Å	2.80Å													
GF-1L2R2									2.64Å	2.58Å				3.27Å		3.22Å
GF-1L2R3		2.90Å														
GF-1L2R4		2.67Å														
GF-1L2R5									2.53Å	2.64Å			3.29Å			
GF-1L2R6	3.81Å	2.83Å	2.94Å						4.12Å	3.10Å		3.06Å		3.91Å	3.72Å	3.17Å
GF-1L2R7		2.74Å														
GF-1L2R8		3.11Å								2.80Å		2.78Å			3.21Å	
GF-NP1											3.48Å					
GF-NP2									2.83Å	3.25Å						
GF-NP3						2.97Å										
GF-NP4							3.31Å	3.57Å			2.98Å					
GF-NP5			3.60Å	3.42Å					2.39Å	2.22Å						
GF-NP6		3.33Å					3.63Å		3.87Å							
GF-NP7						3.23Å										
GF-NP8									2.33Å	3.41Å		2.99Å	3.30Å			
GF-NP9	3.52Å			3.54Å	2.98Å		3.26Å		2.40Å	3.27Å			3.21Å			
GF-NP10	3.00Å															
GF-NP11		2.66Å														

Supplementary Table 3: **Summary of the 126 fibrils resolved from the 13 water samples and their corresponding resolutions.** Ca and Cb represent two batches of water samples collected from Site C. Similarly, Fa, Fb, Ga, and Gb denote two batches of water samples collected from Sites F and G, respectively. The highlighted entry denotes the density map with the highest resolution among those acquired for this type of fibrils. The density maps corresponding to this highlighted entry are presented in Figure 2, while the respective resolution, absolute hand, and helical parameters are listed in Supplementary Table 2.

Supplementary Algorithm S1 Initial monosaccharide placement using CNN and NMS

Input: Cryo-EM density map $\rho : \Omega \subset \mathbb{R}^3 \to \mathbb{R}$;
Trained CNN f_θ for carbohydrate density prediction;
Density threshold τ;
D-Xylopyranose template $T^{\mathrm{Xyl}} = \{\mathbf{x}_i \in \mathbb{R}^3\}_{i=1}^{N_{\mathrm{atom}}}$;
Set of rotation samples $\mathcal{R} = \{R_k \in \mathrm{SO}(3)\}_{k=1}^{N_R}$ (Fibonacci-sphere based);
Translation radius $r_{\mathrm{trans}} = 1\,\text{Å}$;
Clash distance threshold d_{clash}.
Output: Initial monosaccharide model $\mathcal{S}_0 = \{p_j\}_{j=1}^{N_0}$,
where each pose $p_j = (\mathbf{c}_j, R_j, \mathbf{t}_j)$ encodes ring center, rotation and translation.

// 1. Predict carbohydrate-specific density
$p_{\mathrm{gly}}(\mathbf{x}) \leftarrow f_\theta(\rho)(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega$
// 2. Thresholding to obtain candidate carbohydrate regions
$\mathcal{M}_{\mathrm{gly}} \leftarrow \{\mathbf{x} \in \Omega \mid p_{\mathrm{gly}}(\mathbf{x}) \geq \tau\}$ Decompose $\mathcal{M}_{\mathrm{gly}}$ into connected components $\{\mathcal{C}_u\}_{u=1}^{U}$
// 3. Generate candidate monosaccharide poses by rotation+translation sampling
$\mathcal{P}_{\mathrm{cand}} \leftarrow \emptyset$ **foreach** $\mathcal{C}_u$ **do**
  Sample ring centers $\{\mathbf{c}_{u,n}\}_{n=1}^{N_u}$ on $\mathcal{C}_u$ (e.g., voxel centers or local maxima) **foreach** $\mathbf{c}_{u,n}$ **do**
    foreach $R_k \in \mathcal{R}$ **do**
      Sample translations $\{\mathbf{t}_{u,n,k,m}\}_{m=1}^{M}$ with $\|\mathbf{t}_{u,n,k,m}\| \leq r_{\mathrm{trans}}$ **foreach** $\mathbf{t}_{u,n,k,m}$ **do**
        Define pose $p = (\mathbf{c}_{u,n}, R_k, \mathbf{t}_{u,n,k,m})$ $\mathcal{P}_{\mathrm{cand}} \leftarrow \mathcal{P}_{\mathrm{cand}} \cup \{p\}$
// 4. Score each candidate pose using grid_sample
foreach $p = (\mathbf{c}, R, \mathbf{t}) \in \mathcal{P}_{\mathrm{cand}}$ **do**
  // Transform template atoms into map coordinates
  $\mathbf{x}_i^{(p)} \leftarrow R\mathbf{x}_i + \mathbf{c} + \mathbf{t}, \quad i = 1, \ldots, N_{\mathrm{atom}}$
  // Sample density at all atom positions
  $\{d_i^{(p)}\}_{i=1}^{N_{\mathrm{atom}}} \leftarrow \mathrm{grid\_sample}\big(\rho, \{\mathbf{x}_i^{(p)}\}_{i=1}^{N_{\mathrm{atom}}}\big)$
  // Restrict to ring atoms $\mathcal{A}_{\mathrm{ring}} \subset \{1, \ldots, N_{\mathrm{atom}}\}$
  $\mu_{\mathrm{ring}}(p) \leftarrow \frac{1}{|\mathcal{A}_{\mathrm{ring}}|} \sum_{i \in \mathcal{A}_{\mathrm{ring}}} d_i^{(p)}$ $\sigma_{\mathrm{ring}}(p) \leftarrow \sqrt{\frac{1}{|\mathcal{A}_{\mathrm{ring}}|} \sum_{i \in \mathcal{A}_{\mathrm{ring}}} \big(d_i^{(p)} - \mu_{\mathrm{ring}}(p)\big)^2}$
  // Mean minus standard deviation as pose score
  $s(p) \leftarrow \mu_{\mathrm{ring}}(p) - \sigma_{\mathrm{ring}}(p)$
// 5. Non-maximum suppression within each density component
$\mathcal{S}_0 \leftarrow \mathrm{NMS}\big(\mathcal{P}_{\mathrm{cand}}, \{s(p)\}, \{\mathcal{C}_u\}, d_{\mathrm{clash}}\big)$ // $\mathcal{S}_0$ is an assembly of D-Xylopyranose poses without glycosidic bonds
// 6. Real-space refinement of the initial monosaccharide assembly
$\mathcal{S}_0 \leftarrow \mathrm{RealSpaceRefine}\big(\mathcal{S}_0, \rho\big)$
return $\mathcal{S}_0$

Supplementary Algorithm S2 NMS for monosaccharide pose selection within each density component

Input: Candidate poses $\mathcal{P}_{\mathrm{cand}}$;
Pose scores $\{s(p) \mid p \in \mathcal{P}_{\mathrm{cand}}\}$;
Connected components $\{\mathcal{C}_u\}$ of $\mathcal{M}_{\mathrm{gly}}$;
Clash distance threshold d_{clash}.
Output: Selected pose set $\mathcal{S}_0$.

$\mathcal{S}_0 \leftarrow \emptyset$ **foreach** $\mathcal{C}_u$ **do**
  $\mathcal{P}_u \leftarrow \{p \in \mathcal{P}_{\mathrm{cand}} \mid \mathrm{center}(p) \in \mathcal{C}_u\}$ Sort $\mathcal{P}_u$ in descending order of $s(p)$
  while $\mathcal{P}_u \neq \emptyset$ **do**
    $p^* \leftarrow \arg\max_{p \in \mathcal{P}_u} s(p)$ $\mathcal{S}_0 \leftarrow \mathcal{S}_0 \cup \{p^*\}$
    `// Remove poses that clash with` p^*
    $\mathcal{P}_u \leftarrow \mathcal{P}_u \setminus \{p \in \mathcal{P}_u \mid \mathrm{minDist}(p, p^*) < d_{\mathrm{clash}}\}$

return $\mathcal{S}_0$

Supplementary Algorithm S3 Glycosidic bond building and glycan chain growth

Input: Refined initial monosaccharide assembly $\mathcal{S}_0$ (D-Xylopyranose poses);
Automatically built protein model $\mathcal{P}_{\mathrm{prot}}$;
Cryo-EM density map ρ and glycan mask $\mathcal{M}_{\mathrm{gly}}$;
Set of candidate glycosylation sites $\mathcal{G} = \{g_\ell\}$ inferred from $\mathcal{S}_0$ and $\mathcal{P}_{\mathrm{prot}}$;
Library of monosaccharide templates $\mathcal{T} = \{T^{(m)}\}_{m=1}^{22}$;
Target glycosidic bond length ℓ_0 and tolerance δ;
Distance/connectivity thresholds for density-based linkage tests.
Output: Final glycan model $\mathcal{S}_*$ with glycosidic bonds $\mathcal{E}_*$.

`// 1. Segment glycan density by glycosylation sites`
Partition $\mathcal{M}_{\mathrm{gly}}$ into regions $\{\Omega_g\}_{g\in\mathcal{G}}$ such that each Ω_g is the density region associated with glycosylation site g
`// 2. Assign initial monosaccharides to local regions`
foreach $g \in \mathcal{G}$ **do**
  $\mathcal{S}_0^g \leftarrow \{s \in \mathcal{S}_0 \mid \mathrm{center}(s) \in \Omega_g\}$
$\mathcal{S}_* \leftarrow \mathcal{S}_0$; `// poses will be re-typed later`
$\mathcal{E}_* \leftarrow \emptyset$
`// 3. For each glycosylation site, grow glycan chains`
foreach $g \in \mathcal{G}$ **do**
  `// 3.1 Initialize frontier from the glycosylation site`
  $\mathcal{F} \leftarrow \mathrm{InitFrontierFromSite}(g, \mathcal{S}_0^g)$ $\mathcal{C}_{\mathrm{used}}^g \leftarrow \emptyset$; `// used monosaccharide centers`
  while $\mathcal{F} \neq \emptyset$ **do**
    Select and remove a current monosaccharide s_{cur} from $\mathcal{F}$ $\mathcal{C}_{\mathrm{used}}^g \leftarrow \mathcal{C}_{\mathrm{used}}^g \cup \{\mathrm{center}(s_{\mathrm{cur}})\}$
    `// 3.2 Find nearest unassigned monosaccharide in the same density region`
    $\mathcal{N} \leftarrow \mathrm{NearestCandidates}\big(s_{\mathrm{cur}}, \mathcal{S}_0^g \setminus \mathcal{C}_{\mathrm{used}}^g, \Omega_g\big)$
    foreach $s_{\mathrm{cand}} \in \mathcal{N}$ **do**
      `// 3.3 Check density connectivity between current sugar and candidate`
      if $\mathrm{IsDensityConnected}(s_{\mathrm{cur}}, s_{\mathrm{cand}}, \rho, \Omega_g)$ **then**
        `// 3.4 Enumerate monosaccharide types and score glycosidic configuration`
        $(m^*, s_{\mathrm{new}}, \mathrm{score}^*) \leftarrow \mathrm{SelectBestMonosaccharideTemplate}(s_{\mathrm{cur}}, s_{\mathrm{cand}}, \mathcal{T}, \rho, \ell_0, \delta)$
        if score^* *passes acceptance threshold* **then**
          `// 3.5 Commit glycosidic bond and update model`
          $\mathcal{S}_* \leftarrow \mathcal{S}_* \cup \{s_{\mathrm{new}}\}$ $\mathcal{E}_* \leftarrow \mathcal{E}_* \cup \{(s_{\mathrm{cur}}, s_{\mathrm{new}})\}$
          `// Mark the candidate as consumed and add to frontier`
          $\mathcal{C}_{\mathrm{used}}^g \leftarrow \mathcal{C}_{\mathrm{used}}^g \cup \{\mathrm{center}(s_{\mathrm{cand}})\}$ $\mathcal{F} \leftarrow \mathcal{F} \cup \{s_{\mathrm{new}}\}$
return $(\mathcal{S}_*, \mathcal{E}_*)$

Supplementary Algorithm S4 SelectBestMonosaccharideTemplate

Input: Current sugar s_{cur};
Candidate position s_{cand} (geometric anchor for the next sugar);
Monosaccharide template library $\mathcal{T} = \{T^{(m)}\}_{m=1}^{22}$;
Density map ρ;
Target bond length ℓ_0, tolerance δ.
Output: Best template index m^*;
New sugar pose s_{new} (typed and positioned);
Best score score^*.

$\mathrm{score}^* \leftarrow -\infty$, $m^* \leftarrow$ None, $s_{\mathrm{new}} \leftarrow$ None
foreach $m \in \{1, \ldots, 22\}$ **do**
  `// Superimpose template` $T^{(m)}$ `onto the current sugar and candidate anchor`
  $s_{\mathrm{trial}} \leftarrow \mathrm{SuperimposeTemplate}(T^{(m)}, s_{\mathrm{cur}}, s_{\mathrm{cand}}, \ell_0, \delta)$ **if** s_{trial} *is geometrically valid* **then**
    `// Evaluate density-based score`
    $\mathrm{score}(m) \leftarrow \mathrm{ScoreMonosaccharidePose}(s_{\mathrm{trial}}, \rho)$
    if $\mathrm{score}(m) > \mathrm{score}^*$ **then**
      $\mathrm{score}^* \leftarrow \mathrm{score}(m)$ $m^* \leftarrow m$ $s_{\mathrm{new}} \leftarrow s_{\mathrm{trial}}$

return $(m^*, s_{\mathrm{new}}, \mathrm{score}^*)$