



**HAL**  
open science

## Elucidating the glycan-binding specificity and structure of *Cucumis melo* agglutinin, a new R-type lectin

Jon Lundstrøm, Emilie Gillon, Valérie Chazalet, Nicole Kerekes, Antonio Di Maio, Ten Feizi, Yan Liu, Annabelle Varrot, Daniel Bojar

► **To cite this version:**

Jon Lundstrøm, Emilie Gillon, Valérie Chazalet, Nicole Kerekes, Antonio Di Maio, et al.. Elucidating the glycan-binding specificity and structure of *Cucumis melo* agglutinin, a new R-type lectin. *Beilstein Journal of Organic Chemistry*, 2024, 20, pp.306-320. 10.3762/bjoc.20.31 . hal-04542095

**HAL Id: hal-04542095**

**<https://hal.science/hal-04542095>**

Submitted on 11 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Elucidating the glycan-binding specificity and structure of *Cucumis melo* agglutinin, a new R-type lectin

Jon Lundstrøm<sup>1,2</sup>, Emilie Gillon<sup>3</sup>, Valérie Chazalet<sup>3</sup>, Nicole Kerekes<sup>1,2</sup>, Antonio Di Maio<sup>4</sup>, Ten Feizi<sup>4</sup>, Yan Liu<sup>4</sup>, Annabelle Varrot<sup>3</sup> and Daniel Bojar<sup>\*1,2</sup>

## Full Research Paper

[Open Access](#)

### Address:

<sup>1</sup>Department of Chemistry and Molecular Biology, University of Gothenburg, Medicinaregatan 7B, 413 90 Gothenburg, Sweden, <sup>2</sup>Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, 413 90 Gothenburg, Sweden, <sup>3</sup>Univ. Grenoble Alpes, CNRS, CERMAV, 601 Rue de la Chimie, 38610 Gières, France and <sup>4</sup>Glycosciences Laboratory, Faculty of Medicine, Imperial College London, Du Cane Rd, London W12 0NN, United Kingdom

### Email:

Daniel Bojar<sup>\*</sup> - daniel.bojar@gu.se

\* Corresponding author

### Keywords:

carbohydrate; glycan array; melon; plant lectin; R-type

*Beilstein J. Org. Chem.* **2024**, *20*, 306–320.

<https://doi.org/10.3762/bjoc.20.31>

Received: 01 December 2023

Accepted: 09 February 2024

Published: 19 February 2024

This article is part of the thematic issue "Chemical glycobiology".

Guest Editor: E. Fadda



© 2024 Lundstrøm et al.; licensee Beilstein-Institut.  
License and terms: see end of document.

## Abstract

Plant lectins have garnered attention for their roles as laboratory probes and potential therapeutics. Here, we report the discovery and characterization of *Cucumis melo* agglutinin (CMA1), a new R-type lectin from melon. Our findings reveal CMA1's unique glycan-binding profile, mechanistically explained by its 3D structure, augmenting our understanding of R-type lectins. We expressed CMA1 recombinantly and assessed its binding specificity using multiple glycan arrays, covering 1,046 unique sequences. This resulted in a complex binding profile, strongly preferring C2-substituted, beta-linked galactose (both GalNAc and Fuca1-2Gal), which we contrasted with the established R-type lectin *Ricinus communis* agglutinin 1 (RCA1). We also report binding of specific glycosaminoglycan subtypes and a general enhancement of binding by sulfation. Further validation using agglutination, thermal shift assays, and surface plasmon resonance confirmed and quantified this binding specificity in solution. Finally, we solved the high-resolution structure of the CMA1 N-terminal domain using X-ray crystallography, supporting our functional findings at the molecular level. Our study provides a comprehensive understanding of CMA1, laying the groundwork for further exploration of its biological and therapeutic potential.

## Introduction

Lectins have long been the subject of intense scientific scrutiny, serving as molecular bridges that span the realms of biochemistry, cellular biology, and biomedicine. These carbohydrate-

binding proteins boast a range of functions, acting as recognition modules in cell–molecule and cell–cell interactions, thereby playing vital roles in immune defense, regulation of

growth, and apoptosis [1]. In plants, they serve as essential components in development, immunity, and stress signaling [2,3].

In light of the burgeoning interest in the intersection of glycobiology and biomedicine, the characterization of new lectins has carved out a significant niche in scientific research. Specifically, lectins have emerged as invaluable tools for staining cells and tissues, thereby offering insights into cellular heterogeneity and function. For instance, the use of wheat germ agglutinin (WGA) and concanavalin A (ConA) has been instrumental in selectively staining cells based on their glycan expression [4], including single-cell approaches [5,6]. In the realm of therapeutics, lectins such as mistletoe lectins have shown promise in cancer therapy, by virtue of their ability to induce apoptosis in malignant cells [7]. Further, the creation of lectin arrays [8,9], which employ a diverse set of characterized lectins, has enabled high-throughput glycan profiling, thereby advancing both diagnostic methods and biomarker discovery. Examples include arrays that can rapidly profile alterations in glycosylation patterns, pivotal in many diseases and inflammatory changes [10,11].

Traditionally, lectins are divided into classes based on structural similarity and, by extension, common folds [12]. Still, shared binding specificity does not always follow from structural similarity, exemplified by divergent evolution within lectin families as well as independent emergence of similar binding patterns [13]. Many of the most commonly used lectins for the abovementioned applications are R-type lectins, especially those derived from plants. Examples include SNA (from *Sambucus nigra*, binding Neu5Aca2-6 [14]) or RCA1 (from *Ricinus communis*, binding terminal  $\beta$ -linked galactose [15]).

Yet, despite the extensive studies on plant lectins, particularly R-type lectins, there are still significant gaps in our understanding. Further, in general, few melon lectins have been studied in detail. Some reports indicate the presence of chitoooligosaccharide-binding (i.e.,  $\beta$ 1-4 GlcNAc oligomers) lectins from phloem exudates of melons [16,17], as well as R-type lectins in bitter melon [18], yet not much else is known about binding specificities exhibited by lectins derived from melons. In particular, existing research in this area often lacks a comprehensive characterization that includes both functional and structural analysis of these lectins.

Here, we introduce a novel member of characterized melon lectins, namely the *Cucumis melo* agglutinin (CMA1), an R-type lectin derived from melon. Prior to our study, CMA1 was only a predicted protein from genomic sequencing, with

moderate certainty scores on lectin-specific databases. Our comprehensive analysis using glycan array experiments, thermal shift assays, and high-resolution X-ray crystallography not only confirms its classification as a functional R-type lectin but also provides a deep dive into its unique glycan-binding profile and high-resolution 3D structure. Overall, we present a deeply characterized new lectin with a unique binding profile of specifically recognizing C2-substituted galactose in the context of glycans.

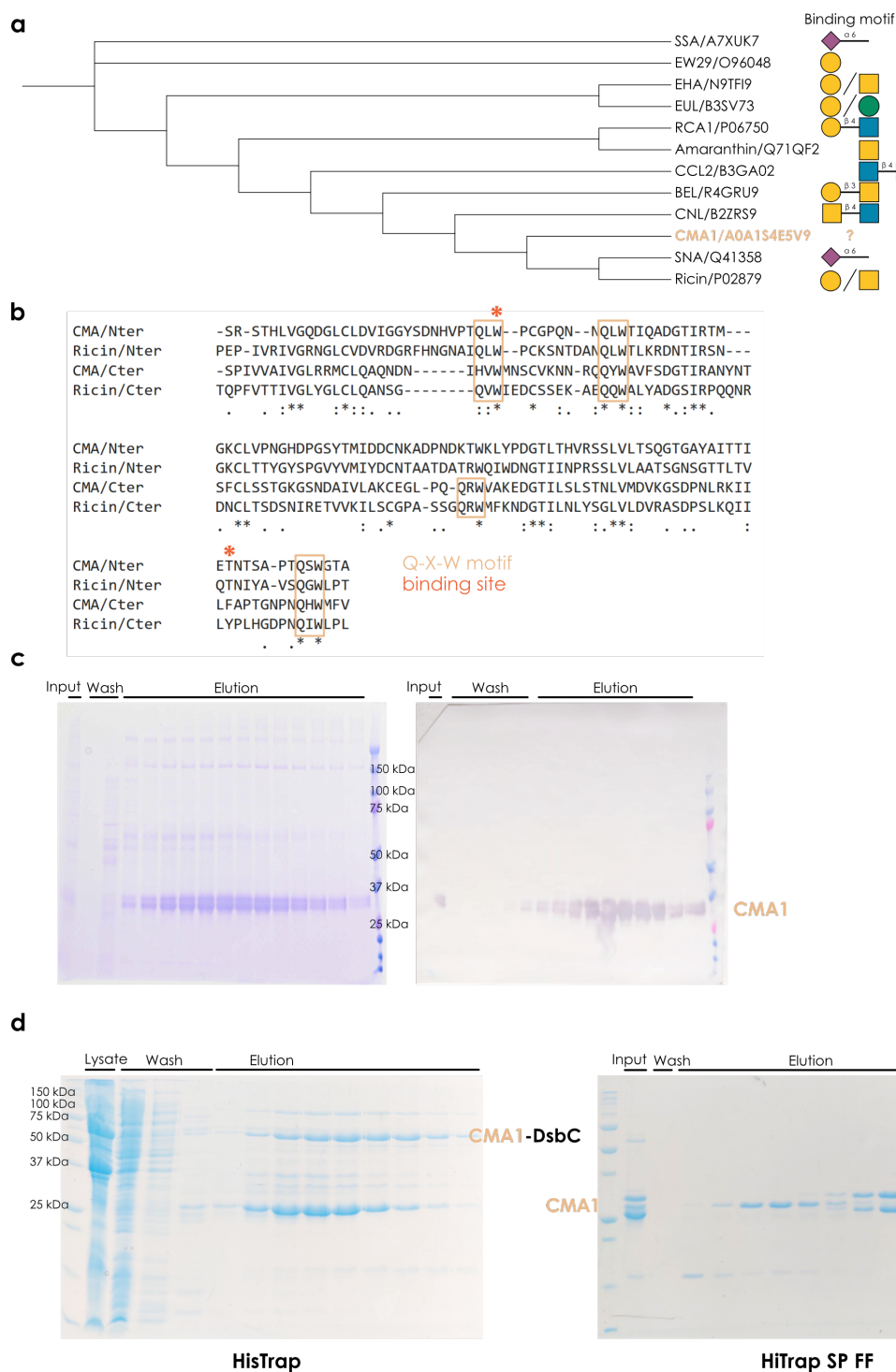
## Results and Discussion

### Identification and production of a new lectin from the melon *Cucumis melo*

CMA1 is a predicted protein from whole-genome shotgun sequencing of leaves from the melon plant *Cucumis melo* (variant *makuwa*, taxon ID: 1194695) [19] and has, to our knowledge, never been studied before. With prediction scores of 0.453 on LectomeXplore [12] and 0.251 on TrefLec [20] (from 0, lowest, to 1, highest), CMA1 is moderately certain in its prior classification as a lectin. CMA1 comprises 291 amino acids and is predicted to fold into two linked  $\beta$ -trefoil domains belonging to carbohydrate-binding module family 13 (CBM13) and placing it into the group of R-type lectins. Both CBM13 domains are likely to exhibit carbohydrate-binding activity due to the conservation of key amino acids in at least one of the three potential binding sites. In contrast to other R-type lectins such as ricin, it lacks a catalytic domain.

As R-type lectins are both a well-investigated family of lectins and widely used in research and beyond, we first wanted to analyze where CMA1 would be situated in the broader context of R-type lectins. A multiple sequence alignment of binding domains of representative R-type lectins (Figure 1a) showed that CMA1 exhibited a binding domain with a sequence relatively similar to those of the plant lectins SNA and ricin. However, we note that, in general, the substantial heterogeneity of binding motifs of even closely related lectins (SNA: Neu5Aca2-6, ricin: Gal/GalNAc) does not allow for a strong a priori hypothesis of what CMA1 would bind, even though R-type lectins in general are thought to prefer the Gal/GalNAc type motif mentioned in the context of ricin [21].

We next aligned the individual units of the tandem repeat CBM13 domains, indicated by the N-terminal (34-158) and C-terminal units (162-286) and compared those to the domains of ricin (Figure 1b). R-type lectins have a characteristic Q-x-W structural motif close to their binding site, which is highly conserved [21]. We report that CMA1 largely follows this trend, with three such binding sites in both N- and C-terminal domain, albeit with imperfect overlap. Based on the location of



**Figure 1:** Characterizing a new lectin from the melon *Cucumis melo*. (a) Evolutionary relationships of common R-type lectins. For a range of representative R-type lectins, we aligned their protein sequences via MUSCLE [22] and built a neighbor-joining tree with the resulting alignment distances, which is shown as a cladogram. For each protein, we only used the lectin domain, as annotated by UniProt or InterPro. For each protein, a representative binding specificity, based on literature reports, is provided. (b) Similarity of the two CBM13 domains in CMA1. Using MUSCLE to align the N-terminal (34–158) and C-terminal domains (162–286) of CMA1 and ricin (321–448 and 451–575), we indicated the position of the conserved Q-x-W motif in R-type lectins. (c) Recombinant expression of CMA1 in mammalian cells. SDS-PAGE and anti-His-tag Western blot of fractions from the expression of CMA1 protein in CHO-S cells. Note the smeared band indicating the presence of glycosylation. (d) Recombinant expression of CMA1 in bacteria. SDS-PAGE gels of the His-tag affinity chromatography and cation exchange chromatography from the expression of CMA1 protein in *E. coli* BL21\* cells.

the known binding pocket of the R-type lectin ricin and the respective sequence conservation in CMA1, we postulate binding sites around W<sup>63</sup> for the N-terminal domain and F<sup>273</sup> for the C-terminal domain of CMA1.

As binding specificities of melon lectins in general (beyond chitooligosaccharides), and CMA1 in particular, are still unknown, we set out to measure, quantify, and understand the glycan-binding properties of CMA1 in depth, as an archetypal example of melon lectins. For this, we needed to express the lectin recombinantly. As it is a secreted plant protein, we elected to express it in mammalian cell lines, to maximize the chances of a functional protein, because of post-translational modifications that would be lacking in bacteria as well as the oxidative environment of the secretory pathway, as CMA1 exhibits predicted disulfide bridges. A single step of His-tag affinity chromatography was sufficient to yield protein of adequate purity and good yield ( $\approx 15$  mg of eluted protein from 800 mL of cell culture, Figure 1c).

In parallel, we also expressed CMA1 in a bacterial expression system, which allowed us to ascertain whether binding was influenced by lectin glycosylation. The full-length mature protein (6–264) and individual N- or C-terminal domains were expressed using a N-terminal fusion comprising DsbC and a hexa-His tag, cleavable by TEV (*Tobacco etch virus*) protease. Despite the presence of the DsbC signal peptide, we did not observe periplasmic localization, and all proteins were instead purified from the cytoplasm. Ni-NTA affinity chromatography followed by TEV protease cleavage of the fusion construct and subsequent reverse Ni-NTA affinity chromatography resulted in significant co-purification of *E. coli* contaminants, necessitating an extra purification step, where cation exchange chromatography allowed us to obtain pure fractions of CMA1<sup>6–291</sup>. Of note, this additional purification step was not necessary for the purification of the CMA1 N-terminal domain (Figure 1d). Expression of the CMA1 C-terminal domain did not yield sufficiently pure and monodisperse protein for further biochemical and structural analyses.

### *Cucumis melo* agglutinin binds C2-substituted, beta-linked galactose

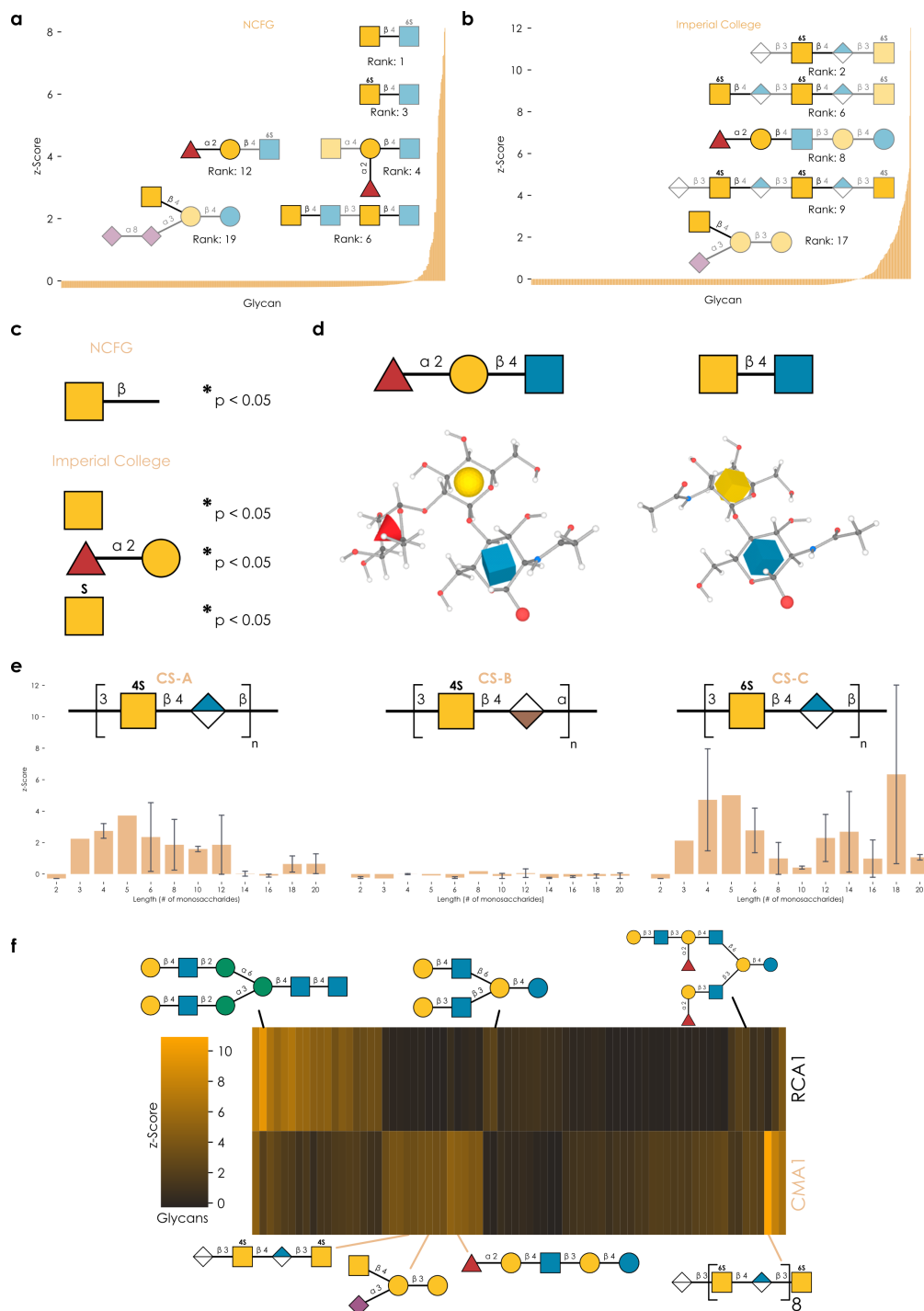
We then set out to answer the question whether CMA1 was a functional lectin and, if yes, what its binding specificity was. The standard approach to elucidate lectin binding specificity is via glycan array experiments. Here, tagged soluble lectin is added to, often, immobilized glycans and bound lectin is quantified via fluorescence scanners, which can be paired with glycan information due to the known arrangements of immobilized glycans on the plate. To cover the broadest possible sequence space, we tested our eukaryotically produced CMA1 protein

against the two largest glycan arrays at the National Center for Functional Glycomics (NCFG, Figure 2a) and the Glycosciences Laboratory at Imperial College London (ICL, Figure 2b). We note that, together, this encompasses 1,046 unique glycan sequences, spanning all major glycan classes and substantial taxonomic diversity. Next to these unique sequences, even more effects stem from a variety of linkers with which these molecules are immobilized.

In general, we observed two binding preferences that were strongly enriched among bound sequences, namely glycans containing Fuc $\alpha$ 1-2Gal epitopes and glycans containing terminal GalNAc residues (Figure 2c). Amongst the bound sequences, these substructures occurred in many different contexts, such as blood group H, LacdiNAc, or the Sd<sup>a</sup> motif, and particularly in sequences resembling O-glycans, milk oligosaccharides, and glycosphingolipids. At first glance, these two binding specificities may seem unconnected, indicating a rather broadly binding lectin. However, we noticed that the commonality of these two epitopes is hidden in the IUPAC-condensed nomenclature: Both substructures exhibited a bulky substituent on C2 of galactose, either a fucosyl (Fuc $\alpha$ 1-2Gal) or *N*-acetyl (GalNAc) moiety (Figure 2d). We thus conclude that CMA1 is highly specific for C2-substituted galactose. We further argue for a preference for a beta-linked epitope as, while we do observe binding to structures containing  $\alpha$ -linked GalNAc, the binding to their  $\beta$ -linked counterparts was generally stronger (e.g., GalNAc $\alpha$ : 1.57 vs GalNAc $\beta$ : 2.21, in z-scores (see Experimental section)). In part, this is reminiscent to the LacdiNAc binding specificity of *Clitocybe nebularis* lectin (CNL; Figure 1a) [27].

An important finding from the ICL array was that CMA1 exhibited robust binding to glycosaminoglycans (GAGs; Figure 2e; Supporting Information File 1, table “imperial”), in particular chondroitin sulfate (CS) C and A. Given the preference for terminal binding epitopes described above, the question naturally arose how the binding to these longer-chain glycans works. On the ICL array, CS sequences are typically capped with 4,5-unsaturated hexuronic acid derivatives on their non-reducing end and, thus, do not provide terminal GalNAc epitopes for binding. Further, while CMA1 did also bind to GalNAc-terminated GAGs (e.g., CSC-5, CSA-5), we measured higher binding to similar GAGs without the terminal GalNAc in several cases (Figure 2d,e). While some of the GAG probes varied in their immobilization amounts, we confirmed these results in a GAG-focused array (data not shown). We thus posit a binding to internal GalNAc epitopes for the case of GAG binding, potentially mediated by several binding sites.

This argument is strengthened by the observation that the highest observed binding to CSC and CSA was not with the



**Figure 2:** Characterizing the binding specificity of CMA1. (a, b) Lectin produced in mammalian cells was analyzed on the NCFG array (a) and the ICL array (b). Representative structures bound by CMA1 are shown via the “Symbol Nomenclature For Glycans” (SNFG), drawn with GlycoDraw [23]. Everything except the assigned binding motif is shown with added transparency. Full array data are available in Supporting Information File 1, tables “cfg” and “imperial”. (c) Enrichment analysis of glycan array data. For both NCFG and ICL array data, we used the *get\_pvals\_motif* function from glycowork [24] (version 0.8.1) with the keywords ‘terminal’ and ‘exhaustive’, to obtain significantly enriched motifs. \* $p < 0.05$ . (d) Common binding motif on the atomic level. Glycan 3D structures for the binding motifs were obtained from the GLYCAM web server [25,26]. (e) Binding of CMA1 to glycosaminoglycans. We grouped chondroitin sulfate (CS) types (A, B, and C) and plotted CMA1 binding against CS chain length. Shown are mean values with their 95% confidence interval. (f) Comparison of CMA1 and RCA1 binding. Glycans with a z-score of at least 0.5 in at least one lectin were retained and plotted as a hierarchically clustered heatmap via the *get\_heatmap* function of glycowork. Representative glycans are shown.

shortest sequences and required at least three repeats, with longer sequences such as CSC-18 even exhibiting the highest binding on the entire array (although we note that the longest GAG sequences were not generally the best binders, potentially hinting at steric clashes or density effects). Another supporting finding can be seen in the fact that CSB (exhibiting iduronic acid in  $\alpha$ -configuration, rather than its epimer, glucuronic acid, in  $\beta$ -configuration) showed virtually no binding to CMA1, further arguing for contacts of the GAG chain with the binding site. Lastly, we note that both CSC and CSA contain sulfated GalNAc, which, together with the observation of GalNAc6S $\beta$ 1-4GlcNAc as one of the highest binders on the NCFG array, leads us to speculate that sulfation further enhances CMA1 binding, a pattern that has been observed for several lectins [28].

Overall, this characterized binding specificity seemed distinct from other R-type lectins and we thus further compared it to a typical R-type lectin, *Ricinus communis* agglutinin (RCA1), on the ICL array. Canonically, RCA1 binds  $\beta$ -linked terminal galactose residues, which is generally what we also found in our array experiments, with Gal $\beta$  in various substructures and glycan types, particularly in those with multiple branches (Figure 2f). At best, the same sequences showed weak binding to CMA1, as they lacked a C2-substitution (Figure S1, Supporting Information File 2). Conversely, CMA1-favored sequences, containing Fuc $\alpha$ 1-2Gal or GalNAc epitopes, were on average not bound by RCA1 (the exception being sequences in which there was an additional free Gal $\beta$  terminus). Similarly, most chondroitin sulfate probes were not bound by RCA1. This gives rise to the conclusion that CMA1 does not merely tolerate but rather actively and strongly prefers C2-substituted Gal, while RCA1 does not even tolerate these substitutions. Interestingly, we also find that fucosylation of the GlcNAc residue (as in Lewis antigen motifs) completely abrogates CMA1 binding (Figure S1, Supporting Information File 2), despite the presence of Fuc $\alpha$ 1-2Gal, likely due to steric clashes in the binding pocket. We thus conclude that the binding profile of CMA1 is distinct from that of the typical R-type lectin RCA1 and unusual for a R-type lectin in general. We also note that the flexibility of accommodated C2 substituents (from *N*-acetyl moieties to whole monosaccharides), could make CMA1 an interesting candidate for probing synthetically produced glycans with novel substituents.

It is of course interesting to speculate about the physiological role of CMA1 in melons, yet this is hard to probe. It is noteworthy, however, that the glycan types in which its preferred binding motifs occur (*O*-glycans, milk glycans, GAGs) are absent from most plants, including melons. We thus hypothesize that the role of this lectin might be to recognize non-self

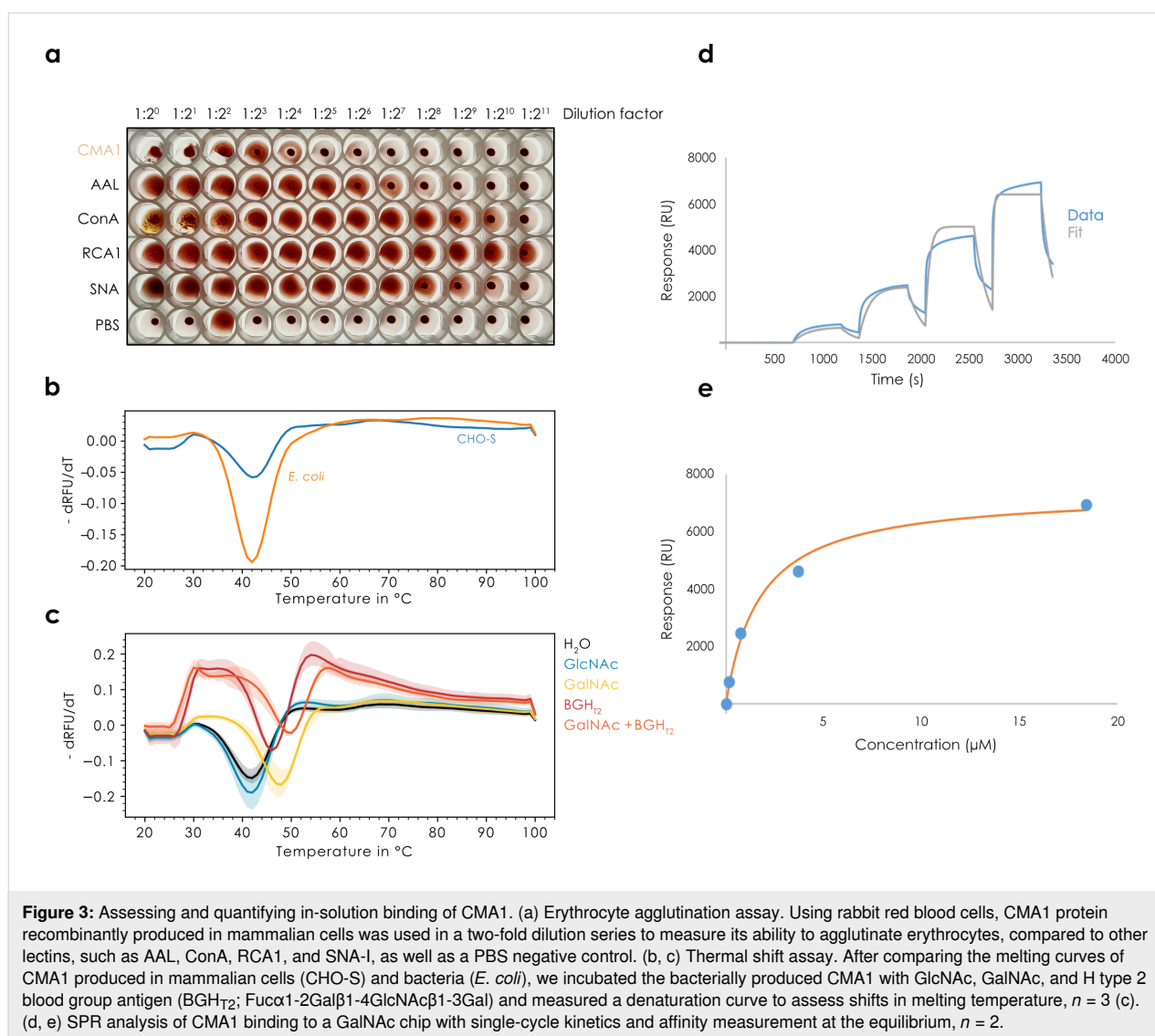
epitopes, such as for protection against pathogens, which is a common function in plant lectins [3].

## Validating binding in solution and assessing binding affinity

As CMA1 both exhibited multiple binding sites and robust binding to blood group epitopes (H-antigen), we hypothesized that it would be capable of agglutinating red blood cells, justifying its new name. When testing the protein recombinantly produced in mammalian cells, incubation with rabbit erythrocytes indeed resulted in moderate agglutination (Figure 3a), which also demonstrated the binding to these glycan substructures in a physiological context.

To further strengthen the case for CMA1 binding glycans in solution, and corroborate its binding specificity with orthogonal methods, we used a thermal shift assay. Herein, the binding of ligands is assessed by the stabilization of the protein, measured by a denaturation curve. Both the protein produced in mammalian and in bacterial cells exhibited similar melting temperatures here, of approximately 42 °C (Figure 3b). Then, we tested the binding of CMA1 to GlcNAc, GalNAc, and H type 2 blood group antigen (BGH<sub>T2</sub>; Fuc $\alpha$ 1-2Gal $\beta$ 1-4GlcNAc $\beta$ 1-3Gal; Figure 3c). This resulted in clear melting points shifts for both GalNAc and BGH<sub>T2</sub> to up to 50 °C, yet importantly not for GlcNAc, demonstrating both binding in solution and a further confirmation of the binding specificity obtained by the array experiments. We note that the functional activity of bacterially produced CMA1 indicates that potential modification by glycosylation is not required for ligand binding.

Next, we set out to quantify the binding affinity of CMA1 to its ligands. Lectins often only exhibit weak to moderate binding affinities, which is somewhat ameliorated by an increased avidity on the side of the lectin but also a dense presentation of the bound glycan epitope on the cell surface. We therefore used surface plasmon resonance (SPR) spectroscopy to derive binding constants for the interaction between CMA1 and GalNAc. A single cycle kinetics approach was applied, resulting in a measured  $K_D$  of  $1.66 \pm 0.08 \mu\text{M}$  (Figure 3d,e). Inhibiting binding of CMA1 to the GalNAc chip through a dilution series of *N*-acetyllactosamine (LacNAc) via multicycle kinetics allowed us to derive an  $\text{IC}_{50}$  of 1.4  $\mu\text{M}$  (Figure S2a,b; Supporting Information File 2). No inhibition was observed with chondroitin 6-sulfate tetrasaccharide (CSC), and only very weak inhibition for BGH<sub>T2</sub> but no  $\text{IC}_{50}$  could be determined as we could not increase the concentration to reach the plateau. For the recombinant CMA1-Nter, no binding could be observed on the GalNAc chip. This suggests either avidity effects in conjunction with the C-terminal domain or a high-affinity site on the C-terminal domain, giving rise to the measured  $K_D$  of the



full-length protein. Still, we were able to measure the affinity of CMA1-Nter to GalNAc in solution by isothermal calorimetry (ITC), obtaining a  $K_D$  of 940  $\mu$ M, confirming the low affinity (Figure S2c,d; Supporting Information File 2).

### Structural insights from the N-terminal domain of CMA1

Given the unusual binding specificity exhibited by CMA1, we were intrigued to elucidate the molecular mechanism that would enable the specific binding of C2-substituted galactose. The natural hypothesis here would be the creation of an additional pocket in the 3D structure of the binding site, accommodating the additional substituent at C2. However, as we observed little to no binding to unsubstituted galactose, we rather hypothesized the existence of specific interactions made with the C2-substituents, that did not exist in other R-type lectins such as RCA1. To determine this, we set out to resolve the detailed

three-dimensional structure of CMA1 via X-ray crystallography.

We obtained several hits for the full-length protein after sparse screening using a crystallization robot at the HTX platform, EMBL, Grenoble. Pill-shaped crystals obtained under conditions of a high salt concentration, in particular ammonium sulfate (Figure S3, Supporting Information File 2), did not give rise to any diffraction. Multiple layer plate or needles clusters were obtained in the presence of PEGs, but only showed weak diffraction ( $\approx 3.5$  Å). Finally, in the presence of 20% PEG 8K, 0.2 M MgCl<sub>2</sub>, and 0.1 M Tris HCl pH 8.5, single diamond-shaped crystals were obtained after 1–2 days for the N-terminal domain (Figure S3, Supporting Information File 2). High-resolution diffraction of the crystals allowed us to solve the CMA1-Nter structure in complex with LacNAc at 1.3 Å and GalNAc at 1.55 Å (see data and refinement statistics in Table 1). All



**Table 1:** Data collection and refinement statistics.

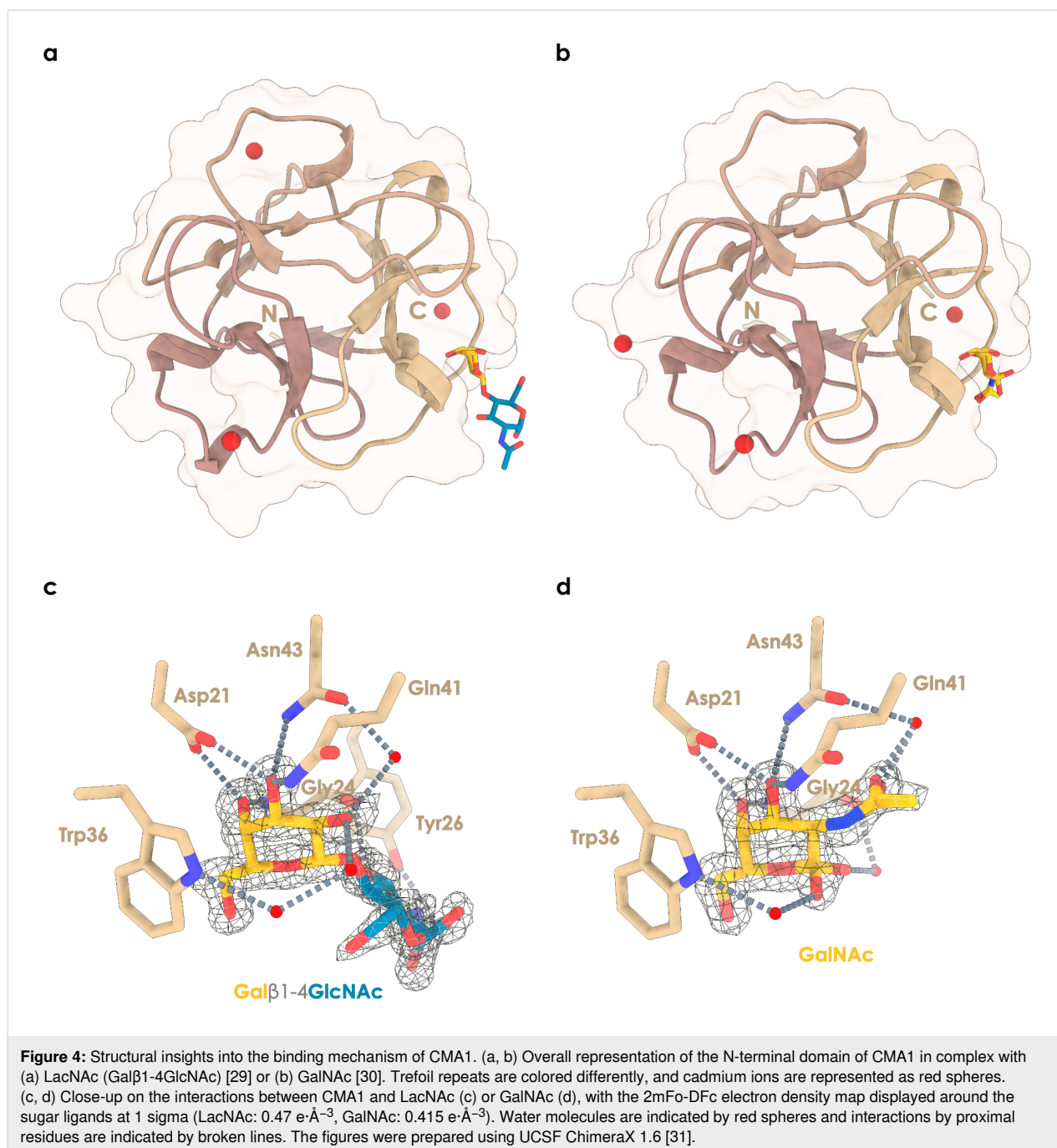
Complex	CMA1-Nter-LacNAc	CMA1-Nter-GalNAc
Data collection		
beamline	Soleil PX1	Soleil PX2
wavelength (Å)	0.97856	0.98011
space group	<i>I</i> 2	<i>I</i> 2
cell parameters <i>a</i> , <i>b</i> , <i>c</i> (Å)	36.70 36.78 94.79	36.61 36.86 94.81
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00 99.24 90.00	90.00 99.17 90.00
protein chains in a.u.	1	1
resolution (Å) <sup>a</sup>	46.78–1.32 (1.34–1.32)	35.68–1.55 (15.8–1.55)
CC1/2 (%) <sup>a</sup>	99.9 (96.9)	99.8 (85.7)
$R_{\text{merge}}$ (within I+/I-) <sup>a</sup>	0.055 (0.369)	0.052 (0.496)
$R_{\text{meas}}$ (within I+/I-) <sup>a</sup>	0.059 (0.400)	0.064 (0.618)
$R_{\text{pim}}$ (within I+/I-) <sup>a</sup>	0.022 (0.153)	0.037 (0.364)
mean I/ $\sigma$ (I) <sup>a</sup>	25.2 (5.7)	14.4 (2.9)
completeness (%) <sup>a</sup>	99.8 (96.0)	99.7 (99.9)
number reflections <sup>a</sup>	399970 (18410)	95115 (4581)
number of unique reflections <sup>a</sup>	29695 (1434)	18279 (911)
multiplicity <sup>a</sup>	13.5 (12.8)	5.2 (5.0)
Wilson <i>B</i> -factor (Å <sup>2</sup> )	14.1	19
Refinement		
resolution (Å)	46.78–1.32	35.69–1.55
no. reflections/no. free reflections	28192/1503	17373/905
$R_{\text{work}}/R_{\text{free}}$ (%)	14.35/18.58	16.3/20.4
R.m.s. bond lengths (Å)	0.0130	0.0127
Rmsd bond angles (°)	1.721	1.893
Rmsd chiral (Å <sup>3</sup> )	0.097	0.092
no. atoms / Bfac (Å <sup>2</sup> )		
protein	1029/15.1	985/19.95
ligand	26/20.3	30/22.3
cadmium	3/21.9	3/27.0
water	248/28.7	176/31.8
Ramachandran allowed (%)	100	100
favored (%)	99	100
outliers	0	0

<sup>a</sup>Values in parenthesis refer to the highest-resolution shell.

residues of the N-terminal construct (Val<sup>6</sup> to Asp<sup>132</sup>) could be modelled, and unambiguous electron density permitted us to locate and model four cation binding sites (three in each structure) and one sugar binding site (Figure 4a,b and Figure S4, Supporting Information File 2).

The complexed structures allowed us to shed light on the arrangement of the ligand in the binding site (Figure 4c,d). While lectins such as CMA1 typically can present three binding pockets in their CBM13 domain, we hypothesized that the N-terminal half of CMA1 would in fact only exhibit two func-

tional binding sites. However, only the alpha site was found occupied with a carbohydrate here. It is found in a shallow groove, supporting our data on the lack of a distinct distal binding specificity. We report a tight coordination of the O3 and O4 hydroxy groups of the galactose residue involving Asp<sup>21</sup>, Asn<sup>43</sup>, and Gln<sup>41</sup> side chains, as well as the Gly<sup>24</sup> main chain nitrogen. CH- $\pi$  stacking and hydrophobic interactions occur between the aromatic ring of Trp<sup>36</sup> and the alpha face of the ring as well as the hydroxymethyl moiety of the galactose residue, additionally ensuring specificity for galactoside over glucoside as an equatorial conformation of the O4 hydroxy



group would lead to steric clashes and loss of strong hydrogen bonding.

In the LacNAc-complexed structure (PDB ID 8R8A) [29], the GlcNAc residue did not seem to engage in extensive interactions, with only a hydrogen bond between the *N*-acetyl moiety and the main chain oxygen of Gly<sup>24</sup> and hydrophobic interaction with the aromatic ring of Tyr<sup>26</sup> (Figure 4c). Further, beyond the C2 position of galactose, a cavity filled with coordinated water molecules hinted at the binding mode for C2-substi-

tuted galactose. Notably, the seemingly inactive beta site was found to be occupied by a cadmium ion (Figure S4, Supporting Information File 2), supporting our ITC and SPR data where no multivalent binding effects were observed for the single-domain N-terminal construct.

In the GalNAc-complexed structure (PDB ID 8R8C) [30], the *N*-acetyl group of GalNAc extended beyond C2 into the cavity noted in the LacNAc complex. While no direct interactions with the protein backbone were observed, we found one water mole-

cule to mediate hydrogen bonding between the oxygen of the *N*-acetyl group and the Asn<sup>43</sup> side chain oxygen (Figure 4d). Both GalNAc anomers could be observed, showing interactions through water molecule coordination with the Trp<sup>36</sup> ring nitrogen (alpha anomer) or the Gly<sup>24</sup> main chain oxygen (beta anomer).

## Conclusion

Our work presents a substantial exploration of the binding specificity and mechanism of the hitherto uncharacterized lectin CMA1 from melons. The binding specificity of CMA1, C2-substituted galactose that is preferentially presented in a  $\beta$ -configuration, enables it to bind to a range of biologically relevant epitopes, such as LacdiNAc, Sd<sup>a</sup>, blood group H, and chondroitin sulfate motifs. Further, the inhibition of binding by the presence of Lewis antigen motifs additionally narrows its binding specificity. Our binding data and structural information lead us to the conclusion that crucially positioned asparagine residues facilitate this unusual binding specificity that delineates CMA1 from typical R-type lectins such as RCA1. Together, these results advance our knowledge of R-type lectins in general and the range of their binding specificities, but also our knowledge of melon lectins in particular, which has remained limited so far. Further experiments are still required to determine the role of the C-terminal domain, as well as the physiological function of the full-length CMA1 protein.

## Experimental

### Recombinant protein expression

For mammalian expression, the gene for CMA1 (A0A1S4E5V9) was synthesized with human-optimized codons and a C-terminal hexa-His tag (GSHHHHHH). We then cloned this gene into a pCI backbone (U47119; Promega GmbH) for expression in mammalian cells under a constitutive cytomegalovirus (CMV) promoter. Then, the Mammalian Protein Expression core facility at the University of Gothenburg transfected this plasmid into FreeStyle™ CHO-S cells (Cat nr R80007, ThermoFisher Scientific). Cells were cultured in Freestyle™ CHO medium at 37 °C in 5% CO<sub>2</sub> in Optimum Growth™ flasks (Thomson instrument company) at 130 rpm in a Multitron 4 incubator (Infors) and transfected at  $2 \times 10^6$  cells/mL using FectoPro transfection reagent (Polyplus). Protein-containing culture supernatant (0.8 L) was harvested after 120 h, filtered using Polydisc AS 0.45  $\mu$ m (Whatman, Cytiva) and loaded onto a 5 mL HisExcel column (GE healthcare) at 5 mL/min. The column was washed with 10 mM phosphate-buffered saline (Medicago), 500 mM NaCl and 50 mM imidazole before elution of the protein using the same buffer with a gradient from 50 mM to 500 mM imidazole (G-Biosciences) over 15 column volumes. Pooled fractions were concentrated using Vivaspin concentrators (MWCO 10 kDa, Sartorius

Stedim), passed over a HiPrep 26/10 desalting column (GE Healthcare) in phosphate-buffered saline (Medicago), and finally concentrated again.

For bacterial expression, the gene of CMA1 (33–291, corresponding to residues 6–291 of the mature protein) with optimized codons for *Escherichia coli* was synthesized flanked by *Nco*I and *Xho*I restriction sites where L<sup>6</sup> was mutated to valine. The gene was inserted in the homemade plasmid pET40b-TEV where the enterokinase cleaving site was replaced by a TEV cleavage site by site directed mutagenesis. This plasmid was obtained by PCR using pET-40b(+) (Novagen, Merck, #70091) as template and the following primers: forward (gccagatctgggtac\_cGAAAACCTGTATTTTCAGGGCGccatggcgatcg) and reverse (GGTACCCAGATCTGGGCTGTCCATGTGCTGGC) with complementary sequence underlined. PCR was performed using PrimeSTAR DNA polymerase (Takara #TAKR045A); then the product was digested by *Dpn*I and finally transformed in NEB5 $\alpha$  strain (New England Biolabs, #C2992H). Both gene and vector were digested by *Nco*I and *Xho*I restriction enzymes (New England Biolabs) prior to purification on agarose gel using Monarch Gel extraction kit and supplier instructions (New England Biolabs, #T1020S) and ligation using the DNA ligation kit, Mighty Mix (Ozyme, Takara, #TAK6023Z), at room temperature to form the pET40b-TEV-CMA11 plasmid.

The N-terminal domain of CMA1 (6–132 in mature protein) was amplified by PCR using the following primers: forward (ACGCCATGGTGAGCCGTTCTACGC) and reverse (ATATCTCGAGTTAATCTG CCGTACCCCAGGATTGTGTAGG) and pET40b-TEV-CMA1 plasmid as template. Similarly, the C-terminal domain of CMA1 (136–264 in mature protein) was amplified by PCR using the subsequent primers: forward (ATTCCATGGGTCCGATTGTGGTTGC-CATTGTTGG) and reverse (ACACCTCGAGTTAGGTTGTACTGTGTACGAACATCC). The primers contained the restriction sites (underlined) *Nco*I (sense) and *Xho*I (antisense) on their 5'-ends for further sub-cloning. PCR was performed using PrimeSTAR DNA polymerase. The purified PCR fragment of 395 bp was digested by *Nco*I and *Xho*I restriction enzymes, then ligated into pET40b-TEV vector, and finally transformed in NEB5 $\alpha$  strain to form the pET40b-TEV-CMA1-Nter and pET40b-TEV-CMA1-Cter plasmids. All plasmids and new vectors were verified by sequencing (Eurofins Genomics, Ebersberg, Germany). Primers were purchased from Eurofins Genomics (Ebersberg, Germany).

*E. coli* BL21\*(DE3) [Invitrogen, #C601003] cells were transformed by heat shock at 42 °C with pET40b-TEV-CMA1 and Tuner(DE3) [Novagen, #70623] cells with pET40b-TEV-CMA1Nter prior pre-culturing in lysogeny broth (LB) [Invit-

rogen, #12780052] media containing 25 µg/mL kanamycin [Euromedex, #UK0015-A] at 37 °C, 180 rpm overnight. Then, 1 L LB medium supplemented with 25 µg/mL kanamycin was inoculated with 25 mL of the pre-culture and incubated at 37 °C, 180 rpm. When OD<sub>600nm</sub> reached 0.4, the temperature was lowered to 16 °C, and when OD<sub>600nm</sub> reached 0.8, protein expression was induced by the addition of 0.1 mM isopropyl β-D-thiogalactoside (IPTG) [Euromedex, #EU0008-C]. After 20 h, the cells were harvested by centrifugation at 5,000g for 10 min at 4 °C.

For purification of bacterial recombinant CMA1, each gram of cell pellet was resuspended with 5 mL of buffer A (20 mM Tris-HCl pH 7.5, 500 mM NaCl). After addition of 1 µL of Denarase® (C-LEcta GmbH, #20804) and moderate agitation on a rotating wheel for a period of 30 min at room temperature, cells were lysed using a cell disruptor (Constant Systems Ltd, UK) under a pressure of 2.5 kbar. The lysate was cleared by centrifugation at 24,000g for 30 min at 4 °C and passed through a 0.45 µm syringe filter prior to affinity chromatography purification using 1 mL HisTrap™ HP column (Cytiva) preequilibrated with buffer A and an NGC chromatography system (Bio-Rad). After loading the cleared lysate, the column was washed with buffer A + 50 mM imidazole (Sigma-Aldrich, Merck, #56749) to remove all contaminants and unbound proteins. CMA1 was eluted by a 20 mL linear gradient from 50 mM to 500 mM imidazole in buffer A. The fractions were analyzed by SDS-PAGE with 15% gel and those containing CMA1 were collected and deprived of imidazole by buffer exchange in buffer A using a Macro and Microsep Advance Spin 3 kDa MWCO centrifugal filter (Pall). The N-terminal His-tag was removed by TEV cleavage in the presence of 1 mM EDTA (Euromedex, #EU0084.B) overnight at 10 °C, using a TEV/CMA1 ratio of 1:50. TEV was prepared in-house. The protein mixture was then purified on a 1 mL HisTrap column, where pure CMA1 protein was collected in the flowthrough and column wash. Full-length CMA1 (6-291) was purified from remaining *E. coli* contaminants using a 1 mL HiTrap™ SP Sepharose FF column (Cytiva) preequilibrated with 50 mM sodium acetate pH 5.5. After loading, the column was washed, and CMA1 was eluted by a 20 mL linear gradient from 0 to 700 mM NaCl in 50 mM sodium acetate pH 5.5. The protein was concentrated and the buffer exchange to 20 mM HEPES pH 8, 100 mM NaCl using a 3 kDa MWCO centrifugal filter and stored at 4 °C.

For CMA1-Nter, the same protocol was followed, with the following changes: Purification was carried out by exploiting gravity using 1 mL of Ni Sepharose High Performance resin (Cytiva, #17.5268.01) and an Econo-Pac® Chromatography Column (Bio-Rad, #7321010). Buffer A was exchanged to

buffer B (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 500 mM urea, and 5 mM imidazole). Washing steps were performed using buffer B and buffer B containing 50 mM imidazole. Elution was performed using buffer B plus 250 mM imidazole. The buffer was exchanged with 20 mM HEPES pH 7.5, 100 mM NaCl by three times 10× dilution and the sample was concentrated to at least 1 mg/mL using a 3 kDa MWCO centrifugal filter prior to TEV cleavage.

## Glycan array experiments

### NCFG array

For the NCFG array, data was collected by the National Center for Functional Glycomics (NCFG) at Beth Israel Deaconess Medical Center, Harvard Medical School. For experiments, a standard binding buffer (20 mM Tris-HCl pH 7.4, 150 mM NaCl, 2 mM CaCl<sub>2</sub>, 2 mM MgCl<sub>2</sub>, 0.05% Tween 20, 1% BSA) was used. CMA1 binding was probed by incubation with a penta-His-488 antibody (5 µg/mL). CMA1 was tested in two concentrations (5 and 50 µg/mL) on Version 5.4 of the printed CFG array, consisting of 585 printed glycans in replicates of six. Results from replicates were combined as average RFU (raw fluorescence unit). For this average, the highest and lowest value was removed for each glycan, mitigating the effects of outliers. The results can be found in Supporting Information File 1, table “cfg”.

### ICL array

For experiments, a standard binding buffer (10 mM HEPES, 150 mM NaCl, 1% BSA, 0.02% casein blocker (Pierce), 5 mM CaCl<sub>2</sub>) was used. CMA1 was tested at 100 µg/mL for 1 h on the broad spectrum screening array (in house designation ‘Array Sets 42–56’) of the Glycoscience Laboratory at Imperial College London, consisting of 866 lipid-linked glycans. Then the detecting solution composed of anti-polyHistidine (Sigma-Aldrich, Merck, SAB4200620) and biotin anti-mouse IgG (Sigma-Aldrich, Merck, B7264) antibodies (10 µg/mL, precomplexed in a ratio of 1:1) was overlaid onto the arrays for 1 h. The final detection was with a 30 min overlay of streptavidin-Alexa Fluor 647 (Molecular Probes) at 1 µg/mL. The microarray slides were scanned with GenePix 4300A scanner instrument (50% laser power at PMT 350), and the image analysis (quantitation) was performed with GenePix® Pro 7 software. The results can be found in Supporting Information File 1, table “imperial” and “rca\_imperial”, with the array generation in Supporting Information File 3 according to the MIRAGE guidelines (Minimum Information Required for A Glycomics Experiment) [32].

For both array types, data were transformed into z-scores by subtracting the mean value across the array and dividing the results by the standard deviation.

## Agglutination assay

The hemagglutinating activity of CMA1 was determined in V-bottom 96-well plates by a twofold serial dilution procedure in PBS using rabbit red blood cells (Atlantis France). 25  $\mu$ L of 4% erythrocyte suspension was added to an equal volume of the sample, and the mixture was incubated for 60 min at room temperature. Starting concentrations were: CMA1 0.6 mg/mL, AAL 0.5 mg/mL, ConA 2.5 mg/mL, RCA1 2.5 mg/mL, and SNA 0.5 mg/mL.

## Thermal shift assay

Thermal shift assays were performed using a Mini Opticon Real Time PCR machine (BioRad). 0.6 mg/mL protein in PBS was mixed with SYPRO Orange (Sigma-Aldrich, Merck, #S5692) and glycan ligand (10 mM GalNAc; Carbosynth, #MA04390; 10 mM GlcNAc, Carbosynth, #MA00834; 10 mM blood group H type-2 tetrasaccharide; Elicityl, GLY032-2) in a total reaction volume of 25  $\mu$ L. The temperature was raised by 1  $^{\circ}$ C/min from 25 to 100  $^{\circ}$ C, and fluorescence readings were taken at each step.

## Surface plasmon resonance spectroscopy

Experiments were performed using a Biacore X100 instrument (Cytiva) at 25  $^{\circ}$ C in HBS-T running buffer (10 mM HEPES pH 7.4, 150 mM NaCl and 0.05% Tween 20). Biotinylated PAA-GalNAc (Lectinity, GlycoNZ, #0031-BP) was immobilized on CM5 chips (Cytiva #BR100012) that were coated previously with streptavidin (Sigma-Aldrich, Merck, #S4762), following standard protocol. Biotinylated GalNAc was diluted to 2  $\mu$ g/mL in HBS-T before being injected into one of the flow cells of the chip. An immobilization level of 900 response units (RU) was obtained. A reference surface was always present in flow cell 1, allowing for the subtraction of bulk effects and non-specific interactions with streptavidin. The mammalian-produced CMA1 was injected in single cycle kinetic over the flow cell surface at 10  $\mu$ L/min at increasing concentrations with a contact time of 500 s. Dissociation was achieved by passing running buffer for 2 min. Surfaces were regenerated with four consecutive 30 s injections of 50 mM NaOH and 1 M NaCl. Binding affinity ( $K_D$ ) was measured after subtracting the channel 1 reference (streptavidin only) and subtracting a blank injection (running buffer – zero analyte concentration). Data evaluation and curve fitting was performed using the provided BIACORE X100 evaluation software (version 2.0). Measurements were at least done in duplicate.

Then, to perform competition experiments, nine concentrations of LacNAc (Elicityl, #GLY008) from 10 to 0 mM with a dilution coefficient of two supplemented with a fixed concentration of 0.8  $\mu$ M was injected into the cell surface in multiple cycle kinetic with an association time of 500 s and a dissociation time

of 12 s at a flow rate of 10  $\mu$ L/min. Surfaces were regenerated with 30 s injections of 50 mM NaOH and 1 M NaCl.  $IC_{50}$  was measured using the response at equilibrium for each concentration of competitive sugar that were translated in percentage of inhibition, then plotted against the molar concentration of competitive sugar using the free software “data entry”. The  $IC_{50}$  was calculated using <https://www.aatbio.com/tools/ic50-calculator>.

## X-ray crystallography

All consumables for crystallization and crystal handling were purchased at Molecular Dimensions, Calibre Scientific, Rotherham, UK, unless stated otherwise. CMA1 concentrated at 5.7 or 3.5 mg/mL in 20 mM HEPES pH 8, 100 mM NaCl, and 14 mM GalNAc was subjected to crystallization screening using the robotized HTXlab platform (EMBL, Grenoble, France) with 200 nL sitting drops at 20  $^{\circ}$ C using a 1:1 ratio. Wizard I and II screen (Rigaku) and SaltRX (Hampton Research) screens were used and led to more than 30 hits after one to three days. Pill-like crystals were obtained with high salt concentration that could be reproduced by hand in the laboratory. Plates and needles clusters were obtained with PEG containing solutions. For CMA1-Nter, protein at a concentration of 2.9–3.5 mg/mL was crystallized using hanging drop and vapor diffusion methods with a 2  $\mu$ L drop in 1:1 ratio at 20  $^{\circ}$ C. Bipyramidal single crystals were obtained after one or two days in a solution containing 10–12% PEG Smear Medium, 0.1 M MES pH 6.5, 1 $\times$  divalent (5 mM of CaCl<sub>2</sub>, MgCl<sub>2</sub>, CsCl<sub>2</sub>, CdCl<sub>2</sub>, NiCl<sub>2</sub>, and zinc acetate), or 5 mM CdCl<sub>2</sub>, and in the presence or not of 5 mM GalNAc. Cocrystals of CMA1-Nter in complex with LacNAc (Gal $\beta$ 1-4GlcNAc, Elicityl, #GLY008) were obtained by the addition of 5 mM LacNAc to the protein solution and incubation at room temperature for 30 min prior to crystallization. For both complexes, single crystals were mounted in a cryoloop after transfer in a cryoprotectant solution, composed of 30% PEG Smear Medium and 5 mM CdCl<sub>2</sub>, and flash-cooled in liquid nitrogen. Crystal diffraction was evaluated, and data were collected on the Proxima 1 and 2 beamlines at the synchrotron SOLEIL, Saint Aubin, France using an Eiger 16M or 9M detector (Table 1) for LacNAc and GalNAc complexed structures, respectively. XDS and XDSME were used to process the data and all further steps were performed using programs of the CCP4 suite version 8.25–27 [33–35]. The model coordinates predicted by AlphaFold [36] Monomer v2.0 for the monomer of CMA1 (A0A1S4E5V9) were trimmed to only include the N-terminal domain (residues 33–159), with all B-factors reset to 15  $\text{Å}^2$ , to be subsequently used as a search model to solve the structure of CMA1-Nter by molecular replacement using PHASER [37]. Multiple iterations of anisotropic restrained maximum likelihood refinement using REFMAC 5.8 [38] and manual building using Coot [39] were performed.

Hydrogen atoms were added in their riding positions during refinement and 5% of the observations were set aside for cross-validation analysis. Upon inspection of the electron density maps, carbohydrate moieties were introduced and checked using Privateer [40]. The final model was validated using the wwPDB validation server (<https://validate-rcsb-1.wwpdb.org>). Structure figures were made using PyMol 2.5.7 and ChimeraX 1.6 [31]. The parameters for CH– $\pi$  interactions were defined as previously reported [41,42].

## Supporting Information

### Supporting Information File 1

Full array data regarding the binding specificity of CMA1.

[<https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-31-S1.xlsx>]

### Supporting Information File 2

Additional figures.

[<https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-31-S2.pdf>]

### Supporting Information File 3

Supplementary glycan microarray document (MIRAGE) for the ICL glycan arrays.

[<https://www.beilstein-journals.org/bjoc/content/supplementary/1860-5397-20-31-S3.pdf>]

## Acknowledgements

We acknowledge support from the Mammalian Protein Expression core facility at the University of Gothenburg via the Protein Production Sweden (PPS) framework as well as the synchrotron SOLEIL (Saint Aubin, France) for access to beamline Proxima 1 and 2 (Proposal Number 20210859) and for the technical support of Pierre Legrand and Martin Savko, respectively. The authors would like to thank Iris Lopez and Federico Musso for their technical help in the expression assays of CMA1-Nter and CMA1 purification trials, respectively, and Wengang Chai for preparing the GAG probes on the ICL array.

## Funding

This work was funded by a Branco Weiss Fellowship – Society in Science awarded to D.B., by the Knut and Alice Wallenberg Foundation, and the University of Gothenburg, Sweden as well as support from the GLYCONanoPROBES (CA18132) and INNOGLY (CA18103) COST actions awarded to J.L. This work was further supported by the Protein-Glycan Interaction Resource of the CFG and the National Center for Functional Glycomics (NCFG) at Beth Israel Deaconess Medical Center,

Harvard Medical School (supporting grant R24 GM137763). The glycan microarray studies were performed in the Carbohydrate Microarray Facility at the ICL Glycosciences Laboratory, which is supported by Wellcome Trust biomedical resource grants (099197/Z/12/Z, 108430/Z/15/Z, and 218304/Z/19/Z) and partially by the March of Dimes Prematurity research centre grant (22-FY18-82). The sequence-defined glycan microarrays contain many saccharides provided by collaborators whom we thank, as well as members of the Glycosciences Laboratory for their contribution in the establishment of the NGL-based microarray system. This work benefited from access to EMBL HTX lab, which has been supported by iNEXT-Discovery, project number 871037, funded by the Horizon 2020 program of the European Commission.

## ORCID® iDs

Nicole Kerekes - <https://orcid.org/0000-0001-7065-8092>

Antonio Di Maio - <https://orcid.org/0000-0002-2740-9098>

Ten Feizi - <https://orcid.org/0000-0001-6495-0329>

Annabelle Varrot - <https://orcid.org/0000-0001-6667-8162>

Daniel Bojar - <https://orcid.org/0000-0002-3008-7851>

## Data Availability Statement

All generated data here can be found in the Supporting Information. The coordinates of CMA1 in complex with LacNAc (PDB ID 8R8A), <https://doi.org/10.2210/pdb8R8A/pdb>, and GalNAc (PDB ID 8R8C) <https://doi.org/10.2210/pdb8R8C/pdb>, have been deposited in the Protein Data Bank (PDB).

## Preprint

A non-peer-reviewed version of this article has been previously published as a preprint: <https://doi.org/10.1101/2023.11.30.569503>

## References

- Sharon, N. *Glycobiology* **2004**, *14*, 53R–62R. doi:10.1093/glycob/cwh122
- Damme, E. J. M. V.; Peumans, W. J.; Barre, A.; Rougé, P. *Crit. Rev. Plant Sci.* **1998**, *17*, 575–692. doi:10.1080/07352689891304276
- Lannoo, N.; Van Damme, E. J. M. *Front. Plant Sci.* **2014**, *5*, 397. doi:10.3389/fpls.2014.00397
- Keller, L.-A.; Niedermeier, S.; Claassen, L.; Popp, A. *Acta Histochem.* **2022**, *124*, 151877. doi:10.1016/j.acthis.2022.151877
- Kearney, C. J.; Vervoort, S. J.; Ramsbottom, K. M.; Todorovski, I.; Lelliott, E. J.; Zethoven, M.; Pijpers, L.; Martin, B. P.; Semple, T.; Martelotto, L.; Trapani, J. A.; Parish, I. A.; Scott, N. E.; Oliaro, J.; Johnstone, R. W. *Sci. Adv.* **2021**, *7*, eabe3610. doi:10.1126/sciadv.abe3610
- Minoshima, F.; Ozaki, H.; Odaka, H.; Tateno, H. *iScience* **2021**, *24*, 102882. doi:10.1016/j.isci.2021.102882
- Choi, S. H.; Lyu, S. Y.; Park, W. B. *Arch. Pharmacol. Res.* **2004**, *27*, 68. doi:10.1007/bf02980049
- Hirabayashi, J.; Yamada, M.; Kuno, A.; Tateno, H. *Chem. Soc. Rev.* **2013**, *42*, 4443. doi:10.1039/c3cs35419a

9. Pilobello, K. T.; Slawek, D. E.; Mahal, L. K. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 11534–11539. doi:10.1073/pnas.0704954104
10. Qin, R.; Meng, G.; Pushalkar, S.; Carlock, M. A.; Ross, T. M.; Vogel, C.; Mahal, L. K. *J. Proteome Res.* **2022**, *21*, 1974–1985. doi:10.1021/acs.jproteome.2c00251
11. Heindel, D. W.; Chen, S.; Aziz, P. V.; Chung, J. Y.; Marth, J. D.; Mahal, L. K. *ACS Infect. Dis.* **2022**, *8*, 1075–1085. doi:10.1021/acsinfectdis.2c00082
12. Bonnardel, F.; Mariethoz, J.; Pérez, S.; Imberty, A.; Lisacek, F. *Nucleic Acids Res.* **2021**, *49*, D1548–D1554. doi:10.1093/nar/gkaa1019
13. Taylor, M. E.; Drickamer, K. *Curr. Opin. Struct. Biol.* **2014**, *28*, 14–22. doi:10.1016/j.sbi.2014.07.003
14. Bojar, D.; Meche, L.; Meng, G.; Eng, W.; Smith, D. F.; Cummings, R. D.; Mahal, L. K. *ACS Chem. Biol.* **2022**, *17*, 2993–3012. doi:10.1021/acscchembio.1c00689
15. Wu, A. M.; Wu, J. H.; Singh, T.; Lai, L.-J.; Yang, Z.; Herp, A. *Mol. Immunol.* **2006**, *43*, 1700–1715. doi:10.1016/j.molimm.2005.09.008
16. Swamy, M. J.; Bobbili, K. B.; Mondal, S.; Narahari, A.; Datta, D. *Phytochemistry* **2022**, *201*, 113251. doi:10.1016/j.phytochem.2022.113251
17. Allen, A. K. *Biochem. J.* **1979**, *183*, 133–137. doi:10.1042/bj1830133
18. Wang, H.; Ng, T. B. *Biochem. Biophys. Res. Commun.* **1998**, *253*, 143–146. doi:10.1006/bbrc.1998.9765
19. Shin, A.-Y.; Koo, N.; Kim, S.; Sim, Y. M.; Choi, D.; Kim, Y.-M.; Kwon, S.-Y. *Sci. Data* **2019**, *6*, 220. doi:10.1038/s41597-019-0244-x
20. Notova, S.; Bonnardel, F.; Rosato, F.; Siukstaite, L.; Schwaiger, J.; Lim, J. H.; Bovin, N.; Varrot, A.; Ogawa, Y.; Römer, W.; Lisacek, F.; Imberty, A. *Commun. Biol.* **2022**, *5*, 954. doi:10.1038/s42003-022-03869-w
21. Cummings, R. D.; Schnaar, R. L.; Ozeki, Y. R-Type Lectins. In *Essentials of Glycobiology*; Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Mohnen, D.; Kinoshita, T.; Packer, N. H.; Prestegard, J. J.; Schnaar, R. L.; Seeberger, P. H., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2022.
22. Edgar, R. C. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. doi:10.1093/nar/gkh340
23. Lundström, J.; Urban, J.; Thomès, L.; Bojar, D. *Glycobiology* **2023**, *33*, 927–934. doi:10.1093/glycob/cwad063
24. Thomès, L.; Burkholz, R.; Bojar, D. *Glycobiology* **2021**, *31*, 1240–1244. doi:10.1093/glycob/cwab067
25. GLYCAM-Web | Utilities for molecular modeling of carbohydrates. <https://glycam.org/> (accessed Feb 8, 2024).
26. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–655. doi:10.1002/jcc.20820
27. Pohleven, J.; Obermajer, N.; Sabotič, J.; Anžlovar, S.; Sepčić, K.; Kos, J.; Kralj, B.; Štrukelj, B.; Brzin, J. *Biochim. Biophys. Acta, Gen. Subj.* **2009**, *1790*, 173–181. doi:10.1016/j.bbagen.2008.11.006
28. Jung, J.; Enterina, J. R.; Bui, D. T.; Mozaneh, F.; Lin, P.-H.; Nitin; Kuo, C.-W.; Rodrigues, E.; Bhattacharjee, A.; Raeisimakiyani, P.; Daskhan, G. C.; St. Laurent, C. D.; Khoo, K.-H.; Mahal, L. K.; Zandberg, W. F.; Huang, X.; Klassen, J. S.; Macauley, M. S. *ACS Chem. Biol.* **2021**, *16*, 2673–2689. doi:10.1021/acscchembio.1c00501
29. Lundström, J.; Varrot, A. Structure of the N-terminal domain of CMA in complex with N-acetyllactosamine. [https://www.wwpdb.org/pdb?id=pdb\\_00008r8a](https://www.wwpdb.org/pdb?id=pdb_00008r8a) (accessed Feb 12, 2024). doi:10.2210/pdb8r8a/pdb
30. Varrot, A. Structure of the N-terminal domain of CMA from *Cucumis melo* in complex with N-acetylgalactosamine. [https://www.wwpdb.org/pdb?id=pdb\\_00008r8c](https://www.wwpdb.org/pdb?id=pdb_00008r8c) (accessed Feb 12, 2024). doi:10.2210/pdb8r8c/pdb
31. Meng, E. C.; Goddard, T. D.; Pettersen, E. F.; Couch, G. S.; Pearson, Z. J.; Morris, J. H.; Ferrin, T. E. *Protein Sci.* **2023**, *32*, e4792. doi:10.1002/pro.4792
32. Liu, Y.; McBride, R.; Stoll, M.; Palma, A. S.; Silva, L.; Agravat, S.; Aoki-Kinoshita, K. F.; Campbell, M. P.; Costello, C. E.; Dell, A.; Haslam, S. M.; Karlsson, N. G.; Khoo, K.-H.; Kolarich, D.; Novotny, M. V.; Packer, N. H.; Ranzinger, R.; Rapp, E.; Rudd, P. M.; Struwe, W. B.; Tiemeyer, M.; Wells, L.; York, W. S.; Zaia, J.; Kettner, C.; Paulson, J. C.; Feizi, T.; Smith, D. F. *Glycobiology* **2017**, *27*, 280–284. doi:10.1093/glycob/cww118
33. Kabsch, W. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 125–132. doi:10.1107/s0907444909047337
34. legrandp/xdsme: March 2019 version working with the latest XDS version (Jan 26, 2018). <https://zenodo.org/records/2613389> (accessed Feb 8, 2024). doi:10.5281/zenodo.837885
35. Agirre, J.; Atanasova, M.; Bagdonas, H.; Ballard, C. B.; Baslé, A.; Beilsten-Edmands, J.; Borges, R. J.; Brown, D. G.; Burgos-Mármol, J. J.; Berrisford, J. M.; Bond, P. S.; Caballero, I.; Catapano, L.; Chojnowski, G.; Cook, A. G.; Cowtan, K. D.; Croll, T. I.; Debreczeni, J. É.; Devenish, N. E.; Dodson, E. J.; Drevon, T. R.; Emsley, P.; Evans, G.; Evans, P. R.; Fando, M.; Foadi, J.; Fuentes-Montero, L.; Garman, E. F.; Gerstel, M.; Gildea, R. J.; Hatti, K.; Hekkelman, M. L.; Heuser, P.; Hoh, S. W.; Hough, M. A.; Jenkins, H. T.; Jiménez, E.; Joosten, R. P.; Keegan, R. M.; Keep, N.; Krissinel, E. B.; Kolenko, P.; Kovalevskiy, O.; Lamzin, V. S.; Lawson, D. M.; Lebedev, A. A.; Leslie, A. G. W.; Lohkamp, B.; Long, F.; Malý, M.; McCoy, A. J.; McNicholas, S. J.; Medina, A.; Millán, C.; Murray, J. W.; Murshudov, G. N.; Nicholls, R. A.; Noble, M. E. M.; Oeffner, R.; Pannu, N. S.; Parkhurst, J. M.; Pearce, N.; Pereira, J.; Perrakis, A.; Powell, H. R.; Read, R. J.; Rigden, D. J.; Rochira, W.; Sammito, M.; Sánchez Rodríguez, F.; Sheldrick, G. M.; Shelley, K. L.; Simkovic, F.; Simpkin, A. J.; Skubak, P.; Sobolev, E.; Steiner, R. A.; Stevenson, K.; Tews, I.; Thomas, J. M. H.; Thorn, A.; Valls, J. T.; Uski, V.; Usón, I.; Vagin, A.; Velankar, S.; Vollmar, M.; Walden, H.; Waterman, D.; Wilson, K. S.; Winn, M. D.; Winter, G.; Wojdyr, M.; Yamashita, K. *Acta Crystallogr., Sect. D: Struct. Biol.* **2023**, *79*, 449–461. doi:10.1107/s2059798323003595
36. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Židek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohli, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. *Nature* **2021**, *596*, 583–589. doi:10.1038/s41586-021-03819-2
37. McCoy, A. J. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63*, 32–41. doi:10.1107/s0907444906045975

38. Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 355–367. doi:10.1107/s0907444911001314
39. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486–501. doi:10.1107/s0907444910007493
40. Agirre, J.; Iglesias-Fernández, J.; Rovira, C.; Davies, G. J.; Wilson, K. S.; Cowtan, K. D. *Nat. Struct. Mol. Biol.* **2015**, *22*, 833–834. doi:10.1038/nsmb.3115
41. Hudson, K. L.; Bartlett, G. J.; Diehl, R. C.; Agirre, J.; Gallagher, T.; Kiessling, L. L.; Woolfson, D. N. *J. Am. Chem. Soc.* **2015**, *137*, 15152–15160. doi:10.1021/jacs.5b08424
42. Brandl, M.; Weiss, M. S.; Jabs, A.; Sühnel, J.; Hilgenfeld, R. *J. Mol. Biol.* **2001**, *307*, 357–377. doi:10.1006/jmbi.2000.4473

## License and Terms

This is an open access article licensed under the terms of the Beilstein-Institut Open Access License Agreement (<https://www.beilstein-journals.org/bjoc/terms>), which is identical to the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0>). The reuse of material under this license requires that the author(s), source and license are credited. Third-party material in this article could be subject to other licenses (typically indicated in the credit line), and in this case, users are required to obtain permission from the license holder to reuse the material.

The definitive version of this article is the electronic one which can be found at:  
<https://doi.org/10.3762/bjoc.20.31>