Current Research in Biotechnology 8 (2024) 100255

Contents lists available at ScienceDirect

Current Research in Biotechnology

journal homepage: www.elsevier.com/locate/crbiot

Expanding the chitin oligosaccharide portfolio by engineering NodC chitin synthases in *Escherichia coli*

Chiara Guidi [a], Xevi Biarnés [b], Antoni Planas [b,1], Marjan De Mey [a,*,1]

[a] Centre for Synthetic Biology, Ghent University, Coupure Links 653 9000, Ghent, Belgium
[b] Laboratory of Biochemistry, Institut Químic de Sarrià, Universitat Ramon Llull, Via Augusta 390 08017, Barcelona, Spain

ARTICLE INFO

Keywords:
Chitin oligosaccharide synthase
Molecular dynamics
NodC
Protein engineering
Synthetic biology

ABSTRACT

Synthetic biology greatly accelerated the building process of potential microbial cell factories for the production of industrially relevant compounds, e.g., chitooligosaccharides (COS) which have an enormous application potential in multiple industries, i.e., pharma, cosmetics and agrifood. COS are produced by the heterologous expression of the chitin oligosaccharide synthase, NodC, in *Escherichia coli,* mainly yielding mixtures of chitintetraose (A4) and/or chitinpentaose (A5). We rationalised here product formation limitations based on molecular modelling of the structures of several NodC enzymes. We used this information to protein engineer NodC, rendering longer COS. Hence, an in vivo platform of defined COS-producing strains with different degrees of polymerisation was developed and experimentally characterised. Significantly, several strains were producing long COS, such as chitinhexaose (A6) and −heptaose (A7), not identified in any other natural producer. Additionally, other engineered strains efficiently produce almost 100% specific A4 or A5 product. Altogether, our results indicate that electrostatics-driven dynamics effects are to be considered in the molecular ruler hypothesis. Charge density at the transmembrane helices of NodC affects the opening of the integral binding pocket and in this way the length of the produced chitin oligomers can be modulated. As a result, the internal ruler mechanism elaborated and validated in this manuscript can serve as a guideline to perform site-directed mutagenesis at positions in related NodC and chitin synthase enzymes for both industrial applications as for identification of therapeutic targets.

Introduction

Chitooligosaccharides (COS) are bioactive molecules with potential applications in the food (Rakkhumkaew and Pengsuk, 2018; Rao et al., 2008) and feed industry (Duan et al., 2020; Osho and Adeola, 2020; Wan et al., 2017), cosmetics (Aranaz et al., 2018), and health care (Wolinsky et al., 2012; Xu et al., 2020). Many of their bioactivities appear to reflect highly specific effector-receptor interactions (Hayafune et al., 2014). This umbrella term, for homo- and hetero-oligosaccharides composed of β-(1,4)-linked *N*-acetylglucosamine (GlcNAc, A) and/or glucosamine (GlcN, D) units, hides an incredibly diverse palette of molecules in terms of degree of polymerisation (DP), degree of acetylation (DA) and pattern of acetylation (PA), rendering them with unique characteristics. However, elucidating structure–function relationships of these oligosaccharides is severely hindered by their structural complexity and microheterogeneity. Expanding the COS-portfolio from low (DP 4–5) to higher DPs could extend this range of bioactivities. For example, COS with DPs of 7–8 directly affect plants by increasing their resistance to fungal and bacterial diseases (Basa et al., 2020; Feng et al., 2019; Hayafune et al., 2014; Wan et al., 2008). COS-derived products could also be developed for applications in medicine and drug delivery, e.g., increased ability to cross the blood–brain barrier (BBB), which could improve the efficiency of drugs targeting Alzheimer's disease (Akhlaghi et al., 2013; Ojeda-Hernández et al., 2020; Ouyang et al., 2017; Yu et al., 2019). In addition to an expanded COS-portfolio, industry would greatly benefit from the production of specific and well-defined COS with DP4 or 5, due to the opportunities for the feed market, e.g., accelerated piglet growth rate and improved piglet immune response (Duan et al., 2020). COS can be produced by chemical synthesis, enzymatic degradation or biotechnologically by the heterologous expression of the enzyme NodC

* Corresponding author.
E-mail addresses: chiara.guidi@ugent.be (C. Guidi), xevi.biarnes@iqs.url.edu (X. Biarnés), antoni.planas@iqs.url.edu (A. Planas), marjan.demey@ugent.be (M. De Mey).
[1] These authors jointly supervised this work.

https://doi.org/10.1016/j.crbiot.2024.100255
Received 29 April 2024; Received in revised form 29 August 2024; Accepted 10 September 2024
Available online 13 September 2024

in *Escherichia coli* (*E. coli*) (Fig. 1a) (Samain et al., 1999). More specifically, current chemical or enzymatic synthesis or degradation technologies are not able to produce pure oligomers with fully defined architecture and in sufficient amounts to study their bioactivities (Aam et al., 2010; Alsina et al., 2021; Dong et al., 2014; Hamed et al., 2016; Hao et al., 2021; Liaqat and Eltem, 2018). Current industrial production of COS involves harsh chemical (or enzymatic) incomplete hydrolysis of shrimp shells or insect cuticula producing a COS mixture as an intermediate step, purified via extensive (& expensive) downstream processing (DSP) to generate COS molecules (fully acetylated or fully deacetylated) or a mixture of partially acetylated COS (paCOS). Thus, current technologies work as complete black boxes: final product varies from batch-to-batch yielding (pa)COS-mixtures for which results may fluctuate without knowing why and how to improve this. Furthermore, this conventional route is subjected to regional and seasonal fluctuations (crustacea harvest dependent) and to a potential allergen threat (animal-origin). Industries across diverse sectors are actively seeking a cost-effective, sustainable, and efficient method for obtaining defined and pure COS.

The *nod*C gene is part of the *nod*ABC cluster in *Rhizobia*, and is essential for the synthesis and secretion of *Rhizobial* lipo-chitin oligosaccharides (LCO), also known as Nod factors, which induce nodule formation for nitrogen fixation in plant roots. In return, the plant provides nutrients for the bacteria (Fliegmann and Bono, 2015; Kamst et al., 1999; Schultze and Kondorosi, 1996). The heterologous expression of NodC in *E. coli* produces fully acetylated COS and the enzyme can therefore be classified as a β-*N*-acetyl-glucosaminyltransferase, representing the glycosyltransferase 2 family (GT2). There are many different types of glycosyltransferases (EC2.4.1.x), and these are assigned to approximately 100 families based on sequence similarities in the CAZy database (Saxena and Brown, 1997). GT2 enzymes are inverting GTs with a GT-A fold composed of a single α/β/α sandwich structure (Rossmann-type fold) that catalyse glycosyl transfer from a sugar nucleotide donor to a sugar acceptor (Lairson et al., 2008). In the case of NodC, UDP-GlcNAc is the donor and GlcNAc is transferred to a growing chain of β-1,4-linked GlcNAc. The NodC proteins of *Rhizobium* sp. GRH2 (*R. sp.* GRH2) and *Sinorhizobium meliloti* (*S. meliloti*) yield chitinpentaose and −tetraose as their main products, respectively (Kamst and Breek, 2000). Several research groups refer to the C-terminus of these proteins as an important part to determine the COS-chain length by forming a cleft in the inner-membrane (IM), prohibiting further elongation of the chain (Dorfmueller et al., 2014; Kamst et al., 1997; Kamst and Breek, 2000). A 3D-model was built for NodC from *Sinorhizobium meliloti* (strain SM11) by (Dorfmueller et al., 2014) taking *R. sphaeroides* cellulose synthase (BcsA, PDB code 4HG6) as structural template. This structural model suggested two residues defining a molecular ruler for oligomer production of limited length in contrast to chitin or cellulose synthases which encode a transporter channel (Dorfmueller et al., 2014) (Fig. 1b). In addition to the multiple application opportunities, engineering NodC for the microbial production of COS with a higher or specific DP would help to circumvent production technologies based on carbohydrate chemistry or the enzymatic/chemical depolymerisation of chitin and chitosan. Neither of these methods sustainably produce COS with a completely defined architecture (Dong et al., 2014; Hao et al., 2021).

The engineering of NodC is challenging because no crystal structure is yet available and the correlation between NodC structure and DP is therefore unknown. However, it is known that GTs generally possess one of two major folds (GT-A and GT-B) and that NodC enzymes adopt a GT-A fold, comprising a central continuous β-sheet surrounded by α-helices. To address the lack of structural data, we therefore constructed a homology model of *Rhizobium* sp. GRH2 and *Sinorhizobium meliloti* strain 1021 NodC based on multiple templates from the GT2 family. We used these models to predict amino acid exchanges that would modify the products length generated by these enzymes. Mutational analysis and molecular dynamics simulations provided structural insights on a GT2 synthase producing oligomers instead of polymers. New NodC enzyme variants rendering COS with defined DP of 5 or 6 and up to 7 are reported. The employed protein engineering strategies will allow for the development of specific chitin (oligosaccharide) synthase engineering guidelines for the in vivo production of a wide range of COS.

Results

Sequence and structure analysis of GT2 enzymes for NodC homology modelling

The current knowledge on the chitin synthase mechanism is mainly based on in vitro/in vivo enzyme studies or bioinformatic analysis

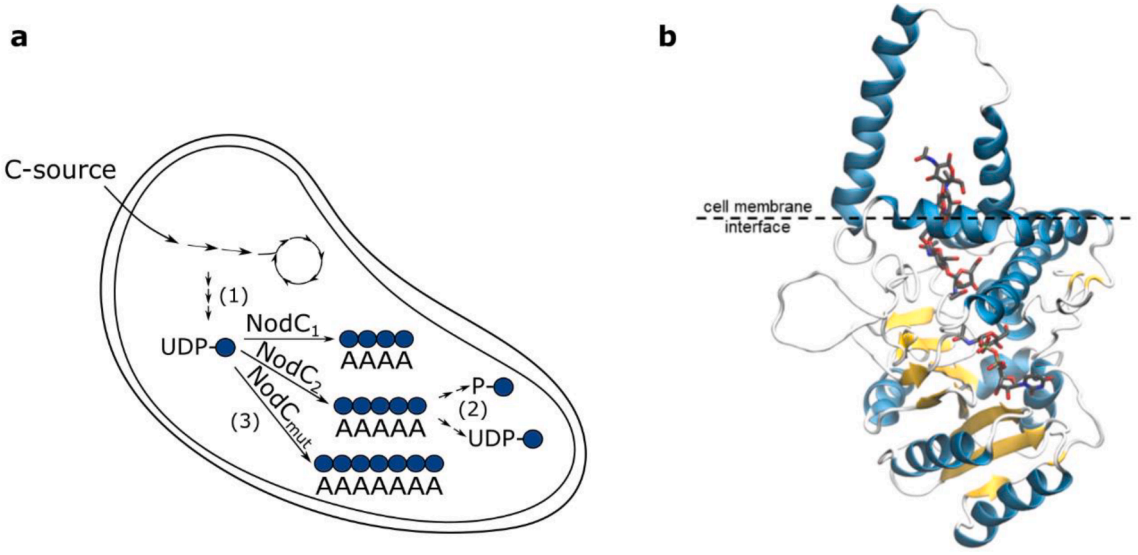

Fig. 1. Microbial production of fully acetylated chitin oligosaccharides (COS). **(a)** Metabolic engineering strategies for in vivo synthesis of fully acetylated COS. Previous engineering strategies include efficiently directing cellular resources (1) and eliminating COS-modifying enzymes (WO2020/058493 A1) (2). The strategy proposed here is protein engineering of specific chitin oligosaccharide synthases, i.e., NodC, to further expand the COS portfolio (EP 21208979.1) (3). **(b)** NodC homology model (this work) suggesting the presence of a catalytic cleft A=*N*-acetylglucosamine (A from acetylated), C=carbon, NodC=chitin oligosaccharide synthase, P=phosphate, R=arginine, UDP=uridine-di-phosphate.

(Dorfmueller et al., 2014; Gohlke et al., 2017). Recent work elucidated several structural insights on the chitin biosynthesis in both *Candida albicans* and *Phytophtora sojae* (Chen et al., 2022; Ren et al., 2022). However, no structural information, crystal structure, is available for any NodC enzyme making NodC engineering strategies, to expand the currently available COS-portfolio, very challenging. For this reason, a thorough sequence and structure analysis has been performed for modelling NodC structure. The GT2 enzymes characterised so far with available structural information are listed in Table S4. We only included structures that align properly with the NodC GT-A domain. This list was complemented with non-bacterial glycosyltransferases (GTs) accessed using the HHpred server (Söding et al., 2005). All these enzymes share a consensus topology of secondary structure elements (Romero-García et al., 2013). The region between conserved β-strands 5 and 6 is a highly variable region among the different GT-A fold enzymes, difficult to align, and is also present in NodC (Romero-García et al., 2013). The C-terminal region of NodC is probably required to determine the COS-chain length (Kamst et al., 1997; Kamst and Breek, 2000). However, a good template for the C-terminus (Ct) was not found in a straightforward manner. The Ct of other GT2 enzymes did not align properly with the Ct of NodC. For this reason, two separate alignments were made: one for the N-terminus (Nt) and central region of NodC and another for the Ct. Both alignments were based on sequence profile analysis and guided by structural information. In the end, the two alignments were blended into one reference multiple sequence alignment which reflects the consensus topology of this family of enzymes (Fig. 2). The final complete alignment was used to create a homology model of respectively the NodCs of *Rhizobium* sp. GRH2 (*Rs*NodC) and *S. meliloti* strain 1021 (*Sm*NodC).

Structural modelling of RsNodC and SmNodC

Different automatic modelling servers, e.g., HHpred (Söding et al., 2005), I-TASSER (Yang and Zhang, 2015), AlphaFold (Abramson et al., 2024; Jumper et al., 2021) or YASARA (Krieger et al., 2002), were initially tested to model both NodC structures. However, the final models were strongly dependent on the server used and did not take into account the binding of substrate and/or product in the catalytic pocket of NodC. This resulted in secondary structures which interfered with substrate and/or product binding. Therefore, models of *Rs*NodC and *Sm*NodC were built based on the reference sequence alignment (Fig. 2) by combination of multiple templates (2) in MODELLER (Larsson et al., 2008). The use of templates 5EJ1 (cellulose synthase subunit A) and 5EKE (undecaprenyl-P β-glycosyltransferase) rendered NodC models with a predicted tertiary structure compatible with the GT-A fold. However, our models lack a segment of the N-terminus, predicted to be located in the membrane (Dorfmueller et al., 2014; Weyer et al., 2022), due to the lack of proper templates. For each NodC enzyme, 20 different models were generated and one representative structure was selected for *Sm*NodC and *Rs*NodC for further docking and modelling purposes. Representative structures (Figure S1) were chosen based on two criteria: (i) matching of the Ct-domain with that of cellulose synthase subunit A, and (ii) the side chain of *Sm*NodC Arg349 and *Rs*NodC Arg346 pointed towards the product ligand as described in the active site model of (Dorfmueller et al., 2014). An overall view of the generated model is shown in Fig. 1.

To understand the determinants of substrate length specificity of *Rs*NodC and *Sm*NodC, COS of different length were docked to both modelled structures. Two different docking software were used:

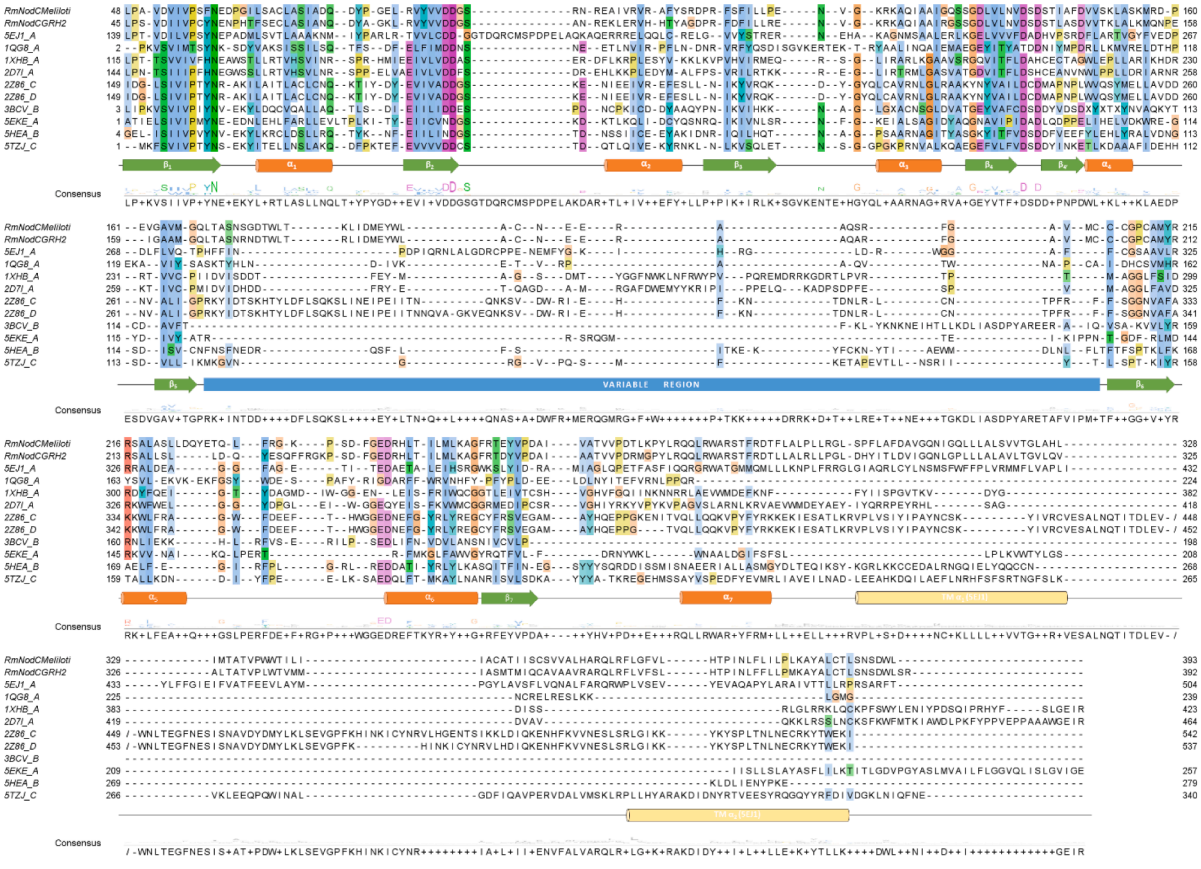

Fig. 2. MSA of *Rs*NodC and *Sm*NodC with 10 GT-A templates. Accession names of the templates in the PDB were used. The consensus topology, as described by (Romero-García et al., 2013), is illustrated and also the secondary structure of the transmembrane domains of cellulose synthase (5EJ1 A). GT=glycosyltransferase, MSA=multiple sequence alignment, PDB=protein data bank, *Rs*NodC=chitin oligosaccharide synthase from *Rhizobium* sp. GRH2, *Sm*NodC=chitin oligosaccharide synthase from *Sinorhizobium meliloti* strain 1021.

AutoDock4 (Morris et al., 2009) and AutoDock Vina (Trott and Olson, 2009). The former is appropriate for the docking of the shorter COS (A2-A3). For the docking of longer COS (A4-A7), AutoDock Vina was applied. The docking search space covered the acceptor and product binding site and part of the catalytic cleft (Figure S2). The donor substrate, UDP-GlcNAc, was added to both homology models to allow checking for the proper relative orientation to the COS ligands (reducing end of the donor towards the non-reducing end of the COS) (Dorfmueller et al., 2014). Docking results showed that short chain COS (A2-A4) can bind and fit in the catalytic pocket for both *Rs*NodC and *Sm*NodC enzymes (Figure S3). In contrast, docking results for longer chain COS (A5-A7) suggested two possible binding sites: (i) binding along the catalytic cleft and extending in the two transmembrane helices ("integral binding pocket", Fig. 3a and Figure S3), and (ii) binding in the active site but extending away from the enzyme along an amphipathic helix presumably located at the membrane interface ("export route cleft", Fig. 3b and Figure S3). The extension of longer COS in the integral binding pocket is apparently restricted by bulky side chain amino acids such as *Sm*NodC Arg349/Met345 and *Rs*NodC Arg346/Met342. In contrast, binding along the export route cleft, predominantly observed for *Sm*NodC models, does not show any apparent restriction to longer COS.

Mutational analysis of NodC's internal ruler mechanism

The presented structural modelling of NodC enzymes in complex with COS ligands has suggested some determinants that may limit their product length. According to these, several amino acid exchanges were proposed to test the hypothesis of enlarging the catalytic cleft in the "integral binding pocket" or favoring the access through the "export route cleft".

Integral binding pocket

(Dorfmueller et al., 2014) already hypothesised that the predicted product-binding site for *Sm*NodC is limited to five sugar units by amino acids Arg349 and Leu19, although their impact on product length formation was not assessed experimentally. In accordance with their findings, our models suggest that amino acids Arg349 and Arg346 in *Sm*NodC and *Rs*NodC, respectively, seem crucial to define the product length profile. Mutants at these positions were prepared by site-directed mutagenesis on *Rs*NodC and *Sm*NodC genes. These were tested for *in vivo* production of COS in an *E. coli* strain with 3 knock-outs that counters COS-interfering background mechanisms (see Experimental section). Single *Rs*NodC mutant R346S produced mainly A6 (92.06 ± 2.5 %)

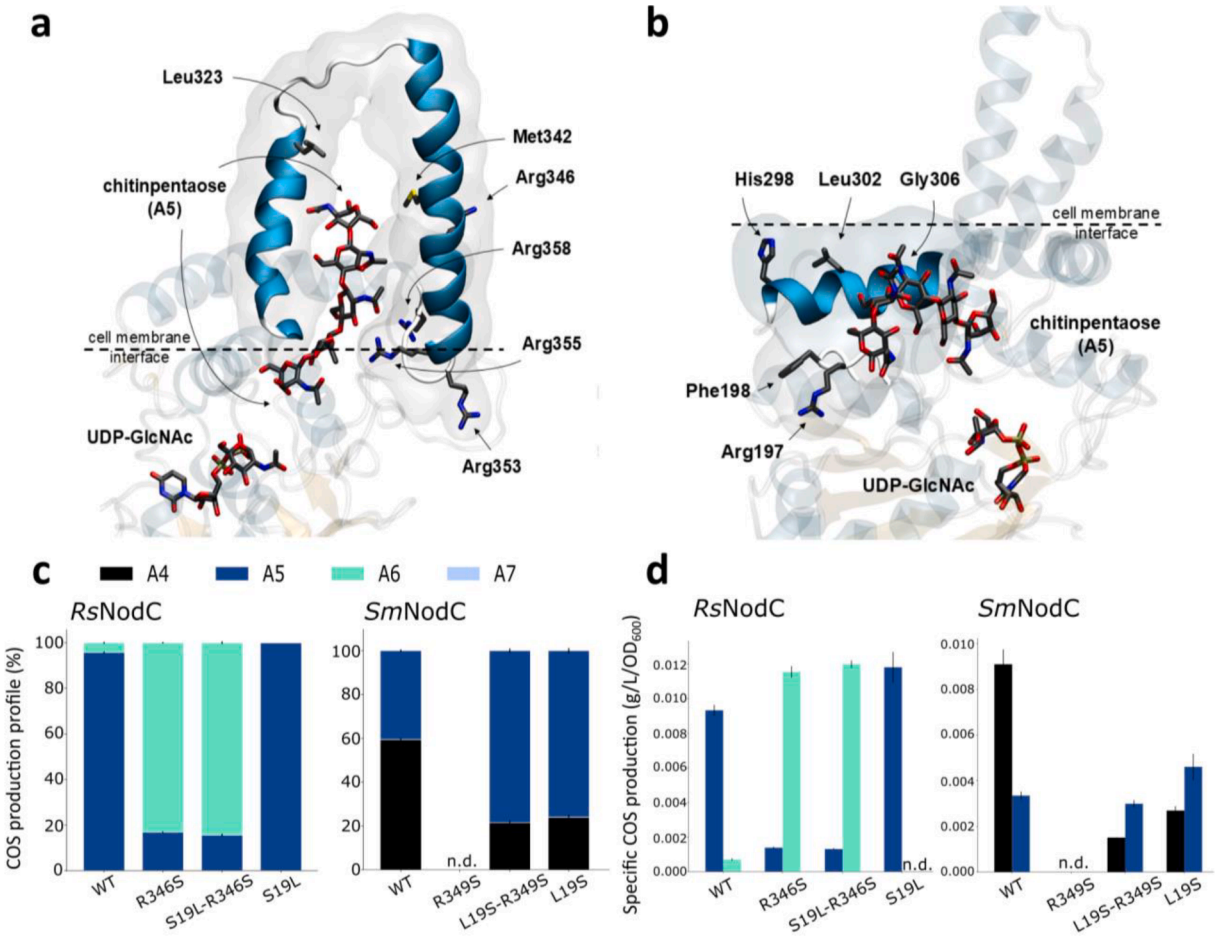

Fig. 3. Binding modes of long chain COS in *Rs*NodC as suggested by homology modelling and virtual docking (Figure S4 for *Sm*NodC) and in vivo characterisation of COS production by *Rs*NodC and *Sm*NodC and derived mutants to study structural changes in the catalytic cleft. **(a)** Integral membrane binding pocket defined by two transmembrane α-helices, and **(b)** export route cleft defined by an amphipathic α-helix. In both models the non-reducing end is in a proper orientation to accept GlcNAc transfer. Ligands chitinpentaose (A5) and UDP-GlcNAc are colored according to atom type (C, black; O, red; N, blue). Amino acids targeted for mutagenesis are shown in thick lines. Protein α-helices are shown in blue cartoons and their corresponding solvent accessible surface area is represented as a shaded surface. **(c)** COS production profile of *Rs*NodC and *Sm*NodC mutants. **(d)** Specific A4, A5 and A6 production of *Rs*NodC and *Sm*NodC mutants. The specific COS-production is represented as a calculated average of three biological replicates (n = 3) and error bars represent the calculated standard deviations. All statistical tests (ANOVA) and p-values are depicted in Table S9. n.d. = non-detectable, A=*N*-acetylglucosamine (A from acetylated), COS=chitooligosaccharides, OD_{600} = optical density measured at 600 nm, *Sm*NodC=chitin oligosaccharide synthase from *Sinorhizobium meliloti* strain 1021, *Rs*NodC=chitin oligosaccharide synthase from *Rhizobium* sp. GRH2, WT=wild-type. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

causing a remarkable COS-production switch from A5 (wild-type (WT) *Rs*NodC produces mainly A5 with 98.2 ± 0.2 %) to the longer A6 (Fig. 3c). In contrast, *Sm*NodC R349S resulted in a non-functional enzyme, probably due to local misfolding of the enzyme. Regarding the other position suggested by (Dorfmueller et al., 2014), the single mutant L19S in *Sm*NodC yielded only a moderate shift from A4 towards A5 (Fig. 3c), and the single mutant *Rs*NodC S19L resulted in specific A5 production. Thus, the product length profile was not remarkably impacted by mutations at position 19 in both enzymes, and no COS with higher DP were observed (Fig. 3c and d). The double mutant *Rs*NodC R346S-S19L was also constructed which resulted in the same product profile as for *Rs*NodC R346S, i.e., higher production of the long A6 product (Fig. 3c). The double mutant *Sm*NodC R349S-L19S also generated a similar product profile as the single L19S mutant with a moderate shift from A4 towards A5 (Fig. 3c).

The molecular ruler hypothesis postulated by (Dorfmueller et al., 2014) seems reasonable although not conclusive. *Rs*NodC R346S variant produce long A6, but the role of L19 is not crucial in terms of product length. Moreover, these alterations in *Sm*NodC affected protein folding more than the equivalent mutations in *Rs*NodC yielding non-functional enzymes. Thus, following mutational studies were done exclusively in *Rs*NodC. Additional positions along the integral binding pocket (R346W, R346S/M342W, R346S/M342S, R346S/M342S/L323W) were mutated aimed at increasing space in the cleft or increase product-enzyme interactions in *Rs*NodC following the molecular ruler hypothesis (Dorfmueller et al., 2014). Unfortunately, these attempts failed and neither increased A6 production nor an increase in DP was observed (Figure S5). Alternatively, position 346 in *Rs*NodC was further mutated into amino acids (AAs) with varying size and charge. Retaining the positive charge, i.e., R346K, resulted in the same product profile as for *Rs*NodC WT (Fig. 4a). In contrast to what was expected, substitution of Arg346 by neutral side chain AAs with similar side-chain length (e.g., glutamine, Q), also increased the DP towards A6 production. It seemed that deleting the positive charge was the main reason for an increase in DP and not removing the steric hindrance in the cleft. Surprisingly, upon introduction of a negative charge (aspartate, D or glutamate, E side chains) at position 346, an additional increase in DP was obtained: A7 production (Fig. 4a and Table S7). In the past, this DP was never observed to be produced by any *Rhizobial* species.

To rationalise this change in product length formation, the impact of these mutations on the protein structure was assessed by atomistic molecular dynamics simulations of WT *Sm*NodC and *Rs*NodC, and of *Rs*NodC R346E and R346K mutants. Even though NodC enzymes are often claimed to be membrane anchored proteins, microscopic visualisation of recombinantly expressed *Rs*NodC in *E. coli* localised the enzyme (at least partially) in inclusion bodies (Figure S6). Whether both membrane-bound and cytosolic (as aggregates) forms co-exist or are in dynamic equilibrium is unknown. Here, we evaluate the impact of

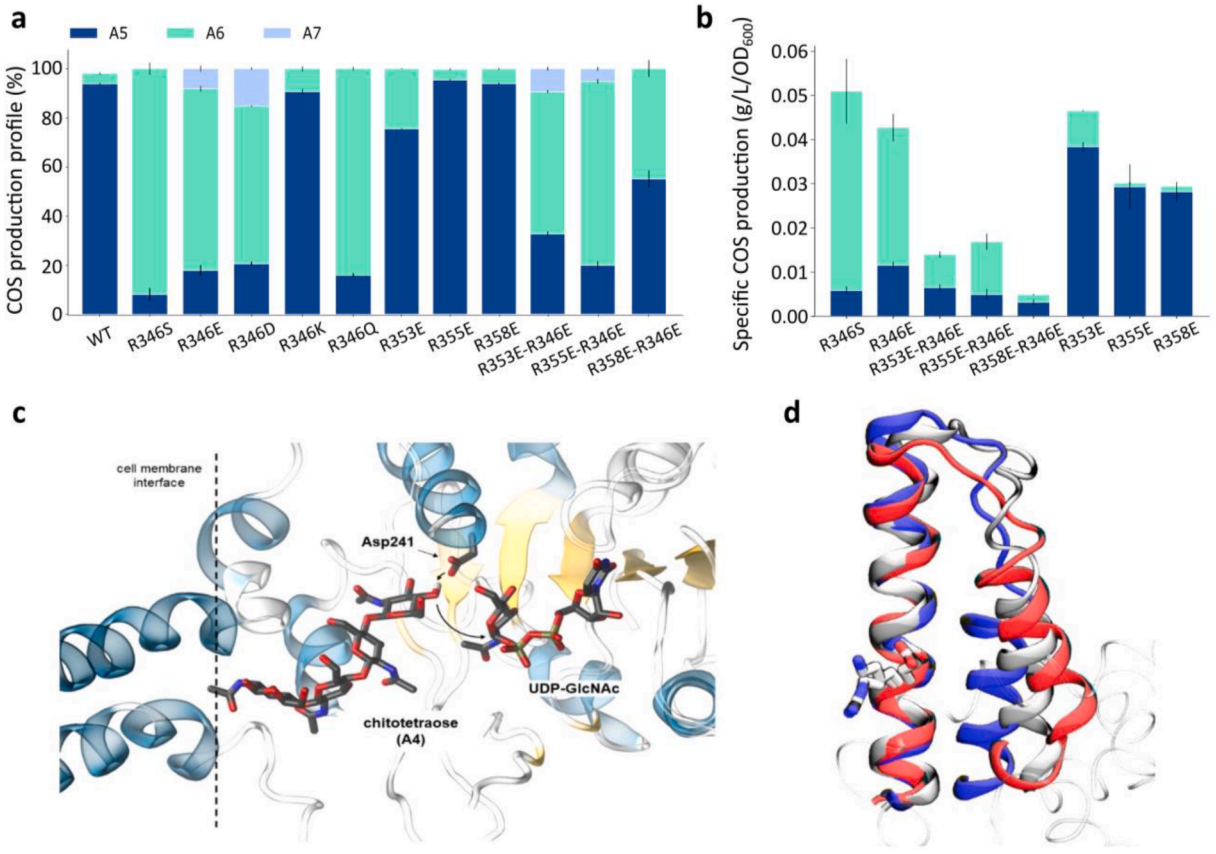

Fig. 4. In vivo characterisation of COS production by *Rs*NodC and derived mutants and results of molecular dynamics studies. **(a)** COS production profile for several single and double *Rs*NodC mutants. **(b)** Specific A5 and A6 production for multiple *Rs*NodC mutants with significant increase in A6 production compared to the *Rs*NodC WT. COS production is represented as a calculated average of three biological replicates (n = 3) and error bars represent the calculated standard deviation. All statistical tests (ANOVA) and p-values are depicted in Table S9. A=*N*-acetylglucosamine (A from acetylated), COS=chitooligosacchardes, *Rs*NodC=chitin oligosaccharide synthase from *Rhizobium* sp. GRH2 **(c)** Precatalytic ternary complex of *Sm*NodC with UDP-GlcNAc (substrate donor) and chitintetraose (A4, substrate acceptor) refined by molecular dynamics simulations. Substrate ligands and catalytic base (Asp241) are shown in thick lines colored according to atom type (C, black; O, red; N, blue). Protein α-helices and β-sheets are shown in blue and yellow cartoons respectively. Curved arrows indicate the catalytic mechanism for retaining glycosyl-transferases. **(d)** Rearrangement of the "membrane integral binding pocket" upon mutation at *Rs*NodC Arg346 position, evidenced by molecular dynamics simulations. Transmembrane α-helices are shown in cartoons colored in white for the wild-type enzyme, red for the R346E mutant and blue for the R346K variant. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

protein dynamics on substrate specificity in the cytosolic (non-membrane bound) form. The overall fold of the enzyme was maintained in all simulations. Docking of A4 and A5 COS into this equilibrated structure rendered catalytically competent geometries at the active site, thus validating the models (Fig. 4c). The side chain of aspartic acid in the conserved 'EDR' motif was closely positioned to the 4-hydroxyl group of the acceptor *N*-acetylglucosamine unit, activating it for nucleophilic attack (Fig. 4c). For both *Sm*NodC WT and *Rs*NodC WT, the two transmembrane helices defining the integral binding pocket got slightly closer at a distance around 10 Å (Fig. 4b and Figure S8). Notably, simulations show that for the R346E variant, the two TM helices got separated at around 14 Å. This separation generates enough space at the integral binding pocket to accommodate longer COS, such as A7 produced by *Rs*NodC R346E. In contrast, *Rs*NodC R346K mutant kept the two helices even closer (Fig. 4b and Figure S8) reducing the accessibility of the pocket. Indeed, *Rs*NodC R346K mainly produces the shorter A5 product. It can therefore be inferred that neutral- or negatively charged amino acids at this position shift the dynamics towards an open form in the catalytic cleft, yielding higher DP products. For this reason, we concluded that loss of the positive charge modulates the DP and not deleting steric hinderance as suggested before. As electrostatic interactions have been shown to play a vital role in enzyme catalysis (Kosugi and Hayashi, 2012; Liu et al., 2014), additional positively charged AAs, along the same α-helix (residues 320 to 370), were substituted with the negatively charged AA glutamic acid (E): R353E, R355E and R358E. All three single mutants did produce some amount of A6, with single mutant R353E significantly enhancing specific A6 production compared to *Rs*NodC WT (Fig. 4b and Table S9).

Altogether, our results indicate that electrostatics-driven dynamics effects are to be considered in the molecular ruler hypothesis. Charge density at the transmembrane helices of *Rs*NodC affects the opening of the integral binding pocket and in this way the length of the produced

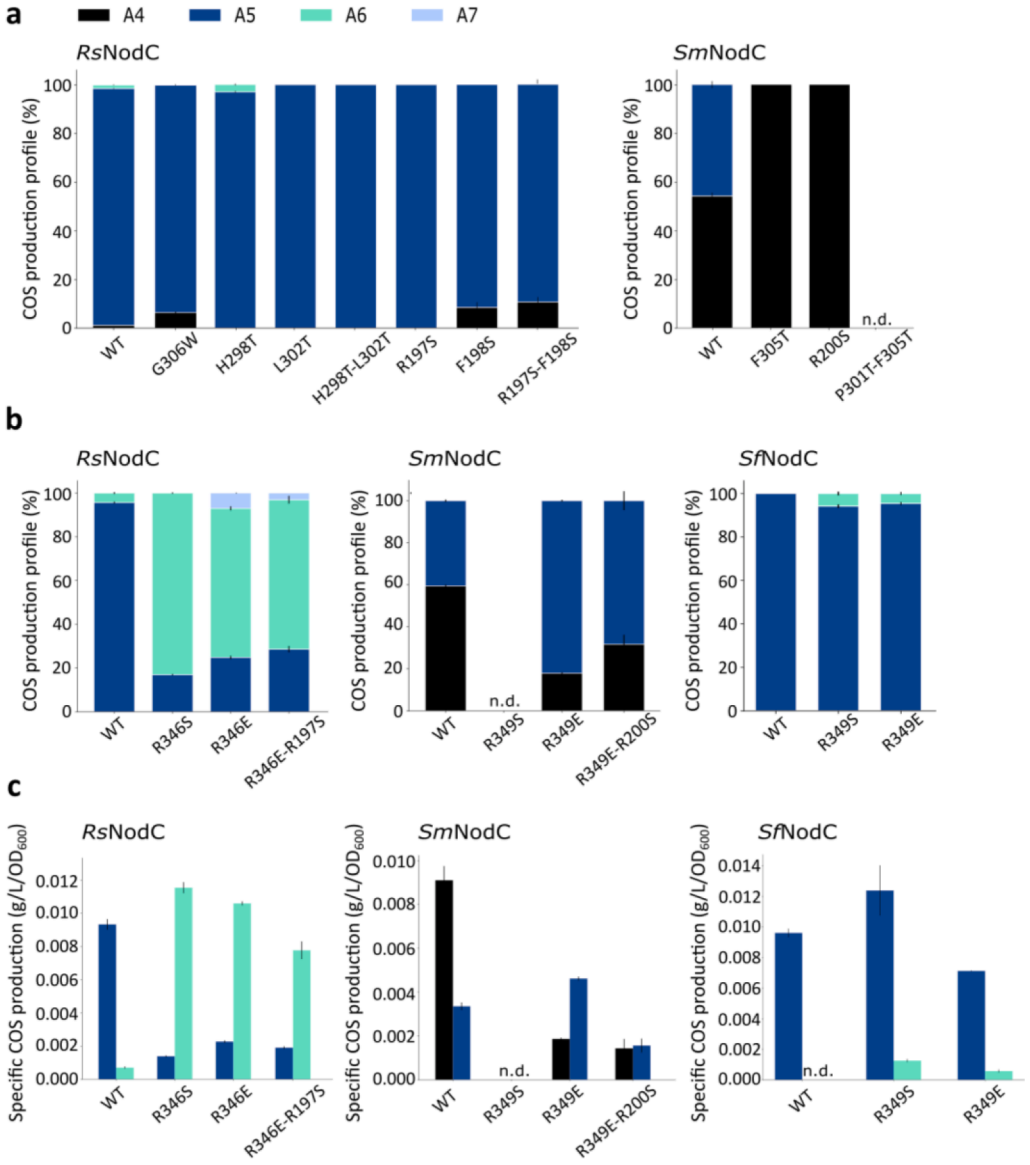

Fig. 5. In vivo characterisation of COS production by *Rs*NodC, *Sm*NodC and *Sf*NodC and derived mutants to study structural changes in the catalytic cleft. **(a)** COS-production profile of *Rs*NodC and *Sm*NodC mutants. **(b)** COS production profile of *Rs*NodC, *Sm*NodC and *Sf*NodC mutants. **(c)** Specific A4, A5 and A6 production of *Rs*NodC, *Sm*NodC and *Sf*NodC mutants. The specific COS-production is represented as a calculated average of three biological replicates (n = 3) and error bars represent the calculated standard deviations. All statistical tests (ANOVA) and p-values are depicted in Table S9. n.d. = non-detectable, A=*N*-acetylglucosamine (A from acetylated), COS=chitooligosaccharides, OD_{600} = optical density measured at 600 nm, *Sm*NodC=chitin oligosaccharide synthase from *Sinorhizobium meliloti* strain 1021, *Rs*NodC=chitin oligosaccharide synthase from *Rhizobium* sp. GRH2, *Sf*NodC=chitin oligosaccharide synthase from *Sinorhizobium fredii* USDA 191, WT=wild-type.

chitin oligomers can be modulated: short products (A5) are obtained when high positive charge densities are present in the binding pocket (Arg346 and Arg353) and long products (A6) are achieved when these charges are neutralised or inverted (R346E and R353E). Lastly, these single mutants were combined with the A7 producer mutation, R346E. Compared to single mutant R346E, double mutant R346E/R353E did significantly increase A5 production while still A6 and A7 were observed (Fig. 4a and Table S9). The effect of single and double mutations was not that pronounced for the other positively charged positions along the TM α-helix (Arg355, Arg358).

Export route cleft

Our modelling and docking studies also revealed a second main binding platform besides the catalytic cleft (Fig. 3b): a route for the product to move out of the enzyme ("export route cleft"). Based on this, we hypothesised that if the growing COS-chain could be directed outside the protein core, while keeping the non-reducing end at the catalytic center, the resulting products length would not be limited by the dimensions of the catalytic cleft. The putative export route would involve an amphipathic α-helix which indicates the position of the cytoplasm-membrane interface (Figure S7). Several residues along this exit route were targeted: R197, F198, H298, L302 and G306 in *Rs*NodC, and equivalent residues R200 and F305 in *Sm*NodC (Fig. 3). These positions were mutated to short and polar amino acids (serine and threonine), except G306 which was substituted to a tryptophan. Bulkier G306W substitution may either establish a stabilizing interaction with the product and hence the growing chain stay longer for further extension (higher DP) or introduce some steric hindrance leading to faster product release and hence shorter DP. Interestingly, *Rs*NodC G306W produced relatively more A4 in comparison to *Rs*NodC WT (Fig. 5a). This could illustrate the improved ability of the product to move out of the enzyme before additional glycosylations, and hence increased A4 accumulation.

An increase in product formation specificity was observed for several of the rest of tested positions. Remarkably, three *Rs*NodC mutants (R197S, L302T and the double mutant R197S/L302T) produced 100 % specific A5, whereas two *Sm*NodC mutants (F305T and R200S) produced 100 % specific A4 products (Fig. 5b). In both cases, it is the accumulation of the main product rendered by the WT enzyme which is achieved. Therefore, modulation of the suggested export route seems to have a clear impact on product specificity rendering useful variants for practical production of defined and pure COS.

At an attempt to combine both observed phenomena so far, the most promising mutations were combined: *Rs*NodC R346E (or R349S in *Sm*NodC) for product elongation and R197S (or R200S in *Sm*NodC) for product specificity. Unfortunately, *Rs*NodC R346E/R197S and *Sm*NodC R349E/R200S did not improve specificity towards A7 or A5 production, respectively, and rendered similar product profiles than single mutants *Rs*NodC R346E (or *Sm*NodC R346E) (Fig. 5b, c and Table S9).

Additionally, with the aim of validating the proposed ruler mechanism, a third NodC enzyme was studied to evaluate if transferring the sensible mutations do produce equivalent phenotypes. *S. fredii* USDA 191 (*Sf*NodC) was cloned using gene walking (Martin-Harris et al., 2010) (See Table S6 for all primers used). Interestingly, *Sf*NodC WT produces exclusively A5 product, no other DPs being detected (Fig. 5 b and c). This is unique among characterised NodC enzymes where the WT enzymes produce mainly A4 or A5 but always with some proportion of other DPs. Next, mutations were introduced at the equivalent *Rs*NodC R346 position. Indeed, deletion of the positive charge at this position in *Sf*NodC (R349S, E), induced the production of longer A6 (Fig. 5 b, c and Table S9). However, productivity was lower in contrast to what was observed for RsNodC R346S or R346E. Again, charge modulation at the membrane integral binding pocket seems to impact the degree of polymerisation of the products in NodC enzymes.

Discussion

Our model for ligand binding to NodC proposes two binding modes for chitooligosaccharide products: (i) binding along the catalytic cleft and extending into the two transmembrane helices ("integral binding pocket"), and (ii) binding in the active site but extending away from the enzyme along an amphipathic helix presumably located at the membrane interface ("export route cleft"). The procedure to build the models in this work differs from the models obtained by (Dorfmueller et al., 2014; Weyer et al., 2022). In more detail, (i) our models were not generated with an automatic modelling online tool (e.g., RaptorX for Dorfmueller et al. (2014) versus SWISS-MODEL for Weyer et al. (2022)), rather (ii) our models were generated using a curated multiple sequence alignment using multiple templates (cellulose synthase subunit A and undecaprenyl-P β-glycosyltransferase), and finally (iii) previous works did not take into account the binding of substrate and/or product in the catalytic pocket before model generation. This will result in secondary structures that interfere with substrate and/or product that were, in the case of Weyer et al. (2022), curated after model generation. To minimise required model corrections, we selected proper models which were able to accommodate substrates after docking and included templates for which substrates were present. However, our models lack a segment of the N-terminus due to the lack of proper templates, hindering the location of residue S19 in *Rs*NodC (L19 in *Sm*NodC). Nevertheless, mutants considering amino acids at position 19 for both proteins were generated to allow a correct comparison with the results of (Dorfmueller et al., 2014; Weyer et al., 2022). In contrast to Weyer et al. (2022), our single mutant *Rs*NodC S19L resulted in pure A5 production after 24 h, no other or shorter COS could be detected. This observation was strengthened by the reverse single mutant in *Sm*NodC (L19S), yielding a shift from A4 to A5 (Fig. 3c). The double mutant *Rs*NodC R346S-S19L yielded a similar result to *Rs*NodC R346S and double mutant *Sm*NodC R349S-L19S also generated a switch to longer oligos (Fig. 3c). We do acknowledge that both amino acids are important for modulating the COS chain length, as reported by (Dorfmueller et al., 2014; Weyer et al., 2022). However, our results indicate that improving space in the integral binding pocket is not sufficient to increase COS chain length. We discovered that deleting the positive charge of crucial amino acids (not only R349, R346) in the integral binding pocket affects the opening of the pocket.

From the mutational analysis here reported, it is proposed that preferential binding to the integral binding pocket defines product length (DP), whereas the export route cleft modulates product specificity. In this way the length of the produced chitin oligomers can be modulated: short products (A5) are obtained when high positive charge densities are present in the binding pocket (Arg346 and Arg353) and long products (A6) are achieved when these charges are neutralised or inverted. Our results indicate that electrostatics-driven dynamics effects are to be considered in the molecular ruler hypothesis of chitin oligosaccharide synthases.

Next, a potential export route for the synthesised chitin oligomers was discovered. In contrast to full processive enzymes, e.g., cellulose or chitin synthases, chitin oligosaccharide synthases, NodC enzymes do not form a transmembrane channel and it is not known what mechanism and exit route results in product release (Guidi et al., 2023). Our models suggest the presence of an amphipathic α-helix at the cytoplasm-membrane interface. Several residues along this export route cleft were targeted which resulted in an almost 100 % product specificity for A4 and A5 products. We hypothesise that if the intermediary product was to be immediately directed through the export route and the enzyme only kept hold of the non-reducing end, the COS-product would not be limited by the dimensions of the integral binding pocket, leading to faster and uniform product release instead of elongating COS chains further into the integral binding pocket.

However, extensive molecular dynamics (MD) simulations, focusing on the export route cleft, using a resolved NodC crystal structure are

needed to fully understand and unravel the exit mechanism in chitin oligosaccharide synthases.

No additional hints towards improved product specificity were found when comparing the *Sf*NodC sequence to *Rs*NodC and *Sm*NodC (Figure S10). For the mutated positions that yielded 100 % product specificity, i.e., L302/R197 and F305/R200 for *Rs*NodC and *Sm*NodC respectively, similar amino acids (L305/R200) are present in *Sf*NodC. Again, we acknowledge that – similar to what was indicated by Weyer et al. (2022) – modulating chain length also strongly depends on the choice of host organism and the genetic context of the NodC gene. In more detail, all experiments were performed using *E. coli* K-12 MG1655 3KO (Coussement, 2016) which counters COS-interfering background mechanisms, improving product purity to a certain extent (data not shown). It is clear that *Sf*NodC has a preferred product specificity towards chitinpentaose, however, the reason could be a combined effect of protein sequence and genomic context.

Based on this model, several guidelines for future NodC engineering can be postulated (Fig. 6 and Figure S9). If integral membrane binding pocket is targeted, an increase in DP can be obtained (Fig. 6a). For this, substitution of positively charged AAs, present in one of the two TM helices, by a non-positively charged AA and, preferably, a negatively charged one is desired. If the export route is targeted, an increase in specificity towards the main produced COS-product can be obtained (Fig. 6b). For this, AAs positioned along the exit route need to be substituted with stability enhancing AAs, e.g., threonine, or positively charged AAs with non-positively charged ones. Of course, it needs to be noted that this study focused on three NodC enzymes which were successfully expressed and functional in *E. coli*. Depending on the genetic context of the NodC enzyme, protein engineering can be more challenging.

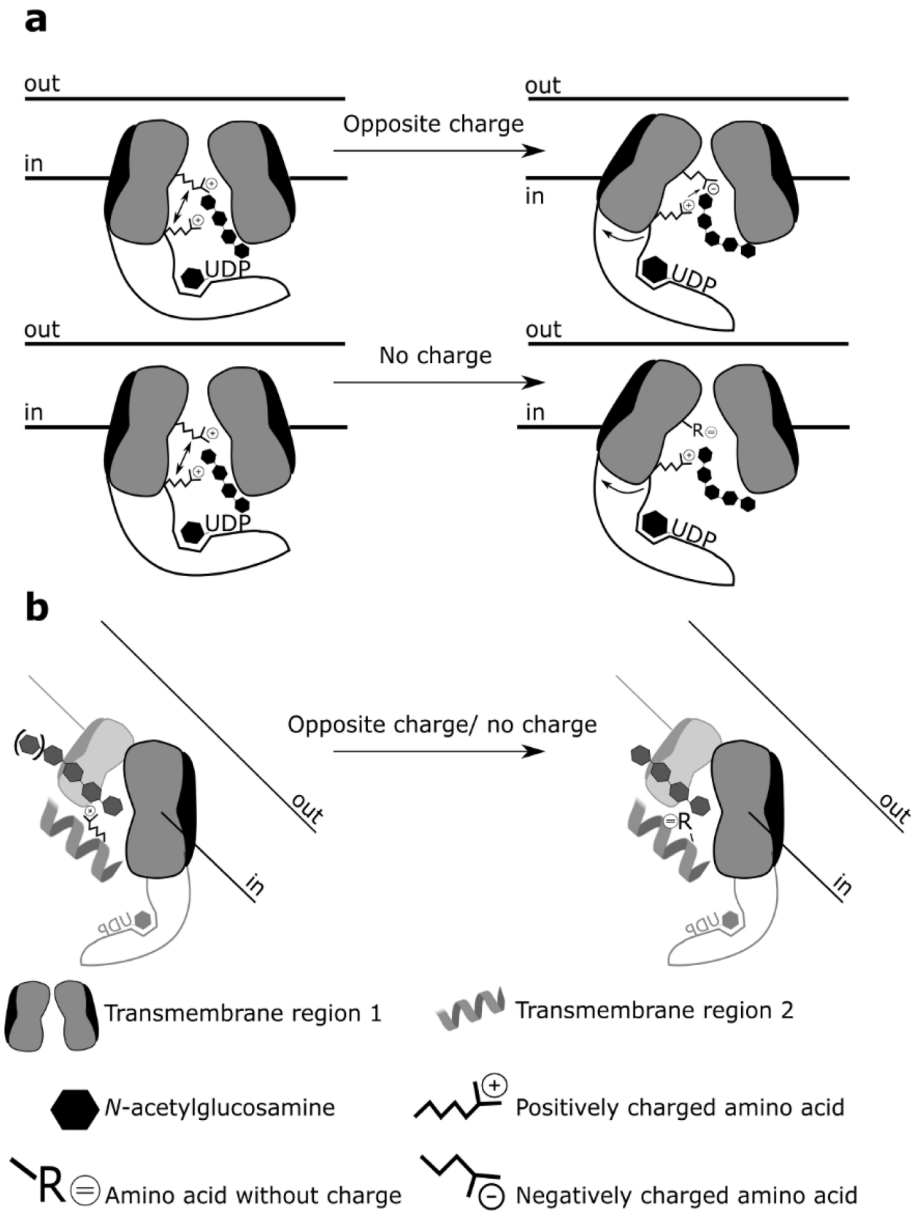

Fig. 6. Schematic representation of the protein engineering strategy of chitin oligosaccharide synthases to synthesise chitin oligosaccharides with a controllable chain length. (a) Example in which a positive charged amino acid in binding pocket 1 is replaced by an amino acid with the opposite charge (negatively charged amino acid) or a neutral amino acid (uncharged amino acid) yielding a chitin oligosaccharide product portfolio with increased fraction of chitin oligosaccharides with higher degree of polymerisation. Protein is depicted in front view. (b) Example in which a positively charged amino acid in binding pocket 2 is replaced by an amino acid with the opposite charge (negatively charged amino acid) or a neutral amino acid (uncharged amino acid) yielding a chitin oligosaccharide product portfolio with increased product specificity (higher proportion of the main native product by the WT enzyme). Protein rotated over 180°. WT=wild-type.

Our findings are of crucial importance for further research to elucidate the highly specific effector-receptor interactions of chitin oligomers. Traditionally, the active product in complex mixtures of chitin oligomers is purified using extensive and costly downstream processing efforts. Therefore, industry is searching for scalable, reproducible and economically viable production technologies that can deliver structurally perfectly defined chitin oligomers in a cost competitive manner. Based on obtained OD_{600}-values and COS-production titers in stationary phase (Table S8), an average COS-productivity of 0.01–0.15 g/OD was synthesised with intracellular product purities ranging from 80 to 100 %, which was never reported before by chemical or enzymatic degradation methods, starting from chitin or chitosan substrates. As a result, the *E. coli*-based microbial cell factories for the production of pure COS-molecules from glucose, combine superior performance in terms of product quality and quantity on one hand, and feasibility of industrial handling on the other hand.

Our results will significantly contribute to biotechnological production of structurally perfectly defined chitin oligomers with unprecedented characteristics in terms of product purity, yield, productivity and titers, allowing testing in real-life industrial scenarios.

Methods

Data retrieval

The list of GT2 family enzymes characterised so far was obtained from the CAZy database (Cantarel et al., 2009). This list was complemented with sequences accessed using the HHpred server (Söding et al., 2005). The three dimensional structures of these enzymes were downloaded from the Protein Data Bank (PDB) (Burley et al., 2017) and their corresponding amino acid sequences were obtained from UniProt (Consortium and T.U., 2008) (Table S4).

Structural and sequence alignment

A structural alignment of the N-terminus of the proteins listed in Table S4 was prepared using VMD software (Humphrey et al., 1996). HMMER software v3.0 was used to build a Hidden Markov Model (HMM) profile of the GT-A fold clan (NodC templates) (Finn et al., 2011). A new multiple sequence alignment, with the protein sequences from UniProt and the NodC sequences from *Rhizobium* sp. GRH2 (Uniprot ID A0A0N7ARR3) and *Sinorhizobium meliloti* strain 1021 (Uniprot ID P04341), was generated using this HMM profile. Although the variable region could not be incorporated into the profile, the HMM captured both conserved regions of the alignment that flank the variable region. In this way, the profile allowed detecting and aligning these regions properly in any member of the GT-A fold clan of proteins, also for the NodC enzymes. The alignment of the variable region was refined to reflect the protein–ligand interactions of selected structures with donor and/or acceptor substrates as inspected using the VMD software (Hsin et al., 2008). Also, the variable regions of NodC enzymes was kept aligned (90.5 % sequence identity). The sequences in the alignment were clustered using the Neighbor-joining algorithm and a BLOSUM62 scoring matrix implemented in JALVIEW (Saitou and Nei, 1987). For the proper alignment of C-terminus region, a second alignment of the selected sequences together with the NodC sequences was generated using the PROMALS server which performs profile-based multiple sequence alignments incorporating secondary structures (Pei et al., 2014). As the Ct is predicted to be located in the inner-membrane (IM) (Dorfmueller et al., 2014; Kamst et al., 1999, 1997; Kamst and Breek, 2000; Kamst and Herman, 1999), we identified if the PROMALS alignment conserved this transmembrane (TM) topology by means of TMHMM server. Indeed, two TM helices were identified in the Ct region of the alignment. The two MSA alignments (N-terminal part and C-terminus) were blended into one MSA and was manually adjusted by visual inspection and superposition of several templates with donor/and substrate using the VMD software (Hsin et al., 2008), e.g., in the variable region.

Homology modelling

The structural models of NodC were constructed by comparative homology modelling using MODELLER v9.8 (Sali and Blundell, 1993). Multiple GT-A structures were used as templates (PDB codes: 5EJ1 and 5EKE). The multiple sequence alignment used to guide the modelling is shown in Fig. 2. 20 different replicates of each model were generated. The N-terminus fragments of NodC sequences to be modelled did not align properly with the templates. As a consequence, the structure of the first 47 and 44 amino acids for *Sm*NodC and *Rs*NodC was discarded as this part of the model only yielded a large unstructured coil. An ensemble of 20 different structural models were generated for each protein. Models are first optimised with the variable target function method Discrete Optimized Protein Energy (assess.DOPE) and GA341 method (assess.GA341). Original ligand geometry is preserved in the model using intra-ligand, inter-ligand, and ligand–protein distance constrains as implemented in MODELLER (AutoModel.non-std_restraints) method. Short missing template fragments (insertions in the sequence alignment) were explicitly modelled with the "auto_loop" library as implemented in MODELLER. For these, an ensemble of 5 replicates for each starting model structure is generated and optimised using the internal function energy for loop optimisation in MODELLER (Fiser et al., 2000). Models were further refined by simulated annealing molecular dynamics in "slow_md" mode as implemented in MODELLER. Both models were attached to Supplementary Information.

Acceptor and donor binding site

Autodock v4.2.6 was used to predict the putative binding sites between chitooligosaccharide (COS) products and NodC protein structures (Morris et al., 1998; Seeliger and De Groot, 2010). A docking strategy was followed in which the catalytic pocket and cleft of the protein structure were scanned for putative binding sites for COS of different chain lengths (DP2, DP3 and DP4). We computed solvation, electrostatic and affinity grid potentials on a specific Grid Box (center x = 11.984, center y = 12.431, center z = 53.321, size x = 72, size y = 102 and size z = 92) of the protein structure for each atom type in the substrate using Autogrid v4.2.6 (Morris et al., 1998). The single bonds of the substrate molecule were considered flexible during the docking search and 100 rounds of a genetic algorithm were performed for docking. Docking with COS DP2 and COS DP3 was achieved using AutoDock4 whereas docking with COS DP4-DP7 was achieved AutoDock Vina (Trott and Olson, 2009). Both sets of results were analysed using VMD software (Hsin et al., 2008).

Molecular dynamics simulations

The molecular dynamics (MD) simulations were initiated using one of the 20 models generated for NodC from *S. meliloti* strain 1021 and NodC from *R. sp.* GRH2. The models were chosen based on the docking results. GROMACs v4.5.3 was used to apply a long MD simulation to each of the structures and no ligand was incorporated (Berendsen et al., 1995). The simulations were performed with an Amber force field, cubic box, solvent explicit treatment and neutral charge in the system using the Amber parameters as previously described (Petrova and Koc, 1999). The systems were equilibrated following the same protocol as in (Romero-García et al., 2013) and simulations were conducted under the NPT ensemble. The MD simulation was extended to 5 ns for each of the structures. At the end of each simulation, all the structures generated during the trajectory were aligned using the (root-mean-square) RMSD as a metric (Daura et al., 1999). A consensus between the temporal location (the closest to 5 ns), and state of the TM helices was used to choose the representative structure for each MD simulation.

Chemicals, oligonucleotides and molecular biology

Products were purchased from Sigma-Aldrich (Diegem, Belgium) unless stated otherwise. Agarose and ethidium bromide were purchased from Thermo Fisher Scientific (Erembodegem, Belgium). Standard molecular biology procedures were conducted as described by Sambrook et al. (1989). A variety of polymerases were used for the different types of PCR reactions: PrimeSTAR HS DNA polymerase (Takara, Westburg, Leusden, The Netherlands) and Q5 polymerase (New England Biolabs, County Road, Ipswich, MA, USA) were used for short DNA fragments (<4kb). PS GxL polymerase (Takara Bio, France) was used for long DNA fragments (>4kb). All DNA fragments were purified using the innuPREP PCR-pure Kit (Analytik Jena AG, Germany). All plasmids were isolated from bacterial cultures using the innuPREP Plasmid Mini Kit (Analytik Jena AG, Germany). Genomic DNA was isolated using the PurelinkTM Microbiome DNA purification kit (Invitrogen Carlsbad, California, USA). Oligonucleotides were purchased from Integrated DNA Technologies (Leuven, Belgium). Sequencing services were conducted by Macrogen (Amsterdam, The Netherlands).

Strains, media and culture conditions

E. coli One Shot Top10 Electrocomp^{TM} (Invitrogen, Carlsbad, California, USA) were used for the construction and maintenance of all plasmids. For all COS production assays throughout this contribution, the *E. coli* K-12 MG1655 (ATCC) 3KO (Coussement, 2016) was used as it counters COS-interfering background mechanisms. Expression vectors were introduced into their host cells by electroporation (Sambrook et al., 1989). *Sinorhizobium fredii* USDA 191 was purchased from the BCCM/LMG bacterial culture collection (LMG 6216 – *Ensifer fredii*). The culture medium Lysogenic Broth (LB) was used for cloning purposes. LB was composed of 1 % tryptone peptone (Difco, Erembodegem, Belgium), 0.5 % yeast extract (Difco) and 1 % (v/v) sodium chloride (VWR, Leuven, Belgium). Lysogenic broth agar (LBA) has the same ingredients as LB, with the addition of 10 g/l bacteriological agar type A (BIOKAR Diagnostics, Beauvais, France). Transformed cultures were plated out on LBA and grown overnight at 30 °C. If required, LB and LBA medium were supplemented with appropriate antibiotics (100 μg/mL ampicillin). For COS production, strains were grown in Minimal Medium (MM) with glucose as carbon source, and was supplemented with appropriate antibiotics (100 μg/mL ampicillin).

MM was composed of 2 g/L NH_4CL, 5 g/L $(NH_4)_2SO_4$, 3 g/L KH_2PO_4, 7.3 g/L K_2HPO_4, 8.4 g/L MOPS, 0.5 g/L NaCl, 0.5 g/L $MgSO_4{\cdot}7H_2O$, and 16.5 g/L glucose as carbon source, 1 mL/L trace element solution, 100 μL/L molybdate solution. Trace element solution consisted of 3.6 g/L $FeCl_2{\cdot}4H_2O$, 5 g/L $CaCl_2{\cdot}2H_2O$, 1.3 g/L $MnCl_2{\cdot}2H_2O$, 0.38 g/L $CuCl_2{\cdot}2H_2O$, 0.5 g/L $CoCl_2.6H_2O$, 0.94 g/L $ZnCl_2$, 0.0311 g/L H_3BO_4, 0.4 g/L $Na_2EDTA{\cdot}2H_2O$, 1.01 g/L thiamine.HCl. The molybdate solution contained 0.967 g/L $Na_2MoO_4{\cdot}2H_2O$ To avoid Maillard reaction and precipitation during sterilisation of the shake flask medium, the glucose and magnesium sulphate were autoclaved separately from the remaining salts. Glucose and magnesium sulphate were autoclaved in a 200 mL solution, the remaining salts in an 800 mL solution. Prior to autoclaving (121 °C, 30 min, 15 psi), the latter was set to a pH of 7 with 1 M KOH. After autoclaving, these two solutions were cooled down and mixed. Subsequently, the trace element and molybdate solutions were added and filter-sterilised with a bottle top filter (Corning PTFE filter, 0.22 μm).

Precultures were grown overnight (16 h) at 30 °C with agitation (200 rpm) in 5 mL LB (20 mL inoculation tube) with appropriate antibiotics and were used for 1 % inoculation of 5 mL fresh MM for 24 h until stationary phase. These MM precultures were used for 1 % inoculation of 50 mL fresh MM in 250 mL-shake flasks and subsequently cultivated at 30 °C and 200 rpm (LS-X AppliTek, Nazareth, Belgium). Cultures were sampled after 24 h of growth for COS analysis and OD_{600} measurement.

Site-directed mutagenesis and plasmid construction

All plasmids constructed in this study are listed in Table S1. All coding, promoter, 5′UTR and transcriptional terminator sequences are listed in Table S5. All plasmids used in this study were constructed using Circular Polymerase Extension Cloning (CPEC) (Quan and Tian, 2009). All plasmids containing the *nod*C gene from *R. sp.* GRH2 and *S. fredii* USDA 191 are medium-copy vectors (pET22b) with an origin of replication from pBR322 (Coussement et al., 2017) and an ampicillin resistance marker. Both plasmid constructions were based on the pCXhP14-mKate2 expression vector (P_{14} promoter and RBS (De Mey et al., 2007; Shcherbo et al., 2007)) in which the mKate2 gene was replaced with a gene coding for the chitin oligosaccharide synthase *Rs*NodC and *Sf*NodC, respectively. All plasmids consisting of the *nod*C gene from *S. meliloti* strain 1021 are high-copy vectors origination from plasmid pUC57 with a pMB1 origin of replication (ori) and an ampicillin resistance marker (Coussement et al., 2017). These plasmids also carries an operon devoid of any coding sequences, which comprises the constitutive P_{22} promoter and RBS from (Aerts et al., 2011; Shcherbo et al., 2007). The coding sequence for the chitin oligosaccharide synthase *Sm*NodC was placed in this operon. Site-directed mutagenesis, one- and multi-site, to introduce specific point mutations was performed as described by (Liu and Naismith, 2008). The sequences of the two chitin oligosaccharide synthases (*Rs*NodC and *Sm*NodC) were obtained from *R. sp.* GRH2 (*Rs*NodC, Genbank access code: AJW76243) and from *S. meliloti* strain 1021 (*Sm*NodC, Genbank access code: AAK65131.1). The sequence of the chitin oligosaccharide synthase from *S. fredii* USDA 191 (*Sf*NodC) was obtained through genome walking with degenerate primers (Table S6) and subsequent sequencing.

COS sample preparation and analysis

Strains were cultured in biological triplicate as described above, unless stated otherwise. At stationary phase, 1.5 mL samples were taken, centrifuged at 18,000 g for 15 min and pellets stored at −80 °C. OD_{600} was measured using a V-630Bio spectrophotometer (Jasco, Easton, UK) after 24 h. To stay within the linear range for measuring OD_{600} (0.1–1.0) samples were diluted in fresh MM. Pellets were resuspended in 250 μL 60 % acetonitrile in water for cell lysis and subsequent intracellular COS detection. This resuspension was vortexed and centrifuged at 18,000 g for 15 min. The supernatant was stored at −80∘C prior to analysis. COS samples were analysed on a Shimadzu HPLC system (Shimadzu, Jette, Belgium) or a Waters ACQUITY UPLC system (Waters, Milford, Masssechusetts, USA). Both were connected to an ELSD detector. Fully acetylated chitintetraose, −pentaose, −hexaose and −heptaose congeners were separated by hydrophilic interaction chromatography (HILIC) using an ACQUITY UPLC BEH Amide 1.7 μm column (2.1 x 100 mm) and a Kinetix 2.6 μm HILIC 100A column (2.6 μm, 4.6 mm x 150 mm; Phenomenex, Utrecht, The Netherlands) for respectively UPLC or HPLC analysis. Chitintetraose, −pentaose and −hexaose standards were used (Neogen, Ireland). Process details, flow rate and elution profile for HPLC analysis were identical as described in Table S3. Process details, flow rate and elution profile for UPLC analysis are summarised in Table S2.

Mass spectrometry analysis

Detection of samples containing chitinheptaose, were analysed by HPLC-MS (Agilent 1260 HPLC-MS, electrospray ionisation (ESI+), single quadrupole MS detector) using a X-Bridge BEH Amide 2.5 μm 3.0x100 mm XP Column (Waters), 5 μL injection, and isocratic elution at 60 °C with acetonitrile/water 65:35 v/v, 0.1 % formic acid, at a flow rate of 0.4 ml/min. First, a SCAN mode analysis was performed to detect the different masses of the samples (increasing the *m/z* to 1600). Secondly a SIM mode was performed looking for the DP7, DP6 and DP5 (N) (SIM for [M+H]+, *m/z* 1441,4, 1238,2 and 993).

Data and statistical analysis

Data were analysed using pandas (https://www.pandas.pydata.org) unless stated otherwise. Chromatograms were analysed using the OpenChrom v1.1 software packages. Final COS titers were determined based on a calibration curve. For determination of specific COS production, COS titers were corrected for biomass by OD_{600} measurements. Error bars represent the calculated standard deviations of three biological replicates. Pairwise comparisons between different strains were done by a two-sided T-test using the scipy.stats package in Python. One-way ANOVA was performed using the scipy.stats package in Python. In all cases a significance level of 0.05 was applied. All statistical tests (ANOVA) and p-values are depicted in Table S9.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: A patent application (WO2023089009A1) for protecting the method to synthesize chitin oligosaccharides has been filed by Ghent University and Instituto Químico de Sarrià with Chiara Guidi, Xevi Biarnés, Antoni Planas, Jo Maertens and Marjan De Mey named as inventors.

Data availability

No data was used for the research described in the article.

Acknowledgements

C. Guidi acknowledges FWO (Fonds Wetenschappelijk Onderzoek) for the PhD grant (1S16017N). This research was also supported by the BOF-IOP project "MLSB" (BOF16/ IOP/040) and BOF-IOP "Immuno-Keys" (BOF/IOP/2022/069) of the Bijzonder Onderzoeksfonds and the FWO research project "SynSysBio4COS" (G0B8118) of the Research Foundation—Flanders (FWO) (to M. De Mey). This work was also funded by Grants PID2019-104350RB-I00 and PID2022-138252OB-I00 from the Ministry of Science and Innovation (MICINN), Spain (to A. Planas).

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.crbiot.2024.100255.

References

Aam, B.B., Heggset, E.B., Norberg, A.L., Sø, M., Vå, K.M., 2010. Production of chitooligosaccharides and their potential applications in medicine. Mar. Drugs 8, 1482–1517.

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., Bodenstein, S.W., Evans, D.A., Hung, C.C., O'Neill, M., Reiman, D., Tunyasuvunakool, K., Wu, Z., Žemgulytė, A., Arvaniti, E., Beattie, C., Bertolli, O., Bridgland, A., Cherepanov, A., Congreve, M., Cowen-Rivers, A.I., Cowie, A., Figurnov, M., Fuchs, F.B., Gladman, H., Jain, R., Khan, Y.A., Low, C.M.R., Perlin, K., Potapenko, A., Savy, P., Singh, S., Stecula, A., Thillaisundaram, A., Tong, C., Yakneen, S., Zhong, E.D., Zielinski, M., Žídek, A., Bapst, V., Kohli, P., Jaderberg, M., Hassabis, D., Jumper, J.M., 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630, 493–500.

Aerts, D., Verhaeghe, T., De Mey, M., Desmet, T., Soetaert, W., 2011. A constitutive expression system for high-throughput screening. Eng. Life Sci. 11, 10–19.

Akhlaghi, S.P., Berry, R.C., Tam, K.C., 2013. Surface modification of cellulose nanocrystal with chitosan oligosaccharide for drug delivery applications. Cellul. 20, 1747–1764.

Alsina, C., Sancho-Vaello, E., Aranda-Martínez, A., Faijes, M., Planas, A., 2021. Auxiliary active site mutations enhance the glycosynthase activity of a GH18 chitinase for polymerization of chitooligosaccharides. Carbohydr. Polym. 252.

Aranaz, I., Acosta, N., Civera, C., Elorza, B., Mingo, J., Castro, C., Gandía, M. d. l. L., Caballero, A.H., 2018. Cosmetics and cosmeceutical applications of chitin, chitosan and their derivatives. Polymers (Basel) 10.

Basa, S., Nampally, M., Honorato, T., Das, S.N., Podile, A.R., El Gueddari, N.E., Moerschbacher, B.M., 2020. The pattern of acetylation defines the priming activity of chitosan tetramers. J. Am. Chem. Soc. 142, 1975–1986.

Berendsen, H.J.C., van der Spoel, D., van Drunen, R., 1995. GROMACS: A message-passing parallel molecular dynamics implementation. Comput. Phys. Commun. 91, 43–56.

Burley, S.K., Berman, H.M., Kleywegt, G.J., Markley, J.L., Nakamura, H., Velankar, S., 2017. Protein data bank (PDB): The single global macromolecular structure archive. Methods Mol. Biol. 1607, 627–641.

Cantarel, B.I., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., Henrissat, B., 2009. The carbohydrate-active EnZymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res. 37, 233–238.

Chen, W., Cao, P., Liu, Y., Yu, A., Wang, D., Chen, L., Sundarraj, R., Yuchi, Z., Gong, Y., Merzendorfer, H., Yang, Q., 2022. Structural basis for directional chitin biosynthesis. Nature 610.

Consortium, T.U., 2008. The Universal Protein resource (UniProt). Nucleic Acids Res 36, D190–D195.

Coussement, P., Bauwens, D., Maertens, J., De Mey, M., 2017. Direct combinatorial pathway optimization. ACS Synth. Biol. 6, 224–232.

Coussement, P., 2016. Pathways to chito-oligosaccharides production: Integrating synthetic biology and metabolic engineering for pathway optimization. Ph.D. thesis Ghent University.

Daura, X., Gademann, K., Jaun, B., Gunsteren, W.F.V., Mark, A.E., 1999. Peptide folding: When simulation meets experiment. Angew. Chem. Int. Ed. 38, 236–240.

De Mey, M., Maertens, J., Lequeux, G.J., Soetaert, W.K., Vandamme, E.J., 2007. Construction and model-based analysis of a promoter library for *E. coli*: An indispensable tool for metabolic engineering. BMC Biotech. 7, 1–14.

Dong, H., Wang, Y., Zhao, L., Zhou, J., Xia, Q., Jiang, L., Fan, L., 2014. Purification of DP 6 to 8 chitooligosaccharides by nanofiltration from the prepared chitooligosaccharides syrup. Bioresour. Bioprocess. 1, 1–12.

Dorfmueller, H.C., Ferenbach, A.T., Borodkin, V.S., Van Aalten, D.M.F., 2014. A structural and biochemical model of processive chitin synthesis. J. Biol. Chem. 289, 23020–23028.

Duan, X., Tian, G., Chen, D., Yang, J., Zhang, L., Li, B., Huang, L., Zhang, D., Zheng, P., Mao, X., Yu, J., He, J., Huang, Z., Yu, B., 2020. Effects of diet chitosan oligosaccharide on performance and immune response of sows and their offspring. Livest. Sci. 239.

Feng, F., Sun, J., Radhakrishnan, G.V., Lee, T., Bozsóki, Z., Fort, S., Gavrin, A., Gysel, K., Thygesen, M.B., Andersen, K.R., Radutoiu, S., Stougaard, J., Oldroyd, G.E.D., 2019. A combination of chitooligosaccharide and lipochitooligosaccharide recognition promotes arbuscular mycorrhizal associations in *Medicago truncatula*. Nat. Commun. 10.

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: Interactive sequence similarity searching. Nucleic Acids Res. 39, 29–37.

Fiser, A., Do, R.K.G., Šali, A., 2000. Modeling of loops in protein structures. Protein Sci. 9, 1753–1773.

Fliegmann, J., Bono, J., 2015. Lipo-chitooligosaccharidic nodulation factors and their perception by plant receptors. Glycoconj. J. 455–464.

Gohlke, S., Muthukrishnan, S., Merzendorfer, H., 2017. In vitro and in vivo studies on the structural organization of Chs3 from *Saccharomyces cerevisiae*. Int. J. Mol. Sci. 18, 1–17.

Guidi, C., Biarnés, X., Planas, A., De Mey, M., 2023. Controlled processivity in glycosyltransferases: A way to expand the enzymatic toolbox. Biotechnol. Adv. 63.

Hamed, I., Özogul, F., Regenstein, J.M., 2016. Industrial applications of crustacean by-products (chitin, chitosan, and chitooligosaccharides): A review. Trends Food Sci. Technol. 48, 40–50.

Hao, W., Li, K., Li, P., 2021. Review: Advances in preparation of chitooligosaccharides with heterogeneous sequences and their bioactivity. Carbohydr. Polym. 252, 117206.

Hayafune, M., Berisio, R., Marchetti, R., Silipo, A., Kayama, M., Desaki, Y., 2014. Chitin-induced activation of immune signaling by the rice receptor CEBiP relies on a unique sandwich-type dimerization. PNAS 111, E404–E413.

Hsin, J., Arkhipov, A., Yin, Y., Stone, J.E., Schulten, K., 2008. Using VMD - An introductory tutorial. Bone.

Humphrey, W., Dalke, A., Schulten, K., 1996. Visual molecular dynamics. J. Mol. Graph. 14, 33–38.

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, 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., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589.

Kamst, E., Bakkers, J., Quaedvlieg, N.E.M., Pilling, J., Kijne, J.W., Lugtenberg, B.J.J., Spaink, H.P., 1999. Chitin oligosaccharide synthesis by *rhizobia* and zebrafish embryos starts by glycosyl transfer to O4 of the reducing-terminal residue. Biochemistry 38, 4045–4052.

Kamst, E., Breek, C.K.D., 2000. Functional analysis of chimeras derived from the *Sinorhizobium meliloti* and *Mesorhizobium loti nod*C genes identifies regions controlling chitin oligosaccharide chain length. Mol. Genet. Genomics 264, 75–81.

Kamst, E., Herman, P., 1999. Functional domains in the chitin oligosaccharide synthase NodC and related β-polysaccharide synthases. Trends Glycosci. Glycotechnol. 11, 187–199.

Kamst, E., Pilling, J., Raamsdonk, L.M., Lugtenberg, B., Spaink, H.P., 1997. *Rhizobium* nodulation protein NodC is an important determinant of chitin oligosaccharide chain length in Nod factor biosynthesis. J. Bacteriol. 179, 2103–2108.

Kosugi, T., Hayashi, S., 2012. Crucial role of protein flexibility in formation of a stable reaction transition state in an α-amylase catalysis. J. Am. Chem. Soc. 134, 7045–7055.

Krieger, E., Koraimann, G., Vriend, G., 2002. Increasing the precision of comparative models with YASARA NOVA - A self-parameterizing force field. Proteins: Structure, Function and Genetics 47, 393–402.

Lairson, L.L., Henrissat, B., Davies, G.J., Withers, S.G., 2008. Glycosyltransferases: Structures, functions, andm. Annu. Rev. Biochem. 77, 521–555.

Larsson, P., Wallner, B., Lindahl, E., Elofsson, A., 2008. Using multiple templates to improve quality of homology models in automated homology modeling. Protein Sci. 17, 990–1002.

Liaqat, F., Eltem, R., 2018. Chitooligosaccharides and their biological activities: A comprehensive review. Carbohydr. Polym. 184, 243–259.

Liu, C.T., Layfield, J.P., Stewart, R.J., French, J.B., Hanoian, P., Asbury, J.B., Hammes-Schiffer, S., Benkovic, S.J., 2014. Probing the electrostatics of active site microenvironments along the catalytic cycle for *Escherichia coli* dihydrofolate reductase. J. Am. Chem. Soc. 136, 10349–10360.

Liu, H., Naismith, J.H., 2008. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. BMC Biotech. 8, 91.

Martin-Harris, M.H., Bartley, P.A., Morley, A.A., 2010. Gene walking using sequential hybrid primer polymerase chain reaction. Anal. Biochem. 399, 308–310.

Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K., Olson, A. J., 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comput. Chem. 19, 1639–1662.

Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., Olson, A.J., 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J. Comput. Chem. 30, 2785–2791.

Ojeda-Hernández, D.D., Canales-Aguirre, A.A., Matias-Guiu, J., Gomez-Pinedo, U., Mateos-Díaz, J.C., 2020. Potential of chitosan and its derivatives for biomedical applications in the central nervous system. Front. Bioeng. Biotechnol. 8, 1–15.

Osho, S.O., Adeola, O., 2020. Chitosan oligosaccharide supplementation alleviates stress stimulated by in-feed dexamethasone in broiler chickens. Poult. Sci. 99, 2061–2067.

Ouyang, Q.Q., Zhao, S., Li, S.D., Song, C., 2017. Application of chitosan, chitooligosaccharide, and their derivatives in the treatment of Alzheimer's disease. Mar. Drugs 15, 1–15.

Pei, J., Grishin, N.V., Road, F.P., 2014. Multiple Sequence Alignment Methods. Methods Mol. Biol. 1079, 263–271.

Petrova, P., Koc, J., 1999. Potential energy hypersurfaces of nucleotide sugars : Ab initio calculations, force-field parametrization, and exploration of the flexibility. J. Am. Chem. Soc. 121, 5535–5547.

Quan, J., Tian, J., 2009. Circular polymerase extension cloning of complex gene libraries and pathways. PLoS One 4, e6441.

Rakkhumkaew, N., Pengsuk, C., 2018. Chitosan and chitooligosaccharides from shrimp shell waste: Characterization, antimicrobial and shelf life extension in bread. Food Sci. Biotechnol. 27, 1201–1208.

Rao, M.S., Chander, R., Sharma, A., 2008. Synergistic effect of chitooligosaccharides and lysozyme for meat preservation. LWT Food Sci. Technol. 41, 1995–2001.

Ren, Z., Chhetri, A., Guan, Z., Suo, Y., Yokoyama, K., Lee, S.Y., 2022. Structural basis for inhibition and regulation of a chitin synthase from *Candida albicans*. Nat. Struct. Mol. Biol. 29, 653–664.

Romero-García, J., Francisco, C., Biarnés, X., Planas, A., 2013. Structure-function features of a mycoplasma glycolipid synthase derived from structural data integration, molecular simulations, and mutational analysis. PLoS One 8, 1–14.

Saitou, N., Nei, M., 1987. The neighbour-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evo. 4, 406–425.

Sali, A., Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815.

Samain, E., Chazalet, V., Geremia, R.A., 1999. Production of O -acetylated and sulfated chitooligosaccharides by recombinant *Escherichia coli* strains harboring different combinations of nod genes. J. Biotechnol. 72, 33–47.

Sambrook, J., Fritisch, E.F., Maniatis, T., Al., E., 1989. Molecular cloning: A laboratory manual. (Cold spring harbor laboratory press) No. Ed. 2.

Saxena, I.M., Brown, R.M., 1997. Identification of cellulose synthase(s) in higher plants: Sequence analysis of processive β-glycosyltransferases with the common motif 'D, D, D35Q(R, Q)XRW'. Cellul. 4, 33–49.

Schultze, M., Kondorosi, A., 1996. The role of Nod signal structures in the determination of host specificity in the *Rhizobium*-legume symbiosis. World J. Microbiol. Biotechnol. 12, 137–149.

Seeliger, D., De Groot, B.L., 2010. Ligand docking and binding site analysis with PyMOL and Autodock/Vina. J. Comput. Aided Mol. Des. 24, 417–422.

Shcherbo, D., Merzlyak, E.M., Chepurnykh, T. V, Fradkov, A.F., Ermakova, G. V, Solovieva, E. a, Lukyanov, K. a, Bogdanova, E. a, Zaraisky, A.G., Lukyanov, S., Chudakov, D.M., 2007. Bright far-red fluorescent protein for whole-body imaging. Nat Methods 4, 741–746.

Söding, J., Biegert, A., Lupas, A.N., 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 33, 244–248.

Trott, O., Olson, A.J., 2009. Software news and update AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem. 31, 455–461.

Wan, J., Zhang, X.C., Neece, D., Ramonell, K.M., Clough, S., Kim, S.Y., Stacey, M.G., Stacey, G., 2008. A LysM receptor-like kinase plays a critical role in chitin signaling and fungal resistance in *Arabidopsis*. Plant Cell 20, 471–481.

Wan, J., Jiang, F., Xu, Q., Chen, D., Yu, B., Huang, Z., Mao, X., Yu, J., He, J., 2017. New insights into the role of chitosan oligosaccharide in enhancing growth performance, antioxidant capacity, immunity and intestinal development of weaned pigs. RSC Adv. 7, 9669–9679.

Weyer, R., Hellmann, M.J., Hamer-Timmermann, S.N., Singh, R., Moerschbacher, B.M., 2022. Customized chitooligosaccharide production—controlling their length via engineering of *rhizobial* chitin synthases and the choice of expression system. Front. Bioeng. Biotechnol. 10, 1–17.

Wolinsky, J.B., Colson, Y.L., Grinstaff, M.W., 2012. Local drug delivery strategies for cancer treatment: Gels, nanoparticles, polymeric films, rods, and wafers. J. Control. Release 159, 3333.

Xu, X., Sun, L., Zhou, L., Cheng, Y., Cao, F., 2020. Functional chitosan oligosaccharide nanomicelles for topical ocular drug delivery of dexamethasone. Carbohydr. Polym. 227, 115356.

Yang, J., Zhang, Y., 2015. Protein Structure and function prediction using I-TASSER. Curr Protoc Bioinformatics 52, 5.8.1-5.815.

Yu, S., Xu, X., Feng, J., Liu, M., Hu, K., 2019. Chitosan and chitosan coating nanoparticles for the treatment of brain disease. Int. J. Pharm. 560, 282–293.