



HAL
open science

In silico and in vitro analysis of an *Aspergillus niger* chitin deacetylase to decipher its subsite sugar preferences

Martin Bonin, Lisanne Hameleers, Lea Hembach, Thomas Roret, Stefan Cord-Landwehr, Gurvan Michel, Bruno M Moerschbacher

► **To cite this version:**

Martin Bonin, Lisanne Hameleers, Lea Hembach, Thomas Roret, Stefan Cord-Landwehr, et al.. In silico and in vitro analysis of an *Aspergillus niger* chitin deacetylase to decipher its subsite sugar preferences. *Journal of Biological Chemistry*, 2021, pp.101129. 10.1016/j.jbc.2021.101129 . hal-03354312

HAL Id: hal-03354312

<https://hal.sorbonne-universite.fr/hal-03354312v1>

Submitted on 24 Sep 2021

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.

Journal Pre-proof

In silico and in vitro analysis of an *Aspergillus niger* chitin deacetylase to decipher its subsite sugar preferences

Martin Bonin, Lisanne Hamелеers, Lea Hembach, Thomas Roret, Stefan Cord-Landwehr, Gurban Michel, Bruno M. Moerschbacher

PII: S0021-9258(21)00930-3

DOI: <https://doi.org/10.1016/j.jbc.2021.101129>

Reference: JBC 101129

To appear in: *Journal of Biological Chemistry*

Received Date: 28 May 2021

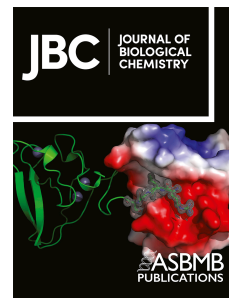
Revised Date: 21 August 2021

Accepted Date: 26 August 2021

Please cite this article as: Bonin M, Hamелеers L, Hembach L, Roret T, Cord-Landwehr S, Michel G, Moerschbacher BM, In silico and in vitro analysis of an *Aspergillus niger* chitin deacetylase to decipher its subsite sugar preferences, *Journal of Biological Chemistry* (2021), doi: <https://doi.org/10.1016/j.jbc.2021.101129>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology.



In silico and in vitro analysis of an *Aspergillus niger* chitin deacetylase to decipher its subsite sugar preferences

Running title: *In silico* and *in vitro* analysis of AngCDA

Martin Bonin¹, Lisanne Hameleers^{1*}, Lea Hembach¹, Thomas Roret², Stefan Cord-Landwehr¹, Gervan Michel³, Bruno M. Moerschbacher^{1†}

¹ University of Münster, Institute for Biology and Biotechnology of Plants, 48143 Münster, Germany

² Sorbonne Université, CNRS, FR 2424, Station Biologique de Roscoff (SBR), 29680 Roscoff, France

³ Sorbonne Université, CNRS, Integrative Biology of Marine Models (LBI2M), Station Biologique de Roscoff (SBR), 29680 Roscoff, France

*Present address: Department of Bioproduct Engineering, University of Groningen, 9747 AG Groningen, The Netherlands

†Corresponding author e-mail: moersch@uni-muenster.de

Abstract

Chitin deacetylases (CDAs) are found in many different organisms ranging from marine bacteria to fungi and insects. These enzymes catalyze the removal of acetyl groups from chitinous substrates generating various chitosans, linear co-polymers consisting of *N*-acetylglucosamine (GlcNAc) and glucosamine (GlcN). CDAs influence the degree of acetylation (DA) of chitosans as well as their pattern of acetylation (PA), a parameter which was recently shown to influence the physicochemical properties and biological activities of chitosans. The binding site of CDAs typically consists of around four subsites, each accommodating a single sugar unit of the substrate. It has been hypothesized that the subsite preferences for GlcNAc or GlcN units play a crucial role in the acetylation pattern they generate, but so far, this characteristic was largely ignored, and still lacks structural data on the involved residues.

Here, we determined the crystal structure of an *Aspergillus niger* CDA (AngCDA). Then, we used molecular dynamics simulations, backed up with a variety of *in vitro* activity assays using different well-defined polymeric and oligomeric substrates, to study this CDA in detail. We found that AngCDA strongly prefers a GlcNAc sugar unit at its -1 subsite and shows a weak GlcNAc preference at the other non-catalytic subsites, which was apparent both when de- and *N*-acetylating oligomeric substrates. Overall, our results show that the combination of *in vitro* and *in silico* methods used here enables the detailed analysis of CDAs, including their subsite preferences, which could influence their substrate targets and the characteristics of chitosans produced by these species.

Keywords: chitosan, carbohydrate function, carbohydrate biosynthesis, crystal structure, enzyme mechanism, molecular dynamics, molecular docking

1 Introduction

Chitosans are highly versatile and promising biopolymers, consisting of β -1,4 linked *N*-acetylglucosamine (GlcNAc, A) and glucosamine (GlcN, D) units. They are found in the cell wall of several pathogenic fungi, possibly masking the fungal chitin to evade the host's immune system^{1,2,3,4}, but also in other, non-pathogenic fungi⁵. Furthermore, chitosans can be used in a variety of applications, for example in agriculture, where they show plant strengthening and plant protecting effects, or in the medical field in drug delivering nanoparticles^{6,7,8}. Their utilization in these areas as well as, presumably, their biological functions highly depend on their physicochemical properties, which are known to be strongly influenced by the percentage of acetylated units (degree of acetylation, DA) and the length of the polymer (degree of polymerization, DP)^{9,10}. Beyond these two parameters, whose control was the critical step in developing reliably performing second generation chitosans, the pattern of acetylation (PA) is currently gaining increasing attention¹¹. A deep influence of the distribution of GlcNAc and GlcN units along the chain, ranging from alternating to random and

block-wise, has recently been shown for the physicochemical properties as well as the biological activities of partially acetylated chitosans^{12,13}. The DP and DA can be controlled in chemical chitosan production, when either highly acetylated chitin polymers are partially deacetylated e.g. using sodium hydroxide at high temperatures or fully deacetylated polyglucosamines are partially *N*-acetylated e.g. using acetic anhydride^{14,15}. However, regarding the PA, only random distributions can be achieved using chemical production methods¹⁶. Therefore, enzymatic production routes using chitin deacetylases (CDAs) have been proposed, that may yield polymers with defined non-random PA and thus, may open the way to third generation chitosans^{12,17}.

According to the carbohydrate-active enzymes (CAZy) database, CDAs (EC 3.5.1.41) are classified in the carbohydrate esterase 4 (CE4) family, together with four other deacetylase and esterase activities¹⁸. *In vivo*, CDAs catalyze the release of acetate from *N*-acetylglucosamine units of chitins and chitosans (forward mode) but *in vitro*, they are also able to catalyze the reverse reaction, thus *N*-acetylating chitosan polymers and oligomers (reverse mode)¹⁹. They are often found in multigene families suggesting different physiological functions such as the production of cell wall chitosan or the deacetylation of oligomers released from the cell wall^{2,3,20}. All CE4 enzymes are metalloenzymes that share a similar fold and five conserved motifs forming the substrate binding site. This binding site is comprised of several subsites, with subsite 0 binding the sugar unit being deacetylated and the minus and plus subsites accommodating the units towards the non-reducing and reducing end, respectively²¹. Depending on the number and accessibility of these subsites, the enzymes can act on oligomeric or polymeric substrates and thereby deacetylate either one or several sugar units²². While it seems that CDAs keep their regioselectivity on chitooligosaccharides (COS) and partially acetylated COS (paCOS)^{23,24}, they appear to generate different patterns on polymeric substrates when acting in forward¹⁷ or reverse mode¹². For the latter, the subsite preferences for an acetylated or deacetylated unit was proposed to play a crucial role¹¹.

However, subsite preferences have rarely been studied in CDAs. Only a *Colletotrichum lindemuthianum* and a *Cryptococcus neoformans* CDA (CICDA and CnCDA4) were described to have a clear preference for a GlcNAc unit at subsite -2 or a GlcN unit at subsite -1, respectively^{2,21}. So far, CDAs and other CE4 enzymes were mainly tested for their activity on different polymeric substrates like chitins, chitosans, acetyl xylans, peptidoglycans and their oligomeric counterparts. In recent years, for an increasing number of CE4 enzymes, the mode of action, referring to the different products that are generated over time, was investigated as well^{2,4,25,26,27,28}. Computational methods, such as sequence and structure alignments as well as homology modelling and docking studies were used to complement these *in vitro* assays^{2,27,28}. To our knowledge, only a few CDAs were additionally studied by molecular dynamics (MD) simulations^{22,29} and only for one CDA, from *Cryptococcus laurentii*, MD simulations were used for a more in-depth analysis³⁰. The latter, however, was not studied *in vitro*.

In this study, we analyzed AngCDA, an *Aspergillus niger* CDA that is strongly expressed in the mutant *scl-2*, which in contrast to the wild type forms sclerotia, a survival structure for harsh environmental conditions and a prerequisite for sexual reproduction³¹. We solved the 3D structure by X-ray crystallography and analyzed the enzyme using classical *in vitro* assays. In addition, we performed extensive MD simulations, especially focusing on the subsite preferences, which we then validated *in vitro*.

2 Results

2.1 Sequence analysis and initial protein characterization

The protein sequence of AngCDA reveals the presence of a CE4 superfamily domain (residues 33-223) including the zinc binding site and all four catalytic residues previously described for CICDA³². Further bioinformatic predictions, including signal peptide, GPI anchor and transmembrane domains, suggest that the enzyme has a signal peptide (residue 1-19) and is secreted into the extracellular space, but is not anchored or attached to the membrane (see fig. S1-S4).

For protein expression in *E. coli* and subsequent purification, two expression constructs without the predicted signal peptide were generated, one with an N-terminal pelB sequence for protein secretion

and a C-terminal Strep-tag II for purification and one without a pelB sequence and both an N- and a C-terminal Strep-tag II. The X-ray crystal structure was determined with the first construct, while the second construct was used for all activity assays. This latter construct was used as a (lab-internal) standard construct for better comparability between different CDAs. For the same reason, all experiments were carried out at 37 °C and pH 7. These parameters allowed us to determine the optimum temperature and pH as 50 °C and 8, respectively, using chitotetraose (A4) as a substrate (fig. S5).

The enzyme activity was first tested on different polymeric and oligomeric substrates, including chitosans with different degrees of acetylation, insoluble α - and β -chitin, colloidal chitin (fig. S6) and COS of DP1-6 (fig. 1). AngCDA showed only weak activity on α - and β -chitin, while the activity slightly increased on colloidal chitin, probably due to a larger accessible surface area. Overall, the activity on these crystalline substrates was low, as generally seen with other CDAs, too^{4, 26, 33, 34}. When chitosans with DA12, DA32 and DA46 were used as water soluble substrates 62 %, 60 % and 48 % of the available acetyl groups were removed during 24 hours of incubation, respectively. While on chitosan with DA46, the Δ DA increase was linear within the first 4 h, it slowed down towards 24 h, an effect already visible at earlier time points for the other chitosans tested.

When using COS of different DP as a substrate, AngCDA was inactive on monomeric GlcNAc (A1), but active on DP ≥ 2 with an increase towards larger substrates (fig. 1). For all substrates (An), all intermediate products were produced in succession until only one GlcNAc unit was left (A1Dn-1). The fully deacetylated product (Dn) was only produced in very low amounts. Considering that for all tests the same enzyme and substrate concentrations were used, it is striking that the A1Dn-1 concentration after 72 h always ranged between 60 % and 80 %, apparently regardless of the number of deacetylations needed until that point. The more detailed time courses for A4 and chitopentaose (A5) reveal that the first deacetylation was the fastest with An-1D1 occurring as the main product after 4 h, whereas An-2D2 was the main product 8 h later for both substrates. This can also be seen when looking at the corresponding peak of the first product (fig. 1, shown in red), which is much sharper compared to the following ones which get wider with decreasing degree of acetylation of the corresponding product. The overall activity increased towards larger DPs. However, upon closer inspection of the initial slope of the curves, it increased only up to A5, while the initial slope of A6 is similar to that of A5 (see fig. 1 smaller graph in overall activity).

2.2 Crystal structure and multiple structure alignment

To further elucidate the substrate binding site, the X-ray crystal structure of AngCDA was determined at 1.81 Å resolution. The enzyme crystallized in the space group *P432* with one monomer per asymmetric unit. This oligomeric state correlates with the results obtained by gel filtration and DLS measurements (28.6 ± 6.5 kDa). AngCDA adopts a distorted (α/β)₈ fold, as typically found in CE4 enzymes. An intramolecular disulfide bridge (C36 - C226), which tethers the N- and C-terminal ends, stabilizes the structure as already observed in the crystal structure of chitin deacetylases from the fungal pathogens *Colletotrichum lindemuthianum* and *Aspergillus nidulans* (pdb entries 2IW0 and 2Y8U)^{32, 34}. The conserved Asp-His-His triad (N ϵ 2-H97, N ϵ 2-H101 and O δ 1-D48 atoms) plus a malonate ion (O6 and O7 atoms) from the crystallization solution and a water molecule coordinate a zinc(II) ion to form an octahedral coordination geometry with a metal-ligand distance of 2.2 Å. The malonate ion is found at the expected subsite 0, where the acetyl group of the chitinous substrate is normally placed and hydrolyzed (fig. 2 B, C).

To compare the structure to other CE4 enzymes, a multiple structure alignment was generated including only those enzymes for which CDA activity has been described in the literature (fig. 2 A). All five conserved motifs can be found in AngCDA, with motif 1 harboring the catalytic base (D47) and the metal-binding aspartate (D48) and motif 2 containing both metal binding histidines (H97, H101). Motif 3 includes an arginine (R135) which properly orients the catalytic aspartate (D47) and a tyrosine (Y138) forming a hydrogen bond with the acetyl oxygen at subsite 0. Motif 4 contains an aspartate (D165) that enables the protonation of the catalytic histidine (H195) and motif 5 contains a

leucine (L193) forming one side of the hydrophobic pocket for the acetamido's methyl group at subsite 0 and the catalytic acid (H195).

All residues mentioned are conserved in all CE4 enzymes included in the alignment, except for the aspartate (D165 in AngCDA), which is a proline in the *Vibrio cholerae* CDA (VcCDA). Both on the sequence and on the structural level, VcCDA differs most from all the other CE4 enzymes (see tab. S2 & S3). Its active site is surrounded by six loops, defined by the subsite capping model²². These loops which are so prominent in VcCDA are very small in the other enzymes, with the exception of a short loop 1 in CICDA and short loops 4 in AngCDA, AnCDA and CICDA (see grey boxes in fig. 2 A). The subsite capping model suggests that the specific pattern of acetylation generated by VcCDA on oligomeric substrates can be explained by these loops positioning the substrate in a certain way. Of the CE4 enzymes shown in the alignment, solely VcCDA is reported to deacetylate only one sugar unit in the substrate, namely the penultimate unit from the nonreducing end²². As far as sequence data of the products generated by the other enzymes are available (AnCDA³⁴, CICDA³², ArCE4A³³, BsPdaC²⁷, SICE4³⁵), they are all described to deacetylate several positions in oligomeric substrates. Since for most of these enzymes, different methods and protocols were applied to determine the PA of their products, they are not easily comparable. Nonetheless, their mode of action clearly differs from what is described for VcCDA, since they have a much more open binding site and thus, their mode of action cannot or only partially be explained by the subsite capping model alone. We therefore assume that in addition to the loops, several key residues along the binding site contribute to define the mode of action and regioselectivity of these more open CE4 enzymes.

2.3 Docking studies and MD simulations with A4, A5 and A3D1

To identify amino acids along the AngCDA binding site which interact with the substrate, A4, A5 and different mono- deacetylated tetramers (A3D1) were docked into the active site of the enzyme. Binding modes spanning from the hypothetical -3 to the hypothetical +3 subsite were chosen, resulting in four different binding modes for A4, sequentially placing each unit at subsite 0, and three different binding modes for A5, placing all internal units at subsite 0. For the A3D1 substrates (DAAA, ADAA, AADA, AAAD), different binding modes were chosen to create comparable binding modes to A4 (see fig. 3). Since the amino groups of GlcN units are mainly not protonated at pH 7, these substrates were uncharged³⁶. For each binding mode and substrate, three different conformations with the highest docking scores were chosen from the *in silico* docking. The corresponding docking scores can be found in table S1.

From hereon, the binding modes will be indicated by the distance of the non-reducing and reducing end sugar units to subsite 0, respectively. For example, an A4 bound with its non-reducing end unit at subsite 0 will be denoted as binding mode [0,+3]. This however, does not suggest that a subsite +3 is actually existing. In the following, the term 'subsite' refers to the region of the enzyme which interacts with the corresponding sugar unit of the substrate, without implying that these interactions substantially contribute to substrate binding.

The selected complexes from the substrate dockings of all substrates in different binding modes were subjected to molecular dynamics simulations. As a first indication of the substrate stability in the binding site, the root mean square fluctuation (rmsf) was calculated for each sugar unit and for the complete substrate (fig. 3 A). At first, it appears that the fluctuation of each sugar unit solely depends on the subsite at which it is situated and is mostly independent of the substrate length, the binding mode and even the type of sugar. The sugar unit bound at subsite 0 always shows the lowest fluctuation, while fluctuation increases towards both the plus and minus subsites. The average rmsf of the substrate and the fluctuation of the different sugar units suggest that sugar units bound at the minus subsites fluctuated less compared to those at the plus subsites, resulting in an overall more stable binding for binding modes spanning up to subsite -3. On closer inspection, more differences are visible between the different substrates and the different binding modes. The most visible difference is the increased rmsf for paCOS with a GlcN unit at subsite 0, leading to a higher fluctuation of the whole substrate as well. With GlcNAc at subsite 0, the zinc ion coordinates the acetyl oxygen, the O3 of GlcNAc and the catalytic water. With GlcN at subsite 0, different coordinations were observed. Either zinc coordinated the amino group and/or the O3 of GlcN, the O6 of the neighboring sugar unit

bound at subsite -1 or no part of the substrate. In contrast, a GlcN unit at subsite +1 or +2 seems to slightly decrease the rmsf compared to a GlcNAc unit at the same subsite, while this was not the case at subsite +3. The reason for the reduced rmsf might be the missing acetyl group itself, as acetylated amino groups show more rotational movement as long as they do not form a stable interaction with any residue. In all binding modes, the fluctuation of the sugar unit at subsite -2 differs most strongly from that at all other subsites. It appears to be stabilized if another sugar unit is bound at subsite -3, but destabilized if a deacetylated unit is bound at subsite -1. However, a GlcN unit at subsite -2 itself does not seem to influence the fluctuation.

The fluctuation of the substrate serves as a valuable first comparison of the different substrates in their different binding modes. Still, it is not a quantitative measure for the strength of the interaction between substrate and enzyme. Therefore, the binding energy was estimated using the molecular mechanics generalized Born surface area (mmgbsa) approach, which calculates not only the total binding free energy, but also the contribution of each amino acid and the different sugar units of the substrates. The calculated binding free energy of the sugar units and the total binding free energy for all substrates and binding modes are shown in figure 3 B.

As already indicated by the rmsf results, the calculated binding free energy is the lowest and the binding therefore the strongest, for the subsite 0 sugar unit, and the main differences are not between the different substrates or binding modes, but between the different sugar unit positions. Somewhat contradicting the higher rmsf of the sugar units bound towards the plus subsites, the +1 sugar unit strongly contributes to the binding energy. The highest total binding energy, and therefore the most unfavorable binding, was calculated for binding modes [0,+3] and for the binding modes where a GlcN unit is placed at subsite 0. The latter may indicate that deacetylated products may bind in an unproductive way (although a productive binding mode is strongly preferred), leading to product inhibition. The former can be expected to lead to slower rates of deacetylation of the sugar unit at the non-reducing end, as described for the closely related AnCDA³⁴. Moreover, an acetylated unit appears to be beneficial not only at subsite 0, but also at the other subsites, since the total binding free energy increases in all A3D1 substrates compared to A4 with the same binding mode. This effect is most pronounced if a deacetylated unit was placed at subsite -1, which increased the binding energy from -21.21 kcal/mol to -18.42 kcal/mol and from -27.19 kcal/mol to -21.58 kcal/mol for the binding modes [-1,+2] and [-2,+1], respectively. Together with the slightly increased fluctuations of these substrates, this suggests a preference for acetylated units at subsite -1. Interestingly, for ADAA placed with the GlcN unit at subsite -1, the fluctuation and binding energy at this subsite are comparable to A4 in the same binding mode. However, it appears that the GlcN unit at subsite -1 influences the binding of the neighboring sugar unit at subsite -2. Taking a closer look at these simulations, it turns out that the deacetylated sugar unit at subsite -1 is slightly tilted compared to an acetylated unit at the same position, which forces the sugar unit at subsite -2 to fold out of the binding site, allowing a stronger fluctuation (fig. 4).

To identify interesting residues which contribute to these observed differences, we took a closer look at the amino acids which showed the strongest influence on the binding energy (see below, fig. 6). Additionally, we calculated the average number of hydrogen bonds between substrate and enzyme throughout all simulations and identified the residues with the highest values for each subsite (fig. 5). It should be noted that a simple cut-off for the hydrogen bond angle between donor, acceptor and hydrogen atom and for the distance between donor and acceptor was applied. Thus, no direct assumption can be made about the strength of these bonds. The simulations where a GlcN unit is placed at subsite 0, with more movement in the substrate overall, show quite different interactions and are not discussed further. All other simulations show the most stable hydrogen bonds at subsite 0, with the backbone nitrogen of Tyr138 forming a hydrogen bond with the acetyl oxygen. Moreover, the second oxygen from Asp48, the one that does not coordinate the zinc ion, forms a hydrogen bond with O3 of the GlcNAc unit at subsite 0, which is also coordinated by the zinc ion. At subsites -1 and -2, the Thr197 side and main chain oxygens interact with the O3 and O6 of the sugar unit, respectively. This interaction is strongly reduced at subsite -2 if a GlcN unit is present at subsite -1 or -2 since the orientation of these sugar units differs from the others (see fig. 4). Furthermore, a new hydrogen bond between the Tyr138 side chain oxygen and the O6 of GlcN unit situated at subsite -1

appears, which stabilizes this orientation. At subsite -3, the backbone nitrogen of Ser52 forms a stable hydrogen bond with the acetyl group oxygen. If the acetyl group is missing, only occasional hydrogen bonds with the O4 and O5 were observed. On the plus subsites, mainly Asp162 and Lys164 seem to be involved in hydrogen bond formation with the substrate, while no stable hydrogen bonds are formed with the +3 sugar units. It appears, that the hydrogen bond with Asp162 is more stable if it interacts with the reducing end sugar, which also leads to a decreased binding free energy at the +1 subsite.

All the aforementioned residues belong to those residues that most strongly influence the binding energy (fig. 6). Besides Asp48 and Tyr138, the catalytic Asp47, the metal binding His97 and His101, the hydrophobic Phe139 and the aromatic Tyr166 contribute to the substrate binding at subsite 0. Tyr166, as studied in detail for the equivalent Trp151 of the cryptococcal CDA by Sarkar et al.³⁰, probably stacks with the substrate. As previously described for CICDA³², Phe139 forms a hydrophobic pocket together with Leu193 to accommodate the acetyl methyl group at subsite 0. Leu73 might have a similar role for the acetyl group at the -2 subsite. The positively charged amino acids Arg135 and the catalytic His195 at subsite 0 as well as Lys164 and Lys198 show a positive binding energy, due to their high desolvation penalty. This effect, where the exchange of surrounding water molecules by the substrate is energetically disfavored, is by far the strongest for the zinc ion, leading to a strong increase of the binding energy. Upon closer inspection of the values for the different binding modes, it again becomes visible that those with a GlcN unit at subsite 0 stand out. As expected, residues such as Ser51 or Lys164 mostly influence the binding energy in binding modes where their subsites are occupied. However, it also shows, that one residue cannot be simply ascribed to only a single subsite. Lys164 shows the strongest influence for binding modes [0,+3], [-1,+2], [-1,+3] and [-2,+2], i.e. those involving subsite +2, but also affects, to a weaker extent, binding modes [-2,+1] and [-3,+1], and slightly even [-3,0], i.e. binding modes not involving subsite +2. And even the subsite 0 residues show differences, especially for those binding modes where a terminal unit of the substrate is positioned at subsite 0. Besides the binding modes with a GlcN unit at subsite 0, also ADAa in binding mode [-2,+1] (ADaA, i.e. with a GlcN unit at subsite -1 and an GlcNAc unit at subsite 0, as indicated by the lower case letter a) shows some differences compared to the other [-2,+1] binding modes, especially for Leu73, Tyr166 and Thr197. This seems surprising for Tyr166, as it appears to normally interact with the sugar ring at subsite 0 (see fig. 4). While examining the trajectories with a GlcN unit at subsite -1 more closely, we saw that this tyrosine can rotate towards the minus subsites, occupying the freed-up space of the missing acetyl group.

Overall, many more details can be observed from these simulations, but for space constraints, it is not possible to discuss all of them in detail here. Therefore, all tables generated for the binding free energy calculations and hydrogen bonds are summarized in the two spreadsheets included in the supporting information.

2.4 Mode of action on A4, A5 and D4 studied *in vitro*

To validate the conclusions drawn from the MD simulations and to gain a better understanding of how AngCDA de- and *N*-acetylates its substrates, the enzyme was incubated with A4, A5 (as shown in figure 1) and D4. For A4 and A5, samples were taken at those time points where the different products had the highest relative concentration, to determine their pattern of acetylation and, thus, the mode of action of the enzyme (fig. 7). In both acetylated substrates, the internal units were deacetylated first, step by step generating ADDA and ADDDA, before finally DDDA and DDDDA were produced, as previously described for other CDAs such as AnCDA and ArCE4A^{34,33}. While AngCDA prefers to deacetylate the third unit from the non-reducing end in both substrates first, this preference seems to be more pronounced for A5, resulting in AADAA in rather pure form. To our knowledge, this makes AngCDA the first CDA to mainly produce this acetylation pattern. On closer inspection of the A4 deacetylation, the comparably high standard deviation for A3D1 suggests, that the exact time point influences which products are present. If the reaction was slightly less advanced, more AADA was found, while later, ADAa seemed to accumulate as further discussed in section 2.5.

In contrast to the deacetylation of fully acetylated COS (forward mode, fig. 1 & fig. 7), the *N*-acetylation of GlcN-tetraose (D4) (reverse mode, fig. 8) looks rather different. While deacetylating,

AngCDA generates products with decreasing DA in succession (see section 2.1), whereas during the *N*-acetylation of D4, none of the intermediate products show a distinct peak (fig. 8 A). As no clear peaks were visible for the intermediates regarding their DA, the PA of these products was determined at early time points during the reaction (fig. 8 B). No clear preference can be observed for the first *N*-acetylation, where any sugar unit except for the reducing end unit was *N*-acetylated to an equal extent. Then, the neighboring unit towards the reducing end was *N*-acetylated, apparently also *N*-acetylating the reducing end unit itself. However, as the final product A3D1 clearly consisted of mainly AAAD, the DDAA detected might represent an artifact from the difficult sequencing of the low amounts of A2D2 at early time points.

All in all, for both de- and *N*-acetylation, AngCDA seems to prefer substrates with a higher DA. In forward mode, first the substrate itself and then the following products are preferably deacetylated, while the enzyme avoids placing a deacetylated unit at subsite -1. In reverse mode, the first products are preferably *N*-acetylated further, presumably placing the already acetylated unit at subsite -1. To further investigate these subsite preferences, the AngCDA activity on the four different, defined A3D1 paCOS was tested.

2.5 Activity on defined A3D1 paCOS

AngCDA was incubated with the four mono- deacetylated paCOS DAAA, ADAA, AADA and AAAD in direct comparison to A4 and the substrate and product concentrations were monitored during the deacetylation reaction (fig. 9). Only the product development on AAAD was similar to that on A4. The other substrates were deacetylated at lower rates, with ADAA showing the slowest acetate release with only 0.5 mM after 72 hours, corresponding to one deacetylation per substrate, while these levels were reached for A4 already within the first 12 hours. As expected from the MD simulations and the activity assay on D4, partially deacetylated substrates are not preferred by AngCDA. However, the position of the GlcN unit seems to play a major role. If the reducing end of the substrate is a deacetylated GlcN unit, a position which cannot easily be deacetylated by AngCDA, the enzyme is still able to deacetylate the remaining three units, i.e. the ones it also deacetylates in A4 (see fig. 7 A). For all the other mono- deacetylated substrates (DAAA, ADAA, AADA), the GlcN unit is in a position at which it would be deacetylated by AngCDA in A4. Since in the observed time period, only up to two units were deacetylated, it could have been assumed that the activity on DAAA should be comparable to the activity on A4, as the first two deacetylation on A4 occur at the internal units only. However, this was not the case. The reduced activity on DAAA compared to A4, suggests a preference for a GlcNAc unit at either or both subsite -1 or/and -2. This would also explain the strongly reduced activity on ADAA, since according to the mode of action (see fig. 7 A), it is quite likely that the GlcN unit in this substrate would be placed at subsite -1. This would also explain, why ADAA accumulated when AngCDA deacetylates A4, as the other mono- deacetylated product, AADA, is a preferred substrate for the second deacetylation step. Indeed, this was also observed here, when AngCDA was more active on AADA compared to ADAA.

3 Discussion

We have performed a more in depth *in vitro* and *in silico* analysis on AngCDA than previously reported for any other CDA. Concerning the pH and temperature optima, activity on chitin and chitosan polymers and activity on COS DPI-6, AngCDA is similar to already described CDAs^{26, 27, 34, 33}. We thus assume that our insight into this fungal CDA allows a deeper understanding of other CDAs of both bacterial and fungal origin.

For the bacterial VcCDA and the fungal PcCDA, critical loops were identified that shape the substrate binding site and, thus, determine the substrate binding mode and, consequently, the PA of the products generated^{22, 28}. According to the subsite capping model, these loops block parts of the binding site, forcing the substrate to bind in a particular position and preventing deacetylation of polymeric substrates. CDAs such as the fungal AngCDA, AnCDA and CICDA as well as the bacterial ArCE4A, which are active on polymeric substrates, have much smaller loops which do not block parts of the binding site, leaving it more open and accessible for polymers^{32, 34, 33}. Nonetheless, their corresponding smaller loops form the majority of the binding site and most likely contribute to the

different modes of action observed for these CDAs. Based on recent results¹², we assume that in addition, and probably even more dominantly in the case of the more open CDAs, the preference for acetylated or deacetylated sugar units at the different subsites of the enzymes also contribute to defining the PA of the products. To our knowledge, clear subsite preferences have so far only been described for ClCDA (reporting a preference for GlcNAc at subsite -2)²¹ and CnCDA4 (reporting a preference for GlcN at subsite -1)². For other CDAs, such as PesCDA and PgtCDA, such preferences can only be presumed based on the available data¹².

3.1 Comparison between best binding modes *in silico* and *in vitro*

To obtain deeper insights into substrate binding and possible preferences for acetylated or deacetylated units at the different subsites of AngCDA, we performed *in vitro* studies of the mode of action on fully acetylated substrates A4 and A5, on the four mono- deacetylated substrates A3D1 and on the fully deacetylated substrate D4, the latter in reverse *N*-acetylation mode, and compared the results with the detailed *in silico* analysis focusing on the same substrates. Based on the *in silico* data, we would have expected six subsites ranging from -3 to +2, but we did not observe a significantly faster first deacetylation of A6 than of A5. And contrary to the binding energy calculated for the three different binding modes of A5, where the lowest energy was calculated for binding mode [-3,+1], the mode of action shows that the middle unit is deacetylated first, i.e. that the preferred binding mode is in fact [-2,+2]. This hints at a minor role for subsite -3, if any. Similarly, the rmsf values, which are lowest for binding mode [-3,0] and [-3,+1] for A4 and A5, respectively, suggests a stable binding in these orientations, but the corresponding products AAAD and AAADA were not or only rarely observed *in vitro*.

It is important to highlight that (i) before the start of the simulations, the substrates were already positioned in their binding modes and (ii) standard MD simulations are not able to simulate the catalytic reaction. Thus, either the entry of the substrate into the binding site or the reaction itself might account for the differences between best binding modes observed *in vitro* and *in silico*. Since the *in silico* data suggest that the sugar unit at subsite 0 is still correctly positioned for deacetylation in binding modes ending at subsite 0, we expect that the substrate entry into the binding site might have a higher energy barrier for binding modes mainly including the minus subsites. As already reported or suggested for other CDAs, the +1 subsite might play an important role for efficient catalysis, as it appears to be necessary for substrate binding. For ArCE4A, the only CDA with an open binding site which has been crystallized with its substrate, only two units of the chitotetraose substrate were resolved, at subsites 0 and +1, indicating a very stable binding at these positions³³. And for all CDAs classified in the CE4 family, at least two sugar units are needed for activity^{2, 4, 26, 27, 33, 34}. The importance of subsite +1 is also visible for AngCDA in the binding free energy calculations, where, besides the sugar unit in subsite 0, the sugar unit at subsite +1 shows the lowest binding energy. The most consistently observed hydrogen bond at subsite +1 involves Asp162, which is also highly conserved in other CDAs (first D in motif 4, see fig. 2). It should be noted here that this hydrogen bond was often present at the beginning, while it tended to break at some point during the simulation and was rarely reestablished. This may indicate that it is primarily important for substrate entry and may play a minor role in keeping the substrate in the binding site, explaining the differences observed between *in silico* and *in vitro* results. Furthermore, the energy contribution of the zinc ion is highly positive in all simulations (see tab. 6), suggesting that the zinc ion prefers to be surrounded by water instead of interacting with the substrate. In conclusion, we thus hypothesize that the substrate first interacts with Asp162 at subsite +1 and possibly with Lys164 at subsite +2, allowing it to then displace the water molecules surrounding the zinc ion.

3.2 Subsite preferences for GlcNAc and GlcN units

In addition to hydrogen bonds, which contribute to substrate binding and possibly substrate entry, other interactions certainly play a role as well. Hydrophobic interactions between Phe139 and Leu193 on the one hand and the acetyl methyl group on the other hand strongly contribute to the substrate binding at subsite 0. Stacking interactions between aromatic residues, such as Tyr138 and Tyr166, further contribute to a strong binding at subsite 0. Only salt bridges were not observed *in silico*, as all substrates were uncharged. Given that the MD simulations do not cover the substrate entry into the

binding cleft, as discussed above, the preferred binding mode is difficult to predict from these simulations. Nonetheless, they provide valuable insights into the subsite preferences, namely a strong preference for a GlcNAc unit at subsite -1 and probably weak preferences for GlcNAc units at the other subsites. But even with these detailed simulations, it remains difficult to develop a hypothesis that would clearly explain these GlcNAc preferences. The main difference between A4 and ADA4 in binding mode [-2,+1] is the rotation of the GlcN unit at subsite -1 resulting in a completely different orientation of the neighboring non-reducing end unit. This is enabled by the missing acetyl group and residues Thr197 and Tyr138, which seem to stabilize this orientation (fig. 4). However, compared to the A4 simulations where the positioning of the GlcNAc unit at subsite -1 is highly reproducible, this reorientation of the GlcN unit is not visible in all ADA4 trajectories. This would suggest that contrary to a GlcN unit, a distinct energy minimum for a GlcNAc unit exists at subsite -1, resulting in a GlcNAc preference. A weak GlcNAc preference at subsite +1, deduced from both the *in silico* and *in vitro* results, could be attributed to Asn167, which occasionally formed a hydrogen bond with the acetyl oxygen but did not interact with a GlcN unit (see the hbonds occupancy spreadsheet in the supporting information). At subsite -2, Leu73 might contribute to a weak preference for acetylated units, but this preference is difficult to confirm with our *in vitro* results. As proposed by Wattjes et al.¹², these subsite preferences are thought to influence the acetylation pattern generated especially on polymeric substrates. They describe *N*-acetylation of fully deacetylated polyglucosamine polymers using different CDAs including the here characterized *Aspergillus niger* CDA. Their results show, that AngCDA (named AnCDA in their paper) generates a more block-wise acetylation pattern, preferably *N*-acetylating neighboring units of already acetylated ones. Our results suggest that this *N*-acetylation occurs towards the reducing end.

4 Conclusion & Outlook

Our results show that the combination of *in silico* and *in vitro* methods can unveil more details regarding the mechanisms underlying the generation of different acetylation patterns by CDAs. On the one hand, without simulating substrate entry into the active site, the *in silico* predictions concerning the most favored binding mode need to be validated experimentally and evaluated carefully. On the other hand, the computational comparison between COS and paCOS seems to open the door for a detailed analysis of subsite preferences in CDAs, which would be very tedious in the lab. However, given the current state of the art, any conclusions from the *in silico* data will require final *in vitro* validation by testing the activity on different defined paCOS selected based on the computational results. As a further proof of concept, the *in silico* approach introduced here should be tested with other CDAs, such as the chitosan deacetylase CnCDA4 with its uniquely strong preference for GlcN at its -1 subsite².

While the acetylation pattern generated on smaller oligomers seems to be defined by the available and accessible subsites, it appears to be strongly influenced by subsite preferences when acting on larger oligomers or on polymeric substrates. While CDAs were reported to retain their regioselectivity on A4 and D4 in forward and reverse mode, respectively²³, this does not appear to be the case for polymers. Here, a GlcN preference next to subsite 0, i.e. at either subsite -1 or +1, leads to an alternating or block-wise acetylation pattern when the substrate is *N*- or deacetylated, respectively (see studies on PgtCDA^{12,17}).

Based on these recent findings and our results presented here, we suggest that any detailed characterization of a CDA needs to include a thorough analysis of its subsite preferences, like in the case of chitinases and chitosanases which can even be classified based on their subsite specificities and preferences³⁷. A CDA like AngCDA which apparently favors GlcNAc at all subsites is expected to produce a random acetylation pattern when deacetylating high DA chitosans and rather large GlcNAc blocks when *N*-acetylating polyglucosamine. A hypothetical CDA, with a different preference at subsite +1 (e.g. a GlcNAc preference at subsite -1 and a GlcN preference at subsite +1) would presumably produce small GlcN or GlcNAc blocks when de- or *N*-acetylating the substrate, since single GlcNAc or GlcN units between these blocks would not be further de- or *N*-acetylated, respectively.

Our combined *in vitro* and *in silico* approach can help to elucidate these subsite preferences. The detailed understanding thus gained of the most influential residues, can then be used for protein engineering to tailor the subsite preferences, improving access to a broader range of chitosan oligomers with fully defined architecture and of chitosan polymers with defined, non- random patterns of acetylation. Such third generation chitosans are a prerequisite to understand the physiological roles of CDA-generated PAs in natural chitosans and they are promising the next breakthrough in the development of reliable chitosan-based applications e.g. in agriculture or biomedicine.

5 Experimental procedures

5.1 *In silico*

5.1.1 Sequence analysis

The AngCDA sequence (UniProt ID: A2QZC8) was analyzed using several online tools including the conserved domain databank^{38,39} for domain prediction, SignalP- 5.0⁴⁰ for signal peptide prediction, PredGPI⁴¹ and NetGPI- 1.1⁴² for GPI anchor prediction and TMHMM⁴³ for transmembrane region prediction.

5.1.2 Multiple structure alignment

To our knowledge, the crystal structure of eight CE4 enzymes acting on chitinous substrates are described so far, including AnCDA (2Y8U)³⁴, ArCE4A (5LFZ)³³, CICDA (2IW0)³², SICE4 (2CC0)³⁵, VcCDA (4NY2)²², SpPgdA (2C1G)⁴⁴, BsPdaC (6H8L)²⁷ and BmCDA8 (5Z34)²⁹. All of them, except for BmCDA8, were used for a multiple structure alignment using the PyMOL plugin PyMod3 with SALIGN^{45,46}. If multiple chains were present in the crystal structure, only chain A was used for the alignment. For CICDA, the His- tag was removed, for VcCDA the chitin binding domains (residues 336- 433) and for SpPgdA the two additional domains (residues 46- 266) were deleted from the structure. The 3D structures were colored by conservation with the CAMPO score⁴⁷ using the Blosum62 scoring matrix⁴⁸.

5.1.3 Ligand generation and docking

The 3D structures of chitotetraose (A4) and chitopentaose (A5) were created using the carbohydrate builder from GLYCAM Web⁴⁹. To create ligands with deacetylated units, the acetyl group was removed using the builder function in PyMOL⁴⁵ and the residues and atom names were adjusted to fit the GLYCAM names⁵⁰.

For docking, the ligands (A4, A5, DAAA, ADAA, AADA and AAAD) and the receptor (AngCDA crystal structure) were prepared using the `prepare_ligand4.py` and `prepare_receptor4.py` scripts included in the AutoDockTools⁵¹. The histidine protonation was set to HID for the metal coordinating His97 and His101. The charge of the zinc ion was set to +2. Docking was performed using AutoDock VinaCarb v1.0 with the default parameters for `chi_coeff` and `chi_cutoff` suggested by the authors^{52,53}.

5.1.4 Molecular dynamics simulations

All simulations were run with the GROMACS 2019 package^{54,55,56,57,58,59,60,61} using the Amber force field ff14SB⁶² for the protein, the compatible GLYCAM06 force field^{50,63} for the ligands and TIP3P as the water model⁶⁴. The histidine protonation was set to HID for the metal coordinating His97 and His101 and to HIP for the catalytic His195. Since the GLYCAM06 force field is not available in the GROMACS package, the ligands were first prepared using the LEaP program included in AmberTools20⁶⁵ and then converted into GROMACS format using the GLYCAM06 compatible ACPYPE^{66,67}. Water molecules were added in a dodecahedron- shaped box with 10 Å distance to each side. The charge was neutralized with sodium ions before running the energy minimization using steepest descent. Before the production run, NVT and NPT equilibrations were conducted using the leap- frog integrator in 2 fs steps for 100 ps with a V- rescale thermostat at 310 K and a Berendsen barostat at 1 bar. The same integrator and step size were used for the final NPT production run, using a V- rescale thermostat at 310 K as well, but the more precise Parrinello-

Rahman barostat at 1 bar. A Verlet cutoff- scheme was used for van der Waals interactions and the Ewald summation was used for long- range electrostatic interactions. Distance restraints between the metal ion and the metal binding triad (D48, H97, H101), the catalytic aspartate (D47) and the catalytic water molecule and a dihedral restraint of the catalytic aspartate (D47) were applied to stabilize the metal binding during the simulation.

The resulting trajectories were analyzed using the GROMACS implemented rmsf tool to detect the fluctuation of the ligand in the active site after the complex was centered in the simulation box and rotational and translational movements of the complex were removed. These trajectories were inspected with vmd⁶⁸ and the vmd hbond plugin 1.2 was used with a distance cutoff of 3.5 nm and an angle cutoff of 35° to detect all hydrogen bonds between the enzyme and the ligand during the simulations. In addition, the molecular mechanics generalized Born surface area (mmgbsa) approach was applied to calculate the average binding energy of the ligand throughout the whole trajectory using gmx_MMPBSA⁶⁹, a GROMACS implementation of MMPBSA.py⁷⁰ from AmberTools20⁶⁵. The GB implicit solvent model was used in a single trajectory approximation using standard parameters (igb = 5, saltcon = 0.150) as described in the online documentation. All docked A4 substrates were chosen to find any residue within 6 Å around the substrates and these residues were used for a decomposition analysis of all trajectories.

Representative structures for a visual comparison (fig. 4) were generated using the Jarvis-Patrick clustering algorithm implemented in GROMACS. First, one snapshot was taken from the main cluster of each trajectory, which is most similar to all other snapshots from this cluster. Then, these snapshots from the nine replicates were aligned and again the one most similar to all others was used as a representative structure for the corresponding substrate and binding mode.

5.2 *In vitro*

5.2.1 Cloning

The AngCDA gene (UniProt ID: A2QZC8), without the sequence encoding the 19 amino acid signal peptide, was codon- optimized for *Pichia pastoris* and synthesized by GeneArt (Regensburg, Germany). It was amplified with corresponding overlaps to be cloned via Gibson assembly⁷¹ into a previously generated pET- 22b(+) plasmid (Merck KGaA, Darmstadt, Germany), already containing either an N- terminal pelB and a C- terminal Strep- tag II sequence or both an N- and C- terminal Strep- tag II. Both constructs were transformed into *E. coli* Rosetta™ 2 (DE3) cells (Merck KGaA) for protein expression.

5.2.2 Protein expression and purification

For both constructs, *E. coli* Rosetta™ 2 (DE3) cells carrying the desired plasmid were grown in 500 ml LB auto- induction medium as described by Studier⁷² at 26 °C (for activity tests) or 30 °C (for protein crystallization) for 48 h. Then, cells were harvested at 4000 g for 20 min at 4 °C, resuspended in 20- 30 ml fast protein liquid chromatography (FPLC) washing buffer (20 mM TEA, 400 mM NaCl, pH 8) and stored at - 20 °C. For protein crystallography, lysozyme (1.5 g/l final concentration) and NaCl (10 g/l final concentration) were added and the cells were thawed at room temperature (RT) before they were further lysed by five 15 s pulses at 40% amplitude using a Branson Digital Sonifier model 250- D (Emerson, St Louis, Missouri, USA). For all activity assays, the cells were thawed at room temperature before 3 µl benzonase (Merck KGaA, 25 U/µl) in 250 µl 2 M MgCl₂ were added and incubated at RT for 15 min, shaking slightly. Then, 2 ml high salt buffer (1 M TEA, 1 M NaCl, pH 8) were added before the cells were lysed by sonication as described above. For both objectives, the lysed cells were centrifuged for 60 min at 40,000 g at 4 °C and the AngCDA was purified from the supernatant by affinity chromatography using the Strep- Tactin®XT purification system (IBA, Göttingen, Germany). Finally, the enzymes were concentrated with Amicon Ultra- 15 centrifugal filters (Merck KGaA) and rebuffered into Tris- HCl pH 8 and 100 mM NaCl for protein crystallization or into 50 mM TEA pH 7 for activity assays.

5.2.3 Protein crystallization

For further purification before crystallization, a size- exclusion chromatography was performed at 20 ° C on a Superdex 75 16/600 column (GE Healthcare, Boston, Massachusetts, USA) equilibrated with 50 mM Tris -HCl pH 8 and 100 mM NaCl. Peak fractions corresponding to the AngCDA monomer were pooled and concentrated with an Amicon Ultra- 15 centrifugal filter to 4.57 mg/ml before the crystallization screening. Crystals of AngCDA were found in sitting- drop vapor- diffusion experiments at 20 ° C. Drops were prepared by mixing 0.2 μ l of AngCDA and 0.1 μ l of reservoir containing 1.4 M sodium malonate and 0.1 M Bis- Tris propane pH 7. Crystals were cryo- protected by soaking in 10 % (v/v) glycerol in crystallization buffer, before flash- freezing in liquid nitrogen. Diffraction data were collected at 100 K at microfocus beamline Proxima- 2A (Soleil, France). The collected images were processed using XDS⁷³ and scaled in the CCP4 suite program AIMLESS⁷⁴. MOLREP⁷⁵ was used for molecular replacement using *Aspergillus nidulans* chitin deacetylase (AnCDA, 2Y8U)³⁴. The structure was rebuilt with Coot⁷⁶ and refined with PHENIX⁷⁷. The final model was evaluated with MolProbity⁷⁸. All values obtained and generated are shown in table 1.

5.2.4 pH and temperature optimum

The enzyme activity was tested from pH 2 to 12 in increments of 1 and at pH 7.5 and 8.5 using 50 % (v/v) Teorell Stenhagen buffer^{79,80} with 18.975 μ g/ml purified AngCDA and 0.5 mM A4. In addition, the activity was also tested for 50 mM TEA buffer at pH 7. After 30 min incubation at 37 ° C, the reaction was stopped with one volume 0.1 M HCl and the products were measured via HILIC-ESI- MS as described in section 5.2.6.

5.2.5 Activity assay on polymeric substrates

Chitosan with DA 0 % as well as α - or β - chitin were kindly provided by Dominique Gillet, Gillet Chitosan (Plumaudan, France). The DA 0 % chitosan was *N*- acetylated to DA 12 %, 32 % and 46 % according to Lamarque et al.¹⁵. The DA was analyzed using 400 MHz 1H NMR⁸¹ and the DP and dispersity (DP 1600 and \bar{D} 1.88 for the starting material) were analyzed using SEC- RI- MALLS^{17,82}. Colloidal chitin was prepared as described by Hsu et al.⁸³.

To determine the activity on these substrates, 18.975 μ g/ml purified AngCDA was incubated with 1 mg/ml substrate in 50 mM TEA buffer (pH 7) at 30 ° C for 24 h. Samples were taken after 2, 4 and 24 h and the reaction was stopped with one volume of 66.6 mM HCl. Samples containing insoluble chitin substrates were shaken during incubation to avoid precipitation of the substrates. The released acetate was quantified with the acetic acid kit from R- Biopharm (Darmstadt, Germany) using reduced volumes (100 μ l sample, 63 μ l water, 100 μ l solution 1, 20 μ l solution 2, 20 μ l 1:20 diluted solution 3 and 20 μ l 1:10 diluted suspension 4) to fit in a microtiter plate. The Δ Acetic acid was calculated as described in the kit. A standard curve was generated with 0.05, 0.03, 0.015 and 0.003 g/l acetic acid to calculate the acetate concentration in each sample, which directly correlates with the Δ DA.

5.2.6 Activity assays on oligomeric substrates

Oligomeric substrates chitobiose (A2), chitotriose (A3) and chitohexaose (A6) were purchased from Megazyme (Wicklow, Ireland). Chitosantetraose (D4) was purchased from Biosynth (Thal, Switzerland). A mixture of freeze- dried chitotetraose (A4) and chitopentaose (A5) was kindly provided by the Bio Base Europe Pilot Plant (Gent, Belgium), produced as described in Hamer et al.²⁴. The mixture was dissolved in 1:1 H₂O:ACN (15 mg/ml) and purified via hydrophilic liquid interaction chromatography (HILIC) using the modular LC- 20A Prominenc HPLC system (Shimadzu Deutschland GmbH, Duisburg, Germany) with a BEH Amide column (5 μ m, 10 x 250 mm, Waters Corporation, Milford, Massachusetts, USA). A gradient shifting from solvent A (80 % (v/v) ACN, 20 % (v/v) H₂O) to solvent B (80 % (v/v) H₂O, 20 % (v/v) ACN), both with 10 mM NH₄HCO₂ and 0.1 % (v/v) formic acid, was used to separate the two oligomers. Collected fractions were freeze- dried and dissolved in a final concentration of 10 mM. A3D1 substrates were generated by incubating A4 with NodB (DAAA)²⁴, VcCDA (ADAA)²² or PesCDA (AADA)⁴ in 50 mM TEA buffer (pH 7) or

by incubating D4 with AngCDA (AAAD) in 2.25 M NaAc buffer (pH 7) at 37 °C. All samples were freeze-dried and the paCOS generated were separated as described for A4 and A5.

For all tests, 0.5 mM substrate was incubated at 37 °C, pH 7 either using 50 mM TEA buffer for deacetylation or 2 M NH₄Ac for *N*-acetylation. The deacetylation and *N*-acetylation reactions were stopped with one volume of 1 % formic acid or by freezing the samples, respectively. The enzyme concentration was adjusted to an activity resulting in 10- 20 % A3D1 after 30 min incubation with A4 for deacetylation reactions and in 50- 60 % A3D1 after 72 h incubation with D4 for *N*-acetylation reactions.

All paCOS generated from these activity assays were chemically *N*-acetylated with acetic anhydride-d₆ (Merck KGaA) to allow quantification by MS analysis as described by Cord- Landwehr et al.⁸⁴. To this end, the samples were freeze-dried and resolved in 10 µl H₂O before 10 µl 100 mM NaHCO₃ were added, followed by 20 µl methanol with 1 µl acetic anhydride- d₆. After 30 min incubation at 30 °C and 1200 rpm, 10 µl methanol with 1 µl acetic anhydride- d₆ were added before a second 30 min incubation under the same conditions. Finally, the samples were freeze-dried again and resolved in their initial volume for quantitative MS analysis according to Hamer et al.²⁴.

5.2.7 Sequencing of A4, A5 and D4 products

To determine the PA of paCOS generated by A4 and A5 deacetylation, samples corresponding to the time points with the highest concentration of the respective paCOS were chosen (e.g. 6 h for A3D1, see fig. 1). As products generated from D4 *N*-acetylation did not peak one after the other, early time points were chosen for A1D3 and A2D2 pattern analyses to avoid the influence of isotope peaks from the following products. The pattern was determined as described by Cord- Landwehr et al.⁸⁴. To this end, the samples were chemically *N*-acetylated as described above, before reducing end labeling was performed in two steps with 3 and 10 µl H₂¹⁸O at 70 °C for 3 and 18 h, respectively. Finally, the respective paCOS for the chosen time points were analyzed via HILIC- ESI- MS² and fragment intensities were used to determine the pattern of acetylation.

Data availability

Data which is not included in the main article can be found in the supporting information pdf document and the two spreadsheets. The crystal structure of AngCDA can be found in the protein databank with the ID 7BLY.

Acknowledgment

A special thanks goes to Mario Sergio Valdes who helped a lot with the binding energy calculation using his GROMACS implementation (gmx_MMPBSA) of MMPBSA.py from the AmberTools. We thank David Thieker for providing missing force field parameters for the substrates used for the simulations via the glycam mailing list. The authors appreciated the access to the CristalO platform (FR2424, Station Biologique de Roscoff), which is part of the Biogenouest core facility network. The authors thank Dr. Elizabeth Ficko-Blean and the staff for beam time and help during data collection on beamline Proxima- 2A (Soleil, France). GM is grateful for the support of the Alexander von Humboldt foundation during his stay at the University of Münster. Furthermore, we would like to thank Rita Weyer and Margareta Hellmann for their valuable feedback on this manuscript.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations

Abbreviations used are: GlcNAc/A, *N*- acetylglucosamine; GlcN/D, glucosamine; DA, degree of acetylation; DP, degree of polymerization; PA, pattern of acetylation; CDA, chitin deacetylase; COS, chitoooligosaccharide; paCOS, partially acetylated COS; CE4, carbohydrate esterase 4; MD, molecular

dynamics; mmgbsa, molecular mechanics generalized Born surface area; rmsf, root mean square fluctuation; A4, chitotetraose; A5, chitopentaose, D4, GlcN- tetraose or chitosantetraose.

References

- [1] Baker, L. G., Specht, C. A., and Lodge, J. K. (2011). Cell Wall Chitosan Is Necessary for Virulence in the Opportunistic Pathogen *Cryptococcus neoformans*. *Eukaryot. Cell*, 10(9):1264 – 1268.
- [2] Hembach, L., Bonin, M., Gorzelanny, C., and Moerschbacher, B. M. (2020). Unique subsite specificity and potential natural function of a chitosan deacetylase from the human pathogen *Cryptococcus neoformans*. *Proc. Natl. Acad. Sci.*, 117(7):3551–3559.
- [3] Rizzi, Y. S., Happel, P., Lenz, S., Urs, M. J., Bonin, M., Cord-Landwehr, S., Singh, R., Moerschbacher, B. M., and Kahmann, R. (2021). Chitosan and chitin deacetylase activity are necessary for development and virulence of *Ustilago maydis*. *MBio*, 12(2):1–18.
- [4] Cord-Landwehr, S., Melcher, R. L. J., Kolkenbrock, S., and Moerschbacher, B. M. (2016). A chitin deacetylase from the endophytic fungus *Pestalotiopsis* sp. efficiently inactivates the elicitor activity of chitin oligomers in rice cells. *Sci. Rep.*, 6(1):38018.
- [5] Mishra, C., Semino, C. E., McCreath, K. J., De La Vega, H., Jones, B. J., Specht, C. A., and Robbins, P. W. (1997). Cloning and expression of two chitin deacetylase genes of *Saccharomyces cerevisiae*. *Yeast*, 13(4):327–336.
- [6] Malerba, M. and Cerana, R. (2018). Recent Advances of Chitosan Applications in Plants. *Polymers*, 10(2):118.
- [7] Younes, I., Sellimi, S., Rinaudo, M., Jellouli, K., and Nasri, M. (2014). Influence of acetylation degree and molecular weight of homogeneous chitosans on antibacterial and antifungal activities. *Int. J. Food Microbiol.*, 185:57–63.
- [8] Bernkop-Schnürch, A. and Dünnhaupt, S. (2012). Chitosan-based drug delivery systems. *Eur. J. Pharm. Biopharm.*, 81(3):463–469.
- [9] Kauss, H., Jeblick, W., and Domard, A. (1989). The degrees of polymerization and N-acetylation of chitosan determine its ability to elicit callose formation in suspension cells and protoplasts of *Catharanthus roseus*. *Planta*, 178(3):385–392.
- [10] Vander, P., Våain, K. M., Domard, A., El Gueddari, N. E., and Moerschbacher, B. M. (1998). Comparison of the ability of partially N-acetylated chitosans and chitooligosaccharides to elicit resistance reactions in wheat leaves. *Plant Physiol.*, 118(4):1353–1359.
- [11] Wattjes, J., Sreekumar, S., Richter, C., Cord-Landwehr, S., Singh, R., El Gueddari, N. E., and Moerschbacher, B. M. (2020). Patterns matter part 1: Chitosan polymers with non-random patterns of acetylation. *React. Funct. Polym.*, 151(March):104583.
- [12] Wattjes, J., Sreekumar, S., Niehues, A., Mengoni, T., and Mendes, A. C. (2021). Biotechnology-derived chitosans with non-random patterns of acetylation differ from conventional chitosans in their properties and activities. *ChemRxiv. Prepr.*
- [13] Gubaeva, E., Gubaev, A., Melcher, R. L. J., Cord-Landwehr, S., Singh, R., El Gueddari, N. E., and Moerschbacher, B. M. (2018). 'Slipped Sandwich' Model for Chitin and Chitosan Perception in *Arabidopsis*. *Mol. Plant-Microbe Interact.*, 31(11):1145–1153.
- [14] Synowiecki, J. and Al-Khateeb, N. A. (2003). Production, Properties, and Some New Applications of Chitin and Its Derivatives. *Crit. Rev. Food Sci. Nutr.*, 43(2):145–171.
- [15] Lamarque, G., Lucas, J.-M., Viton, C., and Domard, A. (2005). Physicochemical Behavior of Homogeneous Series of Acetylated Chitosans in Aqueous Solution: Role of Various Structural Parameters. *Biomacromolecules*, 6(1):131–142.

- [16] Weinhold, M. X., Sauvageau, J. C., Kumirska, J., and Thöming, J. (2009). Studies on acetylation patterns of different chitosan preparations. *Carbohydr. Polym.*, 78(4):678–684.
- [17] Wattjes, J., Niehues, A., Cord-Landwehr, S., Hoßbach, J., David, L., Delair, T., and Moerschbacher, B. M. (2019). Enzymatic Production and Enzymatic-Mass Spectrometric Fingerprinting Analysis of Chitosan Polymers with Different Nonrandom Patterns of Acetylation. *J. Am. Chem. Soc.*, 141(7):3137–3145.
- [18] Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.*, 42(D1):D490–D495.
- [19] Tokuyasu, K., Ono, H., Mitsutomi, M., Hayashi, K., and Mori, Y. (2000). Synthesis of a chitosan tetramer derivative, β -d-GlcNAc-(1-4)- β -d-GlcNAc-(1-4)- β -d-GlcNAc-(1-4)-d-GlcN through a partial N-acetylation reaction by chitin deacetylase. *Carbohydr. Res.*, 325(3):211–215.
- [20] Baker, L. G., Specht, C. A., Donlin, M. J., and Lodge, J. K. (2007). Chitosan, the deacetylated form of chitin, is necessary for cell wall integrity in *Cryptococcus neoformans*. *Eukaryot. Cell*, 6(5):855–867.
- [21] Tokuyasu, K., Mitsutomi, M., Yamaguchi, I., Hayashi, K., and Mori, Y. (2000). Recognition of Chitoooligosaccharides and Their N-Acetyl Groups by Putative Subsites of Chitin Deacetylase from a Deuteromycete, *Colletotrichum lindemuthianum*. *Biochemistry*, 39(30):8837–8843.
- [22] Andrés, E., Albesa-Jové, D., Biarnés, X., Moerschbacher, B. M., Guerin, M. E., and Planas, A. (2014). Structural Basis of Chitin Oligosaccharide Deacetylation. *Angew. Chemie Int. Ed.*, 53(27):6882–6887.
- [23] Hembach, L., Cord-Landwehr, S., and Moerschbacher, B. M. (2017). Enzymatic production of all fourteen partially acetylated chitosan tetramers using different chitin deacetylases acting in forward or reverse mode. *Sci. Rep.*, 7(1):17692.
- [24] Hamer, S. N., Cord-Landwehr, S., Biarnés, X., Planas, A., Waegeman, H., Moerschbacher, B. M., and Kolkenbrock, S. (2015). Enzymatic production of defined chitosan oligomers with a specific pattern of acetylation using a combination of chitin oligosaccharide deacetylases. *Sci. Rep.*, 5(1):8716.
- [25] Hoßbach, J., Bußwinkel, F., Kranz, A., Wattjes, J., Cord-Landwehr, S., and Moerschbacher, B. M. (2018). A chitin deacetylase of *Podospora anserina* has two functional chitin binding domains and a unique mode of action. *Carbohydr. Polym.*, 183:1–10.
- [26] Naqvi, S., Cord-Landwehr, S., Singh, R., Bernard, F., Kolkenbrock, S., El Gueddari, N. E., and Moerschbacher, B. M. (2016). A Recombinant Fungal Chitin Deacetylase Produces Fully Defined Chitosan Oligomers with Novel Patterns of Acetylation. *Appl. Environ. Microbiol.*, 82(22):6645–6655.
- [27] Grifoll-Romero, L., Sainz-Polo, M. A., Albesa-Jové, D., Guerin, M. E., Biarnés, X., and Planas, A. (2019). Structure-function relationships underlying the dual N-acetylmuramic and N-acetylglucosamine specificities of the bacterial peptidoglycan deacetylase PdaC. *J. Biol. Chem.*, 294(50):19066–19080.
- [28] Aranda-Martinez, A., Grifoll-Romero, L., Aragunde, H., Sancho-Vaello, E., Biarnés, X., Lopez-Llorca, L. V., and Planas, A. (2018). Expression and specificity of a chitin deacetylase from the nematophagous fungus *Pochonia chlamydosporia* potentially involved in pathogenicity. *Sci. Rep.*, 8(1):2170.
- [29] Liu, L., Zhou, Y., Qu, M., Qiu, Y., Guo, X., Zhang, Y., Liu, T., Yang, J., and Yang, Q. (2019). Structural and biochemical insights into the catalytic mechanisms of two insect chitin deacetylases of the carbohydrate esterase 4 family. *J. Biol. Chem.*, 294(15):5774–5783.
- [30] Sarkar, S., Gupta, S., Chakraborty, W., Senapati, S., and Gachhui, R. (2017). Homology modeling, molecular docking and molecular dynamics studies of the catalytic domain of chitin deacetylase from *Cryptococcus laurentii* strain RY1. *Int. J. Biol. Macromol.*, 104:1682–1691.

- [31] Jørgensen, T. R., Burggraaf, A.-M., Arentshorst, M., Schutze, T., Lamers, G., Niu, J., Kwon, M. J., Park, J., Frisvad, J. C., Nielsen, K. F., Meyer, V., van den Hondel, C. A., Dyer, P. S., and Ram, A. F. (2020). Identification of Sc1B, a Zn(II)2Cys6 transcription factor involved in sclerotium formation in *Aspergillus niger*. *Fungal Genet. Biol.*, 139(February):103377.
- [32] Blair, D. E., Hekmat, O., Schüttelkopf, A. W., Shrestha, B., Tokuyasu, K., Withers, S. G., and van Aalten, D. M. F. (2006). Structure and Mechanism of Chitin Deacetylase from the Fungal Pathogen *Colletotrichum lindemuthianum*. *Biochemistry*, 45(31):9416–9426.
- [33] Tuveng, T. R., Rothweiler, U., Udatha, G., Vaaje-Kolstad, G., Smalås, A., and Eijsink, V. G. H. (2017). Structure and function of a CE4 deacetylase isolated from a marine environment. *PLoS One*, 12(11):e0187544.
- [34] Liu, Z., Gay, L. M., Tuveng, T. R., Agger, J. W., Westereng, B., Mathiesen, G., Horn, S. J., Vaaje-Kolstad, G., van Aalten, D. M. F., and Eijsink, V. G. H. (2017). Structure and function of a broad-specificity chitin deacetylase from *Aspergillus nidulans* FGSC A4. *Sci. Rep.*, 7(1):1746.
- [35] Taylor, E. J., Gloster, T. M., Turkenburg, J. P., Vincent, F., Brzozowski, A. M., Dupont, C., Shareck, F., Centeno, M. S., Prates, J. A., Puchart, V., Ferreira, L. M., Fontes, C. M., Biely, P., and Davies, G. J. (2006). Structure and Activity of Two Metal Ion-dependent Acetylxylylan Esterases Involved in Plant Cell Wall Degradation Reveals a Close Similarity to Peptidoglycan Deacetylases. *J. Biol. Chem.*, 281(16):10968–10975.
- [36] Delas, T., Mock-Joubert, M., Faivre, J., Hofmaier, M., Sandre, O., Dole, F., Chapel, J., Crépet, A., Trombotto, S., Delair, T., and Schatz, C. (2019). Effects of Chain Length of Chitosan Oligosaccharides on Solution Properties and Complexation with siRNA. *Polymers*, 11(8):1236.
- [37] Weikert, T., Niehues, A., Cord-Landwehr, S., Hellmann, M. J., and Moerschbacher, B. M. (2017). Reassessment of chitosanase substrate specificities and classification. *Nat. Commun.*, 8(1):1698.
- [38] Lu, S., Wang, J., Chitsaz, F., Derbyshire, M. K., Geer, R. C., Gonzales, N. R., Gwadz, M., Hurwitz, D. I., Marchler, G. H., Song, J. S., Thanki, N., Yamashita, R. A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C. J., and Marchler-Bauer, A. (2020). CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res.*, 48(D1):D265–D268.
- [39] Marchler-Bauer, A. and Bryant, S. H. (2004). CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.*, 32(Web Server):W327–W331.
- [40] Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., and Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol.*, 37(4):420–423.
- [41] Pierleoni, A., Martelli, P., and Casadio, R. (2008). PredGPI: a GPI-anchor predictor. *BMC Bioinformatics*, 9(1):392.
- [42] Gíslason, M. H., Nielsen, H., Almagro Armenteros, J. J., and Johansen, A. R. (2021). Prediction of GPI-Anchored proteins with pointer neural networks. *Curr. Res. Biotechnol.*
- [43] Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001). Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes. *J. Mol. Biol.*, 305(3):567–580.
- [44] Blair, D. E. and van Aalten, D. M. (2004). Structures of *Bacillus subtilis* PdaA, a family 4 carbohydrate esterase, and a complex with N-acetyl-glucosamine. *FEBS Lett.*, 570(1-3):13–19.
- [45] Schrödinger, L. (2020). The PyMOL Molecular Graphics System, Version 2.5.
- [46] Janson, G. and Paiardini, A. (2020). PyMod 3: a complete suite for structural bioinformatics in PyMOL. *Bioinformatics*.
- [47] Paiardini, A., Bossa, F., and Pascarella, S. (2005). CAMPO, SCR_FIND and CHC_FIND: a suite of web tools for computational structural biology. *Nucleic Acids Res.*, 33(Web Server):W50–W55.

- [48] Henikoff, S. and Henikoff, J. G. (2000). Amino acid substitution matrices. *Biochemistry*, 89(November):73–97.
- [49] Woods Group (2005). GLYCAM Web. Complex Carbohydrate Research Center, University of Georgia, Athens, GA. (<http://www.glycam.com>).
- [50] Kirschner, K. N., Yongye, A. B., Tschampel, S. M., González-Outeiriño, J., Daniels, C. R., Foley, B. L., and Woods, R. J. (2008). GLYCAM06: A generalizable biomolecular force field. *Carbohydrates. J. Comput. Chem.*, 29(4):622–655.
- [51] Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.*, 30(16):2785–2791.
- [52] Trott, O. and Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, 31(2):455–461.
- [53] Nivedha, A. K., Thieker, D. F., Makeneni, S., Hu, H., and Woods, R. J. (2016). Vina-Carb: Improving Glycosidic Angles during Carbohydrate Docking. *J. Chem. Theory Comput.*, 12(2):892–901.
- [54] Bekker, H., Berendsen, H. J. C., Druenen, R. V., and Spoel, D. V. D. (1993). Gromacs : A parallel computer for molecular dynamics simulations. *Phys. Comput.* 92, 92:252–256.
- [55] Berendsen, H., van der Spoel, D., and van Druenen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.*, 91(1-3):43–56.
- [56] Lindahl, E., Hess, B., and van der Spoel, D. (2001). GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.*, 7(8):306–317.
- [57] Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *J. Comput. Chem.*, 26(16):1701–1718.
- [58] Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.*, 4(3):435–447.
- [59] Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., van der Spoel, D., Hess, B., and Lindahl, E. (2013). GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854.
- [60] Páll, S., Abraham, M. J., Kutzner, C., Hess, B., and Lindahl, E. (2015). Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In *Solving Software Challenges for Exascale*, volume 8759, pages 3–27. Springer International Publishing.
- [61] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25.
- [62] Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015). ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.*, 11(8):3696–3713.
- [63] Singh, A., Tessier, M. B., Pederson, K., Wang, X., Venot, A. P., Boons, G.-J., Prestegard, J. H., and Woods, R. J. (2016). Extension and validation of the GLYCAM force field parameters for modeling glycosaminoglycans. *Can. J. Chem.*, 94(11):927–935.
- [64] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926–935.
- [65] Case, D., Belfon, K., Ben-Shalom, I., Brozell, S., Cerutti, D., Cheatham, T., Cruzeiro, V., Darden, T., Duke, R., Giambasu, G., Gilson, M., Gohlke, H., Goetz, A., Harris, R., Izadi, S., Izmailov,

- S., Kasavajhala, K., Kovalenko, A., Krasny, R., Kurtzman, T., Lee, T., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Man, V., Merz, K., Miao, Y., Mikhailovskii, O., Monard, G., Nguyen, H., Onufriev, A., F.Pan, Pantano, S., Qi, R., Roe, D., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C., Skrynnikov, N., Smith, J., Swails, J., Walker, R., Wang, J., Wilson, L., Wolf, R., Wu, X., Xiong, Y., Xue, Y., York, D., and Kollman, P. (2020). AMBER 2020, University of California, San Francisco.
- [66] Sousa da Silva, A. W. and Vranken, W. F. (2012). ACPYPE - AnteChamber PYthon Parser interface. *BMC Res. Notes*, 5(1):367.
- [67] Bernardi, A., Faller, R., Reith, D., and Kirschner, K. N. (2019). ACPYPE update for nonuniform 1-4 scale factors: Conversion of the GLYCAM06 force field from AMBER to GROMACS. *SoftwareX*, 10:100241.
- [68] Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD – Visual Molecular Dynamics. *J. Mol. Graph.*, 14:33–38.
- [69] Tresanco, M. S. V., Valdes-Tresanco, M. E., Valiente, P. A., and Frias, E. M. (2021). gmx_MMPBSA (Version v1.4.1).
- [70] Miller, B. R., McGee, T. D., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012). MMPBSA.py : An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.*, 8(9):3314–3321.
- [71] Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, 6(5):343–345.
- [72] Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.*, 41(1):207–234.
- [73] Kabsch, W. (2010). XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 66(2):125–132.
- [74] Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011). Overview of the CCP 4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 67(4):235–242.
- [75] Vagin, A. and Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 66(1):22–25.
- [76] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 66(4):486–501.
- [77] Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010). PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 66(2):213–221.
- [78] Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010). MolProbity : all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, 66(1):12–21.
- [79] Teorell, T. and Stenhagen, E. (1938). Ein Universalpuffer für den pH-Bereich 2.0 bis 12.0. *Biochem. Z.*, 299:416–419.
- [80] Östling, S. and Virtama, P. (1946). A Modified Preparation of the Universd Buffer Described by Teorell and Stenlingen . *Acta Physiol. Scand.*, 11(4):289–293.

[81] Hirai, A., Odani, H., and Nakajima, A. (1991). Determination of degree of deacetylation of chitosan by ¹H NMR spectroscopy. *Polym. Bull.*, 26(1):87–94.

[82] Schatz, C., Viton, C., Delair, T., Pichot, C., and Domard, A. (2003). Typical Physicochemical Behaviors of Chitosan in Aqueous Solution. *Biomacromolecules*, 4(3):641–648.

[83] Hsu, S. C. and Lockwood, J. L. (1975). Powdered Chitin Agar as a Selective Medium for Enumeration of Actinomycetes in Water and Soil. *Appl. Microbiol.*, 29(3):422–426.

[84] Cord-Landwehr, S., Ihmor, P., Niehues, A., Luftmann, H., Moerschbacher, B. M., and Mormann, M. (2017). Quantitative Mass-Spectrometric Sequencing of Chitosan Oligomers Revealing Cleavage Sites of Chitosan Hydrolases. *Anal. Chem.*, 89(5):2893–2900.

Figure 1: Product development over time using different COS (0.5 mM) of DP1- 6 (A1 not shown) as substrates ($n = 3$). Graphs A2 to A6 show the relative substrate and all product concentrations for a 72 h reaction period determined using mass spectrometry. The last graph shows the overall activity determined as the acetate release calculated from the MS results, with the insert giving the first 2 h in more detail.

Figure 2: A: Structural alignment of CE4 enzymes which are active on chitinous substrates, ranked by their sequence identity to AngCDA. Conserved motifs (MT), containing the metal binding (circles) and catalytic (asterisks) residues, are highlighted by black boxes. The six loops proposed for VcCDA²² are indicated by a grey background. α helices and β sheets in AngCDA are indicated by red and blue boxes, respectively. B: AngCDA crystal structure with a chitin pentamer docked *in silico* shown as a surface representation, showing the sugar units -2 to +2. This binding mode corresponds to the first deacetylation of the chitin pentamer *in vitro* as shown in 7 B. C: AngCDA shown as a ribbon diagram, highlighting the zinc ion which is coordinated by the metal binding triad (Asp48, His97, His101), a malonate ion and a water molecule in an octahedral coordination geometry. Additionally, the catalytic base (Asp47) and acid (His195) are shown. In both the structural alignment (A) and the 3D representations (B & C), the residues are colored according to the CAMPO score⁴⁷, representing the structural conservation between these enzymes. Further details about the alignment can be found in section 5.1.2.

Figure 3: A: Root mean square fluctuation (rmsf) of the different substrates and their indicated individual sugar units in the given binding modes. B: Binding free energy contribution of the different substrates and their indicated individual sugar units in the given binding modes. A more detailed overview of the binding free energy of all active site residues is available in the MMGBSA residue contribution spreadsheet found in the supporting information. The values for GlcN units are highlighted in red. All values are means \pm SD of nine replicates based on three different starting structures each, except for AADA* with $n = 6$ where the substrate from one starting structure always detached from the enzyme during simulation.

Figure 4: Comparison of representative snapshots for A4 (left) and ADA4 (right) in the binding mode [-2,+1]. The AngCDA is shown as a ribbon diagram, with some key residues being highlighted as sticks and the zinc ion shown as a sphere. Dashed yellow lines show interactions with these key residues. The subsites to which the sugar units are bound are indicated next to each unit.

Figure 5: Average number of hydrogen bonds of the indicated amino acids with the sugar units at the corresponding subsites and the total number of hydrogen bonds between substrate and enzyme throughout the simulation for the different substrates and binding modes. Unless otherwise specified in the corresponding field, the value gives the average number of hydrogen bonds between the amino acid named on top and the sugar unit at the indicated subsite. Values for deacetylated GlcN units are highlighted in red. All values are means \pm SD of nine replicates based on three different starting structures each, except for AADA* with $n = 6$, where the substrate from one starting structure always detached from the enzyme during simulation. A more detailed overview of all hydrogen bonds is available in the hbonds occupancy spreadsheet found in the supporting information.

Figure 6: Average energy contributions of all amino acid residues at the substrate binding site with values below -1 kcal/mol or above +1 kcal/mol, given for all binding modes. The subsite(s) to which

each amino acid residue contributes is/are given in parentheses. The different binding modes are grouped by color, with the subsite 0 sugar unit being indicated by a lower case 'a' or 'd'. Error bars show SD of nine replicates based on three different starting structures each, except for AAdA with $n = 6$, where the substrate from one starting structure always detached from the enzyme during the simulation. A more detailed overview about all active site residues is available in the MMGBSA residue contribution spreadsheet found in the supporting information.

Figure 7: A: Mode of action of AngCDA deacetylating chitotetraose (A4). B: Mode of action of AngCDA deacetylating chitopentaose (A5). Products detected in relative amounts below 5 % and 7 % are not shown for A4 and A5, respectively. Times shown on the left indicate at which time points the samples were taken (see fig. 1). Orange arrows indicate the main path. The binding mode for each step is indicated next to the corresponding arrow ($n = 3$).

Figure 8: A: Product development over time of AngCDA *N*-acetylating GlcN- tetraose (D4). B: Mode of action of AngCDA *N*-acetylating D4. Products detected in relative amounts below 7 % are not shown. Times shown on the left indicate at which time points the samples were taken during the reaction shown in A. The binding mode for each step is indicated next to the corresponding arrow ($n = 3$).

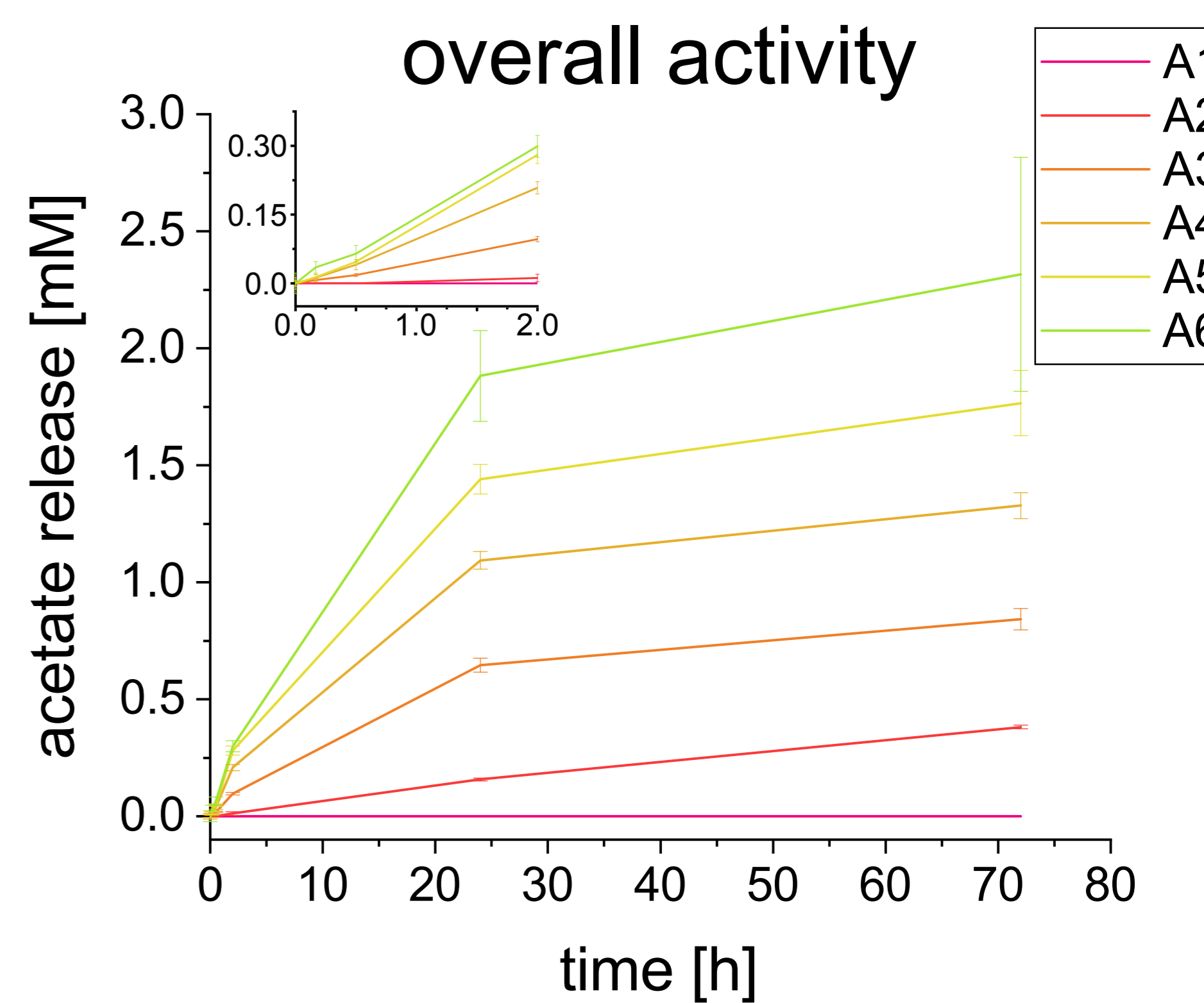
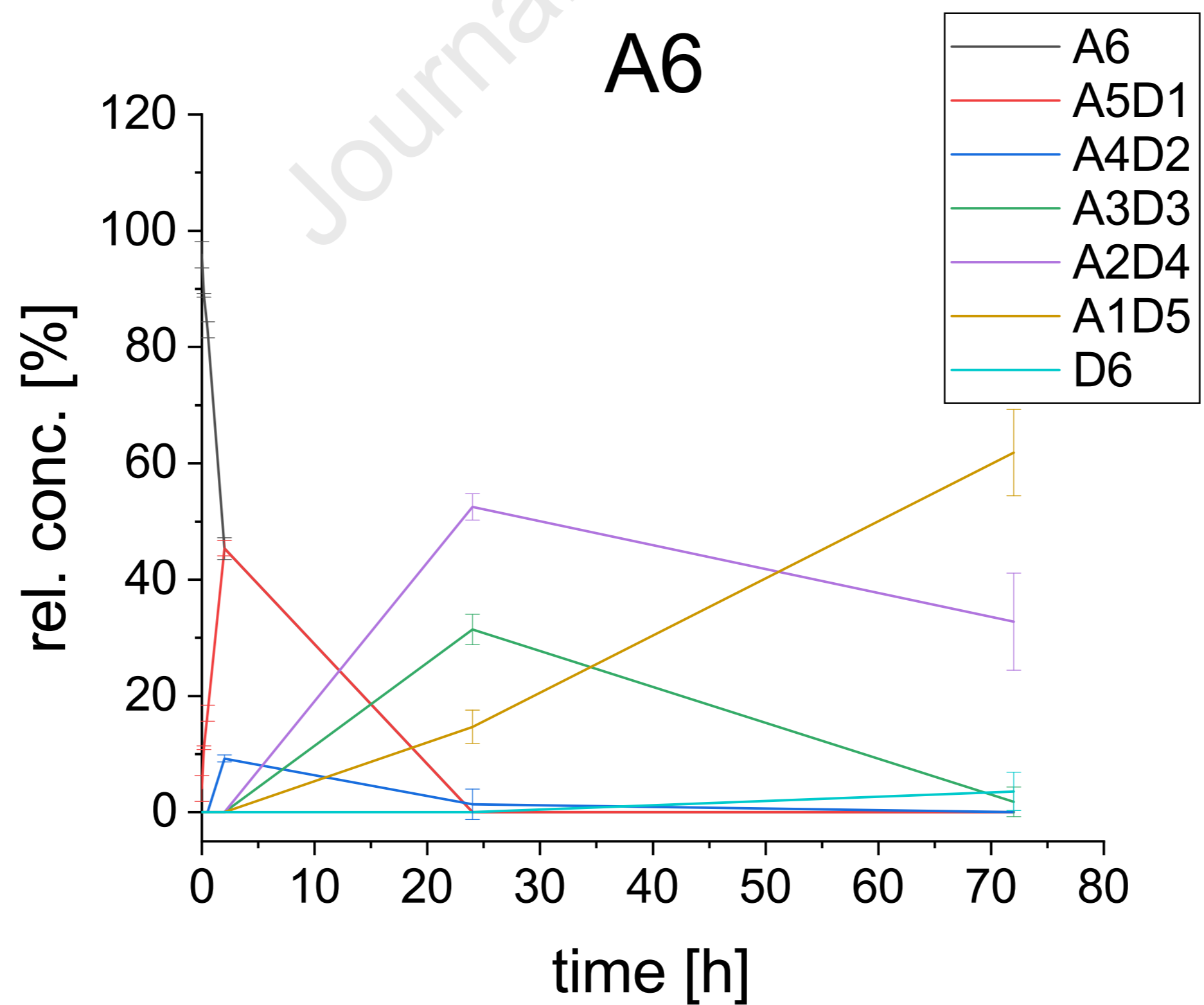
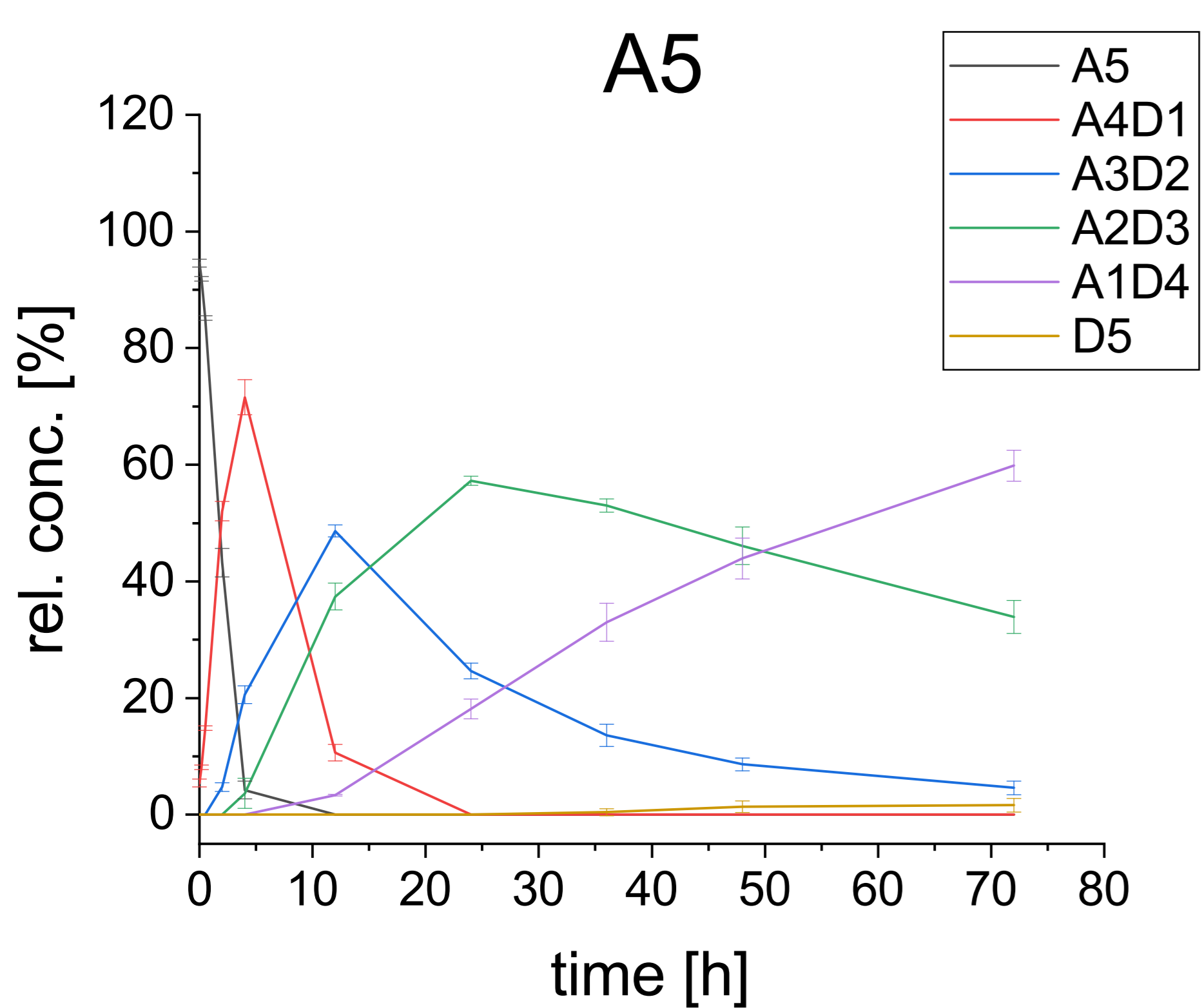
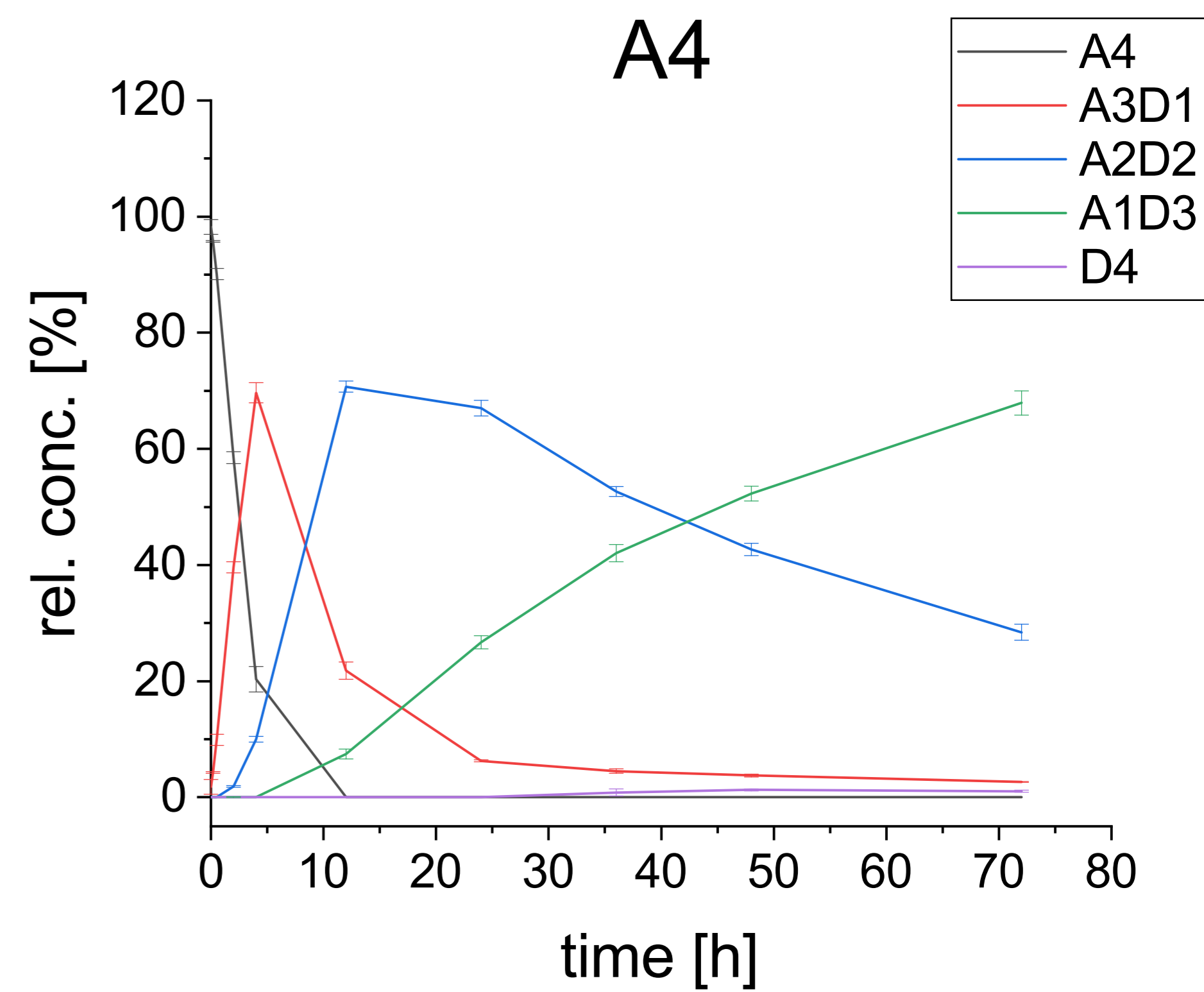
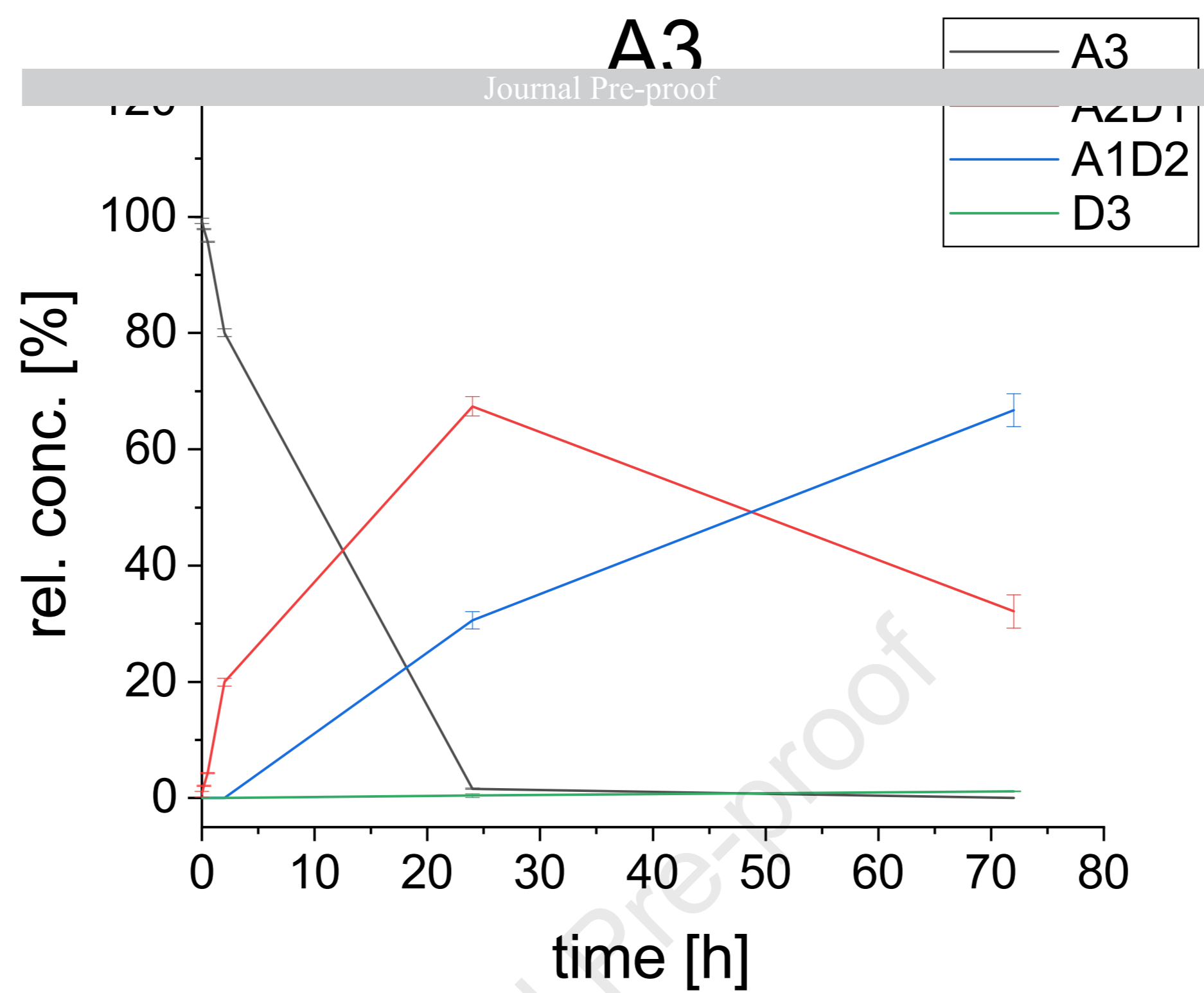
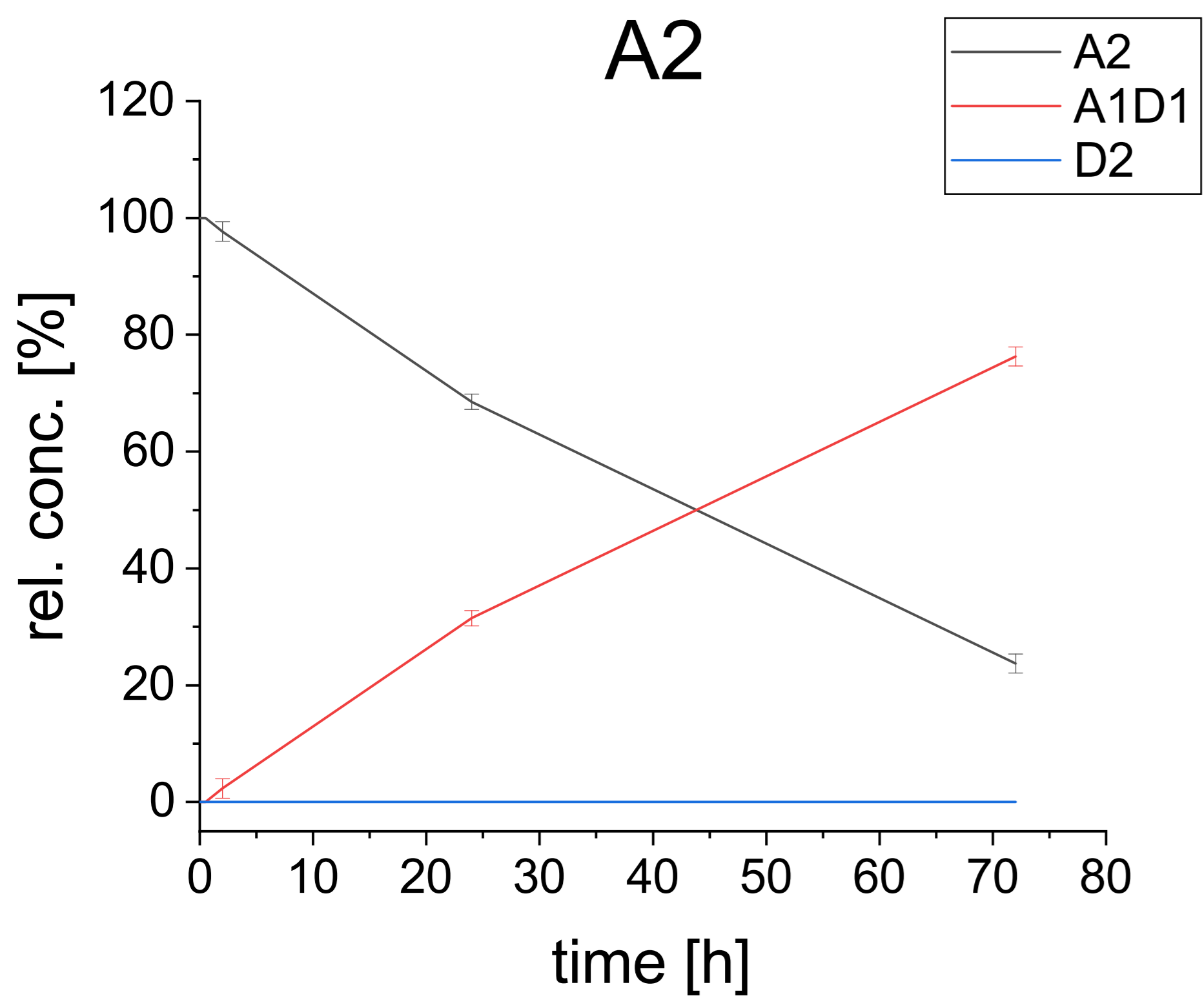
Figure 9: Product development over time using different A3D1 paCOS. Graphs DAAA, ADAA, AADA, AAAD and A4 show the relative substrate (0.5 mM) and all product concentrations for a 72 h reaction period. The final graph gives the overall activity in terms of the calculated total acetate release. Since this is an independent experiment with a different enzyme concentration, the activity on A4 is slightly different from that shown in figure 1 ($n = 3$).

Table 1: Data collection and refinement statistics for the crystal structure of the chitin deacetylase of *Aspergillus niger* CBS 513.88. Values in parentheses refer to the outer resolution shell.

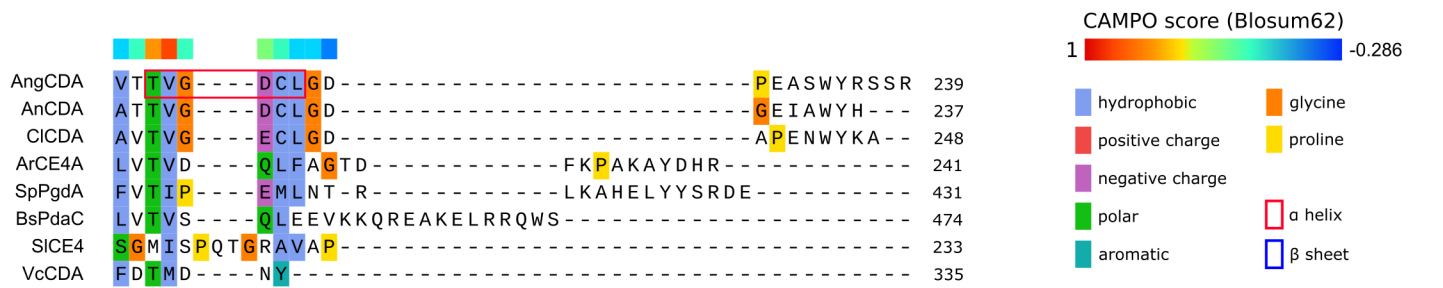
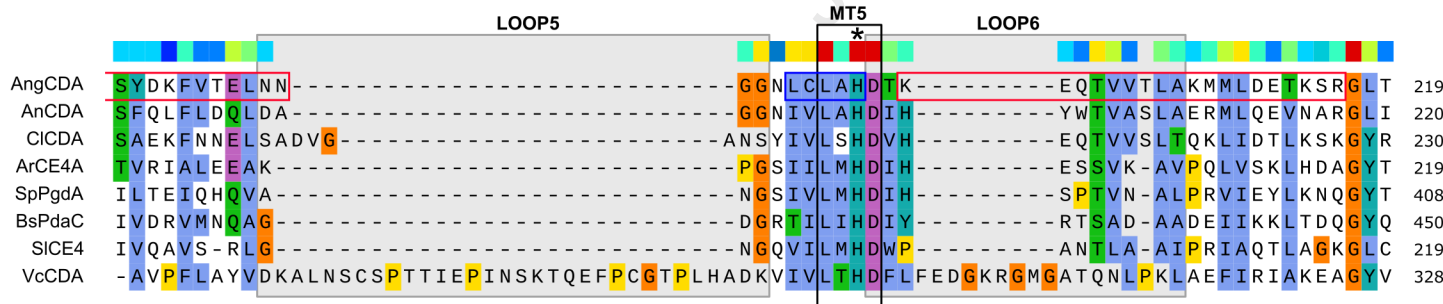
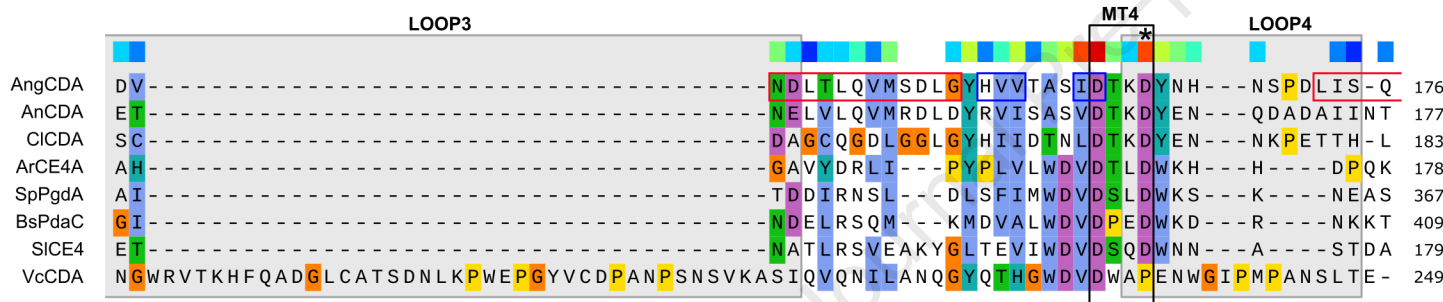
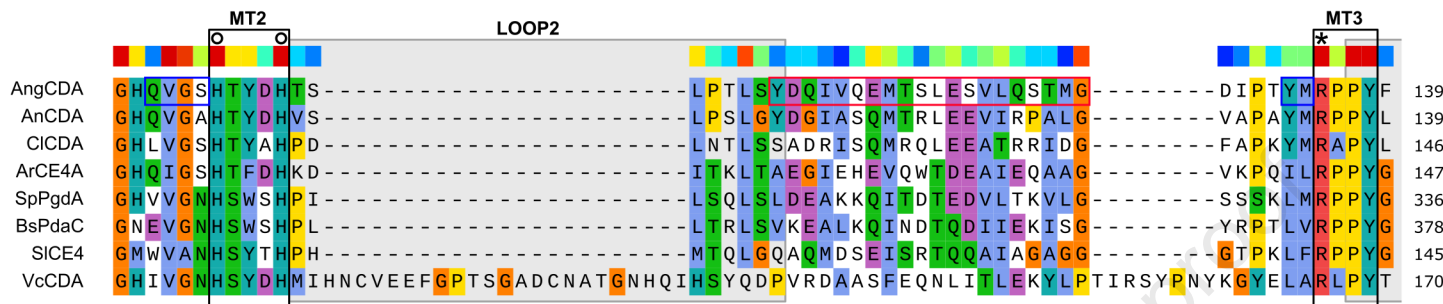
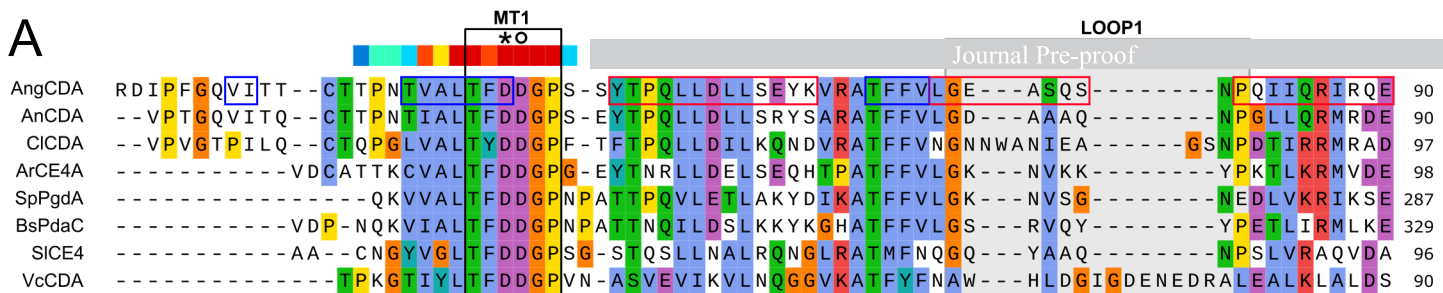
<i>Data collection</i>	AngCDA
Beam line	PROXIMA- 2A
Space group	<i>P</i> 432
Average unit cell (Å)	a = b = c = 119.87
Wavelength (Å)	0.98012
Resolution (Å)	48.94- 1.81 (1.85- 1.81)
R _{pim}	0.043 (0.771)
No. unique reflections	27,485 (1,613)
Mean I/σ I	18.6 (1.9)
CC1/2	0.999 (0.716)
Completeness (%)	100.0 (100.0)
Average redundancy	77.5 (79.6)
<i>Refinement</i>	
Resolution (Å)	42.38- 1.81
R _{free} /R _{work}	18.55/15.26
Total number of atoms	1,965
Water	237
Average B factor	26.56

Ligands	ZN; MLI; CL
<i>R.m.s deviations</i>	
Bonds	0.006
Angles	0.889
<i>MolProbity analysis</i>	
Clashscore, all atoms	0.29 (100 %)
MolProbity score	0.86 (100 %)
PDB entry	7BLY

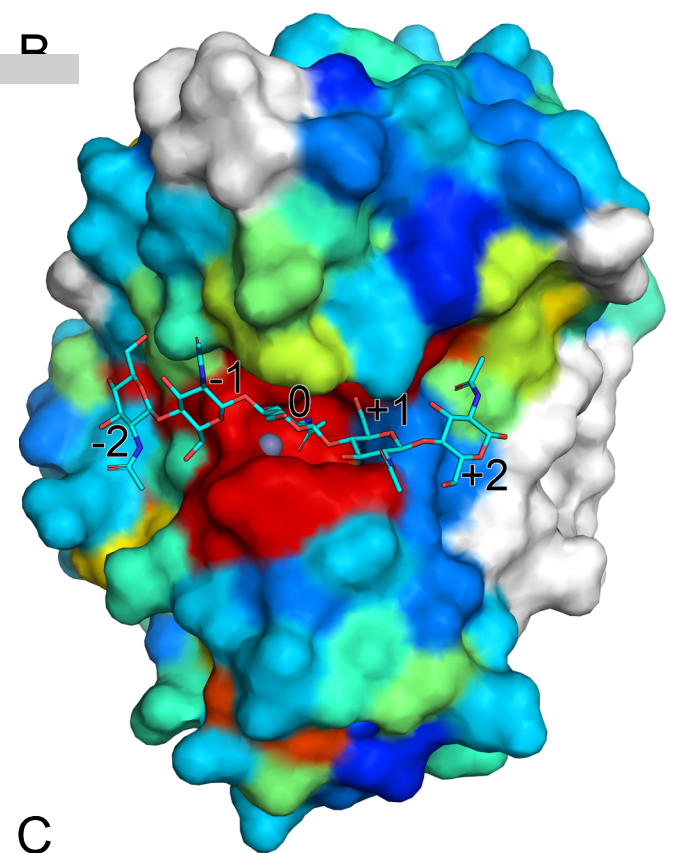
Journal Pre-proof



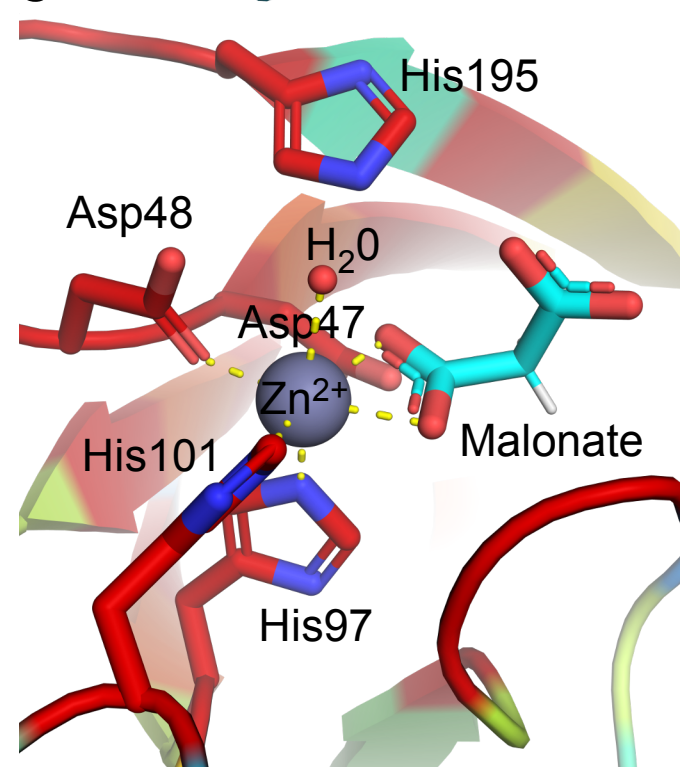
A

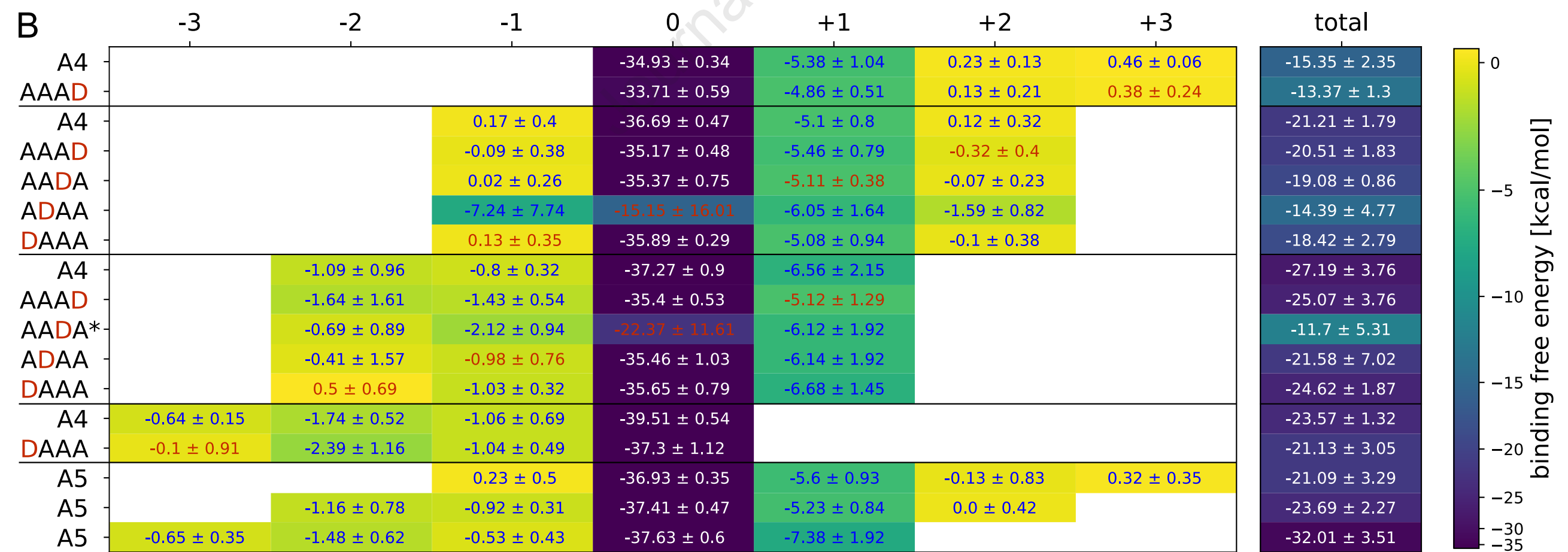
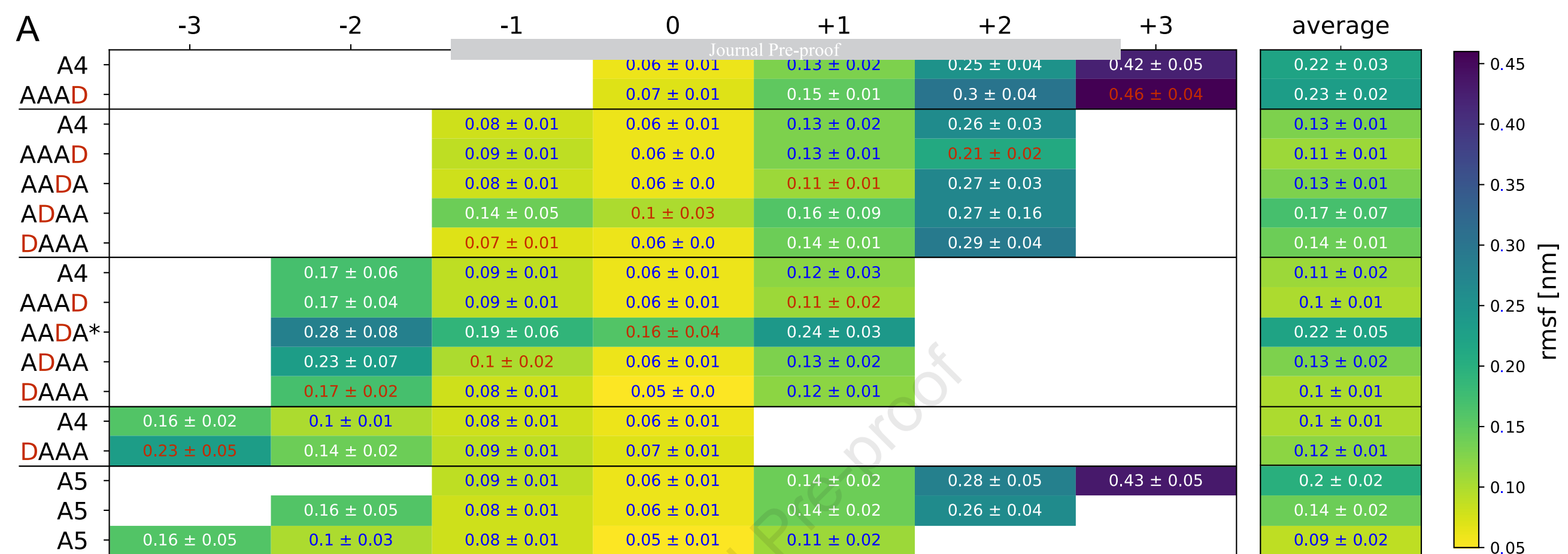


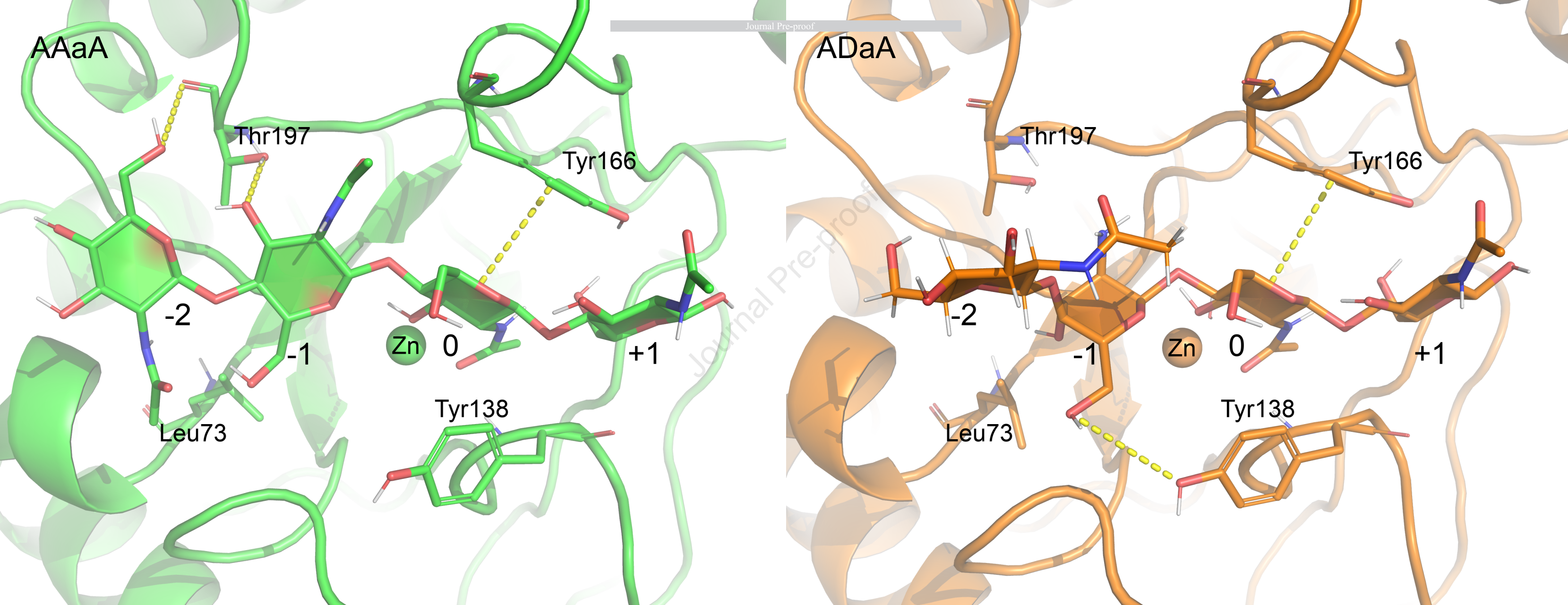
D

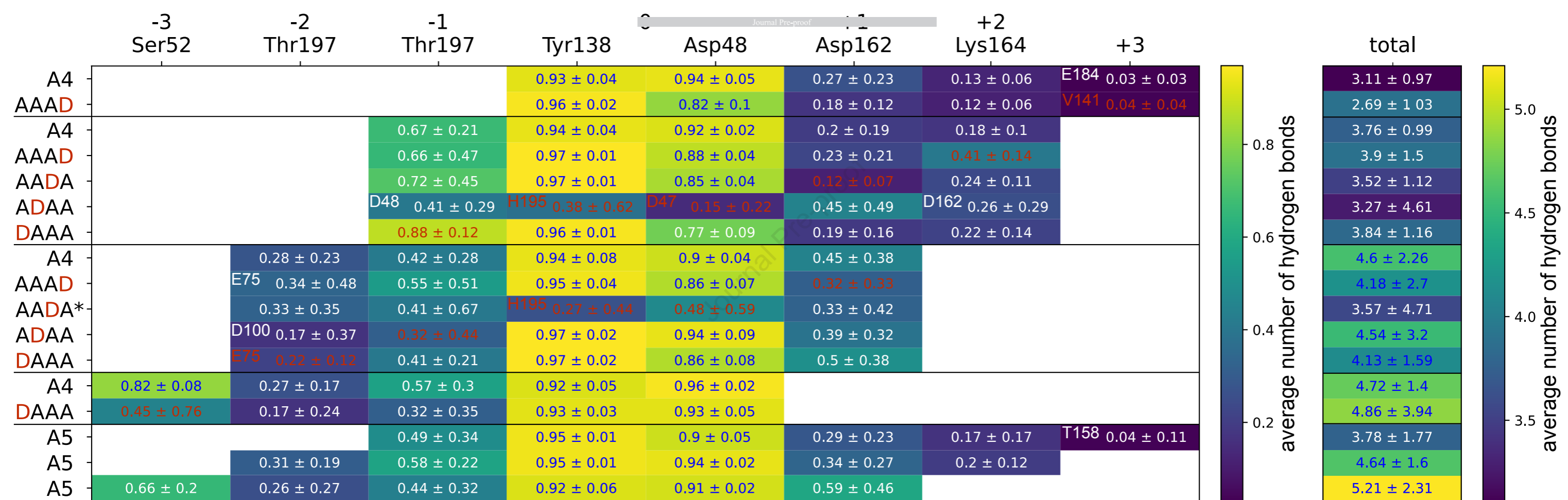


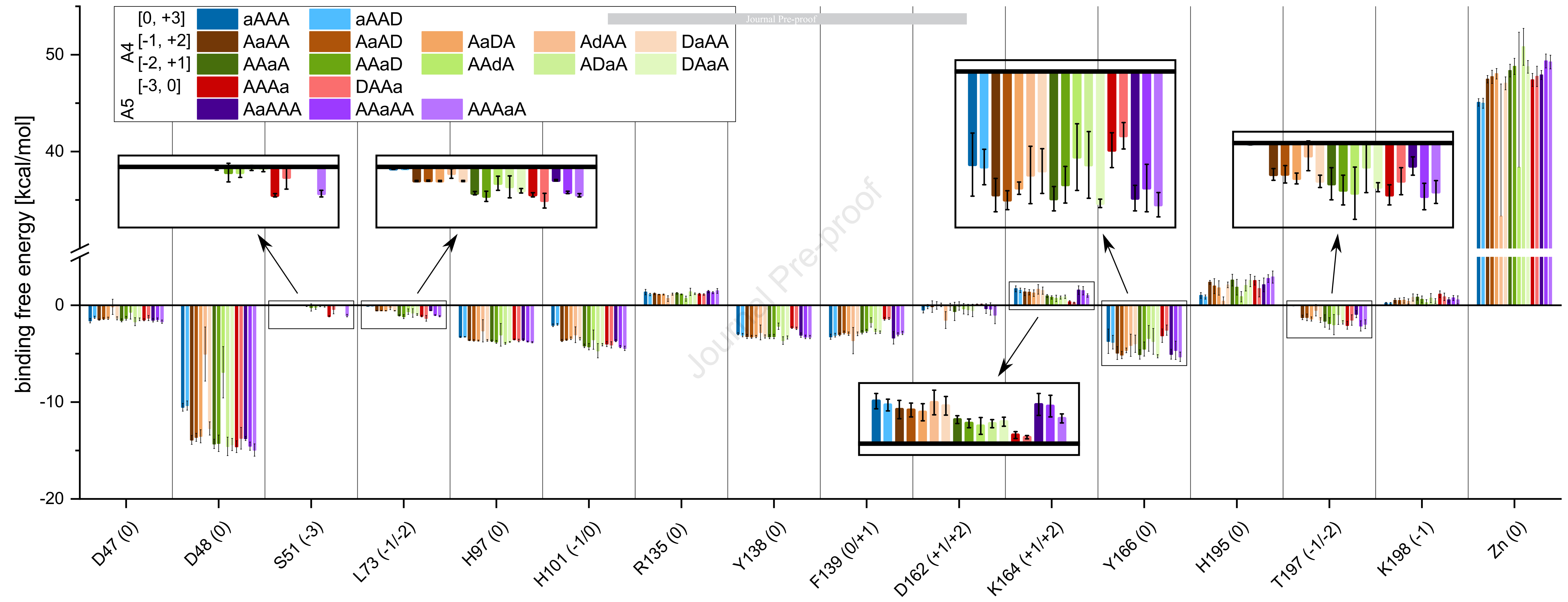
C



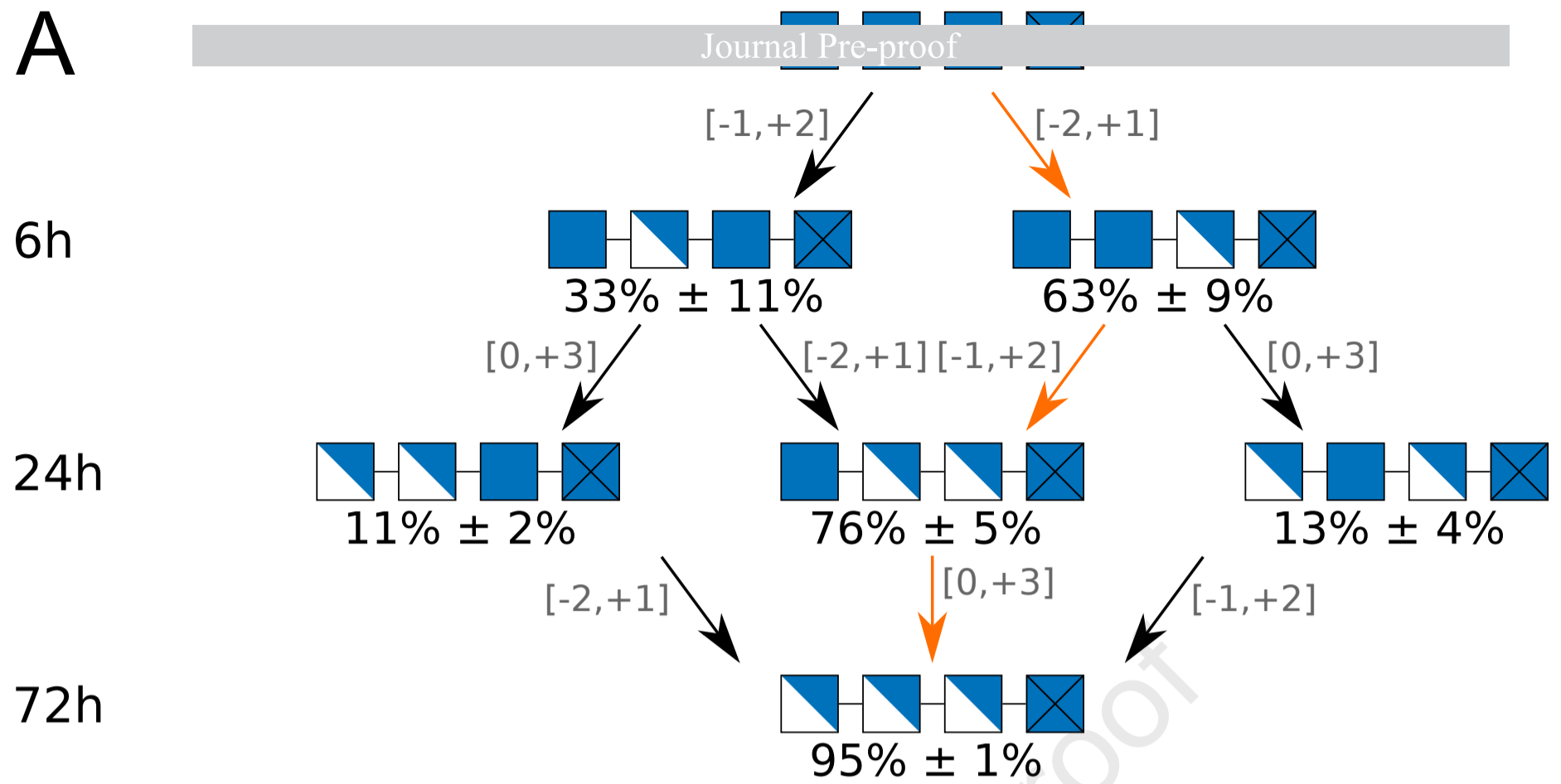




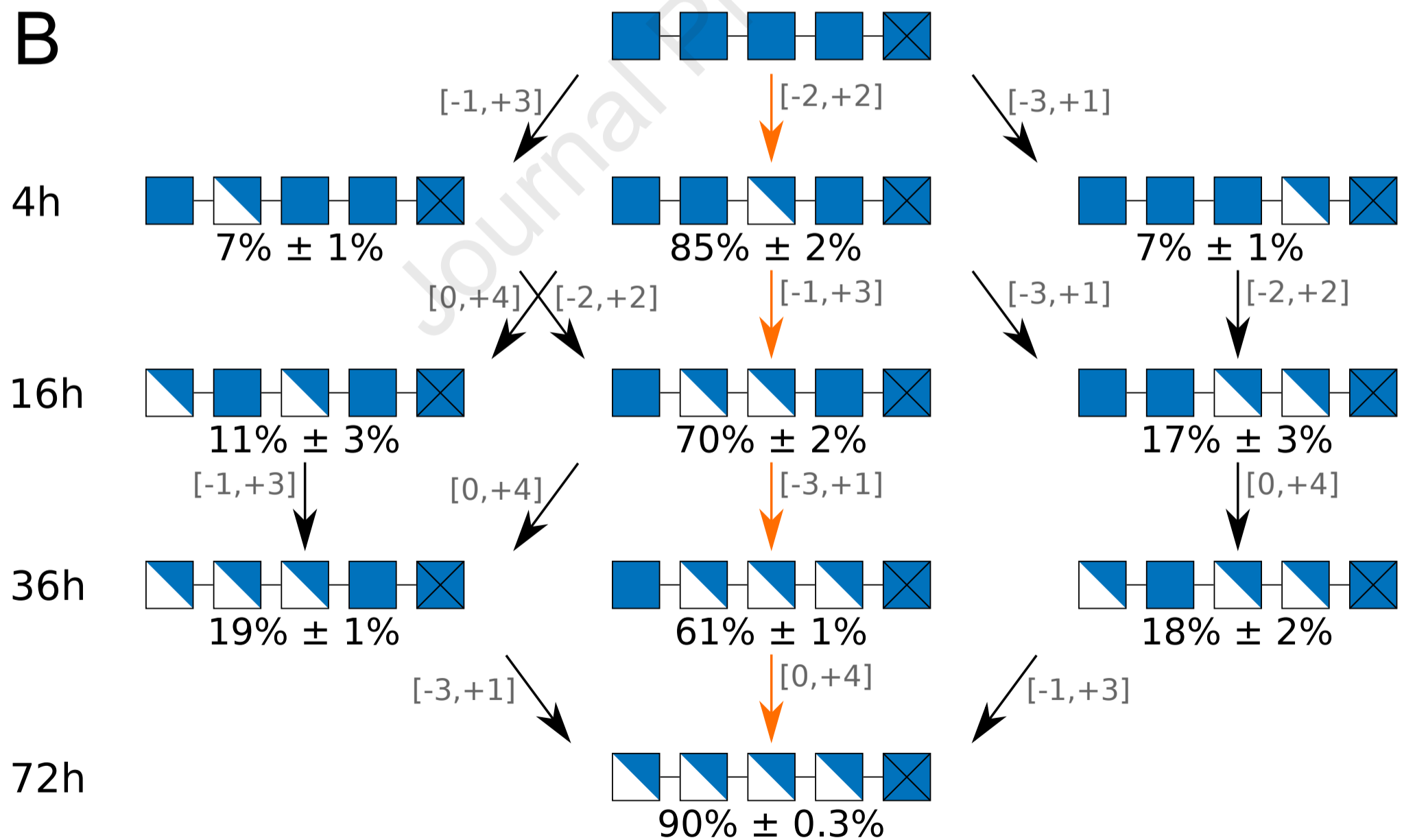




A



B



= *N*-acetylglucosamine
 = glucosamine
 = reducing end

