RESEARCH ARTICLE

The structural principles underlying molybdenum insertase complex assembly

Ahmed H. Hassan[1,2] | Christian Ihling[3,4] | Claudio Iacobucci[3,4,5] |
Panagiotis L. Kastritis[6] | Andrea Sinz[3,4] | Tobias Kruse[1]

[1]TU Braunschweig, Institute of Plant Biology, Braunschweig, Germany

[2]Central European Institute of Technology, Masaryk University, Brno, Czech Republic

[3]Department of Pharmaceutical Chemistry & Bioanalytics, Institute of Pharmacy, Halle (Saale), Germany

[4]Center for Structural Mass Spectrometry, Halle (Saale), Germany

[5]Department of Physical and Chemical Sciences, University of L'Aquila, L'Aquila, Italy

[6]ZIK HALOmem and Institute of Biochemistry and Biotechnology, Martin-Luther University Halle-Wittenberg, Halle (Saale), Germany

Correspondence
Tobias Kruse, TU Braunschweig, Institute of Plant Biology, Spielmannstrasse 7, 38106 Braunschweig, Germany.
Email: t.kruse@tu-bs.de

Review Editor: Nir Ben-Tal

Abstract

Within the cell, the trace element molybdenum (Mo) is only biologically active when complexed either within the nitrogenase-specific FeMo cofactor or within the molybdenum cofactor (Moco). Moco consists of an organic part, called molybdopterin (MPT) and an inorganic part, that is, the Mo-center. The enzyme which catalyzes the Mo-center formation is the molybdenum insertase (Mo-insertase). Mo-insertases consist of two functional domains called G- and E-domain. The G-domain catalyzes the formation of adenylated MPT (MPT-AMP), which is the substrate for the E-domain, that catalyzes the actual molybdate insertion reaction. Though the functions of E- and G-domain have been elucidated to great structural and mechanistic detail, their combined function is poorly characterized. In this work, we describe a structural model of the eukaryotic Mo-insertase Cnx1 complex that was generated based on cross-linking mass spectrometry combined with computational modeling. We revealed Cnx1 to form an asymmetric hexameric complex which allows the E- and G-domain active sites to align in a catalytic productive orientation toward each other.

KEYWORDS

biosynthesis complex, molybdenum cofactor, molybdenum insertase

1 | INTRODUCTION

For most species, Molybdenum (Mo) is an essential element (Zhang et al., 2011; Zhang & Gladyshev, 2010) as it owns an irreplaceable function for enzymes involved in the global C, N, and S metabolism (Stiefel, 2002). Here Mo is complexed within the molybdenum cofactor (Moco) whose synthesis involves a set of evolutionary highly conserved enzymes and reaction steps (Mendel & Kruse, 2012). Synthesis of Moco is commonly subdivided into four steps. In the first step of Moco biosynthesis, GTP is converted into cyclic pyranopterin monophosphate (cPMP; Hover et al., 2013; Wuebbens & Rajagopalan, 1993). Two enzymes, historically termed cPMP synthase, catalyze this reaction. In eukaryotes, these were localized in the mitochondria of the cell (Mendel, 2013; Mendel & Kruse, 2012; Teschner et al., 2010), while all subsequent pathway steps are localized in the cytoplasm (Mendel, 2013; Mendel & Kruse, 2012). Recent work identified $3',8\text{-}cH_2GTP$ as initial, short-life, first step intermediate (Hover et al., 2013, 2015). Subsequent work identified its formation to be under tight transcriptional control—at least in the fungus *Neurospora crassa* (Wajmann et al., 2020), suggesting that regulation of Moco biosynthesis is realized at the earliest stage possible. In the second pathway step cPMP is converted into molybdopterin (MPT), a tricyclic pterin derivative (Mendel & Kruse, 2012; Wuebbens & Rajagopalan, 2003). MPT is the organic part of Moco serving as Mo-chelating moiety. As such, MPT serves as a substrate for the molybdenum-insertase (Mo-insertase, recently reviewed in Kruse, 2022a) which catalyzes the third and fourth step of Moco biosynthesis. Mo-insertases are each composed of a functional G- and E-domain (Kruse, 2020; Mendel & Kruse, 2012). In prokaryotes, these occur as separate entities (Leimkühler et al., 2011), while—with the known exception of the green alga *Chlamydomonas reinhardtii* (Llamas et al., 2007)—they are fused in eukaryotes. Here an unstructured linker region of variable length connects the E- and G-domain (Heck et al., 2002; Probst et al., 2014; Smolinsky et al., 2008; Stallmeyer et al., 1995). For the human Mo-insertase gephyrin, shortening of the linker resulted in the depletion of Moco-synthesis activity, hence documenting its importance for proper domain interaction, metabolite-transfer and finally Moco-synthesis activity (Belaidi & Schwarz, 2013). Using the plant Mo-insertase Cnx1 as a model enzyme, the G-domain was shown to be required for MPT-adenylylation, yielding adenylated MPT (MPT-AMP; Kruse, 2022a; Kuper et al., 2004; Llamas et al., 2004). The reaction of ATP with MPT yields the AMP derivative MPT-AMP which should hence be named adenylyl MPT instead of adenylated MPT. However, as the term adenylated MPT has been established upon the identification of MPT-AMP in 2004 (Kuper et al., 2004) and is commonly used in the literature, throughout this work, MPT-AMP is termed adenylated MPT (Kruse, 2022a). The E-domain was found to be involved in the subsequent Mo-insertion reaction (Llamas et al., 2006), yielding the recently identified and unexpected novel intermediate Moco-AMP (Probst et al., 2021), which throughout this work will be termed adenylated Moco (see above; Kruse, 2022a). The liberation of Moco is then achieved by the Mg^{2+} dependent hydrolysis of Moco-AMP (Figure 1; Llamas et al., 2006; Probst et al., 2021).
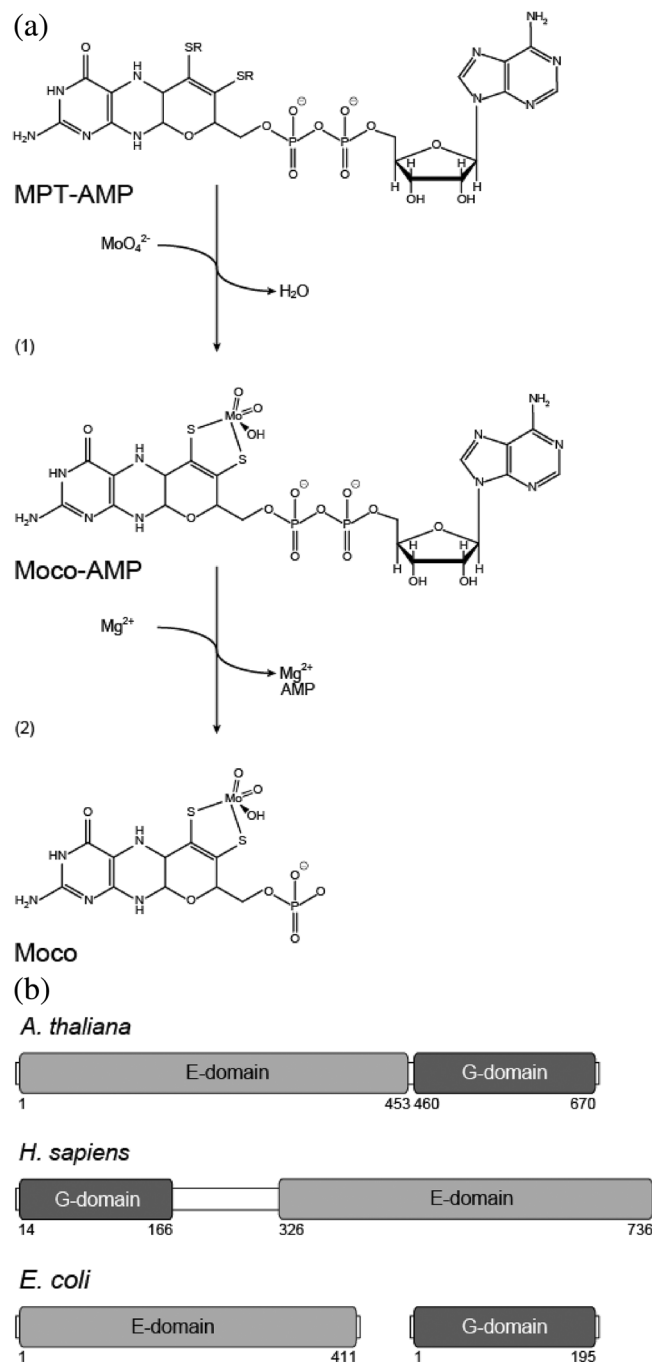

FIGURE 1 Legend on next page.

Subsequently, Moco may either be transferred to the user enzymes directly (Llamas et al., 2006), or is handed over to proteins of the cellular Moco transfer/storage-system (Fischer et al., 2006; Hercher, Krausze, Yang, et al., 2020; Krausze et al., 2020; Kruse et al., 2010; Mendel & Kruse, 2012), recently reviewed in (Kruse, 2022b). However, the successful crystallization of a full-length Mo-insertase (eukaryotes) or a Mo-insertase complex assembled from E- and G-domain(s) (prokaryotes) has not yet been reported. Therefore, the principles behind the interaction of its domains and the linked metabolite-transfer are open to speculation. The mammalian Mo-insertase gephyrin was suggested to adopt mesh-like structures in the cell. This theory was supported by a SAXS study; however, there are no high-resolution structures available so far confirming this model (Kneussel & Betz, 2000; Sander et al., 2013). For the plant Mo-insertase Cnx1, it is generally assumed that its structure conforms with that of full-length gephyrin. Here we report the first evidence-based structural model of the plant (*A. thaliana*) full-length Mo-insertase Cnx1. In this work, we gained insight into the Mo-insertase complex assembly by employing a cross-linking mass spectrometry (XL-MS) approach in combination with computational modeling. The details behind domain–domain interaction within the Cnx1 complex have been deciphered which suggests the hexameric complex to possess two functionally active sites.

2 | RESULTS

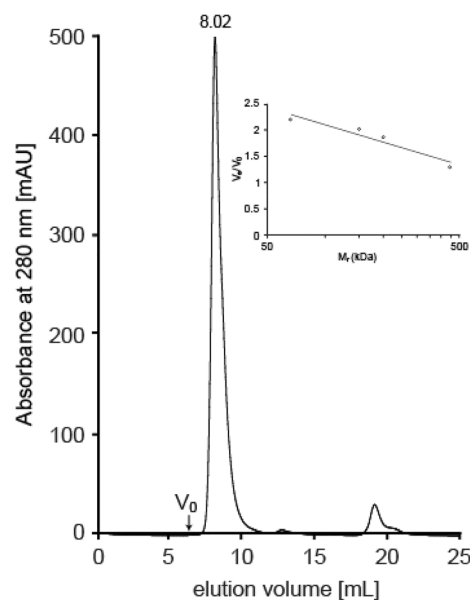### 2.1 | XL-MS of recombinant full-length Cnx1 protein

Structure-based work identified the final Moco biosynthesis intermediates MPT-AMP (Kuper et al., 2004) and Moco-AMP (Probst et al., 2021) bound to the trimeric Cnx1G (Kuper et al., 2004; Llamas et al., 2004) and dimeric Cnx1E (Krausze et al., 2017; Llamas et al., 2006) domain and hence unambiguously confirmed the active sites of these. However, as yet it is not understood how MPT-AMP is transferred within the Cnx1 full-length protein (i.e., from the G-domain active site to the E-domain active site). To shed light on the principle(s) behind the transient E- and G-domain interaction, we employed a structural proteomics approach relying on XL-MS. Using the plant (*A. thaliana*, Cnx1) Mo-insertases as its separately expressed domains are the established model enzymes to study eukaryotic Mo-insertase functionality (Kruse, 2022a). Upon cross-linking of the complex, we determined the molecular weight of cross-linked Cnx1 full-length preparations and found the Cnx1 full-length protein to form a complex of ca. 475 kDa (Figure 2).

Next, protein assemblies were analyzed via SDS-PAGE, which revealed the formation of two higher molecular weight complexes (ca. 200 kDa, Cnx1 complex 1 and ca. 350 kDa, Cnx1 complex 2) documenting the formation of two covalently linked E-G protein complexes as confirmed by MS-based analysis (Figure 3).

For the Cnx1 full-length complexes 1 and 2 in total 87 and 21 cross-links were identified (Figure 3, Tables S1 and S2). All Cnx1E/G-residues of the high-molecular weight complex 1 that were found to be involved in inter-molecular

FIGURE 1 The final step of molybdenum cofactor biosynthesis. (a) The reaction intermediates of the last two steps within molybdenum cofactor (Moco) biosynthesis are shown. These final steps are catalyzed by the molybdenum-insertase (Mo-insertase) E-domain which is required to insert molybdate into adenylated molybdopterin (MPT-AMP). Subsequently, Moco-AMP (adenylated Moco) is cleaved in a Mg^{2+} dependent reaction yielding physiologically active Moco (Llamas et al., 2006). (b) Mo-insertases possess two functional domains, namely G- and E-domain. The G-domain is required for the adenylylation of MPT (Kruse, 2022a; Kuper et al., 2004; Llamas et al., 2004). Both domains are either fused together (most eukaryotes; Llamas et al., 2007; Mendel & Kruse, 2012) or are expressed as separate entities (prokaryotes; Leimkühler et al., 2011). Domain annotation was carried out according to Krausze et al. (2017) (Cnx1, CAB38312), Schwarz et al. (2001) (Gephyrin, Q9NQX3), Xiang et al. (2001) (MoeA, P12281), Liu et al. (2000) (MogA, P0AF03). In Kruse (2022a), the function of Mo-insertases is summarized.

FIGURE 2 Size determination of the recombinant Cnx1 protein complex. Gel-filtration chromatogram from the Cnx1 full-length complex. The complex elutes at 8.02 mL. The cross-linked complex was loaded on a calibrated Superdex 200 increase analytical column using 100 mM HEPES-KOH, 300 mM NaCl, 1 mM EDTA, 5% glycerol as running buffer. The inset shows the calibration of the column with standard proteins.
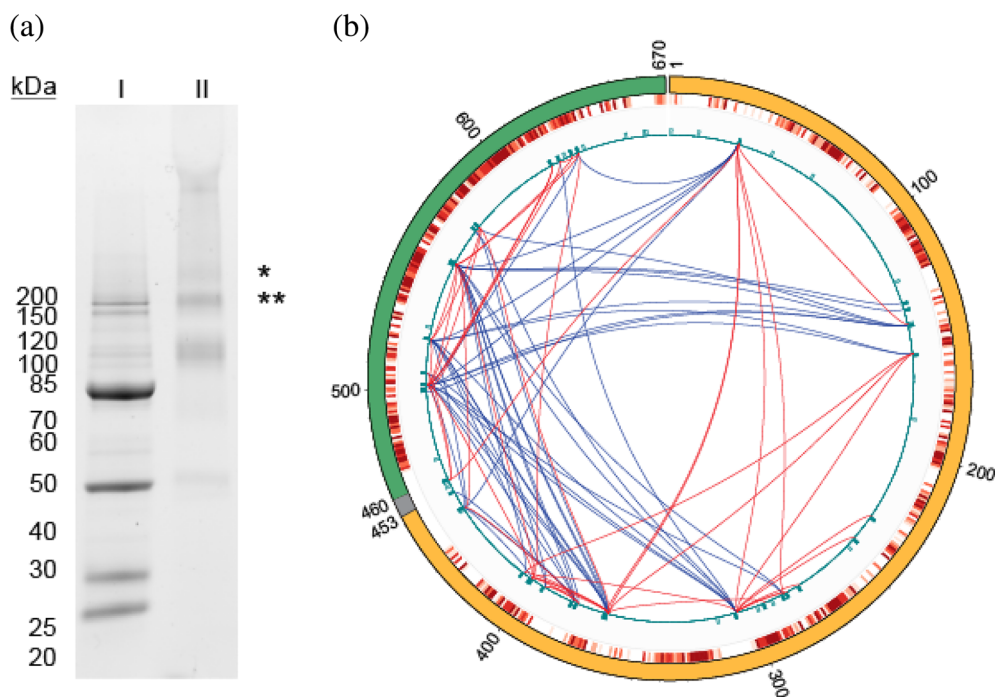
FIGURE 3 In vitro cross-linking of recombinant Cnx1 full-length. (a) TGX-stained SDS PA-gel: Lane I documents the purity of the recombinant Cnx1 protein used for XL-MS experiments. Ten micrograms were loaded. Lane II documents the assembly of high-molecular weight complexes (marked by asterisks) upon addition of the cross-linker DSBU to recombinant Cnx1 full-length protein (see Section 4 for details). (b) The circular plot shown was generated based on the XL-MS data obtained for the ca. 350 kDa complex (a, labeled with*). All inter-domain cross-links are highlighted in blue, and the intra-domain cross-links are highlighted in red. All lysine residues are highlighted in blue. The degree of conservation of the residues is colored from highly conserved (red) to non-conserved (white) (see Section 4 for details; Ashkenazy et al., 2016). The circular plot shows a total of 87 cross-links, either inter cross-links between E-domain (bright orange), G-domain (teal) and/or linker (gray), or intra cross-links within the E- and G-domains itself. Cnx1 homologs in other species were identified with the HMMER homolog search algorithm (1 iteration, E-value cut-off = 0.0001) using the UniRef90 protein database (Suzek et al., 2015). The result obtained was used as input file for the circular plot which was created using the software CX-Circos (http://cx-circos.net).

cross-links (50 out of 87) were mapped on the E- and G-domain protein surface. Therefore the Cnx1E structure with the highest percentage of structurally solved residues amongst all available Cnx1E structures (Hercher, Krausze, Hoffmeister, et al., 2020; Krausze et al., 2017, 2018; Probst et al., 2021) and the Cnx1G structure co-crystallized with the E-domain substrate MPT-AMP (Kuper et al., 2004) were used (Figure 4c). Doing so revealed that a subset of E-domain residues involved in inter-domain cross-links surround the active site (Figure 4c, part I). Notably, this was found to be not the case for the G-domain where involved amino acid residues were found to be scattered across two opposing sites of the G-domain trimer, with the notable difference, that interactions primarily involved one site of the trimer (please compare Figure 4c, part I vs. part II).

From the 87 cross-links obtained for full-length Cnx1 complex 1, the 50 inter-domain cross-links (between the two different Cnx1 domains) were considered in a modeling approach devoted to resolve the Cnx1E/G-domain interaction. However, as 13 inter-domain cross-links involve residues fall into the structurally unresolved parts of Cnx1E (Figure 4a, 5G2S; Krausze et al., 2017; residue

K444, S452) and Cnx1G (1UUY; Kuper et al., 2004; K630) or the likewise structurally unresolved linker region (K458), these were not used for model building. Notably, amino acid residues involved in inter-domain cross-links are found on both sites of the E-domain dimer which holds true to lesser extent also for the G-domain trimer. Consequently, we conclude that more than one orientation may be existent in which E- and G-domain interact with each other. Consistently, it was not possible to establish a single (E + G) interaction model that includes all obtained inter-domain cross-links. Therefore, in the following, we subdivided the cross-linking set assuming that each side of the G-domain interacts with two specific sets of E-domain residues. Doing so revealed two interaction models of G- and E-domain (Figure 5).

Notably, the cross-links used for model building were all derived from the high molecular weight complex 1 - (Figure 3). However, also the cross-links obtained from the high molecular weight complex 2 (Figure 3) are fully consistent with our interaction models shown in Figure 5. However, overall significantly less inter-domain cross-links were obtained from high molecular weight complex

2. Subsequently, we analyzed the two Cnx1E-G interaction models. The interaction interface derived from interaction model 1 was also found to be largely provided by Cnx1E subdomains I + II (23/55 residues) and subdomain IV: 25/55 residues, Figure 6a,c). However, to a minor extent also subdomain III contributes here (7/55 residues). In total 16 H-bond and 7 salt bridge interactions were identified within the modeled complex (Figure 5c).
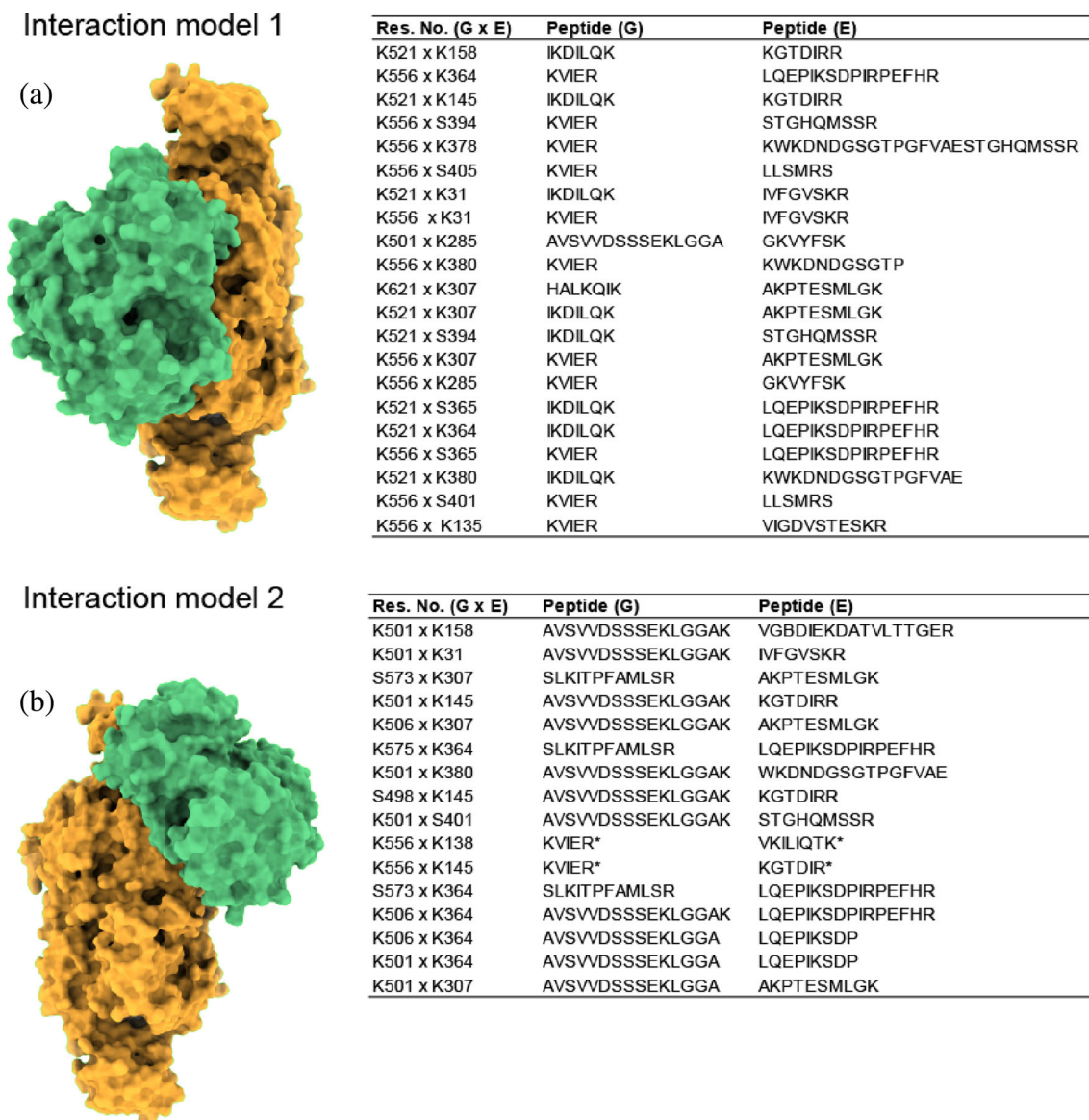
The interaction surface of interaction model 2 involves in total 58 residues (from the E-domain) which were found to be distributed in about equal parts to Cnx1E subdomains I + II (31/58 residues) and IV (25/58 residues) (Figure 6b,c), while only 2 residues were provided from subdomain III. In total 18 H-bond interactions and four salt bridge interactions were identified within the modeled complex (Figure 6c). Importantly, as the full-length Cnx1 protein consists of one E- and one G-domain, the proposed interaction models (Figure 5a,b) cannot occur separately. However, assembly of the two

interaction models within one Cnx1 complex results in a two (trimeric G-domain) to three (dimeric E-domain) stoichiometry which corresponds to a calculated molecular weight of 438 kDa conforming with our results from size exclusion chromatography (Figure 2). Thus formed, the complex does not allow the further recruitment of Cnx1 full-length proteins (Figure 7).

For modeling of the interaction models 1 and 2 (Figure 5) and the complex assembled from these (Figure 7), hitherto exclusively the obtained inter-domain cross-links were used. In a following step, we tested the remaining intra-domain cross-linking set (i.e., the "E–E" and "G–G" cross-links) for consistency with our Cnx1 complex models (Figures 5 and 7). From the 37 intra-domain cross-links, 31 are fully consistent with interaction models 1 and 2, while the remaining 6 intra-domain cross-links became explainable considering that the Cnx1 complex assembles from both interaction models (Figure 5, Figure S1, Tables S6 and S7), thus providing an additional
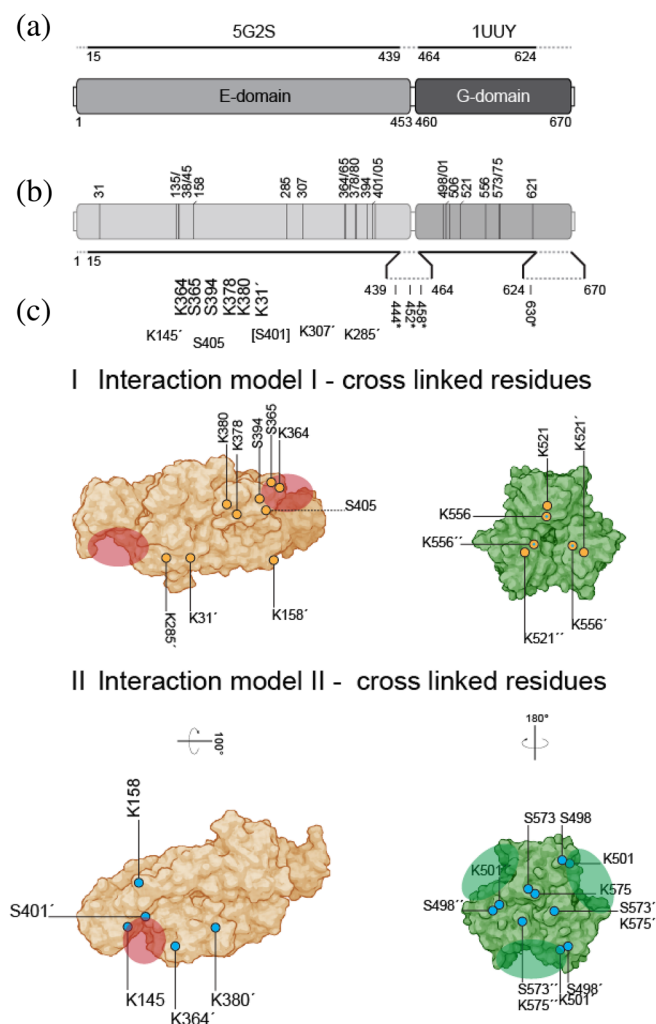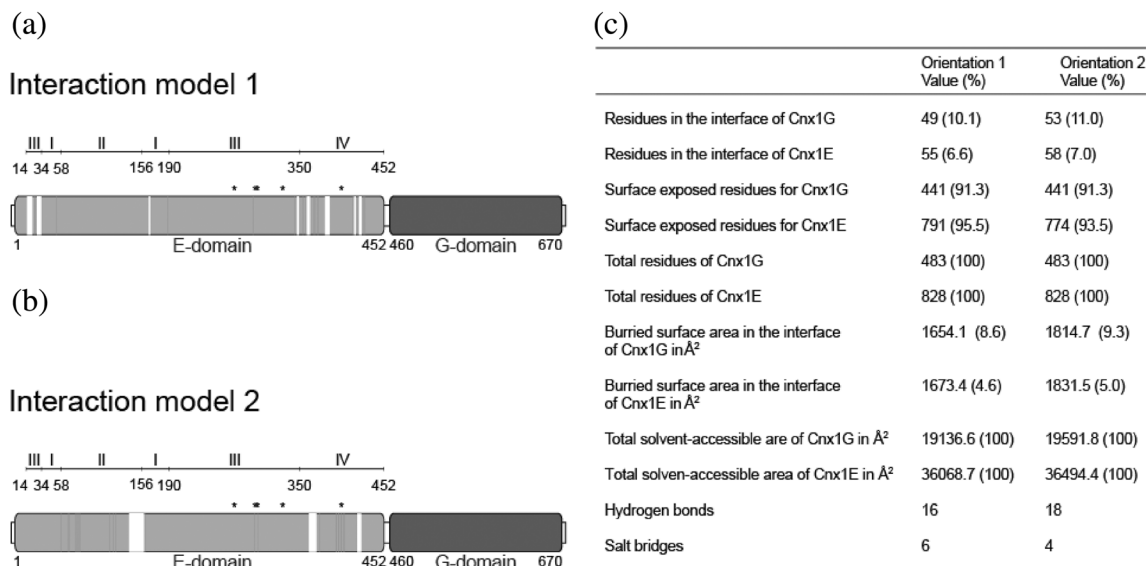

FIGURE 4 In vitro cross-linking of recombinant full-length Cnx1. (a) Schematic representation of full-length Cnx1 with the E- and G-domain colored gray and dark gray respectively. Numbers refer to the amino acid positions (Q39054; Stallmeyer et al., 1995). The PDB codes for Cnx1E (Krausze et al., 2017) and Cnx1G (Kuper et al., 2004) are given above. Residues that have been resolved in the structures are indicated. (b) Position of the Cnx1 amino acids identified in cross-linked Cnx1E-G peptides. Below the structurally resolved parts of the Cnx1 domains are indicated by a black line. Residues that fall into the structurally unresolved parts of the protein are indicated with an asterisk. The first and last resolved amino acid of Cnx1E (Krausze et al., 2017; PDB entry 5G2S) and Cnx1G (Kuper et al., 2004; PDB entry 1UUY) are given. (c) Part I and II: Front view (I) and back view (II) of Cnx1E monomers a and b (PDB entry 5G2S) and Cnx1G monomers a, b, and c (PDB entry 1UUY) with the cross-linked residues indicated. Only residues that were involved in cross-links classified as satisfied (see Table S6) are shown. Orange = residues used for modeling of interaction model 1, cyan = residues used for modeling of interaction model 2. A dotted line indicates that the residue is shielded by the protein surface as a result from the viewing angle. Residue K556 lies at the edge on one side of the trimer and hence has been used to model both, the interaction 1 and the interaction 2 models as indicated by an additional, cyan-colored circle. Interaction model II: The Cnx1E dimer has been additionally rotated by 100° as indicated in the figure. Numbering refers to the numbering of the full-length enzyme (Q39054; Stallmeyer et al., 1995). The Cnx1E and Cnx1G active sites are indicated by a colored and transparent circle for Cnx1E (red) and G (green) respectively. Note that all Cnx1G active sites are shielded by the protein as a result from the viewing angle in the interaction model I representation. Same for one of the two Cnx1E active sites in the dimer (Krausze et al., 2017; Probst et al., 2021) in the interaction model II representation. E- and G-domain are not shown in scale.

Interaction model 1

(a)

Res. No. (G x E)	Peptide (G)	Peptide (E)
K521 x K158	IKDILQK	KGTDIRR
K556 x K364	KVIER	LQEPIKSDPIRPEFHR
K521 x K145	IKDILQK	KGTDIRR
K556 x S394	KVIER	STGHQMSSR
K556 x K378	KVIER	KWKDNDGSGTPGFVAESTGHQMSSR
K556 x S405	KVIER	LLSMRS
K521 x K31	IKDILQK	IVFGVSKR
K556 x K31	KVIER	IVFGVSKR
K501 x K285	AVSVVDSSSEKLGGA	GKVYFSK
K556 x K380	KVIER	KWKDNDGSGTP
K621 x K307	HALKQIK	AKPTESMLGK
K521 x K307	IKDILQK	AKPTESMLGK
K521 x S394	IKDILQK	STGHQMSSR
K556 x K307	KVIER	AKPTESMLGK
K556 x K285	KVIER	GKVYFSK
K521 x S365	IKDILQK	LQEPIKSDPIRPEFHR
K521 x K364	IKDILQK	LQEPIKSDPIRPEFHR
K556 x S365	KVIER	LQEPIKSDPIRPEFHR
K521 x K380	IKDILQK	KWKDNDGSGTPGFVAE
K556 x S401	KVIER	LLSMRS
K556 x K135	KVIER	VIGDVSTESKR

Interaction model 2

(b)

| Res. No. (G x E)	Peptide (G)	Peptide (E)
K501 x K158	AVSVVDSSSEKLGGAK	VGBDIEKDATVLTTGER
K501 x K31	AVSVVDSSSEKLGGAK	IVFGVSKR
S573 x K307	SLKITPFAMLSR	AKPTESMLGK
K501 x K145	AVSVVDSSSEKLGGAK	KGTDIRR
K506 x K307	AVSVVDSSSEKLGGAK	AKPTESMLGK
K575 x K364	SLKITPFAMLSR	LQEPIKSDPIRPEFHR
K501 x K380	AVSVVDSSSEKLGGAK	WKDNDGSGTPGFVAE
S498 x K145	AVSVVDSSSEKLGGAK	KGTDIRR
K501 x S401	AVSVVDSSSEKLGGAK	STGHQMSSR
K556 x K138	KVIER*	VKILIQTK*
K556 x K145	KVIER*	KGTDIR*
S573 x K364	SLKITPFAMLSR	LQEPIKSDPIRPEFHR
K506 x K364	AVSVVDSSSEKLGGAK	LQEPIKSDPIRPEFHR
K506 x K364	AVSVVDSSSEKLGGA	LQEPIKSDP
K501 x K364	AVSVVDSSSEKLGGA	LQEPIKSDP
K501 x K307	AVSVVDSSSEKLGGA	AKPTESMLGK

FIGURE 5 Cnx1E possesses two Cnx1G interaction sites. Cnx1E/G-domain interaction model 1 (a) and 2 (b). The models were generated based upon a subset of inter-domain cross-links tabulated next to the image of the respective complex. The cross-linked amino acid sequences were included (peptide G/peptide E). Modeling was carried out using the HADDOCK server (Koukos et al., 2020) and as described in Section 4. Cnx1E (orange 5G2S; Krausze et al., 2017) and Cnx1G (green, 1UUY; Kuper et al., 2004) are shown in surface representation. The image was prepared using Chimera (Pettersen et al., 2004). *These cross-links were included into modeling based on the high flexibility of the E-domain part targeted by the cross-linker (see Figure S2).

line of evidence for the accuracy of the model. Finally, we modeled the unresolved linker region connecting the Cnx1E$_{C-Terminus}$ and Cnx1G$_{N-Terminus}$ employing de novo modeling (Kim et al., 2004; Mashiach et al., 2010). Next to the annotated Cnx1 linker region (spanning amino acids 454–459, Q39054; Krausze et al., 2017; Stallmeyer et al., 1995) the lacking parts of the Cnx1E (Cnx1E$_{440-453}$; Krausze et al., 2017) and the Cnx1G (Cnx1G$_{460-463}$; Kuper et al., 2004) structures were likewise modeled. Subsequently, the Cnx1 complex (Figure 7) was completed with these modeled parts (Figure 8).

No clashes of these within the complex were observed which we consider as additional supportive evidence for the accuracy of our Cnx1 model. Most importantly, none of the modeled protein parts within the Cnx1 complex was found to shield the Cnx1E- and G-domain active sites here, hence being not in conflict with the metabolite-transfer. The arrangement of the different Cnx1 full-length molecules within the complex is illustrated in Figure 9, which illustrates that the hexameric Cnx1 complex is asymmetric due to a high inter-domain flexibility of the assembling monomers.

(a)

Interaction model 1

(b)

Interaction model 2

(c)

	Orientation 1 Value (%)	Orientation 2 Value (%)
Residues in the interface of Cnx1G	49 (10.1)	53 (11.0)
Residues in the interface of Cnx1E	55 (6.6)	58 (7.0)
Surface exposed residues for Cnx1G	441 (91.3)	441 (91.3)
Surface exposed residues for Cnx1E	791 (95.5)	774 (93.5)
Total residues of Cnx1G	483 (100)	483 (100)
Total residues of Cnx1E	828 (100)	828 (100)
Burried surface area in the interface of Cnx1G in Å²	1654.1 (8.6)	1814.7 (9.3)
Burried surface area in the interface of Cnx1E in Å²	1673.4 (4.6)	1831.5 (5.0)
Total solvent-accessible are of Cnx1G in Å²	19136.6 (100)	19591.8 (100)
Total solven-accessible area of Cnx1E in Å²	36068.7 (100)	36494.4 (100)
Hydrogen bonds	16	18
Salt bridges	6	4

FIGURE 6   Interaction surfaces within the Cnx1E-G complex. (a + b) Schematic representation of Cnx1E with the subdomains (I, II, III, and IV) indicated. Subdomains were annotated as published earlier (Krausze et al., 2017), whereby the numbers below indicate the first residue of each subdomain. Residue # 452 is the last residue of subdomain IV. Residues that contribute to the interaction interface of interaction models 1 (a) and 2 (b) (see Figure 5) are indicated by white lines or boxes respectively. Asterisks indicate the positions of the Cnx1E active site residues K294, K297, S328, R369, S400 (Krausze et al., 2018). (c) Tabulated properties of the interaction interfaces derived from interaction models 1 and 2 (Figure 3). The data were derived from a PISA (Krissinel & Henrick, 2007) based analysis of interaction models 1 and 2 (see Figure 5 for comparison).
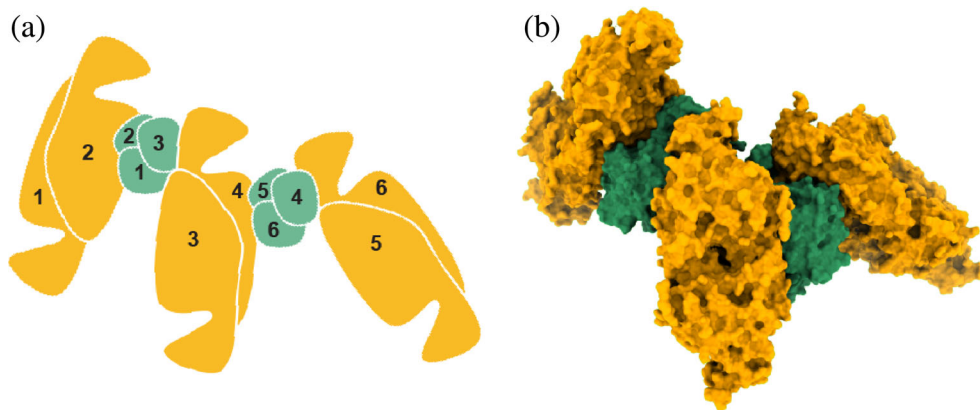

FIGURE 7   The Cnx1 complex. (a) Schematic representation of dimeric Cnx1E and trimeric Cnx1G assembly within the Cnx1 complex. Numbers refer to the respective full-length proteins involved in complex assembly (see Figure 9 for comparison). (b) Surface representation of the Cnx1 complex schematically shown in (a). Cnx1E is shown in orange, Cnx1G is shown in green. The image was generated with Chimera (Pettersen et al., 2004).

3 | DISCUSSION

Within this work, we employed a XL-MS approach to reveal the assembly of Cnx1E and G-domains within the Mo-insertase complex. The current model of eukaryotic Mo-insertase complex interaction suggests that E- and G-domains interact via a central G-domain trimer which attaches to distal E-domain dimers, thus forming a mesh-like structure (Kneussel & Betz, 2000; Sander et al., 2013).

Notably, our data for the plant Mo-insertase Cnx1 does not confirm this model of E- and G-domain complex formation but suggests that the Mo-insertase forms a compact complex that assembles from six full-length proteins. The majority of the cross-links satisfied the data-driven-derived biomolecular model; however, some cross-link distances were violated. Upon model inspection and distance calculations, it is evident that these cross-links may well be satisfied in the hexameric state upon minor
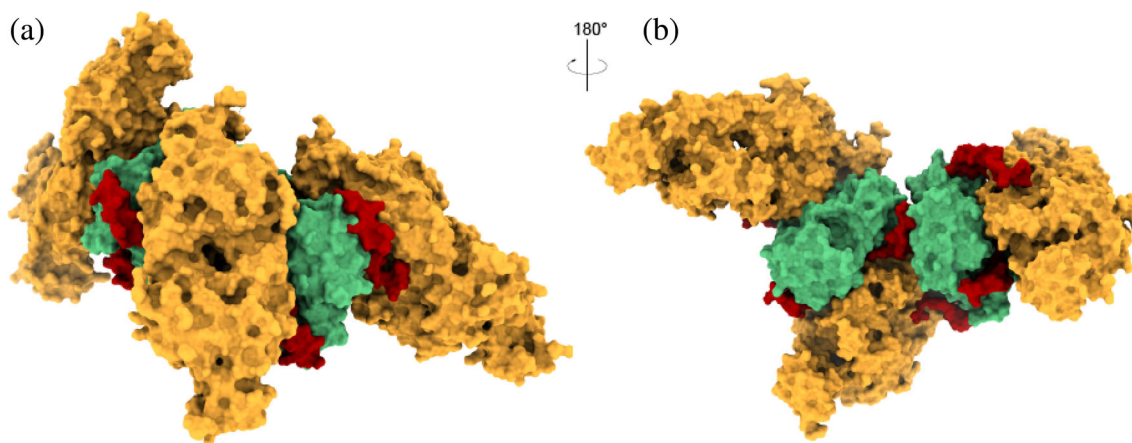
FIGURE 8 De novo modeling of the structurally unresolved parts of the Cnx1 complex. Surface representation of the Cnx1 complex shown in Figure S1 including the de novo modeled, unresolved Cnx1E (PDB entry 5G2S; Krausze et al., 2017) and Cnx1G (PDB entry 1UUY; Kuper et al., 2004) parts. Cnx1E is shown in orange, Cnx1G is shown in green, and the modeled parts are shown in red. Modeling was done using Robetta (Kim et al., 2004); the image was generated with Chimera (Pettersen et al., 2004). The orientation of each domain within the respective Cnx1 full-length monomers is detailed in Figure 9.

conformational re-arrangements within the complex. Consistent data comes from size exclusion chromatography which assigned a molecular mass of ca. 475 kDa to the complex which is in good agreement with its calculated molecular mass (438 kDa). However, as a 350 kDa sized cross-linked complex was identified via SDS PAGE, we assume that here next to the full-length protein unknown stoichiometries of separately occurring E- and G-domain assemble within the complex and consistently upon cross-linking; these were no longer detectable in the analyzed sample. Our detailed investigation of the Cnx1E/G interaction interfaces within this complex revealed that these involve Cnx1E subdomains I + II and IV, but to a significantly less extent subdomain III. Notably, this attributes a novel function to these subdomains which were hitherto assigned to assist in Moco-AMP binding (subdomains I + II; Probst et al., 2021) and to provide part of the molybdate binding site (subdomain IV; Krausze et al., 2017; Probst et al., 2021).

However, the distance between the E- and G-domain active sites within the complex (interaction model 2, Figure 10a) excludes a direct handover of MPT-AMP from G- to E-domain active site, raising the question how metabolite-transfer can nonetheless occur. As concluded from the local RMSD values of the two structurally most diverting Cnx1E structures 5G2S and 6RMS (Figure 10b, Figure S2) we identified Cnx1E subdomains I + II to possess a highly flexible region right adjacent/partially overlapping to the primary G-domain interaction site (Figure 10). We suggest, that this flexibility is required for the MPT-AMP-transfer to occur as part of the subdomains I and II shield the Cnx1E MPT-AMP binding site (Figures 6b and 10a,c; Probst et al., 2021). Since these subdomains possess a high flexibility, we suggest, that within the metabolite-transfer reaction, both domains expose the MPT-AMP binding site. This in turn may allow the E- and G-domain active sites to come into closer proximity as determined for our interaction 2 model (Figure 10a,c). Any movement of Cnx1E subdomains I + II has not been considered for model building so far. Notably, the restrains used for model building (interaction model 2) allow the face-to-face positioning of the E- and G-domain within the complex, however only when subdomains I + II expose the substrate binding site (data not shown). Considering all E/G-domain interactions within the Cnx1 complex reveals overall two Cnx1E active sites involved in catalysis, while the remaining four active sites are not (Figure 7). It is tempting to assume, that in the cell under conditions of high Moco demand, separate Cnx1G may provide MPT-AMP to the remaining Cnx1E active sites thus enhancing the synthesis capacity of the complex. However, as yet it has not been investigated if Cnx1 cleavage occurs in the plant. At least recombinant full-length Cnx1 protein preparations contain significant amounts of free E- and G-domain arising from in cell linker cleavage, documenting both, accessibility and sensitivity of the linker toward cleavage.

4 | MATERIALS AND METHODS

4.1 | Cloning of cnx1

*cnx1* was cloned by PCR using *cnx1* full-length cDNAs as templates (Krausze et al., 2017). Full-length *cnx1* was

amplified by PCR using the Phusion® High-Fidelity DNA Polymerase (NEB) and appropriate primers (Cnx1_BamHI_for 5′-attaggatccatggaaggtcaaggttgttg-3′ and Cnx1_PmlI_rev 5′-taatcacgtggtgagtacaagaacatccag-3′). Subcloning of the PCR products was carried out using the CloneJET™ PCR Cloning Kit according to the manufacturer's instructions. Sequencing revealed a single base exchange in codon 279 (NM_122108.4) which results in a silent mutation and hence the construct was considered to be correct. Next full-length *cnx1* was subcloned into the BamHI and PmlI sites of a His$_6$ tag/Twin-Strep-tag® encoding vector described earlier (Ringel et al., 2013).
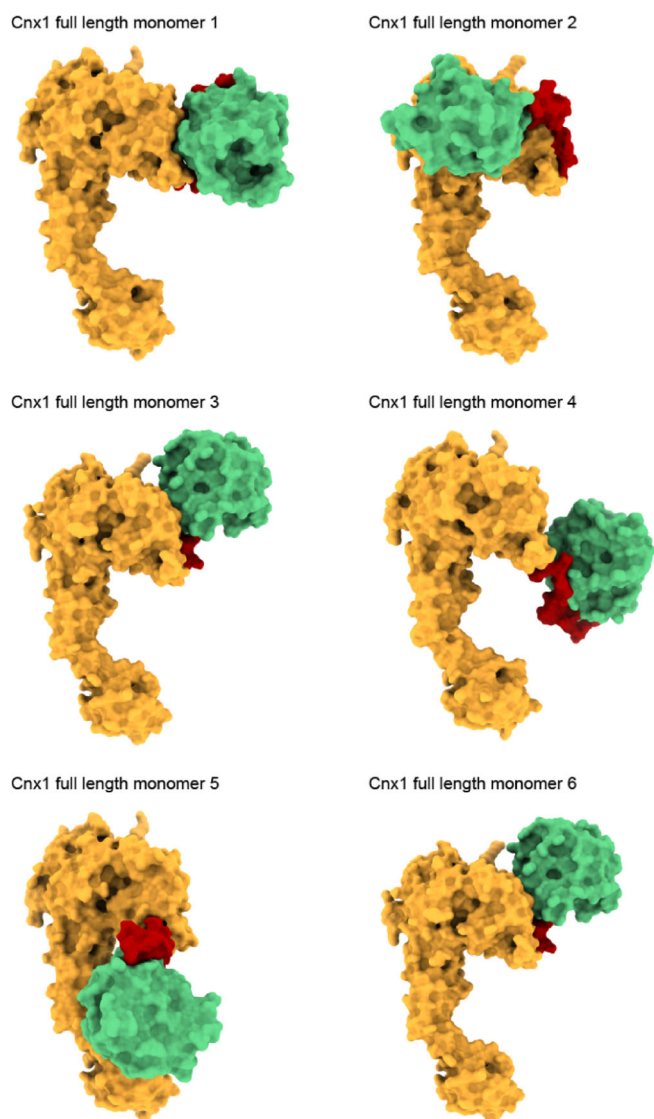
FIGURE 9 Orientation of E- and G-domain within full-length Cnx1 monomers forming the complex. The six full-length monomers were derived from the complex shown in Figure 8. The complex building Cnx1 monomers have been named according to Figure 7. Cnx1E (orange, PDB entry 5G2S) and Cnx1G (green, PDB entry 1UUY) are shown in surface representation. The image was prepared using Chimera (Pettersen et al., 2004).

4.2 | Expression and purification of recombinant proteins

Routinely, full-length Cnx1 was recombinantly expressed in *E. coli* strain TP1000 (Palmer et al., 1996). All expressions were carried out in 2YT media containing 50 μg/mL ampicillin. Recombinant gene expression was induced with 1 mM isopropyl 1-thio-β-D-galactopyranoside and cells were grown aerobically for
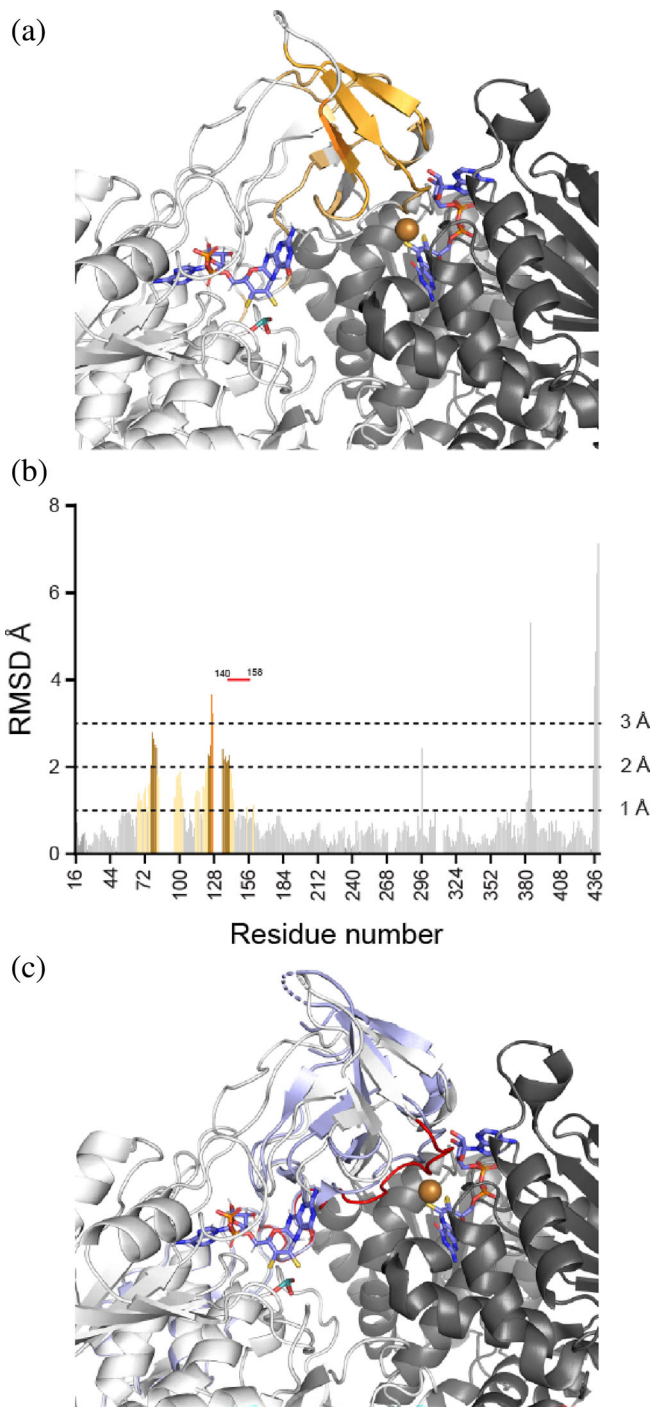
FIGURE 10 Legend on next page.

19 h and at 22°C. Upon harvesting, cell lysis was carried out using an Avestin Emulsiflex C5 homogenizer. For lysis of Cnx1 full-length expressing cells, Roche cOmplete™, EDTA-free protease inhibitor was added. To clarify the lysates, a subsequent centrifugation step was carried out (22,000×g, 60 min, 4°C). The following purification regime was employed for the purification of full-length Cnx1: Initially, recombinant Cnx1 was purified by immobilized metal affinity chromatography (IMAC). IMAC was carried out at 4°C, using Ni-NTA Agarose (Qiagen). Upon loading of the lysate, the gravity flow column was washed once using a wash buffer composed of 100 mM Tris–HCl pH 7.5, 150 mM NaCl, 20 mM imidazole, and 5% glycerol (v/v). For elution, a buffer composed of 100 mM Tris–HCl pH 7.5, 150 mM NaCl, 250 mM imidazole, and 5% glycerol has been used. Next, NiNTA purified protein was loaded onto Strep-Tactin® Superflow® high-capacity resin (IBA), following the manufacturer's description. As washing buffer, 100 mM HEPES-KOH pH 7.5, 300 mM NaCl, 1 mM EDTA, and 5% glycerol were used. For elution, the washing buffer was amended with 5 mM D-Desthiobiotin (IBA). Upon elution, the obtained protein was further purified employing gel filtration chromatography using an analytical Superdex 10/300 increase column (GE Healthcare) connected to an Äkta purifier system (Amersham Biosciences). The running buffer contained 100 mM HEPES-KOH pH 8.0, 300 mM NaCl, and 1 mM EDTA. For calibrating the column, molecular weight standards (GE Healthcare) were used according to the manufacturer's instructions. Fractions containing Cnx1 full-length protein were pooled and buffer exchanged, using PD10

columns (Amersham Biosciences) according to the manufacturer's description. The buffer was exchanged for 20 mM Tris–HCl pH 7.0. The final purification step employed anion exchange chromatography which was carried out using a self-packed XK16 column with SOURCE™ 15Q resin (GE Healthcare). As low salt buffer 20 mM Tris–HCl pH 7.0 was used. As high salt buffer, 20 mM Tris–HCl pH 7.0, 1 M NaCl was used. For purification of full-length Cnx1, an 80-min gradient was applied. Cnx1 full-length eluted over a range of 12.95–25.98 mS cm^{-1}. Cnx1-containing fractions were collected separately and analyzed via SDS PAGE prior to pooling. Twelve percent SDS-Page gels were TGX (Tris-Glycine eXtended)-stained and processed using ImageLab 6.0 with a ChemiDoc XRS+ system (Biorad). Pure protein fractions were concentrated (Vivaspin 6, 10 kDa cut-off, Sartorius AG) and stored in 20 µL aliquots in liquid nitrogen. For the structural proteomics approach, a shortened protocol was applied for the purification of full-length Cnx1, which comprised the Strep-Tactin® based purification and quality controlling steps described above using HEPES-KOH containing buffers.

4.3 | Cross-linking mass spectrometry

For cross-linking mass spectrometry (XL-MS) experiments, the homobifunctional, amine-reactive cross-linker disuccinimidyl dibutyric urea (DSBU, CF PLUS Chemicals) was used. Cross-linking reaction mixtures contained 10 µM Cnx1 that were incubated in 100 mM HEPES-KOH pH 7.5, 300 mM NaCl, 1 mM EDTA. The stock solution of DSBU (1 M) was freshly prepared in DMSO. The cross-linking reaction was carried out at room temperature and for 30 min using a 100-fold molar excess of DSBU over Cnx1E. The reaction was stopped by the addition of 20 mM Tris–HCl (pH 8.0). The reaction mixture was then loaded onto a 10% polyacrylamide gel. SDS-gels were then TGX-stained and processed using ImageLab 6.0 with a ChemiDoc XRS+ system (Biorad). High-molecular weight protein species (i.e., putative Cnx1E-G complexes) were excised from the gel, reduced with 10 mM dithiothreitol, and carbamidomethylated with 55 mM iodoacetamide. Afterwards, the samples were incubated overnight with GluC (1:20 (w/w) enzyme: protein ratio) at 37°C and digested with trypsin (Promega) for 4 h at 37°C.

4.4 | Liquid chromatography–tandem mass spectrometry

Liquid chromatography–tandem mass spectrometry (LC–MS/MS) analyses were performed using an Ultimate 3000

FIGURE 10 Orientation of Cnx1E and G active sites toward each other within the interaction 2 model. (a) Ribbon representation of the Cnx1E-G interaction 2 model. The active site bound metabolites MPT-AMP (Cnx1G, PDB entry 1UUY; Kuper et al., 2004) and Moco-AMP (Cnx1E, PDB entry 6Q32; Probst et al., 2021) are shown in stick representation. The distance between the Moco/MPT C2 atoms was defined and found to span 17.3 Å. Cnx1E is shown in light gray, Cnx1G is shown in dark gray. Flexible residues are colored in orange (see b). (b) Individual Cα RMSD values derived from the superimposition of Cnx1E structures 5G2S and 6RMS are shown in (c). Colors highlight residues that possess RMSD values >1, >2, and >3 Å respectively. The major interfacing part of subdomains I + II (residues 140–158, see Figure 5) within the Cnx1E-G complex is indicated by a red bar. (c) Same view of the Cnx1E-G interaction 2 model as shown in (a). Cnx1E (5G2S) is shown in light gray, the structurally most diverting Cnx1E structure (6RMS, see Figure S2) is shown in light blue. The major interfacing part of subdomains I + II (residues 140–158, see Figure 6) is colored in red. (a) and (c) images were generated with PyMOL (Lilkova et al., 2015). (b) The individual Cα RMSD values were calculated using Chimera (Pettersen et al., 2004).

RSLC nano-HPLC system coupled to an Orbitrap Fusion mass spectrometer equipped with nano-electrospray ionization (ESI) source (Thermo Fisher Scientific). For separation, a C8 reversed-phase (RP) precolumn and a C18 RP separation column were used. To desalt the samples, peptides were washed with 0.1% trifluoroacetic acid (TFA) for 15 min. 0.1% formic acid (FA) in water was used as solvent A, solvent B contained 0.08% FA in 80% acetonitrile (MS-grade, VWR). A linear 90 min gradient (0%–35% solvent B) with a flow rate of 300 nL/min was applied to elute the peptides. Data were acquired in data-dependent MS/MS mode using stepped HCD (high energy collisional dissociation, normalized collision energies [NCEs]: 27%, 30%, 33%). For data acquisition, each high-resolution full scan (m/z 300–1500, $R = 120{,}000$, target value [AGT] 400,000, max. injection time (IT) 50 ms) in the orbitrap was followed by high-resolution product ion scans ($R = 15{,}000$, AGT 50,000, IT 200 ms, isolation window 2 Th) for charge states 2+ to 7+ within 5 s, starting with the most intense signal in the full scan mass spectrum. Dynamic exclusion (duration 60 s, window ±2 ppm) was enabled. Data acquisition was controlled by Xcalibur software (version 4.1, Thermo Fisher Scientific).

4.5 | Identification of cross-linked products

Cross-linked products were identified using MeroX (version 1.6.6; Iacobucci et al., 2018). *C*-terminal cleavage at K and R was set for trypsin, while *C*-terminal cleavage at D and E was set for GluC. Three missed cleavage sites were allowed. Cross-linking site 1 was defined as K, cross-linking site 2 was defined as K, S, T, Y, and N-terminus. All cross-links generated by the software were manually inspected and visualized on the 3D protein structure by PyMOL (Lilkova et al., 2015). Results are tabulated in Tables S1 and S2.

4.6 | Modeling of the Cnx1 complex

Modeling of the complex was done using HADDOCK (Dominguez et al., 2003; van Zundert et al., 2016). For modeling, PDB file 5G2S (Cnx1E; Krausze et al., 2017) and PDB file 1UUY (Cnx1G; Kuper et al., 2004) were used. Therefore, initially, all water molecules and ligands were removed and chains and identifiers were renumbered using Coot (Emsley et al., 2010). Next, we used the validated set of cross-links (Table S1) to generate two groups based on their location on the E- and G-domain surface respectively (see Figure 4 for comparison). These have been included into the HADDOCK input files (table files; Bonvin et al., 2018). The restraint distance of Cα to Cα was set to a range of 4–30 Å (Bonvin et al., 2018; Iacobucci et al., 2018) for all of the cross-linked residues. Subsequently, the input files were included as ambiguous restraints active and passive residues were determined using Cport (de Vries & Bonvin, 2011; Table S3). For group 1, in total, six similar clusters were obtained from which cluster 2 was chosen for subsequent characterization upon manual assessment (see Table S4). For group 2, in total, four similar clusters were obtained from which cluster 1 was chosen for subsequent characterization upon manual assessment (see Table S5). From both clusters, the best scoring model was used for manual inspection which included an initial check for restrain accuracy. Further, the distances between the N-terminus of the G-domain and the C-terminus of the E-domain were determined and checked for consistency with the estimated maximum distance possible (defined by the unresolved structural parts of Cnx1E; Krausze et al., 2017; PDB 5G2S$_{440-453}$, the linker [Cnx1$_{454-459}$; Krausze et al., 2017] and Cnx1G [Kuper et al., 2004] PDB 1UUY$_{460-463}$; *in summa* 23 aa, 87.4 Å when assuming a Cα–Cα distance of 3.8 Å). Doing so revealed both models (group 1 = model 1; group 2 = model 2) to fulfill the criteria set as all retrains were within the range set and the distances between Cnx1E$_{C-terminus}$ and Cnx1G$_{N-terminus}$ were found to range between 40.7 Å and 66.8 Å. Results from the complex modeling are available as supplemental data (Data S1). In Table S6, the classification of cross-links (violated, satisfied) used to establish the interaction models 1 and 2 is tabulated. For the intra-domain complex, this is tabulated in Table S7.

4.7 | De novo modeling

The unresolved parts of Cnx1 (Cnx1$_{439-464}$) were modeled de novo using Robetta (Kim et al., 2004) by applying the TrRosetta de novo modeling approach. The modeled protein part was then positioned manually within the Cnx1 complex before a refinement was done for each Cnx1 full-length protein (composed of Cnx1E, modeled unresolved parts, and Cnx1G) of the complex using FiberDock (Mashiach et al., 2010).

ORCID

Ahmed H. Hassan 🆔 https://orcid.org/0000-0001-8463-8877

Tobias Kruse 🆔 https://orcid.org/0000-0002-2716-3227

REFERENCES

Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res. 2016;44(W1):W344–50.

Belaidi AA, Schwarz G. Metal insertion into the molybdenum cofactor: product-substrate channelling demonstrates the functional origin of domain fusion in gephyrin. Biochem J. 2013; 450(1):149–57.

Bonvin A, Karaca E, Kastritis PL, Rodrigues J. Defining distance restraints in HADDOCK. Nat Protoc. 2018;13(7):1503–5.

de Vries SJ, Bonvin AM. CPORT: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK. PLoS One. 2011;6(3):e17695.

Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc. 2003;125(7):1731–7.

Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr. 2010;66 (Pt 4):486–501.

Fischer K, Llamas A, Tejada-Jimenez M, Schrader N, Kuper J, Ataya FS, et al. Function and structure of the molybdenum cofactor carrier protein from *Chlamydomonas reinhardtii*. J Biol Chem. 2006;281(40):30186–94.

Heck IS, Schrag JD, Sloan J, Millar LJ, Kanan G, Kinghorn JR, et al. Mutational analysis of the gephyrin-related molybdenum cofactor biosynthetic gene cnxE from the lower eukaryote *Aspergillus nidulans*. Genetics. 2002;161(2):623–32.

Hercher TW, Krausze J, Hoffmeister S, Zwerschke D, Lindel T, Blankenfeldt W, et al. Insights into the Cnx1E catalyzed MPT-AMP hydrolysis. Biosci Rep. 2020;40(1):BSR20191806.

Hercher TW, Krausze J, Yang J, Kirk ML, Kruse T. Identification and characterisation of the *Volvox carteri* Moco carrier protein. Biosci Rep. 2020;40(11):BSR20202351.

Hover BM, Loksztejn A, Ribeiro AA, Yokoyama K. Identification of a cyclic nucleotide as a cryptic intermediate in molybdenum cofactor biosynthesis. J Am Chem Soc. 2013;135(18):7019–32.

Hover BM, Tonthat NK, Schumacher MA, Yokoyama K. Mechanism of pyranopterin ring formation in molybdenum cofactor biosynthesis. Proc Natl Acad Sci U S A. 2015;112(20):6347–52.

Iacobucci C, Gotze M, Ihling CH, Piotrowski C, Arlt C, Schafer M, et al. A cross-linking/mass spectrometry workflow based on MS-cleavable cross-linkers and the MeroX software for studying protein structures and protein–protein interactions. Nat Protoc. 2018;13(12):2864–89.

Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 2004;32:W526–31.

Kneussel M, Betz H. Clustering of inhibitory neurotransmitter receptors at developing postsynaptic sites: the membrane activation model. Trends Neurosci. 2000;23(9):429–35.

Koukos PI, Roel-Touris J, Ambrosetti F, Geng C, Schaarschmidt J, Trellet ME, et al. An overview of data-driven HADDOCK strategies in CAPRI rounds 38-45. Proteins. 2020;88(8):1029–36.

Krausze J, Hercher TW, Archna A, Kruse T. The structure of the Moco carrier protein from *Rippkaea orientalis*. Acta Crystallogr F Struct Biol Commun. 2020;76(Pt 9):453–63.

Krausze J, Hercher TW, Zwerschke D, Kirk ML, Blankenfeldt W, Mendel RR, et al. The functional principle of eukaryotic molybdenum insertases. Biochem J. 2018;475(10):1739–53.

Krausze J, Probst C, Curth U, Reichelt J, Saha S, Schafflick D, et al. Dimerization of the plant molybdenum insertase Cnx1E is required for synthesis of the molybdenum cofactor. Biochem J. 2017;474(1):163–78.

Krissinel E, Henrick K. Inference of macromolecular assemblies from crystalline state. J Mol Biol. 2007;372(3):774–97.

Kruse T. Eukaryotic molybdenum insertases. In: Scott RA, editor. Encyclopedia of inorganic and bioinorganic chemistry; 2020. Wiley.

Kruse T. Function of molybdenum insertases. Molecules. 2022a; 27(17):5372.

Kruse T. Moco carrier and binding proteins. Molecules. 2022b; 27(19):6571.

Kruse T, Gehl C, Geisler M, Lehrke M, Ringel P, Hallier S, et al. Identification and biochemical characterization of molybdenum cofactor-binding proteins from *Arabidopsis thaliana*. J Biol Chem. 2010;285(9):6623–35.

Kuper J, Llamas A, Hecht HJ, Mendel RR, Schwarz G. Structure of the molybdopterin-bound Cnx1G domain links molybdenum and copper metabolism. Nature. 2004;430(7001):803–6.

Leimkühler S, Wuebbens MM, Rajagopalan KV. The history of the discovery of the molybdenum cofactor and novel aspects of its biosynthesis in bacteria. Coord Chem Rev. 2011;255(9-10): 1129–44.

Lilkova E, Petkov P, Ilieva A, Litov L. The PyMOL Molecular Graphics System, Version 2.0, Schrödinger, LLC. 2015

Liu MT, Wuebbens MM, Rajagopalan KV, Schindelin H. Crystal structure of the gephyrin-related molybdenum cofactor biosynthesis protein MogA from *Escherichia coli*. J Biol Chem. 2000; 275(3):1814–22.

Llamas A, Mendel RR, Schwarz G. Synthesis of adenylated molybdopterin: an essential step for molybdenum insertion. J Biol Chem. 2004;279(53):55241–6.

Llamas A, Otte T, Multhaup G, Mendel RR, Schwarz G. The mechanism of nucleotide-assisted molybdenum insertion into

molybdopterin. A novel route toward metal cofactor assembly. J Biol Chem. 2006;281(27):18343–50.

Llamas A, Tejada-Jimenez M, Gonzalez-Ballester D, Higuera JJ, Schwarz G, Galvan A, et al. *Chlamydomonas reinhardtii* CNX1E reconstitutes molybdenum cofactor biosynthesis in *Escherichia coli* mutants. Eukaryot Cell. 2007;6(6):1063–7.

Mashiach E, Nussinov R, Wolfson HJ. FiberDock: flexible induced-fit backbone refinement in molecular docking. Proteins. 2010; 78(6):1503–19.

Mendel RR. The molybdenum cofactor. J Biol Chem. 2013;288(19): 13165–72.

Mendel RR, Kruse T. Cell biology of molybdenum in plants and humans. Biochim Biophys Acta. 2012;1823(9):1568–79.

Palmer T, Santini C-L, Iobbi-Nivol C, Eaves DJ, Boxer DH, Giordano G. Involvement of the *narJ* and *mob* gene products in distinct steps in the biosynthesis of the molybdoenzyme nitrate reductase in *Escherichia coli*. Mol Microbiol. 1996;20: 875–84.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. J Comput Chem. 2004; 25(13):1605–12.

Probst C, Ringel P, Boysen V, Wirsing L, Alexander MM, Mendel RR, et al. Genetic characterization of the *Neurospora crassa* molybdenum cofactor biosynthesis. Fungal Genet Biol. 2014;66:69–78.

Probst C, Yang J, Krausze J, Hercher TW, Richers CP, Spatzal T, et al. Mechanism of molybdate insertion into pterin-based molybdenum cofactors. Nat Chem. 2021;13(8):758–65.

Ringel P, Krausze J, van den Heuvel J, Curth U, Pierik AJ, Herzog S, et al. Biochemical characterization of molybdenum cofactor-free nitrate reductase from *Neurospora crassa*. J Biol Chem. 2013;288(20):14657–71.

Sander B, Tria G, Shkumatov AV, Kim EY, Grossmann JG, Tessmer I, et al. Structural characterization of gephyrin by AFM and SAXS reveals a mixture of compact and extended states. Acta Crystallogr D Biol Crystallogr. 2013;69(Pt 10): 2050–60.

Schwarz G, Schrader N, Mendel RR, Hecht HJ, Schindelin H. Crystal structures of human gephyrin and plant Cnx1 G domains: comparative analysis and functional implications. J Mol Biol. 2001;312(2):405–18.

Smolinsky B, Eichler SA, Buchmeier S, Meier JC, Schwarz G. Splice-specific functions of gephyrin in molybdenum cofactor biosynthesis. J Biol Chem. 2008;283(25):17370–9.

Stallmeyer B, Nerlich A, Schiemann J, Brinkmann H, Mendel RR. Molybdenum co-factor biosynthesis: the *Arabidopsis thaliana* cDNA cnx1 encodes a multifunctional two-domain protein homologous to a mammalian neuroprotein, the insect protein Cinnamon and three *Escherichia coli* proteins. Plant J. 1995;8(5):751–62.

Stiefel EI. The biogeochemistry of molybdenum and tungsten. Met Ions Biol Syst. 2002;39:1–29.

Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, UniProt C. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. Bioinformatics. 2015; 31(6):926–32.

Teschner J, Lachmann N, Schulze J, Geisler M, Selbach K, Santamaria-Araujo J, et al. A novel role for Arabidopsis mitochondrial ABC transporter ATM3 in molybdenum cofactor biosynthesis. Plant Cell. 2010;22(2):468–80.

van Zundert GCP, Rodrigues J, Trellet M, Schmitz C, Kastritis PL, Karaca E, et al. The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. J Mol Biol. 2016;428(4):720–5.

Wajmann S, Hercher TW, Buchmeier S, Hansch R, Mendel RR, Kruse T. The first step of *Neurospora crassa* molybdenum cofactor biosynthesis: regulatory aspects under N-derepressing and nitrate-inducing conditions. Microorganisms. 2020;8(4).

Wuebbens MM, Rajagopalan KV. Structural characterization of a molybdopterin precursor. J Biol Chem. 1993;268(18):13493–8.

Wuebbens MM, Rajagopalan KV. Mechanistic and mutational studies of *Escherichia coli* molybdopterin synthase clarify the final step of molybdopterin biosynthesis. J Biol Chem. 2003;278(16): 14523–32.

Xiang S, Nichols J, Rajagopalan KV, Schindelin H. The crystal structure of *Escherichia coli* MoeA and its relationship to the multifunctional protein gephyrin. Structure. 2001;9(4):299–310.

Zhang Y, Gladyshev VN. General trends in trace element utilization revealed by comparative genomic analyses of Co, Cu, Mo, Ni, and Se. J Biol Chem. 2010;285(5):3393–405.

Zhang Y, Rump S, Gladyshev VN. Comparative genomics and evolution of molybdenum utilization. Coord Chem Rev. 2011; 255(9–10):1206–17.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.