

Crystal structure of chalcone synthase, a key enzyme for isoflavonoid biosynthesis in soybean

Riki Imaizumi¹, Ryo Mameda², Kohei Takeshita³, Toshiyuki Waki², Hiroki Kubo¹, Shun Nakata¹, Seiji Takahashi², Kunishige Kataoka¹, Masaki Yamamoto³, Satoshi Yamashita¹, and Toru Nakayama²

¹Kanazawa University

²Tohoku University

³SPring-8

April 28, 2020

Abstract

Isoflavonoid is one of the groups of flavonoids that play pivotal roles in the survival of land plants. Chalcone synthase (CHS), the first enzyme of the isoflavonoid biosynthetic pathway, catalyzes the formation of a common isoflavonoid precursor. We have previously reported that an isozyme of soybean CHS (termed GmCHS1) is a key component of the isoflavonoid metabolon, a protein complex to enhance efficiency of isoflavonoid production. Here, we determined the crystal structure of GmCHS1 as a first step of understanding the metabolon structure, as well as to better understand the catalytic mechanism of GmCHS1.

INTRODUCTION

Flavonoids are a class of specialized plant metabolites that play pivotal roles in the survival of land plants in terrestrial environments. Currently, more than 6900 flavonoids with different structures have been identified and are known to have indispensable roles in land plants, not only owing to their antioxidant or UV-screening functions but also for their reproductions or defense mechanisms against various stresses.¹ Isoflavonoids are one of the flavonoid groups produced by almost all legumes such as *Glycine max* (L.) Merr (soybean). Chalcone synthase (CHS), the first committed enzyme of the flavonoid biosynthetic pathway, catalyzes the consecutive condensations of three malonyl-CoA molecules into one *p*-coumaroyl-CoA molecule to produce 2',4,4',6'-tetrahydrochalcone, which serves as a common flavonoid precursor (Fig. S1). We have previously reported that an isozyme of soybean CHS (termed GmCHS1) is a component of the soybean isoflavonoid metabolon and interacts with other components such as 2-hydroxyisoflavanone synthase isozymes, providing mechanistic evidence regarding the importance of the protein-protein interactions for the efficient production of 5-deoxyisoflavonoids in the soybean.^{2,3} More recently, we reported the interactions of CHS, including GmCHS1, with its nonenzymatic interacting partner, chalcone isomerase-like protein, in a wide variety of land plants.⁴ Interactions of this protein with CHS are widely conserved among land plants and reduce the intrinsic catalytic promiscuity of CHS to enhance the efficiency of flavonoid production *in planta*.⁴ Although these observations indicate the functional significance of protein-protein interactions among CHS and other enzymes and proteins in the metabolons, there is no atomic-level structural information for protein-protein interactions that mediate flavonoid metabolon formation. Thus, in the first step of addressing this issue, we determined the crystal structure of recombinantly expressed GmCHS1 as one of the key components of a well-characterized flavonoid metabolon. We also determined the complex structure of GmCHS1 with bound naringenin, a product isomer of the CHS-catalyzed reaction, to better understand the catalytic mechanism of GmCHS1.

MATERIALS AND METHODS

Purification of recombinantly expressed GmCHS1

Heterologous expression and purification of GmCHS1 were reported previously.² Briefly, the *GmCHS1* ORF was introduced into pCold I vector to express the GmCHS1 with an *N*-terminal extension including His₆-tag. The *N*-terminal extension sequence includes MNHKVHHHHHHIEGRHM, where the underlined Met indicates the original initiating Met of GmCHS1. The resultant plasmid was introduced into *Escherichia coli* BL21 strain. The transformed *E. coli* was cultured in LB medium containing 100 µg/mL ampicillin and the expression of GmCHS1 was induced by cold-shock induction. The harvested cells were disrupted and the lysate was centrifuged at 100,000 × *g* for 1hr at 4°C. The supernatant from the centrifugation was applied to a column filled with Ni-NTA agarose (5 mL; Qiagen, Hilden, Germany) equilibrated with 50 mM Tris-HCl (pH 7.5) (buffer A) containing 150 mM NaCl. Then, proteins were eluted with buffer A containing 200 mM imidazole. The eluted GmCHS1 was then applied to a size-exclusion column (Superdex 200 increase 10/300, GE Healthcare, Chicago, IL, USA).

Crystallization, data collection and structure determination

GmCHS1 crystals were grown at 20°C by vapor diffusion, mixing 1 µL protein solution (6.7 mg/mL) with 1 µL well solution containing 30% PEG4000, 0.2M ammonium acetate and 0.1M sodium citrate (pH 5.6). The typical crystals grew to needle-like structures with 0.5 mm length in average within 1 week. GmCHS1-naringenin complex was prepared by incubating an apo-GmCHS1 crystal with naringenin (Nacalai tesque, Kyoto, Japan) with 2 mM final concentration for 1 hr at 20°C prior to be flash-frozen in liquid N₂. X-ray diffraction data were collected at wavelengths of 1.00000 Å using an EIGER 9M detector (DECTRIS Ltd., Baden-Daettwil, Switzerland) on BL32XU beamline installed ZOO system, SPring-8, Harima, Japan.⁵ These completed reflection data were collected using a helical data collection scheme, and were indexed and integrated with the KAMO pipeline, using DIALS.^{6,7} Integration intensities and merging were performed using XSCALE with outlier rejections implemented in KAMO.⁸ Structures were solved by molecular replacement using MOLREP with *Pinus sylvestris* CHS (PDB ID: 6DXA) as a search model.⁹ GmCHS1 crystallized with two molecules in the asymmetric unit. Manual model building and refinement of both structures were performed using Coot and REFMAC.^{10,11} 3D molecular images were created using the PyMOL program.¹² Analysis of interior cavities and surface pockets were performed by the CASTp 3.0 server program and the result was visualized by the PyMOL program.¹³ The atomic coordinates and structural factors have been deposited in the Protein Data Bank with PDB codes 7BUS (apo) and 7BUR (complex).

RESULTS AND DISCUSSION

Recombinantly expressed GmCHS1 was purified using nickel affinity chromatography with an *N*-terminal His₆-tag, and by using size-exclusion chromatography (Fig. S1). The purified GmCHS1 was crystallized, and the crystal diffracted to 2.5 Å resolution (crystallographic statistics, Table 1; electron density quality, see Fig. S2). GmCHS1 crystallized in space group P2₁2₁2₁ with two molecules per asymmetric unit. The structures were solved using a molecular replacement method, with *Pinus sylvestris* CHS (PDB ID: 6DXA, 85.9% sequence identity) as a search model. The refined structure included all residues from the original sequence of GmCHS1 (residues 1-388). However, the electron densities corresponding to *N*-terminal extension residues including the His₆-tag were not visible owing to the highly disordered structures of the *N*-terminal regions. The overall structure of GmCHS1 exhibits the αββα thiolase fold with a pseudo-symmetric dimer with 12 α-helices and 11 β-strands (Fig. 1A) per monomer, as previously reported for the crystal structure of CHS from *Medicago sativa* (alfalfa, PDB ID: 1BI5).¹⁴ In fact, our size-exclusion chromatography experiment suggests that GmCHS1 mainly exists as a homodimer in solution (Fig. S1). The homodimeric GmCHS1 contains two active sites in the middle of the thiolase fold (Figure 1A, one is highlighted in a dotted circle). When viewed in surface representation, there is a tunnel-like structure, which has been termed the CoA-binding tunnel in previous studies, making a connection between the external solvent environment and the active site in each monomer (Fig. 1B). Further analysis of the interior cavity, including the Co-A binding tunnel, indicates that there is a large cavity around the active site wherein the condensation reactions should occur

(Fig. 1C). It is noteworthy that the large cavity forms another tunnel downward to the external solvent (Fig. 1C, indicated by a dotted arrow). Although the mechanism for the product export of CHS is not completely understood, it will be of interest in the future to evaluate whether this additional tunnel functions as a metabolic channel. A close-up view of the active site showed that the catalytic residues (Cys164, His302 and Asn335) are located in similar orientations to those of other CHS structures, with the electron densities for Cys164 sidechains being in their doubly oxidized forms (Fig. 1D). The doubly oxidized catalytic cysteine sulfinic acid is widely observed in the crystal structures of CHS from euphylllophytes (ferns and seed plants), but not in the CHS from either lycophyte or moss species. Because of the clarity of electron densities at the positions 164 and the fact that soybean belongs to euphylllophytes, we modeled S-cysteinesulfinic acid (ligand ID: CSD) molecules for the corresponding positions. Weng *et al.* hypothesized that the oxidation of catalytic Cys residues was due to the highly reactive environment around Cys during the generation of nucleophilic thiolate anion that was assisted by the conserved His residue (His302 in GmCHS1).¹⁵ In the GmCHS1 structure, distances between the sulfur atom of CSD164 and the N ϵ of His302 are 3.55 and 3.69 Å for molecule A and B in the asymmetric unit, respectively (Fig. 1D). In addition to the oxidation of Cys164, the electron densities for Met70 sidechains are observed in their oxidized forms. Although these Met oxidations are not reported for any other CHS structures, we modeled the methionine sulfoxide (ligand ID: SME) for the positions 70. Further biochemical or physiological investigations will be needed to decipher the importance of the Met oxidation.

To better understand the catalytic mechanism of GmCHS1, newly purified GmCHS1 was crystallized, and the apo-crystals were soaked with naringenin, a product analog of the CHS-catalyzed reaction. The naringenin-soaked crystals diffracted to 1.8 Å resolution, and the structure was determined in a similar procedure to that of the apo-structure. The refined complex structure includes a dimer of GmCHS1, two naringenin molecules, as well as three of acetate ions and three citrate ions, which are contained in the mother liquor. Moreover, SME and CSD are modeled for the positions 70 and 164, respectively as well as the apo-structure. In the complex structure, naringenin molecules are located in the vicinity of the catalytic cysteine sulfinic acids (the residue 164, Fig. 2A). The positions of naringenin molecules were found to be almost equivalent to those of previously reported for the crystal structure of *M. sativa* CHS (Fig. S3).¹⁴ In addition to the CSD164, Phe215 and Phe264 are located in the vicinity of naringenin. The importance of the conserved Phe215 was previously proposed as an involvement in the orientation of substrates and reaction intermediates at the active site.¹⁴ We found that one of the citrate ions was located at the end of the CoA binding tunnel, also in the vicinity of Phe264 (Fig. 2A, 2B). The corresponding position of the citrate ion of our structure was occupied by the CoA moiety of malonyl-CoA or 1,4-piperazine di-ethane sulfonic acid (C₈H₁₈N₂O₆S₂) in the complex structure of the *M. sativa* CHS Cys164Ala mutant (Fig. S3, PDB ID: 1CML), suggesting that a negative charge of the citrate ion could potentially mimic that of either CoA or 1,4-piperazine di-ethane sulfonic acid. Although citric acid has not been studied as a potential inhibitor of CHS, our complex structure suggests the possibility of inhibitor binding upon the existence of the product at the active site. Superposition of apo- and naringenin-bound structures of GmCHS1 shows that the Phe265 is likely to change its orientation through the CHS-catalyzed reaction (Fig. 2C). This is supported by the fact that the average temperature factor for the Phe264 sidechain is higher than that of Phe215 (44.2 Å² and 33.4 Å² for the Phe264 and the Phe215, respectively). It is an intriguing question whether the flexibility or the bulkiness of position 264 is related to the catalytic efficiency of GmCHS1. In conclusion, the crystal structures of GmCHS1 provide a structural basis for further improvement of our understanding of the catalytic mechanism of CHS. Further efforts will include structural characterization of a complex between GmCHS1 and other components of the soybean isoflavonoid metabolon to better understand how GmCHS1 has a distinct role in the formation of the isoflavonoid metabolon. Coordinates and structural factors of apo-GmCHS1 and GmCHS1/naringenin have been deposited in the Protein Data Bank under accession codes 7BUS and 7BUR, respectively.

References

1. Mouradov A, Spangenberg G. Flavonoids: a metabolic network mediating plants adaptation to their real estate. *Front Plant Sci.* 2014;5:620.
2. Waki T, Yoo D, Fujino N, Mameda R, Denessiouk K, Yamashita S, Motohashi R, Akashi T, Aoki T,

- Ayabe S, Takahashi S, Nakayama T. Identification of protein-protein interactions of isoflavonoid biosynthetic enzymes with 2-hydroxyisoflavanone synthase in soybean (*Glycine max* (L.) Merr.). *Biochem Biophys Res Commun.* 2016;469:546-51.
3. Mameda R, Waki T, Kawai Y, Takahashi S, Nakayama T. Involvement of chalcone reductase in the soybean isoflavone metabolon: identification of GmCHR5, which interacts with 2-hydroxyisoflavanone synthase. *Plant J.* 2018;96:86-74.
 4. Waki T, Mameda R, Nakano T, Yamada S, Terashita M, Ito K, Tenma N, Li Y, Fujino N, Uno K, Yamashita S, Aoki Y, Denessiouk K, Kawai Y, Sugawara S, Saito K, Yonekura-Sakakibara K, Morita Y, Hoshino A, Takahashi S, Nakayama T. A conserved strategy of chalcone isomerase-like protein to rectify promiscuous chalcone synthase specificity. *Nat Commun.* 2020;11:870.
 5. Hirata K, Yamashita K, Ueno G, Kawano Y, Hasegawa K, Kumasaka T, Yamamoto M. ZOO: an automatic data-collection system for high-throughput structure analysis in protein microcrystallography. *Acta Crystallogr D Struct Biol.* 2019;75:138-150.
 6. Yamashita K, Hirata K, Yamamoto M. KAMO: towards automated data processing for microcrystals. *Acta Crystallogr D Struct Biol.* 2018;74:441-449.
 7. Winter G, Waterman DG, Parkhurst JM, Brewster AS, Gildea RJ, Gerstel M, Fuentes-Montero L, Vollmar M, Michels-Clark T, Young ID, Sauter NK, Evans G. DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr D Struct Biol.* 2018;74:85-97.
 8. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Cryst D. Biol Crystallogr.* 2010;66:125-132.
 9. Vagin A, Teplyakov A. Molecular replacement with MOLREP. *Acta Crystallogr D Biol Crystallogr.* 2010;66:22-25.
 10. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr.* 2010;66:486-501.
 11. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr.* 2011;67:355-67.
 12. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
 13. Tian W, Chen C, Lei X, Zhao J, Liang J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* 2018;46:363-367.
1. Ferrer JL, Jez JM, Bowman ME, Dixon RA, Noel JP. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nat Struct Biol.* 1999;6:775-84.
 2. Liou G, Chiang YC, Wang Y, Weng JK. Mechanistic basis for the evolution of chalcone synthase catalytic cysteine reactivity in land plants. *J Biol Chem.* 2018;293:18601-18612.

Figure legend

Figure 1. The crystal structure of apo-GmCHS1. A: The overall structure of apo-GmCHS1 homodimer, with each monomer is colored in either cyan or green. Dotted circle indicates the active site of the monomer, and the catalytic residues are represented in stick models as will be shown in D. B: Surface representations of apo-GmCHS1 homodimer. Left: The view from same direction with panel A. Right: The view of 30° rotated around the y-axis. The arrow and dotted circle (orange) indicates the entrance of the putative CoA-binding tunnel.¹⁴ C: An interior cavity of apo-GmCHS1, which is predicted by the CASTp program, is shown in the surface representation. Black arrow indicates the entrance of the CoA-binding tunnel as described above, whereas dotted arrow (red) indicates another tunnel connecting to the active site. Dotted circle highlights the active site, as is shown in panel A. D: A close-up view of the active site of GmCHS1. The $2Fo - Fc$ electron density maps contoured at 1σ are represented for the catalytic cysteine sulfenic acid (CSD164), His302 and Asn335. Distance (Å) between the sulfur atom of CSD164 and the N_ϵ of His302 is labeled and shown as a dotted line.

Figure 2. The crystal structure of GmCHS1 complexed with naringenin. A: A close-up view of the active site of GmCHS1 with bound naringenin (NAR) and citrate ion (CIT). The $2Fo - Fc$ electron

density maps contoured at 1σ are shown for naringenin, Phe215 and Phe264. B: Stereo view of the important residues, the naringenin and the citrate ion in the active site of GmCHS1. C: Superposition of the apo- and naringenin-bound structures of GmCHS1 (green: the apo-structure, cyan: the complex structure). The active sites from both structures are shown without indicating the citrate ion in the complex structure. The $2Fo - Fc$ electron density map for naringenin is shown as in panel A except indicating the stick model. Dotted double arrow indicates the possible movement of Phe264. See also Figure S2.

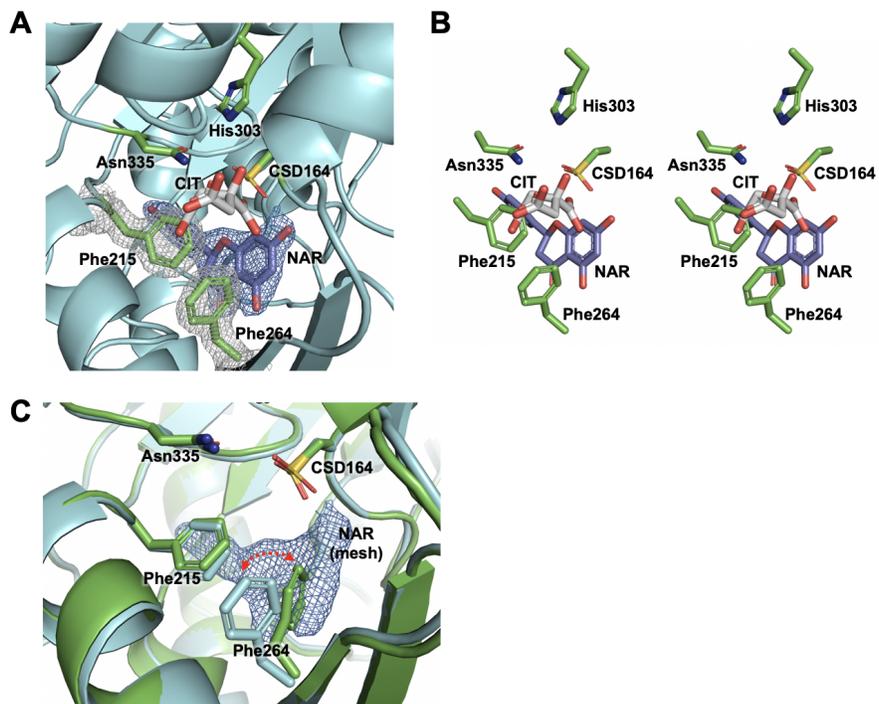
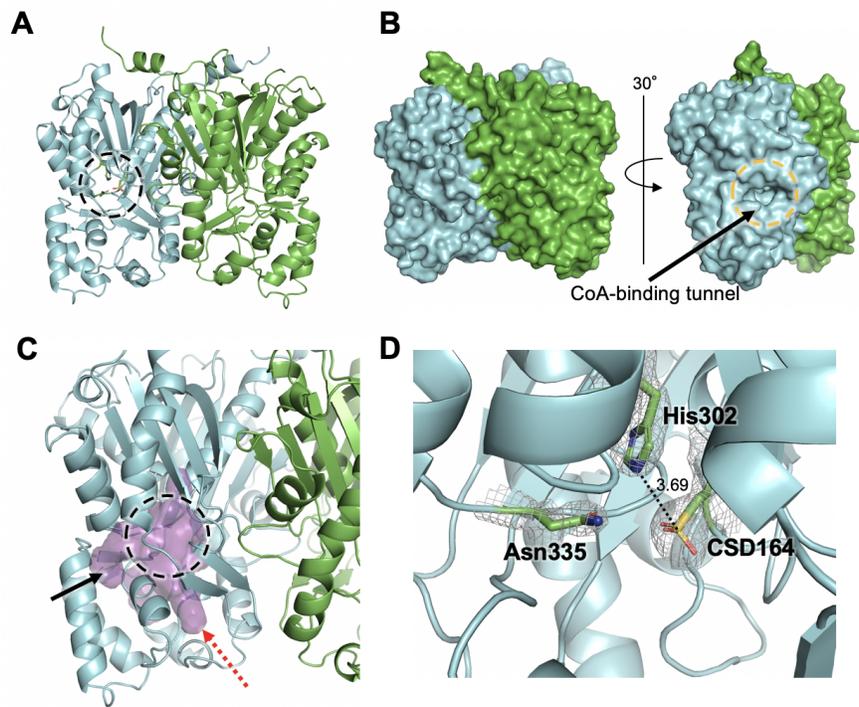
Supplemental figure legend

Figure S1. Structural formulas and protein purification. (a) Structural formulas of 2',4,4',6'-tetrahydroxychalcone (THC) and naringenin. (b) SDS-PAGE of the purified GmCHS1. The calculated molecular mass of the recombinant GmCHS1 including its *N*-terminal tag is approximately 44 kDa. (c) The chromatogram of size-exclusion chromatography of GmCHS1. The arrowhead indicates a major fraction containing GmCHS1. The molecular mass is estimated to be approximately 80 kDa from the elution volume of the major peak, suggesting GmCHS1 exists as a homodimer in solution.

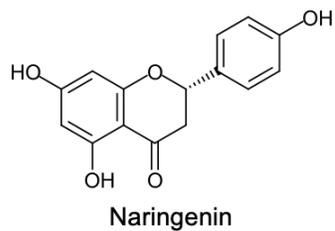
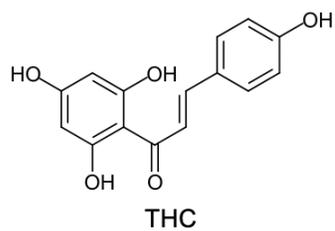
Figure S2. Representative electron densities for apo-GmCHS1 and GmCHS/naringenin complex structures (related to figure 2). Stereo images of the $2Fo - Fc$ electron density maps contoured at 1σ for the β -strands region ($\beta 7$ and $\beta 8$) for (a) apo-GmCHS1 and (b) GmCHS/naringenin complex structures. Phe264 residues are labeled in both (a) and (b). The bound naringenin is shown behind the protein structure with a white stick model in (b).

Figure S3. Structure-based alignment of GmCHS1 and CHS from *Medicago sativa* (alfalfa).

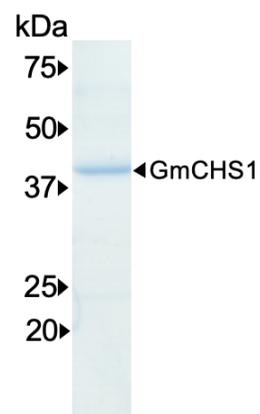
(a) Superposition of GmCHS1/naringenin complex (cyan) and *M. sativa* CHS/naringenin complex (wheat, PDB ID: 1CGK), without indicating bound naringenin molecules. Important residues in both active sites are shown in stick models as will be shown in (b). (b) The super position of important residues and naringenin (NAR) molecules. Colors of the residues and naringenin molecules are corresponding to (a). The residues are labeled only for GmCHS1. (c) Surface representation of *M. sativa* CHS/malonyl-CoA complex (PDB ID: 1CML) with showing the bound malonyl-CoA in stick model, although its malonyl moiety is buried in the active site and thus not visible in this panel. Dotted circle indicates the position where the 1,4-piperazine di-ethane sulfonic acid molecule partially occupies in the coordinate. (d) Surface representation of GmCHS1/naringenin complex with showing the bound citrate ion in stick model.



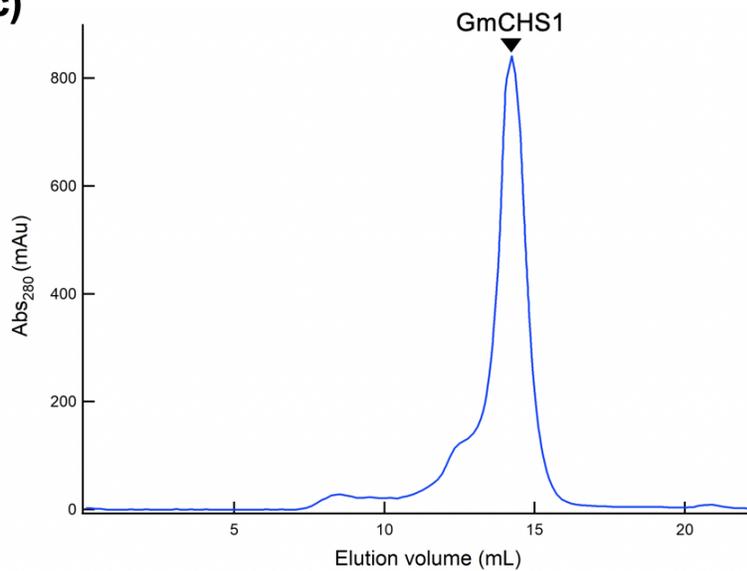
(a)



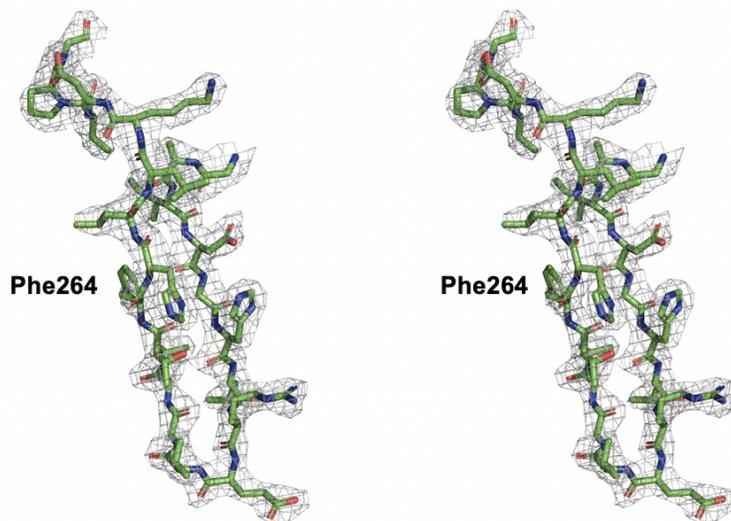
(b)



(c)



(a)



(b)

