Fusion Protein Modeling Supplementary Data

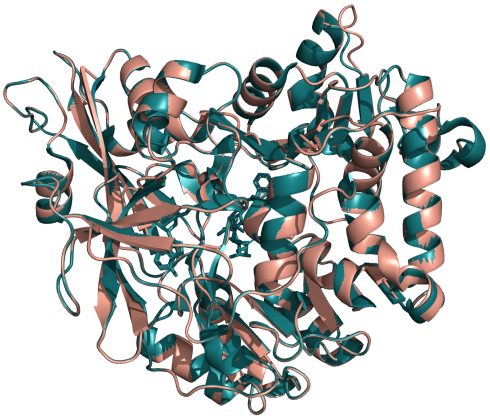

Fig. 1. 4CL2 from *Glycine max* (**pink**) with 4CL2 from *Nicotiana tabacum* (**electric**)

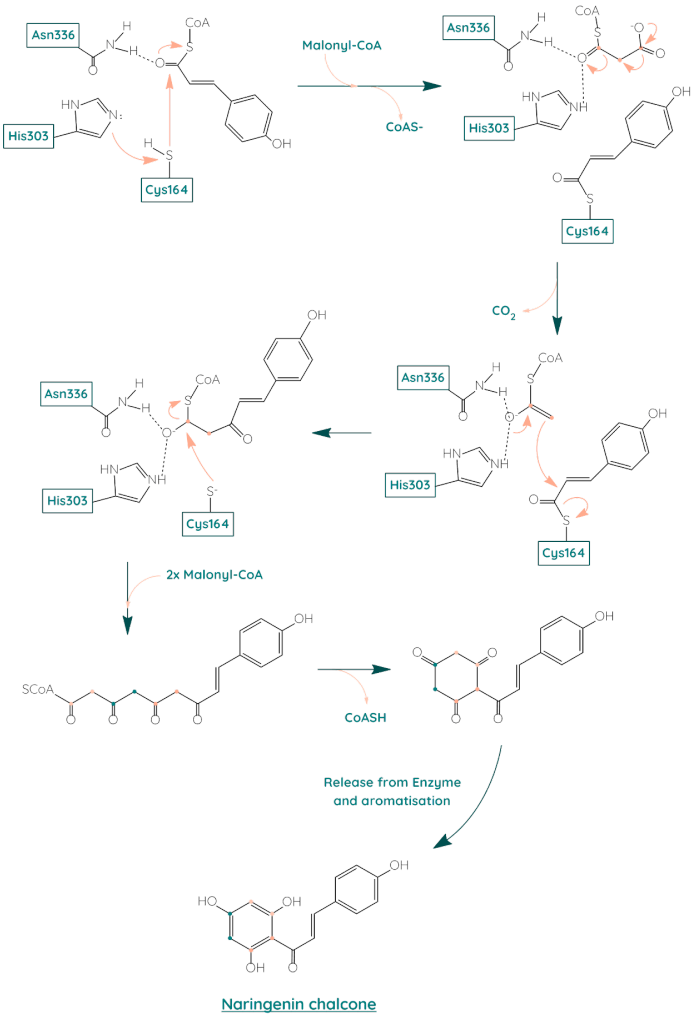

Fig. 2. CHS catalyzed reaction

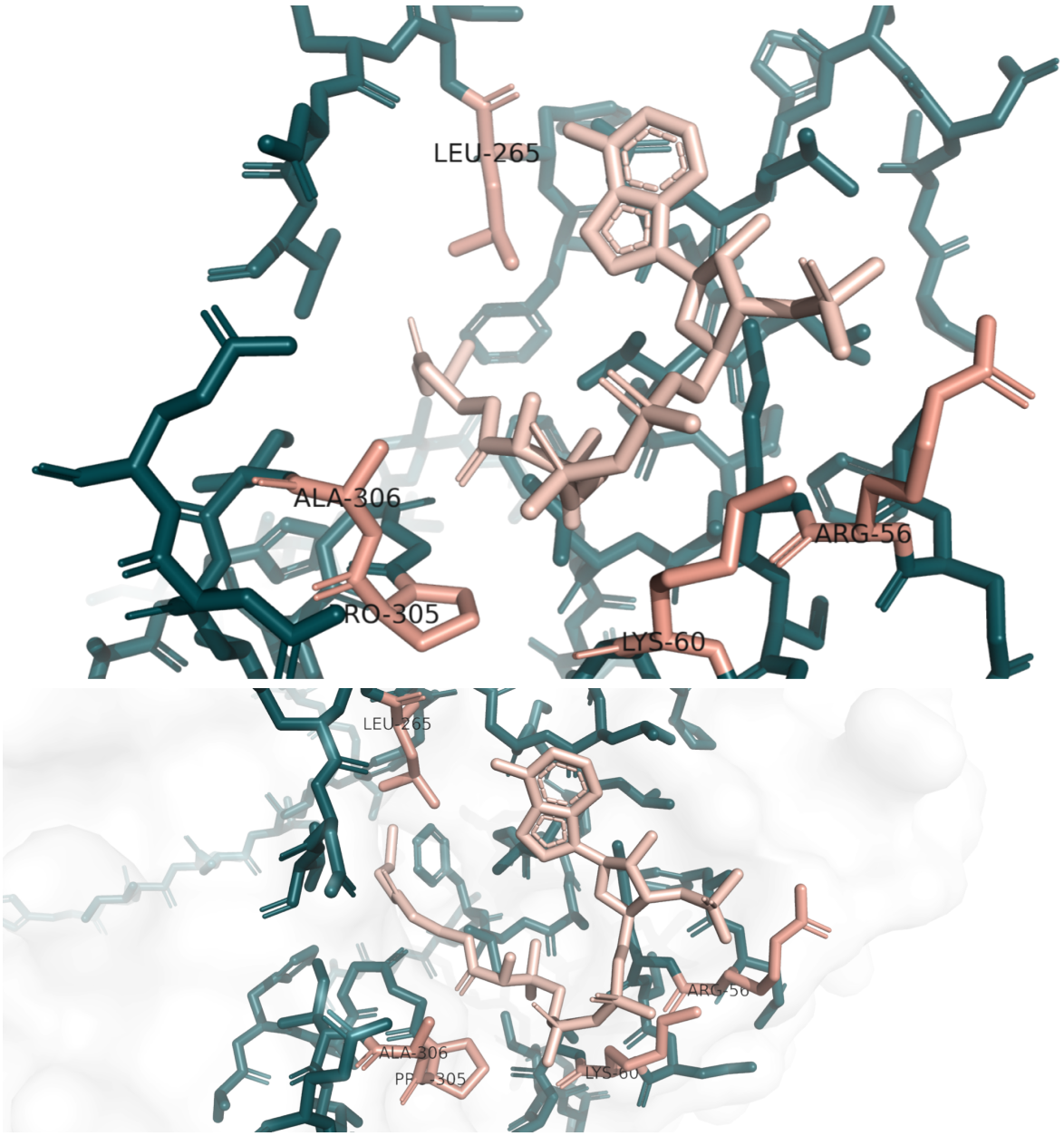

Fig. 3. Active site residues of CHS

Table 1. Characterisation of linkers

Linker	Length	Flexibility	Hydrophobicity	Sensitivity to proteases	Secondary structure
GSG	3	Flexible	Neutral	-	Flexible
GGGGS	5	Flexible	Neutral	-	Flexible
(GGGGS)2	10	Flexible	Neutral	-	Flexible
(GGGGS)3	15	Flexible	Neutral	-	Flexible
EAAAK	5	Rigid	60% hydrophobic 20% acidic 20% basic	Glutamyl endopeptidase LysC LysN Proteinase K Staphylococcal peptidase I Thermolysin Trypsin	α-helical
(EAAAK)2	10	Rigid	60% hydrophobic 20% acidic 20% basic	Asp-N endopeptidase + N-terminal Glu Glutamyl endopeptidase LysC LysN Proteinase K Staphylococcal peptidase I Thermolysin Trypsin	α-helical
(EAAAK)3	15	Rigid	60% hydrophobic 20% acidic 20% basic	Asp-N endopeptidase + N-terminal Glu Glutamyl endopeptidase LysC LysN Proteinase K Staphylococcal peptidase I Thermolysin Trypsin	α-helical

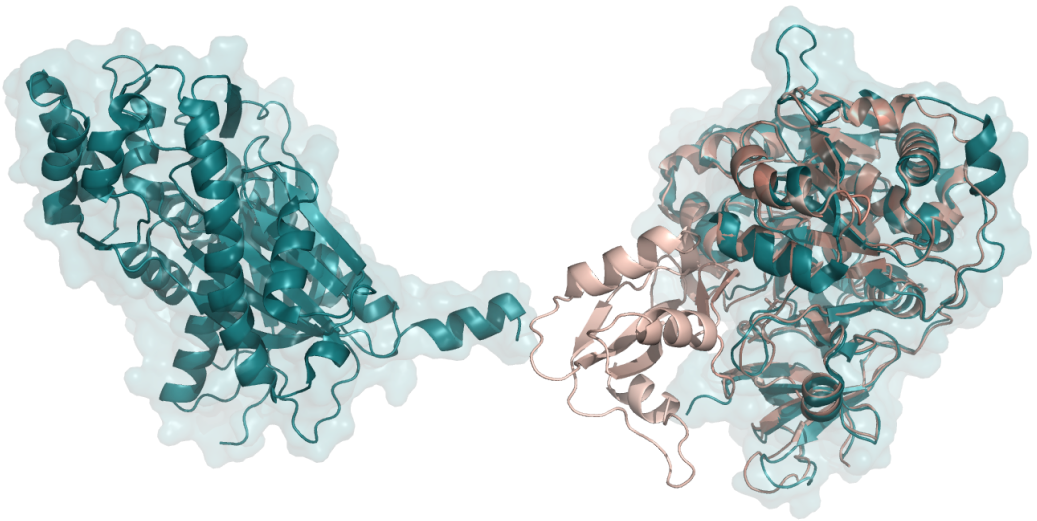

Fig. 4. Modelled 4CL (**pink**) overlay with monomeric enzymes 4CL and STS (PDB 3TSY, **electric**), disordered domain (**light pink**)

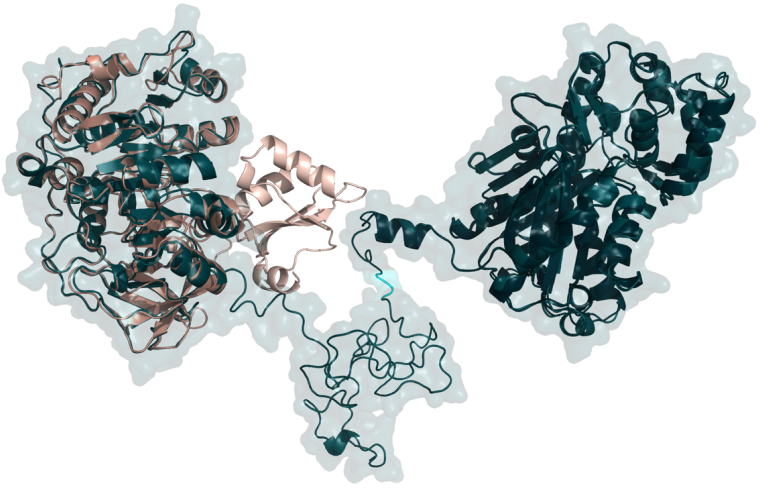

Fig. 5. Modeled 4CL:GSG:CHS (fusion - **electric**, linker - **cyan**) overlay with template structures of enzymes 4CL (**pink**) with disordered domain (**light pink**) and CHS (**dark green**)

Table 2. Evaluating 4CL:GSG:CHS structure that was modeled using homology-based method SWISS-MODEL

SWISS-MODEL					
Model	Template	Sequence identity	QMEAN	QMEANDisCo	VoroMQA
model_1.pdb	3TSY	66.42%	**-2.19**	**0.71**	**0.504224**
model_2.pdb	3TSY	**69.36%**	-2.41	0.71	0.496802

Table 3. Review of the scoring functions to evaluate modeled protein structures

Scoring	Description
VoroMQA[10]	The method is based on **partitioning** space into Voronoi cells. Two adjacent Voronoi cells share a set of points that form a surface called a **Voronoi face**, which can be viewed as a geometric representation of **contact** between two atoms. VoroMQA determines the quality of protein structural models using **inter-atomic** and **solvent** contact areas, including the knowledge-based statistical potential.
ProSA[11]	This program uses knowledge-based potentials of **mean force** to evaluate the **accuracy** of the input model. The structure's energy is evaluated using a **distance-based** pair potential and a potential that captures the solvent **exposure** of protein residues. From these energies, two characteristics are derived: a ***z*-score** and a plot of its **residue energies**. The *z*-score indicates overall model **quality** and measures the structure's total energy **deviation** concerning an energy **distribution** derived from **random conformations**. *Z*-scores outside a range characteristic determined by structures of **native** proteins indicate **errors** and **low quality** of the model that is being evaluated.
TM-Score[12]	Template modeling score is a variation of the Levitt-Gerstein (LG) score, which weights **shorter** distances between corresponding residues more **strongly** than longer distances. This function requires template - experimental structure as a mandatory input.
TM-align[46]	The algorithm employs **TM-Score** and **dynamic programming** to generate the **best** structural alignment between two protein structures. This function requires two input structures.
CAD-Score[13]	The definition of CAD-score is based on the three considerations:

	1. contacts in the model should be evaluated according to the **contacts** in the **reference** structure (target); 2. any **missing** residues in the model should be treated in the same way as if **none** of their contacts were **correctly** predicted; 3. strong over-prediction (nonphysical overlap) of a particular contact should be equivalent to **missing** that contact entirely (in other words, there should be no **false-positive** predictions in terms of contact). This function requires two input structures: model and target.
QMEAN[48]	**QMEAN** is an acronym for **Q**ualitative **M**odel **E**nergy **AN**alysis. There are five different descriptors used for this evaluation approach, which are: 1. **Torsion angle** potential 2. Distance-dependent pairwise **residue** potential 3. **Solvation** potential 4. The term describing **agreement** of predicted and calculated **secondary structure** 5. The term describing **solvent accessibility**
QMEANDisCo[14]	In addition to QMEAN composite scoring, **QMEANDisCo** introduces a new distance constraint (DisCo) that evaluates the **agreement** of pairwise **distances** and **constraints** taken from homologous experimentally determined protein structures.
ProQ3D[15]	**ProQ3D** is a **single-model** quality estimation program. ProQ estimated the quality of the whole model using a machine learning approach - **support-vector machine** - by summing up the predicted qualities for each residue. ProQ2 included **profile weights** to improve the predictions. ProQ3 included **energy terms** calculated from Rosetta. Overall, **ProQ3D** uses the same inputs as in ProQ3. Although, a **neural network** is used to assess the quality of the protein structure instead of a support

	vector machine.

Table 4. Adjustments to the molecular dynamics simulation steps

Before consultation	After consultation	Command
Preparing protein file by **removing** other **chains** and **heterogenous** atoms		`grep`
Converting protein's PDB file to Gromacs format GRO		`pdb2gmx`
Choosing **force-field** CHARMM27	Choosing **force-field** AMBER94	
Choosing **water model** SPC	Choosing **water model** TIP3	
Defining a box with **2 Å** distance between protein each wall of the **dodecahedron** type box	Defining a box with **1.2 Å** distance between protein and each wall of the **dodecahedron** type box	`editconf`
Solvating protein		`solvate`
Adding **ions** to **neutralize** the system, selecting group SOL	Adding **ions** to **neutralize** the system	`grompp`, `genion`
	Adding **ions** to simulate water with dissolved **salt** (concentration 150 mM)	
Energy **minimization** step to reach negative potential energy and maximum force to be less than 1000 kJ/mol/nm		`grompp`, `mdrun`, `energy`
Temperature equilibration for 100 ps, 300 K		`grompp`,`mdrun` `energy`
Pressure equilibration for 100 ps, 300 K		`grompp`,`mdrun` `energy`
Production run for **1 ns**	Production run for **100 ns**	grompp`, `mdrun`

Table 5. Evaluation of 4CL modeled structures

trRosetta			
Model	QMEAN4	QMEANDisCo	VoroMQA score
4CL.sqit.pdb	-0.94	0.67	0.478739

SWISS-MODEL			
Model	QMEAN4	QMEANDisCo	VoroMQA score
4CL_1.pdb	-0.73	0.84	0.541558
4CL_2.pdb	-0.61	0.82	0.519713
4CL_3.pdb	-0.34	0.85	0.549002

RoseTTAFold (without MSA)			
Model	QMEAN4	QMEANDisCo	VoroMQA score
MODEL_1_4CL2_SOYBN.pdb	1.71	0.77	0.540512
MODEL_5_4CL2_SOYBN.pdb	1.44	0.77	0.540305
MODEL_2_4CL2_SOYBN.pdb	1.6	0.76	0.539742
MODEL_3_4CL2_SOYBN.pdb	1.61	0.77	0.533458
MODEL_4_4CL2_SOYBN.pdb	1.79	0.76	0.530254

Table 6. Evaluation of 4CL modeled using RoseTTAFold with MSA structures

RoseTTAFold (with MSA)			
Model	QMEAN4	QMEANDisCo	VoroMQA score
MODEL_1_4CL_MSA.pdb	1.88	0.77	0.541581
MODEL_2_4CL_MSA.pdb	1.12	0.75	0.526281
MODEL_3_4CL_MSA.pdb	1.09	0.76	0.529501
MODEL_4_4CL_MSA.pdb	0.95	0.76	0.534835
MODEL_5_4CL_MSA.pdb	1.41	0.76	0.529382

Table 7. Evaluation of 4CL modeled using AlphaFold2

AlphaFold2				
Model	QMEAN4	QMEANDisCo	VoroMQA score	average plDDT
4CL_unrelaxed_model_1.pdb	-0.62	0.79	0.546974	**89.651979**
4CL_unrelaxed_model_2.pdb	-1.01	0.79	0.549646	89.276523
4CL_unrelaxed_model_3.pdb	-1.02	0.79	0.546767	88.780811
4CL_relaxed_model_1.pdb	0.13	**0.8**	0.551421	-
4CL_relaxed_model_2.pdb	**0.3**	**0.8**	**0.552571**	-
4CL_relaxed_model_3.pdb	-0.86	**0.8**	0.549808	-

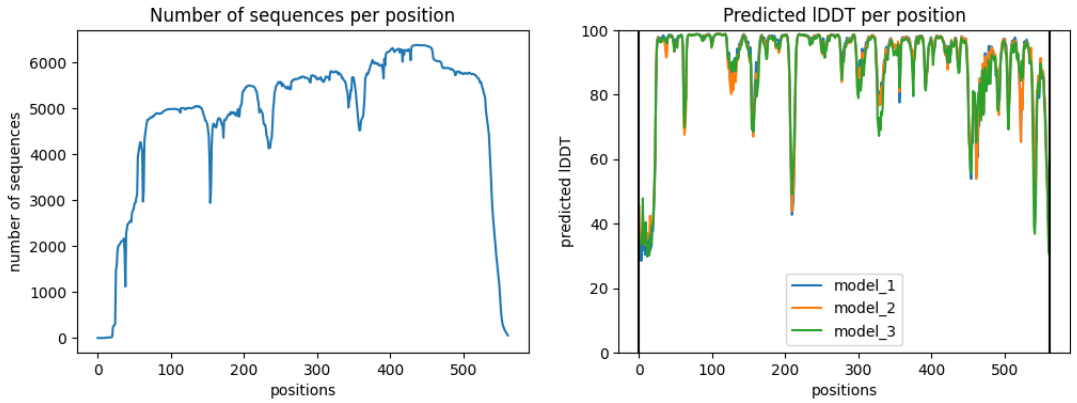

Fig. 6. Coverage and predicted lDDT per position of 4CL structure modeled with AlphaFold2

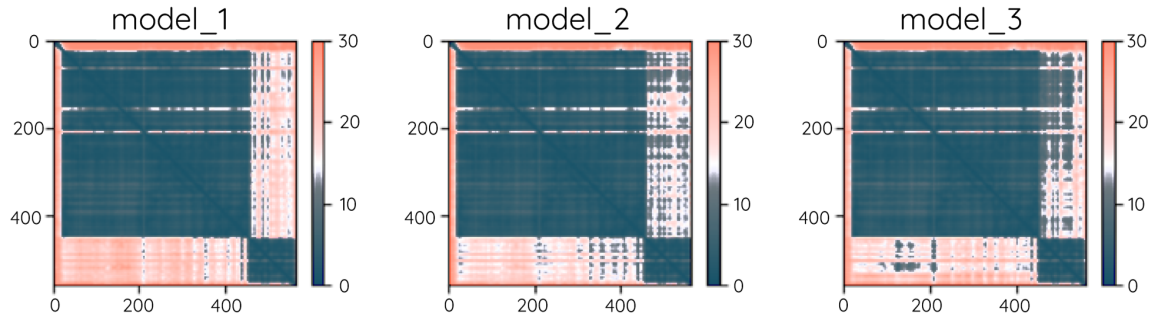

Fig. 7. Predicted alignment error of 4CL structure modeled with AlphaFold2

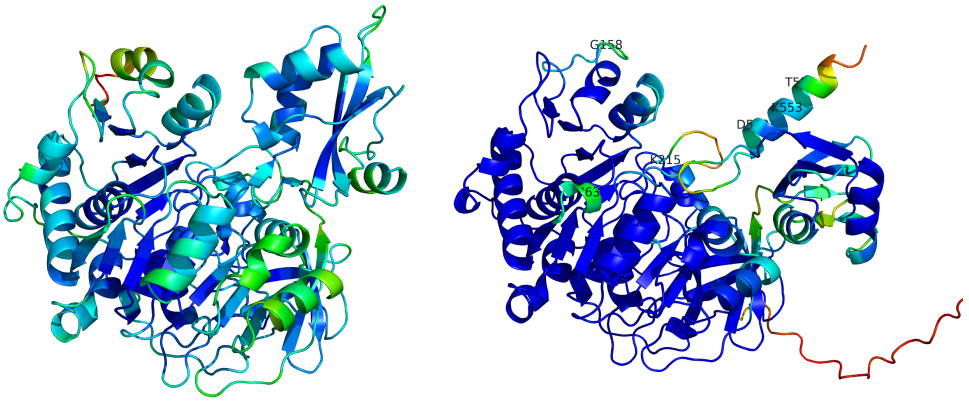

Fig. 8. 4CL structures modeled with SWISS-MODEL (left) and AlphaFold2 (right). The darker shade of blue indicates higher structure determination quality (B-factor spectrum)

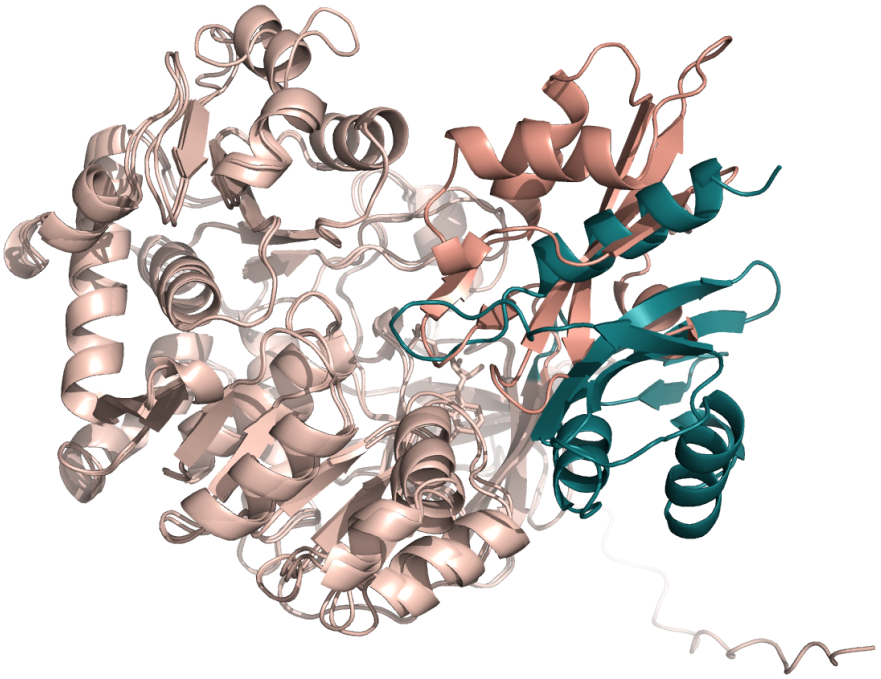

Fig. 9. Disordered region of 4CL structure modeled with AlphaFold2 (**electric**) overlay with SWISS-MODEL model (**pink**)

Table 8. Evaluation of CHS modelled using AlphaFold2

AlphaFold2				
Model	QMEAN4	QMEANDisCo	VoroMQA score	average plDDT
CHS_unrelaxed_model_1.pdb	0.08	0.85	0.535824	**96.86796**
CHS_unrelaxed_model_2.pdb	-0.13	0.85	0.542367	96.62286
CHS_unrelaxed_model_3.pdb	-0.45	0.84	0.536282	96.34417
CHS_relaxed_model_1.pdb	0.2	**0.86**	**0.542643**	-
CHS_relaxed_model_2.pdb	**0.21**	**0.86**	0.539895	-
CHS_relaxed_model_3.pdb	0.2	**0.86**	0.540226	-

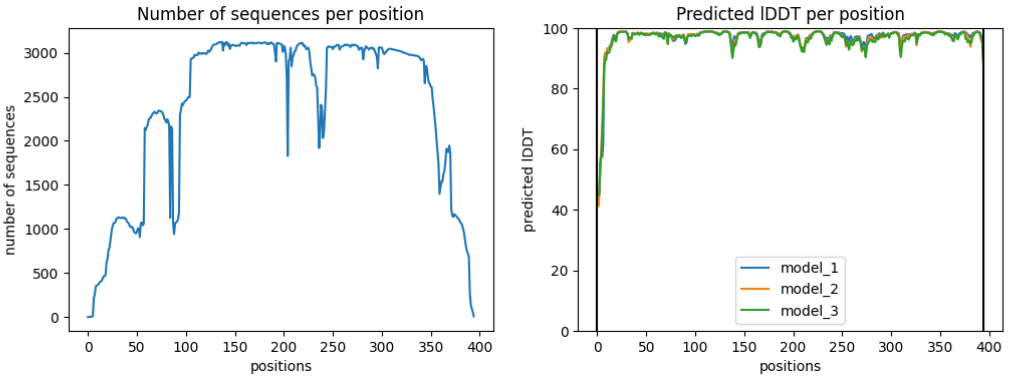

Fig. 10. Coverage and predicted lDDT per position of CHS structure modelled with AlphaFold2

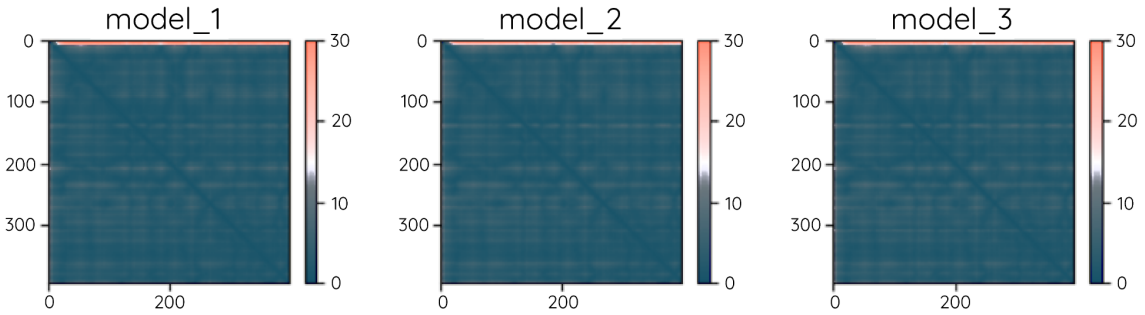

Fig. 11. Predicted alignment error of CHS structure modelled with AlphaFold2

Table 9. Inaccurate residues according to PyMOL B-factor coloring

Experimental structure	Modeled structure
7-8	1-8
67	-
83-84	-
208-209	210-212
271-276	-
295-299	-
327-328	-

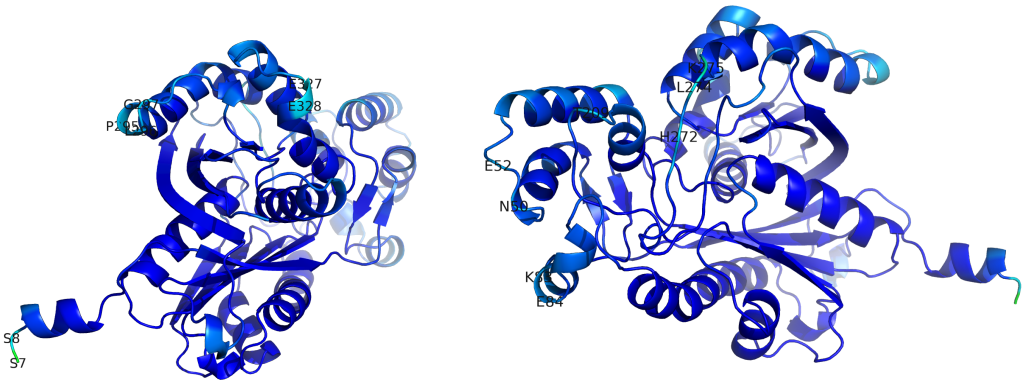

Fig. 12. B-factor coloring for CHS experimental structure front (left) and back (right)

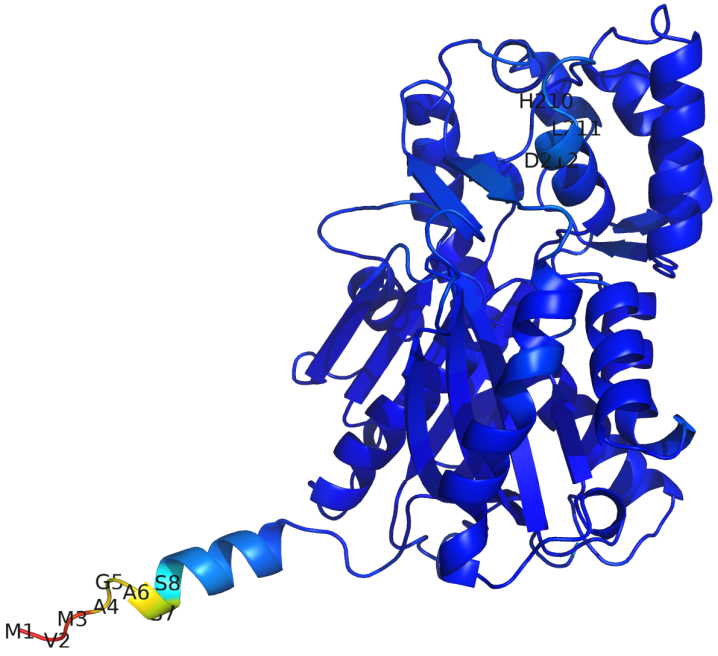

Fig. 13. B-factor coloring for CHS structure modeled with AlphaFold2

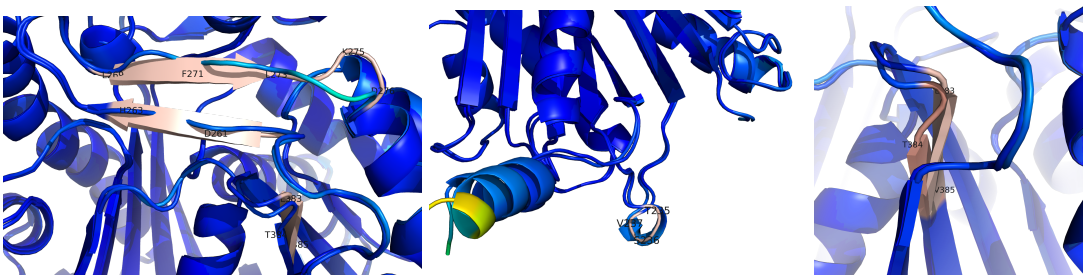

Fig. 14. Structural differences between experimental CHS structure and the best model of AlphaFold2

Initial modeling of linker GSG case

Table 10. VoroMQA evaluation of trRosetta modelled 4CL:GSG:CHS structures

trRosetta		
Model	VoroMQA score	Z-score (ProSA)
model1.pdb	**0.408**	-14.35
model2.pdb	0.390	-12.04
model3.pdb	0.374	-12.17
model4.pdb	0.357	-11.13
model5.pdb	0.378	-11.1

Table 11. VoroMQA evaluation of RoseTTAFold modelled 4CL:GSG:CHS structures

RoseTTAFold		
Model	VoroMQA score	Z-score (ProSA)
model01.pdb	0.512065	-16.07
model02.pdb	0.511473	-16.01
model03.pdb	0.50706	-16.17
model04.pdb	**0.516429**	-16.18
model05.pdb	0.513875	-15.75

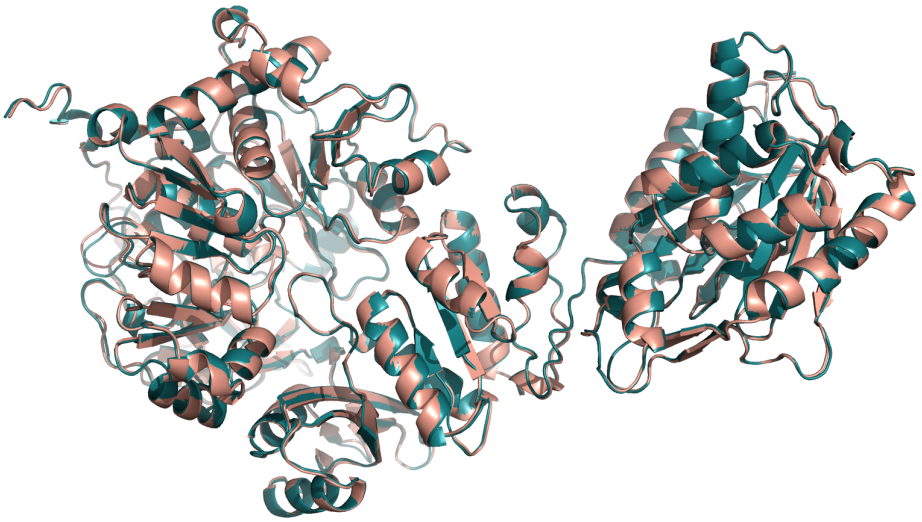

Fig. 15. Structural alignment using PyMOL of 4CL:GSG:CHS structures before (electric) and after (pink) energy minimization with Yasara

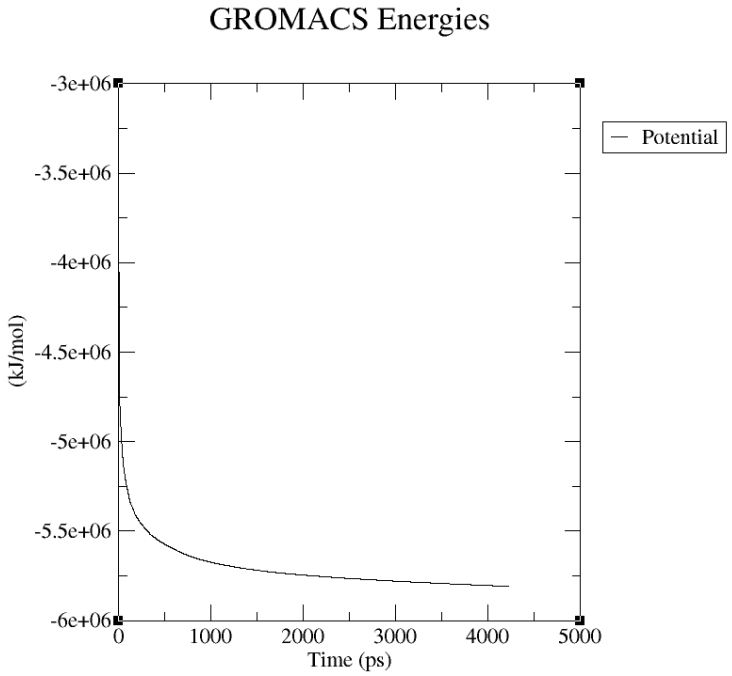

Fig. 16. xmgrace energy graph of 4CL:GSG:CHS structure after molecular dynamics (energy minimization) simulation

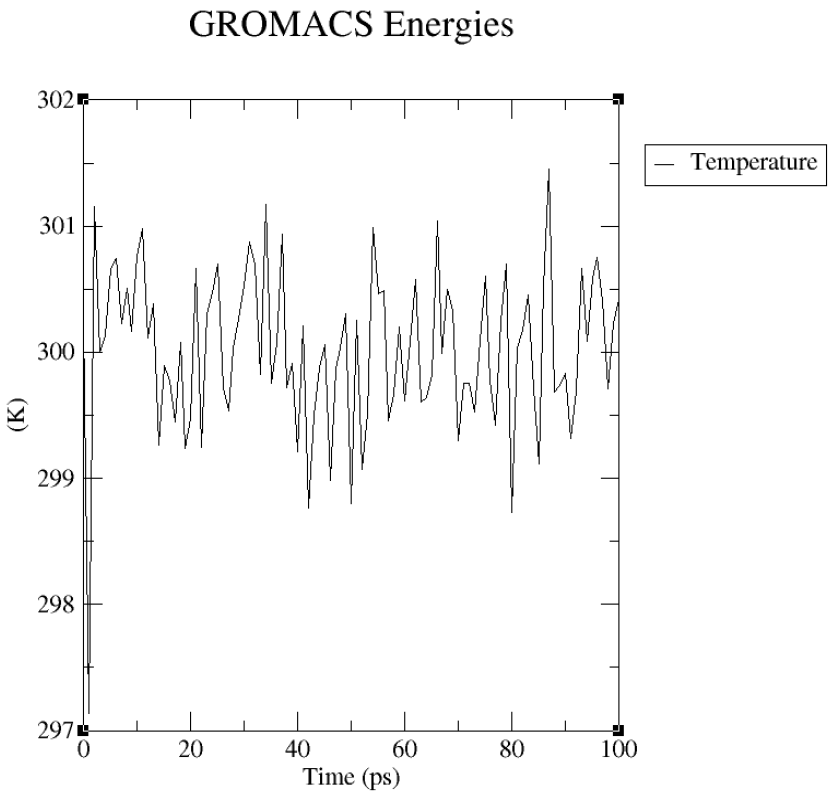

Fig. 17. xmgrace energy graph of 4CL:GSG:CHS structure after molecular dynamics temperature equilibration

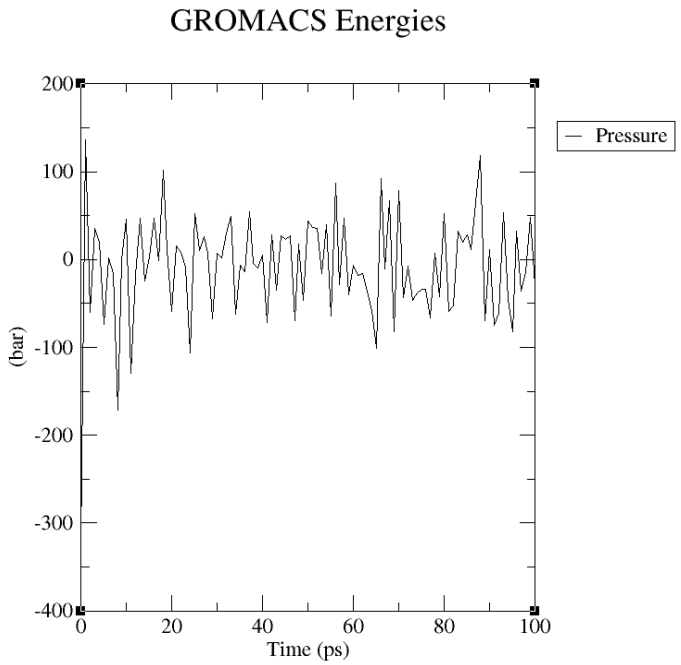

Fig. 18. xmgrace energy graph of 4CL:GSG:CHS structure after molecular dynamics pressure equilibration simulation

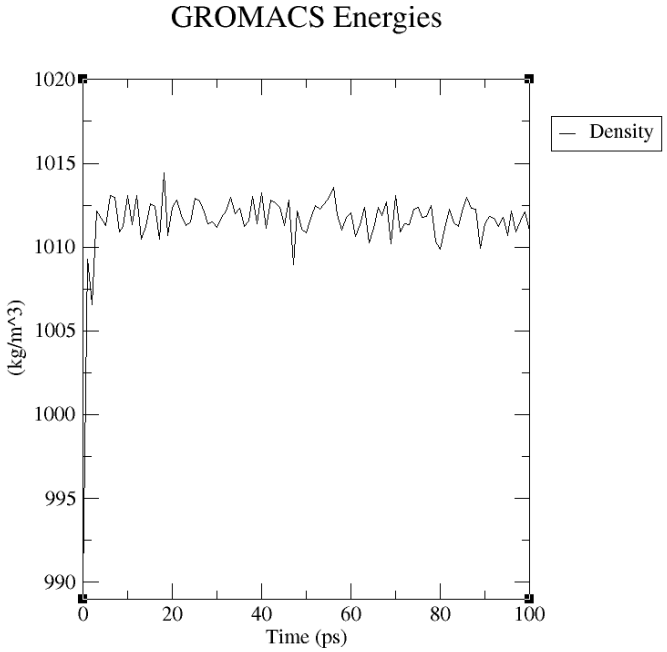

Fig. 19. xmgrace energy graph of 4CL:GSG:CHS structure after molecular dynamics (production run) simulation

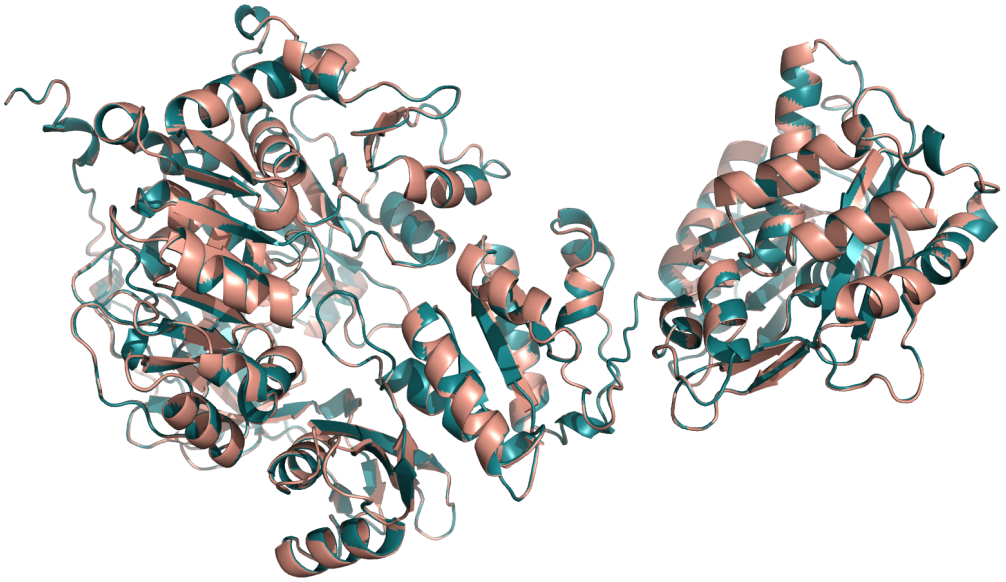

Fig. 20. 4CL:GSG:CHS model structural alignment before (**electric**) MD and after (**pink**)

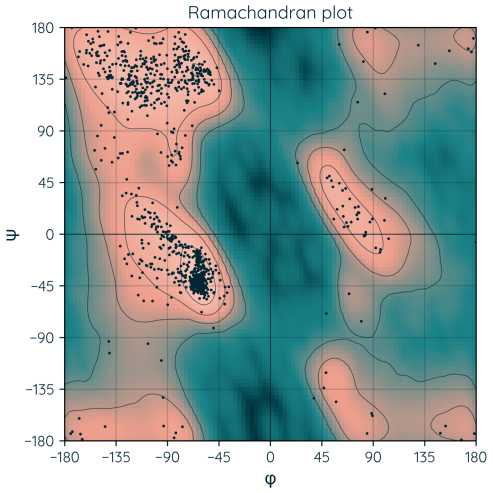

Fig. 21. Ramachandran plot of the best model of 4CL:GSG:CHS before MD

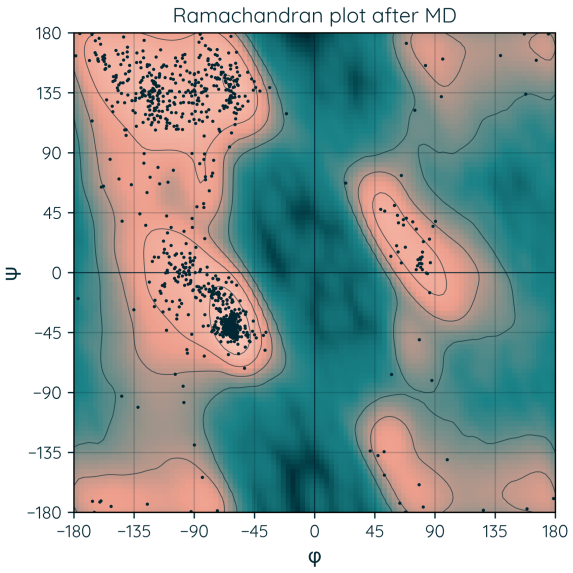

Fig. 22. Ramachandran plot of the best model of 4CL:GSG:CHS after MD

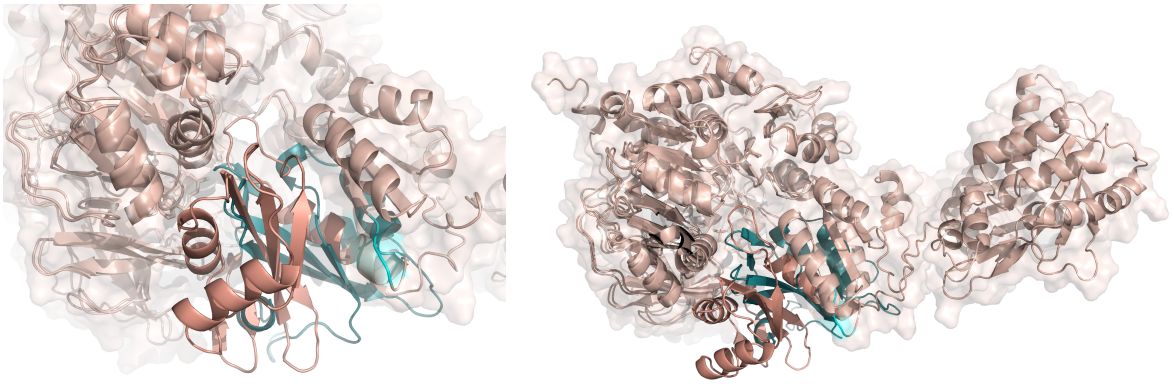

Fig. 23. Disordered domain of 4CL:GSG:CHS domain in modeled structure (**pink**), domain in reference structure of 4CL (**electric**), GSG linker (**cyan**).

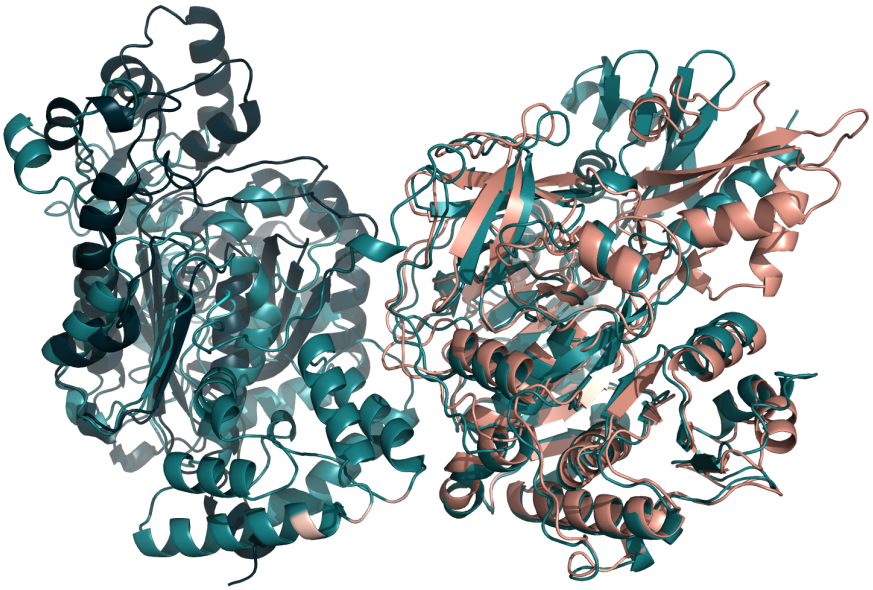

Fig. 24. Modeled CHS:GSG:4CL (fusion - **electric**, linker - **cyan**) overlay with template structures of enzymes 4CL (**pink**) with disordered domain (**light pink**) and CHS (**dark green**)

Table 12. Evaluation of CHS:GSG:4CL structures

RoseTTAFold	
Model	VoroMQA score
MODEL_1_CHS_GSG_4CL.pdb	0.487163
MODEL_2_CHS_GSG_4CL.pdb	0.487654
MODEL_3_CHS_GSG_4CL.pdb	**0.490014**
MODEL_4_CHS_GSG_4CL.pdb	0.469502
MODEL_5_CHS_GSG_4CL.pdb	0.448322

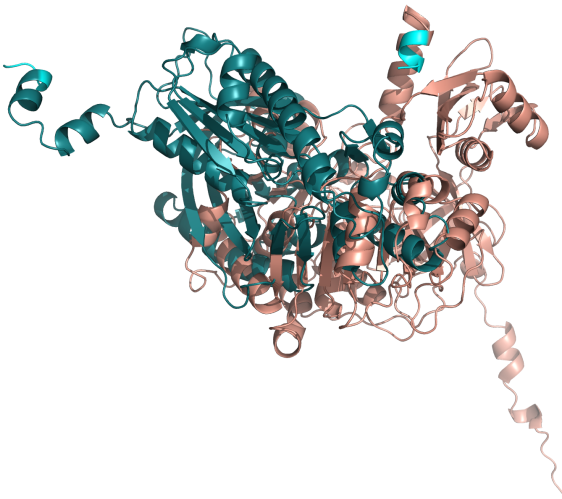

Fig. 25. Modeled 4CL:EAAAK (**electric**) and EAAAK:CHS (**pink**) overlay (EAAAK colored **cyan**)

Table 13. Evaluation of 4CL:EAAAK structures

RoseTTAFold	
Model	VoroMQA score
MODEL_1_4CL_EAAAK.pdb	**0.535044**
MODEL_2_4CL_EAAAK.pdb	0.530567
MODEL_3_4CL_EAAAK.pdb	0.532237
MODEL_4_4CL_EAAAK.pdb	0.53092
MODEL_5_4CL_EAAAK.pdb	0.530439

Table 14. Evaluation of EAAAK:CHS structures

RoseTTAFold	
Model	VoroMQA score
MODEL_1_EAAAK_CHS.pdb	0.517958
MODEL_2_EAAAK_CHS.pdb	**0.531011**
MODEL_3_EAAAK_CHS.pdb	0.51568
MODEL_4_EAAAK_CHS.pdb	0.512811
MODEL_5_EAAAK_CHS.pdb	0.516752

Results of modeling approach with MSA

Linker GSG

Robetta's confidence score was **0.8**. The higher angstroms error estimate is observed in the N terminal of the complex and the linker region (**532-666**).

Table 15. Evaluation of 4CL:GSG:CHS structures that were modeled with provided MSA

RoseTTAFold			
Model	QMEAN	QMEANDisCo	VoroMQA score
MODEL_1_4CL_GSG_CHS_MSA.pdb	**1.15**	**0.71**	0.50925
MODEL_2_4CL_GSG_CHS_MSA.pdb	0.7	**0.71**	0.499289
MODEL_3_4CL_GSG_CHS_MSA.pdb	0.87	**0.71**	0.505087
MODEL_4_4CL_GSG_CHS_MSA.pdb	0.93	**0.71**	**0.511291**
MODEL_5_4CL_GSG_CHS_MSA.pdb	0.89	**0.71**	0.494692

Linker EAAAK

Robetta's confidence score is **0.8**. The higher angstroms error estimate is observed in the N terminal of the complex and the linker region (**549 - 664**).

Table 16. Evaluation of 4CL:EAAAK:CHS structures that were modeled with provided MSA

RoseTTAFold			
Model	QMEAN	QMEANDisCo	VoroMQA score
MODEL_1_4CL_EAAAK_CHS_MSA.pdb	0.95	**0.72**	**0.509592**
MODEL_2_4CL_EAAAK_CHS_MSA.pdb	0.85	**0.72**	0.50054
MODEL_3_4CL_EAAAK_CHS_MSA.pdb	**1.31**	**0.72**	0.507024
MODEL_4_4CL_EAAAK_CHS_MSA.pdb	0.58	0.71	0.502846
MODEL_5_4CL_EAAAK_CHS_MSA.pdb	1.34	**0.72**	0.49208

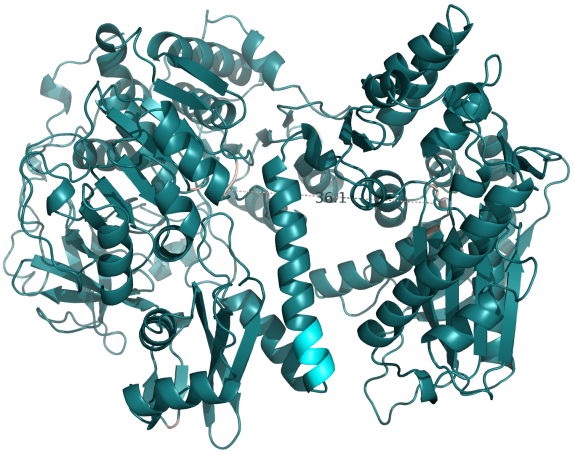

Fig. 26. Measured distance between active sites of 4CL:EAAAK:CHS (4CL on the left, CHS on the right)

Linker (EAAAK)2

Robetta's confidence score is **0.81**. The higher angstroms error estimate is observed in the N terminal of the complex and the linker region (**530 - 674**).

Table 17. Evaluation of 4CL:(EAAAK)2:CHS structures that were modeled with provided MSA

RoseTTAFold			
Model	QMEAN	QMEANDisCo	VoroMQA score
MODEL_1_4CL_EAAAK_2_CHS.pdb	1.01	**0.72**	**0.51103**
MODEL_2_4CL_EAAAK_2_CHS.pdb	**1.39**	**0.72**	0.501628
MODEL_3_4CL_EAAAK_2_CHS.pdb	0.67	0.71	0.50042
MODEL_4_4CL_EAAAK_2_CHS.pdb	1.03	**0.72**	0.504132
MODEL_5_4CL_EAAAK_2_CHS.pdb	0.76	0.71	0.508873

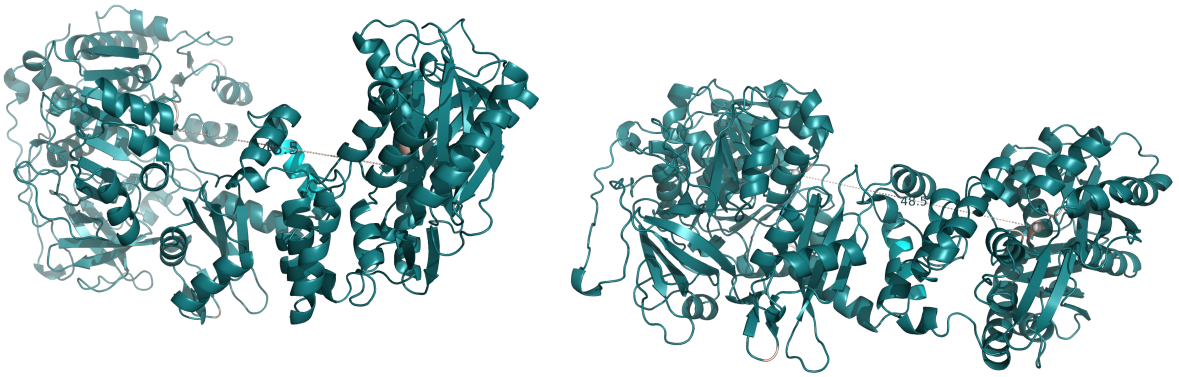

Fig. 27. Measured distance between active sites of 4CL:(EAAAK)2:CHS (4CL on the left of each picture, CHS on the right of each picture)

Linker (EAAAK)3

Robetta's confidence score is **0.79**. The higher angstroms error estimate is observed in the N terminal of the complex and the linker region (**553 - 702**).

Table 18. Evaluation of 4CL:(EAAAK)3:CHS structures that were modeled with provided MSA

RoseTTAFold			
Model	QMEAN	QMEANDisCo	VoroMQA score
MODEL_1_4CL_EAAAK_3_CHS.pdb	**1.39**	**0.72**	0.491909
MODEL_2_4CL_EAAAK_3_CHS.pdb	0.9	0.71	0.500159
MODEL_3_4CL_EAAAK_3_CHS.pdb	1.13	**0.72**	0.494661
MODEL_4_4CL_EAAAK_3_CHS.pdb	0.72	0.71	0.487712
MODEL_5_4CL_EAAAK_3_CHS.pdb	0.87	0.71	**0.503178**

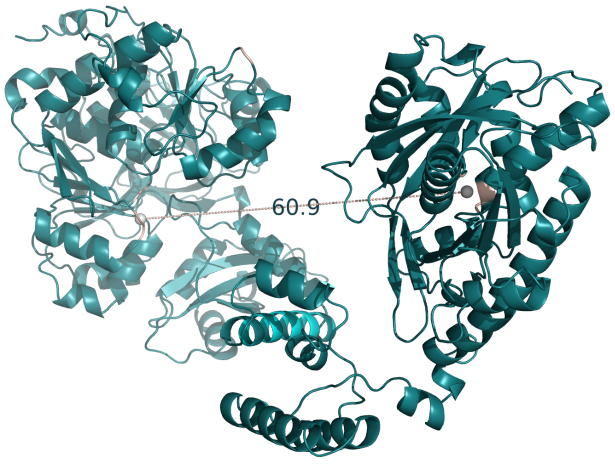

Fig. 28. Measured distance between active sites of 4CL:(EAAAK)3:CHS before structure refinement with molecular dynamics (4CL on the left, CHS on the right)

Linker GGGGS

Robetta's confidence score is **0.81**. The higher angstroms error estimate is observed in the N terminal of the complex and the linker region (**535-668**).

Table 19. Evaluation of 4CL:GGGGS:CHS structures that were modeled with provided MSA

RoseTTAFold			
Model	QMEAN	QMEANDisCo	VoroMQA score
MODEL_1_4CL_GGGGS_CHS_MSA.pdb	1.23	**0.72**	0.509968
MODEL_2_4CL_GGGGS_CHS_MSA.pdb	**1.25**	**0.72**	0.506824
MODEL_3_4CL_GGGGS_CHS_MSA.pdb	1.22	**0.72**	0.503358
MODEL_4_4CL_GGGGS_CHS_MSA.pdb	0.73	**0.72**	**0.510973**
MODEL_5_4CL_GGGGS_CHS_MSA.pdb	0.53	0.71	0.508495

Linker (GGGGS)2

Robetta's confidence score is **0.78**. The higher angstroms error estimate is observed in the N terminal of the complex and the linker region (**552 - 670**).

Table 20. Evaluation of 4CL:(GGGGS)2:CHS structures that were modeled with provided MSA

RoseTTAFold			
Model	QMEAN	QMEANDisCo	VoroMQA score
MODEL_1_4CL_GGGGS_2_CHS_MSA.pdb	**0.97**	0.71	**0.509898**
MODEL_2_4CL_GGGGS_2_CHS_MSA.pdb	0.86	0.71	0.490104
MODEL_3_4CL_GGGGS_2_CHS_MSA.pdb	0.95	0.71	0.495492
MODEL_4_4CL_GGGGS_2_CHS_MSA.pdb	0.5	**0.72**	0.500793
MODEL_5_4CL_GGGGS_2_CHS_MSA.pdb	0.22	0.71	0.49708

Linker (GGGGS)3

Robetta's confidence score is **0.79**. The higher angstroms error estimate is observed in the N terminal of the complex and the linker region (**535-679**).

Table 21. Evaluation of 4CL:(GGGGS)3:CHS structures that were modeled with provided MSA

RoseTTAFold			
Model	QMEAN	QMEANDisCo	VoroMQA score
MODEL_1_4CL_GGGGS_3_CHS_MSA.pdb	0.65	0.71	**0.502113**
MODEL_2_4CL_GGGGS_3_CHS_MSA.pdb	**1.58**	**0.72**	0.494236
MODEL_3_4CL_GGGGS_3_CHS_MSA.pdb	0.25	0.71	0.497905
MODEL_4_4CL_GGGGS_3_CHS_MSA.pdb	0.64	0.71	0.495384
MODEL_5_4CL_GGGGS_3_CHS_MSA.pdb	0.65	0.71	0.488994

Modeling with AlphaFold2

Linker GSG

Table 22. Evaluation of 4CL:GSG:CHS structures modeled with AlphaFold2

AlphaFold2							
Model	QMEAN	QMEAN DisCo	VoroMQA	ProQ2D	ProQRosCenD	ProQRosFAD	ProQ3D
4CL_GSG_CHS_unrelaxed_model_1.pdb	**-1.01**	**0.78**	**0.539915**	**0.709**	0.625	**0.807**	0.705
4CL_GSG_CHS_unrelaxed_model_2.pdb	-1.18	0.77	0.537669	0.701	**0.633**	0.801	**0.721**

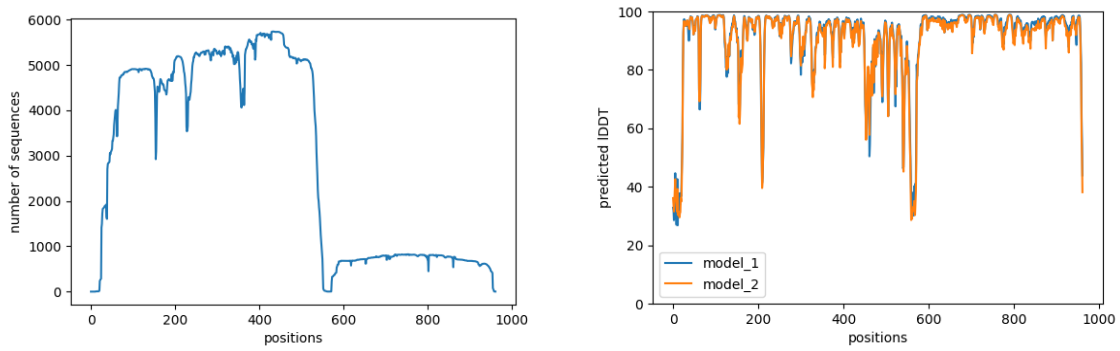

Fig. 29. Coverage and predicted lDDT per position of 4CL:GSG:CHS structure modeled with AlphaFold2

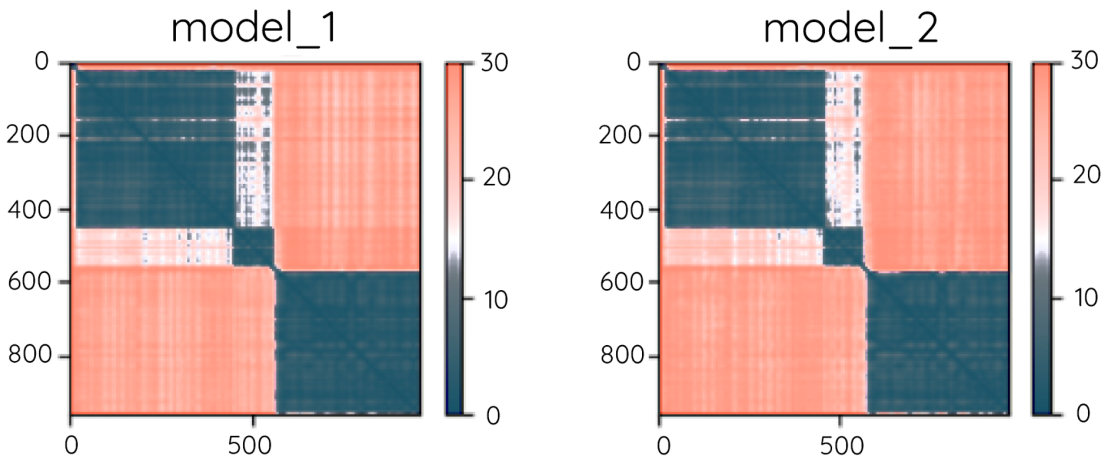

Fig. 30. Predicted alignment error of 4CL:GSG:CHS structure modeled with AlphaFold2

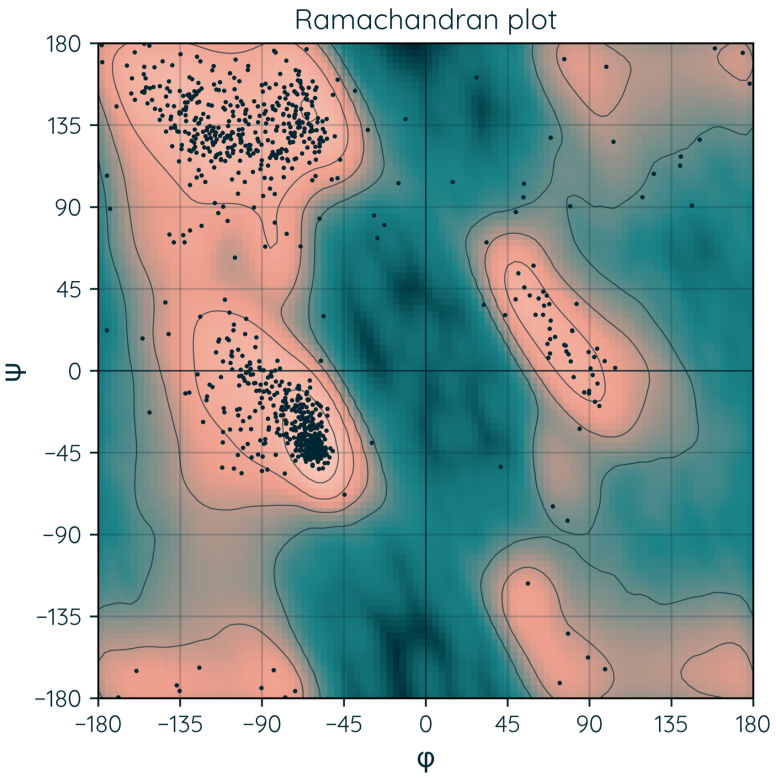

Fig. 31. Ramachandran plot of the best 4CL:GSG:CHS structure modeled with AlphaFold2

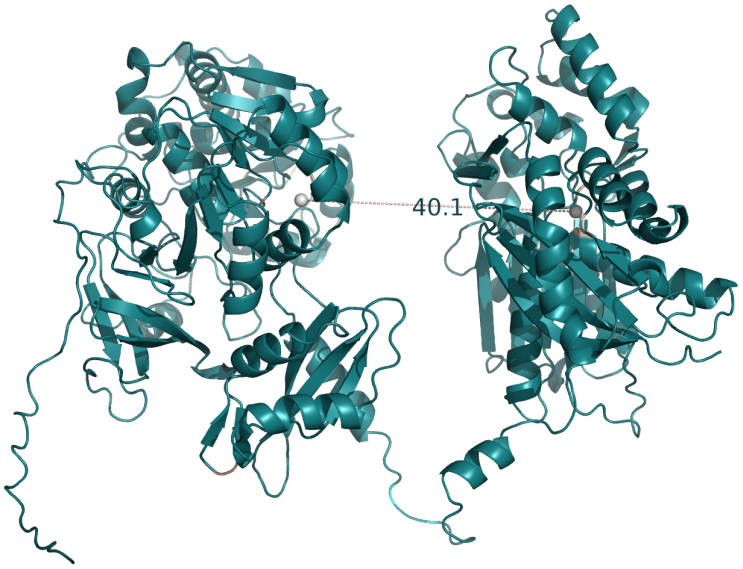

Fig. 32. Illustrating distance between 4CL:GSG:CHS active sites

Linker EAAAK

Table 23. Evaluation of 4CL:EAAAK:CHS structures modeled with AlphaFold2

AlphaFold2							
Model	QMEAN	QMEAN DisCo	VoroMQA	ProQ2D	ProQRosCenD	ProQRosFAD	ProQ3D
4CL_EAAAK_CHS_unrelaxed_model_1.pdb	**-0.69**	**0.79**	**0.542077**	**0.708**	0.624	0.794	0.710
4CL_EAAAK_CHS_unrelaxed_model_2.pdb	-1.03	**0.79**	0.536476	0.7	**0.63**	**0.798**	**0.717**

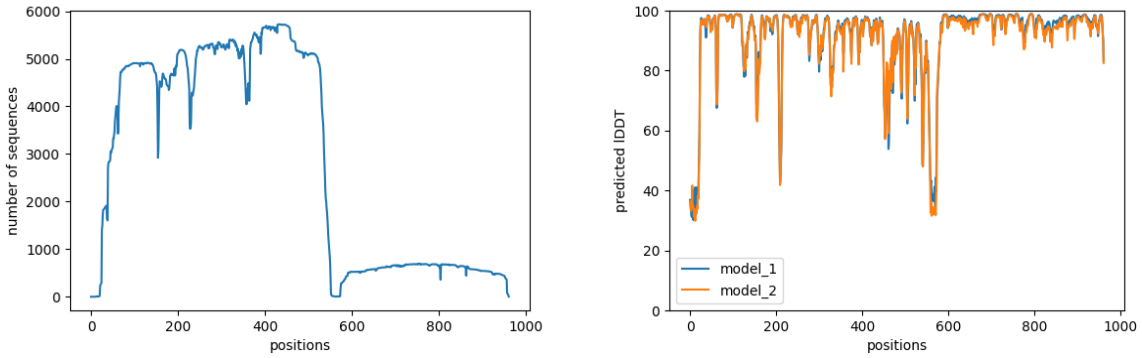

Fig. 33. Coverage and predicted lDDT per position of 4CL:EAAAK:CHS structure modeled with AlphaFold2

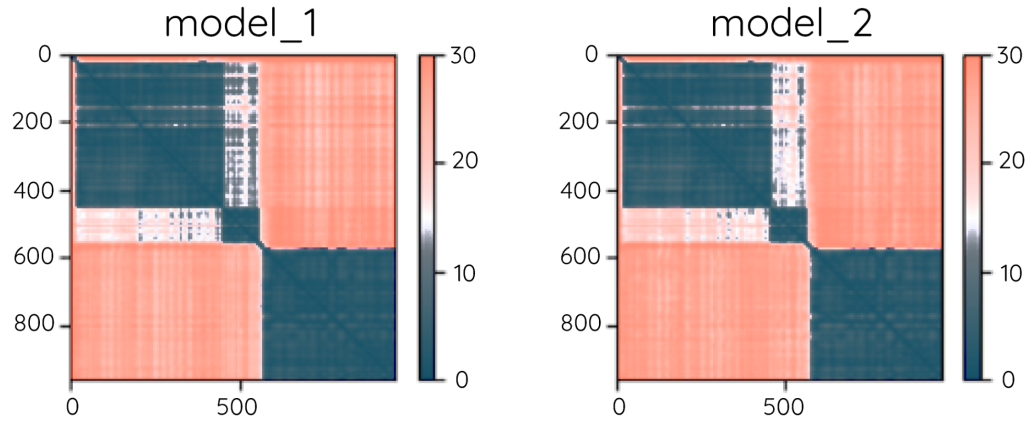

Fig. 34. Predicted alignment` error of 4CL:EAAAK:CHS structure modeled with AlphaFold2

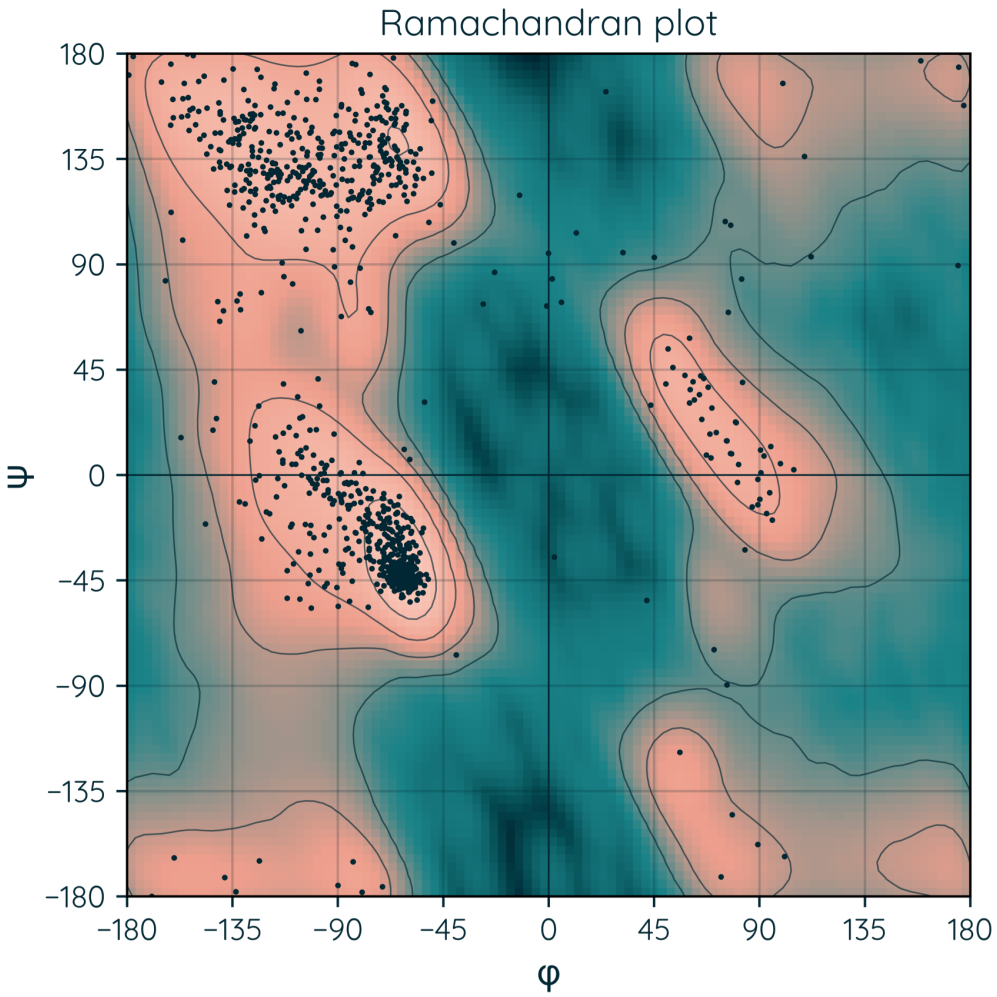

Fig. 35. Ramachandran plot of the best 4CL:EAAAK:CHS structure modeled with AlphaFold2

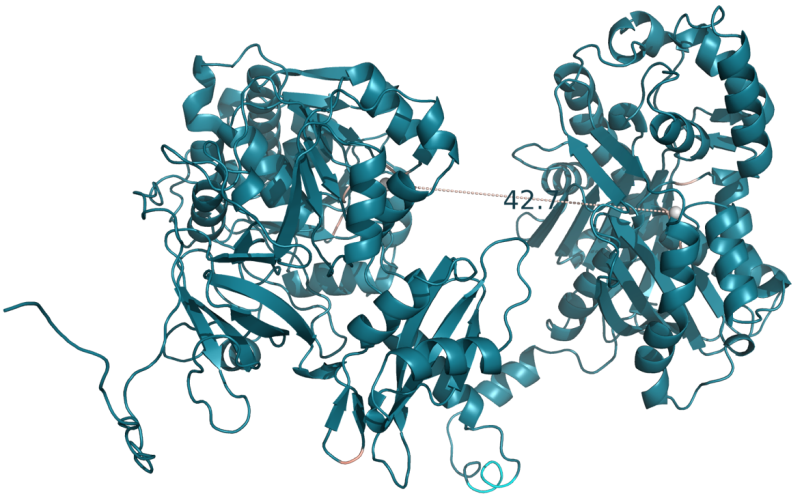

Fig. 36. Illustrating distance between 4CL:EAAAK:CHS active sites

Linker (EAAAK)2

Table 24. Evaluation of 4CL:(EAAAK)2:CHS structures modeled with AlphaFold2

AlphaFold2							
Model	QMEAN	QMEAN DisCo	VoroMQA	ProQ2D	ProQRosCenD	ProQRosFAD	ProQ3D
4CL_EAAAK_2_CHS_unrelaxed_model_1.pdb	**-0.78**	0.79	**0.541834**	0.69	**0.638**	**0.81**	**0.718**
4CL_EAAAK_2_CHS_unrelaxed_model_2.pdb	-1.16	**0.78**	0.536158	**0.693**	0.616	0.789	0.691

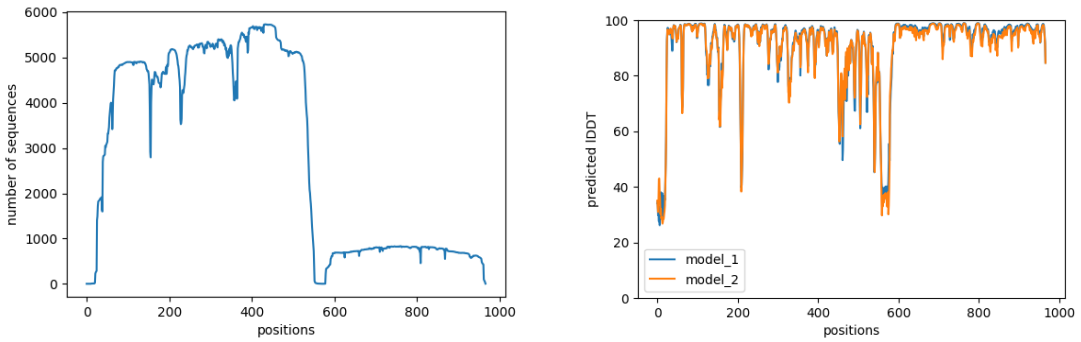

Fig. 37. Coverage and predicted lDDT per position of 4CL:(EAAAK)2:CHS structure modeled with AlphaFold2

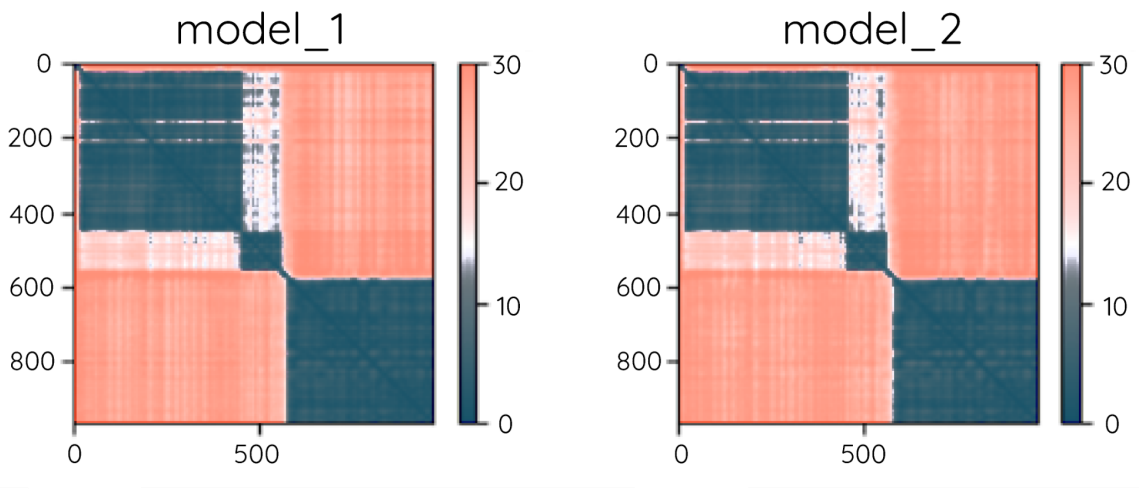

Fig. 38. Predicted alignment error of 4CL:(EAAAK)2:CHS structure modeled with AlphaFold2

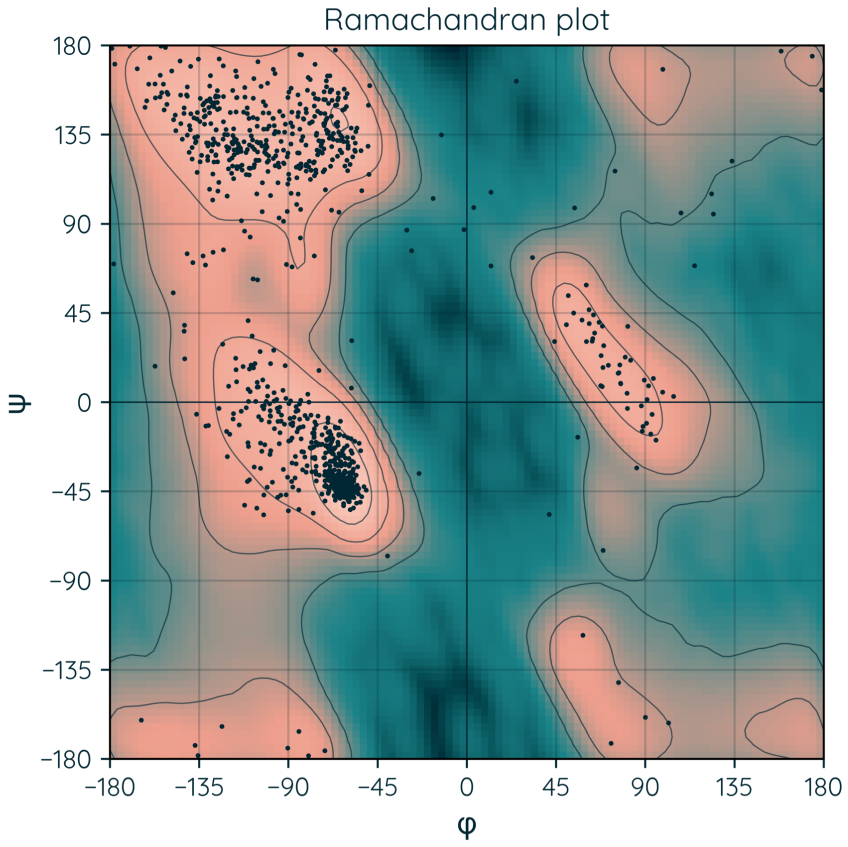

Fig. 39. Ramachandran plot of the best 4CL:(EAAAK)2:CHS structure modeled with AlphaFold2

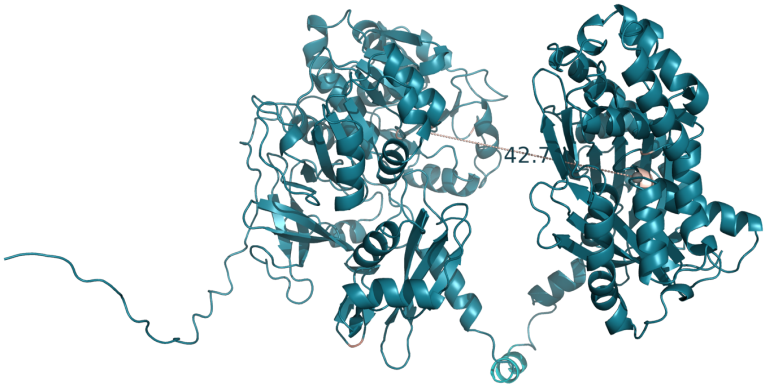

Fig. 40. Illustrating distance between 4CL:(EAAAK)2:CHS active sites

Linker (EAAAK)3

Table 25. Evaluation of 4CL:(EAAAK)3:CHS structures modeled with AlphaFold2

AlphaFold2							
Model	QMEAN	QMEAN DisCo	VoroMQA	ProQ2D	ProQRosCenD	ProQRosFAD	ProQ3D
4CL_EAAAK_3_CHS_unrelaxed_model_1.pdb	**-1.36**	**0.78**	**0.537969**	**0.693**	0.617	0.795	0.697
4CL_EAAAK_3_CHS_unrelaxed_model_2.pdb	-1.38	0.77	0.533997	0.693	**0.634**	**0.803**	**0.713**

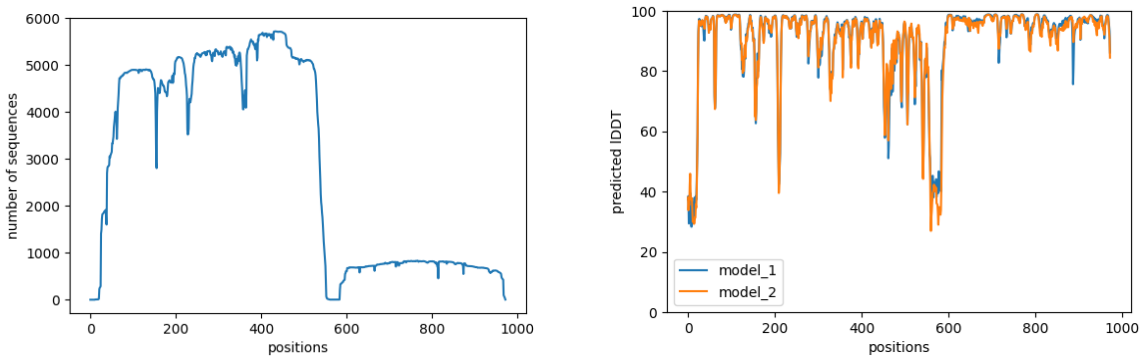

Fig. 41. Coverage and predicted lDDT per position of 4CL:(EAAAK)3:CHS structure modeled with AlphaFold2

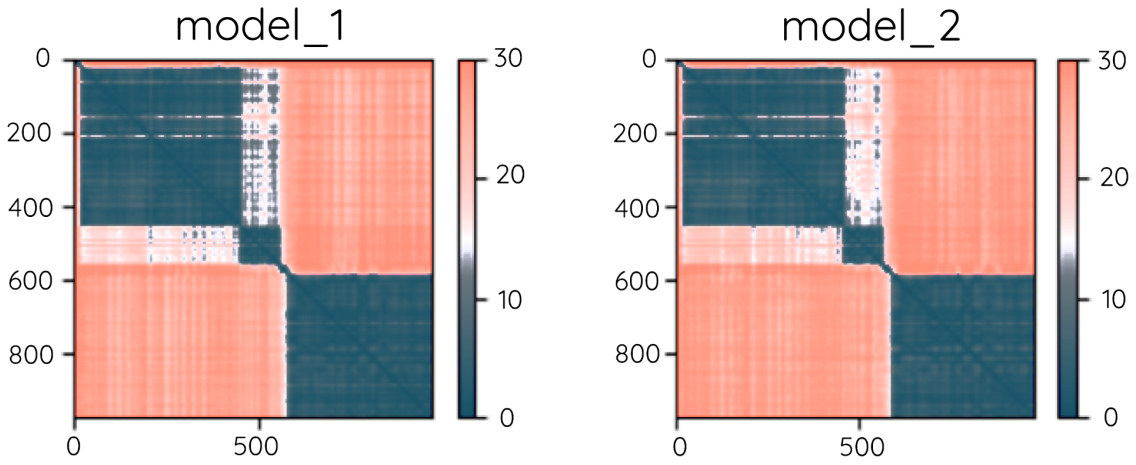

Fig. 42. Predicted alignment error of 4CL:(EAAAK)3:CHS structure modeled with AlphaFold2

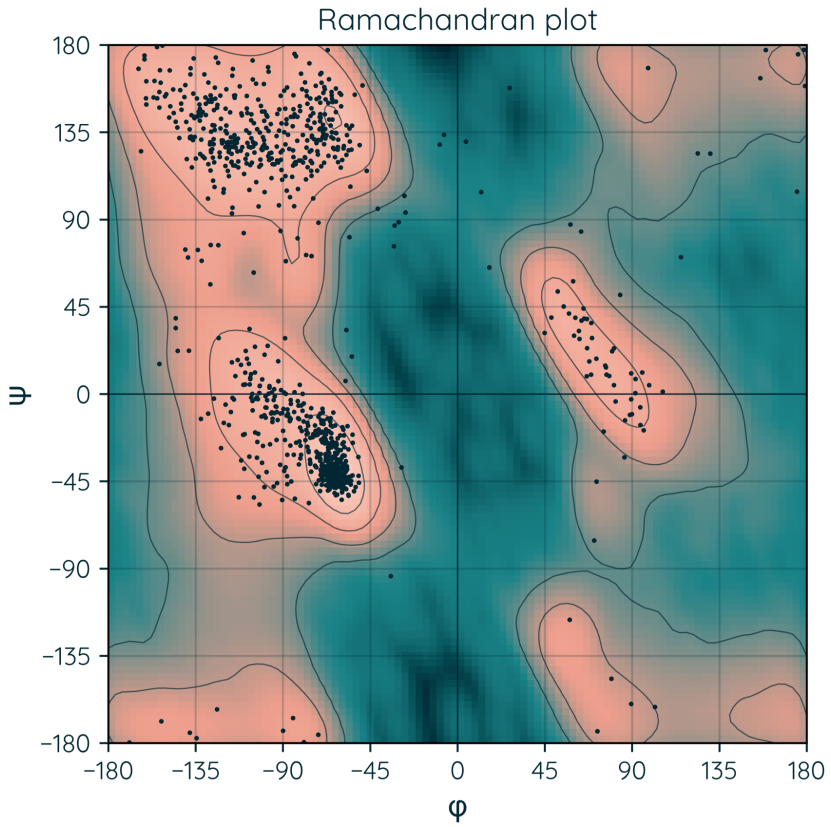

Fig. 43. Ramachandran plot of the best 4CL:(EAAAK)3:CHS structure modeled with AlphaFold2

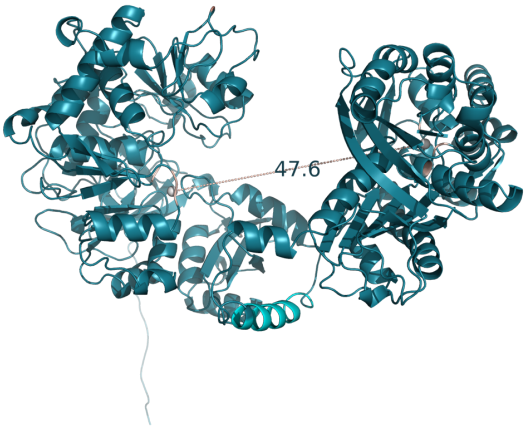

Fig. 44. Illustrating distance between 4CL:(EAAAK)3:CHS active sites

Linker GGGGS

Table 26. Evaluation of 4CL:GGGGS:CHS structures modeled with AlphaFold2

AlphaFold2							
Model	QMEAN	QMEAN DisCo	VoroMQA	ProQ2D	ProQRosCenD	ProQRosFAD	ProQ3D
4CL_GGGGS_CHS_unrelaxed_model_1.pdb	**-0.75**	**0.78**	**0.542**	**0.691**	0.617	0.779	0.698
4CL_GGGGS_CHS_unrelaxed_model_2.pdb	-1.16	0.77	0.536974	0.691	**0.640**	**0.809**	**0.725**

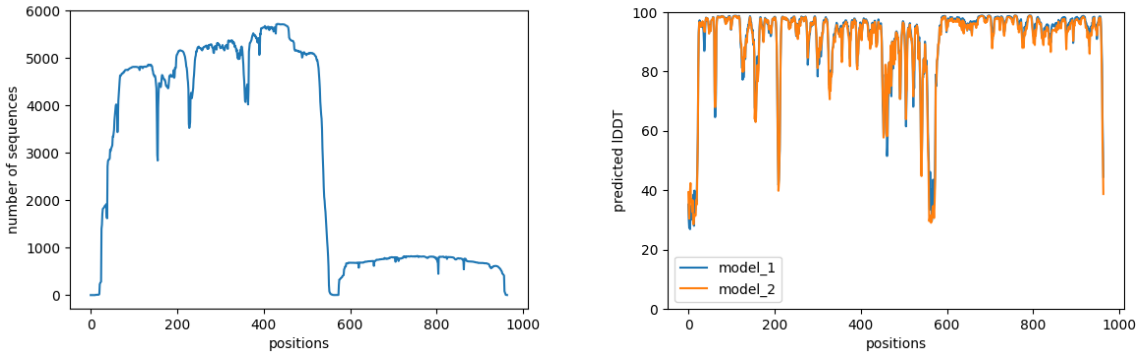

Fig. 45. Coverage and predicted lDDT per position of 4CL:GGGGS:CHS structure modeled with AlphaFold2

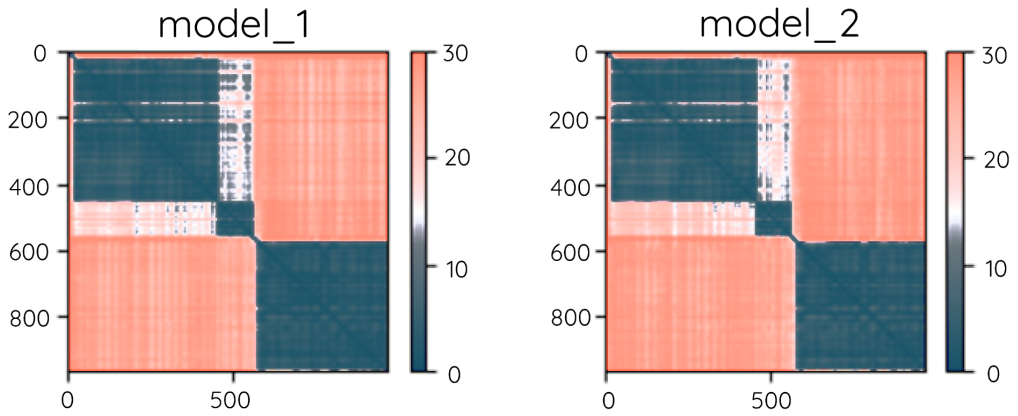

Fig. 46. Predicted alignment error of 4CL:GGGGS:CHS structure modeled with AlphaFold2

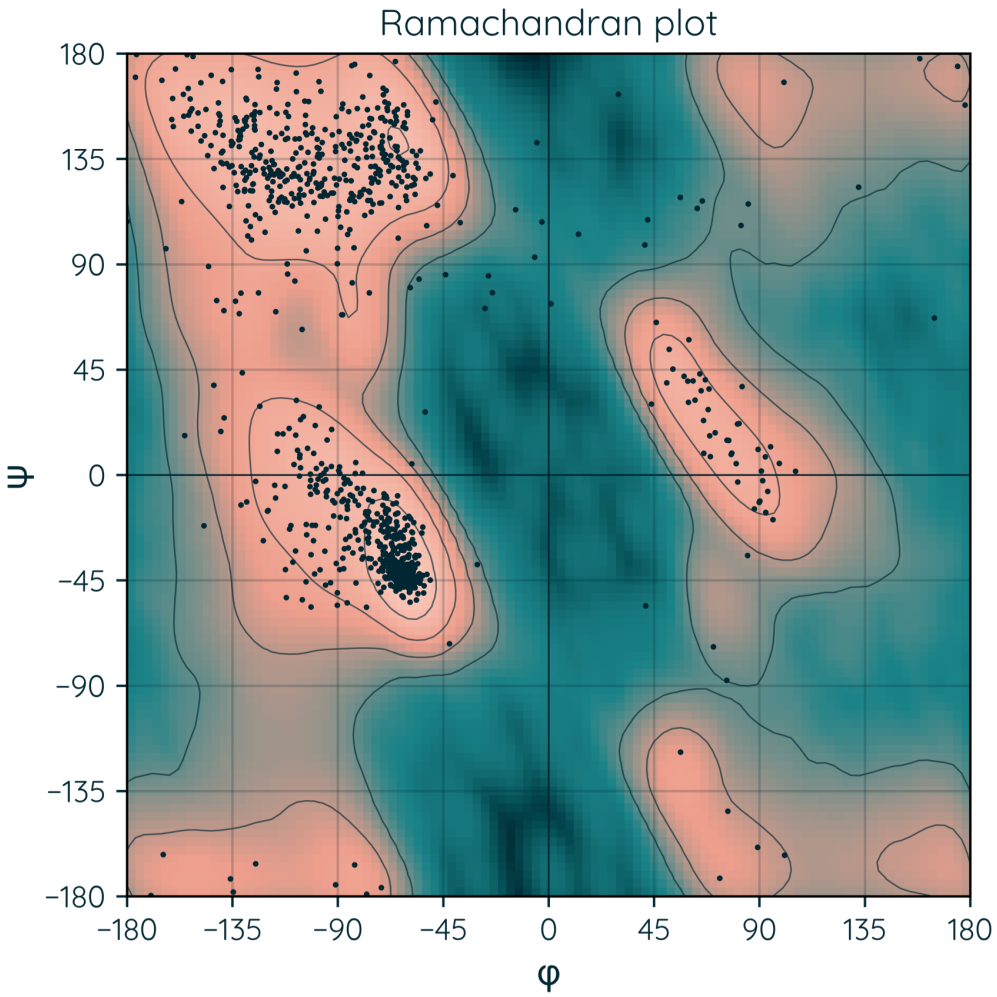

Fig. 47. Ramachandran plot of the best 4CL:GGGGS:CHS structure modeled with AlphaFold2

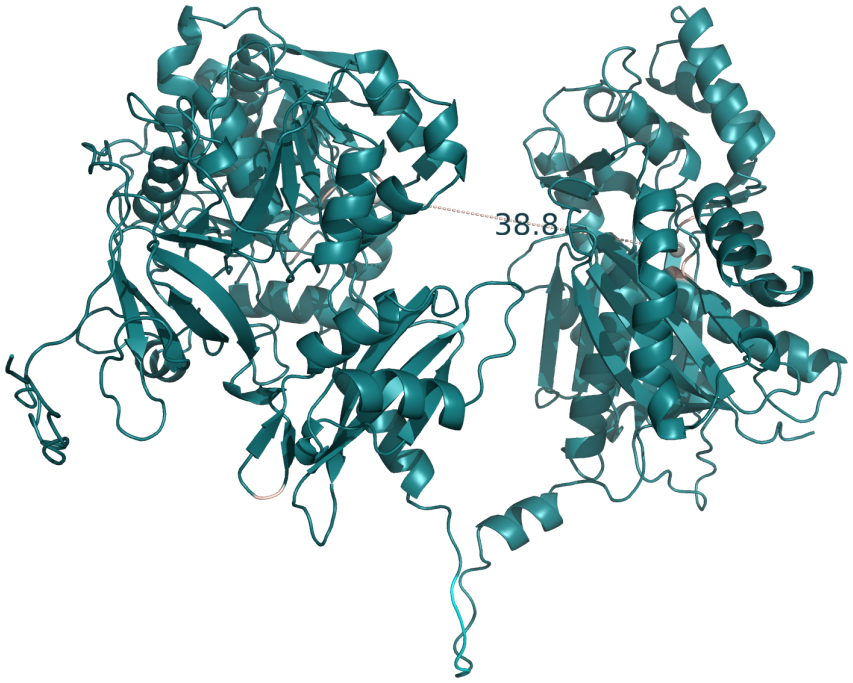

Fig. 48. Illustrating distance (38.8 Å) between 4CL:GGGGS:CHS active sites

Linker (GGGGS)2

Table 27. Evaluation of 4CL:(GGGGS)2:CHS structures modeled with AlphaFold2

AlphaFold2							
Model	QMEAN	QMEAN DisCo	VoroMQA	ProQ2D	ProQRosCenD	ProQRosFAD	ProQ3D
4CL_GGGGS_2_CHS_unrelaxed_model_1.pdb	-1.02	**0.78**	**0.540667**	**0.697**	**0.635**	**0.807**	0.716
4CL_GGGGS_2_CHS_unrelaxed_model_2.pdb	**-0.9**	**0.78**	0.532784	**0.697**	0.629	0.802	**0.725**

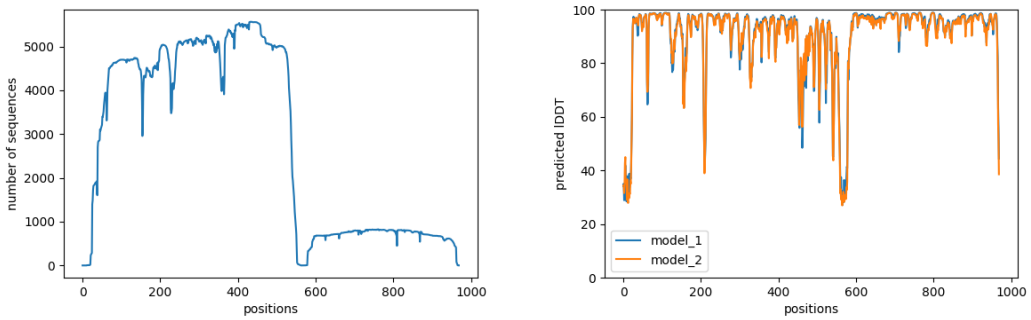

Fig. 49. Coverage and predicted lDDT per position of 4CL:(GGGGS)2:CHS structure modeled with AlphaFold2

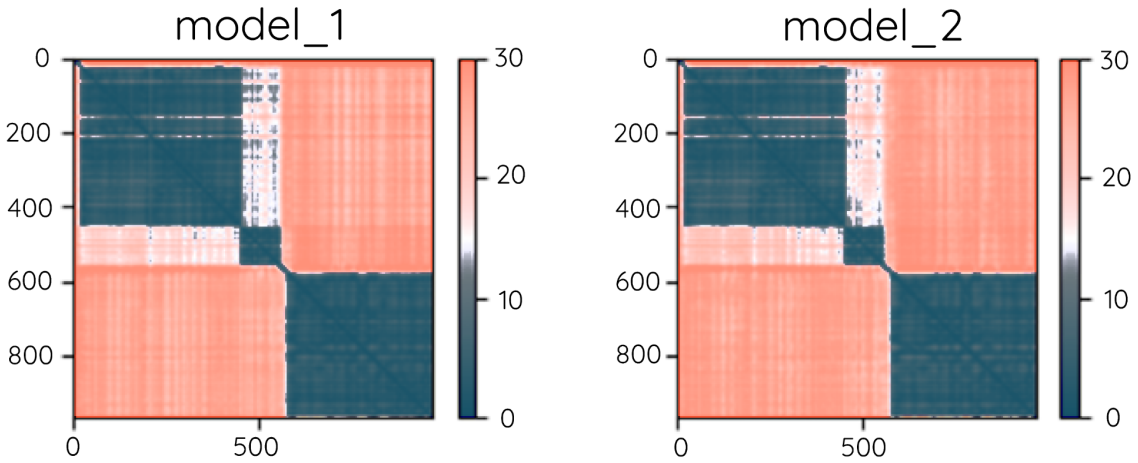

Fig. 50. Predicted alignment error of 4CL:(GGGGS)2:CHS structure modeled with AlphaFold2

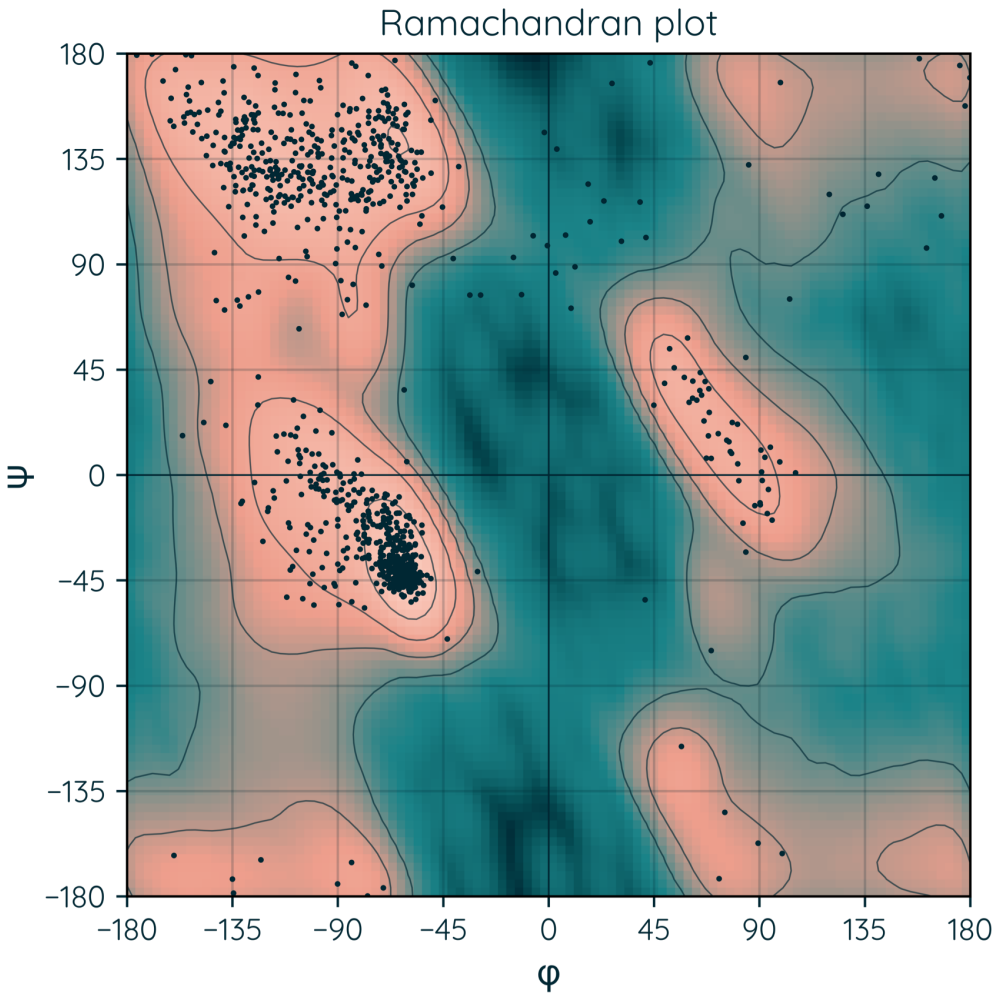

Fig. 51. Ramachandran plot of the best 4CL:(GGGGS)2:CHS structure modeled with AlphaFold2

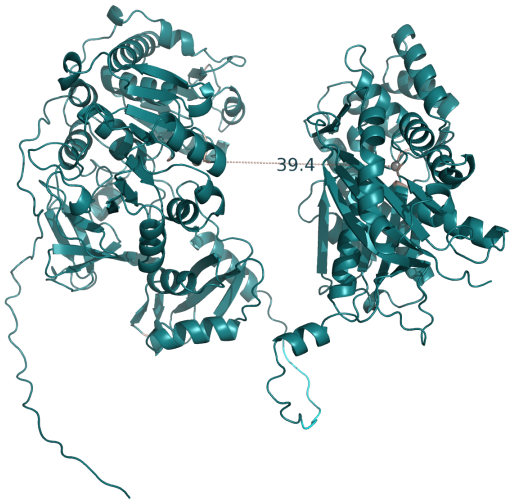

Fig. 52. Illustrating distance (39.4 Å) between 4CL:(GGGGS)2:CHS active sites

Linker (GGGGS)3

Table 28. Evaluation of 4CL:(GGGGS)3:CHS structures modeled with AlphaFold2

AlphaFold2							
Model	QMEAN	QMEAN DisCo	VoroMQA	ProQ2D	ProQRosCenD	ProQRosFAD	ProQ3D
4CL_GGGGS_3_CHS_unrelaxed_model_1.pdb	**-1.25**	**0.78**	**0.540568**	**0.696**	**0.627**	0.801	0.707
4CL_GGGGS_3_CHS_unrelaxed_model_2.pdb	-1.36	0.77	0.531981	0.687	0.626	**0.804**	**0.719**

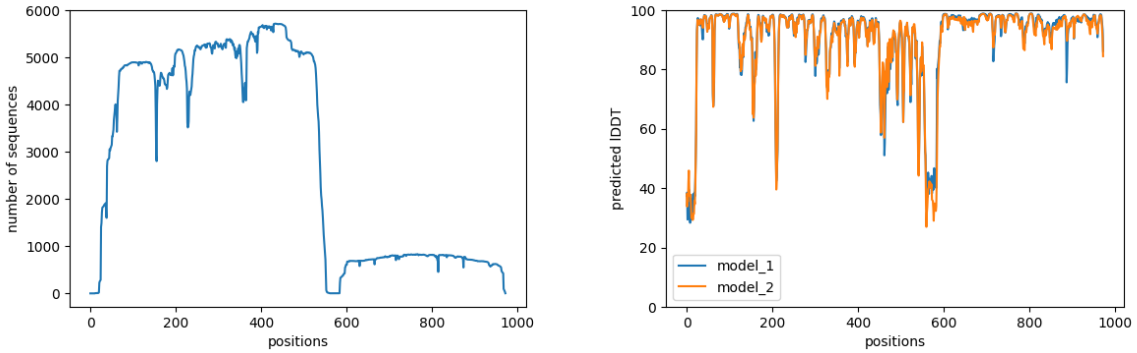

Fig. 53. Coverage and predicted lDDT per position of 4CL:(GGGGS)3:CHS structure modeled with AlphaFold2

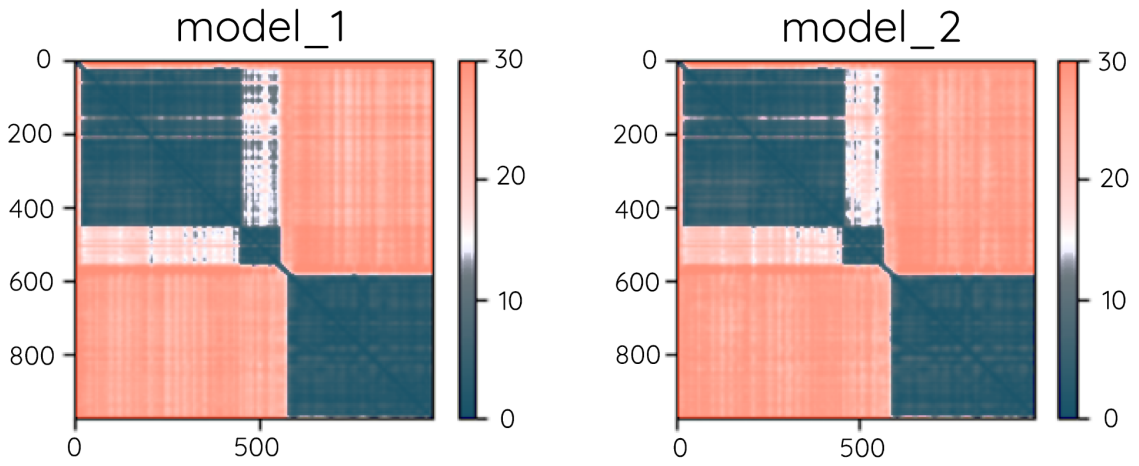

Fig. 54. Predicted alignment error of 4CL:(GGGGS)3:CHS structure modeled with AlphaFold2

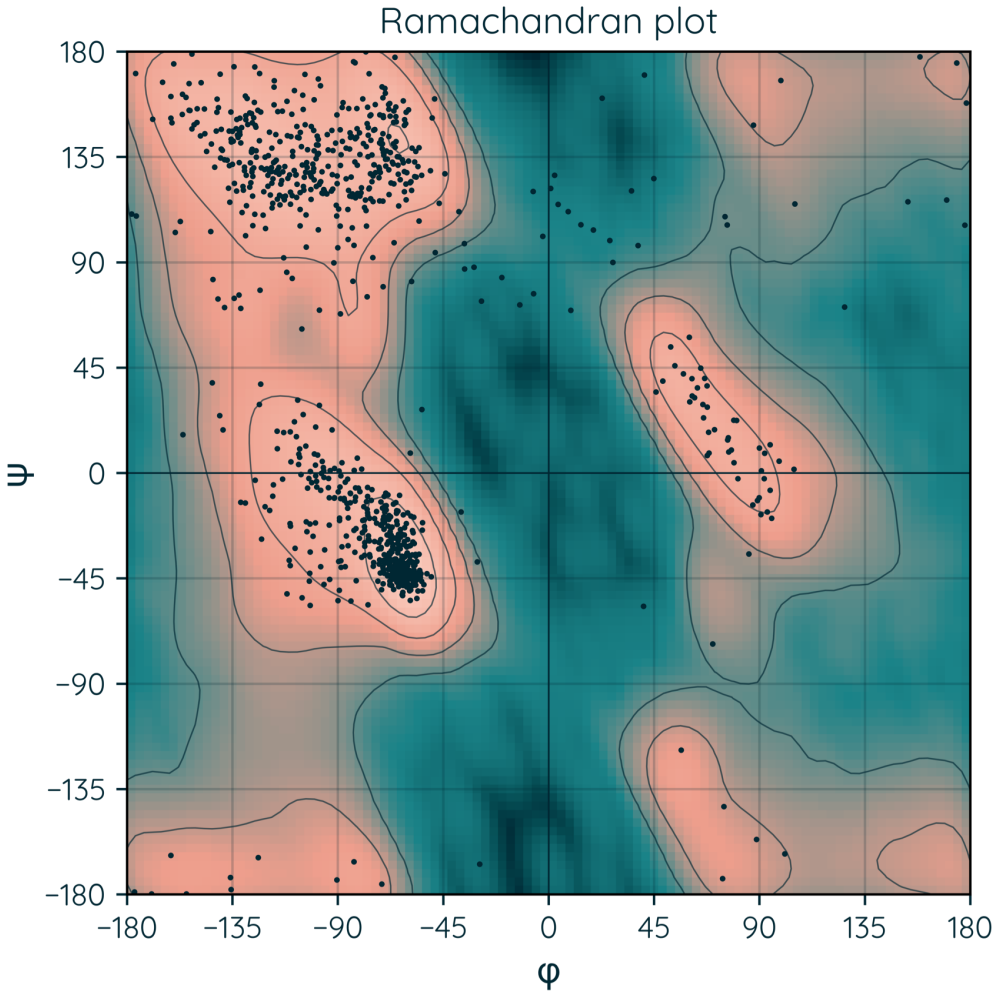

Fig. 55. Ramachandran plot of the best 4CL:(GGGGS)3:CHS structure modeled with AlphaFold2

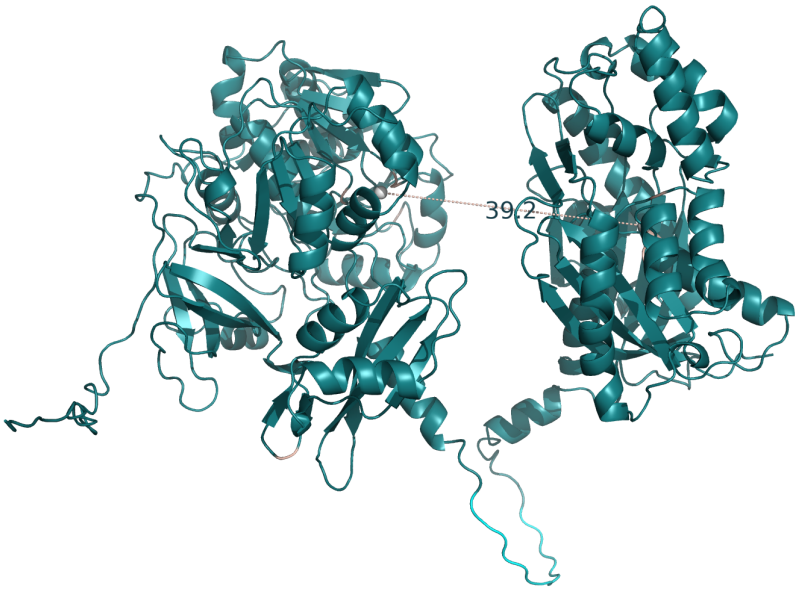

Fig. 56. Illustrating distance (39.2 Å) between 4CL:(GGGGS)3:CHS active sites

Conclusions

Table 29. Summary of distance between active sites and naringenin concentration

Linker	Distance between active sites (Å)	Naringenin concentration
GSG	40.1	
GGGGS	38.8	
(GGGGS)2	39.4	
(GGGGS)3	39.2	
EAAAK	42.7	
(EAAAK)2	42.7	
(EAAAK)3	47.6	

Modeling linkers as unknown amino acids

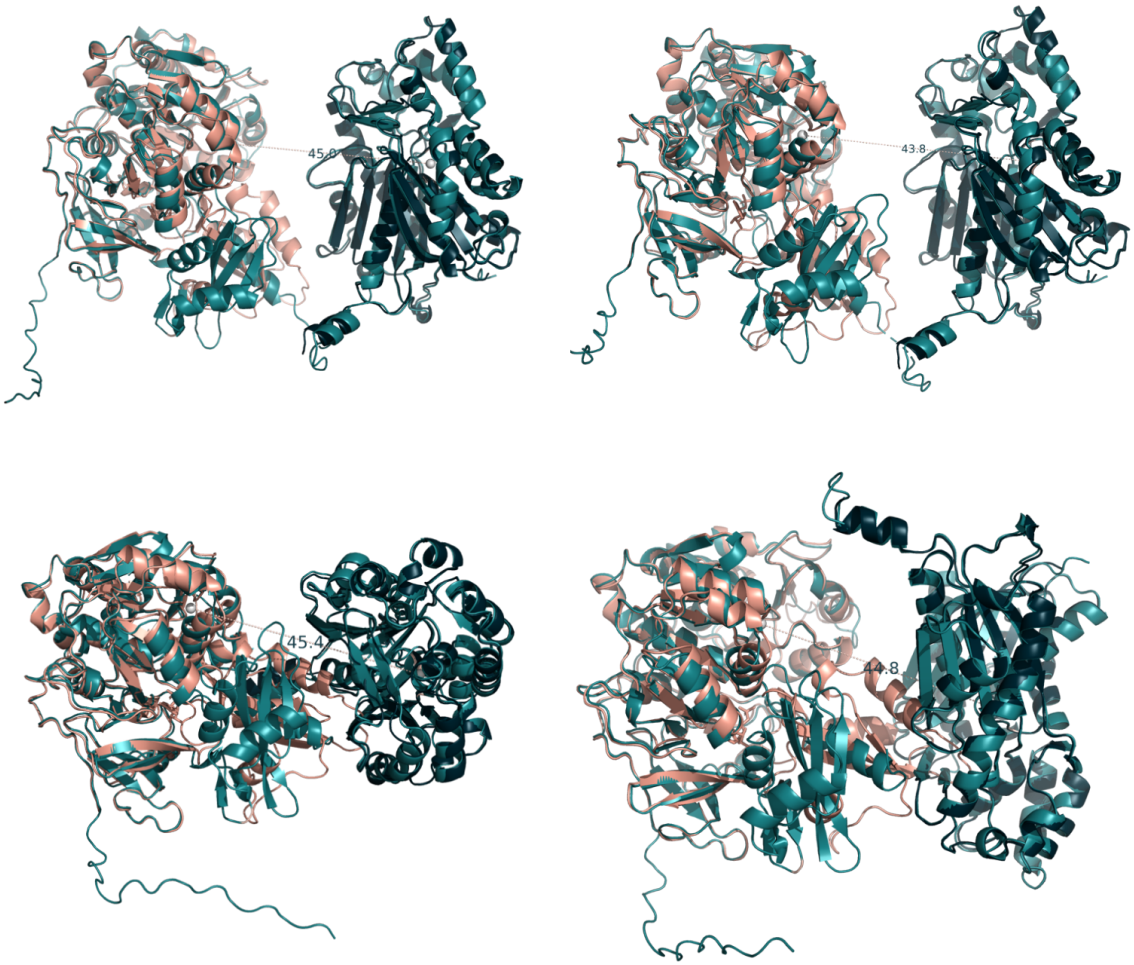

Fig. 57. Fusion protein structures modeled with 3 (upper left), 5 (upper right), 10 (lower left), 15 (lower right) unknown amino acids, 4CL (**pink**), CHS (**dark green**) with their distances