

Integrative Structure Validation Report ?

September 24, 2026 - 12:52 PM PDT

The following software was used in the production of this report:

IHMValidation Version 3.2

Python-IHM Version 2.9

MolProbity Version 4.5.2

PDB ID	9AAU pdb_00009aau
Structure Title	CYP2E-CYP3A4 (proximal side) complex
Structure Authors	Davydov, D.R.; Singh, D.K.; Guha Neogi, A.; Prasad, B.
Deposited on	2026-06-09

This is a PDB-IHM Structure Validation Report.

We welcome your comments at helpdesk@pdb-ihm.org

A user guide is available at https://pdb-ihm.org/validation_help.html with specific help available everywhere you see the ? symbol.

List of references used to build this report is available [here](#).

1. Overview ?

1.1. Summary ?

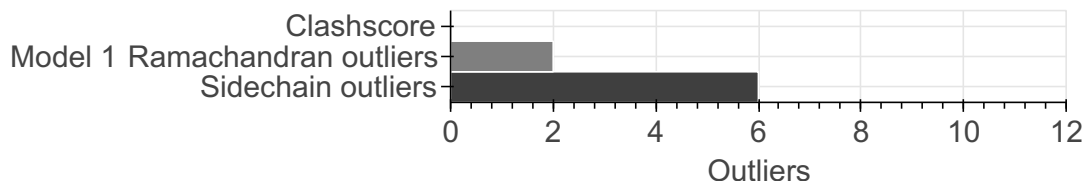
This entry consists of 1 model(s). A total of 3 dataset(s) were used to build this entry.

Name	Type	Count
Crosslinking-MS data	Experimental data	1
De Novo model	Starting model	2

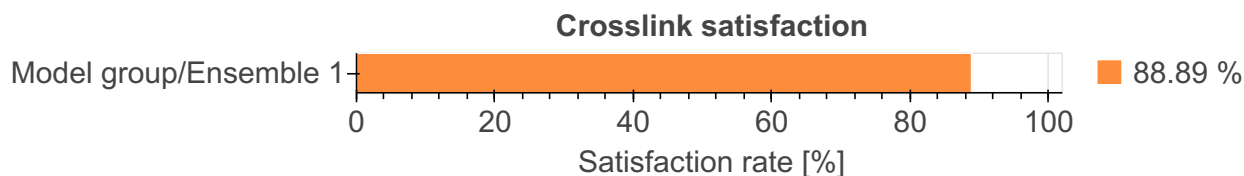
1.2. Overall quality ?

This validation report contains model quality assessments for all structures, data quality and fit to model assessments for SAS and crosslinking-MS datasets. Data quality and fit to model assessments for other datasets and model uncertainty are under development. Number of plots is limited to 256.

Model Quality: MolProbity Analysis ?



Fit to Data Used for Modeling ?



2. Model Details ?

2.1. Ensemble information ?

This entry consists of 0 distinct ensemble(s).

2.2. Representation ?

This entry has 1 representation(s).

ID	Model(s)	Entity ID	Molecule name	Chain(s) [auth]	Total residues	Rigid segments	Flexible segments	Model coverage/ Starting model coverage (%)	Scale
1	1	1	Cytochrome P450 2E1	A	493	1-493	-	100.00 / 100.00	Atomic
		2	Cytochrome P450 3A4	B	503	1-503	-	100.00 / 100.00	Atomic
		3	PROTOPORPHYRIN IX CONTAINING FE	C [A]	Non-polymeric	-	-	Not available / Not available	Atomic
				D [B]					

2.3. Datasets used for modeling ?

There are 3 unique datasets used to build the models in this entry.

ID	Dataset type	Database name	Data access code
1	Crosslinking-MS data	PRIDE	PXD079247
2	De Novo model	AlphaFoldDB	AF-P05181-F1-v6
3	De Novo model	AlphaFoldDB	AF-P08684-F1-v6

2.4. Methodology and software ?

This entry is a result of 1 distinct protocol(s).

Step number	Protocol ID	Method name	Method type	Method description	Number of computed models	Multi state modeling	Multi scale modeling
1	1	Step 1	Manual selection of the initial pose	Selection of the initial pose for proper orientation of both molecules in the membrane and to position some of the surface-exposed cysteines of CYP2E1 in proximity to the TFMD-tagged regions of CYP3A4	Not available	False	False
2	1	Step 2	Restrained docking using HEX 8.0.0	Applied restrains: 15 deg. turning-angle range for both the receptor (CYP2E1) and the ligand (CYP3A4), and a 0 deg. twist range. Docking was based on the combined optimization of shape and DARS potential.	100	False	False
3	1	Step 3	Restrained docking with HADDOCK 2.4	Active residues: CYP2E1 cysteine(s) closest to the contact region and the closest TFMD-tagged residues in CYP3A4. The passive residues were automatically selected within a 12 Å zone around the active residues.	100	False	False
4	1	Step 4	Model refinement with AlphaFold-Multimer	The final, AlphaFold-refined model is characterized with ipTM score of 0.6325	1	False	False

There are 4 software packages reported in this entry.

ID	Software name	Software version	Software classification	Software location
1	UCSF Chimera	1.19	Protein structure manipulation and visualization	https://www.rbvi.ucsf.edu/chimera/
2	HEX 8.0.0	8.0.0	Protein docking	https://hex.loria.fr/
3	HADDOCK	2.40	Protein docking	https://wenmr.science.uu.nl/haddock2.4/
4	Unmasked version of AlphaFold-Multimer	Not available	Protein structure modelling	https://github.com/clami66/AF_unmasked

3. Data quality ?

3.2. Crosslinking-MS

At the moment, data validation is only available for crosslinking-MS data deposited as a fully [compliant](#) dataset in the [PRIDE Crosslinking](#) database. Correspondence between crosslinking-MS and entry entities is established using [pyHMMER](#). Only residue pairs that passed the reported threshold are used for the analysis. The values in the report have to be interpreted in the context of the experiment (i.e. only a minor fraction of in-situ or in-vivo dataset can be used for modeling).

Crosslinking-MS dataset is not available in the [PRIDE Crosslinking](#) database.

4. Model quality ?

For models with atomic structures, MolProbity analysis is performed. For models with coarse-grained or multi-scale structures, excluded volume analysis is performed.

4.1b. MolProbity Analysis ?

Excluded volume satisfaction for the models in the entry are listed below. The Analysed column shows the number of particle-particle or particle-atom pairs for which excluded volume was analysed.

Standard geometry: bond outliers ?

There are 1 bond length outliers in this entry (0.01% of 8362 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
C	1	HEM	C2D-C3D	9.88	1.53	1.33	1	1

Standard geometry: angle outliers ?

There are 28 bond angle outliers in this entry (0.25% of 11356 assessed bonds). A summary is provided below.

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	405	ASP	CA-CB-CG	5.59	118.19	112.60	1	1
B	484	GLN	OE1-CD-NE2	5.19	117.41	122.60	1	1
B	332	GLN	OE1-CD-NE2	4.99	117.61	122.60	1	1
A	249	ARG	NE-CZ-NH2	4.93	123.63	119.20	1	1
A	345	GLN	OE1-CD-NE2	4.87	117.73	122.60	1	1
B	270	ASP	CA-CB-CG	4.67	117.27	112.60	1	1
B	265	GLN	OE1-CD-NE2	4.57	118.03	122.60	1	1
B	201	ASP	CA-CB-CG	4.56	108.04	112.60	1	1
A	400	ASN	OD1-CG-ND2	4.54	118.06	122.60	1	1
A	47	GLN	OE1-CD-NE2	4.50	118.10	122.60	1	1
B	200	GLN	OE1-CD-NE2	4.48	118.12	122.60	1	1
A	255	GLN	OE1-CD-NE2	4.43	118.17	122.60	1	1
A	295	ASP	CA-CB-CG	4.42	117.02	112.60	1	1
B	217	ASP	CA-CB-CG	4.41	117.01	112.60	1	1
A	254	HIS	CB-CG-CD2	4.38	125.50	131.20	1	1
A	337	ARG	NE-CZ-NH1	4.38	125.88	121.50	1	1
A	25	GLN	OE1-CD-NE2	4.30	118.30	122.60	1	1
B	328	GLN	OE1-CD-NE2	4.22	118.38	122.60	1	1
A	216	GLN	OE1-CD-NE2	4.22	118.38	122.60	1	1
B	237	ASN	OD1-CG-ND2	4.19	118.41	122.60	1	1

Chain	Res	Type	Atoms	Z	Observed (Å)	Ideal (Å)	Model ID (Worst)	Models (Total)
A	220	ASN	OD1-CG-ND2	4.18	118.42	122.60	1	1
A	162	GLN	OE1-CD-NE2	4.18	118.42	122.60	1	1
B	151	GLN	OE1-CD-NE2	4.12	118.48	122.60	1	1
B	370	ALA	C-N-CA	4.10	129.09	121.70	1	1
A	238	ASN	CA-CB-CG	4.08	116.68	112.60	1	1
B	159	ASN	OD1-CG-ND2	4.08	118.52	122.60	1	1
A	81	HIS	CB-CG-CD2	4.07	125.91	131.20	1	1
A	141	GLN	OE1-CD-NE2	4.07	118.53	122.60	1	1

Too-close contacts ?

The following all-atom clashscore is based on a MolProbity analysis. All-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The table below contains clashscores for all atomic models in this entry.

Model ID	Clash score	Number of clashes
1	0.00	0

There are no too-close contacts.

Torsion angles: Protein backbone ?

In the following table, Ramachandran outliers are listed. The Analysed column shows the number of residues for which the backbone conformation was analysed.

Model ID	Analysed	Favored	Allowed	Outliers
1	992	964	26	2

There are 2 unique backbone outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
B	6	ASP	1
B	502	GLY	1

Torsion angles : Protein sidechains ?

In the following table, sidechain rotameric outliers are listed. The Analysed column shows the number of residues for which the sidechain conformation was analysed.

Model ID	Analysed	Favored	Allowed	Outliers
1	899	887	6	6

There are 6 unique sidechain outliers. Detailed list of outliers are tabulated below.

Chain	Res	Type	Models (Total)
A	4	LEU	1
A	249	ARG	1
A	310	TYR	1

Chain	Res	Type	Models (Total)
B	3	LEU	1
B	290	LEU	1
B	499	THR	1

5. Fit to Data Used for Modeling Assessment ?

5.2. Crosslinking-MS ?

5.2.1. Restraint types ?

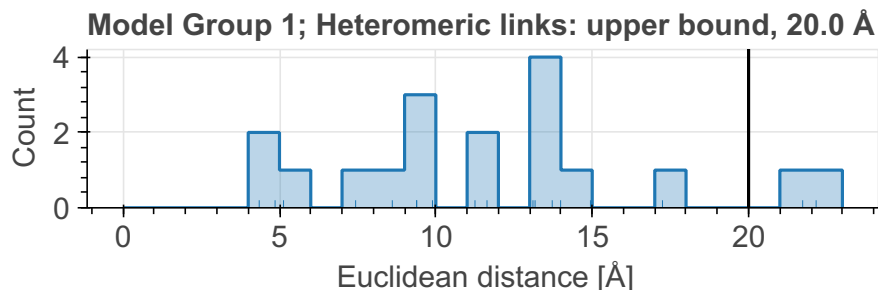
This table summarizes information about crosslinker(s) used for data generation, and how crosslinking information was translated into actual modeling restraints. Restraints assigned "*by-residue*" are interpreted as between CA atoms. Restraints between coarse-grained beads are indicated as "*coarse-grained*". *Restraint group* represents a set of crosslinking restraints applied collectively in the modeling.

There are 18 crosslinking restraints combined in 18 restraint groups.

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
Other	ALA	CA	CYS	CA	upper bound	20.00	1
Other	ARG	CA	CYS	CA	upper bound	20.00	1
Other	CYS	CA	LEU	CA	upper bound	20.00	6
Other	CYS	CA	SER	CA	upper bound	20.00	4
Other	CYS	CA	PHE	CA	upper bound	20.00	2
Other	CYS	CA	THR	CA	upper bound	20.00	2
Other	CYS	CA	LYS	CA	upper bound	20.00	2

Distograms of individual restraints

Distograms (i.e., histogram plots of distances) provide an overview of distributions of distances between residues for which chemical crosslinks were identified. The shift of the distogram relative to the threshold value may indicate a poor model. Restraints with identical thresholds are grouped into one plot. Only the best distance per restraint per model group/ensemble is plotted. Inter- and intramolecular (including self-links) restraints are also grouped into one plot. Distance for a restraint between coarse-grained beads is calculated as a minimal distance between shells; if beads intersect, the distance will be reported as 0.0. A bead with the highest available resolution for a given residue is used for the assessment.



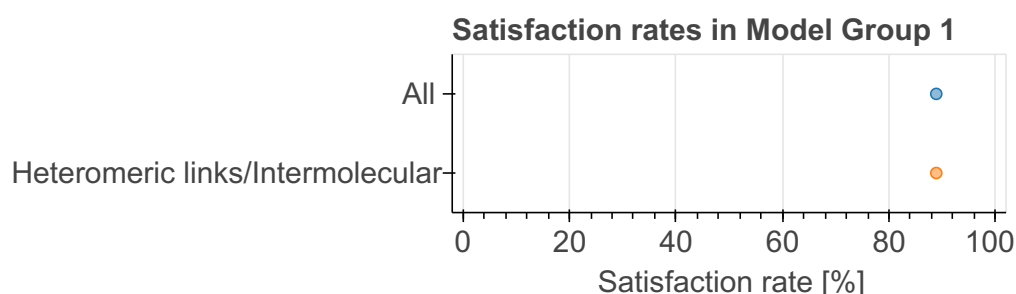
5.2.2. Satisfaction of restraints ?

Satisfaction of restraints is calculated on a **restraint group** (a set of crosslinking restraints applied collectively in the modeling) level. Satisfaction of a restraint group depends on satisfaction of individual restraints in the group and the conditionality (all/any). A restraint group is considered satisfied, if the condition was met in at least one model of the model group/ensemble. The number of measured restraints can be smaller than the total number of restraint groups if crosslinks involve non-modeled residues. Only deposited models are used for validation right now.

State group	State	Model group	# of Deposited models/Total	Restraint group type	Satisfied (%)	Violated (%)	Count (Total=18)
1	1	1	1/1	All	88.89	11.11	18
				Heteromeric links/Intermolecular	88.89	11.11	18

Per-model satisfaction rates in ensembles

Every point represents one model in a model group/ensemble. Where possible, boxplots with quartile marks are also plotted.



6. Fit to Data Used for Validation Assessment

Validation for this section is under development.

Acknowledgments

The development of integrative model validation metrics, implementation of a model validation pipeline, and creation of a validation report for integrative structures are funded by NSF awards to the [PDB-IHM team](#) (DBI-1756248, DBI-2112966, DBI-2112967, DBI-2112968, and DBI-1756250) and awards from NSF, NIH, and DOE to the [RCSB PDB](#) (DBI-2321666, R01GM157729, and DE-SC0019749). The PDB-IHM team and members of the [Sali lab](#) contributed model validation metrics and software packages.

Dr. Jill Trewhella, Dr. Dina Schneidman, and members of the [SASBDB](#) repository are acknowledged for their advice and support in implementing SAS validation methods. Team members from the labs of Dr. Juri Rappsilber, Dr. Alexander Leitner, Dr. Andrea Graziadei, and members of [PRIDE](#) database are acknowledged for their advice and support in implementing crosslinking-MS validation methods. We are grateful to Dr. Shruthi Viswanath for discussions about uncertainty assessment of integrative structural models.

Members of the [wwPDB Integrative/Hybrid Methods Task Force](#) provided recommendations and community support for the project.