

Integrative Structure Validation Report

September 10, 2024 - 11:39 PM PDT

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

Python-IHM Version 1.3


MolProbity Version 4.5.2

Integrative Modeling Validation Version 1.2

PDB ID	9A62
PDB-Dev ID	PDBDEV_00000295
Structure Title	Integrative model of GATA-GATC by crosslinking MS and deep learning
Structure Authors	Kolja Stahl; Oliver Brock; Juri Rappsilber

This is a PDB-Dev IM Structure Validation Report for a publicly released PDB-Dev entry.

We welcome your comments at pdb-dev@mail.wwpdb.org

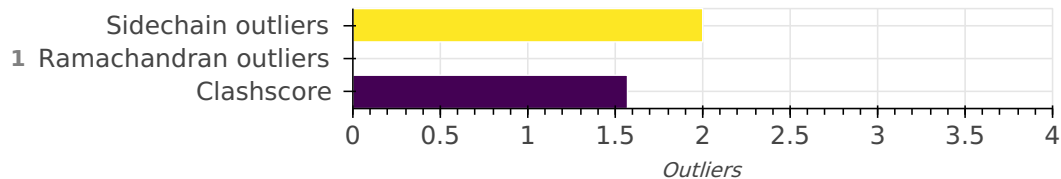
A user guide is available at https://pdb-dev.wwpdb.org/validation_help.html with specific help available everywhere you see the  symbol.

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

Overall quality

This validation report contains model quality assessments for all structures, data quality assessment for SAS datasets and fit to model assessments for SAS 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



Ensemble information ?

This entry consists of 0 distinct ensemble(s).

Summary ?

This entry consists of 1 unique models, with 2 subunits in each model. A total of 1 datasets or restraints were used to build this entry. Each model is represented by 0 rigid bodies and 2 flexible or non-rigid units.

Entry composition ?

There is 1 unique type of models in this entry. This model is titled None/None.

Model ID	Subunit number	Subunit ID	Subunit name	Chain ID	Chain ID [auth]	Total residues
1	1	1	GATA_BACSU	A	A	485
1	2	2	GATC_BACSU	B	B	96

Datasets used for modeling ?

There is 1 unique dataset used to build the models in this entry.

ID	Dataset type	Database name	Data access code
1	Crosslinking-MS data	PRIDE	PXD035508

Representation ?

This entry has only one representation and includes 0 rigid bodies and 2 flexible units.

Chain ID	Rigid bodies	Non-rigid segments

Chain ID	Rigid bodies	Non-rigid segments
A	-	1-485
B	-	1-96

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	AlphaLink2	AlphaLink2	None	1	False	False

There is 1 software package reported in this entry.

ID	Software name	Software version	Software classification	Software location
1	AlphaLink2	1.0	model building	https://github.com/Rappsilber-Laboratory/AlphaLink2

Data quality ?

Crosslinking-MS

Validation for this section is under development.

Model quality ?

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

Standard geometry: bond outliers ?

There are 4465 bond outliers in this entry. A summary is provided below, and a detailed list of outliers can be found [here](#).

Bond type	Observed distance (Å)	Ideal distance (Å)	Number of outliers
CG2--HG22	1.09	0.97	112

Bond type	Observed distance (Å)	Ideal distance (Å)	Number of outliers
CA--HA	1.09	0.97	536
CB--HB2	1.09	0.97	424
CG2--HG21	1.09	0.97	112
NZ--HZ3	1.01	0.89	42
CA--HA3	1.09	0.97	45
CE--HE3	1.09	0.97	54
CG--HG3	1.09	0.97	157
NZ--HZ1	1.01	0.89	42
CB--HB3	1.09	0.97	424
OG1--HG1	0.96	0.84	34
OG--HG	0.96	0.84	40
CG--HG2	1.09	0.97	157
CD2--HD22	1.09	0.97	54
CE--HE1	1.09	0.97	12
CB--HB	1.09	0.97	112
CG2--HG23	1.09	0.97	112
CG1--HG12	1.09	0.97	78
CD1--HD13	1.09	0.97	88
CD2--HD23	1.09	0.97	54
CG1--HG13	1.09	0.97	78
CA--HA2	1.09	0.97	45
CD1--HD12	1.09	0.97	88
NZ--HZ2	1.01	0.89	42

Bond type	Observed distance (Å)	Ideal distance (Å)	Number of outliers
CD1--HD11	1.09	0.97	88
CG--HG	1.09	0.97	54
CD--HD2	1.09	0.97	89
CD2--HD21	1.09	0.97	54
CB--HB1	1.09	0.97	53
CD--HD3	1.09	0.97	89
CE--HE2	1.09	0.97	54
OH--HH	0.96	0.84	21
CG1--HG11	1.09	0.97	44
N--H2	1.01	0.89	2
N--H3	1.01	0.89	2
N--H1	1.01	0.89	2
SG--HG	1.34	1.20	3
N--H	1.01	0.86	555
CE1--HE1	1.08	0.93	53
NH1--HH11	1.01	0.86	23
NE--HE	1.01	0.86	23
CD2--HD2	1.08	0.93	53
NE2--HE21	1.01	0.86	13
CZ--HZ	1.08	0.93	20
NH1--HH12	1.01	0.86	23
CD1--HD1	1.08	0.93	43
ND2--HD22	1.01	0.86	20
NE2--HE22	1.01	0.86	13

Bond type	Observed distance (Å)	Ideal distance (Å)	Number of outliers
NH2--HH21	1.01	0.86	23
CE2--HE2	1.08	0.93	41
ND2--HD21	1.01	0.86	20
CZ3--HZ3	1.08	0.93	2
NH2--HH22	1.01	0.86	23
ND1--HD1	1.01	0.86	11
CZ2--HZ2	1.08	0.93	2
NE2--HE2	1.01	0.86	1
NE1--HE1	1.01	0.86	2
CH2--HH2	1.08	0.93	2
CE3--HE3	1.08	0.93	2

Standard geometry: angle outliers?

There are 10 angle outliers in this entry. A summary is provided below, and a detailed list of outliers can be found [here](#).

Angle type	Observed angle (°)	Ideal angle (°)	Number of outliers
OE1-CD-NE2	122.60	116.35	1
OE1-CD-NE2	122.60	117.93	1
OE1-CD-NE2	122.60	117.96	1
OE1-CD-NE2	122.60	117.97	1
OE1-CD-NE2	122.60	118.21	1
OE1-CD-NE2	122.60	118.35	1
CD-NE-CZ	124.40	130.21	1
OE1-CD-NE2	122.60	118.48	1
OE1-CD-NE2	122.60	118.50	1
OE1-CD-NE2	122.60	118.59	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 the models in this entry.

Model ID	Clash score	Number of clashes
1	1.57	14

All 14 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Model ID	Atom-1	Atom-2	Clash overlap (Å)
1	A:360:MET:HE1	B:28:PHE:CG	0.864
1	A:122:LEU:HD13	A:158:SER:HA	0.851
1	A:303:LYS:HE3	A:304:TYR:CZ	0.717
1	A:360:MET:HE1	B:28:PHE:CD2	0.631
1	A:360:MET:HE1	B:28:PHE:CD1	0.556
1	A:303:LYS:HE3	A:304:TYR:CE1	0.533
1	A:4:PHE:HB3	A:218:ARG:HH22	0.514
1	B:78:MET:HE3	B:87:GLY:O	0.455
1	A:246:PRO:HG2	A:248:PHE:CE1	0.431
1	A:154:SER:H	A:178:SER:HB3	0.425
1	A:303:LYS:CE	A:304:TYR:CZ	0.422
1	A:140:LEU:HD21	A:411:ILE:HG21	0.414
1	A:274:LYS:HE2	A:278:GLU:OE1	0.408
1	A:351:PHE:O	A:356:LYS:HE3	0.406

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	577	559	18	0

Detailed list of outliers are tabulated below.

Torsion angles: Protein sidechains ?

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

Model ID	Analyzed	Favored	Allowed	Outliers
1	483	469	12	2

Detailed list of outliers are tabulated below.

Model ID	Chain	Residue ID	Residue type
1	A	109	GLN
1	B	6	ILE

Fit of model to data used for modeling ?

Crosslinking-MS

Validation for this section is under development.

Fit of model to data used for validation ?

Validation for this section is under development.

Acknowledgements

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 ABI awards (DBI-1756248, DBI-2112966, DBI-2112967, DBI-2112968, and DBI-1756250). The [PDB-Dev team](#) and members of [Sali lab](#) contributed model validation metrics and software packages.

Implementation of validation methods for SAS data and SAS-based models are funded by [RCSB PDB](#) (grant number DBI-1832184). Dr. Stephen Burley, Dr. John Westbrook, and Dr. Jasmine Young from [RCSB PDB](#), Dr. Jill Trehwella, Dr. Dina Schneidman, and members of the [SASBDB](#) repository are acknowledged for their advice and support in implementing SAS validation methods.

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