# Integrative Structure Validation Report ?

## February 18, 2025 - 08:44 AM PST

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

*Integrative Modeling Validation Version 2.0*
*Python-IHM Version 1.8*
*MolProbity Version 4.5.2*
*pyHMMER Version 0.11.0*

| | |
|---|---|
| PDB ID | 9A7C |
| PDB-Dev ID | PDBDEV_00000341 |
| Structure Title | Integrative model of NUSA-RPOB by crosslinking MS and deep learning |
| Structure Authors | Kolja Stahl; Oliver Brock; Juri Rappsilber |
| Deposited on | 2024-01-23 |

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

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

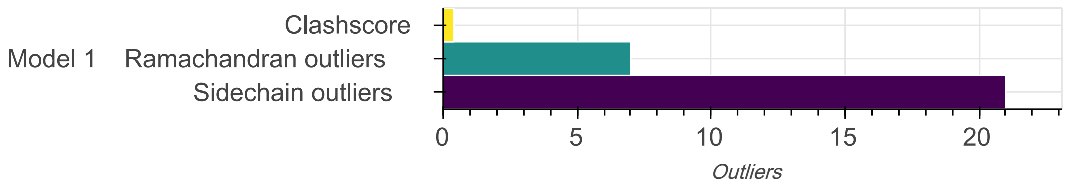
*A user guide is available at https://pdb-ihm.org/validation_help.html with specific help available everywhere you see the* ? symbol.

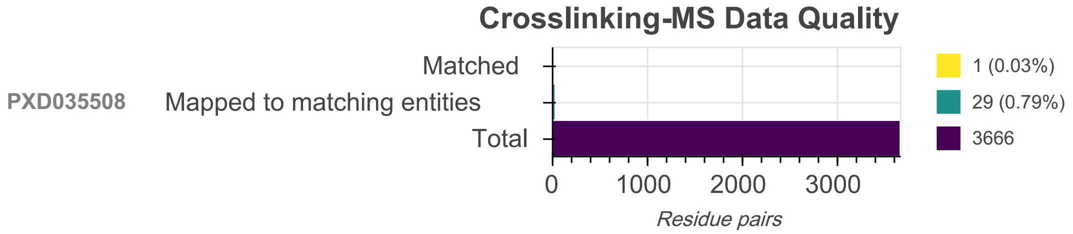
*List of references used to build this report is available here.*

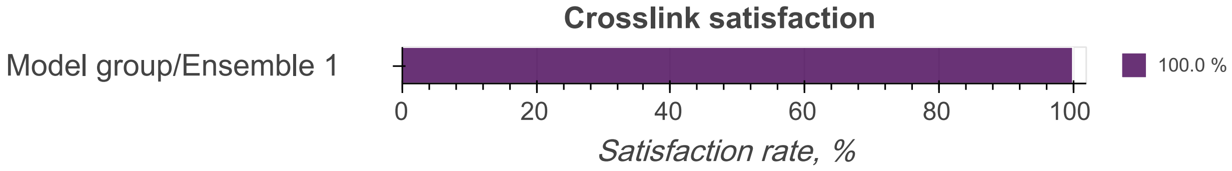
## Overall quality ?

*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







## Ensemble information

*This entry consists of 0 distinct ensemble(s).*

## Summary

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

## 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 | NUSA_BACSU | A | 371 | - | 1-371 | 100.00 / 0.00 | Atomic |
| | | 2 | RPOB_BACSU | B | 1193 | - | 1-1193 | 100.00 / 0.00 | Atomic |

## 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 |

## 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 | 1.00 | model building | https://github.com/Rappsilber-Laboratory/AlphaLink2 |

---

## Data quality ?

### 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 (PRIDE ID) | PXD035508 |
| Number of entities in the crosslinking-MS dataset: | 810 |
| Number of entities in the entry: | 2 |
| Matching entities: | |

| Entity ID | Molecule name | Crosslinking-MS Entity ID | E-value | Exact match |
|---|---|---|---|---|
| 1 | NUSA_BACSU | dbseq_P32727_target | 0.00 | True |
| 2 | RPOB_BACSU | dbseq_P37870_target | 0.00 | True |

Residue pairs stats:

| Source | Total | In matched entities | Total matched |
|---|---|---|---|
| 9A7C | 1 | 1 (100.00%) | 1 (100.00%) |
| PXD035508 | 3666 | 29 (0.79%) | 1 (0.03% ) |

---

## 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.

### Standard geometry: bond outliers ?

*There are 1 bond length outliers in this entry (0.01% of 12540 assessed bonds). A summary is provided below.*

| Chain | Res | Type | Atoms | \|Z\| | Observed (Å) | Ideal (Å) | Model ID (Worst) | Models (Total) |
|---|---|---|---|---|---|---|---|---|
| B | 1186 | GLU | CA-C | 4.07 | 1.61 | 1.52 | 1 | 1 |

### Standard geometry: angle outliers ?

*There are 88 bond angle outliers in this entry (0.52% of 16958 assessed bonds). A summary is provided below.*

| Chain | Res | Type | Atoms | \|Z\| | Observed (Å) | Ideal (Å) | Model ID (Worst) | Models (Total) |
|---|---|---|---|---|---|---|---|---|
| B | 1185 | VAL | C-N-CA | 12.72 | 144.59 | 121.70 | 1 | 1 |
| B | 1182 | SER | C-N-CA | 10.89 | 141.31 | 121.70 | 1 | 1 |
| B | 1188 | ASP | C-N-CA | 10.73 | 141.02 | 121.70 | 1 | 1 |
| B | 1183 | ALA | C-N-CA | 10.60 | 140.77 | 121.70 | 1 | 1 |
| B | 1186 | GLU | C-N-CA | 9.74 | 139.24 | 121.70 | 1 | 1 |
| B | 1186 | GLU | CA-C-N | 8.87 | 133.94 | 116.20 | 1 | 1 |
| B | 1185 | VAL | CA-C-N | 8.53 | 133.26 | 116.20 | 1 | 1 |
| B | 1183 | ALA | CA-C-N | 7.80 | 131.80 | 116.20 | 1 | 1 |
| B | 1185 | VAL | O-C-N | 7.66 | 110.74 | 123.00 | 1 | 1 |
| B | 1188 | ASP | CA-C-N | 7.61 | 131.41 | 116.20 | 1 | 1 |
| B | 1186 | GLU | O-C-N | 7.35 | 111.24 | 123.00 | 1 | 1 |

| Chain | Res | Type | Atoms | \|Z\| | Observed (Å) | Ideal (Å) | Model ID (Worst) | Models (Total) |
|---|---|---|---|---|---|---|---|---|
| B | 1179 | GLU | C-N-CA | 6.88 | 134.08 | 121.70 | 1 | 1 |
| A | 362 | THR | C-N-CA | 6.71 | 133.78 | 121.70 | 1 | 1 |
| B | 1183 | ALA | O-C-N | 6.58 | 112.47 | 123.00 | 1 | 1 |
| B | 1188 | ASP | O-C-N | 6.54 | 112.53 | 123.00 | 1 | 1 |
| B | 280 | HIS | CB-CG-CD2 | 6.04 | 123.35 | 131.20 | 1 | 1 |
| B | 1189 | VAL | CA-CB-CG2 | 6.04 | 120.66 | 110.40 | 1 | 1 |
| A | 312 | GLN | OE1-CD-NE2 | 5.39 | 117.21 | 122.60 | 1 | 1 |
| B | 403 | HIS | CB-CG-CD2 | 5.22 | 124.41 | 131.20 | 1 | 1 |
| B | 773 | ASP | CA-CB-CG | 5.10 | 117.70 | 112.60 | 1 | 1 |
| B | 175 | ASP | CA-CB-CG | 5.05 | 117.65 | 112.60 | 1 | 1 |
| B | 128 | ASN | CA-CB-CG | 5.00 | 117.60 | 112.60 | 1 | 1 |
| B | 438 | GLN | OE1-CD-NE2 | 4.94 | 117.66 | 122.60 | 1 | 1 |
| B | 1182 | SER | CA-C-N | 4.89 | 125.98 | 116.20 | 1 | 1 |
| B | 390 | ASN | CA-CB-CG | 4.87 | 117.47 | 112.60 | 1 | 1 |
| B | 967 | ASN | OD1-CG-ND2 | 4.84 | 117.76 | 122.60 | 1 | 1 |
| B | 355 | GLN | OE1-CD-NE2 | 4.81 | 117.79 | 122.60 | 1 | 1 |
| B | 1011 | ARG | CD-NE-CZ | 4.73 | 131.02 | 124.40 | 1 | 1 |
| B | 1190 | VAL | N-CA-C | 4.72 | 124.21 | 111.00 | 1 | 1 |
| B | 1182 | SER | O-C-N | 4.69 | 115.50 | 123.00 | 1 | 1 |
| A | 41 | GLN | OE1-CD-NE2 | 4.67 | 117.93 | 122.60 | 1 | 1 |
| B | 701 | GLN | OE1-CD-NE2 | 4.63 | 117.97 | 122.60 | 1 | 1 |
| A | 271 | ASP | CA-CB-CG | 4.62 | 117.22 | 112.60 | 1 | 1 |
| B | 1180 | THR | CA-CB-OG1 | 4.62 | 116.53 | 109.60 | 1 | 1 |
| B | 426 | ARG | NH1-CZ-NH2 | 4.61 | 113.31 | 119.30 | 1 | 1 |
| B | 469 | GLN | OE1-CD-NE2 | 4.60 | 118.00 | 122.60 | 1 | 1 |
| B | 280 | HIS | CB-CG-ND1 | 4.60 | 129.60 | 122.70 | 1 | 1 |
| B | 576 | GLN | OE1-CD-NE2 | 4.60 | 118.00 | 122.60 | 1 | 1 |
| A | 84 | HIS | CB-CG-CD2 | 4.58 | 125.25 | 131.20 | 1 | 1 |
| A | 193 | GLN | OE1-CD-NE2 | 4.54 | 118.06 | 122.60 | 1 | 1 |
| B | 657 | PHE | CA-CB-CG | 4.53 | 118.33 | 113.80 | 1 | 1 |
| B | 1191 | THR | N-CA-C | 4.51 | 123.64 | 111.00 | 1 | 1 |
| B | 1101 | ASP | CA-CB-CG | 4.50 | 117.10 | 112.60 | 1 | 1 |
| A | 359 | PRO | C-N-CA | 4.47 | 129.74 | 121.70 | 1 | 1 |
| B | 50 | GLN | OE1-CD-NE2 | 4.44 | 118.16 | 122.60 | 1 | 1 |
| B | 1179 | GLU | O-C-N | 4.43 | 115.91 | 123.00 | 1 | 1 |

| Chain | Res | Type | Atoms | \|Z\| | Observed (Å) | Ideal (Å) | Model ID (Worst) | Models (Total) |
|---|---|---|---|---|---|---|---|---|
| A | 43 | GLN | OE1-CD-NE2 | 4.43 | 118.17 | 122.60 | 1 | 1 |
| B | 308 | GLN | OE1-CD-NE2 | 4.42 | 118.18 | 122.60 | 1 | 1 |
| B | 1165 | GLN | OE1-CD-NE2 | 4.41 | 118.19 | 122.60 | 1 | 1 |
| B | 538 | ASN | CA-CB-CG | 4.40 | 117.00 | 112.60 | 1 | 1 |
| B | 791 | HIS | CB-CG-CD2 | 4.38 | 125.50 | 131.20 | 1 | 1 |
| B | 359 | GLN | OE1-CD-NE2 | 4.36 | 118.24 | 122.60 | 1 | 1 |
| B | 280 | HIS | CA-CB-CG | 4.36 | 118.16 | 113.80 | 1 | 1 |
| B | 1042 | HIS | CB-CG-CD2 | 4.35 | 125.54 | 131.20 | 1 | 1 |
| B | 1184 | ASP | C-N-CA | 4.35 | 129.54 | 121.70 | 1 | 1 |
| B | 507 | HIS | CB-CG-CD2 | 4.34 | 125.56 | 131.20 | 1 | 1 |
| A | 361 | PHE | CA-CB-CG | 4.33 | 118.13 | 113.80 | 1 | 1 |
| A | 200 | HIS | CB-CG-CD2 | 4.32 | 125.58 | 131.20 | 1 | 1 |
| A | 257 | ARG | NH1-CZ-NH2 | 4.32 | 113.68 | 119.30 | 1 | 1 |
| B | 582 | ASP | CA-CB-CG | 4.31 | 116.91 | 112.60 | 1 | 1 |
| B | 1187 | ARG | CA-CB-CG | 4.31 | 105.49 | 114.10 | 1 | 1 |
| B | 419 | GLN | OE1-CD-NE2 | 4.26 | 118.34 | 122.60 | 1 | 1 |
| B | 121 | ASP | CA-CB-CG | 4.25 | 116.85 | 112.60 | 1 | 1 |
| B | 12 | ARG | CD-NE-CZ | 4.25 | 130.35 | 124.40 | 1 | 1 |
| B | 51 | ASP | CA-CB-CG | 4.24 | 116.84 | 112.60 | 1 | 1 |
| A | 259 | GLN | OE1-CD-NE2 | 4.23 | 118.37 | 122.60 | 1 | 1 |
| A | 256 | GLN | OE1-CD-NE2 | 4.22 | 118.38 | 122.60 | 1 | 1 |
| A | 175 | HIS | CB-CG-CD2 | 4.21 | 125.72 | 131.20 | 1 | 1 |
| B | 652 | GLN | OE1-CD-NE2 | 4.21 | 118.39 | 122.60 | 1 | 1 |
| B | 375 | ASN | OD1-CG-ND2 | 4.21 | 118.39 | 122.60 | 1 | 1 |
| A | 143 | GLN | OE1-CD-NE2 | 4.20 | 118.40 | 122.60 | 1 | 1 |
| B | 402 | ASP | CA-CB-CG | 4.20 | 108.40 | 112.60 | 1 | 1 |
| B | 689 | ASN | CA-CB-CG | 4.20 | 116.80 | 112.60 | 1 | 1 |
| B | 825 | ASP | CA-CB-CG | 4.19 | 116.79 | 112.60 | 1 | 1 |
| B | 362 | ASN | CA-CB-CG | 4.18 | 116.78 | 112.60 | 1 | 1 |
| B | 28 | ASN | OD1-CG-ND2 | 4.17 | 118.43 | 122.60 | 1 | 1 |
| B | 13 | GLN | OE1-CD-NE2 | 4.17 | 118.43 | 122.60 | 1 | 1 |
| B | 11 | HIS | CB-CG-CD2 | 4.16 | 125.79 | 131.20 | 1 | 1 |
| B | 446 | GLN | OE1-CD-NE2 | 4.12 | 118.48 | 122.60 | 1 | 1 |
| B | 1049 | HIS | CB-CG-CD2 | 4.11 | 125.86 | 131.20 | 1 | 1 |
| A | 112 | GLN | OE1-CD-NE2 | 4.07 | 118.53 | 122.60 | 1 | 1 |

| Chain | Res | Type | Atoms | \|Z\| | Observed (Å) | Ideal (Å) | Model ID (Worst) | Models (Total) |
|---|---|---|---|---|---|---|---|---|
| A | 116 | GLN | OE1-CD-NE2 | 4.05 | 118.55 | 122.60 | 1 | 1 |
| A | 108 | GLN | OE1-CD-NE2 | 4.04 | 118.56 | 122.60 | 1 | 1 |
| B | 1069 | GLN | OE1-CD-NE2 | 4.04 | 118.56 | 122.60 | 1 | 1 |
| B | 578 | ASN | OD1-CG-ND2 | 4.03 | 118.57 | 122.60 | 1 | 1 |
| B | 606 | ASN | CA-CB-CG | 4.02 | 116.62 | 112.60 | 1 | 1 |
| B | 720 | GLN | OE1-CD-NE2 | 4.00 | 118.60 | 122.60 | 1 | 1 |
| B | 10 | ARG | NE-CZ-NH1 | 4.00 | 125.50 | 121.50 | 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 | 0.41 | 10 |

*There are 10 clashes. The table below contains the detailed list of all clashes based on a MolProbity analysis. Bad clashes are >= 0.4 Angstrom.*

| Atom 1 | Atom 2 | Clash(Å) | Model ID (Worst) | Models (Total) |
|---|---|---|---|---|
| B:1188:ASP:HB3 | B:1189:VAL:HG22 | 0.55 | 1 | 1 |
| B:791:HIS:CE1 | B:1043:MET:HE1 | 0.50 | 1 | 1 |
| B:471:MET:HE1 | B:483:LYS:HE3 | 0.48 | 1 | 1 |
| B:777:MET:HG3 | B:781:LEU:HD12 | 0.45 | 1 | 1 |
| B:781:LEU:HD11 | B:954:ASP:HB3 | 0.44 | 1 | 1 |
| B:849:LYS:HE2 | B:852:THR:HG21 | 0.43 | 1 | 1 |
| B:928:ARG:HH21 | B:973:GLU:CD | 0.42 | 1 | 1 |
| B:976:MET:HE1 | B:987:ILE:HG21 | 0.42 | 1 | 1 |
| A:245:ASP:CG | A:248:GLY:H | 0.42 | 1 | 1 |
| A:232:ARG:HH21 | A:271:ASP:CG | 0.41 | 1 | 1 |

## 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 | 1560 | 1495 | 58 | 7 |

*There are 7 unique backbone outliers. Detailed list of outliers are tabulated below.*

| Chain | Res | Type | Models (Total) |
|---|---|---|---|
| A | 364 | PRO | 1 |

| Chain | Res | Type | Models (Total) |
|---|---|---|---|
| B | 268 | ALA | 1 |
| B | 1041 | ALA | 1 |
| B | 1061 | GLN | 1 |
| B | 1187 | ARG | 1 |
| B | 1189 | VAL | 1 |
| B | 1191 | THR | 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 | 1352 | 1289 | 42 | 21 |

*There are 21 unique sidechain outliers. Detailed list of outliers are tabulated below.*

| Chain | Res | Type | Models (Total) |
|---|---|---|---|
| A | 98 | THR | 1 |
| A | 348 | ILE | 1 |
| B | 48 | MET | 1 |
| B | 57 | ASP | 1 |
| B | 159 | THR | 1 |
| B | 223 | LYS | 1 |
| B | 311 | ASP | 1 |
| B | 319 | LEU | 1 |
| B | 519 | THR | 1 |
| B | 753 | LEU | 1 |
| B | 802 | THR | 1 |
| B | 976 | MET | 1 |
| B | 1011 | ARG | 1 |
| B | 1044 | VAL | 1 |
| B | 1169 | LEU | 1 |
| B | 1179 | GLU | 1 |
| B | 1180 | THR | 1 |
| B | 1185 | VAL | 1 |
| B | 1188 | ASP | 1 |
| B | 1189 | VAL | 1 |

| Chain | Res | Type | Models (Total) |
|---|---|---|---|
| B | 1191 | THR | 1 |

# Fit of model to data used for modeling ?

## Fit of model(s) to crosslinking-MS data

### Restraint types

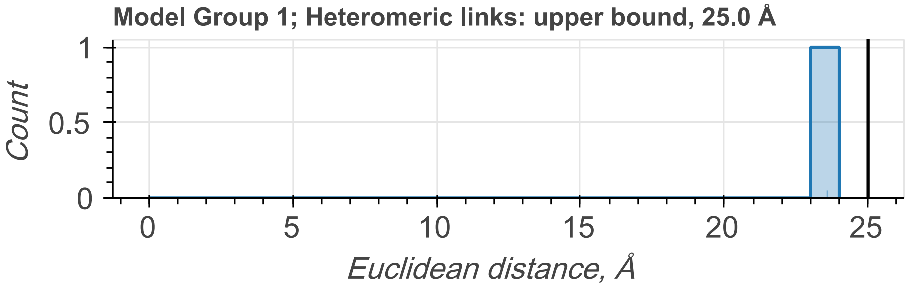
*Restraint types are summarized in the table below. Restraints assigned "by-residue" are interpreted as between CA atoms. Restraints between coarse-grained beads are indicated as "coarse-grained". Restraint group represents a set of crosslinking restraints applied collectively in the modeling.*

There are 1 crosslinking restraints combined in 1 restraint groups.

| Linker | Residue 1 | Atom 1 | Residue 2 | Atom 2 | Restraint type | Distance, Å | Count |
|---|---|---|---|---|---|---|---|
| SDA | LYS | CA | LYS | CA | upper bound | 25.0 | 1 |

### Distograms of individual restraints

*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.*



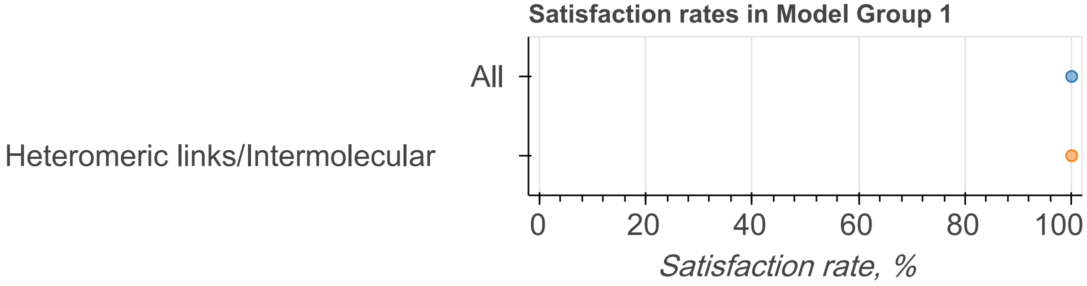
### Satisfaction of restraints

*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=1) |
|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1/1 | All | 100.00 | 0.00 | 1 |
| | | | | Heteromeric links/ Intermolecular | 100.00 | 0.00 | 1 |

### 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.*



## Fit of model to data used for validation

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.*