

# Integrative Structure Validation Report ?

February 27, 2025 - 12:19 PM PST

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

*Integrative Modeling Validation Version 2.0*

*Python-IHM Version 1.8*

*PyMOL Version 2.5.0*

PDB ID	9A3Q
PDB-Dev ID	PDBDEV_00000211
Structure Title	Modeling hLINE1 ORF2p
Structure Authors	Baldwin, E.T.; van Eeuwen, T.; Hoyos, T.; Zalevsky, A.; Tchesnokov, E.P; Sanchez, R.; DiStefano, L.; Ruiz, F.X; Hancock, M.; Walpole, T.; Nichols, C.; Wan, P.; Riento, K.; Kass, R.-H.; Augustin, M.; Lammens, A.; Jestel, A.; Upla, P.; Xibinaku, K.; Congreve, S.; Hennink, M.; Rogala, K.B.; Schneider, A.M.; Fairman, J.E.; Christensen, S.M.; Miao, W.; Zaller, D.M.; Sali, A.; Weichenrieder, O.; Burns, K.H.; Gotte, M.; Rout, M.P.; Arnold, E.; Greenbaum, B.D.; Romero, D.L.; LaCava, J.; Taylor, M.S.
Deposited on	2023-05-25

*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](mailto:helpdesk@pdb-ihm.org)*

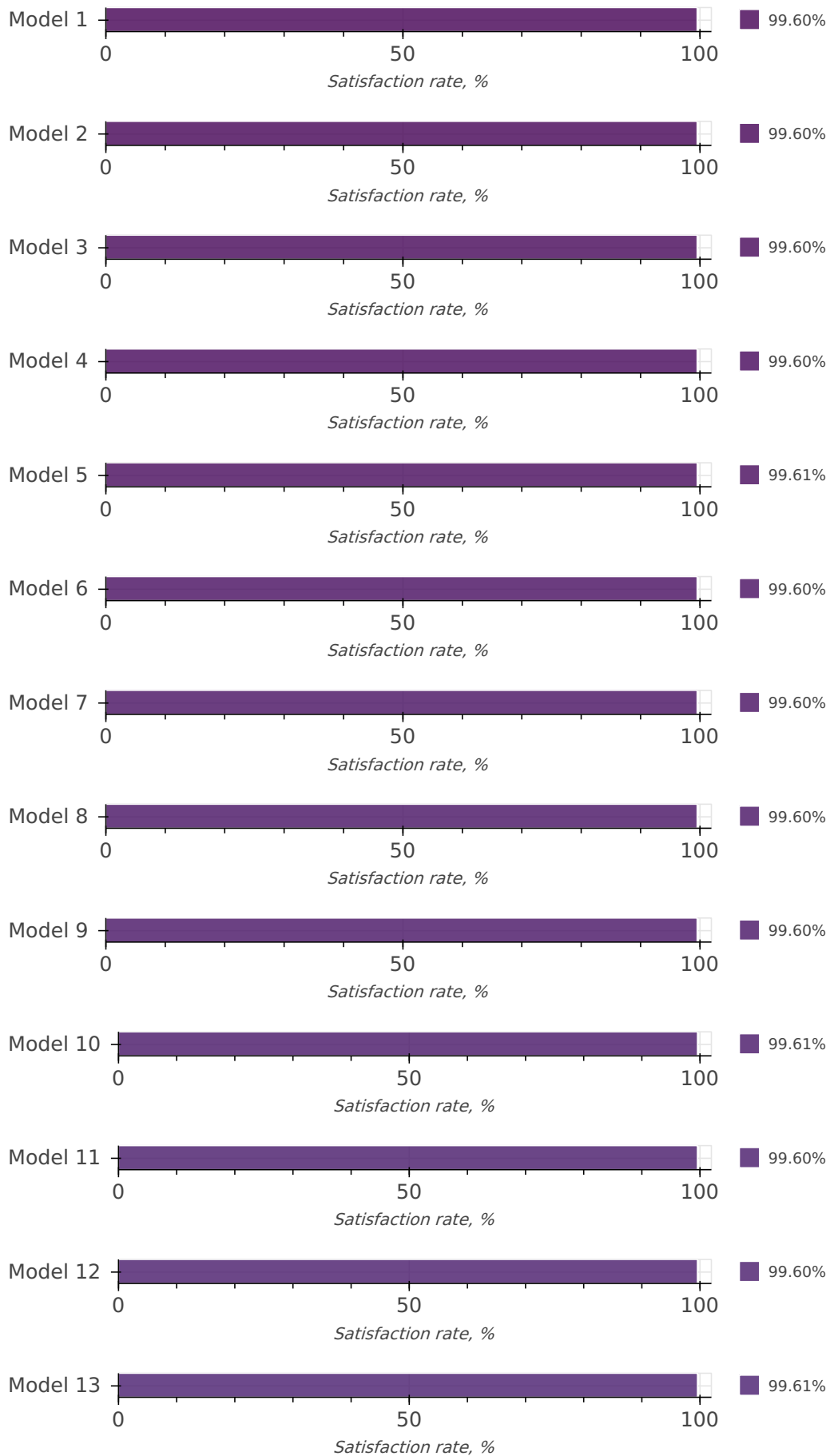
*A user guide is available at [https://pdb-ihm.org/validation\\_help.html](https://pdb-ihm.org/validation_help.html) with specific help available everywhere you see the ? symbol.*

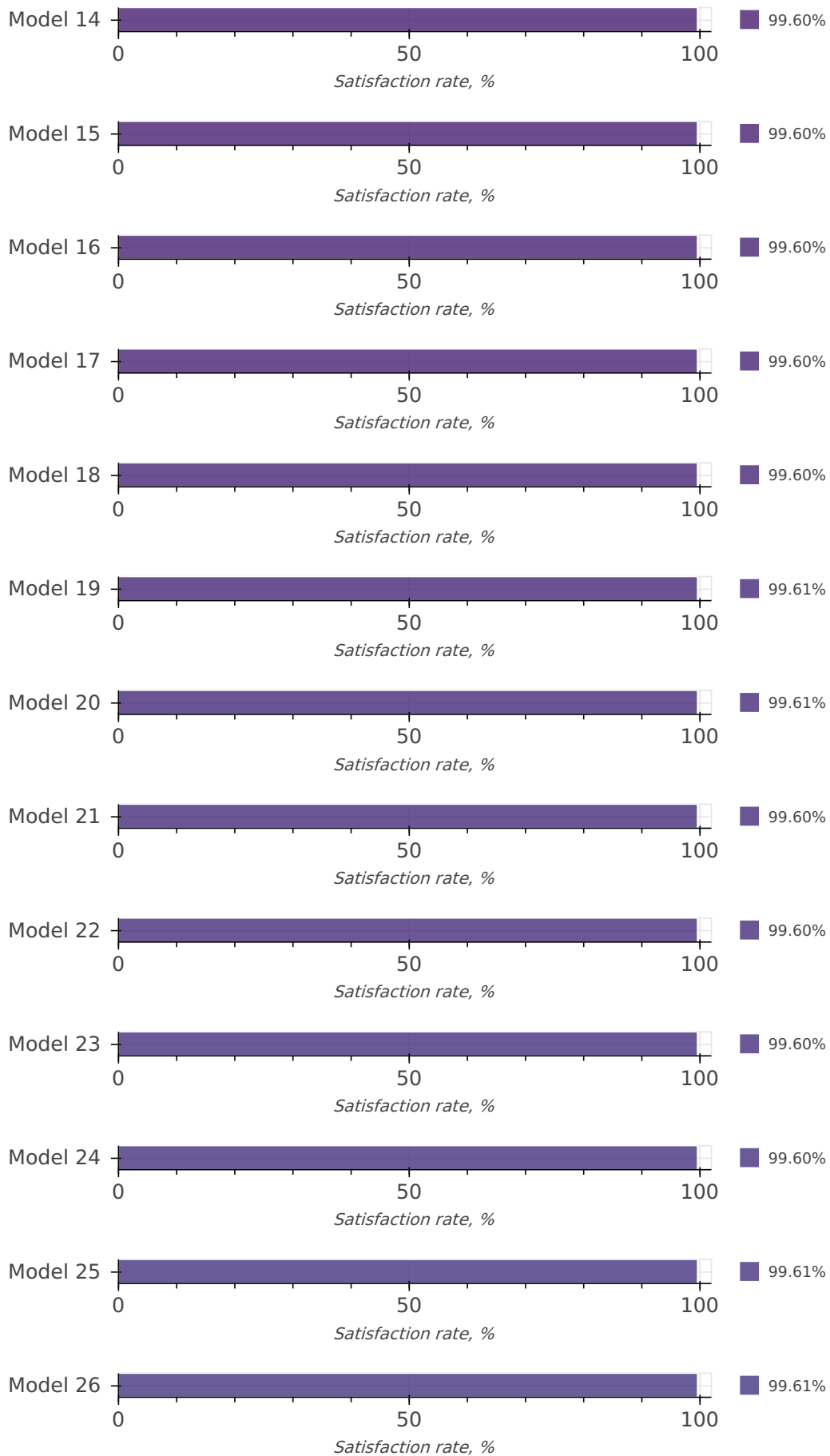
*List of references used to build this report is available [here](#).*

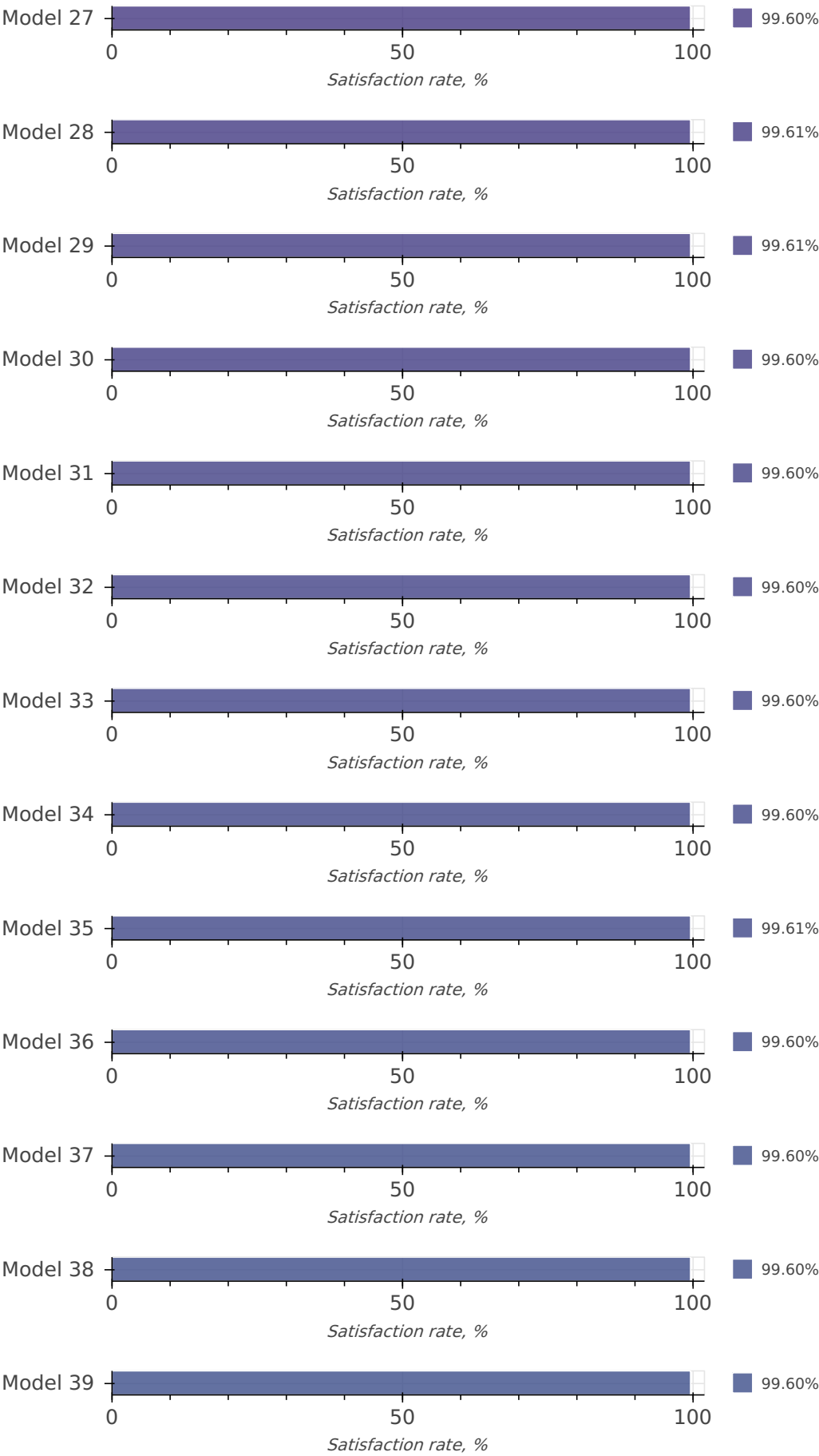
## Overall quality ?

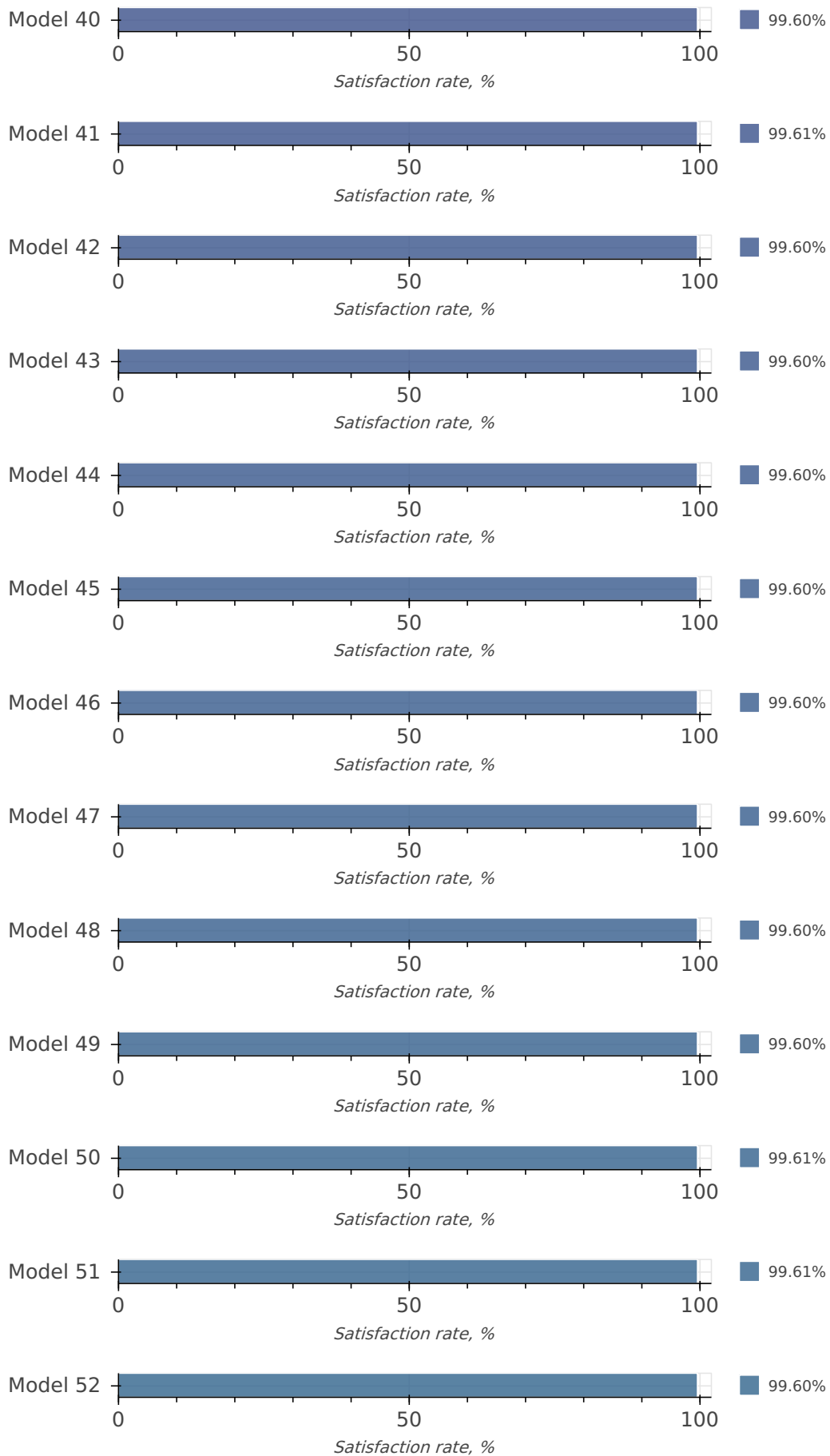
*This validation report contains model quality assessments for all structures, data quality and fit to model assessments for SAS and crosslinking-MS datasets. Data quality and fit to model assessments for other datasets and model uncertainty are under development. Number of plots is limited to 256.*

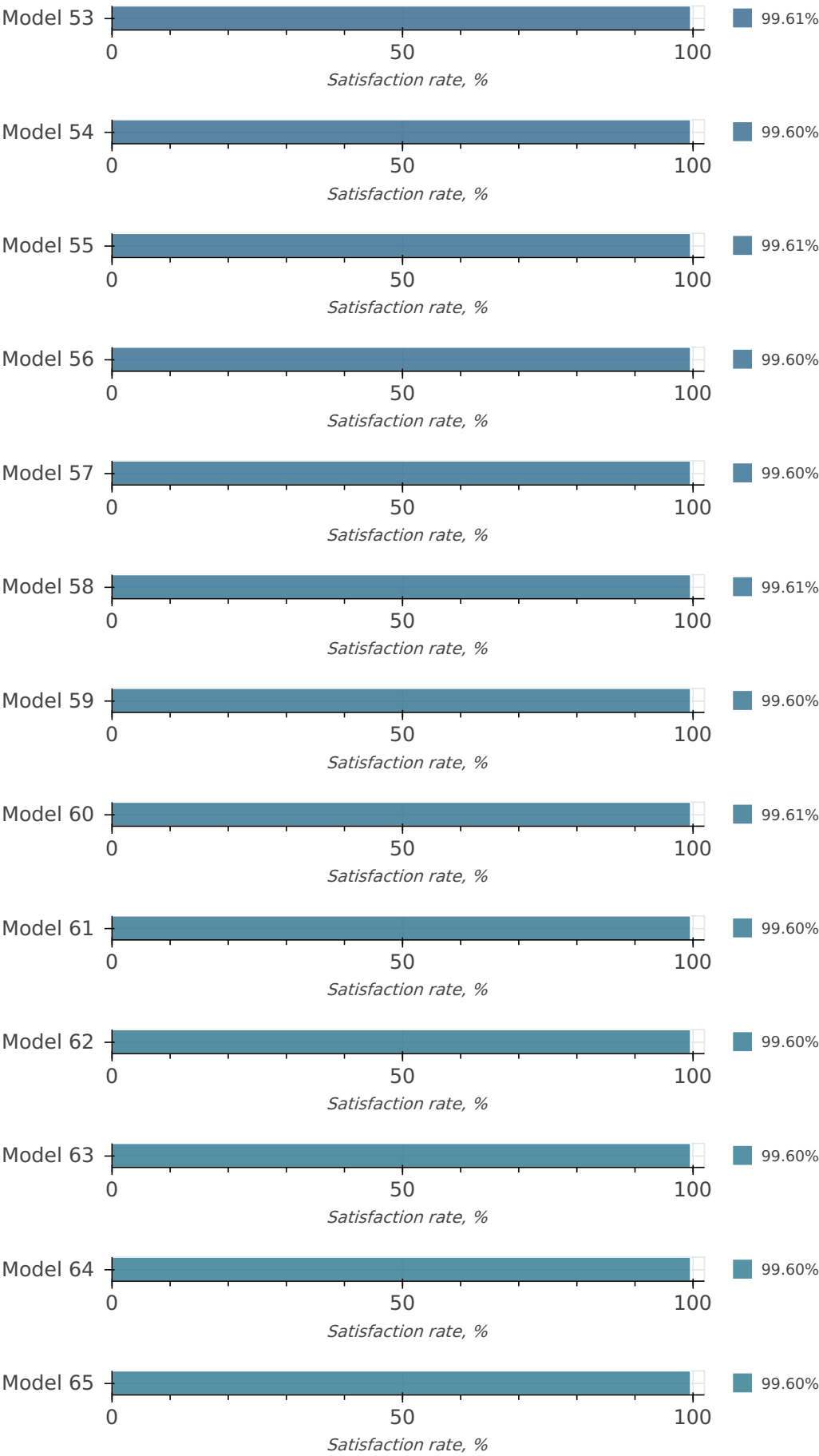
Model Quality: Excluded Volume Analysis

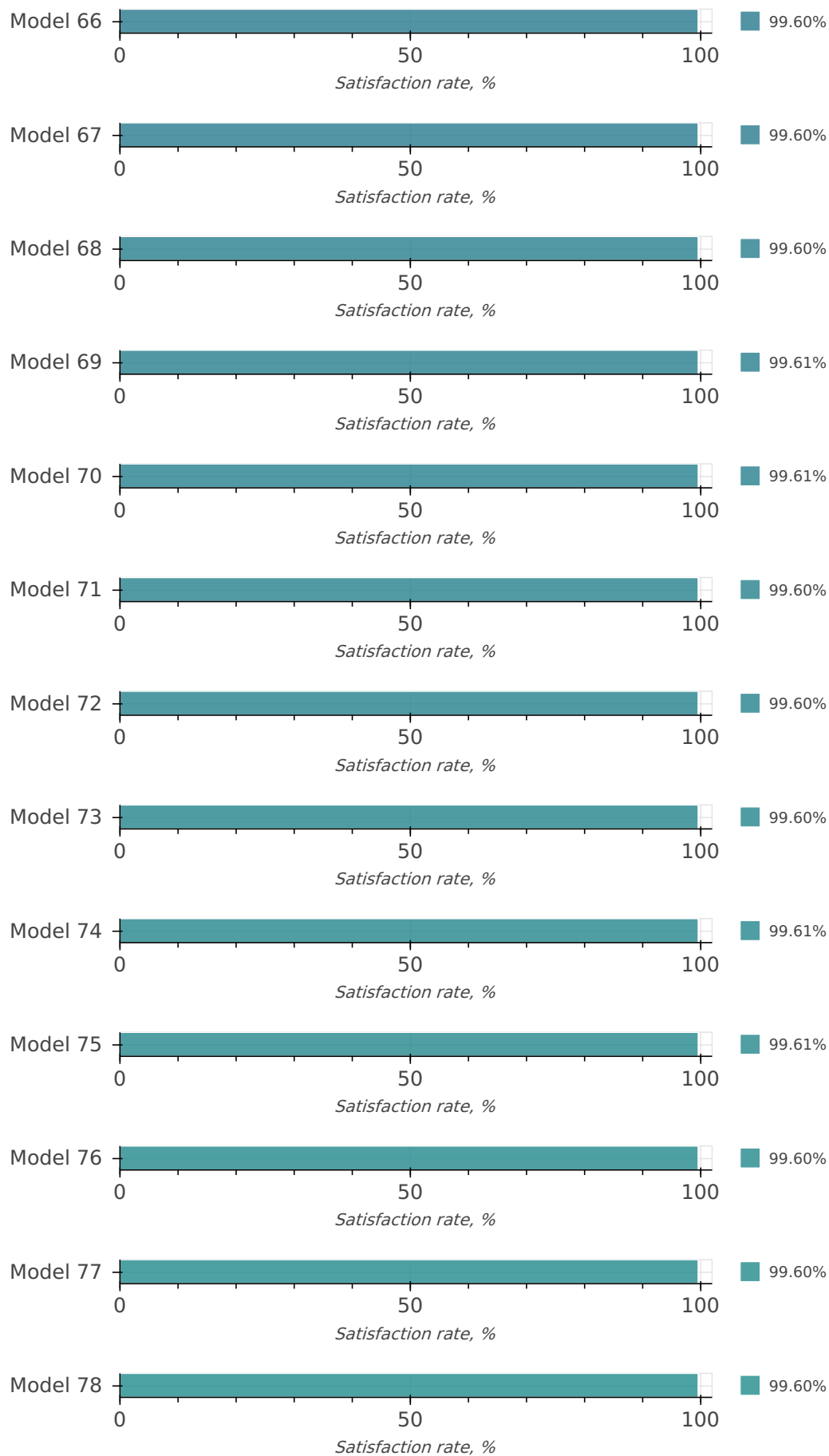


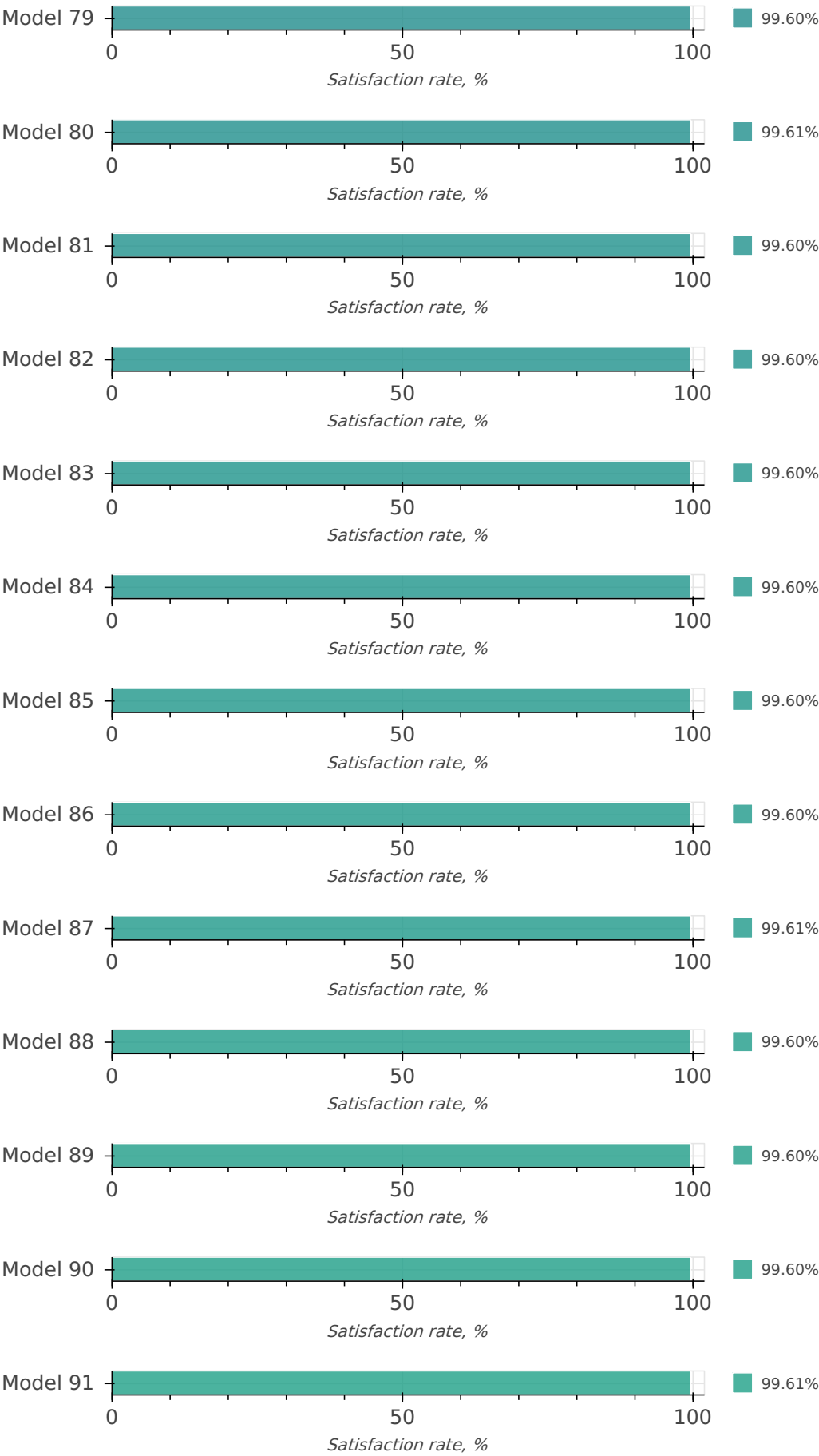




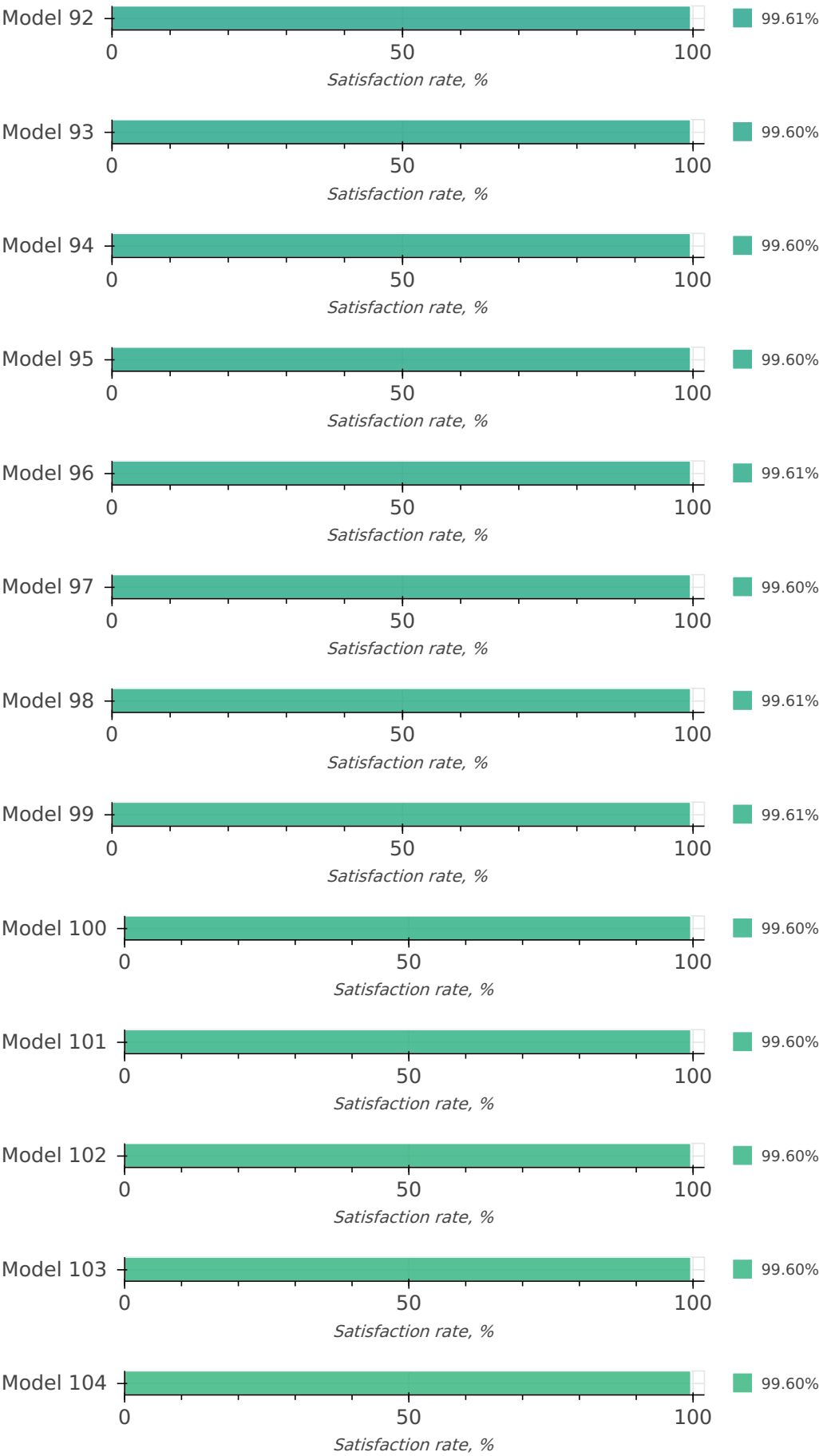


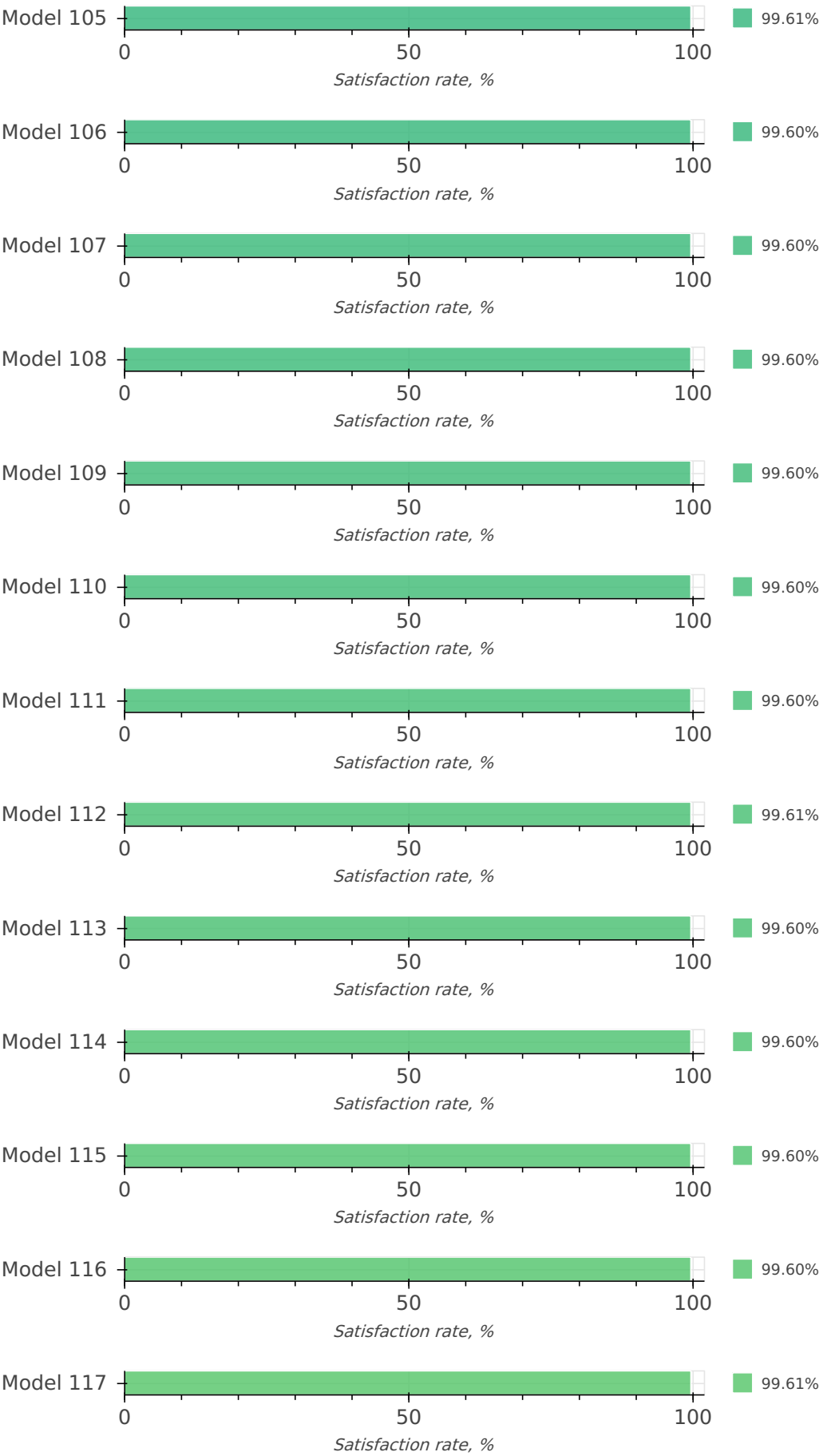


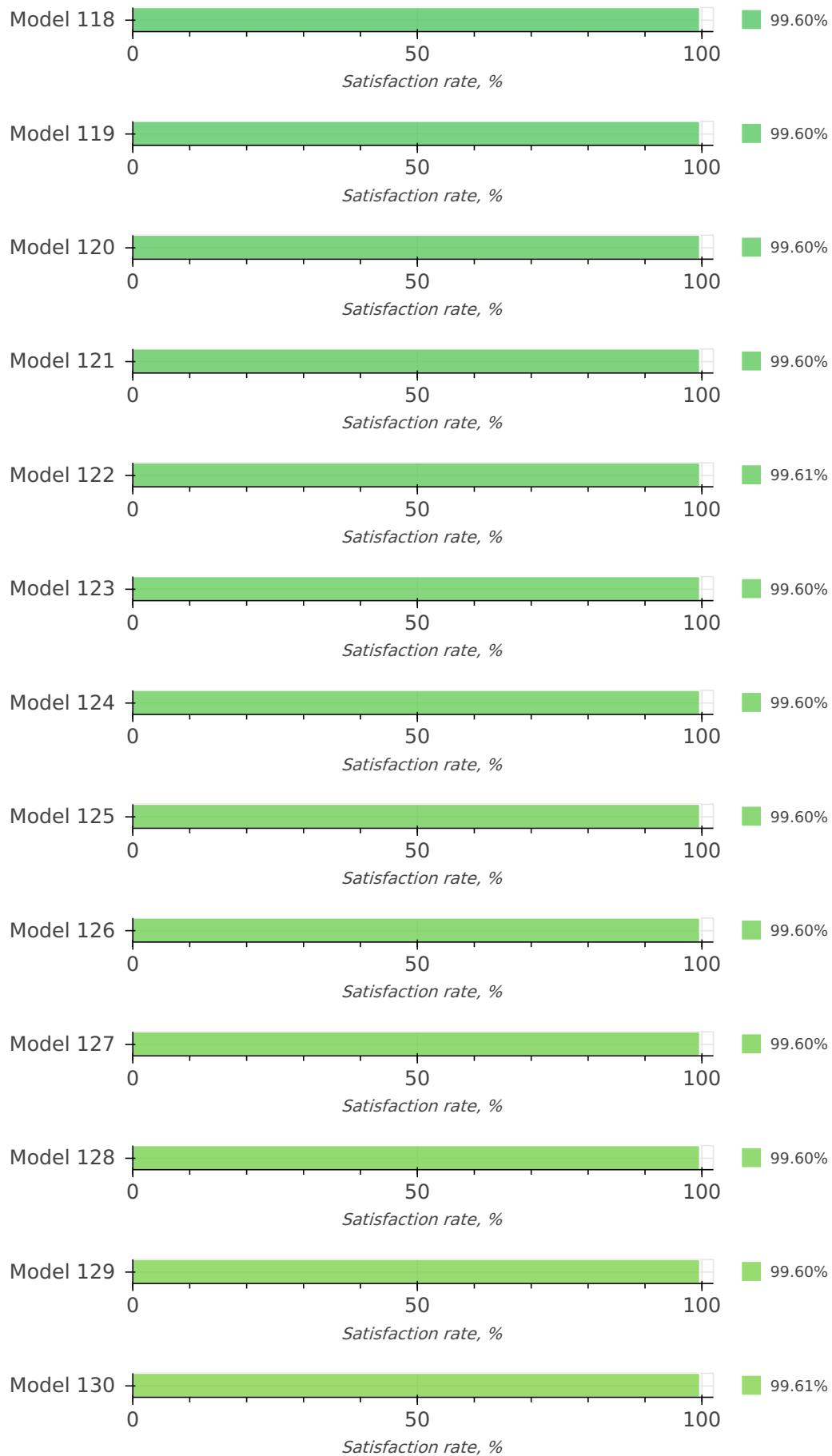


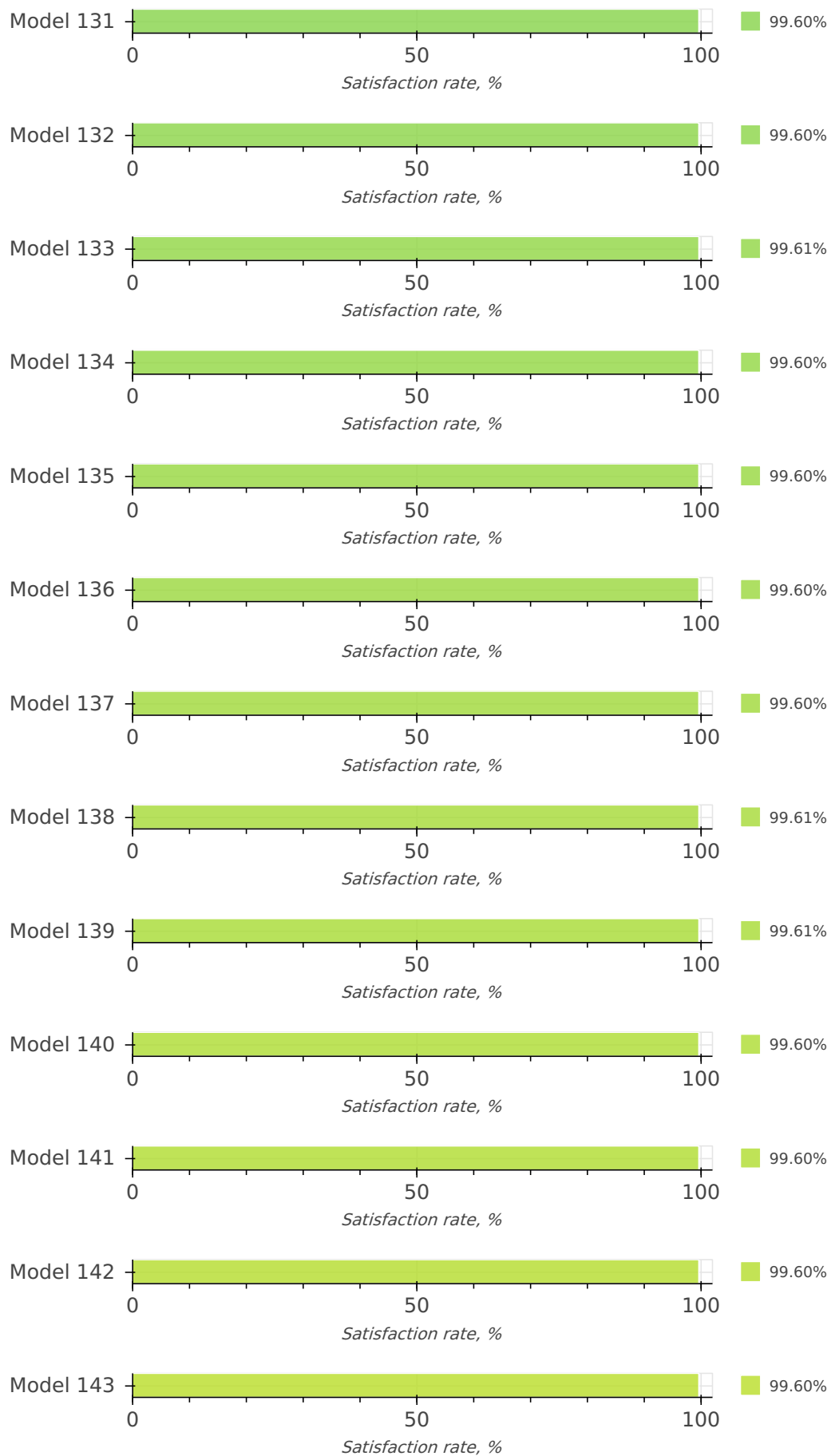


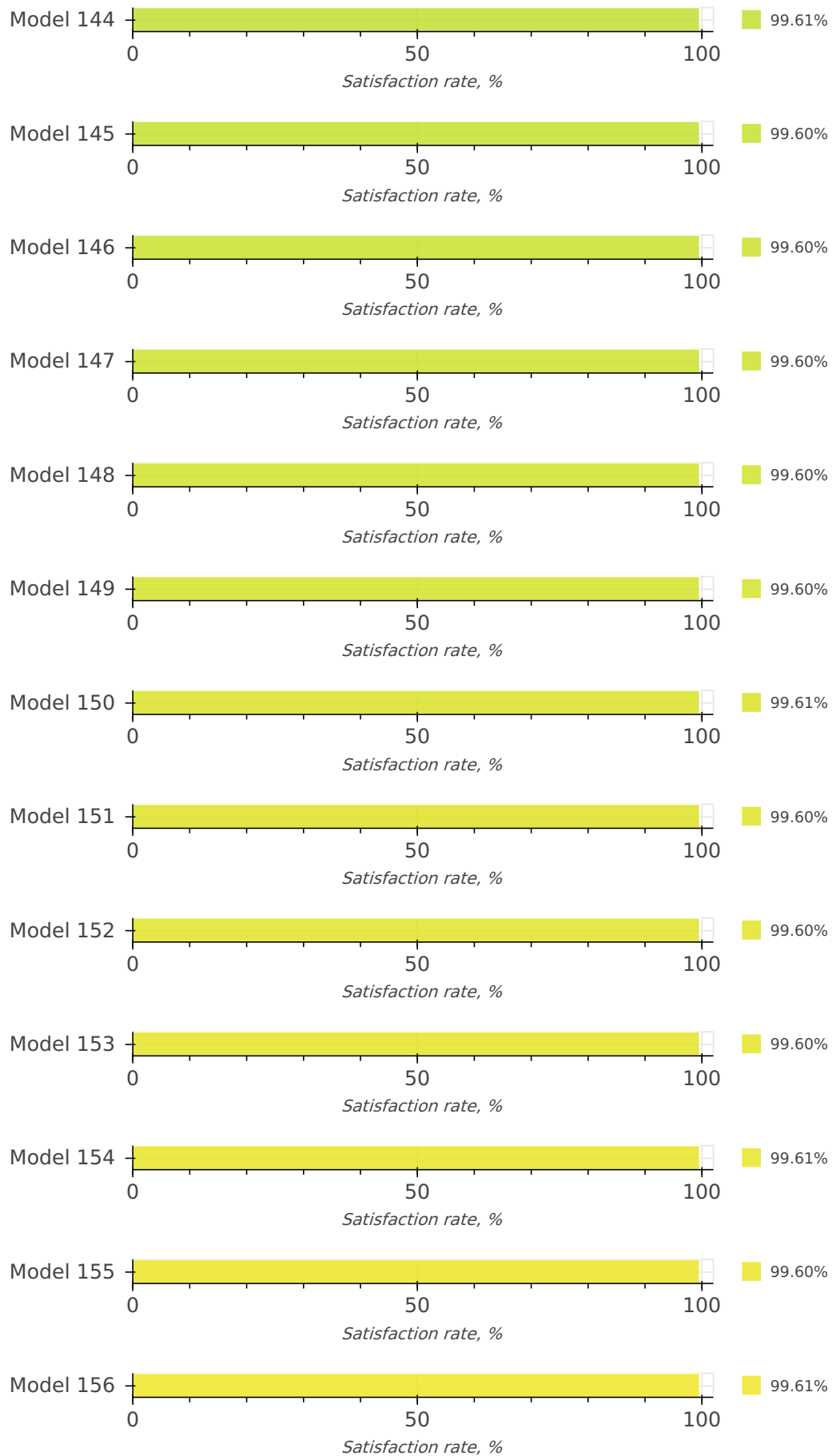


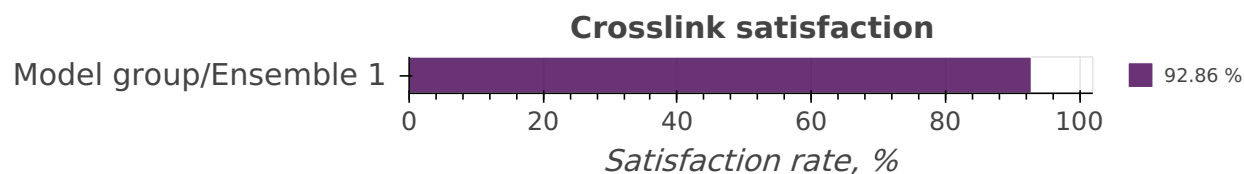












### Ensemble information ?

*This entry consists of 1 distinct ensemble(s).*

### Summary ?

*This entry consists of 159 model(s). A total of 16 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

ID	Model(s)	Entity ID	Molecule name	Chain(s) [auth]	Total residues	Rigid segments	Flexible segments	Model coverage/ Starting model coverage (%)	Scale
1	1-159	1	ORF2	A	1275	8-237, 250-258, 260-277, 284-310, 313-352, 353-359, 362-370, 375-381, 393-849, 857-862, 864-868, 873-955, 960-1030, 1033-1061, 1068-1275	1-7, 238-249, 259, 278-283, 311-312, 360-361, 371-374, 382-392, 850-856, 863, 869-872, 956-959, 1031-1032, 1062-1067	100.00 / 100.00	Coarse-grained: 1 residue(s) per bead

### Datasets used for modeling ?

There are 16 unique datasets used to build the models in this entry.

ID	Dataset type	Database name	Data access code
12	3DEM volume	Zenodo	<a href="https://zenodo.org/record/10377421">10.5281/zenodo.10377421</a>
7	Crosslinking-MS data	Zenodo	<a href="https://zenodo.org/record/10377421">10.5281/zenodo.10377421</a>
8	Crosslinking-MS data	Zenodo	<a href="https://zenodo.org/record/10377421">10.5281/zenodo.10377421</a>
9	Crosslinking-MS data	Zenodo	<a href="https://zenodo.org/record/10377421">10.5281/zenodo.10377421</a>
14	2DEM class average	Zenodo	<a href="https://zenodo.org/record/10377421">10.5281/zenodo.10377421</a>
15	2DEM class average	Zenodo	<a href="https://zenodo.org/record/10377421">10.5281/zenodo.10377421</a>
16	2DEM class average	Zenodo	<a href="https://zenodo.org/record/10377421">10.5281/zenodo.10377421</a>
1	De Novo model	AlphaFoldDB	<a href="https://alphafolddb.org/entry/AF-O00370-F1">AF-O00370-F1</a>
2	De Novo model	MODEL ARCHIVE	<a href="https://modelarchive.org/entry/ma-fejd6">ma-fejd6</a>
3	De Novo model	MODEL ARCHIVE	<a href="https://modelarchive.org/entry/ma-joo4d">ma-joo4d</a>
4	De Novo model	MODEL ARCHIVE	<a href="https://modelarchive.org/entry/ma-lzyrq">ma-lzyrq</a>
5	De Novo model	MODEL ARCHIVE	<a href="https://modelarchive.org/entry/ma-xlzzy">ma-xlzzy</a>
13	De Novo model	MODEL ARCHIVE	<a href="https://modelarchive.org/entry/ma-9wovj">ma-9wovj</a>
6	Mass Spectrometry data	PRIDE	<a href="https://www.ebi.ac.uk/pride/entry/PXD038615">PXD038615</a>
10	EM raw micrographs	EMPIAR	<a href="https://www.ebi.ac.uk/empair/entry/EMPIAR-11556">EMPIAR-11556</a>

ID	Dataset type	Database name	Data access code
11	3DEM volume	EMDB	<a href="#">40856</a>

## 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	Sampling	AlphaFold2	Modeling of full-length ORF2p with AlphaFold2 using varying alignment depth. Details of the simulations are available in the ModelArchive entry ma-fejd6	None	False	False
2	1	Sampling	Molecular Dynamics simulations	Details of molecular dynamics simulations are available ModelArchive entries	None	False	False
3	1	Sampling	Replica Exchange Gibbs sampling, based on Metropolis Monte Carlo	20 replicas; 3 runs; 10000 models per run	30000	False	True
4	1	Refinement	Steepest descent	Conversion of a Ca-model to a full backbone model	159	False	False
5	1	Refinement	SCWRL	Conversion of a backbone model to a full-atom model	159	False	False
6	1	Refinement	Geometry optimization	Conversion of a backbone model to a full-atom model	159	False	False

*There are 7 software packages reported in this entry.*



ID	Software name	Software version	Software classification	Software location
7	<a href="#">Sampcon</a>	2.18.0	validation	<a href="https://github.com/salilab/imp-sampcon">https://github.com/salilab/imp-sampcon</a>
3	<a href="#">ColabFold</a>	1.3.0	model building	<a href="https://github.com/sokrypton/ColabFold">https://github.com/sokrypton/ColabFold</a>
4	<a href="#">GROMACS</a>	2022.30	model building	<a href="https://www.gromacs.org/">https://www.gromacs.org/</a>
1	<a href="#">IMP PMI module</a>	2.19.0	integrative model building	<a href="https://integrativemodeling.org">https://integrativemodeling.org</a>
5	<a href="#">PULCHRA</a>	3.04	model building	<a href="https://sites.gatech.edu/cssb/pulchra/">https://sites.gatech.edu/cssb/pulchra/</a>
6	<a href="#">SCWRL4.0</a>	4.00	model building	<a href="http://dunbrack.fccc.edu/lab/scwrl">http://dunbrack.fccc.edu/lab/scwrl</a>
2	<a href="#">Integrative Modeling Platform (IMP)</a>	2.19.0	integrative model building	<a href="https://integrativemodeling.org">https://integrativemodeling.org</a>

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

### 2DEM class average

Validation for this section is under development.

### 3DEM volume

Validation for this section is under development.

### EM raw micrographs

Validation for this section is under development.

### Mass Spectrometry

Validation for this section is under development.

## 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.

## Excluded volume satisfaction ?

*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.*

Model ID	Analysed	Number of violations	Excluded Volume Satisfaction (%)
1	812175	3213	99.60
2	812175	3209	99.60
3	812175	3214	99.60
4	812175	3213	99.60
5	812175	3208	99.61
6	812175	3218	99.60
7	812175	3219	99.60
8	812175	3225	99.60
9	812175	3219	99.60
10	812175	3207	99.61
11	812175	3218	99.60
12	812175	3214	99.60
13	812175	3207	99.61
14	812175	3215	99.60
15	812175	3217	99.60
16	812175	3210	99.60
17	812175	3215	99.60
18	812175	3214	99.60
19	812175	3207	99.61
20	812175	3208	99.61
21	812175	3212	99.60
22	812175	3214	99.60
23	812175	3215	99.60
24	812175	3211	99.60
25	812175	3208	99.61
26	812175	3201	99.61
27	812175	3212	99.60
28	812175	3203	99.61
29	812175	3207	99.61
30	812175	3220	99.60
31	812175	3214	99.60
32	812175	3216	99.60


Model ID	Analysed	Number of violations	Excluded Volume Satisfaction (%)
33	812175	3218	99.60
34	812175	3213	99.60
35	812175	3207	99.61
36	812175	3217	99.60
37	812175	3219	99.60
38	812175	3215	99.60
39	812175	3211	99.60
40	812175	3212	99.60
41	812175	3206	99.61
42	812175	3221	99.60
43	812175	3214	99.60
44	812175	3211	99.60
45	812175	3218	99.60
46	812175	3218	99.60
47	812175	3218	99.60
48	812175	3212	99.60
49	812175	3220	99.60
50	812175	3202	99.61
51	812175	3206	99.61
52	812175	3219	99.60
53	812175	3205	99.61
54	812175	3218	99.60
55	812175	3208	99.61
56	812175	3211	99.60
57	812175	3214	99.60
58	812175	3202	99.61
59	812175	3210	99.60
60	812175	3203	99.61
61	812175	3209	99.60
62	812175	3221	99.60
63	812175	3220	99.60
64	812175	3220	99.60
65	812175	3209	99.60
66	812175	3213	99.60
67	812175	3212	99.60

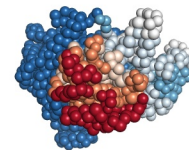
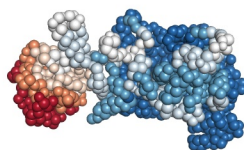
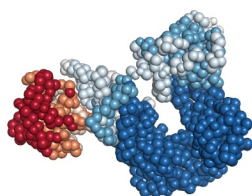
Model ID	Analysed	Number of violations	Excluded Volume Satisfaction (%)
68	812175	3212	99.60
69	812175	3207	99.61
70	812175	3202	99.61
71	812175	3214	99.60
72	812175	3217	99.60
73	812175	3215	99.60
74	812175	3207	99.61
75	812175	3207	99.61
76	812175	3210	99.60
77	812175	3219	99.60
78	812175	3214	99.60
79	812175	3215	99.60
80	812175	3204	99.61
81	812175	3212	99.60
82	812175	3220	99.60
83	812175	3215	99.60
84	812175	3209	99.60
85	812175	3209	99.60
86	812175	3214	99.60
87	812175	3208	99.61
88	812175	3215	99.60
89	812175	3215	99.60
90	812175	3218	99.60
91	812175	3207	99.61
92	812175	3202	99.61
93	812175	3213	99.60
94	812175	3214	99.60
95	812175	3216	99.60
96	812175	3206	99.61
97	812175	3218	99.60
98	812175	3207	99.61
99	812175	3208	99.61
100	812175	3211	99.60
101	812175	3210	99.60
102	812175	3217	99.60

Model ID	Analysed	Number of violations	Excluded Volume Satisfaction (%)
103	812175	3211	99.60
104	812175	3220	99.60
105	812175	3205	99.61
106	812175	3215	99.60
107	812175	3216	99.60
108	812175	3219	99.60
109	812175	3210	99.60
110	812175	3213	99.60
111	812175	3212	99.60
112	812175	3197	99.61
113	812175	3212	99.60
114	812175	3214	99.60
115	812175	3224	99.60
116	812175	3214	99.60
117	812175	3207	99.61
118	812175	3211	99.60
119	812175	3209	99.60
120	812175	3211	99.60
121	812175	3210	99.60
122	812175	3206	99.61
123	812175	3216	99.60
124	812175	3216	99.60
125	812175	3213	99.60
126	812175	3221	99.60
127	812175	3211	99.60
128	812175	3212	99.60
129	812175	3218	99.60
130	812175	3207	99.61
131	812175	3217	99.60
132	812175	3213	99.60
133	812175	3207	99.61
134	812175	3217	99.60
135	812175	3214	99.60
136	812175	3217	99.60
137	812175	3219	99.60

Model ID	Analysed	Number of violations	Excluded Volume Satisfaction (%)
138	812175	3206	99.61
139	812175	3205	99.61
140	812175	3214	99.60
141	812175	3214	99.60
142	812175	3209	99.60
143	812175	3211	99.60
144	812175	3205	99.61
145	812175	3219	99.60
146	812175	3228	99.60
147	812175	3218	99.60
148	812175	3214	99.60
149	812175	3213	99.60
150	812175	3203	99.61
151	812175	3216	99.60
152	812175	3216	99.60
153	812175	3214	99.60
154	812175	3207	99.61
155	812175	3216	99.60
156	812175	3206	99.61
157	812175	3217	99.60
158	812175	3217	99.60
159	812175	3217	99.60

### PrISM precision analysis ?

Regions of **low**  **high** precision, defined as the variability among the models that satisfy the input data and calculated as the density-weighted root mean-square fluctuation (RMSF) from the bead/atom center of density, annotated and visualized using PrISM. The per-bead precision is computed from the deposited ensemble of superposed integrative models. High- and low-precision regions are then determined by clustering beads of similar precision based on their proximity in the structure. Only coarse-grained beads (or CA atoms for atomic models) of deposited models are used for assessment and visualization, and three projections for each representative model are generated. PrISM analysis for Ensemble 1 (models deposited/total: 159/1383).



## Fit of model to data used for modeling ?

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

#### Restraint types

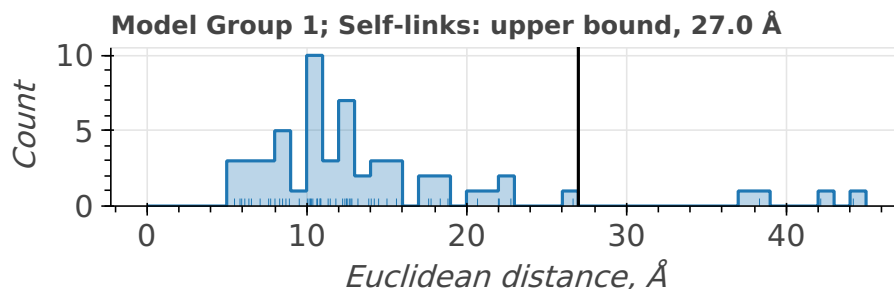
Restraint types are summarized in the table below. Restraints assigned "*by-residue*" are interpreted as between CA atoms. Restraints between coarse-grained beads are indicated as "*coarse-grained*". *Restraint group* represents a set of crosslinking restraints applied collectively in the modeling.

There are 56 crosslinking restraints combined in 56 restraint groups.

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
BS3	LYS	CA	LYS	CA	upper bound	27.0	56

#### 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=56)
1	1	1	159/1383	All	92.86	7.14	56
				Self-links/Intramolecular	92.86	7.14	56

#### 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.



### 2DEM class average

Validation for this section is under development.

### 3DEM volume

Validation for this section is under development.

### EM raw micrographs

Validation for this section is under development.

### Mass Spectrometry

Validation for this section is under development.

## 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.*



