Integrative Structure Validation Report

July 22, 2024 - 03:54 PM PDT

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

Python-IHM Version 1.3
MolProbity Version 4.5.2
ATSAS Version 3.2.1 (r14885)
Integrative Modeling Validation Version 1.2

PDB ID	8ZZZ
PDB-Dev ID	PDBDEV_00000035
Structure Title	Rosetta model of human LRH-1 nuclear receptor generated with XL-MS, HDX-MS, and SAXS data
Structure Authors	Seacrist CD; Kuenze G; Hoffmann R; Burke J; Meiler J; Blind RD

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

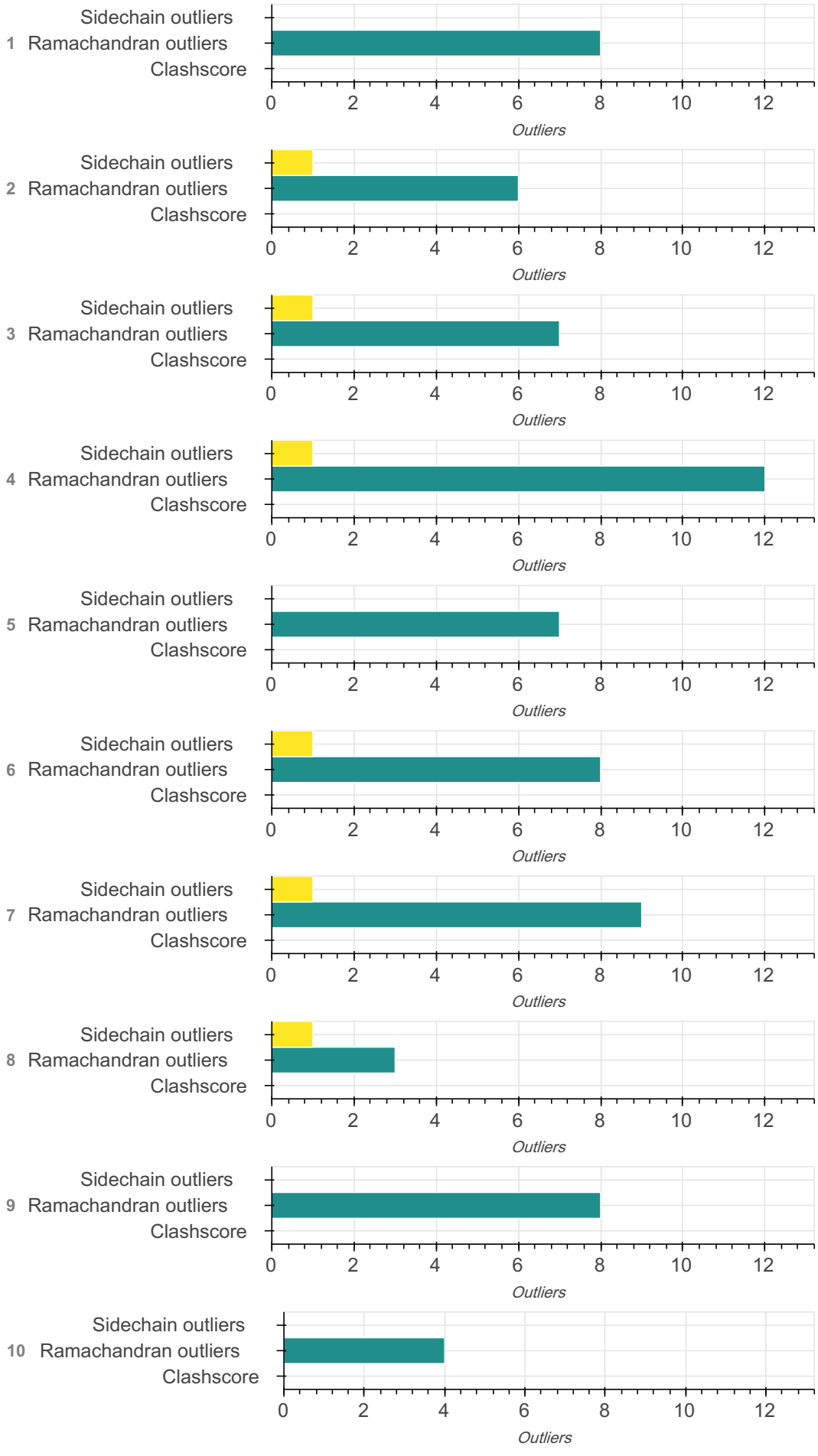
A user guide is available at https://pdb-dev.wwpdb.org/validation_help.html with specific help available everywhere you see the ? symbol.

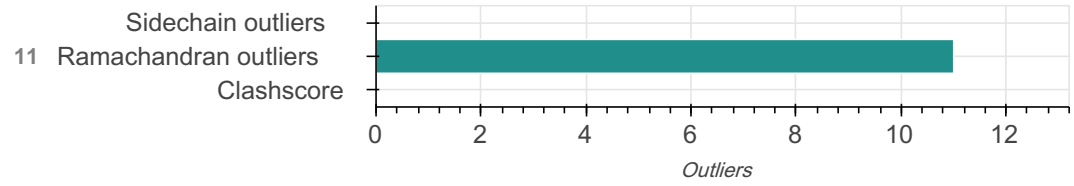
List of references used to build this report is available here.

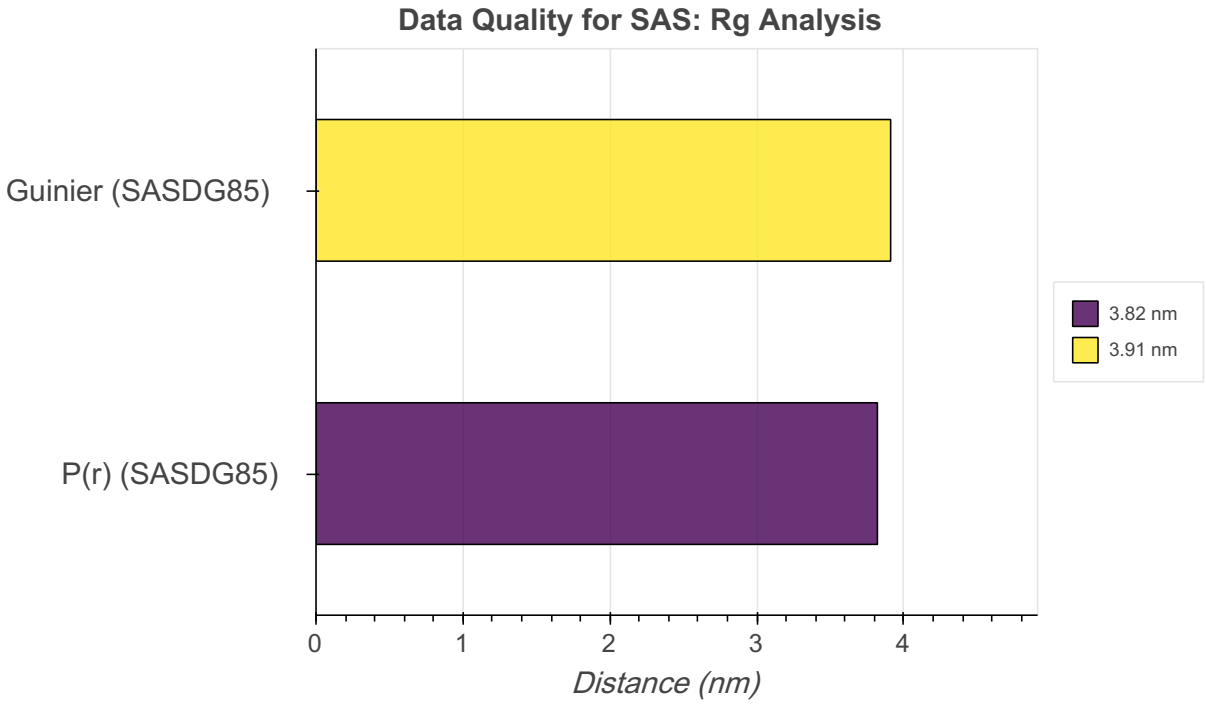
Overall quality

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


1
2
3
4
5
6
7
8
9
10
Sidechain outliers
Ramachandran outliers
Clashscore
0
2
4
6
8
10
12
Outliers





Ensemble information

This entry consists of 1 distinct ensemble(s).

Summary

This entry consists of 11 unique models, with 7 subunits in each model. A total of 9 datasets or restraints were used to build this entry. Each model is represented by 10 rigid bodies and 5 flexible or non-rigid units.

Entry composition

There are 11 unique types of models in this entry.These models are titled Best scoring LRH-1 docking model, LRH-1 full-length model 1, LRH-1 full-length model 2, LRH-1 full-length model 3, LRH-1 full-length model 4, LRH-1 full-length model 5, LRH-1 full-length model 6, LRH-1 full-length model 7, LRH-1 full-length model 8, LRH-1 full-length model 9, LRH-1 full-length model 10 respectively.

Model ID	Subunit number	Subunit ID	Subunit name	Chain ID	Chain ID [auth]	Total residues
1	1	1	LRH-1 DNA-binding domain	A	A	102
1	2	2	LRH-1 Ligand-binding domain	B	B	240
1	3	4	PGC1-alpha coactivator peptide	D	D	10
1	4	5	DNA strand 1	E	E	12
1	5	6	DNA strand 2	F	F	12
1	6	7	Phospholipid ligand	X	X	Not available
1	7	8	Zinc ion	Y	Y	Not available
1	8	8	Zinc ion	Z	Z	Not available
2	1	3	LRH-1 full length polypeptide	C	C	538
2	2	4	PGC1-alpha coactivator peptide	D	D	10
2	3	5	DNA strand 1	E	E	12
2	4	6	DNA strand 2	F	F	12
2	5	7	Phospholipid ligand	X	X	Not available
2	6	8	Zinc ion	Y	Y	Not available
2	7	8	Zinc ion	Z	Z	Not available
3	1	3	LRH-1 full length polypeptide	C	C	538

Model ID	Subunit number	Subunit ID	Subunit name	Chain ID	Chain ID [auth]	Total residues
3	2	4	PGC1-alpha coactivator peptide	D	D	10
3	3	5	DNA strand 1	E	E	12
3	4	6	DNA strand 2	F	F	12
3	5	7	Phospholipid ligand	X	X	Not available
3	6	8	Zinc ion	Y	Y	Not available
3	7	8	Zinc ion	Z	Z	Not available
4	1	3	LRH-1 full length polypeptide	C	C	538
4	2	4	PGC1-alpha coactivator peptide	D	D	10
4	3	5	DNA strand 1	E	E	12
4	4	6	DNA strand 2	F	F	12
4	5	7	Phospholipid ligand	X	X	Not available
4	6	8	Zinc ion	Y	Y	Not available
4	7	8	Zinc ion	Z	Z	Not available
5	1	3	LRH-1 full length polypeptide	C	C	538
5	2	4	PGC1-alpha coactivator peptide	D	D	10
5	3	5	DNA strand 1	E	E	12
5	4	6	DNA strand 2	F	F	12

Model ID	Subunit number	Subunit ID	Subunit name	Chain ID	Chain ID [auth]	Total residues
5	5	7	Phospholipid ligand	X	X	Not available
5	6	8	Zinc ion	Y	Y	Not available
5	7	8	Zinc ion	Z	Z	Not available
6	1	3	LRH-1 full length polypeptide	C	C	538
6	2	4	PGC1-alpha coactivator peptide	D	D	10
6	3	5	DNA strand 1	E	E	12
6	4	6	DNA strand 2	F	F	12
6	5	7	Phospholipid ligand	X	X	Not available
6	6	8	Zinc ion	Y	Y	Not available
6	7	8	Zinc ion	Z	Z	Not available
7	1	3	LRH-1 full length polypeptide	C	C	538
7	2	4	PGC1-alpha coactivator peptide	D	D	10
7	3	5	DNA strand 1	E	E	12
7	4	6	DNA strand 2	F	F	12
7	5	7	Phospholipid ligand	X	X	Not available
7	6	8	Zinc ion	Y	Y	Not available

Model ID	Subunit number	Subunit ID	Subunit name	Chain ID	Chain ID [auth]	Total residues
7	7	8	Zinc ion	Z	Z	Not available
8	1	3	LRH-1 full length polypeptide	C	C	538
8	2	4	PGC1-alpha coactivator peptide	D	D	10
8	3	5	DNA strand 1	E	E	12
8	4	6	DNA strand 2	F	F	12
8	5	7	Phospholipid ligand	X	X	Not available
8	6	8	Zinc ion	Y	Y	Not available
8	7	8	Zinc ion	Z	Z	Not available
9	1	3	LRH-1 full length polypeptide	C	C	538
9	2	4	PGC1-alpha coactivator peptide	D	D	10
9	3	5	DNA strand 1	E	E	12
9	4	6	DNA strand 2	F	F	12
9	5	7	Phospholipid ligand	X	X	Not available
9	6	8	Zinc ion	Y	Y	Not available
9	7	8	Zinc ion	Z	Z	Not available
10	1	3	LRH-1 full length polypeptide	C	C	538

Model ID	Subunit number	Subunit ID	Subunit name	Chain ID	Chain ID [auth]	Total residues
10	2	4	PGC1-alpha coactivator peptide	D	D	10
10	3	5	DNA strand 1	E	E	12
10	4	6	DNA strand 2	F	F	12
10	5	7	Phospholipid ligand	X	X	Not available
10	6	8	Zinc ion	Y	Y	Not available
10	7	8	Zinc ion	Z	Z	Not available
11	1	3	LRH-1 full length polypeptide	C	C	538
11	2	4	PGC1-alpha coactivator peptide	D	D	10
11	3	5	DNA strand 1	E	E	12
11	4	6	DNA strand 2	F	F	12
11	5	7	Phospholipid ligand	X	X	Not available
11	6	8	Zinc ion	Y	Y	Not available
11	7	8	Zinc ion	Z	Z	Not available

Datasets used for modeling

There are 9 unique datasets used to build the models in this entry.

ID	Dataset type	Database name	Data access code
1	Experimental model	PDB	2A66

ID	Dataset type	Database name	Data access code
2	Experimental model	PDB	1YOK
3	Comparative model	File	10.5281/zenodo.3405545
4	Comparative model	File	10.5281/zenodo.3405545
5	Crosslinking-MS data	File	10.5281/zenodo.3405545
6	H/D exchange data	File	10.5281/zenodo.3405545
7	Integrative model	File	10.5281/zenodo.3405545
8	Crosslinking-MS data	File	10.5281/zenodo.3405545
9	SAS data	SASBDB	SASDG85

Representation

This entry has only one representation and includes 10 rigid bodies and 5 flexible units

Chain ID	Rigid bodies	Non-rigid segments
A	-	1-102
B	-	1-240
C	-	1-538
D	1-10	1-10
E	1-121-12	-
F	1-121-12	-
X	None-None	None-None
Y	None-NoneNone-None	-
Z	None-NoneNone-None	-

Methodology and software

This entry is a result of 2 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	RosettaRemodel	Loop modeling	None	100	False	False
2	1	RosettaDock	Docking	None	40000	False	False
1	2	Ranch and RosettaMinimize	Linker modeling	None	45000	False	False

There are 3 software packages reported in this entry.

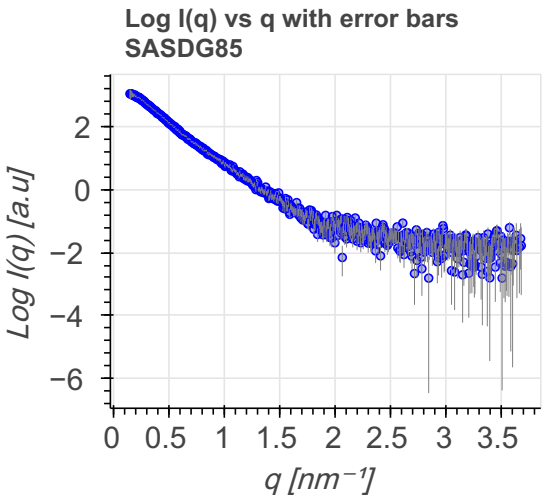
ID	Software name	Software version	Software classification	Software location
1	Rosetta	3.10	model building, model validation	https://www.rosettacommons.org/
2	ATSAS	2.8.4	model building, model validation	https://www.embl-hamburg.de/biosaxs/software.html
3	REMO	1	model building	https://zhanglab.ccmb.med.umich.edu/REMO/

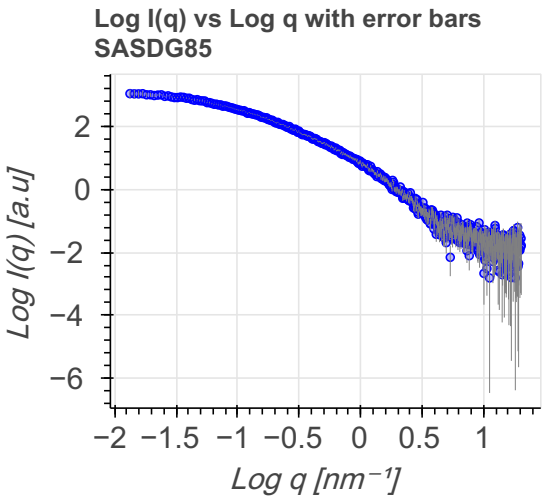
Data quality

Scattering profile

SAS data used in this integrative model was obtained from 1 deposited SASBDB entry (entries).

Scattering profile for SASDG85: data from solutions of biological macromolecules are presented as both log I(q) vs q and log I(q) vs log (q) based on SAS validation task force (SASvtf) recommendations. I(q) is the intensity (in arbitrary units) and q is the modulus of the scattering vector.





Key experimental estimates

Molecular weight (MW) estimates from experiments and analysis: true molecular weight can be compared to the Porod estimate from scattering profiles.

SASDB ID	Chemical composition MW	Standard MW	Porod Volume/MW
SASDG85	67.0 kDa	N/A	N/A

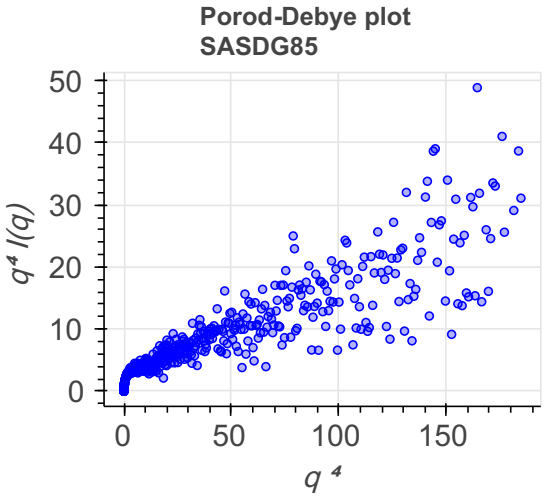
Volume estimates from experiments and analysis: estimated volume can be compared to Porod volume obtained from scattering profiles.

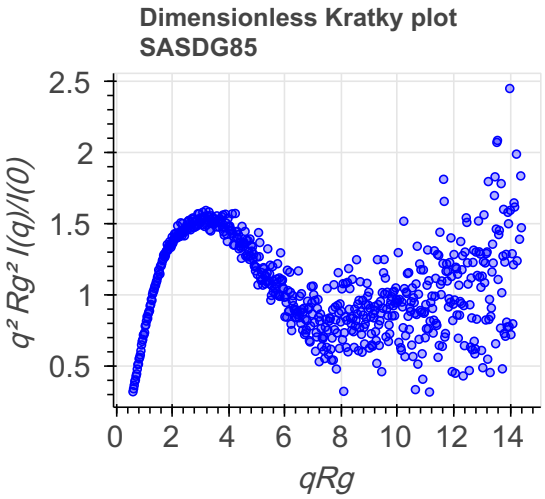
SASDB ID	Estimated Volume	Porod Volume	Specific Volume	Sample Contrast	Sample Concentration
SASDG85	N/A	76.00 nm³	N/A	N/A	7.50 mg/mL

Flexibility analysis

Flexibility analysis for SASDG85: In a Porod-Debye plot, a clear plateau is observed for globular (partial or fully folded) domains, whereas, fully unfolded domains are devoid of any discernable plateau. For details, refer to Figure 5 in Rambo

and Tainer, 2011 . In a Kratky plot, a parabolic shape is observed for globular (partial or fully folded) domains and a hyperbolic shape is observed for fully unfolded domains.



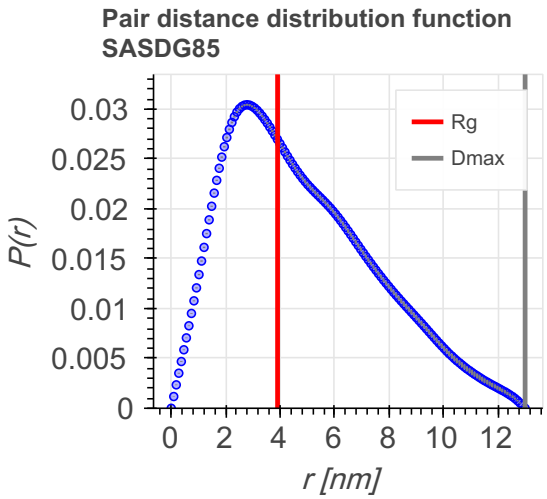


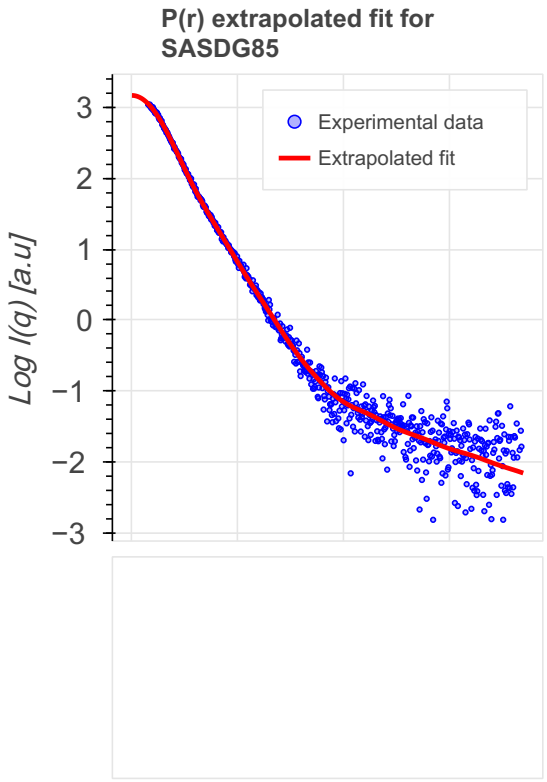
Pair-distance distribution analysis

P(r) analysis: P(r) represents the distribution of distances between all pairs of atoms within the particle weighted by the respective electron densities. P(r) is the Fourier transform of I(s) (and vice versa). R_g can be estimated from integrating the P(r) function. Agreement between the P(r) and Guinier-determined R_g (table below) is a good measure of the self-consistency of the SAS profile. R_g is a measure for the overall size of a macromolecule; e.g. a protein with a smaller R_g is more compact than a protein with a larger R_g, provided both have the same molecular weight (MW). The point where P(r) is decaying to zero is called D_{max} and represents the maximum size of the particle.

SASDB ID	Software used	Dmax	Dmax error	Rg	Rg error
SASDG85	GNOM 4.6	13.000 nm	N/A	3.907 nm	0.020 nm

P(r) for SASDG85: The value of P(r) should be zero beyond r=D_{max}.





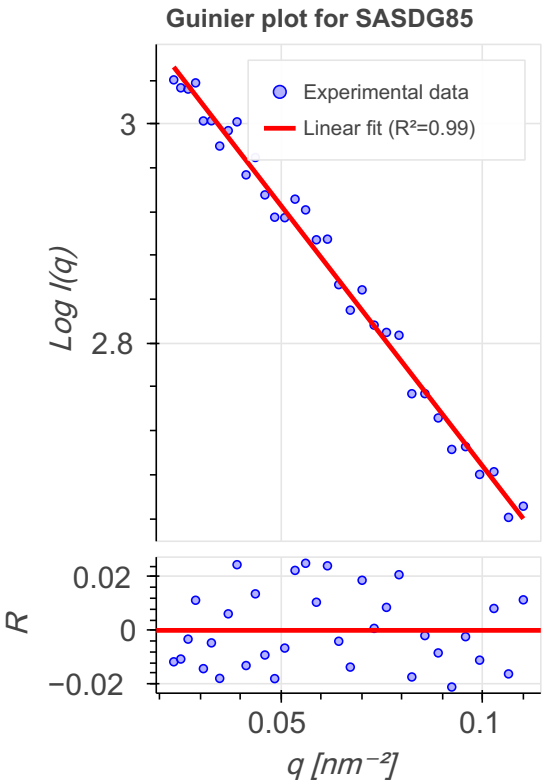
Guinier analysis

Guinier analysis: agreement between the P(r) and Guinier-determined R_g (table below) is a good measure of the self-consistency of the SAS profile. Molecular weight estimates can also be compared to Porod and sample molecular weights for consistency.

SASDB ID	Rg	Rg error	MW	MW error
SASDG85	3.82 nm	0.05 nm	N/A	N/A

Guinier analysis for SASDG85: the linearity of the Guinier plot is a sensitive indicator of the quality of the experimental SAS data; a linear Guinier plot is a necessary but not sufficient demonstration that a solution contains monodisperse particles of the same size. Deviations from linearity usually point to strong interference effects, polydispersity of the

samples or improper background subtraction. Residual value plot and coefficient of determination (R^2) are measures to assess linear fit to the data. A perfect fit has an R^2 value of 1. Residual values should be equally and randomly spaced around the horizontal axis.



H/D exchange

Validation for this section is under development.

Crosslinking-MS

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.

Standard geometry: bond outliers

Bond length outliers can not be evaluated for this model

Standard geometry: angle outliers

Bond angle outliers do not exist or can not be evaluated for this model

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	0.00	0
2	0.00	0
3	0.00	0
4	0.00	0
5	0.00	0
6	0.00	0
7	0.00	0
8	0.00	0
9	0.00	0
10	0.00	0
11	0.00	0

All 0 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

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	Analyzed	Favored	Allowed	Outliers
1	526	490	28	8
2	526	490	30	6
3	526	489	30	7
4	526	481	33	12
5	526	495	24	7
6	526	489	29	8
7	526	490	27	9
8	526	493	30	3

Model ID	Analyzed	Favored	Allowed	Outliers
9	526	491	27	8
10	526	493	29	4
11	526	481	34	11

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	469	464	5	0
2	469	465	3	1
3	468	463	4	1
4	469	464	4	1
5	469	464	5	0
6	468	464	3	1
7	469	463	5	1
8	469	463	5	1
9	468	465	3	0
10	469	463	6	0
11	468	463	5	0

Detailed list of outliers are tabulated below.

Model ID	Chain	Residue ID	Residue type
2	C	298	PRO
3	C	82	LEU
4	C	278	THR
6	C	239	PRO

Model ID	Chain	Residue ID	Residue type
7	C	43	PRO
8	C	250	PRO

Fit of model to data used for modeling

Fit of model(s) to SAS data

χ^2 goodness of fit and cormap analysis

Model(s) and/or fit for this entry have not been deposited.

H/D exchange

Validation for this section is under development.

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 labcontributed model validation metrics and software packages.

Implementation of validation methods for SAS data and SAS-based models are funded byRCSB PDB (grant number DBI-1832184). Dr. Stephen Burley, Dr. John Westbrook, and Dr. Jasmine Young from RCSB PDB, Dr. Jill Trewhella, 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.