Integrative Structure Validation Report ?

October 09, 2025 - 04:37 PM PDT

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

IHMValidation Version 3.0
Python-IHM Version 2.5
ATSAS Version 3.2.1 (r14885)
EMDB validation analysis Version 0.0.1.dev127
ChimeraX Version 1.9
Chimera Version 1.19
MapQ Version 1.8.1

PDB ID	8ZZA | pdb_00008zza
PDB-Dev ID	PDBDEV_00000010
Structure Title	Integrative structure and functional anatomy of a single spoke of a nuclear pore complex
Structure Authors	Kim, S.J.; Fernandez-Martinez, J.; Nudelman, I.; Shi, Y.; Zhang, W.; Raveh, B.; Herricks, T.; Slaughter, B.D.; Hogan, J.A.; Upla, P.; Chemmama, I.E.; Pellarin, R.; Echeverria, I.; Shivaraju, M.; Chaudhury, A.S.; Wang, J.; Williams, R.; Unruh, J.R.; Greenberg, C.H.; Jacobs, E.Y.; Yu, Z.; de la Cruz, M.J.; Mironska, R.; Stokes, D.L.; Aitchison, J.D.; Jarrold, M.F.; Gerton, J.L.; Ludtke, S.J.; Akey, C.W.; Chait, B.T.; Sali, A.; Rout, M.P.
Deposited on	2018-01-04

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 ?

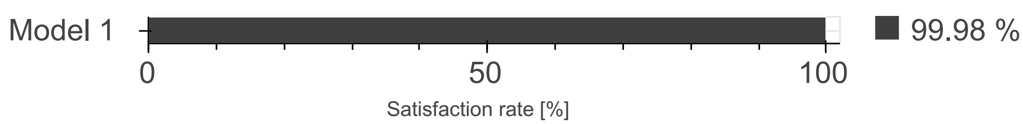
This entry consists of 1 model(s). A total of 65 dataset(s) were used to build this entry.

Name	Type	Count
2DEM class average	Experimental data	2
3DEM volume	Experimental data	2
Crosslinking-MS data	Experimental data	2
EM raw micrographs	Experimental data	2
Mass Spectrometry data	Experimental data	1
SAS data	Experimental data	36
Experimental model	Starting model	7
Comparative model	Starting model	9
Integrative model	Starting model	4

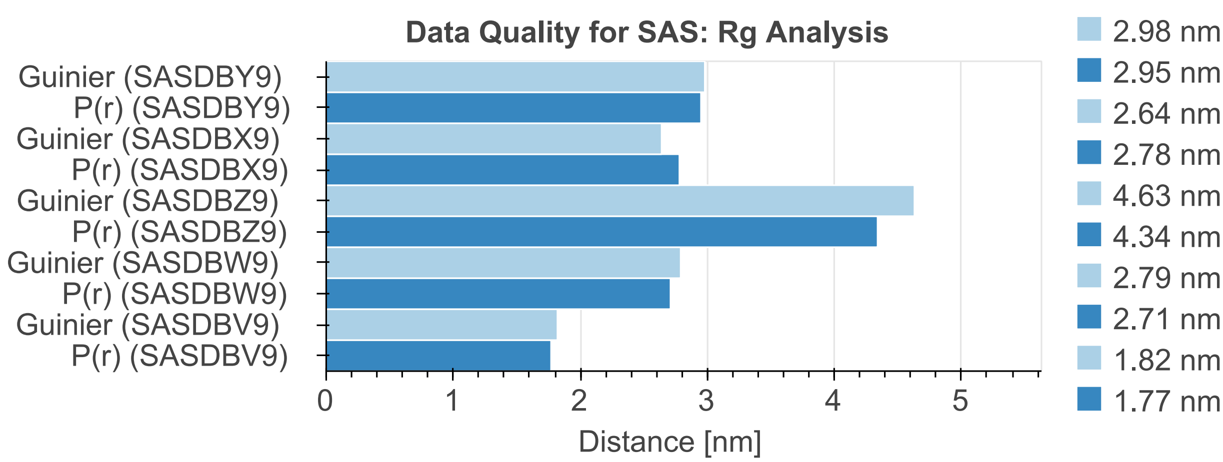
1.2. Overall quality ?

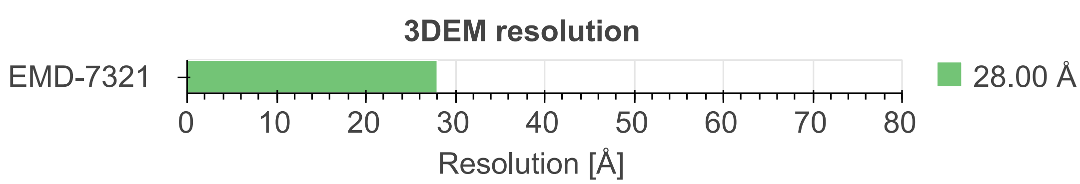
This validation report contains model quality assessments for all structures, data quality and fit to model assessments for SAS and crosslinking-MS datasets. Data quality and fit to model assessments for other datasets and model uncertainty are under development. Number of plots is limited to 256.

Model Quality: Excluded Volume Analysis ?

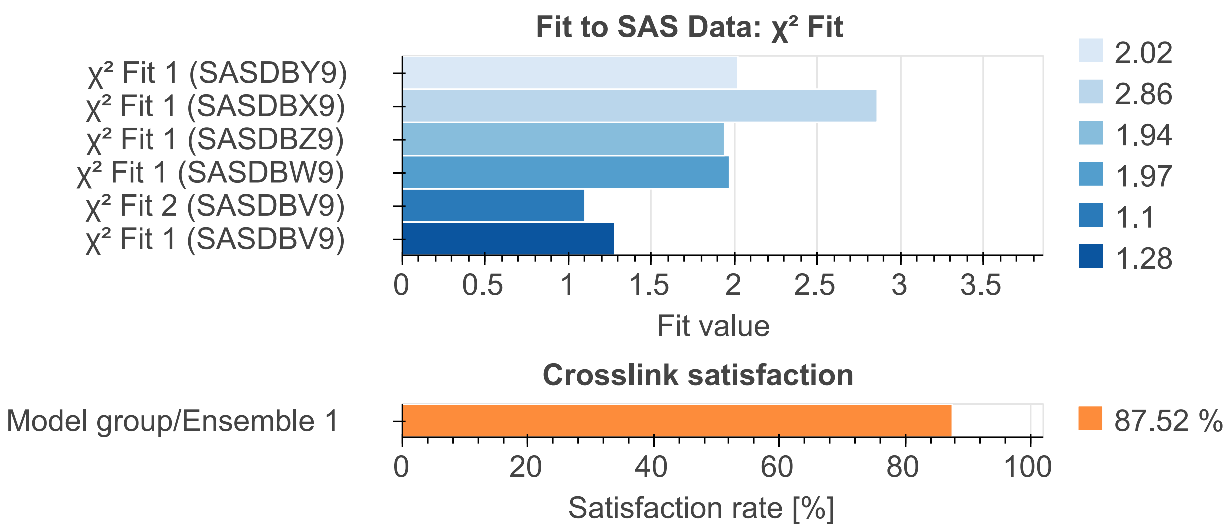


Data Quality ?





Fit to Data Used for Modeling ?



2. Model Details

2.1. Ensemble information

This entry consists of 1 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	Nup84	A	726	-	1-6, 7-20, 21-26, 27-80, 81-95, 96-126, 127-135, 136-364, 365-371, 372-483, 484-505, 506-562, 563-574, 575-726	100.00 / 89.39	Multiscale: Coarse-grained: 1 - 10 residue(s) per bead
				H					
		2	Nup85	B	744	-	1-46, 47-126, 127-131, 132-230, 231-234, 235-436, 437-450, 451-492, 493-495, 496-544, 545-552, 553-560, 561-566, 567-585, 586-589, 590-597, 598-602, 603-612, 613-615, 616-634, 635-637, 638-655, 656-660, 661-675, 676-684, 685-699, 700-706, 707-719, 720-724, 725-744	100.00 / 82.93	Multiscale: Coarse-grained: 1 - 10 residue(s) per bead
				I					

ID	Model(s)	Entity ID	Molecule name	Chain(s) [auth]	Total residues	Rigid segments	Flexible segments	Model coverage/ Starting model coverage (%)	Scale
		3	Nup120	C	1037	-	1, 2-29, 30-52, 53-305, 306-310, 311-711, 712-714, 715-726, 727-732, 733-746, 747-753, 754-766, 767-769, 770-781, 782-806, 807-818, 819-820, 821-833, 834-837, 838-853, 854-861, 862-879, 880-883, 884-895, 896-900, 901-913, 914-916, 917-931, 932-942, 943-955, 956-959, 960-971, 972-975, 976-987, 988-993, 994-1008, 1009-1024, 1025-1036, 1037	100.00 / 86.40	Multiscale: Coarse-grained: 1 - 10 residue(s) per bead
				J					
		4	Nup133	D	1157	-	1-55, 56-78, 79-85, 86-125, 126-132, 133-144, 145-161, 162-184, 185-192, 193-200, 201-205, 206-249, 250-257, 258-480, 481-489, 490-763, 764-771, 772-1155, 1156-1157	100.00 / 89.11	Multiscale: Coarse-grained: 1 - 10 residue(s) per bead
				K					
		5	Nup145c	E	712	-	1-91, 92-99, 100-125, 126-144, 145-148, 149-550, 551-553, 554-560, 561-565, 566-576, 577-586, 587-602, 603-611, 612-624, 625-630, 631-645, 646-653, 654-673, 674-680, 681-689, 690-702, 703-712	100.00 / 74.44	Multiscale: Coarse-grained: 1 - 10 residue(s) per bead
				L					
		6	Seh1	F	349	-	1-248, 249-287, 288-346, 347-349	100.00 / 87.97	Multiscale: Coarse-grained: 1 - 10 residue(s) per bead
				M					
		7	Sec13	G	297	-	1-9, 10-158, 159-165, 166-296, 297	100.00 / 94.28	Multiscale: Coarse-grained: 1 - 9 residue(s) per bead
				N					
		8	Dyn2	O	92	-	1-6, 7-92	100.00 / 93.48	Multiscale: Coarse-grained: 1 - 6 residue(s) per bead
				P					

ID	Model(s)	Entity ID	Molecule name	Chain(s) [auth]	Total residues	Rigid segments	Flexible segments	Model coverage/ Starting model coverage (%)	Scale
		9	Nup82	Q R	713	-	1-6, 7-16, 17-22, 23-120, 121-122, 123-452, 453-521, 522-612, 613-624, 625-669, 670-677, 678-713	100.00 / 85.55	Multiscale: Coarse-grained: 1 - 10 residue(s) per bead
		10	Nup159	S T	1460	-	1082-1116, 1117-1126, 1127-1210, 1211-1239, 1240-1265, 1266-1321, 1322-1331, 1332-1372, 1373-1381, 1382-1412, 1413-1428, 1429-1456, 1457-1460	25.96 / 51.45	Multiscale: Coarse-grained: 1 - 35 residue(s) per bead
		11	Nsp1	U V AF AJ	823	-	601-636, 637-727, 728-741, 742-778, 779-787, 788-823	27.10 / 73.54	Multiscale: Coarse-grained: 1 - 36 residue(s) per bead
		12	Nic96	W AA	839	20-56	1-19, 57-204, 205-360, 361-365, 366-374, 375-404, 405-444, 445-454, 455-515, 516-532, 533-747, 748-752, 753-835, 836-839	100.00 / 71.63	Multiscale: Coarse-grained: 1 - 25 residue(s) per bead
		11	Nsp1	X AB	823	637-727, 742-778, 788-823	601-636, 728-741, 779-787	27.10 / 73.54	Multiscale: Coarse-grained: 1 - 36 residue(s) per bead
		13	Nup49	Y AC	472	270-359, 369-407, 433-472	201-269, 360-368, 408-432	57.63 / 62.13	Multiscale: Coarse-grained: 1 - 69 residue(s) per bead

ID	Model(s)	Entity ID	Molecule name	Chain(s) [auth]	Total residues	Rigid segments	Flexible segments	Model coverage/ Starting model coverage (%)	Scale
		14	Nup57	Z AD	541	287-423, 433-476, 505-540	201-286, 424-432, 477-504, 541	63.03 / 63.64	Multiscale: Coarse-grained: 1 - 86 residue(s) per bead
		12	Nic96	AE AI	839	-	1-19, 20-56, 57-204, 205-360, 361-365, 366-374, 375-404, 405-444, 445-454, 455-515, 516-532, 533-747, 748-752, 753-835, 836-839	100.00 / 71.63	Multiscale: Coarse-grained: 1 - 25 residue(s) per bead
		13	Nup49	AG AK	472	-	201-269, 270-359, 360-368, 369-407, 408-432, 433-472	57.63 / 62.13	Multiscale: Coarse-grained: 1 - 69 residue(s) per bead
		14	Nup57	AH AL	541	-	201-286, 287-423, 424-432, 433-476, 477-504, 505-540, 541	63.03 / 63.64	Multiscale: Coarse-grained: 1 - 86 residue(s) per bead
		15	Nup157	AM AQ	1391	-	1-87, 88-289, 290-300, 301-309, 310-338, 339-457, 458-480, 481-515, 516-534, 535-679, 680-703, 704-730, 731-743, 744-775, 776-785, 786-830, 831-835, 836-892, 893-899, 900-916, 917-920, 921-933, 934-943, 944-1016, 1017-1038, 1039-1141, 1142-1154, 1155-1390, 1391	100.00 / 80.01	Multiscale: Coarse-grained: 1 - 25 residue(s) per bead
		16	Nup170	AN AR	1502	-	1-97, 98-299, 300-310, 311-319, 320-352, 353-471, 472-504, 505-537, 538-573, 574-717, 718-764, 765-791, 792-830, 831-862, 863-883, 884-916, 917-918, 919-930, 931-935, 936-992, 993-999, 1000-1016, 1017-1020, 1021-1033, 1034-1043, 1044-1116, 1117-1140, 1141-1191, 1192-1194, 1195-1243, 1244-1256, 1257-1502	100.00 / 74.37	Multiscale: Coarse-grained: 1 - 25 residue(s) per bead

ID	Model(s)	Entity ID	Molecule name	Chain(s) [auth]	Total residues	Rigid segments	Flexible segments	Model coverage/ Starting model coverage (%)	Scale
		17	Nup188	AO AS	1655	-	1-11, 12-34, 35-39, 40-91, 92-100, 101-123, 124-130, 131-166, 167-173, 174-224, 225-255, 256-282, 283-287, 288-304, 305-317, 318-434, 435-438, 439-479, 480-492, 493-508, 509-514, 515-530, 531-550, 551-577, 578-583, 584-605, 606-607, 608-619, 620-631, 632-785, 786-792, 793-889, 890-891, 892-1100, 1101-1118, 1119-1133, 1134-1156, 1157-1241, 1242-1246, 1247-1265, 1266-1275, 1276-1292, 1293-1302, 1303-1322, 1323-1331, 1332-1354, 1355-1382, 1383-1567, 1568-1592, 1593-1628, 1629-1632, 1633-1652, 1653-1655	100.00 / 82.18	Multiscale: Coarse-grained: 1 - 25 residue(s) per bead
		18	Nup192	AP AT	1683	-	1-362, 363-416, 417-574, 575-601, 602-798, 799-813, 814-849, 850-856, 857-953, 954-960, 961-1126, 1127-1136, 1137-1226, 1227-1233, 1234-1258, 1259-1271, 1272-1366, 1367-1370, 1371-1418, 1419-1420, 1421-1502, 1503-1510, 1511-1559, 1560-1583, 1584-1590, 1591-1596, 1597-1619, 1620-1622, 1623-1644, 1645-1650, 1651-1683	100.00 / 88.53	Multiscale: Coarse-grained: 1 - 25 residue(s) per bead
		19	Nup53	AU AZ	475	-	1-247, 248-284, 285-303, 304-360, 361-475	100.00 / 19.79	Multiscale: Coarse-grained: 1 - 50 residue(s) per bead
		20	Nup59	AV BA	528	-	1-265, 266-302, 303-345, 346-402, 403-528	100.00 / 17.80	Multiscale: Coarse-grained: 1 - 50 residue(s) per bead
		21	Ndc1	AW BB	655	-	1-655	100.00 / 0.00	Multiscale: Coarse-grained: 55 - 100 residue(s) per bead

ID	Model(s)	Entity ID	Molecule name	Chain(s) [auth]	Total residues	Rigid segments	Flexible segments	Model coverage/ Starting model coverage (%)	Scale
		22	Pom34	AX	299	-	1-299	100.00 / 0.00	Multiscale: Coarse-grained: 49 - 50 residue(s) per bead
				BC					
		23	Pom152	AY	1337	-	1-378, 379-472, 473-519, 520-611, 612-615, 616-714, 715-721, 722-818, 819-823, 824-918, 919-930, 931-1026, 1027-1035, 1036-1141, 1142-1149, 1150-1229, 1230-1243, 1244-1337	100.00 / 63.80	Multiscale: Coarse-grained: 1 - 50 residue(s) per bead
				BD					
		24	Nup100	BE	959	816-958	551-815, 959	42.65 / 34.96	Multiscale: Coarse-grained: 1 - 25 residue(s) per bead
				BF					
		25	Nup116	BG	1113	-	751-965, 966-1111, 1112-1113	32.61 / 40.22	Multiscale: Coarse-grained: 1 - 25 residue(s) per bead
				BH					
		27	Gle1	BJ	538	-	1-120	22.30 / 0.00	Multiscale: Coarse-grained: 20 - 50 residue(s) per bead
		28	Nup145	BK	1317	459-605	201-458	30.75 / 36.30	Multiscale: Coarse-grained: 1 - 25 residue(s) per bead
				BL					

ID	Model(s)	Entity ID	Molecule name	Chain(s) [auth]	Total residues	Rigid segments	Flexible segments	Model coverage/ Starting model coverage (%)	Scale
		29	Nup1	BM	1076	-	1-351	32.62 / 0.00	Multiscale: Coarse-grained: 1 - 50 residue(s) per bead
		30	Nup60	BN BO	539	-	1-398	73.84 / 0.00	Multiscale: Coarse-grained: 48 - 50 residue(s) per bead
		31	Mlp1	BP	1875	-	238-716	25.55 / 0.00	Multiscale: Coarse-grained: 29 - 50 residue(s) per bead
		32	Mlp2	BQ	1679	-	215-690	28.35 / 0.00	Multiscale: Coarse-grained: 26 - 50 residue(s) per bead

2.3. Datasets used for modeling

There are 65 unique datasets used to build the models in this entry.

ID	Dataset type	Database name	Data access code
1	Integrative model	Not available	10.1016/j.cell.2016.10.028
2	Integrative model	Zenodo	10.5281/zenodo.1194547
3	Experimental model	PDB	pdb_00005cws
4	Comparative model	Zenodo	10.5281/zenodo.1194547
5	Experimental model	PDB	pdb_00002qx5
6	Experimental model	Zenodo	10.5281/zenodo.1194547
7	Experimental model	Zenodo	10.5281/zenodo.1194547
8	Comparative model	Zenodo	10.5281/zenodo.1194547
9	Comparative model	Zenodo	10.5281/zenodo.1194547

ID	Dataset type	Database name	Data access code
10	Comparative model	Zenodo	10.5281/zenodo.1194547
11	Comparative model	Zenodo	10.5281/zenodo.1194547
12	Comparative model	Zenodo	10.5281/zenodo.1194547
13	Comparative model	Zenodo	10.5281/zenodo.1194547
14	Integrative model	Not available	10.1016/j.str.2017.01.006
15	Integrative model	Zenodo	10.5281/zenodo.1194547
16	Experimental model	PDB	pdb_00003nf5
17	Comparative model	Zenodo	10.5281/zenodo.1194547
18	Comparative model	Zenodo	10.5281/zenodo.1194547
19	Experimental model	PDB	pdb_00003kep
20	Experimental model	Zenodo	10.5281/zenodo.1194547
21	Mass Spectrometry data	Zenodo	10.5281/zenodo.1149746
22	Crosslinking-MS data	Zenodo	10.5281/zenodo.1194547
23	Crosslinking-MS data	Zenodo	10.5281/zenodo.1194547
24	EM raw micrographs	EMPIAR	EMPIAR-10155
25	3DEM volume	EMDB	EMD-7321
26	3DEM volume	Zenodo	10.5281/zenodo.1194547
27	SAS data	Zenodo	10.5281/zenodo.1194547
28	SAS data	Zenodo	10.5281/zenodo.1194547
29	SAS data	Zenodo	10.5281/zenodo.1194547
30	SAS data	Zenodo	10.5281/zenodo.1194547
31	SAS data	Zenodo	10.5281/zenodo.1194547
32	SAS data	Zenodo	10.5281/zenodo.1194547
33	SAS data	Zenodo	10.5281/zenodo.1194547
34	SAS data	Zenodo	10.5281/zenodo.1194547
35	SAS data	Zenodo	10.5281/zenodo.1194547
36	SAS data	Zenodo	10.5281/zenodo.1194547
37	SAS data	Zenodo	10.5281/zenodo.1194547
38	SAS data	Zenodo	10.5281/zenodo.1194547
39	SAS data	Zenodo	10.5281/zenodo.1194547
40	SAS data	Zenodo	10.5281/zenodo.1194547
41	SAS data	Zenodo	10.5281/zenodo.1194547
42	SAS data	Zenodo	10.5281/zenodo.1194547
43	SAS data	Zenodo	10.5281/zenodo.1194547
44	SAS data	Zenodo	10.5281/zenodo.1194547

ID	Dataset type	Database name	Data access code
45	SAS data	Zenodo	10.5281/zenodo.1194547
46	SAS data	Zenodo	10.5281/zenodo.1194547
47	SAS data	Zenodo	10.5281/zenodo.1194547
48	SAS data	SASBDB	SASDBV9
49	SAS data	SASBDB	SASDBW9
50	SAS data	SASBDB	SASDBZ9
51	SAS data	SASBDB	SASDBX9
52	SAS data	SASBDB	SASDBY9
53	SAS data	Zenodo	10.5281/zenodo.1194547
54	SAS data	Zenodo	10.5281/zenodo.1194547
55	SAS data	Zenodo	10.5281/zenodo.1194547
56	SAS data	Zenodo	10.5281/zenodo.1194547
57	SAS data	Zenodo	10.5281/zenodo.1194547
58	SAS data	Zenodo	10.5281/zenodo.1194547
59	SAS data	Zenodo	10.5281/zenodo.1194547
60	SAS data	Zenodo	10.5281/zenodo.1194547
61	SAS data	Zenodo	10.5281/zenodo.1194547
62	SAS data	Zenodo	10.5281/zenodo.1194547
63	EM raw micrographs	EMPIAR	EMPIAR-10162
64	2DEM class average	Zenodo	10.5281/zenodo.1194547
65	2DEM class average	Zenodo	10.5281/zenodo.1194547

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	Sampling	Replica exchange monte carlo	Not available	500	False	True
2	1	Sampling	Replica exchange monte carlo	Not available	3000	False	True
3	1	Sampling	Replica exchange monte carlo	Not available	1000	False	True

There are 13 software packages reported in this entry.

ID	Software name	Software version	Software classification	Software location
1	Integrative Modeling Platform (IMP)	develop-0a5706e202	integrative model building	https://integrativemodeling.org
2	IMP PMI module	67456c0	integrative model building	https://integrativemodeling.org

ID	Software name	Software version	Software classification	Software location
3	HHpred	2.0.16	protein homology detection	https://toolkit.tuebingen.mpg.de/hhpred
4	PSIPRED	4.00	secondary structure prediction	http://bioinf.cs.ucl.ac.uk/psipred/
5	DISOPRED	3	disorder prediction	http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1
6	DomPred	Not available	domain boundary prediction	http://bioinf.cs.ucl.ac.uk/dompred
7	COILS/PCOILS	Not available	coiled-coil prediction	https://toolkit.tuebingen.mpg.de/#/tools/pcoils
8	EMAN2	2.20	image processing	http://blake.bcm.edu/emanwiki/EMAN2
9	RELION	1.40	image processing	https://www2.mrc-lmb.cam.ac.uk/relion/
10	SGD	Not available	database	https://www.yeastgenome.org/
11	HeliQuest	Not available	helix prediction	http://heliquest.ipmc.cnrs.fr/
12	MODELLER	9.15	comparative modeling	https://salilab.org/modeller/
13	MODELLER	9.13	comparative modeling	https://salilab.org/modeller/

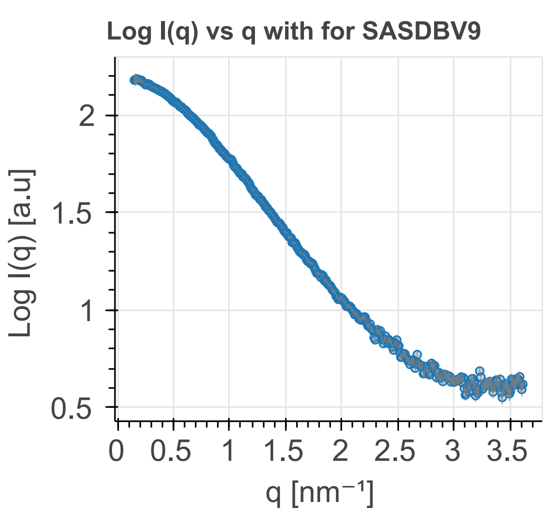
3. Data quality

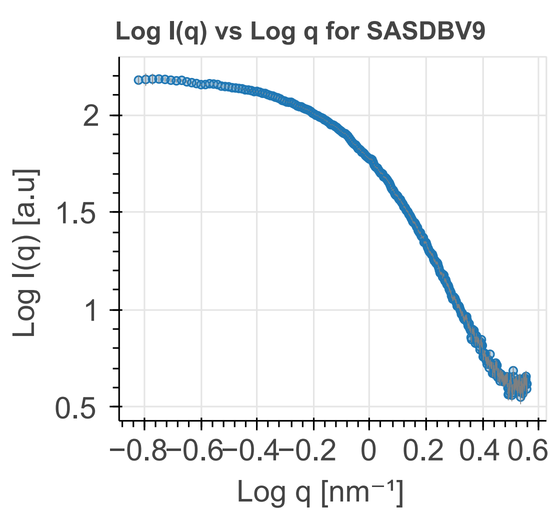
3.1. SAS

3.1.1. Scattering profile

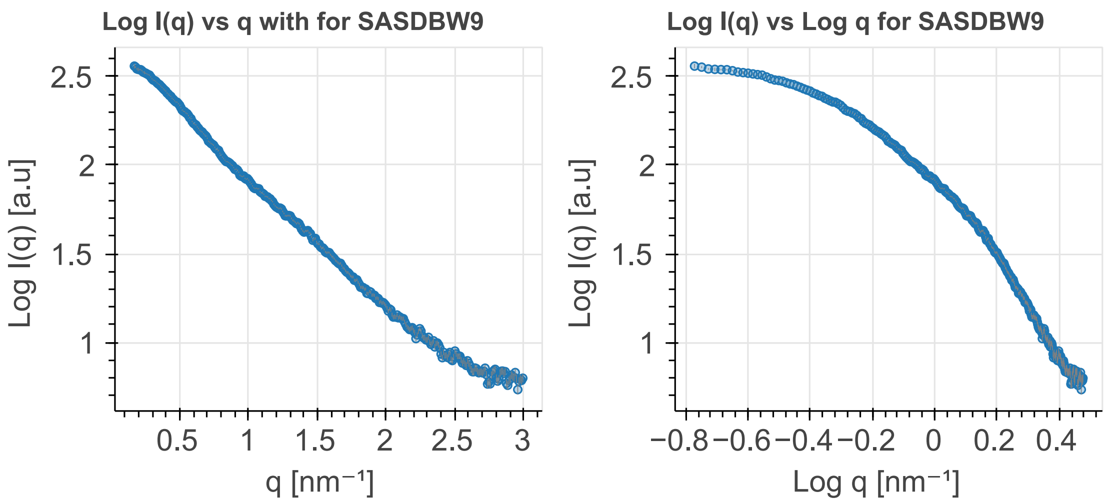
SAS data used in this integrative model was obtained from 5 deposited SASBDB entry (entries).

Scattering profile for SASDBV9: 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.

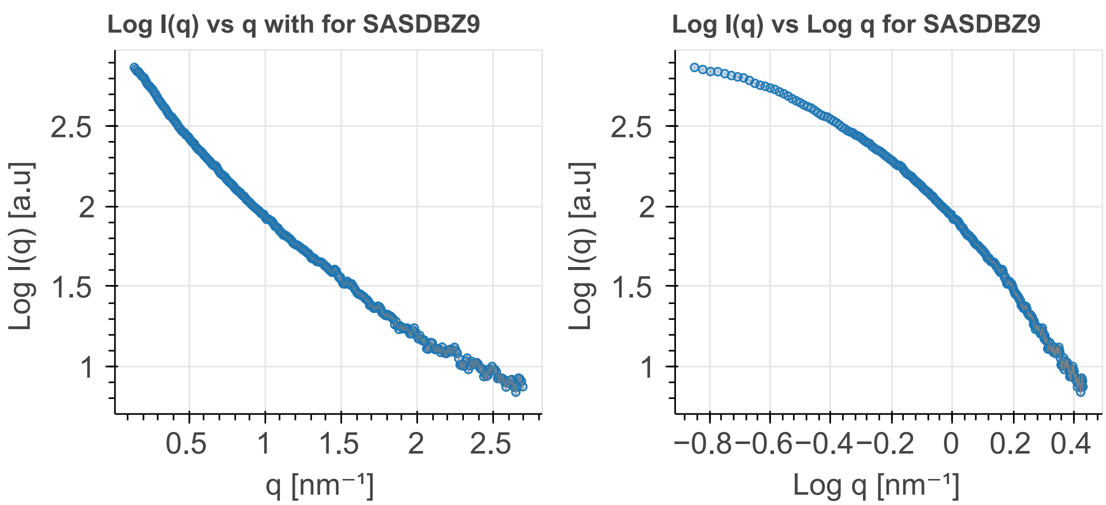




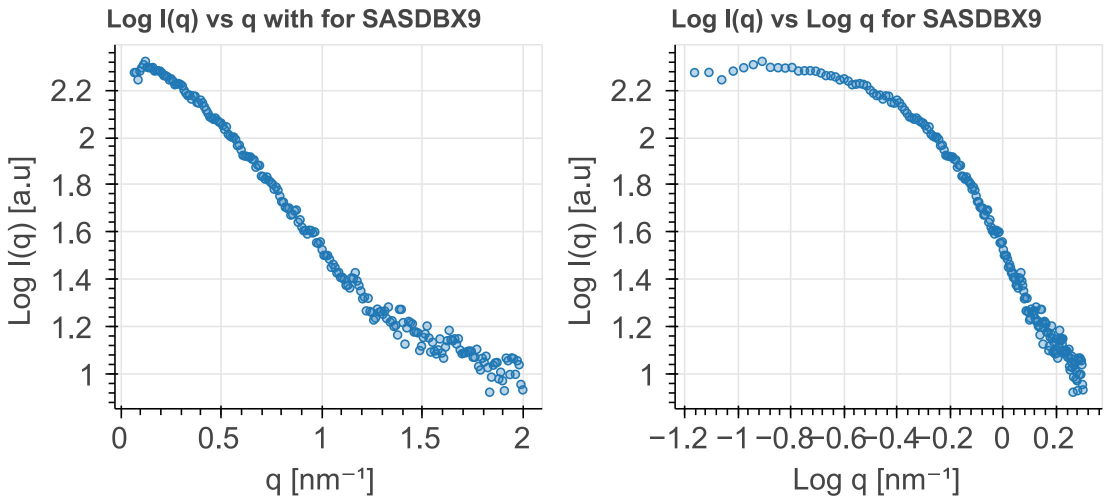
Scattering profile for SASDBW9: data from solutions of biological macromolecules are presented as both log I(q) vs q and log I(q) vs log (q) based on SAS validation task force (SASvtf) recommendations. I(q) is the intensity (in arbitrary units) and q is the modulus of the scattering vector.



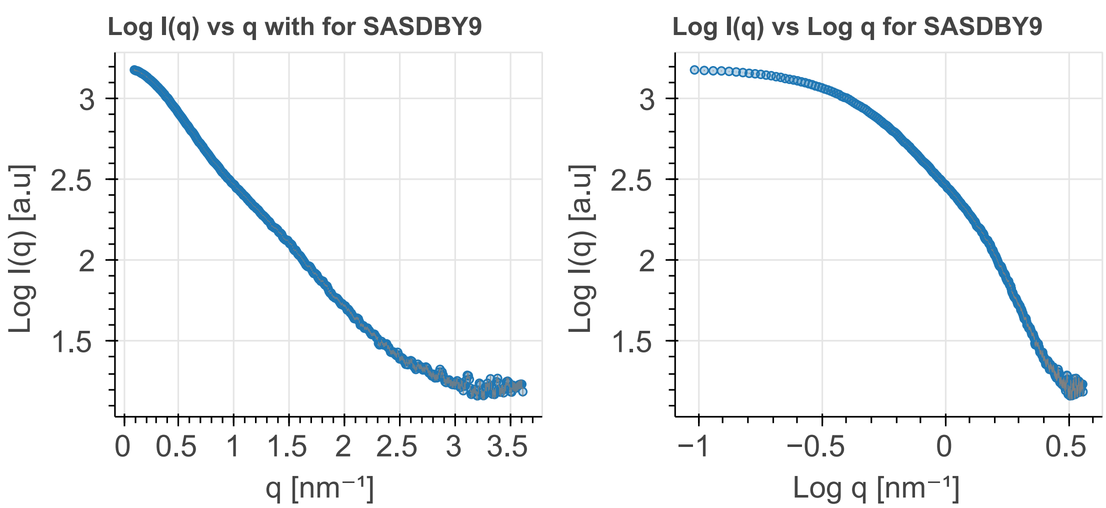
Scattering profile for SASDBZ9: data from solutions of biological macromolecules are presented as both log I(q) vs q and log I(q) vs log (q) based on SAS validation task force (SASvtf) recommendations. I(q) is the intensity (in arbitrary units) and q is the modulus of the scattering vector.



Scattering profile for SASDBX9: 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.



Scattering profile for SASDBY9: 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.



3.1.2. Key experimental estimates

Molecular weight (MW) estimates from experiments and analysis: Theoretical MW can be compared to SAS-derived values using the forward scatter (*I(0)*) and the known concentration and partial specific volume of the scattering particle, or as estimated from the Porod volume and partial specific volume (Trewhella et al., 2017, Trewhella et al., 2023).

SASDB ID	Chemical composition MW	Standard MW	Porod Volume/MW
SASDBV9	12.6 kDa	12.2 kDa	Not available
SASDBW9	24.1 kDa	25.2 kDa	Not available
SASDBZ9	49.4 kDa	48.3 kDa	Not available
SASDBX9	12.5 kDa	14.7 kDa	Not available
SASDBY9	25.9 kDa	25.2 kDa	Not available

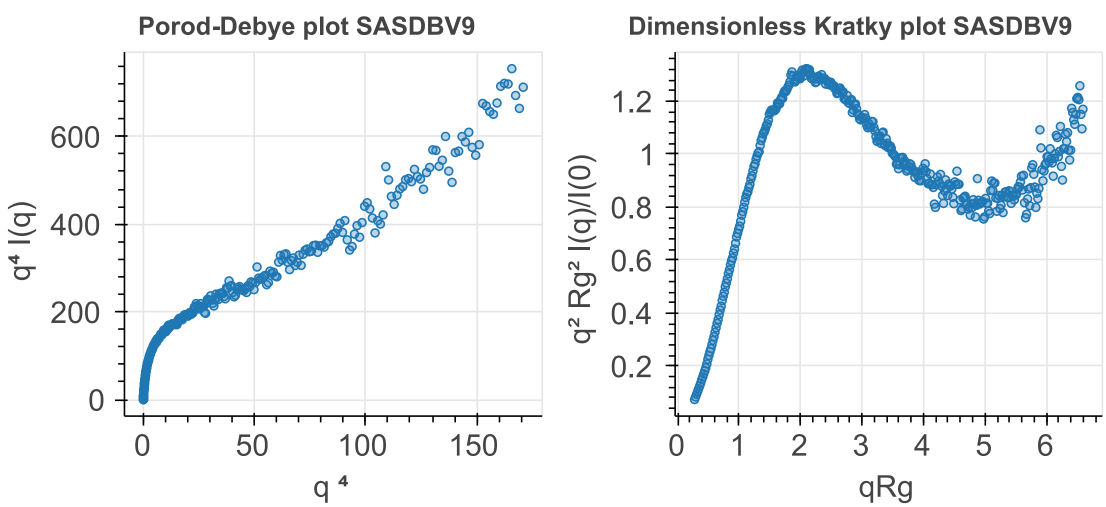
Volume estimates from experiments and analysis: estimated volume can be compared to Porod volume obtained from scattering profiles.

SASDB ID	Estimated Volume	Porod Volume	Specific Volume	Sample Contrast	Sample Concentration
SASDBV9	Not available	17.94 nm³	Not available	Not available	Not available
SASDBW9	Not available	22.50 nm³	Not available	Not available	Not available
SASDBZ9	Not available	66.59 nm³	Not available	Not available	Not available
SASDBX9	Not available	56.68 nm³	Not available	Not available	Not available
SASDBY9	Not available	27.97 nm³	Not available	Not available	Not available

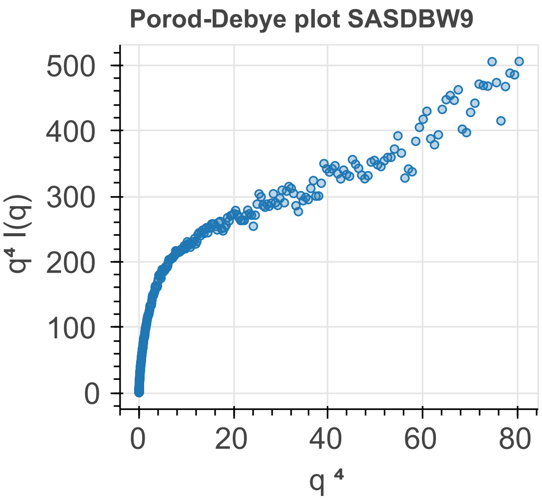
3.1.3. Flexibility analysis

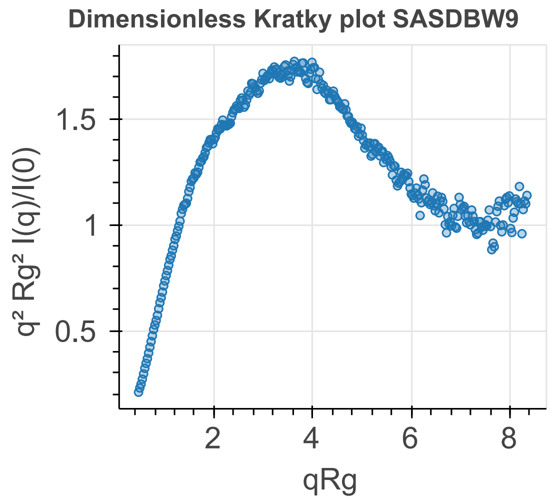
In a Porod-Debye plot, a clear plateau is observed for globular (partial or fully folded) domains, whereas flexible-modular, fully unfolded domains or extended/stiff rodshaped domains lack a discernible plateau (Rambo and Tainer 2013). A bell-shaped Kratky plot ($q^2I(q)$ *vs.* q) with a well-defined maximum is observed for compact/folded structures. For partially flexible/modular or extended structures the Kratky plot can show multiple maxima and/or an increase in intensity at higher q-values depending on the degree of flexibility and extension. Fully intrinsically disordered structures yield a Kratky plot that systematically increases with increasing q values and will be near linear for highly extended molecules. The dimensionless Kratky plot ($(qR_g)^2I(q)$ *vs.* qR_g) is useful for quantifying differences in shape and foldedness among scattering objects of different sizes (Trewhella et. al., 2023).

Flexibility analysis for SASDBV9

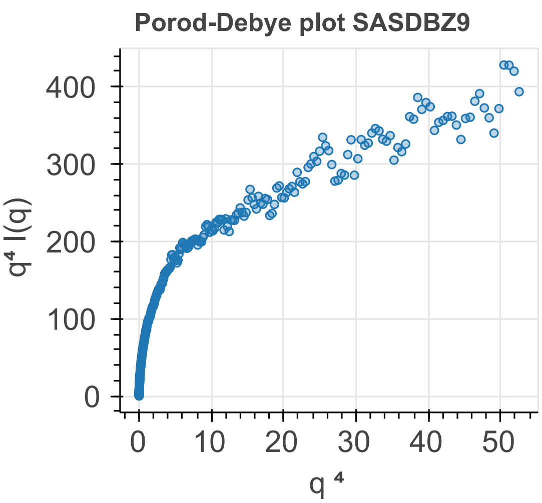


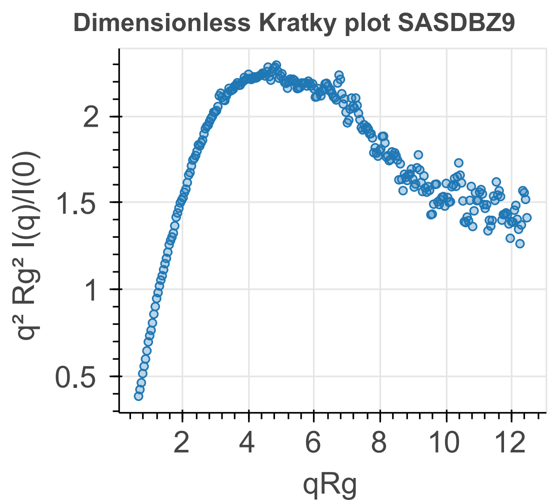
Flexibility analysis for SASDBW9



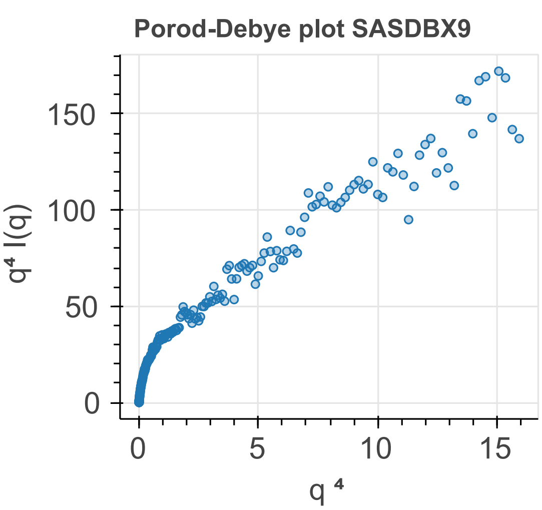


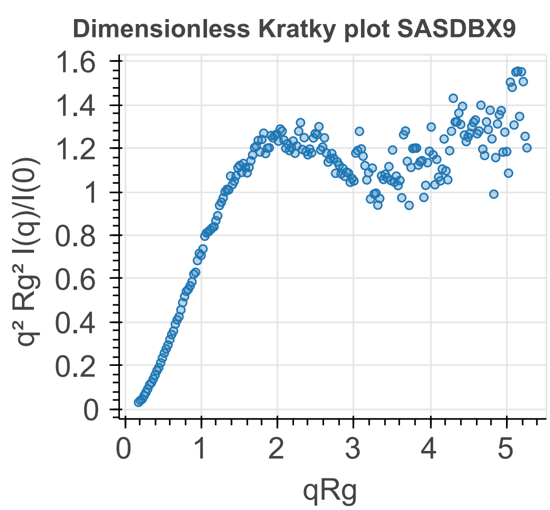
Flexibility analysis for SASDBZ9



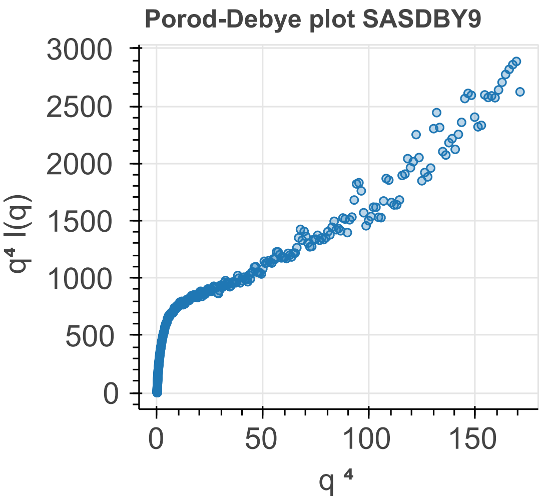


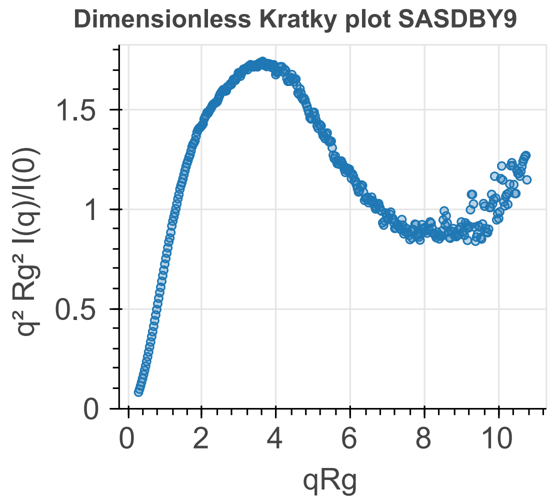
Flexibility analysis for SASDBX9





Flexibility analysis for SASDBY9



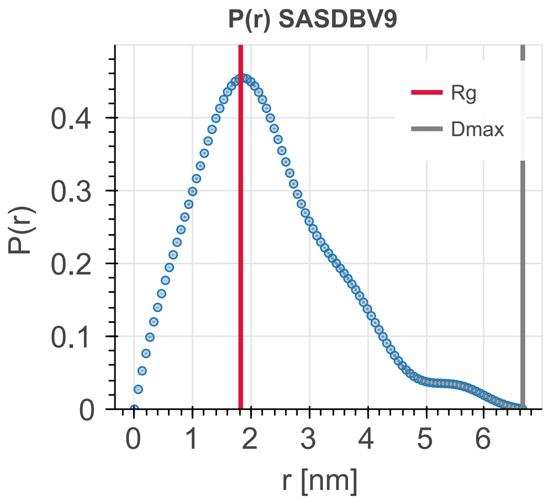


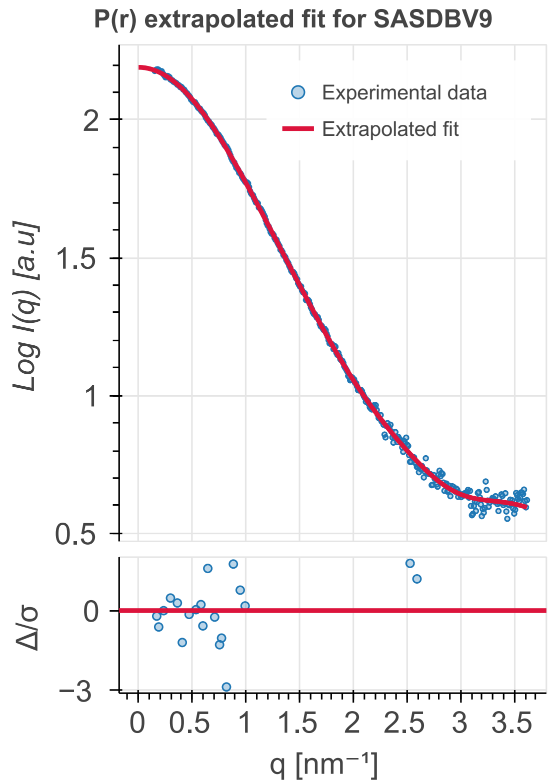
3.1.4. Pair-distance distribution analysis

The the atom-pair distance distribution function (PDDF) or *P(r)* represents the distribution of distances between all pairs of atoms within the particle weighted by the respective scattering contrasts (Moore, 1980). The second moment of *P(r)* yields the radius of gyration (Rg), which is a measure of the overall size and shape of a macromolecule (i.e. the spatial distribution of volume elements). 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.

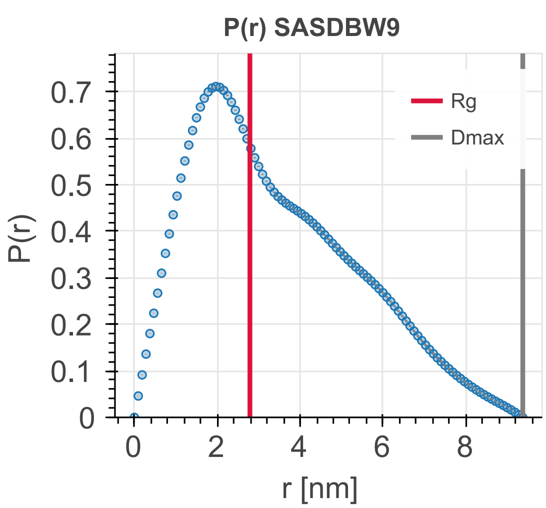
SASDB ID	Software used	D_{max}	D_{max} error	R_g	R_g error
SASDBV9	GNOM 4.5a	6.660 nm	Not available	1.824 nm	0.006 nm
SASDBW9	GNOM 4.5a	9.370 nm	Not available	2.787 nm	0.007 nm
SASDBZ9	GNOM 4.5a	15.430 nm	Not available	4.629 nm	0.011 nm
SASDBX9	GNOM 4.5a	7.930 nm	Not available	2.636 nm	0.008 nm
SASDBY9	GNOM 4.5a	10.450 nm	Not available	2.976 nm	0.005 nm

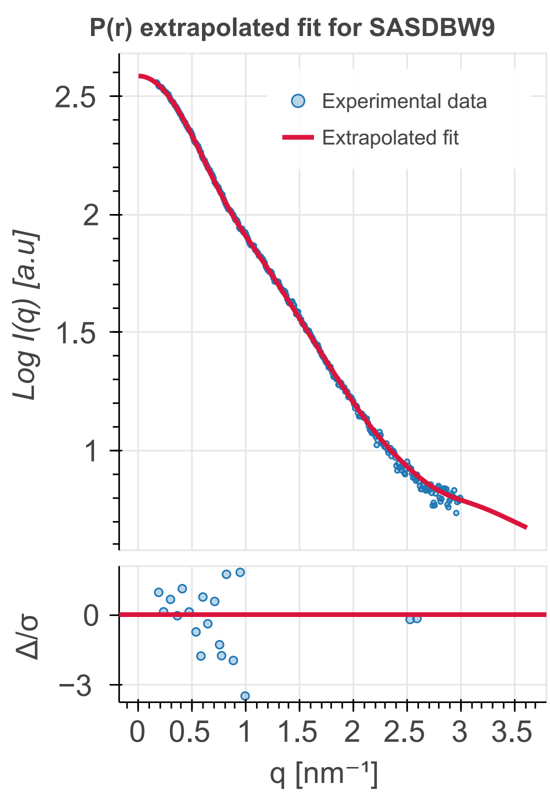
P(r) for SASDBV9: The value of P(r) should be zero beyond $r=D_{max}$.



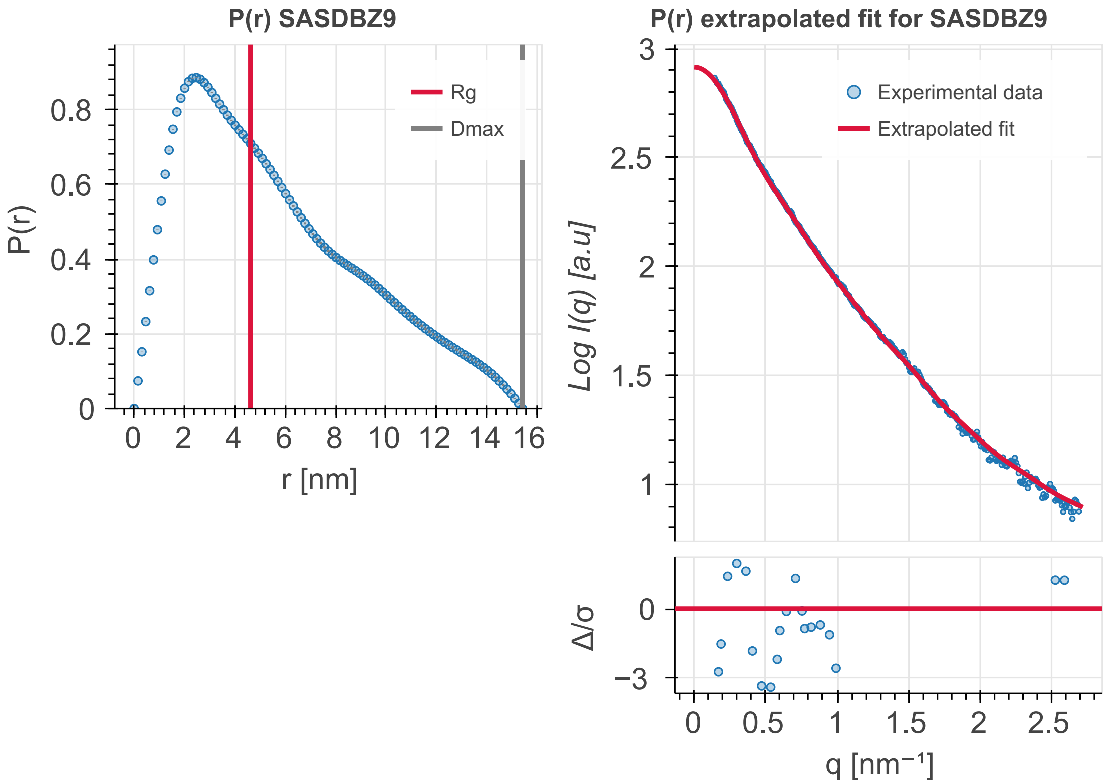


P(r) for SASDBW9: The value of P(r) should be zero beyond r=D_{max}.

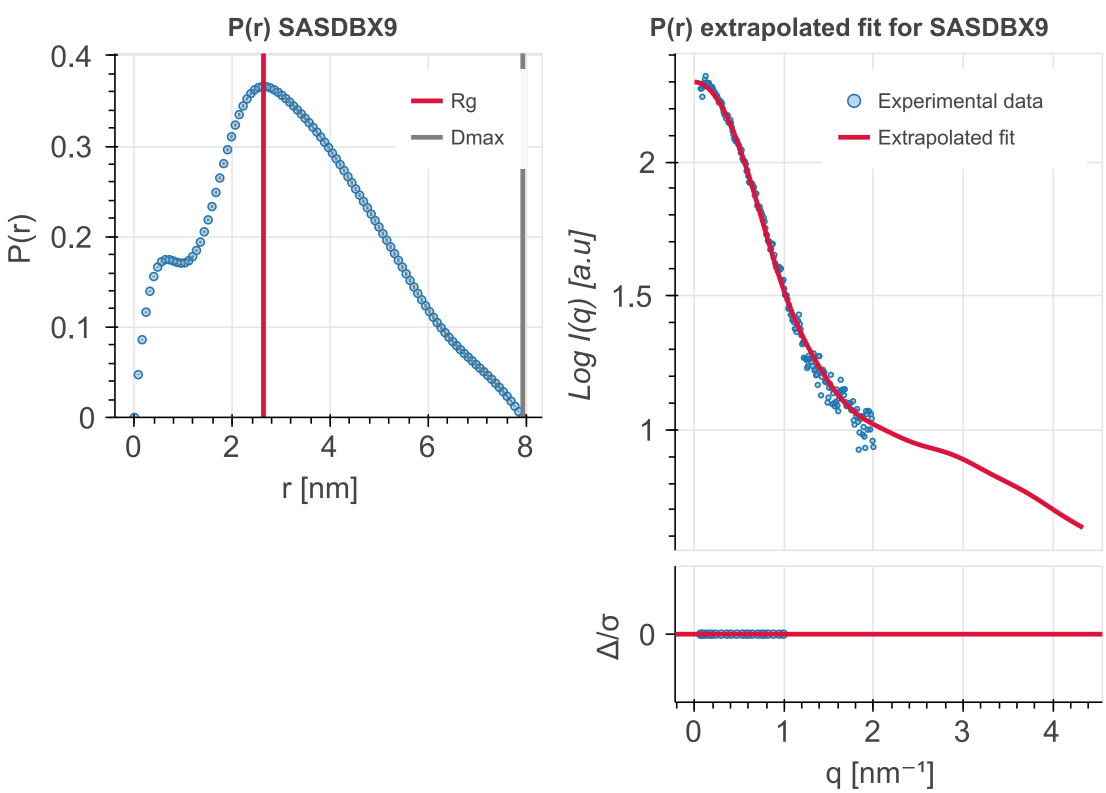




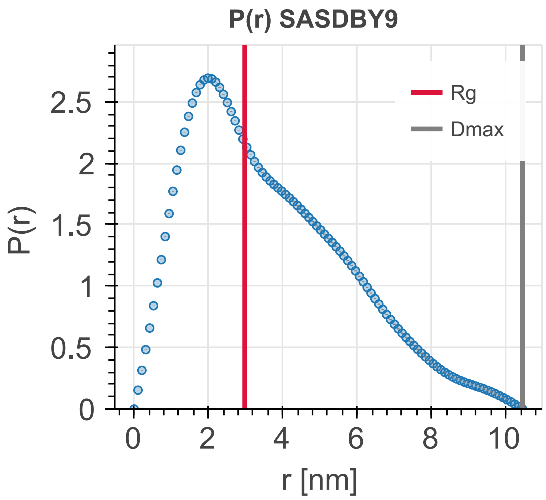
P(r) for SASDBZ9: The value of P(r) should be zero beyond r=D_{max}.

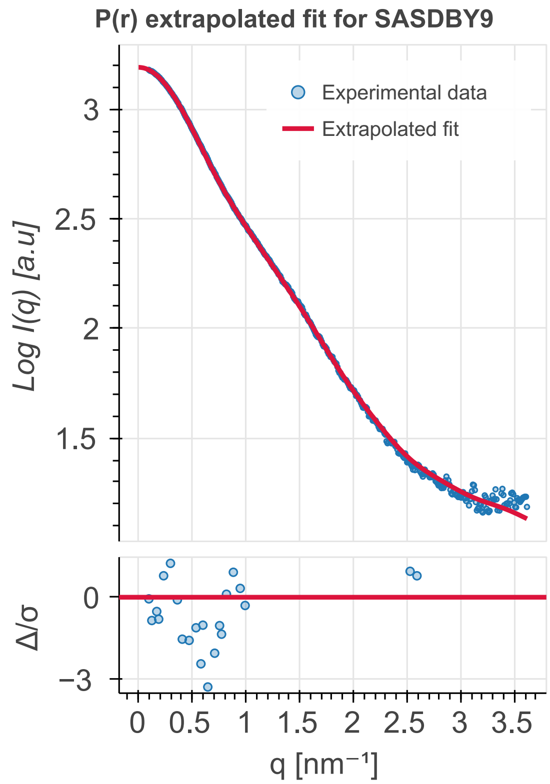


P(r) for SASDBX9: The value of P(r) should be zero beyond r=D_{max}.



P(r) for SASDBY9: The value of P(r) should be zero beyond r=D_{max}.

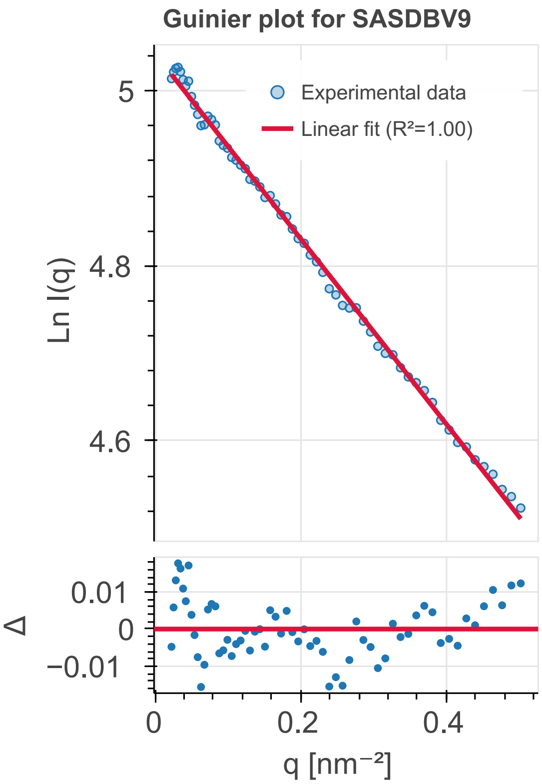


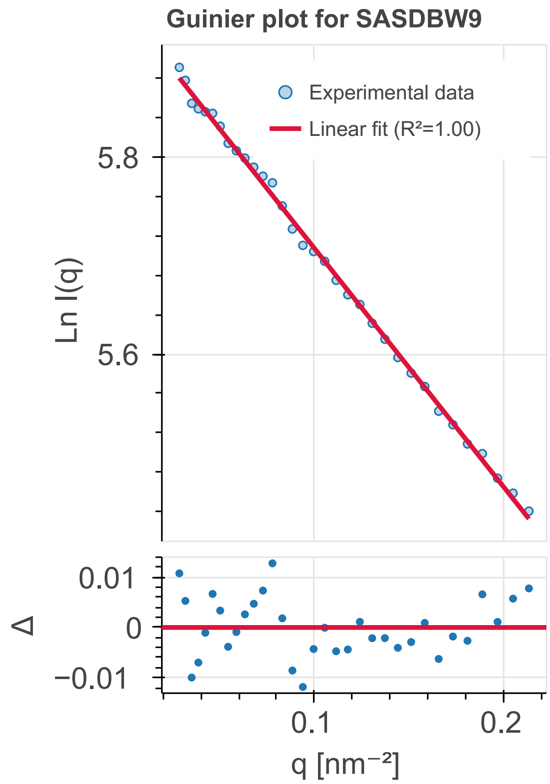


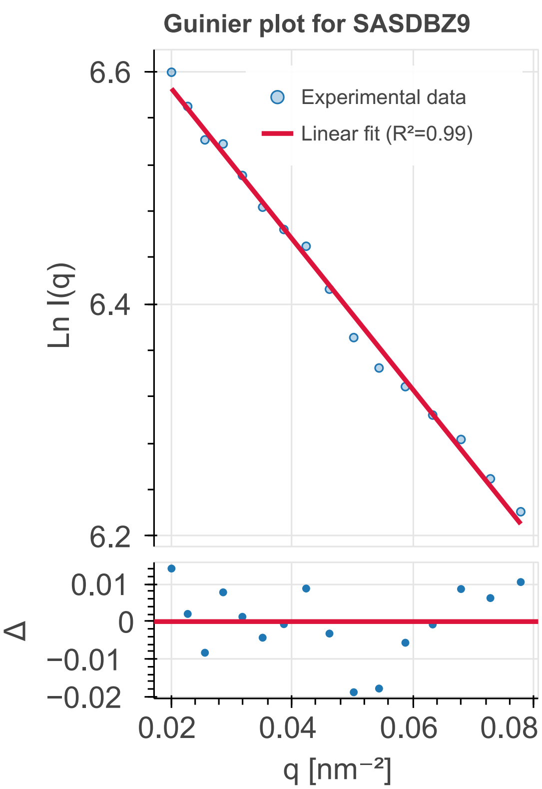
3.1.5. Guinier analysis

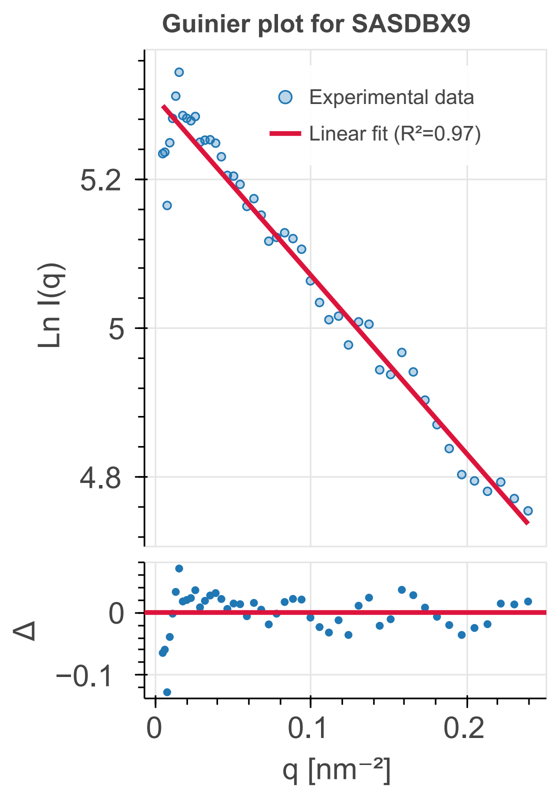
The linearity of the Guinier plot (ln(q) *vs.* q^2) at very-low angle (qRg < 1.3) is a sensitive indicator of the quality of the sample in relation to its homogeneity; a linear Guinier plot is a necessary but not sufficient demonstration that a solution contains monodisperse particles of the same size. Deviations from linearity can point to strong interference effects from particle attraction or repulsion, polydispersity of the samples, or improper background subtraction (Feigin et al., 2013). Residual difference plots and Pearson correlation coefficient determination (R^2) are measures to assess quality of the linear fit to the Guinier region. A perfect fit has an R^2 value of 1. Residual values should be equally and randomly spaced around the horizontal axis with no evident systematic upward or downward curvature.

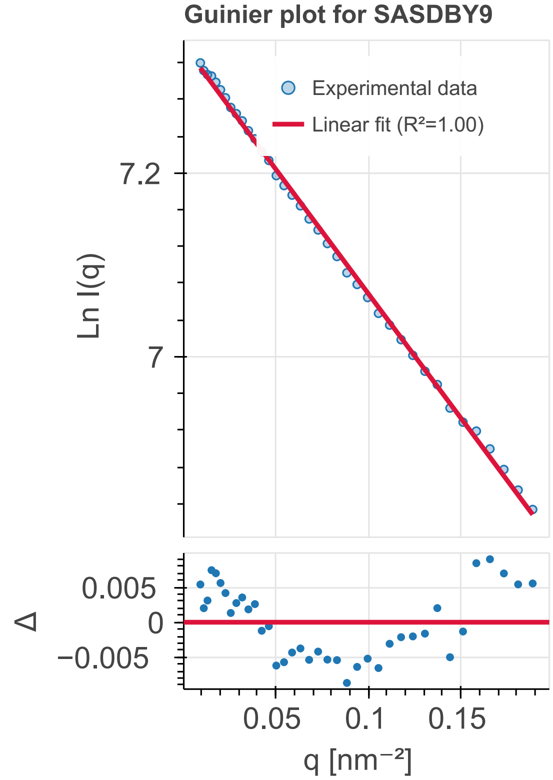
SASDB ID	R_g	R_g error	MW	MW error
SASDBV9	1.77 nm	0.05 nm	12.2 kDa	Not available
SASDBW9	2.71 nm	0.06 nm	25.2 kDa	Not available
SASDBZ9	4.34 nm	0.17 nm	48.3 kDa	Not available
SASDBX9	2.78 nm	0.18 nm	14.7 kDa	Not available
SASDBY9	2.95 nm	0.11 nm	25.2 kDa	Not available


Guinier plot for SASDBV9
Experimental data
Linear fit (R²=1.00)
Ln I(q)
5
4.8
4.6
Δ
0.01
0
−0.01
0
0.2
0.4
q [nm⁻²]


Guinier plot for SASDBW9
Experimental data
Linear fit (R²=1.00)
Ln I(q)
5.8
5.6
Δ
0.01
0
−0.01
0.1
0.2
q [nm⁻²]


Guinier plot for SASDBZ9
Experimental data
Linear fit (R²=0.99)
Ln I(q)
6.6
6.4
6.2
Δ
0.01
0
−0.01
−0.02
0.02
0.04
0.06
0.08
q [nm⁻²]


Guinier plot for SASDBX9
Experimental data
Linear fit (R²=0.97)
Ln I(q)
5.2
5
4.8
Δ
0
−0.1
0
0.1
0.2
q [nm⁻²]



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.

3.3. 3DEM

This section describes quality of the 3DEM datasets

EMD-7321

3.3.1. Experimental information

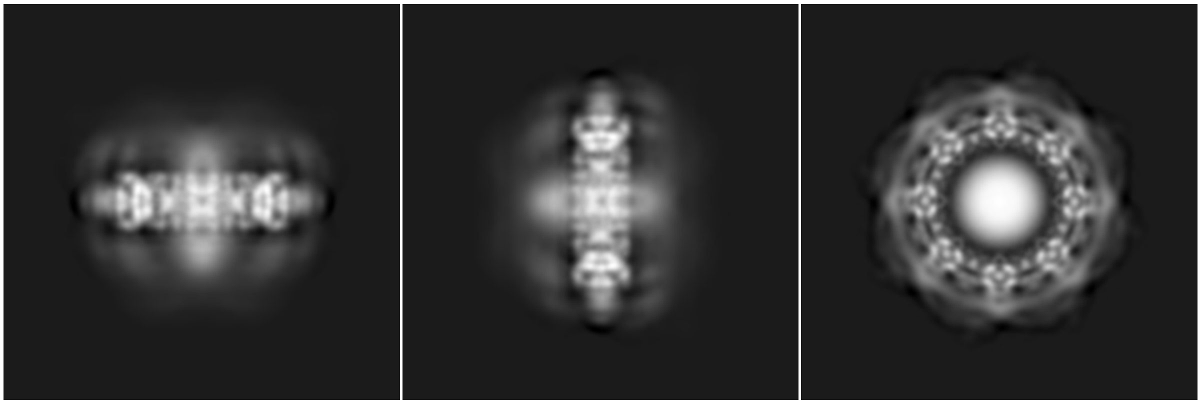
EM reconstruction method:	SUBTOMOGRAM AVERAGING
Resolution:	28.00 Å
Recommended level:	0.015
Estimated volume:	214542.20 nm³
Specimen preparation:	Preparation ID 1 Vitrification
Map-only validation report:	wwPDB validation report

3.3.2. Map visualisation

This section contains visualisations of the EMDB entry EMD-7321. These allow visual inspection of the internal detail of the map and identification of artifacts. Images derived from a raw map, generated by summing the deposited half-maps, are presented below the corresponding image components of the primary map to allow further visual inspection and comparison with those of the primary map.

3.3.2.1. Orthogonal projections

Primary map



X **Y** **Z**

The images above show the map projected in three orthogonal directions.

3.3.2.2. Central slices

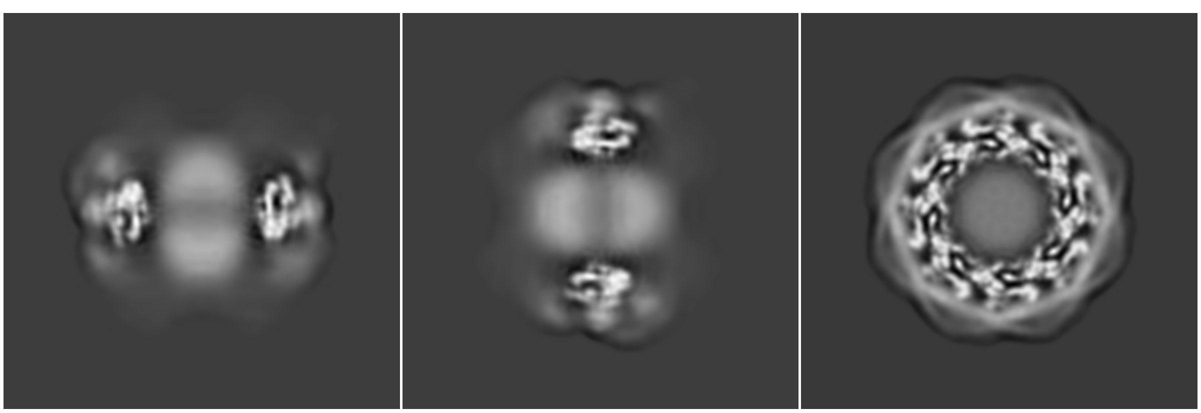
Primary map

X Index: 150 **Y Index: 150** **Z Index: 150**

The images above show central slices of the map in three orthogonal directions.

3.3.2.3. Largest variance slices

Primary map

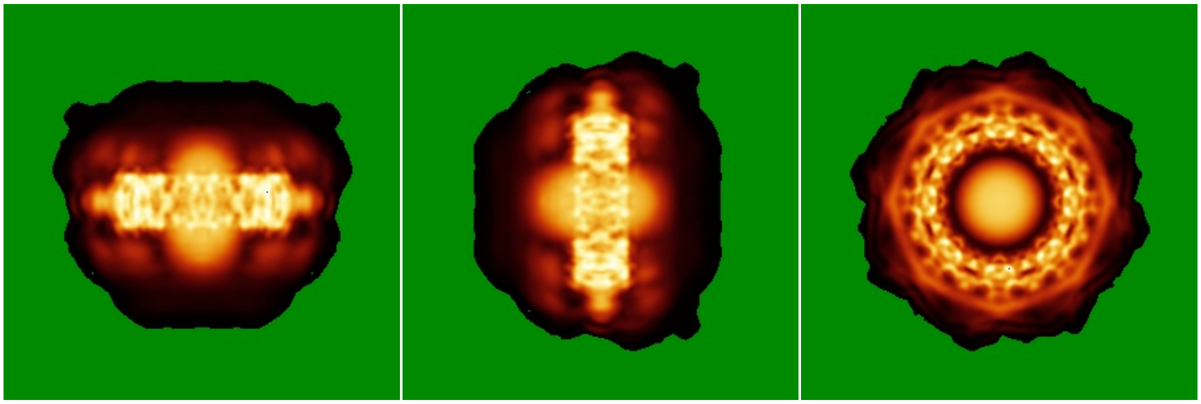


X Index: 144 **Y Index: 156** **Z Index: 154**

The images above show the largest variance slices of the map in three orthogonal directions.

3.3.2.4 Orthogonal standard-deviation projections (false-color)

Primary map

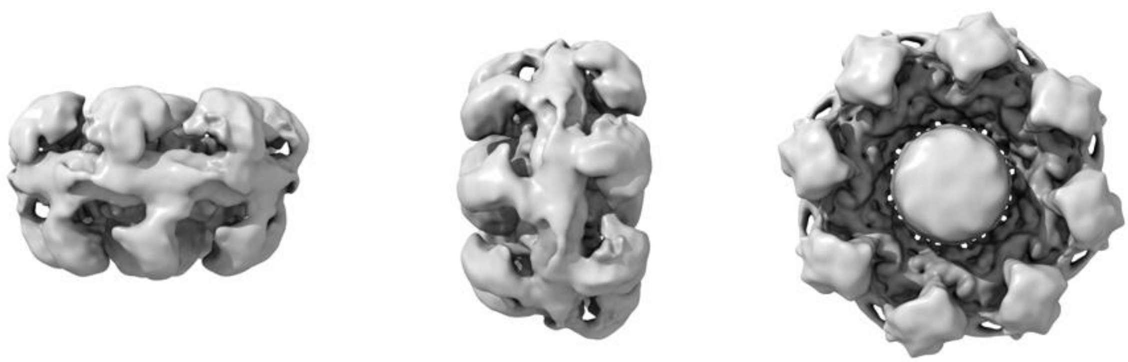


X **Y** **Z**

The images above show the map standard deviation projections with false color in three orthogonal directions. Minimum values are shown in green, max in blue, and dark to light orange shades represent small to large values respectively.

3.3.2.5. Orthogonal surface views

Primary map



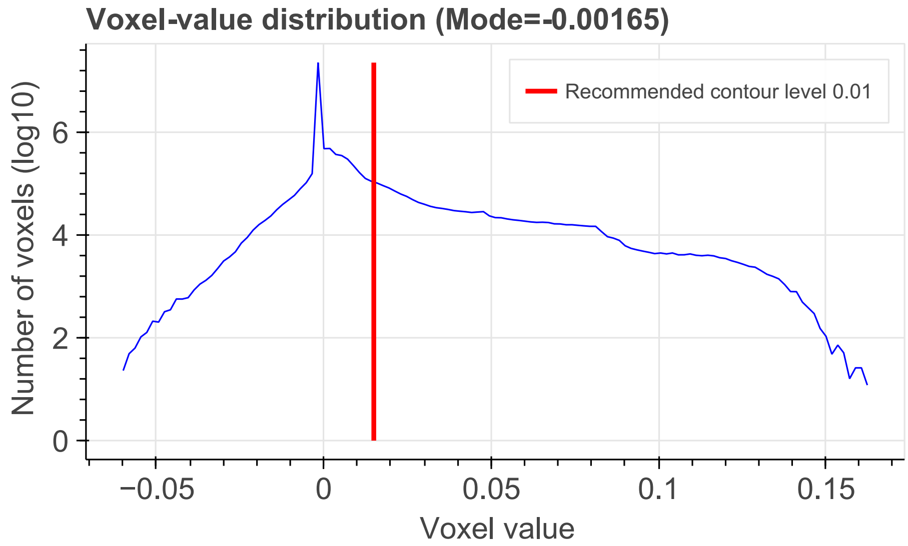
X **Y** **Z**

The images above show the 3D surface view of the map at the recommended contour level 0.015 . These images, in conjunction with the slice images, may facilitate assessment of whether an appropriate contour level has been provided.

3.3.3. Map analysis

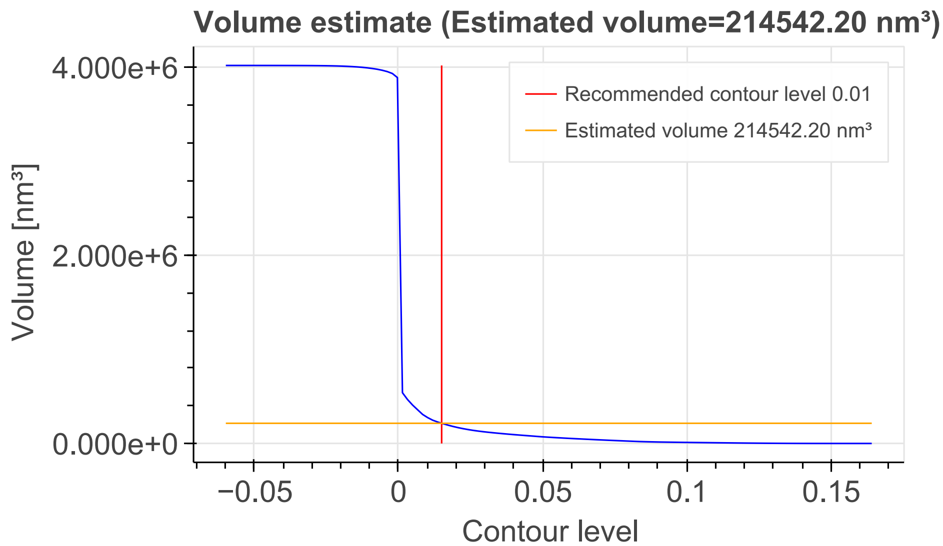
This section contains the results of statistical analysis of the map.

3.3.3.1. Map-value distribution



The map-value distribution is plotted in 128 intervals along the x-axis. The y-axis is logarithmic. A spike in this graph at zero usually indicates that the volume has been masked.

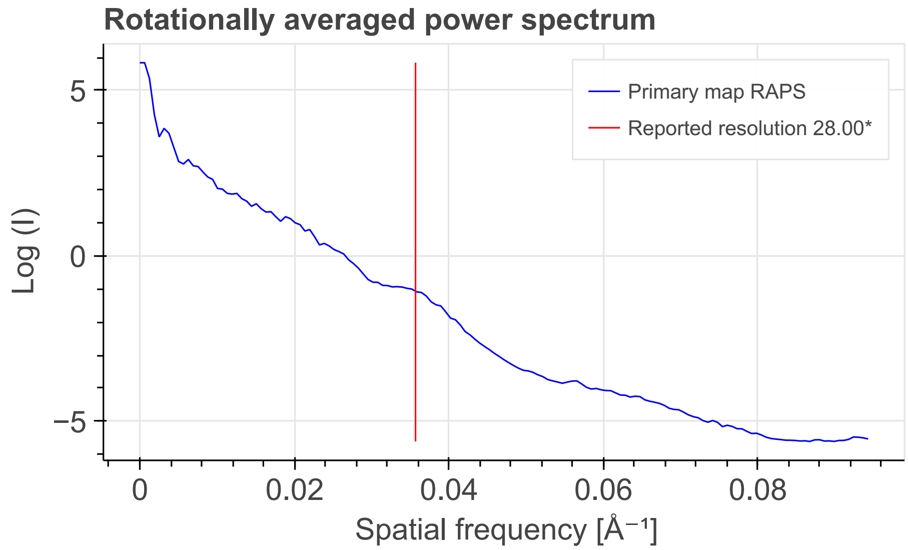
3.3.3.2. Volume estimate ?



The volume at the recommended contour level is 214542.20 nm^3.

The volume estimate graph shows how the enclosed volume varies with the contour level. The recommended contour level is shown as a vertical line and the intersection between the line and the curve gives the volume of the enclosed surface at the given level.

3.3.3.3. Rotationally averaged power spectrum ?



*Reported resolution corresponds to spatial frequency of 0.036 $Å^{-1}$

3.3.4. Fourier-Shell correlation

3.3.4.2. Resolution estimates

Resolution estimate (Å)	Estimation criterion (FSC cut-off)		
	0.143	**0.5**	**Half-bit**
Reported by author	28.00	-	-

Author-provided FSC curve is not available.

3.4. Mass Spectrometry

Validation for this section is under development.

3.4. 2DEM class average

Validation for this section is under development.

3.4. EM raw micrographs

Validation for this section is under development.

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.

4.1a. Excluded Volume Analysis

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

Model ID	Analysed	Number of violations	Excluded Volume Satisfaction (%)
1	428439628	76321	99.98

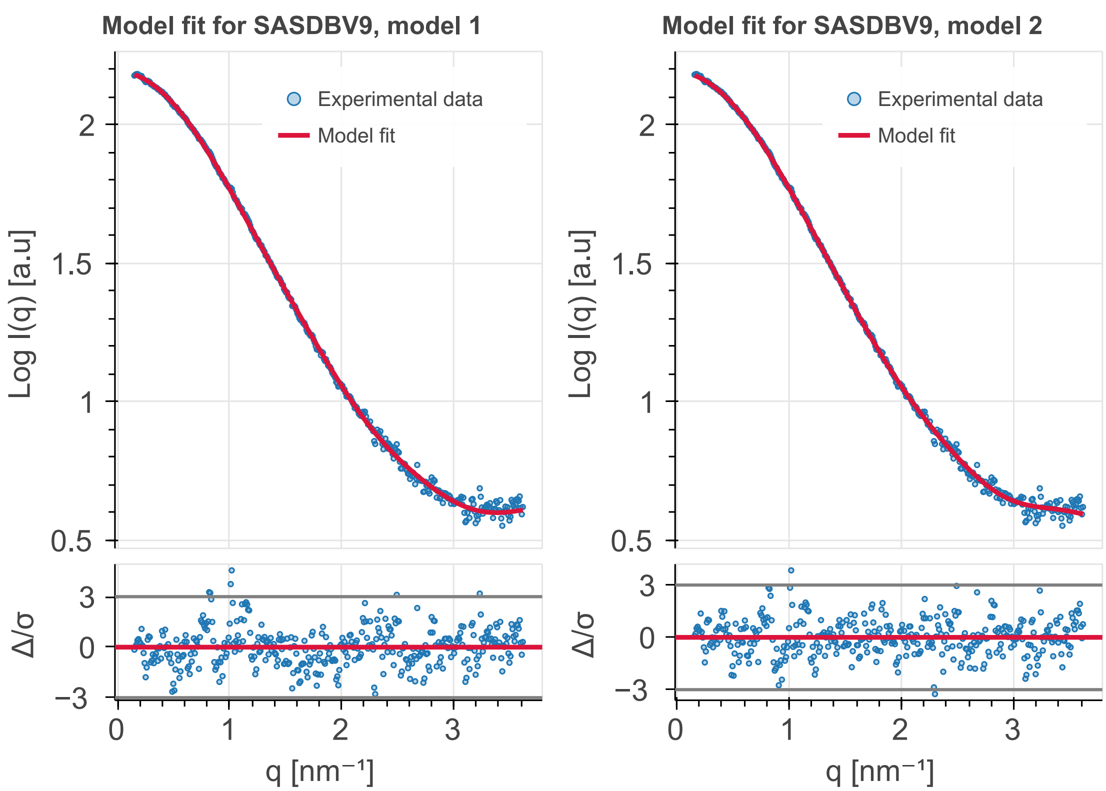
5. Fit to Data Used for Modeling Assessment

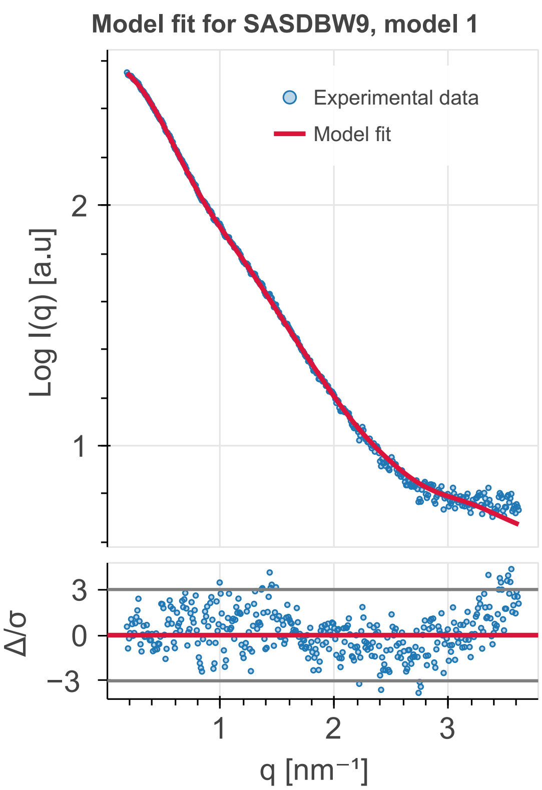
5.1. SAS

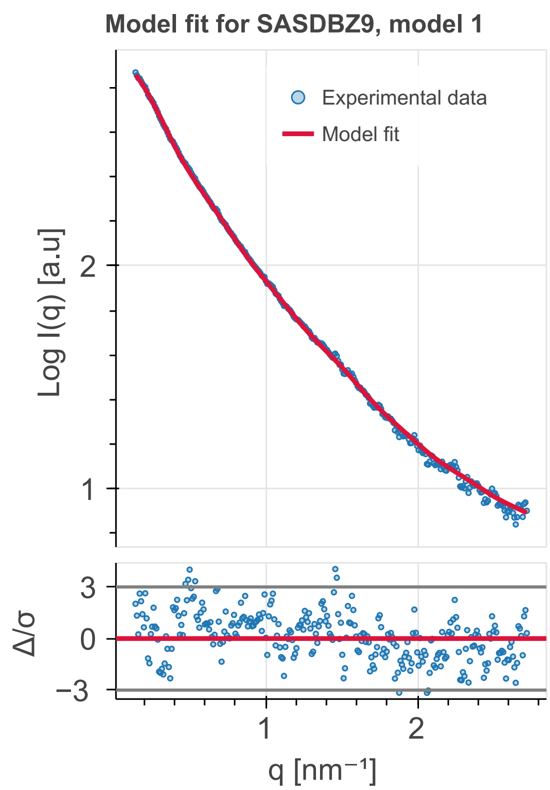
Model and fits displayed below were obtained from SASBDB.

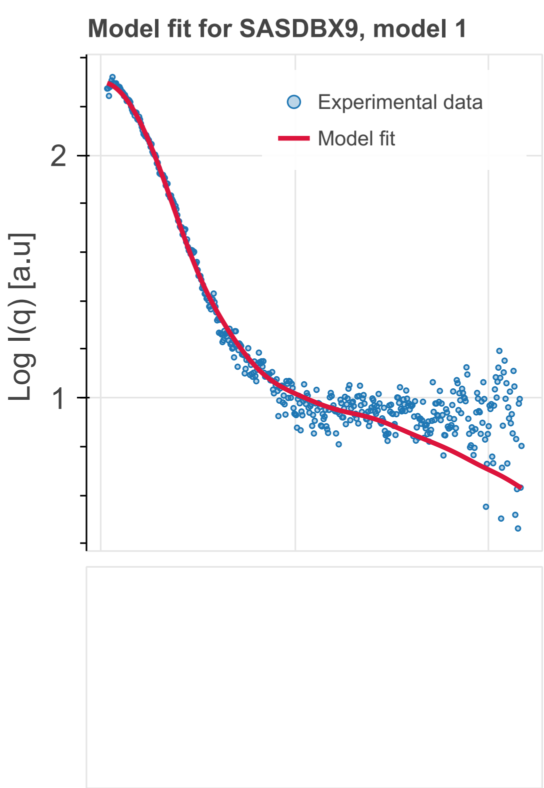
5.1.1 Model versus Experimental Scattering Profiles

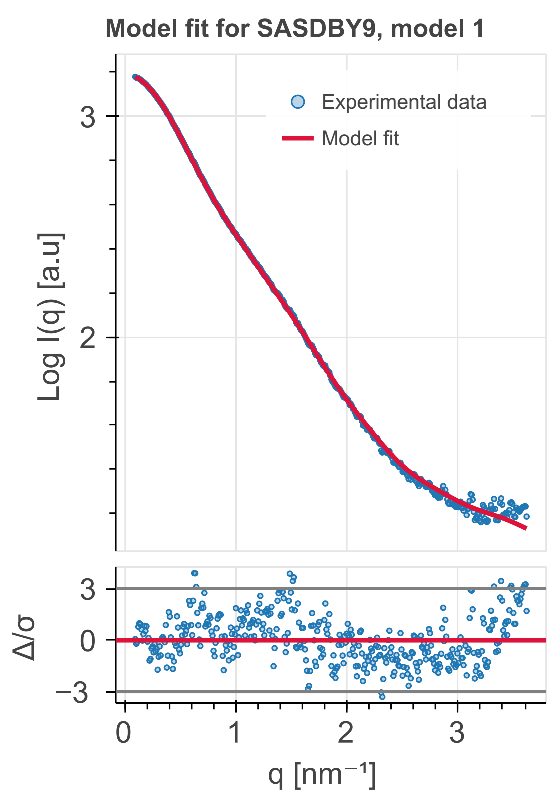
Experimental (blue) and model (red) scattering profiles are presented as log *I(q) vs. q* together with error weighted residual difference plot between the experimental and model *I(q) vs q*. The *I(q)* is the intensity (preferably in absolute units of cm-1 or arbitrary units) and *q* is the modulus of the scattering vector. For a good fit, residual values should be equally and randomly spaced around the horizontal axis largely within +/- three standard deviations.











5.1.2 χ^2 goodness of fit and 5.1.3 cormap analysis

χ^2 values are a measure of the overall fit of the model to the 1D scattering profile. A model that fits the data within its error estimates will have a χ^2 value close to one, provided that the dominant errors are the random statistical errors (i.e. no systematic errors) from the SAS measurement that are correctly propagated. Correlation Map (CorMap) test is a variance-covariance analysis on the scattering intensities comparing two (or more) scattering profiles (e.g. model versus experiment or multiple measures from the same sample). The CorMap test complements χ^2 and importantly is independent of the reported errors. The method assigns a probability (P-value based on a 1-tailed Schilling test) for finding the longest string of experimental data points that lie systematically above (+1) or below (-1) the model profile. The P-value lies between 0 – 1 and a significance threshold is chosen below which the model fit is judged to show systematic deviation from experiment. A typical range statisticians use to indicate significant deviation is 0.01 - 0.05. As implemented in the ATSAS suite, the reported CorMap P-value is green (model fit is good) for P > 0.05, yellow for 0.01 < P < 0.05, and red (model deviates significantly) for P < 0.01.

SASDB ID	Model	χ^2	P-value
SASDBV9	1	1.28	0.02
SASDBV9	2	1.10	0.01
SASDBW9	1	1.97	0.00
SASDBZ9	1	1.94	0.00
SASDBX9	1	2.86	0.00
SASDBY9	1	2.02	0.00

5.2. Crosslinking-MS

5.2.1. Restraint types

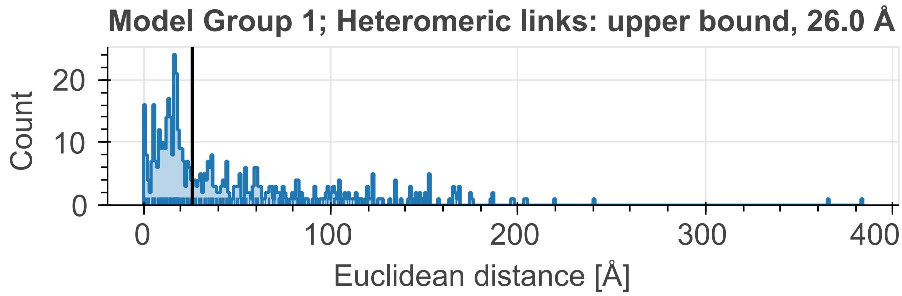
This table summarizes information about crosslinker(s) used for data generation, and how crosslinking information was translated into actual modeling restraints. Restraints assigned "by-residue" are interpreted as between CA atoms. Restraints between coarse-grained beads are indicated as "coarse-grained".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

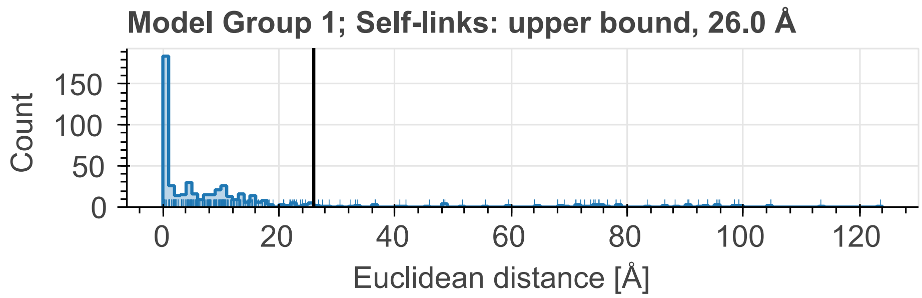
There are 1071 crosslinking restraints combined in 615 restraint groups.

Linker	Residue 1	Atom 1	Residue 2	Atom 2	Restraint type	Distance, Å	Count
DSS	LYS	coarse-grained	LYS	coarse-grained	upper bound	26.00	916
DSS	LYS	CA	LYS	CA	upper bound	26.00	102
DSS	LYS	coarse-grained	THR	coarse-grained	upper bound	26.00	13
DSS	LYS	coarse-grained	MET	coarse-grained	upper bound	26.00	27
DSS	ASN	coarse-grained	LYS	coarse-grained	upper bound	26.00	2
DSS	ARG	coarse-grained	LYS	coarse-grained	upper bound	26.00	2
DSS	GLN	coarse-grained	LYS	coarse-grained	upper bound	26.00	2
DSS	ALA	coarse-grained	LYS	coarse-grained	upper bound	26.00	2
DSS	LYS	coarse-grained	VAL	coarse-grained	upper bound	26.00	3
DSS	GLN	coarse-grained	MET	coarse-grained	upper bound	26.00	1
DSS	LYS	CA	MET	CA	upper bound	26.00	1

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.





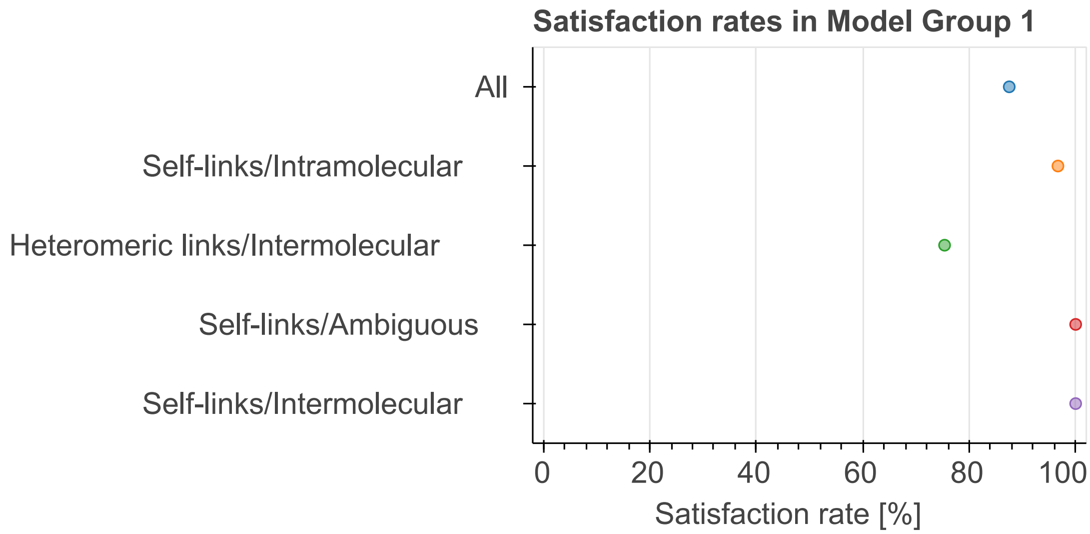
5.2.2. Satisfaction of restraints

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=615)
1	1	1	1/5	All	87.52	12.48	609
				Self-links/ Intramolecular	96.67	3.33	300
				Heteromeric links/ Intermolecular	75.37	24.63	268
				Self-links/ Ambiguous	100.00	0.00	37
				Self-links/ Intermolecular	100.00	0.00	4

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.



5.3. 3DEM

This section describes fit of models to the 3DEM data. Only results for the representative model, selected as a first model with the largest number of asymmetric units.

3DEM validation for coarse-grained structures is under development.

5.4. Mass Spectrometry ?

Validation for this section is under development.

5.4. 2DEM class average ?

Validation for this section is under development.

5.4. EM raw micrographs ?

Validation for this section is under development.

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.