

Integrative Structure Validation Report

July 22, 2024 - 03:36 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 | 8ZZ9 |
| PDB-Dev ID | PDBDEV_00000009 |
| Structure Title | Structure of the human Rev7 dimer |
| Structure Authors | Alessandro A. Rizzo; Faye-Marie Vassel; Nimrat Chatterjee; Sanjay D'Souza; Yunfeng Li; Bing Hao; Michael T. Hemann; Graham C. Walker; Dmitry M. Korzhnev |

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

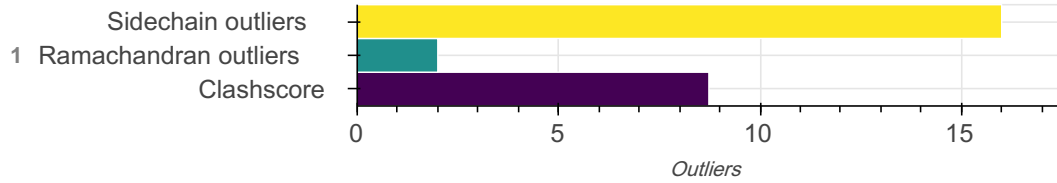
A user guide is available at https://pdb-dev.wwpdb.org/validation_help.html with specific help available everywhere you see the  symbol.

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

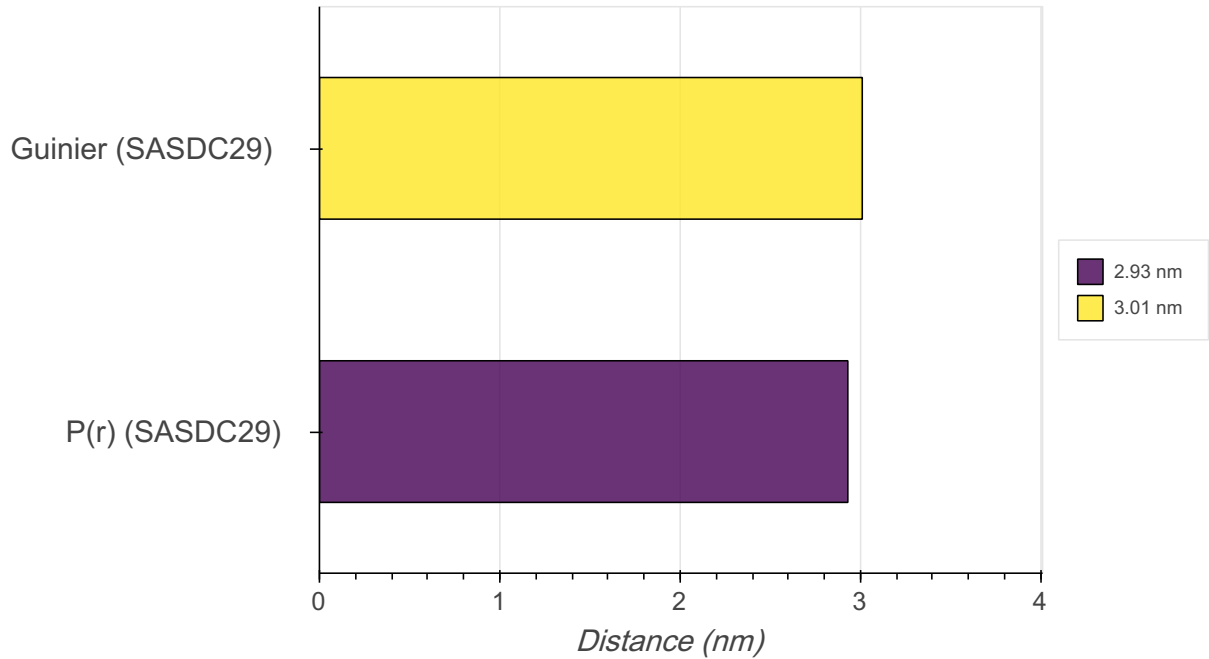
Overall quality

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.

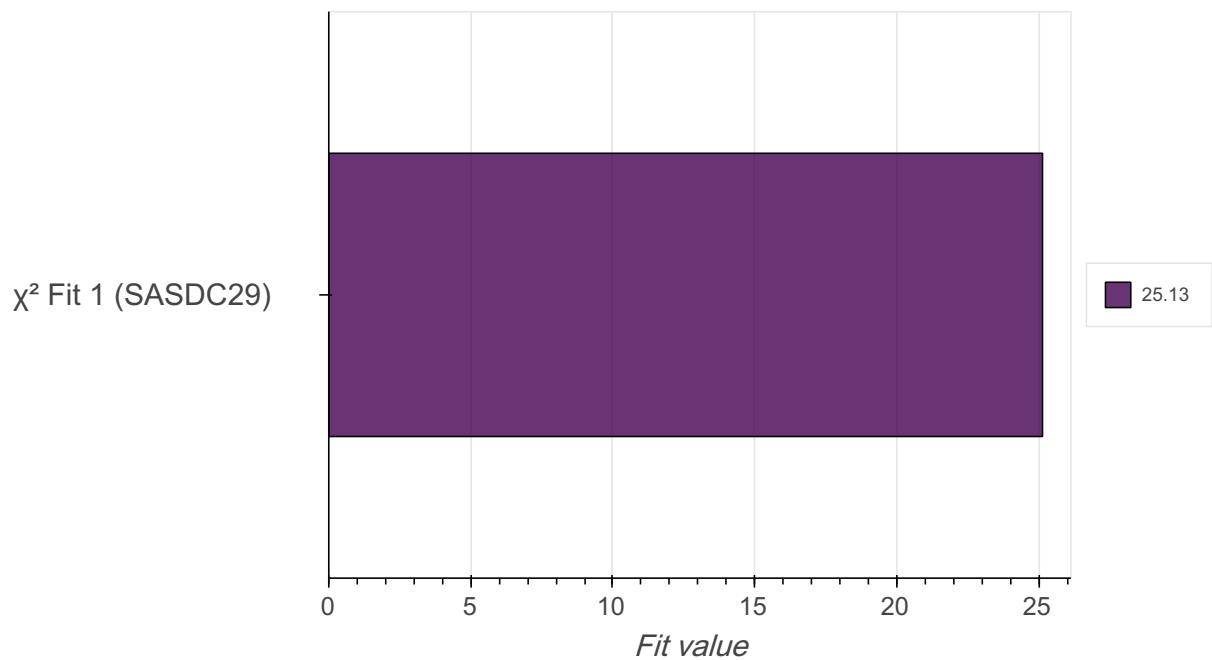
Model Quality: MolProbity Analysis



Data Quality for SAS: Rg Analysis



Fit to SAS Data: χ^2 Fit



Ensemble information ?

This entry consists of 0 distinct ensemble(s).

Summary ?

This entry consists of 1 unique models, with 4 subunits in each model. A total of 3 datasets or restraints were used to build this entry. Each model is represented by 0 rigid bodies and 4 flexible or non-rigid units.

Entry composition ?

There is 1 unique type of models in this entry. This model is titled None/Best scoring model, N/C-termini built.

| Model ID | Subunit number | Subunit ID | Subunit name | Chain ID | Chain ID [auth] | Total residues |
|----------|----------------|------------|--------------|----------|-----------------|----------------|
| 1 | 1 | 1 | Rev7-monomer | A | A | 212 |
| 1 | 2 | 1 | Rev7-monomer | C | C | 212 |
| 1 | 3 | 2 | Rev3-RBM2 | B | B | 28 |
| 1 | 4 | 2 | Rev3-RBM2 | D | D | 28 |

Datasets used for modeling ?

There are 3 unique datasets used to build the models in this entry.

| ID | Dataset type | Database name | Data access code |
|----|--------------------|---------------|------------------------|
| 1 | SAS data | SASBDB | SASDC29 |
| 2 | Experimental model | PDB | 6BC8 |
| 3 | Mutagenesis data | File | 10.5281/zenodo.1323686 |

Representation ?

This entry has only one representation and includes 0 rigid bodies and 4 flexible units

| Chain ID | Rigid bodies | Non-rigid segments |
|----------|--------------|--------------------|
| A | - | 1-212 |
| B | - | 1-28 |
| C | - | 1-212 |
| D | - | 1-28 |

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 | None | None | None | None | False | False |

There is 1 software package reported in this entry.

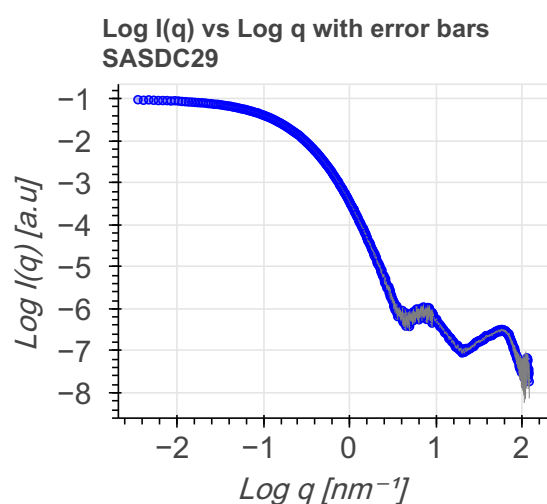
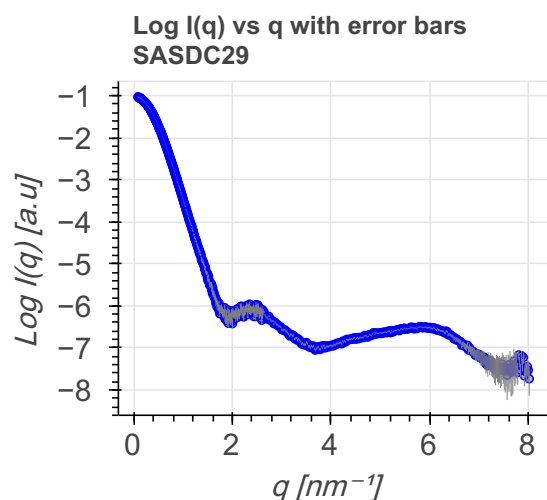
| ID | Software name | Software version | Software classification | Software location |
|----|-------------------------|------------------|-------------------------|---|
| 1 | HADDOCK | Not available | model building | http://haddock.science.uu.nl/services/HADDOCK/ |

Data quality ?

Scattering profile ?

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

Scattering profile for [SASDC29](#): 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 |
|----------|-------------------------|-------------|---------------------------|
| SASDC29 | 53.4 kDa | 53.4 kDa | 1.70 nm ³ /kDa |

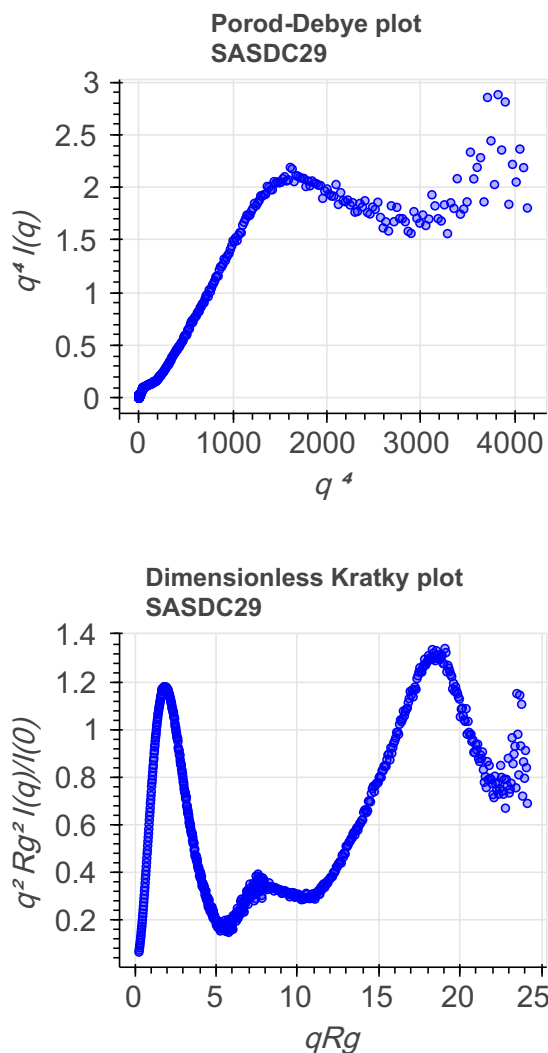
Volume estimates from experiments and analysis estimated volume can be compared to Porod volume obtained from scattering profiles.

| SASDB ID | Estimated Volume | Porod Volume | Specific Volume | Sample Contrast | Sample Concentration |
|----------|------------------|------------------------|-----------------|-----------------|----------------------|
| SASDC29 | N/A | 108.40 nm ³ | N/A | N/A | 10.60 mg/mL |

Flexibility analysis ?

Flexibility analysis for SASDC29: 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.

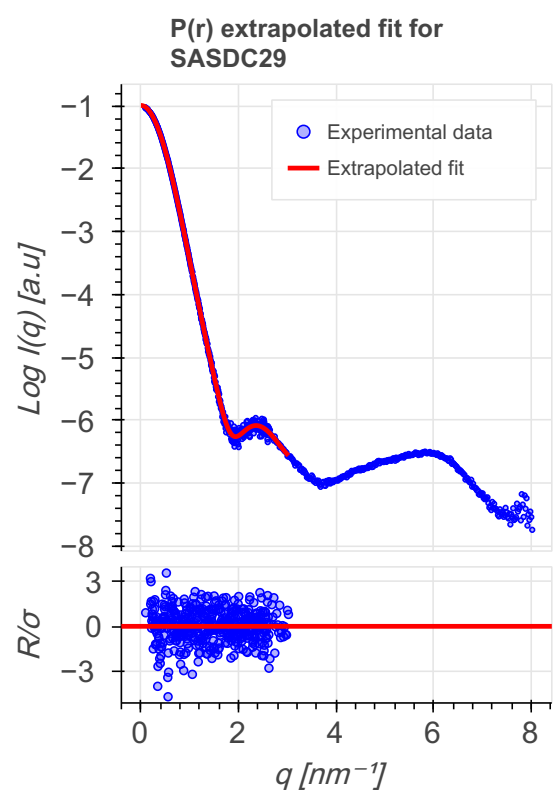
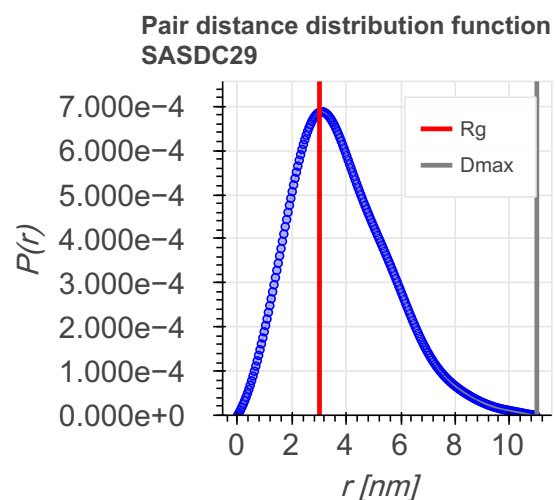


Pair-distance distribution analysis

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 | R_g | R_g error |
|----------|---------------|-----------|------------|----------|-------------|
| SASDC29 | GNOM 5.0 | 11.000 nm | N/A | 3.010 nm | 0.004 nm |

P(r) for SASDC29: The value of P(r) should be zero beyond $r=D_{max}$.



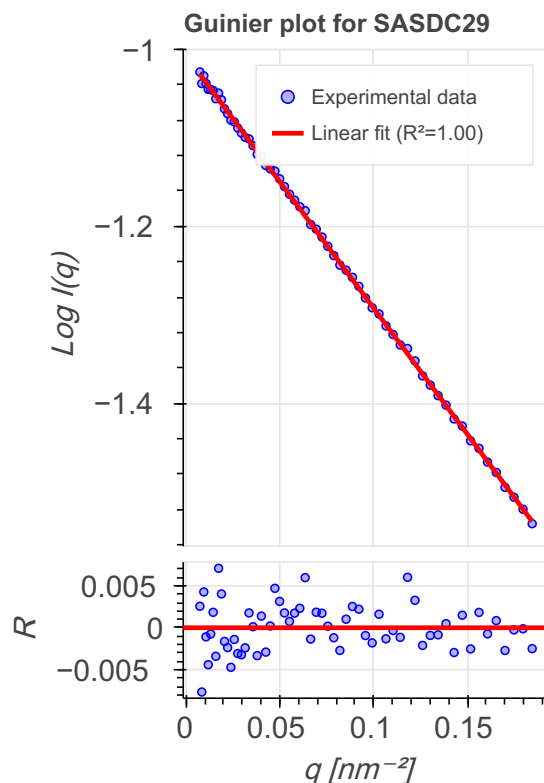
Guinier analysis ?

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 | R_g | R_g error | MW | MW error |
|----------|---------|-------------|----------|----------|
| SASDC29 | 2.93 nm | 0.00 nm | 53.4 kDa | 0.0 kDa |

Guinier analysis for SASDC29: 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.



Mutagenesis

Validation for this section is under development.

Model quality ?

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

Standard geometry: bond outliers ?

There are 123 bond outliers in this entry. A summary is provided below, and a detailed list of outliers can be found [here](#).

| Bond type | Observed distance (Å) | Ideal distance (Å) | Number of outliers |
|-----------|-----------------------|--------------------|--------------------|
| CA--HA3 | 1.05 | 0.97 | 3 |
| NZ--HZ3 | 1.01 | 0.89 | 2 |
| OG1--HG1 | 0.96 | 0.84 | 2 |
| N--H1 | 1.08 | 0.96 | 4 |

| Bond type | Observed distance (Å) | Ideal distance (Å) | Number of outliers |
|-----------|-----------------------|--------------------|--------------------|
| OG--HG | 0.96 | 0.84 | 2 |
| CG--HG3 | 1.10 | 0.97 | 6 |
| CG--HG2 | 1.10 | 0.97 | 2 |
| CE--HE3 | 1.10 | 0.97 | 4 |
| CA--HA | 1.10 | 0.97 | 8 |
| CE--HE2 | 1.10 | 0.97 | 4 |
| CE--HE1 | 1.10 | 0.97 | 3 |
| NZ--HZ2 | 1.02 | 0.89 | 2 |
| CA--HA2 | 1.10 | 0.97 | 2 |
| CG2--HG22 | 1.10 | 0.97 | 1 |
| CG2--HG22 | 1.11 | 0.97 | 1 |
| CA--HA | 1.11 | 0.97 | 6 |
| CB--HB2 | 1.11 | 0.97 | 12 |
| CG--HG3 | 1.11 | 0.97 | 2 |
| CE--HE1 | 1.11 | 0.97 | 1 |
| CA--HA2 | 1.11 | 0.97 | 2 |
| CG2--HG21 | 1.11 | 0.97 | 2 |
| CD--HD3 | 1.11 | 0.97 | 2 |
| CG2--HG23 | 1.11 | 0.97 | 2 |
| CB--HB3 | 1.11 | 0.97 | 12 |
| CG--HG2 | 1.11 | 0.97 | 6 |
| NZ--HZ1 | 1.03 | 0.89 | 2 |
| CD--HD2 | 1.11 | 0.97 | 2 |

| Bond type | Observed distance (Å) | Ideal distance (Å) | Number of outliers |
|-----------|-----------------------|--------------------|--------------------|
| CE--HE2 | 1.11 | 0.97 | 2 |
| CE--HE3 | 1.11 | 0.97 | 2 |
| CB--HB | 1.11 | 0.97 | 2 |
| CE1--HE1 | 1.08 | 0.93 | 2 |
| CD2--HD2 | 1.09 | 0.93 | 2 |
| NE2--HE2 | 1.03 | 0.86 | 2 |
| N--H | 1.08 | 0.86 | 14 |

Standard geometry: angle outliers

There are 44 angle outliers in this entry. A summary is provided below, and a detailed list of outliers can be found [here](#).

| Angle type | Observed angle (°) | Ideal angle (°) | Number of outliers |
|----------------|--------------------|-----------------|--------------------|
| ND1-CE1-NE2 | 108.40 | 118.08 | 1 |
| ND1-CE1-NE2 | 108.40 | 118.05 | 1 |
| C-N-C-N-C-N-CA | 121.70 | 106.28 | 1 |
| C-N-CA | 121.70 | 106.29 | 1 |
| C-N-CA | 121.70 | 106.30 | 2 |
| C-N-CA | 121.70 | 106.31 | 2 |
| C-N-CA | 121.70 | 106.32 | 1 |
| C-N-CA | 121.70 | 106.33 | 2 |
| C-N-CA | 121.70 | 106.35 | 1 |
| C-N-CA | 121.70 | 106.36 | 1 |
| C-N-CA | 121.70 | 106.37 | 1 |
| C-N-CA | 121.70 | 109.48 | 1 |
| C-N-C-N-C-N-CA | 121.70 | 109.51 | 1 |

| Angle type | Observed angle (°) | Ideal angle (°) | Number of outliers |
|-------------|--------------------|-----------------|--------------------|
| CB-CG-CD2 | 131.20 | 122.65 | 1 |
| CB-CG-CD2 | 131.20 | 122.73 | 1 |
| CD2-NE2-CE1 | 109.00 | 102.59 | 1 |
| CD2-NE2-CE1 | 109.00 | 102.61 | 1 |
| CG-ND1-CE1 | 109.30 | 99.85 | 1 |
| CG-ND1-CE1 | 109.30 | 99.88 | 1 |
| ND1-CG-CD2 | 106.10 | 111.47 | 2 |
| CG-CD-NE | 112.00 | 102.05 | 1 |
| CG-CD-NE | 112.00 | 102.09 | 1 |
| CA-CB-CG | 114.10 | 122.52 | 1 |
| CA-CB-CG | 114.10 | 122.50 | 1 |
| CA-CB-CG | 114.10 | 105.85 | 1 |
| CA-CB-CG | 114.10 | 105.87 | 1 |
| CA-N-H | 126.77 | 114.00 | 1 |
| CA-N-H | 126.78 | 114.00 | 2 |
| CA-N-H | 126.80 | 114.00 | 1 |
| CA-N-H | 126.81 | 114.00 | 3 |
| CA-N-H | 126.83 | 114.00 | 1 |
| CA-N-H | 126.84 | 114.00 | 1 |
| CA-N-H | 126.85 | 114.00 | 2 |
| CA-N-H | 126.87 | 114.00 | 2 |
| CA-N-H | 126.91 | 114.00 | 1 |

Too-close contacts

The following all-atom clashscore is based on a MolProbity analysis. All-atom clashscore is defined as the number of

clashes found per 1000 atoms (including hydrogen atoms). The table below contains clashscores for all the models in this entry.

| Model ID | Clash score | Number of clashes |
|----------|-------------|-------------------|
| 1 | 8.73 | 69 |

All 69 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

| Model ID | Atom-1 | Atom-2 | Clash overlap (Å) |
|----------|----------------|----------------|-------------------|
| 1 | A:47:LYS:NZ | C:140:HIS:NE2 | 1.297 |
| 1 | A:47:LYS:NZ | C:140:HIS:CE1 | 1.193 |
| 1 | A:47:LYS:HZ1 | C:140:HIS:CE1 | 1.178 |
| 1 | A:133:VAL:HB | C:133:VAL:HG12 | 1.071 |
| 1 | A:33:TYR:OH | C:40:VAL:HB | 1.048 |
| 1 | A:47:LYS:HZ1 | C:140:HIS:CD2 | 1.026 |
| 1 | A:133:VAL:HG11 | C:133:VAL:HA | 0.949 |
| 1 | A:133:VAL:CB | C:133:VAL:HG12 | 0.870 |
| 1 | A:133:VAL:CG1 | C:133:VAL:HA | 0.858 |
| 1 | A:47:LYS:CE | C:140:HIS:CE1 | 0.826 |
| 1 | A:36:GLU:OE2 | C:45:LYS:CB | 0.789 |
| 1 | A:36:GLU:OE2 | C:45:LYS:HB2 | 0.765 |
| 1 | A:47:LYS:HE2 | C:140:HIS:CE1 | 0.737 |
| 1 | A:47:LYS:HZ3 | C:140:HIS:CE1 | 0.726 |
| 1 | A:133:VAL:HB | C:133:VAL:CG1 | 0.720 |
| 1 | A:133:VAL:CG2 | C:133:VAL:HG12 | 0.719 |
| 1 | A:45:LYS:HE2 | C:36:GLU:HB3 | 0.679 |
| 1 | A:36:GLU:OE2 | C:45:LYS:HB3 | 0.670 |
| 1 | A:45:LYS:HE2 | C:36:GLU:CB | 0.637 |

| Model ID | Atom-1 | Atom-2 | Clash overlap (Å) |
|----------|----------------|----------------|-------------------|
| 1 | C:44:GLN:CD | C:57:BCYS:SG | 0.619 |
| 1 | A:44:GLN:CD | A:57:BCYS:SG | 0.618 |
| 1 | A:49:TYR:HA | A:121:GLU:HG3 | 0.556 |
| 1 | C:49:TYR:HA | C:121:GLU:HG3 | 0.548 |
| 1 | A:133:VAL:CB | C:133:VAL:HA | 0.547 |
| 1 | A:36:GLU:CD | C:45:LYS:HB3 | 0.526 |
| 1 | A:33:TYR:CE2 | C:33:TYR:OH | 0.510 |
| 1 | A:133:VAL:HG12 | C:132:SER:O | 0.491 |
| 1 | A:33:TYR:HE2 | C:33:TYR:OH | 0.489 |
| 1 | A:116:LEU:HD23 | A:197:ILE:HD11 | 0.466 |
| 1 | A:33:TYR:HH | C:40:VAL:HB | 0.458 |
| 1 | A:135:ASP:O | C:52:PRO:HG2 | 0.458 |
| 1 | C:116:LEU:HD23 | C:197:ILE:HD11 | 0.454 |
| 1 | A:24:LEU:O | A:28:VAL:HG23 | 0.453 |
| 1 | C:24:LEU:O | C:28:VAL:HG23 | 0.450 |
| 1 | A:193:MET:HB3 | A:193:MET:HE2 | 0.439 |
| 1 | C:98:LYS:O | C:204:VAL:HA | 0.434 |
| 1 | A:98:LYS:O | A:204:VAL:HA | 0.433 |
| 1 | C:158:THR:H | C:161:MET:HE2 | 0.422 |
| 1 | A:158:THR:H | A:161:MET:HE2 | 0.417 |
| 1 | C:79:ASN:O | C:154:ARG:HD3 | 0.415 |
| 1 | A:168:LYS:O | A:169:ASP:HB2 | 0.414 |
| 1 | A:79:ASN:O | A:154:ARG:HD3 | 0.413 |

| Model ID | Atom-1 | Atom-2 | Clash overlap (Å) |
|----------|----------------|----------------|-------------------|
| 1 | A:33:TYR:OH | C:33:TYR:HE1 | 0.413 |
| 1 | A:133:VAL:HG23 | C:133:VAL:HG12 | 0.408 |
| 1 | C:168:LYS:O | C:169:ASP:HB2 | 0.406 |

Torsion angles: Protein backbone ?

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

| Model ID | Analysed | Favored | Allowed | Outliers |
|----------|----------|---------|---------|----------|
| 1 | 482 | 472 | 8 | 2 |

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 | Analysed | Favored | Allowed | Outliers |
|----------|----------|---------|---------|----------|
| 1 | 454 | 416 | 22 | 16 |

Detailed list of outliers are tabulated below.

| Model ID | Chain | Residue ID | Residue type |
|----------|-------|------------|--------------|
| 1 | A | 4 | THR |
| 1 | A | 111 | ILE |
| 1 | A | 121 | GLU |
| 1 | A | 159 | ARG |
| 1 | A | 169 | ASP |
| 1 | A | 209 | HIS |
| 1 | B | 4 | LYS |
| 1 | B | 25 | GLN |
| 1 | C | 4 | THR |
| 1 | C | 111 | ILE |

| Model ID | Chain | Residue ID | Residue type |
|----------|-------|------------|--------------|
| 1 | C | 121 | GLU |
| 1 | C | 159 | ARG |
| 1 | C | 169 | ASP |
| 1 | C | 209 | HIS |
| 1 | D | 4 | LYS |
| 1 | D | 25 | GLN |

Fit of model to data used for modeling ?

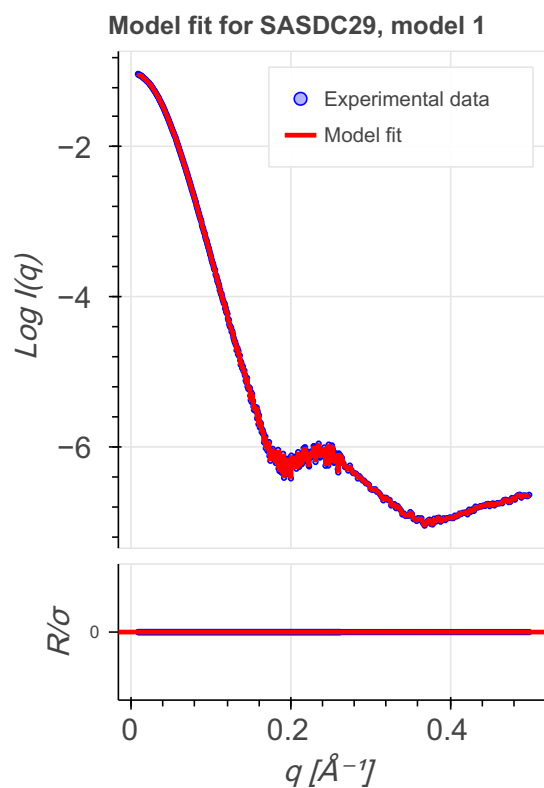
Fit of model(s) to SAS data

χ^2 goodness of fit and cormap analysis ?

Model and fits displayed below were obtained from SASBDB. χ^2 values are a measure of fit of the model to data. A perfect fit has a χ^2 value of 1.0. ATSAS datcmp was used for hypothesis testing. All data sets are similar (i.e. the fit and the data collected) is the null hypothesis. p-value is a measure of evidence against the null hypothesis, smaller the value, the stronger the evidence that you should reject the null hypothesis.

| SASDB ID | Model | χ^2 | p-value |
|----------|-------|----------|----------|
| SASDC29 | 1 | 25.13 | 0.00E+00 |

Model fit for SASDC29 (fit/model number 1): Residual value plot is a measure to assess fit to the data. Residual values should be equally and randomly spaced around the horizontal axis.



Mutagenesis

Validation for this section is under development.

Fit of model to data used for validation ?

Validation for this section is under development.

Acknowledgements

Development of integrative model validation metrics, implementation of a model validation pipeline, and creation of a validation report for integrative structures, are funded by NSF ABI awards (DBI-1756248, DBI-2112966, DBI-2112967, DBI-2112968, and DBI-1756250). The [PDB-Dev team](#) and members of [Sali lab](#) contributed model validation metrics and software packages.

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

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