wwPDB X-ray Structure Validation Summary Report (i)

Jun 17, 2024 – 02:56 PM EDT

PDB ID	:	3BIJ
Title	:	Crystal structure of protein GSU0716 from Geobacter sulfurreducens. Northeast Structural Genomics target GsR13
Authors	:	Forouhar, F.; Neely, H.; Su, M.; Seetharaman, J.; Benach, J.; Conover, K.; Fang, Y.; Xiao, R.; Owen, L.A.; Maglaqui, M.; Cunningham, K.; Baran, M.C.; Acton, T.B.; Montelione, G.T.; Tong, L.; Hunt, J.F.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2007-11-30
Resolution	:	2.50 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

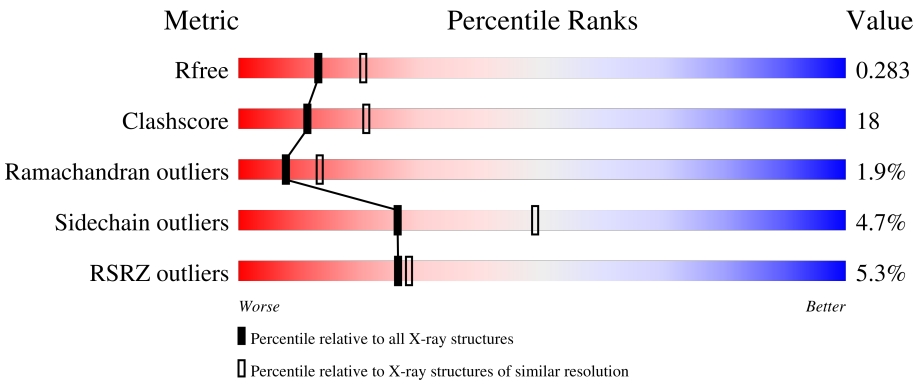
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.50 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	285	5% 61% 28% • 7%
1	B	285	3% 59% 30% • 8%
1	C	285	6% 61% 29% • 7%

Validation Pipeline (wwPDB-VP) : 2.37.1

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 6200 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Uncharacterized protein GSU0716.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	265	2028	1286	352	379	4	7	0	0	0
1	B	262	2007	1273	349	374	4	7	0	0	0
1	C	264	2024	1284	351	378	4	7	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	278	LEU	-	EXPRESSION TAG	UNP Q74F93
A	279	GLU	-	EXPRESSION TAG	UNP Q74F93
A	280	HIS	-	EXPRESSION TAG	UNP Q74F93
A	281	HIS	-	EXPRESSION TAG	UNP Q74F93
A	282	HIS	-	EXPRESSION TAG	UNP Q74F93
A	283	HIS	-	EXPRESSION TAG	UNP Q74F93
A	284	HIS	-	EXPRESSION TAG	UNP Q74F93
A	285	HIS	-	EXPRESSION TAG	UNP Q74F93
B	278	LEU	-	EXPRESSION TAG	UNP Q74F93
B	279	GLU	-	EXPRESSION TAG	UNP Q74F93
B	280	HIS	-	EXPRESSION TAG	UNP Q74F93
B	281	HIS	-	EXPRESSION TAG	UNP Q74F93
B	282	HIS	-	EXPRESSION TAG	UNP Q74F93
B	283	HIS	-	EXPRESSION TAG	UNP Q74F93
B	284	HIS	-	EXPRESSION TAG	UNP Q74F93
B	285	HIS	-	EXPRESSION TAG	UNP Q74F93
C	278	LEU	-	EXPRESSION TAG	UNP Q74F93
C	279	GLU	-	EXPRESSION TAG	UNP Q74F93
C	280	HIS	-	EXPRESSION TAG	UNP Q74F93
C	281	HIS	-	EXPRESSION TAG	UNP Q74F93
C	282	HIS	-	EXPRESSION TAG	UNP Q74F93
C	283	HIS	-	EXPRESSION TAG	UNP Q74F93
C	284	HIS	-	EXPRESSION TAG	UNP Q74F93

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	285	HIS	-	EXPRESSION TAG	UNP Q74F93

- Molecule 2 is water.

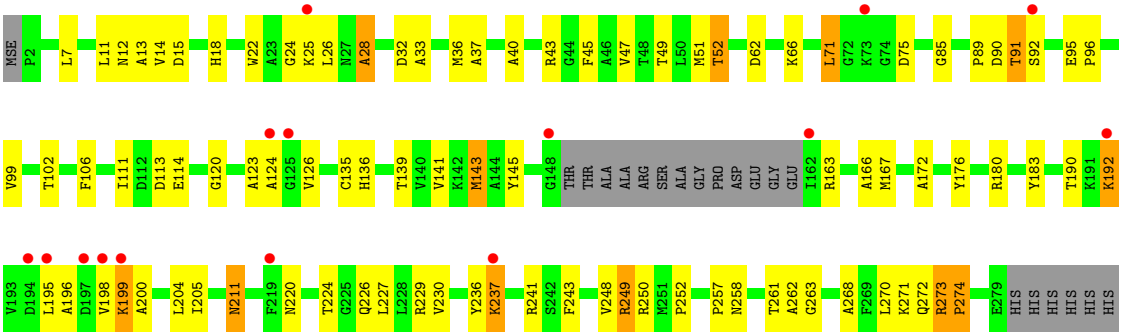
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	43	Total O 43 43	0	0
2	B	63	Total O 63 63	0	0
2	C	35	Total O 35 35	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

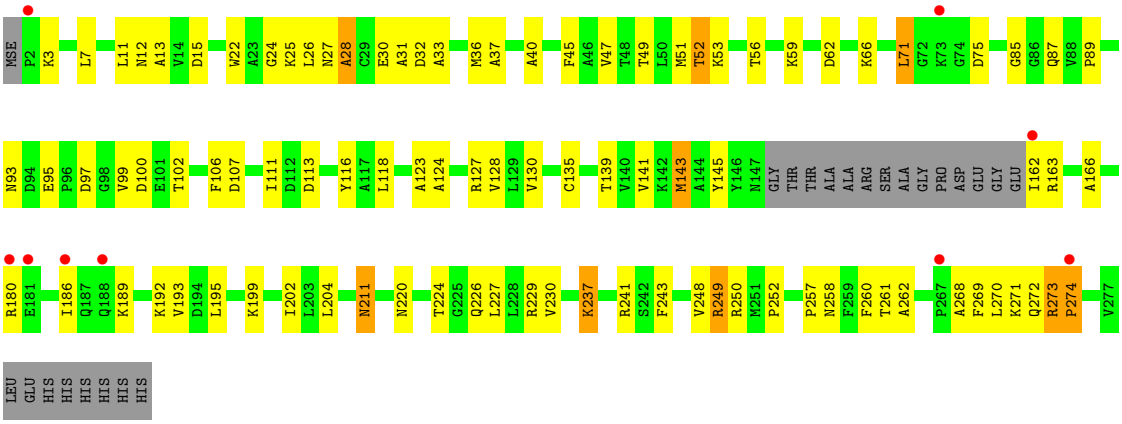
• Molecule 1: Uncharacterized protein GSU0716

Chain A:



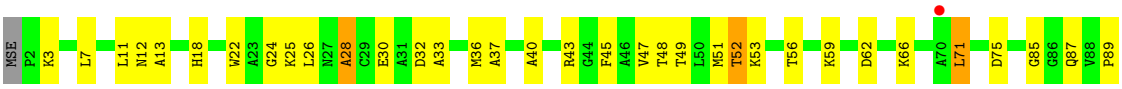
• Molecule 1: Uncharacterized protein GSU0716

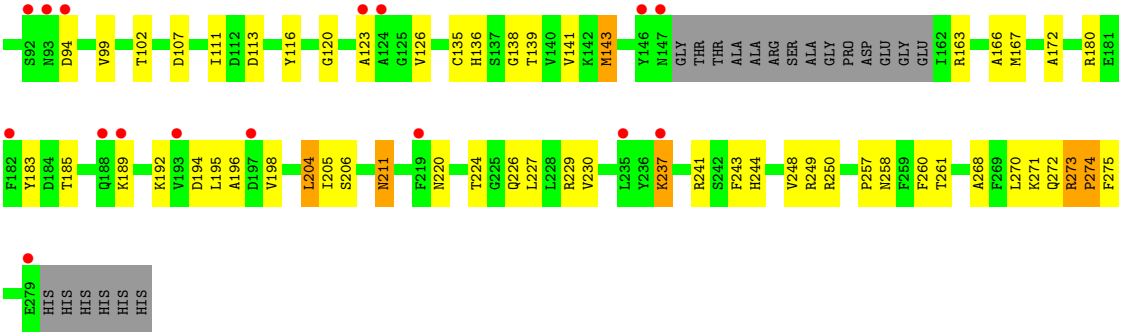
Chain B:



• Molecule 1: Uncharacterized protein GSU0716

Chain C:




S92
N93
D94
V99
T102
D107
I111
D112
D113
Y116
G120
A123
A124
G125
V126
C135
H136
S137
G138
T139
V140
V141
K142
M143
Y146
N147
GLY
THR
THR
ALA
ALA
ARG
SER
ALA
GLY
PRO
ASP
GLU
GLY
GLU
I162
R163
A166
M167
A172
R180
E181
F182
Y183
D184
T185
Q188
K189
K192
V193
D194
L195
A196
D197
V198
L204
I205
S206
N211
F219
N220
T224
G225
Q226
L227
L228
R229
V230
L235
Y236
K237
R241
S242
F243
H244
V248
R249
R250
P257
N258
F259
F260
T261
A268
F269
L270
K271
Q272
R273
P274
F275
E279
HIS
HIS
HIS
HIS
HIS
HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 4 2 2	Depositor
Cell constants a, b, c, α, β, γ	121.72Å 121.72Å 219.04Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	19.98 – 2.50 29.77 – 2.50	Depositor EDS
% Data completeness (in resolution range)	86.0 (19.98-2.50) 92.2 (29.77-2.50)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	5.64 (at 2.51Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.207 , 0.267 0.225 , 0.283	Depositor DCC
R_{free} test set	5383 reflections (9.89%)	wwPDB-VP
Wilson B-factor (Å^2)	46.8	Xtriage
Anisotropy	0.192	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 54.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	6200	wwPDB-VP
Average B, all atoms (Å^2)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 11.73% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality (i)

5.1 Standard geometry (i)

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\#\|Z\|>5$	RMSZ	$\#\|Z\|>5$
1	A	0.42	0/2063	1.10	9/2775 (0.3%)
1	B	0.40	0/2042	0.74	7/2747 (0.3%)
1	C	0.40	0/2059	0.73	6/2770 (0.2%)
All	All	0.41	0/6164	0.87	22/8292 (0.3%)

There are no bond length outliers.

The worst 5 of 22 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	180	ARG	NE-CZ-NH1	-19.67	110.47	120.30
1	A	241	ARG	NE-CZ-NH1	-18.84	110.88	120.30
1	A	273	ARG	NE-CZ-NH2	-18.79	110.90	120.30
1	A	241	ARG	NE-CZ-NH2	18.34	129.47	120.30
1	A	273	ARG	NE-CZ-NH1	18.00	129.30	120.30

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts (i)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2028	0	2007	74	0
1	B	2007	0	1987	76	0
1	C	2024	0	2004	73	0
2	A	43	0	0	2	0
2	B	63	0	0	5	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	35	0	0	2	0
All	All	6200	0	5998	215	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 18.

The worst 5 of 215 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:12:ASN:HB2	1:B:26:LEU:O	1.66	0.96
1:A:12:ASN:HB2	1:A:26:LEU:O	1.66	0.95
1:C:12:ASN:HB2	1:C:26:LEU:O	1.64	0.95
1:C:36:MSE:HE1	1:C:227:LEU:HD22	1.62	0.80
1:A:36:MSE:HE1	1:A:227:LEU:HD22	1.65	0.77

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	A	261/285 (92%)	246 (94%)	10 (4%)	5 (2%)	8 13
1	B	258/285 (90%)	244 (95%)	9 (4%)	5 (2%)	8 13
1	C	260/285 (91%)	241 (93%)	14 (5%)	5 (2%)	8 13
All	All	779/855 (91%)	731 (94%)	33 (4%)	15 (2%)	8 13

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	52	THR
1	B	52	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	189	LYS
1	B	274	PRO
1	C	52	THR

5.3.2 Protein sidechains (i)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	207/214 (97%)	196 (95%)	11 (5%)	22	43
1	B	205/214 (96%)	197 (96%)	8 (4%)	32	57
1	C	207/214 (97%)	197 (95%)	10 (5%)	25	48
All	All	619/642 (96%)	590 (95%)	29 (5%)	26	49

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	143	MSE
1	C	249	ARG
1	B	249	ARG
1	C	204	LEU
1	B	237	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 30 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	169	GLN
1	C	226	GLN
1	B	226	GLN
1	C	258	ASN
1	C	169	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å^2)	Q<0.9
1	A	258/285 (90%)	0.26	15 (5%)	23	24	32, 48, 73, 93	0
1	B	255/285 (89%)	0.12	9 (3%)	44	47	31, 43, 65, 87	0
1	C	257/285 (90%)	0.37	17 (6%)	18	19	34, 54, 80, 93	0
All	All	770/855 (90%)	0.25	41 (5%)	26	28	31, 49, 73, 93	0

The worst 5 of 41 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	188	GLN	3.6
1	A	192	LYS	3.5
1	C	146	TYR	3.5
1	A	195	LEU	3.4
1	C	70	ALA	3.4

6.2 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.